



Réduction de dimension avec des modèles auto-associatifs

Rencontres INRIA LJLL en calcul scientifique

2 mars 2020

Sylvain Girard

girard@phimeca.com

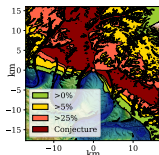
Décider avec un modèle

- ▶ Simulation coûteuse
- ▶ Dimension d'entrée ou de sortie élevée

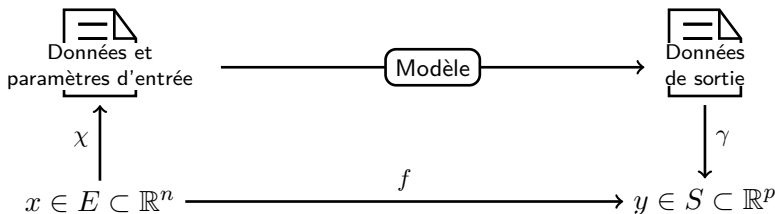
1. Réduction de dimension **linéaire supervisée** pour l'inversion



2. Réduction de dimension **non-linéaire** pour l'émulation

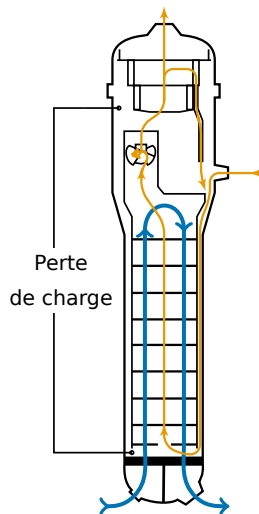


Expériences numériques

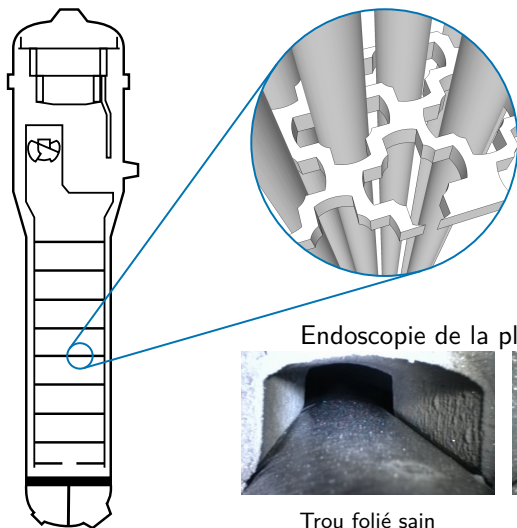


- ▶ Le **choix** de E et S définit f .
- ▶ Propagation d'**incertitude** : simuler Y aléatoire, à partir de X
- ▶ **Inversion** : inférer x à partir de y
- ▶ **Émulation** : approximation de f
- ★ Trouver E et S intéressants, avec f régulière, n et p petits.

1. Colmatage des générateurs de vapeur



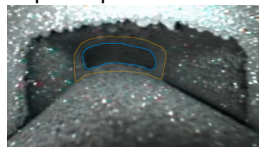
Le phénomène de colmatage



Endoscopie de la plaque supérieure :

