

INSTITUT NATIONAL DE LA RECHERCHE SCIENTIFIQUE

CENTRE EAU TERRE ENVIRONNEMENT

QUÉBEC

**Revue des modèles appliqués dans la modélisation de l'effet des sources
de pollution ponctuelles sur la qualité de l'eau**

Par

Khawla Riahi

André St-Hilaire

Référence à Citer :

Riahi, K. et St-Hilaire, A. (2023). Revue des modèles appliqués dans la modélisation de l'effet des sources de pollution ponctuelles sur la qualité de l'eau. Québec, Institut national de la recherche scientifique, Centre Eau Terre Environnement, 40 pages (INRS - Centre Eau Terre Environnement, rapport de recherche R2151).

© INRS, Centre - Eau Terre Environnement, 2023
Tous droits réservés

ISBN : 978-2-89146-975-3 (version électronique)

Dépôt légal - Bibliothèque et Archives nationales du Québec, 2023
Dépôt légal - Bibliothèque et Archives Canada, 2023



Table des matières

Liste des figures.....	iii
Liste des tableaux	iii
Liste des abréviations	iv
1. INTRODUCTION.....	7
2. LA MODÉLISATION DE LA QUALITÉ DE L’EAU	8
3. MODÈLES DE LA QUALITÉ DE L’EAU	10
3.1. Les modèles déterministes.....	10
3.2. Les modèles statistiques	23
3.2.1. Approche d’analyses multivariées.....	23
3.2.2. Les modèles statistiques spatiaux	26
3.2.3. Approches probabilistes	27
3.3. Les modèles basés sur l’intelligence artificielle	28
3.3.1. Les réseaux de neurones artificiels (RNA).....	28
3.3.2. Les différentes architectures	29
4. CONCLUSION	33
5. BIBLIOGRAPHIE	36

Liste des figures

Figure 1: Méthode de choix du modèle de la qualité de l'eau approprié.....	9
Figure 2: Représentation schématique d'un réseau de neurones artificiels.....	29

Liste des tableaux

Tableau 1: Modèles de qualité de l'eau recensés, leurs points forts et leurs limitations.....	11
Tableau 2: Utilisations des modèles de RNA dans la modélisation de la qualité de l'eau	31

Liste des abréviations

AC: Analyse en cluster

ACP: Analyse en composante principale

AD: Analyse discriminante

AF: Analyse factorielle

ANN : Artificial Neural Network (Réseau de neurones artificiels)

ARIMA: Autoregressive integrated moving average (modèle autorégressif avec moyenne mobile)

CCNN: Convolutional neural network (réseau de neurones avec convolution)

Chl-a: Chlorophylle a

COT: Carbone organique total

CSS: Concentration de sédiments en suspension

DBO: Demande Biologique en Oxygène

DCO: Demande Chimique en Oxygène

EC: Conductivité Électrique

ELM: Extreme Learning Machine

ESN: Echo State Network

GRNN: GPU-based RNN inference library (bibliothèque pour réseau de neurones artificiels)

IA: Intelligence Artificielle

IQE: Indice de Qualité de l'Eau

LSTM: Long short-term memory (mémoire à long et court terme)

MLP: Multiple Linear regression (régression linéaire multiple)

MNN: Minimum error Neural Network (réseau de neurones minimisant l'erreur)

NARX: Nonlinear Autoregressive Exogenous (modèle autorégressif non linéaire avec variables explicatives exogènes)

NGL: Azote Total

NH₃-N: Azote ammoniacal

NO₂: Nitrite

NO₃: Nitrate

OD: Oxygène Dissous

P: Phosphore

PH: Potentiel hydrogène

PT: Phosphore Total

RBFNN: Radial Basis Function Neural Network (Réseau de neurones avec fonctions radiales de base)

RLM: Régression Linéaire Multiple

RNA: Réseau de Neurones Artificiels

TA: Total Alkalinity (alcalinité totale)

TDNN: Time Delay neural network (réseau de neurones artificiels avec délai)

TDS: Total Dissolved Solids (solides dissous totaux)

TH: Total Hardness (dureté totale)

TLRN: Time Lag Recurrent Networks

Tur: turbidité

TW: Température de l'eau

WNN: Wavelet Neural Network

1. INTRODUCTION

La dégradation de la qualité de l'eau des bassins versants est principalement due au développement socio-économique, dont découlent la croissance démographique, l'urbanisation, l'industrialisation et l'intensification de l'agriculture. Au cours des dernières décennies, les cours d'eau du monde entier sont entre autres pollués par plusieurs sources ponctuelles (les eaux usées industrielles et domestiques...) et non ponctuelles (les champs agricoles, les fosses septiques défectueuses, le dépôt direct de matières fécales animales...).

L'eau de surface est l'une des principales sources d'eau douce utilisées pour satisfaire les innombrables besoins de la société, Il est donc obligatoire de maintenir la qualité de l'eau aux normes prescrites dans les rivières pour assurer un approvisionnement ininterrompu en eau douce qui répond aux normes (Chen et al., 2016).

Pour relever ce défi, les outils de modélisation de la qualité de l'eau peuvent être particulièrement utiles. Ils sont à la fois nécessaires pour décrire et prédire les conditions de la qualité de l'eau et comprendre le fonctionnement de l'écosystème aquatique.

Les modèles de pollution sont devenus des outils importants pour comprendre, évaluer et prévoir les impacts négatifs de la pollution sur la qualité de l'eau. Aujourd'hui, il existe de nombreux modèles disponibles pour les utilisateurs. Cependant, les différents types de modèles présentent des formes et des structures différentes ainsi que des complexités de calcul différentes. Il est difficile pour les utilisateurs de sélectionner un modèle approprié pour une application spécifique sans une compréhension claire des limites ou des forces de chaque modèle ou outil.

La présente revue propose un recensement des outils de modélisation de l'évolution de la qualité de l'eau dans les cours d'eau. Divers documents ont été consultés : bibliographie, rapports d'études, sites internet, revues spécialisées, dont la liste figure dans la bibliographie (p. 35).

2. LA MODÉLISATION DE LA QUALITÉ DE L'EAU

La modernisation de l'agriculture et l'urbanisation rapide furent largement responsables du déclin de la qualité de l'eau. L'afflux massif des matières organiques qui provenaient de la fertilisation intensive des milieux agricoles et des rejets municipaux enrichis en phosphates (détergents) ont favorisé et accéléré l'eutrophisation des rivières et de plans d'eau. Afin de pallier au problème de la détérioration des milieux aquatiques, des investissements ont été faits dans de nombreux programmes de simulation de la qualité de l'eau afin de décrire l'évolution spatiale et temporelle d'un certain nombre de paramètres physico-chimiques. L'un des premiers modèles de qualité des eaux fut celui de Streeter et Phelps (1925), dont l'objectif était d'évaluer l'impact d'un rejet sur les concentrations en oxygène à l'aval d'un point de rejet à partir de deux équations et deux paramètres à caler. Puis le nombre de paramètres pris en compte a progressivement augmenté. De plus en plus de modèles ont été développés afin de prendre en compte toutes sortes de substances chimiques potentiellement toxiques et représenter des écosystèmes dans leur ensemble. Les modèles de qualité des eaux de rivières sont donc nombreux et ont chacun leurs caractéristiques propres qu'il faut prendre en compte lors du choix d'un modèle pour un projet donné. La dernière phase de développement moderne des modèles de qualité de l'eau a commencé avec le début d'utilisation d'internet au début des années 2000. L'amélioration de la technologie informatique a permis l'expansion et l'amélioration des modèles de qualité de l'eau. Cela comprend des ordinateurs plus rapides capables de gérer des analyses environnementales plus détaillées (spatialement et temporellement), de meilleures interfaces utilisateur et logiciels SIG, une plus grande accessibilité aux données environnementales à partir de référentiels en ligne et des cadres de modélisation plus robustes reliant l'hydrodynamique, la qualité de l'eau, les bassins versants et les modèles atmosphériques.

Un modèle de qualité de l'eau des rivières est une formulation mathématique des processus biologiques et physico-chimiques (diffusion, dispersion, advection, sédimentation, resuspension...) ainsi que du transport dans les rivières. Il existe trois grandes familles de modèles, les modèles stochastiques (basés sur des statistiques), les modèles basés sur

l'intelligence artificielle (IA) et les modèles déterministes (basés sur les équations de la physique/chimie/biologie).

Ces outils ne sont pertinents que si les questions auxquelles ils doivent apporter des éléments de réponse sont bien identifiées et que le modèle choisi représente bien les phénomènes principaux grâce à des données suffisantes et un calage préalable.

La figure 1 présente l'une des méthodes de choix de modèle de qualité de l'eau. Dans cette figure, les données réfèrent aux intrants nécessaires, ex., les Réseau hydrographiques, météorologiques, les rejets ponctuels et diffus, les mesures de débits, les informations géographiques, etc. Le terme « variables » réfère à la variable qu'on cherche à modéliser : contaminants conventionnel (l'azote ammoniacal, coliformes fécaux (CF), DBO, Pt, MES ou turbidité, etc.) Le terme « système » réfère au type d'écosystème aquatique recevant les rejets : lac, rivière, réservoir. Le terme « processus » réfère aux mécanismes physiques, chimiques, biochimiques mis en cause. Le terme « ressource » réfère aux ressources informatiques (par ex. temps de calcul), humaines, ou financières...

