Re-parameterized Weibull distribution for modeling metocean extremes of multiple hazards with the Rosenblatt transformation

C Qiao* and A T Myers

1 Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA
* Email: qiao.c@husky.neu.edu

Abstract. The construction of environmental contours using the Rosenblatt transformation requires the parameters defining the conditional distribution to be determined for all values of the conditioned random variable, including extreme values which are extrapolated from data in most cases. In the typical case where the conditional distribution is described by many parameters, the fitting process of these parameters is cumbersome, and the resultant extrapolation is prone to error. A re-parameterization technique is proposed here to approximate such a conditional distribution, which depends on many parameters, with a two-parameter Weibull distribution, so that fewer distribution parameters are involved in the fitting and extrapolation processes. This technique is shown to have a better performance compared to conventional approaches, as demonstrated here for a numerical example involving the construction of a 50-year environmental contour for wind and wave at offshore locations.

1. Introduction

Environmental contours are important for designing offshore structures to characterize the metocean conditions for multiple correlated hazards (e.g., wind speed, wave height, etc.) and to associate these conditions with a mean return period (MRP). For example, in offshore wind turbine design standards, environmental contours are used to describe combinations of the 1-hour averaged mean wind speeds and significant wave heights with a 50-year MRP. A typical wind-wave contour is given in Figure 1, with the part that is of engineering interest indicated using a solid curve. For a given wind speed, structural demand increases as the wave height increases, so the lower value of the wave height (indicated using the dashed curve) is not considered in the structure design; while for a given wave height, both moderate and extreme values of wind speed need to be considered because the aerodynamic shapes of the rotor blades vary drastically for different wind speeds (e.g., if the rotor is operational or feathered). As such, the contour for design purposes (i.e., the blue solid curve in Figure 1) depends on both moderate and extreme values of wind speed, but only on extreme values of significant wave height conditioned on wind speed.

The Inverse First Order Reliability Method (IFORM) [1] is a popular approach to construct such a contour, and an important step is to transform the joint probability distribution of metocean conditions into an uncorrelated standard joint normal distribution with the same dimension. A common way to achieve this is the Rosenblatt transformation [2], which requires an accurate estimation at the tail of
each distribution. Extreme Value Theory (EVT) [3, 4] is well-established to estimate the tail behavior of an independent and identically distributed (IID) random variable, and an innovative technique is proposed by Qiao et al. (the manuscript is under review) to extend EVT to continuous dependent data, from which environmental contours are constructed. In this technique, the tail of the dependent data is described by six parameters (three parameters for the Generalized Extreme Value (GEV) distribution fit to the extreme values and three parameters for the MRP ratio curve describing the relationship of the tails between the dependent data and the corresponding extreme values). For the Rosenblatt transformation, the six distribution parameters of one variable need to be fit as functions of the variable on which it is conditioned, increasing the number of parameters by multiple-fold, depending on the complexity of the fitting functions. The fitting of these parameters is cumbersome and can introduce errors in the resulting contour if the fitting functions are not chosen properly and extrapolated accurately. A solution is provided in this paper to improve the fitting process. The idea is to re-parameterize the tail of the conditional distribution using a two-parameter distribution so that fewer parameters are required for extrapolation. Compared to other distribution fitting techniques (e.g., Maximum Likelihood), this method discards all the information of low values to ensure a good fit for the conditional distribution at extreme values, and thus is suitable for constructing environmental contours where the random variable modeled with the marginal distribution must be modeled fully (wind speed in the case of Figure 1) and the random variable modeled with the conditional distribution need only be modeled in the upper tail (significant wave height in the case of Figure 1).

Section 2 provides background on IFORM, the Rosenblatt transformation, and the technique proposed by Qiao et al. for extrapolating the tail of dependent data. The details of the re-parameterization of the tail are provided in Section 3. Numerical examples, which compare this technique with conventional approaches, are given in Section 4, and conclusions are summarized in Section 5.

