

Université Mohamed Khider – Biskra
Faculté des Sciences et de la technologie
Département de Génie civil et d'Hydraulique
Réf :/2014.....



جامعة محمد خيضر بسكرة
كلية العلوم و التكنولوجيا
قسم: الهندسة المدنية والري
المرجع :/2014.....

Thèse présentée en vue de l'obtention du diplôme de

Doctorat en sciences

Spécialité: Hydraulique

Thème

Prévision du taux d'échec avec les réseaux neurones artificiels dans une station de traitement des eaux résiduaires

Présenté par :

DJEDDOU Messaoud

Soutenu publiquement le : *04 Décembre 2014*

Devant le jury composé de :

Pr. ACHOUR Samia	Professeur	Présidente	Université de Biskra
Pr. ACHOUR Bachir	Professeur	Rapporteur	Université de Biskra
Pr. BOUDOUKHA Abderrahmane	Professeur	Examineur	Université de Batna
Pr. CHAFFAI Hicham	Professeur	Examineur	Université d'Annaba
Pr. DJABRI Larbi	Professeur	Examineur	Université d'Annaba
Pr. GUERGAZI Saadia	Professeur	Examinatrice	Université de Biskra

DEDICACE

À mon Père,

À ma Mère,

À ma femme,

À mon fils Yanis,

À mes frères et sœurs,

À la mémoire de mon oncle DJEDDOU Lamine,

À toutes les personnes que je porte dans mon cœur.

Je vous dédie cette thèse en guise de reconnaissance.

Djeddou Messaoud

REMERCIEMENTS

Tout d'abord je remercie « ALLAH » pour m'avoir donné le courage de finir ce travail, et essayer d'apporter une petite contribution dans l'effort d'améliorer et protéger notre environnement.

J'exprime ma profonde gratitude et mes sincères remerciements à mon directeur de

Thèse, Mr. ACHOUR Bachir professeur à l'université de Biskra, qui m'a accordé sa confiance pour diriger cette recherche, et m'a permis d'être l'un de ces disciples.

Sa contribution hautement distinguée dans la réalisation de ce travail me restera à l'esprit, son sens d'appréciation scientifique, sa rigueur pour le travail bien fait, ses suggestions pertinentes, et surtout sa disponibilité.

Ma reconnaissance et mes vives remerciements vont également au Dr. MARTAUD Maurice co-directeur de thèse et expert au CTA de la Lyonnaise des Eaux (France) pour ses différentes orientations et ces questions qui me pousser à aller apprendre et comprendre pour pouvoir trouver des réponses.

Je tiens à remercier également:

- Madame ACHOUR Samia, Professeur à l'Université de Biskra, qui m'a fait l'honneur de présider le jury.
- Monsieur BOUDOUKHA Abderrahmane, Professeur à l'Université de Batna, qui a bien voulu examiner ce travail.
- Monsieur CHAFFAI Hicham, Professeur à l'Université d'Annaba, qui a bien voulu examiner ce travail.
- Monsieur DJABRI Larbi, Professeur à l'Université d'Annaba, qui a bien voulu examiner ce travail.
- Madame GUERGAZI Saadia, Professeur à l'Université de Biskra, qui a bien voulu examiner ce travail.

Je tiens à présenter mes remerciements au personnel de la station de traitement des eaux usées municipales de la ville de Khenchela, particulièrement Mr. **BELAIDI Hichem** et Mme **MESSERA Akila** pour leurs disponibilités et leurs aides pour l'accès aux données d'exploitation.

Je tiens à témoigner ma gratitude à Mr. Marco Von Sperling (Professeur à l'Université fédérale de Minas Gerais. Brésil) pour ces conseils et orientations.

Je ne saurais clôturer cette liste de remerciements et de reconnaissances, sans exprimer ma profonde gratitude à tous mes enseignants de primaire, moyen,

secondaire, et universitaire ceux qui ont contribué à mon éducation et ma formation à la recherche scientifique.

RESUME

La performance et la fiabilité d'une usine de traitement des eaux usées est une considération critique, particulièrement si ces eaux usées traitées seront récupérées pour une réutilisation ultérieure. Les marges de sécurité pour la santé publique et la protection de l'environnement doivent être assurées et respectées.

Les performances des processus de traitement sont généralement influencées par de nombreux facteurs tels que les changements qualitatifs et quantitatifs dans les eaux usées et la variabilité inhérente du procédé de traitement. La législation algérienne a établi des critères de qualité de l'eau traitée et rejetée, donc il devrait être possible d'évaluer la performance du processus et la fiabilité des installations pour assurer leurs conformités.

La performance de la station de traitement des eaux usées municipales de la ville de Khenchela a été évaluée en fonction de la fiabilité. Les données étudiées et analysées statistiquement incluent les débits des eaux usées et les paramètres importants de qualité de l'eau telles que la demande chimique en oxygène (DCO), la demande biochimique en oxygène après cinq jours (DBO_5), et les matières en suspension (MES).

Les résultats des tests d'adéquation (Kolmogorov-Smirnov, test d'Anderson-Darling, le test de Cramer-von Mises) ont indiqué que, parmi plusieurs modèles de distribution étudiés, les relevés quotidiens des paramètres de qualité de l'eau traitée, la distribution log-normale étaient appropriés pour les données d'auto-surveillances, et la plus adéquate pour décrire les données recueillies.

L'utilisation de la fonction log-normal a permis le développement d'un modèle de fiabilité probabiliste pour l'évaluation des performances du procédé de traitement. Une caractéristique

importante de ce modèle est que les paramètres du modèle sont basés sur les propriétés des données d'auto-surveillance de la station d'épuration (moyenne arithmétique et écart type).

Le développement de ce modèle de fiabilité est la première contribution de cette étude, un modèle qui peut fournir une représentation quantitative des performances de la station d'épuration des eaux usées étudié, et peut également être utilisé pour estimer les valeurs moyennes de qualité des eaux usées traitées.

Afin d'élaborer une stratégie de contrôle du processus de boues activées dans une station de traitement des eaux usées, la compréhension des variations des niveaux de fiabilité à la station d'épuration est nécessaire.

Les systèmes biologiques sont parmi les plus difficiles à contrôler et à prévoir. En raison des mécanismes de réaction biologique complexes, hautement variables dans le temps et les aspects multi variables de la station de traitement des eaux usées (STEU), le diagnostic de la station d'épuration sont encore difficiles dans la pratique.

L'application des techniques intelligentes, qui peut analyser les données de processus non-linéaire multidimensionnelle en utilisant une technique de visualisation, peut être utile pour l'analyse et le diagnostic du procédé à boues activées dans une station d'épuration. Cette capacité complexe de la représentation de la non-linéarité combiné avec le fait qu'il n'existe pas de modèle déterministe de la fiabilité pour les stations d'épurations. Les réseaux de neurones artificiels sont un choix idéal comme solution.

La prévision du comportement des systèmes complexes a été un vaste domaine d'application pour les réseaux de neurones artificiels. Des applications telles que la prévision économique, la charge d'électricité / prévision de la demande, et la prévision des phénomènes naturels et physiques ont été largement étudiés, d'où les nombreux articles présentés lors des conférences annuelles dans ce domaine d'intervention. La capacité cognitive des réseaux de neurones artificiels pour la cartographie des entrées-sortie complexes et les relations non linéaires, ce qui permettrait une meilleure prévision du processus et de contrôle, les rendent particulièrement attrayant.

La deuxième contribution de ce travail de recherche est le développement de modèles de réseaux de neurones artificiels pour la prédiction des taux de fiabilité et taux de défaillance basée sur les données d'auto-surveillance de la station d'épuration, et un modèle probabiliste spécialement développé pour cette étude .

Six différents réseaux de neurones artificiels répartis en deux types ; un modèle simple RNA et un modèle complet RNA, basé sur un Perceptron Multicouche (PMC) sont développés pour la prédiction du niveau de la fiabilité, et les taux de défaillance du processus basée sur les paramètres de qualité des eaux usées traitées suivants : la demande chimique en oxygène (DCO), la demande biochimique en oxygène après cinq jours(DBO_5), et les matières en suspension (MES).

Le domaine d'application de ces modèles proposés est la prédiction du niveau de fiabilité d'une station de traitement des eaux usées. Le résultat de la prédiction est utilisé pour la détermination les performances du processus, de manière à optimiser le rendement de la station.

Les résultats devrait permettre de fournir des informations utiles sur la portée et les possibilités d'application des réseaux de neurones artificiels dans le domaine du traitement des eaux usées et de servir comme un outil de diagnostic des défaillances du processus et en particulier d'aider les exploitant dans la gestion quotidienne de la station d'épuration.

Développer des modèles de réseaux de neurones artificiels en fonction d'un modèle probabiliste pour la prédiction de la fiabilité et les taux de défaillance d'un procédé de traitement basée sur les boues activées est la première tentative dans la littérature scientifique et de la recherche à ce jour.

RESUME EN ANGLAIS (ABSTRACT)

The performance and reliability of a wastewater treatment plant is an important consideration especially if treated wastewater will be recovered for future reuse. Safety margins for public health and environmental protection must be ensured.

The performances of treatment process are usually influenced by many factors such as qualitative and quantitative changes in waste water and the inherent variability of the process wastewater. Algerian legislation has established criteria of quality of treated water and discarded, thus it should be possible to evaluate the performance of the process and the reliability of facilities to ensure compliance.

The performance of municipal wastewater treatment plant in Khenchela was assessed based on reliability. Data studied and analyzed statistically included wastewater flow rates and some important water quality parameters such as Chemical Oxygen Demand (COD), five-day Biochemical Oxygen Demand (BOD₅), and Total Suspended Solids (TSS).

Results of goodness of fit tests (Kolmogorov-Smirnov test, Anderson-Darling test, Cramer-von Mises test) indicated that among the several distribution models investigated to fit the daily records of water quality parameters, the lognormal distribution were appropriate for the data, and the most adequate to describe collected data.

Using the lognormal function, a probabilistic reliability model was developed and found suitable to quantitatively the performance of the plant. A significant feature of this model is that the model parameters are based on properties of original data.

The development of this reliability model is the first contribution of this study; a model that can provide a quantitative performance of the studied wastewater treatment plant, and can also be used to estimate mean values of effluent quality.

In order to develop an effective control strategy for the activated sludge process of wastewater treatment plant, an understanding of the reliability disturbances to the wastewater treatment plant is necessary. Biological systems are among the most difficult to control and predict. Due to the complex biological reaction mechanisms, the highly time-varying and multivariable aspects of the wastewater treatment plant (WWTP), the diagnosis of the WWTP are still difficult in practice. The application of intelligent techniques, which can analyze the multidimensional nonlinear process data using a visualization technique, can be useful for analyzing and diagnosing the activated-sludge process in the WWTP.

This complex capability for nonlinearity representation combined with the fact that no reliability model exists for the WWTP, makes neural networks an ideal choice for a solution.

Forecasting the behavior of complex systems has been a broad application area for neural networks. Applications such as economic forecasting, electricity load / demand forecasting, and forecasting natural and physical phenomena have been extensively studied, hence the numerous papers presented at annual conferences in this focus area. The cognitive ability of artificial neural networks to map nonlinear complex input-output relationships, which would allow for better prediction and corrective control of processes, make them particularly attractive.

The second contribution of this work presents the development of neural network models for prediction of reliability and process rate of failure based on historical plant data and a probabilistic model especially developed for this study.

Six different neural networks divided on two types; simple model and complete model, based on Multi-Layer Perceptrons (MLP) are developed for the prediction of reliability level, and process rate of failure based on the following quality parameters of treated wastewater, Chemical Oxygen Demand (COD), five-day Biochemical Oxygen Demand (BOD₅) and Total Suspended Solids (TSS).

The application area is the prediction of the reliability level at a local municipal wastewater treatment plant. The forecast result is used for the determination of the set point to control the process, in order to optimize plant performance.

The results will hopefully provide useful information about the scope and possibilities for the application of neural networks in the field of wastewater treatment and serve as a diagnostic tool of process failures and especially help operator in the daily management of the plant.

Develop artificial neural networks models based on a probabilistic model for predicting the reliability and failure rate of wastewater treatment process based on activated sludge is the first attempt in the literature and research.

RESUME EN ARABE (الملخص)

ان أداء وموثوقية محطة معالجة مياه الصرف الصحي هو أحد الاعتبارات الهامة وخاصة إذا كان سيتم استرداد مياه الصرف الصحي المعالجة لإعادة استخدامها في المستقبل مع الأخذ بعين الاعتبار ضمان الصحة العامة وحماية البيئة.

عادة ما يتأثر أداء عملية المعالجة من قبل العديد من العوامل مثل التغيرات النوعية والكمية في مياه الصرف الصحي. لقد وضع التشريع الجزائري معايير لنوعية المياه المعالجة و التخلص منها، وعليه فإنه من الممكن تقييم أداء و موثوقية محطة معالجة مياه الصرف الصحي بالاعتماد على تلك التشريعات.

جرى تقييم أداء محطة معالجة مياه الصرف الصحي لمدينة خنشلة على أساس الاعتمادية . بدراسة البيانات وتحليلها إحصائيا و قد شملت معدلات تدفق مياه الصرف الصحي وبعض الخصائص الهامة لنوعية المياه مثل ، الطلب الكيميائي للأكسجين (DCO) ، الأكسجين الحيوي الممتص بعد خمسة أيام (DBO₅) و المواد الصلبة العالقة (MES).

أشارت نتائج الاختبارات الإحصائية (إختبار كولموجوروف - سميرنوف ، و إختبار أندرسون دارلينج ، و إختبار كريمة - فون ميزس). أن من بين العديد من النماذج المدروسة التي تتناسب مع التوزيع الحقيقي للسجلات اليومية لنوعية المياه ، تبين ان التوزيع اللوغاريتمي هو الأكثر ملائمة لوصف البيانات التي تم جمعها.

باستعمال التوزيع اللوغاريتمي تم وضع نموذج احتمالي للموثوقية مناسباً لتقييم أداء المحطة. وهناك ميزة كبيرة هي أن خصائص هذا النموذج تقوم على خصائص البيانات الأصلية.

تطوير النموذج الاحتمالي للموثوقية هي المساهمة الأولى من هذه الدراسة ؛ هذا الأخير يمكن ان يوفر الأداء الكمي لمحطة معالجة مياه الصرف الصحي التي تمت دراستها، و يمكن أيضا أن يستخدم لتقدير القيم المتوسطة من نوعية النفايات السائلة، ولهذا فان وضع استراتيجية السيطرة الفعلية لعملية الحمأة المنشطة وفهم الاضطرابات الاعتمادية للمحطة أمراً ضرورياً.

النظم البيولوجية هي من بين النظم الأكثر صعوبة في السيطرة عليها و التنبؤ بها. ويرجع ذلك إلى آليات رد الفعل البيولوجية المعقدة، و جوانب متعددة المتغيرات في محطة معالجة مياه الصرف الصحي ، لا تزال صعبة في الممارسة في اوقات متفاوتة .

ان تطبيق تقنيات ذكية، والتي يمكنها تحليل بيانات عملية متعددة الأبعاد غير خطية باستخدام تقنية رسم التفاصيل، التي يمكن أن تكون مفيدة لتحليل و تشخيص عملية تنشيط الحمأة في محطات معالجة مياه

الصرف الصحي. هذه القدرة معقدة للتمثيل استقامة جنبا إلى جنب مع حقيقة أنه لا يوجد نموذج الموثوقية لهذه المحطات معالجة مما يجعل الشبكات العصبية خيارا مثاليا للتوصل إلى حل.

لقد تم التنبؤ بسلوك الأنظمة المعقدة بتوسيع مجال تطبيق الشبكات العصبية الاصطناعية مثل التنبؤ الاقتصادي ، تحميل الكهرباء / التنبؤ بالطلب، والتنبؤ بالظواهر الطبيعية و المادية ، وبالتالي فإن العديد من الأوراق التي قدمت في المؤتمرات السنوية في هذا المجال تركز على القدرة المعرفية للشبكات العصبية الاصطناعية لرسم خريطة العلاقات بين المدخلات و المخرجات غير الخطية المعقدة ، التي من شأنها أن تسمح بالتنبؤ ومراقبة أفضل للعمليات التصحيحية ، وجعلها جذابة بشكل خاص .

المساهمة الثانية من هذا العمل تتلخص في تطوير نماذج للشبكات العصبية الاصطناعية للتنبؤ بمعدلات الموثوقية و معدلات فشل عملية المعالجة استنادا إلى البيانات التاريخية للمحطة و النموذج الاحتمالي الذي وضع خصيصا لهذه الدراسة.

تم تطوير ستة شبكات عصبية مختلفة مقسمة على نوعين ؛ نموذج بسيط و نموذج كامل ، مبنية على شبكة ذات الطبقات المتعددة الأمامية (MLP) للتنبؤ بمستوى الموثوقية، و معدل فشل عملية المعالجة استنادا على معايير جودة مياه الصرف الصحي المعالجة ، والمتمثلة في طلب الأكسجين الكيميائي (DCO)، الأكسجين الحيوي الممتص بعد خمسة أيام (DBO₅) و المواد الصلبة العالقة (MES).

تم استخدام نتيجة التوقعات بمستوى الموثوقية من أجل تحسين أداء المحطة . وعليه فالنتائج تقدم معلومات مفيدة حول نطاق وإمكانات تطبيق الشبكات العصبية في مجال معالجة مياه الصرف الصحي ، التي تكون بمثابة أداة تشخيصية لإخفاقات عملية معالجة مياه الصرف الصحي و ادوات مساعدة في الإدارة اليومية لمحطة معالجة المياه لمدينة خنشلة.

تطوير نماذج الشبكات العصبية الاصطناعية على أساس نموذج احتمالي لتوقع الموثوقية ومعدل فشل عملية معالجة مياه الصرف الصحي على أساس الحمأة المنشطة هي المحاولة الأولى في المؤلفات والبحوث العلمية.

SOMMAIRE

DEDICACE	II
REMERCIEMENTS	III
RESUME	V
RESUME EN ANGLAIS (ABSTRACT)	VIII
RESUME EN ARABE (الملخص)	X
INTRODUCTION GENERALE	1
CHAPITRE I: LES PROCEDES DE TRAITEMENT DES EAUX USEES	5
I. 1 INTRODUCTION	5
I.2 PRETRAITEMENTS	5
I.2.1 DEGRILLAGE	5
I.2.2 POSTE DE RELEVAGE	6
I.2.3 DESSABLAGE	7
I.2.3.1 Dessableur aéré	7
I.2.3.2 Durée de détention.	8
I.2.3.3 Dégraissage déshuillage	8
I.3 TRAITEMENTS PRIMAIRES	9
I.3.1 DECANTATION PRIMAIRE SIMPLE SANS REACTIF	9
I.3.2 ELIMINATION DE LA POLLUTION COLLOÏDALE.....	10
I.4 TRAITEMENTS SECONDAIRES (TRAITEMENT BIOLOGIQUE)	10
I.4.1 TRAITEMENTS ANAEROBIES	11
I.4.2 TRAITEMENTS AEROBIES	11
I.4.2.1 Cultures fixes (lits bactériens et disques biologiques)	11
I.4.2.2 Les cultures libres (boues activées).....	12
I.4.2.3 Le lagunage.....	13
I.4.2.4 Filtration/percolation.....	13
I.4.2.5 Filtres plantés.....	14
I.5 TRAITEMENTS TERTIAIRES	14
I.5.1 ELIMINATION BIOLOGIQUE DE L'AZOTE	14
I.5.2 ELIMINATION BIOLOGIQUE DU PHOSPHORE	15
I.5.3 ELIMINATION ET TRAITEMENT DES ODEURS	16
I.5.4 PROCEDES DE DESINFECTIONS.....	16
I.6 TRAITEMENT DES BOUES	17
I.7 CONCLUSION	18

CHAPITRE II: DESCRIPTION GENERALE DE LA STATION D'EPURATION DE LA VILLE DE KHENCHELA.....	19
II.1 LOCALISATION DE LA STATION D'EPURATION.....	19
II.2 DONNEES SPECIFIQUES DE LA STATION.....	19
II.3 PRESENTATION DE LA FILIERE DE TRAITEMENT DE LA STATION	20
II.3.1 FILIERE DE TRAITEMENT DES EAUX	20
II.3.2 FILIERE DES BOUES	20
II. 4 CARACTERISTIQUES TECHNIQUES DES OUVRAGES DE LA STATION.....	21
II.4.1 CARACTERISTIQUES DES OUVRAGES EXISTANTS.....	21
CHAPITRE III: ETUDE DE LA FIABILITÉ D'UNE STATION DE TRAITEMENT DES EAUX USEES	25
III. 1 INTRODUCTION	25
III. 2 NOTION DE FIABILITE.....	25
III.2.1 DEFINITION DE LA NOTION DE FIABILITE	25
III.3 METHODE DES COURBES DE DEFAILLANCE.....	26
III.3.1 LES DIFFERENTES FORMES DE COURBES DE DEFAILLANCE.....	27
III.3.2 ANALYSE DE LA COURBE DE DEFAILLANCE D'UN GROUPE DE SYSTEME.....	28
III.3.3 DETERMINATION DE LA FORME DE LA COURBE DE DEFAILLANCE.....	29
III.4 METHODE DU COEFFICIENT DE FIABILITE (C.D.F)	31
III.4.1 CONCEPT DE FIABILITE DANS UNE STATION DE TRAITEMENT DES EAUX USEES.....	31
III.4.2 EVALUATION DE LA FIABILITE D'UNE STATION DE TRAITEMENT DES EAUX USEES PAR LA METHODE DU COEFFICIENT DE FIABILITE (C.D.F).....	32
III.5 FACTEURS AFFECTANT LES PERFORMANCES D'UNE STATION DE TRAITEMENT DES EAUX USEES.....	36
III.6 NORMES DE REJET DES EAUX USEES EN ALGERIE.....	37
III.7 EVALUATION DE LA FIABILITE D'UNE STATION D'EPURATION EN EMPLOYANT D'AUTRES APPROCHES	38
III.6 CONCLUSION.....	38
CHAPITRE IV : EVALUATION DE LA FIABILITE DE LA STATION D'EPURATION DE KHENCHELA PAR MODELE PROBABILISTE	39
IV.1 ORGANISATION DES DONNEES.....	39
IV.1.1 ACQUISITION DES DONNEES	39
IV.1.2 TRAITEMENT PRIMAIRE DES DONNEES.....	40
IV.1.3 CONTROLE DES DONNEES.....	40
IV.1.4 MODELE NUMERISE DES DONNEES	40
IV.2 RECHERCHE DES ERREURS DANS LES DONNEES D'AUTO-SURVEILLANCES.....	41

IV.3 CORRECTION DES DONNES D'AUTO-SURVEILLANCES.....	41
IV.4 TRAITEMENT STATISTIQUE DES DONNES DE LA STATION D'EPURATION DE KHENCHELA.....	43
IV.4.1 ANALYSE STATISTIQUE DES DONNEES DE LA STATION DE KHENCHELA.....	44
IV.4.2 ETUDE DE LA VARIABILITE DES PARAMETRES DE POLLUTION.....	46
IV.5 CHOIX DE LA LOI DE DISTRIBUTION APPROPRIEE.....	48
IV.5.1 ANALYSE DES DONNEES A L'ENTREE DE LA STATION.....	50
IV.5.2 ANALYSE DES DONNEES A LA SORTIE DE LA STATION D'EPURATION.....	56
IV.6 EVALUATION DE LA FIABILITE GLOBALE DE LA STATION DE KHENCHELA.....	62
IV.6.1 EVALUATION DE LA FIABILITE GLOBALE DE L'ELIMINATION DES MES.....	62
IV.6.2 EVALUATION DE LA FIABILITE GLOBALE DE L'ELIMINATION DE LA DCO.....	63
IV.6.3 EVALUATION DE LA FIABILITE GLOBALE DE LA DEGRADATION DE LA DBO ₅	64
IV.7 DETERMINATION DU TAUX DE FIABILITE PAR MODELE PROBABILISTE.....	65
IV.7.1 EVALUATION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO ₅	67
IV.7.2 EVALUATION DU TAUX DE FIABILITE DE L'ELIMINATION DE LA DCO.....	68
IV.7.3 EVALUATION DU TAUX DE FIABILITE DE L'ELIMINATION DES MES.....	68
IV.8 DETERMINATION DES TAUX DE DEFAILLANCES.....	69
IV.8.1 EVALUATION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅	70
IV.8.2 DETERMINATION DU TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO.....	70
IV.8.3 DETERMINATION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES.....	71
IV.9 CONCLUSION.....	72
CHAPITRE V : RESEAUX DE NEURONES ARTIFICIELS.....	74
V.1 INTRODUCTION.....	74
V.2 MODELE DE NEURONE.....	74
V.2.1. NEURONE BIOLOGIQUE.....	74
V.2.2. NEURONE FORMEL (ARTIFICIEL).....	75
V.2.3 RESEAU DE NEURONES ARTIFICIELS.....	76
V.3 HISTORIQUE DES RECHERCHES SUR LES RESEAUX DE NEURONES.....	77
V.4 ARCHITECTURE DES RESEAUX DE NEURONES ARTIFICIELS.....	79
V.4.1 RESEAU DE NEURONES ARTIFICIELS DIRECT (A ALIMENTATION AVANT).....	79
V.4.2 RESEAU DE NEURONES ARTIFICIELS RECURRENT.....	80
V.4.3 FONCTION DE TRANSFERT.....	81
V.5 ENTRAINEMENT ET APPRENTISSAGE DU RESEAU DE NEURONES.....	83
V.5.1 APPRENTISSAGE SUPERVISE.....	83
V.5.2 APPRENTISSAGE NON SUPERVISE.....	83
V.5.3 APPRENTISSAGE PAR RENFORCEMENT.....	83
V.5.4 REGLE D'APPRENTISSAGE DU PERCEPTRON.....	84
V.6 ALGORITHME D'APPRENTISSAGE.....	86

V.6.1 ALGORITHME D'APPRENTISSAGE DE RETRO PROPAGATION (BP)	86
V.6.2 LIMITES DE L'ALGORITHME D'APPRENTISSAGE RETRO PROPAGATION	87
V.6.3 AMELIORATION HEURISTIQUE POUR L'ALGORITHME DE RETRO PROPAGATION	88
V.7 GENERALISATION	89
V.7.1 SUR-APPRENTISSAGE ET SOUS-APPRENTISSAGE	90
V.7.2 ARRET PREMATURE	91
V.7.3 REGULARISATION	92
V.8 AVANTAGES DES RESEAUX DE NEURONES ARTIFICIELS.....	94
V.9. STRUCTURES DE RESEAUX DE NEURONES ARTIFICIELS POUR LA PREDICTION	94
V.10 REVUE DE LITTERATURE	96
V.11. CONCLUSION.....	99
CHAPITRE VI : PREDICTION DU TAUX DE FIABILITE PAR RESEAUX DE NEURONES ARTIFICIELS (CAS D'ETUDE : STATION D'EPURATION DE KHENCHELA).....	100
VI.1 INTRODUCTION	100
VI.2 PRESENTATION DES CARACTERISTIQUES DU LOGICIEL	100
VI.2.1 METHODE DE CONSTRUCTION DES MODELES RNA	100
VI.2.2 SELECTIONS DES VARIABLES D'ENTREE DU RESEAU DE NEURONES ARTIFICIELS (RNA)	101
VI.2.3 TAILLE DE LA BASE DES DONNEES	102
VI.2.4 PARTAGE DE L'ENSEMBLE DES DONNEES	103
VI.2.5 FONCTION D'ACTIVATION	103
VI.2.6 CRITERE DE L'ARRET D'ENTRAINEMENT	104
VI.2.7 MESURE DES PERFORMANCES DE PREDICTION.....	104
VI.3 STRATEGIE MISE EN OEUVRE POUR LA CONSTRUCTION DES MODELES RNA	105
VI.3.1 MODELE SIMPLE DE RESEAU DE NEURONES ARTIFICIELS	105
VI.3.2 MODELE COMPLET DE RESEAU DE NEURONES ARTIFICIELS	106
VI.4 PREDICTION DU TAUX DE FAIBILITE DES PARAMETRES DE QUALITE DES EAUX TRAITEES PAR MODELE SIMPLE	109
VI.4.1 PREDICTION DE LA FIABILITE DE LA DEGRADATION DE LA DBO5 PAR MODELE SIMPLE	109
VI.4.2 PREDICTION DE LA FIABILITE D'ELIMINATION DE LA DCO PAR MODELE SIMPLE.....	111
VI.4.3 PREDICTION DE LA FIABILITE D'ELIMINATION DES MES PAR MODELE SIMPLE	114
VI.5 PREDICTION DU TAUX DE FAIBILITE DES PARAMETRES DE QUALITE DES EAUX TRAITEES PAR MODELE COMPLET	117
VI.5.1 PREDICTION DE LA FIABILITE DE LA DEGRADATION DE LA DBO5 PAR MODELE COMPLET.....	117
VI.5.2 PREDICTION DE LA FIABILITE D'ELIMINATION DE LA DCO PAR MODELE COMPLET	120
VI.5.3 PREDICTION DE LA FIABILITE D'ELIMINATION DES MES PAR MODELE COMPLET.....	123
VI.6 DISCUSSION DES RESULTATS DE LA PREDICTION DES TAUX DE FIABILITE	128
VI.6.1 RESULTATS DES MODELES RNA _{DBO5}	129
VI.6.1.1 Prédiction du taux de fiabilité par modèle simple MS-RNA _{DBO5} (4-3-1)	129
VI.6.1.2 Prédiction du taux de fiabilité par modèle complet MC-RNA _{DBO5} (14-3-1)	129
VI.6.2 RESULTATS DES MODELES RNA _{DCO}	130

VI.6.2.1 Prédiction du taux de fiabilité par modèle simple MS-RNA _{DCO} (4-2-1)	130
VI.6.2.2 Prédiction du taux de fiabilité par modèle complet MC-RNA _{DCO} (14-8-1).....	131
VI.6.3 RESULTATS DES MODELES RNA _{MES}	131
VI.6.3.1 Prédiction du taux de fiabilité par modèle simple MS-RNA _{MES1} (4-4-1).....	131
VI.6.3.2 Prédiction du taux de fiabilité par modèles complets MC-RNA _{MES1,2} (10-3-1 ; 10-6-1)	132
VI.7. CONCLUSION	133
CHAPITRE VII : PREDICTION DU TAUX DE DEFAILLANCES PAR LES RESEAUX DE NEURONES ARTIFICIELS (CAS D'ETUDE : STATION D'EPURATION DE KHENCHELA).....	134
VII.1 INTRODUCTION.....	134
VII.2 CONSTRUCTION DES MODELES RNA	134
VII.2.1 MODELE SIMPLE RNA UTILISEES POUR LA PREDICTION DU TAUX DE DEFAILLANCE	134
VII.2.2 MODELES COMPLETS RNA UTILISES POUR LA PREDICTION DU TAUX DE DEFAILLANCE.....	135
VII.3 PREDICTION DU TAUX DE DEFAILLANCE DU PROCEDE DE TRAITEMENT	136
VII.3.1 PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO5 PAR MODELE SIMPLE	136
VII.3.2 PREDICTION DU TAUX DE DEFAILLANCE DE L'ELIMINATION DE LA DCO PAR MODELE SIMPLE	139
VII.2.3 PREDICTION DU TAUX DE DEFAILLANCE DE L'ELIMINATION DES MES PAR MODELE SIMPLE.....	142
VII.3 PREDICTION DU TAUX DE DEFAILLANCE DU PROCEDE DE TRAITEMENT DES EAUX USEES PAR MODELE COMPLET	145
VII.3.1 PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO5 PAR MODELE COMPLET..	145
VII.3.2 PREDICTION DU TAUX DE DEFAILLANCE DE L'ELIMINATION DE LA DCO PAR MODELE COMPLET	148
VII.3.3 PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES PAR MODELE COMPLET.....	151
VII.4 DISCUSSION DES RESULTATS DE LA PREDICTION DES TAUX DE DEFAILLANCE	154
VII.4.1 RESULTATS DES MODELES RNA _{DBO5}	155
VII.4.1.1 Prédiction du taux de défaillance par modèle simple MS-RNA _{DBO5} (5-2-1).....	155
VII.4.1.2 Prédiction du taux de défaillance par modèle complet MC-RNA _{DBO5} (10-4-1).....	155
VII.4.2 RESULTATS DES MODELES RNA _{DCO}	156
VII.4.2.1 Prédiction du taux de défaillance par modèle simple MS-RNA _{DCO} (5-8-1).....	156
VII.4.2.2 Prédiction du taux de défaillance par modèle complet MC-RNA _{DCO} (10-4-1).....	157
VII.4.3 RESULTATS DES MODELES RNA _{MES}	157
VII.4.3.1 Prédiction du taux de défaillance par modèle simple MS-RNA _{MES} (4-3-1)	157
VII.4.3.2 Prédiction du taux de défaillance par modèle complet MC-RNA _{MES} (14-5-1)	158
VII.5 CONCLUSION.....	159
CONCLUSION GENERALE ET ORIENTATIONS FUTURES DE LA RECHERCHE.....	160
REFERENCES BIBLIOGRAPHIQUES	165

LISTE DES FIGURES

FIGURE 1: APERÇU GENERAL DE LA STATION D'EPURATION DE LA VILLE DE KHENCHELA.	20
FIGURE 2: SCHEMA GENERAL DE LA FILIERE DE TRAITEMENT DES EAUX USEES DE LA STATION DE KHENCHELA... ..	24
FIGURE 3: LES DIFFERENTS COURBES DE DEFAILLANCES (MOUBRAY, 1997)	27
FIGURE 4: PROBABILITE CONDITIONNELLE DE DEFAILLANCE EN FONCTION DU TEMPS	30
FIGURE 5: DIAGRAMME BOX-WHISKER DES DIFFERENTS PARAMETRES DES EAUX USEES A L'ENTREE DE LA STATION DE KHENCHELA (2009-2012)	47
FIGURE 6: DIAGRAMME BOX-WHISKER DES DIFFERENTS PARAMETRES DES EAUX TRAITES A LA SORTIE DE LA STATION DE KHENCHELA	48
FIGURE 7 : COURBE DE PROBABILITE DE LA DBO ₅ A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT UNE LOI DE DISTRIBUTION GAMMA.....	52
FIGURE 8: COURBE DE PROBABILITE DE LA DBO ₅ A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT UNE LOI DE DISTRIBUTION LOG-NORMALE.....	52
FIGURE 9: COURBE DE PROBABILITE DE LA DCO A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT UNE LOI DE DISTRIBUTION NORMALE.	53
FIGURE 10: COURBE DE PROBABILITE DE LA DCO A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT UNE LOI DE DISTRIBUTION GAMMA.....	53
FIGURE 11: COURBE DE LA FONCTION CUMULATIVE DE DENSITE DE LA DCO A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT LES LOIS DE DISTRIBUTION NORMALE ET GAMMA.....	54
FIGURE 12: HISTOGRAMME DE DISTRIBUTION DES MES A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT LES LOIS DE DISTRIBUTION LOG-NORMALE ET GAMMA.....	55
FIGURE 13: COURBE DES QUANTILES DES MES A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT LES LOIS DE DISTRIBUTIONS LOG-NORMAL ET GAMMA.....	55
FIGURE 14 : HISTOGRAMME DE LA DBO ₅ A LA SORTIE DE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DISTRIBUTION.....	57
FIGURE 15: COURBE DES QUANTILES DE LA DBO ₅ A LA SORTIE DE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DEDISTRIBUTION.....	57
FIGURE 16: HISTOGRAMME DE LA DCO A LA SORTIE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DISTRIBUTION.....	58
FIGURE 17: COURBE DES QUANTILES DE LA DCO A LA SORTIE DE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DISTRIBUTION.....	58
FIGURE 18: CAPTURE ECRAN LORS DU TRAITEMENT STATISTIQUE DES DONNEES DE LA STATION DE KHENCHELA	59
FIGURE 19: HISTOGRAMME DE DISTRIBUTION DES MES A LA SORTIE DE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DISTRIBUTION.....	60
FIGURE 20: COURBE DES QUANTILES VERSUS LA PROBABILITE CUMULATIVE DES MES SUIVANT DIFFERENTS LOIS DE DISTRIBUTIONS.....	60
FIGURE 21: COURBE DE PROBABILITE DES MES A LA SORTIE DE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DISTRIBUTION.....	61
FIGURE 22: COURBE DE CONCENTRATION DES MES A LA SORTIE DE LA STATION DE KHENCHELA (2009-2012)... ..	62
FIGURE 23: COURBE DE CONCENTRATION DE LA DCO A LA SORTIE DE LA STATION DE.....	63
FIGURE 24: COURBE DE CONCENTRATION DE LA DBO ₅ A LA SORTIE DE LA STATION DE.....	64
FIGURE 25: COURBE DE LA VARIATION DU TAUX DE FIABILITE DU PROCEDE DE DEGRADATION DE LA DBO ₅ DANS LA STATION DE KHENCHELA (2009-2012).....	67
FIGURE 26: COURBE DE LA VARIATION DU TAUX DE FIABILITE DE L'ELIMINATION DE LA DCO DANS LA STATION D'EPURATION DE KHENCHELA (2009-2012).	68
FIGURE 27: COURBE DE LA VARIATION DU TAUX DE FIABILITE DU PROCEDE D'ELIMINATION DES MES DANS LA STATION D'EPURATION DE KHENCHELA (2009-2012).....	69
FIGURE 28: COURBE DE LA VARIATION DU TAUX DE DEFAILLANCE DU PROCEDE DE TRAITEMENT DE LA DBO ₅ DANS LA STATION D'EPURATION DE KHENCHELA (2009-2012).....	70

FIGURE 29: COURBE DE LA VARIATION DU TAUX DE DEFIALANCE DU PROCEDE DE TRAITEMENT DE LA DCO DANS LA STATION D'EPURATION DE KHENCHELA (2009-2012).....	71
FIGURE 30: COURBE DE LA VARIATION DU TAUX DE DEFIALANCE DU PROCEDE D'ELIMINATION DES MES DANS LA STATION D'EPURATION DE KHENCHELA (2009-2012).....	72
FIGURE 31: STRUCTURE D'UN NEURONE BIOLOGIQUE (GRAUPE, 2007).....	74
FIGURE 32: NEURONE FORMEL (ARTIFICIEL).....	76
FIGURE 33: RESEAU DE NEURONES MULTICOUCHES (PMC) AVEC DEUX COUCHES CACHEES (RUSTUM, 2009).....	77
FIGURE 34: STRUCTURE D'UN NEURONE ARTIFICIEL QUI CALCULE LA SOMME DE SES ENTREES PUIS CETTE VALEUR PASSE A TRAVERS LA FONCTION D'ACTIVATION POUR PRODUIRE SA SORTIE (KRIGER, 2007).....	77
FIGURE 35: RESEAU DE NEURONES ARTIFICIELS DIRECTS (PAR ALIMENTATION AVANT) (FEED-FORWARD) (KRIGER, 2007).....	79
FIGURE 36: RESEAU DE NEURONES RECURRENT (KRIGER, 2007).....	80
FIGURE 37: ALGORITHME D'APPRENTISSAGE COINCE DANS UN MINIMUM LOCAL	88
FIGURE 38 : PRESENTATION GRAPHIQUE DU PHENOMENE DE SUR-APPRENTISSAGE.....	90
FIGURE 39 : EXEMPLE DU PHENOMENE DE SOUS-APPRENTISSAGE (PARIZEAU, 2006).	91
FIGURE 40 : ILLUSTRATION DU CRITERE DE LA VALIDATION CROISEE (PARIZEAU, 2006).....	92
FIGURE 41 : RESEAU DE NEURONES ARTIFICIELS D'ELMAN	95
FIGURE 42 : RESEAU DE NEURONES ARTIFICIELS POUR DES APPLICATIONS DE PREDICTION AVEC DES ENTREES TEMPORISES	96
FIGURE 43: PRINCIPE GENERAL DES ALGORITHMES GENETIQUES (ALLIOT ET DURAND, 2005).....	102
FIGURE 44: SCHEMA GENERALE DU MODELE SIMPLE RNA_{DBO_5} (4-3-1) POUR LA PREDICTION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO_5	109
FIGURE 45: TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO_5 PREDITE PAR MODELE SIMPLE $MS-RNA_{DBO_5}$ (4-3-1) VERSUS FIABILITE PAR MODELE PROBABILISTE (VALEUR CIBLE).	110
FIGURE 46: TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO_5 PREDIT PAR RESEAU DE NEURONES ARTIFICIELS SIMPLE (4-3-1), STATION D'EPURATION DE KHENCHELA (2009-2012).	111
FIGURE 47: SCHEMA GENERALE DU MODELE SIMPLE $MS-RNA_{DCO}$ (4-2-1) POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DE LA DCO.	112
FIGURE 48: TAUX DE FIABILITE D'ELIMINATION DE LA DCO PREDIT PAR MODELE SIMPLE $MS-RNA_{DCO}$	113
FIGURE 49: TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO_5 PREDIT PAR RESEAU DE NEURONES ARTIFICIELS SIMPLE (4-2-1), STATION D'EPURATION DE KHENCHELA (2009-2012)	114
FIGURE 50: SCHEMA GENERALE DU MODELE SIMPLE $MS-RNA_{MES}$ (4-4-1) POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES.....	115
FIGURE 51: TAUX DE FIABILITE D'ELIMINATION DES MES PREDIT PAR MODELE SIMPLE $MS-RNA_{MES}$	116
FIGURE 52: TAUX DE FIABILITE D'ELIMINATION DES MES PREDIT PAR RESEAU DE NEURONES ARTIFICIELS SIMPLE (4-4-1), STATION D'EPURATION DE KHENCHELA (2009-2012)	117
FIGURE 53 : SCHEMA GENERALE DU MODELE COMPLET $MC-RNA_{DBO_5}$ (14-3-1) POUR LA PREDICTION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO_5	118
FIGURE 54: FIABILITE DE LA DEGRADATION DE LA DBO_5 PREDITE PAR MODELE COMPLET $MC-RNA_{DBO_5}$	119
FIGURE 55: TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO_5 PREDIT PAR RESEAU DE NEURONES ARTIFICIELS COMPLET (14-3-1), STATION D'EPURATION DE KHENCHELA (2009-2012).....	120
FIGURE 56: SCHEMA GENERALE DU MODELE COMPLET $MC-RNA_{DCO}$ (14-8-1) POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DE LA DCO.	121
FIGURE 57: TAUX DE FIABILITE D'ELIMINATION DE LA DCO PREDIT PAR MODELE COMPLET $MC-RNA_{DCO}$	122
FIGURE 58: TAUX DE FIABILITE D'ELIMINATION DE LA DCO PREDIT PAR RESEAU DE NEURONES ARTIFICIELS COMPLET (14-8-1), STATION D'EPURATION DE KHENCHELA (2009-2012).....	123
FIGURE 59: SCHEMA GENERALE DU MODELE COMPLET $MC-RNA_{MES1}$ (10-3-1) POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES.....	124
FIGURE 60: TAUX DE FIABILITE D'ELIMINATION DES MES PREDIT PAR MODELE COMPLET RNA_{MES1} (10-3-1) VERSUS FIABILITE PAR MODELE PROBABILISTE (VALEUR CIBLE).	125
FIGURE 61: TAUX DE FIABILITE D'ELIMINATION DES MES PREDIT PAR RESEAU DE NEURONES ARTIFICIELS COMPLET (10-3-1), STATION D'EPURATION DE KHENCHELA (2009-2010).....	126

FIGURE 62: SCHEMA GENERALE DU MODELE COMPLET MC-RNA _{MES2} (10-6-1) POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES.....	126
FIGURE 63: TAUX DE FIABILITE D'ELIMINATION DES MES PREDIT PAR MODELE COMPLET RNA _{MES2} (10-6-1) VERSUS TAUX DE FIABILITE PAR MODELE PROBABILISTE (VALEUR CIBLE).....	127
FIGURE 64: TAUX DE FIABILITE D'ELIMINATION DES MES PREDIT PAR RESEAU DE NEURONES ARTIFICIELS PAR MODELE COMPLET RNA _{MES2} (10-6-1), STATION D'EPURATION DE KHENCHELA (2011-2012).....	128
FIGURE 65: SCHEMA GENERALE DU MODELE SIMPLE MS-RNA _{DBO5} (5-2-1) POUR LA PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅	137
FIGURE 66 : TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅ PREDIT PAR MODELE SIMPLE MS-RNA _{DBO5} (5-2-1) VERSUS TAUX DE DEFAILLANCE PAR MODELE PROBABILISTE (VALEUR CIBLE).	138
FIGURE 67 : TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅ PREDIT PAR RESEAU DE NEURONES ARTIFICIELS SIMPLE (5-2-1), STATION D'EPURATION DE KHENCHELA (2009-2012).....	139
FIGURE 68 : SCHEMA GENERALE DU MODELE SIMPLE MS-RNA _{DCO} (5-8-1) POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO.	140
FIGURE 69 : TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO PREDIT PAR MODELE SIMPLE MS-RNA _{DCO}	141
FIGURE 70 : TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO PREDIT PAR RESEAU DE NEURONES ARTIFICIELS SIMPLE (5-8-1), STATION D'EPURATION DE KHENCHELA (2009-2012).	142
FIGURE 71: SCHEMA GENERALE DU MODELE SIMPLE MS-RNAMES (4-3-1) POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES.....	143
FIGURE 72 : TAUX DE DEFAILLANCE D'ELIMINATION DES MES PREDIT PAR MODELE SIMPLE MS-RNA _{MES} (4-3-1) VERSUS TAUX DE DEFAILLANCE PAR MODELE PROBABILISTE (VALEUR CIBLE).	144
FIGURE 73 : TAUX DE DEFAILLANCE D'ELIMINATION DES MES PREDIT PAR RESEAU DE NEURONES ARTIFICIELS SIMPLE (4-3-1), STATION D'EPURATION DE KHENCHELA (2009-2012).	145
FIGURE 74 : SCHEMA GENERALE DU MODELE COMPLET MC-RNA _{DBO5} (10-2-1) POUR LA PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅	146
FIGURE 75 : TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅ PREDIT PAR MODELE COMPLET MC-RNAD _{BO5} (10-2-1) VERSUS TAUX DE DEFAILLANCE PAR MODELE PROBABILISTE (VALEUR CIBLE).	147
FIGURE 76 : TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅ PREDIT PAR RESEAU DE NEURONES ARTIFICIELS COMPLET (10-2-1), STATION D'EPURATION DE KHENCHELA (2009-2012).	148
FIGURE 77 : SCHEMA GENERALE DU MODELE COMPLET MC-RNA _{DCO} (10-4-1) POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO.	149
FIGURE 78 : TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO PREDIT PAR MODELE COMPLET MC-RNA _{DCO} (10-4-1) VERSUS TAUX DE DEFAILLANCE PAR MODELE PROBABILISTE (VALEUR CIBLE).....	150
FIGURE 79 : TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO PREDIT PAR RESEAU DE NEURONES ARTIFICIELS COMPLET (10-4-1), STATION D'EPURATION DE KHENCHELA (2009-2012).....	151
FIGURE 80 : SCHEMA GENERALE DU MODELE COMPLET MC-RNAMES (14-5-1) POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES.....	152
FIGURE 81 : TAUX DE DEFAILLANCE D'ELIMINATION DES MES PREDIT PAR MODELE COMPLET MC-RNA _{MES} (14-5-1) VERSUS TAUX DE DEFAILLANCE PAR MODELE PROBABILISTE (VALEUR CIBLE).....	153
FIGURE 82 : TAUX DE DEFAILLANCE D'ELIMINATION DES MES PREDIT PAR RESEAU DE NEURONES ARTIFICIELS SIMPLE (14-5-1), STATION D'EPURATION DE KHENCHELA (2009-2012).	154

LISTE DES TABLEAUX

TABLEAU N° 1: DIFFERENTS TYPES DE DEGRILLAGE (WEF, 1998; DAUKES, 2006)	6
TABLEAU N° 2: AVANTAGES ET INCONVENIENTS DES DESSABLEURS AERES (WEF, 1998 ; SPANGLER, 2006)	8
TABLEAU 3: DONNEES SPECIFIQUES DE LA STATION D'EPURATION DE KHENCHELA (HYDRO-PROJET EST, 2005).	19
TABLEAU 4: DONNEES DE LA CHARGE POLLUANTE (HYDRO PROJET EST, 2005).	21
TABLEAU 5: CHOIX DU PROCEDE DE TRAITEMENT SELON LA QUALITE DE REJET CONFORME AUX NORMES ALGERIENNES (HYDRO PROJET EST, 2005).	21
TABLEAU 6: CARACTERISTIQUES DU DEGRILLAGEUR GROSSIER (HYDRO PROJET EST, 2005)	22
TABLEAU 7: CARACTERISTIQUES DU POSTE DE RELEVAGE DES EAUX USEES (HYDRO PROJET EST, 2005).	22
TABLEAU 8: CARACTERISTIQUES DU DEGRILLEUR FIN (HYDRO PROJET EST, 2005).	22
TABLEAU 9: CARACTERISTIQUES DU DESSABLEUR-DESHUILEUR AERE (HYDRO PROJET EST, 2005).	23
TABLEAU 10: CARACTERISTIQUES DU TRAITEMENT BIOLOGIQUE (HYDRO PROJET EST, 2005).	23
TABLEAU 11: NOMBRE DE STATIONS CONSTRUIT ET NOMBRE DE STATIONS EN SERVICE A LA FIN DE CHAQUE ANNEE D'EXPLOITATION (HUDSON, 1986).	29
TABLEAU 12: TABLEAU RECAPITULATIF DU TAUX D'ECHEC (HUDSON, 1986).	30
TABLEAU 13: VALEURS DES CENTILES D'UNE DISTRIBUTION NORMALE STANDARD (NIKU ET AL., 1981)	35
TABLEAU 14: COEFFICIENT DE FIABILITE (CDF) EN FONCTION DU COEFFICIENT DE VARIATION (CV) ET DU TAUX	35
TABLEAU 15: NORMES ALGERIENNES DES REJETS D'EAUX USEES TRAITES (JORA, 2006)	37
TABLEAU 16: SYNTHESE DE L'ANALYSE STATISTIQUE DES DONNEES DE LA STATION D'EPURATION DE KHENCHELA (2009-2012).	46
TABLEAU 17: RESULTATS DU TEST D'ADEQUATION (TEST DE KOLMOGROV-SMIRNOV) DES DIFFERENTS PARAMETRES DE POLLUTION DES EAUX USEES A L'ENTREE DE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DISTRIBUTION.	51
TABLEAU 18: RESULTATS DU TEST D'ADEQUATION (TEST DE KOLMOGROV-SMIRNOV) DES DIFFERENTS PARAMETRES DE POLLUTION DES EAUX TRAITES A LA SORTIE DE LA STATION DE KHENCHELA SUIVANT DIFFERENTS LOIS DE DISTRIBUTION.	56
TABLEAU 19: EXEMPLE DE DETERMINATION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO ₅ PAR MODELE PROBABILISTE, STATION D'EPURATION DE KHENCHELA (2012).	66
TABLEAU 20: LES PRINCIPALES FONCTIONS DE TRANSFERT	82
TABLEAU 21: DONNEES ENTREES-SORTIE DES DIFFERENTS MODELES SIMPLES RNA	106
TABLEAU 22: DONNEES ENTREES-SORTIE DES DIFFERENTS MODELES COMPLETS RNA	108
TABLEAU 23: DONNEES D'ENTREES-SORTIE DU MODELE SIMPLE MS-RNA _{DBO5} POUR LA PREDICTION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO ₅	109
TABLE 24: PERFORMANCES DE LA PREDICTION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO ₅ PAR MODELE SIMPLE MS-RNA _{DBO5}	110
TABLEAU 25: DONNEES D'ENTREES-SORTIE DU MODELE SIMPLE MS-RNA _{DCO} POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DE LA DCO.	111
TABLEAU 26: PERFORMANCES DE LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DE LA DCO PAR MODELE SIMPLE MS-RNA _{DCO}	112
TABLEAU 27: DONNEES D'ENTREES-SORTIE DU MODELE SIMPLE MS-RNA _{MES} POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES.	114
TABLEAU 28: PERFORMANCES DE LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES PAR MODELE SIMPLE MS-RNA _{MES}	115
TABLEAU 29: DONNEES D'ENTREES-SORTIE DU MODELE COMPLET MC-RNA _{DBO5} POUR LA PREDICTION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO ₅	117
TABLEAU 30: PERFORMANCES DE LA PREDICTION DU TAUX DE FIABILITE DE LA DEGRADATION DE LA DBO ₅ PAR LE MODELE COMPLET MC-RNA _{DBO5}	119
TABLEAU 31: DONNEES D'ENTREES-SORTIE DU MODELE COMPLET MC-RNA _{DCO} POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DE LA DCO.	120

TABLEAU 32: PERFORMANCES DE LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DE LA DCO LA PAR LE MODELE COMPLET MC-RNA _{DCO}	121
TABLEAU 33: DONNEES D'ENTREES-SORTIE DES MODELES COMPLETS MC-RNA _{MES1} ET MC-RNA _{MES2} POUR LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES.....	123
TABLEAU 34: PERFORMANCES DE LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES PAR MODELE COMPLET MC-RNA _{MES1}	124
TABLEAU 35: PERFORMANCES DE LA PREDICTION DU TAUX DE FIABILITE D'ELIMINATION DES MES PAR MODELE COMPLET MC-RNA _{MES2}	127
TABLEAU 36: DONNEES ENTREES-SORTIE DES DIFFERENTS MODELES SIMPLES RNA (PREDICTION DU TAUX DE DEFAILLANCE).....	135
TABLEAU 37: DONNEES ENTREES-SORTIE DES DIFFERENTS MODELES COMPLETS RNA (PREDICTION DU TAUX DE DEFAILLANCE).....	136
TABLEAU 38: DONNEES D'ENTREES-SORTIE DU MODELE SIMPLE MS-RNA _{DBO5} POUR LA PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅	137
TABLEAU 39 : PERFORMANCES DE LA PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅ PAR LE MODELE SIMPLE MS-RNA _{DBO5}	138
TABLEAU 40 : DONNEES D'ENTREES-SORTIE DU MODELE SIMPLE MS-RNA _{DCO} POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO.....	139
TABLEAU 41 : PERFORMANCES DE LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO PAR LE MODELE SIMPLE MS-RNA _{DCO}	140
TABLEAU 42 : DONNEES D'ENTREES-SORTIE DU MODELE SIMPLE MS-RNAMES POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES.....	142
TABLEAU 43 : PERFORMANCES DE LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES PAR LE MODELE SIMPLE MS-RNA _{MES}	143
TABLEAU 44 : DONNEES D'ENTREES-SORTIE DU MODELE COMPLET MC-RNA _{DBO5} POUR LA PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅	145
TABLEAU 45 : PERFORMANCES DE LA PREDICTION DU TAUX DE DEFAILLANCE DE LA DEGRADATION DE LA DBO ₅ PAR LE MODELE COMPLET MC-RNA _{DBO5}	146
TABLEAU 46 : DONNEES D'ENTREES-SORTIE DU MODELE COMPLET MC-RNA _{DCO} POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO.....	148
TABLE 47 : PERFORMANCES DE LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DE LA DCO PAR LE MODELE COMPLET MC-RNA _{DCO}	149
TABLEAU 48 : DONNEES D'ENTREES-SORTIE DU MODELE COMPLET MC-RNAMES POUR LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES.....	151
TABLEAU 49 : PERFORMANCES DE LA PREDICTION DU TAUX DE DEFAILLANCE D'ELIMINATION DES MES PAR LE MODELE COMPLET MC-RNA _{MES}	152

INTRODUCTION GENERALE

Les eaux usées représentent une des composantes de l'offre globale en eau au même titre que les eaux superficielles et souterraines. En Algérie, leur volume annuel est estimé à 800 millions de m³, dont quelque 650 millions de m³ correspondant aux agglomérations de taille supérieure à 50 000 habitants (MRE, 2014). Notons que 60 % de ces eaux sont soit rejetés dans les cours soit en mer (Medkour, 2003).

Pour ce qui est des ouvrages de traitement et d'épuration des eaux, il existe à ce jour 145 stations de traitement et d'épuration avec une capacité totale de traitement estimé à plus de 8 millions d'équivalent-habitants, avec une capacité mensuelle de collecte estimée à 66,77 millions de m³, et un volume d'eau usées épurées de 13,6 millions de m³, soit un taux d'utilisation des capacités des installations de 45% (ONA, 2014). Il faut ajouter 435 bassins de décantation qui reçoivent les eaux usées domestiques (Medkour, 2003). En raison d'un cout moyen de l'épuration de 5,94 DA/m³ incluant la réhabilitation des équipements, l'amortissement et les frais d'exploitation, l'aspect financier reste le principal frein à la remise en service des stations d'épuration à l'arrêt.

Si les eaux usées non traitées sont autorisés à s'accumuler, la décomposition de la matière organique peut conduire à la production de grandes quantités de gaz malodorants (Belanche et al., 1992). Les stations de traitement des eaux usées sont importantes pour le développement économique et social de tout pays. Ils protègent la santé de la population contre les bactéries pathogènes et des virus. Pour ces raisons, le retrait immédiat et sans nuisance des eaux usées à partir de ses sources de production , suivi d'un traitement et d'élimination , est non seulement souhaitable mais aussi nécessaire dans une société industrialisée (Metcalf et Eddy, 2003).

Les stations de traitement des eaux usées sont considérées comme la première ligne dans la bataille de l'humanité pour éliminer la pollution de l'eau et restaurer la qualité de l'eau dans les cours d'eaux.

Pourquoi les stations d'épuration ne fonctionnent pas comme prévu?

On constate généralement que ces manques ne sont pas dus à un seul facteur, mais plutôt due à une combinaison de problèmes qui limite la capacité d'une station de traitement des eaux usées. Ces problèmes se répartissent généralement en une ou plusieurs des catégories suivantes:

- Défauts de conception.
- Les lacunes d'équipement.
- Surcharges hydrauliques.

- Les surcharges de déchets industriels.
- Les carences de fonctionnement et d'entretien.

Les stations de traitement des eaux usées sont conçues et construites de manière à éliminer une quantité prédéterminée des polluants contenus dans les eaux usées brutes. La quantité de polluant à enlever repose sur des critères standards de qualité de l'eau correspondant à la quantité de polluant qui peut être contenue dans l'eau et ne pas avoir un impact négatif surtout si on projette sa réutilisation. Une défaillance dans l'élimination de la quantité de polluants requis peut signifier non seulement que l'eau ne peut pas être utilisée comme prévu mais aussi des millions de dinars algériens perdus.

Un dysfonctionnement d'une installation de traitement des eaux usées est un problème social et biologique délicat (Belanche et al., 1992).

L'évaluation des performances est un facteur important dans l'optique d'une réutilisation des eaux usées traitées afin d'assurer des marges de sécurité pour le respect de la santé publique et la protection de l'environnement. La fiabilité d'un système peut être définie comme «la probabilité d'une performance adéquate pour une période de temps déterminée dans les conditions de fonctionnement spécifiées » ou " le pourcentage du temps que les concentrations des eaux usées traitées et rejetées sont conformes aux exigences du permis précisées" (Kottegoda et Rosso, 2009 ; Metcalf et Eddy, 2003). La probabilité de bonne exécution est sensible à la fonction de distribution des concentrations de l'eau traitée à la sortie de la station. Ainsi, afin de développer un modèle de fiabilité, la distribution des concentrations des effluents doit d'abord être modélisée. Cette distribution de probabilité caractérise la variabilité de la performance du traitement.

Le processus de traitement des eaux usées est un processus assez complexe et dynamique. Selon Theilliol et al., (2003), certains facteurs doivent être pris en considération lorsque l'on tente de mettre en œuvre un contrôle avancé en temps réel du processus :

- Le processus est variable dans le temps (Carlsson et Lindberg, 2004). Des paramètres tels la température, la composition de l'eau entrante, la quantité de biomasse, des flux, pH, etc., varient au cours d'une période de temps.
- Le processus est non linéaire et dynamique.
- Le processus est dynamique raides. Les constantes de temps impliquées dans le processus, allant de quelques secondes à quelques mois.
- Le processus est multi variée : Existence de plusieurs entrées et sorties.

Le recours à l'application de techniques intelligentes pour tenter de modéliser un processus de traitement des eaux usées reste une alternative prometteuse rendu possible avec le développement des capacités de calculs des ordinateurs.

Les réseaux de neurones sont une technique d'intelligence artificielle qui utilise un ensemble d'équations non linéaires pour imiter les connexions neuronales des systèmes biologiques. Ils se sont révélés être utiles pour la reconnaissance des formes et des applications de prédiction de résultat.

Le potentiel des réseaux de neurones pour la surveillance et le contrôle des systèmes dynamiques a été découvert il y'a une dizaine d'années (Narendra et Parthasarathy, 1990) et de nombreuses applications dans les procédés industriels existent.

A en juger par le nombre de publications dans des revues et à des conférences, il y'a beaucoup de chercheurs travaillant dans les domaines des réseaux de neurones et les processus de traitement des eaux usées. Ces informations ont été glanées dans des livres de conseils, publications dans les revues, Internet, et des actes de conférences. De nombreux auteurs ont vues sur de nombreuses questions et de nombreux « règles empiriques » différentes ont été proposées pour résoudre les différents problèmes pratiques de mise en œuvre. L'expérience de l'auteur a consisté à utiliser ces règles comme des lignes directrices, mais de résoudre de nombreux problèmes pratiques avec une expérimentation rigoureuse.

La station de traitements des eaux usées de la ville de Khenchela a pour but de traiter les eaux usées du chef-lieu de la wilaya ainsi que sa banlieue proche. Elle a été mise en service le 20 Octobre 2008 avec une capacité de 192000 Eq. Hab, avec un débit d'effluents à traiter en temps de pluie estimé à 23000 m³/jour.

Les travaux présentés dans cette thèse ont pour objectifs de développer des méthodes combinées basés sur un modèle probabiliste et puis la conception d'un modèle réseau de neurones artificiels pour prédire les taux de fiabilités et les taux de défaillances d'un procédé de traitement des eaux usées utilisant les boues activées.

Les principaux objectifs de cette étude sont :

- La collecte et l'analyse des données d'auto-surveillances;
- Construction d'un support numérisé des données et rendre leurs exploitation plus facile;
- Analyse des lois de distributions des différents paramètres de qualité des eaux usées traitées;
- Evaluer statistiquement les performances de la station de traitement des eaux usées de la ville de Khenchela en termes de fiabilité;
- Construire des modèles de réseaux de neurones artificiels pour la prédiction des taux de fiabilités

- Construire des modèles de réseaux de neurones artificiels pour la prédiction des taux de défaillances
- développer les modèles pour être utilisés pour fournir une prédiction quantitative de la performance future du procédé de traitement, et peut être les utilisés pour d'autres stations similaires.

Pour atteindre les objectifs ci-dessus, l'étude a été organisée en sept chapitres.

Le premier chapitre présente les différents précédés de traitement des eaux usées. Une description générale de la de la station de traitement de la ville de Khenchela est présenté dans le deuxième chapitre.

Les différentes méthodes d'évaluation de la fiabilité d'une station d'épuration sont énumérées dans le troisième chapitre.

La première contribution de cette thèse est détaillé dans le chapitre quatre avec l'application du modèle probabiliste pour l'évaluation des niveaux de performances de la station. Une revue de littérature portant sur les réseaux de neurones artificiels est consacrée dans le cinquième chapitre.

La deuxième contribution de cette étude est partagée en deux parties, le sixième chapitre qui se consacre à l'application des RNA pour la prédiction des taux de fiabilités du procédé de traitement suivi d'une discussion des différents résultats.

Le septième chapitre englobe l'application des RNA pour la prédiction des taux de défaillance des différents paramètres de qualité des eaux usées traitées.

Une conclusion générale décrit les principales conclusions et recommandations. Enfin, les références bibliographique.

CHAPITRE I: LES PROCEDES DE TRAITEMENT DES EAUX USEES

I. 1 INTRODUCTION

Le rôle d'une station de traitement des eaux usées est l'élimination de la pollution jusqu'à un niveau définie par la réglementation en vigueur pour assurer que le rejet des eaux traitées n'affecte pas le milieu récepteur, et selon cette réglementation, les procédés de traitement sont mis en œuvre selon plusieurs niveaux de traitements.

Les niveaux de traitement d'une station sont définis selon la succession suivante:

Les prétraitements, le traitement primaire et le traitement secondaire. Lorsque l'eau traitée est être rejetée en milieu particulièrement sensible, un traitement tertiaire est nécessaire.

