



# Modélisation multiéchelle de la relation entre une exposition et une issue sanitaire

## SURVEILLANCE DES MALADIES CHRONIQUES

Numéro 25

### PRINCIPAUX CONSTATS

**+** Les séries météorologiques et sanitaires contiennent un ensemble de cycles complexes qui expliquent les tendances temporelles et rendent plus difficile la compréhension d'une relation de régression.

La séparation des cycles permet de scanner la relation entre une issue sanitaire et une ou plusieurs expositions. La méthodologie détaille également la relation non linéaire existant avec certaines expositions.

La régression-EMD permet donc l'étude approfondie d'une relation à différentes échelles temporelles souvent peu étudiées en surveillance (jours, mois, décades). En prenant en compte les cycles de longue périodicité, elle permet de plus une anticipation à long terme de l'évolution d'une issue sanitaire, ce qui s'avère utile en matière de changements climatiques.

### Introduction

En épidémiologie, l'analyse de la relation entre une exposition environnementale et une issue sanitaire se fait généralement avec un modèle de régression. Les expositions peuvent être des polluants atmosphériques ou des conditions météorologiques, alors que des décès ou des hospitalisations toutes causes ou pour causes spécifiques, comme les maladies respiratoires ou cardiovasculaires (MCV), comptent parmi les issues sanitaires. Les données utilisées consistent, généralement, en des mesures ayant un pas de temps régulier (quotidiennes dans la plupart des cas) des variables d'intérêt et constituent donc des séries temporelles.

Les variables considérées sont le reflet d'interactions complexes entre un grand nombre de facteurs tels que la météorologie, les cycles climatiques ou encore l'activité humaine. Il en résulte des séries temporelles présentant des cycles complexes avec des échelles temporelles très différentes. L'imbrication de tous ces cycles rend difficile la compréhension globale d'une relation santé/climat et peut même, parfois, cacher certains aspects de la relation. D'un point de vue statistique, la complexité de ces séries temporelles cause la présence d'autocorrélation (la dépendance d'une observation sur les suivantes), voire de non-stationnarité (l'évolution temporelle de la distribution statistique des données). De plus, lorsque plusieurs expositions sont considérées, elles sont souvent interdépendantes à cause d'interactions naturelles (p. ex. : la température et les précipitations). Tous ces aspects violent les hypothèses classiques de la régression, entraînant des estimations peu fiables (Schwartz et collab., 1996).

Il est proposé dans ce projet de décomposer les séries de données par la méthode de décomposition modale empirique (EMD pour *empirical mode decomposition* en anglais) afin d'en extraire leurs cycles (Huang et collab., 1998), et par la suite d'utiliser ces cycles basiques dans des modèles de régression.

L'objectif d'une telle démarche (appelée régression-EMD, ou R-EMD) est double. Premièrement, elle permet de séparer les cycles les uns des autres pour pouvoir les étudier séparément et ainsi comprendre à quelles échelles la relation santé/climat se joue. Deuxièmement, pour revenir au plan statistique, les cycles basiques consistent en des « sous-séries » plus régulières que les séries observées et atténuent donc en pratique les problèmes d'autocorrélation de non-stationnarité et d'interdépendance.

Une méthodologie similaire à la R-EMD n'a pour l'instant été considérée que par Yang et collab., (2011) pour l'étude des maux de tête à Taiwan. Or, elle a le potentiel d'augmenter la connaissance concernant la relation recherchée et donc d'affiner les alertes sanitaires en fonction des variations météorologiques. La méthodologie proposée, les outils statistiques utilisés ainsi que les résultats complets sont présentés de façon détaillée dans le rapport Masselot et collab., (2015;2018).

## Décomposer les données : la méthode EMD

La décomposition de séries de données est très utilisée dans certains domaines comme les géosciences, et de nombreuses méthodes mathématiques existent à ce sujet. Les plus connues sont les décompositions de Fourier, d'ondelettes ou, plus simplement, la décomposition tendance-saisonnalité par Loess (plus connue sous le nom de STL pour *seasonal-trend decomposition*). Cependant, la méthode retenue ici est EMD, car contrairement à celles citées ci-dessus, elle est entièrement empirique et s'adapte aux données. Les sous-sections suivantes présentent brièvement la méthode EMD, puis son extension multivariée, et l'extension assistée par bruit qui répond à une lacune de l'algorithme de base. Le détail des algorithmes se trouve en annexe 1.

