



Cas d'étude d'une installation industrielle

COURS "PRATIQUE DE LA FIABILITÉ"

RAPPORT DE PROJET

Mars 2023

Haythem Boucharif, Ferdinand Genans Boiteux, Khoa Anh Ngo

Table des matières

1 Détermination de la hauteur de la digue à partir des relevés de mesure historiques	3
1.1 Traiter les données manquantes	3
1.1.1 Modèle de régression	4
1.2 Modéliser les données de H	5
1.3 Trouver une hauteur de digue h_d de façon à "minimiser le risque d'inondation"	6
2 Détermination de la hauteur de la digue à partir à partir du modèle hydraulique	8
2.1 Calcul par le modèle hydraulique et les hypothèses probabilistes	8
2.2 Calcul par le modèle hydraulique et les données historiques	9
2.3 Calcul par estimation fréquentiste	11
3 Détermination de la hauteur de la digue à partir du modèle économique	13
3.1 Les fonctions de coût	13
3.1.1 Coût d'investissement	14
3.1.2 Coût des dommages de la digue	14
3.1.3 Coût des dommages du site	15
3.2 Exploitation du modèle hydraulique	15
3.2.1 Génération de notre échantillon	16
3.3 Résolution du problème économique	16

1 Détermination de la hauteur de la digue à partir des relevés de mesure historiques

Le but de cette section est de réduire au minimum la probabilité $P(H > h_d)$ en fonction de h_d en utilisant les données historiques de mesure de 1849 à 1997. Ces données se composent de deux ensembles distincts :

- La première colonne est constituée des valeurs de débit maximal annuel de crue Q observées sur une année, sans aucune donnée manquante.
- La deuxième colonne se compose des valeurs de hauteur de l'eau observées H lors d'une année de crue associé à un débit Q. Certaines de ces données sont manquantes.

La démarche suivie est la suivante :

- Traiter les données manquantes en choisissant une méthode appropriée.
- Modéliser les données de H comme des réalisations d'une variable aléatoire avec une loi à déterminer.
- Choisir un quantile de la loi pour évaluer le risque acceptable dans cette problématique. Ainsi, on cherche une hauteur de digue telle que la probabilité $P(X > h_d) = q$, où H est modélisé par la variable aléatoire X, et q est faible.

1.1 Traiter les données manquantes

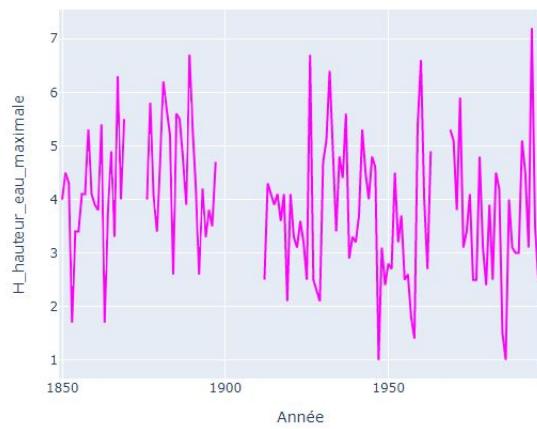
Nous devons trouver un moyen de traiter les données manquantes de H car la hauteur de la digue h_d est directement liée aux données de H. Étant donné que nous avons un petit nombre de données, nous privilégierons l'ajout des données manquantes plutôt que la suppression de ces données.

Année <int>	Q <int>	H <dbl>
1849	3854	NA
1850	1256	4.0
1851	1649	4.5
1852	1605	4.3
1853	341	1.7

FIGURE 1 – 5 premières lignes de notre donnée

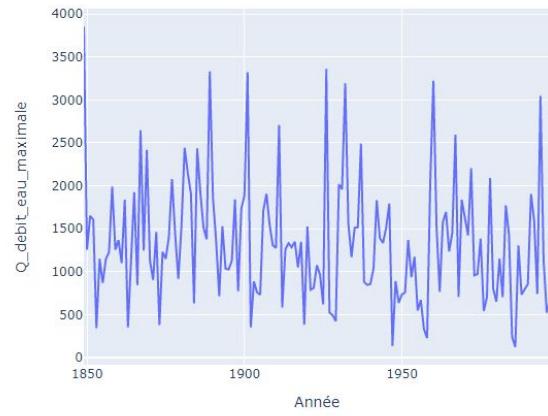
On a constaté qu'il y avait 26 données manquantes sur un total de 149, ce qui représente environ 17%. De plus, avec les observations initiales, on constate aussi une relation entre H et Q, où lorsque Q augmente, H augmente également et vice versa. Pour vérifier cette relation entre Q et H, nous allons maintenant visualiser l'évolution de Q et H par rapport au temps.

Evolution de la hauteur d'eau maximale



(a) Graphe de H

Evolution du débit d'eau maximale



(b) Graphe de Q

FIGURE 2 – Évolution de Q et H au cours du temps

On peut voir graphiquement que l'évolution de Q et de H ont de même tendance. Pour renforcer cette affirmation, nous allons calculer la corrélation entre ces deux variables. En calculant la corrélation à partir de la base de données après avoir supprimé les valeurs manquantes, nous obtenons le résultat qui est un peu près à 0.965, on déduit une forte corrélation empirique entre Q et H . Il y a, en ce sens, des possibilités de bien reformer les données manquantes de H à partir de Q .

