

DÉVELOPPEMENT D'UN MODÈLE HYBRIDE POUR L'OPTIMISATION DE LA MODÉLISATION DU TRAITEMENT DES EAUX USÉES (2020 – 2023)

Application à la station d'épuration Seine-Aval (Achères) du SIAAP

11 et 12/07/2022

Leesuriales 2022

Marcello Michael Serrao

ÉCOLE DOCTORALE UNIVERSITÉ — PARIS-EST
Sciences, Ingénierie et Environnement



Peter Vanrolleghem



Bruno Tassin



Ilan Juran



Vincent Jauzein & Vincent Rocher

SOMMAIRE

Présentation du projet de recherche

Objectifs

Plan du travail

Contexte

Méthodologie appliquée

Résultats obtenus

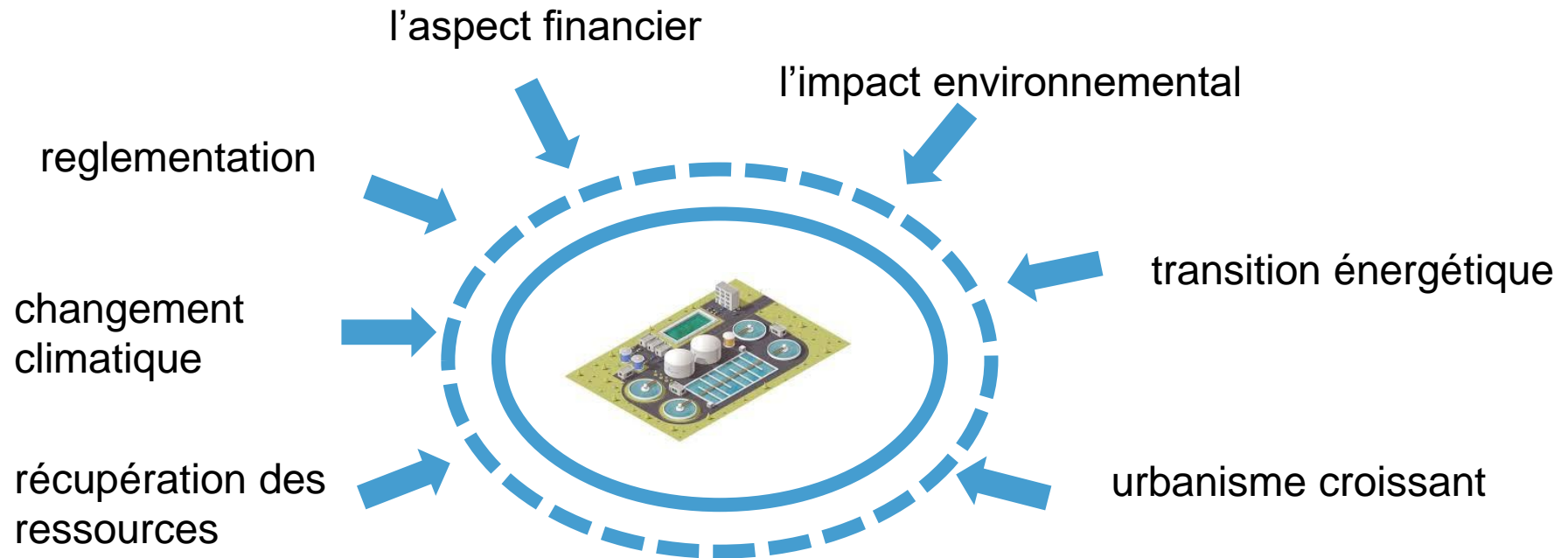
Conclusions & perspectives

01/ MON PROJET DE RECHERCHE

Contexte . Objectifs . Plan du travail

POURQUOI DES MODÈLES ?

- Traitement des eaux usées urbaines est pertinente pour l'environnement et la santé humaine
- La conception et l'exploitation d'une STEP sont souvent (très) conservatrices (avec des marges de sécurité élevées), conduisant à des installations surdimensionnées et à des procédés énergivores

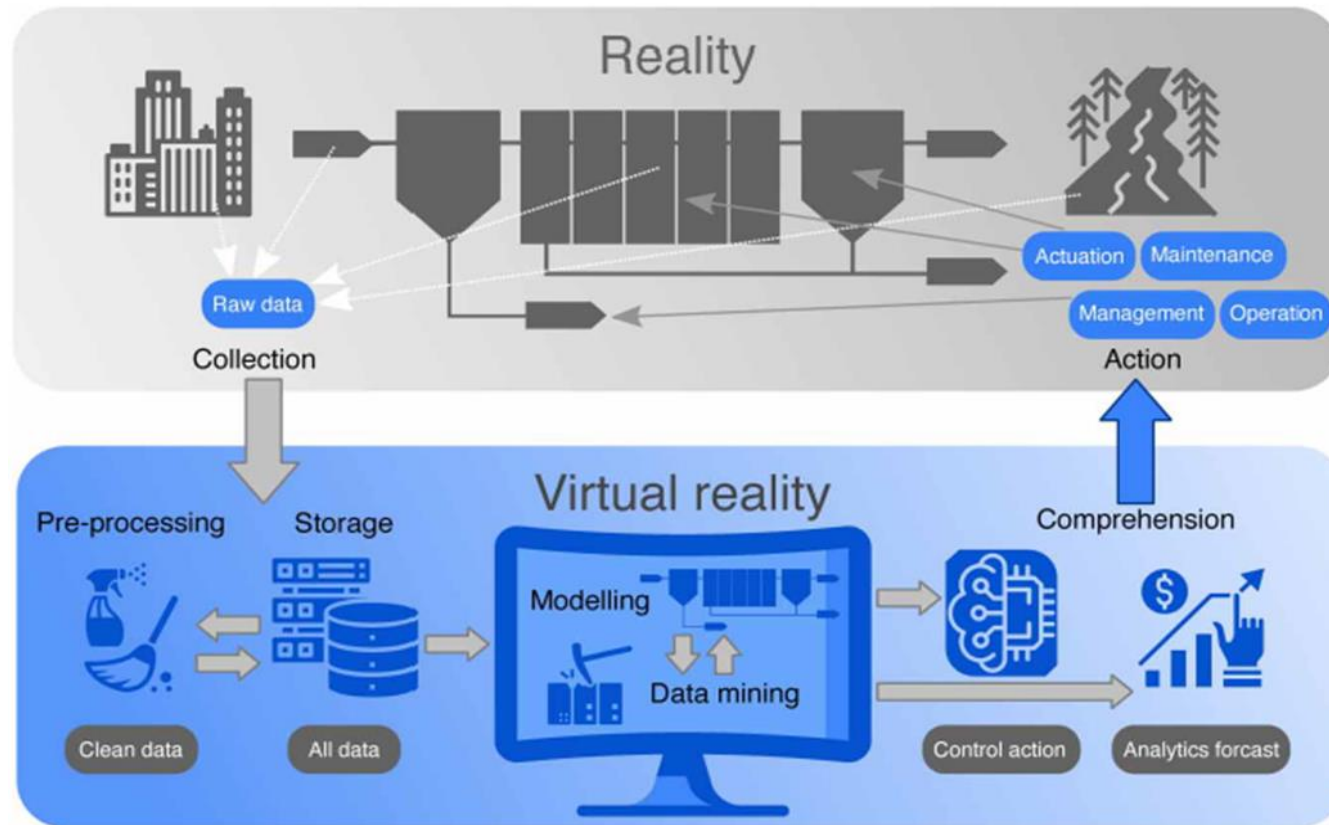


Comment relever ces défis et améliorer la durabilité et la résilience?

POURQUOI DES MODÈLES ?

La modélisation des processus biologiques, chimiques et physiques est utile pour

- acquérir la compréhension des processus
- simuler et tester les stratégies de contrôle
- prédire le comportement futur dans des conditions changeantes
- un contrôle prédictif avec des règles de commande modélisées



POURQUOI DES MODÈLES HYBRIDES?

Pour se faire compléter !

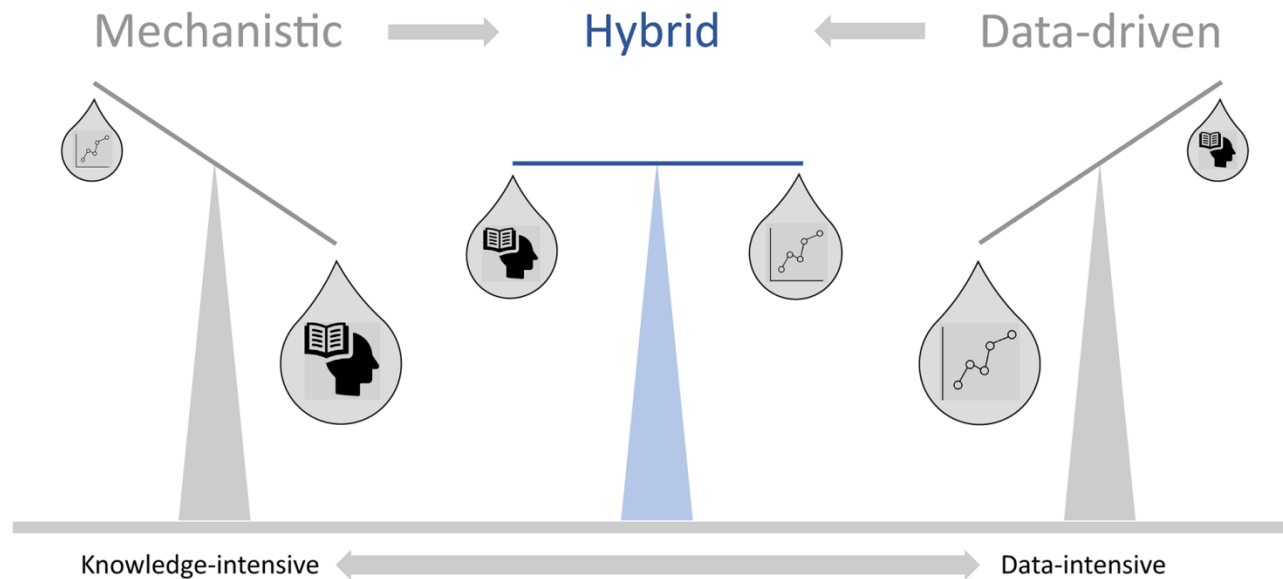
basé sur des connaissances

hypothèses et simplifications

nécessite d'une paramétrage approfondis



Extrapolation & Prédiction



basé sur des schémas de données

digitalisation de l'industrie > 'big data'

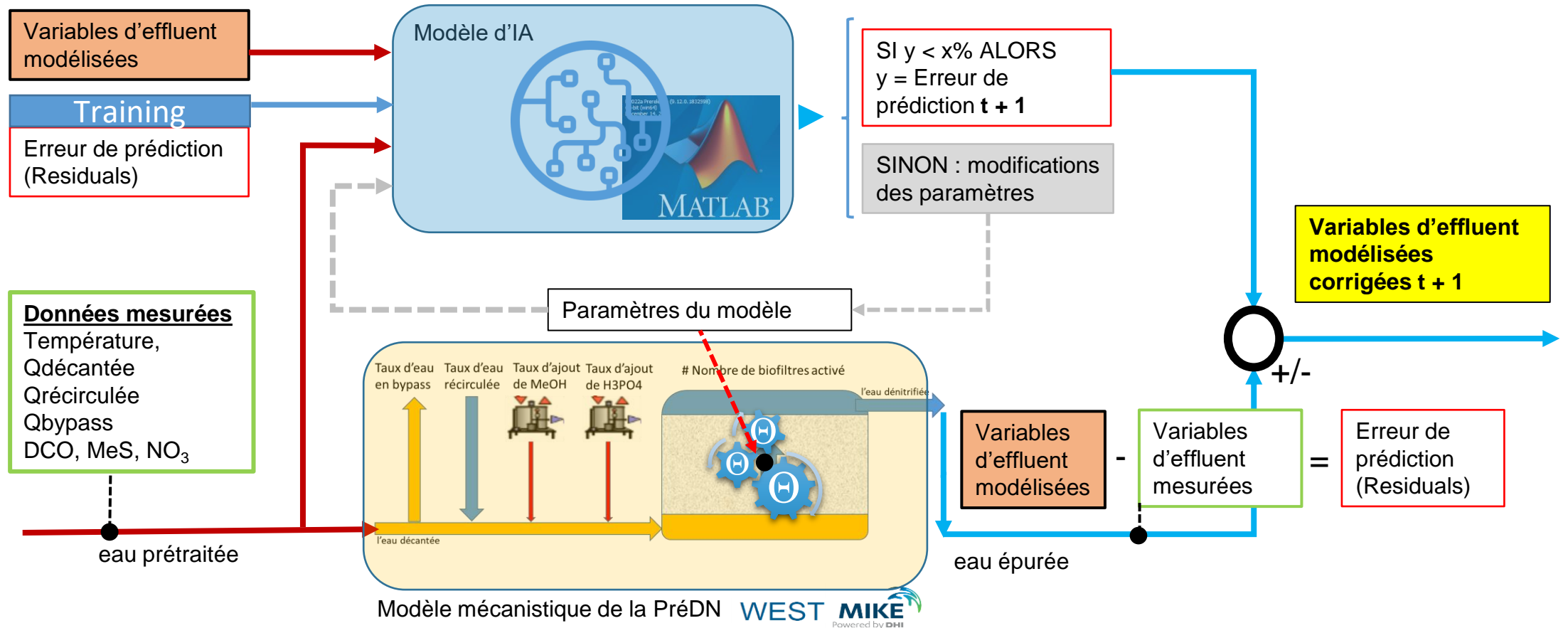
paramétrage rapide



interpolation



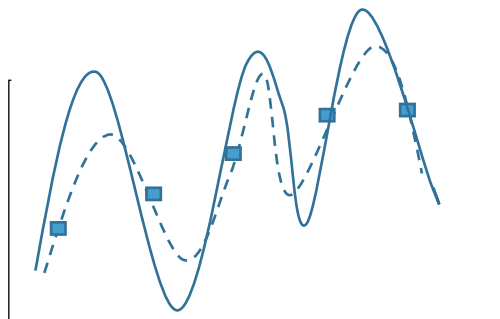
CONCEPT D'OBJECTIF DU PROJET



CONCEPT D'OBJECTIF DU PROJET

Objectif conceptuel

- Données modélisées
- - - Données mesurées par capteur
- Données laboratoire

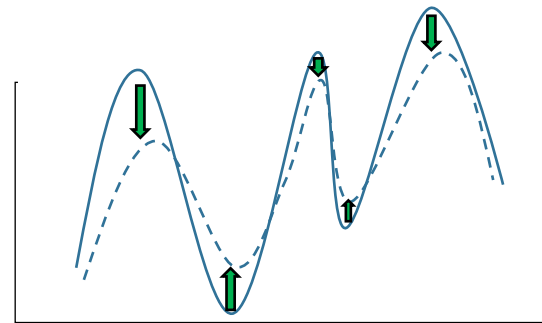


Résultat du modèle mécanistique

ME
MAE
RMSE
R²



+

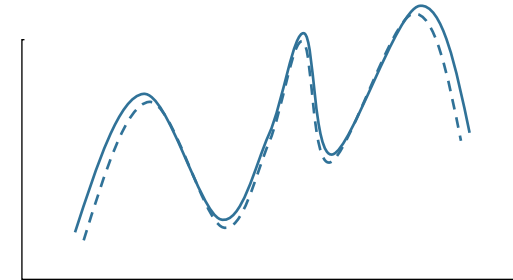


Résultat du modèle d'IA

ME
MAE
RMSE
R²



=



Résultat du modèle hybride

ME
MAE
RMSE
R²



OBJECTIFS & PLANIFICATION

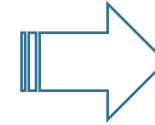
Objectif 1 : modélisation

(Re-)calibration et validation des modèles phénoménologiques / mécanistiques



Objectif 2 : l'IA

Développement et testing des algorithmes d'apprentissage automatique



Objectif 3 final : Jumeau Numérique Adaptatif Hybride

- Effectuer une analyse de la qualité des données des mesures en continu
- Générer des données de modèle d'affluent
- Recalibrer les valeurs des paramètres

« Un modèle hybride peut-il améliorer la précision et la stabilité des prédictions ? »

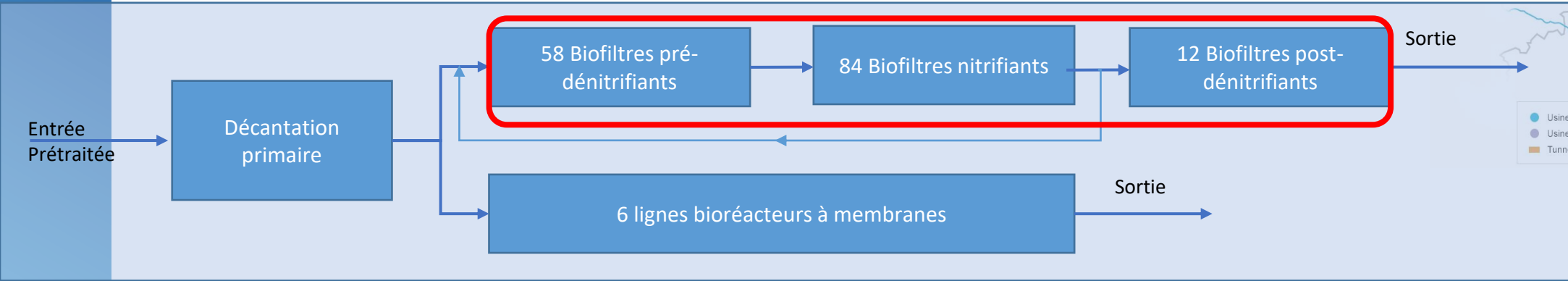
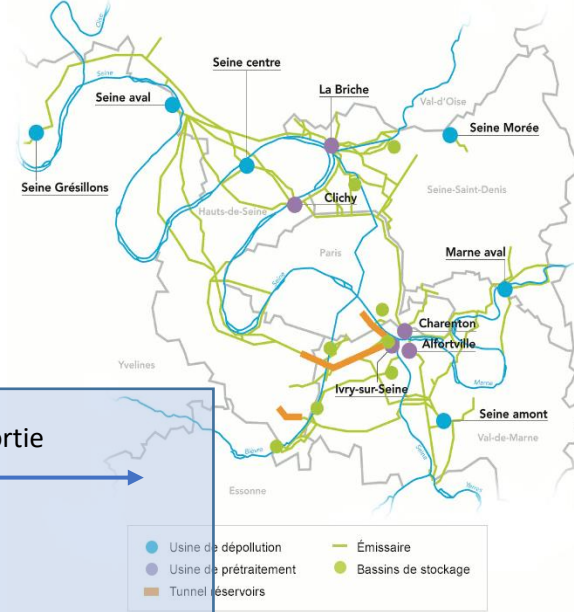
Planification du projet

2020				2021				2022				2023	
3	6	9	12	3	6	9	12	3	6	9	12	3	6
étude de la littérature		Calibration de modèle PrédN		Calibration du modèle filière				Machine Learning sur le modèle PrédN		Modèle Hybride	Ecrire la thèse	Ecrire la thèse	Soutenance de la thèse

SITE D'ETUDE

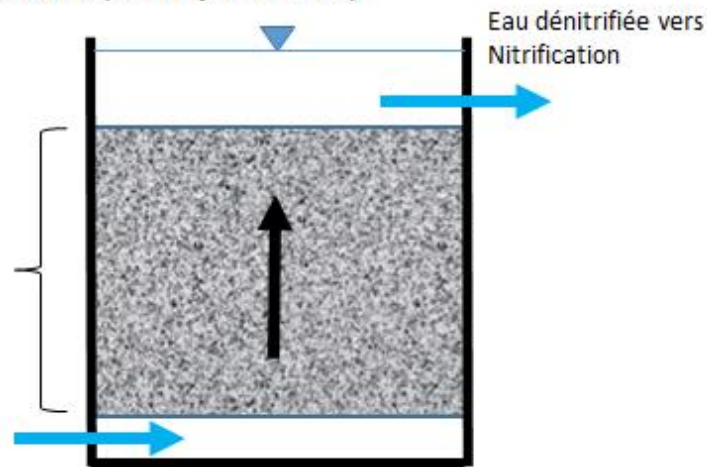
STEP Seine-Aval (SIAAP) Achères : 1,5 million m³/jour (6 million IE)

- 2018 nouveaux biofiltres en service en mode prédénit - nit - postdénit
- données mesurées en ligne et laboratoire

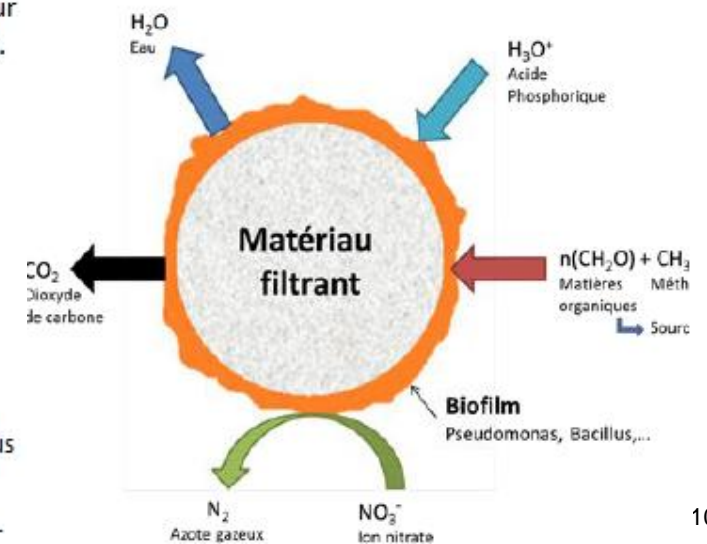
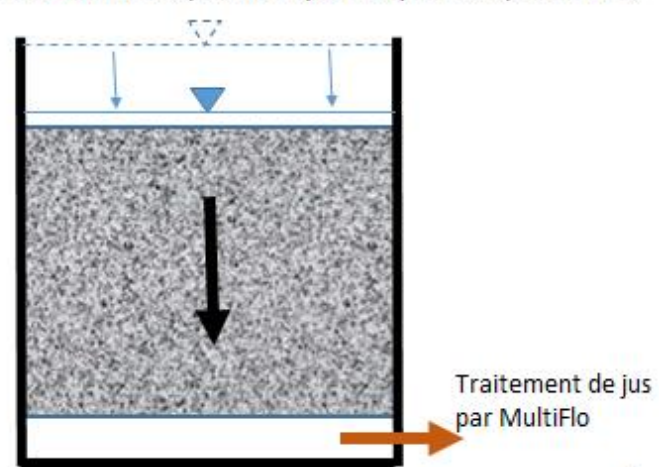


Concept de la biofiltration

La phase de **filtration/traitement biologique** : transformation nitrates $\text{NO}_3^- \rightarrow$ azote gazeux N_2 (sans injection d'air)



La phase de **lavage** : retirer l'excès de biomasse pour maintenir les capacités hydrauliques et épuratoires.



