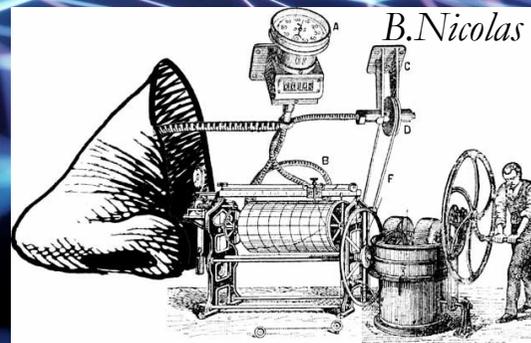


À LA CROISÉE DE L'IA ET DE L'ANALYSE MULTIVARIÉE POUR L'IDENTIFICATION D'ATMOSPHÈRES COMPLEXES PAR DES RÉSEAUX DE CAPTEURS

PROF. A-C ROMAIN

RESPONSABLE LABO SAM (SENSING OF ATMOSPHERES AND MONITORING)



OUTLINE

IA dans ma recherche

Imiter le sens de l'odorat

PARC et Machine Learning

Illustrations pour des atmosphères complexes

IA DANS MA RECHERCHE?

- Euh...non, pourquoi?



IA ?

(<https://www.coursera.org/articles/what-is-artificial-intelligence>)

“Artificial intelligence is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as **recognizing** speech, **making decisions**, and **identifying patterns**”.

- **umbrella term** that encompasses a wide variety of technologies, including **machine learning**, deep learning, and natural language processing (NLP).
- suite of machine learning-powered technologies, such as Chat GPT or computer vision, that enable machines to perform tasks that previously **only humans can do** like generating written content, steering a car, or **analyzing data**.

IA DANS MA RECHERCHE?

- Euh...non, pourquoi?

mais finalement quand est née l'IA (Chat GPT)?

- À l'Antiquité, références à des automates et à des entités artificielles intelligentes
- Domaine de recherche scientifique qui a émergé au milieu du 20e siècle.
- Le terme "intelligence artificielle" : à la Conférence de Dartmouth en 1956 (J, McCarthy, M, Minsky, N, Rochester et C, Shannon)
- l'IA a connu des périodes de succès et de ralentissement, souvent liées à l'évolution de **la puissance informatique disponible et aux avancées dans les algorithmes.**
- renouveau de l'intérêt pour l'IA, en particulier avec le développement de techniques d'apprentissage profond (deep learning) qui ont conduit à des avancées significatives dans des domaines tels que la reconnaissance d'image, la traduction automatique, et d'autres applications.

IA DANS MA RECHERCHE?

- Euh...non, pourquoi?

(ChatGPT)

...l'IA utilise souvent des **techniques d'apprentissage supervisé** pour apprendre à effectuer des tâches spécifiques. Dans le cadre de l'analyse supervisée, le modèle apprend à associer des entrées données à des sorties connues, en généralisant à partir des exemples fournis lors de l'entraînement.

...l'analyse supervisée est une méthode d'apprentissage machine utilisée dans le domaine de l'IA, où les modèles sont formés à partir de données étiquetées pour effectuer des tâches spécifiques.

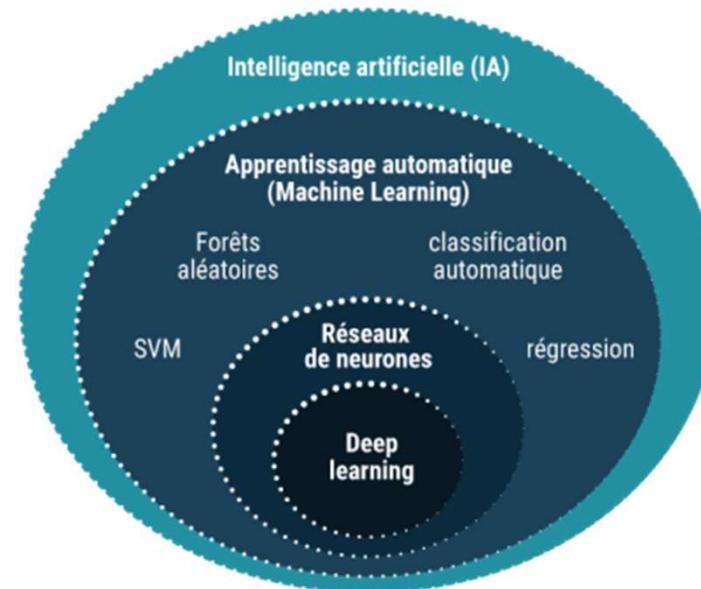
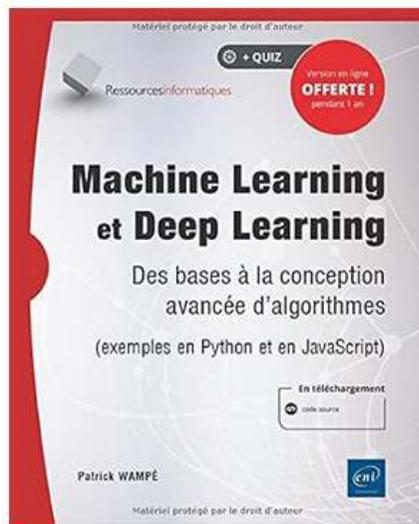
mais l'analyse supervisée fait bien partie de ma recherche...

Enfinement peut-être que oui...



IA DANS MA RECHERCHE?