1.1.1 Modèle de régression

Après visualisations des données H en fonction de Q , nous avons décidé de modéliser leur relation par une régression linéaire.

$$H = a + b \cdot Q$$

avec a, b les coefficients à déterminer. Cette méthode à l'avantage d'être simple à calculer ainsi qu'à interpréter. De plus, cette méthode est robuste vis à vis des valeurs manquantes et/ou aberrantes. Par cette méthode, on trouve un coefficient de détermination $R^2 = 0.962$, ce qui indique que 96,2% de la variance de H est expliquée par Q dans le modèle de régression. Cela suggère que le modèle est assez bon pour expliquer les données et qu'il y a une forte corrélation entre H et Q .

Voici la visualisation d'un modèle de régression sur notre base de donnée après cette implémentation :

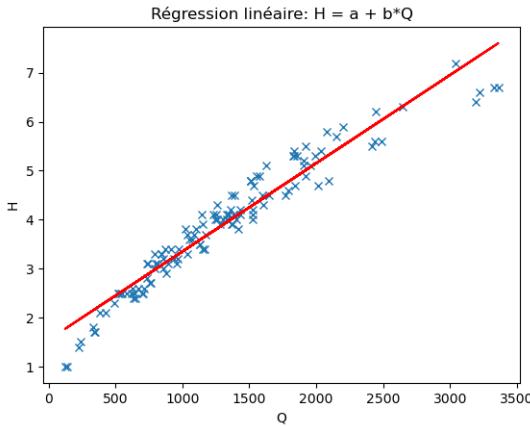
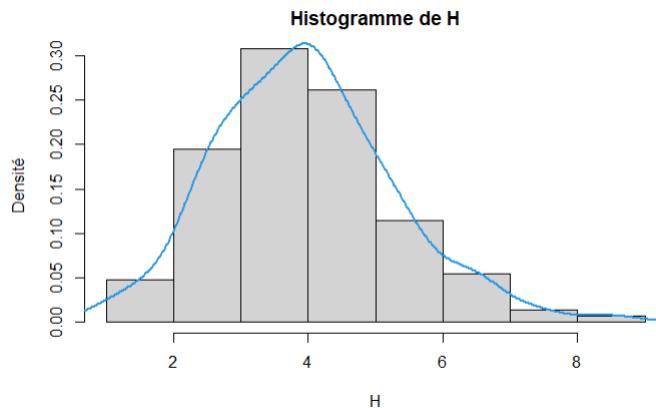


FIGURE 3 – Régression linéaire de H en fonction de Q

Remarque : d'autres méthodes et modèles auraient pu être utilisées pour cette imputation de données. Par exemple, une régression polynomiale collerait mieux à nos données. Cependant, comme certains facteurs non déterministes peuvent rentrer en jeu pour calculer H , nous gardons notre modèle robuste plutôt qu'un autre qui serait trop proche des données historiques.

1.2 Modéliser les données de H

L'objectif de cette partie est de trouver une loi qui modélise bien les données de H . On trace tout d'abord l'histogramme de H pour avoir un vue générale.



En observant notre histogramme, la hauteur H semble suivre une loi normale. Pour renforcer cette opinion, on utilise donc le test adéquation Kolmogorov-Smirnov, avec un risque de 5% et le test Shapiro-Wilk pour la loi normale.

Principe du test Kolmogorov-Smirnov :

Le principe est de choisir comme statistique une distance entre fonctions que l'on applique à F_n et F_0 .

$$D_n = \sup_{t \in R} |F_n(t) - F_0(t)|$$

On applique maintenant ce test sur nos données :

```

one-sample kolmogorov-smirnov test

data: data$H
D = 0.062803, p-value = 0.5993
alternative hypothesis: two-sided

```

FIGURE 4 – Résultat du test Kolmogorov-Smirnov

On trouve le **p-value** qui est supérieur à 0.05, cela indique de ne pas rejeter hypothèse nulle, ca veut dire que cette distribution suit la loi normale par rapport à ce test.

Principe du test Shapiro-Wilk

Soit L'échantillon X_1, \dots, X_n et 2 hypothèses

H_0 : La loi de l'échantillon est normale

H_1 : La loi de l'échantillon n'est pas normale

Le test de Shapiro-Wilk consiste à considérer le rapport entre l'estimation de la variance suivant la droite d'Henry et l'estimation de la variance par l'estimateur habituel.

$$W = \frac{(\sum_{i=1}^n a_i x_i)^2}{(\sum_{i=1}^n x_i - \bar{x})^2}$$

ou les a_i sont des constantes tabulées. Si la distribution est bien normale, ce rapport doit être proche de 1.