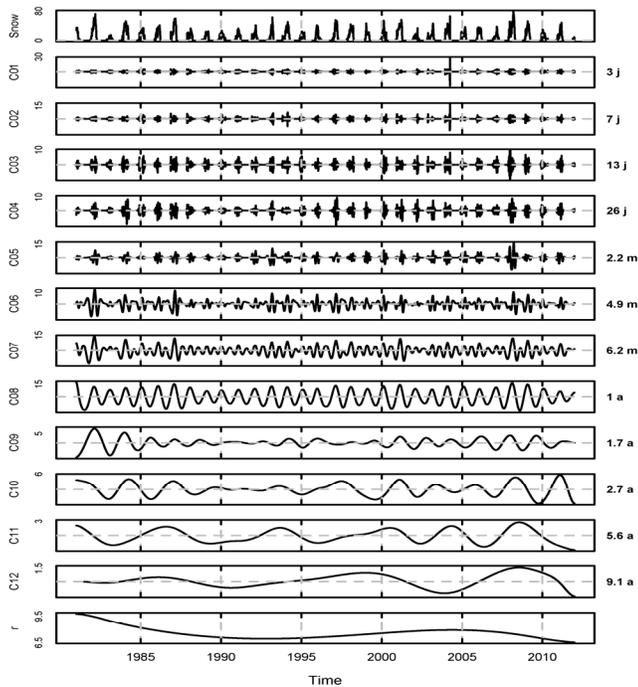
## EMD de base

Les modes d'oscillations obtenus par EMD sont appelés IMFs (fonctions modales intrinsèques). Une IMF est une fonction ayant deux contraintes : 1) tous ses maxima sont positifs et ses minima négatifs et 2) la moyenne entre son enveloppe supérieure et son enveloppe inférieure est zéro partout. Il s'agit donc de fonctions présentant des oscillations régulières autour de zéro, mais dont la flexibilité permet de capter des irrégularités dans un cycle. D'une décomposition EMD, il résulte donc un faible nombre de composantes facilement interprétables. La figure 1 illustre les avantages d'EMD en montrant la décomposition de la série de données de hauteur de neige de la région métropolitaine de recensement (RMR) de Montréal entre 1981 et 2011.

Par exemple, les composantes  $C_1$  à  $C_4$  illustrent bien l'adaptabilité des IMFs, car elles sont nulles hors des périodes estivales, et la composante  $C_8$  représente bien le cycle annuel de la neige.

À partir d'une série, les IMFs correspondantes sont obtenues itérativement en commençant par celui de plus haute fréquence jusqu'à obtenir un reste qui n'est pas une IMF et qui est assimilé à la tendance de la série. Pour plus de détails, voir l'algorithme 1 qui se trouve dans l'annexe 1.

**Figure 1** Décomposition de la série de données de hauteur de neige de la RMR de Montréal entre 1981 et 2011 inclusivement.



## EMD multivarié

La principale extension de l'algorithme EMD présenté ci-dessus est l'EMD multivarié (MEMD, Rehman et Mandic, 2010). Il vise à décomposer une série multivariée, c.-à-d. décomposer conjointement un ensemble de séries que l'on considère comme liées les unes aux autres. La météorologie en est un bon exemple dans la mesure où les différentes variables (température, précipitations, etc.) peuvent être vues comme faisant partie d'un seul grand phénomène qui est la météorologie. L'avantage d'une telle approche est qu'elle résulte en un nombre identique d'IMFs pour chaque variable de la série multivariée. De plus, toutes les IMFs de mêmes ordres ont une périodicité similaire, ce qui permet une cohérence entre les variables (p. ex. : toutes les IMFs d'ordre 8 représentent les variations annuelles). Cette extension est donc retenue et recommandée dans l'application de la régression-EMD. La procédure est décrite par l'algorithme 2 de l'annexe 1.

