



HAL
open science

Évènements Rares sur des Séries Temporelles Environnementales

Gilles Durrieu, Ion Grama, Véronique Le Tilly, J.C. Massabuau, Quang-Khoai
Pham

► **To cite this version:**

Gilles Durrieu, Ion Grama, Véronique Le Tilly, J.C. Massabuau, Quang-Khoai Pham. Évènements Rares sur des Séries Temporelles Environnementales. 44e Journées de Statistique, May 2012, Bruxelles, Belgique. pp.1-6. hal-00905437

HAL Id: hal-00905437

<https://hal.science/hal-00905437>

Submitted on 18 Nov 2013

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.

ÉVÈNEMENTS RARES SUR DES SÉRIES TEMPORELLES ENVIRONNEMENTALES

Gilles Durrieu ¹ & Ion Grama ¹ & Véronique Le Tilly ² & Jean-Charles Massabuau ³ & Quang Khoai Pham ¹

¹ *Laboratoire de Mathématiques de Bretagne Atlantique, Université de Bretagne Sud et UMR CNRS 6205*

² *Laboratoire d'Ingénierie des MATériaux de Bretagne, Université de Bretagne Sud*

³ *Université Bordeaux et CNRS UMR 5805-EPOC*

Résumé. La mesure de l'activité de mollusques bivalves est un moyen d'enregistrer le comportement de bivalves *in situ* et donc d'évaluer des changements de la qualité de l'eau. Nous proposons une méthode statistique basée sur la théorie des valeurs extrêmes permettant d'estimer des changements globaux (pollution, changement de température) et ainsi d'aider à la surveillance de systèmes aquatiques.

Mots-clés. Estimateur à noyau, Haute fréquence, Valeurs extrêmes

Abstract. Measurement of mollusks bivalves activity is a way to record the animal behavior and so to evaluate possible changes in the water quality. We propose a statistical method based on the theory of extreme values to estimate global changes (pollution, change of temperature) and so to contribute to the survey of aquatic systems.

Keywords. Kernel Estimator, High frequency, Extreme values.

1 Introduction

Les activités humaines sont responsables d'importants rejets d'agents polluants dans le milieu naturel. Ces polluants entraînent la dégradation de nombreux biotopes perturbant les écosystèmes et posant également des problèmes en termes de santé publique. Des réglementations et des contrôles sur la qualité des eaux ont été mis en place. Parmi ces contrôles, les bioindicateurs sont de plus en plus utilisés car ils peuvent se révéler très efficaces par leurs capacités à révéler la présence de traces (concentrations très faibles) de contaminant et par leur capacité d'intégration au cours du temps. Les mollusques bivalves ventilent tout au long de la journée pour se nourrir et respirer. Ils sont équipés de récepteurs qui leur permettent en permanence d'estimer la qualité de l'eau dans laquelle ils baignent, de façon à pouvoir réagir immédiatement face à une eau qu'ils jugent nocive pour leur intégrité. Le cas des mollusques bivalves est particulièrement intéressant en tant qu'espèce bioindicatrice car ce sont des animaux sédentaires qui peuvent être témoins de changement locaux de la qualité de l'eau. Le suivi du comportement de bivalves

