

# REVOLECO : ruptures dans l'évolution à long terme des écosystèmes côtiers

Atelier du 8 au 11 juillet 2019

## Compte rendu

### Contexte et objectifs

De nombreuses séries de données (pluri-)décennales existent au sein de la communauté océanographique côtière. Un premier regard collectif sur ces données (colloque EVOLECO 2017 ; <http://www.evoleco-2017.com/>), a permis de mettre en évidence l'existence de ruptures dans de nombreuses séries, en particulier entre le milieu des années 90 et le milieu des années 2000. Ces séries portent sur des paramètres et des compartiments très divers (climat, paramètres hydrologiques, structure des communautés zooplanctoniques ou macrozoobenthiques, isotopie de la macrofaune benthique, etc...) et sur des sites et empreintes géographiques diverses allant d'un écosystème à l'ensemble du littoral national selon les paramètres ou les compartiments.

L'objectif général de REVOLECO est, suite à l'identification des jeux de données présentant une rupture entre mi-90 et mi-2000, 1) la comparaison des méthodes statistiques permettant la mise en évidence de telles ruptures, 2) la mise en relation des jeux de données afin d'identifier leurs liens éventuels et 3) la mise en évidence des causalités et des mécanismes associés. L'atelier de juillet 2019 s'est focalisé sur le premier objectif.

### Organisation et déroulement

L'atelier a été co-financé par le RESOMAR et l'AFB. Il a été co-organisé par Eric Thiébaud (Sorbonne Université), Cédric Bacher (Ifremer) et Nicolas Savoye (Université de Bordeaux).

Suite à un appel à manifestation d'intérêt, 31 personnes se sont déclarées intéressées par l'atelier. 14 ont pu y participer (voir ci-dessous)

L'atelier s'est déroulé du 8 juillet 13h30 au 11 juillet 12h30 sur le campus de Jussieu de Sorbonne Université à Paris.

La première demi-journée a été consacrée à des présentations (voir ci-dessous). Les quatre demi-journées suivantes ont été consacrées à des séances de co-working, généralement en binôme, ponctuées de synthèses et de discussions collectives. Ces demi-journées ont constitué le cœur de l'atelier. Elles ont eu pour objectif majeur de tester les différents outils statistiques sur des séries temporelles de fréquence (hebdomadaire, mensuelle, annuelle), de durée (une à plusieurs décennies) et de nature (paramètres environnementaux, taxa) différentes. La dernière demi-journée a été consacrée à un bilan à chaud de l'atelier et à des perspectives concernant la suite à donner à REVOLECO.

### Présentations

- Présentation de l'atelier, de sa genèse, des attentes et synthèse des jeux de données des participants par les organisateurs.
- Retour sur l'atelier 'série chronologiques' (29-31/10 2018) par Dominique Soudant (Ifremer)
- Quelques méthodes de détection des changements écosystémiques abrupts par Eric Goberville (Sorbonne Université)
- Les modèles hiérarchiques multivariés pour mieux comprendre les changements de régimes par Guillaume Blanchet (Université de Sherbrooke au Canada)
- Utilisation des DLM (Dynamic Linear Model) dans la détection de ruptures Dominique Soudant (Ifremer)
- Dynamique temporelle de la biodiversité : vers la détection d'alarmes précoces de l'impact du changement global ? par David Carayon (IRSTEA)

## Participants

Maïa Akopian (Agence Française pour la Biodiversité), Elvire Antajan (Ifremer), Cédric Bacher (Ifremer), Guillaume Blanchet (Université de Sherbrook), Henrique Cabral (IRSTEA), David Carayon (IRSTEA), Eric Goberville (Sorbonne Université), Francis Gohin (Ifremer), Arnaud Lheureux (Université de Bordeaux), Maud Pierre (IRTSEA), Nicolas Savoye (Université de Bordeaux), Dominique Soudant (Ifremer), Eric Thiébaud (Sorbonne Université), Dorothée Vincent (Agence Française pour la Biodiversité).

