

SANDIA REPORT

SAND2014-17550

Unlimited Release

Printed September 2014

Modified Inverse First Order Reliability Method (I-FORM) for Predicting Extreme Sea States

Aubrey C. Eckert-Gallup, Cédric J. Sallaberry, Ann R. Dallman, and Vincent S. Neary

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.



Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: <http://www.osti.gov/bridge>

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



Modified Inverse First Order Reliability Method (I-FORM) for Predicting Extreme Sea States

Aubrey C. Eckert-Gallup¹, Cédric J. Sallaberry², Ann R. Dallman³,
and Vincent S. Neary³

Departments 06233¹, 06224², and 06122³
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185 MS0744

Abstract

Environmental contours describing extreme sea states are generated as the input for numerical or physical model simulations as a part of the standard current practice for designing marine structures to survive extreme sea states. Such environmental contours are characterized by combinations of significant wave height (H_s) and energy period (T_e) values calculated for a given recurrence interval using a set of data based on hindcast simulations or buoy observations over a sufficient period of record. The use of the inverse first-order reliability method (IFORM) is standard design practice for generating environmental contours. In this paper, the traditional application of the IFORM to generating environmental contours representing extreme sea states is described in detail and its merits and drawbacks are assessed. The application of additional methods for analyzing sea state data including the use of principal component analysis (PCA) to create an uncorrelated representation of the data under consideration is proposed. A reexamination of the components of the IFORM application to the problem at hand including the use of new distribution fitting techniques are shown to contribute to the development of more accurate and reasonable representations of extreme sea states for use in survivability analysis for marine structures.

Keywords: Inverse FORM, Principal Component Analysis, Environmental Contours, Extreme Sea State Characterization, Wave Energy Converters

Acknowledgments

This research was made possible by support from the U.S. Department of Energy's (DOE) Energy Efficiency and Renewable Energy (EERE) Office's Wind and Water Power Technologies Office. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. This paper is an independent product of the authors and does not necessarily reflect views held by either SNL or the U.S. Department of Energy. (SAND2014-17550)

CONTENTS

1	Introduction.....	9
2	Study of the data and original approach	11
2.1	Review of the initial code.....	11
2.2	Study of the wave data	16
3	Description of specific mathematical methods	21
3.1	Inverse FORM.....	21
3.2	Principal component analysis.....	23
4	Creation of extreme sea state contour	27
4.1	Distribution and parameter fitting	27
4.2	Application of inverse FORM methodology.....	32
5	Comparison of results with initial code	33
6	Conclusion and perspective	35
7	Bibliography	37
8	Distribution	39

Figures

Figure 1:	Representation of the 3-parameter Weibull distribution used to fit the H_s data.....	12
Figure 2:	Entire CDF representation (top) and zoom in on the highest quantiles (bottom) for significant wave height (H_s).	13
Figure 3:	Relationship between lognormal parameters (top: mean of log, bottom: standard deviation of log) and significant wave height.	14
Figure 4:	100-year contour (and expansion) around a scatterplot of significant wave height vs. energy period for NDBC 46212.	15
Figure 5:	100-year contour (and expansion) around a scatterplot of significant wave height vs. energy period for (a) 46022, (b) 51202, and (c) 46050.	16
Figure 6:	Representation of data density for four study sites (a) NDBC 46212, (b) NDBC 46022, (c) NDBC 51202, and (d) NDBC 46050.	18
Figure 7:	Representation of the standard normal space used by reliability techniques.	21
Figure 8:	Transposition of an isoline (and center point) from the standard normal space (left) into the original sample space (right).	22
Figure 9:	Representation of the axes of the new basis developed using principal component analysis for NDBC 46212.....	25
Figure 10:	Component one CDF for all data (blue) and inverse Gaussian model (red) for NDBC 46212.....	27
Figure 11:	CDFs for all bins of component two for NDBC 46212.	28
Figure 12:	CDFs for selected bins of component two with normal distribution fits for NDBC 46212.	29
Figure 13:	Estimates of the component two normal distribution parameters μ (top) and σ (bottom) as a function of component one for each bin for site 46212.	30

Figure 14: Estimates of component two normal distribution parameters μ (top) and σ (bottom) as a function of component one for each bin with related fitting functions for NDBC 46212. 31

Figure 15: 100-year extreme sea state contour for NDBC 46212. 32

Figure 16: Extreme sea state contour for NDBC 46212 created by the original methodology (left) and the new methodology shown with data density (right)..... 33

Figure 17: Extreme sea state contour for NDBC 46022 created by the original methodology (left) and the new methodology shown with data density (right)..... 33

Figure 18: Extreme sea state contour for NDBC 51202 created by the original methodology (left) and the new methodology shown with data density (right)..... 34

Figure 19: Extreme sea state contour for NDBC 46050 created by the original methodology (left) and the new methodology shown with data density (right)..... 34

Nomenclature

CDF	Cumulative Distribution Function
FORM	First Order Reliability Method
H_s	Significant Wave Height
IFORM	Inverse FORM
NDBC	National Data Buoy Center
PCA	Principal Component Analysis
T_e	Energy Period
T_p	Peak Period
WEC	Wave Energy Converter

This page intentionally left blank.

1 INTRODUCTION

The current practice for designing marine structures to survive extreme sea states is to apply nonlinear time domain numerical simulations to predict the structural response to a short-term extreme wave or wave group. Extreme wave design generally includes the following steps as outlined in (Coe, et al. 2014) (1) Application of hindcast simulations or buoy observations of a sufficient duration (twenty years preferred) and an appropriate location; (2) Application of extreme value theory and models used for extrapolation to events more extreme than those observed in a shorter period of record; (3) Generation of environmental contours consisting of H_s , T_p pairs that elicit extreme structural responses for a given return period ; (4) Identification of one or more extreme sea states, which can be used with a wave spectrum, appropriate for the location of interest, to reconstruct a single extreme wave or wave group as input for numerical or physical model simulation.

(Vanem and Bitner-Gregersen 2014) summarize methods for generating environmental contours, including the traditional inverse first-order reliability method (IFORM) by Winterstein (Winterstein, et al. 1993), which uses the Rosenblatt transformation (Rosenblatt 1952), and the more recent methods, which avoid the Rosenblatt transformation by employing Monte Carlo simulations of a joint probability model (Vanem and Bitner-Gregersen 2014).

It is recognized that the environmental load associated with the largest significant wave height on the environmental contour is not necessarily the one that will cause failure (Baarholm, Haver and Økland 2010). In fact, marine structures can fail due to resonant oscillations of waves and wave groups associated with smaller significant wave heights waves with periods that match the natural frequencies of motion of the structure or its subsystems. Although 100-year recurrence intervals (return periods) are common for marine structures, lower return periods can be used, if acceptable for survivability, when the design service life is less than 100 years (DNV 2005).

The IFORM continues to be standard design practice for generating environmental contours used for estimating extreme sea states of a given recurrence interval or return period, e.g., 100 years (DNV 2014). Environmental loads associated with these extreme sea states are used to design various marine structures, including ships (DNV 2002), dynamic risers (DNV 2001), position moorings (DNV 2010a), offshore floating platforms (DNV 2010b), and wave energy converters (WEC) (DNV 2008). For this reason, (Dallman and Neary 2014) constructed environmental contours to characterize extreme sea states at sites identified for testing or commercially developing WEC technologies.

The purpose of this study is to investigate techniques for improving the traditional IFORM through improved extreme value models for the correlated random variables and principal component analysis (PCA). Berg (2011) estimated extreme sea states using the traditional IFORM for a wave energy site located in Humboldt Bay, California (Neary, et al. 2014), but many measured data points fell outside the calculated contour, even when inflated by 20% to account for approximations using the traditional IFORM. A study of the observations density reveals that the problem was

partly coming from fitting significant wave height and energy period separately. Principal component analysis was used to capture the relation between these two parameters, allowing for a better coverage of the data in the period of record under consideration as well as for the creation of a more appropriately shaped extreme sea state contour.

Data from four buoys was used for analysis: National Data Buoy Center (NDBC) 46212 offshore of Northern California in 40 m depth, NDBC 46022, also offshore of Northern California in 675 m depth, NDBC 51202 offshore of Oahu in 82 m depth, and NDBC 46050 offshore of Oregon in 128 m depth.

2 STUDY OF THE DATA AND ORIGINAL APPROACH

2.1 Review of the initial code

As a part of the effort to characterize sea states relevant for studying the response of a wave energy conversion device under extreme events, a code was initially developed in Matlab in 2011 (Berg 2011) using the equations proposed in (Haver and Winterstein 2008) to implement the IFORM as recommended in the DNV standard on position mooring (DNV 2010). The method developed in this existing code consists of a two-step process used to characterize extreme sea states. The first step includes fitting distributions to the significant wave height (H_s) and either the energy period (T_e) or peak period (T_p) observations. The second step uses these fitted distributions to estimate extreme sea states. Note that in this study and for the application of the wave resource catalogue, the energy period (T_e) is used because it is widely used in wave energy applications (Lenee-Bluhm, Paasch and Özkan-Haller 2011), and has been found to be more robust than the peak period (T_p), due to a high sensitivity to spectral peak.

2.1.1 Application of current approach to extreme sea state contour development

The approach presented in the original code uses the same traditional monodimensional fitting technique that is presented in the literature (Haver and Winterstein 2008). The consequence of using a monodimensional fitting approach as opposed to a more multidimensional consideration will be discussed in section 2.2.2. A more multidimensional approach, developed using principal component analysis, will be described in chapter 3.2.

The first step of the current method includes fitting the existing significant wave height (H_s) and energy period (T_e) data using probability distributions. A least squares technique is applied to a 3-parameter Weibull distribution in order to fit the cumulative distribution function (CDF) of the H_s data. An optimization is performed on the two classical scale (λ) and shape (k) parameters of a Weibull distribution (Johnson, Kotz and Balakrishnan 1994) along with an additional third parameter (α) that serves as an offset on x . This third parameter allows the Weibull distribution to be shifted on the x axis, as is shown in Figure 1.

