

# Improving Surrogate Model-based Uncertainty Quantification in a Costly Numerical Environment

OpenTURNS Users Day #10 | 6th June 2017

**Pamphile ROY**

PhD student supervised by  
B. Cuenot and J-C. Jouhaud

in collaboration with:

GLOBE: S. Ricci, R. Mélanie, N. El Moçayd, M. De Lozzo, V. Laborie (CEREMA)  
EDF LNHE: Équipe MASCARET, N. Goutal, F. Zaoui  
EDF MRI



# Un contexte HPC avec diverses problématiques

## Plusieurs axes de travail

- ◆ Étude de sensibilité
- ◆ Assimilation de données
- ◆ Optimisation robuste



## Quelques applications d'intérêt pour le CERFACS

- ◆ UQ sur des configurations LES industrielles -Thèse de Pamphile Roy (CERFACS, débutée en 2016)
- ◆ Optimisation pour le craquage du fioul - Thèse de Robin Campet (CERFACS, débutée en 2015)
- ◆ Système d'alerte sur la Gironde via la modélisation 2-D avec TELEMAC - Thèse de Vanessya Laborie (CEREMA-SCHAPI, débutée en 2016), Post-doctorat de Matthias De Lozzo (CERFACS, 2016-2018)
- ◆ Prévision des crues via la modélisation 1-D avec MASCARET - Thèse de Nabil El Moçayd (EDF-CNES, 2017)

# BATMAN, du Python pour HPC



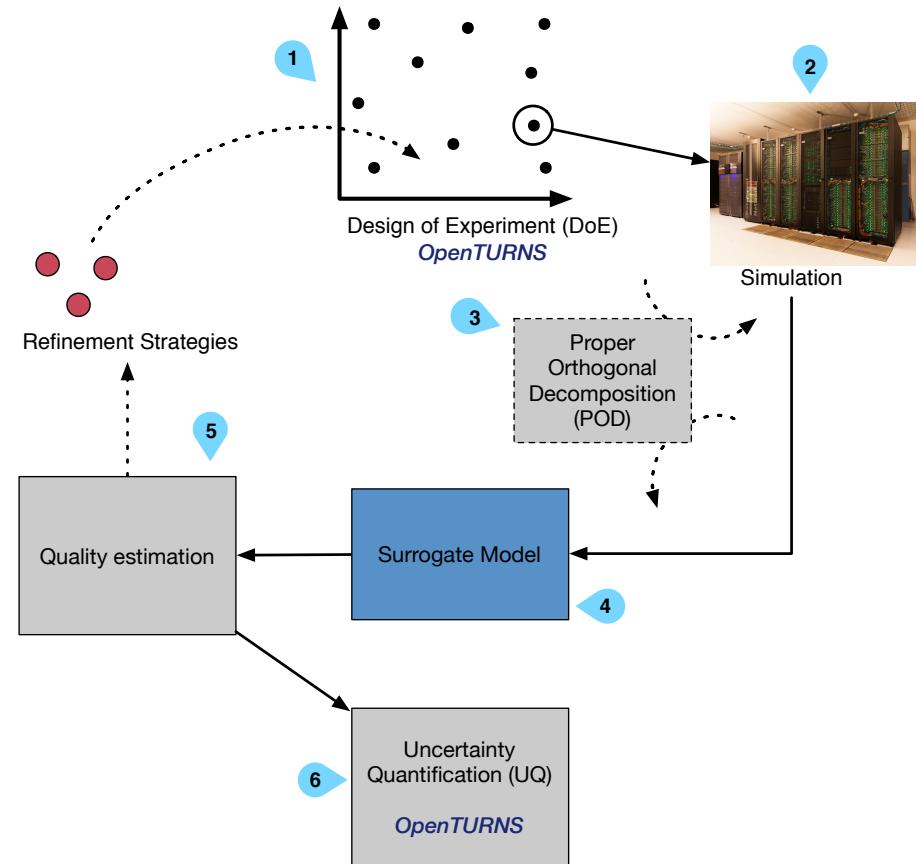
Plateforme UQ en Python qui repose sur OpenTURNS

## Objectifs

- ◆ Mettre en œuvre OT sur nos cas
- ◆ Agnostique des codes de calculs
- ◆ Implémentation et test de nouvelles méthodes