permet donc de rendre compte jour après jour de leur état de santé et au-delà, de l'évolution de la qualité de l'eau. Le modèle biologique utilisé pour cette étude est un bivalve filtreur : l'huître. L'huître *Crassostrea gigas* représente une ressource économique importante dans le Morbihan. Depuis de nombreuses années, le fonctionnement des entreprises ostréicoles se traduit par le transfert des huîtres d'un bassin ostréicole à un autre (étang de Thau, bassin d'Arcachon, Marennes-Oléron, Morbihan, Normandie) mais également au sein d'un même bassin. En parallèle, l'origine des naissains a fortement évolué. Ces différentes pratiques ont permis dans un premier temps de réduire les risques et donc de réguler l'économie des entreprises. Mais depuis une dizaine d'années, des épisodes de mortalités d'adultes, puis depuis 3 ans de naissains, ont fortement déstabilisé les pratiques ostréicoles et donc l'équilibre économique des entreprises. Le fonctionnement d'une entreprise ostréicole dépend donc de nombreux paramètres dont l'imbrication complexe ne permet plus à l'ostréiculteur de s'adapter rapidement à ces variations : marché, comportement des autres ostréiculteurs, qualité de l'environnement (sédiments et masses d'eau), stabilité biologique et physiologique des huîtres (croissance, résistance, mortalité). Dans ce travail, le contrôle permanent de l'état physiologique des huîtres est assuré par un suivi continu à haute fréquence (10 Hz) de l'ouverture des valves par valvométrie. La valvométrie (mesure de l'activité des valves de mollusques) est une technique qui permet d'enregistrer *in situ* les réactions de bivalves, face aux changements de la qualité de l'eau dans laquelle ils vivent. Le principe de ce biocapteur est le suivant : nous collons sur les valves des électrodes (électro-aimants) légères (200 mg) et non invasives (l'animal ne doit ni les sentir, ni être entravé) qui permettent de mesurer en permanence leur état d'ouverture. Ces électrodes sont gérées par une carte analogique immergée, à côté des animaux, en boîtier étanche. Le tout est relié à une seconde carte électronique en surface pilotée sous Linux. Le dispositif expérimental consomme seulement 1 watt et est alimentable par panneaux solaires, batteries ou source de courant. Actuellement dans ce projet, l'acquisition, le transfert et le traitement des données fonctionnent de manière automatique dans le secteur ostréicole de Locmariaquer situé à l'entrée du Golfe du Morbihan depuis le 4 mars 2011. Ce dispositif de suivi est installé au niveau de concessions ostréicoles accessibles à pieds à marée basse ($N47^{\circ}34,075'$ / $W2^{\circ}55,970'$). Les données brutes sont alors transférées tous les jours du terrain à une station de travail sous linux par le réseau de la téléphonie mobile (GPRS) et internet (FTP). Pour avoir un nombre d'animaux représentatif, nous travaillons sur un groupes de 16 animaux. Nous disposons donc d'un total de 864000 triplets de points par jour. Une fois les données brutes arrivées au laboratoire, elles sont modélisées et traitées statistiquement par des méthodes à noyaux (Durrieu et Briollais (2009) ; Schmitt et al. (2011) ; Sow et al. (2011) ; Tran et al. (2011) ; Schwartzmann et al. (2011)). Quelques résultats sont déjà mis automatiquement en accès public et professionnel sur un site web (l'Œil du Mollusque <http://molluscan-eye.epoc.u-bordeaux1.fr>). Dans cette communication, nous proposons une méthode statistique basée sur la théorie des valeurs extrêmes permettant d'estimer des changements globaux comme par exemple une pollution ou un changement de température et ainsi d'aider à la surveillance.

2 Modélisation dynamique d'évènements rares

On considère X_{t_1}, \dots, X_{t_n} une suite de variables aléatoires indépendantes où X_{t_i} a une loi F_{t_i} avec $t_i = \frac{i}{n}T$ pour $i = 1, \dots, n$ et T correspondant au dernier temps de la période d'analyse (24h, 48h, ..., un mois, une année, etc). On suppose que F_t est définie sur $[x_0, +\infty[$ avec $x_0 \geq 0$, et qu'elle appartient à l'ensemble \mathcal{F} des fonctions de répartition strictement croissantes admettant une densité continue par rapport à la mesure de Lebesgue μ . Un problème classique en statistique des extrêmes concerne l'estimation de la queue de distribution de F_t . La méthode proposée est ici de déterminer automatiquement un seuil τ et d'ajuster sur $[\tau, +\infty[$ une famille paramétrique de modèles $\{\mathcal{P}_\theta; \theta > 0\}$. On choisit ici pour \mathcal{P}_θ une distribution de Pareto définie par

$$G_{\tau, \theta}(x) = 1 - \left(\frac{x}{\tau}\right)^{-\frac{1}{\theta}}, x \in [\tau, +\infty[,$$

où le paramètre $\theta > 0$ et $\tau \geq x_0$ est la valeur seuil. L'espace des décisions est l'espace \mathcal{D} des couples $d = (\tau, \theta)$, où $\tau \in [x_0, +\infty[$ et $\theta \in]0, +\infty[$. La fonction de répartition F_t est modélisée en utilisant le modèle semi-paramétrique

$$P_{F_t, \tau, \theta}(x) = \begin{cases} F_t(x) & \text{if } x < \tau \\ 1 - (1 - F_t(\tau))(1 - G_{\tau, \theta}(x)) & \text{if } x \geq \tau. \end{cases} \quad (1)$$

