Probabilistic assessment of erosion and flooding risk in the northern Gulf of Mexico

Thomas Wahl[1], Nathaniel G. Plant[2], and Joseph W. Long[2]

[1]College of Marine Science, University of South Florida, St. Petersburg, Florida, USA

[2]USGS Coastal and Marine Geology Program, St. Petersburg Coastal and Marine Science Center, St. Petersburg, Florida, USA

Corresponding author: T. Wahl, College of Marine Science, University of South Florida, 140 7th Avenue South, St. Petersburg, FL 33701, US, thomaswahl@mail.usf.edu

Key points

- Sea-storm time series are simulated with a multivariate probabilistic model
- Erosion and flooding risk are assessed accurately with a joint probability approach
- Return water levels and impact hours could be larger than recently observed

Abstract

We assess erosion and flooding risk in the northern Gulf of Mexico by identifying interdependencies among oceanographic drivers and probabilistically modeling the resulting potential for coastal change. Wave and water level observations are used to determine relationships between six hydrodynamic parameters that influence total water level and therefore erosion and flooding, through consideration of a wide range of univariate distribution functions and multivariate elliptical copulas. Using these relationships, we explore how different our interpretation of the present-day erosion/flooding risk could be if we had seen more or fewer extreme realizations of individual and combinations of parameters in the past by simulating 10,000 physically and statistically consistent sea-storm time series. We find that seasonal total water levels associated with the 100-year return period could be up to 3 m higher in summer and 0.6 m higher in winter relative to our best estimate based on the observational records. Impact hours of collision and overwash – where total water levels exceed the dune toe or dune crest elevations – could be on average 70% (collision) and 100% (overwash) larger than inferred from the observations. Our model accounts for non-stationarity in a straightforward, non-parametric way that can be applied (with little adjustments) to many other coastlines. The probabilistic model presented here, which accounts for observational uncertainty, can be applied to other coastlines where short record lengths limit the ability to identify the full range of possible wave and water level conditions that coastal mangers and planners must consider to develop sustainable management strategies.

Key words

Multivariate sea-storm model; elliptical copulas; coastal erosion and flooding; northern Gulf of Mexico

1. Introduction

Erosion and flooding occur on sandy coastlines when the total water level (TWL) exceeds critical thresholds of backshore features. *Ruggiero* [2013] defined TWL as the superposition of astronomical tide (η_A), storm surge (or non-tidal residual; η_{NTR}), and the extreme wave runup statistic (R2%; e.g., *Stockdon et al.* [2014]), all of which can be derived with various numerical or empirical models. The impacts of extreme oceanographic events, in terms of erosion of barrier islands and sandy beaches and flood damages in low-lying coastal areas, also strongly depend on how long critical TWL thresholds are exceeded (i.e. the event duration is important). For long-term simulations of erosion, the duration of calm periods between successive sea-storm events are also relevant since they determine how much the beach or dune can recover before the next extreme event occurs.

All of the above-mentioned variables are modulated by climate variability, including climate change (e.g., trends), introducing non-stationarity into the system. For our case study site, Dauphin Island in the northern Gulf of Mexico, *Wahl and Plant* [2015] (hereafter referred to as WP15) found, for example, significant trends in mean sea level (MSL), significant wave height (Hs), and peak wave period (Tp) between 1980 and 2013; this led to an increase in the erosion and flooding risk (from hereon we refer to erosion risk for both) by ~30% over this three-decade period. WP15 also reported significant changes in the amplitudes of the seasonal cycles of MSL and Hs, resulting in an additional increase in erosion risk in summer of ~30% and similar decrease in winter. For the next 30 years they projected that erosion risk may increase by up to 300% under a high sea-level rise scenario and assuming that observed trends in wave parameters continue. In the present study we shift our focus from “climate” to the role of “weather” (and the associated sea-state conditions), acting on much shorter time-scales than the seasonal to decadal changes considered in WP15. We do this by analyzing individual extreme oceanographic events observed within the 1980 and 2013 time period.

Observational records are often used to quantify erosion risk; for example by determining impact hours [*Rugiero*, 2001] when TWLs resulted in collision (where waves reach dunes), overwash (where waves overtop dunes), or inundation (where the wave-averaged water level exceeds the dune crest elevation), according to the storm impact scales defined by *Sallenger* [2000]), or by performing extreme value analysis on TWL time series. However, observations comprising simultaneous water levels and waves sample a limited number of locations and limited periods of time, ranging from several days or weeks for the most detailed studies to a few decades for long-term stations. Hence, we should not assume that we have already seen the highest possible realizations of the individual variables contributing to TWL and erosion, nor should we assume that we have seen all possible extreme event combinations, and we expect that there will be uncertainty in any estimate of future extreme values of erosion [*Serafin and Ruggiero*, 2014]. We explicitly account for this by assessing the erosion risk in the northern Gulf of Mexico in a probabilistic way by developing and applying a multivariate sea-storm model (MSSM). Such a model should account for the non-stationarity in the different variables and the dependencies, represented through joint correlations, between variables.

Several authors used statistical approaches to determine joint probabilities of multivariate sea-storms or to investigate past and/or future erosion risk at different coastline stretches around the globe: e.g., *DeMichele et al.* [2007] for Italy, *Callaghan et al.* [2008] for south-east Australia, *Wahl et al.* [2012] for the German Bight, *Corbella & Stretch* [2012a, 2012b, 2013] for South Africa, *Li et al.* [2014a, 2014b] for the Dutch coast, and *Serafin and Ruggiero* [2014] for Oregon on the north-west coast of the United States. The models considered in those studies all differ in the way they account for non-stationarity and/or interdependencies between variables and also in their definition of "sea-storm events". Here we make use of these earlier applications and develop a generic model that is functionally similar to several of the earlier models. Our approach is unique in the sense that it is developed for, and applied to, a region that experienced significant long-term changes in seasonal cycles of sea level and wave heights [*Wahl et al*., 2014; WP15]. Including this form of non-stationarity, in addition to other variations and trends, is important because it affects vulnerability estimates of sandy beaches, dunes, and built infrastructure that are threatened when specific morphological elevation thresholds are exceeded. While driven by different meteorological forcing, these elevation thresholds can be exceeded by both tropical and extra-tropical events depending on the superposition of waves and water levels; hence models must be capable of considering both types of storms. Specifically, because extreme-value distributions are fit to historical data sets, estimates of vulnerability depend on the actual extreme events that were observed. We assume that these events were drawn from some random distribution and that, in an alternate realization of our universe, a different set of events would have been observed. And, in the future, different events will be observed.

The different steps that are involved in the data pre-processing, the model development, and (selected) model applications are summarized in Figure 1 and described in more detail in the following sections. In section 2 we describe the available observational data for a case study site (Dauphin Island, Alabama) and summarize the different steps of the MSSM development in Section 3. In Section 4 we apply the model with the main objective of exploring how different our interpretation of the present-day erosion risk in the northern Gulf of Mexico (defined through TWL return periods and impact hours of collision and overwash under stationary morphological conditions) could be if more or less extreme realizations of individual variables and/or their combinations had occurred in the observational period or if we had much longer data sets available. The results are briefly discussed in Section 5 and conclusions drawn in Section 6.

2. Data

Our case study site is Dauphin Island, a barrier island off the coast of Alabama in the northern Gulf of Mexico. The oceanographic data we use for the model development were the same that were used in WP15 and we refer to this earlier paper for details on how the time series of the different variables were derived. The final data set comprises hourly records of water levels (at a tide gauge on Dauphin Island) and the wave parameters Hs, Tp, and direction θ (at a wave buoy offshore in 28m water depth) for the period 1980 to 2013 (Figures 2a to 2d). Those variables (except for θ) exhibit significant decadal trends, inter-annual variability, and seasonal cycles whose amplitudes also changed through time in case of MSL and Hs (see WP15). We account for this non-stationarity by removing 30-day running medians (shifted by one hour each time step) from the hourly time series of water levels, Hs, and Tp (blue lines in Figures 2a to 2c). We then apply parallel offsets so that the medians of the last three years in the observed and "corrected" (de-trended and de-seasonalized) time series are the same and the corrected time series is representative of the recent climate. This non-parametric approach to account for non-stationarity is straightforward and effectively removes all linear and non-linear, long-term, and cyclic trends, whose role in altering the erosion risk was already assessed in WP15. The corrected time series fulfill the stationarity criteria for the subsequent statistical analysis and trends and variability can easily be re-included at a later step in the model application. Alternatively, one can model the non-stationarity through parametric functions within the statistical analysis [e.g., *Méndez et al*., 2006; *Serafin and Ruggiero*, 2014]; this may, however, introduce additional uncertainties (especially when the selected functions represent a poor fit) because it requires the estimation of more parameters, particularly in our case where changing seasonal cycles have been observed.

