

Correlation models in flood risk analysis

F.L.M. Diermanse

Deltares, Delft, The Netherlands

C.P.M. Geerse

HKV_{LIN IN WATER}, Lelystad, The Netherlands

ABSTRACT: Statistical dependence among random hydraulic variables generally stem from a common meteorological cause. Usually, this dependence increases the probability of occurrence of floods. Therefore, in many (potential) applications in flood risk analysis there is a need for techniques that properly describe the statistical dependence among random variables. This study makes an inventory of existing bivariate correlation models as available from international literature and/or ‘Dutch practice’ in the design and testing of flood defence structures. Differences in correlation structures are described and quantified. Subsequently, their practical applicability is tested in two case studies.

1 INTRODUCTION

Statistical dependence among random variables representing hydraulic processes generally stem from a common meteorological cause. This dependence increases the probability of occurrence of extreme combinations of threats that may cause floods. Therefore, in many (potential) applications in flood risk analysis there is a need for techniques that properly describe the statistical dependence among random variables. A problem in this respect is that a multivariate distribution function that is derived directly from multivariate observation data usually conflicts with the previously established univariate distribution functions of the individual variables. This effect is especially profound in the tail ends of distribution functions, which are so crucial in risk analysis.

This paper describes a number of methods that incorporate statistical dependence of random variables in probabilistic risk analyses. These methods have the advantage that the derived bivariate distribution functions do not conflict with the univariate distribution functions of the individual random variables. The application of these techniques is demonstrated for two case studies.

2 DUTCH PRACTICE IN TESTING FLOOD DEFENCES

Most techniques that are described in this paper have been applied in the models that form the technical basis of the procedure to test flood defences in the Netherlands. Therefore, a brief description

of this ‘Dutch practice’ in testing flood defences is given.

Without adequate flood defence structures, almost 60% of the Netherlands would be regularly flooded. To prevent floods by the sea, the large rivers and lakes, an extensive system of dikes and dunes has been constructed in the past. In the design and testing of dikes and dunes a safety level is defined of the order of 10^{-3} to 10^{-4} “failures” per year. This frequency is based on both the economic value of the protected area and the extent of the threat.

To test whether the flood defences fulfil the required safety level, representative hydraulic conditions are derived. For this purpose a number of probabilistic models have been developed over the past decades. Different models have been developed for different water systems, such as Hydra-B for the tidal area of the Rhine-Meuse delta, Hydra-K for the coastal areas and Hydra-VIJ for the deltas of the rivers Vecht and IJssel. Furthermore, the model PC-Ring has been developed, which considers entire ring dikes and therefore integrates the various water systems. Currently, PC-Ring still has the status of “research tool”, whereas the Hydra-models are “officially approved” models.

3 MATHEMATICAL DESCRIPTION OF BIVARIATE CORRELATION MODELS

3.1 Introduction

For a mathematical description of bivariate correlation models in this paper, two random variables V and W are considered, with known univariate

distribution functions $F_V(v)$ and $F_W(w)$. The correlation models describe the mutual dependence of these variables, resulting in a bivariate distribution function $F_{V,W}(v, w) = P(V \leq v, W \leq w)$. It is noted that some models in this paper can be extended to the multivariate case with 3 or more variables, but this is not always the case (at least: it does not always lead to sensible models). The latter is of little concern, though, since the quantification of statistical dependence is generally limited to two variables in most practical applications in flood risk analysis.

To apply the models in this paper, a dataset (v_i, w_i) , $i = 1, 2, \dots, n$ is assumed to be available, for which a proper correlation model needs to be derived.

3.2 Statistical transformations

The correlation models in this paper transform random variables into variables with “convenient” distribution functions such as the standard normal, standard uniform or standard exponential distribution function. The correlation structure is then described in the transformed space and subsequently transformed back to the real space. This section describes basic transformations and the most relevant associated features, as obtained e.g. from Ditlevsen & Madsen (1996).

