



**Mieux comprendre pour mieux mesurer.
Défis et avancées en calibration de capteurs de
pollution en environnement ouvert.**

G. Perrin¹, M. Dumon¹, B. Lebental¹

ETICS, 5-10 Octobre, Evian-les-Bains, France | June 2025

¹Université Gustave Eiffel, COSYS, 77420 Champs-sur-Marne, France



La pollution de l'air en France en quelques chiffres

- **Amélioration nette depuis 20 ans** : entre 2005 et 2024, les concentrations en PM_{2,5} et NO₂ ont chuté de respectivement 55% et 50%, entraînant une baisse d'un tiers des décès prématurés liés à la pollution de l'air.
- **L'Ozone, une exception persistante** : contrairement aux autres polluants, l'ozone de basse altitude reste problématique, en raison du réchauffement climatique et du transport de pollution à longue distance.
- **Des expositions très localisées** : en 2024, 800 Franciliens étaient encore exposés à des dépassements réglementaires (vs 5 000 en 2023), principalement près des grands axes routiers et en petite couronne.
- **Enjeu des futures normes** : si les seuils OMS 2030 étaient déjà en vigueur, 2.6M d'habitants d'IDF seraient concernés par un dépassement en 2024.
- **Un poids sanitaire toujours lourd** : en 2019, la pollution entraînait encore ~10 mois de vie perdus par adulte et contribuait à 10-20% des maladies respiratoires chroniques et 5-10% des maladies card. ou métaboliques.

⇒ besoin d'un suivi beaucoup plus **précis** de la pollution à l'échelle **locale**.

La pollution de l'air à Paris en images¹

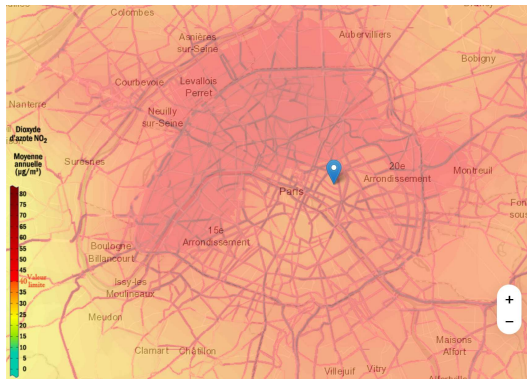


Figure – Concentration moyenne annuelle en NO_2 en 2007.

1. <https://www.airparif.fr/surveiller-la-pollution/bilans-et-cartes-annuels-de-pollution>

La pollution de l'air à Paris en images¹

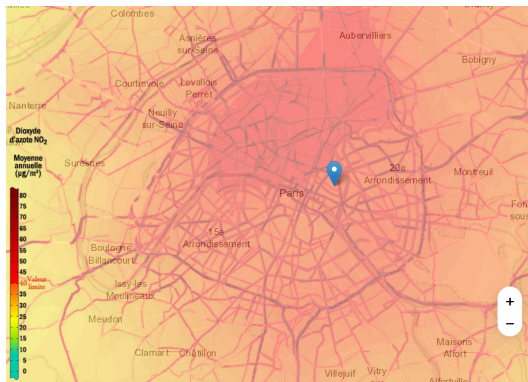


Figure – Concentration moyenne annuelle en NO_2 en 2010.

1. <https://www.airparif.fr/surveiller-la-pollution/bilans-et-cartes-annuels-de-pollution>

La pollution de l'air à Paris en images¹

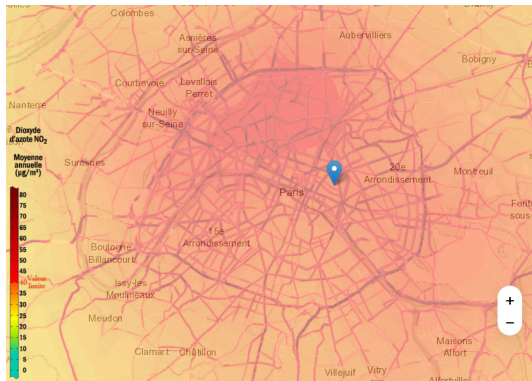


Figure – Concentration moyenne annuelle en NO_2 en 2013.

1. <https://www.airparif.fr/surveiller-la-pollution/bilans-et-cartes-annuels-de-pollution>

La pollution de l'air à Paris en images¹

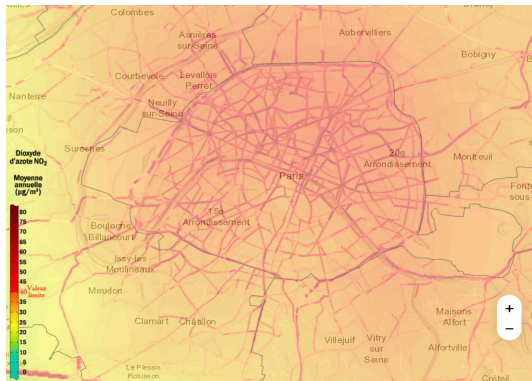


Figure – Concentration moyenne annuelle en NO_2 en 2016.

1. <https://www.airparif.fr/surveiller-la-pollution/bilans-et-cartes-annuels-de-pollution>

La pollution de l'air à Paris en images¹



Figure – Concentration moyenne annuelle en NO_2 en 2019.

1. <https://www.airparif.fr/surveiller-la-pollution/bilans-et-cartes-annuels-de-pollution>

La pollution de l'air à Paris en images¹

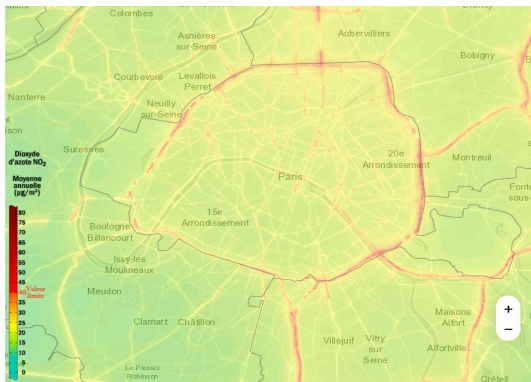


Figure – Concentration moyenne annuelle en NO_2 en 2024.

1. <https://www.airparif.fr/surveiller-la-pollution/bilans-et-cartes-annuels-de-pollution>



Des inégalités d'exposition "très" locales

- **Des différences d'exposition très marquées à l'échelle locale** : un simple changement de parcours – voire de trottoir – ainsi qu'une variation de hauteur dans la rue (enfants/adultes) peut réduire fortement l'inhalation de particules fines et de gaz.
- **Influence du contexte urbain** : la configuration des rues, la présence de bâtiments rapprochés ("rues canyons"), ou encore le sens du vent créent des micro-zones où la pollution stagne et s'accumule.
- **Le déploiement d'applications de guidage** : AirCheck (CEA), Walkable ou Bikeable proposent de comparer des itinéraires non seulement sur la distance, mais aussi sur la qualité de l'air et le confort urbain.



