

VERS UNE OPTIMISATION DE LA SURVEILLANCE DE LA QUALITÉ DES SOURCES D'EAU POTABLE PAR LA VALORISATION DE DONNÉES HAUTE FRÉQUENCE

Ianis Delpla, Mihai Florea, Manuel J. Rodriguez

Contexte

- Un cadre réglementaire de plus en plus strict au Québec (RQEP, RPEP,...)
- Production importante de données :
 - Suivi réglementaire, autosurveillance
 - Projets de recherche (polluants émergents, variabilité spatio-temporelle)
- MAIS utilisation ponctuelle des données & difficultés d'interprétations des grandes bases de données ^[1]
- Les stratégies actuelles de suivi de la qualité de l'eau brute permettent seulement de réagir à des épisodes de contamination à posteriori
 - Délai entre la prise d'un échantillon d'eau et son analyse, ainsi qu'entre la réaction des opérateurs et ses effets
 - Supervision manuelle : basé sur le suivi visuel des variables de qualité
 - Anticipation requise → intérêt d'outils d'optimisation de la surveillance

Outils d'optimisation de la surveillance de la qualité de l'eau potable

- Optimisation de la surveillance de la qualité de l'eau potable
 - Analyse rapide (Surveillance en ligne)
 - Systèmes intelligents
 - Modélisation spatiale (représentation)
 - Modélisation temporelle (alerte précoce) : réseaux de neurones, analyse de tendance, régression, logique floue...
- Intérêt d'une approche combinant l'analyse rapide d'indicateurs et la modélisation temporelle pour réduire le temps de détection d'une contamination

Paramètre indicateur de la qualité de l'eau

- Turbidité
- Indicateur de qualité microbiologique et chimique de l'eau
 - Présence de solides en suspension, matière organique, microorganismes potentiellement pathogènes (Ex : Milwaukee, USA, 1993)
- Influence l'efficacité du traitement et ses coûts associés
- Contrôlée par plusieurs facteurs :
 - Pédologie
 - Occupation du sol (agricole, urbain)
 - Climat
- Influence du climat sur la turbidité :
 - Épisodes de pluie peuvent conduire à des hausses rapides de la turbidité
 - Effets négatifs sur l'efficacité de traitement, coûts de traitement (quantité de réactifs) et la qualité de l'eau

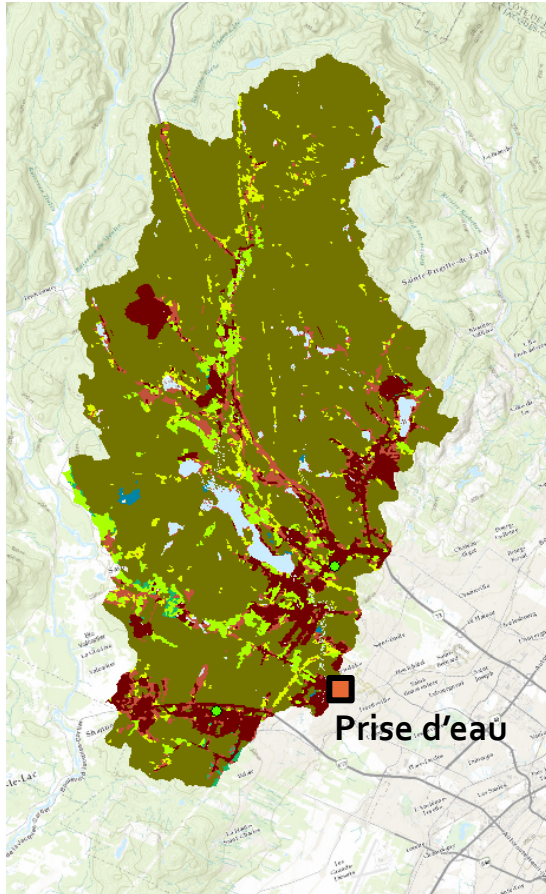


Hypothèse : les techniques d'exploration de données peuvent servir à définir un système d'alerte pour le suivi de la qualité de la ressource en eau pour le paramètre turbidité ⁴

