



**HAL**  
open science

## Metamodélisation d'un outil de dimensionnement de zones tampons enherbées

C Lauvernet, C Helbert, N Carluer, C Catalogne, E Leblois, R Muñoz-Carpena

► **To cite this version:**

C Lauvernet, C Helbert, N Carluer, C Catalogne, E Leblois, et al.. Metamodélisation d'un outil de dimensionnement de zones tampons enherbées. Séminaire HYPOPO, Riverly, Jun 2021, Villeurbanne, France. hal-03807261

**HAL Id: hal-03807261**

**<https://hal.inrae.fr/hal-03807261>**

Submitted on 9 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Metamodélisation d'un outil de dimensionnement de zones tampons enherbées

C. Lauvernet<sup>1</sup>, C. Helbert<sup>2</sup>  
N. Carluer<sup>1</sup>, C. Catalogne<sup>1</sup>, E. Leblois<sup>1</sup>, R. Muñoz-Carpena<sup>3</sup>

<sup>1</sup>RIVERLY, INRAE Lyon-Villeurbanne

<sup>2</sup>Univ. Lyon, UMR CNRS 5208, Ecole Centrale de Lyon

<sup>3</sup>Univ. Florida ; Agricultural and Biological Engineering unit - USA

# Problématiques en modélisation de l'environnement

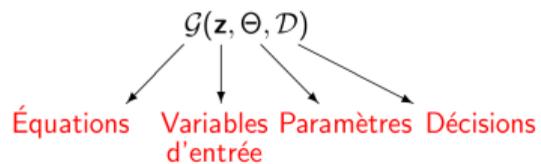
- un grand nombre de processus donc de paramètres, certains corrélés
- des processus dynamiques (équations différentielles ou aux différences) ou spatiaux
- des entrées temporelles (météo par exemple) ou spatiales (paysage)
- des fonctions à seuil (combinaison modèles physiques/empiriques)
- temps de calcul modèle-dépendant (< 1s à 1h voire plus)
- un grand nombre de variables de sortie

# Gestion des incertitudes dans un modèle

Phénomène physique  $\Rightarrow$  **Modèle physique**  $\Rightarrow$  Modèle numérique

simplification  $\Rightarrow$  **approximation numérique**  $\Rightarrow$  incertitudes liées entrées du modèle

Incertitude sur par. d'entrée  $\Rightarrow$  **Incertitude sur sorties**  $\Rightarrow$  Incertitude sur prise de décision finale



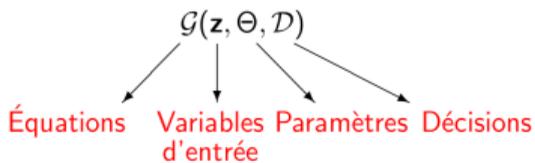
R. Faivre

## Gestion des incertitudes dans un modèle

Phénomène physique  $\Rightarrow$  **Modèle physique**  $\Rightarrow$  Modèle numérique

simplification  $\Rightarrow$  **approximation numérique**  $\Rightarrow$  incertitudes liées entrées du modèle

Incertitude sur par. d'entrée  $\Rightarrow$  **Incertitude sur sorties**  $\Rightarrow$  Incertitude sur prise de décision finale