Des inégalités d'exposition "très" locales

- **Des différences d'exposition très marquées à l'échelle locale** : un simple changement de parcours – voire de trottoir – ainsi qu'une variation de hauteur dans la rue (enfants/adultes) peut réduire fortement l'inhalation de particules fines et de gaz.
- **Influence du contexte urbain** : la configuration des rues, la présence de bâtiments rapprochés ("rues canyons"), ou encore le sens du vent créent des micro-zones où la pollution stagne et s'accumule.
- **Le déploiement d'applications de guidage** : AirCheck (CEA), Walkable ou Bikeable proposent de comparer des itinéraires non seulement sur la distance, mais aussi sur la qualité de l'air et le confort urbain.

⇒ **de nouvelles attentes de personnalisation** : en exploitant différentes données en temps réel (capteurs, météo, trafic), on pourrait "imaginer" proposer des recommandations ultra-locales, au jour le jour.

Quelle instrumentation actuelle ?

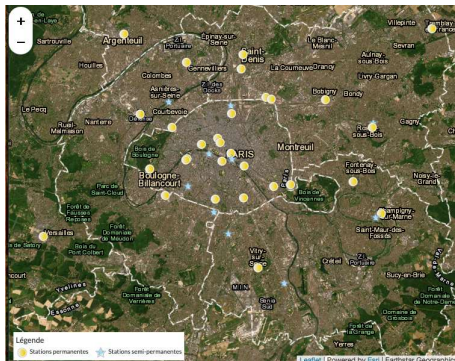


Figure – (seulement) 70 stations de mesure, dont plus d'une cinquantaine de stations permanentes.

⇒ de grandes attentes pour le déploiement de capteurs bas-coûts, performants et fiables.



Les capteurs low-cost

- Exemples : capteurs optiques, électrochimiques, semi-conducteurs, nanocapteurs (nanoparticules métalliques, **nanotubes de carbone**, graphène...),...

(+) **Avantages** :

- grande sensibilité (détection à faible concentration, jusqu'à qq ppb),
- consommation énergétique faible,
- miniaturisation possible,
- faible coût, production de masse possible.

(-) **Limites** :

- faible sélectivité,
- sensibilité à l'humidité, à la température, ...
- calibration difficile,
- dérives dans le temps.

⇒ des capteurs souvent encore en développement, en phase de transfert entre recherche et marché.

Un exemple de capteur développé à UGE

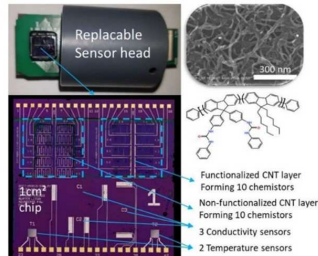


Figure – Sensibilités probables d'après la littérature : Température, Humidité relative (RH), CO , CO_2 , O_3 , NO , NO_2 ,...

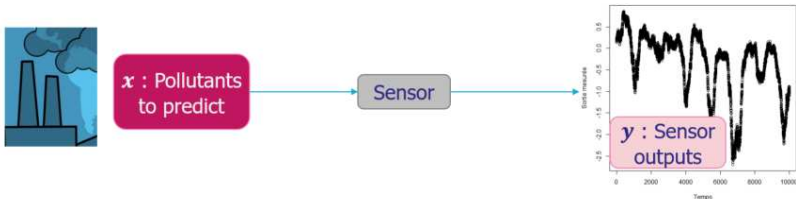
- Comment exploiter les sorties renvoyées par de tels capteurs ?
- Peut-on évaluer l'effet de la fonctionnalisation ?
- Quelles futures fonctionnalisations proposer ?



Plan de la présentation

- 1 Introduction
- 2 Formalisation du déploiement d'un capteur innovant
- 3 Difficulté 1 : une manque de sélectivité
- 4 Difficulté 2 : des entrées incertaines
- 5 Difficulté 3 : la prise en compte du temps
- 6 Conclusions

Formalisation du déploiement d'un capteur innovant



Une approche en deux-temps

1. Calibration du capteur $\leftrightarrow$ Problème **direct**.

- Recueil d'observations $\mathcal{D}_n = (x_i, y_i)_{i=1}^n$.
- Approximation de la relation $x \rightarrow y$ (loi de calibration).

2. Exploitation du capteur $\leftrightarrow$ Problème **inverse**.

- Observation d'une sortie capteur y_* .
- Estimation de la valeur associée de x_* à partir de la loi de calibration.

Un exemple simple

Cas linéaire gaussien univarié : $y_i = ax_i + b + \varepsilon_i$, $1 \leq i \leq n$. $\mathcal{D}_n = (x_i, y_i)_{i=1}^n$.

Calibration (Formalisme bayésien)

- Priors : $\beta = (a, b) \sim \mathcal{N}(\beta_0, C_0) \perp \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.
- Posterior : $\mathbf{y}_n = (y_1, \dots, y_n)$, $\mathbf{H}_n = [\mathbf{x}_n \ \mathbf{1}_n]$, $\mathbf{x}_n = (x_1, \dots, x_n)$.

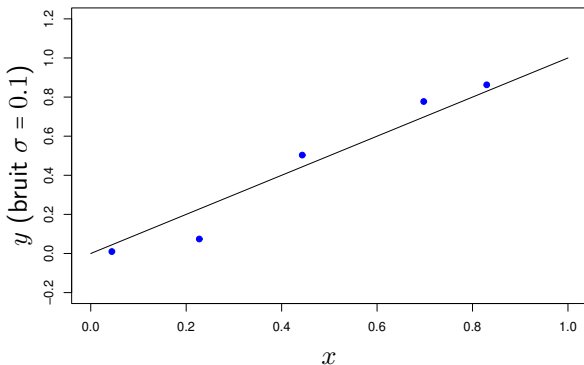
$$\beta | \mathcal{D}_n \sim \mathcal{N}(\mathbf{R}^{-1}(\sigma^{-2} \mathbf{H}_n^T \mathbf{y}_n + \mathbf{C}_0^{-1} \beta_0), \mathbf{R}^{-1}), \quad \mathbf{R} = \sigma^{-2} \mathbf{H}_n^T \mathbf{H}_n + \mathbf{C}_0^{-1}.$$

Exploitation (Formalisme bayésien)

- Priors : $x_* \sim \mathcal{U}(0, 1) \perp \varepsilon_* \sim \mathcal{N}(0, \sigma^2)$.
- Posterior : $\mathbf{h}_* = (x_*, 1)$, $\mathbf{Q}(x_*) = \mathbf{R} + \sigma^{-2} \mathbf{h}_* \mathbf{h}_*^T$,
 $\tilde{\beta}(x_*, y_*) = \sigma^{-2} (\mathbf{H}_n^T \mathbf{y}_n + \mathbf{h}_* y_*) + \mathbf{C}_0^{-1} \beta_0$.

