



M2 Data Science

# Rapport de Projet Pratique de la Fiabilité

Sarah OUAHAB, Sheïma MEBARKA, Erivan INAN

*Professeur :* Emmanuel REMY

Un rapport soumis en vue de l'accomplissement des exigences de  
l'ISUP et Sorbonne Université pour l'obtention du diplôme de  
*Master 2 Data Science*

14 mars 2025

## Table des matières

<b>1</b>	<b>Contexte industriel du problème</b>	<b>2</b>
<b>2</b>	<b>Détermination de la hauteur de la digue</b>	<b>3</b>
2.1	À partir des relevés de mesure historiques . . . . .	3
2.1.1	Méthode naïve . . . . .	4
2.1.2	Méthode des quantiles : estimation par moyenne . . . . .	4
2.1.3	Méthode des quantiles : estimation par régression . . . . .	5
2.1.4	Comparaison des résultats . . . . .	8
2.2	À partir du modèle hydraulique . . . . .	8
2.2.1	Prise en compte des incertitudes . . . . .	9
2.2.2	Estimation de $P(S > 0)$ par Simulation Monte-Carlo . . . . .	9
2.3	À partir du modèle économique . . . . .	14
2.3.1	Modélisation des coûts . . . . .	14
2.3.2	Prise en compte des incertitudes et choix de la simulation de Monte-Carlo . . . . .	14
2.3.3	Interpolation . . . . .	15
<b>3</b>	<b>Conclusion</b>	<b>20</b>
3.1	Résumé des différentes approches . . . . .	20
3.2	Justification du choix final . . . . .	20

## 1 Contexte industriel du problème

Les infrastructures situées à proximité de cours d'eau sont particulièrement vulnérables aux phénomènes de crue, qui peuvent entraîner des inondations.

Dans cette étude, une installation industrielle se trouve en bordure d'une rivière et doit faire face au risque de débordement des eaux en cas de crue exceptionnelle. La présence d'une digue de protection constitue une première ligne de défense contre ces événements, mais son efficacité repose sur un dimensionnement optimal prenant en compte les incertitudes hydrologiques et économiques.



Digue anti-inondation de la centrale nucléaire de Gravelines (nord de la France), source Nord Littoral

La gestion de ce risque soulève deux enjeux majeurs : d'une part, garantir la sûreté de l'installation industrielle en minimisant la probabilité de débordement et, d'autre part, optimiser les coûts liés à la digue en intégrant les coûts de construction et de maintenance de la digue aux pertes potentielles en cas d'inondation.

L'objectif est donc de proposer un dimensionnement optimal de la digue selon trois approches : en tenant compte des aléas hydrauliques, des données historiques de crues et des contraintes budgétaires, tout en s'appuyant sur des méthodes statistiques et de fiabilité.

Tous les fichiers code générés pendant l'élaboration de ce projet sont consultables sur [ce repository Github](#).

## 2 Détermination de la hauteur de la digue

### 2.1 À partir des relevés de mesure historiques

Dans cette première approche, on cherche à déterminer la hauteur optimale de la digue  $h_d$  afin de minimiser le risque d'inondation en nous basant uniquement sur les relevés historiques des hauteurs de crue.

On commence par réaliser une rapide analyse du jeu de données, en affichant un résumé des valeurs manquantes à l'aide de la fonction Python `describe()` :

```
print("\nRésumé des valeurs manquantes :")
print(df.isnull().sum())
print(df["Hauteur associée mesurée (en m)"].describe())
```

```
['Données débit-hauteur', 'Données coûts']
```

```
Résumé des valeurs manquantes :
Année                                0
Débit mesuré maximal annuel (en m3/s)  0
Hauteur associée mesurée (en m)       26
dtype: int64
count    123.000000
mean      3.896748
std       1.272337
min       1.000000
25%       3.100000
50%       3.900000
75%       4.750000
max       7.200000
Name: Hauteur associée mesurée (en m), dtype: float64
```

FIGURE 1 – Statistiques descriptives du jeu de données

On a 26 valeurs manquantes dans la colonne hauteur de crue. Nous allons traiter ces valeurs manquantes de deux manières :

- Une première méthode, la **méthode naïve**, consistant à ignorer les valeurs censurées et se basant uniquement sur les observations disponibles.
- Une seconde, la **méthode des quantiles avec estimation par moyenne**, consistant à remplacer les valeurs censurées par la moyenne des hauteurs connues avant le calcul des quantiles.
- Une troisième et dernière, la **méthode des quantiles avec estimation par régression**, qui consiste en la modélisation de la relation entre le débit et la hauteur pour estimer les valeurs censurées avant de calculer les quantiles.

### 2.1.1 Méthode naïve

Dans cette première approche, on supprime les valeurs censurées et on travaille uniquement avec les hauteurs de crue observées. La hauteur de la digue  $h_d$  est alors définie comme la valeur maximale observée, augmentée d'une marge de sécurité de 5% afin de prendre en compte les incertitudes liées à cette méthode.

```
df_naive = df.dropna(subset=["Hauteur associée mesurée (en m)"])

h_max = df_naive["Hauteur associée mesurée (en m)"].max()

# Marge de sécurité de 5% (ajout)
h_empirique = h_max * 1.05

print(f"Hauteur maximale observée : {h_max:.2f} m")
print(f"Hauteur empirique (max + 5%) : {h_empirique:.2f} m")
```

On obtient les résultats suivants :

```
Hauteur maximale observée : 7.20 m
Hauteur empirique (max + 5%) : 7.56 m
```

FIGURE 2 – Résultats de l'approche naïve

L'idée derrière cette approche est de fixer une hauteur de digue garantissant la protection contre toutes les crues enregistrées, tout en ajoutant une marge de sécurité de 5% pour anticiper d'éventuelles variations futures.

Cette méthode repose sur l'hypothèse que la hauteur maximale observée est un bon indicateur du risque extrême et qu'une simple majoration permettrait d'éviter toute inondation.