Objectifs de l'étude

- Elaborer un système d'alerte pour le paramètre turbidité utilisant des données facilement accessibles et peu coûteuses à acquérir et en appliquant différentes techniques d'analyse statistique des données
 - Prévision court terme : épisodes de forte turbidité à la prise d'eau (analyse de tendance)
 - Prévision long terme : séries temporelles (réseau de neurones avec indicateurs pluviométriques)

Zone d'étude

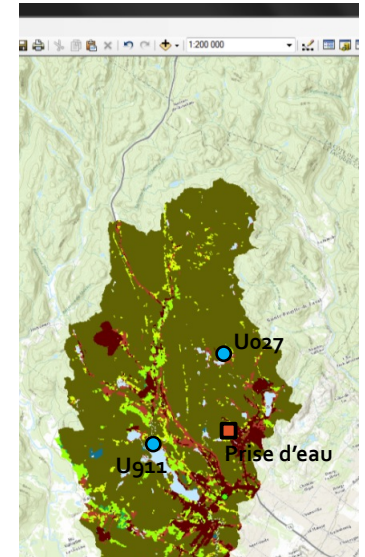


- Bassin versant de la rivière Saint-Charles
- Occupation du sol sur le BV :
 - forte pression urbaine (étalement urbain, présence d'usines de traitement des eaux usées, ...)
 - forestière (amont)
- Hausse de la turbidité dans les affluents ^[1]
- Source d'eau potable principale de la Ville de Québec (alimente 306 000 habitants)



Type de données

- Qualité de l'eau
 - Turbidité en ligne
 - Prise d'eau potable : 1 donnée/minute
- Pluviométrie:
 - 2 stations pluviométriques
 - Réseau de pluviomètres de la Ville de Québec : 1 donnée/5min
- Durée : 54 mois (01/2012- 06/2016)
- Correction des données : suppression des valeurs aberrantes et calcul des valeurs manquantes par interpolation



Méthodes d'analyse statistique

Analyse de tendance

- **Analyse de tendance**
- Anticipation des pics : Permet de détecter la direction, la vitesse et l'amplitude du changement de la variable de qualité de l'eau
 - Étude des chroniques de turbidité, pour des épisodes de turbidité de niveau élevé (>99^{ème} percentile de la distribution des données : 12,1 NTU)
 - Analyse sur des pas de temps courts (5 minutes)
- Calcul de l'indice de tendance : $T_i = \frac{1}{2}(X + dX/dt)$
Avec $X(t) = X(t) - X(t-n)/n$, avec un pas $n=10$

