

# «Catégorisation et indexation de la pollution de l'air extérieur par apprentissage supervisé neuronal appliqué à des systèmes multi-capteurs »

**Département Sciences de  
l'Atmosphère et Génie de  
l'Environnement (S.A.G.E.)**

Nathalie Redon

# Objectifs

- Offrir un outil alternatif, complémentaire des stations fixes de surveillance de la qualité de l'air

Collaboration avec

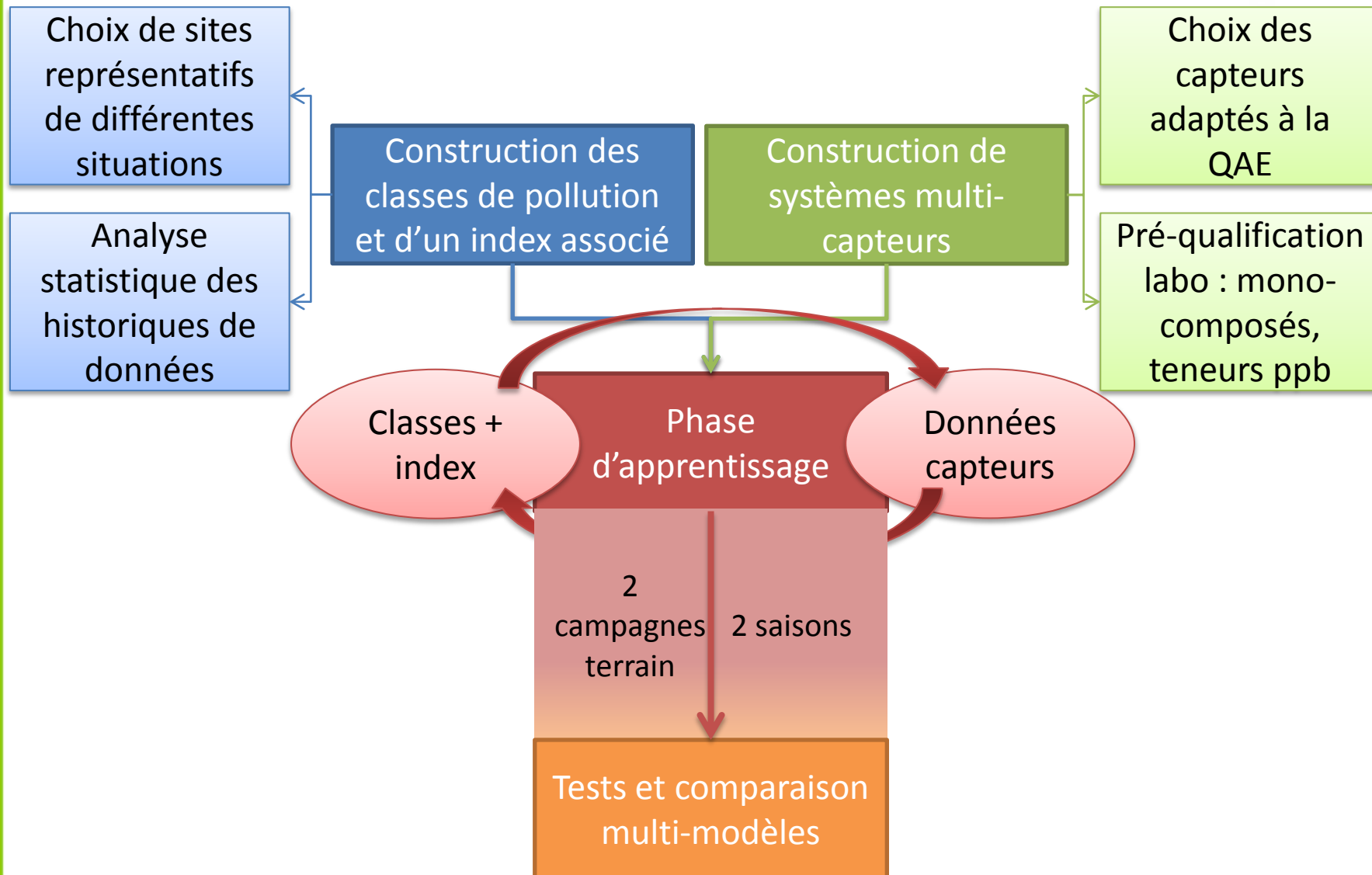


et

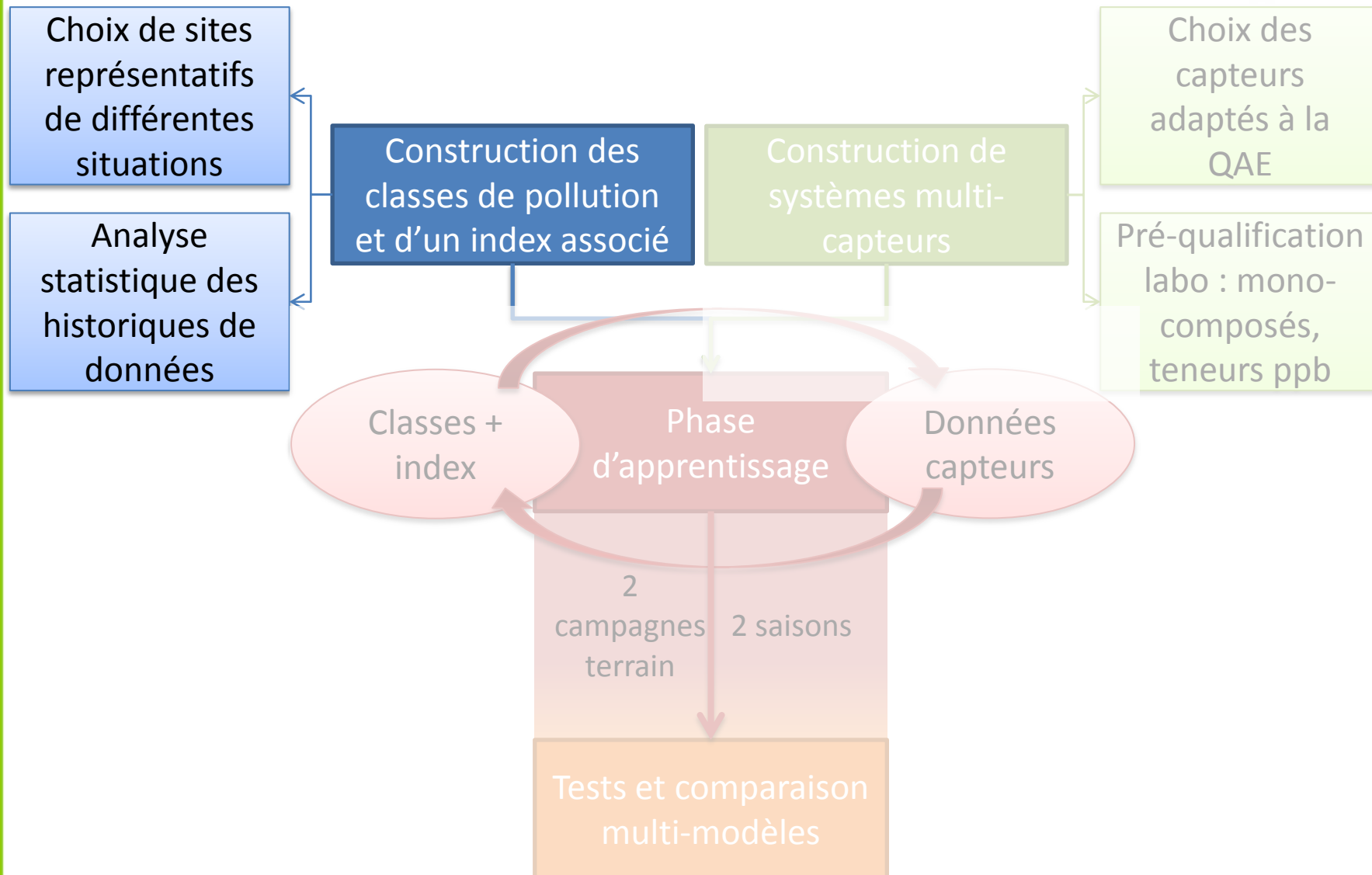


- Améliorer la résolution spatio-temporelle des mesures
- Cartographie qualitative/semi-quantitative

# Démarche



# Démarche



# Classes de pollution + index

Collaboration avec



et



Choix de 7 sites représentatifs de différentes situations



-  Roubaix: Trafic
-  Valenciennes: Trafic
-  Béthune: Urbain
-  Creil: Urbain
-  Douai: Urbain
-  Nogent sur Oise : Périurbain
-  Saint Amand les Eaux: Périurbain

# Classes de pollution + index

Collaboration avec



et



Choix de 7 sites représentatifs de différentes situations

Période	Site	Type de station	Gaz mesurés par ATMO
Eté	Creil	Urbain	NO-NO <sub>2</sub> -O <sub>3</sub> -PM <sub>2,5</sub> -PM <sub>10</sub> -CO-SO <sub>2</sub>
	Nogent	Périurbain	NO-NO <sub>2</sub> -O <sub>3</sub> -PM <sub>10</sub>
Automne	Roubaix	Trafic	NO-NO <sub>2</sub> -O <sub>3</sub> -CO-PM <sub>10</sub>
	Valenciennes	Trafic	NO-NO <sub>2</sub> -O <sub>3</sub> -PM <sub>10</sub>
	Saint Amand	Périurbain	NO-NO <sub>2</sub> -O <sub>3</sub>
	Douai	Urbain	NO-NO <sub>2</sub> -O <sub>3</sub> -PM <sub>2,5</sub> -PM <sub>10</sub>
	Béthune	Urbain	NO-NO <sub>2</sub> -O <sub>3</sub> -PM <sub>2,5</sub> -PM <sub>10</sub>