Toutefois, cette méthode présente une limite majeure : elle **ignore** totalement les valeurs censurées. Cela peut donc mener à une **surestimation** de la hauteur  $h_d$  si la crue maximale observée est exceptionnelle, ou au contraire à une **sous-estimation** si les crues les plus extrêmes sont absentes des données disponibles.

### 2.1.2 Méthode des quantiles : estimation par moyenne

Dans cette approche, les valeurs censurées sont estimées par la moyenne des hauteurs observées :

$$\bar{h} = \frac{1}{n} \sum_{i=1}^n h_i$$

Une fois cette estimation réalisée, nous appliquons les quantiles 95% et 99% pour définir un seuil en fonction du risque d'inondation acceptable :

$$Q_{95} = F^{-1}(0.95) \quad \text{et} \quad Q_{99} = F^{-1}(0.99)$$

où  $F^{-1}$  est la fonction quantile empirique de la distribution des hauteurs après l'imputation des valeurs manquantes. Ces quantiles sont utilisés pour définir des seuils en fonction du risque d'inondation acceptable.

```
mean_height = df["Hauteur associée mesurée (en m)"].mean()
# Remplissage des valeurs manquantes avec la moyenne
df_filled = df.copy()
df_filled["Hauteur associée mesurée (en m)"].fillna(mean_height, inplace=True)

# Calcul des quantiles 95% et 99% après avoir rempli les valeurs manquantes
h_q95 = df_filled["Hauteur associée mesurée (en m)"].quantile(0.95)
h_q99 = df_filled["Hauteur associée mesurée (en m)"].quantile(0.99)

print(f"Quantile 95% après estimation : {h_q95:.2f} m")
print(f"Quantile 99% après estimation : {h_q99:.2f} m")
```

On obtient les résultats suivants :

```
Quantile 95% après estimation : 5.86 m
Quantile 99% après estimation : 6.70 m
```

FIGURE 3 – Résultats de la méthode des quantiles avec estimation par moyenne

On peut ensuite afficher les intervalles de confiance associés à ces quantiles :

```
Quantile 95% après estimation : 5.86 m
IC pour le quantile 95% : [5.41 m ; 6.31 m]
Quantile 99% après estimation : 6.70 m
IC pour le quantile 99% : [6.02 m ; 7.38 m]
```

FIGURE 4 – Intervalles de confiance après estimation associés aux quantiles 95% et 99%

En estimant les valeurs censurées par la moyenne des observations disponibles, nous avons supposé que ces valeurs étaient réparties de manière homogène, sans être nécessairement plus élevées que les autres.

Cependant, cette approche lisse les extrêmes, ce qui peut conduire à une **sous-estimation** des quantiles élevés. On constate que le quantile 99% issu de cette méthode (6.70 m) est significativement inférieur à l'approche naïve, traduisant un possible **biais** vers le bas.

Cela nous a amené à chercher une alternative permettant une estimation plus fidèle des valeurs censurées.

### 2.1.3 Méthode des quantiles : estimation par régression

Une solution consiste alors à modéliser la relation entre le débit et la hauteur de crue afin d'estimer ces valeurs.

L'idée principale est que le débit  $Q$  et la hauteur  $H$  sont fortement corrélés : lorsqu'un débit important est observé, la hauteur de crue est généralement plus élevée. En exploitant cette relation, nous pouvons estimer les hauteurs manquantes à partir des débits mesurés.

**Modélisation par régression linéaire** On suppose une relation linéaire entre le débit maximal annuel  $Q$  et la hauteur associée  $H$ , exprimée sous la forme d'un modèle de régression linéaire :

$$H = \alpha + \beta Q + \varepsilon$$

où :

- $\alpha$  est l'ordonnée à l'origine,
- $\beta$  est le coefficient de régression représentant l'effet du débit sur la hauteur,
- $\varepsilon$  est un terme d'erreur supposé centré.

L'estimation des paramètres  $\alpha$  et  $\beta$  est obtenue par la méthode des moindres carrés ordinaires (MCO) :

$$\hat{\beta} = \frac{\sum_{i=1}^n (Q_i - \bar{Q})(H_i - \bar{H})}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \quad \text{et} \quad \hat{\alpha} = \bar{H} - \hat{\beta}\bar{Q}$$

où  $\bar{Q}$  et  $\bar{H}$  sont les moyennes des valeurs observées du débit et de la hauteur, respectivement.

**Estimation des valeurs censurées** Une fois le modèle ajusté sur les données observées, on utilise les valeurs de débit  $Q$  pour estimer les hauteurs de crue correspondantes :

$$\hat{H}_{\text{manquants}} = \hat{\alpha} + \hat{\beta}Q$$

**Calcul des quantiles** Après imputation des valeurs censurées, on calcule les quantiles empiriques à 95% et 99% de la distribution des hauteurs :

$$Q_{95}^{(\text{reg})} = F^{-1}(0.95) \quad \text{et} \quad Q_{99}^{(\text{reg})} = F^{-1}(0.99)$$

Ces quantiles permettent de fixer un seuil en fonction du risque d'inondation acceptable.

```
from sklearn.linear_model import LinearRegression

# On prend uniquement les lignes avec des valeurs observées pour entraîner le modèle
df_observed = df.dropna(subset=["Hauteur associée mesurée (en m)"])

# Modèle de régression entre le débit et la hauteur observée
X_train = df_observed["Débit mesuré maximal annuel (en m3/s)"].values.reshape(-1, 1)
y_train = df_observed["Hauteur associée mesurée (en m)"].values

# Entraînement du modèle de régression linéaire
model = LinearRegression()
model.fit(X_train, y_train)

# Prédiction des valeurs censurées
df_missing = df[df["Hauteur associée mesurée (en m)"].isna()]
X_missing = df_missing["Débit mesuré maximal annuel (en m3/s)"].values.reshape(-1, 1)

df.loc[df["Hauteur associée mesurée (en m)"].isna(), "Hauteur associée mesurée (en m)"] =
    model.predict(X_missing)
```

```

# Calcul des quantiles après estimation
h_q95_r = df["Hauteur associée mesurée (en m)"].quantile(0.95)
h_q99_r = df["Hauteur associée mesurée (en m)"].quantile(0.99)

print(f"Quantile 95% après estimation (régression) : {h_q95_r:.2f} m")
print(f"Quantile 99% après estimation (régression) : {h_q99_r:.2f} m")

