Joint probability analysis of extreme wave heights and storm surges in the Aegean Sea in a changing climate

Panagiota Galiatsatou[1,a] and Panayotis Prinos[2]

[1]*Researcher, Department of Civil Engineering, A.U.Th, 54124, Thessaloniki, Greece*

[2]*Professor, Department of Civil Engineering, A.U.Th, 54124, Thessaloniki, Greece*

Abstract. Joint probability analysis is most often conducted within a stationary framework. In the present study a nonstationary bivariate approach is used to investigate the changes in the joint probabilities of extreme wave heights and corresponding storm surges with time. The dependence structure of the studied variables is modelled using copulas. The nonstationary Generalized Extreme Value (GEV) distribution is utilized to model the marginal distribution functions of the variables, within a 40-year moving time window. All parameters of the GEV are tested for statistically significant linear and polynomial trends over time. Then different copula functions are fitted to model the dependence structure of the data. The nonstationarity of the dependence structure of the studied variables is also investigated. The methods and techniques of the present work are implemented to wave height annual maxima and corresponding storm surges at two selected areas of the Aegean Sea. The analysis reveals the existence of trends in the joint exceedance probabilities of the variables, in the most likely events selected for each time interval, as well as in a defined hazard series, such as the water level at the coastline.

1 Introduction

Extreme marine events can give rise to serious flooding and can have severe impacts on the human society, as well as on the environment. The general inception of a changing climate, with extreme meteorological events of higher frequency and intensity increases the exposure of the human society and the environment to severe damages. Therefore, the analysis of extreme marine events under present and future climate conditions is of great significance.

The study of the climate change effects on mean sea level, storm surge and waves became a subject of systematic research in the recent past. Although the majority of the studies focused on mean sea level variability and trends [1, 2] storm surges and waves and more specifically their extreme values in a changing climate were also considered. Significant fluctuations in the frequency and the intensity of storms, as well as in the wave climate [3, 4] were observed in the recent past in the North Sea, without however identifying significant general trends. Studies conducted in larger areas, e.g. in the Northern Atlantic, proved certain changes in the wind fields, in storm surge levels, as well as in the wave climate [5, 6]. Woth et al. [7] studied the effects of climate change on storm surge extreme values in the North Sea, observing a statistically significant increase at the end of the 21st century. De Winter et al. [8] examined the effects of climate change on extreme waves in front of the Dutch coast, identifying no significant changes

between the return values at the end of the 21st and those at the end of the 20th century. Weisse et al. [9] reviewed the knowledge about long-term changes in sea level components and pointed out that most future projections in the North Sea area identify a moderate increase in storm activity with changes in storm surge and wave climate. Evidence of climate change effects on the marine climate has also been observed in the Mediterranean area [10, 11, 12]. Gaertner et al. [13] detected for the first time the danger of a tropical cyclone above the Mediterranean accounting for future climate change, using different high-resolution Regional Climate Models (RCMs). Martucci et al. [14] studied wave height extremes in the Italian Seas, identifying decadal negative trends during the second half of the 20th century. Galiatsatou and Prinos [15, 16] studied the effects of climate change on wave height and storm surge extremes in selected areas of the Aegean and the Ionian Seas, identifying a significant increase in extreme wave and storm surge climate in the North Aegean and Ionian Seas during the first part of the 21st century.

Recent studies on extreme value analysis for variables associated with the marine and coastal environment have been published by different researchers. Sánchez-Arcilla et al. [17] studied extreme wave events at the Spanish coast, as well as at the Dutch coast on the North Sea, assessing return level confidence intervals using a conventional extreme value and a Bayesian approach, indicating how the introduction of a priori knowledge in extreme value analysis helps to reduce uncertainty. Van

[a] Corresponding author: pgaliats@civil.auth.gr

Gelder and Mai [18] identified the main methods for estimating the distribution functions for wave height and storm surge extremes at the Dutch coast in the North Sea area, implementing Extreme Value Theory (EVT). Bulteau et al. [19] performed spatial extreme value analysis of significant wave height along the French coast using different extreme value techniques. Extreme value methods have been implemented for studying the statistical characteristics of storm surge, mainly in the North Sea area [20, 21]. Galiatsatou and Prinos [22] studied extreme storm surge events in selected locations of the Dutch coast, comparing the conventional maximum likelihood estimation procedure with techniques implemented within the Bayesian framework. Bardet et al. [23] presented a regional frequency analysis of extreme storm surges along the French coast, leading to more reliable estimates compared to at-site analysis.

Although extreme wave heights and water levels have been studied by numerous authors, studies on the combined impact of extreme marine variables are more limited. Galiatsatou and Prinos [24] studied the bivariate process of extreme wave heights and storm surges, using different methods of selecting concurrent observations as well as different measures of extremal dependence of the two variables involved. Morton and Bowers [25], De Haan and De Ronde [26], Ferreira and Guedes Soares [27] and Repko et al. [28] described the joint probability distribution function of long-term hydraulic conditions. Yeh et al. [29] examined the joint probabilities of high waves and water levels and compared results of design water level with estimates from the traditional empirical design approach by frequency analysis. Galiatsatou [30] compared different pairs of bivariate observations of extreme waves and surges with reference to joint exceedance probabilities, in order to find the most severe sea state caused by the two variables. Wahl et al. [31] jointly analyzed storm surge parameters, such as highest turning point and intensity with the significant wave height, by means of Archimedean Copulas, resulting in reliable exceedance probability estimates. Corbella and Stretch [32] investigated dependencies between wave height, wave period, storm duration, water level and storm inter-arrival time and used trivariate copulas to jointly analyse the variables that are significantly associated. Masina et al. [33] produced the joint probability distribution of extreme water levels and wave heights at Ravenna coast in Italy and used the direct integration method to assess the flooding probability.

Copulas were widely used in the analysis of multivariate extreme values both in hydrology and in marine studies (e.g. [34, 35, 36]). However, the majority of the studies considered stationarity of the marginal parameters and of the dependence structure of the copula. Zhang [37] investigated the use of nonstationary marginal distributions within a multivariate hydrological frequency analysis based on copulas. Corbella and Stretch [32] developed multivariate models of sea storms using copulas, considering the influence of nonstationary marginal distributions. Chebana et al. [38] investigated the inclusion of a changing dependence structure between the studied variables modeled by means of a copula function, within a general framework of hydrologic frequency analysis. Bender et al. [39] analysed the joint extremes of flood peak and flood discharge in the Rhine River, introducing a multivariate nonstationary approach based on copulas. The latter study considered nonstationarity both in the marginal distributions of the variables involved, as well as in their dependence structure.

In the present work a nonstationary multivariate approach [39] has been implemented to wave height annual maxima and corresponding sea level height data at two selected areas of the Aegean Sea. In Section 2 the GEV distribution, used to model the marginal distributions of the variables, is introduced and described. In Section 3, a short introduction to the copula theory is provided, while Section 4 deals with the technique used to select design events from the bivariate models constructed. Section 5 describes the study areas and the datasets available. Section 6 includes the main results of the nonstationary analysis, while Section 7 summarizes its main findings.