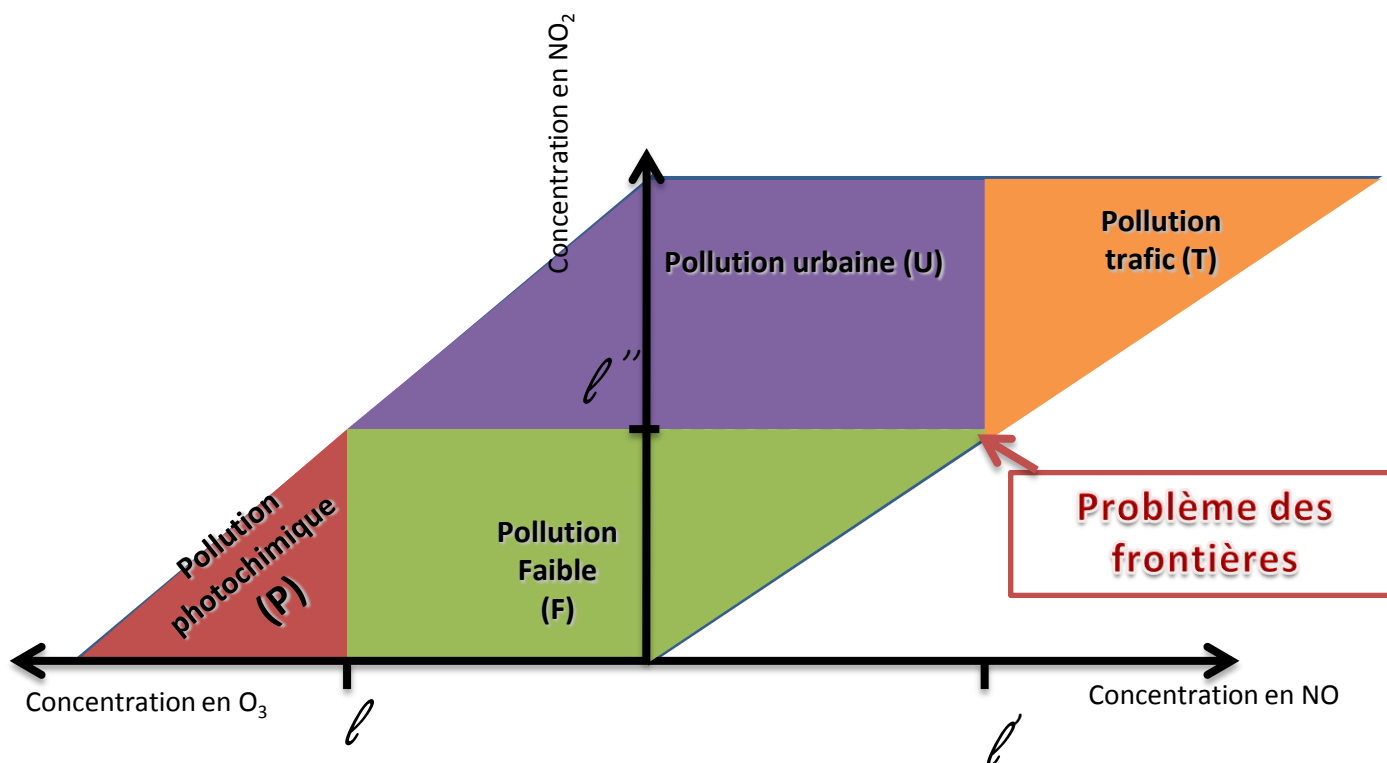
Analyse statistique des historiques de données

1 an de données analysées dont 3 communes à l'ensemble des sites : NO, NO<sub>2</sub>, O<sub>3</sub>



# Classes de pollution + index

Il existe une forte relation entre ces 3 polluants: NO, NO<sub>2</sub>, O<sub>3</sub>  
 → Il est possible de définir des seuils pour catégoriser la pollution

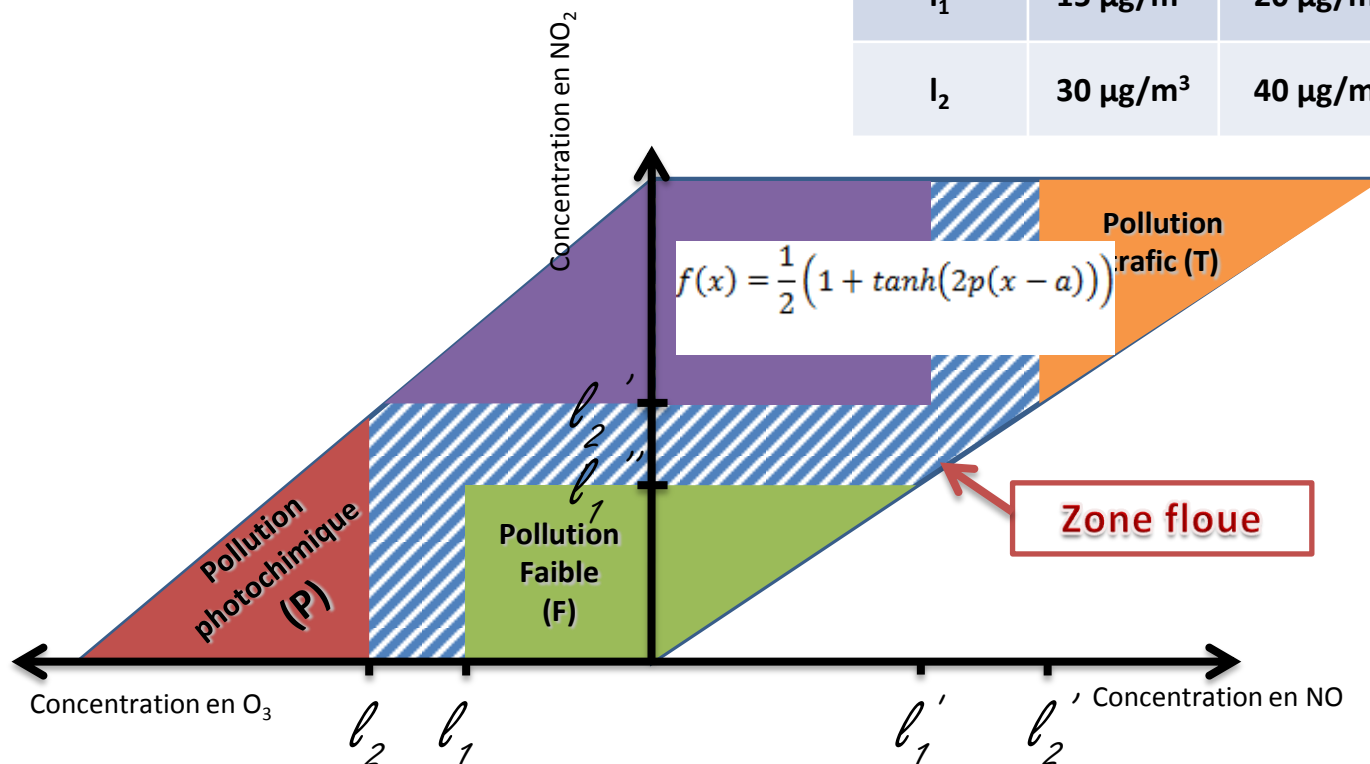


# Classes de pollution + index

La logique floue au service de la catégorisation :

→ 2 niveaux de seuils sont introduits

Seuils	NO	NO <sub>2</sub>	O <sub>3</sub>
$l_1$	15 µg/m <sup>3</sup>	20 µg/m <sup>3</sup>	30 µg/m <sup>3</sup>
$l_2$	30 µg/m <sup>3</sup>	40 µg/m <sup>3</sup>	50 µg/m <sup>3</sup>



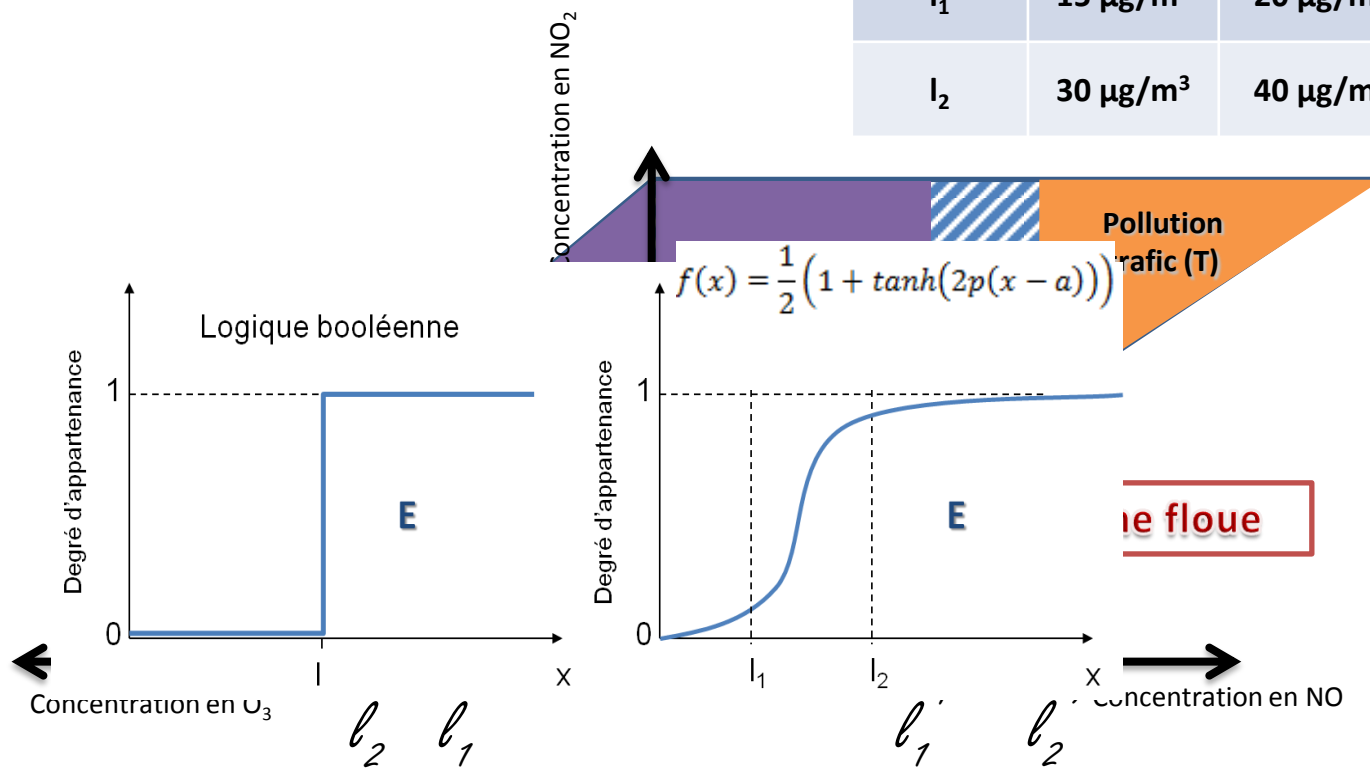


# Classes de pollution + index

La logique floue au service de la catégorisation :