```

On obtient les résultats suivants :

```

Quantile 95% après estimation (régression) : 6.36 m
Quantile 99% après estimation (régression) : 7.37 m

```

FIGURE 5 – Résultats de la méthode des quantiles avec estimation par régression

On peut ensuite afficher les intervalles de confiance associés à ces quantiles :

```

Quantile 95% après estimation (régression) : 6.36 m
IC pour le quantile 95% : [5.95 m ; 6.77 m]
Quantile 99% après estimation (régression) : 7.37 m
IC pour le quantile 99% : [6.66 m ; 8.09 m]

```

FIGURE 6 – Intervalles de confiance après estimation associés aux quantiles 95% et 99%

Ces résultats indiquent que, bien que les estimations centrales fournissent une indication de la hauteur critique à considérer, une marge d'incertitude subsiste, soulignant la nécessité d'intégrer ces variations dans la prise de décision pour garantir un niveau de protection adapté.

Pour améliorer l'estimation des valeurs censurées, nous avons cherché à exploiter une **relation physique** existante entre le débit et la hauteur de crue, plutôt que de leur attribuer une valeur fixe. L'idée est d'utiliser une **régression linéaire** pour modéliser cette relation à partir des données observées, puis d'**extrapoler** les valeurs censurées.

Cette approche repose sur l'hypothèse que le débit maximal annuel est un bon prédicteur de la hauteur de crue associée, ce qui est cohérent avec la physique des crues. En estimant les hauteurs manquantes à partir de cette relation, nous avons obtenu une meilleure représentation des valeurs extrêmes, **évitant l'effet de lissage** de la méthode d'estimation par moyenne.



## 2.1.4 Comparaison des résultats

Approche	Prise en compte des censures ?	Hauteur ( $h_d$ ) obtenue (m)
Méthode naïve (Max + 5%)	❌ Non	7.56
Quantile 95% (moyenne)	✅ Oui (Estimation moyenne)	5.86
Quantile 99% (moyenne)	✅ Oui (Estimation moyenne)	6.70
Quantile 95% (régression)	✅ Oui (Estimation régression)	6.36
Quantile 99% (régression)	✅ Oui (Estimation régression)	7.37

TABLE 1 – Tableau comparatif des différentes approches

Les trois approches testées aboutissent à des estimations de  $h_d$  comprises entre 5.86 m et 7.56 m, ce qui met en évidence l'impact du traitement des valeurs censurées.

La méthode naïve et l'approche par régression donnent des valeurs relativement proches (7.56 m vs 7.37 m). On peut donc suggérer que la modélisation du lien entre débit et hauteur permet d'approcher les crues extrêmes de manière cohérente avec les observations historiques.

En revanche, l'approche par moyenne sous-estime  $h_d$  (6.70 m), car elle lisse les valeurs censurées sans refléter la distribution réelle des hauteurs extrêmes.

Ainsi, le quantile 99% estimé par régression (7.37 m) constitue un **bon compromis**, en tenant compte des censures tout en restant plus réaliste que la méthode naïve. Si une sécurité absolue est requise, la méthode naïve reste plus conservatrice. Le choix final dépend donc du niveau de protection souhaité.

## 2.2 À partir du modèle hydraulique

Après avoir estimé la hauteur optimale de la digue à partir des données historiques, cette seconde approche repose sur un modèle hydraulique **intégrant les lois physiques** de l'écoulement et les **incertitudes** environnementales.

L'objectif est de déterminer la hauteur  $h_d$  en minimisant la probabilité de surverse  $P(S > 0)$ , c'est-à-dire la probabilité que l'eau dépasse la digue :

$$P(S > 0) = P(H > h_d + Z_b - Z_v)$$

où :

- H est calculé à partir du débit de crue Q selon l'équation :  $H = \left( \frac{Q}{K_s \sqrt{\frac{Z_m - Z_v}{L} \cdot B}} \right)^{\frac{3}{5}}$ ,
- Q est le débit maximal annuel (m³/s), qui varie selon les crues,
- $K_s$  est un coefficient de frottement lié aux caractéristiques du terrain,
- $Z_m$  et  $Z_v$  sont respectivement les côtés amont et aval,
- L est la longueur du tronçon de rivière étudié,
- B est la largeur du lit du fleuve.

Il faut donc choisir la hauteur optimale  $h_d$  de sorte à **minimiser**  $P(S > 0)$ , tout en tenant compte des **incertitudes** affectant les variables du modèle.

### 2.2.1 Prise en compte des incertitudes

Dans un environnement naturel, les paramètres du modèle ne sont pas parfaitement connus et varient au cours du temps. Plusieurs sources d'incertitude influencent directement la hauteur d'eau et la probabilité de surverse.

D'une part, la variabilité naturelle du débit  $Q$ , du coefficient de frottement  $K_s$  et des cotes du lit du fleuve ( $Z_m$  et  $Z_v$ ) est liée aux fluctuations climatiques, météorologiques et hydrologiques.

En effet, le débit d'un fleuve est influencé par des événements extrêmes, tels que les fortes précipitations ou la fonte des neiges, ce qui rend ses valeurs imprévisibles d'une année à l'autre. De même, l'état du lit du fleuve évolue en raison de l'érosion et du dépôt de sédiments, modifiant ainsi les conditions d'écoulement.

D'autre part, il existe des incertitudes de mesure et de modélisation.