2. Background

The $N$-year value of a single random variable $X$ is calculated through the definition of MRP as,

$$X_N = F^{-1}\left(1 - \frac{1}{\nu \cdot N}\right)$$

(1)

where $\nu$ is the annual occurrence rate, and $F$ denotes the cumulative distribution function (CDF). IFORM [1] calculates combinations of multiple random variables with the same MRP (i.e., the environmental contour) by transforming the joint probability distribution into an uncorrelated standard joint normal distribution with the same dimension, through which Equation (1) is extended to the multivariate case as a hypersphere (or a circle for the two-dimensional case) with radius $\beta = \Phi^{-1}\left(1 - 1/(\nu \cdot N)\right)$, where $\Phi^{-1}$ denotes the inverse CDF of the normal distribution. This hypersphere is then transformed back to the original variables and the result is the so-called environmental contour.

The Rosenblatt transformation [2] is a popular approach for the implementation of IFORM. For example, consider a dataset with measurements of the hourly averaged wind speed $V$ and the hourly significant wave height $H_S$; Pairs of $(V, H_S)$ are transformed into $(U_1, U_2)$ which follow an uncorrelated standard joint normal distribution through $U_1 = \Phi^{-1}\left(F_V(V)\right)$ and $U_2 = \Phi^{-1}\left(F_{H_S|V}(H_S)\right)$, i.e., $V$ is

![Figure 1. A typical environmental contour for averaged wind speed and significant wave height. The solid curve indicates the part that is of engineering interest, and the dashed curve indicates the part that is not considered in the structure design. The red and green ×'s indicate independent MRP values of wind and wave respectively.](image-url)
transformed independently and $H_g$ is transformed conditioned on $V$. As such, for a given wind speed $V_0$, the corresponding $N$-year conditional wave height is expressed as,

$$H_{s,N|V_0} = F_{H_s|V_0}^{-1}\left(\beta^2 - \left(\Phi^{-1}(F_{V_0})\right)^2\right)$$  

(2)

The independent $N$-year wind speed (i.e., the MRP value obtained using the marginal wind speed distribution) corresponds to the right apex of the $N$-year environmental contour (the red × in Figure 1). And because this transformation is independent of the order of the transformation (i.e., transforming $H_g$ first and then $V|H_g$ yields the same result), the top apex of the environmental contour corresponds to the independent $N$-year wave height (the green × in Figure 1).

To obtain an accurate environmental contour, $F_V$ and $F_{H_s|V}$ need to be estimated accurately for a large range of quantiles. This is particularly challenging for the tail, where quantile estimations are significantly influenced by sampling error and EVT cannot be directly applied because IID is not satisfied. In a manuscript currently under review, a method is proposed by Qiao et al. to extrapolate the tail of a continuous dependent dataset $X_c$. This method converts the EVT result, which characterizes the tail behaviour of the corresponding annual maxima dataset $X_m$ using the GEV distribution, back to represent the tail behaviour of $X_c$. The conversion is achieved through the MRP ratio defined as $R(x) = MRP_c(x)/MRP_m(x)$, where $MRP_c(x)$ and $MRP_m(x)$ denote the MRP as functions of $X_c$ and $X_m$. $R(x)$ is first calculated from the empirical MRP values of the two datasets and then extrapolated through curve fitting. As such, the tail of $X_c$ is expressed as,

$$F_c(x) = 1 - \frac{1}{v_c} \left(1 - F_m(x; \xi, \mu, \sigma)\right) \cdot R(x; a, b, c)$$  

(3)

where $F_m$ follows the GEV distribution according to the EVT, and $R$ is approximated by a function with three parameters $a$, $b$, and $c$. For the rest part of $F_c(x)$, traditional approaches still provide an accurate estimation, and the results using the empirical CDF is adopted in this paper for simplicity.

An illustration of this method is given in Figure 2. The annual maxima are first extracted (red dots) and fit to the GEV distribution (red curves). The MRP ratio $R(x)$ between the hourly dataset (grey dots) and the annual maxima dataset is quantified, and the extrapolation results of the GEV distribution are then converted to the hourly data (blue dashed curves).

