Trivariate copula to design coastal structures

Olivier Orcel[1], Philippe Sergent[1], François Ropert[1]

[1]Cerema, Margny-Lès-Compiègne, 60280, France

Correspondence to: Philippe Sergent (philippe.sergent@cerema.fr)

Abstract. Some coastal structures must be redesigned in the future due to rising sea levels caused by global warming. The design of structures subjected to the actions of waves requires an accurate estimate of the long return period of such parameters as wave height, wave period, storm surge and more specifically their joint exceedance probabilities. The simplified Defra method that is currently used in particular for European coastal structures makes it possible to directly connect the joint exceedance probabilities to the product of the univariate probabilities by means of a single factor. These schematic correlations do not, however, represent all the complexity of the reality because of the use of this single factor. That may lead to damaging errors in coastal structure design. The aim of this paper is therefore to remedy the lack of robustness of these current approaches. To this end, we use copula theory with a copula function that aggregates joint distribution function to its univariate margins. We select a bivariate copula that is adapted to our application by the likelihood method with a copula parameter that is obtained by the error method. In order to integrate extreme events, we also resort to the notion of tail dependence. We select the copulas with the same tail dependence as data. In the event of an opposite tail dependence structure, we resort to the survival copula. The tail dependence parameter makes it possible to estimate the optimal copula parameter. The most robust copulas for our practical case with applications in Saint-Malo and Le Havre (in Northern France) are the Clayton normal copula and the Gumbel survival copula. The originality of this paper is the creation of a new and robust trivariate copula with an analysis of the sensitivity to the method of construction and to the choice of the copula. Firstly, we select the best fitting of the bivariate copula with its parameter for the two most correlated univariate margins. Secondly, we build a trivariate function. For this purpose, we aggregate the bivariate function with the remaining univariate margin with its parameter. We show that this trivariate function satisfies the mathematical properties of the copula. We finally represent joint trivariate exceedance probabilities for a return period of 10, 100 and 1000 years. We finally conclude that the choice of the bivariate copula is more important for the accuracy of the trivariate copula than its own construction.

1 Introduction

The design of coastal structures requires the multiplicity of variables and their degree of correlation to be taken into account. We must therefore address the lack of robustness in the modelling procedure of the dependencies between the different variables characterizing the sea state (Sergent *et al.*, 2014; Hawkes, 2005) such as wave height H, wave period T and storm surge S. The design of coastal structures is based in particular on the return periods of wave overtopping or of armour damage (Ciria *et al.*, 2007). Since the applications on wave overtopping and armour damage depend on the parameters of the coastal structure, we will not deal with the return periods of these quantities. The aim of this paper is however to improve the methods of estimating them in order to avoid costly and inappropriate decisions (Li *et al.*, 2008). To this end, we provide accurate estimates of the correlations between the variables H, T and S and obtain reliable return periods. Currently, in reference manuals such as the Rock Manual (Ciria *et al.*, 2007), it is recommended that a factor be applied to the product of univariate survival functions in order to determine the joint period. Copulas are mathematical tools for modelling the dependence structure of several random variables. The theory of copulas was developed by the mathematician Sklar (1959). The copula is a written form of the joint distribution function that provides all the information on the dependency structure. The recent interest in copulas started in financial risk management and insurance. Its use in environmental science especially concerns hydrology with the works for example of De Michele and Salvadori (2003), Favre *et al.* (2004), Grimaldi and Serinaldi (2006), Genest and Favre (2007), Zhang and Singh (2007), Aghakouchak *et al.* (2010), Lee *et al.* (2013), Chang *et al.* (2016).

In coastal engineering, in order to estimate the probability of failure of coastal or offshore structures caused in particular by the critical appearance of the combinations of parameters during a storm, Salvadori *et al.* (2007) use a copula in order to link the intensity of storm surge to its duration. Using the copula theory, Hawkes (2005) obtains, for example, all the pairs of variables wave height H and surge S for a given return period. The bivariate return period can be generalized to the multivariate case (Charpentier, 2014).

In this paper we propose the use of copulas to take into account the dependence between three variables H, T and S. Copulas generally aggregate only two random variables. The purpose of this article is the creation of a new trivariate copula and the evaluation of its robustness. Nelsen (1985) mentions that the construction of a trivariate copula requires a specific attention. In the literature the Chakak and Koehler (1995) method is commonly used and in particular by Joe (1997) and Salvadori *et al.* (2007). This method is based on bivariate conditional distributions and requires the use of three bivariate copulas. The method has a compatibility problem. There is no guarantee that the method gives the same result when the order of variables is changed. Aas and Berg (2009) propose copula construction with conditional sets : the pair copula construction (PCC). As the bivariate copulas that are selected as the most promising in our application are Archimedean copulas, simpler methods of construction are available.

Gouldby et al. (2014) propose a methodology for deriving extreme nearshore sea conditions for structural design with waves, winds and sea levels as offshore variables using also conditional distributions.

Corbella and Strech (2013) nevertheless study trivariate copula based on storm magnitude, storm duration and wave height. They show that the fully nested method of creating hierarchical copulas provides the best results for their case study. This method appears moreover to be simpler than the Chakak and Koehler (1995) procedure and the conditional mixture with its complicated integral to solve. According to Corbella and Stretch (2013), the conditional mixture is conceptually similar to that of Chakak and Koehler (1995). Based on these conclusions concerning results and complexity, we propose to use a fully nested hierarchical trivariate copulas and to test the sensitivity of the results to the method of construction and to the choice of the copula. Showing that Archimedean copulas give the best results, we can indeed adopt a fully nested hierarchical copula.

In a first part, we define the theory by presenting, partly in appendix, the marginal distribution, the recommended method of the Rock Manual, the normal copula, the bivariate copula, the tail dependence, the survival copula, the trivariate copula and isovalue lines for different return periods. We obtain a bivariate copula and the copula parameter by the method of maximum likelihood and the method of the error. We show that the trivariate function that is obtained satisfies the mathematical properties of a copula.

In a second part, we present the isovalue lines for applications at the ports of Le Havre and Saint-Malo (Northern France) with bivariate copulas corresponding to different return periods. We show that the Clayton and Gumbel copulas are the most robust copulas for our practical applications of coastal engineering.

Finally, in a third part, we apply trivariate copulas in Le Havre.

2 Theoretical approach

The notations and the main notions of copula for a bivariate distribution function are recalled in appendix A. In order to determine the return period of events that lead to wave overtopping or armour damages, we choose to use survival functions. As mentioned by Serinaldi (2015), this option is not unique and will lead to a specific return period that he denotes T_{AND}. We present here the sets of data on the sites, the selection of the best bivariate copula and the construction of trivariate copulas.

2.1 Sets of data

The approach is applied in two ports in Northern France, Saint-Malo and Le Havre that are presented in Figure 1.

Figure 1 : The Saint–Malo and Le Havre sites.

To characterize oceanic forcing, we introduce three random variables, wave height H, wave period T and storm surge S. The wave height is the significant wave height that is noted H in order to simply the notation.

By convention, the random variables are written in capital letters and the realizations of these random variables are written in lowercase (h, t, s).

The probability density functions (PDF) f_H, f_T and f_S are the result of calibrations of statistic exponential laws of data recorded at high tide and collected by the Candhis wave buoy network for waves and by tide gauge measurements recorded in the ports for storm surge.

As the study focuses on the integration of tidal range in the macrotidal environment in the calculation of the probability of joint occurrence of waves and water levels, the used data are those of waves and surges taken at high tide. The sample is made of 706 events per year using the same definition as in the Rock Manual. The independence assumption is not completely valid when two tuples per day are selected but that is an approximation commonly used. Another approximation is the assumption of the presence of a unique wave population. This assumption is also not completely valid when we consider the wave direction of extreme events. The topic has already been discussed by Hawkes (2002) and Mazas (2017, 2019), among others. The treatment of wave direction can also be considered as a fourth random variable of the oceanic forcing but has not been included in this work.

For low and moderate values the density functions are the empirical density functions. For the strongest and extreme values, the density functions result from an adjustment of the exponential law.

Kergadallan (2015) recommends selecting the maximum H value within a time window centered on the time of high water. Using the same data, this recommendation is followed.

2.2 Selection of the best bivariate copula by two methods

2.2.1 The error method

We illustrate the method for the random variables wave height H and storm surge S. This method consists in determining the mean error e between the calculated joint cumulative distribution function $F_{cal}(h, s, \theta)$ with the copula C and its parameter θ and the observed joint cumulative distribution function $F_{mes}(h, s)$.

$$e = \frac{1}{n} \sum_{i=1,n} \left| \ln \frac{F_{cal}(h_i, s_i, \theta)}{F_{mes}(h_i, s_i)} \right| \quad (1)$$

with n the number of pairs of values (h_i, s_i).

For each copula, we first determine the parameter θ that minimizes the error *e*. We then select the copula with the lowest minimum error.

2.2.2 The maximum likelihood method

Let us call *X* the sample of measures (x_1, x_2, ..., x_n) with bivariate $x_i = (h_i, s_i)$, i = 1, .., *n*. The likelihood function *L* is defined by equation (2) :

$$L(X,\theta) = \prod_{i=1}^{n} f_{cal}(h_i, s_i, \theta) \tag{2}$$

where f_{cal} is the probability density function of the bivariate cumulative distribution function F_{cal}. θ is the parameter of the copula.

The maximum likelihood method consists in finding the parameter θ, which maximizes the probability of obtaining the sample (Tassi, 2004). Since likelihood is a product of density we take its log-likelihood in order to facilitate calculations. We can thus work with the sum and derive it with respect to θ.

$$\frac{\partial}{\partial\theta} \ln L(X\,,\theta) = \frac{\partial}{\partial\theta} \ln \sum_{i=1}^{n} f_{cal}(h_i, s_i, \theta) \tag{3}$$

The best copula is the copula with the largest likelihood.

2.3 Construction of a trivariate copula

For more than two variables, *C* is not generally a copula (impossibility theorem of Genest (1995)). According to Nelsen (2006), it is difficult to construct n-order copulas from n-1 copulas. We present two methods for the construction of trivariate copulas. In the first method, a trivariate copula generalizes the bivariate copula with three random variables and one parameter. In the second method, a trivariate copula associates two bivariate copulas with their two respective parameters.

2.3.1 Definition of a copula in dimension *d* > 2

A copula in dimension *d* is a distribution function on $[0,1]^d$ whose marginal laws are uniform on [0,1].

A copula is a function C: $[0,1]^d$ —> [0,1], which satisfies the following three conditions :

$$\begin{array}{lll} i) & C(u_1,..,u_{i-1},0,u_{i+1}\,,..,u_d) = 0 & \forall u_i \in [0,1] \\ ii) & C(1,\dots,1,u_i,1,\dots,1) = u_i & \forall u_i \in [0,1] \\ iii) & C\ is\ d-growing & \end{array} \tag{4}$$

A function h : $[0,1]^d$ —> R is called *d*-growing if for any hyper-rectangle [*a*,*b*] of R^d, $V_h([a,b]) \geq 0$, where

$$V_h([a,b]) = \Delta_a^b h(t) = \Delta_{a_d}^{b_d}\Delta_{a_{d-1}}^{b_{d-1}} \dots \dots \Delta_{a_2}^{b_2}\Delta_{a_1}^{b_1} h(t) \tag{5}$$

For each t, $\Delta_{a_i}^{b_i} h(t) = h(t_1, \dots, t_{i-1}, b_i, t_{i+1}, \dots., t_n) - h(t_1, \dots, t_{i-1}, a_i, t_{i+1}, \dots., t_n)$.