2 The GEV distribution function

The univariate Extreme Value Theory (EVT) includes models for block maxima and models for exceedances over appropriately defined thresholds (Peak Over Threshold - POT models). The former correspond to the Generalised Extreme Value (GEV) distribution function. The GEV is a three parameter distribution, including the location, μ, the scale, $\sigma > 0$, and the shape, ξ, parameters. Within a stationary context, the cumulative distribution function of the GEV for $\xi \neq 0$ is given by the following formula [40]:

$$G(x) = \exp[-\{1+\xi\frac{(x-\mu)}{\sigma}\}^{-1/\xi}], \ 1+\xi\frac{(x-\mu)}{\sigma} > 0 \qquad (1)$$

The special case with $\xi=0$ corresponds to the Gumbel distribution function. The parameters of the GEV distribution can be assessed using different estimation techniques. Among them, the maximum likelihood estimation (MLE) procedure is a common and easy to apply procedure. However, the method of L-moments (LM) introduced by Hosking [41], has been identified in the literature as more reliable and robust, even for relatively small sample sizes. The L-moments are analogous to ordinary moments and can be computed from linear combinations of probability weighted moments. In fact, they provide measures of the basic aspects of the shape of distributions or data samples, such as location, dispersion, skewness and kurtosis. For a data sample X_1, X_2, …, X_n arranged in increasing order, the sample probability weighted moments are [21]:

$$b_o = \frac{1}{n}\sum_{j=1}^{n} X_j \qquad (2)$$

$$b_r = \frac{1}{n}\sum_{j=r+1}^{n}\frac{(j-1)(j-2)....(j-r)}{(n-1)(n-2)....(n-r)}X_j \qquad (3)$$

The first three L-moments of the data sample are defined as:

$$l_1 = b_o \qquad (4)$$

$$l_2 = 2b_1 - b_o \qquad (5)$$

$$l_3 = 6b_2 - 6b_1 + b_o \qquad (6)$$

For the probability distribution of Eq. (1), the probability weighted moments of order r are defined [21]:

$$\beta_r = \int x\{G(x)\}^r \, dG(x) \quad r = 0,\ 1,\ 2,\ldots \qquad (7)$$

The respective L-moments can be defined, as for the data samples as:

$$\lambda_1 = \beta_o \qquad (8)$$

$$\lambda_2 = 2\beta_1 - \beta_o \qquad (9)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_o \qquad (10)$$

Most marine variables, especially at their extreme levels exhibit phenomena of nonstationaritty. Natural climatic variability and climate change are some of the prominent causes of such nonstationarities. Natural climatic variability is mainly associated with internal interactions between components of the climate system. Such components are among others the El Nino Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO) and the North Atlantic Oscillation (NAO), acting on different time scales. The aforementioned oscillations can have a significant impact on the magnitude of extreme marine events and on the occurrence of flood events. Climate change, with strong evidence existing nowadays on its existence and impacts, is associated in the literature with extreme events of higher intensity and frequency. Therefore, climate change can also be considered as a major cause of nonstationarity of hydro-meteorological or marine extremes. The aforementioned causes contribute significantly to the nonstationary behaviour of extreme marine events and the incorporation of such techniques in the extreme value models, for the process of extrapolation to be more reliable and unbiased. To incorporate nonstationarity in modelling the univariate extreme values, the three parameters of the GEV are assumed to vary as functions of time. Therefore the nonstationary version of the GEV becomes [40]:

$$G(x) = \exp[-\{1 + \xi(t)\frac{(x - \mu(t))}{\sigma(t)}\}^{-1/\xi(t)}], \ 1 + \xi(t)\frac{(x - \mu(t))}{\sigma(t)} > 0 \qquad (11)$$

Therefore, within a nonstationary context, the return level x_p corresponding to a return period of $1/p$, is assessed as a function of time and it represents the quantile of the distribution function of the studied variable in a given year:

$$x_p(t) = \mu(t) - \frac{\sigma(t)}{\xi(t)}[1 - \{-\log(1 - p)\}^{-\xi(t)}] \qquad (12)$$

To estimate the parameters for both studied variables, the methodology presented in [39] is implemented. A moving time window of length n is shifted by one year each time and the parameters μ, σ and ξ are estimated using the method of L-moments (Eqs. (2)-(10)) for each such time period. The parameter estimates correspond to the last year of each studied n-years period (moving window).

3 Modeling dependence using copulas

The main advantage of copulas over other multivariate distributions focuses on the fact that the dependence structure of the variables can be modeled independently from their marginal distributions. Copulas are multivariate distribution functions with uniform margins over (0, 1). Considering a two dimensional vector **U** on the unit cube, a copula can be defined as:

$$C(u_1, u_2) = \Pr(U_1 \le u_1, U_2 \le u_2) \qquad (13)$$

If F is a two dimensional distribution with marginals F_1 and F_2, there exists a two dimensional copula, such that for all x in the domain of F [39]:

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \qquad (14)$$

An one-parameter Archimedean copula is constructed through a generator φ as:

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) \qquad (15)$$

In the present work three one-parameter Archimedean copulas are utilised, namely the Clayton, the Frank and the Gumbel. The former is characterised by lower tail dependence, the middle one with no tail dependence and the Gumbel with strong upper tail dependence. For the abovementioned Archimedean copulas, the copula function $C(u_1, u_2)$ can be respectively given by [39]:

$$C_{clayton}(u_1, u_2) = (u_1^{-a} + u_2^{-a} - 1)^{-(1/a)} \qquad (16)$$

$$C_{frank}(u_1, u_2) = -\frac{1}{a}\ln[1 + \frac{(e^{-a \cdot u_1} - 1)(e^{-a \cdot u_2} - 1)}{e^{-a} - 1}] \qquad (17)$$

$$C_{gumbel}(u_1, u_2) = \exp\{-[(-\ln u_1)^a + (-\ln u_2)^a]^{1/a}\} \qquad (18)$$

The dependence parameter a ranges in (0, ∞) for the Clayton, in –ln(-∞, ∞)/{0} for the Frank and in [1, ∞) for the Gumbel copula. In the present work, the dependence parameter of the copulas has been assumed to vary with time, $a(t)$.

Apart from the one-parameter Archimedean copulas, an Elliptical copula was also implemented in the present work, namely the bivariate Student's t copula [42]:

$$C_t(u_1, u_2) = t_{v,a}(t_v^{-1}(u_1), t_v^{-1}(u_2)) \qquad (19)$$

where $t_{v,a}$, with $a \in [0, 1]$ and v the degrees of freedom, is the bivariate distribution corresponding to the univariate t Student distribution, t_v [42]:

$$t_{v,a}(x_1,x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-a^2}}(1 + \frac{s^2 + t^2 - 2ast}{v(1-a^2)})^{-\frac{(v+2)}{2}} \, dsdt \quad (20)$$

The generator for the Student's *t* is regularly varying. The degrees of freedom were considered fixed in the present work (d*f*=4). The dependence parameter of the copula can be also considered to vary in time.

To estimate the dependence parameter of the copula functions, Joe and Xu [43] proposed a two-stage procedure known as Inference Functions for Margins (IFM). The marginal parameters are first estimated:

$$\hat{\theta}_{IFM} = \arg\max_\theta \sum_{i=1}^{n} \sum_{j=1}^{p} \log f_i(X_{ij};\theta) \quad (21)$$

where f_i is the probability density function used for the marginals of the bivariate distribution. Then the dependence parameter of the copula is estimated given the marginal parameters:

$$\hat{a}_{IFM} = \arg\max_a \sum_{i=1}^{n} \log c(F_1(X_{i1};\hat{\theta}_{IFM}), F_2(X_{i2};\hat{\theta}_{IFM}); a) \quad (22)$$

When a consistent estimation of the dependence parameter *a,* is of importance, the canonical maximum likelihood method (CML) can be utilised, without first specifying the marginal distributions. In this method, the marginals are first transformed to pseudo-observations with uniform margins $(U_{i1}, U_{i2})^T$ and then the dependence parameter, *a*, is estimated as:

$$\hat{a}_{CML} = \arg\max_a \sum_{i=1}^{n} \log c(U_{i1}, U_{i2}; a) \quad (23)$$

To select an appropriate copula function among the four presented (Clayton, Frank, Gumbel and *t* copula), the parametric bootstrap procedure proposed by Genest et al. [44] has been utilised. The tests computes the Cramér-von Mises functional S_n, comparing the empirical copula of the observations with a parametric estimate of the copula derived under the null hypothesis. Approximate *p*-values for the test have been computed using the parametric bootstrap procedure. Large values of S_n result in the rejection of the null hypothesis that the bivariate data result from the tested copula function.

