

Titre de la thèse : Evaluation et prédiction de la qualité de l'eau du territoire de l'Adour basées sur le traitement et l'analyse de données

Type de financement : Région Occitanie – Agence de l'Eau Adour – Garonne

Encadrement : Prof. Kamal Medjaher, Dr. Raymond Houé Ngouna, Dr. Fabien Dauriac

E-mail : kamal.medjaher@enit.fr, raymond.houe-ngouna@enit.fr, f.dauriac@hautes-pyrenees.chambagri.fr

Laboratoire : Laboratoire Génie de Production (LGP)

Etablissement : Ecole Nationale d'Ingénieurs de Tarbes (www.enit.fr)

Candidature : envoyer par mail un CV et une lettre de motivation

Date limite de candidature : 15/05/2017

Date de début de la thèse : 01/10/2017

Mots-clés : Traitement et analyse de données, Données massives (big data), apprentissage automatique (Machine Learning), Détection, Diagnostic, Pronostic de défaillances.

Profil du candidat :

- Formation : Master 2 Recherche ou Diplôme d'Ingénieur avec une expérience recherche obligatoire (PFE effectué dans un laboratoire de recherche).
- Disciplines : Sciences pour l'ingénieur ou Mathématiques appliquées.
- Connaissances et compétences : Traitement et analyse de données, Données massives (big data), apprentissage automatique (Machine Learning). Des connaissances en détection et diagnostic de défaillances seraient un plus.
- Programmation : Matlab ou autres langages équivalents.
- Autre : Esprit de synthèse, bonne rédaction, bon niveau d'anglais, aptitude à travailler en groupe.

1. Problématique de recherche

Depuis 2008, l'Agence de l'eau Adour Garonne accompagne un Plan d'Action Territorial (PAT) de la nappe alluviale de l'Adour, avec comme objectif de mobiliser tous les acteurs du territoire afin de reconquérir la qualité de l'eau de la nappe de l'Adour. Ce PAT se traduit par la protection des captages d'eau potable contre les pollutions diffuses (pesticides, nitrates) ainsi que la préservation globale du milieu naturel en réduisant ces mêmes polluants.

De nombreuses actions ont été conduites depuis 8 ans sur le territoire, parmi lesquelles :

- l'accompagnement de la majorité des collectivités du territoire (plans de désherbage dans les communes, zéro phyto pour l'entretien des routes, meilleure fertilisation des stades...),
- le changement de pratiques des agriculteurs (réduction des quantités d'engrais et de pesticides, modernisation du matériel, développement de l'agriculture biologique...),
- la sensibilisation du grand public aux bonnes pratiques (éducation à l'environnement dans les écoles, jardinage sans pesticides...).

Ces changements durables de pratiques doivent être évalués au regard de l'amélioration de la qualité de l'eau sur le territoire. Une telle évaluation est délicate aujourd'hui et fait appel à des études scientifiques, car le milieu naturel (climat, sol) joue un rôle important dans la régulation de la qualité de l'eau. Des variabilités spatiales et temporelles sont observées et méritent d'être analysées.

Ce travail de recherche qui sera mené s'intéresse à la problématique de l'évaluation de la qualité de l'eau sur le territoire de l'Adour. Cette problématique soulève plusieurs questions : **par quoi est influencée la qualité de l'eau de la nappe de l'Adour ? Le climat est-il prépondérant ou non ? Dans quelle proportion les pratiques agricoles ou les autres activités humaines jouent-elles un rôle ? Peut-on prédire la qualité de l'eau en fonction des pratiques agricoles et autres facteurs climatiques ?**

L'enjeu est donc de croiser entre elles de données massives, archivées depuis plus de 30 ans, afin de répondre à ces questions : données sur la qualité de l'eau (utilisation des nitrates, des pesticides...), sur la quantité d'eau utilisée (débits et hauteurs de nappe), sur le climat (pluviométrie, températures, évapotranspiration...), sur les pratiques agricoles (assolement, consommation d'eau pour l'irrigation...), et sur d'autres données issues des études spécifiques d'hydrogéologie sur certains captages. Cela permettrait de définir dans le futur les plans d'action prioritaires à conduire.

2. Objectif de la thèse

Le travail de recherche de cette thèse vise à développer et à mettre en œuvre une méthode de surveillance, d'évaluation et de prédiction de la qualité de l'eau, basée sur le traitement et l'analyse des données disponibles.

L'objectif est de permettre :

- la détection d'anomalies (présence de taux élevés de nitrates dans l'eau, par exemple),
- l'identification des facteurs influents (diagnostic des causes),
- et la prédiction de la qualité de l'eau en fonction des conditions climatiques et des pratiques humaines (irrigation, traitement des sols, etc.).

