DETERMINATION OF COMBINED PROBABILITIES OF OCCURRENCE OF WATER LEVELS AND WAVE HEIGHTS FOR THE SAFTEY VALIDATION OF DIKES AT THE BALTIC SEA

M. Sc. C. Kaehler[1], Dipl.-Ing. C. Schlamkow[2] and Prof. Dr.-Ing. F. Saathoff[3]

Large parts of the Baltic Sea coast in Germany are protected by dikes against storm surges and floods. The dikes are designed to resist storm surges and floods, also taking into consideration of climate changes and sea level rise. To ensure the protective function the safety standards of the dikes are validated in regular intervals. This paper presents an approach to determine combined probabilities of occurrence of water level and wave heights for three selected sections. The probabilities of occurrence for defined return periods have been calculated by comparing several Copula models from the Archimedean Copula family.

Keywords: Copula, Combined probability, Risk management, Sea state modelling, Extreme value theory

INTRODUCTION

The Baltic Sea coast in Northern Germany has a length of nearly 2500 km. About 240.000 people live in low-lying areas and are endangered by floods. Storm surges and floods are a constant threat to the coastal zones, in particular for the settlement and industrial areas and the infrastructure. Great parts of the hinterland are protected by various types of dikes. The dikes are often designed to resist storm and flood events with a certain return period. Depending on the safety standard, common return periods are 50, 100 or 200 years. The safety standard is validated in regular intervals by the appropriate authority to adapt the dikes if needed, especially with regard to climate change and sea level rise.

The most important design parameters in this validation process are water level and sea state. Usually these parameters are calculated separately (univariate) by determining extreme values. In the Baltic Sea the maximum values of water level and wave height often do not occur at the same moment. So, for the safety validation and design of coastal protection measures, different water levels with corresponding sea state parameters of the same probability of occurrence often more useful. A method to get these values is the application of combined probabilities of occurrence of water level and wave heights. In this paper an approach is presented to determine combined probabilities of occurrence.

The State Office for Coastal Protection, National Park and Marine Protection Schleswig-Holstein engaged the Chair of Geotechnics and Coastal Engineering at the University of Rostock to run an initial project to determine the combined probabilities of occurrence of water levels and wave heights for three dikes at the western Baltic Sea coast.

The project is divided into two parts. The first part contains the estimating of the basic data (especially sea state) and the second part considers the determination of combined probabilities of occurrence (Saathoff et al. 2015).

The determination of basic data includes the univariate design parameters and distribution functions of water levels and sea state. The design parameters and distribution functions of water levels already existed and were not part of the presented investigation. To determine the basic sea state data a method has been developed to combine local wind measurements with stationary wave simulations. The wave simulations for the western Baltic Sea have been implemented by using the SWAN model. With this approach time series of sea state parameters of 30 to 40 years have been calculated. These time series have been improved by applying location- and direction-specific adjustment factors based on local sea state measurements. In the first step of the evaluation of these basis data, different univariate extreme value distribution functions of wave heights have been determined. In the second step the distribution functions of water levels and wave heights (so called marginal distributions) have been combined by using Copula functions to derive combined probabilities of occurrence.

---

[1] Chair of Geotechnics and Coastal Engineering, University of Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
[2] Chair of Geotechnics and Coastal Engineering, University of Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
[3] Chair of Geotechnics and Coastal Engineering, University of Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany

RESEARCH AREA

The research sections are located directly Baltic Sea at the in northern Germany in the federal state of Schleswig Holstein. Each research section contains one so called Regionaldeich[4].
The state office distinguished between Regionaldeichen and Landesschutzdeichen[5]. In contrast to a Landesschutzdeich that protect inhabited areas, like villages and cities, the function of a Regionaldeich is the protection of mostly uninhabited areas like agricultural areas or industrial sections. In consequence to that, Regionaldeiche are often smaller and have a limited protection against floods. Safety validation with independent calculated water levels and wave heights that are regularly performed for Landeschutzdeiche, leads to implausible results. A better approach is the application of Copula models to determine the combined occurrence of water level and wave heights.
Future studies will contain more sections with a various number of Regionaldeiche and Landesschutzdeiche.

Due the varying exposed geographical locations (see Figure 1) and fetch lengths of the three dikes, the wave climate differs for each section and has to be examined separately.

Figure 1: Location of the research sections Schuby / Damp, Behrensdorf / Lippe and Süssau.