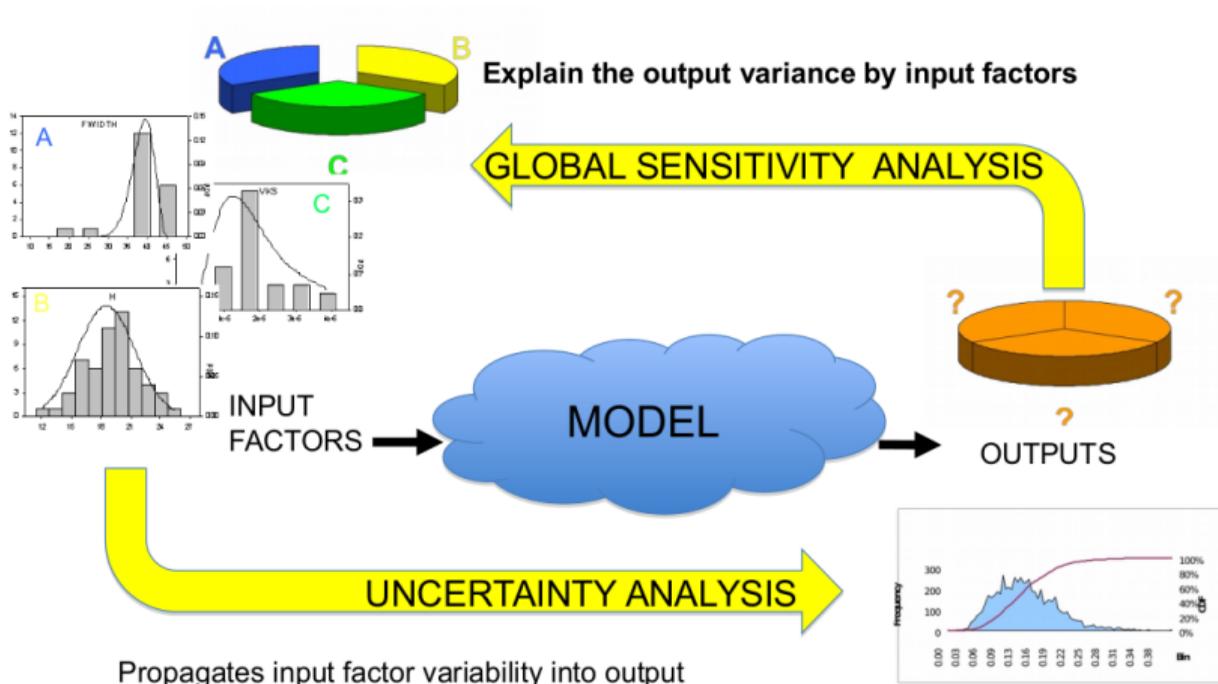


R. Faivre

- Comment propager les incertitudes sur la sortie du code ?
- Quelles sont les incertitudes les plus préjudiciables ?
- Quelles sont les variables les plus influentes ?
- Incertitude au final sur la sortie ?
- Estimation de la marge de confiance sur la prise de décision ?

$\Rightarrow$  **propagation d'incertitudes, analyse de sensibilité, problème inverses, ...**

# Sources d'incertitude dans un modèle



# Quantification de l'incertitude dans un modèle

Problème :

- Code souvent complexe et coûteux
- Grand nombre de variables d'entrée
- Grand nombre de simulations nécessaires pour réaliser les études de sensibilité, de propagation d'incertitude ou de calibration
- Etude souvent « multi-objectif »

# Quantification de l'incertitude dans un modèle

Problème :

- Code souvent complexe et coûteux
- Grand nombre de variables d'entrée
- Grand nombre de simulations nécessaires pour réaliser les études de sensibilité, de propagation d'incertitude ou de calibration
- Etude souvent « multi-objectif »

⇒ Exploitation directe du code difficile  
⇒ Utilisation d'un métamodèle

# Qu'est-ce qu'un métamodèle/surrogate model/émulateur/surface de réponse ?

- C'est un modèle statistique construit entre les entrées et les sorties du modèle
- un "modèle de modèle" (2ème niveau d'abstraction)
- permettant de prédire avec une « bonne précision » de nouvelles réponses dans le domaine de variation des paramètres incertains

# Qu'est-ce qu'un métamodèle/surrogate model/émulateur/surface de réponse ?

- C'est un modèle statistique construit entre les entrées et les sorties du modèle
- un "modèle de modèle" (2ème niveau d'abstraction)
- permettant de prédire avec une « bonne précision » de nouvelles réponses dans le domaine de variation des paramètres incertains

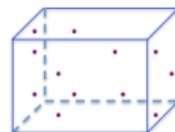
- ↪ dépend de beaucoup moins de paramètres d'entrée
- ↪ temps d'exécution négligeable
- ↪ analyses de sensibilité/propagation d'incertitude/optimisation réalisables à moindre coût
- ↪ transfert d'outils de recherche vers les utilisateurs de terrain

## Construction d'un métamodèle

$\mathcal{G}$  la fonction code,  $\mathcal{M}$  le métamodèle,  $K \ll N$

$$Y = \mathcal{G}(X_1, \dots, X_N) \approx \mathcal{M}(X_i, \dots, X_K)$$

1. Plan d'expérience = échantillonnage dans l'espace des input

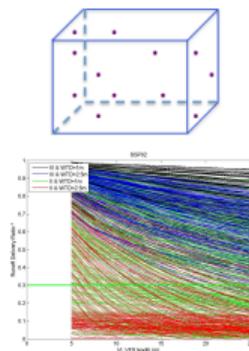


## Construction d'un métamodèle

$\mathcal{G}$  la fonction code,  $\mathcal{M}$  le métamodèle,  $K \ll N$

$$Y = \mathcal{G}(X_1, \dots, X_N) \approx \mathcal{M}(X_i, \dots, X_K)$$

1. Plan d'expérience = échantillonnage dans l'espace des input
2. Evaluation du modèle sur l'échantillon

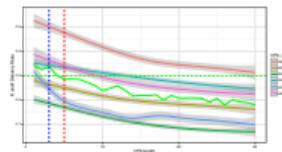
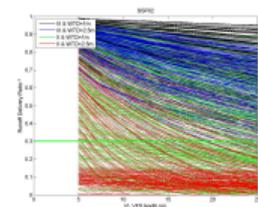
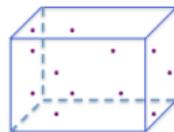


## Construction d'un métamodèle

$\mathcal{G}$  la fonction code,  $\mathcal{M}$  le métamodèle,  $K \ll N$

$$Y = \mathcal{G}(X_1, \dots, X_N) \approx \mathcal{M}(X_i, \dots, X_K)$$

1. Plan d'expérience = échantillonnage dans l'espace des input
2. Evaluation du modèle sur l'échantillon
3. Approximation du modèle par  $\mathcal{M}$  sur ce PLEX (surface de réponse) :  $Y = \mathcal{M}(X_{i1}, X_{i2}, \dots, X_{iJ}) + \eta$