## Un package

- ◆ Commun aux équipes
- ◆ Parallèle
- ◆ Documenté et testé



# BATMAN est utilisé dans plusieurs études

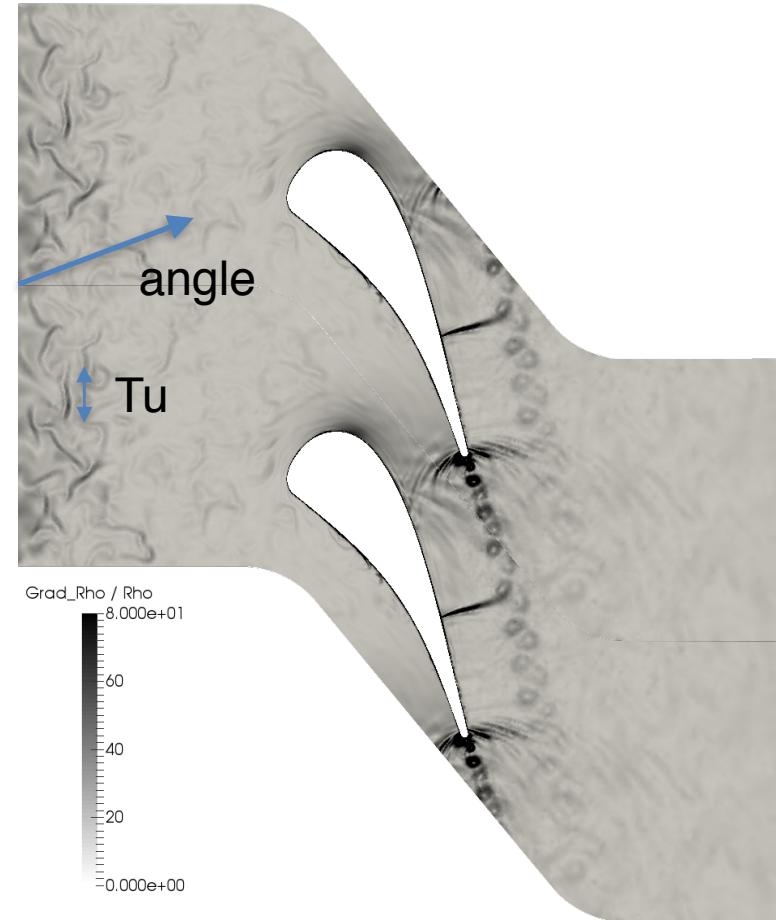
1. Quantification d'incertitudes sur un cas LES
2. Comparaison de modèles réduits
3. Optimisation géométrique pour le craquage du fioul
4. Quantification d'incertitudes géométriques sur un tourbillonneur

# Résultats sur le cas LS89



Évaluer l'influence des conditions amonts sur le flux de chaleur

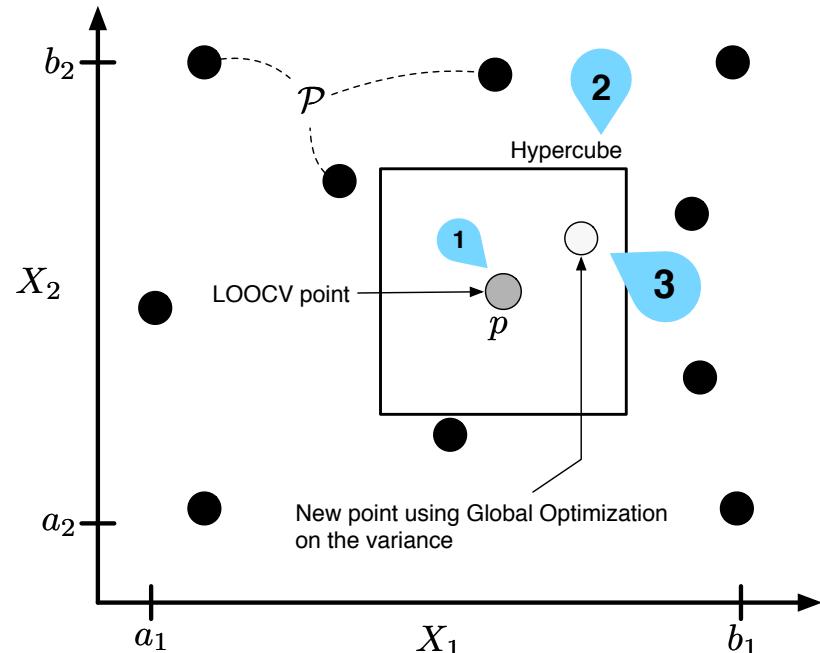
- ◆ Première étude UQ en LES de la configuration LS89
- ◆ 2 paramètres incertains : angle incident du flux et intensité de la turbulence
- ◆ 3-D instationnaire :  $T_{CPU} \sim 7000$  h
- ◆ Code utilisé : AVBP (CERFACS)



1. LOOCV pour obtenir le point le plus sensible
2. Optimisation discrète en utilisant les points du DoE, puis optimisation continue pour trouver le plus grand hypercube qui l'inclut
3. Ajout d'un point dans l'hypercube