Une station d'épuration comporte généralement une phase de prétraitement, pendant laquelle les éléments les plus grossiers sont éliminés par dégrillage (pour les solides de grandes tailles), puis par flottaison/décantation (pour les sables et les graisses). Vient ensuite un traitement dit primaire, une décantation plus longue, pour éliminer une partie des MES. Des traitements physico-chimiques et/ou biologiques sont ensuite appliqués afin d'éliminer la matière organique. Ils sont généralement suivis d'une phase de clarification qui est encore une décantation. Enfin, un traitement des nitrates et des phosphates est exigé en fonction de la sensibilité du milieu récepteur. Il existe également des traitements dits extensifs, comme le lagunage, qui combinent des traitements biologiques, physiques et naturels.

I.2 PRETRAITEMENTS

Les eaux usées brutes à leur arrivée à la station doivent généralement subir, un prétraitement qui composé d'un certain nombre d'opérations successives, uniquement physiques ou mécaniques. Il est destiné à extraire de l'eau usée, la plus grande quantité possible d'éléments dont la nature ou la dimension constitueront une gêne pour les traitements ultérieurs. Selon la nature des eaux à traiter et la conception des installations, le prétraitement peut comprendre les opérations : (le dégrillage), principalement pour les déchets volumineux, (le dessablage) pour les sables et graviers et (le dégraissage-déshuilage ou d'écumage-flottation) pour les huiles et les graisses.

I.2.1 Dégrillage

Lors de l'opération de dégrillage, les eaux usées passent au travers d'une grille dont les barreaux, dont l'espacement est déterminé de sorte qu'il puisse retenir matières grossières les plus volumineuses et flottantes charriées par l'eau brute, qui pourraient nuire, obstruer ou provoquer des bouchages dans conduites d'alimentation de l'installation, et nuire à l'efficacité

de la station. Le dégrillage permet aussi de protéger la station contre l'arrivée intempestive des gros objets, les éléments retenus sont, ensuite, éliminés avec les ordures ménagères.

Cette opération est effectuée avant le poste de relevage afin de protéger les pompes ou les vis d'Archimède et de ne pas gêner leur fonctionnement.

Les grilles peuvent être verticales, mais sont le plus souvent inclinées de 60° à 80° sur l'horizontale (Dufournet, 1974), l'espacement est définis comme suit:

Tableau n° 1: Différents types de dégrillage (WEF, 1998; Daukes, 2006)

	Espacement	Rôle
Pré-dégrillage (dégrillage grossier)	40-150 mm	Empêcher que les bûches et les gros débris lourds d'entrer dans les procédés de traitement. Principalement utilisé dans le poste d'arrivage des eaux brutes avant le poste de relevage
Dégrillage moyen	6-75 mm	Pour éliminer les grosses matières solides, les chiffons et les débris. Généralement utilisé dans les stations d'épuration.
Dégrillage fin	1,5-6 mm	Enlever les petits solides. Suit généralement un tamis grossier
Dégrillage très fin	0,25-1,5 mm	Réduire les matières en suspension au niveau du traitement primaire.

I.2.2 Poste de relevage

La conception du poste de pompage (poste de relevage) d'une station de traitement des eaux usées est en grande partie similaire à la conception de poste de relèvement placé dans le réseau d'assainissement. Les principales différences est le choix du type de pompes utilisées pour assurer cette tâche. Généralement le poste de relevage est équipé de pompes centrifuges, par contre dans le cas d'une station d'épuration plus importante (débit plus important et variable), une pompe à vis (Vis d'Archimède) est un choix plus judicieux.

Les pompes à vis (Vis d'Archimède) peuvent pomper une variété des eaux usées brutes contenant des solides et des débris. L'un des avantages majeurs de ces pompes est la variabilité du débit de pompage à une vitesse constante, puisque le niveau d'eau dans le puisard est le facteur qui influence sur les performances de pompage. Les opérateurs constatent qu'une fois installées, les pompes à vis d'Archimède sont rarement victimes de problèmes de défaillances (Garbus, 2006).

I.2.3 Dessablage

Le dessablage a pour but d'extraire des eaux brutes les graviers, les sables, les verres brisés, les coquilles d'œufs, et les particules minérales plus ou moins fines ayant une vitesse de sédimentation sensiblement supérieure à la matière organique. Le dessablage est prévu pour protéger les équipements mécaniques à l'abrasion et à l'usure, de réduire la formation de dépôts dans les canalisations et les canaux, et de réduire la fréquence de nettoyage du digesteur qui est nécessaire en raison des particules accumulés.

Un but secondaire, mais cependant pas le moins extrêmement souhaitable du système d'élimination du sable est de séparer les grains de la matière organique dans les eaux usées. Cette séparation permet à la matière organique d'être traité dans les processus subséquents.

Dans un dessableur à flux horizontal, pour assurer l'élimination des grains et empêcher que la matière organique se dépose, trois conditions doivent être remplies (Steel et McGhee, 1979) :

1. La vitesse d'écoulement à la sortie du dessableur doit être égale à la vitesse de sédimentation des particules inertes.
2. La vitesse horizontale doit être inférieure à la vitesse d'érosion des particules inertes.
3. La vitesse horizontale doit être supérieure à la vitesse de décantation des particules organiques.

Dans les années 1950, la philosophie de conception des dessableurs implique que l'écoulement horizontal doit capturer les particules de sable de diamètre de 200 μm et une densité proche de 2,65 (Fair et Geyer, 1954).

Des recherches ont révélés que la gamme de densités des grains de sable est de l'ordre de 1,1 à 2,7 (Metcalf & Eddy, 2003). En outre, il est reconnu que la graisse et autres matières organiques fréquemment recouvre le grain de particules inorganiques. Ainsi, ni la densité, ni la taille d'une particule de grain ne peut être décrite en termes d'un grain de sable seul. Une mesure plus réaliste appelé la taille équivalente de sable (TES) est préféré (Wilson et al., 2007a).

I.2.3.1 Dessableur aéré

On insuffle de l'air qui provoque une rotation du liquide et crée ainsi une vitesse constante de balayage du fond, perpendiculaire à la vitesse de transit, laquelle beaucoup plus faible.

La vitesse d'eau à travers le fond du bassin est influencée par la taille des particules et la densité des particules qui vont se déposer (Albrecht, 1967 ; Sawicki, 2004). L'action de roulement induit par les diffuseurs d'air est indépendante du débit de transit. La vitesse de diffusion de l'air et de la forme du bassin régit la vitesse du rouleau. Les particules sont déplacées par l'écoulement en spirale vers le fond de la cuve.

Le sable est extrait soit mécaniquement ou par raclage vers un poste de réception, puis repris par pompage, soit directement par pompe suceuse montée sur pont roulant.

Les avantages et les inconvénients du dessableur aéré sont résumés dans le tableau ci-dessous.

Tableau n° 2: Avantages et inconvénients des dessableurs aérés (WEF, 1998 ; Spangler, 2006)

Avantages	Inconvénients
<ul style="list-style-type: none"> • Efficacité constante sur une large gamme de débit • Perte de charge minimale • Teneur en matière organique peut être contrôlé par le taux de l'air • Bassin peut être utilisé pour ajouter et mélanger les produits chimiques. 	<ul style="list-style-type: none"> • Consommation électrique élevée • Main-d'œuvre nécessaire pour l'entretien du système d'air • Les composés organiques volatils peuvent être libérés • Présence d'odeur nuisible • Entretien de l'équipement d'enlèvement

Bien qu'il n'y ait pas d'arguments formels pour l'efficacité requise de dessablage, généralement il a été supposé qu'un dessableur aéré fonctionne correctement devrait supprimer 100 % de la fraction de sable supérieur à 200 µm de diamètre et 65 à 75 % de la fraction de sable entre 100µm et 200µm. En outre, la teneur en matière organique du grain capturée ne devra pas dépasser 10% (Imhoff et Imhoff, 2007).

I.2.3.2 Durée de détention.

L'expérience a montré que le temps de séjour de 2 à 5 minutes au débit horaire de pointe est suffisante pour obtenir plus de 95 % de retrait de la conception traditionnelle de la taille des particules (200 µm de diamètre, le poids spécifique de 2,65). Typiquement, le temps de détention est fixé à 3 minutes à débit horaire de pointe (Metcalf et Eddy, 2003; WEF, 1998). Un temps de détention de 3,6 à 7,8 minutes dans des conditions d'écoulement moyenne ont été signalés à réaliser des économies allant de 94 à 98,8% (Morales et Reinhart, 1984). Un temps de détention plus long améliore le dessablage et peut être nécessaires pour capturer les petites particules abrasives (WEF 1998).

I.2.3.3 Dégraissage déshuilage

Le déshuilage-dégraissage se rapporte à l'extraction de toutes les matières flottantes d'une densité inférieure à celle de l'eau. Ces matières sont de natures très diverses et leurs quantités s'estime par la mesure des «matières extractibles par solvants». La teneur des eaux usées en matières extractibles est de l'ordre de 30 à 75 mg/L.

Néanmoins, certains rejets industriels (abattoirs, laiteries...) peuvent élever ces valeurs à 300-350 mg/L.

Les huiles et graisses, lorsqu'elles ne sont pas émulsionnées, sont séparées sous forme de boues flottantes dans des ouvrages comportant une zone d'aération où les bulles d'air augmentent la vitesse de montée des particules grasses et une zone de tranquillisation où s'effectue la récupération.

Le temps de séjour dans ce type d'ouvrage est de 5 à 12 min. Le débit d'air insufflé est de l'ordre de 0,2 m³ par mètre cube d'eau et par heure.

Le plus souvent, les fonctions de dessablage et de déshuilage sont combinées dans un même ouvrage qui met en œuvre les principes de fonctionnement cités précédemment (Gaïd, 1993).

I.3 TRAITEMENTS PRIMAIRES

Le traitement "primaire" fait appel à des procédés physiques naturels, filtration et décantation plus ou moins aboutie, éventuellement assortie de procédés physico-chimiques, tels que la coagulation- floculation.

I.3.1 Décantation primaire simple sans réactif

La décantation se distingue du dessablage en ce sens qu'elle enlève plus de 50% des matières en suspension, alors que le dessablage ne vise que l'élimination du sable et des autres matières minérales (plus lourdes que les matières organiques) (Gaid, 1993).

Elle permet d'alléger les traitements biologiques et physico-chimiques ultérieurs, en éliminant une partie des solides en suspension. L'efficacité du traitement dépend du temps de séjour et de la vitesse ascensionnelle (qui s'oppose à la décantation). La décantation primaire permet d'éliminer, pour une vitesse ascensionnelle de 1,2 m/h, 40 à 60 % des MES, soit 40% de matière organique, 10 à 30 % des virus, 50 à 90 % des helminthes et moins de 50 % des kystes de protozoaires et entraîne également avec elle une partie des micropolluants, (Faby et Brissaud, 1997).

Le dépôt des matières solides au fond d'un ouvrage appelé "décanteur" produit boues "primaires". Le raclage des boues s'effectue au moyen d'un racleur de fond entraîné par chaînes qui les ramène vers 1 ou 2 trémies. Avec un décanteur classique ou lamellaire sur des eaux domestiques, le pourcentage de matières totales en suspension éliminées est de 50 à 65 %, celui de la DBO₅ éliminée est de 20 à 35 % (Gaïd, 1993).

I.3.2 Elimination de la pollution colloïdale

Dans certaines conditions, le traitement primaire simple est insuffisant. L'amélioration du rendement d'épuration de la décantation se fait alors par ajout de réactifs de coagulation floculation. Pour certains effluents industriels, la flottation peut être utilisée.

La coagulation permet l'agglomération directe de particules colloïdales, alors que la floculation, qui suit, fait chuter les agrégats déjà formés par coagulation.

Les principaux coagulants minéraux utilisés en eaux résiduaires urbaines sont le sulfate d'alumine, le chlorure ferrique. Les floculants organiques les plus employés sont des polymères synthétiques de haut poids moléculaire. En général, des essais de laboratoire sont indispensables pour sélectionner le ou les réactifs à utiliser (Jar-test).

Les techniques utilisées sont extrêmement diverses. Elles vont d'appareils directement extrapolés des décanteurs primaires (décanteur-floculateur) à des appareils plus complexes et plus performants.

Les performances dépendent très largement de la composition des eaux brutes du choix et du dosage des réactifs. On obtient en moyenne une réduction de (Sadowski, 2002):

- DBO₅ : 70 à 80 %;
- MES : 90 %.

Ces traitements sont généralement utilisés dans les stations d'épuration de grande capacité, ou dans celles ayant à faire face à de grandes variations de charge dans l'année (zone touristique). La séparation du floc a lieu pendant la phase de clarification (décantation secondaire). Les procédés les plus modernes utilisent du micro sables injectés dans l'effluent afin d'accélérer la décantation des floes. On parle alors d'élimination à floes lestés (Lazarova et al., 2003).

I.4 TRAITEMENTS SECONDAIRES (TRAITEMENT BIOLOGIQUE)

L'épuration biologique a pour but d'éliminer la matière polluante biodégradable contenue dans l'eau domestique (décantée ou non) en la transformant en matières en suspension : micro-organismes et leurs déchets, plus facilement récupérables (Dufournet, 1974 ; Gaïd, 1993).

La dégradation peut se réaliser par voie aérobie (en présence d'oxygène), pour les eaux usées et en anaérobie (en l'absence d'oxygène) pour le traitement des boues.

Le traitement biologique classique des eaux domestiques s'effectue par voie aérobie, qui consiste à dégrader les impuretés grâce à l'action d'une biomasse épuratrice, à laquelle doit être fourni l'oxygène nécessaire à son développement (Gaïd, 1993).

Selon (Faby et Brissaud, 1997), une épuration biologique (boues activées, puis bassin de clarification) permet d'éliminer 90 % des virus, 60 à 90 % des bactéries, mais par contre a peu d'effet sur les kystes de protozoaires et les œufs d'helminthes. Selon (Rowe et Abdel Magid, 1995 ; Asano, 1998) un traitement par boues activées élimine 90 % des bactéries entériques, 80 à 99 % des entérovirus et des rotavirus. L'élimination a lieu grâce à la sédimentation des MES, la compétition avec les micro-organismes non pathogènes et la température ; la part la plus importante est due à la sédimentation.

Ces traitements conçus à l'origine essentiellement pour l'élimination de la pollution carbonée et des matières en suspension, ainsi pour poursuivre l'épuration de l'effluent provenant du décanteur primaire ; par voie biologique le plus souvent.

Les micro-organismes, les plus actifs, sont les bactéries qui conditionnent en fonction de leur modalité propre de développement, deux types de traitements :

I.4.1 Traitements anaérobies

Les traitements anaérobies font appel à des bactéries n'utilisant pas de l'oxygène, en particulier, aux bactéries méthanogènes qui conduisent, comme leur nom l'indique, à la formation du méthane à partir de la matière organique, et à un degré moindre de CO₂. Ce traitement. Ce type de traitement appelé aussi « digestion anaérobie » est une opération délicate qui demande une surveillance importante. En effet, la température doit être maintenue à un niveau très stable et suffisamment élevé. Il faut aussi éviter les écarts brutaux de pH et les substances inhibitrices du développement bactérien. Ce traitement est utilisé pour des effluents très concentrés en pollution carbonée, de type industriel (Gaïd, 1993; Sadowski, 2002).

I.4.2 Traitements aérobies

Les micro-organismes utilisés exigent un apport permanent d'oxygène. On distingue cinq méthodes essentielles :

I.4.2.1 Cultures fixes (lits bactériens et disques biologiques)

Le lit bactérien est le plus ancien procédé à biomasse fixée est le lit bactérien. La biomasse est fixée sur un matériau de grosse granulométrie (3 à 8 cm) sur lequel percole l'effluent à traiter. L'air est transféré par diffusion à travers le film d'eau ruisselant à la surface du matériau. Une vitesse hydraulique suffisante, assurée par un recyclage d'eau traitée, permet l'évacuation des boues en excès qui sont séparées de l'effluent traité dans un ouvrage de décantation situé en aval (Gaïd, 1993).

Le principe du disque biologique consiste en l'utilisation de disques tournant autour d'un axe horizontal et baignant en partie dans l'eau à traiter. De par la rotation, la biomasse fixée sur les disques se trouve alternativement en contact avec l'eau à traiter et l'oxygène de l'air.

Comme dans le cas du lit à ruissellement, la biomasse se développe sur le support et le transfert d'oxygène se fait directement à travers la couche liquide. La régulation de l'épaisseur du biofilm se fait naturellement dès que l'assise biologique en contact avec le support passe sous des conditions anaérobies liées à une limitation de transfert de l'oxygène. Le décrochage de la boue en excès nécessite la présence en aval d'un dispositif de clarification de l'eau traitée.

La vitesse de rotation de ces disques (1 à 2 tours par minute) ne permet pas de générer des énergies de circulation capables de maintenir en suspension des matières solides. Le risque de dépôts en fond de bassin oblige donc à un prétraitement de l'eau brute par décantation primaire et empêche la recirculation de la boue.

Les disques sont réalisés en polystyrène, PVC ou feuilles de polyéthylène, leur diamètre est généralement compris entre 2 et 3 m. Les disques utilisés sont plats ou présentent des ondulations ou une rugosité créée par des reliefs de façon à accroître la surface de fixation de la biomasse. Ils sont espacés de 2 à 3 cm et leur vitesse de rotation est de 1 à 2 tours par minute. Les surfaces développées sont de 150 à 200 m²/m³ de disque (Sadowski, 2002).

I.4.2.2 Les cultures libres (boues activées)

Le terme « cultures libres » regroupe les procédés où l'on provoque le développement d'une culture bactérienne dispersée sous forme de flocons au sein du liquide à traiter. Le procédé par « boues activées » est le plus commun des procédés par « cultures libres » (Dufournet, 1974 ; Gaïd, 1993 ; Faby et Brissaud, 1997 ; Sadowski, 2002).

Pour cela, on utilise un bassin brassé, pour conserver en suspension la culture, dans lequel est maintenue. Dans ce bassin, le brassage a pour but d'éviter les dépôts et d'homogénéiser le mélange des flocons bactériens et de l'eau usée (liqueur mixte) ; l'aération peut se faire à partir de l'oxygène de l'eau, du gaz enrichi en oxygène par (le brassage, l'injection d'air comprimé, voire même d'oxygène pur), a pour but de dissoudre ce gaz dans la liqueur mixte, afin de répondre aux besoins des bactéries épuratrices aérobies.

Après un temps de contact suffisant, les bactéries épuratrices disposent d'un temps d'action suffisant appelée "temps de contact", ensuite la liqueur mixte est envoyée dans un clarificateur appelé parfois décanteur secondaire, pour séparer l'eau épurée des boues. Une partie de ces dernières sont recyclées dans le bassin d'aération pour y maintenir une concentration suffisante en bactéries épuratrices.

L'excédent (boues en excès) est extrait du système et évacué vers la filière de traitement des boues.

I.4.2.3 Le lagunage

La technique de traitement des eaux usées par lagunage se sert des mécanismes naturels de l'environnement où l'eau est épurée par des communautés de micro-organismes variés.

L'épuration par lagunage naturel repose sur la présence de bactéries aérobies en cultures libres et d'algues. L'oxygène nécessaire à la respiration bactérienne est produit par des végétaux en présence du rayonnement solaire. Les bactéries présentent dans le système consomment la pollution dissoute dans l'eau pour respirer. L'oxygène est produit grâce aux mécanismes photosynthétiques des algues qui poussent et se développent grâce aux engrais qu'apportent les eaux d'égout (Racault et al., 1997).

Le lagunage aéré permet d'épurer des rejets peu chargés en matières en suspension dans des bassins de 2 à 3 m de profondeur. Comme il n'y a pas de recirculation de boues à partir d'un clarificateur, il se crée un équilibre entre l'apport de pollution biodégradable et la masse de bactéries qui se développe à partir de cette pollution. Les dispositifs d'aération sont calculés sur la base des besoins en oxygène. Compte tenu des grands volumes mis en jeu, les puissances spécifiques appliquées sont faibles (2 à 5 W/m³) ce qui se traduit par une décantation des bactéries au fond des bassins où elles forment un dépôt (Sadowski, 2002).

Il simule, en l'amplifiant, l'action auto-épuratrice des étangs ou des lacs. Associés aux systèmes conventionnels de traitement secondaire, ils constituent aussi d'excellents dispositifs tertiaires aptes à réduire les risques liés aux micro-organismes pathogènes.

Les mécanismes de l'épuration et le fonctionnement d'un lagunage consiste, à retenir les effluents dans des bassins pendant une période plus ou moins longue au cours de laquelle les organismes présents permettent d'éliminer 50 à 60 Kg de DBO₅/hectare par jour (Gaïd, 1993; Sadowski, 2002)

I.4.2.4 Filtration/percolation

La filtration ou percolation consiste à traiter l'eau par l'intermédiaire du sol ou d'un massif filtrant (Vasel, 2007). On filtre les effluents à raison de quelques centaines de litres d'effluent par mètre carré de massif filtrant et par jour. Deux mécanismes entrent en jeu :

1) La filtration des MES : plus le sable est grossier, plus la fixation des MES se fera en profondeur. Les MES finissent par colmater le filtre. Pour lutter contre le bouchage du massif filtrant, il faut donc alterner phase de filtration et phase de séchage, l'élimination des MES permet également l'élimination des micro-organismes qui y sont fixés ;

2) L'adsorption des bactéries libres par les grains de sable du filtre : il se forme alors un film biologique contaminé, surtout dans la partie supérieure, ce film va permettre une dégradation microbienne de la matière organique et des substances dissoutes dans l'effluent (phosphates, nitrates, etc.). Cette dégradation consomme de l'O₂ et produit du CO₂, il faut donc aérer régulièrement le film pour éviter l'asphyxie du milieu.

Les techniques de filtration/percolation permettent l'élimination des « gros » micro-organismes (protozoaires et helminthes) par filtration/adsorption au début du massif filtrant.

L'élimination des virus et des bactéries est fonction du milieu poreux, de la vitesse de percolation, de l'épaisseur du massif filtrant et du niveau d'oxydation de l'eau filtrée (Faby, 1997).

I.4.2.5 Filtres plantés

Le sol et les rivières, ont été depuis toujours les milieux récepteurs naturels de la pollution engendrée par l'activité humaine. Les techniques sophistiquées actuellement mises en œuvre dans les stations d'épuration des eaux résiduaires urbaines, ne constituent qu'une optimisation de l'auto épuration naturelle qui s'exerce dans ces deux milieux (Guilloteau, 1992).

Plusieurs chercheurs proposent une épuration par les procédés de type extensif qui sont des systèmes d'épuration d'eaux usées par voie naturelle (ou zones humides artificielles), dont le plus connu est le lagunage à microphytes et à macrophytes (bassin de stabilisation, bassin facultatif ou bassin de maturation). Ce système est expérimenté depuis la première moitié du siècle sous climat tempéré et présente de nombreuses possibilités d'adaptation au climat des pays en développement qui est généralement chaud et donc favorable à l'activité bactérienne dont dépend son bon fonctionnement. Ces stations d'épuration par voie naturelle sont nombreuses, mais elles ne sont pas souvent connues du grand public. Leur principe de fonctionnement s'inspire de celui des écosystèmes rencontrés dans les zones humides naturelles (Radoux, 1989).

I.5 TRAITEMENTS TERTIAIRES

Les traitements tertiaires ou d'affinage visent principalement l'élimination de la pollution l'azotée et phosphatée ainsi que la pollution bactériologique des eaux usées domestiques, ayant déjà subi au préalable des traitements primaires et secondaires qui s'avèrent insuffisants pour répondre aux normes de rejet. Pour cela les traitements tertiaires s'imposent et deviennent plus que nécessaires, afin de garantir une meilleure protection des milieux récepteurs.

Les traitements tertiaires souvent considérés comme complémentaire permettent d'affiner ou d'améliorer le traitement secondaire. De telles opérations sont nécessaires pour assurer une protection complémentaire de l'environnement ou en vue d'une réutilisation de l'eau traitée en agriculture ou en industrie. On distingue généralement les opérations suivantes :

- La nitrification-dénitrification et déphosphatation biologique ou mixte (biologique et physico-chimique) ;
- La désinfection bactériologique et virologique.

I.5.1 Elimination biologique de l'azote

L'impact de l'azote sur le milieu récepteur explique le souci de l'élimination de l'azote qui se retrouve dans les textes réglementant les rejets d'effluents dans le milieu.

L'azote présent dans les rejets domestiques est essentiellement d'origine urinaire. Il est apporté pour près des 3/4 par l'urée qui s'hydrolyse rapidement pour donner NH_4 en présence d'une enzyme qui est l'uréase. Les bactéries qui possèdent cette enzyme sont en abondance dans les égouts. Par contre, les formes oxydées de l'azote (NO_2 , NO_3) seront toujours en faible concentration, inférieure en général à 1 mg/l, en entrée de station. En plus de NH_4 , les eaux résiduaires contiennent dans des proportions variables de l'azote organique. On rappellera que l'azote total Kjeldahl ou NTK est la somme de l'azote organique et ammoniacal exprimée en N. (Les nitrites (NO_2) et les nitrates (NO_3) ne sont pas pris en compte par ce dosage).

L'élimination de l'azote fait intervenir quatre processus :

1. Ammonification : transformation de l'azote organique en azote ammoniacal par des bactéries banales;
2. Assimilation : utilisation d'une partie de l'azote ammoniacal et éventuellement organique pour la synthèse bactérienne;
3. Nitrification : oxydation de l'azote ammoniacal en nitrites et en nitrates. Cette réaction fait intervenir des bactéries très particulières : Nitrosomonas pour l'oxydation de NH_4 en NO_2 , Nitrobacter pour l'oxydation de NO_2 en NO_3 ;
4. Dénitrification : réduction des nitrates en azote gazeux qui retournera après un stade intermédiaire NO_2 sous sa forme primitive dans l'atmosphère. Cette réaction fait intervenir des bactéries banales, abondantes dans le milieu naturel (Sadowski, 2002).

I.5.2 Elimination biologique du phosphore

L'élimination du phosphore, ou "déphosphatation", peut être réalisée par des voies biologiques ou physico-chimiques.

Le principe de la déphosphatation biologique consiste en une suraccumulation de phosphore dans une biomasse. Si des teneurs de 2 à 3 % en phosphore dans les boues sont obtenues sous des conditions normales de dégradation d'un substrat organique, le mécanisme de suraccumulation nécessite de placer la biomasse alternativement en phase anaérobie (sans oxygène même lié à un composé chimique) et aérobie. Dans les grosses installations d'épuration, ce procédé est souvent couplé à une déphosphatation physico-chimique, pour atteindre les niveaux de rejets requis.

Les sels de fer ou d'aluminium sont également capables de se combiner avec les ions phosphate pour former un précipité de phosphate de fer ou d'aluminium (FePO_4 ou AlPO_4).

Du fait de la compétitivité des réactions de formation de précipités d'hydroxyde ou de phosphate, le rapport molaire à mettre en œuvre entre Fe/P ou Al/P varie de 1 à 3. Cela signifie que, pour éliminer 1 g de P, il faut :

1,8 à 5,4 g de Fe soit, exprimé par exemple en FeCl_3 pur, de 5,2 à 15,7 g ou 0,87 à 2,61 g d'Al soit, exprimé en sulfate d'alumine, de 9,3 à 28 g de produit commercial [$\text{Al}_2(\text{SO}_4)_3, 18 \text{H}_2\text{O}$].

La précipitation chimique du phosphore peut se réaliser à plusieurs étapes dans la filière de traitement :

1. Au niveau de la décantation primaire ;
2. Dans le bassin de boues activées, le précipité décantant alors dans le clarificateur ;
3. Sur l'eau épurée biologiquement avec un ouvrage spécifique de décantation (Gaïd, 1993).

I.5.3 Elimination et traitement des odeurs

Les eaux usées, chargées en matières organiques particulières et dissoutes, en composés azotés et phosphorés, peuvent dégager des odeurs désagréables suivant un processus biologique bien connu qui se déclenche en milieu réducteur. Par ailleurs, certains rejets industriels contiennent des composés très volatils utilisés dans les procédés de fabrication comme des sulfures, des aldéhydes, des alcools ou encore de l'ammoniaque.

Les principaux composés odorants rencontrés dans les stations d'épuration font partie essentiellement des familles des produits soufrés et azotés, ainsi que des composés organiques tels les acides gras volatils. Pour éviter la propagation des mauvaises odeurs émises aux différents postes de traitement, il convient d'isoler les sources odorantes dans des enceintes hermétiques. La couverture complète de tous les ouvrages est la technique la plus adaptée avec, en plus, un souci d'intégration au site. La couverture du poste de relevage, des prétraitements et de la filière de traitement des boues reste toutefois suffisante dans la plupart des cas. Après confinement, les odeurs doivent être évacuées par ventilation forcée et les composants odorants traités (Gaïd, 1993).

Des installations de désodorisation chimique ou biologique sont également mises en place, au sein des stations d'épuration. La désodorisation chimique est la technique la plus utilisée. Les gaz malodorants sont captés puis envoyés dans des tours de lavage, où un liquide désodorisant est pulvérisé. Ces lavages peuvent comporter de la soude, de l'acide et/ou de l'hypochlorite de sodium (eau de javel), réactifs qui captent ou neutralisent les mauvaises odeurs.

I.5.4 Procédés de désinfections

A l'issue des procédés décrits précédemment, les eaux sont normalement rejetées dans le milieu naturel. Dans le cadre d'une réutilisation, les eaux usées nécessitent des traitements supplémentaires, essentiellement pour éliminer les micro-organismes qui pourraient poser des problèmes sanitaires. Ce ne sont pas des traitements d'épuration classiques ; par contre ils sont fréquemment utilisés dans les usines de production d'eau potable. On peut donc supposer qu'ils constituent l'aménagement technique minimum d'une station d'épuration en vue d'une réutilisation.

Trois produits ont été testés pour la désinfection d'eaux résiduaires urbaines : le chlore, le dioxyde de chlore et les permanganates de potassium (Spellman, 2003).

Le chlore présent dans l'eau sous forme ionique pénètre dans la cellule après une altération de la membrane cytoplasmique, réagit avec les acides aminés et enzymes et bloque ainsi le

métabolisme du glucose. La membrane étant altérée et ne pouvant plus assurer sa "production d'énergie", la cellule meurt.

La réalisation d'un poste de chloration doit répondre à un dimensionnement précis. Le point d'injection du réactif doit être dans une zone de mélange parfait. De plus, le temps de contact entre le désinfectant et l'effluent à traiter ne peut être assuré que par une hydraulique de type flux piston.

Le dioxyde de chlorure est un réactif très instable et doit être fabriqué sur place à partir de l'oxydation de chlorites (ClO_2^-) par le chlore. Il est alors injecté dans l'eau sous forme gazeuse ou liquide.

Le chlorure de brome est livré sous forme de liquide pressurisé. Sa vaporisation s'effectue à température ambiante. Il est injecté sous forme gazeuse (Sadowski, 2002).

Les désinfections utilisant des produits chimiques sont efficaces, sauf contre les *Cryptosporidium*. Il a été montré que des kystes de *Cryptosporidium* pouvaient résister à des traitements à $\text{pH} = 11,2$, à la chloration et à d'autres traitements chimiques (Rose et al, 1999). Cependant, la plus grande partie des kystes de *Cryptosporidium* sont éliminés pendant les phases primaires de décantation et coagulation/floculation. Par ailleurs, il faut trouver l'équilibre entre le risque posé par les désinfectants en eux-mêmes, et le risque lié aux microorganismes pathogènes (Asano et Levine, 1995). C'est essentiellement le cas pour le chlore dont l'utilisation crée des dérivés halogénés potentiellement cancérigènes. Pour les ultraviolets, ce problème ne se pose pas. Leur action sur les virus et les coliformes fécaux est bonne. Seules les formes de résistances, comme les œufs d'helminthes, ne sont pas trop affectées (Cauchi et al, 1996).

Le traitement de désinfection par les rayons UV est plus économique et pose moins de problèmes de toxicité que le chlore. Il est largement utilisé dans les moyennes et grandes stations (Metcalf & Eddy, 2003.).

I.6 TRAITEMENT DES BOUES

La réduction de volume est classiquement obtenue à travers des opérations de séparation de phases liquide/solide par décantation, filtration ou évaporation rencontrées dans les techniques d'épaississement, de déshydratation et de séchage thermique. La dégradation des matières organiques de la boue par des procédés biologiques (digestion, compostage) ou thermiques (incinération à $850\text{ }^\circ\text{C}$, oxydation par voie humide de boues liquides épaissies sous 45 bar à $250\text{ }^\circ\text{C}$) conduira également à un volume final moindre.

La stabilisation sera concrètement obtenue en ralentissant, voire en supprimant, la biodégradation putride des matières organiques de la boue, à travers différentes voies, biologique, chimique ou physique avant (phase liquide) ou après (phase pâteuse) l'étape de déshydratation.

La déshydratation constitue souvent l'étape ultime de la filière de traitement des boues:

Une siccité minimale peut en effet être imposée contractuellement (généralement > 30 %) en vue de l'évacuation de la boue ou être requise en vue d'une incinération dans des conditions d'auto combustibilité.

Les principaux débouchés des boues produites par les stations d'épuration restent la possibilité d'une valorisation agricole (Gaïd, 1993).

I.7 CONCLUSION

Les différents constituant d'un procédé de traitement des eaux usées à fait l'objet d'une présentation sommaire.

Grâce aux développements des procédés de traitements, la production des eaux usées issue de l'activité humaine, n'est plus un problème, mais une source alternative pour une récupération des eaux traitées afin de les réutiliser pour répondre aux besoins croissant dans l'agriculture et l'industrie.

CHAPITRE II: DESCRIPTION GENERALE DE LA STATION D'EPURATION DE LA VILLE DE KHENCHELA

II.1 LOCALISATION DE LA STATION D'EPURATION

La station de traitement des eaux usées municipale de la ville de Khenchela est située à 2,5 Km au Nord de la ville et à une l'altitude de 1045m dans la vallée d'OUED BOUGHEGAL.

Les coordonnées de localisation géographiques selon le système GPS sont :

- Latitude : N 35°27'37.11"
- Longitude : E 7° 8'55.06"

La station s'étend sur une superficie de 8 hectares (Hydro Projet Est, 2005). La figure 1 représente un aperçu général de la station d'épuration de la ville de Khenchela

II.2 DONNEES SPECIFIQUES DE LA STATION

La station est conçue pour répondre aux exigences de traitement de la ville de Khenchela (chef-lieu de wilaya et de sa banlieue), avec une capacité de 192000 Eq. Hab (Hydro-projet Est, 2005).

Tableau 3: Données spécifiques de la station d'épuration de Khenchela (Hydro-projet Est, 2005).

	Unités	Horizon 2015
Equivalent .habitant	E.H	192000
Volume journalier	m ³ /j	23000
Débit moyen de temps sec	m ³ /h	959
Débit de pointe de temps sec	m ³ /h	1516
Débit maximal admis en temps de pluie	m ³ /h	4446,5



Figure 1: Aperçu général de la station d'épuration de la ville de Khenchela.

II.3 PRESENTATION DE LA FILIERE DE TRAITEMENT DE LA STATION

II.3.1 Filière de traitement des eaux

La filière de traitement des eaux usée de la station de Khenchela est constituée de:

- Dégrillage grossier manuel;
- Quatre dégrilleurs mécanisés;
- Deux dessaleurs-déshuileurs, rectangulaires avec aération immergés et pompes de reprise des sables, et égouttoir à graisses;
- Six lits de sable;
- Huit bassins d'aération par turbines de surface de dimension unitaire répartiteur avec quatre vannes batardeaux pour régler;
- Quatre clarificateur de 32 m de diamètre;
- Un bassin de désinfection (de contact) (Hydro-projet Est, 2005).

II.3.2 Filière des boues

La filière boue est constituée de:

- Un Poste de recirculation et extraction des boues comprenant chacun trois pompes de recirculation d'une capacité unitaire de 720 m³/h en amont du bassin.
- Un épaisseur gravitaire équipé de deux pompes volumétriques (pompe à rotor excentré) pour l'extraction et le refoulement vers les lits de séchages.
- Huit lits de séchage solaire (Hydro-projet Est, 2005).

II. 4 CARACTERISTIQUES TECHNIQUES DES OUVRAGES DE LA STATION

Les bases retenue pour le dimensionnement de la station d'épuration de la ville de Khenchela sont conformes aux charges définies dans le dossier d'appel d'offre et sont récapitulées dans les tableaux suivants :

Tableau 4:Données de la charge polluante (Hydro Projet Est, 2005).

Charge journalière en DCO (Kg/j)	3542
Charge journalière en DBO ₅ (Kg/j)	3404
Charge journalière en MES (Kg/j)	17871

Tableau 5:Choix du procédé de traitement selon la qualité de rejet conforme aux normes Algériennes (Hydro Projet Est, 2005).

Paramètres	Concentration en (mg/l)	Rendement minimum d'élimination en (%)
DBO ₅	≤ 40	≥ 90
DCO	≤ 130	≥ 90
MES	≤ 40	≥ 90
NTK	≤ 07	≥ 80

II.4.1 Caractéristiques des ouvrages existants

Les caractéristiques techniques des dégrilleurs sont présentées dans le tableau ci-dessous.

Tableau 6:Caractéristiques du dégrillageur grossier (Hydro Projet Est, 2005)

Désignation	Unité	Horizon 2015
Type de grille	/	Barreau incliné à 60°
Nombre de grille	/	02
Espacement entre fer	mm	100
Epaisseur des barreaux	mm	10
Matériau de construction du dégrilleur	/	acier

Les déchets recueillis du dégrillage grossier seront acheminés manuellement (Hydro-projet Est, 2005).

Tableau 7:Caractéristiques du poste de relevage des eaux usées (Hydro Projet Est, 2005).

Le nombre des pompes	02 + 02
Temps de fonctionnement des pompes	2 à 8 h
Débit des pompes	2 pompes de 800 m ³ /h 2 pompes de 1600 m ³ /h

Tableau 8: Caractéristiques du dégrilleur fin (Hydro Projet Est, 2005).

Désignation	Unité	Horizon 2015
Nombre de grilles	/	04
Ecartement entre barreaux	mm	10
Matériaux de construction châssis et grille		Acier galvanisé

Le dessableur-déshuileur aéré installé dans la station d'épuration de Khenchela présente les caractéristiques suivantes.

Tableau 9: Caractéristiques du dessableur-déshuileur aéré (Hydro Projet Est, 2005).

Désignation	Unité	Horizon 2015
Forme géométrique de l'ouvrage	Rectangulaire	Rectangulaire
Nombre de files d'eau	/	02
Surface totale de l'ouvrage	m ²	76,95 × 2
Volume total de l'ouvrage	m ³	238,5 × 2
Longueur	m	27
largeur chambre à sable	m	1,9
largeur chambre à graisse	m	0,95
Hauteur d'eau dans l'ouvrage	m	3,1

Remarque: Le système d'aération pour la flottation des graisses est constitué de supprimeurs et diffuseurs, l'extraction de sable est réalisée par des pompes de reprise.

Tableau 10: Caractéristiques du traitement biologique (Hydro Projet Est, 2005).

Désignation	Unité	Horizon 2015
Débit max a traité	m ³ /h	467
Nombre des bassins	/	8
Longueur de chaque bassin	m	23
largeur de chaque bassin	m	18
Forme géométrique de l'ouvrage	Rectangulaire	Rectangulaire
Norme de filles	U	02
Charge massique (cm)	Kg DBO ₅	0,15
Surface totale des bassins	m ²	3375
Volume total des bassins	m ³	15187,5
Hauteur d'eau dans l'ouvrage	m	4,5
Concentration en MVS dans le bassin	g/l	04
Puissance installé	kW	37

Remarque: L'aération des bassins est assurée par des agitateurs (aérateur de surface).

II. 3 SCHEMA GENERAL DU PROCEDE DE TRAITEMENT

Le schéma général du procédé de traitement des eaux usées de la ville de Khenchela (type boue activée) est présenté sur la figure 2 (ci-dessous).

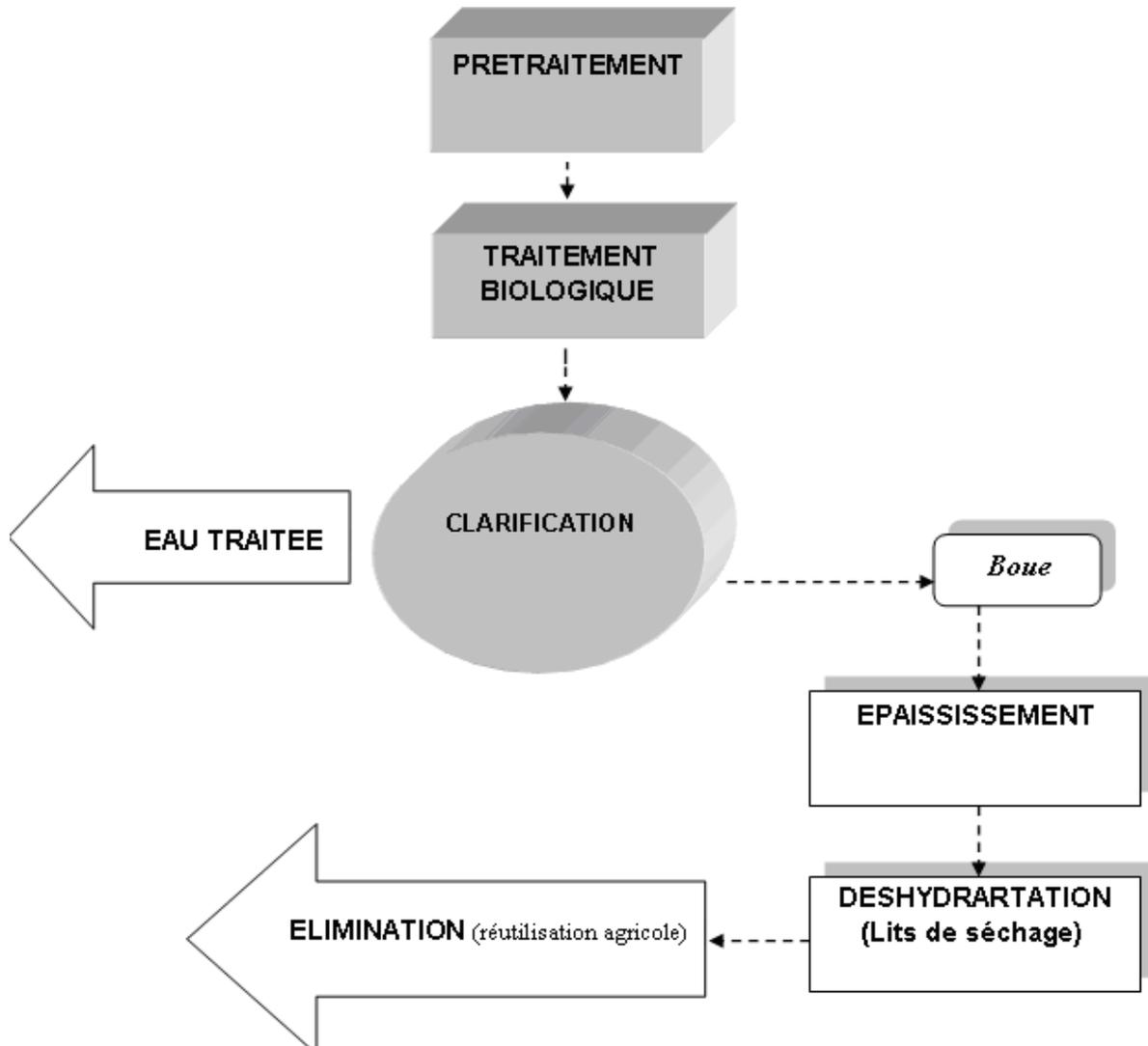


Figure 2: Schéma général de la filière de traitement des eaux usées de la station de Khenchela.

CHAPITRE III: ETUDE DE LA FIABILITÉ D'UNE STATION DE TRAITEMENT DES EAUX USEES

III. 1 INTRODUCTION

Qu'est-ce qu'une norme de performance que la réglementation signifie?

Par exemple, si une norme de rejet des effluents traités est fixée à 20 mg/L, est-ce que cela signifie-t-il que le système de traitement devrait respecter cette norme tout le temps? La moitié du temps? Quatre-vingt-dix-neuf pour cent du temps?

L'étude de la fiabilité fournit le langage et les concepts pour répondre à ces questions et d'autres au sujet de la performance d'un système de traitement des eaux usées.

La fiabilité des composants spécifiques d'un système de traitement des eaux usées est essentielle pour la protection de la santé humaine et du milieu récepteur, cette tâche n'est garantie que par le respect des normes légales de rejet.

Les outils de fiabilité aident aussi à améliorer la conception initiale de nouveaux systèmes, aident les opérateurs dans la résolution des problèmes quotidiens rencontrés durant l'exploitation, et fournissent aux dirigeants et les organismes de réglementation des informations essentielles pour diagnostiquer les points sensibles et / ou de localiser les parties du procédé de traitement nécessitant une amélioration appropriée. Les outils présentés dans ce chapitre permettent d'explorer chacun de ces domaines. Ces outils ont été testés dans des applications de la vie réelle, et la plupart ont été appliqués dans les applications de traitement des eaux usées.

Il est important de noter que la suite d'outils présentés dans ce chapitre est une sélection de deux méthodes utilisées dans l'industrie de traitement des eaux pour évaluer la fiabilité de ces systèmes.

III. 2 NOTION DE FIABILITE

III.2.1 Définition de la notion de fiabilité

La fiabilité d'un système peut être définie comme la capacité d'exécuter les exigences spécifiées libres de défaillance ou de la probabilité d'une performance adéquate pour une période minimum de temps spécifié dans des conditions spécifiées (Chorafas, 1960 in Saleem, 2007), ou "la probabilité d'une performance adéquate pour une période de temps spécifiée dans les conditions spécifiées" ou "le pourcentage du temps que les concentrations des effluents traités sont conformes aux exigences du permis précisées" (Metcalf & Eddy 2003).

Dans cette section, les normes de rendement utilisées comme exemples sont établis, et plusieurs outils sont décrits qui peut être utilisé pour déterminer si les systèmes répondent à ces normes.

III.3 METHODE DES COURBES DE DEFAILLANCE

Combien de systèmes de traitement des eaux usées ne parviennent plus à assurer son rôle dans un an? Combien de composants s'usent après un certain temps de service et, si oui, quelle est leur durée de vie utile ? Quel est le rôle du manque d'entretien régulier des systèmes de traitement des eaux usées dans les taux d'échec ?

La méthode des courbes de défaillance (aussi appelés courbes de décroissance) aident à répondre à ces questions.

Une courbe de défaillance illustre le nombre d'unités (systèmes ou composants) dans une population qui ne parviennent pas à un moment donné de la durée de vie de cette population, et il donne par la probabilité de défaillance. Une fois les attributs de base des courbes de défaillance sont établis, alors le temps moyen de fonctionnement avant panne (MTTF), le temps moyen entre pannes (MTBF) et d'autres mesures potentiellement utiles peuvent être facilement calculés.

Les courbes de défaillance sont tracées par l'analyse de données actualisées des échecs, les données requises pour chaque échec sont, au minimum, des informations sur le moment où l'appareil a été installé, quand il a échoué (si jamais). D'autres informations supplémentaires sur l'unité pourraient également être importantes. L'information supplémentaire contribue à définir le groupe d'unités devant être analysés. Par exemple, Hudson (1986) soutient que le lieu et la date de la construction du système (et donc la réglementation en vertu de laquelle le système a été conçu et construit), ainsi que le type de sol sur lequel le système est installé, donnent un pouvoir prédictif suffisant pour le taux d'échec.

Hudson (1986) affirme que l'ajout de plusieurs variables à la définition d'un groupe augmente la complexité des calculs, et ainsi augmente le pouvoir prédictif de l'analyse.

Les changements technologiques depuis les travaux d'Hudson en 1986 peuvent le rendre plus attractif d'utiliser plusieurs variables dans la définition d'un groupe. Les progrès dans le matériel et les logiciels permettent de réaliser des calculs compliqués pour réaliser une prédiction du taux d'échec. Une grande variété de systèmes de traitement a été utilisée, et ils peuvent avoir sensiblement différentes courbes de défaillance.

Il n'y a pas de moyens généralement reconnus pour définir un groupe des systèmes de traitement des eaux usées à des fins d'analyse de la courbe de l'échec. Si des informations détaillées telles que le type du système, le nom du concepteur, le nom de l'installateur, la fréquence d'entretien, ou d'autres informations est associé à d'autres données sur chaque système, il devient alors possible d'effectuer des analyses pour savoir si chaque facteur à un

effet significatif sur le rendement du système de traitement et par conséquent peut être incluse dans la définition du groupe. Sans informations sur les modes de défaillance.

III.3.1 Les différentes formes de courbes de défaillance

Moubray (1997) dans ces travaux présente des exemples de distributions de taux de défaillance dans le temps pour les systèmes ou les composants d'un système (voir Figure 3). Une discussion sur les différentes courbes et leurs implications pour les procédures de maintenance est présentée pour mettre à jour l'importance de cette analyse.

Dans toutes ces courbes de défaillance, l'axe des abscisses représente le temps et l'axe des ordonnées représente la probabilité conditionnelle de l'échec.

La probabilité conditionnelle de l'échec (défaillance) est définie comme la probabilité pour qu'un membre de la population, au début d'une période de temps (par exemple, une année) échoue à la fin de cette période de temps (Barlow et Proschan, 1996; Moubray, 1997).

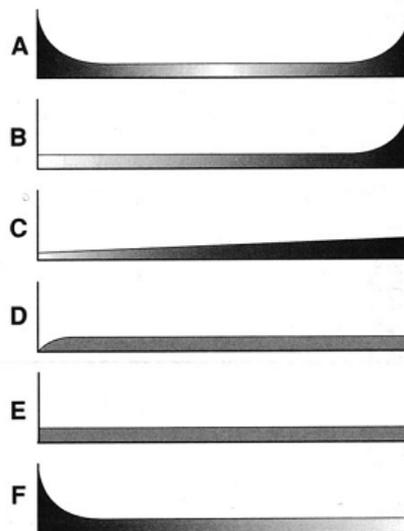


Figure 3: Les différents courbes de défaillances (Moubray, 1997)

- **Modèle A**, la courbe dite de baignoire, présente un nombre relativement élevé d'échecs en début de vie d'une unité, suivie d'une période de défaillance à peu près constante, suivi d'un temps de port de départ. La courbe représentée dans le schéma A est une combinaison d'au moins deux modes de défaillance, celui du (modèle F), et celle affichée lors de la phase d'usure (modèle B).
- **Modèle B** s'applique aux unités qui ne parviennent pas à un nombre relativement faible sur une durée de vie utile, après quoi le taux d'échec augmente rapidement. Instituant la durée de vie utile d'une unité se fait en rassemblant suffisamment de données actuarielles pour tracer une courbe comme motif B. Les procédures d'entretien varient, en fonction du coût de la détection de défauts, conséquences d'un échec, et les coûts de remise en état ou de remplacement. Pour les unités avec des

conséquences élevés d'échec, le point de référence clé pour les tâches de maintenance de planification est la longueur de la durée de vie utile.

- **Modèle C** a une probabilité de plus en plus régulière d'échec, sans aucune fin précise de durée de vie utile. Il peut être associé à la fatigue du matériau. La pente, qui peut varier aura une forte influence sur la stratégie de maintenance choisi.
- **Modèle D** est similaire au modèle C, sauf qu'il ya une brève période de très faible probabilité de défaillance précoce dans la vie de l'unité.
- **Modèle E** montre un taux constant d'échec au cours du temps. La distribution de l'échec dans le temps ne suggère pas de temps critique pour les interventions de maintenance. En fait, dans de nombreuses industries, l'entretien réduit les taux d'échec.
- **Modèle F** le début de la courbe baignoire est caractérisée par un taux élevé de défaillance sans la période d'usure est suivie par une probabilité de défaillance constante ou en augmentation progressive. La maintenance réactive est généralement le moyen le plus efficace pour remédier aux défaillances précoces. Dans le traitement des eaux usées, les modèles F et A indiquent des problèmes due à l'implantation, la conception et / ou de construction, soit parce que des procédures standards n'ont pas été suivies ou qu'elles étaient insuffisantes.

La forme de la courbe de défaillance permet de déterminer lequel des mesures peuvent donner des informations importantes. Le temps moyen entre pannes (MTBF) est une mesure couramment discuté. Cette mesure peut donner des informations importantes sur la durée de vie utile d'une unité qui présente un modèle d'échec suivant le modèles A ou B. Par conséquent, les unités peuvent être ciblées pour des interventions (remplacement, des inspections plus fréquentes, ou d'autres mesures) au moment où l'usure de l'ensemble des composants du système comment à se manifester.

Pour les unités conformes aux autres modèles, qui n'ont pas un temps distinct d'usure, le temps moyen entre pannes (MTBF) ne donne pas d'information utile sur le moment d'accélérer les interventions. Le temps moyen entre pannes (MTBF) peut toutefois encore être utile dans le cycle de vie des calculs de coûts (Moubray 1997).

III.3.2 Analyse de la courbe de défaillance d'un groupe de système

Une première étape dans l'application de l'analyse de la courbe de défaillance des systèmes de traitement des eaux usées est de décider lequel des systèmes à analyser. Les réglementations régissant la conception et l'installation des systèmes de traitement des eaux usées ont changé à plusieurs reprises au fil des ans.

Hudson (1986) identifie un éventail de méthodes de prévision des taux d'échec. La méthode la plus simple, "vielle ux est inacceptable», implique un état 100% réussi pour les systèmes

installés avant une certaine date et nécessite le remplacement de tous les systèmes plus anciens. Comme le souligne Hudson (1986), l'application de cette méthode peut conduire à une meilleure efficacité des systèmes remplacés.

Selon la complexité du système de traitement étudié, une "Analyse statistique complète" peut mettre en évidence que d'autres modèles sont disponibles.

L'analyse statistique complète nécessite plus de données et une maîtrise des outils statistiques pour aboutir à d'autres méthodes de prévision.

Le calcul du taux de défaillance brute est indépendant du moment de la panne. La méthode recommandée par Hudson permet d'établir des courbes de défaillance. Toutefois, la division en groupes est aussi une première étape dans l'établissement de la forme des courbes de défaillance.

III.3.3 Détermination de la forme de la courbe de défaillance

Hudson (1986) donne un exemple hypothétique d'une population de 74 systèmes autonomes de traitement des eaux usées (fosses septiques) construits sur une période de six ans. Ces 74 systèmes comprennent un groupe qui leur permet d'être relativement par rapport à l'autre. Pour les fins d'Hudson, ce qui signifie qu'ils sont construits sous le même modèle et le même type de sol. Le tableau 1 résume les stations toujours en service à la fin de chaque année d'exploitation.

Tableau 11: Nombre de stations construit et nombre de stations en service à la fin de chaque année d'exploitation (Hudson, 1986)

Année de mise en service	1975	1976	1977	1978	1979	1980	Total
<i>Systèmes en service</i>	10	8	17	22	5	12	74
1 ^{ère} Année de service	9	8	17	19	5	10	68
2 ^{ème} Année de service	8	8	15	19	5		55
3 ^{ème} Année de service	8	7	14	17			46
4 ^{ème} Année de service	8	7	13				28
5 ^{ème} Année de service	7	6					13
6 ^{ème} Année de service	7						7

A partir des informations ci-dessus, Hudson construit un tableau indiquant le taux d'échec de chaque année de service (tableau 12).

Tableau 12: Tableau récapitulatif du taux d'échec (Hudson, 1986).

	Systèmes en service	Défaillances	Taux de défaillance
1 ^{ère} Année de service	74	6	0,0811
2 ^{ème} Année de service	68	3	0,0441
3 ^{ème} Année de service	55	4	0,0727
4 ^{ème} Année de service	46	1	0,0217
5 ^{ème} Année de service	28	2	0,0714
6 ^{ème} Année de service	13	0	0,0000

Un graphique de la probabilité conditionnelle de l'échec (le taux d'échec de chaque année) est présenté sur la figure 4.

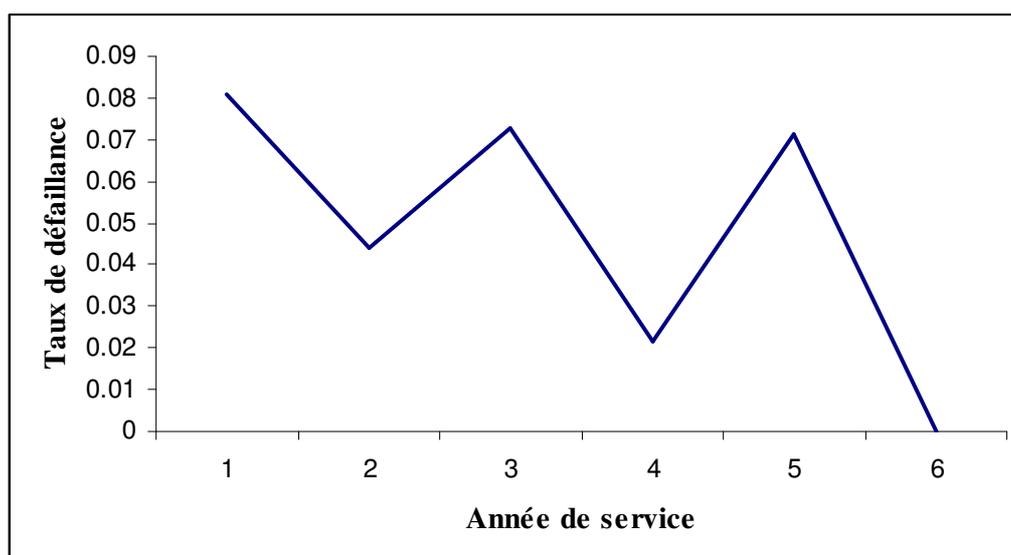


Figure 4: Probabilité conditionnelle de défaillance en fonction du temps de service (Hudson, 1986).

L'analyse de la courbe n'est pas aussi claire que les modèles de courbe de défaillance (modèle A et F) présenté à la figure 3. Avec 74 systèmes et après 6 ans de service, on constate que la population est trop petite et/ou le temps d'observation est trop court pour permettre la détermination de la forme de la courbe de l'échec.

Pour répondre à la question pour les données au cours de cette courte période de temps de service, Hudson (1986) demande si le taux d'échec moyen est supérieur au cours des trois premières années d'exploitation qu'au cours des trois dernières années d'exploitation.

Le taux d'échec lors des trois premières années d'exploitation est le nombre de défaillances au cours de cette période divisée par le nombre des systèmes en services:

$$\frac{6+3+4}{74+68+55} = 0,066 = 6,6\%$$

Utilisant le même procédé pour déterminer le taux d'échec des trois dernières années d'exploitation:

$$\frac{1+2+0}{46+28+13} = 0,034 = 3,4\%$$

Un taux de d'échec 6,6% lors des trois premières années d'exploitation est plus élevé que le taux d'échec de 3,4% des trois dernières années d'exploitation. Même un taux d'échec apparemment deux fois plus élevé pendant la première moitié de la durée de service que la deuxième moitié n'est pas statistiquement significatif avec cette petite quantité de données. Cette absence de signification statistique montre l'importance d'inclure un grand nombre d'années d'exploitation dans l'étude pour assurer un échantillon de taille suffisante pour les tests statistiques

La signification statistique est plus facile à réaliser avec de plus grandes quantités de données, mais il est également important de voir comment les échantillons sont distribués. Pour une étude statistique significative, il est important d'avoir accès aux grandes bases de données (Groves et al., 2005; Converse et Nordheim,2004).

III.4 METHODE DU COEFFICIENT DE FIABILITE (C.d.F)

La notion de **fiabilité** peut être employée pour prévoir l'efficacité d'un processus de traitement des eaux usées, que ce soit en phase de conception ou en exploitation a été développée par Niku et al. (1979) qui ont appliqué le concept de fiabilité pour évaluer l'efficacité de 37 stations de traitement des eaux usées aux USA utilisant le procédé de boue activée pour former une base de corrélation statistique.

L'analyse des données avait eu la conclusion que la distribution log-normale la DBO₅ et les MES pour l'effluent d'entrée. Cette distribution peut être employée pour prévoir l'efficacité, la qualité de l'eau traitée, et la fiabilité des stations de traitement des eaux résiduaires.

III.4.1 Concept de fiabilité dans une station de traitement des eaux usées

Les principes fondamentaux de la technologie de fiabilité peuvent être appliqués à la quantification de la probabilité que les événements indésirables peuvent se produire. En conséquence, l'analyse de la fiabilité des équipements de traitement des eaux résiduaires permet à l'ingénieur d'exploiter la structure statistique des données des effluents a l'entrée et à la sortie de la station en prévoyant la probabilité des événements indésirables.

La nature variable des caractéristiques des eaux usées, et de leurs qualités pendant le cycle de vie, ainsi que la conception de la station, peut résulter en fluctuations des efficacités prévues lors de la conception. Cependant, viser des niveaux de rejet des effluents devraient être choisis ont se basant sur la probabilité admise de réaliser uniformément ces derniers valeurs (Ellis et al., 1993).

Mathématiquement, une installation de traitement est complètement fiable s'il n'y a aucun échec dans le processus d'épuration (par exemple, violations des normes de rejet). L'échec d'un processus de traitement est mentionné quand les normes de rejet de la station d'épuration dépassent les normes de rejet standard mentionnées par la législation en vigueur. (Niku et al., 1979) a simplifié l'échec par l'équation suivante :

$$E = C_{\text{sortie}} > C_{\text{Normes rejet}}$$

E: échec (défaillance)

C_{sortie} : concentration des différents paramètres physico-chimiques de l'eau rejetée après traitement

$C_{\text{Normes rejet}}$: normes légales de concentration des différents paramètres physico-chimiques pour qu'une eau soit rejetée dans le milieu naturel.

Dans une définition technique, le concept essentiel de la fiabilité est la « *probabilité de succès* » ou la « *probabilité d'exécution proportionnée* », qui est le pourcentage de temps que la concentration des eaux rejetées répond aux exigences légale (Niku et al., 1979) :

$$M = 1 - P(E)$$

Où:

M = Maintenabilité

La valeur de M est égale à:

$$M = 1 - P(C_{\text{sortie}} > C_{\text{Normes rejet}})$$

P(E) est extrêmement sensible à la fonction de répartition de la probabilité de la concentration de l'effluent. Ainsi, pour calculer la maintenabilité (M), la distribution de la concentration de l'effluent devrait être modélisée. Pour effectuer une telle analyse, les données des concentrations des différents paramètres physico-chimiques de l'effluent doivent adapter une distribution donnée, pour cela on devra utiliser plusieurs techniques d'analyses statistiques pour la détermination de la loi de distribution appropriée.

III.4.2 Evaluation de la fiabilité d'une station de traitement des eaux usées par la méthode du coefficient de fiabilité (C.d.F)

La fiabilité d'une station de traitement est basée sur la connaissance du comportement du processus. (Niku et al., 1979) a réalisé une analyse des données de 37 stations de traitement des eaux usées et a abouti à la conclusion suivante:

Pour les paramètres de DBO₅ et les MES de l'effluent à l'entrée de la station, la distribution est le choix adéquat pour modéliser ces paramètres, et donc, la distribution log-normale peut être utilisée pour prédire la performance de la qualité des effluents et la fiabilité des installations de traitement des eaux usées.

En raison des variations dans les performances de la qualité des effluents, une station de traitement doit être conçue pour respecter une concentration moyenne de l'effluent en dessous des normes de rejet. La question est, quelle valeur moyenne garantit une concentration de l'effluent constamment inférieure à une norme avec une certaine fiabilité?

Le coefficient de fiabilité peut être introduit en relation avec le rapport des valeurs moyennes (valeur de dimensionnement) à la norme de rejet qui doit être réalisé sur une base de probabilité. Par exemple, pour 90% du temps de fonctionnement du procédé de traitement, le procédé doit être conçu pour atteindre une valeur moyenne m_x obtenue à partir de l'équation suivante:

$$m_x = (CdF) \times C_{NR}$$

Avec:

m_x : Moyenne de la concentration du paramètre considéré;

CdF: Coefficient de fiabilité;

C_{NR} : Concentration de la norme de rejet du paramètre considéré

$$m_x = (Cv^2 + 1)^{1/2} \times Exp\{-Z_{1-\alpha} \times [\ln(Cv^2 + 1)]^{1/2}\} \times (C_{NR})$$

Avec:

$$CdF = (Cv^2 + 1)^{1/2} \times Exp\{-Z_{1-\alpha} \times [\ln(Cv^2 + 1)]^{1/2}\}$$

Le coefficient de fiabilité CdF rapporte la valeur moyenne m_x constituant le niveau de fiabilité du paramètre de qualité X. pour un niveau de $(1-\alpha)$, on note que CdF est exprimé en fonction des propriétés des données et non pas du logarithme des données. Deux paramètres statistiques sont utilisés dans la détermination du coefficient de fiabilité:

1. Le coefficient de variation Cv qui est le rapport de l'écart type (σ_x) sur la moyenne arithmétique m_x

$$Cv = \frac{\sigma_x}{m_x}$$

2. Les centiles $Z_{1-\alpha}$ de la distribution normale : c'est le nombre de l'écart type par lequel X diffère de la moyenne

$$Z_{1-\alpha} = -\frac{\ln\left[\frac{m_X}{C_{NR}} \times (Cv^2 + 1)^{-1/2}\right]}{[\ln(Cv^2 + 1)]^{1/2}}$$

Les différentes valeurs des percentiles pour une distribution normale standard sont représentées dans le tableau 13.

