

# Caractérisation de la contamination des eaux de surface par le couplage d'analyses non-ciblées en spectrométrie de masse avec des analyses d'écotoxicologie

Présenté par SADE Julien - 11 juillet 2022

## LEESURIALES

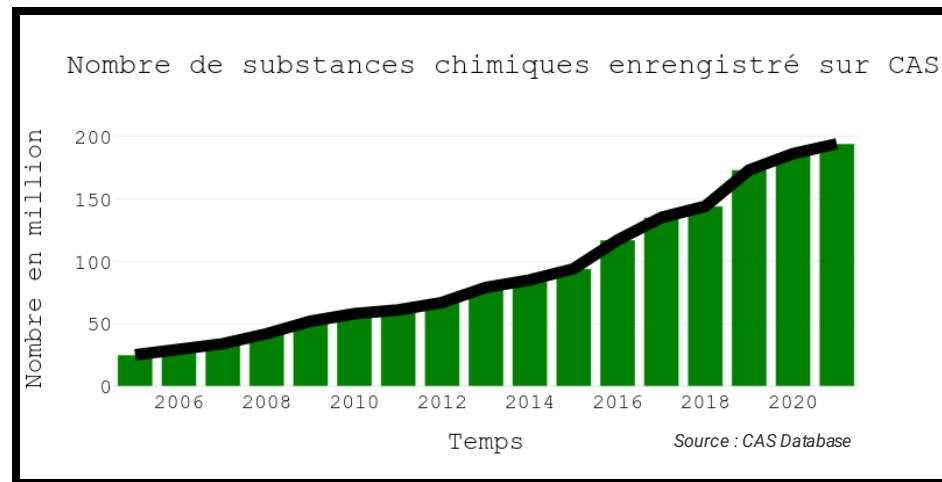
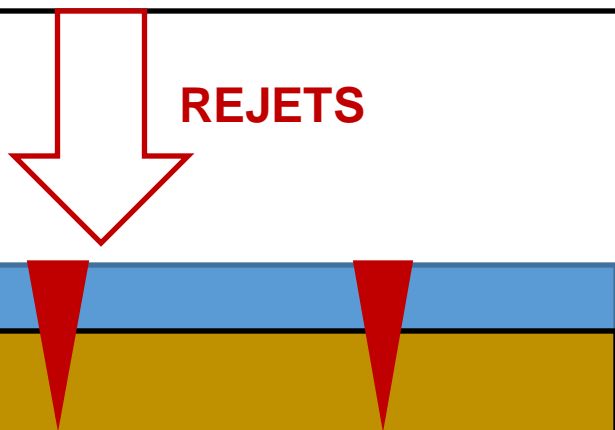
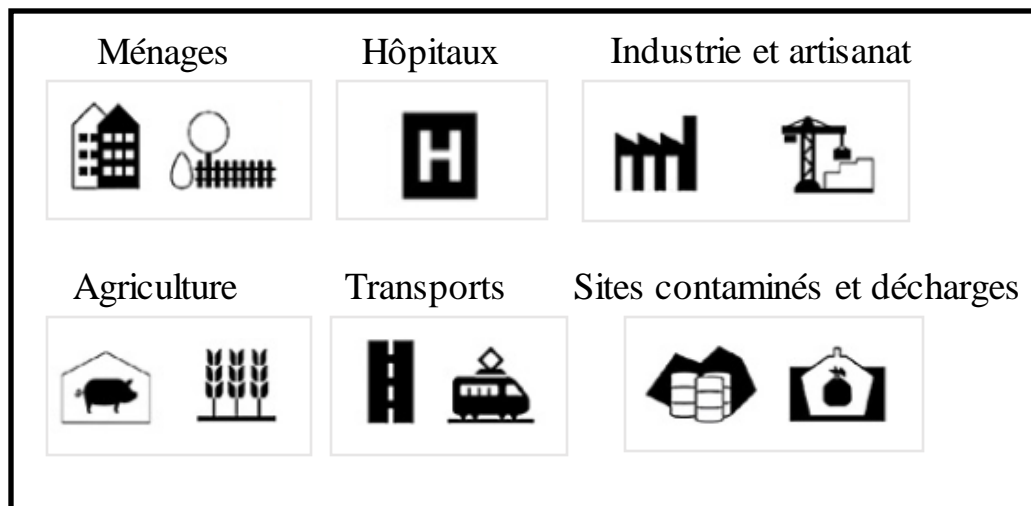


Encadrement : Régis MOILLERON, Julien LE ROUX, Sabrina GUERIN, Vincent ROCHER, Stéphane MOTTELET



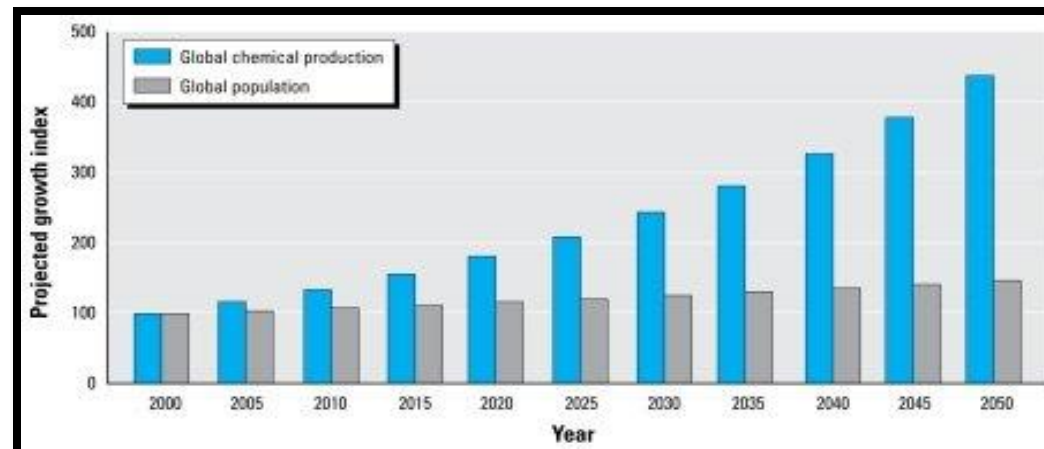
# Contexte: Contamination des eaux de surfaces

## Source des micropolluants



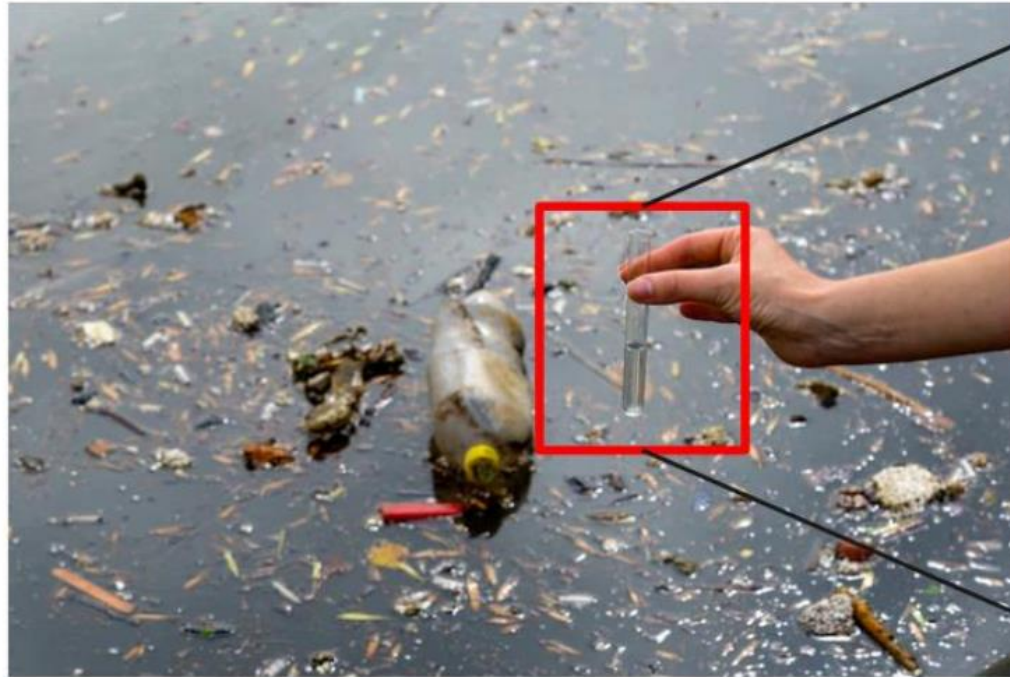
*“ Global chemical production is projected to grow at a rate of 3% per year, rapidly outpacing the rate of global population growth ”*