Afin d'intégrer ces incertitudes dans notre analyse, nous modélisons chaque variable aléatoire selon les distributions de probabilités repérées dans le tableau ci-dessous :

Variable	Type de Loi	Paramètre 1	Paramètre 2
Débit ( $Q$ ) ( $\text{m}^3/\text{s}$ )	Gumbel	1013 (mode)	558 (échelle)
Coefficient de frottement ( $K_s$ )	Normale	30 (espérance)	7.5 (écart-type)
Cote aval ( $Z_v$ ) (m NGF)	Triangulaire	50 (espérance)	1 (demi-étendue)
Cote amont ( $Z_m$ ) (m NGF)	Triangulaire	55 (espérance)	1 (demi-étendue)

TABLE 2 – Distributions pour les variables d'entrée considérées comme aléatoires

Les autres paramètres du modèle sont considérés comme déterministes, notamment la longueur du tronçon ( $L = 5000$  m), la largeur du fleuve ( $B = 300$  m) et la cote de la berge ( $Z_b = 55.5$  m).

### 2.2.2 Estimation de $P(S > 0)$ par Simulation Monte-Carlo

Le modèle hydraulique implique des relations non-linéaires entre les variables d'entrée et la surcharge  $S$ . En raison de cette complexité, il est difficile d'obtenir une formule analytique exacte pour estimer la probabilité  $P(S > 0)$ .

Pour pallier à cela, on réalise une simulation de Monte-Carlo, qui consiste à générer un grand nombre de scénarios aléatoires en échantillonnant  $Q$ ,  $K_s$ ,  $Z_v$ ,  $Z_m$  selon leurs distributions respectives, puis à calculer  $S$  pour chaque réalisation.

$$Q \sim \mathcal{G}(\mu_Q, \sigma_Q), \quad K_s \sim \mathcal{N}(\mu_{K_s}, \sigma_{K_s}), \quad Z_m \sim \mathcal{T}(c_m, a_m, b_m), \quad Z_v \sim \mathcal{T}(c_v, a_v, b_v)$$

où :

—  $\mathcal{G}(\mu, \sigma)$  représente une loi de Gumbel de paramètres  $\mu$  et  $\sigma$ ,

- $\mathcal{N}(\mu, \sigma)$  est une loi normale de moyenne  $\mu$  et écart-type  $\sigma$ ,
- $\mathcal{T}(c, a, b)$  est une loi triangulaire de paramètres  $c$  (mode),  $a$  (min) et  $b$  (max).

Ensuite, on calcule la hauteur d'eau  $H$  pour chaque réalisation en utilisant la relation hydraulique suivante :

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

où :

- $L$  est la longueur de la rivière,
- $B$  est la largeur du lit,
- $Z_b$  est l'altitude du lit de la rivière.

**Optimisation de la hauteur critique  $h_d$**  On cherche à optimiser la hauteur  $h_d$  de la digue en minimisant la probabilité de surcharge définie comme :

$$P_{\text{surverse}}(h_d) = P(H - h_d - (Z_b - Z_v) > 0)$$

On réalise une recherche numérique en testant différentes valeurs de  $h_d$  dans un intervalle donné. L'optimum est obtenu en sélectionnant la valeur de  $h_d$  qui minimise l'écart avec le seuil critique :

$$h_d^{\text{opt}} = \arg \min_{h_d} |P_{\text{surverse}}(h_d) - \alpha|$$

où  $\alpha$  représente un seuil acceptable de probabilité d'inondation.

**Choix du nombre de simulations** Nous avons fixé  $n = 10000$  simulations pour cette seconde approche, en partie car l'estimation de la hauteur de  $h_d$  repose sur un quantile et selon la règle  $n = 10(k + 2)$ , avec  $k = 2$  pour un quantile, cela garantit une incertitude d'environ 10%, assurant une précision suffisante, mais qui reste raisonnable en coût computationnel.

Cette méthode présente plusieurs avantages :

- Elle ne nécessite pas d'hypothèse restrictive sur la forme de la distribution de  $S$ , contrairement aux méthodes analytiques.
- Elle gère la complexité du modèle et sa non-linéarité.
- Elle prend en compte les valeurs extrêmes.
- Elle capture directement les incertitudes en respectant les distributions de chaque paramètre.
- Elle permet d'optimiser la hauteur  $h_d$  en testant plusieurs valeurs et en observant leur impact sur la probabilité de surverse.

```

from scipy.stats import gumbel_r, norm, triang

# Paramètres fixes
L, B, Z_b = 5000, 300, 55.5

# Lois de probabilité des variables aléatoires
n_simulations = 100000

Q_sim = gumbel_r.rvs(loc=1013, scale=558, size=n_simulations)
Ks_sim = norm.rvs(loc=30, scale=7.5, size=n_simulations)
Zm_sim = triang.rvs(c=0.5, loc=54, scale=2, size=n_simulations)
Zv_sim = triang.rvs(c=0.5, loc=49, scale=2, size=n_simulations)

# Calcul des hauteurs d'eau H
H_sim = (Q_sim / (Ks_sim * np.sqrt(((Zm_sim - Zv_sim)/L))*B)) ** (3/5)

# Optimisation de h_d
h_d_values = np.arange(1, 9, 0.1)
P_surverse_values = [np.mean(H_sim - h_d - (Z_b - Zv_sim) > 0) for h_d in h_d_values]

# Sélection de la hauteur optimale
h_d_optimal = h_d_values[np.argmin(np.abs(np.array(P_surverse_values)))]
print(f"Hauteur optimale de la digue : {h_d_optimal:.2f} m")
    
```

À l'issue duquel on obtient :

Hauteur optimale de la digue : 7.60 m

Une fois la simulation effectuée, on peut afficher la distribution probabiliste des hauteurs de crues :