Tableau 13: Valeurs des centiles d'une distribution normale standard (Niku et al., 1981)

Probabilité cumulative (1- α)	Centiles $Z_{1-\alpha}$
99.9	3.090
99	2.326
98	2.054
95	1.645
92	1.405
90	1.282
80	0.842
70	0.525
60	0.253
50	0

Les valeurs du coefficient de fiabilité (CdF) pour les concentrations des effluents pour les différents coefficients de variation à différents niveaux de fiabilité sont présentées dans le tableau 14.

Tableau 14: Coefficient de fiabilité (CdF) en fonction du coefficient de variation (Cv) et du taux

Taux de fiabilité	50%	80%	90%	92%	95%	98%	99%	99.9%
Cv								
0.3	1.04	0.81	0.71	0.69	0.64	0.57	0.53	0.42
0.4	1.08	0.78	0.66	0.63	0.57	0.49	0.44	0.33
0.5	1.12	0.75	0.61	0.58	0.51	0.42	0.37	0.26
0.6	1.17	0.73	0.57	0.54	0.47	0.37	0.32	0.21
0.7	1.22	0.72	0.54	0.50	0.43	0.33	0.28	0.17
0.8	1.28	0.71	0.52	0.48	0.40	0.30	0.25	0.15
0.9	1.35	0.70	0.50	0.46	0.38	0.28	0.22	0.12
1.0	1.41	0.70	0.49	0.44	0.36	0.26	0.20	0.11
1.2	1.56	0.70	0.46	0.41	0.33	0.22	0.17	0.08
1.5	1.80	0.72	0.45	0.39	0.30	0.19	0.14	0.06

Les valeurs présentées dans le tableau 14 peuvent être utilisées dans la conception d'une station de traitement prévu pour des niveaux de fiabilité données.

Le coefficient de variation de concentration des effluents (Cv) doit être estimé selon une sélection de la valeur raisonnable de performances souhaitées.

Pour avoir une fiabilité de 95%, la concentration de la DBO₅ devra être inférieure à une certaine concentration C_{NR}, lorsque le coefficient de variation (Cv) de la station est estimé à 0,70, la station doit être conçu pour une valeur moyenne égale ou inférieure à 0,43xC_{NR}.(si

par exemple on a la concentration de norme de rejet $C_{NR} = 30\text{mg/l}$, alors on obtient la moyenne $DBO_5 = 0,43 \times 30 = 12,9 \text{ mg/l}$, même pour une norme moins strict $C_{NR} = 40\text{mg/l}$, la moyenne de la DBO_5 devra être inférieur a la valeur suivante ($0,43 \times 40 = 17,2 \text{ mg/l}$).

Comme on pouvait s'y attendre, la valeur moyenne de la concentration requise de l'effluent varie inversement avec le niveau de fiabilité et avec le coefficient de variation (C_v) de la concentration de l'effluent (Niku et al., 1979).

III.5 FACTEURS AFFECTANT LES PERFORMANCES D'UNE STATION DE TRAITEMENT DES EAUX USEES

L'un des problèmes commun aux stations de traitement des eaux résiduaires est que l'efficacité des systèmes n'est pas toujours au niveau prévu lors de la conception (Drakatos et al., 1992).

Selon (Metcalf et Eddy, 2003) ce problème est du a plusieurs facteurs affectant l'exécution est principalement, du a trois catégories de variabilité:

1. La variabilité des débits des eaux usées et de leurs caractéristiques;
2. La variabilité inhérente dans un processus de traitement de l'eau usée ou fiabilité inhérente;
3. La variabilité provoquée par les pannes mécaniques.

La variabilité de la charge polluante à l'entrée de la station dues à la croissance démographique, activité industriel, changements de régime, croissance du niveau de vie, nature cyclique des activités humaines, vieillissement du réseau d'assainissement, variabilité des conditions environnementales dues au caractère saisonnier de la température et précipitations, effets des chocs, niveau de rejet des matériaux toxique, l'échec mécanique du système, et les erreurs humaines causent fréquemment le processus renversement et instabilité (Niku et Schroeder, 1981 ; Berthouex et Fan, 1986 ; Ossenbruggen et al., 1987 ; Drakatos et al., 1992 ; Leitao et al., 2006).

Quelques fois, la conception incertaine donne des résulte typiquement de l'utilisation des eaux à traiter on se basant sur la moyenne des paramètres de qualités de eaux usées sans prendre en considération les conditions maximales (Ellis et al., 1993). Tous ces paramètres ont des différents impacts sur l'efficacité du processus. Les échecs mécaniques et opérationnels sont d'autres sources importantes de variabilité dans le traitement des eaux résiduaires plantent l'exécution, mais ce ne sont pas inclus dans cette étude.

Il convient noter que l'échec d'un composant du processus de traitement ne doit en aucun cas empêcher la station de continuer sa mission et de respecter les normes de rejets légales (AWWA, 2012).

III.6 NORMES DE REJET DES EAUX USEES EN ALGERIE

L'eau usée traitée, récoltée à l'aval des systèmes d'assainissement urbains représente une eau renouvelable non conventionnelle, qui pourrait être une source attrayante et bon marché à employer en agriculture, au voisinage des centres urbains (Vasel, 2007). Toutefois, en raison de la nature variable de la composition de cette eau (sa charge en constituants minéraux, organiques et biologiques); sa réutilisation devrait être gérée soigneusement, surveillée et contrôlée par des spécialistes, afin d'éviter les risques et menaces potentiels sur les usagers, le sol et les cultures irriguées, ainsi que sur l'environnement dans son ensemble (Blumenthal, 1989).

En Algérie le volume global d'eaux usées rejetées annuellement est évalué à près de 800 millions de m³, dont 600 millions de m³ pour les seules agglomérations du Nord, sur ce volume global seul environ 30% sont traitées. Ces effluents, d'origine domestique, constitueraient une ressource d'eau importante qui peut être prise en considération pour diverses utilisations. Pour cela, il est primordial d'augmenter le nombre de stations d'épuration dotées d'équipements de traitements tertiaires qui permettraient un meilleur abattement de la pollution notamment azotée et phosphatée (Metahri, 2012).

Par ailleurs l'Algérie est classée parmi les pays hydro-sensibles confrontés à la rareté des eaux naturelles conventionnelles due à l'insuffisance et à l'irrégularité des précipitations dans le temps et dans l'espace. Le climat aride et semi-aride qui sévit sur une grande partie du territoire réduit également les disponibilités de cette ressource.

Le taux de satisfaction en eau en Algérie est en moyenne de 500 m³/habitant/an, qu'est inférieur à la norme internationale estimé à 1700 m³/habitant/an. Par ailleurs l'introduction et la réutilisation des eaux non conventionnelles sont devenues une alternative incontournable pour satisfaire les besoins qui ne cessent de croître et réserver les eaux conventionnelles pour les vocations nobles comme l'AEP et l'abreuvement des cheptels. A cet effet le recyclage des eaux usées traitées peut constituer une alternative attrayante à plusieurs titres (agricole, économique, sociale et environnementale). La législation algérienne d'après le décret N°06-141 du 19/04/2006 stipule que les normes de rejet des eaux usées traitées doivent respecter les limites suivantes:

Tableau 15: Normes Algériennes des rejets d'eaux usées traités (JORA, 2006)

Paramètres de pollution	Limites acceptables
DBO ₅	35 – 40 mg/l
DCO	120 – 130 mg/l
MES	35 – 40 mg/l
Azote Kjeldahl	30 – 40 mg/l
Phosphore total	10 – 15 mg/l

III.7 EVALUATION DE LA FIABILITE D'UNE STATION D'EPURATION EN EMPLOYANT D'AUTRES APPROCHES

La stabilité d'un processus de traitement a été employée pour décrire la conformité d'efficacité d'une station de traitement basée sur les mesures statistiques des différents paramètres de qualité de l'eau usée, on a constaté que l'écart type est l'indicateur le plus approprié de la stabilité de processus. Niku et Schroeder (1981) ont trouvé une forte corrélation linéaire entre l'écart type et d'autres indicateurs de stabilité.

La stabilité d'une station d'épuration a été défini comme étant la valeur d'écart type au-dessous du quelle, une station est considérée comme statistiquement stable, dépassant cette valeur limite, la station est considérée comme statistiquement instable. Ils ont conclu que les stations de traitement ayant un écart type de moins de 10 mg/l pour la DBO₅ et de 70 mg/l pour les MES peuvent être statistiquement considérées comme stations stables (Niku et Schroeder, 1981 ; Niku et al., 1982).

L'instabilité d'une station, ou la variabilité excessive de la qualité des effluents à l'entrée, peut faire échouer le processus de traitement et ne plus répondre uniformément aux normes de rejet de la DBO₅. La plupart des violations des normes des rejets peuvent être liées à l'absence de possibilités à régulariser les variabilités excessives de la qualité des eaux usées brutes à l'entrée de la station (Ossenbruggen et al., 1987).

III.6 CONCLUSION

La méthode des courbes de défaillance est un outil simple qui trace la fréquence de défaillance des composants du système ou des systèmes entiers dans le temps. Cet outil permet à l'opérateur de comprendre quand certains éléments sont susceptibles d'échouer et de planifier en conséquence leur maintenance ou leur remplacement.

Grâce à cet outil de gestion, une réduction des coûts globaux se manifeste sensiblement par des pannes en moins du système, et une amélioration du niveau de protection de l'environnement (Etnier et al., 2005).

Malheureusement, lors de l'élaboration de cette étude, on a constaté l'absence flagrante des données historique des défaillances, en effet les gestionnaires de la station de Khenchela ne gardent aucune copie de registre ou support informatique sur lequel figure les interventions et les causes des défaillances des composants du systèmes rendant ainsi l'élaboration des courbes de défaillances du système de traitement une tâche impossible.

La disponibilité des données d'auto-surveillance de la station de traitement des eaux usée de la ville de Khenchela englobant les données d'entrée-sortie, des débits, ainsi que les concentrations des différents paramètres de qualité (MES, DCO et DBO₅), on permet l'orientation de notre recherche pour l'évaluation des performances de la station par l'utilisation de la méthode du coefficient de fiabilité.

CHAPITRE IV : EVALUATION DE LA FIABILITE DE LA STATION D'EPURATION DE KHENCHELA PAR MODELE PROBABILISTE

Ce chapitre est divisé en neuf sections, affaire à la détermination et l'analyse de la fiabilité du procédé de traitement des eaux usées de la station de Khenchela.

D'abord un aperçu sur l'organisation des données, la deuxième section évoque la recherche des erreurs dans les données, la troisième section est consacrée la correction des données.

La quatrième section dédiée au traitement statistique des données d'entrée et de sortie de la station, de l'étude de la variabilité des paramètres de pollution.

Le choix de la loi de distribution approprié des données d'entrée et de sortie de la station fait l'objet de la cinquième section. L'évaluation de la fiabilité globale de la station de Khenchela est présentée dans la sixième section.

L'application de la méthode du coefficient de fiabilité pour déterminer le taux de fiabilité du procédé de traitement des eaux usées de la station est l'objet de la septième section.

L'évaluation du taux de défaillance des procédés de traitement des eaux usées est présentée dans la huitième section. En conclusion, une présentation des différents paramètres influençant les performances de la station sont présentées dans la dernière section.

IV.1 ORGANISATION DES DONNEES

La compréhension des processus intervenant dans cycle de traitement des eaux usées ainsi que l'étude des différents paramètres liés procédé, de leurs variations spatiales et temporelles nécessitent de disposer de données. Celles-ci sont essentielles et constituent un préalable à toute analyse statistique, que ce soit dans le but de procéder à une étude du procédé de traitement de l'eau usée, d'impacts environnementaux ou pour procéder à l'évaluation de la fiabilité de la station.

De façon générale, pour permettre le passage de l'acquisition des données à leur utilisation effective dans le cadre d'une analyse statistique on distingue les étapes suivantes : acquisition, traitement, contrôle et validation, organisation. Une collaboration entre l'université d'Oum El-Bouaghi et la station de traitement des eaux usées de la ville de Khenchela a permis de numériser les données d'auto-surveillance de la station durant les 5 dernières années sous format EXCEL.

IV.1.1 Acquisition des données

L'acquisition de données consiste à procéder, par le biais d'un instrument de mesure, à acquérir de l'information (par exemple : débit d'eau usée à l'entrée de la station, mesure des

différents paramètres de pollution, température, pH etc...). Le procédé peut être automatisé ou non. Ceci aura une influence sur le type d'erreurs que l'on peut commettre.

IV.1.2 Traitement primaire des données

Les données acquises précédemment nécessitent souvent un traitement préalable – ou traitement primaire afin de les rendre pertinentes et exploitables. Il s'agit pour l'essentiel de la conversion de la mesure effectuée en une grandeur qui soit statistiquement significative (par exemple : transformation d'impulsions électriques en débits, génération de données à pas de temps constant à partir de mesures effectuées à pas de temps variables etc).

Le traitement des données inclut aussi le contrôle primaire des données qui comprend les contrôles de cohérence à l'exclusion de tous traitements statistiques. Il s'agit par exemple, dans le cas d'une acquisition manuelle des données, de les convertir en fichiers informatiques. Dans ce cas, on procède généralement à une double saisie des données puis les fichiers sont comparés afin de déceler d'éventuelles erreurs de saisie.

IV.1.3 Contrôle des données

Avant de pouvoir exploiter les données et bien qu'elles soient dans un format adéquat, il importe de contrôler la fiabilité et la précision de ces dernières. Le contrôle permet de valider les données avant leur organisation au sein d'une base de données pour leur mise à disposition à des fins opérationnelles.

IV.1.4 modèle numérisé des données

Au vu de l'importance quantitative et qualitative des données, il importe de les organiser avec soin. Ceci se fait à partir d'un corpus de documents originaux (registre d'exploitation, rapport de gestion informatisé) constituant les archives qui sont en règle générale accessibles uniquement à un personnel spécifique (responsable d'exploitation, responsable du laboratoire...). La traduction des archives sous la forme de fichiers de base génère les "fichiers en l'état" et fournit une indication sur la provenance de la donnée (mesure, calcul), ainsi que sur sa qualité (fiable, complète ou non) et sa précision. Enfin, on constitue un fichier de travail provisoire permettant une visualisation des données et permettant de procéder aux différents tests de qualité et de précision des données qui seront développés tout au long de ce chapitre. L'élément ultime de cette chaîne opératoire est la constitution des fichiers d'auto-surveillance englobant tous les paramètres indicateurs du procédé de traitement pouvant être publiés et les rendre exploitables par les dirigeants de la station.

IV.2 RECHERCHE DES ERREURS DANS LES DONNEES D'AUTO-SURVEILLANCES

La constitution d'une série de valeurs, constituant un échantillon au sens statistique du terme, est un processus long, parsemé d'embûches, et au cours duquel de nombreuses erreurs, de nature fort différentes, sont susceptibles d'être commises.

Des erreurs peuvent en effet être perpétrées lors de l'une ou de l'autre des quatre phases du déroulement classique des opérations, à savoir : la mesure ; la transmission de l'information ; le stockage de l'information ; le traitement de l'information (prétraitement et analyse). Il est donc indispensable, avant d'utiliser des séries de données, de se préoccuper de leur qualité et de leur représentativité en utilisant diverses techniques en général de type statistique ou graphiques.

Une erreur de mesure est définie comme étant la différence entre la vraie valeur (qui est l'idéal recherché, mais qui n'est en principe et malheureusement jamais connue) et la valeur mesurée. Il est commode, tant pour les présenter que pour différencier la façon de les aborder, de considérer deux types d'erreur :

Les erreurs aléatoires (ou accidentelles): Elles affectent la précision des données et sont non corrélées. Ce type d'erreur est dû à des raisons nombreuses et variées, généralement inconnues, affectant différemment chaque mesure individuelle. Ces erreurs étant inévitables, il faut en estimer l'importance afin de pouvoir en tenir compte lors de l'évaluation de l'incertitude finale. Dans la mesure du possible, la technique de mesure induisant les erreurs aléatoires les plus faibles devrait être préférée.

Les erreurs systématiques: Elles affectent la fiabilité des données et sont totalement corrélées. On parle aussi d'inconsistance. Supposons qu'aucune erreur aléatoire n'affecte les mesures. La différence entre la vraie valeur et la valeur mesurée, si elle existe, est alors due à une erreur systématique. L'origine des erreurs systématiques est le plus souvent liée à la calibration de l'appareil de mesure qui n'est pas parfaite ou à un phénomène extérieur qui perturbe la mesure (erreur d'appareillage, changement d'observateur...) (Saporta, 2006).

IV.3 CORRECTION DES DONNES D'AUTO-SURVEILLANCES

Selon la nature des erreurs constatées ou supposées la recherche de ces dernières fait appel à différentes techniques et méthodes

- « in situ » qui consiste à vérifier sur place la manière dont les données ont été organisées, traitées et/ou transformées.
- Investigation de bureau qui consiste à vérifier la chaîne de traitement de la mesure/donnée à chaque étape de son élaboration, tout comme la manière dont on a constitué les séries de données soumises à contrôle et/ou sauvegarde.
- Investigation statistique qui, à l'aide d'outils spécifiques, permet de mettre en évidence certaines erreurs ou inconsistance. Ces techniques efficaces ont largement été

utilisées dans la pratique professionnelle et se basent sur des hypothèses spécifiques qu'il convient de bien connaître.

Hypothèses de l'analyse statistique:

Les calculs statistiques sont basés sur un certain nombre d'hypothèses qui doivent en principe être vérifiées. Parmi celles-ci, citons :

1. Les mesures reflètent les vraies valeurs: Cette hypothèse n'est malheureusement jamais réalisée en pratique, du fait des erreurs systématiques ou aléatoires.
2. Les données sont consistantes: Aucune modification dans les conditions internes du système n'intervient durant la période d'observation (position du pluviomètre, procédures d'observation, observateur unique).
3. La série de données est stationnaire: Les propriétés de la loi statistique qui régit le phénomène (moyenne, variance ou moments d'ordre supérieur) sont invariantes au cours du temps.

Les données sont homogènes: Une série de données est réputée non homogène lorsque:

- Elle provient de la mesure d'un phénomène dont les caractéristiques évoluent durant la période de mesure; le phénomène est alors dit "non-stationnaire" (par exemple: variations climatiques, variations du régime des débits). Il est également possible d'observer des signes d'une non stationnarité apparente lorsque l'électronique intégrée à l'équipement de mesure présente une dérive temporelle ou lors du changement de l'observateur.
- La série de données est aléatoire et simple: Le caractère aléatoire et simple d'une série d'observations est une hypothèse fondamentale pour l'analyse statistique. Un échantillon aléatoire signifie que tous les individus de la population ont la même probabilité d'être prélevés. Un échantillon simple signifie que le prélèvement d'un individu n'influe pas la probabilité d'apparition des individus suivants. Autrement dit, si toutes les observations de la série sont issues de la même population et qu'elles sont indépendantes entre elles, la série est alors aléatoire et simple. La non vérification du caractère aléatoire et simple peut avoir plusieurs causes, parfois simultanément. Ces causes se groupent en deux catégories, les défauts d'auto corrélation d'une part (caractère non aléatoire des séries) et les défauts de stationnarité du processus d'autre part (dérive à long terme et dérive cyclique).
- La série doit être suffisamment longue: La longueur de la série influe sur les erreurs d'échantillonnage, notamment sur le calcul des moments d'ordre supérieurs donc sur les tests inhérents à leur fiabilité (Saporta, 2006).

IV.4 TRAITEMENT STATISTIQUE DES DONNES DE LA STATION D'EPURATION DE KHENCHELA

Les données recueillies ont été analysées statistiquement pour déterminer la nature et le type des fonctions de distribution qui représentent adéquatement les données observées et leurs paramètres caractéristiques suivant différents modèles de distribution.

Le choix de la distribution hypothétique pour ajuster les données est important afin d'obtenir la fiabilité. Parfois, les analystes peuvent utiliser leur connaissance des phénomènes physiques pour choisir une distribution pour modéliser les données. Parfois l'expérience peut suggérer le choix d'une autre distribution. Dans la situation où il n'y a pas d'expérience ou de théorie précédente (comme notre situation) on suggère une répartition qui décrit les données.

La théorie sous-jacente de tests d'adéquation sort du cadre de cette étude, mais une courte description des composants de base est décrite ci-dessous.

D'Agostino et Stephens (1986) fournissent l'explication suivante de ces tests:

"Supposons qu'un échantillon aléatoire donnée de taille n est $X_1; \dots; X_n$ soit $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ les statistiques d'ordre. Supposons aussi que la distribution de X est $F(x)$. La fonction de répartition empirique (FRE) est $F_n(x)$ définie par:

$$F_n(x) = \begin{cases} 0 & \dots & x < X_{(1)}; \\ i/n & \dots & X_{(i)} \leq x < X_{(i+1)}; i = 1; \dots; n-1, \\ 1 & \dots & X_{(n)} \leq x. \end{cases}$$

Ainsi $F_n(x)$ est une fonction en escalier, calculée à partir des données, lorsque x augmente, il fait un pas vers le haut d'une hauteur de $1/n$ que chaque observation de l'échantillon est atteinte. Pour tout x , $F_n(x)$ enregistre le nombre d'observations inférieure ou égale à x , tandis que $F(x)$ est la probabilité d'une observation inférieure ou égale à x . Nous pouvons nous attendre $F_n(x)$ pour estimer $F(x)$, et il est en fait un estimateur de $F(x)$, comme $n \rightarrow \infty$, $|F_n(x) - f(x)|$ tend vers zéro avec une probabilité. Empiriques fonction statistique de distribution (par exemple, D , A^2) mesure le décalage entre la fonction de répartition empirique (FRE) et une fonction de distribution donnée, et sont utilisés pour tester l'ajustement de l'échantillon à la distribution choisie. La distribution peut être complètement spécifiée ou peut contenir des paramètres qui doivent être estimés à partir des données.

Quelques exemples de statistiques des fonctions de distribution empiriques qui mesurent la distance entre $F_n(x)$ et $F(x)$ sont:

Test statistique de Kolmogorov-Smirnov (D)

$$D = \sup_x |F_n(x) - f(x)|$$

Test statistique de Kuiper (K):

$$K = \sup_x \{F_n(x) - f(x)\} + \sup_x \{F(x) - F_n(x)\}$$

Test statistique de Cramer-von Mises (W^2)

$$W^2 = n \int_{-\infty}^{+\infty} \{F_n(x) - F(x)\}^2 dF(x)$$

Test statistique d'Anderson-Darling (A^2)

$$A^2 = n \int_{-\infty}^{+\infty} \{F_n(x) - F(x)\}^2 \left[\frac{F(x)}{1-F(x)} \right]^{-1} dF(x)$$

Les formules de calcul pour les tests statistiques ci-dessus peuvent être trouvées en utilisant la transformation intégrale de probabilité (TIP), $Z = F(x)$; lorsque $F(x)$ est la vraie distribution de X , la nouvelle variable aléatoire Z est uniformément réparti entre 0 et 1. Ensuite, Z à la distribution fonction $F^*(z) = z$, $0 \leq z \leq 1$. Supposons qu'un échantillon X_1, \dots, X_n donne des valeurs $Z_i = F(X_i)$; $i = 1, \dots, n$, et $F_n^*(z)$ la fonction de distribution empirique des valeurs Z_i .

Les statistiques des fonctions de distribution empirique peuvent désormais être calculés à partir d'une comparaison de $F_n^*(z)$, avec $F^*(z)$:

$$F_n(x) - F(x) = F_n^*(z) - F^*(z) = F_n^*(z) - z$$

Cela conduit des formules de calcul suivantes:

$$D = \max \left\{ \max_i \left\{ \frac{i}{n} - Z_{(i)} \right\}, \max_i \left\{ Z_{(i)} - \frac{(i-1)}{n} \right\} \right\}$$

$$V = \max_i \left\{ \frac{i}{n} - Z_{(i)} \right\} + \max_i \left\{ Z_{(i)} - \frac{(i-1)}{n} \right\}$$

$$W^2 = \sum_i \left\{ Z_{(i)} - \frac{(2i-1)}{(2i)} \right\}^2 + 1/(12n)$$

$$A^2 = -n - (1/n) \sum_i (2i-1) \left[\ln Z_{(i)} + \ln \left\{ 1 - Z_{(n+&-i)} \right\} \right]$$

IV.4.1 Analyse statistique des données de la station de Khenchela

Cette étude est réalisé sur la base des données d'auto-surveillance de quarte années d'exploitation de 2009 jusqu'en 2012 pour les débits journaliers (Q), et les trois paramètres de pollution (DCO, DBO₅, et les MES) à l'entrée et des même paramètres de qualité des eaux rejetés, pour la bonne réalisation de cette étude une combinaison de trois logiciel de calcul statistique été nécessaire (EXCEL®, MINTAB®.16 et STATGRAPHICS®).

Les données des paramètres de qualité sont absents particulièrement les week-ends et les jours fériés, par contre les données concernant les débits d'entrée, sorties sont pratiquement complets sauf pour quelques mois. Le tableau 13 résume les statistiques des données comprenant le nombre d'observations, valeur maximum, valeur minimum, moyenne et l'écart type.

De 2009 à 2012, la moyenne des débits à l'entrée de la station de Khenchela étaient de 12268,50 m³/j, la station fonctionne au-dessous de sa capacité maximale de conception.

L'analyse de l'écart type annuelle de la DBO₅ à la sortie montre que la station de Khenchela est statistiquement instable pour les années 2009 et 2010, par contre pour les années 2011 et 2012 aucune conclusion ne peut être tirée dû au manque cruel de données. (Une station est statistiquement stable si l'écart type pour la DBO₅ des effluents est inférieur à 10 mg/l, et si l'écart type pour les MES des effluents est inférieur à 70 mg/l) (Niku et al., 1982).

Tableau 16: Synthèse de l'analyse statistique des données de la station d'épuration de Khenchela (2009-2012).

	Débit (m³/j)	MES (mg/l)	DCO (mg/l)	DBO₅ (mg/l)
Entrée de la station				
N ^{br} d'observations	972	776	217	209
Min.	888	33	33	70
Max.	19079	1162	1650	1332
Moy.	12268,50	216,11	561,40	244,90
Ecart-type	4350	118,45	308	137,43
Sortie de la station				
N ^{br} d'observations	974	542	247	97
Min.	836	10	13	10
Max.	14904	452	165	128
Moy.	9240,8	26,10	52,25	27,15
Ecart-type	3532,6	24,44	33,80	28,88

Pour les données d'entrée de la station, on constate que la valeur de l'écart type du débit est très importante, due principalement à la grande variabilité du débit d'entrée entre un débit minimum de 888 m³/j et un maximum de 19079 m³/j.

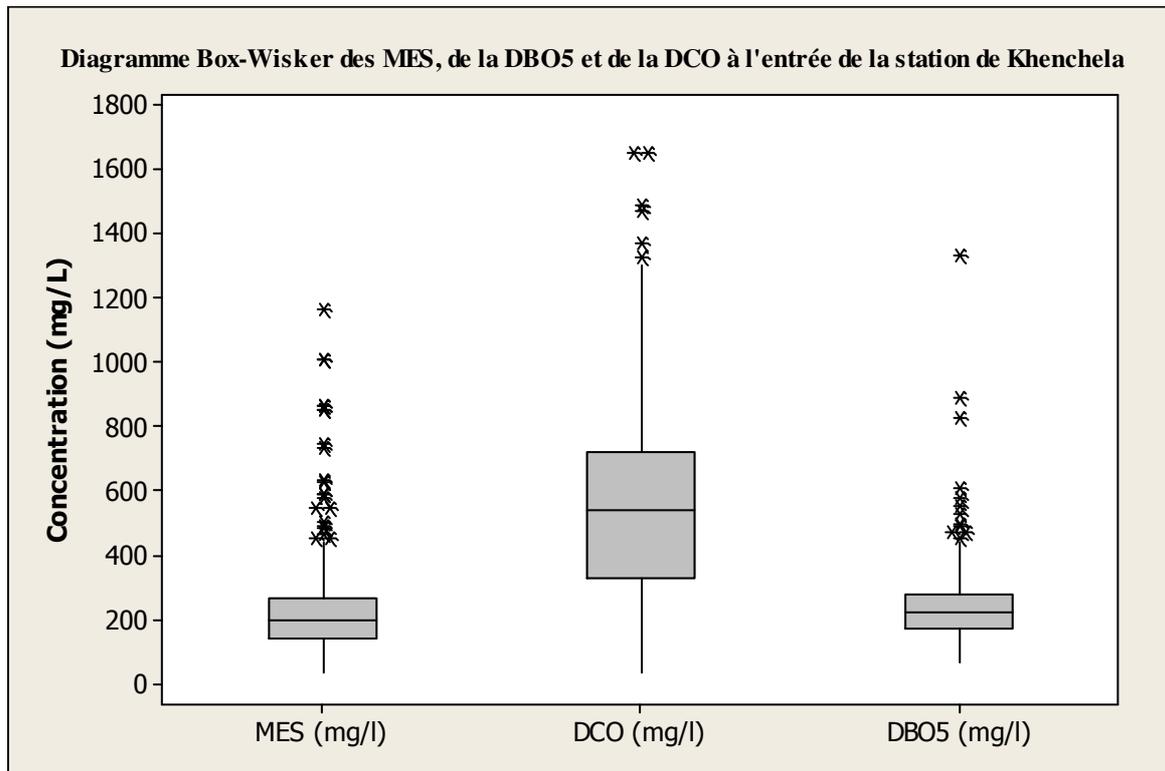
Même constat pour les autres paramètres de pollution, une valeur élevée de l'écart type observée est synonyme de grande variation, ce qui nous laisse conclure que la série d'observation est étalés.

IV.4.2 Etude de la variabilité des paramètres de pollution

Les données rassemblées ont montré que les valeurs des paramètres de pollution diffèrent de manière significative pour les eaux usées à l'entrée et à la sortie. Les moyennes des paramètres de pollution des eaux usées à traiter sont étendus entre 33-1650 mg/l pour la DCO, 70 - 1332 mg/l pour la DBO₅, 33 - 1162 mg/l pour les MES.

En outre, les moyennes des concentrations de l'effluent à la sortie sont étendus entre 13 - 165 mg/l pour la DCO, 10 - 128 mg/l pour la DBO et 10 - 452 pour les MES.

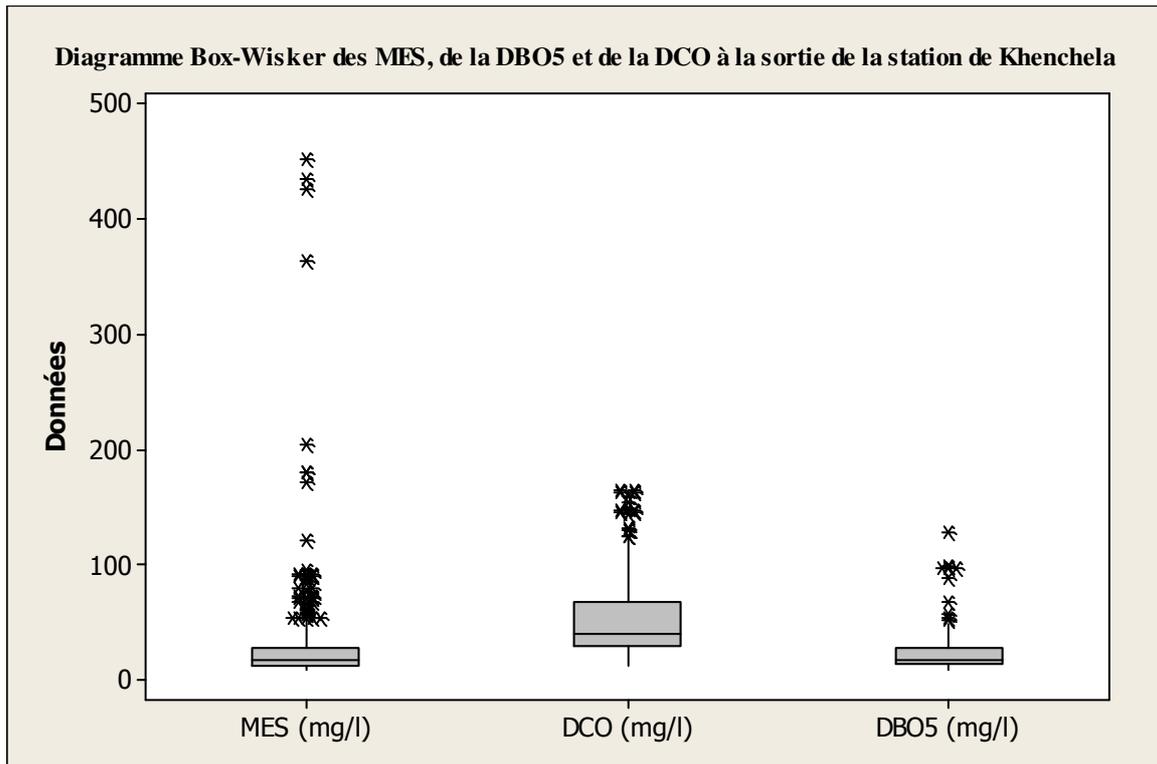
Pour montrer les niveaux et la variabilité différents dans la DCO entrée et sortie, la DBO, les MES, les différents données sont regroupées et illustrées sur la figure 5.



Les médianes (traits horizontaux à l'intérieur des boîtes), les diffusions (distances verticales), les annexes (points), et la gamme entre le 25^{ème} et le 75^{ème} percentile (distance entre les côtés supérieurs et inférieurs des boîtes).

Figure 5: Diagramme Box-Whisker des différents paramètres des eaux usées à l'entrée de la station de Khenchela (2009-2012)

Du diagramme de Box-Whiskers on peut constater que les paramètres MES et DBO₅ sont très variables et que leurs domaine de variation est très large.



Les médianes (traits horizontaux à l'intérieur des boîtes), les diffusions (distances verticales), les annexes (points), et la gamme entre le 25^{ème} et le 75^{ème} percentile (distance entre les côtés supérieurs et inférieurs des boîtes).

Figure 6: Diagramme Box-Whisker des différents paramètres des eaux traitées à la sortie de la station de Khenchela

Du diagramme de Box-Whiskers on peut constater que la concentration des MES est très variable et que son domaine de variation est très large.

IV.5 CHOIX DE LA LOI DE DISTRIBUTION APPROPRIÉE

Dans une analyse des données d'auto-surveillance d'une station de traitement des eaux résiduaires, il y a habituellement des données suffisantes pour caractériser la variabilité des chargements des différents paramètres caractérisant la qualité d'effluent qui doivent être traité. La variabilité des débits et des paramètres de qualité sont caractérisés avec des fonctions de densité de probabilité, le choix adéquat de la loi de distribution est important pour la détermination des différents paramètres statistiques.

Avant d'employer une distribution de probabilité pour représenter les données fondamentales d'une population de données, il est important d'examiner l'adéquation de cette distribution. Pour faire cela il faut étudier la qualité des tests d'ajustement. La qualité de ces tests statistiques convenables détermine s'il y a d'une différence significative entre une distribution de fréquence observée et une distribution de probabilité théorique présumé pour décrire la distribution observée (Levin, 1984).

Par ceci nous voulons dire qu'à quel point la population de l'échantillon de données est conformes à une distribution donnée. Cependant, les tests statistiques convenables sont divisés en deux types :

- Les tests paramétriques
- Les tests non paramétriques

On parle de *tests paramétriques* (les données observées issu d'une loi gaussienne) lors qu'on stipule que les données sont issues d'une distribution paramétrée. Dans ce cas, les caractéristiques des données peuvent être résumées à l'aide de paramètres estimés sur l'échantillon, la procédure de test subséquente ne porte alors que sur ces paramètres. L'hypothèse de normalité sous-jacente des données est le plus souvent utilisée, la moyenne et la variance suffisent pour caractériser complètement la distribution. Concernant les tests d'homogénéité par exemple, pour éprouver l'égalité des distributions, il suffira de comparer les moyennes et/ou les variances. En revanche *les tests non paramétriques* (les données observées ne sont pas issu d'une loi gaussienne) ne font aucune hypothèse sur la distribution sous-jacente des données. On les qualifie souvent de tests distribution free. L'étape préalable consistant à estimer les paramètres des distributions avant de procéder au test d'hypothèse proprement dit n'est plus nécessaire.

Les fonctions de densité de probabilité peuvent être caractérisées avec une fonction de probabilité de différents types, par exemple, normal, log-normal, gamma, logistique, Laplace, distributions de Weibull. Ces distributions pour des variables continues et chacune a ses caractéristiques particulières. Deux paramètres décrivent habituellement la distribution normale, à savoir la moyenne et l'écart type. Un paramètre détermine la forme de la distribution tandis qu'un autre détermine la diffusion des données (Wadsworth, 1990 ; Kottegoda et Rosso, 2009).

Malheureusement, les observations faites sur les différents variables (paramètres) de qualité d'eaux usées ne font pas souvent suivre une distribution normale. Puisque les variables telles que les niveaux de concentrations sont souvent positif (aucunes valeurs négatif), cette remarque font on sorte que la courbe oblique se décale vers la droite, incluent des observations extrêmes à des niveaux élevés.

En conséquence, la distribution log-normale est utilisée généralement dans le traitement des eaux résiduaires pour la caractérisation des effluents. Une justification qui fait penser que le processus donne à l'effluent une concentration ayant une loi de distributions log-normales (Niku et al., 1979) ou une loi de distribution du type gamma (Kottegoda et Rosso, 2009), le processus aléatoire fondamental est multiplicatif dans la nature et la fluctuation de la moyenne peut être proportionnel plutôt qu'additif. En d'autres termes, la concentration des effluents peut être prédite avec une équation non linéaire. Dans de telles circonstances, si les paramètres sont des variables aléatoires sans se soucier de leurs distributions, la distribution du produit, par exemple, la concentration de l'effluent, est log-normal (Burmester et Hall, 1997).

Niku et al. (1982) ont examiné statistiquement les concentrations d'écoulement des effluents de processus de filtre biologique et de boue activée, et ils ont constaté que les données se sont adaptées non seulement à une distribution log-normale mais également d'autres types de distributions telles que la distribution gamma.

La distribution log-normale des concentrations des effluents était également en accord avec d'autres distributions (Cohen et al., 1975 ; Culp et al., 1980 ; Ossenbruggen et al., 1987 ; Kahn et Rubin, 1989 ; USEPA, 1991).

Sherwani et Moreau (1975) cité par Loftis et al. (1983) trouvent utiles de proposer les lois de distributions normal, log-normal, et gamma pour modéliser les variables aléatoires de la qualité de l'eau. Salle et al., (1981) ont suggéré que les distributions normales et log-normales s'appliquent le plus largement pour la qualité des eaux, mais aussi que la loi gamma, et la loi de Gumble sont parfois les plus appropriées.

On, constate d'après les études mentionnées dans cette recherche que la modélisation des variables aléatoires (paramètres de qualités des eaux) peuvent être modélisés par une distribution adéquate mais reste que l'effet de l'originalité des données joue un rôle important, éliminant de ce fait l'idée de généraliser le choix d'une meilleure distribution, les paramètres de qualités peuvent être modélisés par une loi de distribution donnée, mais peuvent dans d'autres cas d'étude pourrait être mieux modéliser par une autre distribution.

Les distributions normales et log-normales sont d'intérêt particulier pour deux raisons : D'abord, elles modélisent le comportement de plusieurs paramètres de qualités des eaux usées. En second lieu, les paramètres de ces distributions (la moyenne et l'écart-type sont facile à estimer), des tables de la fonction de distribution normale cumulative sont largement disponible (Loftis et al., 1983 ; USEPA, 1991). Ainsi, ces distributions sont faciles à utiliser.

La distribution de Weibull, qui est baptisée du nom du physicien suédois W. Weibull fournit une approximation étroite aux lois de probabilité de plusieurs phénomènes. Elle a été employée pour modéliser le temps de défaillance des éléments électriques, et systèmes mécaniques (Kottegoda et Rosso, 2009).

D'une part, il se peut qu'une partie des données des paramètres de qualités d'une eau usée ne suive aucune distribution. Par exemple, les données de la DBO₅ et les MES des effluents de 28 stations de traitement des eaux dans le Colorado (USA), ces données ont été analysées, mais aucune fonction de densité de probabilité n'a justement décrit les concentrations en DBO₅ et MES de ces effluents (Culp et al., 1980).

IV.5.1 Analyse des données à l'entrée de la station

Un premier aperçu se dégage de l'analyse des données à l'entrée de la station de Khenchela, on constate la variabilité de ces eaux.

Les données de qualité des eaux usées brute à l'entrée à fait l'objet de tentative de modélisation pour le choix d'une loi de distribution appropriée de probabilité on utilisant plusieurs tests statistiques adaptées avec différentes tests de distributions (par exemple, distribution normale, log-normale et Gamma).

Initialement, un test de Kolmogorov-Smirnov a été appliqué sur tous les données brutes d'auto-surveillance de la station incluant tous les paramètres de pollution (DCO, DBO₅ et MES).

Tableau 17: Résultats du test d'adéquation (Test de Kolmogorov-Smirnov) des différents paramètres de pollution des eaux usées à l'entrée de la station de Khenchela suivant différents lois de distribution.

	P-valeurs des lois de distribution à deux paramètres			Distribution acceptées		
	Normale	Log normale	Gamma	Normale	Log normale	Gamma
MES (mg/l)	0	0,0718185	0,103602		X	X
DCO (mg/l)	0,106485	0,00315011	0,144785	X		X
DBO ₅ (mg/l)	0	0,215521	0,0246568		X	

Remarque: Les p-valeurs obtenues sont issu de l'application du Test Kolmogorov-Smirnov (D test).

Si Les p-valeurs sont inférieures à 0,05 indiquerait que le paramètres de pollution ne vient pas de la distribution choisie avec 95% de confiance.

Le tableau 17 indique, que la DBO₅ à l'entrée de la station de Khenchela suit la distribution log-normale, alors que la DCO à l'entrée de la station suit une loi normale et une loi gamma. Pour illustrer ces résultats graphiquement, les courbes de probabilité, et la fonction de densité empirique (FDE) versus la fonction de répartition cumulative de la loi log- normale et gamma des données choisis aléatoirement sont présentés sur les figures 7, 8, 9 et 10, respectivement. La courbe de la fonction cumulative de densité de la DCO à l'entrée de la station de Khenchela suivant les lois de distribution normale et Gamma est représenté dans la figure 11.

Les MES à l'entrée de la station de Khenchela suivent une loi log-normal et une loi de gamma

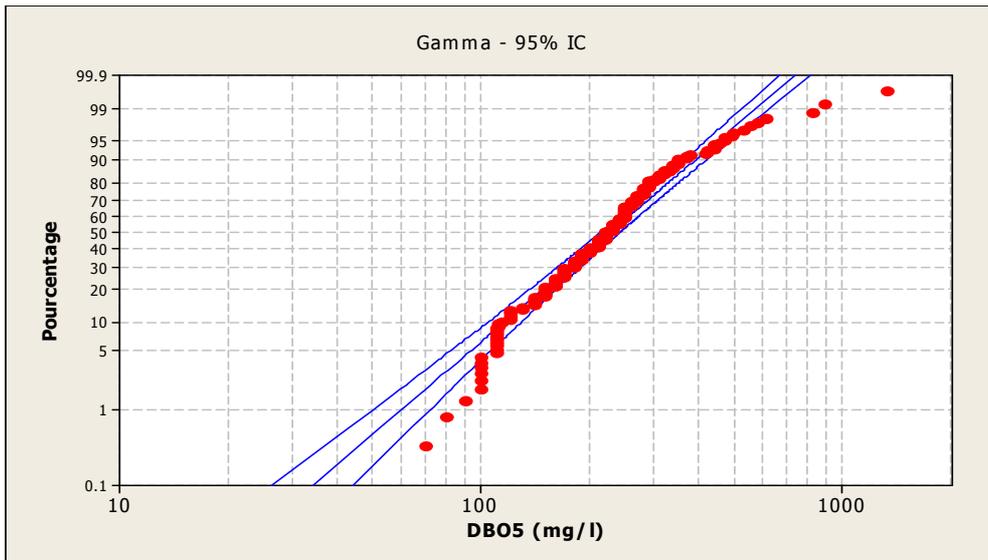


Figure 7 : Courbe de probabilité de la DBO₅ à l'entrée de la station de Khenchela suivant une loi de distribution Gamma.

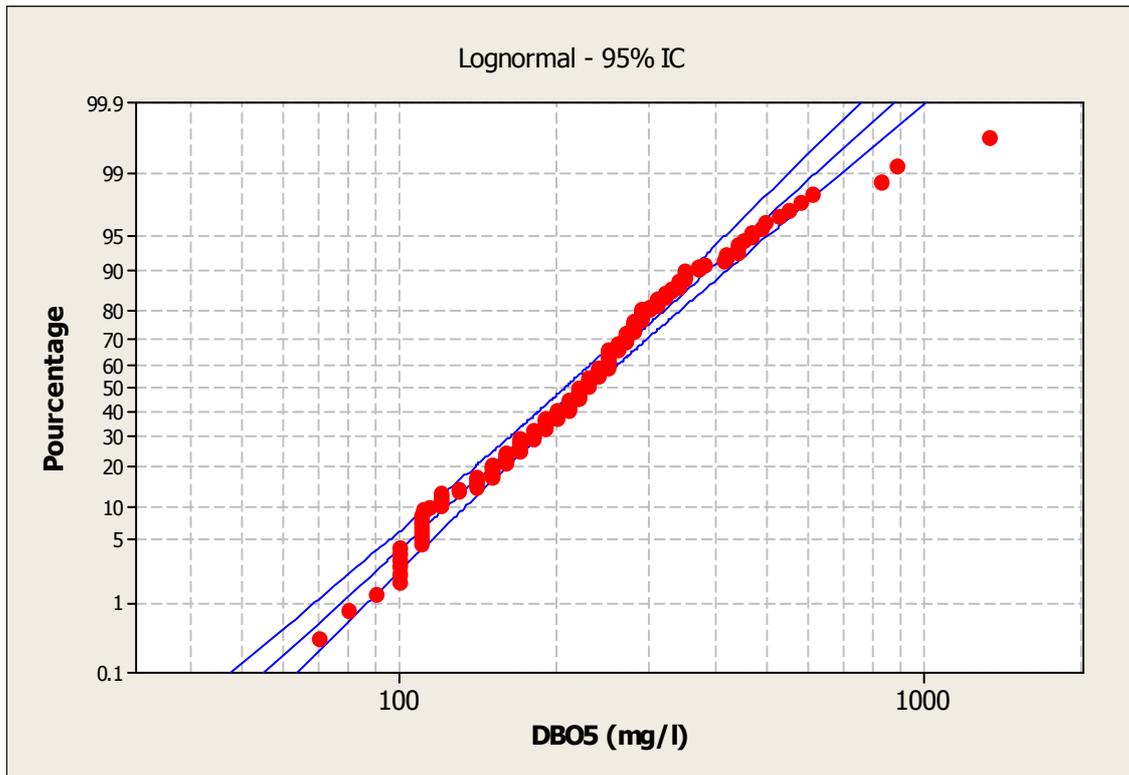


Figure 8 : Courbe de probabilité de la DBO₅ à l'entrée de la station de Khenchela suivant une loi de distribution Log-normale.

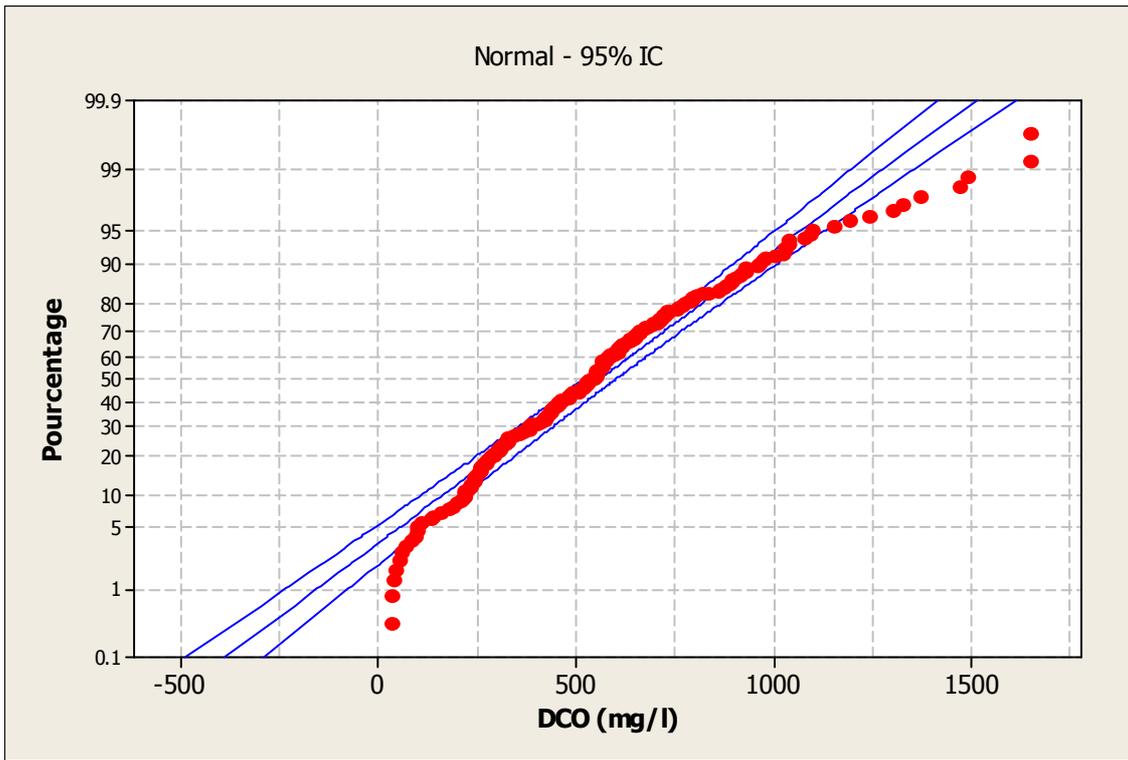


Figure 9: Courbe de probabilité de la DCO à l'entrée de la station de Khenchela suivant une loi de distribution normale.

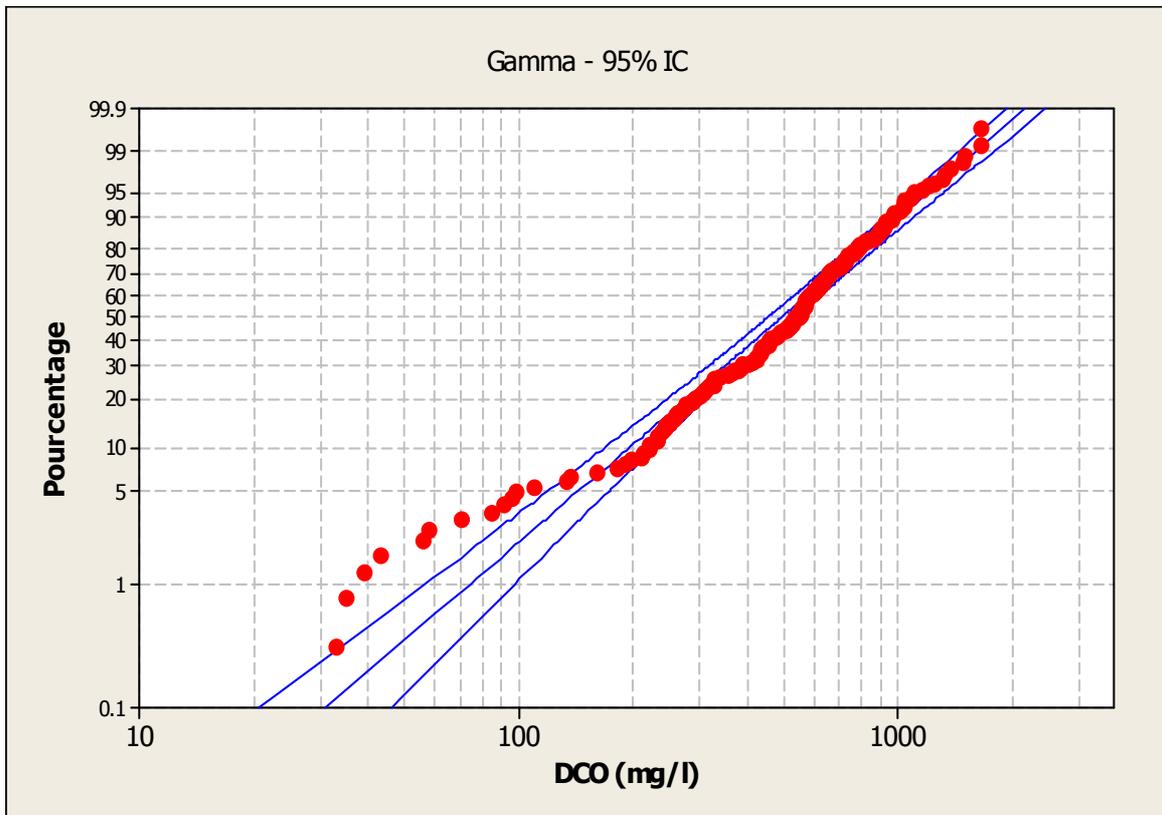


Figure 10: Courbe de probabilité de la DCO à l'entrée de la station de Khenchela suivant une loi de distribution Gamma.

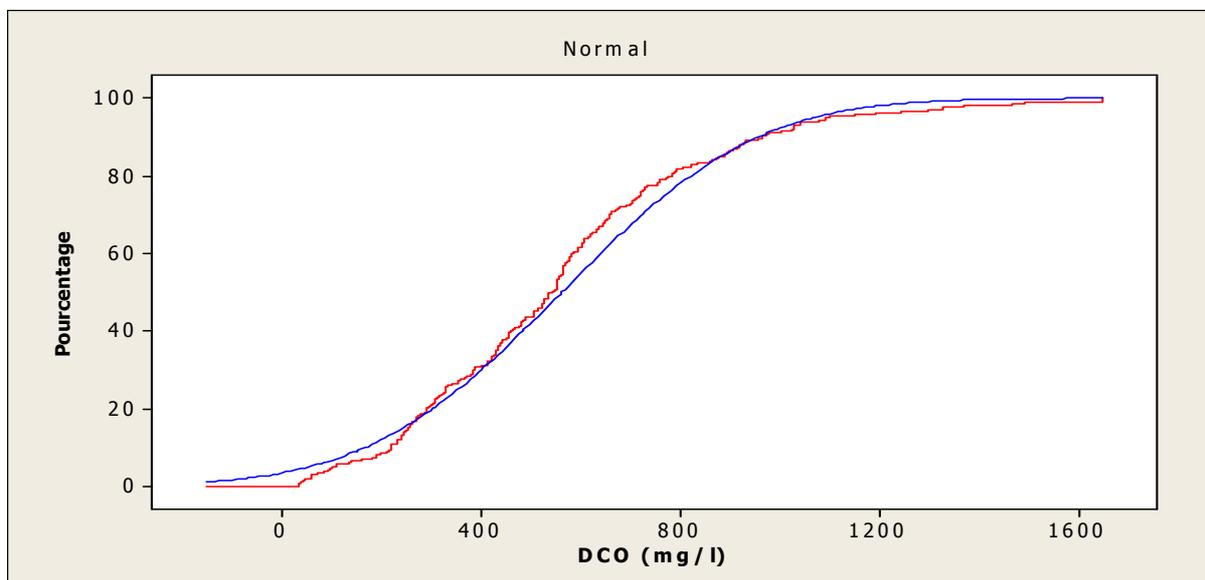
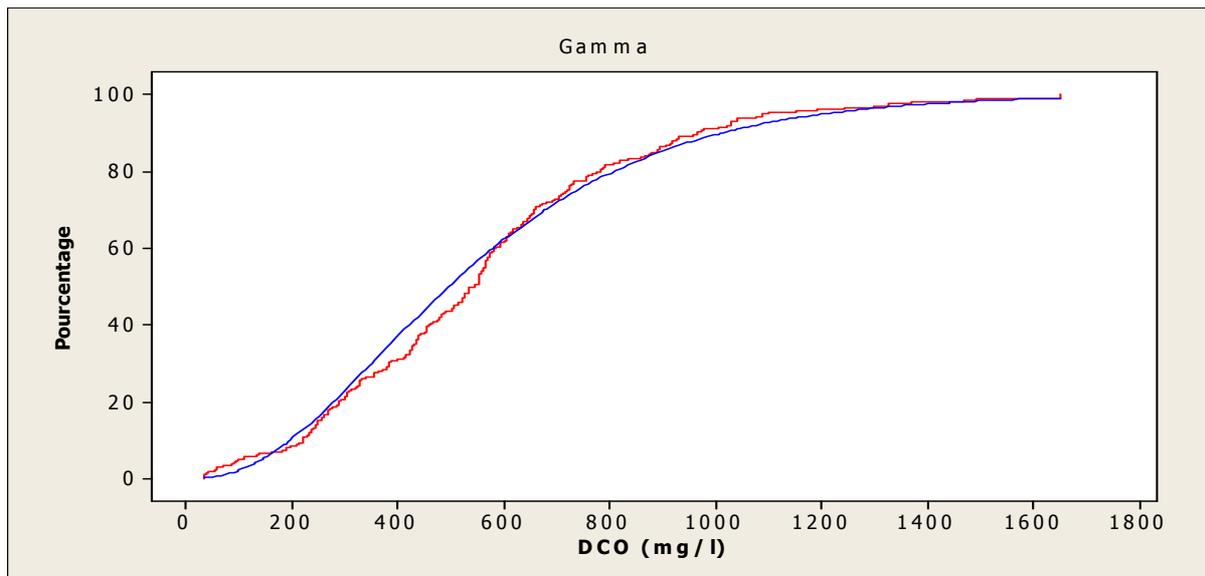


Figure 11: Courbe de la fonction cumulative de densité de la DCO à l'entrée de la station de Khenchela suivant les lois de distribution normale et Gamma.

La figure 12 représente l'histogramme de distribution versus la fonction de répartition de probabilité des MES suivant la loi log-normale et Gamma. La figure 13 représente la courbe des quantiles (concentration des MES versus la probabilité cumulative) suivant la loi log-normale et Gamma. Ces figures indiquent clairement que les données des MES peuvent être décrites par la loi de distribution gamma.

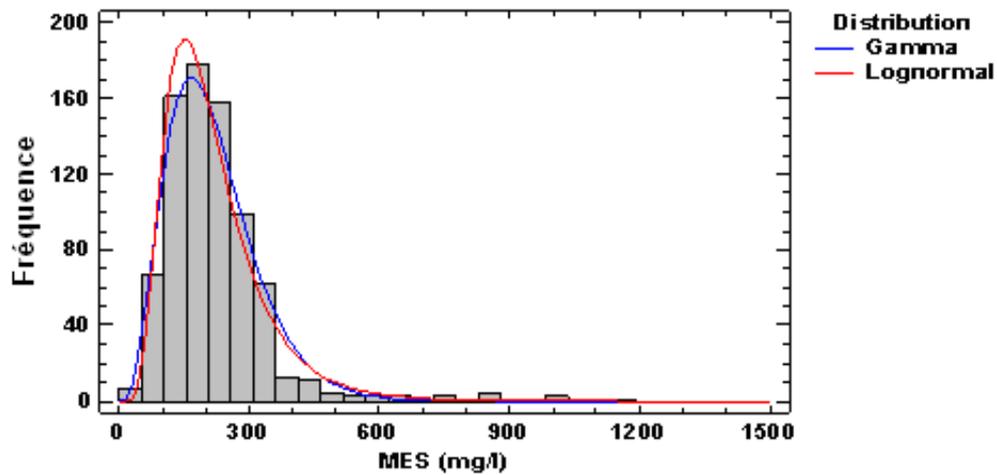


Figure 12: Histogramme de distribution des MES à l'entrée de la station de Khenchela suivant les lois de distribution log-normale et Gamma

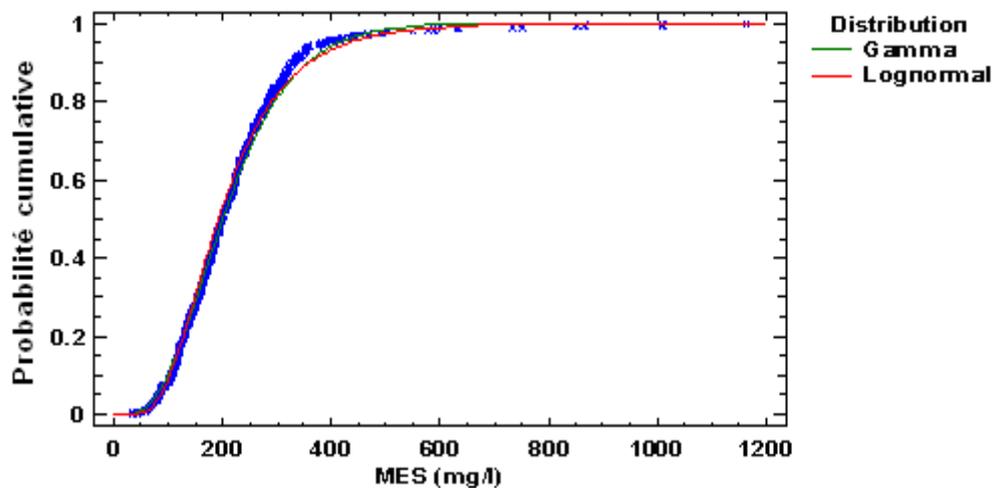


Figure 13: Courbe des quantiles des MES à l'entrée de la station de Khenchela suivant les lois de distributions log-normal et Gamma

IV.5.2 Analyse des données à la sortie de la station d'épuration

Le but de la deuxième partie de l'analyse statistique des paramètres de pollution des eaux usées traités par la station de Khenchela, et la détermination de la loi de distribution de chaque paramètre est indispensable pour la détermination des taux de fiabilité de la station.

Les données de qualité des eaux traitées à la sortie ont fait l'objet de tentative de modélisation adaptées sur différentes tests de distributions (par exemple, distribution normale, log-normale et gamma).

Comme mentionné précédemment, un test de Kolmogrov-Smirnov a été appliqué sur tous les données d'auto-surveillance à la sortie de la station incluant tous les paramètres de pollution (DCO, DBO₅ et MES).

Tableau 18: Résultats du test d'adéquation (Test de Kolmogrov-Smirnov) des différents paramètres de pollution des eaux traités à la sortie de la station de Khenchela suivant différents lois de distribution.

	P-valeurs de lois de distribution à deux paramètres			Distribution acceptées		
	Normale	Log normale	Gamma	Normale	Log normale	Gamma
MES (mg/l)	0	0,0053075	0		X*	
DCO (mg/l)	0,0000122273	0,0953478	0,00622771		X	
DBO ₅ (mg/l)	0,00000803763	0,0469248	0,00411114		X	

Remarque: Les p-valeurs obtenues sont issus de l'application du test Kolmogorov-Smirnov (D). Si Les p-valeurs <0,05 alors la loi de distribution choisi est rejetée avec 95% de confiance.

*: On combine d'autres méthodes statistiques pour le choix de la loi de distribution.

Les résultats affichés dans le tableau 18 indique, que la DCO et DBO₅ à la sortie de la station de Khenchela suit une distribution log-normale. Pour illustrer ces résultats graphiquement, les courbes de probabilité, et la courbe de la fonction de densité empirique (FDE) suivant la loi log-normale est présenté sur les figures 14, 15, 16 et 17, respectivement.

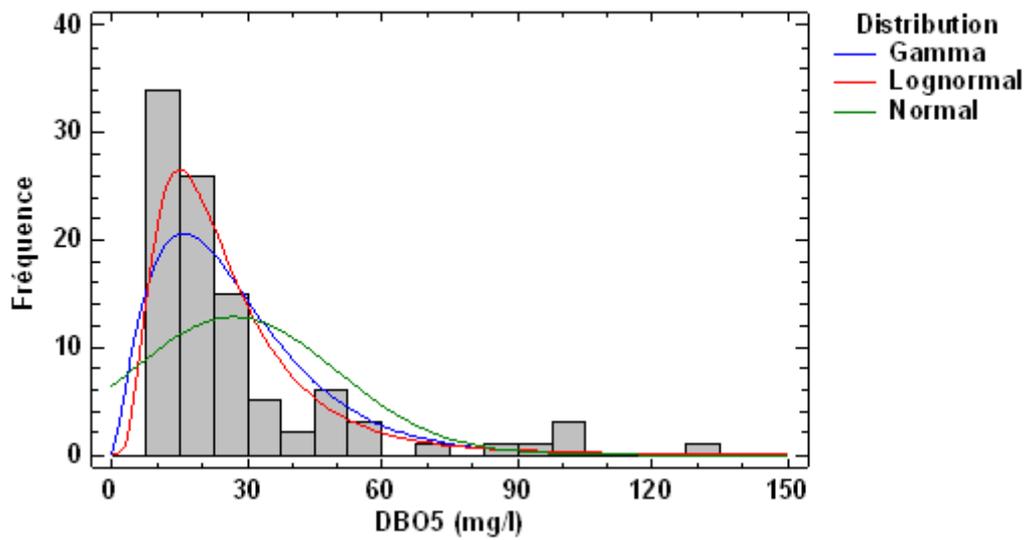


Figure 14 : Histogramme de la DBO₅ à la sortie de la station de Khenchela suivant différents lois de distribution.