From the corrected observational records we obtain η_{NTR} and η_A by performing a year-by-year tidal analysis with the Matlab t_Tide package [*Pawlowitz et al.*, 2002] (Figures 2e and 2f). We also determine R2% with the empirical formulation of *Stockdon et al.* [2006]:

$$R_{2\%} = 1.1\left\{0.35\tan\beta(H_0L_0)^{1/2} + \frac{\left[H_0L_0\left(0.563\tan\beta^2+0.004\right)\right]^{1/2}}{2}\right\} \quad (1)$$

where tanβ represents the foreshore beach slope, H_0 is the offshore significant wave height Hs, and L_0 is the offshore wave length given by Airy's linear wave theory as $(g/2\pi)Tp^2$. We use the average present-day beach slope on Dauphin Island of 0.07 (the spatial variability ranges from 0.04 to 0.18) to derive an hourly R2% time series which is superimposed onto the observed water levels to obtain TWL (Figure 2h). We want the individual sea-storm events used in the development of the MSSM to be representative for the entire barrier island and therefore use the average beach slope at this stage of the analysis. Later, in the model application when we calculate impact hours for Dauphin Island (Section 4) we use detailed spatially variable morphological data, including foreshore beach slopes and elevation of the dune toe and dune crest. The data were extracted using a standard methodology [*Stockdon et al.,* 2009] applied to a lidar survey data set conducted by the U.S. Geological Survey in July 2013 [*Guy and Plant,* 2014] and were smoothed and interpolated in the alongshore direction every 10 meters as described in WP15. Dune toe and crest heights are used here as proxies for backshore elevations that are relevant to assess coastal erosion risk. Water levels that exceed the dune-toe elevation are required to initiate dune erosion whereas water levels that exceed the dune crest elevation lead to potential changes in the dune position, dune height, and, due to overwash, changes in the island morphology landward of the dune. In this study we want to explore only the effects of altering the oceanographic forcing variables; therefore we assume stationary present-day morphological conditions throughout the analysis.

3. Model development and validation

3.1 Threshold selection and event definition

The first step of developing the MSSM consists of extracting sea-storm events from the observational records. Here, we want to identify events where TWL exceeded a critical threshold making it likely for morphological change to occur; and we also want to know how much the different variables contributed to those events and if there is a dominant driver that can be used for the event selection. Therefore, for each individual year, we find the hourly values when TWL exceeded 1.2 m above the North American Vertical Datum of 1988 (NAVD88; in Dauphin Island NAVD88 lies 18 cm below MSL and 20.7 cm below mean high water for the 1983 to 2001 epoch). This represents approximately the 5th percentile of dune toe heights (ranging from 1.1 m to 2.6 m across the island with an average of 1.75 m) and it is likely that dune erosion occurs somewhere on Dauphin Island when TWL (assuming the average beach slope of 0.07) exceeds this threshold.

The annually averaged TWL from all threshold exceedances from a given year are then computed. We also want to assess the relative importance of the contributions of MSL (here the 30 day running median of hourly water levels), η_A, η_{NTR}, and R2%. Thus, we also calculate annual averages for each individual variable during the TWL threshold exceedances (Figures 3a and 3b). This is done separately for summer and winter seasons in order to isolate tropical and extra-tropical weather events. We define the summer from June through November (i.e. the Atlantic Hurricane season) and the winter from December through May. In both seasons the wave contribution dominates the TWL threshold exceedances, explaining

73% and 79% of the average TWL exceedance events for summer and winter, respectively (Figures 3a and 3b). The other variables played a less important role in pushing TWL beyond critical thresholds.

With the wave contribution being so dominant we also calculated average Hs values for all times when TWL exceeded the 1.2 m threshold. Based on the results shown in Figures 3c and 3d we select Hs thresholds of 1.4 m and 1.6 m for summer and winter, respectively, to identify "extreme events" directly from the wave heights, which are independent of the beach slope. In some instances waves higher than the thresholds were observed offshore but coincided with negative surge values at the tide gauge. This suggests that winds blowing away from the shoreline were responsible for the high offshore waves and it is likely that smaller waves occurred close to shore and did not result (in combination with the water level) in collision or overwash. To account for this we only consider events where the Hs thresholds were exceeded and the simultaneous surge was positive. A careful screening of the initial data set, including the wave direction, confirmed that there were no relevant events where high waves and negative surge resulted in high TWL.

Based on these definitions we apply the following approach to identify individual multivariate sea-storms: we search for the Hs threshold exceedances and select the concomitant Tp and θ values; the duration (D) of an event is defined as the time period where Hs remains above the threshold, if it drops below the threshold and stays there for more than 24 hours (that is the same value used for example by *Li et al.* [2014b]) we assume a new event, otherwise we assume that the threshold exceedance is still associated with the same large scale weather system. This assures that events are approximately independent for the subsequent statistical analysis. Once we know the start and end dates of individual events we select the largest (positive) η_{NTR} and simultaneous η_A values. This event definition approach is outlined by the schematic in Figure 4. In total, we use six variables to define individual multivariate sea-storm events: η_A, $\eta_{NTR,}$ Hs, Tp, θ, and D. We find 358 events in summer (~1.8 events/month) and 421 events in winter (~2 events/month) for the 1980 to 2013 period.

3.2 Modelling marginal variables

In Section 3.3 we describe a copula-based approach to model the interdependency between the different sea-storm variables. One of the advantages of using copulas is the decoupling of the marginal and dependence problem [e.g., *Nelson*, 2006]. This allows us to be flexible in selecting (and mixing) various marginal distributions that are most suitable to capture the behavior (mainly in the tail regions) of the underlying data sets. In terms of the marginal distributions, the six variables are treated differently as described below. We fit the following distributions, which are typically used in (coastal) hydrologic applications, to the summer and winter samples of $\eta_{NTR,}$ Tp, and D: generalized extreme value (GEV), exponential, gamma, inverse Gaussian, logistic, log-logistic, lognormal, Rayleigh, t location-scale, and Weibull. The distributions that fit best to the underlying data (see Figure 5) are identified by minimizing the root mean squared error (RMSE) between theoretical and empirical (obtained here with Weibull's plotting position formula [*Chow*, 1964]) non-exceedance probabilities (P_u).

As described in the previous section, the maximum Hs associated with each event are derived with a peaks-over-threshold approach and therefore – instead of testing different distributions – we assume, similar to *Serafin and Ruggiero* [2014], that the generalized Pareto distribution is capable of modelling the tail behavior of the samples (Figure 5b). The Quantile-Quantile plots in Figure 5 show that, in general, the selected distributions fit well to the summer and winter samples of $\eta_{NTR,}$ Hs, Tp, and D. The GEV distribution selected to model summer η_{NTR} underestimates the largest values, which were the result of strong hurricanes that are typically underrepresented in (short) observational records [e.g., *Haigh et al.*, 2014; *Nadal-Caraballo* *et al.*, 2015]. However, it has been shown here that the wave contribution dominates most of the large TWL values, and therefore we expect the moderate underestimation of the most extreme η_{NTR} to have only a negligible effect on the overall results. Furthermore, all selected distributions (including the GEV distribution for summer η_{NTR}) pass the Kolmogorov-Smirnov goodness-of-fit test at the 95% confidence level.

The two remaining sea-storm variables, η_A and θ, vary within restricted ranges, and therefore – instead of fitting parametric distributions that can be extrapolated beyond the observed values – we use their respective empirical distributions (ECDFs) to draw samples within the Monte-Carlo simulation (Section 3.4). Histograms derived from the observed samples of all variables are shown on the diagonal of Figure 6. This highlights the advantage of separating summer and winter samples as some of them (especially η_{NTR} and Hs) clearly stem from different populations.

3.3 Dependence analysis and modelling

Next, we want to identify dependencies between variables. Therefore, we calculate Kendall's rank correlation τ for all data pairs in our sample of extreme events; we prefer the rank correlation over the widely applied linear correlation coefficient because it also captures potential non-linear relationships. When τ is significant (95% confidence) for a given variable pair the respective scatter plot in Figure 6 is colored, when τ is insignificant the data are plotted in grey. In both seasons η_{NTR}, Hs, Tp, and D share significant (95% confidence) dependency, with τ ranging from 0.34 to 0.58. We want our model to account for those interdependencies and different multivariate approaches exist (and have been applied in the past) for this purpose, most notably Archimedean [e.g., *DeMichele et al.*, 2007; *Corbella and Strecth*, 2013] and elliptical [e.g., *Li et al.*, 2014b] copulas, the multivariate logistics model [e.g., *Callaghan et al.*, 2008; *Serafin and Ruggiero*, 2014], and a conditional model introduced by *Heffernan and Tawn* [2004] [e.g., *Gouldby et al.*, 2014]. *Li et al.* [2014a] considered and compared the first three approaches using a data set from the Dutch coast and concluded – based on a goodness-of-fit test – that the Gaussian copula was most suitable to model the interdependencies. For the Australian coast *Li* [2014] identified Archimedean copulas and the Gaussian copula to be applicable to model the interdependencies, while the logistics model failed the goodness-of-fit test. The Gaussian and *t*-Student copulas belong to the class of elliptical copulas [e.g., *Embrechts et al.*, 2003] arising from elliptical distributions via Sklar's theorem [*Sklar*, 1959]. They are restricted in the sense that they are not capable of modelling tail dependence (stronger/weaker dependence in the upper/lower tails or vice versa), but they also have the advantage of being easily constructed and applied to *d*-dimensional data sets. Archimedean and extreme value copulas are capable of modelling tail dependence but their extension to higher dimensions is more complicated. Based on the conclusions drawn by *Li et al.* [2014a] and *Li* [2014], and as a trade-off between capturing much of the relevant interdependencies and reducing model complexity, we test the ability of the two elliptical copulas to model the observed interdependencies.