On applique maintenant ce test :

```

shapiro-wilk normality test

data: data$H
W = 0.99088, p-value = 0.4519

```

FIGURE 5 – Résultat du test Shapiro-Wilk

On trouve le **p-value** qui est supérieur à 0.05, cela indique de ne pas rejeter hypothese nulle, ca veut dire que cette distribution suit la loi normale par rapport à ce test. Donc, avec ces 2 tests, on peut maintenant d'affirmer que :

$$H \sim \mathcal{N}(\mu, \sigma^2), \text{ avec } \mu = 3.95 \text{ et } \sigma^2 = 1.32$$

où μ et σ^2 sont respectivement la moyenne empirique et la variance empirique de nos données historiques.

1.3 Trouver une hauteur de digue h_d de façon à "minimiser le risque d'inondation"

Dans le premier cas d'étude, où l'enjeu premier est relatif à la sûreté, la variable d'intérêt est directement la hauteur de la surverse et on cherche à dimensionner la hauteur de la digue h_d de sorte à minimiser le risque de débordement, c'est-à-dire de minimiser la probabilité

$P(H > h_d)$. L'enjeu étant ici de définir un site “hautement protégé”, nous choisissons un risque de 1%, qui est un risque classique pour les enjeux de sécurité. On obtient alors en prenant le quantile 0.99 de la loi normale avec les paramètres ajustés, la hauteur de digue h_d qui est égale **7.03** environ pour que la probabilité $P(H \leq h_d) = 99\%$.

2 Détermination de la hauteur de la digue à partir à partir du modèle hydraulique

Dans cette section, nous allons nous intéresser de manière plus détaillée aux probabilités d'inondations qu'on appellera par la suite S en utilisant un modèle plus complexe que dans la première partie. On se basera non seulement seulement sur les données historiques brutes de la hauteur maximale annuelle d'eau H , mais aussi sur un modèle appelé hydraulique, qui vise à calculer avec plus de précision cette hauteur.

La formule du modèle inclut certaines variables aléatoires qui suivent des lois de probabilités courantes, tandis que d'autres composantes du modèle sont déterministes pour simplifier les calculs. Ainsi, la hauteur maximale d'eau H peut s'exprimer comme :

$$H = \left(\frac{Q}{K_s \times \sqrt{\frac{Z_m - Z_v}{L}} \times B} \right)^{\frac{3}{5}} \quad (1)$$

Nous rappelons que le débit maximal Q du cours d'eau (en m^3/s) suit une loi de Gumbel de paramètres $\alpha = 1013$ et $\beta = 558$. Le coefficient de frottement K_s suit quant à lui une loi normale de moyenne 30 et d'écart-type 7.5. Enfin Z_v et Z_m modélisant la cote du fond suivent tout deux une loi triangulaire d'espérance respective 50 et 55 et de même demi-étendue valant 1.

En ajoutant les variables d'environnement/physiques, le calcul de la hauteur de surverse S s'effectue comme suit :

$$S = Z_c - Z_d = Z_v + H - h_d - Z_b \quad (2)$$

l'objectif devient donc de trouver h_d tel que :

$$\arg \min_{h_d \in \mathbb{R}^+} \mathbb{P}(S > 0)$$

2.1 Calcul par le modèle hydraulique et les hypothèses probabilistes

Pour estimer la probabilité que la hauteur de surverse est positive en fonction de h_d , nous pouvons supposer que nous n'avons pas de données historiques, mais que nous connaissons les lois des variables aléatoires qui composent la hauteur maximale d'eau grâce à l'équation (1), que nous allons utiliser par la suite dans l'équation (2).

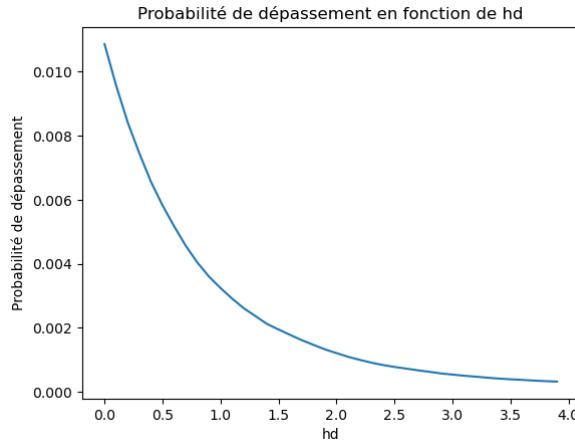
Nous pouvons alors commencer par simuler des échantillons $(h^{(1)}, \dots, h^{(K)}) \stackrel{iid}{\sim} H$ de manière itérative selon les étapes suivantes :

Boucle : Calcul de la hauteur d'eau maximale

1. $L = 5000, B = 300, Z_b = 55.5$
2. Pour i allant de 1 à 100000 :
 - i. Générer Q avec $Q \sim Gumbel(\alpha = 1031, \beta = 558)$
 - ii. Générer K_s avec $K_s \sim N(30, 7.5)$.
 - iii. Générer Z_m avec $Z_m \sim Triang(a = 54, c = 55, b = 56)$
 - iv. Générer Z_v avec $Z_m \sim Triang(a = 49, c = 50, b = 51)$

$$v. \text{ Calculer } H \text{ en utilisant la formule hydraulique : } H = \left(\frac{Q}{K_s \sqrt{\frac{Z_m - Z_v}{L}} B} \right)^{3/5}$$

Avec notre échantillon $(h^{(1)}, \dots, h^{(K)})$ et en gardant l'échantillon $(Z_v^{(1)}, \dots, Z_v^{(K)})$ avec l'étape iii de notre boucle, on calcule la surverse en fonction de différentes valeurs d'hauteur de digue h_d . Nous pouvons ensuite tracer un graphique de la probabilité empirique de surverse en fonction de la hauteur de digue.



