



# Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a multivariate statistical approach based on Copula functions

T. Wahl, C. Mudersbach, and J. Jensen

Research Institute for Water and Environment, University of Siegen, Germany

Correspondence to: T. Wahl (thomas.wahl@uni-siegen.de)

Received: 9 August 2011 – Revised: 11 January 2012 – Accepted: 11 January 2012 – Published: 29 February 2012

**Abstract.** This paper presents an advanced approach to statistically analyse storm surge events. In former studies the highest water level during a storm surge event usually was the only parameter that was used for the statistical assessment. This is not always sufficient, especially when statistically analysing storm surge scenarios for event-based risk analyses. Here, Archimedean Copula functions are applied and allow for the consideration of further important parameters in addition to the highest storm surge water levels. First, a bivariate model is presented and used to estimate exceedance probabilities of storm surges (for two tide gauges in the German Bight) by jointly analysing the important storm surge parameters “highest turning point” and “intensity”. Second, another dimension is added and a trivariate fully nested Archimedean Copula model is applied to additionally incorporate the significant wave height as an important wave parameter. With the presented methodology, reliable and realistic exceedance probabilities are derived and can be considered (among others) for integrated flood risk analyses contributing to improve the overall results. It is highlighted that the concept of Copulas represents a promising alternative for facing multivariate problems in coastal engineering.

## 1 Introduction

Scenario- or event-based flood risk analyses in coastal areas are often performed by following the so-called Source-Pathway-Receptor-Concept (e.g. Oumeraci, 2004). One of the main challenges consists in estimating the hydrodynamic boundary conditions. Possible future sea level changes have to be taken into account, as well as storm surge scenarios and wind waves. The latter may coincide with the high storm surge water levels and play an important role for some investigation areas, while they can be neglected for others (e.g. lee

side of islands). All of the different loading factors for the existent coastal defence structures have to be jointly examined within integrated risk analyses, as performed in the German joint research project XtremRisK (<http://www.xtremrisk.de>) for the city of Hamburg and Sylt Island in the German North Sea.

Results from recently analysing observed mean sea level changes in the German Bight are summarised in Wahl et al. (2010a, 2011a). The interaction between mean sea level changes and changes in storm surge heights (i.e. total water levels arising from a combination of astronomical tides and a meteorological induced surge component) have recently been investigated by Mudersbach et al. (2012). The present study focuses on the multivariate statistical assessment of storm surge events, including the wave conditions where necessary. Most former studies only considered the storm surge water levels to determine exceedance probabilities (e.g. Jensen et al., 2006; Haigh et al., 2010; Mudersbach and Jensen, 2010). However, especially for risk analyses, where the complete storm surge curve is used to identify the initial conditions for flood propagation in the hinterland, the temporal behaviour of the storm surge water levels should also be taken into account for the statistical assessment. When exclusively analysing the maximum storm surge water levels, a storm surge event with two or more high tides in a row has the same exceedance probability as a storm surge event with only one high tide and the same maximum water level. At the same time the temporal behaviour of storm surge water levels may significantly affect the potential losses along the coastal defence line or in the hinterland. Cai et al. (2008), for example, refer to a flood event which occurred in North Wales in February 1990 (namely the Towyn flood) and where defence breaches occurred during the initial high tide while flooding arose from three successive high tides (see also HR Wallingford, 1990). In a recent

study, Ruocco et al. (2011) first ranked flood events that occurred along the south coast of the UK (Southampton and Portsmouth) by looking at the storm surge water levels. Secondly, they reconstructed coastal flooding based on the information from media sources (first of all newspapers). They also report a flood event (from December 1989) that was ranked to be only the 70th highest for Portsmouth, but still resulted in significant flooding due to the long duration of elevated high waters. This highlights the necessity to include such information also into statistical assessments (especially within flood risk analyses) to be able to determine realistic exceedance probabilities. A storm surge event with a moderate maximum water level, but consisting of two or three high tides in a row, should have a small exceedance probability. The same applies for an extremely high storm surge event, where the maximum water level is only reached for a short time period and the surrounding high tides are much lower.

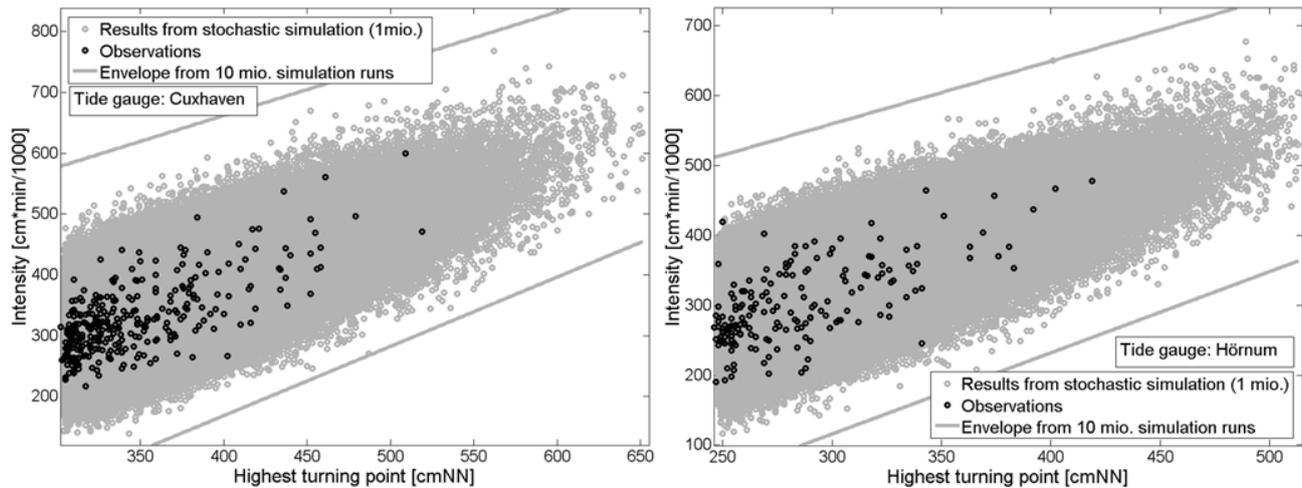
Hence, more storm surge parameters have to be included within the statistical assessment. This requires the calculation of joint exceedance probabilities. If the parameters are not independent from each other multivariate statistical models have to be applied or the data sets have to be filtered as, for example, described by Tawn (1988). Many different models, which are first of all bivariate, are described in literature: for example, the bivariate (or trivariate) Normal, the bivariate Gumbel, or the bivariate Gamma models (see e.g. Kotz and Nadarajah, 2000 for an overview). Applications of bivariate models in coastal engineering mostly focussed on the joint analysis of high water levels and wave heights or wave periods (e.g. Coles and Tawn 1994; HR Wallingford, 1990, 2000; Hawkes et al., 2002; Galiatsatou and Prinos, 2007; Hanson and Larson, 2008; Hawkes, 2008) or astronomical tidal water levels and surges (e.g. Pugh and Vassie, 1979; Tawn and Vassie, 1989; Tawn, 1992; Dixon and Tawn, 1994; McMillan et al., 2011).

Most of the multivariate models suffer from the drawback that the marginal parameters need to be independent from each other or that the marginal distributions need to be from the same family. When working with a bivariate Gumbel model, it is, for example, assumed that both of the considered parameters are Gumbel distributed. However, in reality this assumption does often not apply and one has to deal with dependent marginal parameters with different distributions. In such cases Copula functions, first mentioned by Sklar (1959), may be used. Copulas are very flexible joint distributions. They are able to handle mixed marginal distributions and account for the structure of dependence overlooking the margins. When using Copulas, the dependence function is studied separately from the marginal distributions (Salvadori et al., 2007). Although the theory of Copulas is not new, the number of papers dealing with Copula functions in many different ways has significantly increased over the last decade or so. Mikosch (2006) reports that a Google search for the word “Copula” resulted in 10 000 responses in 2003, while today – in July 2011 –

the same search leads to more than 2.6 million responses. Many of the available journal papers are related to mathematical finance, risk management, insurance, econometrics and hydrology (first of all multivariate hydrological frequency analyses, e.g. DeMichele and Salvadori, 2003; Favre et al., 2004; Salvadori and De Michele, 2004; Klein et al., 2008; Karmakar and Somonovic, 2009). An interesting website providing an overview of available papers dealing with Copulas in water science is hosted by the “Statistics in Hydrology Group” (<http://www.stahy.org>). The website contains only 2 references from the year 2003, 9 references from 2004 and more than 30 references from 2010, which again highlights the explosion of activity in this field. In contrast, very few authors addressed coastal engineering problems by using bivariate Copulas. So did, for example, de Waal and van Gelder (2005) and Serinaldi and Grimaldi (2007) to model wave heights and periods. Sto. Domingo et al. (2010) used Copulas to calculate the joint probabilities of storm surge water levels and durations. Some of the ideas forming the basis for the methodology that is presented here in detail have already been summarised by Wahl et al. (2010b). The approach described in there has recently been adopted by Salecker et al. (2012) to perform similar analyses (i.e. bivariate statistical storm surge analyses with Copulas) in the Baltic Sea. One of the advantages of Copulas is given by the possibility to extend the models and add further dimensions. Different authors recently considered higher-dimensional (first of all trivariate) Copula functions with respect to hydrological data sets to perform flood frequency or rainfall frequency analyses, respectively (e.g. Grimaldi and Serinaldi, 2006a, b; Serinaldi and Grimaldi, 2007; Zhang and Singh, 2007). Wong et al. (2010) performed drought analyses and De Michele et al. (2007) analysed sea storms based on a trivariate Copula model by taking the parameters wave height, storm duration and storm direction into account. Pinya et al. (2009) applied a nested Copula model to assess the risk of flooding in a tidal sluice regulated catchment. A comprehensive review of multivariate Archimedean Copula models is provided by Berg and Aas (2007).

Although Copulas have many advantages when addressing multivariate problems and performing statistical assessments (especially within risk analyses), Mikosch (2006) provides a very critical review on Copulas. This resulted in an extensive expert discussion (Genest and Remillard, 2006; de Vries and Zhou, 2006; Segers, 2006). However, as outlined in the following sections, Copula functions represent a powerful alternative to deal with various multivariate problems in coastal engineering.