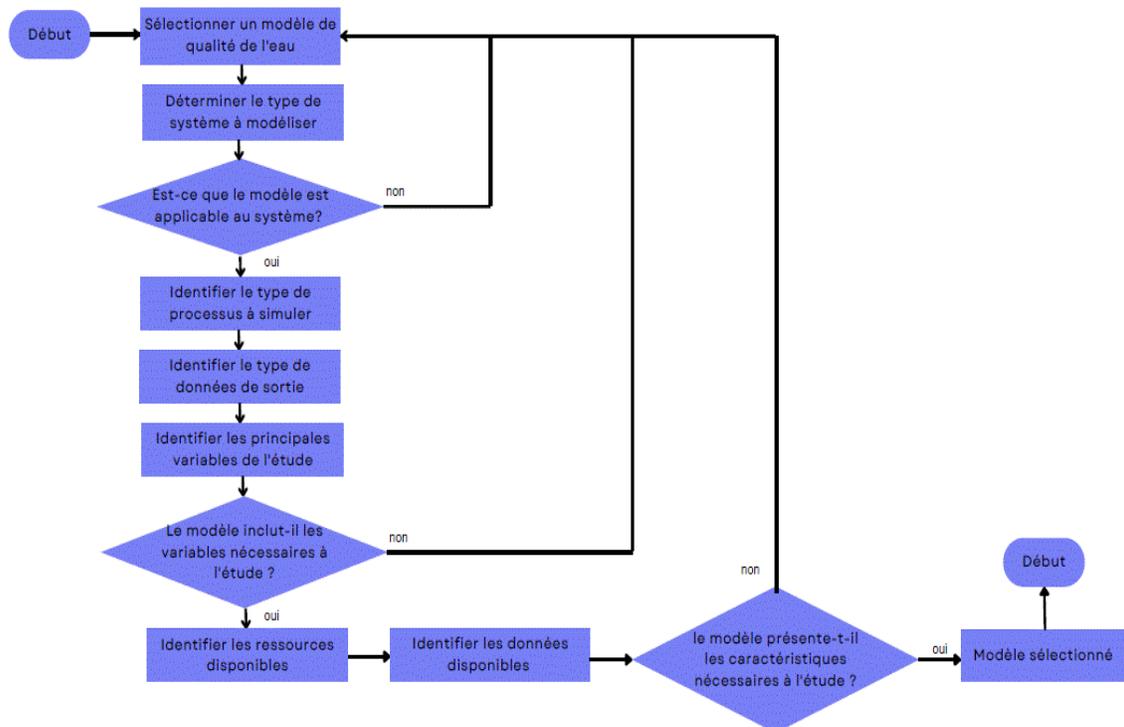


Figure 1: Méthode de choix du modèle de la qualité de l'eau approprié

3. MODÈLES DE LA QUALITÉ DE L'EAU

3.1. Les modèles déterministes

Afin de répondre aux objectifs de cette étude, les principaux critères de sélection des modèles étaient les suivants :

- La capacité du modèle à traiter les sources de pollution ponctuelles et diffuses;
- La capacité du modèle à simuler les nutriments, les sédiments et les agents pathogènes;
- La considération ou non du cycle saisonnier et des débits d'étiage;
- la capacité du modèle à s'adapter face aux changements climatiques;
- La complexité du modèle par rapport à sa précision;
- La disponibilité du modèle et de sa documentation;
- La disponibilité des données d'entrée;
- L'applicabilité du modèle à différents types de bassins versants au Québec.

Le tableau ci-dessous contient les différents modèles recensés en décrivant chacun d'entre eux.

Tableau 1: Modèles de qualité de l'eau recensés, leurs points forts et leurs limitations

Modèle	Type	Polluants simulés	Points forts et applications	Limitations
QUALs	1D	DBO, DCO, TW, chl-a, NH ₃ -N, NO ₂ , Ni, NO ₃ , P, OD, bactéries coliformes, sédiments...	<ul style="list-style-type: none"> • Les QUALs sont des modèles unidimensionnels de qualité de l'eau en régime permanent, largement utilisés pour les simulations de rivières (Ning et al., 2000; Pelletier et al., 2006; Zhu et al., 2015). • L'USEPA a publié une série de modèles QUAL : QUAL2E, QUAL2E-UNCAS, QUAL2K, et QUAL2KW. • Les QUALs simulent les charges polluantes ponctuelles et non ponctuelles. • Le modèle QUAL2E est l'un des modèles de qualité de l'eau les plus utilisés dans le monde (Costa et al., 2019), c'est un modèle en régime permanent, du débit et de la qualité de l'eau dans les cours d'eau. Il simule l'OD et (jusqu'à 15 paramètres) déterminants de la qualité de l'eau associés le long d'une rivière et de ses affluents (Kannel et al., 2011). • Le modèle est capable de prendre en compte les effets des variations météorologiques diurnes (par exemple le rayonnement) sur certains déterminants de la qualité de l'eau tels que l'OD et la température (Brown & Barnwell, 1987). 	<ul style="list-style-type: none"> • En tant que modèles en régime permanent, les QUALs (à l'exception de QUAL2KW) sont limités aux périodes où les débits des cours d'eau et les rejets sont essentiellement constants (régime permanent, ou « steady-state »). Ils sont donc incapables de simuler la variation temporelle du débit, ce qui affecte énormément les principales conditions de la qualité de l'eau • Ne considèrent pas le mouvement des sédiments en suspension, les macrophytes et la dénitrification. • Lors de l'élaboration de QUALs, les tronçons et les éléments de calcul

			<ul style="list-style-type: none"> • QUAL2E-UNCAS est une amélioration de QUAL2E, permettant une analyse d'incertitude en même temps que la simulation de la qualité de l'eau. Le modèle QUAL2K est également une version mise à jour de QUAL2E où la plupart des équations des modèles sont les mêmes, à l'exception de celles pour l'OD, la DBO et le nitrate, qui ont été améliorées (Angello et al., 2021). • QUAL2Kw est capable de simuler des conditions d'écoulement permanent ou dynamique. • Les modèles QUALs sont : largement documenté dans le manuel de l'utilisateur (Brown et Barnwell, 1987) et ils ne nécessitent pas un grand nombre de données d'entrée. Ils sont disponibles gratuitement. • La principale différence entre QUAL2E et QUAL2E-UNCAS est que ce dernier propose une analyse de l'incertitude des résultats simulés afin d'évaluer les performances du modèle et également d'identifier les paramètres les plus critiques pour la détermination de la qualité de l'eau. 	<p>ne doivent pas être supérieurs à 25 et 20 par tronçon ou à un total de 250, respectivement.</p> <ul style="list-style-type: none"> • Le modèle est conçu pour des simulations à petite échelle qui se font sur quelques dizaines de kilomètres. La simulation se fait dans des tronçons composés d'un certain nombre d'éléments unitaires Δx définis par l'utilisateur. La longueur de Δx est ajustée dépendamment de la vitesse d'écoulement et en se basant sur le critère de Fisher, afin de minimiser les erreurs numériques associées à la résolution des équations (Fisher et al., 1979).
Environmental fluid dynamics code (EFDC)	2D/3D	OD, DCO, NH ₃ -N, NO ₂ , Ni, NO ₃ , P, algues, solides dissous, EColi,	<ul style="list-style-type: none"> • EFDC est un modèle numérique à base physique applicable à l'analyse de l'hydrodynamique et de la qualité de l'eau dans les lacs, les estuaires, les rivières, les zones humides et les réservoirs (Hamrick, 1992). 	<ul style="list-style-type: none"> • Toutes les références datent de plus de 10 ans. • Un modèle assez complexe

		sédiments et autres.	<ul style="list-style-type: none"> • Le modèle EFDC est un système polyvalent de modélisation des eaux de surface, composé de quatre modules permettant de simuler des composantes d'hydrodynamique, de transport de sédiments, de transport de contaminants toxiques et d'eutrophisation de la qualité de l'eau (Wu et Xu 2011). • Le modèle EFDC résout les équations tridimensionnelles du mouvement dans la colonne d'eau, à la surface libre, les équations de turbulence et de transport. Il fournit une grande précision pour la simulation numérique (Wang et al., 2013) • La grande souplesse du modèle EFDC dans sa mise en place, permettant l'ajout ou la soustraction des couches de discrétisation verticales et l'utilisation d'un pas de temps dynamique pour une meilleure stabilité numérique. • Capable de simuler l'élévation de surface, la vitesse d'écoulement, la température la salinité, ainsi que 21 paramètres de qualité de l'eau 	
Water Quality Analysis Simulation Program (WASP)	1D/2D/3D	NH3-N, NO ₂ , Ni, NO ₃ , P, température, ph, alcalinité, OD, DBOC, algues, bactéries (coliformes fécaux), métaux, sédiments...	<ul style="list-style-type: none"> • WASP est un modèle de qualité des eaux de surface qui a été développé par l'Agence américaine de protection de l'environnement (EPA). • WASP 6 est un logiciel modulaire de modélisation dynamique des systèmes qui permet de simuler différents types de pollutions en fonction des modules. Module « Eutrophisation » Module « produits chimiques organiques » Module « mercure ». 	<ul style="list-style-type: none"> • Le WASP peut être considéré comme un modèle de haute complexité. Les modèles purement statistiques ont tendance à être plus simples (Schwarz et al. 2006).

