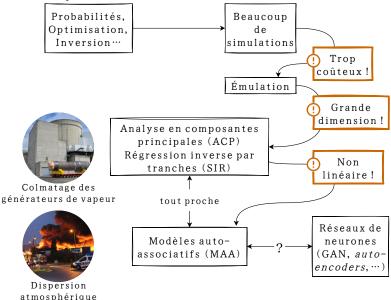


Réduction de dimension par des modèles auto-associatifs

22 novembre 2021

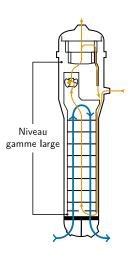
Sylvain Girard (girard@phimeca.com)

Problématique

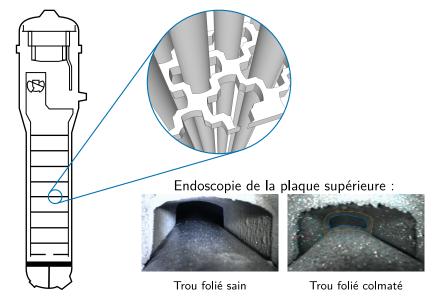


1. Colmatage des générateurs de vapeur

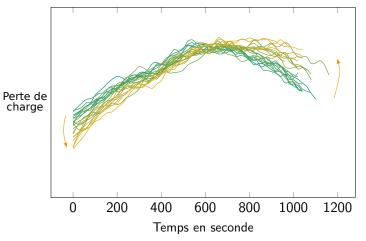




Le colmatage obstrue le passage de la vapeur

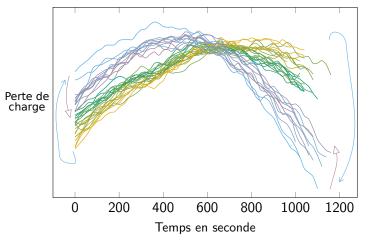


Le colmatage perturbe la réponse dynamique



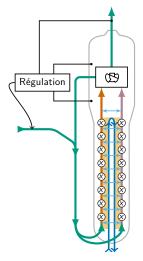
: évolution avant nettoyage chimique

Le colmatage perturbe la réponse dynamique



✓ : nettoyage chimique

Modèle de générateur de vapeur



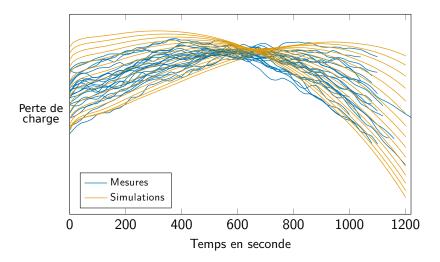
- Entrée : 16 taux de colmatages
- Sortie : signal temporel de pression

Programmé en Modelica avec ThermoSysPro (https://thermosyspro.com/)

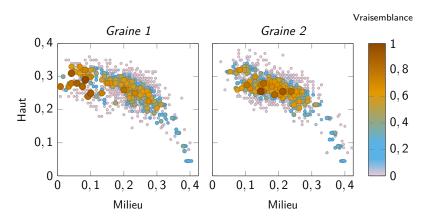




Le modèle reproduit l'effet du colmatage



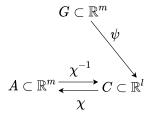
Tentative d'inversion par filtrage particulaire



Le problème inverse de prédiction des taux colmatage à partir des signaux de pression est mal posé.

Réduction de dimension

Réduire la dimension de G plongé dans \mathbb{R}^m , c'est construire un ensemble approché $A \subset \mathbb{R}^m$ doté d'un système de coordonnées de basse dimension $C \subset \mathbb{R}^l$ (l petit).



Réduction non supervisée : A est une (plus ou moins) « bonne approximation » de G (\rightarrow ACP)

Réduction supervisée : C est un « bon point de départ » pour prédire autre chose $(\rightarrow SIR)$

Analyse en composantes principales (ACP)

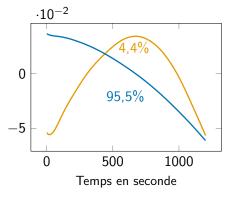
Étant donné z_1,\ldots,z_N de G, l'ACP produit une famille d'espaces vectoriels A_1,\ldots,A_N par ajout successif d'un vecteur de base orthogonal.

- ightharpoonup 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.

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 pente, et en moindre mesure la courbure.