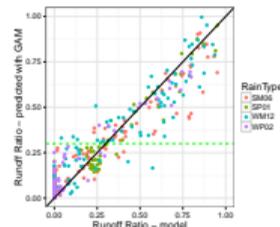
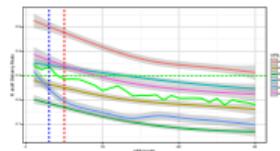
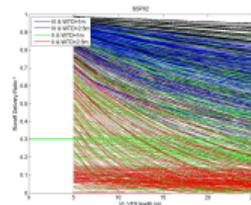
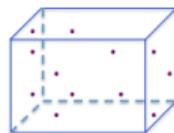


# Construction d'un métamodèle

$\mathcal{G}$  la fonction code,  $\mathcal{M}$  le métamodèle,  $K \ll N$

$$Y = \mathcal{G}(X_1, \dots, X_N) \approx \mathcal{M}(X_i, \dots, X_K)$$

1. Plan d'expérience = échantillonnage dans l'espace des input
2. Evaluation du modèle sur l'échantillon
3. Approximation du modèle par  $\mathcal{M}$  sur ce PLEX (surface de réponse) :  $Y = \mathcal{M}(X_{i1}, X_{i2}, \dots, X_{ij}) + \eta$
4. Validation de  $\mathcal{M}$  sur un PLEX indépendant



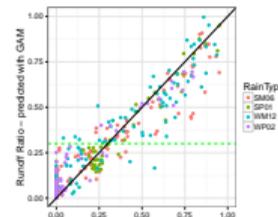
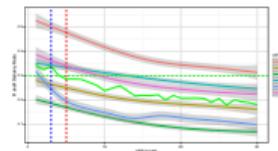
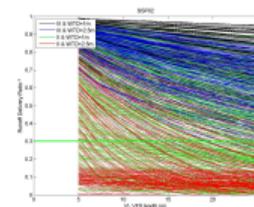
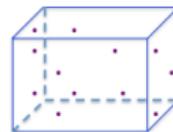
# Construction d'un métamodèle

$\mathcal{G}$  la fonction code,  $\mathcal{M}$  le métamodèle,  $K \ll N$

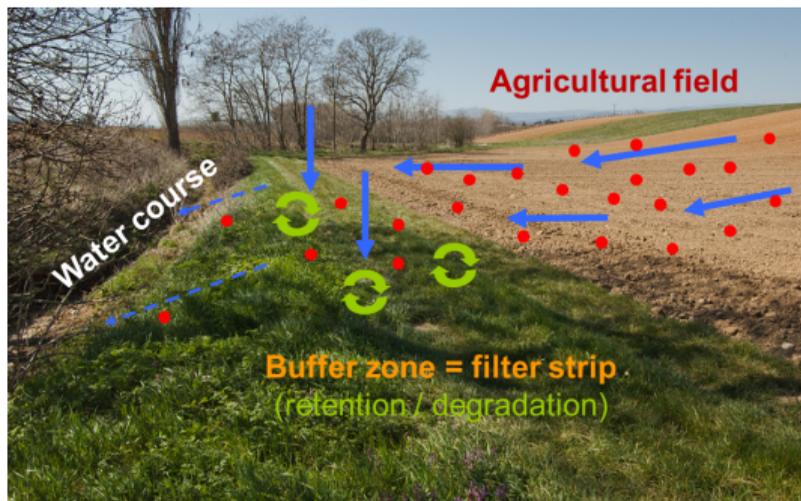
$$Y = \mathcal{G}(X_1, \dots, X_N) \approx \mathcal{M}(X_i, \dots, X_K)$$

## Différentes méthodes:

- Parametric regression (linear, polynomial)
- Nonparametric regression : **GAM**, Spline,...
- **Krigeage** ou processus gaussiens conditionnels (PG)
- Learning methods : NN, SVM, Random Forests,...



## Zones tampons enherbées pour atténuer les transferts



[ C.Catalogne-Irstea ]

Forte perméabilité / forte rugosité /  
teneur élevée en matière organique



Augmentation de la résistance  
hydraulique du flux

Augmentation de l'**infiltration** du sol



Réduction de la capacité de transport  
des sédiments/particules + polluants  
dissous

+ décalage (eau + sédiments)



**Adsorption** des pesticides solubles



**Dégradation** des pesticides par  
micro-organismes

## Efficacité des BE pour réduire le transfert de pesticides : 0-99%

- Efficacité sur des événements particuliers = 0-99%.
- L'efficacité dépend de la dimension de la BE, du placement dans le bassin versant et des conditions locales (climat, sol, etc.)
- De nombreux processus entrent en jeu : hydrologie, transport de sédiments, flux de polluants
- Les dimensions actuelles préconisées/obligatoires ne sont pas du tout adaptées

## Efficacité des BE pour réduire le transfert de pesticides : 0-99%

- Efficacité sur des événements particuliers = 0-99%.
- L'efficacité dépend de la dimension de la BE, du placement dans le bassin versant et des conditions locales (climat, sol, etc.)
- De nombreux processus entrent en jeu : hydrologie, transport de sédiments, flux de polluants
- Les dimensions actuelles préconisées/obligatoires ne sont pas du tout adaptées

Besoin d'un outil opérationnel prenant en compte la complexité des processus en jeu

