ERDC/CHL CHETN-VIII-9
January 2018

US Army Corps
of Engineers®

Identification and Selection of Representative Storm Events from a Probabilistic Storm Data Base

by Mark B. Gravens and Dylan R. Sanderson

PURPOSE: This Coastal and Hydraulics Engineering Technical Note (CHETN) provides guidance on the use of a probabilistic storm data base for the development of environmental forcing information for use in probabilistic life-cycle analysis (PLCA) tools such as Beach-*fx* (Gravens et al. 2007) or Generation Two Coastal Risk Model (G2CRM) (currently under development at the U.S. Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory [ERDC-CHL], and Institute for Water Resources [IWR]). In recent years large probabilistic storm data sets including storm surge hydrographs and coincident wind wave information have been generated using high-fidelity numerical models (e.g., ADCIRC, STWAVE) for purposes of characterizing the storm climatology in support of coastal storm risk assessments including updates of the Federal Emergency Management Agency Flood Hazard Mapping Program and the U.S. Army Corps of Engineers North Atlantic Comprehensive Coastal Study (NACCS). The high-fidelity numerical hydrodynamic and wind wave modeling results from these studies are archived for analysis and project use purposes in the Coastal Hazards System (CHS) (https://chs.erdc.dren.mil/default.aspx). The CHS is a national, coastal, storm-hazard data storage and mining system. It stores comprehensive, high-fidelity, storm-response computer modeling results including climatology, surge, total water levels, and waves as well as measurements. Extremal statistics and epistemic uncertainties on the processes are also stored, and the data are easily accessed, mined, plotted, and downloaded through a user-friendly web interface. The purpose of this CHETN is to provide a methodology to identify and select a small number (12 to 36) of representative storm events from the considerably larger probabilistic storm data base. This paring of the number of storm events is necessary for computational efficiency related to the application of subsequent coastal process response models such as SBEACH (Larson et al. 1990) or STWAVE (Massey et al. 2011).

INTRODUCTION: Beach-*fx* and G2CRM are PLCA tools that implement Monte-Carlo methods to predict the evolution and estimate the uncertainty of a project's physical behavior and economic benefits over the project's life-cycle. These PLCA models are data-driven models and rely on a relational data base that is accessed from within the model's computational kernel (Gravens et al. 2007) to estimate project evolution and the economic consequence of storm events occurring throughout the project life cycle. The relational data bases are populated from external process-based models that compute storm responses needed by the PLCA model. For example, in Beach-*fx*, a beach profile response to each event in the storm suite is required for the full range of pre-storm upper beach configurations (dune height, dune width, berm width) that may be encountered throughout the simulated life cycle (typically on the order of 200 unique upper beach profiles). Separately, each storm surge hydrograph must be combined with 12 representations of the astronomical tide to cover variations in tidal range and phasing of the storm surge hydrograph with the tide signal. Given that approximately 200 unique beach profiles are required, and that each storm is combined with 12 representations in tide, limiting the number of unique storm surge hydrographs in the storm suite is necessary for computational efficiency purposes.

The typical Beach-*fx* shore response database (SRD) includes between 30,000 to 90,000 unique storm/profile responses for each representative submerged profile; project study areas typically involve from 3 to 9 representative submerged profiles resulting in an SRD containing somewhere between 90,000 and 810,000 unique storm/profile responses. If the full probabilistic storm population were used, which may involve 500 or more storms, this number would geometrically increase to between 3.6 and 10.8 million unique storm/profile responses and makes the problem untenable from a computational perspective. For these reasons it is necessary that a limited number of selected representative storms be identified and selected from the probabilistic storm data base for use in PLCA models. Guidance for the identification and selection of representative storms from a probabilistic data base is provided in procedures outlined in the following sections.

METHOD: The process of identifying and selecting representative storms can be divided into four steps:

- Group storms into clusters based on the magnitude of the peak surge generated.
- Further sub-divide storm clusters if appropriate, based on duration of storm surge hydrograph or low versus high peak amplitude.
- Select representative storm events within each storm cluster.
- Assign appropriate relative probability to each selected representative storm.