Consider two random variables V and W with given univariate cumulative distribution functions (CDF’s) $F_V(v)$ and $F_W(w)$. The variables V and W are transformed to dimensionless random variables X and Y as follows:

$$\begin{aligned} F_V(v) &= F_X(x) \\ F_W(w) &= F_Y(y) \end{aligned} \quad (1)$$

where F_X and F_Y are univariate cumulative distribution functions of the transformed variables. Equation (1) can be rewritten as:

$$\begin{aligned} x &= F_X^{-1}(F_V(v)) = J(v) \\ y &= F_Y^{-1}(F_W(w)) = K(w) \end{aligned} \quad (2)$$

where function names J and K have been introduced for convenience. This transformation means that for a given value v^* , the transformed value $x^* = J(v^*)$ has the same probability of (non-)exceedance as v^* . The major benefit of this transformation is that a wide variety of correlation models can be applied on the transformed variables x and y , while at the same time the univariate distribution functions of the real variables V and W are preserved.

To derive the correlation model, the observed data (v_i, w_i) also needs to be transformed, into $(x_i, y_i) = (J(v_i), K(w_i))$. This transformation will

distort the cloud of data-points to some extent and thus has some effect on the observed correlation structure of the data. For instance, the value of Pearson’s correlation coefficient in the transformed space (x, y) may differ from its corresponding value in the real space (v, w) . However, these differences are generally small. Furthermore, specific features of the correlation structure, such as “correlation increases for decreasing probability of exceedance”, are generally preserved. The equations below describe the relation between the correlation structures in the real and transformed space.

Define $F_{X,Y}(x,y)$ as the bivariate CDF for transformed variables x and y , and $f_{X,Y}(x,y)$ is the associated PDF. The corresponding functions $F_{V,W}(v, w)$ and $f_{V,W}(v,w)$ in the real space can be derived from $F_{X,Y}$ and $f_{X,Y}$ by applying the following two equations:

$$F_{V,W}(v, w) = F_{X,Y}(J(v), K(w)) \quad (3)$$

$$\begin{aligned} f_{V,W}(v, w) &= f_{X,Y}(J(v), K(w)) \\ &\times \frac{dJ(v)}{dv} \frac{dK(w)}{dw} \end{aligned} \quad (4)$$

where the latter follows by differentiation of Equation (3) with respect to v and w . Other relevant information that can be obtained are the conditional CDF and PDF:

$$F_{W|V}(w | v) = F_{Y|X}(K(w) | J(v)) \quad (5)$$

$$f_{W|V}(w | v) = f_{Y|X}(K(w) | J(v)) \frac{dK(w)}{dw} \quad (6)$$

3.3 Copulas

3.3.1 General

An effective and relatively new approach to incorporate specific correlation structures are the copula functions or copulas (see e.g. Kallenberg 2009, Kole et al, 2006, Panchenko, 2005). In Dutch flood risk analysis, copulas have been applied in a probabilistic model to derive statistics of the spatially averaged water level of the IJssel Lake (Diermanse & Van der Klis, 2005).

Consider two random variables V and W with univariate distribution functions F_V and F_W and bivariate distribution function $F_{V,W}(v,w)$. A function $C: [0, 1]^2 \rightarrow [0, 1]$, is called a *copula function* of $F_{V,W}$ if:

$$C(F_V(v), F_W(w)) = F_{V,W}(v, w) \quad (7)$$

For example, if V and W are statistically independent, the copula function C is as follows:

$$C(F_V(v), F_W(w)) = F_V(v) F_W(w) \quad (8)$$

This function is referred to as the *product Copula*. Another example is the Gaussian copula. In 2 dimensions, this copula is as follows:

$$C(x, y) = \int_{-\infty}^{\Phi^{-1}(x)} \int_{-\infty}^{\Phi^{-1}(y)} \frac{\exp\left\{-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right\}}{2\pi(1-\rho^2)^{0.5}} ds dt \quad (9)$$

where ρ is the correlation coefficient and Φ is the standard normal distribution function.

According to Sklar's theorem (Sklar, 1959), the bivariate distribution function $F_{V,W}$ has *exactly* one copula function C if the univariate distribution functions F_V and F_W are both continuous. Equation (7) shows that a bivariate distribution function can easily be obtained if the copula-function and the univariate distribution functions are known.

