Sensitivity analysis of erosion on the landward slope of an earthen flood defence submitted to wave overtoppings

Clément Lutringer[1,2], Adrien Poupardin[1], Philippe Sergent[3], Abdelkrim Bennabi[1], and Jena Jeong[1,2]

[1]ESTP Paris, 28 Avenue du Président Wilson, 94230 Cachan, France
[2]Gustave Eiffel University, Cité Descartes, 77200 Marne-la-Vallée, France
[3]CEREMA Risques Eau Mer, 134 rue de Beauvais, 60280 Margny-les-Compiègne, France

Correspondence: Clément Lutringer (clutringer@estp-paris.eu)

Abstract. The study aims to provide a complete analysis framework applied to an earthen dyke located in Camargue, France. This dyke is regularly submitted to erosion on the landward slope that needs to be repaired. Improving the resilience of the dyke calls for a reliable model of damage frequency. The developed system is a combination of copula theory, empirical wave propagation and overtopping equations as well as a global sensitivity analysis in order to provide the return period of erosional damage on a set dyke while also providing recommendations in order for the dyke to be reinforced as well the model to be self-improved. The results give a good correspondence, within uncertainty range, between the model prediction of return periods and the on-site observation ($\approx$ two-year return period). The mean of the return periods is slightly higher with an average return period of six years but the peak of the distribution is located around the two years mark. The sensitivity analysis shows that the geometrical characteristics of the dyke - slope angles and dyke height - are the ones carrying the highest amount of uncertainty into the system, showing that maintaining a homogeneous dyke is of great importance. Some empirical parameters intervening inside the propagation and overtopping process are also fairly uncertain and suggest that using more robust methods at their corresponding steps could improve the reliability of the framework. The obtained return periods have been confirmed by current *in situ* observations but the uncertainty increases for the most severe events due to the lack of long-term data.

1 Introduction

The site of the Salin-de-Giraud located in the Camargue area in southern France is an historically low-lying region and is thus frequently exposed to numerous storms. Due to the climate change, this region is among the most endangered by the future rise in sea level and the increase in storm intensity described in the latest Intergovernmental Panel on Climate Change report (Pörtner et al., 2022). In fact, all the infrastructures on the site as well as the land itself must be maintained in order to ensure its exploitation in the future. An earthen dyke, named Quenin, has been constructed on the site in order to protect the salt marshes during storm surges. The structure is quiet large, covering a few kilometers along the coastline. The dyke is approximately 2 meters high with large rocks on the seaward slope while the landward slope is only covered by sand (picture in appendix A). The erosion problem of the dyke is common in this area and therefore assessment of erosion is necessary.

The semi-empirical approach based on the hydraulic loading has been well established and traditionally used. Wave propaga-
tion from deep water to the surf-zone has been well explored both analytically, numerically and experimentally in the literature.
A large overview of theory surrounding random sea wave propagation theory was provided by Goda (2000) and brought ad-
vices on coastal protection. An evaluation of the different methods available on the subject has also been given by Liu and
Han (2017). Overtopping appears to be more complex and relies mainly up to now to experiments from which empirical laws
are deducted, as from van der Meer (2011) as well as Hughes and Nadal (2008); Hughes et al. (2012) with use of the Wave
Overtopping Simulator. Numerical simulations have also been explored by Li et al. (2003) using the Volum of Fluid method.
More recently, the EurOtop manual (van der Meer et al., 2018) laid an extensive set of recommendations and experimentally
based equations in order to functionnally model the overtopping phenomenon .Bergeijk et al. (2019) also provided a more
refined analytical model of overtopping using a set of coupled equations validated by numerical simulations and experiments.
Regarding the statistical tool to predict a higher risk, Copula theory has been well accepted and used to calculate multivari-
ate return periods of natural hazards.De Michele et al. (2007); Bernardara et al. (2014) wrote extensively on the subject with
guidelines on using copulas to predict sea storms. More specifically, Kole et al. (2007) found that the Student's and Gumbel
copulas are particularly interesting for risk management applications. Liu and Han (2017) deemed that the Clayton and Gumbel
Copulas are to be preferred for calculating multivariate joint return periods of natural hazards. Bivariate copulas combining
wave height and sea elevation are the main method in use, as seen in Salvadori and Michele (2007) but Orcel et al. (2020)
expanded the method to trivariate copulas, allowing the method to yield the probability of structural failure.

As indicated by many sources, we have a large choice of different copulas to link our different deep water conditions (Durante
and Sempi, 2010, 2016; Tootoonchi et al., 2022) the survival Gumbel copula would be one of the best candidates to estimate
the return periods of the defined events when it comes to prediction (Kumar and Guloksuz, 2021). As mentioned by Orcel et al.
(2020), this will lead to the calculation of an "and" return period, denoted RP, yielding the conjunction of multiple events
(as opposed to the occurrence of one or more of the events). However, there are very few researches on assessment of erosion
of dyke combining statistical and probability approach and theoretical and semi-empirical approach as well. Mehrabani and
Chen (2015) worked on joint probabilistic approach for assessment of climate change effect on hydraulic loading. However,
the authors constrained themselves to the frame of copula theory, assessing the risk to offshore conditions. That approach
has not considered an interactions with a dyke nor propagation of deep water wave, but used a physical erosion criteria to
put a threshold metric. In the present study, we used global sensitivity analysis to assess the most important parameters in
the framework as the ones that contribute the most to the variance of the system in order to provide self-improvement to the
framework as well as recommendations to improve the resiliency of the dyke. Combining different approaches, sensitivity
analysis, a fully functional and modular overtopping framework and copula theory into a full stack has not been explored
before and might provide use for the practitioner. We start by recalling the theoretical foundations of our study, regarding
copula theory, propagation and overtopping equations as well as the process of global sensitivity analysis in the first section.
The second section is dedicated to the structure of our data and its associated preprocessing in order to render them exploitable.
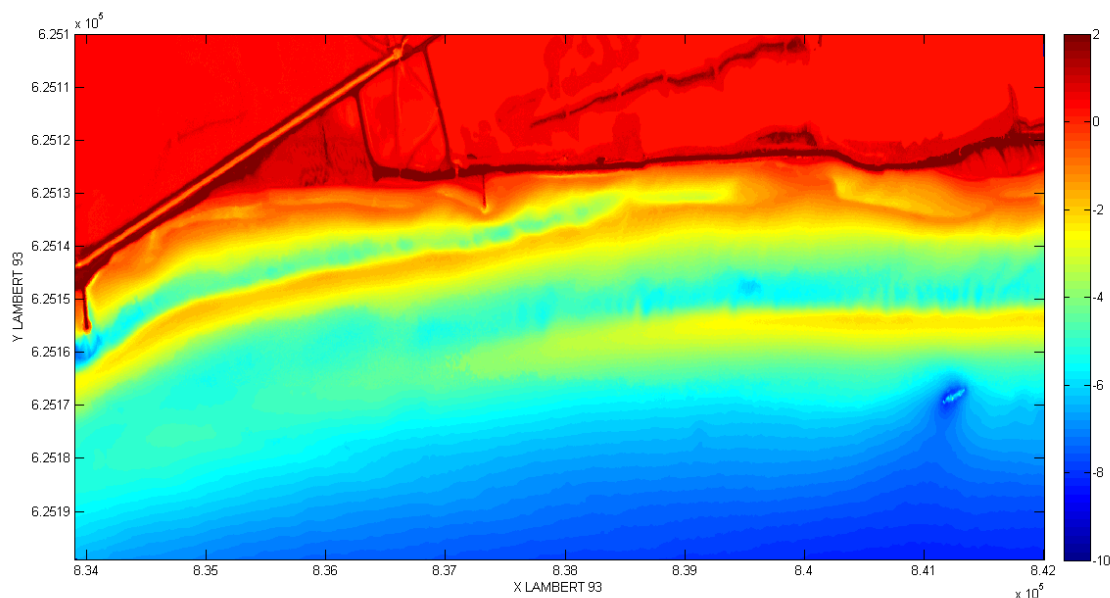
Figure 1. 2D bathymetry map of the Salin-de-Giraud area.

Finally, the third section provides the results of the study, the return periods obtained with details of the sensitivity analysis and our deduced recommendations extracted from the conclusions of the framework in order to improve the resilience of the dyke.

2 Meteocean Data Preprocessing

2.1 Data Description

The framework is supposed to take as inputs at least three sets of data :

- An accurate measure of the 2D bathymetry grid (as seen in Figure 1) is important as it provides the user a useful approximation of the average steepness of the seabed between the open sea and the toe of dyke. It also allows the calculation of coefficient such as γ_b, the contribution of the berm. If such measure is not available, an estimation of the mean seabed steepness can be provided instead.

- The significant wave height, noted H_0.

- The water level N.