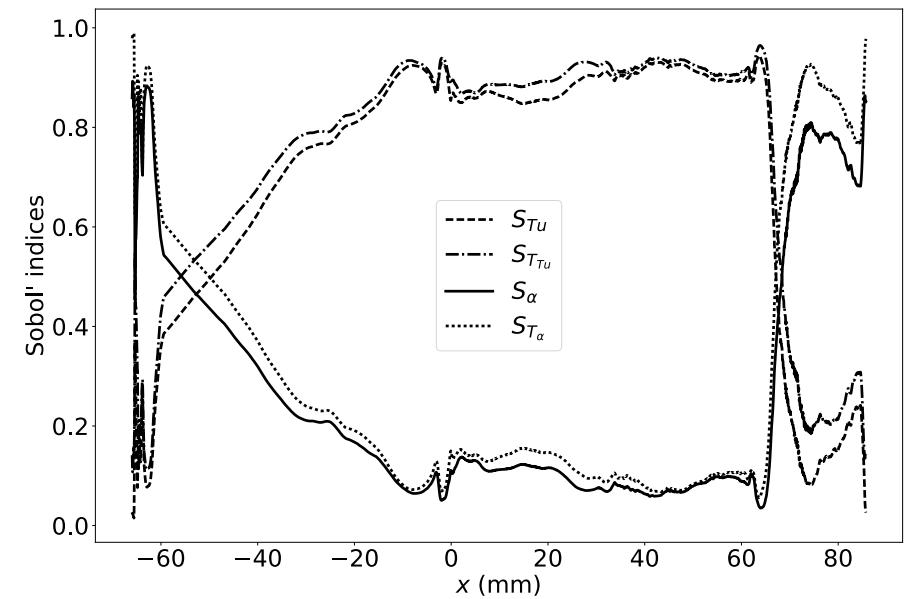
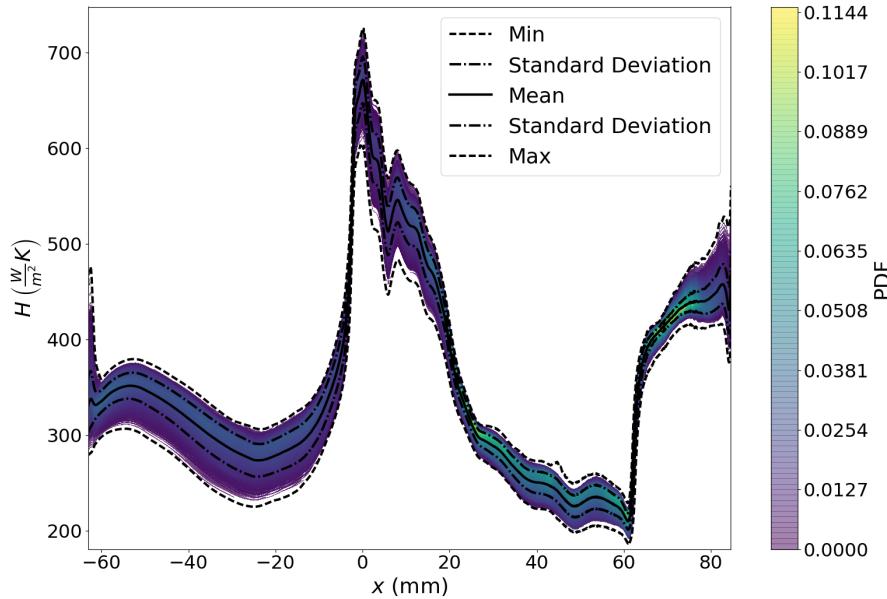


*LowDiscrepancySequence*



P.T. Roy, L. Segui *et al.* Resampling Strategies to Improve Surrogate Model-based Uncertainty Quantification - Application to LES of LS89. Submitted to Computers and Fluids, 2017.

# Résultats sur le cas LS89



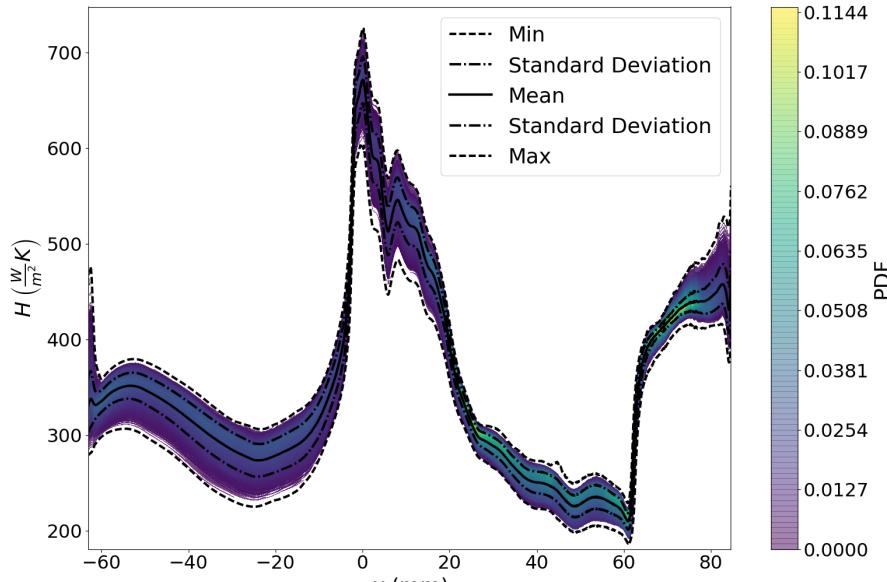
- Influence spatiale de la contribution des paramètres sur le flux de chaleur