Remarque : bien qu'ici une simulation itérative n'est pas nécessaire dû à l'hypothèse d'indépendance des lois, cette façon de faire pourrait être meilleure avec un modèle hydraulique plus complexe.

Il est évident que la probabilité de submersion diminue avec l'augmentation de la hauteur de la digue. Toutefois, la décision finale dépendra du niveau de risque que l'on est prêt à accepter. Si l'on souhaite éviter tout risque de submersion et que l'on cherche la hauteur de digue qui annule la probabilité $P(S > 0)$, alors cette hauteur sera de **3.9 m**. En revanche, si l'on accepte un niveau de confiance de 99.99% correspondant à une probabilité de surverse $P(S > 0) = 0.1\%$, alors la hauteur de digue optimale sera de **2 m**.

2.2 Calcul par le modèle hydraulique et les données historiques

Dans cette partie, on fait le choix d'utiliser à la fois le modèle hydraulique et les données historiques à disposition. Nous allons utiliser le modèle hydraulique pour compléter nos données, remplaçant la hauteur H manquante par le modèle hydraulique quand c'est nécessaire. Nous avons les données historiques du débit maximal d'eau, nous connaissons les valeurs déterministes de B et L , ainsi que les lois de probabilités de K_s , Z_m et Z_v .

Par analogie avec le 2.1, nous allons encore utiliser une boucle itérative pour nos données et imputer les valeurs manquantes.

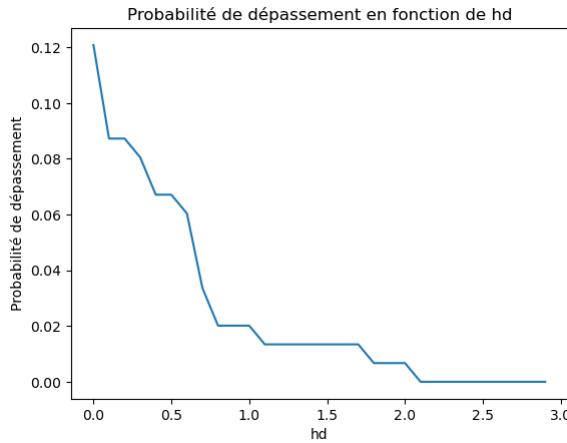
Boucle : Calcul de la hauteur d'eau maximale manquante

1. $L = 5000$, $B = 300$, $Z_b = 55.5$
2. Pour i parcourant les années données :
 - (a) Si $H_i = null$:
 - i. Générer K_s en utilisant une distribution normale avec une moyenne de 30 et un écart-type de 7.5.
 - ii. Générer Z_m en utilisant une distribution triangulaire avec une moyenne de 55 et demi-étendue de 1.

iii. Générer Z_v en utilisant une distribution triangulaire avec une moyenne de 50 et demi-étendu de 1.

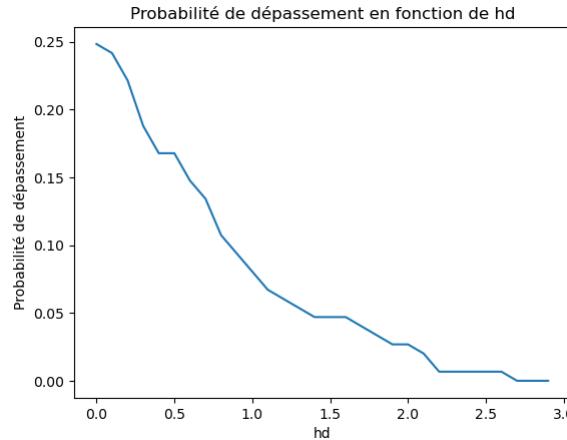
iv. Calculer H en utilisant la formule hydraulique : $H_i = \left(\frac{Q_i}{Ks\sqrt{\frac{Z_m - Z_v}{L}}B} \right)^{3/5}$

Finalement, maintenant que nous avons les valeurs manquantes de H nous pouvons tracer un graphique empirique de la probabilité de surverse, tout comme en 2.1 en utilisant l'équation (2). Pour cela, on génère une unique observation 'aléatoire' donnant la cote du fond du cours d'eau Z_v pour chaque année, et la substituer dans chaque opération pour chaque h_d . Nous pourrons alors obtenir les probabilités que la hauteur de surverse soit positive :



Avec une simple opération d'optimisation, nous constatons que la hauteur h_d optimale est de **2.1 m**.