# Une chaîne de modèles pour dimensionner les BE

## BUVARD (Carlier et al. 2017)

### Buffer strip for runoff Attenuation and pesticides Retention Design tool

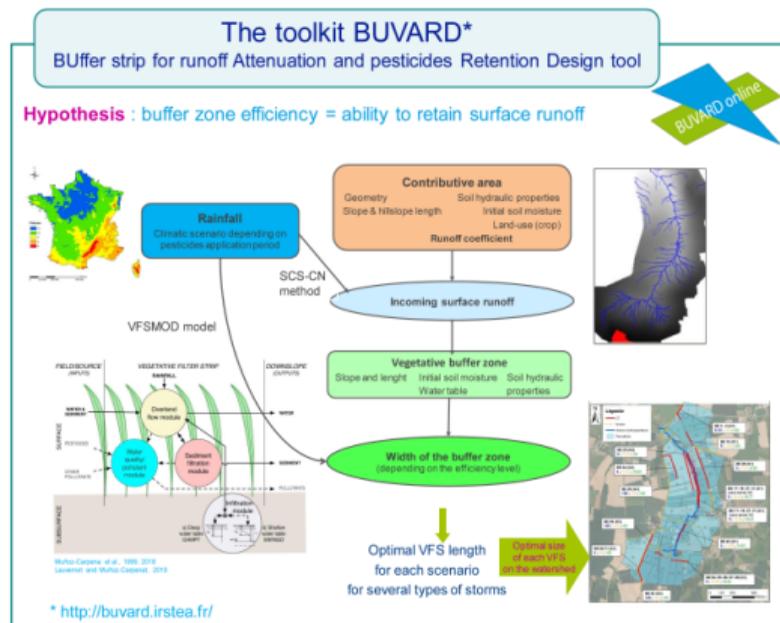
Definition d'un scénario (agronomie, sol, climat) :

- 1 Generation de plusieurs événements-types de pluie de la zone et de ruissellement entrant sur la BE <sup>a</sup>
- 2 Processus au sein de la BE avec le modèle mécaniste VFSMOD <sup>b</sup>

⇒ Efficacité de la BE pour chaque scénario

<sup>a</sup>Catalogne, C. Lauvernet, and Leblois 2016

<sup>b</sup>Rafael Muñoz-Carpena, Parsons, and Gilliam 1999; C. Lauvernet and R. Muñoz-Carpena 2018; R. Muñoz-Carpena, C. Lauvernet, and Carlier 2018



# Une chaîne de modèles pour dimensionner les BE

## BUVARD (Carluer et al. 2017)

BUffer strip for runoff Attenuation and pesticides Retention Design tool

⇒ Un outil complet, mécaniste, mais peu de succès en pratique...7 finishers depuis 2011 ?

- 1 nécessite connaissances détaillées du terrain
- 2 lecture du guide (120 p.) pour renseigner les bons paramètres
- 3 prise en main de plusieurs interfaces/langages
- 4 ne permet pas de faire de déduire l'incertitude associée

⇒ il faut simplifier BUVARD !  
Base de scénarios simulés Métamodèle



# Metamodèle de BUVARD: en pratique

VFS efficiency

$$RDR = \frac{\text{Runoff}[out]}{\text{Runoff}[in]}$$


=

Modeling toolkit  $\mathcal{G}$

- Rainfall hyetograph
- Runoff dynamic hydrograph
- Season
- Curve Number
- Slope, Area
- VFS Water table depth
- Soil type ( $K_{sat}$ ,  $\theta_s$ , VG par., ...)
- Sediments characteristics
- Roughness, grass height
- ...

≈

Metamodel  $\mathcal{M}$

- Curve Number
- Slope
- Contributive area
- VFS width
- VFS Water table depth
- VFS Soil type
- Rainfall typical event

7 parameters

>70 parameters

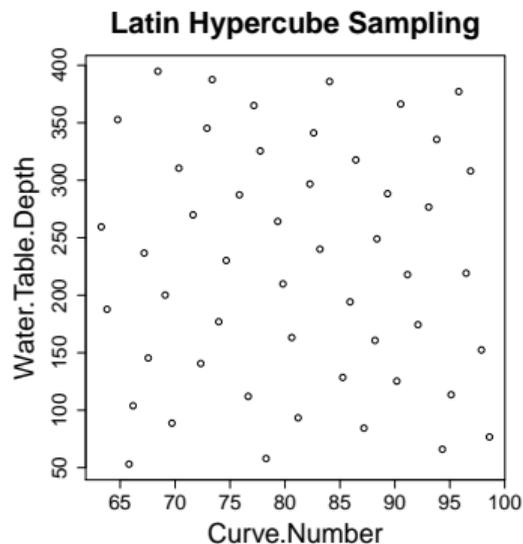
# Stratégie d'échantillonnage

en fonction des possibilités des utilisateurs de l'outil:

Key variable	Sampling range
Curve Number	[ 63 , 99 ]
Slope	[ 2% , 20% ]
Contributive area length(m)	[ 25 , 300 ]
VFS Water table depth (m)	[ 0.5 , 4 ]
Rainfall typical event	Summer/Winter short/long
VFS Soil type	clo, scl, SIL, CLO , SCL, SAL