*Wilson & Schwarzman, Environ. Health Persp. (2009) 117:1202-1209*

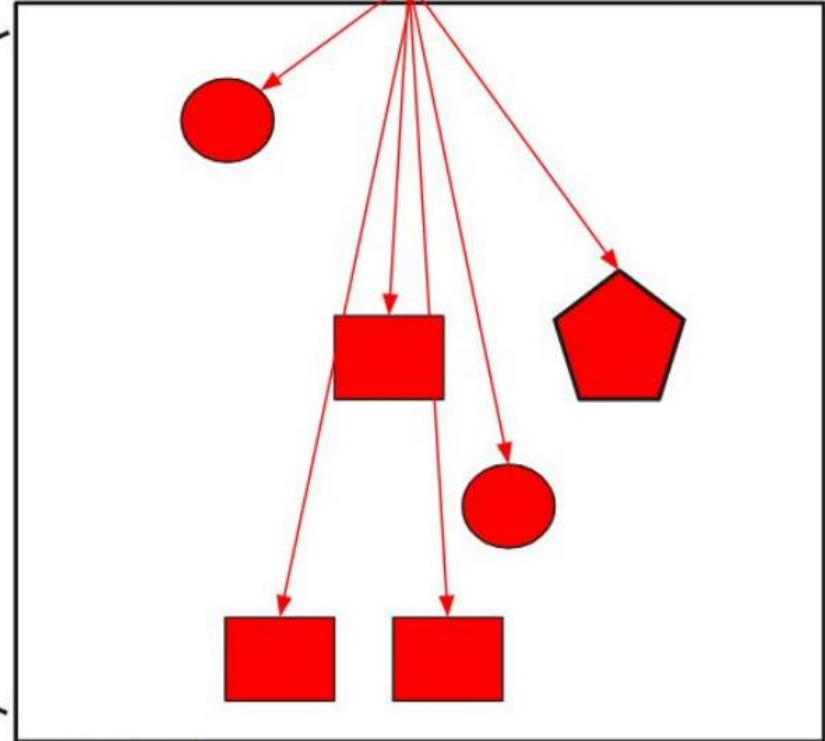


Nécessité de développer des méthodes robustes et exhaustives pour l'analyse de micropolluants