			<ul style="list-style-type: none"> • La dernière version du modèle (WASP7) est sortie en 2019, elle comprend deux modules cinétiques généraux pour simuler la pollution conventionnelle (oxygène dissous, demande biochimique en oxygène, nutriments et eutrophisation) et la pollution toxique (produits chimiques organiques, métaux et sédiments) (Ambrose et al. 2001). • WASP peut aussi être couplé avec des modèles d'hydrodynamique ou de transport sédimentaire qui fournissent débits, profondeurs, vitesses, température, salinité et flux de sédiments. Les résultats peuvent être fournis sous deux formes : tableau et/ou graphique. • WASP est un modèle déterministe tridimensionnel bien établi (Di Toro et al. 1983), fournissant une modélisation dynamique des compartiments pour les systèmes aquatiques, y compris la colonne d'eau et le benthos sous-jacent. (Zhang et al. 2008). • L'équation générale du transport des solutés tient compte de l'advection-dispersion (Zhang et al. 2008). • La capacité de fonctionner de manière stable sans données d'entrée très détaillées (et souvent non disponibles) est l'une des principales forces de WASP (Rose et Pedersen 2005). 	<ul style="list-style-type: none"> • Selon Kannel et al. (2011), le modèle WASP7 est très complexe et il est difficile de justifier son temps et son coût. • Processus laborieux de paramétrage, d'étalonnage et de validation ; nécessite une analyse de l'incertitude et une interprétation attentive des résultats ; • Forte courbe d'apprentissage ou formation nécessaire pour l'application du modèle.
SPAtially Referenced Regression On	2D	NH3-N, NO2, Ni, NO3, température, coliformes	<ul style="list-style-type: none"> • Le modèle SPARROW est un modèle hybride statistique/mécaniste de modélisation des bassins versants et des eaux de surface développé par Smith et al. (1997). 	<ul style="list-style-type: none"> • La description du processus est trop limitée et les résultats

Watershed attributes (SPARROW)		fécaux, pesticides...	<ul style="list-style-type: none"> • Le modèle SPARROW possède une infrastructure unique qui consiste en un réseau détaillé de tronçons de cours d'eau avec des bassins versants délimités par des modèles d'élévation numériques (DEM), comprenant des stations de mesure et des données SIG à référence spatiale (Schwarz et al. 2006). • Les équations de régression non linéaire décrivent le transport des contaminants depuis les sources ponctuelles et diffuses à travers le réseau de cours d'eau et de rivières. Le modèle prédit le flux, la concentration et le rendement des contaminants dans les cours d'eau et a été utilisé pour évaluer des hypothèses alternatives sur les sources importantes de contaminants et les propriétés des bassins versants qui contrôlent le transport sur de grandes échelles spatiales (Schwarz et al. 2006). • SPARROW analyse la qualité de l'eau des ruisseaux, des rivières et des lacs en fonction de l'emplacement et de l'intensité relative des sources de contaminants, du relief et de facteurs environnementaux. SPARROW permet de suivre le parcours des contaminants et des nutriments qui traversent les bassins versants. • SPARROW peut être classé comme un niveau de complexité inférieur. 	<p>statistiques sont difficiles à interpréter.</p> <ul style="list-style-type: none"> • Le modèle exige un nombre suffisamment important de stations de contrôle de la qualité de l'eau pour la modélisation. Cependant, il a déjà été utilisé pour étudier la qualité des eaux dans divers bassins versants au Canada, donc certains avec des informations limitées (Wellen et al, 2014).
AQUATOX	1D/2D/3D	NGL, P, OD, sédiments organiques et inorganiques,	<ul style="list-style-type: none"> • AQUATOX est un modèle d'écosystème qui simule les polluants (nutriments et sédiments) et plusieurs niveaux trophiques (algues fixées et planctoniques, végétation aquatique, invertébrés et poissons). 	<ul style="list-style-type: none"> • N'est pas capable de modéliser les métaux et impossible à coupler

		produits organiques toxiques (plus de 20 produits), biomasse algale	<ul style="list-style-type: none"> • C'est un modèle mécanistique ou « process-based model ». La simulation se fait par calcul simultané de chacun des processus chimiques et biologiques. Les données d'entrée nécessaires sont les charges apportées au milieu aquatique, les caractéristiques chimiques des produits toxiques, ainsi que caractéristiques biologiques des plantes et animaux. • Ce modèle fait partie des rares modèles qui expriment le devenir environnemental des produits chimiques dans les écosystèmes aquatiques et leurs effets sur les organismes résidents. AQUATOX est le dernier d'une série de modèles conçus pour simuler le devenir des polluants, à commencer par le modèle d'écosystème aquatique CLEAN publié en 1974 (Park 1974) et ensuite PEST développé en 1982 (Park et al. 1982) • Peut être appliqué aux bassins expérimentaux, aux étangs et aux enceintes d'étangs, aux ruisseaux, aux petites rivières, aux segments de rivières reliés, aux lacs, aux réservoirs, aux segments de réservoirs reliés et aux estuaires (Shoemaker et al., 2005 et Stolarska et Skrzypski, 2012). • AQUATOX utilise la méthode numérique de Runge-Kutta de quatrième et cinquième ordre (Park et al. 2008) et une méthode d'étape adaptative pour résoudre les équations différentielles basées sur un calcul du bilan massique. • Le modèle simule l'impact de la pollution sur les organismes vivants et requiert des données hydrauliques, 	avec les modèles hydrodynamiques
--	--	---	---	----------------------------------

			<p>géométriques, ainsi que des données sur les facteurs abiotiques et les caractéristiques de l'environnement (Park et Clough, 2004, Bingli et al., 2008, Stolarska et Skrzypski, 2012)</p> <ul style="list-style-type: none"> • Le modèle peut expliquer les fluctuations de la demande en oxygène et de la toxicité résultant de faibles niveaux d'oxygène et de la présence d'ammoniac (Park et al.1988) • Les charges de pollution sont des paramètres clés du modèle AQUATOX. Les charges sont classées en charges de source ponctuelle et charges de source diffuse. 	
MIKE	1D/2D/3D	DBO, NH ₃ -N, NO ₂ , Ni, NO ₃ , phytoplancton, zooplancton, chlorophylle, carbone, oxygène, P, T, métaux lourds dissous, sédiments...	<ul style="list-style-type: none"> • MIKE est un outil de simulation des processus hydrodynamiques, hydrologiques, sédimentaires et de qualité des eaux (Danish Hydraulic Institute, 1993). • Le modèle comprend des modules permettant de traiter différents types de données. L'un de ces modules est le module de qualité de l'eau ECOLab qui est une boîte à outils de processus écologiques adimensionnels. ECOLab peut être appliqué à MIKE 11, MIKE 21, MIKE 3 et MIKE SHE et constitue la base de la simulation de la qualité de l'eau qui nécessite généralement des entrées hydrodynamiques (HD) et d'advection/dispersion (AD) (Q. Wang et al., 2013). • MIKE offre plusieurs approches différentes allant des approches simples, forfaitaires et conceptuelles aux approches avancées, distribuées et physiques. 	<ul style="list-style-type: none"> • L'installation du modèle nécessite un expert. • Nécessite beaucoup d'informations/données pour le fonctionnement et le calibrage du modèle (Kayode et Muthukrishna, 2018).

			<ul style="list-style-type: none"> • Il est largement utilisé pour l'analyse de la qualité de l'eau et permet de simuler le transport des sédiments, des nutriments, et des pesticides (Stolarska et Skrzypski, 2012). 	
<p>BASINS</p> <p>Better Assessment Science Integrating Point and Nonpoint Sources (BASINS)</p>		<p>Sédiment, pesticides, NGL, TP, DBO, phytoplankton, zooplankton, OD</p>	<ul style="list-style-type: none"> • BASINS est un système d'analyse environnemental permettant la simulation de la qualité de l'eau à l'échelle du bassin versant (Q. Wang et al., 2013 et USEPA (Agence américaine de protection de l'environnement), 2019). • La dernière version 4.1 de BASINS a été publiée en 2013. C'est un modèle SIG libre et open-source, compatible avec Windows et est disponible à l'adresse suivante : https://www.epa.gov/ceam/basinsdownload-and-installation • Il permet d'estimer rapidement les effets de la pollution causée par des sources ponctuelles et diffuses dans un format simple à utiliser et comprendre. (Wang et al., 2013). • BASINS est un modèle polyvalent qui intègre plusieurs modèles de bassin versant tels que HSPF, SWAT, SWMM, des modèles de qualité de l'eau tels que AQUATOX et WASP ainsi que des outils permettant l'analyse des systèmes environnementaux. • Les objectifs primordiaux de BASINS sont les suivants : (1) identifier et hiérarchiser les cours d'eau; (2) évaluer l'ampleur et l'importance potentielle de la pollution de sources ponctuelles et diffuses; (3) simuler les charges de nutriments de sources ponctuelles et diffuses ainsi que 	<ul style="list-style-type: none"> • Les modèles nécessitent de grandes quantités de données d'entrée. • La base de données utilisée dans BASINS est uniquement compatible avec les formats utilisés au États-Unis. Cela s'est avéré un problème sérieux lors de la mise en place du (des) modèle(s) pour les bassins versants canadiens. • Pour les applications canadiennes, les données sur la qualité de l'eau devront être extraites des bases de données fédérales et provinciales existantes et formatées pour fonctionner dans BASINS (Booty et al, 2009).