2.3.2 Trivariate copula with one parameter : a multi-level Archimedean trivariate

Since we are looking for the correlation between three variables, the first idea is to generalize the bivariate copula $C(u_1, u_2)$ to obtain $C(u_1, u_2, u_3)$. We must check that $C(u_1, u_2, u_3)$ is a copula, which is difficult. However Archimedean copulas like Gumbel and Clayton can be extended to an order greater than 2 using the property of Archimedean copulas (see appendix A).

For a Clayton copula of order *n*, this gives :

$$C(u_1, \dots, u_n) \ = \ [{u_1}^{-\frac{1}{\theta}} + {u_2}^{-\frac{1}{\theta}} + \cdots + {u_n}^{-\frac{1}{\theta}} - (\mathrm{n}-1)]^{-\theta} \tag{6}$$

For Clayton copula of order 3, it gives :

$$\mathrm{C}(u_1, u_2, u_3) \ = \ [{u_1}^{-\frac{1}{\theta}} + {u_2}^{-\frac{1}{\theta}} + {u_3}^{-\frac{1}{\theta}} - 2]^{-\theta} \tag{7}$$

For Gumbel copula of order *n*, it gives :

$$C(u_1, \dots, u_n) = \exp\left(-\left[(-\mathrm{Ln}\, u_1)^\theta + (-\mathrm{Ln}\, u_2)^\theta + \dots + (-\mathrm{Ln}\, u_n)^\theta\right]^{\frac{1}{\theta}}\right) = \exp\left(-\left[\sum_i (-\mathrm{Ln}\, u_i)^\theta\right]^{\frac{1}{\theta}}\right) \qquad (8)$$

For Gumbel copula of order 3, it gives :

$$C(u_1, u_2, u_3) = \exp\left(-\left[(-\mathrm{Ln}\, u_1)^\theta + (-\mathrm{Ln}\, u_2)^\theta + (-\mathrm{Ln}\, u_3)^\theta\right]^{\frac{1}{\theta}}\right) \qquad (9)$$

By taking a single copula parameter for the three variables, we do not differentiate the two-to-two correlations of the variables even though some variables may be more correlated than others.

2.3.3 Trivariate copula with two parameters : a fully nested hierarchical copula

To better take into account the correlations of variables two by two, one option is to build trivariate functions from bivariate copulas as a fully nested hierarchical copula:

$$C(u_1, u_2, u_3) = C_1(C_2(u_1, u_2), u_3) \qquad (10)$$

Corbella (2013) tests a fully nested hierarchical copula but he uses a unique bivariate copula and does not distinguish the two bivariate copulas C_1 and C_2. C_1 is a bivariate copula with θ_1 as copula parameter. C_2 is a bivariate copula with θ_2 as copula parameter. We must check that this function (10) is a copula and satisfies the properties of equations (4). We first aggregate the two most correlated variables with the copula C_2 and its copula parameter. We then add the third random variable with the copula C_1 and its copula parameter. We will show later that this order provides the most robust copula.

2.3.4 Validity of copula properties for *2.3.3*

We do not know any general methods to build high order copulas from low order copulas (Durrleman, 2010). Generally $C(u_1, u_2, u_3) = C_1(C_2(u_1, u_2), u_3)$ is not a copula. To prove that $C(u_1, u_2, u_3)$ is a copula, we must check that $C(u_1, u_2, u_3)$ satisfies the three properties of equation (4) with d = 3, which is difficult. However Charpentier (2014) points out that C is a copula if it satisfies i) or ii).

i) C_1 and C_2 are both Clayton or Gumbel copulas with parameters θ_1 for C_1 and θ_2 for C_2 positive and growing.

ii) C_1 and C_2 are both Archimedean copulas of respective generator ϕ_1, ϕ_2 with $\phi_2 \mathrm{o} \phi_1^{-1}$ being the inverse of a Laplace transform.

For Gumbel and Clayton copulas C_1 and C_2 that are Archimedean copulas we check the condition (ii) that there is a function f for which the inverse Laplace transform T_L^{-1} satisfies :

$$T_L^{-1}[f] = \phi_2 \mathrm{o} \phi_1^{-1} \qquad (11)$$

with ϕ_1, ϕ_2 generators of the copulas C_1 and C_2. $T_L[f](s) = \int_0^{+\infty} e^{-st} f(t)dt$ is the Laplace transform of f.

For C_1 and C_2 Clayton copulas we have as the generator of C_2 and as the inverse generator of C_1 :

$$\phi_2(t) = \frac{t^{-\theta_2 - 1}}{\theta_2} ; \phi_1^{-1}(t) = (1 + \theta_1 \mathrm{t})^{-\frac{1}{\theta_1}} \qquad (12)$$

This gives :

$$\phi_2 \mathrm{o} \phi_1^{-1}(t) = \frac{\left[(1 + \theta_1 t)^{\frac{\theta_2}{\theta_1}} - 1\right]}{\theta_2} \qquad (13)$$

We can find that :

$$T_L[\phi_2 \mathrm{o} \phi_1^{-1}](s) = \frac{\left[e^{\frac{s}{\theta_1}} \Gamma(\frac{\theta_2}{\theta_1} + 1, \frac{s}{\theta_1}) - 1\right]}{s\theta_2} \qquad (14)$$

With $\Gamma(a, x)$ the incomplete Gamma function set by for a complex with real part(a) > 0 :

$$\Gamma(a, x) = \int_x^{+\infty} t^{a-1} e^{-t}\, dt \qquad (15)$$

We conclude that there is a function f such that $\phi_2 \mathrm{o} \phi_1^{-1} = T_L^{-1}[f]$:

$$f = \frac{\left[e^{\frac{s}{\theta_1}}\Gamma(\frac{\theta_2}{\theta_1}+1,\frac{s}{\theta_1})-1\right]}{s\theta_2} \tag{16}$$

For C_1 and C_2 Gumbel copulas we have as generator of C_2 and as the inverse generator of C_1 :

$$\phi_2(t) = (-\ln t)^{\theta_2}\ ;\ \phi_1^{-1}(t) = e^{-t^{\frac{1}{\theta_1}}} \tag{17}$$

This gives :

$$\phi_2 o \phi_1^{-1}(t) = \left[-\ln\left(e^{-t^{\frac{1}{\theta_1}}}\right)\right]^{\theta_2} = t^{\frac{\theta_2}{\theta_1}} \tag{18}$$

We can find that :

$$T_L[\phi_2 o \phi_1^{-1}](s) = T_L\left(t^{\frac{\theta_2}{\theta_1}}\right) = \Gamma\left(\frac{\theta_2}{\theta_1}\right)s^{-\frac{\theta_2+\theta_1}{\theta_1}} \tag{19}$$

With Γ Gamma function, defined by :

$$\Gamma(a) = \int_0^{+\infty} y^{a-1}e^{-y}\,dy \tag{20}$$

We conclude that there is a function *f* such that $\phi_2 o \phi_1^{-1} = T_L^{-1}[f]$:

$$f = \Gamma\left(\frac{\theta_2}{\theta_1}\right)s^{-\frac{\theta_2+\theta_1}{\theta_1}} \tag{21}$$

2.4 Determination of the contour of equal joint exceedance probability

The determination of the contour of equal joint exceedance probability consists in obtaining all the variables (*H*, *T*, *S*) associated with different return periods : T_{10} (10-year event), T_{100} (100-year event) and T_{1000} (1000-year event).

2.4.1 Bivariate probability without tide

We deal with a set of pairs of values (*h*, *s*) that satisfy :

$$\bar{C}[\bar{F}_H,\bar{F}_s] = f_{10}, f_{100} \text{ or } f_{1000} \tag{22}$$

$\bar{C}$ is the selected bivariate survival copula. $\bar{F}_H, \bar{F}_s$ are survival functions associated with the variables. The values f_{10}, f_{100} or f_{1000} are the frequencies corresponding to the ten-year, hundred-year and thousand-year periods.

2.4.2 Bivariate probability with tide

The bivariate probability with tide requires the development of the copula connecting wave height and storm surge. We can then define the joint survival function of the wave height and the storm surge. The chosen calculation method favors high tide. The sea levels considered are therefore the sums of the astronomical high tide (generated by the attraction of the moon and the sun without weather disturbance) and the storm surges raised at the time of these astronomical high tides. This method is of course valid only for macrotidal seas. The equation (23) established by Simon (1994) gives the probability that the sea level at high tide *N e*xceeds a given value *n* :

$$\mathrm{P}(n) = \mathrm{P}[N > n] = \int_{M_{min}}^{M_{max}} f_M(z)\bar{F}_S(n-z)\mathrm{d}z \tag{23}$$

z is the height of the high tide, between the minimum and maximum values M_{min} and M_{max} respectively at high tide.
$f_M(z)dz$ is the probability that the high tide is between *z* and *z* + d*z*.
$\bar{F}_S(s)$ is the probability of observing a storm surge *S* larger than s, thus $\bar{F}_S(s)$ = P(*S* > *s*).
The bivariate survival function for wave height *H* and sea level *N* is therefore written as follows :

$$\bar{F}_{HN}(h,n) = \int_{M_{min}}^{M_{max}} f_M\,(z)\bar{F}_{HS}(h,n-z)dz \tag{24}$$

This can be written by introducing the survival copula $\bar{C}$:

$$\bar{F}_{HN}(H,N) = \int_{M_{min}}^{M_{max}} f_M\,(z)\bar{C}(\bar{F}_H(h),\bar{F}_S(n-z))dz \tag{25}$$

The set of pairs (*h*, *n*) corresponding to the different return periods, the ten-year, hundred-year and thousand-year periods, satisfies :

$$\int_{M_{min}}^{M_{max}} f_M\,(z)\bar{C}(\bar{F}_H(h),\bar{F}_S(n-z))dz = f_{10}, f_{100} \text{ or } f_{1000} \tag{26}$$

It is thus possible to represent the contour of equal joint exceedance probability associated with the variables wave height and sea level .

2.4.3 Trivariate probability without tide

Here we have chosen the method of construction of a trivariate copula with two parameters known as fully nested hierarchical copula. We have :

$$\bar{F}_{HT}(h,t) = \bar{C}_1(\bar{F}_H(h),\bar{F}_T(t)) \tag{27}$$

$$\bar{F}_{HTS}(h,t,s) = \bar{C}_2(\bar{F}_{HT}(h,t),\bar{F}_S(s)) \tag{28}$$

with $\bar{C}_1$ and $\bar{C}_2$ the selected bivariate survival copula. From equations (27) and (28) we therefore obtain the equation (29) :

$$\bar{F}_{HTS}(h,t,s) = \bar{C}_2(\bar{C}_1(\bar{F}_H(h),\bar{F}_T(t)),\bar{F}_S(s)) \tag{29}$$

The triplets of values (*h*, *t*, *s*) corresponding to the different return periods, T_{10} (10-year event), T_{100} (100-year event) and T_{1000} (1000-year event) satisfy :

$$\bar{C}_2(\bar{C}_1(\bar{F}_H(h),\bar{F}_T(t)),\bar{F}_S(s)) = f_{10}, f_{100} \text{ or } f_{1000} \tag{30}$$

It is thus possible to represent the contours of equal joint exceedance probability associated with the variables wave height, wave period and sea level.