Ce travail de recherche peut être ainsi inscrit dans la thématique du PHM (Prognostics and Health Management) où le système physique est l'eau, surveillée par différents capteurs. Dans le domaine du PHM, trois approches peuvent être utilisées : l'approche basée sur un modèle physique, l'approche guidée par des données et l'approche hybride combinant les deux. Dans ce travail de thèse, et vu la connaissance et les données disponibles sur le sujet, l'approche guidée par les données semble la plus adéquate à mettre en œuvre. Celle basée sur un modèle physique, au vu de la complexité du système étudié (eau de l'Adour combinée aux différents facteurs influents), et à la lumière des résultats de l'approche basée sur les données, pourra être abordée par la suite.

Différents outils peuvent être utilisés par l'approche guidée par les données : outils statistiques, outils issus de l'intelligence artificielle, outils stochastiques, outils de Machine Learning, etc. Pour répondre à la problématique et aux questions posées précédemment, et vu la nature et la quantité de données ainsi que les différents facteurs mis en jeu, les outils issus du Machine Learning seraient de « bons » candidats, correspondant au choix actuellement privilégié.

3. Développements scientifiques envisagés

- **Etat de l'art** : Il s'agira dans un premier temps d'étudier le système « eau et son environnement » et à faire une étude bibliographique approfondie des méthodes

proposées dans la littérature pour le traitement et l'analyse de données massives. L'objectif est de mieux cerner le système étudié, de bien définir les verrous scientifiques et de positionner scientifiquement le travail de recherche.

- **Traitement et analyse des données** : l'objectif est d'établir une première cartographie des données disponibles, d'identifier les relations entre les différents facteurs influents et de définir les hypothèses de travail pour la suite des développements de la thèse. Des méthodes d'analyse, de fusion et de sélection de données identifiées suite à l'étude de l'état de l'art seront utilisées.
- **Détection d'anomalies et diagnostic des causes** : il s'agit de développer une méthode guidée par les données pour l'évaluation de la qualité de l'eau, la détection d'anomalies et le diagnostic des causes probables. Les outils d'apprentissage automatique, notamment les méthodes supervisées et non supervisées, seront particulièrement utilisés.
- **Etude de l'impact de l'activité humaine sur la qualité de l'eau** : il s'agira d'étendre la méthode développée précédemment en vue d'intégrer l'influence du facteur humain (actions des agriculteurs) sur la qualité de l'eau.
- **Prédiction de la qualité de l'eau** : cela concerne le développement d'une méthode de prédiction de la qualité de l'eau qui tient compte des différents facteurs influents. Cette méthode s'appuiera sur les données collectées. Les méthodes de régression non linéaires, comme celle par machines à vecteur de support, pourraient être de bons candidats.

4. Références

- Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- C. Cortes and V. Vapnik. Support-vector networks. Machine Learning, 20(3): 273–297, 1995.
- A. Smola and B. Schölkopf. A Tutorial on Support Vector Regression. NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK, 1998.
- J.Z. Sikorska, M Hodkiewicz, and L Ma. Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing, 25(5): 1803–1836, 2011.
- A. Heng, S. Zhang, Andy C.C. Tan, and J. Mathew. Rotating machinery prognostics: State of the art, challenges and opportunities. Mechanical Systems and Signal Processing, 23(3) :724 – 739, 2009.
- Andrew. K.S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7): 1483 – 1510, 2006.
- A. Soualhi, K. Medjaher, and N. Zerhouni. Bearing health monitoring based on Hilbert-Huang transform, support vector machine and regression. IIEEE Transactions on Instrumentation and Measurement, 2014. Article published online 01 July 2014, DOI: 10.1109/TIM.2014.2330494.
- T. Benkedjough, K. Medjaher, N. Zerhouni, S. Rechak. Remaining useful life estimation based on nonlinear feature reduction and support vector regression. Engineering Applications of Artificial Intelligence, vol. 26, no. 7, pp. 1751-1760, 2013.
- A. Mosallam, K. Medjaher, and N. Zerhouni. Data-driven prognostic method based on bayesian approaches for direct remaining useful life prediction. Journal of Intelligent

Manufacturing, 2014. Article published online 13 June 2014, DOI: 10.1007/s10845-014-0933-4.

- A. Mosallam, K. Medjaher, and N. Zerhouni. Nonparametric time series modelling for industrial prognostics and health management. *The International Journal of Advanced Manufacturing Technology*, 6: 1685 – 1699, 2013.
- F. Camci, K. Medjaher, Zerhouni, and P. Nectoux. Feature evaluation for effective bearing prognostics. *Quality and Reliability Engineering International*, 29: 477–486, 2013.
- K. Medjaher, D.A. Tobon-Mejia, and N Zerhouni. Remaining useful life estimation of critical components with application to bearings. *IEEE Transactions on Reliability*, pages 292–302, 2012.