Une autre approche possible, consiste à éviter tout risque en prenant la valeur maximale que Z_v peut atteindre pour chaque année. Cela nous donnerait évidemment un résultat différent :



Il est clair que la probabilité que la hauteur de surverse S soit positive augmente considérablement en prenant le pire scénario. Cela rend la hauteur de la digue qui minimise ce risque beaucoup plus élevée : $h_d = 2.7 \text{ m}$.

En conclusion, nous avons décidé de garder une hauteur de digue $h_d = 2.1m$, qui semble être la solution la plus adaptée à la réalité et à nos hypothèses de travail.

2.3 Calcul par estimation fréquentiste

Dans le même objectif, nous allons tenter de modéliser la problématique de manière similaire à la première partie. Nous savons que la hauteur de surverse S est composée d'un mélange de variables aléatoires dont nous connaissons les lois. Supposons que $F = S - h_d$ soit une variable aléatoire, pour laquelle nous avons des observations de 1849 à 1997. L'objectif serait ensuite de supposer une loi proche de la distribution historique, puis d'effectuer un test statistique pour estimer le niveau de confiance que notre variable aléatoire suit effectivement la distribution empirique de F . Ensuite, nous estimerons ses paramètres et déterminerons la valeur de h_d correspondant à un certain niveau de risque α , qui vérifie $P(F > h_d) = \alpha$.

De la même manière que les approches précédentes, nous avons déjà les données complètes de H , il suffit de faire un calcul annuel de $F = Z_v + H_{Zb}$ pour disposer de 149 observations annuelles. Nous allons regarder l'histogramme des données et faire une estimation non paramétrique de la densité de F à l'aide de la méthode du noyau. Le graphique suivant illustre cela :

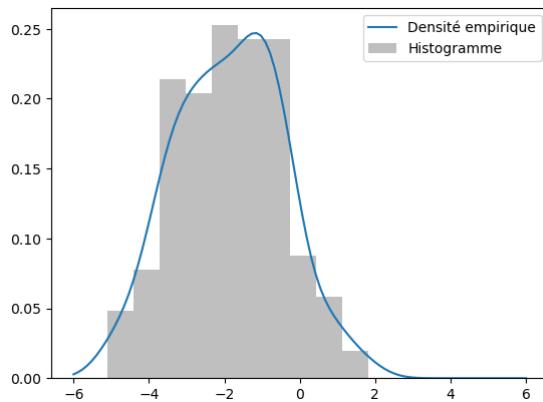


FIGURE 6 – Histogramme et densité empirique des observations de F

En examinant la figure ci-dessous, nous pouvons supposer que F suit une distribution issue de la famille des lois d'extrême généralisées, typiquement une loi de Gumbel. Cela pourrait correspondre avec notre densité empirique qui présente une queue légèrement lourde à droite. Cela signifie qu'elle attribue parfois une probabilité non négligeable aux valeurs extrêmes.

Pour vérifier notre hypothèse, nous allons utiliser le test statistique de Kolmogorov-Smirnov, détaillé dans la première partie. Avec un niveau de signification $\alpha = 5\%$, nous obtenons les résultats suivants :

t -test = 0.08

p -value = 0.27

Étant donné que la valeur p est supérieure au niveau de signification, nous ne pouvons pas rejeter l'hypothèse nulle et concluons que la variable aléatoire F ne diffère pas significativement d'une loi de Gumbel.

La densité de probabilité de la loi de Gumbel est donnée par la formule suivante :

$$f(x) = \frac{1}{\beta} \exp\left(-\frac{x-\mu}{\beta} - \exp\left(-\frac{x-\mu}{\beta}\right)\right) \quad (3)$$

où μ est le paramètre de localisation et β est le paramètre d'échelle. Nous les estimons par la méthode du maximum de vraisemblance :

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \quad \hat{\mu} = \bar{x} - \frac{0.5772}{\hat{\beta}} \quad (4)$$

où x_i sont les observations, \bar{x} est la moyenne des observations et n est le nombre d'observations.

Finalement, après inférence, nous obtenons les estimations :

$$\hat{\beta} = 1.35 \quad \hat{\mu} = -2.57$$

Si nous choisissons un niveau de risque de $\alpha = 1\%$, nous pourrions rechercher h_d tel que $P(F > h_d) = 0.01$. Ainsi, nous obtenons la hauteur de digue h_d d'environ **3.6 m**.

On constate une valeur de h_d très proche de celle trouvée en 2.1. Finalement, nos trois manières différentes trouvent une hauteur de digue optimale vers les 3 mètres.

3 Détermination de la hauteur de la digue à partir du modèle économique

Cette section se concentre sur l'aspect économique de la hauteur de la digue, dans le but de déterminer la hauteur h_d optimale qui permettrait de minimiser le risque économique associé à la surverse. Plus particulièrement, en gardant les notations du paragraphe 1.6, on s'intéresse au problème suivant :

$$(P) \arg \min_{h_d \in \mathbb{R}^+} C_{c,moyenne}$$

Comme indiqué en 1.6, nous allons étudier ce problème pour une période de 30 ans, c'est à dire $T = 30$. Ainsi, on a :

$$\arg \min_{h_d \in \mathbb{R}^+} \frac{C_c(30)}{30}$$

où C_c est le coût complet de l'installation industrielle sur une période donnée.