# L'eau, un mélange complexe de polluants

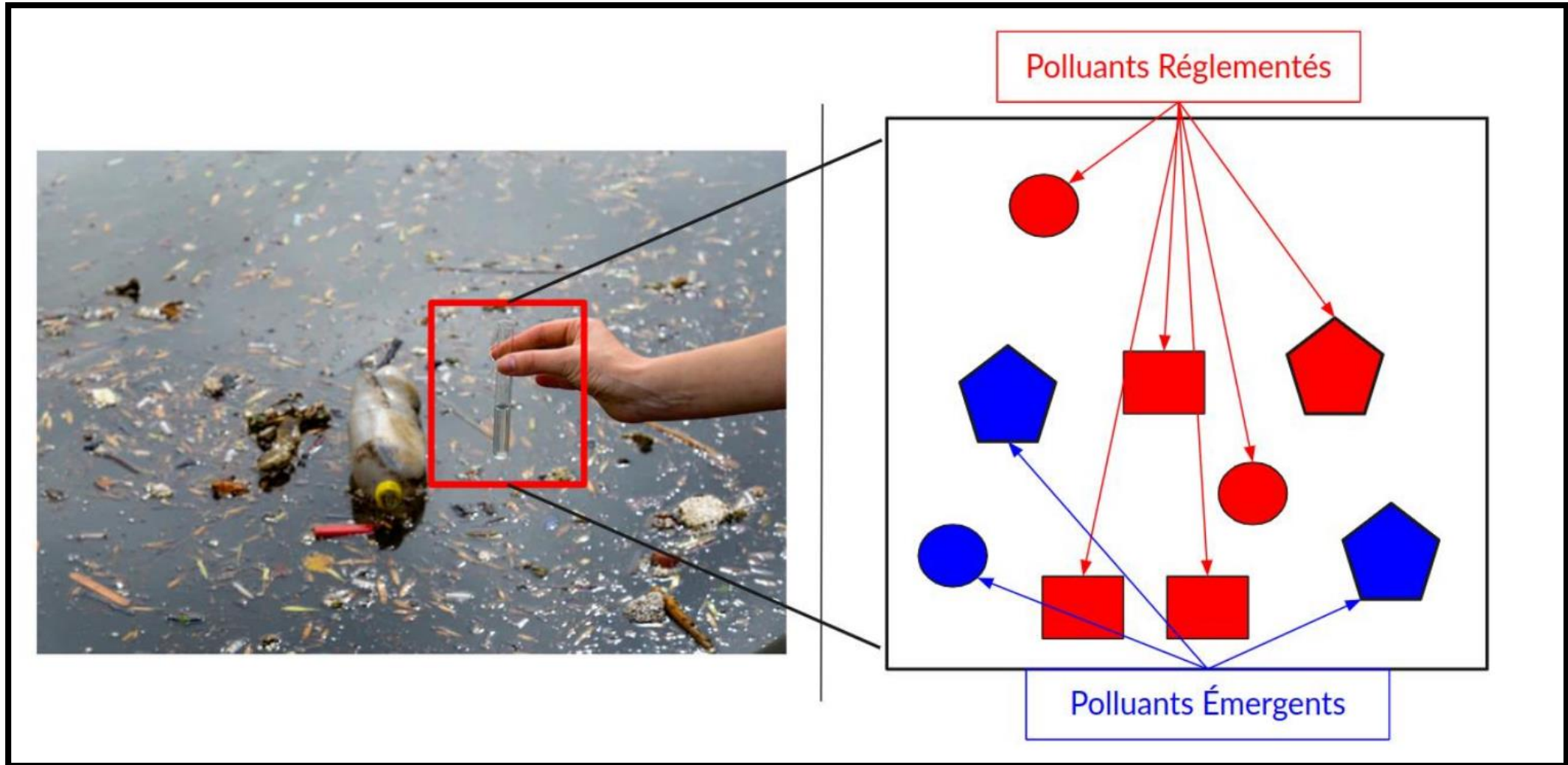


Polluants Réglementés



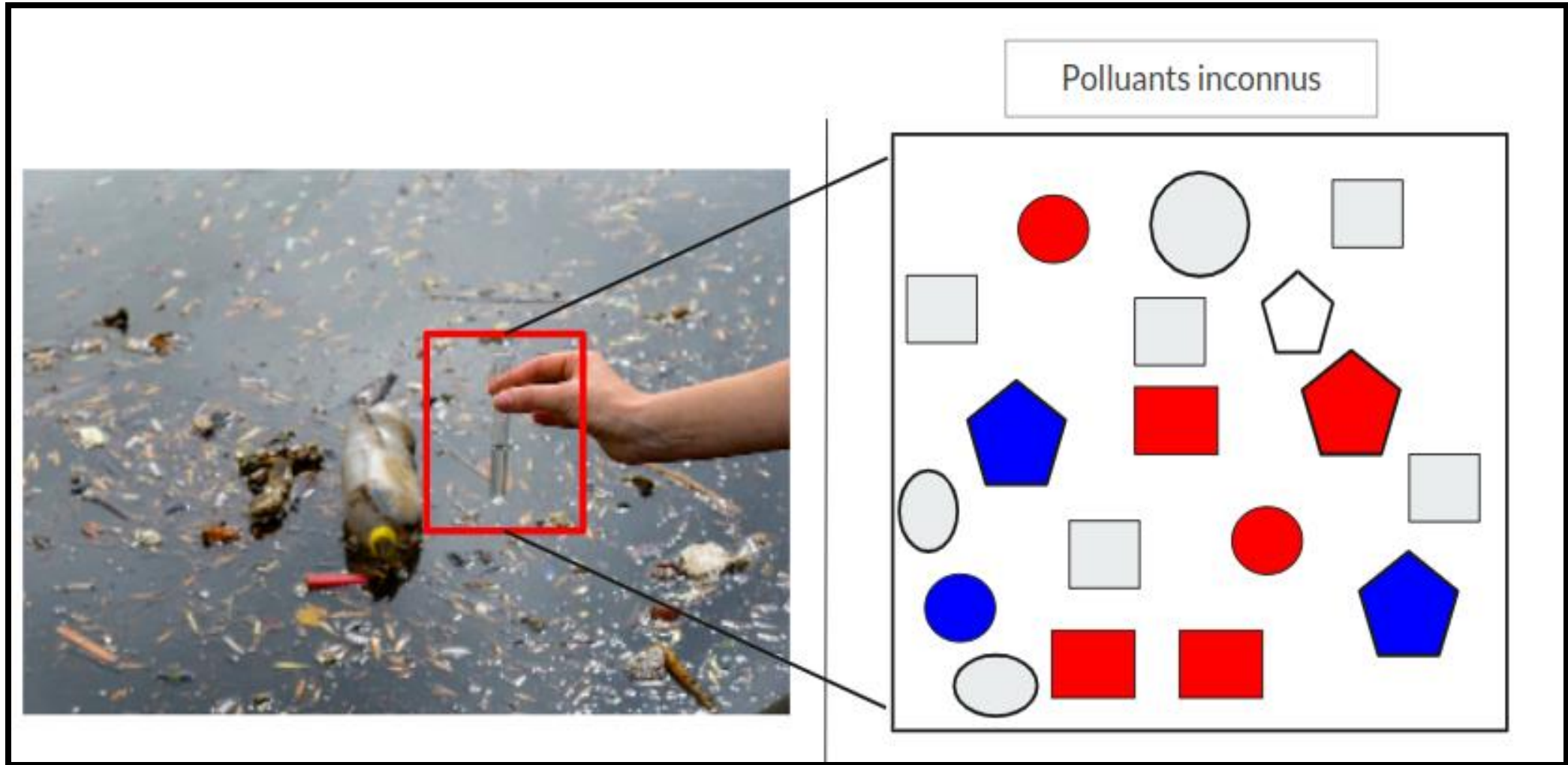
Echantillon d'eau

# L'eau, un mélange complexe de polluants

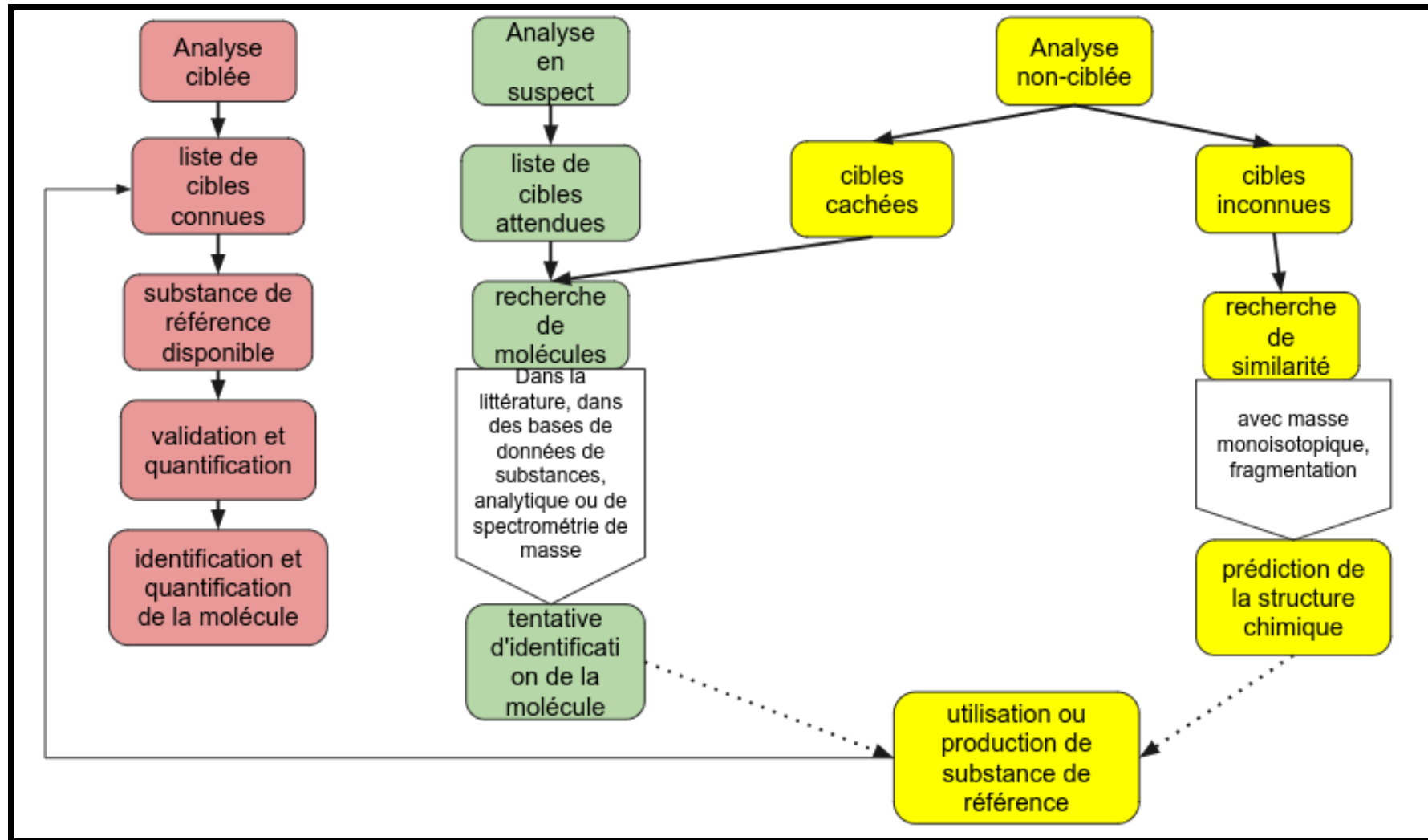




# L'eau, un mélange complexe de polluants



# Les différents types d'analyse d'identification de composé

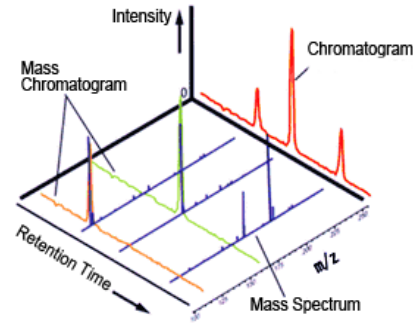


*" Due to the complexity of environmental samples and site-specific chemical distributions, targeted chemical analysis-driven monitoring strategies are not suitable for identifying key toxic chemicals and assessing risk." Brack et al 2003*

# Analyse non-ciblée par HRMS



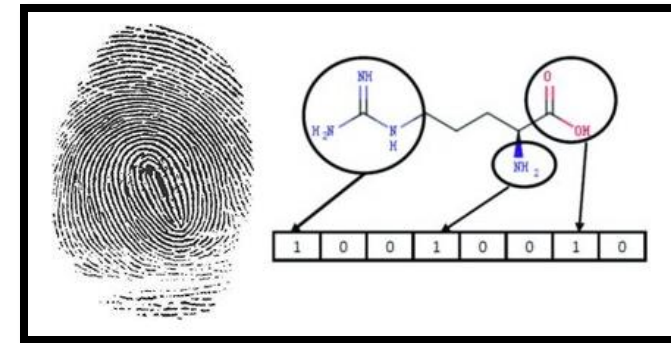
UPLC-IMS-QTOF (HRMS)



+ mobilité ionique

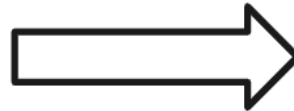


Empreinte digitale chimique

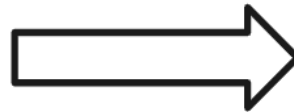


Traitement et analyse des données

Analyse non-ciblée



Identification de polluants émergents



Identification de nouveaux dangers



Comparer les empreintes chimiques/pattern pour différents échantillons :  
eau surface vs eau de ruissellement vs eau usée, etc..