→ 2 niveaux de seuils sont introduits

Seuils	NO	NO <sub>2</sub>	O <sub>3</sub>
$l_1$	15 µg/m <sup>3</sup>	20 µg/m <sup>3</sup>	30 µg/m <sup>3</sup>
$l_2$	30 µg/m <sup>3</sup>	40 µg/m <sup>3</sup>	50 µg/m <sup>3</sup>

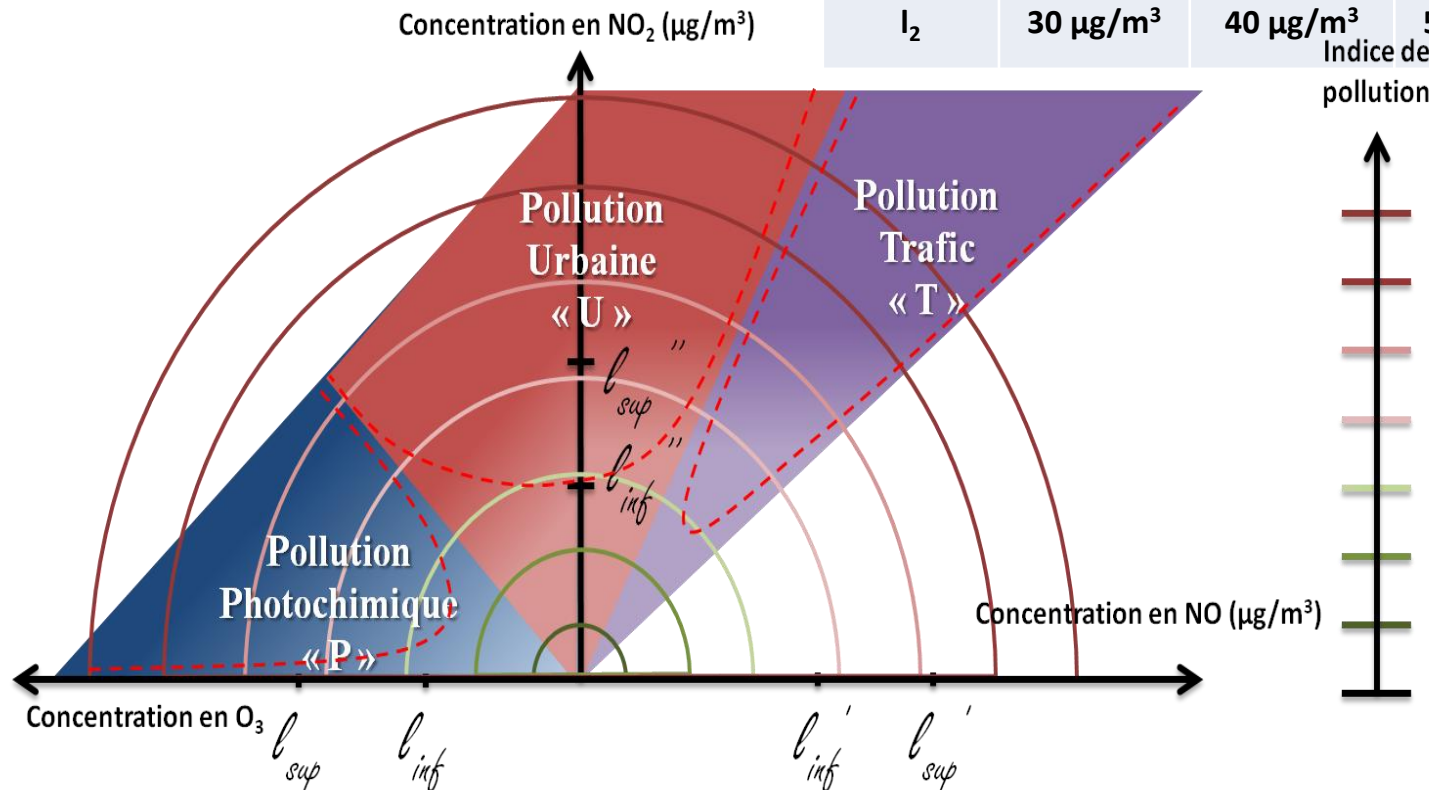


# Classes de pollution + index

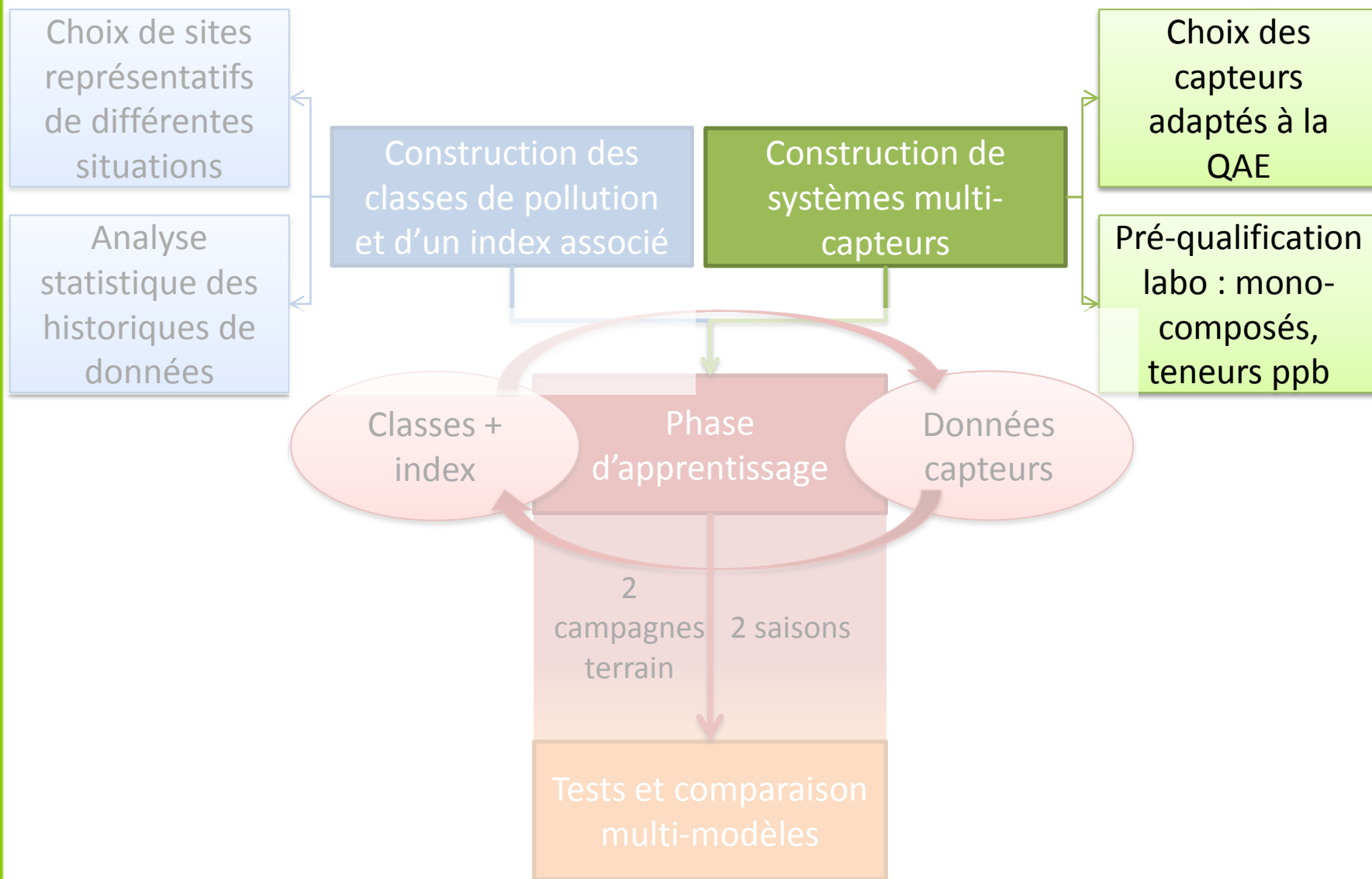
La logique floue au service de la catégorisation :

→ 2 niveaux de seuils sont introduits

Seuils	NO	NO <sub>2</sub>	O <sub>3</sub>
I <sub>1</sub>	15 µg/m <sup>3</sup>	20 µg/m <sup>3</sup>	30 µg/m <sup>3</sup>
I <sub>2</sub>	30 µg/m <sup>3</sup>	40 µg/m <sup>3</sup>	50 µg/m <sup>3</sup>



# Démarche



# Systèmes multi-capteurs

Choix des capteurs adaptés à la Qualité de l'Air Extérieur

polluants gazeux réglementés + COV + gaz non réglementés

Capteurs	Gaz cibles	Gamme constructeurs
MICS5525	Bonne sensibilité au CO Filtre charbon actif pour éliminer l'éthanol et COV	1-1000 ppm
MICS2710	NO <sub>2</sub>	0,5-5 ppm
MICS2611	O <sub>3</sub>	0,01-1 ppm
TGS2600	Large spectre des contaminants de l'air	500-10000 ppm
TGS2602	COV– odeurs, H <sub>2</sub> S	30-10000 ppm
TGS2620	Alcool et vapeurs solvants	50-5000 ppm
TGS2611	Gaz combustibles (Méthane, isobutane...)	500-10000 ppm
MICS5521	CO-COV et hydrocarbures.	1-1000 ppm
MICS5914	NH <sub>3</sub>	0,1-100 ppm