3.3.2 Archimedean copulas

A well-known and widely used type of copula-functions are the Archimedean copulas. This group of copulas is related to a generator function $\varphi: [0, 1] \rightarrow [0, \infty]$. This function should be continuous, strictly decreasing and convex. Furthermore, $\varphi(0) = \infty$ and $\varphi(1) = 0$. For two

standard uniformly distributed random variables x and y the following function is an Archimedean copula:

$$C_\varphi(x, y) = \varphi^{-1}[\varphi(x) + \varphi(y)] \quad (10)$$

Equations (11)–(13) shows generator functions of three Archimedean copulas:

$$\text{Frank: } \varphi(t) = -\ln\left[\frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}\right], \quad \alpha \in (-\infty, \infty) \setminus \{0\} \quad (11)$$

$$\text{Clayton: } \varphi(t) = \frac{1}{\alpha}(t^{-\alpha} - 1), \quad \alpha \in (-1, \infty) \setminus \{0\} \quad (12)$$

$$\text{Gumbel: } \varphi(t) = (-\ln t)^\alpha, \quad \alpha \in (1, \infty) \quad (13)$$

The choice of parameter α determines the overall correlation for each Archimedean copula.

Figure 1 shows scatter-plots of four copulas: Gumbel, Frank, Clayton and Gaussian. In these Figures the correlation coefficient is taken to be equal to 0.7 by selecting the proper value of α and ρ in equations (9), (11), (12) and (13). A close look at Figure 1 reveals the specific features of these four copulas. Application of the Gumbel copula results in high correlation for values of x and y close to 1.

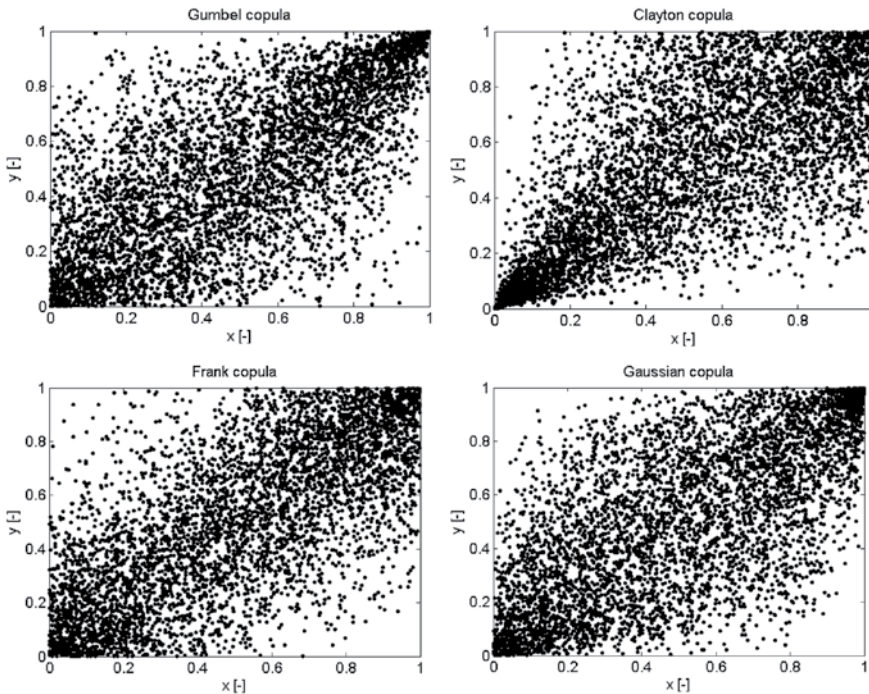


Figure 1. 5.000 samples of 2 standard uniformly distributed random variables with four different copula structures. In all cases the correlation coefficient is taken equal to 0.7.

Since x and y are standard uniformly distributed, this implies the correlation increases in the right tails of the univariate distribution functions. In other words, if a sample of the first random variable is relatively large, the accompanying sample of the second variable most likely is relatively large as well. Clayton's copula on the other hand generates high correlations in the left tails of the univariate CDFs. In that respect Frank's copula and the Gaussian copula are much more symmetrical, where the latter shows stronger correlation in the extremes (both left and right).