We present here the data sets that are going to be used to illustrate the framework. We start by showcasing the univariate data distribution of the significant wave height. Then, we present the univariate data distribution of the water level. Finally, we
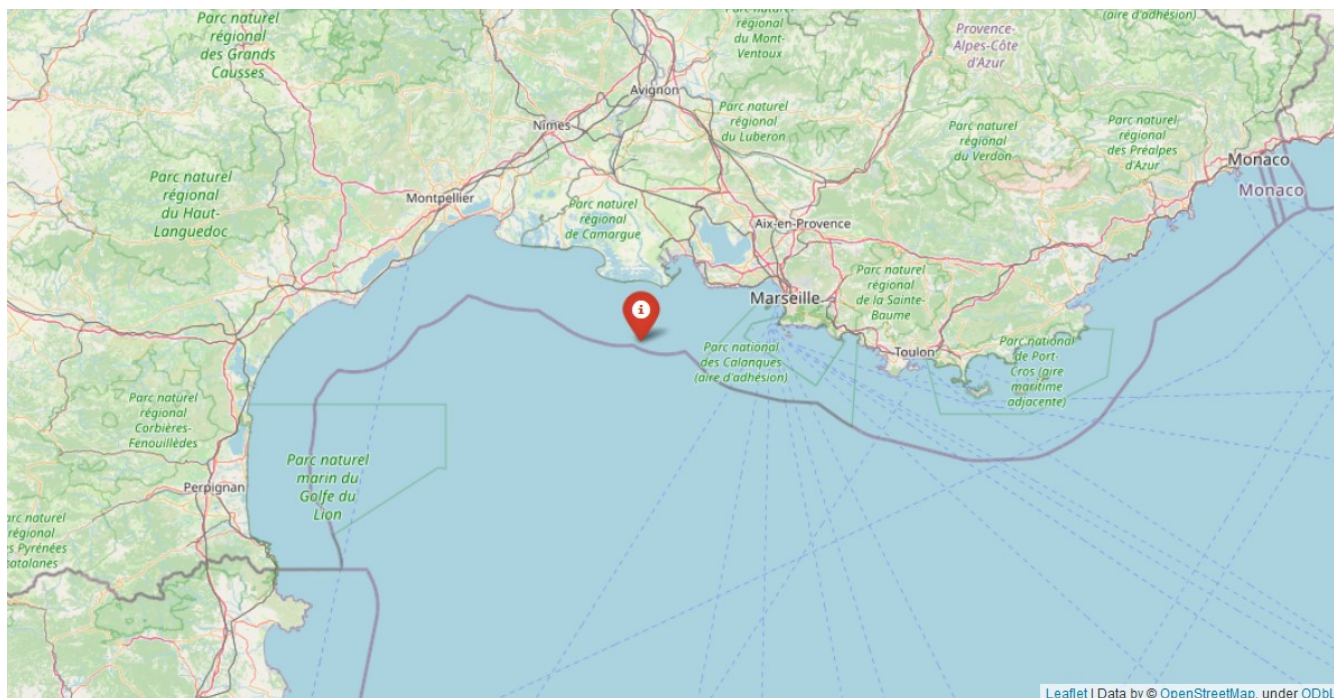
Figure 2. Location of the point where the significant wave height is extracted from.

introduce the methods that are used to generate the joint probability distribution from the univariate distributions using copula theory.

2.2 Univariate Wave Height Distribution

The significant wave height is essential and is provided here using the ANEMOC-2 database currently maintained by the CEREMA[1], reproducing numerically the sea conditions over a long period. the significant wave height is estimated by calculating the mean value of the upper third of the recorded waves every hour. Thus, one value is given hourly at a single location. A data point has been chosen at the location of interest (see Figure 2). We have at our disposal around 30 years of data using this source.

The data can be represented as a time series (Figure 3, bottom) to visualize its evolution.

---

[1]https://www.cerema.fr/fr

2.3 Univariate Water Level Distribution

The water level, noted N, is then extracted from the *in situ* data provided by the Marseille port's sea gauge (REFMAR database) maintained by SHOM [2]. This location is quite far off the actual place of interest but it has the advantage of being located inside a port, providing protection from sea waves, and has many years of extracted data to use (around 30 years).

The data can be represented as a time series (Figure 3, top) to visualize its evolution.

2.4 Bivariate Joint Distribution

2.4.1 Peak selection

The copula that we want to generate is based on the identification of extreme events implying a locally high value of both N and H_0 that we will call here "storms". We use the same protocol as in Kergadallan (2015) which is :

- Search for high peak values V_i on data set A;

- Associate each value V_i with its time of occurrence t_i;

- Define a time window Δt which would be the expected mean duration of a storm;

- For each peak, look for the maximum value W_i of data set B during the $[t_i - \Delta t/2; t_i + \Delta t/2]$;

- create the couple (V_i, W_i) as the characteristics for storm i.

Using the data sets described in part 3.1, we generated the full sample and selected the meaningful events (the storms) using the described protocol. The representation of the selection process is given in Figures 3 and 4. Note that since our protocol is getting rid of the redundancy of certain peaks on the data sets as they are considered a part of the same events, not all extreme events are retained as their information is contained inside a single meaningful point.

The data can also be represented as couples of values showcasing the selected values that we will be using as significant data points (Figure 4). Note that the storm events (in blue) are meant to be representative of a whole event in order to avoid redundancy. The aim is not to select purely high values.

This means that the events can then be sorted into an histogram for us to observe their respective distributions. In this case, the sample limits us to events that can happen up to once every 20 years since we have no data covering a larger period. In this case, we can obtain information about more extreme events by extrapolating the data using a fitted distribution. The Generalized Extreme Value distribution is particularly adapted for this kind of problem where cumulative distribution function is formulated as :

$$[H]F(x) = \exp(t(x)) \qquad \text{with} \qquad t = \begin{cases} \left(1 + \xi * \left(\frac{x - \mu*}{\sigma*}\right)\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ \exp\left(-\left(\frac{x - \mu*}{\sigma*}\right)\right) & \text{if } \xi = 0 \end{cases} \tag{1}$$

---

[2]data.shom.fr
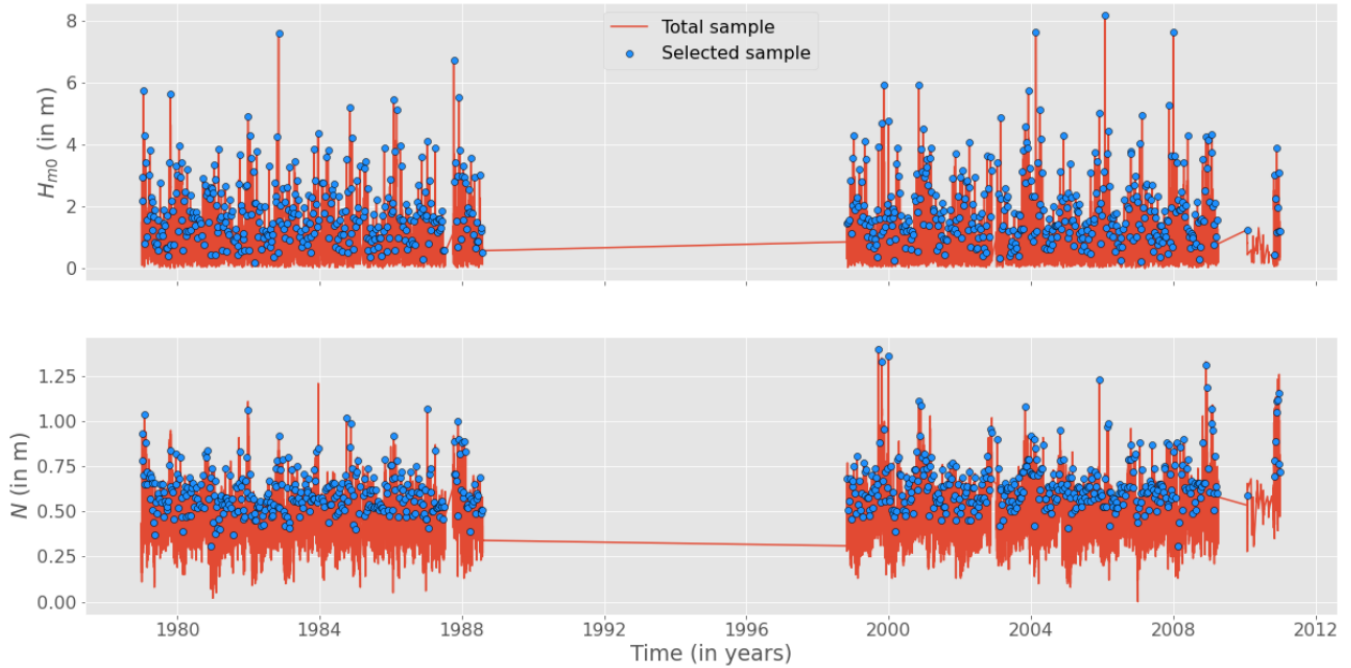
Figure 3. Temporal representation of the offshore significant wave height (bottom) and the still water level (top) as well as the selected events.

with (μ, σ, ξ) the location, scale and shape factor, respectively. The results are displayed in Figure 5. The laws are fitted using the maximum-likelihood method.