DATA BASIS AND STATISTICAL ANALYSES

To estimate the combined probability of occurrence of water levels and wave heights for the safety validation using Copula models it is necessary to know the so called marginal distribution of both parameters. The term "marginal distributions" describes univariate distributions that the Copula models put together to a combined statistical model. The extreme distribution function for the water level in the research area is given by the State Office and can be calculated with the quantile function of the Kappa distribution (1). Table 1 contains the given parameter definitions for the examined section.

$$x(P) = a + \frac{b}{c} \cdot \left(1 - \left[\frac{1 - p^d}{d} \right]^c \right) \quad (1)$$

Table 1: Parameters for the Kappa-quantile-function.

Parameter	Value
a	0.87644502
b	0.26176439
c	0.15767087
d	0.05941047

---

[4] The word by word translation of "Regionaldeich" is "regional dike."
[5] Landesschutzdeiche can be translated as protection dikes.

The extreme distribution for the significant wave height is not known and has to be determined within this project. In order to make qualitative predictions about wave heights with a certain return period, it is necessary to analyze long term time series to fit the extreme value functions. Since sufficient measured of sea state parameters in the research area are not available, they have to be determined.

In general, different methods can be used to predict sea state. In addition to direct measurement, established methods are simple empirical sea state approaches (e.g. Richter, Wagner), wind-wave-correlation or numerical sea state models. For this project the Simulating WAves Nearshore (SWAN) model has been selected to determine long term sea state time series.

The SWAN wave model is appropriable for wave calculations in deep and shallow water and it takes account all relevant physical processes. A border condition of the model for the Baltic Sea is the bathymetry (see Figure 2) of the western Baltic Sea (Seifert et al. 2001). For areas with a highly structured topography, the model is nested with additional bathymetries of smaller spatial resolutions.

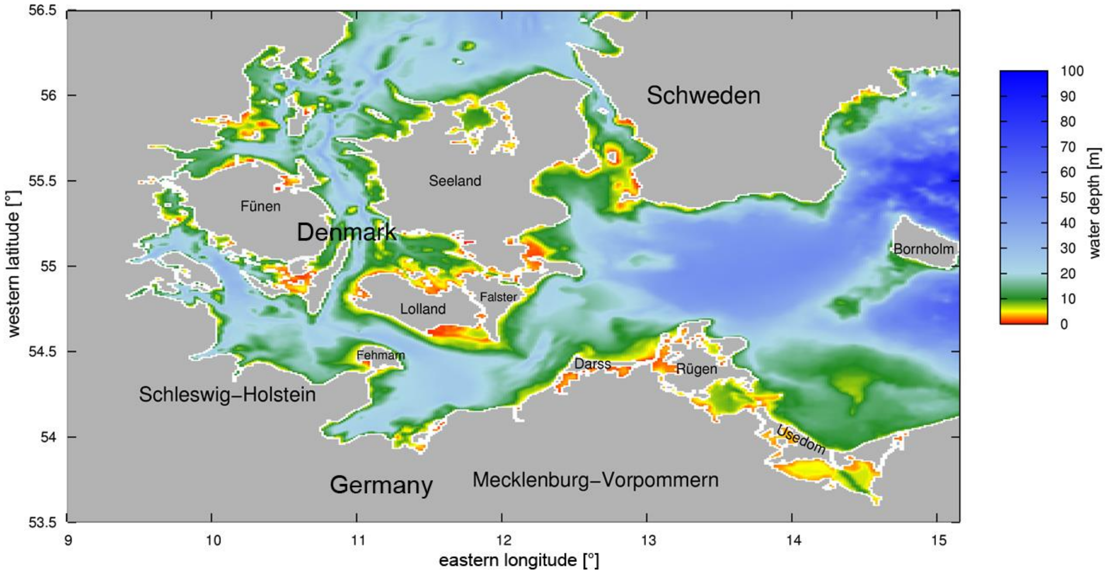

Figure 2: Bathymetry for the stationary sea state simulations.

The Model was configured with the following setup:

- Spatial resolution: $d_x=d_y \approx 1$ km.
- Spectral resolution: $f_{min}=0.02$ Hz, $f_{max}=1$ Hz, 20 frequency grid points, 36 directional grid points.
- Physical effects: Quadruplets, Triads, wave breaking and bottom friction.

Long term time series of wind data (wind velocity and wind direction) and water levels are used as boundary conditions of the model. The wind data were provided by the German Meteorological Office[6] and includes up to 41 years. Water levels long term time series has been made available by the Federal Maritime and Hydrographic Agency of Germany[7]. All data were checked with respect to outliers and trends. Gaps were filled with near station data (if available) or were interpolated.

