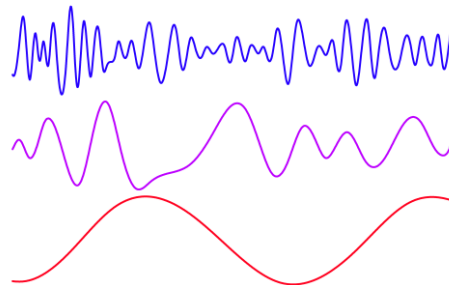


Régression EMD avec application à la relation entre les maladies cardiovasculaires et le climat



© INRS, Centre - Eau Terre Environnement, 2015
Dépôt légal, Bibliothèque nationale du Québec
Dépôt légal, Bibliothèque et archives Canada
ISBN : 978-2-89146-853-4

Auteurs

Pierre Masselot

Institut national de la recherche scientifique, Centre Eau-Terre-Environnement, Québec, Canada

Fateh Chebana

Institut national de la recherche scientifique, Centre Eau-Terre-Environnement, Québec, Canada

Diane Bélanger

Institut national de la recherche scientifique, Centre Eau-Terre-Environnement, Québec, Canada

Centre hospitalier universitaire de Québec, Centre de Recherche, Québec, Canada

André St-Hilaire

Institut national de la recherche scientifique, Centre Eau-Terre-Environnement, Québec, Canada

Belkacem Abdous

Centre hospitalier universitaire de Québec, Centre de Recherche, Québec, Canada

Université Laval, Département de médecine sociale et préventive, Québec, Canada

Pierre Gosselin

Institut national de la recherche scientifique, Centre Eau-Terre-Environnement, Québec, Canada

Centre hospitalier universitaire de Québec, Centre de Recherche, Québec, Canada

Institut national de santé publique du Québec (INSPQ), Québec, Canada

Taha B.M.J. Ouarda

Institut national de la recherche scientifique, Centre Eau-Terre-Environnement, Québec, Canada

Institute center for water and environment (iWATER), Masdar institute of science and technology, Abu Dhabi, United Arab Emirates

Avril 2015

Cette étude a été financée par le Fonds vert, dans le cadre de l'Action 21 du Plan d'action 2006-2012 sur les changements climatiques du gouvernement du Québec.

Crédit images : <http://pixabay.com/>

Table des matières

Liste des sigles et acronymes	7
Remerciements	7
Résumé	8
1. Introduction	9
1.1. Problématique sanitaire	9
1.2. Problématique méthodologique	11
1.3. But, objectifs et plan du rapport	15
2. Données	16
2.1. Données sanitaires	16
2.1.1. Définition d'un cas MCV	16
2.1.2. Décès	17
2.1.3. Hospitalisations	17
2.1.4. Considérations éthiques.....	18
2.2. Données météorologiques.....	18
3. Méthodes statistiques	19
3.1. Connaissances nécessaires.....	20
3.1.1. L'algorithme EMD	20
3.1.1.1. Les bases de l'algorithme EMD.....	20

3.1.1.2.	L'EMD multivarié	23
3.1.2.	La méthode du Lasso pour la régression	25
3.2.	La régression EMD.....	27
3.2.1.	R-EMD1 : modélisation avec météorologie décomposée	28
3.2.2.	R-EMD2 : modélisation avec santé et météorologie décomposées	29
3.2.3.	Détails méthodologiques supplémentaires	31
3.2.3.1.	La validation croisée pour séries temporelles	31
3.2.3.2.	L'utilisation des effets retardés (lags).....	32
3.3.	Application de la régression EMD	33
3.3.1.	Démarche	33
3.3.2.	Implémentation.....	34
4.	Résultats	35
4.1.	Communauté métropolitaine de Montréal	37
4.1.1.	Analyse préliminaire des IMF.....	37
4.1.2.	Régression des décès.....	40
4.1.2.1.	R-EMD1.....	40
4.1.2.2.	R-EMD2.....	42
4.1.3.	Régression des hospitalisations.....	44
4.1.3.1.	R-EMD1.....	44
4.1.3.2.	R-EMD2.....	46

4.2. Communauté métropolitaine de Québec	48
4.2.1. Analyse préliminaire des IMF.....	48
4.2.2. Régression des décès.....	50
4.2.2.1. R-EMD1.....	50
4.2.2.2. R-EMD2.....	51
4.2.3. Régression des hospitalisations.....	53
4.2.3.1. R-EMD1.....	53
4.2.3.2. R-EMD2.....	54
5. Discussion	56
5.1. Résumé des résultats.....	56
5.2. Évaluation de la régression EMD.....	60
6. Conclusion.....	63
Bibliographie.....	65
Tableaux	73
Figures.....	77
Annexes.....	110
A. Description des données.....	110
B. IMF obtenus	119
C. Figures supplémentaires pour la R-EMD.....	131
C.1 R-EMD1.....	131

C.2 R-EMD2.....	134
D. Filtrage des séries	141
E. Résultats R-EMDF1	143
F. Résultats R-EMDF2	147
G. Résultats R-EMDF3	151
H. Lags utilisés.....	154
I. Résultats des modèles classiques	167
I.1. GAM.....	167
I.2. DLNM.....	175

Liste des sigles et acronymes

CCF	Fonction de corrélation croisée (<i>cross-correlation function</i>)
CMM/CMQ	Communauté métropolitaine de Montréal/Québec
CV	Validation croisée (<i>cross validation</i>)
DLM	Modèles à effets retardés distribués (<i>distributed lag models</i>)
DLNM	Modèles non linéaires à effets retardés distribués (<i>distributed lag nonlinear models</i>)
EEMD	Décomposition modale empirique d'ensemble (<i>ensemble empirical mode decomposition</i>)
EMD	Décomposition modale empirique (<i>empirical mode decomposition</i>)
GAM	Modèles additifs généralisés (<i>generalized additive models</i>)
GLM	Modèles linéaires généralisés (<i>generalized linear models</i>)
HHT	Transformée de Hilbert-Huang (<i>Hilbert-Huang transform</i>)
IMF	Fonction modale intrinsèque (<i>intrinsic mode function</i>)
LARS	Régression par moindre angle (<i>least angle regression</i>)
Lasso	Opérateur de sélection et de rétrécissement par moindre valeur absolue (<i>least absolute shrinkage and selection operator</i>)
MCV	Maladies cardiovasculaires
MEMD	Décomposition modale empirique multivariée (<i>multivariate empirical mode decomposition</i>)
NA-MEMD	Décomposition modale empirique multivariée assistée de bruit (<i>noise assisted multivariate empirical mode decomposition</i>)
R-EMD1	Régression EMD où seule la météorologie est décomposée
R-EMD2	Régression EMD où la santé et la météorologie sont décomposées
R-EMDF1	Régression EMD où la météorologie est filtrée
R-EMDF2	Régression EMD modélisant la santé filtrée par les IMF météorologiques
R-EMDF3	Régression EMD où la santé et la météorologie sont filtrées

Remerciements

Les auteurs remercient le Fonds Vert du Québec pour le financement de cette étude et l'Institut national de santé publique du Québec pour l'accès aux données sanitaires. Les auteurs souhaitent également remercier Jean-Xavier Giroux (INRS-ETE) pour son aide lors de la constitution de la banque de données et Yohann Chiu (INRS-ETE) pour ses commentaires judicieux tout au long du projet.

Résumé

Ce rapport présente une méthodologie statistique quantifiant à différentes échelles temporelles la relation entre la mortalité et morbidité pour cause de MCV et la météorologie. Cette méthodologie est nommée « régression EMD » et consiste en une décomposition préliminaire des séries de données avant d'être intégrées à un modèle de régression.

L'application de la méthodologie de régression EMD aux données du Québec a permis d'obtenir les principaux résultats suivants :

- L'augmentation de la mortalité et de la morbidité en hiver sont dues aux faibles températures de la saison.
- Le rôle des fortes chaleurs estivales et des tempêtes de neige hivernales dans une augmentation de mortalité et de la morbidité à court terme à été mis en évidence.
- L'humidité joue un rôle important dans l'accentuation des effets dus à la température, notamment l'humidité automnale sur la mortalité et morbidité hivernales.

Une comparaison avec les modèles statistiques les plus utilisés dans la littérature épidémiologique a montré que la régression EMD a de bonnes performances explicatives et prédictives.

1. Introduction

Cette section présente la problématique ayant motivé cette étude, aussi bien du point de vue sanitaire que méthodologique. Elle présente également les objectifs de l'étude ainsi que le plan du rapport.

1.1. Problématique sanitaire

L'adaptation aux changements climatiques sera un enjeu majeur des décennies à venir. Parmi leurs conséquences néfastes, une hausse de la mortalité et de la morbidité pour cause de maladies cardiovasculaires (MCV) lors d'extrêmes météorologiques a déjà été observée (p. ex. Braga et al. 2002; Lan Chang et al. 2004; Houck et al. 2005; Patz et al. 2005; Knowlton et al. 2009; Nitschke et al. 2011). Au Québec, l'augmentation des taux d'hospitalisation pour cardiopathies ischémiques durant les mois très chauds ou très froids de 1989 à 2006 en est un exemple, de même que la relation entre ces taux et les conditions météorologiques jusqu'à deux semaines avant le jour d'admission hospitalière (Bayentin et al. 2010). De fait, les MCV mettent à rude épreuve la réactivité vasculaire et la stabilité cardiovasculaire (pression sanguine) (Matthies et al. 2008). Lors d'une canicule, la chaleur ajoute donc un stress à des organismes déjà stressés par ces maladies et peut précipiter ainsi les personnes qui en sont atteintes vers le décès. Sans compter que toutes les classes de médicaments thérapeutiques utilisées pour traiter les MCV, tout particulièrement les diurétiques, sont susceptibles d'aggraver le syndrome d'épuisement-déshydratation et de contribuer au coup de chaleur (Blachère et Perreault 2012). Enfin, le réchauffement climatique ne s'effectue pas nécessairement de façon linéaire (MacCracken et al. 2001). Au Québec, il y aura donc encore des vagues de froid intense, bien que leur durée ait légèrement diminué de 1951 à 2010 (Donat et al. 2013). Or, le froid intense crée aussi un stress

sur l'organisme, d'autant plus ressenti chez une personne atteinte de certains problèmes de santé comme les MCV (p. ex. Huynen et al. 2001; Keatinge 2002).

Au Canada, les maladies cardiovasculaires (MCV) constituent la première cause de mortalité; elles représentaient 32 % de l'ensemble des décès au pays en 2004 (Wielgosz et al. 2009). La prévalence de MCV y est d'ailleurs très élevée, avec 1,3 million de Canadiens ayant déclaré en 2009 avoir reçu un diagnostic de MCV de la part d'un professionnel de la santé. Le Québec ne fait pas exception. Certes, la mortalité par MCV y a nettement diminué entre 1995-1999 et 2005-2008 (Ministère de la santé et des services sociaux 2011), en raison vraisemblablement de traitements plus efficaces. Malgré cela, en 2007, plus du quart des décès à l'échelle de la province étaient attribuables aux MCV, qui occupaient ainsi le 2^e rang des causes de mortalité (Institut de la statistique du Québec [ISQ] 2009). De plus, comme dans le reste du Canada, la prévalence des maladies cardiaques et de certains de leurs facteurs de risque (p. ex., diabète, surpoids/obésité) y est en augmentation (Lee et al. 2009), notamment en raison du vieillissement de la population, et le profil clinique de la maladie est plus difficile à traiter à l'âge avancé (Tu et al. 2009).

Ainsi, les répercussions des MCV sur la vie des Canadiens et de leurs familles (ex. besoin d'aide pour accomplir les activités de la vie quotidienne) sont loin d'être négligeables, tout comme leurs répercussions sur le système de santé et l'économie (Wielgosz et al. 2009). À ce sujet, en 2000, les coûts engendrés par les MCV se chiffraient à 22,2 milliards de dollars (G\$) et occupaient le deuxième rang à ce chapitre au pays (Wielgosz et al. 2009). De ce total, 7,6 G\$ représentaient des coûts directs, soit 4,0 G\$ en soins hospitaliers, 2,1 G\$ en médicaments et 1,5 G\$ en soins médicaux, et 14,6 G\$ des coûts indirects, c'est-à-dire 9,3 G\$ en pertes de productivité, 4,2 G\$ en incapacités de longue durée et 1,2 G\$ en incapacités de courte durée. Les coûts de 2015 ne sont pas disponibles, mais tout porte à croire qu'ils sont encore plus élevés, comme le suggère la

hausse considérable du nombre d'ordonnances (en millions, M) délivrées pour le traitement de MCV au fil du temps (2003 : 52,5 M, 2005 : 57,4 M, 2009 : 65,7 M) et le fait que la prise en charge des MCV était le motif de 34,6 M de consultations (soit 10 % de toutes les consultations) chez le médecin et de 17 % de tous les cas d'hospitalisation.

En bref, au Québec, il est donc possible que la tendance à la baisse de la mortalité par MCV s'inverse dans un contexte de changements climatiques. D'autant que la prévalence des MCV, tout comme celle de l'obésité et du diabète (deux facteurs de risque de MCV) (Wielgosz et al. 2009), s'accroît. Il en est de même de la part dans la population des 65 ans et plus qui ne cesse d'augmenter (passant de 12 % en 2001, à 17 % en 2014 et à 23 % en 2026, Institut de la statistique du Québec [ISQ] 2000), lesquels s'avèrent plus vulnérables au réchauffement climatique que celui des 15 à 65 ans (Doyon et al. 2008). Tant sur le plan individuel et familial que sur le plan sociétal, il est donc des plus importants de mieux comprendre le lien existant entre la météorologie et la mortalité ou la morbidité pour cause de MCV, notamment en améliorant la méthodologie utilisée généralement pour ce faire.

1.2. Problématique méthodologique

Afin de bien comprendre le lien entre la santé et la météorologie, de plus en plus d'études épidémiologiques et en santé publique ont vu le jour à l'international (p. ex. Curriero et al. 2002; Bassil et al. 2009; Gasparrini et Armstrong 2011; Morabito et al. 2012) et au Québec (p. ex. Bayentin et al. 2010; Bustinza et al. 2013; Vutcovici et al. 2013). L'analyse des relations entre la santé et la météorologie doit faire face à la difficulté de travailler avec des séries chronologiques qui sont des données présentant une structure de dépendance temporelle qui peut être complexe. Dans un modèle de régression, cette structure peut avoir pour effet de faire diminuer la performance des modèles de régression utilisés et même d'induire des conclusions peu réalistes.

De fait, la comparaison entre diverses études épidémiologiques récentes met en évidence des façons très différentes de tenir compte des données météorologiques et de définir des périodes de latence, tant pour la mortalité que la morbidité. Les conclusions varient aussi de manière importante entre les multiples recherches étudiant des phénomènes a priori comparables, et il n'existe encore aucun guide méthodologique sur les meilleures façons de faire, notamment pour la mesure et le traitement des variables météorologiques (p. ex. Huang et al. 2011; Ye et al. 2012).

La structure temporelle des données sanitaires et météorologiques induit deux principaux problèmes préjudiciables à tout modèle linéaire, soit l'interdépendance des variables explicatives et la non-stationnarité. Par exemple, les différentes composantes de la météorologie (p. ex. température, précipitations, pression) sont intimement liées les unes aux autres. Dans un modèle de régression, cette interdépendance est problématique, car elle cause de la colinéarité qui fait augmenter la variabilité des estimations et les rend donc moins fiables (voir p. ex. Cornillon et Matzner-Løber 2007 chapitre 8). De plus, les séries sanitaires et météorologiques présentent des cycles complexes avec des non-stationnarités telles que des saisonnalités ou des tendances pluriannuelles. Ceci a également pour effet d'augmenter la variabilité des estimations et donc d'augmenter drastiquement la probabilité de trouver à tort une relation significative entre une variable météorologique et la variable sanitaire d'intérêt (Granger et Newbold 1974).

La colinéarité entre les variables est généralement traitée en utilisant des méthodes statistiques dites « biaisées » telles que la régression Ridge (Hoerl et Kennard 1970), le Lasso (*least absolute shrinkage and selection operator*, Tibshirani 1996) ou encore la régression PLS (*partial least squares*, Wold et al. 2001). Cependant, ces méthodes traitent du cas de dépendance linéaire entre

les variables explicatives ce qui n'est pas toujours le cas avec les séries météorologiques (p. ex. Stuart et Isaac 1994).

Le cas d'un modèle de régression utilisant des variables (réponses et explicatives) non stationnaires est connu sous le nom de « régression fallacieuse » (Phillips 1986). L'approche communément utilisée pour éviter ce problème est de stationnariser les séries, soit en retranchant les tendances et saisonnalités, soit en différenciant les séries. Cependant, ceci revient à se concentrer sur les motifs *court terme* des phénomènes modélisés. Or les motifs *long terme* représentés par les saisonnalités et les tendances ont également leur importance et se doivent d'être considérés dans les modèles.

Les deux problèmes d'interdépendance et de non-stationnarité peuvent être gérés en utilisant une décomposition spectrale des séries de données. Une telle approche permet de séparer les motifs *long terme* et *court terme* les uns des autres afin que leurs effets ne soient pas confondus. À ces fins, les méthodes traditionnelles sont les séries de Fourier (p. ex. Harvey 1978) et les ondelettes (p. ex. Daubechies 1992). Cependant, elles présentent certaines lacunes, notamment la subjectivité et le manque de flexibilité. En effet, les séries de Fourier décomposent une série en une somme de sinus et de cosinus de différentes fréquences, ce qui ne permet pas de prendre en compte l'éventuelle non-stationnarité (p. ex. tendance monotone) de la série. Les ondelettes peuvent décomposer les signaux non stationnaires, mais imposent un choix subjectif d'une « ondelette mère » qui définit la forme des composantes finales.

Afin de traiter cette problématique tout en étant flexible et objectif, il est proposé ici d'utiliser la décomposition modale empirique (EMD pour *empirical mode decomposition*, Huang et al. 1998) pour décomposer les séries de données. La méthode EMD permet de décomposer les séries en un

ensemble de composantes oscillantes appelées fonctions modales intrinsèques (*intrinsic mode functions*, IMF) qui représentent séparément les différents niveaux de variations de la série décomposée. La méthode EMD a été très utilisée depuis son introduction pour étudier toutes sortes de signaux, que ce soit en océanographie (Huang et al. 1999), hydrologie (Lee et Ouarda 2010), sismologie (Loh et al. 2001) ou encore en économétrie (Zhang et al. 2008). De plus, la méthode elle-même a fait l'objet de beaucoup de recherches (p. ex. Flandrin et al. 2004b; Haiyong et Qiang 2006; Rato et al. 2008) et d'améliorations (p. ex. Junsheng et al. 2006; Rilling et al. 2007; Wu et al. 2009) lui permettant d'être plus complète et adaptée pour les différents besoins existants.

L'algorithme EMD est ici utilisé dans la régression en remplaçant les variables d'origine par des IMF issus de la décomposition. Dans ce cadre, l'utilisation des IMF peut concerner la variable réponse (Xie et al. 2014), une seule ou plusieurs variables explicatives (Yang et al. 2011a,b), ou bien les deux à la fois (Hu et Si 2013). Un modèle similaire utilisant, au lieu des IMF, des équivalents construits par ondelettes peut se retrouver dans la littérature hydrologique (p. ex. Kişi 2009; Sahay et Sehgal 2013). Le type de modèle utilisé ne doit pas être confondu avec la régression non paramétrique par ondelettes qui travaille sur la fonction de régression et non sur les variables elles-mêmes (p. ex. Donoho et Johnstone 1994; Antoniadis et al. 2001). Les IMF représentant chacun un motif d'oscillation propre, un modèle de régression avec des IMF permet de déterminer plus précisément à quelle échelle exacte se trouve l'influence de chaque variable météorologique. Ainsi, l'effet d'une variable météorologique peut être dû entièrement à la tendance *long terme*, mais pas à des variations *court terme*. Ceci permet notamment d'inclure les tendances dans le modèle sans qu'elles occultent d'autres effets de plus *court terme*. Sur un plan statistique, les IMF sont quasi orthogonaux entre eux (donc non corrélés linéairement). De plus,

leur périodicité permet de conserver des caractéristiques statistiques quasi constantes au cours de la période étudiée. Ces deux propriétés permettent, en pratique, d'atténuer significativement les problèmes de colinéarité et de régression fallacieuse. Enfin, la définition d'un IMF est moins restrictive que celle des composantes de Fourier et ondelettes. En effet, les IMF s'adaptent à la série de départ afin de capter ses irrégularités et ainsi de conserver entièrement l'information de la série avec un nombre limité de composantes.

1.3. But, objectifs et plan du rapport

Le but ultime de cette étude est de soutenir la surveillance en santé publique dans un contexte de changements climatiques et de mieux évaluer les effets de la météorologie sur le comportement temporel des MCV. Ses objectifs sont de présenter une méthodologie de régression EMD et de l'appliquer à l'exploration de la relation (1) entre les décès pour cause principale de MCV et la météorologie au Québec de 1981 à 2011 et (2) entre les hospitalisations pour cause principale ou secondaire (morbidité dans ce rapport) de MCV et la météorologie au Québec de 1996 à 2007. La méthodologie vise à exploiter les points forts de l'algorithme EMD pour inclusion dans un modèle explicatif.

Ainsi, après avoir introduit les données disponibles en section 2, la section 3 présente les étapes essentielles ainsi que les détails de la méthodologie de régression EMD. La section 4 rapporte ensuite les résultats obtenus lors de l'application de cette méthodologie au problème de la relation entre les MCV et la météorologie. Les possibilités d'interprétations offertes par cette méthodologie, ainsi que ses points forts et faibles, seront notamment mis en évidence dans la discussion de la section 5, pour ensuite apporter une conclusion dans la section 6. Des annexes sont également jointes à ce rapport pour compléter les figures et tableaux présentés.

2. Données

Le projet utilise deux types de données : les données sanitaires et les données météorologiques. Les données sanitaires regroupent les décès et les hospitalisations pour cause de MCV et représentent les variables à expliquer. Les données météorologiques sont les variables explicatives des modèles et correspondent aux valeurs quotidiennes brutes des variables météorologiques.

Dans ce projet, les données touchent les communautés métropolitaines de Montréal (CMM) et de Québec (CMQ), toutes deux localisées sur la carte de la Figure 1. La communauté métropolitaine représente un bassin dense de population, ce qui permet d'obtenir suffisamment de cas quotidiens de mortalité par MCV. De plus, cela permet de relier les séries sanitaires à des séries météorologiques ayant du sens, car ces bassins sont suffisamment restreints pour être considérés comme homogènes en matière de météorologie.

Les séries de données décrites dans la présente section peuvent être visualisées dans l'annexe A du rapport. Cette annexe comprend de plus quelques statistiques descriptives de base de ces données.

2.1. Données sanitaires

Les données sanitaires correspondent aux cas de décès et hospitalisations pour cause de MCV.

2.1.1. Définition d'un cas MCV

Les définitions de cas de MCV ont été adoptées par les experts en surveillance des provinces et territoires du Groupe de travail sur les MCV coordonné par l'Agence de santé publique du Canada (ASPC). Dans cette étude, elles regroupent les maladies coronariennes, l'insuffisance cardiaque, les maladies vasculaires cérébrales et l'ischémie cérébrale transitoire. Ces définitions

ont été validées et publiées (Tu et al. 2010; Schultz et al. 2013; Tu et al. 2013); elles sont présentées au Tableau 1, ainsi que leurs codes selon la classification internationale des maladies (CIM). Le passage de la version 9 à la version 10 de la CIM a été effectué en janvier 2000 pour les décès et en avril 2006 pour les hospitalisations. Ce passage ne nécessite pas l'application d'un facteur de correspondance, car la différence est négligeable dans le cas des MCV et ne provoque donc pas de saut de moyenne (Statistics Canada 2005).

2.1.2. Décès

Les données de décès ont été générées par l'INSPQ, dans le cadre de l'entente du Plan ministériel de surveillance multithématique du Québec. Ils correspondent à tous les Québécois décédés ayant comme cause principale une MCV du 1^{er} janvier 1981 au 31 décembre 2011. La date d'un cas correspond à la date du décès.

2.1.3. Hospitalisations

Les hospitalisations (ou MED-ECHO) sont issues de la banque de données « recherche » constituée pour l'INSPQ aux fins de surveillance des maladies cardiovasculaires, constitution entérinée par la Commission d'accès à l'information (CAI) du Québec. Elles regroupent tous les cas hospitalisés pour cause principale ou secondaire de MCV, tels que définis précédemment, de janvier 1996 à mars 2007 au Québec, sauf quelques exceptions. Ces exceptions, déterminées sur la base de la littérature (Pelletier 1999) et recommandées par l'INSPQ, sont : (1) les hospitalisations en soins d'un jour (chirurgie d'un jour depuis 1995) et en soins de longue durée, (2) les hospitalisations en soins de longue durée dans des unités de courte durée, (3) les hospitalisations de type hôpital à domicile, (4) les hospitalisations de nouveau-nés en bonne santé, (5) les centres hospitaliers dont la vocation première est la psychiatrie, la réadaptation ou l'hébergement et les soins de longue durée, et (6) les non-résidents québécois.

La date d'un cas correspond à la date d'admission du séjour hospitalier. Dans cette étude, la règle des 30 jours a été appliquée. Plus spécifiquement, pour chaque individu et une même cause médicale (ex., maladie coronarienne), une hospitalisation est considérée comme une nouvelle hospitalisation seulement s'il s'est écoulé plus de 30 jours entre la date de départ de l'hospitalisation précédente et la date d'admission de la présente hospitalisation (Ross et al. 2009; Ricciardi et al. 2012).

2.1.4. Considérations éthiques

Les données sanitaires brutes journalières ont été fournies par l'Institut national de santé publique du Québec (INSPQ). Le transfert de données d'hospitalisation de l'INSPQ vers l'INRS-ETE a été entériné par la Commission d'accès à l'information du Québec (N/Réf. : 11 09 45 (08 19 12 07 00 76)), alors que le transfert de données de mortalité est couvert par l'entente du Plan ministériel de surveillance multithématique (PMSM), sous l'aspect du développement d'indicateurs de surveillance applicables au domaine de la santé et de l'environnement (avis favorable du Comité d'éthique de la santé publique dans le cadre de son examen du PCS et du PMSM en 2010).

2.2. Données météorologiques

Les données météorologiques actuellement disponibles sont issues du portail de téléchargement DAI d'Environnement Canada (*Données Accès et Intégration*, cf. <http://loki.qc.ec.gc.ca/DAI/>, Tableau 2). Les données utilisées sont la température (minimale T_{min} , maximale T_{max} , moyenne T_{moy}), les précipitations totales ($Prec$), l'épaisseur de neige au sol ($Snow$), la pression atmosphérique et l'humidité relative. Pour ces deux dernières, les relevés étant horaires, des données journalières en sont extraites pour cette étude afin de disposer d'une série de pression quotidienne moyenne (P_{moy}) et des séries d'humidité quotidienne moyenne (H_{moy}), minimale (H_{min}) et maximale (H_{max}). Les données de température (T_{moy} , T_{min} et T_{max}), de

précipitations totales et d'épaisseur de neige au sol sont quant à elles journalières. Toutes ces séries couvrent la période 1981 à 2011. Toutes les séries météorologiques utilisées dans cette étude, ainsi que leur abréviation, sont recensées dans le Tableau 2.

Les données proviennent à la base d'un grand nombre de stations de mesures réparties sur le territoire. Afin d'obtenir une seule série pour un territoire, la moyenne spatiale est calculée sur les stations faisant partie du territoire. En effet, pour les grandes municipalités comme la CMM et la CMQ, Giroux et al. (2013) ont conclu que l'utilisation d'une moyenne spatiale donnait des résultats aussi probants qu'en interpolant par krigeage au Québec.

La pression au niveau de la mer sera utilisée au lieu de la pression à la station, car il s'agit de la pression à la station ajustée pour tenir compte de l'altitude¹. Puisque les stations se situent à des altitudes différentes, les mesures de pression sont normalisées pour obtenir les pressions équivalentes à un point de référence (ici, le niveau de la mer). Ces nouvelles valeurs normalisées facilitent ainsi les comparaisons entre pressions.

3. Méthodes statistiques

Cette section décrit la méthodologie statistique développée et appliquée dans cette étude. Dans un premier temps, les méthodes statistiques utilisées, c.-à-d. l'algorithme EMD et la méthode du Lasso pour la régression, sont introduites. Puis la méthodologie de régression EMD pour explorer la relation entre les décès ou hospitalisations pour cause de MCV (regroupés sous le terme MCV dans cette section) et la météorologie est détaillée. Enfin, son application sera discutée.

¹ La formule internationale du nivellement barométrique : http://fr.wikipedia.org/wiki/Formule_du_nivellement_barom%C3%A9trique#Formule_internationale_du_nivellement_barom.C3.A9trique.

3.1. Connaissances nécessaires

Cette section présente l'algorithme EMD et la méthode du Lasso pour la régression qui sont les deux composantes de base de la méthodologie de régression EMD.

3.1.1. L'algorithme EMD

Cette section introduit d'abord les bases de l'EMD puis son extension multivariée qu'est la MEMD (*multivariate EMD*, Rehman et Mandic 2010).

3.1.1.1. Les bases de l'algorithme EMD

La méthode EMD a été développée dans le but d'identifier les différents modes d'oscillations d'une série temporelle. Ces modes d'oscillations permettent ensuite d'obtenir un spectre de fréquences appelé *transformée de Hilbert-Huang (HHT)*, Huang et al. 1998). Le spectre HHT s'interprète similairement à un spectre d'ondelettes mais est plus flexible pour des données empiriques. Les modes d'oscillations recherchés sont représentés par des IMF qui sont des fonctions dont la définition répond à deux contraintes : a) tous les maxima locaux et minima locaux doivent être positifs et négatifs respectivement, et b) la moyenne entre les enveloppes supérieures et inférieures, formées à partir des maxima et minima locaux, doit être nulle partout. Selon sa définition, un IMF est donc une fonction oscillant symétriquement autour de zéro, ce qui lui assure une certaine stabilité et facilité d'interprétation. Cette définition permet de plus aux IMF d'accepter des modulations en amplitude et en fréquence pour s'adapter plus efficacement aux données.

Dans la décomposition EMD, les IMF sont obtenus successivement en commençant par les hautes fréquences. Le premier IMF est obtenu par le processus itératif suivant, à partir de la série

X :

- 1) Ajuster les enveloppes de X sur les maxima et minima par spline naturelle cubique;
- 2) Retrancher la moyenne m_1 des enveloppes supérieures et inférieures à la série de base pour obtenir le prototype IMF h_1 ;
- 3) Recommencer les étapes 1) et 2) sur h_1 pour obtenir h_2 , et ainsi de suite jusqu'à obtenir un prototype h_k satisfaisant la définition des IMF.

Le prototype h_k devient donc le premier IMF C_1 . L'IMF suivant C_2 s'obtient ensuite en appliquant le même processus au reste $r_1 = X - C_1$ et ainsi de suite jusqu'à obtenir un reste r_n contenant au plus un extremum. Ce reste r_n est assimilé à la tendance de la série.

Le seul paramètre important à choisir pour l'utilisateur est le critère d'arrêt du processus itératif d'obtention de chaque IMF (le 3e point de l'algorithme ci-dessus) qui a une influence déterminante sur la forme des IMF obtenus (Wu et Huang 2010). Il est choisi ici d'arrêter le processus lorsque la moyenne m_k des enveloppes est suffisamment proche de zéro partout, rejoignant Rilling et al. (2003) qui conseillent un seuil de 5% de l'amplitude du signal. Ce critère permet d'obtenir des composantes respectant les contraintes des IMF en forçant les composantes à osciller symétriquement autour de zéro, ce qui implique que tous les maxima soient positifs et les minima négatifs (Yang et Yang 2009).

Bien que l'algorithme EMD a fait l'objet de diverses modifications telles que l'*orthogonal EMD* (Huang et al. 2008) et le *statistical EMD* (SEMD, Kim et al. 2012), seule l'extension EEMD (*ensemble EMD*, Wu et Huang 2009) est ici utilisée. L'EEMD vise à gérer le problème du mélange de modes, qui apparaît souvent lors de la décomposition de séries de données empiriques. Il y a mélange de modes lorsqu'un IMF C_j contient des fréquences très différentes

en son sein, ou lorsqu'une fréquence d'oscillation est présente dans plusieurs C_j . Dans un tel cas les C_j sont plus difficilement interprétables. Le mélange de modes survient généralement lorsque certains modes d'oscillations sont intermittents dans la série de base (Huang et al. 1999), ce qui est notamment le cas des hautes fréquences de hauteur de neige, présentes dans la série uniquement l'hiver.

L'EEMD apporte une solution au problème du mélange de modes en ajoutant un bruit blanc z_e à la série afin de peupler le spectre de fréquences de la série X avant de la décomposer pour obtenir ses IMF ($C_1^{(e)}, \dots, C_n^{(e)}$). Comme l'ajout de bruit blanc change le résultat de la décomposition, ce processus est répété un grand nombre N_e de fois avec un bruit blanc z_e différent à chaque fois.

Les composantes finales C_j^* (appelées « pseudo-IMF ») sont la moyenne de toutes les

composantes de même ordre, c.-à-d.
$$C_j^* = \frac{1}{N_e} \sum_{e=1}^{N_e} C_j^e.$$

L'utilisation d'EEMD nécessite de choisir les paramètres supplémentaires que sont le nombre d'ensembles N_e et l'écart-type σ_z du bruit blanc ajouté. Le choix de N_e est assez simple dans la mesure où il doit théoriquement être le plus grand possible, mais en pratique $N_e \geq 100$ donne de bons résultats (Zhang et al. 2010). Le choix de σ_z est en revanche plus critique, car une valeur trop basse rend l'EEMD inefficace alors qu'une valeur trop élevée résulte en une trop grande erreur de décomposition. Les valeurs recommandées pour σ_z vont généralement de 1% (p. ex. Zhang et al. 2010) à 20% de l'écart-type de X (p. ex. Wu et Huang 2009; Colominas et al. 2013). Sans consensus sur la valeur optimale de σ_z , différentes valeurs seront testées dans cette

étude pour retenir celle qui résulte en le meilleur compromis entre résolution du mélange de modes et faible erreur de décomposition.

L'inconvénient majeur de l'EEMD est le fait que sommer des IMF ne résulte pas forcément en une fonction respectant les deux contraintes des IMF (présentées au premier paragraphe de la section 3.1.1.1). Les composantes C_j^* obtenues par EEMD ne sont donc pas forcément strictement des IMF (d'où l'utilisation du terme « pseudo-IMF »). Il convient donc d'effectuer un post-traitement qui consiste en une application supplémentaire d'EMD sur certaines composantes C_j^* s'éloignant trop d'IMF. Le post traitement consiste à analyser la corrélation entre les C_j^* adjacents afin de sommer ceux qui sont trop corrélés. De plus, EMD est appliqué aux C_j^* présentant trop de minima positifs ou maxima négatifs. La procédure de post-traitement est détaillée par Wu et Huang (2009). Les tentatives d'automatisation de ce processus de post-traitement (Peel et al. 2011; Wu et al. 2012) s'étant malheureusement révélées insatisfaisantes, dans cette étude le post-traitement sera réalisé « à la main » en analysant les C_j^* et leurs corrélations. Ceci aura cependant très peu d'influence sur les résultats.

3.1.1.2. L'EMD multivarié

L'algorithme EMD permet de décomposer des séries univariées. Or, un phénomène comme la météorologie est en fait constitué d'un ensemble de variables liées les unes aux autres. L'extension multivariée de l'algorithme EMD (MEMD, Rehman et Mandic 2010) permet de décomposer un ensemble de séries comme une seule série multivariée. Cette approche permet d'obtenir des IMF multivariés \tilde{C}_j ayant donc une interprétation globale plutôt qu'intrinsèque à chaque série. Son principal avantage réside dans l'alignement des modes. En effet, en utilisant

MEMD, les IMF \tilde{C}_j obtenus contiennent des fréquences similaires sur toutes les séries, ce qui n'est pas garanti lorsque l'on décompose les séries séparément (Rehman et al. 2013).

L'algorithme MEMD est une généralisation du cas bidimensionnel développé par Rilling et al. (2007). Le but du cas bidimensionnel est de déterminer les modes de rotations intrinsèques d'un signal. Ces rotations sont obtenues en ajustant un tube autour de la série, qui correspond à la généralisation des enveloppes traditionnellement ajustées dans l'algorithme EMD. Le centre de ce tube fait office de moyenne locale du signal bidimensionnel, pour ensuite la retrancher et obtenir les \tilde{C}_j successifs (qui sont maintenant des modes de rotation). Afin de déterminer son enveloppe, le signal bidimensionnel est projeté sur un plan afin de retrouver un signal unidimensionnel sur lequel il est possible de construire des enveloppes tel qu'expliqué dans la section 3.1.1.1. En faisant tourner ce plan, des enveloppes pour différentes projections du signal bidimensionnel sont obtenues afin de construire un tube autour de celui-ci. Finalement, ce principe de « plan qui tourne » pour obtenir l'enveloppe d'un signal bidimensionnel se généralise ensuite à un plus grand nombre de dimensions (Rehman et Mandic 2010).

Comme pour l'EMD classique, le mélange de modes peut subvenir dans les résultats, ce qui a motivé l'élaboration de la MEMD assistée de bruit (noise assisted MEMD, NA-MEMD). Le NA-MEMD est une méthode équivalente à l'EEMD dans le cas multidimensionnel (Rehman et Mandic 2011). Ici encore, le but est de s'aider d'un bruit blanc pour exacerber les fréquences présentes dans la série de base. Cependant, au lieu d'être sommé à la série, le bruit blanc est rajouté comme variable supplémentaire à la série multivariée. Ceci permet d'éviter le mélange de modes aussi bien que l'EEMD dans le cas unidimensionnel. Comme le bruit blanc n'est pas ajouté directement à la série, l'algorithme NA-MEMD a également la vertu de ne pas créer de

bruit résiduel dans la décomposition, ce qui était le principal problème de l'EEMD, et donc de redevenir une décomposition sans perte d'information (dite « complète »). Les composantes obtenues sont donc bien des IMF (alors que l'on parle de « pseudo-IMF » pour l'EEMD). Cet algorithme permet également un gain de temps de calcul conséquent par rapport l'EEMD (Rehman et al. 2013).

3.1.2. La méthode du Lasso pour la régression

L'utilisation d'IMF météorologiques comme variables explicatives, au lieu des séries météorologiques en tant que telles, apporte deux difficultés aux modèles de régression, soit (i) la décomposition peut donner un grand nombre d'IMF, qui augmente très rapidement avec le nombre de variables (p. ex. 10 IMF par série résulte en presque 100 IMF au total sur les données utilisées ici), et (ii) des colinéarités persistent entre des IMF issus de séries différentes, mais de fréquences similaires (p. ex. les oscillations de période un an pour la température et la hauteur de neige). Afin de gérer ces deux difficultés, la régression EMD développée dans ce rapport utilise la méthode du Lasso (Tibshirani 1996).