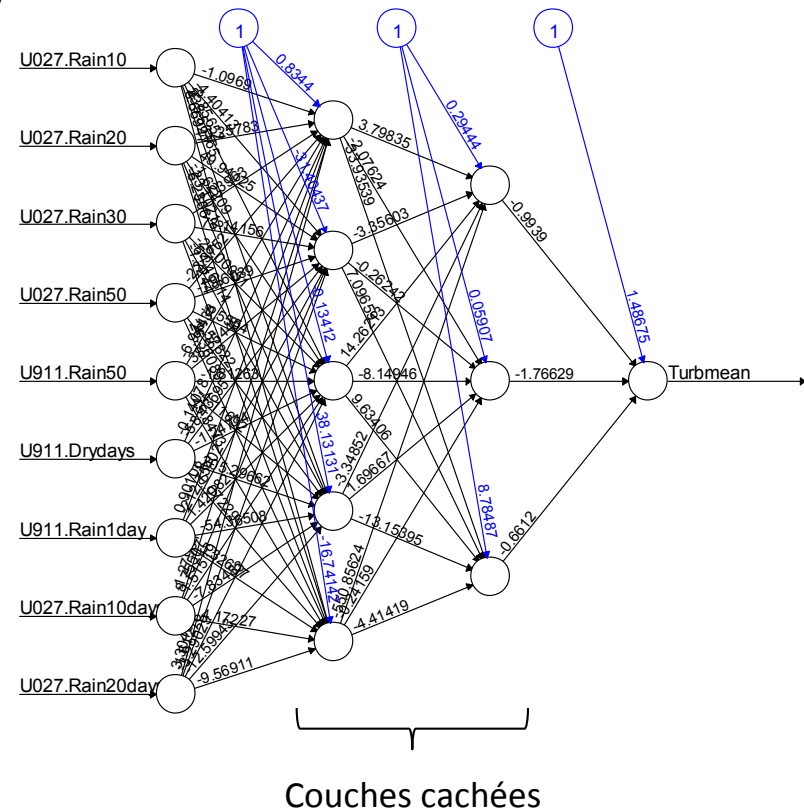
Indices pluviométriques

- Création d'indices pluviométriques pour servir de variables explicatives de la turbidité
- 2 stations pluviométriques
- Types d'indices pluviométriques
 - Nombre de jours secs depuis la dernière pluie (jours secs : pluvio 24h < 1mm)
 - Nombre de jours depuis une pluviométrie journalière ≥ 10 , ≥ 20 , ≥ 30 , ≥ 50 mm
 - Somme des précipitations journalières sur 1, 2, 5, 10 ou 20 jours
- Réduction du nombre de variables explicatives :
 - Calcul par régression linéaire multiple du poids de chaque variable avec la variable réponse (turbidité) et évaluation de l'autocorrélation entre les variables
 - Exclusion des variables avec un facteur d'inflation de la variance (VIF) >5 et ayant un coefficient de corrélation partielle avec la turbidité <1
 - 14 variables retenues

Méthodes d'analyse statistique

Réseau de neurones (1)

- **Réseau de neurones artificiel**
 - Méthode supervisée (rétropropagation)
 - 1 couche d'entrée (indices pluviométriques)
 - 1 couche de sortie (turbidité)
 - 1 ou plusieurs couches dites cachées
 - Bias layers (bleu) appliquent une constante (pondération) à chaque neuron
 - Chaque neuron somme l'information pondérée provenant de l'amont qui est ensuite transformée par une fonction d'activation
 - Processus itératif où l'erreur de prédiction est calculée puis minimisée



Méthodes d'analyse statistique

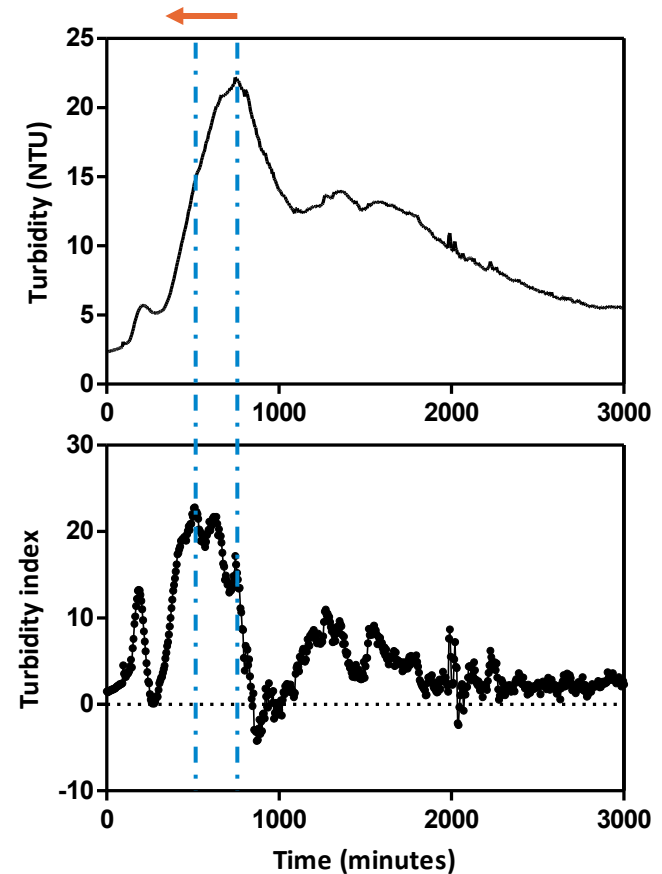
Réseau de neurones (2)

- Appliqué sur des données journalières
- Normalisation des données : Méthode min-max: mise à l'échelle des données dans l'intervalle [0,1]
- Optimisation de la configuration du réseau de neurones
 - Test des différentes configurations : nombre de couches cachées et nombre de neurons par couche
 - Choix de la configuration minimisant l'erreur quadratique moyenne (MSE) : $\sum(\text{Obs-mod})^2/N$
 - Validation par comparaison avec une méthode de régression linéaire généralisée