Copulas are distributions over the unit hypercube $[0,1]^d$; therefore, we transform the observations by rescaling their ranks by a factor $1/(N+1)$, where N is the number of events (358 in summer and 421 in winter). For a given linear correlation matrix $\sum \in R^{d \times d}$ (in our case $d = 4$) the multivariate Gaussian copula can be written as:

$$C_{\Sigma}(u) = \Phi_{\Sigma}\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right) \qquad (2)$$

with $u_j \sim U(0,1)$ for $j = 1,\ldots,d$, where $U(0,1)$ represents the uniform distribution on the [0,1] interval, Φ^{-1} is the inverse distribution function of a standard normal random variable, and Φ_{Σ} is the d-variate standard normal distribution function.

The t-Student copula can be expressed as:

$$C_{\mathrm{v},\Sigma}(u) = t_{v,\Sigma}\left(t_v^{-1}(u_1), \ldots, t_v^{-1}(u_d)\right) \quad (3)$$

where t_v is the one dimensional t distribution with v degrees of freedom and $t_{v,\Sigma}$ is the multivariate t distribution with a correlation matrix Σ and v degrees of freedom.

After fitting copulas to the transformed 4-dimensional data sets, we can simulate a large number of quadruplets of η_{NTR}, Hs, Tp, and D in the unit hypercube while preserving the interdependencies between them. Using the inverse of the marginal cumulative distribution functions (CDFs) identified in Section 3.2, the simulated data can be transformed from the unit hypercube space to real units.

We follow this procedure with the two elliptical copulas and compare 3000 simulated quadruplets to the observations in order to identify the most suitable one for modelling the underlying dependence structure. In addition to the visual comparison of the scatter plots we also compare τ and non-parametric tail dependence coefficients (TDCs; for a threshold of 0.5) [*Schmidt and Stadtmüller*, 2006] derived from observations and simulations (Figure 7). Results obtained with the *t*-Student copula are slightly better than those derived with the Gaussian copula (not shown) and it also passes the formal goodness-of-fit test proposed by *Genest at al.* [2009] at the 95% confidence level. η_A is found to be independent from all the other variables (Figure 6) and can be simulated randomly from its ECDF (as outlined in Section 3.2). When comparing the circular data of θ with the other variables the τ values are insignificant, but it is obvious from Figure 6 that large waves typically approach from an angle between ~100 and 210 degrees in nautical convention; this is the south-east direction where the fetch is relatively open to Dauphin Island. We account for this by obtaining two different ECDFs for θ, one from values when Hs < 2.7 m and another one from values when Hs > 2.7 m (marked with dashed green lines in the respective sub-panels in Figure 6). In the Monte-Carlo framework (Section 3.4) we then sample θ from one of the two ECDFs, conditional on Hs. This effectively captures the dependency between θ and the other variables (Figure 6).

Now we are able to simulate a large number of sea-storm events comprised of 6 variables: values for η_{NTR}, Hs, Tp, and D come from the copula model and inverse CDFs described in Section 3.2, whereas η_A and θ are simulated independently from their ECDFs (θ conditioned on Hs).

3.4 Time series simulation

We want to use the multivariate model to simulate – in a Monte-Carlo sense – long (or many) time series of sea-storms which can then be used for the probabilistic erosion risk analysis. Thereby, we have to keep in mind that our statistical model has no knowledge about physical mechanisms constraining some of the variables. When the univariate distributions (Section 3.2) of certain variables are unbounded we may sample unrealistically large values of those variables in the simulations. Hence, we control the model by defining upper boundary values for some variables. Hs is constrained by the water depth *z* [e.g., *Thornton and Guza*, 1982] and we use the following simple relationship to derive Hs,max:

$$Hs, \max = 0.5 \cdot \sqrt{2} \cdot z \tag{4}$$

For a water depth of 28 m where the waves are measured off the coast of Dauphin Island we obtain $H_{s,max}$ = 19.8 m (for comparison, the highest observed Hs value is 14.58 m). Sensitivity tests with smaller thresholds revealed that the effect on the overall results is small, which means that although we put a high threshold to constrain the model only very few of the simulated Hs values get close to this value. For Tp we only allow wave periods of up to 25 seconds; waves with periods larger than that are typically allocated to the infragravity wave energy band(e.g. Munk 1949; Tucker 1950).

For D we tested different thresholds (8, 9, 10, and 20 days) and repeated the application (and validation) described in Section 4 to find that the differences in the results are relatively small and that 9 days seems to be the most reasonable choice (the longest observed event lasted D = 6.5 days according to our event definition in Figure 4).

Before we are able to simulate sea-storm time series we also need to know how many events need to be generated for each year. A simple approach would be to use the average of the observations which would result in ~23 events per year. This does not, however, account for the fact that it was only by chance that we observed exactly 358 summer events and 421 winter events between 1980 and 2013 (according to our definition). To allow the model to be more flexible we calculate the numbers of storms observed in each month between 1980 and 2013 (Figure 7a) and obtain monthly time series (all January values, all February values, etc.). We then fit Poisson distributions to the monthly data sets and use those to obtain a varying number of storms for each simulation month. When the simulated time series is long enough the average number of simulated events converges with the observations (Figure 7b). When we simulate 10,000 34-year long time series (i.e. the length of the observed record) we obtain the min/max ranges shown as vertical bars in Figure 7b; instead of 779 events (i.e. the total number of observed storms), the model generates between 675 and 893 events across the 10,000 simulated time series and it resembles the seasonal cycle in the frequency of events.

Finally, we follow *Li at al.* [2014b] in distributing the simulated storms randomly within a month (i.e. each event gets assigned a time stamp) accounting for their duration and making sure that there are at least 24 hours between successive events.

4. Model application

4.1 TWL return periods

As outlined in the introduction we want to use the MSSM to explore how different our interpretation of the erosion risk could be – defined through TWL return periods (in this section) and impact hours (in the next section) – if more or less extreme realizations of the different sea-storm variables and/or their combinations had occurred between 1980 and 2013. Therefore, as already mentioned in the previous section, we simulate 10,000 sea-storm time series, each one comprising 34 years, and derive TWLs for all events assuming an average beach slope of 0.07. We then fit GEV distributions to the observed and each of the simulated TWL time series. From the 10,000 GEVs derived from the simulations we also obtain 95% confidence levels (Figures 8a and 8b). In summer, the GEV distribution from the observations is closer to the upper end of the 95% level of the simulations, but taking the 100-year return level as an example the simulations are still more than 0.6 m higher. Focusing on the full range of simulation results, the TWL associated with a 100-year return period could be up to 3 m higher than our best estimate based on the observational data. In winter the 100-year TWL could be ~0.6 m higher than the best estimate derived from the observations when looking at the full range of simulation results, and ~0.1 m higher when focusing on the upper 95% confidence level.

For comparison – and noting that the results are unrealistic but represent simplifications that reduce model complexity and have been applied in previous studies – we repeat the same analysis with three different model assumptions to generate synthetic sea-storm time series. In this case we only show the upper end of the range of GEV distributions fitted to the simulation results. For the first experiment we assume that all six sea-storm variables are completely independent from each other. The GEV distributions for summer and winter, shown as black dashed lines in Figures 8a and 8b, reveal that this assumption leads to an underestimation of the TWLs associated with a 100-year return period of ~1.65 m in summer and ~0.25 m in winter relative to those derived from the observed TWL time series. For the second experiment we assume that most sea-storm variables are independent but that Hs and Tp are fully dependent. We derive Tp using the following regression model that was also used by *Stockdon et al.* [2012] to construct extreme event scenarios for the Gulf of Mexico.

$$Tp = 3.846 + 1.7812\,Hs - 0.012Hs^2 - 0.0049\,z \qquad (5)$$

where z is the water depth. Consistent with the model development, we do not allow Tp values larger than 25 seconds. In this case the return TWLs (shown as green dashed lines in Figures 8a and 8b) are significantly overestimated relative to the ones obtained from the observations and also compared to the ones derived from the simulations that used the more realistic interdependencies. For the third and final experiment we account for the interdependencies between the different sea-storm variables as explained in the previous sections but do not allow the individual variables to reach values larger than their observed maxima. The results (shown as dashed brown lines in Figures 8a and 8b) reveal that 100-year TWLs could be approximately 2.6 m higher in summer and 0.25 m in winter only due to different (but according to our model realistic) extreme event combinations where none of the individual variables exceeds its observed maximum.