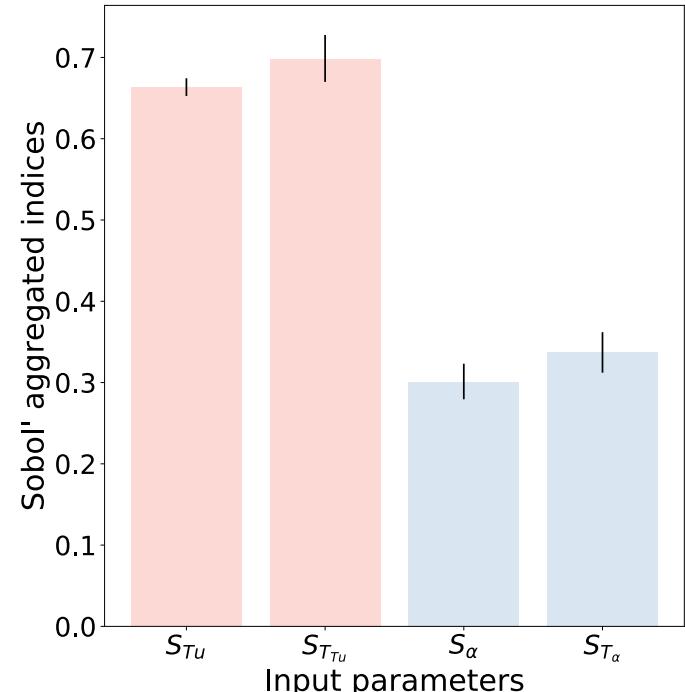
*ComposedDistribution ;  
SobolIndicesAlgorithm; otwraphy  
KernelSmoothing*

P.T. Roy, L. Segui *et al.* Resampling Strategies to Improve Surrogate Model-based Uncertainty Quantification - Application to LES of LS89. Submitted to Computers and Fluids, 2017.

# Résultats sur le cas LS89



*Propagation d'incertitudes*



*Analyse de sensibilité*

- Influence spatiale de la contribution des paramètres sur le flux de chaleur



*ComposedDistribution ;  
SobolIndicesAlgorithm; otwraphy  
KernelSmoothing*

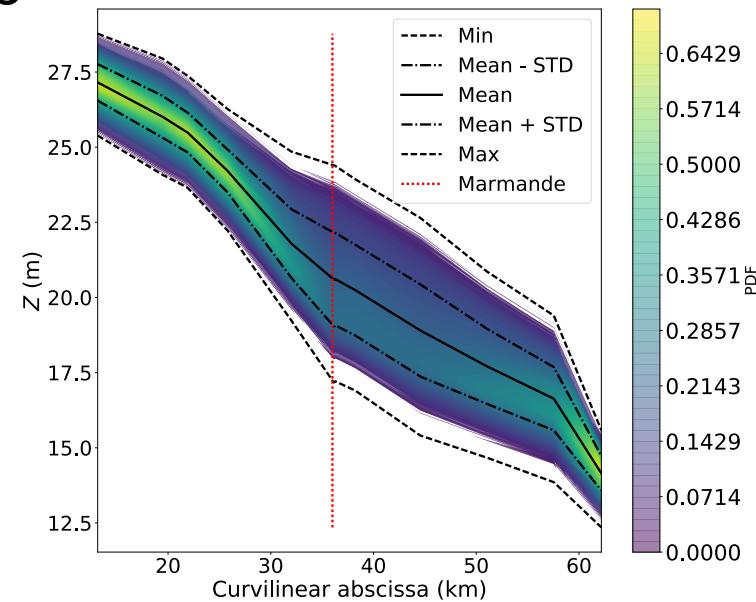
P.T. Roy, L. Segui *et al.* Resampling Strategies to Improve Surrogate Model-based Uncertainty Quantification - Application to LES of LS89. Submitted to Computers and Fluids, 2017.

# Comparaison entre PC et GP



Comparer la performance des métamodèles PC vs GP dans un cadre hydraulique

- ◆ Étude stationnaire 1-D sur la Garonne
- ◆ 2 paramètres incertains : coefficient de frottement et débit amont
- ◆ Code : MASCARET



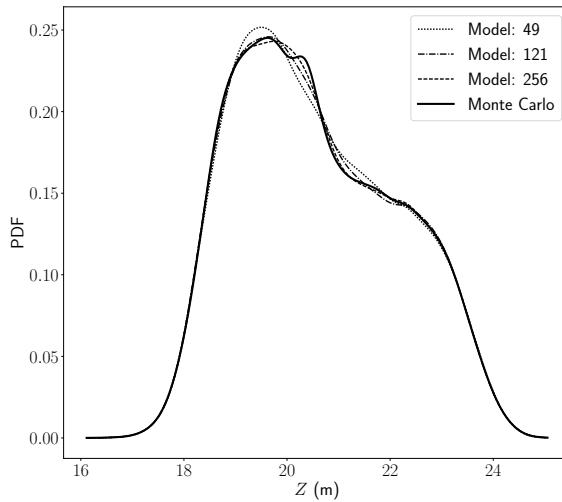
N. El Moçayd *et al.* Polynomial surrogates for open-channel flows in steady state. Under review in ENMO, 2017