02/ LA MÉTHODOLOGIE

Calibration des modèles mécanistiques

Validation des modèles statistiques

Méthodologie d'un modèle hybride

MODÈLE MÉCANISTIQUE DE BASE

Modèle ASM1 => ASM1JZ (Jialu Zhu, 2020)
5 fichiers d'affluent + 1 série de température

1 biofiltre => 7 réacteurs => 5 couches biofilm (1D)

- Transfert et transport des composants solubles par diffusion -> JZ: couche limite (MTBL) avec épaisseur variable (plus réaliste due au changement de la vitesse de filtration)
- Transfert et transport des composants particulaires par filtration (Ives), détachement et échange
- Mélange de média dans un filtre

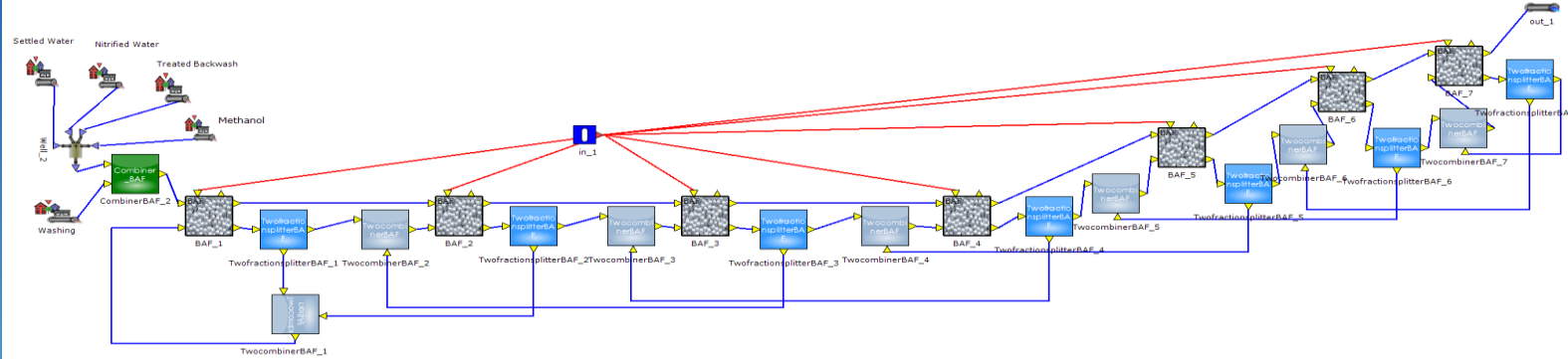
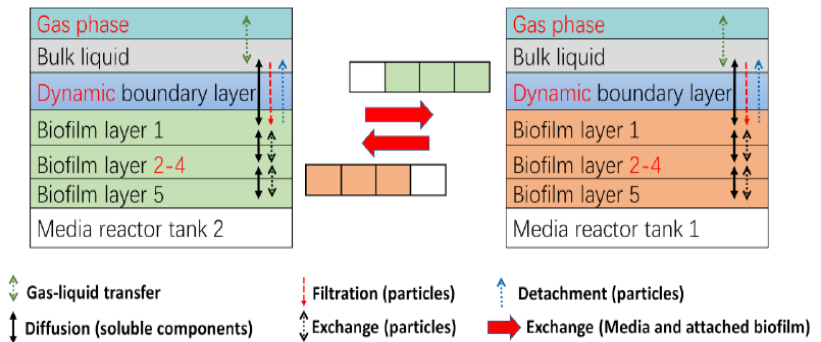


Tableau 4.1 Processus de conversion biologique pris en compte par le modèle développé (les lignes en gris sont les procédés qui étaient déjà considérés dans le modèle SimBio)

j	Processus
1	Croissance aérobie de la biomasse hétérotrophe
2	Croissance anoxique de la biomasse hétérotrophe sur le nitrate
3	Croissance anoxique de la biomasse hétérotrophe sur le nitrite
4	Croissance anoxique de la biomasse hétérotrophe sur l'oxyde nitrique
5	Croissance anoxique de la biomasse hétérotrophe sur l'oxyde nitreux
6	Oxydation de l'ammoniac en hydroxylamine
7	Croissance aérobie de la biomasse nitrifiante
8	Oxydation de l'oxyde nitrique en nitrite
9	Réduction de l'oxyde nitrique en oxyde nitreux
10	Réduction de l'acide nitreux en oxyde nitreux
11	Croissance aérobie de la biomasse nitrifiante
12	Mort de la biomasse hétérotrophe
13	Mort de la biomasse nitrifiante
14	Mort de la biomasse nitrifiante
15	Ammonification
16	Hydrolyse du substrat particulaire biodégradable
17	Hydrolyse de l'azote organique particulaire

Tableau 4.2 Composants solubles et particulaires pris en compte par le modèle développé.

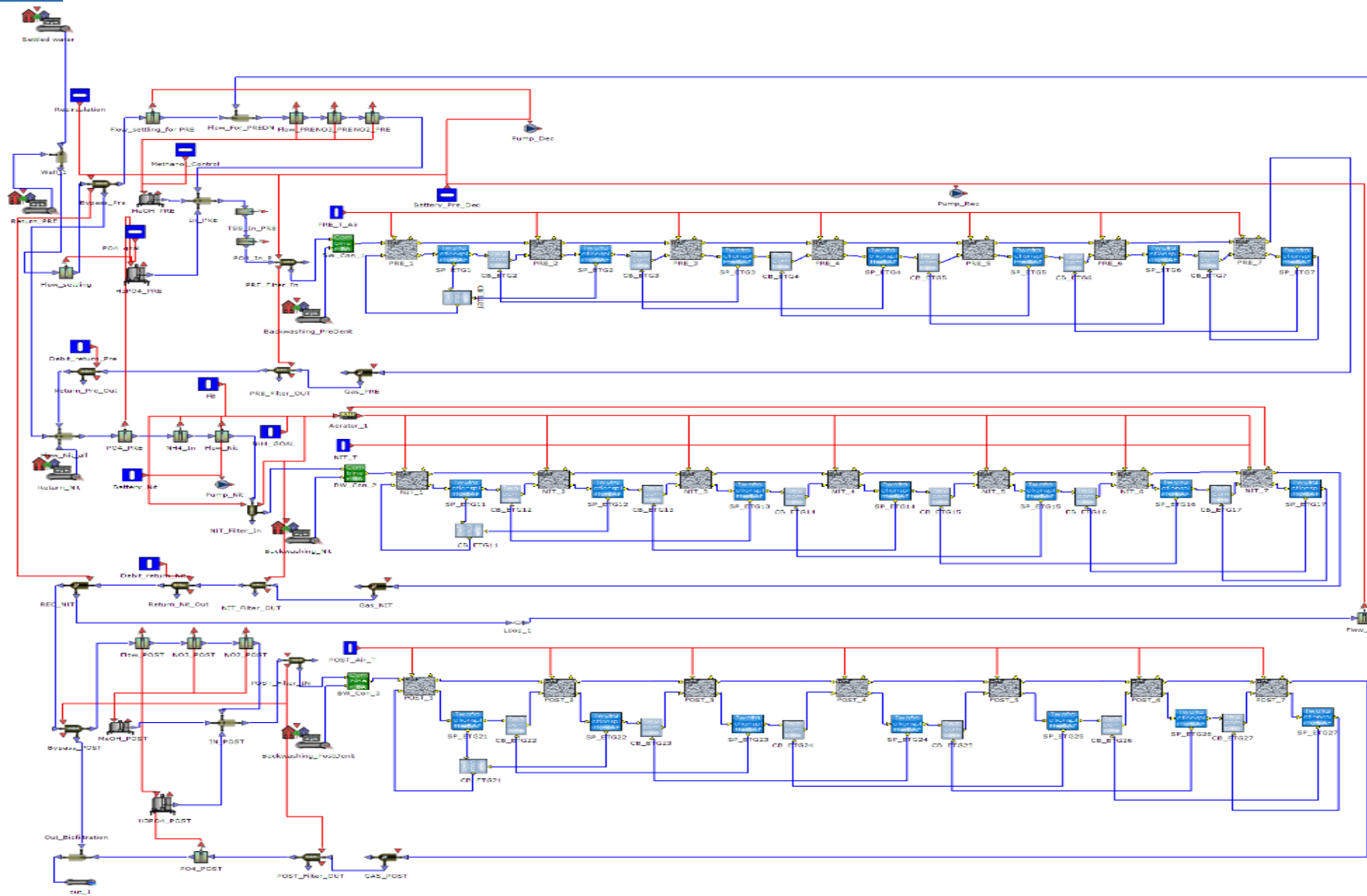
Composants solubles	Symboles	Composants particulaires	Symboles
Alcalinité	S_{ALK}	Azote organique particulaire	$X_{CB,N}$
DCO soluble biodégradable	S_B	Biomasse hétérotrophe	X_{CHO}
DCO soluble inerte	S_U	Biomasse nitrifiante	X_{AOB}
Azote organique soluble	$S_{B,N}$	Biomasse nitrifiante	X_{NOB}
Azote ammoniacal	S_{NH_4}	DCO particulaire biodégradable	X_{CB}
Hydroxylamine	S_{NH_2OH}	DCO particulaire inerte	X_U
Nitrate	S_{NO_3}	DCO particulaire inerte issue de la mort cellulaire	X_{UE}
Nitrite	S_{NO_2}		
Oxyde nitrique	S_{NO}		
Oxyde nitreux	S_{N_2O}		
Diazote dissous	S_{N_2}		
Oxygène dissous	S_{O_2}		
Orthophosphate	S_{PO_4}		



RÉSULTAT D'OBJECTIF 1.2 - (RE)CALIBRER LE MODÈLE BIOFILTRATION COMPLÈTE (PRÉDN + NIT + POSTDN)

Ce modèle comprend :

- les paramètres calibrés du modèle PréDN (de l'objectif 1.1)
- les paramètres de control recalculés pour la période 20190112 - 20191231



Modèle de biofiltration en 3 étapes successives comprenant la recirculation de l'eau nitrifiée, l'ajout des réactifs et d'air et les dérivations hydrauliques :

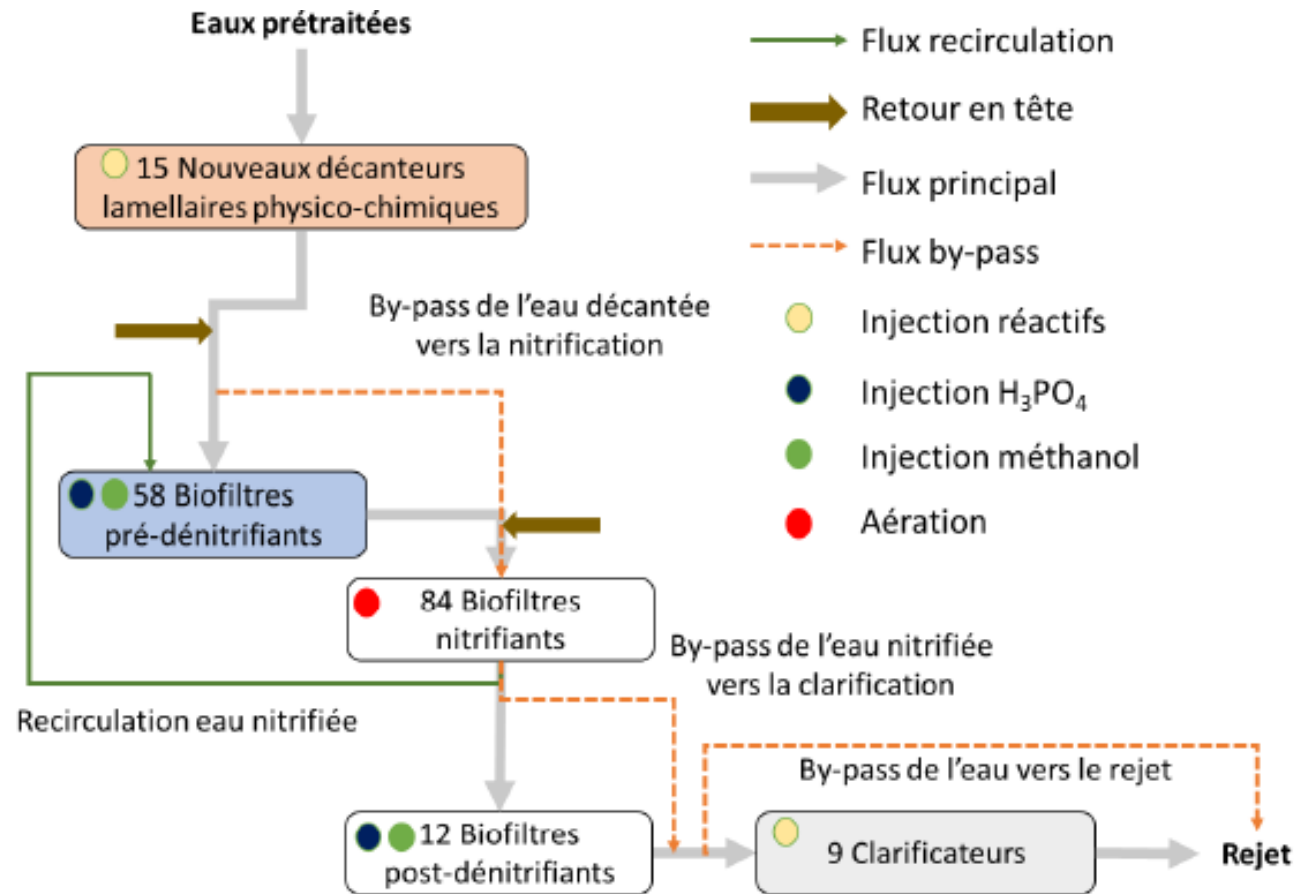
Dénitification amont (DAM)

Nitrification

Dénitification aval (DAV)

RÉSULTAT D'OBJECTIF 1.2 - (RE)CALIBRER LE MODÈLE BIOFILTRATION COMPLÈTE (PRÉDN + NIT + POSTDN)

Les commandes de procédées dans le modèle



Pré-dénitrification	By-pass de l'eau décantée vers la nitrification	m^3/j
	Recirculation de l'eau nitrifiée	m^3/j
	Nombre de biofiltres actifs	-
	Taux d'injection méthanol	$g/gN-NO_x$
Nitrification	Taux d'injection H_3PO_4	g/m^3
	Nombre de biofiltres actifs	-
	Débit d'air injecté	Nm^3/h
Post-dénitrification	By-pass de l'eau nitrifiée vers la clarification tertiaire	m^3/j
	Taux d'injection méthanol	$g/gN-NO_x$
	Taux d'injection H_3PO_4	g/m^3

PROTOCOL POUR LA CALIBRATION

IWA Guidelines for using Activated Sludge Models

Bonnes pratiques de la modélisation (IWA) :

Étape 1. Définition du projet

Étape 2. Collecte et rapprochement des données

-> Alfères & Vanrolleghem (2016)

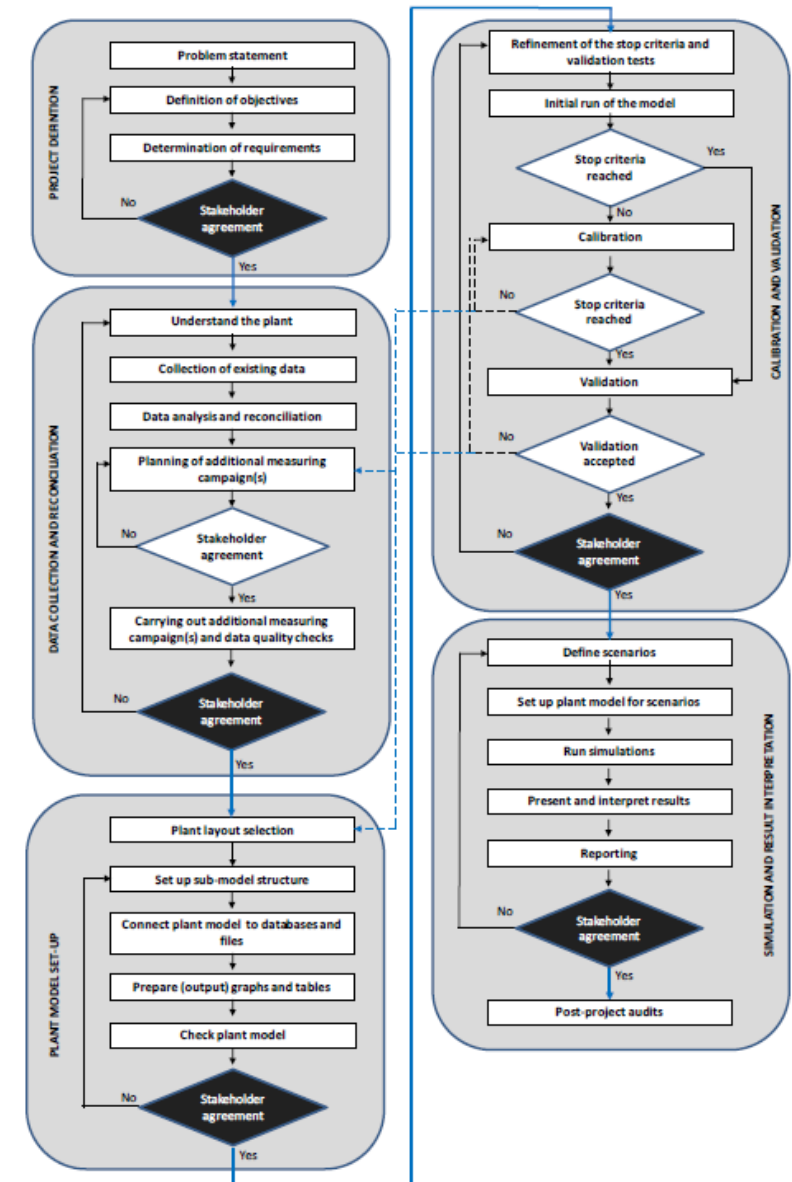
Étape 3. Configuration du modèle d'usine

Étape 4. Calibration et validation du modèle

-> Mannina et al 2011

Étape 5. Simulation et interprétation des résultats

The GMP Unified Protocol



ETAPE 2 - (RE)CALIBRER UN MODÈLE

Traitement des données cfr. Alfères et Vanrolleghem (2016)