To estimate the dependence structure of the data, within the introduced nonstationarity framework, the moving time windows, utilized to estimate the marginal parameters of the variables involved, have also been applied. Pseudo-observations of the variables were first extracted. The dependence parameters for all selected copula functions were then extracted for each moving window using the CML procedure (Eq. (23)). After estimating the copula parameters, the statistic S_n and its associated *p*-value were estimated for all moving windows and all candidate copula functions. The copula with the highest *p*-values, that exceeded the 5% significance level for the entire interval of the study, was selected and applied for joint exceedance probability estimation.

4 Selection of design events

In the framework of multivariate statistics, the joint return period can be estimated from the joint exceedance probability of a pair of events. In the present work, the joint return period of the studied marine variables can be given by [39]:

$$T_{X_1,X_2} = \frac{1}{P(X_1 \geq x_1 \cap X_2 \geq x_2)} = \frac{1}{1 - F_1(x_1) - F_2(x_2) + C(u_1, u_2)} \quad (24)$$

The denominator of the joint return period of Eq. (24) is the joint exceedance probability, P_E.

The implementation of multivariate extreme value models results in an infinite number of combinations of the variables involved corresponding to each joint return period. Therefore, for a given joint exceedance probability, an infinite number of data couples (u_1, u_2) can be equally selected to be used in the design process. To overcome the selection problem, Salvadori et al. [45] presented the most likely design event method. This method identifies the multivariate event with the highest joint probability density, among events belonging to the same probability isoline [46]:

$$(u_1, u_2) = \arg\max_{T_{X_1,X_2}} f_{XY}(F_1^{-1}(u_1), F_2^{-1}(u_2)) \quad (25)$$

The resulting design values (x_1, x_2) can then be estimated using the inverse of the cumulative distribution functions of the marginals:

$$x_1 = F_1^{-1}(u_1) \text{ and } x_2 = F_2^{-1}(u_2) \quad (26)$$

Inundation of coastal areas is caused by the combined effect of high water levels (storm surges and astronomical tides) and wave heights. The wave-induced run-up at the coast can be calculated using the Stockdon et al. [47] formula:

$$R_{2\%} = 1.1(0.35\tan\beta(H_s L_o)^{\frac{1}{2}} + \frac{(H_s L_o(0.563\tan^2\beta + 0.004))^{\frac{1}{2}}}{2}) \quad (27)$$

where H_s is the deepwater significant wave height, L_o is the deepwater wave length associated to the wave peak period T_p and tanβ is the beachface slope. To estimate the wave-induced run-up, in the present work the wave period was modeled by means of a conditional GEV distribution function (Eq. (1)) with parameters depending on the significant wave height estimates. The parameters μ and σ of the fitted distribution were modelled using empirical regression functions [28]:

$$g(H_s)_i = a\log(H_s) + b$$
$$\text{or} \quad g(H_s)_i = aH_s^b \quad (28)$$

where $g(H_s)_i$ correspond to the location (*i*=1) and scale (*i*=2) parameters and *a*, *b* are parameters to be estimated. The shape parameter, ξ, of the GEV, which determines the tail behaviour of the distribution, has been considered constant.

5 Study areas and datasets

The methods and techniques of the present work have been implemented to wave height annual maxima and associated sea level height values at selected locations of the Aegean Sea. Two different areas have been selected (Figure 1), one in the North Aegean Sea (area 1) close to the coastal areas of Thrace and one in the South Aegean Sea (area 2), in the marine area of Heraklion in Crete.
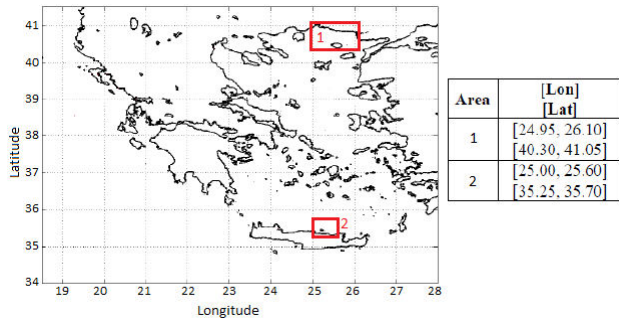

Figure 1. Selected areas of the Aegean Sea

The marine data of the present work are predictions of wave height and sea level height at selected grid points of the two study areas. The wave data resulted from a wave prediction system formulated for the Greek Seas, based on the wave model SWAN [48]. Sea level height simulations resulted from a high resolution two-dimensional model of hydrodynamic ocean circulation also formulated for the Greek seas at large [49]. The validation and calibration of the models was performed using historic data from insitu stations. The datasets cover a period of 150 years (1951- 2100) and were produced within the research project "CCSEAWAVS: Estimating the effects of climate change on sea level and wave climate of the Greek seas, coastal vulnerability and safety of coastal and marine structures". The forcing of the models consisted of simulated data of wind and pressure fields derived from a Regional Climate Model (RCM), RegCM3. RegCM3 was built upon the NCAR-Pennsylvania State University (PSU) Mesoscale Model version 4 (MM4) [50]. The spatial resolution of the model is 10x10 km and its future projections were forced by the A1B emissions scenario.

The selection of representative points of the wave height and storm surge model grid in the study areas, has been performed utilizing the homogeneity measures of Hosking and Wallis [51]. For each one of the two marine variables examined, if the homogeneity measures are lower than unity for a particular study area, then this area can be characterised as adequately homogeneous. In the present work, the homogeneity measures were assessed based on the annual and also on the monthly maxima of the wave height in the period 1951-2000, for all the studied grid points and for the two study areas. Regarding the extreme storm surge climate, the study areas were judged to be acceptably homogeneous. For wave height extremes, two distinct groups of grid points were recognized in area 1, with large differences in extreme wave height quantiles, while a single homogeneous group of points was identified in area 2. In the present work, the

joint probability analysis has been conducted for a selected grid point belonging to the homogeneous group of area 1 with the highest wave height quantiles, grid point P1 [25.30^o, 40.65^o] and for a grid point in area 2, P2 [25.15^o, 35.70^o]. Concomitant data of wave height and sea level height at the selected points were used in the joint probability analysis. For the wave height data, annual maxima corresponding to a period of 150 years were selected. Due to the quite low values of the storm surge in the Aegean Sea, sea level heights corresponding to the respective wave height annual maxima were used in the bivariate analysis.