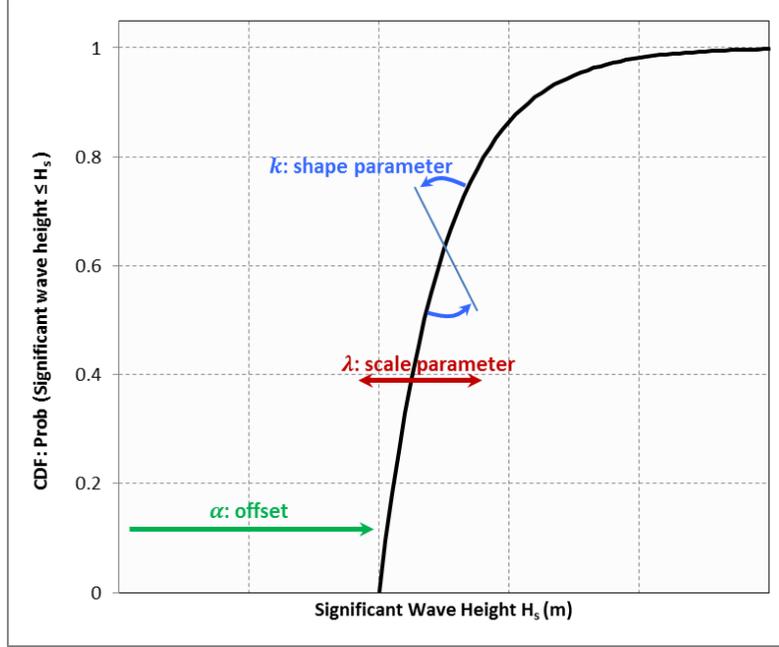


Figure 1: Representation of the 3-parameter Weibull distribution used to fit the H_s data.

The CDF and inverse CDF of the 3-parameter Weibull distribution are given by:

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x-\alpha}{\lambda}\right)^k} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

and

$$F^{-1}(q) = \alpha + \lambda[-\ln(1 - q)]^{1/k} \text{ for } q \in]0,1[$$

In order to increase the speed of the fitting calculation, the observations are first grouped into a set of 49 bins using a constant significant wave height increment. The subsequent 3-parameter Weibull fitting resulting from this binning approach for NDBC 46212 can be seen in Figure 2.

The binning approach used in this methodology slightly underestimates the distribution; for a given wave height, the quantile value is overestimated, meaning that the likelihood of occurrence for a certain wave height being equal to or less than the quantile value is lower for the binned data than the likelihood that is actually observed. While the corresponding 3-parameter Weibull fitting provides a good representation of the binned distribution, its accuracy drops for the highest quantiles, as is shown in Figure 2. The consequence of this loss of accuracy is that, for a high value of significant wave height, the fitted distribution will associate a higher quantile value, and therefore a lower likelihood of occurrence, than what was observed, as was stated above. Thus, this underestimation may lead to a prediction of maximum significant wave height for a 100-year return period that is significantly smaller than the significant wave height values observed in the period of record.

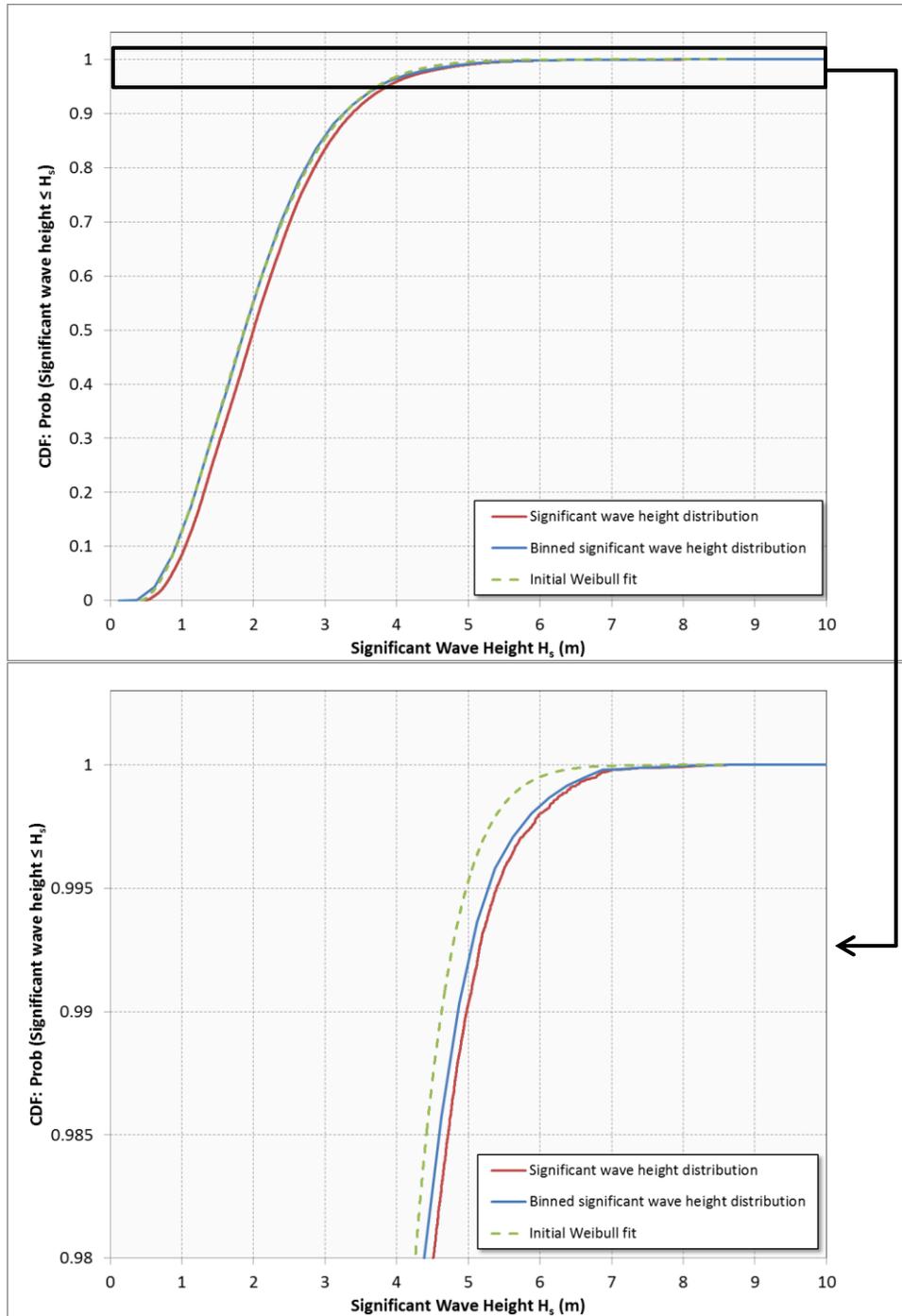


Figure 2: Entire CDF representation (top) and zoom in on the highest quantiles (bottom) for significant wave height (H_s).

The energy period data is split into bins based on corresponding significant wave height values. Within each bin, the approach for fitting the energy period data is similar to that used for the significant wave height. The energy period values in each bin are fitted with a lognormal distribution. The mean (μ) and standard deviation (σ) of the log (the traditional lognormal parameters (Johnson, Kotz and Balakrishnan 1994)) are then estimated. As a result, distributions of

μ and σ as a function of significant wave height are obtained. These distributions are displayed in Figure 3.

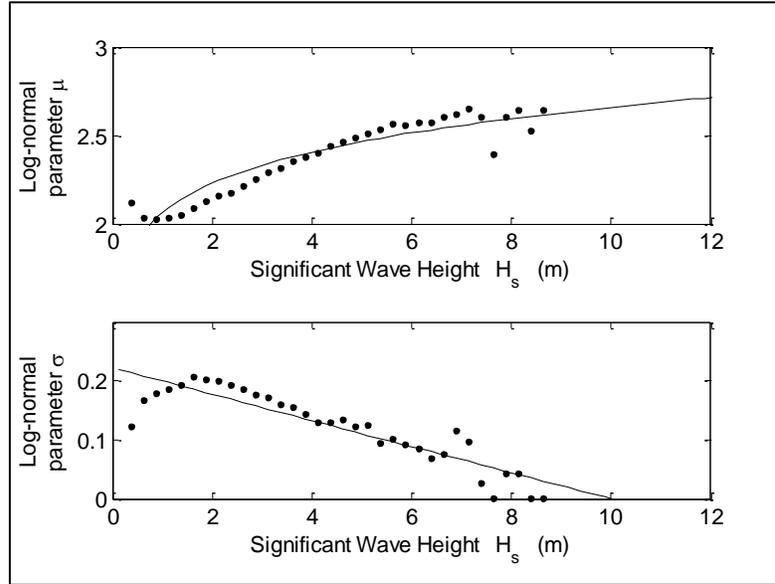


Figure 3: Relationship between lognormal parameters (top: mean of log, bottom: standard deviation of log) and significant wave height.

The use of a binning scheme based on a decomposition of the domain of significant wave height creates uneven samples of energy period with very few observations for the highest wave height intervals. As a result, the behavior of the lognormal parameters for each bin of energy period is unstable for high values of significant wave height, as is seen above.

The sets of lognormal parameters μ and σ are fit with models that describe their behavior as a function of significant wave height. The data fitting models used for the μ and σ parameters shown in Figure 3, proposed in (Haver and Winterstein 2008) and based on (Nygaard and Johannessen 2000), are given by:

$$\begin{aligned}\mu(H_s) &= m_1 + m_2(H_s)^{m_3} \\ \sigma^2(H_s) &= s_1 + s_2 \exp(-s_3 H_s)\end{aligned}$$

As seen in Figure 3, the fitting models for these parameters do not fit the data, especially for the highest significant wave height intervals. These poor fits contribute to inaccuracies in the final extreme sea state contour.

Following the creation of fitting models for μ and σ , the IFORM is used to calculate the extreme sea state contour. The details of this method are described in section 3. The result of the application of the inverse FORM approach to the problem of interest using the fitting models described above is presented in Figure 4. The 100-year contour estimates a smaller wave height than is actually observed in the 8 years of data used in the analysis. This is due to the loss of accuracy in the data fitting models for high quantile values previously described. This contour also fails to estimate the total extent of the energy period when the value of significant wave height is high.