Similar to *Serafin and Ruggiero* [2014], we perform a second analysis where we use 500 simulated time series with a time-span of 500 years (instead of 34 years). From such long time series we can obtain the relevant return water levels empirically so that no uncertainties are involved from fitting parametric distributions; however, we still have to use the GEV distribution for the observations to facilitate comparison of the results (Figures 8c and 8d). The medians of relevant return TWLs (10-, 20-, 50-, and 100-years) from the simulations (grey circles) are similar to those derived from the observations, highlighting that our model does a good job in capturing the behavior of the underlying sea-storm variables and their interdependencies. The interpretation of how much larger TWLs associated with different return periods could be are, however, slightly different to those derived earlier by fitting GEV distributions to both the simulated and observed TWL time series. The 100-year TWL could be almost 2.2 m higher in summer and 0.25 m in winter (black dots and light shaded bands). Results from the three additional experiments (only the upper ranges are shown) confirm the underestimation when assuming fully independent sea-storm variables (black crosses) and overestimation when assuming independency between most variables but full dependence between Hs and Tp (green crosses). Accounting for the interdependencies but constraining the model with observed maxima of the individual variables (brown circles) leads to slightly smaller values as derived with the optimal model setup and from the observations. The differences increase for larger return periods and may stem from the uncertainties when fitting the GEV to the observations or from the fact that we have already seen a larger number of extreme event combinations over the last three decades than our model predicts (especially in summer).

4.2 Impact hours

The TWL time series used in the previous section were derived from η_A, $\eta_{NTR,}$ Hs, and Tp, and the extreme value analysis is also affected by the frequency of events. The duration D was, however, not included in the analysis. Therefore, and as an alternative way of assessing erosion risk we calculate average impact hours for Dauphin Island (1980 to 2013) when TWL exceeded the height of the dune toe (collision) or dune crest (overwash) [e.g., *Ruggiero*, 2013, WP15]. Impact hours are affected by the oceanographic forcing variables but also by beach slope and the elevation of backshore features. Therefore, we no longer assume the foreshore beach slope (and dune characteristics) are uniform alongshore and instead perform the model simulations using the spatially variable measured values at each location. We use the observed and the 10,000 simulated sea-storm data sets and derive TWL time series for each of the transects along Dauphin Island and compare them to dune toe and crest elevations. The impact hours of collision and overwash are determined from both observations and simulations under the assumption that if TWL associated with a particular event exceeds a critical threshold, then it is exceeded for the entire event duration. Because of this assumption the absolute values of impact hours presented here overestimate the "true" values. The latter could be derived from hourly observations of the different variables accounting for the fact that critical TWL thresholds are not necessarily exceeded throughout an entire event (as defined here). However, we are only interested in the relative comparison between observations and simulations and since impact hours are derived under the same assumption the direct comparison is valid. Similar to the previous analysis of TWL return periods we repeat the analysis with the three additional model assumptions (independence assumption; full dependence between Hs and Tp; and individual variables constrained with observed maxima). At this stage of the analysis we can also re-introduce the seasonal cycles, inter-annual variability, and decadal trends that were removed earlier. We add the running medians that were subtracted at the beginning of the analysis to the simulated time series and assess relative changes between observations and simulations. This provides information about the importance of the timing of extreme events relative to the seasonal cycle or longer-term variability.

For collision (Figure 9a) we find that the number of impact hours could have been up to 70% larger (or up to 60% smaller) than inferred from the observed time series in both seasons; overwash impact hours (Figure 9b) could have been twice as high. For both collision and overwash the 68% and 95% confidence intervals reach values that are ~20% and ~40% larger, respectively. Note that the overall number of impact hours decreases after trends and variability are re-included because we corrected the time series earlier in a way that they resemble the present-day climate. Accordingly, the heights of TWL events earlier in the records decrease when trends and variability are re-included. For overwash in summer the number of impact hours inferred from the observations becomes considerably larger (close to the upper 68% confidence level) than the median derived from the simulations after trends and variability of the different variables are re-included. This suggests that the timing of extreme events relative to the seasonal cycle and climate related variations is important, especially when focusing on the most extreme events leading to overwash (and inundation).

The results obtained from the three additional experiments with varying model setups (only maxima values are shown in Figure 9 for the simulations without trend and variability) are somewhat different compared to those derived in the previous section for TWL return periods. Under the full independence assumption the erosion risk is still underestimated; the same is also true now for the assumption that Hs and Tp are fully dependent (whereas we found overestimation in the previous section). This is because impact hours are strongly affected by the duration D, and by assuming that it is independent from the other variables the highest modelled TWL events do not tend to have longer durations (contrary to what is inferred from the observations). When we constrain the model so that none of the individual variables can reach values larger than the observed maxima we find that the number of impact hours could have been ~50% larger for collision and overwash in summer and ~30% for collision and ~45% for overwash in winter. These higher values solely stem from a larger (but physically consistent) number of extreme event combinations of the different sea-storm variables over the 34 year analysis period, instead of more extreme realizations of the individual variables.

5. Discussion

When assessing erosion or flooding risk for certain regions we often rely on the observational data sets that are available to estimate extreme-value statistics. The observational records of wave properties rarely go back more than a few decades, which we show has limited the accuracy of such estimates. In this present study we focus on how different our interpretation of the erosion/flooding risk could be if observations had sampled different realizations of the individual sea-storm parameters and their combinations over the last few decades (or if we had hundreds of years of observations available instead of only 34 years). If we use very long simulated data sets return water levels become more stable and in our case the range of results from the simulations proceeds within the theoretical uncertainties from fitting a GEV to the short observational record (Figures 8c and 8d). On the other hand, if we use the exact same approach for both observations and simulations of fitting the GEV to 34 year long records the range of results is much wider (Figures 8a and 8b) and would exceed the theoretical uncertainties that are shown in Figures 8c and 8d. This highlights the importance of a detailed uncertainty assessment in extreme value analysis and its inclusion into engineering design concepts.

Based on our analysis we cannot say which variable contributes most to the identified differences in estimates of return water levels and impact hours from simulations and observations. Future work is needed to explore the role of individual sea-storm variables and identify those which need to be carefully constrained in future applications when the methodology is for example transferred to other regions with different oceanographic (and morphologic) conditions.

The MSSM output derived here may be used for various future applications, including long-term simulation of erosion and/or recession under different sea level rise scenarios (similar to *Corbella and Strecth* [2012a] and *Li et al.* [2014b]). The results may also be used along with more sophisticated (but computationally demanding) numerical models (e.g. XBeach) that include estimates of morphological change and/or flood impacts and require, as boundary conditions, a meaningful selection of extreme events and, depending on the application, their (joint) return periods in order to perform a full risk analysis; this is something we will explore in a future investigation. The uncertainties in return TWL estimates stemming from short observational record availability can furthermore be incorporated into more robust and risk aversive design strategies for coastal infrastructure and/or restoration of backshore features.

There are two quantities that we derive through the MSSM that are not used in the TWL-based applications presented above. By assigning time stamps to the simulated sea-storms we implicitly derive the time spans between the end of one extreme event and beginning of the next which are relevant, for example, for long-term simulation of erosion/recession and accretion. The wave direction θ is directly simulated in the MSSM but also not used here. Depending on the purpose of the application it can, however, be an important variable, e.g., when quantifying morphological change including long-shore sediment transport.

Results from assessing impact hours with and without trends and variability of the underlying variables included reveal the importance of the timing of extreme events within the seasonal cycle and relative to monthly mean sea level anomalies. Therefore, for the future it would be interesting to explore those relations in more detail and ultimately include them in the analysis either by directly modelling mean sea level anomalies (and their dependence with other sea-storm variables) or including climate indices as covariates [e.g., *Serafin and Ruggiero*, 2014].

The MSSM developed here is generic and can – with very few adjustments – be applied to other coastlines. Assessing erosion risk in a probabilistic way and the results derived in the present study in combination with our knowledge on the effects of climate variability and change can help decision makers and planners to account for previously unseen, but possible, events when planning for long-term sustainability of beaches and barrier islands.