3.4 Homoscedastic model (HOS)

The name *homoscedastic* refers to the fact that the variance of the conditional distribution of y , given x , is constant. This as opposed to the heteroscedastic model where the variance varies with x , see section 3.5. This model (Geerse, 2004) is applied in the probabilistic model Hydra-VIJ (Section 2) to describe the correlation between the discharge of the river IJssel and the spatially averaged water level of the IJssel lake into which the river discharges. Furthermore, it is used in the probabilistic model Hydra-B to describe the correlation between sea water level and wind speed. In this model, transformed variable x is standard exponentially distributed:

$$F(x) = 1 - e^{-x} \tag{14}$$

Variable y is related to x as follows:

$$F_Y(y) = \int_0^{\infty} e^{-x} \Lambda_{\sigma}(y - x - \delta) dx \tag{15}$$

where δ is a constant and Λ_{σ} is a CDF with mean 0 and standard deviation σ . Figure 2 presents a schematic view of this correlation model. It shows the conditional CDF of y has a constant variance around the line $y = x - \delta$. Small values of σ are associated with high correlation and vice versa.

The fact that y is described as a function of x has a relevant practical implication with regard to the use of transformations. Section 3.2 describes how observations of variables V and W are transformed to variables X and Y , using the four respective univariate distribution functions (see Equation (1)). Subsequently, the correlation structure is derived from the transformed observations. In the HOS model, however, the transformation can only be applied once the correlation structure is defined, because otherwise the univariate distribution function of variable Y is not available (see Equation (15)). In practice this means the best combination of δ and Λ_{σ} is

determined to some extent by “trial-and-error”, using the observed correlation structure of variables V and W as a valuable starting point. A useful criterion for goodness-of-fit is to verify whether the confidence intervals of the assumed correlation model are in accordance with the observations. In section 4, some examples of this comparison are presented.

The conditional distribution of y , given x , is:

$$f(y | x) = \lambda_{\sigma}(y - x - \delta) \tag{16}$$

where λ_{σ} is the PDF that is associated with Λ_{σ} . In this model, variable y is not standard exponentially distributed like x . However, asymptotically (if $y \rightarrow \infty$) variable y is exponentially distributed. With the appropriate selection for constant δ , y is even *standard* exponentially distributed in the limit. For instance, if Λ_{σ} is a Gaussian distribution, δ should be taken equal to:

$$\delta = -\frac{\sigma^2}{2} \tag{17}$$

Figure 3 illustrates that, with this choice of parameter δ , y is indeed standard exponentially distributed

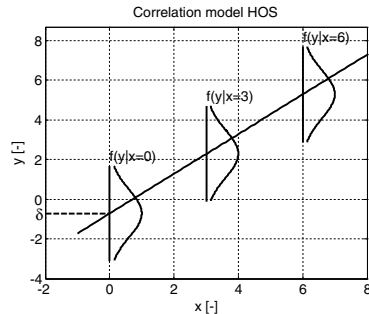


Figure 2. Schematic view of the HOS model in the transformed space.

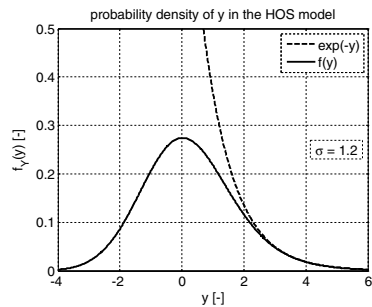


Figure 3. Probability density function of y in the HOS model, compared with the standard exponential density function.

distributed in the tail. We note that the choice of δ will have no influence at all on the resulting bivariate distribution $F_{V,W}(v,w)$ in the real space. In that sense it is a redundant parameter and serves only to obtain the convenient property of y being asymptotically standard exponentially distributed in the transformed space.

3.5 Heteroscedastic model (HES)

In the heteroscedastic model (HES) the variance of variable y is dependent on x . This model, described in Geerse & Diermanse (2006), is very similar to the model of the previous section. In this case, however, σ is a function of variable x :

$$F_Y(y) = \int_0^{\infty} e^{-x} \Lambda_{\sigma(x)}(y-x-\delta) dx \quad (18)$$

Function $\sigma(x)$ can be any positive function, which makes this model very flexible. With this model an increase (or decrease) of the correlation with increasing values of x (and, consequently, v), can be simulated (e.g. see Figure 4).