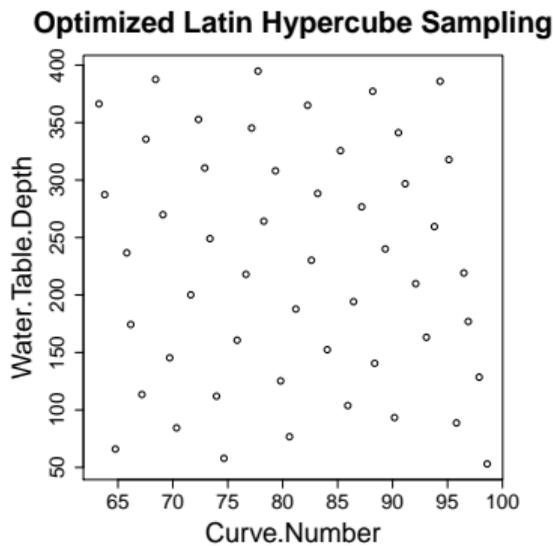
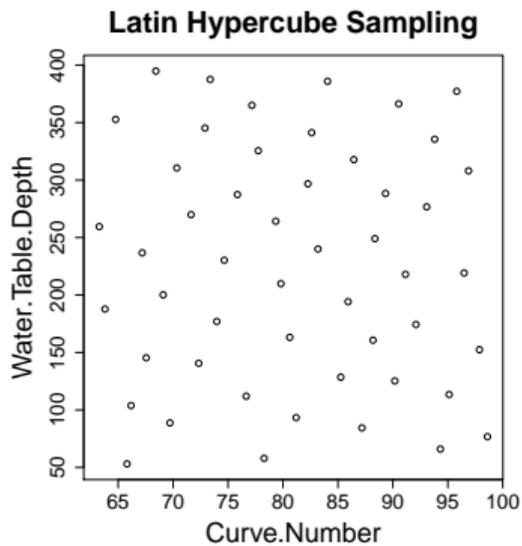
# Stratégie d'échantillonnage : Latin Hypercube Sampling en 7D

Projections uniformes dans chaque direction (un unique point par colonne et par ligne).

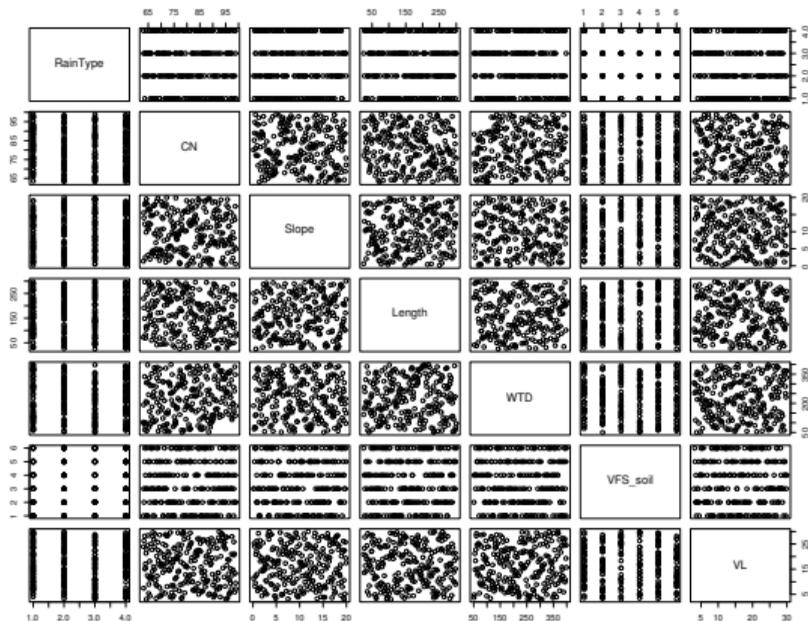


# Stratégie d'échantillonnage : Latin Hypercube Sampling en 7D

Optimisation du LHS via un critère de distance (maximin): on maximise la plus petite distance entre les points.



# Stratégie d'échantillonnage : Latin Hypercube Sampling en 7D



Approche d'échantillonnage pas trop coûteuse, et adaptée à des modèles irréguliers.