3 Systèmes identiques ont été créés

# Systèmes multi-capteurs

Choix des capteurs adaptés à la Qualité de l'Air Extérieur

polluants gazeux réglementés + COV + gaz non réglementés

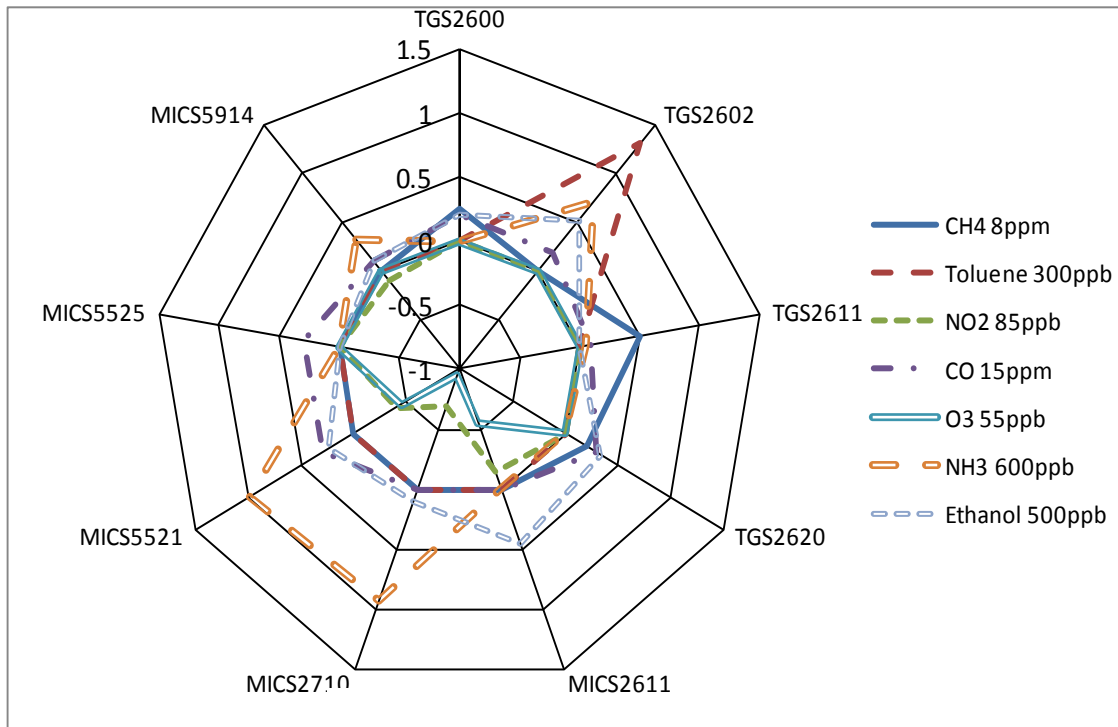


3 Systèmes identiques ont été créés



# Systèmes multi-capteurs

## Pré-qualification labo : mono-composés, teneurs ppb



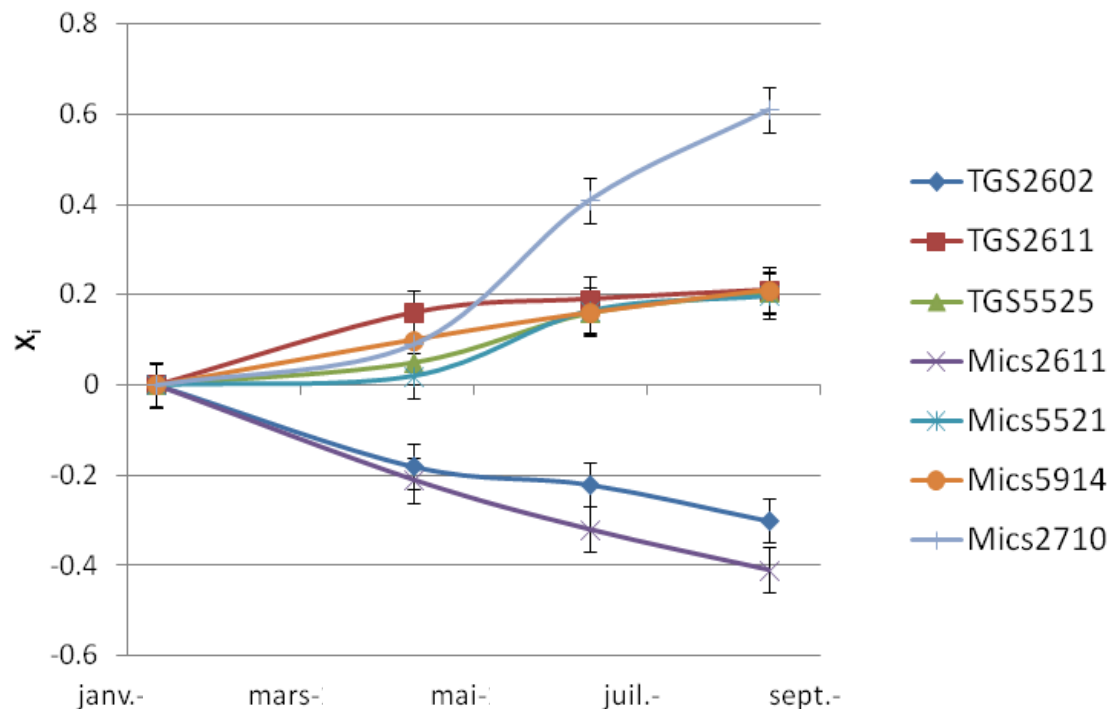
➤ Sensibilité à faibles teneurs démontrée



➔ 3 systèmes opérationnels pour le déploiement sur le terrain

# Systèmes multi-capteurs

Pré-qualification labo : mono-composés, teneurs ppb



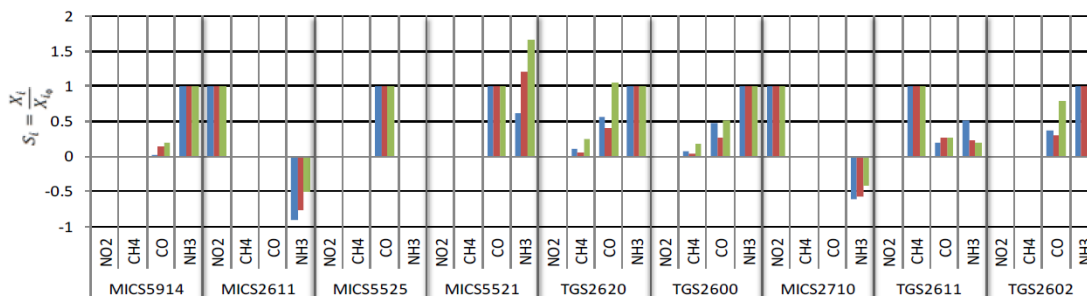
- Sensibilité à faibles teneurs démontrée
- Bonne stabilité long terme mais...
- ... correction de la dérive de la ligne de base sous air de référence à prendre en compte si nécessaire
- 

→ 3 systèmes opérationnels pour le déploiement sur le terrain



# Systèmes multi-capteurs

## Pré-qualification labo : mono-composés, teneurs ppb

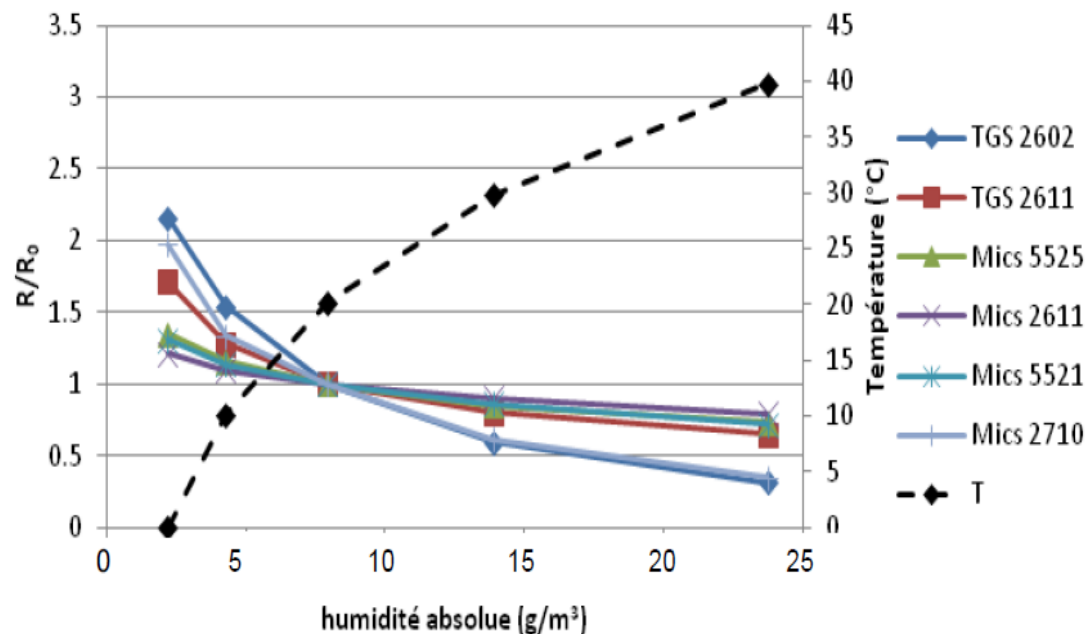


- Sensibilité à faibles teneurs démontrée
- Bonne stabilité long terme mais...
- ... correction de la dérive de la ligne de base sous air de référence à prendre en compte si nécessaire
- Bonne reproductibilité capteurs du même type (sous réserve des standardisation des données)
- 