- Les principes ne sont pas nouveaux
- Le vocabulaire change
- La puissance informatique permet le traitement d'un nombre de données quasiment infini

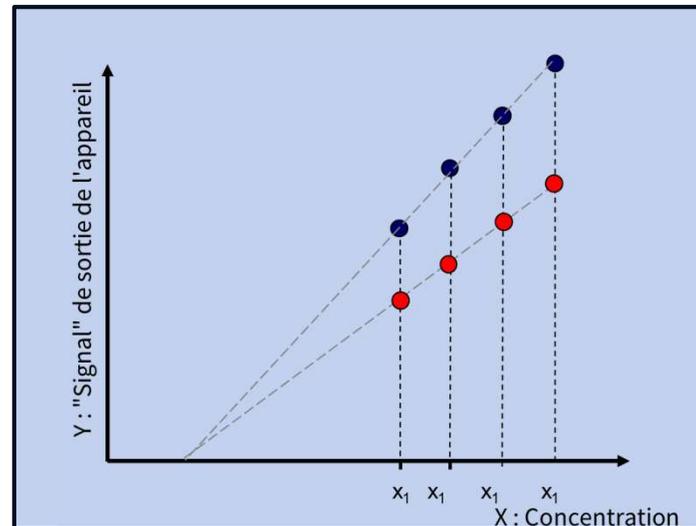
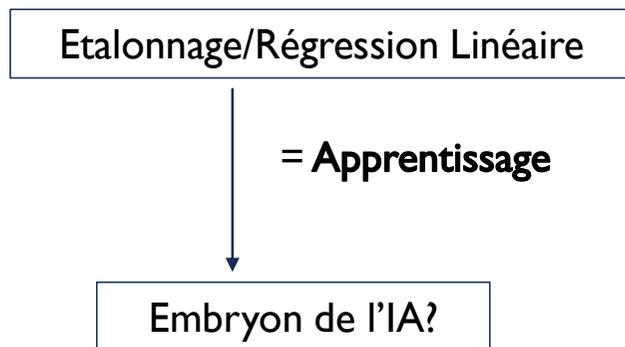


IA DANS MA RECHERCHE?

- ... cela reste une question philosophique et de définition qui demande à être débattue...



Tout traitement de données avec « apprentissage » devrait-il être considéré comme un élément de l'IA?



IMITER LE SENS DE L'ODORAT ?

“machines to perform tasks that previously only humans can do”

« En 1982, Persaud and Dodd, de l'université de Warwick, introduisent le concept d'**olfaction artificielle**. Ils présentent cette nouvelle technologie comme **un système intelligent basé sur un réseau de capteurs chimiques pour la classification des odeurs**. Mais le pionnier est sans doute, Moncrieff, qui en 1961 propose un instrument destiné à la détection des odeurs, basé sur un thermistor, recouvert d'un film adsorbant, intégré dans un pont de Wheatstone.

C'est seulement en 1991 que le terme « **nez électronique** » est accepté et qu'une première conférence lui est dédiée [NATO Advanced research workshop on sensors and sensory systems].

Gardner en 1994 propose une définition de ce nez :

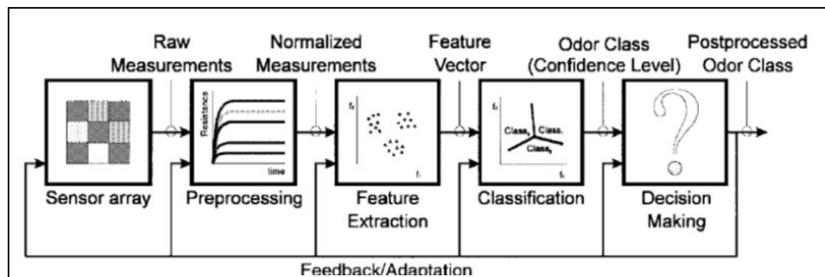
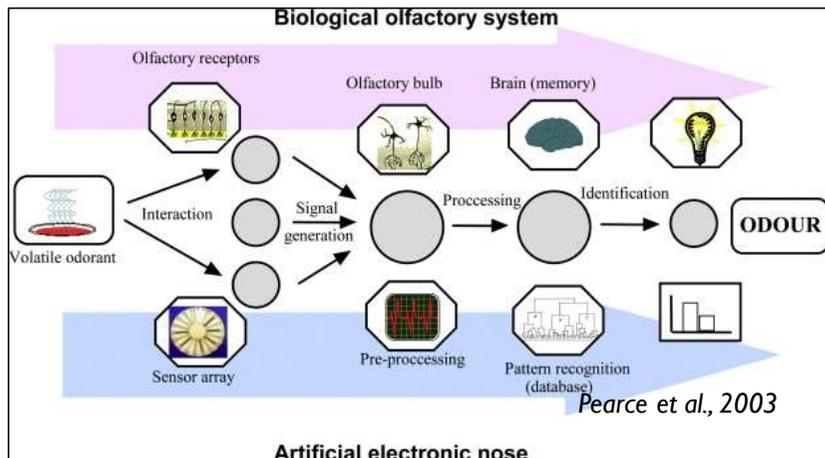
*instrument qui comprend un réseau de capteurs chimiques à spécificité partielle et
un système de reconnaissance de formes,
capable de reconnaître des odeurs simples ou complexes.*

...

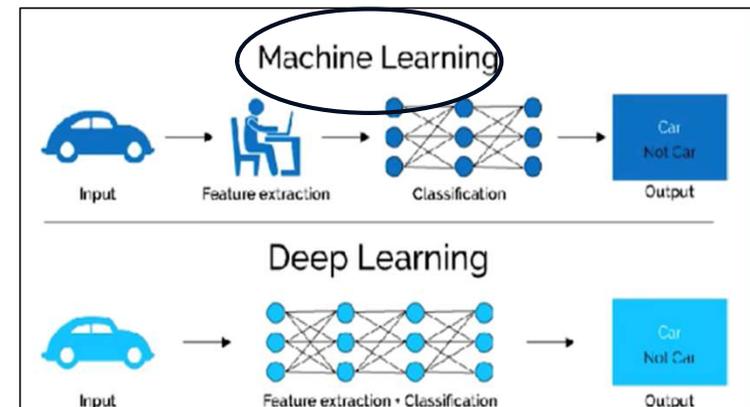
Plagiat de ma Thèse, 2006

8

IMITER LE SENS DE L'ODORAT ?



