

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE D'OUUM EL BOUAGHI
Faculté : Sciences et Sciences Appliquées



Thèse

POUR OBTENIR LE DIPLOME DE

Doctorat 3^{ème} Cycle

Filière : Hydraulique

Spécialité : Sciences Hydrauliques et Développement Durable

Thème

**Contribution à la modélisation prédictive d'une station d'épuration
grandeur nature par les réseaux de neurones artificiels**

Cas d'étude : Station d'Épuration de la ville de Ain Beida (Algérie)

Présenté Par

ZAIDI Kaddour

Thèse soutenue le 21/11/2024 devant le jury composé de

N°	Nom et prénom	Grade	Etablissement	Qualité
01	Zeroual Abdelatif	MCA	Université d'Oum El Bouaghi	Président
02	Djeddou Messaoud	Prof.	Université d'Oum El Bouaghi	Rapporteur
03	Toumi Abdelouaheb	Prof.	Université de Guelma	Examineur
04	Loukam Imed	MCA	Université de Souk-Ahras	Examineur
05	Merrouchi Farida	MCA	Université d'Oum El Bouaghi	Examinatrice

Dédicace

À la ma mémoire de mon Père,

À ma Mère,

À mes Frères,

À mes Sœurs,

À ma famille,

À mes amis,

Je dédie ce modeste travail.

Remerciements

Tout d'abord, nous exprimons nos louanges à Dieu tout-puissant pour nous guider, nous accorder la patience, et nous fournir les connaissances nécessaires pour achever cette thèse et rendre nos familles fières.

Nous tenons à exprimer notre sincère gratitude envers notre superviseur, Pr. Messaoud Djeddou, pour son orientation, ses précieuses remarques et sa capacité à nous pousser au-delà de nos limites. Sa patience, son soutien et son inspiration ont grandement contribué à nos réussites.

Nous remercions également Pr. Khadidja Zaidi pour sa présence lorsque de nombreuses portes étaient fermées, pour son guidage et son soutien dans nos recherches académiques. Nous sommes profondément reconnaissants envers les membres du jury pour avoir consacré leur temps à la lecture et à l'évaluation de cette thèse.

Je tiens à remercier le personnel de la station d'épuration des eaux usées de la ville d'Aïn Beïda, en particulier le directeur et les ingénieurs, pour leur accueil chaleureux et leur gentillesse, pour nous avoir aidés à atteindre nos objectifs et pour avoir facilité l'accès aux données d'exploitation.

Enfin, nous souhaitons sincèrement remercier tous nos enseignants du primaire, du moyen, du lycée et de l'université qui ont joué un rôle essentiel dans notre parcours académique. Nous sommes reconnaissants pour votre orientation, vos encouragements et les informations précieuses que vous avez partagées avec nous. Nous sommes honorés d'avoir été vos étudiants.

Table des matières

Dédicace	I
Remerciements	II
Table des matières	III
Liste des figures	VI
Liste des tableaux	VII
Abréviations	VIII
Résumé	IX
Abstract	X
المخلص	XI
Introduction générale	1
Chapitre 1 : Modèles de Boues Activées dans les Stations d'Épuration	3
1.1 Introduction	3
1.2 Historique de la modélisation des procédés de boues activées	3
1.3 Objectifs de la modélisation des procédés de boues activées	5
1.4 Modélisation des procédés de boues activées dans les stations d'épuration	5
1.4.1 Modèle de Boue Activée 1	5
1.4.2 Modèle de Boues Activées 2	10
1.4.3 Modèle de Boues Activées 2 Réduit	10
1.4.5 Modèle de Boues Activées 3	11
1.5 Techniques Alternatives de Modélisation	12
1.6 Conclusion	13
Chapitre 2 : Réseaux de Neurones Artificiels : Principes, Modèles et Applications	14
2.1 Introduction	14
2.2 Historique des Réseaux Neurones Artificiels	14
2.3 Principe des Réseaux de Neurones Artificiels	15
2.4 Réseaux de neurones artificiels	15
2.4.1 Neurone biologique	16
2.4.2 Neurone formel	17
2.4.2.1 Réseaux neurone à perceptron à couche unique	18
2.4.2.2 Réseaux neurone à perceptron multi couches	19
2.4.3 Fonction de transfert	19
2.4.3.1 La fonction linéaire	21
2.4.3.2 La fonction sigmoïde	21
2.4.3.3 La tangente hyperbolique	21
2.4.3.4 La fonction seuil	21
2.5 Architecture des réseaux de neurones artificiels	21
2.5.1 Réseaux déverrouillés	21
2.5.2 Réseaux verrouillés	22
2.5.3 Réseaux avec des connexions complexes	22
2.6 Apprentissage des réseaux de neurones artificiels	22
2.6.1 Apprentissage supervisé	22
2.6.2 Apprentissage non supervisé	22
2.6.3 Apprentissage par renforcement	22
2.7 Règles d'apprentissage	23
2.7.1 Règle de Hebb	23
2.7.2 Règle de Widrow-Hoff	23
2.7.3 Règle de rétropropagation	23
2.7.4 Règle de compétition	23

2.7.5 Règle de correction d'erreur	23
2.8 Paramètres d'apprentissage	24
2.9 Critères d'arrêt de l'apprentissage	24
2.10 Phénomènes d'apprentissage des réseaux de neurones artificiels	24
2.11 Propriétés des réseaux de neurones artificiels	25
2.12 Classification des réseaux de neurones artificiels	26
2.13 Application des réseaux de neurones artificiels	26
2.14 Avantages des réseaux de neurones artificiels	26
2.15 Limites opérationnelles des réseaux de neurones artificiels	27
2.16 Comparaison des réseaux de neurones artificiels avec des modèles alternatifs	28
2.17 Conclusion	28
Chapitre 3 : Techniques de Décomposition : Ondelettes Discrètes, Modes Empiriques et Variationnels	29
3.1 Introduction	29
3.2 Transformation en ondelettes	29
3.2.1 Historique de la transformation en ondelettes	29
3.2.2 Transformation en Ondelettes Continue	30
3.2.3 Transformation en Ondelettes Discrètes	31
3.2.3.1 Principe de la transformation en ondelettes	33
3.2.3.2 Propriétés des familles d'ondelettes	33
3.2.3.3 Différentes familles d'ondelettes	35
3.2.3.4 Avantages de la transformation en ondelettes discrètes	36
3.3 Décomposition en Modes Empiriques	37
3.3.1 Principe de la décomposition en modes empiriques	37
3.3.2 Propriétés de la décomposition en modes empiriques	39
3.3.3 Avantages de la décomposition en modes empiriques	39
3.4 Décomposition en Modes Variationnels	39
3.4.1 Principe de la décomposition en modes variationnels	40
3.4.2 Propriétés de la décomposition en modes variationnels	41
3.4.3 Avantages de la décomposition en modes variationnels	43
3.5 Comparaison entre les techniques utilisées de décomposition	43
3.6 Conclusion	44
Chapitre 4 : Présentation de la Station d'Epuration d'Aïn Beïda	45
4.1 Introduction	45
4.2 Localisation géographique de la station	45
4.3 Caractéristiques techniques de la station	46
4.4 Présentation des procédés de traitement de la station	46
4.4.1.2 By-pass	47
4.4.1.3 Dégrilleur grossier	48
4.4.1.4 Poste de relevage	48
4.4.1.5 Dégrilleur fin	48
4.4.1.6 Dessableur-déshuileur	48
4.4.1.7 Bassin biologique	49
4.4.1.8 Décanteur	50
4.4.1.9 Bassin de chloration	50
4.4.2 Filière de traitement des boues	51
4.4.2.1 Epaississeur primaire	51
4.4.2.2 Digesteur aérobie	51
4.4.2.3 Epaississeur secondaire	51

4.4.2.4 Lits de séchage	51
4.4.3 Laboratoire	51
4.4.4 Salle de commande	52
4.5 Modélisation par RNA et techniques de décomposition dans les STEP (Études Antérieures)	52
4.6 Conclusion	54
Chapitre 5 : Analyse des Résultats des Modèles Prédicatifs pour la STEP d'Aïn Beïda	55
5.1 Introduction	55
5.2 Description et collecte de données	55
5.3 Prétraitement des données	55
5.4 Analyse statistique des données	56
5.5 Critères d'évaluation des performances	60
5.6 Construction des modèles	62
5.6.1 Modélisation de la configuration des modèles à sortie unique	63
5.6.2 Modélisation des modèles à sorties multiples	89
5.7 Conclusion	92
Conclusion générale	93
Références bibliographiques	95

Liste des figures

Figure 2.1: La structure d'un neurone biologique	16
Figure 2.2: Modèle mathématique d'un neurone artificiel	17
Figure 2.3: Phénomène du sous-apprentissage et du sur-apprentissage	25
Figure 3.1: Présentation de différentes ondelettes mères	35
Figure 3.2: Organigramme de la technique de décomposition en modes empiriques	38
Figure 3.3: Organigramme de la technique de décomposition en modes variationnels	42
Figure 4.1: Localisation de la station d'épuration d'Aïn Beïda	45
Figure 4.2: Schéma de procédé de la station d'épuration d'Aïn Beïda	47
Figure 5.1: Nuage de points des modèles utilisés pour la prédiction de la DBO ₅	65
Figure 5.2: Diagramme de Taylor des différentes modèles utilisés pour la prédiction de la DBO ₅	66
Figure 5.3: Nuage de points des modèles utilisés pour la prédiction de la DCO	69
Figure 5.4: Diagramme de Taylor des différentes modèles utilisés pour la prédiction de la DCO	70
Figure 5.5: Nuage de points des modèles utilisés pour la prédiction du MES	72
Figure 5.6: Diagramme de Taylor des différentes modèles utilisés pour la prédiction de MES	73
Figure 5.7: Nuage de points des modèles utilisés pour la prédiction du NT	76
Figure 5.8: Diagramme de Taylor des différentes modèles utilisés pour la prédiction du NT	77
Figure 5.9: Nuage de points des modèles utilisés pour la prédiction de PO ⁻³ ₄	80
Figure 5.10: Diagramme de Taylor des différentes modèles utilisés pour la prédiction de PO ⁻³ ₄	81
Figure 5.11: Nuage de points des modèles utilisés pour la prédiction de l'IB	84
Figure 5.12: Diagramme de Taylor des différentes modèles utilisés pour la prédiction de l'IB	85
Figure 5.13: Nuage de points des modèles utilisés pour la prédiction de QBS	88
Figure 5.14: Diagramme de Taylor des différentes modèles utilisés pour la prédiction du QBS	89

Liste des tableaux

Tableau 1.1: Variables d'état pour MBA1	7
Tableau 1.2: Expression des taux de traitement pour MBA1	9
Tableau 2.1: Analogie entre le neurone biologique et le neurone formel	18
Tableau 2.2: Principales fonctions de transfert	20
Tableau 3.1: Principales propriétés des familles d'ondelettes	35
Tableau 3.2: Comparaison des différentes techniques de décomposition	43
Tableau 4.1: Différents débits de dimensionnement de la station	46
Tableau 4.2: Concentrations des paramètres de pollution	46
Tableau 4.3: Charge journalière des paramètres de pollution	46
Tableau 5.1: Synthèse statistique des différents paramètres de la station	57
Tableau 5.2: Paramètres choisis des différents modèles des différentes configurations	62
Tableau 5.3: Evaluation des performances des modèles utilisées pour la prédiction de la DBO ₅	63
Tableau 5.4: Evaluation des performances des modèles utilisées pour la prédiction de la DCO	67
Tableau 5.5: Evaluation des performances des modèles utilisés pour la prédiction de MES	70
Tableau 5.6: Evaluation des performances des modèles utilisés pour la prédiction de l'azote totale	74
Tableau 5.7: Evaluation des performances des modèles pour la prédiction de PO ⁻³ ₄	78
Tableau 5.8: Evaluation des performances des différents modèles pour la prédiction IB	82
Tableau 5.9: Evaluation des performances des différents modèles pour la prédiction QBS	86
Tableau 5.10: Evaluation de la performance des différents modèles pour la prédiction des paramètres de qualité des effluents	89

Abréviations

Sigles

AIE : Association internationale de l'eau
AM : Arbre de modèle
Cap : Coefficient d'aplatissement
Cas : Coefficient d'asymétrie
CE : Conductivité électrique
Cv : Coefficient de variation
DBO₅ : Demande biochimique en oxygène
DCO : Demande chimique en oxygène
DME : Décomposition en modes empiriques,
DMV : Décomposition en modes variationnels,
EMA : Erreur moyenne absolue
ENS : Coefficient d'efficacité de Nash-Sutcliffe
ET : Ecart-type
FBR : Fonction à base radiale
FMI : Fonction de mode intrinsèque
IA : Intelligence artificielle
IB : Indice de boues
MBA : Modèle de boues activées

Formules chimiques et notations

CaCO₃ : Carbonate de calcium
FeCl₃ : Chlorure de fer
NaOCl : Hypochlorite de sodium
NH₄⁺ : Ammonium
NO₂⁻ : Nitrites; NO₃⁻ : Nitrates, NT: Azote totale
PO₄⁻³ : Ortho-phosphates, PT : Phosphore total
S_{ALK} : Alcalinité
S_I : Matière organique inerte soluble
S_{ND} : Azote organique soluble biodégradable

Liste des symboles

$\| \cdot \|_2^2$: Norme quadratique
 \bar{Y}_o, \bar{Y}_p : Moyenne des valeurs observées, prédites
 $\Phi_k(t)$: La phase croissante définissant la modulation en fréquence
 $h_{0,j}$: Filtre passe-bas, $h_{1,j}$: Filtre passe-haut
 A_j, D_j : Les approximations et détails
 $r_n(t)$: La composante résiduelle
 $u_k(t)$: Composantes modales
 $\psi_{i,j}(t)$: Ondelette fille obtenue par translation et dilatation de l'ondelette mère $\psi(t)$
 $\psi_{i,j}^*(t)$: Conjugué complexe de $\psi_{i,j}(t)$
 ω_k : Fréquences centrales des modes
 ∂_t : Dérivée temporelle

MBA2r : Modèle de boues activées 2 réduit
MES : Matières en suspension
MSM : Modèle à sortie multiple
MSU : Modèle à sortie unique
PEG : Programmation par expression génétique
pH : Potentiel hydrogène
PMC : Perceptron multi couche
QBS : Débit de boue séchée
R : Coefficient de corrélation
R² : Coefficient de détermination
RÉQM : Racine de l'écart quadratique moyenne
REQM : Racine de l'erreur quadratique moyenne
RG : Régression générale
RNA : Réseaux neurones artificiels
RPE : Régression polynomiale évolutionnaire
STEP : Station d'épuration des eaux usées
TO : Transformation en ondelettes
TOC : Transformation en ondelettes continue
TOD : Transformation en ondelettes discrète

S_{NH} : Azote ammoniacal
S_{NO} : Nitrates et nitrites
S_O : Oxygène dissous
S_S : Substrat rapidement biodégradable
X_{BA} : Biomasse autotrophe
X_{BH} : Biomasse hétérotrophe
X_I : Matière organique inerte particulaire
X_{ND} : Azote organique particulaire biodégradable
X_P : Biomasse particulaire inerte
X_S : Substrat lentement biodégradable

$A_k(t)$: l'enveloppe du signal
 $C_i(t)$: Les i -èmes
 $S_{i,j}$: Coefficient d'ondelette calculé à l'échelle j et à la position i .
 $Y_{o,i}$ et $Y_{p,i}$: La valeur observée et prédite
 $d_k(t)$: La moyenne des enveloppes
 $\frac{j}{\pi t}$: Partie imaginaire du noyau de Hilbert
 η : Taux d'apprentissage, η_h : Taux d'hydrolyse
 η_g : Cinétique de Monod
 $x(t)$: Signal observé
 Ω : Momentum
 α : Facteur de régularisation
 $\delta(t)$: Fonction impulsionnelle de Dirac
 $\lambda(t)$: Multiplicateurs de Lagrange

Résumé

Cette thèse vise à appliquer des techniques d'intelligence artificielle (IA) pour modéliser et prédire les performances de la station d'épuration des eaux usées (STEP) de la ville d'Aïn Beïda, en Algérie. Trois types de modèles de réseaux de neurones artificiels (RNA) – Perceptron Multi-Couche (PMC), Fonction à Base Radiale (FBR) et Régression Générale (RG) – ont été combinés avec trois techniques avancées de décomposition : la Transformation en Ondelettes Discrètes (TOD), la Décomposition en Modes Empiriques (DME) et la Décomposition en Modes Variationnels (DMV), afin de développer une approche hybride.

L'objectif principal est de prédire plusieurs paramètres clés liés à la qualité des effluents, notamment la demande biochimique en oxygène après cinq jours (DBO_5), la demande chimique en oxygène (DCO), les matières en suspension (MES), l'azote total (NT), les orthophosphates (PO_4^{3-}), l'indice de boue (IB) et le débit journalier de boues séchées (QBS). Deux configurations fonctionnelles ont été testées : le modèle à sortie unique (MSU) et le modèle à sorties multiples (MSM).

Les résultats obtenus montrent que le MSU fournit des prédictions significativement plus précises et fiables que le modèle à sorties multiples. Les simulations comparatives révèlent que les modèles hybrides surpassent systématiquement les modèles simples, avec des améliorations notables des performances (coefficient de corrélation (R) et coefficient d'efficacité de Nash-Sutcliffe (ENS)) ainsi qu'une réduction significative des erreurs (racine de l'erreur quadratique moyenne (REQM) et erreur moyenne absolue (EMA)).

Parmi les modèles testés, les combinaisons DME-RNPMC, TOD-RNPMC, DMV-RNPMC et TOD-RNFBR ont offert les meilleures performances dans les différentes phases de modélisation. Ces résultats soulignent l'efficacité des approches hybrides pour évaluer les performances des STEP et fournissent des outils précieux pour optimiser leur gestion et leur fonctionnement.

Mots Clés : qualité des effluents, réseaux de neurones artificiels, station d'épuration des eaux usées, techniques de décomposition, modèle à sortie unique, modèle à sorties multiples.

Abstract

This thesis aims to apply artificial intelligence (AI) techniques to model and predict the performance of the wastewater treatment plant (WWTP) in the city of Aïn Beïda, Algeria. Three types of artificial neural network (ANN) models—Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), and General Regression (GR)—were combined with three advanced decomposition techniques: Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD), and Variational Mode Decomposition (VMD) to develop a hybrid approach.

The main objective is to predict several key parameters related to effluent quality, including the five-day biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), total suspended solids (TSS), total nitrogen (TN), orthophosphates (PO_4^{3-}), sludge volume index (SVI), and daily dried sludge flow (QDS). Two functional configurations were tested: the single-output model (SOM) and the multi-output model (MOM).

The results show that the single-output model provides significantly more accurate and reliable predictions than the multi-output model. Comparative simulations reveal that hybrid models consistently outperform simple models, with notable improvements in performance metrics (correlation coefficient (R) and Nash-Sutcliffe efficiency coefficient (NSE)) and significant reductions in errors (root mean square error (RMSE) and mean absolute error (MAE)).

Among the tested models, the DME-MLP, DWT-MLP, VMD-MLP, and DWT-RBF combinations delivered the best performance across different modeling phases. These findings highlight the effectiveness of hybrid approaches in evaluating WWTP performance and provide valuable tools for optimizing their management and operation.

Keywords: effluent quality, artificial neural networks, wastewater treatment plant, decomposition techniques, single-output model, multi-output model.

الملخص

تهدف هذه الأطروحة إلى تطبيق تقنيات الذكاء الاصطناعي (ذ إ) لنمذجة والتنبؤ بأداء محطة تصفية مياه الصرف الصحي (م ت م ص ص) في مدينة عين البيضاء، الجزائر. ثلاثة أنواع من نماذج الشبكات العصبية الاصطناعية (ش ع إ): الطبقات المتعددة الأمامية (ط م أ)، دالة القاعدة الشعاعية (د ق ش) والانحدار المعمّم (إ م) تم دمجها مع ثلاث تقنيات تحلل متقدمة: التحويل الموجي المتقطع (ت م م)، التحلل في الوضع التجريبي (ت و ت) و التحلل في الوضع المتغير (ت و م)، من أجل تطوير نموذج هجين. الهدف الرئيسي هو التنبؤ بالعديد من العوامل الرئيسية المرتبطة بجودة النفايات السائلة، بما في ذلك الطلب على الأكسجين الكيميائي الحيوي بعد 5 أيام (ط أ كج5)، والطلب على الأكسجين الكيميائي (ط أ ك)، المواد الصلبة العالقة (م ص ع)، النيتروجين الكلي (ن ك)، والأورثوفوسفات (أ ف)، مؤشر الحمأة (م ح) والتدفق اليومي للحمأة المجففة (ت ح م). تم اختبار تكوينين وظيفيين: نموذج المخرج الواحد (ن م و) ونموذج متعدد المخارج (ن م م). تظهر النتائج التي تم الحصول عليها أن نموذج المخرج الواحد يوفر تنبؤات أكثر دقة وموثوقية بكثير من نموذج متعدد المخارج. تكشف عمليات المحاكاة المقارنة أن النماذج الهجينة تتفوق باستمرار على النماذج البسيطة، مع تحسينات ملحوظة في الأداء (معامل الارتباط (م إ) ومعامل كفاءة ناش ساتكليف (م ك ن س) بالإضافة إلى انخفاض كبير في الأخطاء (جذر متوسط الخطأ المربع (ج م م) ومتوسط الخطأ المطلق (م م)).

من بين النماذج التي تم اختبارها، قدمت النماذج الهجينة: ت و ت- ش ع ط م أ، ت م م- ش ع ط م أ، ت و م- ش ع ط م أ، ت م م- ش ع د ق ش أفضل الأداء في مراحل النمذجة المختلفة. تؤكد هذه النتائج فعالية النماذج الهجينة لتقييم أداء محطات معالجة مياه الصرف الصحي وتوفير أدوات قيمة لتحسين إدارتها وتشغيلها.

الكلمات المفتاحية: جودة النفايات السائلة، الشبكات العصبية الاصطناعية، محطة تصفية مياه الصرف الصحي، تقنيات التحلل، نموذج المخرج الواحد، نموذج متعدد المخارج.

Introduction générale

Depuis son introduction au début du 20^e siècle, le processus des boues activées (PBA) s'est imposé comme une solution efficace, économique et polyvalente pour l'épuration des eaux usées. Il permet de produire des effluents de haute qualité tout en stabilisant les boues générées. Celles-ci, riches en matière organique et en nutriments, sont souvent utilisées comme amendements agricoles. Cependant, leur gestion soulève des préoccupations, notamment en raison de la présence de métaux lourds et de leur impact environnemental.

L'évaluation des performances des stations d'épuration des eaux usées (STEP) repose sur l'analyse de plusieurs paramètres clés, tels que la demande biochimique en oxygène après cinq jours (DBO_5), la demande chimique en oxygène (DCO) et les matières en suspension (MES). Avec l'augmentation de la population mondiale et l'urbanisation croissante, la gestion des boues est devenue un enjeu crucial pour limiter les émissions de polluants et préserver la santé publique ainsi que l'environnement.

Dans cette recherche, 531 données collectées entre décembre 2014 et avril 2017 à la STEP d'Aïn Beïda, en Algérie, ont été exploitées pour développer des modèles prédictifs basés sur des réseaux de neurones artificiels (RNA). Ces données ont subi des prétraitements visant à réduire les fluctuations, éliminer les valeurs extrêmes et normaliser les variables. Des RNA simples et hybrides, combinés à des techniques de décomposition avancées telles que la transformée en ondelettes discrète (TOD), la décomposition en modes empiriques (DME) et la décomposition en modes variationnels (DMV), ont été utilisés pour prédire plusieurs paramètres essentiels, notamment la qualité des effluents (DBO_5 , DCO, MES, azote total (NT), orthophosphates (PO_4^{3-})), l'indice de boue (IB) et le débit des boues séchées (QBS). L'ensemble des données a été aléatoirement divisé en deux sous-groupes : 80 % pour l'entraînement des modèles et 20 % pour leur validation.

L'objectif principal de cette thèse est de contribuer à l'amélioration de la gestion des STEP en explorant l'application des techniques d'intelligence artificielle (IA). Les modèles développés visent à optimiser les processus de traitement et à évaluer la pertinence de ces approches innovantes dans le domaine du traitement des eaux usées. Plus spécifiquement, cette étude poursuit les objectifs suivants :

1. Développer des modèles de RNA capables de prédire les paramètres de qualité des effluents, l'IB et le QBS.
2. Coupler les RNA avec des techniques de décomposition (TOD, DME, DMV) afin d'améliorer la précision des prédictions.
3. Évaluer les performances des modèles en s'appuyant sur des critères statistiques et des analyses comparatives.

Afin d'atteindre ces objectifs, cette thèse est structurée en cinq chapitres :

- **Chapitre 1** : Historique, objectifs et principes de la modélisation des stations d'épuration, avec un accent particulier sur les modèles basés sur les boues activées.
- **Chapitre 2** : Description détaillée des RNA, de leurs principes, caractéristiques et domaines d'application.
- **Chapitre 3** : Exploration des techniques de décomposition des données, incluant la TOD, la DME et la DMV, avec une analyse de leurs propriétés respectives.
- **Chapitre 4** : Présentation de la STEP d'Aïn Beïda et de ses caractéristiques.
- **Chapitre 5** : Analyse des résultats obtenus à partir des modèles développés pour la STEP d'Aïn Beïda, accompagnée d'une comparaison des performances des approches simples et hybrides.

Enfin, la conclusion générale résume les principales contributions de cette recherche, met en évidence les avantages des techniques d'intelligence artificielle dans le domaine des stations d'épuration des eaux usées et propose des perspectives pour des travaux futurs.

Chapitre 1 : Modèles de Boues Activées dans les Stations d’Epuraton

1.1 Introduction

Face à la croissance démographique et à la raréfaction des ressources en eau, une gestion optimisée des eaux usées est devenue une nécessité absolue. Parmi les technologies de traitement disponibles, le procédé des boues activées s'est imposé comme une référence incontournable en raison de son efficacité et de sa polyvalence. Il permet non seulement d'éliminer les polluants présents dans les eaux usées, mais aussi de récupérer des éléments essentiels tels que l'azote et le phosphore, contribuant ainsi à une gestion plus durable des ressources.

Une étape clé de ce procédé est la clarification secondaire, qui assure la séparation des boues biologiques de l'effluent traité (Chen et al., 2023). Cependant, la complexité des phénomènes biologiques impliqués rend sa modélisation particulièrement délicate. En effet, la nature vivante des micro-organismes, la diversité des réactions biochimiques et les limites inhérentes aux mesures expérimentales représentent autant de défis pour l'élaboration de modèles fiables et précis. De plus, les variations temporelles des paramètres et la complexité des processus biologiques à l'échelle microscopique dépassent souvent les capacités descriptives des modèles macroscopiques classiques.

Malgré ces contraintes, la modélisation demeure un outil indispensable pour comprendre les mécanismes fondamentaux du procédé de boues activées, optimiser son fonctionnement et prédire son comportement en cas de perturbations. En simplifiant la réalité, les modèles permettent d'analyser les interactions entre les différents paramètres du système et d'évaluer l'impact de diverses stratégies de gestion.

Pour relever les défis liés à la complexité des processus biologiques et à l'incertitude des mesures expérimentales, il est essentiel de s'appuyer sur une compréhension approfondie des bases historiques et conceptuelles des modèles de boues activées. Cette approche permet d'apprécier les avancées technologiques dans leur contexte et d'identifier les solutions développées pour améliorer la performance des stations d'épuration.

Ce chapitre est structuré comme suit : il retrace l'historique des modèles de boues activées, définit les objectifs de la modélisation, décrit les différentes approches existantes et explore les techniques alternatives de modélisation.

1.2 Historique de la modélisation des procédés de boues activées

La modélisation des procédés de boues activées (PBA) s'est développée en réponse aux besoins croissants d'optimisation et de contrôle des stations d'épuration. L'objectif initial était de concevoir des outils prédictifs permettant de mieux comprendre le comportement des micro-organismes et les interactions complexes régissant le traitement biologique des eaux usées.

Les premières tentatives de modélisation, initiées dans les années 1960, ont cherché à représenter de manière simplifiée les variations de concentration des substrats et de la biomasse (McKinney, 1962). Toutefois, la complexité des PBA et la variabilité des conditions opérationnelles ont rapidement mis en évidence les limites de ces modèles statiques (Smith, 1969). Progressivement, les recherches ont évolué, aboutissant à des modèles plus sophistiqués intégrant des dimensions écologiques, structurales et cinétiques.

Les années 1970 ont marqué un tournant avec l'émergence de modèles incorporant des relations entrée-sortie, des structures d'âge et des fonctions de transfert (Lacroix et Bloodgood, 1972 ; Ranta et al., 1977 ; Berthouex et al., 1978). Ces approches ont permis de mieux appréhender l'influence de facteurs tels que les charges organiques, la disponibilité en oxygène et les conditions environnementales sur les performances des PBA (Kessler et Nicholas, 1985).

Parallèlement, la question du gonflement des boues, un phénomène affectant la performance des stations d'épuration, a attiré l'attention des chercheurs. Des études ont mis en évidence le rôle de l'aération, de la nature des charges organiques et des conditions hydrodynamiques dans la survenue de ce phénomène (Ullrich et Smith, 1951 ; Chen, 1993).

Dans les années 1980, l'Association Internationale de l'Eau (AIE) a lancé un programme de standardisation des modèles de boues activées, conduisant au développement des modèles MBA (Modèles de Boues Activées). Les modèles MBA1 et MBA2, élaborés par Henze et ses collaborateurs, ont constitué une avancée majeure en proposant une représentation détaillée des processus biologiques fondamentaux, notamment la dégradation de la matière organique, la nitrification et la dénitrification (Henze et al., 1987 ; Henze et al., 1995). Ces modèles ont été largement adoptés et régulièrement améliorés pour intégrer les nouvelles connaissances scientifiques.

Le MBA2 réduit a introduit des perfectionnements significatifs, notamment en intégrant la dynamique de croissance et de déclin des micro-organismes impliqués dans la déphosphatation biologique, ainsi qu'une modélisation plus détaillée des processus anoxiques (Henze et al., 1999).

Le MBA3 a marqué une évolution importante en distinguant clairement les processus de stockage des substrats avant leur utilisation pour la croissance bactérienne (Gujer et al., 1999). Contrairement au MBA1, il a remplacé le concept de mort-régénération par celui de respiration endogène, simplifiant ainsi la modélisation de la mortalité bactérienne.

Ces avancées ont permis d'améliorer la précision des simulations, notamment en ce qui concerne les variations des concentrations de phosphate et de nitrate dans des environnements complexes (Du et al., 1999).

1.3 Objectifs de la modélisation des procédés de boues activées

La modélisation est un outil essentiel pour la conception, l'optimisation et la gestion des stations d'épuration. Elle permet une meilleure compréhension du fonctionnement de ces systèmes complexes, une évaluation précise de l'impact des modifications des paramètres opérationnels et une amélioration des performances globales du traitement. Selon Hauduc (2010), la modélisation permet notamment de :

- Simuler la croissance de la biomasse bactérienne et l'évolution des substrats ;
- Optimiser la qualité des effluents, réduire le volume de boues et minimiser la consommation énergétique ;
- Anticiper l'efficacité du traitement face aux variations des conditions opératoires ;
- Dimensionner les infrastructures de manière optimale lors de la conception ou de la réhabilitation d'une station d'épuration.

1.4 Modélisation des procédés de boues activées dans les stations d'épuration

L'évolution des systèmes de traitement des eaux usées s'appuie aujourd'hui sur une approche scientifique avancée grâce à la modélisation. En intégrant des modèles mathématiques décrivant les processus biologiques, physiques et chimiques, il est possible de simuler de manière dynamique le fonctionnement des stations d'épuration. Les modèles de boues activées (MBA), développés par l'Association Internationale de l'Eau (AIE), constituent désormais des références incontournables. Ils permettent d'améliorer la compréhension des mécanismes en jeu et d'optimiser les performances des installations (Henze et al., 1987 ; Henze et al., 1995 ; Henze et al., 1999 ; Gujer et al., 1999).

1.4.1 Modèle de Boue Activée 1

Le Modèle de Boues Activées 1 (MBA1), développé par Henze et al. (1987), constitue une référence majeure dans la modélisation des procédés de traitement des eaux usées. Il fournit une représentation détaillée des processus biologiques impliqués dans l'élimination des matières azotées et carbonées, notamment l'oxydation du carbone, la nitrification et la dénitrification. Ce modèle repose sur une approche cinétique et stœchiométrique rigoureuse, décrivant les interactions complexes entre les différents composants du système.

Le MBA1 s'appuie sur treize (13) variables d'état, classées en plusieurs catégories :

1. Substrats carbonés:

- Demande Chimique en Oxygène (DCO) biodégradable, subdivisée en :
 - Substrat rapidement biodégradable (S_s),
 - Substrat lentement biodégradable (X_s).

2. Biomasses microbiennes :

- Biomasse hétérotrophe (X_{BH}) : responsable de l'oxydation de la matière organique et de la dénitrification,
- Biomasse autotrophe (X_{BA}) : responsable de la nitrification,
- La croissance des biomasses est modélisée par la loi de Monod,
- Matière organique particulaire inerte (X_P), issue de la mortalité des biomasses.

3. Composés azotés :

- Azote organique particulaire biodégradable (X_{ND}),
- Azote organique soluble biodégradable (S_{ND}),
- Azote ammoniacal (S_{NH}),
- Nitrates et nitrites (S_{NO}).

4. Oxygène dissous :

- Concentration en oxygène dissous (S_O).

5. Matières biologiquement inertes :

- Fraction non biodégradable de la matière organique, divisée en :
 - Matière organique inerte soluble (S_I),
 - Matière organique inerte particulaire (X_I),
 - Alcalinité (S_{ALK}).

Le modèle MBA1 considère ainsi sept substrats carbonés et six substrats azotés, définis comme paramètres fondamentaux. Le tableau 1.1 détaille la répartition et l'évolution de ces matières dans le système modélisé. Les interactions entre ces variables sont régies par différents processus biologiques et chimiques, décrits de manière détaillée par Dispan (2003). Ces mécanismes sont essentiels pour comprendre le fonctionnement des stations d'épuration et constituent une base solide pour l'intégration d'approches modernes d'optimisation des procédés.

Tableau 1.1: Variables d'état pour MBA1 (Henze et al., 1987)

Description de la variable d'état	Symbole de l'État	Unités de mesure
Matière organique soluble inerte	S_I	g.DCO/m ³
Substrat rapidement biodégradable	S_S	
Matière organique particulaire inerte	X_I	
Substrat lentement biodégradable	X_S	
Biomasse active hétérotrophe	$X_{B,H}$	
Biomasse active autotrophe	$X_{B,A}$	
Matière organique particulaire résultant de la décomposition de la biomasse	X_P	
Oxygène dissous	S_O	g O ₂ /m ³
Nitrate et nitrite	S_{NO}	g N/m ³
Azote NH ₄ ⁺ et NH ₃	S_{NH}	
Azote organique soluble biodégradable	S_{ND}	
Azote organique particulaire biodégradable	X_{ND}	
Alcalinité	S_{ALK}	mol HCO ₃ ⁻ /m ³

➤ **Croissance aérobie des bactéries hétérotrophes**

Les substrats facilement biodégradables (S_S) constituent la principale source d'énergie pour la croissance des bactéries hétérotrophes, tandis que les substrats lentement biodégradables (X_S) sont progressivement hydrolysés avant d'être oxydés en CO₂. Ce processus suit une cinétique de Monod, dans laquelle la disponibilité des substrats influence directement la vitesse de croissance bactérienne. Par ailleurs, l'ammonium (S_{NH}) est utilisé comme source d'azote et incorporé dans la biomasse en cours de formation. La concentration en S_S et X_S peut ainsi constituer un facteur limitant pour la prolifération des bactéries hétérotrophes.

➤ **Croissance anoxique des bactéries hétérotrophes**

En conditions anoxiques, les bactéries hétérotrophes exploitent les nitrates (S_{NO}) comme accepteurs d'électrons pour dégrader les substrats biodégradables (S_S). Ce mécanisme, connu sous le nom de dénitrification, permet non seulement la croissance bactérienne, mais entraîne également la libération d'azote gazeux (N₂), contribuant ainsi à l'élimination de l'azote des effluents. La cinétique de Monod s'applique également dans ce contexte, avec un facteur correctif ($\eta_g < 1$) qui traduit une croissance plus lente en absence d'oxygène moléculaire. Ce facteur reflète la capacité limitée de certaines bactéries à utiliser les nitrates comme accepteurs d'électrons, réduisant ainsi leur taux de croissance comparé aux conditions aérobies.