			<p>les processus de transport; (4) évaluer la valeur relative d'une stratégie de contrôle potentielle. Permet de visualiser les conditions environnementales pour le public à l'aide de tableaux, de graphiques et de cartes.</p> <ul style="list-style-type: none"> • Les données d'entrée comprennent des données cartographiques (MNT, les cartes d'occupation des sols.), des données météorologiques et des données sur les sources ponctuelles. Les données sorties comprennent des cartes, des graphiques et des tableaux résumant la pollution ponctuelle et non ponctuelle dans un bassin versant. 	
HSPF		<p>Sédiments, pesticides, NGL, PT, DBO, phytoplankton, zooplankton, OD, pesticides, Coliformes fécaux, ammoniacque, nitrate-nitrite</p>	<ul style="list-style-type: none"> • HSPF est un modèle de bassin versant complet et semi-distribué qui est inclus dans BASINS et a été largement utilisé dans l'analyse hydrologique et de la qualité de l'eau à l'échelle du bassin versant (Zhang et Ross 2012). L'interface Windows interactive de HSPF, qui peut être lancée via BASINS, est appelée WinHSPF et comprend trois modules : PERLND (terres perméables), IMPLND (terres imperméables), et RCHRES (bief et réservoir). Les simulations hydrologiques et de la qualité de l'eau sont effectuées via les modules PERLND et IMPLND tandis que les processus hydrauliques et de la qualité de l'eau dans les cours d'eau sont simulés via le module RCHRES. • HSPF est un modèle détaillé de bassin versant avec de nombreux paramètres • HSPF utilise des données météorologiques pour calculer les hydrogrammes de débit et les pollutogrammes. 	<ul style="list-style-type: none"> • Le modèle n'est pas sensible à la variation spatiale, il ne considère pas la morphologie des bassins versants. • Limité aux rivières bien mélangées, aux réservoirs et aux écoulements unidimensionnels. • Nécessite un haut niveau de compétences pour être appliqué. • De nombreux processus physiques sont basés sur des relations

			<ul style="list-style-type: none"> • HSPF est généralement utilisé pour évaluer les effets des changements d'utilisation des terres, de l'exploitation des réservoirs, des alternatives de traitement de sources ponctuelles ou non ponctuelles, des déviations de flux, etc. Des programmes, disponibles séparément, permettent le prétraitement et le post-traitement des données pour les statistiques et les graphiques (Stolarska & Skrzypski, 2012) • Applicable aux systèmes naturels et artificiels pour simuler l'hydrologie et la qualité de l'eau au niveau des surfaces terrestres perméables et imperméables, des cours d'eau et des bassins de retenue bien mélangés. • HSPF est un modèle de simulation en continu qui nécessite comme entrées, des séries chronologiques. Il permet de modéliser le mouvement des polluants dans les bassins versants par des équations de dispersion/convection. • Il permet d'estimer le mouvement des polluants dans les bassins versants à l'aide de modèles mathématiques (Stolarska et Skrzypski, 2012). Les modèles mathématiques reposent sur les concepts fondamentaux des systèmes physiques. Un système physique se décrit par quelques variables mesurables et des limites bien précises 	<p>empiriques, et non sur des mécanismes.</p> <ul style="list-style-type: none"> • Un grand nombre d'éléments sont nécessaires et il est difficile d'assurer leur exactitude, d'où la nécessité d'effectuer un calage efficace. • La précision du modèle dépend de facteurs météorologiques (Liz, Liu et Li, 2012).
SWAT Soil and Water Assessment Tool	2-D	Sédiments, NGL, PT, pesticides, bactéries, CO, OD, DBO	<ul style="list-style-type: none"> • Le modèle SWAT est un modèle hydrologie/qualité de l'eau, conçu pour simuler l'impact des pratiques et gestions des terres sur l'hydrologie, les sédiments et les transports de polluants dans les bassins versants (Arnold et al., 1998 ; Neitsch et al., 2005). 	<ul style="list-style-type: none"> • Un effort considérable est requis pour préparer les données d'entrée nécessaires à

			<ul style="list-style-type: none"> • SWAT est couplé avec un système d'information géographique ARCGIS (ArcSWAT 2012.10.5.1) et utilise le type de sol, l'utilisation du sol, les caractéristiques topographiques et les données climatiques en intrants. • Le modèle le plus utilisé dans le monde pour évaluer l'impact de la pollution diffuse à l'échelle du bassin versant, il a été utilisé dans plus de 68 pays (Costa et al., 2019). • Les données d'entrée sont majoritairement disponibles auprès des agences gouvernementales • SWAT peut également être utilisé pour simuler l'impact des sources de pollution ponctuelles (Lam et al., 2010). 	<p>l'exécution du modèle SWAT.</p> <ul style="list-style-type: none"> • Les charges de sources ponctuelles sont limitées à des taux de charge constants pour toute la période de simulation.
CE-QUAL-W2	2D	TW, TDS, coliformes, sédiments inorganiques en suspension, matière organique dissoute, DBO, algues, matière détritique, P, NGL, OD, fer, alcalinité, pH.	<ul style="list-style-type: none"> • CE-QUAL-W2 est un modèle open source d'hydrodynamique et de qualité des eaux qui peut être appliqué aux rivières, lacs, réservoirs, estuaires et bassins versants (Masoumi et al., 2016; Shokri et al., 2014). • Le modèle suppose une homogénéité latérale ; il est donc adapté aux plans d'eau longs et étroits présentant des gradients de qualité de l'eau longitudinaux et verticaux (Cole et Wells, 2017). • Le modèle simule les vitesses horizontales et verticales, les niveaux d'eau, la température et 21 autres paramètres de qualité de l'eau (Stolarska et Skrzypski, 2012). • CE-QUAL-W2 comporte plus de 40 variables d'état de la qualité de l'eau, notamment les nutriments, la 	<ul style="list-style-type: none"> • Son application est une tâche complexe et chronophage (Cole et Wells, 2013). • Charge de calcul et de stockage sur un ordinateur lors de simulations continues. • L'accessibilité des données d'entrée est le problème limitant pour l'application (Cole et Wells, 2018).

			température, les sédiments, les algues, le périphyton et le phytoplancton. De plus, plus de 60 variables dérivées peuvent être calculées en interne à partir des variables d'état et produites en sortie (Cole et Wells, 2018).	
--	--	--	---	--

3.2. Les modèles statistiques

L'importance des approches statistiques pour révéler la variation spatiale de la qualité des eaux de surface a été largement citée dans la littérature (Singh et al. 2004 ; Shrestha et Kazama 2007 ; Li et al. 2009). L'utilisation d'outils statistiques dans l'analyse de la qualité de l'eau reflète plus précisément la nature multivariée de l'écosystème naturel et fournit un moyen efficace pour traiter et analyser les grands ensembles de données avec un grand nombre de paramètres. (McGarigal et al. 2000; Mustapha et Aris 2012).

Dans cette section on cite les différentes approches statistiques les plus utilisées:

3.2.1. Approche d'analyses multivariées

Les modèles statistiques multivariés sont largement appliqués pour évaluer les variations spatio-temporelles des paramètres de qualité de l'eau et aussi pour interpréter les ensembles de données vastes et complexes. Ces analyses permettent également d'identifier les facteurs/sources possibles responsables des variations de la qualité de l'eau permettant ainsi de développer des stratégies appropriées pour une gestion efficace des ressources en eau. Ces approches sont le plus souvent utilisées pour faire de la caractérisation et des regroupements homogènes plutôt que de l'estimation ou de la prévision.

L'analyse en Cluster AC

L'analyse en cluster ou analyse par grappes est une technique statistique généralement utilisée pour déterminer la variation spatiale et temporelle significative entre les différentes séries de données. C'est une technique de classification statistique dans laquelle un ensemble d'objets ou de points ayant des caractéristiques similaires sont regroupés en grappes. Elle englobe un certain nombre d'algorithmes et de méthodes différents qui sont tous utilisés pour regrouper des objets de même nature dans des catégories respectives. Les groupes d'objets qui en résultent doivent présenter une grande homogénéité interne (inter-groupes) et une grande hétérogénéité externe (intra- groupes) (McGarigal et al. 2000, Al-Odaini et al. 2012). L'objectif de cette analyse est d'organiser les données observées en structures significatives afin d'en tirer un meilleur aperçu.

Parmi toutes les analyses de clusters, le cluster agglomératif hiérarchique est l'approche la plus courante. L'analyse en clusters agglomératifs hiérarchiques (HACA) est utilisée pour agréger un groupe aux caractéristiques identiques de manière séquentielle en se basant sur la méthode de Ward qui utilise la distance euclidienne au carré comme mesure de similarité pour construire un dendrogramme (Lau et al. 2009). Un dendrogramme permet une réduction de la dimensionnalité et de la complexité des ensembles de données. Il fournit également une représentation visuelle du niveau d'uniformité dans le modèle des paramètres d'observation (Ogwueleka 2015, Kazi et al. 2009). Dans le clustering hiérarchique, les clusters sont formés séquentiellement en commençant par la paire d'objets la plus similaire et en formant des clusters plus élevés étape par étape.

Cette technique a été largement utilisée afin de regrouper les sites d'échantillonnages ou les stations de contrôle de qualité de l'eau en fonction des similitudes des caractéristiques de la qualité de l'eau échantillonnée.

Ogwueleka et Toochukwu Chibueze (2015) ont utilisé les techniques statistiques multivariées pour étudier les variations temporelles et spatiales et pour interpréter des ensembles de données volumineux et complexes sur la qualité de l'eau collectés dans la rivière Kaduna au Nigeria. L'AC hiérarchique a permis de regrouper l'ensemble des données en fonction des différences saisonnières et des différents niveaux de pollution.

(Mustapha et al, 2019) a fourni une étude détaillée de la qualité de l'eau dans la rivière Kano au Nigeria. L'AC hiérarchique a permis de regrouper les sites d'échantillonnage en trois grappes statistiquement significatives en fonction des similitudes des caractéristiques de la qualité de l'eau.