Depending on the considered section, the long term time series covers periods of 33, 39 and 41 years. The required samples for the fitting process were generated by the block maxima method. The following extreme value distributions were fitted to the sample:

- Generalized extreme value distribution (GEV).
- Weibull distribution.
- Log-normal-distribution.
- Gumbel distribution.

For each section all distribution adaptation were graphically compared and valued by the mean absolute deviation between the plotting positions and the predicted wave heights. An example is illustrated in Figure 3 for the section Süssau. It shows the annual maximum of the wave heights as well as the fitted

[6] Deutsche Wetterdienst (DWD)
[7] Bundesamt für Seeschifffahrt und Hydrographie (BSH)

extreme value distributions. In all three sections the best results can be achieved with the GEV distribution.

This distribution can be characterized by the three parameters location μ, scale σ and shape ξ. All parameters are given in Table 2. It is possible to calculate the expected wave heights with a certain probability for each section, using the quantile function of the GEV distribution (2) and the parameters from Table 2.

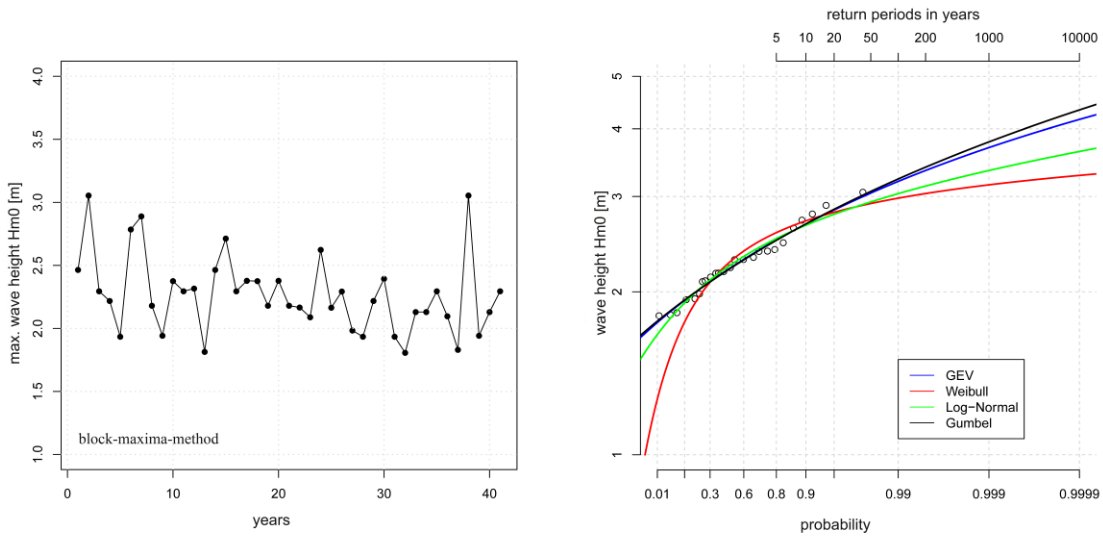

Figure 3: Identification of the proper extreme value distribution for the wave heights for the section Süssau.

$$x(P) = \mu + \sigma + \left(\frac{-\log(P)^{-\xi}}{\xi} \right) \qquad (2)$$

Table 2: Parameters for the GEV distribution.			
Section	location μ	scale σ	shape ξ
Süssau	2.133459264	0.243363857	-0.021995879
Behrensdorf/Lippe	2.174204837	0.284800679	-0.114337922
Schuby/Damp	2.265240834	0.160722949	-0.188314836

Dikes are often designed to resist extreme events with a return period of 200 years which is equal to a non-exceedance probability of P=0.995. Table 3 shows the design water level and the design wave height for the sections Süssau, Behrensdorf/Lippe and Schuby/Damp.

Table 3: Water level and wave heights with a return period of 200 years (P=0.995).		
Section	water level	wave height
Süssau	2,40 m	3,46 m
Behrensdorf/Lippe	2,36 m	3,45 m
Schuby/Damp	2,45 m	2,31 m

DETERMINATION OF COMBINED PROBABILITIES OF OCCURRENCE

Copula models

Copula models are flexible tools to describe the dependence between two or more random variables. In case of bivariate variables, the statistical correlation can be defined with Equation (3).

$$F_{X,Y}(x,y) = C[F_X(x), F_Y(y)] = C(u,v) \qquad (3)$$

$F_{X,Y}$ denotes the common cumulative distribution function for both variables which is descripted by the Copula model C. u and v represent the marginal distributions of both variables.