Here, for the first time the two important storm surge parameters “highest turning point” ( $S$ ) (i.e. the maximum storm surge water level) and “intensity” ( $F$ ) (i.e. the area between the observed storm surge water level and the German ordinance datum NN) are taken into account for statistical storm surge analyses. The bivariate model is then extended to additionally include selected wave parameters. The three main



**Fig. 1.** Results from stochastic storm surge simulation performed by Wahl et al. (2011b) for the tide gauges of Cuxhaven (left) and Hörnum (right).

objectives of the paper are to: (1) present a bivariate statistical approach based on Copula functions to jointly analyse the two storm surge parameters  $S$  and  $F$  (including the application of goodness of fit (GoF) tests for the model selection), (2) present a trivariate (nested) Copula model to jointly analyse the two storm surge parameters and the significant wave height  $H_s$  (including GoF tests) and (3) present results for selected investigation areas in the German Bight (i.e. Cuxhaven, Hörnum and Westerland). The presented results are used to perform scenario-based risk analyses within the XtremRisK project.

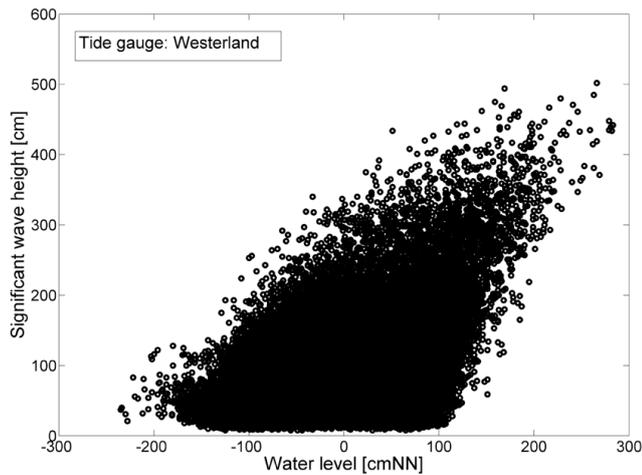
The paper is organised as follows: in Sect. 2 the considered data sets are introduced. Section 3 contains detailed information about the applied methodology based on the Copula theory (Sklar, 1959). The overall results are presented and discussed in Sect. 4 and the paper closes with a summary of the key findings and the conclusions in Sect. 5.

## 2 Data

In a companion paper, Wahl et al. (2011b) present a methodology to stochastically simulate a large number of storm surge scenarios for flood risk analyses. Simulated storm surge events cover three tidal cycles and have a temporal resolution of 1-min. From the study, 10 million storm surge scenarios are available for the tide gauges of Cuxhaven (located in the Elbe estuary; coordinates:  $53^{\circ}52'04''$  N,  $8^{\circ}43'03''$  E) and Hörnum (located in the southeast of Sylt Island; coordinates:  $54^{\circ}45'29''$  N,  $8^{\circ}45'29''$  E). These scenarios are used here as a data basis for bivariate (and trivariate) statistical storm surge analyses. The simulation results are based on 314 observed events for Cuxhaven (from 1900 to 2008) and 175 for Hörnum (from 1936 to 2008). The available data set is shown in Fig. 1, where the storm surges are represented by the two parameters “highest turning point” and “intensity” (the unit for the intensity has been divided by thousand for

presenting purposes). These parameters are also taken into account for the statistical assessment. The observed storm surge events are shown as black dots. One million of the simulated events are shown as grey dots and envelopes calculated from the 10 million storm surge scenarios are displayed for presenting purposes.

In Sect. 3.3 a trivariate statistical model is introduced, which allows the inclusion of selected wave parameters in the statistical analyses. Therefore, observational records for different wave parameters are needed. For both investigation areas, Cuxhaven and Hörnum, wind waves do not play an important role due to the locations in an estuary and the lee of an island, respectively. Furthermore, no observational data sets are available. Hence, an empirical model to transfer the simulated storm surge events from the tide gauge of Hörnum to the tide gauge of Westerland is introduced in Sect. 3.2. The tide gauge of Westerland is located on the west side of Sylt Island (coordinates:  $54^{\circ}54'31''$  N,  $8^{\circ}16'16''$  E), where high wind waves occur due to the exposed location. The tide gauge provides high resolution sea level data for the period from 1988 to 2007 and wave data are available for the same time period from a measuring station near the tide gauge (coordinates:  $54^{\circ}55'2''$  N,  $8^{\circ}13'18''$  E; water depth: 13 m). The sea level data set is used to compile the empirical transfer model for the storm surges (see Sect. 3.2) and the wave measurements are considered for the trivariate (or 3-dimensional) statistical assessment (see Sect. 3.3). Figure 2 shows a scatter plot of simultaneously measured water levels (referred to the German ordnance datum NN) and significant wave heights  $H_s$ . Wave heights of up to 5 m have been observed in Westerland in the past and the scatter plot highlights an existent dependency between the two parameters. High waves tend to coincide with high water levels raising the overall flood risk. This has to be taken into account for the trivariate statistical analyses.



**Fig. 2.** Observed water levels (from tide gauge) and significant wave heights (from wave measurement station) in Westerland on the west side of Sylt Island.

### 3 Methods

#### 3.1 Bivariate statistical model and Goodness of Fit tests

The exceedance probabilities of the available storm surge events are expressed as joint probabilities of the two storm surge parameters  $S$  and  $F$ . Figure 1 shows that these two parameters are not independent from each other. Thus, the simple multiplication of the exceedance probabilities of the margins does not represent the joint probabilities and a multivariate model has to be applied. From testing different parametric distribution functions for the marginal parameters  $S$  and  $F$ , it was found that the Generalized Pareto distribution (GPD) fits best for the parameter  $S$ . This is not surprising, as the GPD has been used by Wahl et al. (2011b) to simulate this parameter. The parameter  $F$  is not directly simulated but is given implicitly by the stochastic storm surge model. The LogNormal distribution is found to be the most appropriate parametric distribution function for this parameter. Hence, one has to deal with dependent marginal parameters with different distributions. As outlined in Sect. 1, Copula functions are valuable to analyse such multivariate data sets and are applied in the following. Before the theoretical background of Copulas is briefly introduced, appropriate univariate marginal distributions for the two storm surge parameters  $S$  and  $F$  have to be defined. It has been mentioned above that the GPD has been identified to be the most suitable distribution for  $S$  and the LogNormal distribution for  $F$ . As the margins are analysed separately from the dependence function when using Copulas, this also allows for taking nonparametric marginal distributions, such as Kernel Density Functions (KDFs) into account. Especially when large numbers of realisations are available for the marginal parameters, as is the case for the present study, such nonparametric

functions lead to good results. The uncertainties are smaller as when fitting certain parametric functions to the available data sets. Here, for both of the parameters  $S$  and  $F$  the following Gaussian (or Normal) Kernel Density Function  $K(x)$  is applied (see e.g. Karmakar and Simonovic, 2008):

$$K(x) = (2\pi)^{-1/2} e^{-(x^2)/2}, \text{ with } \int_{-\infty}^{\infty} K(x) dx = 1 \quad (1)$$

To calculate joint probabilities from the marginal distributions, different Copula functions belonging to different Copula families are available. In the relevant literature, the applied Copulas often belong to the Elliptical, the Normal, the  $t$ -Student or the Archimedean family. Especially Copula functions belonging to the Archimedean family are often used for hydrological analyses (e.g. Favre et al., 2004), as they are flexible and easy to construct. Here, three Archimedean Copula functions – namely the Clayton, Frank and Gumbel Copulas – are considered. These Copulas were chosen as many of the authors mentioned in Sect. 1 outlined their applicability for multivariate frequency analyses and more important, the three Copulas cover the full range of tail behaviour. The Clayton Copula has lower tail dependence, while the Frank Copula has no tail dependence, and the Gumbel Copula has only upper tail dependence (Schölzel and Friedrichs, 2008).

Sklar (1959) describes the connection between a Copula  $C$  and a bivariate cumulative distribution function (cdf)  $F_{XY}(x, y)$  of any pair  $(X, Y)$  as follows (also known as Sklar's theorem):

$$F_{XY}(x, y) = C[F_X(x), F_Y(y)] \quad (2)$$

where  $F_X(x)$  and  $F_Y(y)$  are the univariate marginal distributions. The bivariate probability density function (pdf) reads as:

$$f_{XY}(x, y) = c[F_X(x), F_Y(y)] f_X(x) f_Y(y) \quad (3)$$

where  $f_X(x)$  and  $f_Y(y)$  represent the pdf's of the margins. Let  $U$  and  $V$  be uniformly distributed random variables defined as  $U = F_X(x)$  and  $V = F_Y(y)$ , then the function  $c(u, v)$  (sometimes referred to as the Copula density function) is given by:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \quad (4)$$

Further important features of Copulas and information about the theoretical background can be found in Nelsen (1999), who provides a detailed introduction to the subject.

The Archimedean Copulas considered for the present study are constructed based on the so-called Copula generator  $\phi: [0, 1] \rightarrow [0, \infty]$ , being a strictly monotonically decreasing function with  $\phi(1) = 0$  (e.g. Nelsen, 1999). The general form of a one-parametric Archimedean Copula is:

$$C_\theta(u, v) = \phi^{-1}[\phi(u) + \phi(v)] \quad (5)$$

**Table 1.** Archimedean Copula functions considered for the present study and their generator functions, ranges for the Copula parameters  $\theta$  and connections to Kendall’s  $\tau$ .

Copulafunction $C_\theta$	Generator $\phi(t)^{**}$	Range $\theta \in$	Kendall’s $\tau$
Clayton (or Cook-Johnson) $\left[ u^{-\theta} + v^{-\theta} - 1 \right]^{-\frac{1}{\theta}}$	$t^{-\theta} - 1$	$[0, \infty)$	$\frac{\theta}{\theta+2}$
Frank $-\frac{1}{\theta} \ln \left[ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right]$	$-\ln \left( \frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)$	$(-\infty, \infty) \setminus \{0\}$	$1 - \frac{4}{\theta} [1 - D_1(\theta)]^*$
Gumbel (or Gumbel-Hougaard) $\exp \left\{ - \left[ (-\ln u)^\theta + (-\ln v)^\theta \right]^{\frac{1}{\theta}} \right\}$	$(-\ln t)^\theta$	$[1, \infty)$	$1 - \theta^{-1}$

\* 1. Debye Function  $D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt$ ; \*\*  $t = u$  or  $t = v$ .

Table 1 contains an overview of the Copula functions used for the present study. Functions of the generators  $\phi(t)$  are displayed, as well as the connection between the Copula parameter  $\theta$  and the rank correlation Kendall’s  $\tau$ . The latter represents a well-known nonparametric measure of dependence and is calculated from the available observations as:

$$\tau = \binom{N}{2}^{-1} \sum_{j=1}^N \sum_{i=1}^j \text{sign}[(x_i - x_j)(y_i - y_j)] \tag{6}$$

with  $\text{sign} = 1$  if  $[(x_i - x_j)(y_i - y_j)] > 0$ ,  $\text{sign} = -1$  if  $[(x_i - x_j)(y_i - y_j)] < 0$  and  $i, j = 1, 2, \dots, N$  (e.g. Kamarkar and Simonovic, 2009). Another nonparametric measure of dependence is given by Spearman’s rank correlation  $\rho$  (Spearman, 1904), which is not used for the present study. Table 1 shows that Kendall’s  $\tau$  (here, for the storm surge parameters  $S$  and  $F$ ) is the only parameter, which is required (in addition to the marginal distributions) to solve the Copula functions. The values for Kendall’s  $\tau$  are found to be  $\tau = 0.43[-]$  for both of the tide gauges Cuxhaven and Hörnum.