L'analyse discriminante (AD)

L'analyse discriminante est une méthode d'analyse multivariée qui sert à identifier les variables qui permettent de distinguer deux ou plusieurs groupes/clusters prédéfinis selon l'équation suivante :

$$f(G_i) = k_i + \sum_{j=1}^n w_{ij} + P_{ij} \quad (1)$$

Où i est le nombre de groupes (G), k_i la constante inhérente à chaque groupe, n le nombre de paramètres utilisés pour classer un ensemble de données dans un groupe donné, w_j le coefficient de pondération, attribué par l'AD à un paramètre sélectionné donné (p_j)

L'analyse discriminante linéaire a été appliquée par (Han et al, 2020) pour sélectionner les paramètres qui différencient le plus efficacement les groupes temporels (période humide contre période sèche) et les groupes spatiaux (parties est et ouest du lac) dans la rivière Fu et le lac Baiyangdian, tous deux situés dans une région très peuplée du nord de la Chine. L'analyse discriminante des variations temporelles est effectuée après que l'ensemble des données ait été divisé en deux groupes saisonniers (saison des pluies et saison sèche).

Les résultats de l'analyse discriminante permettent d'identifier les paramètres les plus significatifs pour discriminer entre deux périodes et responsable de la plupart des variations temporelles attendues dans la qualité de l'eau de la rivière. Diverses recherches menées dans le monde entier ont utilisé l'analyse discriminante (AD) pour révéler les variables les plus significatives qui entraînent une variation de la qualité des eaux de surface (Mustapha et al, 2019, Avila et al, 2018, Hajigholizadeh, 2016).

L'analyse Factorielle (FA)/ L'analyse des composantes principales (ACP)

Les analyses ACP/FA permettent de réduire la dimensionnalité en créant des nouvelles variables étant des combinaisons linéaires des variables précédentes, remplaçant les anciennes coordonnées de l'espace factoriel. Les nouvelles coordonnées sont appelées facteurs latents ou composantes principales.

L'ACP permet d'identifier un nombre réduit de facteurs latents aux sources de pollution telles que les sources de variations spatiales (pollution d'origine anthropique) et temporelles (saisonniers et climatiques) affectant la qualité des eaux (Kowalkowski et al. 2006). L'AF réduit la contribution des variables moins significatives obtenues à partir de l'analyse en composantes principales (ACP). Cette réduction est obtenue en transformant l'ensemble de données en un nouvel ensemble de variables, les composantes principales (CP), qui sont orthogonales (non corrélées) et classées par ordre décroissant d'importance. Mathématiquement, les CP sont calculés à partir de la covariance ou d'autres facteurs de pondération.

Les détails mathématiques de l'ACP/AF peuvent être trouvés dans Morrison (2004).

L'analyse en composantes principales et/ou l'analyse factorielle (ACP et AF) ont été utilisées par les chercheurs pour réduire le nombre de variables nécessaires pour décrire la variation observée dans les ensembles de données. Xu et al. (2009), dans leur étude sur la région du lac Taihu en Chine, ont découvert que les eaux de surface de la région étaient progressivement sensibles à la pollution anthropique. Cela a été révélé par l'ACP et l'AF. Dans l'ACP, trois composantes principales (CP) correspondaient à la subsistance résidentielle urbaine, à l'élevage et au ruissellement des terres agricoles respectivement. Huang et al. (2010) ont appliqué l'ACP et l'AF pour identifier les facteurs latents de pollution et ont produit trois sources potentielles de pollution, les résultats de cette étude ont démontré que la variation de la qualité de l'eau était principalement influencée par la pollution due aux eaux usées industrielles, au ruissellement agricole et aux activités domestiques. PCA et FA se sont avérés robustes pour dévoiler les sources de pollution

Dans l'AF, le concept de base a été exprimé en utilisant l'équation suivante:

$$Z_{ij} = af_1f_{1i} + af_2f_{2i} + \dots + af_mf_{mi} + e_{fi} \quad (2)$$

Où z est la valeur mesurée d'une variable, a est la saturation factorielle, f est le score factoriel, e est le terme résiduel tenant compte des erreurs ou d'autres sources de variation, i est le numéro de l'échantillon, j le numéro de la variable, et m est le nombre total de facteurs.

3.2.2. Les modèles statistiques spatiaux

Mainali et al. (2019) ont complété une revue des approches statistiques spatiales servant à modéliser la qualité de l'eau. Ils ont répertorié plusieurs approches, incluant celles qui sont basés sur les vecteurs propres spatiaux, les régressions pondérées géographiquement et les modèles basés sur le réseau spatial des cours d'eau. Les auteurs rappellent dans un premier temps, l'importance de la prise en compte de l'autocorrélation spatiale dans ces modèles (les points situés à proximité l'un de l'autre sont plus similaires que des points séparés par une plus grande distance). Le I de Moran est un de ces outils. Cependant il est calculé à partir de longues séries chronologiques et produit une seule valeur entre les paires de points.

Les modèles les plus simples sont basés sur l'approche par régression. Cependant ces modèles doivent être adaptés à la présence d'autocorrélation spatiale. Par exemple, le modèle avec pondération spatiale utilise une matrice de pondération dans laquelle chaque point indique l'influence relative des unités spatiales. Afin de limiter la taille de la matrice, il est fréquent de définir un voisinage autour du point considéré. Lorsque le modèle calé est imparfait d'erreur, il est habituel d'analyser la structure de cette incertitude. Si une structure d'autocorrélation est présente dans les résidus (la différence entre les valeurs estimées et observées) on peut inclure cette autocorrélation dans le modèle. Ce modèle est appelé en anglais : « spatial error model ».

Les modèles pour lesquels une pondération spatiale existe sont ponctuels (i.e., ils estiment la variable de qualité de l'eau en un seul point) et doivent être recalibrés pour chaque point d'intérêt. De plus, dans la majorité des cas, la direction n'est pas prise en compte: la distance est considérée comme un scalaire et non un vecteur. Pour pallier cette lacune les modèles de type « Spatial Stream Network » ou modèles de réseau de rivières, ont été proposés. Ces modèles prennent en compte la nature dendritique des réseaux hydriques (e.g., Isaak et al., 2014). Pour ce faire, la structure d'autocorrélation spatiale du modèle est modifiée. La pondération spatiale prend en compte une (ou plusieurs) variables suivantes : Le kilomètre de rivière, le volume ou le débit, la superficie drainée.

3.2.3. Approches probabilistes

L'approche probabiliste consiste à regrouper toutes les observations et de construire un histogramme de fréquence (diagramme de fréquence cumulée) ou une forme analytique d'un modèle de probabilité et à ensuite estimer les paramètres de plusieurs distributions couramment utilisées. Ward et al. (1981) suggèrent que les distributions normale et log-normale sont les plus largement applicables à la qualité de l'eau, mais que les distributions gamma, Gumbel et log-Gumbel sont parfois les plus appropriées. Les distributions normale et log-normale présentent un intérêt particulier pour le personnel chargé de la réglementation pour deux raisons. Premièrement, elles modélisent le comportement de nombreuses variables de qualité de l'eau de manière acceptable. Deuxièmement, les paramètres de ces distributions (la moyenne et la variance) sont faciles à estimer, et les

tables de la fonction de distribution normale cumulative sont largement disponibles. Ainsi, ces distributions sont faciles à utiliser (Loftis et al, 1983).

3.3. Les modèles basés sur l'intelligence artificielle

Le besoin de prédictions précises de la qualité de l'eau dans les rivières a encouragé les chercheurs à développer de nouvelles méthodes et à améliorer la capacité prédictive des modèles conventionnels. Ces dernières années, les méthodes basées sur l'intelligence artificielle (IA) ont été reconnues comme très puissantes à cette fin.

3.3.1. Les réseaux de neurones artificiels (RNA)

Les RNA sont des modèles d'entrée-sortie basés sur une représentation mathématique simpliste des neurones biologiques qui font spécifiquement référence au cerveau humain. Le but initial de cette modélisation est de reproduire les capacités du cerveau humain à interpoler ou à classifier. En général, les RNA peuvent être représentés comme un système de "neurones" interconnectés qui constituent la base du fonctionnement des réseaux de neurones, les poids des connections entre les neurones sont utilisés pour mémoriser les connaissances. Les RNA sont généralement divisés en trois couches : entrée, la ou les couche(s) cachée(s) et sortie.

Selon (Haykin, 1994), un RNA est un processus distribué de manière massivement parallèle, conçu pour mémoriser des connaissances de façon expérimentale et de les rendre disponibles pour utilisation.

Les RNA dépendent fortement de la quantité de données. Il n'est donc pas recommandé d'utiliser des données de taille relativement faible pour les prédicteurs (entrées). En effet, certaines informations utiles sont perdues dans les données à court terme, ce qui peut conduire à de mauvais résultats de prédiction. En outre, la division des données ainsi que le choix de l'algorithme d'apprentissage sont des étapes primordiales dans le processus de modélisation qui permettent de calibrer les paramètres du modèle (par exemple, les poids de connexion) afin que le réseau puisse approximer une relation entrée-sortie non linéaire complexe. L'algorithme de Levenberg-Marquardt et l'algorithme de rétropropagation (BP) sont les algorithmes les plus couramment utilisés.