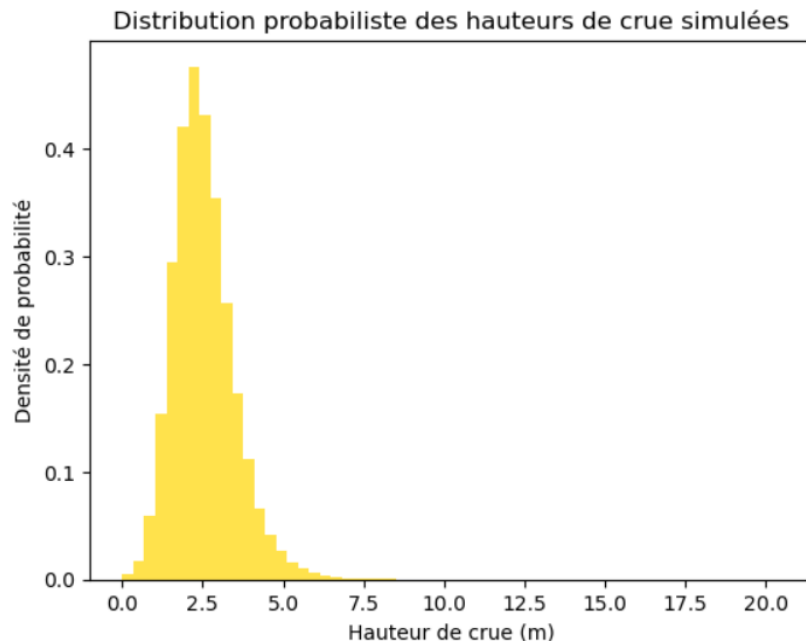


FIGURE 7 – Distribution probabiliste des hauteurs de crue obtenues par simulation

La distribution des hauteurs de crue simulées met en évidence une majorité de crues comprises entre 2 m et 5 m, avec une faible probabilité d'événements extrêmes dépassant 7 m. Cette répartition permet de mieux appréhender la variabilité des crues et d'évaluer le niveau de protection nécessaire. En complément des quantiles et des intervalles de confiance, elle offre une vision globale du risque de surverse et valide la pertinence de l'approche Monte-Carlo pour représenter l'incertitude hydraulique. La présence de crues extrêmes, bien que peu fréquentes, justifie une hauteur de digue optimisée, assurant un équilibre entre sécurité et maîtrise des coûts.

Afin d'évaluer la variabilité de la hauteur optimale  $h_d$ , nous utilisons une méthode de *bootstrap* avec  $n_{\text{bootstrap}}$  itérations. Pour chaque itération, nous générons un nouvel échantillon de tailles  $n$  à partir des valeurs simulées  $Q, K_s, Z_m, Z_v$  en appliquant un rééchantillonnage avec remise. Cela permet de recalculer la hauteur d'eau  $H$  et d'estimer une distribution empirique de la hauteur optimale de la digue  $h_d$ .

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

où  $b$  désigne une itération bootstrap.

Ensuite, on recalcule la probabilité de surverse pour chaque hauteur de digue candidate :

$$P_{\text{surverse}}^{(b)}(h_d) = P(H^{(b)} - h_d - (Z_b - Z_v^{(b)}) > 0)$$

et on sélectionne la hauteur optimale  $h_d^{(b)}$  comme :

$$h_d^{(b)} = \arg \min_{h_d} |P_{\text{surverse}}^{(b)}(h_d) - \alpha|$$

**Calcul de l'intervalle de confiance à 95%** Une fois la distribution bootstrap obtenue, on estime l'intervalle de confiance à 95% pour  $h_d$  en supposant une distribution normale des estimations :

$$IC_{95\%}(h_d) = \left[ h_d^{\text{opt}} - z_{\alpha} \frac{\sigma_{h_d}}{\sqrt{n_{\text{bootstrap}}}} ; h_d^{\text{opt}} + z_{\alpha} \frac{\sigma_{h_d}}{\sqrt{n_{\text{bootstrap}}}} \right]$$

où :

- $z_{\alpha}$  est le quantile de la loi normale standard à  $\alpha = 0.975$  (soit  $z_{\alpha} \approx 1.96$ ),
- $\sigma_{h_d}$  est l'écart-type empirique des valeurs de  $h_d$  obtenues par bootstrap.

```

n_simulations = 10000 # Nombre d'échantillons Monte Carlo
n_bootstrap = 1000
h_d_bootstrap = []

for _ in range(n_bootstrap):
    # Recalculer H_sim avec un échantillonnage aléatoire (bootstrap)
    Q_resample = np.random.choice(Q_sim, size=n_simulations, replace=True)
    Ks_resample = np.random.choice(Ks_sim, size=n_simulations, replace=True)

```

```
Zm_resample = np.random.choice(Zm_sim, size=n_simulations, replace=True)
Zv_resample = np.random.choice(Zv_sim, size=n_simulations, replace=True)

H_resample = (Q_resample / (Ks_resample * np.sqrt(((Zm_resample - Zv_resample)/L))*B))
              ** (3/5)

P_surverse_resample = [np.mean(H_resample - h_d - (Z_b - Zv_resample) > 0)
                        for h_d in h_d_values]
h_d_bootstrap.append(h_d_values[np.argmin(np.abs(np.array(P_surverse_resample)))])

# Calcul des IC à 95%
z_alpha = norm.ppf(0.975) # 1.96 pour IC à 95%
sigma_h_d = np.std(h_d_bootstrap)
IC_h_d_lower = h_d_optimal - z_alpha * sigma_h_d / np.sqrt(n_bootstrap)
IC_h_d_upper = h_d_optimal + z_alpha * sigma_h_d / np.sqrt(n_bootstrap)

# Résultats
print(f"Hauteur optimale de la digue : {h_d_optimal:.2f} m")
print(f"IC à 95% pour h_d optimal : [{IC_h_d_lower:.2f} m ; {IC_h_d_upper:.2f} m]")
```

On obtient les résultats suivants :

```
Hauteur optimale de la digue : 7.60 m
IC à 95% pour h_d optimal : [7.52 m ; 7.68 m]
```

FIGURE 8 – Intervalle de confiance associé au quantile 95%

L'intervalle de confiance, relativement étroit ( $\pm 0.08$  m autour de la hauteur optimale), témoigne de la précision de l'estimation, avec 95 % des simulations Monte-Carlo situées dans cette plage de valeurs. La hauteur optimale estimée à 7.60 m assure ainsi une protection efficace contre le risque de surverse.

Les deux approches développées jusqu'ici aboutissent à des estimations relativement proches de la hauteur optimale  $h_d$  de la digue - comprises entre 5.86 m et 7.56 m - en fonction de la méthode utilisée pour traiter les valeurs censurées.