6 Nonstationary analysis

6.1. Estimation of the margins

Annual maximum wave height data and simultaneous sea level heights have been processed at the two selected areas using a moving time window of forty years length. The length of the window was selected short enough for the assumption of stationarity to be quite sound and adequate for the fitting of extreme value models and more particularly for identifying the dependence structure of the bivariate data. The GEV distribution was fitted to all time windows for both the wave height and sea level height data. The goodness of fit of the GEV distribution function has been checked by means of the Kolmogorov-Smirnov test and the model was identified as the most suitable for both studied variables and for both areas. The selection of the GEV as the marginal distribution for both the wave height and the sea level height, has been performed among different fitted models. The extracted time dependent parameters for the wave height maxima, μ, σ and ξ, from the 40-years moving time windows for grid points P1 (N. Aegean Sea) and P2 (S. Aegean Sea) are presented in Figure 2. The ordinary least squares method has been utilised to fit linear and polynomial trends to the parameter estimates. The significance of linear trends has been assessed using the Mann-Kendall test [52]. Polynomial trends have also been fitted to the parameters of the GEV distribution function. The statistical significance of polynomial terms has been judged using the *t*-test [53]. An analysis of variance (ANOVA) was then utilized to compare between trend models with statistically significant polynomial terms, to identify the simplest one that can provide an adequate description of the inherent trend in the GEV parameters. In Figure 2 the dashed blue line corresponds to the statistically significant linear trends (5% significance level), while dashed red lines represent statistically significant polynomial trends. The order of the fitted polynomials has been selected by means of the ANOVA. For grid point P1, a statistically significant linear negative trend has been detected in the location and the shape parameter of the GEV, while the respective trend in the scale had a positive sign. However, the analysis of variance revealed the existence of polynomial trends of fourth order for the location parameter and of third order for the scale and the shape parameters. For grid point P2, statistically significant linear trends have been detected
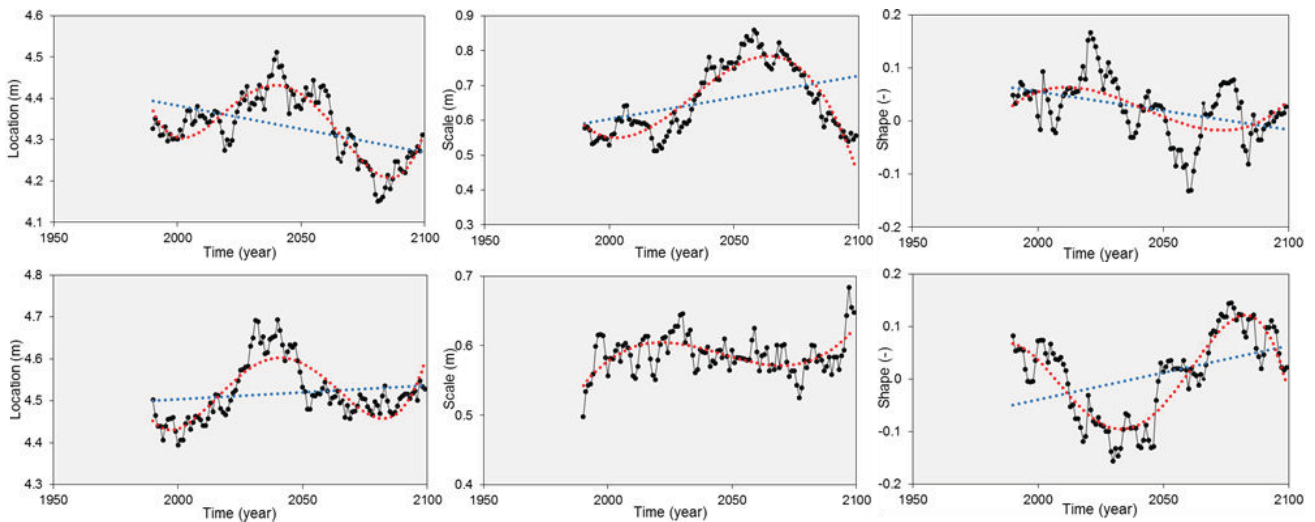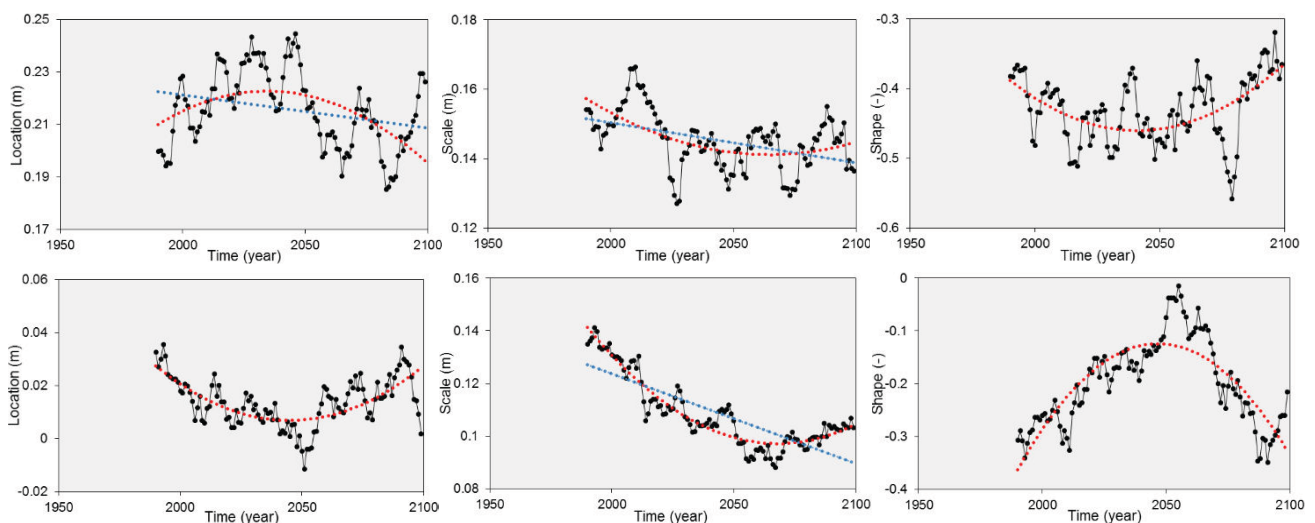
Figure 2. Time dependent series of location (first column), scale (second column) and shape (third column) parameters of the GEV for wave heights at grid point P1 (first row) and P2 (second row). Black dotted lines correspond to estimates from the 40-years moving time window, dashed blue and red lines represent statistically significant linear and polynomial trends, respectively.

only in the location and shape parameters of the GEV for wave height annual maxima. These linear trends are positive for both parameters. The polynomial models fitted to the three parameters, revealed statistically significant trends of fourth order for the location and shape parameters and of third order for the scale.

Figure 3 presents the respective time dependent GEV parameter estimates for sea level height (storm surge) for grid points P1 in the North Aegean Sea and P2 in the South Aegean Sea. The dashed blue line corresponds to the statistically significant linear trends, while dashed red lines represent statistically significant polynomial trends, based on the extracted results of the ANOVA. For grid

point P1 statistically significant linear negative trends have been detected in the location and scale parameters of the GEV, while for grid point P2 only the scale parameter has a statistically significant linear negative trend. The polynomial trends selected for the sea level height data are mostly of second order. More specifically, for grid point P1 in the N. Aegean Sea, a concave trend has been detected in the location parameter, while convex second order trends have been found for the scale and the shape parameters. For grid point P2 in the S. Aegean Sea, statistically significant concave trends have been identified for the location and scale parameters of the GEV, while convex trends have been found in the shape parameter.

Figure 3. Time dependent series of location (first column), scale (second column) and shape (third column) parameters of the GEV for sea level heights at grid point P1 (first row) and P2 (second row). Black dotted lines correspond to estimates from the 40-years moving time window, dashed blue and red lines represent statistically significant linear and polynomial trends, respectively.