Because the overall frequency of storm occurrence is spatially dependent, the first step in identifying representative storms is to identify from the available population of storms those storms whose track passes within a 200 kilometer (km) radius of the project study area. The CHS provides a unique estimate of frequency of storm occurrence for both low-intensity and high-intensity storms as part of the attribute information for each save point in the data base. An example of this step is illustrated in Figure 1, which shows the NACCS storm tracks (white lines) and the tracks of those storms that pass within 200 km of an example project location (red lines). Use of a 200 km radius centered on the project study area is consistent with the methodology used to compute to the frequency of storm occurrence rates at the NACCS save points (Nadal-Caraballo et al. 2015). Storm frequency becomes input to the PLCA tools and is used to control the sampling of storm events in the Monte Carlo simulation.

The second step involves bin sorting the storms identified in the first step into clusters of storms based on the peak surge generated by the storm event. The number of clusters and the specific peak storm surge limits defining each cluster of storms are arbitrary, and in this procedure the choice was made to use the available stage frequency relationship (annual exceedance probability) to define the clusters. The CHS provides an estimated stage frequency relationship at each save point in the NACCS database. The annual exceedance probability for tropical storms at ADCIRC save point 6125 of the NACCS database is provided in Table 1. This save point is located in Tangier Sound of the Chesapeake Bay near Crisfield, MD, and will be used to illustrate the representative storm selection procedure. The data at ADCIRC save point 6125 will also be used to investigate the influence of using a representative storm suite versus the full probabilistic storm suite in PLCA models.

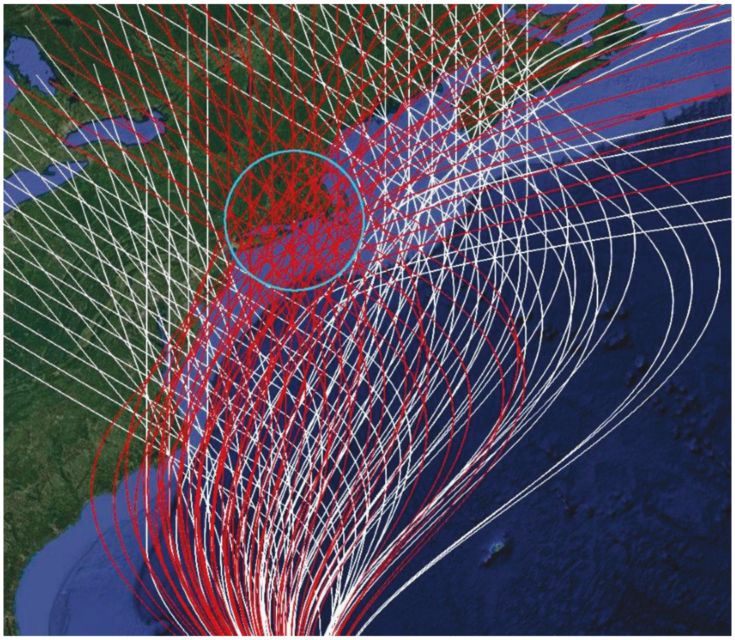

Figure 1. NACCS storm tracks and identification of project relevant storm tracks of interest.

Table 1. Annual exceedance probability for ADCIRC save point 6125 (tropical storm surge with tide).

Return Year	Stage (meters, MSL*)
1	-
2	0.73
5	0.89
10	1.04
20	1.22
50	1.49
100	1.68
200	1.88
500	2.13
1000	2.30
2000	2.46
5000	2.64
10000	2.77

*mean sea level

Each storm is assigned to one of the storm clusters by determining the storm's peak surge elevation (from the storm surge hydrograph) then adding a statistically determined mean high tide amplitude to the peak surge value (adding a representative tidal contribution to the peak surge elevations is necessary because the CHS annual exceedance probability curve includes tidal contributions to total stage) and assigning the storm to the cluster whose upper and lower limits bound the computed peak surge plus tide value. The lower and upper limits of the storm clusters are computed as the midpoint value between the current storm cluster stage value and the adjacent (lower and upper) storm cluster stage value. Table 2 lists the stage plus tide boundaries of the storm clusters in this example for

NACCS ADCIRC save point 6125. Note that the first storm cluster starts at the 2-year return interval and the lower stage value for this storm cluster is assigned a value of 0.5 meter (m) because a flood stage below this level at the study area is not damage producing and consequently not of interest for the PLCA purposes. The twelfth storm cluster, corresponding to the 10,000-year return interval, contains all storms producing a stage of at least 2.70 m.