2.4.2 Copula Generation

110   Copula theory has been first introduced by applied mathematician Abe Sklar who developped the eponym Sklar's theorem which is the foundation of copulas (Sklar, 1959; Durante et al., 2013). Copulas have been since used widely in quantitative finance as portfolio-diversification recommendation tools and more recently in extreme natural events prediction as a mean of risk management. It appears that the practitioner has a large selection of copulas to choose depending of the nature of the data. According to Brunner et al. (2016), the Gumbel copula is particularly adapted in our case as follows:

115
$$F(u, v \,|\, \theta) = \exp\left[- \left[(-\log(u))^{\theta} + (-\log(v))^{\theta} \right]^{1/\theta} \right] \tag{2}$$

where u and v are the cumulative distribution functions of the histograms originated from the data sets. The copula parameter θ represents the interdependency of the data.

  The value of this copula parameter is important, and can be calculated using a panel of different methods, ie. the Error method (see Appendix B for the equation as written by (Capel, 2020)) and Maximum-Likelihood method.
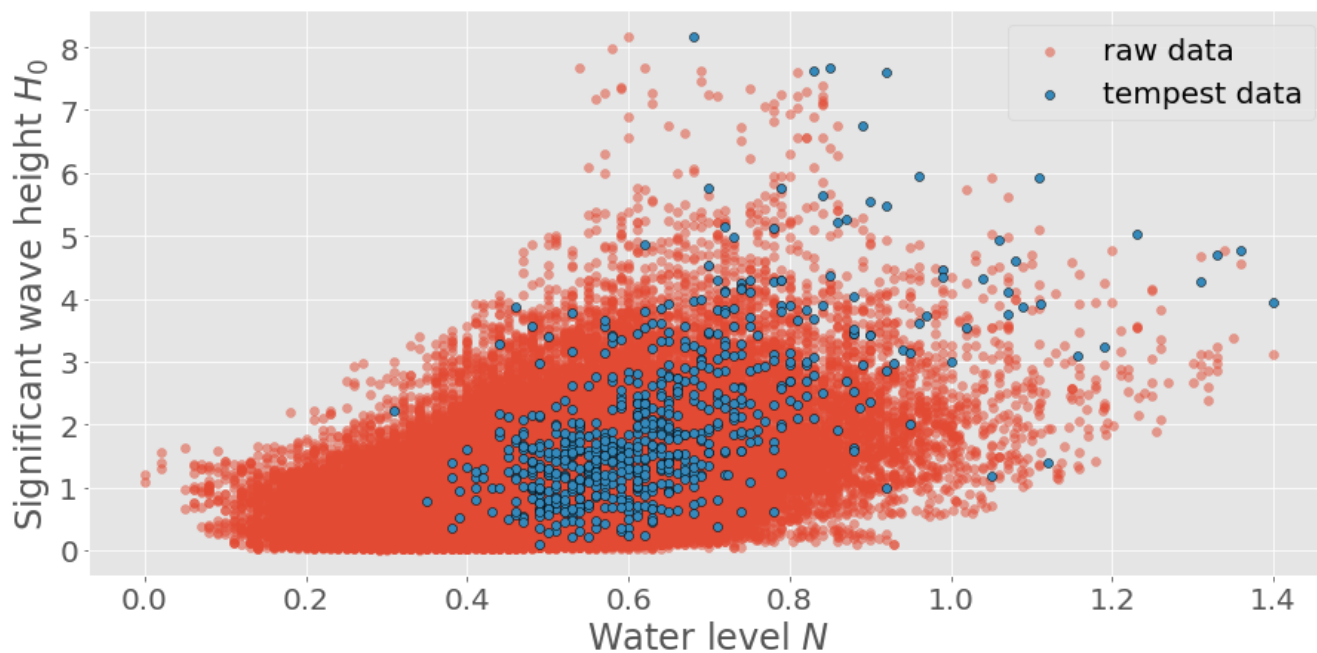
Figure 4. Representation of the selected events over the full sample. We can see that not all extreme events are selected as they are redundant and yielded by other selected events.
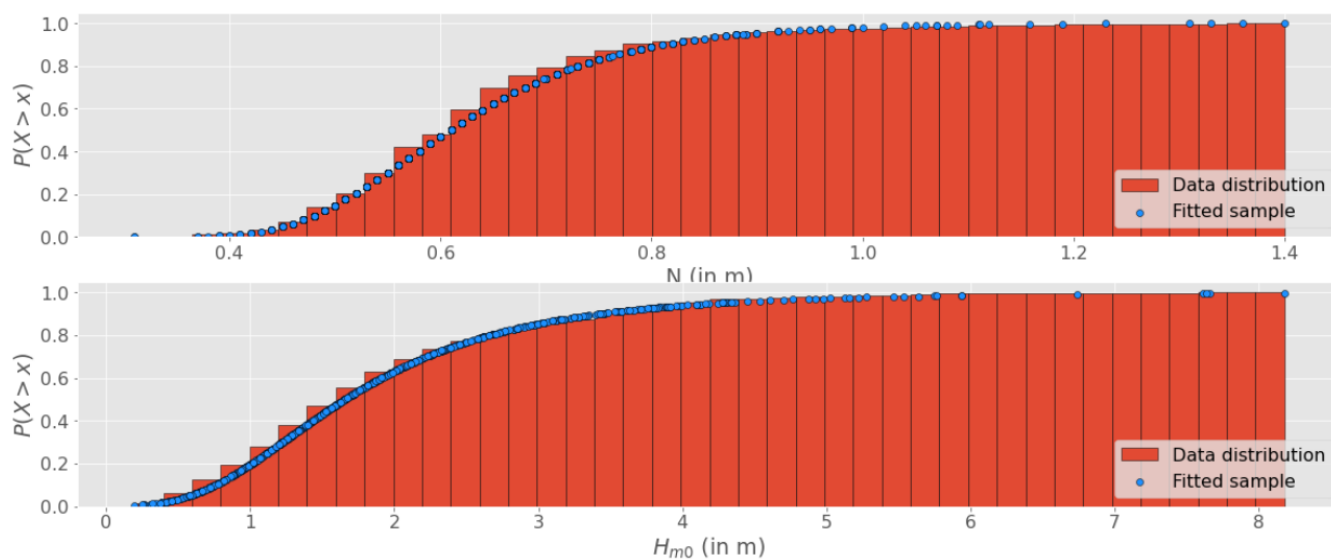
Figure 5. Cumulative Distribution Functions of the offshore significant wave height (bottom) and the still water level (top) as well as the selected events with their respective fitted functions.

120   In fact, the parameter θ yields the degree of correlation between the two variables linked through the copula. Its value is
estimated by minimizing the function $L(\theta)$ defined as follows:

$$L(\theta) = \sum_{i=0}^{n} c_\theta(u(i), v(i)) \tag{3}$$

where c_θ is the copula density, which can be obtained by calculating the derivative of the copula function with respect to its
CDFs in equation 4:

125    $$c_\theta(u,v) = \frac{\partial^2 C_\theta(u,v)}{\partial u \partial v} \tag{4}$$

Once done, the copula can be calculated using equation 2, attributing a probability of occurrence of any event E with one of
the variables having a value smaller of equal to the defined ones, noted $E(H \leq h || N \leq n)$. The logical inverse $E(H > h, N >$
130    $n)$ can then be obtained by calculating the survival copula C_θ^{-1} defined as :

$$C_\theta^{-1} = C_\theta + u + v - 1 \tag{5}$$

Finally, we can associate to each value of C_θ^{-1} a return period using the formula provided in (Salvadori and Michele, 2007)
135   :

$$RP = \frac{\mu}{C_\theta^{-1}} \tag{6}$$

where μ is the average interarrival time between two events of interest i.e., the storms. The offshore conditions have been
determined by a couple (N, H), the water level and the significant wave height respectively) with an associated return period.
This gives us the properties of an offshore wave.

140   The principle of the maximum-likelihood method that we use is that we try to maximize a function (equation 3) yielding
the likelihood of generating the observed data for a set value of θ. The results are shown is Figure 6, showing that a value of
$\theta \approx 1.6$ is optimal, which we will use thereafter.