## Outils testés

Les différents outils testés sont : Rodionov-stars et équivalents (T-tests), DLM (Dynamic Linear Model), AEM (Asymmetric Eigenvector Maps), MHM (Modèles Hiérarchiques Multidimensionnels), Ondelettes. Ils sont présentés ci-dessous.

## Bilan et perspectives

Les discussions de la dernière demi-journée 1) ont fait le constat que l'atelier s'est uniquement focalisé sur des questions méthodologiques et techniques, 2) ont fait apparaître différentes questions en suspens et quelques commentaires techniques et 3) ont finalement fait émerger le souhait de poursuivre les travaux REVOLECO entrepris lors de l'atelier de juillet 2019.

### Questions et commentaires techniques :

- Désaisonnaliser ou pas ? comment ? Faut-il enlever le 'bruit' ? comment ? La saisonnalité peut varier au cours du temps (échelle interannuelle à décennale) : à prendre en compte.
- Améliorer le Rodionov-stars pour le rendre robuste
- Transformer les données (log, autre) : comment ? pour quelles variables ?
- Ne pas regarder que la moyenne : les quantiles et autres
- Univarié versus multivarié et la combinaison des deux
- Espèces rares : à prendre en compte ou pas ?
- Faire des modèles mécanistiques simples pour tester des hypothèses
- Il semble y avoir un équivalent Rodionov-stars en multivarié
- DLM : passer à un vrai formalisme bayésien ; prise en compte les limites de quantification prise en compte de la variance d'observation variable dans le temps ; vers du multivarié ?

### Suite de REVOLECO

Une suite à REVOLECO est très clairement souhaitée d'une part pour finaliser les discussions méthodologiques (objectifs 1) et 2) ci-dessus) et d'autre part pour démarrer le travail 'de recherche' (objectifs 2) et 3) ci-dessus). Cette suite pourra se faire dans un premier temps avec l'organisation d'un autre atelier similaire avant le printemps 2020 (utilisation des reliquats de crédits) mais devra également passer par le dépôt des projets pour des financements à partir de 2020 ou 2021. Les programmes EC2CO, LEFE et ANR sont évoqués, de même que l'Appel à Manifestation d'Intérêt de l'AFB. Un GDR est également évoqué mais l'idée n'est pas retenue (trop lourd/rigide).

Le souhait d'une vitrine internet et d'un support pour diffusion, partage et stockage de documents sont également abordés. Le site web de l'infrastructure de recherche ILICO ([www.ir-ilico.fr/](http://www.ir-ilico.fr/)) est cité [accord de principe donné par la direction d'ILICO à l'automne 2019].

## Méthode STARS/SRSD (Rodionov)

Rodionov (2004, 2005) a mis au point la méthode STARS (Sequential T-test Analysis of Regime Shifts), renommée SRSD (Sequential Regime Shift Detection), basée sur une analyse séquentielle du F-test (la version STARS étant basée sur le t-test). Cette méthode permet de détecter des décalages dans trois statistiques d'une série temporelle (moyenne, variance et corrélation) et il est recommandé de travailler sur une série chronologique filtrée (option « Prewhitening » de traitement du bruit rouge ; Rodionov, 2006), afin de minimiser l'effet de l'autocorrélation temporelle : application d'un modèle autorégressif d'ordre 1 pour lequel deux méthodes d'estimation du coefficient sont proposées, i.e. MPK (Marriott and Pope, 1954 ; Kendall, 1954) et IP4 (Inverse Proportionality with 4 corrections ; Rodionov, 2006). La méthode STARS/SRSD est présentée par comme :

- Pouvant détecter automatiquement les changements de moyenne et de variance,
- Etant performante sur les « fins » de série temporelle, ce qui sous-entend une détection de changement en temps quasi-instantané,
- Pouvant détecter les changements à différentes échelles de temps,
- Ne nécessitant pas d'a priori sur les changements à détecter,
- Pouvant gérer l'autocorrélation

L'intensité et l'échelle temporelle de détection des changements peuvent être contrôlées par l'utilisateur, afin d'éviter - par exemple - d'identifier les cycles saisonniers comme étant des ruptures. L'algorithme permet le calcul d'un seuil de significativité associé au changement et donc de comparer, lorsque le cas se présente, les différentes ruptures identifiées au sein d'une même série chronologique. La méthode - qui tolère les données manquantes - permet de minimiser l'influence d'observations extrêmes, en leur attribuant un « poids » dans l'analyse (estimateur de Huber, 1981).