Table 2. Lower and upper stage limits for ADCIRC save point 6125 (surge and tide).

Lower Stage (m, MSL)	Return Year/Cluster Number	Upper Stage (m, MSL)
0.50	2 / 1	0.81
0.81	5 / 2	0.96
0.96	10 / 3	1.13
1.13	20 / 4	1.35
1.35	50 / 5	1.59
1.59	100 / 6	1.78
1.78	200 / 7	2.01
2.01	500 / 8	2.21
2.21	1000 / 9	2.38
2.38	2000 /10	2.55
2.55	5000 / 11	2.70
2.70	10000 / 12	

After the storms have been assigned to the storm clusters, the storm surge hydrograph time series of all storms in each cluster are translated along the time axis such that peak surge is aligned at a single time. Examples of this are shown in Figure 2 for the 10-year and 50-year return interval storm clusters for NACCS ADCIRC save point 6125.

Based on visual inspection of the aligned storm surge hydrographs, the storms in the storm cluster may be further divided into sub-clusters to capture different storm surge hydrograph characteristics present in the original storm cluster. That is, if there is a notable population of both long- and short-duration hydrographs in the storm cluster, then the storm cluster should be split into two separate storm clusters segregating the long duration hydrographs from the short duration hydrographs. Likewise, if the storm cluster contains a large number of storms as occurs for high-frequency storm events, the storm cluster can be sub-divided again based on peak surge generated by the storms. In this example, the 10-year return interval storm cluster contains 56 storm events and was sub-divided into 2 sub-clusters, one involving long-duration storms (12 storms) and another involving short-duration storms (44 storms). One storm was selected to represent the 12 long-duration storms (Figure 3-a), and 3 storms were selected to represent the 44 short-duration storms (Figure 3-b), one with a peak storm surge near the high-surge boundary, one with a peak surge in the middle of the surge range, and another with a peak surge near the low-surge boundary. Likewise, the 50-year return interval storm cluster contains 25 storms events and was sub-divided into 2 sub-clusters, one involving long-duration storms (5 storms) and another involving short-duration storms (20 storms). One storm was selected to represent the 5 long-duration storms (Figure 3-c), and one was selected to represent the 20 short-duration storms (Figure 3-d). The thick color hydrographs in Figure 3 denote the selected representative storm(s) in the illustrated storm cluster.

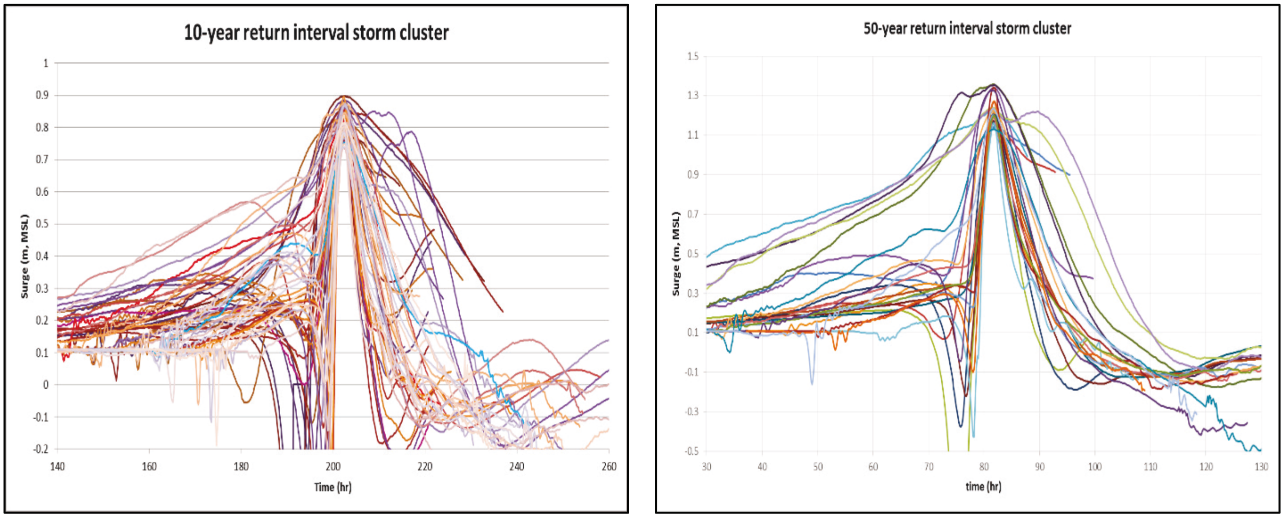

Figure 2. 10- and 50-year return interval storm clusters; storm surge hydrographs aligned at peak surge.