2.4.4 Trivariate joint exceedance probability with tide

The trivariate survival function for wave height *H*, wave period *T* and sea level *N* is written as follows:

$$\bar{F}_{HTN}(h,t,n) = \int_{M_{min}}^{M_{max}} f_M\,(z)\bar{F}_{HTS}(h,t,n-z)dz \tag{31}$$

This can be written by introducing the selected survival copula $\bar{C}_2$:

$$\bar{F}_{HTN}(h,t,n) = \int_{M_{min}}^{M_{max}} f_M(z)\,\bar{C}_2(\bar{F}_{H,T}(h,t),\bar{F}_S(n-z))dz \tag{32}$$

This expression can be written by introducing the survival copula $\bar{C}_1$ connecting $\bar{F}_H$ and $\bar{F}_T$.

$$\bar{F}_{HTN}(h,t,n) = \int_{M_{min}}^{M_{max}} f_M(z)\,\bar{C}_2(\bar{C}_1(\bar{F}_H(h),\bar{F}_T(t)),\bar{F}_S(n-z))dz \tag{33}$$

The triplets of values (*h*, *t*, *n*) corresponding to the different return periods, T_{10} (10-year event), T_{100} (100-year event) and T_{1000} (1000-year event) satisfy :

$$\int_{M_{min}}^{M_{max}} f_M(z)\,\bar{C}_2(\bar{C}_1(\bar{F}_H(h),\bar{F}_T(t)),\bar{F}_S(n-z))dz = f_{10}, f_{100} \text{ or } f_{1000} \tag{34}$$

It is thus possible to represent the contours of equal joint exceedance probability associated with the variables wave height, wave period and sea level with tide.

2.5 Tail dependence of the sample

It is necessary to treat the extreme events that are characterized by a very low occurrence. The difficulty of taking them into account is of a statistical nature: the scarcity of observations. In order to take the extreme events into account, we introduce the concept of tail dependence. For a bivariate copula, it measures the probability of simultaneous extreme realizations (Clauss, 2009). It describes the dependences of distribution tails for the simultaneous occurrence of extreme values. It is a highly relevant tool for the study of extreme values. We distinguish lower and upper tail dependences. They are characterized by their lower and upper tail dependence coefficients that are deduced from the following conditional probabilities, whose value is given by equations (35) and (36) that, in tum, are given by (Clauss, 2009) :

$$\mathrm{P}(U_1 \leq u_1 | U_2 \leq u_2) = \frac{\mathrm{P}(U_1 \leq u_1, U_2 \leq u_2)}{P(U_2 \leq u_2)} = \frac{C(u_1, u_2)}{u_2} \quad (35)$$

$$\mathrm{P}(U_1 > u_1 | U_2 > u_2) = \frac{P(U_1 > u_1, U_2 > u_2)}{P(U_1 > u_1)} = \frac{1 + C(u_1, u_2) - u_1 - u_2}{1 - u_2} \quad (36)$$

Since we fix the lower tail dependence coefficient λ_L and upper tail dependence coefficient λ_U by equations (37) and (38) :

$$\lambda_L = lim_{u \to 0}\, \mathrm{P}(U_1 \leq u_1 | U_2 \leq u_2) \quad (37)$$

$$\lambda_U = lim_{u \to 1}\, \mathrm{P}(U_1 > u_1 | U_2 > u_2) \quad (38)$$

We deduce the definitions of tail dependence coefficients.

Definition: The lower tail dependence coefficient is defined by :

$$\lambda_L = \lim_{u \to 0} \frac{C(u,u)}{u} \quad (39)$$

The copula *C* has a lower tail dependence if λ_L exists with λ_L Є]0,1].

If λ_L = 0 then it does not have a lower tail dependence.

Definition : The upper tail dependence coefficient is defined by :

$$\lambda_U = lim_{u \to 1} \frac{1 + C(u,u) - 2u}{1 - u} \quad (40)$$

The copula *C* has an upper tail dependence if λ_U exists with λ_U Є]0,1].

If λ_U = 0 then it does not have an upper tail dependence.

The tail dependences of the different copulas are determined in (Nelsen, 2006) and (Roncalli, 2002) from their tail dependence coefficients. They are expressed in Table 1.

Copula	λ_L	λ_U
Fréchet	*0*	*0*
Marshall-Olkin	Min*(α,β)*	*0*
Plackett	*0*	*0*
Clayton	$2^{-\frac{1}{\theta}}$	*0*
Franck	*0*	*0*
Gumbel	*0*	$2 - 2^{-\frac{1}{\theta}}$
Joe	$2 - 2^{-\frac{1}{\theta}}$	*0*
Ali-Mikhail-Haq	*0*	*0*
Gauss	*0*	*0*

Table 1 : Tail dependence coefficients.

We find that some copulas do not have lower and upper tail dependence coefficients. They cannot deal with extreme dependence. Some copulas have a lower tail dependence, others have an upper tail dependence.

The tail dependence of the sample is firstly checked. For this we graphically represent the evolution of *C(u, u)/u* and determine its limit when *u* tends to 0. We can therefore decide whether the sample has or has not a lower or upper tail dependence.

In choosing the copula, it is essential to satisfy the tail dependence of the sample.

If the sample does not have a tail dependence, then the use of Gaussian copula or other copula with the same tail dependence characteristics is recommended.

If the sample has a lower tail dependence, the use of a copula with a lower tail dependence or the survival copula of a copula with an upper tail dependence is recommended.

If the sample has an upper tail dependence, the use of a copula with an upper tail dependence or the survival copula of a copula with a lower tail dependence is recommended.

We can also deduce the parameter of the copula from the tail dependence coefficient given by the sample.

3 Results for bivariate copulas

We select the most appropriate copulas at both the Le Havre and Saint-Malo (Northern France) sites using two methods. We analyze the tail dependence of the two samples. We represent the contour of equal joint exceedance probability with the selected copulas for three return periods in order to assess the relevance of the copulas.

3.1 Statistical law for adjusting wave height, wave period and storm surge

The representation of the contours requires knowledge of the statistical laws of adjustment of the different parameters. We therefore present these laws. For the two sites of Saint Malo and Le Havre we have used data files that provide the values for wave height, wave period and storm surge at high tide over a time period of about twenty years. The file for Le Havre site includes, for example, around 15.000 values. The wave data are extracted from the Anemoc digital database. Sea levels at high tide are extracted from tide gauge measurements. The astronomical tide is obtained from the Shom Predit software.

Adjustments of the statistical laws are made according to the POT method on the basis of the exponential law.

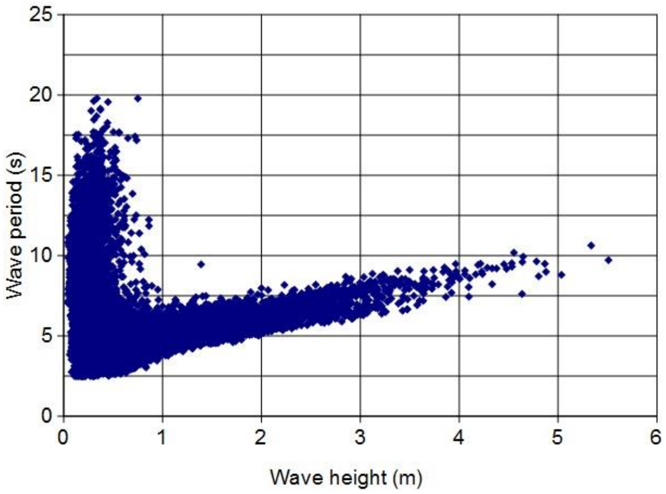


Figure 2 : Set of wave data in Le Havre (1979 – 2002).

The copula parameters were calibrated from samples where wave height values less than one meter were excluded (see Figure 2), thus reducing the sample size to about 3.000 values. The copulas are fitted to all pairs/triplets of observations where the wave height exceeds one meter.

3.2 Current pratice : Defra method

The use of the simplified Defra method in Ciria *et al.* (2007) is common among European coastal engineers for the study of wave overtopping or armor damages in coastal structures. It refers to the Defra method presented for example by Hawkes (2005) that is based on the Gauss copula. The simplified Defra method refers to univariate survival functions $\bar{F}_H$ and $\bar{F}_S$ of wave height and storm surge. The reason is that coastal engineers usually work with exceedance probability rather than with non exceedance probability. In this simplified method, the bivariate survival function is related to univariate survival functions by expression (41). In France, the order of magnitude for the FD coefficient is about 20. Kergadallan (2013) recommends however a minimum value of 25.

$$\bar{F}_{HS} = \mathrm{FD}\ \bar{F}_H \bar{F}_S \tag{41}$$

The equation (41) is used to determine the table 4.15 of Rock Manual (Ciria *et al.*, 2007). Figure 3 shows the differences between observed bivariate survival functions and calculated bivariate survival functions using the simplified Defra method. The points of calculations in blue lie far from the first bisector in black in the figure. This shows that the use of the Defra simplified method is inappropriate. This is due mostly to the use of the simplified Defra method of equation (41) but the complete Defra method with Gauss copula would not represent also perfectly the extreme events because Gauss copula has not tail dependence as we will see later.

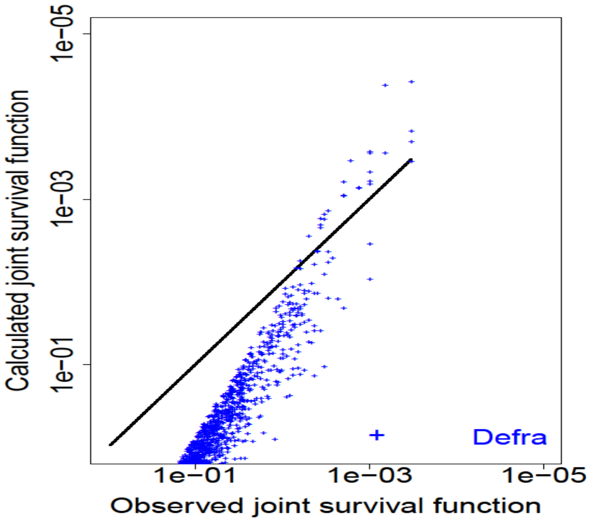


Figure 3 : Comparison of calculated (with Defra method) and observed joint frequency for Le Havre.

In order to improve the results we now introduce the copula theory.

3.3 Analysis of the tail dependence

The sample is analyzed in order to determine its tail dependence. This will affect the choice of copula. Since the sample has a tail dependence, it should be known whether it has a lower tail dependence or an upper tail dependence. Indeed, the result will condition the choice of the copula depending on whether the sample has the same tail dependence as the copula or not. To simplify the notation, we will use the survival copula $\bar{C}$ of equations (22), (26), (30), (34) as copula C. We determine its limit for u tending to 0.

This choice of the survival copula $\bar{C}$ enables to simplify the equations (22), (26), (30), (34). If we kept the standard notations, we would deal with the upper tail dependence and the chosen copulas (for example Clayton and survival Gumbel) would be said survival Clayton and Gumbel.