The interdependency parameter can take values in the interval $[1, +\infty[$, where 1 is the independent copula and $+\infty$ means
absolute correlation. A value of 1.6 means that there is a mild but significant correlation between the two distributions. Hence,
145   we can generate the copula using equation 2. The cumulative distribution function yields the probability of a value laying under
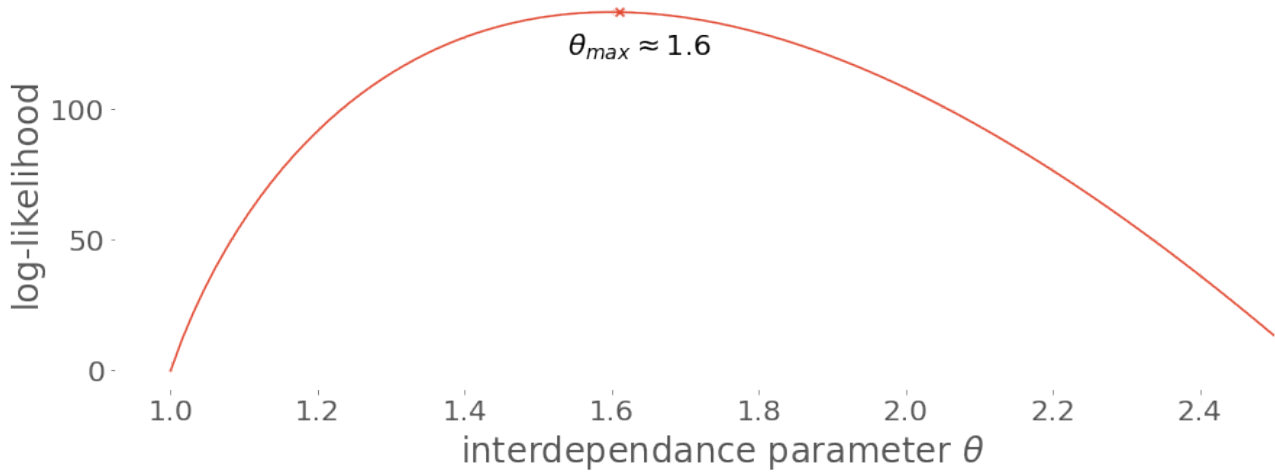
Figure 6. Value of the maximum-likelihood estimator with respect to the interdependence parameter θ.
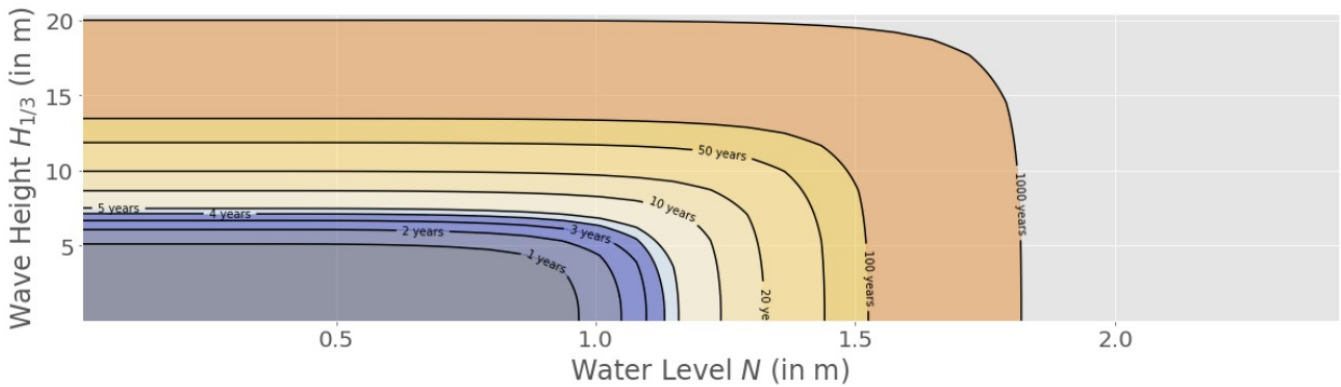
Figure 7. Return period of an event composed of a couple (N, H_0) or with a higher value of N or H_0, noted as $RP(E(N > x || H_0 > y))$ with the interdependence factor having a value of $\theta \approx 1.6$.

a threshold. Hence, we use equation 5 to inverse the copula and obtain the survival copula (Figure 7). This allows us to evaluate the return period of any event E so that $E(N = x || H_0 = y)$.

The contour lines of the the copula in Figure 7 show that the data are coupled to some degree. Indeed, since the data are correlated, a both high value of the water level N and the significant wave height H should be more probable than if the data were uncorrelated, thus decreasing the return period and driving the contour lines towards the higher values.

3 Landward Slope Flow

We are able to link a deep water state to a return period. However, this does not give us any information on the probability of occurrence of an event that would provoke erosion. Hence, we need to assess what kind of event provokes erosion using equations 7 - 12 to calculate the terminal velocity of the flow on the landward slope.

155 ### 3.1 Propagation

The offshore significant wave height can be propagated up to the toe of the dyke. Among the numerous methods, the most convenient to use is the propagation formula written in equation 7 extracted from (Goda, 2000) allowing us to calculate the significant wave height $H_{1/3}$, the mean of the third of the highest wave height over a set period of time, as follows. Note that refraction is neglected in this case :

160
$$H_{1/3} = \begin{cases} K_s H_0 & \text{for } \frac{d}{L_0} > 0.2 \\ \min\left[\beta_0 H_0 + \beta_1 d; \beta_{max} H_0; K_s H_0\right] & \text{for } \frac{d}{L_0} < 0.2 \end{cases} \tag{7}$$

where the coefficients β_0, β_1 and β_{max} can be calculated as detailed in Goda (2000)[3] :

3.2 Overtopping Equations

Once the wave reaches the toe of the dyke, the wave will start interacting with the dyke in what is called the overtopping phase. This phenomenon is divided into 3 steps with equations detailed in (van der Meer et al., 2018). We give a brief summary here

165 of the used equations :

– **Run-up** : The wave reaches the dyke and flows up towards the crest. The run-up height reached by 2% of the incoming waves is calculated using equation 8

$$RU_{2\%} = \gamma_f \cdot \gamma_\beta \cdot \left(4 - \frac{1.5}{\sqrt{\gamma_b \cdot \xi}}\right) \cdot H \tag{8}$$

where ξ is the Irribarren Number, H the wave height at the toe of the dyke, we will use $H_{1/3}$ instead. The γ factors γ_b,

170 γ_f and γ_β yield the contribution of the berm, the roughness and porosity of the seaward slope and the obliquity of the waves, respectively

– **Crest flow** : The water flows on the crest up to the landward slope. We calculate the flow velocity and thickness as the beginning of the crest using equations 9 and 10, respectively:

$$v_{A,2\%} = c_{v2\%} \left(g(RU_{2\%} - z_A)\right)^{0.5} \tag{9}$$

---

[3]This method is convenient and easy to use but can be imprecise, especially if the deepwater steepness is highly irregular and not constantly positive. The results can then be confirmed using numerical simulations using a wave propagator such as Tomawac. Sergent et al. (2015) gave an estimation of the reliability of the simplified Goda modal compared to numerical methods (BEACH and SWAN for instance), they obtained a reasonable concordance for a steepness inferior to 7%, which corresponds to our case study.

175
$$h_{A,2\%} = c_{h2\%}(RU_{2\%} - z_A) \tag{10}$$

With $c_{v2\%}$ and $c_{h2\%}$ arbitrary coefficients that are used as fitting parameters. Z_A is the height of the dyke above the still water level and g the gravitational acceleration. These equations where compiled in (van der Meer, 2011; van der Meer et al., 2012) from the works led by Shüttrumpf and van Gent (2003) and Lorke et al. (2012).

The flow velocity will then decay along the crest following equation 11, which is a function of distance from the seaward
180 side of the crest (x_c). Note that this formula is only valid for a crest a few meters long as the formula becomes less precise for higher values of x_c.

$$\frac{v_{2\%}(x_c)}{v_{2\%}(x_c = 0)} = exp(-1.4x_c/L_0) \tag{11}$$

With $L_0 = g \cdot T_0^2$ the deep water wavelength of the incoming waves.

According to (van der Meer et al., 2012), the decrease of flow thickness upon reaching the crest is about one third and
185 can be attributed to the change of direction of the flow and stays relatively constant along the crest.

– **Landward slope flow** : The water trickles down the landward slope, this is where erosion usually happens. Since we quickly reach the maximum velocity of the flow on a slope such as ours, we directly use Eq. 12 to compute the terminal velocity of the flow.

190
$$v_b = \sqrt[3]{\frac{2 \cdot g \cdot h_{b0} \cdot v_{b0} \cdot \sin\beta}{f}} \tag{12}$$

with h_{b0} and v_{b0} are the flow thickness and velocity at the entry of the slope, respectively. f is the friction coefficient, which is determined experimentally when possible and estimated otherwise, g the gravity acceleration and β the slope angle.

These equations rely on a large number of parameters that are detailed in table 1.

Variable Name	Description	Value in Figure 8	Source
H_{dyke}	Height of the dyke	2.2	*in situ* data
f	Friction coefficient	0.02	EurOtop (2018)
β	Landward slope	30°	*in situ* data
α	Seaward slope	30°	*in situ* data
γ_f	Influence of roughness and porosity	0.6	EurOtop (2018)
γ_b	Influence of berm	1.0	EurOtop (2018)
d	Water depth at the toe of the dyke	0.54	*in situ* data
C_{h2}	Arbitrary coefficient of equation 10	0.2	EurOtop (2018)
C_{v2}	Arbitrary coefficient of equation 9	1.4	EurOtop (2018)

Table 1. Main control parameters in the equation system of the framework.