The theoretical basic research on Copula models was made by the american mathematician Abe Sklar in the year 1958. First used in the insurance and banking sectors, the use of Copulas for engineering questions is increasing.

The strength of dependence between the considered variable is controlled by the parameter Θ, which can be estimated by mathematical methods such as the maximum likelihood estimation.
There are several copula families available like the Gaussian Copulas and the Archimedean Copulas. Common Copulas and related methods are implemented in libraries for the software “R[8]” which was used in this project (Yan 2006).

Methods for generating samples

Using the fitted extreme value distributions of water level and wave height and a Copula model, the combined probability of occurrence of both parameters can be estimated. In the first step, it is necessary to generate a sample of extreme combinations of water levels and wave heights. For that purpose several methods had been developed and tested. The best results can be achieved with a peak over threshold sample method (POTS) and a sum based sample method (SBS). Both methods have been implemented in a library for “R”.

The POTS method is searching for flood events in which a defined water level is exceeded. Within these events the maximum water level and the maximum wave height will be extracted and saved as a sample member. The algorithm takes into account that the maximum of both parameters are not necessary occur at the same time. Events with a too low wave height will be removed from the sample automatically, by using a threshold. The sample size is controlled by the event limits which have to be defined separately for every section. Event limits are the water levels which defines the start and the exit level of an event. Especially the start level has to be adapted with respect to local conditions. In case of a high start level the sample size may be too small or even zero. A small start level could return a sample which includes non-extreme events. The samples for the considered sections where generated with a start level between 0.9 m and 1 m. This coincides with the start of flooding of the dike toe. The exit level was set to the average of the water level time series. In some cases the sample scale is too small for the fitting process of the Copula. The reasons for this could be a too small number of flood events in the time series or flood events with a too small number of sufficiently high waves.

For this case a different approach is taken with the SBS method. In this approach the sample is generated by selecting the largest sums of water levels and wave heights. The first step is merging the water levels and sea state time series by time. Subsequently both parameters are summed. The extreme combinations of water level and wave heights are now selected by a threshold. Initial studies on the threshold have shown that sum above the 0.999 quantile return acceptable results.

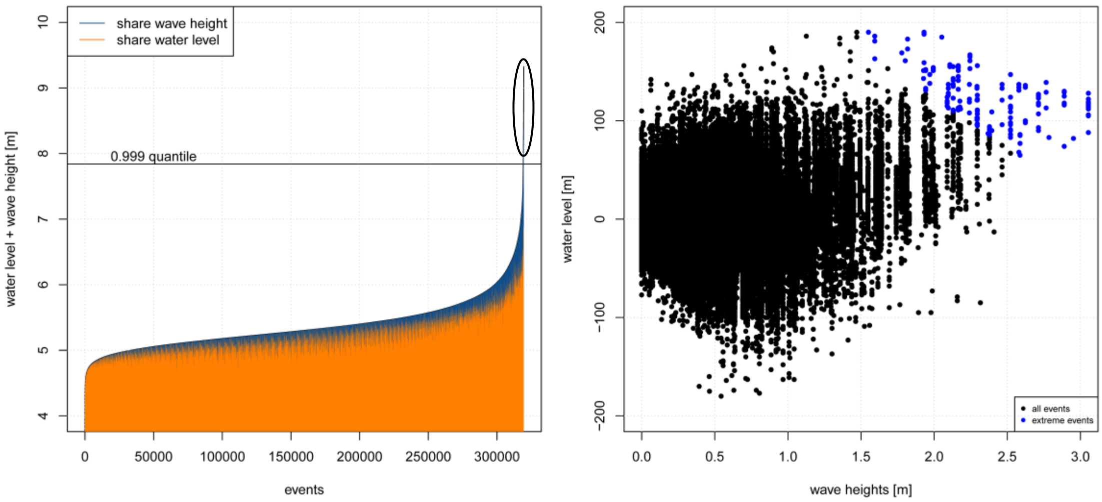

Figure 4: Summed water levels and wave heights and selected events for the section Süssau.

Figure 4 shows the sorted summed parameter and the resulting sample of the SBS method for the section Süssau. The advantage of this approach is an improved flexibility regarding to the considered

[8] “R” is a free open source programming language for statistical computing and graphics.

location. By using the quantile as a threshold, the generation of a sample is not based on flood events. This approach selects the highest combinations of water level and wave heights which are defined by the 0.999 quantile. Compared to the POTS method, the time related reference is lost. It is possible that multiple events from are selected from only one flood event.