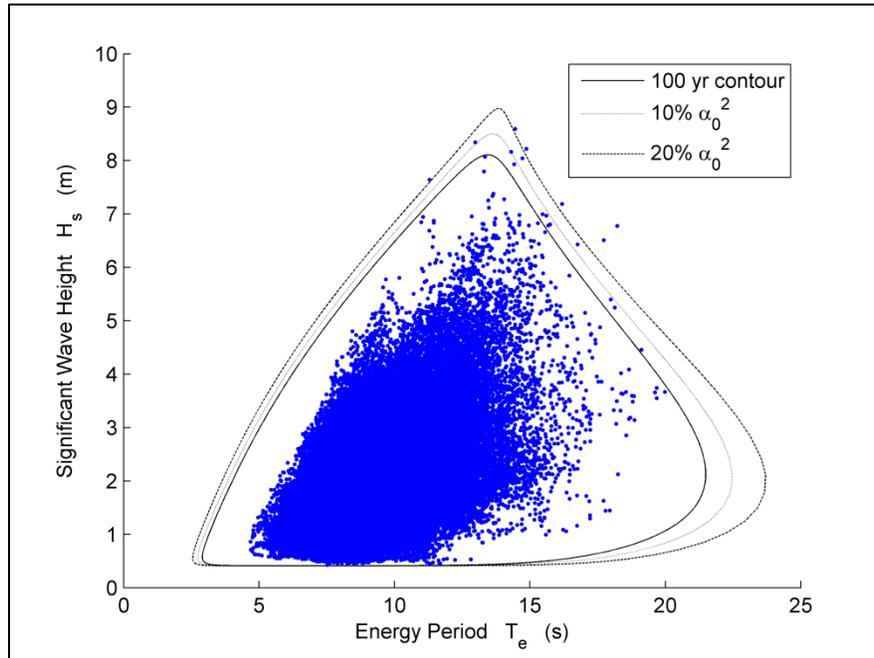


Figure 4: 100-year contour (and expansion) around a scatterplot of significant wave height vs. energy period for NDBC 46212.

The 100-year contour was expanded by both ten and twenty percent in order to create a better coverage of the data from the period of record under consideration. Although this expansion does allow for the inclusion of some of the points falling outside of the original 100-year contour, the selection of these expansion contours is arbitrary and does not reflect any true description of the problem at hand.

The extreme sea state contours created for three additional study sites (NDBC 46022, 51202, 46050) using the original code are shown in Figure 5 below.

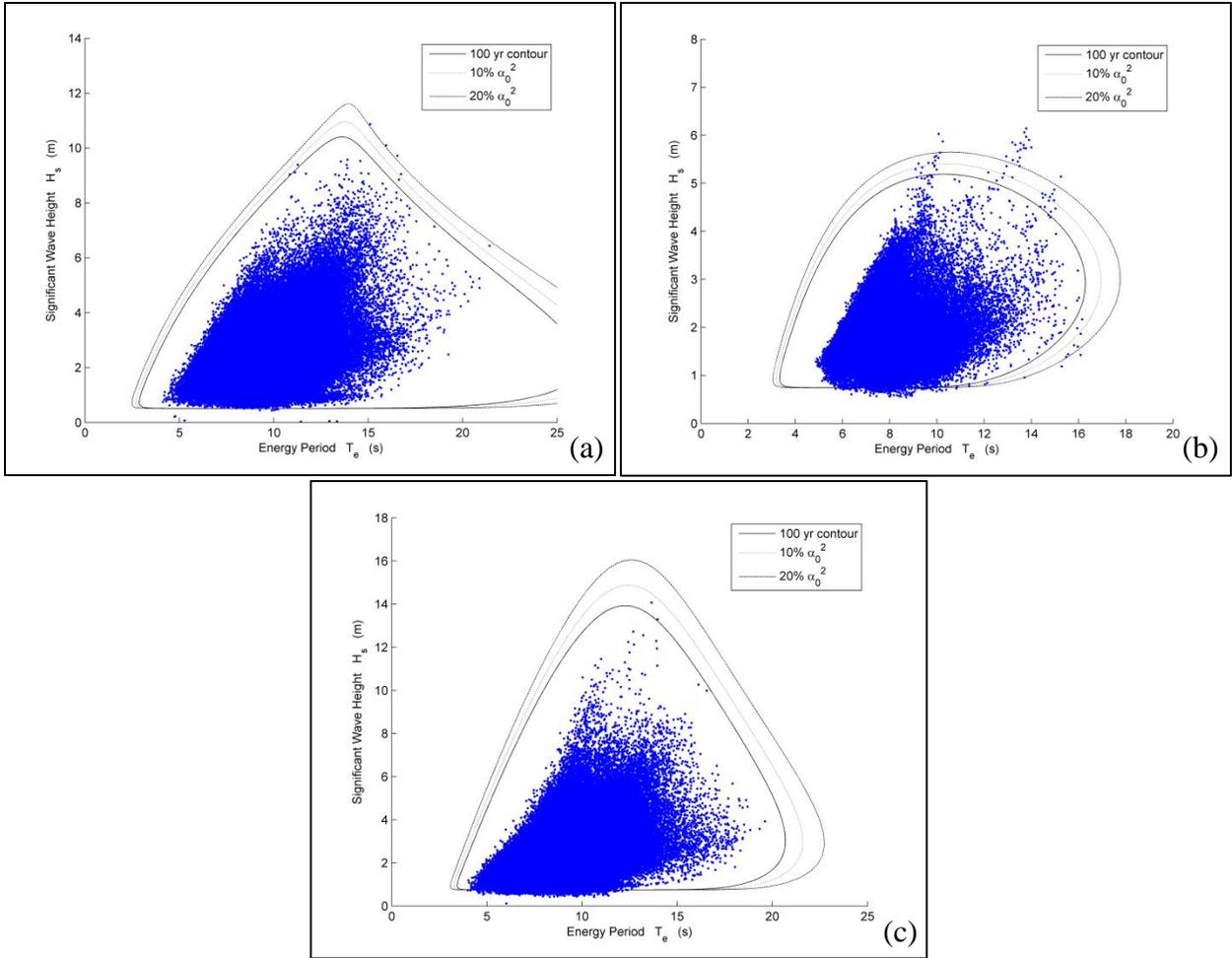


Figure 5: 100-year contour (and expansion) around a scatterplot of significant wave height vs. energy period for (a) 46022, (b) 51202, and (c) 46050.

2.2 Study of the wave data

The wave data across four sites of interest was studied and compared as a first step in the process of improving upon the traditional method demonstrated in the original code. This study included the creation of a representation of data density, the development of an understanding of the conjoint influence of energy period and wave height, and the characterization of differences between the data sets under consideration.

2.2.1 Representation of plot density

A representation of the density of the data at each study site was created in order to understand the underlying patterns and trends masked by a traditional scatterplot representation of the data. In order to estimate the density, significant wave height and energy period data are first normalized for each value X as follows:

$$X' = \frac{X - \mu}{\sigma}$$

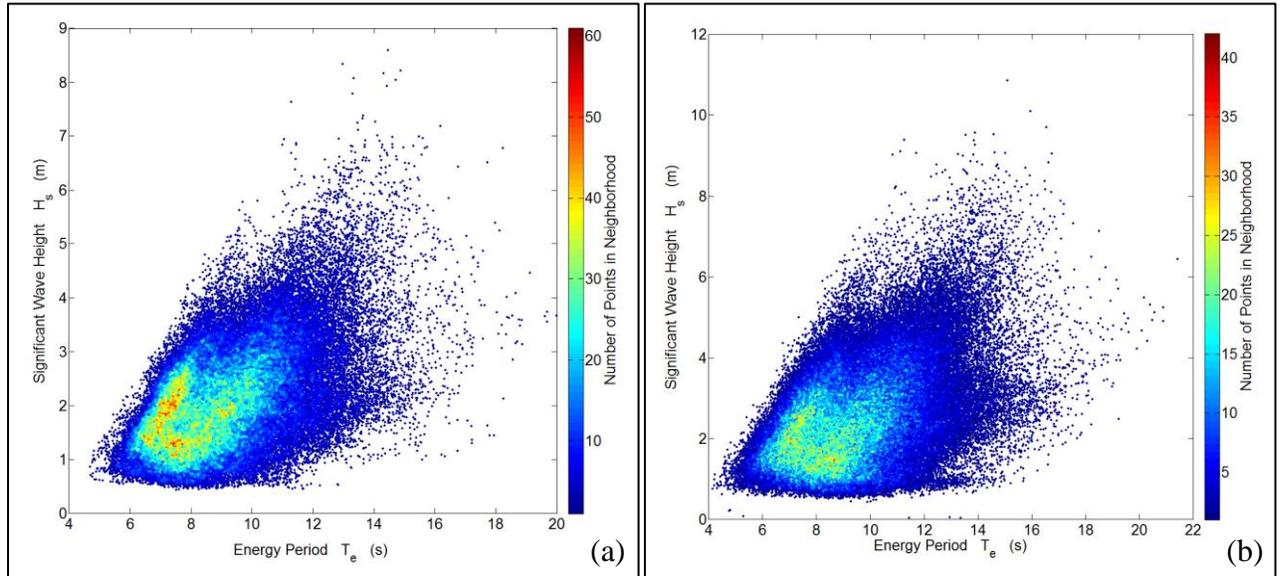
Where μ and σ are the respective mean and standard deviation of the set of data that X belongs to (H_s or T_e).

A radius ε is estimated using the total number of points N in the dataset. The radius of the neighborhood defined by $\varepsilon = 2000/N$ was found to represent the density well based on a trial and error approach. Following this radius estimation, a circular neighborhood is constructed around each point $p_i = (x_i, y_i)$. For each neighborhood, the number of subsequent points N_i in the dataset that falls inside of this radius is counted for each individual point using the formula presented below:

$$N_i = \sum_{j=1}^N I \left(\sqrt{(x_i^2 - x_j^2) + (y_i^2 - y_j^2)} < \varepsilon \right)$$

where I is the indicator function equal to 1 if the condition is true and 0 if otherwise.

This provides an estimate of data density for each study site under consideration. Examples of the data density calculated for each of the four sites studied during the development of this methodology are shown in Figure 6.