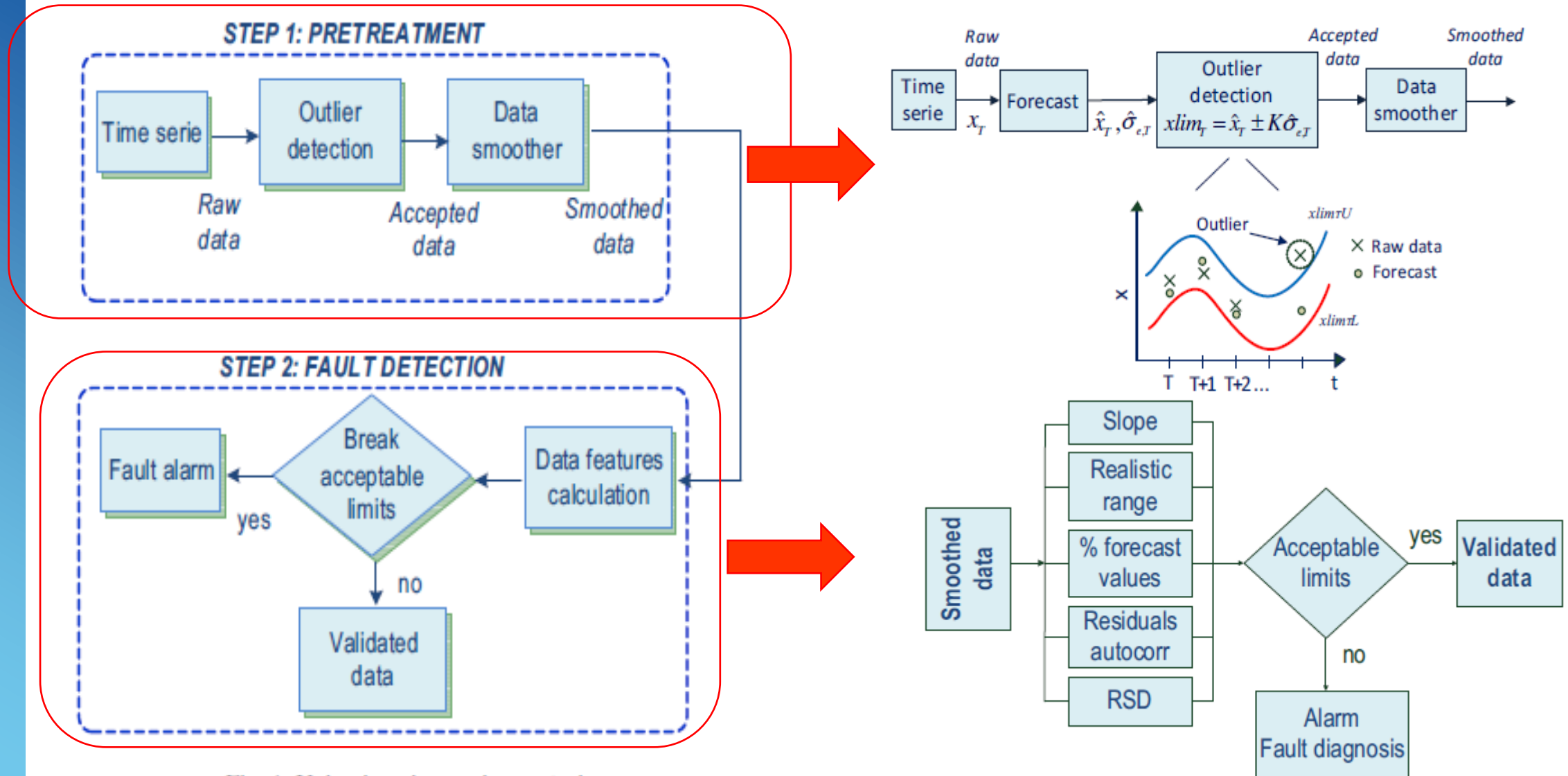
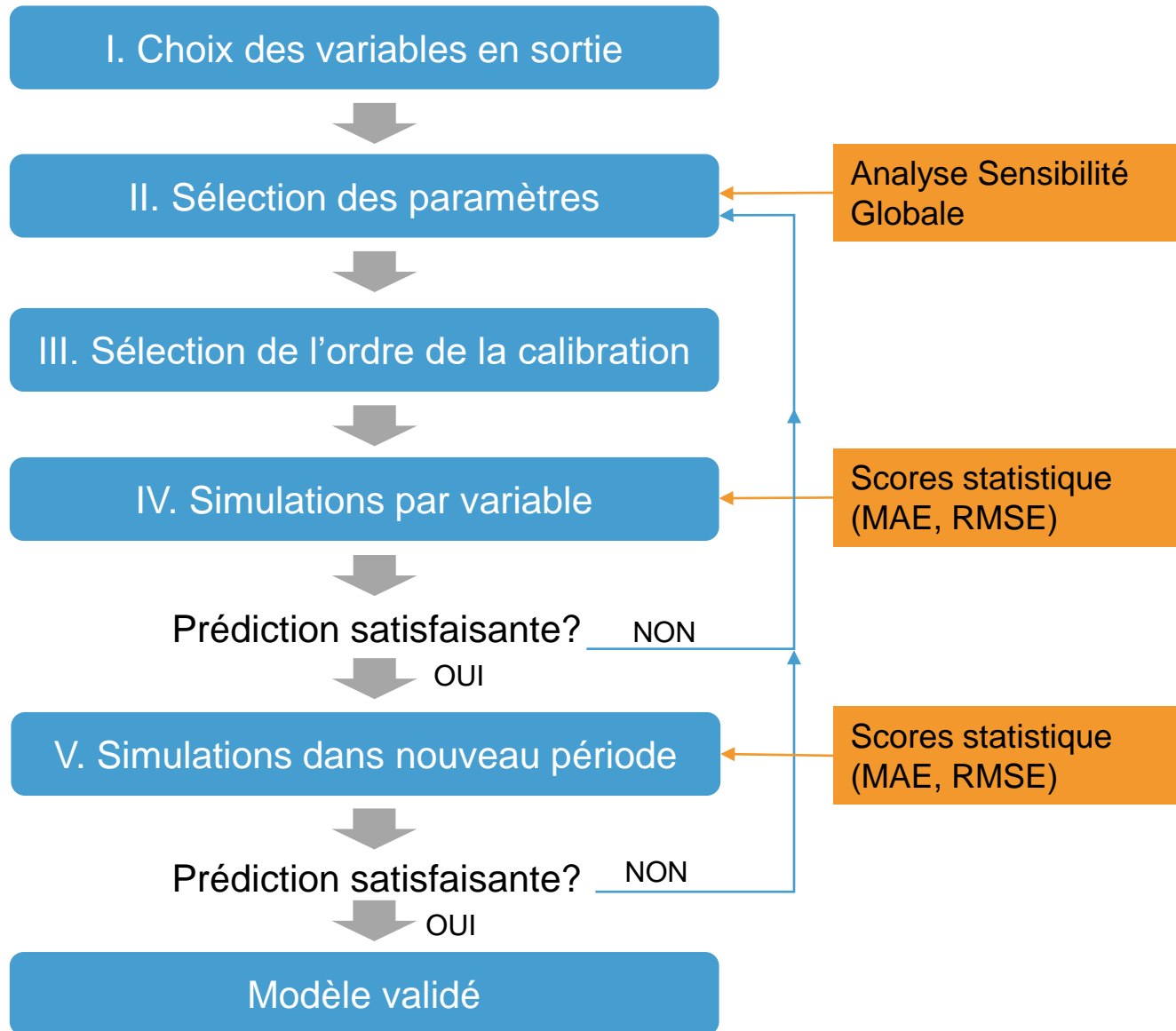


Fig. 1. Univariate time series analysis

ETAPE 4 - (RE)CALIBRER UN MODÈLE

Méthodologie de la calibration d'un modèle phénomologique cfr Mannina et al. (2011)



Calibration (étapes I. -> IV.) : ajuster les valeurs des certains paramètres **uniques pour une variable** afin que les résultats de simulation soient les plus précises aux observations faites.

Validation (étape V.) : une vérification de la performance après calibration avec des jeux de données différents

L'OBJECTIF 2 - DÉVELOPPEMENT DES ALGORITHMES DE MACHINE LEARNING

Méthodologie générique d'entraînement d'un modèle d'IA

Preparation de données ('preprocessing')

1. Par variable d'effluent individuelle (= univariate) : NOx, DCO, MeS
2. Data preprocessing :
 - i. Data cleaning, outliers, smoothing cfr Alferes & Vanrolleghem 2016
 - ii. Data filling or removing NaN
 - iii. Normalization (range 0 -> 1)
 - iv. Diviser l'ensemble de données en jeux de validation (80%) et testing (20%)

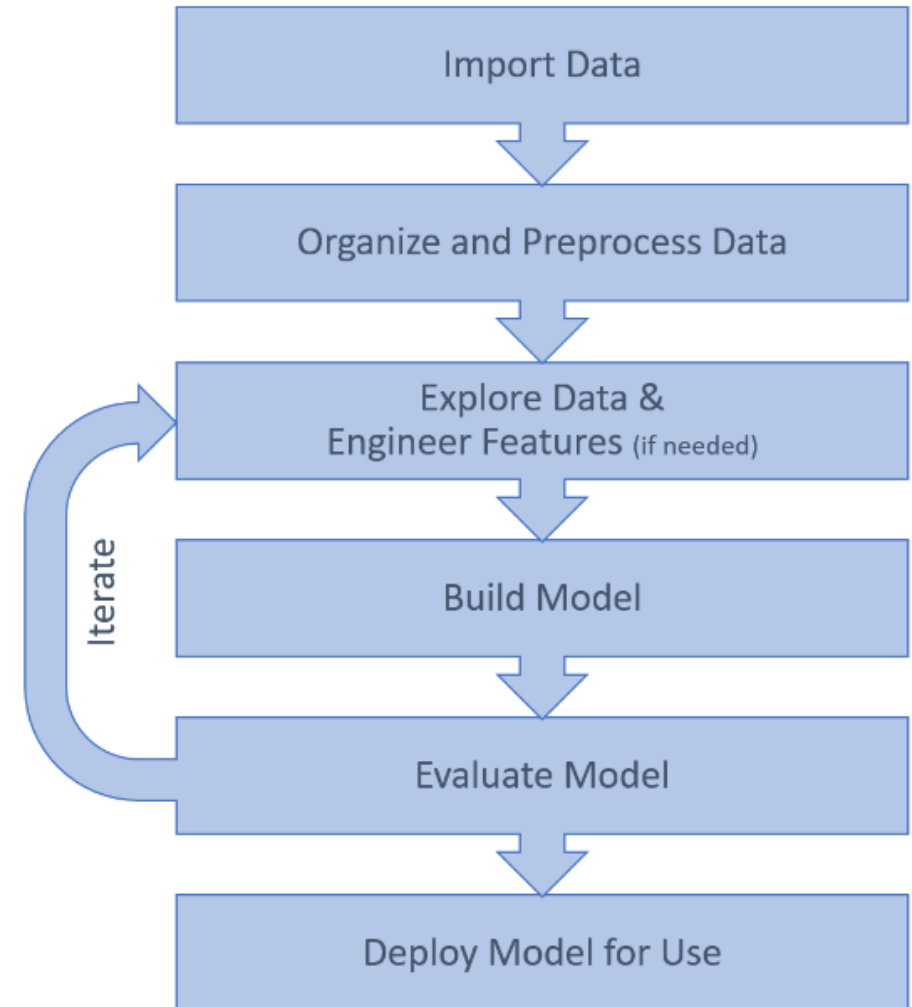
Features Selection & reduction ('extraction')

1. Testing of different ML models on the features
2. Evaluation of performance (RMSE, MAE)
3. Transform Features with PCA
 - PCA analysis to reduce dimensionality of components (95%)

Build & Evaluate model

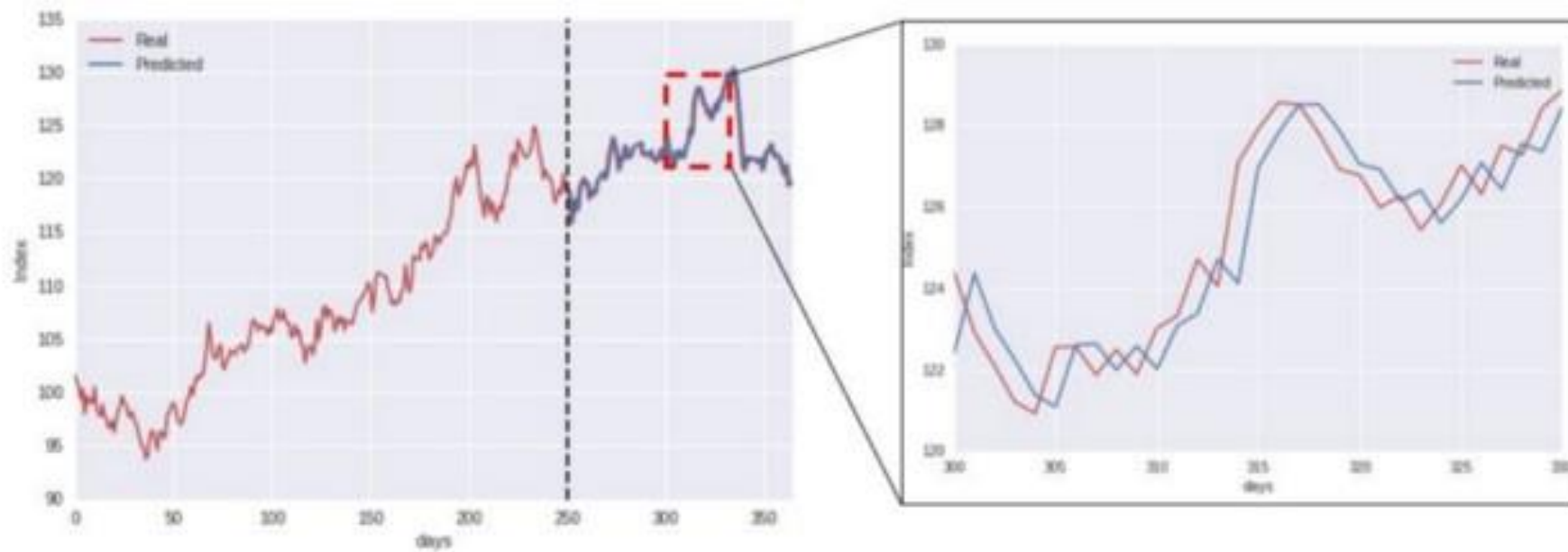
- Compare RMSE, R^2
- Hyperparameter optimisation

Check for correlations between output variables; if so => multivariate analysis



Attention aux données fournies aux algorithmes

Check for auto-correlations !!

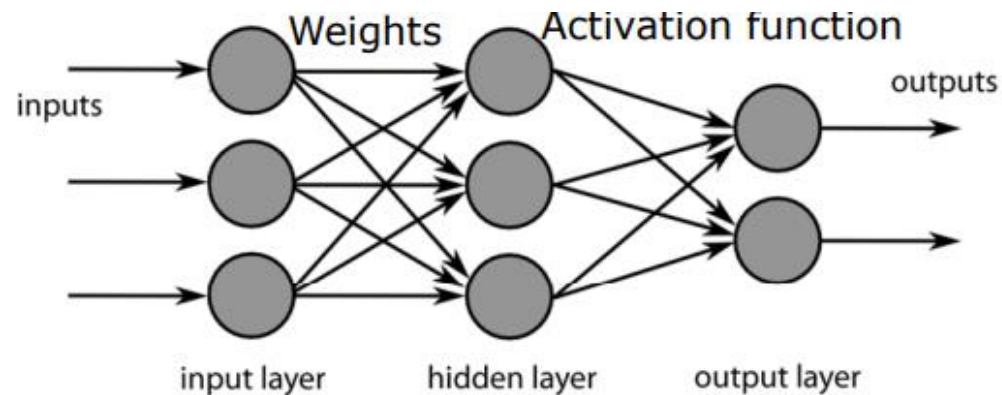


Exemple d'un auto-correlation (1)

(1) <https://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424>

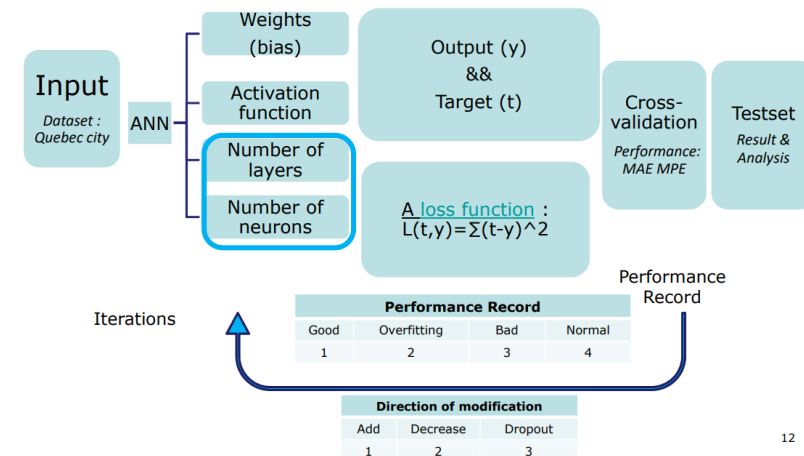
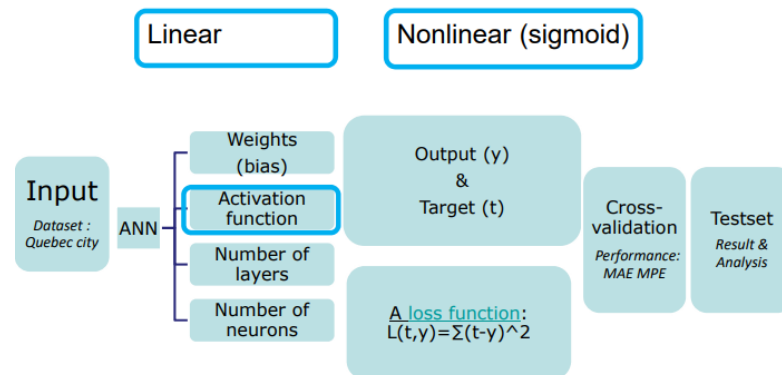
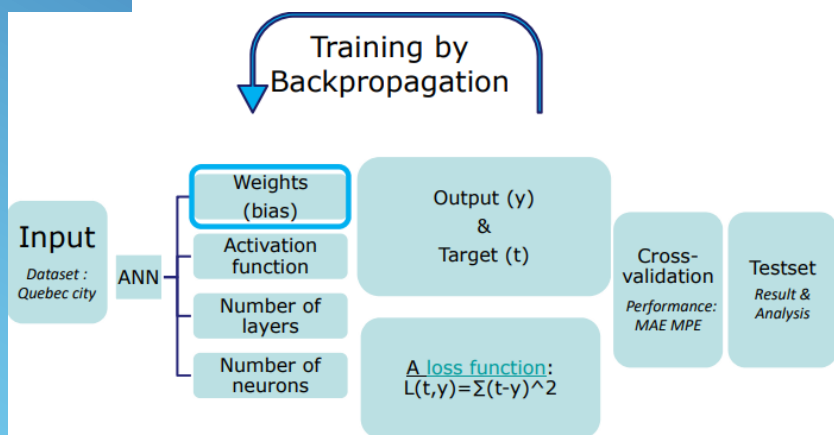
(2) <https://machinelearningmastery.com/persistence-time-series-forecasting-with-python/>

L'OBJECTIF 2 - DÉVELOPPEMENT DES ALGORITHMES DE MACHINE LEARNING



Li et al., 2019

Tuning an ANN with Hyper Parameters (Li et al., 2019)



12

MÉTHODOLOGIE D'UN MODÈLE HYBRIDE

Formes de l'architecture

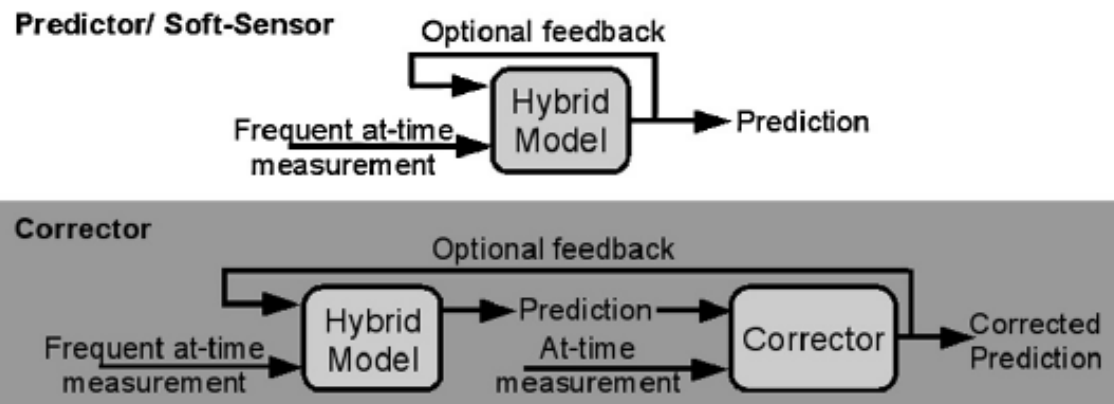
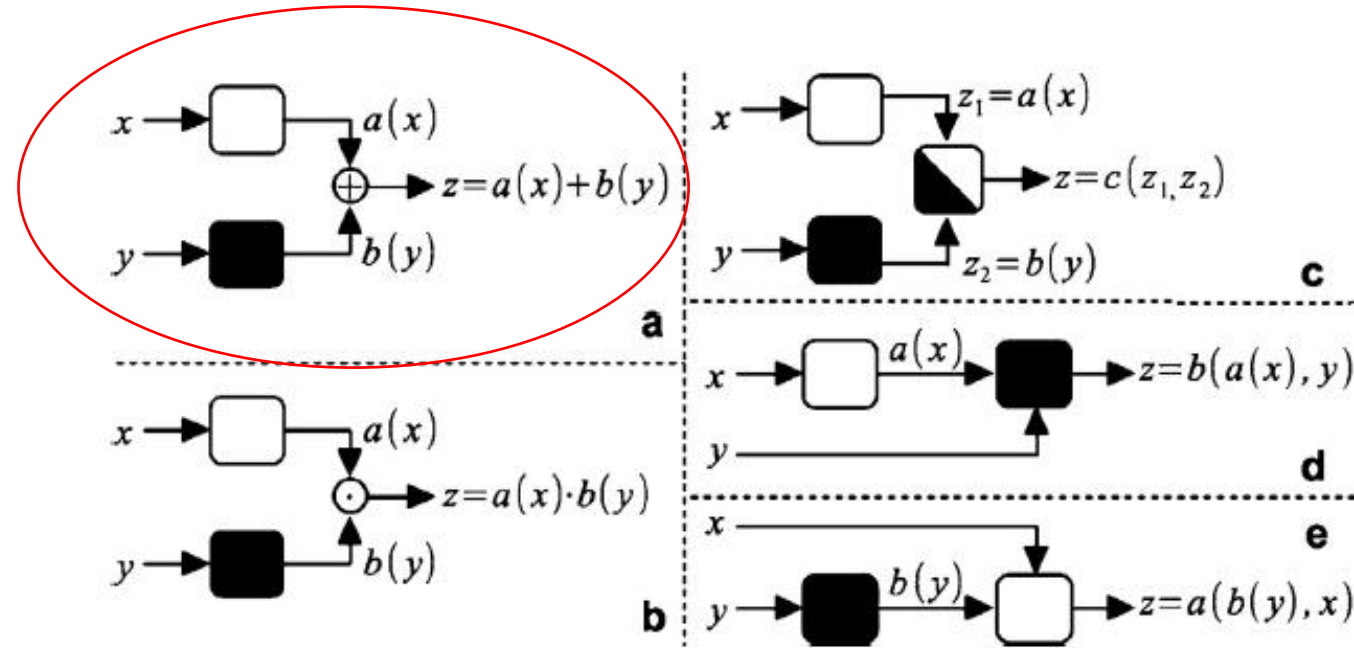


Fig. 7. Diagram of two possibilities to use hybrid semi-parametric modeling for monitoring.