6. Conclusions

Based on 34 years of wave and water level observations from Dauphin Island in the northern Gulf of Mexico we develop a copula-based MSSM to simulate a large number of synthetic time series of the six most relevant (for driving erosion/flooding) sea-storm parameters, the interrelationships among them, and derive TWLs with the empirical formulation of *Stockdon et al.* [2006]. We quantify the erosion and flooding risk by calculating return periods of TWLs and impact hours of collision and overwash. Our results indicate, for example, that the 100-year return TWLs (often used for design purposes of coastal infrastructure or restoring dunes) could be more than 3 m higher in summer and 0.6 m in winter relative to our best estimate based on the observational records. The number of impact hours of collision and overwash could have been up to 70% and 100% larger, respectively, than inferred from the observations. Many of these differences are explained by an increase in the total number of extreme events that can occur from plausible combinations of different sea states even when none of the individual sea-state variables exceed the highest value from the observational record. This demonstrates why incorporating joint correlations is essential in performing coastal risk analyses rather than only relying on historical conditions derived from short observational records.

Acknowledgments

The hourly water level data used in the analysis were downloaded from the data base of the University of Hawaii Sea Level Center (http://uhslc.soest.hawaii.edu/) and NOAA's Tides and Currents website (http://tidesandcurrents.noaa.gov/). Wave data were obtained from NOAA's National Data Buoy Center (http://www.ndbc.noaa.gov/) and USACE's Wave Information Studies (http://wis.usace.army.mil/). Lidar data were analyzed and provided to us by K. Doran and H. Stockdon (USGS). Erika Lentz and two anonymous reviewers provided thoughtful comments that helped improving the statistical analysis, presentation of the results, and overall quality of the manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Callaghan, D. P., P. Nielsen, A. Short, and R. Ranasinghe (2008), Statistical simulation of wave climate and extreme beach erosion, *Coasal Eng.*, 55, 375–390, doi:10.1016/j.coastaleng.2007.12.003.

Chow, V. T. (1964), Handbook of Applied Hydrology, McGraw-Hill, New York.

Corbella, S., and D. D. Stretch (2012a), Predicting coastal erosion trends using non-stationary statistics and process-based models, *Coastal Eng.*, 70, 40–49, doi:10.1016/j.coastaleng.2012.06.004.

Corbella, S., and D. D. Stretch (2012b), Multivariate return periods of sea-storms for coastal erosion risk assessment, *Nat. Hazards Earth Syst. Sci*., 12, 2699–2708, doi:10.5194/nhess-12–2699-2012.

Corbella, S., and D. D. Stretch, D. D. (2013), Simulating a multivariate sea-storm using Archimedean Copulas, *Coastal Eng.*, 76, 68–78, doi:10.1016/j.coastaleng.2013.01.011.

De Michele, C., G. Salvadori, G. Passoni, and R. Vezzoli, (2007), A multivariate model of sea-storms using copulas, *Coastal Eng.*, 54, 734–751, doi:10.1016/j.coastaleng.2007.05.007.

Guy, K. K., and N.G. Plant (2014), Topographic lidar survey of Dauphin Island, Alabama and Chandeleur, Stake, Grand Gosier and Breton Islands, Louisiana, July 12–14, 2013: U.S. Geological Survey Data Series 838, http://dx.doi.org/10.3133/ds838.

Embrechts P., F. Lindskog, and A. J. McNeil (2003), Modelling Dependence with Copulas and applications to Risk Management, In: *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, Chapter 8, pp. 329–384.

Genest, C., B. Rémillard, and D. Beaudoin (2009), Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44, 199–214, doi:10.1016/j.insmatheco.2007.10.005.

Gouldby, B., F. Mendez, Y. Guanche, A. Rueda, and R. Minguez, R. (2014), A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis, *Coastal Engineering*, 88, 15–26.

Haigh, I. D., L. R. MacPherson, M. S. Mason, E.M.S. Wijeratne, C. B. Pattiaratchi, R. P. Crompton, and S. George (2014), Estimating present day extreme water level exceedance probabilities around the coastline of Australia: tropical cyclone-induced storm surges, *Clim. Dynam.*, 42(1–2), 139–157. doi:10.1007/s00382-012-1653-0.

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

Li, F. (2014), Probabilistic estimation of dune erosion and coastal zone risk, PhD thesis, doi:10.4233/uuid:221611ff-8a69-40f3-86df-7143e182619c

Li, F., P. van Gelder, R. Ranasinghe, D. Callaghan, and R. Jongejan (2014a), Probabilistic modelling of extreme storms along the Dutch coast, *Coastal Eng.*, 86, 1–13, doi:10.1016/j.coastaleng.2013.12.009.

Li, F., P. H. A. J. M. van Gelder, J. K. Vrijling, T. B. Callaghan, R. B. Jongejan, and R. Ranasinghe (2014b), Probabilistic estimation of coastal dune erosion and recession by statistical simulation of storm events, *Applied Ocean Research*, 47, 53–62, doi:10.1016/j.apor.2014.01.002.

Méndez, F. J., M. Menéendez, A. Luceño, and I. J. Losada (2006), Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model, *J. Geophys. Res. Oceans*, 111, C07024, doi:10.1029/2005JC003344.

Munk, W. H. (1949), Surf beats. *EOS*, 30(6), 849-854.

Nadal-Caraballo, N.C., J. A. Melby, and V. M. Gonzalez, (2015), Statistical analysis of historical extreme water levels for the U.S. North Atlantic coast using Monte Carlo life-cycle simulation, *J. Coastal Res*., online first.

Nelsen, R. B (2006), *An introduction to copulas,* Lecture Notes in Statistics, 139, Springer, New York, 2nd ed..

Pawlowicz, R., B. Beardsley, and S. Lentz (2002), Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE, *Comput. Geosci.*, 28, 929–937, doi:10.1016/S0098-3004(02)00013-4.

Ruggiero, P., P. D. Komar, W. G. McDougal, J. J. Marra, and R. A. Beach (2001), Wave runup, extreme water levels and the erosion of properties backing beaches, *J. Coastal Res*., 17(2), 407–419.

Ruggiero, P. (2013), Is the intensifying wave climate of the U.S. Pacific Northwest increasing flooding and erosion risk faster than sea-level rise?, *J. Waterw. Port Coast. Ocean Eng.*, 139(2), 88–97, doi: /10.1061/(ASCE)WW.1943-5460.0000172.

Sallenger, A. H., Jr. (2000), Storm impact scale for barrier islands, *J. Coast. Res.*, 16(3), 890–895.

Schmidt, R., and U. Stadmüller (2006), Nonparametric estimation of tail dependence, Scand. J. Stat., 33, 307–335, doi: 10.1111/j.1467-9469.2005.00483.x

Serafin, K. A., and P. Ruggiero (2014), Simulating extreme total water levels using a time-dependent, extreme value approach, *J. Geophys. Res. Oceans*, 119(9), 6305–6329, doi10.1002/2014JC010093.

Sklar, A. (1959), Fonctions de Répartition à n Dimensions et Leurs Marges, *Publications de Institut de Statistique Universite de Paris*, 8, 229–231.

Stockdon, H. F., R. A. Holman, P. A. Howd, and A. H. Sallenger Jr. (2006), Empirical parameterization of setup, swash, and runup, *Coastal Eng.*, 53, 573–588, doi:10.1016/j.coastaleng.2005.12.005.

Stockdon, H. F., K. S. Doran, and A. H. Sallenger (2009), Extraction of lidar-based dune-crest elevations for use in examining the vulnerability of beaches to inundation during hurricanes, *J. Coastal Res.*, 25(6), 59–65, doi: 10.2112/SI53-007.1.

Stockdon, H. F., K. J. Doran, D. M. Thompson, K. L. Sopkin, N. G. Plant, and A. H. Sallenger (2012), National assessment of hurricane-induced coastal erosion hazards–Gulf of Mexico: U.S. Geological Survey Open-File Report 2012–1084, 51 p.

Stockdon, H. F., D. M. Thompson, N. G. Plant, and J. W. Long (2014), Evaluation of wave runup predictions from numerical and parametric models, *Coastal Eng.*, 92, 1–11, doi:10.1016/j.coastaleng.2014.06.004.

Thornton, E. B., and R. T. Guza (1982), Energy saturation and phase speeds measured on a natural beach, *J. Geophys. Res. Oceans*, 87(C12), 9499-9508.

Tucker, M. J. (1950), Surf beats: sea waves of 1 to 5 min. period, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 202(1071) , 565-573.

Wahl, T., C. Mudersbach, and J. Jensen (2012), Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a multivariate statistical approach based on Copula functions, *Nat. Hazards Earth Syst. Sci.*, 12, 495-510, doi:10.5194/nhess-12-495-2012.

Wahl, T., F.M. Calafat, and M.E. Luther, M.E. (2014), Rapid changes in the seasonal sea level cycle along the US Gulf coast from the late 20th century, *Geophys. Res. Lett*., 41, 491–498, 2014.