## EMD assisté par bruit

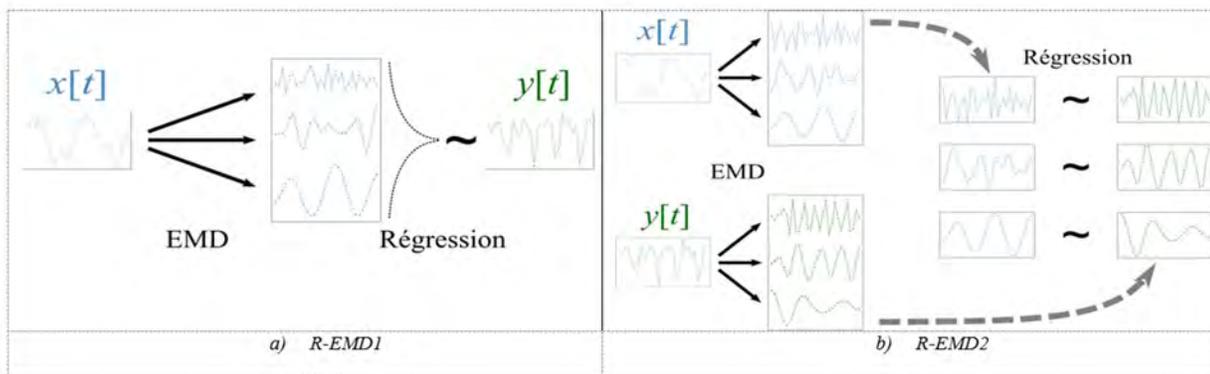
L'autre extension importante de la méthode EMD est l'assistance par bruit pour régler le problème du mélange de modes. Ce dernier terme désigne une décomposition dans laquelle une ou plusieurs IMFs présentent des changements importants de périodicité, entravant ainsi une bonne interprétation de la décomposition. Le mélange de modes survient lorsque des cycles sont intermittents comme les hautes fréquences dans la hauteur de neige qui n'existent que durant l'hiver (voir figure 1). Ce problème est réglé en utilisant la méthode MEMD et en ajoutant une ou plusieurs variables supplémentaires à la série, sous la forme de bruits blancs. Un bruit blanc étant une série dans laquelle toutes les échelles de variation sont représentées. Sa présence enlève le problème d'intermittence, ce qui évite donc le mélange de modes. Les IMFs issues de ces variables bruits blancs supplémentaires sont, à la fin, retirées des résultats. L'algorithme 3 en annexe 1 détaille la procédure à appliquer.

## Régression EMD

### Quelles variables décomposer?

Deux cas ont été retenus pour présenter la régression-EMD : 1) seules les expositions sont décomposées (R-EMD1), et 2) à la fois les expositions et la réponse sont décomposées (R-EMD2). La figure 2 illustre la différence entre les deux cas.

**Figure 2** Illustration des deux méthodes de régression-EMD



Dans le premier cas, les IMFs des variables explicatives sont utilisées pour expliquer la réponse telle que mesurée. Elle permet de mettre en évidence les fréquences des expositions les plus importantes pour l'étude de la réponse sanitaire. Il s'agit donc d'une modélisation qui a pour but d'avoir une vision globale de la relation.

Le deuxième cas vise à décomposer la variable sanitaire et à expliquer chacune de ses IMFs en fonction des IMFs météorologiques de même ordre. La R-EMD2 comprend donc un modèle de régression pour chaque IMF obtenue. Elle permet d'explorer la relation à toutes les échelles temporelles présentes dans les données. Les deux cas sont résumés en annexe 2.

### Décaler les IMFs

Avant d'utiliser les IMFs dans la régression, il peut être nécessaire de leur appliquer un retard (*lag*). En effet, il est possible qu'il y ait un temps de latence entre un cycle météorologique et son impact sur l'issue sanitaire. Le décalage d'une IMF explicative permet donc de prendre en compte ce potentiel temps de latence.

Le décalage à appliquer à chaque IMF est déterminé en analysant la fonction de corrélation croisée entre l'IMF et la réponse (soit l'issue sanitaire directe, soit l'une de ses IMFs si l'on est dans la R-EMD1 ou R-EMD2 respectivement). La fonction de corrélation croisée indique la corrélation entre deux séries selon le décalage appliqué. Ainsi, le décalage qui maximise la fonction de corrélation croisée (en valeur absolue) est celui qui doit être appliqué. Ce décalage est cependant restreint aux valeurs inférieures à la périodicité de l'IMF explicative, car il est peu probable que le temps de latence soit supérieur. Appliquer la régression : le Lasso

La méthode du Lasso (Tibshirani, 1996) est retenue pour la régression dans la R-EMD1 et la R-EMD2. Cette méthode a été développée pour la régression à très grande dimension. Deux objectifs sont ainsi visés : 1) réduire la variance induite par la colinéarité entre les variables explicatives et 2) opérer une sélection des variables directement lors de l'estimation des coefficients de régression. Le principe du Lasso est d'estimer les coefficients par moindres carrés (comme dans la régression classique) en ajoutant une pénalisation sur leur amplitude. Cette pénalisation permet de privilégier les modèles avec peu de coefficients différents de zéro et donc d'écarter du modèle les variables peu importantes dans l'explication de la réponse. Les détails de l'estimation par Lasso sont donnés dans l'annexe 1.

Le Lasso permet donc de gérer le grand nombre potentiel d'IMFs qui résulte de la décomposition de plusieurs expositions en ne conservant dans le modèle que celles qui expliquent le mieux la réponse. En outre, le Lasso permet de gérer la colinéarité potentielle entre IMFs de même ordre. Il est donc choisi comme modèle de régression à la fois pour la R-EMD1 et la R-EMD2.