For the further investigation of selected sections the generation of the sample was performed by the POTS method. If the sample scope was too small the SBS method was used instead.

Used Copula models

The combined probabilities of occurrence of water levels and wave heights can be described by several Copula models. The following Copula models have been considered:

- Gumbel-Copula.
- Frank-Copula.
- Clayton-Copula.
- AMH-Copula.

All listed models belong to the Archimedean Copula family. Archimedean Copulas can be specified by a generator function and are flexible to capture various dependence structures. For that reason this family is suitable for modeling extreme events (Hofert 2007). The Copula parameter Θ can be fitted by the following methods:

- Maximum likelihood method (mpl).
- Pseudo maximum likelihood method (pmpl).
- Inversion of Kendal's τ (itau).
- Inversion of Spearman's ρ (iroh).

Θ was fitted by all methods and compared by the standard deviation. Subsequently Θ was selected by the minimum deviation (see Table 4). Depending on the sample, not all fitting methods, which are implemented in the package "copula[9]" for "R", are appropriable to estimate the Copula parameter. A solution is to try different start parameters for the fitting method.

Table 4: Estimated parameters for different Copula models.

Copula model	Parameter Θ
Clayton	0.61
Gumbel	1.30
Frank	2.08
AMH	0.56

Based on Copula distribution function, the combined probability of occurrence of water level and wave height can be estimated by the so called survival function which is given with Equation (4).

$$p(x,y) = 1 - u(x) - v(x) - w(x,y) \qquad (4)$$

p is the combined probability of occurrence of x and y, u(x) and v(y) are the distribution functions for x and y and finally, w(x, y) is the multivariate distribution via Copula.

Validating the goodness of fits

The validation of the goodness of fits can be realized by graphical and mathematical approaches. One mathematical approach is to compare the empirical Copula with the theoretical Copulas. The empirical Copula can be estimated with Equation (5). R_i and S_i corresponds to rank of the sample element, n represents the sample size and I is the indicator function.

$$C_{empiric}(u,v) = \frac{1}{n}\sum_{i=1}^{n} I \cdot \left(\frac{R_i}{n+1} \le u, \frac{S_i}{n+1} \le v \right) \qquad (5)$$

The non-exceedance probability of the empirical Copula can be compared to the probabilities of the fitted theoretical Copulas. A possibility to evaluate the differences is to plot the results in a scatter plot as shown in Figure 5 for the section Süssau. Concordant probabilities are represented by points on the diagonal line. If the points are located above the line, probabilities are overestimated by the

[9] The package is documented and can be downloaded from https://cran.r-project.org/web/packages/copula/ .

theoretical Copula. By contrast, points under the line represent underestimated values. In addition to the graphical evaluation the RMSE[10] value (see Equation 6) can be used to evaluate the goodness of fit. The value measures the error between the modeled value and the empirical value. The smaller an RMSE value, the closer modeled and empirical values are.

Based on the RMSE for the sample of the section of Süssau, the minimal deviations (see Table 5) are achieved by the Gumbel Copula.