Résultats

Analyse de tendance

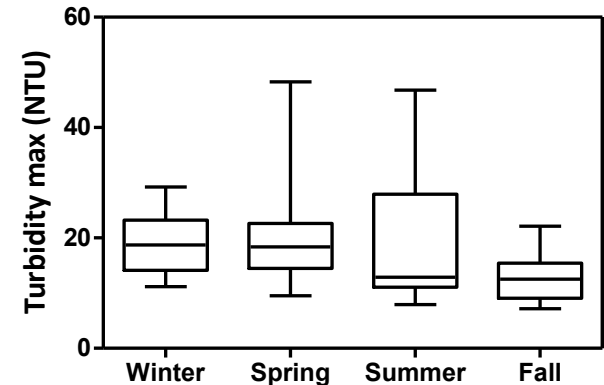
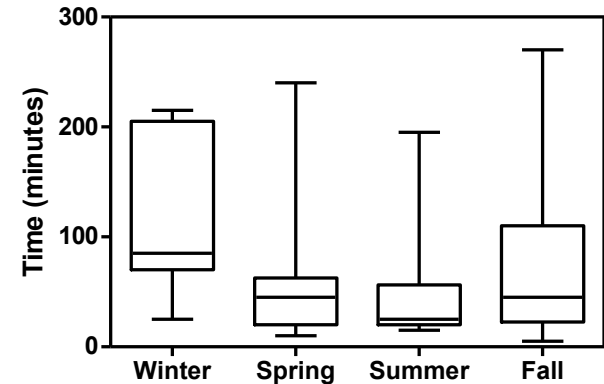
- Valeurs extrêmes de la distribution de données sur la période 2012-2016 (percentile 99th = 12.1 NTU)
- Au total, 44 pics de turbidité retenus
- Anticipation du pic de turbidité:
 - le pic de turbidité est à 745 min et l'indice de turbidité indique que le 1er pic est atteint à 510 min, soit une anticipation de 235 minutes



Résultats

Analyse de tendance

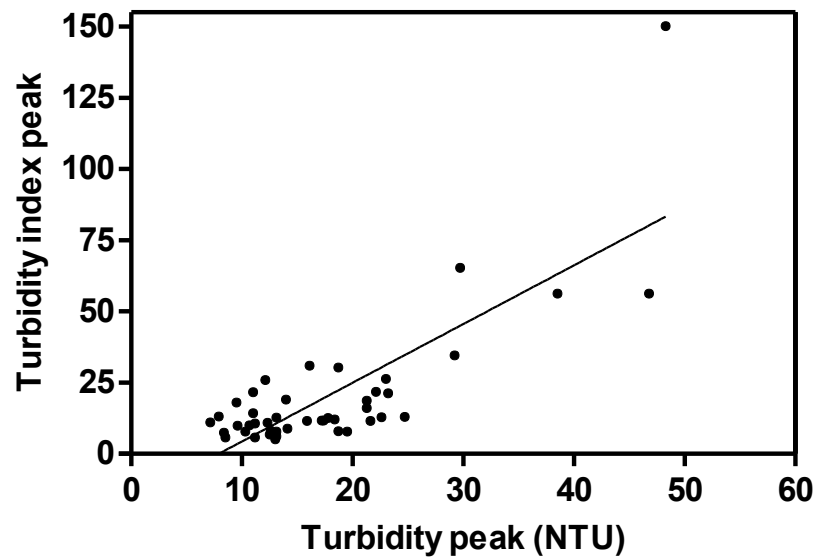
- Délais d'anticipation :
 - Entre 15 - 270 minutes saison "froide" (Hiver -Automne)
 - Entre 15 - 240 minutes saison "chaude" (Printemps - Été)
- Anticipation plus faible des pics en saison chaude
- Causes :
 - Phase d'augmentation des valeurs de turbidité plus courte
 - Pics de valeur plus élevée