Fig. 4. Schematic representation of white box and black box model combination schema. (a) Superposition; (b) multiplication; (c) weighting of the predictions of the same quantity by using either a black or a white box model; (d) weighting of the white box model predictions using a black box model; (e) weighting of the black box model predictions using a white box model.

STRATÉGIE

Du développement d'un modèle hybride pour l'élimination de l'azote par dénitrification

● Etape 1

Calibration et validation des paramètres du modèle WEST pour la PrédN à la SAV pour 2019 - 2020



● Etape 2

Estimation des erreurs de prédiction du modèle WEST



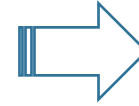
● Etape 3

Entraînement des algorithmes de *Machine Learning* à prédire les erreurs de prédiction du modèle WEST par uni-variable



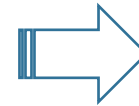
● Etape 4

Intégration des résultats du modèle WEST et du modèle de *Machine Learning* dans un modèle hybride



● Résultats:

- ❑ Variables en sortie pour NO3, NO2, NH4, MeS, DCO, DCOs, PO4



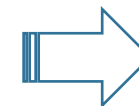
● Résultats:

- ❑ E_pred_NO3
- ❑ E_pred_NO2
- ❑ E_pred_NH4
- ❑ E_pred_MeS
- ❑ E_pred_DCO
- ❑ E_pred_DCOs
- ❑ E_pred_PO4



● Résultats:

- ❑ E_pred_NO3
- ❑ E_pred_NO2
- ❑ E_pred_NH4
- ❑ E_pred_MeS
- ❑ E_pred_DCO
- ❑ E_pred_DCOs
- ❑ E_pred_PO4



● Résultats:

- ❑ NO3_corrigée
- ❑ NO2_corrigée
- ❑ NH4_corrigée
- ❑ MeS_corrigée
- ❑ DCO_corrigée
- ❑ DCOs_corrigée
- ❑ PO4_corrigée

03/ LES RÉSULTATS

Calibration des modèles mécanistiques

Validation des modèles statistiques

Testing d'un modèle hybride

ETAPE 2 - (RE)CALIBRER UN MODÈLE

Traitement des données cfr. Alfères et Vanrolleghem (2016)

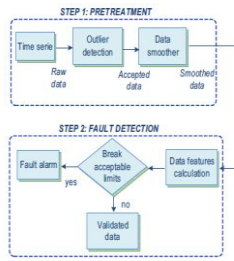
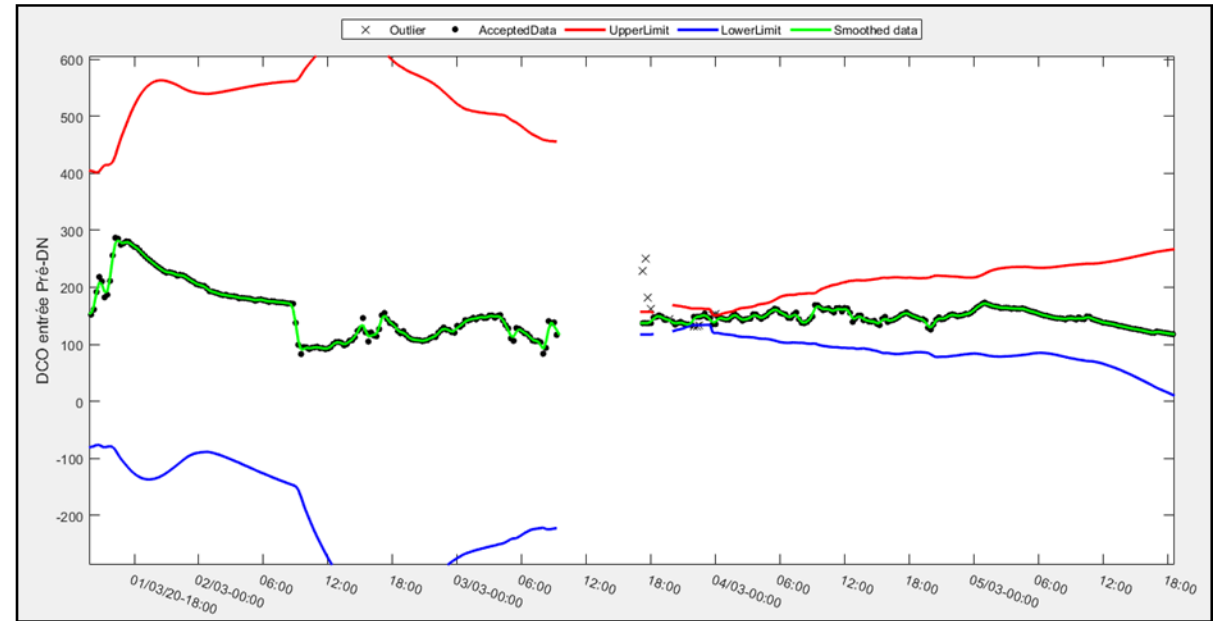
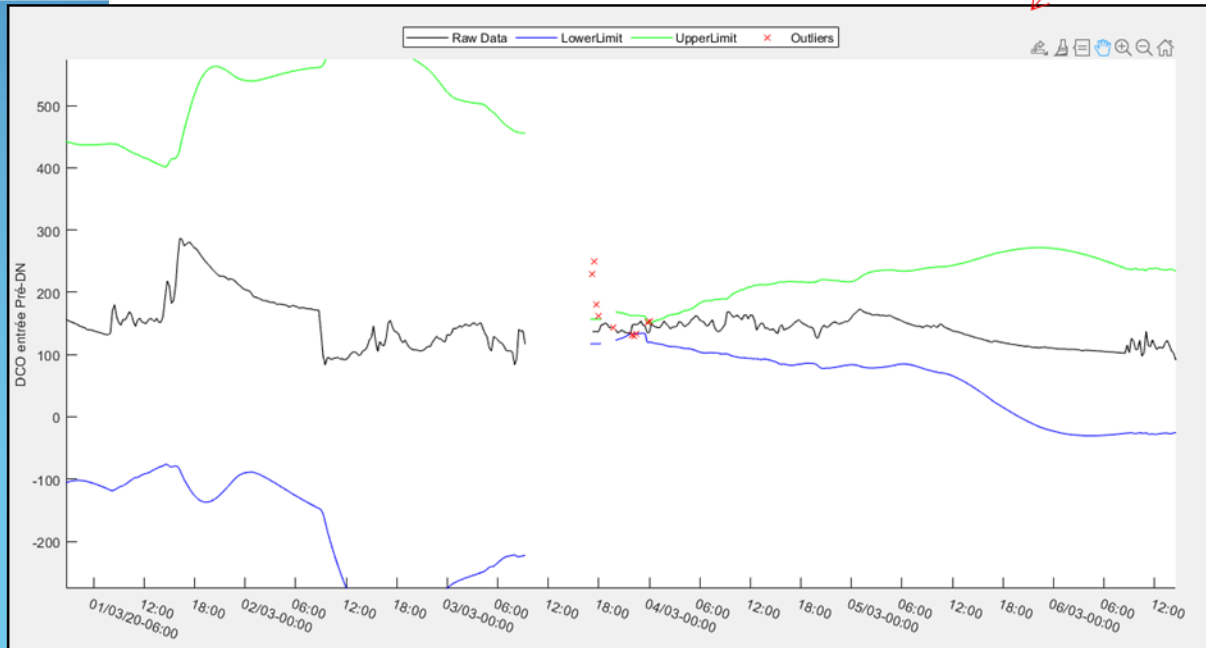
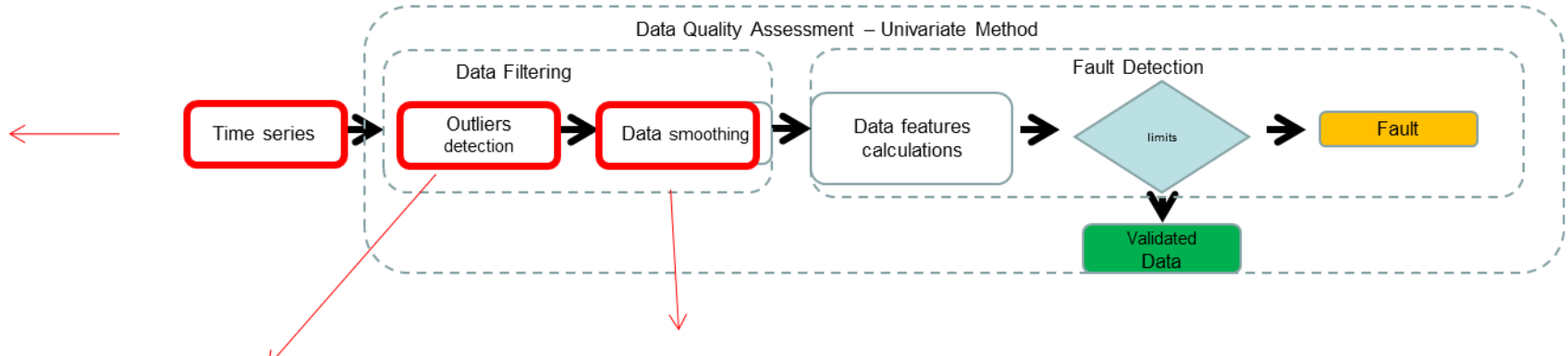
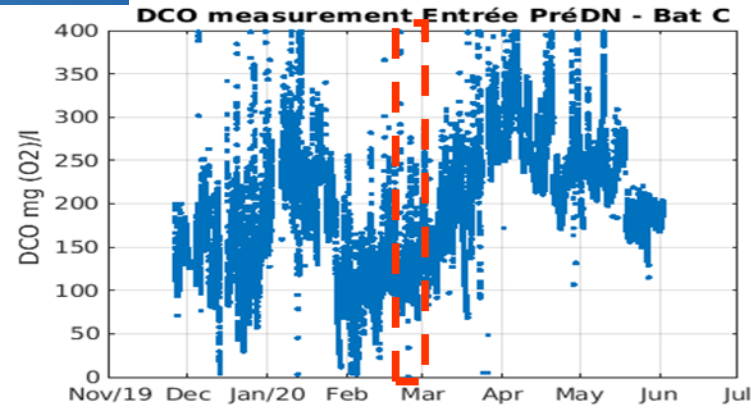
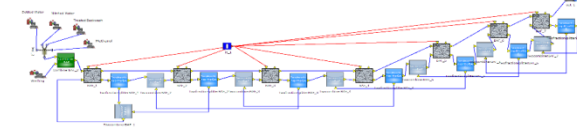


Fig. 1. Univariate time series analysis



Alfères, J., & Vanrolleghem, P. A. (2016). Efficient automated quality assessment: Dealing with faulty on-line water quality sensors. AI Communications, 29(6), 701-709.

RÉSULTAT D'OBJECTIF 1.1 - (RE)CALIBRER LE MODÈLE POUR LA PRÉDN DE LA SAV



L'analyse de sensibilité globale GSA cfr Maninna et al (2011)

- 140 => 89 sélectionnés pour la GSA => 28 paramètres retenues
- 900 Monte Carlo simulations en mode stationnaire avec Latin Hypercube Sampling par variable en sortie : DCO, DCOs, NO3, NO2, NH4, TKN, PO4, MeS
- effectuer une régression linéaire multiple pour identifier les paramètres les plus pertinents
- calibrer les paramètres retenues par variable en sortie dans l'ordre : MeS, (NO3 + DCOs), NO2, NH4, PO4

Tableau avec les paramètres les plus pertinents

PM	Group	Rang	Valeur calibrée	Valeur de Jialu Zhu
lami	TSS	1	0,4	0,6
F_TSS_COD	TSS	1	0,7	0,8
xx	TSS	1	0,6	1
yy	TSS	1	1,02	0,83
zz	TSS	1	1,2	1
rho	TSS	1	180000	133600
WashRate	TSS	1	0,63	0,55
theta_mu_H	CODs + NO3	2	1,082	1,072
n_h1	CODs + NO3	2	0,28	0,28
K_S_NO3	CODs + NO3	2	41,37	28,59
k_h	CODs + NO3	2	2,6	2,5
mu_H	CODs + NO3	2	4,74	5,18
theta_k_h	CODs + NO3	2	0,978	1,116
n_hy	CODs + NO3	2	0,72	0,87
Diffusion_Reduction	CODs + NO3	2	0,55	0,7
K_NO3_H	CODs + NO3	2	0,16	0,2
theta_k_a	SNH	4	1,194	1,072
theta_b_H	SNH	4	1,03	1,12
k_a	SNH	4	0,065	0,075
i_X_B	SNH	4	0,109	0,098
b_H	SNH	4	0,73	0,7
n_h2	NO2	5	0,19	0,14
K_S_NO2	NO2	5	16,06	13,78
K_NO2_H	NO2	5	0,29	0,2
n_h3	NO2	5	0,31	0,47
K_OH_NO3	NO2	5	0,1	0,1
i_X_BP	PO4	6	0,0046	0,0055
i_X_UP	PO4	6	0,008	0,006