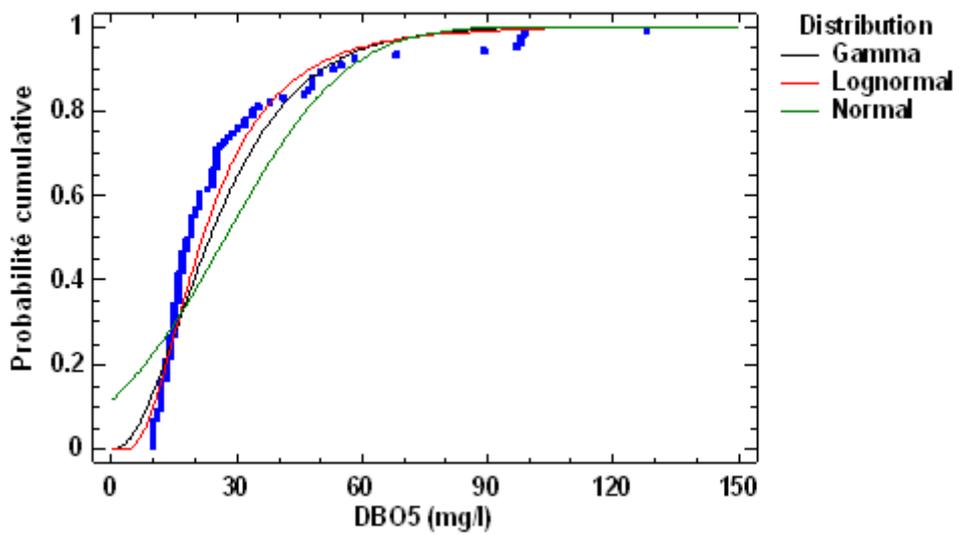


Figure 15: Courbe des quantiles de la DBO₅ à la sortie de la station de Khenchela suivant différents lois de distribution.

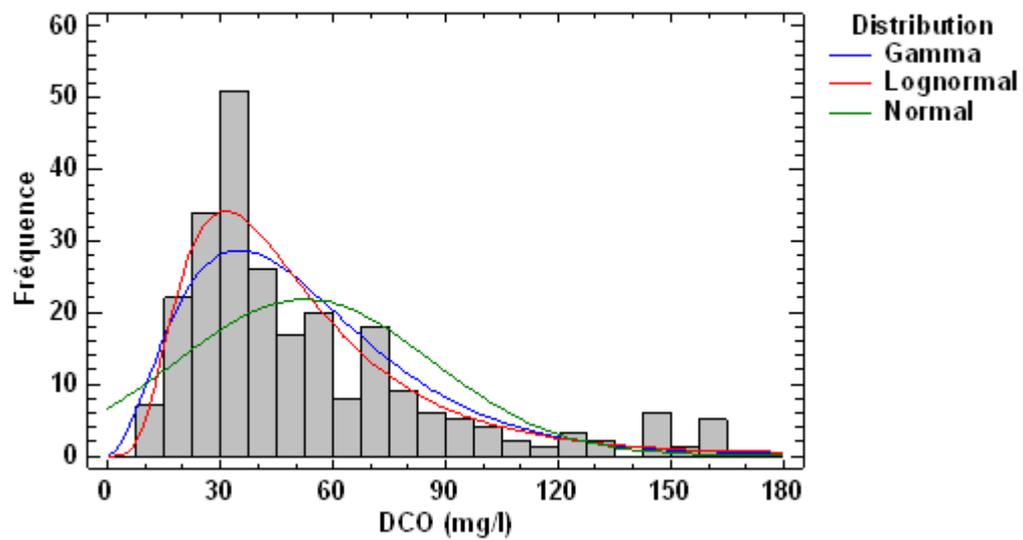


Figure 16: Histogramme de la DCO à la sortie la station de Khenchela suivant différents lois de distribution.

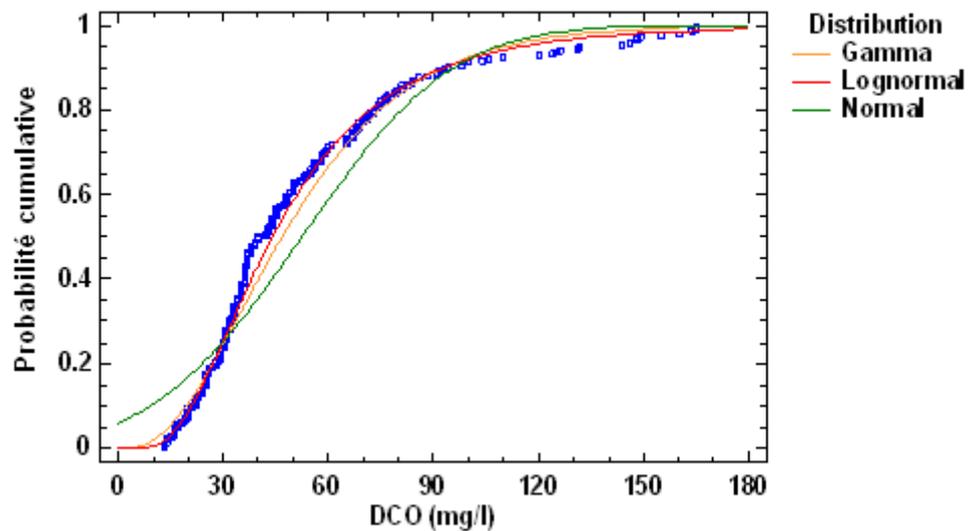


Figure 17: Courbe des quantiles de la DCO à la sortie de la station de Khenchela suivant différents lois de distribution.

La figure suivante représente une capture d'écran du logiciel de traitement statistique STAGRAPHICS utilisé pour le traitement des données.

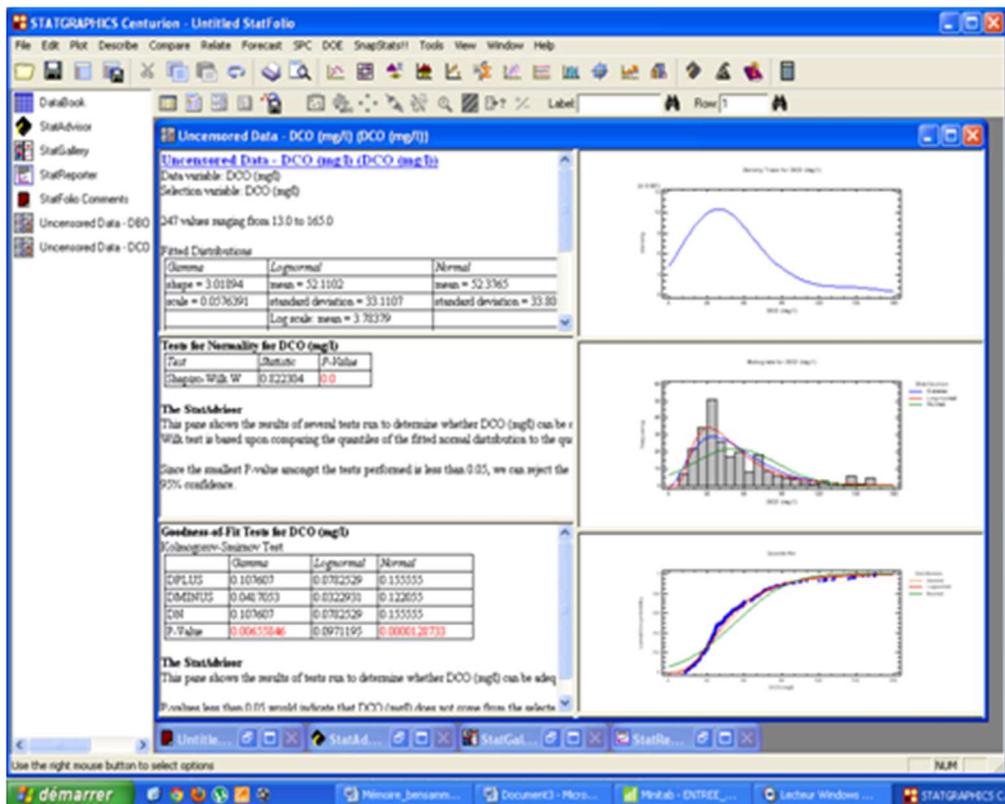


Figure 18: Capture écran lors du traitement statistique des données de la station de Khenchela

L'ajustement faible des MES à la sortie de la station de Khenchela, le choix de la loi de distribution n'est possible qu'avec une visualisation des courbes quantiles et de l'histogramme.

La figure 19 représente l'histogramme des MES avec un essai de modélisation avec différents lois de distribution. Une remarque se dégage de cette courbe, les MES suivent une loi de distribution log-normale.

La courbe des quantiles confirme notre remarque de distribution versus la fonction de répartition de probabilité des MES suivant la loi log-normale et Gamma. La figure 20 représente la courbe des quantiles (concentration des MES versus la probabilité cumulative) suivant la loi log-normale et Gamma. Ces figures indiquent clairement que les données des MES peuvent être décrites par la loi de distribution log-normale.

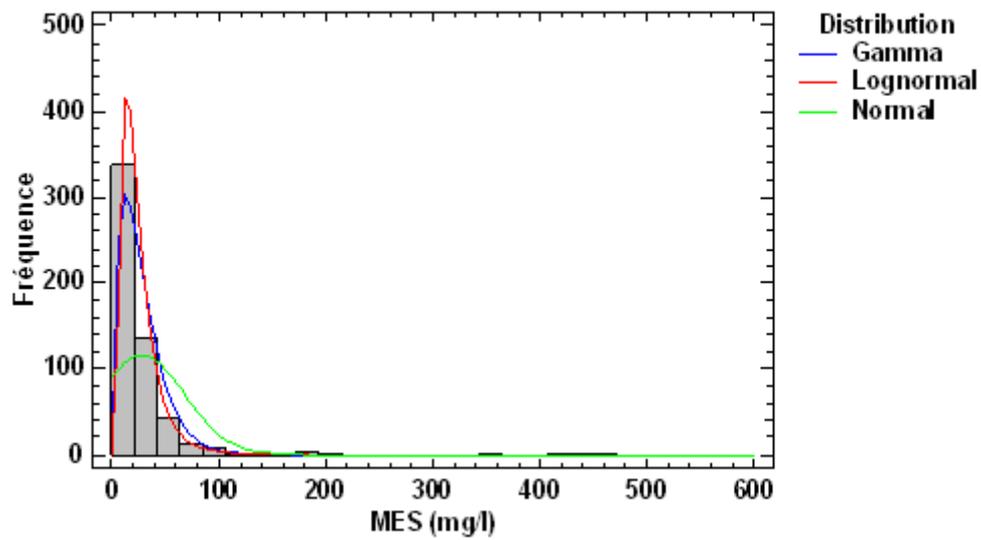


Figure 19: Histogramme de distribution des MES à la sortie de la station de Khenchela suivant différents lois de distribution.

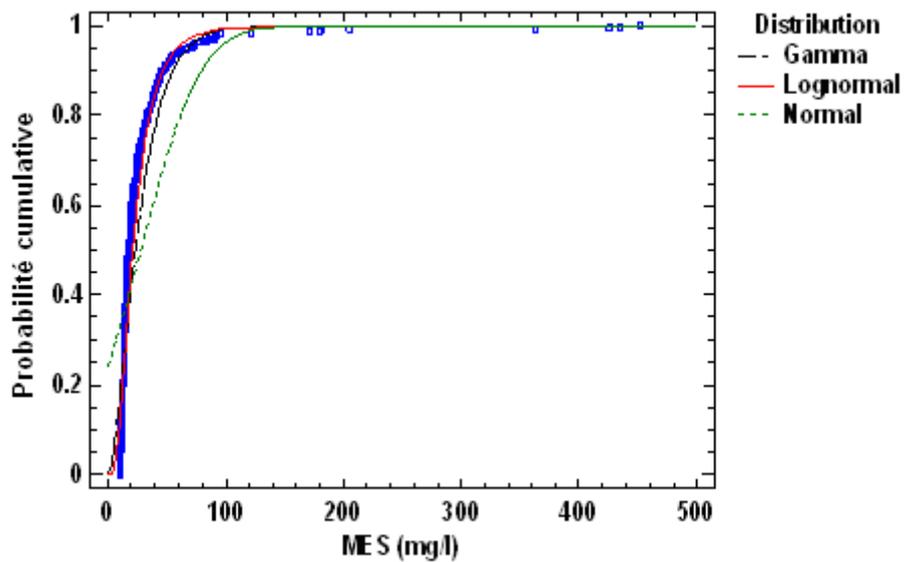


Figure 20: Courbe des quartiles versus la probabilité cumulative des MES suivant différents lois de distributions.

Pour justifier notre choix de la distribution log-normale, les courbes de probabilité des MES à la sortie de la station suivant différentes lois de distribution sont présentées dans la figure 21.

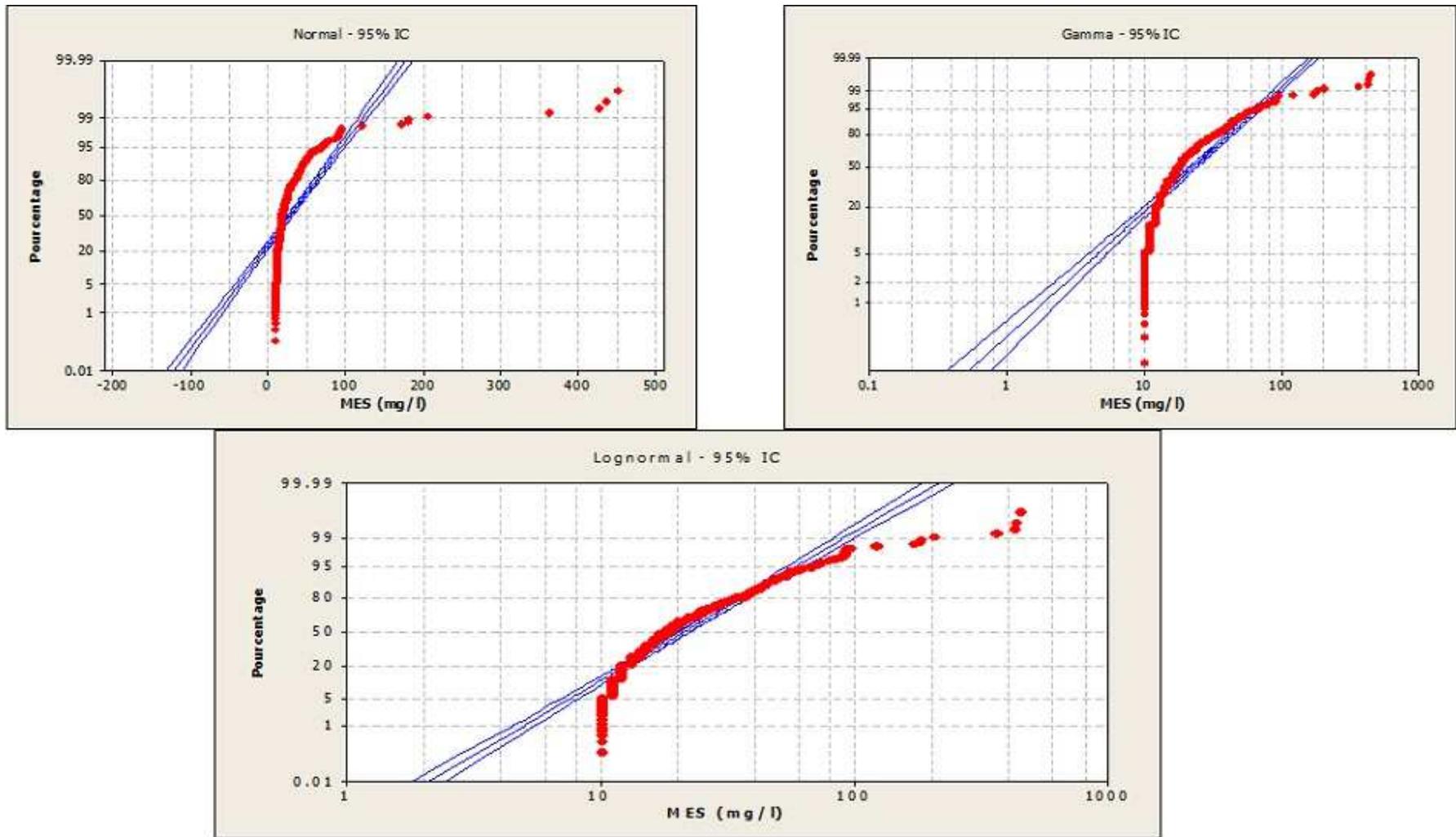


Figure 21: Courbe de probabilité des MES à la sortie de la station de Khenchela suivant différents lois de distribution.

IV.6 EVALUATION DE LA FIABILITE GLOBALE DE LA STATION DE KHENCHELA

L'évaluation globale de la fiabilité du procédé de traitement de la station de Khenchela est une approche simplifiée se basant sur la comparaison des performances d'exécution du procédé de traitement et le respect des normes de rejet.

La station est considéré comme fiable si la concentration des paramètres de pollution (MES, DCO, et la DBO_5) des eaux traités rejetés respectent les normes de rejet (concentration inférieur à la limite décrite par la législation).

Dans cette section, nous évaluerons la fiabilité de différents paramètres basée sur les données rassemblées et pertinentes durant la période 2009-2012.

IV.6. 1 Evaluation de la fiabilité globale de l'élimination des MES

Un premier approche de mesure le la fiabilité globale de la concentration des MES à la sortie de la station de Khenchela consiste a réaliser une comparaison simple de la conformité des eaux traitées rejetés avec les normes de rejet algériennes, et estimer le temps de respect de la législation.

La figure 22 est une représentation graphique de la variabilité de la concentration des MES à la sortie de la station de Khenchela et leur conformité avec la législation.

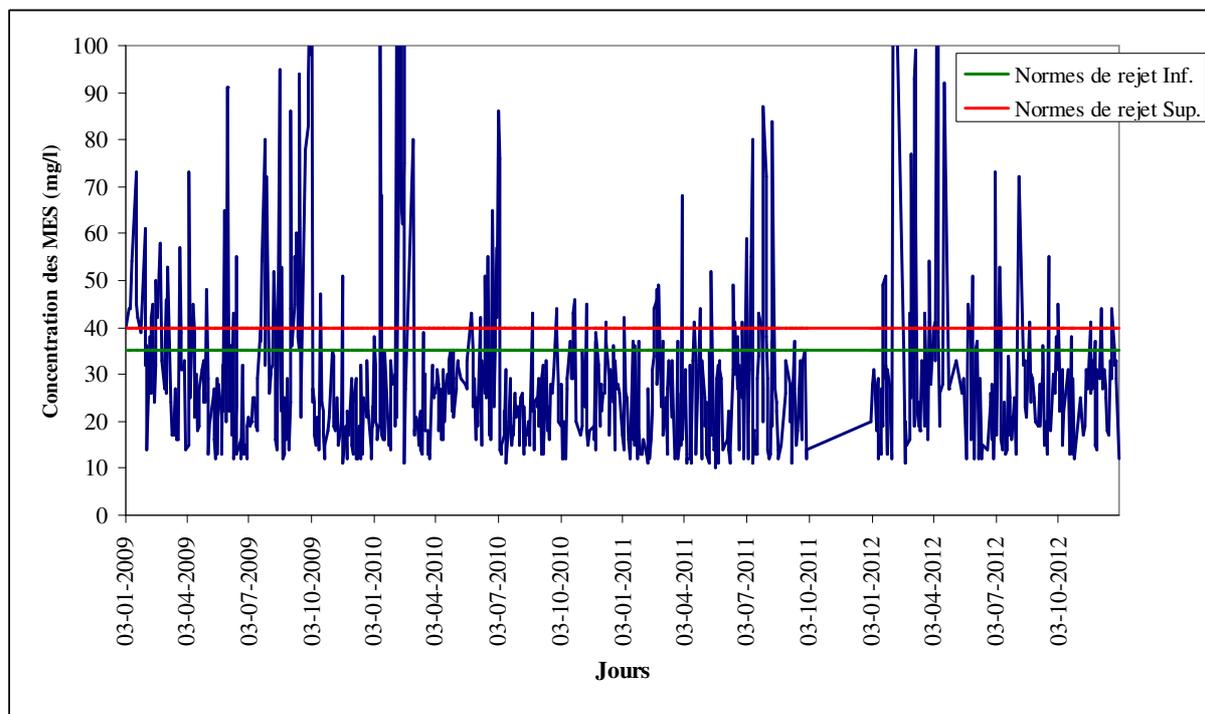


Figure 22: Courbe de concentration des MES à la sortie de la station de Khenchela (2009-2012).

Sur la base des données pris en considérations soit 784 jours de données disponibles, la station a respecté la réglementation en 662 jours. Pour un total de 122 jours sur la période d'observation, la station n'a pas respecté la réglementation.

On constate que pour ce paramètre de qualité des eaux usées traitées rejetés, la station affiche un taux de fiabilité relatif estimé à 84,43 %.

Sur la base du même diagnostic et durant les quatre dernières années d'exploitations (4 ans = 1460 jours), la concentration des MES des eaux traitées rejetées ont respecté la législation durant 662 jours, soit un taux global de fiabilité de 45,34%. Ce taux de fiabilité globale est médiocre du principalement à l'absence des données de plusieurs mois d'exploitations.

IV.6.2 Evaluation de la fiabilité globale de l'élimination de la DCO

La figure 23 illustre la variabilité des concentrations de la DCO à la sortie de la station de Khenchela et leur conformité avec la législation.

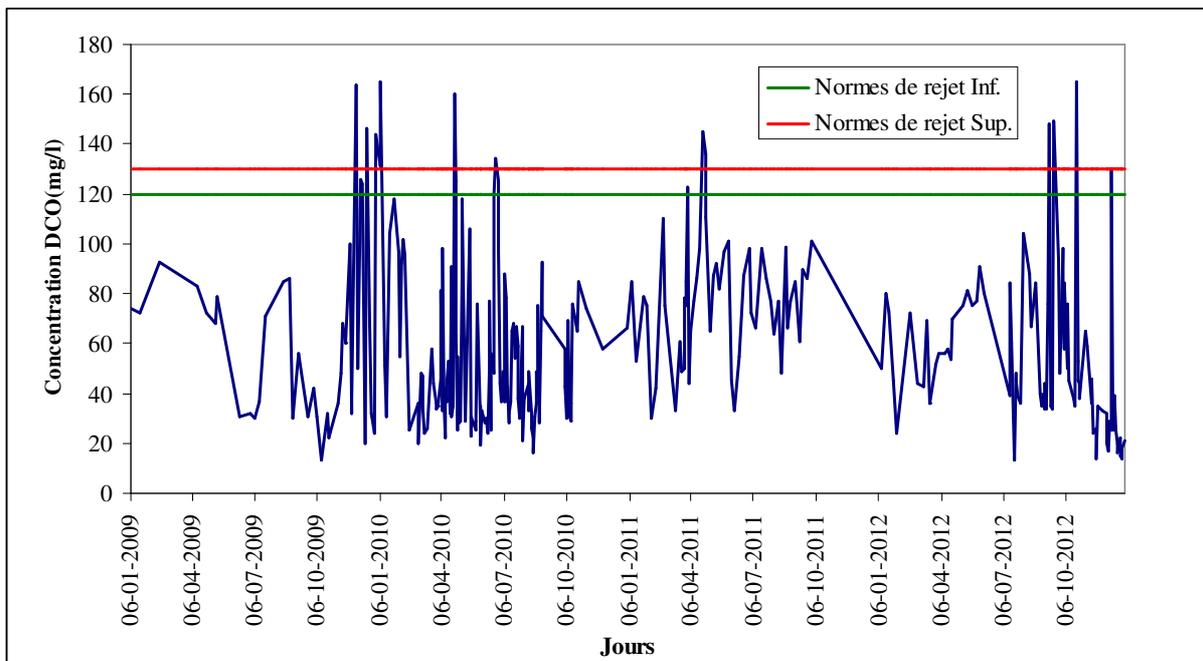


Figure 23: Courbe de concentration de la DCO à la sortie de la station de Khenchela (2009-2012).

On prenant en considération les données disponible sur une période de 300 jours d'observation, l'analyse statistique montre que la station a respecté la réglementation durant 288 jours, la station a enregistré un dépassement des normes de rejet seulement durant 14 jours sur la période d'observation, ainsi la station affiche un taux de fiabilité relatif estimé à 96 %.

Ce constat n'est que relatif, puisque sur la durée des quatre dernières années d'exploitations (4 ans = 1460 jours), la concentration de la DCO des eaux traitées rejetées ont respecté la législation que durant 288 jours, soit un taux de fiabilité globale estimé à 19,72 %. Un taux insuffisant due principalement à l'absence des données puisque la station fonctionnée durant des mois sans enregistré les données de la concentration de la DCO des eaux usées traitées.

V.6.3 Evaluation de la fiabilité globale de la dégradation de la DBO₅

La variabilité des concentrations de la DBO₅ à la sortie de la station de Khenchela et leur conformité avec la législation est illustrée dans la figure 24.

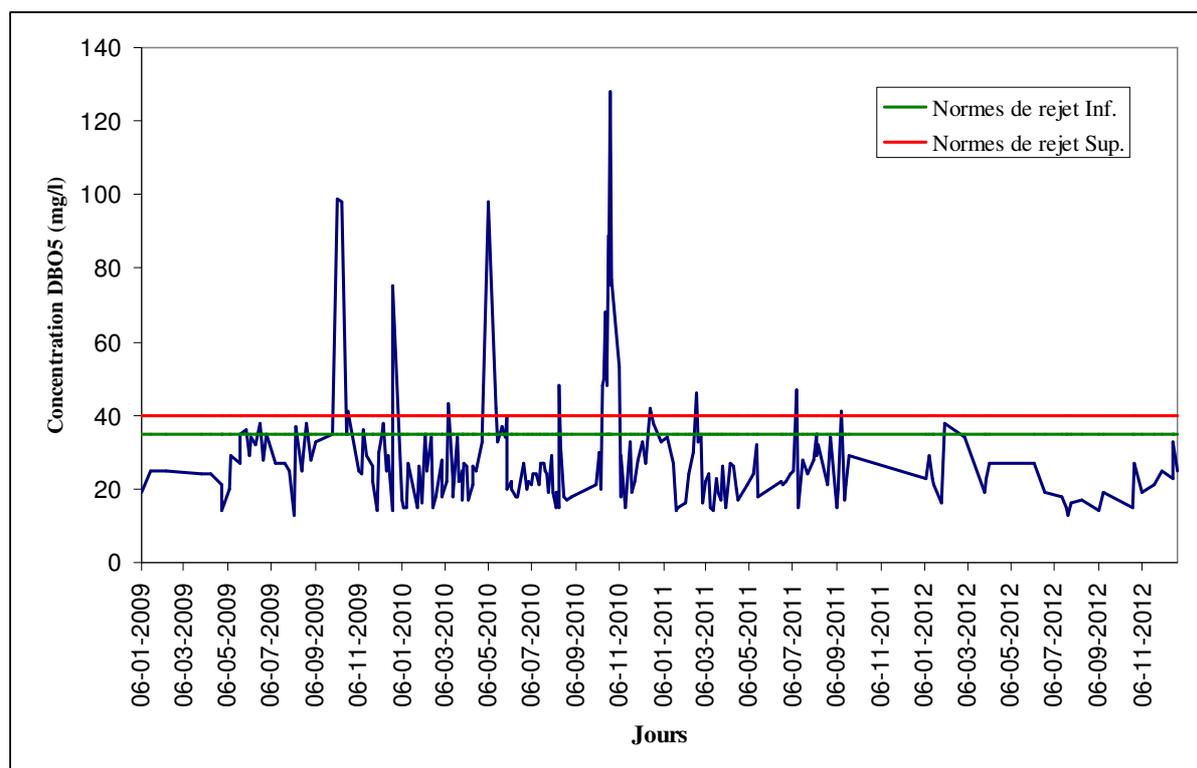


Figure 24: Courbe de concentration de la DBO₅ à la sortie de la station de Khenchela (2009-2012).

Les données pris en considération ne sont disponible que sur 214 jours d'observation, l'analyse statistique montre que la station a respecté la réglementation durant 191 jours, mais sur la même période d'observation, la station a enregistré un dépassement de la norme de rejet de la DBO₅ durant 14 jours. La station affiche un taux relatif de fiabilité estimé à 89,25 %.

Le taux globale de fiabilité sur la durée des quatre années d'exploitations (4 ans = 1460 jours) est estimé à 13,08 %. L'absence des données à la sortie de la station est la principale cause de ce taux médiocre.

IV.7 DETERMINATION DU TAUX DE FIABILITE PAR MODELE PROBABILISTE

Les distributions de probabilité caractérisent explicitement la variabilité des performances du procédé de traitement, et permettent ainsi d'estimer la probabilité que les buts du traitement (concentration des paramètres de pollution) seraient dépassés. Dans cette partie de notre étude, nous présentant un essai d'évaluation de la variabilité du taux de fiabilité du procédé de traitement de la station de Khenchela basée sur les travaux de recherche de Niku et al. (1979,1981,1982) mentionnée dans le chapitre précédent.

La loi de distribution des données à la sortie de la station de traitement des eaux usées de la ville de Khenchela suivent une loi de distribution type log-normale.

Le tableau 19 présente un exemple de calcul du taux de fiabilité de la DBO_5 par la méthode du coefficient de fiabilité (CdF) pour l'année 2012.

Tableau 19: Exemple de détermination du taux de fiabilité de la dégradation de la DBO₅ par modèle probabiliste, station d'épuration de Khenchela (2012).

Date	DOB5 (mg/L)	Taux de Fiabilité (%) Prob (Cs < Cn)	Centiles Z1-α	Coefficient de Fiabilité	Date	Taux de Fiabilité
09-01-2012	23	5	1.645	1.655	09-01-2012	99,00
12-01-2012	29	10	1.282	1.494	12-01-2012	90,00
16-01-2012	23	15	1.036	1.394	16-01-2012	99,00
19-01-2012	21	20	0.842	1.319	19-01-2012	99,00
30-01-2012	16	25	0.674	1.259	30-01-2012	99,00
02-02-2012	38	30	0.524	1.207	02-02-2012	65,00
01-03-2012	34	35	0.385	1.160	01-03-2012	80,00
29-03-2012	19	40	0.253	1.118	29-03-2012	99,00
02-04-2012	23	45	0.126	1.078	02-04-2012	99,00
05-04-2012	27	50	0.000	1.041	05-04-2012	95,00
07-06-2012	27	55	-0.126	1.004	07-06-2012	95,00
21-06-2012	19	60	-0.253	0.969	21-06-2012	99,00
15-07-2012	18	65	-0.385	0.933	15-07-2012	99,00
22-07-2012	15	70	-0.524	0.897	22-07-2012	99,00
24-07-2012	13	75	-0.674	0.860	24-07-2012	99,00
29-07-2012	16	80	-0.842	0.821	29-07-2012	99,00
13-08-2012	17	85	-1.036	0.777	13-08-2012	99,00
05-09-2012	14	90	-1.282	0.725	05-09-2012	99,00
12-09-2012	19	95	-1.645	0.654	12-09-2012	99,00
23-10-2012	15	96	-1.751	0.635	23-10-2012	99,00
25-10-2012	27	97	-1.881	0.612	25-10-2012	95,00
04-11-2012	19	98	-2.054	0.583	04-11-2012	99,00
22-11-2012	21	99	-2.326	0.540	22-11-2012	99,00
03-12-2012	25	Norme de rejet de la DBO5 (mg/L)		40	03-12-2012	97,00
18-12-2012	23				18-12-2012	99,00
19-12-2012	33	Taux de Fiabilité (%)			19-12-2012	80,00
24-12-2012	25	Prob Cs < Cn			24-12-2012	97,00
Min.	13,00	40				
Max.	38,00	45				
Moy.	22,19	50				
Ecart-type	6,18	55				
Cv	0,29	60				
		65				
		70				
		75				
		80				
		85				
		90				
		95				
		96				
		97				
		98				
		99				
			Concentration requise (mg/L)			
			44,71			
			43,13			
			41,62			
			40,17			
			38,75			
			37,34			
			35,90			
			34,41			
			32,83			
			31,07			
			29,00			
			26,17			
			25,40			
			24,49			
			23,32			
			21,59			

IV.7.1 Evaluation du taux de fiabilité de la dégradation de la DBO₅

Les calculs des taux de fiabilité du procédé de traitement de la station d'épuration de Khenchela pour la dégradation de la DBO₅ sont schématisés sur la figure 25.

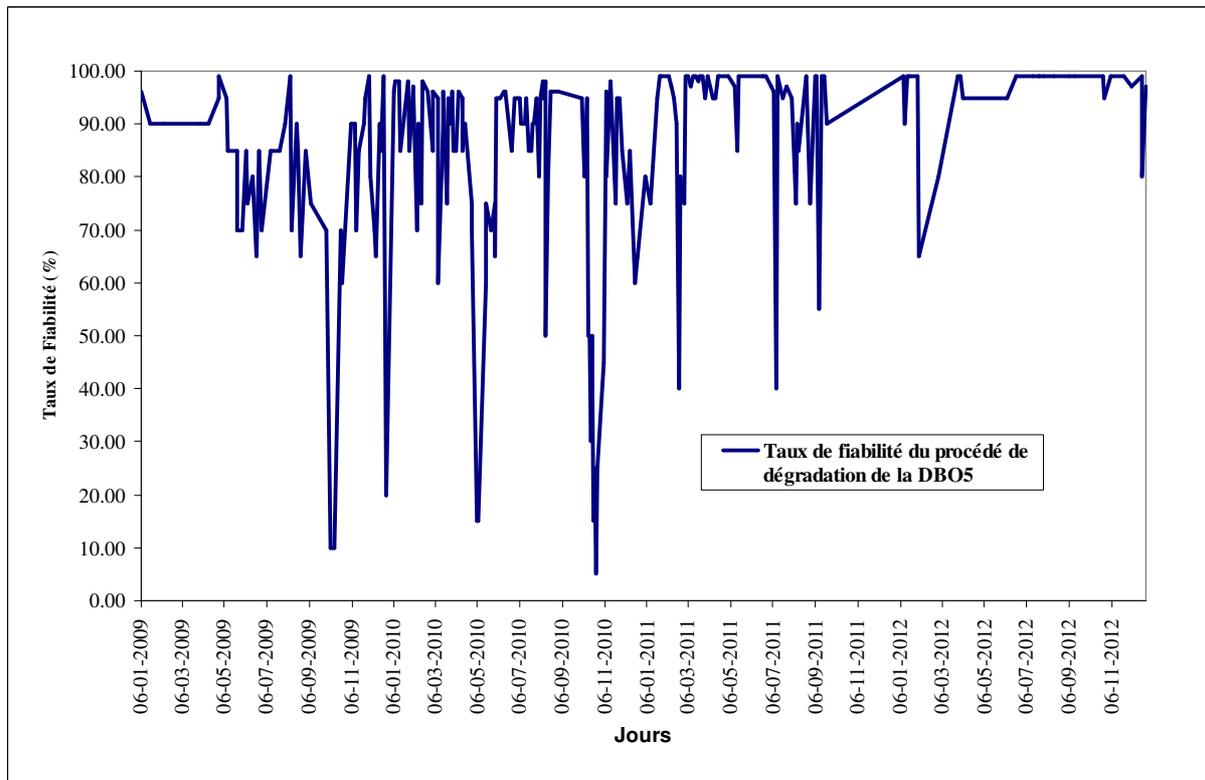


Figure 25: Courbe de la variation du taux de fiabilité du procédé de dégradation de la DBO₅ dans la station de Khenchela (2009-2012).

La variabilité du taux de fiabilité est due principalement à la variation qualitative et quantitative des eaux usées brutes à l'entrée de la station, de la présence de forte pollution organique, et des différents paramètres qui régissent le processus de dégradation la pollution par les bactéries.

Les facteurs suivant peuvent être la cause de cette variation:

- Le manque d'oxygénation (vitesse de rotation des agitateurs inadéquats, mauvaise répartition de l'oxygène dans le bassin,...);
- Changement climatiques (la variation saisonnière de la température influence sur le rendement de la biomasse);
- Non maîtrise du taux de recirculation;
- Défaut de conception.

IV.7.2 Evaluation du taux de fiabilité de l'élimination de la DCO

Les calculs des taux de fiabilité du procédé de traitement de la station d'épuration de Khenchela pour l'élimination de la DCO sont schématisés sur la figure 26.

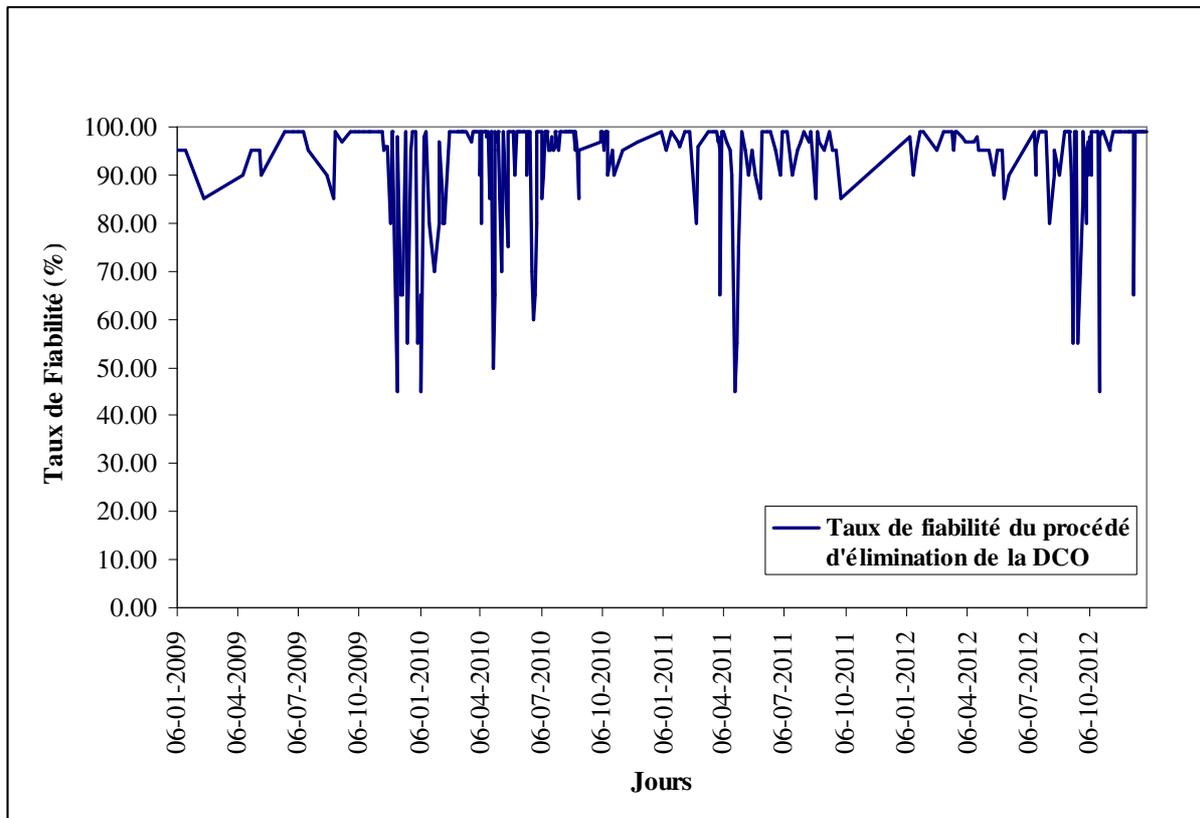


Figure 26: Courbe de la variation du taux de fiabilité de l'élimination de la DCO dans la station d'épuration de Khenchela (2009-2012).

La variabilité du taux de fiabilité est influencé par la variation qualitative et quantitative des eaux usées brutes à l'entrée de la station, la présence d'une pollution chimique (réseau d'assainissement unitaire), lors et après des phénomènes pluvieux (lessivage des sols) rendant ainsi la tâche très difficile aux exploitants combinés a cela une conception inadéquate de la station.

IV.7.3 Evaluation du taux de fiabilité de l'élimination des MES

Les calculs des taux de fiabilité du procédé d'élimination des MES dans la station d'épuration de Khenchela sont schématisés sur la figure 27.

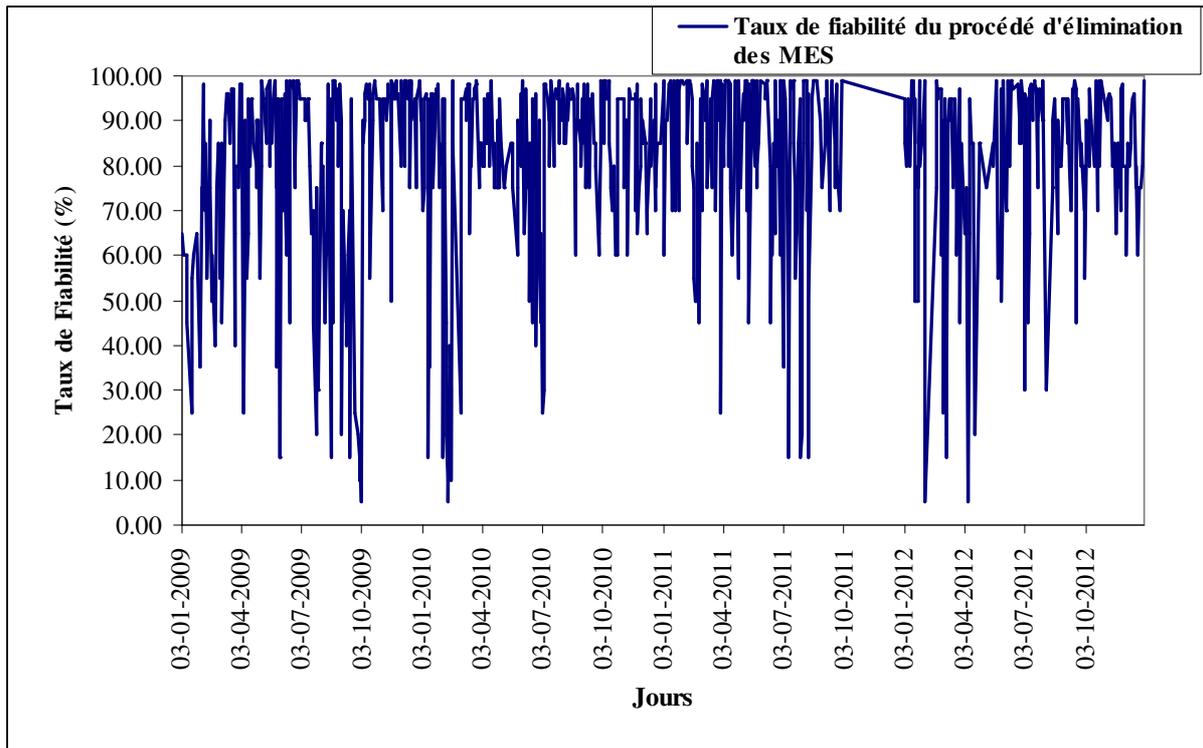


Figure 27: Courbe de la variation du taux de fiabilité du procédé d'élimination des MES dans la station d'épuration de Khenchela (2009-2012).

On constate la variabilité des taux de fiabilité, cette variabilité est due principalement à la variabilité qualitative et quantitative des eaux usées brutes à l'entrée de la station.

Les paramètres suivants peuvent influencer la fiabilité du procédé d'élimination des MES:

- Temps de séjours dans le bassin de dessablage insuffisant;
- Injection de l'air dans le bassin de dessablage inadéquate;
- Performances médiocre du dégrilleur fin;
- Conception inadéquate.

IV.8 DETERMINATION DES TAUX DE DEFAILLANCES

Dans cette section, nous évaluerons la variabilité du taux de défaillance du procédé de traitement de la station basée sur les travaux de recherche de Niku et al., (1979,1981,1982) mentionnée dans le chapitre IV.

Une station est dite défaillante si : $C_{\text{sortie}} > C_{\text{Normes rejet}}$

$$\text{Taux de défaillance} = 1 - \text{Taux de fiabilité}$$

IV.8.1 Evaluation du taux de défaillance de la dégradation de la DBO₅

Les calculs du taux de défaillance du procédé de dégradation de la DBO₅ sont schématisés sur la figure 28.

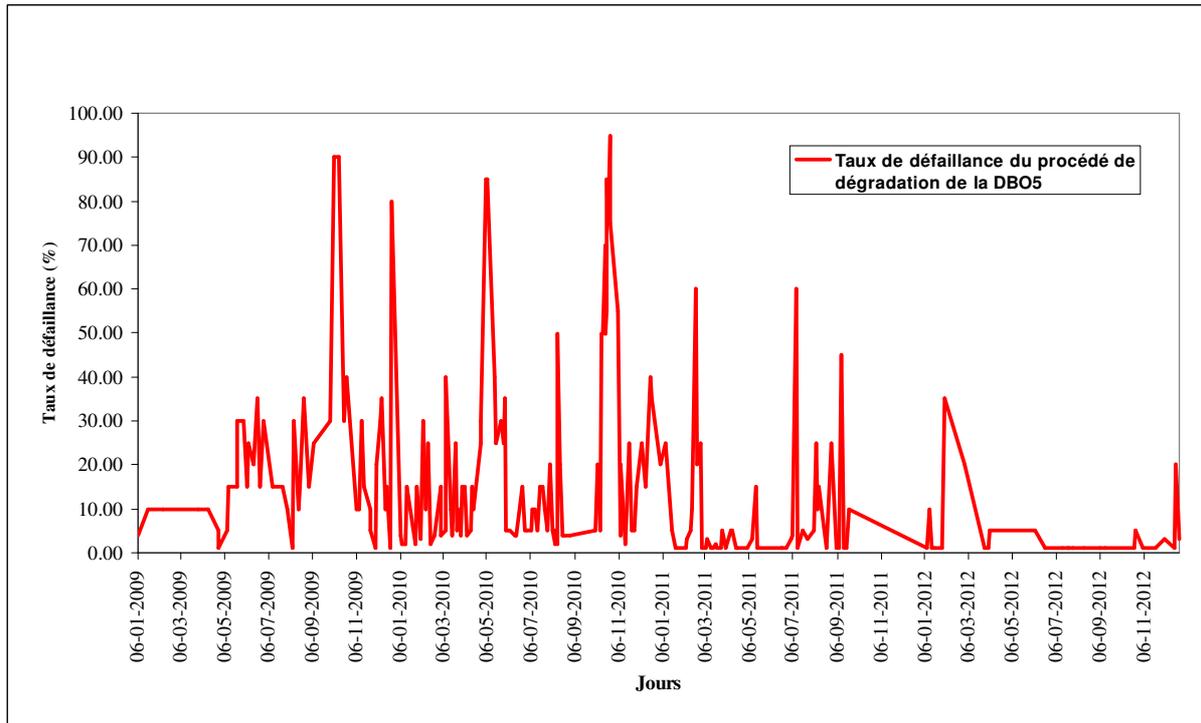


Figure 28: Courbe de la variation du taux de défaillance du procédé de traitement de la DBO₅ dans la station d'épuration de Khenchela (2009-2012).

L'absence d'historique de défaillance mécanique ou de notes de phénomènes exceptionnelles (crues, variation excessives des paramètres de qualités des eaux usées,...) dans le registre d'exploitation de la station rend l'explication des pics de taux de défaillance une tâche difficile.

IV.8.2 Détermination du taux de défaillance d'élimination de la DCO

Les calculs du taux de défaillance du procédé d'élimination de la DCO dans la station de Khenchela sont schématisés sur la figure 29.

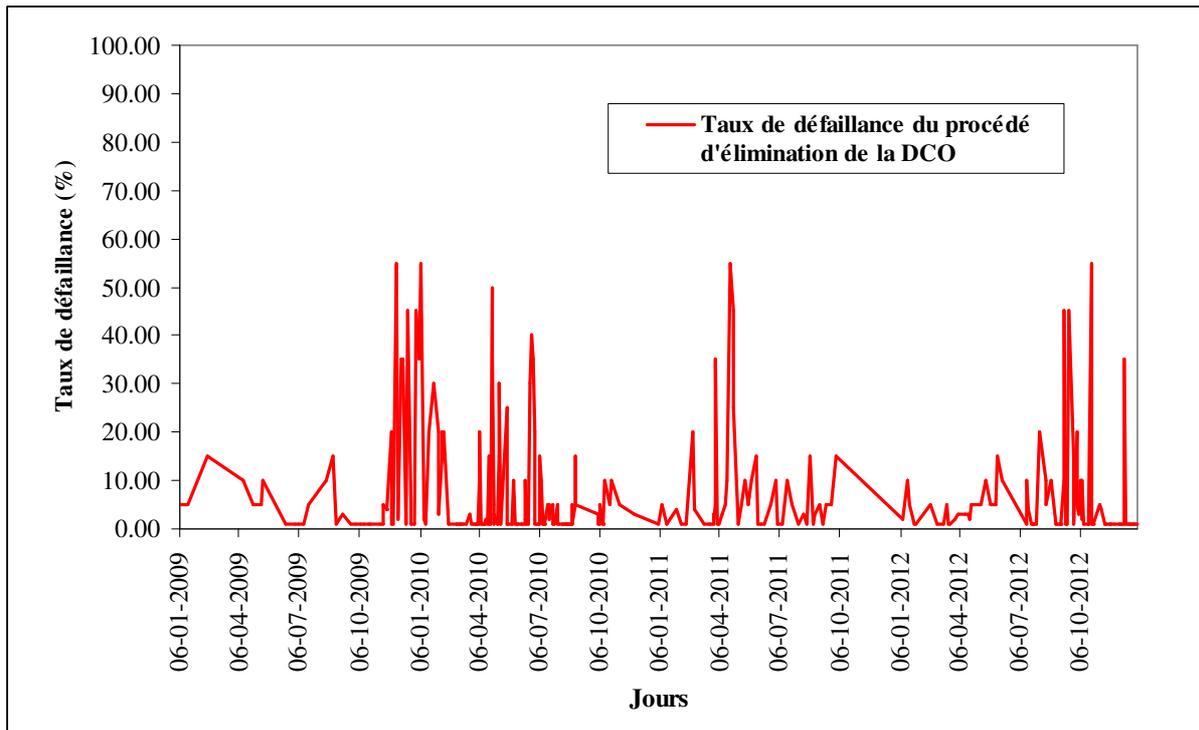


Figure 29: Courbe de la variation du taux de défaillance du procédé de traitement de la DCO dans la station d'épuration de Khenchela (2009-2012).

La variation du taux de défaillance du procédé d'élimination de la DCO est influencée par la variation des paramètres de la pollution des eaux usées brutes, et une aération insuffisante pour l'oxydation des composants organiques.

IV.8.3 Détermination du taux de défaillance d'élimination des MES

Les calculs du taux de défaillance du procédé d'élimination des MES dans la station de Khenchela sont schématisés sur la figure 30.

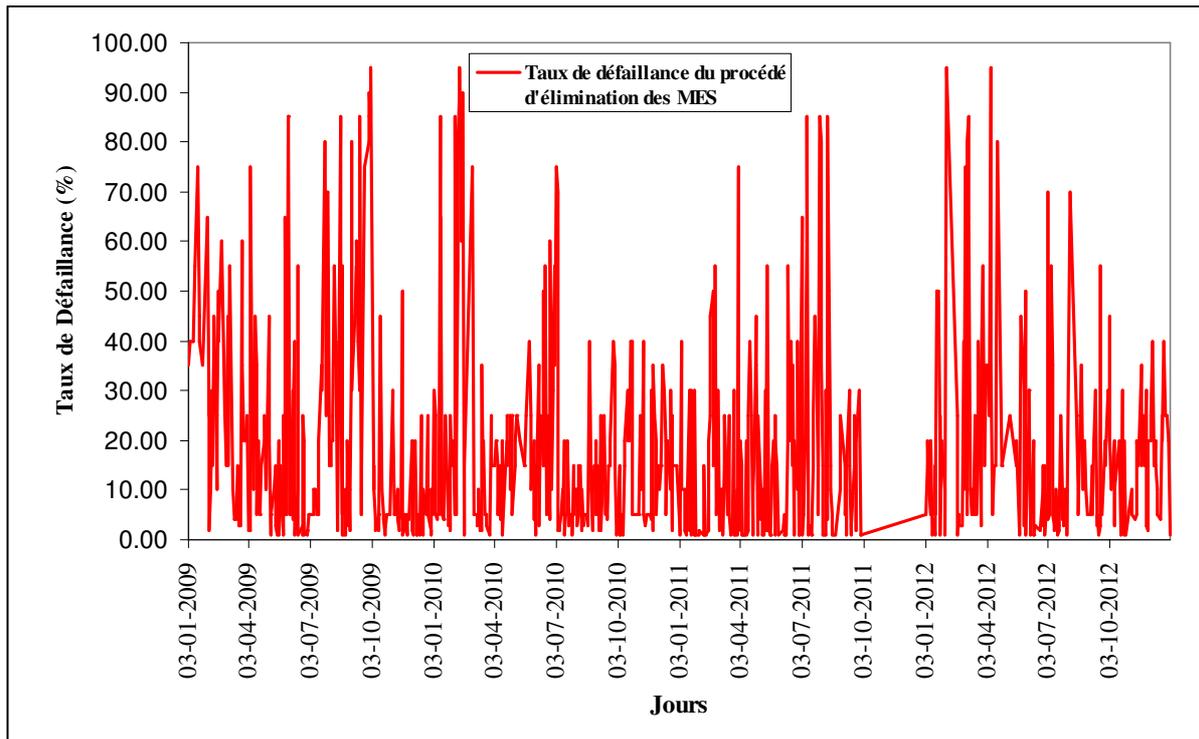


Figure 30: Courbe de la variation du taux de défiailance du procédé d'élimination des MES dans la station d'épuration de Khenchela (2009-2012).

Une variabilité du taux de défaillance est un aperçu du comportement du procédé de traitement de la station de Khenchela, qui se manifeste par un manque de rendement souhaité de la part de la station du a la variabilité excessive des paramètres de pollution à l'entrée de la station, et la présence de nombreux problèmes de fonctionnement du bassin de dessablage déshuilage (trois des quatre bassin sont hors service).

IV.9 CONCLUSION

L'analyse statistique des données de concentrations des différents paramètres de pollution des eaux traitées et rejetés montre que durant la période d'observation, le taux de fiabilité de la station est acceptable. Malheureusement ce constat n'est que relatif, puisque une grande partie des données est manquantes sur les quatre années d'exploitation ciblé par notre étude. La variabilité du taux de fiabilité du procédé de traitement de la station de Khenchela est très variable, cette variabilité quantitatives et surtout qualitatives des eaux usées brutes combiné avec une conception inadéquate avec des hypothèses de conception qui ne reflète pas la situation réelle de la station (station à faible charge). Une étude récente a montré que la station fonctionne à forte charge. Le manque d'expérience des gestionnaires de la station est un facteur important à prendre en considération.

La variabilité des taux de défaillance des différents composants du système de traitement donne un aperçu général du comportement instable de la station, malheureusement l'absence

de note de gestion rend toute tentative d'explication de l'instabilité des performances de la station une tâche rude.

CHAPITRE V : RESEAUX DE NEURONES ARTIFICIELS

V.1 INTRODUCTION

Un réseau de neurones est un système de traitement d'informations qui tente de simuler la façon dont un cerveau humain traite les signaux électriques.

Grace à leur traitement parallèle de l'information et à leurs mécanismes inspirées des cellules nerveuses (neurones), ils infèrent des propriétés émergentes permettant de solutionner des problèmes qualifiées de complexes. Ils ont été développés pour résoudre les problèmes de contrôle, de la reconnaissance des formes et des mots, de prise de décision, de mémorisation comme une alternative à l'intelligence artificielle. Les travaux dans le domaine des réseaux de neurones artificiels ont été motivés par la fascination de l'humanité à la complexité du cerveau humain

V.2 MODELE DE NEURONE

V.2.1. Neurone biologique

Le système nerveux humain est constitué de milliards de neurones de divers types et de longueurs utiles à leur emplacement dans le corps (Schalkoff,1997). Les cellules nerveuses, appelées neurones, sont les éléments de base du système nerveux central.

Les neurones possèdent de nombreux points communs dans leur organisation générale et leur système biochimique avec les autres cellules. Ils présentent cependant des caractéristiques qui leur sont propres et qui se retrouvent au niveau des cinq fonctions spécialisées qu'ils assurent, c'est-à-dire, la fonction de recevoir des signaux en provenance de neurones voisins, de les intégrer, d'engendrer un influx nerveux, de le conduire et le transmettre à un autre neurone capable de le recevoir.

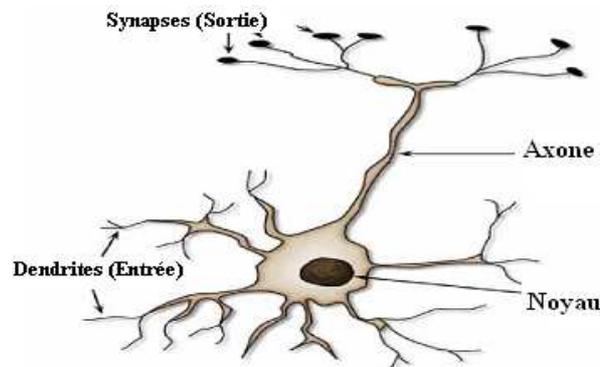


Figure 31: Structure d'un neurone biologique (Graupe, 2007).

La structure d'un neurone est constituée de trois principales parties qui sont (voir figure 31):

- Le corps cellulaire;
- Les dendrites;
- L'axone.

Le corps de la cellule a un noyau qui contient des informations sur les caractéristiques héréditaires, et un plasma qui contient le matériel moléculaire utilisée pour produire le matériau nécessaire par le neurone (Jain et al., 1996).

Les dendrites reçoivent des signaux d'autres neurones et les passer sur le corps de la cellule. La surface totale de réception des dendrites d'un neurone typique est d'environ 0,25 mm (Zupan et Gasteiger, 1993).

L'axone, qui est à proprement parler la fibre nerveuse, sert de moyen de transport pour les signaux émis par le neurone. Il se distingue des dendrites par sa forme et par les propriétés de sa membrane externe. En effet, il est généralement plus long que les dendrites, et se ramifie à son extrémité, là où il communique avec d'autres neurones, alors que les ramifications des dendrites se produisent plutôt près du corps cellulaire. La quantité de signal qui passe par un neurone récepteur dépend de l'intensité du signal issu de chacun des neurones d'alimentation, leurs forces synaptiques, et le seuil du neurone récepteur (Basheer et Hajmeer, 2000).

Le cerveau humain est composé de dix milliards de neurones et chaque neurone peut interagir avec aussi peu qu'un mille ou moins de deux cent mille autres neurones. En raison du grand nombre d'interconnexions, le cerveau est capable de s'adapter à des signaux externes et faire des réactions raisonnables (Park, 1996).

Le cerveau est capable de réagir aux signaux qu'il n'a jamais connu auparavant en raison des connexions entre les neurones massives interagissant les uns avec les autres. Les bonnes réactions sont dues à la capacité de programmation génétique et la capacité d'apprentissage du cerveau à réagir aux événements. Ce fait est l'objectif principal de l'élaboration d'un réseau de neurones artificiels (RNA), ou simplement un réseau de neurones (RN) (Zupan et Gasteiger, 1993).

V.2.2. Neurone formel (artificiel)

Pour reproduire le neurone biologique, on se sert d'un modèle mathématique La figure 32 illustre la structure d'un neurone artificiel. Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones amont. A chacune de ces entrées est associé un poids w représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones avals. A chaque connexion est associé un poids.

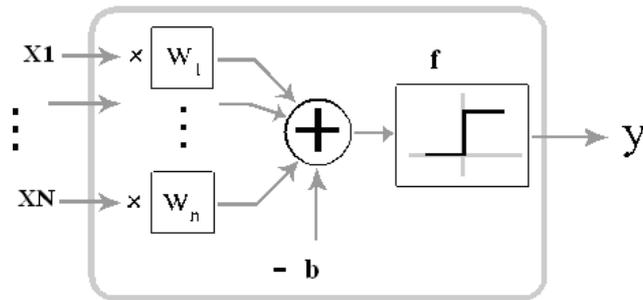


Figure 32: Neurone formel (artificiel).

Le neurone formel. Il doit être capable de :

1. Recevoir en entrée différentes informations provenant des neurones environnants ;
2. Analyser ces informations, de manière à envoyer en sortie une réponse ;
3. Ajuster cette réponse avant de l'envoyer aux neurones suivants.

Un neurone est l'unité de traitement de base d'un réseau de neurones. La figure 32 illustre un neurone d'entrée avec Les entrées x_1 à x_n constituent le vecteur d'entrée x . La fonction f est appelée fonction d'activation du neurone. Chaque entrée x_i est pondérée par un poids w_i , ce sont les poids des entrées qui vont permettre au neurone d'apprendre et de modifier sa sortie au fur et à mesure de l'apprentissage. La sortie du neurone est $y = f(\bar{w} \cdot \bar{x})$, où $\bar{w} \cdot \bar{x}$ représente le produit scalaire du vecteur poids par le vecteur des entrées. Le biais b fournit une variable supplémentaire qui peut être ajusté pour obtenir la performance du réseau désiré. Cette somme est l'argument de la fonction de transfert f , qui prend l'argument et produit une sortie. La fonction de transfert qui limite la gamme d'amplitude admissible du signal de sortie à des valeurs finies dans l'intervalle $[-1, 1]$. Les poids w_i et le biais b sont des paramètres ajustables qui permettent au réseau de neurones de présenter le comportement désiré (Haykin 1999 ; Krenker et al., 2011 ; Demuth, Beale et Hagan, 2013).

V.2.3 Réseau de neurones artificiels

Haykin (1999), définit un réseau de neurones, comme un processeur distribué massivement parallèle qui est composé d'unités de traitement simples, ayant la capacité de stocker des connaissances empiriques et de les rendre disponible pour l'utilisation.

Un réseau de neurones artificiel (figure 33) est en général composé d'une succession de couches dont chacune prend ses entrées sur les sorties de la précédente. Chaque couche (i) est composée de N_i neurones, prenant leurs entrées sur les N_{i-1} neurones de la couche précédente. À chaque synapse est associé un poids synaptique, de sorte que les N_{i-1} sont multipliés par ce poids, puis additionnés par les neurones de niveau i, ce qui est équivalent à multiplier le vecteur d'entrée par une matrice de transformation. Mettre l'une derrière l'autre les différentes couches d'un réseau de neurones reviendrait à mettre en cascade plusieurs matrices de transformation et pourrait se ramener à une seule matrice, produit des autres, s'il n'y avait à

chaque couche, la fonction de sortie qui introduit une non linéarité à chaque étape (Touzet, 1992). Ceci montre l'importance du choix judicieux d'une bonne fonction de sortie : un réseau de neurones dont les sorties seraient linéaires n'aurait aucun intérêt.

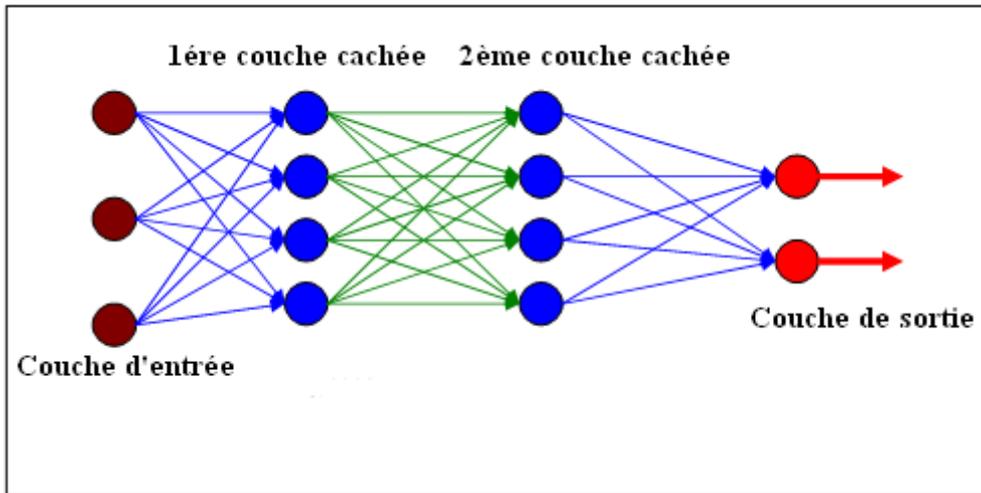


Figure 33: Réseau de neurones multicouches (PMC) avec deux couches cachées (Rustum, 2009).