With this information the Copula parameters  $\theta$  (as shown in Table 1) and joint probabilities for the parameters  $S$  and  $F$  can be calculated. Alternative methods to derive the Copula parameters  $\theta$ , such as the maximum pseudolikelihood estimator, are described, for example, by Genest and Favre (2007). Corresponding to the univariate case, appropriate Copula functions have to be identified based on GoF tests before starting the statistical analyses. As the margins are analysed separately, the GoF tests aim to identify a Copula function, which is able to capture the existent structure of dependence between the considered parameters. Extensive research efforts have been undertaken in recent years to improve GoF tests for multivariate problems and Copulas, respectively. A comprehensive review including a power study is provided by Genest et al. (2009), as well as by Berg (2009). Some results from sensitivity studies can be found in Berg

and Quessy (2009). For the present study, two GoF tests are applied to identify proper Copula functions. The first test is based on a comparison of the theoretical and empirical joint non-exceedance probabilities. The theoretical joint non-exceedance probabilities are calculated with the three Copula functions from Table 1, where the marginal distributions are KDFs (see Eq. 1) derived from the available observations. The empirical joint non-exceedance probabilities are calculated as shown in Eq. (7) (see e.g. Yue et al., 1999; Kamarkar and Simonovic, 2009). This is an extension of the approach introduced by Gringorten (1963) to derive unbiased plotting positions for the Gumbel distribution. For univariate statistical storm surge analyses in the German Bight area Gringorten’s approach was already used by Jensen et al. (2006) and Mudersbach and Jensen (2010) to estimate empirical probabilities (alternative approaches for the univariate case are summarised in Stedinger et al., 2008).

$$F_{XY}(x_i, y_i) = P(X \leq x_i, Y \leq y_i) = \frac{\sum_{m=1}^i \sum_{l=1}^i N_{ml} - 0.44}{N + 0.12} \tag{7}$$

where the pairs  $(x_m, y_l)$  are arranged in ascending order with respect to  $x_m$  and  $N_{ml}$  represents the number of occurrences of  $(x_m, y_l)$  with  $x_m < x_i$  and  $y_l < y_i$ ,  $i = 1, \dots, N$  and  $1 \leq m, l \leq i$ .  $N$  is the sample size (here: 314 for Cuxhaven and 175 for Hörnum). The simulation results are not considered for this test as the structure of dependence between  $S$  and  $F$  is found to be the same in the observations and the simulation results (see Fig. 1). The root mean squared errors (RMSEs) and the maximum distances (for a Kolmogorov-Smirnov (KS)-test) are calculated from the theoretical and the empirical joint non-exceedance probabilities to identify the most appropriate Copula function (e.g. Zhang and Singh, 2007). The RMSEs and the KS-values are determined for all of the observed events, as well as for rare events

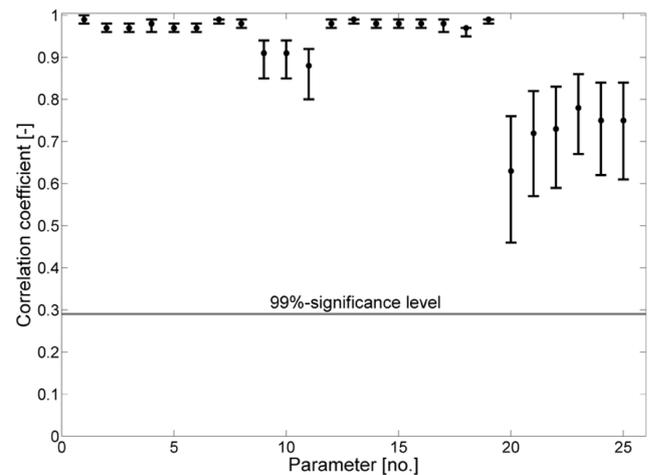
with empirical non-exceedance probabilities  $>0.8 [1/a]$  (see Sect. 4.1).

A second GoF test is applied. This test compares observed and simulated (with different theoretical Copula functions) pairs of the two marginal parameters (see e.g. Serinaldi and Grimaldi, 2007; Klein et al., 2008). In contrast to most other GoF tests, this test does not calculate any critical value of a statistic. First, large numbers (here: 1 million) of random pairs  $(u_i, v_i)$  are generated based on the different theoretical Copula functions shown in Table 1. With the marginal distributions  $F_X(x)$  and  $F_Y(y)$  the generated pairs are subsequently transformed into the original units of the considered parameters (here  $S$  and  $F$ ). Finally, the simulated pairs are superimposed by the observed pairs of  $S$  and  $F$  (or here by pairs of  $S$  and  $F$  from the stochastic storm surge simulation). From the resulting plots it is easy to identify a theoretical Copula function, which is able to model the structure of dependence between the marginal parameters.

### 3.2 Empirical transfer functions

As it has been mentioned in Sect. 2, wind waves do not represent an important loading factor for existent coastal defence structures (i.e. first of all dikes) in the investigation areas Cuxhaven and Hörnum. However, in other areas, for example Westerland on the west side of Sylt Island, high wind waves may coincide with storm surge water levels. For Westerland, a long record (from 1988 to 2007; see Sect. 2) of simultaneously measured wave parameters and water levels is available (data sets from after 2007 were not available at the time of the analyses), but no stochastically simulated storm surge events. To transfer 10 million storm surges from Hörnum to Westerland, an empirical model is introduced in the following.

The model is mainly based on regression functions, which are compiled for the 25 storm surge parameters considered by Wahl et al. (2011b) to parameterise an observed storm surge event consisting of three tides. The parameterisation scheme is based on 19 sea level parameters (i.e. the tidal high and low waters and the water levels one hour before and one hour afterwards) and 6 time parameters (i.e. the time periods between two adjacent high and low waters). The available observations from Westerland and Hörnum include 64 storm surge events, which have been recorded at both sites. From parameterising the 64 storm surge events for Westerland and Hörnum, regression functions are derived for all of the 25 parameters. In addition to the 64 observed storm surge events, selected extreme events from hydrodynamic model runs are considered to build the model (Jensen et al., 2006). Therewith, a wider range of possible values is covered and regression functions can be reliably estimated. Correlation coefficients  $r$  are calculated for the 25 parameter time series from Hörnum and Westerland (see Fig. 3). All correlation coefficients are significant (on the 99 %-significance level from  $t$ -test statistics) with most of them being larger than



**Fig. 3.** Correlation coefficients for the 25 storm surge parameters calculated based on the parameter time series from the tide gauges of Hörnum and Westerland and 99 %-significance level from  $t$ -test statistics.

$r = 0.9[-]$ . Slightly weaker correlations are found for the parameters 9 to 11, where parameter 10 is the “highest turning point” and the parameters 9 and 11 are the surrounding parameters (i.e. the water levels one hour before and one hour afterwards). High frequency variations often occur in the water level time series around the storm surge peaks. These variations mainly result from small wind generated waves, which are poorly damped by the tide gauges. This increases the uncertainties for the affected parameters when parameterising the observed storm surge events. Lowest coefficients (in the order of  $r = 0.6[-]$  to  $r = 0.8[-]$ ) are found for the time parameters (i.e. the parameters 20 to 25). Uncertainties from automatically parameterising observed storm surge events are in general higher for the time parameters compared to the sea level parameters.

Based on the 25 regression functions, all of the 25 parameters of a particular storm surge event are transferred from Hörnum to Westerland, where the storm surge curve is reconstructed by applying piecewise cubic hermite interpolation. As an example, for the important parameter 10 the following second order polynomial function is considered as a transfer function:

$$P_{10, \text{Westerland}} = 0.0013 \times (P_{10, \text{Hörnum}})^2 + 0.0464 \times P_{10, \text{Hörnum}} + 113.57 \quad (8)$$

With the empirical model all 10 million storm surge scenarios are transferred from Hörnum to Westerland. By taking these storm surge events and the available wave measurements into account, a trivariate model as described in the following Sect. 3.3 is used to calculate joint probabilities. The value of Kendall’s  $\tau$  for the parameters  $S$  and  $F$  is found to be  $\tau = 0.46[-]$  for the tide gauge of Westerland.

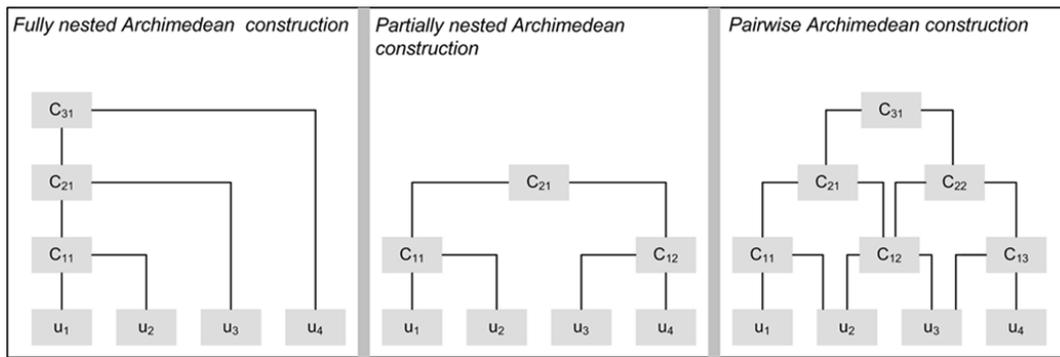


Fig. 4. Selected models to construct higher-dimensional Copulas.