6.2. Estimation of the dependence structure

After estimating the marginal distributions for both the wave height and the sea level height data, copula functions have been fitted to the pseudo-observations of the different 40-years moving time windows. The copulas fitted to the bivariate samples are the one-parameter Archimedean copulas (Clayton, Frank, Gumbel), as well as the t Copula. The goodness of fit test of Genest et al. [44] has been applied to select the best fitted copula (Section 3) among the different candidate models. Figure 4 presents results of the parametric bootstrap goodness of fit test for grid points P1. The upper part of the Figure illustrates the results of the statistic S_n for the different copulas, while the lower part presents the corresponding p-values, together with the level of statistical significance 5% (represented as a solid black line).
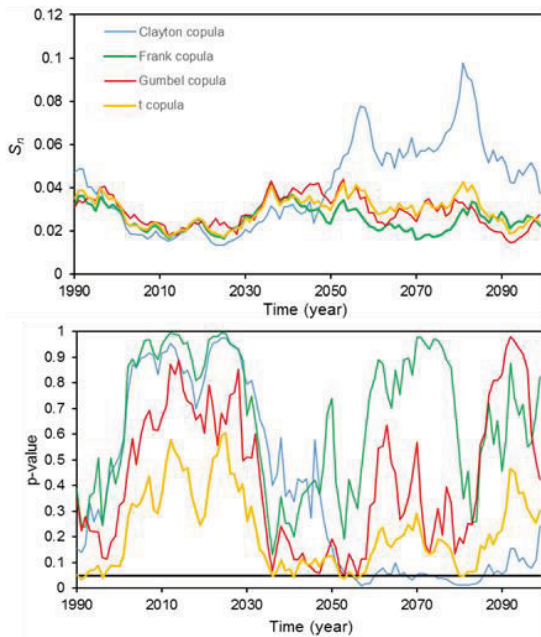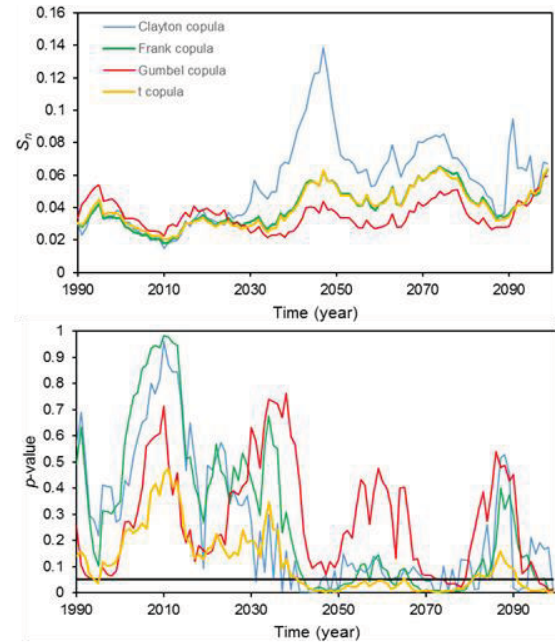

Figure 4. Parametric goodness of fit results for grid point P1. The upper panel shows results of the statistic S_n for different copulas. The lower panel shows the corresponding p-values.

For grid point P1, the Frank copula provides the lowest values of the S_n for a large part of the studied time interval. Regardless of the fact that it does not always yield the lowest values of the statistic, it leads to the best fit in the majority of cases, with p-values estimated high enough for the entire time interval. It should also be noted that p-values of the Frank copula are estimated above the 5% significance level for all time steps.

Figure 5 provides similar information to Figure 4 for grid P2. For grid point P2, the selection of the appropriate copula function is not as evident as in the case of P1. The fact is that for P2, none of the four tested models results in the lowest values for S_n for the largest part of the studied period and none of them is associated with p-values higher than the significance level 5% for all time steps. However, the Gumbel copula seems to result in the lowest S_n values for the last eighty years, while there are only very few time steps, where the p-

value for the Cramér-von Mises statistic falls below 0.05. Therefore the Frank and the Gumbel copulas have been selected for grid points P1 and P2, respectively.

Figure 5. Parametric goodness of fit results for grid point P2. The upper panel shows results of the statistic S_n for different copulas. The lower panel shows the corresponding p-values.

After selecting an appropriate copula function, the dependence structure has been fitted to the bivariate pseudo-observations of each moving 40-year time window and the dependence parameter of the copula has been calculated. Figure 6 presents the time dependent parameter of the selected copula functions for annual maximum wave heights and associated sea level heights, for grid points P1 (top) and P2 (bottom). The dashed blue and red lines correspond to statistically significant linear or polynomial trends, respectively.
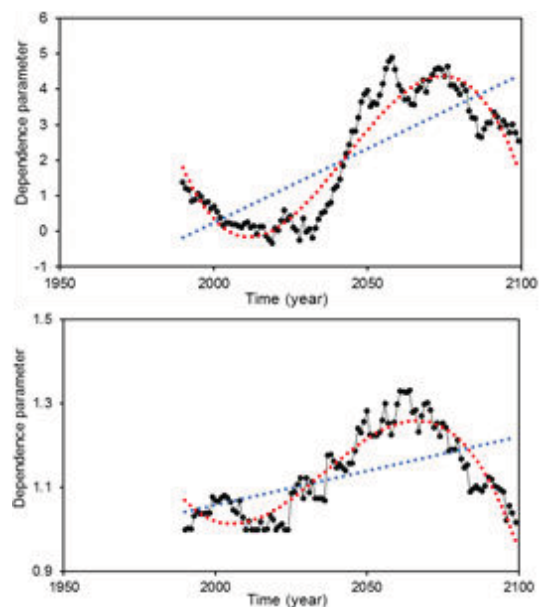

Figure 6. Dependence parameter of the Frank and Gumbel copulas for grid points P1 (top) and P2 (bottom), respectively, and fitted linear (blue) and polynomial (red) trends.

For grid point P1, a statistically significant linear positive trend has been detected in the dependence parameter of the Frank copula. A statistically significant linear positive trend has also been detected in the dependence structure of the Gumbel copula for grid point P2. However, a polynomial function of third order has been judged to describe the variation of the Frank or the Gumbel dependence parameter in a more detailed way (dashed red line), for grid points P1 and P2, respectively.

With the time dependent marginal distributions for wave height and sea level height data and the time dependent copula parameters, joint exceedance probabilities can be assessed for any joint event. Considering events with a low probability of exceedance, the temporal variation of the joint exceedance probability of wave heights and storm surges can be extracted by constructing joint exceedance probability isolines, considering the temporal variations detected in the marginal distributions of the studied variables as well as in the dependence structure. In the present work, the polynomial functions selected for the marginal distributions of wave heights and sea level heights, and for the dependence structure of the data, were used to assess the joint estimates of the studied variables

corresponding to a return period of 100 years (joint exceedance probability P_E=0.01). Figures 7 and 8 present time dependent joint exceedance probability isolines for each year in the interval 1990-2100 for grid points P1 in the N. Aegean Sea and P2 in the S. Aegean Sea, respectively. For all the studied years the joint exceedance probability is P_E=0.01. The left part of Figures 7 and 8 corresponds to the period 1990-2050, while the right to 2051-2100. The points correspond to the bivariate observations of the entire sample (1951-2100).

For grid point P1, for the period 1990-2050 (left panel of Figure 7), the probability isolines cover a wide range of values especially for the wave height data. Wave height marginal values increase within this interval from 7.2m to almost 8m, while sea level height marginal values decrease from 0.54m to almost 0.49m. For the period 2025-2050, a significant increase of extreme wave heights corresponding to a joint exceedance probability P_E=0.01 has been observed. During the years 1990-2000, a decrease has been observed in the bivariate estimates of both the wave height and the sea level height extremes corresponding to P_E=0.01. Wave height marginal estimates progressively increased shortly after year 2000,
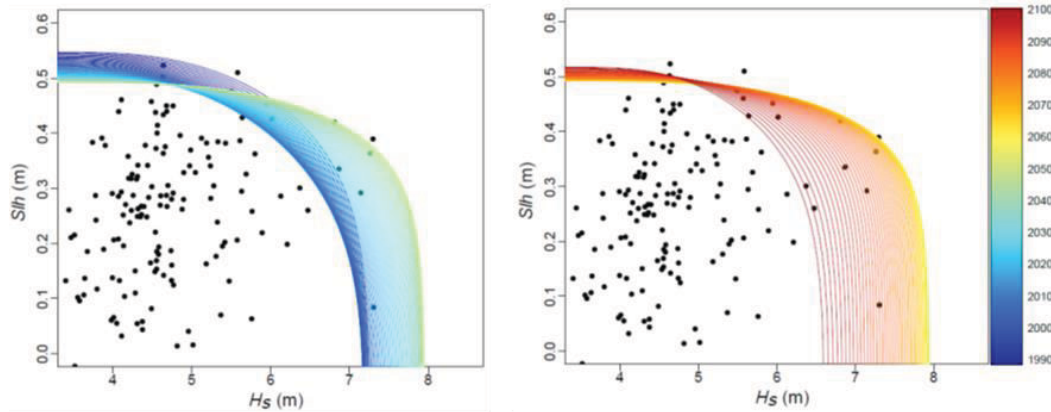
Figure 7. Time dependent joint exceedance probability isolines for P_E=0.01 for bivariate data of wave height and sea level height at grid point P1. The left panel corresponds to the period 1990-2050, and the right to 2051-2100. The colour bar refers to the last year of each moving time window.
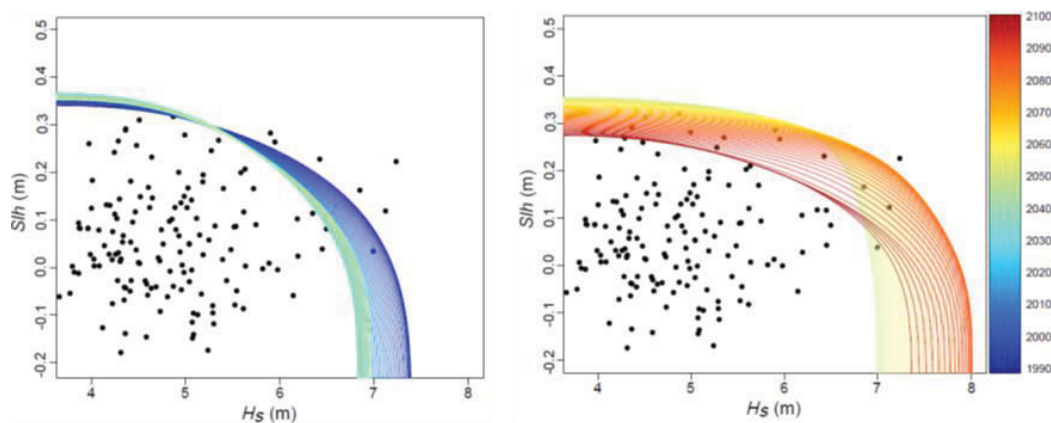
Figure 8. Time dependent joint exceedance probability isolines for P_E=0.01 for bivariate data of wave height and sea level height at grid point P2. The left panel corresponds to the period 1990-2050, and the right to 2051-2100. The colour bar refers to the last year of each moving time window.