### Choisir la sévérité du Lasso : la validation croisée

L'estimation d'une régression par Lasso dépend d'un paramètre (généralement appelé  $\lambda$ ) qui contrôle la sévérité de la sélection de variables. Plus ce paramètre est important, plus le nombre de variables restantes dans le modèle final est faible. Ce paramètre ayant une grande influence sur les résultats, il est déterminé objectivement en utilisant un critère d'erreur de prédiction par validation croisée.

Le principe de la validation croisée est de segmenter le jeu de données en sous-échantillons disjoints. Ensuite, les sous-échantillons (sauf un) sont regroupés (appelés échantillons d'apprentissage) et utilisés pour ajuster le Lasso. Puis, ce modèle ajusté est utilisé pour prédire les données du sous-échantillon restant (appelé échantillon de validation). On recommence cette procédure plusieurs fois, en changeant chaque fois d'échantillon de validation. En comparant les prédictions obtenues et les vraies observations des échantillons de validation, on peut calculer une valeur d'erreur de prédiction. Le meilleur modèle est ainsi, celui qui minimise cette erreur.

En général, les échantillons sont déterminés aléatoirement, mais une hypothèse de la validation croisée est l'indépendance des échantillons d'apprentissage et de validation. Dans le cas de la régression-EMD, cette hypothèse est violée, car les données sont des séries ayant une dépendance temporelle. La validation croisée utilisée ici est donc la validation croisée par *hv*-bloc (Racine, 2000). Cette technique découpe les échantillons en blocs de données successives et écarte à chaque fois les blocs entourant le bloc de validation, tel qu'illustré sur la figure 3.

**Figure 3 Illustration de la validation croisée par *hv*-bloc pour les séries temporelles. Chaque bloc représente une partie de la série. Le bloc vert représente le bloc de validation courant; les blocs rouges, les blocs d'apprentissage courants. Les blocs gris ne sont utilisés ni pour l'apprentissage ni pour la validation**



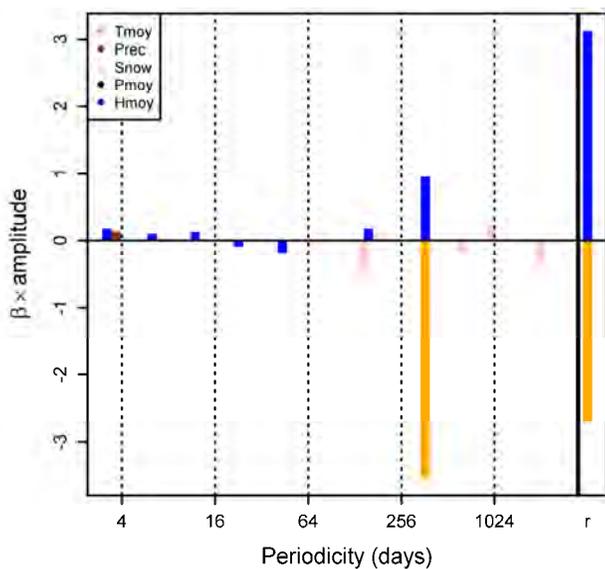
### Représenter les résultats : la sensibilité

L'application de la régression donne un ensemble de coefficients indiquant l'influence de chaque IMF sur la réponse sanitaire (R-EMD1) ou sur l'une de ses IMFs (R-EMD2). Cependant, comme les IMFs obtenues ont toutes des amplitudes différentes (voir p. ex. : la 0, il est peu pertinent de comparer directement les coefficients. Il est donc recommandé de calculer des valeurs de « sensibilité ». La sensibilité consiste en une standardisation des coefficients et plus précisément, à multiplier un coefficient estimé par l'amplitude pic-à-pic moyenne de l'IMF correspondante. L'amplitude pic-à-pic moyenne d'une IMF correspond à la différence moyenne entre un maximum et un minimum (elle est, par exemple, d'environ 30 pour l'IMF  $C_8$  dans la figure 1, où cette IMF oscille entre -15 et 15). La sensibilité ainsi calculée indique l'amplitude des variations de mortalité expliquée par l'IMF correspondante.

La figure 4 montre un exemple de résumé des résultats en présentant les sensibilités obtenues en fonction de la période moyenne des IMFs correspondantes. Dans cette figure, la sensibilité à un an de périodicité égale à -3,5

(barre orange négative) indique que le cycle annuel de température explique une différence de -3,5 décès quotidiens entre le maximum (l'été) et le minimum (l'hiver). Il est à noter que cette figure indique également une relation entre mortalité et humidité à une périodicité comprise entre 1 et 2 mois, ainsi qu'une relation avec la hauteur de neige à l'échelle de plusieurs années.