V.3 HISTORIQUE DES RECHERCHES SUR LES RESEAUX DE NEURONES

L'histoire des recherches sur les réseaux de neurones peut être retracée par les tentatives des chercheurs d'essayer de modéliser le neurone biologique. La motivation originale pour les réseaux de neurones découle des travaux de McCulloch et Pitts (1943). Le modèle qu'ils ont créé a deux entrées et une seule sortie. Les poids pour les entrées ont été maintenues égaux et la sortie était soit logique 1 ou 0 logique. La sortie restera à zéro jusqu'à ce que la somme pondérée des entrées ait atteint un certain seuil. Le neurone de McCulloch et Pitts est connu comme un circuit logique.

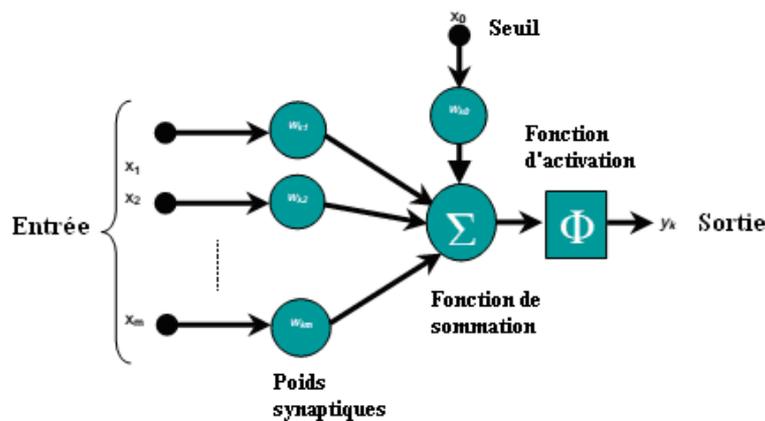


Figure 34: Structure d'un neurone artificiel qui calcule la somme de ses entrées puis cette valeur passe à travers la fonction d'activation pour produire sa sortie (Krieger, 2007).

La sortie du neurone est donné par:

$$y_k = f(w_{k0}x_0 + \sum_{j=1}^m w_{kj}x_j) \dots\dots(V.1)$$

où:

y_k est la sortie du neurone, et $w_{k0} x_0$ est le poids de sollicitation et de l'ampleur, x_j est le signal d'entrée et w_{kj} sont les poids. Φ est la fonction d'activation calculée au niveau du nœud de sommation.

Donald Hebb (1949) a été préoccupé par les lois d'adaptation impliquée dans les systèmes neuronaux, et en 1949 a conçu une règle d'apprentissage pour adapter les connexions au sein de neurones artificiels.

Le Perceptron fut le modèle de neurone suivant développé par Rosenblatt (1958). Les perceptrons sont particulièrement adaptés pour des problèmes simples dans la classification des modèles.

Rosenblatt a interconnecté au hasard les perceptrons et procéder à des essais pour changer aléatoirement les poids pour parvenir à un "apprentissage". Le neurone de McCulloch et Pitts est cependant un meilleur modèle pour simuler le processus électrochimique qui se passe à l'intérieur d'un neurone biologique, à l'inverse du perceptron, qui est la base des réseaux de neurones des temps modernes (Anderson, Rosenfeld et Pellionisz, 1993). Malheureusement, l'inconvénient majeur du perceptron est qu'il est seulement capable de classer les problèmes qui sont linéairement séparables à la sortie (Masters 1993 ; Demuth, Beale et Hagan, 2013).

Vers 1954, Minsky introduit l'apprentissage par renforcement .L'une des évolutions majeures a été la formulation de nouvelles règles d'apprentissage par Widrow et Hoff (1960) dans leur application à une simple adaptation linéaire neurone (Adaline) (Widrow et Hoff, 1960). La règle était aussi connue sous le nom d'algorithme des moindres carrés (LMS).

Une percée importante dans les réseaux de neurones a été faite en 1967 avec l'introduction de la fonction d'activation sigmoïde lisse par Cowan. Cette fonction a la capacité d'approximativement une fonction non linéaire. Au lieu de passage ou hors comme le perceptron, cette fonction d'activation active la sortie progressive lorsqu'il est activé.

Minsky et Papert (1969) publièrent un livre dans lequel ils ont montré les lacunes des modèles de perceptron, ayant pour effet l'orientation des fonds de recherches sur les réseaux de neurones vers d'autres domaines de recherches, ce qui a conduit de nombreux chercheurs à changer de domaine de recherche. Seuls quelques chercheurs ont continué notamment Teuvo Kohonen, Stephen Grossberg, James Anderson et Kunihiko Fukushima (Kröse and van der Smagt, 1996). L'intérêt pour les réseaux de neurones artificiels a connu une résurrection après quelques résultats théoriques importants dans les années 1980 (notamment l'algorithme de retro propagation).

Werbos (1974) présente la méthode d'apprentissage par rétro-propagation, mais cela n'a pas reçu beaucoup de publicité. Cette méthode d'apprentissage traite la propagation de l'information sur les erreurs ont la réinjectant dans la couche cachée.

Parker (1985) et Le Cun (1986) rajeuni l'intérêt pour l'algorithme de rétro-propagation. La méthode de rétro-propagation d'erreur publiée par Rumelhart et Williams (1986), pourrait former des réseaux de neurones multicouches pour effectuer des tâches où le perceptron avait échoué. C'est après cette publication que la rétro-propagation a été considéré comme une percée dans la littérature des réseaux de neurones artificiels avec l'introduction de la règle de delta généralisée. La méthode de rétro-propagation a été utilisé avec succès dans de nombreuses applications telles que les opérations de filtrage, de reconnaissance de chiffres, de manuscrits, des systèmes de contrôle (Ng, 1997).

V.4 ARCHITECTURE DES RESEAUX DE NEURONES ARTIFICIELS

En générale, Il y'a essentiellement deux types de topologies de réseau de neurones artificiels (RNA), à savoir *réseau de neurones artificiel directe* ou (par alimentation avant) (Feed-forward) et le *réseau de neurones artificiels récurrent* (RNAR) (Bishop, 1995; Demuth, Beale et Hagan, 2013).

V.4.1 Réseau de neurones artificiels direct (à alimentation avant)

Un réseau de neurones artificiels direct est un réseau neuronal artificiel où des connexions entre les unités ne forment pas un cycle dirigé. Ceci est différent de réseaux de neurones récurrents.

Le réseau neuronal direct a été le premier et le plus simple type de réseau de neurones artificiels imaginés. Dans ce réseau, l'information se déplace dans une seule direction, vers l'avant, à partir des nœuds d'entrée, par l'intermédiaire des nœuds cachés (le cas échéant) et de nœuds de sortie. Il n'y a pas de cycles ou des boucles dans le réseau.

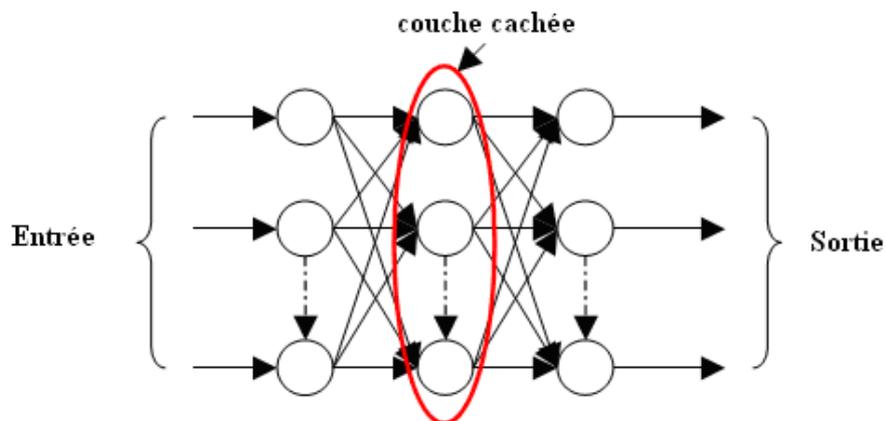


Figure 35: Réseau de neurones artificiels directs (par alimentation avant) (Feed-Forward) (Kriger, 2007).

V.4.2 Réseau de neurones artificiels récurrent

Un réseau de neurones artificiels récurrent (RNAR) est une classe de réseau de neurones, où les connexions entre les unités forment un cycle dirigé. Ceci crée un état interne du réseau qui lui permet un comportement temporel dynamique. Contrairement aux réseaux de neurones directs (feed-forward), les RNAR peuvent utiliser leur mémoire interne pour traiter les séquences arbitraires d'intrants. Cela les rend applicable à des tâches telles que la reconnaissance de l'écriture manuscrite connecté non segmenté, où ils ont obtenu les meilleurs résultats connus (Graves et al., 2009).

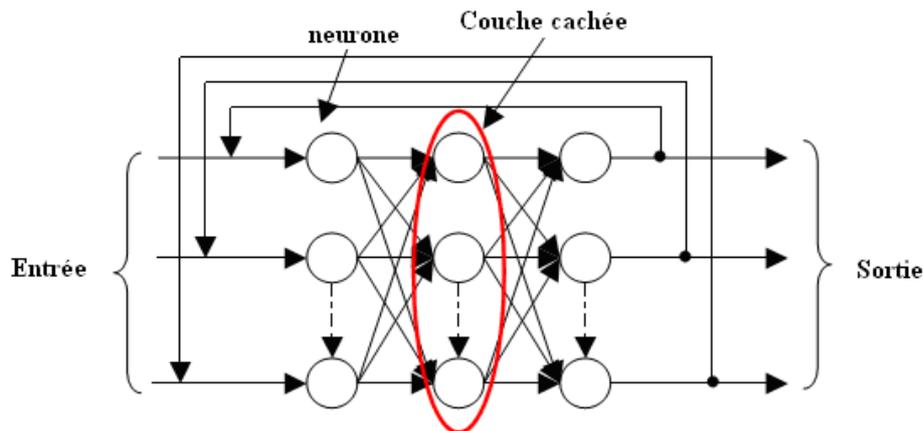


Figure 36: Réseau de neurones récurrent (Kriger, 2007).

Il existe un grand nombre de topologies de réseau neuronal différent. Pour une liste plus complète des différentes topologies de réseau neuronal avec les sources consultées voir l'annexe A. Cette liste est loin d'être les seuls types de réseaux neuronaux qui sont disponibles.

Trouver une configuration optimale du réseau est l'un des problèmes les plus importants dans la conception d'un réseau de neurones artificiels. Les paramètres du réseau de neurones suivants doivent être sélectionnés:

- le nombre de couches cachées;
- le nombre de neurones dans chaque couche cachée;
- le type de fonction d'activation, qui peut varier dans la même couche.

Ces paramètres doivent être choisis avec soin particulier pour éviter les phénomènes de sur-apprentissage et sous-apprentissage et réaliser une phase d'apprentissage convergent (Khan et Ondrůšek 2000). En raison d'un manque d'informations sur les variables représentées dans la couche cachée, les réseaux de neurones artificiels sont considérés par la plupart des chercheurs comme un modèle « boîte noire » (Hu, 1997 ; Moshiri et Cameron, 2000 ; Olden et Jackson, 2002).

V.4.3 Fonction de transfert

La sortie du neurone est déterminé par l'activation (ou transfert) en fonction du neurone. Le but de la fonction d'activation est d'introduire une non-linéarité dans le réseau. Cette non-linéarité associée aux interconnexions des neurones accomplit une cartographie des données propre à partir de signaux d'entrée pour les activités de sortie correspondants (Zupan et Gasteiger, 1993). Sans cette non-linéarité du perceptron multicouche ne serait pas fonctionnel différent d'un filtre linéaire et ne serait pas en mesure d'effectuer la séparation non linéaire et la trajectoire d'apprentissage pour les signaux non linéaires et non stationnaires (Mandic et Chambers, 2001).

La fonction de transfert peut être aussi simple comme la fonction binaire appelé aussi limiteur dur ou la fonction par étage. Le signal de sortie dépend du fait que le produit est positif ou négatif. La sortie est à '1' pour un produit positif et '0' pour un produit négatif. La sortie peut également être bipolaire entre $[-1, +1]$. Les principales fonctions de transfert sont présentées dans le tableau 20 ci-dessous.

Tableau 20: Les principales fonctions de transfert

Nom	Valeur	Représentation
Seuil	$f(x) = 0$ si $x \leq 0$ $f(x) = 1$ si $x > 0$	
Seuil symétrique	$f(x) = -1$ si $x < 0$ $f(0) = 0$ $f(x) = 1$ si $x > 0$	
Linéaire saturée	$f(x) = 0$ si $x < 0$ $f(x) = x$ si $0 \leq x \leq 1$ $f(x) = 1$ si $x > 1$	
Linéaire	$f(x) = x$	
Sigmoïde	$f(x) = \frac{1}{1 + e^{-x}}$	
Tangente hyperbolique	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	

Il existe d'autres fonctions d'activation moins populaires tels que la fonction satlins (linéaire saturée symétrique), la fonction poslin (linéaire positive), et la fonction Softmax (Burton, 2006). Le succès ou l'échec du processus d'apprentissage dépend du fait que la fonction d'activation est dérivable ou non. Sur la base de cette exigence, la fonction non-linéaire la plus populaire est la fonction sigmoïde pour la couche cachée en raison de son dérivé facilement calculé (Rustum, 2009). Parmi les fonctions non-linéaires les plus populaires, la fonction logistique et la fonction tangente hyperbolique (Demuth, Beale et Hagan, 2013).

V.5 ENTRAINEMENT ET APPRENTISSAGE DU RESEAU DE NEURONES

L'entrainement est la façon dont un RNA apprend. Les algorithmes d'entrainement ou d'apprentissage sont des procédures de modification des poids sur les liaisons dans la structure du réseau de neurones artificiels, il existe 3 types d'algorithmes d'apprentissage (Ng, 1997; Haykin, 1999; Krenker, 2011):

- L'apprentissage supervisé ;
- L'apprentissage non supervisé ;
- Apprentissage par renforcement.

V.5.1 Apprentissage supervisé

L'apprentissage supervisé, un ensemble d'entrées et sorties adéquates sont utilisées pour former le réseau. Le réseau produit ensuite ses propres sorties. Ces sorties sont comparées avec les sorties correctes et la différence (erreur) est utilisée pour modifier les poids. Ce type d'apprentissage est utilisé dans la thèse en raison de la nature des données disponibles (Haykin, 1999; Krenker, 2011).

V.5.2 Apprentissage non supervisé

Apprentissage non supervisé ou auto-organisation est l'endroit où un réseau développe ses propres règles de classification par l'extraction d'informations à partir des entrées présentées au réseau.

1. La formation des lots est l'endroit où les modifications apportées aux poids et les préjugés sont prises en fonction de l'application de l'ensemble des vecteurs de données sur le réseau.
2. Formation progressive est l'endroit où les modifications apportées aux poids et les préjugés sont effectués après l'application de chaque vecteur de données d'entrée individuel. Formation incrémentielle est également considérée comme une formation en ligne ou de formation adaptative.

V.5.3 Apprentissage par renforcement

L'apprentissage par renforcement est un cas particulier de l'apprentissage supervisé où il n'existe pas d'ensemble des entrées et sorties adéquates. Le réseau est juste dit si une sortie produite est «bon» ou «mauvais», selon un critère défini.

L'apprentissage par renforcement fait référence à une classe de problèmes d'apprentissage automatique, dont le but est d'apprendre, à partir d'expériences, ce qu'il convient de faire en différentes situations, de façon à optimiser une récompense numérique au cours du temps.

Un paradigme classique pour présenter les problèmes d'apprentissage par renforcement consiste à considérer un agent autonome, plongé au sein d'un environnement, et qui doit prendre des décisions en fonction de son état courant. En retour, l'environnement procure à l'agent une récompense, qui peut être positive ou négative.

L'agent cherche, au travers d'expériences itérées, un comportement décisionnel appelé "stratégie", et qui est une fonction associant à l'état courant l'action à exécuter) optimal, en ce sens qu'il maximise la somme des récompenses au cours du temps.

Des algorithmes de formation utilisent des données historiques (ensemble de données d'apprentissage) pour régler automatiquement les coefficients de pondération et des seuils afin de minimiser l'erreur de prédiction. La méthode utilisée pour rajuster les poids sur les connexions s'appelle une règle d'apprentissage. Il ya beaucoup de règles d'apprentissage couramment utilisés, mais il ya certains qui se sont révélés être efficaces dans la résolution de certains types de problèmes (Haykin, 1999 ; Krenker, 2011).

V.5.4 Règle d'apprentissage du perceptron

Un perceptron est un réseau de neurones avec des neurones «produit scalaire» fonctionnant avec une fonction par étage ou sigmoïde (la fonction est un cas particulier de la fonction sigmoïde). Faire apprendre un neurone, consiste à régler (estimation adéquate) de ses poids de manière à ce que la sortie du neurone évolue dans le sens que l'on souhaite sans bien sûr changer les entrées. Il s'agit dans ce cas d'un apprentissage supervisé (on spécifie la sortie que l'on souhaiterait voir, et le neurone va adapter ses poids pour essayer de s'approcher de la valeur de la sortie. On appelle cette sortie la *cible* du neurone). La règle d'apprentissage du perceptron va donc faire évoluer les poids du neurone vers une cible que l'on a spécifiée (Heaton, 2008 ; Prevost, 2012).

1- Initialisation

Mettre les poids initiaux w_1, w_2, \dots, w_n ainsi que le seuil θ à des valeurs aléatoires de l'intervalle $[-0,5, 0,5]$. Mettre le taux d'apprentissage α à une petite valeur positive

2- Activation

Activer le perceptron en appliquant les intrants $x_1(p), x_2(p), \dots, x_n(p)$ et l'extrant désiré $Y_d(p)$. Calculer l'extrant actuel à l'itération $p = 1$.

$$Y(p) = \text{étage} \left[\sum_{i=1}^n x_i(p)w_i(p) - \theta \right]$$

où n est le nombre de signaux intrants, et *étage* est la fonction d'activation par étage.

3- Entraînement des poids

Mettre à jour les poids du perceptron

$$w_i(p+1) = w_i + \Delta w_i$$

où $\Delta w_i(p)$ est la correction de poids à l'itération p.

La correction de poids est calculée par la loi delta :

$$\Delta w_i = \alpha \times x_i(p) \times e(p)$$

où α est le taux d'apprentissage et $e(p)$ l'erreur à l'itération p (la différence entre l'extrait désiré et l'extrait actuel du perceptron)

4- Itération

Augmenter p de 1, retourner à l'étape 2 et répéter le procédé jusqu'à convergence (jusqu'à ce que les erreurs atteignent un niveau satisfaisant).

V.5.5 L'algorithme des moindres carrés (LMS)

L'algorithme d'apprentissage des moindres carrés (LMS) est parfois connu comme la règle delta ou la règle d'apprentissage Widrow-Hoff. La correction des coefficients de pondération est déterminé par:

$$\Delta w = \frac{\beta \cdot E \cdot x}{|x|^2} \dots\dots(V.2)$$

$$E = y_d - y_m \dots\dots(V.3)$$

Où:

y_d est la sortie désirée;

y_m est la sortie du modèle réel.

L'erreur est basée sur la somme des entrées de l'unité de S.

$$e_j(k) = O_{dj}(k) - S_j(k) \dots\dots(V.4)$$

La somme linéaire des entrées S est passée à travers une fonction tangente qui produit la sortie "+1" ou "-1" en fonction de la polarité de la somme.

L'algorithme LMS est basé sur les valeurs instantanées de la fonction de coût (Haykin, 1999):

$$\zeta(w) = \frac{1}{2} e^2(n) \dots\dots(V.5)$$

où $e(n)$ est le signal d'erreur mesuré à l'instant n.

Différencier $\zeta(w)$ par rapport au rendement du vecteur de poids w .

$$\frac{\partial \zeta(w)}{\partial w} = e(n) \frac{\partial e(n)}{\partial w} \dots \dots \dots (V.6)$$

V.6 ALGORITHME D'APPRENTISSAGE

V.6.1 Algorithme d'apprentissage de retro propagation (BP)

La rétro-propagation est un exemple d'apprentissage supervisé. Dans de nombreuses applications des réseaux neuronaux sont essentiellement des réseaux fonction de cartographie. Le problème avec la cartographie est de découvrir la relation entre les entrées et les sorties du réseau cible. Cette relation est selon toute vraisemblance, probablement non linéaire, comme c'est le cas avec le procédé par boues activées. L'algorithme de rétro propagation a été effectivement utilisé dans de nombreuses applications nécessitant une cartographie d'entrée-sortie non linéaire. Avec la possibilité d'approcher une fonction non-linéaire continue, le réseau de rétro propagation a cartographie extraordinaire (prévision) capacités (Rui et El-Keib, 1995).

Le terme se rapporte à rétro-propagation de la manière dont le gradient est calculé pour les réseaux multicouches non linéaires. La généralisation de la règle d'apprentissage Widrow-Hoff aux réseaux multicouches et des fonctions de transfert non linéaires différentiables créé l'algorithme de rétro propagation.

L'algorithme est basé sur la plus grande pente (descente de gradient) techniques accordés à chacune des couches du réseau par la règle de la chaîne (Ng, 1997). La dérivée partielle de la fonction d'erreur par rapport aux poids est calculée en tant que:

$$\nabla E(k) = \frac{\partial E(k)}{\partial W(k-1)} \dots \dots \dots (V.7)$$

avec:

W: Vecteur des poids.

E(k) : Fonction d'erreur

La fonction d'erreur est définie comme:

$$E(k) = \frac{1}{2} [y_p(k) - y_m(k)]^2 \dots \dots (V.8)$$

où y_p est la sortie du réseau et $e_m(k) = y_p(k) - y_m(k)$.

La constante $\frac{1}{2}$ est choisie pour faciliter le calcul d'une dérivée de la fonction de coût qui est essentiel dans l'estimation des paramètres.

L'objectif est de minimiser la fonction d'erreur $E(k)$ en prenant le gradient d'erreur par rapport au vecteur des poids, W doit être adapté Les pondérations sont mises à jour à l'aide:

$$W(k) = W(k-1) + \eta \frac{\partial E(k)}{\partial W(k-1)} \dots\dots\dots(V.9)$$

avec:

η est le taux d'apprentissage et:

$$\frac{\partial E(k)}{\partial W(k-1)} = -e_m(k) \frac{\partial y_m(k)}{\partial W(k-1)} = -e_m(k) \frac{\partial O_o(k)}{\partial W(k-1)} \dots\dots(V.10)$$

Il devient évident que le succès de la technique de rétro-propagation dépend de trois facteurs: le taux d'apprentissage η , la distance entre la production réelle et la production prévue, et la fonction d'activation. Le taux d'apprentissage détermine la vitesse du réseau se règle en réponse à des erreurs. Si le taux d'apprentissage est trop élevé, les poids ajustés oscillent entre des points avec des hauteurs similaires sur la surface d'erreur (Park, 1996) et le réseau ne converge pas vers un minimum. Si le taux d'apprentissage est trop faible, l'entraînement est lent et le réseau est piégé dans un minimum local. Le taux d'apprentissage devrait se situer dans la gamme $0 \leq \eta \leq 1$.

L'algorithme de rétro propagation pourrait se résumer comme suit:

- Un ensemble d'entrées et de sorties d'entraînement cible correspondant sont fournis sur le réseau;
- Le réseau calcule les signaux d'erreur, et il est utilisé pour ajuster les poids;
- Après avoir beaucoup en amont et en passes, l'erreur de réseau atteint un minimum sur les données d'apprentissage;
- Le réseau est testé sur des données expérimentales, il n'a pas vu avant de mesurer la capacité de généralisation du réseau.

Un seul passage des deux premières étapes ci-dessus est connu comme une époque. La formation dure habituellement dans certains cas, des milliers d'époques que l'erreur du réseau est inférieur à un certain seuil. De cela, il est évident que la modélisation des réseaux de neurones est extrêmement chronophage (Haykin, 1999 ; Krenker, 2011).

V.6.2 Limites de l'algorithme d'apprentissage retro propagation

Limites de l'algorithme d'apprentissage de rétro propagation sont donnés ci-dessous:

- Des temps d'entraînement long, donc une vitesse de convergence relativement lente;
- Si l'ensemble des données d'entraînement et la stratégie d'apprentissage inadéquate aura pour résultat la paralysie du réseau. Si les poids deviennent trop grands, les changements dans les poids deviennent minimes (le dérivé de la fonction sigmoïde est très faible);
- Aucune garantie pour la convergence vers un minimum global; parfois la méthode peut rester coincé dans un minimum local (voir figure 37 ci-dessous);
- Il n'existe pas de règle précise pour définir le nombre de neurones, et le nombre de couches pour une meilleure performance;
- Présence du phénomène d'instabilité si le taux d'apprentissage est trop grand.

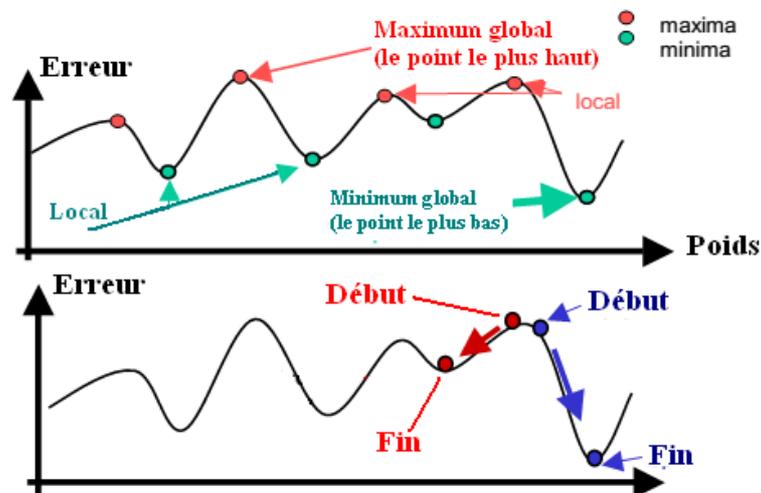


Figure 37: Algorithme d'apprentissage coincé dans un minimum local

Malgré les limites ci-dessus à l'algorithme de rétro-propagation, il est toujours très populaire, en particulier dans le milieu du contrôle des systèmes. Il existe cependant de nombreuses améliorations à l'algorithme de rétro propagation commun. Une amélioration est l'ajout d'un terme du moment (Rustum, 2009, Kriger, 2007).

V.6.3 Amélioration heuristique pour l'algorithme de rétro propagation

Etant les poids oscillant, et pour éviter une trajectoire d'erreur dans l'espace, une méthode est proposé par l'ajout d'un terme de mouvement (Papliński, 2004). Le moment est introduit pour réduire le risque d'oscillations pour un taux d'apprentissage donné. Le moment permet à un réseau de répondre non seulement à la pente locale, mais aussi à l'évolution de la surface d'erreur (Demuth, Beale et Hagan, 2013).

Le terme de moment est une modification de la règle delta généralisée en incluant un terme de mise à jour dans l'équation des poids (équation V.11). L'ajout de ce terme peut réduire le temps d'entraînement du réseau et accélérer sa convergence.

$$W(k) = W(k-1) + (-\eta \frac{\partial E(k)}{\partial W(k-1)}) + \alpha \Delta W(k-1) \dots\dots(V.11)$$

Le paramètre α (terme de moment) est choisi entre $0 < \alpha < 1$. Si le terme de moment est réglé sur 0, le changement de poids est uniquement basé sur le gradient. Si le terme de moment est mis à 1, le nouveau changement de poids est égal à la dernière modification du poids, alors que le gradient est ignoré. La pente est calculée en faisant la somme des gradients calculés après chaque exemple d'entraînement.

Une telle modification de la loi d'apprentissage agit comme un filtre passe-bas pour le lissage de la trajectoire d'erreur. En conséquence, il est possible d'appliquer un taux d'apprentissage (η) plus élevé,

Nagata et al. (1990) ont proposé une modification de l'équation (V.11), où β est une constante décidée par l'utilisateur.

$$W(k) = W(k-1) + (-\eta \frac{\partial E(k)}{\partial W(k-1)}) + \alpha \Delta W(k-1) + \beta \Delta W(k-2) \dots\dots(V.12)$$

Toutefois, il s'agit d'une répétition du rôle de α et l'avantage d'ajouter β n'est pas très clair (Ng, 1997).

Malgré les tentatives d'amélioration de l'algorithme de rétro propagation traditionnel en ajoutant des termes de moment et d'autres méthodes, il existe d'autres méthodes de gradient plus efficace tel que le gradient conjugué à l'échelle.

V.7 GENERALISATION

Il est rarement utile d'avoir un simple réseau de neurones artificiels qui mémorise un ensemble de données (approximation de fonction), puisque de nombreux algorithmes pour consultation peuvent faire la mémorisation beaucoup plus efficacement.

En règle générale, il est nécessaire que l'RNA soit capable d'effectuer avec précision une généralisation sur de nouvelles données.

La généralisation peut être définie comme la capacité du réseau à faire une réponse hypothétique à une entrée qui n'a pas été vu avant, et de faire une réponse raisonnable (Park, 1996).

V.7.1 Sur-apprentissage et sous-apprentissage

Sur-apprentissage ou sous-apprentissage est essentiellement la façon dont le réseau de neurones artificiels réalise des prédictions pour les cas qui ne sont pas dans le jeu de données d'entraînement, comme d'autres méthodes d'estimation non linéaire.

En règle générale, le but de l'entraînement est de faire des prédictions pour données de sortie à venir dans lequel seules les entrées du réseau sont connues. Le résultat de la formation du réseau conventionnel est un ensemble unique de poids qui peuvent être utilisés pour faire de telles prédictions (Kriger, 2007).

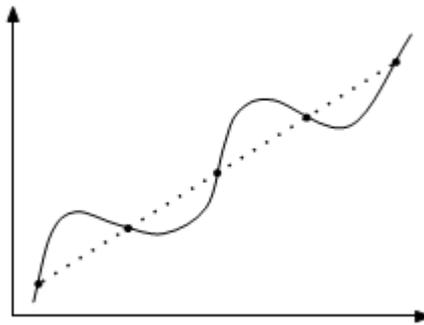


Figure 38 : Présentation graphique du phénomène de sur-apprentissage
"Cas d'approximation d'une fonction" (Parizeau, 2006).

Par exemple, la figure 38 illustre le problème du sur-apprentissage dans le contexte d'une tâche d'approximation d'une fonction. La droite en pointillé montre une fonction linéaire que l'on voudrait approximer en ne connaissant que les points noirs. La courbe en trait plein montre ce qu'un réseau hypothétique pourrait apprendre. On constate que la courbe passe par tous les points d'entraînement et donc que l'erreur est nulle. De toute évidence, ce réseau ne généralisera pas bien si l'on échantillonne d'autres points sur la droite (Parizeau, 2006).

Le phénomène de sous-apprentissage est le contraire du sur-apprentissage. Cela se produit lorsque le modèle est incapable de capturer la variabilité des données. Le réseau de neurones artificiels n'aura aucun pouvoir prédictif, ni qu'il sera en mesure de cartographier correctement les données d'entraînement. Ceci est le résultat d'un sous-apprentissage, ou de tenter d'utiliser ce modèle qui est trop simple pour décrire un ensemble de données. La figure 39 présente graphiquement le phénomène de sous-apprentissage.

La compréhension de ces deux phénomènes permet d'aller dans l'espace entre les deux extrêmes. C'est dans cet intervalle lorsque le modèle a un pouvoir prédictif de l'ensemble des données de validation (Parizeau, 2006).

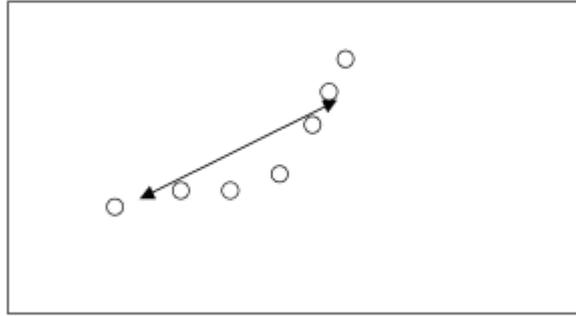


Figure 39 : Exemple du phénomène de sous-apprentissage (Parizeau, 2006).

Une solution à ce problème consiste à utiliser un autre critère d'arrêt et basée sur une technique dite de validation croisée (en anglais «cross-validation»). Cette technique consiste à utiliser deux ensembles indépendants de données pour entraîner le réseau : un pour l'apprentissage (l'ajustement des poids) et l'autre pour la validation, c'est-à-dire le calcul d'un indice de performance (Parizeau, 2006).

V.7.2 Arrêt prématuré

Le but de l'entraînement du réseau n'est pas d'apprendre la représentation exacte des données d'entraînement par lui-même, mais de construire un modèle du processus qui génère ces données, il est important que le réseau présente une bonne généralisation. L'arrêt prématuré est la technique la plus utilisée pour résoudre le problème du sur-apprentissage et de trouver le réseau ayant la meilleure performance sur de nouvelles données. Cette arrêt prématuré implique la séparation des données disponibles en 3 sous-ensembles: *ensemble d'apprentissage*, *ensemble de validation* et *ensemble de test*. Au cours de l'entraînement du réseau, le message d'erreur sur l'ensemble de validation est contrôlé, ainsi que l'erreur sur l'ensemble d'apprentissage (Bishop, 1995; Demuth, Beale et Hagan, 2013).

Le critère d'arrêt consiste alors à stopper l'apprentissage lorsque l'indice de performance calculée sur les données de validation cesse de s'améliorer pendant plusieurs périodes d'entraînement. La figure 40 illustre le critère de la validation croisée dans le cas d'un indice de performance que l'on cherche à minimiser. La courbe en pointillées de ce graphique représente l'indice de performance d'un réseau Hypothétique calculée sur les données d'apprentissage, alors que la courbe en trait plein montre le même indice mais calculée sur les données de validation. On voit qu'il peut exister un moment au cours de l'apprentissage où l'indice en validation se détériore alors que le même indice continue à s'améliorer pour les données d'entraînement. C'est alors le début du «sur-apprentissage» (Parizeau, 2006; Adeloje et De Munari, 2006).

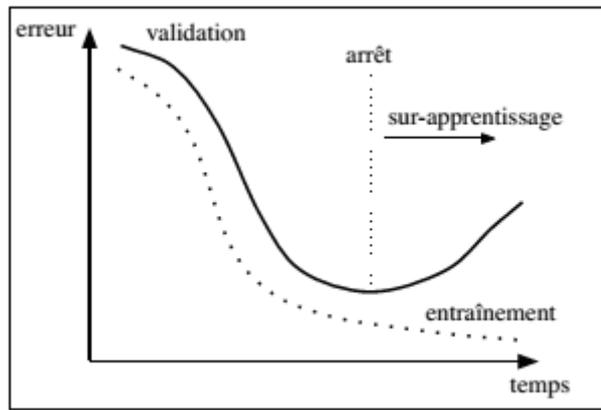


Figure 40 : Illustration du critère de la validation croisée (Parizeau, 2006)

Il ya aussi beaucoup d'autres problèmes importants qui sont si difficiles qu'un réseau de neurones est incapable de les apprendre sans avoir à mémoriser l'ensemble de la formation, tels que:

- Prévoir des nombres aléatoires ou pseudo-aléatoires.
- Factoriser de grands nombres entiers.
- Déterminer si un grand nombre entier est premier ou composite.
- Décryptage rien chiffré par un bon algorithme.

V.7.3 Régularisation

La régularisation est une méthode pour améliorer la capacité de généralisation d'un RNA. Les statisticiens utilisent cette technique pour trouver l'équilibre entre le nombre de paramètres et la bonté de l'ajustement en pénalisant les gros modèles. La fonction de performance est modifiée de telle sorte que l'algorithme élague le réseau par la conduite estimations non pertinents à zéro (Rech, 2002). Cela entraîne une modification de la fonction de performance telle que la somme des carrés (erreur quadratique moyenne) de l'erreur du réseau sur la base d'apprentissage.

$$F = mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 \dots (V.13)$$

Il est possible d'améliorer la généralisation en modifiant la fonction des performances en ajoutant un terme qui se compose de la moyenne de la somme des carrés des poids du réseau et les biais (Demuth, Beale et Hagan, 2013).

$$msereg = \gamma.mse + (1 - \gamma).msw \dots (V.14)$$

Où:

γ : Indice de performance,

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \dots\dots\dots(V.15)$$

Grâce à cette fonction de performance provoque le réseau d'avoir moins de poids et les biais, et cela force la réponse du réseau pour être plus lisse et moins susceptibles de sur ajustement.

Le principal avantage de l'utilisation de la régularisation, c'est que même si le RNA est sur-paramétré, les estimations des paramètres pertinents sont susceptibles d'être proche de zéro et le RNA se comporte comme un petit réseau (Rech, 2002).

V.8 AVANTAGES DES RESEAUX DE NEURONES ARTIFICIELS

Les réseaux de neurones sont caractérisés par les capacités positives suivantes:

1. Pour résoudre des tâches où il est avantageux d'utiliser une machine, mais où il est impossible de répondre du programme à toutes les issues possibles.
2. Apprendre des exemples limités «les réseaux de neurones artificiels sont particulièrement adaptés à des problèmes où la solution est complexe et difficile à définir, mais fournit une abondance de données à partir de laquelle une réponse peut être apprise (Tarassenko, 1998).
3. Capacité à généraliser, à savoir la carte entrées semblables aux sorties similaires; les réseaux de neurones sont capables d'interpoler à partir d'une expérience d'apprentissage précédent. Avec une conception soignée, un RNA peut être formé pour donner la bonne réponse à des données qu'il n'a pas préalablement rencontré (Tarassenko, 1998).
4. Possibilité de cartographier les fonctions linéaires et non linéaires : la cartographie non linéaire donne souvent aux réseaux de neurones artificiels l'avantage de traiter les problèmes complexes du monde réel.
5. Efficacité de calcul : la formation d'un RNA pour un calcul intensif peuvent être modeste (Tarassenko, 1998), mais les exigences de calcul d'un RNA entièrement formés peuvent être amélioré pour des problèmes plus importants, grâce à un traitement parallèle (Haykin, 1999).
6. Robuste en présence de bruit.
7. Capacités multi variables.

V.9. STRUCTURES DE RESEAUX DE NEURONES ARTIFICIELS POUR LA PREDICTION

La structure d'un prédicteur sous-entend sa capacité à représenter les propriétés dynamiques d'un signal d'entrée à temps discret statistiquement non stationnaire et donc sa capacité à prédire une valeur future (Mandic et Chambers, 2001). Un temps continu ou discret, est un aspect important du processus d'apprentissage du réseau de neurones (Elman, 1990).

Les prédictions de séries chronologiques les plus élémentaires sont faites en utilisant des décalages à partir d'une seule série chronologique de toutes les variables (Masters 1993).

Les perceptrons multicouches (PMC) ont d'abord été introduits pour résoudre des problèmes complexes de classification. Mais en raison de leur propriété d'approximateur universel (Hornik, Stinchcombe et White, 1990 ; Funahashi, Bruce et Rakic, 1989), ils ont été rapidement utilisés comme modèles de régression non linéaire, et ensuite pour la modélisation des séries temporelles et la prévision. (Gershenfeld et Weigend, 1993).

Cependant, l'estimation et l'identification de ces modèles utilisent des techniques sophistiquées et il n'est pas facile de déterminer l'architecture adéquate. En effet, ces modèles sont par définition sur-paramétrés, les fonctions d'erreur à minimiser ont de nombreux minima locaux, et la mise en œuvre s'avère souvent délicate.

De nombreux articles portent sur les techniques d'élagage des paramètres inutiles, en particulier dans le cadre des modèles de régression, appliquées aux séries temporelles (Moody et Daken, 1989 ; Reed et Marks, 1999).

La prédiction des séries temporelles à traditionnellement été réalisée par l'utilisation des modèles autorégressif (AR), des modèles autorégressif à moyenne mobile (ARMA, modèle de Box-Jenkins), ou des modèle à moyenne mobile (MA). Un réseau de neurones artificiels constitué de perceptrons multicouches (PMC) est un réseau statique, car elle n'a pas de lien de retour physique. Pour le transformer en un réseau dynamique, une temporisation (mémoire) dans les données d'entrée est introduite (Elman, 1990). La structure d'un RNA pour la prédiction est illustrée à la figure 41

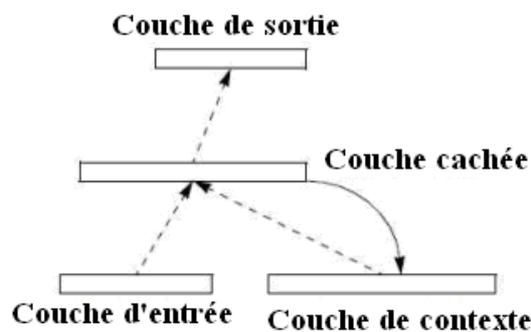


Figure 41 : Réseau de neurones artificiels d'Elman

Elman (1990) présente une nouvelle approche de la manière suivante. Supposons un réseau est augmentée au niveau d'entrée par des unités supplémentaires; appeler "unités de contexte". Ces unités sont également "caché" dans le sens où ils interagissent exclusivement avec les autres nœuds internes au réseau, et non pas à l'extérieur. Imaginer qu'il existe une entrée séquentielle à traiter, et une horloge qui règle la présentation de l'entrée sur le réseau. Le traitement consisterait alors à la séquence d'événements suivante. Au moment t , les unités d'entrée reçoivent la première entrée dans la séquence. Chaque entrée peut être une valeur scalaire unique ou un vecteur, selon la nature du problème. Les unités de contexte sont initialement configurées pour 0,5.

La structure du réseau de neurones artificiels pour la prédiction est illustrée sur la figure 42, où la variable à prédire est temporisée par un nombre fini d'étapes et utilisé à l'entrée. Il n'y a pas de chemins physiques de rétroaction de la sortie vers l'entrée dans cette structure, mais il a été appliqué avec succès dans de nombreuses applications telles que la prédiction de séries temporelles (Anava, Hazan et Mannor, 2013), la prévision des inondations (Chang, Liang et

Chen, 2001), prévisions à court terme de charge (Khan et Ondrůšek, 2000), et la prévision des séries chronologiques (Malasri et Malasri, 2002).



Figure 42 : Réseau de neurones artificiels pour des applications de prédiction avec des entrées temporisées

Le réseau de neurones artificiels est entraîné pour prédire $y(t+1)$ à partir de l'un ou plus connu précédemment observations historiques (par exemple, $y(t-n)$, $y(t-2)$, $y(t-1)$, $y(t-1)$, et $y(t)$ ou t est le pas de temps).

V.10 REVUE DE LITTÉRATURE

Un fonctionnement plus sûr et un contrôle adéquat d'une station d'épuration peuvent être atteints par le développement d'un outil de modélisation pour prédire la performance de la station sur la base des observations passées de certains paramètres clés des eaux usées. Une station de traitement des eaux usées implique plusieurs processus complexes, biologique et chimique.

Souvent, ces procédés présentent des comportements non linéaires qui sont difficiles à décrire par des modèles mathématiques linéaires (Plazl et al., 1999) . En outre, la variabilité des caractéristiques influentes, en termes de taux de composition, de débit, pourrait influencer les paramètres du modèle, et du contrôle opérationnel de façon significative (Hamoda et al., 1999). Par conséquent, la modélisation d'une station d'épuration est une tâche difficile et la plupart des modèles disponibles ne sont que des approximations basées sur des hypothèses (Lee et Park, 1999 ; Cote et al , 1995. ; Hamoda et al , 1999. ; Plazl et al, 1999 ; Bhat et al. , 1990).

Les techniques employant les réseaux de neurones artificiels peuvent être utilisées pour modéliser ces processus de stations d'épuration. Il peut être utilisé pour une meilleure prédiction de la performance du processus en raison de leur grande précision, d'adéquation et d'applications très prometteuses en génie (Govindaraju, 2000 ; Maier et Dandy, 2000 ; Neelakantan, Brion et Lingireddy, 2001). Elles s'appuient normalement sur des données historiques représentatives du processus. Dans une station d'épuration, il y a certaines variables explicatives clés qui peuvent être utilisées pour évaluer le rendement de la station.

Ces variables comprennent la demande biologique en oxygène (DBO), la demande chimique en oxygène (DCO), et matières en suspension (MES).

La plupart des études disponibles sur l'application des RNA dans les stations d'épuration ont pour but la modélisation de ces variables. On constate que les modèles à base de RNA constituent un outil efficace et robuste dans la prédiction de la performance d'une station d'épuration.

Häck et Köhne (1996) dans leurs travaux pour la prédiction permettant le calcul approximatif ou l'estimation des paramètres du processus de traitement qui sont temporairement non disponible en cas de la rupture d'un analyseur en ligne dans la station de traitement des eaux usées municipales de Siegen (Allemagne). Les résultats de la prédiction présente des coefficients de corrélation de 0,92 pour la prévision de la DCO et de 0,82 pour la prévision de l'ammonium (NH_4^+).

Zhu et al. (1998) ont proposé un réseau de neurones retard temporisé (TDNN) pour prédire l'efficacité d'un procédé de traitement biologique en termes de mesures de DBO. Les résultats de la simulation des auteurs à partir des données de processus réels ont montré que la performance du modèle réseau de neurones à retard temporisé (coefficient de corrélation) est améliorée à 0,845 contre 0,80 pour un réseau PMC initiale.

Belanche et al. (1999) proposent deux modèle RNA à retard temporisé pour la prédiction de la DBO et la DCO de la station de traitement des eaux usées de la Costa Brava (Espagne), Les résultats de prédiction ont trouvé un indice de corrélation de 0,504 pour la prédiction de la demande chimique en oxygène (DCO). Ils mentionnent qu'un modèle hybride RNA-Logique flou améliore considérablement la précision de la prédiction.

Choi et Park (2001) proposent un modèle RNA associé à une analyse en composantes principales (ACP) pour extraire les données d'une station de traitement des eaux usées industrielles. Sur les onze paramètres de qualités des eaux usées industrielles sont réduits à seulement cinq principaux composants. Ces composants principaux sont devenus des entrées du modèle RNA.

Les dispositifs d'extractions sont alors employés pour prévoir l'azote total kjeldahl (NTK) des eaux usées industrielles à l'entrée de la station on utilisant un RNA. Le système hybride montre un perfectionnement en matière de prédiction et réduit le problème du sur-apprentissage.

Oliveira-Esquerre et al. (2002) ont obtenu des prédictions satisfaisantes de la concentration de la DBO à la sortie d'une station de traitement des eaux usées industrielles au Brésil. Les résultats de la simulation des auteurs à partir des données de processus réels ont montré que la performance du modèle RNA associé à une analyse en composantes principales (coefficient de corrélation) c'est améliorée à 0,80 contre 0,74 pour un modèle RNA initiale.

Chen et al. (2003) développent un modèle réseau de neurones artificiels pour prédire la concentration de l'azote dans les effluents traités qui pourront être réutilisé pour la recharge des nappes souterraines. L'exactitude de ce modèle était supérieure à 90%. Par contre dans ce travail, aucune mention n'est faite sur la qualité des données.

Hamed et al. (2004) appliquent un modèle de réseau de neurones artificiels pour la prédiction des performances d'une station de traitement des eaux usées résiduaires avec un procédé de traitement conventionnelle importante dans la grande zone du Caire (Egypte). Les résultats obtenus à partir de cette étude ont indiqué que les valeurs R^2 varie de 0,63 à 0,81 pour la DBO₅ et de 0,45 à 0,65 pour les MES. Les auteurs ont conclu que le modèle RNA a été gêné par la limitation des données, les données corrompu, et la restriction du modèle basée uniquement sur deux paramètres.

Mjalli et al. (2007) ont employé un modèle RNA pour prévoir la qualité des effluents traités et rejetés (DBO₅, DCO et MES) à la sortie de la station de traitement des eaux résiduaires de Doha, (Qatar). Ils ont employé les données brutes des paramètres de pollutions qui caractérisaient la qualité d'eaux usées à savoir DBO₅, DCO et les MES.

Les auteurs ont employé les données sur un an qui ont été prélevées tous les 5 jours. Ils ont obtenu un coefficient de corrélation supérieur à 0,7, mais ils n'ont pas distingué si ces résultats étaient obtenus de l'apprentissage, de la validation ou du test de l'ensemble des données.

Vyas et al. (2011) appliquent deux modèles de réseau de neurones artificiels pour la prédiction des performances d'une station de traitement des eaux usées industrielles dans la zone industrielle de Bhopal (Inde). La prédiction donne des résultats très satisfaisants pour les deux modèles. Pour le premier modèle, la valeur de R est de 0,90 ce qui montre une bonne corrélation entre DBO_{eq} réelles et prévues avec une précision de 90%. De même le deuxième montre de meilleurs résultats, la valeur de R est de 0,73 et une précision de 88%.

Nasr et al., (2012) proposent un modèle RNA pour prédire le rendement, construire une base de données pour effectuer un contrôle du fonctionnement du processus de traitement de la station d'épurations des eaux usées d'El-Agamy (Egypt). L'étude indique que le model RNA

peut prédire le rendement de la station entre les variables de sortie observées et prédites avec un coefficient de corrélation (R) jusqu'à 0,90. En outre, le modèle RNA fournit une analyse efficace et peut être considéré comme un outil de diagnostic pour comprendre et simuler le comportement non linéaire de la station.

V.11. CONCLUSION

Ce chapitre fournit une brève introduction théorique aux réseaux neuronaux et commence par examiner les différences et les similitudes entre le neurone biologique et artificiel.

Les réseaux de neurones artificiels possèdent une propriété remarquable qui est à l'origine de leur intérêt pratique dans des domaines très divers.

Les RNA sont des approximateurs universels (Funahashi, Bruce et Rakic, 1989 ; Hornik, Stinchcombe et White, 1990 ; Haykin, 1999). La propriété d'approximation peut être énoncée de la manière suivante : toute fonction bornée suffisamment régulière peut être approchée avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire. Cette propriété de parcimonie est précieuse dans les applications industrielles.

Les réseaux de neurones artificiels (RNA) sont des structures mathématiques non linéaires qui sont capables de représenter le rapport fonctionnel non linéaire arbitraire et complexe entre les données d'entrée et les données de sortie de n'importe quel système. Les RNA ont été employés avec succès pour modéliser les rapports non linéaires complexes de séries chronologiques d'entrée-sortie dans une large variété de domaines.

CHAPITRE VI : PREDICTION DU TAUX DE FIABILITE PAR RESEAUX DE NEURONES ARTIFICIELS (Cas d'étude : station d'épuration de Khenchela)

VI.1 INTRODUCTION

Les réseaux de neurones offrent la possibilité de modéliser des données arbitraires d'entrée au moyen de l'ajustement des connexions internes au réseau, de sorte que pour une entrée donnée de la différence entre la sortie du réseau et la réponse souhaitée, l'erreur est minimisée. C'est le processus de l'entraînement du réseau au moyen d'un apprentissage supervisé, de sorte que l'erreur du réseau est réduit de manière itérative sur un ensemble de vecteurs correspondant à des entrées-sorties d'entraînement. Pourvu que l'ensemble des données d'entraînement soit représentatif du processus, le modèle RNA résultant saisit la relation inhérente entre l'entrée et la sortie et la possibilité de généraliser pour les entrées invisibles futures (Garvey, 1997).

L'objectif principal de cette recherche est de trouver des modèles RNA pouvant prédire les taux de fiabilité des principaux paramètres de contrôles des eaux usées traités dans une station d'épuration (DBO₅, DCO et les MES).

L'un des défis est de trouver une topologie de réseau optimale en faisant varier les types d'entrées du réseau et ainsi de résoudre le problème particulier de la cartographie d'entrée-sortie. L'architecture du réseau approprié pour une application donnée est fortement dépendante du problème, de la diversité de l'ensemble d'apprentissage et de la complexité de la fonction sous-jacente (Garvey, 1997).

VI.2 PRESENTATION DES CARACTERISTIQUES DU LOGICIEL

VI.2.1 Méthode de construction des modèles RNA

Le logiciel Predict est basé sur une approche constructive dans la création du modèle RNA appelée cascade corrélation. Développés par Fahlman et Lebiere (1989) cette technique est basée sur deux parties.

La partie "Cascade" qui se réfère à l'architecture et son mode de construction qui implique l'ajout de couches cachées (nœuds de réseau située entre l'entrée et la sortie), une à la fois, et toujours reliant toutes les unités précédentes de l'unité en cours. La deuxième partie "corrélacion" fait référence à la façon dont les unités cachées ont été formés en essayant de

maximiser la corrélation entre la sortie de la couche cachée et la sortie désirée du réseau de neurones artificiels à travers les données de formation.

Le logiciel Predict offre un choix de deux règles d'apprentissage. La première est la règle d'apprentissage du gradient adaptative. La deuxième règle d'apprentissage est la règle de Kalman qui est applicable à des problèmes de type de régression dans lequel le nombre d'entrées n'est pas trop grand.

Dans notre étude, le choix s'est porté sur la première règle d'apprentissage qui est largement applicable (Lecun et al., 1998 ; User guide, 2009).

VI.2.2 Sélections des variables d'entrée du réseau de neurones artificiels (RNA)

Le logiciel Predict utilise un algorithme génétique pour la sélection de variables d'entrée. L'algorithme génétique (AG) est un algorithme d'optimisation s'appuyant sur des techniques dérivées de la génétique et de l'évolution naturelle (croisements, mutations et sélection) (Dawkins, 1996). Un algorithme génétique recherche le ou les extrema d'une fonction définie sur un espace de données (Goldberg, 1989 ; Holland, 1992 ; Coley, 1999).

Le principe général du fonctionnement d'un algorithme génétique est représenté sur la figure 43: on commence par générer une population d'individus de façon aléatoire. Pour passer d'une génération k à la génération $k+1$, les trois opérations suivantes sont répétées pour tous les éléments de la population k . Des couples de parents $P1$ et $P2$ sont sélectionnées en fonction de leurs adaptations. L'opérateur de croisement leur est appliquée avec une probabilité Pc (généralement autour de 0,6) et génère des couples d'enfants $C1$ et $C2$. D'autres éléments P sont sélectionnés en fonction de leur adaptation. L'opérateur de mutation leur est appliquée avec la probabilité Pm (Pm est généralement très inférieur à Pc) et génère des individus mutées P' . Le niveau d'adaptation des enfants ($C1, C2$) et des individus mutées P' sont ensuite évaluées avant insertion dans la nouvelle population. Différents critères d'arrêt de l'algorithme peuvent être choisis (Alliot et Durand, 2005)

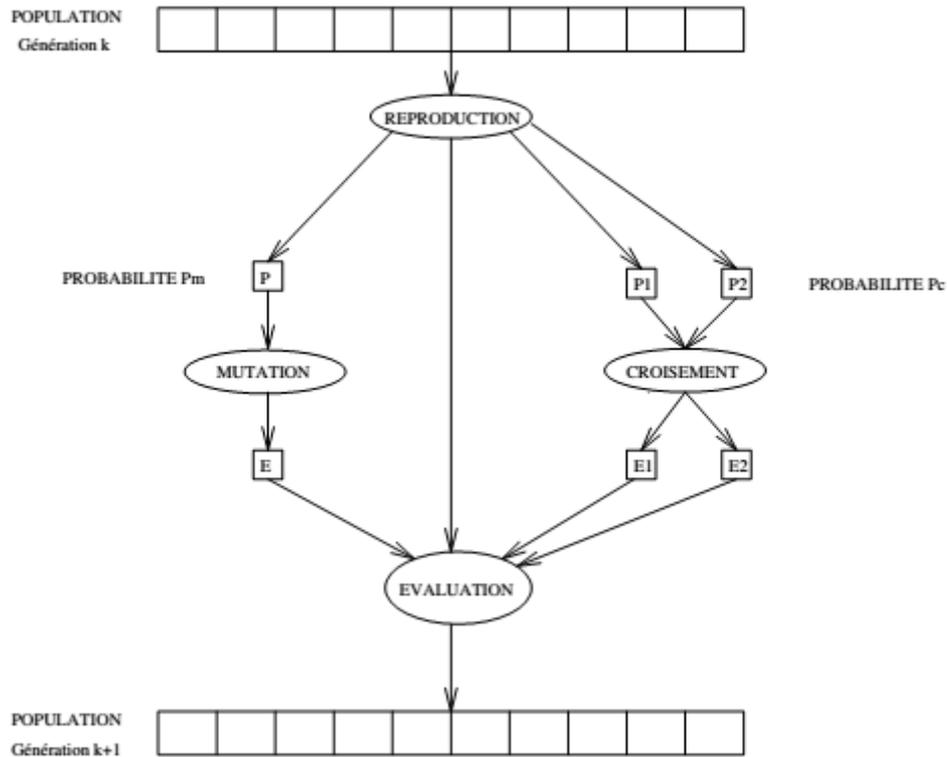


Figure 43: Principe général des algorithmes génétiques (Alliot et Durand, 2005)

Le logiciel Predict construit le modèle RNA progressivement, en utilisant la technique de cascades de corrélation. Les couches cachées sont ajoutés périodiquement, habituellement un ou deux à la fois. Chaque fois qu'une couche cachée ou une paire de couches cachées est ajouté, les poids sont formés de plusieurs valeurs d'initialisation différentes. Chaque initialisation est désignée comme un nouveau candidat. Le meilleur candidat est établi dans le réseau, puis tous les poids vers la couche de sortie sont recyclés.

Le nombre d'entrées du modèle RNA diffère souvent du nombre des données choisies au départ en raison du processus de transformation des données pour la construction du modèle de prédiction (User guide, 2009). Cette technique fera l'objet de futures études de recherches. Le choix s'est porté dans cette étude sur la nécessité de prendre en considérations tous les paramètres jugés pouvoir modéliser les performances du procédé de traitement.

Les modèles RNA proposés dans cette étude n'incluent pas les propositions fournis par l'algorithme génétique.

VI.2.3 Taille de la base des données

La taille de la base des données est déterminée par l'expérimentation. Une faible erreur quadratique moyenne (EQM) au cours de l'entraînement suggère que la construction du réseau est de bonne qualité. Cependant, ce n'est généralement pas le cas. Un cas de sur-apprentissage peut se produire. La capacité de généralisation du réseau doit être surveillée par rapport à la

croissance des données. Le nombre varie de 500 à 1000 données. Il y'a des résultats où les époques sont variés aussi bas que 100 et la performance du réseau est similaire à celle où le nombre d'époques est élevé (Kriger, 2007).

Dans le cas de notre étude, on a choisi d'utiliser l'ensemble des données des paramètres DCO DBO₅ et MES (à savoir 123 données), pour les modèles simples et les modèles complets des paramètres DBO₅ et DCO. Par contre pour les modèle complet du paramètres MES, l'ensemble des données utilisées dans cette études est constitué d'un premier groupe de 374 données (2009-2010), et le deuxième groupe de 410 données (2011-2012).

VI.2.4 Partage de l'ensemble des données

Il est conseillé de partager l'ensemble des données en deux sous-ensembles.

- Le premier sous ensemble servira lors de la construction du modèle RNA à apprentissage, il représentera 70% de l'ensemble des données.
- Le deuxième sous ensemble qui représente les 30% restante de l'ensemble des données aura pour mission d'effectuer le test de validation. Il faut être prudent toutefois d'assurer que l'ensemble de test n'inclut pas les données de l'ensemble de la formation. Si toutefois le modèle RNA est formé d'un ensemble de données particulier, encore testé avec un autre ensemble de données complètement différentes, le réseau n'est pas en mesure de généraliser (Haykin, 1999 ; Kriger, 2007).

VI.2.5 Fonction d'activation

Le choix de la fonction d'activation pour les couches cachées est nécessaire pour introduire la non-linéarité dans le réseau. Sans la non-linéarité, les couches cachées se comportent comme des perceptrons simples.

Cependant, la non-linéarité (c'est-à-la capacité de représenter des fonctions non linéaires) qui rend les réseaux multicouches puissant. Pour l'apprentissage de rétro-propagation, la fonction d'activation doit être différentiables, et ça aide si la fonction est limitée, les fonctions sigmoïdes telles que la logistique, la fonction tangente hyperbolique (tanh), et la fonction de Gauss sont les choix les plus courants. Pour les couches cachées, les fonctions d'activation sigmoïde sont généralement préférables (LeCun, 1998 ; Mjalli, 2007 ; User guide, 2009 ; faqs, 2012).

Lors de la conception des réseaux de neurones artificiels, la fonction de transfert initiale est appliquée à chacune des couches du réseau. Pour la plupart des réseaux. La fonction d'activation la plus couramment utilisée dans les modèle RNA à base Perceptron Multi couche (PMC) est la fonction d'activation sigmoïde et donne généralement de bons résultats (Joarder et aziz, 2002 ; Piekiewicz et Tybicki, 2004). La fonction d'activation sigmoïde est limitée dans l'intervalle [0 , 1], qui sont les valeurs utilisées pour indiquer l'appartenance à une classe

de sortie . L'entrée pour cette fonction varie de $-\infty$ à $+\infty$ mais généralement limitée à certaine valeur (Isa et al., 2010).

Dans cette étude notre choix c'est orienté pour la fonction d'activation *tanh* pour les neurones de la couche cachés, et pour la couche de sortie la fonction sigmoïde pour modéliser le comportement non linéaires du procédé de traitement.

VI.2.6 Critère de l'arrêt d'entraînement

Afin d'assurer que le modèle RNA est capable de généraliser, il est parfois nécessaire d'arrêter la formation (l'apprentissage) avant que le réseau converge vers un minimum global.

La phase d'apprentissage ou de formation d'un RNA peut être interrompue à trois conditions:

- Le nombre maximal d'époques a été atteint;
- L'erreur quadratique moyenne a convergé vers un minimum, ou
- L'algorithme d'apprentissage a atteint un minimum de pente.

VI.2.7 Mesure des performances de prédiction

Les performances de prédiction des RNA peuvent être mesurées en utilisant un certain nombre de facteurs différents. Il faut être prudent quant à ces critères pour décider de la mesure qui détermine la performance du RNA. Dans cette étude trois méthodes sont utilisées pour mesurer la performance prédictive du RNA.

1. Coefficient de corrélation linéaire (**R**) entre les valeurs cibles et les valeurs de sortie prédites correspondantes.

$$R = \frac{N \sum \hat{y} - (\sum y)(\sum \hat{y})}{\sqrt{(N \sum y^2 - (\sum y)^2) \times (N \sum \hat{y}^2 - (\sum \hat{y})^2)}}$$

N: Nombre de paires de valeurs,

\hat{y} : Valeur de sortie prédite par le modèle RNA,

y: Valeur cible correspondante (mesurée).

2. Coefficient de détermination (**R²**) entre les valeurs cibles et les valeurs de sortie correspondantes (valeurs de sortie prédites).

3. L'erreur moyenne absolue (**EMA**) entre les valeurs cibles et les valeurs de sortie correspondantes (valeurs de sortie prédites).

$$EMA = \frac{1}{N} \sum_{i=1}^N \left| \hat{y}_i - y_i \right|$$

N: Nombre de paires de valeurs,

\hat{y} : Valeur de sortie prédite par le modèle RNA,

y: Valeur cible correspondante (mesurée).

1. L'erreur quadratique moyenne (**EQM**) entre les valeurs de sortie prédites et les valeurs cibles correspondantes.

$$EQM = \frac{1}{N} \sum_{i=1}^N \left[\hat{y}_i - y_i \right]^2$$

N: Nombre de paires de valeurs,

\hat{y} : Valeur de sortie prédite par le modèle RNA,

y: Valeur cible correspondante (mesurée).

La performance est mesurée à la fois pendant la phase d'apprentissage (formation) et la phase de validation.

VI.3 STRATEGIE MISE EN OEUVRE POUR LA CONSTRUCTION DES MODELES RNA

La quête de trouver la bonne architecture implique la tâche fastidieuse d'analyser l'effet de varier certains paramètres lors de l'apprentissage du RNA, la généralisation, et finalement la capacité de prévision. Dans cette étude le choix s'est porté sur une seule architecture (topologie) à savoir le Perceptron multicouches (PMC). Cette section présente la stratégie mise en œuvre pour les deux groupes de modèles RNA (Modèle simple et Modèle complet). Chaque groupe est composé de trois modèles (RNA_{MES}, RNA_{DCO}, et RNA_{DBO5}).

VI.3.1 Modèle simple de réseau de neurones artificiels

L'approche choisie dans la construction de ces modèles simples est présenté dans le tableau 21. Les attentes des exploitants de la station d'épuration de Khenchela ont été prises en charge

par la construction de modèle simple à utiliser et basé sur des données quotidiennes et faciles à mesurer.