Le paramètre θ est estimé par la méthode du maximum de vraisemblance, tandis que le paramètre de nuisance τ est déterminé par une procédure de sélection de modèle basée sur le maximum de vraisemblance pénalisée. Soit $\mathcal{K}(P, Q) = \int \log \frac{dP}{dQ} dP$ l'entropie de Kullback-Leibler entre les mesures équivalentes P et Q :

$$\mathcal{K}(F_t, P_{F_t, \tau, \theta}) = (1 - F_t(\tau)) \mathcal{K}(F_{t, \tau}, G_{\tau, \theta}).$$

Pour chaque $\tau \geq x_0$ et $t \in [0, T]$ on définit

$$\theta_{t, \tau} = \theta(\tau, F_{t, \tau}) = \arg \min_{\theta \in \Theta} \mathcal{K}(F_{t, \tau}, G_{\tau, \theta}).$$

$G_{\tau, \theta(\tau, F_{t, \tau})}$ est donc le plus proche élément $F_{t, \tau}$ dans la famille défini par (1) pour chaque τ et t . Le paramètre $\theta_{t, \tau}$ appelé paramètre de Pareto ajusté est donné par

$$\theta_{t, \tau} = \int_{\tau}^{\infty} \log \frac{x}{\tau} \frac{F_t(dx)}{1 - F_t(\tau)}, \tau \geq x_0.$$

Pour chaque t fixé dans $[0, T]$ et pour tout $\tau \geq x_0$, nous construisons un estimateur du paramètre $\theta_{t, \tau}$. Comme les observations X_{t_1}, \dots, X_{t_n} ne sont pas identiquement distribuées, $\theta_{t, \tau}$ est une fonction inconnue de t . Nous estimons dans un premier temps la fonction $\theta_{t, \tau}$ au point t en utilisant un estimateur usuel à noyau d'une taille de fenêtre h et dans un second temps nous donnons une procédure de sélection du seuil τ .

Nous définissons la quasi-log-vraisemblance par

$$\begin{aligned}\mathcal{L}_{t,h}(\tau, \theta) &= \sum_{i=1}^n w_{i,h}(t) \log \frac{dP_{F_t, \tau, \theta}}{d\mu}(X_{t_i}) \\ &= \sum_{X_i \leq \tau} w_{i,h}(t) \log f_t(X_{t_i}) + \sum_{X_{t_i} > \tau} w_{i,h}(t) \log \left(\frac{1}{\theta \tau} \left(\frac{X_{t_i}}{\tau} \right)^{-\frac{1}{\theta}-1} \right),\end{aligned}\quad (2)$$

où $w_{i,h}(t) = \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^n K(\frac{t_i-t}{h})}$ sont les poids avec K le noyau Gaussien et h la taille de la fenêtre. La fonction de répartition F_t est estimée par la fonction de répartition empirique pondérée

$$\widehat{F}_{t,h}(x) = \sum w_{i,h}(t) 1(X_{t_i} \leq x), \quad x \geq x_0.$$

En maximisant (2) par rapport à θ , nous obtenons l'estimateur

$$\widehat{\theta}_{t,h,\tau} = \frac{1}{n_{t,h,\tau}} \sum_{X_{t_i} > \tau} w_{i,h}(t) \log \frac{X_{t_i}}{\tau},$$

où $n_{t,h,\tau} = \sum_{i=1}^n w_i 1\{X_{t_i} > \tau\}$. Alors, l'estimateur semi-paramétrique de la fonction de survie est donné par

$$\widehat{F}_{t,h,\tau}(x) = \begin{cases} \widehat{F}_{t,h}(x), & x \in [x_0, \tau], \\ 1 - \left(1 - \widehat{F}_{t,h}(\tau)\right) \left(\frac{x}{\tau}\right)^{-\frac{1}{\widehat{\theta}_{t,h,\tau}}}, & x > \tau. \end{cases}$$

Nous remarquons que l'estimateur $\widehat{F}_{t,h,\tau}$ dépend des paramètres inconnus h et τ . La taille de la fenêtre h peut être estimée de manière classique en utilisant la méthode de la validation croisée. Pour le paramètre τ , nous proposons dans la suite une version pondérée de la procédure présentée dans Grama et Spokoiny (2008).