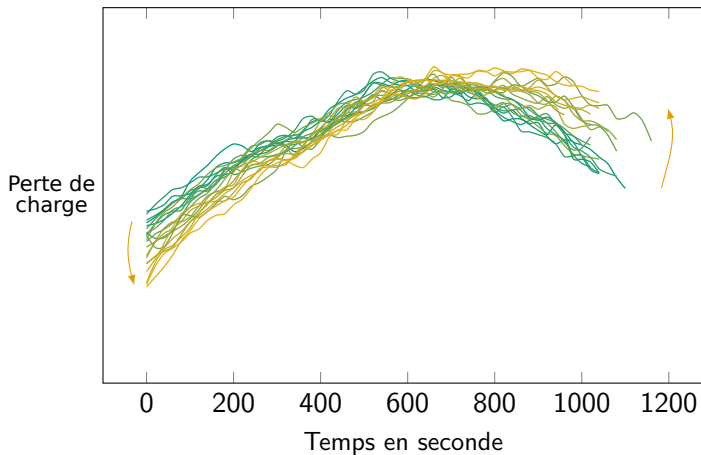


Trou folié sain



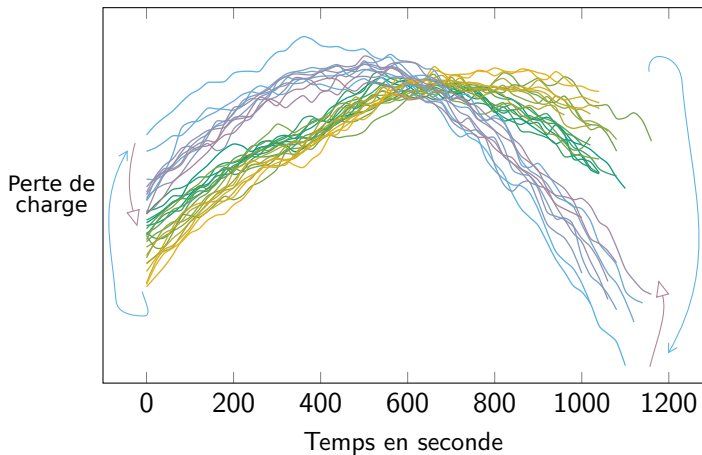
Trou folié colmaté

Effet du colmatage



↗ : évolution avant nettoyage chimique

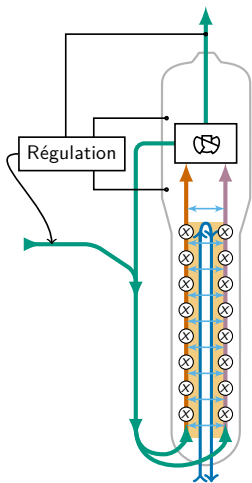
Effet du colmatage



↗ : nettoyage chimique

↘ : évolution après nettoyage chimique

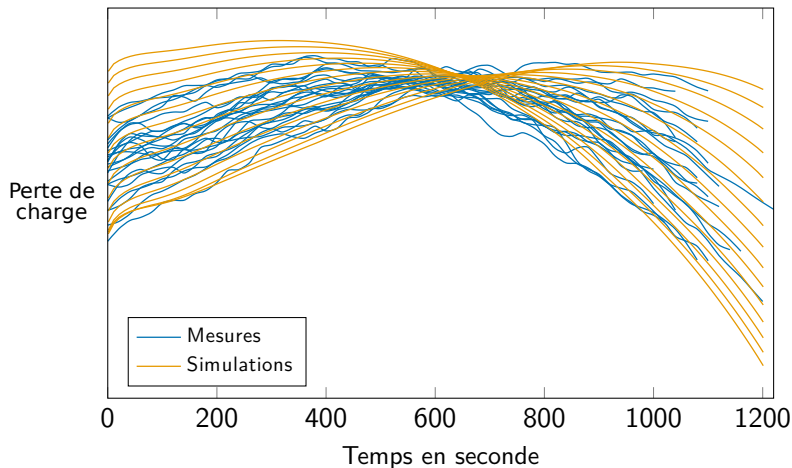
Modèle thermo-hydraulique



- ▶ Système algébro-différentiel programmé en Modelica
- ▶ Écoulement diphasique : modèle de mélange avec glissement de phase
- ▶ Nombreuse relations de fermeture empiriques : expériences et comparaison à des modèles 3D