➤ **Croissance aérobie des micro-organismes autotrophes**

Les micro-organismes autotrophes, principalement des bactéries nitrifiantes, assurent l'oxydation de l'ammoniac (S_{NH}) en nitrates (S_{NO}), un processus clé du cycle de l'azote, connu sous le nom de nitrification. Cette réaction se déroule en deux étapes successives :

1. L'oxydation de l'ammoniac (NH_4^+) en nitrites (NO_2^-) par les bactéries du genre Nitrosomonas.
2. L'oxydation des nitrites en nitrates (NO_3^-) par les bactéries du genre Nitrobacter.

Ce processus entraîne une faible production de biomasse autotrophe, mais il s'accompagne d'une consommation élevée d'oxygène, essentielle pour la synthèse énergétique. De plus, la nitrification influence fortement l'alcalinité du milieu, car elle entraîne une acidification progressive due à la consommation d'ions bicarbonates.

➤ **Mortalité des bactéries hétérotrophes et autotrophes**

La mortalité des bactéries joue un rôle clé dans le cycle de la matière organique au sein des systèmes de boues activées. Lorsqu'elles meurent, les bactéries libèrent une fraction de leur biomasse sous forme de matière organique particulaire inerte (X_P), immédiatement disponible pour d'autres processus biologiques. Une autre fraction est progressivement dégradée et incorporée aux substrats lentement biodégradables (X_S), influençant ainsi la dynamique globale du système.

➤ **Hydrolyse de la matière organique particulaire**

L'hydrolyse est une étape clé dans la dégradation de la matière organique lentement biodégradable (X_S), permettant sa conversion en substrat rapidement biodégradable (S_S). Ce processus est modélisé comme une réaction de surface, se produisant aussi bien en conditions aérobies qu'anoxiques.

Cependant, en milieu anoxique, le taux d'hydrolyse est réduit en raison du facteur correctif ($\eta_h < 1$), traduisant une efficacité moindre dans ces conditions.

La cinétique d'hydrolyse dépend fortement de l'équilibre entre la biomasse hétérotrophe (X_B) et le substrat (X_S) :

- Lorsque la biomasse est abondante ($X_B \gg X_S$), l'hydrolyse suit une cinétique de premier ordre, où le taux de transformation est proportionnel à la concentration en substrats.
- À l'inverse, en présence d'un excès de matière organique lentement biodégradable, le processus peut atteindre un état de saturation, limitant ainsi l'efficacité de l'hydrolyse.

Ce mécanisme joue un rôle crucial dans le traitement biologique des eaux usées, influençant directement la biodisponibilité des substrats pour les bactéries hétérotrophes et, par conséquent, les performances du procédé de boues activées.

➤ **Hydrolyse de l'azote organique**

L'azote organique particulaire biodégradable (X_{ND}) subit un processus d'hydrolyse, conduisant à sa conversion en azote organique soluble (S_{ND}). Cette transformation, modélisée par une réaction d'hydrolyse, est un élément clé du cycle de l'azote dans les systèmes de boues activées.

Ce processus est crucial, car il permet de libérer l'azote organique, qui sera ensuite converti en ammonium (S_{NH}) lors de l'ammonification. Cette dernière étape rend l'azote disponible pour les processus biologiques

ultérieurs, notamment la nitrification et la dénitrification, influençant ainsi l'efficacité globale du traitement des eaux usées.

➤ **Ammonification de l'azote organique soluble**

L'azote organique soluble (S_{ND}) est transformé en ammonium (S_{NH}) sous l'action de la biomasse hétérotrophe active. Ce processus joue un rôle fondamental dans le cycle de l'azote, car il conditionne la disponibilité de l'ammonium pour les étapes ultérieures de nitrification et de dénitrification. Par ailleurs, cette transformation affecte l'alcalinité du milieu, ce qui peut influencer l'équilibre des autres réactions biologiques et chimiques dans le système.

Les équations du modèle de boues activées 1 (MBA1) décrivent en détail ces transformations, dont un récapitulatif est présenté dans le tableau 1.2.

Tableau 1.2: Expression des taux de traitement pour MBA1 (Henze et al., 2006)

Taux de traitement	Expression mathématique
Croissance hétérotrophe aérobie	$\mu_H = \left(\frac{S_S}{K_S + S_S}\right) \left(\frac{S_O}{K_{OH} + S_O}\right) X_{BH}$
Croissance hétérotrophe anaérobie	$\mu_H = \left(\frac{S_S}{K_S + S_S}\right) \left(\frac{K_{OH}}{K_{OH} + S_O}\right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}}\right) \eta_g X_{BH}$
La croissance autotrophe aérobie	$\mu_A = \left(\frac{S_{NH}}{K_{NH} + S_{NH}}\right) \left(\frac{S_O}{K_{OA} + S_O}\right) X_{BA}$
La lyse hétérotrophe	$b_H X_{BH}$
La lyse autotrophe	$b_A X_{BA}$
L'ammonification	$k_A S_{ND} X_{BH}$
Hydrolyse de la matière organique particulaire	$k_h \left(\frac{X_S/X_{BH}}{K_X + X_S/X_{BH}}\right) \left(\left(\frac{S_O}{K_{OH} + S_O}\right) + \eta_h \left(\frac{K_{OH}}{K_{OH} + S_O}\right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}}\right)\right) X_{BH}$
Hydrolyse de l'azote organique particulaire	$k_h \left(\frac{X_S/X_{BH}}{K_X + X_S/X_{BH}}\right) \left(\left(\frac{S_O}{K_{OH} + S_O}\right) + \eta_h \left(\frac{K_{OH}}{K_{OH} + S_O}\right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}}\right)\right) X_{BH} \left(\frac{X_{ND}}{X_S}\right)$

Le MBA1 modélise avec précision les principaux processus biologiques des boues activées, notamment la dégradation de la matière organique, la nitrification et la dénitrification. Toutefois, une limite majeure réside dans l'absence de prise en compte de l'élimination du phosphore, comme le soulignent Gernaey et al. (2004). Cette lacune rend le modèle inadapté aux contextes nécessitant un contrôle strict du phosphore afin de prévenir des phénomènes tels que l'eutrophisation des milieux aquatiques.

Pour pallier cette limitation, plusieurs versions simplifiées du MBA1 ont été développées (Steffens et al., 1997). Ces modèles allégés visent à réduire le nombre de variables non mesurables, tout en conservant une précision suffisante pour les simulations appliquées. L'approche proposée par Bassompierre (2007) illustre bien cette démarche, en adaptant le MBA1 à des contextes spécifiques et à des analyses plus opérationnelles.

1.4.2 Modèle de Boues Activées 2

Le modèle de boues activées 2 (MBA2), proposé par Henze et al. (1995), étend les fonctionnalités du MBA1 en intégrant des variables d'état et des processus supplémentaires. Il permet notamment de mieux prendre en compte la déphosphatation biologique et physico-chimique. Ce modèle offre une représentation plus précise des interactions entre les phases anaérobies et aérobies, essentielles au cycle du phosphore.

Les principales contributions du MBA2 sont les suivantes :

➤ Description interne de la biomasse

Le phosphore est intégré dans la structure des cellules bactériennes, reflétant ainsi les mécanismes biologiques réels.

➤ Précipitation chimique du phosphore

En plus des processus biologiques, deux mécanismes chimiques sont introduits pour simuler la précipitation du phosphore.

La déphosphatation repose sur l'activité des bactéries hétérotrophes qui, en conditions anaérobies, dégradent leurs réserves de polyphosphates pour accumuler des substrats carbonés, libérant ainsi du phosphore dans le milieu. Sous conditions aérobies, ces bactéries réassimilent le phosphore et suraccumulent les polyphosphates dans leur structure intracellulaire, atteignant des niveaux supérieurs à ceux initiaux (Henze et al., 1995).

Malgré ses avancées, le MBA2 ne parvient pas à représenter de manière satisfaisante les dynamiques complexes de croissance et de déclin des bactéries impliquées dans l'élimination du phosphore. Pour répondre à ce défi, une version améliorée, le MBA2 réduite, a été développée afin d'affiner la modélisation de ces mécanismes.

1.4.3 Modèle de Boues Activées 2 Réduit

Le modèle de boues activées 2 réduit (MBA2r), proposé par Henze et al. (1999), améliore le MBA2 en intégrant deux processus supplémentaires permettant de mieux modéliser les dynamiques de croissance et de déclin des micro-organismes impliqués dans la déphosphatation. Ce modèle offre une représentation plus précise des interactions entre le phosphore et le nitrate en tenant compte de la structure interne de la biomasse.

Les principales améliorations du MBA2r incluent :

➤ Nouveaux facteurs stœchiométriques et de conversion

Ces facteurs prennent en compte la complexité accrue de la structure cellulaire interne.

➤ Concentration individuelle des composés

Les calculs des concentrations de phosphore et de nitrate sont ajustés pour intégrer des processus biologiques spécifiques et une notation matricielle.

Ces ajustements rendent le MBA2r plus performant pour modéliser les dynamiques réelles des systèmes biologiques complexes, notamment en matière de gestion du phosphore.

Cependant, bien que le MBA2r améliore significativement la modélisation du phosphore, il reste principalement axé sur la déphosphatation et ne traite pas de manière exhaustive d'autres aspects, tels que le stockage des substrats ou les mécanismes de respiration endogène. Ces limites ont conduit à l'élaboration du MBA3, une version plus globale et flexible.

1.4.5 Modèle de Boues Activées 3

Le modèle de boues activées 3 (MBA3), développé par Gujer et al. (1999), constitue une évolution du MBA1 en intégrant une nouvelle approche pour le stockage des substrats et en révisant le concept de mortalité bactérienne. Contrairement au MBA2, ce modèle se focalise sur la pollution carbonée et azotée, sans inclure directement le traitement du phosphore.

Les améliorations majeures du MBA3 incluent :

- **Stockage des substrats**

Le substrat rapidement biodégradable (S_S) est d'abord transformé en composés de stockage intracellulaires avant d'être assimilé par la biomasse.

- **Révision de la mortalité bactérienne**

La respiration endogène remplace le concept de mort-régénération du MBA1, simplifiant ainsi le modèle.

- **Fractionnement simplifiée de l'azote**

Les composants S_{ND} et X_{ND} sont supprimés, et l'azote est désormais fractionné proportionnellement à la demande chimique en oxygène (DCO).

De plus, le MBA3 distingue clairement les phénomènes de croissance bactérienne et de stockage des polymères intracellulaires. Cette distinction permet une meilleure modélisation des comportements microbiens dans des conditions environnementales instables, telles que celles rencontrées dans les systèmes anoxiques ou aérobies.

En réduisant la complexité du modèle, les équations et les paramètres sont simplifiés, ce qui facilite son utilisation pour des simulations rapides et des analyses de scénarios dans les stations de traitement des eaux usées.

Ces évolutions successives des modèles à boues activées illustrent la nécessité d'une approche modulaire permettant d'adapter la modélisation aux spécificités de chaque processus biologique et chimique. Le choix entre MBA1, MBA2, MBA2r, et MBA3 dépend des priorités du système étudié, qu'il s'agisse de la gestion du phosphore, du nitrate ou des dynamiques de stockage.

Si les modèles classiques comme les MBA offrent une base solide pour comprendre les processus biologiques, les exigences actuelles en matière de gestion des stations d'épuration nécessitent des approches plus avancées. L'intégration de techniques modernes, telles que la modélisation basée sur les données, les algorithmes d'optimisation et l'intelligence artificielle, permet d'affiner les prédictions et d'améliorer les performances opérationnelles. Ces outils complémentaires s'appuient sur des mesures en temps réel et adaptent les processus aux conditions environnementales variables.

1.5 Techniques Alternatives de Modélisation

Les modèles analytiques, comme ceux décrivant les processus biologiques de traitement abordés précédemment, peuvent désormais être simulés ou complétés grâce à des outils d'intelligence artificielle (IA) (Gernaey et al., 2004). Ces approches modernes intègrent diverses techniques de modélisation basées sur les données, parmi lesquelles :

- **Régression linéaire multivariée (RLM)**

Permet d'établir des relations entre plusieurs variables d'entrée et de sortie à partir de données mesurées.

- **Analyse en composantes principales (ACP)**

Utilisée pour réduire la dimensionnalité des données tout en préservant l'essentiel des informations.

- **Machines à vecteurs de support (MVS)**

Modèle statistique performant pour l'analyse de données complexes et non linéaires.

- **Système d'inférence adaptative neuro-floue (SIANF)**

Combine la logique floue et les réseaux neuronaux pour modéliser des processus incertains ou mal définis.

- **Réseaux neuronaux artificiels (RNA)**

Reconnus pour leur capacité à prédire les paramètres des processus de traitement dans les stations d'épuration.

Les études montrent que les réseaux neuronaux artificiels surpassent souvent les autres méthodes en termes de fiabilité et de précision pour prédire les paramètres clés du traitement biologique. Ces modèles, basés sur des algorithmes d'apprentissage automatique, peuvent traiter de grandes quantités de données, capturer des relations complexes et s'adapter à des environnements changeants.

En complément, des modèles hybrides, combinant un modèle analytique classique et un RNA, se révèlent particulièrement efficaces. Cette approche permet de décrire certaines parties du processus de manière analytique tout en exploitant les RNA pour modéliser des phénomènes plus complexes ou mal compris. Ainsi, elle optimise à la fois la précision de la modélisation et la simplicité d'utilisation, rendant ces outils particulièrement adaptés aux besoins modernes des stations d'épuration.

Ces techniques modernes et hybrides s'inscrivent dans une démarche d'innovation visant à améliorer la gestion des stations d'épuration. Elles permettent de surmonter les limites des modèles classiques tout en

exploitant le potentiel des nouvelles technologies pour atteindre une efficacité opérationnelle accrue et une meilleure durabilité environnementale.

1.6 Conclusion

Les modèles à boues activées constituent un outil essentiel pour la conception, l'optimisation et la gestion des stations d'épuration des eaux usées. En décomposant les processus biologiques et chimiques complexes, ils permettent de mieux comprendre le fonctionnement des systèmes de traitement et d'anticiper leurs performances face aux variations des paramètres opérationnels.

Les MBA, à travers leurs évolutions successives (MBA1, MBA2, MBA2r, et MBA3), ont progressivement intégré de nouvelles dynamiques, telles que la déphosphatation et le stockage des substrats, tout en simplifiant certains aspects pour des applications spécifiques. Cependant, malgré leur efficacité, les limites des MBA classiques, notamment en termes de complexité et d'adaptabilité aux environnements changeants, ont conduit au développement de techniques alternatives de modélisation.

Les approches modernes, basées sur l'exploitation des données en temps réel, telles que les réseaux neuronaux artificiels, vont au-delà des modèles analytiques traditionnels. Ces outils d'intelligence artificielle offrent une capacité accrue à capturer des interactions complexes et à s'adapter à des conditions opérationnelles variées, améliorant ainsi la précision des prédictions.

En parallèle, les techniques de décomposition permettent de simplifier des processus multidimensionnels en réduisant la complexité des données tout en préservant l'essentiel de l'information. Ces approches offrent une perspective complémentaire, facilitant l'intégration de modèles hybrides combinant des bases analytiques et des algorithmes d'apprentissage automatique.

L'avenir de la modélisation des stations d'épuration repose donc sur une synergie entre modèles classiques et outils d'intelligence artificielle, permettant d'optimiser les performances des installations tout en garantissant une gestion plus efficace et durable des ressources.

Chapitre 2 : Réseaux de Neurones Artificiels : Principes, Modèles et Applications

2.1 Introduction

Les réseaux de neurones artificiels (RNA) sont des systèmes informatiques conçus sur la base du fonctionnement biologique du cerveau humain (Krose et Smagt, 1996). Ils se distinguent par leur capacité à modéliser des systèmes complexes et non linéaires en s'appuyant sur des données d'entraînement. Grâce à leur adaptabilité et à leur puissance de calcul, les RNA sont employés dans de nombreux domaines, tels que la reconnaissance de formes, la classification, et la prédiction. Leur performance découle de leur capacité à représenter les relations entre plusieurs variables tout en s'adaptant à des environnements changeants.

Ce chapitre explore en détail le fonctionnement des RNA, en abordant leur structure, leurs principes fondamentaux, leurs architectures et leurs méthodes d'apprentissage. Toutefois, bien que les RNA offrent une flexibilité et une performance impressionnantes, ils présentent également des limites opérationnelles, notamment en termes de sur-apprentissage, de complexité computationnelle et de manque d'interprétabilité. Par ailleurs, une comparaison avec des modèles alternatifs tels que les réseaux bayésiens et d'autres approches probabilistes met en évidence les forces et faiblesses relatives des RNA, permettant ainsi de contextualiser leur utilisation et de guider le choix des outils adaptés aux différentes applications. Ce chapitre propose également une transition vers des techniques avancées, notamment les techniques de décomposition, pour améliorer l'efficacité et la robustesse des RNA dans des environnements complexes et dynamiques.

2.2 Historique des Réseaux Neurones Artificiels

L'étude des RNA a commencé en 1943 lorsque McCulloch et Pitts ont proposé un modèle simplifié du neurone biologique, nommé neurone formel. Ils ont montré théoriquement que des réseaux de neurones simples pouvaient réaliser des opérations logiques, arithmétiques et symboliques complexes. En 1949, Hebb a formulé une théorie clé sur l'apprentissage, qui est devenue une base essentielle pour l'évolution des RNA. En 1957, Rosenblatt a développé le perceptron, un type de réseau de neurones qui a permis des applications pratiques. Durant les années 1960, Widrow et Hoff ont proposé un algorithme pour la formation de réseaux neurones linéaires adaptatifs, qui partageaient des capacités similaires à celles du perceptron.

Cependant, en 1969, un ouvrage de Minsky et Papert a remis en question les avancées de ces années-là. Ils ont mis en évidence les limitations des réseaux proposés par Rosenblatt, Widrow et Hoff, limitations en grande partie liées aux capacités de calcul limitées des machines de l'époque, mais aussi à des interprétations erronées des théories sous-jacentes. Ces découvertes ont conduit à un ralentissement de la recherche dans le domaine des RNA.

Malgré ce recul, Kohonen et Anderson ont proposé des architectures similaires dès 1972. En 1990, Kohonen a publié ses célèbres cartes auto-organisatrices, qui constituaient une avancée majeure dans l'étude des réseaux de neurones. En parallèle, Grossberg (1976) a également proposé d'autres réseaux auto-organiseurs.

L'apparition de l'algorithme de rétropropagation a marqué un tournant dans le développement des RNA. Selon Rumelhart (1986), cet algorithme a gagné en popularité en raison de ses performances remarquables, contribuant ainsi à un renouveau dans l'activité des réseaux de neurones. Bien que Parker et LeCun aient exploré cette approche en 1985, Werbos l'avait déjà décrit dans sa thèse en 1974 (Viana Di Prisco, 1984). Ces avancées ont ouvert la voie à une adoption croissante des RNA dans la modélisation des systèmes complexes, en surmontant certaines limitations initiales liées à la capacité de calcul des machines.

2.3 Principe des Réseaux de Neurones Artificiels

Le principe des réseaux de neurones artificiels repose sur la modélisation de systèmes complexes à l'aide de données d'apprentissage. Les RNA imitent les mécanismes du cerveau humain pour établir des relations entre des variables d'entrée et de sortie. Leur objectif principal est de créer un système de raisonnement statistique, capable de prendre des décisions ou de faire des prédictions en se basant sur des données d'entraînement.

Les réseaux de neurones artificiels fonctionnent en modélisant des systèmes complexes grâce à l'utilisation de données d'apprentissage. Ils reproduisent les processus du cerveau humain pour identifier des liens entre des variables d'entrée et de sortie. Leur but principal est de développer un système de raisonnement statistique, apte à prendre des décisions ou à effectuer des prédictions en s'appuyant sur des données d'entraînement.

Un réseau de neurones artificiels fonctionne grâce à un processus d'apprentissage par l'expérience, où il ajuste ses paramètres internes pour optimiser ses performances. Les décisions ou classifications produites par le RNA sont basées sur une probabilité de ressemblance avec des catégories qu'il a apprises.

Les principales étapes du fonctionnement d'un RNA sont les suivantes :

- ✓ Entrée des données : les données brutes sont fournies au modèle ;
- ✓ Analyse des données : les données sont pondérées et transformées via des fonctions d'activation ;
- ✓ Sortie : une classification ou une prédiction est réalisée avec une probabilité de certitude associée.

2.4 Réseaux de neurones artificiels

Un réseau de neurones artificiels est un modèle computationnel inspiré du fonctionnement des neurones biologiques. Il se compose de nombreuses unités de traitement simples, appelées neurones artificiels, interconnectées pour former une structure réseau. Ces neurones artificiels sont capables d'apprendre des relations complexes à partir de données d'exemple, ce qui rend les RNA adaptés à des tâches comme la

classification, la prédiction, et la reconnaissance de motifs non linéaires. Un RNA typique est constitué de trois types de couches (Demuth et Beale, 1998) :

- La couche d'entrée : elle reçoit les données brutes. Aucune transformation des données n'est réalisée à ce niveau. Les entrées doivent être pertinentes, car des informations manquantes ou inutiles peuvent introduire du bruit et nuire à la performance du modèle.
- Les couches cachées : elles effectuent des transformations complexes grâce aux fonctions d'activation, qui permettent au RNA de capturer des relations non linéaires dans les données.
- La couche de sortie : elle fournit le résultat final du traitement, qu'il s'agisse d'une classification, d'une prédiction ou d'une autre sortie.

Les fonctions d'activation utilisées dans les couches cachées et la couche de sortie jouent un rôle crucial pour introduire la non-linéarité dans le modèle, rendant le RNA capable de modéliser des systèmes complexes (Haykin, 1999).

2.4.1 Neurone biologique

Un neurone biologique est l'unité de base du système nerveux humain. Il est conçu pour traiter et transmettre des signaux électriques et des messages nerveux. Le cerveau humain contient plusieurs milliards de neurones interconnectés, chacun étant relié à des milliers d'autres neurones (Park, 1996). Un neurone biologique est composé de quatre parties principales (Figure 2.1), chacune jouant un rôle clé dans le traitement des informations (Gurney, 1997) :

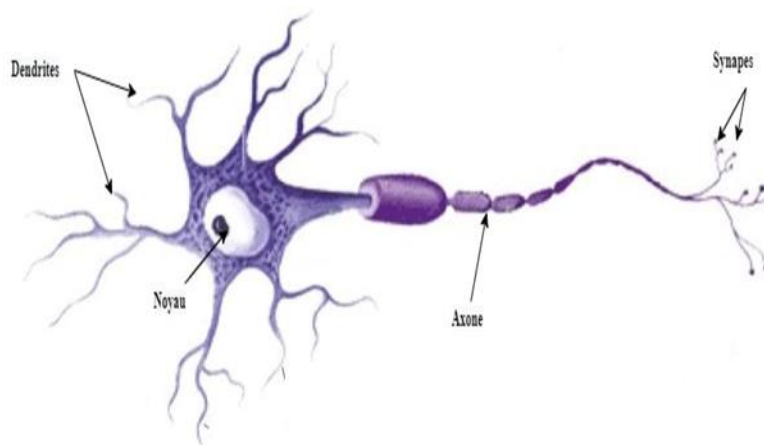


Figure 2.1: La structure d'un neurone biologique (Parizeau, 2004)

➤ Dendrites

Les dendrites sont des extensions tubulaires qui servent de principaux récepteurs pour capter les signaux provenant d'autres neurones. Leur longueur varie de quelques micromètres à plusieurs dizaines de micromètres (Zupan et Gasteiger, 1993).

➤ Corps cellulaire

Le corps cellulaire, ou soma, contient le noyau du neurone. Il est responsable des transformations biochimiques nécessaires à la synthèse des molécules essentielles à la vie du neurone, comme les enzymes. Sa forme est généralement sphérique ou pyramidale (Jain et al., 1996).

➤ Axone

L'axone est une extension longue et fine qui transmet les signaux générés dans le corps cellulaire vers d'autres neurones ou cellules. Il se termine par une arborisation composée de boutons synaptiques, autour desquels se trouvent les synapses. La longueur des axones peut varier, allant de quelques millimètres à plusieurs mètres. En général, l'axone est plus long que les dendrites, bien que cela puisse varier selon le type de neurone.

➤ Synapses

Les synapses sont des connexions entre deux neurones. Elles permettent la transmission des signaux nerveux, jouant un rôle fondamental dans le fonctionnement du système nerveux.

Le mécanisme des neurones biologiques a servi de modèle pour la création des neurones artificiels, qui forment les éléments de base des RNA. À l'instar des neurones biologiques, ces neurones artificiels reçoivent des entrées, les pondèrent et les transforment à l'aide de fonctions d'activation avant de produire une sortie. La section suivante examine les points communs et les distinctions entre les neurones biologiques et artificiels, ainsi que leur importance dans les RNA.

2.4.2 Neurone formel

Le neurone formel, ou neurone artificiel est l'unité fondamentale de traitement dans les RNA. Il reçoit un ensemble d'entrées pondérées et génère une sortie réelle. Cette sortie est déterminée par les valeurs d'entrée, les poids attribués à ces connexions, et une fonction d'activation qui transforme la somme pondérée des entrées en une sortie non linéaire (Figure 2.2).

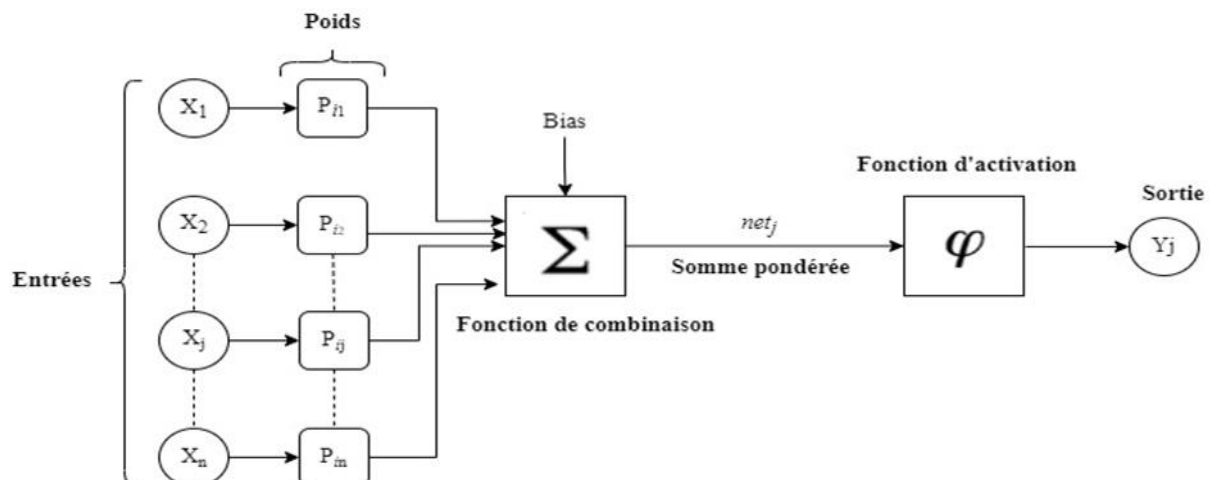


Figure 2.2: Modèle mathématique d'un neurone artificiel (Parizeau, 2004)

Chaque neurone est un processeur élémentaire qui :

- ✓ Reçoit des signaux d'un nombre variable de neurones en amont ;
- ✓ Associe chaque entrée (x_i) à un poids (p_{ij}) représentant l'importance ou la force de la connexion ;
- ✓ Produit une sortie unique (y_j), qui est transmise à d'autres neurones situés en aval (McCulloch et Pitts, 1943).

Contrairement aux neurones biologiques, dont les états sont binaires (actif ou inactif), les neurones artificiels utilisent des fonctions d'activation continues, offrant une plage de valeurs dans des intervalles tels que $[-1, +1]$ ou $[0,1]$. Ces fonctions permettent de modéliser des relations non linéaires complexes.

Le neurone formel est souvent utilisé dans des modèles à temps discret, où les variations temporelles ne sont pas prises en compte (Parizeau, 2004). Pendant le processus d'apprentissage, les poids p_{ij} et les biais b_j sont ajustés pour améliorer les performances du réseau (Bishop, 1995).

Les étapes principales pour l'utilisation des RNA sont :

- ✓ Initialisation : assigné des valeurs initiales aux poids et biais ;
- ✓ Activation : calculer les sorties du réseau pour des entrées données ;
- ✓ Apprentissage des poids : ajuster les poids pour réduire l'erreur entre les prédictions et les résultats réels.

Le tableau 2.1 présente une comparaison entre les neurones biologiques et formels, illustrant leur correspondance fonctionnelle.

Tableau 2.1: Analogie entre le neurone biologique et le neurone formel (Bishop, 1995)

Neurone biologique	Neurone formel
Synapses	Connexion synaptique (poids)
Dendrite	Signal d'entrée
Corps cellulaire	Fonction d'activation
Axones	Signal de sortie

2.4.2.1 Réseaux neurone à perceptron à couche unique

Le perceptron est l'un des premiers modèles de réseaux de neurones artificiels, proposés par Rosenblatt (1958). Il est constitué d'une seule couche de neurones, où chaque neurone :

1. Reçoit N entrées ($x_{i...n}$).
2. Produit une sortie en fonction des poids associés à ces entrées.

L'apprentissage supervisé du perceptron implique l'ajustement des poids à chaque itération pour minimiser l'erreur entre les prédictions et les valeurs observées. Cependant, ce modèle est limité aux problèmes linéairement séparables (Rosenblatt, 1962).

2.4.2.2 Réseaux neurone à perceptron multi couches

Pour surmonter les limites des perceptrons à une couche, les perceptrons multi couches (PMC) ont été introduits. Ces réseaux comportent (Haykin, 2009) :

- ✓ Une couche d'entrée : qui reçoit les données brutes ;
- ✓ Une ou plusieurs couches cachées : responsables des transformations non linéaires grâce à des fonctions d'activation ;
- ✓ Une couche de sortie : qui fournit les résultats finaux.

Le perceptron multicouche (PMC) est un réseau à propagation avant, ce qui signifie que les informations circulent dans un seul sens, de la couche d'entrée à la couche de sortie. Ce modèle est particulièrement utile pour résoudre des problèmes non linéaires complexes grâce à sa capacité à établir des limites décisionnelles plus sophistiquées, définies par (Minsky et Papert, 1969) :

- ✓ Le nombre de couches cachées ;
- ✓ Le nombre de neurones dans chaque couche.

Les relations entre les entrées, les poids et les biais sont exprimées par l'équation :

$$y_j = f \sum_{i=1}^n x_i w_{ij} + b_j \quad (2.1)$$

Où x_i représentent les entrées, y_j la sortie du neurone, w_{ij} les poids, b_j le biais, et f la fonction d'activation.

Grâce à l'algorithme de rétropropagation, les PMC permettent l'ajustement itératif des poids et biais, ce qui améliore leur performance dans les tâches complexes. Ce modèle reste l'un des plus couramment utilisés dans les applications modernes (Haykin, 2009).

2.4.3 Fonction de transfert

La fonction de transfert (ou d'activation) est un composant essentiel des neurones artificiels. Elle transforme la somme pondérée des entrées en une sortie, introduisant ainsi la non-linéarité nécessaire pour modéliser des relations complexes.

Dans leur modèle initial, McCulloch et Pitts (1943) utilisaient une fonction d'activation basée sur un seuil, où la sortie d'un neurone était binaire (Parizeau, 2004). Aujourd'hui, les fonctions de transfert sont plus flexibles et peuvent adopter différentes formes.

La sélection de la fonction d'activation repose sur les exigences spécifiques de la tâche ainsi que sur les propriétés des données. Chaque type de fonction de transfert est conçu pour répondre à des besoins spécifiques dans la modélisation des RNA. Par exemple, une fonction binaire est particulièrement efficace pour des tâches de classification et d'organisation, tandis que les fonctions continues et différentiables, comme les fonctions linéaires, sigmoïdes ou à base radiale, sont mieux adaptées pour approximer des

fonctions continues (Zupan et Gasteiger, 1993). Ces différentes fonctions, essentielles pour la transformation des données dans les RNA, sont décrites ci-dessous.

Tableau 2.2: Principales fonctions de transfert (Parizeau, 2004)

Fonctions de transfert	Équation	Graphiques
Seuil (hardlim)	$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	
Seuil symétrique (hardlims)	$f(x) = \begin{cases} -1 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	
Linéaire saturante (satlin)	$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$	
Linéaire à saturation symétrique (satlins)	$f(x) = \begin{cases} -1 & \text{for } x < -1 \\ x & \text{for } -1 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$	
Linéaire (purelin)	$f(x) = x$	
Linéaire positive (poslin)	$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	
Log sigmoïde (logsig)	$f(x) = \frac{1}{1 + e^{-x}}$	
Tangente hyperbolique sigmoïde (tansig)	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	
Base radiale	$f(x) = e^{-x^2}$	
Unité Linéaire Rectifiée (ReLU)	$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	
Unité Linéaire Exponentielle Scalable (SELU)	$f(x) = \lambda \begin{cases} x & \text{for } x > 0 \\ \alpha(e^x - 1) & \text{for } x \leq 0 \end{cases}$	

2.4.3.1 La fonction linéaire

Également appelée fonction identité, cette fonction attribue directement son entrée à sa sortie sans appliquer de transformation. Elle est idéale pour des tâches où la relation entre les variables est simple et linéaire. Cependant, en raison de son absence de non-linéarité, elle est rarement utilisée seule dans des applications complexes.

2.4.3.2 La fonction sigmoïde

La fonction sigmoïde est largement utilisée pour interpréter les sorties des neurones comme des probabilités. Elle génère des valeurs dans l'intervalle $[0,1]$, ce qui est particulièrement utile pour des tâches de classification binaire ou d'interprétation probabiliste. Elle atténue l'impact des valeurs extrêmes, ce qui réduit le risque de divergence dans les calculs (Rumelhart et al., 1986). Cependant, elle présente une limite importante : dans les cas où les valeurs d'entrée sont très grandes ou très petites, la dérivée devient proche de zéro, ce qui peut entraîner un problème de gradient évanescent dans les réseaux profonds.

2.4.3.3 La tangente hyperbolique

Contrairement à la fonction sigmoïde, la fonction tangente hyperbolique (\tanh) est centrée sur zéro, produisant des sorties dans l'intervalle $[-1, 1]$. Cette caractéristique la rend particulièrement adaptée aux modèles où les données d'entrée peuvent contenir des valeurs négatives, car elle traite ces dernières plus efficacement. De grandes entrées négatives tendent vers -1 , tandis que de grandes entrées positives tendent vers $+1$ (Demuth et al., 2013).

2.4.3.4 La fonction seuil

La fonction seuil est une fonction simple qui applique un seuil binaire à l'entrée. Elle renvoie une sortie de 0 si l'entrée est négative, et 1 si l'entrée est positive ou nulle (Demuth et al., 2004). Bien qu'efficace pour des tâches de classification élémentaire, sa nature discrète limite son utilisation dans des applications nécessitant des ajustements fins ou des transitions continues entre les sorties.

2.5 Architecture des réseaux de neurones artificiels

Les réseaux de neurones artificiels sont des systèmes formels, hautement connectés et fonctionnant en parallèle. Chaque neurone reçoit des informations d'autres neurones, les traite, et produit une sortie unique. L'architecture d'un réseau est définie par la manière dont ses neurones sont connectés (Bishop, 1995). Il existe principalement trois types d'architectures de réseaux de neurones :

2.5.1 Réseaux déverrouillés

Aussi appelés réseaux unidirectionnels ou réseaux à propagation avant, ces réseaux sont organisés en une ou plusieurs couches successives. Les données circulent exclusivement de la couche d'entrée vers la couche de sortie, traversant les couches intermédiaires sans possibilité de rétroaction. Le perceptron est un exemple de réseau adoptant cette architecture (Jain et al., 1996).

2.5.2 Réseaux verrouillés

Également appelés réseaux récurrents, ils ressemblent aux réseaux unidirectionnels mais incluent des connexions entre les neurones d'une même couche ou vers les couches précédentes. Cette structure confère aux réseaux verrouillés un comportement séquentiel et dynamique, leur permettant d'analyser des données temporelles ou séquentielles avec une grande efficacité (Dreyfus, 2011; Mandic et Chambers, 2001).

2.5.3 Réseaux avec des connexions complexes

Ces réseaux, désignés comme réseaux de Hopfield, sont des réseaux récurrents entièrement connectés où chaque neurone est connecté à tous les autres. Il n'existe pas de distinction entre les neurones d'entrée et de sortie. Cette catégorie englobe les réseaux qui ne s'inscrivent pas dans les deux architectures précédentes (Hopfield, 1984).