Le Lasso a été développé pour la régression à très grande dimension avec le double objectif de réduire la variance induite par la colinéarité entre les variables explicatives et d'opérer une sélection des variables directement lors de l'estimation des coefficients. Elle repose sur l'hypothèse de parcimonie qui veut que peu de coefficients de régression β_j sont véritablement différents de zéro (c.-à-d., peu des variables explicatives sont réellement influentes). Le Lasso permet donc à la fois de sélectionner un sous-ensemble d'IMF significatifs et de gérer les colinéarités persistantes entre les IMF de même fréquence.

Le principe du Lasso est d'estimer les coefficients par moindres carrés (comme pour une régression linéaire classique, p. ex. Cornillon et Matzner-Løber 2007) en ajoutant la contrainte

$\sum_{j=1}^p |\beta_j| \leq t$. Cette contrainte ajoute un biais à l'estimation, ce qui a pour effet de faire diminuer sa

variance (voir Friedman et al. 2009 pour une discussion sur le compromis biais-variance). De

plus, le fait de limiter l'estimation au domaine défini par la contrainte permet de forcer un certain

nombre de coefficients à 0, et donc d'opérer une sélection des variables (Tibshirani 1996).

Le Lasso est plus connu sous sa forme lagrangienne, c.-à-d.

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (1)$$

ce qui signifie que les β_j sont estimés en minimisant les moindres carrés soumis à une

pénalisation sur la valeur absolue des β_j . Le paramètre λ contrôle la sévérité de la pénalisation :

lorsqu'il augmente, la pénalisation devient plus forte et le modèle obtenu est très parcimonieux.

À l'inverse, lorsque $\lambda = 0$, il n'y a plus de contrainte et l'estimation est celle des moindres carrés

ordinaires. En pratique, λ est estimé par validation croisée (CV pour *cross validation*) en prenant

la valeur correspondant au modèle le plus parcimonieux ayant un score de CV inférieur au

minimum global plus son écart-type (p. ex. Friedman et al. 2009). La section 3.2.3.1 détaille

l'utilisation de la CV dans ce projet.

Le Lasso nécessite que la variable réponse soit gaussienne pour que le critère des moindres carrés

de l'équation (1) soit optimal. Or, comme la variable réponse est ici un nombre de cas

d'hospitalisations ou de décès quotidiens, il s'agit d'une variable de comptage mieux modélisée

par une loi de Poisson. Cependant, si la moyenne de cette loi de Poisson est suffisamment élevée,

elle converge théoriquement vers une gaussienne (p. ex. McCullagh et Nelder 1989, chapitre 6). Dans ce cas, la variable réponse peut ainsi être considérée gaussienne. Dans les autres cas, Park et Hastie (2007) ont étendu la méthode pour les GLM (*generalized linear models*). Le principe reste le même sauf que le critère à minimiser pour l'estimation des β_j est la log-vraisemblance pénalisée de la même manière (en ajoutant le terme $\lambda \sum_{j=1}^p |\beta_j|$). Enfin, le même principe existe également pour les modèles additifs généralisés (GAM) (Ravikumar et al. 2009).

D'un point de vue pratique, l'algorithme « glmnet » (Friedman et al. 2010) calcule très efficacement la solution du Lasso pour l'ensemble des λ de zéro (solution des moindres carrés) à sa valeur maximale permettant un modèle final non vide (les valeurs plus élevées donnent des modèles avec aucune variable). Il en résulte un « chemin de régularisation » montrant l'évolution des coefficients $\hat{\beta}_j$ estimés à mesure que λ grandit.

3.2. La régression EMD

Cette section présente la modélisation de la régression EMD. Deux modélisations complémentaires sont considérées. La première, appelée « R-EMD1 », consiste à expliquer une série sanitaire (c.-à-d. les décès ou hospitalisation par MCV) en fonction des IMF météorologiques. Cette modélisation permet de mettre en évidence les principales échelles d'influence de la météorologie sur les MCV. La deuxième modélisation, appelée « R-EMD2 », décompose à la fois la santé et la météorologie et modélise chaque IMF sanitaire en fonction des IMF météorologiques de fréquences similaires. La R-EMD2 a pour but de préciser la relation entre les MCV et la météorologie à toutes les échelles temporelles, même celles qui pourraient être « cachées » et ne pas ressortir dans la R-EMD1.

Au cours du développement et de l'application de la régression EMD, d'autres modélisations ont été considérées. En particulier, EMD permet de filtrer (ou « débruiter ») les séries sanitaires et météorologiques pour donner une régression avec séries filtrées (R-EMDF). Le filtrage d'une série se réalise en la décomposant, puis en la reconstruisant en sommant uniquement les IMF porteurs d'information selon le test de Wu et Huang (2004). Ce filtrage EMD a été jugé le plus pertinent parmi ceux ayant été développés (p. ex. Flandrin et al. 2004a; Boudraa et Cexus 2007; Wu et al. 2007; Moghtaderi et al. 2013). Le filtrage a mis en évidence plusieurs possibilités de modélisation, à savoir i) l'explication de la série sanitaire d'origine en fonction des séries météorologiques filtrées (R-EMDF1), ii) la série sanitaire filtrée en fonction des IMF météorologiques (R-EMDF2) et iii) la série sanitaire filtrée en fonction des séries météorologiques filtrées (R-EMDF3). Néanmoins, ces modélisations avec séries filtrées ont apporté des résultats redondants avec ceux de la R-EMD1, tout en ayant des performances statistiques moindres. Ils ne seront donc pas plus développés ici.

Cette section présente donc la R-EMD1 puis la R-EMD2, suivies d'une discussion sur des détails méthodologiques ayant nécessité une attention particulière. Le processus global de régression EMD est résumé dans la Figure 2. Les différences entre R-EMD1, R-EMD2 ainsi que les modèles avec séries filtrées (R-EMDF1, R-EMDF2 et R-EMDF3) sont résumées dans le Tableau 3.

3.2.1. R-EMD1 : modélisation avec météorologie décomposée

Dans cette modélisation, les IMF météorologiques sont utilisés pour expliquer la variable sanitaire d'origine. Cette modélisation, inspirée de Yang et al. (2011a), permet de mettre en évidence les fréquences des séries météorologiques les plus importantes pour l'étude des MCV. Il s'agit donc d'une modélisation qui vise à avoir une vision globale de la relation.

Dans cette modélisation, les IMF d'une série météorologique sont obtenus par EEMD (cf. section 3.1.1.1) indépendamment des autres séries météorologiques. Ceci permet de ne pas aligner les modes, contrairement à la MEMD, et donc de réduire légèrement le nombre total d'IMF météorologiques explicatifs ainsi que les colinéarités entre IMF de différentes variables météorologiques.

En dénotant $X^{(1)}, \dots, X^{(p)}$ les p séries météorologiques à disposition, la R-EMD1 s'écrit donc

$$Y \sim \sum_{l=1}^p \left(\sum_{j=1}^{n_l} C_j^{X^{(l)}} + r^{X^{(l)}} \right) + \varepsilon \quad (2)$$

où Y représente la variable sanitaire réponse, $C_j^{X^{(l)}}$ et $r^{X^{(l)}}$ représentent respectivement le $j^{\text{ème}}$ IMF et la tendance de la série $X^{(l)}$, n_l représente le nombre final d'IMF de la série $X^{(l)}$ et ε est le résidu. Comme expliqué dans la section 3.1.2, la régression est faite par la méthode du Lasso pour sélectionner les IMF les plus pertinents dans l'explication de Y .

3.2.2. R-EMD2 : modélisation avec santé et météorologie décomposées

Cette modélisation, inspirée par Hu et Si (2013), vise à décomposer la variable sanitaire et à l'expliquer en fonction des IMF météorologiques. La R-EMD2 comprend donc n_y modèles, où n_y est le nombre d'IMF sanitaires, c-à-d un modèle pour chaque IMF sanitaire.

Dans cette modélisation, les séries météorologiques sont décomposées conjointement avec la réponse sanitaire (c'est-à-dire que la série sanitaire est rajoutée comme variable supplémentaire au signal multivarié qui regroupe déjà les séries météorologiques) par MEMD. Cette façon de faire permet d'aligner les modes météorologiques aux modes sanitaires. Ainsi, les IMF \tilde{C}_j^X auront une périodicité similaire à l'IMF \tilde{C}_j^Y de même indice j . Ceci permet d'expliquer chaque \tilde{C}_j^Y

avec des IMF de fréquences similaires pour chaque variable météorologique, car les IMF météorologiques de fréquences très différentes n'apportent pas d'information au modèle. Ainsi, ce modèle s'écrit finalement

$$\tilde{C}_j^Y \sim \tilde{C}_j^{X^{(1)}} + \dots + \tilde{C}_j^{X^{(p)}} + \varepsilon \quad (3)$$

où \tilde{C}_j^Y est le $k^{\text{ème}}$ IMF de la variable sanitaire réponse obtenu conjointement avec les $\tilde{C}_j^{X^{(i)}}$. La R-EMD2 permet d'explorer la relation à toutes les échelles temporelles présentes dans les données.

Afin de pouvoir comparer directement les résultats de tous les modèles de la R-EMD2, l'élasticité des coefficients obtenus sera calculée. L'élasticité est définie par

$$\xi_j^X = \beta_j^X \frac{\bar{C}_j^X}{\bar{C}_j^Y} \quad (4)$$

où β_j^X est le coefficient associé à \tilde{C}_j^X dans le modèle explicatif de \tilde{C}_j^Y (p. ex. Lipfert 1993).

L'élasticité peut être interprétée comme un coefficient de régression sans dimension.

À l'instar de la R-EMD1, la régression de la R-EMD2 est effectuée par la méthode du Lasso. En effet, comme les IMF explicatifs de chaque modèle sont tous de fréquences similaires, ils sont probablement très fortement corrélés entre eux. Cette corrélation peut être gérée grâce au Lasso qui permet de sélectionner un sous-ensemble de variables (p. ex. Zou et Hastie 2005). Par ailleurs, Wu et Huang (2004) ont montré qu'un IMF tend à être normalement distribué, ce qui permet l'utilisation du Lasso classique dans tous les modèles de la R-EMD2.

Pour les IMF réponse \tilde{C}_j^y ayant une périodicité inférieure à 6 mois, un modèle sera appliqué pour la saison estivale et un modèle pour la saison hivernale. L'été comprend les mois de mai à septembre (inclusivement) et les autres mois font partie de l'hiver, similairement à Chebana et al. (2013).

3.2.3. Détails méthodologiques supplémentaires

Dans cette section se trouvent des précisions concernant deux détails de la méthodologie, à savoir la CV et l'utilisation des effets retardés (*lags*). Ils font l'objet de cette section, car leur utilisation a nécessité une attention particulière, pour les raisons développées dans les deux sous sections suivantes.

3.2.3.1. La validation croisée pour séries temporelles

Pour choisir ou évaluer les performances d'un modèle, un critère très utilisé est l'erreur de prédiction, traditionnellement estimée par validation croisée (Stone 1974) ou validation croisée généralisée (Craven et Wahba 1978). La première approche est utilisée dans cette étude pour choisir le paramètre λ du Lasso et pour comparer les différents modèles entre eux.

L'utilisation de la validation croisée sous-entend l'hypothèse d'indépendance des observations, afin que les groupes formés soient indépendants (Arlot et Celisse 2010). Cette hypothèse est ici violée en raison de l'utilisation de séries temporelles. L'effet pratique d'une telle violation est de privilégier des modèles très complexes (p. ex. Opsomer et al. 2001). Ce problème sera contourné en utilisant une CV par blocs, qui consiste à définir les groupes comme des blocs d'observations successives et à retirer les blocs précédant et suivant le bloc de test (Burman et al. 1994; Racine 2000; Bergmeir et Benítez 2012). Ainsi, comme illustré sur la Figure 3, les données de validation sont indépendantes des données d'ajustement (ou d'apprentissage) du modèle.

3.2.3.2. *L'utilisation des effets retardés (lags)*

Les effets retardés (lags) ont été considérés dans diverses études faisant le lien entre la santé et la météorologie (p. ex. Alberdi et al. 1998; Armstrong 2006). Lorsque les MCV sont les maladies d'intérêt, les lags reportés vont généralement de 3 jours (p. ex. Braga et al. 2002) à environ une semaine (p. ex. Ebi et al. 2004; Törő et al. 2010). Ces lags peuvent être déterminés de deux façons : avec les modèles à effets retardés distribués (DLM, p. ex. Schwartz 2000), ou par l'étude de la fonction de corrélation croisée (*cross correlation function*, CCF, p. ex. Cryer et Chan 2008).

Les DLMs consistent à considérer plusieurs lags des séries explicatives et à déterminer les lags significatifs dans l'estimation des coefficients de régression. Cependant, les DLMs ne sont pas utilisés dans cette méthodologie pour deux raisons. La première raison concerne la simplicité du modèle, car l'ajout de la contrainte des DLMs, en plus de la contrainte due au Lasso, rendrait l'estimation compliquée. Deuxièmement, comme les IMF sont des oscillations régulières, les jours successifs sont extrêmement similaires (surtout pour les IMF à basses fréquences) et n'apportent donc pas d'information significative.

La présente méthodologie ne considère donc qu'un lag pour chaque IMF explicatif. Ce lag est choisi par la deuxième méthode, l'étude de la CCF. Le processus est de tracer la CCF et de choisir le lag correspondant à la CCF maximale (en valeur absolue), ce qui revient à décaler la série explicative pour maximiser sa corrélation avec la réponse. Or, les séries explicatives sont ici des fonctions oscillantes et il en résulte une CCF qui oscille également autour de zéro. La Figure 4, qui illustre ce phénomène, montre également que la CCF peut être maximisée pour des lags très grands lorsque l'IMF en question est de basse fréquence (en l'occurrence, pour des IMF de période 13 ans, un lag d'environ 2 ans et demi soit un cinquième de la période). Ainsi, l'heuristique utilisée pour le choix du lag est de prendre celui qui correspond au premier

extremum de la CCF, si celui-ci est suffisamment élevé par rapport au maximum de la CCF (en valeur absolue toujours). Pour fixer les idées, la Figure 5 illustre le décalage de 2 ans et demi entre l'IMF réponse des décès et l'IMF explicatif des précipitations.

3.3. Application de la régression EMD

L'application de la méthodologie R-EMD aux données décrites dans la section 2 permet d'illustrer son intérêt et de l'évaluer en pratique. Cette section présente la démarche appliquée dans ce rapport et précise les logiciels statistiques utilisés.

3.3.1. Démarche

La démarche appliquée dans le présent rapport peut être résumée en trois points clés :

1) Décomposition des données et interprétation. La première partie consiste à déterminer les paramètres de décomposition optimaux (notamment l'amplitude du bruit pour l'EEMD et la NAMEMD). Une fois la décomposition effectuée, une étude de l'amplitude des IMF obtenus donne des informations préliminaires sur les modes de variation dominants de chaque variable, ainsi que sur leur évolution temporelle. Une première intuition des variations communes entre la mortalité ou morbidité par MCV et la météorologie peut donc être obtenue à cette étape.

2) Régression et interprétation. En premier lieu, le lag optimal de chaque IMF explicatif doit être déterminé en examinant la CCF. L'algorithme GLMNET peut ainsi être appliqué sur les variables explicatives retardées pour obtenir le chemin de régularisation du Lasso. Le modèle optimal est ensuite déterminé par CV pour obtenir les coefficients de régression dont seul un sous-ensemble est non-nul. L'interprétation de ces résultats permet enfin une première évaluation de la R-EMD en vérifiant leur cohérence avec la littérature, ainsi que de potentielles nouveautés dans les connaissances sur la relation entre la mortalité ou morbidité par MCV et la météorologie.

3) Évaluation des performances statistiques. L'évaluation numérique de la régression EMD se fait en utilisant les critères R^2 et CV. Le R^2 mesure la qualité d'ajustement des modèles comparés et le score de CV mesure la capacité du modèle à rester pertinent pour les données non encore mesurées. Ces critères sont utilisés pour comparer la R-EMD aux modèles les plus utilisés dans la littérature des relations entre la santé et le climat. Ces modèles sont les GAM (Hastie et Tibshirani 1986) et les modèles non linéaires à effet retardé distribué (DLNM, Gasparrini et al. 2010). Les GAM ont notamment permis de mettre en évidence des relations en forme de J, U ou V entre la mortalité et la température, et sont donc maintenant très utilisés (p. ex. Braga et al. 2001; Samoli et al. 2001; Dominici et al. 2002; Doyon et al. 2006; Kan et al. 2010; Dukić et al. 2012; Qiu et al. 2012). Les DLNM incluent en plus la possibilité de considérer plusieurs lags dans le modèle, et permettent ainsi de prendre en compte le temps d'induction des expositions (p. ex. Huang et al. 2012; Li et al. 2013; Vutcovici et al. 2013; Wu et al. 2013; Vanos et al. 2014). Ainsi, les GAM et DLNM sont appliqués aux données utilisées dans cette étude à des fins de comparaison.

3.3.2. Implémentation

La quasi-totalité du projet est réalisée avec le logiciel R (R Core Team 2014). Notamment, les données sont décomposées en utilisant les package « EMD » (Kim et Oh 2009) et « hht » (Bowman et Lees 2014). Seule la décomposition MEMD est faite en utilisant la version 2012a du logiciel matlab (The MathWorks Inc. 2012) et plus précisément un ensemble de programmes écrits par Naveed ur Rehman et Danilo P. Mandic disponibles librement à l'adresse <http://www.commsp.ee.ic.ac.uk/~mandic/research/emd.htm> (consulté le 25 février 2015).

La régression par la méthode du Lasso est réalisée à partir du package R « glmnet » (Friedman et al. 2010). Enfin, les applications des régression GAM et DLNM sont réalisées à partir des packages « mgcv » (Wood 2006) et « dlnm » (Gasparrini 2011).

4. Résultats

Cette partie décrit les résultats obtenus avec l'utilisation de la méthodologie détaillée dans la section 3. Une première section décrit les résultats pour la communauté métropolitaine de Montréal, puis une deuxième décrit les résultats pour la communauté métropolitaine de Québec. Au sein de chacune de ces sections, les caractéristiques des IMF obtenus sont d'abord présentées (tous les IMF sont présentés dans l'annexe B), suivies des modèles de régression sur les décès puis sur les hospitalisations.

Pour plus de clarté, dans la suite un IMF est dénoté $C_{j,f}^X$ (ou $\tilde{C}_{j,f}^X$ dans le cas des IMF issus de la NA-MEMD) où j correspond à l'ordre d'obtention de l'IMF (par fréquence descendante, cf. section 3.1.1.1), f correspond à sa périodicité moyenne et X à la série dont il est issu. L'unité de la périodicité sera « j » pour « jour », « m » pour « mois » ou « a » pour « année ». Par exemple, $C_{7,1a}^{Tmoy}$ est le 7e IMF obtenu dans la décomposition de la série de températures moyennes et représente ses variations annuelles.

Concernant le modèle de régression, le Lasso classique (gaussien) est utilisé pour les décès et hospitalisations de la CMM et pour les hospitalisations de la CMQ. Seule la variable des décès de la CMQ n'a pas une moyenne suffisamment élevée pour permettre la convergence de la loi de Poisson vers une loi normale, ce qui est confirmé par les histogrammes de l'annexe A. Ainsi un Lasso avec loi de Poisson est appliqué pour les décès de la CMQ. L'annexe A montre également

qu'une correction de l'effet fin de semaine est nécessaire pour les variables d'hospitalisations. Enfin, l'utilisation des GAM (Hastie et Tibshirani 1986) ne sera pas évoquée dans la présentation des résultats des modélisations ci-dessous, car l'analyse des résidus partiels (p. ex. Cornillon et Matzner-Løber 2007, chap. 4) ne suggère jamais de relation non linéaire entre la réponse et les IMF dans la présente étude. Il n'est donc pas nécessaire de complexifier davantage le modèle.

Tous les détails des résultats ne sont pas présentés dans la présente section. Par exemple, la présentation des résultats des modélisations R-EMD1 et R-EMD2 donne directement les IMF explicatifs présents dans le modèle final. Ainsi, les courbes de CV pour déterminer le meilleur modèle et les chemins de pénalisations du Lasso ne sont pas présentés directement dans les résultats. Ces détails sont néanmoins accessibles dans l'annexe C. Il en est de même pour les résultats des modèles R-EMDF1, R-EMDF2 et R-EMDF3 qui, s'ils ne sont pas développés dans le texte, sont présentés dans les annexes D, E, F, et G.

Si la présentation de la R-EMD1 est directe et peut être résumée sous la forme d'une seule équation, la R-EMD2 est plus complexe. Ainsi, un résumé graphique de cette modélisation sera présenté, en mettant l'accent sur les modèles les plus pertinents. Ces modèles sont notamment ceux présentant les valeurs les plus élevées de variance expliquée (R^2). Dans le cas où le modèle n'est pas gaussien (les décès de la CMQ), le R^2 correspond à un rapport de déviance, mais s'interprète de la même manière que le R^2 classique. Les résultats présentés ici prennent directement compte des lags dans la régression, selon l'heuristique décrite dans la section 3.2.3.2. Les lags associés aux $C_{j,f}^x$ explicatifs seront précisés dans le texte. Le détail de lags utilisés est néanmoins présent dans l'Annexe H.

4.1. Communauté métropolitaine de Montréal

La communauté métropolitaine de Montréal (CMM) représente le plus important bassin de population du Québec. Sa densité de population fait en sorte que de nombreux cas de MCV sont survenus sur un territoire suffisamment restreint pour que la météorologie y varie peu spatialement. Cette section commence par présenter les IMF météorologiques ainsi que sanitaires obtenus pour la CMM, puis les régressions sur les décès et les hospitalisations par MCV sont présentées.

4.1.1. Analyse préliminaire des IMF

Avant d'être utilisés comme variables de régression, les IMF obtenus peuvent être analysés pour comprendre les variations météorologiques et sanitaires de la CMM. Que ce soit pour l'EEMD ou le NA-MEMD, le critère d'arrêt utilisé pour l'obtention d'un IMF est celui de Rilling et al. (2003). Pour l'ajout du bruit, le nombre d'ensembles est fixé à $N_e = 500$ dans le cas de l'EEMD et $N_e = 100$ pour le NA-MEMD. Le niveau de bruit σ_e utilisé est fixé à 2% de l'écart-type de la série multivariée pour le NA-MEMD; pour l'EEMD est reporté dans le Tableau 4, car il est différent selon la série. Les σ_e de l'EEMD sont plus élevés que ce qui est recommandé dans la littérature (cf. section 3.1.1.1), mais l'utilisation de σ_e plus faibles ne réglait par complètement le mélange de modes. Une explication possible est le fait que les séries de données utilisées ici contiennent beaucoup plus de valeurs que les séries simulées utilisées lors des études de l'algorithme EEMD (p. ex. Wu et Huang 2009; Zhang et al. 2010; Colominas et al. 2013). Les valeurs de σ_e utilisées restent néanmoins cohérentes avec les résultats de Colominas et al. (2013) et n'ont pas de conséquences sur la modélisation qui s'en suit.

Pour bien voir ce que donne une décomposition EMD, la Figure 6 montre l'ensemble des IMF $C_{j,f}^{Snow}$ et la tendance r^{Snow} obtenus à partir de la série de hauteur de neige. Un ensemble d'IMF permet ainsi de voir la localisation temporelle de tous les types de variation. La série de hauteur de neige en est un cas d'école à cause de son intermittence, très bien représentée sur les IMF de très haute fréquence ($C_{1;3j}^{Snow}$ à $C_{5;2m}^{Snow}$). Ces IMF illustrent également l'intérêt d'utiliser l'EEMD pour gérer le mélange de modes. La décomposition de la Figure 6 met également en évidence le caractère saisonnier de la neige avec l'IMF $C_{8;1a}^{Snow}$ dont l'oscillation est très régulière avec le maximum au milieu de l'hiver et le minimum en été. Enfin, la tendance r^{Snow} de la Figure 6 suggère une diminution de la hauteur de neige moyenne depuis les années 80 avec malgré tout un sursaut au milieu des années 2000. Les autres IMF obtenus sont montrés en annexe B.

Afin de résumer l'information de tous les $C_{j,f}^X$ obtenus par EEMD, la Figure 7 montre la relation entre la variance et la période moyenne de chaque $C_{j,f}^X$. Plus la variance est élevée et plus le $C_{j,f}^X$ en question est considéré comme un motif important de la série X dont il est issu. Il est immédiatement visible sur cette figure que la plupart des séries sont dominées par leur cycle annuel avec une série de points parfaitement alignés. C'est notamment vrai pour les trois séries de températures où $C_{7;1a}^{Tmoy}$, $C_{7;1a}^{Tmin}$ et $C_{7;1a}^{Tmax}$ (respectivement en orange, jaune et rouge sur la Figure 7) contiennent une grande partie de la variance des températures.

La seule série qui n'est pas dominée par un cycle annuel est la série des précipitations totales (en marron sur la Figure 7) qui ne dispose pas d'IMF de période 1 an. De plus, tous les $C_{j,f}^{Prec}$ sont proches de la ligne grise, indiquant que la série des précipitations montre un comportement très proche de celui d'un bruit blanc. Les précipitations ne recèlent donc pas de régularité évidente.

La Figure 7 montre que les séries sanitaires (en vert pour les décès et vert pâle pour les hospitalisations) contiennent des variations annuelles fortes, à l'instar de la météorologie. Pour les hospitalisations, il est également visible que l'IMF $C_{2;7j}^{Hosp}$ contient beaucoup de variance. Il contient donc l'effet fin de semaine souvent très présent dans les séries d'hospitalisations (p. ex. Sarmiento et al. 2011).

La Figure 8 montre les tendances r^X des séries météorologiques et sanitaires. Elle met en évidence la hausse des températures (T_{moy} , T_{max} et T_{min}) de ces 30 dernières années avec une différence d'environ 2°C en moyenne entre 1981 et 2011, ce qui pourrait être du à un effet d'urbanisation. Une hausse significative est également visible pour les précipitations ainsi que pour les hospitalisations. À l'inverse, les décès par MCV ont chuté depuis le début des années 90. Enfin, les tendances de la pression et l'humidité (H_{min} , H_{moy} et H_{max}) évoluent peu au cours de la période étudiée.

Enfin, la Figure 9 montre la relation entre période moyenne et variance des IMF $\tilde{C}_{j,f}^X$ obtenus par NA-MEMD (les IMF sont montrés en annexe B). Cette figure illustre bien l'alignement des modes de l'algorithme MEMD, car les IMF de même ordre sont très bien alignés verticalement. Seuls les IMF $\tilde{C}_{j,f}^{Snow}$ de haute fréquence ont un comportement différent à cause de leur intermittence qui fait augmenter artificiellement leur période moyenne. L'IMF $\tilde{C}_{3;7j}^{Hosp}$ ressort également afin de mettre en évidence la prédominance de l'effet fin de semaine. Le reste est similaire à la Figure 7 et la même analyse peut donc y être appliquée.

En résumé, la décomposition EMD permet d'observer la domination du cycle annuel sur les séries sanitaires et météorologiques, sauf en ce qui concerne les précipitations. Pour les

hospitalisations, l'effet fin de semaine est également observé. Enfin, les températures, les précipitations et les hospitalisations évoluent généralement à la hausse, alors que la hauteur de neige au sol ainsi que les décès tendent à diminuer.

4.1.2. Régression des décès

Cette section présente les résultats des deux modélisations décrites en 3.2 pour les décès par MCV de la CMM.

4.1.2.1. R-EMD1

La Figure 10 montre les IMFs présents dans le modèle obtenu, c.-à-d. r^{Tmax} , $C_{7;1a}^{Tmoy}$, $C_{7;1a}^{Tmin}$, $C_{8;1a}^{Pmoy}$, r^{Hmoy} et $C_{8;1a}^{Hmoy}$. La régularité des oscillations de $C_{7;1a}^{Tmoy}$ et $C_{7;1a}^{Tmin}$ en fait des bons prédicteurs des variations annuelles des décès. Leurs variations sont quasiment simultanées avec les variations annuelles des décès, car leurs lags respectifs sont de 7 et 5 jours. De plus, leurs coefficients sont négatifs, indiquant une augmentation des décès lorsque les températures sont basses, c'est-à-dire l'hiver. Quant à $C_{8;1a}^{Hmoy}$ et $C_{8;1a}^{Pmoy}$, ils présentent des modulations dans leur cycle, qui permettent d'expliquer les modulations se trouvant dans le cycle annuel des décès. Leurs coefficients sont positifs avec des lags respectifs de 3 mois et demi et 1 mois et demi environ. Ainsi, l'augmentation des décès l'hiver peut être accentuée par un automne particulièrement humide, mais également par des pressions élevées au début de l'hiver.

La forte association entre les décès par MCV et r^{Tmax} s'explique par le fait que cette dernière possède un point de rupture simultané à la rupture de la tendance des décès par MCV, peu avant l'année 1995 (cf. Figure 10). Cette simultanéité est confirmée par le lag associé à r^{Tmax} qui est de

zéro. Quant à r^{Hmoy} , il est significatif, mais avec un lag d'environ 8 ans. En effet, en considérant ce lag, sa forme est très similaire à la tendance des décès par MCV (cf. Figure 10).

Le modèle retenu ici peut ainsi s'écrire

$$\begin{aligned}
 Deces = & 19,180 - 0,086 \times C_{7;1a}^{Tmoy} - 0,001 \times C_{7;1a}^{Tmin} - 4,022 \times r^{Tmax} \\
 & + 0,680 \times C_{8;11m}^{Pmoy} + 0,037 \times C_{8;1a}^{Hmoy} + 0,570 \times r^{Hmoy} + \varepsilon
 \end{aligned} \tag{5}$$

avec des lags respectifs de 7 jours, 5 jours et 0 jour, 3 mois et demi, 1 mois et demi et 8 ans (cf. Annexe H). L'amplitude des IMFs (cf. Annexe B) reportés ici et leurs coefficients du modèle (5) permettent de calculer l'amplitude des variations des décès qu'ils induisent. Ainsi, comme les IMF $C_{7;1a}^{Tmoy}$ et $C_{7;1a}^{Tmin}$ ont une amplitude comprise entre 30 et 40°C, les coefficients associés indiquent que les différences de températures entre l'été et l'hiver sont responsables d'une différence d'environ 3 décès en moyenne entre les deux saisons (représentant presque 20% de la moyenne annuelle). Cette différence peut encore être exacerbée de 0,75 décès par les variations de $C_{8;1a}^{Hmoy}$ (entre -10 et 10% d'humidité relative) et de 0,5 décès par les variations de $C_{8;11m}^{Pmoy}$ (entre -0,4 et 0,4). Enfin, la tendance r^{Tmax} représentant un accroissement de 1°C de température, elle a permis de faire diminuer le nombre moyen de décès de 4 environ en 30 ans, en plus d'une diminution de 1 décès due à r^{Hmoy} (la tendance diminue de 2 points d'humidité relative).

La Figure 11 montre la série des décès prédite par le modèle (5) et résume bien ce qui a été présenté dans cette section. Les variations annuelles sont bien retranscrites par le modèle et sont dues à la baisse de température l'hiver. Par contre, les variations de l'humidité et de la pression sont très faiblement retranscrites et sont largement dominées par les variations des températures. Enfin, la tendance à la baisse de la mortalité par MCV est également bien expliquée par la

tendance croissante des températures maximales. L'explication de ces motifs est importante dans la mesure où ce modèle a un R^2 de 18,2% de variance expliquée, ce qui est beaucoup considérant que l'exposition à la météorologie est loin des facteurs principaux de décès par MCV.

4.1.2.2. R-EMD2

La R-EMD2 consiste en plusieurs modèles à différentes échelles temporelles, générant un grand nombre de résultats. Afin de résumer les résultats de ces modèles, la Figure 12 trace l'élasticité des coefficients en fonction de la périodicité f de l'IMF réponse $\tilde{C}_{j,f}^{Deces}$. Cette figure permet d'avoir une vision globale des variables météorologiques les plus influentes aux différentes échelles. Seuls quelques modèles intéressants seront détaillés ici. Ceux-ci peuvent être déterminés par la Figure 13 qui montre le R^2 de chaque modèle.

Les deux modèles de plus haute fréquence donnent des résultats assez similaires en mettant en évidence l'influence positive de la température minimale ainsi que celle des précipitations. Ces deux modèles traduisent en fait à peu près le même phénomène, mais dans les deux saisons. En effet, la Figure 13 montre que le R^2 est plus élevé l'hiver pour la période 3 jours, et l'été pour la période 5 jours. Le modèle hivernal à 3 jours contient notamment la pression moyenne avec un coefficient négatif traduisant ainsi les tempêtes hivernales (chute de pression, hausse des précipitations et de la neige au sol). Les lags associés étant de 1 jour, la mortalité augmente ainsi le lendemain d'une tempête de neige. Le modèle à 5 jours traduit lui plutôt les orages estivaux avec une très forte élasticité positive pour les précipitations et l'humidité moyenne. Les températures moyennes et minimales, ainsi que la pression présentent aussi une forte élasticité. Le lag associé à la pression est de 1 jour alors que les autres sont nuls, ce qui suggère une augmentation des décès pour les journées chaudes et orageuses suivant des journées ensoleillées.

Le *moyen terme* semble dominé par la présence des variables d'humidité dans les modèles. Par exemple, à l'échelle d'un mois, il y a une association positive avec les températures maximales en même temps qu'une association négative avec la hauteur de neige et l'humidité. L'effet météorologique à cette échelle est donc plutôt estival, ce qui est confirmé par un R^2 plus élevé (Figure 13). Ceci traduit notamment une influence des périodes particulièrement chaudes et sèches. Les lags se situent entre 5 et 7 jours indiquant un délai d'une semaine pour observer l'effet. À l'inverse, l'échelle des trois mois traduit plutôt un effet hivernal (avec un R^2 plus élevé l'hiver d'après la Figure 13), car l'hiver est plus long que l'été. Ici, ce sont plutôt les périodes chaudes avec des précipitations neigeuses qui ont un impact sur la hausse de la mortalité. Toujours dans le *moyen terme*, la Figure 13 indique un R^2 très élevé pour le modèle de période 1 an, ce qui confirme la variation conjointe des décès et de la météorologie à cette échelle. La Figure 12 montre que les meilleurs prédicteurs de ces variations sont la température minimale et la température moyenne.

Le *long terme* est plus compliqué à interpréter dans la mesure où les R^2 sont tous très proches de 1 à cause du peu de variance se trouvant aux basses fréquences. De plus, le peu d'oscillations complètes disponibles empêche une inférence précise de l'influence à cette échelle. L'élasticité négative des précipitations à l'échelle de 7 ans peut toutefois être notée, avec un retard d'environ 6 mois (soit $1/14^{\text{ème}}$ de période environ). Il en est de même pour l'humidité et la pression moyennes à l'échelle de 13 ans avec des lags respectifs d'environ 4 et 9 mois (soit $1/39^{\text{ème}}$ et $1/20^{\text{ème}}$ de période respectivement).

Pour évaluer globalement la R-EMD2, tous les modèles la constituant sont sommés afin d'obtenir un grand modèle explicatif de la série des décès entière (les modèles saisonniers sont utilisés pour les hautes fréquences). Il s'agit d'une façon alternative d'ajuster la R-EMD1. La Figure 14

montre que la prédiction implique beaucoup plus de motifs que dans la R-EMD1. De plus, cette prédiction résulte en un R^2 de 29,1%, ce qui constitue une amélioration par rapport au score de 18,2% pour la R-EMD1.

En résumé, la R-EMD2 des décès de la CMM met en évidence le rôle des tempêtes dans l'augmentation des MCV à *court terme*. En hiver, un lag d'un jour est mis en évidence alors qu'en été, la mortalité commence à augmenter dès la déclaration de l'évènement orageux. À l'échelle mensuelle, les périodes particulièrement sèches en été et humides en hiver semblent être associées à une augmentation de la mortalité. De plus, cette modélisation confirme le rôle prééminent des variations annuelles.

4.1.3. Régression des hospitalisations

Cette section présente les résultats obtenus pour les modèles de régression appliqués sur les hospitalisations de la CMM. Pour les hospitalisations, les données sont disponibles uniquement pour les années 1996 à 2006 inclusivement.

4.1.3.1. R-EMD1

La Figure 15 montre les IMF du modèle optimal pour expliquer les hospitalisations de la CMM. Les 5 IMF explicatifs présents dans le modèle sont $C_{7,1a}^{Tmin}$, $C_{11,13a}^{Tmin}$, $C_{7,6m}^{Snow}$, $C_{8,1a}^{Hmoy}$ et $C_{10,4a}^{Hmoy}$. Cette figure montre que $C_{7,1a}^{Tmin}$ et $C_{8,1a}^{Hmoy}$ permettent d'expliquer les variations annuelles avec des lags respectifs de 6 jours et presque 3 mois. L'IMF $C_{7,1a}^{Tmin}$ explique principalement le cycle annuel de base des hospitalisations (fort en hiver et faible en été) et $C_{8,1a}^{Hmoy}$ explique plus précisément les modulations. En effet, la Figure 15 montre que les années où $C_{8,1a}^{Hmoy}$ a une plus grande amplitude correspondent également à une plus grande amplitude des variations annuelles des décès. À une

échelle similaire, le modèle contient également $C_{7,6m}^{Snow}$ avec un coefficient positif et un lag de 1 jour. Cet IMF est maximal au début et à la fin de l'hiver et minimal au milieu de l'hiver. Ainsi, $C_{7,6m}^{Snow}$ pourrait expliquer par les neiges précoces et tardives le fait que les hospitalisations sont élevées même au printemps et automne (cf. Figure 15).

À plus *long terme*, les IMF $C_{11,13a}^{Tmin}$ et $C_{10,4a}^{Hmoy}$ sont les deux autres variables explicatives présentes dans le modèle choisi. D'après la Figure 15, $C_{11,13a}^{Tmin}$ s'ajuste très bien avec la tendance à la hausse des hospitalisations avec un coefficient positif et un lag d'un peu plus de 3 ans. Cette augmentation des températures minimales peut, à terme, expliquer le fait que la différence de moyenne du nombre d'hospitalisations s'amenuise au milieu des années 2000. L'IMF $C_{10,4a}^{Hmoy}$ semble avoir un rôle similaire, mais avec un coefficient négatif.

Le modèle final est

$$\begin{aligned} Hospitalisations = & 136,322 - 0,253 \times C_{7;1a}^{Tmin} + 6,693 \times C_{11;13a}^{Tmin} \\ & + 0,010 \times C_{7;6m}^{Snow} + 0,076 \times C_{8;1a}^{Hmoy} - 0,671 \times C_{10;4a}^{Hmoy} + \varepsilon \end{aligned} \quad (6)$$

où ε représente l'erreur gaussienne. Le coefficient de -0,253 associé à $C_{7;1a}^{Tmin}$ indique que cet IMF est responsable d'une différence d'environ 8 hospitalisations entre l'hiver et l'été (car $C_{7;1a}^{Tmin}$ varie entre -20°C et 20°C environ) augmenté d'environ 1,5 hospitalisation en prenant en compte $C_{8;1a}^{Hmoy}$. L'IMF $C_{7;6m}^{Snow}$ est responsable de plus légères variations dans les hospitalisations, car il peut augmenter le nombre d'hospitalisations d'environ 0,3 entre son minimum (au milieu de l'hiver) et son maximum (début et fin de l'hiver). $C_{11;13a}^{Tmin}$ a une amplitude d'environ 1°C et est donc

responsable d'une augmentation de presque 7 hospitalisations entre 1996 et 2006. Enfin, $C_{10;4a}^{Hmoy}$ ayant un coefficient de -0,671, il induit des variations *long terme* d'environ 2 hospitalisations.