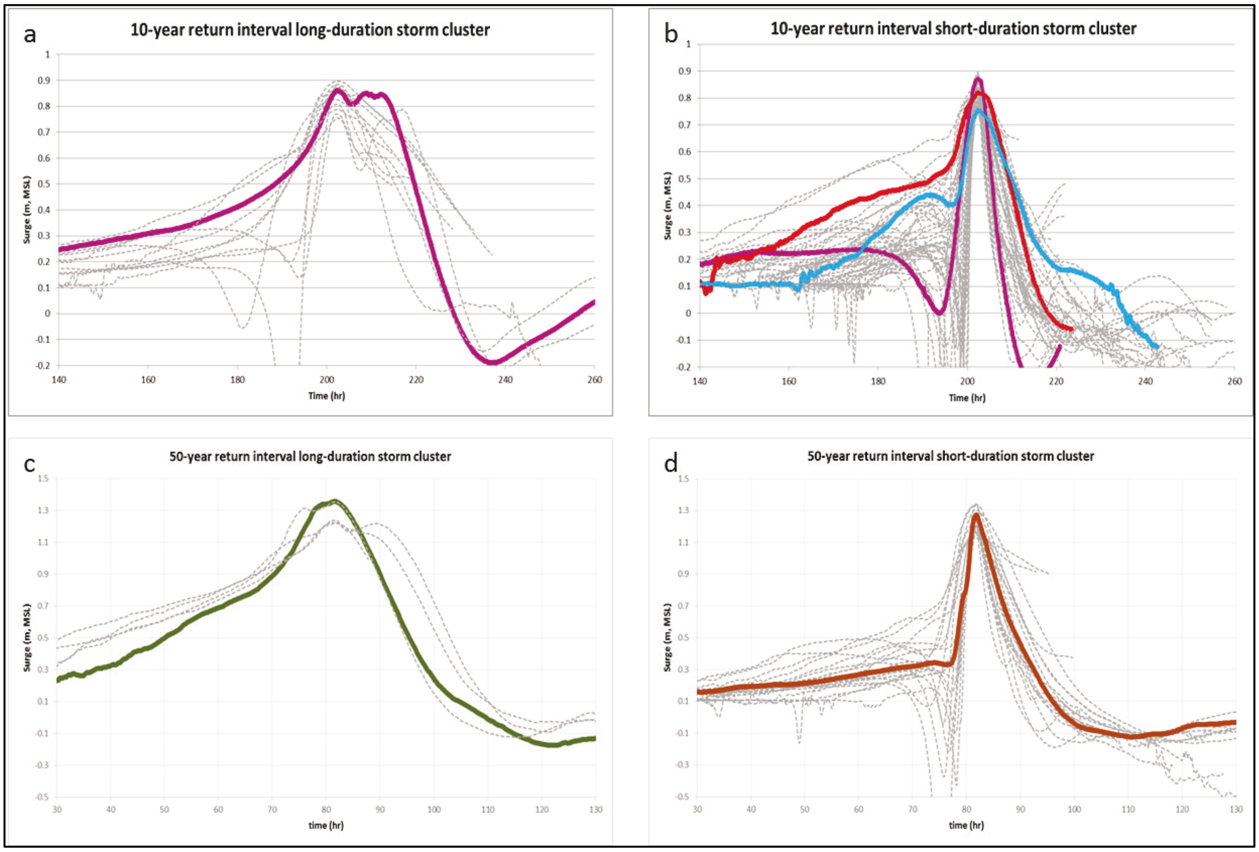

Figure 3. Selected representative storms for the 10- and 50-year return interval storm clusters. (a) 10-year return interval, long-duration storms; one representative storm, (b) 10-year return interval, short-duration storms; three representative storms for low, medium, and high amplitude, (c) 50-year return interval, long-duration storms; one representative storm, (d) 50-year return interval, short-duration storms; one representative storm.