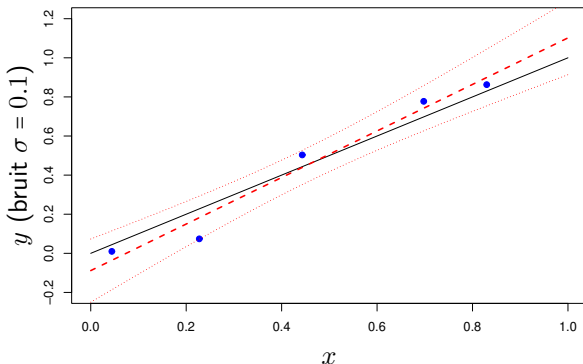
$$\pi[x_* | y_*, \mathcal{D}_n] \propto 1_{[0,1]}(x_*) \sqrt{\det(\mathbf{Q}(x_*))} \exp\left(\frac{1}{2} \tilde{\beta}(x_*, y_*)^T \mathbf{Q}(x_*)^{-1} \tilde{\beta}(x_*, y_*)\right).$$

Illustration graphique - approximation de $x \mapsto y(x)$



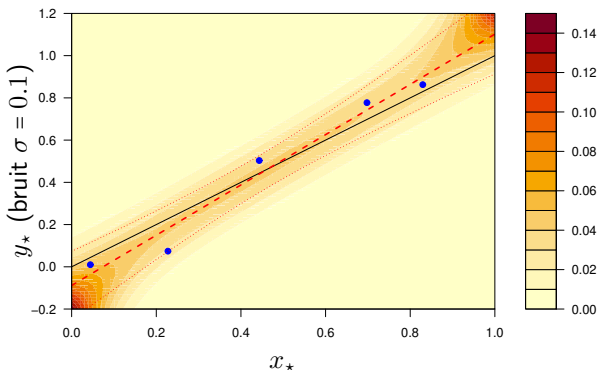
■ Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.

Illustration graphique - approximation de $x \mapsto y(x)$



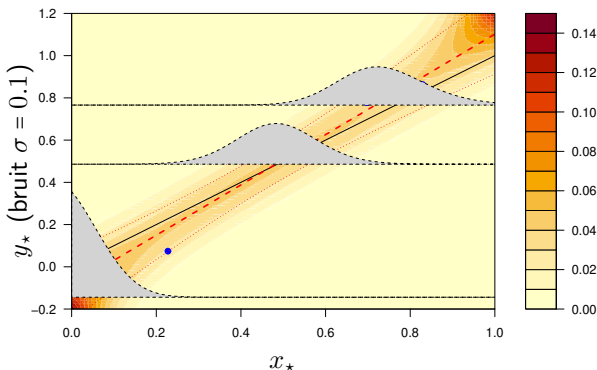
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
Rouge $\leftrightarrow$ fonction de calibration approchée + IC 95%.

Illustration graphique - inversion de $x \mapsto y(x)$



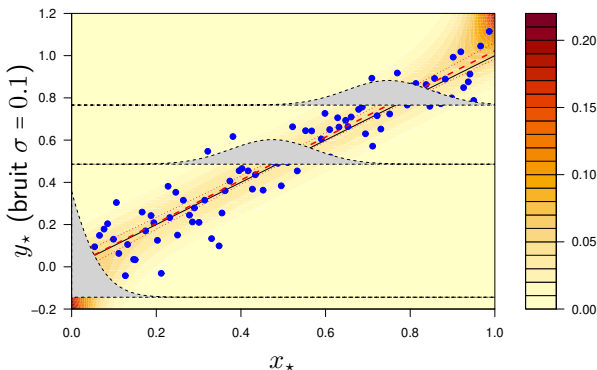
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration. Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_*|y_*, \mathcal{D}_n$. Aires grisées $\leftrightarrow$ zooms sur les PDF de $x_*|y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

Illustration graphique - inversion de $x \mapsto y(x)$



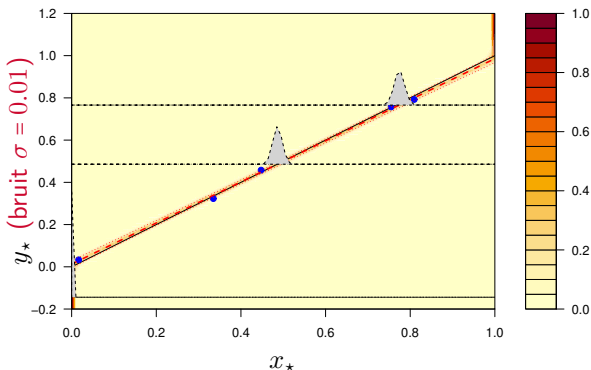
- Points bleus ↔ Info. dispo. / Ligne noire ↔ "vraie" fonction de calibration.
Rouge ↔ fonction de calibration approchée. Isoniveaux ↔ PDF de $x_*|y_*, \mathcal{D}_n$.
Aires grisées ↔ zooms sur les PDF de $x_*|y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

Illustration graphique - inversion de $x \mapsto y(x)$



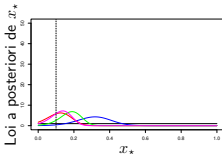
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration. Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_\star | y_\star, \mathcal{D}_n$. Aires grises $\leftrightarrow$ zooms sur les PDF de $x_\star | y_\star, \mathcal{D}_n$ pour des valeurs spécifiques de $y_\star$.

Illustration graphique - inversion de $x \mapsto y(x)$

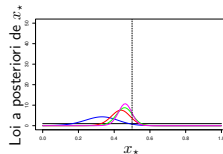


- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_*|y_*, \mathcal{D}_n$.
Aires grises $\leftrightarrow$ zooms sur les PDF de $x_*|y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

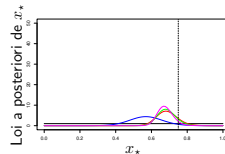
Illustration graphique - extension multi-capteurs



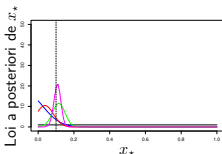
(a) $n = 5, \sigma = 0.1$



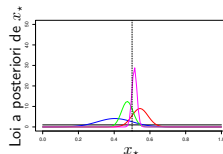
(b) $n = 5, \sigma = 0.1$



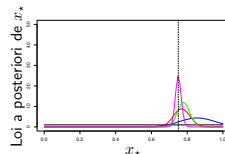
(c) $n = 5, \sigma = 0.1$



(d) $n = 100, \sigma = 0.1$



(e) $n = 100, \sigma = 0.1$

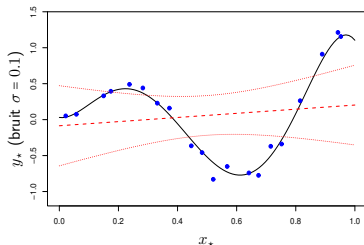


(f) $n = 100, \sigma = 0.1$

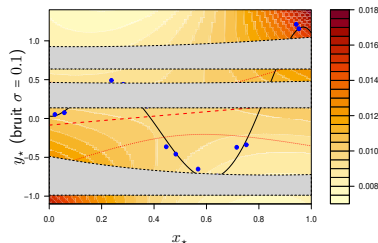
$n \leftrightarrow$ nombre de points pour l'apprentissage. Ligne noire verticale $\leftrightarrow x_{\star}^{\text{vrai}}$ / bleu $\leftrightarrow$ une seule mesure / rouge $\leftrightarrow$ 5 répétitions de mesure / vert $\leftrightarrow$ 10 répétitions de mesure / magenta $\leftrightarrow$ 100 répétitions de mesure.