P.T. Roy, N. El Moçayd *et al.* Comparison of Polynomial Chaos and Gaussian Process surrogate for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows. Submitted to SERRA, 2017

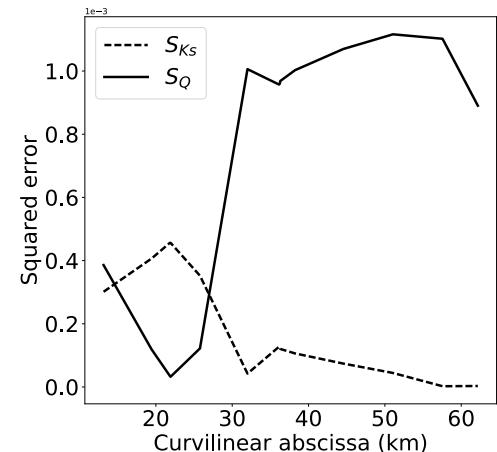
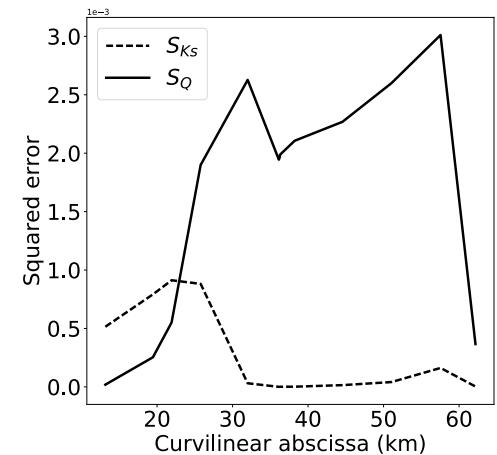
# Comparaison entre PC et GP

- Les deux métamodèles sont performants
- À partir de 121 simulations, on retrouve le caractère bi-modal pour GP et il faut >256 simulations pour PC
- Le PC localise mieux le mode principal de la PDF

*Propagation  
des incertitudes*



*LowDiscrepancySequence  
FunctionalChaosAlgorithm*



*Analyse de sensibilité : erreur*

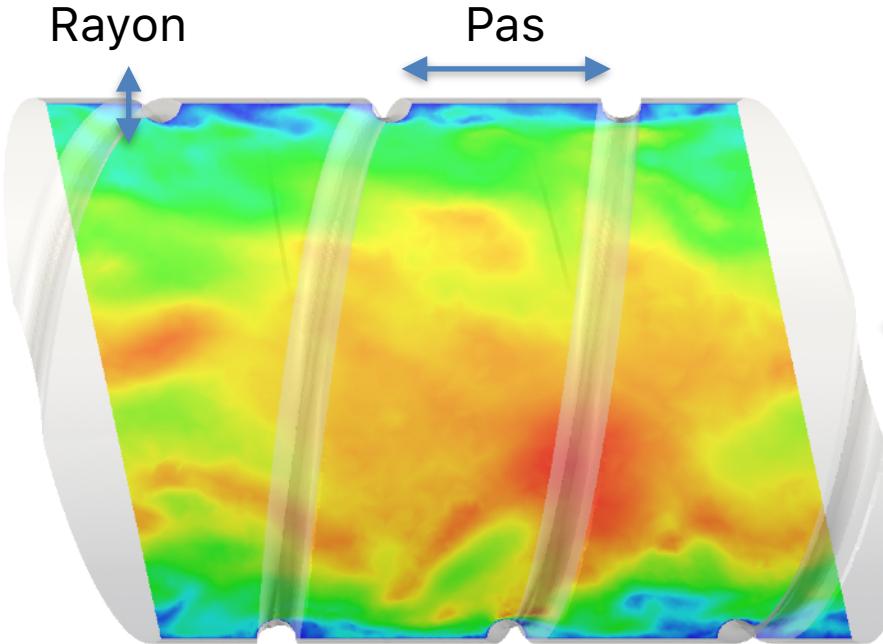
# Optimisation robuste



Augmenter la sélectivité chimique en augmentant les échanges thermiques (thèse de Robin Campet)

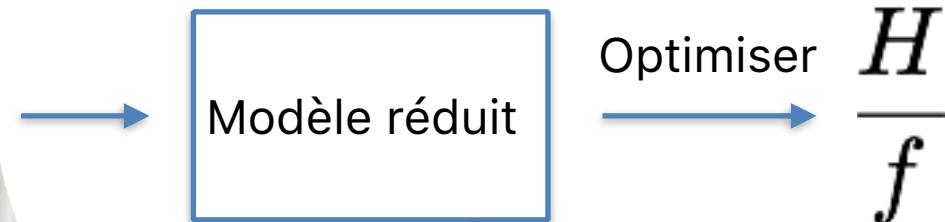


otmorris



Tube nervuré : vitesse instantanée, AVBP

- Les simulations sont coûteuses, 3D instationnaire et réactives : TCPU ~ 20 000 h
- 4 paramètres dont 1 discret
- Code utilisé : AVBP (CERFACS)