In this model, there is no choice of parameter δ that will result in variable y being asymptotically standard exponentially distributed. Since the choice of this parameter also has no effect on the resulting bivariate CDF $F_{V,W}(v,w)$ in the real space, it is set equal 0.

3.6 De Haan method (DH)

The correlation model ‘‘De Haan’’ (DH) is applied in the probabilistic model Hydra-K (see Section 2) to describe the correlation between wind speed and sea water level. It differs from the previously described methods since it does not use a parametric description of the correlation

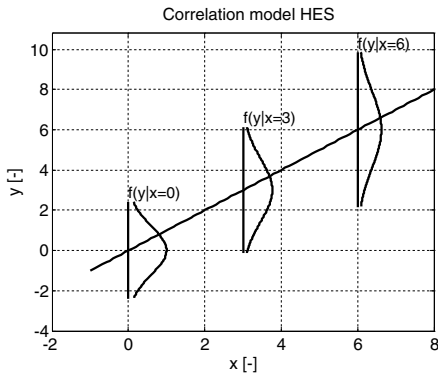


Figure 4. Schematic view of the HES model in the transformed space.

structure. Instead, it translates the observations to more extreme conditions and as such assumes the observed correlation structure is applicable for the entire domain of possible extreme events.

The theoretical foundation of this method is introduced in De Haan & Resnick (1977) and further developed for the Dutch part of the North Sea by De Valk (1996). In this model, paired observations $(v_i, w_i)_{i=1..n}$ of variables V and W are translated to the standard exponential space:

$$\begin{aligned} x_i &= -\ln[1 - F_V(v_i)]; \quad i = 1, 2, \dots, n \\ y_i &= -\ln[1 - F_W(w_i)]; \quad i = 1, 2, \dots, n \end{aligned} \quad (19)$$

Where n is the number of observations. The variables V and W are assumed to be *asymptotically dependent*. In the transformed (x,y) -space this means:

$$\lim_{u \rightarrow \infty} P(Y > u | X > u) = c > 0 \quad (20)$$

This implies a relatively strong correlation for extreme events.

In the standard exponential space, data is shifted along a line of 45°. This means for a user defined value of λ , the following translation is applied:

$$\begin{aligned} x_i^* &= x_i + \lambda, \quad i = 1, 2, \dots, n \\ y_i^* &= x_i + \lambda, \quad i = 1, 2, \dots, n \end{aligned} \quad (21)$$

The value of λ is chosen such that the observations are shifted towards the area of interest. In flood risk analysis, this means λ should be chosen such that for a substantial percentage of the shifted pairs of observations failure of the flood defence occurs. This is schematically depicted in Figure 5, where the closed circles are the observations and the open circles are the shifted observations. This Figure also shows the limit state of the flood defence, i.e. the transition line between ‘failure’ and non-failure of the flood defence. Combinations of x and y to the upper right of the limit state will lead to failure, so this area is called the *failure domain*. The position of the limit state results from a comparison of the hydraulic load and the resistance of the flood defence. For this paper, the limit state and failure domain are assumed to be known. This shift of observations can be used, provided Equation (20) holds, to obtain an estimate of the frequency of failure of the flood defence:

$$\mu = \frac{\kappa(\lambda)}{T} e^{-\lambda} \quad (22)$$

where μ is the frequency of failure (per year), $\kappa(\lambda)$ is the number of shifted observations in the

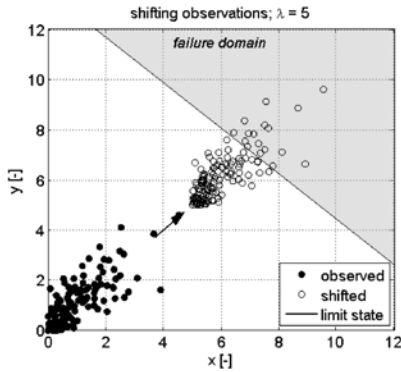


Figure 5. Shifting observations in the standard exponential space along the line of 45 degrees.

failure domain, T is the length of the period of observations (in years) and λ is as defined before. The value of λ should be chosen in such a way that the value of μ is rather insensitive to small changes in λ . As long as Equation (20) holds, it is possible to find such a value of λ (De Haan & Resnick, 1977).