195 3.3 Resulting Terminal Velocity

We use the terminal velocity on the landward slope v_t as a criteria of erosion. Meaning that damage starts to occur when $v_t > v_c$ where v_c is the critical velocity which has to be determined using the literature. The results are shown in Figure 8.
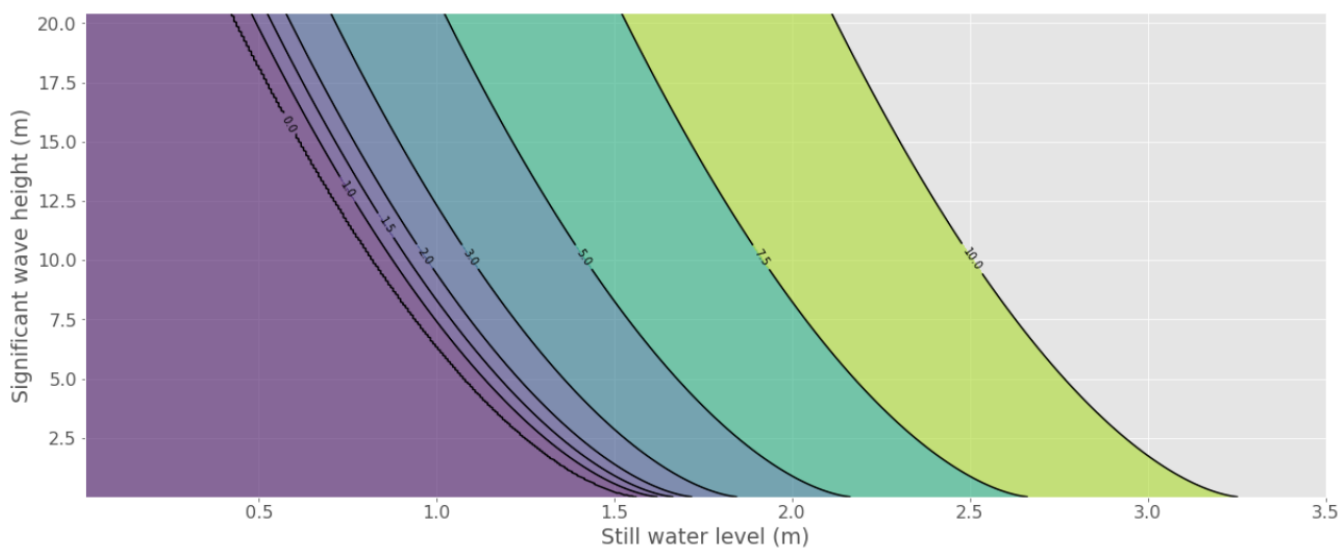
Figure 8. Terminal velocity along the landward slope for any couple (N, H_0).

Unsurprisingly, higher values of both N or H_0 induce higher values of terminal velocities. All values below the "0.0" line in Figure 8 failed to produce overtopping and thus generate a null value while in fact there is no water flowing on the slope.

4 Erosion Probability on Landward Slope

4.1 Combining Copula and Terminal Velocity

We can now associate a terminal velocity to a set $\mathcal{S}_{vt} = \{(N, H_0), f(N, H_0) = v_t\}$ that is the set of couples (N, H_0) which are associated through the function f to a terminal velocity v_t.

By integrating the derivative of the copula with respect to H_0 along the isoline $\mathcal{S}_{vt}$, we can obtain the return period of event $E_{vt} = \{v_t* > v_t\}$ which is any event implying a terminal velocity equal or higher than v_t (see equations 13 to 15).

$$P(v_t^* > v_t) = \iint\limits_{\mathcal{C}} \left(\frac{\partial^2 C_{N,H_0}}{\partial N \partial H_0} \right) dN dH_0 \tag{13}$$

$$P(v_t^* > v_t) = \int\limits_0^\infty \left[\frac{\partial C_{N,H_0}}{\partial H_0} \right]_{\mathcal{S}(H_0)}^\infty dH_0 \tag{14}$$

$$P(v_t^* > v_t) = -\int\limits_0^\infty \left(\frac{\partial C_{N,H_0}}{\partial H_0}(\mathcal{S}(H_0), H_0) \right) dH_0 \tag{15}$$

Where C is the surface of integration, which is the area below the velocity curve and $S(H_0)$ the velocity curve. This means that we can calculate the return period associated with a certain terminal velocity threshold for a defined dyke by fixing the parameters in Table 1.

Typically, we observe that for a dyke with characteristics close to the one located on the Quenin site in Camargue (see Table 1) a flow velocity of $2\,\mathrm{m/s}$, which is considered here as the critical flow velocity, will occur on average once every 5.86 years. This result is higher than what is observed on site (≈ 2 years) by the company. This gap can be caused by uncertainty on the parameters which will be further estimated via sensitivity analysis.

5 Sensitivity Analysis

5.1 Uncertainty Parameters

The showcased system is indeed able to provide return periods associated to events leading to erosion or any dangerous event defined as a criteria on flow velocity. However, added to the deep water conditions used to generate the copula, are the characteristics associated to the dyke as well as many empirical parameters used to fit the laws allowing the calculations leading to the landward terminal velocity of the dyke. All of these parameters carry an intrinsic amount of uncertainty which has a non-negligible impact on the results. This calls for an accurate quantification on the whole potential range of variation of each parameter. Global sensitivity analysis through the computation of global sensitivity indices will be our tool of choice.

A combination of the 1-st order and total effect sensitivity indices defined in equations (23 - 24) is a principled and classical
225 approach that encapsulates a useful enough amount of information on the variation of system's characteristics.

We estimate the value of the indices using the Saltelli estimator defined in equations (23 - 24). The number of dimensions being high, we accelerate the convergence of the estimator using a pseudo-random sampler, in our case the Sobol sequence, which generates a low discrepancy sample of points. The performance comparison of the Monte-Carlo process against the improved Quasi-Monte-Carlo estimations has been extensively discussed, noticeably in (Sobol', 1990, 1998; Sobol´ and Kucherenko,
230 2005; Acworth et al., 1998). The improvement in performance is unanimously in favour of the Quasi-Monte-Carlo Method.

The first step is to define the parameters used in equations 2 to 12 that we are going to consider as relevant sources of uncertainty. They are compiled into Table 2 where we associate a potential range of variation that is deemed as reasonable with its source. Each parameter is further described in its associated description below.

Variable Name	Description	Range of variation	Source of interval
H_{dyke}	Height of the dyke	$[1.89, 2.47]$	*in situ* data
f	Friction coefficient	$[0.01, 0.03]$	EurOtop (2018)
β	Landward Slope	$[20^o, 50^o]$	*in situ* data
α	Seaward Slope	$[20^o, 50^o]$	*in situ* data
γ_f	Influence of roughness and porosity	$[0.4, 0.8]$	EurOtop (2018)
γ_b	Influence of berm	$[0.75, 1.0]$	EurOtop (2018)
C_{h2}	Arbitrary coefficient in equation 10	$[0.1, 0.4]$	Bosman (2007) + Schüttrumpf (2001,2005)
C_{v2}	Arbitrary coefficient in equation 9	$[0.7, 2.1]$	Bosman (2007) + Schüttrumpf (2001,2005)
θ	Interdependency parameter (copula)	$[1.45, 1.75]$	Numerical Estimator
v_c	Critical erosion velocity	$[1.0, 4.0]$	Hughes (2012)
d	Water depth at the toe of the dyke	$[0.47, 0.82]$	*in situ* data
b_0	First coefficient of equation B1	$[0.028, 0.052]$	Goda (2000)
b_1	First coefficient of equation B2	$[0.52, 0.63]$	Goda (2000)

Table 2. Characteristics of the parameters used during the sensitivity analysis.

Parameters description

235 – **The height of the dyke** H_{dyke} is defined as the vertical distance between the still water level in a calm sea condition and the culminating point of the dyke. Using *in situ* data from a Litto3D bathymetry map. We are able to assess a reasonable range of variation of the dyke height.

– **The friction coefficient** f yields the resistance of contact between two materials, in our case between the landward slope of the dyke and water. A higher coefficient brings a slower flow velocity but also more shear stress. The EurOtop gives
240 many different values but raw clayey soil and mild vegetation generally fall in the given range of variation.

- **The landward slope** β is defined from the end of the crest which is considered as flat. The steeper the slope, the higher the terminal velocity.

- **The seaward slope** α is defined as the mean slope from the toe of the dyke to the beginning of the crest, assuming that the crest is flat. Its value is important as the behaviour of the up-rushing wave may change drastically for different values of α. Both angles are determined using the Litto3D map as previously cited.

- **The influence of roughness and porosity** on the seaward slope γ_f is a factor with value in the range 0 to 1 scaling how much the run-up will be attenuated thanks to the slope surface characteristics (1 means no influence). This is difficult to estimate as it relies on *in situ* experiments. Evaluating this parameter is not easy. Hence, we chose a relatively large range as the rocks on the slope are expected to have an influence of the same order of magnitude as other structures described in the EurOtop.

- **The influence of the berm** γ_b with value between 0 and 1 indicating the attenuation of the wave due to the presence of a berm. This value can be estimated using the geometry of the dyke if it is simple. It is more uncertain for a more complicated geometry. We calculate this factor using equations given in the EurOtop to which we add some variability as the dyke is not consistent through its length.

- **The depth at the toe of the dyke** b is calculated *in situ* using the Litto3D map as previously cited. Its value is registered for every transversal slice of the dyke. The range of variation yields the recorded minimum and maximum.

- **The scaling coefficients of the input crest velocity and thickness** C_{h2} and C_{v2}, respectively, are scaling factors on the equations calculating the velocity and thickness of the flow at the beginning of the crest from the run-up. The range is estimated as a variation of $+/-50\%$ from their suggested values in the EurOtop (2018).