2.6 Apprentissage des réseaux de neurones artificiels

La phase d'apprentissage est cruciale dans la conception des réseaux de neurones artificiels. Elle permet au réseau d'apprendre et d'améliorer ses performances en ajustant les poids et biais des connexions pour converger vers des valeurs optimales. Grâce à cette phase, les RNA deviennent des outils puissants, capables de modéliser des systèmes complexes avec précision (Touzet, 1992). Deux éléments clés déterminent le processus d'apprentissage :

- ✓ Le paradigme d'apprentissage, qui dépend des données disponibles ;
- ✓ La règle d'apprentissage, qui définit comment les poids des connexions sont ajustés (Touzet, 1992).

2.6.1 Apprentissage supervisé

Dans ce paradigme, les paramètres du réseau sont ajustés pour reproduire des réponses souhaitées. Les exemples fournis sont des couples (entrée, sortie souhaitée). Parmi les algorithmes les plus utilisés pour cet apprentissage, l'algorithme de rétropropagation du gradient (Basheer et Hajmeer, 2000; Haykin, 1999).

2.6.2 Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, l'adaptation des poids dans ce cas repose uniquement sur les données d'entrée. Le réseau n'a ni signal d'erreur ni sortie souhaitée, ce qui le rend idéal pour le regroupement et la classification (Dreyfus, 1998).

2.6.3 Apprentissage par renforcement

Aussi appelé apprentissage semi-supervisé, ce paradigme permet au réseau d'interagir avec l'environnement. L'environnement fournit des récompenses pour les réponses correctes et des pénalités pour les réponses erronées. Cela pousse le réseau à maximiser ses récompenses et à affiner ses décisions (White, 1992).

2.7 Règles d'apprentissage

L'algorithme d'apprentissage est une méthode mathématique qui ajuste les poids des connexions pour permettre au réseau d'accomplir sa tâche de manière optimale. Dans le contexte des réseaux de neurones, la connaissance est distribuée et encodée dans plusieurs éléments : les poids des connexions, la structure du réseau, les fonctions de transfert de chaque neurone, et les seuils associés à ces fonctions. La méthode d'apprentissage utilisée joue également un rôle crucial dans ce processus (Deryfus et al., 2008).

2.7.1 Règle de Hebb

Inspirée de la biologie, cette règle renforce les connexions entre neurones simultanément actifs, laissant inchangées celles où l'un des deux neurones est inactif, les poids restent inchangés (Hebb, 1949).

2.7.2 Règle de Widrow-Hoff

Également appelée règle delta, elle ajuste les poids en fonction de la différence entre la sortie réelle et la sortie désirée. Elle repose sur la méthode des moindres carrés moyens pour minimiser l'erreur (Widrow et Hoff, 1960).

2.7.3 Règle de rétropropagation

Cette règle, une généralisation de la règle delta, ajuste les poids pour les réseaux multicouches et les fonctions d'activation non linéaires. Elle repose sur la propagation de l'erreur à travers les couches du réseau (Rumelhart et al., 1986).

2.7.4 Règle de compétition

Ce processus s'applique à un seul neurone : les données sont d'abord classées en catégories, puis les neurones similaires sont regroupés en fonction de leurs corrélations et représentés par un neurone unique. Ce type de réseau comprend une couche d'entrée et une couche compétitive. Lorsqu'une forme est introduite en entrée du réseau, elle est projetée sur chaque neurone de la couche compétitive. Le neurone vainqueur est celui dont le vecteur de poids est le plus proche de la forme initiale. Chaque neurone de sortie est connecté aux neurones de la couche d'entrée, à d'autres neurones de sortie, ou à lui-même. La sortie résulte alors de la compétition entre les connexions inhibitrices et excitatrices (Haykin, 1999 ; Parizeau, 2004).

2.7.5 Règle de correction d'erreur

Adaptée à l'apprentissage supervisé, cette règle ajuste les poids en fonction d'une erreur calculée entre la sortie désirée et la sortie réelle, modifiant les connexions pour diminuer l'erreur globale jusqu'à ce que la sortie obtenue corresponde à la sortie souhaitée. Les poids sont modifiés jusqu'à ce que cette erreur soit minimisée (Deryfus et al., 2008).

2.8 Paramètres d'apprentissage

Différents paramètres influencent l'efficacité du processus d'apprentissage, notamment la vitesse de convergence et la performance de généralisation. Les principaux paramètres sont :

➤ Taux d'apprentissage η

Le taux d'apprentissage contrôle la vitesse de convergence. Une valeur initiale trop élevée accélère le processus, peut entraîner des oscillations empêchant l'algorithme d'atteindre un minimum global, tandis qu'un taux trop faible ralentit excessivement l'apprentissage. Il est recommandé de débiter avec un faible η pour atteindre un minimum global, même si cela allonge le processus d'apprentissage.

➤ Momentum Ω

Le momentum permet d'incorporer une fraction de l'ancien poids dans le calcul du nouveau, ce qui aide à éviter de rester bloqué dans un minimum local et augmente la vitesse de convergence. Une valeur de Ω est généralement inférieure à 1.

➤ Seuil de tolérance

Ce paramètre détermine la précision souhaitée pour la réponse du réseau. L'apprentissage est arrêté lorsque l'erreur globale du réseau passe sous un seuil prédéfini ou après un nombre maximal d'itérations.

2.9 Critères d'arrêt de l'apprentissage

Les critères d'arrêt indiquent le moment où il est pertinent de stopper l'apprentissage, soit parce que le réseau a atteint une performance satisfaisante, soit parce que poursuivre l'apprentissage devient inefficace, pour éviter une perte de temps ou des résultats contre-productifs.

➤ Nombre maximum d'itérations

L'apprentissage est arrêté lorsqu'un nombre prédéfini d'itérations est atteint. Cependant, utiliser ce critère seul est déconseillé, car le réseau peut atteindre une performance optimale bien avant ce seuil.

➤ Seuil de gradient

La norme du gradient est utilisée comme critère pour stopper l'apprentissage lorsque celle-ci devient inférieure à un seuil donné, indiquant que l'erreur ne diminue plus de manière significative.

➤ Seuil de variation de la valeur des poids

Ce critère repose sur la mesure des variations des poids entre deux itérations consécutives. Si ces variations sont inférieures à une valeur prédéfinie, l'apprentissage est arrêté.

2.10 Phénomènes d'apprentissage des réseaux de neurones artificiels

Deux phénomènes importants influencent la qualité de l'apprentissage :

1. Sous-apprentissage

Le sous-apprentissage survient lorsque le réseau n'a pas suffisamment appris, souvent en raison d'un nombre insuffisant d'itérations ou d'une architecture trop simple. Cela se traduit par des erreurs élevées aussi bien sur les données d'apprentissage que sur celles de test (Haykin, 1999).

2. Sur-apprentissage

Le sur-apprentissage se manifeste par une capacité excessive du réseau à mémoriser les données d'apprentissage, entraînant une faible généralisation sur les données de test. Cela se produit généralement en raison d'un nombre excessif de neurones dans les couches cachées ou d'itérations trop nombreuses (Baughman et Liu, 1995).

La distinction entre ces phénomènes est illustrée dans la figure 2.3, qui met également en évidence l'apprentissage optimal.

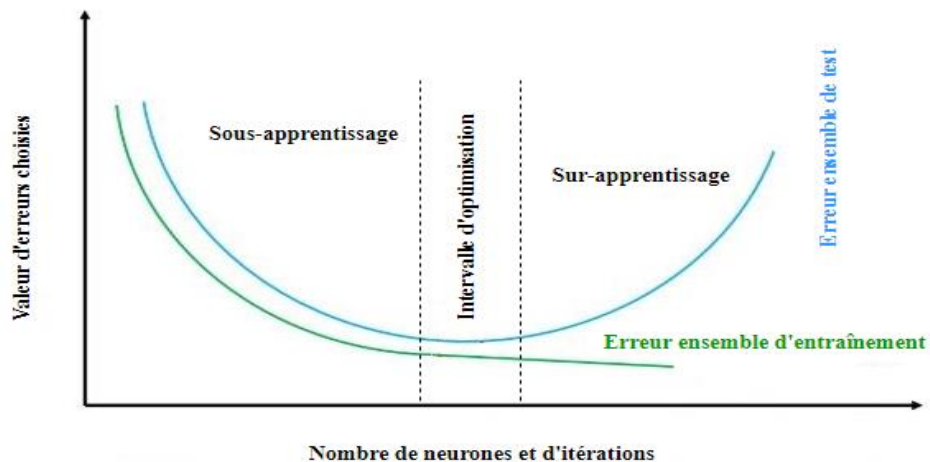


Figure 2.3: Phénomène du sous-apprentissage et du sur-apprentissage (Haykin, 1999)

2.11 Propriétés des réseaux de neurones artificiels

Les réseaux de neurones artificiels se caractérisent par les propriétés principales suivantes :

1. Capacité d'apprentissage

Les RNA sont capables d'apprendre à partir d'exemples et de résoudre des problèmes de manière similaire aux humains ou aux animaux.

2. Capacité de généralisation

Ils peuvent extraire, à partir d'un ensemble d'exemples, des règles permettant de résoudre un problème donné.

3. Parallélisme

Cette propriété est fondamentale dans l'architecture des RNA, qui sont constitués d'un ensemble d'entités élémentaires fonctionnant simultanément. Cela permet une vitesse de calcul accrue, mais nécessite une approche conceptuelle différente.

2.12 Classification des réseaux de neurones artificiels

Les nombreux modèles connexionnistes ont été classés par les chercheurs en fonction de leurs applications. En effet, divers modèles de RNA ont été développés selon les caractéristiques adoptées, telles que :

- ✓ Les cellules élémentaires : selon leur nature, elles peuvent être binaires ou continues ;
- ✓ Architecture et le nombre d'automates : cela inclut la structure globale et la taille du réseau;
- ✓ La nature des connexions et des méthodes d'apprentissage : les modèles se distinguent par la façon dont les neurones interagissent et s'adaptent ;
- ✓ La performance des mécanismes de mémorisation : cette caractéristique évalue la capacité du réseau à classer ou mémoriser des informations.

2.13 Application des réseaux de neurones artificiels

Les réseaux de neurones artificiels ont récemment connu un développement significatif dans de nombreux domaines. Cette expansion s'explique par leurs capacités de calcul élevées, leurs performances remarquables en apprentissage et leur aptitude à estimer des paramètres, indépendamment de la complexité des problèmes accueillis. L'étude des RNA constitue une voie prometteuse pour le développement de l'IA, car elle se situe à l'intersection de différents champs disciplinaires. On peut résumer leurs principaux domaines d'application comme suit (Personnaz et Rivals, 2003) :

- ✓ Environnement : évaluation des risques, analyse chimique, prévisions météorologique, gestion des ressources ;
- ✓ Défense : guidage d'armes, suivi de cibles, systèmes radars ;
- ✓ Industrie : contrôle qualité, gestion des processus, diagnostic des pannes ;
- ✓ Finance : prévisions des marchés, modélisations d'indicateurs économiques, sélection des investissements, prévision des prix ;
- ✓ Télécommunications et informatique : analyse des signaux, réduction du bruit, compression des données ;
- ✓ Médical : analyse des signaux biologiques, diagnostic de maladies comme le cancer.

Ces applications variées illustrent la polyvalence et la puissance des RNA dans des domaines divers. Cependant, leurs avantages intrinsèques nécessitent également une analyse approfondie.

2.14 Avantages des réseaux de neurones artificiels

Les réseaux de neurones artificiels présentent de nombreux avantages, parmi lesquels on peut citer (Bishop, 1995) :

- ✓ Adaptabilité aux prédictions diverses : les RNA peuvent prédire une grande diversité de variables;

- ✓ Simplicité d'utilisation et rapidité : ils sont faciles à manipuler et permettent une consultation rapide des résultats ;
- ✓ Robustesse face au manque de données : ils présentent moins de comportements indésirables lorsqu'ils disposent de peu de données ;
- ✓ Simplicité conceptuelle de l'apprentissage : l'idée d'apprentissage est plus intuitive que les complexités des méthodes statistiques multivariées ;
- ✓ Capacité à modéliser toute dépendance fonctionnelle : ils peuvent représenter une large variété de relations fonctionnelles entre les variables ;
- ✓ Résistance au bruit : les RNA restent performants même lorsque les données sont bruitées ou peu fiables.

2.15 Limites opérationnelles des réseaux de neurones artificiels

Les réseaux de neurones artificiels sont des outils puissants pour la modélisation et l'apprentissage, mais ils présentent plusieurs limitations opérationnelles qui devraient être abordées pour offrir une vue complète de leur utilisation et de leurs contraintes. Une nouvelle section pourrait inclure :

➤ **Problèmes liés à l'apprentissage**

- Sur-apprentissage (overfitting) : Lorsque le RNA s'adapte trop aux données d'entraînement, ce qui nuit à sa capacité de généralisation sur des données non vues.
- Sous-apprentissage (underfitting) : Lorsque le modèle est trop simple ou mal entraîné, il n'arrive pas à capturer les relations sous-jacentes.

➤ **Exigences en données**

Les RNA nécessitent de grandes quantités de données pour obtenir de bons résultats, ce qui peut être un obstacle dans les cas où les données disponibles sont limitées ou difficiles à collecter.

➤ **Problèmes de convergence**

Difficulté à ajuster des hyperparamètres comme le taux d'apprentissage ou le momentum, pouvant entraîner une convergence lente ou des minima locaux.

➤ **Interprétabilité limitée**

Les RNA fonctionnent souvent comme une "boîte noire", rendant difficile l'explication des décisions prises par le modèle.

➤ **Coût computationnel**

Les réseaux profonds, en particulier, nécessitent une puissance de calcul importante et des ressources matérielles coûteuses pour leur entraînement et leur déploiement.

2.16 Comparaison des réseaux de neurones artificiels avec des modèles alternatifs

Pour contextualiser les performances des RNA et proposer des perspectives comparatives, une nouvelle section pourrait comparer les RNA avec d'autres modèles comme :

➤ Les réseaux bayésiens

Gestion explicite de l'incertitude grâce à une approche probabiliste.

Capacité à intégrer des connaissances préalables (a priori).

Moins adaptés aux problèmes avec de grandes quantités de données ou des relations complexes.

➤ Machines à vecteurs de support (MVS)

Bonne performance sur des ensembles de données de petite taille avec une séparation linéaire ou non linéaire.

Moins sujets au sur-apprentissage grâce à leur régularisation intégrée.

Moins performant pour les données volumineuses par rapport aux RNA.

➤ Réseaux de régression ou modèles statistiques

Utilisés pour des problèmes linéaires simples ou des données structurées.

Performances limitées pour des systèmes hautement non linéaires.

2.17 Conclusion

Les réseaux de neurones artificiels ont considérablement évolué depuis leur introduction, devenant des outils robustes pour traiter des systèmes complexes. Leur structure flexible, associée à des approches d'apprentissage diversifiées, leur confère la capacité de modéliser des relations non linéaires et de résoudre des problèmes complexes. Cependant, leurs limites opérationnelles, telles que la nécessité de volumes importants de données, le risque de sur-apprentissage et le manque de transparence dans les décisions, constituent des défis majeurs à surmonter.

La comparaison avec d'autres modèles, tels que les réseaux bayésiens, souligne l'importance de choisir le bon outil en fonction des besoins spécifiques de l'application. Les RNA excellent dans les tâches où les données massives et la complexité sont dominantes, tandis que des approches probabilistes offrent une meilleure interprétation et gestion de l'incertitude dans certains cas.

Enfin, pour renforcer les performances des RNA, l'intégration de techniques avancées, comme les techniques de décomposition, est essentielle. Ces méthodes permettent un prétraitement efficace des données, améliorant leur qualité et facilitant leur traitement dans des environnements complexes. Cette approche marque une transition clé vers des systèmes hybrides, combinant les forces des RNA avec des outils d'analyse avancés, offrant ainsi des solutions plus précises, robustes et adaptatives pour répondre aux défis des systèmes modernes.

Chapitre 3 : Techniques de Décomposition : Ondelettes Discrètes, Modes Empiriques et Variationnels

3.1 Introduction

Les signaux non stationnaires, souvent rencontrés dans divers domaines d'application, posent des défis significatifs en termes d'analyse et de traitement. Ces signaux, caractérisés par des variations complexes dans le temps et en fréquence, nécessitent des techniques de décomposition adaptées pour en extraire efficacement les informations essentielles. Parmi les méthodes contemporaines, la transformation en ondelettes, la décomposition en modes empiriques, et la décomposition en modes variationnels se démarquent par leurs capacités à offrir des résolutions multi-échelles et multi-résolutions. Ces techniques, développées à partir d'avancées mathématiques et computationnelles, ont permis de dépasser les limites des approches traditionnelles, telles que la transformée de Fourier. Elles s'appuient sur des principes distincts.

Dans ce chapitre, nous examinerons en détail ces trois techniques, en mettant en évidence leurs principes fondamentaux, leurs propriétés et leurs avantages. Nous proposerons également une comparaison entre ces techniques.

3.2 Transformation en ondelettes

Les transformations en ondelettes (TO) ont émergé dans les années 80, grâce aux travaux de Jean Morlet, qui les a initialement développées pour des applications dans l'analyse sismique (Meyer et al., 1987). Par la suite, les ondelettes ont connu des avancées majeures grâce à des contributions significatives dans le domaine des mathématiques, notamment celles de Yves Meyer (1990, 1993), Ingrid Daubechies (1992), et Stéphane Mallat (2009). Ces chercheurs ont élargi les applications des ondelettes au-delà du traitement sismique, en établissant un cadre rigoureux pour l'analyse multi-résolution et les décompositions de signaux complexes.

3.2.1 Historique de la transformation en ondelettes

L'analyse des données par ondelettes est aujourd'hui une technique incontournable dans de nombreux domaines, aussi bien théoriques que pratiques. Elle est particulièrement prisée dans le traitement du signal en raison de sa capacité à s'adapter aux données et de sa simplicité de mise en œuvre. Cependant, pour comprendre son évolution, il est nécessaire de revenir sur son développement historique.

Au 19^e siècle, l'analyse de Fourier était la seule méthode disponible pour décomposer un signal et le reconstruire sans perte d'information. Bien qu'efficace pour fournir une analyse en fréquence, cette méthode présentait une limitation majeure : elle ne permettait pas de localiser temporellement les variations brusques du signal. Cette lacune a conduit à l'exploration d'approches alternatives.

En 1909, Alfréd Haar a introduit la première fonction d'ondelette, caractérisée par une courte impulsion négative suivie d'une courte impulsion positive. Cette contribution, bien que rudimentaire, a jeté les bases du développement des ondelettes modernes.

Ensuite, en 1946, Dennis Gabor a proposé une transformation inspirée des travaux de Joseph Fourier. Cette transformation, qui repose sur une fenêtre temporelle définie par une fonction gaussienne, a marqué une avancée en introduisant la notion de localisation temporelle dans l'analyse en fréquence.

Le concept d'ondelette, tel que nous le connaissons aujourd'hui, a véritablement émergé en 1984, grâce à Jean Morlet et Alex Grossmann, qui ont introduit le terme ondelette dans le langage mathématique. Cette étape a marqué une révolution dans le domaine en proposant une méthode capable de représenter un signal à la fois dans les domaines temporel et fréquentiel.

À partir de ces fondations, plusieurs chercheurs ont perfectionné la théorie des ondelettes :

- En 1986, Yves Meyer a consolidé les travaux antérieurs en définissant les ondelettes orthogonales, qui garantissent une reconstruction exacte du signal sans redondance.
- La même année, Stéphane Mallat a établi un lien entre les ondelettes et l'analyse multi-résolution, offrant ainsi une méthode structurée pour décomposer un signal en différentes échelles de détail.
- En 1987, Ingrid Daubechies a introduit les célèbres ondelettes de Daubechies, qui sont orthogonales et particulièrement adaptées à l'analyse des signaux complexes en raison de leurs propriétés mathématiques remarquables.

Grâce à ces contributions successives, les ondelettes sont devenues une méthode puissante et polyvalente pour l'analyse des signaux. Contrairement à l'analyse de Fourier, les ondelettes permettent une localisation précise dans le temps et en fréquence, ce qui les rend particulièrement adaptées à des applications comme le traitement d'images, la compression de données, et la détection de singularités.

Deux principales méthodes de transformations en ondelettes ont été développées :

3.2.2 Transformation en Ondelettes Continue

La transformation en ondelettes continue (TOC) est une méthode linéaire qui décompose un signal en composantes élémentaires appelées ondelettes, obtenues par translation et dilatation d'une ondelette mère. Cette technique offre une résolution élevée en fréquence pour les basses fréquences et une résolution temporelle fine pour les hautes fréquences.

L'équation de la transformation en ondelettes continue s'exprime comme suit :

$$S_{i,j} = \int_{-\infty}^{+\infty} \psi_{i,j}^*(t) X(t) dt \quad (3.1)$$

Où : $S_{i,j}$: Coefficient d'ondelette calculé à l'échelle j et à la position i .

$\psi_{i,j}(t)$: Ondelette fille obtenue par translation et dilatation de l'ondelette mère $\psi(t)$

$\psi_{i,j}^*(t)$: Conjugué complexe de $\psi_{i,j}(t)$

L'ondelette fille est définie comme suit :

$$\psi_{i,j}(t) = \frac{1}{\sqrt{j}} \psi\left(\frac{t-i}{j}\right) \quad (3.2)$$

Où : i : Paramètre de translation, indiquant la position temporelle de l'ondelette.

j : Paramètre d'échelle, contrôlant la dilatation ou la compression de l'ondelette.

L'un des principaux avantages de la TOC réside dans sa capacité à produire :

- ✓ Une résolution fréquentielle élevée aux basses fréquences, adaptée aux signaux de longue durée ;
- ✓ Une résolution temporelle élevée aux hautes fréquences, idéale pour les signaux de courte durée.

Cette caractéristique rend la transformation en ondelettes continues particulièrement efficace pour analyser les signaux non stationnaires et localiser précisément les événements dans le temps. De plus, la transformation en ondelettes continues peut être utilisée pour réduire le bruit dans les signaux, améliorant ainsi leur qualité sans perte significative d'information.

Dans la transformation en ondelettes continues, les paramètres de dilatation (échelle) et de translation (position) sont continus, ce qui entraîne une infinité de coefficients. Toutefois, le nombre de coefficients peut être restreint tout en conservant les informations essentielles du signal original.

3.2.3 Transformation en Ondelettes Discrètes

La transformation en ondelettes discrète (TOD) repose sur l'utilisation d'ondelettes orthogonales, permettant une localisation précise en temps et en fréquence. Contrairement au TOC, les paramètres d'échelle et de translation sont discrétisés, ce qui réduit la redondance des coefficients.

La transformation en ondelettes discrète offre une analyse multi-résolution efficace et est largement utilisée dans des applications comme la compression d'images ou la détection de singularités.

➤ Paramètres d'échelle et de position

La transformation en ondelettes discrète repose sur la discrétisation des paramètres d'échelle (a) et de translation (b) définis comme suit :

$$\begin{cases} a = a_0^j, a_0 > 1 \\ b = ib_0 a_0^j, b_0 > 0 \end{cases} \quad i, j \in Z \quad (3.3)$$

Les ondelettes filles sont obtenues par dilatation et translation de l'ondelette mère $\psi(t)$ selon la relation:

$$\psi_{i,j}(t) = a_0^{-j/2} \psi\left(\frac{t}{a_0^j} - ib_0\right) \quad (3.4)$$

Où : a_0 et b_0 sont des constantes. Selon Yves Meyer, les valeurs courantes utilisées sont $a_0 = 2$ et $b_0 = 1$, ce qui donne :

$$\psi_{i,j}(t) = 2^{-j/2} \psi\left(\frac{t}{2^j} - i\right) \quad (3.5)$$

Cette discrétisation est appelée analyse dyadique et constitue la base de nombreuses applications en traitement du signal.

➤ **Algorithme de Mallat**

La méthode la plus couramment utilisée pour appliquer cette transformation est basée sur des bancs de filtres pyramidaux (Mallat, 1992). Cet algorithme de filtrage décompose un signal en :

1. Approximations (A_j)

- ✓ Représentent les composantes de basse fréquence du signal ;
- ✓ Fournissent une vue d'ensemble ou une forme générale du signal.

2. Détails (D_j)

- ✓ Contiennent les composantes de haute fréquence ;
- ✓ Révèlent les variations fines et les singularités du signal.

Les approximations et détails détaillent la méthodologie de décomposition d'un signal à travers des filtres passe-bas et passe-haut, montrant comment les signaux peuvent être analysés et reconstruits. Cette décomposition s'effectue en utilisant :

- Filtre passe-bas ($h_{0,j}$) : Projette les approximations d'un niveau $j-1$ vers l'espace d'approximation du niveau j , en conservant les basses fréquences.
- Filtre passe-haut ($h_{1,j}$) : Capture les détails du signal en projetant les approximations sur l'espace des détails, correspondant aux hautes fréquences.
- Ces composants permettent de reconstruire le signal d'origine en combinant les approximations et les détails à différents niveaux de résolution.

Les approximations (A_j) et les détails (D_j) sont calculés comme suit :

$$A_j[K] = \sum_{i=1}^n X[i] \times h_{0,j}[2K - i] \tag{3.6}$$

$$D_j[K] = \sum_{i=1}^n X[i] \times h_{1,j}[2K - i] \tag{3.7}$$

Ce processus est itératif, permettant de décomposer les approximations (A_j) en niveaux de résolution de plus en plus fins.

➤ **Reconstruction d'un signal**

Le processus de reconstruction consiste à reconstituer le signal original à partir des coefficients d'ondelettes. Cette étape inverse la phase de décomposition en utilisant les filtres de reconstruction, qui sont les inverses des filtres utilisés pour la décomposition :

$$h_{0,j}R = \text{inverse}(h_{0,j}) \tag{3.8}$$

$$h_{1,j}R = \text{inverse}(h_{1,j}) \tag{3.9}$$

La relation entre l'approximation et les détails permet de reconstituer le signal :

$$A^{j-1} = A^j + D^j \quad (3.10)$$

Où : $h_{0,j}$: Filtre passe-bas qui projette l'approximation de l'échelle $j-1$ sur l'espace d'approximation de l'échelle j . Cela nous donne l'approximation à l'échelle j .

$h_{1,j}$: Filtre passe-haut qui projette l'approximation de l'échelle $j - 1$ sur l'espace d'ondelettes (détail) de l'échelle j . Cela nous donne les coefficients d'ondelettes de l'échelle j .

3.2.3.1 Principe de la transformation en ondelettes

Le principe de la transformation en ondelettes repose sur la décomposition d'un signal en une famille d'ondelettes, générées par dilatation (échelle) et translation (position) d'une ondelette mère (Daubechies, 1992). Cette approche met en évidence l'adaptabilité de la méthode, permettant d'analyser un signal à différentes échelles et positions.

- Une ondelette courte, caractérisée par fréquence centrale élevée, offre une meilleure résolution temporelle mais une résolution fréquentielle plus faible ;
- À l'inverse, une ondelette longue avec une fréquence centrale basse fournit une meilleure résolution fréquentielle, mais au détriment de la résolution temporelle.

Cette flexibilité rend la transformation en ondelettes particulièrement adaptée à l'analyse de signaux non stationnaires, dont les caractéristiques varient dans le temps.

La sélection de l'ondelette d'analyse explique l'importance du choix de l'ondelette mère pour garantir une décomposition précise et efficace. Le choix de l'ondelette mère est une étape essentielle dans l'application de la TO, car il détermine la qualité de la décomposition. Cependant, il n'existe pas de critère universel pour sélectionner l'ondelette optimale.

Les critères de sélection dépendent des caractéristiques du signal et des objectifs de l'analyse, et les chercheurs s'appuient souvent sur une évaluation visuelle des résultats pour ajuster leur choix. Une ondelette bien adaptée garantit une décomposition précise et cohérente du signal à différentes échelles.

3.2.3.2 Propriétés des familles d'ondelettes

Avant de procéder à une décomposition par ondelettes, il est essentiel de choisir une ondelette mère appropriée. Ce choix doit prendre en compte de la durée et de la largeur de bande de l'ondelette, car ces paramètres impactent les résolutions temporelle et fréquentielle de la transformation.

Les familles d'ondelettes sont variées et peuvent être classées en deux grandes catégories (Bryson et Ho, 1969 ; Wallis, 1976) :

1. Ondelettes associées à des filtres

- ✓ Utilisées principalement avec la TOD ;
- ✓ Incluent les ondelettes de Daubechies, Symlet, Coiflet, et Haar ;

- ✓ Ces ondelettes sont adaptées aux analyses multi-résolutions et permettent des calculs rapides grâce à l'algorithme de Mallat.

2. Ondelettes sans filtres

- ✓ Utilisées principalement avec la TOC ;
- ✓ Comprennent des ondelettes comme l'ondelette de Morlet, le chapeau mexicain, les ondelettes gaussiennes et leurs variantes complexes ;
- ✓ Ces ondelettes, dotées d'une régularité infinie et d'une symétrie parfaite, offrent des reconstructions explicites et des translations constantes.

Les principales caractéristiques des familles d'ondelettes se distinguent par :

- ✓ La présence ou non de filtres associés ;
- ✓ Leur biorthogonalité ou orthogonalité ;
- ✓ Leur support (compact ou non compact) ;
- ✓ Leur nature (réelle ou complexe).

➤ **Support compact**

Une ondelette est dite à support compact dans le domaine temporel si sa durée est limitée et si elle présente une atténuation rapide. Cette propriété réduit la complexité de calcul, améliore la résolution temporelle, mais peut entraîner une résolution fréquentielle médiocre.

Les ondelettes de Daubechies, Symlet et Coiflet ont un support compact dans le domaine temporel, tandis que les ondelettes de Meyer ont un support compact dans le domaine fréquentiel.

➤ **Régularité**

La régularité d'une ondelette correspond à sa capacité à détecter les singularités dans un signal. Une ondelette est d'autant plus régulière qu'elle possède un grand nombre de moments nuls, ce qui lui permet de ne pas détecter les polynômes de degré inférieur à son ordre. La corrélation entre régularité et moments nuls est particulièrement utile dans l'analyse des signaux irréguliers.

➤ **Symétrie**

La symétrie d'une ondelette garantit une réponse équilibrée autour d'un centre. Cette propriété dépend de la régularité de l'ondelette et du nombre de moments nuls qu'elle possède. Les ondelettes symétriques jouent un rôle crucial pour éviter les déphasages lors de la décomposition et de la reconstruction.

➤ **Orthogonalité**

La propriété d'orthogonalité d'une ondelette garantit qu'aucune information n'est redondante lors de la décomposition d'un signal. Cela permet une reconstruction parfaite du signal original, sans perte ni duplication d'information.

Le tableau 3.1 démontre les propriétés principales différentes des familles d'ondelettes, en distinguant celles avec et sans filtres. Ces caractéristiques influencent leur choix pour différentes applications :

Tableau 3.1: Principales propriétés des familles d'ondelettes (Misiti et al., 2003)

Type d'ondelettes	Support compact	Nature	Exemples
Ondelettes avec filtres	Oui	Réelles	Daubechies (db), Haar, Symlet (sym), Coiflet (coif)
	Oui	Réelles	Biorthogonales (Bior)
Ondelettes sans filtres	Non	Réelles	Morlet, Gaussienne (gaus), Chapeau mexicain
	Non	Complexes	Cgaus, Shannon, Cmor

3.2.3.3 Différentes familles d'ondelettes

Les familles d'ondelettes regroupent une variété d'ondelettes mères utilisées pour appliquer la TO à des signaux analysés. Chaque famille est adaptée à des types de signaux et à des domaines d'application spécifiques. Le choix de l'ondelette mère a un impact direct sur la précision et la qualité de la décomposition, ce qui en fait une décision clé dans le processus d'analyse. Les principales familles d'ondelettes sont illustrées dans la figure 3.1.

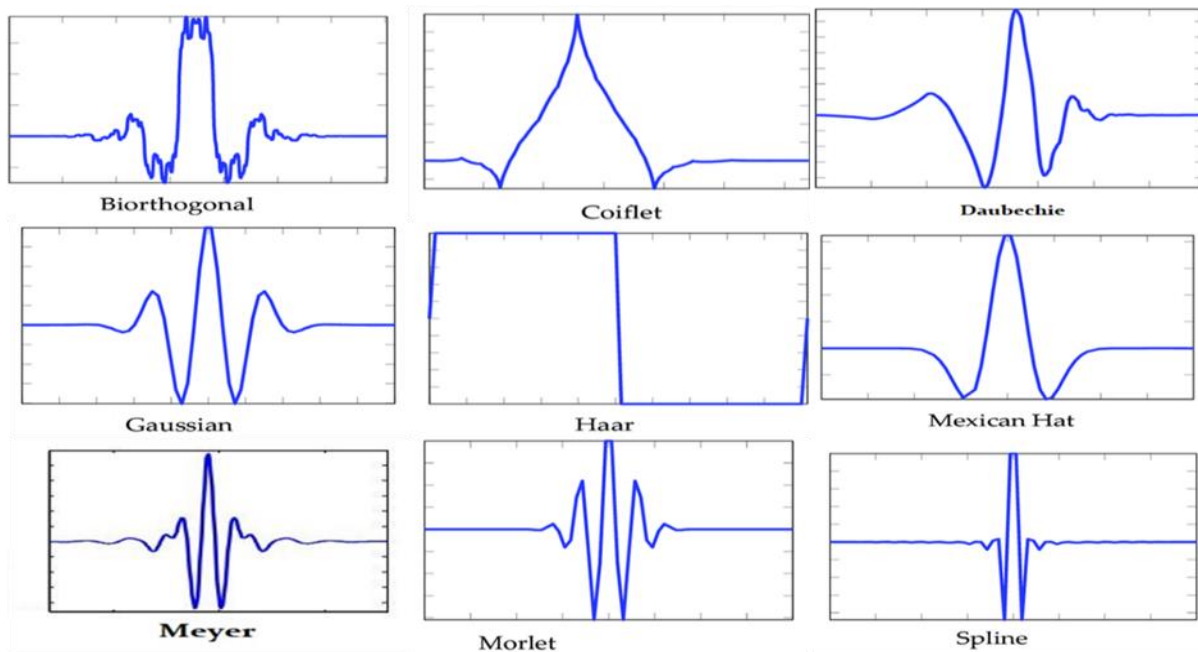


Figure 3.1: Présentation de différentes ondelettes mères

1. Ondelettes de Coiflets

Les ondelettes de Coiflets se distinguent par une propriété particulière : leur fonction d'échelle associée ϕ possède $2N-1$ moments nuls. Cela les rend particulièrement adaptées pour des analyses nécessitant une localisation précise des singularités dans les signaux.

2. Ondelettes de Daubechies

Les ondelettes de Daubechies (dbN) sont orthogonales, avec un support compact et une régularité variable.

- ✓ Leur ordre (N) détermine leur niveau de régularité ;
- ✓ L'ondelette $db1$, connue comme l'ondelette de Haar, est la forme la plus simple mais elle manque de continuité ;
- ✓ Les ondelettes dbN (avec $N > 1$) offrent une meilleure régularité et permettent des analyses orthogonales efficaces.

Les ondelettes de cette famille sont définies par:

$$h(t) = \begin{cases} 1 & \text{pour } 0 \leq t \leq \frac{1}{2} \\ -1 & \text{pour } \frac{1}{2} \leq t \leq 1 \end{cases} \quad (3.11)$$

3. Ondelettes de Meyer

L'ondelette de Meyer, l'une des premières ondelettes orthogonales, a été développée au milieu des années 80 par Yves Meyer, elle est caractérisée par un support compact en fréquence, mais pas en temps. Cela la rend utile pour des analyses nécessitant une représentation précise dans le domaine fréquentiel.

L'ondelette de Mexican, dérivé de la seconde dérivée de la fonction gaussienne, est définie comme suit

4. Ondelettes de Mexican

L'ondelette Mexican, dérivée de la seconde dérivée de la fonction gaussienne, est définie comme suit :

$$\psi(x) = \left(\frac{2}{\sqrt{3\pi}}\right)^{-1/4} (1 - x^2)e^{-x^2/2} \quad (3.12)$$

Cette ondelette est idéale pour l'analyse des singularités et des transitions rapides dans les signaux.

5. Ondelettes de Symlets

Les Symlets ($symN$) sont des variantes presque symétriques des ondelettes de Daubechies. Bien qu'elles partagent des propriétés similaires, leur construction modifiée garantit une meilleure symétrie, ce qui est particulièrement utile dans les applications nécessitant une faible distorsion des signaux.

6. Ondelettes de Morlet

L'ondelette de Morlet est définie par :

$$\psi(x) = C_e^{-x^2/2} \cos(5x) \quad (3.13)$$

Où C : est une constante de normalisation.

Elle est souvent utilisée dans les analyses temps-fréquence en raison de sa localisation simultanée dans les deux domaines.