while sea level heights continued to decrease until the middle of the 21st century. For the period 2051-2100 (right panel of Figure 7), a significant decrease in marginal wave height estimates has been noticed. Wave height marginals in the year 2100 range at almost 6.6m. During the same period, sea level height marginals slightly increase to about 0.52m. The effect of the dependence structure on the joint estimates of wave heights and sea level heights can also be noticed. The dependence structure decreases until the beginning of the century, then there is a subsequent quite steep increase up to year 2080 followed by a decrease up to 2100.

For grid point P2, the marginal estimates of the wave height also cover a wide range of values. In the period 1990-2050 (left panel of Figure 8) the wave height decreases from 7.4m to 6.8m, while sea level height slightly increases. In the period 2050-2100, the marginal wave height estimates increase up to almost 8m in the early 2080's and then decrease again to almost 7.4m. On the other hand, sea level heights decrease from 0.36m to 0.27m. The shape parameter of both the wave height maxima and the concomitant sea level height determines the aforementioned variations in a very significant way. It should be noted that for grid point P1, the scale parameter variations seemed to influence more the time dependent joint exceedance probability estimates for the two variables. For both grid points in the second half of the century the range of variation of wave height marginals is higher than the one for the first half. For sea level height, the highest variation in the marginal distribution has been noticed in the first half for grid point P1 and in the second one for grid point P2.

The most likely design event, corresponding to the event with the highest likelihood to occur is then defined for each joint exceedance probability isoline. Figure 9 presents the time dependent design estimates of both the wave height and the associated sea level height. The upper panel shows the variation of the most likely wave event for grid points P1 (blue curve) and P2 (red curve). The figure includes the most likely events extracted using the parametric trends in the marginals and in the dependence parameter (solid lines), as well as the events extracted without considering the parametric trends, but by just using the results extracted from applying the moving time windows for estimating the marginals and the dependence function of the data (dotted lines). It should be noted that the approach used here to select the so called most likely design event is not the most appropriate one. Instead, the response to the defined joint flooding source could be used, and the "design event" could be found as the point of intersection of the response and the survivor function of the source. Several combinations of the relevant variables could also be used and the system's response to the bivariate flooding source could be simulated, to finally select the bivariate data that maximise the cost-benefit ratio.

For both P1 and P2 the wave height most likely events present an almost polynomial variation. In fact they can be fitted by fourth order polynomials quite satisfactory. For grid point P1, the most likely wave event ranges between 4.40m and 5.72m, while for grid point P2 this range becomes 4.63m to 5.51m. For grid

point P1, the wave height event takes its maximum value in the second half of the 21st century (before 2070), while for grid point P2 the maximum wave height is noticed in the first half (just after 2020). For the sea level height the variations are not that significant. At both grid points, a parabolic polynomial can represent these variations quite well. For grid point P1, the sea level height ranges between 0.44m and 0.54m, while for P2 this range becomes 0.23m to 0.40m.
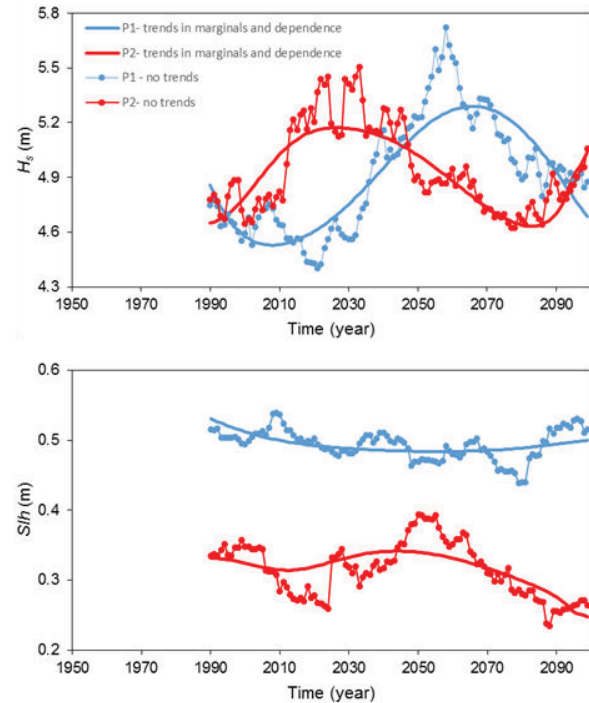
Figure 9. Time dependent development of the most likely event for wave heights (upper panel) and sea level heights (lower panel) at grid points P1 (blue line) and P2 (red line).

The total water level at the coast is approximated in the present work as a sum of the wave-induced run-up at the coast (Eq. (27)) and of the sea level height in the nearshore area. It has been assumed that the sea level height near the coast is almost equal to the one estimated in the deeper water. However, this is just an approximation and more detailed analysis is necessary to extract more reliable estimates of the storm surge at the coastal zone. The wave induced run-up has been estimated for two beach profiles, one for each selected grid point at the coastal area of Thrace (N. Aegean Sea) and Heraklion, Crete (S. Aegean Sea). The selected beach profile in the coastal area of Thrace [25.21°, 40.94°] has a beach slope of almost 4%. The beach width at the selected location is 28m, while the beach berm height is almost 1.1m. In the coastal area of Heraklion, the selected profile [25.36°, 35.34°] is characterised by a slope of 5%, a beach width of 40m and a berm height of 2m. Figure 10 presents the water level, defined as the sum of wave induced run-up and sea level height, near the above mentioned coastal areas in the interval 1990-2100. The upper panel corresponds to the estimated water level for the first selected profile in the coastal area of Thrace, while the lower panel corresponds to the second profile in the coastal area of Heraklion. To