Illustration graphique - extension non-linéaire

Approximation poly. d'ordre 1.



(g) Pb. direct

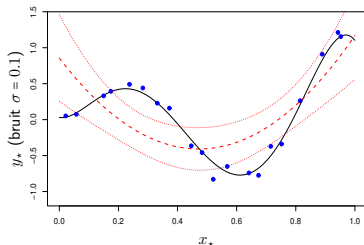


(h) Pb. inverse

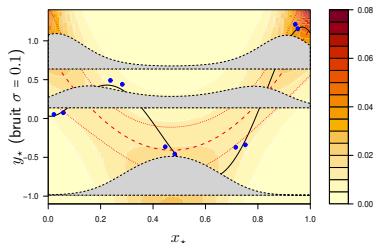
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
- Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_*|y_*, \mathcal{D}_n$.
- Aires grises $\leftrightarrow$ zooms sur les PDF de $x_*|y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

Illustration graphique - extension non-linéaire

Approximation poly. d'ordre 2.



(i) Pb. direct

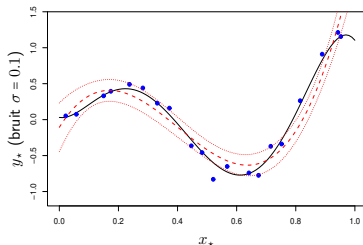


(j) Pb. inverse

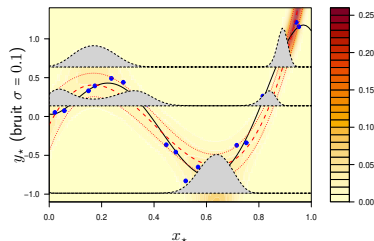
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
- Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_*|y_*, \mathcal{D}_n$.
- Aires grises $\leftrightarrow$ zooms sur les PDF de $x_*|y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

Illustration graphique - extension non-linéaire

Approximation poly. d'ordre 3.



(k) Pb. direct

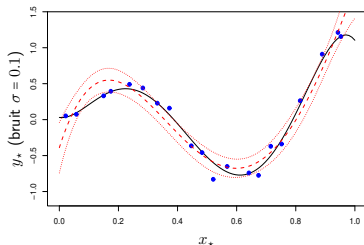


(l) Pb. inverse

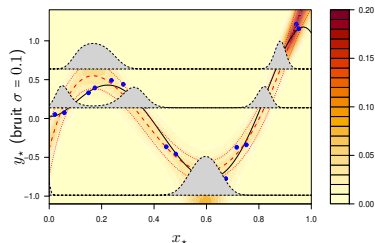
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
- Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_* | y_*, \mathcal{D}_n$.
- Aires grises $\leftrightarrow$ zooms sur les PDF de $x_* | y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

Illustration graphique - extension non-linéaire

Approximation poly. d'ordre 4.



(m) Pb. direct

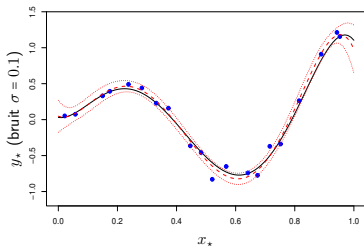


(n) Pb. inverse

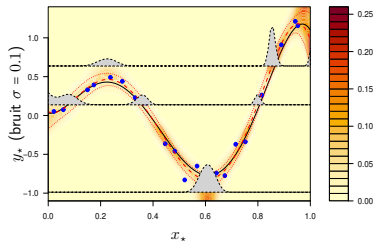
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
- Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_* | y_*, \mathcal{D}_n$.
- Aires grisées $\leftrightarrow$ zooms sur les PDF de $x_* | y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

Illustration graphique - extension non-linéaire

Approximation poly. d'ordre 5.



(o) Pb. direct

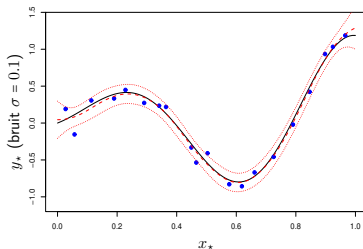


(p) Pb. inverse

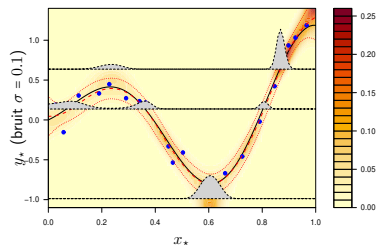
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
- Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_*|y_*, \mathcal{D}_n$.
- Aires grisées $\leftrightarrow$ zooms sur les PDF de $x_*|y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .

Illustration graphique - extension non-linéaire

Approximation par régression par processus gaussien (moyenne nulle, Matern-5/2).



(q) Pb. direct



(r) Pb. inverse

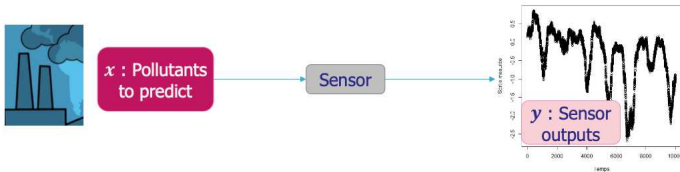
- Points bleus $\leftrightarrow$ Info. dispo. / Ligne noire $\leftrightarrow$ "vraie" fonction de calibration.
- Rouge $\leftrightarrow$ fonction de calibration approchée. Isoniveaux $\leftrightarrow$ PDF de $x_*|y_*, \mathcal{D}_n$.
- Aires grises $\leftrightarrow$ zooms sur les PDF de $x_*|y_*, \mathcal{D}_n$ pour des valeurs spécifiques de y_* .