In the two methods, we are interested in the extreme events with large wave heights and water levels.

a) Saint-Malo b) Le Havre

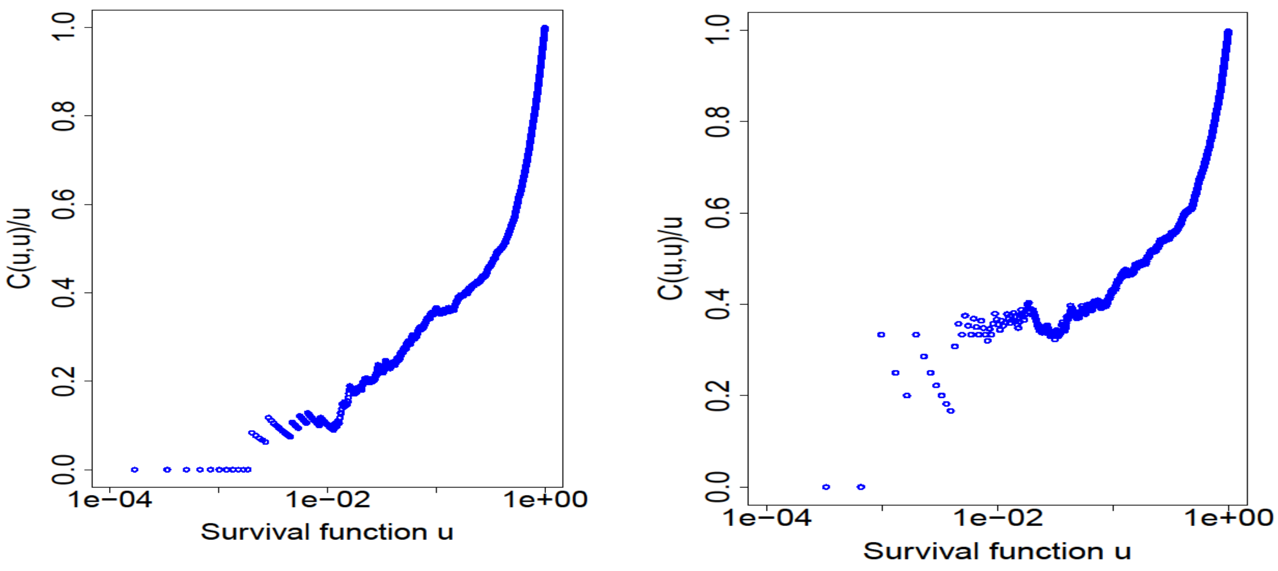


Figure 4 : $\frac{C(u,u)}{u}$ for a) Saint-Malo and b) Le Havre samples.

For the Saint-Malo sample, $\frac{C(u,u)}{u}$ tends to around 0.2 when u tends to 0.

For the Le Havre sample, $\frac{C(u,u)}{u}$ tends to around 0.4 when u tends to 0.

These two samples have a lower tail dependence which justifies the use of the Clayton copula. We determine the Clayton copula parameter from the lower tail dependence coefficient of the sample. With the Clayton copula, we can determine the value of its copula parameter in Saint-Malo and Le Havre with equation (42). This copula parameter is 0.43 and 0.76 respectively.

$$\theta = -\frac{\ln 2}{\ln \lambda_L} \tag{42}$$

Note : as the Gumbel copula has an upper tail dependence, the use of its survival copula is recommended. This analysis of the sample makes it possible to understand why the Gumbel survival copula gives a minimum of error much close to the minimum error of the Clayton copula. We can therefore expect Gumbel survival copula results to be close to the results obtained by Clayton copula.

3.4 Selection of the best bivariate copula for Le Havre and Saint-Malo samples

3.4.1 The log-likelihood method

Copula	**Copula Parameter**	**Copula Parameter**	**Maximum likelihood**	**Maximum likelihood**
Sites	**Saint-Malo**	**Le Havre**	**Saint-Malo**	**Le Havre**
Gumbel	1.09	1.29	52	185
Survival Gumbel	1.18	1.39	243	372
Clayton	0.38	0.74	**291**	**387**
Gauss	0.22	0.42	149	297
Franck	1.25	2.67	124	271
Student	0.22	0.42	157	303
Plackett	1.88	3.58	127	277
Joe	1.03	1.21	4	76
AMH	0.71	0.96	196	375
Galambos	0.31	0.54	41	175

Table 2 : Copula parameter and maximum likelihood for the different copulas in Saint-Malo and Le Havre.

For the set of copulas we determine their maximum likelihood with their parameter. We will select the copula that has the same tail dependence as the sample with the largest likelihood.

For the Saint-Malo sample, we choose the Clayton copula, which has the same tail dependence as the sample, with a log-likelihood of 291 in table 2. For the Le Havre sample, we also choose the Clayton copula, which has the same tail dependence as the sample, with a log-likelihood of 387.

The Clayton copula parameters obtained by the tail dependence coefficients come close to those obtained by the log-likelihood method for the Le Havre sample (3.040 values) and the Saint-Malo sample (5.888 values).

For Saint-Malo, we obtain as 0.38 the parameter of the Clayton copula using the method of maximum likelihood and 0.43 with the tail dependence coefficient.

For Le Havre, we obtain 0.74 as the parameter of the Clayton copula using the method of maximum likelihood and 0.76 with the tail dependence coefficient.

The value of the log-likelihood of the Gumbel survival copula is as large as the log-likelihood of the Clayton copula. In addition, the Gumbel survival copula has the same tail dependence as the Clayton. It is therefore as suitable as the Clayton copula.

The Gauss and especially the AMH copula have a relatively large likelihood. However, they do not have a correct tail dependence. They cannot therefore correctly represent the tail dependence. We will come back later to the AMH copula which has a special property.

3.4.2 The error method for the Clayton, Gumbel and survival Gumbel Copula

In order to select the most relevant copula, we represent the mean error e between the calculated survival function $F_{cal}(h, s, \theta)$ with the copula C and its parameter and the measured $F_{mes}(h, s)$.

a) Saint-Malo b) Le Havre

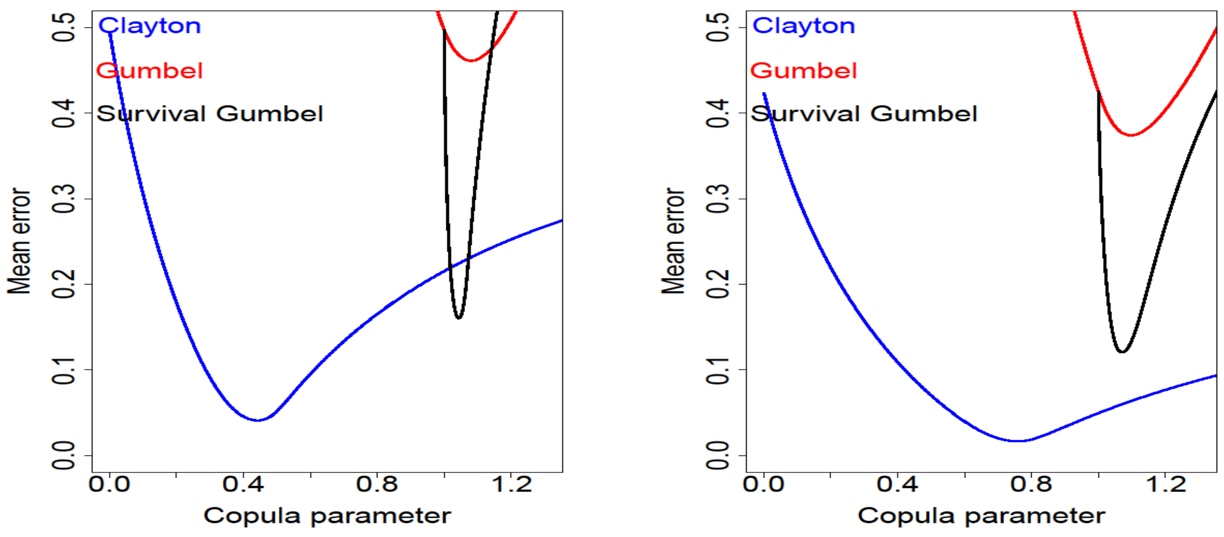


Figure 5 : Evolution of the error according to the Clayton, Gumbel and survival Gumbel copula parameter in a) Saint-Malo and b) Le Havre.

Figure 5 for the ports of Saint-Malo and Le Havre shows that the error that is obtained with the Gumbel survival copula is very close to that obtained with the Clayton copula. The curve of the error obtained by the Gumbel copula survival however has a very acute minimum. Obtaining the parameter of this copula will therefore be very sensitive to the value of its minimum error. It will therefore be necessary to determine it very precisely.

Note: Gumbel and Clayton copula parameter supports are different and are [1, + ∞ [and] 0, + ∞ [respectively.

We note Emin the minimum of the mean error e and Error rate = exp (Emin) - 1. Table 3 below shows the results obtained for Saint Malo and Le Havre.

Copula	Emin	Emin	Error rate	Error rate	Parameter	Parameter
Sites	Saint-Malo	Le Havre	Saint-Malo	Le Havre	Saint-Malo	Le Havre
Gumbel	0.45	0.37	57 %	44 %	1.03	1.10
Survival Gumbel	0.18	0.12	20 %	13 %	1.02	1.07
Clayton	0.05	0.03	5 %	3 %	0.40	0.76

Table 3 : Emin, error rate and copula parameter for the Clayton, Gumbel and Gumbel survival copula in the ports of Le Havre and Saint Malo.

Table 3 is used to verify that Clayton copula is the most robust copula. It also appears that Gumbel survival copula is also an appropriate option.

We have therefore shown by two methods that the Clayton copula is the most relevant for the Saint-Malo and Le Havre sites. The parameters of the copula obtained by the error method are close to those obtained by the method of maximum likelihood for the Clayton copula.

3.5 Comparison of observed and calculated joint frequencies

In order to assess the robustness of the copulas, we show in Figure 6 the observed and calculated joint frequencies for the Le Havre sample (3.040 pairs of values). The copula represents reality more closely as the points approach the bisector $y = x$.

The simplified Defra method currently in use does not give a good representation of the reality of the joint frequencies for wave height and storm surge. The points obtained by this simplified Defra method are very far from the bisector.

The Clayton copula provides a good representation of the reality of joint frequencies for wave height and storm surge. The points obtained by the Clayton copula come close to the bisector.

In contrast, the Gumbel copula does not give a good representation of the reality of the joint frequencies for wave height and storm surge. The points obtained by the Gumbel copula move away from the bisector. The explanation is therefore in the analysis of the sample carried out in section 3.3: we showed that the sample had a lower tail dependence whereas the Gumbel copula has an upper tail dependence.

The Gumbel survival copula provides a good representation of the reality of joint frequencies for wave height and storm surge. The points obtained by the Gumbel survival copula come close to the bisector. The explanation lies in the fact of introducing the survival copula. The tail dependence of the Gumbel survival copula is opposite to the tail dependence of the Gumbel copula. We therefore reestablish a right tail dependence which gives correct results.

The results obtained by the AMH are surprisingly correct. Kumar (2010) shows that the AMH copula does not have tail dependence except if the copula parameter is equal to 1. In our case, the copula parameter is close to 1. The copula seems therefore to behave like a copula with a lower tail dependence.

We show the utility of the Clayton copula in comparison with the Gumbel copula and the Defra method that is currently in use.

The results highlight the importance in copula selection of the tail dependence analysis of the sample. If the sample has a tail dependence it is necessary to select a copula with the same tail dependence. The Clayton copula that has the same tail dependence as the sample gives a calculated joint frequency close to the observed joint frequency. Conversely the Gumbel copula does not correctly represent the observed joint frequency: it moves away from the bisector for the extreme points. This is because the sample has a tail dependence opposite to that of the Gumbel copula. In order to restore the proper tail dependence, we resort to the survival copula. The latter comes close the bisector but is slightly less robust than the Clayton copula. It should be noted that calibration is performed on the entire sample. By truncating the sample for joint frequency values below 0.01, we would have obtained a much larger parameter for the Gumbel copula with results that are closer to measurements.