En utilisant les équations (3) à (5) du paragraphe 1.6, on a :

$$\begin{aligned} C_c(T) &= C_t(T) + \sum_{j=1}^T C_{d,j}(S_j, h_d) \\ &= C_i(h_d) + T \cdot C_m(h_d) + \sum_{j=1}^T C_{d,j}(S_j, h_d) \\ &= C_i(h_d)(1 + \frac{T}{100}) + \sum_{j=1}^T C_{d,j}(S_j, h_d) \\ &= C_i(h_d)(1 + \frac{T}{100}) + \sum_{j=1}^T (C_s(S_j) + C_g(S_j, h_d)) \end{aligned}$$

Nous avons utilisé le fait que le coût de maintenance C_m est égale à 1 % du coût d'investissement C_i , c'est-à-dire : $C_m(h_d) = \frac{C_i(h_d)}{100}$. En utilisant cette expression de C_c et en enlevant la fraction $\frac{1}{T}$ qui ne change pas la valeur de notre argmin on réécrit (P) :

$$(P) \arg \min_{h_d \in \mathbb{R}^+} C_i(h_d)(1 + \frac{30}{100}) + \sum_{j=1}^{30} (C_s(S_j) + C_g(S_j, h_d))$$

Pour l'instant (P) n'est pas encore résolvable car nous n'avons pas de formules pour les fonctions de coût C_i , C_s et C_g . De plus, la surverse S n'est pas exprimée en fonction de h_d . Nous allons étudier ces points par la suite.

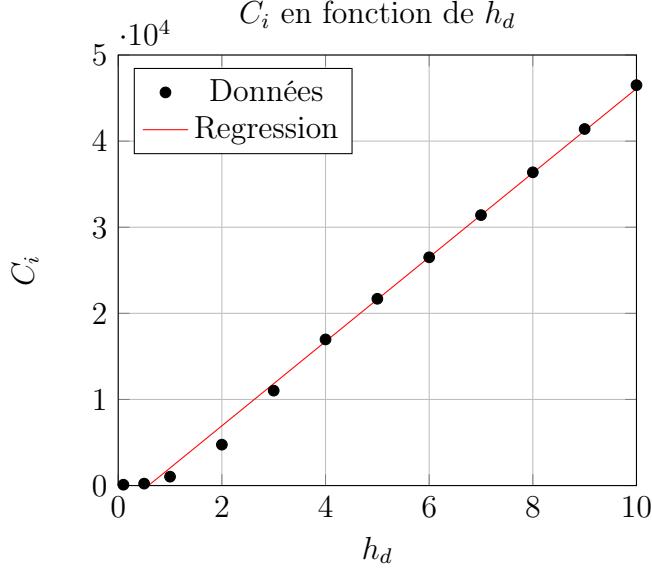
3.1 Les fonctions de coût

Outre les valeurs numériques données pour les différents coûts en fonction de S et de h_d , nous avons les informations suivantes :

- C_i est déterministe, fonction de h_d , l_d et C_l où C_l est le coût linéaire. De plus C_l croît plus fortement que linéairement par rapport à h_d , à l'inverse de l_d .
- C_g est aussi déterministe, fonction de S .
- C_s est croissant par rapport à S mais garde une composante aléatoire.

3.1.1 Coût d'investissement

On sait que le coût d'investissement est une fonction se calculant à partir de h_d , l_d et C_l et nous avons des valeurs numériques à notre disposition. De plus, l_d et C_l peuvent être reformulées par rapport à h_d ce qui nous donne finalement que C_i est de la forme $C_i(h)$.

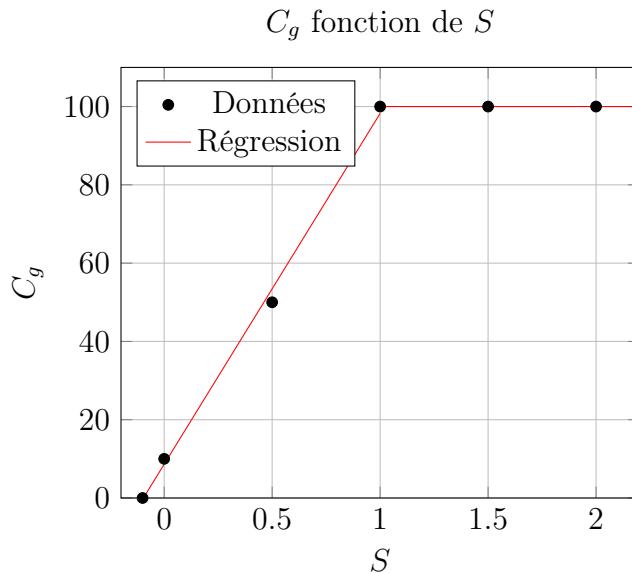


Au vu des valeurs de C_i en fonction de h_d , nous avons opté pour une simple régression linéaire simple. Le coefficient R^2 obtenu est de 0.996 avec le coefficient : 4895.872 et l'intercepte -2851.205 . Pour palier l'erreur d'approximation quand h_d est trop petit, on remplace les coûts négatifs par 0. On a donc finalement (en milliers d'euros) :

$$C_i(h_d) = \max\{0, 4895 \cdot h_d - 2851.05\}$$

3.1.2 Coût des dommages de la digue

Voici le tableau des coûts des dommages de la digue en fonction de la surverse S :