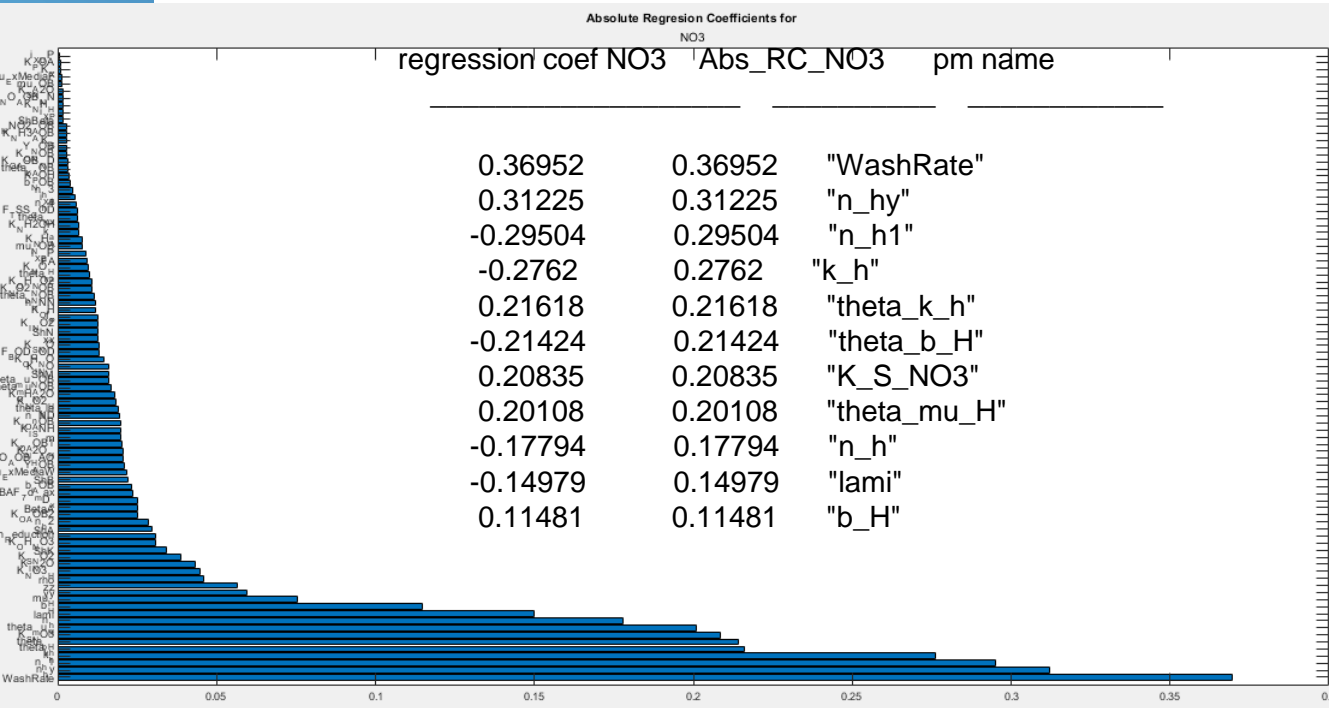
La filtration

Croissance de la biomasse & Consommation de substrat S_NO3

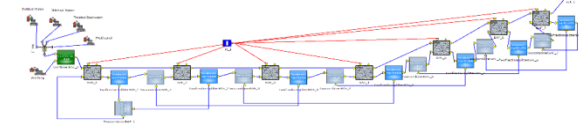
L'ammonification

Consommation de substrat S_NO2

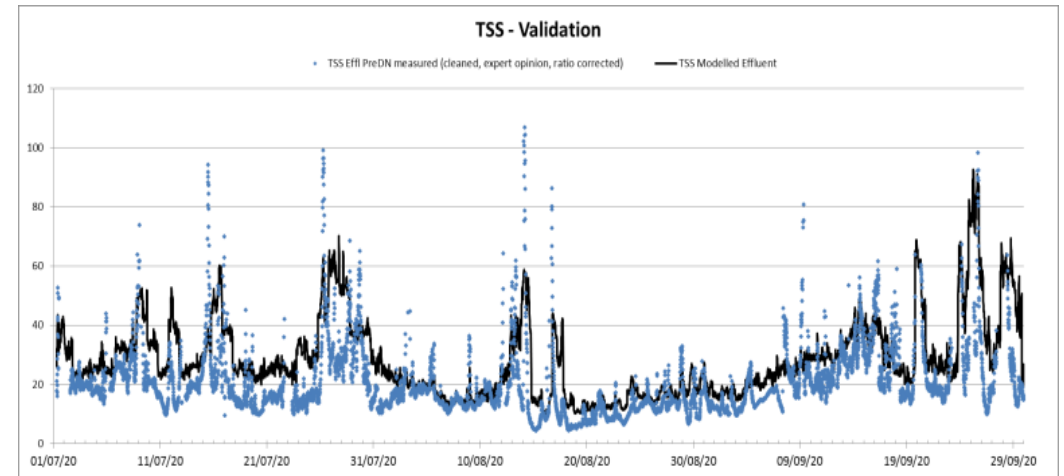
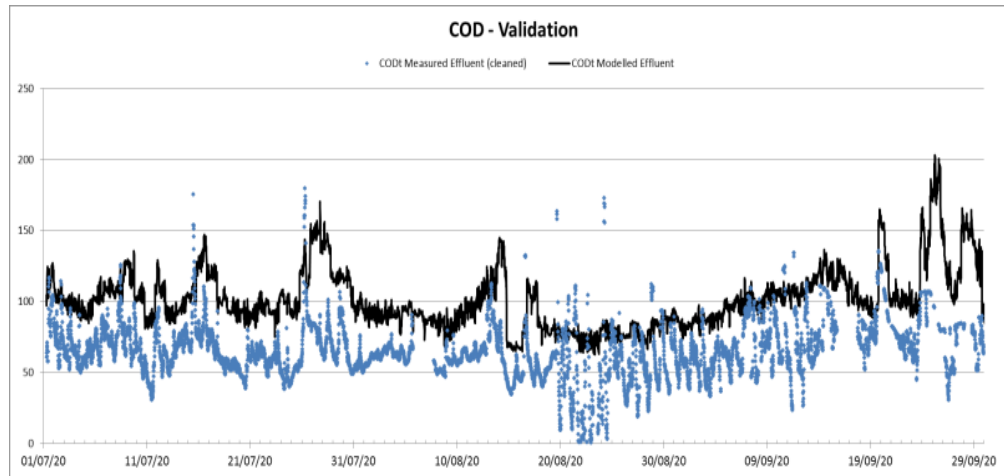
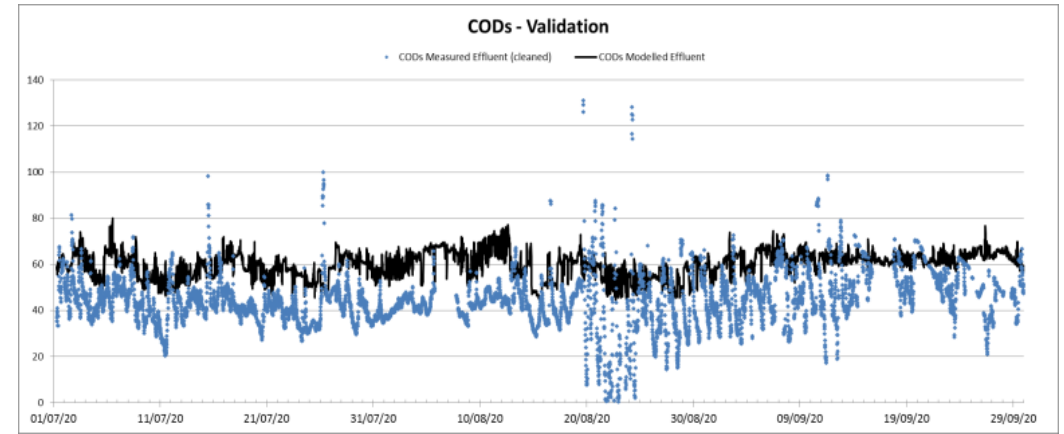
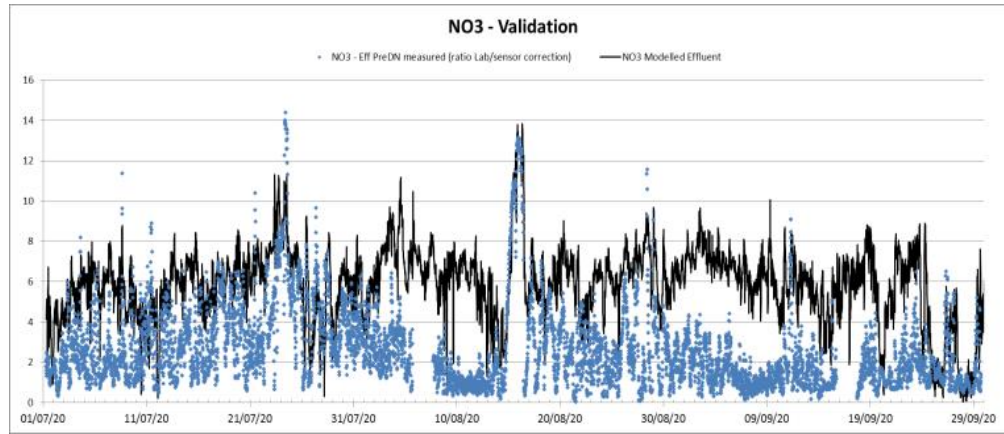
L'absorption de phosphores



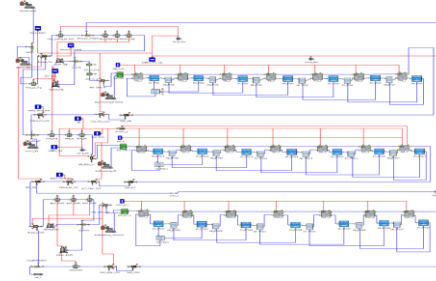
RÉSULTAT D'OBJECTIF 1.1 - (RE)CALIBRER LE MODÈLE POUR LA PRÉDN DE LA SAV



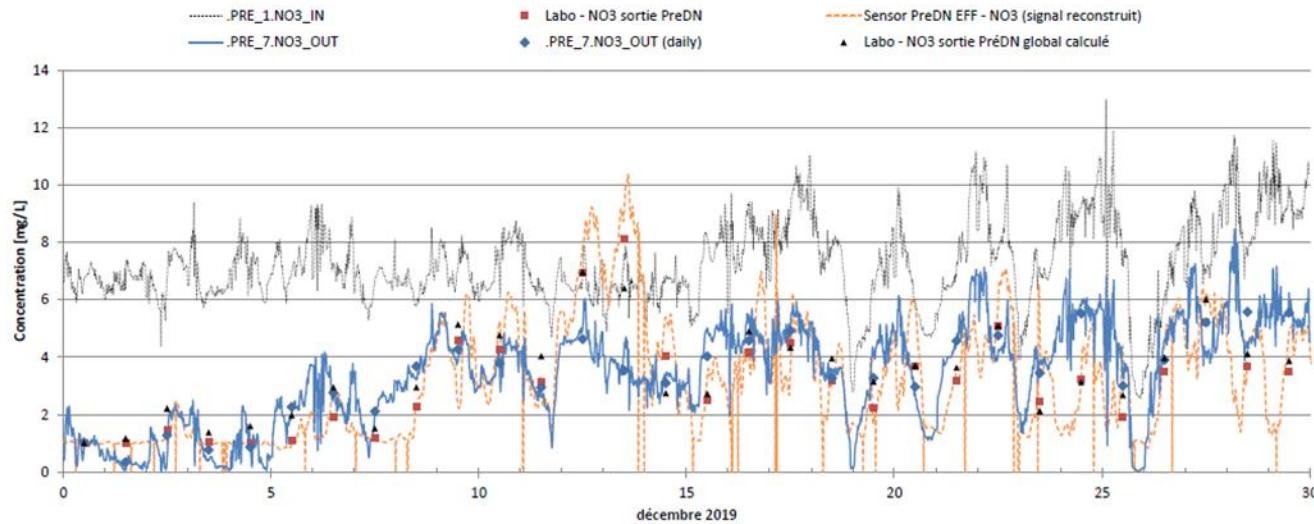
Validation – période 01/07/2020 – 01/10/2020 (= 92 jours)



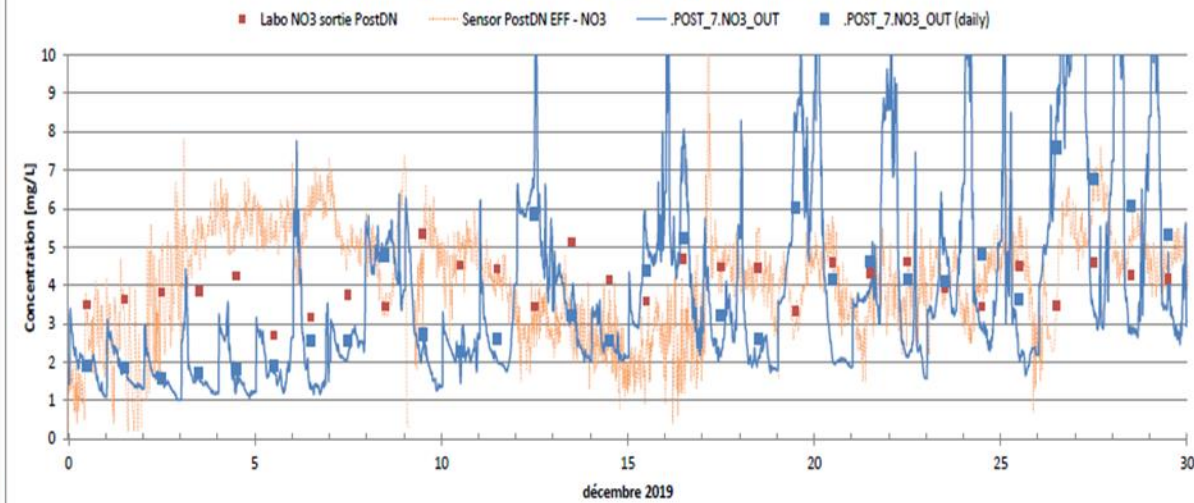
RÉSULTAT D'OBJECTIF 1.2 - (RE)CALIBRER LE MODÈLE BIOFILTRATION (PRÉDN + NIT + POSTDN)



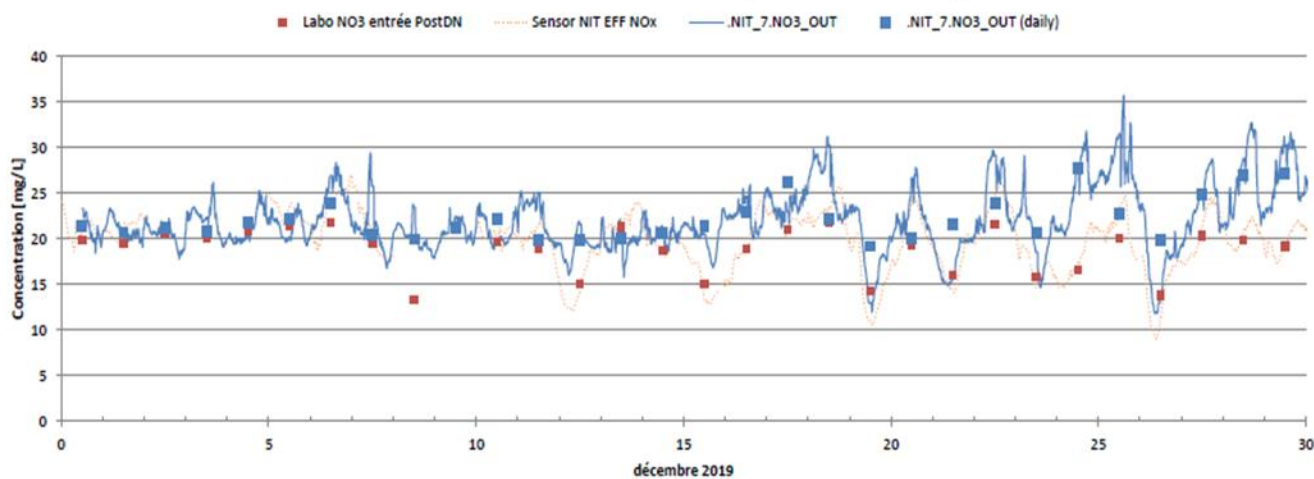
NO3 - Effluent PréDN modèle Biofiltration (01122019 - 31122019)



NO3 - Effluent PostDN modèle Biofiltration (01122019 - 31122019)



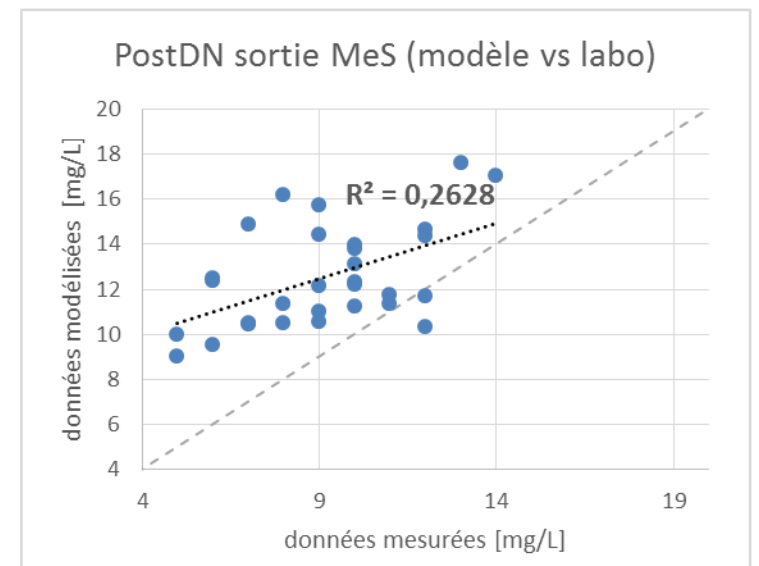
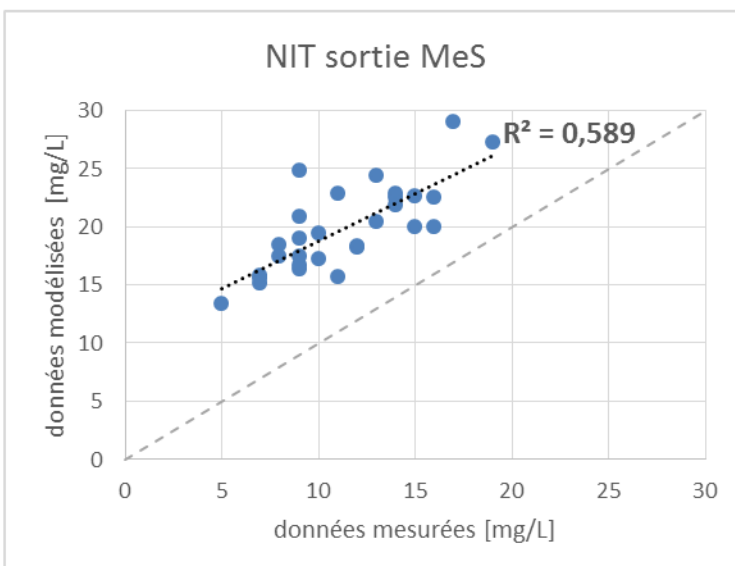
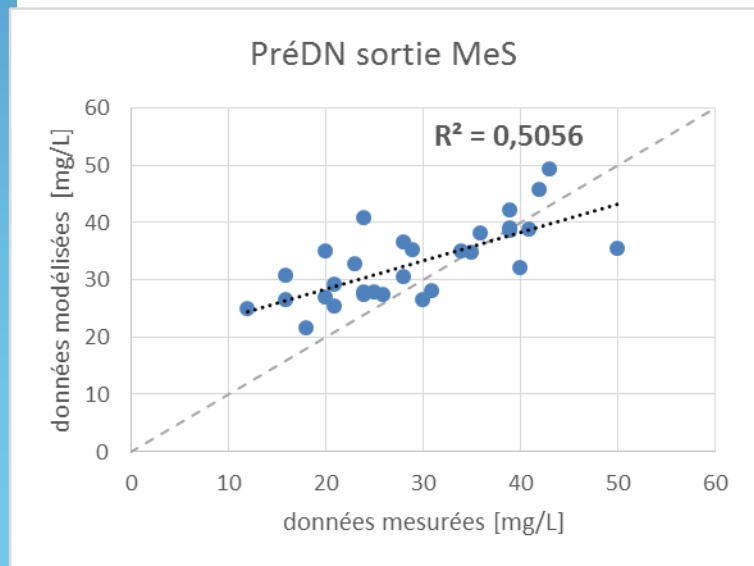
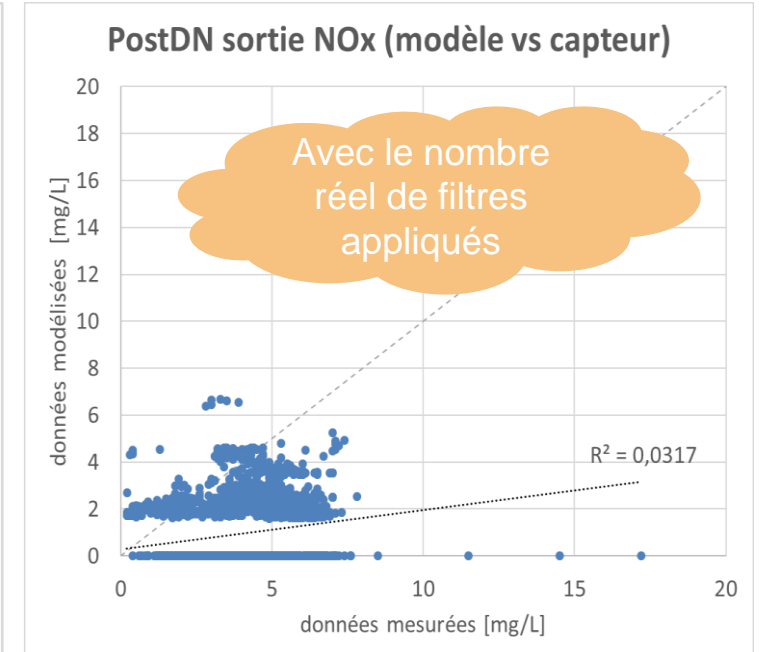
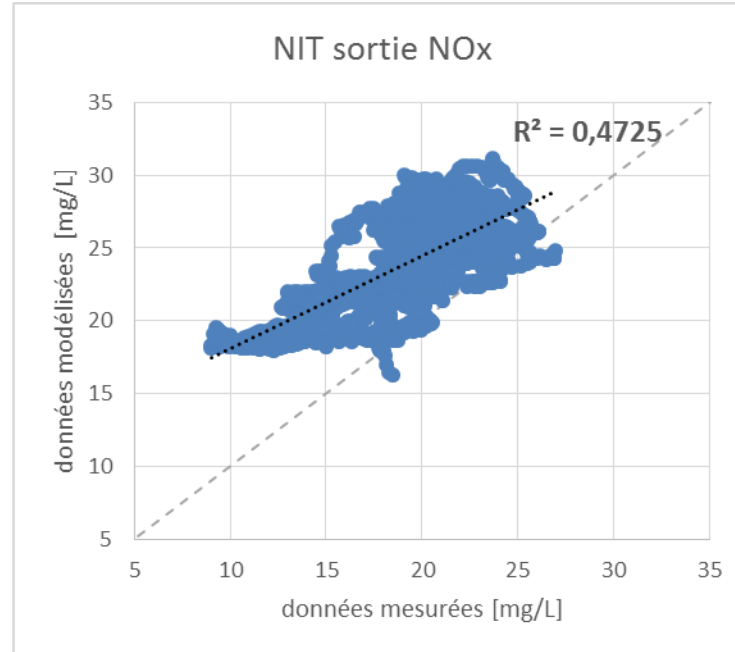
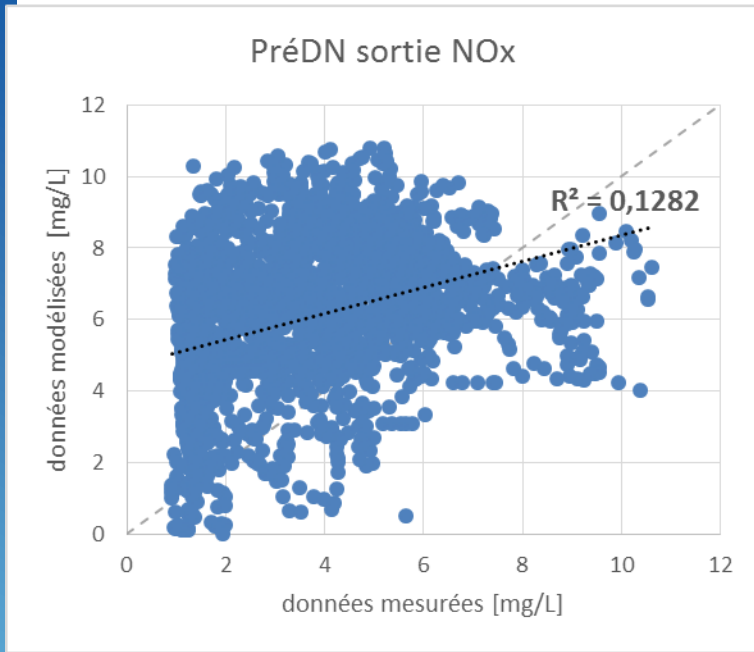
NO3 - Effluent NIT modèle Biofiltration (01122019 - 31122019)