![Figure 2. Tail extrapolation of the hourly $H_g$ data conditioned on (a) $V = 5$ m/s, and (b) $V = 20$ m/s for an offshore site near Virginia following the method proposed by Qiao et al. Detailed procedure and notations are introduced in Section 2.](image)

3. Re-parameterization of the distribution tail

Following the Rosenblatt transformation, $F_V$ and $F_{H_s|V}$ need to be estimated from the metocean measurements. For high wind speeds, which is of interest for risk analysis, there are very few $H_g$ measurements, so the corresponding $F_{H_s|V_{high}}$ has to be extrapolated from $F_{H_s|V_{low}}$. Using Equation (3), $F_{H_s|V}$ is characterized by at least six parameters $\theta_{6\times1} = [\xi, \mu, \sigma, a, b, c]$ ($v_m$ is usually set in EVT as a
constant, and $v_c$ can be derived from $F_V$), each of which must be represented as a function of $V$, which makes the extrapolation process cumbersome and prone to error, because proper curve functions need to be carefully selected for each parameter, and extrapolation errors from any of the parameters would influence the resultant environmental contour.

One way to diminish parameter extrapolation errors of the conditional distribution is to approximate $F_{H_s|V}(H_s; \theta_{6\times1})$ with a simpler two-parameter Weibull distribution $\tilde{F}_{H_s|V}(H_s; \theta_{2\times1})$ using the Weibull chart [5]. The Weibull chart plots $\ln(-\ln(1-F))$ as the vertical coordinate and $\ln(x)$ as the horizontal coordinate. As such, a straight line in the Weibull chart indicates a goodness-of-fit using the Weibull distribution with slope $k$ and intercept $-k \cdot \ln(\lambda)$, where $k$ and $\lambda$ are the parameters to define the CDF of the Weibull distribution as $F(x) = 1 - \exp(-x/\lambda^k)$. In figure 3, where the hourly MRP results in Figure 2 are expressed using the Weibull chart, linearity is observed in both cases for relatively high wave heights, while $H_s|V(V = 5 \, m/s)$ deviates from this linear trend at low wave heights. Since the environmental contour considered here involves only the upper tail of the conditional wave height, re-parameterizing only the tail of $H_s|V$ using the Weibull distribution reduces the number of conditional distribution parameters $\theta$ and maintains a good fit. $k$ and $\lambda$ are estimated for each bin of wind speeds from the linear fitting results (shown in cyan in Figure 3), and thus only two parameters are to be fit as functions of $V$ instead of six parameters ($\xi, \mu, \sigma, a, b, c$).

It is shown in Figure 3 that the Weibull distribution can be used to re-parameterize $H_s|V$ as long as a linear tail in the Weibull chart exists for different values of $V$. Then the question arises on how to determine the range of the tail for fitting. This question becomes more urgent if the overall behavior of $H_s|V$ shows curvature (see Figure 3(a)) rather than linearity (see Figure 3(b)). As a rule of thumb, the range of quantiles involved in the construction of the environmental contour should be examined carefully for linearity. Specifically, the minimum quantile of $H_s|V$ is 0.5, corresponding to $F_V = 1 - 1/(v \cdot N)$ (i.e., the independent $N$-year wind speed), and the maximum quantile of $H_s|V$ is 1 − 1/(v \cdot N), corresponding to $F_V = 0.5$.

The re-parameterization introduced here is a general idea to reduce the number of parameters of a distribution. It is applied to the distribution tail in this paper because it serves the purpose of constructing environmental contours, which characterize the tail behaviour of the conditional distribution. This idea can be applied with large flexibility. For example, a different distribution other than the Weibull distribution can be selected to re-parameterize the distribution tail, and the tail can be estimated using other methods with a different set of parameters, or even estimated empirically.