La Figure 16 montre l'ajustement obtenu par le modèle (6). Le cycle annuel est principalement prédit, mais avec des modulations, induites par $C_{8,1a}^{Hmoy}$, $C_{10;4a}^{Hmoy}$ et $C_{7,6m}^{Snow}$. La tendance est également bien ajustée avec le coefficient très élevé attribué à $C_{11,13a}^{Tmin}$ (car les hospitalisations ne couvrent qu'une période de 11 ans). Cette prédiction est cependant moins bonne que pour les décès, car elle explique $R^2 = 4,1\%$ de la variabilité des hospitalisations. En modélisant sans tendance, le R^2 monte à 5,7% confirmant ainsi que son retrait ne change pas fondamentalement la qualité du modèle. La météorologie semble donc moins pertinente dans l'explication des hospitalisations que dans l'explication des décès.

4.1.3.2. R-EMD2

Les élasticités (4) de la Figure 17 résument l'ensemble de la R-EMD2 des hospitalisations de la CMM. La Figure 18 montre la variance expliquée de chacun des modèles de cette R-EMD2 afin d'indiquer quels sont les motifs des hospitalisations de la CMM qui sont le mieux expliqués par la météorologie. En premier lieu, les variations à une échelle de 5 jours sont bien expliquées l'été. Il y a notamment une forte association négative avec l'humidité moyenne, suggérant une hausse des hospitalisations pour les jours très secs (avec un lag nul). La Figure 18 indique logiquement de très faibles valeurs de R^2 pour les modèles de périodicité 7 jours, car il s'agit principalement de l'effet fin de semaine, qui ne peut donc être bien modélisé par la météorologie.

À l'échelle de deux semaines (15 jours), le modèle estival a un R^2 élevé (presque 10% comparativement à une valeur très proche de 0 pour l'hiver) avec une forte association négative avec les variables de températures minimales et maximales et une association positive avec

l'humidité. Les lags étant compris entre 2 et 6 jours, il y a donc une hausse des hospitalisations suivant des périodes avec beaucoup d'humidité et des températures faibles pour la saison.

À plus *long terme*, les modèles saisonniers de l'échelle 3 mois ont tous deux une variance expliquée élevée avec seulement deux variables explicatives. Le modèle hivernal montre une association positive avec la pression moyenne et une association négative avec l'humidité maximale. Les lags associés sont respectivement de 1 et 5 jours ce qui signifie qu'il y a une hausse des hospitalisations pour de longues périodes sèches et ensoleillées en hiver. Le modèle estival indique un phénomène similaire en contenant uniquement les variables de précipitations et hauteurs de neige avec des coefficients négatifs. Les variations décrites dans ce modèle ont plus d'amplitude en début d'été (non montré) ce qui explique la présence de la variable de hauteur de neige dans un modèle estival (car il peut rester de la neige au sol début mai, ou y en avoir un peu fin septembre). Les deux variables Prec et Snow ont des lags de 0 et 22 jours respectivement, ce qui peut indiquer une forte morbidité pour les printemps secs suivant une fin d'hiver avec peu de neige.

Les R^2 des basses fréquences atteignent des valeurs élevées (supérieures à 50%, alors que les R^2 dépassent rarement 10% pour les hautes fréquences), mais, à l'instar des décès, cela est surtout dû au peu d'oscillations présentes à ces échelles. C'est surtout vrai pour les hospitalisations pour lesquelles il y a seulement 11 ans de données. Le modèle annuel ressort cependant particulièrement cette fois encore avec une influence négative des températures minimales et une influence positive de l'humidité moyenne. Les lags étant de 1 semaine et presque 3 mois respectivement, ce modèle est parfaitement cohérent avec la R-EMD1. Le modèle de la tendance est également cohérent avec la R-EMD1 en indiquant que la hausse de la morbidité puisse être

liée à l'augmentation des températures, mais également à une baisse de l'humidité et des précipitations hivernales.

La Figure 19 montre la prédiction réalisée en sommant tous les modèles présentés (seulement les modèles saisonniers pour les périodicités inférieures à 3 mois) dans la Figure 17. Elle montre que la tendance et le cycle annuel sont bien ajustés, mais que la variabilité est largement sous-estimée. Cette variabilité est principalement due à l'effet fin de semaine qui n'est pas lié à la météorologie. Cet ajustement entraîne une variance expliquée $R^2 = 9\%$ ce qui est là encore plus élevé que pour la R-EMD1.

Pour résumer, la R-EMD2 des hospitalisations de la CMM met principalement en évidence des modèles estivaux. En effet, les jours très secs (au contraire des périodes plus longues particulièrement humides) semblent associés à des hausses de morbidité. À plus *long terme*, les hivers secs et ensoleillés (généralement associés à des températures très froides) suivis de printemps toujours secs augmentent la morbidité. Le rôle de l'hiver de tant que tel avec les variations annuelles est toujours souligné, avec l'influence d'un automne humide.

4.2. Communauté métropolitaine de Québec

Cette section présente les résultats de la modélisation pour la communauté métropolitaine de Québec (CMQ). La première section présentera les $C_{j,f}^X$ obtenus par EEMD et les deux sections suivantes présenteront les résultats de régression pour les décès puis les hospitalisations.

4.2.1. Analyse préliminaire des IMF

La plupart des paramètres des décompositions EEMD et NA-MEMD sont les mêmes pour la CMQ que pour la CMM (cf. section 4.1.1). La seule exception étant le niveau de bruit pour EEMD, le Tableau 4 présente les valeurs utilisées pour la CMQ. Le niveau de bruit utilisé est

souvent plus élevé pour la CMQ que pour la CMM (cf. Tableau 4). Il en résulte en un nombre de $C_{j,f}^X$ légèrement plus élevé que pour la CMM (par exemple, il y a 112 $C_{j,f}^X$ météorologiques dans la CMM et 116 dans la CMQ). Les IMF ainsi obtenus sont montrés en annexe B.

La relation variance-période moyenne des $C_{j,f}^X$ est reportée sur la Figure 20. Les décompositions EEMD des séries de la CMQ sont très similaires à celles de la CMM, car la variance des séries est globalement répartie de la même façon sur les échelles temporelles. En effet, les variations annuelles sont toujours le motif principal, de même que la faible variance des hautes fréquences de températures et des hauteurs de neige qui sont des séries particulièrement dominées par le cycle annuel. Ensuite $C_{2;7j}^{Hosp}$, qui transcrit l'effet fin de semaine des hospitalisations, porte moins de variance que dans la CMM, car il se détache maintenant très peu des autres $C_{2,f}^X$.

Sur la Figure 21, qui montre les tendances r^X , se retrouve la tendance à la hausse des trois séries de températures. Plus surprenant, les trois séries d'humidité présentent cette fois également une tendance croissante indiquant une augmentation de 5~6% en moyenne depuis les années 80. Du côté des variables sanitaires, les mêmes tendances que pour la CMM, c'est-à-dire à la baisse pour les décès et à la hausse pour les hospitalisations, se retrouvent dans les données de la CMQ.

Les IMF $\tilde{C}_{j,f}^X$ obtenus par NA-MEMD permettent d'obtenir les mêmes conclusions et la principale différence réside dans l'alignement des modes. Ils ne sont donc pas présentés en détail ici.

En résumé, à l'instar des séries de la CMM, les variations annuelles dominent les séries de la CMQ. Les tendances à la hausse des températures et des hospitalisations, ainsi que la baisse des décès sont toujours observables. Cependant, contrairement à la CMM, les tendances des

précipitations et de la hauteur de neige ne mettent pas en évidence de changement net. Cette fois, ces sont les variables d'humidité qui montrent une hausse importante.

4.2.2. Régression des décès

Cette section présente les résultats des deux modélisations appliquées à la série des décès par MCV de la CMQ.

4.2.2.1. R-EMD1

Comme expliqué dans l'introduction de la section 4, la présente modélisation est la seule qui a nécessité l'utilisation d'une loi autre que la gaussienne pour la réponse, en l'occurrence une Poisson (voir annexe A pour plus de détails). Les IMF présents dans le modèle obtenu pour expliquer les décès de la CMQ sont $C_{7;1a}^{Tmoy}$, r^{Pmoy} , $C_{8;1a}^{Hmin}$ et $C_{13;20a}^{Hmoy}$ qui représentent deux cycles annuels, une tendance et un cycle de 20 ans, montrés dans la Figure 22. Ce résultat est très similaire à celui de la R-EMD1 des décès de la CMM (cf. section 4.1.2.1). En effet, le modèle contient à nouveau deux cycles annuels représentant la température moyenne et une variable d'humidité ($Hmin$ au lieu de $Hmoy$), avec des lags respectifs de 7 jours et 2 mois environ. En revanche, ici la chute des décès est mieux expliquée par r^{Pmoy} (alors que c'était r^{Tmax} pour la CMM) et par $C_{13;20a}^{Hmoy}$ avec des lags respectifs nuls et d'environ 12 ans.

Le modèle obtenu pour expliquer les décès de la CMQ s'exprime finalement

$$\begin{aligned} \log(Décès) = & -416,999 - 0,005 \times C_{7;1a}^{Tmoy} + 4,121 \times r^{Pmoy} \\ & + 0,001 \times C_{8;1a}^{Hmin} - 0,010 \times C_{13;20a}^{Hmoy} + \varepsilon \end{aligned} \quad (7)$$

où ε représente l'erreur gaussienne du modèle. Ici, les variations annuelles de températures et d'humidité sont chacune responsables d'environ 1 décès de différence entre l'hiver et l'été

(représentant environ 25% de la moyenne globale qui est de 4 décès quotidiens). Pour la tendance des décès, $C_{13;20a}^{Hmoy}$ et r^{Pmoy} sont respectivement associés à une baisse de 1 et 1,5 décès durant la période étudiée.

L'ajustement résultant du modèle (7) est montré dans la Figure 23. Il s'agit d'une faible oscillation annuelle avec une tendance descendante. Le cycle annuel est plus faible que pour les cas de la CMM, car il y a peu de décès et ce cycle est donc moins évident dans les données que pour la CMM. Le cycle annuel constitue cependant toujours un motif important, car ce modèle présente une valeur de $R^2 = 4,9\%$.

4.2.2.2. R-EMD2

Bien que la série des décès par MCV de la CMQ ne suit pas bien une loi normale (cf. annexe A), ses IMF $\tilde{C}_{j:f}^{Deces}$ sont cependant distribués normalement comme affirmé par Wu et Huang (2004). Cette propriété des IMF permet d'utiliser l'algorithme classique du Lasso. L'ensemble des résultats de la R-EMD2 des décès de la CMQ se trouve dans la Figure 24. De plus, la Figure 25 montre le R^2 associé à chacun des modèles de la R-EMD2.

La Figure 24 montre que les modèles de plus hautes fréquences (3 et 5 jours) sont très liés avec la pression moyenne. Ainsi, le rôle des changements de pression indiquant des passages de « beau temps » ou « mauvais temps » est souligné dans la CMQ. Le modèle hivernal de 3 jours inclut notamment les variables d'humidité avec des coefficients négatifs. Ainsi les jours secs couplés à des chutes de pression semblent avoir un effet positif sur les décès par MCV. À 5 jours, le modèle estival montre une influence positive des températures minimales, soulignant le stress lié aux jours très chauds. Pour une période un peu plus élevée, le modèle à 15 jours reporte également une influence instantanée des chutes de pression.

À *moyen terme*, le modèle estival de périodicité 2 mois montre une influence des périodes avec peu de précipitations et des températures maximales basses. Les lags associés sont de 10 et 13 jours pouvant indiquer une hausse des décès lorsque de telles conditions suivent une période particulièrement douce et ensoleillée.

Enfin, à plus *long terme*, la Figure 25 indique à nouveau un très bon ajustement pour le modèle de période 1 an ainsi que pour le modèle de période 8 ans. Pour le modèle à un an, la Figure 24 montre que la température moyenne a une élasticité particulièrement éloignée de zéro, traduisant le phénomène d'augmentation des décès l'hiver, déjà observé dans la R-EMD1 (cf. section 4.2.2.1). Le modèle de période 8 ans attribue une forte élasticité positive à l'humidité minimale avec un lag d'un peu plus de 2 ans. Ceci suggère qu'une succession d'années particulièrement humides (2 ou 3 ans) pourrait contribuer à affaiblir certaines personnes et les rendre plus à risque de décéder lors d'un autre stress thermique.

La Figure 26 montre la prédiction des décès de la CMQ obtenue en réunissant tous les modèles de la R-EMD2. Les variations annuelles sont bien obtenues avec certaines de leurs modulations. Des extrêmes sont également modélisés (par le modèle à 3 jours), mais ne semblent pas correspondre à des extrêmes particuliers dans la série observée. De tels extrêmes sont cependant difficiles à voir dans une série discrète avec des valeurs de si faible amplitude. Cette modélisation permet d'obtenir une variance expliquée globalement de $R^2 = 5,6\%$.

Pour résumer, en plus de l'influence des variations annuelles, la R-EMD2 des décès de la CMQ met en évidence le rôle des variations de pression à *court terme*, avec les passages très secs et ensoleillés en hiver et au contraire les orages humides et chaud en été. De plus, une succession d'années particulièrement humides peut augmenter la morbidité des années suivantes.

4.2.3. Régression des hospitalisations

Cette section présente les résultats des modélisations sur les hospitalisations de la CMQ.

4.2.3.1. R-EMD1

La Figure 27 montre les IMF présents dans le modèle pour expliquer les hospitalisations de la CMQ. On y trouve notamment trois IMF représentant des variations annuelles, c.-à-d. $C_{8;1a}^{Snow}$, $C_{9;1a}^{Hmoy}$ et $C_{7;1a}^{Tmax}$. Ainsi, les variations annuelles sont encore très importantes pour expliquer les variations des hospitalisations par MCV. Les lags associés à $C_{8;1a}^{Snow}$, $C_{9;1a}^{Hmoy}$ et $C_{7;1a}^{Tmax}$ sont respectivement nul, de 3 mois et de 5 jours. Ainsi, cela correspond à une augmentation de la morbidité lors de l'hiver, représentée par la chute des températures et l'augmentation de la neige, modulée par l'humidité automnale. De plus, à l'instar de la R-EMD1 des hospitalisations de la CMM, l'IMF $C_{7;6m}^{Snow}$ est présent dans le modèle, pour expliquer le fait que les hospitalisations sont encore élevées au début et à la fin de l'hiver quand la température n'est pas encore trop basse.

Pour la CMQ, comme pour la CMM, deux IMF sont associés à la tendance des hospitalisations. Cependant, cette fois-ci, ce sont $C_{11;10a}^{Pmoy}$ et $C_{11;6a}^{Hmin}$ qui représentent des changements moins forts que dans le cas de la CMM. Ainsi, l'augmentation moyenne de la morbidité de la CMQ peut être associée notamment à une augmentation moyenne de la pression, mais également à une diminution de l'humidité minimale (avec cependant un retard de deux ans ici). Il est possible que cette période ait pu être plus sèche qu'habituellement.

Finalement, le modèle obtenu s'écrit

$$\begin{aligned} Hospitalisations = & 32,006 - 0,038 \times C_{7;1a}^{Tmax} + 0,027 \times C_{7;6m}^{Snow} + 0,016 \times C_{8;1a}^{Snow} \\ & + 4,282 \times C_{11;10a}^{Pmoy} - 0,143 \times C_{11;6a}^{Hmin} + 0,057 \times C_{9;1a}^{Hmoy} + \varepsilon \end{aligned} \quad (8)$$

où ε représente le terme d'erreur gaussienne. Les variations annuelles de la météorologie sont responsables de différences d'hospitalisations entre l'hiver et l'été, à hauteur de 1 hospitalisation pour $C_{7;1a}^{Tmax}$, 0,5 pour $C_{9;1a}^{Hmoy}$ et 1,2 pour $C_{8;1a}^{Snow}$. Ainsi, ces trois phénomènes mis ensemble mènent à une différence de moyenne d'environ 3 hospitalisations entre l'hiver et l'été. L'IMF $C_{7;6m}^{Snow}$ augmente également le nombre d'hospitalisations d'environ 0,5 sur le début et la fin de l'hiver. Enfin $C_{11;10a}^{Pmoy}$ et $C_{11;6a}^{Hmin}$ sont respectivement responsable d'une chute de 1,5 et d'environ 0,1 hospitalisation sur la période étudiée.

La Figure 28 montre l'ajustement obtenu avec le modèle (8). Le cycle annuel est une fois encore le plus représenté, car il représente des variations d'environ 3 hospitalisations. Le reste des variations de la série des hospitalisations semble être principalement du bruit. Cet ajustement correspond à une variance expliquée de $R^2 = 2,3\%$, ce qui est très faible.

4.2.3.2. R-EMD2

La Figure 29 récapitule les modèles de la R-EMD2 des hospitalisations de la CMQ. Afin de déterminer quels sont les modèles les plus intéressants, la Figure 30 montre les taux de variance expliquée R^2 de chacun des modèles. Elle indique notamment que les hospitalisations par MCV sont mieux expliquées l'été.

Pour la périodicité 5 jours, le modèle estival est le plus intéressant avec une forte influence positive des précipitations ayant un lag nul. De plus les variables de température moyenne et de pression moyenne ont une influence positive avec des lags de 1 et 3 jours, et l'humidité moyenne a une influence négative avec un lag de 3 jours. Ceci suggère une augmentation de la morbidité pour les jours pluvieux et chauds qui suivent des jours secs et ensoleillés (les lags de 3 jours de la pression et de l'humidité). Le modèle hivernal souligne également le rôle des précipitations avec

cette fois une influence négative et un lag de 1 jour. La température maximale est également présente avec une influence positive et un lag nul ce qui pourrait indiquer une augmentation de la morbidité pour les jours chauds qui suivent un jour sec (donc potentiellement plus froid). Mais le modèle hivernal ressort surtout pour la périodicité à 7 jours avec une influence positive de la température maximale et une influence négative de la pression moyenne. Le lag est nul pour les deux variables ce qui indique une augmentation de la morbidité pour les tempêtes, lorsque la température augmente et qu'il y a une dépression.

À *moyen terme* (c.-à-d. 2 et 3 mois) les modèles estivaux ressortent particulièrement. À une périodicité de 2 mois, il y a une influence positive de l'humidité maximale avec un lag de 2 semaines et demie indiquant une augmentation de la morbidité suivant des périodes particulièrement humides. À 3 mois, il y a une influence négative de l'humidité moyenne indiquant une augmentation de la morbidité pour de longues périodes sèches. L'effet se fait sentir à partir de 10 jours après le début de l'épisode sec.

Pour les modèles de plus *long terme*, la Figure 30 indique à nouveau un très bon ajustement pour le modèle annuel, et la Figure 29 montre qu'il indique les mêmes conclusions que la R-EMD1, c'est-à-dire que les variations annuelles des décès sont induites par la baisse des températures et les chutes de neige hivernales ainsi que par la hausse d'humidité automnale. De plus, ici la tendance à la hausse des hospitalisations peut être associée à la hausse de l'humidité constatée sur la Figure 21.

La Figure 31 indique l'ajustement obtenu en combinant tous les modèles de la R-EMD2. La variabilité est encore plus sous-estimée que pour la CMM. La tendance et le cycle annuel sont par

contre une fois de plus très bien ajustés. Cette modélisation entraîne une variance expliquée de $R^2 = 4,7\%$.

5. Discussion

Cette section discute les résultats présentés dans la section 4 à travers deux axes. Dans un premier temps, les résultats des modèles sont discutés d'un point de vue sanitaire et météorologique afin de les comparer aux résultats obtenus dans la littérature. Le Tableau 5 offre un résumé des résultats les plus importants à retenir. Dans un deuxième temps, le point de vue statistique est discuté afin d'évaluer les performances de la régression EMD relativement aux modèles plus souvent utilisés dans les études liant la santé et le climat.

5.1. Résumé des résultats

La décomposition des séries météorologiques et sanitaires, préliminaire à la régression EMD, permet d'avoir une idée des variations de la météorologie et de la santé dans la CMM et la CMQ. Globalement, la décomposition de la météorologie suggère largement les changements climatiques au Québec. En effet, les tendances obtenues pour les variables de températures (CMM et CMQ, voir Figure 8 et Figure 21) sont toutes croissantes avec des hausses de 1 à 2 degrés Celsius durant la période étudiée. Ces observations sont cohérentes avec le dernier rapport du GIEC (2013) qui indique que le sud du Québec est une des régions du globe ayant connu les plus grandes augmentations de température au cours du dernier siècle. L'amplitude de cette augmentation suggère également un effet d'urbanisation dans la mesure où les données utilisées sont celles des deux plus grandes villes du Québec. À plus haute fréquence, la variabilité de la température diminue l'hiver, comme l'illustre le spectre HHT (cf. section 3.1.1.1) des températures minimales de la CMQ (Figure 32). En effet, l'hiver présente plus de variations de

température (ce qui est indiqué par des couleurs qui s'éloignent du bleu dans la Figure 32). Or les hivers semblent plus courts durant les années 2000 que durant les années 80 (il y a plus de bleu clair et de vert dans les années 80 que dans les années 2000). D'après le rapport Ouranos (2010), la diminution de la variabilité hivernale des températures devrait s'intensifier au Québec durant le 21e siècle. La cohérence des résultats de décompositions montrés ici avec les rapports GIEC (2013) ainsi que Ouranos (2010) se voit également au niveau des précipitations, à la hausse dans la CMM (Figure 8). Si cette hausse des précipitations est moins évidente pour la CMQ, elle peut cependant être déduite des tendances croissantes des variables d'humidité relative (Figure 21).

La météorologie de la CMQ semble plus variable que la météorologie de la CMM, comme suggéré par la nécessité d'utiliser un bruit de plus grande amplitude pour la décomposition EEMD (Tableau 4). Une plus grande variabilité de la météorologie pourrait intuitivement se traduire par un effet plus fort de la météorologie sur les MCV. Cependant, les résultats de la section 4 montrent des taux de variance expliquée (R^2) faibles pour les modèles de la CMQ comparés aux modèles appliqués à la CMM. Ainsi, le faible nombre de cas de la CMQ (surtout pour les décès) rend moins visible l'effet de la météorologie sur les MCV.

La R-EMD1 montre que les hautes fréquences des séries sanitaires sont difficiles à expliquer en utilisant uniquement la météorologie. Elles semblent donc principalement liées à l'activité humaine (p. ex. organisation des hôpitaux), avec notamment une très forte variabilité due à l'effet fin de semaine (Figure 7). Ainsi, les R-EMD1 effectuées expliquent essentiellement les variations à *long terme* que sont le cycle annuel et les tendances des différentes séries sanitaires. En effet, la principale leçon apportée par la R-EMD1 est la hausse importante du nombre moyen de cas MCV l'hiver, liée principalement à la chute des températures. Dans tous les modèles, le meilleur prédicteur est un IMF de période 1 an issu d'une série de températures.

Dans un but de prédiction de la mortalité et de la morbidité hospitalière, aucune des trois séries de températures utilisées dans cette étude (T_{moy} , T_{min} et T_{max}) ne semble être meilleure que les autres, car les trois variables sont également présentes dans les modèles. Similairement, Barnett et al. (2010) n'avaient pu mettre en évidence une mesure de température optimale pour prédire la mortalité.

L'effet annuel de l'hiver peut être accentué ou atténué par l'humidité de l'automne ayant précédé. Dans les quatre R-EMD1, le cycle annuel de l'humidité relative était présent dans le modèle choisi avec des lags allant de 1 à 3 mois. De plus, les modulations de l'amplitude de ce cycle s'ajustent bien avec les modulations d'amplitude de la hausse de la mortalité et morbidité en hiver (voir p. ex. Figure 15). Dans le cas des hospitalisations, la neige peut allonger l'effet hivernal en induisant une hausse du nombre de cas pour les neiges prématurées et tardives, à des moments où la population n'est pas forcément préparée pour faire face à ce genre d'évènements.

Les séries sanitaires présentent des tendances fortes, à la baisse pour les décès et à la hausse pour les hospitalisations. Dans un premier temps, l'analyse des données sanitaires (annexe A) montre que ces tendances ne sont pas dues à l'évolution démographique des villes ni à l'évolution du nombre de personnes âgées de 65 ans et plus, strate particulièrement vulnérable à la fois aux MCV (Institut national de santé publique du Québec 2006) et aux changements climatiques (Doyon et al. 2006). Ceci permet de donner du crédit aux relations trouvées avec la météorologie. Ainsi, pour la CMM, à la fois la baisse des décès et la hausse des hospitalisations sont conjointes à la hausse des températures. Dans le cas des décès, la hausse des températures peut atténuer l'effet hivernal par rapport à l'effet estival (cf. Figure 10) et dans le cas des hospitalisations, le nombre de cas augmente l'été pour rattraper l'effet hivernal (cf. Figure 15). Dans le cas de la CMQ, la pression est plus présente dans les modèles avec des coefficients positifs. Ainsi, l'effet à

long terme peut être affecté par une variation de la moyenne de pression. Ces conclusions doivent tout de même être atténuées par le fait que les changements climatiques représentent un phénomène à très grande échelle temporelle dont il est difficile d'observer les conséquences sur seulement 30 ans de données (voire même 11 ans dans le cas des hospitalisations).

La R-EMD2, permet de considérer l'effet à plus *court terme* de la météorologie sur les MCV. Les évènements de « mauvais temps » avec des chutes de pression et des températures élevées (relativement à la moyenne saisonnière) sont associés avec des hausses de mortalité. L'effet est plus rapide l'hiver, car il se trouve au jour même, alors que l'été demande une exposition d'un jour supplémentaire. Les hospitalisations sont plutôt touchées l'été par la météorologie avec une augmentation de la morbidité lorsque des jours chauds et pluvieux suivent des journées plus « agréables », c'est-à-dire ensoleillées et douces.

À plus *long terme*, l'humidité semble avoir un rôle prépondérant dans les variations des décès et d'hospitalisations par MCV. En effet, dans toutes les R-EMD1 présentées ici, des périodes humides sont associées à une augmentation de la mortalité ou de la morbidité par MCV. Elle est de plus un bon prédicteur des tendances, notamment celles des hospitalisations de la CMQ. Ainsi, l'humidité semble avoir un rôle prépondérant alors que cette variable semble très peu présente dans la littérature. Elle est par exemple absente des revues de Ye et al. (2012) et de Patz et al. (2014). Ceci peut être dû à la longueur des séries dont bénéficie cette étude. De manière similaire, on peut trouver un effet des variations de précipitations (p. ex. pour les décès de la CMM, cf. section 4.1.2.2). Couplé au lien trouvé avec la pression dans le cas de la CMQ, ces résultats peuvent suggérer un effet de variations climatiques lentes telles que le phénomène ENSO ou l'oscillation nord-atlantique (NAO).

Enfin, plusieurs modèles mettent en évidence le rôle des événements « inhabituels ». Par exemple, la R-EMD1 des hospitalisations (pour la CMM, section 4.1.3.1, et la CMQ, 4.2.3.1) incluent l'IMF $C_{7;6m}^{Snow}$ qui est maximal à l'automne et au printemps et donc qui correspond à des chutes de neige précoces ou tardives à des périodes où la population est moins préparée qu'au cœur de l'hiver. C'est également le cas du modèle estival de périodicité 3 mois de la R-EMD2 (cf. Figure 17) qui traduit les périodes de transitions entre l'été et l'hiver qui peuvent durer jusqu'à début mai ou dès fin septembre (mois faisant partie de la saison estivale). Ces observations vont dans le même sens que la littérature qui indique que l'augmentation des températures en hiver n'entraînera vraisemblablement pas de diminution de la mortalité, car les populations seront moins bien préparées (voir Kinney et al. 2012), revenant ainsi au rôle des événements inhabituels.

5.2. Évaluation de la régression EMD

Cette section discute des performances numériques de la méthodologie R-EMD, comparativement aux GAM et DLNM (cf. section 3). Les résultats de ces deux derniers modèles sont présentés dans l'annexe I. Dans ces modélisations, les températures sont représentées par $Tmoy$ et l'humidité par $Hmoy$ seulement, car les très fortes colinéarités sont mal gérées par les GAM et DLNM.

Le R^2 , mesurant la qualité d'ajustement des modèles comparés, est présenté par la Figure 33 pour les quatre cas considérés dans cette étude. Elle présente des différences entre les décès et les hospitalisations, sans grande différence entre les cas de la CMM et de la CMQ. Pour les décès, la régression EMD apparaît très performante, car les GAM sont surpassés dans les deux régions (18,2% et 29,1% pour la R-EMD1 et la R-EMD2 contre 10,7% pour GAM puis 4,9% et 5,6%

contre 2,8%). La R-EMD2 surpasse également les DLNM, en ayant même un R^2 particulièrement élevé dans le cas de la CMM (29,1% de variance expliquée). La R-EMD1 apparaît cependant trop peu précise comparée aux DLNM. Dans le cas des hospitalisations, aucune des deux modélisations de la régression EMD ne parvient à atteindre des R^2 aussi élevés que les DLNM. La R-EMD2 a tout de même des scores plus élevés que la R-EMD1 et les GAM, ces deux derniers ayant des performances équivalentes.

Le CV est le deuxième critère utilisé pour la comparaison; il est montré dans la Figure 34. Cette figure permet de voir qu'au niveau de l'erreur de prédiction, les modèles de régression EMD surpassent dans tous les cas par les GAM et les DLNM. Ainsi malgré les très bons scores de R^2 des DLNM, ils sont très peu performants en termes d'erreur de prédiction. Il en est de même pour la R-EMD2, qui présente des scores de CV moins bons que ceux de la R-EMD1 (excepté pour la mortalité de la CMM) malgré des valeurs de R^2 plus élevées. Ce phénomène s'explique par la propension de la CV à privilégier des modèles simples au lieu de modèles plus compliqués (comme la R-EMD2 ou les DLNM) (p. ex. Friedman et al. 2009, chap. 7). C'est d'autant plus vrai dans le cas du lien entre la morbidité par MCV et la météorologie, cette dernière n'expliquant que peu les variations de morbidité.

La méthodologie de régression EMD présentée ici n'utilise que des modèles linéaires. Or, ses performances sont meilleures que celles des GAM qui sont des modèles non linéaires, alors même que la relation entre la température et la mortalité est souvent reportée comme étant non linéaire (p. ex. Kunst et al. 1993; Huynen et al. 2001; Doyon et al. 2006; Anderson et Bell 2009). Il en est de même ici, car la Figure 35 montre la relation en J entre la température moyenne et la mortalité de la CMM. Ce résultat suggère donc que la décomposition des séries permet de compenser le manque de flexibilité des modèles linéaires comparativement aux modèles non

linéaires, comme le suggère l'analyse des résidus partiels (Larsen et McCleary 1972) qui ne montre jamais de relation effectivement non linéaire (p. ex. Figure 36, mais tous les résidus partiels sont de la même forme). Gardant T_{moy} en tant qu'exemple, la relation en J de la Figure 35 peut être retrouvée dans la R-EMD2 des décès de la CMM (les points orange de la Figure 12). En effet, la phase descendante presque linéaire de la fonction entre -20 et 20°C est représentée par le coefficient négatif du modèle de périodicité 1 an qui représente les variations saisonnières entre -15 et 15°C environ. Pour la phase montante de la fonction GAM au-delà de 20°C, elle est représentée dans la R-EMD2 par les coefficients positifs de T_{moy} se retrouvant dans les modèles estivaux des périodicités 5 jours, 15 jours et 1 mois, qui représentent donc des variations autour de 15 à 20°C. Ainsi, une modélisation GAM permet de résumer la relation entre une exposition météorologique et la réponse sanitaire, mais la régression EMD, et en particulier la R-EMD2, permet de détailler cette relation.

La régression EMD permettant de retrouver les différentes composantes d'une relation non linéaire, les meilleures performances des DLNM en termes de variance expliquée sont alors dues à l'utilisation de lags distribués. Cependant, lorsque les expositions sont des IMF, l'utilisation de lags distribués résulte en une fonction périodique sur les lags, comme l'illustre la Figure 37. Ces modèles s'avèrent également complexes et donc peu performant dans la prédiction de nouvelles données. De plus, les DLNM ne peuvent gérer le grand nombre de variables et les colinéarités induites par l'utilisation d'IMF. Il n'existe pour le moment pas de modèle capable de gérer les aspects de colinéarité entre les variables explicatives et la distribution des lags simultanément.

6. Conclusion

Le présent rapport présente une méthodologie appelée régression EMD qui vise à utiliser les IMF des séries météorologiques et sanitaires dans les modèles de régression afin d'expliquer la relation à différentes échelles sanitaires. Cette modélisation a également la vertu d'apporter une solution aux problèmes statistiques posés par l'interdépendance et la non-stationnarité des séries sanitaires et météorologiques. Cette méthodologie consiste en deux grands axes appelés R-EMD1 et R-EMD2. La R-EMD1 modélise la variable sanitaire réponse d'origine en fonction des IMF météorologiques afin de mettre en évidence les principales échelles d'influence de la météorologie sur les MCV. La R-EMD2 décompose également la variable sanitaire, tout en expliquant des motifs pouvant être cachés dans la R-EMD1.

Les résultats de la régression EMD sont cohérents avec ce qui est connu dans la littérature épidémiologique des relations entre la santé et la météorologie. En effet, la modélisation a pu mettre en évidence l'augmentation de la mortalité et de la morbidité l'hiver, l'influence des fortes chaleurs en été, ou encore des tempêtes de neige l'hiver. La régression EMD apporte en sus une plus-value aux études épidémiologiques en ayant un bon pouvoir explicatif par rapport aux GAM, mais également en mettant en évidence des relations difficiles à voir jusqu'ici. C'est par exemple le cas des variables d'humidité qui semblent jouer un rôle prédominant dans les variations des décès ou d'hospitalisations à *moyen terme* (échelle mensuelle environ). De plus, elle permet de décomposer les relations déjà connues comme la relation en forme de J entre les températures et la mortalité. Ces aspects font de la régression EMD une approche complémentaire aux approches classiquement utilisées telles que les GAM ou les DLNM. En effet, les relations trouvées dans ces modèles peuvent être détaillées grâce à la régression EMD, tout en bénéficiant de la simplicité d'interprétation des coefficients des modèles linéaires.

Les résultats obtenus suggèrent cependant que l'utilisation de lags distribués pourrait encore augmenter les performances de la régression EMD, en témoignent les bonnes performances explicatives des DLNM par rapport aux modèles de régression EMD. Le problème réside dans l'utilisation conjointe du Lasso pour gérer un grand nombre de variables et des effets retardés distribués, qui amène à une estimation complexe n'ayant pas encore été considérée dans la littérature statistique. La version la plus proche d'un Lasso-DLM réside dans l'approche de Park et Sakaori (2013) qui ajoute des poids à la pénalisation Lasso en fonction des lags considérés. Le développement d'un tel modèle est d'un grand intérêt en épidémiologie, car, au-delà de la régression EMD, cela permettrait de considérer simultanément un grand nombre d'expositions, de gérer leurs dépendances et de considérer l'ensemble de l'intervalle temporel d'une exposition.

Bibliographie

- Alberdi, J., J. Díaz, J. Montero et I. Mirón (1998). "Daily mortality in Madrid community 1986–1992: Relationship with meteorological variables." European Journal of Epidemiology **14**(6): 571-578.
- Anderson, B. G. et M. L. Bell (2009). "Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States." Epidemiology (Cambridge, Mass.) **20**(2): 205.
- Antoniadis, A., J. Bigot et T. Sapatinas (2001). "Wavelet estimators in nonparametric regression: a comparative simulation study." Journal of Statistical Software **6**(6): 1-83.
- Arlot, S. et A. Celisse (2010). "A survey of cross-validation procedures for model selection." 40-79.
- Armstrong, B. (2006). "Models for the Relationship Between Ambient Temperature and Daily Mortality." Epidemiology **17**(6): 624-631
610.1097/1001.ede.0000239732.0000250999.0000239738f.
- Barnett, A. G., S. Tong et A. C. A. Clements (2010). "What measure of temperature is the best predictor of mortality?" Environmental Research **110**(6): 604-611.
- Bassil, K. L., D. C. Cole, R. Moineddin, A. M. Craig, W. Y. Wendy Lou, B. Schwartz et E. Rea (2009). "Temporal and spatial variation of heat-related illness using 911 medical dispatch data." Environmental Research **109**(5): 600-606.
- Bayentin, L., S. El Adlouni, T. Ouarda, P. Gosselin, B. Doyon et F. Chebana (2010). "Spatial variability of climate effects on ischemic heart disease hospitalization rates for the period 1989-2006 in Quebec, Canada." International Journal of Health Geographics.
- Bergmeir, C. et J. M. Benítez (2012). "On the use of cross-validation for time series predictor evaluation." Information Sciences **191**(0): 192-213.
- Blachère, J.-C. et S. Perreault (2012). Médicaments des systèmes cardiovasculaire et rénal et canicules: rapport et recommandations, Direction de la santé environnementale et de la toxicologie, Institut national de santé publique.
- Boudraa, A. O. et J. C. Cexus (2007). "EMD-Based Signal Filtering." Instrumentation and Measurement, IEEE Transactions on **56**(6): 2196-2202.
- Bowman, D. C. et J. M. Lees (2014). "The Hilbert-Huang Transform: Tools and Methods."
- Braga, A. L. F., A. Zanobetti et J. Schwartz (2001). "The Time Course of Weather-Related Deaths." Epidemiology **12**(6): 662-667.
- Braga, A. L. F., A. Zanobetti et J. Schwartz (2002). "The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities." Environmental health perspectives.
- Burman, P., E. Chow et D. Nolan (1994). "A cross-validatory method for dependent data." Biometrika **81**(2): 351-358.
- Bustanza, R., G. Lebel, P. Gosselin, D. Belanger et F. Chebana (2013). "Health impacts of the July 2010 heat wave in Quebec, Canada." BMC Public Health **13**(1): 56.
- Chebana, F., B. Martel, P. Gosselin, J.-X. Giroux et T. B. Ouarda (2013). "A general and flexible methodology to define thresholds for heat health watch and warning systems, applied to the province of Québec (Canada)." International journal of biometeorology **57**(4): 631-644.
- Cleveland, W. S. et S. J. Devlin (1988). "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." Journal of the American Statistical Association **83**(403): 596-610.