### 3.3 Trivariate statistical model and Goodness of Fit tests

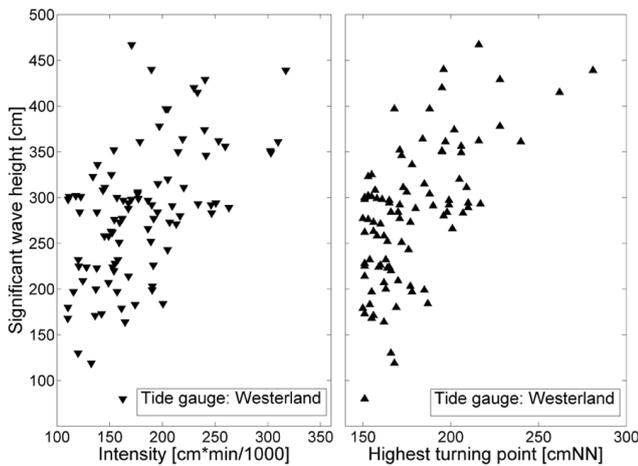
It has been outlined in the introduction, that for some investigation areas not only storm surge scenarios, but also reasonable wave conditions have to be taken into account for integrated flood risk analyses. Especially during storm surges, high wind waves may occur and potentially lead to damages along the coastal defence line or in the hinterland due to high overtopping rates. Here, a trivariate (or 3-dimensional) model is applied to jointly analyse selected wave parameters and the storm surge parameters  $S$  and  $F$ . For the present study, the significant wave height  $H_s$  is considered as it represents one of the most important wave parameters (alternatively other parameters, such as the wave period can be used). In Sect. 1, it has been outlined that one of the advantages of Copulas is given by the possibility to extend the models to the  $d$ -dimensional case, where Eq. (5) becomes:

$$C(u_1, u_2, \dots, u_d) = \varphi^{-1}[\varphi(u_1) + \varphi(u_2) \dots + \varphi(u_d)] \quad (9)$$

Three higher-dimensional Copula models discussed by Berg and Aas (2007) are briefly introduced in the following. They are subsequently used to calculate exceedance probabilities of combinations of storm surges (represented by the parameters  $S$  and  $F$ ) and wave conditions (represented by  $H_s$ ). The first model is referred to as fully nested Archimedean Copula construction (FNAC). Figure 4 (left) shows the construction scheme for a 4-dimensional Copula, where one dimension is added step by step. The 4-dimensional Copula consists of three bivariate Copulas ( $C_{11}, C_{21}, C_{31}$ ) with three generators ( $\phi_{11}, \phi_{21}, \phi_{31}$ ) (note that the subscripts of  $C$  and  $\phi$  do not refer to the dimensions). This approach is restricted in its flexibility, as four parameters are analysed, but only three mutual bivariate dependence structures are freely specified. The remaining Copula and distribution parameters are implicitly given through the construction. In Fig. 4 the pairs  $(u_1, u_3), (u_1, u_4), (u_2, u_3)$  and  $(u_2, u_4)$  all have Copula  $C_{21}$  with the related dependence parameter  $\theta_{21}$  (i.e. all of the mentioned pairs need to have a similar rank correlation). Furthermore, it is required that

the degree of dependence decreases with the level of nesting. The same constraints as for the FNAC model also apply for the partially nested construction (PNAC) (Fig. 4, middle). First, the parameters  $(u_1, u_2)$  and  $(u_2, u_3)$  are coupled via Copulas  $C_{11}$  and  $C_{12}$ , respectively, before the two bivariate Copulas are coupled via a third bivariate Copula  $C_{21}$ . Again, the pairs  $(u_1, u_3), (u_1, u_4), (u_2, u_3)$  and  $(u_2, u_4)$  all have Copula  $C_{21}$  with the related parameter  $\theta_{21}$ . The third model is referred to as pairwise Archimedean construction (PAC). This model is very flexible as more Copulas than parameters are considered (see Fig. 4, right). The Copulas for each pair of variables can be freely chosen and do not even have to belong to the same family. However, the construction is more complicated and requires more computational efforts. For more information on the described models and the theoretical background see Berg and Aas (2007) and the literature referenced therein.

Here, the parameters  $S, F$  and  $H_s$  are jointly analysed. The parameters are all cross correlated with each other (see detailed explanation below). This requires the application of a trivariate model and for the present study a fully nested approach (FNAC) as shown in Fig. 4 (left) is considered. First, it has to be tested whether the multivariate data set faces the abovementioned criteria for the application of a FNAC model. Hence, it has to be proven, whether the parameter pairs  $(F, H_s)$  and  $(S, H_s)$  have at least similar rank correlations or not. Further, the values for Kendall's  $\tau$  have to be smaller than  $\tau = 0.46[-]$ , which is the rank correlation calculated for the pair  $(S, F)$  for the tide gauge of Westerland (see Sect. 3.2). It has been mentioned that the degree of dependence must decrease with the level of nesting. To measure the dependence for the parameter pairs  $(S, H_s)$  and  $(F, H_s)$  wave events which occurred simultaneously with high water levels have to be analysed. Here, a comparable small threshold for the water level of  $W = 85$  cm above MHW is chosen to identify storm surge events from the available water level time series for the tide gauge of Westerland. On the one hand, this threshold still represents a high water level, which has been exceeded 95 times (two events have to be at least 30 h apart from each other) during the 20 yr of



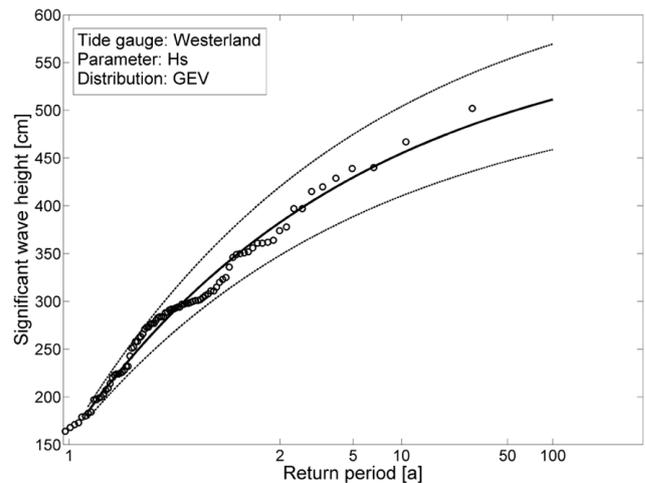
**Fig. 5.** Scatter plots for the parameter pairs  $(F, H_s)$  (left) and  $(S, H_s)$  (right).

observations (number of occurrences per year is  $<5$ ). On the other hand, a number of 95 events is large enough to capture the existent structure of dependence between the different parameters. The value for Kendall's  $\tau$  for the parameters  $S$  and  $F$  is slightly higher when the smaller threshold is considered ( $\tau = 0.48[-]$ ) but does not change significantly. Figure 5 shows scatter plots for the parameter pairs  $(F, H_s)$  and  $(S, H_s)$  and it is obvious that both parameter pairs have similar structures of dependence. The values for Kendall's  $\tau$  are found to be  $\tau = 0.36[-]$  for the pair  $(F, H_s)$  and  $\tau = 0.35[-]$  for the pair  $(S, H_s)$ . There is no physical reason for the two parameter pairs to have the same (or at least similar) rank correlations, but it is required for the FNAC model (if the rank correlations were different, a more complex PAC approach could be used instead of the FNAC model). Here, both criteria for the application of a FNAC model are fulfilled: the pairs  $(F, H_s)$  and  $(S, H_s)$  have the same Copula (i.e. the same structure of dependence) and the existent dependency is weaker than for the pair  $(S, F)$ .

Again, KDFs are used as marginal distributions for the parameters  $S$  and  $F$ , resulting from transferring 10 million storm surge events from Hörnum to Westerland. The KDFs are shown in Fig. 11. For the third parameter  $H_s$ , 95 realisations are available and a Generalized Extreme Value distribution (GEV) of the following form is fitted to the data set:

$$GEV(x) = \exp \left[ - \left( 1 + k \times \frac{x-a}{b} \right)^{-1/k} \right] \quad (10)$$

where  $a$ ,  $b$  and  $k$  represent location, scale and shape parameters, respectively. Figure 6 shows the results from fitting the GEV (with 95 %-confidence levels) to the available data set of  $H_s$ . The distribution parameters are estimated with maximum likelihood approach (e.g. Rao and Hamed, 2000) and the plotting positions are derived by following the approach proposed by Gringorten (1963). A significant wave height of



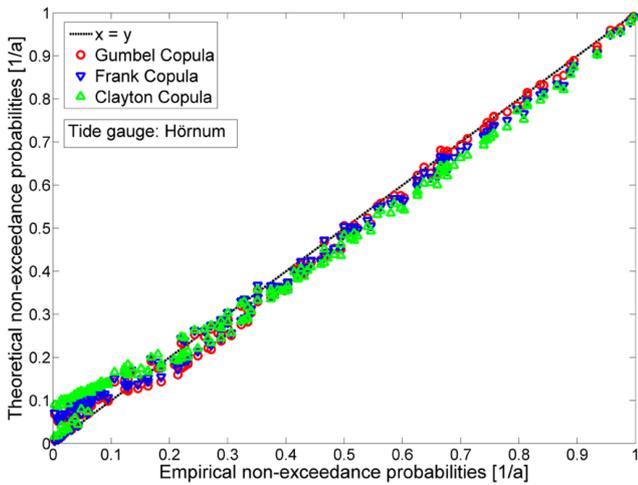
**Fig. 6.** Marginal distribution for the parameter  $H_s$  based on a GEV (with 95 %-confidence bounds).

$H_s = 300$  cm has a return period of about 1 yr, an event with  $H_s = 400$  cm of about 2.7 yr and a 100-yr event is represented by a significant wave height of about  $H_s = 510$  cm.

Now that all three marginal distributions, as well as the dependence structures between the considered parameters are known, a trivariate FNAC model can be applied to estimate the joint exceedance probabilities. According to the 1- and 2-dimensional cases, GoF tests are applied to identify appropriate Archimedean Copulas for the model. Here, the same GoF tests as described in Sect. 3.1 are taken into account. The first test compares theoretical and empirical joint non-exceedance probabilities. Theoretical probabilities are calculated with the Copulas in Table 1, whereas the Gumbel Copula is used to combine the parameters  $S$  and  $F$  and the Clayton, Frank and Gumbel Copulas are tested to incorporate the third parameter  $H_s$ . Marginal distributions are KDFs for the parameters  $S$  and  $F$  (as in Sect. 3.1) and a GEV for the parameter  $H_s$ . Empirical joint non-exceedance probabilities are calculated with an extension of Eq. (7) to the trivariate case (e.g. Zhang and Singh, 2007), which reads:

$$F_{XYZ}(x_i, y_i, z_i) = P(X \leq x_i, Y \leq y_i, Z \leq z_i) = \frac{\sum_{m=1}^i \sum_{l=1}^i \sum_{p=1}^i N_{mlp} - 0.44}{N + 0.12} \quad (11)$$

The second GoF test is applied in the same way as described in Sect. 3.1, but this time for the parameter pairs  $(F, H_s)$  and  $(S, H_s)$ , which both have a similar structure of dependence and the same bivariate Copula in the FNAC model.



**Fig. 7.** Q-Q-plot for the tide gauge of Hörnum with the theoretical and empirical joint exceedance probabilities of the observed storm surge events.