5.2 Sobol indices

If we provide our framework inputs that are uncertain, it should be expected that the uncertainty will be carried through the system up to the outputs. We rely on sensitivity analysis to quantify such uncertainty by comparing the influence of each parameter on the variation of the outputs relative to their respective range of variation. Since there may be a lot of interaction between parameters and we need to assess the influence of the parameters over their whole range of variation, we use global sensitivity analysis.

Let $Y = f(X_1, ..., X_n)$ be a function of the X_i parameters with $i = 1, ..., n$. The uncertainty of the parameters X_i will carry over the uncertainty of the output Y. Therefore, it would be necessary to estimate the impact of parameters on the output Y.

In order to quantify the influence of a single parameter X_i on a complex system, a good starting point can be to fix this parameter to a defined value x_i. Logically, freezing a parameter, which is a potential source of variation, should reduce the variance $V(Y)$ of the output Y. Hence, a small value of variance $V_{X_{\sim i}}(Y|X_i = x_i)$ would imply a high influence of the parameter X_i. We can globalize the approach by calculating the average value of the variance over all valid values of x_i, preventing the dependence on x_i. This is written as :

$$E_{X_i}(V_{X_{\sim i}}(Y|X_i = x_i)) < V(Y) \tag{16}$$

The following relation is also useful in our case :

$$E_{X_i}(V_{X_{\sim i}}(Y|X_i = x_i)) + V_{X_i}(E_{X_{\sim i}}(Y|X_i = x_i)) = V(Y) \tag{17}$$

The conditional variance $V_{X_i}(E_{X_{\sim i}}(Y|X_i = x_i))$ is called the first-order effect of X_i on Y. We can then use the sensitivity measure called the sensitivity index or Sobol index (see (Sobol, 2001)) defined as :

$$S_i = \frac{V_{X_i}(E_{X_{\sim i}}(Y|X_i = x_i))}{V(Y)} \tag{18}$$

This gives the proportion of contribution of the parameter X_i alone on the total variance of the output Y relatively to the other parameters $X_{\sim i}$. The main drawback of this measure is that the interaction of the parameters between themselves is not taken into account. These measures are contained in higher-order indices. However, this may become quite time-consuming and impractical if the amount of parameter is high as the total number of Sobol indices that could be calculated grows as $n!$ with n the number of parameters.

Let us imagine what would happen if we were to have all the X_j with $j \neq i$ parameters frozen while only X_i can vary. The corresponding Sobol index can be written as :

$$\frac{V(E(Y|X_{\sim i}))}{V(Y)} = \frac{V(E(Y|X_1, ..., X_{i-1}, X_{i+1}, ..., X_n))}{V(Y)} \tag{19}$$

This term should include any Sobol index that does not yield the index i. Since the sum of all Sobol indices must be 1, we introduce the difference :

$$1 - \frac{V(E(Y|X_{\sim i}))}{V(Y)} \tag{20}$$

We then use equation 17 to simplify the expression :

$$E_{X_i}(V_{X_{\sim i}}(Y|X_i)) + V_{X_i}(E_{X_{\sim i}}(Y|X_i)) = V(Y) \tag{21}$$

Hence, dividing by V(Y) gives :

$$S_{Ti} = 1 - \frac{V_{X_i}(E_{X_{\sim i}}(Y|X_i))}{V(Y)} = \frac{E_{X_i}(V_{X_{\sim i}}(Y|X_i))}{V(Y)} \tag{22}$$

This is called the total effect Sobol' index, which measures the influence of a parameter i on the variance as well as its
295   interaction with every other parameters.

Although concise, equation 18 and 22 are difficult to calculate analytically. We circumvent the problem by using the method developed by (Sobol, 2001) and further improved by (Saltelli et al., 2008) using the Monte-Carlo method to estimate these parameters.

The protocol works as follows :

300   1. Generate a $(N, 2k)$ matrix of random values extracted from the distributions of the parameters, with k the number of parameters. N is called the base sample and varies between a few hundreds to thousands.

$$
\begin{bmatrix}
x_1^{(1)} & x_2^{(1)} & \dots & x_{k-1}^{(1)} & x_k^{(1)} & x_{k+1}^{(1)} & \dots & x_{2k}^{(1)} \\
x_1^{(2)} & x_2^{(2)} & \dots & x_{k-1}^{(2)} & x_k^{(2)} & x_{k+1}^{(2)} & \dots & x_{2k}^{(2)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_1^{(N-1)} & x_2^{(N-1)} & \dots & x_{k-1}^{(N-1)} & x_k^{(N-1)} & x_{k+1}^{(N-1)} & \dots & x_{2k}^{(N-1)} \\
x_1^{(N)} & x_2^{(N)} & \dots & x_{k-1}^{(N)} & x_k^{(N)} & x_{k+1}^{(N)} & \dots & x_{2k}^{(N)}
\end{bmatrix}
$$

2. Split the matrix into two (N, k) matrices, this gives 2 separate samples of parameters A and B. Each line can be computed by your system to give a specific output.

$$
A = \begin{bmatrix}
x_1^{(1)} & x_2^{(1)} & \dots & x_{k-1}^{(1)} & x_k^{(1)} \\
x_1^{(2)} & x_2^{(2)} & \dots & x_{k-1}^{(2)} & x_k^{(2)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_1^{(N-1)} & x_2^{(N-1)} & \dots & x_{k-1}^{(N-1)} & x_k^{(N-1)} \\
x_1^{(N)} & x_2^{(N)} & \dots & x_{k-1}^{(N)} & x_k^{(N)}
\end{bmatrix}
\quad
B = \begin{bmatrix}
x_{k+1}^{(1)} & x_{k+2}^{(1)} & \dots & x_{2k-1}^{(1)} & x_{2k}^{(1)} \\
x_{k+1}^{(2)} & x_{k+2}^{(2)} & \dots & x_{2k-1}^{(2)} & x_{2k}^{(2)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{k+1}^{(N-1)} & x_{k+2}^{(N-1)} & \dots & x_{2k-1}^{(N-1)} & x_{2k}^{(N-1)} \\
x_{k+1}^{(N)} & x_{k+2}^{(N)} & \dots & x_{2k-1}^{(N)} & x_{2k}^{(N)}
\end{bmatrix}
$$

3. From A and B, generate k matrice C_i which is composed of the matrix B with the i-th column that is replaced by the i-th column of matrix A.

$$
C_1 = \begin{bmatrix}
x_1^{(1)} & x_{k+2}^{(1)} & \dots & x_{2k-1}^{(1)} & x_{2k}^{(1)} \\
x_1^{(2)} & x_{k+2}^{(2)} & \dots & x_{2k-1}^{(2)} & x_{2k}^{(2)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_1^{(N-1)} & x_{k+2}^{(N-1)} & \dots & x_{2k-1}^{(N-1)} & x_{2k}^{(N-1)} \\
x_1^{(N)} & x_{k+2}^{(N)} & \dots & x_{2k-1}^{(N)} & x_{2k}^{(N)}
\end{bmatrix}
\quad
C_2 = \begin{bmatrix}
x_{k+1}^{(1)} & x_2^{(1)} & \dots & x_{2k-1}^{(1)} & x_{2k}^{(1)} \\
x_{k+1}^{(2)} & x_2^{(2)} & \dots & x_{2k-1}^{(2)} & x_{2k}^{(2)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_{k+1}^{(N-1)} & x_2^{(N-1)} & \dots & x_{2k-1}^{(N-1)} & x_{2k}^{(N-1)} \\
x_{k+1}^{(N)} & x_2^{(N)} & \dots & x_{2k-1}^{(N)} & x_{2k}^{(N)}
\end{bmatrix}
$$

4. Run the system for each line of matrices A, B and C_i, giving the output matrices $f(A)$, $f(B)$ and $f(C_i)$. This gives a total of $k(N+2)$ runs. This is significantly more efficient than brute-force which would require N^2 runs.

$$
Y_A = f(A) = \begin{bmatrix} y_A^{(1)} \\ y_A^{(2)} \\ \vdots \\ y_A^{(N-1)} \\ y_A^{(N)} \end{bmatrix}
\qquad
Y_B = f(B) = \begin{bmatrix} y_B^{(1)} \\ y_B^{(2)} \\ \vdots \\ y_B^{(N-1)} \\ y_B^{(N)} \end{bmatrix}
\qquad
Y_{C_i} = f(C_i) = \begin{bmatrix} y_{C_i}^{(1)} \\ y_{C_i}^{(2)} \\ \vdots \\ y_{C_i}^{(N-1)} \\ y_{C_i}^{(N)} \end{bmatrix}
$$

5. use the matrices to calculate the Sobol indices through the following estimators :

$$
S_i = \frac{V(E(Y|X_i))}{V(Y)} = \frac{Y_A \cdot Y_{C_i} - f_0^2}{Y_A \cdot Y_A - f_0^2} = \frac{\sum_{j=1}^{N} Y_A^{(j)} \cdot Y_{C_i}^{(j)} - f_0^2}{\sum_{j=1}^{N} Y_A^{(j)} \cdot Y_A^{(j)} - f_0^2}
\tag{23}
$$

$$
S_{Ti} = 1 - \frac{E(V(Y|X_i))}{V(Y)} = 1 - \frac{Y_B \cdot Y_{C_i} - f_0^2}{Y_A \cdot Y_A - f_0^2} = 1 - \frac{\sum_{j=1}^{N} Y_B^{(j)} \cdot Y_{C_i}^{(j)} - f_0^2}{\sum_{j=1}^{N} Y_A^{(j)} \cdot Y_A^{(j)} - f_0^2}
\tag{24}
$$

with

$$
f_0^2 = \left(\frac{1}{N} \sum_{j=1}^{N} Y_A^{(j)} \right)^2
\tag{25}
$$

which is the mean of the output sample.