![Figure 3](image_url)

**Figure 3.** The re-parameterization using the Weibull chart for the hourly MRP results of (a) $H_s|V(V = 5 \, m/s)$ and (b) $H_s|V(V = 20 \, m/s)$ in Figure 2. The lines in cyan are the linear fitting results, and the rest of the notations are the same as in Figure 2.
4. Numerical examples

A 37-year (1979-2015) hourly metocean hindcast is used here to illustrate the application of the re-parameterized Weibull distribution. The hourly averaged wind speed $V$ is at a 10-meter elevation and is taken from the Climate Forecast System Reanalysis (CFSR) [6, 7], and the hourly significant wave height $H_s$ is taken from a hindcast study [8] using the Mike 21 numerical metocean model driven by the CFSR wind. The 50-year environmental contour for the hourly $V$ and $H_s$ is constructed for three offshore sites, one near Massachusetts, one near New York, and one near Virginia (see Table 1). These locations are all within proposed areas for the development of offshore wind farms [9]. Due to the page limitation, only results for the Virginia site are shown in the following discussion, and the resultant contours for Massachusetts and New York sites are given in Figure 11.

To construct the 50-year hourly $V - H_s$ contour, $F_V$ and $F_{H_s|V}$ are estimated, as part of the Rosenblatt transformation. $F_{H_s|V}$ is estimated for $1 \text{ m/s} \leq V \leq 25 \text{ m/s}$ with 1 m/s intervals, and with the conditional $H_s$ data extracted using a wind speed window of 3 m/s. MRP results of $V$, $H_s$ are given in Figure 4, and the MRP result of $H_s|(V = 20 \text{ m/s})$ and $H_s|(V = 5 \text{ m/s})$ are given in Figure 2. The corresponding CDFs are then calculated from MRP results through Equation (1). The extrapolation of the tail for dependent data follows the method proposed by Qiao et al. (see Section 2).

The 50-year environmental contours using three approaches are considered in the rest of this section. A conventional implementation of the Rosenblatt transformation is given in Section 4.1; results using the Nataf transformation, a simpler transformation technique compared to the Rosenblatt transformation, are given in Section 4.2; and lastly, results using the re-parameterized Weibull distribution are given in Section 4.3. Note that $F_V$ used in these three approaches is the same, while the estimation of $F_{H_s|V}$ differs.

Two values are used here to validate the resultant 50-year contour. For low wind speed values, the 50-year conditional $H_s$ estimated directly from data (i.e., $H_s$ data within a certain range of $V$) using Equation (2) is a good reference for validation. For high wind speed values where the 50-year conditional $H_s$ is estimated from $F_{H_s|V}$ with its parameters $\theta(V)$ extrapolated, the independent 50-year $H_s$ estimated from the marginal $H_s$ distribution is a good reference, as the apex of the 50-year environmental contour should ideally match the independent 50-year $H_s$.

| State        | Lon (°W) | Lat (°N) | Water depth (m) |
|--------------|----------|----------|-----------------|
| Massachusetts| 70.46    | 40.89    | 50              |
| New York     | 73.53    | 40.37    | 26              |
| Virginia     | 75.42    | 36.89    | 27              |

Table 1. Site information for the numerical example

![Figure 4](image-url)  
**Figure 4.** The MRP results for (a) $V$ and (b) $H_s$ for the Virginia site. Notations are the same as in Figure 2.
4.1. Rosenblatt transformation using prescribed distributions

As a conventional implementation of the Rosenblatt transformation, a prescribed distribution is selected to fit \( F_{H,v} \), and then the parameters of that distribution are fit as functions of \( V \) and are extrapolated for high wind speeds where insufficient data are available to estimate \( F_{H,v} \) directly. In Johannessen et al. [10], \( H_v \) is assumed to follow the Weibull distribution, and two parameters are fit as
\[
\theta(V) = a_1 + a_2 \cdot V \\
\lambda(V) = b_1 + b_2 \cdot V^{b_3}
\]
respectively. In IEC 61400-3 [11], the log-normal distribution is suggested as a candidate to estimate \( H_v \). As such, both the Weibull and log-normal distributions are tested here to construct the 50-year contour. And each of the distribution parameters is fit as
\[
\theta(V) = c_1 + c_2 \cdot V^{c_3}
\]
where \( c_1, c_2, \) and \( c_3 \) are fitting coefficients.