3.2.3.4 Avantages de la transformation en ondelettes discrètes

La transformation en ondelettes discrètes est particulièrement utile pour :

- Analyse multi-résolution : Elle permet d'identifier les structures globales et les détails locaux dans les signaux.
- Compression des données : Elle réduit la taille des fichiers tout en conservant les informations essentielles.
- Réduction du bruit : Isolation des composantes bruitées dans les détails, permettant une reconstruction propre.

Elle offre une localisation précise en temps et en fréquence tout en étant plus efficace en termes de calcul que la TOC. Grâce à sa nature discrète, elle est idéale pour les implémentations numériques et les applications pratiques en traitement du signal.

3.3 Décomposition en Modes Empiriques

La décomposition en modes empiriques (DME), développée par Huang et al. (1998), représente une avancée majeure dans l'analyse des signaux non linéaires et non stationnaires. Cette technique innovante, initialement introduite pour résoudre les limites des méthodes classiques comme la transformée de Fourier, s'est rapidement imposée comme un outil puissant dans divers domaines du traitement du signal (Huang et al., 1998 ; Li et al., 2009).

Elle décompose un signal en une somme de fonctions de mode intrinsèques (FMI), obtenues par un processus de tamisage itératif. Les étapes de la décomposition en modes empiriques sont :

- ✓ Identification des données initiales ;
- ✓ Extraction des enveloppes des maxima et minima locaux par interpolation ;
- ✓ Calcul de la moyenne des enveloppes pour isoler une FMI ;
- ✓ Répétition du processus jusqu'à ce que toutes les composantes oscillatoires soient extraites.

La décomposition en modes empiriques permet une analyse locale des signaux et est particulièrement efficace pour des applications telles que la détection de défauts, l'analyse des vibrations, la surveillance des machines, l'analyse des séries temporelles, et bien d'autres domaines où les signaux présentent des variations complexes.

3.3.1 Principe de la décomposition en modes empiriques

Cette technique repose sur un processus itératif appelé *tamisage*, qui décompose un signal en une somme de composantes oscillantes appelées FMI. Ces FMI sont des fonctions de moyenne nulle avec des amplitudes et des fréquences modulées.

La méthode fonctionne de manière locale, en analysant le signal en fonction de ses caractéristiques propres, sans hypothèses prédéfinies sur sa linéarité ou sa stationnarité. La décomposition suit les étapes suivantes illustrées à la figure 3.2 :

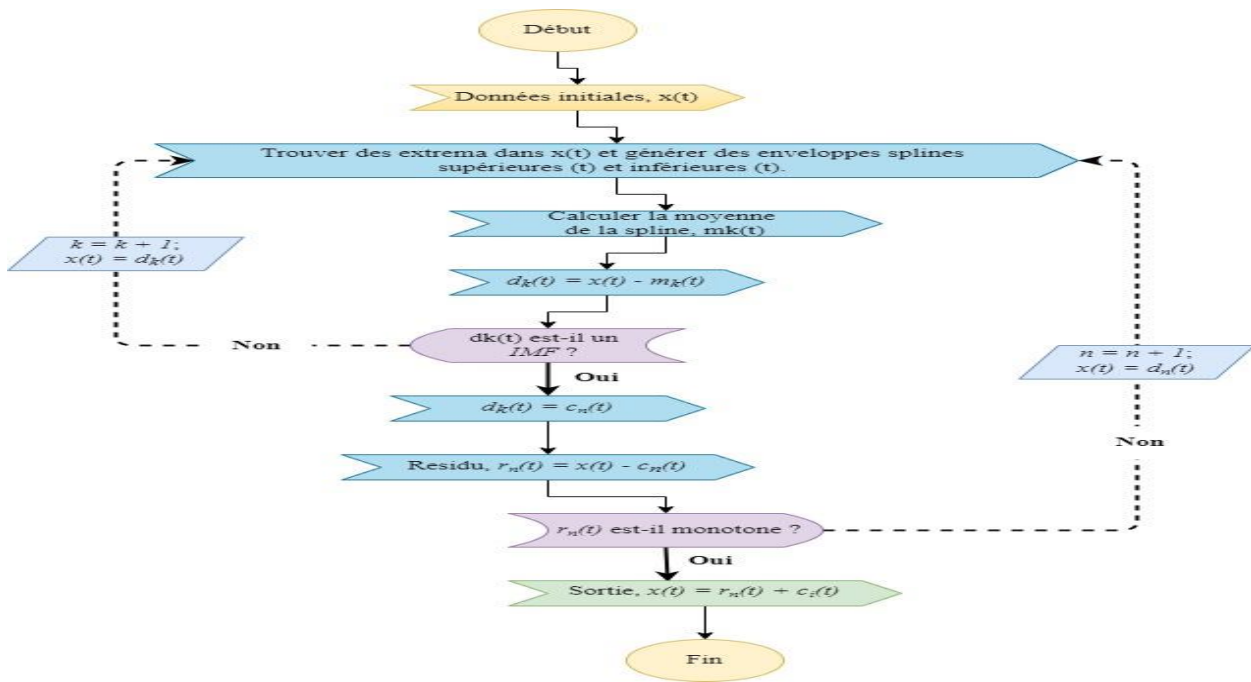


Figure 3.2: Organigramme de la technique de décomposition en modes empiriques

1. Identification des données initiales ($x(t)$).
2. Détection des maxima et minima locaux pour générer des enveloppes par interpolation spline.
3. Calcul de la moyenne des enveloppes ($d_k(t)$).
4. Tant que $d_k(t)$ n'est pas une FMI :
Calculer $d_k(t)$, incrémenter k , et réinitialiser $x(t) = d_k(t)$.
5. Une fois $d_k(t)$ validé comme FMI, extraire :
 $c_n(t) = d_k(t)$, et calculer le résidu $r_n(t) = x(t) - c_n(t)$.
6. Si $r_n(t)$ n'est pas monotone, reprendre à l'étape 2, sinon la décomposition s'arrête.

L'équation générale d'un signal décomposé par la DME est donnée par :

$$x(t) = r_n(t) + \sum_{i=1}^n C_i(t) \quad (3.14)$$

où : $x(t)$ est le signal d'origine,

$C_i(t)$ représente les i -èmes FMI,

$r_n(t)$ est la composante résiduelle (signal restant après décomposition),

n est le nombre total de FMI.

La décomposition en modes empiriques décompose ainsi un signal des fréquences les plus élevées vers les fréquences les plus basses, offrant une analyse multi-échelle naturelle.

3.3.2 Propriétés de la décomposition en modes empiriques

Les propriétés expliquent les critères techniques essentiels à une décomposition réussie.

➤ Critère d'arrêt

La décomposition en modes empiriques utilise deux boucles principales :

1. La première boucle (indexée par j) s'arrête lorsque le résidu actuel $r_j(t)$ ne peut plus être décomposé, c'est-à-dire lorsqu'il reste moins de deux extrema dans le signal.
2. La seconde boucle (indexée par i) vérifie si $h_{j,i}(t)$, une composante intermédiaire, répond aux critères d'un FMI.

Pour garantir un arrêt efficace du processus de tamisage, un critère d'écart type (ET) est introduit:

$$ET(i) = \sum_{i=0}^n \frac{|h_{j,i-1}(t) - h_{j,i}(t)|^2}{(h_{j,i-1}(t))^2} \quad (3.15)$$

Le processus s'arrête lorsque $ET(i)$ devient inférieur à un seuil prédéterminé ε , généralement compris entre 0,2 et 0,3 (Huang et al., 1998, 2003).

➤ Interpolation

L'interpolation joue un rôle crucial dans la génération des enveloppes des maxima et minima.

- Interpolation spline cubique : Fournit des résultats précis et limite l'apparition de modes excessifs.
- Comparée aux méthodes d'interpolation linéaire ou de plus proche voisin, la spline cubique offre une meilleure estimation des FMI, garantissant une décomposition plus fidèle du signal (Huang et al., 1998).

3.3.3 Avantages de la décomposition en modes empiriques

La technique présente plusieurs avantages qui expliquent son adoption dans divers domaines :

1. **Adaptabilité** : Elle ne nécessite aucune hypothèse préalable sur la linéarité ou la stationnarité du signal.
2. **Décomposition multi-échelle** : La DME sépare les composantes oscillantes dans un ordre décroissant de fréquence, permettant une analyse détaillée des signaux complexes.
3. **Application pratique** : Souvent utilisée comme filtre passe-bande, elle permet de détecter les fréquences de résonance dans les systèmes mécaniques et d'identifier des défauts de roulement.
4. **Flexibilité** : Elle peut être combinée avec la transformée de Hilbert pour analyser les *amplitudes instantanées* et les *fréquences instantanées* des FMI.

3.4 Décomposition en Modes Variationnels

La décomposition en modes variationnels (DMV), développée par Dragomiretskiy et Zosso (2013), est une méthode récente qui décompose un signal d'entrée $x(t)$ en un ensemble discret de modes intrinsèques quasi-orthogonaux u_k ayant des bandes limitées autour de fréquences centrales ω_k . Contrairement à la

DME, la DMV est basée sur une optimisation variationnelle qui minimise la bande passante de chaque mode autour d'une fréquence centrale.

Chaque mode est modélisé comme un signal modulé en amplitude et en fréquence, exprimé par :

$$u_k(t) = A_k(t) \cos(\phi_k(t)) \quad (3.16)$$

Où : $A_k(t) \geq 0$: l'enveloppe du signal, toujours non négative.

$\phi_k(t)$: la phase croissante définissant la modulation en fréquence.

$\omega_k(t) = \frac{d\phi_k(t)}{dt}$: la fréquence instantanée est définie comme la dérivée temporelle de la phase.

3.4.1 Principe de la décomposition en modes variationnels

La décomposition en modes variationnels repose sur la minimisation d'un modèle variationnel contraint, qui divise adaptativement le spectre fréquentiel en sous-bandes, en maximisant la séparation modale des composantes spectrales.

- ✓ **Transformation de Hilbert** : Permet de calculer les signaux analytiques pour chaque mode u_k .
- ✓ **Correction exponentielle** : Recentre chaque mode autour de sa fréquence centrale ω_k .
- ✓ **Lissage gaussien** : Réduit la bande passante et isole les composantes spectrales.

Dans la décomposition en modes variationnels, la décomposition d'un signal $x(t)$ en k composantes modales (u_k) repose sur la minimisation d'un modèle variationnel contraint. Ce modèle est formulé par les équations suivantes :

$$\min_{\{u_k\}, \{\omega_k\}} = \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} s. t. \sum_k u_k = x(t) \quad (3.17)$$

Pour résoudre ce problème, on utilise des multiplicateurs de Lagrange λ et une pénalité quadratique pour obtenir la solution optimale.

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \quad (3.18)$$

1. $x(t)$ (Signal observé)

Le signal $x(t)$ est le signal d'entrée que l'on souhaite décomposer. Il est supposé être constitué de plusieurs composantes modulées en fréquence et en amplitude.

2. $u_k(t)$ (Composantes modales)

Chaque $u_k(t)$ représente une composante oscillatoire du signal, appelée Fonction de Mode Intrinsèque (FMI). Ces modes ont les propriétés suivantes :

- Amplitudes modulées ($A_k(t)$) et fréquences modulées ($\omega_k(t)$).
- Confinement dans une bande spectrale autour de la fréquence centrale ω_k .
- Quasi-orthogonalité avec les autres modes.

3. ω_k (Fréquences centrales des modes)

Les ω_k sont les fréquences centrales autour desquelles chaque composante modale $u_k(t)$ est compactée dans le domaine spectral.

- Elles sont ajustées adaptativement pendant le processus itératif pour garantir une séparation spectrale optimale.

4. ∂_t (Dérivée temporelle)

La dérivée temporelle (∂_t) mesure les variations des modes dans le temps, permettant d'évaluer la bande passante de chaque mode. Une dérivée faible indique une bande passante étroite, tandis qu'une dérivée élevée indique une bande plus large.

5. $(\delta(t) + \frac{j}{\pi t})$ (Noyau analytique basé sur la transformation de Hilbert)

- $\delta(t)$ est la fonction impulsionnelle de Dirac.
- $\frac{j}{\pi t}$ est la partie imaginaire du noyau de Hilbert.

Ce noyau permet de calculer le signal analytique associé à chaque composante $u_k(t)$, ce qui facilite l'analyse des amplitudes et fréquences instantanées.

6. α (Facteur de régularisation)

Le paramètre α contrôle l'équilibre entre la bande passante des modes et leur ajustement au signal.

- ✓ Une valeur élevée de α favorise des modes plus lisses avec des bandes passantes plus étroites ;
- ✓ Une valeur faible permet une meilleure adaptation locale au signal mais peut augmenter la redondance.

7. $\lambda(t)$ (Multiplicateurs de Lagrange)

Les multiplicateurs $\lambda(t)$ imposent la contrainte de reconstruction du signal $x(t)$:

$$\sum_k u_k = x(t) \quad (3.19)$$

Ils assurent que la somme des composantes modales est égale au signal d'origine à chaque étape de l'optimisation.

8. $\| \cdot \|_2^2$ (Norme quadratique)

La norme quadratique mesure l'énergie des résidus dans les domaines temporel et spectral. Minimiser cette norme permet d'obtenir des modes optimaux en termes de bande passante et de reconstruction.

3.4.2 Propriétés de la décomposition en modes variationnels

Le processus itératif de la décomposition en modes variationnels repose sur la Méthode des Multiplicateurs de Direction Alternée pour trouver une solution optimale au problème variationnel. La convergence est garantie lorsque :

$$\frac{\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{a}_k^n\|_2^2} < \epsilon \quad (3.20)$$

où : ϵ est un seuil de tolérance prédéfini, généralement fixé entre 10^{-6} et 10^{-4} .

n est le numéro d'itération.

Les étapes clés pour la convergence est :

➤ **Mise à jour des modes (u_k) :**

Les modes u_k sont mis à jour en minimisant l'expression quadratique associée à leur bande passante.

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^{n+1}(\omega) + \left(\frac{\hat{\lambda}(\omega)}{2}\right)}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3.21)$$

➤ **Mise à jour des fréquences centrales (ω_k) :**

Les fréquences ω_k sont recalculées comme barycentres spectraux des modes :

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (3.22)$$

➤ **Mise à jour des multiplicateurs de Lagrange (λ) :**

$$\lambda^{n+1} = \lambda^n + \tau(f - \sum_k u_k^{n+1}) \quad (3.23)$$

où τ est le taux d'apprentissage ajustant la vitesse de convergence.

Cette technique repose sur une approche mathématique rigoureuse et une optimisation itérative pour décomposer un signal en modes distincts. Chaque paramètre joue un rôle essentiel dans la précision et la robustesse du processus. Une bonne compréhension de ces paramètres et des conditions de convergence est cruciale pour garantir des résultats fiables et pertinents dans les applications pratiques. En outre, aussi le diagramme illustré dans la figure 3.3 décrit la technique itérative de la décomposition en modes variationnels qui semble être utilisé pour résoudre un problème numérique spécifique. Chaque étape correspond à une action bien définie dans le processus de résolution.

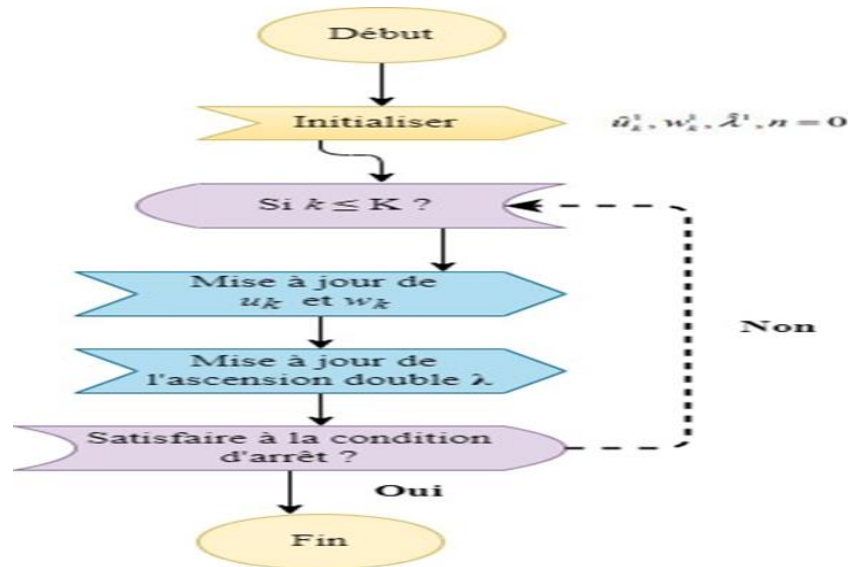


Figure 3.3: Organigramme de la technique de décomposition en modes variationnels

3.4.3 Avantages de la décomposition en modes variationnels

La décomposition en modes variationnels présente de nombreux avantages par rapport à d'autres techniques de décomposition, notamment la DME :

1. Séparation modale efficace : La DMV divise le spectre en bandes distinctes, réduisant les interférences entre modes.
2. Réduction des effets de bord : Les bordures des signaux ne perturbent pas la décomposition.
3. Robustesse au bruit : Contrairement à la DME, la DMV est moins sensible au bruit, ce qui en fait un choix privilégié pour les signaux bruyants.
4. Précision spectrale : La fréquence centrale et la largeur de bande de chaque composante sont déterminées de manière optimale.
5. Flexibilité : Adaptée à une large gamme d'applications, telles que le traitement des vibrations, la détection des défauts mécaniques et l'analyse des séries temporelles complexes.

3.5 Comparaison entre les techniques utilisées de décomposition

Pour mieux illustrer les différences entre ces techniques, nous pouvons nous référer au tableau ci-dessous qui compare la TO, la DME et la DMV selon différents critères.

Tableau 3.2: Comparaison des différentes techniques de décomposition

Critères	TO	DME	DMV
Principe	Analyse basée sur des bases fixes	Décomposition adaptative par tamisage	Décomposition variationnelle et itérative
Adaptativité	Fixe	Très élevée	Moyenne
Résistance au bruit	Moyenne	Faible	Elevée
Effets de bord	Limites dans certains cas	Problèmes fréquents	Minimisés
Robustesse spectrale	Moyenne	Moyenne	Élevée
Complexité algorithmique	Faible	Moyenne	Elevée
Applications	Signaux stationnaires ou quasi-stationnaires	Signaux non linéaires et non stationnaires	Signaux bruités et fortement complexes

3.6 Conclusion

Ce chapitre a présenté et comparé trois techniques de décomposition des signaux, démontrant leur efficacité pour analyser des signaux complexes en temps et en fréquence, tout en offrant des résolutions adaptées aux particularités des données.

La transformation en ondelettes (TO) se distingue pour les signaux stationnaires et quasi-stationnaires, elle est largement utilisée pour ses capacités d'analyse multi-résolution et sa simplicité algorithmique. Cependant, son adaptabilité est limitée, car elle repose sur des bases fixes prédéfinies. En contraste, la DME offre une flexibilité supérieure pour traiter les signaux non stationnaires, elle offre une décomposition naturelle en FMI. Toutefois, la DME est sensible au bruit et souffre d'effets de bord importants, limitant parfois sa robustesse dans des applications complexes. Enfin, la DMV est une méthode robuste qui réduit les effets de bord et s'adapte efficacement aux signaux bruités grâce à une optimisation variationnelle. Sa complexité algorithmique constitue néanmoins une limite pour certaines applications.

En synthèse, bien que chaque technique présente des avantages spécifiques, leur utilisation optimale dépend des caractéristiques du signal étudié et des objectifs de l'analyse. Ces outils, lorsqu'ils sont bien exploités, permettent d'améliorer considérablement la précision et l'efficacité dans le traitement des données complexes.

Chapitre 4 : Présentation de la Station d'Épuration d'Aïn Beïda

4.1 Introduction

La gestion des eaux usées constitue un enjeu majeur pour préserver les écosystèmes et protéger la santé publique. Les stations d'épuration des eaux usées (STEP) jouent un rôle central en traitant les rejets domestiques et industriels afin de garantir leur conformité avec les normes environnementales. La ville d'Aïn Beïda, est située au nord-est de l'Algérie, dispose d'une station d'épuration qui utilise un procédé de boues activées à faible charge.

Ce chapitre présentera les caractéristiques techniques ainsi que les différentes filières d'épuration de notre station, depuis la collecte des eaux usées jusqu'à leur purification finale, et met en évidence l'importance de cette infrastructure dans la gestion durable de l'eau et la protection de l'environnement.

Enfin, il abordera des exemples de modélisation utilisant des RNA combinés à des techniques de décomposition, appliquées aux stations d'épuration.

4.2 Localisation géographique de la station

La station d'épuration des eaux usées municipales de la ville d'Aïn Beïda est située à 5 km au nord-est d'Aïn Beïda, dans l'Oued Azzabi, à une altitude de 930 m (Figure 4.1). Mise en service en 2014, cette station s'appuie sur le procédé des boues activées à faible charge. Elle occupe une superficie de 10 hectares et est conçue pour épurer les eaux usées domestiques de la ville, avec une capacité de 210 000 EH et un débit unitaire de 25 320 m³/j par temps sec, prévu à l'horizon 2033 (Fiche technique, 2015).

Les coordonnées géographiques de la station, selon le système mondial de positionnement (GPS), sont :

Latitude : 35°47'22.24" Nord ;

Longitude : 7°20'27.18" Est.

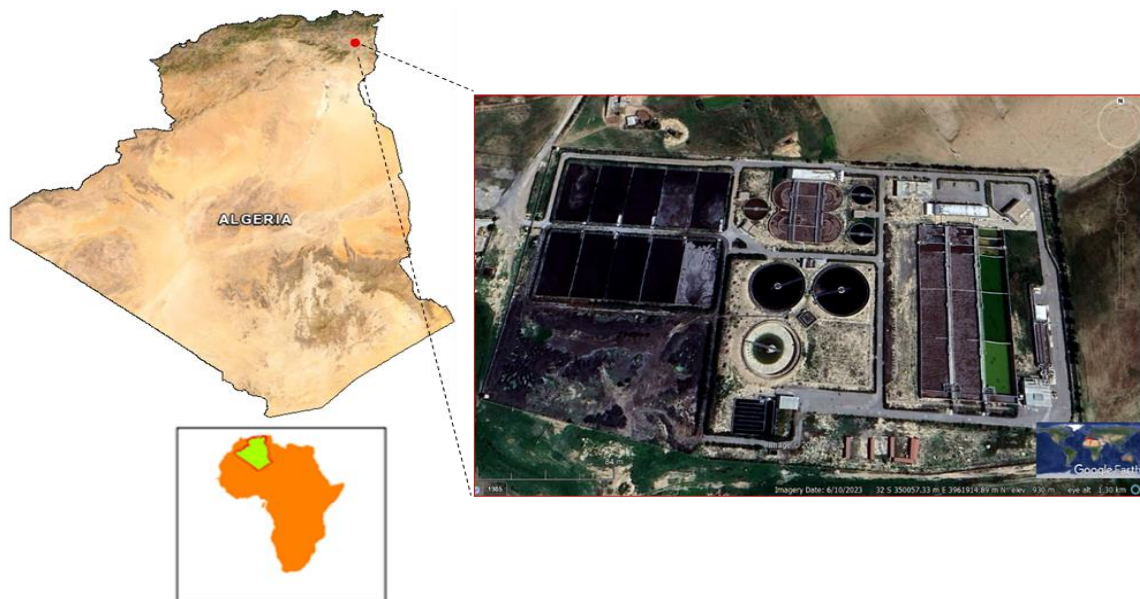


Figure 4.1: Localisation de la station d'épuration d'Aïn Beïda (Google, 2024)

4.3 Caractéristiques techniques de la station

Le tableau 4.1 ci-dessous présente les débits et la capacité de la station étudiée à différents horizons temporels. On observe une augmentation significative des paramètres entre 2015 et 2033, reflétant probablement une croissance démographique et une activité économique accrue dans la ville d'Aïn Beïda.

Tableau 4.1: Différents débits de dimensionnement de la station (Zaidi et al., 2023)

Paramètres	Unités	Horizon 2015	Horizon 2033
Capacité de la station	E H	140000	210000
Débit unitaire	m ³ /j	16840	25320
Débit maximal par temps sec	m ³ /h	1 179	1 741
Débit moyen par temps sec		702	1 055
Débit de pointe par temps de pluie		2 947	4 352

Les concentrations des polluants en eaux brutes et en eaux purifiées, ainsi que les rendements d'élimination attendus sont représentés dans le tableau 4.2. Ces paramètres sont essentiels pour évaluer l'efficacité du traitement et la qualité des effluents rejetés dans le milieu naturel.

Tableau 4.2: Concentrations des paramètres de pollution (Zaidi et al., 2023)

Paramètres	Concentration d'eau brute (mg/L)	Concentration d'eau purifiée (mg/L)	Rendement minimum d'élimination (%)
DBO ₅	449	30	93
DCO	847	80	90
MES	582	30	94
NT	81	40	51
PT	15	2	87

Le tableau ci-dessous, présente les charges journalières estimées pour différents paramètres de pollution dans la station d'épuration d'Aïn Beïda.

Tableau 4.3: Charge journalière des paramètres de pollution (Zaidi et al., 2023)

Paramètres	Charge par personne (g/EH.j)	Charge journalière Horizon 2015 (kg/jour)	Charge journalière Horizon 2033 (kg/jour)
DBO ₅	54	7560	11 340
DCO	102	14280	21 378
MES	70	9800	14 700
NT	9,7	1358	2046
PT	1,8	252	379

4.4 Présentation des procédés de traitement de la station

Le procédé d'épuration utilisé dans cette station est la boue activée à faible charge. Ce système se compose de deux filières principales : la première est dédiée au traitement des eaux usées, tandis que la deuxième

est destinée au traitement des boues (Figure 4.2), lesquelles permettent d'extraire progressivement les polluants présents dans les eaux usées.

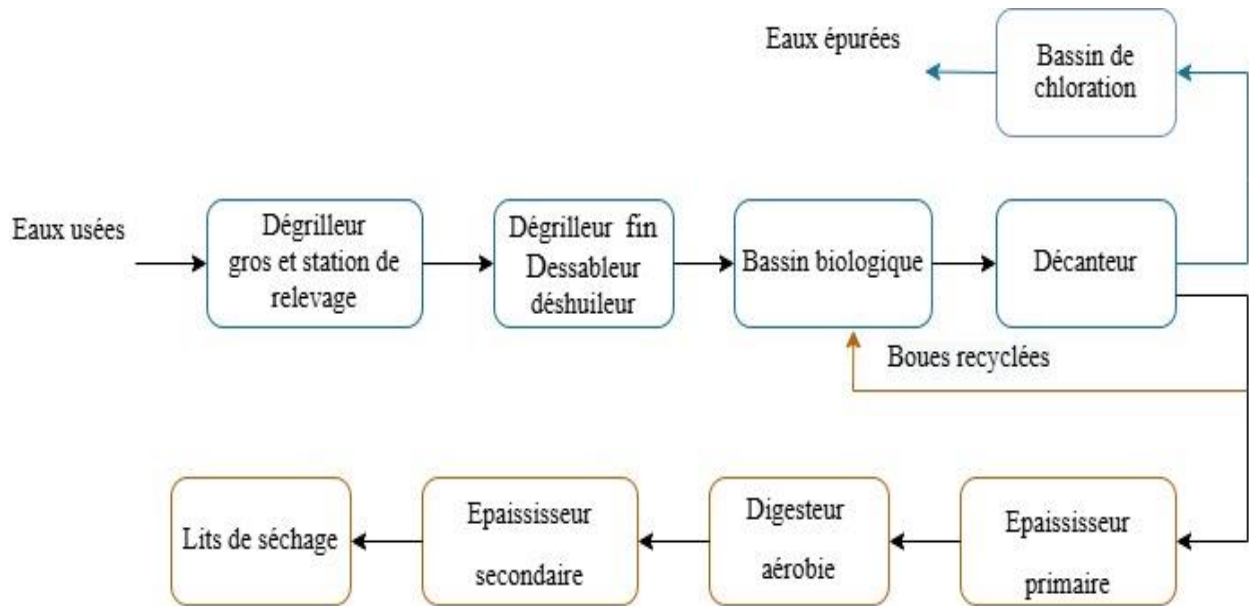


Figure 4.2: Schéma de procédé de la station d'épuration d'Aïn Beïda

4.4.1 Filière de traitement des eaux

Le traitement des eaux usées se déroule en deux grandes phases :

1. La purification physique (prétraitement) : elle regroupe une série d'opérations physiques visant à éliminer les déchets solides, tels que les MES volumineuses, les particules fines, le sable et les graisses ;
2. La purification biologique (traitement secondaire) : elle intervient après le prétraitement et s'appuie sur des procédés biologiques pour dégrader la matière organique dissoute.

4.4.1.1 Panier gros

La première étape du traitement consiste à faire passer l'eau brute à travers un panier gros équipé d'un dispositif manuel, dont les barres sont espacées de 10 cm. Ce dispositif retient environ 20 % des gros déchets, et les dépose dans la zone de récupération. Après cette étape, l'eau s'écoule par gravité vers la tête de station, où convergent trois conduites : l'eau brute, l'eau provenant du classificateur de sable et l'eau de retour.

4.4.1.2 By-pass

Le by-pass est conçu pour évacuer l'excédent d'eau en cas d'urgence, par exemple lors de la maintenance des équipements ou en cas de colmatage. L'eau excédentaire est alors dirigée directement vers l'Oued. Ce système est relié à un canal Venturi, permettant de mesurer le débit de l'eau évacuée.

4.4.1.3 Dégrilleur grossier

Avant le relevage des eaux usées, celles-ci passent à travers deux grandes grilles inclinées à 60°, avec un espacement de 4 cm entre les barreaux. Ces dégrilleurs capturent environ 40 % des déchets grossiers, qui sont ensuite transférés automatiquement vers une zone de récupération. Cet équipement joue un rôle essentiel en protégeant les pompes de relevage contre les obstructions.

4.4.1.4 Poste de relevage

Après l'élimination des gros déchets, les eaux usées sont collectées dans un puits. Ce puits est équipé de quatre pompes submersibles, chacune ayant une capacité de 1100 m³/h et une hauteur manométrique totale de 14 m. Ces pompes permettent de relever les eaux usées, notamment en période de pluie, avec une capacité maximale de 2945 m³/h lorsque trois pompes fonctionnent simultanément. Les eaux ainsi relevées sont ensuite dirigées vers les dégrilleurs fins.

4.4.1.5 Dégrilleur fin

Après le relevage, les eaux usées passent à travers deux dégrilleurs fins inclinés à 60°, avec un espacement de 0,8 cm entre les barreaux. Ces dégrilleurs permettent d'éliminer environ 30 % des déchets fins. Les déchets ainsi capturés sont transférés automatiquement vers une zone de récupération. Ce dispositif est crucial pour protéger les équipements mécaniques en éliminant les matériaux fins susceptibles de les endommager.

4.4.1.6 Dessableur-déshuileur

Les eaux usées sont ensuite dirigées vers un dessableur-déshuileur aéré, de forme trapézoïdale, conçu pour éliminer environ 10 % des sables et des particules minérales. Cet équipement réduit l'abrasion des installations en aval et l'encrassement des canalisations. Cet ouvrage est divisé en deux canaux distincts :

- **Zone d'aération** : l'air est insufflé pour faciliter la séparation des particules de sable et des graisses ;
- **Zone calme** : permet l'accumulation des graisses en surface.

Un pont racleur mobile extrait simultanément les sables et les graisses :

- Les sables sont dirigés vers deux classificateurs, où une vis d'Archimède les sépare de l'eau, avant leur collecte dans la zone de récupération ;
- Les graisses flottantes sont raclées vers une fosse à graisse, puis pompées vers un digesteur pour traitement ;
- L'eau résiduelle est ensuite renvoyée par gravité vers la tête de la station pour les étapes suivantes du traitement.

Après cette étape de prétraitement physique, l'eau est transférée dans le bassin biologique pour un traitement biologique visant à éliminer la matière organique et minérale.

4.4.1.7 Bassin biologique

La station d'épuration fonctionne à faible charge avec un procédé de culture libre par boues activées. Le bassin biologique est conçu pour éliminer la pollution organique et minérale. L'eau prétraitée s'écoule par gravité et rejoint la zone de contact avec la boue de retour provenant du décanteur. Ce bassin est constitué de trois canaux symétriques, chacun composé de trois compartiments interconnectés. Ces compartiments forment un écoulement renforcé dans une configuration géométrique rectangulaire. Chaque compartiment a une fonction spécifique, contribuant à l'épuration de l'eau de manière séquentielle. Les transitions entre les différentes zones sont bien définies :

➤ **Zone anaérobie**

Dans ce réacteur, l'absence d'oxygène favorise l'activité des bactéries anaérobies, qui déphosphatent le phosphore minéral tout en consommant la matière organique. À la fin de cette étape, les bactéries libèrent des molécules de phosphate (PO_4^{3-}). L'eau prétraitée passe ensuite dans une zone légèrement oxygénée, dite "zone anoxique", pour poursuivre le processus d'épuration biologique.

➤ **Zone anoxie**

Cette zone, pauvre en oxygène, est dédiée à la dénitrification. Cette réaction biologique est effectuée par des bactéries dénitrifiantes en condition anoxique (absence d'oxygène). Ces bactéries utilisent les nitrates (NO_3^-) comme source d'oxygène pour leur métabolisme. Ce processus réduit les nitrates en nitrites (NO_2^-), puis en azote gazeux (N_2), qui est libéré dans l'atmosphère. La dénitrification est essentielle pour limiter les excès de nitrates dans l'eau, responsables de la pollution des nappes phréatiques et des phénomènes d'eutrophisation. Après la dénitrification, l'eau passe dans la zone aérobie, où une oxygénation active permet d'éliminer la pollution carbonée et d'achever le processus de nitrification.

➤ **Zone aérobie**

Cette zone, riche en oxygène, est alimentée par un suppresseur d'air injectant de l'oxygène à travers des disques à membrane percés de micro-orifices (1 mm). Dans cette étape, les bactéries éliminent 70 à 80 % de la pollution carbonée, tout en procédant à la nitrification.

La nitrification est une réaction chimique réalisée par des bactéries nitrifiantes en condition aérobie (présence d'oxygène). Elle permet de transformer l'ammonium, souvent toxique pour les organismes aquatiques, en nitrates, une forme moins toxique et plus facilement utilisable par les plantes.

Ces deux processus jouent un rôle crucial dans le cycle de l'azote, en maintenant l'équilibre des composés azotés dans l'environnement et en réduisant leur impact négatif sur les écosystèmes aquatiques.

Les nitrates produits sont ensuite recirculés vers la zone anoxique par pompage pour poursuivre leur élimination. La concentration idéale d'oxygène dissous dans le bassin biologique est de 2 mg/L, avec une concentration de MES des boues activées de 4 g/L.

L'eau purifiée et les boues activées passent ensuite par le déversoir pour une ultime étape de traitement avant d'atteindre le puits de distribution.

➤ **Déversoir (bassin de dégazage)**

Dans cette dernière partie du bassin biologique, le chlorure de fer (FeCl_3), est utilisé pour éliminer les phosphates minéraux résiduels (PO_4^{3-}) par un dosage à l'aide d'un acide fort. Cette étape permet également de libérer les gaz accumulés durant le traitement. Le mélange final, composé d'eau purifiée et de boues activées, s'écoule ensuite par gravité vers le puits de distribution, avant d'être acheminé aux décanteurs pour une séparation finale. Ces étapes successives du bassin biologique assurent une épuration efficace de l'eau, grâce à une gestion optimisée des différentes zones fonctionnelles.

4.4.1.8 Décanteur

Les eaux purifiées sont dirigées vers trois décanteurs (deux en fonctionnement et un en réserve). Ces décanteurs, de forme conique, sont équipés d'un pont racleur rotatif en acier inoxydable. Le pont racleur gratte les flottants vers un puits dédié, où ils sont ensuite pompés vers le digesteur. Il facilite également le dépôt des boues au fond du dispositif, où la concentration de MES atteint environ 8 g/L. Ces décanteurs assurent une séparation efficace des phases :

Les eaux purifiées sont dirigées vers trois décanteurs (deux en fonctionnement et un en réserve). Ces décanteurs, de forme conique, sont équipés d'un pont racleur rotatif en acier inoxydable. Le pont racleur gratte les flottants vers un puits dédié, où ils sont ensuite pompés vers le digesteur. Il facilite également le dépôt des boues au fond du dispositif, où la concentration de MES atteint environ 8 g/L. Ces décanteurs assurent une séparation efficace des phases :

- **Phase liquide** : eau purifiée évacuée directement vers le bassin de chloration ;
- **Phase solide** : boues activées collectée dans un puits de boues décantées, équipé de cinq pompes (trois pour recirculer les boues vers la zone de contact et deux pour les envoyer vers l'épaississeur primaire).

Une fois l'eau purifiée acheminée et les boues séparées, celle-ci passe à l'étape de désinfection dans le bassin de chloration.