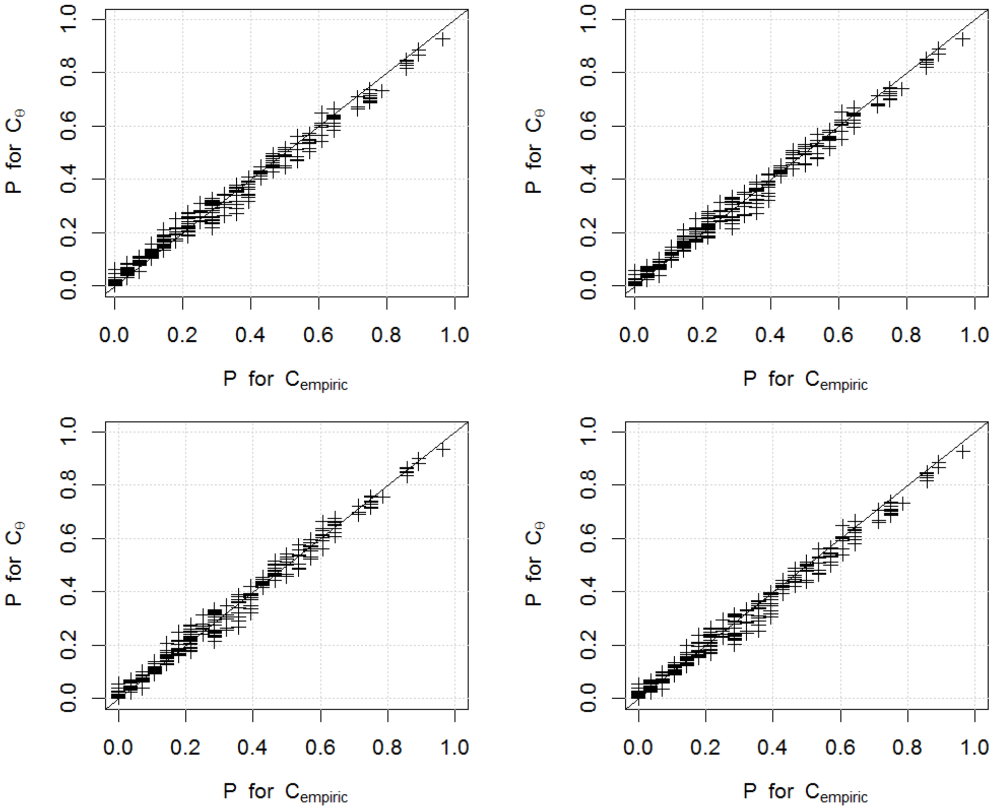

Figure 5: Comparison of the empirical and theoretical probability.

$$RMSE = \sqrt{\frac{1}{n-k}\sum_{i=1}^{n}\left[C_{empiric}(u,v) - C_{\Theta}(u,v)\right]^2} \qquad (6)$$

Table 5: RMSE values for the section Süssau.		
Copula model	RMSE	P-Value
Clayton	0.78	0.00698
Gumbel	0.56	0.1846
Frank	1.34	0.01297
AMH	0.92	0.00499

In principle, the goodness of fit can also be evaluated graphical as it is shown in Figure 6. A necessary condition is a sufficient sample size. In this study the number of observed extreme events is too small to compare the observed events with the Copula generated events. Furthermore, events with a low return period were not measured and cannot be compared with the predicted data.

The isolines in Figure 6 represent combinations of water level and wave heights with the same probability of occurrence, indicated as return periods of 50, 100 and 200 years. To get information about the distribution at the marginal areas, the plot is supplemented by random combinations of water levels and wave heights, based on the fitted distribution functions (black points). Additionally the plot shows random combined water levels and wave heights, generated by the Copula model (grey points).

---

[10] Root Mean Square Error.

In general, the goodness-of-fit can be examined by statistical hypothesis tests. The Copula package in “R” provides a function comparing the empirical process with the parametric estimate of the used Copula under the null hypothesis. Depending on the section and fitted Copula model, the function fails. For that reason the statistical test from the package “copula” was not taken into account.