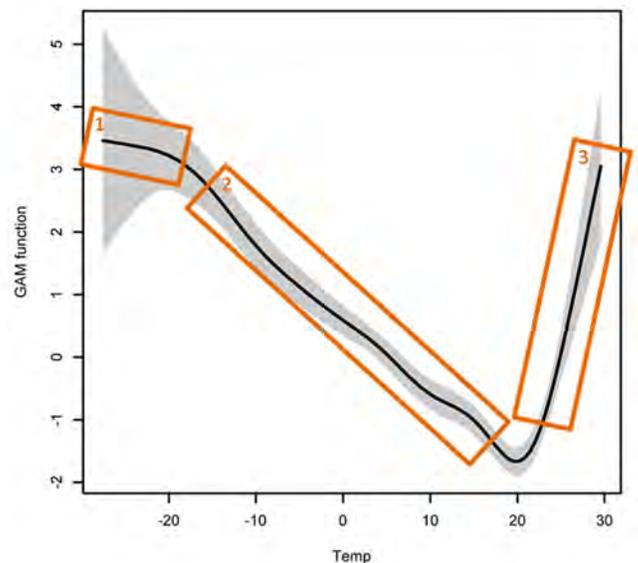
**Figure 4 Sensibilités obtenues par R-EMD1 en fonction de la période moyenne des IMFs pour l'explication des décès par MCV de la RMR de Montréal<sup>1</sup>. La source de données : MSSS, Fichier des décès (produit électronique).**



## La non-linéarité de la relation santé/climat

La modélisation de régression-EMD utilise Le Lasso qui est un modèle de régression linéaire (voir annexe 1). Il est pourtant bien connu que la relation entre une issue sanitaire et une exposition environnementale est souvent non linéaire (Doyon et collab., 2006). Les relations en J, V ou U obtenues peuvent néanmoins être vues comme linéaires par morceaux, comme illustré pour la température sur la figure 5. En utilisant la R-EMD, la pente de chacun des morceaux identifiés dans cette figure correspond au coefficient associé à une IMF. Ainsi la boîte 2 correspond à la différence entre l'été et l'hiver. Obtenue pour l'IMF de période 1 an, alors que les boîtes 1 et 3 correspondent à des IMFs de période très courte, représentant plutôt les vagues de froid et de chaleur.

**Figure 5 Relation en J entre les décès par MCV et la température de la RMR de Montréal<sup>1</sup>, avec les trois morceaux linéaires identifiés.**



<sup>1</sup>. Requête spéciale de données dans le cadre du plan d'action 2013-2020 sur les changements climatiques – volet santé.

## Conclusion

Habituellement, les études d'épidémiologie environnementale se concentrent sur les événements à court terme en utilisant directement les données quotidiennes et en attribuant les cycles de plus longs termes à une composante lisse du temps. Cependant, une telle approche ne permet pas de mettre en évidence des aspects peu visibles d'une relation entre une issue sanitaire et une exposition.

Ce document présente donc une méthodologie de régression-EMD qui permet l'étude approfondie d'une relation à différentes échelles temporelles. Une description de la méthodologie est fournie avec, en plus, un outil de description des résultats : la sensibilité. Cette méthodologie permet de mettre en évidence différentes facettes d'une relation à des échelles de temps peu étudiées. Cette méthodologie permet également de décomposer les différentes parties d'une relation non linéaire (communément retrouvée dans les études santé-climat).

La méthodologie de régression-EMD, fonctionnant comme un scanner, permet d'améliorer la connaissance de la relation étudiée. En prenant en compte les cycles de longue périodicité, elle permet de plus, une anticipation à long terme de l'évolution d'une issue sanitaire. Il s'agit donc d'un outil utile dans un contexte de changement climatique afin de comprendre les impacts sanitaires potentiels des variations du climat à grande échelle et d'améliorer leur surveillance.

## Bibliographie

Doyon, B., D. Bélanger et P. Gosselin (2006). "Effets du climat sur la mortalité au Québec méridional de 1981 à 1999 et simulations pour des scénarios climatiques futurs." Institut national de santé publique du Québec.

Friedman, J., T. Hastie et R. Tibshirani (2010). "CRregularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33(1): 1-22.

"The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454(1971): 903-995.

Masselot, P., F. Chebana, D. Bélanger, A. St-Hilaire, B. Abdous, P. Gosselin et T. B. M. J. Ouarda (2015). Régression EMD avec application à la relation entre les maladies cardiovasculaires et le climat, INRS-ETE.