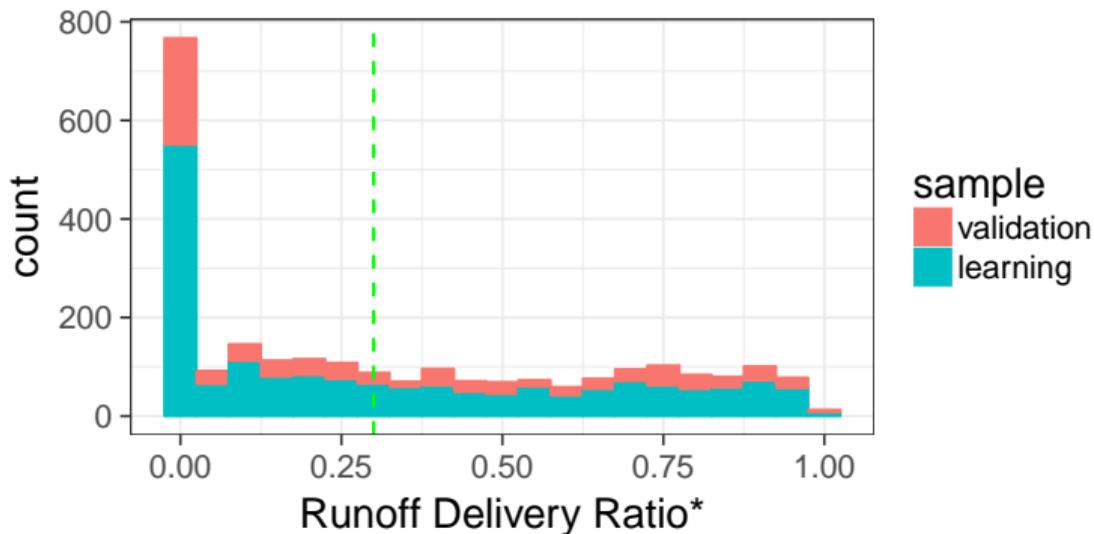
Objectif = assurer de bonnes propriétés de projection sur chaque axe : chaque projection 1D est

Maximin-optimale

# Analyse des simulations sur l'échantillon d'apprentissage

Output variable = Runoff Delivery Ratio

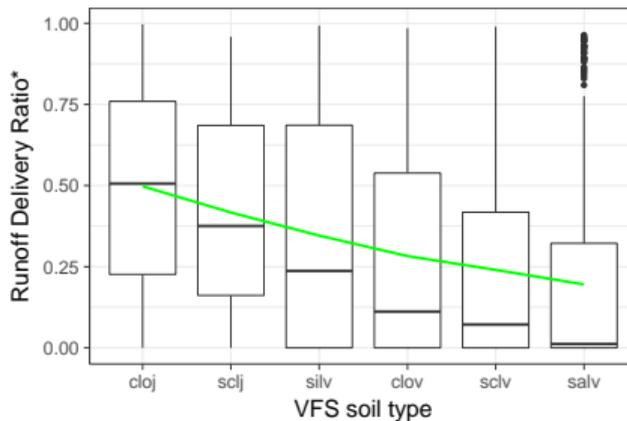
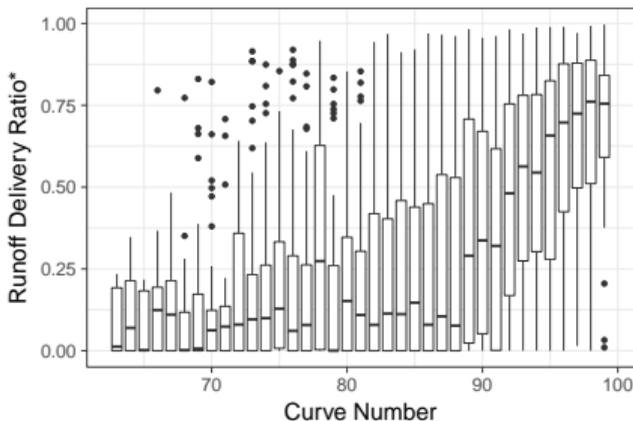
$$\frac{\text{Runoff}[out]}{\text{Runoff}[in]}$$



# Analyse des simulations sur l'échantillon d'apprentissage

Output variable = Runoff Delivery Ratio

$$\frac{\text{Runoff}[out]}{\text{Runoff}[in]}$$



# Métamodèle par krigeage ou Processus gaussiens conditionnels

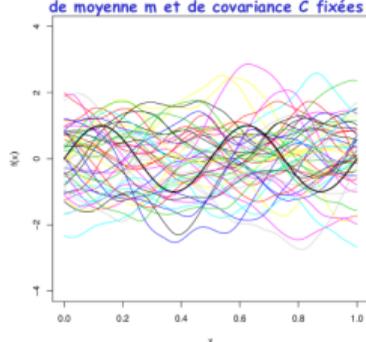
- Hypothèse = la sortie déterministe du modèle est la réalisation d'un processus gaussien :

$$Z(x) = m(x) + W(x)$$

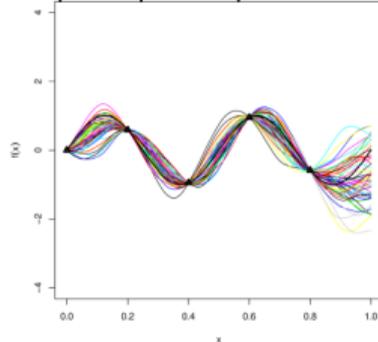
$m$  fonction moyenne,  $W(x)$  PG centré stationnaire ( $\sigma^2, R$ )

- On conditionne le PG à passer par les  $n$  points observés (points de simulations du code)
- ⇒ Le PG  $Z$  conditionné reste un PG (formules analytiques pour moyenne et covariance)
- Prédicteur = moyenne des réalisations du PG conditionnel (Espérance)
  - Le krigeage a été adapté aux var. mixtes quanti/quali (type de sol, typ de pluie) <sup>1</sup>