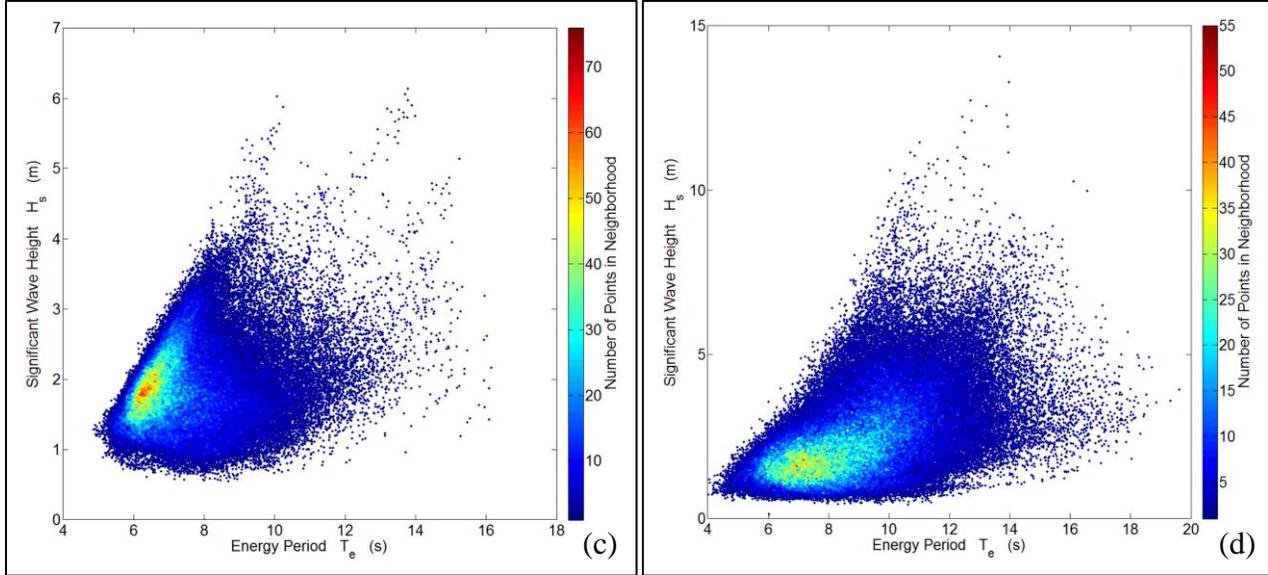


Figure 6: Representation of data density for four study sites (a) NDBC 46212, (b) NDBC 46022, (c) NDBC 51202, and (d) NDBC 46050.

These representations of data density help to characterize the developmental patterns present in the data by showing differences in frequencies across the entire dataset. The shape of the trends shown in the density plots above help to support the new methodology proposed in subsequent sections.

2.2.2 Conjoint effect of energy period and significant wave height

The original methodology used to approach the problem at hand represents the relation between T_e and H_s monodimensionally, meaning that T_e and H_s are treated independently. This monodimensional treatment fails to capture the complex relation between these two datasets and creates a misrepresentation of the worst case scenario calculated in the application of the inverse FORM method. Under this approach, the extreme values of T_e will be considered at the median values of H_s while the extreme values of H_s will be considered at median values of T_e . This representation fails to cover the area of the dataset in which the values of T_e and H_s are both high, an area that the trends present in the density plots shown in Figure 6 demonstrate to be of importance and may lead to the most extreme case for a given return period. As was mentioned in the introduction, it is not necessarily the largest significant wave height or period, but a combination of the two that may be the most critical for WEC survivability. The extreme sea states of most concern may depend on the natural frequencies of the structure or subsystems under consideration.

2.2.3 Differences in behavior between the sites

The density plots provided in Figure 6 show trends in the development of the entire dataset for each study site. The shapes present in these density plots both indicate areas of importance in the relation between T_e and H_s and also provide insight into the different behaviors found at the study sites under consideration.

An example of the type of site-dependent complexity in the relation between H_s and T_e can be seen in Figure 6(c). The data in the period of record for this site displays several dependency tendencies manifested in long fingers of extreme points that appear to be related. An examination of the data for this site as a function of time reveals that these long fingers may be related to individual storm events.

An additional example of a different complex relation between these two datasets can be found in Figure 6(d). The plot of the density for this data shows that the relation between H_s and T_e at this site is curved and, thus, an orthogonal decomposition may not be the best way to capture the intricacies of this relation. However, such decomposition will allow for a better analysis of the data than ignoring the inherent relation between these measurements, as was done in the original methodology.

While the new methodology proposed in this report attempts to capture some of the relationship between H_s and T_e at each site, the complexity of this relation varies from site to site. Additional work attempting to capture the complex relation linking H_s and T_e at each site poses mathematical and computational problems that are beyond the current scope of this report. In order to achieve the development of a method that can be easily applied to a variety of sites of interest rather than focusing on the detailed characterization of site-dependent behaviors, a simple orthogonal decomposition is proposed as a first step towards the desired representation. The method used to apply this decomposition is described in section 3.2.

This page intentionally left blank.

3 DESCRIPTION OF SPECIFIC MATHEMATICAL METHODS

The inverse FORM (First-Order Reliability Method) and principal component analysis are the two main components in the development of extreme sea state contours through the methodology presented in this work. These mathematical methods are described in the following sections in the context of their application to the current problem of interest.

3.1 Inverse FORM

In order to construct a 100-year contour representing extreme wave events, the fitted probability distributions for the variables of interest are used in an inverse FORM (First-Order Reliability Method) approach.

The FORM approach (Zhao and Ono 1999) consists of projecting the input space in a standard multidimensional normal space, meaning that each uncertainty is represented using an uncorrelated normal distribution function. In such a space, the center point O is the most likely area (mode). The further the solution moves from this center point, the less likely the solution will be, as is seen in Figure 7.

The Nataf or Rosenblatt transformations are usually used to de-correlate the data (Liu and Der Kiureghian 1986). The Nataf transformation is based on covariance matrix decomposition. In the current context, the use of principal component analysis, described in section 3.2, will provide an equivalent benefit such that the use of either of these transformations is not necessary.

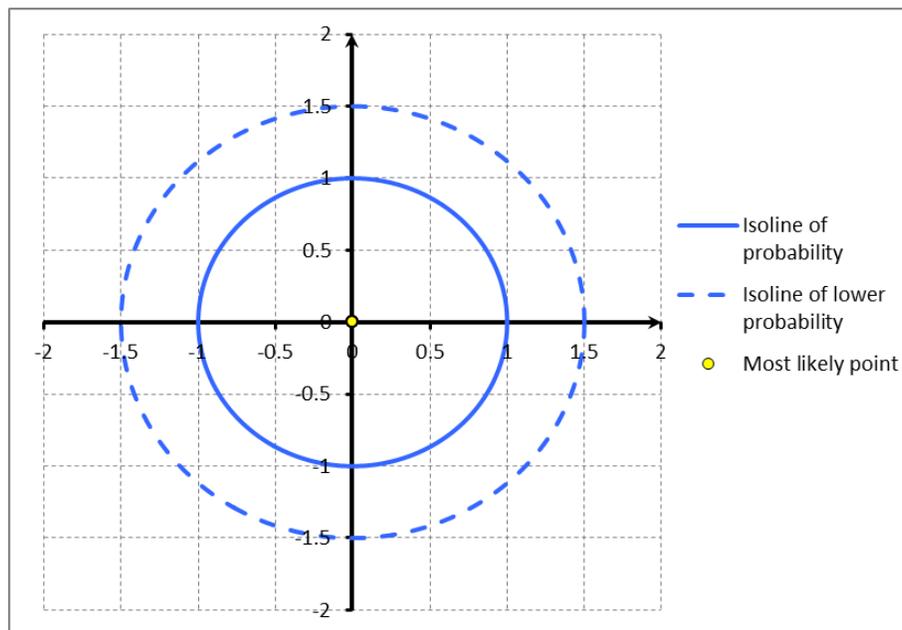


Figure 7: Representation of the standard normal space used by reliability techniques.

In the standard FORM approach, a threshold value is considered and its likelihood is estimated in the standard normal space. The inverse FORM (Winterstein, et al. 1993) approach starts from the

normal space and a probability of likelihood (for instance a return period of 10 or 100-years) that defines an isoline, as is seen in Figure 7. This isoline is then transposed into the original uncertain input space in order to evaluate the potential range of extreme values. Numerically, a discretization is used on the angle θ over the isoline, represented as a parametric function as is shown below:

$$\begin{cases} x_i = \beta \cos(\theta_i) \\ y_i = \beta \sin(\theta_i) \end{cases} \quad \text{with } \theta_i = \frac{i2\pi}{k} \quad i = 1, \dots, k$$

where β represents the radius of the circle, i.e., the distance from the most likely point.

For each value of i , the quantile position of the chosen probability of likelihood is calculated in both directions in the standard normal space. Each resulting quantile is then evaluated using the inverse CDF that represents the input distribution in order to estimate the boundary that would lead to extreme events amongst the input set, as is shown in Figure 8. In the problem of interest, the x axis represents the energy period while the y axis represents significant wave height.

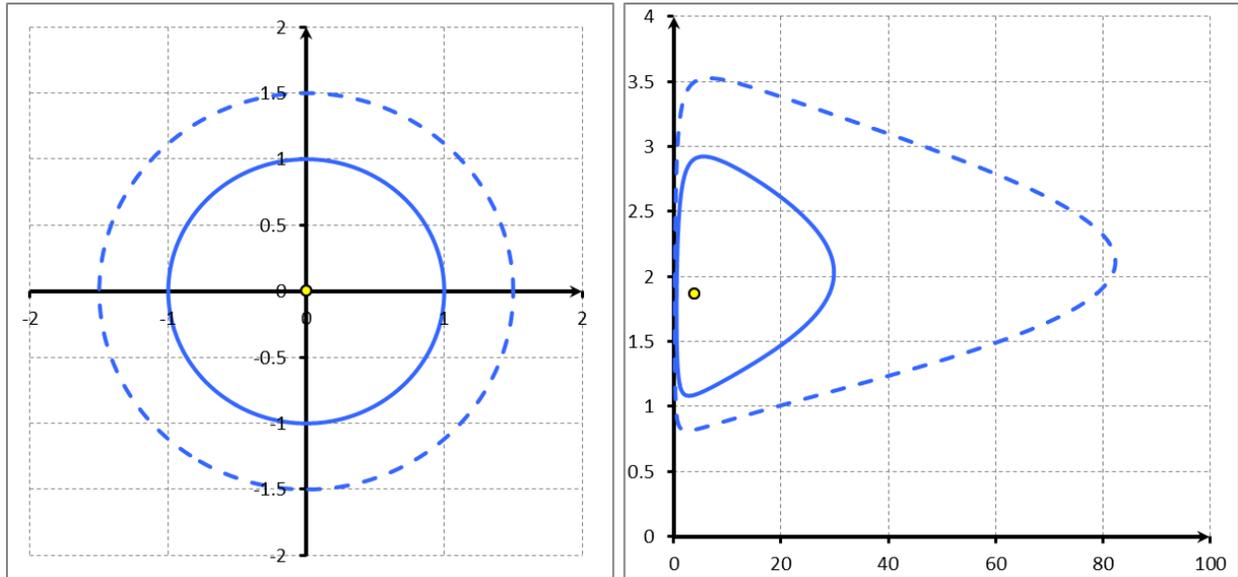


Figure 8: Transposition of an isoline (and center point) from the standard normal space (left) into the original sample space (right).