L'approche hydraulique par Monte-Carlo, quant à elle, estime cette hauteur à 7.60 m, une valeur très proche de celle obtenue avec la méthode naïve. Cette simulation probabiliste capture efficacement la distribution des hauteurs extrêmes en intégrant la variabilité des paramètres hydrauliques et les incertitudes liées aux conditions d'écoulement. Contrairement à l'approche historique, elle ne se limite pas aux observations passées mais modélise également les phénomènes physiques influençant les crues, offrant ainsi une représentation plus complète du risque de surverse.

Cette nouvelle approche permet donc une évaluation plus robuste du risque de surverse.

## 2.3 À partir du modèle économique

L'objectif de cette approche est de déterminer la hauteur optimale  $h_d$  de la digue, en prenant en compte trois composantes essentielles :

- Le coût d'investissement et de maintenance,
- Le coût des dommages liés aux inondations,
- Les incertitudes hydrauliques.

Cette approche se distingue des précédentes en intégrant l'aspect économique dans la prise de décision. Contrairement à l'approche historique, qui analyse les crues passées, et à l'approche hydraulique, qui vise à minimiser le risque de surverse seulement, on cherche ici à minimiser le coût total moyen annuel sur une période de  $T = 30$  ans, en équilibrant protection et viabilité économique.

### 2.3.1 Modélisation des coûts

Le coût total de la digue est composé de deux éléments :

- $C_t(T)$  : le coût total d'investissement et de maintenance sur 30 ans
- $C_d$  : le coût total de dommage (pour une année)

L'objectif est de minimiser le coût moyen annuel :

$$C_{c,moyenne} = C_t(T) + \sum_{j=1}^T C_{d,j}(S_j, h_d)$$

avec :

- $C_t(T)$  le coût d'investissement et de maintenance sur la durée  $T$  :  
 $C_t(T) = C_i(h_d) + T.C_m(h_d)$ , où  $C_i(h_d)$  est le coût initial de construction et  $C_m(h_d) = 0.01.C_i$  est le coût annuel de maintenance (fixé à 1% du coût initial).
- $C_d$  le coût annuel des dommages en cas de surverse :  
 $C_d = C_s + C_g = C_s(S) + C_g(S, h_d)$ , où  $C_s(S)$  est le coût des dommages aux infrastructures et  $C_g(S, h_d)$  est le coût des réparations de la digue après une inondation.

L'idée est donc de trouver la hauteur de digue  $h_d$  qui minimise ce coût total.

### 2.3.2 Prise en compte des incertitudes et choix de la simulation de Monte-Carlo

Tout comme dans l'approche hydraulique, cette méthode repose sur le modèle physique des crues, soumis à d'importantes incertitudes liées aux conditions naturelles. La simulation de Monte-Carlo est donc utilisée pour prendre en compte ces variabilités aléatoires, en générant un grand nombre de scénarios possibles afin de représenter l'ensemble des configurations hydrauliques envisageables.

La principale différence avec l'approche précédente réside dans l'intégration de l'optimisation économique du coût de la digue. En réalisant 10 000 simulations sur une période

de 30 ans, cette approche permet d'estimer non seulement une distribution réaliste des hauteurs de crue, mais aussi leur impact financier, en évaluant à la fois les coûts d'investissement et les dommages potentiels.

Les variables aléatoires utilisées dans le modèle suivent les distributions suivantes :

$$Q \sim \mathcal{G}(\mu_Q, \sigma_Q), \quad K_s \sim \mathcal{N}(\mu_{K_s}, \sigma_{K_s}), \quad Z_m \sim \mathcal{T}(c_m, a_m, b_m), \quad Z_v \sim \mathcal{T}(c_v, a_v, b_v)$$

où les variables sont définies comme précédemment en section 2.2.2

Dans ce cas, la simulation Monte-Carlo présente plusieurs avantages :

- **Robustesse** : Elle permet de modéliser précisément les risques associés aux inondations.
- **Flexibilité** : Elle permet d'intégrer des distributions de probabilité adaptées aux variables hydrauliques et économiques.

### 2.3.3 Interpolation

L'interpolation joue un rôle clé dans cette approche en permettant d'estimer des valeurs intermédiaires qui ne sont pas directement disponibles dans les données initiales. Elle est utilisée à deux niveaux distincts :

#### Interpolation des coûts de dommages

Les valeurs de surverse  $S$  obtenues à partir des simulations de Monte-Carlo ne correspondent pas nécessairement aux hauteurs de surverse définies dans le tableau des coûts économiques. Afin d'estimer les coûts de dommages au site ( $C_s$ ) et à la digue ( $C_g$ ) pour n'importe quelle valeur de  $S$ , une interpolation linéaire est appliquée :

$$C_s(S) = \text{interp1d}(S, S_{\text{réf}}, C_s^{\text{réf}})$$

$$C_g(S, h_d) = \text{interp1d}(S, S_{\text{réf}}, C_g^{\text{réf}}) \times C_i(h_d)$$

où  $S_{\text{réf}}, C_s^{\text{réf}}, C_g^{\text{réf}}$  sont les valeurs tabulées de la hauteur de surverse et des coûts associés.

#### Interpolation pour l'optimisation de la hauteur de la digue

Les hauteurs de digue disponibles dans le fichier Excel sont discrètes, et la hauteur optimale  $h_d$  ne correspond pas toujours à l'une de ces valeurs. Pour affiner la recherche du minimum du coût total moyen, une interpolation cubique est réalisée avec `interp1d()` définie par :

$$C_{c,\text{moyenne}}^{\text{interp}}(h_d) = \text{interp}_{\text{cubique}}(h_d, h_d^{\text{tab}}, C_{c,\text{moyenne}}^{\text{tab}})$$

Cette approche permet de transformer la relation entre la hauteur de la digue et le coût en une fonction continue, facilitant ainsi la recherche d'une hauteur optimale plus précise, non contrainte aux seules valeurs tabulées.