**→ 3 systèmes opérationnels pour le déploiement sur le terrain**

# Systèmes multi-capteurs

Pré-qualification labo : mono-composés, teneurs ppb



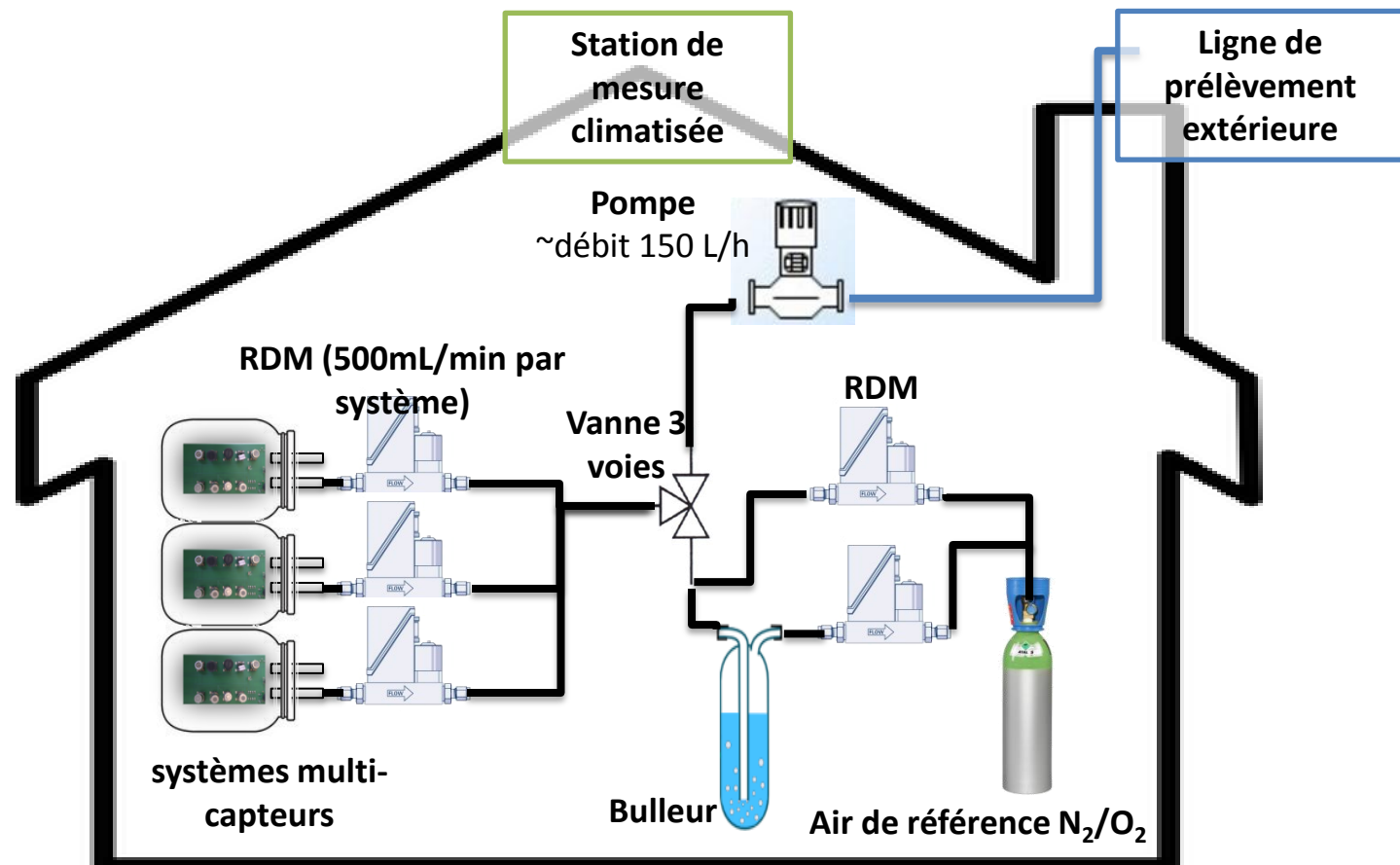
- Sensibilité à faibles teneurs démontrée
- Bonne stabilité long terme mais...
- ... correction de la dérive de la ligne de base sous air de référence à prendre en compte si nécessaire
- Bonne reproductibilité capteurs du même type (sous réserve des standardisation des données)
- Systèmes à maintenir sous T(°C)/RH% constantes

➔ 3 systèmes opérationnels pour le déploiement sur le terrain

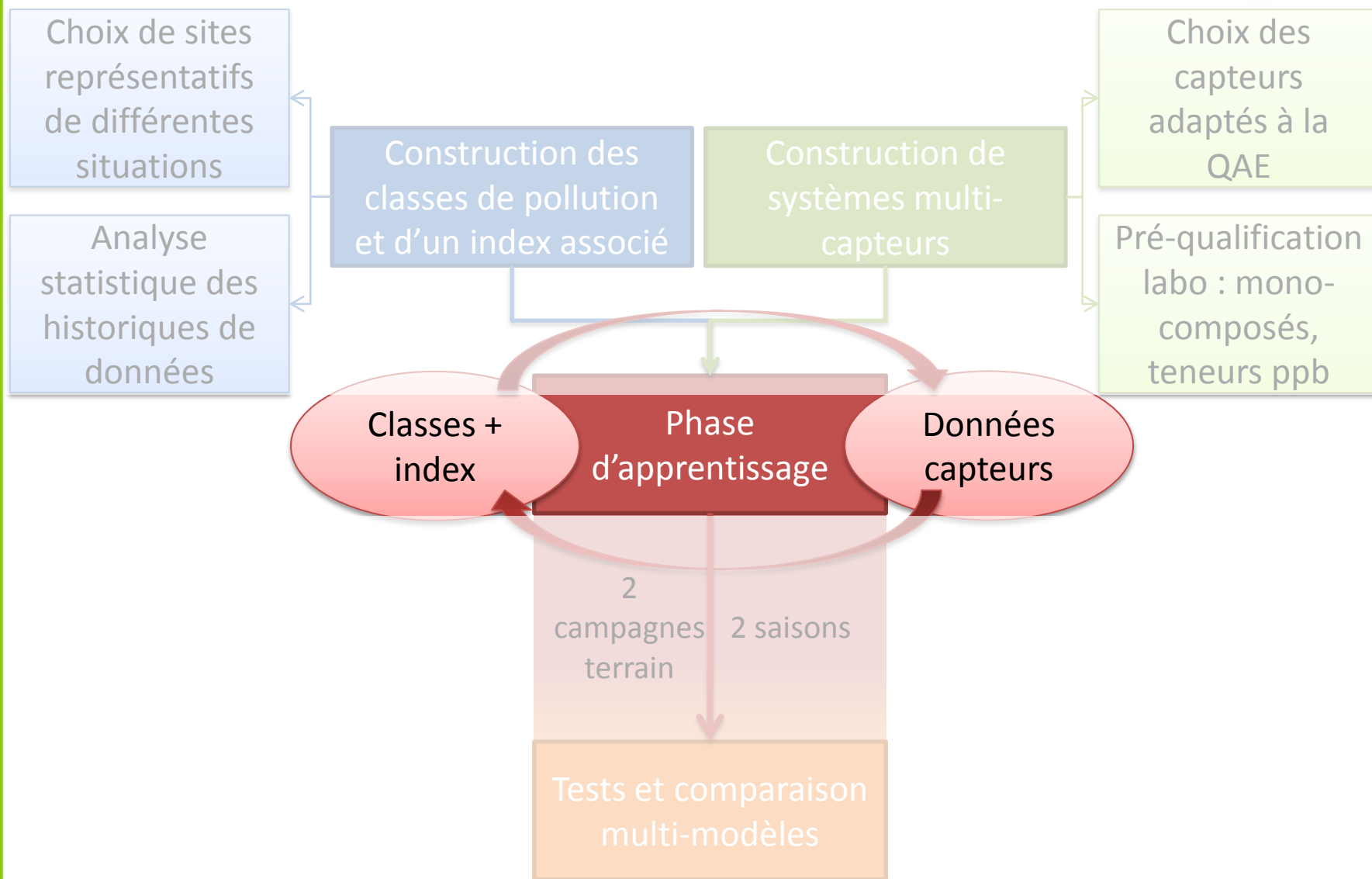
# Systèmes multi-capteurs

Pré-qualification labo : mono-composés, teneurs ppb

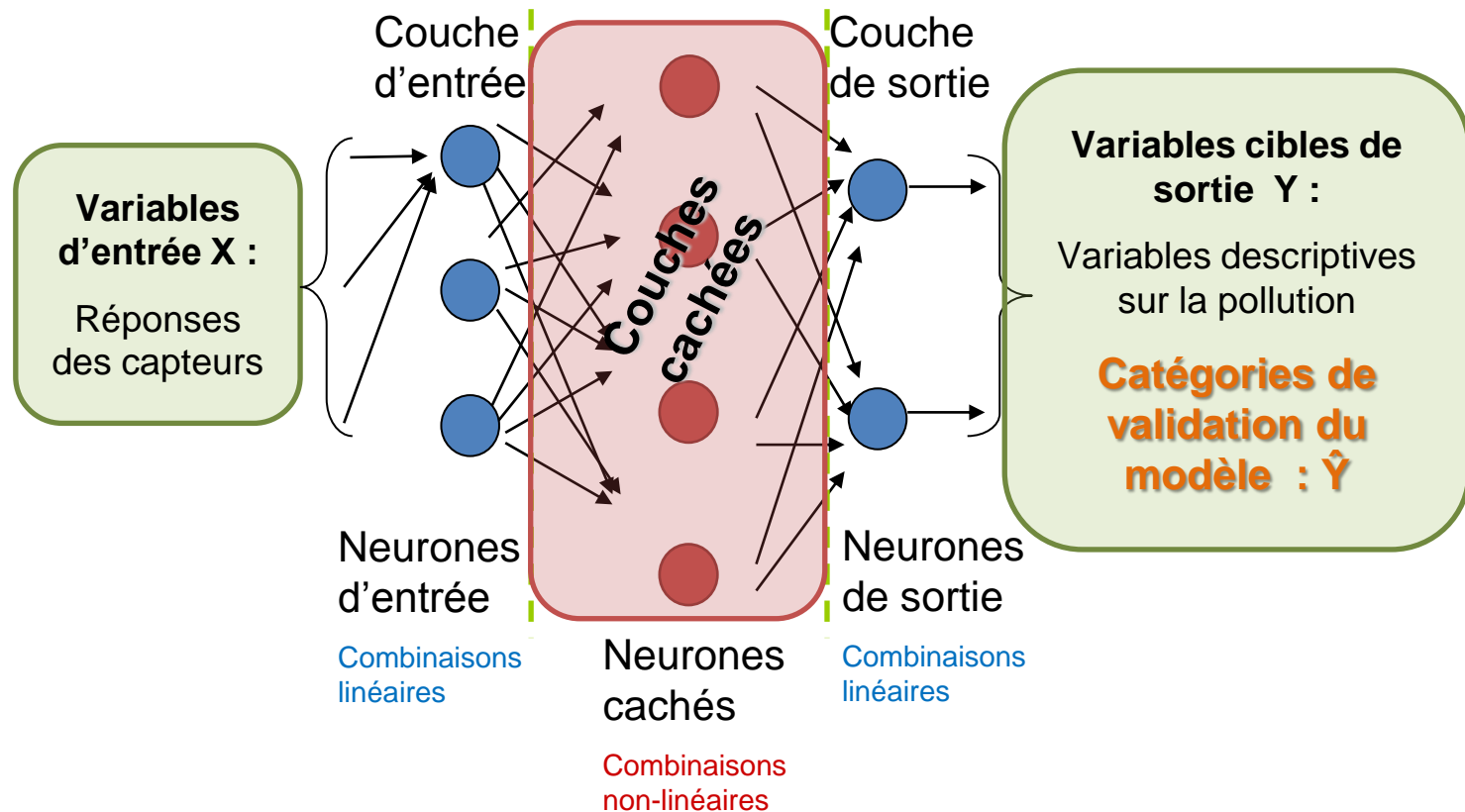
→ 3 systèmes opérationnels pour le déploiement sur le terrain