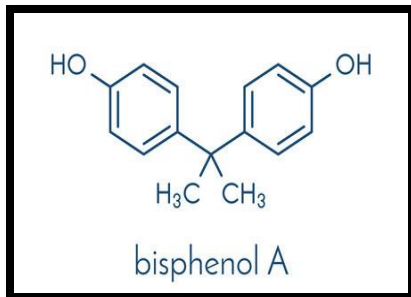
# La présence de polluants n'induit pas systématiquement une toxicité



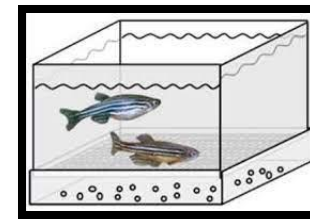
L'échantillon d'eau est un mélange complexe de composés dont des polluants

Limite analyse non-ciblée

- L'identification de polluant, même ceux considérés comme les plus dangereux n'induit pas systématiquement une toxicité au sein de l'échantillon

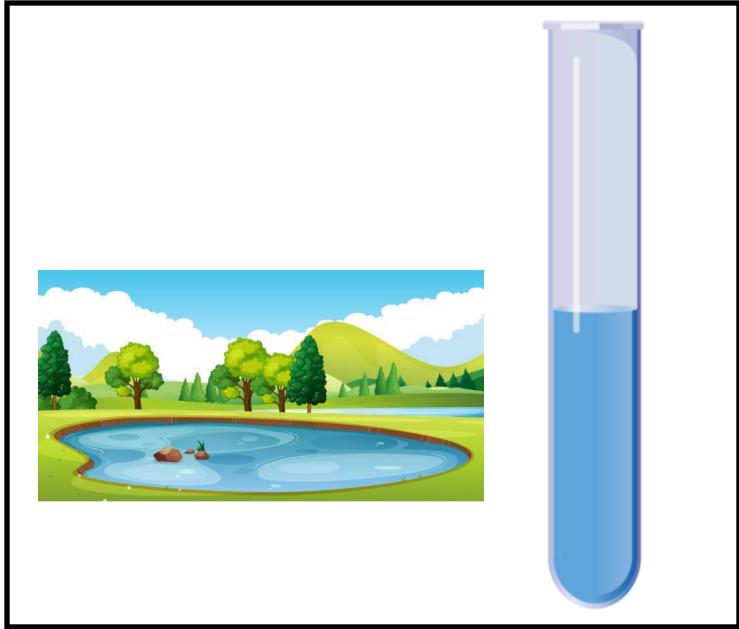


Couplage à des analyses écotoxicologiques afin de déterminer la toxicité de l'échantillon





# Association d'une toxicité à un échantillon d'eau



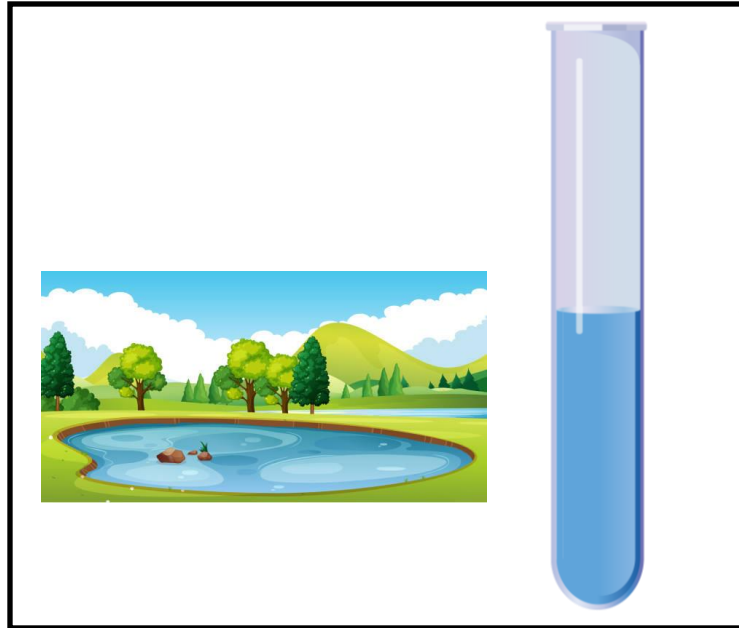
# Association d'une toxicité à un échantillon d'eau



**Recherche dans les bases de données/littérature** (site contaminé connu?)

Mais s'il n'y a pas d'information?

# Association d'une toxicité à un échantillon d'eau



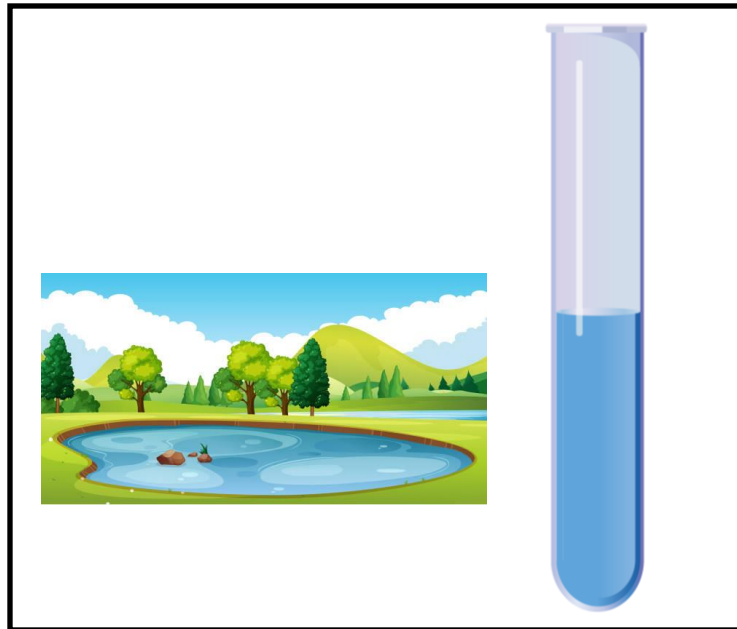
**Recherche dans les bases de données/littérature** (site contaminé connu?)

**Expérimentale** : Analyses éco-toxicologiques in vivo/ in vitro



**Inconvénients** : chronophage, très compliqué d'analyser une grande bibliothèque d'échantillons, couteux

# Association d'une toxicité à un spectre

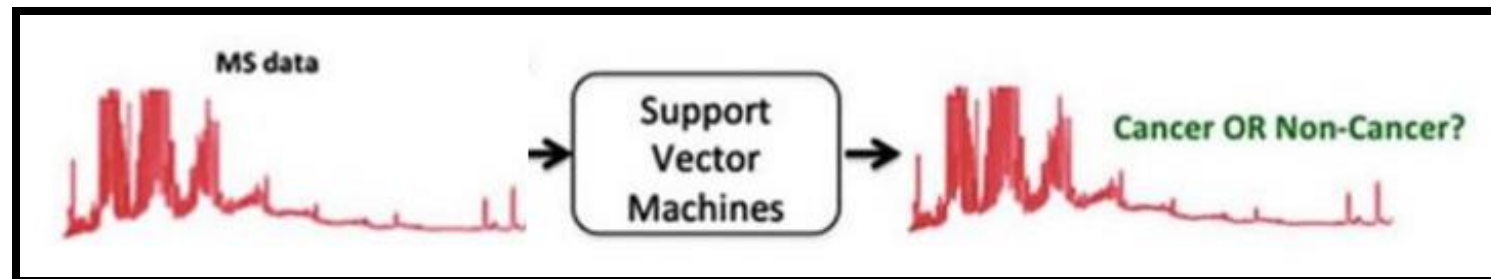


**Recherche dans les bases de données/littérature** (site contaminé connu?)

**Expérimentale** : Analyses éco-toxicologiques in vivo/ in vitro

**Prédictions:** Machine learning / Deep learning, prédiction d'une toxicité d'un échantillon grâce à son spectre de masse

- Méthode utilisée dans le domaine de la santé pour prédire une cellule cancéreuse



# Questionnement scientifique

- Comment tirer le **maximum d'informations** des données acquises en HRMS ?
  - Quel est l'avantage de la mobilité ionique dans l'identification de composés?
  - Comparaison de différentes méthodes de pré-traitement et de traitement des données
  - Développement de méthodes en bio-informatique pour le **traitement de données HRMS**
- Quel est la valeur ajoutée de l'**analyse non-ciblée** pour l'identification de micropolluants?
- L'**analyse non-ciblée** et les **analyses éco-toxicologiques** sont-elles complémentaires?
- Les **modèles** peuvent-ils être performants concernant la **prédiction de la toxicité** pour un spectre?