Dans le code ci-dessous, on va donc rechercher :

$$h_d^{\text{opt}} = \arg \min_{h_d} C_{c,\text{moyenne}}^{\text{interp}}(h_d)$$

```
from scipy.interpolate import interp1d
from scipy.optimize import minimize_scalar

# deuxième feuille contenant les deux tableaux
df_economie = pd.read_excel(xls, sheet_name=xls.sheet_names[1], header=None)

# Identification de la séparation des tableaux
sep_index = df_economie[df_economie[0]=="Coûts des dommages liés à la surverse"].index[0]

# premier tableau (= coût de la digue)
df_digue = df_economie.iloc[1:sep_index, :].dropna(how="all")
df_digue.columns = ["Hauteur de la digue (m)", "Coût du linéaire (€/m)",
                    "Longueur de la digue (m)", "Coût total d'investissement (k€)"]
df_digue = df_digue.iloc[1:].reset_index(drop=True)

# second tableau (= coût des dommages)
df_dommages = df_economie.iloc[sep_index+2:, :].dropna(how="all")
df_dommages = df_dommages.dropna(axis=1, how='all')
df_dommages.columns = ["Hauteur de la surverse (m)", "Coût des dommages au site (M€)",
                       "Coût des dommages à la digue (%)"]
df_dommages["Hauteur de la surverse (m)"] =
    df_dommages["Hauteur de la surverse (m)"].replace(">2", 2.1).astype(float)

# Conversion en tableaux exploitables
hauteurs_digue = df_digue["Hauteur de la digue (m)"].astype(float).values
Ci_values = df_digue["Coût total d'investissement (k€)"].astype(float).values * 1000
Cm_values = 0.01 * Ci_values

hauteurs_surverse = df_dommages["Hauteur de la surverse (m)"].values
Cs_values = df_dommages["Coût des dommages au site (M€)"].astype(float).values * 1e6
Cg_percent = df_dommages["Coût des dommages à la digue (%)"].astype(float).values / 100

# Paramètres du modèle
T = 30
n_simulations = 10000
L, B, Zb = 5000, 300, 55.5

# Fonctions pour générer les échantillons
def generate_samples(n):
    return (
        np.random.gumbel(1013, 558, n),
        np.random.normal(30, 7.5, n),
        np.random.triangular(49, 50, 51, n),
        np.random.triangular(54, 55, 56, n),
    )

# Fonction pour calculer H et S
def calc_hydraulique(Q, Ks, Zv, Zm, h_d):
    H = (Q / (Ks * np.sqrt((Zm - Zv) / L) * B))**(3/5)
    S = np.maximum(Zv + H - h_d - Zb, 0)
    return S
```

```

# Fonction coût de dommage
def C_dommage(S, Ci):
    return (
        np.interp(S, hauteurs_surverse, C_s_values),
        np.interp(S, hauteurs_surverse, C_g_percent) * Ci,
    )

# Optimisation de la hauteur de la digue
Cc_moyenne = []

for h_d, Ci, Cm in zip(hauteurs_digue, Ci_values, Cm_values):
    Ct_T = Ci + (T * Cm)

    # Génération des échantillons
    Q_sim, Ks_sim, Zv_sim, Zm_sim = generate_samples(n_simulations * T)
    Q_sim, Ks_sim, Zv_sim, Zm_sim =
        [x.reshape(T, -1) for x in (Q_sim, Ks_sim, Zv_sim, Zm_sim)]

    # Calcul de S
    S_sim = calc_hydraulique(Q_sim, Ks_sim, Zv_sim, Zm_sim, h_d)
    S_sim = S_sim[~np.isnan(S_sim)]

    # Coût des dommages
    C_s_sim, C_g_sim = C_dommage(S_sim, Ci)
    Cd_T = np.sum(C_s_sim + C_g_sim) / T

    # Coût total moyen
    Cc_moyenne.append((Ct_T + Cd_T) / T)

# Recherche de la hauteur optimale
interp_func = interp1d(hauteurs_digue[:len(Cc_moyenne)], Cc_moyenne, kind='cubic')
res = minimize_scalar(interp_func, bounds=(min(hauteurs_digue), max(hauteurs_digue)),
                      method='bounded')
h_d_optimal = res.x

# Résultats
plt.figure(figsize=(8, 5))
plt.plot(hauteurs_digue[:len(Cc_moyenne)], Cc_moyenne, label="Coût moyen annuel",
         color='b', marker='o')
if h_d_optimal:
    plt.axvline(h_d_optimal, color='r', linestyle="--",
                label=f"Hauteur optimale: {h_d_optimal:.2f} m")

plt.xlabel("Hauteur de la digue (m)")
plt.ylabel("Coût moyen annuel (€)")
plt.legend()
plt.title("Optimisation de la hauteur de la digue")
plt.grid(True)
plt.show()

print(f"Hauteur optimale de la digue: {h_d_optimal:.2f} m")

```

Après exécution de l'optimisation, le code renvoie :

Hauteur optimale de la digue : 7.34 m

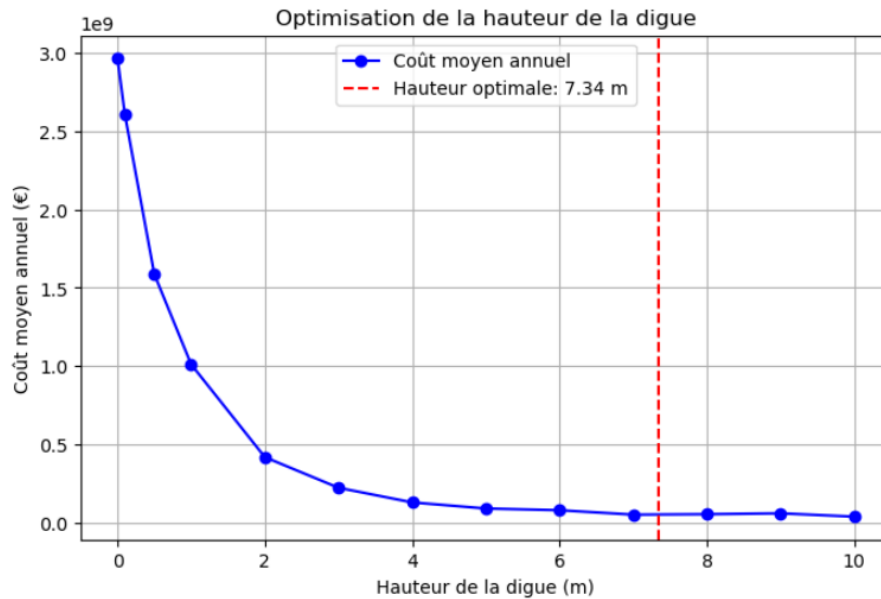


FIGURE 9 – Hauteur de la digue en fonction du coût moyen annuel en euros (échelle 10<sup>9</sup>)