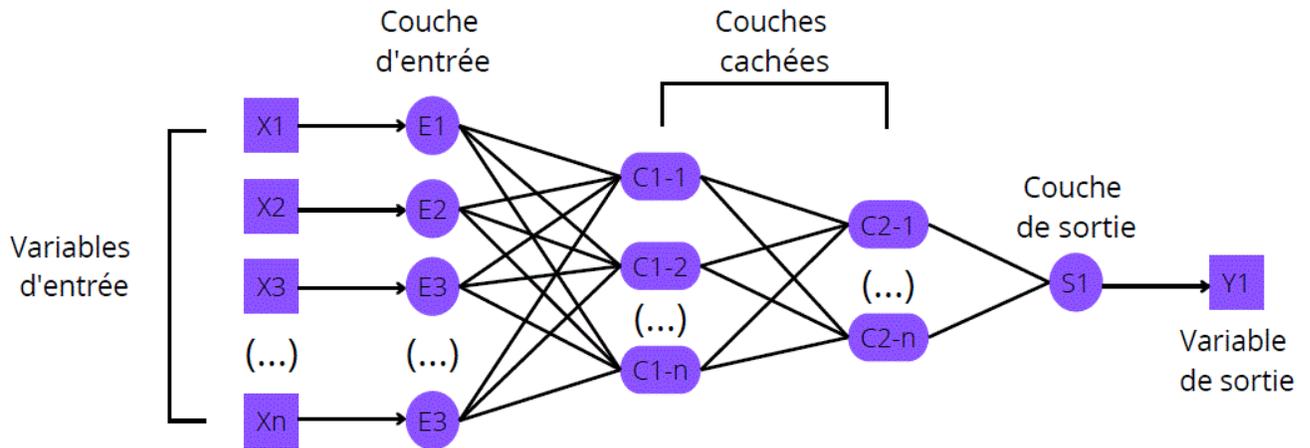


Figure 2: Représentation schématique d'un réseau de neurones artificiels

3.3.2. Les différentes architectures

L'architecture d'un modèle fait référence à la structure globale et à la manière dont les informations circulent d'une couche à l'autre. Il existe plusieurs architectures différentes (Chen et al, 2020). Les trois architectures de modèles les plus utilisées pour la prédiction de la qualité de l'eau sont : les modèles non bouclés connus sous le nom de "feed-forward", les modèles des réseaux récurrents et les modèles hybrides.

Feed-forward

Le réseau feedforward est un réseau non bouclé qui ne présente aucune rétroaction entre les neurones des différentes couches. Le terme "feed-forward" signifie qu'une connexion neuronale n'existe qu'entre un neurone de la couche d'entrée et d'autres neurones de la couche cachée ou entre un neurone de la couche cachée et des neurones de la couche de sortie. Cependant, les neurones d'une couche ne sont pas interconnectés. Il existe plusieurs sous-classes de l'architecture feed-forward dont « MLP » qui fait partie des architectures

les plus utilisées dans de nombreux types d'ARN, suivies des BPNN qui utilisent les algorithmes de rétropropagation pour former les réseaux. Les autres architectures couramment utilisées dans la prédiction de la qualité de l'eau comprennent les TDNN, les RBFNN, les GRNN, les WNN, les ELM, les CCNN et les MNN. Pour plus de détails sur ces architectures, l'excellent ouvrage de référence de Haykin (1994) peut être consulté.

Architectures récurrentes

Les réseaux récurrents RNN, peuvent être totalement ou partiellement bouclés, les neurones d'une couche étant connectés aux noeuds de la couche suivante, de la couche précédente et de la même couche. Les réseaux feed-forward sont donc des cas spéciaux de réseaux récurrents.

Les réseaux récurrents sont particulièrement efficaces pour modéliser les séries temporelles, communes aux applications de prévision et de prédiction (Zhang et Stanley, 1997, Maier et Dandy, 2001, Wilson et Recknagel, 2001). Ils présentent également l'avantage non négligeable de modéliser les propriétés dynamiques implicitement, alors que les réseaux feed-forward ont besoin que les systèmes dynamiques soient traités explicitement par l'ajout de variables de décalage (Maier et Dandy, 2001). Mais ils prennent aussi plus de temps à se développer car leur vitesse d'apprentissage est plus lente.

Différents types de RNN sont développés pour que les réseaux de neurones aient une meilleure capacité de mémoire, tel que les LSTM, dont l'amélioration consiste à l'ajout d'un processeur appelé « état de la cellule mémoire » à sa couche cachée pour déterminer si l'information est utile ou non. Les TLRN caractérisés par une faible sensibilité au bruit et d'une capacité de stockage adaptative. Les réseaux NARX, les réseaux de neurones Elman ainsi que les ESN.

Contrairement aux réseaux de neurones non bouclés dont le graphe de connexions est acyclique, les réseaux de neurones bouclés peuvent avoir une topologie de connexions quelconque, comprenant notamment des boucles qui ramènent aux entrées la valeur d'une ou plusieurs sorties. Un réseau de neurones bouclé est un système dynamique, régi par des équations différentielles.

Architectures hybrides

Des modèles RNA hybrides, sont de plus en plus utilisés et jouent un rôle important dans la modélisation grâce à leur capacité à s'intégrer à d'autres techniques de modélisation conventionnelles et plus avancées ce qui permet de créer des modèles flexibles et efficaces. Les modèles hybrides sont divisés en trois catégories, à savoir les modèles intensifs, les techniques intensives et les données intensives. Les approches à modèles intensifs modélisent les sous-composants du système physique global et agrègent la réponse globale de chaque modèle. Il existe différentes formes de modèles hybrides, tel que LSTM-RNN ou FNN-WNN, qui sont des modèles intensifs. Les modèles de type « technique-intensive » consistent à développer un cadre de modélisation capable d'intégrer différentes technologies. Les méthodes qui combinent des approches d'ensemble ou des modèles de séries chronologiques qui suppriment les tendances ou les périodicités, comme les (ARIMA-RBFNN) ou ARIMA-ANN qui sont des méthodes techniques-intensives.

Tableau 2: Utilisations des modèles de RNA dans la modélisation de la qualité de l'eau

Architecture	Méthode	Avantages	Références
Feed-forward	MLP	Permet de résoudre les problèmes non linéaires.	(Iglesias et al, 2014) (Emamgholizadeh et al, 2014) (Najah et al, 2012) (Gazzaz et al, 2012) (Dogan et al, 2009)
	MNN	Choisit le réseau neuronal qui a le maximum de similarité entre les entrées et les centroïdes du cluster et résout les problèmes de faible précision de prédiction.	(Yeon et al, 2008)
	BPNN	C'est une méthode flexible car elle ne nécessite pas de connaissances préalables sur le réseau. Il s'agit d'une méthode standard qui fonctionne généralement bien	(Miao et al, 2009) (Chen et al, 2010)

		Selon le théorème de Kolmogorov, un BPNN peut approximer n'importe quelle fonction continue. Ainsi, le BPNN est capable de réaliser n'importe quelle cartographie non linéaire complexe.	
	RBFNN	Permet de surmonter les problèmes de minimum local.	(Faramarzi et al, 2014) (Najah et al, 2012)
	GRNN	Une forme modifiée du modèle RBFNN, contient un motif et une couche de sommation entre les couches d'entrée et de sortie, il permet de résoudre les problèmes liés aux petits échantillons.	(Antanasijević et al, 2014)
	ANFIS	Combine les avantages des réseaux de neurones artificiels (ANN) et de la logique floue (FL) dans un cadre unique. Il offre une capacité d'apprentissage accélérée et des capacités d'interprétation adaptative pour modéliser des modèles complexes et appréhender les relations non linéaires.	(Emamgholizadeh et al, 2014) (Yeon et al, 2008)
Récurrents	RNN	Résout les problèmes de dépendance à long terme qui ne sont pas captés par le réseau feedforward.	(Lim et al, 2019)
	LSTM	Sa structure est similaire aux RNN, l'état de la cellule mémoire est ajouté à la couche cachée. Aide à résoudre le problème de la disparition du gradient de fuite des RNN.	(Lim et al, 2019)
	NARX	Résout les problèmes de dépendance à long terme.	(Chang et al, 2016)
	TLRN	Sa structure est similaire aux MLP, il a les connexions récurrentes locales dans la couche cachée. Réduit l'influence du bruit et possède l'avantage de la profondeur de mémoire adaptative.	(Evrendilek and Karakaya, 2014)

	ESN	<p>Sa structure est composée de trois couches : les couches d'entrée, de réservoir et de lecture.</p> <p>Permet de surmonter les problèmes des minima locaux et de la disparition du gradient.</p>	(Liu et al, 2012)
	RESN	<p>Une méthode basée sur la structure de l'ESN.</p> <p>Étant donné que les poids de sortie sont calculés par l'inverse de Moore-Penrose, le problème mal posé peut exister dans l'ESN. Pour surmonter ce problème, RESN est proposé, dans lequel l'algorithme de régression de crête est utilisé pour calculer les poids de sortie au lieu de la régression linéaire.</p>	(Liu et al, 2012)
	Elman	<p>Une couche de contexte qui peut stocker les états internes est ajoutée aux trois couches traditionnelles. Elle est utile dans la modélisation des systèmes dynamiques grâce à la couche de contexte.</p>	<p>(Asadollahfardi et al, 2012)</p> <p>(Antanasijević et al, 2013)</p>
hybrides	<p>LSTM_RNN</p> <p>FNN_WNN</p> <p>ARIMA-RBFNN</p> <p>ARIMA-ANN</p> <p>PCA-BPNN</p> <p>EMD-BPNN...</p>	<p>C'est une combinaison de méthodes conventionnelles. Permet l'intégration interne des méthodes ANN et d'explorer les avantages de chaque méthode.</p>	

4. CONCLUSION

La présente revue permet de mettre en exergue les atouts et les inconvénients de différents modèles de qualité de l'eau. De manière générale, la revue permet de constater la grande

diversité d'outils existants et permet de contraster les approches déterministes et statistiques.