# Collecte des données

## Campagnes de prélèvements sur 7 sites



- Identification de micropolluants => **Analyse non-ciblée**
- Batterie de test éco-toxicologiques sur 15 modèles biologiques différents

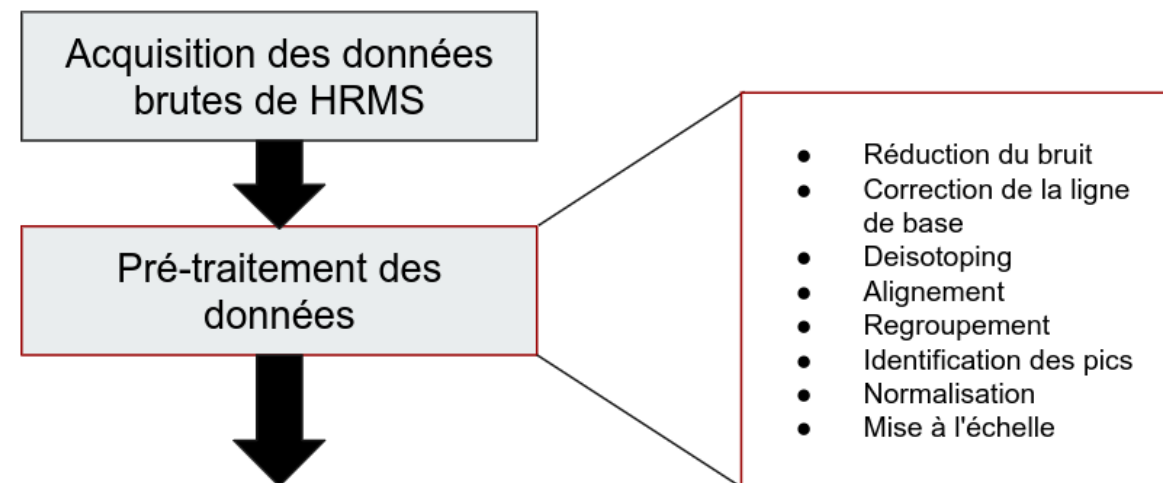
Famille Effets	In vitro/In vivo	Organisme/Cellules	Type de réponse
Perturbateurs endocriniens	In vivo	Oestrogènes (Zebra Fish)	Equivalents [Hormones]
		Androgènes (Zebra Fish)	
		Thyroïdiens (xenopes)	
	In vitro	Oestrogènes (cellules humaines MELN)	Equivalents [Hormones]
		Androgènes (cellules humaines PALM)	
		YES/YAS ( <i>S. cerevisiae</i> )	
Toxicité générale microorganismes	In vitro	Bactéries ( <i>E. coli</i> )	% de réponse + possibilité EC/IC50 sur substances isolées
		Algues ( <i>C. Reinhardt</i> , <i>P. subcapitata</i> , <i>C. vulgaris</i> , <i>S. capricornutum</i> )	
		Levures ( <i>S. cerevisiae</i> )	
		Fungi ( <i>S. tritici</i> )	
		Cellules humaines ("globules blancs" PBMC)	
Génotoxicité	In vitro	SOS Chromotest	
		H2AX (cellules humaines)	

- Corrélation entre nombre de micropolluants détectés et toxicité
- Corrélation entre l'occurrence de certains marqueurs et certains effets de toxicité ?

# Association d'une toxicité à un spectre

Les données issues de spectrométrie de masse sont volumineuses et très complexes

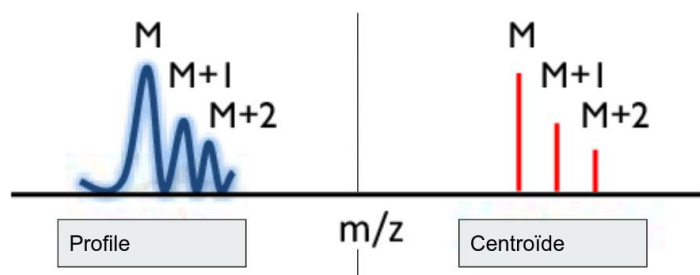
=> **Nécessité de prétraitement des données**



Les descripteurs sont des composés détectés avec un unique RT et masse (+ isotopes et un pattern d'adduits)



Impossible de faire des modèles robustes avec du bruit dans les données ou des descripteurs non-significatifs



Exemple analyse univariée :

- T-test (2 groupes)
- ANOVA (3 groupes ou plus)

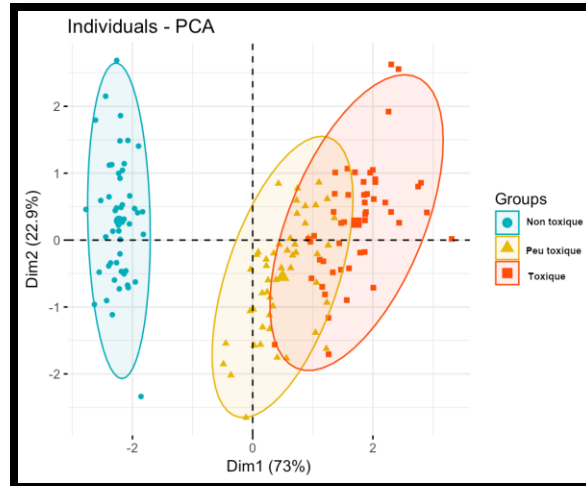
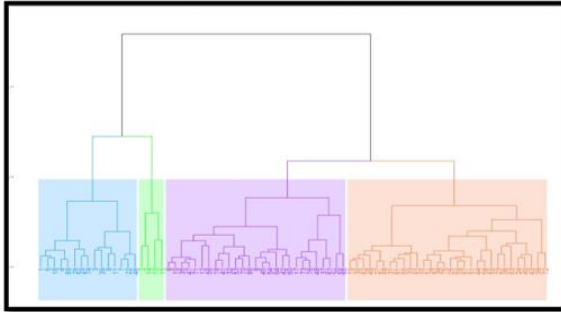
Si possible, les descripteurs doivent avoir une grande différence entre les groupes

# Association d'une toxicité à un spectre

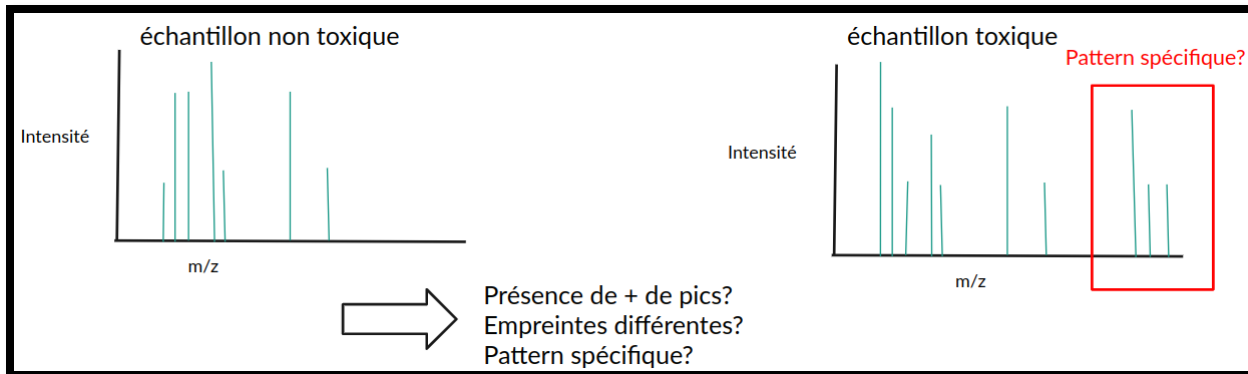
## Non supervisée

Ne s'intéresse pas à la variable cible de toxicité mais à l'identification des clusters ou des groupes en fonction des données non-étiquetées.

Classification ascendante hiérarchique   Analyse en composantes principales



KMEANS, t-SNE, UMAP

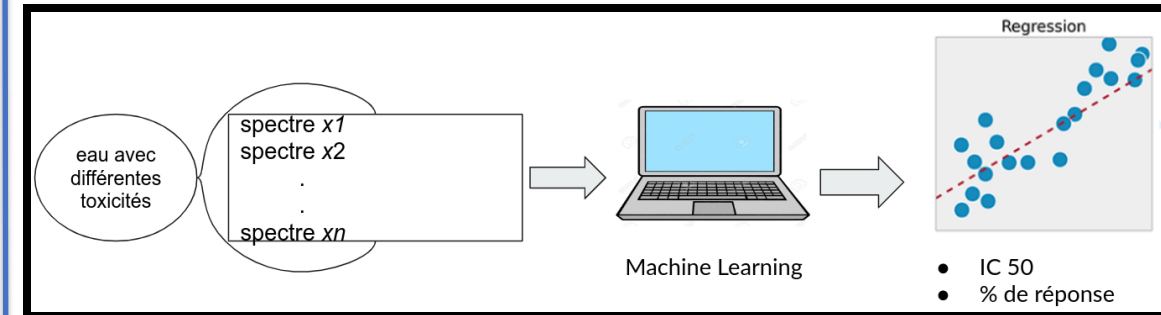


## Supervisée

Prédiction sur la toxicité de l'échantillon associé à son spectre

Différents algorithmes :

- Random Forest
- SVM
- Decision tree regression
- Multiple linear regression
- Etc...



