
HydroQual: Un outil visuel pour analyser la qualité de l'eau des rivières

Pierre Accorsi¹, Mickaël Fabrègue², Arnaud Sallaberry¹, Flavie Cernesson³, Nathalie Lalande³, Agnès Braud⁴, Sandra Bringay¹, Florence Le Ber⁵, Pascal Poncelet¹ and Maguelonne Teisseire³

1. LIRMM, Montpellier

prenom.nom@lirmm.fr

2. IRSTEA Montpellier, Université de Strasbourg/ENGEES,

prenom.nom@teledetection.fr

3. TETIS, Montpellier

prenom.nom@teledetection.fr

4. ICUBE, Université de Strasbourg

prenom.nom@unistra.fr

5. ICUBE, Université de Strasbourg/ENGEES

prenom.nom@unistra.fr

RESUME. La qualité de l'eau des rivières est un problème sociétal aussi bien pour la santé que pour l'environnement et les différents gouvernements ont proposé de nombreuses initiatives pour préserver et restaurer l'écosystème aquatique et les ressources en eau. Les gestionnaires de l'eau ont donc maintenant besoin de nouveaux outils pour les aider à interpréter les informations complexes disponibles pour évaluer la qualité de l'eau. Dans cet article, nous proposons HydroQual une plateforme qui combine à la fois des techniques de fouille de données spatio-temporelles et des techniques de visualisation pour aider les experts.

ABSTRACT. River water quality is a pivotal public health and environmental issue that has prompted governments to plan initiatives for preserving or restoring aquatic ecosystems and water resources. Water managers require operational tools to help interpret the complex range of information available on river water quality functioning. In this paper, we describe HydroQual, a tool to facilitate visual analysis of river water quality. This tool combines spatiotemporal data mining and visualization techniques to perform tasks defined by domain experts.

Mots-clés : Analyse Visuelle, Fouille et visualisation de motifs spatiaux-temporels, Qualité de l'eau.

KEYWORDS : Visual Analytics, Spatiotemporal Data Mining and Visualization, Water Quality

DOI:10.3199/JESA.45.1-n © Lavoisier 2012 [AR](#) [DOI](#)

1. Introduction

L'eau est un composant vital pour toutes les formes connues de vie. La *World Health Organization* (WHO) estime que l'usage domestique (boisson, cuisine, hygiène) nécessite un minimum de 20 litres d'eau de bonne qualité par jour et par personne (Who, 2014). La problématique de la qualité de l'eau est tout d'abord apparue au XIX^{ème} siècle lorsque des scientifiques ont mis en évidence les corrélations qui existaient entre la qualité de l'eau et certains problèmes de santé. Par exemple J. Snow a été le premier à suggérer que le cholera était lié à une épidémie se propageant par l'eau (Snow, 1855). Depuis les scientifiques ont également mis en évidence le fait que la qualité de l'eau est fortement dépendante de la croissance des populations, du développement des villes et des activités économiques (Meybeck and Helmer, 1989).

La pollution est un problème majeur que l'*United Nations Environment Program* (UNEP) et la WHO ont mis en évidence et qui a conduit de nombreux pays à définir de nouveaux programmes ou plans pour préserver et restaurer la qualité de l'eau (Who, 1997). Pour cela, les stratégies mises en place ont nécessité de développer des techniques de suivi de la qualité de l'eau au travers de différentes analyses de terrain. Ainsi, le premier suivi de quelques rivières européennes a commencé à la fin du XIX^{ème} siècle où seulement 5 descripteurs étaient analysés. Aujourd'hui plus de 150 descripteurs sont analysés régulièrement sur de nombreuses rivières (Meybeck and Helmer, 1989). Toutes les données (indices biologiques ou paramètres physico-chimiques) concernant des prélèvements sur des stations géolocalisées sont maintenant collectées dans des bases de données dont l'objectif est de pouvoir caractériser : (1) les conditions globales et les tendances concernant la qualité des rivières ; (2) la possibilité de contrôler l'eau pour un usage particulier ou en analysant l'impact des activités humaines sur la qualité. Cette connaissance a pour but, dans un processus d'aide à la décision, de prendre des mesures de protection ou de restauration de l'environnement et par exemple peut conduire à l'autorisation ou à l'interdiction d'utiliser les ressources.