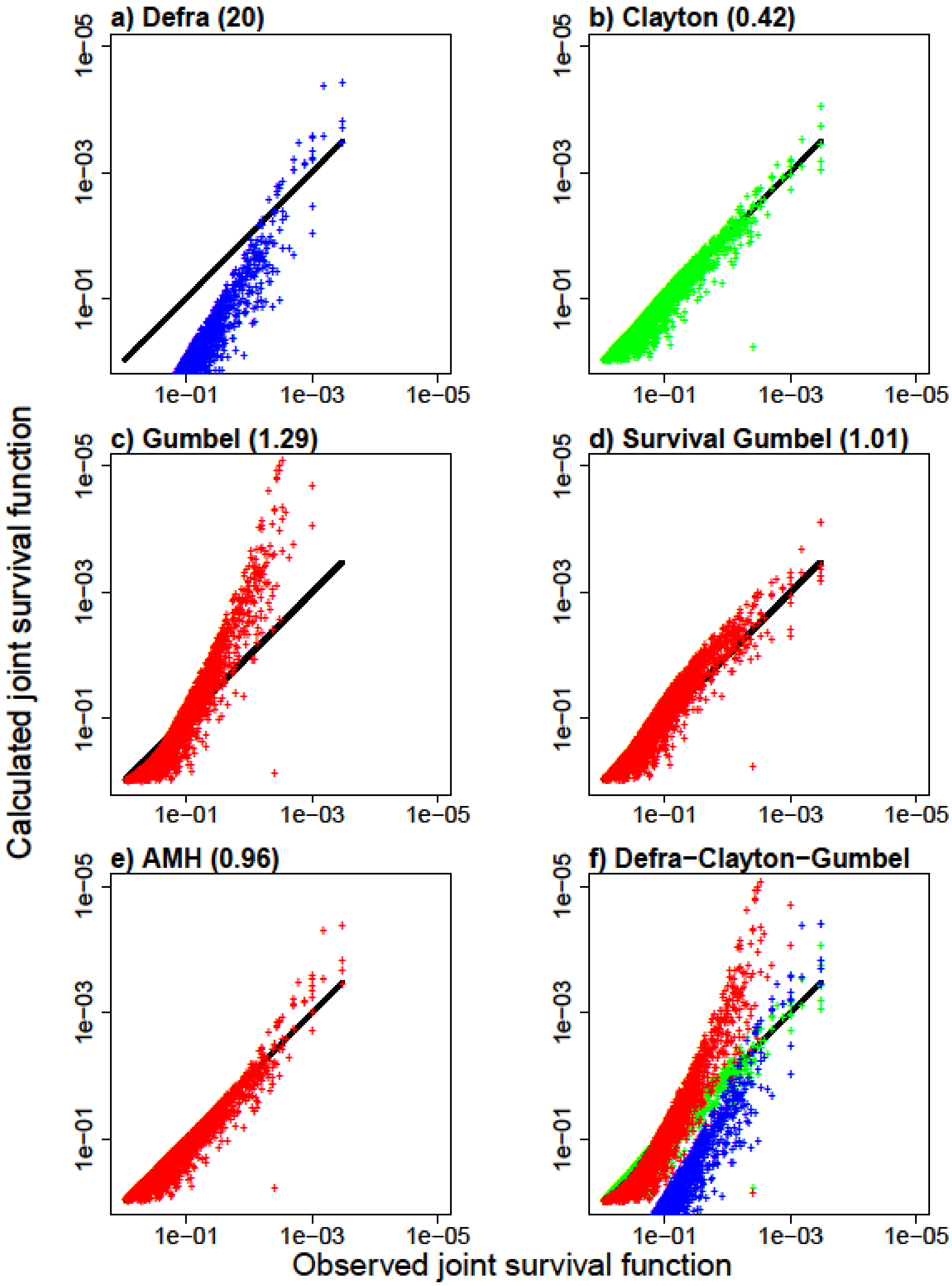
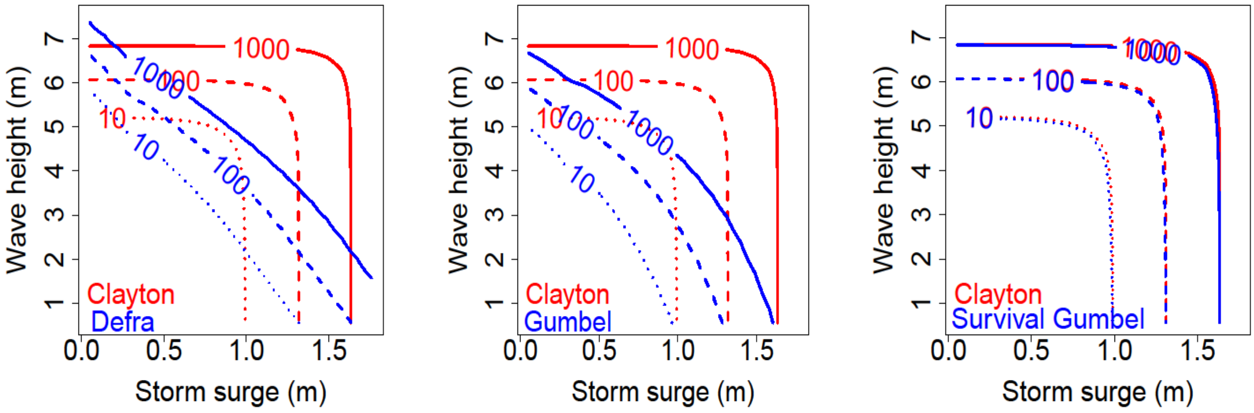


Figure 6 : Comparison of the observed joint survival and the calculated joint survival function for Le Havre with (a) Defra method, (b) Clayton (0.42), (c) Gumbel (1.29), (d) Survival Gumbel (1.01), (e) AMH (0.96) and (f) Defra-Clayton-Gumbel.

380 **3.6 Contours of equal joint exceedance probability with bivariate copula**

3.6.1 Contours without tide for the Clayton, Gumbel, and Survival Gumbel copulas and the Defra method

Figure 7 shows the joint exceedance probability (*H*, *S*) for the Le Havre (3.040 values) samples respectively with Clayton, Gumbel, Gumbel survival copulas and the Defra method.



a) Defra (20)
Clayton (0.74)

b) Gumbel (1.29)
Clayton (0.74)

c) Survival Gumbel (1.39)
Clayton (0.74)

Figure 7 : Contours of equal joint exceedance probability with Clayton (0.74), Defra (20), Gumbel (1.29) and survival Gumbel (1.39) for return periods of 10, 100 and 1000 years for Le Havre.

Figures 7a, 7b and 7c present the comparison of Clayton with respectively Defra, Gumbel and Survival Gumbel. Contours of equal joint exceedance probabilities obtained by Clayton are very far from those obtained by Gumbel and the Defra method. On the contrary, the joint exceedance curves obtained using the Gumbel survival copula are very similar to those obtained with Clayton. Results are therefore very sensitive to the choice of copula. A poor choice may lead to undersizing and may have economic consequences.

3.6.2 Contours with tide for Clayton copula

Figure 8 shows the contours of equal joint exceedance probability respectively for the port of Saint-Malo (5.000 tidal values) and the Le Havre sample (22.000 tidal values) with the Clayton copula.

a) Saint-Malo

b) Le Havre

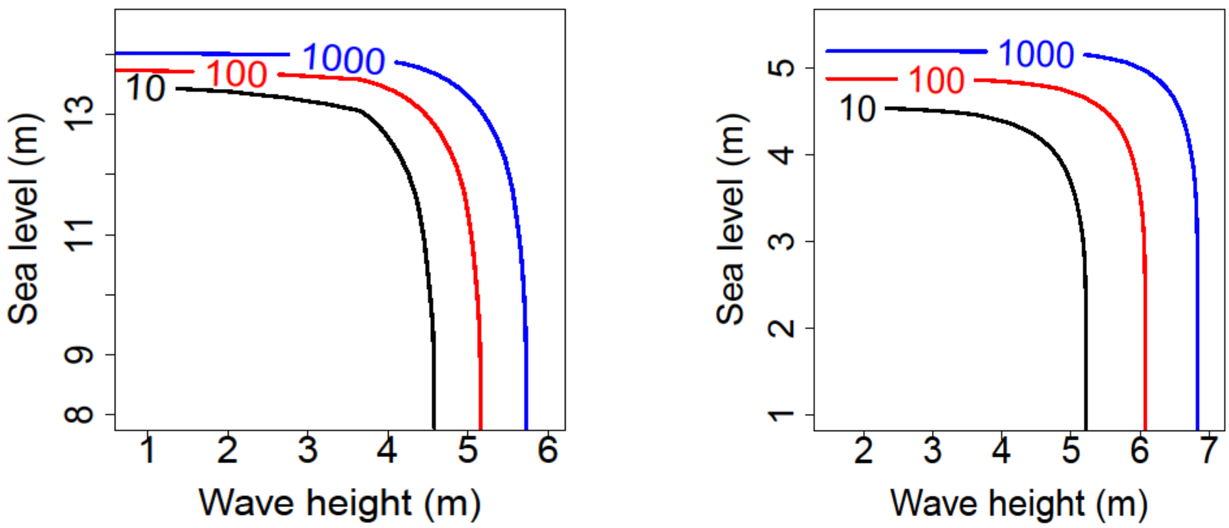


Figure 8 : a) Joint exceedance probability obtained with Clayton copula (0.38) with tide for return periods of 10, 100 and 1000 years for Saint Malo and b) with Clayton copula (0.74) with tide for return periods of 10, 100 and 1000 years for Le Havre.

With tide the effect of storm surge on the sea level is small. The tidal range, which has an amplitude much larger than the storm surge especially for the port of Saint Malo, mitigates the variations due to the storm surge. In particular, for the port of Saint-Malo, it can be seen that sea level is less sensitive to variations in the return periods than storm surge (cf. Figure 8).

3.7 Conclusion on selecting of the best bivariate copula

We selected the Clayton copula for the ports of Le Havre and Saint-Malo using three methods. In order to validate the Clayton copula, we analyzed samples from 19 sites of the French coast along the Atlantic and English Channel with the maximum likelihood method. We always obtained the greatest maximum likelihood with the Clayton copula or the AMH copula (see appendix C). The sample always has a lower tail dependence (see appendix B). We can therefore conclude that the Clayton copula is the most appropriate copula for our application. For this purpose, the Table 4 gives the parameters of the different sites.

Sites	Parameter
Dunkerque	0.67
Calais	0.56
Boulogne-sur-mer	0.77
Dieppe	0.80
Le Havre	0.95
Cherbourg	0.49
Saint-Malo	0.48
Roscoff	0.41
Le Conquet	0.54
Brest	0.55
Concarneau	0.93
Port-Tudy	0.92
Saint-Nazaire	1.05
Saint-Gildas	0.9
La Rochelle	1.00
Bayonne	0.43
Socoa	0.43
Port-Bloc	0.95

Table 4 : Clayton parameters for the different sites.

Even though in some sites the AMH copula provides a larger likelihood than the Clayton copula, it should not be chosen because it has a particular kind of behavior. It has a lower tail dependence if the copula parameter is 1 (or close to 1 in practice). If the parameter is not 1, the AMH copula does not have tail dependence and its interests disappears. Since the robustness depends on the copula parameter and on the site, it cannot be recommended for a general use.

4 Results for trivariate copulas

4.1 State of the art

Corbella (2013) mentions multivariate copulas with the application of a trivariate copula linking wave height, storm surge and storm duration. Comparing different construction methods, he concludes that the Chakak and Koehler (1995) method that is based on bivariate conditional distribution is too complex and not robust enough. Neither is he in favor of the use of the conditional mixtures approach for the same reasons. He therefore recommends the nested hierarchical construction with Archimedean copulas. Based on his guidelines, we have not tested conditional distributions that have been used by other authors like for example Aas and Berg (2009) or Gouldby *et al.* (2014). We have tested hierarchical construction using a fully nested hierarchical Archimedean copula. In this type of construction, we build a bivariate copula between two parameters, then we create a trivariate copula with the previous copula and the third parameter. Unlike Corbella (2013) we introduce two parameters

4.2 Construction of the best trivariate copula for the port of Le Havre

We first determine the most appropriate copula for two parameters: (T, S), (H, T) and then (H, S). We construct the bivariate distribution function using the selected copula for the two most correlated variables. We determine the most relevant copula between the function obtained with the two most correlated variables and the third variable.