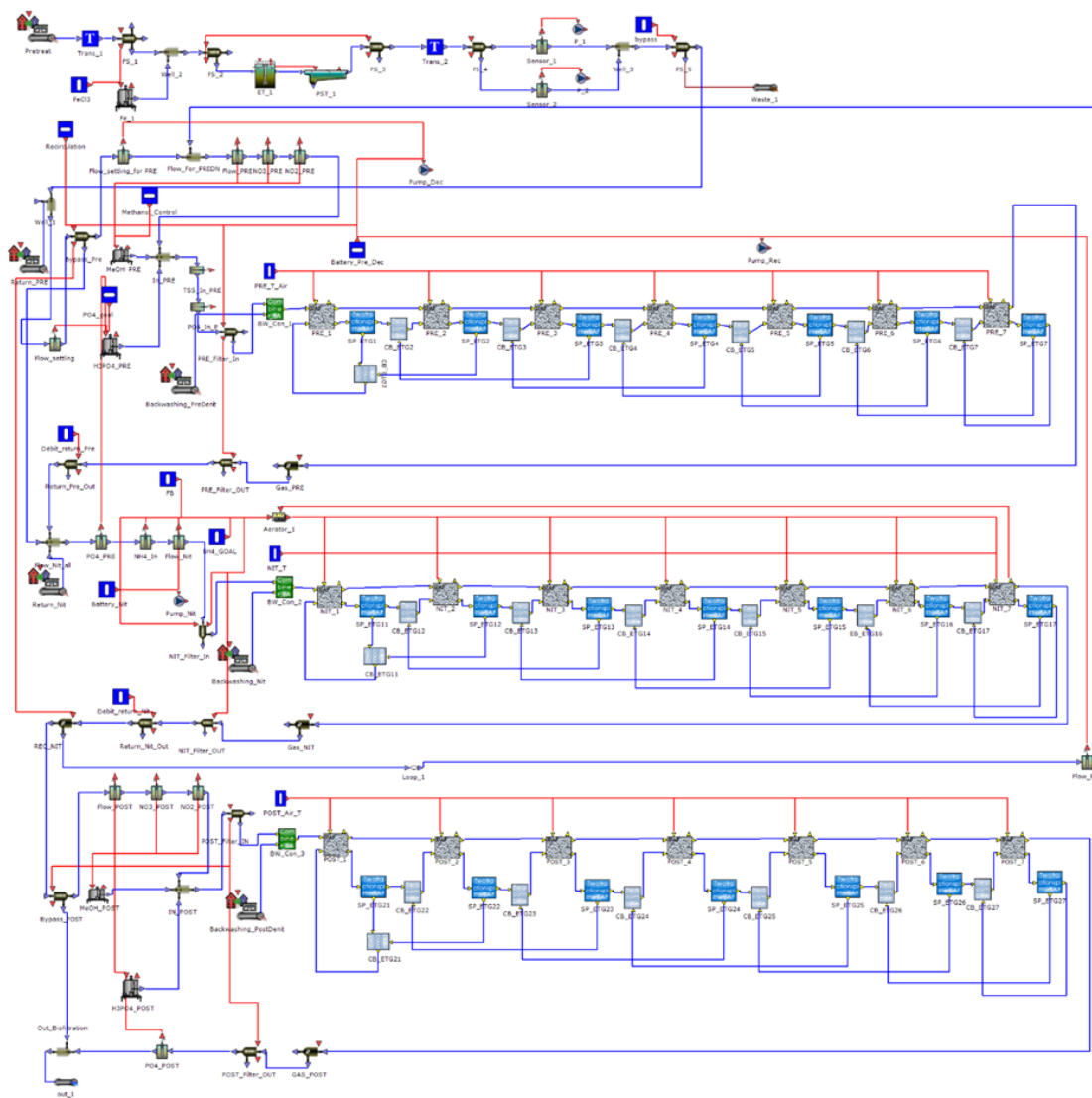
Etape	Variables	Unité	moyenne en sortie observée	moyenne en sortie modélisée	Nombre d'observations	ME	% de la moyenne observée	MAE	% de la moyenne observée	RMSE	% de la moyenne observée
PréDN	NO3	(mgN/L)	3,17	3,4	2881	-0,23	-7%	1,26	40%	1,75	55%
NIT	NO3	(mgN/L)	19,3	22,1	30	-3,37	-17%	3,45	18%	4,39	23%
PostDN	NO3	(mgN/L)	3,59	3,97	30	0,29	8%	1,51	42%	1,75	49%

RÉSULTAT D'OBJECTIF 1.2 - (RE)CALIBRER LE MODÈLE BIOFILTRATION (PRÉDN + NIT + POSTDN)

Sortie PréDN -> sortie NIT -> sortie PostDN



RÉSULTAT D'OBJECTIF 1.3 - VALIDATION DU MODÈLE FILIÈRE (DÉCANTATION PRIMAIRE + BIOFILTRATION)



Décantation primaire

La couplage des deux modèles est fait, mais les temps de calculs est trop lents.

PréDN

Raison: chaque modèle individuel utilise une solveur optimisée différente. Cela n'est pas possible dans un modèle intégré !

Nitrification

Solution proposé : découplage du modèle de la décantation !

PostDN

RÉSULTAT D'OBJECTIF 2.1 - DÉVELOPPEMENT DES ALGORITHMES POUR LA PRÉDN COMME CORRECTOR DES PRÉDICTIONS DU MODÈLE MÉCANISTIQUE

Données brutes SCADA
 $\Delta t = 15$ minutes
 20191201 - 20200315

Nettoyage des données cfr. Alfères & Vanrolleghem (2016)

Machine Learning Toolbox (Matlab)
 Data IN 10560 x 11

B_TRNNG = B(1:7000,:);
 B_TST = B(7001:9000,:);
 B_TST_NW = B(9001:end,:);

Variable d'effluent individuelle : NOx

Fichier d'input pour l'entrainement des algorithmes:

- Variables mesurées à l'entrée
- Variables mesurées à la sortie
- Les erreurs de la simulation mécanistique

Weekday	Temperature	QSW	QREC	QRetours	QBypass	Nr_Filters	MeOHinjec...	NOx_IN	NOx_REC	NH4_REC	NOx_EFF	Epred_NO...
1	18.4	3.6	0.15	0.344	0	17.4	0	2.42				-2.0884
2	18.4	3.9	0.15	0.344	0	8.6	0	2.37				-1.9884
3	18.4	4.5	1.87	0.349	0	14.3	0	3.59				-1.8184
4	18.4	4.3	2.78	0.616	0	17	0	9.05				-1.5784
5	18.4	2.8	2.16	0.713	0	15.8	0	10.72	26.96		1.39	-1.7284
6	18.4	3.1	2.19	0.682	0	14.2	0	8.61	26.67		1.32	-1.9184
7	18.4	3.9	2.61	0.678	0	15.8	0	9.3	26.43		1.23	-2.0884
8	18.4	4.4	2.97	0.689	0	17.2	0	9.09	26.12		1.16	-2.1884
9	18.4	3.4	2.5	0.454	0	16.4	0	9.51	25.9		1.12	-2.2284
10	18.4	3.7	2.53	0.489	0	16.7	0	9.21	25.65		1.09	-2.1884
11	18.4	4.2	2.63	0.269	0	17.4	0	8.11	25.29		1.07	-2.2684
12	18.4	4.5	2.69	0	0	18	0	7.87	24.96		1.04	-2.3184
13	18.4	4.5	2.74	0.2	0	17	0	7.97	24.81		1.02	-2.3484
14	18.4	4	2.67	0.377	0	17	0	8.04	24.68		1.02	-2.3484
15	18.4	4.3	2.82	0.706	0	17	0	7.96	24.54		1.02	-2.3384
16	18.4	4.5	3.05	0.708	0	18	0	8.58	24.28		1.02	-2.3284
17	18.4	4.5	3.17	0.685	0	18	0	8.46	23.79		1.03	-2.3184
18	18.4	4.5	3.16	0.659	0	17.1	0	8.19	23.06		1.02	-2.3184
19	18.5	4.5	3.12	0.523	0	17.3	0	7.77	22.28		1.02	-2.3284

Les prédicteurs

La réponse

Pour chaque sous-jeux

- Interpolation (moving mean)
- Normalization [0...1]

Sélection des modèles

5 Neural Network	RMSE (Validation): 0.055249
Last change: Medium Neural Network	12/12 features
1 Tree	RMSE (Validation): 0.057576
Last change: Fine Tree	12/12 features
2.2 Tree	RMSE (Validation): 0.057576
Last change: Fine Tree	12/12 features
2.3 Tree	RMSE (Validation): 0.059113
Last change: Medium Tree	12/12 features
4 Neural Network	RMSE (Validation): 0.063706
Last change: Narrow Neural Network	12/12 features
2.4 Tree	RMSE (Validation): 0.066307
Last change: Coarse Tree	12/12 features
2.1 Linear Regression	RMSE (Validation): 0.081013
Last change: Linear	12/12 features
3 SVM	RMSE (Validation): 0.081795
Last change: Linear SVM	12/12 features

Sélection des features pertinents

	Select	Features
1	<input checked="" type="checkbox"/>	Weekday
2	<input checked="" type="checkbox"/>	Temperature
3	<input checked="" type="checkbox"/>	QSW
4	<input checked="" type="checkbox"/>	QREC
5	<input checked="" type="checkbox"/>	QRetours
6	<input checked="" type="checkbox"/>	QBypass
7	<input checked="" type="checkbox"/>	Nr_Filters
8	<input checked="" type="checkbox"/>	MeOHInjection
9	<input checked="" type="checkbox"/>	NOx_IN
10	<input checked="" type="checkbox"/>	NOx_REC
11	<input checked="" type="checkbox"/>	NH4_REC
12	<input checked="" type="checkbox"/>	NOx_EFF

RÉSULTAT D'OBJECTIF 2.1 - DÉVELOPPEMENT DES ALGORITHMES POUR LA PRÉDN COMME CORRECTOR DES PRÉDICTIONS DU MODÈLE MÉCANISTIQUE

Testing avec des données nouvelles jeu de données DTST (01.03.2020 – 16.03.2020) taille 1058 x 13

★ 8 Neural Network

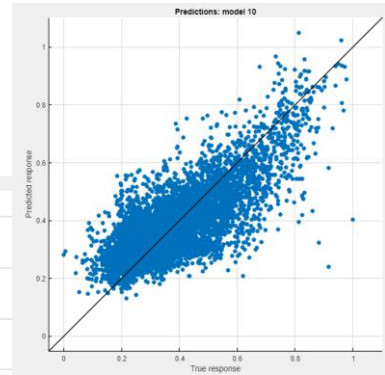
Last change: Medium Neural Network

RMSE (Validation): 0.1001

6/8 features (PCA on)

Hyperparameter Optimization

Different combinations of hyperparameter values using an optimization scheme that seeks to minimize the model mean squared error (MSE)

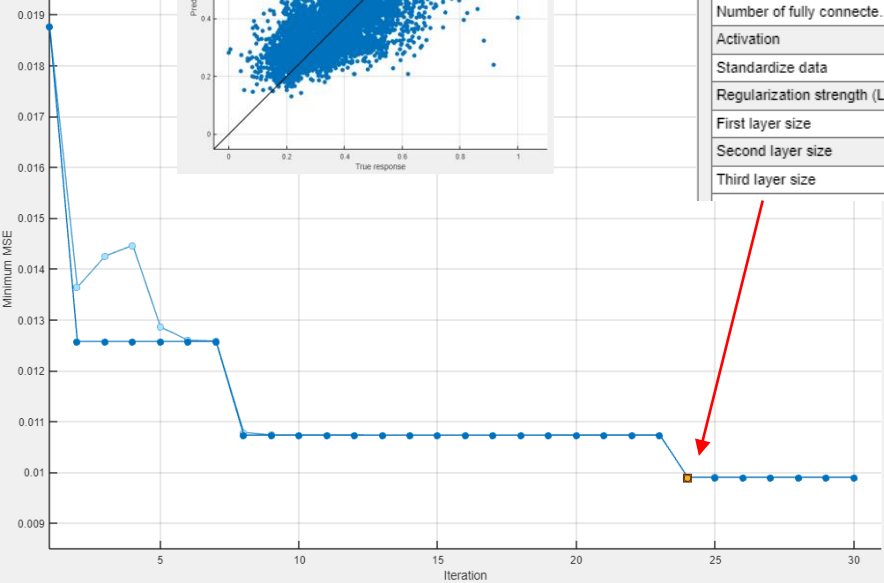


Legend

- Estimated minimum MSE
- Observed minimum MSE
- Bestpoint hyperparameters
- Minimum error hyperparameters

Optimization Results

Hyperparameter	Value
Number of fully connecte...	3
Activation	Sigmoid
Standardize data	Yes
Regularization strength (L...	3.2473e-07
First layer size	31
Second layer size	296
Third layer size	13



Model 10: Optimizable Neural Network
Status: Trained

Training Results

RMSE (Validation) 0.10034
R-Squared (Validation) 0.61
MSE (Validation) 0.010068
MAE (Validation) 0.075927
Prediction speed ~29000 obs/sec
Training time 2345.2 sec

Test Results

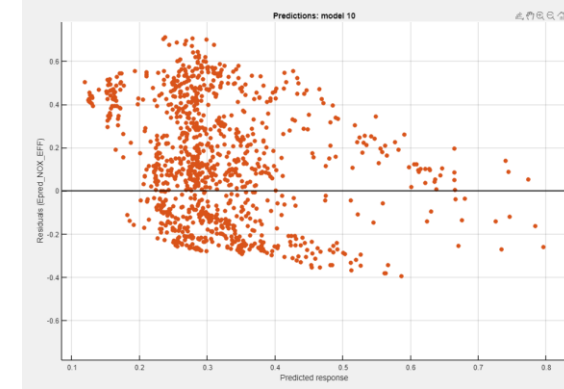
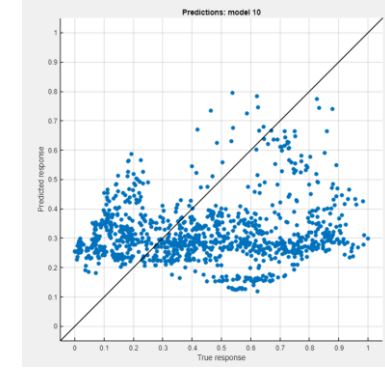
RMSE (Test) 0.2966
R-Squared (Test) -0.29
MSE (Test) 0.087969
MAE (Test) 0.24623

Model Hyperparameters

Preset: Optimizable Neural Network
First layer size: 10
Second layer size: 10
Third layer size: 10
Iteration limit: 750

Feature Selection: 8/12 individual features selected

Select	Features	MRMR
<input type="checkbox"/>	NOx_EFF	0.6672
<input checked="" type="checkbox"/>	NOx_IN	0.3680
<input checked="" type="checkbox"/>	Weekday	0.2465
<input checked="" type="checkbox"/>	Temperature	0.2106
<input checked="" type="checkbox"/>	Nr_Filters	0.1713
<input checked="" type="checkbox"/>	QREC	0.1313
<input checked="" type="checkbox"/>	NH4_REC	0.1310
<input checked="" type="checkbox"/>	NOx_REC	0.1005
<input checked="" type="checkbox"/>	QSW	0.0788
<input type="checkbox"/>	QBypass	0.0471



Over-fitting!

Model 8: Neural Network
Status: Trained

Training Results

RMSE (Validation) 0.059016
R-Squared (Validation) 0.87
MSE (Validation) 0.0034829
MAE (Validation) 0.043575
Prediction speed ~66000 obs/sec
Training time 10.114 sec

Test Results

RMSE (Test) 0.19441
R-Squared (Test) 0.44
MSE (Test) 0.037796
MAE (Test) 0.14829

Model Hyperparameters

Preset: Medium Neural Network
Number of fully connected layers: 1
First layer size: 25
Activation: ReLU
Iteration limit: 1000
Regularization strength (Lambda): 0
Standardize data: Yes

Feature Selection: 9/12 individual features

Select	Features
<input checked="" type="checkbox"/>	NOx_EFF
<input checked="" type="checkbox"/>	NOx_IN
<input checked="" type="checkbox"/>	Weekday
<input checked="" type="checkbox"/>	Temperature
<input checked="" type="checkbox"/>	Nr_Filters
<input checked="" type="checkbox"/>	QREC
<input checked="" type="checkbox"/>	NH4_REC
<input checked="" type="checkbox"/>	NOx_REC
<input checked="" type="checkbox"/>	QSW
<input type="checkbox"/>	QBypass

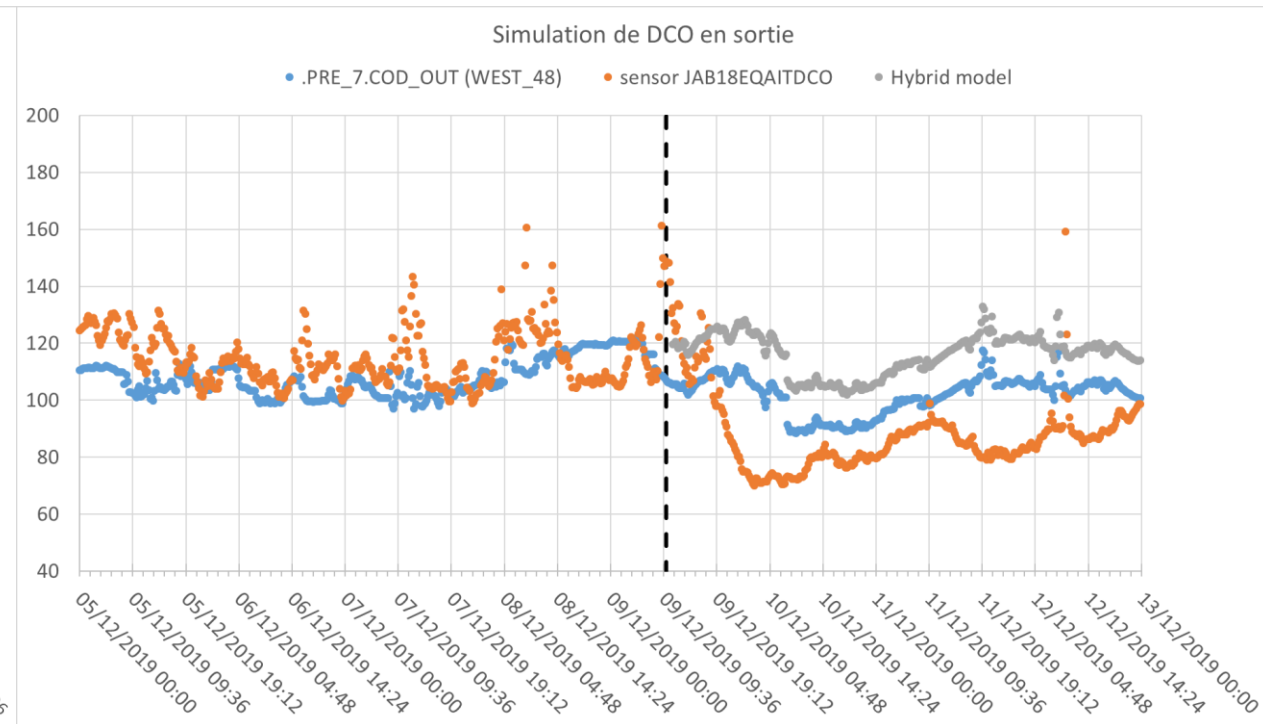
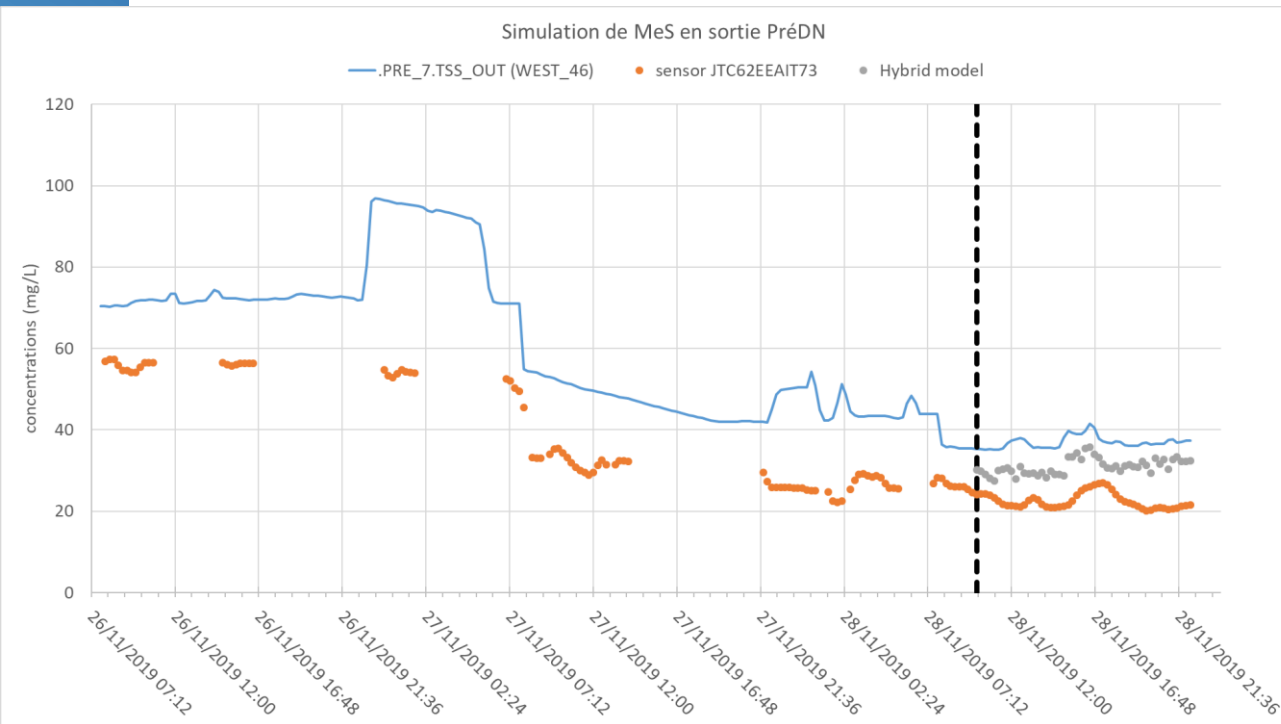
RÉSULTAT D'OBJECTIF 3.1 - DÉVELOPPEMENT D'UN MODÈLE HYBRIDE POUR LA PRÉDN

