

Quantization applied to the visualization of low-probability flooding events

Charlie SIRE^{1,2,3}

Supervisors: R. LE RICHE³, D. RULLIERE³, J. ROHMER², L. PHEULPIN¹, Y. RICHEL¹

¹IRSN

²BRGM

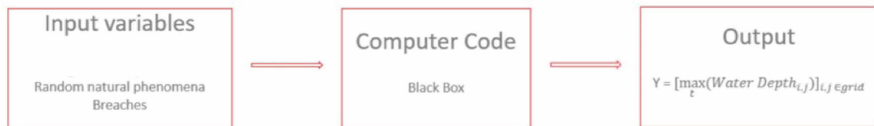
³CNRS,LIMOS

SIAM UQ, April 14

Content

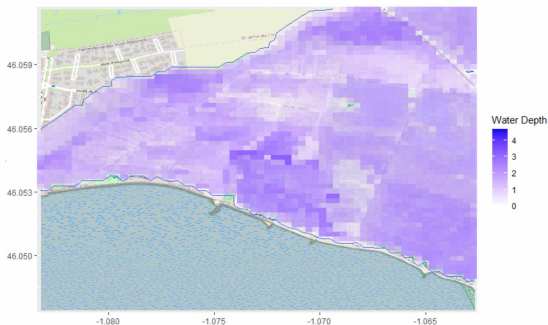
- 1 Introduction
- 2 Quantization
- 3 Importance Sampling
- 4 Metamodeling
- 5 Applications

Hydraulic simulators



One simulation \sim several hours

Output : Flooding map



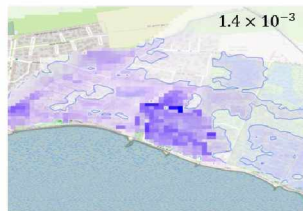
Visualization problem

Flooding map : easy to understand for everyone (Urban Planners, Decision makers)

But how to show a set of flooding maps that best represent the probability law associated to the flooding event ?

⇒ K-Means Clustering : Identify K Voronoi cells such that the mean distance between an observation and its nearest cell centroid is minimized

Objective



Problem formulation I

Input space = \mathcal{X} = Natural Phenoma parameters \times Breach parameters

Output space = \mathcal{Y} the space of pixelated maps, with inner product $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$

We introduce the random field Y :

$$Y: D \rightarrow \mathcal{Y}, \langle \cdot, \cdot \rangle_{\mathcal{Y}} \\ x \mapsto Y(x).$$

Problem formulation II

Quantization problem : Find for a given $\ell \in \mathbb{N}$, $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_\ell\} \in \mathcal{E}^\ell$ ℓ representatives of $Y(X)$

Closest representative map function:

$$q_\Gamma: \mathcal{Y} \rightarrow \Gamma$$
$$y \mapsto q_\Gamma(y) = \arg \min_{\gamma_i \in \Gamma} \|y - \gamma_i\|_{\mathcal{Y}}$$

Quantization error: $e(\Gamma) = [\mathbb{E} [\|Y(X) - q_\Gamma(Y(X))\|_{\mathcal{Y}}^2]]^{\frac{1}{2}}$

Objective: Find

$$\Gamma^* = \{\gamma_1^*, \dots, \gamma_\ell^*\} = \arg \min_{\Gamma \in \mathcal{Y}^\ell} (e(\Gamma))$$
$$= \arg \min_{\Gamma \in \mathcal{Y}^\ell} \left[\mathbb{E} \left[\min_{i \in \{1.. \ell\}} \|Y(X) - \gamma_i\|_{\mathcal{Y}}^2 \right] \right]^{\frac{1}{2}}$$

Key points

Quantization is performed in a specific context :

- 1 Expensive-to-evaluate simulators : metamodels adapted to spatial output
- 2 Low probability event : standard Monte Carlo sampling approach inefficient
- 3 Quantization in a space of maps (Adapted metamodel, storage)

Content

- 1 Introduction
- 2 Quantization
- 3 Importance Sampling
- 4 Metamodeling
- 5 Applications

Theorem Kieffer, Cuesta-Albertos

Theorem

If

- \mathcal{Y} is of finite dimension q
- $\forall z, y \in \mathcal{Y}, \langle z, y \rangle_{\mathcal{Y}} = \sum_{i=1}^n \lambda_i z_i y_i$ with $\forall i, \lambda_i > 0$
- $\mathbb{E} [\|Y(X)\|_{\mathcal{Y}}^2] < +\infty$

then $\forall i \in \{1 \dots \ell\}, \mathbb{E} [Y(X) \mid Y(X) \in C_i^{\Gamma^*}] = \gamma_i^*$

by defining C_j^{Γ} the Voronoi cells associated with a quantization Γ
: $C_j^{\Gamma} = \{y \in \mathcal{Y}, q_{\Gamma}(y) = \gamma_j\}$.