estimate the wave induced run-up, wave periods associated with annual maxima wave heights have been extracted and fitted to a conditional GEV distribution function with parameters given by Eq. (28). Wave period quantiles corresponding to a return period of 100 years (P_E=0.01) were then calculated for the interval 1990-2100, using the estimates of the wave height most likely design events presented in Figure 9 (input wave heights correspond to the dotted lines of Figure 9).
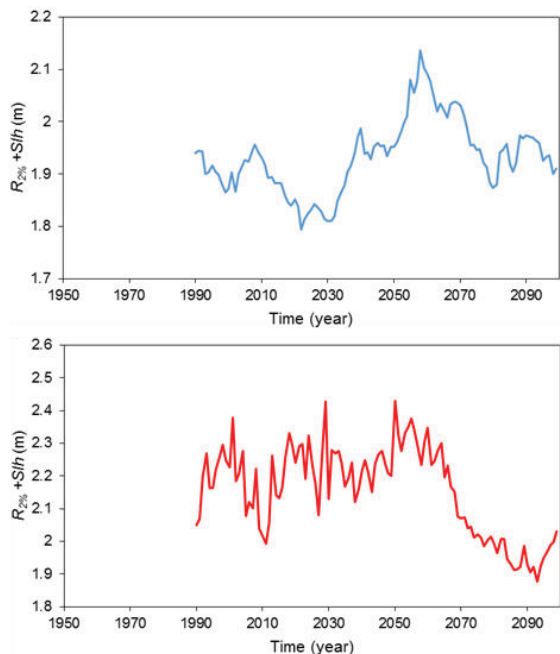
Figure 10. Time dependent development of the water level (sum of wave induced run-up and sea level height) at selected profiles in the coastal areas of Thrace (upper panel) and Heraklion, Crete (lower panel)

For the selected profile in the coastal area of Thrace, the water level in the interval 1990-2100 ranges from 1.79m to 2.14m, with the highest estimates observed in the second half of the 21st century (between 2055-2060). Water levels rise quite steeply after 2030 until 2060 and then decrease again until the early 2080's. For the selected profile in the coastal area of Heraklion, the water level ranges from 1.88m to 2.43m. The highest estimates have been noticed in the period 1990-2060. After 2060, water level estimates decrease quite steeply. In the period 1990-2060, a linear increasing trend can be observed in the estimates of the sum of wave induced run-up and sea level height. The highest water level estimates are assessed around 2030 and 2050.

7 Conclusions

In the present study a novel approach introduced by Bender et al. [39] has been utilised and further developed to investigate the changes in the joint probabilities of extreme wave heights and associated sea level heights with time. The dependence of the studied variables has been modeled using copulas. The nonstationary GEV distribution has been utilized to model the marginal distribution functions of the variables, with a 40-year

moving time window. All parameters of the GEV were tested for statistically significant trends. Then different copula functions were fitted to model the dependence structure of the data. The nonstationarity of the dependence structure of the studied variables was also investigated. Design events of wave height and sea level height were extracted and finally, water level estimates at the coast were produced for selected beach profiles in the study areas.

The nonstationary analysis of the marginals revealed statistically significant trends in all parameters of the GEV for both the wave height and the sea level height at the selected areas of the Aegean Sea. Third or fourth order polynomial trends have been detected in the GEV parameters for the wave height annual maxima, while second order polynomial trends were judged to describe best the variation of the GEV parameters of the sea level heights. Third order polynomials were also fitted to the dependence structure of the studied variables for both areas of the Aegean Sea considered.

For the studied area of the N. Aegean Sea (Thracian Sea), the joint exceedance probability isolines revealed an increase of marginal wave height estimates in the first half of the 21st century from 7.2m to 8m, and a decrease in sea level heights (almost 10%). In the second half of the century, a significant decrease in extreme wave heights has been noticed. Wave heights in the year 2100 were assessed at almost 6.6m. During the same period, sea level height marginal estimates slightly increased. It has been noticed that the scale parameter variations significantly influenced the time dependent joint exceedance probability estimates for the two variables. For the studied area of the S. Aegean Sea (marine area of Heraklion) wave height decreased from 7.4m to 6.8m in the period 1990-2050, while sea level height slightly increased. In the period 2050-2100, the marginal wave height estimates increased up to almost 8m, while sea level heights decreased from 0.36m to 0.27m. The shape parameter of both the wave height maxima and the concomitant sea level height had a significant effect on the aforementioned variations.

For both study areas, the wave height most likely events presented a fourth order polynomial variation. For the selected area in the N. Aegean Sea, the most likely wave event ranged between 4.40m and 5.72m, with the maximum value assessed in the second half of the 21st century (before 2070). For the selected area in the S. Aegean Sea, wave height most likely events varied from 4.63m to 5.51m, with the maximum quantile noticed in the first half of the century (just after 2020). For the sea level height the variations were not that significant.

Finally, water levels at the coastline were assessed for two selected profiles in the two study areas, calculating the sum of the wave induced run-up and the sea level height. For the selected profile in the coastal area of Thrace (N. Aegean Sea), the water level in the interval 1990-2100 ranged from 1.79m to 2.14m, with the highest estimates assessed in the second half of the 21st century. For the selected profile in the coastal area of Heraklion, the water level ranged between 1.88m and 2.43m. The highest estimates were noticed in the period 1990-2060.

This work summarises a first approach to the nonstationary modeling of wave height and sea level height data. The methodology presented can be further evolved by using two parameter copulas, to overcome the problem of goodness of fit of the selected dependence structure to all time windows. The selection of design events could be further examined and more advanced approaches, better oriented to flooding hazards, can be adopted. Finally, a more reliable and robust methodology can be used to assess the water level at the coastline, considering its different components and their transformations from deep water to the nearshore area.

Acknowledgements

References

1. Cabanes, C., Cazenave, A. and Le Provost, C. (2001). Sea level rise during past 40 years determined from satellite and in situ observations. *Science*, **294**, 840-842.
2. Marcos, M. and Tsimplis, M.N. (2008). Coastal sea level trends in Southern Europe. *Geophysical Journal International*, **175(1)**, 70-82.
3. WASA-Group. (1998). Changing waves and storms in the Northeast Atlantic? *Bulletin of the American Meteorological Society*, **79**, 741-760.
4. Weisse, R. von Storch, H. and Feser, F. (2005). Northeast Atlantic and North Sea storminess as simulated by a regional climate model. *Journal of Climate*, **18**, 465-479.
5. Wang, X. and Swail, V. (2006). Climate change signal and uncertainty in projections of ocean wave heights. *Climate Dynamics*, **26**, 109-126.
6. Debernard, J. and Roed, L. (2008). Future wind, wave and storm surge climate in the Northern Seas: a revisit. *Tellus*, **60**, 427-438.
7. Woth, K. Weisse, R. and Storch, H. (2006). Climate change and North Sea storm surge extremes: an ensemble study of storm surge extremes expected in a changed climate projected by four different regional climate models. *Ocean Dynamics*, **56(1)**, 3-15.
8. De Winter, R.C., Sterl, A., de Vries, J.W., Weber, S.L. and Ruessink, G. (2012). The effect of climate change on extreme waves in front of the Dutch coast. *Ocean Dynamics*, **62(8)**, 1139-1152.
9. Weisse, R., von Storch, H., Niemeyer, H.D. and Knaack, H. (2012). Changing North Sea storm surge climate: An increasing hazard? *Ocean and Coastal Management*, **68**, 58-68.
10. Méndez, F.J., Menéndez, M., Luceño, A. and Losada, I.J. (2006). Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *Journal of Geophysical Research*, **111(7)**, C07024.
11. Lionello, P., Cogo, S., Galati, M.B. and Sanna, A. (2008). The Mediterranean surface wave climate inferred from future scenario simulations. *Global and Planetary Change*, **63**, 152-162.
12. Casas-Prat, M. and Sierra, J.P. (2011). Future scenario simulations of wave climate in the NW Mediterranean Sea. *Journal of Coastal Research*, **SI 64**, 200-204.
13. Gaertner, M.A., Jacob, D., Gil, V., Dominguez, M., Padorno, E., Sanchez, E. and Castro, M. (2007). Tropical cyclones over the Mediterranean Sea in climate change simulations. *Geophysical Research Letters*, **34(14)**, L14711.
14. Martucci, G., Carniel, S., Chiggiato, J., Sclavo, M., Lionello, P. and Galati, M. B. (2010). Statistical trend analysis and extreme distribution of significant wave height from 1958 to 1999–an application to the Italian Seas. *Ocean Science*, **6(2)**, 525-538.
15. Galiatsatou, P. and Prinos, P. (2014). Analysing the effects of climate change on wave height extremes in the Greek Seas, *Proceedings of the 11th International Conference on Hydroscience & Engineering* (ICHE 2014), Hamburg - Lehfeldt & Kopmann (eds) - © 2014 Bundesanstalt für Wasserbau ISBN 978-3-939230-32-8, 773-781.
16. Galiatsatou, P. and Prinos, P. (2015). Estimating the effects of climate change on storm surge extremes in the Greek Seas, *36th IAHR World Congress, 28 June - 3 July, The Hague, The Netherlands*
17. Sánchez-Arcilla, A., Gomez-Aguar, J., Egozcue, J. J., Ortego, M. I., Galiatsatou, P. and Prinos, P. (2008). Extremes from scarce data. The role of Bayesian and scaling techniques in reducing uncertainty. *Journal of Hydraulic Research*, **46(2)**, 224-234.
18. van Gelder, P.H.A.J.M. and Mai, C. (2008). Distribution functions of extreme sea waves and river discharges. *Journal of Hydraulic Research*, **46(2)**, 280-291.
19. Bulteau, T., Lecacheux, S., Lerma, A. N. and Paris, F. (2013). Spatial extreme value analysis of significant wave heights along the French coast. *In International short conference on advances in extreme value analysis and application to natural hazards: EVAN2013.*
20. Coles, S. and Tawn, J. (2005). Bayesian modelling extreme surges on the UK east coast. *Philosophical Transactions of the Royal Society of London* (*A: Mathematical, Physical and Engineering Sciences*), **363**, 1387-1406.
21. van Gelder, P.H.A.J.M. (1999). *Risk-based design of civil structures*. PhD-Thesis, University of Technology, Delft, The Netherlands.
22. Galiatsatou, P. and Prinos, P. (2008). Non-stationary point process models for extreme storm surges, *Flood Risk Management Research into Practice*, Oxford, 1045-1054.