3.2 Principal component analysis

The concept of principal components was initially introduced by Karl Pearson in 1901 (Pearson 1901) and more formally developed by H. Hotelling in 1933 (Hotelling 1933). The underlying goal of principal component analysis (PCA) is to develop a new orthogonal basis in which the variables will be (1) uncorrelated and (2) sorted such that the first variable represents the direction in which the data has the largest variance and each subsequent variable leads to the next largest variance.

PCA provides a powerful transformation that works to reduce the dimensionality of a problem considering that the higher order components have a low impact on the variance of the data. Traditionally, the variables in the new basis are called principal components while the values associated to each variable in this new basis for each point are called z-scores. The mathematical tools used to generate this new basis are based on classical linear (matrix) algebra applied to the covariance matrix, taking advantage of its structure (square symmetrical and non-singular matrix) (Jackson 1991).

Principal component analysis was used to remove the correlation between the two variables analyzed in the problem of interest (i.e., significant wave height and energy period) for each dataset because of this method's intrinsic properties and its simplicity of application. The application of PCA to these two variables will generate two new variables that will be called component one (C_1) and component two (C_2). C_1 representing the component with the highest variance. The application of PCA to the original H_s and T_e data yields a coefficient matrix defining a linear combination that allows for rotation into the principal component space. The rotation axes defined by this linear combination are shown in Figure 9. The general form of this coefficient matrix is shown below:

$$\mathbf{V} = \begin{bmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{bmatrix}$$

where $v_{2,2} = -v_{1,1}$ and $v_{1,2} = v_{2,1}$.

The equations for each component based on the application of the coefficient matrix above for each point $p_i = (H_{s_i}, T_{e_i})$ are shown below:

$$\mathbf{U} = \begin{bmatrix} H_{s_1} & T_{e_1} \\ \vdots & \vdots \\ H_{s_i} & T_{e_i} \end{bmatrix}$$

$$\mathbf{C} = \mathbf{UV} = \begin{bmatrix} C_{1_1} = H_{s_1} v_{1,1} + T_{e_1} v_{2,1} & C_{2_1} = H_{s_1} v_{1,2} + T_{e_1} v_{2,2} \\ \vdots & \vdots \\ C_{1_i} = H_{s_i} v_{1,1} + T_{e_i} v_{2,1} & C_{2_i} = H_{s_i} v_{1,2} + T_{e_i} v_{2,2} \end{bmatrix}$$

In order to fulfill the requirements for subsequent elements of the extreme event analysis (i.e., fitting probability distributions to the data), the rotated components must also be shifted upwards along the y axis to ensure that they are entirely positive. This is achieved by simply applying a shift s defined by the following equation:

$$s = |\min(C_2)| + 0.1$$

The final equations for the components defined by a single original data point $p_i = (H_{s_i}, T_{e_i})$ are then given by:

$$\begin{aligned} C_{1_i} &= H_{s_i}v_{1,1} + T_{e_i}v_{2,1} \\ C_{2_i} &= H_{s_i}v_{1,2} + T_{e_i}v_{2,2} + s \end{aligned}$$

The components defined by these equations are used throughout the remainder of the analysis until they are transformed back into the original space in order to show the extreme sea state contour in the input space defined by variables T_e and H_s rather than in the principal component space defined by variables C_1 and C_2 .

An additional benefit of principal component analysis is that it is a bijective transformation, meaning that there is one and only one way to transform back to the original space, and that the inverse transformation is also a simple linear combination. Given a point $p_i = (C_{1_i}, C_{2_i})$ in the principal component space, the transformation back to the corresponding point on the extreme sea state contour in the original space defined by variables H_s and T_e is defined by:

$$\begin{aligned} H_{s_i} &= \frac{v_{2,2}C_1 - v_{2,1}(C_2 - s)}{v_{1,1}v_{2,2} - v_{1,2}v_{2,1}} \\ T_{e_i} &= \frac{v_{1,2}C_1 - v_{1,1}(C_2 - s)}{v_{1,2}v_{2,1} - v_{1,1}v_{2,2}} \end{aligned}$$

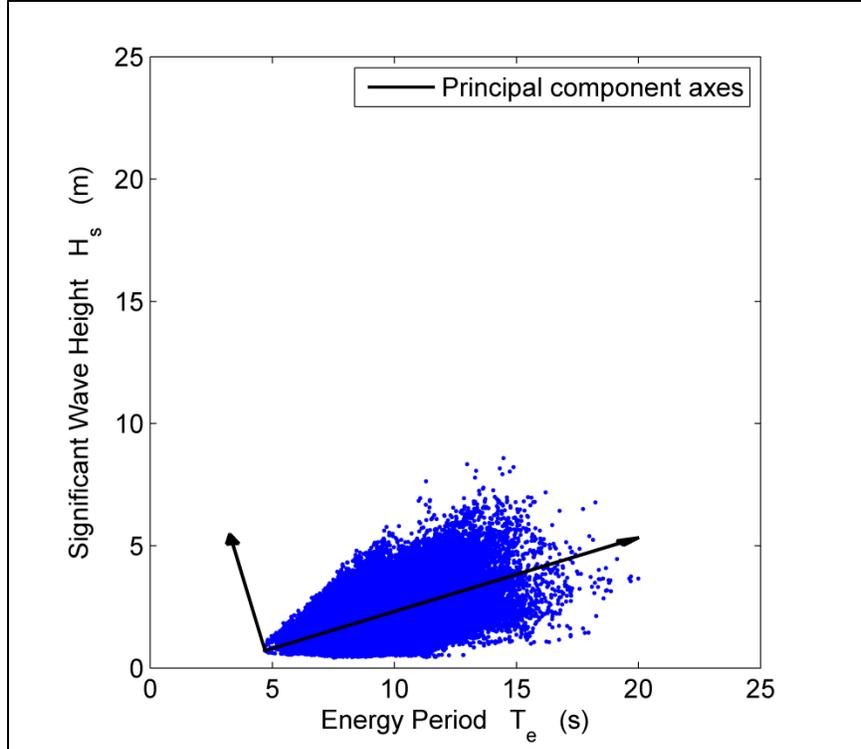


Figure 9: Representation of the axes of the new basis developed using principal component analysis for NDBC 46212.

As can be seen in Figure 6, there is a definite correlation between energy period and significant wave height. Dissociating these two variables and treating them independently, as is done in the traditional approach to extreme sea state characterization, underestimates the inherent dependency between the values. The principal components in the new basis will be uncorrelated and, thus, the rotation from the old basis into the new basis under the methodology described in this work will capture some of the dependency between the initial variables T_e and H_s , as is shown in Figure 9.

Although PCA partially captures the dependency between H_s and T_e , some dependency may still remain due to the complexity of the relation between these two variables. The remaining dependency will be taken into account using the same approach proposed in the original code. In keeping with this approach, C_2 will be split into multiple groups depending on the value of C_1 .

The implementation of this methodology differs in two ways from its application in the original code. First, in the original application of distribution fitting, H_s was fit first with a single distribution while T_e was split into bins to be fit with distributions based on the value of H_s . In the new implementation, C_1 , which is mostly influenced by T_e , will be the data set that is fit first with a single distribution while C_2 , mostly influenced by H_s , will be binned based on the values of C_1 in order to be fit with distributions based on the value of C_1 . Second, the binning of the second group of variables, C_2 , will be based on a discrete number of points in the new implementation rather than the binning scheme based on decomposition of the domain covered by the first component that was applied in the original methodology. The reasoning behind this new binning scheme is linked to the density of points; it is not necessary to have a large number (several thousands) of values to fit a distribution in some areas while some other areas are poorly represented with a distribution

considering only a small number (less than 10 in some cases) of points, as was seen in the application of the original methodology.

The use of principal component analysis can easily be extended to include multiple dimensions with each new set of components ultimately presented as a simple linear combination of the original variables. In future work, this could be used to consider additional variables, e.g., wind and current speed, related to the problem of extreme sea state characterization.

The idea of expressing the data using principal components prior to use the inverse FORM approach has been applied to design current profiles in the past (Forristall and Cooper 1997).

4 CREATION OF EXTREME SEA STATE CONTOUR

Principal component analysis is applied to the T_e and H_s data in order to create a representation of the data in the space defined by the principal components component one, C_1 , and component two, C_2 . As is described in section 3.2, this helps to remove the correlation between the variables in the input space. These components are used throughout the remainder of the analysis in the creation of the extreme event contour until this contour is transformed back into the original input space.

4.1 Distribution and parameter fitting

4.1.1 Fitting of first component

Following the rotation of the dataset into the principal component space, the CDF of C_1 is fitted with an inverse Gaussian distribution. This component was chosen for the initial fitting because it has the largest variance, as can be seen in Figure 9. The inverse Gaussian distribution was chosen from the 23 available fitting distributions in the Matlab Statistics toolbox both because it provides a good fit of the CDF shape observed in the dataset and because of the simplicity of its defining parameters in terms of interpretation. The result of this fitting for NDBC 46212 is shown in Figure 10 below.

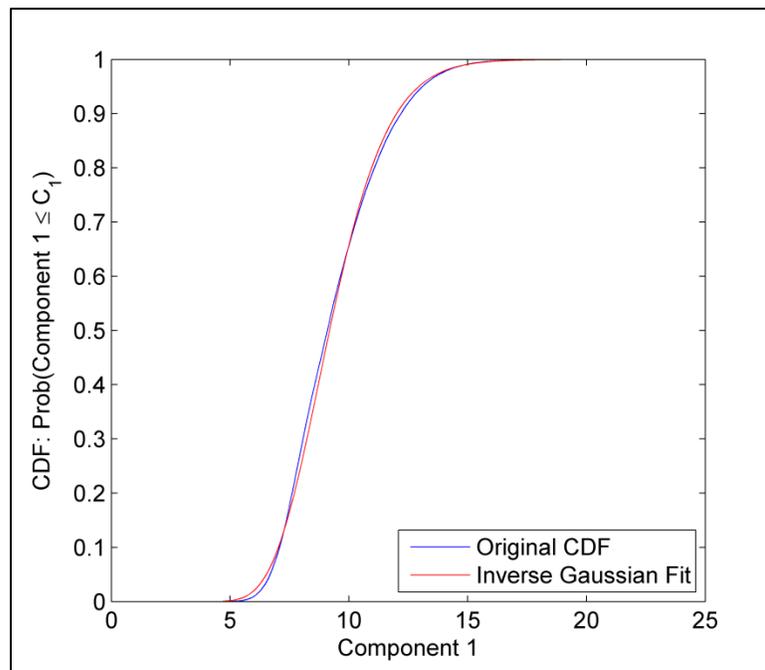


Figure 10: Component one CDF for all data (blue) and inverse Gaussian model (red) for NDBC 46212.