Soit $\mathcal{H}(\tau)$ l'hypothèse nulle stipulant que F_t est défini par (1) et s_1, \dots, s_n une suite d'instantanés triés par ordre décroissant de sorte que $s_1 \geq \dots \geq s_n$. Nous considérons une suite de tests d'adéquation en déterminant le premier instant s_k notée \widehat{s} pour lequel $\mathcal{H}(s_k)$ est rejetée. Le paramètre τ se détermine en maximisant, par rapport à τ , la fonction de vraisemblance pénalisée

$$\mathcal{L}_{\tau,h}(\tau, \widehat{\theta}_{t,h,\tau}) - \text{Pen}_{\tau,h}(\tau, \widehat{\theta}_{t,h,\widehat{s}}), \quad \text{où} \quad \text{Pen}_{t,h}(\tau, \theta) = \mathcal{L}_{t,h}(\tau, \theta).$$

3 Application sur des données environnementales

Nous évaluons ici la méthode statistique sur les données recueillies depuis le 4 mars 2011 au niveau du site Locmariaquer en Bretagne Sud. Les mesures sont collectées par

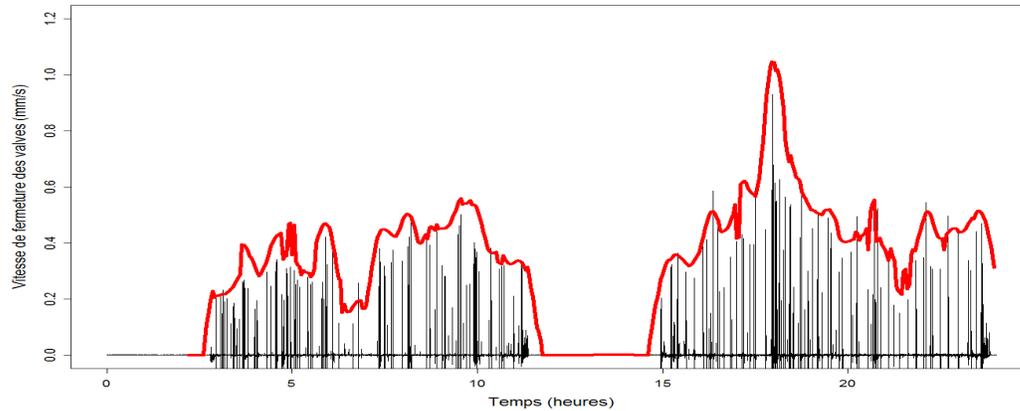


FIGURE 1 – Représentation pour le 20 mai 2011 des vitesses de fermetures en mm/s (en trait plein) et des 0.001-quantiles estimés de la distribution des vitesses de fermetures des valves (en trait plein gras) sur 24 heures.