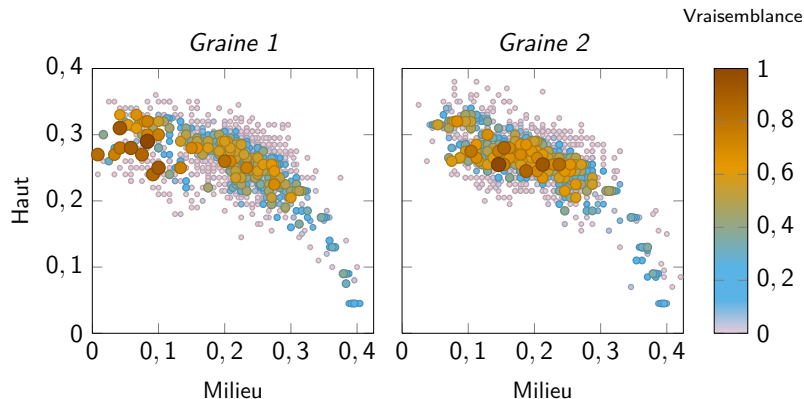
- ▶ $E \subset \mathbb{R}^{16}$: 16 taux de colmatage
- ▶ $S \subset$: courbe de réponse de perte de charge

Simulation de l'effet du colmatage



Objectif : prédire les taux colmatage en observant la réponse dynamique (problème inverse mal posé).

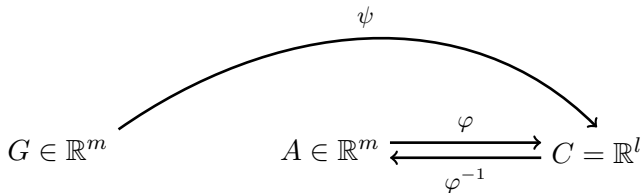
Tentative d'inversion par filtrage particulaire



Problème : trouver

- ▶ une paramétrisation de l'état de colmatage,
- ▶ et une métrique pour comparer les courbes de réponse adaptés au diagnostic.

Réduction de dimension



- ▶ « Réduire la dimension » de l'ensemble G , c'est construire un **ensemble approché** A doté d'un **système de coordonnées** φ de basse dimension ($l \ll m$).
- ▶ En pratique, on construit ψ à partir d'un échantillon z_1, \dots, z_N de G en minimisant une fonction de coût $l_N : (A^N, X^N) \mapsto \mathbb{R}$.
- ★ Garantir la qualité d'approximation : $\forall z \in G, \quad z \approx \varphi^{-1} \circ \psi(z)$

Analyse en composantes principales (ACP)

- ▶ Étant donné z_1, \dots, z_N de G ,
l'ACP produit une famille d'espaces vectoriels A_1, \dots, A_N
par ajout successif d'un vecteur de base orthogonal.

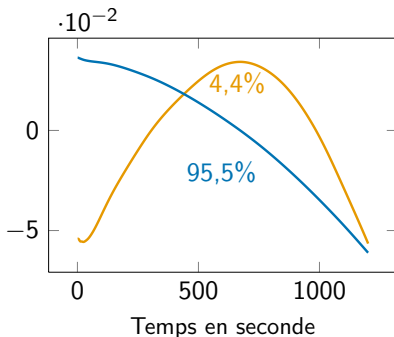
- ▶ La projection sur A_k **préserve les distances euclidiennes**,
 - ⇔ minimise la des normes des résidus,
 - ⇔ maximise les distances entre projections,
 - ⇔ maximise les distances au barycentre.

∴ Th. de Pythagore et identité du variogramme : $\sum_i \sum_j (z_i - z_j) \propto \sum_i (z_i - \bar{z})$

- ▶ **Solution algébrique** : décomposition aux valeurs singulières (SVD).

ACP de la sortie dynamique

- ▶ Les résidus d'approximation en dimension 2 sont presque nuls.
- = Les 2 premières composantes expliquent toute la variance.



- ▶ Le colmatage modifie essentiellement la **pen**te, et en moindre mesure la **cour**bure.