Wahl, T., and N. Plant (2015), Changes in erosion and flooding risk due to long-term and cyclic oceanographic trends, *Geophys. Res. Lett*., 42, doi:10.1002/2015GL063876.

Figures and figure captions

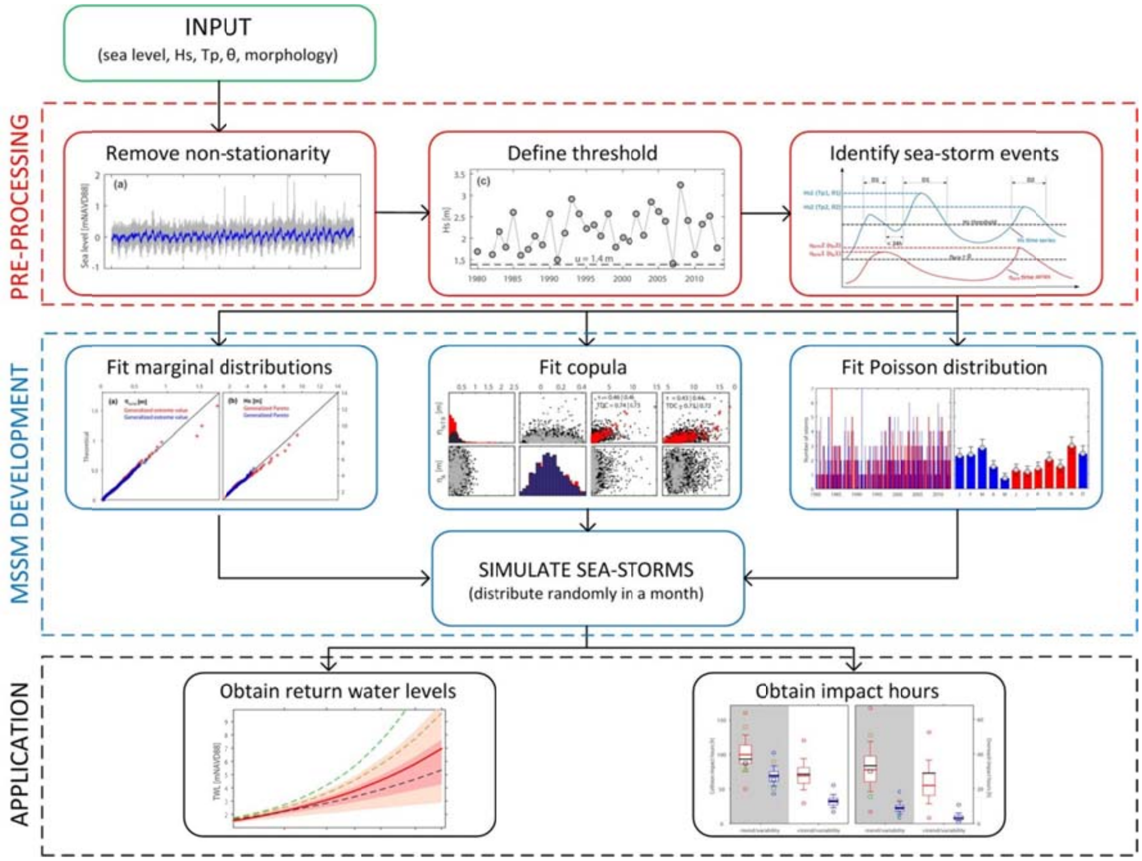

Figure 1. Steps involved in the data pre-processing (Section 2), MSSM model development (Section 3), and model application (Section 4).

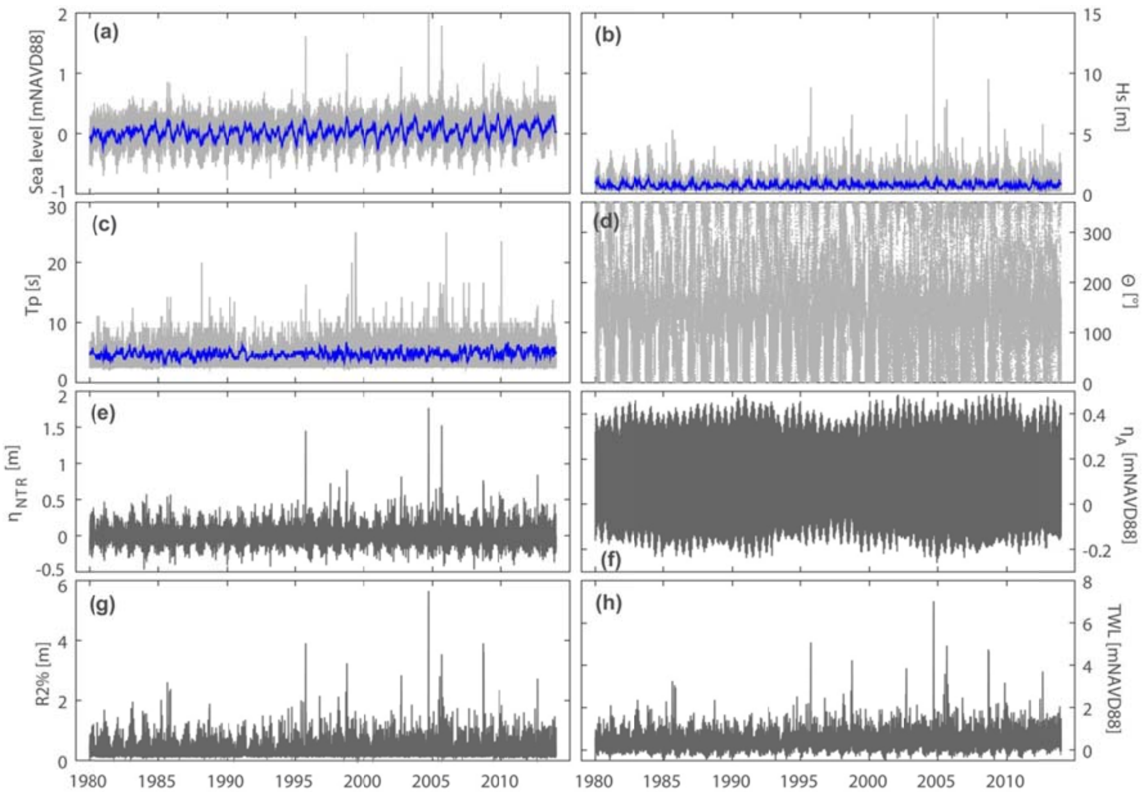

Figure 2. Observed hourly time series of water level, Hs, Tp, and θ (a–d; running 30-day medians are shown in blue); η_{NTR} and η_A derived with a year-by-year tidal analysis (e and f); and R2% (g) and TWL (h).

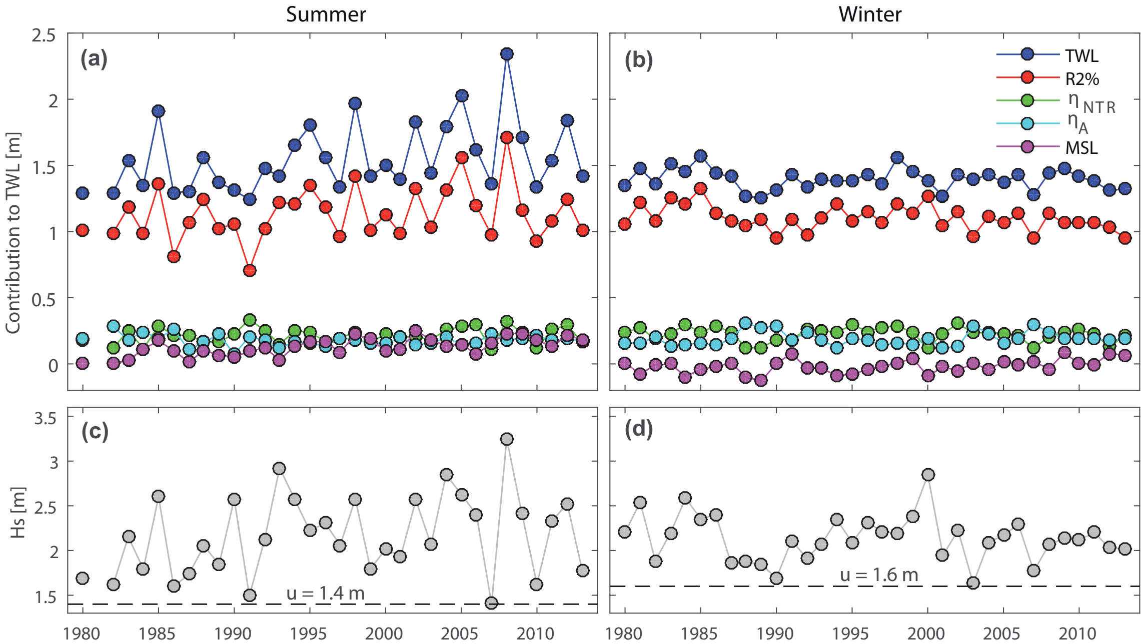

Figure 3. Annual averages of TWL exceedances (1.2 m above NAVD88) and of simultaneous MSL, η_A, η_{NTR}, R2% (a–b) and Hs (c–d); results are shown separately for summer (a, c) and winter (b, d).