4 CASE STUDIES

The application of the correlation models is demonstrated for two case studies. The first case study is about the correlation between the discharge of the IJssel river and the level of the IJssel lake, into which the IJssel river discharges. The second case study is about the correlation between wind speed and water level of the Dutch North Sea. A comparison is made with respect to whether the correlation models are capable of reproducing the observed correlation structures. Furthermore, the influence of the choice of correlation model on computed dike failures is quantified.

4.1 Case study IJssel river and IJssel lake

With a total surface of 1,182 km², the IJssel Lake is the largest lake in the Netherlands. The IJssel river drains at an average rate of approximately 400 m³/s into the lake. The lake is separated from the sea by a dike (the “Afsluitdijk”). High water levels in the IJssel Lake are the result of extended periods during which the discharge of the IJssel river exceeds the outflow through the sluices of the Afsluitdijk into the sea. As a result, peak discharges of the IJssel river and peak levels of the IJssel lake are correlated. This is relevant for dikes along the IJssel delta, because increased lake levels will lead to increased water levels in the river due to backwater effects.

Figure 6 shows observed combinations of IJssel river peak discharges and IJssel lake peak levels in the transformed space. The 3 lines show the 10%,

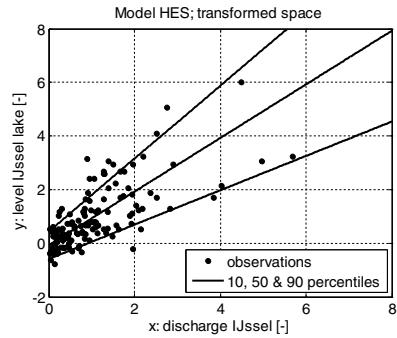


Figure 6. Observed combinations of IJssel river peak discharges and IJssel lake peak level in the transformed space. The 3 lines show the 10%, 50% and 90% exceedance lines for the HES model.

50% and 90% exceedance lines for the HES model (section 3.5) with $\sigma(x) = 0.4 + 0.28x$, where x is the transformed value of the discharge.

Of all the models described in chapter 3 the HES model turned out to be best in accordance with the observed data. The HOS model, for instance, did not perform as well as the HES model because the data shows a clear increase in variance with increasing discharge, which cannot be incorporated in the HOS model. The Frank copula and Clayton copula are inappropriate here, because for these copulas the correlation for higher values is too low. The Gaussian copula is much more capable of reproducing the correlation in the extremes, especially if a high value is chosen for the correlation coefficient ρ (see Equation (9)). However, a high correlation coefficient also results in a relatively narrow margin between the 10% and 90% quantiles which causes far less than the required 80% of the observations to be located inside of this band. For the Gumbel copula, a lower value of ρ is sufficient to model the correlation of the extremes, leading to a better representation of the correlation of the lower values of x and y in comparison with the Gaussian copula. The DH method is considered unsuitable in this case, because the increasing variance for increasing values of x indicates that discharge and lake level are not asymptotically dependent.

For a quantitative comparison of the correlation models, two hypothetical failure domains are defined as depicted in Figure 7. Clearly, these domains represent events that are far more extreme than the observed events. This is typical in Dutch practice in designing and testing flood defences, where safety standards in the order of 10^{-3} to 10^{-4} failures per year are applied. Table 1 shows the selected model parameters and resulting probabilities of failure for the two failure domains. The standard deviation σ in the models HOS and HES was chosen in such a way that it best fitted the data, while Λ_σ

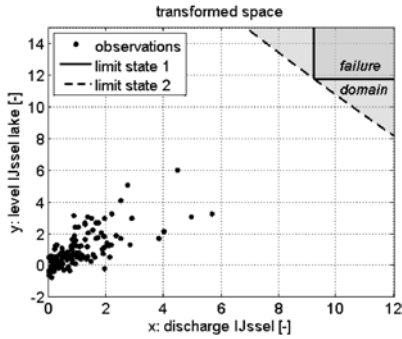


Figure 7. Observed combinations of IJssel river peak discharges and IJssel lake peak level in the transformed space and 2 hypothetical failure domains.