“machines to perform tasks that previously **only humans can do**”



<https://culturesciencesphysique.ens-lyon.fr/ressource/IA-apprentissage-Rousseau.xml>

Gutierrez-Osuna R (2002) Pattern analysis for machine olfaction: A review. *IEEE Sensors Journal* 2 (3):189-202. doi:10.1109/jsen.2002.800688

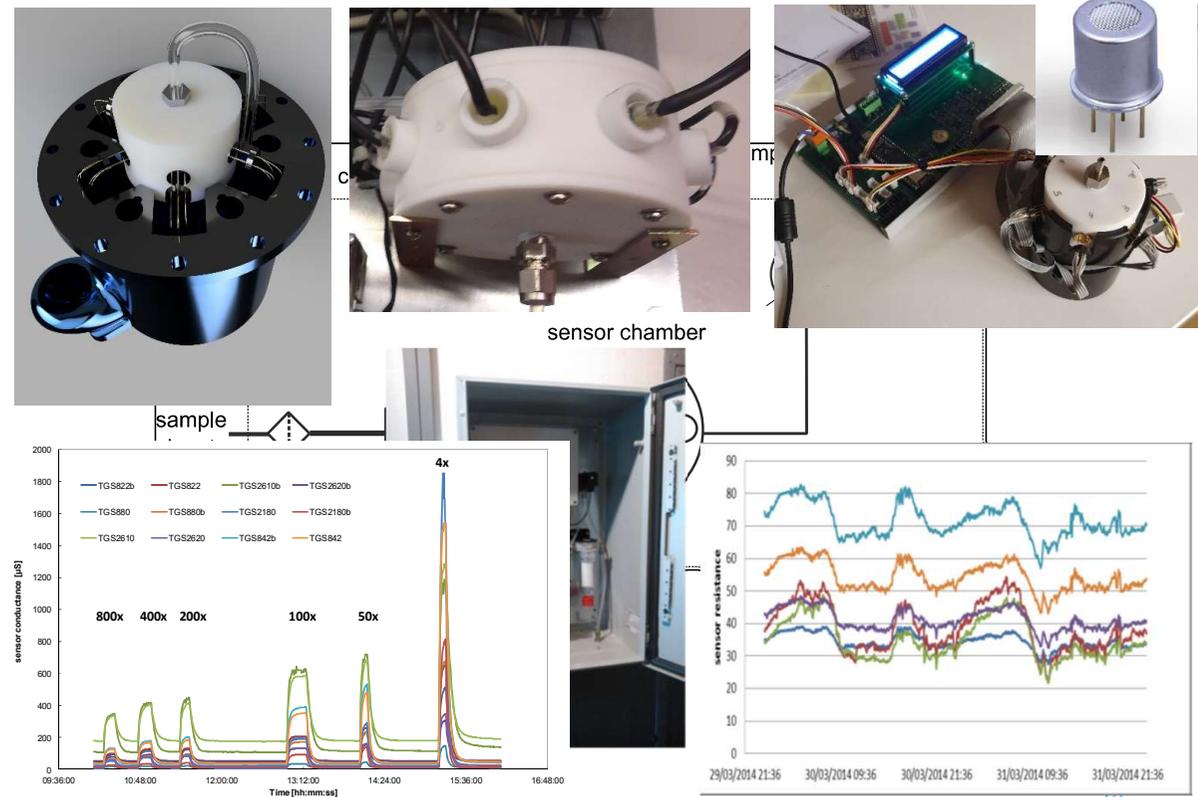
IMITER LE SENS DE L'ODORAT ?