- Colominas, M. A., G. Schlotthauer, M. E. Torres et P. Flandrin (2013). "Noise-assisted EMD methods in action." Advances in Adaptive Data Analysis.
- Cornillon, P.-A. et É. Matzner-Løber (2007). Régression: théorie et applications, Springer Paris.
- Craven, P. et G. Wahba (1978). "Smoothing noisy data with spline functions." Numerische Mathematik **31**(4): 377-403.
- Cryer, J. D. et K.-S. Chan (2008). Time series analysis: with applications in R, Springer-Verlag New York.
- Curriero, F. C., K. S. Heiner, J. M. Samet, S. L. Zeger, L. Strug et J. A. Patz (2002). "Temperature and Mortality in 11 Cities of the Eastern United States." American Journal of Epidemiology **155**(1): 80-87.
- Daubechies, I. (1992). Ten lectures on wavelets, SIAM.
- Dominici, F., A. McDermott, S. L. Zeger et J. M. Samet (2002). "On the Use of Generalized Additive Models in Time-Series Studies of Air Pollution and Health." American Journal of Epidemiology **156**(3): 193-203.
- Donat, M., L. Alexander, H. Yang, I. Durre, R. Vose, R. Dunn, K. Willett, E. Aguilar, M. Brunet et J. Caesar (2013). "Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset." Journal of Geophysical Research: Atmospheres **118**(5): 2098-2118.
- Donoho, D. L. et J. M. Johnstone (1994). "Ideal spatial adaptation by wavelet shrinkage." Biometrika **81**(3): 425-455.
- 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.
- Doyon, B., D. Bélanger et P. Gosselin (2008). "The potential impact of climate change on annual and seasonal mortality for three cities in Québec, Canada." International Journal of Health Geographics **7**: 23.
- Dukić, V., M. Hayden, A. Forgor, T. Hopson, P. Akweongo, A. Hodgson, A. Monaghan, C. Wiedinmyer, T. Yoksas, M. Thomson, S. Trzaska et R. Pandya (2012). "The Role of Weather in Meningitis Outbreaks in Navrongo, Ghana: A Generalized Additive Modeling Approach." Journal of Agricultural, Biological, and Environmental Statistics **17**(3): 442-460.
- Ebi, K. L., K. A. Exuzides, E. Lau, M. Kelsh et A. Barnston (2004). "Weather changes associated with hospitalizations for cardiovascular diseases and stroke in California, 1983–1998." International journal of biometeorology **49**(1): 48-58.
- Flandrin, P., P. Goncalves et G. G. Rilling (2004a). Detrending and denoising with empirical mode decompositions: XXXV-2310 p.
- Flandrin, P., G. Rilling et P. Goncalves (2004b). "Empirical mode decomposition as a filter bank." Signal Processing Letters, IEEE **11**(2): 112-114.
- Friedman, J., T. Hastie et R. Tibshirani (2009). The elements of statistical learning.
- Friedman, J., T. Hastie et R. Tibshirani (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." Journal of Statistical Software **33**(1): 1-22.
- Gasparrini, A. (2011). "Distributed lag linear and non-linear models in R: the package dlnm." Journal of Statistical Software **43**(8): 1.
- Gasparrini, A. et B. Armstrong (2011). "The impact of heat waves on mortality." Epidemiology (Cambridge, Mass.) **22**(1): 68.
- Gasparrini, A., B. Armstrong et M. G. Kenward (2010). "Distributed lag non-linear models." Statistics in Medicine **29**(21): 2224-2234.

- GIEC (2013). *Climate Change: The Physical Science Basis*. C. U. Press.
- Giroux, J.-X., F. Chebana, D. Bélanger, E. Gloaguen, T. B. M. J. Ouarda et S.-H. A. (2013). *Projet M1 : Comparaison de l'utilisation des moyennes spatiales à celle du krigeage, appliquée à la relation mortalité par MCV - météorologie, au Québec, de 1996 à 2007.*, INRS-ETE.
- Granger, C. W. J. et P. Newbold (1974). "Spurious regressions in econometrics." *Journal of Econometrics* **2**(2): 111-120.
- Haiyong, Z. et G. Qiang (2006). *Research on Properties of Empirical Mode Decomposition Method*. Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on.
- Harvey, A. C. (1978). "Linear Regression in the Frequency Domain." *International Economic Review* **19**(2): 507-512.
- Hastie, T. et R. Tibshirani (1986). "Generalized Additive Models." *Statistical Science* **1**(3): 297-310.
- Hoerl, A. E. et R. W. Kennard (1970). "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* **12**(1): 55-67.
- Houck, P. D., J. E. Lethen, M. W. Riggs, D. S. Gantt et G. J. Dehmer (2005). "Relation of Atmospheric Pressure Changes and the Occurrences of Acute Myocardial Infarction and Stroke." *The American Journal of Cardiology* **96**(1): 45-51.
- Hu, W. et B. C. Si (2013). "Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition." *Geoderma* **193-194**(0): 180-188.
- Huang, C., A. G. Barnett, X. Wang et S. Tong (2012). "Effects of Extreme Temperatures on Years of Life Lost for Cardiovascular Deaths: A Time Series Study in Brisbane, Australia." *Circulation: Cardiovascular Quality and Outcomes* **5**(5): 609-614.
- Huang, C., A. G. Barnett, X. Wang, P. Vaneckova, G. FitzGerald et S. Tong (2011). "Projecting future heat-related mortality under climate change scenarios: a systematic review." *Environmental health perspectives*.
- Huang, N. E., Z. Shen et S. R. Long (1999). "A New View of Nonlinear Water Waves : The Hilbert Spectrum1." *Annual Review of Fluid Mechanics* **31**(1): 417-457.
- Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung et H. H. Liu (1998). "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.
- Huang, T.-L., W.-X. Ren et M.-l. Lou (2008). The orthogonal Hilbert-Huang transform and its application in earthquake motion recordings analysis. *The 14th World Conference on Earthquake Engineering*. Beijing, China.
- Huynen, M. M., P. Martens, D. Schram, M. P. Weijenberg et A. E. Kunst (2001). "The impact of heat waves and cold spells on mortality rates in the Dutch population." *Environmental health perspectives* **109**(5): 463-470.
- Institut de la statistique du Québec [ISQ] (2000). *Perspectives démographiques du Québec 1996-2041, régions administratives, régions métropolitaines et municipalités régionales de comté*. Québec.
- Institut de la statistique du Québec [ISQ] (2009). "Décès selon les principaux groupes de causes, sexes réunis, Québec, 2000-2009."
- Institut national de santé publique du Québec (2006). "Les maladies du coeur et les maladies vasculaires cérébrales : prévalence, morbidité et mortalité au Québec." *INSPQ*.

- Junsheng, C., Y. Dejie et Y. Yu (2006). "Research on the intrinsic mode function (IMF) criterion in EMD method." Mechanical Systems and Signal Processing **20**(4): 817-824.
- Kan, H., C.-M. Wong, N. Vichit-Vadakan et Z. Qian (2010). "Short-term association between sulfur dioxide and daily mortality: The Public Health and Air Pollution in Asia (PAPA) study." Environmental Research **110**(3): 258-264.
- Keatinge, W. R. (2002). "Winter mortality and its causes." 2002 **61**(4).
- Kim, D., K. Kim et H.-S. Oh (2012). "Extending the scope of empirical mode decomposition by smoothing." EURASIP Journal on Advances in Signal Processing **2012**(1): 1-17.
- Kim, D. et H.-S. Oh (2009). "EMD: A Package for Empirical Mode Decomposition and Hilbert Spectrum." The R Journal **1**(1): 40-46.
- Kinney, P., M. Pascal, R. Vautard et K. Laaidi (2012). "La mortalité hivernale va-t-elle diminuer avec le changement climatique?" Bulletin Épidémiologique Hebdomadaire **12-13**.
- Kişi, Ö. (2009). "Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting." Hydrological processes **23**(25): 3583-3597.
- Knowlton, K., M. Rotkin-Ellman, G. King, H. G. Margolis, D. Smith, G. Solomon, R. Trent et P. English (2009). "The 2006 California heat wave: impacts on hospitalizations and emergency department visits." Environ Health Perspect **117**(1): 61-67.
- Kunst, A. E., C. W. N. Looman et J. P. Mackenbach (1993). "Outdoor Air Temperature and Mortality in the Netherlands: A Time-Series Analysis." American Journal of Epidemiology **137**(3): 331-341.
- Lan Chang, C., M. Shipley, M. Marmot et N. Poulter (2004). "Lower ambient temperature was associated with an increased risk of hospitalization for stroke and acute myocardial infarction in young women." Journal of Clinical Epidemiology **57**(7): 749-757.
- Larsen, W. A. et S. J. McCleary (1972). "The Use of Partial Residual Plots in Regression Analysis." Technometrics **14**(3): 781-790.
- Lee, D. S., M. Chiu, D. G. Manuel, K. Tu, X. Wang, P. C. Austin, M. Y. Mattern, T. F. Mitiku, L. W. Svenson, W. Putnam, W. M. Flanagan, J. V. Tu et f. t. C. C. O. R. Team (2009). "Trends in risk factors for cardiovascular disease in Canada: temporal, socio-demographic and geographic factors." Canadian Medical Association Journal **181**(3-4): E55-E66.
- Lee, T. et T. B. M. J. Ouarda (2010). "Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes." Journal of Geophysical Research: Atmospheres **115**(D13): n/a-n/a.
- Li, T., R. M. Horton et P. L. Kinney (2013). "Projections of seasonal patterns in temperature-related deaths for Manhattan, New York." Nature Clim. Change **3**(8): 717-721.
- Lipfert, F. W. (1993). "A critical review of studies of the association between demands for hospital services and air pollution." Environ Health Perspect.
- Loh, C.-H., T.-C. Wu et N. E. Huang (2001). "Application of the Empirical Mode Decomposition-Hilbert Spectrum Method to Identify Near-Fault Ground-Motion Characteristics and Structural Responses." Bulletin of the Seismological Society of America **91**(5): 1339-1357.
- MacCracken, M., E. Barron, D. Easterling, B. Felzer et T. Karl (2001). Scenarios for climate variability and change. Climate change impacts on the United States: the potential consequences of climate variability and change. N. A. S. Team. Cambridge, Cambridge University Press.
- Matthies, F., G. Bickler, N. C. Marín et S. Hales (2008). Heat-health action plans: guidance, World Health Organization.
- McCullagh, P. et J. A. Nelder (1989). "Generalized linear models."

- Ministère de la santé et des Services sociaux, Institut national de santé publique du Québec [INSPQ] et Institut de la statistique du Québec [ISQ] (2011). Pour guider l'action - Portrait de santé du Québec et de ses régions : les statistiques. L. g. d. Québec.
- Moghtaderi, A., P. Flandrin et P. Borgnat (2013). "Trend filtering via empirical mode decompositions." Computational Statistics & Data Analysis **58**(0): 114-126.
- Morabito, M., F. Profili, A. Crisci, P. Francesconi, G. Gensini et S. Orlandini (2012). "Heat-related mortality in the Florentine area (Italy) before and after the exceptional 2003 heat wave in Europe: an improved public health response?" International journal of biometeorology **56**(5): 801-810.
- Nitschke, M., G. R. Tucker, A. L. Hansen, S. Williams, Y. Zhang et P. Bi (2011). "Impact of two recent extreme heat episodes on morbidity and mortality in Adelaide, South Australia: a case-series analysis." Environ Health **10**(1): 42.
- Opsomer, J., Y. Wang et Y. Yang (2001). "Nonparametric regression with correlated errors." Statistical Science **16**(2): 134-153.
- Ouranos (2010). "Savoir s'adapter aux changements climatiques." Ouranos.
- Park, H. et F. Sakaori (2013). "Lag weighted lasso for time series model." Computational Statistics **28**(2): 493-504.
- Park, M. Y. et T. Hastie (2007). "L1-regularization path algorithm for generalized linear models." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **69**(4): 659-677.
- Patz, J. A., D. Campbell-Lendrum, T. Holloway et J. A. Foley (2005). "Impact of regional climate change on human health." Nature **438**(7066): 310-317.
- Patz, J. A., H. Frumkin, T. Holloway, D. J. Vimont et A. Haines (2014). "Climate change: challenges and opportunities for global health." JAMA.
- Peel, M., R. Srikanthan, T. McMahon et D. Karoly (2011). Ensemble Empirical Mode Decomposition of monthly climatic indices relevant to Australian hydroclimatology, MODSIM.
- Pelletier, G. (1999). L'hospitalisation pour soins de courte durée au Québec: statistiques évolutives 1982-1983 à 1997-1998, [Québec]: Ministère de la santé et des services sociaux, Direction générale de la planification stratégique et de l'évaluation.
- Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics." Journal of Econometrics **33**(3): 311-340.
- Qiu, H., I. T.-s. Yu, L. Tian, X. Wang, L. A. Tse, W. Tam et T. W. Wong (2012). "Effects of coarse particulate matter on emergency hospital admissions for respiratory diseases: a time-series analysis in Hong Kong." Environmental health perspectives **120**(4): 572-576.
- R Core Team (2014). "R: A Language and Environment for Statistical Computing."
- Racine, J. (2000). "Consistent cross-validated model-selection for dependent data: hv-block cross-validation." Journal of Econometrics **99**(1): 39-61.
- Rato, R. T., M. D. Ortigueira et A. G. Batista (2008). "On the HHT, its problems, and some solutions." Mechanical Systems and Signal Processing **22**(6): 1374-1394.
- Ravikumar, P., J. Lafferty, H. Liu et L. Wasserman (2009). "Sparse additive models." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **71**(5): 1009-1030.
- 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. et D. P. Mandic (2011). "Filter Bank Property of Multivariate Empirical Mode Decomposition." Signal Processing, IEEE Transactions on **59**(5): 2421-2426.

- 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.
- Ricciardi, M. J., F. Selzer, O. C. Marroquin, E. M. Holper, L. Venkitachalam, D. O. Williams, S. F. Kelsey et W. K. Laskey (2012). "Incidence and predictors of 30-day hospital readmission rate following percutaneous coronary intervention (from the National Heart, Lung, and Blood Institute Dynamic Registry)." Am J Cardiol.
- Rilling, G., P. Flandrin, P. Gonalves et J. M. Lilly (2007). "Bivariate Empirical Mode Decomposition." Signal Processing Letters, IEEE **14**(12): 936-939.
- 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.
- Ross, J. S., J. Chen, Z. Q. Lin, H. Bueno, J. P. Curtis, P. S. Keenan, S.-L. T. Normand, G. Schreiner, J. A. Spertus, M. T. Vidán, Y. Wang, Y. Wang et H. M. Krumholz (2009). "Recent National Trends in Readmission Rates after Heart Failure Hospitalization." Circulation: Heart Failure.
- Sahay, R. R. et V. Sehgal (2013). "Wavelet regression models for predicting flood stages in rivers: a case study in Eastern India." Journal of Flood Risk Management **6**(2): 146-155.
- Samoli, E., J. Schwartz, B. Wojtyniak, G. Touloumi, C. Spix, F. Balducci, S. Medina, G. Rossi, J. Sunyer et L. Bacharova (2001). "Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA project: a sensitivity analysis for controlling long-term trends and seasonality." Environmental health perspectives **109**(4): 349.
- Sarmiento, S. M., T. G. Verburg, S. M. Almeida, M. C. Freitas et H. T. Wolterbeek (2011). "Robustness of different regression modelling strategies in epidemiology: a time-series analysis of hospital admissions and air pollutants in Lisbon (1999–2004)." Environmetrics **22**(1): 86-97.
- Schultz, S., D. Rothwell, Z. Chen et K. Tu (2013). "Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records." managed-care **10**: 11.
- Schwartz, J. (2000). "The distributed lag between air pollution and daily deaths." Epidemiology **11**(3): 320-326.
- Statistics Canada (2005). Comparability of ICD-10 and ICD-9 for Mortality Statistics in Canada. Statistics Canada.
- Stone, M. (1974). "Cross-Validatory Choice and Assessment of Statistical Predictions." Journal of the Royal Statistical Society. Series B (Methodological) **36**(2): 111-147.
- Stuart, R. A. et G. A. Isaac (1994). "A Comparison of Temperature–Precipitation Relationships from Observations and As Modeled by the General Circulation Model of the Canadian Climate Centre." Journal of Climate **7**(2): 277-282.
- The MathWorks Inc. (2012). "MATLAB, version R2012a."
- Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) **58**(1): 267-288.
- Törő, K., J. Bartholy, R. Pongrácz, Z. Kis, É. Keller et G. Dunay (2010). "Evaluation of meteorological factors on sudden cardiovascular death." Journal of Forensic and Legal Medicine **17**(5): 236-242.
- Tu, J. V., L. Nardi, J. Fang, J. Liu, L. Khalid, H. Johansen et f. t. C. C. O. R. Team (2009). "National trends in rates of death and hospital admissions related to acute myocardial

- infarction, heart failure and stroke, 1994–2004." Canadian Medical Association Journal **180**(13): E118-E125.
- Tu, K., T. Mitiku, D. S. Lee, H. Guo et J. V. Tu (2010). "Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD)." Canadian Journal of Cardiology **26**(7): e225-e228.
- Tu, K., M. Wang, J. Young, D. Green, N. M. Ivers, D. Butt, L. Jaakkimainen et M. K. Kapral (2013). "Validity of Administrative Data for Identifying Patients Who Have Had a Stroke or Transient Ischemic Attack Using EMRALD as a Reference Standard." Canadian Journal of Cardiology **29**(11): 1388-1394.
- Vanos, J. K., S. Cakmak, L. S. Kalkstein et A. Yagouti (2014). "Association of weather and air pollution interactions on daily mortality in 12 Canadian cities." Air Quality, Atmosphere & Health: 1-14.
- Vutcovici, M., M. Goldberg et M.-F. Valois (2013). "Effects of diurnal variations in temperature on non-accidental mortality among the elderly population of Montreal, Québec, 1984–2007." International journal of biometeorology: 1-10.
- Wielgosz, A., M. Arango, C. Bancej, A. Bienek, H. Johansen, P. Lindsay, W. Luo, A. Luteyn, C. Nair, P. Quan, P. Stewart, P. Walsh et G. Webster (2009). Suivi des maladies du coeur et des accidents vasculaires cérébraux au Canada. A. d. l. s. p. d. Canada.
- Wold, S., M. Sjöström et L. Eriksson (2001). "PLS-regression: a basic tool of chemometrics." Chemometrics and Intelligent Laboratory Systems **58**(2): 109-130.
- Wood, S. (2006). Generalized additive models: an introduction with R.
- Wu, T.-Y., H.-C. Hong et Y.-L. Chung (2012). "A looseness identification approach for rotating machinery based on post-processing of ensemble empirical mode decomposition and autoregressive modeling." Journal of Vibration and Control **18**(6): 796-807.
- Wu, W., Y. Xiao, G. Li, W. Zeng, H. Lin, S. Rutherford, Y. Xu, Y. Luo, X. Xu, C. Chu et W. Ma (2013). "Temperature–mortality relationship in four subtropical Chinese cities: A time-series study using a distributed lag non-linear model." Science of The Total Environment **449**(0): 355-362.
- Wu, Z. et N. E. Huang (2004). "A study of the characteristics of white noise using the empirical mode decomposition method." Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **460**(2046): 1597-1611.
- Wu, Z. et N. E. Huang (2009). "Ensemble empirical mode decomposition : A noise-assisted data analysis method." Advances in Adaptive Data Analysis **01**(01): 1-41.
- Wu, Z. et N. E. Huang (2010). "On the Filtering Properties of the Empirical Mode Decomposition." Advances in Adaptive Data Analysis **02**(04): 397-414.
- Wu, Z., N. E. Huang et X. Chen (2009). "The Multi-Dimensional Ensemble Empirical Mode Decomposition Method." Advances in Adaptive Data Analysis **01**(03): 339-372.
- Wu, Z., N. E. Huang, S. R. Long et C.-K. Peng (2007). "On the trend, detrending, and variability of nonlinear and nonstationary time series." Proceedings of the National Academy of Sciences **104**(38): 14889-14894.
- Xie, G., Y. Guo, S. Tong et L. Ma (2014). "Calculate excess mortality during heatwaves using Hilbert-Huang transform algorithm." BMC Medical Research Methodology.
- Yang, A. C., J.-L. Fuh, N. E. Huang, B.-C. Shia, C.-K. Peng et S.-J. Wang (2011a). "Temporal Associations between Weather and Headache: Analysis by Empirical Mode Decomposition." PLoS ONE **6**(1): e14612.

- Yang, A. C., S.-J. Tsai et N. E. Huang (2011b). "Decomposing the association of completed suicide with air pollution, weather, and unemployment data at different time scales." Journal of Affective Disorders **129**(1–3): 275-281.
- Yang, Z. et L. Yang (2009). "A new definition of the intrinsic mode function." World academy of science.
- Ye, X., R. Wolff, W. Yu, P. Vaneckova, X. Pan et S. Tong (2012). "Ambient temperature and morbidity: a review of epidemiological evidence." Environmental health perspectives **120**(1): 19.
- Zhang, J., R. Yan, R. X. Gao et Z. Feng (2010). "Performance enhancement of ensemble empirical mode decomposition." Mechanical Systems and Signal Processing **24**(7): 2104-2123.
- Zhang, X., K. K. Lai et S.-Y. Wang (2008). "A new approach for crude oil price analysis based on Empirical Mode Decomposition." Energy Economics **30**(3): 905-918.
- Zou, H. et T. Hastie (2005). "Regularization and variable selection via the elastic net." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**(2): 301-320.

Tableaux

Tableau 1 : Définitions MCV, incluant leurs codes CIM-9, CIM-10 et de traitements, d'après la définition de MCV par l'ASPC.

MCV	Définitions (un ou l'autre des deux critères)	CIM-9 ²	CIM-10	Traitement		Référence	Sensibilité (%)	Spécificité (%)	VPP (%)	VPN (%)
MCV LES PLUS MORTELLES :										
Maladies coronariennes (MC)	Deux diagnostics ou plus de MC inscrits au fichier des services médicaux rémunérés à l'acte à l'intérieur d'une période de 1 an (365 jours) OU un diagnostic (principal ou secondaire) ou un code de traitement de MC enregistré au fichier MED-ÉCHO (≥ 20 ans).	410-414	I20-I25	CCADTC* 48.02 48.03 48.11- 48.19	CCI† 1.IJ.50 1.IJ.57.GQ 1.IJ.54 1.IJ.76	Tu K et al. Can J Cardiol 2010;26(7) :e225-e228	77.0	97.5	75.3	97.7
Insuffisance cardiaque (IC)	Deux diagnostics ou plus d'IC inscrits au fichier des services médicaux rémunérés à l'acte à l'intérieur d'une période de 1 an (365 jours)	428	I50			Schultz SE et al. Chronic Diseases and Injuries in Canada 33 :3 June 2013	84.8	97.0	56.5	

² La Classification internationale des maladies (CIM) est passée de la neuvième (CIM-9) à la dixième version (CIM-10) au cours des dernières décennies. Au Québec, le passage en CIM-10 est survenu en avril 2006 pour le fichier MED-ÉCHO et en janvier 2000 pour le fichier des décès, alors que le fichier des services médicaux est toujours codifié en CIM-9.

	OU un diagnostic (principal ou secondaire) d'IC enregistré au fichier MED-ÉCHO (≥ 40 ans).								
Maladies vasculaires cérébrales (MCV) et ischémie cérébrale transitoire. (Définition du 29 avril 2013, Agence de la santé publique du Canada)	Deux diagnostics ou plus de MCV inscrits au fichier des services médicaux rémunérés à l'acte à l'intérieur d'une période de 1 an OU un diagnostic (principal ou secondaire) de MCV enregistré au fichier MED-ÉCHO (≥ 20 ans).	362.3 430 431 434.x 435.x 436	G45.x (excluant G45.4) H34.0 H34.1 I60.x I61.x I63.x (excluant I63.6) I64		Tu et al., 2013 The Canadian journal of cardiology, 29:1388-1394	68.0	98.9	66.2	99.0
		Codes dans les services médicaux 436, 432, 435							

Tableau 2 : Données météorologiques brutes. Toutes les séries sont disponibles du 1^{er} janvier 1981 au 31 décembre 2011.

Données	Code / Type	Unités
Température maximale (Tmax)	001 / Journalière	Degrés Celsius (°C)
Température minimale (Tmin)	002 / Journalière	
Température moyenne (Tmoy)	003 / Journalière	
Précipitations totales (Prec)	012 / Journalière	Millimètres (mm)
Épaisseur de neige au sol (Snow)	013 / Journalière	Centimètres (cm)
Pression au niveau de la mer (Pmoy)	073 / Horaire	Kilopascal (kPa)
Humidité relative (Hmin, Hmoy, Hmax)	080 / Horaire	Pourcentages (%)

Tableau 3 : Récapitulatif des méthodes statistiques considérées. R-EMD1 et R-EMD2 sont présentées dans le rapport alors que R-EMDF1, R-EMDF2 et R-EMDF3 sont présentées en Annexes.

	R-EMD1 : météorologie décomposée	R-EMD2 : santé et météorologie décomposées	R-EMDF1 : météorologie filtrée	R-EMDF2 : santé filtrée	R-EMDF3 : santé et météorologie filtrées
<i>Variable réponse(s)</i>	Série sanitaire Y	IMF sanitaires \tilde{C}_k^Y	Série sanitaire Y	Série sanitaire filtrée Y^F	Série sanitaire filtrée Y^F
<i>Variables explicatives</i>	IMF météorologiques $C_k^{X^{(j)}}$	IMF météorologiques $\tilde{C}_k^{X^{(j)}}$	Séries météorologies filtrées $X^{F(j)}$	IMF météorologiques $\tilde{C}_k^{X^{(j)}}$	Séries météorologies filtrées $X^{F(j)}$
<i>Décomposition</i>	EEMD	NA-MEMD	EEMD	EEMD	EEMD
<i>Régression</i>	Lasso				
<i>Évaluation des performances</i>	R^2 , CV par blocs				

Tableau 4 : Récapitulatif du niveau de bruit utilisé pour la décomposition EEMD des séries de la CMM et de la CMQ. Le niveau de bruit reporté est relatif à la variance de la série, p. ex. $\sigma_e = 0,50$ signifie que le bruit ajouté a un écart-type égal à 50% de l'écart-type de la série de données.

Variable	Niveau de bruit σ_e	
	CMM	CMQ
<i>Tmoy</i>	1,00	1,00
<i>Tmin</i>	1,00	1,00
<i>Tmax</i>	0,50	1,00
<i>Prec</i>	0,20	2,00
<i>Snow</i>	0,20	0,50
<i>Pmoy</i>	0,01	0,20
<i>Hmin</i>	0,20	0,50
<i>Hmoy</i>	0,20	2,00
<i>Hmax</i>	2,00	2,00
<i>Décès</i>	0,20	0,20
<i>Hosp</i>	2,00	1,50

Tableau 5 : Tableau résumant les résultats obtenus à l'application de la R-EMD pour examiner la relation entre les MCV et la météorologie. Les résultats de la CMM et de la CMQ sont regroupés, car ils ne présentent pas de différence flagrante.

Variations	Décès	Hospitalisations
<i>Court terme</i> (période < 1mois)	- Hausses pendant précipitations neigeuses en hiver	- Hausse pour évènements chauds et humides l'été
<i>Moyen terme</i> (1 mois < période < 1 an)	- Augmentation hivernale due aux basses températures - Hausse hivernale accentuée par l'humidité automnale	- Augmentation hivernale due aux basses températures - Hausse hivernale accentuée par l'humidité automnale - Hausse hivernale allongée par des chutes de neige automnales et printanières
<i>Long terme</i> (période > 1 an)	- Tendence décroissante due à la hausse des températures et l'augmentation de l'humidité	- Tendence croissante due à la hausse des températures et l'augmentation de l'humidité

Figures

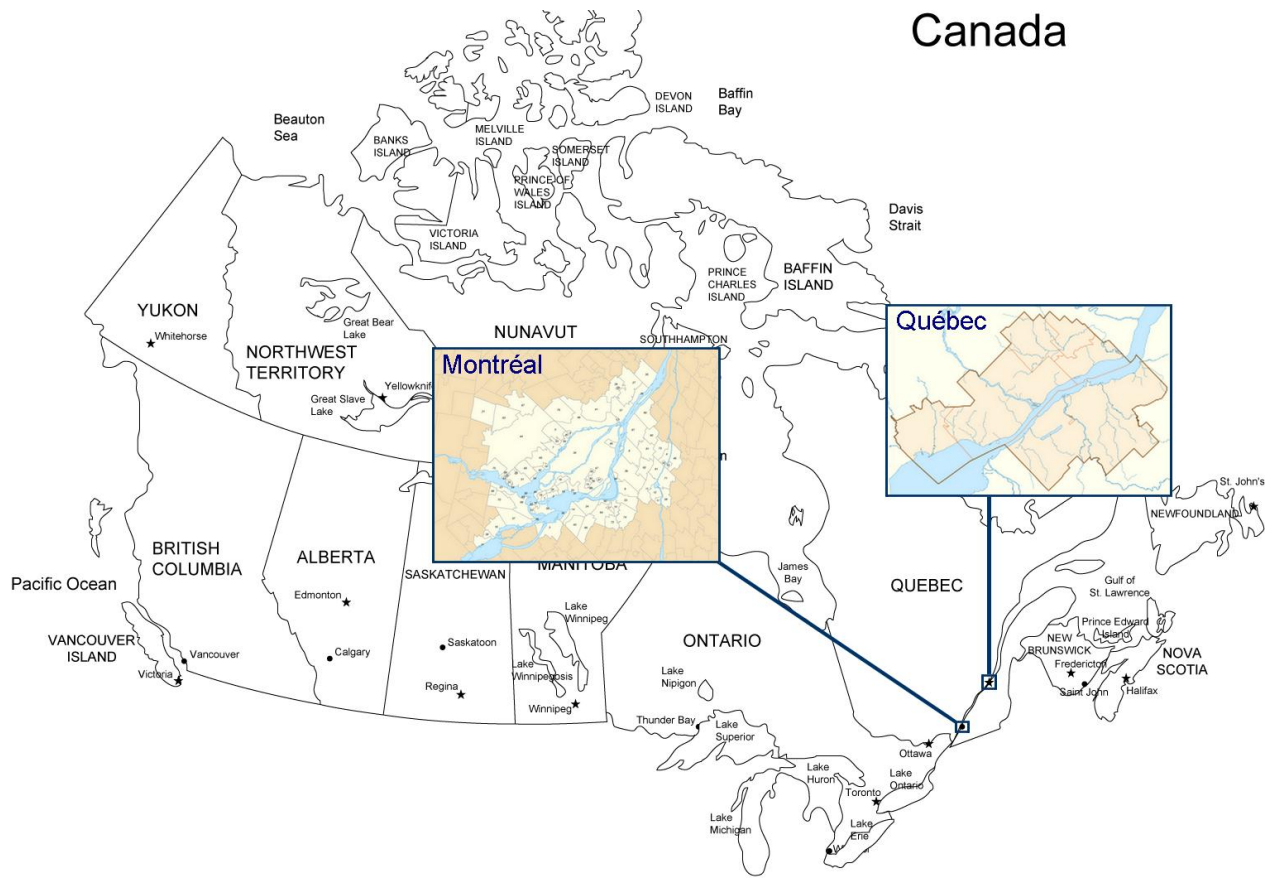


Figure 1 : Carte des communautés métropolitaines de Montréal (CMM) et de Québec (CMQ).

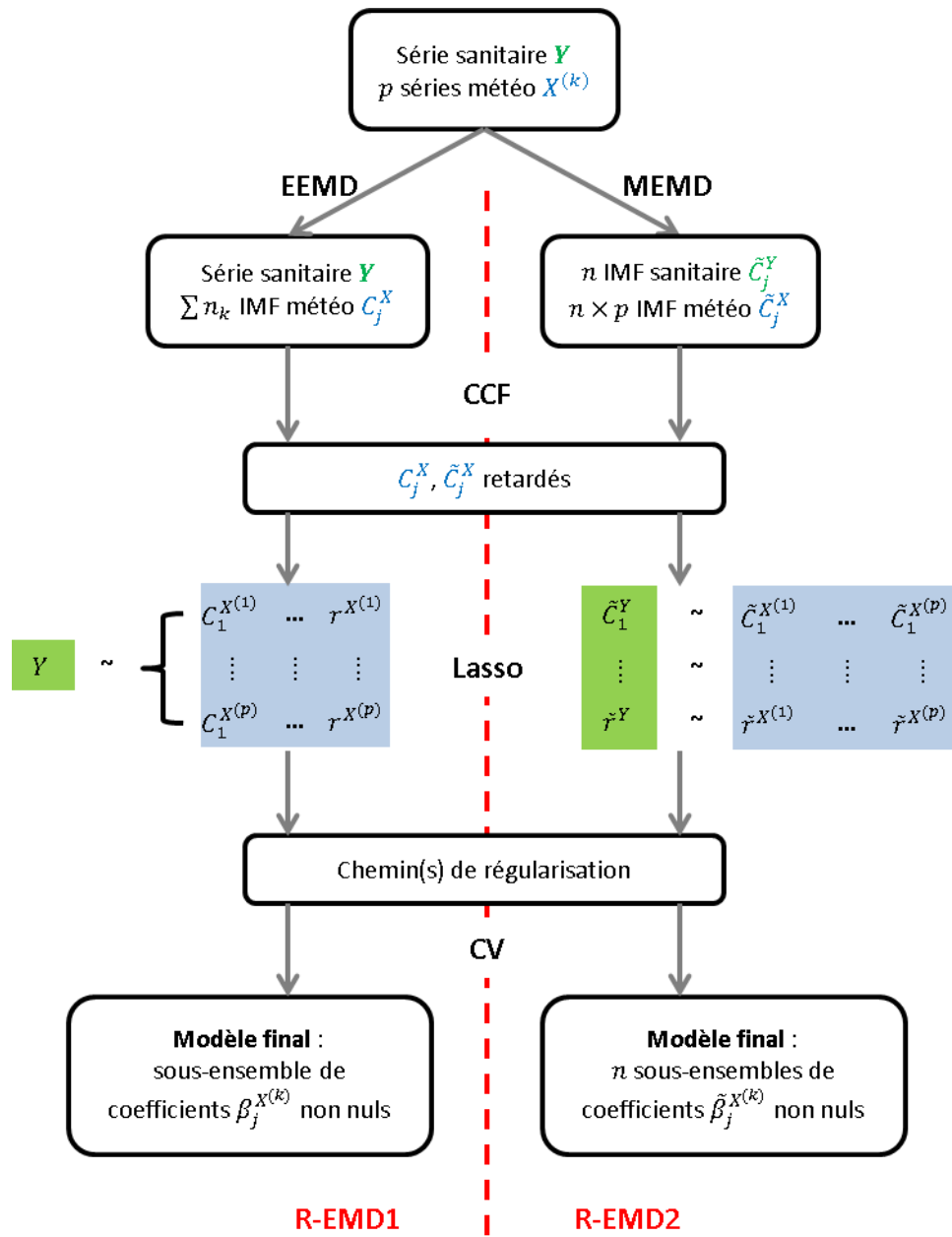


Figure 2 : Résumé de la méthodologie de régression EMD (R-EMD).

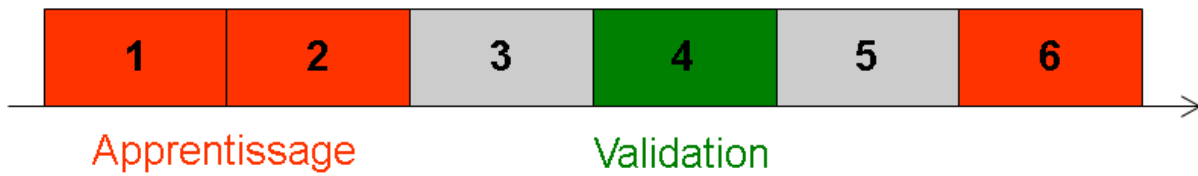


Figure 3 : Illustration de la validation croisée par bloc pour les séries temporelles. Chaque bloc représente une partie de la série. Le bloc vert représente le bloc de validation courant et les blocs rouges les blocs d'apprentissage courant. Les blocs gris ne sont pas utilisés, ni pour l'apprentissage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

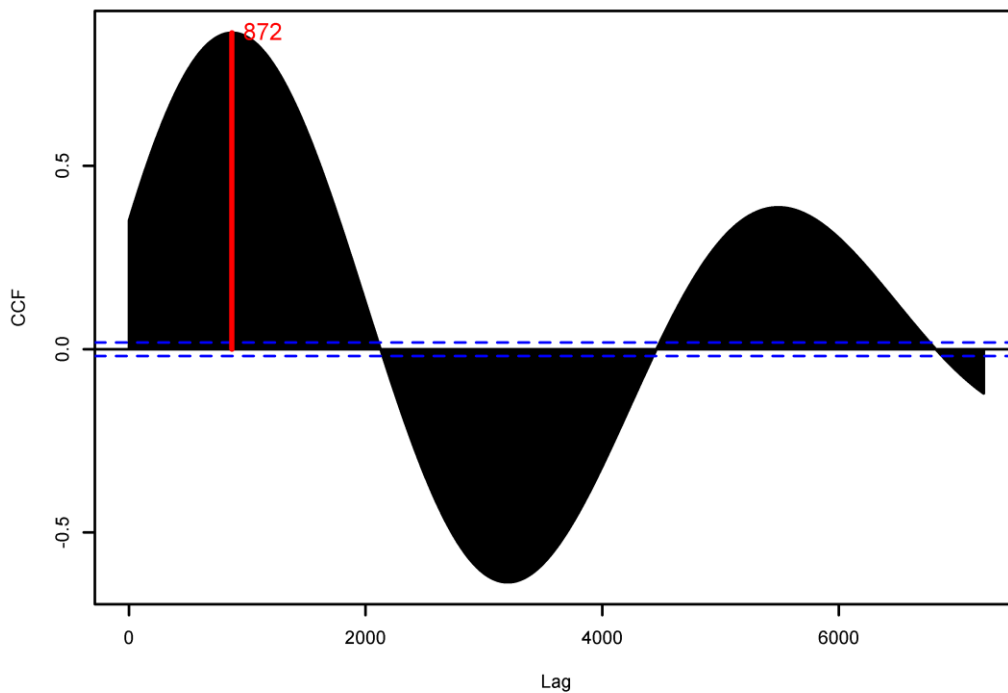


Figure 4 : Illustration de l'oscillation de la CCF entre une série sanitaire et un IMF météorologique. La CCF reportée ici concerne les IMF \tilde{C}_{13}^{Deces} et \tilde{C}_{13}^{Prec} , tous deux de période 13 ans, pour la CMM. Les pointillés bleus indiquent la limite au-delà de laquelle la corrélation est significative au seuil $\alpha = 5\%$. Le premier extremum correspond à un lag de 872 jours soit environ 2 ans et demi.