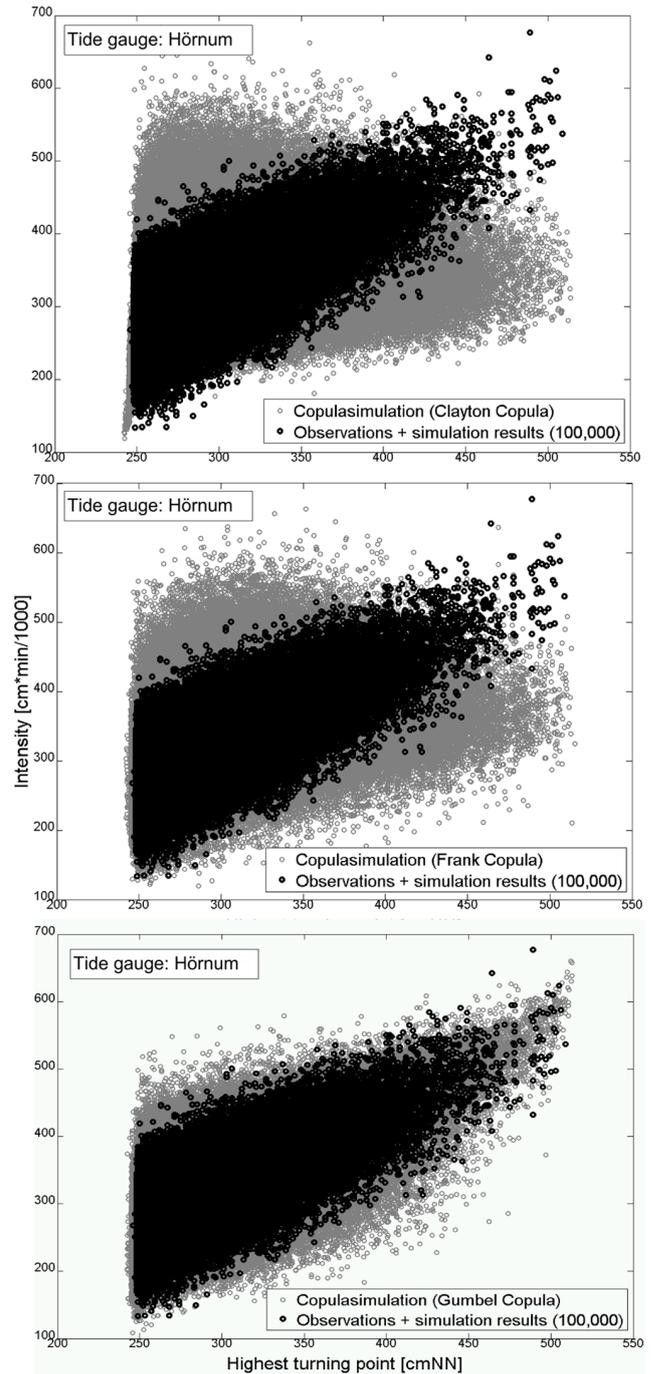
## 4 Results

### 4.1 Bivariate statistical analyses

Results from comparing theoretical and empirical joint probabilities for the parameters  $S$  and  $F$  are shown in Fig. 7 (only for Hörnum) and Table 2 (for Cuxhaven and Hörnum). The results in Fig. 7 highlight that the Gumbel Copula fits best to the data set. The same conclusion is drawn from the results presented in Table 2, where the RMSEs and the values for the KS-statistic (i.e. the maximum distance between the theoretical and empirical joint probabilities) are shown for Cuxhaven and Hörnum. For both tide gauges the Gumbel Copula leads to the smallest RMSEs and the smallest values for the KS-statistic when all events are taken into account, as well as when only extreme events (return period > 5 yr) are analysed. The Gumbel Copula is followed by the Frank Copula, while the Clayton Copula leads to the largest RMSEs and the largest values for the KS-statistic.

Results from a second and graphical based GoF test (see Sect. 3.1) are shown in Fig. 8 (only for the tide gauge of Hörnum). Again, the test results clearly point to the Gumbel Copula to be the most appropriate one to calculate the joint exceedance probabilities. It is the only Copula being able to model the existent structure of dependence between the parameters  $S$  and  $F$  with a clear tail dependency in the upper right. The same is found for the tide gauge of Cuxhaven (not shown here).

Thus, the Gumbel Copula as shown in Table 1 is chosen for both gauges to calculate the joint exceedance probabilities of the available storm surge events (from the parameters  $S$  and  $F$ ). Selected storm surges may directly be considered as scenarios for event-based risk analyses. Therefore, the so-called “AND” case is considered and it is assumed that both



**Fig. 8.** Results from a graphical based GoF test for the tide gauge of Hörnum and for the Clayton Copula (top), the Frank Copula (middle) and the Gumbel Copula (bottom).

of the parameters  $S$  and  $F$  exceed a given value. The joint exceedance probabilities are given by:

$$\begin{aligned}
 P(X > x \wedge Y > y) &= 1 - F_X(x) - F_Y(y) + F_{XY}(x, y) \\
 &= 1 - F_X(x) - F_Y(y) + C[F_X(x), F_Y(y)]
 \end{aligned}
 \tag{12}$$

**Table 2.** RMSEs and values for the KS-statistic from comparing theoretical and empirical joint probabilities (Cuxhaven | Hörnum).

Function	RMSE	RMSE (extreme events*)	KS-statistic	KS-statistic (extreme events*)
Clayton Copula	0.036   0.043	0.033   0.032	0.071   0.087	0.068   0.056
Frank Copula	0.023   0.030	0.026   0.026	0.062   0.068	0.051   0.045
Gumbel Copula	0.018   0.029	0.011   0.012	0.060   0.065	0.028   0.023

\* Events with an empirical non-exceedance probability of  $P_u > 0.8$  (equals a return period of 5 yr).

For other studies, for example when designing reservoirs or performing safety checks, the “OR” case might also be interesting. It assumes that only one of the parameters exceeds a given threshold, while the other parameter does not (see e.g. Klein et al., 2008). The joint exceedance probabilities are calculated as follows:

$$P(X > x \vee Y > y) = 1 - F_{XY}(x, y) = 1 - C[F_X(x), F_Y(y)] \quad (13)$$

The overall results from statistically analysing the available storm surge events (observations + results from stochastic storm surge simulation) are presented in Fig. 9 for Cuxhaven (top) and Hörnum (bottom). For both gauges the marginal distributions (i.e. KDFs) are shown in the figure as well the contours of some relevant joint exceedance probabilities calculated with the Gumbel Copula. With the results it is easily possible to estimate the exceedance probability of any given storm surge event (either from stochastic simulation, empirical studies or numerical model runs) by taking into account the two important storm surge parameters “highest turning point” and “intensity”.

Furthermore, it is possible to extract a specified number of storm surge events with the same exceedance probability but different values for  $S$  and  $F$ . Hence, the selected events, although having the same exceedance probability, may have significantly different characteristics and implications in terms of the flood risk. Figure 9 (right) shows selected storm surge events for both of the considered tide gauges. All of these storm surges have a joint exceedance probability of  $P_e \approx 0.001$  [ $1/a$ ] (i.e. a return period of 1000 yr) and different values for  $S$  and  $F$ . For the tide gauge of Hörnum for instance (Fig. 9, bottom), the storm surge event shown in the middle panel on the right (red curve) has a “highest turning point” of  $S = 468$  cmNN and an “intensity” of  $F = 529$  [cm\*min/1000]. The exceedance probabilities for the marginal parameters are found to be  $P_{e,S} = 0.0016$  [ $1/a$ ] (equals a return period of 625 yr) and  $P_{e,F} = 0.0025$  [ $1/a$ ] (equals a return period of 400 yr). If one would mistakenly assume independency between the parameters  $S$  and  $F$ , the joint exceedance probability would be calculated by simply multiplying the exceedance probabilities of the margins and be  $P_{e,\text{independency}} = 0.000004$  [ $1/a$ ] (equals a return period of 250 000 yr). Using this exceedance

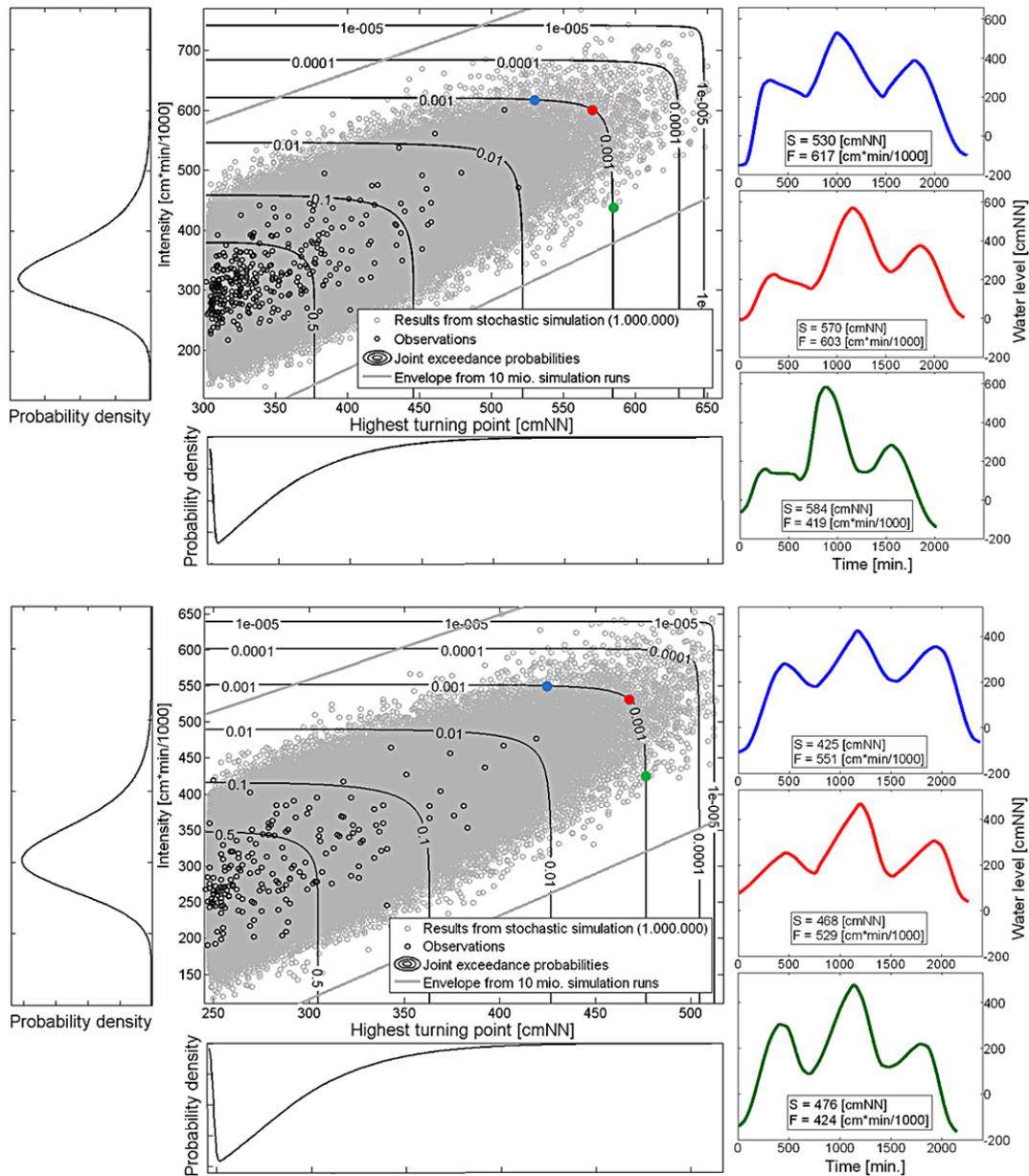
probability for a risk analysis would result in a significant underestimation of the flood risk. By assuming perfect dependency, the exceedance probability would be  $P_{e,\text{dependency}} = 0.0025$  [ $1/a$ ], resulting in an overestimation of the flood risk. Only analysing the parameter  $S$  within a risk assessment also leads to an overestimation of the flood risk.