Les données d'auto-surveillances journalières serviront de données d'entrées aux modèles RNA pour prédire les différents taux de fiabilité du procédé de traitement.

Tableau 21: Données entrées-sortie des différents modèles simples RNA.

Modèle Simple RNA	Données d'entrée				Donnée de sortie
MS-RNA _{DBO5} <i>(123 données)</i>	Débit _{entrée}	DBO ₅ entrée	DCO _{entrée}	DBO ₅ sortie	Fiabilité-DBO ₅ prédite
MS-RNA _{DCO} <i>(123 données)</i>	Débit _{entrée}	DCO _{entrée}	DBO ₅ entrée	DCO _{sortie}	Fiabilité-DCO prédite
MS-RNA _{MES} <i>(123 données)</i>	Débit _{entrée}	MES _{entrée}	Débit _{sortie}	MES _{sortie}	Fiabilité-MES prédite

Dans la quête de construire des modèles RNA, chaque paramètre de qualité étudié est modéliser par une combinaison de paramètres d'entrées, chaque paramètre de qualité étudié est modéliser par la concentration à la sortie de la station, ainsi que les taux de fiabilité estimé par la méthode probabiliste de Niku et al. (1979) qui serviront de données cibles utilisées par le modèle RNA lors de la phase d'apprentissage et de validation.

VI.3.2 Modèle complet de réseau de neurones artificiels

Un approche original a été choisi pour la construction de ces modèles complets est présenté dans le tableau 22. L'intérêt principal de la construction de ces modèles est de modéliser l'effet de la variation des quantités et de la qualité des effluents à l'entrée de la station d'épuration de Khenchela, et de prendre en considération tous les paramètres qui influencent sur les performances du procédé de traitement.

De nombreux problèmes dans des compartiments spécifiques sont rencontrés dans la station d'épuration de Khenchela. Suite aux différents discussions avec les exploitants, un intérêt c'est dégagé pour l'étude et l'analyse des performances de l'élimination des MES dans la station. Concernant ce paramètre, les données sont devisés en deux parties vue le nombre

important (784 données), et due au fait que le nombre maximale des données acceptées par ce logiciel est limité à 512 données.

Tableau 22: Données entrées-sortie des différents modèles complets RNA.

Modèle Complet RNA	Données d'entrée				Donnée de sortie
MC-RNA _{DBO5} <i>(123 données)</i>	Débit _{entrée}	Débit _{sortie}	pH _{entrée}	pH _{sortie}	Fiabilité- DBO ₅ prédite
	DBO ₅ entrée	DBO ₅ sortie	Température _{entrée}	Température _{sortie}	
	MES _{entrée}	MES _{sortie}	Conductivité _{entrée}	Conductivité _{sortie}	
	DCO _{entrée}	DCO _{sortie}			
MC-RNA _{DCO} <i>(123 données)</i>	Débit _{entrée}	Débit _{sortie}	pH _{entrée}	pH _{sortie}	Fiabilité- DCO prédite
	DBO ₅ entrée	DBO ₅ sortie	Température _{entrée}	Température _{sortie}	
	MES _{entrée}	MES _{sortie}	Conductivité _{entrée}	Conductivité _{sortie}	
	DCO _{entrée}	DCO _{sortie}			
MC-RNA _{MES1} <i>(374 données)</i> (2009-2010)	Débit _{entrée}		Débit _{sortie}	Température _{entrée}	Fiabilité- MES ₁ prédite
	MES _{entrée}		MES _{sortie}	Température _{sortie}	
MC-RNA _{MES2} <i>(410 données)</i> (2011-2012)	Débit _{entrée}		Débit _{sortie}	Température _{entrée}	Fiabilité- MES ₂ prédite
	MES _{entrée}		MES _{sortie}	Température _{sortie}	

Tous les modèles de réseaux de neurones artificiels dans le présent travail sont mises en œuvre en utilisant la version gratuite (Limited version) du logiciel NeuralWorks Predict® (Vyas et al., 2011).

VI.4 PREDICTION DU TAUX DE FAIBILITE DES PARAMETRES DE QUALITE DES EAUX TRAITES PAR MODELE SIMPLE

VI.4.1 Prédiction de la fiabilité de la dégradation de la DBO5 par modèle simple

Les données d'entrée du modèle simple RNA_{DBO_5} sont le débit d'entrée, la concentration de la DBO₅ à l'entrée de la station, la concentration de la DCO à l'entrée de la station, et la concentration de la DBO₅ à la sortie de la station, présentés dans le tableau ci-dessous.

Tableau 23: Données d'entrées-sortie du modèle simple MS- RNA_{DBO_5} pour la prédiction du taux de fiabilité de la dégradation de la DBO₅.

Entrées	Sortie
Débit _{entrée} , DBO ₅ _{entrée} , DCO _{entrée} , DBO ₅ _{sortie}	Fiabilité-DBO ₅

Le modèle construit pour prédire le taux de fiabilité de la dégradation de la DBO₅ du procédé de traitement dans la station d'épuration de Khenchela est constitué de 4 neurones dans la couche d'entrées, 3 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

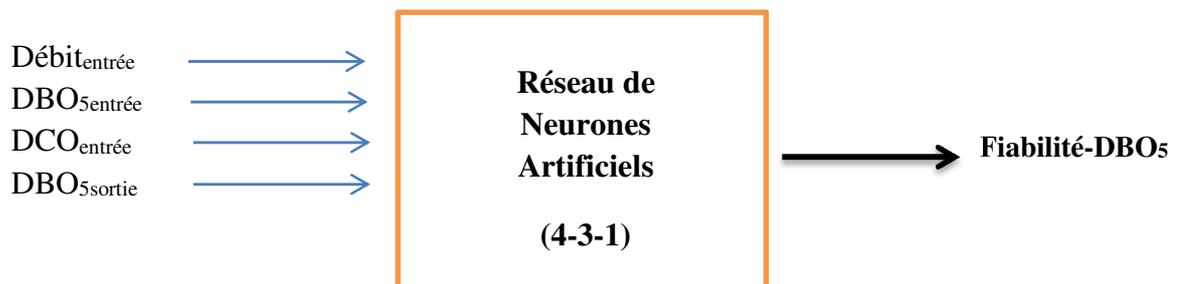


Figure 44: Schéma générale du modèle simple RNA_{DBO_5} (4-3-1) pour la prédiction du taux de fiabilité de la dégradation de la DBO₅.

Les performances de la prédiction sont présentées dans le tableau 24.

Table 24: Performances de la prédiction du taux de fiabilité de la dégradation de la DBO₅ par modèle simple MS-RNA_{DBO5}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,94	0,88	4,23	7,61
² Données d'apprentissage	0,94	0,88	4,17	7,52
³ Données de validation	0,93	0,86	4,37	7,83

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction de la fiabilité de la dégradation de la DBO₅ du procédé de traitement de la station de Khenchela est représentée sur la figure 45.

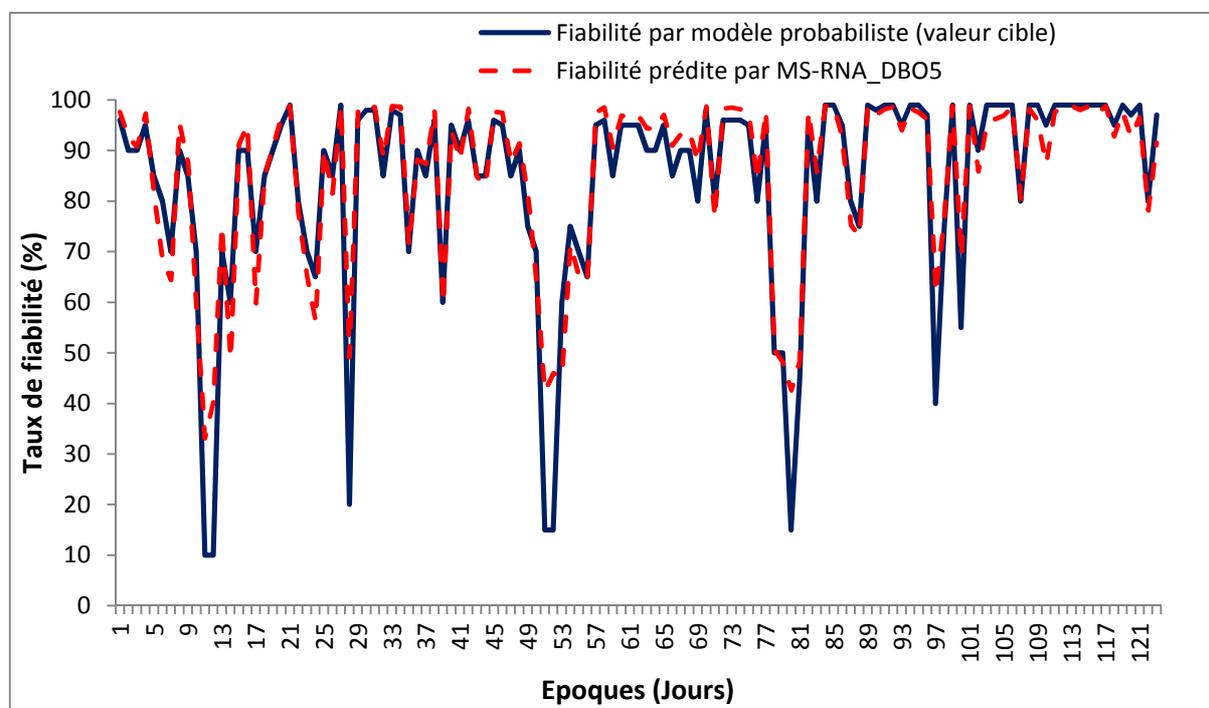


Figure 45: taux de fiabilité de la dégradation de la DBO₅ prédite par modèle simple MS-RNA_{DBO5} (4-3-1) versus fiabilité par modèle probabiliste (valeur cible).

L'application du modèle simple RNA_{DBO_5} pour la prédiction du taux de fiabilité de la dégradation de la DBO_5 dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 46.

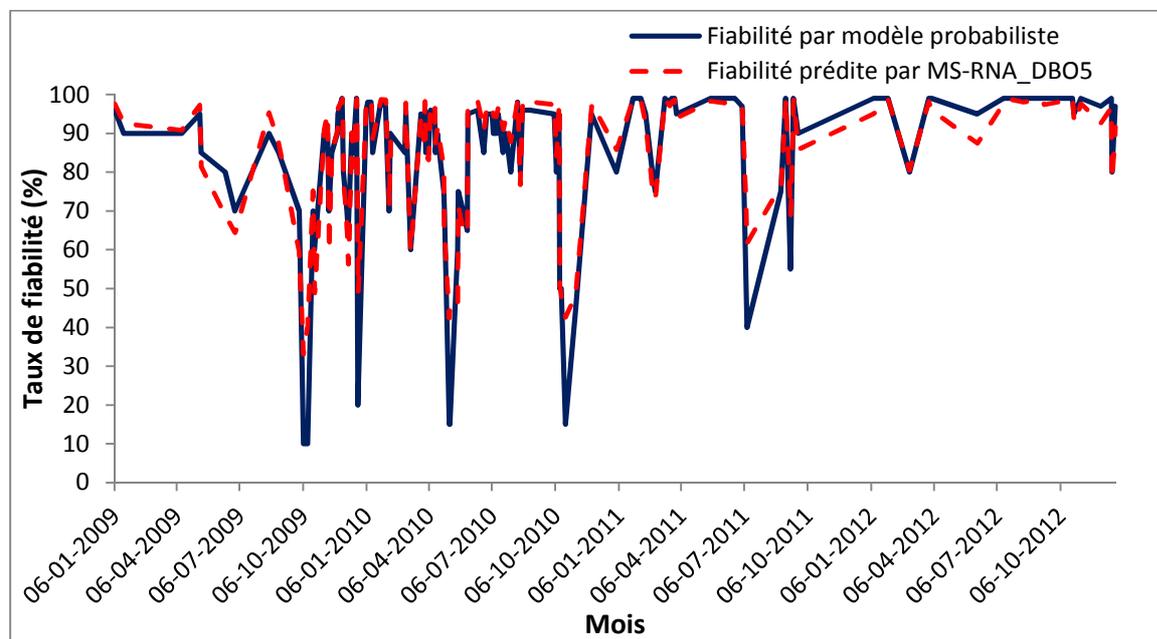


Figure 46: Taux de fiabilité de la dégradation de la DBO_5 prédit par réseau de neurones artificiels simple (4-3-1), station d'épuration de Khenchela (2009-2012).

VI.4.2 Prédiction de la fiabilité d'élimination de la DCO par modèle simple

Les données d'entrée du modèle simple RNA_{DCO} sont le débit d'entrée, la concentration de la DCO à l'entrée de la station, la concentration de la DBO_5 à l'entrée de la station, et la concentration de la DCO à la sortie de la station, présentés dans le tableau ci-dessous.

Tableau 25: Données d'entrées-sortie du modèle simple $MS-RNA_{DCO}$ pour la prédiction du taux de fiabilité d'élimination de la DCO.

Entrées	Sortie
Débit _{entrée} , DCO _{entrée} , DBO ₅ _{entrée} , DCO _{sortie}	Fiabilité-DCO

Le modèle construit pour prédire le taux de fiabilité de l'élimination de la DCO du procédé de traitement de la station d'épuration de Khenchela est constitué de 4 neurones dans la couche d'entrées, 2 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

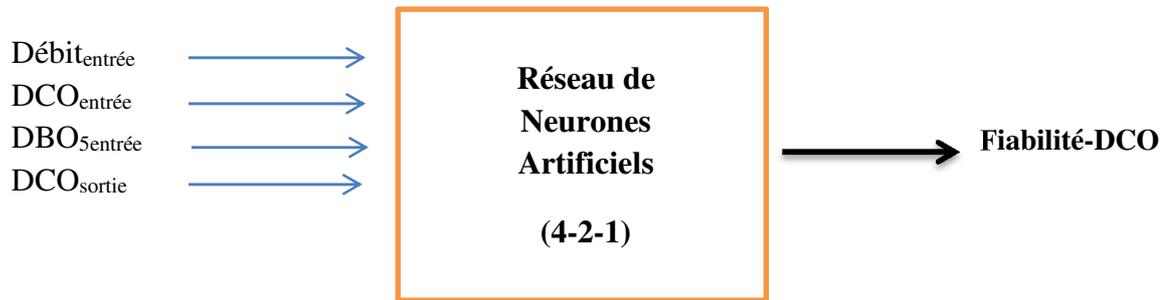


Figure 47: Schéma générale du modèle simple MS-RNA_{DCO} (4-2-1) pour la prédiction du taux de fiabilité d'élimination de la DCO.

Les performances de la prédiction sont présentées dans le tableau 26.

Tableau 26: Performances de la prédiction du taux de fiabilité d'élimination de la DCO par modèle simple MS-RNA_{DCO}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,94	0,88	1,72	4,20
² Données d'apprentissage	0,95	0,90	1,63	3,77
³ Données de validation	0,93	0,86	1,92	5,05

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction de la fiabilité de l'élimination de la DCO du procédé de traitement de la station de Khenchela est représentée sur la figure 48.

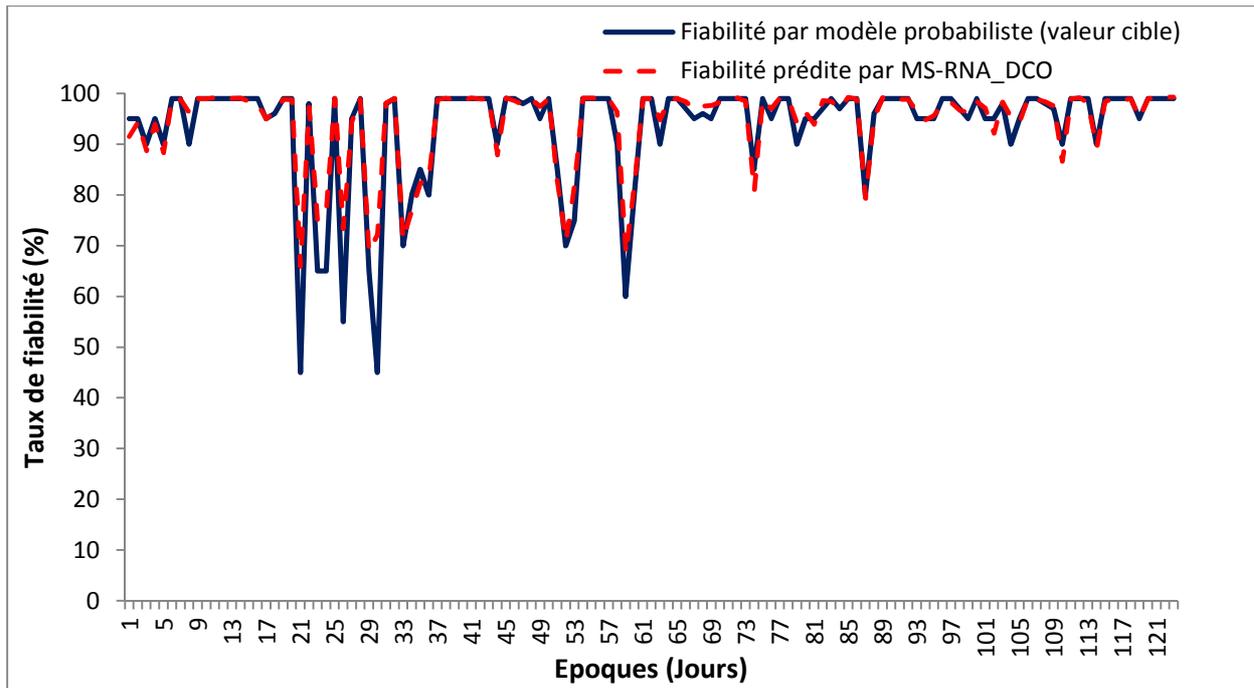


Figure 48: Taux de fiabilité d'élimination de la DCO prédit par modèle simple MS-RNA_{DCO} (4-2-1) versus fiabilité par modèle probabiliste (valeur cible).

L'application du modèle simple RNA_{DCO} pour la prédiction du taux de fiabilité de l'élimination de la DCO dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 49.

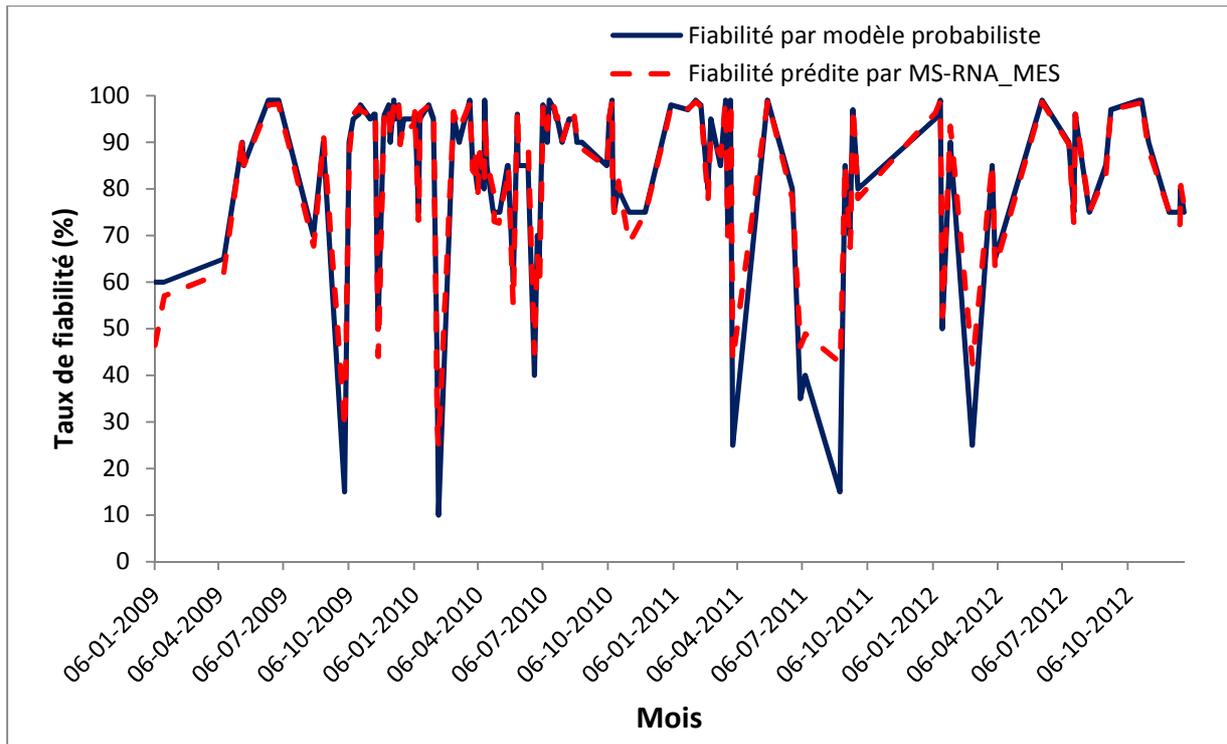


Figure 49: Taux de fiabilité de la dégradation de la DBO₅ prédit par réseau de neurones artificiels simple (4-2-1), station d'épuration de Khenchela (2009-2012)

VI.4.3 Prédiction de la fiabilité d'élimination des MES par modèle simple

Les données d'entrée du modèle simple RNA_{MES} sont : le débit d'entrée, le débit de sortie, la concentration des MES à l'entrée de la station, et la concentration des MES à la sortie de la station, présentés dans le tableau 27.

Tableau 27: Données d'entrées-sortie du modèle simple MS-RNA_{MES} pour la prédiction du taux de fiabilité d'élimination des MES.

Entrées	Sortie
Débit _{entrée} , Débit _{sortie} , MES _{entrée} , MES _{sortie}	Fiabilité-MES

Le modèle construit pour prédire le taux de fiabilité de l'élimination des MES de la station d'épuration de Khenchela est constitué de 4 neurones dans la couche d'entrées, 4 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

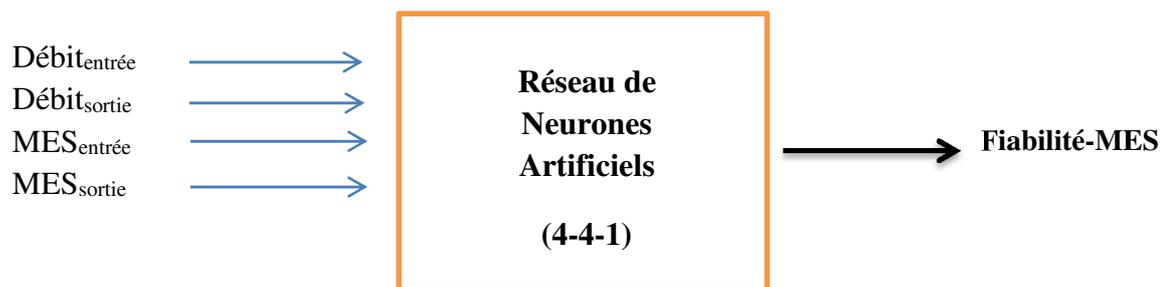


Figure 50: Schéma générale du modèle simple MS-RNA_{MES} (4-4-1) pour la prédiction du taux de fiabilité d'élimination des MES.

Les performances de la prédiction du taux de fiabilité d'élimination des MES sont présentées dans le tableau 28.

Tableau 28: Performances de la prédiction du taux de fiabilité d'élimination des MES par modèle simple MS-RNA_{MES}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,97	0,94	2,49	4,86
² Données d'apprentissage	0,97	0,94	2,50	4,74
³ Données de validation	0,96	0,92	2,47	5,12

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction de la fiabilité de l'élimination des MES du procédé de traitement de la station d'épuration de Khenchela est représentée sur la figure 51.

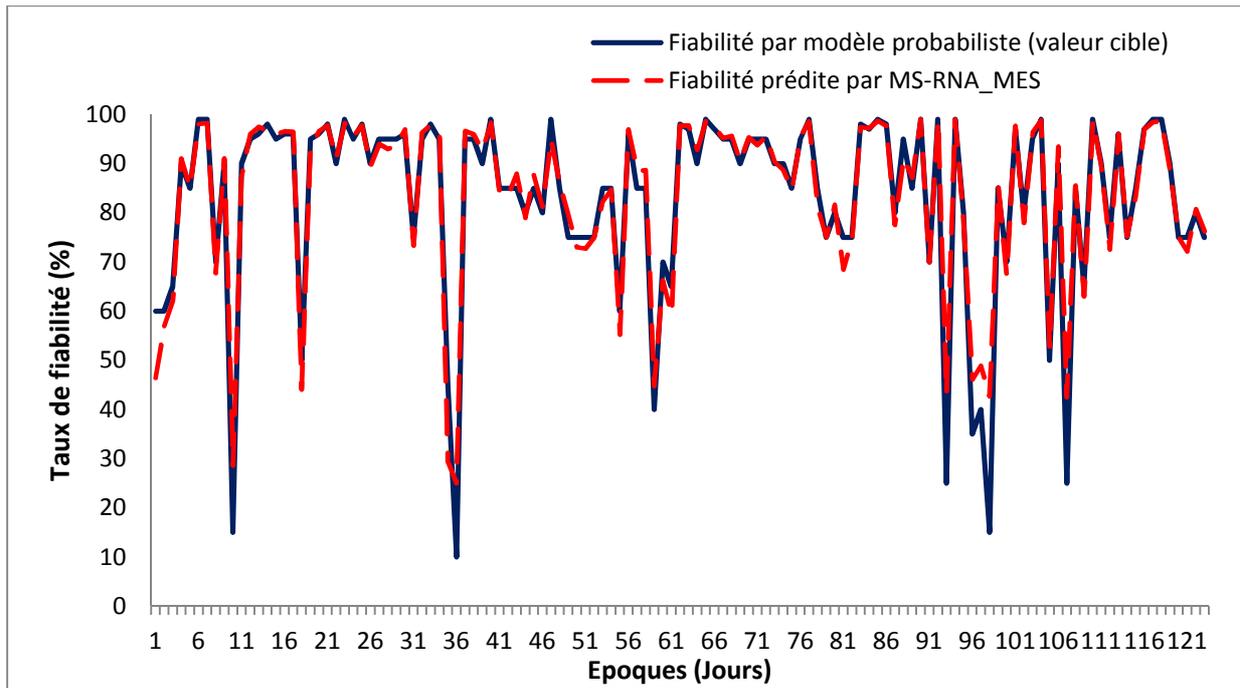


Figure 51: Taux de fiabilité d'élimination des MES prédit par modèle simple MS-RNA_{MES} (4-4-1) versus fiabilité par modèle probabiliste (valeur cible).

L'application du modèle simple RNA_{MES} pour la prédiction du taux de fiabilité de l'élimination des MES dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 52.

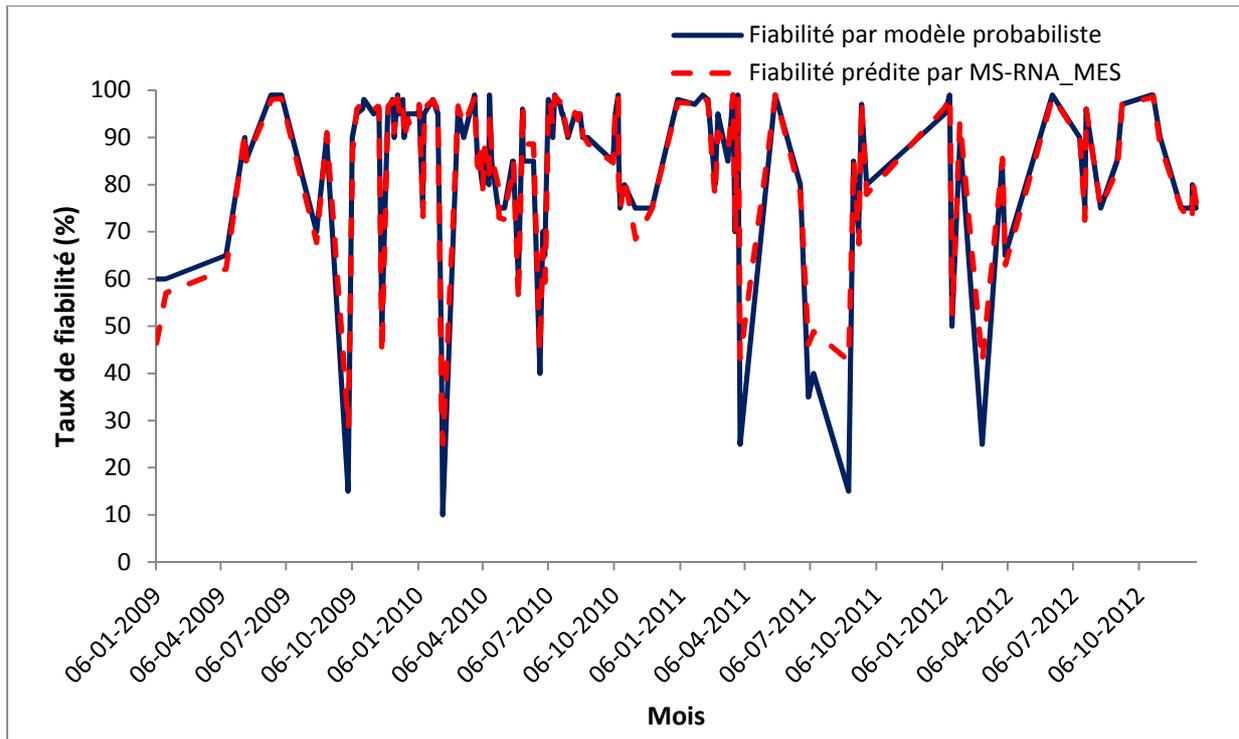


Figure 52: Taux de fiabilité d'élimination des MES prédit par réseau de neurones artificiels simple (4-4-1), station d'épuration de Khenchela (2009-2012)

VI.5 PREDICTION DU TAUX DE FAIBILITE DES PARAMETRES DE QUALITE DES EAUX TRAITEES PAR MODELE COMPLET

VI.5.1 Prédiction de la fiabilité de la dégradation de la DBO5 par modèle complet

Les données d'entrée du modèle complet RNA_{DBO5} sont tous les données disponible à l'entrée et à la sortie de la station d'épuration qui sont les suivant : débit d'entrée et de sortie, les paramètres de qualité des eaux usées brutes et traités (concentration des MES, DCO et DBO₅), température de l'eau à l'entrée et sortie de la station, la conductivité à l'entrée et sortie de la station, et le pH des eaux usées brutes, et traités.

Tableau 29: Données d'entrées-sortie du modèle complet MC- RNA_{DBO5} pour la prédiction du taux de fiabilité de la dégradation de la DBO₅.

Entrées	Sortie
Débit _{entrée} , DBO ₅ _{entrée} , MES _{entrée} , DCO _{entrée} Débit _{sortie} , DBO ₅ _{sortie} , MES _{sortie} , DCO _{sortie} pH _{entrée} , Température _{entrée} , Conductivité _{entrée} pH _{sortie} , Température _{sortie} , Conductivité _{sortie}	Fiabilité-DBO ₅

Le modèle complet construit pour prédire le taux de fiabilité de la dégradation de la DBO₅ de la station d'épuration de Khenchela est constitué de 14 neurones dans la couche d'entrée, 3 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

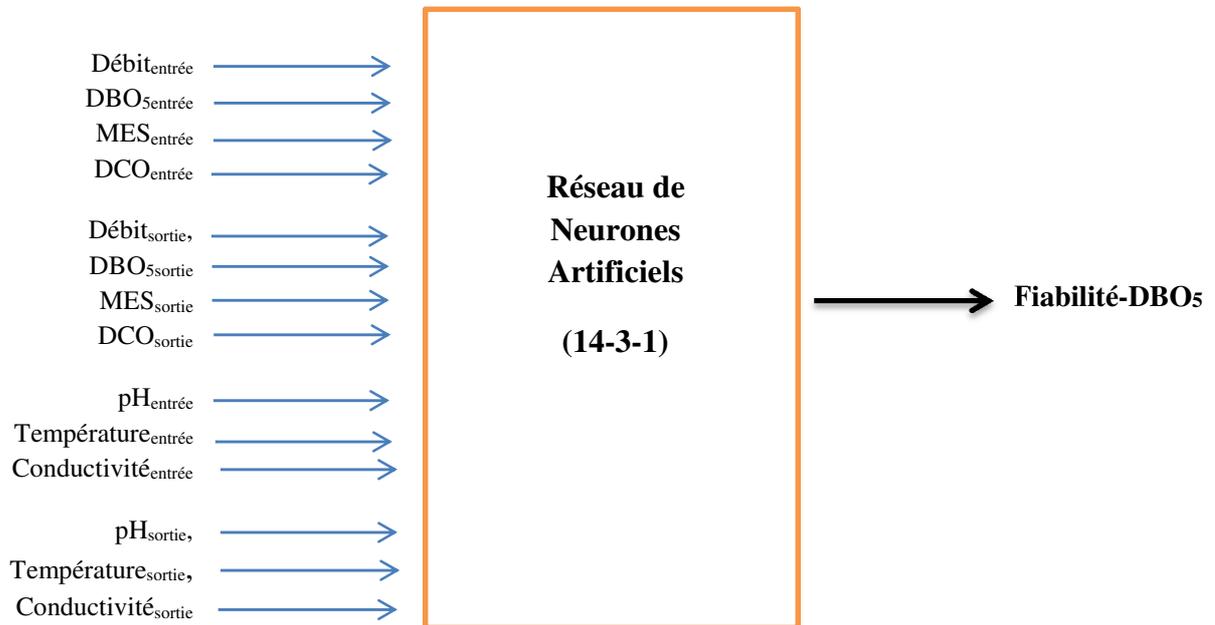


Figure 53 : Schéma générale du modèle complet MC-RNA_{DBO5} (14-3-1) pour la prédiction du taux de fiabilité de la dégradation de la DBO₅.

Les performances de la prédiction de la fiabilité de la dégradation de la DBO₅ par le modèle complet RNA_{DBO5} sont présentées dans le tableau ci-dessous.

Tableau 30: Performances de la prédiction du taux de fiabilité de la dégradation de la DBO₅ par le modèle complet MC-RNA_{DBO5}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,91	0,82	4,89	8,37
² Données d'apprentissage	0,93	0,86	4,61	7,42
³ Données de validation	0,87	0,75	5,54	10,24

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction de la fiabilité de la dégradation de la DBO₅ du procédé de traitement de la station de Khenchela est représentée sur la figure 54.

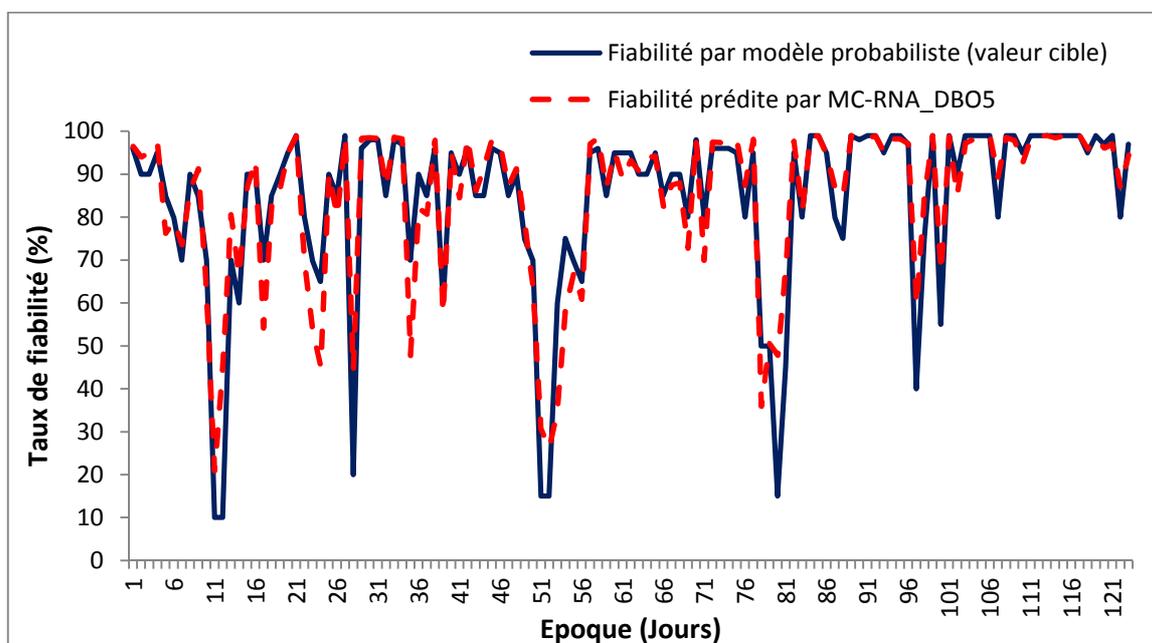


Figure 54: Fiabilité de la dégradation de la DBO₅ prédite par modèle complet MC-RNA_{DBO5} (14-3-1) versus fiabilité par modèle probabiliste (valeur cible).

L'application du modèle complet RNA_{DBO_5} pour la prédiction du taux de fiabilité de la dégradation de la DBO_5 dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 55.

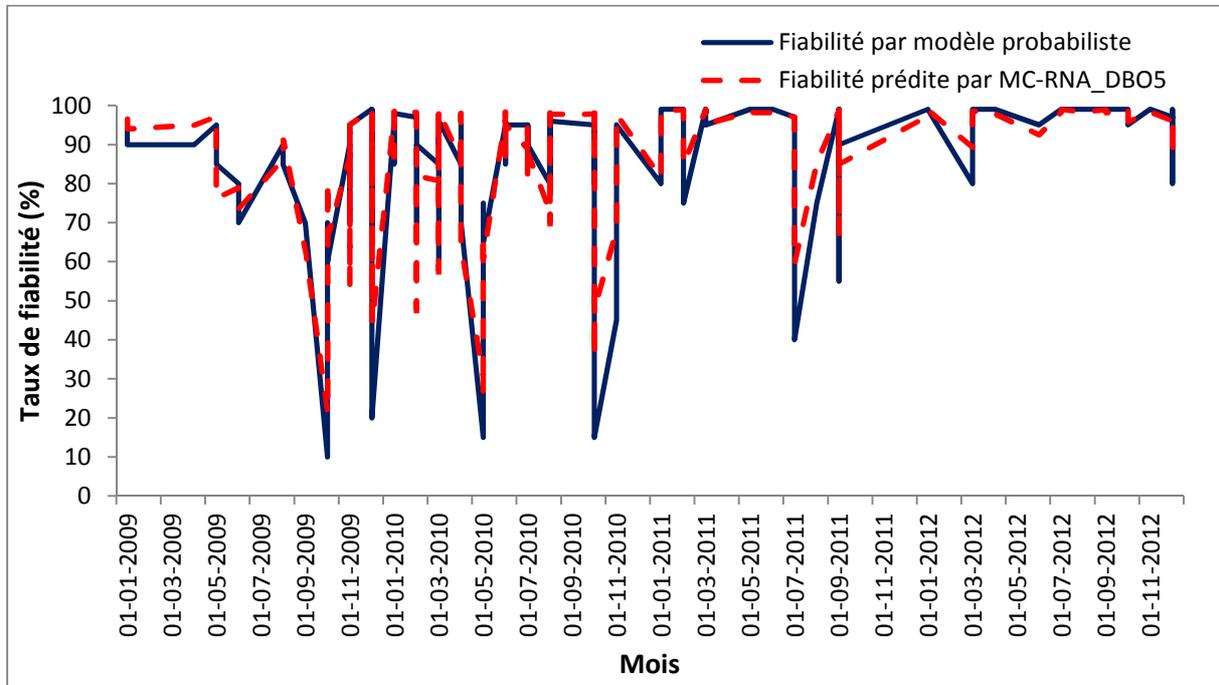


Figure 55: Taux de fiabilité de la dégradation de la DBO_5 prédit par réseau de neurones artificiels complet (14-3-1), station d'épuration de Khenchela (2009-2012)

VI.5.2 Prédiction de la fiabilité d'élimination de la DCO par modèle complet

Les données d'entrée du modèle complet RNA_{DCO} sont tous les données disponibles à l'entrée et à la sortie de la station qui font l'objet de mesures quotidiennes pour le contrôle de la qualité des eaux traitées et assurer un fonctionnement adéquat de la station.

Tableau 31: Données d'entrées-sortie du modèle complet $MC-RNA_{DCO}$ pour la prédiction du taux de fiabilité d'élimination de la DCO.

Entrées	Sortie
Débit _{entrée} , $DBO_{5\text{entrée}}$, $MES_{\text{entrée}}$, $DCO_{\text{entrée}}$ Débit _{sortie} , $DBO_{5\text{sortie}}$, MES_{sortie} , DCO_{sortie} $pH_{\text{entrée}}$, $Température_{\text{entrée}}$, $Conductivité_{\text{entrée}}$, pH_{sortie} , $Température_{\text{sortie}}$, $Conductivité_{\text{sortie}}$	Fiabilité-DCO

Le modèle construit pour prédire le taux de fiabilité de la dégradation de la DCO de la station d'épuration de Khenchela est constitué de 14 neurones dans la couche d'entrée, 8 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

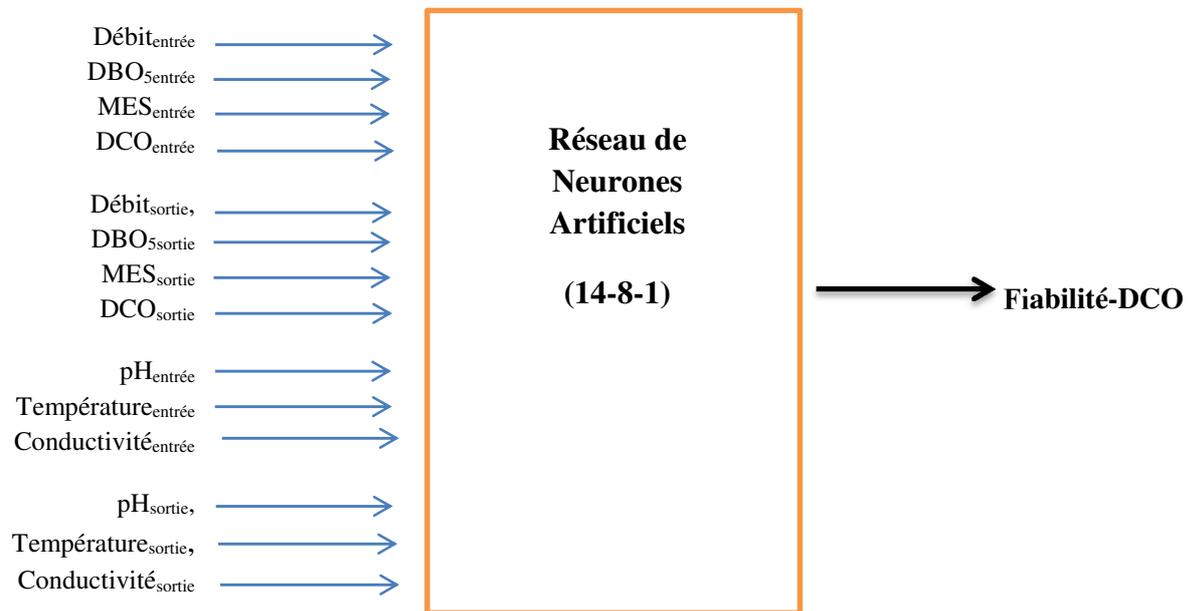


Figure 56: Schéma générale du modèle complet MC-RNA_{DCO} (14-8-1) pour la prédiction du taux de fiabilité d'élimination de la DCO.

Les performances de la prédiction de la fiabilité d'élimination de la DCO du procédé de traitement de la station de Khenchela par le modèle complet RNA_{DCO} sont présentées dans le tableau ci-dessous.

Tableau 32: Performances de la prédiction du taux de fiabilité d'élimination de la DCO la par le modèle complet MC-RNA_{DCO}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,81	0,65	3,09	7,79
² Données d'apprentissage	0,90	0,81	2,83	7,49
³ Données de validation	0,69	0,47	3,70	8,46

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction de la fiabilité de l'élimination de la DCO du procédé de traitement de la station d'épuration de Khenchela est représentée sur la figure 57.

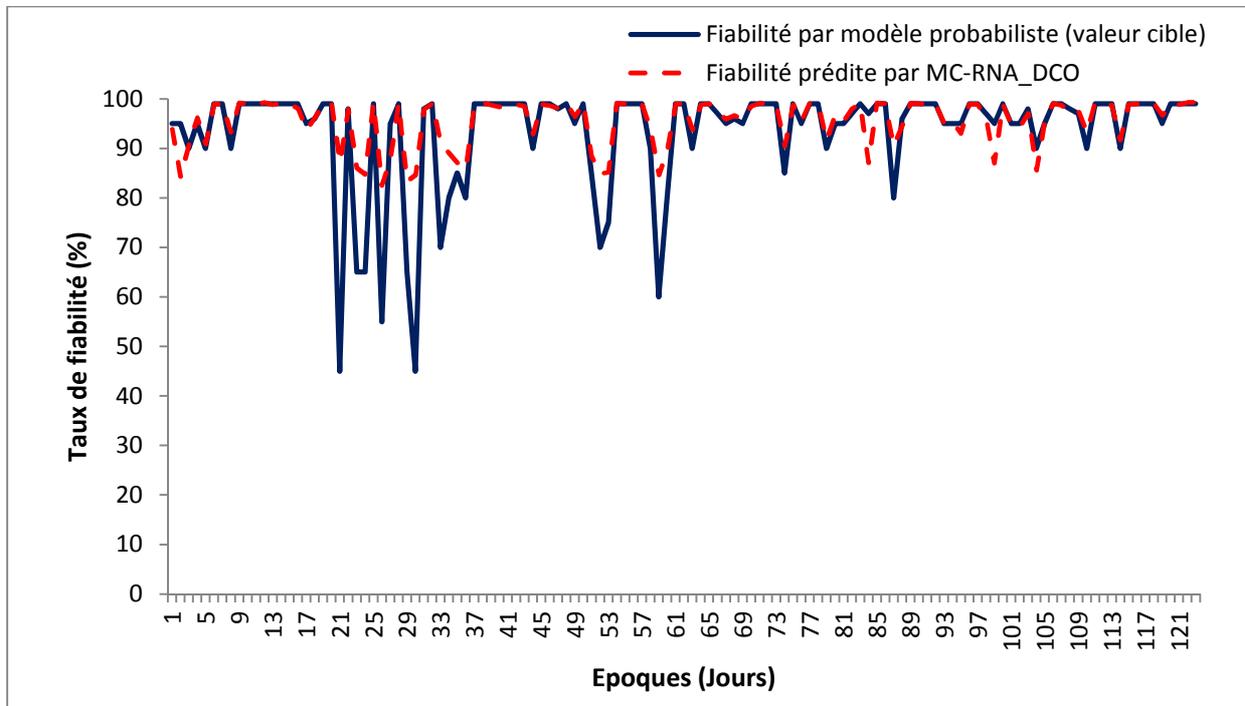


Figure 57: Taux de fiabilité d'élimination de la DCO prédit par modèle complet MC-RNA_{DCO} (14-8-1) versus fiabilité par modèle probabiliste (valeur cible).

L'application du modèle complet RNA_{DCO} pour la prédiction du taux de fiabilité de l'élimination de la DCO dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 58.

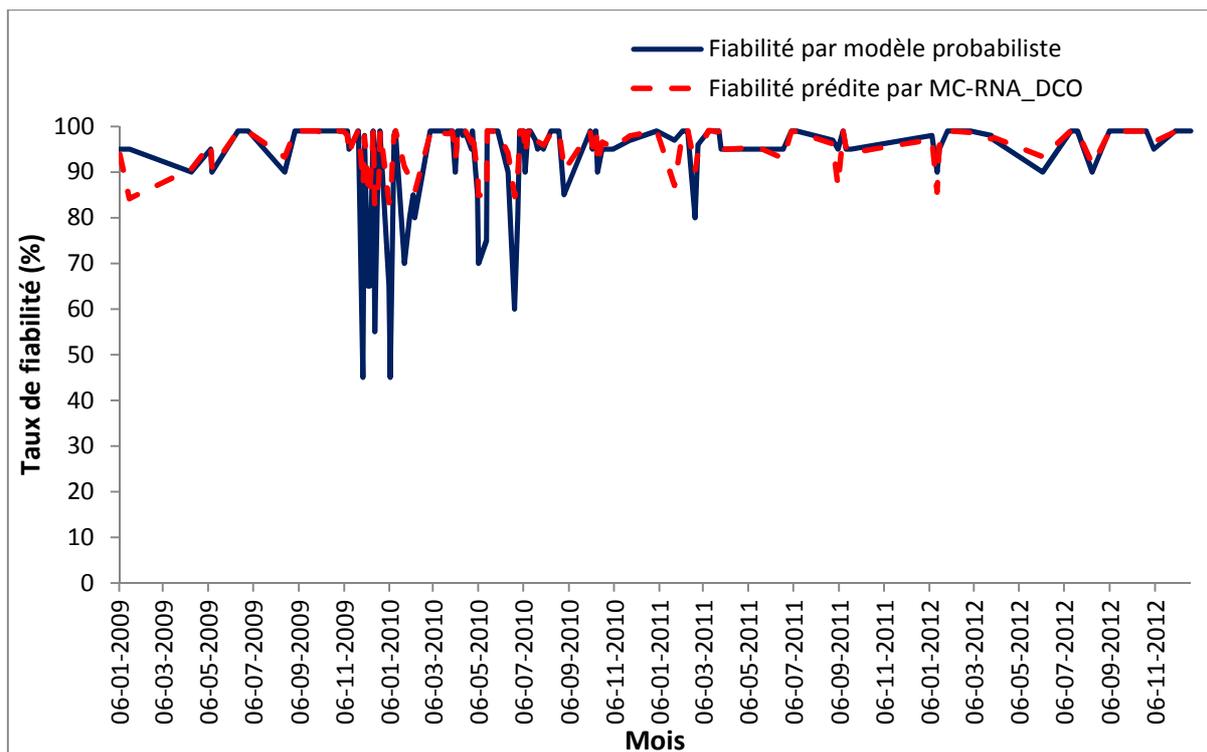


Figure 58: Taux de fiabilité d'élimination de la DCO prédit par réseau de neurones artificiels complet (14-8-1), station d'épuration de Khenchela (2009-2012)

VI.5.3 Prédiction de la fiabilité d'élimination des MES par modèle complet

Lors de cette étude, on a constaté que les données de mesure des MES sont très importante, et les limitations techniques du logiciel Predict nous on contrant à réaliser deux modèles biannuelles (2009-100 avec 436 données), (2011-2012 avec 348 données). Les données d'entrée des deux modèles complets RNA_{MES} sont citées dans le tableau 32.

Tableau 33: Données d'entrées-sortie des modèles complets MC-RNA_{MES1} et MC-RNA_{MES2} pour la prédiction du taux de fiabilité d'élimination des MES.

Entrées	Sortie
Débit _{entrée} , MES _{entrée} , Débit _{sortie} , MES _{sortie} , pH _{entrée} , Température _{entrée} , Conductivité _{entrée} pH _{sortie} , Température _{sortie} , Conductivité _{sortie}	Fiabilité-MES

- **Le premier modèle** construit pour prédire le taux de fiabilité de l'élimination des MES de la station d'épuration de Khenchela est constitué de 10 neurones dans la couche d'entrées, 3 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

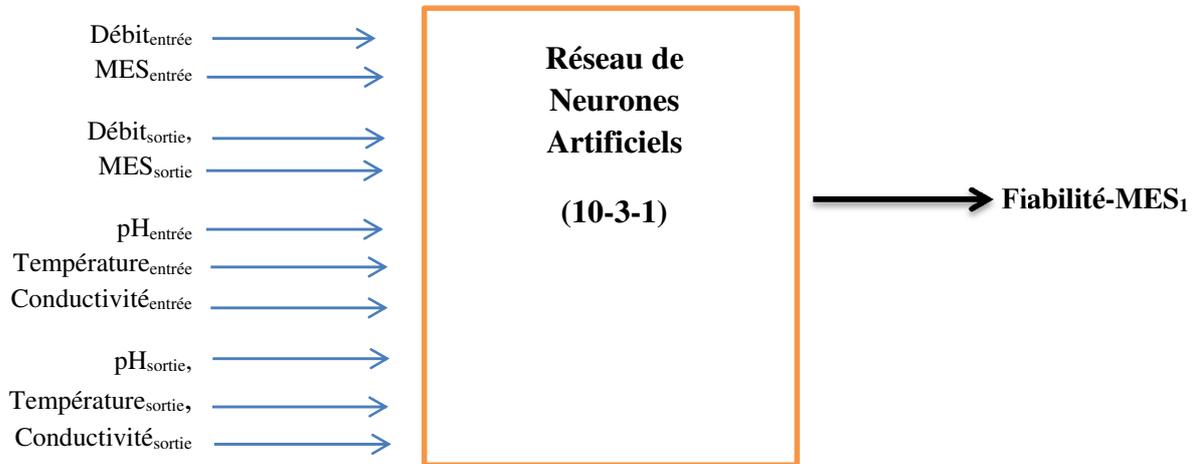


Figure 59: Schéma générale du modèle complet MC-RNA_{MES1} (10-3-1) pour la prédiction du taux de fiabilité d'élimination des MES.

Les performances de la prédiction du modèle RNA_{MES1} sont présentées dans le tableau 34.

Tableau 34: Performances de la prédiction du taux de fiabilité d'élimination des MES par modèle complet MC-RNA_{MES1}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,97	0,94	2,56	4,86
² Données d'apprentissage	0,97	0,94	2,53	4,74
³ Données de validation	0,97	0,94	2,64	5,12

¹ : Ensemble de données (436 données)

² : Données d'apprentissage (305 données)

³ : Données de validation (131 données)

La prédiction de la fiabilité d'élimination des MES du procédé de traitement de la station d'épuration de Khenchela est représentée sur la figure 60.

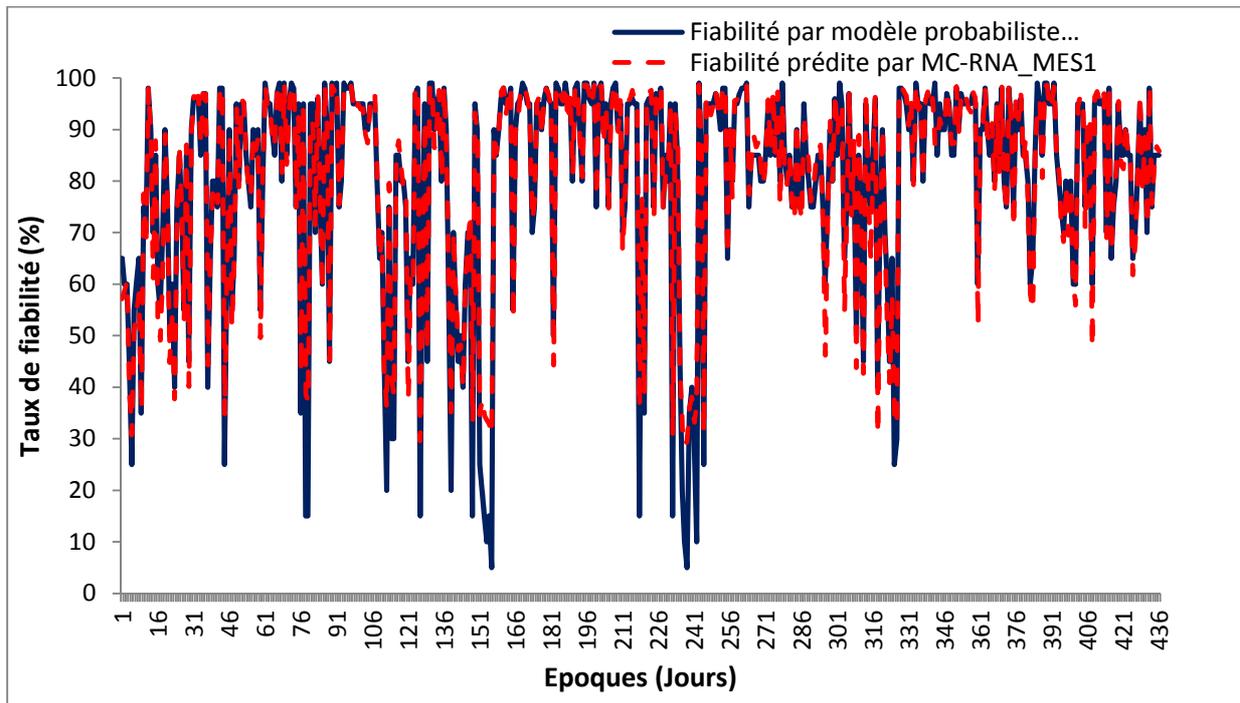


Figure 60: Taux de fiabilité d'élimination des MES prédit par modèle complet RNA_{MES1} (10-3-1) versus fiabilité par modèle probabiliste (valeur cible).

L'application du modèle complet RNA_{MES1} pour la prédiction du taux de fiabilité d'élimination des MES dans la station d'épuration de Khenchela durant une période d'étude de 2 ans (2009-2010) est présenté sur la figure 61.

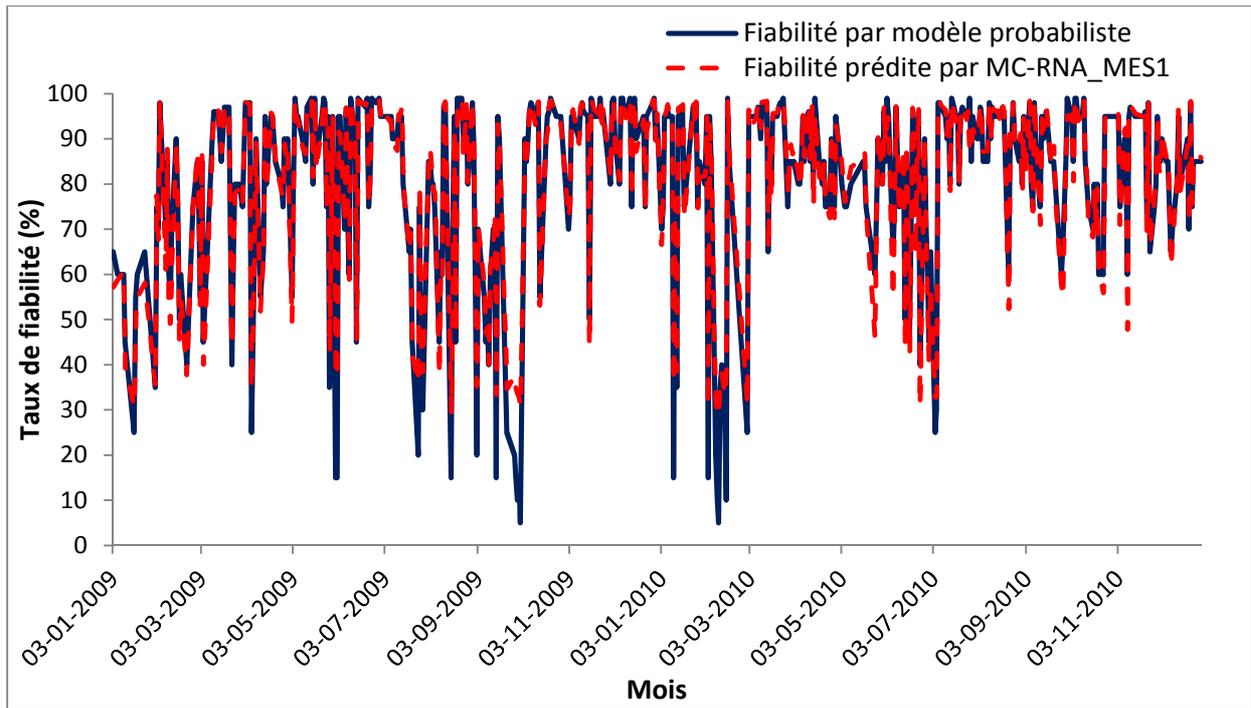


Figure 61: Taux de fiabilité d'élimination des MES prédit par réseau de neurones artificiels complet (10-3-1), station d'épuration de Khenchela (2009-2010).

- **Le deuxième modèle** construit pour prédire le taux de fiabilité d'élimination des MES de la station d'épuration de Khenchela est constitué de 10 neurones dans la couche d'entrées, 6 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

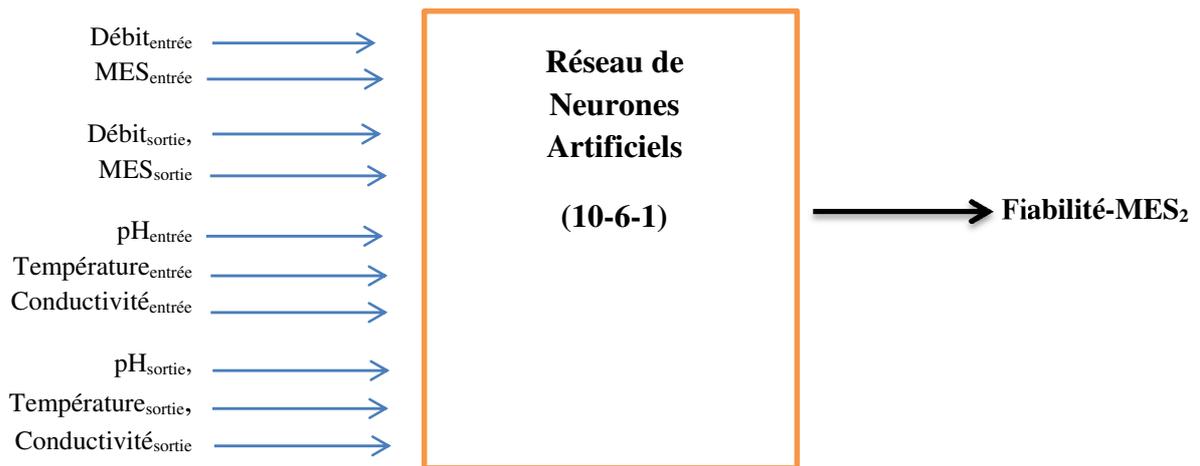


Figure 62: Schéma générale du modèle complet MC-RNA_{MES2} (10-6-1) pour la prédiction du taux de fiabilité d'élimination des MES.

Les performances de la prédiction du modèle RNA_{MES2} sont présentées dans le tableau 35.

Tableau 35: Performances de la prédiction du taux de fiabilité d'élimination des MES par modèle complet MC-RNA_{MES2}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,97	0,94	1,97	4,36
² Données d'apprentissage	0,97	0,94	1,958	4,17
³ Données de validation	0,97	0,94	2,01	4,77

¹ : Ensemble de données (348 données)

² : Données d'apprentissage (243 données)

³ : Données de validation (105 données)

La prédiction de la fiabilité d'élimination des MES du procédé de traitement de la station d'épuration de Khenchela par le modèle complet MC-RNA_{MES2} est représentée sur la figure 63.

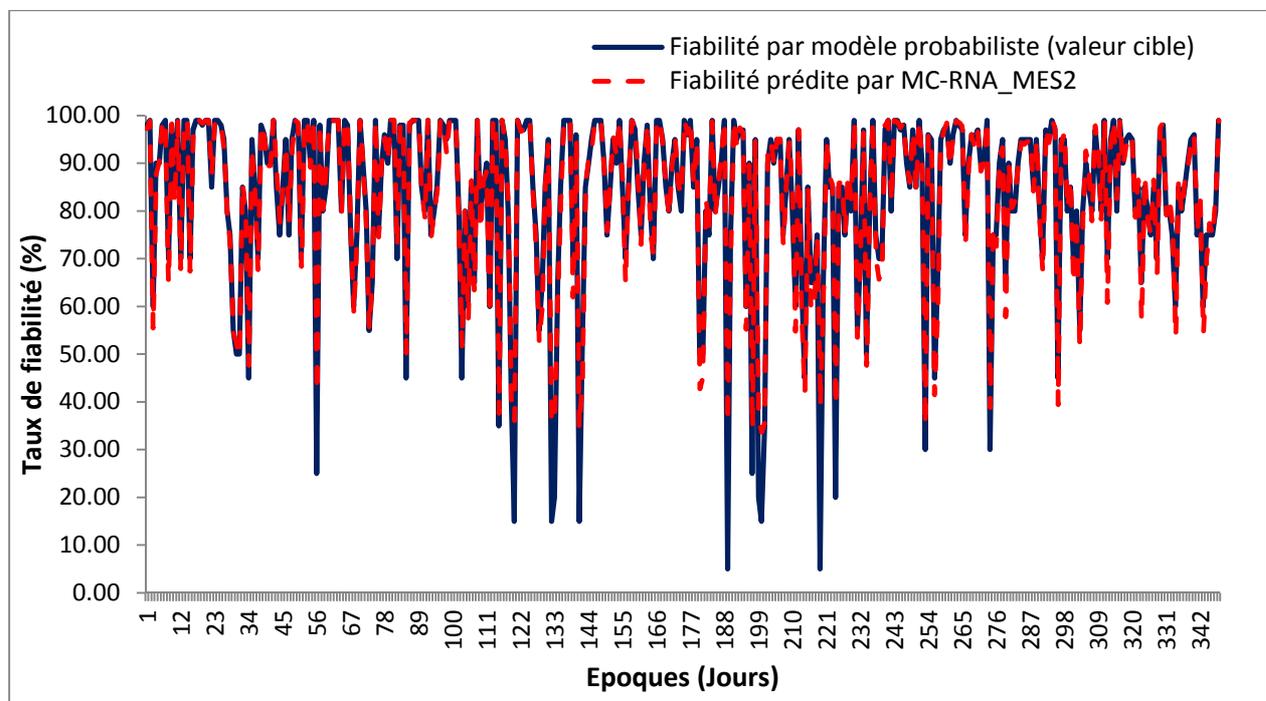


Figure 63: Taux de fiabilité d'élimination des MES prédit par modèle complet RNA_{MES2} (10-6-1) versus taux de fiabilité par modèle probabiliste (valeur cible).