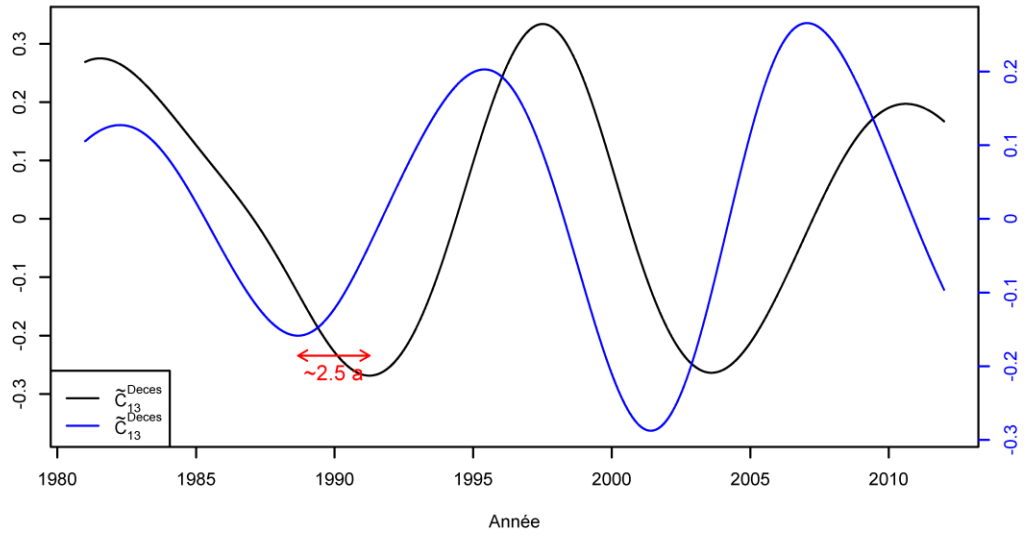


Figure 5 : IMF \tilde{C}_{13}^{Deces} et \tilde{C}_{13}^{Prec} de la CMM afin d'illustrer le lag entre les deux. \tilde{C}_{13}^{Prec} précède \tilde{C}_{13}^{Deces} d'environ 2 ans et demi.

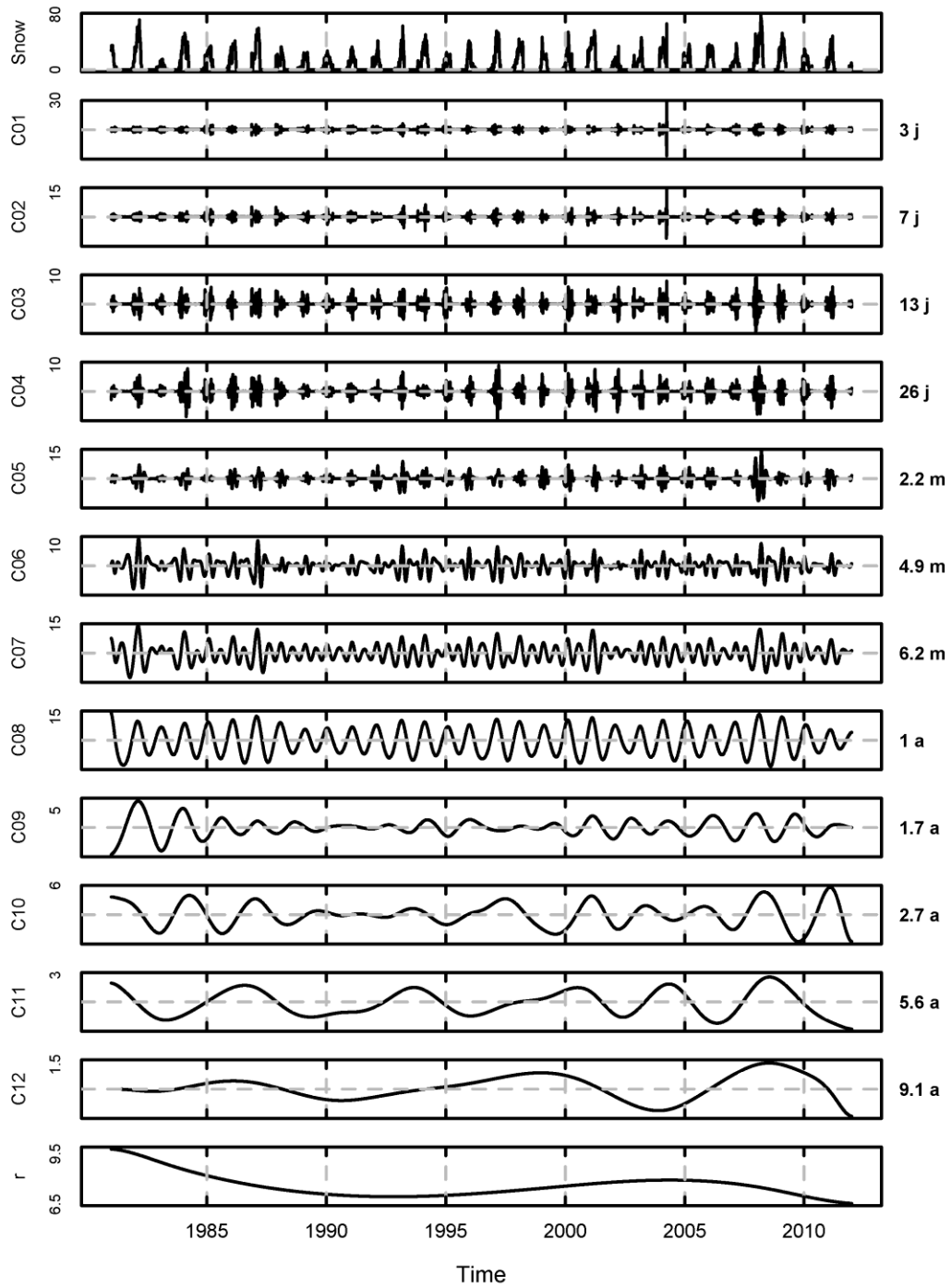


Figure 6 : IMF C_f^{Snow} et tendance r^{Snow} issus de l'algorithme EEMD pour la série de hauteur de neige de la CMM. Les composantes sont rangées par ordre de périodicité ascendante. À gauche de chaque panneau est reporté l'ordre de la composante et la période moyenne de chaque IMF est reportée à droite.

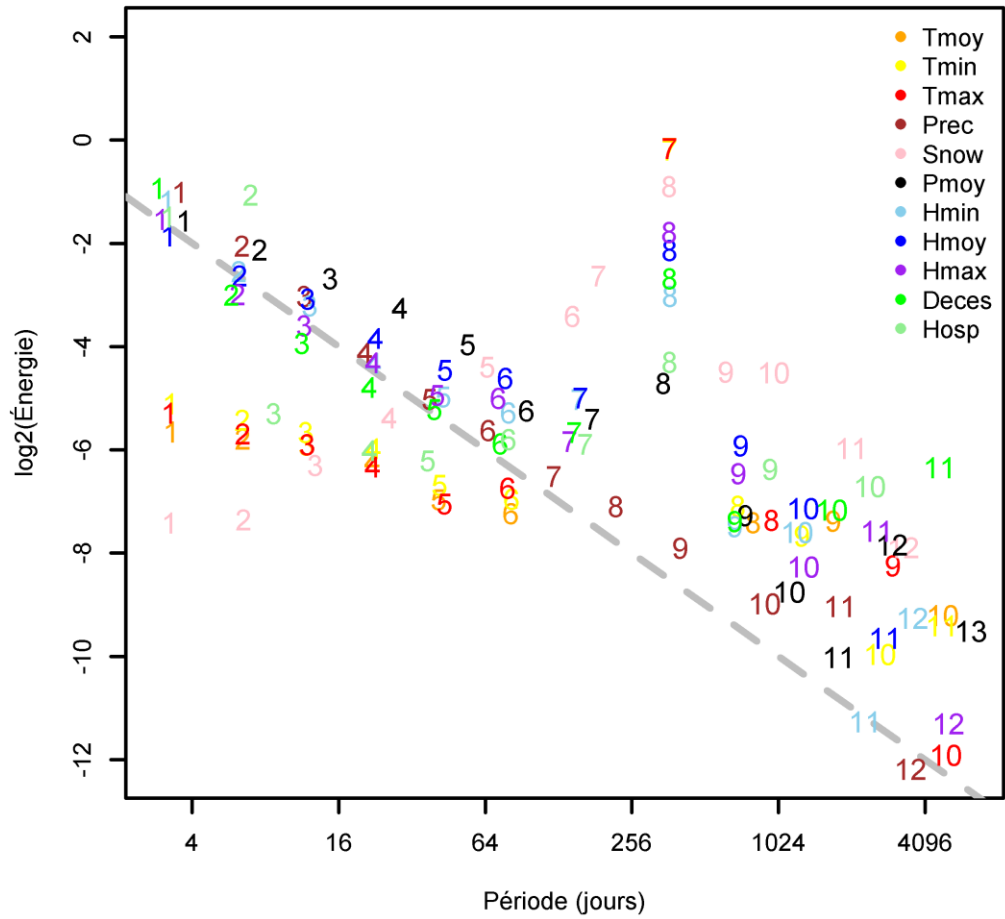


Figure 7 : Graphe reportant la relation entre variance et période moyenne des C_f^x météorologiques et sanitaires de la CMM obtenus par EEMD. Le chiffre reporté correspond à l'ordre de l'IMF et la couleur à la série dont est issu l'IMF. La ligne pointillée grise indique la relation théorique pour un bruit blanc.

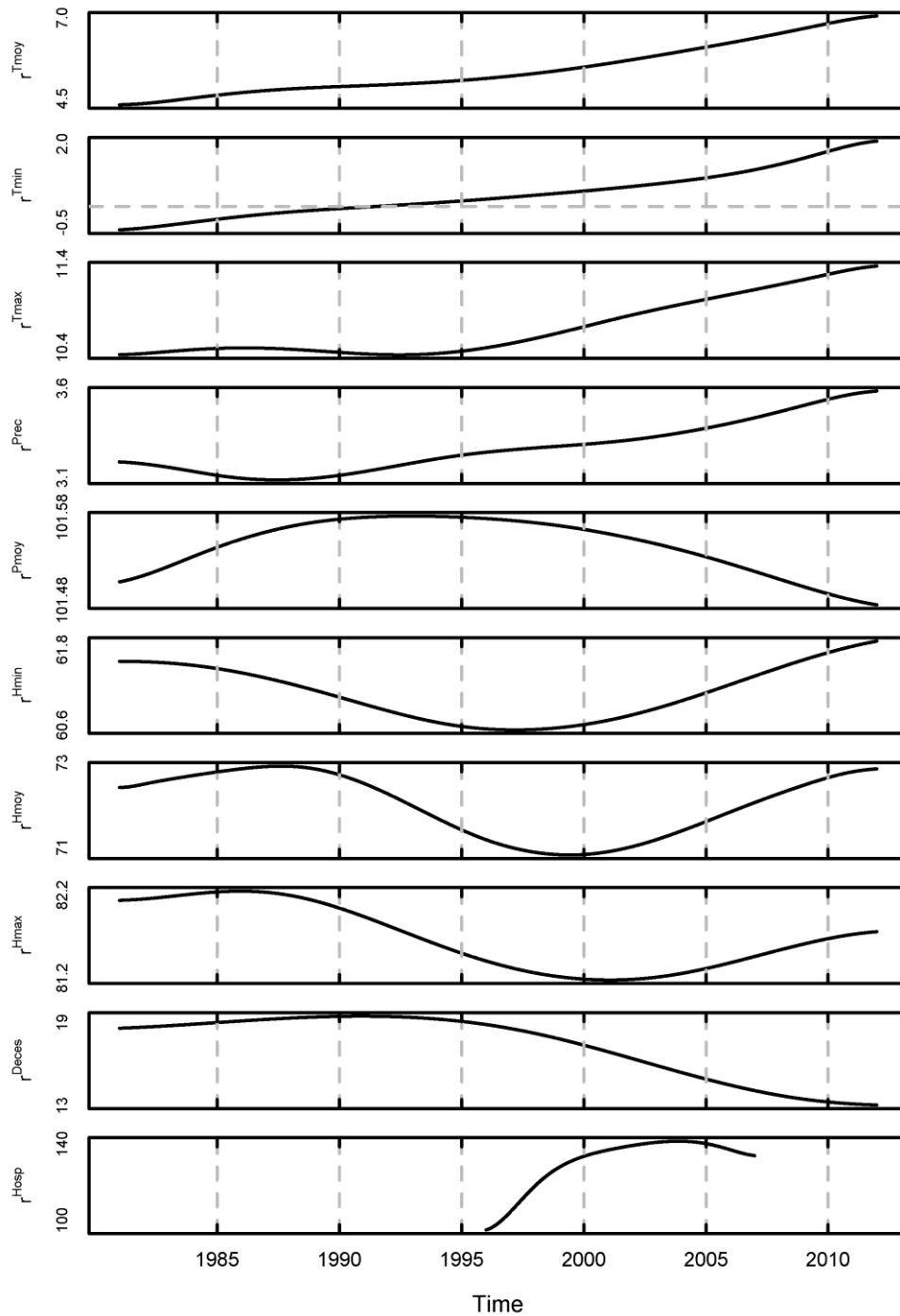


Figure 8 : Tendances r^X obtenues par EEMD des séries météorologiques et sanitaires de la CMM (à l'exception de r^{Snow} , déjà montrée dans la Figure 6). r^{Hosp} est plus courte que les autres à cause du manque de données d'hospitalisations (cf. section 2.1).

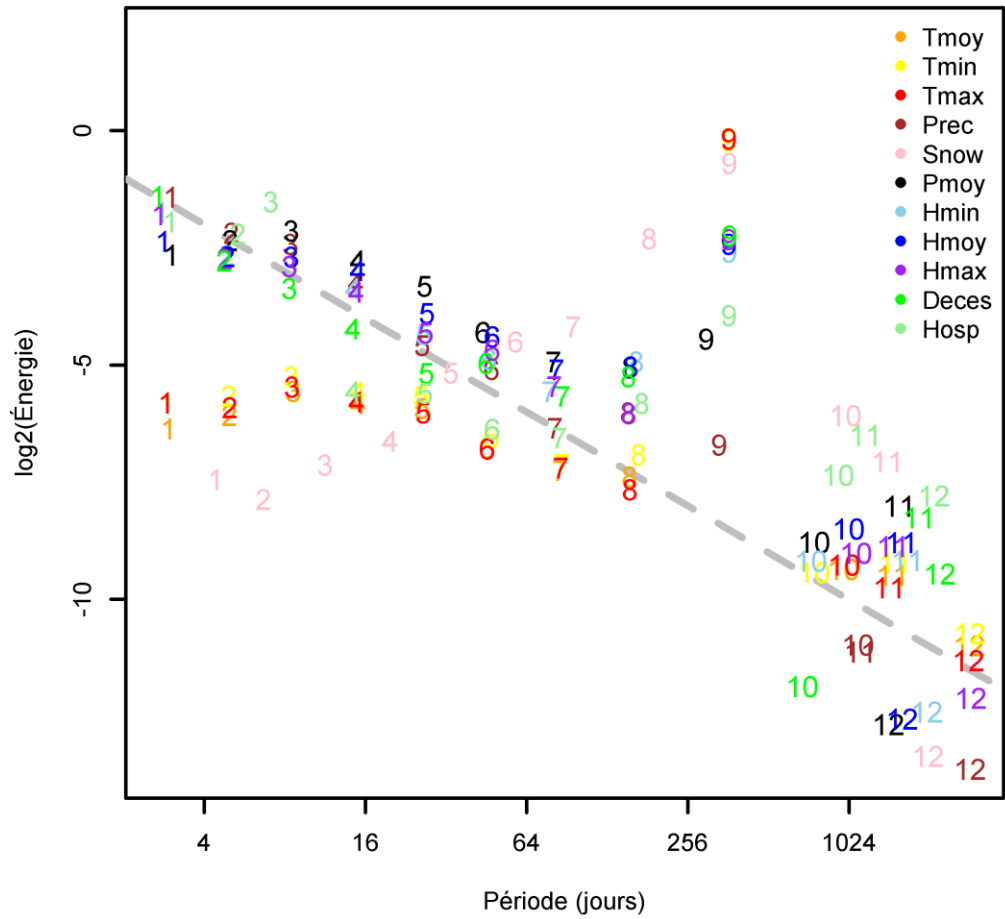


Figure 9 : Graphe reportant la relation entre variance et période moyenne des C_f^x météorologiques et sanitaires de la CMM obtenus par NA-MEMD. Le chiffre reporté correspond à l'ordre de l'IMF et la couleur à la série dont est issu l'IMF. La ligne pointillée grise indique la relation théorique pour un bruit blanc.

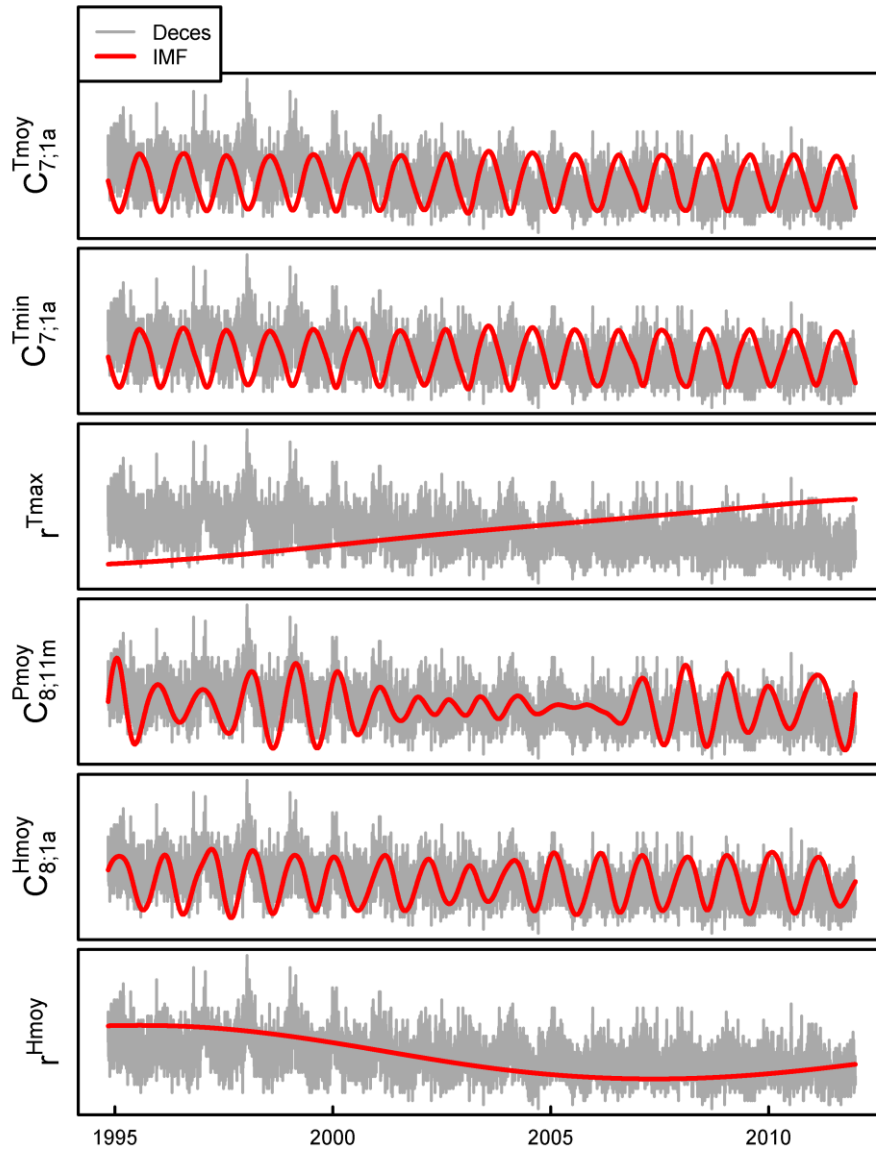


Figure 10 : IMF explicatifs présents dans la R-EMD1 des décès de la CMM (en rouge) et série réponse des décès de la CMM (en gris). Les décès ainsi que les IMF tracés sont standardisés afin de pouvoir les comparer visuellement. Les IMF sont décalés en fonction de leur lag obtenu dans la modélisation afin d’être en phase avec les variations qu’ils expliquent.

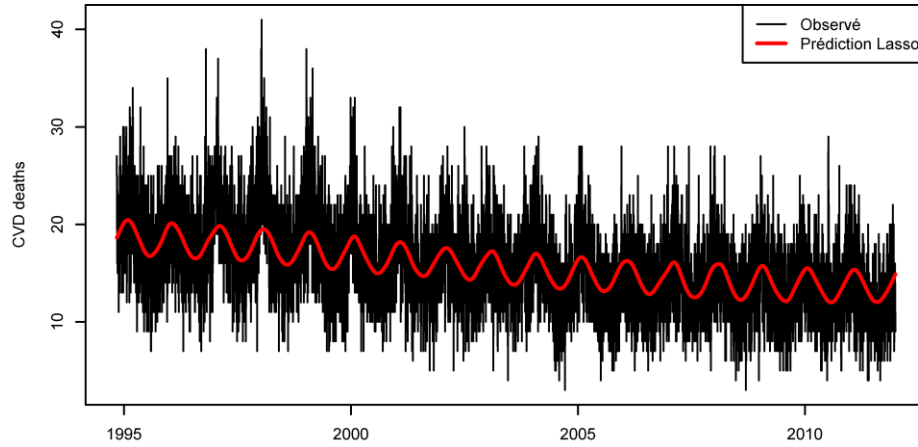


Figure 11 : Série des décès de la CMM prédite par le Lasso de la R-EMD1. La tendance à la baisse et le cycle annuel sont retranscrits par cette modélisation.

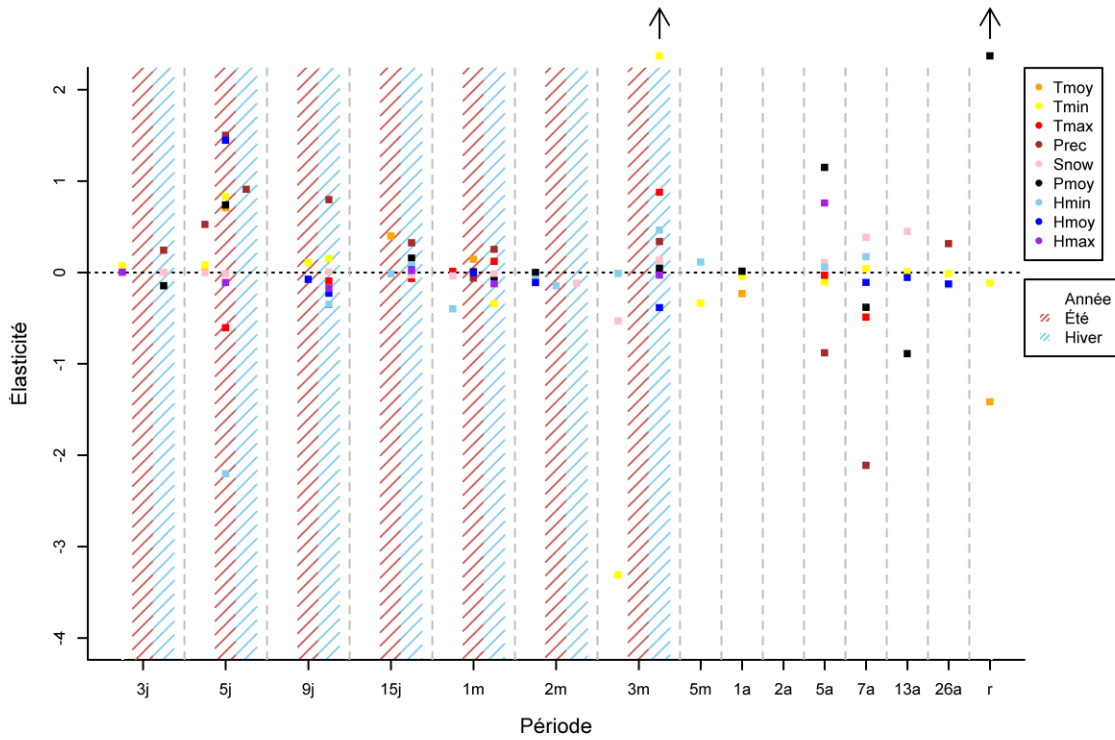


Figure 12 : Graphique récapitulatif des modèles ajusté pour la R-EMD2 des décès de la CMM. L'abscisse correspond à la période moyenne de la réponse (chaque marque correspond à un modèle) et l'ordonnée correspond à l'élasticité (4) des IMF $\tilde{C}_{j,f}^x$ présents dans le modèle. Les points associés aux $\tilde{C}_{j,f}^x$ issus d'une variable météorologique ont tous la même couleur (p. ex. jaune pour $Tmin$). Les hachures rouges et bleues indiquent respectivement les modèles estivaux et hivernaux.

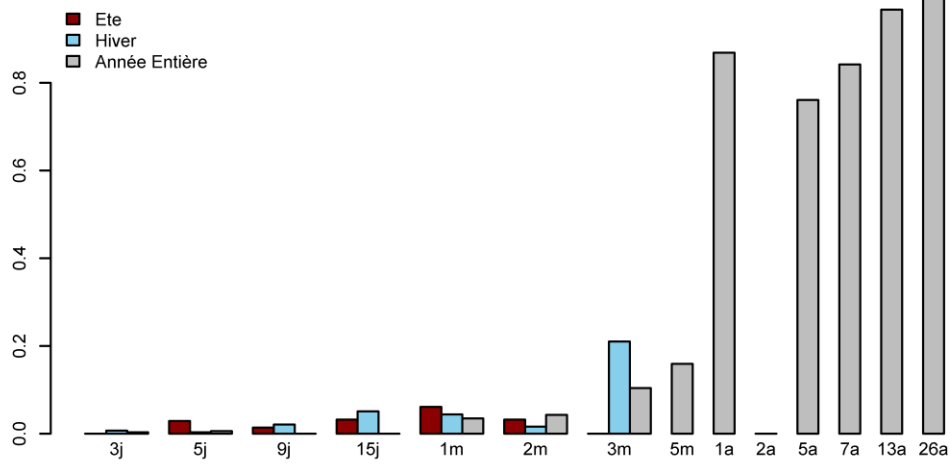


Figure 13 : Variance expliquée R^2 pour chaque modèle de la R-EMD2 des décès de la CMM. À chaque échelle, la barre rouge indique le modèle estival, la barre bleue le modèle hivernal et la barre grise le modèle toutes saisons confondues. L'augmentation du R^2 avec la périodicité découle du fait que la quantité de variance à expliquer décroît lorsque la périodicité augmente.

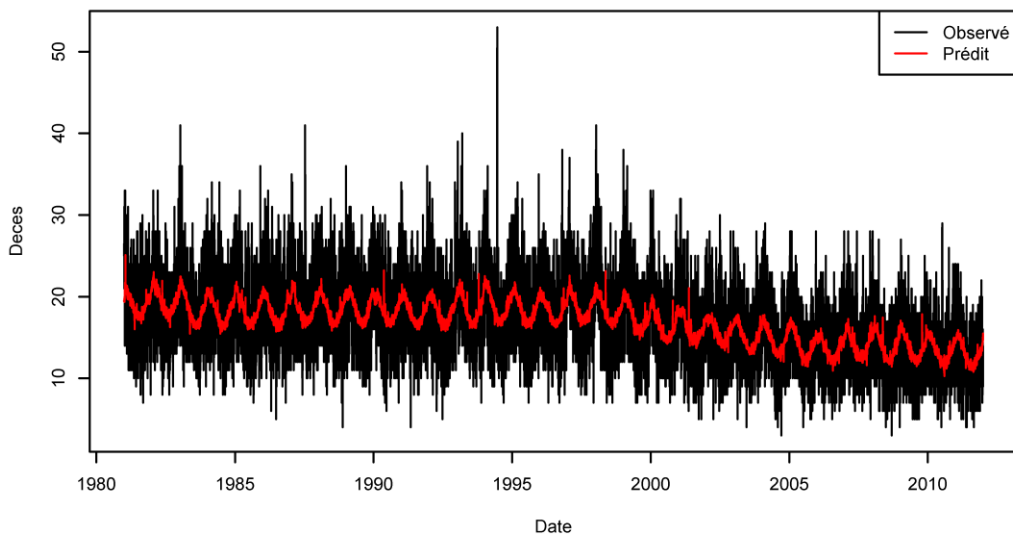


Figure 14 : Série des décès prédite par sommation des modèles individuels de la R-EMD2.

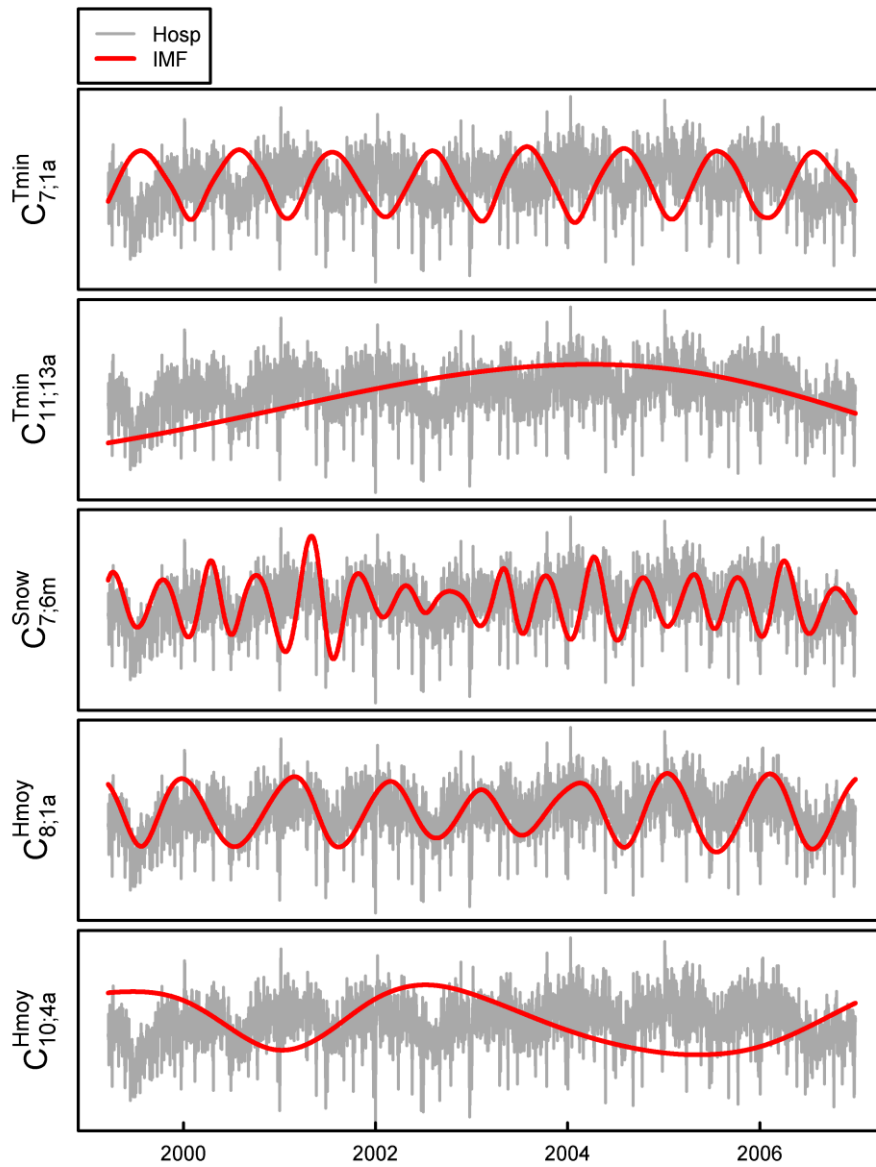


Figure 15 : IMF explicatifs présents dans la R-EMD1 des hospitalisations de la CMM (en rouge) et série réponse des hospitalisations de la CMM (en gris). Les hospitalisations ainsi que les IMF tracés sont standardisés afin de pouvoir les comparer visuellement. Les IMF sont décalés en fonction de leur lag obtenu dans la modélisation afin d'être en phase avec les variations qu'ils expliquent.

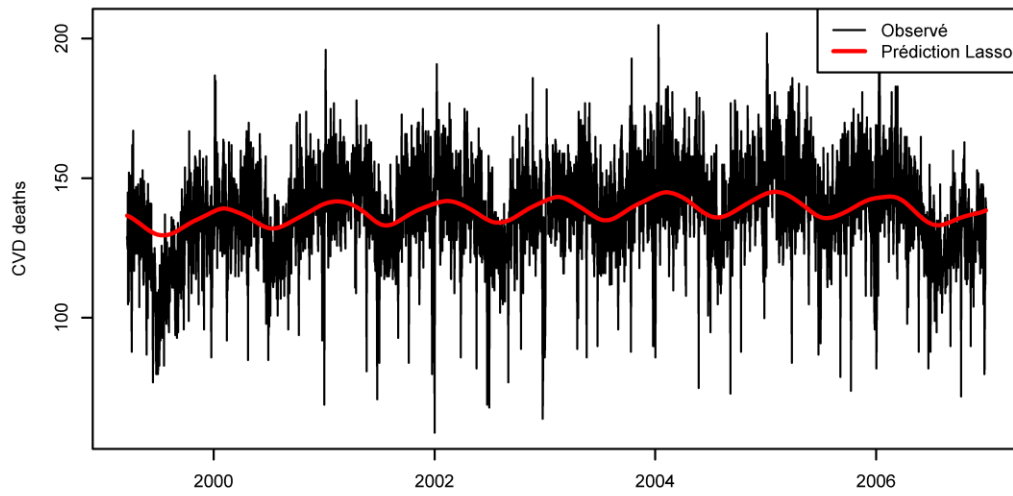


Figure 16 : Série des hospitalisations de la CMM corrigée par rapport à la tendance et à l'effet fin de semaine (en noir) et prédiction par la R-EMD1 (en rouge).

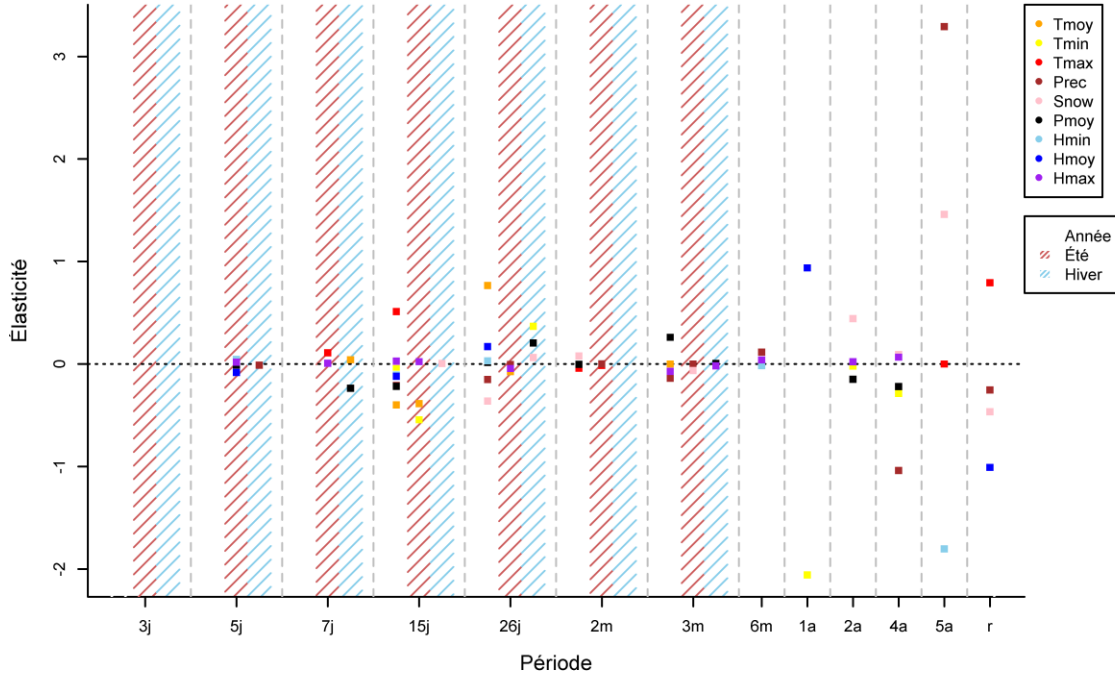


Figure 17 : Graphique récapitulatif des 14 modèles ajustés pour la R-EMD2 des hospitalisations de la CMM. L'abscisse correspond à la période moyenne de la réponse (chaque marque correspond à un modèle) et l'ordonnée correspond à l'élasticité (4) des IMF $\tilde{C}_{j,f}^x$ présents dans le modèle. Les points associés aux $\tilde{C}_{j,f}^x$ issus d'une variable météorologique ont tous la même couleur (p. ex. jaune pour $Tmin$). Les hachures rouges et bleues correspondent respectivement aux modèles estivaux et hivernaux.

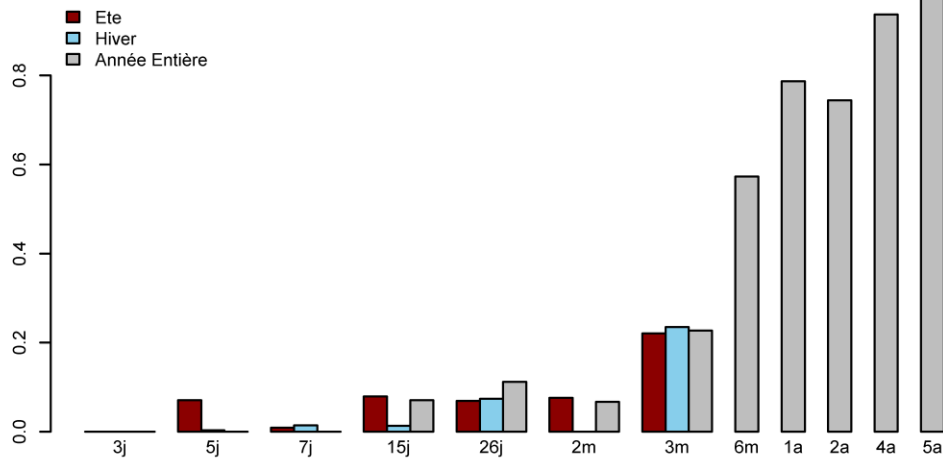


Figure 18 : Variance expliquée R^2 pour chaque modèle de la R-EMD2 des hospitalisations de la CMM. À chaque échelle, la barre rouge indique le modèle estival, la barre bleue le modèle hivernal et la barre grise le modèle toutes saisons confondues. L'augmentation du R^2 avec la périodicité découle du fait que la quantité de variance à expliquer décroît lorsque la périodicité augmente.

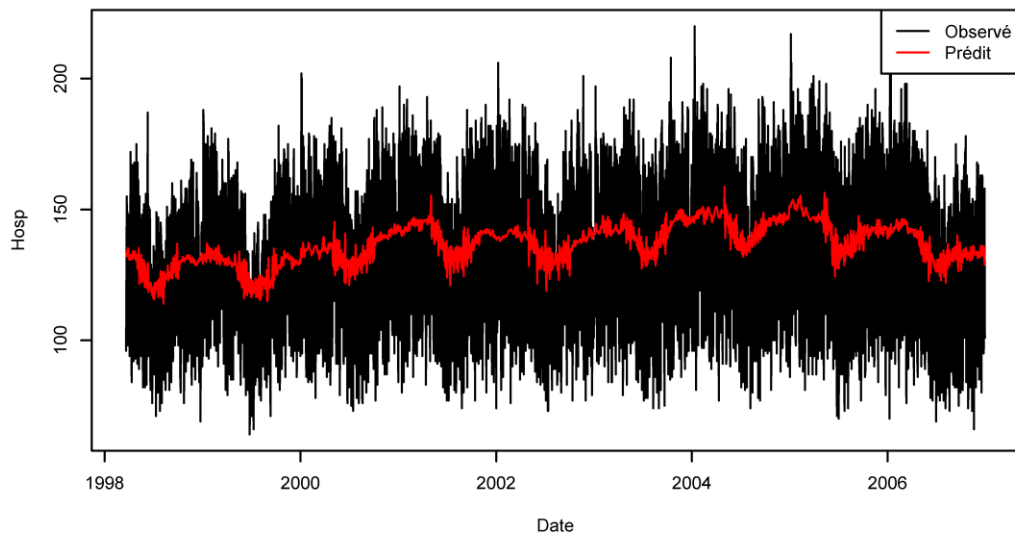


Figure 19 : Série des hospitalisations de la CMM (en noir) avec la prédiction faite par la R-EMD2 (en rouge).