# Comparaison empreintes chimiques de différentes eaux



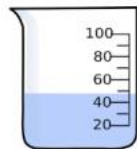
Prélèvements d'eau à Valenton et ajout d'eau usée en attendant les résultats écotoxicologiques.



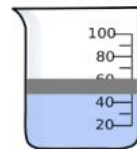
- Comparaison des empreintes digitales chimiques.
- Analyse non supervisée : observation de cluster, séparation entre les groupes
- Analyse supervisée : Création de modèles de prédiction de l'ajout d'eaux usées

Gamme d'échantillonnage à différentes concentrations d'ajout d'eau usée

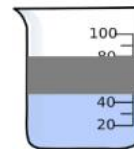
Ajout



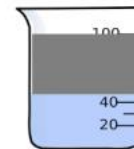
+ 0%



+20%



+50%



+100%

**1**

## **Développement d'outils statistiques et informatiques**

- Développement d'une chaîne analytique (de l'acquisition au traitement de données) en vue d'une application en routine (accessibilité et facilité d'utilisation).

**2**

## **Développements méthodologiques pour le couplage des données (Simulations d'échantillons, développement de modèles)**

- Développement de modèles de prédiction à visée opérationnelle pour la prédiction de la toxicité associée à un spectre (accessibilité et facilité d'utilisation).

**3**

## **Application au suivi de la qualité des eaux**

- Validation du modèle de prédiction en condition "réelle" avec des échantillons issus des campagnes de prélèvements.





# LEESU

laboratoire eau environnement systemes urbains



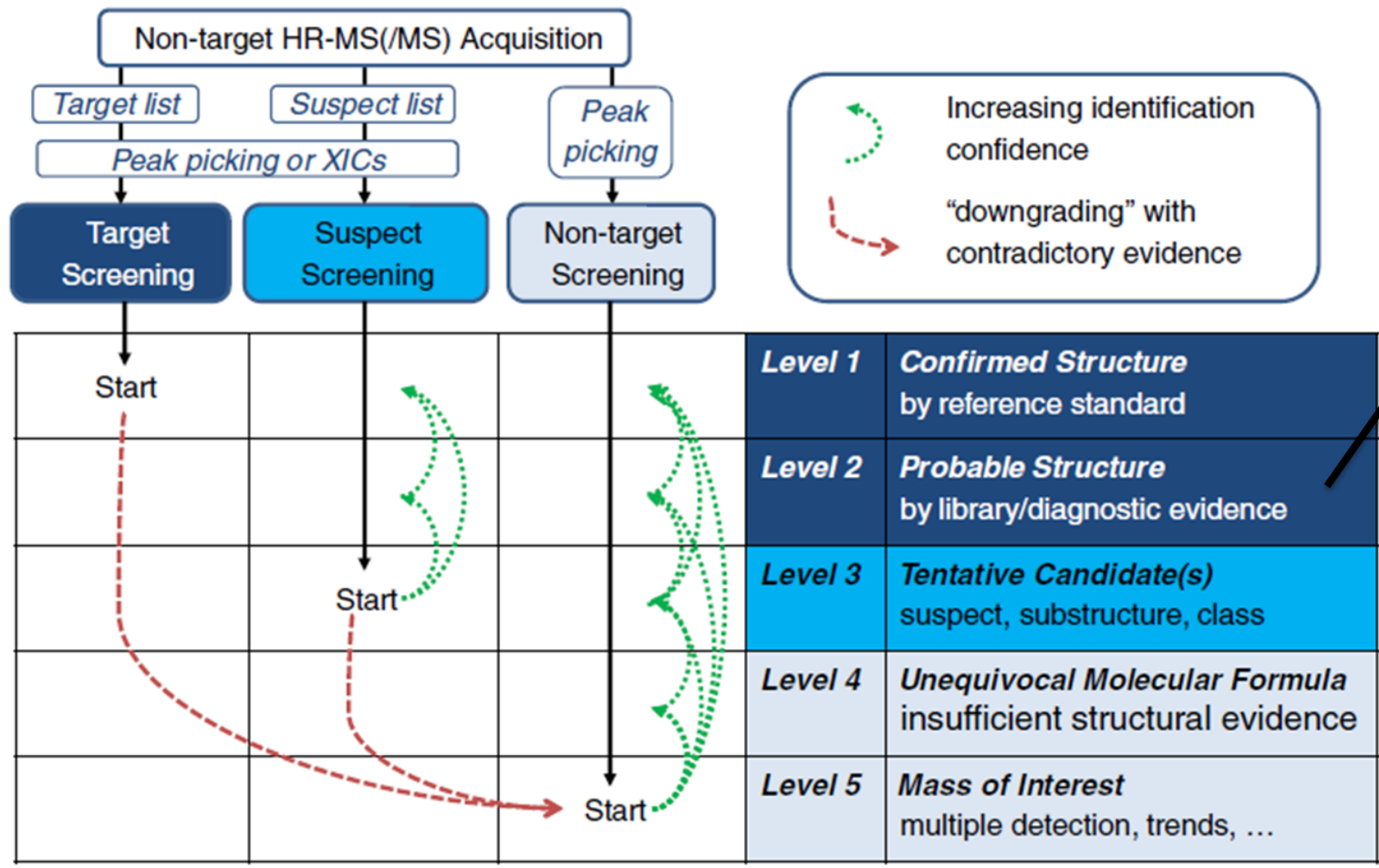
*Merci pour votre attention*

# LEESURIALES



**ANNEXES**

# Les niveaux de confiance d'identification



Base de données :

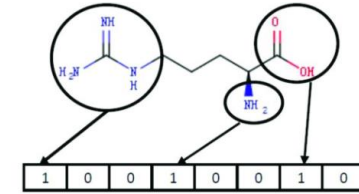
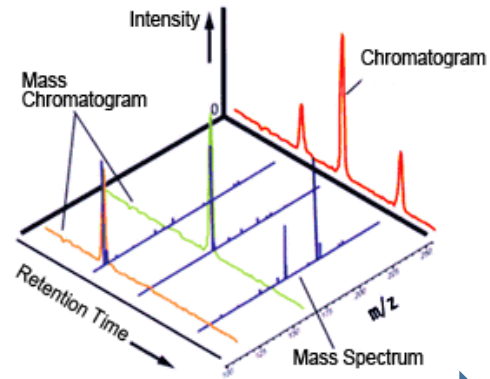


Schymanski et al, Environmental Science and Technology, 2014

# Analyse non-ciblée basée sur les données du QTOF (HRMS)

 PRAMMICS

UPLC-IMS-QTOF (HRMS) Plateforme PRAMMICS

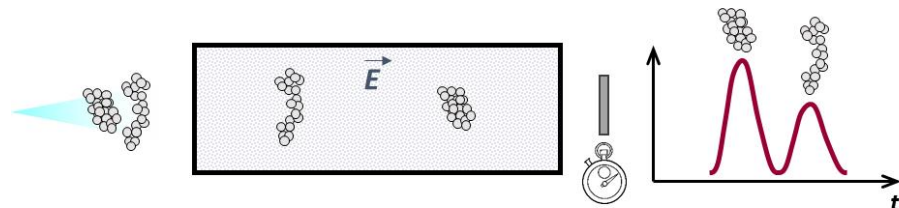


Empreinte digitale chimique

Traitement et analyse des données



Apport d'une 3e dimension : la mobilité ionique



- **+** une molécule est petite ou repliée, **+** elle se déplacera rapidement dans le tube car elle subira moins de friction
- **Inversement** pour une molécule grosse/ largement déployée.