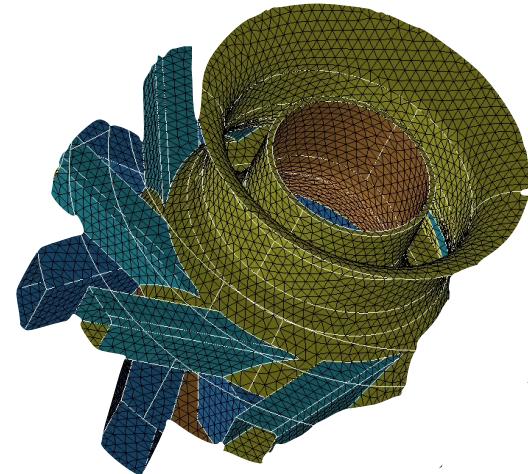
Raffinement :  
EGO, etc.

# Propagation d'incertitudes dans un tourbillonneur



Quantifier l'influence de la géométrie sur l'écoulement

- ◆ Les simulations sont coûteuses, 3D instationnaire : T<sub>CPU</sub> ~ 25 000 h
- ◆ La géométrie est complexe
- ◆ La précision de fabrication est élevée
- ◆ Les paramètres sont nombreux (> 30)
- ◆ Code utilisé : AVBP (CERFACS)



*KarhunenLoeveP1Factory  
otmorris*

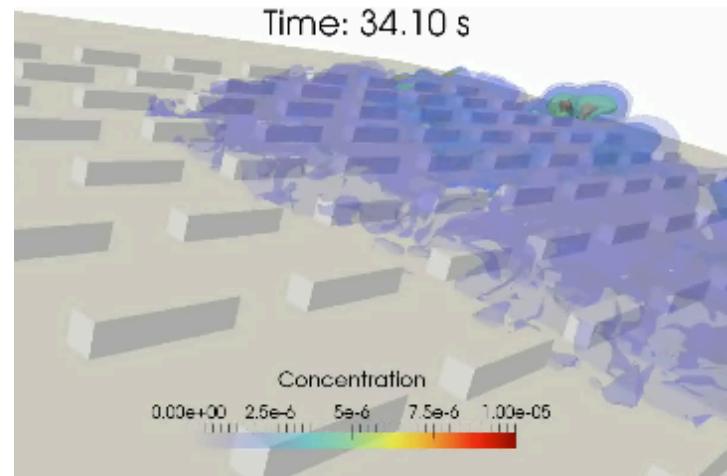
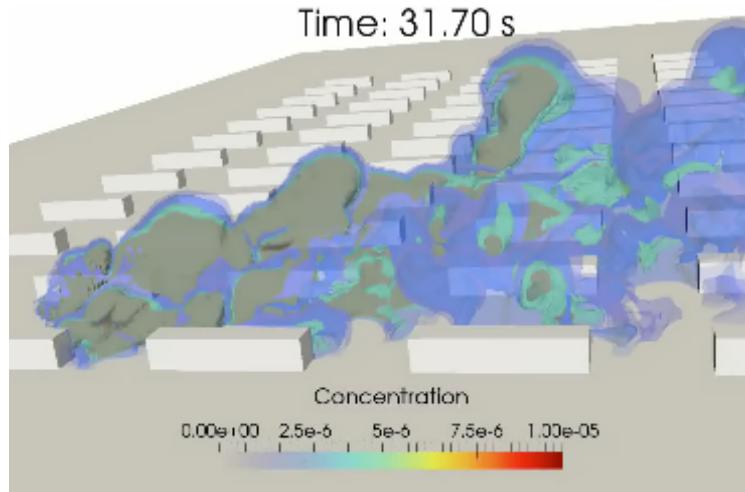
# Perspectives d'utilisations d'OpenTURNS

- ◆ Extension des métamodèles pour des variables incertaines non scalaires
  - Application hydraulique : série temporelle du débit amont, profil spatial de la bathymétrie
  - Application CFD : champ d'incertitudes géométriques
- ◆ Assimilation de données
- ◆ Multi-Level Monte-Carlo

# Dispersion de polluants

Caractériser l'impact des aléas météorologiques sur la concentration instationnaire du traceur (postdoctorat de Géraldine Rea)

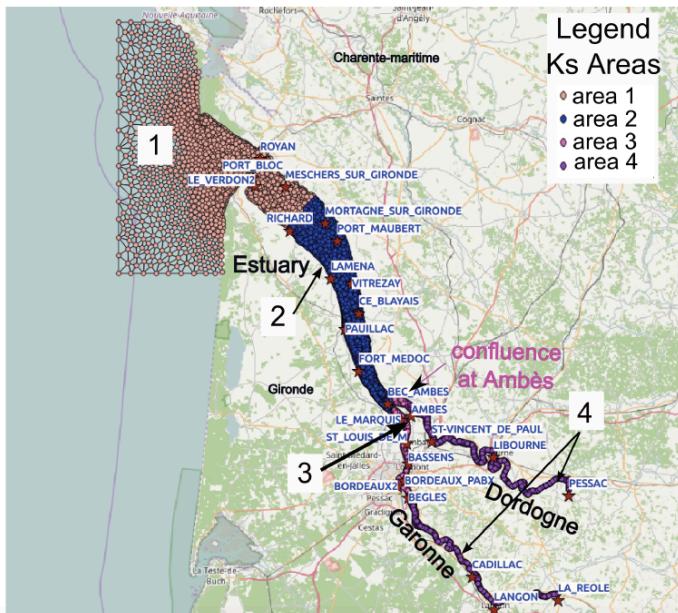
- Les simulations sont coûteuses, 3-D instationnaire : TCPU ~ 30 000 h
- 2 paramètres incertains : intensité et direction du vent à la condition aux limites