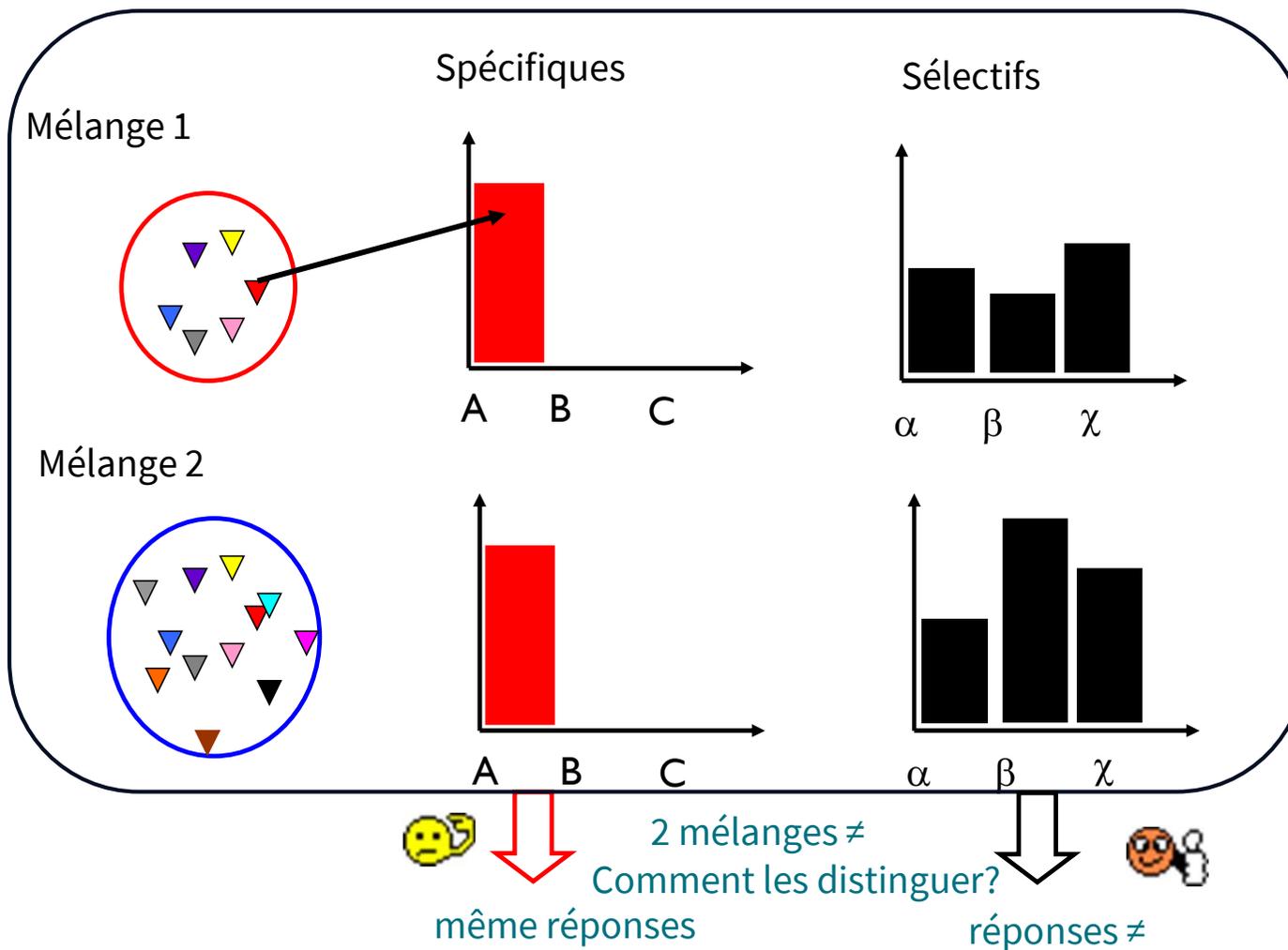
Nez électronique et PARC

Technologie

- ✓ système de préparation d'**échantillon** et d'introduction du flux gazeux (pneumatique -pompe, débitmètres, tuyaux,...-filtres, unité de concentration, générateur d'espace de tête, échantillonneur automatique,) ;
- ✓ **chambre(s)** des **capteurs chimiques** sensibles à la phase gazeuse (contenant le réseau de capteurs, généralement entre 4 et maximum 32 capteurs par réseau) ;
- ✓ **électronique** (alimentation, amplification, déclenchement de vannes, conversion A/D...) ;
- ✓ **système central** de pilotage, d'acquisition et de traitement des mesures (processeur, mémoire, logiciels de traitement des mesures, affichage).

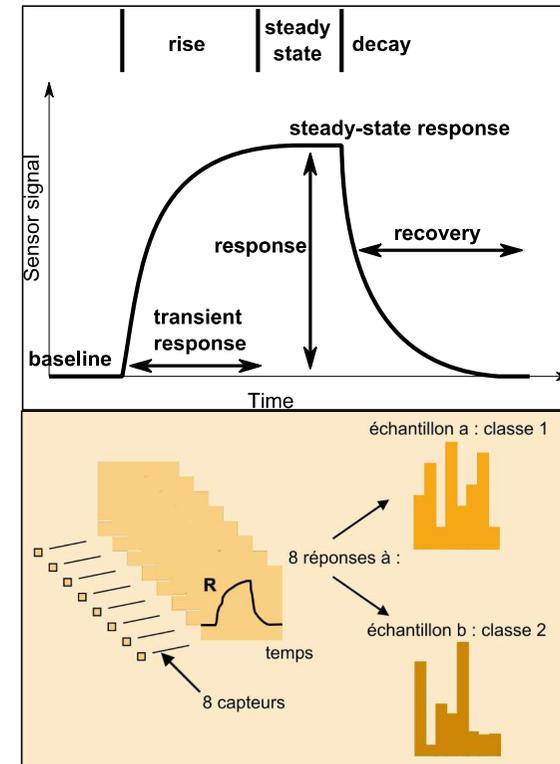
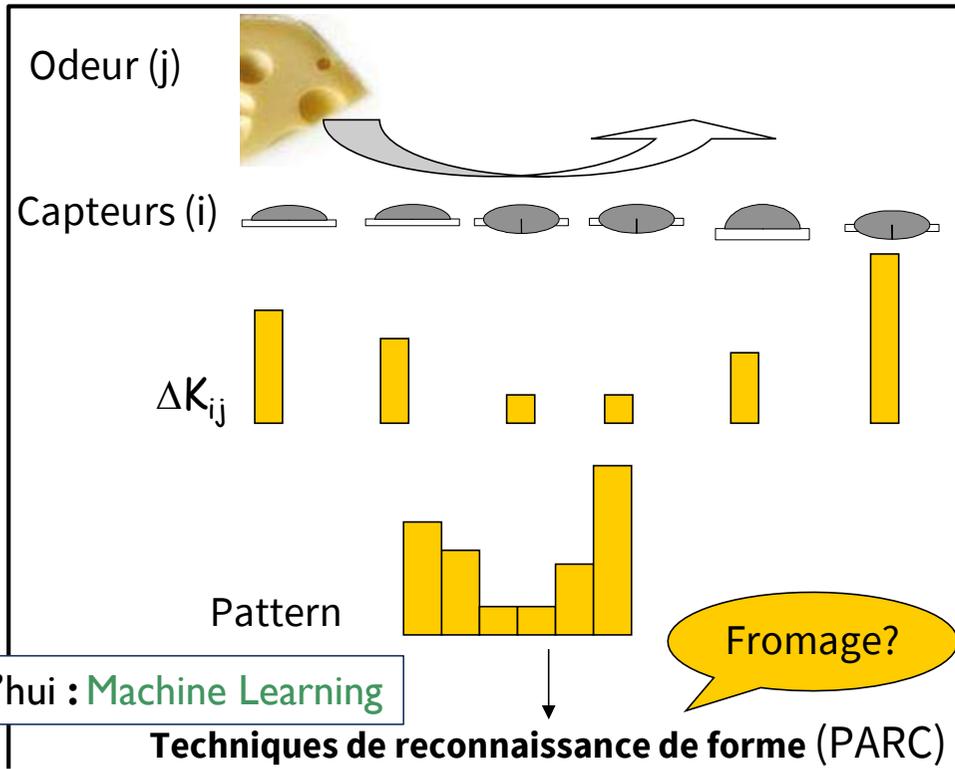


PARC and Cross-Selectivity



IMITER LE SENS DE L'ODORAT ?

Nez électronique et PARC



PARC ET MACHINE LEARNING?

Techniques de reconnaissance de forme (PARC)

Différentes approches - méthodes supervisées, non supervisées, identifier des modèles dans les données

ACP, Réseaux de neurones, LSTM, SVM, K-NN, Random Forest (decision tree), LDA, Classifieur Bayésien,

- ✓ multiples
- ✓ statistiques ou mathématiques (paramétriques ou non paramétriques)
- ✓ **supervisées** ou non supervisées

Analyse supervisée : Sous-catégorie du Machine Learning

« technique spécifique de mise en œuvre de la reconnaissance de forme dans le contexte du Machine Learning »

données d'apprentissage : **entrées avec des sorties connues** (classe, label)

pour créer un modèle et ensuite généraliser à de nouvelles données

faire des prédictions ou des classifications basées sur les modèles appris

13

PARC ET MACHINE LEARNING?