After generating a sample of parameter values, each set is computed through the framework, giving an associated return period from which we calculate the global sensitivity indices of both 1-st order and total effect. The results are compiled in Figure 9.
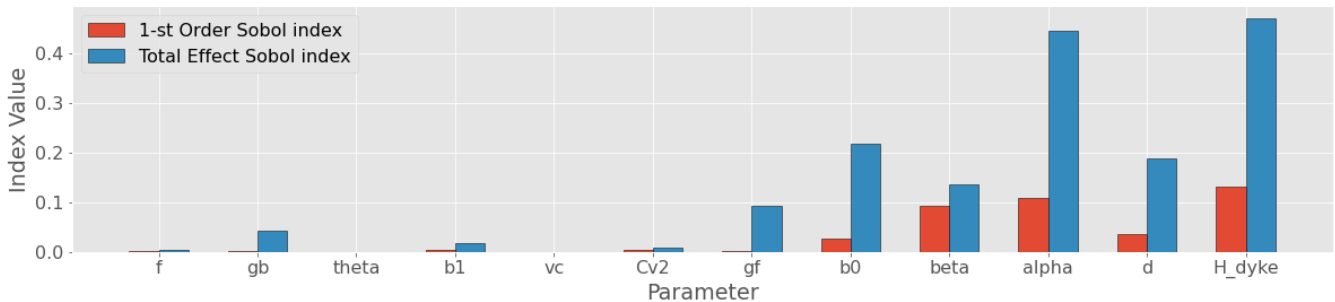
Figure 9. Value of the 1st-order sensitivity (in red) and total effect (in blue) indices for each tested parameter.

The first observation is that some parameters contribute a lot more to the global variance of the system than others. Each parameter lies in four different categories to which we can attribute a degree of importance from the most important to the less important :

325   1. The parameters related to the geometrical features of the dyke (H_{dyke}, α, β, ...) seem to carry on average a lot of uncertainty and should be inspected thoroughly;

    2. The foreshore with parameters d and b_0;

    3. The overtopping process relies on the intervention of many parameters which may have a significant importance (γ_f, C_{v2});

330   4. The erosional process with parameter v_c however looks to be either well-defined or only mildly significant according to the values of the Sobol' indices.

Also, the values of the total effect indices seem to be much higher than for the 1-st order, which indicates that a great high amount of variance is hidden in higher-order indices, proving the presence of strong interactions between the parameters.

5.3 Return periods distribution

335 Launching such a high number of calculations allows us to compile the return periods into a histogram to evaluate the probability of the return periods taking into account uncertainties. The results are compiled in Figure 10.
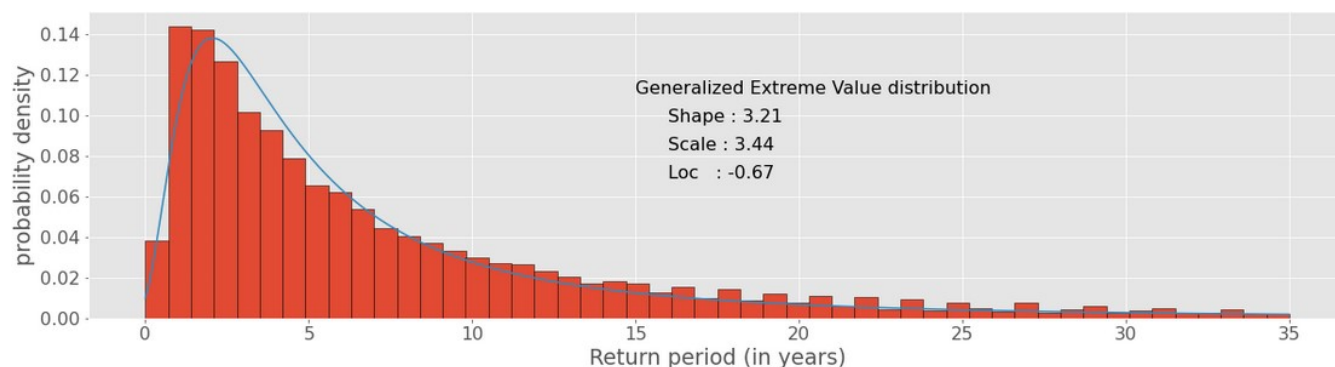
Figure 10. Distribution of the return periods of an event able to provoke some amount of erosion to landward slope a the dyke with random variation of the parameters in Table 2 according to their respective range of variation.

The results show that the distribution can be well fitted using a Generalized Extreme Value distribution which is right-skewed with a peak around the two years value and a long tail in the upper range of the return periods. The mean value is close to ten years.

340 The peak value is more representative that the mean as many of the extreme geometries may not be present in reality and are probably closer to the most frequent configurations. Historical data gathered from the company monitoring the dyke seem to be in accordance with the choice of the peak value as the representative metrics of the distribution.

This asymmetry is expected since a negative return period would not make sense physically while it is not bounded by any high value. It appears however that, sadly although unsurprisingly, a bad design of a dyke, resulting in a small return period,

345 can be more easily reached than a good one.

5.4 Results Validation

In order to make sure that the estimation of the sensitivity indices is accurate we need to ensure that the convergence of the estimator has been reached. We will do this by plotting the values of the indices and incrementally increasing the amount of points generated by the Sobol' sequence, this is called a validation curve. Note that the amount of plotting points is limited

350 because the Sobol' sequence, being a non independent sample, is only valid for 2^n points. The results are displayed in Figures 11 and 12.
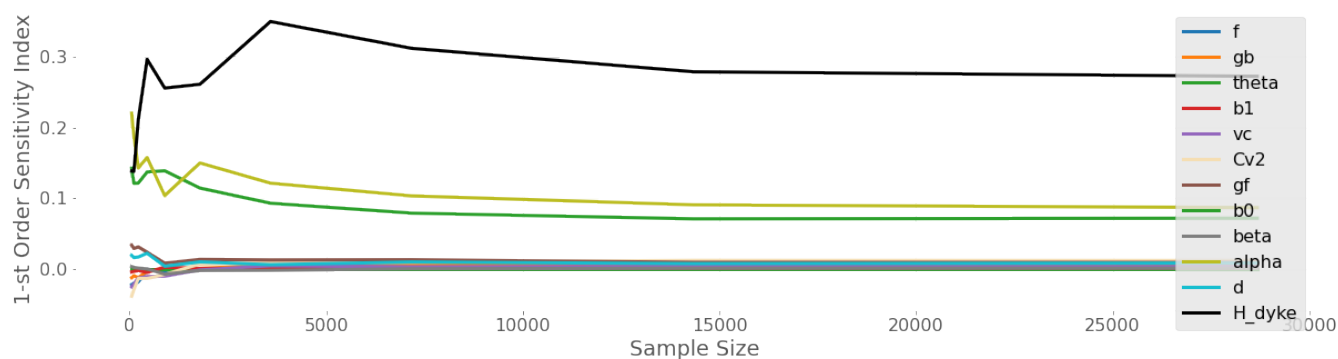
Figure 11. Evolution of the values of the 1-st order sensitivity indices for different sample sizes.
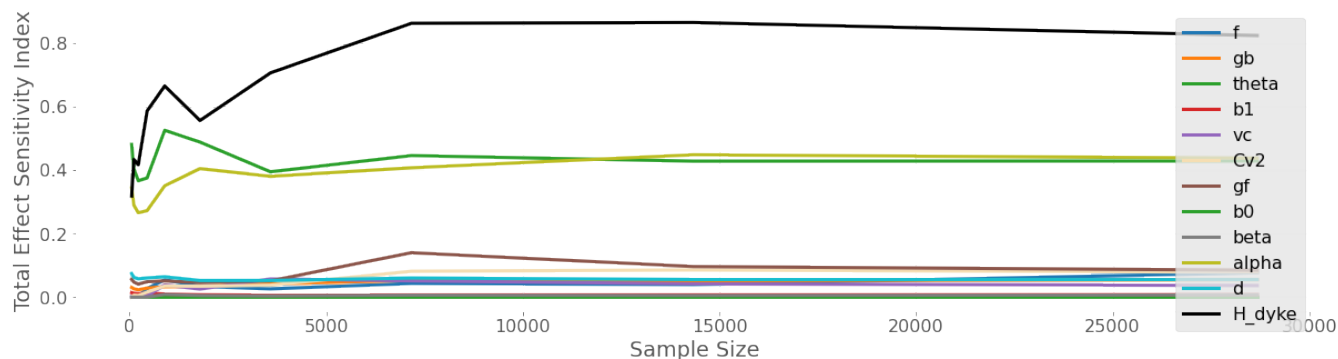
Figure 12. Evolution of the values of the total effect sensitivity indices for different sample sizes.