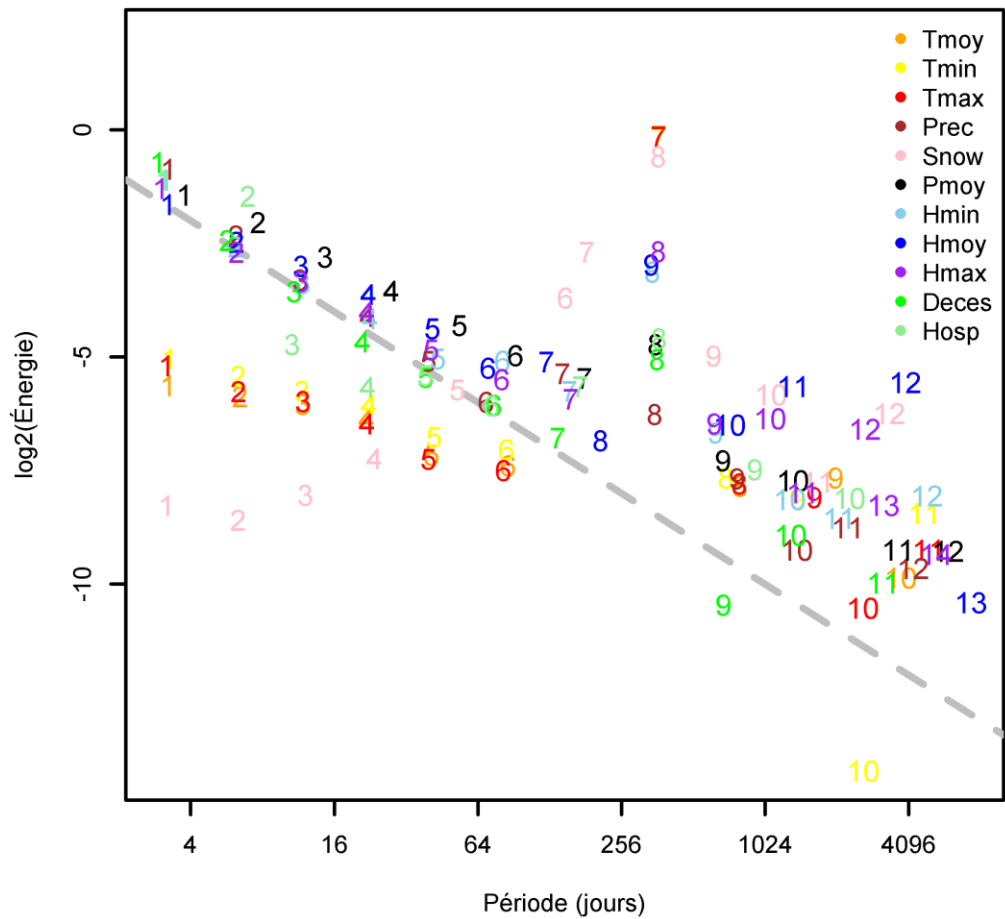


Figure 20 : Graphe reportant la variance des $C_{i,f}^x$ de la CMQ obtenus par EEMD en fonction de leur période moyenne. Chaque variable météorologique est représentée par une couleur et le chiffre correspond à l'ordre de l'IMF. La ligne pointillée grise indique la relation théorique pour un bruit blanc.

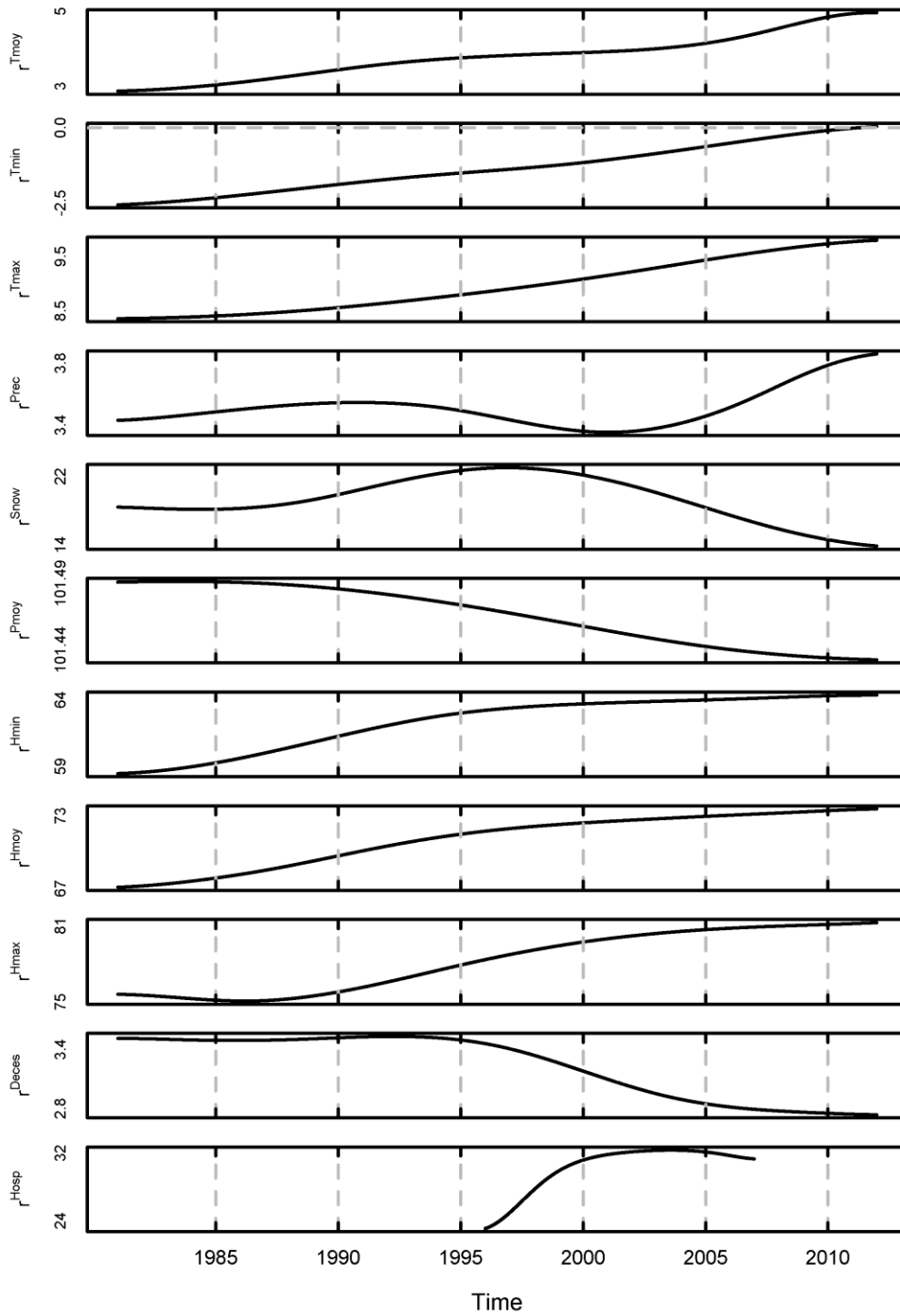


Figure 21 : Tendances r^X obtenues par EEMD des séries météorologiques et sanitaires de la CMQ. r^{Hosp} est plus courte que les autres à cause du manque de données d'hospitalisations (cf. section 2.1).

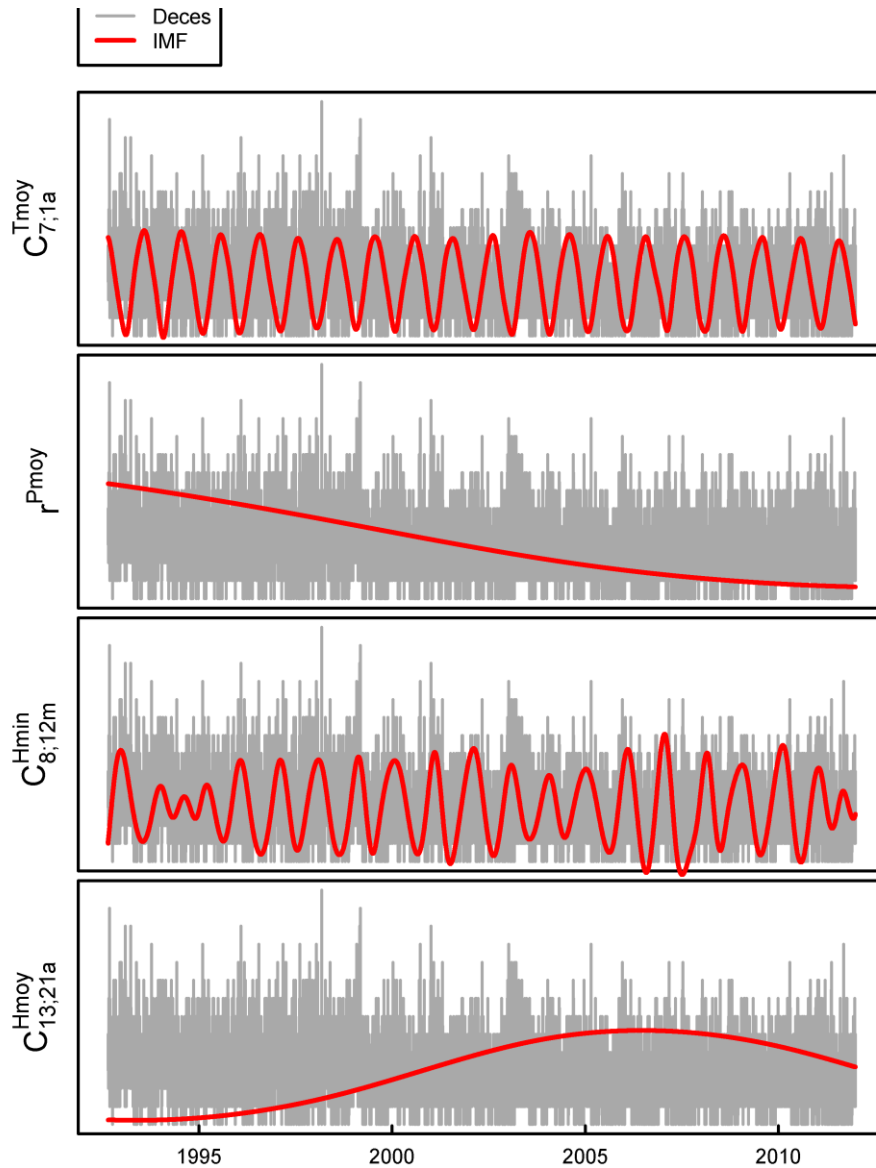


Figure 22 : IMF explicatifs présents dans la R-EMD1 des décès de la CMQ (en rouge) et série réponse des décès de la CMQ (en gris). Les décès ainsi que les IMF tracés sont standardisés afin de pouvoir les comparer visuellement. Les IMF sont décalés en fonction de leur lag obtenu dans la modélisation afin d’être en phase avec les variations qu’ils expliquent.

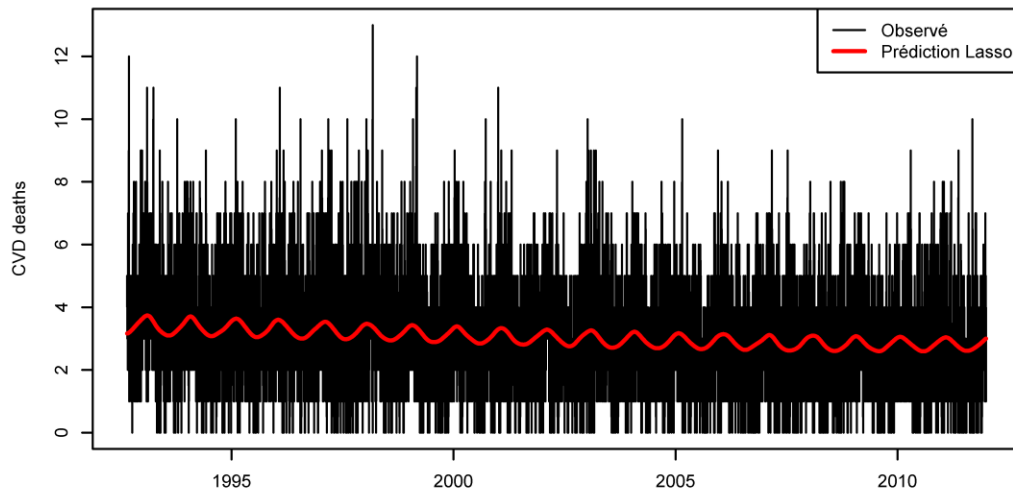


Figure 23 : Séries des décès de la CMQ avec la prédiction effectuée par la R-EMD1.

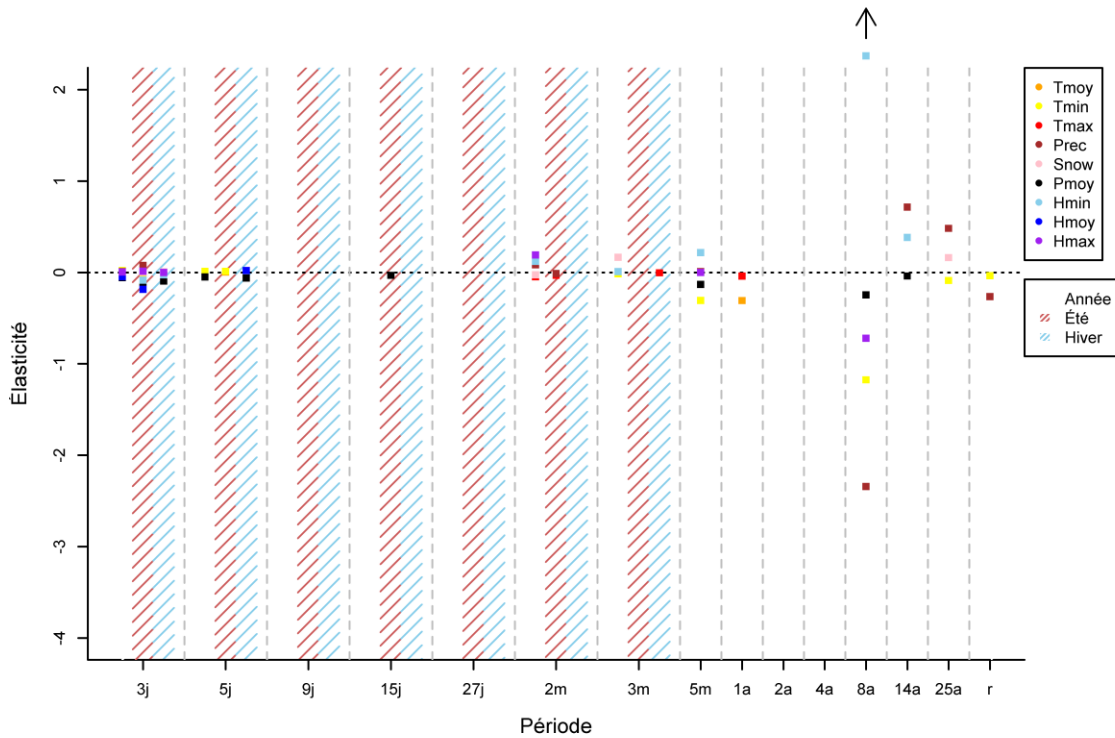


Figure 24 : Graphique récapitulatif des 14 modèles ajustés pour la R-EMD2 des décès de la CMQ. L'abscisse correspond à la période moyenne de la réponse (chaque marque correspond à un modèle) et l'ordonnée correspond à l'élasticité (4) des IMF $\tilde{C}_{j,f}^x$ présents dans le modèle. Les points associés aux $\tilde{C}_{j,f}^x$ issus d'une variable météorologique ont tous la même couleur (p. ex. jaune pour *Tmin*). Les hachures rouges et bleues indiquent respectivement les modèles estivaux et hivernaux.

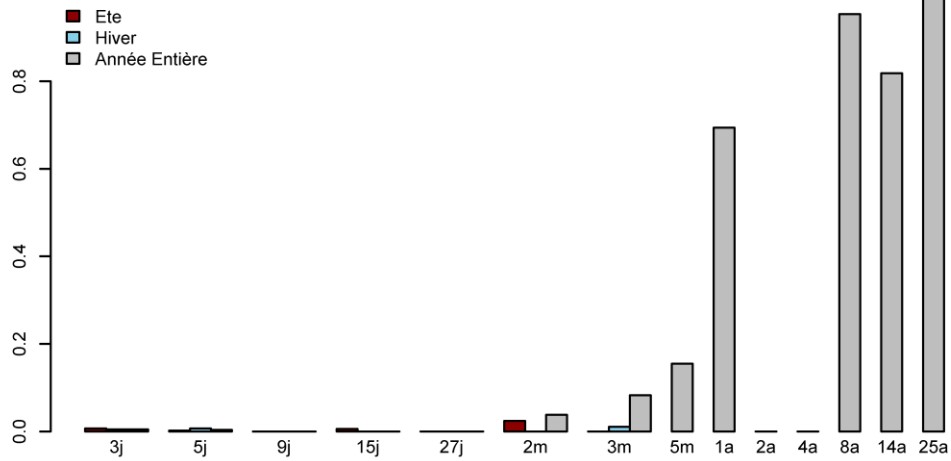


Figure 25 : Variance expliquée R^2 pour chaque modèle de la R-EMD2 des décès de la CMQ. À chaque échelle, la barre rouge indique le modèle estival, la barre bleue le modèle hivernal et la barre grise le modèle toutes saisons confondues. L'augmentation du R^2 avec la périodicité découle du fait que la quantité de variance à expliquer décroît lorsque la périodicité augmente.

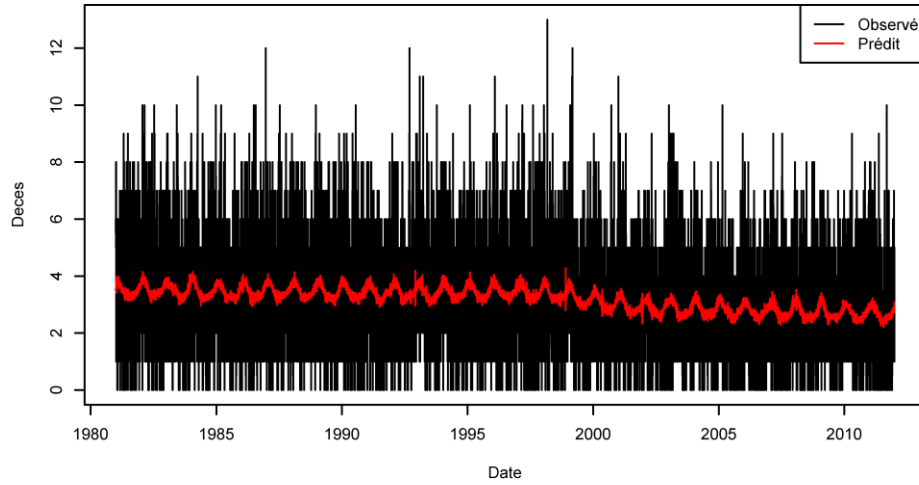


Figure 26 : Série des décès de la CMQ (en noir) avec la prédiction faite par la R-EMD2 (en rouge). La prédiction est réalisée en faisant la somme de tous les modèles obtenus.

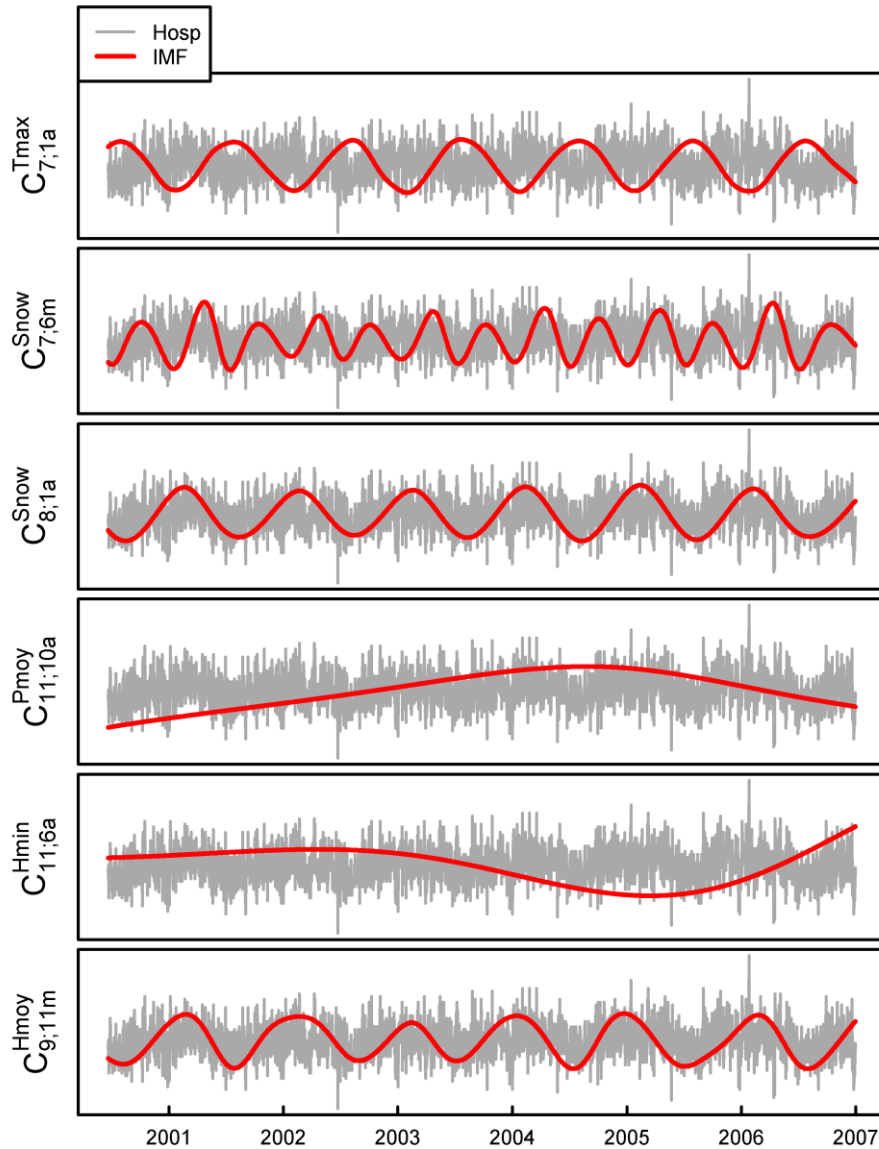


Figure 27 : IMF explicatifs présents dans la R-EMD1 des hospitalisations de la CMQ (en rouge) et série réponse des hospitalisations de la CMQ (en gris). Les hospitalisations ainsi que les IMF tracés sont standardisés afin de pouvoir les comparer visuellement. Les IMF sont décalés en fonction de leur lag obtenu dans la modélisation afin d'être en phase avec les variations qu'ils expliquent

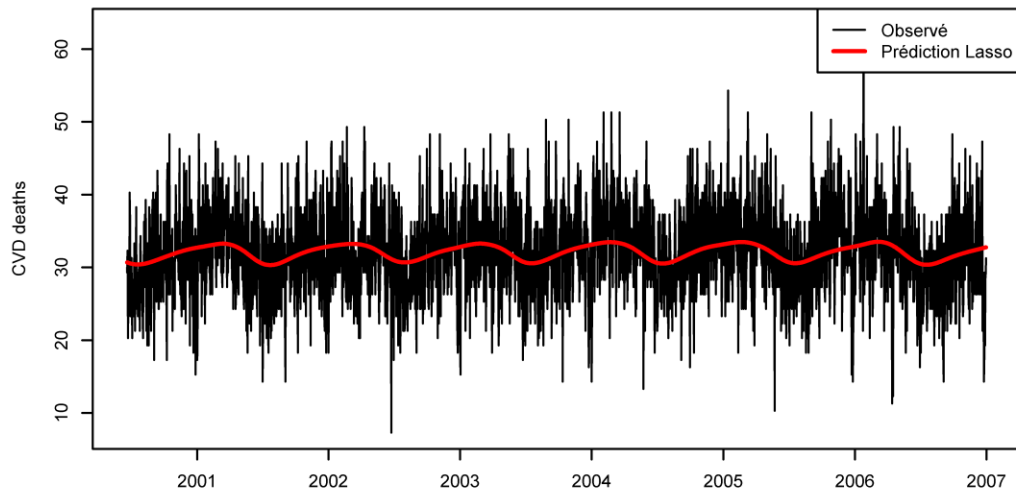


Figure 28 : Hospitalisations de la CMQ dont l'effet « fin de semaine » a été corrigé (en noir) et prédiction du Lasso de la R-EMD1 (en rouge).

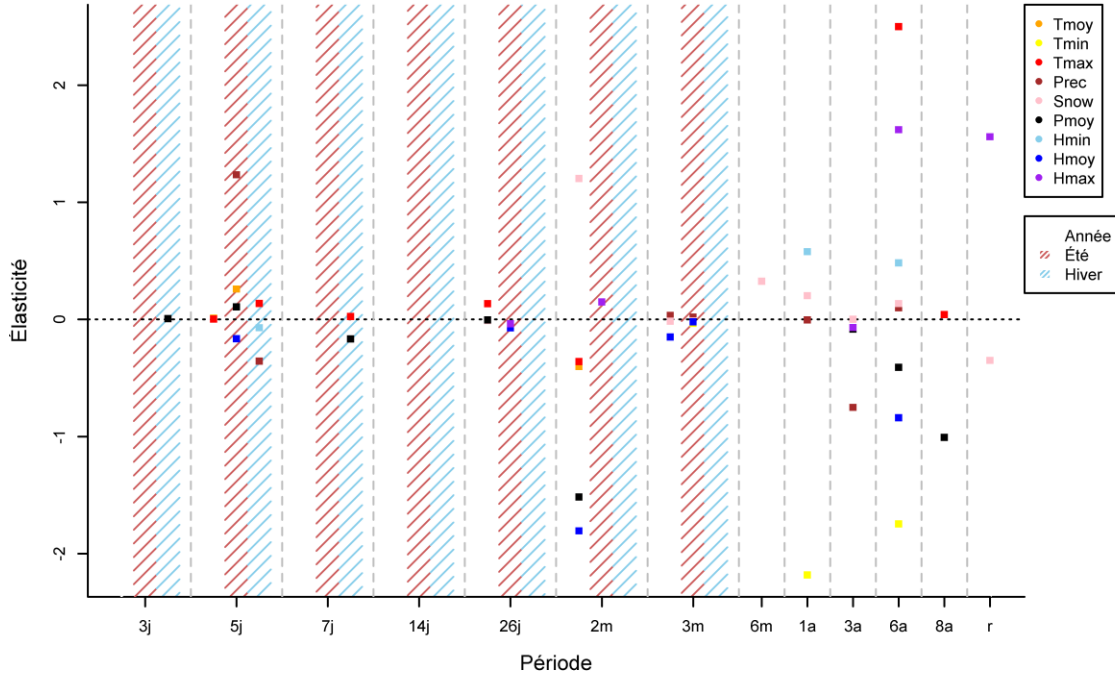


Figure 29 : Graphique récapitulatif des 12 modèles ajustés pour la R-EMD2 des hospitalisations de la CMQ. L'abscisse correspond à la période moyenne de la réponse (chaque marque correspond à un modèle) et l'ordonnée correspond à l'élasticité (4) des IMF $\tilde{C}_{j,f}^x$ présents dans le modèle. Les points associés aux $\tilde{C}_{j,f}^x$ issus d'une variable météorologique ont tous la même couleur (p. ex. jaune pour $Tmin$). Les hachures rouges et bleues indiquent respectivement les modèles estivaux et hivernaux.

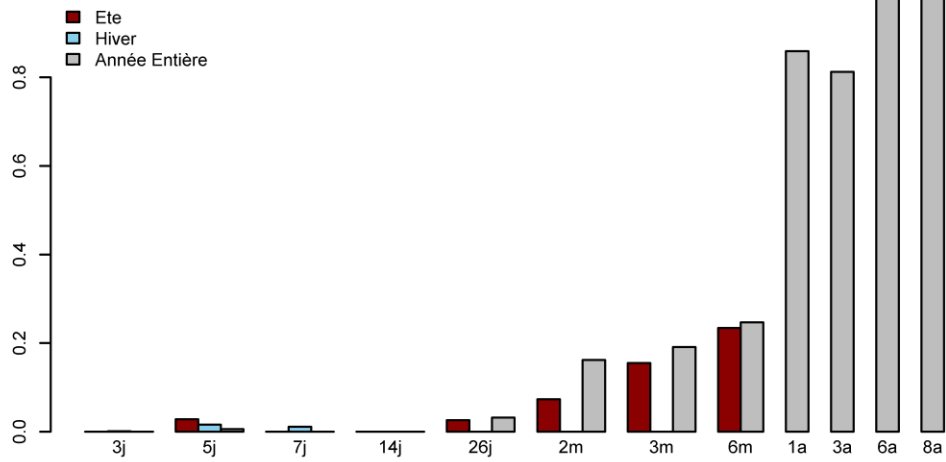


Figure 30 : Variance expliquée R^2 pour chaque modèle de la R-EMD2 des hospitalisations de la CMQ. À chaque échelle, la barre rouge indique le modèle estival, la barre bleue le modèle hivernal et la barre grise le modèle toutes saisons confondues. L'augmentation du R^2 avec la périodicité découle du fait que la quantité de variance à expliquer décroît lorsque la périodicité augmente.

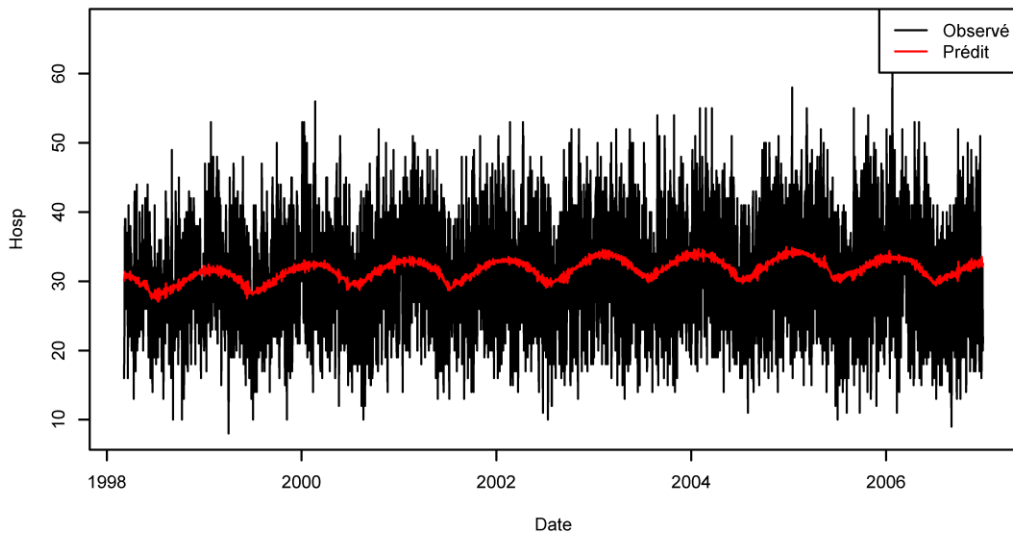


Figure 31 : Série des hospitalisations de la CMQ (en noir) avec la prédiction faite par la R-EMD2 (en rouge).

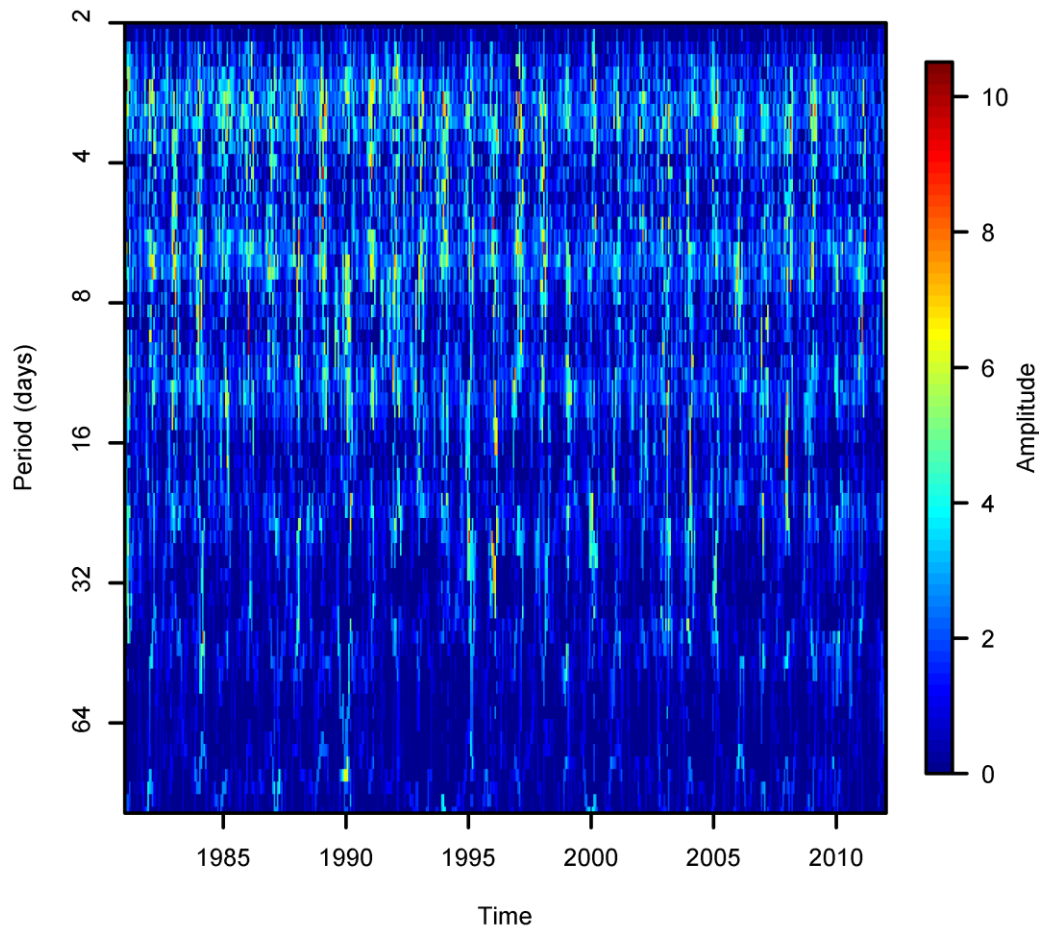
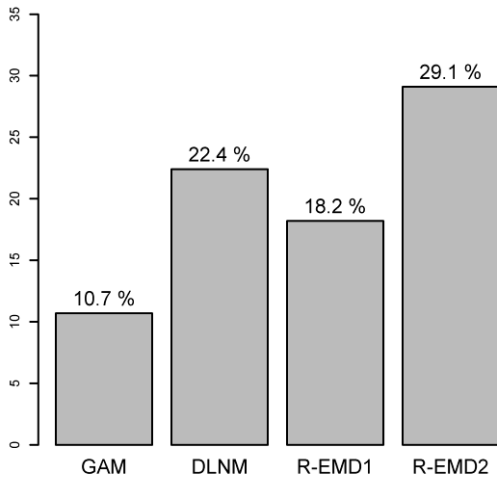
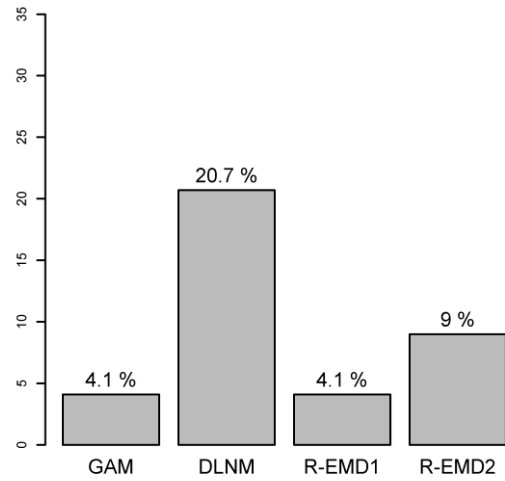


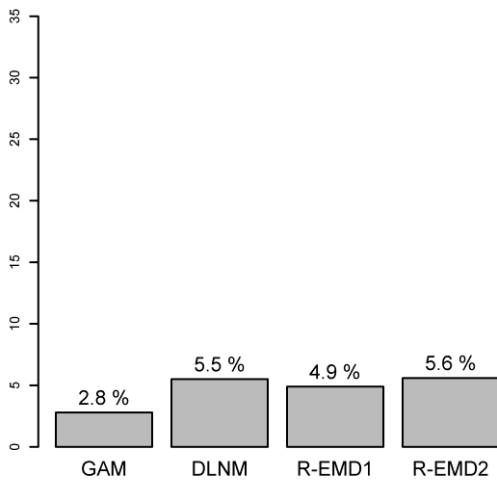
Figure 32 : Transformation Hilbert-Huang (HHT) de la série des températures minimales de la CMQ. La HHT est basée sur les IMF de la série et s'interprète similairement à un spectre d'ondelettes. Plus la couleur s'éloigne du bleu et plus la fréquence représentée en ordonnée a une grande amplitude pour la période représentée en abscisse. Seules les périodicités inférieures à 100 jours sont montrées ici.