Le bon choix d'un modèle fait partie du travail de planification et de la gestion qui compte parmi les étapes les plus importantes d'une étude. Ainsi, ce rapport cherche à aider dans ce processus. Un critère qui est le plus souvent priorisé dans le choix du modèle est le niveau de complexité. La comparaison des modèles déterministes permet au lecteur de prioriser les outils plus ou moins parcimonieux selon le nombre d'intrants et de paramètres et selon le niveau de complexité requis.

Les modèles peuvent être classés par niveau de complexité en tenant compte des critères suivants : nombre de variables analysées, dimensions, types de processus et types de résultats.

Chinyama et al (2013), Sharma et Kansal (2013) et Costa et al, (2019) décrivent les modèles 3D comme trop complexes, ainsi, les modèles WASP, AQUATOX, EFDC, BASINS et HSPF peuvent être considérés comme des modèles de grande complexité. Les modèles CE-QUAL-W2 et SWAT sont des modèles de complexité intermédiaire (Costa et al, 2019). CE-QUAL-W2 présente certains avantages : c'est un modèle en source libre qui permet de simuler environ 40 variables. Il est recommandé quand des fines résolutions temporelles sont demandées mais son application est complexe et chronophage. Le modèle SWAT compte parmi les modèles de qualité de l'eau les plus utilisés au Canada. Il est capable de prédire l'impact des pratiques de gestion des terres sur l'eau, les sédiments dans des grands bassins versants. Il comprend également des interfaces SIG favorisant la visualisation de la relation spatiale entre les sous-bassins.

Les QUALs et SPARROW sont des modèles de complexité inférieur. Les QUALs sont largement utilisés mais plus précisément dans les cas de petits bassins versants et est insensible aux variations des conditions hydrologiques et aux taux de rejets. SPARROW est un modèle relativement simple qui permet de simuler différentes variables, applicable même en cas de données insuffisants et convient à la modélisation de longues périodes allant de plusieurs années jusqu'à des décennies. Ainsi, ce modèle est probablement le modèle déterministe le mieux adapté pour répondre aux exigences de la présente étude.

Parmi les modèles statistiques, il faut rappeler qu'un sous-ensemble de modèles (modèles statistiques spatiaux) sont les seules approches empiriques qui permettent une prise en compte directe et explicite de la variabilité spatiale de la qualité de l'eau, comme le font la majorité des modèles déterministes. Cette catégorie de modèles pourrait être mise en œuvre en complémentarité à un modèle déterministe, selon la variabilité dans la quantité de données dont on dispose d'un endroit à l'autre. Les réseaux de neurones artificiels offrent une grande flexibilité, mais doivent être entraînés avec de très grandes bases de données.

5. BIBLIOGRAPHIE

- Ambrose Jr, R. B., Wool, T. A., & Barnwell Jr, T. O. (2009). Development of water quality modeling in the United States. *Environmental Engineering Research*, 14(4), 200-210.
- Angello, Z. A., Behailu, B. M., & Tränckner, J. (2021). Selection of Optimum Pollution Load Reduction and Water Quality Improvement Approaches Using Scenario Based Water Quality Modeling in Little Akaki River, Ethiopia. *Water*, 13(5), 584.
- Antanasijević, D., Pocajt, V., Perić-Grujić, A., & Ristić, M. (2014). Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis. *Journal of Hydrology*, 519, 1895-1907.
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development 1. *JAWRA Journal of the American Water Resources Association*, 34(1), 73-89.
- Asadollahfardi, G., Taklify, A., & Ghanbari, A. (2012). Application of artificial neural network to predict TDS in Talkheh Rud River. *Journal of Irrigation and Drainage Engineering*, 138(4), 363.
- Avila, R., Horn, B., Moriarty, E., Hodson, R., & Moltchanova, E. (2018). Evaluating statistical model performance in water quality prediction. *Journal of environmental management*, 206, 910-919.
- Bingli, L., Huang, S., Min, Q., Tianyun, L. I., & Zijian, W. A. N. G. (2008). Prediction of the environmental fate and aquatic ecological impact of nitrobenzene in the Songhua River using the modified AQUATOX model. *Journal of Environmental Sciences*, 20(7), 769-777.
- Bo, W., Wang, X., Zhang, Q., Xiao, Y., & Ouyang, Z. (2018). Influence of land use and point source pollution on water quality in a developed region: a case study in Shunde, China. *International Journal of Environmental Research and Public Health*, 15(1), 51.
- Booty, W., & Benoy, G. (2009). Multicriteria review of nonpoint source water quality models for nutrients, sediments, and pathogens. *Water Quality Research Journal*, 44(4), 365-377.
- Borsuk, M. E., Stow, C. A., & Reckhow, K. H. (2002). Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. *Environmental Science & Technology*, 36 (10), 2109-2115
- Brown, L. C., & Barnwell, T. O. (1987). *The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual* (p. 189). Athens: US Environmental Protection Agency. Office of Research and Development. Environmental Research Laboratory.
- Burigato Costa, C. M. D. S., da Silva Marques, L., Almeida, A. K., Leite, I. R., & de Almeida, I. K. (2019). Applicability of water quality models around the world—a review. *Environmental Science and Pollution Research*, 26(36), 36141-36162.

- Chang, F. J., Chen, P. A., Chang, L. C., & Tsai, Y. H. (2016). Estimating spatio-temporal dynamics of stream total phosphate concentration by soft computing techniques. *Science of the Total Environment*, 562, 228-236.
- Chang, F. J., Chen, P. A., Liu, C. W., Liao, V. H. C., & Liao, C. M. (2013). Regional estimation of groundwater arsenic concentrations through systematical dynamic-neural modeling. *Journal of Hydrology*, 499, 265-274.
- Chen, Q., Mei, K., Dahlgren, R. A., Wang, T., Gong, J., & Zhang, M. (2016). Impacts of land use and population density on seasonal surface water quality using a modified geographically weighted regression. *Science of the total environment*, 572, 450-466.
- Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. *Applied Sciences*, 10(17), 5776.
- Chen, W. B., & Liu, W. C. (2014). Artificial neural network modeling of dissolved oxygen in reservoir. *Environmental Monitoring and Assessment*, 186(2), 1203-1217.
- Costa, C. M. D. S. B., Leite, I. R., Almeida, A. K., & de Almeida, I. K. (2021). Choosing an appropriate water quality model—a review. *Environmental Monitoring and Assessment*, 193(1), 1-15.
- Di Toro, D. M., Fitzpatrick, J. J., & Thomann, R. V. (1983). Documentation for water quality analysis simulation program (WASP) and model verification program (MVP).
- Dogan, E., Sengorur, B., & Koklu, R. (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management*, 90(2), 1229-1235.
- Ejigu, M. T. (2021). Overview of water quality modeling. *Cogent Engineering*, 8(1), 1891711.
- Emamgholizadeh, S., Kashi, H., Marofpoor, I., & Zalaghi, E. (2014). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology*, 11(3), 645-656.
- Evrendilek, F., & Karakaya, N. (2014). Monitoring diel dissolved oxygen dynamics through integrating wavelet denoising and temporal neural networks. *Environmental monitoring and assessment*, 186(3), 1583-1591.
- Faruk, D. Ö. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering applications of artificial intelligence*, 23(4), 586-594.
- Gamble, A., & Babbar-Sebens, M. (2012). On the use of multivariate statistical methods for combining in-stream monitoring data and spatial analysis to characterize water quality conditions in the White River Basin, Indiana, USA. *Environmental Monitoring and Assessment*, 184(2), 845-875.
- Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine pollution bulletin*, 64(11), 2409-2420.

- Gronewold, A. D., Qian, S. S., Wolpert, R. L., & Reckhow, K. H. (2009). Calibrating and validating bacterial water quality models: A Bayesian approach. *Water Research*, 43(10), 2688-2698.
- Hajjigholizadeh, M. (2016). Water quality modelling using multivariate statistical analysis and remote sensing in South Florida.
- Hamrick, J. M. (1992). A three-dimensional environmental fluid dynamics computer code: Theoretical and computational aspects. FIU Electronic Theses and Dissertations. 2992.
- Han, Q., Tong, R., Sun, W., Zhao, Y., Yu, J., Wang, G., & Jin, Y. (2020). Anthropogenic influences on the water quality of the Baiyangdian Lake in North China over the last decade. *Science of the Total Environment*, 701, 134929.
- Heddam, S. (2016). Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: a case study from the Klamath River, Oregon, USA. *Modeling Earth Systems and Environment*, 2(3), 1-18.
- Isaak, D. J., Peterson, E. E., Ver Hoef, J. M., Wenger, S. J., Falke, J. A., Torgersen, C. E., ... & Monestiez, P. (2014). Applications of spatial statistical network models to stream data. *Wiley Interdisciplinary Reviews: Water*, 1(3), 277-294.
- Isiyaka, H. A., Mustapha, A., Juahir, H., & Phil-Eze, P. (2019). Water quality modelling using artificial neural network and multivariate statistical techniques. *Modeling Earth Systems and Environment*, 5(2), 583-593.
- Kannel, P. R., Kanel, S. R., Lee, S., Lee, Y. S., & Gan, T. Y. (2011). A review of public domain water quality models for simulating dissolved oxygen in rivers and streams. *Environmental Modeling & Assessment*, 16(2), 183-204.
- Kim, D., Kim, Y., & Kim, B. (2019). Simulation of eutrophication in a reservoir by CE-QUAL-W2 for the evaluation of the importance of point sources and summer monsoon. *Lake and Reservoir Management*, 35(1), 64-76.
- Lafdani, E. K., Nia, A. M., & Ahmadi, A. (2013). Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Journal of Hydrology*, 478, 50-62.
- Lam, Q. D., Schmalz, B., & Fohrer, N. (2010). Modelling point and diffuse source pollution of nitrate in a rural lowland catchment using the SWAT model. *Agricultural Water Management*, 97(2), 317-325.
- Lim, H., An, H., Kim, H., & Lee, J. (2019). Prediction of pollution loads in the Geum River upstream using the recurrent neural network algorithm. *Korean Journal of Agricultural Science*, 46(1), 67-78.
- Liu, S., Yan, M., Tai, H., Xu, L., & Li, D. (2011). Prediction of dissolved oxygen content in aquaculture of *Hyriopsis cumingii* using Elman neural network. In *International Conference on Computer and Computing Technologies in Agriculture* (pp. 508-518). Springer, Berlin, Heidelberg.
- Loftis, J. C., Ward, R. C., & Smillie, G. M. (1983). Statistical models for water quality regulation. *Journal (Water Pollution Control Federation)*, 1098-1104.