Table 1. Probability of failure for the two failure domains of Figure 7.

Model	Parameters	Failure dom. 1	Failure dom. 2
		10^{-4} per year	10^{-4} per year
HOS	$\sigma = 1.2$	0.54	0.82
HES	$\sigma = 0.4 + 0.28x$	0.31	0.63
DH	–	0.78	1.00
Gauss	$\rho = 0.7$	0.10	0.32
Gumbel	$\rho = 0.7$	0.63	0.91
Frank	$\rho = 0.7$	0.00	0.00
Clayton	$\rho = 0.7$	0.00	0.00

was taken to be a normal distribution function. For the copulas a correlation coefficient of 0.7 was selected, which was derived directly from the data.

Application of the DH method gives the highest probability of failure. This is no surprise, since this model assumes the variables are asymptotically dependent, which means a very strong correlation in the extremes. Based on the correlation structure of the observations, this property is considered unrealistic for the two variables in this case study. The Gumbel copula also assumes strong correlation in the extremes, which is why it generates the second highest failure probabilities. The Frank and Clayton copula, on the other hand, assume a weak correlation in the extremes, which means a combined occurrence of extreme values of the two variables is far more unlikely than for the other models. This is why the resulting probability of failure is orders of magnitudes lower for these two copulas than for the other models. The lack of correlation in the extremes in these two models is considered unrealistic in this case study.

Visual inspection yields that the best fit to the data is provided by the HES model. The HOS, DH and Gumbel models did not perform as well

as the HES model, since they seem to overestimate the correlation in the extreme while the Gauss model seems to underestimate the correlation in the extreme. Frank and Clayton are not appropriate at all because of the lack of correlation in the extremes. Note that the HOS, DH and Gumbel models, though not performing very well, only yield differences in failure probabilities of a factor 2.5 or less in comparison with the HES model. Compared to other sources of uncertainties in the probabilistic safety calculations for dikes, an error of such a factor is relatively small. The conclusion, at least for this case study, is that it is relevant to take the correlation between discharge and lake level into account, but that the exact choice of the model and model parameter values are not crucial, as long as the model is reasonably well in accordance with the observations.

4.2 Case study North Sea

The most crucial flood defences in the Netherlands are the dunes and dikes that protect the land from floods of the North Sea and connected estuaries. Potential extreme loads for these defences are a combination of high water levels and waves. Both are generated by the wind, which causes high water levels and high waves to be correlated. The probabilistic model Hydra-K (Section 2) uses the DH method to describe the correlation between water level and wind speed. Wave characteristics are subsequently derived from the water level and wind speed, using the wave simulation model SWAN (Booij et al, 1999).

Figure 8 shows observed combinations of sea water level and wind speed at location Hoek van Holland for wind direction 330 degrees (in the transformed space). The 3 lines show the 10%, 50% and 90% exceedance lines for the HOS model with $\sigma = 1.6$. It shows that the HOS model in this case is in accordance with the observations. This automatically means the same can be said about the HES model (since HOS is a special case of HES). Method DH is also more plausible in this case than in the first case study since the data does not show an increase in variance with increasing value of x . Similar to the previous case, Gumbel showed the best fit of the 4 copulas. Frank and Clayton proved to be inadequate because, again, the lack of correlation in the extremes in these two models is unrealistic.

For a quantitative comparison of the correlation models, two hypothetical failure domains are defined. These failure domains are very similar to the ones shown in Figure 7. Table 2 shows the resulting failure probabilities. The maximum difference between the models that are reasonably in accordance with the data (HOS, HES DH and Gumbel) is a factor 2, comparable to the previous

Table 2. Probability of failure for case 2.