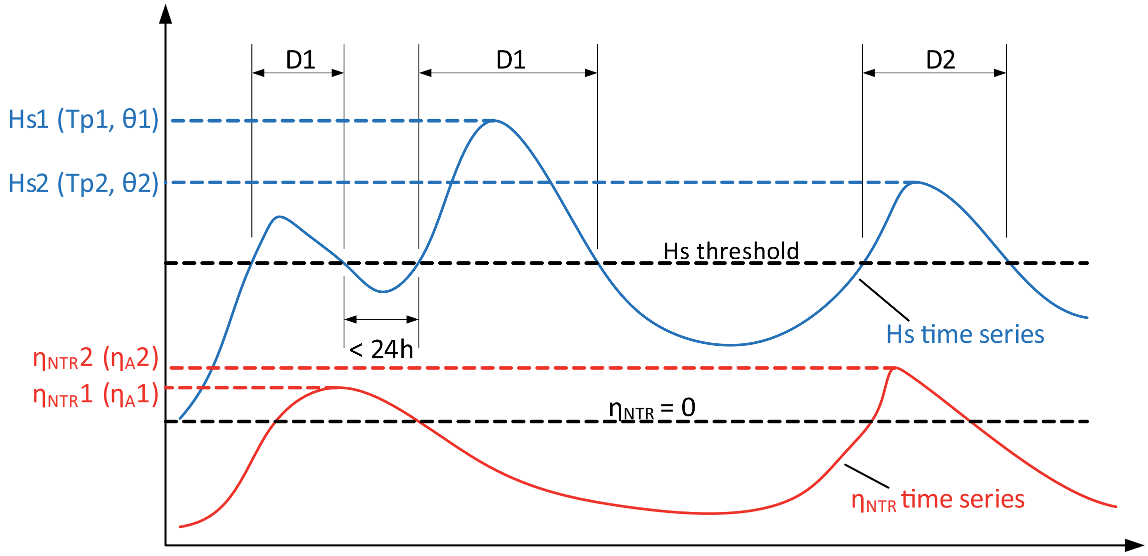

Figure 4. Definition of independent multivariate sea-storm events as used in the present study.

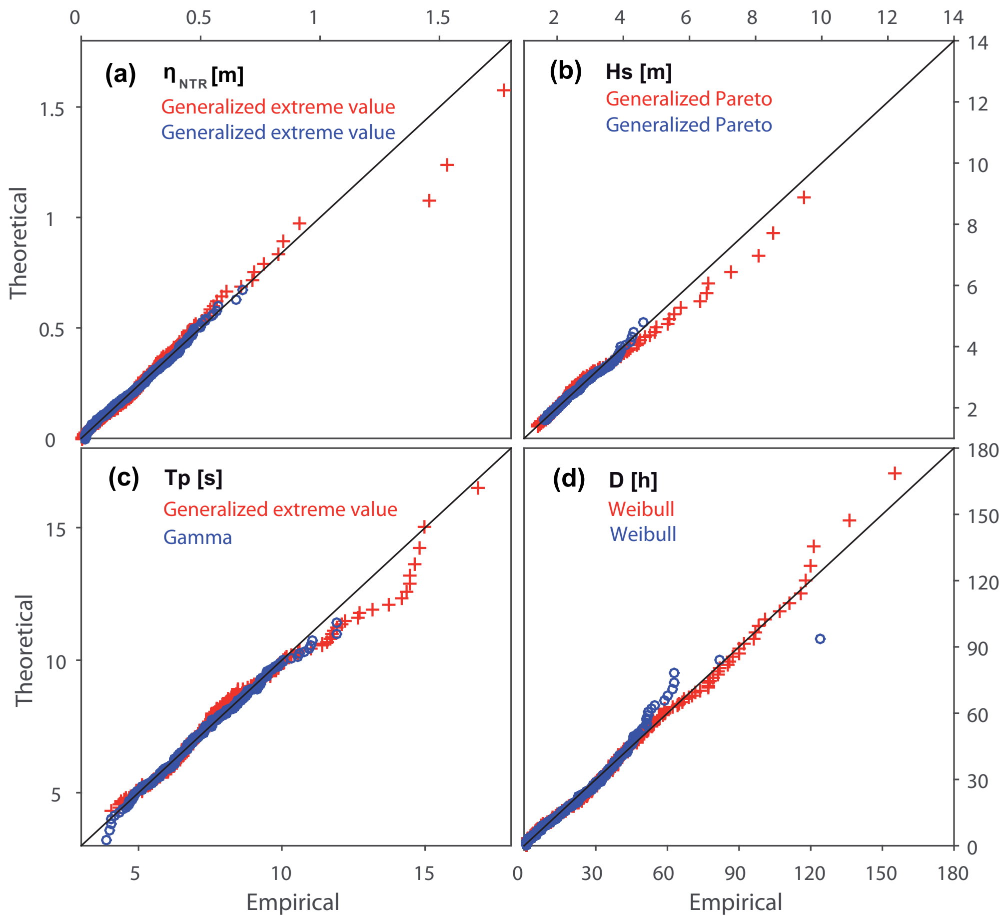

Figure 5. Q-Q plots of parametric distributions fitted to summer (red) and winter (blue) samples of different sea-storm variables.

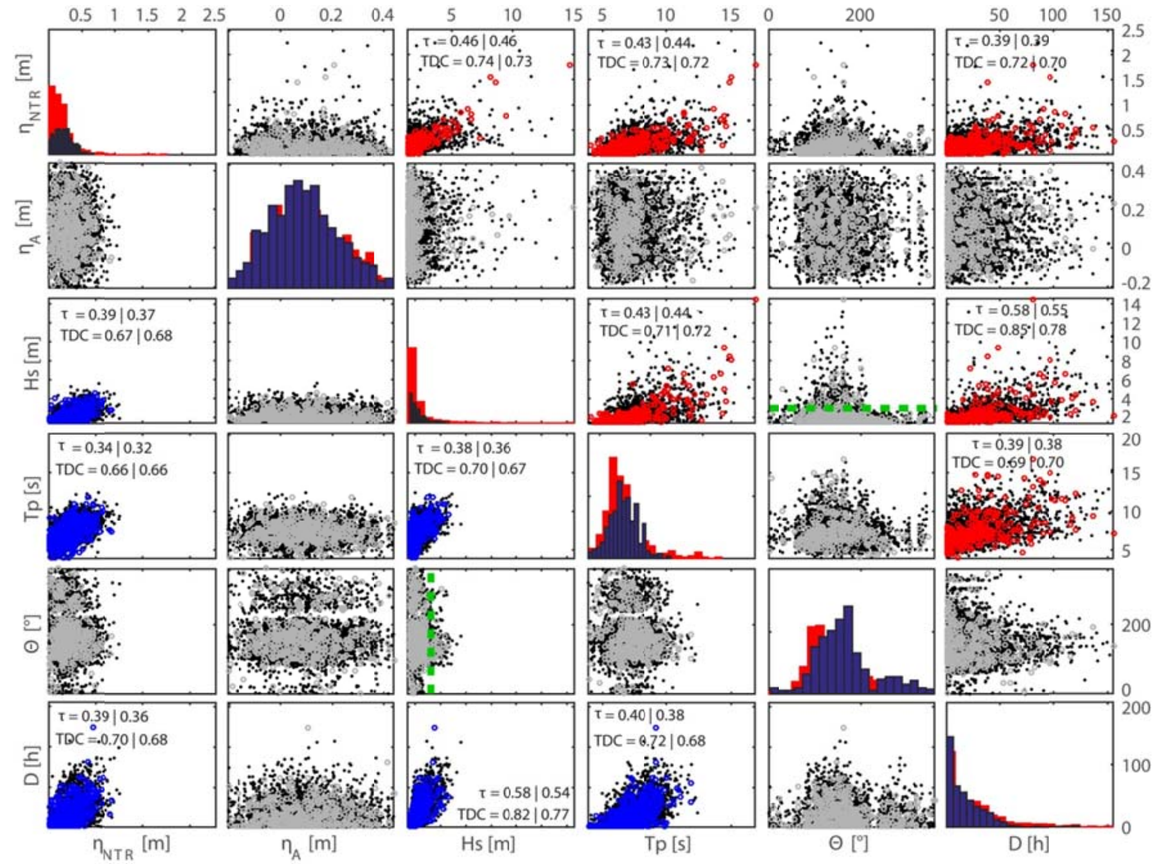

Figure 6. Scatter plots of the six sea-storm variables for summer (panels above the diagonal) and winter (panels below the diagonal); dots are colored (red: summer; blue: winter) when τ is significant and grey otherwise. Simulation results (3000 sea-storms) are shown as black dots; τ and TDC values (observation | simulation) are shown for data pairs where the observed correlation is significant. Green dashed lines mark Hs = 2.7 m used to separate θ values to obtain two ECDFs (see text). Sub-panels on the diagonal show histograms derived from the samples of the six variables.