Régression inverse par tranches (SIR)

Soit un vecteur X et une variable Y aléatoires.

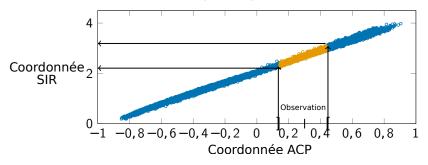
A est un espace de réduction de dimension efficace si la projection ψ est telle que $Y\mid X\sim Y\mid \psi(X).$

Sous des conditions peu contraignantes sur X, la courbe de régression inverse $\mathrm{E}(X\mid Y)$ est contenue dans cet espace.

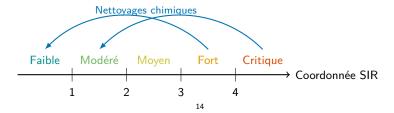
- ► Algorithme :
 - 1. Approximation par morceaux (slices) de la régression inverse
 - 2. Estimation de l'espace la contenant au mieux par ACP.

Résultat : nouvelle méthode de diagnostic

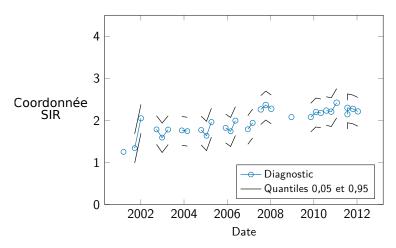
L'inversion est devenue triviale $(1d\rightarrow 1d)$.



Échelle de gravité établie empiriquement

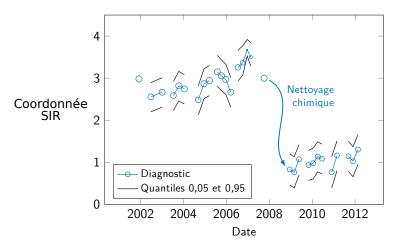


Diagnostic du colmatage 1/2



Générateur de vapeur en bon état.

Diagnostic du colmatage 2/2



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

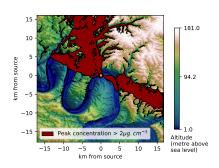
Synthèse intermédiaires

- Les méthodes linéaires fonctionnent souvent très bien !
- Réduction non supervisée : annalyse en composantes principales (ACP)
- Réduction supervisée : régression inverse par tranches (sliced inverse regression, SIR)

Imaginons pixel noir se déplaçant sur un maillage blancs à n cases. Quelle est la dimension de l'espace vectoriel contenant le nuage de point (dans \mathbb{R}^n) de ses positions ?

2. Dispersion de polluant dans l'atmosphère

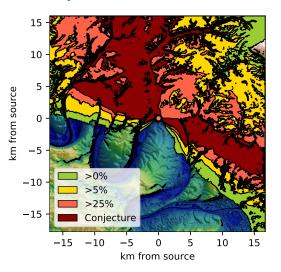




Exemple inspiré de l'accident de l'usine Lubrizol (Rouen, 2013).

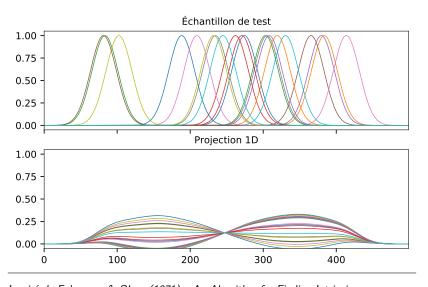
Modèle PMSS

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 »

De l'ACP aux modèles auto-associatifs (MAA)

- Un algorithme pour trouver un espace vectoriel de projection préservant la « topologie »
 - ► =les relations locales de voisinage, là où l'ACP préserve globalement les distances.

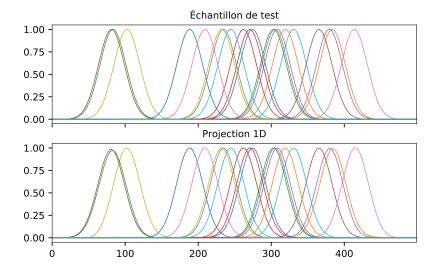
- 2. Utilisation séquentielle de cet algorithme en intercallant à chaque étape l'estimation d'une fonction de « rattrapage » liant les coordonnées de projection aux données de départ
 - On remplace par le résidu à chaque étape

Algorithme des (MAA)

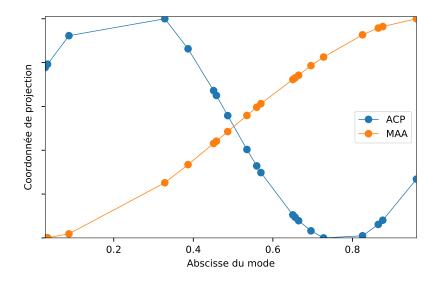
- 1. Ajout d'un vecteur de base. Soit ψ_k la projection associée.
- **2.** Estimation d'une fonction de rattrapage $r_k: \psi_k(x) \mapsto x$,
- **3.** Remplacer les données par les résidus $x r_k(\psi_k(x))$, et itérer.