Résultats

Analyse de tendance

- Anticipation de l'amplitude de la variation du paramètre turbidité

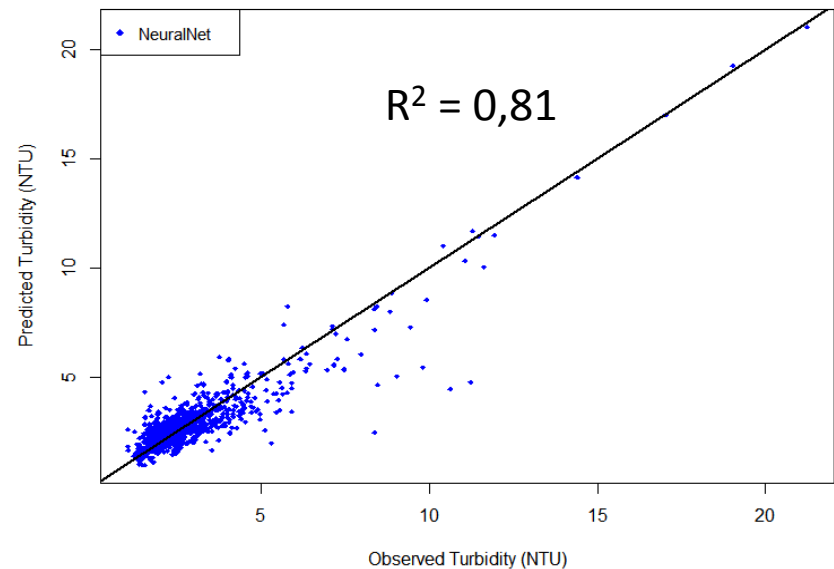
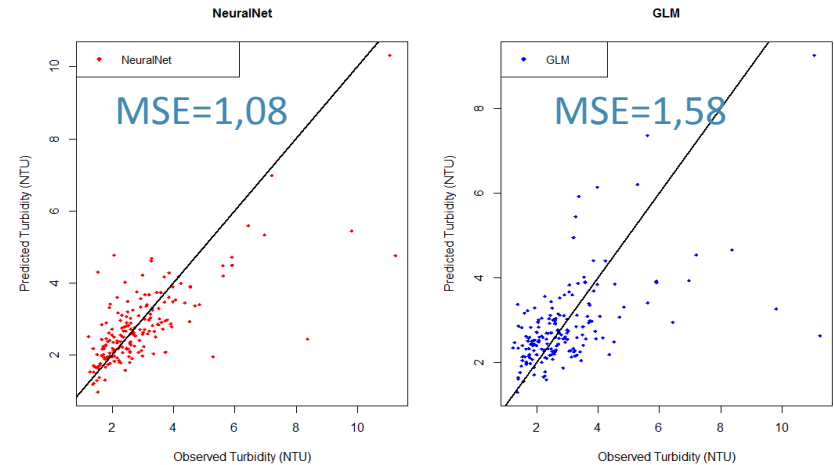


Résultats

Réseau de neurones

- Optimisation du nombre de neurons
 - Test de 196 (14*14) configurations
 - Forme optimale : 11*5 couches cachées
- Estimation adéquate des valeurs de sortie ($R^2 = 0,81$)
- Sous estimation des valeurs intermédiaires (5-10 NTU)

>5NTU	Classification correcte (%)	Faux positifs (%)	Faux négatifs (%)
Réseau de neurones	96,4%	1,3%	2,3%



Conclusion

- Anticipation des pics de turbidité à l'usine de traitement : anticipation possible des modifications de traitements à apporter
- Défi : implémentation des algorithmes « en ligne »
- Utilité pour la prédiction des valeurs de turbidité journalières selon les données climatiques
- Outil complémentaire pour le traitement des données : Utilisation justifiée car on s'intéresse plus à la précision de la prédiction du paramètre turbidité qu'à l'explication des mécanismes contrôlant sa variabilité

Perspectives

- Application sur le terrain en continu
 - Prévision de la qualité de l'eau à la prise d'eau
- Application à d'autres paramètres : conductivité, paramètres microbiologiques (E. Coli, Coliformes fécaux,...), selon la disponibilité des données
- Changements climatiques : Prédiction des valeurs de turbidité à la prise d'eau selon différents scénarios à l'échelle locale

Remerciements

- Techniciens de l'usine de traitement de Québec : Christian Pelletier et Louis Collin
- Ville de Québec : Francois Proulx



Chaire de recherche en eau potable
de l'Université Laval

THALES



Merci!

Questions?

Contact: Ianis.Delpla@crad.ulaval.ca

Méthodes d'analyse de données

- Analyse de régression
- Analyse de tendance
- Réseaux de neurones
- Fuzzy logic
- Analyse de groupement