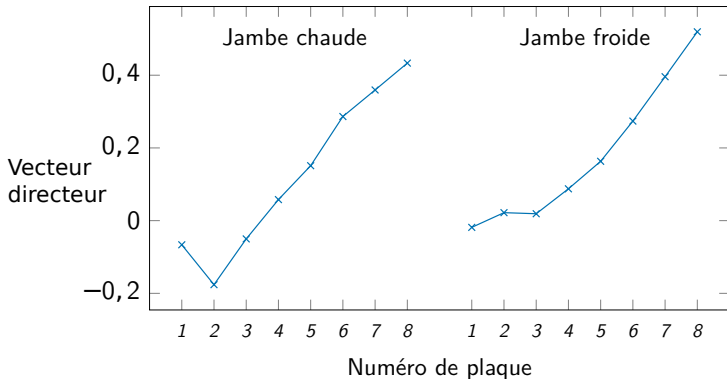
Régression inverse par tranches (SIR)

Ker-Chau Li (1991) « *Sliced Inverse Regression for Dimension Reduction* »

- ▶ Soit un vecteur X et une variable Y aléatoires.
 A est un **espace de réduction de dimension** si la projection ψ est telle que $Y | X \sim Y | \psi(X)$.
- ▶ Sous des conditions peu contraignantes sur X , la courbe de **régression inverse** $E(X | Y)$ est contenue dans cet espace.
- ▶ Algorithme :
 1. Approximation par morceaux (*slices*) de la régression inverse
 2. Estimation de l'espace la contenant au mieux par ACP.

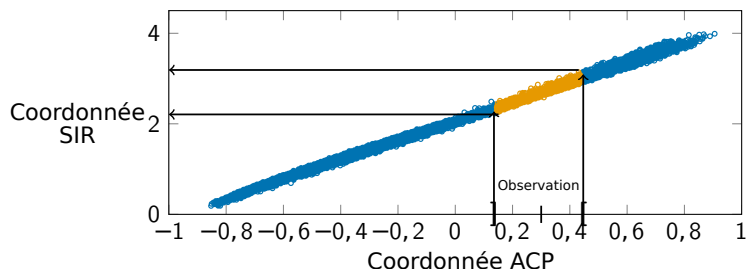
Régression inverse de l'« effet pente »

- ▶ Les points sont très concentrés autour d'une droite.

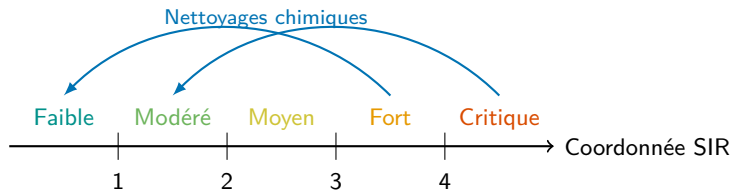


Méthode de diagnostic

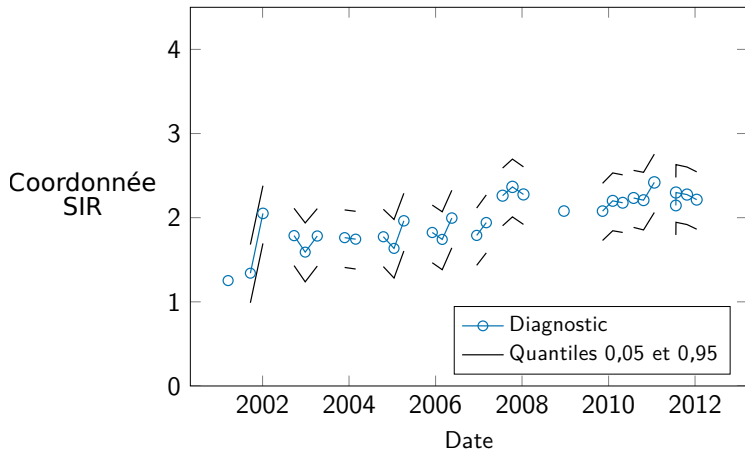
- L'inversion est devenue triviale.