- L'ACP est un MAA
 - préservant les distances euclidienne,
 - avec l'identité comme fonction de rattrapage

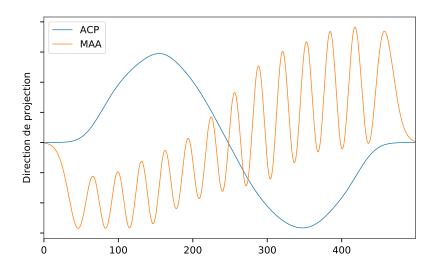
Fukunaga & Olsen avec MAA



Coordonées des projections

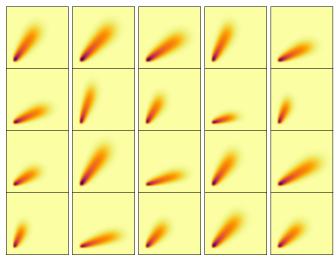


Vecteur directeur des projections



Modèle jouet Ubik

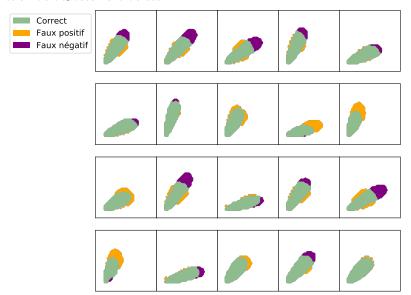
Échantillon de test



Variables : vitesse et direction du vent

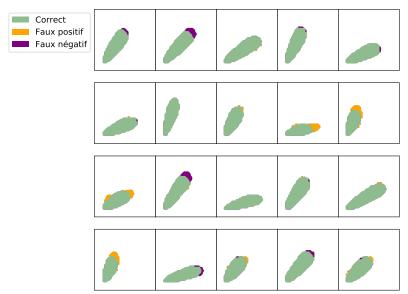
Approximation par ACP (dimension 2)

Prédiction de dépassement de seuil



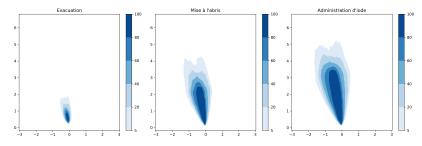
Approximation par MAA (dimension 2)

Réduction de dimension et émulation par krigeage



Émulation du modèle pX IRSN

Nous somme parvennu a émuler par un processus gaussien le modèle à bouffées pX simulant la dispersion à courte distance.



Cartes de probabilité de dépassement de 3 seuils conventionnels ; propagation d'incertitudes de 5 variables d'entrée.

Projet MADiPA

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

- Collaboration avec les équipes Statify et Modal d'Inria (Stéphane Girard et Serge lovleff)
- Subvention AMIES, programme PEPS2 sur l'environnement¹
- Janvier 2022 : embauche de Valentin Pibernus (EC Nantes) pour 6 mois







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

- Amélioration et documentation du code pour faciliter les collaboration
- Combinaison de modèles locaux
- Que faire avec une métrique non euclidienne ?
- Projection préalable pour imposer une régularité
- Version supervisée

Points à retenir

- La réduction de dimension supervisée gagne à être connue
 - Ker-Chau Li (1991) « Sliced Inverse Regression for Dimension Reduction »

- ► Les MAA = réduction de dimesnsion « auto-supervisée » + estimation de « fonctions de rattrapage »
 - Fonctionnent plutôt bien... dans les bons cas.

Notre objectif : aboutir à une « ACP non linéaire » robuste.

► Intérêt de confronter les points de vue : analyse théorie de l'approximation, statistique, algèbre numérique...

Merci de votre attention.

Annexes

Fonction de coût pour les MAA

Remplacement du critère métrique global par un critère topologique local :

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