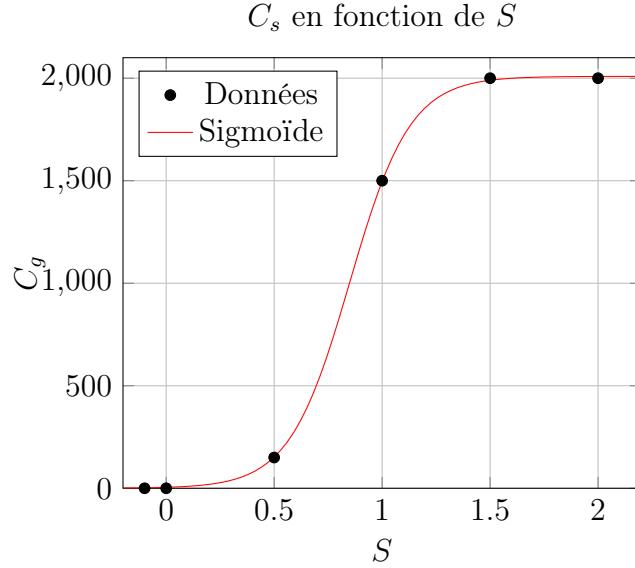


Au vu des données, nous avons fait une régression linéaire pour $S \leq 1$ pour avoir l'expression finale suivante :

$$C_g(S, h_d) = \left(1_{[-0.1, 1]}(S)(89.61 \cdot S + 8.63) + 100 \cdot 1_{]1, \infty[}(S) \right) C_i(h_d)$$

3.1.3 Coût des dommages du site

Comme C_s est croissant par rapport à S mais garde une composante aléatoire, il est difficile de donner une bonne approximation de la fonction. De plus, ayant peu de données et ne connaissant pas la nature de la composante aléatoire, il est difficile de savoir à quoi ressemble réellement la fonction. Néanmoins, nous savons les choses suivantes : $C_s(S) = 0$ pour $S < -0.1$ et $C_s(S) = 2000$ M€ pour $S \geq 2$. Une fonction sigmoïde permet de bien prendre en compte ces informations et nous avons donc opté pour une telle fonction bien qu'elle s'accorde trop avec les données connues.



Ainsi, ne connaissant pas la nature de la composante aléatoire, on retient :

$$C_s(S) = \frac{2009}{1 + \exp\{-(S - 0.85)/0.14\}}$$

3.2 Exploitation du modèle hydraulique

Avant de pouvoir résoudre notre problème d'optimisation, il faut relier la hauteur de digue h_d et le risque de surverse S et donc pouvoir connaître C_s et C_g en fonction de h_d .

Malheureusement l'équation $S = Z_c - Z_d = Z_v + H - h_d - Z_b$ reliant la hauteur h_d à la surverse S comporte des variables stochastiques.

Une première approche serait de faire une minimisation stochastique de la forme :

$$\arg \min_{h_d \in \mathbb{R}^+} \mathbb{E}_S[C_{c,moyenne}]$$

en utilisant les densités des lois aléatoires composants S . Cependant, cette méthode semble difficile d'accès comme S est un mélange de lois. Nous allons donc opter pour une discréétisation du problème par estimateur de Monte Carlo. Grâce aux données apportées par le modèle hydraulique et en générant $(x^{(1)}, \dots, x^{(K)}) \stackrel{\text{iid}}{\sim} Z_v + H - Z_b$ on a la minimisation suivante :

$$(P) \arg \min_{h_d \in \mathbb{R}^+} C_i(h_d) \left(1 + \frac{30}{100}\right) + \frac{30}{K} \sum_{k=1}^K (C_s(x^{(k)} - h_d) + C_g(x^{(k)} - h_d, h_d))$$

Grâce aux lois données dans le paragraphe 1.3 et la partie II de notre projet, nous pouvons simuler nos variables aléatoires et donc résoudre notre problème.

3.2.1 Génération de notre échantillon

Comme dit précédemment, il faut qu'on approche $X = Z_v + H - Z_b$ avec Z_v et H qui sont des variables aléatoires. Un dernier problème qui se pose est le choix du nombre d'échantillons K pour approximer X . Evidemment, plus K est grand, plus nous sommes précis. Cependant cela impact le temps de calcul de la simulation mais surtout celle de la résolution du problème de minimisation.