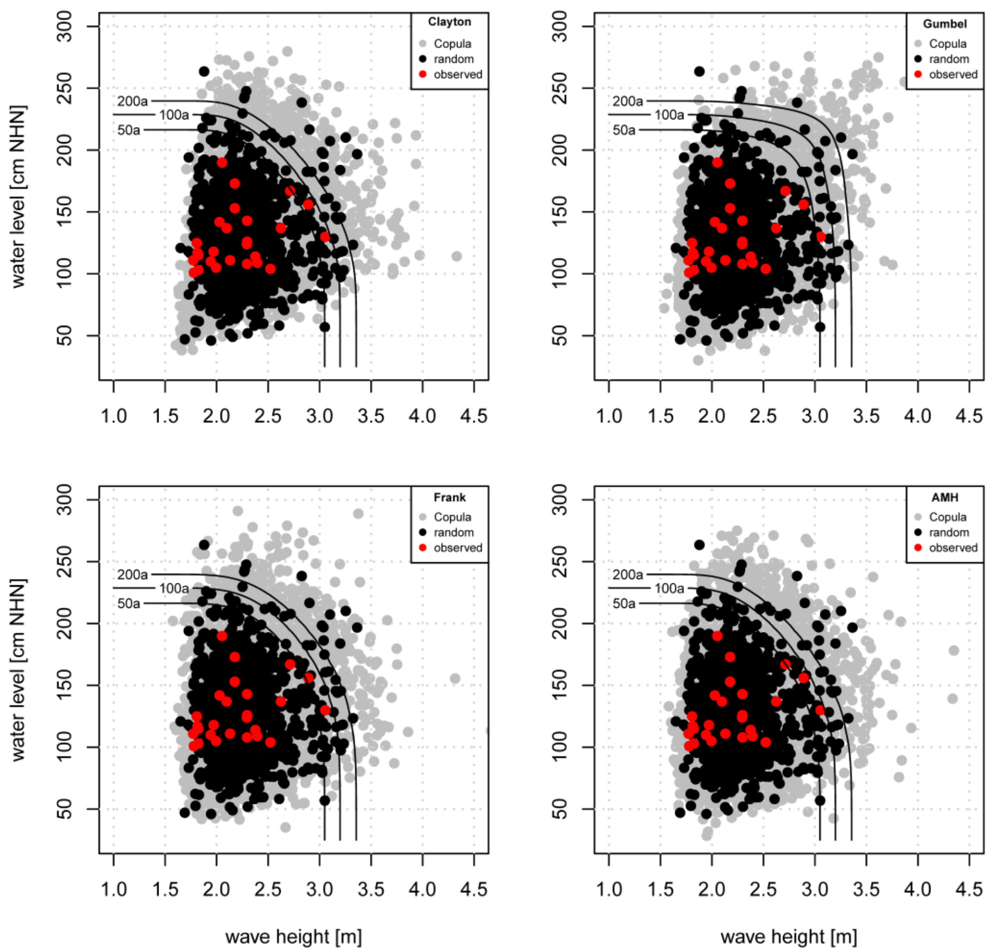

Figure 6: Graphical presentation of the studied Copula models.

RESULTS FOR THE SECTION SÜSSAU

Based on the RMSE value from Table 5, the combined probabilities of occurrence of water level and wave height were estimated by the Gumbel Copula. The occurrence of probability of water levels and wave heights are analyzed in graphical and tabular form (see Figure 7 and Figure 8). The user is able to determine the return period of a selected combination of a certain water level and wave height. To perform the safety validation of the dike, several of these combinations are used to calculate the wave overtopping charges of the dike. In the next step, the combinations of dike failure are selected. The lowest return period of these combinations gives the probability of failure of the selected dike. By comparison of these return periods, the dikes with the lowest safety level have been found.

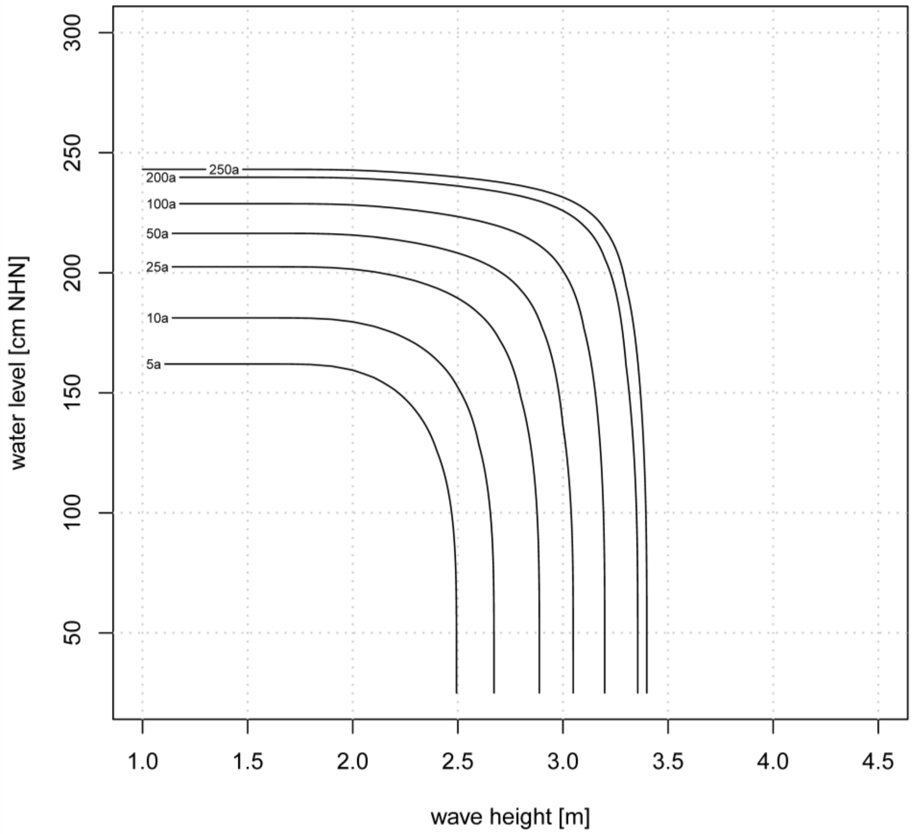

Figure 7: Combinations of water levels and wave heights with return periods of 5a, 10a, 25a 50a, 100a, 200a and 250a.