Validation indispensable

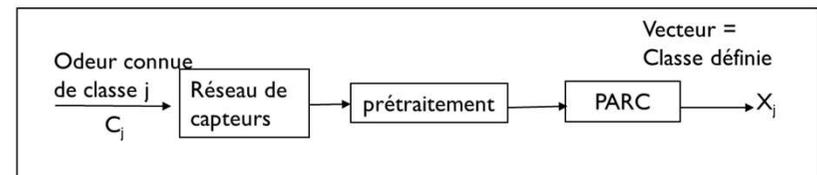
validation croisée, ...
augmenter le jeu de données

+ Test

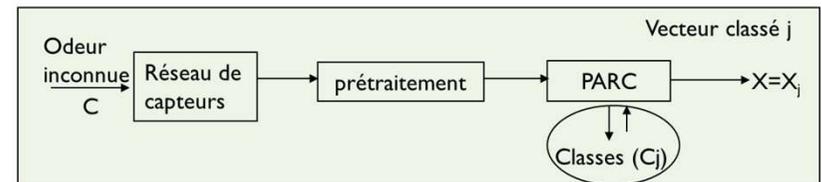
DANGERS DU TRAITEMENT DE DONNÉES

- Overfitting
- (underfitting)
- biais
- Dimensionnalité : excès/lacune de données
- Pas de signification physique des modèles

1. Apprentissage (créer le modèle)



2. Reconnaissance (utilisation du modèle)



mais surtout : **Qualité des données (d'entrée) !!!**

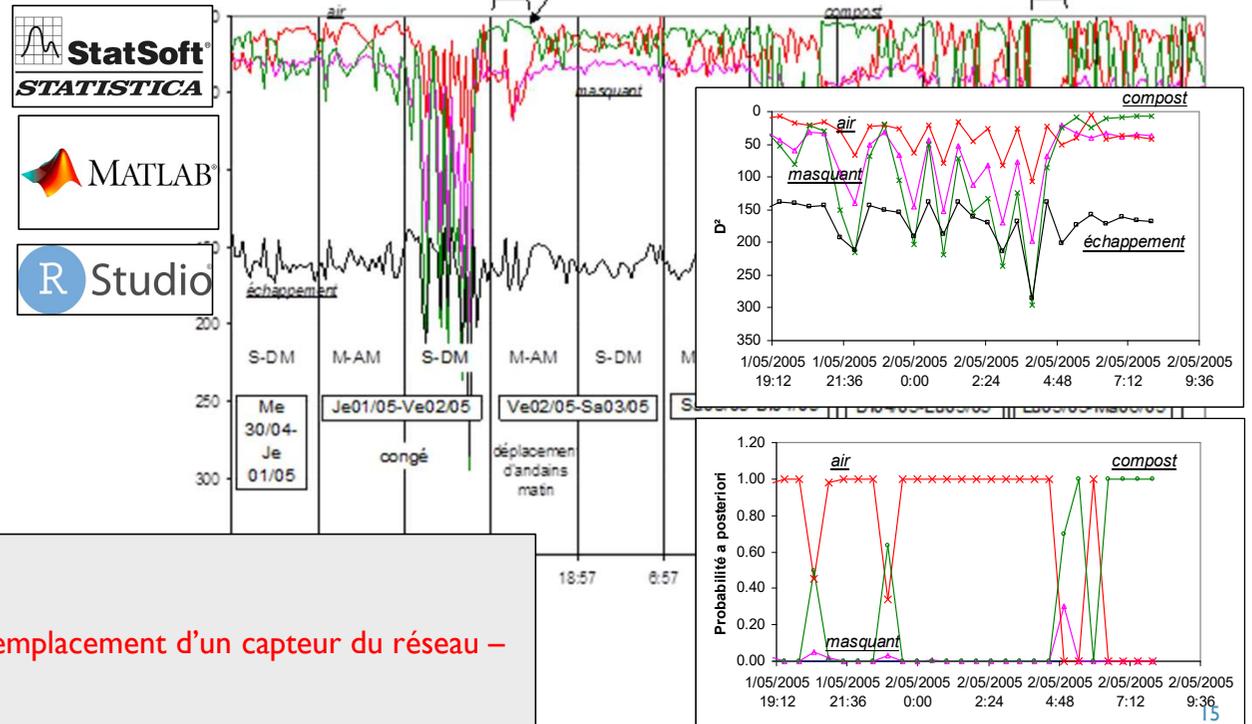
ILLUSTRATION POUR DES MELANGES COMPLEXES

Surveillance des odeurs : réseaux de capteurs à oxydes métalliques (MOS)

1995-

LDA, distance de Mahalanobis, ANN

Variables	lambda Wilks	F	% de classification correcte (fonctions de classification)			
			station d'épuration	carrosserie	imprimerie	centre de compostage
$G / \sqrt{\sum G_i^2}$	0.0003	25	100	100	100	100
G	0.0004	24	100	100	100	100
G/G_0	0.0004	22.7	90	83.3	100	100
$(G - G_0)/G_0$	0.0004	22.7	90	83.3	100	100
Intégrale sur 15 min $(G - G_0)/G_0$	0.001	18.7	90	83.3	100	100
$\sqrt{\sum_n ((G - G_0)/G_0)^2}$	0.001	15.1	90	83.3	100	100
$G_{15 \text{ min}}$	0.002	13.6	80	100	100	80
Pente de recouvrement (ligne de régression sur les trois premiers points)	0.015	6.2	90	100	75	90
Pente de réponse (ligne de régression sur les trois premiers points)	0.02	6	90	83.3	100	90



Super résultats mais pas valorisés à long terme...