4.2.1 Bivariate copula for the three random variables

To determine the best bivariate copula we assess the maximum likelihood between (F_H, F_S), (F_T, F_S) and (F_H, F_T) with the different copulas in Table 5. For all three combinations, the Clayton copula still has the largest maximum likelihood value. In addition, we find that for the combination (H, T) the log-likelihood is significantly higher. As expected, the parameters (H, T) are therefore the most correlated parameters. We can write :

$$F_{H,T} = [(F_H)^{-2.37} + (F_T)^{-2.37} - 1]^{\frac{-1}{2.37}} \tag{43}$$

Copula	Parameter	Parameter	Parameter	Maximum likelihood	Maximum likelihood	Maximum likelihood
	(H,S)	(T,S)	(H,T)	(H,S)	(T,S)	(H,T)
Gumbel	1.29	1.18	1.99	185	82	1059
Survival Gumbel	1.39	1.25	2.37	372	205	1584
Clayton	**0.73**	**0.50**	**2.37**	**387**	**22**	**1565**
Gauss	0.42	0.31	0.77	296	149	1369
Franck	0.67	1.83	7.27	271	139	1333
Student	0.42	0.30	0.77	303	159	1404
Plackett	3.58	2.49	15.64	277	138	1349
Joe	1.26	1.14	2.06	76	26	651
Galambos	0.83	0.61	1.25	175	75	1038

Table 5 : Log-likelihood and copula parameter for the different bivariate copulas between the parameters H and S, T and S then H and T.

4.2.2 Determination of the best trivariate copula

We determine the maximum likelihood between $F_{H,T}$ and F_S with the different copulas in Table 6.

Copula	Parameter	Maximum likelihood
Gumbel	1.25	120
Survival Gumbel	1.29	263
Clayton	**0.56**	**289**
Gauss	0.36	195
Franck	2.08	156
Student	0.35	215
Plackett	2.84	165
Joe	1.72	35
Galambos	0.50	111

Table 6 : Log-likelihood and copula parameter for different bivariate copulas between $F_{H,T}$ *and* F_S.

We obtain the largest log-likelihood for Clayton copula, with a parameter of 0.56, which gives:

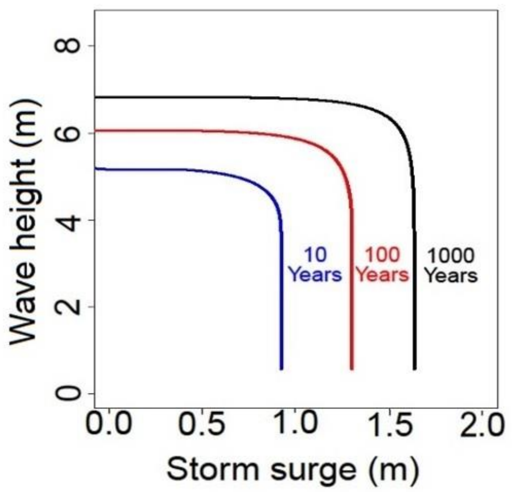
$$F_{H,T,s} = \left[\left(F_{H,T}\right)^{-0.56} + (F_s)^{-0.56} - 1\right]^{\frac{-1}{0.56}} \tag{44}$$

In conclusion, we have thus aggregated the most correlated H and T parameters with the best performing Clayton copula. We also used Clayton copula to aggregate $F_{H,T}$ and F_s. The aggregation requires two different parameters.

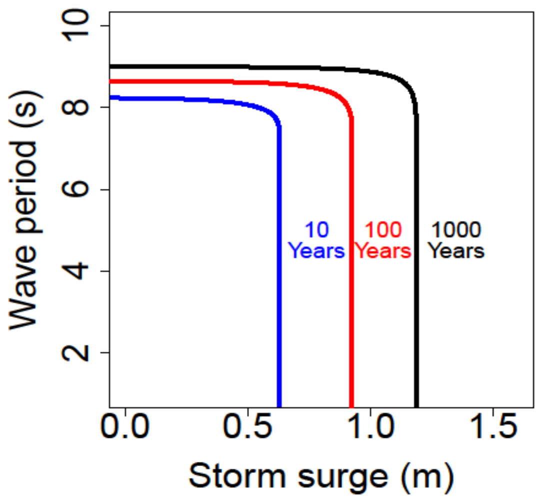
4.3 Contours of equal joint exceedance probability with a trivariate copula

We represent in Figure 9 trivariate joint exceedance probability for return periods of 10, 100 and 1.000 years. The trivariate copula used is therefore constructed from a Clayton copula parameter 2.37 connecting H and T and a copula parameter 0.56 connecting F_{HT} and F_S.

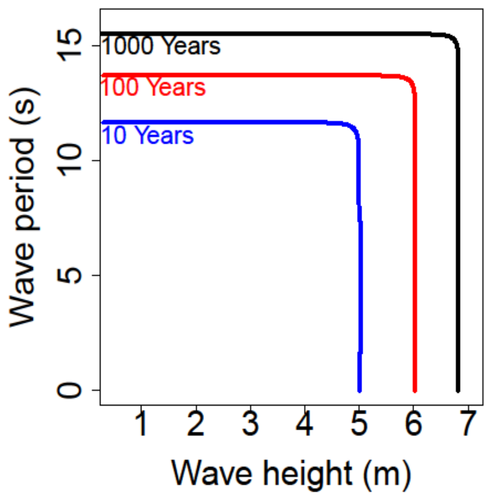
In order to better visualize the incidence of return periods on trivariate joint exceedance probability, cross-sections along (H, T), (H, S) and (T, S) are shown for $T = T_1$, $H = H_1$ and $S = S_1$ in Figures 9a, 9b and 9c respectively.



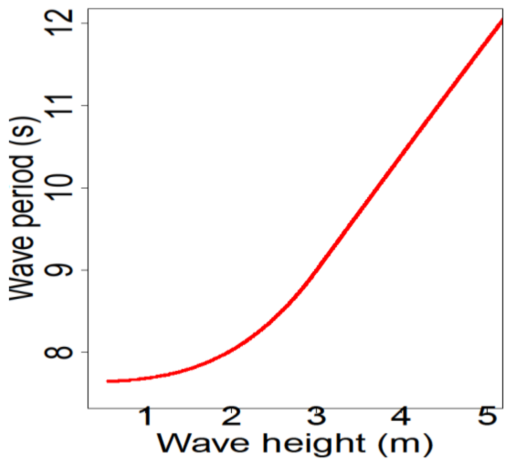
a)



b)



c)



d)

Figure 9 : Contours of equal joint exceedance probability with a trivariate copula.

In Figure 9a, a constant wave period is fixed corresponding to an annual return period. We show the joint exceedance probability of wave height and storm surge for three return periods of 10, 100 and 1.000 years.

In Figure 9b, a constant wave height is fixed corresponding to an annual return period. We show the joint exceedance probability of the storm surge and the wave period for three return periods of 10, 100 and 1.000 years.

In Figure 9c, a constant storm surge is fixed corresponding to an annual return period. We show the joint exceedance probability of the wave height and the wave period for three return periods of 10, 100 and 1.000 years.

In the three latter figures we recognize the usual pattern and the characteristics of a strong correlation for (*H, T*). In Figure 9c we recognize the classic pattern of contours for very dependent variables.

In Figure 9d, a relationship between *H* and *T* is obtained with a trivariate copula with (*H*,*S*) satisfying a joint exceedance probability of 1.000 years and with *T* which maximizes the trivariate joint probability density function. This relationship enables us to obtain the wave period from the wave height and the storm surge.

4.4 Error rate and goodness of fit for trivariate copulas

In order to show the utility of the constructed trivariate copula, we determine the error rate of the different copulas in the Le Havre area using the formula of the error given by equation (1) and the definition of the error rate given by exp(e) – 1 (see Table 7).

Copula	Clayton	Gumbel
$\boldsymbol{C_2(C_1(F_H, F_S), F_T)}$	6.9 %	
$\boldsymbol{C_2(C_1(F_T, F_S), F_H)}$	4.7 %	
$\boldsymbol{C_2(C_1(F_H, F_T), F_S)}$	**3.8 %**	22.2 %
$\boldsymbol{C(F_H, F_S, F_T)}$	8.8 %	169.0 %

Table 7 : Error rate of the different trivariate copulas for the port of Le Havre.

The results obtained by the trivariate copula constructed by two bivariate copulas and two parameters are generally good. However, by aggregating the most correlated variables first, the robustness improves.

As expected, with one parameter Archimedean copula is less robust than fully nested hierarchical copula with two parameters. It can also be seen that by associating the most correlated variables (*H, T*), the Clayton copula gives better results than the Gumbel copula. For a single parameter the trivariate copula constructed with the Clayton copula is significantly more accurate than the Gumbel copula.

Table 7 shows finally that the choice of the copula is much more important than the choice of the trivariate method of construction. This result validates our choice of a simple method of construction that can even lead to the most robust results according to Corbella (2013).

	KHI-2	KS
$\mathbf{C_2(C_1(F_H, F_T), F_S)}$**, Θ1 = 2.37, Θ2 = 0.56**	4.91	0.039
$\mathbf{C(F_H, F_T, F_S)}$**, Θ = 0.56**	5.97	0.098
$\mathbf{C(C(F_H, F_T), F_S)}$**, Θ = 0.56**	5.97	0.098

Table 8 : Goodness of fit of the different trivariate copulas for the port of Le Havre.

The best results are obtained with two parameters. With one parameter Archimedean copula and fully nested hierarchical copula are exactly the same copula as shown in Table 8.

The results highlight the contribution of trivariate copulas constructed as a fully nested hierarchical copula with the help of two Clayton bivariate copulas and two parameters by first aggregating the two most correlated parameters.

5 Conclusion

Wave structure designers must accurately estimate return periods of parameters such as storm surge, wave height and wave period, and more specifically, their joint probabilities of exceedance. In present practice, this joint probability of exceedance is related to the product of univariate probabilities by means of a simple factor. This method can cause damaging design errors. After highlighting the limit of the current simplified Defra method, the theory of copula is introduced. Copulas make it possible to couple the marginal laws in order to obtain a multivariate law.

Analysis of the tail dependence of the sample is used to make an initial selection of the copulas. This is because if the sample has lower tail dependence (upper tail dependence, respectively), the copula with the same tail dependence or an inverse tail dependence is chosen by taking the survival copula. The correlation between the storm surge and wave height is modelled using the Clayton copula and the survival Gumbel copula.

In order to take into account the three variables (wave height, wave period, and storm surge), we show that a fully nested hierarchical trivariate copula with two parameters is the best construction technique. This function satisfies the mathematical properties of the copulas. The error rate of 3.8 % is lower than the trivariate copula obtained by generalizing the Clayton copula with a single parameter (error rate of 8.8 %). We confirm that the best results are obtained by first aggregating the most correlated variables that are here wave height and wave period. Nevertheless, the choice of method of aggregation is much less important than the choice of the copula.

References

Aas, K., Berg, D.: Models for construction of multivariate dependence – a comparison study, The European Journal of Finance, 15, 7-8, 639-659, 2009.

Aghakouchak, A., Bardossy, A., Habib, E.: Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula. Advances in Water Resources, 33(6) :624–634, 2010.

Chakak, A., Koehler, K. J.: A strategy for constructing multivariate distributions. Communications Statistics - Simulation and Computation 24, 537–550, 1995.