- Mainali, J., Chang, H., & Chun, Y. (2019). A review of spatial statistical approaches to modeling water quality. *Progress in Physical Geography: Earth and Environment*, 43(6), 801-826.
- Markus, M., Hejazi, M. I., Bajcsy, P., Giustolisi, O., & Savic, D. A. (2010). Prediction of weekly nitrate-N fluctuations in a small agricultural watershed in Illinois. *Journal of Hydroinformatics*, 12(3), 251-261.
- Masoumi, F., Afshar, A., & Palatkaleh, S. T. (2016). Selective withdrawal optimization in river-reservoir systems; trade-offs between maximum allowable receiving waste load and water quality criteria enhancement. *Environmental monitoring and assessment*, 188(7), 1-16.
- Miao, Q., Yuan, H., Shao, C., & Liu, Z. (2009, June). Water quality prediction of moshui river in china based on BP neural network. In *2009 International Conference on Computational Intelligence and Natural Computing* (Vol. 1, pp. 7-10). IEEE.
- Mustapha, A., Aris, A. Z., Yusoff, F. M., Zakaria, M. P., Ramli, M. F., Abdullah, A. M. & Narany, T. S. (2014). Statistical approach in determining the spatial changes of surface water quality at the upper course of Kano River, Nigeria. *Water Quality, Exposure and Health*, 6(3), 127-142.
- Najah, A., El-Shafie, A., Karim, O. A., & El-Shafie, A. H. (2014). Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environmental Science and Pollution Research*, 21(3), 1658-1670.
- Neitsch SL, Arnold J, Kiniry J, Williams J, King K (2005) Soil and water assessment tool theoretical documentation version 2005 Texas, USA
- Ning, S. K., Chang, N. B., Yang, L., Chen, H. W., & Hsu, H. Y. (2001). Assessing pollution prevention program by QUAL2E simulation analysis for the Kao-Ping River Basin, Taiwan. *Journal of Environmental Management*, 61(1), 61-76.
- Ogwueleka, T. C. (2015). Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. *Environmental monitoring and assessment*, 187(3), 1-17.
- Oliveira Souza da Costa, A., Ferreira Silva, P., Godoy Sabará, M., & Ferreira da Costa, E. (2009). Use of neural networks for monitoring surface water quality changes in a neotropical urban stream. *Environmental monitoring and assessment*, 155(1), 527-538.
- Park RA, Clough JS (2004) a. Modeling environmental fate and ecological effects in aquatic ecosystem. Volume 1: User's Manual, U.S. Environmental Protection Agency Office of Water Office of Science and Technology.
- Park, R. A., Clough, J. S., & Wellman, M. C. (2008). AQUATOX: Modeling environmental fate and ecological effects in aquatic ecosystems. *Ecological Modelling*, 213(1), 1-15.
- Parmar, K. S., & Bhardwaj, R. (2014). Water quality management using statistical analysis and time-series prediction model. *Applied Water Science*, 4(4), 425-434.
- Parmar, K. S., & Bhardwaj, R. (2015). River water prediction modeling using neural networks, fuzzy and wavelet coupled model. *Water resources management*, 29(1), 17-33.

- Pelletier, G. J., Chapra, S. C., & Tao, H. (2006). QUAL2Kw—A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. *Environmental Modelling & Software*, 21(3), 419-425.
- Pohlert, T., Huisman, J. A., Breuer, L., & Frede, H. G. (2005). Modelling of point and non-point source pollution of nitrate with SWAT in the river Dill, Germany. *Advances in Geosciences*, 5, 7-12.
- Ravansalar, M., & Rajaei, T. (2015). Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. *Environmental Monitoring and Assessment*, 187(6), 1-16.
- Rose, P. E., & Pedersen, J. A. (2005). Fate of oxytetracycline in streams receiving aquaculture discharges: model simulations. *Environmental Toxicology and Chemistry: An International Journal*, 24(1), 40-50.
- Schwarz, G. E., Hoos, A. B., Alexander, R. B., & Smith, R. A. (2006). The SPARROW surface water-quality model: theory, application and user documentation.
- Shoemaker, L., Dai, T., Koenig, J., & Hantush, M. (2005). *TMDL model evaluation and research needs*. Washington, DC, USA: National Risk Management Research Laboratory, US Environmental Protection Agency.
- Shokri, A., Bozorg Haddad, O., & Mariño, M. A. (2014). Multi-objective quantity–quality reservoir operation in sudden pollution. *Water Resources Management*, 28(2), 567-586.
- Simeonova, P., Simeonov, V., & Andreev, G. (2003). Water quality study of the Struma river basin, Bulgaria (1989–1998). *Open Chemistry*, 1(2), 121-136.
- Smith, R. A., Schwarz, G. E., & Alexander, R. B. (1997). Regional interpretation of water-quality monitoring data. *Water resources research*, 33(12), 2781-2798.
- Strokal, M., Spanier, J. E., Kroeze, C., Koelmans, A. A., Flörke, M., Franssen, W., & Williams, R. (2019). Global multi-pollutant modelling of water quality: scientific challenges and future directions. *Current Opinion in Environmental Sustainability*, 36, 116-125.
- Van, L. A., Nguyen, K. D., Le Marrec, F., & Jairy, A. (2022). Development of a Tool for Modeling the Fecal Contamination in Rivers with Turbulent Flows—Application to the Seine et Marne Rivers (Parisian Region, France). *Water*, 14(8), 1191.
- Varol, M. (2020). Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: A case study. *Environmental Pollution*, 266, 115417.
- Wang, Q., Li, S., Jia, P., Qi, C., & Ding, F. (2013). A review of surface water quality models. *The Scientific World Journal*, 2013.
- Ward, R. C., Loftis, J. C., & Averett, R. C. (1989). Monitoring systems for water quality. *Critical Reviews in Environmental Science and Technology*, 19(2), 101-118.
- Wellen, C., Arhonditsis, G. B., Labencki, T., & Boyd, D. (2014). Application of the SPARROW model in watersheds with limited information: a Bayesian assessment of the model uncertainty and the value of additional monitoring. *Hydrological Processes*, 28(3), 1260-1283.

Wu, G., & Xu, Z. (2011). Prediction of algal blooming using EFDC model: Case study in the Daoxiang Lake. *Ecological Modelling*, 222(6), 1245-1252.

Yeon, I.S.; Kim, J.H.; Jun, K.W. (2008). Application of artificial intelligence models in water quality forecasting. *Environ. Technol*, 29, 625–631.

Yuan, L., Sinshaw, T., & Forshay, K. J. (2020). Review of watershed-scale water quality and nonpoint source pollution models. *Geosciences*, 10(1), 25.

Zhang, J., & Ross, M. (2012). Hydrologic simulation of clay-settling areas in the phosphate mining district, Florida. *Hydrological Processes*, 26(24), 3770-3778.

Zhang, M. L., Shen, Y. M., & Guo, Y. (2008). Development and application of a eutrophication water quality model for river networks. *Journal of Hydrodynamics*, 20(6), 719-726.

Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q., & Ye, F. (2009). Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. *Environmental monitoring and assessment*, 152(1), 123-131.

Zhang, X., Jiang, H., & Zhang, Y. (2013). Spatial distribution and source identification of persistent pollutants in marine sediments of Hong Kong. *Environmental monitoring and assessment*, 185(6), 4693-4704.

Zhu Y, Liang ZW, Li W, Yang Y, Yang MY, Mao W, Xu HL, Wu WX. (2013). Watershed water environment pollution models and their applications: a review. *Ying Yong Sheng Tai Xue Bao*. 2013 Oct;24(10):3012-8. Chinese. PMID: 24483100.

Zhu, J., Wang, X., Zhang, L., Cheng, H., & Yang, Z. (2015). System dynamics modeling of the influence of the TN/TP concentrations in socioeconomic water on NDVI in shallow lakes. *Ecological Engineering*, 76, 27-35.

Zieminska-Stolarska, A., & Skrzypski, J. (2012). Review of mathematical models of water quality. *Ecological Chemistry and Engineering*, 19(2), 197.