This means that the representatives of an optimal quantification coincide with the cells centroids.

Lloyd's algorithm

Algorithm Lloyd's algorithm

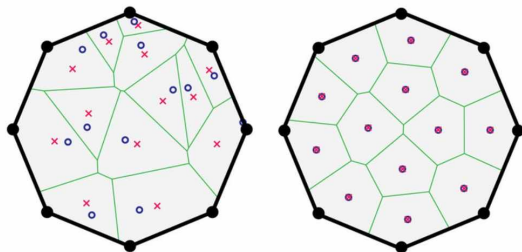
$$\Gamma^{[0]} \leftarrow \{\gamma_0^{[0]}, \dots, \gamma_\ell^{[0]}\}$$

1: **while** not stopping criterion **do**

$$\forall j \in \{1, \dots, \ell\}, \gamma_j^{[k+1]} \leftarrow \mathbb{E} \left[Y(X) \mid Y(X) \in C_j^{\Gamma^{[k]}} \right]$$

$$k \leftarrow k + 1$$

2: **end while**



Lloyd in our case

The main point is to compute at each iteration the conditional expectation $\mathbb{E} \left[Y(X) \mid Y(X) \in C_j^{\Gamma^{[k]}} \right]$

Problem here : In the flooding case, one prevailing Voronoi cell of empty maps (ie without water).

⇒ Monte Carlo sampling techniques not adapted for other clusters

Content

- 1 Introduction
- 2 Quantization
- 3 Importance Sampling**
- 4 Metamodeling
- 5 Applications

Principle of Importance Sampling

Objective : Estimate $\mathbb{E}[g(Y(X))]$ with $g: \mathcal{Y} \rightarrow \mathbb{R}^p$ such as $\mathbb{E}[g(Y(X))^2] < +\infty$

Main idea : The representation of $\mathbb{E}[g(Y(X))]$ as an expectation is not unique :

$$\mathbb{E}[g(Y(X))] = \mathbb{E}\left[g(Y(\tilde{X})) \frac{f_X(\tilde{X})}{\nu(\tilde{X})}\right]$$

with \tilde{X} a random variable with density function ν with $\text{supp}(f_X) \subset \text{supp}(\nu)$

Estimator with importance sampling

From this last representation :

$$\hat{E}_n^{IS} = \frac{1}{n} \sum_{k=1}^n g(Y(\tilde{X}^{(k)})) \frac{f_X(\tilde{X}^{(k)})}{\nu(\tilde{X}_k)}$$

with $(\tilde{X}^{(k)})_{k=1}^n$ be a n -sample of \tilde{X}

Its covariance matrix is : $\mathbb{V}(\hat{E}_n^{IS}) = \frac{1}{n} \mathbb{V}(g(Y(\tilde{X})) \frac{f_X(\tilde{X})}{\nu(\tilde{X})})$

In comparison to $\frac{1}{n} \mathbb{V}(g(Y(X)))$ in a classical MC

Idea : Choose ν that minimises variance

Importance sampling combined with quantization

$$\mathbb{E} [Y(X) \mid Y(X) \in C_j^\Gamma] = \frac{\mathbb{E} [Y(X) \mathbb{1}_{Y(X) \in C_j^\Gamma}]}{\mathbb{E} [\mathbb{1}_{Y(X) \in C_j^\Gamma}]}$$

And an estimator of $\mathbb{E} [Y(X) \mid Y(X) \in C_j^\Gamma]$:

$$\hat{E}_n^{IS}(\Gamma, j) = \frac{\frac{1}{n} \sum_{k=1}^n Y(\tilde{X}^{(k)}) \mathbb{1}_{Y(\tilde{X}^{(k)}) \in C_j^\Gamma} \frac{f_X(\tilde{X}^{(k)})}{\nu(\tilde{X}_k)}}{\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{Y(\tilde{X}^{(k)}) \in C_j^\Gamma} \frac{f_X(\tilde{X}^{(k)})}{\nu(\tilde{X}_k)}}$$

Density function ν : Uniform distribution

Content

- 1 Introduction
- 2 Quantization
- 3 Importance Sampling
- 4 Metamodeling**
- 5 Applications

Method [Perrin et al, 2021]