Selection of the representative storm or storms within the storm clusters using the methodology outlined herein is subjective in that the analyst ultimately selects a representative storm or storms from each of the developed storm clusters. The analyst makes the selection based on judgment of which storm surge hydrograph best characterizes the collection of storm surge hydrographs in the storm cluster. In making the representative storm selection, the analyst should consider both the peak surge value and the overall form or shape of the storm surge hydrograph, aiming to select a storm surge hydrograph that reasonably characterizes the overall shape and duration of the surge hydrographs in the storm cluster. The peak surge produced by the selected storm should be near the median peak surge generated by the collection of hydrographs in the storm cluster.

The final step in this procedure is calculation of the relative probability to be assigned to each of the selected representative storms. This information is required input to PLCA models that will use the specified relative probability data when sampling specific storms from the input storm suite. The relative probability of a selected representative storm is computed by summing the relative probabilities of all the storms in the storm cluster that storm represents.

Continuing with the example using NACCS ADCIRC save point 6125, Table 3 lists the selected representative storms, the computed relative probability, and the number of storms represented for each of the storm clusters listed in Table 2. The maximum peak surge plus mean high tide amplitude fell into the ninth storm cluster; thus, storm clusters 10–12 are not populated with storms. A total of 22 representative storms was identified and selected from the full probabilistic storm suite of 355 unique storms whose track passes within 200 km of the example study area. Note that the full NACCS probabilistic tropical storm suite involves 1050 storms; of those storms, 355 pass within 200 km of the save point of interest and are identified as save-point-relevant storms.

The methodology outlined above for the identification and selection of representative storms from the probabilistic tropical storm suite is similarly followed for extratropical storms. However, because the NACCS extratropical storm suite is comprised of a collection of 100 historical storm events spanning a 75-year period and are not probabilistic synthetic storms (as is the tropical storm suite), the step of eliminating storms that do not pass within 200 km of the save point is not applied in the processing of the extratropical storms. For the extratropical storms, the first step involves eliminating from consideration all storms that produced a peak surge of less than 0.5 m (the threshold for a damage-producing storm event at the study area). For NACCS ADCIRC save point 6125, this step reduced the storm suite of interest from 100 to 90 storms. Extratropical storm frequency is computed by dividing the number of storms by the period of record resulting in an average extratropical storm frequency of 1.2 storms per year. The remainder of the identification and selection of representative extratropical storms is the same as that for the synthetic probabilistic tropical storms. The calculation of the relative probability for the extratropical storms, however, is different because the extratropical storm suite was compiled based on a peaks over threshold analysis. As such, the extratropical storms are presumed to have an equal relative probability. The embedded assumption here is that the 75-year historical record captures the full probability space. Consequently, the relative probability of the selected representative extratropical storms is equal to the number of storms the selected storm represents divided by the total number of storms passing the peaks over threshold analysis (90 storms in this example). Table 4 lists the selected representative extratropical storms, the computed relative probability, and the number of storms represented by the selected storm.

Table 3. Selected representative tropical storms and relative probabilities for ADCIRC save point 6125.

Cluster Number	Representative Storm Number	Relative Probability	Number of Storms Represented	Note
1	240	0.0364289	33	1
1	393	0.0384474	40	2
1	647	0.0469166	62	3
2	198	0.0015040	10	4
2	180	0.0115260	18	6
2	473	0.0115260	18	7
2	1007	0.0115260	18	8
3	277	0.0049432	12	4
3	314	0.0090688	14	6
3	174	0.0097165	15	7
3	92	0.0097165	15	8
4	315	0.0027349	9	4
4	136	0.0186826	35	5
5	195	0.0003630	5	4
5	278	0.0124107	20	5
6	284	0.0006979	8	4
6	524	0.0041390	10	5
7	107	0.0000612	1	4
7	114	0.0025322	7	5
8	271	0.0000775	1	4
8	625	0.0004700	2	5
9	623	0.0006360	2	

Notes

1 – low-amplitude storms
2 – mean-amplitude storms
3 – high-amplitude storms
4 – long-duration storms
5 – short-duration storms
6 – short-duration, low-amplitude storms
7 – short-duration, mean-amplitude storms
8 – short-duration, high-amplitude storms

Table 4. Selected representative extratropical storms and relative probabilities for ADCIRC save point 6125.