4.4.1.9 Bassin de chloration

Ce bassin, conçu en forme de chicane, est installé à la sortie de la station. Il permet un mélange homogène des eaux épurées avec une solution de désinfection à base d'hypochlorite de sodium (NaOCl). Après chloration, l'eau est évacuée vers Oued Azzabi.

4.4.2 Filière de traitement des boues

Parallèlement au traitement des eaux, la station gère également les boues issues du processus d'épuration dans une filière dédiée, ayant pour objectif de réduire le volume des boues produites.

4.4.2.1 Epaisseur primaire

Après le clarificateur, les boues excédentaires sont dirigées vers deux épaisseurs primaires coniques, équipés d'un pont racleur rotatif. Concentration de MES environ 25 g/L dans les boues épaissies. Ces boues sont pompées vers le digesteur, tandis que le surnageant est collecté dans le puits des eaux de retour et pompé vers la tête de la station pour de nouveau être traité. Ces boues passent alors à l'étape de digestion aérobie pour être stabilisées biologiquement.

4.4.2.2 Digesteur aérobie

Ce digesteur, de forme oblongue et aérée, reçoit les boues épaissies, les flottants et les huiles. Une aération est assurée par un surpresseur d'air injectant de l'oxygène à travers des membranes avec des orifices de 2 mm. Les bactéries digèrent les boues et consomment les matières organiques restantes. Une fois les ressources épuisées, ces bactéries meurent, ce qui permet une stabilisation aérobie des boues. Concentration de MES comprise entre 12 et 18 g/L dans les boues digérées, qui sont ensuite envoyées vers un épaisseur secondaire pour une réduction supplémentaire de leur volume.

4.4.2.3 Epaisseur secondaire

Les boues digérées sont traitées dans un épaisseur secondaire, où de la chaux (CaCO_3 : carbonate de calcium) est ajoutée pour accélérer le séchage et minimiser la libération des phosphates. La concentration en MES y est d'environ 45 g/L. Les boues épaissies sont ensuite pompées vers des lits de séchage, tandis que le surnageant est collecté dans un puits des eaux de retour et renvoyé vers la tête de la station pour un retraitement. Ces boues stabilisées sont prêtes pour leur ultime étape : le séchage.

4.4.2.4 Lits de séchage

La station est équipée de dix lits de séchage de forme rectangulaire, avec une base en béton imperméable recouverte de gravier et de sable pour un séchage naturel efficace. Les boues épaissies sont réparties dans ces lits par pompage. L'eau filtrée est collectée dans le radier et acheminée vers un puits d'eaux de retour, où elle est pompée vers la tête de la station. Les boues séchées atteignent 45 % de siccité, facilitant leur manipulation et leur valorisation ultérieure.

4.4.3 Laboratoire

Le laboratoire de la station joue un rôle essentiel dans le suivi de la qualité de l'eau. Des analyses sont régulièrement effectuées pour mesurer les paramètres physico-chimiques, surveiller les paramètres de pollution, et évaluer l'état et la concentration des boues. Ces analyses permettent d'assurer un contrôle rigoureux du processus d'épuration.

Le rôle du laboratoire s'intègre dans une chaîne d'opérations globales qui sont surveillées et contrôlées depuis la salle de commande, qui constitue le centre névralgique de la station.

4.4.4 Salle de commande

La salle de commande constitue le cœur opérationnel des installations automatisées. Elle permet de démarrer, arrêter et contrôler l'état de fonctionnement de chaque équipement de la station. Elle offre également la possibilité de modifier et enregistrer tous les paramètres liés au processus de traitement. Au cœur de cette salle, différents dispositifs coopèrent assurant une gestion optimale des opérations. On y trouve :

- Pupitre de commande ;
- Ecran de surveillance ;
- Système téléphonique ;
- Antenne de la télégestion.

4.5 Modélisation par RNA et techniques de décomposition dans les STEP (Études Antérieures)

L'utilisation des réseaux de neurones artificiels pour modéliser et prédire les performances des STEP a fait l'objet de plusieurs études récentes, mettant en évidence leur efficacité pour traiter des systèmes complexes et non linéaires.

- **Lou et Zhao (2012)**

Ces chercheurs ont appliqué des techniques d'identification basées sur l'analyse des entrées et sorties d'un système de boues activées, incluant des paramètres tels que la température, le PH, la DBO, la DCO, les MES, NH_4^+ , NT, PT, et d'autres indicateurs. Ils ont utilisé l'Analyse en Composantes Principales et les RNA avec des données provenant de la STEP de Chongqing. Les résultats ont montré que le RNA surpassait l'autre modèle dans la prédiction de l'indice de boue (IB), ce qui en fait un outil plus fiable pour le contrôle de l'emballage des boues.

- **Zare Abyaneh (2014)**

Cette étude a comparé l'efficacité des RNA et des modèles de régression linéaire multiple pour prédire deux paramètres clés de la qualité de l'eau : la DBO_5 et la DCO. Les RNA offrant une meilleure précision dans la prédiction de ces paramètres par rapport l'autre modèle.

- **Djeddou et Achour (2015)**

Leur étude a utilisé des RNA pour prédire l'IB dans une STEP algérienne, en intégrant des paramètres comme la température, le pH, la CE et les concentrations en DBO, DCO, MES, NH_4^+ , et PO_4^{3-} . Le meilleur modèle, basé sur un PMC, démontrant la capacité des RNA à estimer avec précision l'IB.

- **Najafzadeh et Zeinolabedini (2018)**

Pour améliorer la précision dans la prédiction du débit de boue séchée, ces chercheurs ont étudié des modèles simples comme la Programmation par Expression Génétique (PEG), l'Arbre de Modèle (AM) et la Régression Polynomiale Evolutionnaire (RPE), ainsi que leurs versions hybrides couplées à l'ondelette (TOD-PEG, TOD-AM et TOD-RPE). Ces modèles utilisaient des données d'entrée transformées par la fonction Discrète Meyer (ondelette mère) pour extraire les informations temporelles et spectrales. L'analyse comparative a montré que les modèles hybrides avec ondelette surpassaient systématiquement les modèles simples. Parmi eux, le TOD-AM a obtenu les meilleures performances. Ces résultats soulignent que l'intégration des ondelettes dans les modèles de prédiction améliore considérablement la précision, en particulier pour des paramètres cruciaux comme le débit de boue séchée dans les stations d'épuration.

- **Zeinolabedini et Najafzadeh (2019)**

En poursuivant cette recherche, les deux chercheurs ont étudié l'intégration de diverses fonctions d'ondelette mère dans les architectures de RNA pour accroître la précision de la prédiction du volume de boue séchée. Les performances des modèles TOD-RNA, incluant le PMC et le FBR, ont été évaluées en utilisant six fonctions d'ondelette mère comme préprocesseurs. Un retard de quatre jours a été considéré comme variable d'entrée pour les phases d'entraînement et de test. Les résultats montrent que les modèles hybrides surpassaient largement les modèles simples. L'utilisation de la fonction DMeyer a particulièrement amélioré la précision des RNA, avec des résultats optimaux pour les modèles hybrides.

Ces études démontrent que l'intégration de fonctions d'ondelette mère dans les modèles RNA améliore significativement leur précision, en faisant des approches TOD-RNA des outils puissants pour prédire le volume de boue séchée dans les STEP. La combinaison des ondelettes et des RNA offre une méthodologie robuste et fiable pour des analyses complexes liées au traitement des eaux usées.

- **Al Saleh (2021)**

Cette étude, des modèles RNA ont été développés pour prédire la DBO₅, la DCO et les MES dans les effluents. Les performances les plus élevées ont été observées pour la DBO₅. Les résultats ont montré que les modèles à sorties multiples étaient supérieurs aux modèles à sortie unique pour la prédiction de certains paramètres, notamment grâce à leur capacité à capturer les relations complexes entre les variables.

- **Wongburi et Park (2022)**

Leur étude a exploré l'utilisation d'un réseau de neurones récurrents pour prédire l'IB dans une STEP aux États-Unis. Le modèle a démontré son efficacité pour gérer des séries temporelles complexes, tout en identifiant les paramètres d'entrée influençant le plus l'IB. Ce travail a mis en lumière l'intérêt des RNA pour améliorer la gestion et la fiabilité des STEP.

- **Sahadev et al. (2023)**

Ces chercheurs ont utilisé un PMC pour prédire les paramètres de qualité de l'eau dans des STEP indienne. Les RNA ont obtenu des performances élevées, démontrant leur capacité à modéliser avec précision les processus complexes de traitement des eaux usées.

Ces travaux montrent une évolution constante dans l'application des RNA pour modéliser et prédire les performances des STEP, avec des approches variées. Les résultats obtenus confirment l'efficacité des RNA dans la gestion des relations non linéaires et des interactions complexes entre les paramètres, offrant ainsi des solutions fiables pour optimiser les processus des STEP. Ces approches, associées à des méthodologies avancées, ouvrent la voie à des améliorations substantielles dans le domaine du traitement des eaux usées.

Les réseaux de neurones artificiels, bien qu'efficaces pour la modélisation des processus complexes, présentent certaines limitations, notamment leur sensibilité à la qualité des données d'entrée et leur difficulté à converger vers des solutions optimales sans ajustement manuel précis. Pour surmonter ces limites, les chercheurs ont exploré des modèles hybrides combinant des RNA avec les techniques de décomposition afin d'améliorer leur précision et leur robustesse.

4.6 Conclusion

La station d'épuration d'Aïn Beïda est une installation moderne et essentielle pour la gestion des eaux usées dans la région. Grâce à son système d'épuration basé sur le procédé de boues activées, elle purifie les effluents domestiques, contribuant ainsi à réduire la pollution des eaux de surface et à préserver les ressources naturelles. La structure rigoureuse de ses filières de traitement, qu'il s'agisse de l'eau ou des boues, assure une élimination progressive et efficace des polluants. Au-delà de son rôle dans la protection de l'environnement et de la santé publique, cette station favorise la valorisation des ressources, notamment par la réutilisation potentielle des eaux épurées et des boues à des fins agricoles. Ce rôle clé en fait un élément essentiel pour une gestion durable des ressources en eau.

En plus de sa contribution à la protection de l'environnement et de la santé publique, cette station favorise la valorisation des ressources, notamment à travers la réutilisation potentielle des eaux épurées et des boues pour des usages agricoles. Ce rôle clé dans la gestion intégrée de l'eau démontre l'importance des infrastructures d'assainissement dans le développement durable des territoires. Ce chapitre offre ainsi une base solide pour comprendre le fonctionnement global de la station et son impact environnemental.

Chapitre 5 : Analyse des Résultats des Modèles Prédicatifs pour la STEP d'Aïn Beïda

5.1 Introduction

Dans ce chapitre, nous présentons une étude comparative de deux approches différentes : la modélisation simple et la modélisation hybride. Ces approches reposent sur l'intelligence artificielle (IA) pour prédire les paramètres de pollution à la sortie de la station, l'indice de boues (IB), et le débit journalier des boues séchées (QBS). Les modèles analysés incluent trois types simples de réseaux de neurones artificiels (RNA) : le perceptron multicouche (PMC), la fonction à base radiale (FBR) et la régression générale (RG). En parallèle, nous étudions des modèles hybrides constitués de trois techniques de décomposition combinées avec les RNA précédemment mentionnés : la transformation en ondelettes discrètes (TOD-RNA), la décomposition en modes empiriques (DME-RNA), et la décomposition en modes variationnels (DMV-RNA).

Dans le cadre de cette thèse, l'idée a été de combiner ces techniques avec les réseaux de neurones artificiels pour modéliser et prédire les paramètres de la qualité des effluents dans les STEP.

5.2 Description et collecte de données

Pour réaliser une analyse statistique, il a été nécessaire de collecter une grande quantité de données afin de constituer une base de données importante pour ce projet de recherche. Nous avons utilisé les données d'auto-surveillance journalière pour établir une base de données des différents paramètres de fonctionnement de la station d'épuration d'Aïn Beïda. Les paramètres étudiés ont été choisis parmi ceux couramment utilisés pour évaluer la performance du processus de traitement des eaux usées. Les données ont été recueillies sur une période de vingt-neuf mois, allant de décembre 2014 à avril 2017, et la base de données construite avec le logiciel Microsoft Office Excel a montré des fluctuations selon les saisons et les périodes de l'année. L'étude a utilisé un total de 531 entrées, qui, après décomposition, ont donné lieu à 144 963 entrées exploitables. Cette base de données constitue la fondation nécessaire pour procéder au prétraitement des données.

5.3 Prétraitement des données

La mise à l'échelle des données est une méthode standard pour préparer les données qui garantit une distribution statistique approximativement uniforme des valeurs entre les entrées et les sorties. La qualité des données de sortie dépend fortement de la qualité des données d'entrée des performances du processus de traitement et de techniques managériales des exploitants de la station.

Le prétraitement des données est une étape très importante en raison de la diversité des informations utilisées et acquises à partir du contrôle opérationnel et de l'échantillonnage manuel. Toutes les données brutes ont été examinées et vérifiées pour garantir leur complétude. Les données expérimentales ont été affinées en éliminant toutes les valeurs aberrantes, considérées comme des points inhabituels (anormaux).

Pour réduire les fluctuations, nous avons utilisé la technique des moyennes des trois valeurs avant et des trois valeurs après, en appliquant un affinage qui élimine les mesures se situant en dehors de la plage de ± 3 écarts-types, selon la technique de pro-procession dans le logiciel Matlab. Cette méthode fait référence à la moyenne du groupe ou de la cellule de conception (Zaidi et al., 2023).

La mise à l'échelle des données est une méthode standard pour préparer celles-ci, garantissant une distribution statistique approximativement uniforme des valeurs entre les entrées et les sorties. La qualité des données de sortie dépend fortement de la qualité des données d'entrée, du processus de traitement et des techniques managériales des exploitants de la station. Les ensembles de données ont été normalisés dans une plage de 0 à 1, en utilisant 80 % des données pour l'entraînement et les 20 % restantes pour la phase de test (Zaidi et al., 2023).

Cette modélisation a été réalisée avec un ensemble de données comportant 425 entrées pour la phase d'entraînement, couvrant la période du 1^{er} décembre 2014 au 26 octobre 2016. Pour la phase de test, 106 données ont été utilisées, allant du 29 octobre 2016 au 26 avril 2017. Il est à noter que les données sont particulièrement manquantes les jours fériés et pendant les week-ends.

5.4 Analyse statistique des données

L'analyse de certains paramètres peut être utilisée comme un outil d'évaluation de la performance de la station d'épuration. Nous avons choisi de nous concentrer sur les paramètres opérationnels et qualitatifs, basés sur les données disponibles, pour évaluer la performance de la station d'Aïn Beïda. Les paramètres suivants ont servi à la construction des modèles de RNA, simples et hybrides :

➤ **Paramètres physico-chimiques des eaux brutes**

On utilise le potentiel hydrogène (pH), la Température (t) et la Conductivité électrique (CE).

➤ **Paramètres de pollution des eaux brutes**

On utilise la Demande biochimique en oxygène (DBO_5), la Demande chimique en oxygène (DCO), les Matières en suspension (MES), l'Azote total (NT) et le Phosphore total (PT),

➤ **Paramètres de pollution des eaux épurées**

On a utilisé les mêmes paramètres que précédemment, mais à la sortie de la station, à l'exception du phosphore total, remplacé par les orthophosphates (PO_4^{3-}).

➤ **Paramètres spécifiques**

On a utilisé l'Indice de boues (IB) et débit journalier de Boues séchées (QBS).

Tableau 5.1: Synthèse statistique des différents paramètres de la station

Paramètres	Minimum	Maximum	Moyenne	ET	Cv	Cas	Cap
Paramètres d'entrées							
pH	7,33	8,7	7,93	0,16	0,02	-5,03 E-06	0,92
T (°C)	8	23,2	16,58	3,35	0,2	0,26	-1,08
CE (mS/cm)	0,64	8,2	1,64	0,38	0,23	9,80	164,37
DBO ₅ (mg/L)	100	742	366,37	84,49	0,23	0,30	0,97
DCO (mg/L)	195	1689	706,54	194,80	0,28	0,88	2,51
MES (mg/L)	100	1820	430,79	154,57	0,36	2,23	13,47
NT (mg/L)	37,8	152	84,16	14,89	0,18	0,72	1,73
PT (mg/L)	2,8	14,6	7,69	1,58	0,21	0,10	1,79
Paramètres de sorties							
DBO ₅ (mg/L)	0,5	17	3,14	2,19	0,7	2,02	6,15
DCO (mg/L)	19,10	97,7	37,7	11,06	0,29	1,56	3,38
MES (mg/L)	2	39	12,87	4,85	0,38	1	2,55
NT (mg/L)	14,8	65,1	33,88	6,74	0,2	0,95	2,87
PO ⁻³ ₄ (mg/L)	1,26	5,2	3,46	0,83	0,24	-0,34	0,43
IB (mL/g)	22,45	81,24	42,5	7,92	0,19	0,51	1,83
QBS (m ³ /j)	4,47	738,58	82,13	71,40	0,87	5,47	36,29

Cette analyse statistique permet de mieux comprendre les performances du système de traitement tout en assurant sa conformité avec les exigences légales. Le tableau 5.1 ci-dessous offre une vue d'ensemble quantitative et qualitative contenant des statistiques descriptives relatives aux paramètres d'entrée et de sortie liés au système de traitement des eaux usées. L'analyse approfondie des variations permet d'évaluer la stabilité des conditions d'entrée et l'efficacité du processus, tout en suggérant des pistes d'amélioration.

1. Paramètres physico-chimiques des eaux brutes

Le potentiel hydrogène (pH) oscille entre 7,33 et 8,7, avec une moyenne légèrement basique de 7,93 et un faible écart type de 0,16, indiquant ainsi une stabilité générale du pH. Le coefficient de variation (0,02) confirme cette faible variabilité. La distribution est légèrement symétrique et peu aplatie (Cas ≈ 0, Cap = 0,92), ce qui est favorable au bon fonctionnement des microorganismes épurateurs. Cependant, la valeur maximale de 8,7, probablement attribuable à des rejets industriels ponctuels, pourrait occasionnellement perturber les performances biologiques du traitement.

La température oscille entre 8 et 23,2 °C, avec une moyenne de 16,58 °C, un écart type de 3,35 °C, et un coefficient de variation modéré (0,2). La distribution présente une légère asymétrie (Cas = 0,26) et une forme plus aplatie (Cap = -1,08). Ces variations sont liées à la saisonnalité, avec des températures plus basses en hiver et plus élevées en été. Cette fluctuation peut affecter les performances biologiques, car la croissance et l'activité des microorganismes épurateurs diminuent lorsque la température diminue.

Les valeurs de la conductivité électrique (CE) s'étendent de 0,64 à 8,2 mS/cm, avec une moyenne de 1,64 mS/cm, un écart type de 0,38, et une variabilité modérée ($C_v = 0,23$). La distribution est fortement asymétrique ($C_{as} = 9,80$) et très pointue ($C_{ap} = 164,37$), traduisant la présence de valeurs élevées et rares. Cette dispersion signale une minéralisation significative, résultant des substances dissoutes (sels), ainsi que des apports d'eaux pluviales et des effluents industriels. Les pics dépassant 3 mS/cm sont probablement liés à des rejets spécifiques ou accidentels, pouvant perturber les microorganismes sensibles à la salinité.

2. Paramètres de pollution des eaux brutes

La demande biochimique en oxygène après cinq jours (DBO_5) varie entre 100 et 742 mg/L, avec une moyenne de 366,37 mg/L et un écart type de 84,49 mg/L, indiquant une variabilité modérée ($C_v = 0,23$). La distribution est légèrement asymétrique ($C_{as} = 0,30$) et peu aplatie ($C_{ap} = 0,97$), ce qui témoigne d'une charge organique significative mais relativement stable. Ces niveaux de DBO_5 signifient une charge biodégradable considérable, typique des eaux usées domestiques et urbaines mixtes.

La demande chimique en oxygène (DCO) présente une grande variabilité, avec des valeurs fluctuantes entre 195 et 1689 mg/L, et une moyenne de 706,54 mg/L. L'écart type est de 194,8 mg/L. La variabilité ($C_v = 0,28$) reste modérée, mais la distribution montre une faible asymétrie (0,88) et un aplatissement élevé (2,51), indiquant la présence sporadique ponctuelle de fortes concentrations de composés chimiques non biodégradables. Ces valeurs dispersées signalent une pollution chimique significative, potentiellement exacerbée par des rejets industriels ou des pollutions diffuses.

Les matières en suspension (MES) montrent également une large variabilité, avec des valeurs allant de 100 à 1820 mg/L, une moyenne de 430,79 mg/L et un écart type de 154,57 mg/L. Le coefficient de variation est élevé ($C_v = 0,36$), et la distribution affiche une forte asymétrie (2,23) et un aplatissement important (13,47). Ces fluctuations sont caractéristiques d'événements ponctuels, tels que des apports pluviaux ou des variations saisonnières des charges organiques et minérales.

Les concentrations d'azote total (NT) varient entre 37,8 et 152 mg/L, avec une moyenne de 84,16 mg/L et un écart type de 14,89 mg/L, indiquant une variabilité faible à modérée ($C_v = 0,18$). La distribution est légèrement asymétrique (0,72) et légèrement aplatie (1,73), montrant une certaine stabilité dans les rejets azotés, bien que des pics modérés puissent survenir, probablement en raison d'apports industriels ou agricoles.

Le phosphore total (PT) oscille entre 2,8 et 14,6 mg/L, avec une moyenne de 7,69 mg/L et un écart type de 1,58 mg/L, indiquant une variabilité modérée ($C_v = 0,21$). La distribution présente une asymétrie faible (0,10) et un aplatissement modéré (1,79), ce qui traduit une légère dispersion des valeurs. Ces niveaux

suggèrent des rejets modérés en phosphore, souvent associés à des détergents ou à des apports d'origine domestique.

Ces données indiquent que les eaux brutes présentent une pollution organique et minérale significative, mais globalement cohérente avec les eaux usées urbaines mixtes. La variabilité des MES, de la DCO et des pics de DBO_5 souligne l'importance de surveiller les apports irréguliers, notamment en liaison avec des événements pluviaux ou des rejets industriels. Les concentrations en azote total et phosphore total, relativement stables, traduisent une charge régulière en nutriments. Ces observations mettent en évidence la nécessité d'une gestion appropriée pour optimiser l'efficacité des traitements et minimiser les perturbations dans les processus épuratoires.

3. Paramètres de pollution des eaux épurées

Les eaux épurées affichent des concentrations de DBO_5 variant de 0,5 à 17 mg/L, avec une moyenne de 3,14 mg/L et un faible écart type de 2,19 mg/L. Le coefficient de variation est relativement élevé ($C_v = 0,7$), et la distribution est marquée par une forte asymétrie (2,02) et un aplatissement significatif (6,15), ce qui indique une réduction substantielle de la charge organique.

La demande chimique en oxygène (DCO) s'étend de 19,1 à 97,7 mg/L, avec une moyenne de 37,7 mg/L et un écart type de 11,06 mg/L. Le coefficient de variation est de 0,29, et la distribution est marquée par une forte asymétrie (1,56) et un aplatissement élevé (3,38), indiquant un rendement global de 94,66 % dans l'élimination de la pollution chimique.

Les rapports DCO/ DBO_5 , compris entre 0,74 et 2,98 pour 97 % des échantillons, se situent dans la plage normale pour les eaux usées municipales. Ces ratios indiquent une bonne biodégradabilité des effluents, justifiant l'utilisation du procédé de boues activées, qui est particulièrement adapté aux effluents contenant une proportion significative de matière organique facilement dégradable (Spellman, 2013 ; Tchobanoglous, 2003).

Les matières en suspension (MES) dans les eaux traitées se situent entre 2 et 39 mg/L, avec une moyenne de 12,87 mg/L et un écart type de 4,85 mg/L. Le coefficient de variation est relativement important (0,38), et la distribution présente une forte asymétrie (1) et un aplatissement significatif (2,55). Les dépassements occasionnels des 30 mg/L sont attribuables à des problèmes opérationnels (changement de gestion ou dysfonctionnements dans le dessableur).

Les concentrations d'azote total (NT) varient entre 14,8 et 65,1 mg/L, avec une moyenne de 33,88 mg/L et un écart type de 6,74 mg/L. Avec un coefficient de variation de 0,2, cette variabilité est considérée comme relativement importante, et la distribution présente une forte asymétrie (0,95) et un aplatissement notable (2,87), ce qui indique une performance modérée en matière d'élimination de l'azote.

Les orthophosphates (PO_4^{3-}) oscillent entre 1,26 et 5,2 mg/L, avec une moyenne de 3,46 mg/L et un écart type de 0,83 mg/L. Le coefficient de variation est de 0,24 indique une variabilité relativement importante, et la distribution affiche une asymétrie marquée (-0,34) et un aplatissement modéré (0,43), signalant un rendement d'élimination de seulement 50,07 %, ce qui reflète des difficultés dans le traitement du phosphore et une faible capacité de dénitrification au sein du bassin d'aération.

4. Paramètres spécifiques

L'indice de boues (IB) varie de 22,45 à 81,24 mL/g, avec une moyenne de 42,5 mL/g et un écart type de 7,92 mL/g. Le coefficient de variation est de 0,19. La distribution présente une asymétrie (0,51) et un aplatissement modéré (1,83), avec des valeurs généralement plus élevées en hiver, probablement en raison de variations de température et de charge organique. Une décantation trop rapide peut entraîner des sédiments indésirables dans les installations.

Le débit journalier des boues séchées (QBS) varie entre 4,47 et 738,58 m³/j, avec une moyenne de 82,13 m³/j et un écart type de 71,4 m³/j. Le coefficient de variation est élevé (0,87), et la distribution montre une forte asymétrie (5,47) et un aplatissement important (36,29), illustrant une variabilité considérable en fonction des débits d'entrée.

5.5 Critères d'évaluation des performances

La performance des réseaux de neurones artificiels est évaluée en recherchant une configuration optimale, notamment le nombre optimal de neurones dans la couche cachée. Une procédure d'essai et d'erreur est adoptée pour déterminer ce nombre. Les performances des RNA sont comparées en utilisant les valeurs d'erreur quadratique moyenne (EQM) et le coefficient de corrélation (R). Lorsqu'un coefficient de corrélation est maximal et que l'erreur quadratique moyenne est minimale, le modèle proposé est considéré comme le meilleur, et vice versa. Pour évaluer les performances des différents modèles utilisés, cinq critères d'évaluation ont été définis comme suit :

➤ Coefficient de corrélation

Le coefficient de corrélation (R) est l'indicateur le plus couramment utilisé pour évaluer la performance des modèles. Il mesure la relation entre les données prédites et observées, avec des valeurs comprises entre -1 et 1. Une valeur proche de 1 indique une forte corrélation positive, tandis qu'une valeur proche de -1 reflète une forte corrélation négative. Ce coefficient est influencé par les données maximales et est calculé selon l'équation suivante :

$$R = \frac{\sum_{i=1}^n (Y_{o,i} - \bar{Y}_o) \times (Y_{p,i} - \bar{Y}_p)}{\sqrt{\sum_{i=1}^n (Y_{o,i} - \bar{Y}_o)^2 \times \sum_{i=1}^n (Y_{p,i} - \bar{Y}_p)^2}} \quad (5.1)$$

➤ Coefficient d'Efficacité de Nash-Sutcliffe

Le coefficient d'efficacité de Nash-Sutcliffe (ENS) est largement utilisé pour évaluer la performance des modèles de simulation. Il varie entre $-\infty$ et 1 (Nash et Sutcliffe, 1970), où une valeur de 1 indique une correspondance parfaite entre les données observées et prédites. Un modèle est généralement considéré comme acceptable lorsque le ENS est supérieur à 0,8 (Garrick et al., 1978). Le coefficient d'efficacité est calculé selon la formule suivante :

$$ENS = 1 - \frac{\sum_{i=1}^n (Y_{p,i} - Y_{o,i})^2}{\sum_{i=1}^n (Y_{o,i} - \bar{Y}_o)^2} \quad (5.2)$$

➤ Racine de l'Erreur Quadratique Moyenne

La racine de l'erreur quadratique moyenne (REQM) est une mesure statistique qui évalue la performance des modèles prédictifs. Elle représente l'écart moyen entre les valeurs prédites et les valeurs observées. La REQM est exprimée dans les mêmes unités que les valeurs mesurées et est calculée selon l'équation suivante :

$$REQM = \sqrt{\frac{\sum_{i=1}^n (Y_{p,i} - Y_{o,i})^2}{n}} \quad (5.3)$$

➤ Erreur Moyenne Absolue

L'erreur moyenne absolue (EMA) mesure l'écart moyen absolu entre les valeurs prédites et observées. Plus l'EMA est faible, plus le modèle est précis. Elle est calculée selon la formule suivante :

$$EMA = \frac{1}{n} \sum_{i=1}^n |Y_{p,i} - Y_{o,i}| \quad (5.4)$$

➤ Coefficient de détermination

Le coefficient de détermination (R^2) représente la proportion de variance des données prédites. Il mesure le niveau de corrélation entre les données observées et prédites, avec des valeurs comprises entre 0 et 1. Une valeur proche de 1 indique une forte précision prédictive, tandis qu'une valeur proche de 0 reflète une mauvaise prédiction (Leggetts et McCabe, 1999). Il est défini par la formule suivante :

$$R^2 = \frac{\sum_{i=1}^n (Y_{o,i} - \bar{Y}_o)^2 - \sum_{i=1}^n (Y_{p,i} - \bar{Y}_p)^2}{\sum_{i=1}^n (Y_{o,i} - \bar{Y}_o)^2} \quad (5.5)$$

où: n: le nombre des valeurs observées;

$Y_{o,i}$ et $Y_{p,i}$: la valeur observée et prédite respectivement;

\bar{Y}_o et \bar{Y}_p : la moyenne des valeurs observées et prédites respectivement.

➤ Diagramme de Taylor

Le diagramme de Taylor est un outil utilisé pour représenter graphiquement, évaluer et comparer les performances des modèles (Taylor, 2001). Il permet une visualisation simultanée de plusieurs indicateurs de performance, tels que le coefficient de corrélation (R), l'écart-type (ET) et la racine de l'écart quadratique moyenne (REQM).

Lorsque les valeurs prédites correspondent étroitement aux valeurs observées, cela signifie une capacité prédictive en termes d'écart-type (ET), avec une corrélation élevée proche de 1 et une valeur faible de la racine de l'écart quadratique moyenne (RÉQM) proche de 0. La qualité du modèle diminue lorsque les valeurs de ces paramètres se déplacent vers des zones plus élevées sur le diagramme.

5.6 Construction des modèles

Les modèles à sortie unique (MSU) et les modèles à sortie multiple (MSM) sont largement utilisés pour l'analyse des procédés de traitement des eaux (Jafar et al., 2022). L'objectif principal de la détermination de la qualité de l'eau traitée est d'évaluer sa conformité avec la législation en vigueur et d'exploiter les résultats de modélisation pour analyser le fonctionnement des procédés.

Le tableau 5.2 présente les données utilisées dans les différents modèles.

Tableau 5.2: Paramètres choisis des différents modèles des différentes configurations

Configuration	Paramètres d'entrée	Paramètres de sortie
Modèle à sortie unique	Modèles simples (MS) : (pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DCO, MES, NT, PO ³⁻⁴)sortie, IB.	DBO ₅ sortie
	Modèles hybrides (MH): Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DCO, MES, NT, PO ³⁻⁴)sortie, IB]	
	MS : (pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , MES, NT, PO ³⁻⁴)sortie, IB.	DCO sortie
	MH : Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , MES, NT, PO ³⁻⁴)sortie, IB]	
	MS : (pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, NT, PO ³⁻⁴)sortie, IB.	MES sortie
	MH : Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, NT, PO ³⁻⁴)sortie, IB]	
	MS : (ppH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, MES, PO ³⁻⁴)sortie, IB.	NT sortie
	Modèles hybrides : Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, MES, PO ³⁻⁴)sortie, IB]	
	MS : (pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, MES, NT)sortie, IB.	PO ³⁻⁴ sortie
	MH : Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, MES, NT)sortie, IB]	
	MS : (pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, MES, NT, PO ³⁻⁴)sortie.	IB
	MH : Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, MES, NT, PO ³⁻⁴)sortie]	
	MS : (pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée. (DBO ₅ , DCO, MES, NT, PO ³⁻⁴)sortie.	QBS
	MH : Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, (DBO ₅ , DCO, MES, NT, PO ³⁻⁴)sortie]	
Modèle à sortie multiple	MS : (pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, IB.	(DBO ₅ , DCO, MES, NT, PO ³⁻⁴)sortie.
	MH : Décomposition [(pH, T, CE, DBO ₅ , DCO, MES, NT, PT)entrée, IB]	

5.6.1 Modélisation de la configuration des modèles à sortie unique

L'évaluation des paramètres de qualité de l'eau est cruciale pour la gestion et l'évaluation des performances des stations d'épuration des eaux usées.

- **Demande Biochimique en Oxygène**

La demande biochimique en oxygène après cinq jours (DBO_5) est un paramètre crucial pour la gestion des stations d'épuration. Elle mesure la quantité de matière organique biodégradable dans les eaux usées. Le test de la DBO, qui s'étale sur une période d'environ cinq jours, nécessite une préparation et une analyse méticuleuse (Rejsek, 2002).

Les résultats qualitatifs des modèles à RNA, simples et hybrides, pour la prédiction de la DBO_5 au cours des deux phases d'évaluation sont présentés dans le tableau 5.3.

Tableau 5.3: Evaluation des performances des modèles utilisées pour la prédiction de la DBO_5

Type de modèle		Entraînement				Test			
		REQM (mg/L)	EMA (mg/L)	R	ENS	REQM (mg/L)	EMA (mg/L)	R	ENS
Modèles simples	RN-PMC	1,0291	0,5080	0,8600	0,7289	1,2377	0,6194	0,8645	0,7336
	RN-FBR	0,7706	0,5828	0,9210	0,8480	0,8808	0,6860	0,9307	0,8651
	RN-RG	1,2989	0,8447	0,7678	0,5681	1,8187	1,3606	0,8068	0,4248
Modèles hybrides	TOD-RNPMC	0,9260	0,2397	0,9039	0,7805	0,7406	0,2212	0,9538	0,9046
	TOD-RNFBR	0,7746	0,5663	0,9209	0,8464	0,8397	0,5619	0,9546	0,8774
	TOD-RNRG	0,0415	0,0122	0,9998	0,9996	0,1055	0,0326	0,9990	0,9981
	DME-RNPMC	0,6174	0,1829	0,9506	0,9024	0,7648	0,2285	0,9519	0,8983
	DME-RNFBR	0,8437	0,6392	0,9056	0,8178	0,8754	0,5990	0,9518	0,8667
	DME-RNRG	0,0652	0,0214	0,9995	0,9989	0,0896	0,0292	0,9993	0,9986
	DMV-RNPMC	0,8068	0,2146	0,9143	0,8334	1,0073	0,3711	0,9102	0,8236
	DMV-RNFBR	0,8885	0,6872	0,8951	0,7979	0,9153	0,6425	0,9490	0,8543
	DMV-RNRG	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000

Lors de la phase d'entraînement, les résultats de modélisation par les modèles RNA simples révèlent que le modèle RN-FBR présente des performances élevées avec une valeur de R égal à 0,9210 et une valeur d'ENS égal à 0,8480, et une erreur plus faible avec une valeur de REQM égal à 0,7706 mg/L. Lors de la phase de test, le modèle RN-FBR continue afficher des performances élevées avec une valeur de R égal à 0,9307 et une valeur d'ENS égal à 0,8651, et l'erreur plus faible avec une valeur de REQM égal à 0,8808 mg/L, par rapport aux modèles RN-PMC et RN-RG.

Pendant la phase d'entraînement, les modèles hybrides TOD-RNA affichent que le modèle TOD-RNFBR avec de meilleures valeurs de performances de $R = 0,9209$ et d'ENS = 0,8464, ainsi que de plus faibles valeurs d'erreur du REQM = 0,7746 mg/L. Par contre au cours de la phase de test, le modèle TOD-RNPMC le meilleur avec de plus faibles valeurs d'erreurs du REQM de 0,7406 mg/L et du EMA de 0,2212 mg/L, et valeur de performance élevée d'ENS = 0,9046, et un $R = 0,9538$ similaire que le $R = 0,9546$ du modèle TOD-RNFBR.

L'étude analytique des modèles DME-RNA au cours de la phase d'entraînement, le modèle DME-RNPMC son efficacité avec des meilleurs valeurs des performances R estimées à 0,9506 et ENS estimé à 0,9024, et des faibles erreurs REQM estimées à 0,6174 mg/L et EMA estimé à 0,1829 mg/L. Lors de la phase du test, le modèle DME-RNPMC confirme sa supériorité avec des valeurs des performances élevées R estimées à 0,9519 et ENS estimés à 0,8983, et avec de faibles valeurs d'erreurs de REQM estimé à 0,7648 mg/L et EMA estimé à 0,2285 mg/L.