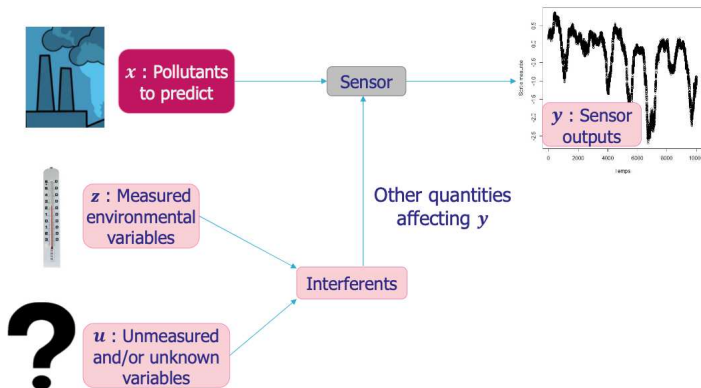
Plan de la présentation

- 1 Introduction
- 2 Formalisation du déploiement d'un capteur innovant
- 3 Difficulté 1 : une manque de sélectivité**
- 4 Difficulté 2 : des entrées incertaines
- 5 Difficulté 3 : la prise en compte du temps
- 6 Conclusions

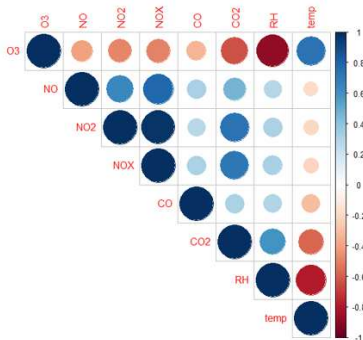
Le problème du manque de sélectivité...



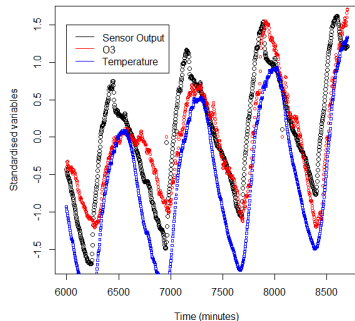
Le problème du manque de sélectivité...



...avec des variables environnementales fortement corrélées



(s) Un exemple de corrélations



(t) Evolutions temporelles associées

⇒ difficile d'identifier les variables ayant un "vrai" impact sur les sorties des capteurs.

Un exemple analytique

$$y = \frac{3}{2} \log(x + 4) + \operatorname{atan}\left(\frac{z_1}{2}\right) \left(1 + 0.1 \frac{z_1}{2}\right) + \operatorname{atan}\left(\frac{z_2}{2}\right) \left(1 + 0.1 \frac{z_2}{2}\right) \\ + 0.1 \operatorname{atan}\left(\frac{z_3}{2}\right) \left(1 + 0.1 \frac{z_3}{2}\right) + 3 \cos\left(\frac{z_1 + z_3}{6}\right) + 2 \cos\left(\frac{x + z_2}{6}\right) + 0z_4 + 0z_5 + \alpha_u u$$

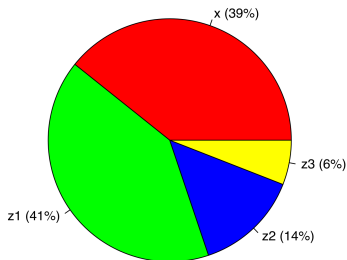


Figure – Décomposition de variance (PME) du vrai modèle

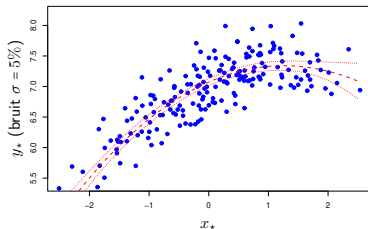
■ Entrées :

- 200 observations (bruitées) de $x, z_1, z_2, z_3, z_4, z_5$
- 5% noise, 0.8 de correlation

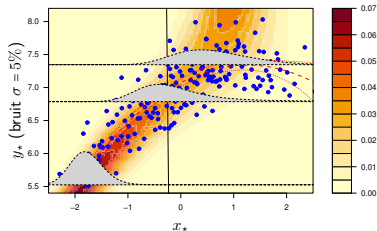
■ Résultats attendus :

- Sélection de $z_1, z_2, z_3,$
- Reconstruction de la décomposition de variance pour identifier les variables attendues.

Un exemple analytique



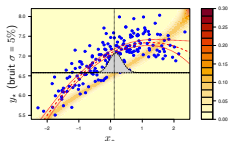
(a) Pb. direct



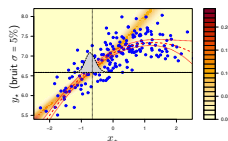
(b) Pb. inverse

⇒ le fait de ne pas intégrer la dépendance en z réduit fortement l'utilité du capteur.

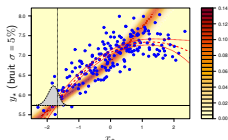
Un exemple analytique



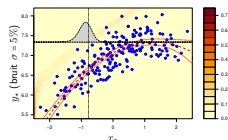
(c) $x_* | y_*(\omega_1), z(\omega_1)$



(d) $x_* | y_*(\omega_2), z(\omega_2)$



(e) $x_* | y_*(\omega_3), z(\omega_3)$



(f) $x_* | y_*(\omega_4), z(\omega_4)$

⇒ en intégrant les mesures de $(z_1, z_2, z_3, z_4, z_5)$, on améliore sensiblement la précision du capteur.



Sélection des paramètres pertinents de z

Pour ne sélectionner *que* les composantes influentes de z , nous avons proposé¹ une méthode reposant sur :

- l'exploration d'un grand nombre de modèles polynomiaux fonction de l'ensemble des sous ensembles de composantes z_α de z , avec $\alpha \subset \{1, \dots, d_z\}$ le vecteur des indices conservées,
- la sélection de modèles *optimaux* via un compromis variance de prédiction/complexité de modèle à travers un critère BIC :

$$\text{BIC}(\alpha, f_\alpha) = n \times \log(V(\alpha, f_\alpha)) + \#f_\alpha \times \log(n),$$

- $V(\alpha, f_\alpha) \leftrightarrow$ variance de prédiction du modèle f_α ,
- $\#f_\alpha \leftrightarrow$ nombre de paramètres sur lequel repose le modèle f_α ,
- $n \leftrightarrow$ nombre de points d'apprentissage.

1. M. Dumon, Addressing uncertainties in uncontrolled environments, Thèse 2025.

Résultats sur le cas analytique

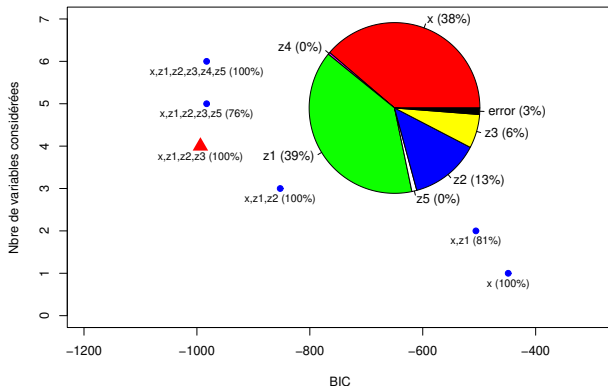


Figure – Fréquence des variables sélectionnées (100 répétitions) : z_1, z_2, z_3 : 100% / z_4 : 8% / z_5 : 10%. Triangle rouge : modèle sélectionné.