The 50-year contours using the Weibull and log-normal distributions for \( H_v \) are shown in Figure 5, using blue and green curves respectively. The value of \( H_{5.50} \) is shown as a red dashed line and the unsmoothed \( H_{5.50} \) at 25 values of \( V \) shown as the red dots and are used here as references. It is interesting to observe that the contour assuming the Weibull distribution underestimates \( H_v \) for low values of \( V \) and becomes more accurate as \( V \) increases; while the contour assuming the log-normal distribution for \( H_v \) fits the unsmoothed \( H_{5.50} \) at low values of \( V \) well and overestimates \( H_{5.50} \) as \( V \) increases. The errors do not arise from fitting the distribution parameters, but rather are caused by the incapability of the distributions in representing the behavior \( H_v \) for a large range of wind speeds. To better understand the cause, MRP results are given for \( H_v (V = 5 \text{ m/s}) \) and \( H_v (V = 20 \text{ m/s}) \) in Figure 6, which clearly shows that at low wind speeds, \( H_v \) follows the log-normal distribution, while at high wind speeds, \( H_v \) follows the Weibull distribution.

Figure 5. The 50-year environmental contour for hourly \( V \) and \( H_v \) for the Virginia site assuming the Weibull distribution (blue curve) and the log-normal distribution (green curve) for \( H_v \) as part of the Rosenblatt transformation. Grey dots are the hourly data. The red dashed line indicates the independent \( H_{5.50} \). Red dots are the unsmoothed \( H_{5.50} \).

Figure 6. The MRP results for (a) \( H_v \) \( (V = 5 \text{ m/s}) \), and (b) \( H_v \) \( (V = 20 \text{ m/s}) \) in Figure 5. Grey dots are MRP values calculated empirically for each hourly \( H_v \). Blue curves are fit using a Weibull distribution. Green curves are fit using a log-normal distribution.

4.2. Nataf transformation

Compared to the Rosenblatt transformation, the Nataf transformation [12] provides a simpler way to implement IFORM. Specifically, each random variable is transformed independently using the marginal
distribution as \( Z_t = \Phi^{-1}(F_X(x)) \), and \( Z \) is then approximated using a correlated standard joint normal distribution. The cost of such a simplification is less accuracy, as the covariance matrix captures only the linear correlation and \( Z \) might not follow a correlated standard joint normal distribution.

The resultant 50-year environmental contour is shown in Figure 7 as a blue curve, and the transformed values of \( Z \) are shown in Figure 8 as grey dots, where \( z_V \) and \( z_{H_s} \) follow the normal distribution marginally; the correlation for data with positive \( z_V \) (i.e., high wind speeds) are noticeably higher than that for data with negative \( z_V \) (i.e., low wind speeds). Using the Nataf transformation, an “averaged” correlation coefficient is used, and as such, the extreme combinations of the 50-year wind and wave are underestimated, because a higher correlation coefficient will reshape the \( V - H_s \) contour towards the upper right corner in Figure 7. Note that using the Nataf transformation, the resultant contour always yields the same independent \( N \)-year hazards as estimated from the marginal distributions, which is reflected in the tangency between the blue curve and the red dashed line in Figure 7.

\[ \text{Figure 7. Similar to Figure 5, but with the blue curve indicating the resultant 50-year environmental contour using the Nataf transformation.} \]

\[ \text{Figure 8. Transformation result for the hourly data in Figure 7. The blue curve is the approximated contour assuming a correlated standard joint normal distribution.} \]