For both tide gauges the “highest turning points” of the three selected storm surge events vary by more than 50 cm (Fig. 9, right). The events in the upper panels (blue curves) have small values for  $S$ , but high values for  $F$  and long durations. In contrast, the storm surge curves in the lower panels (green curves) have large values for  $S$  and small values for  $F$ . This is because the surrounding high water levels (i.e. the first and the third tidal high waters) are comparable low. For some investigation areas (e.g. areas protected by dunes) storm surges with characteristics as shown in the upper panels (i.e. small values for  $S$ , high values for  $F$ ) might have serious implications for the coastal defence line (e.g. risk of erosion) or for the hinterland. For other investigation areas, storm surges with characteristics as shown in the lower panels (i.e. high values for  $S$ , small values for  $F$ ) might have more devastating consequences.

Analysing various storm surge events with a specified exceedance probability but different characteristics might be useful when performing flood risk analyses for investigation areas where special protection is required (e.g. airports). In Germany for instance, dikes in front of nuclear power plants located in tidal influenced areas had to be designed to withstand a storm surge event with a return period of 10 000 yr (KTA 2207, 2004). A design storm surge, causing the highest risk of flooding, can be reliably estimated by analysing a larger number of 10 000-yr storm surge events with different characteristics.

## 4.2 Trivariate statistical analyses

Trivariate statistical analyses are performed for stochastically simulated storm surge events (transferred from Hörnum to Westerland and represented by  $S$  and  $F$ ) and the important wave parameter  $H_s$ . Table 3 shows the results from comparing trivariate theoretical joint probabilities (calculated with a FNAC model as described in Sect. 3.3) and trivariate empirical joint probabilities. The storm surge parameters  $S$  and  $F$  are combined with a Gumbel Copula as described in Sect. 3.1



**Fig. 9.** Results from statistically analysing the observed and stochastically simulated storm surge events for the tide gauges of Cuxhaven (top) and Hörnum (bottom) based on KDFs (as marginal distributions) and the Gumbel Copula and selected simulated storm surge events with a occurrence probability of  $P_e = 0.001 [1/a]$  (right).

and it is tested which Copula is most appropriate to incorporate  $H_s$  into the FNAC model. The results in Table 3 show that the Gumbel Copula leads to the smallest RMSEs and the smallest values for the KS-statistic, closely followed by the Frank Copula. The Clayton Copula leads to the highest values. Hence, the GoF test suggests the FNAC model to jointly analyse the parameters  $S$ ,  $F$  and  $H_s$  to consist of two Gumbel Copulas.

However, conclusions drawn from the second GoF test are different. Figure 10 shows that merely the Frank Copula is able to model the existent structures of dependence between

the parameter pairs  $(F, H_s)$  (left) and  $(S, H_s)$  (right). Simulation results from the Clayton Copula show strong tail dependence in the lower left, which is not present in the observations. Simulation results from the Gumbel Copula in contrast show strong tail dependence in the upper right. Such a dependence structure exists for the parameter pair  $(S, F)$  (see e.g. Fig. 1), but not for the pairs  $(F, H_s)$  and  $(S, H_s)$ . Furthermore, the lower right panel in Fig. 10 shows an outlier in the upper left corner. Hence, this graphical GoF test strongly suggests the Frank Copula could be considered as a second Copula to construct the FNAC model.

**Table 3.** RMSEs and values for the KS-statistic from comparing theoretical and empirical trivariate joint probabilities ( $S$ ,  $F$  and  $H_s$ ) for Westerland.

Function	RMSE	RMSE (extreme events*)	KS-statistic	KS-statistic (extreme events*)
Clayton Copula	0.059	0.047	0.147	0.083
Frank Copula	0.039	0.036	0.116	0.061
Gumbel Copula	0.033	0.020	0.106	0.031

\* Events with an empirical non-exceedance probability of  $P_u > 0.8 [1/a]$  (equals a return period of 5 yr).

As the Frank Copula leads to good results with both GoF tests, the FNAC model is constructed based on a Gumbel and a Frank Copula. First, the parameters  $S$  and  $F$  are analysed based on a bivariate Gumbel Copula. This Copula is subsequently considered as a marginal distribution and a Frank Copula (as shown in Table 1) is used to incorporate the parameter  $H_s$  into the model. In Fig. 11, results from bivariate and trivariate statistical analyses for the tide gauge of Westerland are compared. The parameters  $S$  and  $F$  are taken into account for the bivariate analysis (black contours in Fig. 11). For the trivariate analysis  $S$ ,  $F$  and a significant wave height of  $H_s = 400$  cm (red contours in Fig. 11) are considered. The exceedance probability for  $H_s = 400$  cm is found to be  $P_{e,H_s} = 0.37 [1/a]$ , which equals a return period of approximately 2.7 yr (see Fig. 6). Figure 11 shows that incorporating  $H_s$  into the model reduces the exceedance probabilities for specific storm surge events compared to the bivariate analysis. As an example, a storm surge event with  $S = 425$  cmNN and  $F = 468$  [cm\*min/1000] represents a 1000-yr event ( $P_e = 0.001 [1/a]$ ) for the tide gauge of Westerland (with  $P_{e,S} = 0.0013 [1/a]$  and  $P_{e,F} = 0.0035 [1/a]$ ). When taking into account a significant wave height of  $H_s = 400$  cm, the same storm surge event has an exceedance probability of  $P_e = 0.00078 [1/a]$  (equals a return period of approximately 1280 yr). Again, mistakenly assuming the independent case for all three parameters would lead to a significant underestimation of the flood risk ( $P_{e,\text{independent}} = 0.0000017 [1/a]$ ; return period approx. 590 000 yr). Assuming perfect dependency would lead to a significant overestimation ( $P_{e,\text{dependent}} = 0.37 [1/a]$ ; return period approx. 2.7 yr).

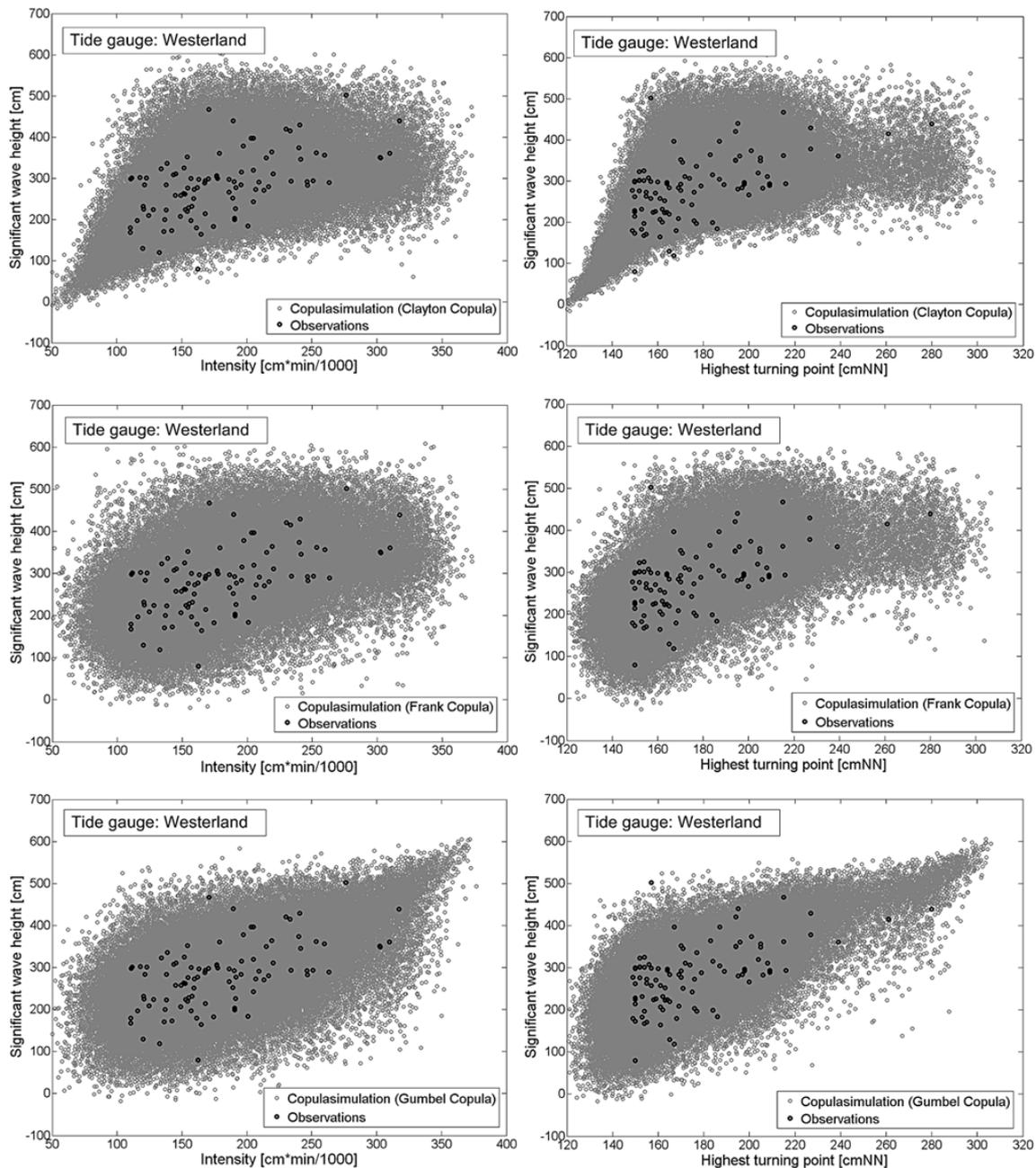
As it has been outlined in Sect. 3.3, Copulas may theoretically be extended to the  $d$ -dimensional case to include further parameters into the statistical assessment. However, this implies more constraints (if a FNAC model or a PNAC model is used), higher uncertainties and increased computational requirements. Here, the two important storm surge parameters  $S$  and  $F$  and the wave parameter  $H_s$  are considered. Alternatively the wave period  $T$  could be used instead of the wave height  $H_s$ .