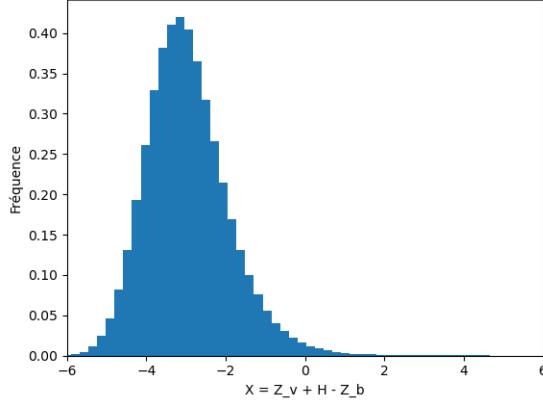


FIGURE 7 – Simulation de 10^6 échantillons de $X = Z_v + H - Z_b$

Remarque : Nous avons tronquer les valeurs extrêmes, pour avoir une valeur maximale à 10, pour des soucis de représentation de l'histogramme.

Informations statistiques : Voici certaines informations sur notre échantillon $\hat{X} = (x^{(1)}, \dots, x^{(K)}) \stackrel{\text{iid}}{\sim} Z_v + H - Z_b$ avec $K = 10^6$.

Moyenne : -2.96 , médiane : -3.06 , déviation standard : 1.20 , max : 146.96 , min : -6.28 . Pourcentage de valeurs au dessus de -0.1 : 1.25% .

Ce dernier pourcentage est important car il correspond à une surverse de -0.1 si $h_d = 0$ et c'est le seuil à partir duquel le coût des dégâts sur la digue est non nul.

3.3 Résolution du problème économique

Maintenant que nous avons l'expression des différents coûts en rapport avec la hauteur de la digue et la surverse, il nous reste plus qu'à résoudre algorithmiquement notre problème de minimisation.

Comme dit précédemment, on va pour cela générer $\hat{X} = (x^{(1)}, \dots, x^{(K)}) \stackrel{\text{iid}}{\sim} Z_v + H - Z_b$ avec $K = 10^6$ pour approximer notre fonction de coût puis trouver son minimum.

Pour tout k tel que $x^{(k)} < -0.1$, on a $C_s(x_k, h_d) = C_d(x_k, h_d) = 0$ quel que soit la hauteur de digue h_d . Pour réduire notre temps de calcul, on note $(y^{(1)}, \dots, y^{(L)}) := \{x^{(k)}; x^{(k)} \geq -0.1, k = 1 \dots K\}$ où L est la longueur de ce vecteur, et $p = L/K$ la proportion des valeurs utiles pour notre calcul de coût.

Finalement en exploitant ces informations on reformule une dernière fois (P) :

$$(P) \arg \min_{h_d \in \mathbb{R}^+} C_i(h_d) \left(1 + \frac{30}{100}\right) + p \frac{30}{L} \sum_{l=1}^L (C_s(y^{(l)} - h_d) + C_g(x^{(l)} - h_d, h_d))$$

Voyons maintenant notre fonction de coût grâce à notre formulation et notre estimateur de Monte Carlo.

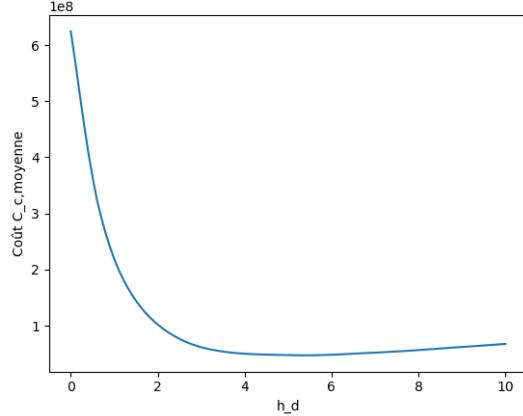


FIGURE 8 – Simulation de Monte Carlo pour le coût, $K = 10^6$

Remarque : Cette dernière formulation de (P) à réduit drastiquement le temps pour évaluer notre fonction de coût, passant de 3 minutes à 3 secondes. Ainsi comme notre estimateur de Monte Carlo est très rapide, nous n'avons pas besoin de réduire le nombre d'échantillons et de faire des compromis entre la précision et le temps de calcul.

Finalement, on trouve que (P) est égale à 4.9 mètres, c'est à dire que d'après le modèle économique étudié sur une période de 30 ans, il faudrait prendre $h_d = 4.9$ mètres.

A première vue, ce résultat semble haut sachant que le modèle hydraulique nous informe que la probabilité d'avoir une surverse quand $h_d > 3.9$ est quasiment nulle. Cependant, nous avons ici étudié le coût $C_{c,moyenne}$ sur une période de 30 ans, ce qui multiplie le risque de surverse sur cette période par 30. De plus, les coûts de dommages sont d'un ordre de grandeur de 10^2 fois plus grands que ceux de construction de la digue. Ainsi dans ce modèle économoqie, les risques de surverses doivent absolument être écartés. Cela se retranscrit aussi sur la forme de notre fonction de coût figure 7.