While the inverse Gaussian distribution fits three of the study sites very well, the fitting is not as good as expected for NDBC 51202, resulting in a less than optimal extreme event estimate as seen in Figure 18. A method to improve this distribution fitting in the future will be discussed in section 6.

4.1.2 Splitting component two according to the value of component one

The values of C_2 were sorted according to their corresponding C_1 values and binned as a first step in the determination of parameter models to be used in the application of the inverse FORM approach to construct a characterization of extreme wave events. This approach attempts to capture some of the necessary dependency between significant wave height and energy period by representing the parameters of the distributions fitted to the C_2 values for each bin as functions of corresponding representative C_1 values.

In order to create a binning scheme that covers the distribution of C_2 values with a more balanced representation, the sorted C_2 data was split into groups of 250 up to the last group, which contains all remaining points. This number is chosen arbitrarily but seems to be reasonable based on the original sample size. This binning scheme allows for a much better coverage of the distribution of the C_2 data, minimizing the errors created by the binning methodology applied in the original problem evaluation, as was described in section 2.1.1. The CDFs for all of the C_2 bins are shown in Figure 11 below.

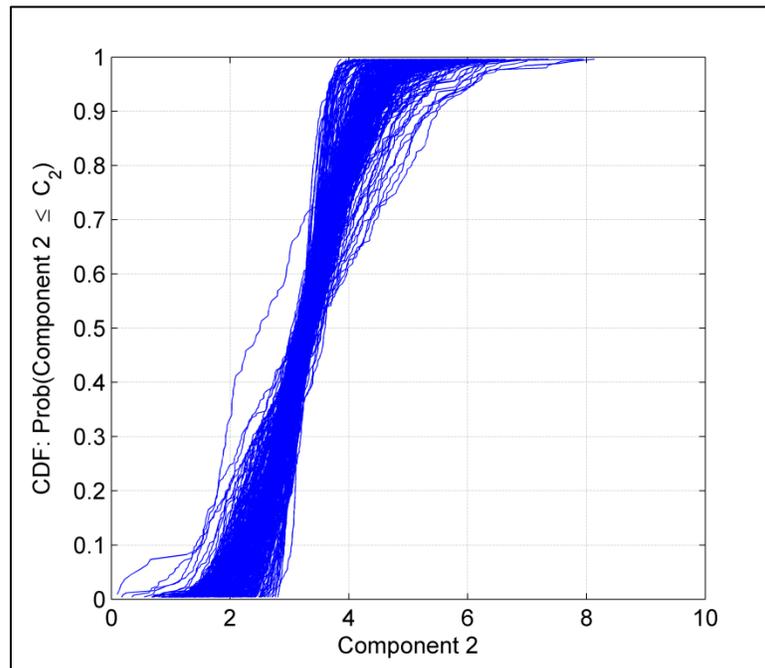


Figure 11: CDFs for all bins of component two for NDBC 46212.

The CDF for C_2 for each bin must be fitted with a distribution. In order to apply the inverse FORM methodology to create an extreme event contour, the parameters for the distribution chosen to fit the binned values of C_2 must be fit as functions of the representative value of C_1 for each bin. Thus, it is important that the distribution chosen to fit the values of C_2 have parameters (e.g., mean and standard deviation) that might be connected to a physical understanding of the problem at hand rather than the shape, scale, and shift parameters that govern many probability distributions. In this manner, an understanding of the relation between the values of C_1 and the overall distribution of C_2 can be used to inform the fitting of C_2 distribution parameters, allowing for the creation of fitting

functions that account for both trends within the binned data and for a consideration of more global trends.

A normal distribution was chosen to fit the distribution of C_2 values for each bin. This fit seems appropriate considering the symmetry of the data, as is shown in Figure 11 above. In addition, the parameters that define the normal distribution are the mean, μ , and standard deviation, σ . These parameters can be used to consider larger trends in the data under consideration in order to inform the construction of their subsequent fitting functions, as was described above. The development of fitting functions for the set of μ and σ values created by fitting a normal distribution to the C_2 CDF for each bin is described in the following section. The CDF for C_2 along with the corresponding normal distribution fit is shown in Figure 12 for a selection of bins.

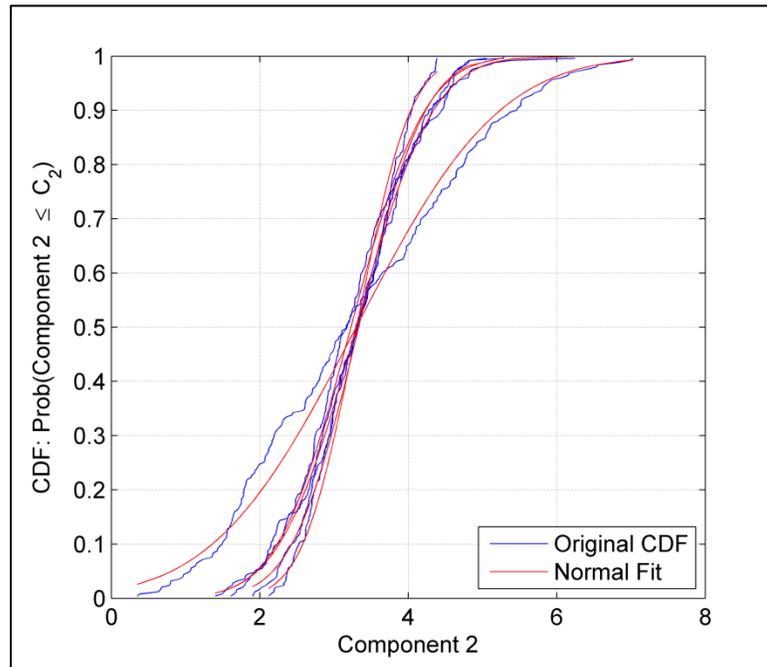


Figure 12: CDFs for selected bins of component two with normal distribution fits for NDBC 46212.

4.1.3 Fitting a function for the variation of mu and sigma

The sets of μ and σ values created following the fitting of a normal distribution to the CDF for the distribution of C_2 values for each bin are represented as functions of the mean value for C_1 for each bin in Figure 13 below.

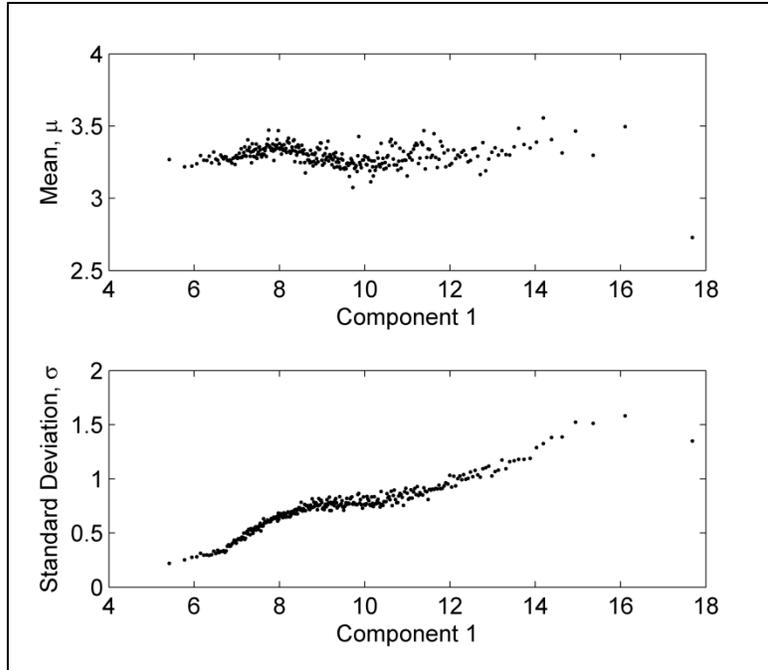


Figure 13: Estimates of the component two normal distribution parameters μ (top) and σ (bottom) as a function of component one for each bin for site 46212.

Based on the trends observed in the data shown above, it was determined that a simple linear approximation could be used to fit μ as a function of C_1 and a quadratic approximation could be used to fit σ . These approximations were fit to the data using a least squares technique applied in Matlab. The fitting functions for μ and σ are as follows:

$$f_{\mu}(C_1) = m_1(C_1) + m_2$$

$$f_{\sigma}(C_1) = s_1(C_1)^2 + s_2(C_1) + s_3$$

The results of these approximations for a selected site are shown in Figure 14.

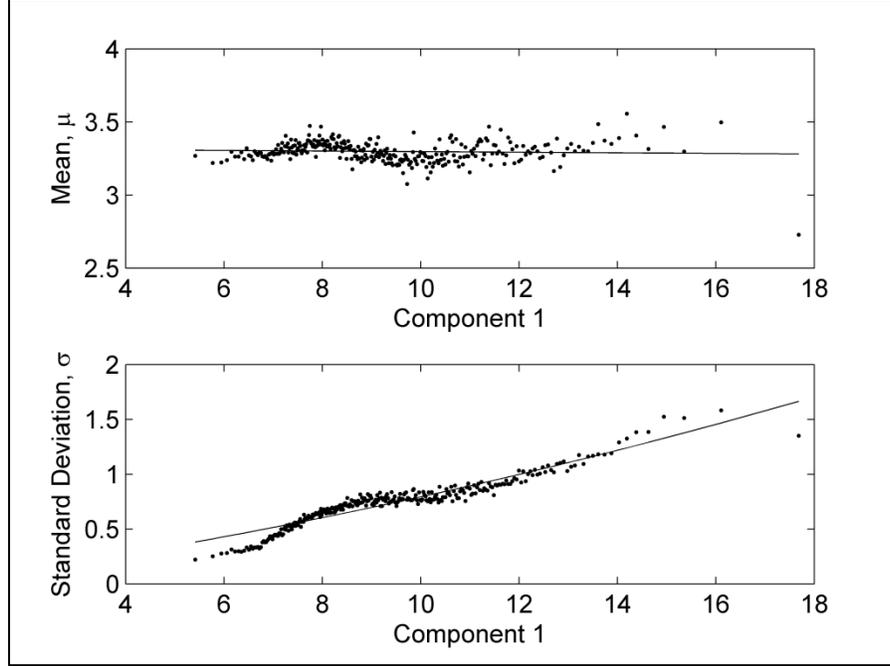


Figure 14: Estimates of component two normal distribution parameters μ (top) and σ (bottom) as a function of component one for each bin with related fitting functions for NDBC 46212.