### 4.3 Uncertainty assessment

Especially when analysing extreme storm surge events, the uncertainties of the statistical assessment are considerable high. In the following, the uncertainties involved in the described methodology are briefly discussed and examples of how to quantify the key uncertainties are provided.

There are two main sources of uncertainties that need to be addressed when applying the presented methodology. First, uncertainties emerge from estimating the structure of dependence or the Copula parameter  $\theta$ , respectively, from a random sample of the considered parameters. By calculating 95 %-confidence bounds of the Copula parameter, uncertainties of the exceedance probabilities for selected storm surge events may be quantified. Taking the storm surge event from the upper panel (blue curve) for Hörnum from Fig. 9 as an example, the exceedance probability is found to be  $P_e = 0.001 [1/a]$  (0.00091 | 0.00104). The numbers in brackets represent the uncertainty range (95 %-confidence level), resulting from varying the Copula parameter ( $\theta = 1.83 \pm 0.25[-]$ ). For the trivariate case, uncertainties may be quantified in the same way when a second Copula (with a second Copula parameter  $\theta$ ) is used to incorporate the wave parameter  $H_s$ .

A second key uncertainty emerges from the fact that the multivariate statistical analyses in this study are based on stochastically simulated storm surge events. As described by Wahl et al. (2011b) different filter functions are applied within the simulations, whereas one of these filters affects the statistical analyses presented here. This filter function removes very extreme events (with water levels exceeding a given threshold) from the simulation results. The water levels which have been chosen are considered currently physically possible for the investigation areas. Thresholds of 651 cmNN (as a result from a large number numerical model runs) and 513 cmNN (as a result from extensive empirical studies) are used for Cuxhaven and Hörnum, respectively. Choosing different thresholds would affect the results of the statistical analyses presented here. The uncertainties for the threshold values are 603 cmNN to 673 cmNN for Cuxhaven and 444 cmNN to 537 cmNN for Hörnum (see Wahl et al., 2011b and literature referenced in there). By choosing the upper and lower values as thresholds for the stochastic simulation, the related uncertainties in the statistical analyses may be

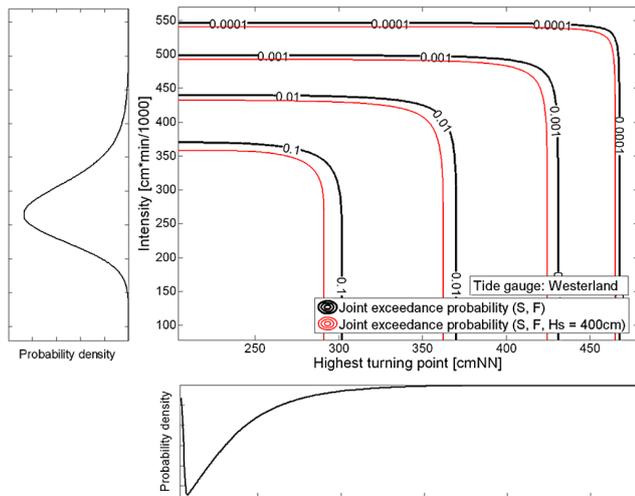


**Fig. 10.** Results from a graphical based GoF test for the tide gauge of Westerland by considering the Clayton Copula (top), the Frank Copula (middle), the Gumbel Copula (bottom) and the parameter pairs ( $F, H_S$ ) (left) and ( $S, H_S$ ) (right).

quantified. Taking the storm surge event in the upper panel (blue curve) for Hörnum from Fig. 9 as an example again, the exceedance probability is found to be  $P_e = 0.001 [1/a]$  (0.00034 | 0.0011). The numbers in brackets represent uncertainties resulting from varying the thresholds within the stochastic simulations. The uncertainty range is much larger compared to the uncertainty range resulting from varying the Copula parameter. Both of the described key uncertainties may affect the results of the statistical assessment at the

same time. Thus, they have to be superimposed to capture the whole range and the exceedance probability (with uncertainties) for the example storm surge from Fig. 9 becomes  $P_e = 0.001 [1/a]$  (0.00025 | 0.00114).

Further uncertainties result from fitting univariate distribution functions to the marginal parameters and from choosing certain bivariate or trivariate models. These uncertainties are not quantified and only briefly discussed in the following. For the parameters  $S$  and  $F$  Kernel Density Functions



**Fig. 11.** Comparison of the results from bivariate statistical analyses for the parameters  $S$  and  $F$  and trivariate analyses for the parameters  $S$ ,  $F$  and  $H_s$  (with  $H_s = 400$  cm).

are considered as marginal distributions, assuring a good fit to the available data sets. Uncertainties may arise for very extreme events from instabilities due to the limited number of events. However, when considering 10 million scenarios, storm surges with exceedance probabilities  $P_e > 5.0 \times 10^{-6}$  [ $1/a$ ] are not affected. A parametric univariate distribution function (namely the GEV) is used here for the marginal parameter  $H_s$ . The uncertainties from fitting the distribution to the data set are expected to be small compared to the key uncertainties discussed above. Furthermore, moderate wave conditions (in terms of the return periods) are usually taken into account when defining scenarios (storm surges + wave conditions) for integrated risk analyses. Here, a significant wave height of 400 cm and with a return period of approx. 2.7 yr is used as an example. Finally, fitting bivariate and trivariate models to the available data sets involves uncertainties. Here, two GoF tests are applied to choose proper Copula functions and to minimise the uncertainties.

## 5 Conclusions

In this paper a Copula based approach to statistically analyse storm surges and wind waves is presented. A bivariate statistical model is applied to jointly analyse the important storm surge parameters “intensity” and “highest turning point”. Especially when performing integrated risk analyses, where breach models are applied to identify the initial conditions for flooding of the hinterland and potential losses are quantified, the “intensity” of a storm surge (as a proxy for the energy input into the defence structures) has significant implications. Many of the bivariate models available from literature suffer from restrictions and constraints regarding the underlying data sets in terms of the dependence or the

marginal distributions. Copulas in contrast are very flexible and can handle dependent parameters with mixed marginal distributions. Moreover, it is possible to construct higher dimensional Copula models to incorporate further important parameters. In the present study, the bivariate model is extended to the trivariate case to additionally take into account the significant wave height  $H_s$  as one of the most important wave parameters. A fully nested Archimedean Copula approach is applied to construct a trivariate model. Alternatively other wave parameters, such as the wave period, may be analysed instead of (or theoretically also in addition to) the significant wave height. To outline the transferability of the model, results from the bivariate analyses are presented for two different investigation areas in the German Bight (i.e. Cuxhaven and Hörnum). The trivariate analyses focus on the investigation area Westerland on the west side of Sylt Island, for which wind waves represent a considerable loading factor. The results presented in Sect. 4 highlight that Copula functions are valuable for statistical assessments (especially within event-based risk analyses), as they are flexible and can analyse a larger number of important parameters. This leads to realistic exceedance probabilities and contributes to improving the overall results from integrated flood risk analyses, as for example performed within the German XtremRisK project for the city of Hamburg and Sylt Island.

Future work to improve the multivariate model(s) may focus on the incorporation of further GoF tests, especially when higher dimensional Copula models are applied. This assures an objective and reliable model selection. Although Copulas provide valuable features compared to traditional multivariate approaches, the adoption of certain models to certain data sets still includes considerable uncertainties. If more than two parameters are jointly analysed, further higher dimensional models might be taken into account as an alternative to the FNAC model used for the present study. Some of the models discussed in Sect. 3.3 (and further approaches described in literature) are more flexible and do not require the marginal parameters to fulfil special criteria as it is the case with the FNAC model.

However, the presented results highlight that Copulas represent a promising alternative to address various multivariate problems. While they are nowadays widely used in hydrology (and other research fields), it is hoped that some of the above remarks and examples will contribute to establish Copulas also in the coastal research and engineering sector.

*Acknowledgements.* We thank our project partners from the XtremRisK project (i.e. colleagues from the Universities of Braunschweig and Hamburg-Harburg and from the Agency of Roads, Bridges and Waters in Hamburg) for a pleasant and trustful cooperation and for providing very helpful comments on our work. The data sets used for the present study were provided by the Shipping Administration of the Federal Government (WSA) and by the Government-Owned Company for Coastal Protection, National Parks and Ocean Protection (LKN). The XtremRisK project is funded by the German Federal Ministry of Education and Research

(BMBF) through the project management of Projektträger Jülich (PTJ) under the grant number 03F043B (subproject 1b). Two anonymous reviewers and the journal editor provided valuable comments and helped us to improve the quality of this paper.