Retour au capteur développé à UGE

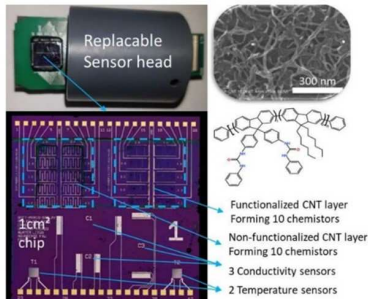
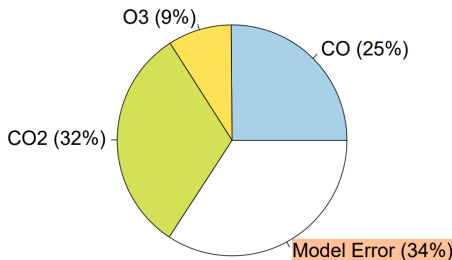


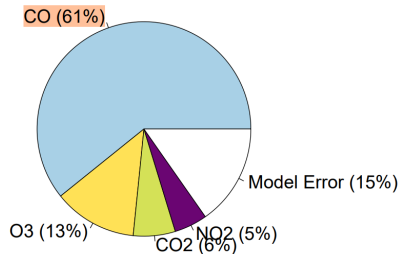
Figure – Variables environnementales mesurées : Température, Humidité relative (RH), CO , CO_2 , O_3 , NO , NO_2 ,..., Vitesse du vent,...

- ⇒ Peut-on évaluer l'effet de la fonctionnalisation ?
- ⇒ Quelle instrumentation complémentaire minimale ?

Retour au capteur développé à UGE



(a) PME - cas non fonctionnalisé



(b) PME - cas fonctionnalisé

⇒ La fonctionnalisation semble avoir amélioré la **sélectivité** du capteur :

- réduction de l'erreur de modèle (moins d'influence des variables non-mesurée ?),
- augmentation de la part attribuée au *CO* (au détriment de *CO₂* ?),
- mais apparition de *NO₂* dans les quantités à mesurer...

Quelques détails sur les PME (1/3)

Le calcul du Proportional Marginal Effect (PME) de la j^{th} composante de $\mathbf{W} = (x, z_1, \dots, z_5)$ est donné par :

$$\text{PME}_j = \sum_{\pi \in \mathcal{S}_D} p(\pi) \left[S^T(C_{\pi(j)}(\pi)) - S^T(C_{\pi(j)-1}(\pi)) \right],$$

- $D = \{1, \dots, d\} \leftrightarrow$ ensemble des indices,
- $\mathcal{S}_D \leftrightarrow$ ensemble des permutations de D ,
- $\pi = (\pi_1, \dots, \pi_d) \leftrightarrow$ élément particulier de $\mathcal{S}_D$,
- $\pi(j)$ est la position de j dans π telle que $\pi_{\pi(j)} = j$,
- $C_j(\pi) = \{\pi_k : k \leq j\}$ est l'ensemble des j premiers indices dans l'ordre fourni par π avec la convention $C_0(\pi) = \emptyset$,
- pour tout $A \subset D$, et $S^T(A)$ est l'**indice de Sobol total** associé à A ,
- $p(\pi)$ est une fonction de poids telle que :

$$p(\pi) = \frac{L(\pi)}{\sum_{\sigma \in \mathcal{S}_D} L(\sigma)}, \quad L(\pi) = \left(\prod_{k \in D} S^T(C_k(\pi)) \right)^{-1}.$$



Quelques détails sur les PME (2/3)

Quelques propriétés intéressantes des PME¹ :

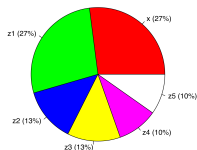
- **Positivité et interprétabilité** : les valeurs PME sont non négatives et somment à la variance totale du modèle.
- **Pertinence causale** : une variable **sans effet causal** aura un PME égal à **zéro**, même si elle est statistiquement corrélée à la sortie.
- **Hiérarchie d'influence** : une valeur PME élevée correspond à une forte influence sur la sortie du modèle, dans le sens où la proportion de variance qui lui est attribuée est élevée.

1. M. Herin et al. PME for Global Sensitivity Analysis, 2024

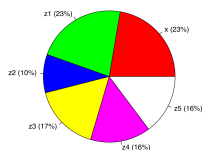
Quelques détails sur les PME (3/3)

$$y = \frac{3}{2} \log(x + 4) + \operatorname{atan}\left(\frac{z_1}{2}\right) \left(1 + 0.1 \frac{z_1}{2}\right) + \operatorname{atan}\left(\frac{z_2}{2}\right) \left(1 + 0.1 \frac{z_2}{2}\right) \\ + 0.1 \operatorname{atan}\left(\frac{z_3}{2}\right) \left(1 + 0.1 \frac{z_3}{2}\right) + 3 \cos\left(\frac{z_1 + z_3}{6}\right) + 2 \cos\left(\frac{x + z_2}{6}\right) + 0z_4 + 0z_5 + \alpha_u u$$

Comparaison à d'autres indices.



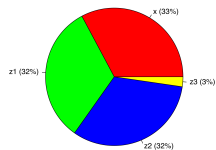
(c) Shapley



(d) Sobol Totaux



(e) PME



(f) Cas indépendant



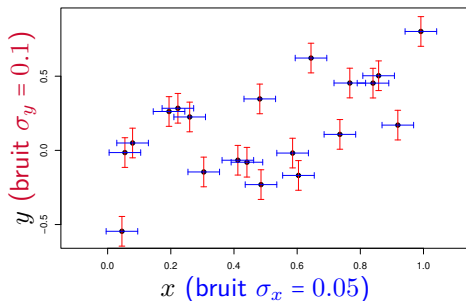
Plan de la présentation

- 1 Introduction
- 2 Formalisation du déploiement d'un capteur innovant
- 3 Difficulté 1 : une manque de sélectivité
- 4 Difficulté 2 : des entrées incertaines**
- 5 Difficulté 3 : la prise en compte du temps
- 6 Conclusions

Un exemple simple

Cas linéaire gaussien : $\mathcal{D}_n = (x_i^{\text{mes}}, z_i^{\text{mes}}, y_i^{\text{mes}})_{i=1}^n$, avec pour tout $1 \leq i \leq n$:

$$\begin{cases} y_i^{\text{mes}} = \beta_0 + \beta_x x_i + \beta_z z_i + \varepsilon_i^y, \\ x_i^{\text{mes}} = x_i + \varepsilon_i^x, \quad z_i^{\text{mes}} = z_i + \varepsilon_i^z, \\ (\varepsilon_i^x, \varepsilon_i^z, \varepsilon_i^y) \sim \mathcal{N}(\mathbf{0}, \text{diag}(\sigma_x^2, \sigma_z^2, \sigma_y^2)), \end{cases} \quad \begin{pmatrix} x_i \\ z_i \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, S^2 \begin{bmatrix} 1 & \rho_{xz} \\ \rho_{xz} & 1 \end{bmatrix}\right).$$