Although the fittings shown above do not perfectly represent the variations present in the data, the smooth extrapolations that these fitting functions allow for creates a more practically applicable extreme sea state contour. This is especially true when the data is unstructured and may create multimodal distributions, as is seen in Figure 6.

Though a linear fitting for μ does not perfectly represent the variations present in the data, this fitting is theoretically sound considering the implications of applying principal component rotation to the data. With this orthogonal decomposition, the mean of C_2 as a function of C_1 should remain relatively linear, as can be seen in Figure 9.

The quadratic fit for σ was chosen as a step towards ensuring that extrapolation towards smaller values for C_1 , as necessarily occurs in the application of the inverse FORM methodology, does not create negative values. A constraint is placed on the optimization of this fitting function in Matlab, forcing the parameter s_3 (y intercept) to be greater than or equal to zero. It can be easily shown that the minimum of the quadratic fitting function for σ occurs at $x = -s_2/2s_1$. Then, the minimum of the function $f_\sigma(C_1)$ is given by:

$$f(C_1) = s_3 - \frac{s_2^2}{4s_1}$$

Thus, even if s_3 is positive, the minimum value of the function could be negative depending on the values of s_2 and s_3 . While this constraint does not necessarily ensure that the quadratic fitting for σ will remain positive, it seems to be sufficient given the trends observed for the study sites under consideration for the present work.

Additional refinements to the fitting functions for μ and σ remain as an area of improvement for future work.

4.2 Application of inverse FORM methodology

Following the creation of an inverse Gaussian distribution fit for component one and the development of models fitting the parameters of the normal distributions fitting bins of C_2 as a function of C_1 , the inverse FORM method is applied in order to construct an extreme sea state contour for the given return period. This application is performed in the exact manner that is described in section 3.1. The quantile position of the chosen probability of likelihood (in this case 100-years) is calculated for the discretized isoline in both directions in the standard normal space. The resulting quantiles are then evaluated using the inverse CDFs that represent the input distribution, creating the extreme sea state contour for the chosen probability of likelihood in the principal component space. The values for each point on the extreme sea state contour must then be transformed from the principal component space into the original sample space defined by variables H_s and T_e using the methodology described in section 3.2. An example of the resulting extreme sea state contour is shown in Figure 15 below.

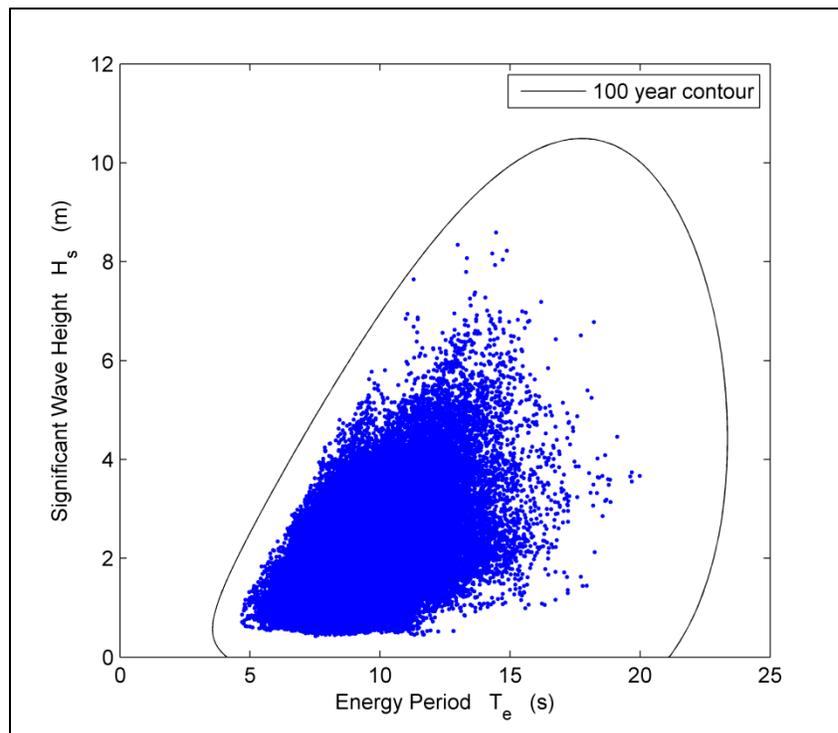


Figure 15: 100-year extreme sea state contour for NDBC 46212.

The rotation of the extreme sea state contour into the original space may create non-physical results if elements of the contour fall below the x axis, indicating that, for a given energy period, a negative value of significant wave height might occur. In order to avoid this non-physical representation of the extreme sea state contour, the contour is truncated such that any elements of the contour including negative values for significant wave height are set to zero.

5 COMPARISON OF RESULTS WITH INITIAL CODE

The results of the extreme contours from four buoy records created using the application of the original methodology are compared with the extreme contours generated using the analysis proposed in this work in the figures below. The extreme contours developed under the new methodology are shown along with a representation of the density of each dataset under consideration, calculated as described in section 2.2.1. This representation helps to emphasize the importance of considering the conjoint influence of energy period and significant wave height in order to create extreme sea state contours that reflect patterns within the data.

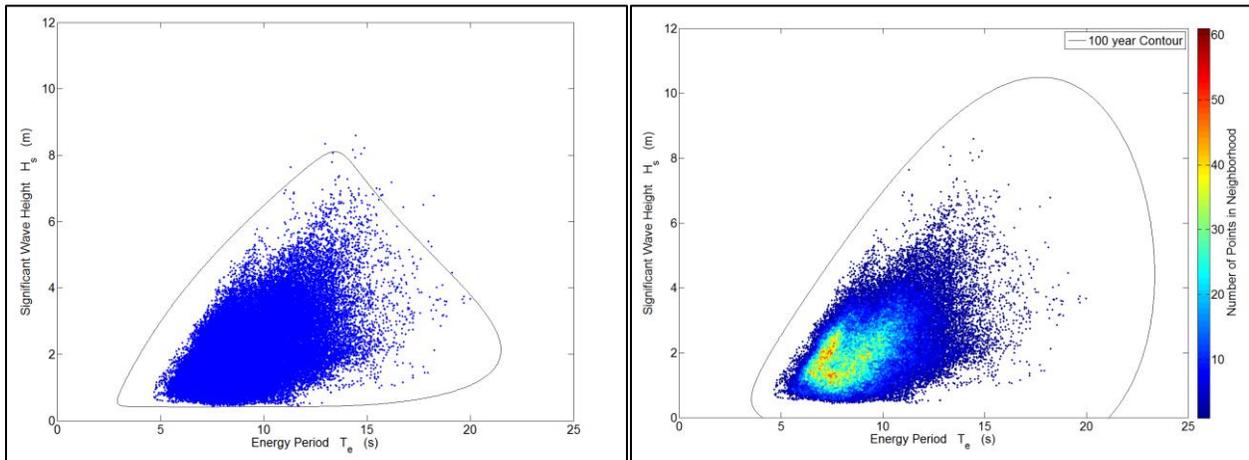


Figure 16: Extreme sea state contour for NDBC 46212 created by the original methodology (left) and the new methodology shown with data density (right).

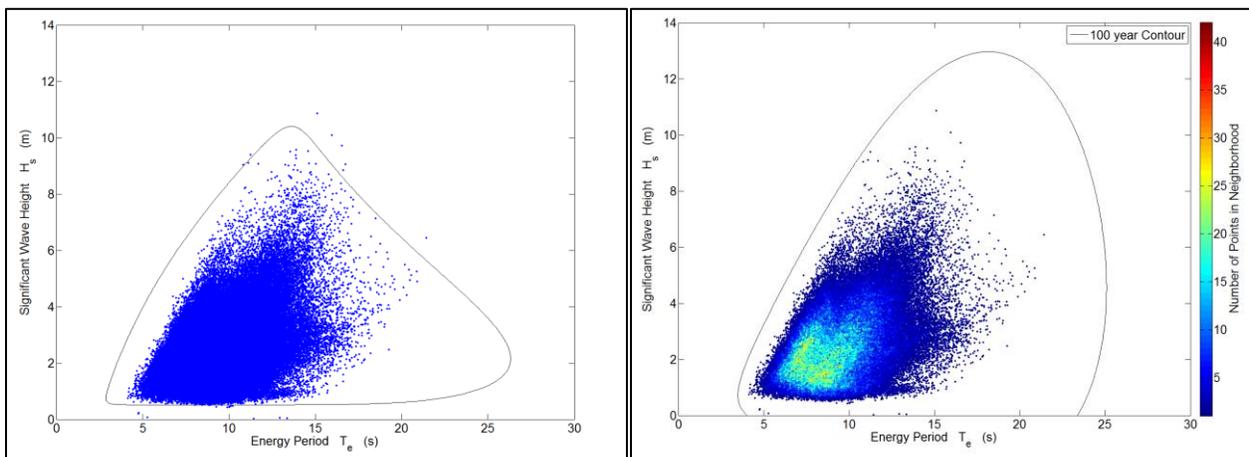


Figure 17: Extreme sea state contour for NDBC 46022 created by the original methodology (left) and the new methodology shown with data density (right).

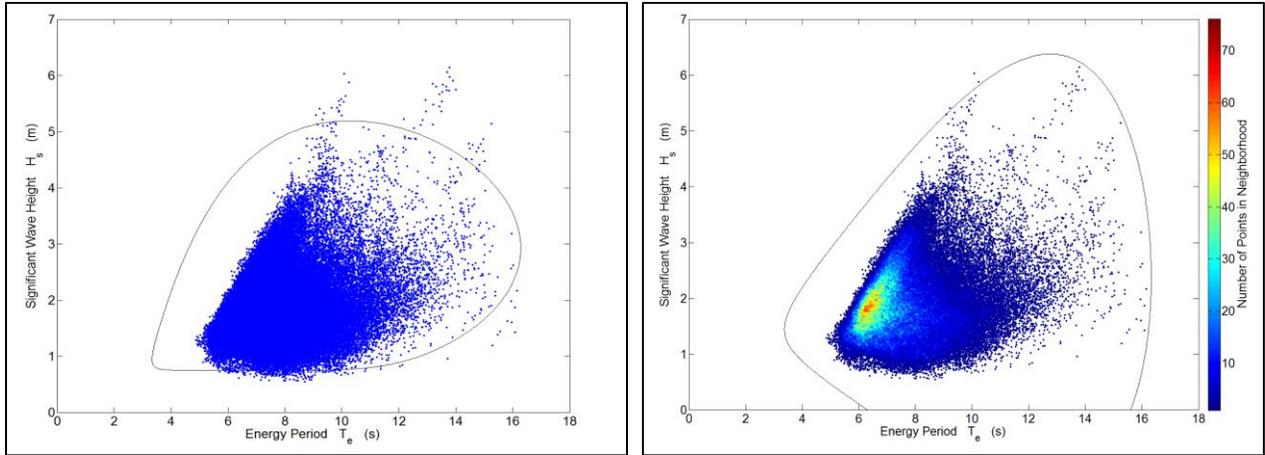


Figure 18: Extreme sea state contour for NDBC 51202 created by the original methodology (left) and the new methodology shown with data density (right).