Les exemples d'utilisation dans la littérature sont nombreux, aussi bien en océanographie, en écologie terrestre, en climatologie, qu'en économie. SRSD est écrit en Visual Basic et peut donc être utilisé sur Excel ; des versions R et Matlab sont disponibles (cf. version d'essai de 30 jours sur <https://sites.google.com/site/climatologic/download>, ou la fonction R sur <http://esapubs.org/archive/ecol/E095/262/STARS.R>). Bien que largement utilisée, elle est également assez critiquée, notamment du fait d'une paramétrisation (e.g. valeurs « target significance level », « cutoff length » ou encore « Huber's weight parameter » à renseigner) par « essais et erreurs », parfois guidée par un a priori sur les observations ou encore de la sensibilité de la méthode dans la détection des ruptures.

### Bibliographie

- Rodionov, S. N. (2004). A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters*, 31(9).
- Rodionov, S. N. (2005). A brief overview of the regime shift detection methods. Large-scale disturbances (regime shifts) and recovery in aquatic ecosystems: challenges for management toward sustainability, 17-24.
- Rodionov, S. N. (2006). Use of prewhitening in climate regime shift detection. *Geophysical Research Letters*, 33(12).
- Marriott, F. H. C., & Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, 41(3/4), 390-402.
- Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3-4), 403-404.
- Huber, P. J., & Ronchetti, E. M. (1981). *Robust Statistics* John Wiley & Sons. New York, 1(1).

### Applications

<https://sites.google.com/site/climatologic/overview>

## Principe général de l'approche *Dynamic Linear Model*

Les *Dynamic Linear Models* (DLM) peuvent être vus comme une généralisation dynamique des modèles linéaires. Ainsi, la plupart des concepts et connaissances relatifs aux seconds (e.g. régression, multivariée, multivarié) restent applicables aux premiers à travers un formalisme très similaire. Le caractère dynamique se manifeste par des paramètres variables dans le temps. Il s'ensuit que ce sont à proprement parler des modèles de traitement de séries temporelles. À ce titre, ils intègrent en particulier l'existence de dépendance entre les observations et de mémoire des processus. Le caractère dynamique permet de s'abstraire de l'hypothèse de stationnarité et facilite l'intégration des informations exogènes sous forme d'interventions. Plus spécifiquement, des approches diagnostiques permettent de suggérer des valeurs exceptionnelles et des changements de niveau qui peuvent être intégrés au modèle et dont la pertinence est évaluée à travers l'estimation des paramètres correspondants par maximum de vraisemblance. Dans l'application proposée, cette procédure a été automatisée. Toutefois, attendu qu'il ne s'agit que d'optimisation numérique, le modèle résultant nécessite une validation experte et thématique (e.g. chimie, hydrologie, biologie). Par ailleurs, le modèle produit un ensemble de quantités (e.g. pente, saisonnalité, variances) essentielles à l'étude des séries temporelles mais également support à de nouvelles interprétations (e.g. phénologie). En dépit de la richesse de l'approche, elle reste rudimentaire relativement aux perspectives méthodologiques (e.g. *clusterisation* des tendances, modélisation spatio-temporelle, variances variables dans le temps, modèles hiérarchiques, modèles robuste, formalisation bayésienne). Enfin, il faut noter que la méthode ne propose pas aujourd'hui de solution opérationnelle au traitement des valeurs au-dessous des limites de quantification, caractéristiques de très nombreuses mesures en écologie.

### Bibliographie

#### *de référence :*

West, M., Harrison, J., 1997. Bayesian forecasting and dynamic models. Springer-Verlag.