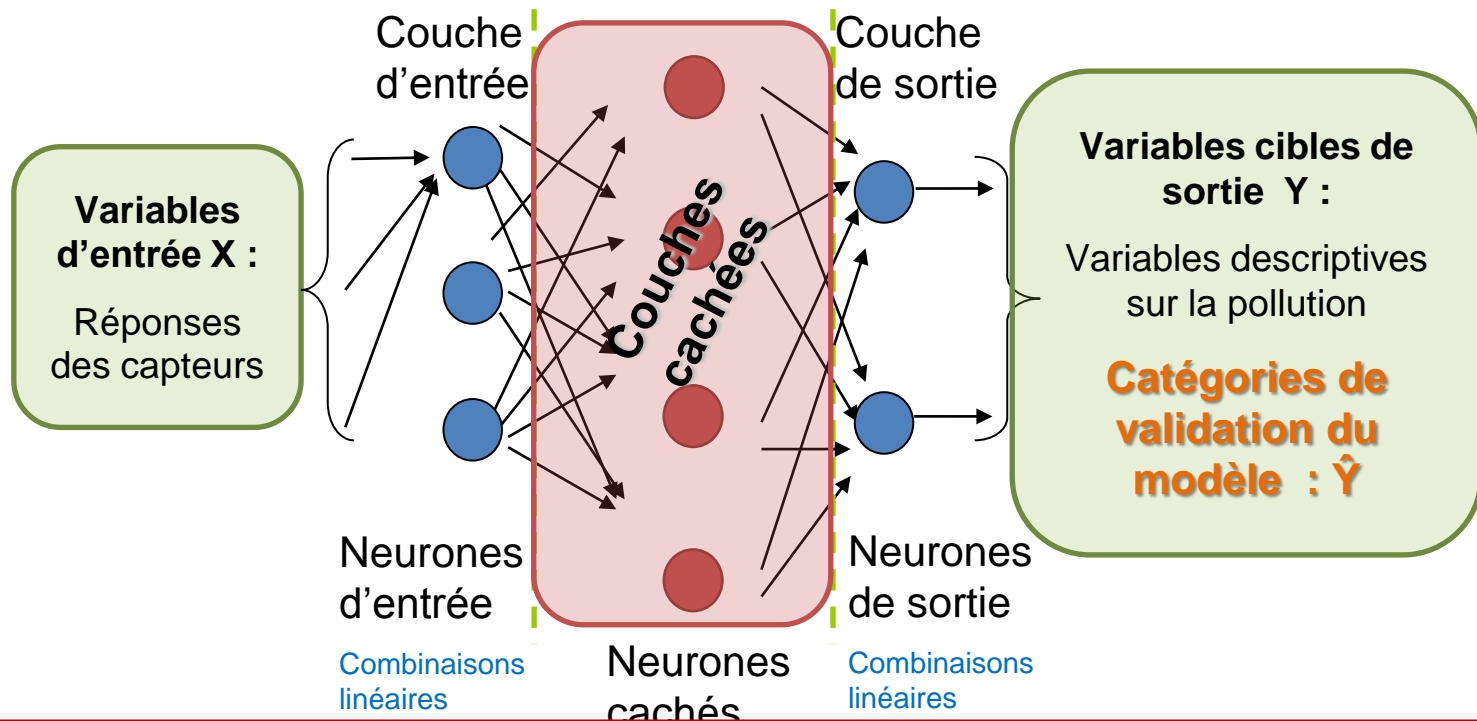
# Démarche



Les réseaux de neurones artificiels:  
adaptés aux réponses des capteurs car utilisation de fonctions non-linéaires



Les réseaux de neurones artificiels:  
adaptés aux réponses des capteurs car utilisation de fonctions non-linéaires



Dans notre cas : → 12 neurones d'entrées, 4 neurones de sortie  
→ une seule couche cachée  
→ fonction non-linéaire utilisée :  $f = \tanh(x_i)$

Réponses des  
systèmes multi-  
capteurs

## Données d'entrée du Réseaux de Neurones :

9 capteurs + 3 données additionnelles

- Réponse du capteur d'humidité
- Horodatage de la mesure
- Variable discrète indiquant les jours ouvrés et les jours non-ouvrés

X Variables

[12 Neurones]

$X_i$

HR%

hh:mm:ss

J-O

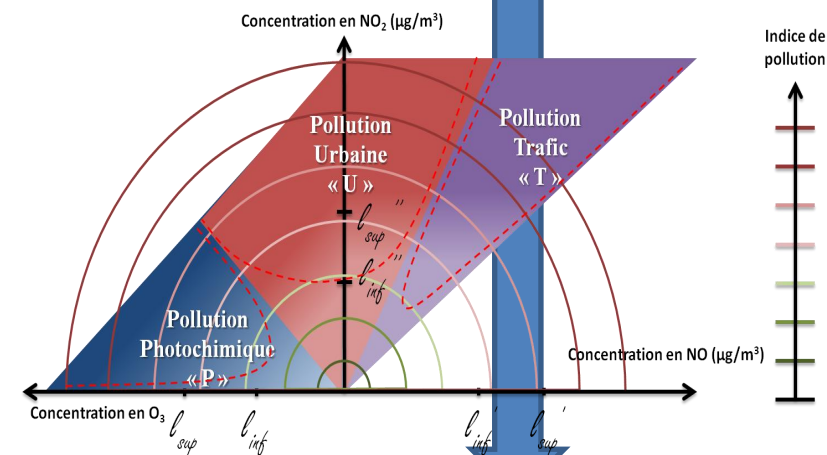
Modèle  
RN

Catégories  
estimées de  
pollution

Y Variables

$\hat{Y}$  Catégories  
[4 neurones]

Etape de validation  
du modèle RN

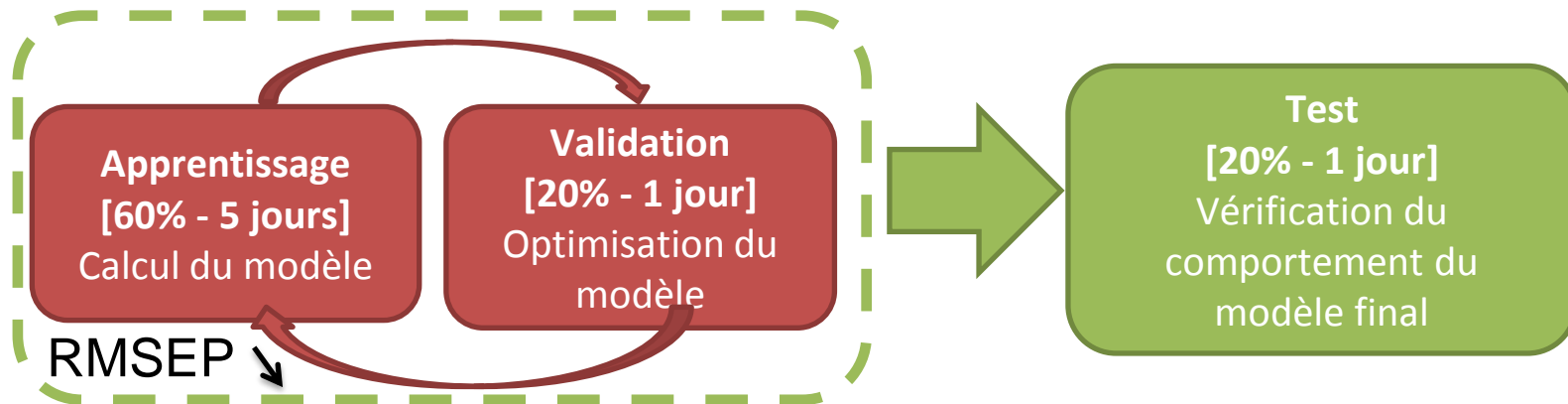




## Etape de validation du modèle RN

3 jeux de données doivent être utilisés:

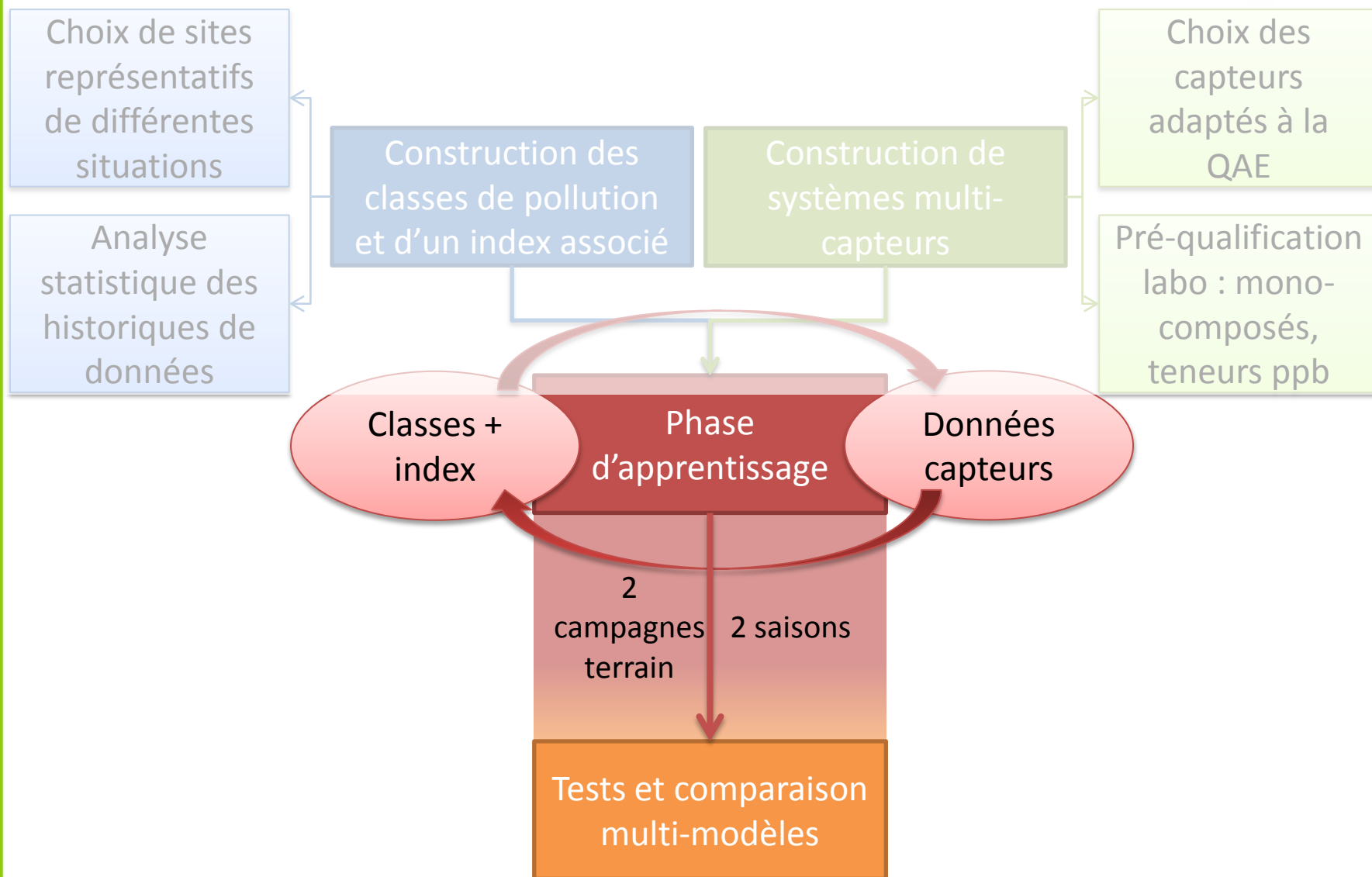
**N-itérations**



RMSEP: *Root mean squared error of prediction*

$$RMSEP = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n}}$$

# Démarche



## Catégorisation

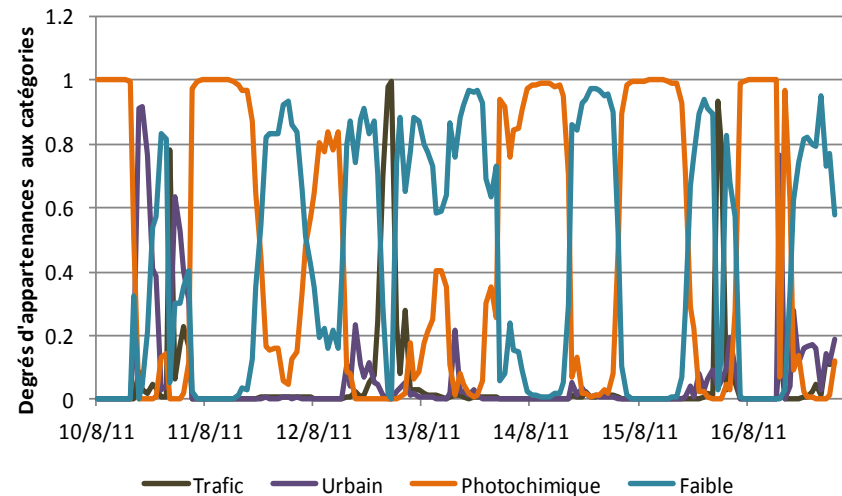
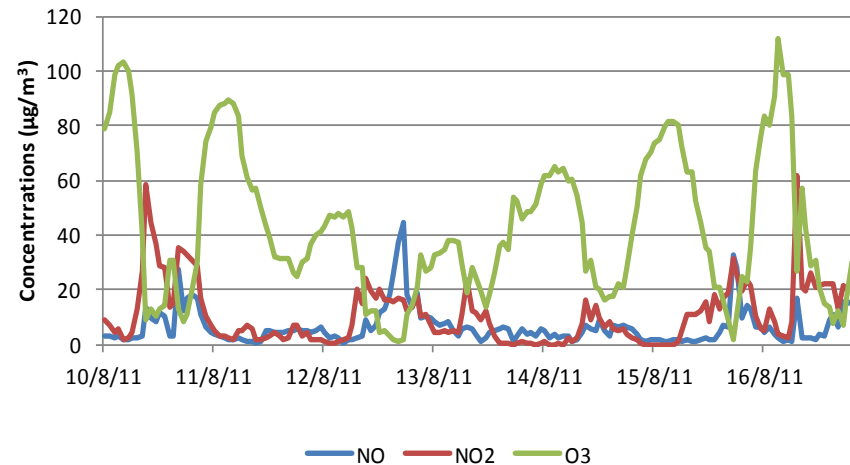
Site urbain Creil  
Système N°1

Données  
Analyseurs ATMO

Logique  
floue

Degrés  
d'appartenance  
aux catégories

**Classe détectée :**  
degré d'appartenance  
le plus élevé.



## Catégorisation

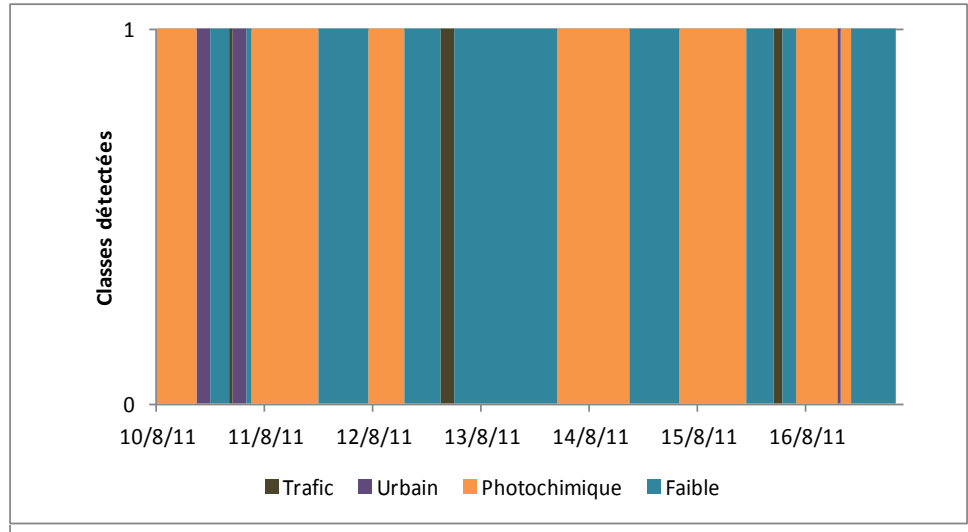
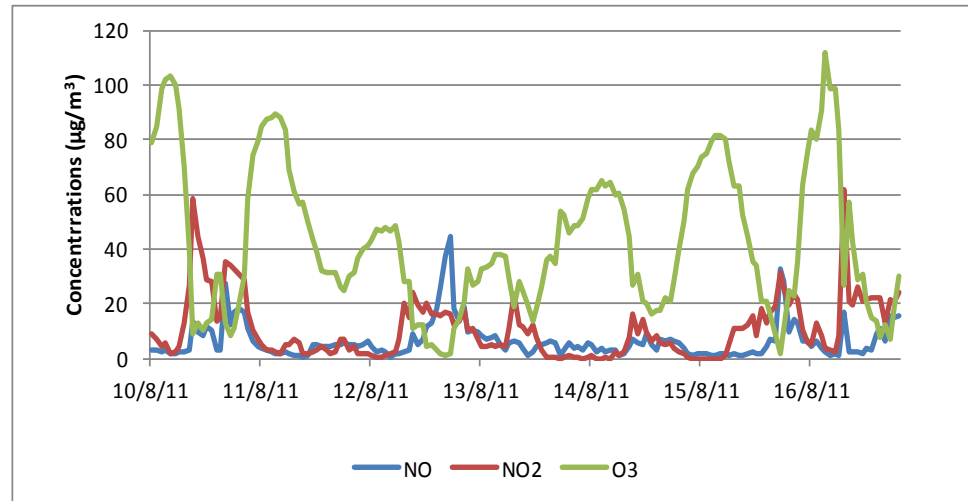
Site urbain Creil  
Système N°1