L'application du modèle complet RNA_{MES2} pour la prédiction du taux de fiabilité d'élimination des MES dans la station d'épuration de Khenchela durant une période d'étude de 2 ans (2011-2012) est présenté sur la figure 64.

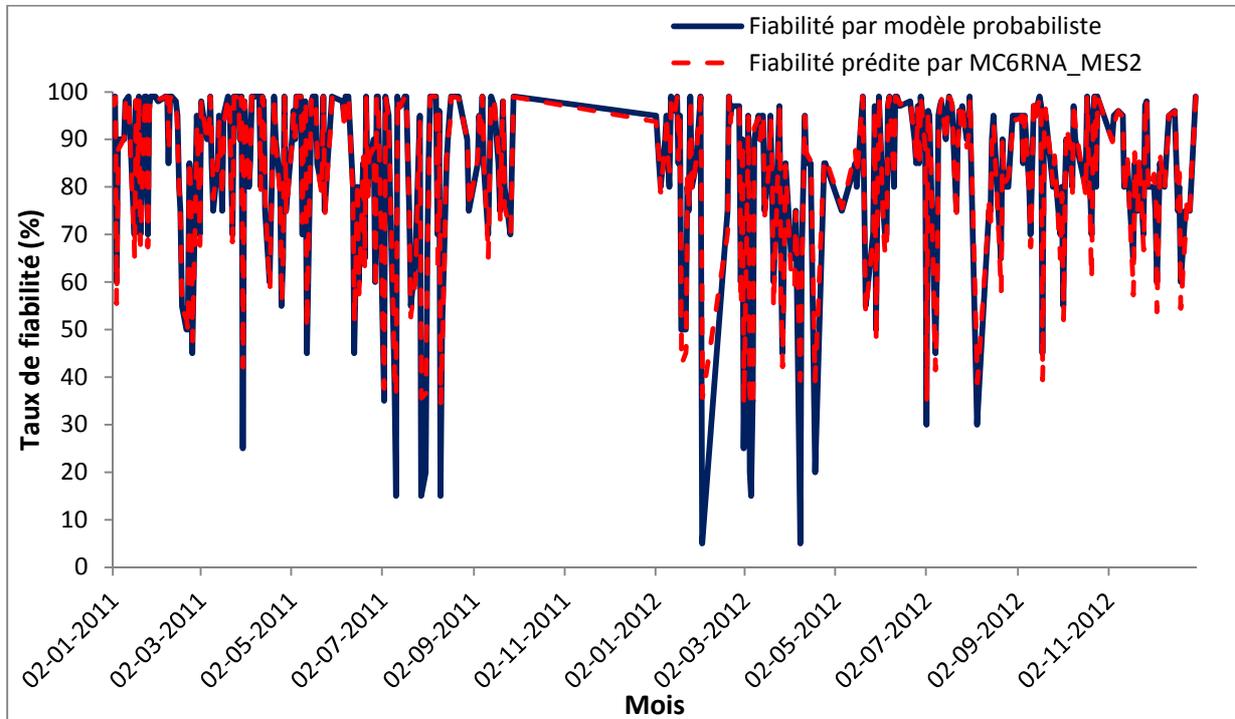


Figure 64: Taux de fiabilité d'élimination des MES prédit par réseau de neurones artificiels par modèle complet RNA_{MES2} (10-6-1), station d'épuration de Khenchela (2011-2012)

VI.6 DISCUSSION DES RESULTATS DE LA PREDICTION DES TAUX DE FIABILITE

Six modèles sont présentées où les entrées sont modifiées pour inclure des combinaisons de différents paramètres pour tenter de modéliser l'effet de la variation de la quantité et les qualités des eaux usées brutes à l'entrée et la sortie de la station d'épuration. Dans cette étude une approche originale a été choisie pour tenter d'inclure l'effet de la température sur la fiabilité du procédé de traitement dans les modèles complets.

La topologie étudiée est Perceptron multicouche PMC un réseau de type Alimentation Avant (en anglais *Feedforward Neural Network*), avec comme règle d'apprentissage l'algorithme de rétro propagation du gradient d'erreur.

La structure physique de tous les réseaux de neurones artificiels considérés dans cette étude dispose d'une seule couche cachée. La fonction d'activation de la couche cachée est la

tangente hyperbolique (\tanh), et la fonction d'activation de la couche de sortie est la fonction sigmoïde.

VI.6.1 Résultats des modèles RNA_{DBO5}

VI.6.1.1 Prédiction du taux de fiabilité par modèle simple MS- RNA_{DBO5} (4-3-1)

Le modèle simple MS- RNA_{DBO5} est présenté dans la sous-section VI.4.1. Les données d'entrées-sortie sont présentées dans le tableau 23. Les performances de prédiction du modèle simple sont présentées dans le tableau 24.

Dans la figure 45, une comparaison graphique entre les taux de fiabilités prédites par le modèle simple RNA_{DBO5} et les taux de fiabilités cible (par modèle probabiliste). En fin une application des taux de fiabilités prédites est représentée pour la période d'études de 4 ans est représenté dans la figure 46.

Les résultats du modèle simple MS- RNA_{DBO5} montrent que lors de la première étape (phase d'apprentissage), l'erreur moyenne absolue de l'apprentissage est de 4,17%, l'erreur quadratique moyenne est de 7,52%. Le coefficient de régression $R= 0,94$, et le coefficient de détermination $R^2= 0,88$. Ces paramètres indiquent que le modèle simple RNA_{DBO5} est en mesure de bien répondre aux données d'apprentissage et capable de les rapprocher. Le modèle simple RNA_{DBO5} est donc en mesure de résoudre le problème particulier de cartographie les données entrée-sortie.

Lors de la deuxième étape (phase de test), les résultats du modèle simple RNA_{DBO5} montrent que l'erreur moyenne absolue de la validation est de 4,37%, l'erreur quadratique moyenne est de 7,835%. Le coefficient de régression $R= 0,93$, et le coefficient de détermination $R^2= 0,86$. Ces résultats indiquent que les performances de généralisation du modèle simple RNA_{DBO5} sont bonnes. Le modèle simple MS- RNA_{DBO5} est en mesure de faire des prédictions précises.

VI.6.1.2 Prédiction du taux de fiabilité par modèle complet MC- RNA_{DBO5} (14-3-1)

Le modèle complet RNA_{DBO5} est présenté dans la sous-section VI.5.1. Les données d'entrées-sortie sont présentées dans le tableau 29. Les performances de prédiction du modèle complet sont présentées dans le tableau 30.

Dans la figure 54, une comparaison graphique entre les taux de fiabilités prédites par le modèle complet MC- RNA_{DBO5} et les taux de fiabilités cible (par modèle probabiliste). En fin une application des taux de fiabilités prédites est représentée pour la période d'études de 4 ans est représentée dans la figure 55.

Les résultats du modèle complet RNA_{DBO5} montrent que lors de la première étape (phase d'apprentissage), l'erreur moyenne absolue de l'apprentissage est de 4,61%, et l'erreur

quadratique moyenne est de 7,42%. Le coefficient de régression $R= 0,93$, et le coefficient de détermination $R^2= 0,86$. Ces paramètres indiquent que le modèle complet RNA_{DBO5} a bien répondu aux données d'apprentissage et qu'il est capable de les rapprocher.

Lors de la deuxième étape (phase de test), les résultats du modèle complet RNA_{DBO5} montrent que l'erreur moyenne absolue de la validation est de 5,54%, l'erreur quadratique moyenne est de 10,24%. Le coefficient de régression $R= 0,87$, et le coefficient de détermination $R^2= 0,76$. Ces résultats indiquent que le modèle complet RNA_{DBO5} trouve des difficultés de généralisation due à la croissance des données d'entrée et du degré de complexité de l'interaction des paramètres d'entrée ce qui rend la tâche de modéliser la fiabilité du procédé encore plus difficile, malgré cet effet de croissance de la complexité de l'interconnexion entre les données d'entrée, le modèle $MC-RNA_{DBO5}$ est en mesure de faire des prédictions de bonne qualité.

VI.6.2 Résultats des modèles RNA_{DCO}

VI.6.2.1 Prédiction du taux de fiabilité par modèle simple $MS-RNA_{DCO}$ (4-2-1)

Le modèle simple RNA_{DCO} est présenté dans la sous-section VI.4.2. Les données d'entrées et de sortie sont présentées dans le tableau 25. Les performances de prédiction du modèle simple sont présentées dans le tableau 26.

Dans la figure 48, une comparaison graphique entre les taux de fiabilités prédites par le modèle simple RNA_{DBO5} et les taux de fiabilités cible (par modèle probabiliste). En fin une application des taux de fiabilités prédites est représentée pour la période d'études de 4 ans est représenté dans la figure 49.

Lors de la première étape (phase d'apprentissage) les résultats du modèle simple $MS-RNA_{DCO}$ montrent que l'erreur moyenne absolue de l'apprentissage est de 1,63%, et l'erreur quadratique moyenne est de 3,77%. Le coefficient de régression $R= 0,95$, et le coefficient de détermination $R^2= 0,90$. Ces paramètres indiquent que le modèle simple RNA_{DCO} a bien répondu aux données d'apprentissage et qu'il est capable de les rapprocher. Le modèle simple RNA_{DCO} est en mesure de cartographier les données entrée-sortie.

Les résultats du modèle MS -simple $MS-RNA_{DCO}$ lors de la deuxième étape (phase de test) montrent une légère croissance de l'erreur moyenne absolue de la validation estimée à 1,92%, ainsi que celle de l'erreur quadratique moyenne avoisinant les 5%. Le coefficient de régression $R= 0,93$, et le coefficient de détermination $R^2= 0,86$. Cette légère détérioration de la performance du modèle est due au fait que les données du set de validation sont trop étendues (l'écart types des données de validation $DCO_{entrée}$ et DCO_{sortie} sont plus importants que ceux des données d'apprentissage). Malgré cet effet, les performances de généralisation du modèle simple sont bonnes, et le modèle $MS-RNA_{DCO}$ est en mesure de faire des prédictions précises.

VI.6.2.2 Prédiction du taux de fiabilité par modèle complet MC-RNA_{DCO} (14-8-1)

Le modèle complet MC-RNA_{DBO5} est présenté dans la sous-section VI.5.2. Les données d'entrées-sortie sont présentées dans le tableau 31. Les performances de prédiction du modèle complet sont présentées dans le tableau 32.

Une étude comparative entre les taux de fiabilités prédites par le modèle complet MC-RNA_{DCO} et les taux de fiabilités cible (par modèle probabiliste) est représentée sur la figure 57. La figure 58 représente une application du modèle de prédiction des taux de fiabilité durant une période d'études de 4 ans.

Les résultats du modèle complet MC-RNA_{DCO} montrent que lors de la première étape (phase d'apprentissage), l'erreur moyenne absolue de l'apprentissage est minimal estimé à 2,83%, alors que l'erreur quadratique moyenne est proche de 7,5%. Le coefficient de régression $R=0,90$, et le coefficient de détermination $R^2=0,81$. Ces paramètres indiquent que le modèle complet RNA_{DCO} est capable de les rapprocher des données d'apprentissage.

Par contre, lors de la deuxième étape (phase de test), les résultats du modèle complet RNA_{DCO} on constate que l'erreur moyenne absolue de la validation a évolué. L'erreur quadratique moyenne dépasse les 8%. Le coefficient de régression $R=0,69$, et la valeur du coefficient de détermination décroît considérablement à 0,47 ce qui signifie que l'ajustement entre les valeurs prédites et les valeurs cibles est médiocre. Ces résultats indiquent que le modèle complet MC-RNA_{DCO} trouve des difficultés de généralisation due à la difficulté de modéliser l'interaction entre les différentes données d'entrée. Malgré l'effet de la complexité de l'interconnexion entre les données d'entrée et la croissance des données d'entrée, le modèle MC-RNA_{DBO5} fourni des prédictions de qualité acceptable.

VI.6.3 Résultats des modèles RNA_{MES}

VI.6.3.1 Prédiction du taux de fiabilité par modèle simple MS-RNA_{MES1} (4-4-1)

Le modèle simple RNA_{MES} est présenté dans la sous-section VI.4.3. Les données d'entrées-sortie sont présentées dans le tableau 27. Les performances de prédiction du modèle simple sont présentées dans le tableau 28.

Une comparaison entre les taux de fiabilités prédites par le modèle simple MS-RNA_{MES} et les taux de fiabilités cibles (par modèle probabiliste) est présentée dans la figure 51. En fin sur la figure 52 une présentation graphique a pour objet la prédiction des taux de fiabilités comparés aux taux de fiabilité par modèle probabiliste durant une période de 4 ans.

Lors de son apprentissage le modèle simple RNA_{MES} affiche une erreur moyenne absolue de l'ordre de 2,5%, et d'une erreur quadratique moyenne inférieur à 5%.

Une très bonne valeur du coefficient de régression R égale à 0,97, et un coefficient de détermination R^2 proche de 0,94. Les paramètres de performances du modèle simple RNA_{MES}

indiquent que le modèle simple MS-RNA_{MES} est capable de cartographier les données entrée-sortie des données d'apprentissage.

Les résultats du modèle simple MS-RNA_{MES} en phase de validation montrent que les performances de généralisation du modèle simple sont très bonnes, et que ce modèle est en mesure de fournir des prédictions précises.

VI.6.3.2 Prédiction du taux de fiabilité par modèles complets MC-RNA_{MES1,2} (10-3-1 ; 10-6-1)

Pour le bon déroulement de cette étude, il a été décidé de choisir un ensemble de données plus important pour tenter d'utiliser un modèle complet. L'ensemble des données est constitué de 784 données, et sont respectivement partagées sous forme biannuelles (2009-12010) et (2011-2012).

Les deux modèles complets MC-RNA_{MES1} et MC-RNA_{MES2} sont présentés dans la sous-section VI.5.3. Les données d'entrées-sortie sont présentées dans le tableau 33. Les performances de prédiction des deux modèles complets sont présentées dans les tableaux 34 et 35. Une étude comparative entre les taux de fiabilités prédites par les modèles complets MC-RNA_{MES1,2} et les taux de fiabilités cibles (par modèle probabiliste) sont représentées sur les figure 60 et 63.

Les figures 61 et 64 englobent une application des deux modèles de prédiction des taux de fiabilité d'élimination des MES en comparaison avec les taux de fiabilités déterminé par modèle probabiliste durant une période d'études de 4 ans.

Les résultats des modèles complets MC-RNA_{MES1,2} montrent que lors de la phase d'apprentissage, l'erreur moyenne absolue est minimal estimé à 2,53% pour le deuxième modèle 2,831%, alors qu'elle est inférieure à 2% pour le premier modèle. L'erreur quadratique moyenne des deux modèles est inférieure à 5%. Les deux modèles ont presque le même coefficient de régression R proche de 0,98, quant au coefficient de détermination des deux modèles, il est proche de 0,95. Ces paramètres indiquent que le modèle complet RNA_{DCO} est capable de les rapprochés les données d'apprentissage.

Les résultats des deux modèles complets MC-RNA_{MES1,2} durant la étape de validation affichent des erreurs moyennes absolues inférieure à 3%. Par contre les erreurs quadratiques moyennes sont légèrement supérieures à 5%. Les coefficients de régression R sont de l'ordre de 0,97, et les coefficients de détermination proche de 0,95 ce qui signifie que les ajustements entre les valeurs prédites et les valeurs cibles sont de très bonne qualité. Ces résultats indiquent que les deux modèles arrivent à généraliser, et sont en mesure de réaliser des prédictions précises.

VI.7. CONCLUSION

L'application des modèles simples pour la prédiction des performances d'une station d'épuration présentent un intérêt spécial vue leurs simplicité et leurs facilité d'utilisation pour l'estimation des taux de fiabilité quotidienne du procédé de traitement. Ces résultats peuvent être un très bon indicateur aux exploitants pour servir de support d'aide à la gestion et l'exploitation des différents compartiments du procédé de traitement.

L'application des modèles complets pour la prédiction des taux de fiabilités représente un intérêt scientifique et académique pour tenter de modéliser des phénomènes non linéaires très complexes qui mettent en contribution un grand nombre de paramètres.

Globalement, les résultats montrent que les modèles RNA sont des alternatives prometteuses pour l'estimation des performances d'une station d'épuration qui peuvent substituer aux méthodes d'estimation des performances traditionnelle basés sur des modèles probabiliste nécessitant une maîtrise des outils statistiques. Toutefois, des améliorations supplémentaires aux structures RNA doivent être étudiées.

CHAPITRE VII : PREDICTION DU TAUX DE DEFAILLANCES PAR LES RESEAUX DE NEURONES ARTIFICIELS (Cas d'étude : station d'épuration de Khenchela)

VII.1 INTRODUCTION

Un dysfonctionnement d'une installation de traitement des eaux usées est un problème social et biologique délicat. Une eau usée mal traitée peut engendrer des conséquences dangereuses pour la santé des êtres humains ainsi que le milieu récepteur (Blanche et al., 1992).

Les réseaux de neurones offrent la possibilité de modéliser des données arbitraires d'entrée au moyen de l'ajustement des connexions internes au réseau, de sorte que pour une entrée donnée de la différence entre la sortie du réseau et la réponse souhaitée, l'erreur est minimisée. C'est le processus de l'entraînement du réseau au moyen d'un apprentissage supervisé, de sorte que l'erreur du réseau est réduit de manière itérative sur un ensemble de vecteurs correspondant à des entrées-sorties d'entraînement. Pourvu que l'ensemble des données d'entraînement soit représentatif du processus, le modèle RNA résultant saisit la relation inhérente entre l'entrée et la sortie et la possibilité de généraliser pour les entrées invisibles futures (Garvey, 1997).

L'objectif principal de cette recherche est de trouver des modèles RNA pouvant prédire les taux de fiabilité, ainsi que les taux de défaillances des principaux paramètres de contrôles des eaux usées traités dans une station d'épuration (MES, DCO et DBO₅).

L'un des défis est de trouver une topologie de réseau optimale en faisant varier les types d'entrées du réseau et ainsi de résoudre le problème particulier de la cartographie d'entrée-sortie. L'architecture du réseau approprié pour une application donnée est fortement dépendante du problème, de la diversité de l'ensemble d'apprentissage et de la complexité de la fonction sous-jacente (Garvey, 1997).

VI.2 CONSTRUCTION DES MODELES RNA

VII.2.1 Modèle simple RNA utilisées pour la prédiction du taux de défaillance

L'approche choisie dans la construction de ces modèles est présenté dans le tableau 36. Les modèles construits seront simples à manipuler et ne nécessitant que quelques données quotidiennes disponibles et faciles à mesurer.

Les données d'auto-surveillances journalières serviront pour prédire les différents taux de défaillance du procédé de traitement.

Tableau 36: Données entrées-sortie des différents modèles simples RNA (prédiction du taux de défaillance).

Modèle Simple RNA	Données d'entrée				Donnée de sortie
MS-RNA _{DBO5} (123 données)	Débit _{entrée} , DBO _{5entrée} , DCO _{entrée} , MES _{entrée} , DBO _{5sortie}				Défaillance- DBO ₅ prédite
MS-RNA _{DCO} (123 données)	Débit _{entrée} , DCO _{entrée} , DBO _{5entrée} , MES _{entrée} , DCO _{sortie}				Défaillance- DCO prédite
MS-RNA _{MES} (123 données)	Débit _{entrée}	MES _{entrée}	Débit _{sortie}	MES _{sortie}	Défaillance- MES prédite

La construction des modèles RNA est basée sur une combinaison de paramètres d'entrée, chaque paramètre de qualité étudié est modélisé par la concentration à la sortie de la station, ainsi que les taux de défaillance estimés par la méthode probabiliste de Niku et al. (1979) qui serviront de données cibles utilisées lors de la phase d'apprentissage et de validation du modèle RNA.

VII.2.2 Modèles complets RNA utilisés pour la prédiction du taux de défaillance

La tentative de modéliser l'effet de la variation des quantités et de la qualité des effluents à l'entrée de la station d'épuration de Khenchela présente un intérêt majeur pour modéliser les performances du procédé et les éventuelles défaillances. L'approche utilisée pour la construction de ces modèles complets est présentée dans le tableau 37.

Tableau 37: Données entrées-sortie des différents modèles complets RNA (prédiction du taux de défaillance).

Modèle Complet RNA	Données d'entrée			Donnée de sortie
MC-RNA _{DBO5} <i>(123 données)</i>	Débit _{entrée}	Débit _{sortie}	Température _{entrée}	Défaillance- DBO ₅ prédite
	DBO ₅ entrée	DBO ₅ sortie	Conductivité _{entrée}	
	MES _{entrée}	DCO _{sortie}		
	DCO _{entrée}	pH _{entrée}		
MC-RNA _{DCO} <i>(123 données)</i>	Débit _{entrée}	Débit _{sortie}	Température _{entrée}	Défaillance - DCO prédite
	DCO _{entrée}	DBO ₅ sortie	Conductivité _{entrée}	
	DBO ₅ entrée	DCO _{sortie}		
	MES _{entrée}	pH _{entrée}		
MC-RNA _{MES} <i>(123 données)</i>	Débit _{entrée}	Débit _{sortie}	Température _{entrée}	Défaillance- MES prédite
	MES _{entrée}	MES _{sortie}	Température _{sortie}	
	DCO _{entrée}	DBO ₅ sortie	Conductivité _{entrée}	
	DBO ₅ entrée	DCO _{sortie}	Conductivité _{sortie}	
		pH _{entrée}		
	pH _{sortie}			

VII.3 PREDICTION DU TAUX DE DEFAILLANCE DU PROCEDE DE TRAITEMENT

VII.3.1 Prédiction du taux de défaillance de la dégradation de la DBO5 par modèle simple

Les données d'entrées du modèle simple RNA_{DBO5} sont le débit d'entrée, la concentration de la DBO₅ à l'entrée de la station, la concentration de la DCO à l'entrée de la station, et la concentration de la DBO₅ à la sortie de la station, présentés dans le tableau ci-dessous.

Tableau 38: Données d'entrées-sortie du modèle simple MS-RNA_{DBO5} pour la prédiction du taux de défaillance de la dégradation de la DBO₅.

Entrées	Sortie
Débit _{entrée} , DBO _{5entrée} , DCO _{entrée} , MES _{entrée} , DBO _{5sortie}	Taux de défaillance-DBO ₅

Le modèle construit pour prédire le taux de la défaillance de la dégradation de la DBO₅ du procédé de traitement de la station d'épuration de Khenchela est constitué de 5 neurones dans la couche d'entrées, 2 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

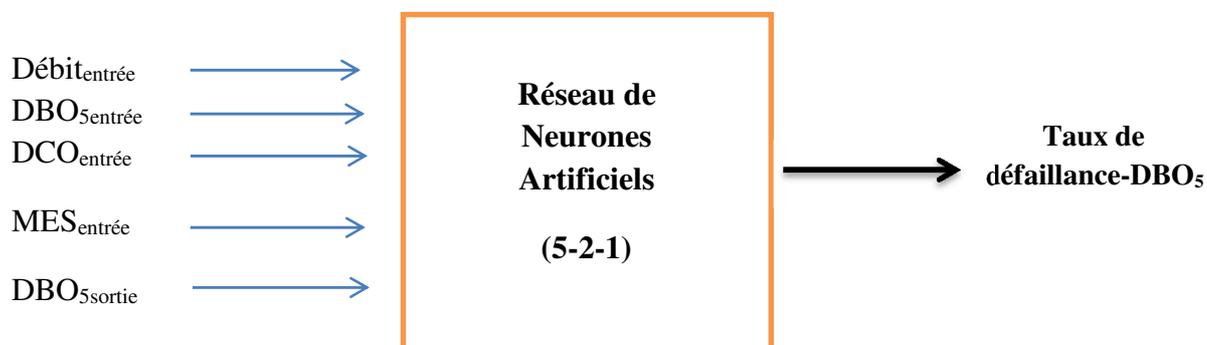


Figure 65: Schéma générale du modèle simple MS-RNA_{DBO5} (5-2-1) pour la prédiction du taux de défaillance de la dégradation de la DBO₅.

Les performances de la prédiction sont présentées dans le tableau ci-dessous.

Tableau 39 : Performances de la prédiction du taux de défaillance de la dégradation de la DBO₅ par le modèle simple MS-RNA_{DBO5}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,92	0,85	4,34	8,03
² Données d'apprentissage	0,94	0,88	3,8	7,48
³ Données de validation	0,90	0,81	5,60	9,19

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction du taux de défaillance de la dégradation de la DBO₅ du procédé de traitement de la station de Khenchela est représentée sur la figure 66.

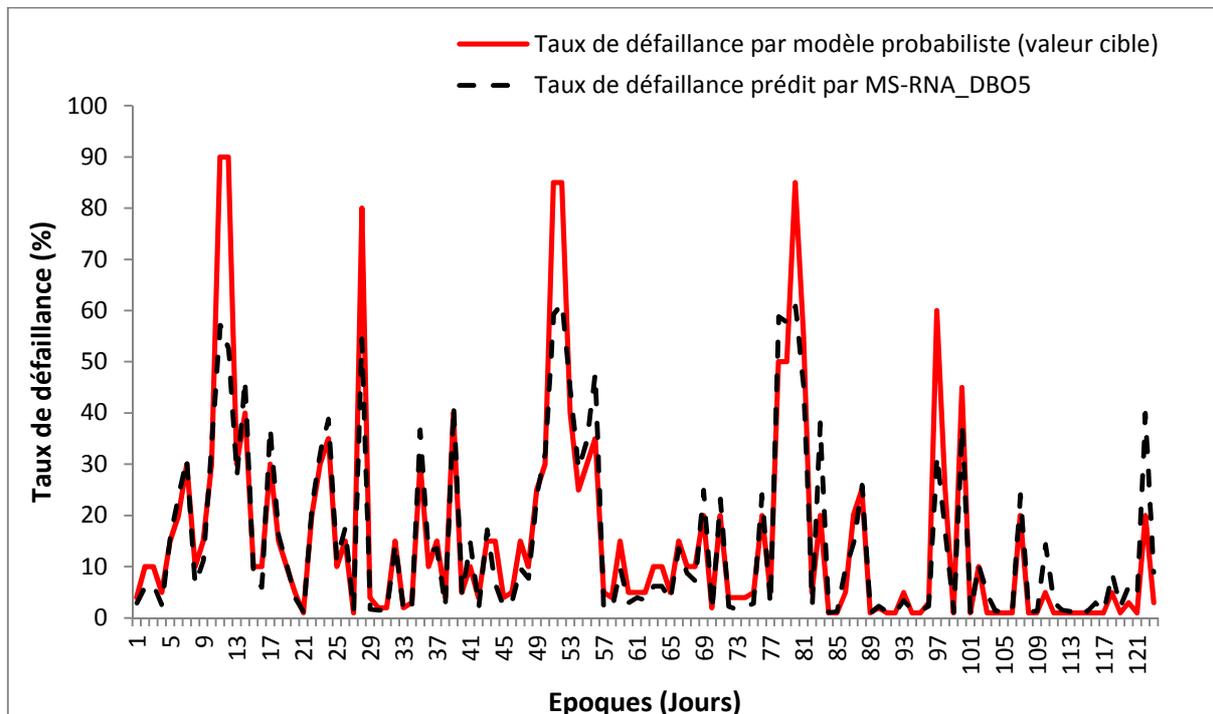


Figure 66 : Taux de défaillance de la dégradation de la DBO₅ prédit par modèle simple MS-RNA_{DBO5} (5-2-1) versus taux de défaillance par modèle probabiliste (valeur cible).

L'application du modèle simple RNA_{DBO_5} pour la prédiction du taux de défaillance de la dégradation de la DBO_5 dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 67.

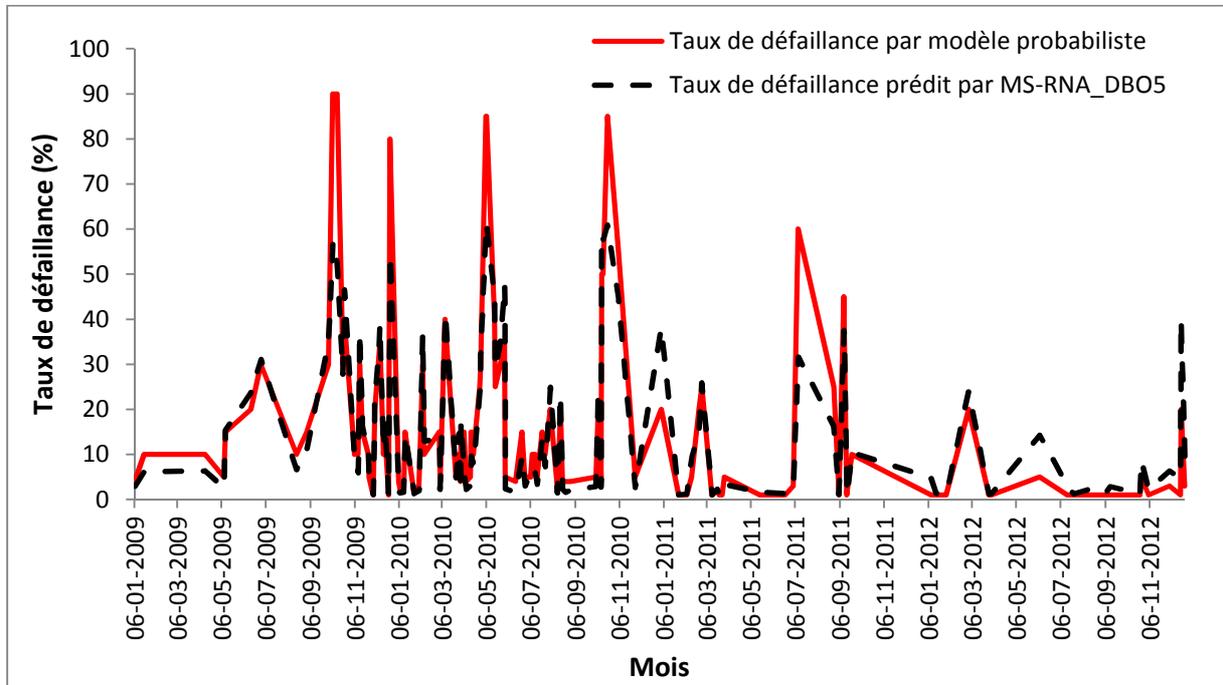


Figure 67 : Taux de défaillance de la dégradation de la DBO_5 prédit par réseau de neurones artificiels simple (5-2-1), station d'épuration de Khenchela (2009-2012).

VII.3.2 Prédiction du taux de défaillance de l'élimination de la DCO par modèle simple

Les données d'entrées-sortie du modèle simple RNA_{DCO} pour la prédiction du taux de défaillance de l'élimination de la DCO du procédé de traitement de la station d'épuration de Khenchela sont présentées dans le tableau ci-dessous.

Tableau 40 : Données d'entrées-sortie du modèle simple $MS-RNA_{DCO}$ pour la prédiction du taux de défaillance d'élimination de la DCO.

Entrées	Sortie
Débit _{entrée} , DCO _{entrée} , MES _{entrée} , DBO ₅ _{entrée} , DCO _{sortie}	Taux de défaillance-DCO

Le modèle construit pour la prédiction du taux de défaillance de l'élimination de la DCO du procédé de traitement de la station d'épuration de Khenchela est constitué de 5 neurones dans la couche d'entrées, 8 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

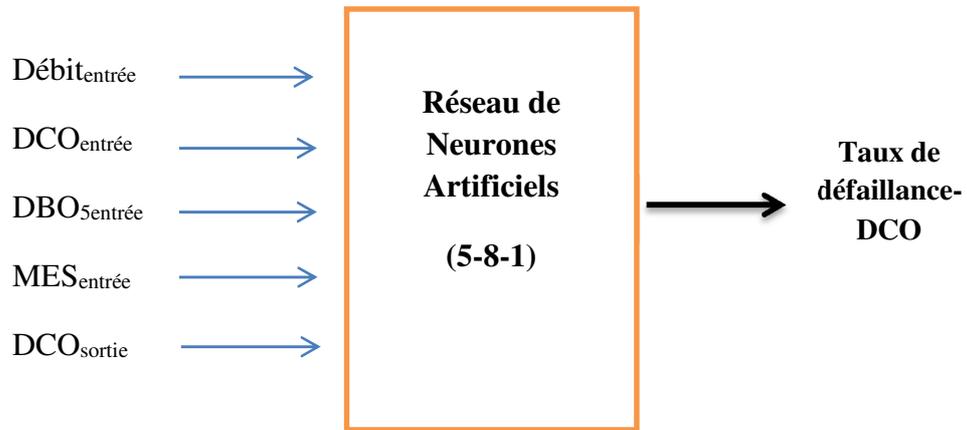


Figure 68 : Schéma générale du modèle simple MS-RNA_{DCO} (5-8-1) pour la prédiction du taux de défaillance d'élimination de la DCO.

Les performances de la prédiction du modèle simple RNA_{DCO} sont présentées dans le tableau ci-dessous.

Tableau 41 : Performances de la prédiction du taux de défaillance d'élimination de la DCO par le modèle simple MS-RNA_{DCO}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,97	0,94	1,46	3,88
² Données d'apprentissage	0,97	0,94	1,45	3,75
³ Données de validation	0,96	0,92	1,48	4,16

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction du taux de défaillance de l'élimination de la DCO du procédé de traitement de la station d'épuration de Khenchela est représentée sur la figure 69.

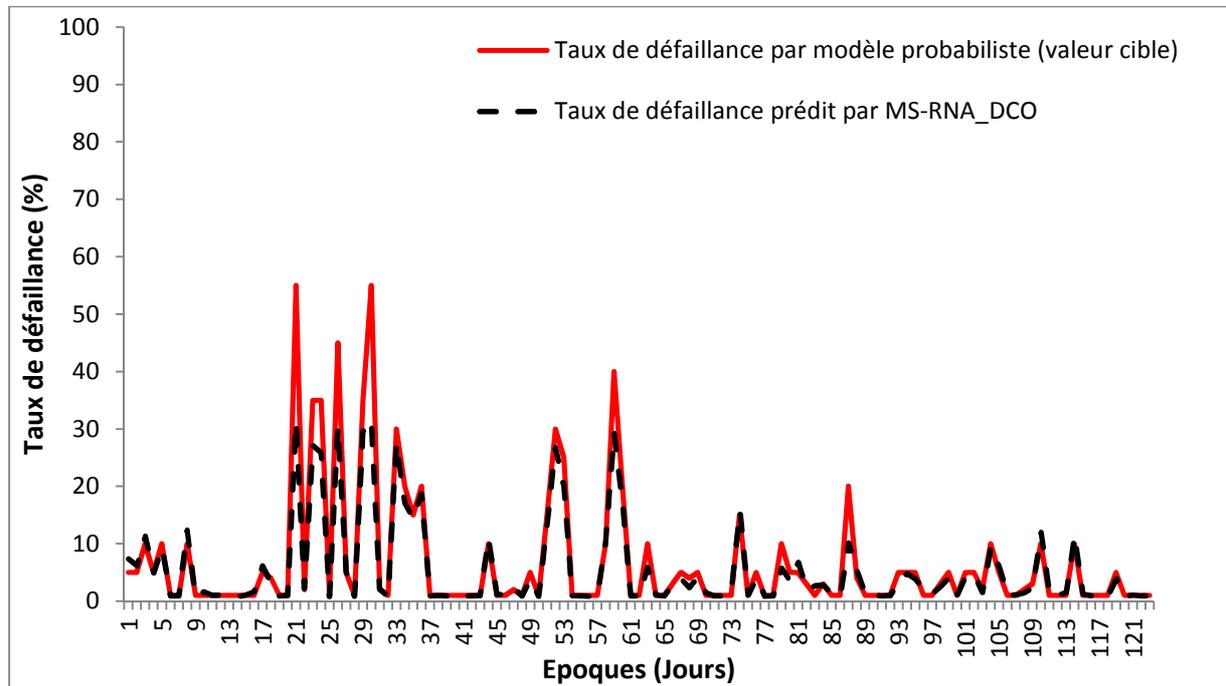


Figure 69 : Taux de défaillance d'élimination de la DCO prédit par modèle simple MS-RNA_{DCO} (5-8-1) versus taux de défaillance par modèle probabiliste (valeur cible).

L'application du modèle simple RNA_{DCO} pour la prédiction du taux de défaillance d'élimination de la DCO du procédé de traitement de la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 70.

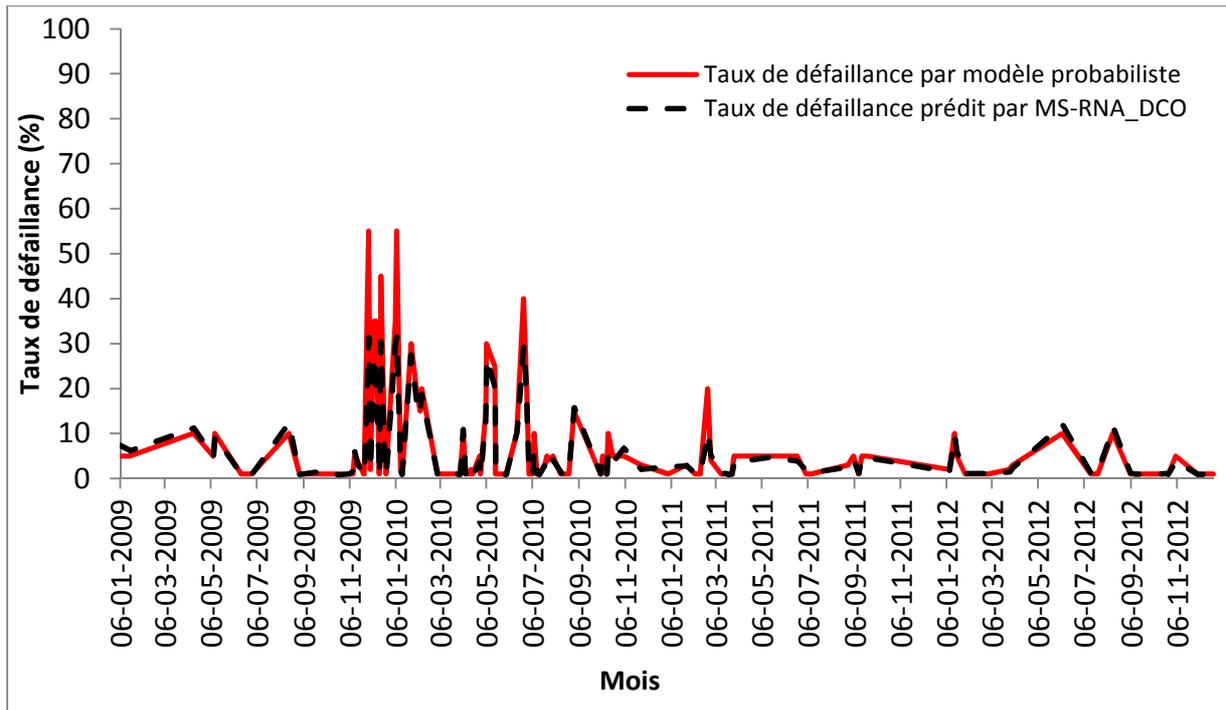


Figure 70 : Taux de défaillance d'élimination de la DCO prédit par réseau de neurones artificiels simple (5-8-1), station d'épuration de Khenchela (2009-2012).

VII.2.3 Prédiction du taux de défaillance de l'élimination des MES par modèle simple

Les données d'entrée du modèle simple RNA_{MES} sont : le débit d'entrée, le débit de sortie, la concentration des MES à l'entrée de la station, et la concentration des MES à la sortie de la station.

Tableau 42 : Données d'entrées-sortie du modèle simple MS-RNAMEs pour la prédiction du taux de défaillance d'élimination des MES.

Entrées	Sortie
Débit _{entrée} , Débit _{sortie} , MES _{entrée} , MES _{sortie}	Taux de défaillance-MES

Le modèle construit pour prédire le taux de fiabilité de l'élimination des MES de la station d'épuration de Khenchela est constitué de 4 neurones dans la couche d'entrées, 3 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.



Figure 71: Schéma générale du modèle simple MS-RNAMES (4-3-1) pour la prédiction du taux de défaillance d'élimination des MES.

Les performances de la prédiction du modèle simple RNA_{MES} sont présentées dans le tableau ci-dessous.

Tableau 43 : Performances de la prédiction du taux de défaillance d'élimination des MES par le modèle simple MS- RNA_{MES} .

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,97	0,94	2,56	4,74
² Données d'apprentissage	0,97	0,94	2,43	4,50
³ Données de validation	0,96	0,92	2,87	5,40

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction du taux de défaillance de l'élimination des MES du procédé de traitement de la station de Khenchela est représentée sur la figure 72.

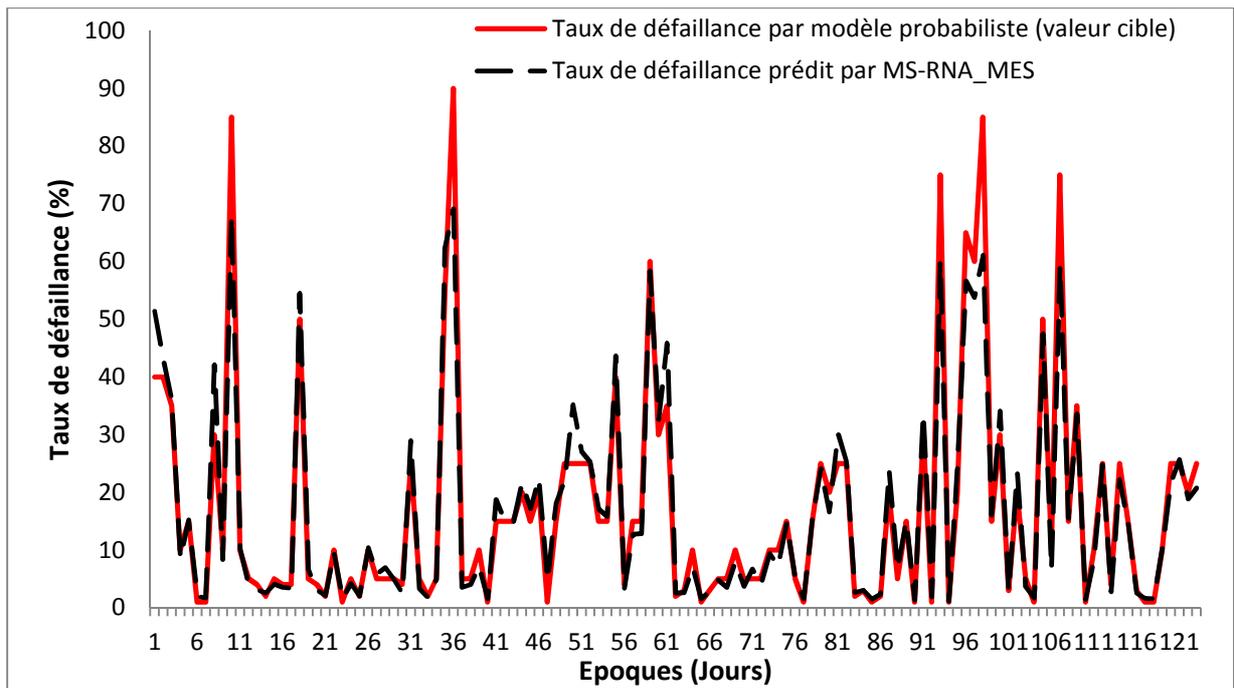


Figure 72 : Taux de défaillance d'élimination des MES prédit par modèle simple MS-RNA_{MES} (4-3-1) versus taux de défaillance par modèle probabiliste (valeur cible).

L'application du modèle simple RNA_{MES} pour la prédiction du taux de fiabilité de l'élimination des MES du procédé de traitement de la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 73.

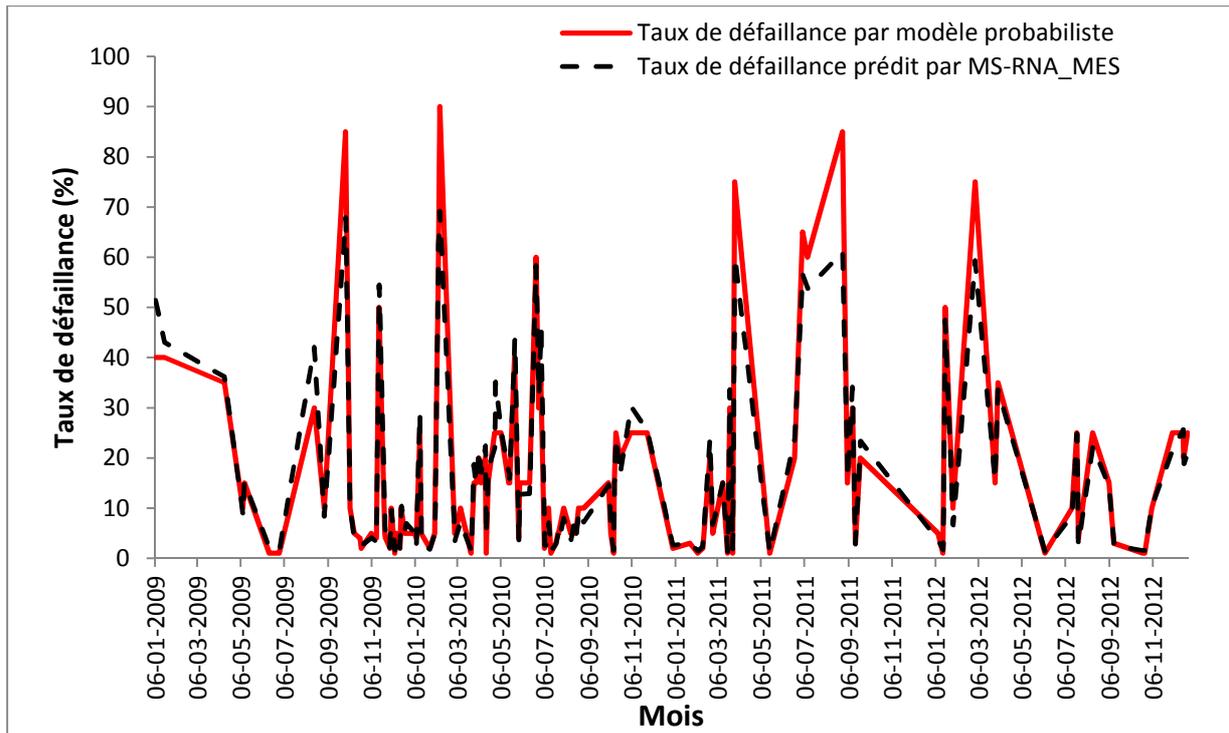


Figure 73 : Taux de défaillance d'élimination des MES prédit par réseau de neurones artificiels simple (4-3-1), station d'épuration de Khenchela (2009-2012).

VII.3 PREDICTION DU TAUX DE DEFAILLANCE DU PROCEDURE DE TRAITEMENT DES EAUX USEES PAR MODELE COMPLET

VII.3.1 Prédiction du taux de défaillance de la dégradation de la DBO5 par modèle complet

Les données d'entrée du modèle complet RNA_{DBO_5} sont tous les données disponible à l'entrée et à la sortie de la station qui sont les suivant : débit d'entrée et de sortie, les paramètres de qualité des eaux usées brutes et traités (concentration des MES, DCO et DBO₅), température de l'eau à l'entrée et sortie de la station, la conductivité à l'entrée et sortie de la station, et le pH des eaux usées brutes, et traités.

Tableau 44 : Données d'entrées-sortie du modèle complet $MC-RNA_{DBO_5}$ pour la prédiction du taux de défaillance de la dégradation de la DBO₅.

Entrées	Sortie
Débit _{entrée} , DBO ₅ _{entrée} , DCO _{entrée} , MES _{entrée} , Débit _{sortie} , DBO ₅ _{sortie} , DCO _{sortie} , pH _{entrée} , Température _{entrée} , Conductivité _{entrée}	Taux de défaillance-DBO ₅

Le modèle construit pour prédire le taux de fiabilité de la dégradation de la DBO₅ de la station d'épuration de Khenchela est constitué de 10 neurones dans la couche d'entrée, 2 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

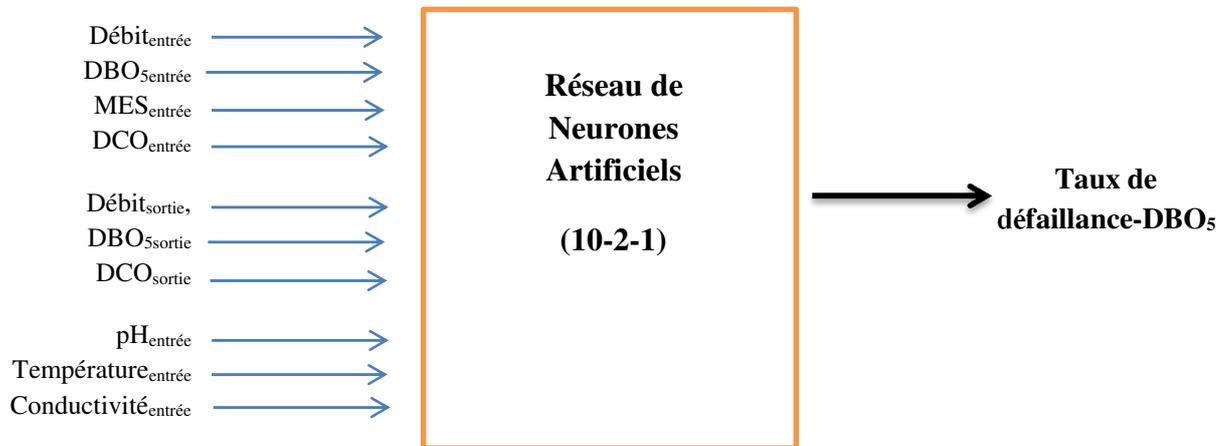


Figure 74 : Schéma générale du modèle complet MC-RNA_{DBO5} (10-2-1) pour la prédiction du taux de défaillance de la dégradation de la DBO₅.

Les performances de la prédiction par le modèle complet RNA_{DBO5} sont présentées dans le tableau 45.

Tableau 45 : Performances de la prédiction du taux de défaillance de la dégradation de la DBO₅ par le modèle complet MC-RNA_{DBO5}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,94	0,88	3,71	7,03
² Données d'apprentissage	0,94	0,88	3,76	7,10
³ Données de validation	0,95	0,90	3,61	6,88

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction du taux de défaillance de la dégradation de la DBO₅ du procédé de traitement de la station d'épuration de Khenchela est représentée sur la figure 75.

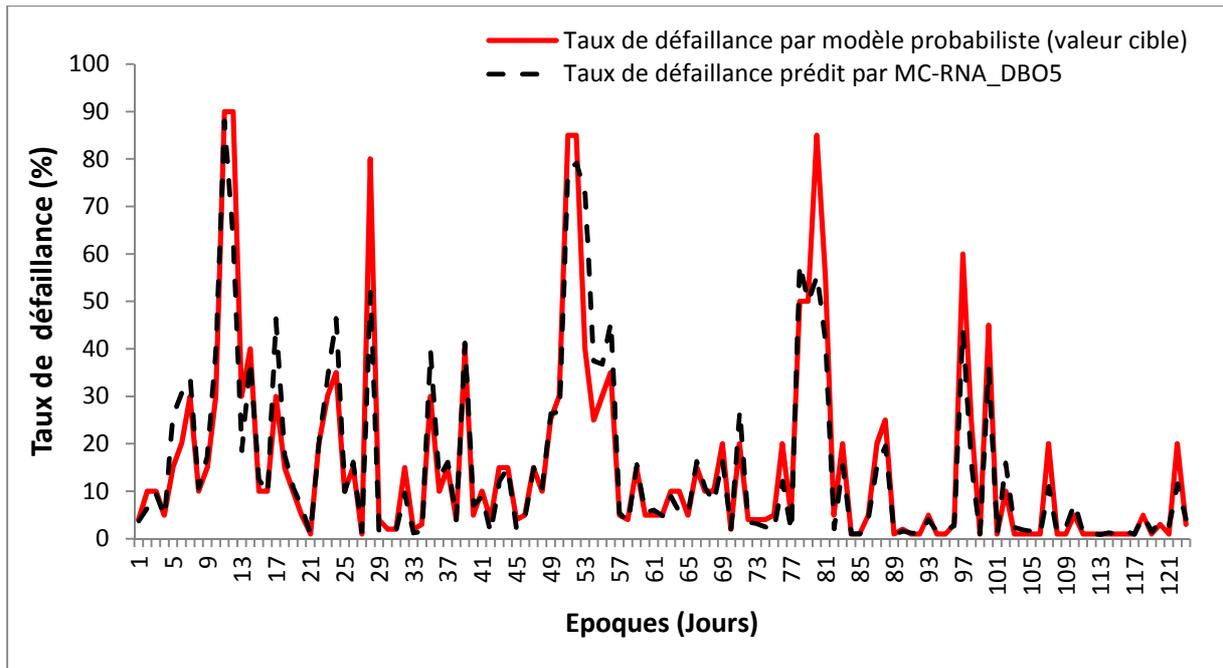


Figure 75 : Taux de défaillance de la dégradation de la DBO₅ prédit par modèle complet MC-RNAD_{BO5} (10-2-1) versus taux de défaillance par modèle probabiliste (valeur cible).

L'application du modèle complet RNA_{DBO5} pour la prédiction du taux de défaillance de la dégradation de la DBO₅ dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 76.

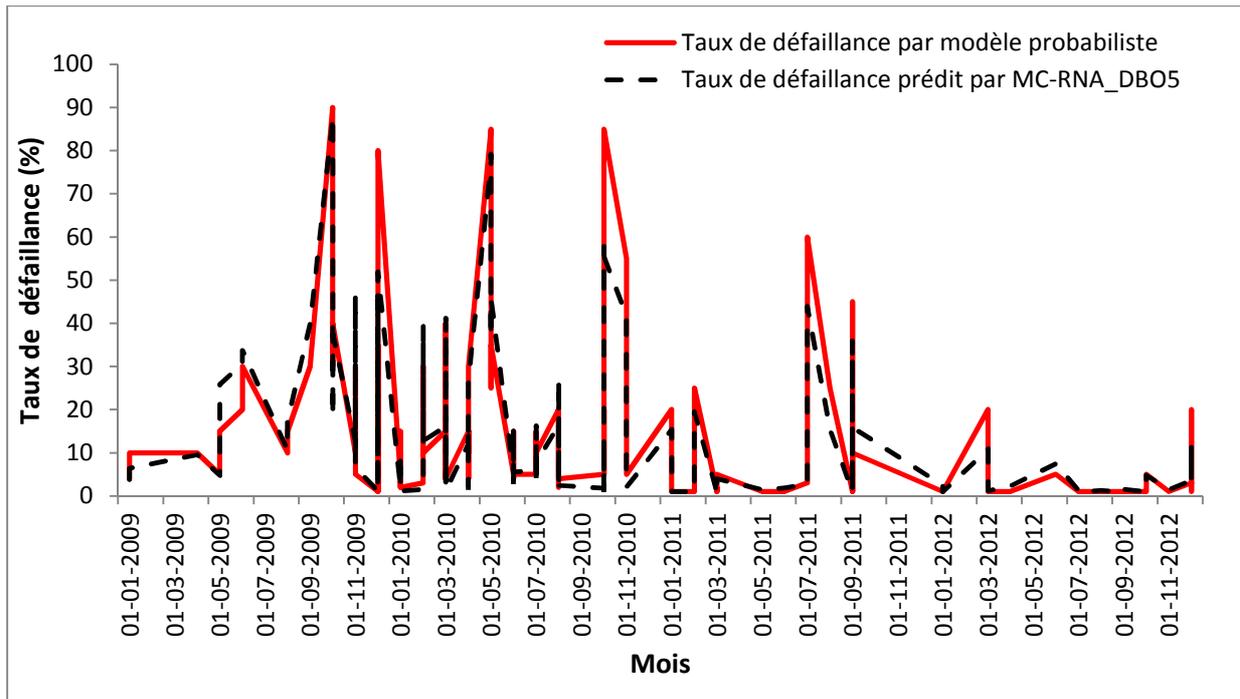


Figure 76 : Taux de défaillance de la dégradation de la DBO₅ prédit par réseau de neurones artificiels complet (10-2-1), station d'épuration de Khenchela (2009-2012).

VII.3.2 Prédiction du taux de défaillance de l'élimination de la DCO par modèle complet

Les données d'entrée du modèle complet RNA_{DCO} sont tous les données disponibles à l'entrée et à la sortie de la station qui font l'objet de mesures quotidiennes pour le contrôle de la qualité des eaux traitées et assurer un fonctionnement adéquat de la station.

Tableau 46 : Données d'entrées-sortie du modèle complet MC-RNA_{DCO} pour la prédiction du taux de défaillance d'élimination de la DCO.

Entrées	Sortie
Débit _{entrée} , DBO ₅ _{entrée} , DCO _{entrée} , MES _{entrée} , Débit _{sortie} , DBO ₅ _{sortie} , DCO _{sortie} , pH _{entrée} , Température _{entrée} , Conductivité _{entrée}	Taux de défaillance-DCO

Le modèle construit pour prédire le taux de fiabilité de la dégradation de la DCO de la station d'épuration de Khenchela est constitué de 10 neurones dans la couche d'entrée, 4 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

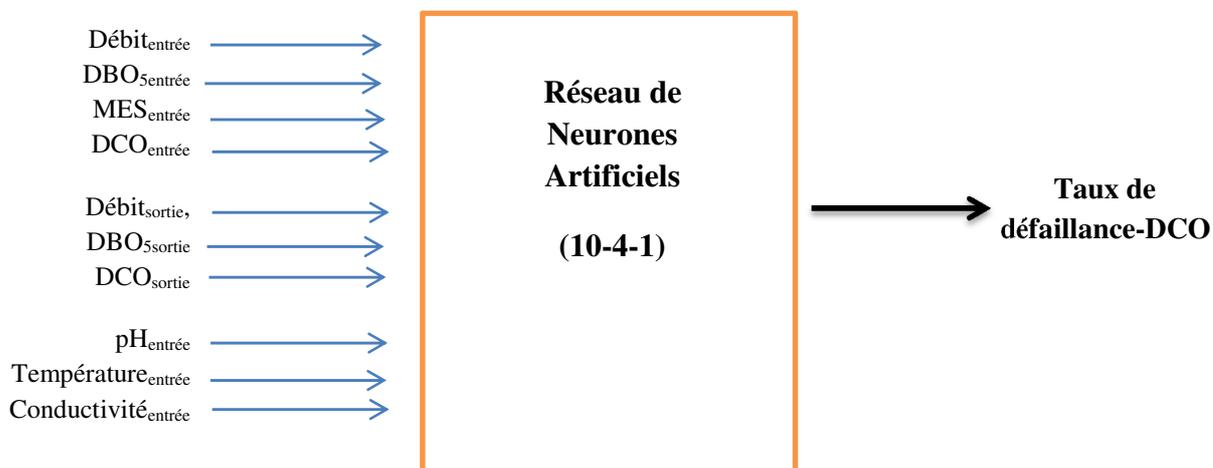


Figure 77 : Schéma générale du modèle complet MC-RNA_{DCO} (10-4-1) pour la prédiction du taux de défaillance d'élimination de la DCO.

Les performances de la prédiction par le modèle complet RNA_{DCO} sont présentées dans le tableau 47 ci-dessous.

Table 47 : Performances de la prédiction du taux de défaillance d'élimination de la DCO par le modèle complet MC-RNA_{DCO}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,93	0,86	2,20	5,54
² Données d'apprentissage	0,95	0,90	1,90	4,78
³ Données de validation	0,93	0,86	2,90	6,98

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction du taux de défaillance de l'élimination de la DCO du procédé de traitement de la station de Khenchela est représentée sur la figure 78.

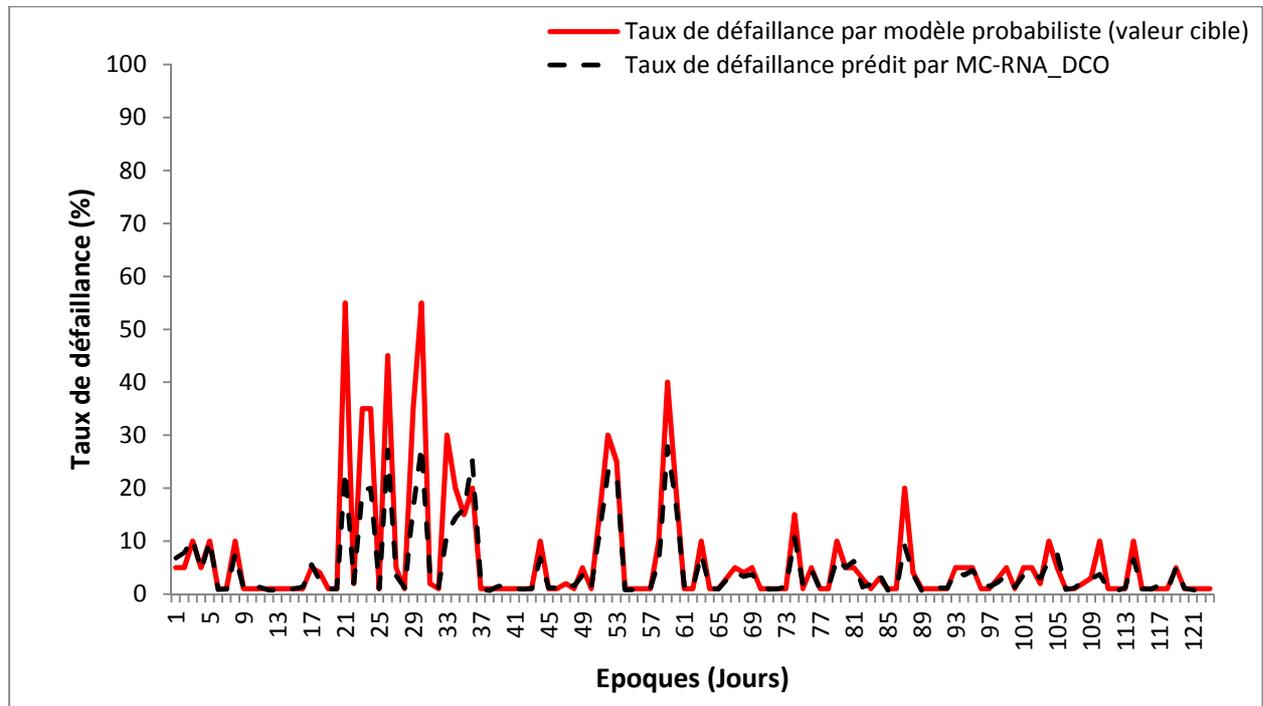


Figure 78 : Taux de défaillance d'élimination de la DCO prédit par modèle complet MC-RNA_{DCO} (10-4-1) versus taux de défaillance par modèle probabiliste (valeur cible).

L'application du modèle complet RNA_{DCO} pour la prédiction du taux de fiabilité de l'élimination de la DCO dans la station d'épuration de Khenchela durant la période d'étude de 4 ans (2009-2012) est présenté sur la figure 79.

