HAL open science

Application de modèles non paramétriques sous R pour l'analyse et le suivi de la qualité de l'eau

Mohamedou Sow

To cite this version:

Mohamedou Sow. Application de modèles non paramétriques sous R pour l'analyse et le suivi de la qualité de l'eau. 1ères Rencontres R, Jul 2012, Bordeaux, France. hal-00717511

HAL Id: hal-00717511

https://hal.science/hal-00717511

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Application de modèles non paramétriques sous R pour l'analyse et le suivi de la qualité de l'eau

M. Sow[a,b], G. Durrieu[c], D. Tran[a,b], P. Ciret[a,b] et J.C. Massabuau[a,b]

[a]Univ. de Bordeaux. EPOC, UMR 5805, F-33120 Arcachon, France
[b]CNRS, EPOC, UMR 5805, F-3312 Arcachon, France
{m.sow, d.tran, p.ciret, jc.massabuau}@epoc.u-bordeaux1.fr

[c] Laboratoire de Mathématiques de Bretagne Atlantique UMR CNRS 6205
Université de Bretagne Sud
Campus de Tohannic, 56017 Vannes
gilles.durrieu@univ-ubs.fr

Mots clefs : Biologie, Environnement, Régression non paramétrique, Estimateur à noyau, Valvometrie HFNI, R.

A l'heure où nos sociétés sont pleinement conscientes que la protection de l'environnement est un enjeu de société majeur, l'étude des milieux aquatiques côtiers apparaît comme une question privilégiée. Il s'agit de préserver des zones riches et sensibles particulièrement à risque. Dans ce contexte, des règlementations et de nombreux contrôles de la qualité de l'eau sont mis en place. De nouvelles techniques et de nouveaux outils doivent être imaginés et développés. Parmi ces techniques de contrôles, les bioindicateurs sont de plus en plus utilisés et sont très efficaces par leurs capacités à révéler la présence de traces (concentrations très faibles) de contaminants comme l'accumulation dans des tissus animaux ou végétaux spécifiques ou des modifications sur des structures populationnelles. Nous travaillons ici sur le développement d'un outil d'analyse haute fréquence (10 Hz en continue pendant au moins un an) du comportement d'huîtres, qui permet d'aborder leur éthologie et l'expression de leurs rythmes biologiques, mais qui est aussi un formidable moyen de surveillance de la qualité de l'eau. Le cas des huîtres est particulièrement intéressant car se sont des animaux sédentaires qui peuvent être témoins de pollutions locales (bioindicateurs) et on les trouve partout dans le monde, des tropiques aux pôles (possibilité d'application très large).

L'objet de ce travail est de modéliser et analyser l'important volumes de données acquis à haute fréquence en implémentant des modèles de régression non paramétriques et des codes de calculs sous R ([1]-[4]) pour l'analyse des données collectées sur différents sites (bassin d'Arcachon, Nouvelle Calédonie, Espagne, Bretagne, Norvège, Russie, …) dans le but de mettre en place un outil de surveillance en ligne de la qualité de l'eau basé sur l'analyse en continue du comportement de bivalves. Plus précisément, à partir d'un échantillon composé de 54000 couples indépendants de variables aléatoires $(T_1, Y_1),\ldots,(T_n, Y_n)$ qui sont le temps en heure et l'écartement valvaire en mm, nous considérons le modèle de régression non paramétrique donné, pour $i = 1,\ldots,54000$, par

$$Y_i=m(T_i)+\varepsilon_i.$$

Dans ce modèle intervient une fonction m inconnue à estimer qui exprime la valeur moyenne de l'écartement valvaire de nos bivalves en fonction du temps T et un terme aléatoire d'erreur ε de loi inconnue et indépendant de T. Nous proposons plusieurs estimateurs non paramétriques de la fonction m.