Un exemple simple

Estimation pratique de $\beta = (\beta_0, \beta_x, \beta_z)$ à partir de $\mathcal{D}_n$:

$$\begin{aligned}
 \mathbf{y}_n^{\text{mes}} &= \beta_0 \mathbf{1}_n + \beta_x \mathbf{x}_n + \beta_z \mathbf{z}_n + \boldsymbol{\epsilon}_n^y \\
 &= \beta_0 \mathbf{1}_n + \beta_x \mathbf{x}_n^{\text{mes}} + \beta_z \mathbf{z}_n^{\text{mes}} + \boldsymbol{\epsilon}_n^y - \beta_x \boldsymbol{\epsilon}_n^x - \beta_z \boldsymbol{\epsilon}_n^z \\
 &= \beta_0 \mathbf{1}_n + \beta_x (\mathbf{x}_n^{\text{mes}} - \bar{x}_n^{\text{mes}} \mathbf{1}_n) + \beta_z (\mathbf{z}_n^{\text{mes}} - \bar{z}_n^{\text{mes}} \mathbf{1}_n) + (\mathbf{e}_n^{x,z,y} - \bar{e}_n^{x,z,y} \mathbf{1}_n) \\
 &= \mathbf{H}_n \beta + (\mathbf{e}_n^{x,z,y} - \bar{e}_n^{x,z,y} \mathbf{1}_n) \quad (\mathbf{e}_n^{x,z,y} = \boldsymbol{\epsilon}_n^y - \beta_x \boldsymbol{\epsilon}_n^x - \beta_z \boldsymbol{\epsilon}_n^z).
 \end{aligned}$$

$$\widehat{\beta} = (\mathbf{H}_n^T \mathbf{H}_n)^{-1} \mathbf{H}_n^T \mathbf{y}_n^{\text{mes}}, \quad \mathbf{H}_n := [\mathbf{1}_n \quad (\mathbf{x}_n^{\text{mes}} - \bar{x}_n^{\text{mes}} \mathbf{1}_n) \quad (\mathbf{z}_n^{\text{mes}} - \bar{z}_n^{\text{mes}} \mathbf{1}_n)].$$

On peut alors montrer que :

$$\text{Var}(\widehat{\beta}_x) \approx \frac{1}{n} \frac{\sigma_y^2 + \sigma_x^2 \beta_x^2 + \sigma_z^2 \beta_z^2}{S^2(1 - \rho_{xz}^2)}.$$

Gestion des entrées incertaines

$$\begin{aligned}
 \text{(Calibration)} \quad y_i^{\text{mes}} &= \mathbf{f}(x_i^{\text{mes}} + \epsilon_i^x, z_i^{\text{mes}} + \epsilon_i^z)^T \boldsymbol{\beta} + \epsilon_i^y + \delta_i^{\text{mod}}, \quad 1 \leq i \leq n, \\
 \text{(Exploitation)} \quad y_\star^{\text{mes}} &= \mathbf{f}(x_\star, z_\star^{\text{mes}} + \epsilon_\star^z)^T \boldsymbol{\beta} + \epsilon_\star^y + \delta_i^{\text{mod}}.
 \end{aligned}$$

La prise en compte des incertitudes sur les entrées a pour conséquence :

→ sur le problème **direct** (calibration) :

- une perte du caractère explicite de la loi de $\boldsymbol{\beta} | \mathcal{D}_n$,
- une augmentation de variance de $\boldsymbol{\beta} | \mathcal{D}_n$, un besoin réduit de δ_i^{mod} ,

→ sur le problème **inverse** (exploitation) :

- une vraisemblance difficile à calculer (espérance sur $\boldsymbol{\beta}$, produit/composition de gaussiennes),
- la possibilité de prédire $x_\star$ plutôt que sa valeur bruitée (et donc une potentielle **réduction** de variance d'estimation).

⇒ des simplifications sont possibles lorsque les erreurs sur les entrées sont petites (développements de Taylor, approximations gaussiennes,...).

Un exemple analytique

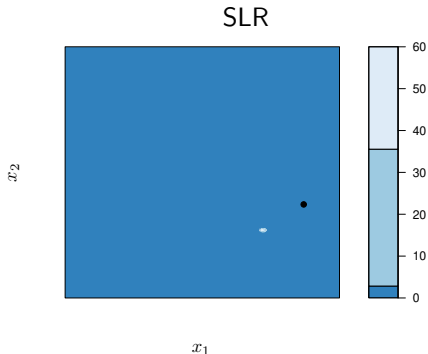


Figure – Point noir $\leftrightarrow$ vraie valeur de x_* . Aires colorées $\leftrightarrow$ régions de prédiction de niveaux 25%, 50%, 75%, 90%, 95%, 99% and 100%.

Un exemple analytique

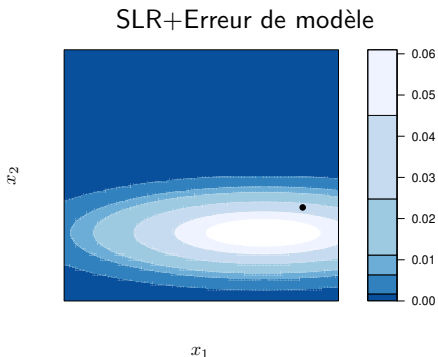


Figure – Point noir $\leftrightarrow$ vraie valeur de x_* . Aires colorées $\leftrightarrow$ régions de prédiction de niveaux 25%, 50%, 75%, 90%, 95%, 99% and 100%.

Un exemple analytique

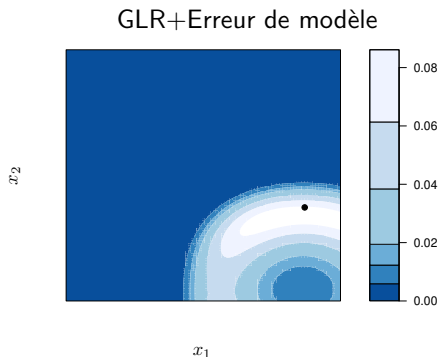


Figure – Point noir $\leftrightarrow$ vraie valeur de x_* . Aires colorées $\leftrightarrow$ régions de prédiction de niveaux 25%, 50%, 75%, 90%, 95%, 99% and 100%.

Un exemple analytique

GLR+Erreur de modèle+Incertitudes sur les entrées

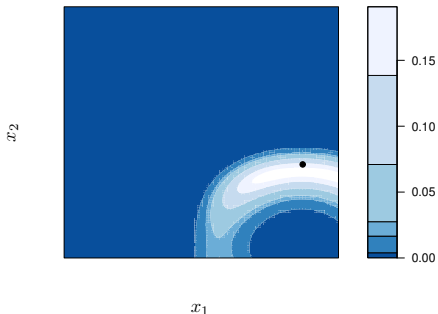


Figure – Point noir $\leftrightarrow$ vraie valeur de x_* . Aires colorées $\leftrightarrow$ régions de prédiction de niveaux 25%, 50%, 75%, 90%, 95%, 99% and 100%.