- Échelle de gravité établie empiriquement

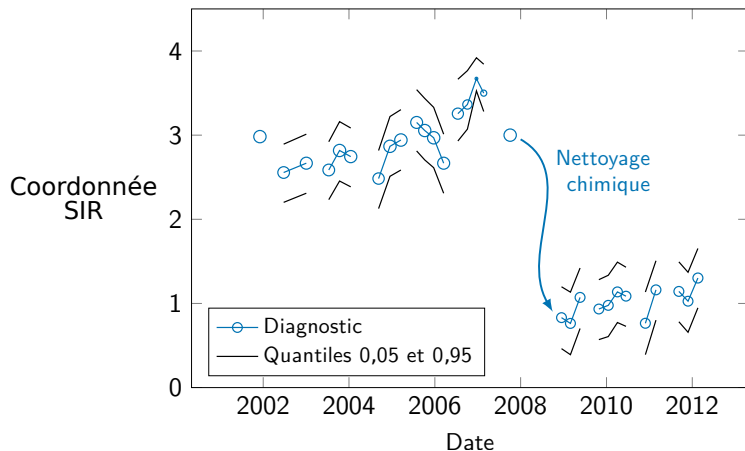


Exemple de diagnostic 1/2



- Générateur de vapeur en bon état.

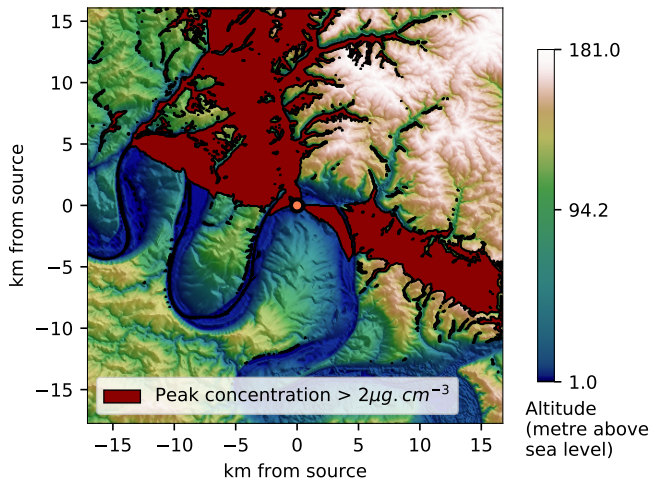
Exemple de diagnostic 2/2



- La fissuration d'un tube a causé un arrêt fortuit.

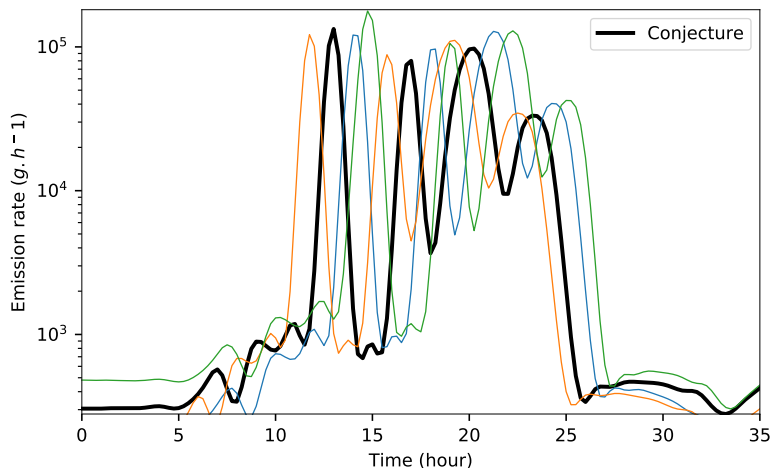
2. Pollution atmosphérique

Carte de décision déterministe (accident de Lubrizol 2013)



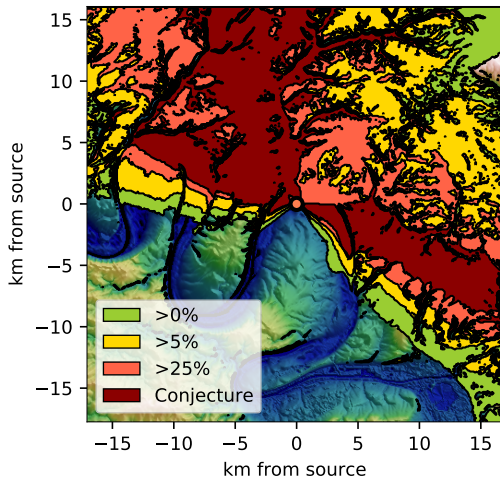
Modèle 3D lagrangien (équation de Langevin) SPRAY (CEA)