Le but des méthodes statistiques et des codes de calculs développés (R et scripts Bash) a d'abord été de mettre en évidence et d'extraire des rythmes biologiques. Elles ont été ensuite utilisées pour les présenter sous forme graphique simple afin de permettre, par la caractérisation de perturbations de ces rythmes, de détecter *in fine* une éventuelle pollution du milieu. Dans un premier temps, nous représentons graphiquement les périodes d'ouvertures et fermetures des bivalves sur l'ensemble des données disponibles depuis mars 2006 au niveau de la jetée d'EYRAC. Ces enregistrements permettent de mettre en évidence des rythmes biologiques liés aux rythmes des marées chez l'huître ([4]). Une de nos hypothèses de travail est que la modification de ces rythmes biologiques pourrait alors être symptomatique d'un problème au niveau de la qualité de l'eau. La figure 1 nous montre que les activités de fermetures sont corrélées à la marée et plus précisément nous pouvons observer que les huîtres ferment leurs valves à l'étale de basse mer.

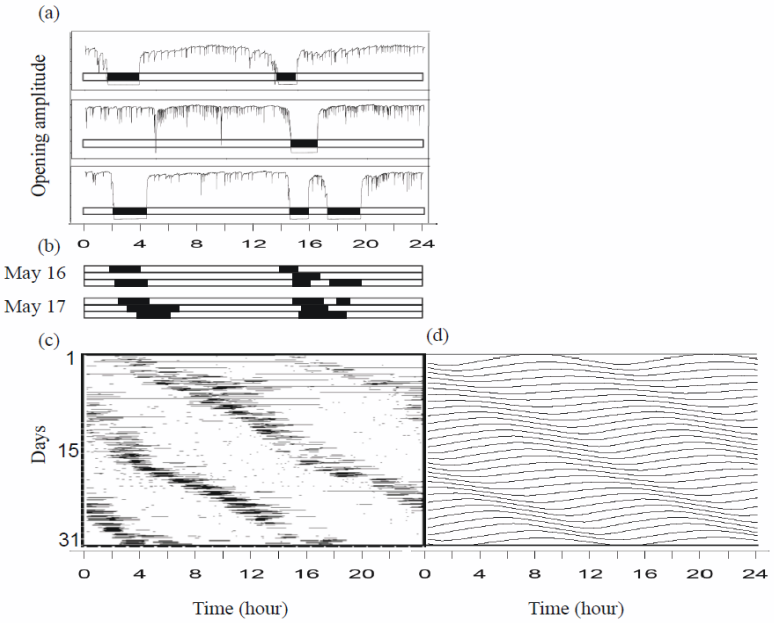

Figure 1. Principe du rythme biologique lié à la marée :
a) activité de fermetures / ouvertures pour 3 différents huîtres (trait noir fermeture et trait blanc ouverture) ;
b) superposition de fermetures / ouvertures de 3 huîtres différentes (16 et 17 Mai) ;
c) représentation de l'activité de fermeture/ouverture sur 31 jours (pour chaque jour, il y'a 16 lignes représentant les périodes de fermetures/ouvertures pour 16 huîtres) ;
d) Représentation de l'évolution de la hauteur d'eau par heure.

Actuellement et grâce aux résultats de nos travaux, l'acquisition, le transfert, et le traitement des données fonctionnent de manière automatique pour les différents sites où un système est positionné dans le monde. Les enregistrements et les résultats du traitement statistique sont accessibles sur le site web "L'œil du mollusque" (http://molluscan-eye.epoc.u-bordeaux1.fr/).

Références

[1] Coudret R., Durrieu G., Saracco J. (2012). Estimateurs a noyau bimodaux d'une densité bimodale et comparaison avec d'autres estimateurs non paramétriques, *Proc de la société Française de Statistique*, sous presse.

[2] Durrieu G., Nguyen T.M.N., Sow M. (2009). Comparaison d'estimateurs de régression non paramétriques : application en valvometrie, *Proc. de la société Française de Statistique*, http://hal.inria.fr/docs/00/38/67/16/PDF/p147.pdf.

[3] Schmitt F.G., De Rosa M., Durrieu G., Sow M., Ciret P., Tran D., Massabuau J.C. (2011). Statistical study of bivalve high frequency microclosing behavior: scaling properties and shot noise analysis, *International Journal of Bifurcation and Chaos*, 21, 3565-3576.

[4] Sow M. Durrieu G., Briollais L., Ciret P., Massabuau J.C. (2011). Water quality assessment by means of HFNI valvometry and high-frequency data modeling, *Environmental Monitoring and Assessment*, 182, 155-170.