Cluster Number	Representative Storm Number	Relative Probability	Number of Storms Represented	Note
1	92	0.1000	9	
2	26	0.1778	16	1
2	47	0.1333	12	2
2	5	0.1222	11	3
3	83	0.1556	14	1
3	15	0.1111	10	3
4	29	0.0556	5	4
4	20	0.0556	5	5
5	1	0.0556	5	
6	60	0.0333	3	

Notes

1 – low-amplitude storms
2 – mean-amplitude storms
3 – high-amplitude storms
4 – long-duration storms
5 – short-duration storms

RESULTS: To investigate the influence of using a representative storm suite as opposed to using the full probabilistic storm suite, G2CRM model simulations were performed, and peak total water levels generated through the Monte Carlo simulation were used to generate a stage frequency relationship. The stage frequency relationship derived from model simulations involving the representative storm suite was compared to the stage frequency relationship derived from model simulations involving the full probabilistic storm suite. Finally, to validate both the representative storm selection procedure and the random storm generation and sampling methods internal to G2CRM, the stage frequency relationships derived from the G2CRM model simulations were compared to the independently developed stage frequency relationship provided in the CHS database.

G2CRM was configured to simulate 75 iterations of a 216-year life cycle, and the randomly sampled storm surge hydrographs were linearly combined with predicted astronomical tides for the random date and time at which the selected storm occurred in the simulation to obtain a total water surface elevation time series. Note that the intra-storm tide is calculated for each time period of the selected storm hydrograph. The peak stage of the total water surface elevation time series was recorded and treated as a gauge observation in a subsequent analysis to estimate the corresponding stage frequency relationship. This procedure resulted in the generation of 75 stage frequency relationships from which a mean stage frequency curve could be generated.

The procedure used to compute the stage frequency relationship of each of the simulated 216-year duration life cycles followed the procedure outlined in Scheffner et al. (1999) and involves first computing the cumulative distribution function (cdf) of the simulated annual maximum stage and then using the computed cdf to interpolate the stage corresponding to annual *n*-year return period events. The cdf is empirically estimated using the Weibull formula:

$$F_X\left(x_{(r)}\right) = \frac{r}{(n+1)} \quad (1)$$

where $F_X(x_{(r)})$ is the cumulative distribution function value, $x_{(r)}$ is the rank-ordered stage elevation (smallest to largest), r is the rank, and n is the number of observations (in this case 216). The cdf value of an n-year return period event (in this case stage) is calculated using the relationship

$$F\left(x_{(n)}\right) = 1 - \frac{1}{n} \quad (2)$$

The computed mean stage frequency based on G2CRM simulations using the representative storm suite and the full probabilistic storm suite, as well as the curve available for ADCIRC Station 6125 of the NACCS database, are plotted in Figure 4. This figure shows that the use of a representative storm suite identified as outlined above, in a PLCA model, results in a nearly identical stage-frequency response out to approximately the 200-year return period as does the use of the full probabilistic storm suite.

Furthermore, it is seen that the NACCS stage frequency curve is above the G2CRM Monte Carlo sampling curve by approximately 0.5 foot (ft) at return periods between 2 and 20 years and increases to nearly 1 ft at the 200-year return period. The NACCS stage frequency curve is computed by integration of the joint probability method (JPM) integral (Nadal-Caraballo et al. 2015). The G2CRM Monte Carlo sampling stage frequency was computed as described above. The differences between the estimated stage frequency curves using these two methods are believed to stem from two sources. The first source of the difference is the JPM integral includes an uncertainty term that contributes to stage elevation when the integration is performed, and the second source of the difference is that the integration is performed across all of the NACCS storms whereas in the Monte Carlo simulation technique, storm selection is limited to only those storms whose track passes within a 200 km radius of the save point.

Although there are differences between the NACCS and the G2CRM sampled curves, a comparison of the curves resulting from the full probabilistic and the representative storm suites provides technical justification for the use of the latter, in addition to the computational justification discussed in the Introduction.