# Assimilation de données sur l'estuaire de la Gironde

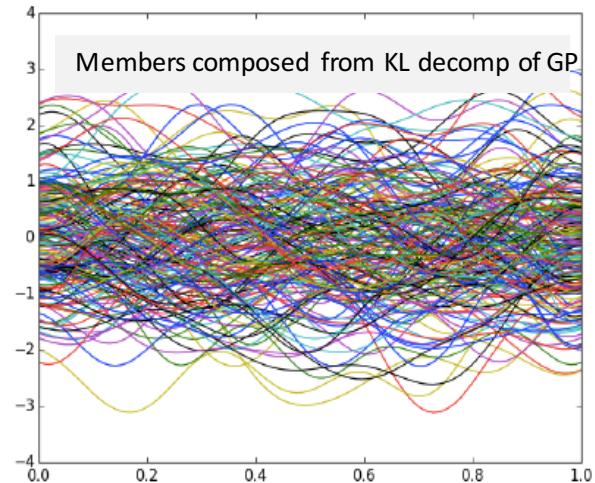


Système de prédition temp réel avec Assimilation de données sur mesures *in-situ* and radar (thèse de Vanessya Laborie)



Estuaire de la Gironde

- Plus de 20 sources d'incertitudes : forçage (hydrologique, pression, vent), conditions maritimes
- UQ en temps et en espace
- Entrées fonctionnelles modélisées par processus gaussien si 1 observation ou par SVD + coefficients aléatoires.



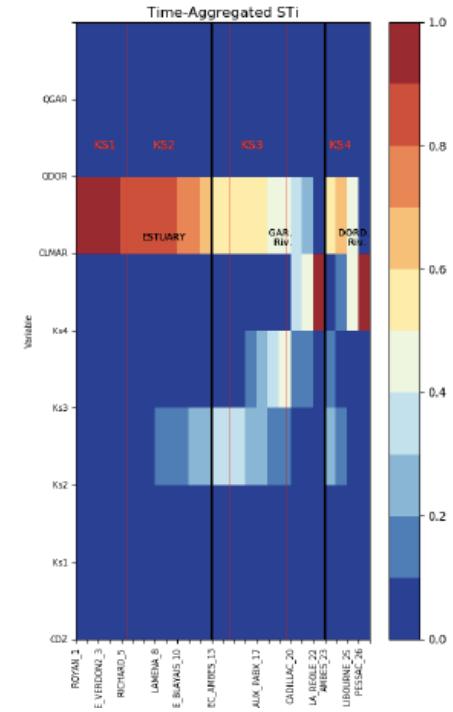
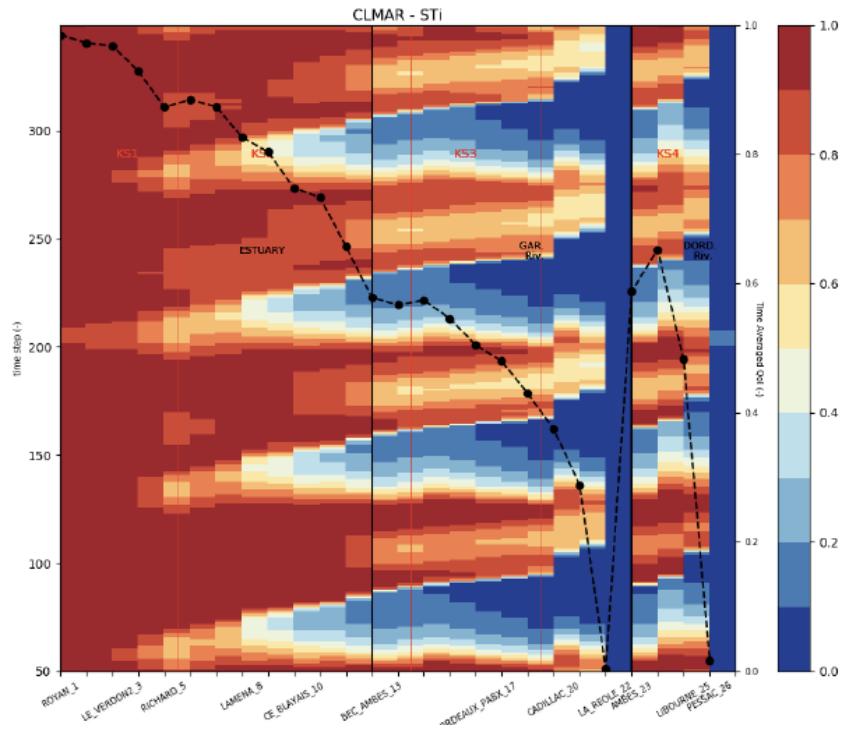
Forçage hydrologique