L'importance d'avoir des outils opérationnels pour interpréter les informations complexes concernant la qualité de l'eau des rivières ainsi que leur fonctionnement a été souligné par la Directive-Cadre européenne sur l'Eau –DCE- (European, 2000). Pour répondre à cette demande de nombreux outils basés sur des Systèmes d'Information Géographiques incluant des analyses statistiques et différents graphiques ont été proposés ces dernières années, e.g. *ArcGIS Geostatistical Analyst* (ArcGIS, 2014). Cependant même si ces outils peuvent aider les experts, ils sont limités parce que les données manipulées sont non seulement complexes mais peuvent être également manquantes. Pour répondre à ce problème, nous proposons dans cet article une plateforme qui combine à la fois des aspects fouille de données et visualisation pour aider l'expert dans l'analyse de la qualité des cours d'eau.

L'article est organisé de la manière suivante. Dans la section 2 nous présentons un aperçu des techniques de fouilles de données qui sont mises en œuvre. La section 3 présente plus en avant la plateforme HydroQual. Enfin dans la section 4 nous concluons.

2. Fouilles de motifs pour la qualité de l'eau

De très nombreux paramètres sont impliqués dans la détermination de l'écosystème de l'eau. Ils sont reliés aux différents aspects biologiques, physico-chimiques ou hydro-morphologiques. La plupart des travaux de recherche se sont intéressés aux corrélations existant dans ces dimensions ou bien entre ces différentes dimensions. Par exemple, certains travaux de recherche se sont intéressés à l'application de techniques de fouille de données pour étudier la faune représentée par les macro invertébrés (Goethals et al, 2007), les communautés de poissons (Yang et al, 2008), etc. Cependant aucune de ces études ne prend en compte l'aspect temporel qui est essentiel pour analyser les dynamiques impliquées dans l'analyse des pollutions. L'une des approches de fouille de données les plus adaptées pour analyser au cours du temps des évolutions est l'extraction de motifs séquentiels. En effet, en recherchant des relations entre des événements au cours du temps, ces approches sont capables de mettre en évidence des motifs du type : $\langle (Pollution) (Animaux Morts) \rangle 30\%$ qui signifient qu'un événement de pollution est suivi, peu de temps après, par la mort d'animaux et que cette séquence se retrouve dans 30% des données de la base. Une telle connaissance est alors utile pour aider l'expert à mieux appréhender la dynamique et les conséquences de l'événement *Pollution*.

Cependant malgré leurs avantages, les motifs séquentiels offrent souvent une information limitée puisque nous disposons seulement d'une information partielle à propos des données. Considérons par exemple, la séquence suivante $\langle (Pollution) (Végétation Morte) \rangle 30\%$ extraite de la même base de données que l'exemple précédent. Ces deux motifs coexistent dans la base de données mais cette coexistence n'est pas prise en compte dans le résultat final. Pour pallier ce problème, nous avons proposé une extension des motifs sous la forme de motifs partiellement ordonnés. Le lecteur intéressé peut se référer à (Fabregues et al, 2013).

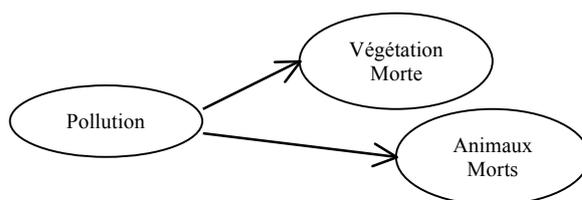


Figure 1. Exemple d'un motif partiellement ordonné avec une fréquence de 30%

La Figure 1 présente un motif partiellement ordonné qui combine les deux motifs précédents. L'interprétation d'un tel motif est qu'une *Pollution* est suivie par deux événements *Animaux Morts* et *Végétation Morte* qui ne sont pas ordonnés dans le temps. Dans le cadre de données hydro-biologiques, ce type de motif possède les avantages suivants : (1) ils sont bien adaptés à l'aspect temporel des données de la base ; (2) ils produisent plus d'informations sur l'ordre entre les éléments que les motifs séquentiels et (3) ils sont représentés comme des graphes orientés acycliques qui sont d'une aide importante pour les experts dans la compréhension des enchaînements d'événements.

3. La plateforme HydroQual

La plateforme HydroQual est le résultat de trois ans de collaborations entre des experts (hydrologistes, hydro-biologistes, gestionnaires de l'eau) et des informaticiens spécialistes en fouille de données et en visualisation d'informations.