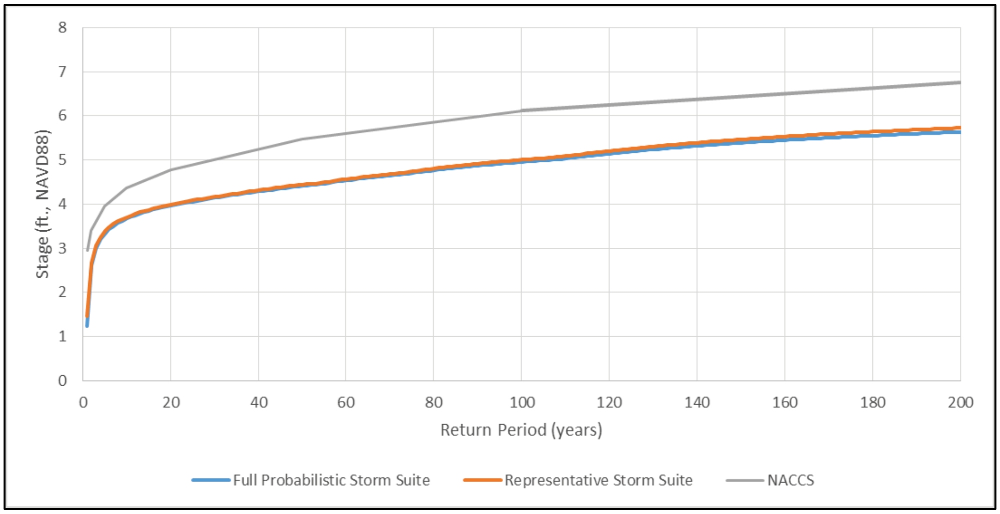

Figure 4. Comparison of stage frequency relationships. Full probabilistic storm suite, representative storm suite, and NACCS.

ADDITIONAL INFORMATION: This Coastal and Hydraulics Engineering Technical Note (CHETN) was prepared by Mark Gravens and Dylan Sanderson, Dylan.R.Sanderson@usace.army.mil, U.S. Army Engineer Research and Development Center. The CHS development team (Dr. Jeffery Melby, Dr. Norberto Nadal-Caraballo, Victor Gonzales Nieves, and Fatima Diop) contributed substantially to the authors' understanding of the various data attributes stored in the CHS database and provided constructive feedback to the authors on the methodology and procedures documented in this CHETN. The study is funded by the USACE Flood Risk Management Research Program. This technical note should be cited as follows:

> Gravens, M. B., and D. R. Sanderson. 2018. *Identification and Selection of Representative Storm Events from a Probabilistic Storm Data Base*. ERDC/CHL CHETN-VIII-9. Vicksburg, MS: U.S. Army Engineer Research and Development Center. http://dx.doi.org/10.21079/11681/26341.

REFERENCES

Gravens, M. B., R. M. Males, and D. A. Moser. 2007. "Beach-*fx*: Monte Carlo Life-Cycle Simulation Model for Estimating Shore Protection Project Evolution and Cost Benefit Analyses." *Shore and Beach* 75(1) Winter 2007: 12–19.

Larson, M., N. C. Kraus, and M. R. Byrnes. 1990. *SBEACH: Numerical Model for Simulating Storm-Induced Beach Change, Report 2 – Numerical Formulation and Model Tests*. Technical Report CERC-89-9. Vicksburg, MS: U.S. Army Engineer Waterways Experiment Station. http://hdl.handle.net/11681/12475.

Massey, T. C., M. E. Anderson, J. M. Smith, J. Gomez, and R. Jones. 2011. *STWAVE: Steady-State Spectral Wave Model User's Manual for STWAVE, Version 6.0*. ERDC/CHL SR-11-1. Vicksburg, MS: U.S. Army Engineer Research and Development Center. http://hdl.handle.net/11681/1850.

Nadal-Caraballo, N. C., J. A. Melby, V. M. Gonzalez, and A. T. Cox. 2015. *Coastal Storm Hazards from Virginia to Maine*. ERDC/CHL TR-15-5. Vicksburg, MS: U.S. Army Engineer Research and Development Center. http://hdl.handle.net/11681/7715.

Scheffner, N. W., J. E. Clausner, A. Militello, L. E. Borgman, B. L. Edge, and P. J. Grace. 1999. *Use and Application of the Empirical Simulation Technique: User's Guide*. Technical Report CHL-99-21. Vicksburg, MS: U.S. Army Engineer Research and Development Center. http://hdl.handle.net/11681/7478.