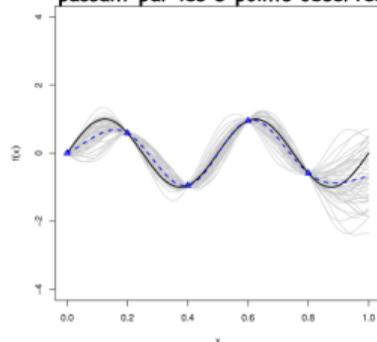
Exemples de réalisations du PG (trajectoires)  
de moyenne  $m$  et de covariance  $C$  fixées



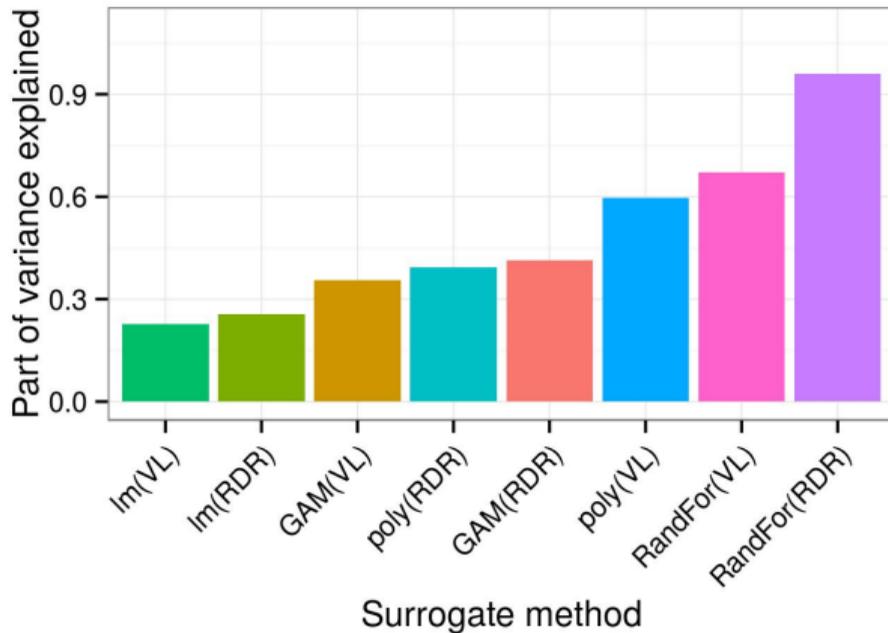
Trajectoires du PG (de moyenne  $m$  et de covariance  $C$  fixées)  
passant par les 5 points observés



Trajectoires du PG (de moyenne  $m$  et de covariance  $C$  fixées)  
passant par les 5 points observés



## Comparison de différentes méthodes classiques



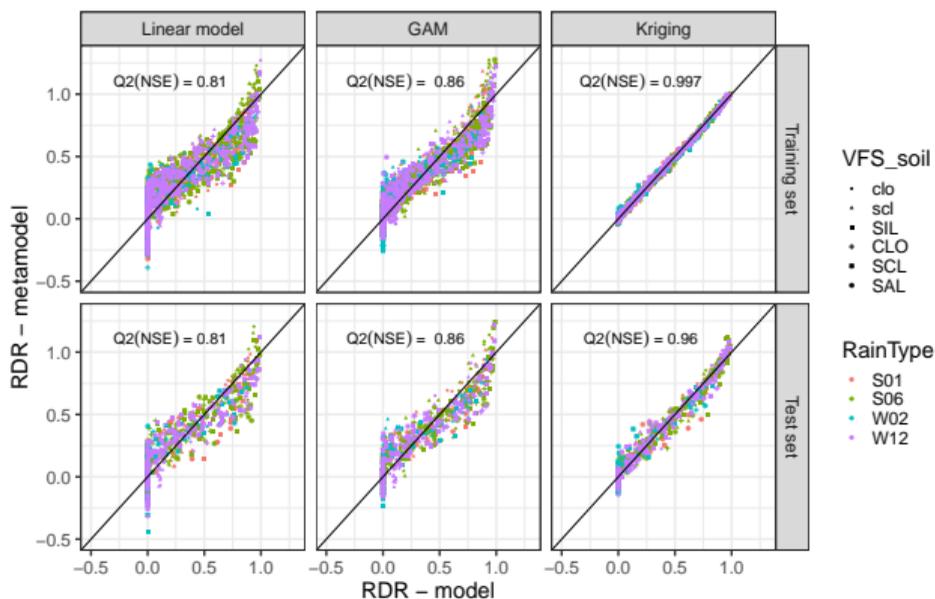
# Metamodel prediction vs model simulation

Output variable = Runoff Delivery Ratio = Runout/Runin

$RDR = 0 \Leftrightarrow \text{efficiency} = 100\%$ ;  $RDR = 1 \Leftrightarrow \text{efficiency} = 0\%$

GAM  $R^2 = 0.84$   
Kriging  $R^2 = 0.96$

- PG est très efficace avec variables quali
- GAM est très stable entre apprentissage et validation
- erreurs de prédiction fortement liées au plateau du RDR