KarhunenLoeveP1Factory

# Assimilation de données sur l'estuaire de la Gironde

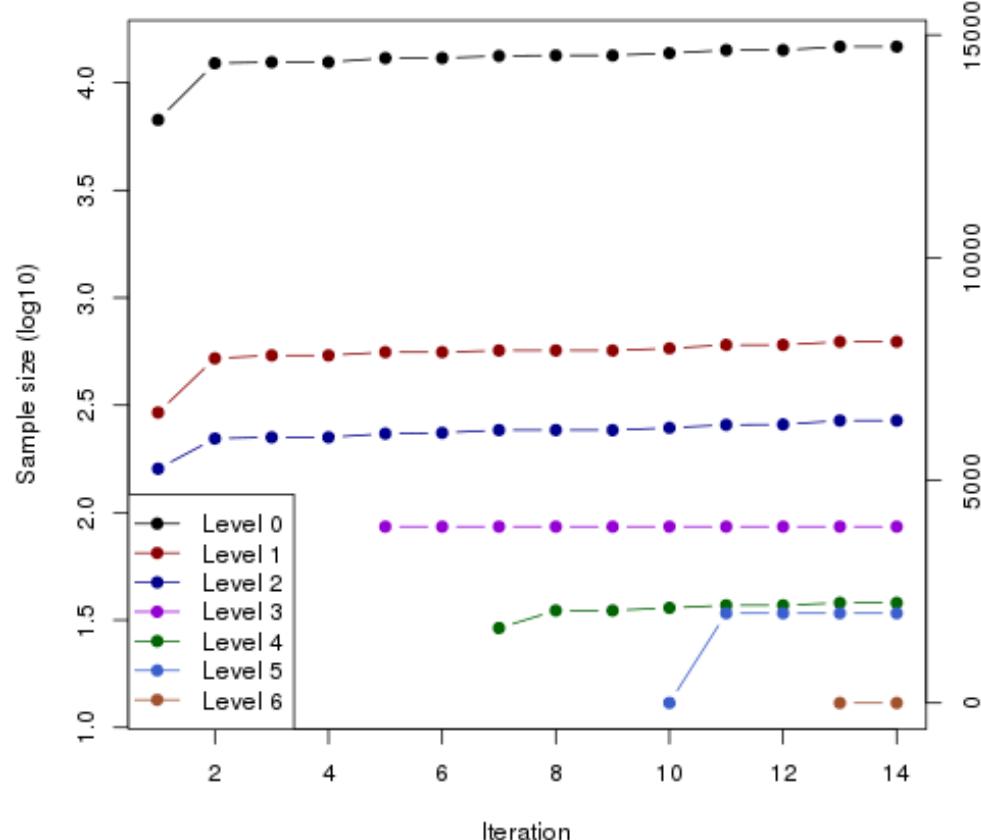
- Cartes des indices de Sobol' en temps et en espace (*data visualization*)
- On voit l'impact des marées sur les liens entrées/sortie



V. Laborie *et al.* Uncertainty Quantification for the Gironde Estuary Hydrodynamics with TELEMAC 2D. SimHydro 2017

# Approches Monte Carlo multi-niveaux

- ◆ Séquence de  $L$  simulateurs de :
  - précision croissante
  - coût croissant
- ◆ Estimateur MLMC : somme télescopique d'estimateurs MC sur niveaux successifs
- ◆ Choix optimal de  $L$  et des tailles des échantillons MC
- ◆ Testé sur l'espérance, la variance et la covariance
- ◆ Appliqué à l'API MASCARET
- ◆ Perspectives : indices de Sobol' et filtre de Kalman d'ensemble



Coût de calcul 36 fois inférieur à MC

# Notre utilisation d'OpenTURNS



*LowDiscrepancySequence  
ComposedDistribution  
SobolIndicesAlgorithm  
PythonFunction  
otwrapy  
KernelSmoothing  
FunctionalChaosAlgorithm  
KarhunenLoeveP1Factory  
otmorris*

## Quelques applications d'intérêt pour le CERFACS

- ◆ UQ sur des configurations LES industrielles -Thèse de Pamphile Roy (CERFACS, débutée en 2016)
- ◆ Optimisation pour le craquage du fioul - Thèse de Robin Campet (CERFACS, débutée en 2015)
- ◆ Système d'alerte sur la Gironde via la modélisation 2-D avec TELEMAC - Thèse de Vanessya Laborie (CEREMA-SCHAPI, débutée en 2016), Post-doctorat de Matthias Delozzo (CERFACS, 2016-2018)
- ◆ Prévision des crues via la modélisation 1-D avec MASCARET - Thèse de Nabil El Moçayd (EDF-CNES, 2017)