La plateforme propose une interface visuelle complète pour explorer la qualité des rivières. Outre la possibilité d'interroger, au travers de l'interface, les différentes données de la base, elle intègre les différentes approches de motifs décrits précédemment et permet à l'expert de pouvoir naviguer entre les connaissances et les données. La figure 2 présente l'aspect général de la plateforme.

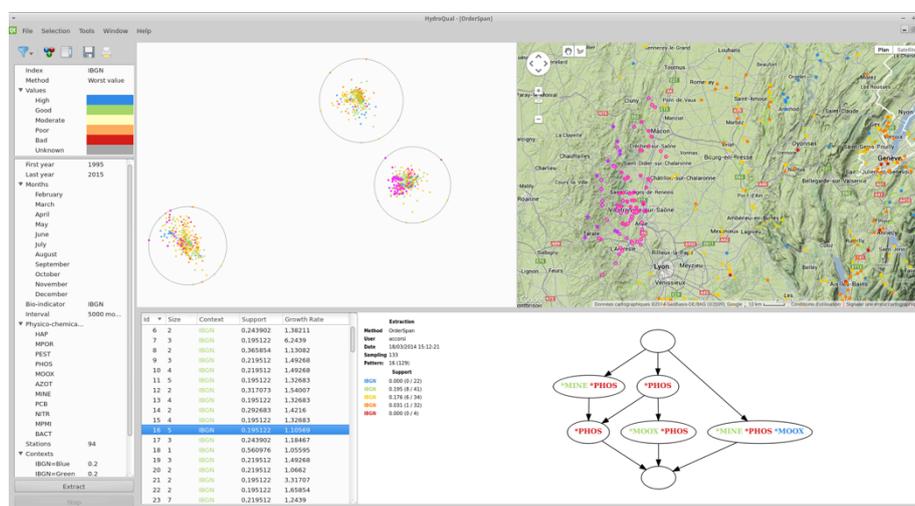


Figure 2. HydroQual est une plateforme qui facilite visuellement l'analyse de la qualité des rivières. Les données d'entrées sont des séquences d'indices biologiques et de valeurs physico-chimiques pour différentes stations de prélèvement géolocalisées. La vue clustérisée (haut gauche) montre les stations regroupées en fonction de leur similarité de comportement. La vue géographique (haut droite) montre la géolocalisation des stations. En sélectionnant un ensemble de stations de ces deux vues, l'expert peut extraire et visualiser les motifs temporels en fonction des indices biologiques et des paramètres physico-chimiques dans la vue des motifs temporels (bas).

L'utilisateur dispose de toutes les fonctionnalités pour extraire les informations relatives aux différentes stations. Différentes possibilités de sélectionner des stations par groupe, par couche, etc. sont offertes et les motifs peuvent alors être extraits en temps réel. Pour faciliter l'interprétation des motifs extraits, différentes vues sont proposées. Par exemple la vue clustérisée (Accorsi et al, 2014) permet de regrouper de manière automatique toutes les stations qui possèdent le même comportement. Ce dernier peut alors être affiché de manière plus détaillée sous la forme d'un graphe dont les couleurs correspondent aux différentes classes de valeurs d'indices (Cf. bas de la Figure 2). Bien entendu toutes les fonctionnalités pour projeter les

informations sur différentes couches (Corine Land Cover, régions administratives, réseau hydrologique, régions hydrologiques, etc.) sont également disponibles (Cf. Figure 3).

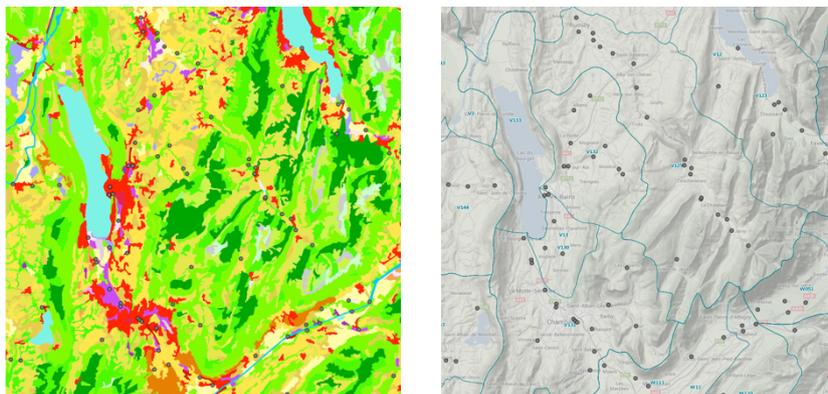


Figure 3. Exemples de couches possibles : (1) à gauche, fond de carte Corine Land Cover, (2) à droite, fond de carte *Google Maps physical* avec les contours des bassins versants en vert clair.