Harvey, A., Koopman, S.J., Penzer, J., 1998. Messy time series : a unified approach. *Advances in Econometrics* 13, 103–143.

#### *pratique :*

Soudant Dominique, Miossec Laurence, Neaud-Masson Nadine, Auby Isabelle, Maurer Daniele, Daniel Anne (2016). Incertitudes des méthodes d'évaluation « eaux littorales » : utilisation de modèles linéaires dynamiques pour l'évaluation des incertitudes (chlorophylle a, phytoplancton). AQUAREF 2015. <https://archimer.ifremer.fr/doc/00345/45594/>

Soudant Dominique, Auby Isabelle, Daniel Anne (2018). Incertitudes des méthodes d'évaluation « eaux littorales » : utilisation de modèles linéaires dynamiques pour l'évaluation des incertitudes des paramètres hydrologiques. Transférer les méthodes de bio indication. Compléments apportés au guide méthodologique et EIL phytoplancton et participation au CST. Rapport AQUAREF 2017. <https://archimer.ifremer.fr/doc/00441/55306/>

#### *de programmation :*

Petris, G., Petrone, S., Campagnoli, P., 2009. Dynamic linear models with R, Use R ! Springer.

R Package « dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models » <https://cran.r-project.org/web/packages/dlm/index.html>

## Cartes de vecteurs propres asymétriques (AEM : asymmetric eigenvector maps)

Les cartes de vecteurs propres asymétriques (*asymmetric eigenvector maps*, AEM) sont des variables construites spécialement pour capturer les variations temporelles d'une (ou d'un groupe) d'espèce(s) ou de tout autre facteur d'intérêt pour lequel on a un intérêt. Techniquement, les AEMs sont des variables orthogonales qui comportent une composante directionnelle imposée lors de leurs constructions. Comme les AEMs sont des variables temporelles, elles n'ont pas besoin d'être utilisées dans un cadre analytique spécifique. On peut utiliser les AEMs comme variables explicatives dans des régressions linéaires (univariées), des analyses canoniques (multivariées) ou tout autre cadre analytique où l'on cherche à explorer une série temporelle. Lorsque l'on construit des AEMs, il en résulte autant d'AEM que d'échantillon dans une série temporelle moins 1. Donc, lorsque l'on utilise des AEMs il faut choisir lesquelles sont intéressantes pour bien expliquer la série temporelle d'intérêt. Ceci peut être fait séquentiellement, en utilisant une technique de sélection de variables (LASSO [*least absolute shrinkage and selection operator*], sélection pas à pas, etc.), ou toute autre approche permettant de réduire le nombre de AEMs à utiliser. De plus, les AEMs n'ont pas besoin d'échantillon distribué régulièrement, elles ont été conçues pour être applicables sur des jeux de données qui présentent des pas de temps irréguliers. Par contre, les AEMs ne sont pas bien adaptés à capturer des ruptures dans les données, elles sont plus efficaces à mesurer des variations dans les gradients temporelles. À ce titre, lorsqu'on cherche des ruptures dans une série temporelle, les AEMs ont avantage à être utilisés à titre exploratoire.

### Bibliographie

- Blanchet, F. G., P. Legendre, and D. Borcard. 2008. Modelling directional spatial processes in ecological data. *Ecological Modelling* 215:325–336.
- Blanchet, F. G., P. Legendre, R. Maranger, D. Monti, and P. Pepin. 2011. Modelling the effect of directional spatial ecological processes at different scales. *Oecologia* 166:357–368.
- Legendre, P., and O. Gauthier. 2014. Statistical methods for temporal and space–time analysis of community composition data. *Proceedings of the Royal Society B: Biological Sciences* 281:20132728.

### Package R

- Dray, S., D. Bauman, F. G. Blanchet, D. Borcard, S. Clappe, G. Guenard, T. Jombart, G. Larocque, P. Legendre, N. Madi, and H. H. Wagner. 2019. *adespatial: Multivariate Multiscale Spatial Analysis*.