L'entraînement des modèles DMV-RNA, a révélé que le modèle DMV-RNPMC avec des performances supérieures avec une valeur du $R = 0,9143$ et une valeur de l'ENS = 0,8334, et des erreurs plus faibles avec une valeur du REQM = 0,8068 mg/L et une valeur du EMA = 0,2146 mg/L. Par contre lors de la phase de test, le modèle DMV-RNFBR a surpassé avec des valeurs de performances de $R = 0,9490$ et d'ENS = 0,8543, et d'erreur plus faible avec une valeur de REQM = 0,9153 mg/L.

Dans le contexte de tous les modèles hybrides combinant la régression générale utilisés dans cette étude, il est important de souligner que les résultats obtenus ne doivent pas être pris en considération ou interprétés automatiquement comme un cas de sur-apprentissage.

Les performances qualitatives des modèles RNA simples et hybrides tout au long des étapes d'entraînement et de test sont présentées dans la figure 5.1, cette figure confirme les résultats obtenus précédemment que le modèle RN-FBR meilleure avec une valeur de coefficient de détermination (R^2) égal à 0,8683. Cela avec une valeur élevée de $R^2 = 0,8611$ du modèle TOD-RNFBR graphiquement confirme aussi la supériorité de ce modèle concernant les modèles du TOD-RNA. L'analyse graphique des modèles DME-RNA confirme aussi la supériorité du modèle DME-RNPMC avec une valeur élevée de R^2 estimé à 0,9136. Les deux modèles DMV-RNPMC et DMV-RNFBR présentent des valeurs de R^2 semblable = 0,8507, et de $R^2 = 0,833$ respectivement.

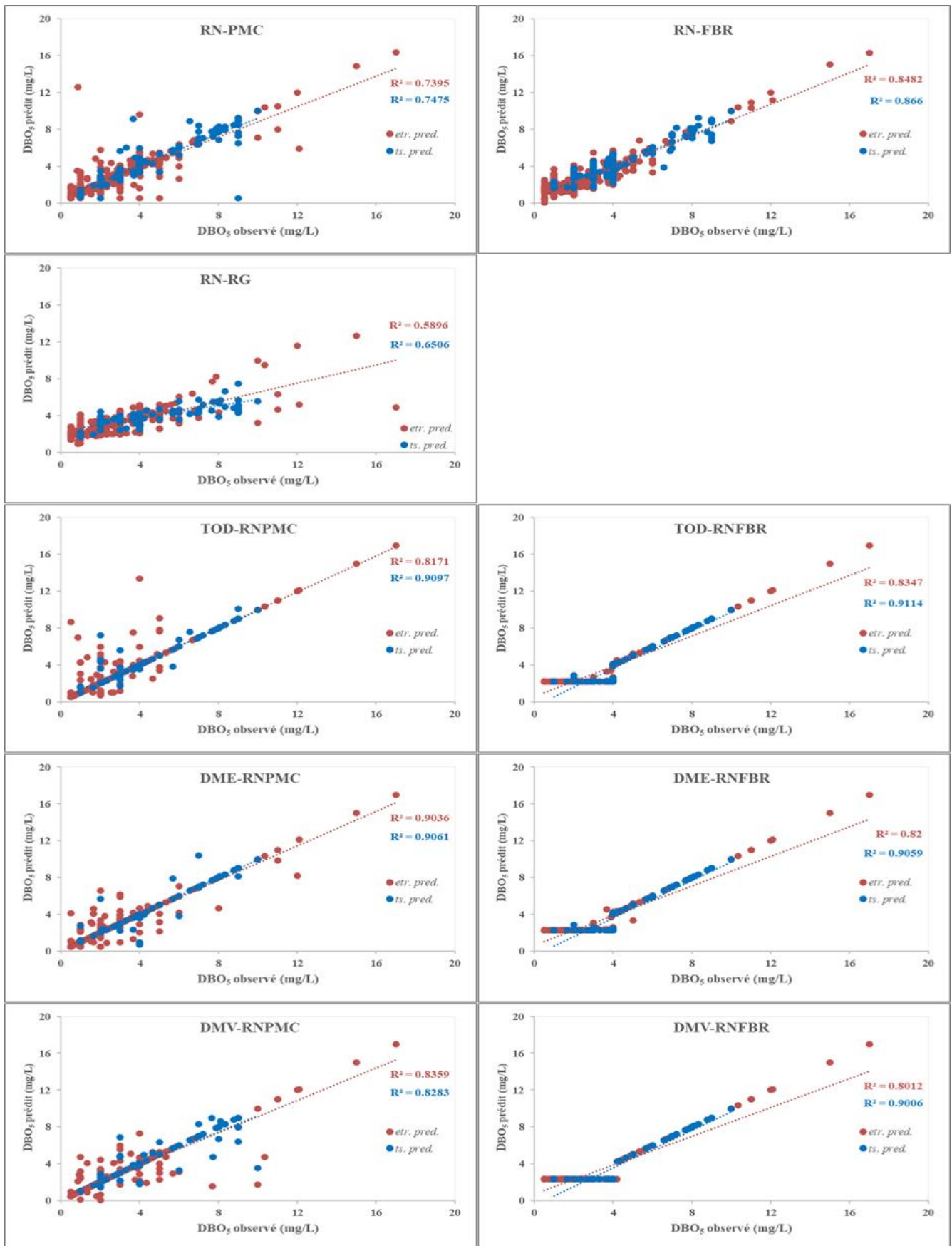


Figure 5.1: Nuage de points des modèles utilisés pour la prédiction de la DBO_5

Il est un peu difficile de faire converger et diverger les résultats dans les deux phases et de choisir le meilleur modèle parmi de différents modèles utilisés, pour valider notre choix du meilleur modèle précédemment, ont utilisé la représentation graphique du diagramme de Taylor, selon la figure 5.2, le modèle DME-RNPMC plus adéquat que celui des modèles pour la prédiction de DBO_5 avec une valeur d'écart type égale à 2,18 mg/L très proche de la valeur d'écart type observée de 2,19 mg/L, ce qui indique une sous-estimation. La corrélation plus élevée R égale à 0,96 démontre un fort accord entre les valeurs observées et prédites, la valeur de $REQM$ plus faible et proche de 1 mg/L.

Ces résultats montrent que le modèle hybride (DME-RNPMC) permet une réduction des erreurs $REQM$ et EMA de 19,88 % et 13,17 %, respectivement. De plus, il augmente les valeurs R et ENS de 3,11 % et 6,03 %, respectivement, lors de l'entraînement. En phase de test, les améliorations sont de 13,17 % pour le $REQM$, 66,69 % pour l' EMA , 2,23 % pour R et 3,7 % pour ENS , comparé au modèle simple (RN-FBR) avec une architecture optimale de TP-97-10-1.

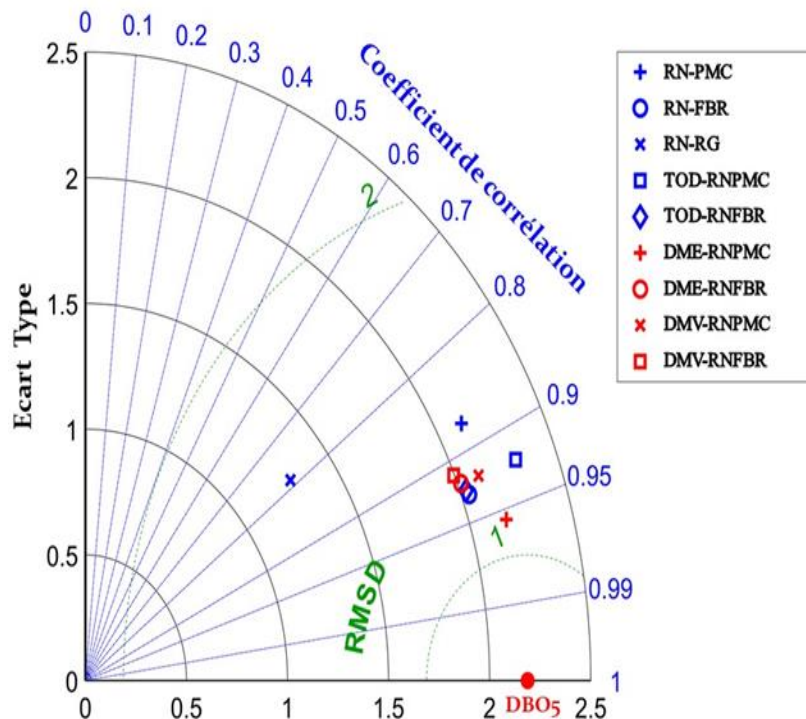


Figure 5.2: Diagramme de Taylor des différents modèles utilisés pour la prédiction de la DBO_5

- **Demande Chimique en Oxygène**

La demande chimique en oxygène (DCO) représente la quantité d'oxygène nécessaire pour oxyder la matière organique en dioxyde de carbone (CO_2) et en eau (H_2O). Elle constitue un indicateur clé pour évaluer la quantité totale de la matière organique des eaux usées, qu'elle soit biodégradable ou non biodégradable (Rodier, 2009). Les performances qualitatives des modèles simples et hybrides appliqués à la prédiction de la DCO sont présentées dans le tableau 5.4.

Tableau 5.4: Evaluation des performances des modèles utilisées pour la prédiction de la DCO

Type de modèle		Entraînement				Test			
		REQM (mg/L)	EMA (mg/L)	R	ENS	REQM (mg/L)	EMA (mg/L)	R	ENS
Modèles simples	RN-PMC	3,7021	2,2709	0,9305	0,8652	2,5615	1,8847	0,8236	0,6294
	RN-FBR	3,3029	2,4644	0,9449	0,8927	2,9187	2,1154	0,7468	0,5189
	RN-RG	5,6110	4,0333	0,8551	0,6904	3,5254	2,8919	0,6206	0,2980
Modèles hybrides	TOD-RNPMC	2,6725	0,7533	0,9649	0,9298	2,3464	0,8820	0,8587	0,6890
	TOD-RNFBR	3,3250	2,2967	0,9474	0,8913	4,4990	3,5464	0,6393	-0,1432
	TOD-RNRG	0,2236	0,0777	0,9998	0,9995	0,1078	0,0509	0,9997	0,9993
	DME-RNPMC	4,0311	0,9494	0,9179	0,8402	1,4582	0,3849	0,9409	0,8799
	DME-RNFBR	3,3473	2,3100	0,9465	0,8898	4,4657	3,5180	0,6427	-0,1264
	DME-RNRG	0,2213	0,1003	0,9998	0,9995	0,1560	0,0698	0,9993	0,9986
	DMV-RNPMC	3,3941	0,8076	0,9420	0,8867	2,2626	0,8239	0,8733	0,7109
	DMV-RNFBR	3,3139	2,2900	0,9479	0,8920	4,4929	3,5012	0,6531	-0,1401
	DMV-RNRG	0,0026	0,0006	1,0000	1,0000	0,0010	0,0001	1,0000	1,0000

Pendant la phase d'entraînement des modèles RNA simples, le modèle RN-FBR a présenté meilleures valeurs de performances R de 0,9449 et ENS de 0,8927, et d'erreur plus faible d'une valeur du REQM égal à 3,3029 mg/L. Par contre, durant la phase de test, le modèle RN-PMC s'affiche des valeurs de performances élevés de R égal à 0,8236 et d'ENS égal à 0,6294, et des valeurs plus faibles d'erreurs de REQM égal à 2,5615 mg/L et EMA égal à 1,8847 mg/L, par rapport aux deux autres modèles.

La supériorité du modèle TOD-RNPMC est confirmée avec des valeurs de performances du coefficient de corrélation R estimé à 0,9649 et d'ENS estimé à 0,9298, et des erreurs plus faibles avec une valeur de REQM estimé à 2,6725 mg/L et une valeur d'EMA estimé à 0,7533 mg/L durant la phase d'entraînement. Pendant la phase de test, la tendance est confirmée puisque le modèle TOD-RNPMC continue d'afficher les meilleures performances avec une valeur de R estimé à 0,8587 et une valeur d'ENS estimé à 0,6890, et des erreurs très faibles avec une valeur de REQM estimé à 2,3464 mg/L et une valeur d'EMA estimé à 0,8820 mg/L par rapport aux autres modèles du TOD-RNA.

Les modèles DME-RNA présentent que pendant la phase d'entraînement, le modèle DME-RNFBR est le plus performant avec des performances élevées avec R de 0,9465 et ENS de 0,8898, et une valeur d'erreur plus faible avec REQM de 3,3473 mg/L. Par contre, en phase de test, les paramètres de performance avec une valeur de R égale à 0,9409 et ENS égale à 0,8799 et des erreurs très faibles avec une valeur de REQM égale à 1,4582 mg/L et EMA égale à 0,3849 mg/L ont montré que le modèle DME-RNPMC est le meilleur.

Les modèles DMV-RNA illustrent que pendant la phase d'entraînement, le modèle DMV-RNFBR est le meilleur grâce aux valeurs de performances de R égale à 0,9479 et d'ENS égale à 0,8920, et à une valeur d'erreur très faible de REQM égale à 3,3139 mg/L. En revanche, lors de la phase de test, les paramètres de performance de R égale à 0,8733 et d'ENS égal à 0,7109, et les erreurs très faibles avec une valeur de REQM égale à 2,2626 mg/L et une valeur de EMA égale à 0,8239 mg/L, suggérant que le modèle DMV-RNPMC se mieux.

La figure 5.3 illustre graphiquement la supériorité du modèle RN-FBR avec le meilleur coefficient de détermination de 0,8772 comparée à celle du modèle semblable RN-PMC. La représentation graphique confirme aussi que le modèle TOD-RNPMC est le choix adéquat avec un R^2 élevé d'une valeur de 0,9218. La représentation graphique montre que le modèle DME-RNPMC est le plus efficace, avec un R^2 élevé de 0,8439, similaire au modèle DME-RNFBR ($R^2 = 0,8476$) et la supériorité du premier modèle durant la phase de test avec une valeur du R^2 égale à 0,8852. La différence des résultats obtenus des modèles DMV-RNA dans les deux phases. Selon la représentation graphique, nous pouvons observer que le modèle DMV-RNPMC a montré une efficacité supérieure avec une valeur de R^2 de 0,8809.

Selon la figure 5.4, le diagramme de Taylor représente que le modèle TOD-RNPMC c'est le modèle le plus performant par rapport aux autres modèles étudiés pour la prédiction de DCO avec une valeur de corrélation élevée de 0,96, et une valeur d'ET égale à 9,26 mg/L, et supérieure à la valeur d'ET observée de 9,22 mg/L, cela indique que ce modèle a tendance à surestimer, la valeur de RÉQM est minimale, près de 2 mg/L.

Le modèle hybride TOD-RNPMC améliore les performances par rapport au modèle simple (RN-FBR) avec une augmentation de R et ENS de 2,07 % et 3,99 %, respectivement, en phase d'entraînement, une réduction des erreurs REQM de 19,09 % et EMA de 69,43 %. Les gains sont encore plus importants en phase de test (R = 13,03 %, ENS = 24,69 %, REQM = 19,61 %, EMA = 58,31 %) avec une architecture optimale de TP-65-10-1.

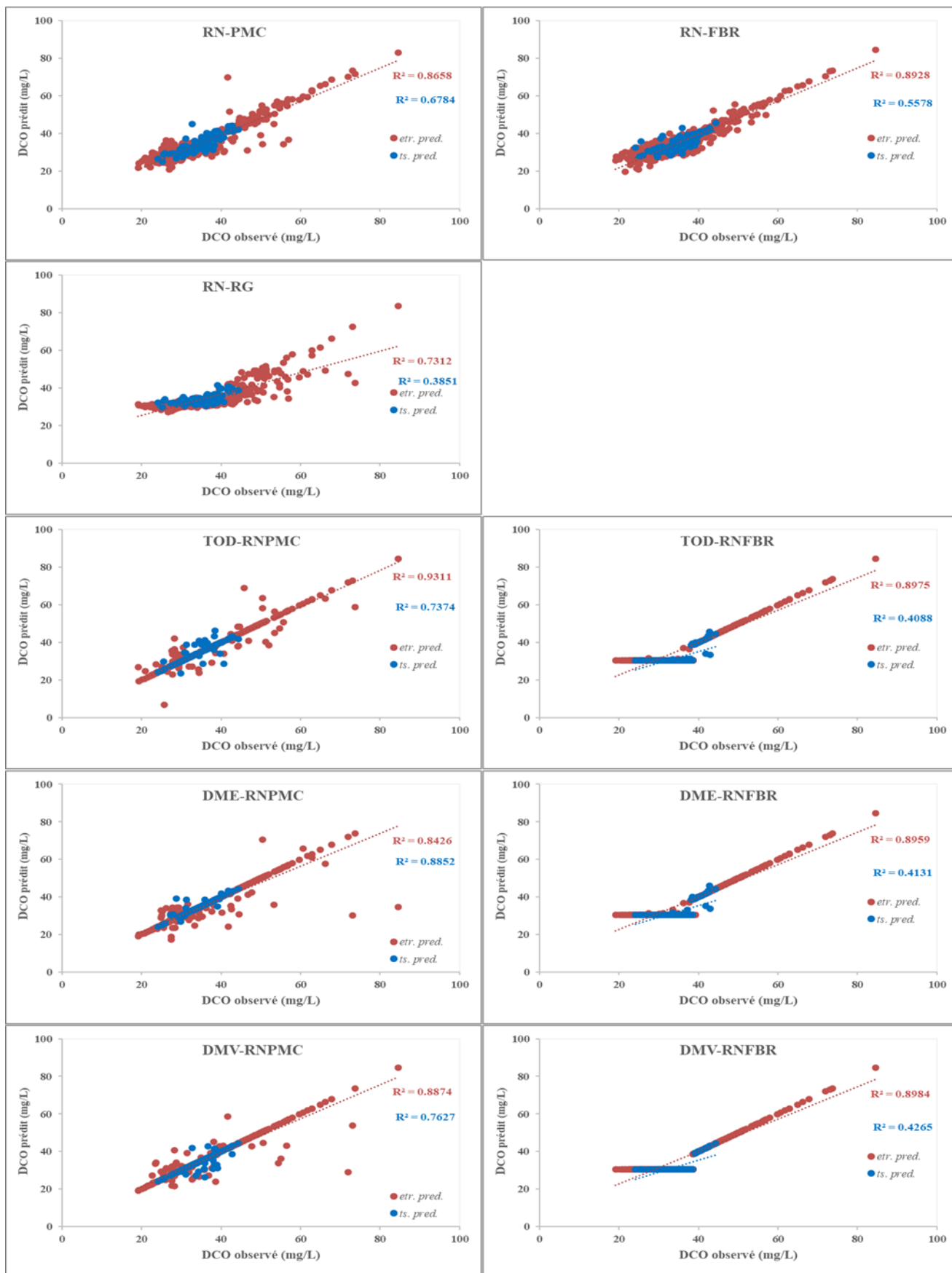


Figure 5.3: Nuage de points des modèles utilisés pour la prédiction de la DCO

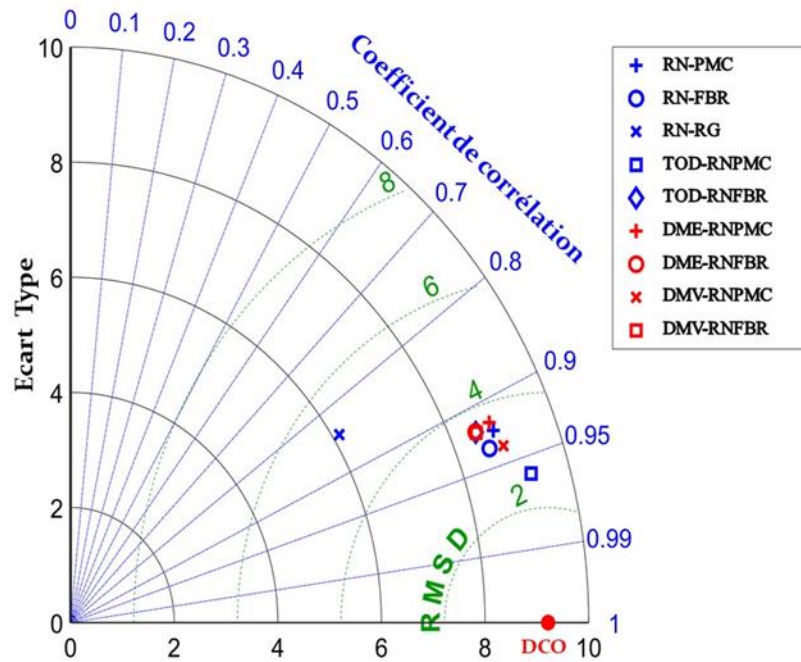


Figure 5.4: Diagramme de Taylor des différents modèles utilisés pour la prédiction de la DCO

- **Matières en Suspension**

Les matières en suspension (MES) représentent les particules solides non dissoutes présentes dans l'eau décantable et non décantable. La modélisation de la concentration totale de MES dans les effluents des stations d'épuration a montré des performances satisfaisantes.

Les résultats qualitatifs obtenus à l'aide de modèles simples et hybrides au cours des phases d'entraînement et de test sont regroupés dans le tableau 5.5.

Tableau 5.5: Evaluation des performances des modèles utilisés pour la prédiction de MES

Type de modèle		Entraînement				Test			
		REQM (mg/L)	EMA (mg/L)	R	ENS	REQM (mg/L)	EMA (mg/L)	R	ENS
Modèles simples	RN-PMC	2,6955	1,5144	0,8567	0,7045	3,4742	1,4962	0,7228	0,3722
	RN-FBR	2,8520	2,0580	0,8180	0,6691	2,6374	1,9149	0,7995	0,6382
	RN-RG	3,6314	2,6023	0,7045	0,4636	3,3551	2,5223	0,7347	0,4145
Modèles hybrides	TOD-RNPMC	4,0622	2,9910	0,5737	0,3288	3,2897	2,4070	0,6697	0,4371
	TOD-RNFBR	2,5162	1,7612	0,8618	0,7425	2,1653	1,6122	0,8716	0,7561
	TOD-RNRG	0,1060	0,0437	0,9998	0,9995	0,0880	0,0392	0,9998	0,9996
	DME-RNPMC	3,2411	0,9146	0,8275	0,5727	2,0015	0,5634	0,9043	0,7916
	DME-RNFBR	2,5144	1,7617	0,8620	0,7428	2,2116	1,6542	0,8667	0,7456
	DME-RNRG	0,2525	0,1219	0,9988	0,9974	0,1453	0,0641	0,9995	0,9989
	DMV-RNPMC	4,1385	3,0368	0,5553	0,3033	3,5245	2,7303	0,5967	0,3539
	DMV-RNFBR	2,5678	1,8200	0,8556	0,7318	2,2975	1,7108	0,8550	0,7255
	DMV-RNRG	0,0021	0,0003	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000

Les résultats des modèles RNA simples montrent que l'efficacité du modèle RN-PMC avec des erreurs plus faibles avec une valeur de REQM égale à 2,6955 mg/L et une valeur de EMA égale à 1,5144 mg/L, et des valeurs élevées de performances de R égale à 0,8567 et ENS égal à 0,7045 lors de la phase d'entraînement. Par contre, lors de la phase de test, le modèle RN-FBR s'illustre avec de meilleures valeurs de performances de R égal à 0,7995 et d'ENS égal à 0,6382, et avec une valeur d'erreur plus faible de REQM égale à 2,6374 mg/L, c'est le meilleur.

L'étude des performances des modèles TOD-RNA durant la phase d'entraînement, indique que le modèle TOD-RNFBR présente des valeurs des paramètres de performances élevées avec un coefficient de corrélation R estimé à 0,8618 et d'ENS estimé à 0,7425, et des valeurs d'erreurs très faibles de REQM estimé à 2,5162 mg/L et d'EMA estimé à 1,7612 mg/L. Lors de la phase de test, les meilleurs paramètres de performances avec une valeur de R estimé à 0,8716 et une valeur d'ENS estimé à 0,7561, et des valeurs d'erreurs très faibles de REQM estimé à 2,1653 mg/L, et d'EMA estimé à 1,6122 mg/L continus montrer que le modèle TOD-RNFBR est le meilleur.

Les résultats des modèles DME-RNA pendant la phase d'entraînement ont montré que l'efficacité du modèle DME-RNFBR avec meilleures performances de R d'une valeur de 0,8620 et d'ENS d'une valeur de 0,7428, et une erreur très faible de REQM d'une valeur de 2,5144 mg/L. Par contre, pendant la phase de test, les paramètres de performance avec une valeur de R égale à 0,9043 et de ENS égal à 0,7916, et des valeurs d'erreurs très faibles de REQM égale à 2,0015 mg/L et d'EMA égal à 0,5634 mg/L ont indiqué que le modèle DME-RNPMC était meilleur.

L'analyse des performances des modèles DMV-RNA a révélé que durant la phase d'entraînement, la supériorité du modèle DMV-RNFBR avec une valeur de R égale à 0,8556 et une valeur d'ENS égal à 0,7318, et des erreurs très faibles avec une valeur de REQM égale à 2,5678 mg/L et avec un valeur d'EMA égal à 1,8200 mg/L, aussi lors de la phase de test, les valeurs des paramètres de performance de R égale à 0,8550 et d'ENS égal à 0,7255, et les valeurs des erreurs très faibles de REQM égale à 2,2975 mg/L et d'EMA égal à 1,7108 mg/L ont que le modèle DMV-RNFBR affiche meilleure performance.

L'analyse des performances des modèles RNA simples vous n'avez pas choisi le meilleur modèle, étant différent à chaque phase. Selon la figure 5.5, la représentation graphique illustre la performance du modèle RN-FBR avec la meilleure valeur du coefficient de détermination R^2 égale à 0,6393 lors de la phase de test, comparée à celle de RN-PMC et RN-RG. La représentation graphique des modèles TOD-RNA, aussi continue, choisit le modèle TOD-RNFBR est le meilleur parce que la valeur de R^2 est élevée (0,7451) par rapport aux autres modèles. Les résultats obtenus des modèles DME-RNA sont différents dans les deux phases. D'après la représentation graphique, il est possible de constater que le modèle DME-RNFBR a

démontré une efficacité accrue avec une valeur de R^2 de 0,7437. La représentation graphique des modèles DMV-RNA illustre que le modèle DMV-RNFBR est le plus performant avec une valeur R^2 de 0,7311.

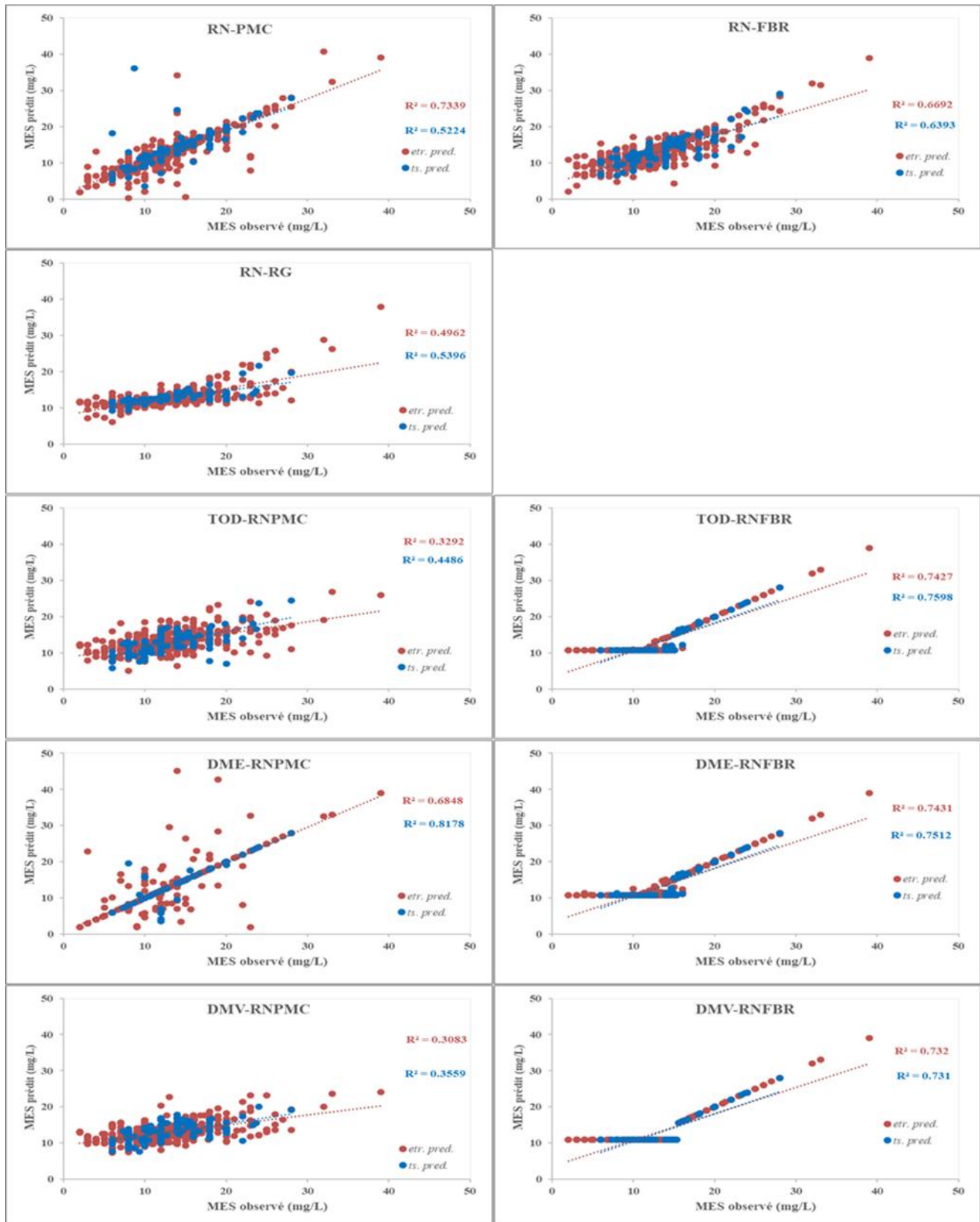


Figure 5.5: Nuage de points des modèles utilisés pour la prédiction du MES

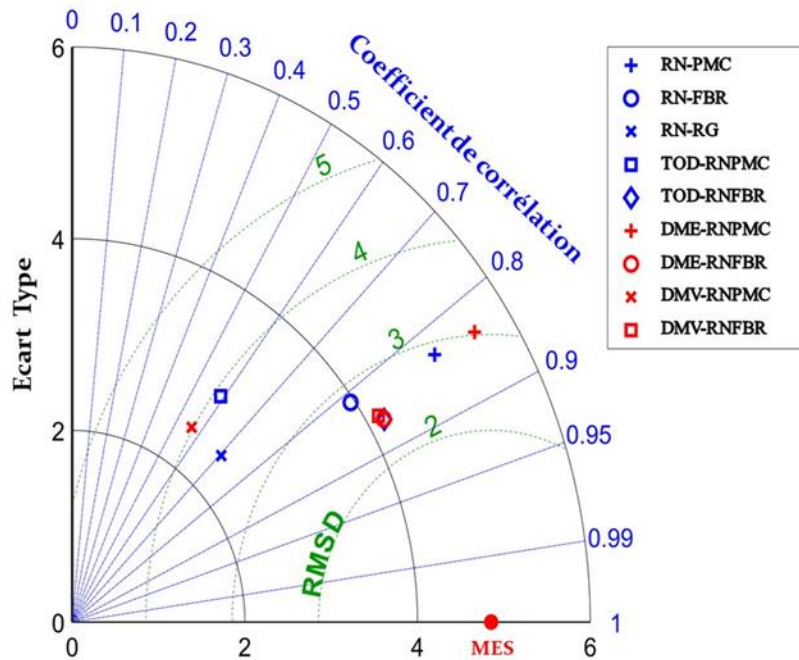


Figure 5.6: Diagramme de Taylor des différents modèles utilisés pour la prédiction de MES

Pour trouver le meilleur modèle pour la prédiction de MES, le diagramme de Taylor (Figure 5.6), présente que les trois modèles TOD-RNFBR, DME-RNFBR, et DMV-RNFBR est plus adéquat que celui des autres modèles à cause de même valeur de corrélation est élevée (0,86), et une valeur d'ET 4,19 mg/L, 4,18 mg/L et 4,15 mg/L respectivement ce qui est une augmentation par rapport aux autres modèles et une diminution par rapport au l'ET observé (4,85 mg/L) cela indique une sous-estimation, et ces modèles affichent des valeurs de RÉQM minimale, d'environ 2 mg/L.

On remarque que le modèle hybride TOD-RNFBR est le meilleur, avec une architecture optimale de 65-130-1. Il améliore les performances avec une augmentation de R et ENS de 2,07 % et 3,99 %, respectivement, une réduction de REQM de 11,77 % et une réduction d'EMA de 14,42 % en phase d'entraînement ; les gains sont encore plus importants durant la phase de test (R = 8,27 %, ENS = 15,59 %, REQM = 17,9 %, EMA = 15,81 %) par rapport au modèle simple RN-FBR.

- **Azote Total**

L'azote total (NT) regroupe les différentes formes d'azote présentes dans l'eau, qu'elles soient organiques ou minérales, comme l'ammonium (NH_4^+), les nitrites (NO_2^-) et les nitrates (NO_3^-). Cet indicateur est essentiel pour évaluer la quantité de nutriments dans les eaux usées (Rodier, 2009).

Les résultats qualitatifs des modèles RNA simples et hybrides appliqués à ce paramètre sont présentés dans le tableau 5.6.

Tableau 5.6: Evaluation des performances des modèles utilisés pour la prédiction du l'azote totale

Type de modèle		Entraînement				Test			
		REQM (mg/L)	EMA (mg/L)	R	ENS	REQM (mg/L)	EMA (mg/L)	R	ENS
Modèles simples	RN-PMC	4,2575	2,0299	0,8271	0,6149	2,9801	1,6189	0,8999	0,7724
	RN-FBR	3,3227	2,3916	0,8749	0,7655	3,2487	2,3385	0,8549	0,7295
	RN-RG	4,9134	3,5019	0,7328	0,4871	4,6280	3,1542	0,8005	0,4510
Modèles hybrides	TOD-RNPMC	4,9829	3,7813	0,6884	0,4725	5,6482	3,9002	0,4301	0,1823
	TOD-RNFBR	3,6689	2,5377	0,8450	0,7140	2,8876	1,4685	0,8871	0,7863
	TOD-RNRG	0,1460	0,0457	0,9998	0,9995	0,2354	0,0704	0,9993	0,9986
	DME-RNPMC	2,8401	0,7757	0,9207	0,8286	4,2684	0,8396	0,7759	0,5330
	DME-RNFBR	3,7473	2,6200	0,8377	0,7017	3,2714	1,7042	0,8527	0,7257
	DME-RNRG	0,2413	0,0829	0,9994	0,9988	0,2841	0,0915	0,9990	0,9979
	DMV-RNPMC	4,8749	3,7067	0,7075	0,4951	5,7042	4,0392	0,4153	0,1660
	DMV-RNFBR	3,7325	2,5973	0,8391	0,7040	3,0414	1,5501	0,8742	0,7629
	DMV-RNRG	0,0027	0,0005	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000

Pour la phase d'entraînement des modèles de RNA simples, le modèle RN-FBR s'affiche comme le meilleur modèle par rapport aux deux autres modèles, avec une valeur d'erreur faible estimée à REQM estimé à 3,3227 mg/L. Concernant les paramètres de performance, une fois de plus le modèle RN-FBR s'affiche avec des valeurs de paramètres de performances élevées avec un coefficient de corrélation d'une valeur estimée à 0,8749 et un coefficient d'efficacité de Nash-Sutcliffe d'une valeur estimée à 0,7655 supérieur aux valeurs des deux autres modèles. Par contre, lors de la phase de test, le modèle RN-PMC décroît considérablement avec des erreurs faibles d'une valeur de REQM égale à 2,9801 mg/L et d'une valeur d'EMA égale à 1,6189 mg/L. Le modèle RN-PMC confirme aussi cette évolution avec les meilleures valeurs des paramètres de performance de R = 0,8999 et d'ENS égal à 0,7724.

Le modèle TOD-RNFBR affiche les meilleures caractéristiques, avec des meilleures les plus faibles erreurs par rapport au modèle TOD-RNPMC, la valeur du REQM et EMA sont respectivement 3,6689 mg/L et 2,5377 mg/L. Les performances du modèle TOD-RNFBR surpassent celle de modèle TOD-RNPMC et affichent des niveaux élevés avec un coefficient de corrélation estimé à 0,8450 et un coefficient d'efficacité de Nash-Sutcliffe d'une valeur de 0,7140 à la phase d'entraînement. Lors de la phase de test, les paramètres de performances élevées, d'une valeur de R égale à 0,8871 et d'une valeur d'ENS égale à 0,7863, et des valeurs les plus faibles d'erreurs de REQM égale à 2,8876 mg/L et EMA égal à 1,4685 mg/L ont montré que le modèle TOD-RNFBR était supérieur.