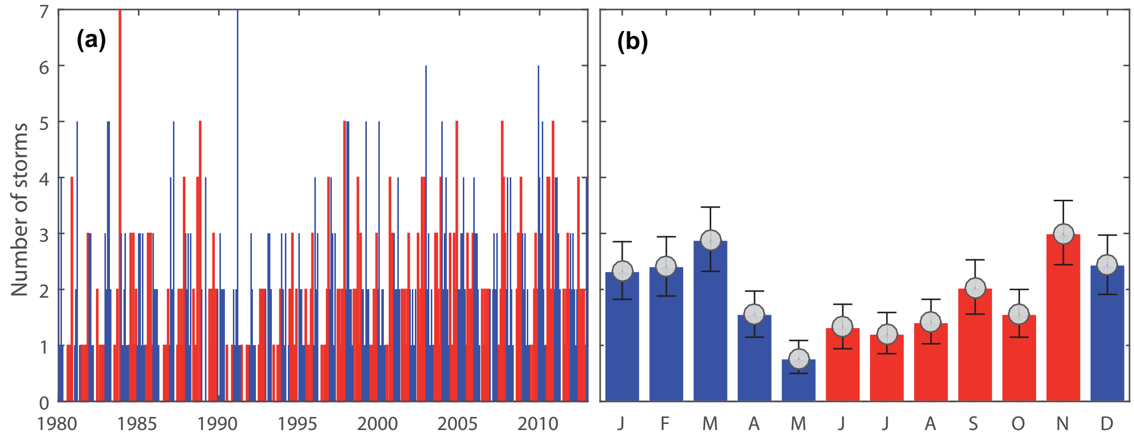

Figure 7. (a) Observed number of storms for each month between 1980 and 2013 (a; red: summer, blue: winter). (b) Average number of storms for each month in a year from observations (colored bars) and average (grey circles) and min/max values (vertical bars) derived from simulating 10,000 sea-storm event time series, each one comprising 34 years.

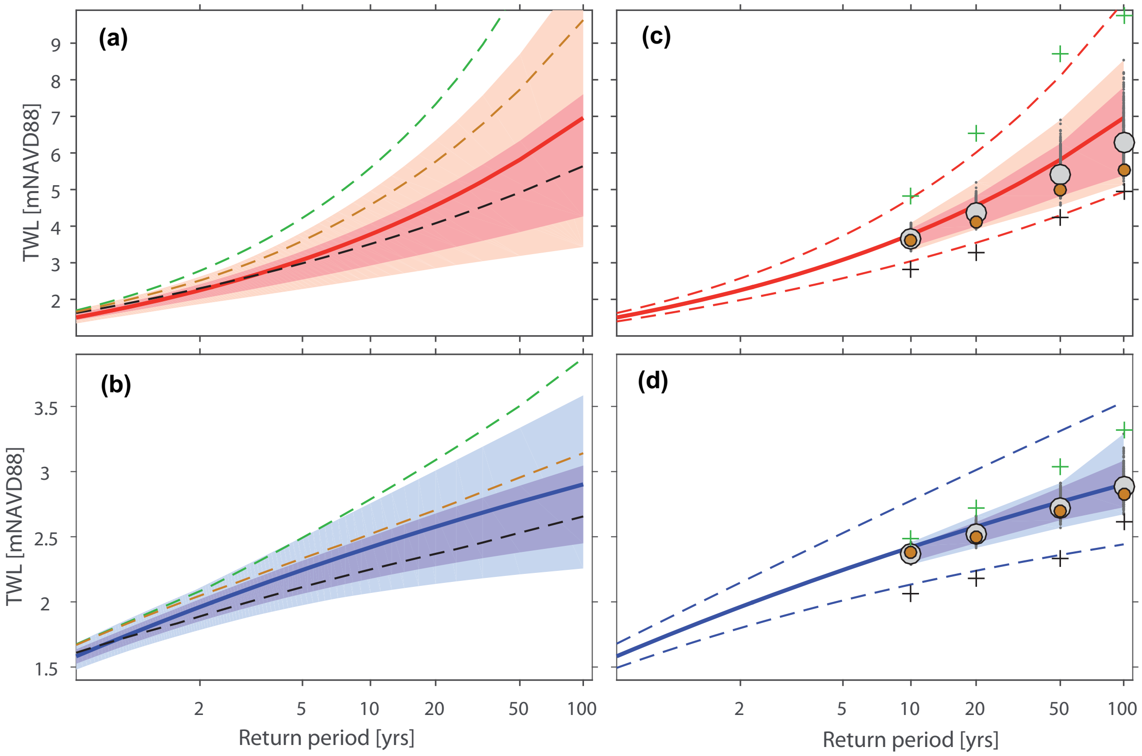

Figure 8. (a–b) Solid lines: GEV fit to TWLs derived from observations; shaded bands: GEV fit to 10,000 simulated time series (each 34-years long), light shading represents full range, dark shading 95% confidence levels; black dashed lines: MSSM assuming independence between all variables; green dashed lines: MSSM assuming independence but full Hs-Tp dependence; brown dashed lines: MSSM capped to observational range of all variables; dashed lines represent upper ends from 10,000 GEV fits. (c-d) Solid and dashed lines: GEV fit to TWL derived from observations (same as in a-b but with 95% confidence levels); black dots and grey circles: empirically derived return TWLs from 500 time series (each 500-years long), grey circles are medians, light and dark shading represent full range and 95% confidence levels; black/green crosses and brown circles are results (only upper end is shown) from three additional MSSM model setups, same color coding as in a-b. Summer results are shown in (a) and (c), winter results in (b) and (d).

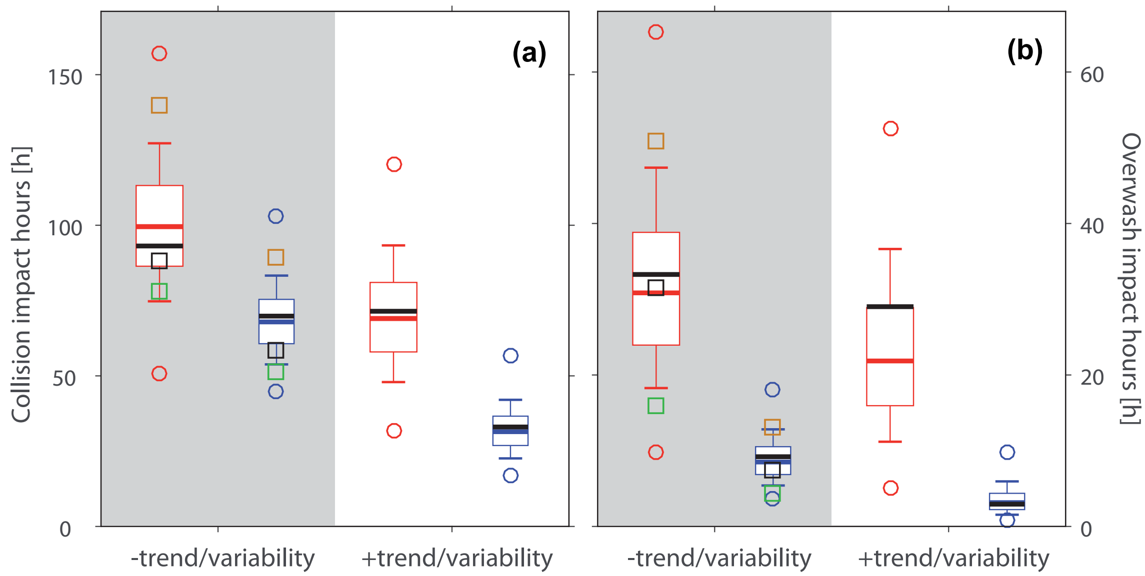

Figure 9. Average number of impact hours for Dauphin Island for collision (a) and overwash (b) as inferred from the observations (black horizontal lines) and 10,000 artificial sea-storm time series derived with the MSSM (box whisker plots show medians and 68% and 95% confidence levels; circles are maxima and minima). Maxima values derived with three additional model setups (see text) are shown as squares (black: independence assumption; green: full dependence between Hs and Tp; brown: individual variables constrained with observed maxima) for the case when trends and variability are not included. Results for the summer half year are shown in red, for the winter half year in blue.