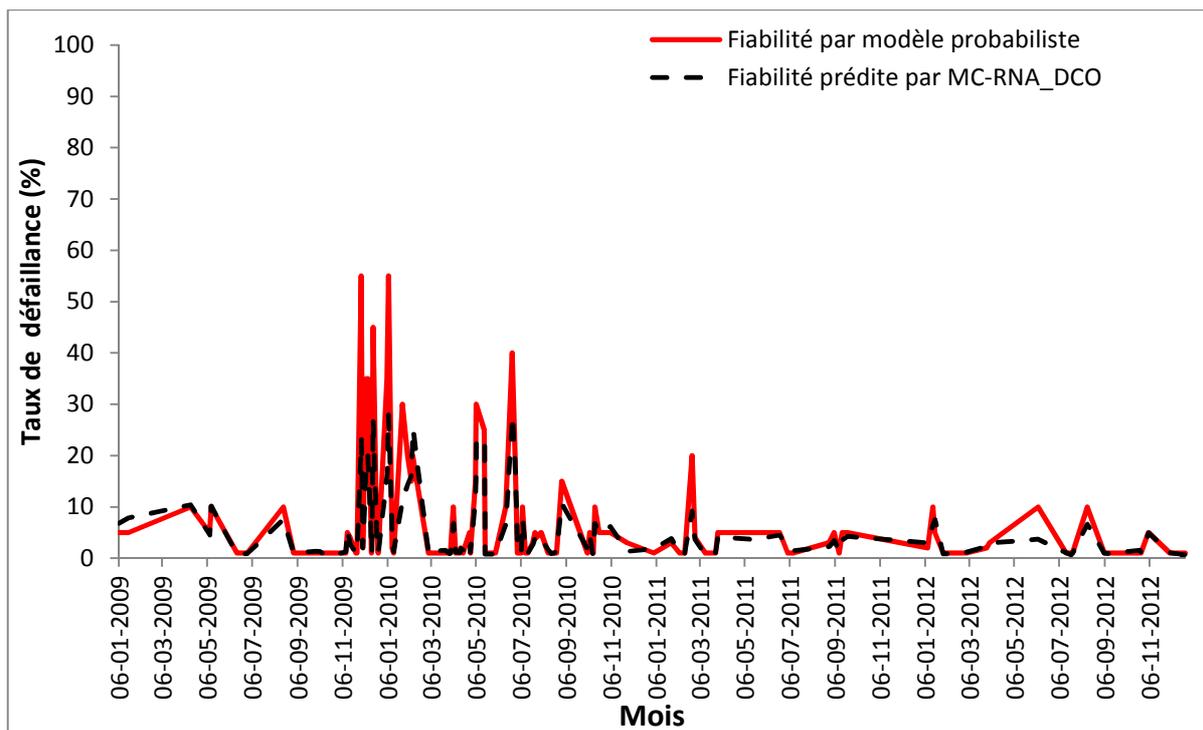


Figure 79 : Taux de défaillance d'élimination de la DCO prédit par réseau de neurones artificiels complet (10-4-1), station d'épuration de Khenchela (2009-2012).

VII.3.3 Prédiction du taux de défaillance d'élimination des MES par modèle complet

Les données d'entrées-sortie du modèle complet MC-RNA_{MES} sont citées dans le tableau 48.

Tableau 48 : Données d'entrées-sortie du modèle complet MC-RNAME_S pour la prédiction du taux de défaillance d'élimination des MES.

Entrées	Sortie
Débit _{entrée} , DBO _{5entrée} , MES _{entrée} , DCO _{entrée}	Taux de défaillance-MES
Débit _{sortie} , DBO _{5sortie} , MES _{sortie} , DCO _{sortie}	
pH _{entrée} , Température _{entrée} , Conductivité _{entrée}	
pH _{sortie} , Température _{sortie} , Conductivité _{sortie}	

Le modèle construit pour prédire le taux de fiabilité de l'élimination des MES de la station d'épuration de Khenchela est constitué de 14 neurones dans la couche d'entrées, 3 neurones dans la couche cachée, et 1 neurone dans la couche de sortie.

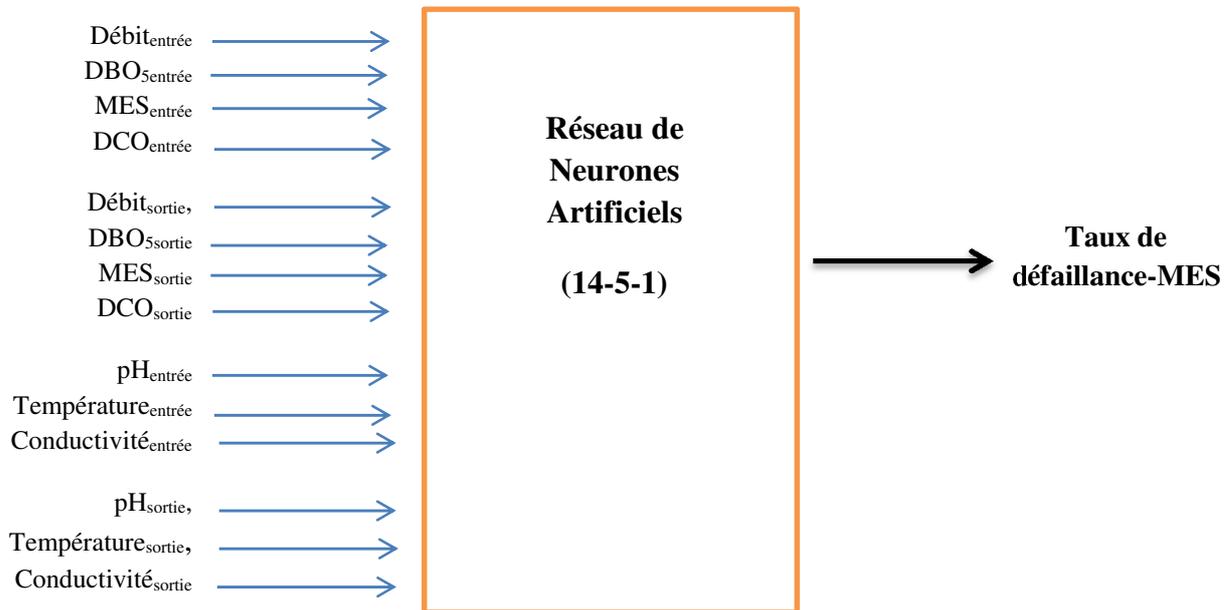


Figure 80 : Schéma générale du modèle complet MC-RNAMES (14-5-1) pour la prédiction du taux de défaillance d'élimination des MES.

Les performances de la prédiction du modèle complet RNA_{MES} sont présentées dans le tableau ci-dessous.

Tableau 49 : Performances de la prédiction du taux de défaillance d'élimination des MES par le modèle complet MC-RNA_{MES}.

	R	R²	Erreur moyenne absolue EMA (%)	Erreur quadratique moyenne EQM (%)
¹ Ensemble de données	0,96	0,92	2,54	5,19
² Données d'apprentissage	0,97	0,94	2,34	4,81
³ Données de validation	0,96	0,92	2,99	5,98

¹ : Ensemble de données (123 données)

² : Données d'apprentissage (86 données)

³ : Données de validation (37 données)

La prédiction du taux de défaillance de l'élimination des MES du procédé de traitement de la station de Khenchela est représentée sur la figure 81.

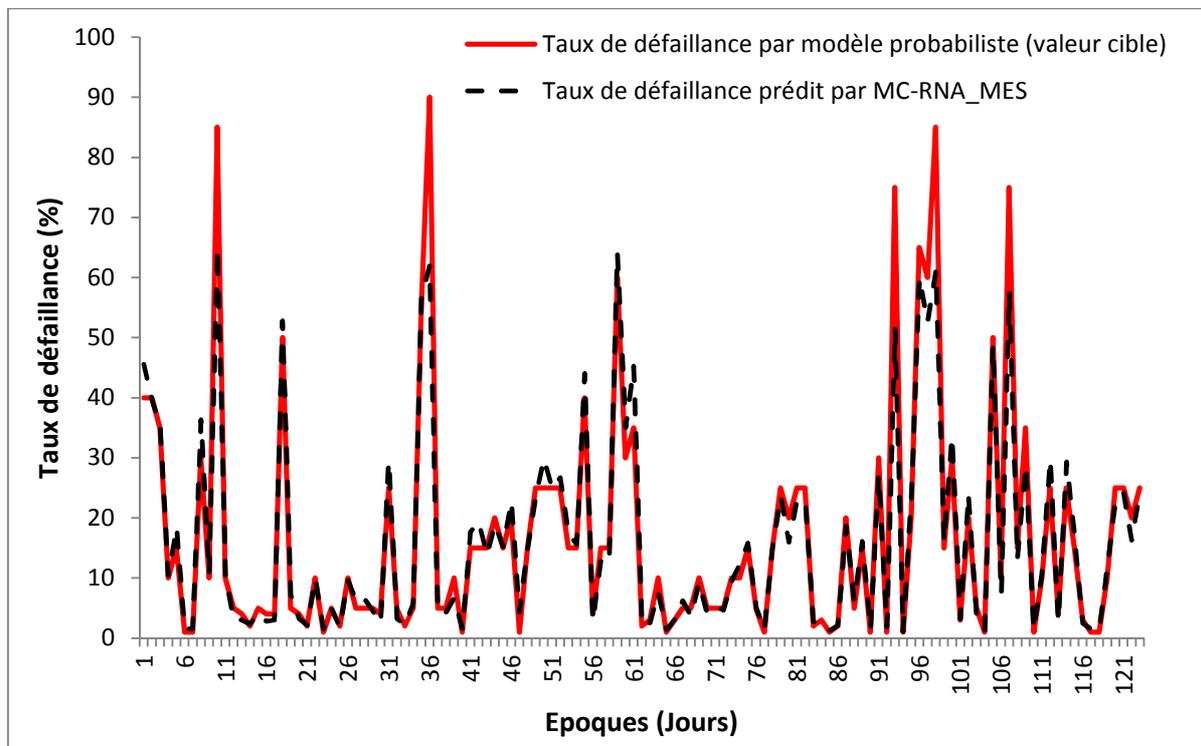


Figure 81 : Taux de défaillance d'élimination des MES prédit par modèle complet MC-RNA_{MES} (14-5-1) versus taux de défaillance par modèle probabiliste (valeur cible).

L'application du modèle complet RNA_{MES} pour la prédiction du taux de défaillance de l'élimination des MES dans la station d'épuration de Khenchela durant une période d'étude de 4 ans (2009-2012) est présenté sur la figure 82.

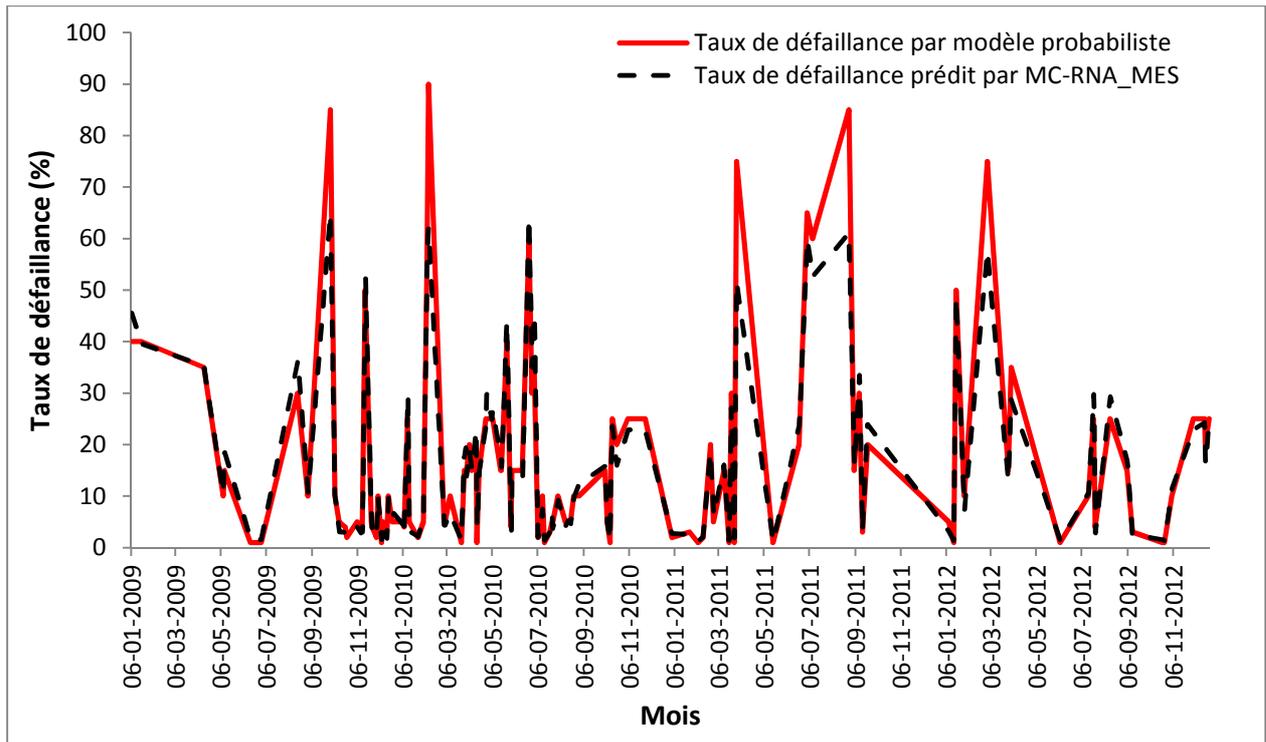


Figure 82 : Taux de défaillance d'élimination des MES prédit par réseau de neurones artificiels simple (14-5-1), station d'épuration de Khenchela (2009-2012).

VII.4 DISCUSSION DES RESULTATS DE LA PREDICTION Des TAUX DE DEFAILLANCE

Six modèles partagés en deux groupes (3 modèles simples et 3 modèles complets) sont présentés dans cette section, où les entrées sont modifiées pour inclure des combinaisons de différents paramètres dans le but de modéliser les variations quantitatives et qualitatives des eaux usées brutes à l'entrée et la sortie de la station d'épuration.

La topologie choisie est le Perceptron multicouche PMC un réseau de type alimentation avant, et avec comme règle d'apprentissage l'algorithme de rétro propagation du gradient d'erreur.

La structure physique tous les réseaux de neurones artificiels construits pour cette étude dispose d'une seule couche cachée. La fonction d'activation de la couche cachée est la tangente hyperbolique (tanh), et la fonction d'activation de la couche de sortie est la fonction sigmoïde.

VII.4.1 Résultats des modèles RNA_{DBO5}

VII.4.1.1 Prédiction du taux de défaillance par modèle simple MS-RNA_{DBO5} (5-2-1)

Le modèle simple MS-RNA_{DBO5} est présenté dans la sous-section VII.2.1. Les données d'entrées-sortie sont présentées dans le tableau 38. Les performances de prédiction du modèle simple sont présentées dans le tableau 39.

Dans la figure 66, une comparaison graphique entre les taux de défaillances prédites par le modèle simple MS-RNA_{DBO5} et les taux de défaillances cible (par modèle probabiliste). En fin une représentation graphique des taux de défaillances fournis par le modèle simple durant une période de 4 ans est illustrée dans la figure 67.

Les résultats du modèle simple MS-RNA_{DBO5} montrent que lors de la (phase d'apprentissage), l'erreur moyenne absolue de l'apprentissage est de 4,17%, l'erreur quadratique moyenne est de 7,52%. Le coefficient de régression $R= 0,94$, et le coefficient de détermination $R^2= 0,88$. Ces paramètres indiquent que le modèle simple RNA_{DBO5} est en mesure de bien répondre aux données d'apprentissage et capable de les rapprocher. Le modèle simple RNA_{DBO5} est donc en mesure de résoudre le problème particulier de cartographie des données entrée-sortie.

Lors de la deuxième étape (phase de test), les résultats du modèle simple RNA_{DBO5} montrent que l'erreur moyenne absolue de la validation est de 4,37%, l'erreur quadratique moyenne est de 7,835%. Le coefficient de régression $R= 0,93$, et le coefficient de détermination $R^2= 0,86$. Ces résultats indiquent que les performances de généralisation du modèle simple RNA_{DBO5} sont bonnes. Le modèle simple MS-RNA_{DBO5} est en mesure de faire des prédictions précises.

VII.4.1.2 Prédiction du taux de défaillance par modèle complet MC-RNA_{DBO5} (10-4-1)

Le modèle complet RNA_{DBO5} est présenté dans la sous-section VII.3.1. Les données d'entrées-sortie sont présentées dans le tableau 44. Les performances de prédiction du modèle complet sont présentées dans le tableau 45.

Dans la figure 75, une comparaison graphique entre les taux de fiabilités prédites par le modèle complet MC-RNA_{DBO5} et les taux de fiabilités cible (par modèle probabiliste). En fin une application des taux de fiabilités prédites est représentée pour la période d'études de 4 ans est représentée dans la figure 76.

Les résultats du modèle complet RNA_{DBO5} montrent que lors de la première étape (phase d'apprentissage), l'erreur moyenne absolue de l'apprentissage est de 4,61%, et l'erreur quadratique moyenne est de 7,42%. Le coefficient de régression $R= 0,93$, et le coefficient de

détermination $R^2= 0,86$. Ces paramètres indiquent que le modèle complet RNA_{DBO5} a bien répondu aux données d'apprentissage et qu'il est capable de les rapprocher.

Lors de la deuxième étape (phase de test), les résultats du modèle complet RNA_{DBO5} montrent que l'erreur moyenne absolue de la validation est de 5,54%, l'erreur quadratique moyenne est de 10,24%. Le coefficient de régression $R= 0,87$, et le coefficient de détermination $R^2= 0,76$. Ces résultats indiquent que le modèle complet RNA_{DBO5} trouve des difficultés de généralisation due à la croissance des données d'entrée et du degré de complexité de l'interaction des paramètres d'entrée ce qui rend la tâche de modéliser la fiabilité du procédé encore plus difficile, malgré cet effet de croissance de la complexité de l'interconnexion entre les données d'entrée, le modèle $MC-RNA_{DBO5}$ est en mesure de faire des prédictions de bonne qualité.

VII.4.2 Résultats des modèles RNA_{DCO}

VII.4.2.1 Prédiction du taux de défaillance par modèle simple $MS-RNA_{DCO}$ (5-8-1)

Le modèle simple RNA_{DCO} est présenté dans la sous-section VII.2.2. Les données d'entrées-sortie sont présentées dans le tableau 40. Les performances de prédiction du modèle simple sont présentées dans le tableau 41.

Dans la figure 69, une comparaison graphique entre les taux de fiabilités prédites par le modèle simple RNA_{DBO5} et les taux de fiabilités cible (par modèle probabiliste). En fin une application des taux de fiabilités prédites est représentée pour la période d'études de 4 ans est représenté dans la figure 70.

Lors de la première étape (phase d'apprentissage) les résultats du modèle simple $MS-RNA_{DCO}$ montrent que l'erreur moyenne absolue de l'apprentissage est de 1,63%, et l'erreur quadratique moyenne est de 3,77%. Le coefficient de régression $R= 0,95$, et le coefficient de détermination $R^2= 0,90$. Ces paramètres indiquent que le modèle simple RNA_{DCO} a bien répondu aux données d'apprentissage et qu'il est capable de les rapprocher. Le modèle simple RNA_{DCO} est en mesure de cartographier les données entrée-sortie.

Les résultats du modèle MS -simple $MS-RNA_{DCO}$ lors de la deuxième étape (phase de test) montrent une légère croissance de l'erreur moyenne absolue de la validation estimée à 1,92%, ainsi que celle de l'erreur quadratique moyenne avoisinant les 5%.

Le coefficient de régression $R= 0,93$, et le coefficient de détermination $R^2= 0,86$. Cette légère détérioration de la performance du modèle est due au fait que les données du set de validation sont trop étendues (l'écart types des données de validation $DCO_{entrée}$ et DCO_{sortie} sont plus importants que ceux des données d'apprentissage). Malgré cet effet, les performances de généralisation du modèle simple sont bonnes, et le modèle $MS-RNA_{DCO}$ est en mesure de faire des prédictions précises.

VII.4.2.2 Prédiction du taux de défaillance par modèle complet MC-RNA_{DCO} (10-4-1)

Le modèle complet MC-RNA_{DBO5} est présenté dans la sous-section VII.3.2. Les données d'entrées-sortie sont présentées dans le tableau 46. Les performances de prédiction du modèle complet sont présentées dans le tableau 47.

Une étude comparative entre les taux de fiabilités prédites par le modèle complet MC-RNA_{DCO} et les taux de fiabilités cible (par modèle probabiliste) est représentée sur la figure 78. La figure 79 représente une application du modèle de prédiction des taux de fiabilité durant une période d'études de 4 ans.

Les résultats du modèle complet MC-RNA_{DCO} montrent que lors de la première étape (phase d'apprentissage), l'erreur moyenne absolue de l'apprentissage est minimal estimé à 2,83%, alors que l'erreur quadratique moyenne est proche de 7,5%. Le coefficient de régression $R=0,90$, et le coefficient de détermination $R^2=0,81$. Ces paramètres indiquent que le modèle complet RNA_{DCO} est capable de les rapprocher des données d'apprentissage.

Par contre, lors de la deuxième étape (phase de test), les résultats du modèle complet RNA_{DCO} on constate que l'erreur moyenne absolue de la validation a évolué. L'erreur quadratique moyenne dépasse les 8%. Le coefficient de régression $R=0,69$, et la valeur du coefficient de détermination (R^2) décroît considérablement à 0,47 ce qui signifie que l'ajustement entre les valeurs prédites et les valeurs cibles est médiocre. Ces résultats indiquent que le modèle complet MC-RNA_{DCO} trouve des difficultés de généralisation due à la difficulté de modéliser l'interaction entre les différentes données d'entrée. Malgré l'effet de la complexité de l'interconnexion entre les données d'entrée et la croissance des données d'entrée, le modèle MC-RNA_{DBO5} fourni des prédictions de qualité acceptable.

VII.4.3 Résultats des modèles RNA_{MES}

VII.4.3.1 Prédiction du taux de défaillance par modèle simple MS-RNA_{MES} (4-3-1)

Le modèle simple RNA_{MES} est présenté dans la sous-section VII.2.3. Les données d'entrées-sortie sont présentées dans le tableau 42. Les performances de prédiction du modèle simple sont présentées dans le tableau 43.

Une comparaison entre les taux de fiabilités prédites par le modèle simple MS-RNA_{MES} et les taux de fiabilités cibles (par modèle probabiliste) est présentée dans la figure 72. En fin sur la figure 73 une présentation graphique a pour objet la prédiction des taux de fiabilités comparés aux taux de fiabilité par modèle probabiliste durant une période de 4 ans.

Lors de son apprentissage le modèle simple RNA_{MES} affiche une erreur moyenne absolue de l'ordre de 2,5%, et d'une erreur quadratique moyenne inférieur à 5%.

Une très bonne valeur du coefficient de régression R égale à 0,97, et un coefficient de détermination R^2 proche de 0,94. Les paramètres de performances du modèle simple RNA_{MES}

indiquent que le modèle simple MS-RNA_{MES} est capable de cartographier les données entrée-sortie des données d'apprentissage.

Les résultats du modèle simple MS-RNA_{MES} en phase de validation montrent que les performances de généralisation du modèle simple sont très bonnes, et que ce modèle est en mesure de fournir des prédictions précises.

VII.4.3.2 Prédiction du taux de défaillance par modèle complet MC-RNA_{MES} (14-5-1)

Le modèle complet MC-RNA_{MES} est présenté dans la sous-section VII.3.3. Les données d'entrées-sortie sont présentées dans le tableau 48. Les performances de prédiction du modèle complet sont présentées dans le tableau 49. Une étude comparative entre les taux de fiabilités prédites par le modèle complet MC-RNA_{MES} et les taux de fiabilités cibles (par modèle probabiliste) sont représentées sur les figure 81.

La figure 82 est la représentation graphique des données fournies par le modèle complet de prédiction durant une période de 4 ans, comparés avec les taux de fiabilités déterminé par modèle probabiliste durant la même durée.

Les résultats du modèles complets MC-RNA_{MES1,2} montrent que lors de la phase d'apprentissage, l'erreur moyenne absolue est minimal estimé à 2,53% pour le deuxième modèle 2,83%, alors qu'elle est inférieure à 2% pour le premier modèle. L'erreur quadratique moyenne des deux modèles est inférieure à 5%. Les deux modèles ont presque le même coefficient de régression R proche de 0,98, quant au coefficient de détermination des deux modèles (R^2), il est proche de 0,95. Ces paramètres indiquent que le modèle complet RNA_{Dco} est capable de les rapprochés les données d'apprentissage.

Les résultats des deux modèles complets MC-RNA_{MES1,2} durant la étape de validation affichent des erreurs moyennes absolues inférieure à 3%. Par contre les erreurs quadratiques moyennes sont légèrement supérieures à 5%. Les coefficients de régression R sont de l'ordre de 0,97, et les coefficients de détermination (R^2) proche de 0,95 ce qui signifie que les ajustements entre les valeurs prédites et les valeurs cibles sont de très bonne qualité. Ces résultats indiquent que les deux modèles arrivent à généraliser, et sont en mesure de réaliser des prédictions précises.

VII.5 CONCLUSION

L'application des modèles simples pour la prédiction des performances d'une station d'épuration présentent un intérêt spécial vue leurs simplicité et leurs facilité d'utilisation pour l'estimation des taux de fiabilité quotidienne du procédé de traitement. Ces résultats peuvent être un très bon indicateur aux exploitants pour servir de support d'aide à la gestion et l'exploitation des différents compartiments du procédé de traitement.

L'application des modèles complets pour la prédiction des taux de défaillances dans le cas de la station de Khenchela présentent des résultats encourageant et constituent une alternative au modèle probabiliste basés sur l'analyse des statistiques des données historiques d'auto-surveillances.

Ces modèles complets sont des outils de modélisation prenant en charge tous les paramètres susceptibles d'influencer par leurs variations qualitatives et quantitatives le fonctionnement du procédé de traitement de la station d'épuration.

CONCLUSION GENERALE ET ORIENTATIONS FUTURES DE LA RECHERCHE

Conclusion

L'analyse statistique des données rassemblées montre que les valeurs des paramètres de pollution diffèrent de manière significative pour les eaux usées à l'entrée de la station s'étend entre 33-1650 mg/l pour la DCO, 70-1332 mg/l pour la DBO₅, 33-1162 mg/l pour les MES.

A la sortie les moyennes des concentrations des paramètres de pollution s'étend entre 13-165 mg/l pour la DCO, 10-128 mg/l pour la DBO₅, 10-452 mg/l pour les MES.

Une approche simple consisté à évaluer les niveaux de performances globale de la station d'épuration de la ville de Khenchela on se basant sur une comparaison des concentrations des différents paramètres et leurs conformités avec la législation en vigueur, on constate que durant la période considéré la station d'épuration présente des taux de fiabilité acceptable.

La première contribution de cette étude été l'évaluation des performances de la station d'épuration de la ville de Khenchela par modèle probabiliste basé sur les travaux de Niku et al. (1979). nous a permis d'estimer le taux de fiabilité des paramètres de pollution on se basant sur la loi de distribution de ces paramètres, tous les paramètres à la sortie de la station suivent une loi log-normale.

Les performances du procédé de traitement de la station d'épuration de Khenchela par modèle probabiliste affichent des résultats de très bons niveaux d'efficacité. Des taux de fiabilité estimés pour les trois paramètres de pollution :

- 94,33% pour l'élimination de la DCO,
- 82,47% pour la dégradation de la DBO₅,
- 85,79% pour l'élimination des MES.

Ce constat positif est entaché par l'absence des données, ce qui rend ces taux de fiabilité relatif à la période considéré seulement, en générale sur les quatre années d'étude (2009-2012) les taux de fiabilité se dégrade et deviennent obsolète, due à la non maîtrise du procédé de traitement, ce manque de performance est le résultat de l'interférence de plusieurs facteurs qui feront l'objet d'autres études complémentaires.

Les fluctuations des taux de fiabilité du procédé de traitement de la station d'épuration est très importante, influencé par la variabilité quantitatives et surtout qualitatives des eaux usées brutes combiné avec une conception inadéquate avec des hypothèses de conception qui ne reflète pas la situation réelle de la station (station à faible charge). Une étude récente à

montrer que la station fonctionne à forte charge. Le manque d'expérience des gestionnaires de la station est un facteur important à prendre en considération.

La variabilité des taux de défaillance des différents composants du système de traitement donne un aperçu général du comportement instable de la station d'épuration, malheureusement l'absence de note de gestion rend toute tentative d'explication de l'instabilité des performances de la station une tâche impossible.

La nécessité d'un contrôle efficace d'une station de traitement des eaux usées utilisant le procédé de boues activées est primordiale pour assurer son efficacité dans la protection de l'environnement.

Le recours à l'application des réseaux de neurones artificiels qui sont une forme d'intelligence artificielle pour tenter de modéliser le comportement d'un procédé de traitement des eaux usées par l'utilisation d'un ensemble d'équations non linéaires pour imiter les connexions neuronales des systèmes biologiques pour la prédictions des performances d'une station de traitement des eaux usées.

Les résultats de l'application de réseau de neurones artificiels présentés dans la deuxième contribution montrent que les six modèles de réseaux de neurones artificiels proposés ont présenté un comportement de prévision précis ou de très bonne qualité. Certains types de réseaux de neurones sont cependant adaptés à certains types de problèmes et obtiennent de meilleurs résultats que d'autres (Haykin , 1999)

Les entrées des différents modèles RNA sont modifiées pour inclure l'effet des différents paramètres sur les performances du procédé de traitement et tenter de modéliser les différents facteurs influençant le fonctionnement de la station.

Le premier ensemble des résultats présentés dans les sous-sections VI.4.1 et VI.5.1 concernent les deux modèles pour la prédiction du taux de fiabilité de la dégradation de la DBO5 montrent qu'en phase d'apprentissage, les coefficients de régression (R) varie sélectivement de 0,94 à 0,93, et que les coefficients de détermination (R^2) varie de 0,88 à 0,86. Ces paramètres indique que les deux modèles RNA_{DBO5} sont en mesure de bien répondre aux données d'apprentissage et capable de les rapprochés. En phase de validation les deux modèles ont des coefficients de régression (R) varie sélectivement de 0,93 à 0,87, et que les coefficients de détermination (R^2) varie de 0,886 à 0,76, des résultats qui indiquent que les

performances de généralisation des deux modèles sont bonnes, et qu'ils sont en mesure de faire respectivement des prédictions précises et de bonne qualité.

Le deuxième ensemble des résultats présentés dans les sous-sections VI.4.2 et VI.5.2 concernent les deux modèles pour la prédiction du taux de fiabilité de l'élimination de la DCO montrent qu'en phase d'apprentissage, les coefficients de régression (R) varient sélectivement de 0,95 à 0,90, et que les coefficients de détermination (R^2) varient de 0,90 à 0,81. Ces paramètres indiquent que les deux modèles RNA_{DCO} sont en mesure de bien répondre aux données d'apprentissage et sont en mesure de cartographier les données entrée-sortie. En phase de validation les deux modèles ont des coefficients de régression (R) varient sélectivement de 0,93 à 0,69, et que les coefficients de détermination (R^2) varient de 0,86 à 0,47, des résultats qui indiquent que les performances de généralisation du modèle simple sont bonnes, par contre celle du modèle complet sont moins bonnes.

Le fait que le modèle complet $MC-RNA_{DCO}$ ne parvient pas à résoudre le problème de généralisation peut être expliqué par l'une des raisons suivantes:

- Le problème est non-déterministe et le modèle complet $MC-RNA_{DCO}$ ne sera jamais le résoudre;
- Les données d'entrées ne sont pas suffisamment représentatives du processus.

Ces deux raisons selon Masters (1993) sont les deux seuls à son avis, responsables du non succès d'un RNA. La première raison est l'option la plus probable et l'expérience de nombreux praticiens des RNA que les applications du monde réel sont les plus difficiles à résoudre.

En outre, la qualité des données de l'installation de traitement des eaux usées peut être un autre facteur à considérer. L'expérience de nombreux exploitants de stations d'épurations et d'ingénieurs a été que, souvent, le matériel de la station échoue, et que les données sur les temps de fonctionnement sont ensuite rapportées à tort, parfois sans le savoir.

Le troisième ensemble des résultats présentés dans les sous-sections VI.4.3 et VI.5.3 concernent les deux modèles RNA_{MES} pour la prédiction du taux de fiabilité de l'élimination des MES montrent que les deux modèles sont en mesure de cartographier les données entrée-sortie et que les performances de généralisation des deux modèles sont très bonnes, et sont en mesure de faire des prédictions précises. Améliorer le système de surveillance et de contrôle d'automatisation de la station de traitement des eaux usées de Khenchela peut apporter une contribution positive à la détection précoce de panne et peut diminuer l'incidence de l'équipement défectueux et le manque de données fiables par rapport aux données rapportées.

L'application des modèles simples pour la prédiction des performances d'une station d'épuration présentent un intérêt spécial vue leurs simplicité et leurs facilité d'utilisation pour l'estimation des taux de défaillances quotidiennes du procédé de traitement. Ces résultats peuvent être un très bon indicateur aux exploitants pour servir de support d'aide à la gestion et l'exploitation des différents compartiments du procédé de traitement. Par contre, l'application des modèles complets pour la prédiction des taux de défaillance représente un intérêt scientifique et académique pour tenter de modéliser des phénomènes non linéaires très complexes qui mettent en contribution un grand nombre de paramètres.

Application des résultats obtenus

Les modèles RNA développés pour la prédiction des taux de fiabilité et taux de défaillances ont les caractéristiques positives suivantes :

- Les structures des RNA sont simples, avec une couche cachée et ne nécessitent pas de calculs complexes;
- Le temps de la formation des réseaux est court;
- Il est capable de résoudre des problèmes non linéaires dans le monde réel.

Les modèles RNA développées peuvent être appliqués dans les directions suivantes :

1. En tant que base pour le développement ultérieur de la théorie pour la prévision des variables influentes sur le fonctionnement de la station d'épuration ;
2. Comme un système expert pour les conseils des opérateurs dans la station de traitement des eaux usées municipales;
3. Comme une boîte à outils en temps réel des systèmes d'auto-surveillances existants dans les stations d'épuration.

Les recherches futures

La recherche a montré que les réseaux neuronaux sont une option fiable à envisager pour la prédiction des performances d'une station de traitement des eaux usées. Toutefois, la mise en œuvre effective peut justifier des recherches supplémentaires. La recherche de la littérature et de l'avis des autres praticiens de réseaux de neurones les plus expérimentés ont indiqué qu'un réseau hybride pourrait être une meilleure solution. Cela pourrait inclure d'autres options de calcul comme un réseau de neurones artificiels à logique floue ou un réseau de neurones artificiels combiné avec un algorithme génétique, qui feront l'objet de futures travaux de recherches, et une continuité du présent travail pour optimiser les modèles construits.

Publications relatives à la thèse

DJEDDOU M., ACHOUR B., et MARTAUD M. (2013), "Determination and analysis of daily reliability level of municipal wastewater treatment plant", Courier du Savoir, N°17, Biskra, Algeria, pp.: 39-46.

DJEDDOU M., ACHOUR B., et MARTAUD M. (2013), "Détermination du niveau quotidien de fiabilité dans une station de traitement des eaux usées municipale", The 4th International Congress Water, Waste & Environment (EDE4), Agadir, Morocco, 8 pages.

DJEDDOU M., ACHOUR B., et MARTAUD M. (2014), "Daily Performances Evaluation of Municipal Wastewater Treatment Plant using Probabilistic Model", accepted for "IWA conference Activated Sludge - 100 Years and Counting", Essen, Germany, 5 pages.

REFERENCES BIBLIOGRAPHIQUES

Adeloye A. J. et De Munari A. (2006), ".Artificial neural network based generalized storage–yield–reliability models using the Levenberg–Marquardt algorithm", *Journal of Hydrology*, 326, (1-4), pp.: 215-230.

Adico U. L. (1997), "Implantation de réseaux de neurones artificiels en S-PLUS. Application : Comparaison de méthodes statistiques dans le choix du nombre de cellules cachées", Mémoire de M. Sc. en Statistique, Université de Montréal, Canada, 155 pages.

Albrecht A. E. (1967), "Aerated Grit Operation, Design and Chamber", *Water and Sewage Works*, vol. 114, no. 9, pp. 331–335.

Alliot J. M. et Durand N. (2005), "Algorithmes génétiques", 34 pages. Disponible sur le site: <http://pom.tls.cena.fr/GA/FAG/ag.pdf>

Anava O., Hazan E. et Mannor S. (2013), "Online Learning for Time Series Prediction", *JMLR: Workshop and Conference Proceedings* vol, pp.: 1-13.

Anderson J. A., Pellionisz A. et Rosenfeld E., (1988), "Neurocomputing: Foundations of Research", The MIT Press, USA, 729 pages.

Asano T. (1998), "Wastewater Reclamation and Reuse", *Water Quality Management Library*, Vol. 10, CRC Press LLC, New York, USA, 1528 pages.

Asano, T. et Levine, A.D. (1995), "Wastewater Reuse: A Valuable Link in Water Resources Management", *Water Quality International*, No. 4, pp.: 20-24.

AWWA (2012), "Water Treatment Plant Design", American Water Works association and American Society of Civil Engineers, Fifth edition, McGraw-Hill, New York, USA, 1376 pages.

Barlow R. E. et Proschan F. (1996), "Mathematical Theory of Reliability", *CLASSICS In Applied Mathematics*, SIAM Society for Industrial and Applied Mathematics, Philadelphia, USA, 277 pages.

Basheer I. A. et Hajmeer M. (2000), "Artificial Neural Networks: Fundamentals, Computing, Design, and Application", *Journal of Microbiological Methods* 43, pp.: 3-31.

Belanche L.B., Valdés J.J., Comas J., Ignasi R Rodab I.R. et Poch M. (1999), "Towards a model of input–output behaviour of wastewater treatment plants using soft computing techniques", *Environmental Modelling & Software* Volume 14, Issue 5, pp.: 409-419.

Baxter C., Lence. B., et Coffey B. (2003), "Analyzing Operational Risk in Potable Water Supply Using Conditional Reliability. *World Water & Environmental Resources Congress 2003*: pp. 1-10.

"Analyzing Operational Risk in Potable Water Supply Using Conditional Reliability", World Water Congress, Pennsylvania, USA.

Berthouex P. M., and fan, Richard, (1986), "Evaluation of Treatment Plant Performance: Causes, Frequency, and Duration of Upsets", Journal Water Pollution Control Federation, Vol. 58, No. 5, pp. 368-375.

Bhat N.V., Minderman P.A., Mcavoy T., et Wang N.S. (1990), "Modeling chemical process systems via neural computation", IEEE Control Systems Magazine 10, Issue 3, pp.: 24 - 30.

Bishop C. M. (1995), "Neural Networks for Pattern Recognition", Oxford University Press, Oxford, UK, 482 pages.

Blumenthal U. J., (1989), "Generalized model of the effect of different control measures in reducing health risks from waste reuse", Water Sci. Tech. 21:567-577.

Burmaster D. E. et Hull D. A., (1997), "Using Lognormal distributions and Lognormal Probability Plots In Probabilistic Risk Assessments", Human and Ecological Risk assessment, Vol. 3, No. 2, pp. 235-255.

Burton M. (2006), "Neural Networks, Course Notes", Rhodes University, RAS, 64 pages.

Carlsson B. et Lindberg C.F. (2004), "Some control strategies for the activated sludge process" Systems and Control Group Uppsala University, Sweden, 22 pages.

Cauchi, H ; Nakache S. ; Zagury B. et al., (1996), "Dossier : La réutilisation des eaux usées après épuration", TSM Techniques, Sciences et Méthodes, N°2, pp. : 81-118.

Chang F. Liang J., et Chen Y. (2001), "Flood forecasting using Radial Basis Function Neural Networks", IEEE Transactions on Systems, Man and Cybernetics, Part C Applications and Reviews, Volume 31 No. 4, pp.: 530 – 535.

Chen J. C., Chang N. B. et Shief W. K. (2003), "Assessing wastewater reclamation potential by neural network model", Engineering Applications of Artificial Intelligence Volume 16, Issue 2, pp.: 149-157.

Choi D. J. et Park H. (2001), "A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process", Water Research Volume 35, Issue 16, pp.: 3959-3967.

Cohen A. I., Bar-Shalom Y. W., et al., (1975), "A Quantitative Method for Effluent Compliance Monitoring Recourses Allocation", EPA-600/5-75-015, U.S. Environmental Protection Agency, Washington, USA, 47 pages. Disponible sur le site:

[http://yosemite.epa.gov/ee/epa/erm.nsf/vwAN/EE-0378_01.pdf/\\$file/EE-0378_01.pdf](http://yosemite.epa.gov/ee/epa/erm.nsf/vwAN/EE-0378_01.pdf/$file/EE-0378_01.pdf)

Coley D. A. (1999), "An Introduction to Genetic Algorithms for Scientists and Engineers", World Scientific Publishing Co. Pte. Ltd., Singapore, 244 pages.

Converse J. C. et Nordheim E. V. (2004), "Sampling Protocol for Evaluating Aeration Systems for Their Ability to Consistently Produce an Effluent to a prescribed Parameter Value", 10th, National symposium on individual and small community sewage systems; On-site wastewater treatment X, pp.: 343-348.

Cote M., Grandjean B. P., Lessard P. et Yhibault J. (1995), "Dynamic modeling of the activated sludge process: improving prediction using neural networks", *Water Research*, 29, pp.: 995-1004.

Culp G., Wesner G., Williams, R. et al., (1980), "Wastewater Reuse and Recycling Technology", Noyes Data Corporation, New Jersey, USA, 838 pages.

D'Agostino R. B., and Stephens M A., (1986), "Goodness-of-fit Techniques", MARCEL DEKKER, Inc, New York, USA, 576 pages.

Dauks P. (2006), "Design Considerations for Screening", Presented at Michigan Water Environment Association Process Seminar, GrandRapids, Michigan, 12 pages.

Dawkins R. (1986), "The Blind Watchmaker", W. W. Norton & Company, USA, 345 pages.

Demuth H., Beale M. et Hagan M. (2013). Neural Network Toolbox For Use with MATLAB, Users Guid, R2013, 430 pages.

Dhillon B. S., (1999), "Design Reliability: Fundamentals and Applications", CRC press, New York, USA, 480 pages.

Drakatos P. A., Lyon R. H., Chyssolouris G., et al., (1992), "Reliability and Exploitation of Wastewater Treatment Plants Using Diagnostic Methods", *International Journal of Environmental Studies*, Vol. 40, pp. 267-280.

Dufournet R. (1974), "Traitement des eaux usées", *Techniques de l'ingénieur*, C670, Editions T. I., Paris, France, 38 pages.

Eisenberge D., Soller J., Sakaji R., et al., (2001), "A Methodology to Evaluate Water and Wastewater Treatment Plant Reliability", *Water Science and Technology*, Vol. 43, No. 10, pp. 91-99.

Ellis G. W., Grasso D., et Ge X., (1993), "ARMA Processes and Reliability-Based Design of Wastewater-Treatment Facilities", *Journal of Environmental Engineering*, Vol. 119, No. 3, pp. 463-477.

Elman J. L. (1990), "Finding Structure in Time", *Cognitive Science*, Vol. 14, Issue 3, pp.: 179-211.

Etnier C., Willetts J., Mitchell C. A. et al., (2005), "Decentralized Wastewater System Reliability Analysis Handbook". Project No. WU-HT-03-57. Prepared for the National Decentralized Water Resources Capacity Development Project, Washington University, St. Louis, MO, by Stone Environmental, Inc., Montpelier, VT, USA, 181 pages.

Faby J. A. et Brissaud F., (1997), "L'utilisation des eaux usées épurées en irrigation", Office International de l'Eau, France, 76 pages.

- Fahlman S. E. et Lebiere C. (1989)**, "The Cascade-Correlation learning architecture", Computer Science Department. Paper 1938, 17 pages. Disponible sur le site:
- Fair, G. M. et Geyer J. C. (1954)**, "Water Supply and Waste-Water Disposal", John Wiley & Sons, New York, USA, 611 pages.
- Funahashi K. (1989)**, "On the approximate realization of continuous mappings by neural networks", Neural Networks, n°2, 1989, pp.: 183-192.
- Funahashi S., Bruce C. J. et Goldman-Rakic P. S. (1989)**, "Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex", J Neurophysiol, 61 (2), pp.: 31-49.
- Gaïd A. (1993)**, "Traitement des eaux usées urbaines", Techniques de l'ingénieur, C5220, Editions T. I., Paris, France, 28 pages.
- Garbus, R. O. (2006)**, "Types of Pumps" in G. M. Jones, (ed.), "Pump Station Design", Elsevier, New York, pp. 1137–1139.
- Garvey E.B. (1997)**, "On-line Quality Control of Injection Molding Using Neural Networks. Minor Thesis Department of Computer Science, Royal Melbourne Institute of Technology, Melbourne, Australia, 82 pages
- Gershenfeld N. A. et Andreas S. Weigend S., (1993)**, "The Future of Time Series", SFI WORKING PAPER: 1993-08-053, Santa Fe Institute, USA, 89 pages.
- Goldberg D.E. (1989)**, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley Publishing Compagny, Inc, USA, 412 pages.
- Govindaraju R.S. (2000)**, "Artificial neural network in hydrology. II: hydrologic application, ASCE task committee application of artificial neural networks in hydrology", Journal of Hydrologic Engineering 5, pp.: 124-137.
- Graupe D. (2007)**, "Principal of Artificial Neural Networks, Advanced Seris In Circuits And Systems", Vol. 6, World Scientific Publishing Co. Pte. Ltd., 2nd Edition, Singapore, Thaïlande, 320 pages.
- Graves A., Liwicki M., Fernandez S., et al., (2009)**, "A Novel Connectionist System for Improved Unconstrained Handwriting Recognition.", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, N° 5, pp.: 855-868.
- Groves T. W., Bowers F., Corriveau J., et al., (2005)**, "Variability and Reliability of Test Center and Field Data: Definition of Proven Technology from a Regulatory Viewpoint, Project No. WU-HT-03-35. Prepared for the National Decentralized Water Resources Capacity Development Project, Washington University, St. Louis, MO, by the New England Interstate Water Pollution Control Commission, Lowell, MA, USA, 175 pages.
- Guilloteau J. A., (1992)**, "Traitement des eaux résiduaires par infiltration percolation", Thèse de doctorat, Université Louis Pastur, Strasbourg, France, 170 pages.

- Häck M. et Köhne M. (1996)**, "Estimation of wastewater process parameters using neural networks", *Water Science and Technology* Volume 33, Issue 1, pp.: 101-115
- Hamed M., Khalafallah M.G. et Hassanein E.A. (2004)**, "Prediction of wastewater treatment plant performance using artificial neural network", *Environmental Modeling and Software*, 19, pp.: 919-928.
- Hamoda M.F., Al-Gusain I.A. et Hassan A.H. (1999)**, "Integrated wastewater treatment plant performance evaluation using artificial neural network", *Water Science and Technology*, 40, pp.:55-69.
- Haykin S. (1999)**, *Neural Networks A Comprehensive Foundation*, Second Edition, Pearson Education, Inc., India, 823 pages.
- Heaton J. (2008)**, "Introduction to Neural Networks for C#", 2nd Edition, Heaton Research, Inc., Saint Louis, USA, 430 pages.
- Holland J. H. (1992)**, "Adaptation in Natural and Artificial Systems", The Mit Press, USA, 245 pages.
- Hornik K., Stinchombe M., et White H. (1990)**, "Universal approximation of an unknown mapping and its derivatives using multilayer feed forward networks", *Neural Networks*, Volume 3, Issue 5, pp.: 551-560. <http://repository.cmu.edu/compsci/1938>
- Hu J. (1997)**, "Research on Hybrid Black-Box Modeling for Nonlinear Systems and Its Applications", thèse de Doctorat, Waseda university, Japan, 139 pages.
- Hudson J. (1986)**, "Forecasting Onsite Soil Absorption System Failure Rates", EPA/600/S2-86-060, OH, USA, 2 pages.
- <http://nepis.epa.gov/Exe/ZyPDF.cgi/2000TL6K.PDF?Dockey=2000TL6K.PDF>
- Hydro Projet Est, (2005)**, " Etude de la station d'épuration des eaux usées de la ville de khenchela wilaya de Khenchela. Phase III. Avant projet détaillé", 100 pages.
- Imhoff, K. et Imhoff K. R. (2007)**, "Taschbuch der stadtentwässerung", [Measurement of Liquid Flow in Open Channels—Parshall and Sanitary Flumes], Ooldenbourg industrie verlag München, Germany, 508 pages.
- Isa I S., Saad Z., Omar S. et al., (2010)**, "Suitable MLP Network Activation Functions for Breast Cancer and Thyroid Disease Detection", Second International Conference on Computational Intelligence, Modelling and Simulation, IEEE Computer Society, pp.: 39-44.
- Jain A.K., Mao J. et Mohiuddin K.M. (1996)**, "Artificial Neural Networks: A Tutorial", *IEEE Computer*, pp.: 31-44.
- Joarder K. et Aziz S. M. (2002)**, "A note on activation function in multilayer feedforward learning", *Proceedings of International Joint Conference on Neural Networks*", IJCNN 02, Vol. 1, pp.: 519-523.

JORA (2006), Décret exécutif n°06-141 du 19 avril 2006 définissant les valeurs limites des rejets d'effluents liquides industriels, Journal Officielle de la République Algérienne N°26 du 23 avril 2006.

Kahn H. D., et Marvin B. R., (1989), "Use of Statistical Methods in Industrial Water Pollution Control Regulations in The United States", Environmental Monitoring and Assessment, Vol. 12, No. 2-3, pp.: 129-148.

Khan, R et Ondrůšek Č. (2000), "Peak electric load forecasting using an artificial neural network", submitted for publishing in Engineering Mechanics International Journal of Theoretical and Applied Mechanics, Technical University of Brno, Czech Republic.

Kottegoda N. T., et Rosso R., (2009), "Statistic, Probability, and Reliability for Civil and Environmental Engineers", John Wiley & Sons, New York, USA, 736 pages.

Krenker A., Bešter J. et Kos A. (2011), "Artificial Neural Networks-Methodological Advances and Biomedical Applications", Publisher InTech, Shanghai, China, 362 pages.

Kruger C. (2007), "Prediction of the influent wastewater variables Using Neural Network theory", thèse de Doctorat, Cape Peninsula University of Technology, RSA, 331 pages.

Kröse B. et Van Der Smagt P. (1996), "An Introduction to Neural Networks", 8th edition, The University of Amsterdam, Netherland, 135 pages.

Lazarova V., Gaïd A., Rodriguez-Gonzales J. et al., (2003), "L'intérêt de la réutilisation des eaux usées: Analyses d'exemples mondiaux", TSM Techniques, Sciences et Méthodes, génie urbain génie rural, n° 9, pp.: 64-85.

Le Cun Y. (1986), "Learning Processes in an Asymmetric Threshold Network", In Bienenstock, E., Fogelman-Souli, F and Weisbuch, G., eds. Disordered Systems and Biological Organization. NATO ASI Series, F20, Berlin: Springer-Verlag, 8 pages.
<http://yann.lecun.com/exdb/publis/pdf/lecun-86.pdf>

LeCun Y., Bottou L., Orr G. B., et Müller K. R. (1998), "Neural networks: tricks of the trade", Springer Lecture Notes in Computer Sciences, Springer Heidelberg, pp.: 5-50.

Lee D.S. et Park J.M. (1999), "Neural network modeling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor", Journal of Biotechnology 75, pp.: 229-239.

Leitao R. C., Haandel, A. C., Zeeman G., et al., (2006), "The Effects of Operational and Environmental Variations on Anaerobic Wastewater Treatment Systems: A review", Bioresource Technology, Vol. 79, pp.: 1105-1118.

Levin R. I., (1984), "Statistics for Management," 2nd editions, Prentice-Hall Inc. New Jersey, USA, 866 pages.

Loftis J. C., Ward R. C., et Smilli G. M., (1983), "Statistical Models for Water Quality Regulation", Journal Water Pollution Control Federation, Vol. 55, No. 8, pp.: 1098-1104.

- Maier H.R. et Dandy G.C. (2000)**, "Neural networks for prediction and forecasting of water resources variables: a review of modeling issues and applications", *Water Resources Research* 15, pp.:101-124.
- Malasri K. et Malasri S. (2002)**, "Back propagation Networks for Time Series Forecasting: Case Studies in Data Modeling", Christian Brothers University, Texas, USA, Proceedings of the MAESC 2002 Conference, 9 pages. <http://www.maesc.org/maesc02/FullPapers/a2-1-doc.pdf>
- Mandic D. P. et Chambers J. A. (2001)**, "Recurrent Neural Networks for Prediction", John Wiley & Sons, Ltd., UK, 295 pages.
- Masters T. (1993)**, "Practical Neural Network Recipes in C++", San Diego Academic Press, USA, 493 pages.
- McCulloch W. S. et Pitts W. (1943)**, "A logical calculus of ideas immanent in nervous activity". *Bulletin of Mathematical Biophysics* 5, pp.: 115-133. Reprinted in Anderson & Rosenfeld (1988).
- Medkour M. (2003)**, "Réutilisation des eaux usées épurées", Séminaire sur le secteur de l'eau en Algérie. Ministère des Ressources en eau, Alger, Algérie. 12 pages.
- Metahri M. S. (2012)**, "Élimination simultanée de la pollution azotée et phosphatée des eaux usées traités, par des procédés mixtes. Cas de la STEP Est de la ville de Tizi-Ouzou", thèse de doctorat, université de Tizi-Ouzou, Algérie, 172 pages.
- Metcalf & Eddy (2003)**, "Wastewater Engineering: Collection, Treatment, Disposal", 4th ed., McGraw-Hill, New York, USA, 1819 pages.
- mfaqs (2012)**, "comp.ai.neural-nets FAQ, Part 2 of 7: Learning", consultable sur: <http://www.faqs.org/faqs/ai-faq/neural-nets/part2/section-10.html>
- Minsky M.L., et Papert S. A. (1969/1988)**, "Perceptions, An Introduction to Computational Geometry", The MIT Press (first edition, 1969; expanded edition, Cambridge, MA, USA, 292 pages.
- Mjalli F., Alasheh S. , et Alfadala H.E. (2007)**, "Use of artificial neural network black-box modelling for the prediction of wastewater treatment plants performance", *Journal Of Environment Management*, 83, pp.: 329-338.
- Modarres R., Gastwirth J. L., et Ewens W., (2004)**, "A Cautionary Note on the Use of Non-parametric Tests in the Analysis of Environmental Data", *Environmetrics*, Vol. 16, Issue 4, pp. 319-326.
- Moody J. et Darken C. J. (1989)**, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, 1, pp.: 281-294.
- Morales L. et Reinhart D. (1984)**, "Full-Scale Evaluation of Aerated Grit Chambers", *Journal of Water Pollution Control Federation*, Vol. 56, no. 4, pp.: 337–343.

Moubray J. (1997), "Reliability-Centered Maintenance". 2nd ed. Industrial Press, New York, USA, 448 pages.

MRE, (2014), "Synthèse sur la situation de l'assainissement et évaluation des besoins en station de traitement et d'épuration en vue de protéger les ressources en eau", Séminaire sur le secteur de l'eau en Algérie. Ministère des Ressources en eau, Alger, Algérie, 11 pages.

Narendra K.S. et Parthasarathy K. (1990), "Identification and Control of Dynamical Systems using Neural Networks", IEEE Transactions on Neural Networks. Vol.1, pp.: 7-27.

Nasr M. S., Medhat A.E. Moustafa M. A. E., Seif H. A. E., El Kobrosy G., (2012), "Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT, [Alexandria Engineering Journal, 51, pp: 37-43](#)

Neelakantan T.R., Brion T.R. et Lingireddy S. (2001), "Neural network modeling of cryptosporidium and giardia concentrations in Delaware River, USA", Water Science and Technology 43, pp.:125-132.

Ng G. W. (1997), "Application of Neural Networks to Adaptive Control of Nonlinear Systems", Research Studies Press Ltd, John Wiley and Sons Inc, .Singapore, 198 pages.

Niku S., et Schroeder, E. D., (1981), "Stability of Activated Sludge Processes Based on Statistical Measures", Journal Water Pollution Control Federation, Vol. 53, No. 4, pp.: 129-143.

Niku S., Schroeder, E. D., et Haugh. R. S., (1982), "Reliability and Stability of Trickling Filter Processes", Journal Water Pollution Control Federation, Vol. 54, No. 2, pp.: 457-470.

Niku S., Schroeder, E. D., et Samango F. J., (1979), "Performance of Activated Sludge Processes and Reliability-Based Design", Journal Water Pollution Control Federation, Vol.51, No. 12, pp. 2841-2857.

Olden J. D. et Jackson D. A. (2002), "Illuminating the 'black box': a randomization approach for understanding variable contributions in artificial neural networks", Ecological Modelling, 154, pp.: 135-150.

Oliveira-Esquerre K.P., Mori M. et Bruns R.E. (2002), "Simulation of an industrial wastewater treatment plant using artificial neural networks and principal components analysis. Brazilian Journal of Chemical Engineering, 19, pp.: 365-370.

ONA (2014), "L'ONA en chiffres, disponible sur le site: <http://www.ona-dz.org/L-ONA-en-chiffres.html>

Ossenbruggen P. J., Constantine K., Collins, M. R., et al., (1987), "Toward Optimum Control of the Activated Sludge Process with Reliability Analysis", Civil Engineering Systems, Vol. 4, No. 2, pp. 77-86.

- Papliński, A. P. (2004)**, "Lecture notes for Course 5301, Neural Networks and Fuzzy Systems (Neuro-Fuzzy Computing), Clayton School of Information Technology, Monash University, Australia. <http://www.csse.monash.edu.au/courseware/cse5301/2006/>
- Parizeau M. (2006)**, "Réseaux de neurones GIF-21140 et GIF-64326", Université de Laval, Canada, 116 pages.
- Park M. H. (1996)**, "Neural Network Control of a Chlorine Basin", University of California, USA, pp.:20-52.
- Parker D. (1985)**, "Learning Logic", Technical Report TR-87, Center for Computational Resesarch in Economics and Management Science MIT, Cambridge, MA, USA, 47 pages.
- PDAU, (2006)**, "Plan directeur d'aménagement et d'urbanisme Inter-communale de Khenchela".
- Piekniowski F. et Tybicki L. (2004)**, "Visual comparison of performance for different activation functions in MLP networks", Proceedings of International Joint Conference on Neural Networks, IJCNN '04, Vol. 4, pp.: 2947-2952.
- Plazl I., Pipus G., Drolka M. et Koloini T. (1999)**, "Parametric sensitivity and evaluation of a dynamic model for single-stage wastewater treatment plant. Acta Chimica Slovenica 42, pp.:289-300.
- Prevost L. (2012)**, "Modèles Connexionnistes Apprentissage Fusion d'informations". 30 pages. site: <http://www.isir.upmc.fr/UserFiles/File/LPrevost/connex%201%20&%202.pdf>
- Racault Y. et al., (1997)**, "Le lagunage naturel, Les leçons tirées de 15 ans de pratique en France", CEMAGREF éditions, Agence de l'eau Loire-Bretagne, France, 64 pages.
- Radoux, M., (1989)**, "Épuration des eaux usées par Hydrosère reconstituée". Tribune de l'eau, Vol.42, N°8, pp.: 62-68.
- Rech, G. (2002)**, "Forecasting with artificial neural networks", SSE/EFI Working Paper Series in Economics and Finance, No 49, 38 pages.
- Reed, R. D. et Marks R. J II (1999)**, "Neural Smthing: Supervised Learning in Feedforward Artificial Neural Networks", Cambridge, MA: The MIT Press, MA, USA, 352 pages.
- Rose J. B, Farrah S. R., Friedman D. et al., (1999)**, "Public evaluation of advanced water for potable application", Water Science and Technology, 40, pp.: 247- 252.
- Rosenblatt F. (1958)**, "The Perceptron: A probabilistic model for information storage and organization in the brain", Psychological Review, 65, pp.:386-408.
- Rowe D. R. et Abdel Magid I. M., (1995)**, "Handbook of Wastewater Reclamation and Reuse", CRC Press, 576 pages.

- Rui, Y. et El-Keib A. A. (1995)**, "A Review of ANN-Based Short-Term Load Forecasting Models", Proceeding of 27th IEEE Southeastern Symposium on System Theory, MS, USA, pp.: 78-82.
- Rumelhart, D. E., Hinton, G. E., et Williams, R. J. (1986)**, "Learning internal representations by error propagation", in Rumelhart, D.E. and McClelland, J. L., eds. (1986), "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", The MIT Press, Cambridge, USA, Volume 1, pp.: 318-362,
- Rustum R. (2009)**, "Modelling Activated Sludge Wastewater Treatment Plants Using Artificial Intelligence Techniques (Fuzzy Logic and Neural Networks)", Thèse de Doctorat, Heriot-Watt University, UK, 271 pages.
- Sadiq R. H., Tahir A., A. M., Sheikh A. K., et al., (2003)**, "Secondary Effluent Treatment by Slow Sand Filters: Performance and Risk analysis", Water, Air, and Soil Pollution, Vol. 143, pp. 41–63.
- Sadowski A. (2002)**, "Traitement des eaux usées urbaines", CIRSEE, ENGEES, Strasbourg, France, 418 pages.
- Saleem A. A. A. (2007)**, "Performance Analysis of Sanitary Wastewater Treatment Plants: Reliability-Based Analysis, Master thesis, King Saud University, KSA, 198 pages.
- Belanche L., Sánchez M. et Cortés U. (1992)**, "Knowledge-Based Techniques in Wastewater Treatment Plants Management", Industrial and Engineering Applications of Artificial Intelligence and Expert Systems Lecture Notes in Computer Science, Volume 604, pp.: 324-336.
- Saporta G. (2006)**, "Analyse des données et statistique", 2^{ème} Edition, Editions TECHNIP, Paris, France, 631 pages.
- Sawicki, J. M. (2004)**, "Aerated Grit Chambers Hydraulic Design Equation", Journal of Environmental Engineering, Vol. 130, N° 9, pp. 1050–1058.
- Schalkoff R.J. (1997)**, "Artificial Neural Networks", McGraw-Hill Series in Computer Science, McGraw-Hill Companies, New York, USA, 448 pages.
- Sellers G. R., Vaderman S. B., et Hackert. A. F., (1992)**, "A first Course in Statistics", 3rd editions, HarperCollins Publishers Inc, New York, USA, 831 pages.
- Spangler, J. (2006)**, "The Grit of the Nitty Gritty", Presented at Michigan Water Environment Association Process Seminar, Grand Rapids, Michigan, USA, 2 pages.
- Spellman F. R. (2003)**, "Handbook of Water and Wastewater Treatment Plant Operations", CRC Press LLC, NY, USA, 669 pages.
- Steel, E. W. et McGhee T. J. (1979)**, "Water Supply and Sewerage", 5th ed., McGraw-Hill, New York, pp.: 456–467.

Tarassenko L. (1998), "A Guide to Neural Computing Applications", John Wiley and Sons, New York, USA, 151 pages.

Tchobanoglous G., Burton F. L., et Stensel H. D. (2003), "Wastewater Engineering: Treatment and Reuse", McGraw-Hill Education, 1819 pages.

Theilliol D., Ponsart J. C., Harmand J., Join C. et al., (2003), "One-line estimation of unmeasured inputs for anaerobic wastewater treatment processes", Control Engineering Practice, vol. 11, pp. :1007-1019.

Touzet C. (1992), "Les réseaux de neurones artificiels: Introduction au connexionnisme", 129 pages.

USEPA, (1991), "Technical Support Document for Water Quality-based Toxics Control", EPA 505/2690-001, U.S. Environmental Protection Agency, Washington, USA.

User guide (2009), "NeuralWorks Predict® The Complete Solution for Neural Data Modeling User Guide", NeuralWare, PA, USA, 402 pages

Vasel J. L. (2007), "Evolution de l'assainissement individuel : Perspectives et questions en suspens", Tribune de l'eau, Vol. 60, N°641, pp.: 3-16.

Vyas M., Modhera B., Vyas V., et A. K. Sharma A. K. (2011), "Performance Forecasting of Common Effluent Treatment Plant Parameters by Artificial Neural Network", ARPN Journal of Engineering and Applied Sciences, Vol. 6, N°1, pp.: 38-42.

Wadsworth H. M., (1990), "Handbook of Statistics Methods for Engieneering and Scientists" McGraw-Hill Companies, New York, USA, 768 pages.

WEF (1998), "Design of Municipal Wastewater Treatment Plants", 4th ed., Water Environment Federation Manual of Practice 8, Alexandria, Virginia, pp. 1001–1070.

Werbos, P.J. (1974/1994), "The Roots of Backpropagation", John Wiley and Sons, New York, USA, 319 pages.

Widrow B. et Hoff M. E. Jr. (1960), "Adaptive switching circuits", IRE WESCON Convention Record. Part 4, pp. 96-104. Reprinted in Anderson and Rosenfeld (1988).

Wilson, G., Tchobanoglous G., et Griffiths J. (2007a), "The Nitty Gritty: Grit Sampling and Analysis", Water Environment & Technology, pp.: 64–68.

Zhu J., Zurcher J., Rao, M. et al. (1998), "An on-line wastewater quality prediction system based on a time-delay neural network", Engineering Application of Artificial Intelligence 11, pp.: 747-758.

Zupan J. et Gasteiger J. (1993), "Neural Networks for Chemist An Introduction", VCH Publishers, New York, USA, 309 pages.