Convergence has evidently been reached. It seems that we can safely use ≈ 1500 points which in our case is still fairly low as the computation of the terminal velocity is pretty fast. However, should the computation time increase by changing the methods of calculation, this could become a problem which would require more intensive optimizations.

6 Conclusion

We have been able to build a complete automated framework allowing the user to estimate the expected return periods of events leading to erosion on the rear side of the earthen dyke submitted to wave overtopping, assuming the correctly assessed ranges of variation of the parameters are provided. The framework itself needs firstly meteocean data in order to create a reliable copula from wave and water level data, then a description of wave propagation to the toe of dyke and finally reliable laws representing wave overtopping process, run-off on the crest then on the landward slope and bottom erosion.

The return period for erosion on the Quenin dyke located in Salin-de-Giraud is firstly estimated from average parameters. This first estimate is equal to six years which is significantly higher than the value of two years written in reports from the operating company. The framework is then able to take parameters' uncertainty into account which provides a Generalized Extreme Value distribution of return periods which is right-skewed with a peak around the two years value and a long tail in the upper range of the return periods. This result shows that a statistical study is necessary to determine a return period of damages in accordance with observed damages. Damages on a long dyke are not observed on an average profile but on the weakest profile. That is why the peak of the statistical analysis is more representative than the first estimate based on average parameters. Sensitivity analysis is implemented into the framework and classifies the dyke's parameters in term of carried uncertainty. In the case of the Quenin dyke, the geometrical features of the dyke are the most important, followed in decreasing order by the foreshore conditions, the overtopping characteristics and finally the erosion process itself. These results are useful and provide new insights to the current state of on-site construction as the focus was previously directed towards the erosional features of the soil, which may not be the most cost-efficient solution. The conclusions about sensitivity should only be used on this particular dyke as they are custom-made. This study case is indeed very specific with a very low return period for damages and large variations of the dyke crest. For any other dyke, the framework is applicable by providing the appropriate input values.

Finally, the results can be provided relatively quickly without an enormous amount of computing power. They can be validated indeed using only a small set of points for the Quasi-Monte-Carlo process (around ten thousand points at most).

Code and data availability. Freely available on demand to the corresponding author

Author contributions. C. Lutringer - Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, Data Curation, Visualization

A. Poupardin - Supervision, Writing - Review and Editing, Methodology, Resources

P. Sergent - Conceptualization, Methodology, Validation, Surveilance, Project Administration, Funding acquisition

A. Bennabi - Conceptualization, Supervision, Project Administration

J. Jeong - Supervision, Writing - Review and Editing, Resources, Funding acquisition, Project Administration

385 *Competing interests.* The authors declare that they have no conflict of interest

Appendix A: Pictures of the dyke

Figure A1. Picture of the landward slope of the dyke

Appendix B: Propagation equations from Goda (2000)

$$\beta_0 = b_0 \cdot \left(\frac{H_0}{L_0}\right)^{-0.38} * e^{20 \cdot m^{1.5}} \tag{B1}$$

$$\beta_1 = b_1 \cdot e^{4.2 \cdot \tan\theta_a} \tag{B2}$$

390 $$\beta_{\max} = \max[0.92, 0.32 \cdot (H_0/L_0)^{0.29} \cdot e^{2.4 \cdot \tan\theta_a} \tag{B3}$$

with m the average steepness of the seabed between the offshore point and the toe of the dyke, θ_a the angle of attack of the oblique waves and L_0 the deep water wavelength. b_0 and b_1 are coefficient determined empiracally from (Goda, 2000) who gives their values of 0.028 and 0.052, respectively.

References

395 Acworth, P., Broadie, M., and Glasserman, P.: A Comparison of Some Monte Carlo and Quasi Monte Carlo Techniques for Option Pricing, Springer New York, 1998.

Bergeijk, V. V., Warmink, J., van Gent, M., and Hulscher, S.: An analytical model of wave overtopping flow velocities on dike crests and landward slopes, Coastal Engineering, 2019.

Bernardara, P., Mazas, F., Kergadallan, X., and Hamm, L.: A two-step framework for over-threshold modelling of environmental extremes,
400   Natural Hazards and Earth Systems Sciences, 14, 635–647, 2014.

Brunner, M., Favre, A.-C., and Seibert, J.: Bivariate return periods and their importance for flood peaks and volume estimations, WIREs Water, 3, 819–833, 2016.

Capel, A.: Wave run-up and overtopping reduction by block revetments with enhanced roughness, Coastal Engineering, 104, 76–92, 2020.

De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R.: A multivariate model of sea storms using copulas, Coastal Engineering, 54,
405   734–751, 2007.

Durante, F. and Sempi, C.: Copula Theory: An Introduction, Lecture Notes in Statistics, 198, 3–31, 2010.

Durante, F. and Sempi, C.: Principles of Copula Theory, CRC Press, 2016.

Durante, F., Fernadez-Sanchez, J., and Sempi, C.: A Topological Proof of Sklar's Theorem, Applied Mathematices Letters, 26, 945–948, 2013.

410 Goda, Y.: Random Seas and Design of Maritime Structures, vol. 15, World Scientific Publishing Co., 2000.

Hughes, S. and Nadal, N.: Laboratory study of combined wave overtopping and storm surge overflow of a levee, Coastal Engineering, 56, 244–259, 2008.

Hughes, S., C.Thornton, der Meer, J. V., and Scholl, B.: Improvements in describing wave overtopping processes, Coastal Engineering, 2012.

Kergadallan, X.: Estimation des niveaux marins extrêmes avec et sans l'action des vagues le long du littoral métropolitain, 2015.

415 Kole, E., Koedijk, K., and Verbeek, M.: Selecting copulas for risk management, Journal of Banking and Finance, 31, 2405–2423, 2007.

Kumar, P. and Guloksuz, C.: Choosing the Best Copula Function in Mathematical Modeling, Springer Proceedings in Mathematics & Statistics, 344, 2021.

Li, T., Troch, P., and Rouck, J. D.: Wave overtopping over a sea dike, Journal of Computational Physics, 198, 686–726, 2003.

Liu, S. and Han, J.: Energy efficient stochastic computing with Sobol sequences, Design, Automation & Test in Europe Conference &
420   Exhibition, pp. 650–653, 2017.

Lorke, S., Borschein, A., Schüttrumpf, H., and Pohl, R.: Influence of wind and current on wave run-up and wave overtopping. Final report., FlowDike-D, 2012.

Mehrabani, M. and Chen, H.: Risk Assessment of Wave Overtopping of Sea Dykes Due to Changing Environments, Conference on Flood Risk Assessment, 2015.

425 Orcel, O., Sergent, P., and Ropert, F.: Trivariate copula to design coastal structures, Nat. Hazards Earth Syst. Sci., 21, 1–22, 2020.

Pörtner, H.-O., Roberts, D., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., and Rama, B.: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2022.

Saltelli, A., Ratto, M., Terry, A., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global Sensitivity Analysis, The
430   Primer, John Wiley and Sons, Ltd, 2008.

Salvadori, G. and Michele, C. D.: On the use of Copulas in Hydrology : Theory and Practice, Journal of Hydrologic Engineering, 12, 2007.

Sergent, P., Prevot, G., Mattarolo, G., Brossard, J., Morel, G., Mar, F., Benoit, M., Ropert, F., Kergadallan, X., Trichet, J., and Mallet, P.: Stratégies d'adaptation des ouvrages de protection marine ou des modes d'occupation du littoral vis-à-vis de la montée du niveau des mers et des océans, Ministère de l'écologie, du développement durable, du transport et du logement, 2015.

435 Shüttrumpf, H. and van Gent, M.: Wave overtopping at seadikes, Proc. Coastal Structures, pp. 431–443, 2003.

Sklar, A.: Fonctions de répartitions à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris, 8, 229–231, 1959.

Sobol', I.: Quasi-Monte Carlo methods, Progress in Nuclear Energy, 24, 55–61, 1990.

Sobol', I.: On quasi-Monte Carlo integrations, Mathematics and Computers in Simulation, 47, 103–112, 1998.

Sobol, I.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in
440   Simulation, 55, 271–280, 2001.

Sobol´, I. and Kucherenko, S.: On global sensitivity analysis of quasi-Monte Carlo algorithms, Monte Carlo Methods and Applications, 11, 83–92, 2005.

Tootoonchi, F., Sadegh, M., Haerter, J., Raty, O., Grabs, T., and Teutschbein, C.: Copulas for hydroclimatic analysis: A practice-oriented overview, WIREs Water, 9, 2022.

445 van der Meer, J.: The Wave Run-up Simulator. Idea, necessity, theoretical background and design, Van der Meer Consulting Report vdm11355, 2011.

van der Meer, J., Provoost, Y., and Steendam, G.: The wave run-up simulator, theory and first pilot test, Proc. ICCE, 2012.

van der Meer, J., Allsop, N., Bruce, T., de Rouck, J., Kortenhaus, A., Pullen, T., Schüttrumpf, H., Troch, P., and Zanuttigh, B.: EurOtop. Manual on wave overtopping of sea defences and related structures. An overtopping manual largely based on European research, but for
450   worldwide application, www.overtopping-manual.com, 2018.