L'étude analytique des modèles DME-RNA pendant la phase d'entraînement, le modèle DME-RNPMC affiche des paramètres de performance plus élevés avec une valeur de R égale à 0,9207 et une valeur d'ENS égale à 0,8286, et des erreurs très faibles avec une valeur de REQM égale à 2,8401 mg/L et une valeur d'EMA égale à 0,7757 mg/L. Par contre, durant la phase de test, le modèle DME-RNFBR surpasse le modèle DME-RNPMC avec des paramètres de performance élevés de R égale à 0,8527 et d'ENS égal à 0,7257, et avec une erreur très faible de REQM égale à 3,2714 mg/L.

L'analyse de performances des modèles DMV-RNA a illustré lors de la phase d'entraînement, les paramètres de performance du modèle DMV-RNFBR enregistrent des valeurs très acceptables avec un R égal à 0,8391 et un ENS égal à 0,7040 et des valeurs d'erreurs acceptables, avec un REQM de valeur égale 3,7325 mg/L, et une EMA de valeur égale 2,5973 mg/L. Lors du process de test, le modèle DMV-RNFBR aussi présente les meilleures performances avec un coefficient de R = 0,8742 et un coefficient d'ENS = 0,7629, associés à des faibles valeurs d'erreurs REQM = 3,0414 mg/L et EMA = 1,5501 mg/L.

L'analyse des modèles RNA simples n'a pas permis de sélectionner le modèle le plus approprié, car les résultats obtenus diffèrent aux deux étapes. La représentation graphique (Figure 5.7) montre que le modèle RN-FBR a mieux à cause d'une valeur élevée du coefficient de détermination R^2 estimé à 0,7593.

Concernant les modèles TOD-RNA obtenus déclarent que le modèle TOD-RNFBR est le meilleur avec les deux phases ; c'est ce que confirme la représentation graphique que ce modèle est plus performant avec une valeur R^2 de 0,7264 élevée.

Les résultats obtenus à partir de l'étude analytique des modèles DME-RNA varient pendant les deux phases. Selon la représentation graphique, illustre que le modèle DME-RNPMC est supérieur par rapport au modèle DME-RNFBR est affiche une valeur élevée de R^2 de 0,8033.

Enfin, la représentation graphique confirme que les résultats obtenus à partir de l'analyse numérique des modèles DMV-RNA, le modèle DMV-RNFBR indiquent une performance supérieure avec une valeur R^2 de 0,7141 élevée.

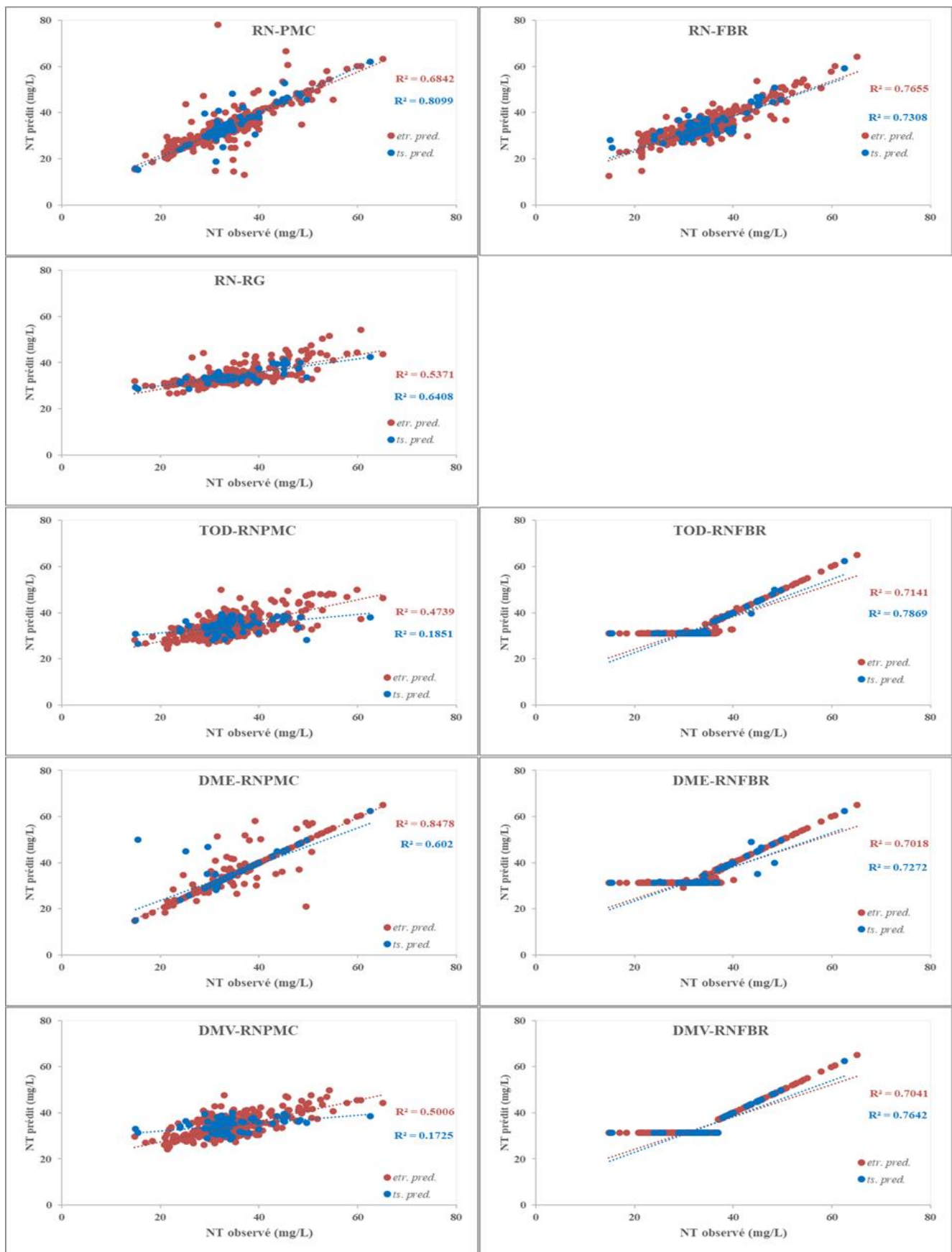


Figure 5.7: Nuage de points des modèles utilisés pour la prédiction du NT

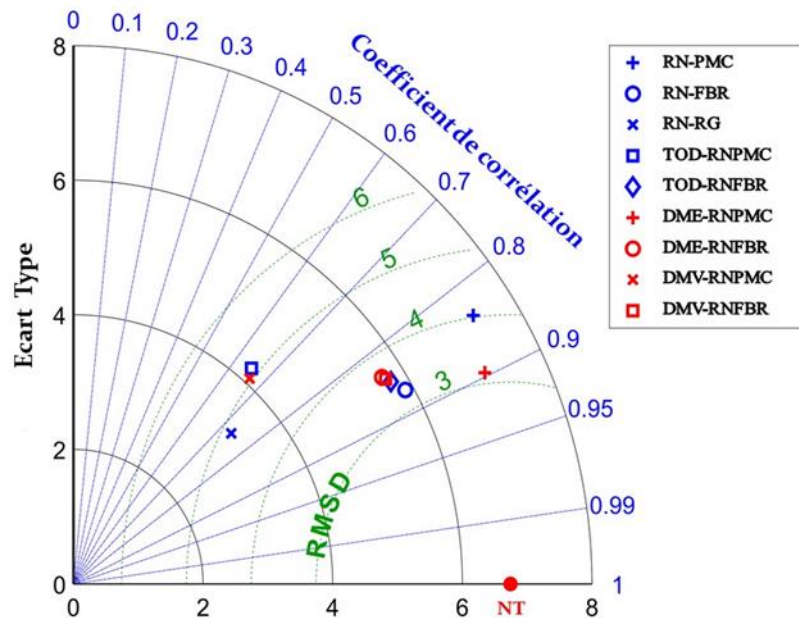


Figure 5.8: Diagramme de Taylor des différents modèles utilisés pour la prédiction du NT

Figure 5.8 illustre graphiquement la performance du modèle RN-FBR avec une valeur de corrélation égale à 0,84 et une valeur de l'ET égale à 7,35 mg/L, ce qui est inférieur à la valeur de l'ET observée de 6,74 mg/L. Il est évident que ce modèle surestime le NT, et avec une valeur de RÉQM minimale, d'environ 3 mg/L.

Le modèle simple RN-FBR améliore les performances avec une augmentation de R et ENS de 3,54 % et 7,21 %, respectivement, une réduction de REQM de 10,42 % et d'EMA de 6,11 % en phase d'entraînement. Cependant, durant la phase de test, le modèle hybride TOD-RNFBR améliore les performances avec une augmentation de R de 3,63 % et d'ENS de 7,22 %, et une réduction de REQM de 11,12 % et d'EMA de 37,21 %. Grâce à ces différences durant les deux phases, nous avons choisi le modèle simple en raison du nombre réduit de paramètres, nécessitant moins de données pour leur entraînement, ce qui limite le risque de sur-apprentissage. De plus, leur faible complexité accélère les calculs et les rend plus accessibles pour des applications avec des contraintes de ressources avec une architecture optimale de 13-130-1.

- **Orthophosphates**

Le phosphore total (PT), principalement issu des détergents et des excréta humains, est un facteur majeur d'eutrophisation. Les orthophosphates (PO_4^{3-}), principale forme de phosphates dans l'eau, proviennent de l'oxydation minérale du phosphore. Ils peuvent se présenter sous forme dissoute, solide ou colloïdale (Rejsek, 2002). Les résultats qualitatifs obtenus à partir des modèles RNA simples et combinés pour la modélisation de ce paramètre sont présentés dans le tableau 5.7.

Tableau 5.7: Evaluation des performances des modèles pour la prédiction de PO^{-3}_4

Type de modèle		Entraînement				Test			
		REQM (mg/L)	EMA (mg/L)	R	ENS	REQM (mg/L)	EMA (mg/L)	R	ENS
Modèles simples	RN-PMC	2,8507	0,7969	0,6506	0,1082	1,4208	0,6239	0,4208	-2,8262
	RN-FBR	0,7491	0,5641	0,9689	0,9384	0,8085	0,5805	0,6564	-0,2390
	RN-RG	2,6707	0,9577	0,5333	0,2173	0,7101	0,5840	0,7826	0,0442
Modèles hybrides	TOD-RNPMC	2,6679	1,1291	0,4749	0,2189	1,1279	0,9277	0,3134	-1,4113
	TOD-RNFBR	0,6323	0,4567	0,9779	0,9561	0,4594	0,3381	0,7939	0,6000
	TOD-RNRG	0,1159	0,0165	0,9994	0,9985	0,0076	0,0024	0,9999	0,9999
	DME-RNPMC	1,2201	0,2171	0,9157	0,8367	0,3501	0,0866	0,8967	0,7677
	DME-RNFBR	0,6684	0,4717	0,9752	0,9510	0,4703	0,3442	0,7827	0,5808
	DME-RNRG	0,2938	0,0324	0,9957	0,9905	0,0051	0,0023	1,0000	0,9999
	DMV-RNPMC	1,1266	0,2071	0,9368	0,8607	0,5174	0,1790	0,8000	0,4925
	DMV-RNFBR	0,6297	0,4486	0,9781	0,9565	0,4662	0,3459	0,7860	0,5881
	DMV-RNRG	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000

Les trois modèles affichent des performances médiocres, la détérioration des performances des modèles RNA simples durant les deux phases indique la difficulté des modèles proposées à généraliser et prédirent des valeurs adéquate, cela indique que tous les modèles sont de très mauvaise qualité.

Pour les modèles TOD-RNA, il a été observé que durant la phase d'entraînement, le modèle TOD-RNFBR se présente comme étant le meilleur modèle avec des performance élevée, une valeur de R estimé à 0,9779 et une valeur d'ENS estimé à 0,9561, et des erreurs très faible estimé à une valeur de REQM estimé à 0,6323 mg/L et une valeur d'EMA estimé à 0,4567 mg/L. Lors de la phase de test, le modèle TOD-RNFBR continue d'affiche les meilleures performances avec une valeur de R estimé à 0,7939 et une valeur d'ENS estimé à 0,6000, et les plus faibles erreurs avec une valeur de REQM estimé à 0,4594 mg/L et une valeur d'EMA estimé à 0,3381 mg/L ont souligné la supériorité du modèle TOD-RNFBR surpassant ainsi le modèle TOD-RNPMC.

L'analyse des performances des modèles hybrides qui combiner des RNA et la technique de décomposition DME montrent que durant la phase d'entraînement, le modèle DME-RNFBR présente les meilleures performances avec une valeur de R égal à 0,9752 et une valeur d'ENS égal à 0,9510, et des erreurs très faibles, avec un REQM de 0,6684 mg/L. Par contre lors de la phase de test, le modèle DME-RNPMC s'illustre avec une amélioration très remarque de ces performances avec R de 0,8967 et ENS de 0,7677, et des erreurs très faibles avec REQM de 0,3501 mg/L et EMA de 0,0866 mg/L.

L'analyse des modèles DMV-RNA a révélé que durant la phase d'entraînement, le modèle DMV-RNFBR a mieux performé avec des paramètres plus élevés : R de 0,9781, ENS de 0,9565 et d'erreur plus faible : REQM de 0,6297 mg/L par rapport à DMV-RNPMC. Par contre en la phase du test, les paramètres de performance de R égale à 0,8000, ENS égale à 0,4925, et une valeur plus faible d'erreur d'EMA égale à 0,1790 mg/L ont indiqué que le modèle DMV-RNPMC surpassait DMV-RNFBR.

La représentation graphique (Figure 5.9) des modèles RNA simples montre que le modèle RN-FBR a mieux à cause d'une valeur élevée du coefficient de détermination R^2 estimé à 0,9600.

La figure si dessous confirme aussi l'efficacité du modèle TOD-RNFBR avec un coefficient de détermination plus élevé (R^2 égale à 0,9511).

Le modèle DME-RNFBR est le meilleur lors de la représentation graphique parce que la valeur de R^2 est élevée égale à 0,9458.

La valeur de R^2 plus élevée du modèle DMV-RNFBR (R^2 de 0,9513) par rapport au modèle DMV-RNPMC (R^2 de 0,8731) indique que le premier modèle est meilleur.

Figure 5.10 indique les modèles TOD-RNFBR, DME-RNFBR et DMV-RNFBR avec des valeurs de corrélation excellente de 0,97, 0,97 et 0,98 respectivement, et même valeur d'ET de 2,65 mg/L, inférieure au ET observée de 2,72 mg/L. Ce modèle semble sous-estimer le PO_4^{3-} , et il présente une valeur minimale de REQM, d'environ 0 mg/L.

On observe que le modèle hybride TOD-RNFBR est le meilleur par rapport au modèle simple RN-FBR, avec une augmentation de R et ENS de 0,92 % et 1,85 %, respectivement, une réduction des erreurs REQM de 15,59 % et d'EMA de 19,04 % en phase d'entraînement ; les gains sont encore plus importants en phase de test (R = 17,32 %, ENS = 139,83 %, REQM = 43,18 %, EMA = 41,76 %) avec une architecture optimale de 65-130-1.

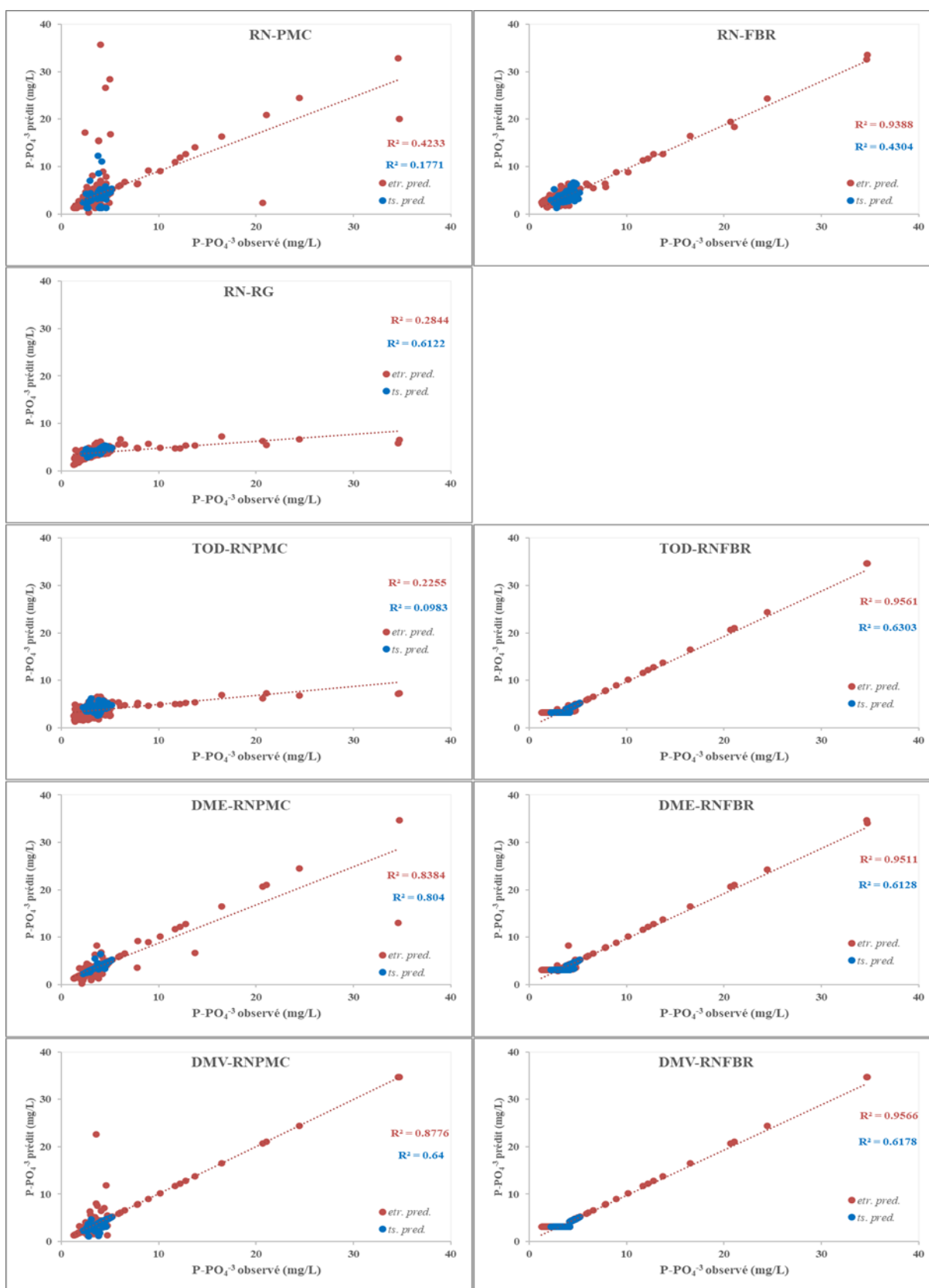


Figure 5.9: Nuage de points des modèles utilisés pour la prédiction de PO_4^{3-}

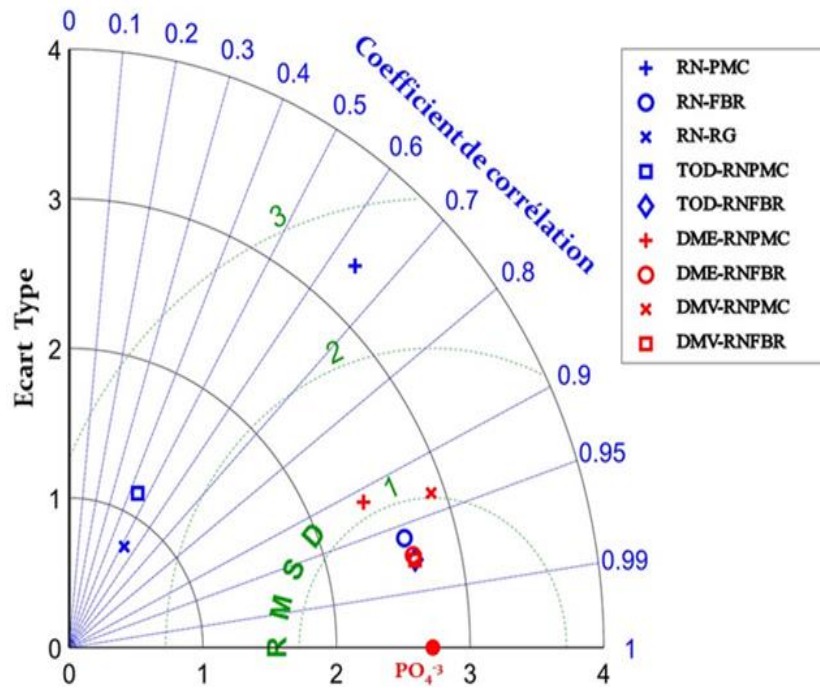


Figure 5.10: Diagramme de Taylor des différents modèles utilisés pour la prédiction de PO_4^{3-}

- **Indice de Boue**

La modélisation de l'indice de boue (IB) dans les stations d'épuration des eaux usées (STEP) est complexe en raison de la non-linéarité des procédés et de la variabilité temporelle des quantités d'influents. L'IB dépend également des pratiques de gestion et d'exploitation de la station.

Les résultats qualitatifs obtenus à l'aide des modèles simples et hybrides au cours des phases d'entraînement et de test sont présentés dans le tableau 5.8.

Tableau 5.8: Evaluation des performances des différents modèles pour la prédiction IB

Type de modèle		Entraînement				Test			
		REQM (mL/g)	EMA (mL/g)	R	ENS	REQM (mL/g)	EMA (mL/g)	R	ENS
Modèles simples	RN-PMC	4,7484	3,2540	0,8265	0,6781	3,0939	2,3163	0,8267	0,6636
	RN-FBR	3,6705	2,7647	0,8989	0,8077	2,8595	2,1221	0,8611	0,7127
	RN-RG	5,8611	4,2260	0,7729	0,5096	3,7125	2,8869	0,7918	0,5157
Modèles hybrides	TOD-RNPMC	3,4555	0,8938	0,9142	0,8295	2,1354	0,7045	0,9236	0,8398
	TOD-RNFBR	5,0608	3,7369	0,7978	0,6344	3,7619	2,8121	0,7782	0,5027
	TOD-RNRG	0,2075	0,0648	0,9997	0,9994	0,1291	0,0399	0,9997	0,9994
	DME-RNPMC	2,8568	0,8448	0,9422	0,8835	1,8253	0,4364	0,9467	0,8829
	DME-RNFBR	5,0320	3,7128	0,8006	0,6385	3,8544	2,9142	0,7708	0,4779
	DME-RNRG	0,2629	0,0890	0,9995	0,9990	0,0955	0,0430	0,9998	0,9997
	DMV-RNPMC	2,9877	0,8838	0,9372	0,8726	1,5503	0,4589	0,9594	0,9155
	DMV-RNFBR	5,0822	3,7374	0,7958	0,6313	3,7025	2,7028	0,7841	0,5183
	DMV-RNRG	0,0007	0,0000	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000

L'étude analytique des modèles RNA simples affiche que le modèle RN-FBR est le meilleur avec des erreurs plus faibles de REQM égale à 3,6705 mL/g et d'EMA égale à 2,7647 mL/g, et avec de meilleurs paramètres de performance des valeurs de R égale à 0,8989 et ENS égale à 0,8077 durant la phase d'entraînement. Lors de la phase de test, encore le modèle RN-FBR c'est le meilleur puisque les erreurs sont très faibles, de REQM égale à 2,8595 mL/g et d'EMA égale à 2,1221 mL/g, et les paramètres de performance sont élevés avec une R égale à 0,8611 et une ENS égale à 0,7127 par rapport aux modèles RN-PMC et RN-RG.

Les modèles TOD-RNA ont affiché que le modèle TOD-RNPMC est le meilleur avait de performance élevée de R égale à 0,9142 et d'ENS égale à 0,8295, et des erreurs très faible de REQM égale à 3,4555 mL/g et d'EMA égale à 0,8938 mL/g à la phase d'entraînement, et lors de la phase de test aussi confirme la supériorité de ce modèle avec des valeurs très faibles des erreurs de REQM égale à 2,1354 mL/g et d'EMA égale à 0,7045 mL/g, et avec valeurs des performances élevées de R et d'ENS sont respectivement 0,9236 et 0,8398.

Le modèle DME-RNPMC indique les meilleures caractéristiques durant la phase d'entraînement avec des niveaux élevés de performance de R égale à 0,9422 et d'ENS égale à 0,8835, et les meilleures plus faibles valeurs des erreurs de REQM égale à 2,8568 mL/g et d'EMA égale à 0,8448 mL/g et aussi en phase de test continue la supériorité de ce modèle avec de meilleurs valeurs de performance de R et d'ENS

respectivement 0,9467 et 0,8829 et avec les erreurs très faibles des valeurs de REQM et d'EMA respectivement 1,8253 mL/g et 0,4364 mL/g que celle des modèles DME-RNA.

Le modèle DMV-RNPMC affiche des performances supérieures avec R estimé à 0,9372 et ENS estimé à 0,8726, et les meilleures très faibles erreurs avec REQM estimé à 2,9877 mL/g et EMA estimé à 0,8838 mL/g à la phase d'entraînement, et durant la phase de test aussi confirme cette supériorité avec des paramètres de performance élevés de R estimé à 0,9594 et d'ENS estimé à 0,9155, et des valeurs d'erreurs très faibles de REQM estimé à 1,5503 mL/g et d'EMA estimé à 0,4589 mL/g.

La supériorité du modèle RN-FBR avec la meilleure valeur de R^2 égale à 0,8019 est confirmée graphiquement par la figure 5.11 que celle des modèles RNA simples.

L'efficacité du modèle TOD-RNPMC tandis que les modèles TOD-RNA clair à travers l'analyse numérique et graphiquement aussi par une meilleure valeur de performance de R^2 égale à 0,8396.

Aussi, en remarque à partir de la représentation graphique, indique que le modèle DME-RNPMC le meilleur parmi les modèles DME-RNA avec une valeur de R^2 égale à 0,8901, cela confirme les résultats obtenus durant l'analyse numérique.

La supériorité du modèle DMV-RNPMC dépassant les autres modèles du DMV-RNA analytiquement, et encore graphiquement, à cause de la valeur élevée de R^2 égale à 0,8834.

En concluant que pour la prédiction du paramètre d'IB, la supériorité des meilleurs modèles obtenue précédemment en les deux phases, il est donc difficile de choisir le modèle le plus approprié. La figure 5.12 illustre que les deux modèles DME-RNPMC et DMV-RNPMC sont les meilleurs puisqu'ils présentent des performances similaires, avec même valeur de corrélation plus élevées de 0,94 et un écart type de 8,02 mL/g et 7,98 mL/g et respectivement, ce qui est supérieur aux autres modèles et supérieur à l'écart type observé de 7,92 mL/g. Il est évident qu'une surestimation à l'IB, et avec des valeurs de REQM minimales d'environ 2 mL/g comparativement aux autres modèles.

Le modèle DMV-RNPMC est le meilleur. Avec une architecture optimale de TP-78-10-1, il améliore les performances avec une augmentation de R et ENS de 4,09 % et 7,44 %, respectivement, une réduction de REQM de 18,6 % et d'EMA de 68,03 % en phase d'entraînement ; les gains sont encore plus importants durant la phase de test (R = 10,25 %, ENS = 22,15 %, REQM = 45,78 %, EMA = 78,38 %) par rapport au modèle simple.

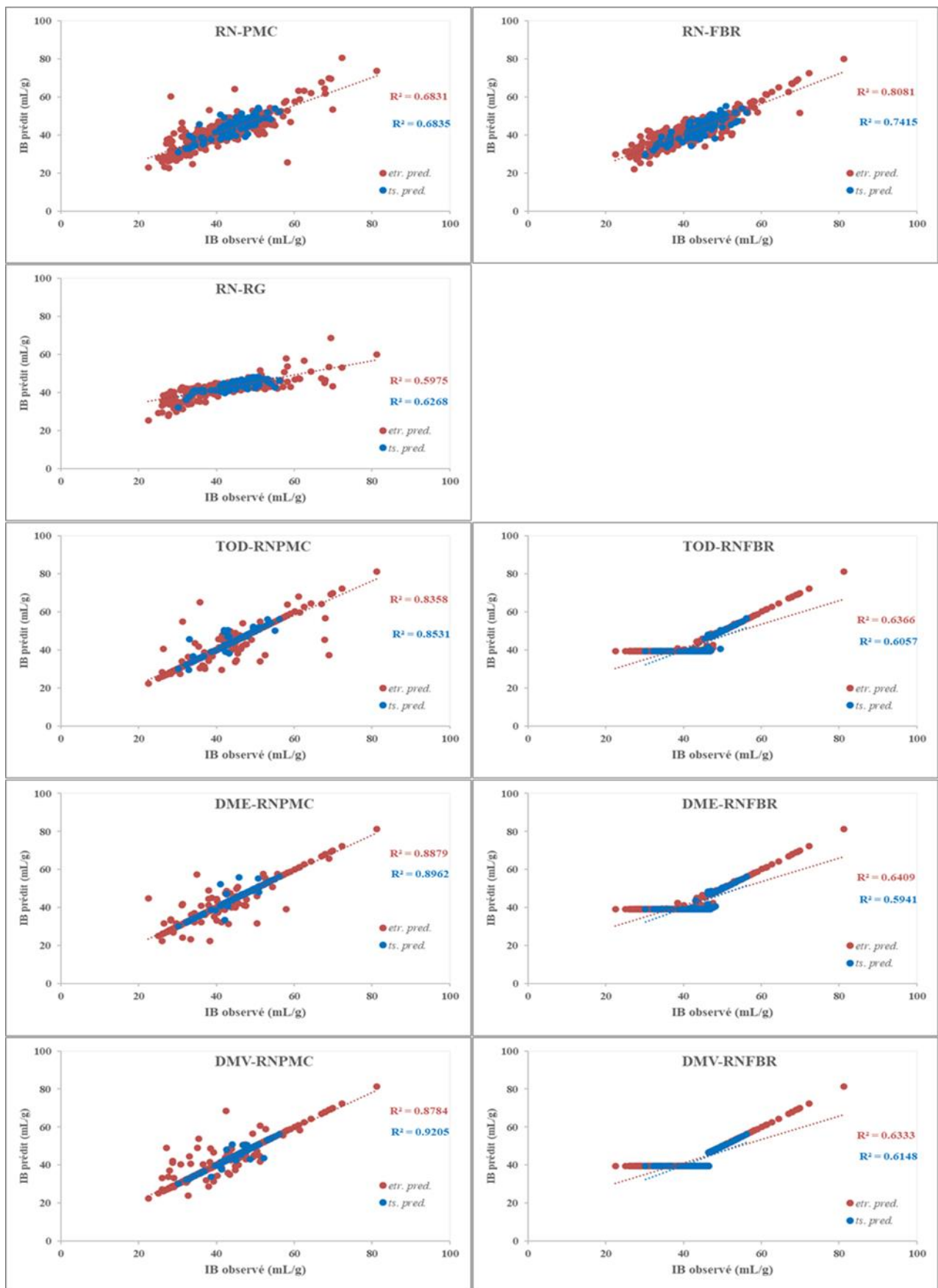


Figure 5.11: Nuage de points des modèles utilisés pour la prédiction de l'IB

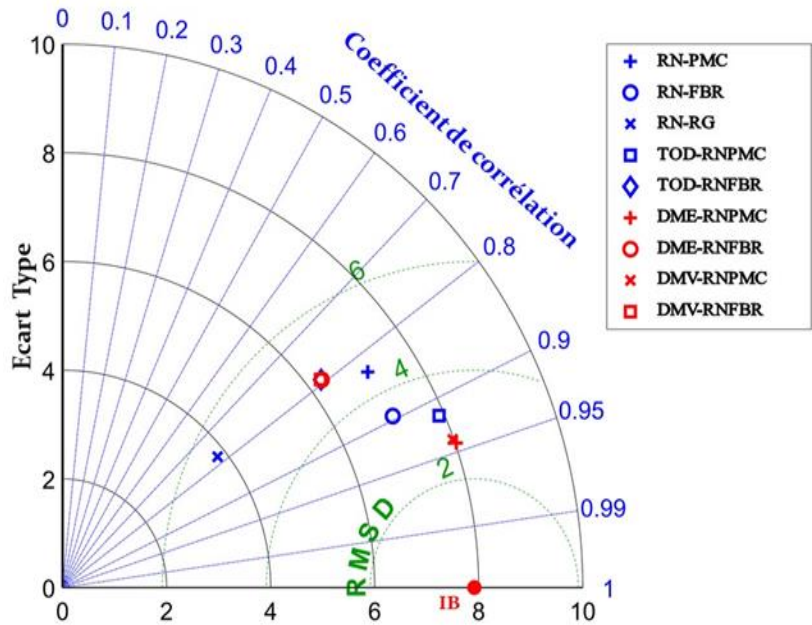


Figure 5.12: Diagramme de Taylor des différents modèles utilisés pour la prédiction de l'IB

- **Débit des Boues Séchées**

Le débit des boues séchées (QBS) représente un défi environnemental majeur en raison de la présence de métaux lourds et de polluants organiques qui entraînent une accumulation excessive (Cloutier et al., 2009). Une gestion efficace de ces boues est essentielle pour réduire les risques environnementaux et optimiser le traitement. Les gestionnaires et exploitants des STEP peuvent utiliser les données sur le séchage des boues pour anticiper les quantités journalières produites et améliorer les performances du processus. Les résultats quantitatifs de la modélisation simple et combinée du QBS sont présentés dans le tableau 5.9.

Tableau 5.9: Evaluation des performances des différents modèles pour la prédiction QBS

Type de modèle		Entraînement				Test			
		REQM (m ³ /j)	EMA (m ³ /j)	R	ENS	REQM (m ³ /j)	EMA (m ³ /j)	R	ENS
Modèles simples	RN-PMC	66,0379	31,2739	0,2376	0,0510	83,1460	39,4100	0,1743	0,0256
	RN-FBR	23,6372	17,8563	0,9375	0,8784	27,7346	20,6526	0,9460	0,8916
	RN-RG	59,5559	24,4037	0,5575	0,2282	75,4144	34,6301	0,6027	0,1984
Modèles hybrides	TOD-RNPMC	21,0312	5,6032	0,9548	0,9038	15,9638	5,5650	0,9834	0,9641
	TOD-RNFBR	17,2073	12,1128	0,9673	0,9356	21,4097	14,9754	0,9672	0,9354
	TOD-RNRG	3,9124	0,5860	0,9987	0,9967	0,4998	0,1550	1,0000	1,0000
	DME-RNPMC	35,7286	8,6218	0,8596	0,7222	33,4765	8,2521	0,9177	0,8420
	DME-RNFBR	17,1439	12,1754	0,9675	0,9360	20,9333	14,6020	0,9687	0,9382
	DME-RNRG	5,4361	0,9545	0,9976	0,9936	0,5388	0,1969	1,0000	1,0000
	DMV-RNPMC	30,4994	7,0460	0,9110	0,7976	49,0746	9,6949	0,8134	0,6606
	DMV-RNFBR	17,0755	12,0119	0,9678	0,9366	20,7423	13,9352	0,9694	0,9394
	DMV-RNRG	0,1361	0,0178	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000

Concernant l'étude analytique des modèles RNA simples, la supériorité du modèle RN-FBR. Lors de la phase d'entraînement, avec des erreurs plus faibles de REQM égale à 23,6372 m³/j et d'EMA égale à 17,8563 m³/j, et des valeurs élevées de performance de R et ENS respectivement égale à 0,9375 et 0,8784. Durant la phase de test, présentent les erreurs très faibles de REQM et d'EMA respectivement 27,7346 m³/j et 20,6526 m³/j, et les meilleures valeurs de performance de R d'une valeur de 0,9460 et d'ENS d'une valeur de 0,8916.

Le modèle TOD-RNFBR est le meilleur avec des paramètres de performance de R égale à 0,9673 et d'ENS égale à 0,9356 élevés, et une erreur de REQM avec une valeur de 17,2073 m³/j plus faible durant la phase d'entraînement. Par contre, les paramètres de performance avec R égale à 0,9834 et avec ENS égale à 0,9641, et les erreurs avec REQM égale à 15,9638 m³/j et avec EMA égale à 5,5650 m³/j indiquent qu'elles ont montré que le modèle TOD-RNPMC était le meilleur que celui des modèles TOD-RNA durant de la phase de test.