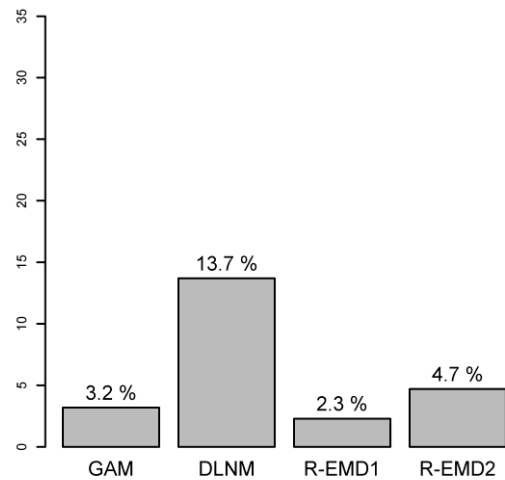
a) Décès de la CMM



b) Hospitalisations de la CMM

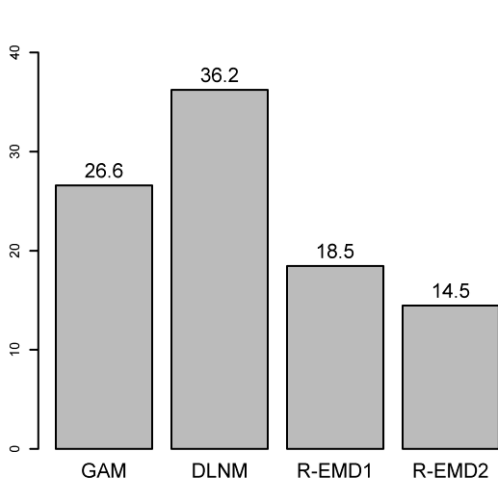


c) Décès de la CMQ

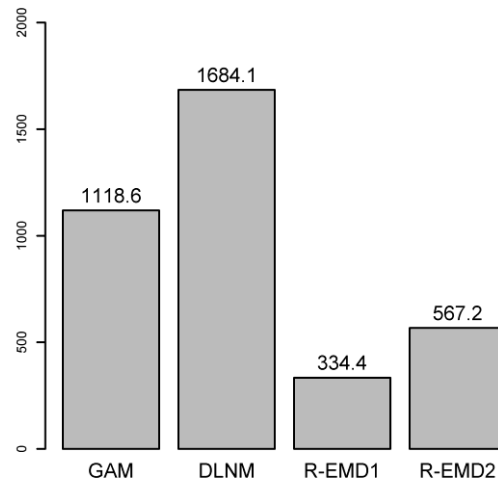


d) Hospitalisations de la CMQ

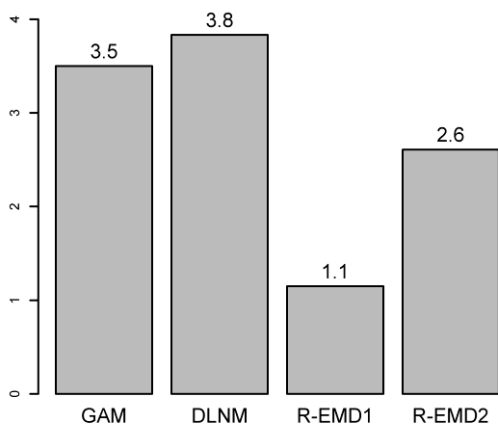
Figure 33 : Valeurs de variance expliquée R² pour les modélisations par modèles additifs généralisés (GAM), modèles non linéaires à effets retardés distribués (DLNM), modélisation avec météorologie décomposée (R-EMD1) et modélisation avec santé et météorologie décomposées (R-EMD2) de la régression EMD.



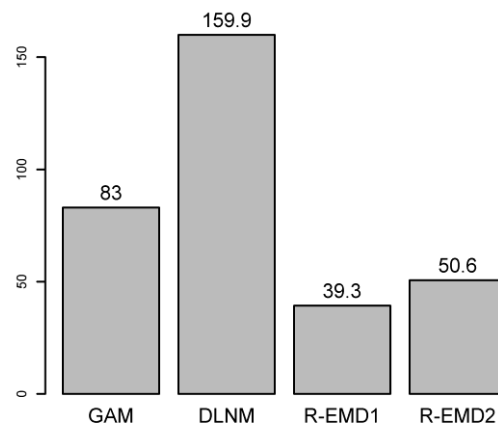
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ



d) Hospitalisations de la CMQ

Figure 34 : Valeurs d'erreur de prédiction estimée par validation croisée (CV) pour les modélisations par modèles additifs généralisés (GAM), modèles non linéaires à effets retardés distribués (DLNM), modélisation avec météorologie décomposée (R-EMD1) et modélisation avec santé et météorologie décomposées (R-EMD2) de la régression EMD.

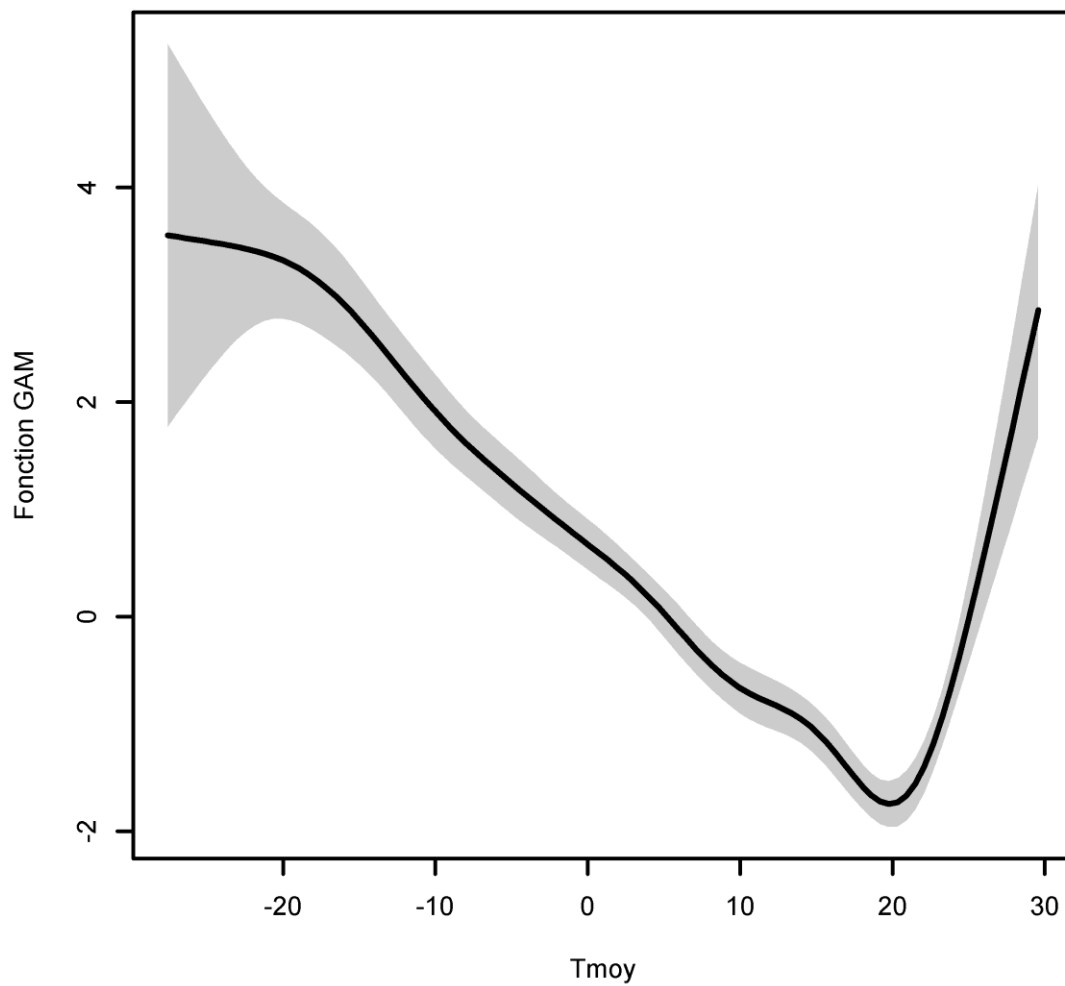


Figure 35 : Fonction de régression estimée par GAM montrant la relation non linéaire entre la température moyenne et les décès de la CMM. La région grise indique l'intervalle de confiance à 95%.

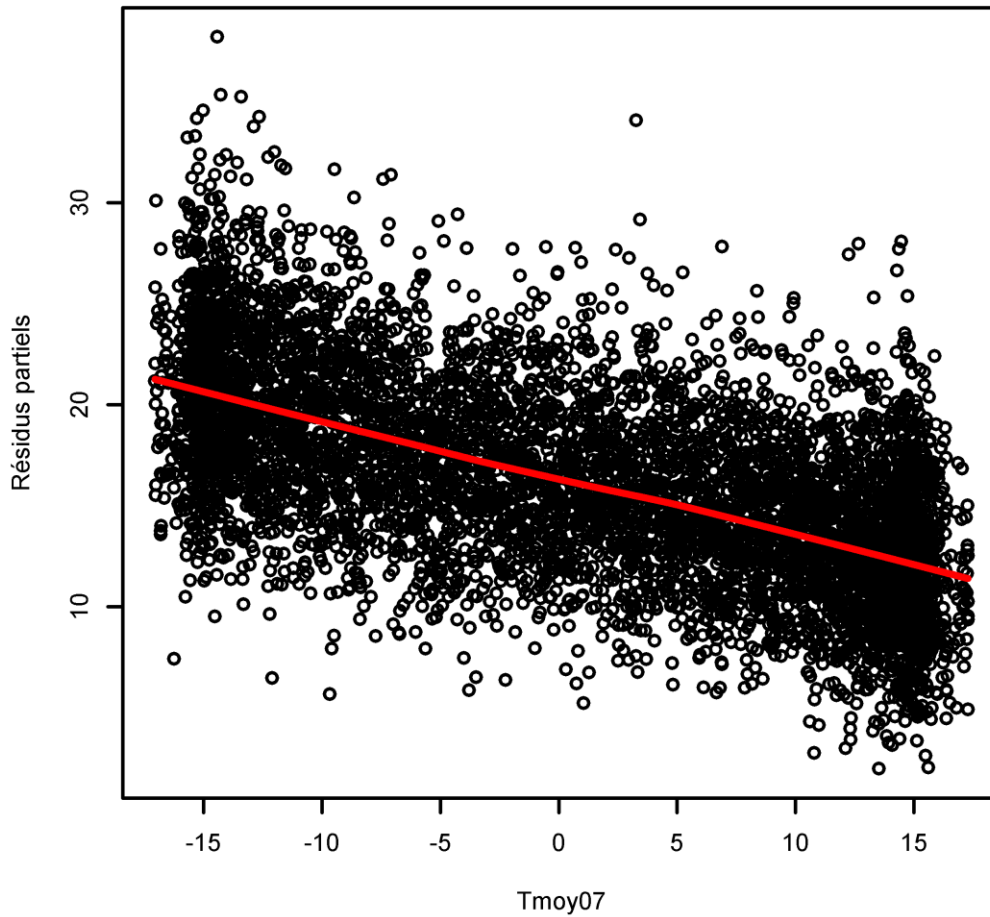


Figure 36 : Résidus partiels associés à $C_{7;1a}^{Tmoy}$ dans la R-EMD1 des décès de la CMM, servant à vérifier l'hypothèse de linéarité dans la régression. La ligne rouge est un ajustement par régression locale (loess, Cleveland et Devlin 1988) des résidus partiels.

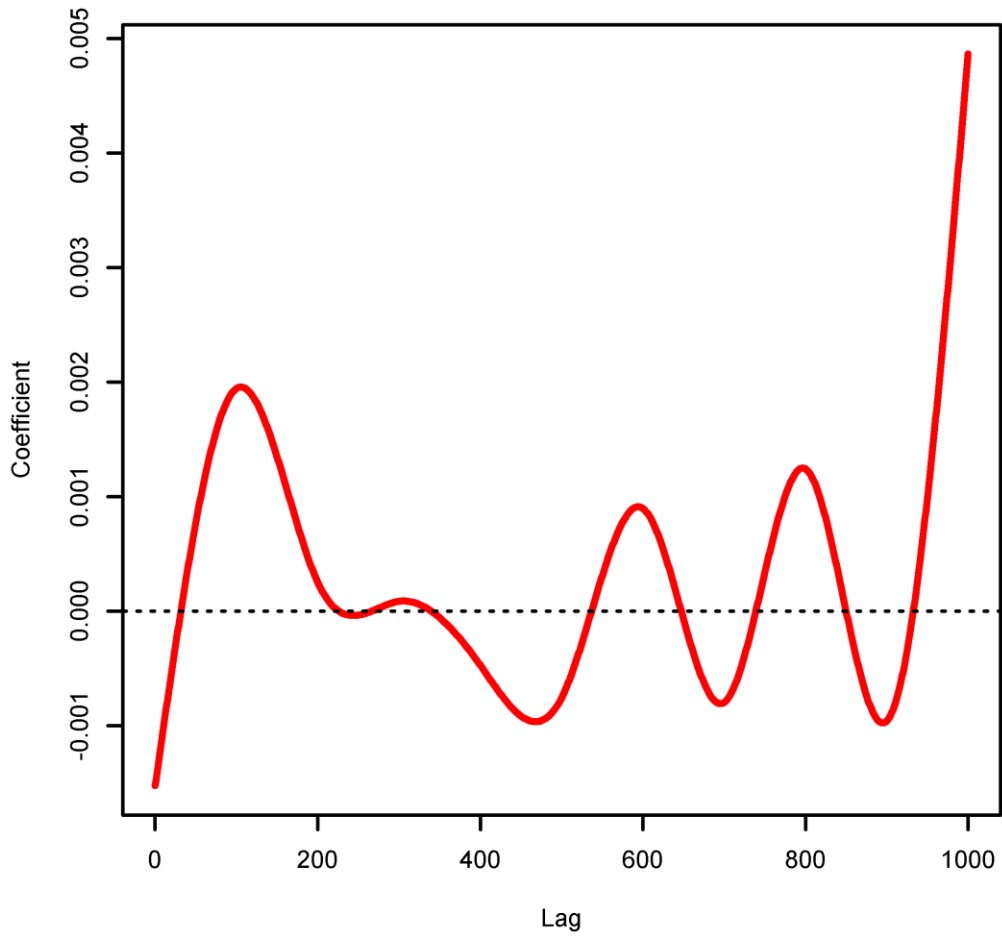


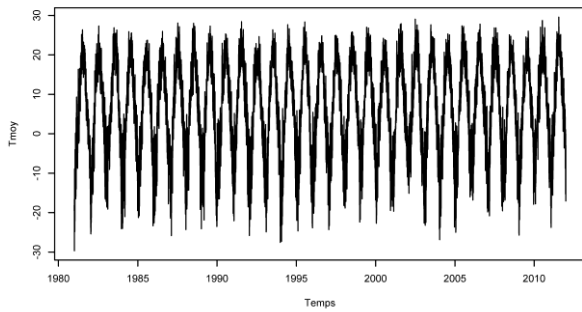
Figure 37 : Fonctions DLM entre l'IMF $C_{7;1a}^{Tmoy}$ et les décès de la CMM.

Annexes

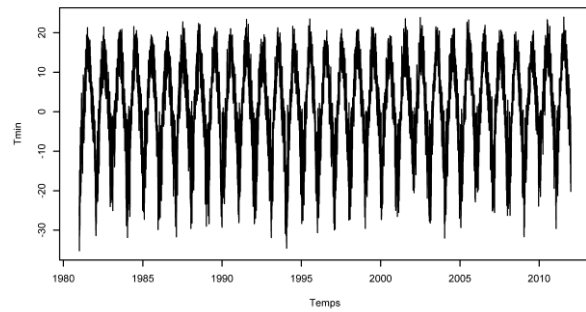
Les annexes recensent l'ensemble des figures non montrées directement dans le corps du rapport.

A. Description des données

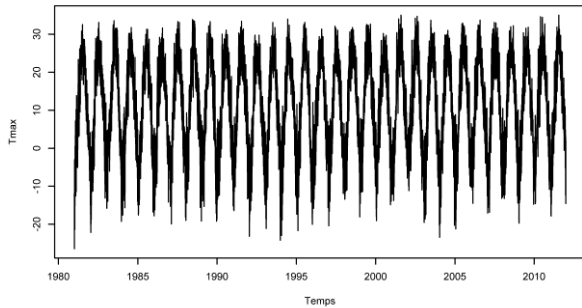
Cette partie de l'annexe montre les données sanitaires et météorologiques avec des statistiques descriptives de base.



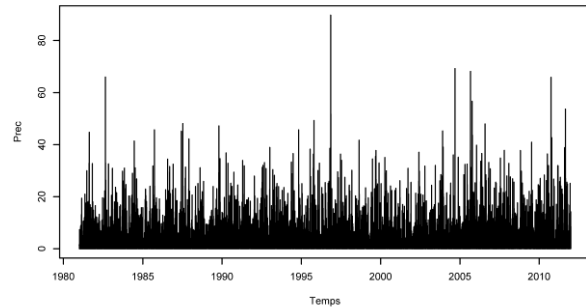
a) Tmoy



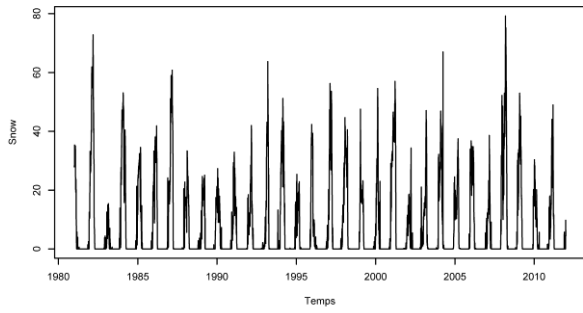
b) Tmin



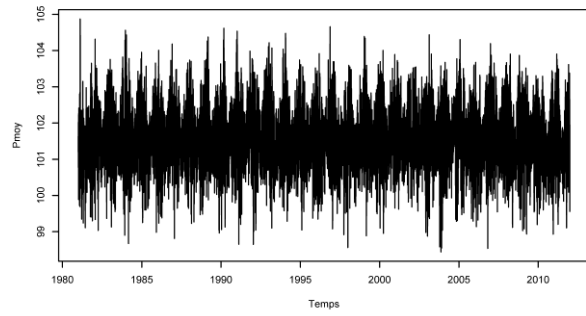
c) Tmax



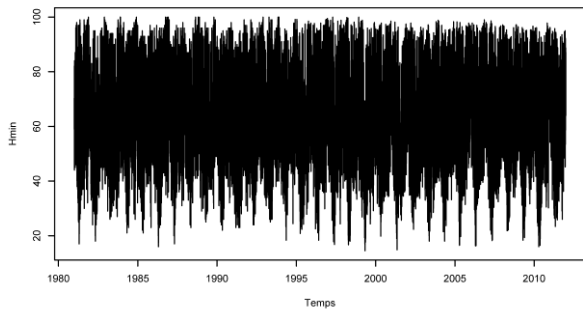
d) Prec



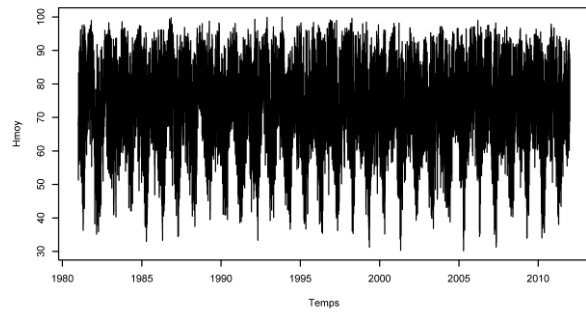
e) Snow



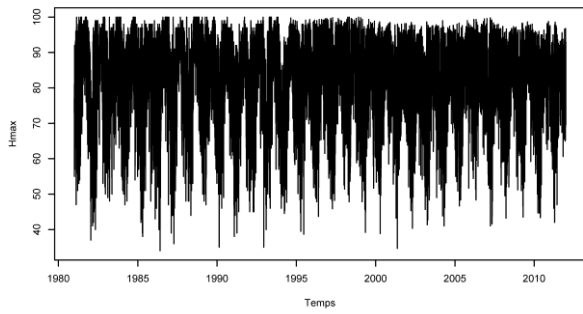
f) Pmoy



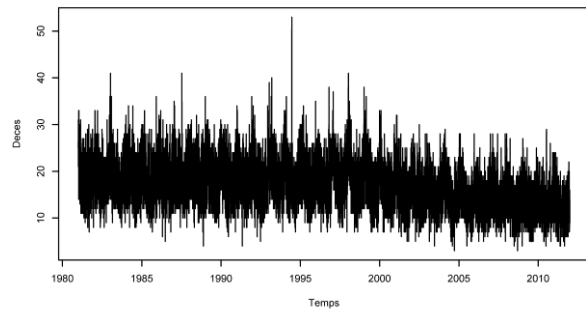
g) Hmin



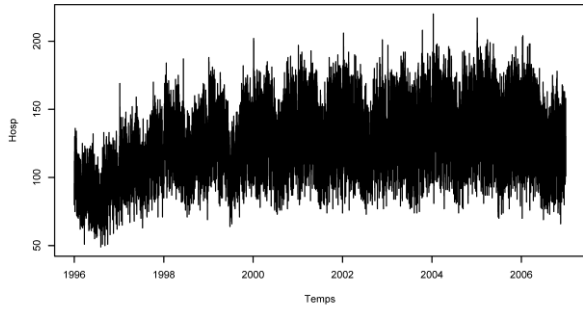
h) Hmoy



i) Hmax

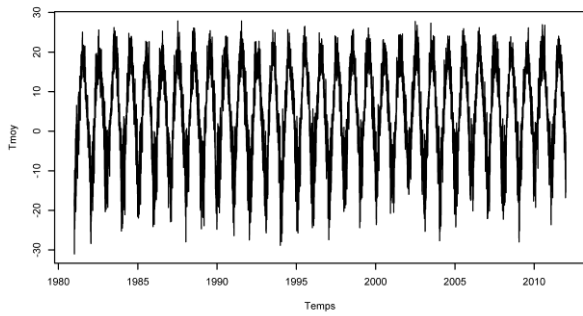


j) Deces

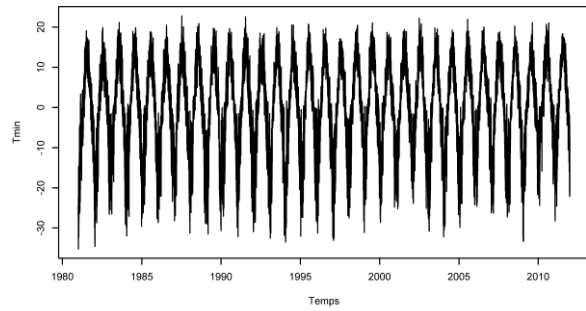


k) Hosp

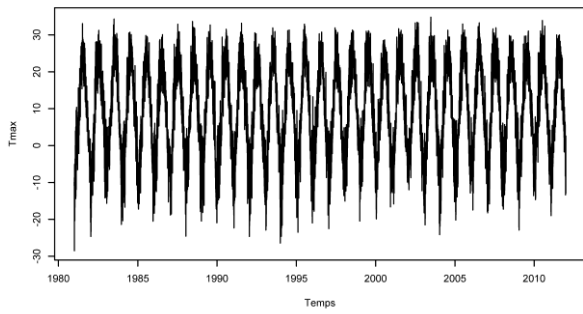
Figure A. 1 : Séries de données météorologiques et sanitaires de la CMM.



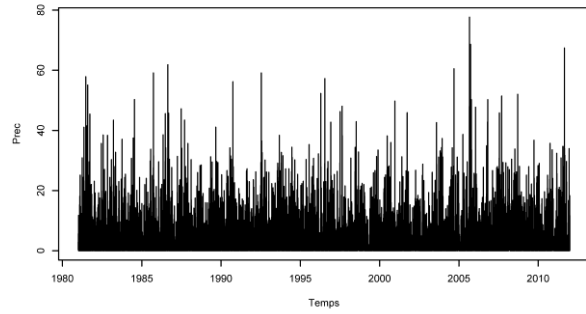
a) Tmoy



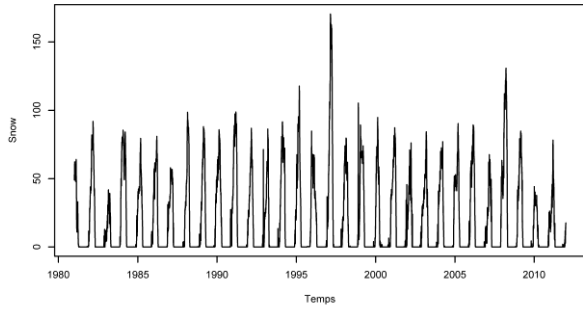
b) Tmin



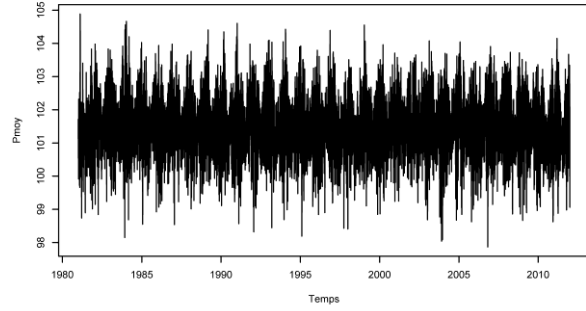
c) Tmax



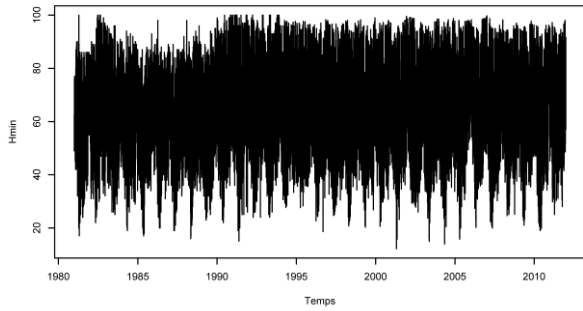
d) Prec



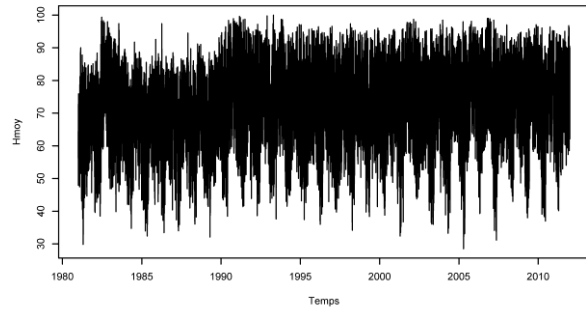
e) Snow



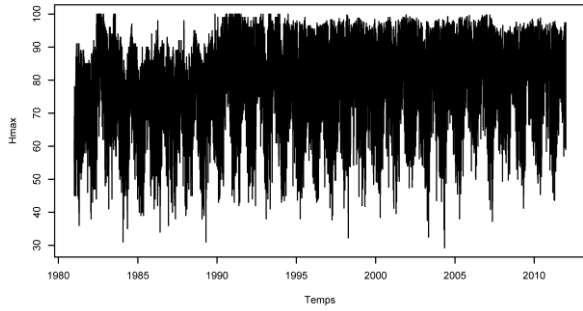
f) Pmoy



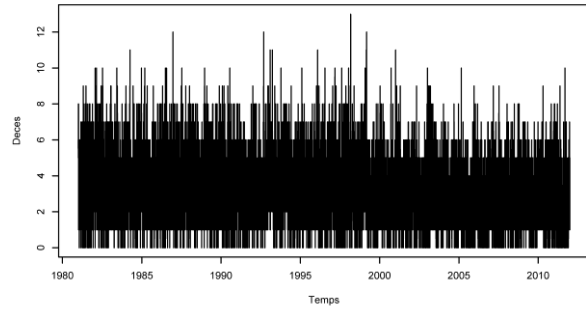
g) Hmin



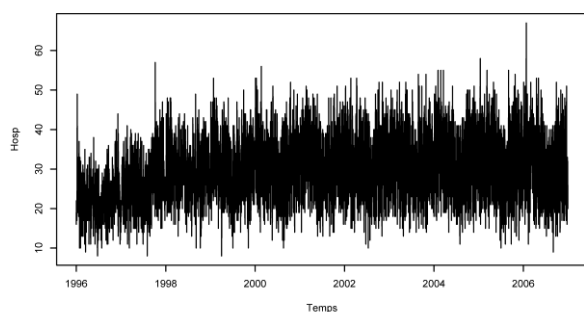
h) Hmoy



i) Hmax



j) Deces



k) Hosp

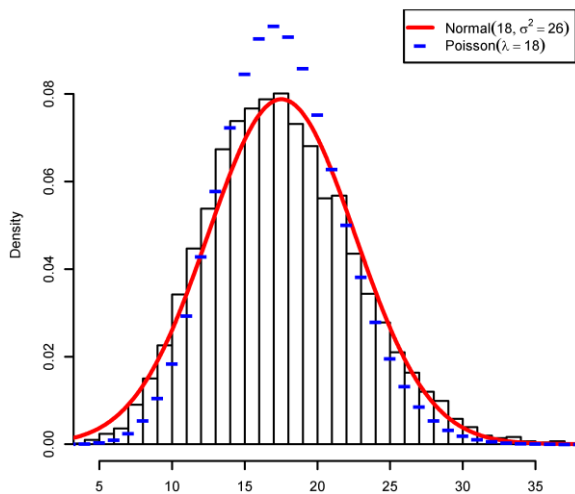
Figure A. 2 : Séries de données météorologiques et sanitaires de la CMQ.

	Tmoy	Tmin	Tmax	Prec	Snow	Pmoy	Hmin	Hmoy	Hmax	Deces	Hosp
Minimum	-29,6	-35,2	-26,5	0,0	0,0	98,4	14,5	30,2	34,0	3,0	49,0
Médiane	7,4	2,6	12,2	0,3	0,0	101,6	58,0	72,95	84,0	17,0	136,0
Moyenne	6,5	1,6	11,4	2,8	6,8	101,6	60,2	72,12	81,6	17,0	131,2
Maximum	29,6	24,0	35,0	89,8	79,2	104,9	100,0	100,0	100,0	53,0	220,0
Variance	140,1	135,0	152,1	32,4	163,0	0,68	358,9	166,6	153,4	25,6	1060,9

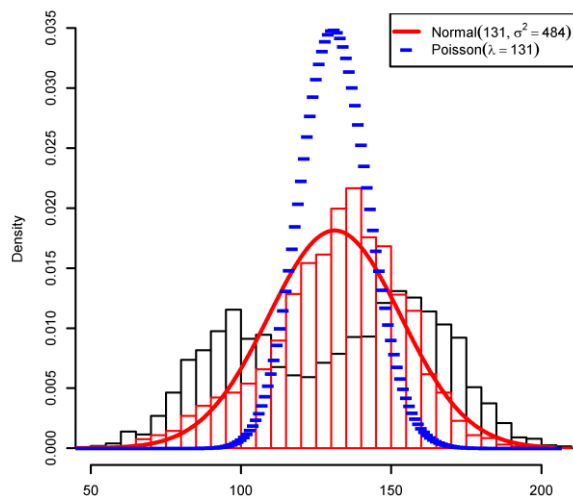
Tableau A. 1 : Statistiques descriptives des données sanitaires et météorologiques de la CMM.

	Tmoy	Tmin	Tmax	Prec	Snow	Pmoy	Hmin	Hmoy	Hmax	Deces	Hosp
Minimum	-31,0	-35,2	-28,5	0,0	0,0	97,9	12,2	28,5	29,3	0,0	8,0
Médiane	5,8	1,3	9,9	0,5	0,0	101,5	60,1	72,3	80,0	3,0	30,0
Moyenne	4,9	0,0	9,7	3,3	17,7	101,5	61,9	71,5	78,3	3,2	30,35
Maximum	27,9	22,8	34,8	77,7	170,5	104,9	100,0	100,0	100,0	13,0	67,0
Variance	133,1	128,6	145,1	39,2	893,7	0,7	357,4	168,4	171,4	3,4	83,4

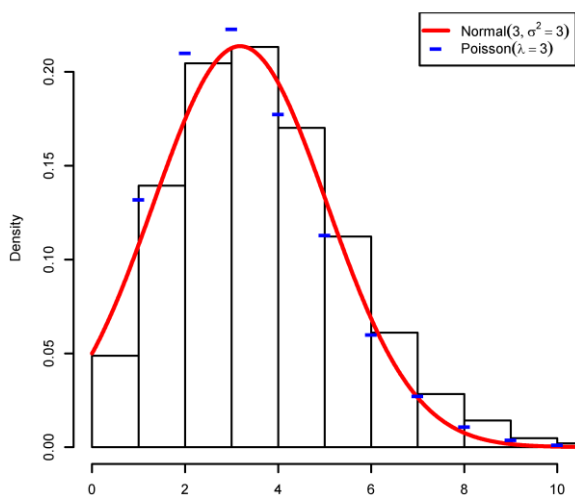
Tableau A. 2 : Statistiques descriptives des données sanitaires et météorologiques de la CMQ.



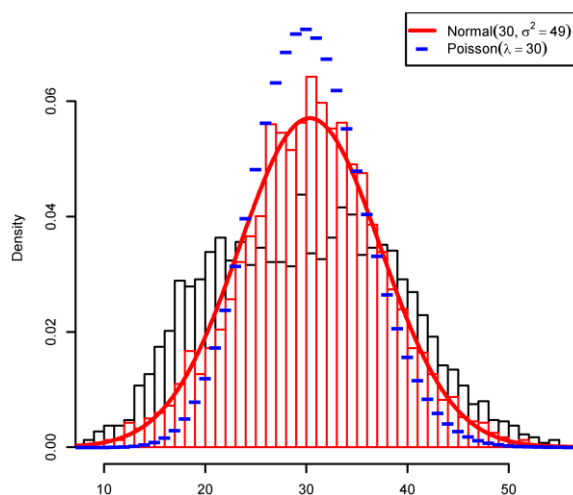
a) Décès de la CMM



b) Hospitalisations de la CMM

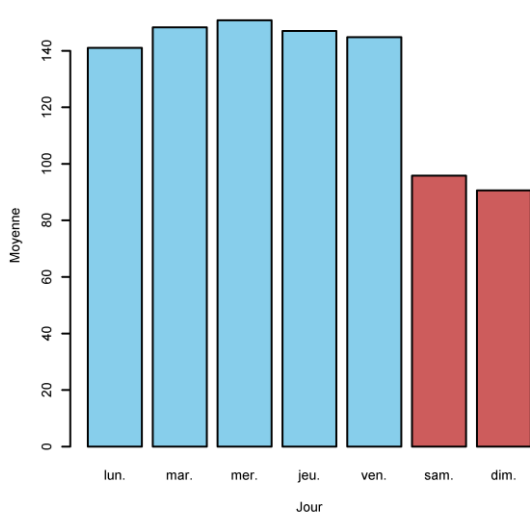


c) Décès de la CMQ

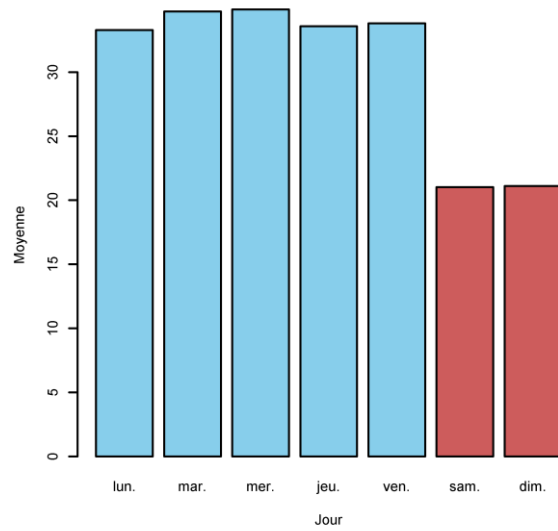


d) Hospitalisations de la CMQ

Figure A. 3 : Histogrammes des séries sanitaires réponses des quatre cas étudiés dans le rapport. La ligne rouge indique l'ajustement de la loi normale et les traits bleus indiquent l'ajustement d'une loi de Poisson. Dans le cas des hospitalisations, l'histogramme rouge indique la distribution de la série corrigée pour l'effet « fin de semaine ».

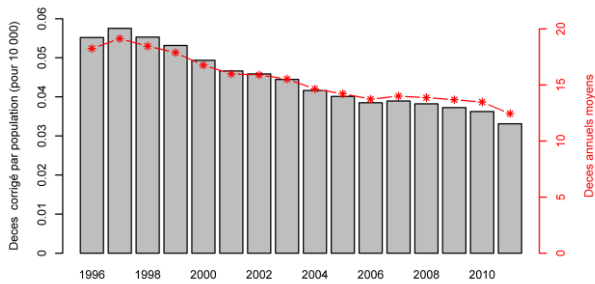


a) CMM

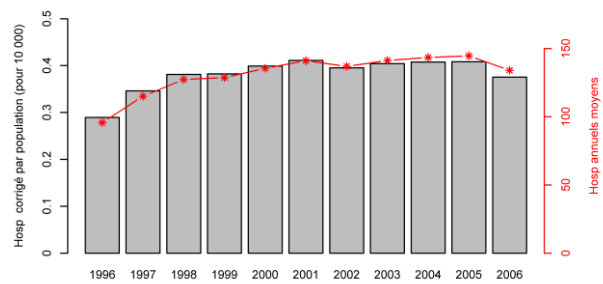


b) CMQ

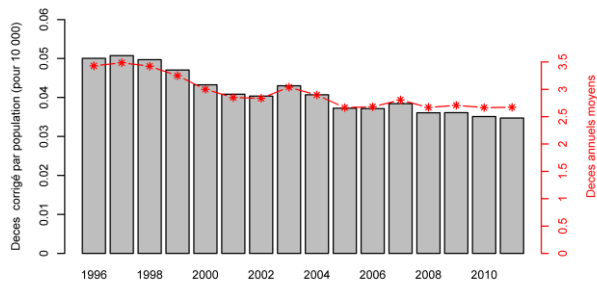
Figure A. 4 : Nombre moyen d'hospitalisations pour chaque jour de la semaine. Les barres rouges indiquent la fin de semaine afin d'illustrer la chute du nombre d'hospitalisations le samedi et le dimanche.



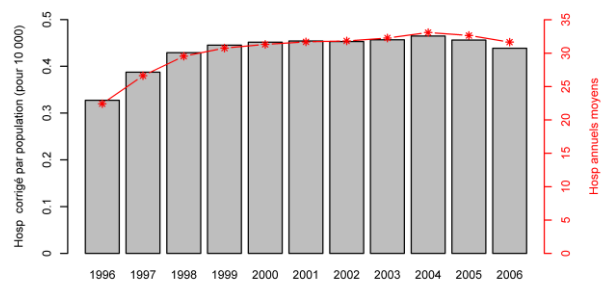
a) Décès de la CMM



b) Hospitalisations de la CMM

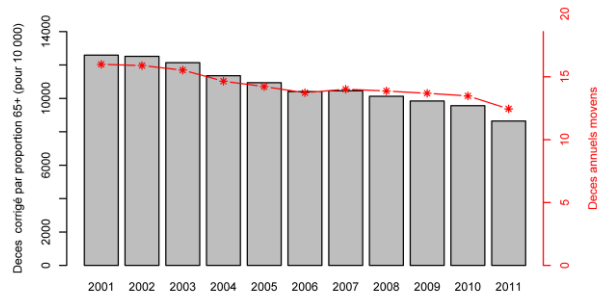


c) Décès de la CMQ

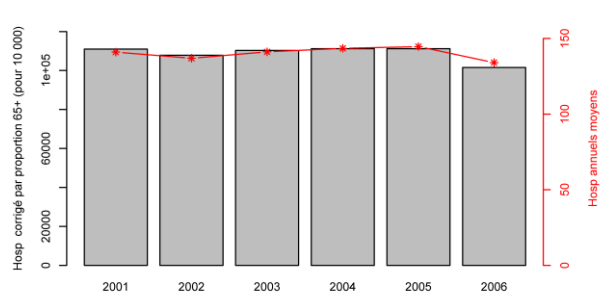


d) Hospitalisations de la CMQ

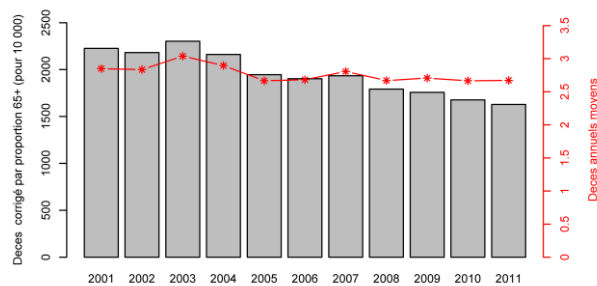
Figure A. 5 : Mortalité et morbidité moyennes annuelles corrigées en fonction de la population. La ligne rouge indique les valeurs moyennes annuelles non corrigées pour comparaisons.