De manière à valider l'utilisation de la plateforme, cinq type de questions ont été exprimées qui intègrent les problématiques spatiales et temporelles et surtout répondent aux attendus des différents experts : (Q1) Pour un objectif donné, quel jeu de données est le plus pertinent ? (Q2) Pour une zone géographique donnée, quelles sont les stations qui possèdent un comportement similaire ? (Q3) A l'inverse où sont situées les stations qui possèdent un même comportement ? (Q4) Pour une station et un indice biologique donné, quelle est la pire valeur qualitative calculée ? (Q5) Pour un ensemble de stations quelles sont les tendances entre les indices biologiques et les paramètres physico-chimiques ?

Les expérimentations menées avec les experts sur l'utilisation d'HydroQual ont montré que l'outil est particulièrement adapté pour répondre à ces différentes questions et que surtout les experts peuvent facilement s'approprier les différentes fonctionnalités de la plateforme.

4. Conclusion

Pouvoir garantir la qualité de l'eau est un enjeu sociétal important. Dans cet article, nous avons montré qu'il était possible d'offrir une nouvelle approche couplant fouille de données et visualisation pour permettre aux experts de mieux analyser les différents paramètres et leur évolution au cours du temps. Via cet outil, les experts disposent de toutes les fonctionnalités pour interroger, analyser, extraire, visualiser les différents éléments de connaissance qui permettent de mieux comprendre l'impact des activités humaines sur la qualité de l'eau.

Remerciements :

Ce projet a été financé par le projet ANR 11 MONU 14 FRESQUEAU. Nous voudrions remercier Corinne Grac (LIVE), Xavier Dolgues (ICUBE) et Danielle Levet (AQUASCOPE) pour leur assistance dans l'obtention des données et la définition des cas d'utilisation. Nous souhaitons également remercier Hugo Alatrística Salas et Vijay Ingalalli (LIRMM) pour leur expertise technique.

Bibliographie

- (Accorsi et al, 2014) P. Accorsi, M. Fabrègue, A. Sallaberry, F. Cernesson, N. Lalande, A. Braud, S. Bringay, F. Le Ber, P. Poncelet, M. Teisseire. HydroQual: Visual Analysis of River Water Quality. *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST 2014)*, 2014.
- (ArcGIS, 2014) ArcGIS Geostatistical Analyst. <http://www.esri.com/software/arcgis/extensions/geostatistical>. [On-line; accessed 16-June-2014].
- (European, 2000) European Union. Directive 2000/60/ec of the European parliament and of the council of 23 october 2000 establishing a framework for community action in the field of water policy. *Official Journal*, OJ L 327:1–73, 2000.
- (Fabregue et al, 2013) M. Fabrègue, A. Braud, S. Bringay, F. Ber, and M. Teisseire. Orderspan: Mining closed partially ordered patterns. In *Proceedings of the 12th International Symposium on Intelligent Data Analysis (IDA'13)*, pages 186–197, 2013.
- (Goehtals et al, 2007) P. L. Goethals, A. Dedecker, W. Gabriels, S. Lek, and N. Pauw. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Aquatic Ecology*, 41:491–508, 2007.
- (Meybeck and Helmer, 1989) M. Meybeck and R. Helmer. The quality of rivers: From pristine stage to global pollution. *Global and Planetary Change*, 1(4):283–309, 1989.
- (Snow, 1855) J. Snow. *On the Mode of Communication of Cholera*. John Churchill, 1855.
- (Who, 2014) World Health Organization. http://www.who.int/water_sanitation_health/en/. [Online; dernier accès 31-March-2014].
- (Who, 1997) WHO/UNEP. *Water Pollution Control - A Guide to the Use of Water Quality Management Principles*. WHO publications, 1997.
- (Yang et al., 2008) Y.-C. E. Yang, X. Cai, and E. E. Herricks. Identification of hydrologic indicators related to fish diversity and abundance: A data mining approach for fish community analysis. *Water Resources Research*, 44, 2008.