Élément limitant = les capteurs

- faible reproductibilité capteurs – nouvel apprentissage si remplacement d'un capteur du réseau –
- baisse de la sensibilité
- mauvaise qualité de la donnée

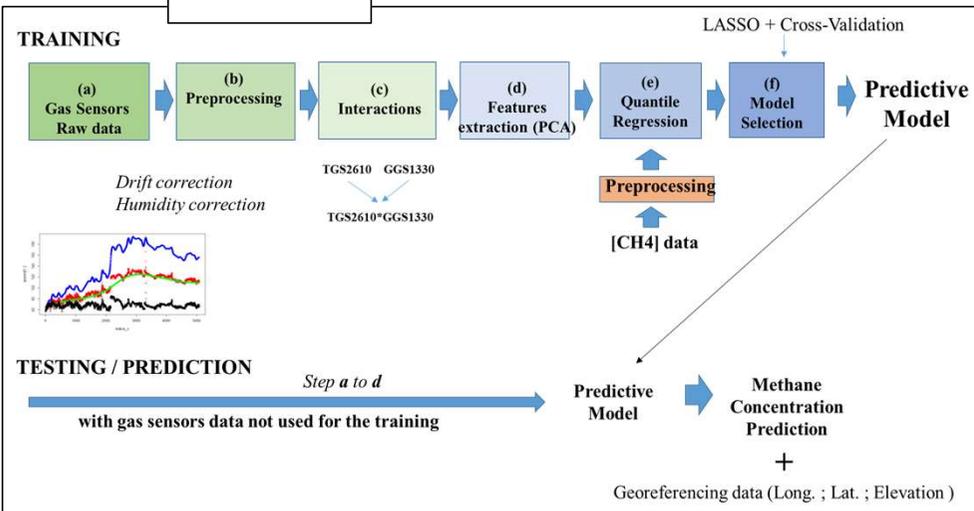
ILLUSTRATION POUR DES MELANGES COMPLEXES

Drone = détection du méthane dans atmo complexe

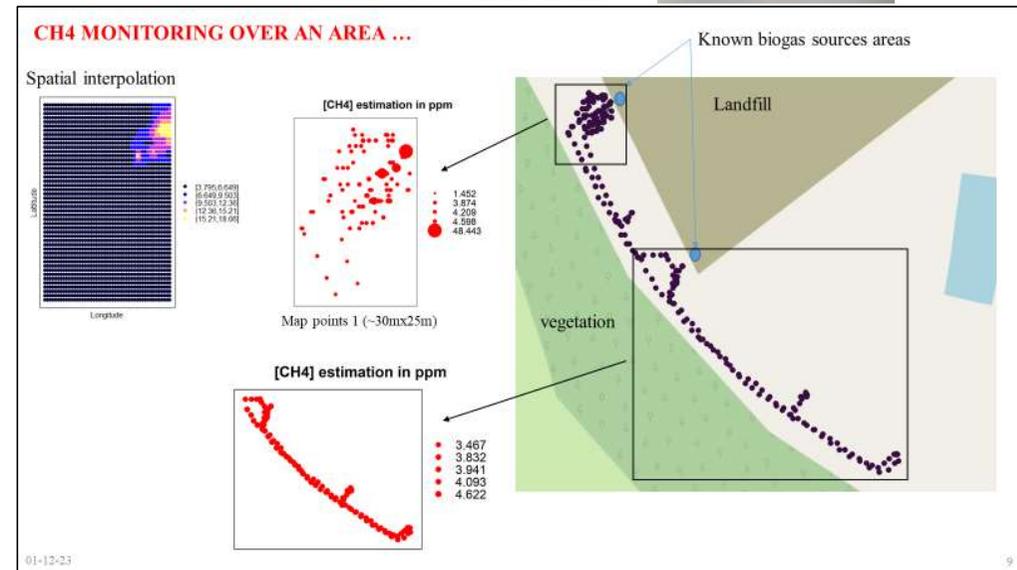
2017



Régression quantile multivariée



Détection d'événements biogaz ok, suffisant pour identifier présence de fuite de biogaz mais erreur de quantification importante
 Élément limitant = capteurs (LOD, sélectivité)
 → mauvaise qualité de la donnée



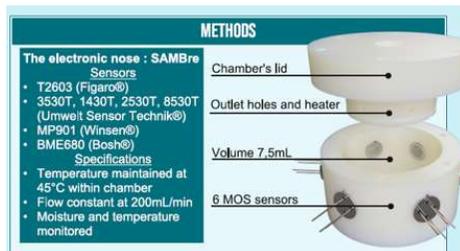
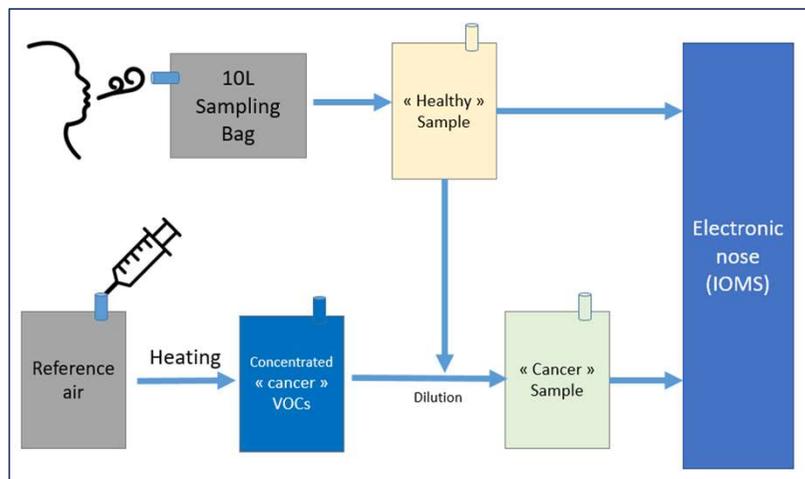
Projet GRONE, Eric Taguem

16

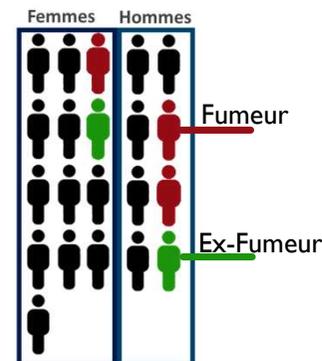
ILLUSTRATION POUR DES MELANGES COMPLEXES