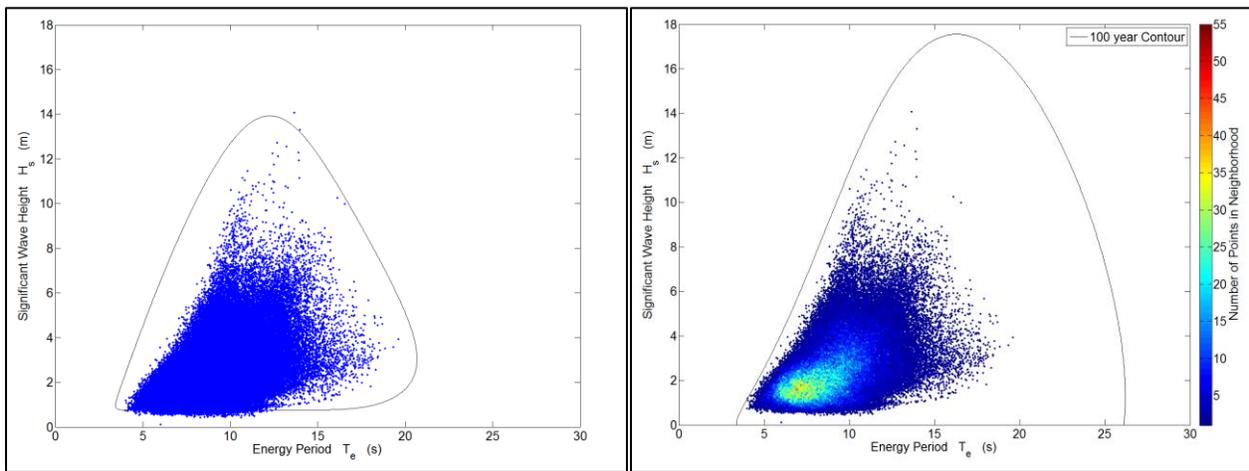


Figure 19: Extreme sea state contour for NDBC 46050 created by the original methodology (left) and the new methodology shown with data density (right).

The extreme sea state contours created using the new methodology appear to follow the shape of trends present within the data, a great improvement upon the contours created using the traditional methodology. This allows for coverage in the area of the input space in which both energy period and significant wave height are high, an area of importance as is discussed in section 2.2.2.

At the most basic level, the extreme sea state contours created using the new methodology create a much better coverage of the data for the given period of record for most of the study sites under consideration. This result is as expected because, given a period of record on the order of tens of years, one would expect the extreme contour for a return period on the order of hundreds of years to include all of the data from the period of record. The exception to this is the contour calculated for NDBC 51202. At this site, the complex relations between H_s and T_e are not entirely captured by the application of principal component analysis and the distribution fitting methodology described in section 3.2. Additional investigation of methodologies that are better able to capture the complex relations found at NDBC 51202 might help to create a better extreme sea state contour for this data set. This remains as an area of future work that is beyond the scope of the current analysis.

6 CONCLUSION AND PERSPECTIVE

The modified version of IFORM developed in this report utilizes several new techniques to generate environmental contours that are more realistic than those created using more traditional methods.

The development of an understanding of trends present within the data under consideration through a representation of density supports the use of principal component analysis. This is reflected in the shape of the extreme sea state contours that are ultimately created in this analysis because these contours follow the overall direction shown in a density-based representation of the study data. Following the application of principal component analysis to transform the original data, the use of an inverse Gaussian distribution to fit the entire distribution of C_1 , the component with the largest variance, creates better fitting for this step in the development of distribution parameters to be used in the inverse FORM process. C_2 is binned based on corresponding sorted values of C_1 using a discrete number of data points, a binning scheme that allows for better coverage across the distribution of this variable. Each bin of C_2 is fit with a normal distribution whose parameters μ and σ are modeled by linear and quadratic functions, respectively, as a function of a representative value of C_1 for each bin. These models allow for an extrapolation in both directions that creates smooth extreme sea state contours when used in the application of the inverse FORM. Finally, the application of the inverse FORM approach and a transformation back into the original input space governed by variables T_e and H_s results in the extreme sea state contour calculated for a given return period.

There are several areas that represent possibilities for future enhancements to the methodology developed in this report. First, the use of principal component analysis to create an orthogonal decomposition of the data such that the values are uncorrelated in each direction only addresses one aspect of the complexity of the relation between energy period and significant wave height, as can be seen in Figure 6 (c) and (d). A more complex decomposition taking into account the varying relations among the study sites shown in these representations of density (e.g., curvature) could lead to a better representation of the data and, therefore, a more accurate approximation of the extreme sea state contour for a given return period. The selection of a more generic distribution or a mixed distribution to fit the variation in behavior for C_1 over the selection of study sites along with refinements in the models developed for the normal distribution parameters μ and σ for C_2 might also contribute to the creation of more accurate extreme sea state contours for all sites.

Overall, this is a significant improvement to the original method of calculating an extreme contour of sea states. The proposed modifications, utilizing principal components, better represent the measured data and provide a more reasonable estimation of environmental contours and extreme sea states. This can better prepare WEC developers for survivability analysis and can be applied to the ship and marine structures industry as well.

This page intentionally left blank.

7 BIBLIOGRAPHY

- Baarholm, G. S., S. Haver, and O. D. Økland. "Combining contours of significant wave height and peak period with platform response distributions for predicting design response." *Marine Structures* 23, 2010: 147-163.
- Berg, J. *Extreme Ocean Wave Conditions for Northern California Wave Energy Conversion Device*. SAND REPORT SAND2011-9034, 2011.
- Coe, R. G., V. S. Neary, M. J. Lawson, Y. Yu, and J. Weber. *Extreme Conditions Modeling Workshop*. Technical Report NREL/TP-5000-62305, SAND2014-16384R, 2014.
- Dallman, A., and V. Neary. "Initial Characterization of the wave resource at several high energy U.S. Sites." *Proceedings of the 2nd Marine Energy Technology Symposium (METS2014)*. Seattle, WA, 2014.
- DNV. *Guidelines on design and operation of wave energy converters*. Det Norske Veritas, The Carbon Trust, 2005.
- DNV. *Offshore service specification - DNV-OSS-312 Certification of Tidal and Wave Energy Converters*. Det Norske Veritas, 2008.
- DNV. *Offshore Standard - DNV-OSF-201 Dynamic Risers*. Hovik Norway: Det Norske Veritas, 2001.
- DNV. *Offshore Standard, DNV-OS-E301 - Position Mooring*. Hovik, Norway: Det Norske Veritas, 2010a.
- DNV. *Recommended practice - DNV-RP-C102 Structural design of offshore ships*. Det Norske Veritas, 2002.
- DNV. *Recommended practice - DNV-RP-C205 Environmental conditions and environmental loads*. Det Norske Veritas, 2014.
- DNV. *Recommended Practice - DNV-RP-F205 Global performance analysis of deepwater floating structures*. Det Norske Veritas, 2010b.
- Forristall, G. Z., and C. K. Cooper. "Design Current Profiles Using Empirical Orthogonal Function (EOF) and Inverse FORM Methods." *Offshore Technology Conference*. Houston, Texas, 1997. 11-21.
- Haver, S, and S.R. Winterstein. "Environmental Contour Lines: A Method for Estimating Long Term Extremes by a Short Term Analysis." *SMTC-067*. 2008.
- Hotelling, H. "Analysis of a complex of statistical variables into principal components." *Journal of educational psychology*, 1933.
- Jackson, J. E. *A User's guide to principal components*. John Wiley and Sons, 1991.
- Johnson, N. L., S. Kotz, and N. Balakrishnan. *Continuous univariate distributions. Vol 1 (2nd ed.)*. New York: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, 1994.
- Lenee-Bluhm, P., R. Paasch, and H. T. Özkan-Haller. "Characterizing the wave energy resource of the US Pacific Northwest." *Renewable Energy* 36, 2011: 2106-2119.
- Liu, P., and A. Der Kiureghian. "Multivariate distribution models with prescribed marginals and covariances." *Probabilistic Engineering Mechanics Vol 1 n. 2*, 1986: 105-112.
- Neary, V. S., et al. *Methodology for Design and Economic Analysis of Marine Energy Conversion (MEC) Technologies*. SANDIA REPORT - SAND2014-9040, 2014.

- Nygaard, E., and K. Johannessen. *Metoccean Design Criteria for Kvitebjorn*. Statoil Report C193-KVB-N-FD-0001, 2000.
- Pearson, K. "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine* 2 (11), 1901: 559-572.
- Rosenblatt, M. "Remarks on a multivariate transformation." *Annals of Mathematical Statistics*, 1952: 470-472.
- Vanem, E., and E.M. Bitner-Gregersen. "Alternative environmental contours for marine structural design - a comparison study." *Proceedings of the ASME 2014 33rd International Conference on Ocean, Offshore and Arctic Engineering*. San Francisco, California, 2014.
- Winterstein, S. R., T. C. Ude, C. A. Cornell, P. Bjerager, and S. Haver. "Environmental Parameters for extreme response: inverse form with omission factors." *ICOSSAR-93*. Innsbruck, AUSTRIA, 1993.
- Zhao, Y.-G., and T. Ono. "A general procedure for First/second-order reliability method (FORM/SORM)." *Structural safety* 21 (2), 1999: 95-112.

8 DISTRIBUTION

3	MS1124	A. Dallman	6122
3	MS1124	V.S. Neary	6122
1	MS0748	J. C. Helton	1341
3	MS0747	C.J. Sallaberry	6224
1	MS0747	R. J. Mackinnon	6224
1	MS0747	R. Dingreville	6233
3	MS0744	A.C. Eckert-Gallup	6233
1	MS0744	P. D. Mattie	6233
1	MS0899	Technical Library	9536 (electronic copy)

This page intentionally left blank.