- 1 FPCA : Write every map $Y(x)$ as a linear combination of n_{pc} maps :

$$Y(x) = t_1(x)Y_1^{pca} + \dots + t_{n_{pc}}(x)Y^{pca}_{n_{pc}}$$

- 2 Gaussian process regression on every axis to predict $(\hat{t}_1(x^*), \dots, \hat{t}_{n_{pc}}(x^*))$ for a new x^*

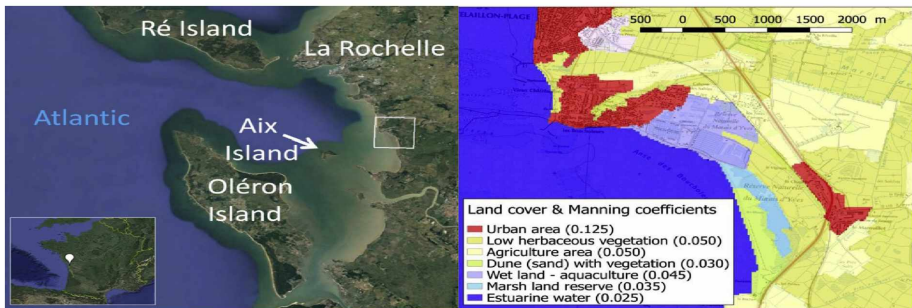
\implies Work with a large sample of predicted maps $\hat{Y}(\tilde{X}^{(k)})_{k=1}^n$

Content

- 1 Introduction
- 2 Quantization
- 3 Importance Sampling
- 4 Metamodeling
- 5 Applications**

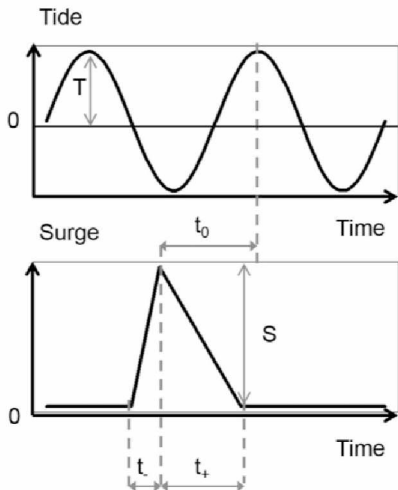
Les Bouchouleurs

- French Atlantic Coast near La Rochelle
- Hit by the storm Xynthia (27-28 February 2010)



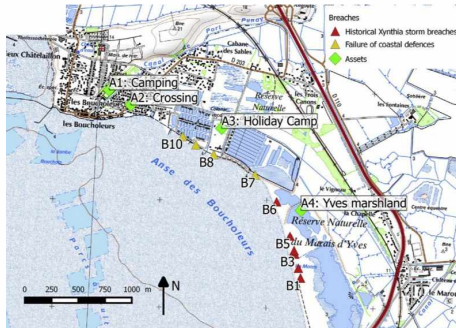
Natural phenomena variables

- High-tide level
- Surge peak amplitude
- Phase difference
- Time duration of the rising part
- Time duration of the falling part

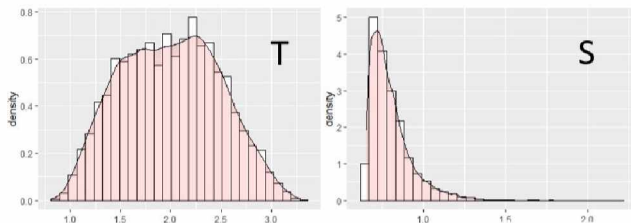


Breach variables

- 1 Location of the breaches : 10 different locations
 - 6 natural protections based on historical observations
 - 4 artificial dykes near vulnerable zones
- 2 Erosion rate : topographic level after failure as a fraction of initial crest level



Density function



Offshore conditions : Historical observations
Important offset added to the Surge that models the rise of the sea level

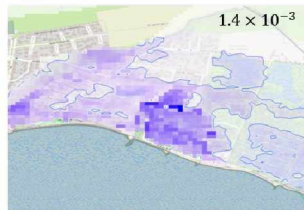
Breach variables :

- Probability of failure : p_{fail} function of T, S, t_0, t_-, t_+
- If failure : Uniform probability law

1300 flood maps simulated as follows :

- Natural phenomena variables sampled as the beginning of the Sobol Sequence
- 500 simulations without breach
- 800 breaches simulated uniformly ($\mathcal{U}_{\{1,\dots,10\}}$ for the breach location and $\mathcal{U}_{[0,1]}$ for the erosion rate)

Results



Future work

- Adapt the number of prototype maps
- Biased density ν
- Add a study of the input space