Enfin, afin d'évaluer l'incertitude sur  $h_d$ , on calcule un intervalle de confiance à 95% en utilisant une approche bootstrap Monte-Carlo. L'intervalle de confiance est donné par :

$$IC(h_d) = h_d \pm z_{\alpha/2} \cdot \frac{\sigma_{h_d}}{\sqrt{n}}$$

où  $\sigma_{h_d}$  est l'écart-type des hauteurs optimales obtenues par simulation.

Cette approche permet de quantifier la variabilité de l'optimisation, en évitant de se baser uniquement sur une valeur unique de  $h_d$ .

```
from scipy.stats import norm

n_bootstrap = 1000
h_d_bootstrap = []

for _ in range(n_bootstrap):
    sample_indices = np.random.choice(len(hauteurs_digue[:len(Cc_moyenne)]),
                                      size=len(Cc_moyenne), replace=True)
    Cc_sample = np.array(Cc_moyenne)[sample_indices]

    interp_func_sample = interp1d(hauteurs_digue[:len(Cc_sample)], Cc_sample, kind='cubic')
    res_sample = minimize_scalar(interp_func_sample, bounds=(min(hauteurs_digue),
                                                            max(hauteurs_digue)), method='bounded')

    h_d_bootstrap.append(res_sample.x)

# Calcul des IC à 95%
z_alpha = norm.ppf(0.975) # 1.96 pour IC à 95%
sigma_h_d = np.std(h_d_bootstrap)
IC_h_d_lower = h_d_optimal - z_alpha * sigma_h_d / np.sqrt(n_bootstrap)
IC_h_d_upper = h_d_optimal + z_alpha * sigma_h_d / np.sqrt(n_bootstrap)
```

```
# Résultats  
print(f"Hauteur optimale de la digue: {h_d_optimal:.2f} m")  
print(f"IC à 95% pour h_d optimal : [{IC_h_d_lower:.2f} m ; {IC_h_d_upper:.2f} m]")
```

Hauteur optimale de la digue: 7.34 m  
IC à 95% pour h\_d optimal : [7.22 m ; 7.46 m]

FIGURE 10 – Intervalle de confiance associé au quantile 95%

L'intervalle de confiance est une nouvelle fois étroit ( $\pm 0.12$  m autour de la valeur optimale), il réside une faible incertitude sur  $h_d$ .

La hauteur optimale de 7.34 m instaure un équilibre entre coût et sécurité. L'optimisation économique réduit les coûts inutiles en garantissant une protection efficace contre les surverses.

## 3 Conclusion

### 3.1 Résumé des différentes approches

n°	Méthode utilisée	Hauteur optimale ( $h_d$ )	Intervalle de Confiance	Avantages	Inconvénients
1	Quantile empirique avec régression linéaire	6.36 m (Q95) 7.37 m (Q99)	[5.95 m ; 6.77 m] (Q95) [6.66 m ; 7.92 m] (Q99)	Simple et rapide. Utilisation des données observées.	Ne prend pas en compte l'incertitude des valeurs manquantes.
2	Simulation Monte Carlo (hydraulique)	7.60 m	[7.52 m ; 7.68 m]	Prend en compte l'incertitude des variables hydrauliques	Ne considère pas les coûts économiques.
3	Simulation Monte Carlo (hydraulique + économique) + interpolation cubique	7.34 m	[7.22 m ; 7.46 m]	Optimise la sécurité et le coût. Prise en compte des incertitudes.	Calcul plus complexe, nécessitant une estimation fiable des coûts.

TABLE 3 – Comparaison des différentes approches

L'Approche 1 (quantile empirique avec régression linéaire) repose sur l'analyse directe des valeurs observées et la régression linéaire pour combler les données manquantes. Cette méthode est simple et rapide mais présente une limite importante : elle ne tient pas compte des incertitudes associées aux hauteurs d'eau futures. De plus, elle ne permet pas d'évaluer les risques économiques liés aux surverses.

L'Approche 2 (Monte-Carlo appliquée au modèle hydraulique) améliore la robustesse de l'estimation de  $h_d$  en prenant en compte l'incertitude des variables hydrauliques. Elle permet d'obtenir une hauteur plus sécurisée (7.60 m), tout en fournissant un intervalle de confiance fiable. Toutefois, cette approche ne prend pas en compte les aspects économiques.

L'Approche 3 (Monte-Carlo appliquée au modèle hydraulique et économique) permet un compromis optimal entre sécurité et coût. En intégrant les coûts de construction et d'entretien, elle permet d'optimiser la hauteur de la digue tout en évitant des dépenses excessives. La hauteur optimale obtenue est 7.34 m, avec un intervalle de confiance plus restreint que celui de l'Approche 2. Cette approche est plus réaliste car elle reflète les contraintes financières et hydrauliques.

### 3.2 Justification du choix final

Après comparaison, l'Approche 3 semble la plus pertinente, car elle permet :

- Une prise en compte complète des incertitudes hydrauliques et économiques, ce qui garantit une solution équilibrée.
- Une optimisation des coûts, évitant un surdimensionnement inutile de la digue tout en assurant une protection efficace.
- Une estimation plus fiable avec un intervalle de confiance plus étroit, réduisant l'incertitude sur  $h_d$ .