Masselot, P., F. Chebana, D. Bélanger, A. St-Hilaire, B. Abdous, P. Gosselin et T. B. M. J. Ouarda (2015). Régression EMD avec application à la relation entre les maladies cardiovasculaires et le climat, INRS-ETE.

Masselot P, Chebana F, Bélanger D, St-Hilaire A, Abdous B, Gosselin P, Ouarda T (2017). EMD-regression for modelling multi-scale relationships, and application to weather-related cardiovascular mortality, *Science of The Total Environment*, Volume 612, 15 January 2018, Pages 1018-1029, ISSN 0048-9697, <https://doi.org/10.1016/j.scitotenv.2017.08.276>.

Racine, J. (2000). "Consistent cross-validated model selection for dependent data: hv-block cross-validation." *Journal of Econometrics* 99(1): 39-61.

Rehman, N. et D. P. Mandic (2010). "Multivariate empirical mode decomposition." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 466(2117): 1291-1302.

Rehman, N. U., C. Park, N. E. Huang et D. P. Mandic (2013). "EMD Via MEMD: Multivariate Noise-Aided Computation of Standard EMD." *Advances in Adaptive Data Analysis* 05(02): 1350007.

Rilling, G., P. Flandrin et P. Gonçalves (2003). On empirical mode decomposition and its algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP*.

Schwartz, J., C. Spix, G. Touloumi, L. Bachárová, T. Barumamdzadeh, A. le Tertre, T. Piekarksi, A. Ponce de Leon, A. Pönkä, G. Rossi, M. Saez et J. P. Schouten (1996). "Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions." *Journal of Epidemiology and Community Health* 50(Suppl 1): S3-11.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1): 267-288.

Yang, A. C., J.-L. Fuh, N. E. Huang, B.-C. Shia, C.-K. Peng et S.-J. Wang (2011). "Temporal Associations between Weather and Headache: Analysis by Empirical Mode Decomposition." *PLoS ONE* 6(1): e1461.

## Annexe 1 Algorithmes et outils mathématiques

### 1 Décomposition modale empirique : algorithmes

#### Algorithme 1 : décomposition modale empirique basique (Huang et collab., 1998)

Soit  $x(t)$ ,  $t = 1, \dots, N$ , une série de données :

1. Déterminer les maxima et minima locaux de  $x(\cdot)$  un maximum local correspond à une valeur  $x(t)$  telle que  $x(t) > x(t-1)$  et  $x(t) > x(t+1)$ , et inversement pour un minimum ;
2. Ajuster l'enveloppe supérieure  $s(t)$  par *spline* cubique sur les maxima locaux et l'enveloppe inférieure  $i(t)$  sur les minima locaux ;
3. Calculer la moyenne locale  $m(t) = (s(t) + i(t)) / 2$  ;
4. Retrancher la moyenne locale à la série pour obtenir le prototype IMF  $h_1(t) = x(t) - m(t)$  ;
5. Recommencer les étapes 1 à 4 sur  $h_1(t)$  jusqu'à obtenir un  $h_k(t)$  satisfaisant le critère d'arrêt ;
6. Fixer la première IMF  $C_1(t) = h_k(t)$ , puis recommencer les étapes 1 à 5 sur  $r_1(t) = x(t) - C_1(t)$ , puis sur  $r_2(t) = r_1(t) - C_2(t)$  et ainsi de suite jusqu'à ce que  $r_n(t)$  contienne moins de deux extrema.

Le critère d'arrêt recommandé à l'étape 5 est celui de Rilling et collab., (2003) qui considère  $h_k(t)$  comme une IMF lorsque le rapport de l'amplitude de  $m(t)$  sur l'amplitude de  $h_k(t)$  est inférieur à 5 %.

#### Algorithme 2 : l'EMD multivarié (Rehman et Mandic, 2010)

Soit  $x(t)$ ,  $t = 1, \dots, N$ , une série de données multivariée contenant  $P > 1$  variables :

1. Déterminer un ensemble  $L$  de directions uniformément sur un espace de dimension  $P - 1$  ;
2. Projeter la série multivariée  $x(t)$  sur chacune des directions déterminées pour obtenir  $L$  projections  $p_l(t)$  unidimensionnelles ;
3. Déterminer l'enveloppe supérieure  $S_l(t)$  de tous les  $p_l(t)$  (tels que dans les étapes 1 et 2 de l'Algorithme 1) ;
4. Calculer la moyenne locale  $m(t) = \frac{1}{L} \sum_{l=1}^L S_l(t)$  ;
5. Appliquer le reste de l'algorithme EMD de base, c.-à-d. les étapes 4 à 6 de l'Algorithme 1.