Un exemple analytique

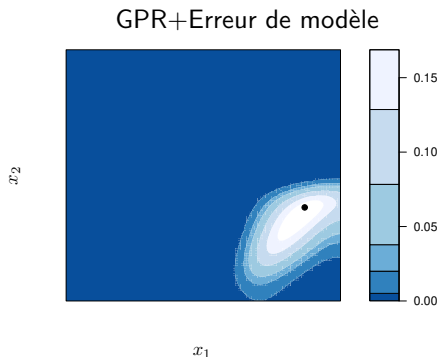


Figure – Point noir $\leftrightarrow$ vraie valeur de x_* . Aires colorées $\leftrightarrow$ régions de prédiction de niveaux 25%, 50%, 75%, 90%, 95%, 99% and 100%.

Un exemple analytique

GPR+Erreur de modèle+Incertitudes sur les entrées

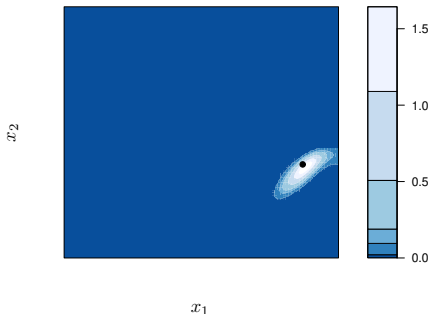


Figure – Point noir $\leftrightarrow$ vraie valeur de x_* . Aires colorées $\leftrightarrow$ régions de prédiction de niveaux 25%, 50%, 75%, 90%, 95%, 99% and 100%.

Retour au capteur développé à UGE

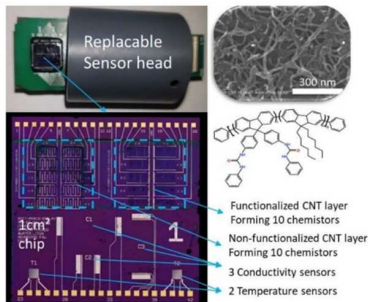


Figure – Variables environnementales considérées : Température, pH, HClO.

- ⇒ Peut-on utiliser le capteur pour surveiller les valeurs de pH et HClO dans l'eau ?
- ⇒ Quel effet du modèle de calibration et de la prise en compte des incertitudes sur les entrées sur les résultats ?

Retour au capteur développé à UGE



Figure – Deux capteurs installés dans une conduite d'eau potable et connectés via un câble USB à des ordinateurs de contrôle.

Method	MAE ₁ (HClO)	MAE ₂ (pH)
SLR	0.056	1.36
SLR+ME	0.054	1.75
GLR+ME	0.064	0.872
GLR+ME+IU	0.068	1.057
GPR+ME	0.054	1.75
GPR+ME+IU	0.039	0.254

Table – Comparaison des valeurs MAE obtenues pour différentes méthodes de calibration (approche LOO).



Plan de la présentation

- 1 Introduction
- 2 Formalisation du déploiement d'un capteur innovant
- 3 Difficulté 1 : une manque de sélectivité
- 4 Difficulté 2 : des entrées incertaines
- 5 Difficulté 3 : la prise en compte du temps
- 6 Conclusions

Sensibilité temporelle des capteurs

- Jusqu'à présent, nous nous sommes limités au cas où

$$y_t = \mathcal{M}(z_t, x_t, u_t), \quad t \geq 0.$$

- la réponse capteur à l'instant t ne dépend **que** des grandeurs environnementales **au même instant** t ,
- **stabilité temporelle** de la loi de calibration.

- Une formulation plus générale (et plus réaliste) serait :

$$y_t = \mathcal{M}_t(\{z_\tau, x_\tau, u_\tau, t - t_{\text{rep}} \leq \tau \leq t\}), \quad t \geq 0.$$

- potentielle **dérive temporelle** : la loi de calibration dépend *a priori* de t (dégradation progressive,...).
- potentiel **temps de réponse** non négligeable : la sortie capteur dépend de l'évolution des variables environnementales sur la durée t_{rep} .

Corrélation des quantités environnementales

- Jusqu'à présent, nous nous sommes limités à l'estimation de la loi de :

$$x_{\star} | y_{\star}, z_{\star}, \mathcal{D}_n.$$

- l'estimation de $x_{\star}$ à l'instant $\star$ ne dépend **que** des mesures disponibles (capteur+variables environnementales) **au même instant** $\star$.
- Pour une estimation plus générale (et sûrement plus précise), on pourrait chercher à estimer :

$$x_{\star} | (y_{\star-k}, z_{\star-k}), \dots, (y_{\star}, z_{\star}), \dots, (y_{\star+k'}, z_{\star+k'}), \mathcal{D}_n.$$

- on utilise pour reconstituer $x_{\star}$, en plus des mesures à $\star$, les k mesures **passées** (et potentiellement les k' mesures **futures**),
- approche "processus" pour exploiter de potentielles **corrélations temporelles** entre variables évoluant dans le temps.

Corrélation des quantités environnementales

- Jusqu'à présent, nous nous sommes limités à l'estimation de la loi de :

$$x_{\star} | y_{\star}, z_{\star}, \mathcal{D}_n.$$

- l'estimation de $x_{\star}$ à l'instant $\star$ ne dépend **que** des mesures disponibles (capteur+variables environnementales) **au même instant** $\star$.
- Pour une estimation plus générale (et sûrement plus précise), on pourrait chercher à estimer :

$$x_{\star} | (y_{\star-k}, z_{\star-k}), \dots, (y_{\star}, z_{\star}), \dots, (y_{\star+k'}, z_{\star+k'}), \mathcal{D}_n.$$

- on utilise pour reconstituer $x_{\star}$, en plus des mesures à $\star$, les k mesures **passées** (et potentiellement les k' mesures **futures**),
 - approche "processus" pour exploiter de potentielles **corrélations temporelles** entre variables évoluant dans le temps.
- ⇒ Présentation de Mouad vendredi pour de premiers travaux sur ce sujet.



Plan de la présentation

- 1 Introduction
- 2 Formalisation du déploiement d'un capteur innovant
- 3 Difficulté 1 : une manque de sélectivité
- 4 Difficulté 2 : des entrées incertaines
- 5 Difficulté 3 : la prise en compte du temps
- 6 Conclusions



Conclusions

- De grandes attentes pour des capteurs bon-marché, précis et fiables, susceptibles de pouvoir être déployés finement sur le territoire.
- Des difficultés de (trop forte) sensibilité et de (trop faible) sélectivité.
- Une difficile intégration des dépendances temporelles.
- Encore beaucoup de questions par rapport au nombre de réponses...



Merci pour votre attention.