Processus stochastique du débit d'émission



- ▶ Sont également perturbés : intensité de la pluie, vitesse et direction du vent

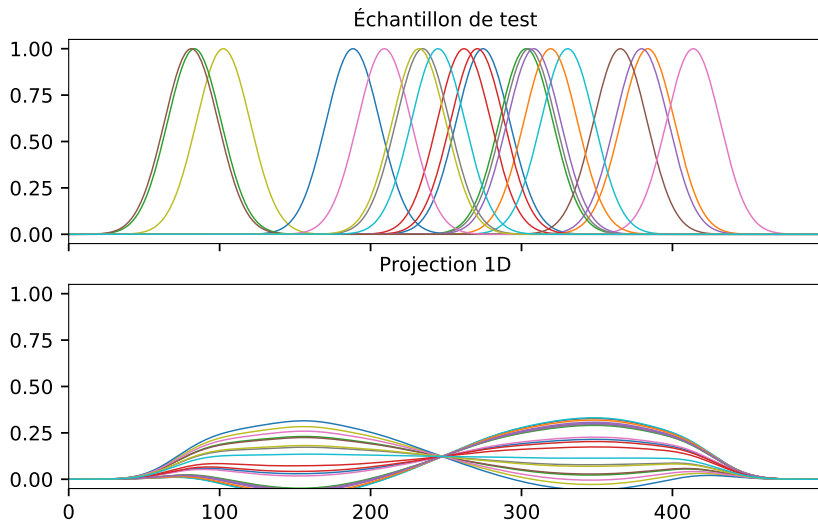
Carte de décision probabiliste



- Probabilités de dépassement estimées avec 100 simulations

Limite de l'ACP

Inspiré de Fukunaga & Olsen (1971) « *An Algorithm for Finding Intrinsic Dimensionality of Data* »



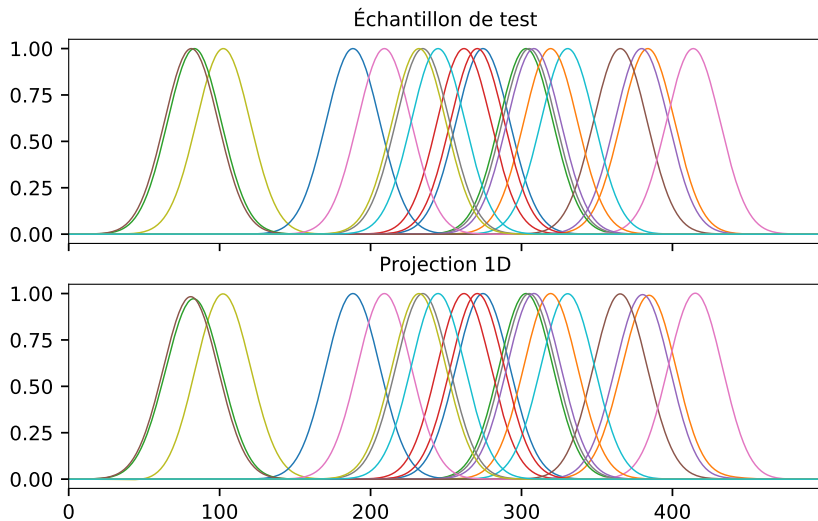
Modèle auto-associatif (MAA)

Stéphane Girard & Iovleff (Auto-Associative Models, Nonlinear Principal Component Analysis, Manifolds and Projection Pursuit) « 2008. »

Algorithme

1. Ajout d'un vecteur de base (minimisation d'une fonction de coût)
Soit ψ_k la projection associée.
 2. Estimation d'une fonction de rattrapage $r_k : \psi_k(x) \mapsto x$,
 3. Remplacer les données par les résidus $x - r_k(\psi_k(x))$, et itérer.
- L'ACP est un MAA
préservant les distances euclidiennes,
avec l'identité comme fonction de rattrapage

Fukunaga & Olsen avec MAA



Fonction de coût

- ▶ Remplacement du critère métrique global par un critère topologique local:

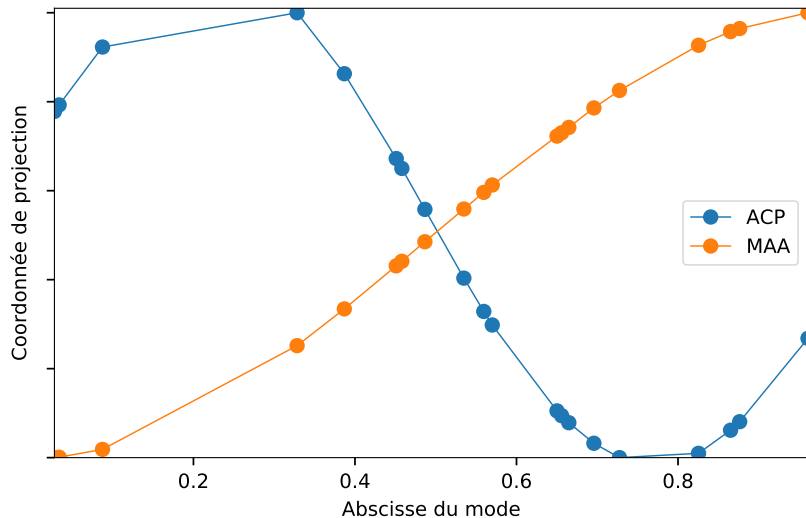
$$\text{maximiser } \frac{\sum_i \psi_k(x_i)^2}{\sum_j \sum_i c_{ij} \psi_k(x_i - x_j)^2} = \frac{V}{V_c}$$

avec $c_{ij} = 1$ si x_i et x_j contigus, 0 sinon.

- ▶ Solution : 1^{er} vecteur propre de $V_c^+ V$
(V_c est singulière)