Les modèles DME-RNA, notamment le modèle DME-RNFBR, ont montré une performance élevée durant la phase d'entraînement, Avec un coefficient R de 0,9675, un indice ENS de 0,9360 et une erreur minimale de REQM de 17,1439 m³/j, ces paramètres démontrent l'efficacité de ce modèle par rapport au DME-RNPMC. Lors de la phase de test, les performances ont été confirmées avec des valeurs de R et d'ENS de

0,9687 et 0,9382 respectivement, et une erreur minimale de REQM de 20,9333 m³/j, positionnant le modèle DME-RNFBR comme le meilleur par rapport au DME-RNPMC.

Les modèles DMV-RNA ont montré, pendant la phase d'entraînement, que le modèle DMV-RNFBR est plus performant avec des résultats de R et ENS plus élevés, soit 0,9678 et 0,9366 respectivement, et une erreur plus faible avec un REQM estimé à 17,0755 m³/j. Pendant la phase de test, il maintient sa performance élevée avec un R de 0,9694 et une ENS de 0,9394, et un faible REQM de 20,7423 m³/j comparé au modèle DMV-RNPMC.

En continuant la supériorité du modèle RN-FBR que celle des modèles RNA simples, voici ce qu'il indique de la figure 5.13, avec la plus grande valeur du coefficient de régression R² égale à 0,8822.

La difficulté rencontrée lors du choix du meilleur modèle dans l'étude analytique réside dans les variations des résultats trouvées au cours des deux phases des modèles TOD-RNA. La représentation graphique a permis de surmonter cette difficulté en mettant en évidence les valeurs élevées du coefficient de corrélation lors des phases d'entraînement et de test, qui étaient respectivement de 0,9356 et 0,9355, cela indique que le modèle TOD-RNFBR est le meilleur choix, notamment en raison de la similarité des valeurs de R² trouvées dans les deux étapes.

Un avantage distinct du modèle DME-RNFBR est sa supériorité en termes de représentation graphique par rapport aux autres modèles de DME-RNA. Cette caractéristique s'est traduite par des valeurs plus élevées de R² lors des phases d'entraînement et de test, atteignant respectivement 0,9361 et 0,9384.

La représentation graphique confirme également l'efficacité continue du modèle DMV-RNFBR par rapport aux modèles DMV-RNA, avec des valeurs de R² similaires de 0,9374 tout au long des deux phases, soit 0,9366 et 0,9397 respectivement pour les phases d'entraînement et de test.

Nous avons facilement choisi le meilleur modèle analytiquement et graphiquement. Cependant, trouver le meilleur parmi ceux-ci est plus complexe. La figure 5.14 indique que les modèles TOD-RNFBR, DME-RNFBR et DMV-RNFBR sont les plus performants, avec même valeur de (R = 0,97), ainsi qu'un écart type de 69,06 m³/j, 69,01 m³/j et 69,13 m³/j respectivement, inférieur à l'écart type observé de 71,4 m³/j. Ces modèles semblent sous-estimer à QBS avec des valeurs minimales de REQM d'environ 20 m³/j.

Le modèle DMV-RNFBR est le meilleur. Avec une architecture optimale de 78-80-1, il améliore les performances avec une augmentation de R et ENS de 3,13 % et 6,21 %, respectivement, une réduction de REQM de 27,76 % et d'EMA de 32,73 % en phase d'entraînement ; les gains sont également présents durant la phase de test (R = 2,41 %, ENS = 5,09 %, REQM = 25,21 %, EMA = 32,53 %) par rapport au modèle simple.

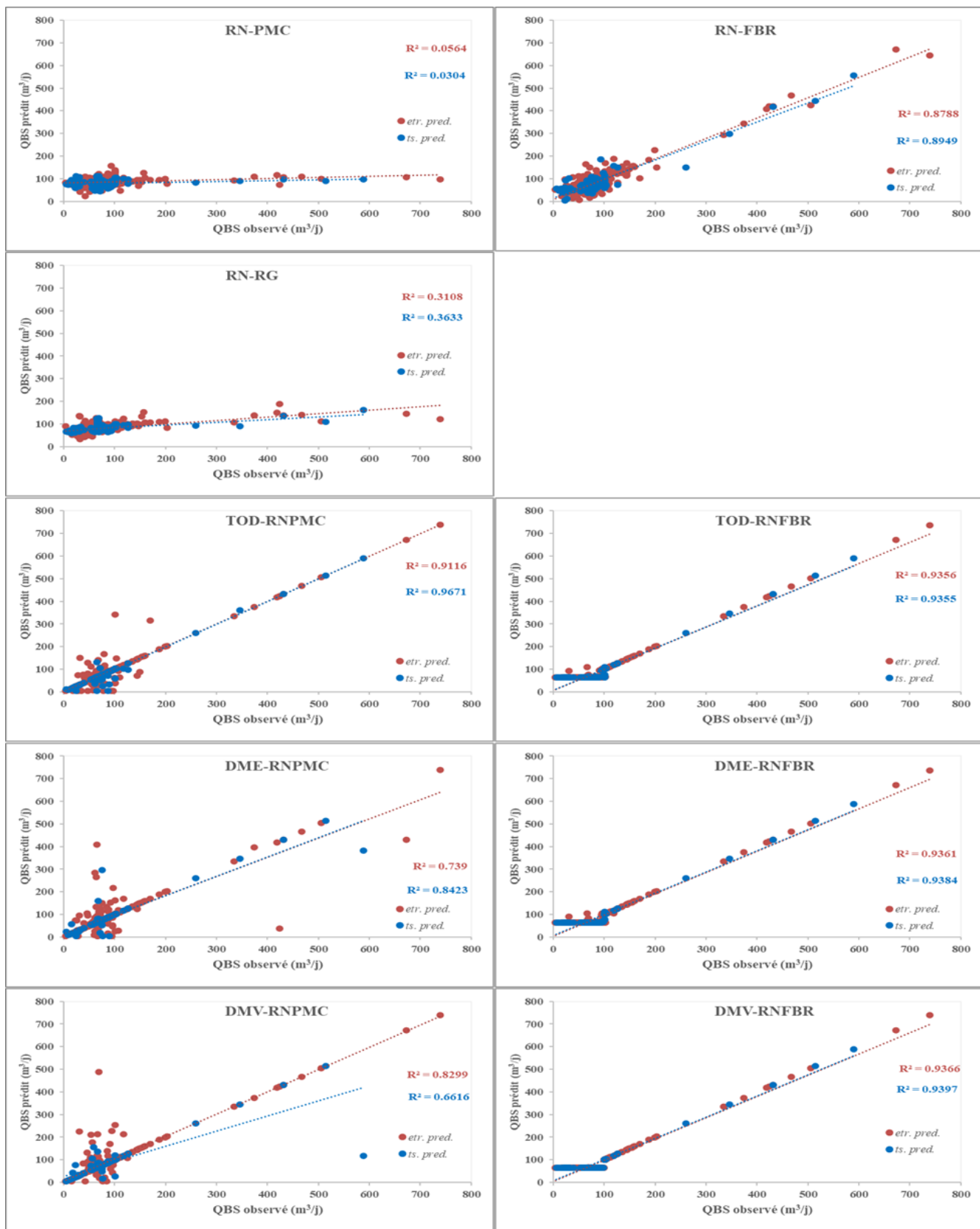


Figure 5.13: Nuage de points des modèles utilisés pour la prédiction de QBS

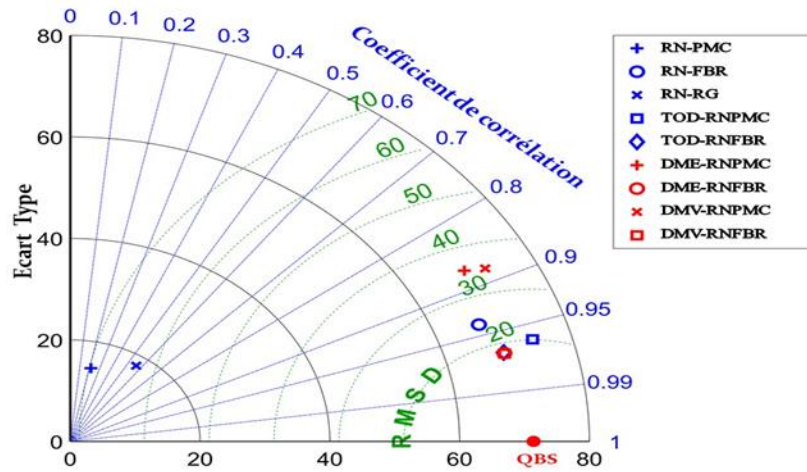


Figure 5.14: Diagramme de Taylor des différents modèles utilisés pour la prédiction du QBS

5.6.2 Modélisation des modèles à sorties multiples

La configuration des modèles à sortie multiple (MSM) permet de prédire simultanément plusieurs paramètres caractérisant la qualité des effluents.

Les performances qualitatives des modèles RNA simples et hybrides pour la prédiction des paramètres de la qualité des effluents au cours des deux phases d'analyse sont présentées dans le tableau 5.10.

Tableau 5.10: Evaluation de la performance des différents modèles pour la prédiction des paramètres de qualité des effluents

Type de modèle		Entraînement				Test			
		REQM (mg/L)	EMA (mg/L)	R	ENS	REQM (mg/L)	EMA (mg/L)	R	ENS
Modèles simples	RN-PMC	14,1296	12,9268	0,2463	-9,4069	13,3605	12,4785	0,1601	-11,0804
	RN-FBR	14,1352	12,9142	0,2763	-9,4547	13,5495	12,6371	0,1244	-11,4132
	RN-RG	13,7461	12,7387	0,2705	-8,6264	13,4680	12,6976	0,0964	-10,9629
Modèles hybrides	TOD-RNPMC	3,0272	2,0446	0,7383	0,5676	2,5300	1,8511	0,7279	0,1961
	TOD-RNFBR	3,4225	2,4370	0,7816	0,6208	2,7418	1,8469	0,6277	0,3565
	TOD-RNRG	0,3263	0,1277	0,9979	0,9954	0,1916	0,0866	0,9986	0,9971
	DME-RNPMC	2,7127	1,7844	0,8005	0,6493	2,2146	1,5405	0,7913	0,5532
	DME-RNFBR	3,4471	2,4537	0,7784	0,6168	2,7951	1,9278	0,6120	0,3308
	DME-RNRG	0,4648	0,1818	0,9940	0,9873	0,2842	0,1341	0,9975	0,9946
	DMV-RNPMC	2,9743	1,9709	0,7691	0,6067	2,6433	1,9500	0,6945	0,3086
	DMV-RNFBR	3,4379	2,4516	0,7749	0,6116	2,6463	1,7132	0,6319	0,3786
	DMV-RNRG	0,0127	0,0033	1,0000	1,0000	0,0015	0,0004	1,0000	1,0000

Pour les modèles RNA simples pendant l'entraînement et le test, on note que les trois modèles montrent des performances médiocres et des erreurs importantes, illustrant la difficulté des modèles à généraliser et prédire correctement les valeurs.

Pour les modèles TOD-RNA, il a été constaté que pendant la phase d'entraînement, le modèle TOD-RNFBR affichait les performances les plus élevées : R de 0,7816 et ENS de 0,6208, tandis que le modèle TOD-RNPMC présentait les erreurs les plus faibles : REQM de 3,0272 mg/L et EMA de 2,0446 mg/L. Lors de la phase de test, le modèle TOD-RNFBR se démarquait avec un R de 0,7279 et un faible REQM de 2,5300 mg/L, tandis que le modèle TOD-RNPMC excellait avec une ENS de 0,3565 et un EMA de 1,8469 mg/L.

En raison des divers résultats de l'étude analytique et de la difficulté à choisir le meilleur modèle, nous avons introduit le coefficient de détermination linéaire pour faciliter notre sélection. Nous avons observé que le modèle TOD-RNFBR est préférable en raison de son R^2 plus élevé (0,6165) par rapport au modèle TOD-RNPMC (0,5778) en ce qui concerne les résultats des modèles TOD-RNA.

Pour l'étude de performance des modèles DME-RNA, le modèle DME-RNPMC se distingue avec les meilleures performances : R d'une valeur estimée 0,8005 et ENS de 0,6493, et les erreurs les plus faibles : REQM : 2,7127 mg/L, EMA : 1,7844 mg/L lors de la phase d'entraînement. Pour la phase de test, il affiche des valeurs de R à 0,7913, ENS à 0,5532, REQM à 2,2146 mg/L et EMA à 1,5405 mg/L, ce qui confirme que ce modèle est le meilleur.

Le coefficient de détermination confirme que le modèle DME-RNPMC est supérieur (R^2 de 0,6610) par rapport au modèle DME-RNFBR (R^2 de 0,6084).

Les modèles DMV-RNA ont démontré des valeurs acceptables, le modèle DMV-RNFBR s'affiche de meilleurs paramètres de performance : R = 0,7749 et ENS égale à 0,6116, en revanche, le modèle DMV-RNPMC s'est distingué par ses faibles erreurs (REQM = 2,9743 mg/L et EMA = 1,9709 mg/L) durant la phase d'entraînement.

Durant la phase de test, le modèle DMV-RNPMC a affiché une valeur acceptable de R et une erreur de REQM réduite (respectivement : 0,6945 et 2,6463 mg/L), tandis que l'efficacité du modèle DMV-RNFBR s'est caractérisée par une mauvaise valeur d'ENS et d'une erreur faible d'EMA (respectivement : 0,3786 et 1,7132 mg/L).

Il convient de noter que le modèle DMV-RNFBR s'est avéré supérieur en raison de son R^2 plus élevé (0,6149) par rapport au modèle DMV-RNPMC (R^2 de 0,6045).

La comparaison de la configuration des modèles à sortie multiple des modèles hybrides précédemment démontre que le modèle DME-RNPMC présente les valeurs les plus élevées des paramètres de performance et les erreurs les plus faibles.

L'étude réalisée par Djeddou (2014) sur la station d'épuration de Khenchela visait à modéliser la fiabilité et les taux de défaillance de la station à l'aide de RNA pour trois paramètres principaux : la DBO₅, la DCO et les MES. La méthodologie impliquait la construction de modèles RNA pour prédire les taux de fiabilité et de défaillance, accompagnée de tests statistiques visant à valider les distributions des données, qui se sont révélées log-normales. Les résultats obtenus ont montré des coefficients de régression R atteignant jusqu'à 0,97, indiquant un excellent ajustement des valeurs prédites et prouvant que les RNA représentent une alternative robuste pour prédire les performances des stations d'épuration et identifier efficacement les défaillances éventuelles du processus. De son côté, l'étude menée par Bekkarri (2020) sur la station d'épuration de Touggourt visait à évaluer les performances de la station via des modèles RNA-PMC. L'objectif principal était de prédire les concentrations des paramètres clés de pollution, notamment la DBO₅, la DCO et les MES. Deux types de modèles ont été adoptés : un modèle simple se basant uniquement sur les paramètres d'entrée et un modèle extensif intégrant également les sorties corrélées. Les résultats ont révélé que le modèle extensif offrait de meilleures performances, avec des coefficients de corrélation R variant entre 0,9 et 0,999 lors des phases d'entraînement, de validation et de test. Les prédictions obtenues étaient non seulement précises, mais également associées à un faible taux d'erreur quadratique moyen (REQM), confirmant la robustesse du modèle RNA même avec des données limitées.

Par ailleurs, notre étude sur la station d'épuration d'Aïn Beïda avait pour but de construire et de comparer deux approches de modélisation : la modélisation simple et la modélisation hybride. Les résultats montrent que les modèles hybrides améliorent significativement la précision par rapport aux modèles simples, avec des paramètres de performance, tout en réduisant les erreurs. De plus, ces modèles hybrides contribuent à diminuer le sur-apprentissage et à optimiser la généralisation des prédictions.

Les trois études soulignent donc l'efficacité des RNA pour prédire les performances des stations d'épuration. Bekkarri (2020) met en avant la robustesse des modèles avec des données limitées, tandis que Djeddou (2014) illustre la fiabilité des RNA pour diagnostiquer et prévenir les défaillances. Ce travail démontre également l'avantage des modèles hybrides pour améliorer à la fois la précision et la généralisation.

Ces résultats confirment que les RNA, notamment sous leurs formes hybrides ou extensives, sont des outils puissants pour l'évaluation, la modélisation et l'optimisation des processus dans les stations d'épuration.

5.7 Conclusion

Ce chapitre présente deux approches de modélisations développées pour prédire les performances du processus de la STEP d' Aïn Beïda :

1. Modèles simples de réseaux de neurones artificiels (RNA) ;
2. Modèles hybrides de RNA, intégrant des techniques de décomposition des données.

Une analyse comparative a été réalisée pour évaluer l'efficacité de ces deux approches, en mesurant les améliorations des paramètres de performances (R, ENS) et la réduction des erreurs (REQM, EMA) des modèles hybrides par rapport aux modèles simples.

Nos simulations comparatives de MSU ont révélé que le modèle à FBR surpasse les autres modèles (PMC et RG) dans la capture de relations complexes. Grâce à sa fonction d'activation localisée, le FBR excelle dans la modélisation des non-linéarités et s'adapte efficacement aux variations locales des données. Cette caractéristique lui confère un avantage significatif en termes de généralisation, réduisant ainsi les risques de sur-apprentissage.

Les résultats obtenus lors de cette étude montrent que les modèles hybrides, qui combinent RNA et techniques de décomposition, ont systématiquement surpassé les modèles simples lors des phases d'entraînement et de test. Cette supériorité s'explique par la capacité des modèles hybrides à capturer des relations complexes et non linéaires présentes dans les données.

En ce qui concerne nos simulations comparatives de modèles à sorties multiples, le modèle hybride DME-RNPMC s'est révélé être le plus performant. Cependant, il demeure moins précis que les MSU.

Les résultats obtenus indiquent clairement que les modèles hybrides offrent une avancée significative dans la prévision des performances des STEP. Les améliorations apportées par ces modèles permettent une meilleure précision dans la prédiction de paramètres critiques, tels que la DBO₅, la DCO, les MES, et d'autres indicateurs de qualité des effluents.

L'utilisation des modèles hybrides contribue à optimiser le contrôle opérationnel des STEP en réduisant les incertitudes liées aux perturbations environnementales. Cette avancée technique constitue un outil puissant pour améliorer les performances des STEP, permettant ainsi une meilleure gestion des ressources en eau et une réduction des impacts environnementaux.

Pour la plupart des paramètres, des tendances similaires ont été observées, où les modèles hybrides surpassent les modèles simples en termes de précision et de qualité des prédictions. Ces améliorations se traduisent par une augmentation notable des paramètres de performances, ainsi que par une réduction significative des erreurs.

Conclusion générale

Actuellement, les modèles d'intelligence artificielle telles que les réseaux de neurones artificiels combinés aux techniques de décompositions ont été largement appliquées dans plusieurs domaines. Cette étude présente une contribution significative à l'adoption de ces approches pour modéliser et prédire les différents paramètres de fonctionnement des STEP, en proposant des méthodologies adaptées pour les systèmes complexes comme celui de la STEP d'Aïn Beïda.

Le contrôle et la possibilité de se doter de modèles pour la prédiction des différents paramètres d'une station d'épuration sont importants pour maintenir la stabilité du système dans un large éventail de circonstances et éviter de perturber l'équilibre environnemental. L'étude de la disponibilité de modèles caractérisant le comportement du STEP en tant que système dynamique est donc une première étape nécessaire. Toutefois, la complexité des processus impliqués, ainsi que l'hétérogénéité, l'incomplétude, le bruit et inexacte des données du STEP, rendent cette tâche particulièrement difficile. Les techniques développées dans ce travail permettent de surmonter ces limitations en exploitant les capacités des RNA et des techniques de décomposition.

Les résultats obtenus démontrent que les modèles RNA hybrides, combinés à des techniques de décomposition, surpassent les modèles simples en termes de précision et de robustesse. Les meilleurs modèles a été choisie pour prédire chaque paramètre utilisé la configuration MSU sont les suivants :

- ✓ DBO₅ : Modèle *DME-RNPMC* avec une corrélation linéaire de 0,96 ;
- ✓ DCO : Modèle *TOD-RNPMC* avec une corrélation de 0,96 ;
- ✓ MES : Modèle *TOD-RNFBR* avec une corrélation de 0,86 ;
- ✓ NT : Modèle *RNFBR* avec une corrélation de 0,85 ;
- ✓ PO₄³⁻ : Modèle *TOD-RNFBR* avec une corrélation de 0,98 ;
- ✓ IB : Modèle *DMV-RNPMC* avec une corrélation de 0,94 ;
- ✓ QBS : Modèle *DMV-RNFBR* avec une corrélation de 0,97.

Le modèle *DME-RNPMC* avec une architecture optimale de TP-68-10-5 a montré les meilleures performances pour prédire simultanément plusieurs paramètres (la configuration MSM), atteignant une corrélation linéaire de 0,8130 entre les valeurs mesurées et prédites.

Les différentes techniques de décomposition appliquées aux données extraites ont montré des performances remarquables pour prédire les paramètres des performances de la station d'Aïn Beïda. Parmi les caractéristiques des techniques de décomposition, la capacité de reconnaître et d'apprendre les relations sous-jacentes entre l'entrée et la sortie sans considération physique explicite, indépendamment de la dimensionnalité du problème et de la non-linéarité du système, et la tolérance élevée aux données contenant des bruits et des erreurs de mesure sont justifiées par les entreprises de cette thèse. En outre,

l'efficacité et la fiabilité des techniques proposées pour la sélection variable des données d'entrée afin d'augmenter la capacité du réseau résultant à prédire ont montré leur efficacité pour optimiser les performances de la STEP d'Aïn Beïda, permettant ainsi aux opérateurs de réagir rapidement aux changements des conditions opérationnelles tout en réduisant les coûts.

Les résultats expérimentaux ont confirmé que la configuration à sortie unique offre une meilleure précision et une meilleure utilité par rapport à la configuration à sorties multiples. Ces modèles permettent d'améliorer le contrôle des STEP, garantissant une meilleure qualité des effluents traités et une exploitation optimale.

Cette étude souligne également que les RNA combinés à des techniques de décomposition constituent une approche fiable pour prédire les performances des STEP. Toutefois, des recherches supplémentaires sont nécessaires pour améliorer encore ces modèles et explorer de nouvelles méthodologies, notamment :

- L'intégration de techniques de décomposition avancées comme l'Ensemble de Décomposition en Mode Empirique, l'Analyse de Décomposition en Mode Complet ou la Décomposition en Mode Variationnel Multiple ;
- L'utilisation de méthodes d'apprentissage automatique (Machine Learning) et d'apprentissage profond (Deep Learning) pour capturer des dynamiques encore plus complexes.

Références bibliographiques

- Al Saleh, H. A. (2021). Wastewater pollutants modeling using artificial neural networks. *Journal of Ecological Engineering*, 22(7), 35-45. <https://doi.org/10.12911/22998993/138872>
- Basheer, I. A., et Hajmeer, M. (2000). Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *Journal of Microbiological Methods*, 43(1), 3-31. [https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3)
- Bassompierre, C. (2007). *Procédé à boues activées pour le traitement d'effluents papetiers: de la conception d'un pilote à la validation de modèles*. Thèse de doctorat, Institut National Polytechnique de Grenoble, France, 230 pages.
- Baughman, D. R. (1995). *Neural Networks in Bioprocessing and Chemical Engineering*. Thèse de doctorat, Virginia Polytechnic Institute and State University, USA, 791 pages.
- Bekkari, N. (2020). *Etude et suivi de procédé d'épuration des eaux usées sous climat aride (simulation par l'application des méthodes d'intelligence artificielle) cas de la station d'épuration de Touggourt*. Thèse de doctorat, Université Kasdi Merbah-Ouargla, Algérie, 133 pages.
- Berthouex, P. M., Hunter, W. G., Pallesen, L., et Shih, C. Y. (1978). Dynamic behavior of an activated sludge plant. *Water Research*, 12(11), 957-972. [https://doi.org/10.1016/0043-1354\(78\)90079-9](https://doi.org/10.1016/0043-1354(78)90079-9)
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Royaume-Uni : Clarendon Press. 482 pages.
- Chen, J., et Beck, M. B. (1993). Modelling, control and on-line estimation of activated sludge bulking. *Water Science and Technology*, 28(11-12), 249-256. <https://doi.org/10.2166/wst.1993.0664>
- Chen G., van Loosdrecht M.C.M., Ekama G.A. et Brdjanovic D. (2023). *Biological Wastewater Treatment: Principles, Modelling and Design*. 2nd Edition. International Water Association Publishing. <https://doi.org/10.2166/9781789060362>
- Cloutier, F., Jalby, G., Lessard, P., et Vanrolleghem, P. A. (2009). Dynamic modelling of heavy metals behavior in wastewater treatment plants. *Journal of Water Science*, 22(4), 461-471. <https://doi.org/10.7202/038325ar>
- Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics. SIAM. Philadelphia. <http://dx.doi.org/10.1137/1.9781611970104>
- Demuth, H. et Beale, M., (1998). *Neural Network Toolbox: for use with MATLAB: user's guide*. The Math Works. Incorporated.
- Dispan, J. (2003). *Biodégradabilité de la matière organique dans le continuum aquatique réseau d'assainissement-station d'épuration-milieu naturel récepteur*. Thèse de doctorat, Marne-la-vallée, Ecole Nationale des Ponts et Chaussées, France, 239 pages.
- Djeddou, M. (2014). *Prévision du taux d'échec avec les réseaux neurones artificiels dans une station de traitement des eaux résiduaires*. Thèse de doctorat, Université Mohammed Khider-Biskra, Algérie, 175 pages.
- Djeddou, M., et Achour, B. (2015). The use of a neural network technique for the prediction of sludge volume index in municipal wastewater treatment plant. *Larhyss Journal*, No 24, 351-370. <http://larhyss.net/ojs/index.php/larhyss/article/view/351>
- Dragomiretskiy, K., et Zosso, D. (2013). Variational mode decomposition. *IEEE transactions on signal processing*, 62(3), 531-544. <https://doi.org/10.1109/TSP.2013.2288675>
- Dreyfus, G., Martinez, J., Samuelides, M., Gordon, M. B., Badran, F., et Thiria, S. (2011). *Apprentissage statistique : Réseaux de neurones - Cartes topologiques - Machines à vecteurs supports*. Eyrolles. 450 p.
- Du, Y. G., Tyagi, R. D., et Bhamidimarri, R. (1999). Use of fuzzy neural-net model for rule generation of activated sludge process. *Process Biochemistry*, 35(1-2), 77-83. [https://doi.org/10.1016/S0032-9592\(99\)00035-7](https://doi.org/10.1016/S0032-9592(99)00035-7)

- Garrick, M., Cunnane, C., et Nash, J. E. (1978). A criterion of efficiency for rainfall-runoff models. *Journal of Hydrology*, 36(3-4), 375-381. [https://doi.org/10.1016/0022-1694\(78\)90155-5](https://doi.org/10.1016/0022-1694(78)90155-5)
- Gernaey, K.V., Van Loosdrecht, M.C.M., Henze, M., et Jorgensen, S.B. (2004). Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling and Software*, 19 (9), 763-783. <https://doi.org/10.1016/j.envsoft.2003.03.005>
- Google (2024). Aïn Beïda wastewater treatment plant cardiff, Available at: <https://earth.google.com/web/@35.7882968,7.33849481,928.45273645a,575.97066821d,35y,0h,00r>, Accessed 25 June 2024.
- Gujer W., Henze M., Mino T. et van Loosdrecht M.C.M. (1999). Activated Sludge Model No. 3. *Water Science Technology*, 39(1), 183-193. [https://doi.org/10.1016/S0273-1223\(98\)00785-9](https://doi.org/10.1016/S0273-1223(98)00785-9)
- Gurney, K. (1997). *An Introduction to Neural Networks*. Royaume-Uni: Taylor & Francis.
- Hauduc, H. (2010). *Modèles biocinétiques de boues activées de type ASM: Analyse théorique et fonctionnelle, vers un jeu de paramètres par défaut*. Thèse de doctorat, Faculté des études supérieures et postdoctorales, Université Laval, Québec-Canada, 222 pages.
- Haykin, S. (2009). *Neural networks and learning machines*. 3rd edition. Pearson Education India.
- Hebb, D. O. (1949). *The organization of the behavior : A neuropsychological theory*. John Wiley et Sons. New York. 335 p. Incorporated. <https://doi.org/10.1002/sce.37303405110>
- Henze M., Grady C.P.L., Gujer W., Marais G.v.R. et Matsuo T. (1987). Activated sludge model No.1. *Water Science and Technology*, 29pp.
- Henze M., Gujer W., Mino T., et van Loosdrecht M.C.M. (2006). *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*. United Kingdom: IWA Publishing. <https://doi.org/10.2166/9781780402369>
- Henze M., Gujer W., Mino T., Matsuo T., Wentzel M.C et Marais, G.v.R. (1995). The activated sludge model No. 2 : Biological Phosphorus Removal. *Water Science and Technology*, Volume 31, 1-11. <https://api.semanticscholar.org/CorpusID:94353095>
- Henze M., Gujer W., Mino T., Matsuo T., Wentzel M.C., Marais G.V.R. et van Loosdrecht M.C.M. (1999). Activated sludge Model No.2d, ASM2d. *Water science and technology*, 39(1), 165-182. <https://doi.org/10.2166/wst.1999.0036>
- Henze M., van Loosdrecht M. C. M., G. A. Ekama, et Brdjanovic D. (2008). *Biological Wastewater Treatment: Principles, Modelling and Design*. International Water Association publishing. <https://doi.org/10.2166/9781780401867>
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10), 3088-3092. <https://doi.org/10.1073/pnas.81.10.3088>
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.C., Tung, C.C. et Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971), 903-995. <https://doi.org/10.1098/rspa.1998.0193>
- Huang, N. E., Wu, M. L. C., Long, S. R., Shen, S. S., Qu, W., Gloersen, P., et Fan, K. L. (2003). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459(2037), 2317-2345. <https://doi.org/10.1098/rspa.2003.1123>
- Jain, A.K., Mao, J., et Mohiuddin, K.M. (1996). Artificial neural networks: a tutorial. *Journal of Computer*, Vol. 29,31-44. <https://api.semanticscholar.org/CorpusID:270618892>
- Kessler, L. H., et Nichols, M. S. (1935). Oxygen utilization by activated sludge. *Sewage Works Journal*, 7 (5), 810-838. <https://www.jstor.org/stable/25028564>
- Krose, B., et Smagt, P. V. (1996). *An Introduction to Neural Network*. 8th edition. University of Amsterdam, Netherland.
- Lacroix, P. G., et Bloodgood, D. E. (1972). Computer simulation of activated sludge plant operation. *Journal of Water Pollution Control Federation*, 44 (9), Annual Conference Issue, 1778-1788.

- Legates, D. R., et McCabe Jr, G. J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233-241. <https://doi.org/10.1029/1998WR900018>
- Li, H., Zhang, Y., et Zheng, H. (2009). Hilbert-Huang transform and marginal spectrum for detection and diagnosis of localized defects in roller bearings. *Journal of mechanical science and technology*, 23(2), 291-301. <https://doi.org/10.1007/s12206-008-1110-5>
- Lou, I., et Zhao, Y. (2012). Sludge bulking prediction using principal component regression and artificial neural network. *Mathematical Problems in Engineering*, 2012(1), 1-17. <https://doi.org/10.1155/2012/237693>
- Mallat, S. G. (2009). *A wavelet tour of signal processing: the sparse way*. 3rd Revised Edition. Academic press. <https://doi.org/10.1016/B978-0-12-374370-1.X0001-8>
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7), 674-693. <https://doi.org/10.1109/34.192463>
- Mandic, D. P., et Chambers, J. A. (2001). *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. John Wiley et Sons. Incorporated.
- McCulloch, W. S., et Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, vol. 5, 115-133. <https://doi.org/10.1007/BF02478259>
- McKinney, R. E. (1962). Mathematics of complete-mixing activated sludge. *Journal of the Sanitary Engineering Division*, 88(3), 87-113.
- Meyer, Y. (1990). *Ondelettes et Opérateurs I*. Hermann. Paris. 532 p.
- Meyer, Y. (1993). *Les ondelettes : algorithmes et applications*. Armand colin, Paris. 196p.
- Meyer, Y., Jaffard, S., et Rioul, O. (1987). L'analyse par ondelettes. *Pour la science*, 28-37.
- Minsky, M. L., et Papert, S. A. (1969). *Perceptrons*. MIT Press, Cambridge, MA. 292p.
- Misiti, M., Misiti, Y., Oppenheim, G., et Poggi, J. M. (2003). *Les ondelettes et leurs applications*. Hermès science publications.
- Najafzadeh, M., et Zeinolabedini, M. (2018). Derivation of optimal equations for prediction of sewage sludge quantity using wavelet conjunction models: an environmental assessment. *Environmental Science and Pollution Research*, 25(23), 22931-22943. <https://doi.org/10.1007/s11356-018-1975-5>
- Nash, J. E., et Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-A discussion of principles. *Journal of hydrology*, 10(3), 282-290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Parizeau, M. (2004). *Réseaux de neurones*. GIF-21140 et GIF-64326, 124. Université Laval, Canada, 117p.
- Personnaz, L., et Rivals, I. (2003). *Réseaux de neurones formels pour la modélisation, la commande et la classification*. CNRS édition (Paris). 464 p.
- Ranta, J., Kaitala, V., et Halme, A. (1977). Modelling and simulation of bacterial population age distribution in activated sludge plant. *IFAC Proceedings Volumes*, 10 (3), 1977, 93-101. <https://doi.org/10.1016/B978-0-08-022029-1.50022-3>
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386-408.
- Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan books.
- Rumelhart, D. E., Hinton, G. E., et Williams, R. J. (1986). Learning representations by back-propagating errors. *Journal of Nature*, 323(6088), 533-536.
- Sahadev, S., Madhu, G., et Thomas, M. R. (2023). Modeling of activated sludge process using multi-layer perceptron neural networks. *Nature Environment and Pollution Technology*, 22(1), 445-461. <https://doi.org/10.46488/NEPT.2023.v22i01.043>
- Smith, R. (1969). Preliminary design of wastewater treatment systems. *Journal of the Sanitary Engineering Division*, 95(1), 117-148. <https://doi.org/10.1061/JSEDAI.0000929>

- Steffens, M. A., Lant, P. A., et Newell, R. B. (1997). A systematic approach for reducing complex biological wastewater treatment models. *Water research*, 31(3), 590-606. [https://doi.org/10.1016/S0043-1354\(96\)00273-4](https://doi.org/10.1016/S0043-1354(96)00273-4)
- Touzet, C. (1992). *Les réseaux de neurones artificiels, introduction au connexionnisme*. 129 pages.
- Ullrich, A. H., et Smith, M. W. (1951). The biosorption process of sewage and waste treatment. *Sewage and Industrial Wastes*, 23 (10), 1248-1253. <https://www.jstor.org/stable/25031721>
- Viana Di Prisco, G. (1984). Hebb synaptic plasticity. *Progress in Neurobiology*, 22 (2), 89-102. [https://doi.org/10.1016/0301-0082\(84\)90021-2](https://doi.org/10.1016/0301-0082(84)90021-2)
- Wallis, R.H. (1976). An approach to the space variant restoration and enhancement of images. *Proceedings of the Symposium on Current Mathematical Problems in Image Science*, Monterey, CA, USA.
- White, H. (1992). *Artificial neural networks: approximation and learning theory*. Blackwell Publishers. 320 p. Incorporated.
- Widrow, B., et Hoff, M. E. (1960). *Adaptive switching circuits*. In IRE WESCON convention record, 4(1), 96-104.
- Wongburi P., et Park J.K. (2022). Prediction of sludge volume index in a wastewater treatment plant using recurrent neural network. *Sustainability*, 14(10), 62-76. <https://doi.org/10.3390/su14106276>
- Zaidi K., Djeddou, M., Sekiou F., Hameed I.A., et Shawaqfah M. (2023). Predictive modelling of daily dried sludge production in full-scale wastewater treatment plant using different machine learning combined with empirical mode decomposition. *Larhyss Journal, Issue (56)*, 77-106. <https://larhyss.net/ojs/index.php/larhyss/article/view/13777>
- Zare Abyaneh, H. (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal of Environmental Health Science and Engineering*, 12(1), 40, 1-8. <https://doi.org/10.1186/2052-336X-12-40>
- Zeinolabedini, M., Najafzadeh, M. (2019). Comparative study of different wavelet-based neural network models to predict sewage sludge quantity in wastewater treatment plant. *Environmental Monitoring and Assessment*, 191(3), 1-25. <https://doi.org/10.1007/s10661-019-7196-7>
- Zupan, J., et Gasteiger, J. (1993). *Neural network for chemists and drug design*. Allemagne : Wiley.