Model	Parameters	Failure dom. 1	Failure dom. 2
		10 ⁻⁴ per year	10 ⁻⁴ per year
HOS	$\sigma = 1.6$	0.43	0.74
HES	$\sigma = 1.8 - 0.09x$	0.57	0.83
DH	–	0.70	1.00
Gauss	$\rho = 0.6$	0.05	0.19
Gumbel	$\rho = 0.6$	0.36	0.63
Frank	$\rho = 0.6$	0.00	0.01
Clayton	$\rho = 0.6$	0.00	0.01

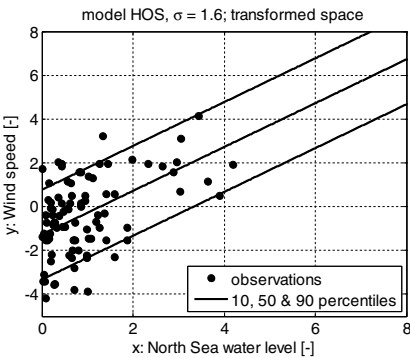


Figure 8. Observed combinations of sea water level and wind speed at location Hoek van Holland for wind direction 330 degrees. The 3 lines show the 10%, 50% and 90% exceedance lines for the HOS model.

case study. So, again, in terms of design and testing of dikes this difference is limited.

5 CONCLUSIONS

A wide variety of correlation structures can be described using the models presented in this paper. The models are very straightforward to apply and have the advantage that the resulting bivariate distribution functions do not conflict with the (previously established) univariate distribution functions of the individual random variables.

In probabilistic applications in hydraulics, generally there is a significant statistical dependence in the extremes. Application of correlation structures that underestimate this dependence structure may lead to a severe underestimation of the probability of failure of a flood defence structure of several orders of magnitude. For the case studies in this paper, this was the case for the Frank and Clayton copula and to a lesser extent also for the Gaussian copula.

The method of De Haan has the advantage that it is “non-parametric” i.e. no model parameters need to be estimated from the observed data. However, it can only be applied to describe

combinations of variables that are asymptotically dependent. This property defines such a strong correlation structure in the extremes, that it generally will not hold in practical applications in hydraulics.

The heteroscedastic (HES) model is the most “flexible” of the models presented here. With this model, a wide variety of correlation structures can be described, due to the variance function that can take any form as long as it is positive.

The results of the case studies showed the importance of taking the correlation structure of the variables under consideration into account in a probabilistic analysis of failure probabilities of flood defences. Application of a badly chosen correlation model might lead to errors in the failure probabilities of several orders of magnitudes. However, the results of the case studies also suggest that the exact choice of the correlation model and associated parameter values seems not very crucial, with regard to the resulting probability of failure, as long as the model is reasonably in accordance with the observed data.

REFERENCES

- Booij, N., Ris, R.C. & Holthuijsen, L.H. 1999: A third generation wave model for coastal regions, Part I, Model description and validation. *J. Geophys. Res.*, 104 (C4), 7649–7666.
- De Haan, L. & Resnick, S.I. 1977: Limit theory for multivariate sample extremes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 40, 317–337.
- De Valk, C.F. 1996: Estimation of the frequencies of failure of coastal structures from offshore data of environmental loads, *Delft Hydraulics report*, H2240.
- Diermanse, F.L.M. & Van der Klis, H. 2005: Modelling statistical dependence using copulas; *Proceedings of the International Symposium on Stochastic Hydraulics*, IAHR, Paseo Bajo Virgen del Puerto 3, 28005 Madrid, Spain, ISBN: 90-805649-9-0.
- Ditlevsen, O. & Madsen, H.O. 1996: *Structural reliability analysis*. O. Ditlevsen, H.O. Madsen. John Wiley & Sons, Chichester, England 1996.
- Geerse, C.P.M. 2004: Bivariate correlatiemodellen met exponentiële en asymptotisch exponentiële marginale verdelingen, *Rijkswaterstaat-RIZA, report*. 2002.104x, may 2004 (in Dutch).
- Geerse, C.P.M. & Diermanse, F.L.M. 2006: Correlaties en meerdimensionale statistiek, *HKV LIJN IN WATER en WL | Delft Hydraulics rapport*, november 2006 (in Dutch).
- Kallenberg, W.C.M. 2009: Estimating copula densities, using model selection techniques, *Insurance: Mathematics and Economics* 45 (2009) 209–223.
- Kole, E., Koedijk, K. & Verbeek, M. 2006: Selecting copulas for risk management, *Journal of Banking & Finance* 31 (2007) 2405–2423.
- Panchenko, V. 2005: Goodness-of-fit test for copulas, *Physica A* 355 (2005) 176–182.
- Sklar, A. 1959: Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite' de Paris* 8, 229–231 (in French).