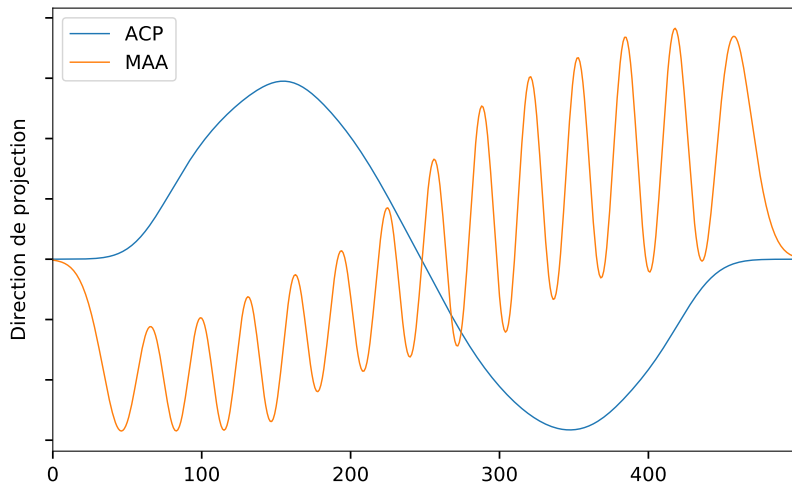
Préservation de la topologie locale 1/2

Coordonnées des projections



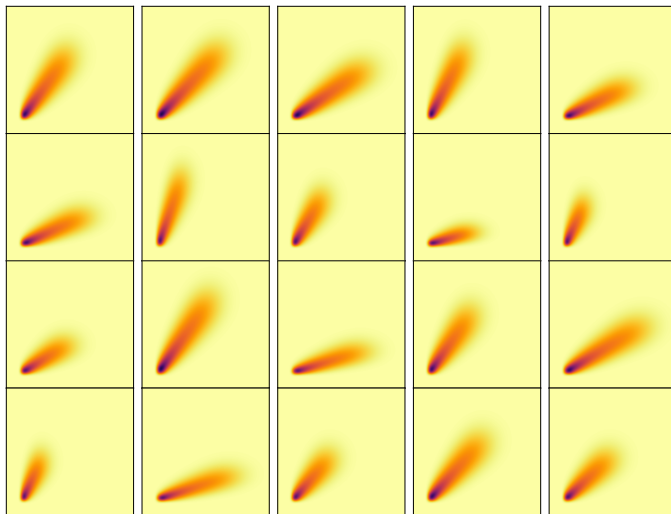
Préservation de la topologie locale 2/2

Vecteur directeur des projections



Modèle jouet Ubik

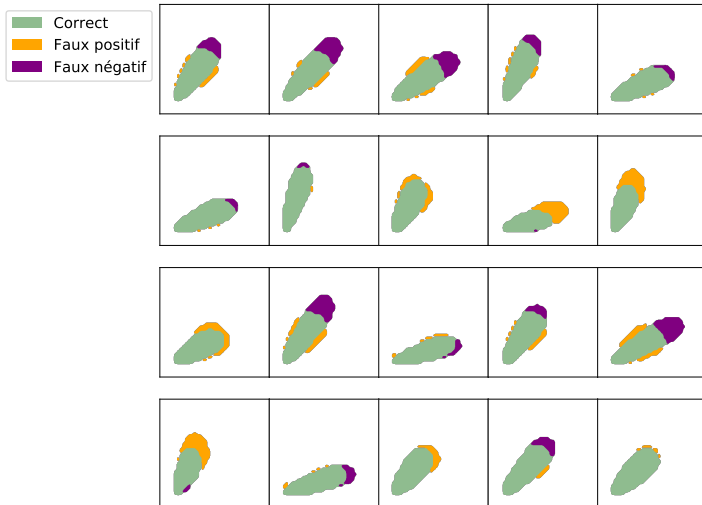
Échantillon de test



Variables : vitesse et direction du vent, diffusion

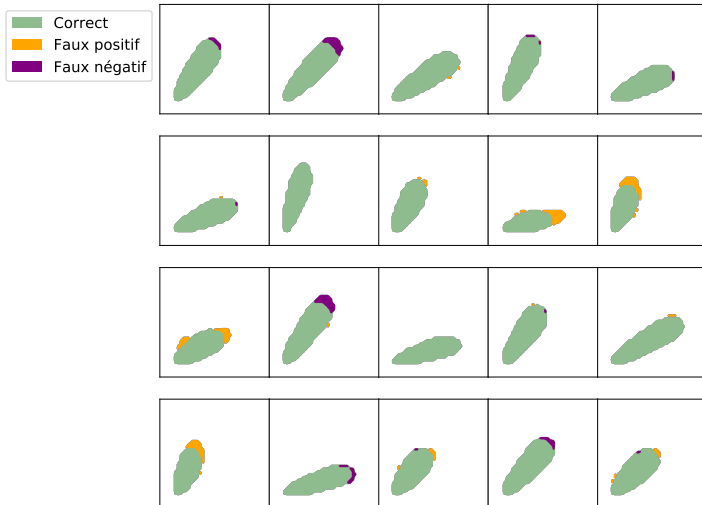
Approximation par ACP (dimension 2)

Prédiction de dépassement de seuil



Approximation par MAA (dimension 2)

Réduction de dimension et émulation par krigeage



Pistes de recherche

- ▶ Projection préalable sur une base fonctionnelle (p. ex. ondelettes)
- ▶ Utiliser d'autres métriques (p. ex. Wasserstein) sans trop compliquer l'optimisation
- ▶ Réduction simultanée de l'entrée et de la sortie : s'inspirer de la SIR?

MADiPA

Modèles Auto-associatifs pour la Dispersion de Polluants dans l'Atmosphère

- ▶ Collaboration avec les équipes Mistis et Modal d'Inria (Stéphane Girard et Serge Iovleff)
- ▶ Embauche d'un post-doctorant (12 mois)
- ▶ Subvention AMIES, programme PEPS2 sur l'environnement¹



Inria



Agence pour les mathématiques
en interaction avec l'entreprise et
la société

- ▶ Je cherche un candidat...

¹<https://www.agence-maths-entreprises.fr/a/?q=fr/node/905>

Merci de votre attention.