Chang, J., Li, Y., Wang, Y., Yuan, M.: Copula-based drought risk assessment combined with an integrated index in the Wei River Basin, China. Journal of Hydrology, 540:824–834, 2016.

Charpentier, A.: Copules et risques multiples, 2014.

Ciria, Cur, Cetmef: Rock Manual, The use of rock in hydraulic engineering (second edition), London, 2007.

Clayton, D. G.: A model for association in bivariate life tables and its application in epidemiological studies of familial tendancy in chronic disease incidence. Biometrika, 65(1), pp 141-151, 1978.

Clauss, P.: Théorie des copules, ENSAI, 2009.

Corbella, S., Stretch, D.: Simulating a multivariate sea storm using Archimedean copulae. Coastal Engineering. 76. 68–78. 10.1016/j.coastaleng.2013.01.011, 2013.

De Michele, C., Salvadori, G.: A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. Journal of Geophysical Research Atmospheres, 108(D2), 2003.

Dodge, D. : Premiers pas en statistiques, Springer, 428 p., 1999.

Durrleman, V.: Les fonctions copules. Définition exemples et propriétés. Cours de l'Ecole Polytechnique, 28 p., 2010.

Genest, C.: De l'impossibilité de construire des lois à marges multidimensionnelles données à partir de copules. Comptes rendus de l'académie des sciences de Paris, 320 : 723-726, 1995.

Favre, A.C., El Adlouni, S., Perreault, L., Thiémonge, N., Bobée B.: Multivariate hydrological frequency analysis using copulas. Water resources research, 40(1), 2004.

Genest C., Favre, A. C.: Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, pages 347–368, 2007.

Grimaldi, S., Serinaldi, F.: Design hyetograph analysis with 3-copula function. Hydrological Sciences Journal, 51(2) pp. 223–238, 2006.

Gouldby, B., Méndez, F.J., Guanche, Y., Rueda, A., Mínguez, R.: A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis. Coastal Engineering, 88, pp.15-26, 2014.

Hawkes, P.J., Gouldby, B.P., Tawn, J.A., Owen, M.W.: The joint probability of waves and water levels in coastal engineering design. Journal of Hydraulic Research 40, 241–251, 2002

Hawkes, P.J.: Use of Joint Probability Methods in Flood Management: A Guide to Best Practice, Defra/Environment Agency and Flood and Coastal Defence R & D Programme, 2005.

Joe, H. : Multivariate models and dependence concepts, 1997.

Kergadallan, X.: Analyse statistique des niveaux d'eau extrêmes. Cetmef, 179 p., 2013.

Kergadallan, X.: Estimation des niveaux marins extrêmes avec et sans l'action des vagues le long du littoral métropolitain (Doctoral dissertation), Université Paris-Est, 2015.

Kumar, P.: Probability Distributions and Estimation of Ali-Mikhail-Haq Copula. Applied Mathematical Sciences, 4, 14, 657 – 666, 2010.

Lee, T., Modarres, R., Ouarda, T. B. M. J.: Data-based analysis of bivariate copula tail dependence for drought duration and severity. Hydrological Processes, 27(10) 1454–1463, 2013.

Li, Y., Simmonds, D., Reeve, D.: Quantifying uncertainty in extreme values of design parameters with resampling techniques. Ocean Eng., 35, 1029-1038, 10.1016/j.oceaneng.2008.02.009, 2008.

Manoukian, E. B.: Guide de statistique appliquée, Herman, 202 p., 1986.

Mazas, F., Hamm, L.: An event-based approach for extreme joint probabilities of waves and sea levels. Coastal Engineering 122, 44–59, 2017

Mazas, F.: Extreme events: a framework for assessing natural hazards. Nat Hazards 98, 823–848, 2019 https://doi.org/10.1016/j.coastaleng.2017.02.003

Nelsen, R. B.: An introduction to Copulae. Springer Series en Statistics, 269 p., 2006.

Ouvrard, J. Y. : Probabilités, Capes et aggregation, Cassini, 247 p., 1998.

Revuz, D.: Probabilités, Herman, 301 p., 1997.

Roncalli, T.: Gestion des risques multiples, cours ENSAI, 219 p., 1997.

Salvadori, G., De Michele, C., Kottengoda, N., Rosso, R.: Extremes in nature. An approach using copulae. Water Science and Technology library. 56, Springer, Dordrecht, 2007.

Serinaldi, F.: Dismissing return periods! Stochastic Environmental Research and Risk Assessment 29, 1179–1189. https://doi.org/10.1007/s00477-014-0916-1, 2015.

Sergent, P., Prevot, G., Mattarolo, G., Brossard, J, Morel, G., Mar, F., Benoit, M., Ropert, F., Kergadallan, X., Trichet, J., Mallet, P.: Adaptation of Coastal Structures to Mean Sea Level Rise. La Houille Blanche, 6, 54-61, 2014.

Simon B.: Statistique des niveaux extrêmes le long de la côte de France. Rapport d'étude n 001/94. Service Hydrographique et Océanographique de la Marine (SHOM), Brest, France, 1994.

Sklar, A.: Fonction de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistiques de l'Université Paris, 229-231, 1959.

Tassi, P.: Méthodes statistiques. Collection Economie et statistiques avancées, 2004.

Zhang, L., Singh, V.P.: Trivariate flood frequency analysis using the Gumbel Hougaard copula. Journal of Hydrologic Engineering, 2007.

Appendices

Appendix A: Outlines of copula theory

A.1 Bivariate cumulative distribution function

We denote by F_X the cumulative distribution function (CDF) of a random variable defined by :

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X\,(y)dy \tag{A.1}$$

where *P* is the probability.

We also introduce the survival function (SF) denoted by $\bar{F}_X$ and defined by :

$$\bar{F}_X(x) = P(X > x) = \int_{x}^{\infty} f_X\,(y)dy = 1 - F_X(x) \tag{A.2}$$

The survival function is related to the probability density function f_X by :

$$f_X(x) = -\frac{d\bar{F}_X\,(x)}{dx} \tag{A.3}$$

Our objective is to obtain the bivariate cumulative distribution function $F_{XY}(x, y) = P(X \le x, Y \le y)$ or the bivariate survival function $\bar{F}_{XY}(x, y) = P(X > x, Y > y)$. For more information, the reader may refer to (Dodge, 1999; Revuz, 1997; Ouvrard, 1998; Manoukian, 1986).

We must model the correlation between, for example, wave heights *H* and storm surges *S* by proposing a relation defining the joint cumulative distribution function from the univariate cumulative distribution functions. We thus seek to obtain a function *C* which links the bivariate cumulative distribution frequency $F_{XY}(x, y)$ to the univariate cumulative distribution frequencies $F_X(x)$ and $F_Y(y)$ by integrating a correlation parameter.

$$F_{XY}(x, y) = C[F_X(x), F_Y(y)] \tag{A.4}$$

A.2 Current practice in coastal engineering

The simplified Defra method that is presented for example in Ciria *et al.* (2007) makes it possible to directly connect the joint probability density function f_{XY} to the product of the univariate probability density functions f_X and f_Y through a dependence factor denoted FD :

$$f_{XY} = \mathrm{FD} f_X f_Y \tag{A.5}$$

The dependence factor FD depends on the correlation coefficient ρ obtained from the Gaussian copula (see definition in section A.3.2). The variables *X* and *Y* for the bivariate analysis are generally wave height *H* and storm surge *S*. The dependence factor is site specific and results from the analysis of the local correlation between wave heights and storm surges.

The correspondence table between the correlation coefficient ρ and the dependence factor FD is given by Kergadallan (2013). This table recommends, for example, for the North Sea, Channel and Atlantic coast the use of a minimum dependence factor FD of 25 that is a weak dependence.

A.3 Copulas

The copula is a statistical tool to characterize the dependence between several random variables where linear correlations are generally not able to represent them accurately (Charpentier, 2014). According to the latter, copulas have become an important tool for modelling a multivariate law that “couples” univariate cumulative distribution functions, hence the Latin name “copula” name chosen by Sklar (1959).

If *C* is the copula associated with a random variable vector (*X*, *Y*) then the copula *C* couples the univariate cumulative distribution functions $F_X(x)$ and $F_Y(y)$ using (A.4).

Survival functions can also be coupled in the sense that there exists a survival copula $\bar{C}$ such that :

$$\bar{F}_{XY}(x,y) = \bar{C}[\bar{F}_X(x), \bar{F}_Y(y)] \quad (A.6)$$

The survival copula $\bar{C}$ is defined from the copula *C*:

$$\bar{C}(\bar{F}_X(x), \bar{F}_Y(y)) = -F_X(x) - F_Y(y) + 1 + C(F_X(x), F_Y(y)) \quad (A.7)$$

In the following description, the univariate cumulative distribution functions $F_X(x)$ and $F_Y(y)$ will be noted u_1 and u_2 respectively. A copula is a function C : $[0,1]^2 \rightarrow [0,1]$ which satisfies the following three conditions :

$$\begin{array}{llr} i) & C(u_1,0) = C(0,u_2) = 0 & \forall u_1,u_2 \in [0,1] \\ ii) & C(u_1,1) = u_1 \text{ and } C(1,u_2) = u_2 & \forall u_1,u_2 \in [0,1] \\ iii) & C(v_1,v_2) + C(u_1,u_2) - C(u_1,v_2) - C(v_1,u_2) \geq 0 & \forall 0 \leq u_i \leq v_i \leq 1 \end{array} \quad (A.8)$$

In the continuation of the paragraph on the description of the copula the functions of distribution $F_X(x)$ and $F_Y(y)$ will be noted $u_1$and u_2.

Sklar (1959) states that there exists a copula *C* such that for each *x* and *y* $F_{XY}(x,y) = C[F_X(x), F_Y(y)]$. If the functions F_X and F_Y are continuous then *C* is unique. There exist four families: Archimedeans, Elliptics, Marshall-Olkin and Archimax.

A.3.1 Archimedean copulas

Archimedean copulas are defined as follows : ϕ is a decreasing function convex on [0,1]→[0,+ ∞[, as ϕ (1) = 0 and ϕ (0) = ∞. We call a strict Archimedean copula of generator ϕ the copula defined by equation (9) :

$$C(u_1,u_2) = \phi^{-1}[\phi(u_1) + \phi(u_2)], u_1,u_2 \in [0,1] \quad (A.9)$$

Archimedean copulas have interesting properties, in particular the possibility of aggregating more than two variables by equation (10) :

$$C(u_1,u_2,\ldots,u_n) = \phi^{-1}[\phi(u_1) + \phi(u_2) + \ldots + \phi(u_n)], u_1,u_2,\ldots,u_n \in [0,1] \quad (A.10)$$

Archimedean copulas are given in table A1.