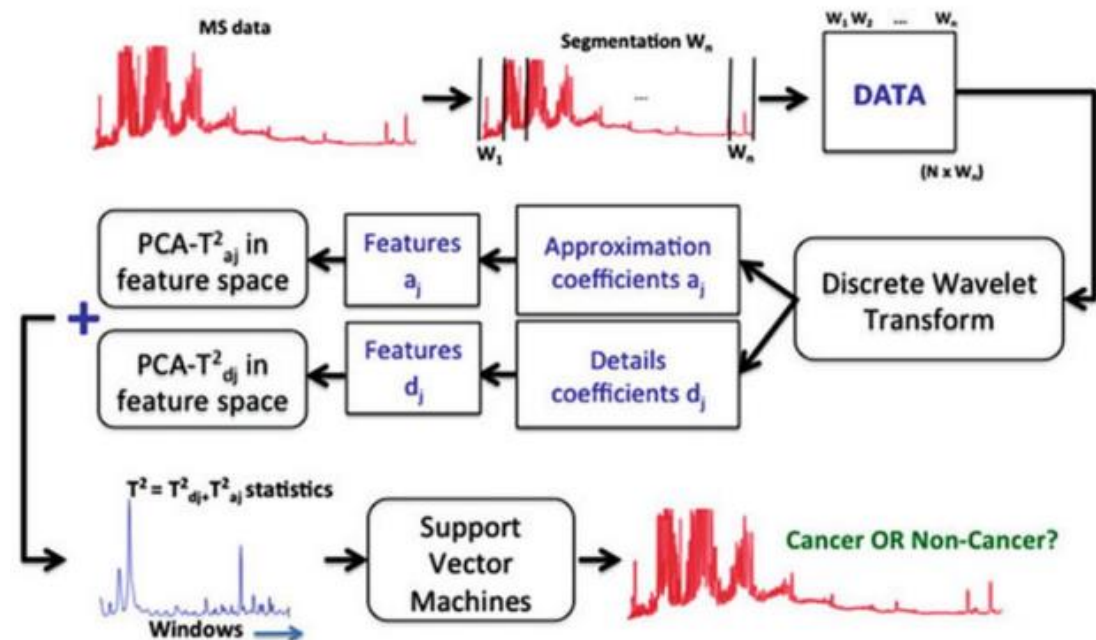


# Association d'une variable cible à un spectre

## Méthodes utilisées dans le domaine de la santé

### Cancer :

- *Petricoin and al (2002)* ont réussi à identifier un cluster de pattern de données de MS spécifiques aux spectres issus du cancer de l'ovaire.
- *Li and al (2016)*. ont crée un modèle de prédiction pour prédire la cancérogénicité.
- *Yildiz et al (2007)*. ont développé un modèle SVM de prédiction pour l'identification de cancer du poumon à partir des données de MS.
- *Cohen, A. and al (2018)*. Utilisation des coefficients du filtre d'ondelette lors de la création de leur modèle afin de réduire la complexité des données.



### Limite

Nécessité d'avoir un jeu de données conséquent et équilibré. Les performances du modèle dépendent de la qualité des données.

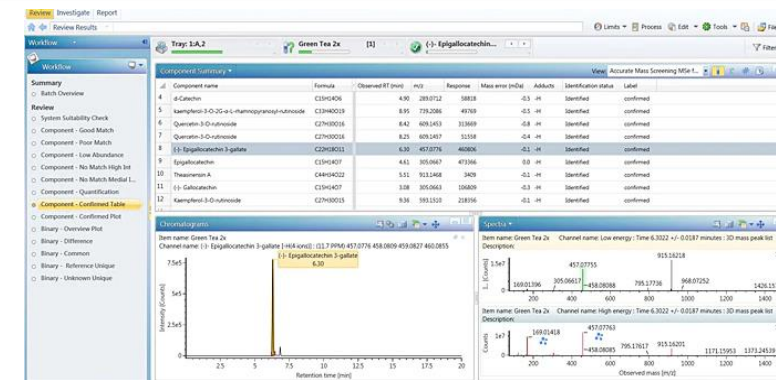
### Avantages :

Possibilité de cribler beaucoup d'échantillon et gain de temps.

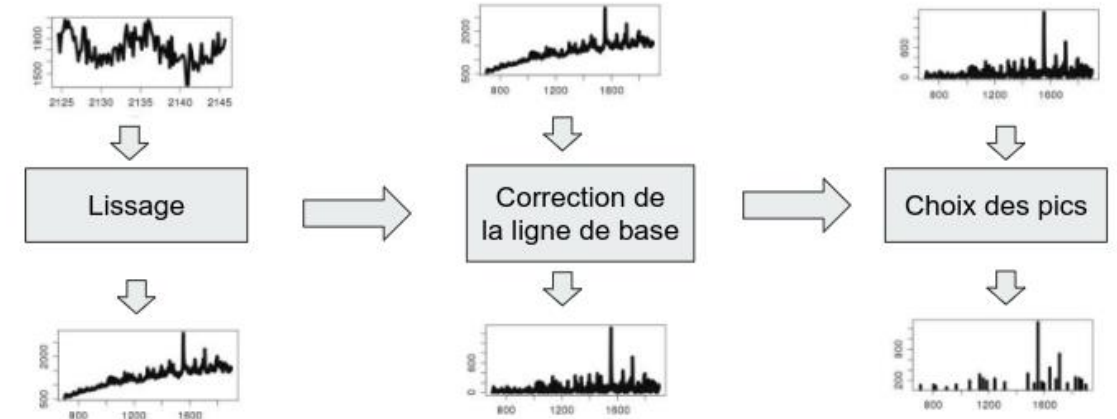
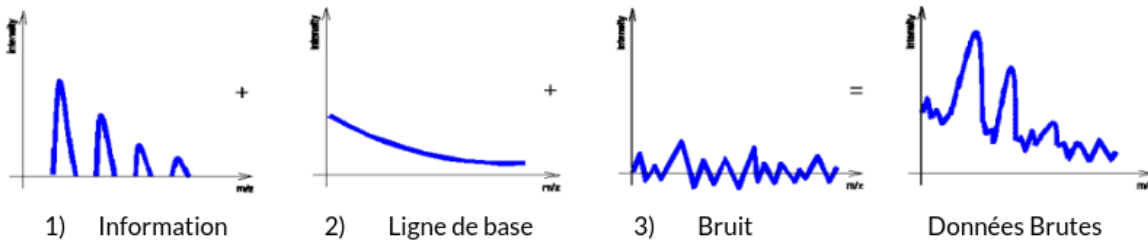


# Traitement du signal acquis

Traitement du signal par l'intermédiaire de l'outil UNIFI de Waters



Un spectre de masse typique peut être décomposé en trois termes additifs



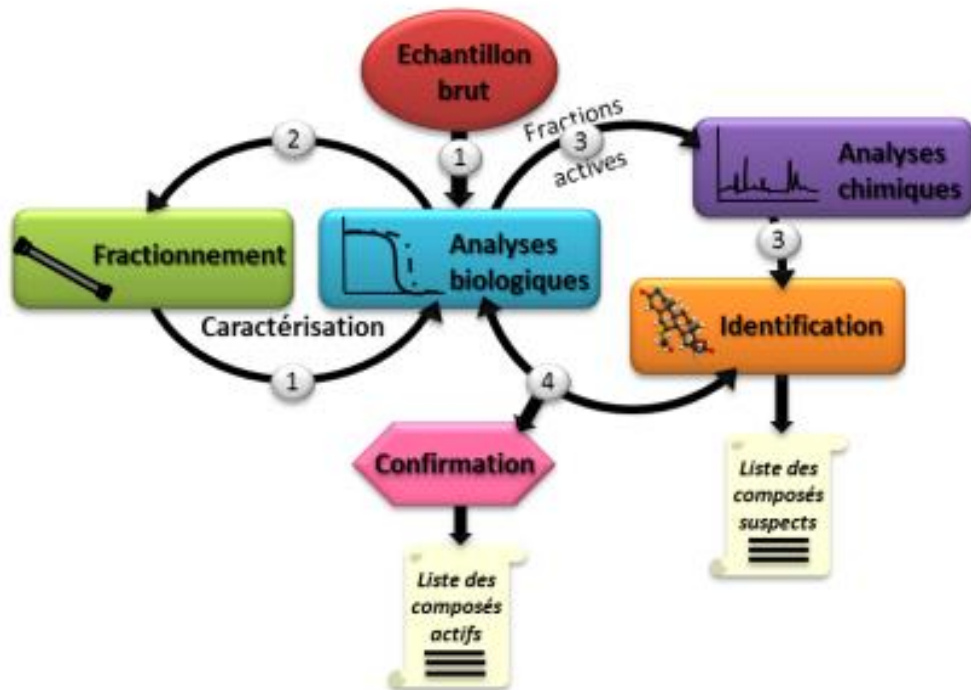
Bauer, Chris & al (2011). Evaluation of Peak-Picking Algorithms for Mass Spectrometry. Methods in molecular biology (Clifton, N.J.)

# L'analyse dirigée par les effets(EDA)

" *Effect-directed analysis (EDA) has become useful for identification of toxicant(s) that occur in mixtures in the environment, especially those that are causative agents of specific adverse effects* "

Seongjin Hong and al. Effect-Directed Analysis: Current Status and Future Challenges (2016). Open science Journal

L'EDA est basée sur une combinaison d'analyses biologiques (tests de toxicité), de procédures de fractionnement, suivies d'une analyse chimique, et facilite l'identification des principales substances toxiques.