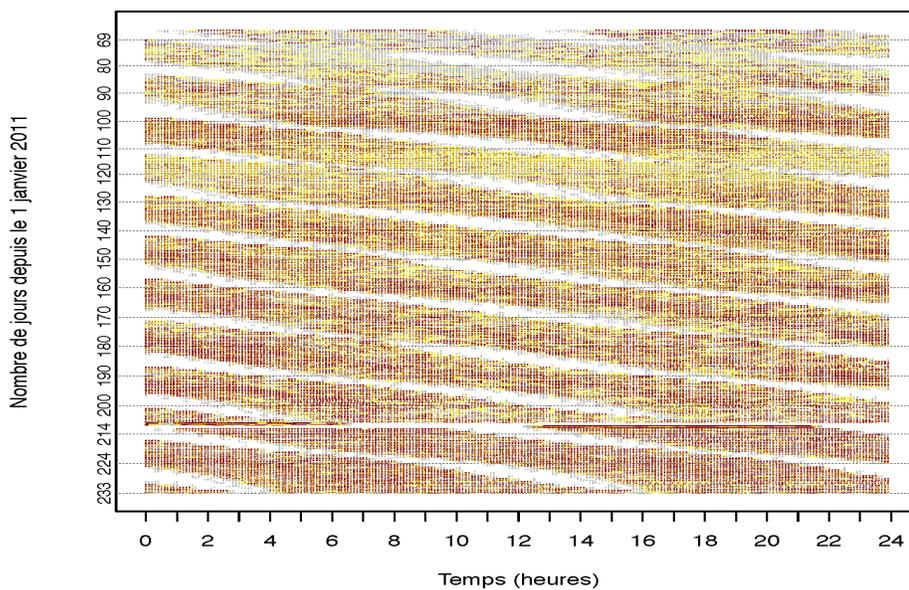


FIGURE 2 – Représentation des 0.001-quantiles estimés de la distribution des vitesses de fermeture entre le 4 mars et le 21 août 2011 sur le site de Locmariaquer. Les 3 couleurs gris, jaune et rouge représentent 3 classes de vitesses des animaux

deux cartes électroniques qui gèrent à la fois l'acquisition (toutes les 1,6 secondes) sur un groupe de 16 huîtres et le transfert des données. La figure 1 montre le comportement de l'estimateur du 0.001-quantile de la distribution des vitesses de fermetures des valves. Cette information de vitesse fournit une indication importante du changement de comportement de l'animale comme par exemple une ponte ou un état de fébrilité (caractérisé par des fermetures-ouvertures partielles rapides) anormal lié à un stress.

Nous représentons ensuite graphiquement les 0.001-quantiles estimés par un gradient de couleurs (gris, jaune et rouge correspondant à 3 classes de mêmes effectifs) pour chaque instant $t \in (0, 24)$ entre le 4 mars et le 21 août sur le site de Locmariaquer. Ces mois d'enregistrements (Figure 2) nous ont permis de mettre en évidence des rythmes biologiques liés aux marées chez l'huître et visualiser des modifications importantes d'activités des animaux. Nous remarquons en particulier une zone jaune (jours 110 à 125) expliquée par un changement brutale des températures observées de manière indépendante par un capteur de température installé à proximité des huîtres et la zone la plus rouge correspondant à une activité plus intense des animaux liés à des activités notamment de reproduction (jours > 210). Ces résultats contribuent ainsi au développement d'un outil de surveillance de la qualité de l'eau basé sur l'analyse en continue du comportement de bivalves (bio-indicateur de pollution sur le terrain).

Remerciements : Nous remercions Pierre Ciret et Damien Tran de l'UMR EPOC pour le prêt du valvomètre et leurs aides lors de l'installation du matériel à Locmariaquer. Nous remercions aussi l'ensemble des collègues impliqués dans le projet ASPEET de l'université de Bretagne Sud.

Bibliographie

- [1] Durrieu, G. and Briollais L. (2009), Sequential Design for Microarray Experiments, *Journal of the American Statistical Association*, 104(486), 650-660.
- [2] Grama, I. and Spokoiny, V. (2008) Statistics of extremes by oracle estimation, *Annals of Statistics*, 36(4), 1619-1648.
- [3] Schmitt, F., De Rosa, M., Durrieu, G., Sow M., Ciret P., Tran D. and 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(12), 3565-3576.
- [4] Schwartzmann, C., Durrieu, G., Sow, M., Ciret, P., Lazareth, C. and Massabuau J.C. (2011) In situ giant clam growth rate behavior in relation to temperature, *Limnology and Oceanography*, 56(5), 1940-1951.
- [5] Sow, M., Durrieu, G., Briollais, L., Ciret, P. et Massabuau, J. C. (2011) Water quality assessment by means of HFNI valvometry and high-frequency data, *Environmental Modeling and Assessment*, 182, 155-170.
- [6] Tran, D., Nadau, A., Durrieu, G., Ciret, P., Parisot, J.C. and Massabuau J.C. (2011) Field Chronobiology of a Molluscan Bivalve : How the Moon and Sun Cycles Interact to Drive Oyster Activity Rhythms, *Chronobiology International*, 28(4), 307-317.