Analyse d'haleines humaines : réseaux MOS

2019



« malades » : haleines saines + biomarqueurs cancer poumon



- 21 volontaires
- 26 jours d'échantillonnage
- 127 échantillons d'haleine
- 236 mesures
 - 117 « sains »
 - 119 « malades »

Réseau de 6 capteurs

ACP – conductances normalisées et correction dérive

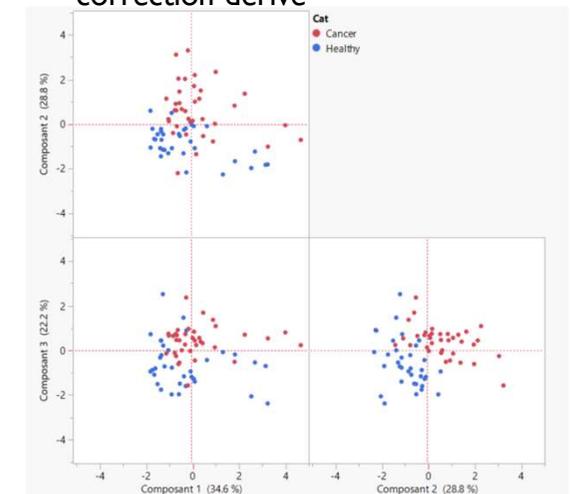


ILLUSTRATION POUR DES MELANGES COMPLEXES

Analyse d'haleine humaine : réseaux MOS



- 176 échantillons pour l'apprentissage (74%),
- 30 pour la validation (13%),
- 30 comme groupe test (13%).



Validation :
87% /cancer
70% /sain
Test :
75% /cancer
74 % classe sain

K-NN

même jeu de données

K = 3 et 3 capteurs

Validation : 95%

Vérification : Données générées aléatoirement avec 2 classes,
résultat de classification 50%



(bibliothèques pour :

- machine learning : « *SKlearn -KNeighborsClassifier* »
- gestion de bases de données : « *Panda* »
- graphiques : « *seaborn* »

Thèse de Justin Martin

Résultats similaires malgré “technique” différentes
Manipulation plus facile du jeu de données avec Anaconda Jupyter (mais plus “boite noire”)

Éléments limitants

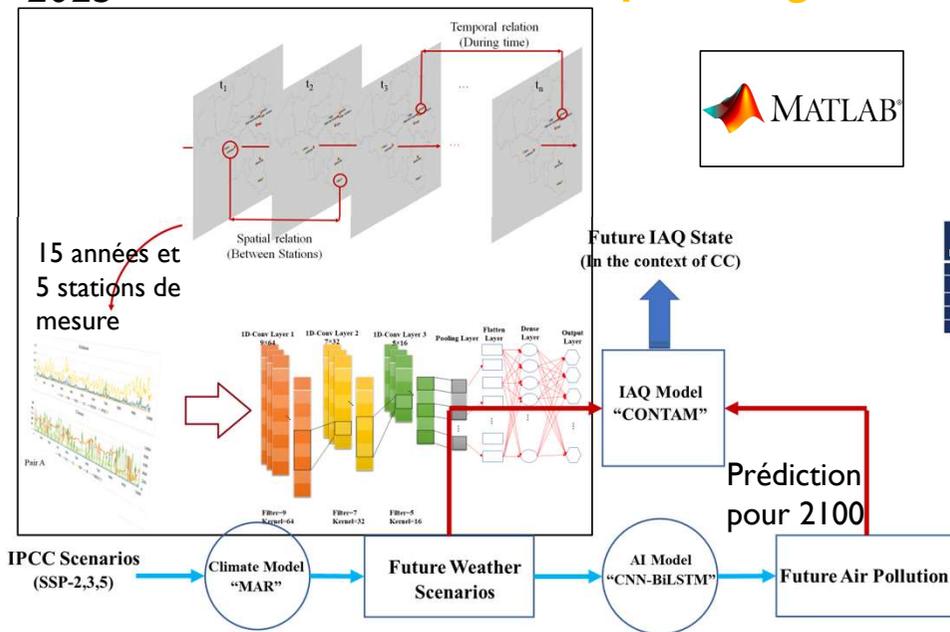
- capteurs (LOD, sélectivité, derive – bien que corrigée -)
- jeu de données insuffisant (à terme 1 donnée = 1 patient malade)
- Domaine médicale : Faux négatifs non autorisés!

ILLUSTRATION POUR DES MELANGES COMPLEXES

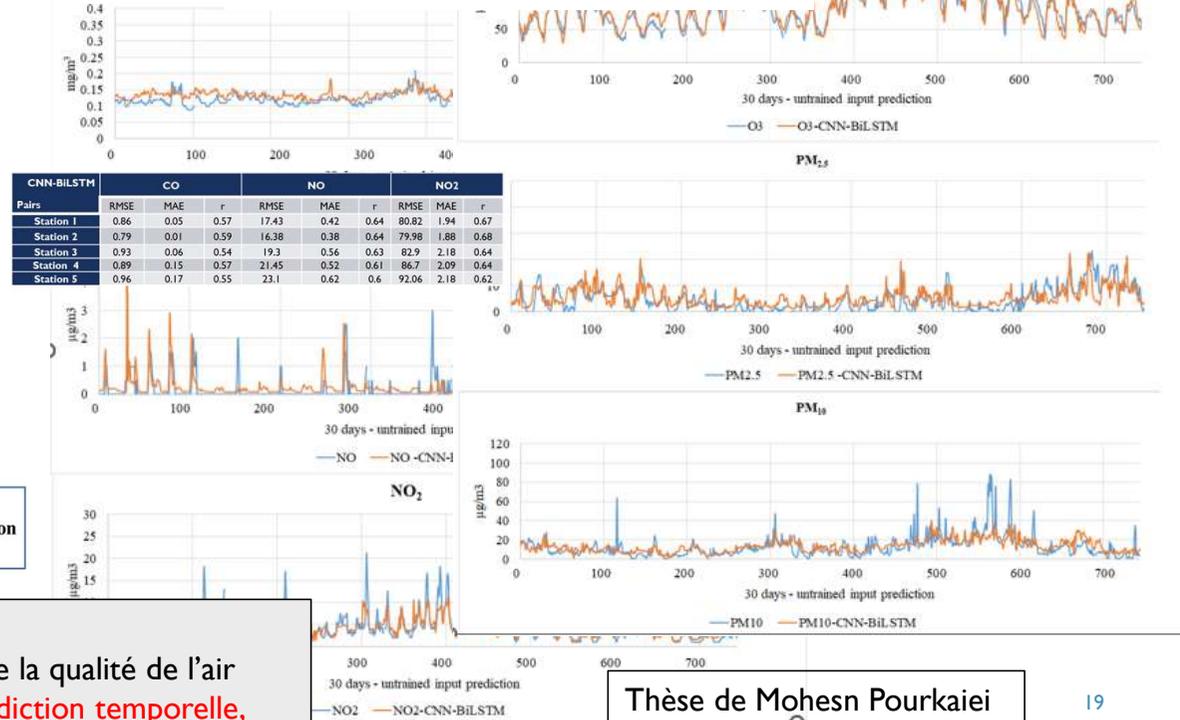
Prédiction de la qualité d'air intérieur dans un contexte de changement climatique