Edited by: S. Tinti

Reviewed by: two anonymous referees

## References

- Berg, D. and Aas, K.: Models for construction of multivariate dependence: A comparison study, *Europ. J. Finance*, 15, 639–659, 2007.
- Berg, D. and Quessy, J.-F.: Local sensitivity analyses of goodness-of-fit tests for copulas, *Scand. J. Statist.*, 36, 389–412, 2009.
- Berg, G.: Copula goodness-of-fit testing: an overview and power comparison. *Europ. J. Finance*, 15, 675–701, 2009.
- Cai, Y., Gouldby, B. P., Hawkes, P. J., and Dunning, P.: Statistical simulation of flood variables: incorporating short-term sequencing, *J. Flood Risk Manage.*, 1, 3–12, 2008.
- Coles, S. G. and Tawn, J. A.: Statistical methods for multivariate extremes: An application to structural design, *J. Appl. Statist.*, 43 1–48, 1994.
- De Michele, C. and Salvadori, G.: A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas, *J. Geophys. Res.*, 108, 4067, doi:10.1029/2002JD002534, 2003.
- De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R.: A multivariate model of sea storms using copulas, *Coast. Eng.*, 54, 734–751, 2007.
- de Vries, C. G. and Zhou, C.: Discussion of “Copulas: Tales and facts”, by Thomas Mikosch, *Extremes*, 9, 23–25, 2006.
- De Waal, D. J. and van Gelder, P. H. A. J. M.: Modelling of extreme wave heights and periods through copulas, *Extremes*, 8, 345–356, 2005.
- Dixon, M. J. and Tawn, J. A.: Extreme sea-levels at the UK A-class sites: site-by-site analysis, POL Internal Document Number 65, 1994.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémondge, N., and Bobée, B.: Multivariate hydrological frequency analysis using copulas, *Water Resour. Res.*, 40, W01101, doi:10.1029/2003WR002456, 2004.
- Galiatsatou, P. and Prinos, P.: Bivariate models for extremes of significant wave height and period – An application to the Dutch Coast., Proc. 2nd IMA Inter-national Conference on Flood Risk Assessment, University of Plymouth, UK, 2007.
- Genest, C. and Favre, A.-C.: Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12, 347–368, 2007.
- Genest, M. and Remillard, B.: Discussion of “Copulas: Tales and Facts”, by Thomas Mikosch, *Extremes*, 9, 27–36, 2006.
- Genest, C., Remillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, *Insurance Math. Econom.*, 44, 199–213, 2009.
- Grimaldi, S. and Serinaldi, F.: Design hyetographs analysis with 3-copula function, *Hydrol. Sci. J.*, 51, 223–238, 2006a.
- Grimaldi, S. and Serinaldi, F.: Asymmetric copula in multivariate flood frequency analysis, *Adv. Water Resour.*, 29, 1115–1167, 2006b.
- Gringorten, I. I.: A plotting rule for extreme probability paper, *J. Geophys. Res.*, 68, 813–814, 1963.
- Haigh, I. D., Nicholls, R. J., and Wells, N. C.: A comparison of the main methods for estimating probabilities of extreme still water levels, *Coast. Eng.*, 57, 838–849, 2010.
- Hanson, H. and Larson, M.: Implications of extreme waves and water levels in the southern Baltic Sea, *J. Hydraul. Res.*, 46, Extra Issue 2 (2008), 292–302, 2008.
- Hawkes, P. J.: Joint probability analysis for estimation of extremes, *J. Hydraul. Res.*, 46, Extra Issue 2 (2008), 246–256, 2008.
- Hawkes, P. J., Gouldby, B. P., Tawn, J. A., and Owen, M. W.: The Joint Probability of Waves and Water Levels in Coastal Defence Design, *J. Hydraul. Res.*, 40, 241–251, 2002.
- HR Wallingford: Joint probability of waves and water levels on the North Wales coast, HR Wallingford Report EX 2331, 1990.
- HR Wallingford and Lancaster University: The Joint Probability of Waves and Water Levels: JOIN-SEA: A Rigorous but Practical New Approach, HR Wallingford Report SR 537, Originally dated November 1998, re-issued with minor amendments in final form in May 2000, HR Wallingford, UK, 2000.
- Jensen, J., Mudersbach, Ch., Bork, I., Müller-Navarra, S. H., Koziar, Ch., and Renner, V.: Modellgestützte Untersuchungen zu Sturmfluten mit sehr geringen Eintrittswahrscheinlichkeiten an der Deutschen Nordseeküste, Die Küste, Heft 71, Boyens Medienverlag, Heide i. Holstein, 2006.
- Karmakar, S. and Simonovic, S. P.: Bivariate flood frequency analysis: Part 1 – Determination of marginals by parametric and non-parametric techniques, *J. Flood Risk Manage.*, 1, 190–200, 2008.
- Karmakar, S. and Simonovic, S. P.: Bivariate flood frequency analysis: Part 2 – A copula-based approach with mixed marginal distributions, *J. Flood Risk Manage.*, 2, 32–44, 2009.
- Klein, B., Pahlow, M., Hundecha, Y., Gattke, C., and Schumann, A.: Probabilistic Analysis of Hydrological Loads to Optimize the Design of Flood Control Systems, 4th International Symposium on Flood Defence: Managing Flood Risk, Reliability and Vulnerability, Toronto, Canada, 2008.
- Kotz, S. and Nadarajah, S.: *Extreme Value Distribution – Theory and Applications*, Imperial College Press, Imperial College Press, London, 2000.
- KTA 2207: Schutz von Kernkraftwerken gegen Hochwasser, Fassung 11/2004, 2004.
- McMillan, A., Batstone, C., Worth, D., Tawn, J., Horsburgh, K., and Lawless, M.: Coastal flood boundary conditions for UK mainland and islands, Report of the Environmental Agency, Project: SC060064/TR2: Design sea levels, Environmental Agency, 2011.
- Mikosch, T.: Copulas: Tales and facts, *Extremes*, 9, 3–20, doi:10.1007/s10687-006-0015-x, 2006.
- Mudersbach, C. and Jensen, J.: Non-Stationarities in Time Series and its Integration in Extreme Value Statistics for Risk Management Issues, Proc. of the 31st International Conference on Coastal Engineering, World Scientific, Singapore, 2, 4109–4120, 2008.
- Mudersbach, C. and Jensen, J.: Nonstationary extreme value analysis of annual maximum water levels for designing coastal structures on the German North Sea coastline, *J. Flood Risk Manage.*, 3, 52–62, doi:10.1111/j.1753-318X.2009.01054.x, 2010.
- Mudersbach, C., Wahl, T., Haigh, I. D., and Jensen, J.: Trends in extreme high sea levels along the German North Sea coastline

- compared to regional mean sea level changes, *Cont. Shelf Res.*, submitted (June 2011), 2012.
- Nelsen, R. B.: An introduction to copulas, *Lecture Notes in Statistics*, 139, Springer, New York, 1999.
- Oumeraci, H.: Sustainable coastal flood defences: scientific and modelling challenges towards an integrated risk-based design concept, *Proc. First IMA International Conference on Flood Risk Assessment*, IMA – Institute of Mathematics and its Applications, Session 1, Bath, UK, 9–24, 2004.
- Pinya, M. A. S., Madsen, H., and Rosbjerg, D.: Assessment of the risk of inland flooding in a tidal sluice regulated catchment using multivariate statistical techniques, *Phys. Chem. Earth*, 34, 662–669, doi:10.1016/j.pce.2009.03.001, 2009.
- Pugh, D. J. and Vassie, J. M.: Extreme sea-levels from tide and surge probability, *Proceedings of the Sixteenth Coastal Engineering Conference*, Hamburg, Germany, 27 Aug–3 Sep 1978, American Society of Civil Engineers, New York, 1, 911–930, 1979.
- Rao, A. R. und Hamed, K. H.: *Flood frequency analysis*, CRC Press, New York, 2000.
- Ruocco, A. C., Nicholls, R. J., Haigh, I. D., and Wadey, M.: Reconstructing coastal flood occurrence combining sea level and media sources: a case study of the Solent, UK since 1935, *Nat. Hazards*, 59, 1773–1796, doi:10.1007/s11069-011-9868-7, 2011.
- Salecker, D., Gruhn, A., Schlamkow, C., and Fröhle, P.: Parameterization of storm surges as a basis for assessment of risks of failure for coastal protection measures, *Proc. of the 5th International short conference on applied coastal research (SCACR)*, Aachen, Germany, accepted.
- Salvadori, G. and De Michele, C.: Frequency analysis via copulas: theoretical aspects and applications to hydrological events, *Water Resour. Res.*, 40, W12511, doi:10.1029/2004wr003133, 2004.
- Salvadori, G., De Michele, C., Kottegoda, N. T., and Rosso, R.: *Extremes in nature: An approach using copulas*, Springer, 2007.
- Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research – introduction to the copula approach, *Nonlin. Processes Geophys.*, 15, 761–772, doi:10.5194/npg-15-761-2008, 2008.
- Segers, J.: Discussion of “Copulas: Tales and facts”, by Thomas Mikosch, Efficient estimation of copula parameters, *Extremes*, 9, 51–53, 2006.
- Serinaldi, F. and Grimaldi, S.: Fully nested 3-copula: procedure and application on hydrological data, *J. Hydrol. Eng.*, 12, 420–430, 2007.
- Sklar, A.: Fonction de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique Université de Paris*, 8, 229–231, 1959.
- Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E.: Frequency Analysis of Extreme Events, Chapter 18, *Handbook of Hydrology*, edited by: Maidment, D., McGraw-Hill, Inc., New York, 1993.
- Sto. Domingo, N. D., Paludan, B., Madsen, H., Hansen, F., and Mark, O.: Climate change and storm surges: Assessing impacts on your coastal city through MIKE flood modeling, *International MIKE by DHI Conference “Modelling in a world of change”*, Copenhagen, 2010.
- Tawn, J.: Bivariate Extreme Value Theory: Models and Estimation, *Biometrika*, 75, 397–415, 1988.
- Tawn, J. A.: Estimating probabilities of extreme sea-levels, *Appl. Stat.*, 41, 77–93, 1992.
- Tawn, J. A. and Vassie, J. M.: Extreme sea-levels: the joint probabilities method revisited and revised, *Proceedings of the Institute of Civil Engineering*, Part 2, 249–442, 1989.
- Wahl, T., Jensen, J., and Frank, T.: On analysing sea level rise in the German Bight since 1844, *Nat. Hazards Earth Syst. Sci.*, 10, 171–179, doi:10.5194/nhess-10-171-2010, 2010a.
- Wahl, T., Jensen, J., and Mudersbach, C.: A multivariate statistical model for advanced storm surge analyses in the North Sea, *Proceedings of the 32nd International Conference on Coastal Engineering*, Shanghai, China, 2010b.
- Wahl, T., Jensen, J., Frank, T., and Haigh, I. D.: Improved estimates of mean sea level changes in the German Bight over the last 166 years, *Ocean Dynam.*, 61, 701–715, doi:10.1007/s10236-011-0383-x, 2011a.
- Wahl, T., Mudersbach, C., and Jensen, J.: Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a stochastic storm surge model, *Nat. Hazards Earth Syst. Sci.*, 11, 2925–2939, doi:10.5194/nhess-11-2925-2011, 2011b.
- Wong, G., Lambert, M. F., Leonard, M., and Metcalfe, A. V.: Drought Analysis Using Trivariate Copulas Conditional on Climatic States, *J. Hydrol. Eng.*, 15, 129–141, 2010.
- Yue, S., Ouarda, T. B. M. J., Bobee, B., Legendre, P., and Bruneau, P.: The Gumbel mixed model for flood frequency analysis, *J. Hydrol.*, 226, 88–100, 1999.
- Zhang, L. and Singh, V. P.: Trivariate flood frequency analysis using the Gumbel-Hougaard copula, *J. Hydrol. Eng.*, 12, 431–439, 2007.