23. Bardet, L., Duluc, C. M., Rebour, V. and L'Her, J. (2011). Regional frequency analysis of extreme storm surges along the French coast. *Natural Hazards and Earth System Sciences*, **11(6)**, 1627-1639.

24. Galiatsatou, P. and Prinos, P. (2005). Analysis of dependence in a bivariate process of extreme waves and surges, *Proceedings of the 1st International Conference on Coastal Zone Management and Engineering in the Middle East*, Dubai, 221-225

25. Morton, I. D. and Bowers, J. (1996). Extreme value analysis in a multivariate offshore environment. *Applied Ocean Research*, **8**, 303-317.

26. De Haan, L. and De Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes*, **1**, 7-45.

27. Ferreira, J.A. and Guedes Soares, C. (2002). Modelling bivariate distributions of significant wave height and mean period. *Applied Ocean Research*, **24**, 31-45.

28. Repko, A., Van Gelder, P.H.A.J.M., Voortman, H.G. and Vrijling, J.K. (2004). Bivariate description of offshore wave conditions with physics-based extreme value statistics, *Applied Ocean Research*, **26**, 162-170.

29. Yeh, S.P., Ou, S.P., Doong, D.J., Kao, C.C. and Hsieh, D.W. (2006). Joint probability analysis of waves and water level during typhoons. *In Proceedings of the Third Chinese-German Joint Symposium on Coastal and Ocean Engineering*.

30. Galiatsatou, P. (2007). Joint exceedance probabilities of extreme waves and storm surges. *XXXIII Congress of IAHR*, pp 780 (abstract), (JFK Competition).

31. Wahl, T., Mudersbach, C. and Jensen, J. (2012). Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a multivariate statistical approach based on copula functions, *Natural Hazards and Earth System Sciences*, **12**, 495-510.

32. Corbella, A. and Stretch, D. D. (2013). Simulating a multivariate sea storm using Archimedean copulas. *Coastal Engineering*, **76**, 68–78.

33. Masina, M., Lamberti, A. and Archetti, R. (2015). Coastal flooding: A copula based approach for estimating the joint probability of water levels and waves. *Coastal Engineering*, **97**, 37-52.

34. Grimaldi, S. and Serinaldi, F. (2006). Design hyetographs analysis with 3-copula function. *Hydrological Sciences Journal*, **51(2)**, 223-238.

35. Shiau, JT. (2006). Fitting drought duration and severity with two dimensional copulas. *Water Resources Management*, **20(5)**, 795-815.

36. Zhong, H., van Overloop, P.-J. and van Gelder, P. (2013). A joint probability approach using a 1-D hydrodynamic model for estimating high water level frequencies in the Dutch Lower Rhine Delta. *Natural Hazards and Earth System Sciences*, **13**, 1841-1852.

37. Zhang, L. (2005). Multivariate hydrological frequency analysis and risk mapping. *Doctoral dissertation*, Beijing Normal University.

38. Chebana, F., Ouarda, T.B.M.J. and Duong, T.C. (2013). Testing for multivariate trends in hydrologic frequency analysis. *Journal of Hydrology*, **486**, 519-530.

39. Bender, J., Wahl, T. and Jensen, J. (2014). Multivariate design in the presence of non-stationarity. *Journal of Hydrology*, **514**, 123-130.

40. Coles, S. (2001). *An introduction to statistical modelling of extreme values*. Springer Series in Statistics, Springer, Berlin.

41. Hosking, J.R.M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society (Series B)*, **52**, 105-124.

42. De Kort, J. (2007). Modeling tail dependence using copulas-literature review, http://ta.twi.tudelft.nl/nw/users/vuik/numanal/kort_scriptie.pdf [accessed 29 February 2016]

43. Joe, H. and Xu, J.J. (1996). The Estimation Method of Inference Functions for Margins for Multivariate Models. Working paper, Department of Statistics, University of British Columbia.

44. Genest, C., Ghoudi, K. and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543-552.

45. Salvadori, G., De Michele C and Durante, F. (2011) Multivariate design via copulas, Hydrology & Earth System Sciences Discussions, **8(3)**, 5523-5558.

46. Gräler, B., van den Berg, M.J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B. and Verhoest, N.E.C. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. Hydrology and Earth System Sciences, **17**, 1281-1296.

47. Stockdon, H. F., Holman, R. A., Howd, P. A. and Sallenger, A. H. (2006). Empirical parameterization of setup, swash, and runup. *Coastal Engineering*, **53**, 573-588.

48. Booij, N., Ris, R.C., and Holthuijsen, L.H. (1999). A Third-Generation Wave Model for Coastal Regions. 1. Model Description and Validation. *Journal of Geophysical Research*, **104**, 7649-7666.

49. Androulidakis, Y.S., Kombiadou, K.D., Makris, C.V., Baltikas, V.N. and Krestenitis, Y.N. (2015). Storm surges in the Mediterranean Sea: Variability and trends under future climatic conditions. *Dynamics of Atmospheres and Oceans*. **71**, 56-82.

50. Dickinson, R, Errico, R., Giorgi, F. and Bates, G. (1989). A regional climate model for the western United States. *Climate Change*, **15(3)**, 383-422.

51. Hosking, J.R.M., Wallis, J.R. (1997). Regional Frequency Analysis: An Approach based on L-Moments. *Cambridge University Press*, 238 p.

52. Hipel, K.W. and McLeod, A.I. (2005). Time Series Modelling of Water Resources and Environmental Systems. Electronic reprint of the book originally published in 1994, http://www.stats.uwo.ca/faculty/aim/1994Book/ [accessed 29 February 2016]

53. Chambers, J. M. (1992). *Linear models*. Chapter 4 of *Statistical Models in S* eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.