wave height	water level [cm NN]																				
[m]	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250
1,0	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	203	412
1,1	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	203	412
1,2	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	203	412
1,3	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	203	412
1,4	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	203	412
1,5	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	203	412
1,6	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	203	412
1,7	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	22	36	61	108	204	412
1,8	1	1	1	1	1	1	1	2	2	3	3	5	7	10	14	23	36	61	109	206	416
1,9	1	1	1	1	1	1	2	2	2	3	4	5	7	10	15	23	38	63	113	212	429
2,0	1	1	1	1	1	1	2	2	2	3	4	5	7	11	16	25	40	68	120	227	459
2,1	1	1	1	2	2	2	2	2	3	3	4	6	8	12	18	28	45	76	135	254	514
2,2	2	2	2	2	2	2	2	3	3	4	5	7	10	14	21	33	53	90	160	301	609
2,3	3	3	3	3	3	3	3	4	4	5	7	9	13	18	27	42	67	113	200	376	761
2,4	4	4	4	4	4	4	4	5	6	7	9	12	17	24	36	55	88	149	263	495	>1000
2,5	5	5	5	5	6	6	6	7	8	10	13	17	23	33	49	76	122	205	362	681	>1000
2,6	8	8	8	8	8	9	9	11	12	15	19	25	34	48	71	108	174	292	517	972	>1000
2,7	11	11	12	12	12	13	14	16	18	22	28	36	50	70	104	159	255	428	758	>1000	>1000
2,8	17	17	18	18	19	20	21	24	28	34	42	55	75	106	156	239	382	642	>1000	>1000	>1000
2,9	27	27	27	28	29	30	33	37	43	51	64	84	114	161	237	364	582	977	>1000	>1000	>1000
3,0	42	42	42	43	45	47	51	57	66	80	100	130	176	249	366	561	899	>1000	>1000	>1000	>1000
3,1	65	66	66	68	70	74	80	90	104	125	156	203	275	389	571	876	>1000	>1000	>1000	>1000	>1000
3,2	103	104	105	107	110	117	126	141	164	196	246	320	433	612	899	>1000	>1000	>1000	>1000	>1000	>1000
3,3	164	165	166	169	175	185	201	224	260	312	390	507	688	971	>1000	>1000	>1000	>1000	>1000	>1000	>1000
3,4	263	263	266	271	280	296	321	359	415	499	623	811	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
3,5	423	424	428	436	451	477	517	578	668	802	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
3,6	685	687	693	706	731	773	837	936	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
3,7	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
3,8	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
3,9	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
4,0	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000

Figure 8: Return periods in years of various combinations of water level and wave heights.

CONCLUSIONS

With regard to the safety validation and the design of coastal protection measures, three sections located directly at the German coast were investigated. The first step was reviewing the existing data and obtaining an overview about the missing data. Long term time series of water level and wind data were provided by the responsible authorities. Long term time series of the local sea state conditions were not available in the research sections. Water level and wind data were used as boundary conditions to set up a numerical sea state model to generate long time series of sea state data. The results, especially the wave heights and periods, where compared to measured date and corrected if necessary.

In the next step, the sea state data (wave height) were fitted to different extreme value distributions. These distributions have been evaluated. The generalized extreme value distribution (GEV) shows the best results. The distributional function for the water levels is given by the authority.

Various Copula models have been investigated to determine combined probabilities of occurrence of water levels and wave heights. The Copula models have been evaluated by different graphical and mathematical methods. On basis of the selected Copula models, tables and graphical representations of the combined probabilities of occurrence has been developed.

Finally, the authors would like to thank the State Office for Coastal Protection in Schleswig Holstein for funding the research project and the good cooperation.

REFERENCES

Yan, J. 2006. Multivariate Modeling with Copulas and Engineering Applications, Springer Handbook of Engineering Statistics, 973-989.

Seifert, F. Tauber, B. Kayser: 2001: A high resolution spherical grid topography of the Baltic Sea - 2nd edition, Baltic Sea Science Congress, Stockholm 25-29. November 2001

Hofert, M. 2007. Sampling Archimedean copulas, Computational Statistics & Data Analysis Volume 52, Issue 12, 5163-5174, August 2008.

Saathoff, F. Schlamkow, C., Völker, A. 2015, Untersuchungen zur Festlegung von Bemessungswerten auf der Grundlage kombinierter Eintritswahrscheinlichkeiten. Lehrstuhl für Geotechnik und Küstenwasserbau, Rostock, 2015 (unpublished)