Variables en sortie du modèle hybride

Variable d'effluent individuelle : MeS

Horizon de prédiction : 6 à 12 heures

Variable d'effluent individuelle : DCO



en cours de développement

Variable d'effluent individuelle : NO3

Variable d'effluent individuelle : NO2

04/ CONCLUSIONS & PERSPECTIVES

Générique

CONCLUSIONS

Calibration du modèle mécanistique

- **Sensible** à la qualité des données observées
- **Conformité** aux règles appliquées par les opérateurs ?
- La **complexité** optimale du modèle ?
- La **connexion** des différentes procédés de traitement n'est toujours pas réalisable dans 1 modèle complet par cause de solveurs mathématiques

Calibration de modèles statistiques développés dans Matlab

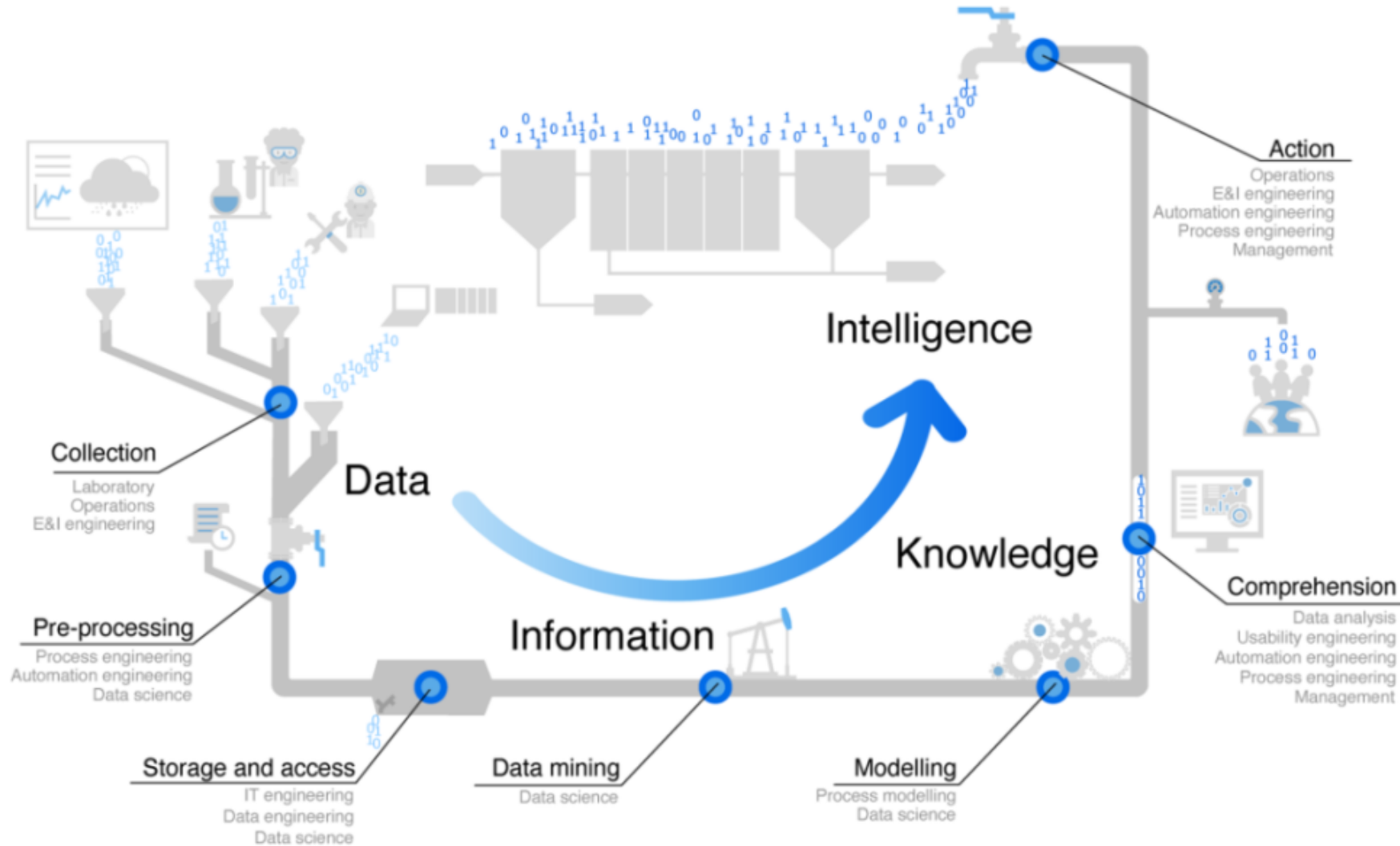
- **Rapide** à apprendre, facile à se suradapter ('overfitting')
- Sélection des '**features**' a une grande impact
- L'ANN est le plus populaire dans la littérature, mais pas toujours le plus efficace !

Développement d'un modèle hybride

- Un **cadre standardisé** pour le développement d'un modèle hybride ?
- Le choix de l'**architecture** dépend du problème, des données disponibles et des connaissances manquantes
- La **précision** des données => à suivre le '*pipeline*' de données de haute qualité

CONCLUSIONS (2)

L'importance d'un 'pipeline' de données de haute qualité pour éviter un scénario « GI-GO »



PROCHAINES ÉTAPES

Travaux pratiques

- Finaliser les testing des algorithmes univariate pour corriger les Erreur des model mécanistique (objectif 2.1)
- Commencer l'entraînement des algorithmes multi-variate pour indiquer les paramètres de modèle WEST (objectif 2.2)
- Finaliser le développement de model hybride (objectifs 3.1 et 3.2)
- Ecrire et réécrire la thèse

Conférences

1. Journées Information Eaux – Poitiers (11 – 13 octobre 2022) – ***candidature accepté***
2. ICA 2022 Instrumentation, Control and Automation – Chine (17 – 22 octobre 2022) - ***candidature envoyée***
3. WRRmod2022+ *par visio* (15 – 18 janvier 2023) – ***candidature envoyée***

Publications

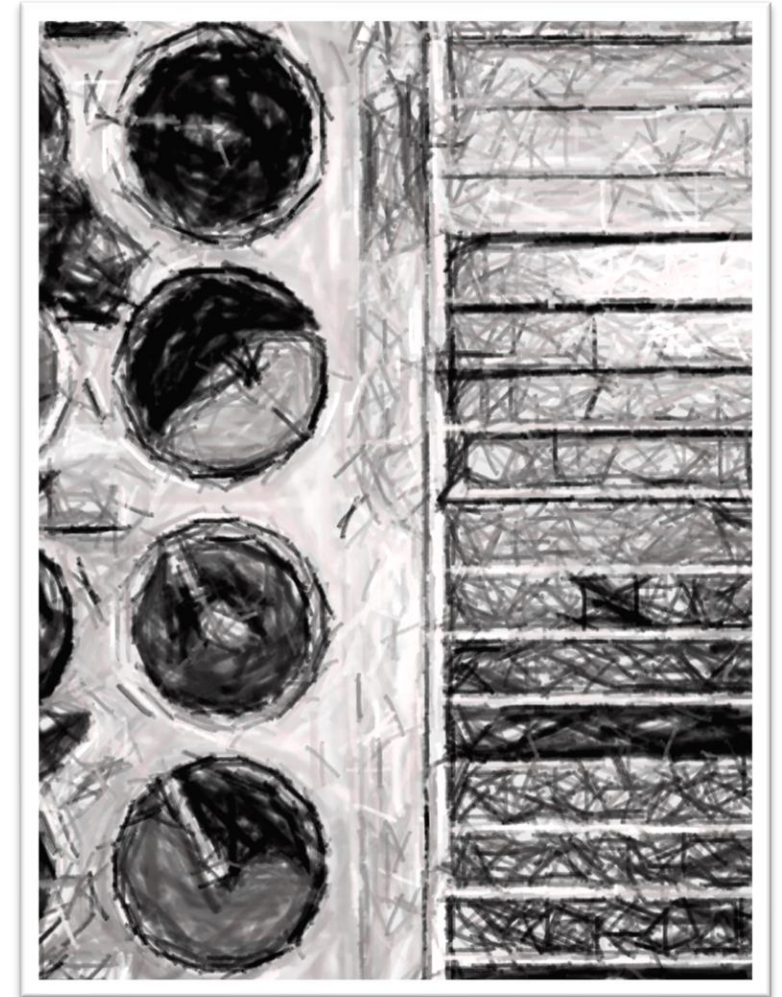
- Smart Water Management in the Greater Paris Region; chapitre dans une publication de l'UNESCO et W-SMART (prevu 2022)

MERCI POUR VOTRE ATTENTION!

Questions ? Suggestions ?

marcello.serrao@enpc.fr

Bruno Tassin
Peter Vanrolleghem
Vincent Jauzein
Vincent Rocher
Sam Azimi
Ilan Juran



Ivan Bandura



RÉFÉRENCES

- Alferes, J., & Vanrolleghem, P. A. (2016). Efficient automated quality assessment: Dealing with faulty on-line water quality sensors. *AI Communications*, 29(6), 701-709.
- Bernier J., Rocher V., Guérin S., Lessard P. (2014b): Modelling of a carbon removal biological aerated filter doing partial nitrification during large-scale secondary treatment; *Water Quality Research Journal of Canada* ; 49, 3 : 245-257.
- Garrido-Baserba, M., Corominas, L., Cortés, U., Rosso, D., & Poch, M. (2020). The fourth-revolution in the water sector encounters the digital revolution. *Environmental Science & Technology*, 54(8), 4698-4705.
- Ingildsen, P. and Olsson, G. (2016). *Smart Water Utilities. Complexity made simple.* IWA Publishing.
- Lee, D. S., Jeon, C. O., Park, J. M., & Chang, K. S. (2002). Hybrid neural network modeling of a full-scale industrial wastewater treatment process. *Biotechnology and bioengineering*, 78(6), 670-682.
- Lee, D. S., Vanrolleghem, P. A., & Park, J. M. (2005). Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. *Journal of Biotechnology*, 115(3), 317-328.
- Li, F. and Vanrolleghem, P. A. (2019) WRRF influent generator model for flowrate and quality prediction from combined sewer systems based on a data-driven methodology. In: *Proceedings 33e Congrès de l'Est du Canada de Recherche sur la Qualité de l'Eau (ACQE)*. Montréal, Québec, Canada, October 25 2019.
- Mannina, G., Cosenza, A., Vanrolleghem, P. A., & Viviani, G. (2011). A practical protocol for calibration of nutrient removal wastewater treatment models. *Journal of hydroinformatics*, 13(4), 575-595.
- Rieger, L., Gillot, S., Langergraber, G., Ohtsuki, T., Shaw, A., Takacs, I., Winkler, S. (2012). *Guidelines for Using Activated Sludge Models.* IWA publishing, London, UK.
- SIAAP (2018). *Innover dans les pratiques de monitoring et d'exploitation des stations d'épuration. Enseignements scientifiques et techniques tirés de la phase I (2014-2017) du programme Mocopée.*
- Schneider, M. Y., Quaghebeur, W., Borzooei, S., Froemelt, A., Li, F., Saagi, R., ... & Torfs, E. (2022). Hybrid modelling of water resource recovery facilities: status and opportunities. *Water Science and Technology*, 85(9), 2503-2524.
- Therrien, J. D., Nicolaï, N., & Vanrolleghem, P. A. (2020). A critical review of the data pipeline: how wastewater system operation flows from data to intelligence. *Water Science and Technology*.
- Von Stosch, M., Oliveira, R., Peres, J., & de Azevedo, S. F. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, 60, 86-101.
- Zhu, J., Bernier, J., Pauss, A., Vanrolleghem, P. A., Rocher, V. (2018a) Modélisation de la station Seine Aval–Vers une optimisation en temps réel des coûts d'exploitation et environnementaux. *Présentation Journée Information Eaux, Poitiers*
- Zhu, J., Bernier, J., Azimi, S., Pauss, A., Rocher, V., Vanrolleghem, P. A. (2018) Comprehensive modelling of full-scale nitrifying biofilters and validation under different configurations. *Presentation IWA Nutrient Removal and Recovery Conference, 18 - 21 November 2018, Brisbane, Australia.*
- Zhu, J. (2020). *Modélisation détaillée du fonctionnement de la filière complète de biofiltration de a station de traitement des eaux usées Seine Aval (PhD Thesis).* Université de Technologie de Compiègne, Sorbonne Université.

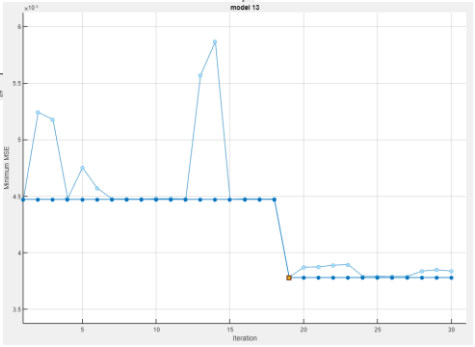
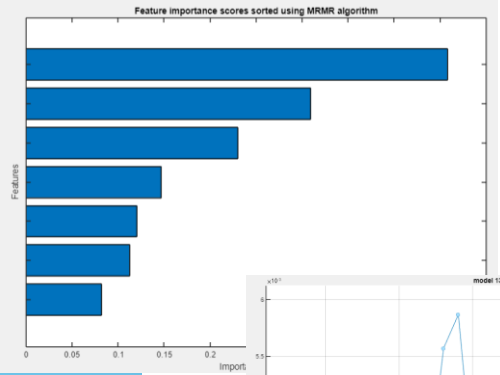
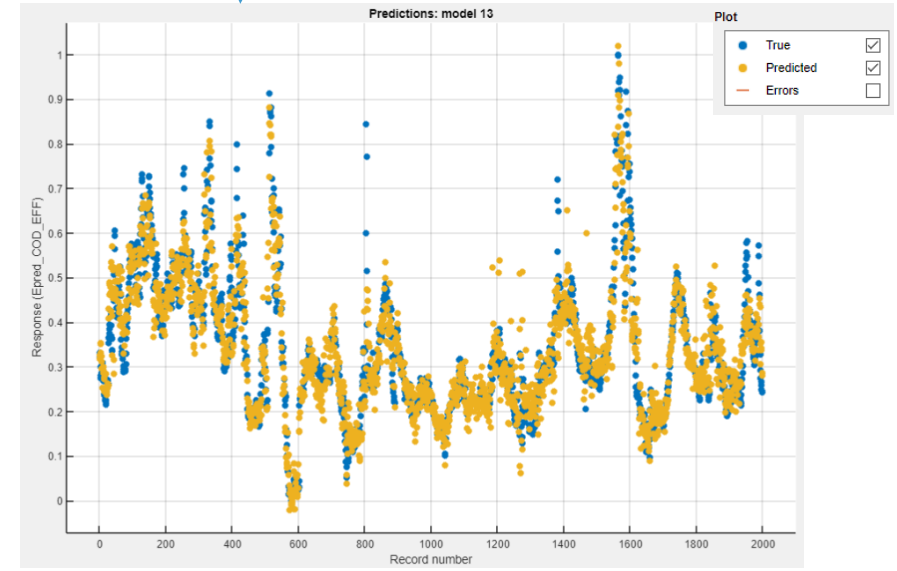
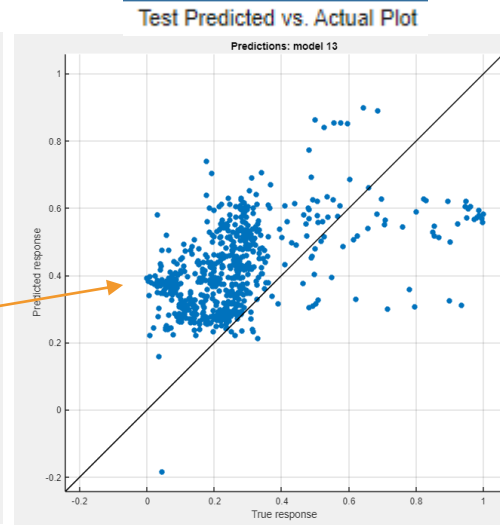
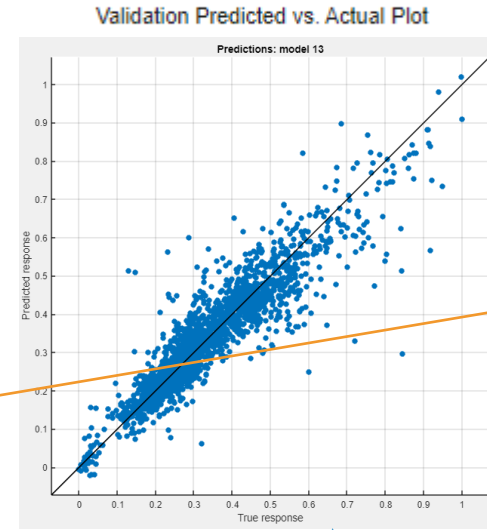
RÉSULTAT D'OBJECTIF 2.1 - DÉVELOPPEMENT DES ALGORITHMES POUR LA PRÉDN COMME CORRECTOR DES PRÉDICTIONS DU MODÈLE MÉCANISTIQUE

Variable d'effluent individuelle : DCO

Model	RMSE (Validation)
1 Tree	0.077159
2.2 Tree	0.077159
4 Neural Network	0.084347
2.3 Tree	0.086708
3 Neural Network	0.092515
2.4 Tree	0.10313
2.1 Linear Regression	0.11391
5 SVM	0.11622

Select	Features	MRMR
<input checked="" type="checkbox"/>	Temperature	0.4583
<input checked="" type="checkbox"/>	QREC	0.3095
<input checked="" type="checkbox"/>	COD_IN	0.2300
<input checked="" type="checkbox"/>	COD_EFF_MOD	0.1472
<input checked="" type="checkbox"/>	Nr_Filters	0.1206
<input checked="" type="checkbox"/>	Weekday	0.1128
<input checked="" type="checkbox"/>	QSW	0.0818
<input type="checkbox"/>	QBypass	0.0308
<input type="checkbox"/>	QRetours	0
<input type="checkbox"/>	MeOHInjection	0

Over-fitting !