Name	Copula	Generator	Inverse generator
Clayton (θ > 0)	$[u_1^{-\theta} + u_2^{-\theta} - 1]^{-1/\theta}$	$\frac{t^{-\theta} - 1}{\theta}$	$(1 + \theta t)^{-1/\theta}$
Franck (θ ≠ 0)	$\frac{1}{\theta} ln\left(\frac{u_1u_2}{[1-\theta(1-u_1)(1-u_2)]}\right)$	$-\ln\left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}\right)$	$\frac{ln(1 + \exp(-t)(\exp(-\theta) - 1))}{\theta}$
Gumbel (θ ≥ 1)	$exp[-(u_1^{\theta} + u_2^{\theta})^{1/\theta}]$	$(-\ln(t))^{\theta}$	$\exp(-t^{1/\theta})$
Independence	u_1u_2	$-\ln(t)$	$\exp(-t)$
Joe (θ ≥ 1)	$1 - [(1-u_1)^{\theta} + (1-u_2)^{\theta} - (1-u_1)^{\theta}(1-u_2)^{\theta})^{\frac{1}{\theta}}]$	$-\ln(1 - (1-t)^{\theta})$	$1 - (1 - \exp(-t))^{1/\theta}$
Ali-Mikhail-Haq (-1 ≤ θ ≤ 1)	$\frac{u_1u_2}{[1-\theta(1-u_1)(1-u_2)]}$	$\ln\left(\frac{1-\theta(1-t)}{t}\right)$	$\frac{1-\theta}{\exp(t) - \theta}$

Table A1: Archimedean copulas

A.3.2 Elliptic copulas

Elliptic copulas are Gaussian and Student's copulas:

The Gaussian copula is written as follows :

$$C(u_1,u_2) = \frac{1}{2\pi\sqrt{1-\theta^2}}\int_{-\infty}^{\phi^{-1}(u_1)}\int_{-\infty}^{\phi^{-1}(u_2)}\frac{1}{2\pi(1-\theta^2)^{0.5}}\exp\left(\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}\right)\mathrm{dxdy}, \theta \in [-1,+1] \qquad (A.11)$$

ϕ is a distribution function of X_i, with X = $(X_1, X_2, \ldots, X_n)$ a Gaussian random vector (X$\sim N_\nu\,(0,\sum)$), where $\sum$ is a covariance matrix.

Student's copula is written as follows :

$$C(u_1,u_2) = \int_{-\infty}^{t_\nu^{-1}(u_1)}\int_{-\infty}^{t_\nu^{-1}(u_2)}\frac{1}{2\pi(1-\theta^2)^{0.5}}\left[1+\frac{s^2-2\theta st+t^2}{2(1-\theta^2)}\right]^{\frac{-(\nu+2)}{2}} dsdt, \theta \in [-1,+1] \qquad (A.12)$$

t_ν is a distribution function of the univariate Student distribution law with ν degrees of freedom.

They are symmetrical copulas. They are widely used in finance. They are implicit and therefore do not have an explicit analytical form.

A.3.3 Marshall-Olkin's copula

Marshall-Olkin's copula is written as follows :

$$C(u_1,u_2) = \min({u_1}^a u_2, u_1{u_2}^b), (a,b) \in [0,1] \qquad (A.13)$$

A.3.4 Archimax copulas

Archimax copulas include a large number of copulas, including Archimedean copulas.

A bivariate function is an Archimax copula if and only if it is of the form :

$$C_{\phi,A}(u_1,u_2) = \phi^{-1}\left[(\phi(u_1)+\phi(u_2))\mathrm{A}\left(\frac{\phi(u_1)}{\phi(u_1)+\phi(u_2)}\right)\right], \forall\, u_1,u_2 \in [0,1]^2 \qquad (A.14)$$

A : [0,1] → [0.5,1] such as max(t,1-t) ≤ A(t) ≤ 1 for each t $0 \le t \le 1$.

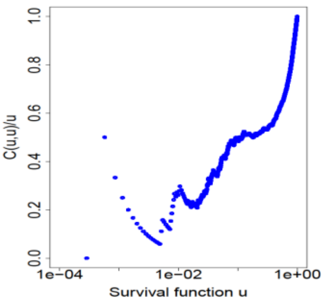
ϕ :]0,1[→ [0,+∞[is a convex, decreasing function that satisfies $\phi(1) = 0$.

We will adopt the following notation $\phi(0) = \lim_{u->0}\phi\,(t)\ et\ \phi^{-1}(s) = 0$, for s $\geq \phi\,(0)$.

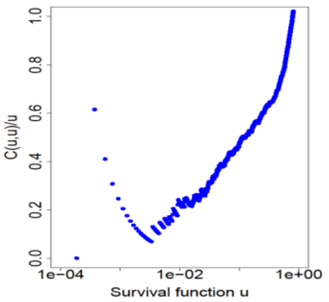
For more information, refer to reference books such as Joe (1997) and Nelsen (1999). The reader may also refer to Clayton (1978).

Appendix B : Tail dependence of the site

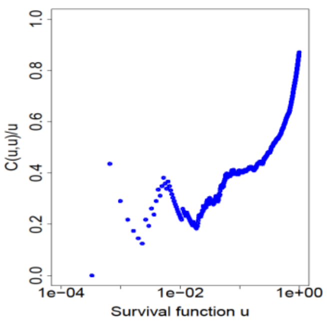
Dunkerque



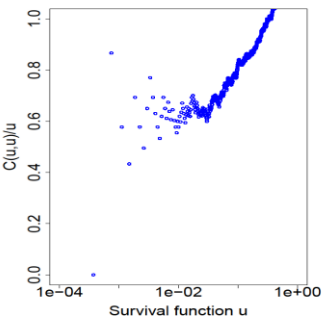
Calais



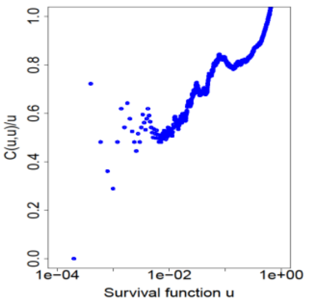
Boulogne



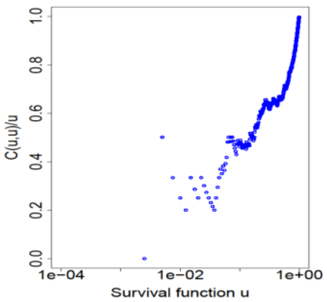
Dieppe



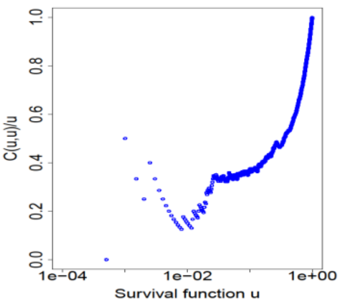
Le Havre



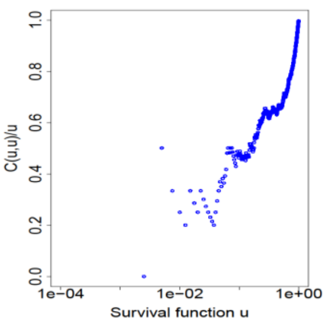
Cherbourg



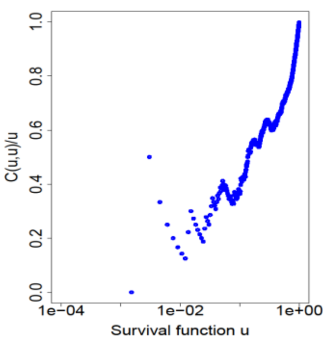
Saint-Malo



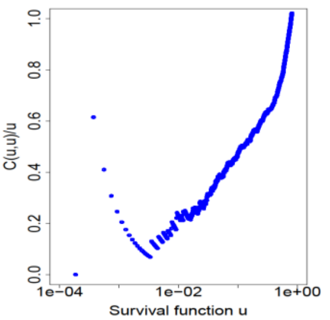
Roscoff



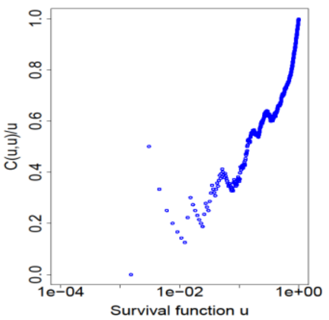
Le Conquet



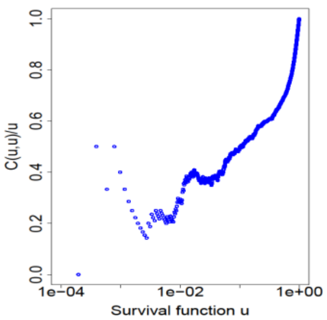
Brest



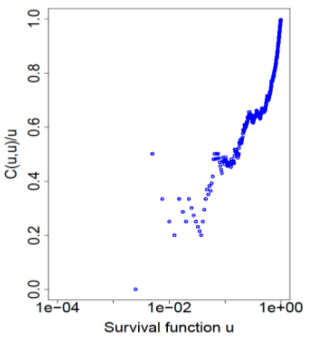
Concarneau



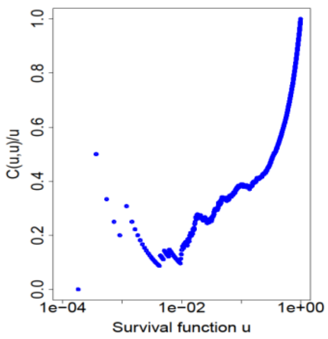
Port Tudy



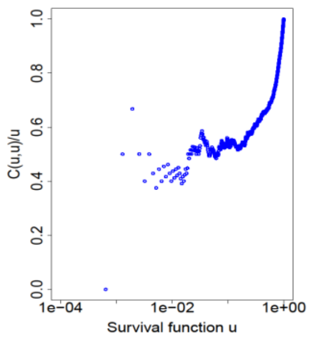
Saint-Nazaire



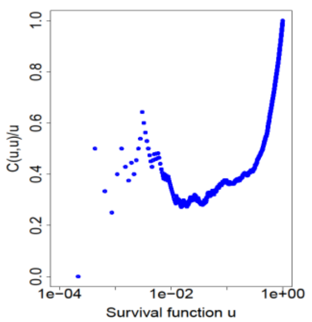
Saint-Gildas



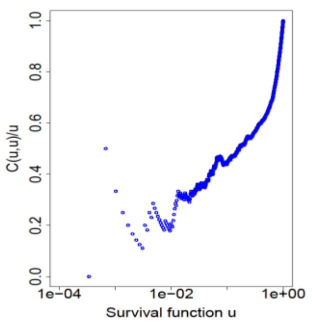
La Rochelle



Bayonne



Socoa



Port Bloc

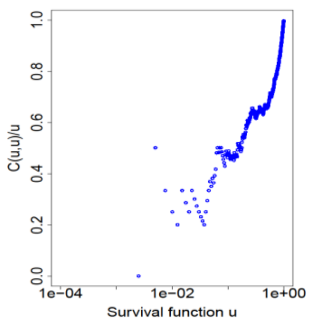


Table B1: Tail dependence of 18 French sites

Appendix C : Likelihood for 18 French sites

Sites	Gumbel	Clayton	Gauss	Franck	Student	Plackett	Joe	AMH	Glambos
Dunkerque	111	**387**	244	214	264	226	38	368	125
Calais	90	**242**	177	172	179	172	23	233	85
Boulogne	174	**393**	287	273	300	279	64	387	164
Dieppe	166	**383**	274	257	286	261	61	379	157
Le Havre	352	**901**	594	551	632	572	117	897	329
Cherbourg	140	**383**	267	224	277	229	44	317	135
Saint Malo	33	**134**	79	65	83	67	5	102	32
Roscoff	92	**273**	178	159	188	164	26	229	81
Le Conquet	160	**389**	28	265	293	268	54	365	150
Brest	178	**439**	322	295	327	299	59	417	168
Concarneau	66	**115**	97	96	98	94	31	117	64
Port Tudy	391	899	653	627	665	635	139	**909**	369
St Nazaire	438	1001	728	713	745	710	159	**1009**	522
Saint Gildas	282	726	492	471	509	479	87	**737**	265
La Rochelle	107	**303**	197	186	199	184	30	**303**	100
Bayonne	75	**275**	153	111	179	116	19	162	67
Soccoa	62	**230**	122	105	155	110	15	163	51
Port Bloc	31	**69**	47	50	52	53	12	69	28.8

Table C1: Likelihood for 18 French sites