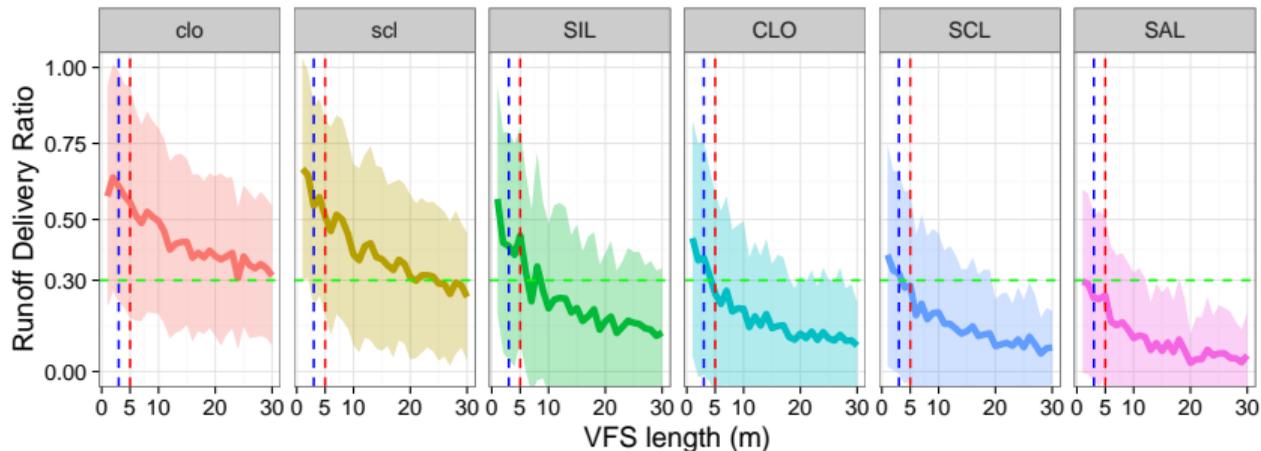


# Application pratique du métamodèle

Quelle longueur pour une efficacité voulue ?

$$RDR = 30\% \Leftrightarrow \text{efficiency} = 70\%$$

Ratio moyen et incertitude sur un large échantillon de simulations

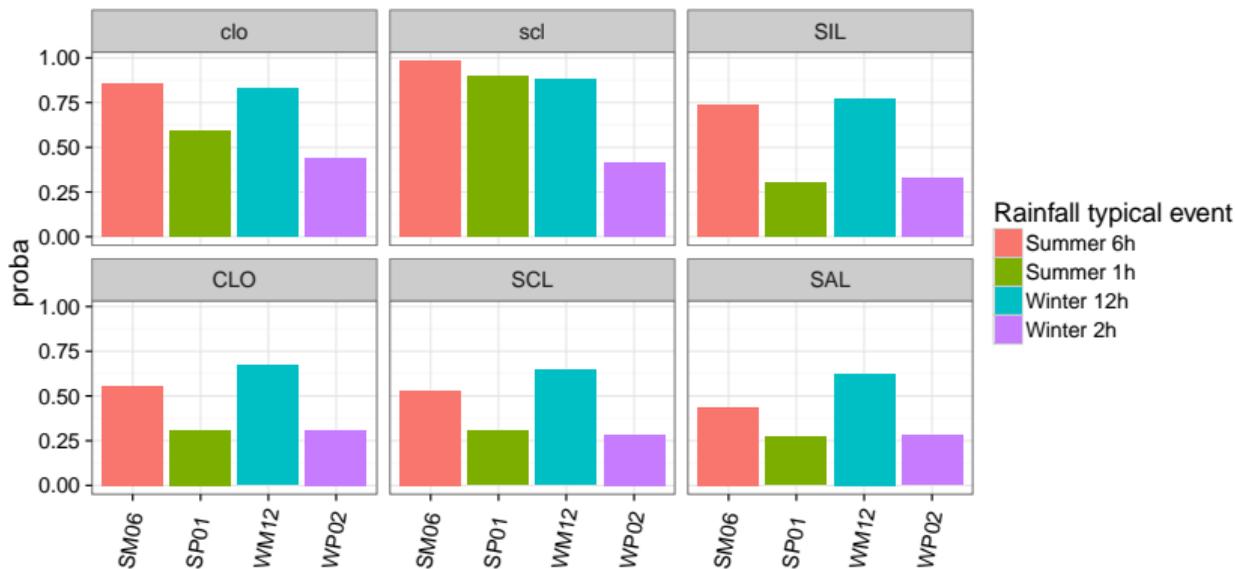


- variabilité très large de la longueur optimale / type de sol
- possibilité de quantifier l'incertitude associée aux prédictions

# Application pratique du métamodèle

Quelle est le risque de prendre la mauvaise décision avec la règle du “one-fits-all”?

$\text{proba}(\text{efficiency} < 70\% \mid \text{filter length} = 3\text{m})$



## Conclusion

- un outil très utilisé en opérationnel pour tous ses avantages de cout de calcul tout en restant très performant : réacteurs nucléaires à EDF, éoliennes à ETZ
- également intéressant pour l'aspect écologique
- pour BUVARD, le MM est bien plus simple à utiliser que la chain d'outils
- mais un problème principal : un plateau important de valeurs très proches de 0
- démarche : toujours discuter en parallèle avec thématiciens pour bien définir son échantillon et l'incertitude des entrées