Le paramètre important de l'Algorithme 2 est le nombre de directions  $L$ . Rehman et Mandic (2010) recommandent de choisir  $L$  largement supérieur à  $P$ , proposant  $L = 64$  comme valeur de base.

#### Algorithme 3 : l'EMD assisté de bruit (Rehman et, 2013)

Soit  $x(t)$ ,  $t = 1, \dots, N$ , une série de données contenant  $P \geq 1$  variables :

1. Générer  $Q$  bruits blancs de taille  $N$  selon une loi normale ;
2. Appliquer l'Algorithme 2 (MEMD) à la série multivariée de taille  $P + Q$  comprenant  $x(t)$  et les bruits blancs générés ;

3. Retirer des résultats les IMFs correspondant aux variables de bruit blanc.

Dans ce cas, les paramètres à déterminer sont le nombre  $Q$  et la variance des bruits blancs générés. Rehman et collab., (2013) recommandent d'utiliser  $Q = 2$  et une variance comprise entre 2 % et 10 % de la variance moyenne de  $x(t)$ .

## 2 Le Lasso

Soit le modèle de régression linéaire  $y_i = \sum_{j=1}^P \beta_j x_{ij} + \varepsilon_i$ ,

où  $y_i$  est la réponse, les  $x_{ij}$  ( $i = 1, \dots, N$  et  $j = 1, \dots, P$ ) sont les variables explicatives,  $\beta_j$  les coefficients de régression et  $\varepsilon_i$  les résidus. Le Lasso (Tibshirani, 1996) estime les coefficients  $\beta_j$  en minimisant un critère de moindres carrés pénalisés, c.-à-d. :

$$\hat{\beta}_j = \arg \min_{\beta_j} \sum_{i=1}^N \left( y_i - \sum_{j=1}^P \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^P |\beta_j| \quad (1)$$

Où  $\lambda$  contrôle la sévérité de la pénalisation  $\sum_{j=1}^P |\beta_j|$ . Plus

$\lambda$  est élevé, plus les modèles complexes sont pénalisés et plus le nombre d'estimations  $\hat{\beta}_j$  nulles est grand. La minimisation du critère (1) se fait par un algorithme développé par Friedman et collab., (2010).

## 3 La sensibilité

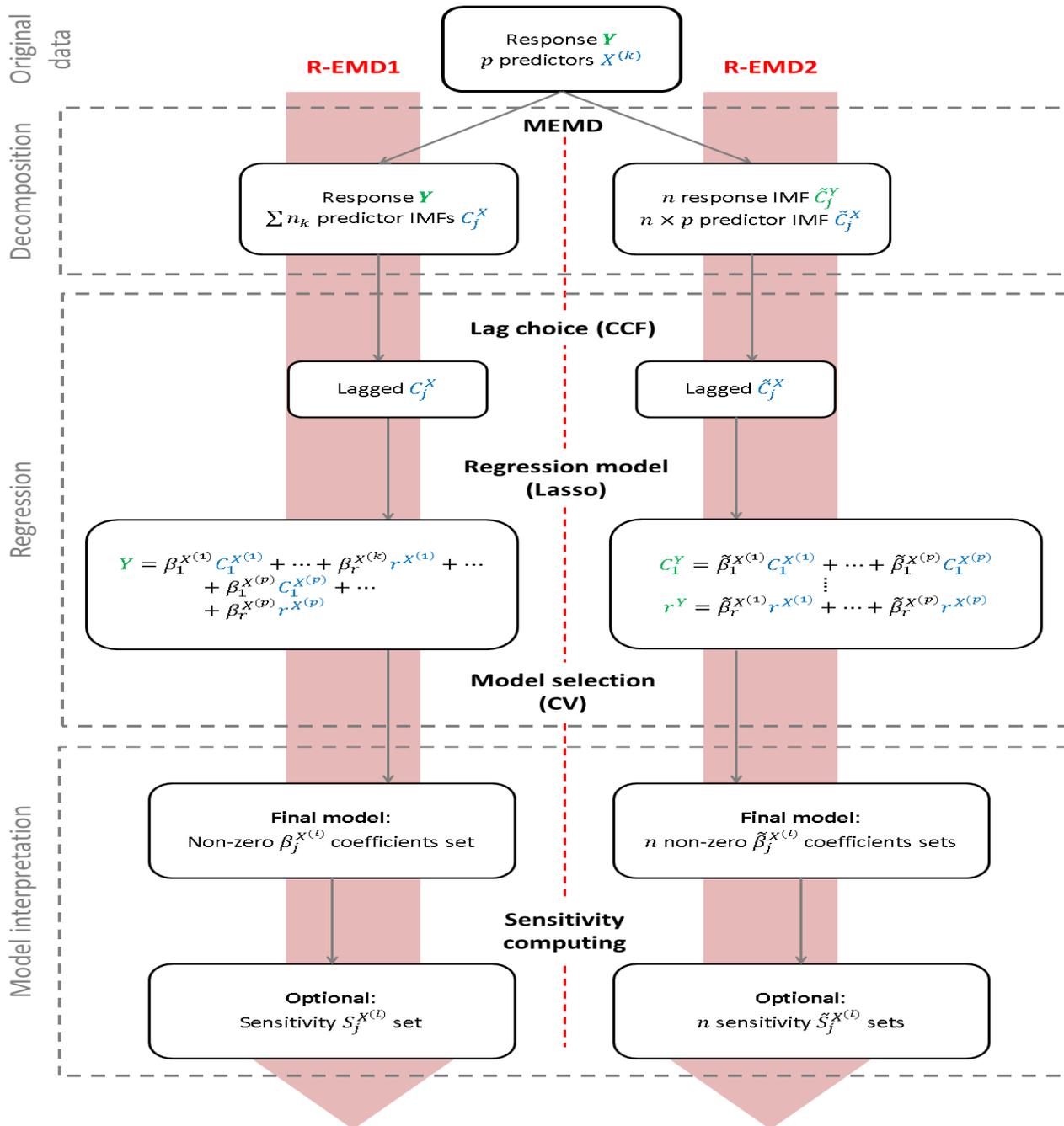
La sensibilité utilisée pour la régression-EMD est définie :