4.3. Rosenblatt transformation using the re-parameterized Weibull distribution

In order to accurately capture the tail behaviour of \( F_{H_s|V} \), and to avoid the extrapolation of too many parameters of \( F_{H_s|V} \) for high wind speeds, the tail of \( F_{H_s|V} \) is re-parameterized using the Weibull distribution (see Figure 3). Compared to Section 4.1, where the Weibull distribution fits the \( H_s|V \) accurately for only high wind speeds; the re-parameterized Weibull distribution here fits the tail of \( H_s|V \) accurately for the whole range of wind speeds. The two parameters of the re-parameterized Weibull distribution at 25 wind speeds are shown in Figure 9, where the function in Equation (4) shows a good fit for both parameters. With the fit parameters, a smooth contour is obtained (see Figure 10), which matches the unsmoothed \( H_{s,50}|V \) better than the results in Sections 4.1 and 4.2. The top apex of this contour is a little higher than the independent \( H_{s,50} \), which might be caused by the parameter extrapolation of the re-parameterized Weibull distribution.

The 50-year \( V - H_s \) contours for Massachusetts and New York sites are given in Figure 11, which match the unsmoothed \( H_{s,50}|V \) well. For the Massachusetts site, the contour also matches the independent \( H_{s,50} \) well, while overestimation of the top apex of the contour is observed for the New York site.
5. Conclusions
In the case where many parameters are used to describe a conditional distribution, the fitting and extrapolation process as part of the Rosenblatt transformation is cumbersome and prone to error. A re-parameterization technique is proposed in this paper, which allows the conditional distribution to be approximated by a prescribed distribution with fewer parameters. In the numerical example, the tail distribution of $H_e|V$ described by six parameters is re-parameterized using the two-parameter Weibull distribution, and the resultant environmental contour is validated using the unsmoothed $N$-year conditional $H_e$ estimated for various wind speed windows and the independent $N$-year $H_e$ estimated from the marginal $H_e$ distribution. Compared to conventional approaches, such as using a prescribed distribution for $H_e|V$ in the Rosenblatt transformation, or using the Nataf transformation, the re-parameterized Weibull distribution performs better.

Acknowledgement
This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1552559, the Massachusetts Clean Energy Center, and Northeastern University. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or other sponsors.

References
[1] S. R. Winterstein, T. C. Ude, C. A. Cornell, P. Bjerager, and S. Haver, "Environmental parameters for extreme response: Inverse FORM with omission factors," *Proceedings of the ICOSSAR-93, Innsbruck, Austria*, pp. 551-557, 1993.
[2] M. Rosenblatt, "Remarks on a multivariate transformation," *The annals of mathematical statistics*, vol. 23, no. 3, pp. 470-472, 1952.

[3] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," in *Mathematical Proceedings of the Cambridge Philosophical Society*, 1928, vol. 24, no. 2, pp. 180-190; Cambridge University Press.

[4] E. Gumbel, "Statistics of extremes. 1958," *Columbia Univ. press, New York*, vol. 247, 1958.

[5] R. B. Abernethy, J. Breneman, C. Medlin, and G. L. Reinman, "Weibull analysis handbook," Pratt and Whitney West Palm beach fl Government Products DIV1983.

[6] S. Saha et al., "NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010," ed. Boulder, CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, 2010.

[7] S. Saha et al., "NCEP Climate Forecast System Version 2 (CFSv2) Selected Hourly Time-Series Products," ed. Boulder, CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, 2011.

[8] C. Qiao, A. Myers, and S. R. Arwade, "Validation of a 37-year Metocean Hindcast along the U.S. Atlantic Coast," presented at the 13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13, Seoul, South Korea, May, 2019.

[9] BOEM, "Wind Planning Areas (Shapefile)," 2018.

[10] K. Johannessen, T. S. Meling, and S. Hayer, "Joint distribution for wind and waves in the northern north sea," in *The Eleventh International Offshore and Polar Engineering Conference, 2001*: International Society of Offshore and Polar Engineers.

[11] "IEC 61400-3: Wind turbines part 3: Design requirements for offshore wind turbines," *International Electrotechnical Commission*, 2009.

[12] A. Nataf, "Determination des Distribution don t les marges sont Donnees," *Comptes Rendus de l Academie des Sciences*, vol. 225, pp. 42-43, 1962.