Legend

- Estimated minimum MSE
- Observed minimum MSE
- Bestpoint hyperparameters
- Minimum error hyperparameters

Optimization Results

Hyperparameter	Value
Number of fully connected layers	2
Activation	ReLU
Standardize data	No
Regularization strength (Lambda)	5.0213
First layer size	186
Second layer size	119

Model 13: Optimizable Neural Network
 Status: Trained
 Model is favorite

Training Results

- RMSE (Validation): 0.061763
- R-Squared (Validation): 0.84
- MSE (Validation): 0.0038147
- MAE (Validation): 0.041306
- Prediction speed: ~30000 obs/sec
- Training time: 683.9 sec

Model Hyperparameters

- Preset: Optimizable Neural Network
- First layer size: 10
- Second layer size: 10
- Third layer size: 10
- Iteration limit: 1000

Optimized Hyperparameters

- Number of fully connected layers: 3
- Activation: ReLU
- Iteration limit: 1000
- Regularization strength (Lambda): 0
- Standardize data: Yes

Hyperparameter Search Range

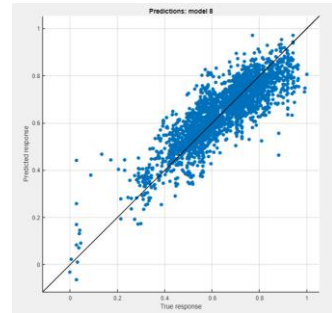
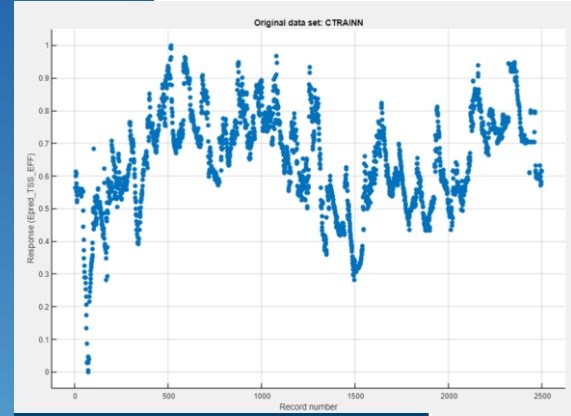
- Feature Selection: 7/10 individual features selected
- PCA: Disabled
- Optimizer: Bayesian optimization

Optimizer: Bayesian optimization

- Acquisition function: Expected improvement per second plus
- Iterations: 30
- Training time limit: false

RÉSULTAT D'OBJECTIF 2.1 - DÉVELOPPEMENT DES ALGORITHMES POUR LA PRÉDN COMME CORRECTOR DES PRÉDICTIONS DU MODÈLE MÉCANISTIQUE

Variable d'effluent individuelle : MeS

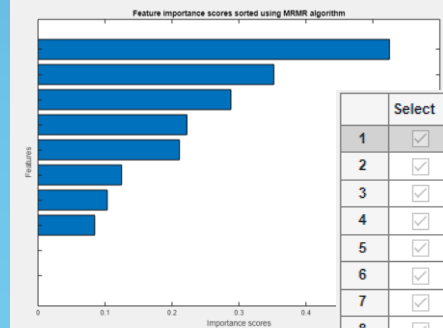


Models

Sort by: RMSE (Validation)

Model	RMSE (Validation)	Last change	Features
1 Tree	0.061989	Fine Tree	10/10 features
2.2 Tree	0.061989	Fine Tree	10/10 features
2.3 Tree	0.065941	Medium Tree	10/10 features
5 Neural Network	0.071063	Medium Neural Network	10/10 features
4 Neural Network	0.077301	Narrow Neural Network	10/10 features
2.4 Tree	0.082166	Coarse Tree	10/10 features
2.1 Linear Regression	0.12316	Linear	10/10 features
3 SVM	0.14471	Linear SVM	10/10 features

Model	RMSE (Validation)	Last change	Features
8 Neural Network	0.078294	Narrow Neural Network	7/10 features
2.4 Tree	0.082166	Coarse Tree	10/10 features
2.1 Linear Regression	0.12316	Linear	10/10 features
3 SVM	0.14471	Linear SVM	10/10 features
7 SVM	0.14798	Linear SVM	7/10 features



Select	Features	MRMR
<input checked="" type="checkbox"/>	1 Temperature	0.5252
<input checked="" type="checkbox"/>	2 QSW	0.3523
<input checked="" type="checkbox"/>	3 TSS_EFF_PR...	0.2880
<input checked="" type="checkbox"/>	4 Nr_Filters	0.2228
<input checked="" type="checkbox"/>	5 TSS_SW_EFF...	0.2110
<input checked="" type="checkbox"/>	6 QREC	0.1247
<input checked="" type="checkbox"/>	7 QBypass	0.1035
<input checked="" type="checkbox"/>	8 Weekday	0.0851
<input checked="" type="checkbox"/>	9 QRetours	0.0000
<input checked="" type="checkbox"/>	10 MeOHInjection	0

Model 10: Optimizable Neural Network
Status: Trained

Training Results
 RMSE (Validation) 0.081729
 R-Squared (Validation) 0.71
 MSE (Validation) 0.0066796
 MAE (Validation) 0.063309
 Prediction speed ~230000 obs/sec
 Training time 69.897 sec

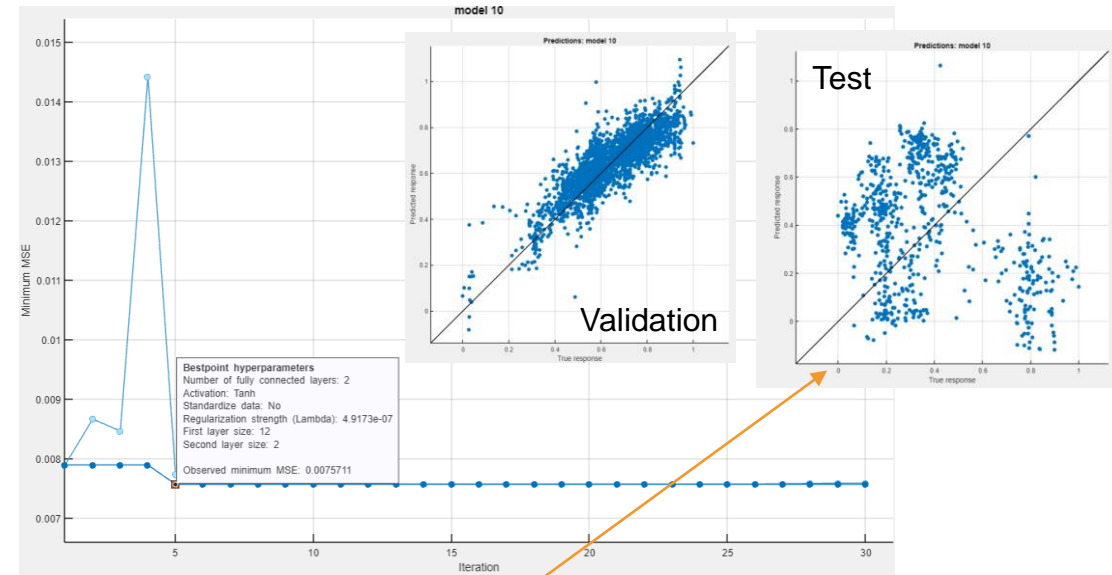
Model Hyperparameters
 Preset: Optimizable Neural Network
 First layer size: 10
 Second layer size: 10
 Third layer size: 10
 Iteration limit: 1000

Optimized Hyperparameters
 Number of fully connected layers: 3
 Activation: ReLU
 Iteration limit: 1000
 Regularization strength (Lambda): 0
 Standardize data: Yes

Hyperparameter Search Range

Feature Selection: Top 7/10 features selected using MRMR
 PCA: Disabled
 Optimizer: Bayesian optimization

Optimizer: Bayesian optimization
 Acquisition function: Expected improvement per second plus iterations: 30
 Training time limit: false



Over-fitting!

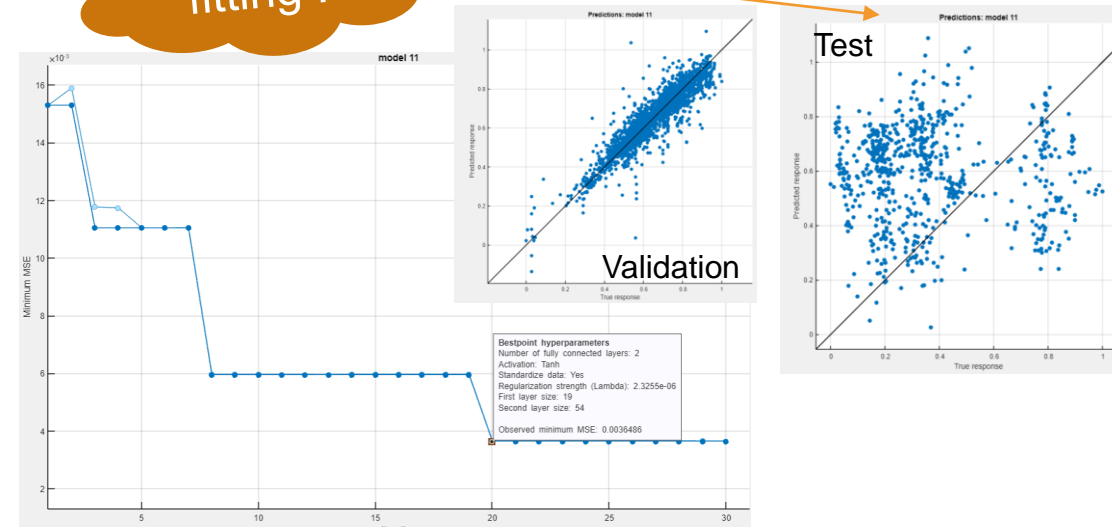
Model 11: Optimizable Neural Network
Status: Trained

Training Results
 RMSE (Validation) 0.059198
 R-Squared (Validation) 0.85
 MSE (Validation) 0.0035044
 MAE (Validation) 0.040605
 Prediction speed ~190000 obs/sec
 Training time 121.87 sec

Test Results
 RMSE (Test) 0.37841
 R-Squared (Test) -1.52
 MSE (Test) 0.14319
 MAE (Test) 0.3281

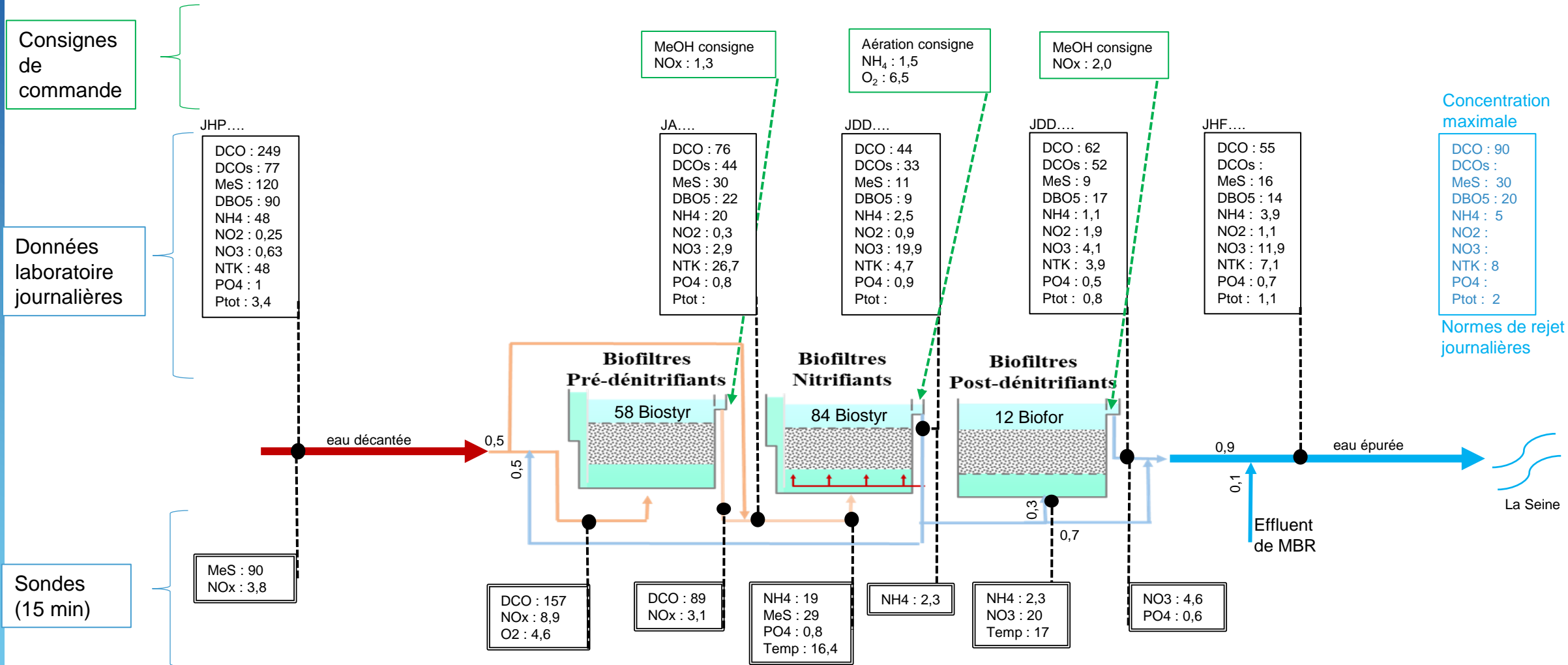
Model Hyperparameters
 Preset: Optimizable Neural Network
 First layer size: 10
 Second layer size: 10
 Third layer size: 10
 Iteration limit: 1000

Feature Selection: Top 7/10 features selected using MRMR
 PCA: Disabled
 Optimizer: Bayesian optimization



DISPONIBILITÉ DES DONNÉES OBSERVÉES

VALEURS MOYENNES 2019.12.01 – 2020.03.15 EN [MG/L]



MÉTHODOLOGIE D'UN MODÈLE HYBRIDE

Etude de cas - Lee et al. (2002 & 2005)

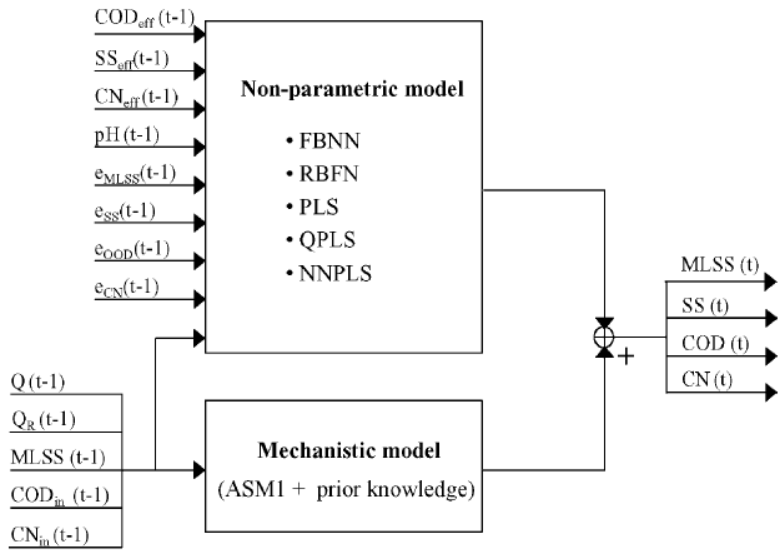
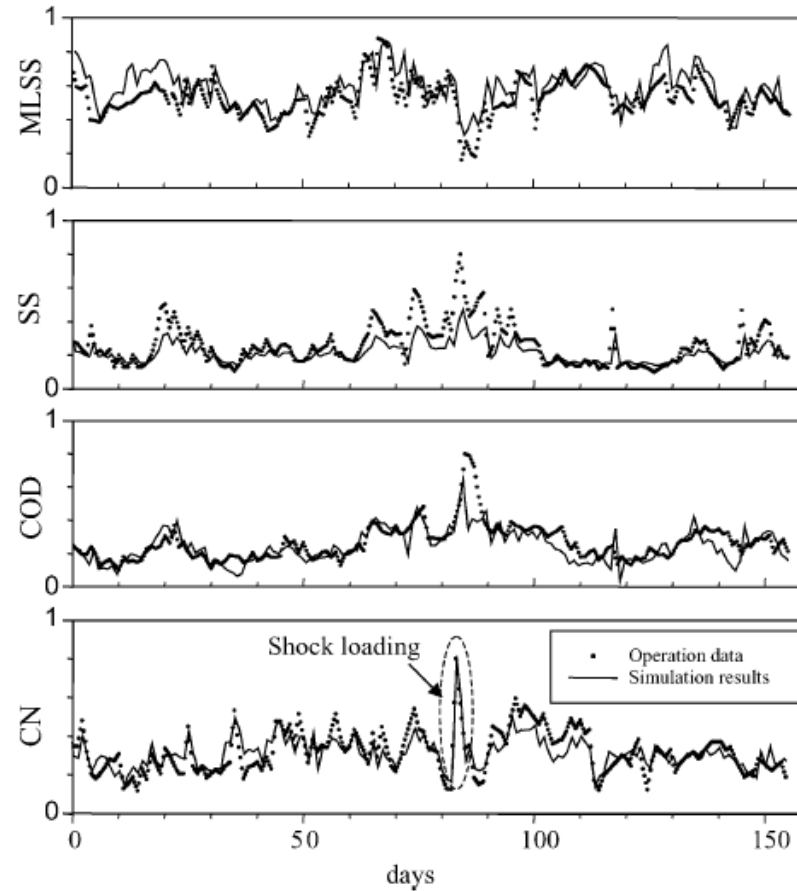


Fig 2. Parallel hybrid modeling approaches (Lee et al., 2005).

Table II. The sum of the squared errors for the recall and generalization of four different models (for scaled data).

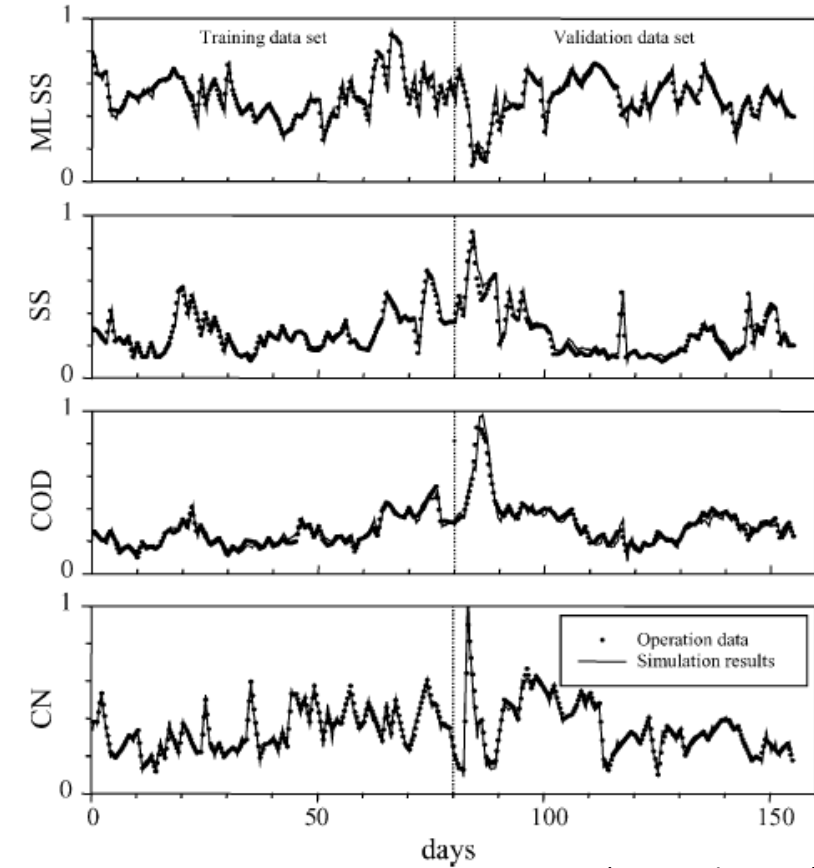
Models	Recall	Validation
Mechanistic model	11.48	15.25
Neural network model	3.08	9.91
Serial hybrid model	9.37	12.43
Parallel hybrid model	3.66	7.58

(Lee et al., 2002).



(Lee et al., 2005).

Fig. 3. Simulation results of the simplified mechanistic model.



(Lee et al., 2005).

Fig. 4. Simulation results of the parallel hybrid model with NNPLS as non-parametric model.

Conclusion : le MH a un fort potentiel d'amélioration de la précision des prédictions, mais testé uniquement pour les modèles les plus simples

OBJECTIFS & PLANIFICATION

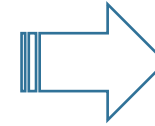
Objectif 1 : modélisation

(Re-)calibration et validation des modèles phénoménologiques / mécanistiques



Objectif 2 : l'IA

Développement et testing des algorithmes d'apprentissage automatique



Objectif 3 final : Jumeau Numérique Adaptatif Hybride

- Effectuer une analyse de la qualité des données des mesures en continu
- Générer des données de modèle d'affluent
- Recalibrer les valeurs des paramètres

- 1.1 (re)calibrer le modèle pour la PrédN à la SAV

- 1.2 (re)calibrer le modèle biofiltration (PrédN + NIT + PostDN) à la SAV

- 1.3 validation du modèle filière (décantation primaire + biofiltration) à la SAV

- 2.1 développement des algorithmes pour la PrédN comme corrector des prédictions du modèle mécanistique

- 2.2 développement des algorithmes pour la PrédN comme corrector des paramètres du modèle mécanistique

- 3.1 développement d'un modèle hybride pour la PrédN comme corrector des prédictions du modèle mécanistique

- 3.2 développement d'un modèle hybride pour la PrédN comme corrector des paramètres du modèle mécanistique

Planification du projet

2020				2021				2022				2023	
3	6	9	12	3	6	9	12	3	6	9	12	3	6
étude de la littérature		Calibration de modèle PrédN		Calibration du modèle filière				Machine Learning sur le modèle PrédN		Modèle Hybride	Ecrire la thèse	Ecrire la thèse	Soutenance de la thèse