Données  
Analyseurs ATMO

Logique  
floue

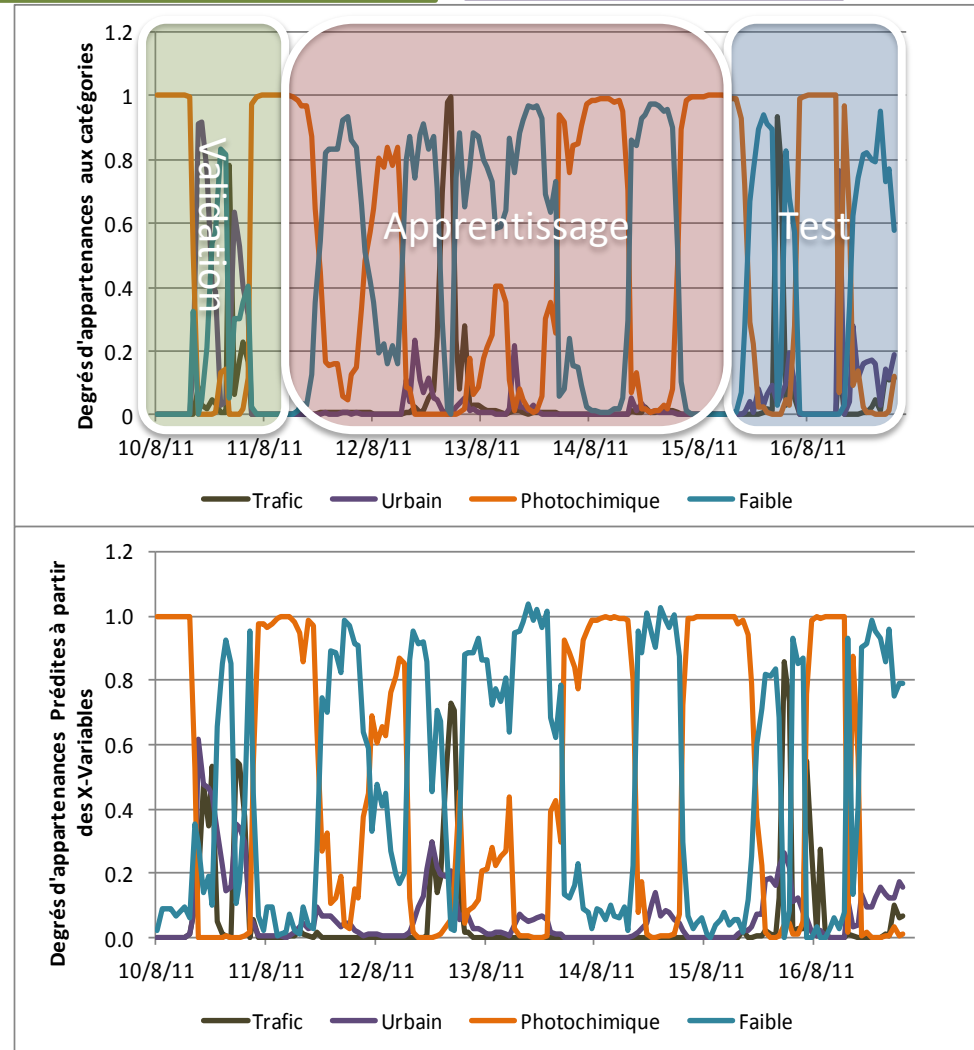
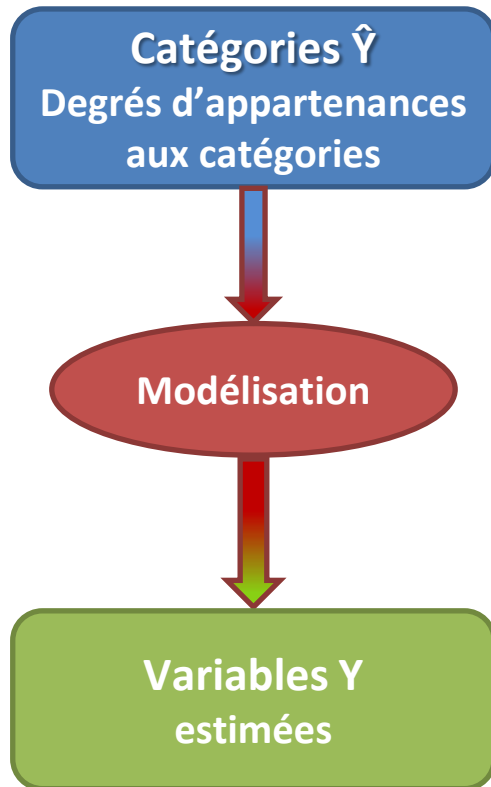
Degrés  
d'appartenances  
aux catégories

**Classe détectée :**  
degré d'appartenance  
le plus élevé.



## Modélisation

Site urbain Creil  
Système N°1



## Modélisation

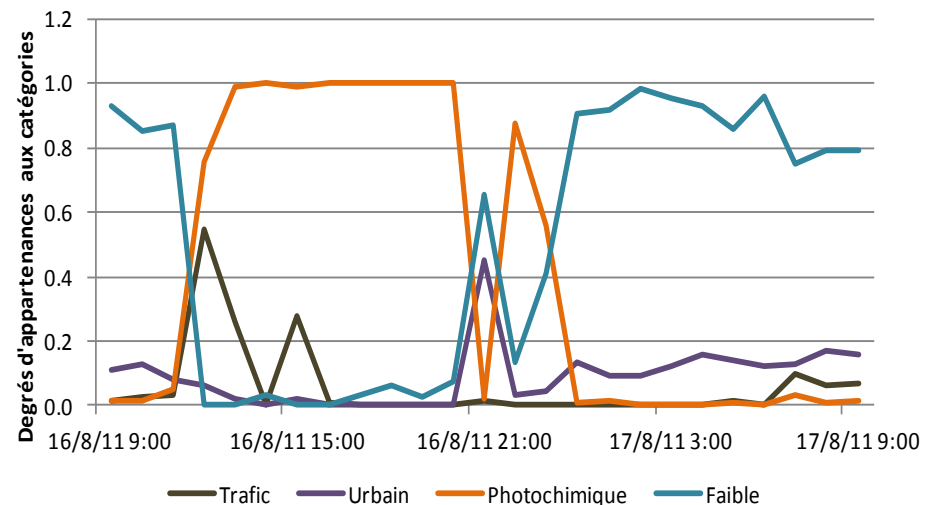
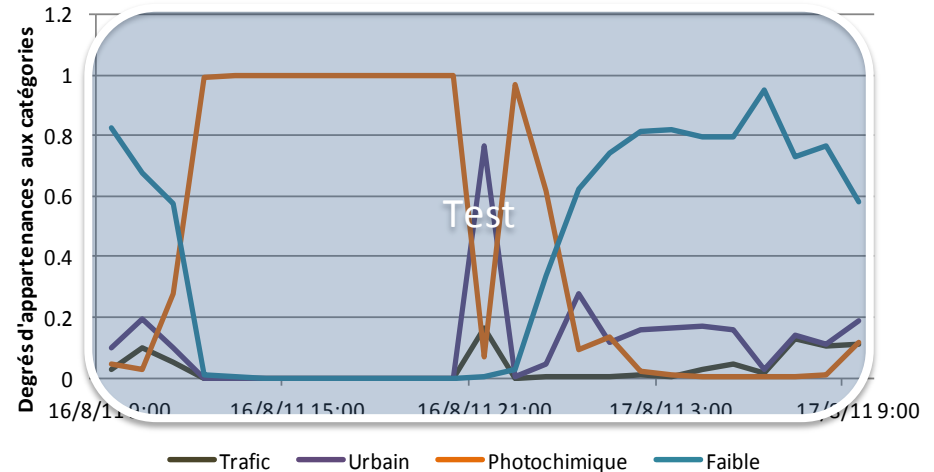
Site urbain Creil  
Système N°1

Etude du  
performance sur le  
jeu de test

Catégories  $\hat{Y}$

Variables  $Y$   
calculées

Taux de  
réussite:  
85%



## Portabilité des modèles ?

→ **Modèle idéal = modèle universel**

➤ **Portabilité inter-systèmes**

Site péri-urbain de Nogent

Modèle	Prédiction	Taux de réussite	RMSEP
Cellule 1	Cellule 1	90%	0.03
Cellule 2	Cellule 2	91%	0.01
Cellule 3	Cellule 3	89%	0.01
Cellule 1	Cellule 2	25%	0.2
Cellule 1+2	Cellule 1+2	83%	0.06

→ **Un modèle optimal doit être estimé sur un nombre suffisant de systèmes multi-capteurs clones**



## Portabilité des modèles ?

→ **Modèle idéal = modèle universel**

➤ **Portabilité multi-sites**

			Cellule 1	
Modèle	Modélisation	Test	Classification	RMSEP
mono cellule	Béthune	Béthune	91%	0.03
mono cellule	Béthune	Nogent	47%	0.15

**Apprentissage + Validation**



**Test**



## Portabilité des modèles ?

→ **Modèle idéal = modèle universel**

➤ **Portabilité multi-sites**

			Cellule 1	
Modèle	Modélisation	Test	Classification	RMSEP
mono cellule	Béthune	Béthune	91%	0.03
mono cellule	Béthune	Nogent	47%	0.15
<b>Mono – cellule et multi-station</b>	<b>Toute station sauf Béthune</b>	<b>Béthune</b>	<b>86%</b>	<b>0.09</b>

→ **Un modèle optimal doit être estimé sur un nombre suffisant de sites représentatifs de l'ensemble des catégories de pollution**

# De conclusions ...

## **Système multi-capteurs pour l'identification de signatures de pollution de l'air extérieur** ✓

- Phase de pré-qualification labo facultative

## **Catégorisation de types et niveaux de pollution** ✓

- Méthode statistique simple associée à une connaissance experte de la pollution à identifier
  - apprentissage supervisé ✓
    - Réseaux de neurones adaptés ✓
    - Excellente portabilité inter-cellules ou inter-sites ✓

## ...en perspectives

### **Système multi-capteurs pour l'identification de signatures de pollution de l'air extérieur**

- Compléter le système multi-capteurs avec des composants spécifiques (EC, optique, ...) + PM

### **Catégorisation de types et niveaux de pollution**

- Apprentissage non supervisé : clustering k-means, classification hiérarchique
- Apprentissage supervisé : Réseaux Bayésiens naïf

### **Domaine d'application en cours : QAI**

# Cartographie qualitative/semi-quantitative de la pollution de l'air