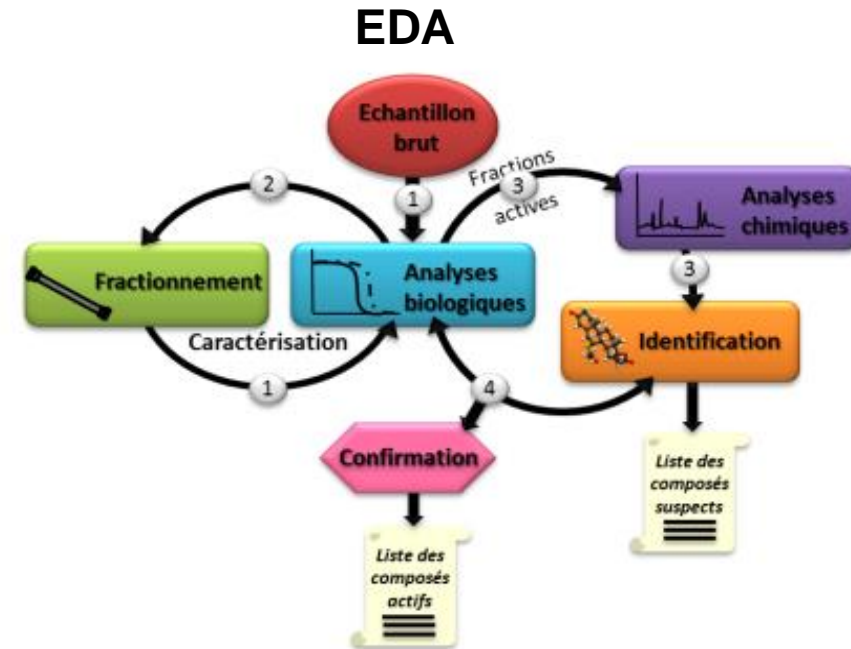
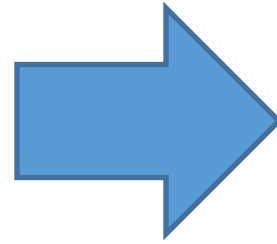
L'hypothèse de base de l'EDA est qu'une partie de composés actifs du mélange est responsable des effets toxiques

## Limites

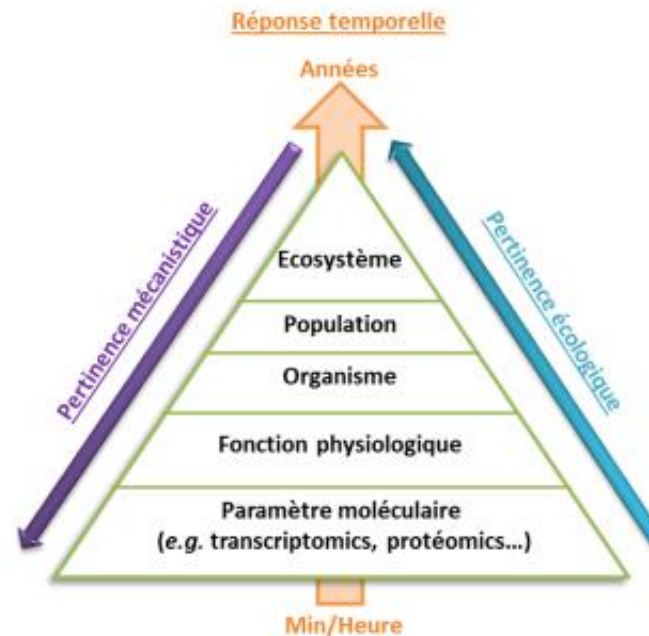
L'EDA est une démarche pluridisciplinaire. Cela nécessite plusieurs étapes afin de simplifier et séparer les molécules ayant un effet induisant une toxicité. Ces nombreuses phases rendent l'approche EDA relativement longue, fastidieuse et coûteuse.

# Campagne toxique : Analyse dirigée par les effets

Si les échantillons d'une campagne sont considérés comme très toxiques, l'EDA permettra de faciliter l'identification des composés induisant la toxicité.

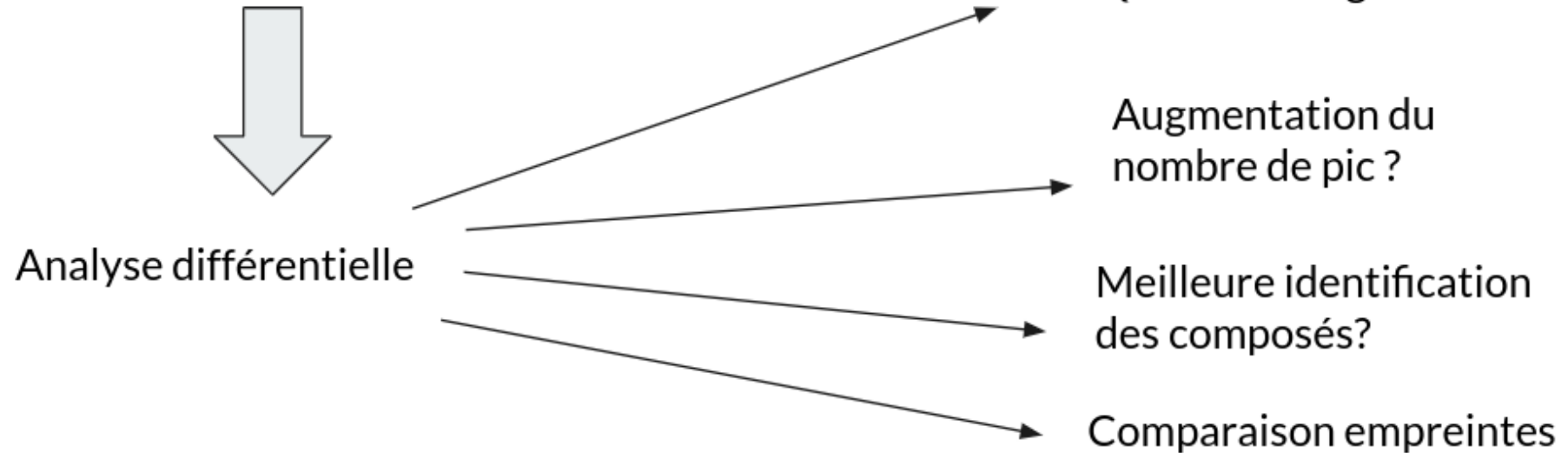


Les outils plus représentatifs de l'impact environnemental : **les modèles in vivo**



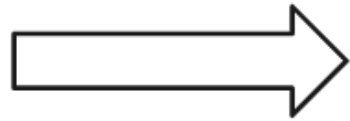
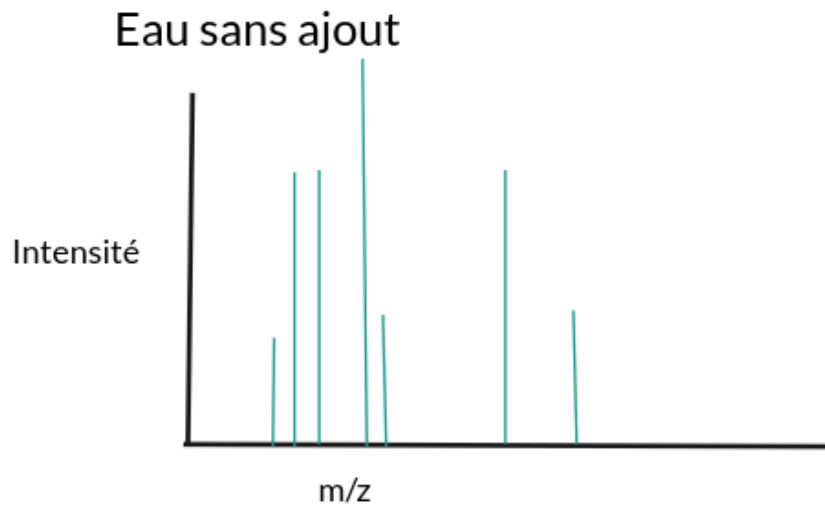
# Rôle de la mobilité ionique sur la qualité des signaux ?

## Comparaison spectre avec IMS et sans IMS



# Comparaison empreintes chimiques de différentes eaux

Simulation laboratoire: eau de seine dont on injecte eau usée ou eau de ruissellement



Impact de l'ajout?  
Présence de + de pic?  
Empreintes différentes?  
Pattern spécifique?