$$S_j = \beta_j A_j$$

Où  $\beta_j$  est le coefficient estimé par Lasso associé à l'IMF  $C_j$  et  $A_j$  est l'amplitude pic-à-pic moyenne d'une même IMF. L'amplitude peut être calculée soit en utilisant le spectre de Hilbert (Huang et collab., 1998) soit en prenant la moyenne de la série  $a_j(t) = s_j(t) - i_j(t)$  où  $s_j(t)$  et  $i_j(t)$  sont respectivement les enveloppes supérieure et inférieure de  $C_j$ .

## Annexe 2 Schéma résumant la méthodologie de régression-EMD

Figure 1 Schéma illustratif de la méthodologie de régression-EMD



# Modélisation multiéchelle de la relation entre une exposition et une issue sanitaire

## AUTEURS

Pierre Masselot<sup>1</sup>  
Fateh Chebana<sup>1</sup>  
Diane Bélanger<sup>1,2</sup>  
André St-Hilaire<sup>1</sup>  
Belkacem Abdous<sup>3</sup>  
Pierre Gosselin<sup>1,2,5</sup>  
Taha B.M.J. Ouarda<sup>1,4</sup>

<sup>1</sup> Institut national de la recherche scientifique, Centre/Eau Terre Environnement, Canada

<sup>2</sup> Centre hospitalier universitaire de Québec, Centre de recherche, Québec, Canada

<sup>3</sup> Université Laval, Département de médecine sociale et préventive, Québec, Canada

<sup>4</sup> Institute Center for Water and Environment (iWATER), Masdar Institute of Science and Technology, Abu-Dhabi, United Arab Emirates

<sup>5</sup> Institut national de santé publique du Québec (INSPQ), Québec, Canada

## RELECTURE DU DOCUMENT

Mariève Doucet  
Bureau d'information et d'études en santé des populations

## MISE EN PAGE

Nabila Haddouche  
Bureau d'information et d'études en santé des populations

*Ce document est disponible intégralement en format électronique (PDF) sur le site Web de l'Institut national de santé publique du Québec au : <http://www.inspq.qc.ca>.*

*Les reproductions à des fins d'étude privée ou de recherche sont autorisées en vertu de l'article 29 de la Loi sur le droit d'auteur. Toute autre utilisation doit faire l'objet d'une autorisation du gouvernement du Québec qui détient les droits exclusifs de propriété intellectuelle sur ce document. Cette autorisation peut être obtenue en formulant une demande au guichet central du Service de la gestion des droits d'auteur des Publications du Québec à l'aide d'un formulaire en ligne accessible à l'adresse suivante :*

*<http://www.droitauteur.gouv.qc.ca/autorisation.php>, ou en écrivant un courriel à : [droit.auteur@cspq.gouv.qc.ca](mailto:droit.auteur@cspq.gouv.qc.ca).*

*Les données contenues dans le document peuvent être citées, à condition d'en mentionner la source.*

Dépôt légal – 2<sup>e</sup> trimestre 2019  
Bibliothèque et Archives nationales du Québec  
ISBN : 978-2-550-84246-0 (PDF)

© Gouvernement du Québec (2019)

N° de publication : 2556