## Modèles hiérarchiques multidimensionnels

Les modèles hiérarchiques multidimensionnels (MHMs) sont une généralisation multidimensionnelle des modèles linéaires généralisés mixtes qui ont la particularité d'être flexibles, permettant donc d'adapter les modèles à diverses questions écologiques d'intérêt. Les MHMs ont la particularité d'être applicable autant pour des données univariées que multivariées, malgré que leurs intérêts s'imposent dans un cadre multivarié. Pour étudier les changements dans les séries temporelles, un MHM peut être utilisé en lien avec des AEMs. Lorsqu'on cherche une rupture dans une série temporelle, on peut aussi tirer avantage de la composante aléatoire de ce type de modèle et ainsi inclure dans un MHM un effet aléatoire pour capturer de potentielle rupture dans une série temporelle multidimensionnelle. Par contre, malgré leurs flexibilités, aucune variante des MHMs n'a été développée pour prendre en compte de façon explicite les ruptures dans les séries temporelles. À ce titre, pour rendre les MHMs réellement attrayants pour l'étude de rupture dans les séries temporelles multidimensionnelles, il deviendrait intéressant de développer une extension des MHMs qui permettrait de répondre à ce type de problème explicitement.

### Bibliographie

- Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters* 20:561–576.
- Ovaskainen, O., N. Abrego, P. Halme, and D. Dunson. 2016. Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution* 7:549–555.

### Package R

- Blanchet, F. G., G. Tikhonov, and A. Norberg. 2019. HMSC: Hierarchical Modelling of Species Community.

## Description des analyses par ondelettes

L'analyse par ondelettes est une méthode développée pour le traitement de signal qui peut être considérée comme une version flexible de l'analyse par série de Fourier. Pour l'étude de série temporelle, l'analyse par ondelettes est très intéressante, car elle permet de visualiser via un diagramme de puissance les régions d'une série temporelle qui présente une régularité en plus d'informer sur la période de la régularité observée. En utilisant cette propriété des ondelettes, il est aussi possible de les utiliser pour déceler les ruptures dans les séries temporelles ; c'est-à-dire en recherchant les régions d'une série temporelle où les régularités disparaissent. Les ondelettes peuvent aussi être utilisées en mettant en relation deux variables entre elles, en étudiant l'évolution dans le temps de la codistribution d'une variable (p.ex. une espèce) avec une autre (p.ex. une autre espèce ou une composante de l'environnement). Il existe aussi des développements proposant une variante de l'analyse par ondelettes multidimensionnelle. L'inconvénient majeur associé à toutes variantes d'analyses par ondelettes est qu'elles nécessitent que les échantillons soient amassés de façon régulière dans le temps (p.ex. chaque jour, semaine, mois, année, etc.), ce qui n'est pas évident pour des données biologiques.

### Bibliographie

#### *Introduction*

Torrence, C., and G. P. Compo. 1998. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society* 79:61–78.

Cazelles, B., M. Chavez, D. Berteaux, F. Ménard, J. O. Vik, S. Jenouvrier, and N. C. Stenseth. 2008. Wavelet analysis of ecological time series. *Oecologia* 156:287–304.

#### *Multidimensionnel*

Park, T., I. A. Eckley, and H. C. Ombao. 2014. Estimating Time-Evolving Partial Coherence Between Signals via Multivariate Locally Stationary Wavelet Processes. *IEEE Transactions on Signal Processing* 62:5240–5250.

### Package R

#### *Général*

Aldrich, E. 2019. wavelets: Functions for Computing Wavelet Filters, Wavelet Transforms and Multiresolution Analyses.

Constantine, W., and D. Percival. 2017. wmts: Wavelet Methods for Time Series Analysis.

Roesch, A., and H. Schmidbauer. 2018. WaveletComp: Computational Wavelet Analysis.

#### *Bidimensionnel*

Gouhier, T. C., A. Grinsted, and V. Simko. 2019. biwavelet: Conduct Univariate and Bivariate Wavelet Analyses.

#### *Multidimensionnel*

Taylor, S., T. Park, I. Eckley, and R. Killick. 2019. mvLSW: Multivariate, Locally Stationary Wavelet Process Estimation.