2023

vers le deep learning...CNN-BiLSTM*



performance for one month prediction
(with untrained input summer 2021)



Résultats moyens satisfaisants, faibles pour les pics de pollution
 A explorer dans le cas de grands jeux de données pour la gestion de la qualité de l'air
 Éléments limitants = absence de modèle physique (boîte noire), prédiction temporelle,
 Incertitude importante sur les données météo

Thèse de Mohesn Pourkaiei

TAKE HOME MESSAGES

- IA dans ma recherche?
- IA et Surveillance d'atmosphères complexes par capteurs low-cost ?

Intérêt certain mais obstacles... souvent indépendants de l'IA

✓ Qualité des données

limitations techniques notamment les performances des capteurs Low cost
variabilité importante de la donnée pour une même « classe »

✓ Validation indispensable

encore trop souvent volontairement oubliée

✓ Nombre de données ?

souvent trop faibles (comparé aux milliards d'infos vérifiées ou non sur le net par ChatGPT par ex)
ou
redondantes (augmentation de la dispersion du jeu de données mais pas info complémentaire utile)

20

RÉFÉRENCES

- [M. Turing \(1950\)](#) Computing Machinery and Intelligence. *Mind* 49: 433-460.
- [H. Shurmer I, A. Fard I, J. Barker I, P. Bartlett I, G. Dodd I and U. Hayat](#), Development of an electronic nose, *Physics in Technology* 1987 18 170 DOI 10.1088/0305-4624/18/4/104
- Moncrieff, R. W., An instrument for measuring and classifying odors, *J. Appl. Physiol.*, 16, (1961) 742-749
- Gardner, J. W., Bartlett, P. N., *Sensors and sensory systems for an electronic nose*, 212, Kluwer Academic, London, 1991.
- Gardner, J. W., A brief history of electronic noses, *Sens. actuators, B, Chem.*, 18-19 (1994) 211-220.
- Bott, B., Jones, T.A., The use of multisensor systems in monitoring hazardous atmospheres, (1986) *Sensors and Actuators*, 9 (1), pp. 19-25. DOI: 10.1016/0250-6874(86)80003-8
- Persaud, K.C., Jagers, H.C., Shaw, J.S., Peacock, S.J., Nicholas, P., Bartlett, P.N., Dodd, G.H., Gardner, J.W., Shurmer, H.V., Slater, J.M., Watt, E.J., Boden, G.F. Odour detection using sensor arrays (1991) *Analytical Proceedings*, 28 (10), pp. 339-344 DOI: 10.1039/AP9912800339
- Romain, A.-C., Nicolas, J., & Andre, P. (1997). In situ measurement of olfactive pollution with inorganic semiconductors: Limitations due to humidity and temperature influence. *Seminars in Food Analysis*, 2, 283-296.
- Romain, A.-C., & Nicolas, J. (2010). Long term stability of metal oxide-based gas sensors for e-nose environmental applications: an overview. *Sensors and Actuators. B, Chemical*, 146, 502-506.
- [Taguem Ngoualadjio, Eric Martial; Mennicken, Luisa; Romain, Anne-Claude](#), Quantile regression with a metal oxide sensors array for methane prediction over a municipal solid waste treatment plant 2021 • In *Sensors and Actuators. B, Chemical*, 334, p. 129590 <https://doi.org/10.1016/j.snb.2021.129590>
- [Martin, Justin; Romain, Anne-Claude](#), Building a Sensor Benchmark for E-Nose Based Lung Cancer Detection: Methodological Considerations, 2022 • In *Chemosensors* <https://doi.org/10.3390/chemosensors10110444>
- Justin D.M. Martina, Falzone Claudia, Anne-Claude Romain 2023 (under review) in *Journal of breath analysis*, In *How Well Does Your E-nose Detects Cancer? Application of Artificial Breath Analysis for Performance Assessment*
- Pourkiaei, Seyed Mohsen; Romain, Anne-Claude 2023 • (under review) In *Journal of Building Engineering A Systematic Framework for Quantitative Assessment of Indoor Air Quality Under Future Climate Scenarios*
- Tong, W. (2020). Machine learning for spatiotemporal big data in air pollution. In *Spatiotemporal Analysis of Air Pollution and Its Application in Public Health* 107-134. Elsevier.
- Masood, A., & Ahmad, K. (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, 322, 129072.

MERCI POUR VOTRE ATTENTION

QUESTIONS?

Merci à mon équipe

Claudia Falzone
Noémie Molitor
Laurent Collard
Justin Martin
Mohsen Pourkiaei
Simon-Pierre liégeois
Bui Thi Ngoc Phuong
Mauri Rosiers
Alexandra Delperdange
Jean-Sébastien Liégeois

(Précédemment dans l'équipe :
Gilles Adam, Eric Taguem et Marie Scheuren)

PROFESSEURE A-C ROMAIN

acromain@uliege.be