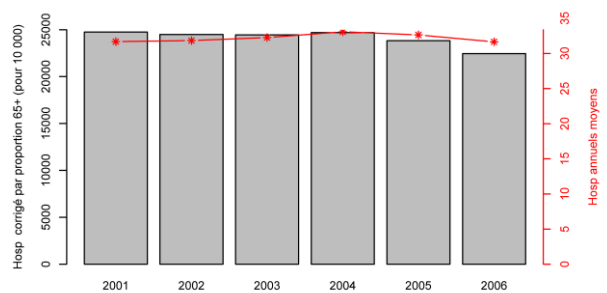
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ



d) Hospitalisations de la CMQ

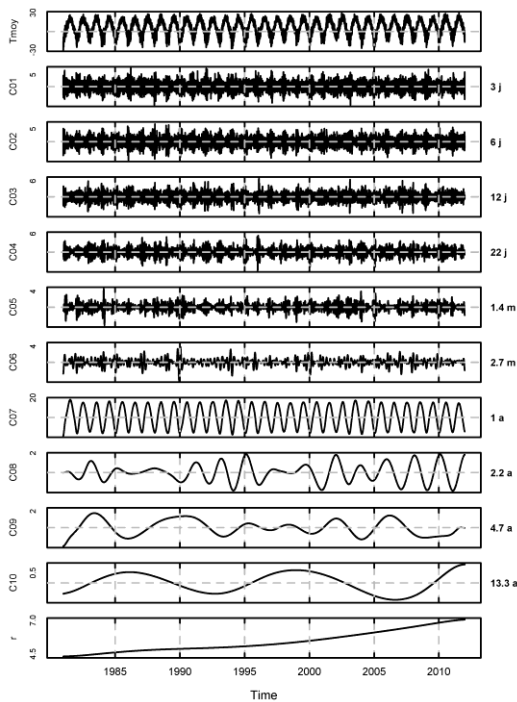
Figure A. 6 : Mortalité et morbidité moyennes annuelles corrigées en fonction de la proportion de personnes âgées de 65 ans et plus. La ligne rouge indique les valeurs moyennes annuelles non corrigées pour comparaisons.

La Figure A. 3 montre que, mis à part pour les décès de la CMQ, les données suivent approximativement une loi normale à cause du grand nombre de cas quotidiens. Les données sont de plus surdispensées par rapport à une loi de Poisson. Dans le cas des hospitalisations, l'ajustement des données à une loi normale nécessite la correction de l'effet « fin de semaine » (illustré par la Figure A. 4) qui entraîne une bimodalité de distribution.

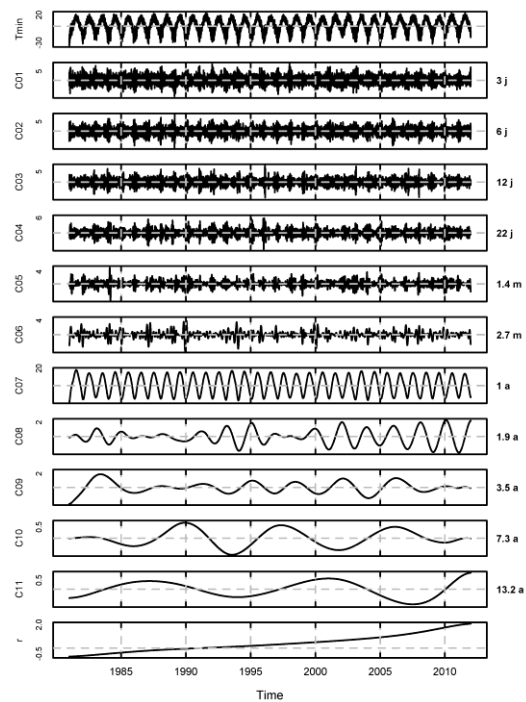
La Figure A. 5 et la Figure A. 6 montrent que les tendances des séries sanitaires ne sont pas dues à des évolutions démographiques de la CMM et de la CMQ. Ceci permet de donner de la force aux relations trouvées avec la météorologie à l'échelle des tendances.

B. IMF obtenus

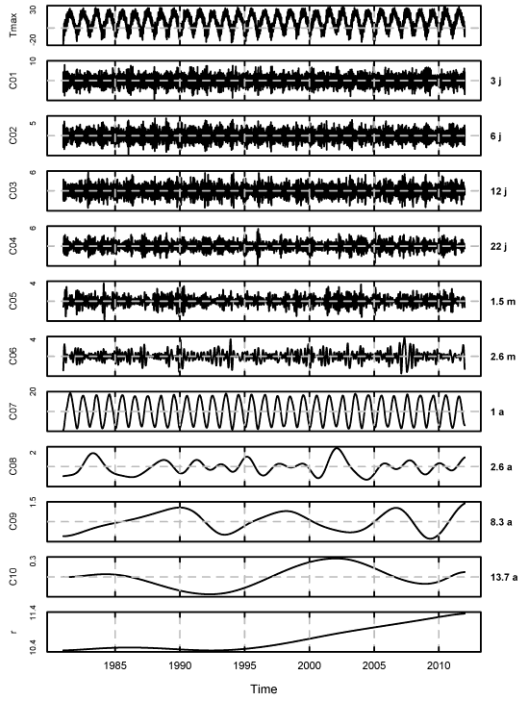
Cette partie de l'annexe montre l'ensemble des IMF utilisés dans l'étude. Cela inclue les IMF obtenus par EEMD pour la R-EMD1, et les IMF obtenus par MEMD pour la ML. De plus, les IMF sont différents pour la CMM et pour la CMQ. La décomposition EEMD de la variable Snow de la CMM n'est pas incluse, car déjà présentée dans la Figure 6. Les IMF obtenus par NAMED ne couvrent que la période 1996-2006 à cause de la présence de la variable des hospitalisations dans la série multivariée décomposée.



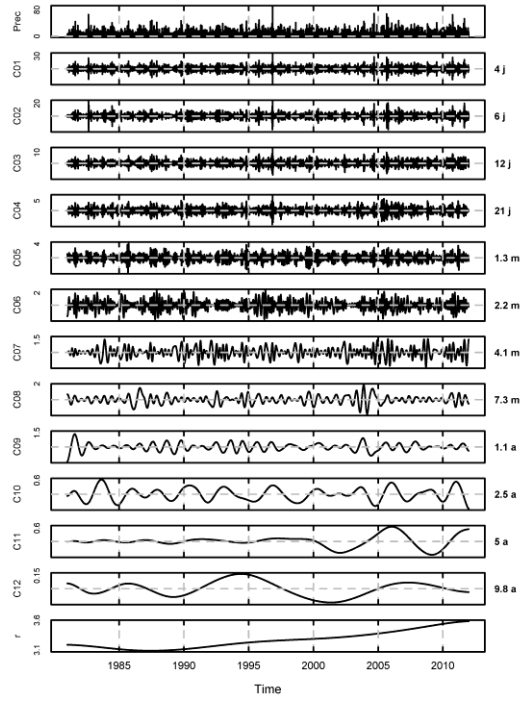
a) *Tmoy*



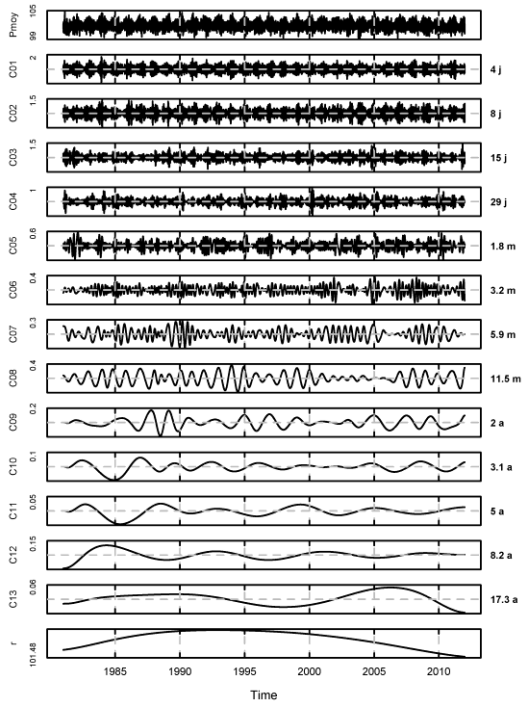
b) *Tmin*



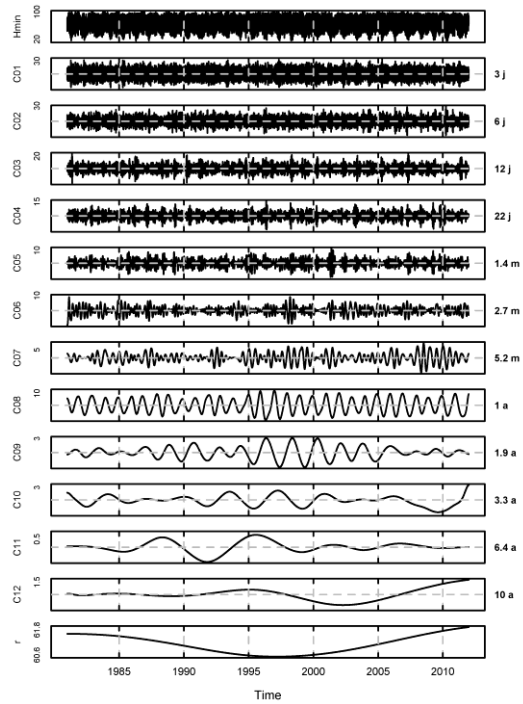
c) T_{max}



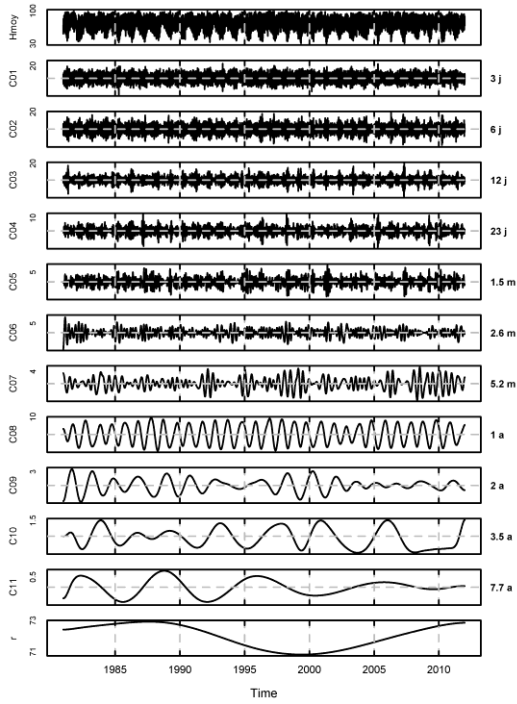
d) $Prec$



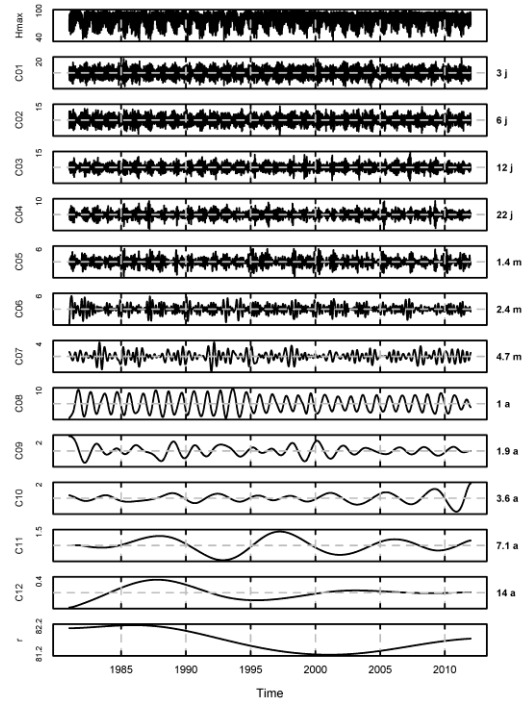
e) P_{moy}



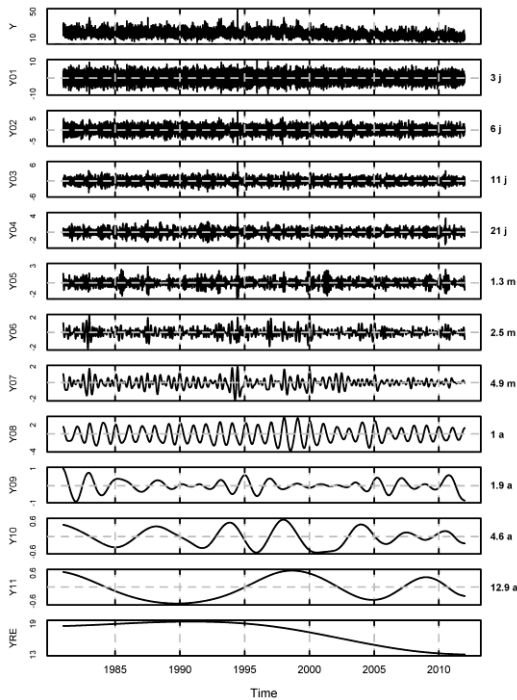
f) H_{min}



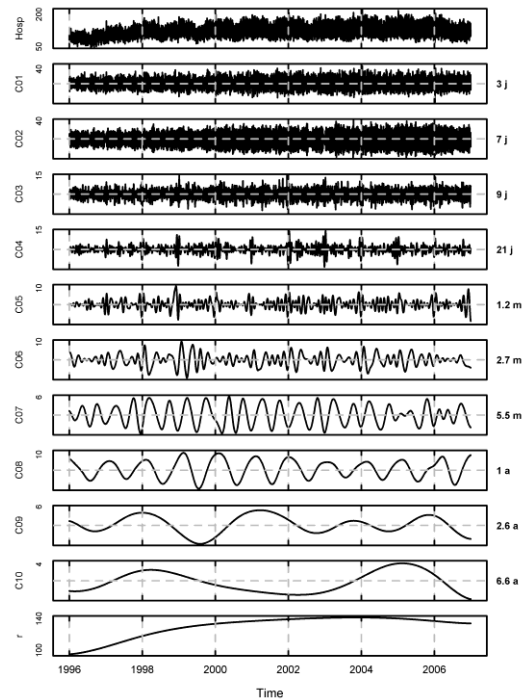
g) Hmoy



h) Hmax

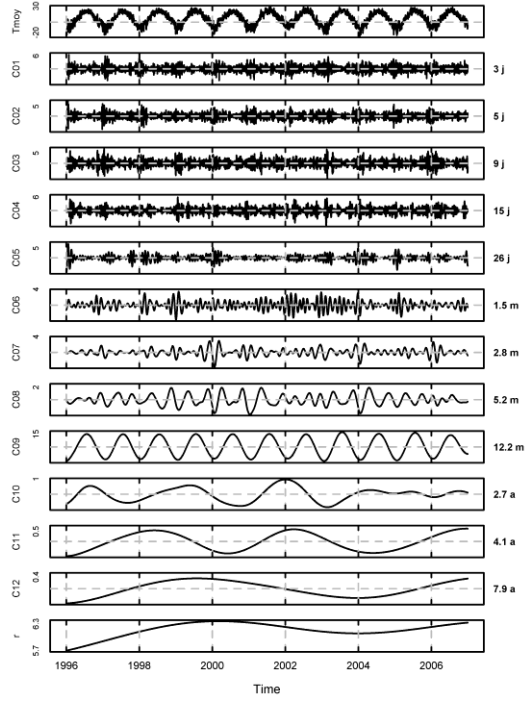


i) Deces

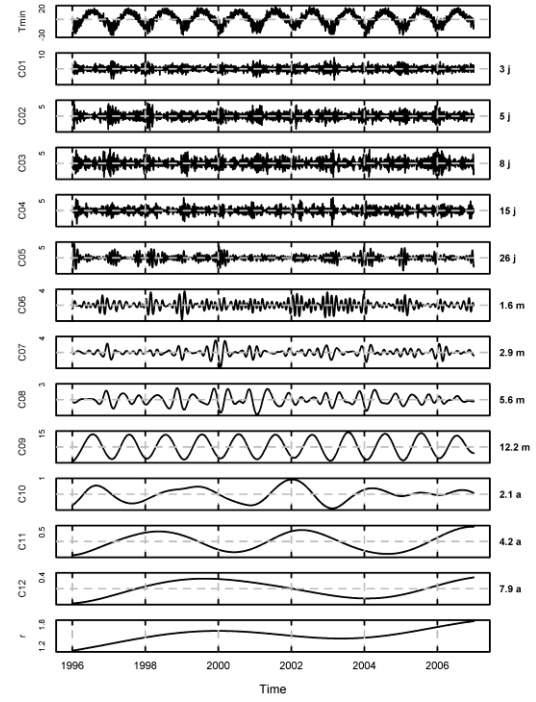


j) Hosp

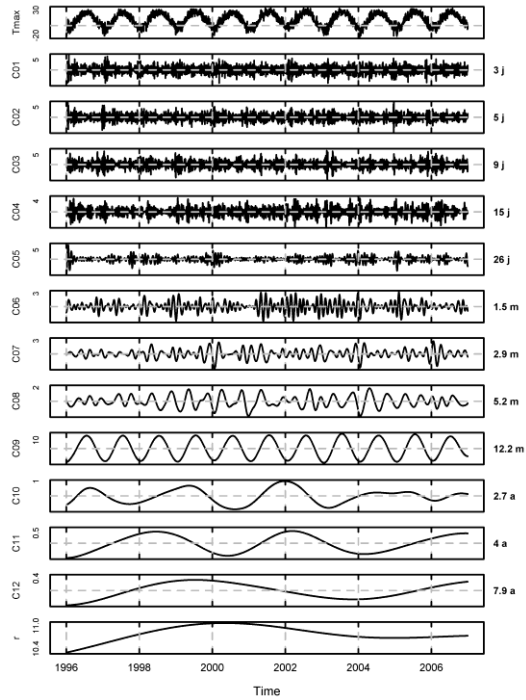
Figure B. 1 : IMF issus de l'EEMD pour la CMM.



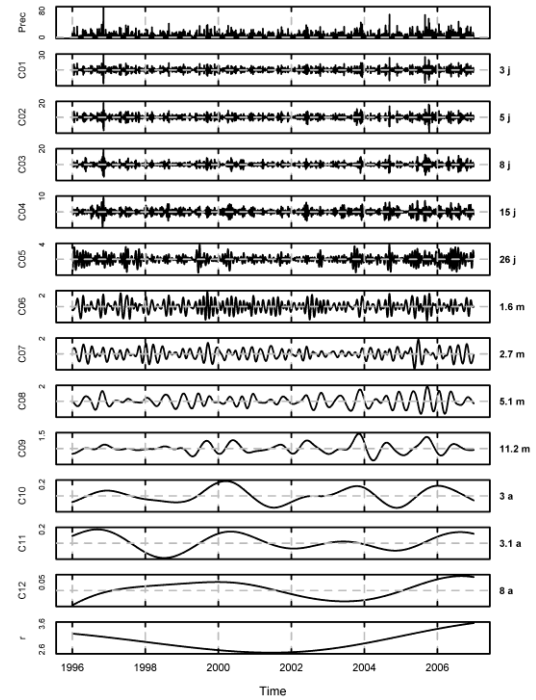
a) T_{moy}



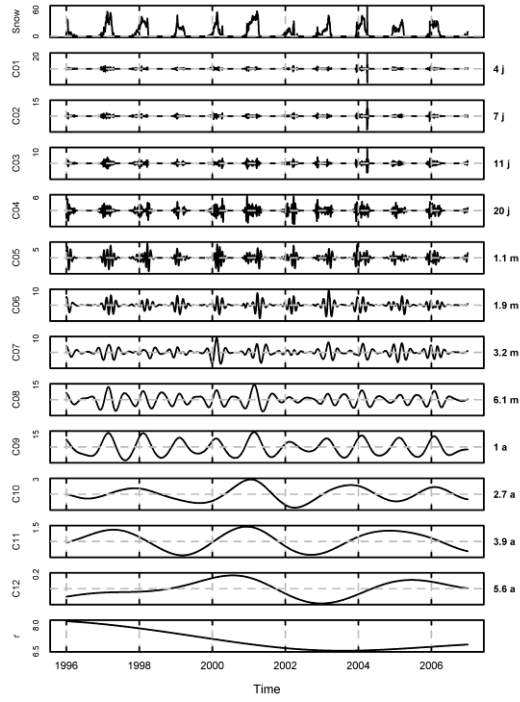
b) T_{min}



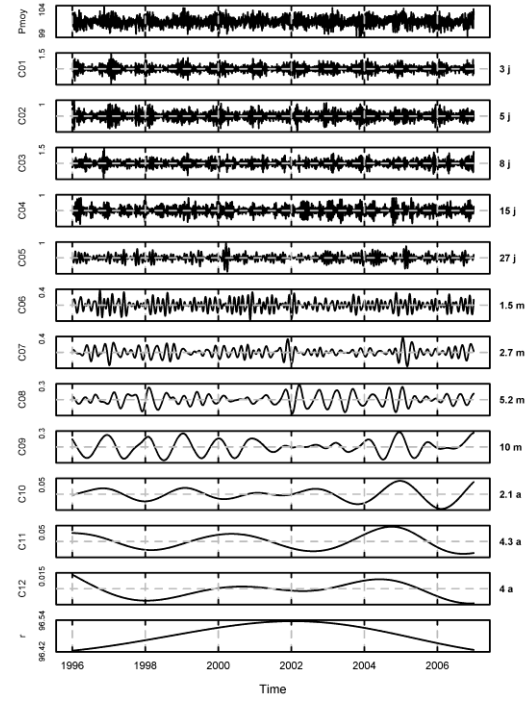
c) T_{max}



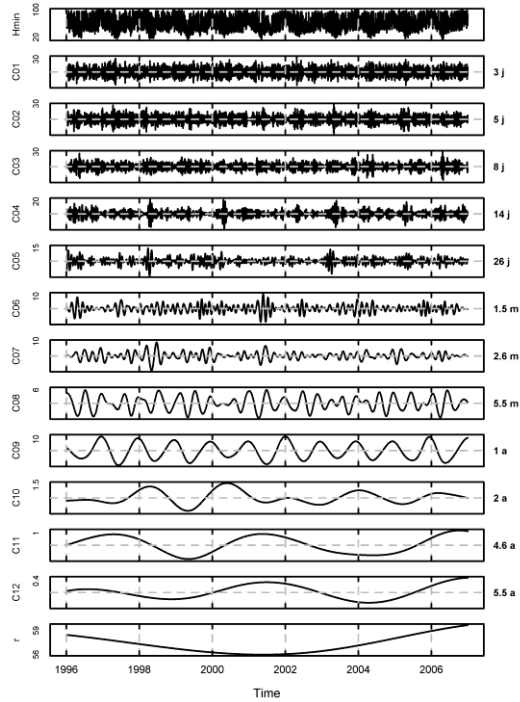
d) $Prec$



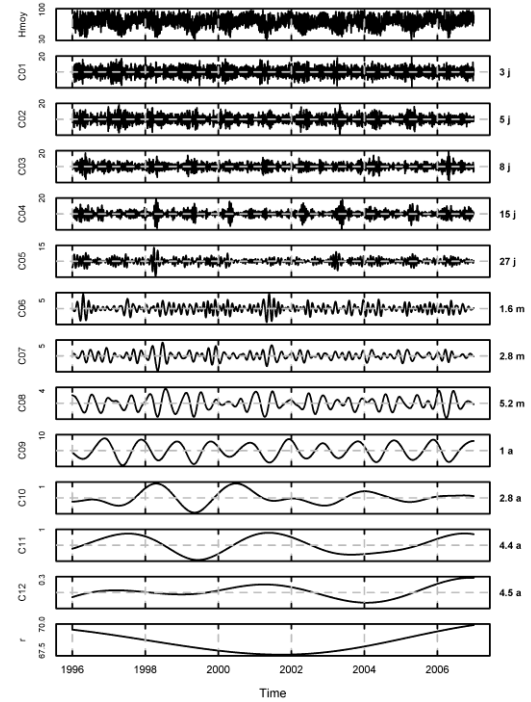
e) *Snow*



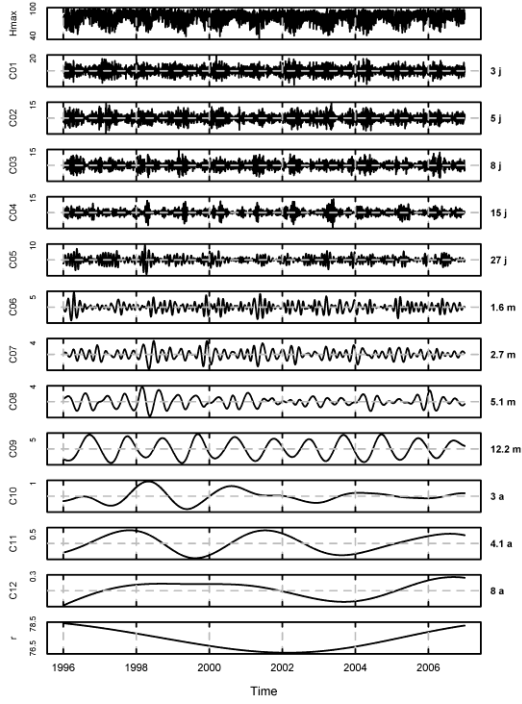
f) *Pmoy*



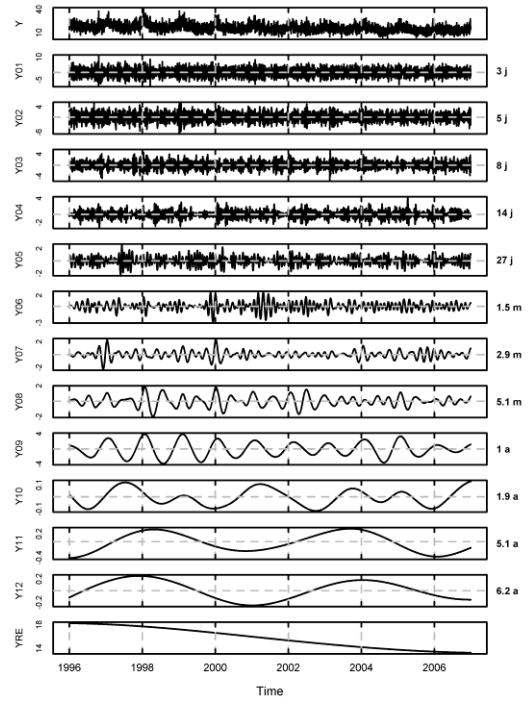
g) *Hmin*



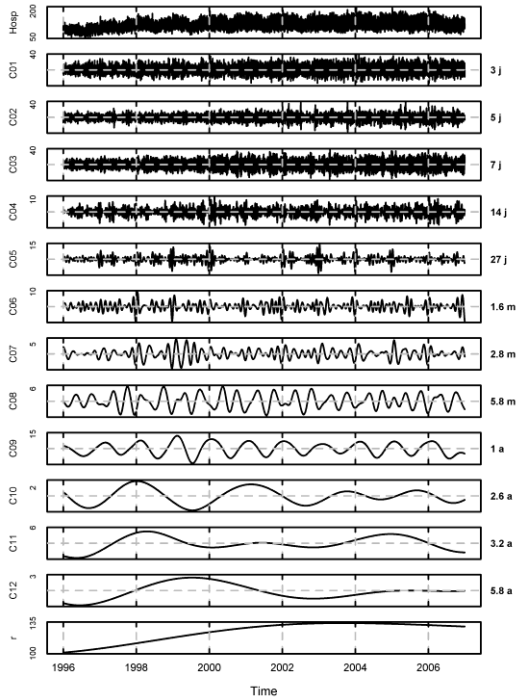
h) *Hmoy*



i) *Hmax*

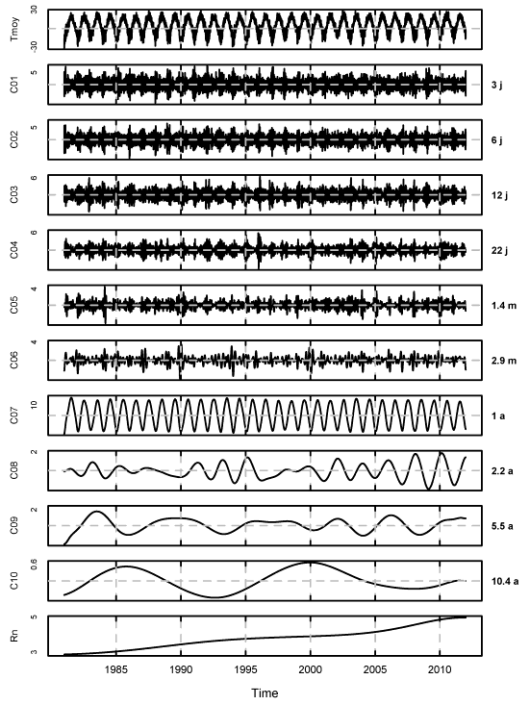


j) *Deces*

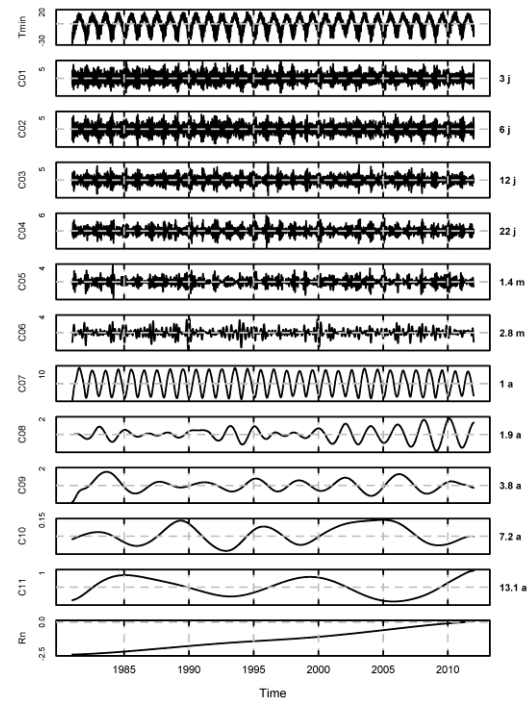


k) *Hosp*

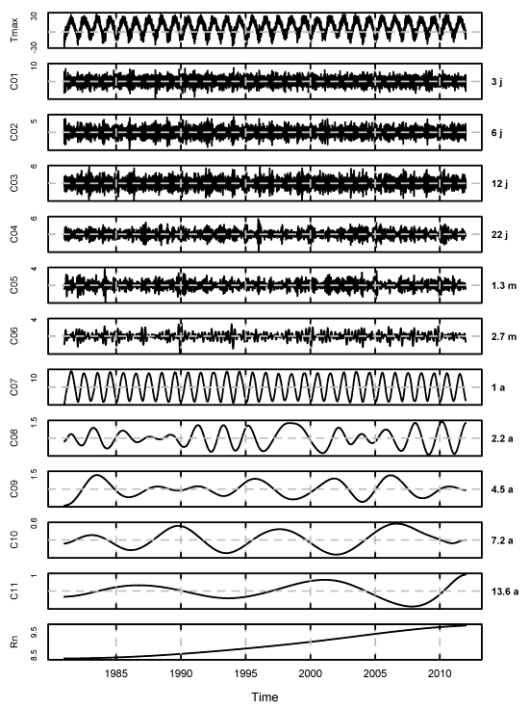
Figure B. 2 : IMF issus du NA-MEMD pour la CMM.



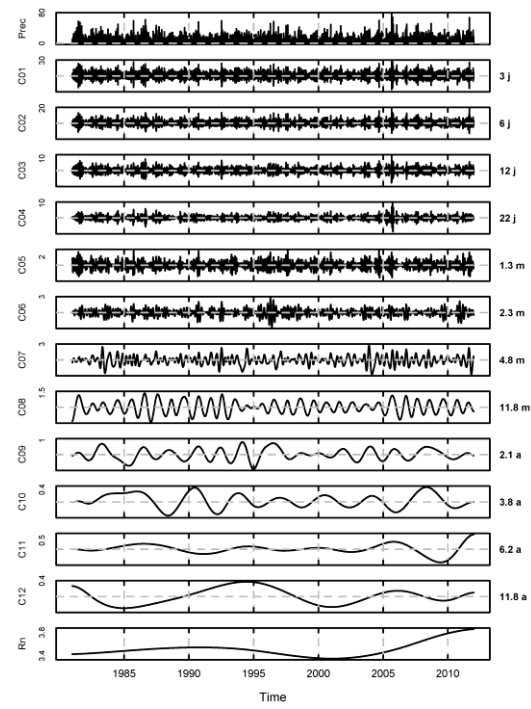
a) T_{moy}



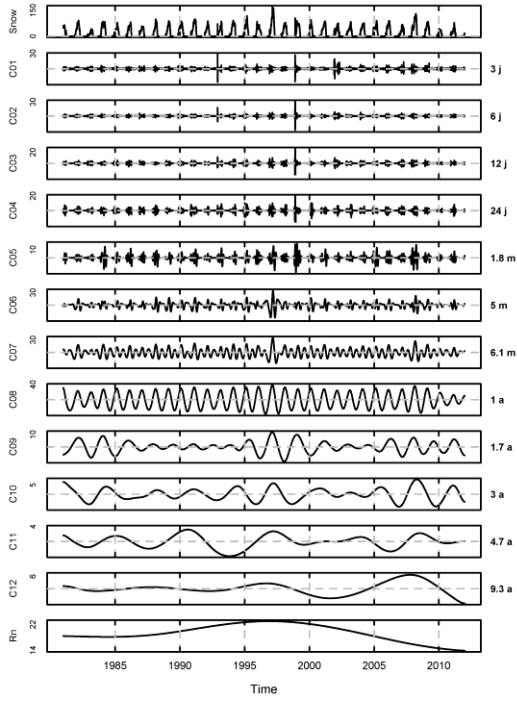
b) T_{min}



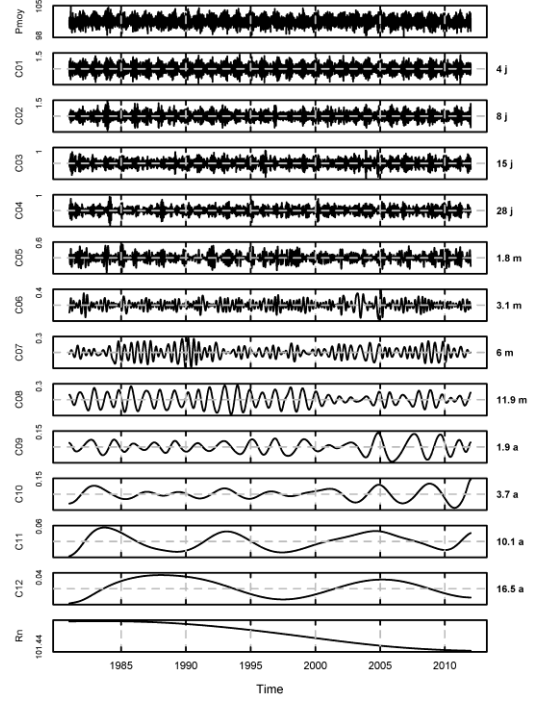
c) T_{max}



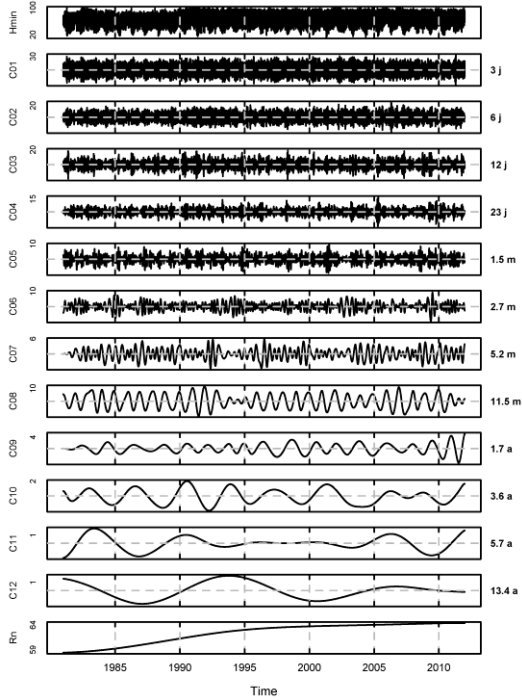
d) $Prec$



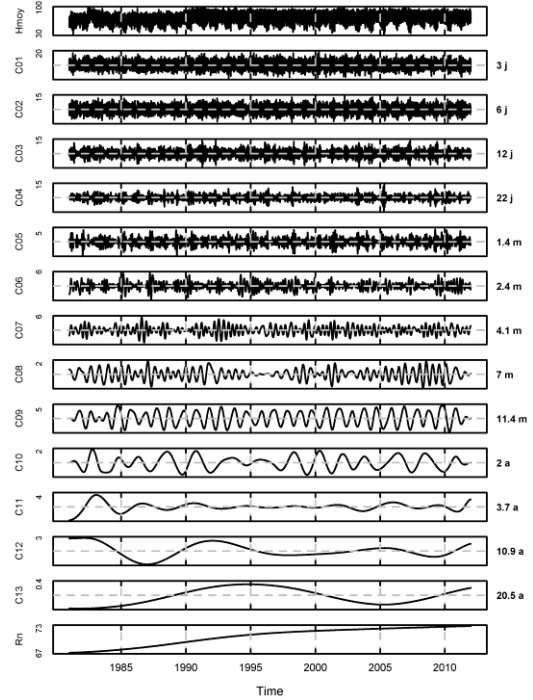
e) Snow



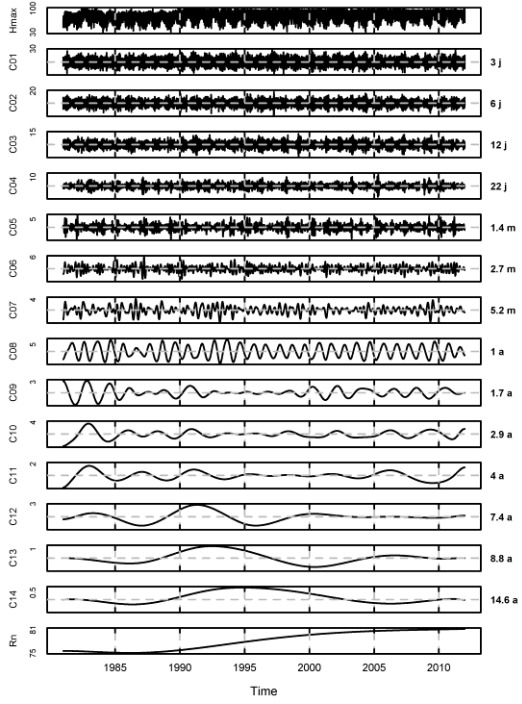
f) Pmoy



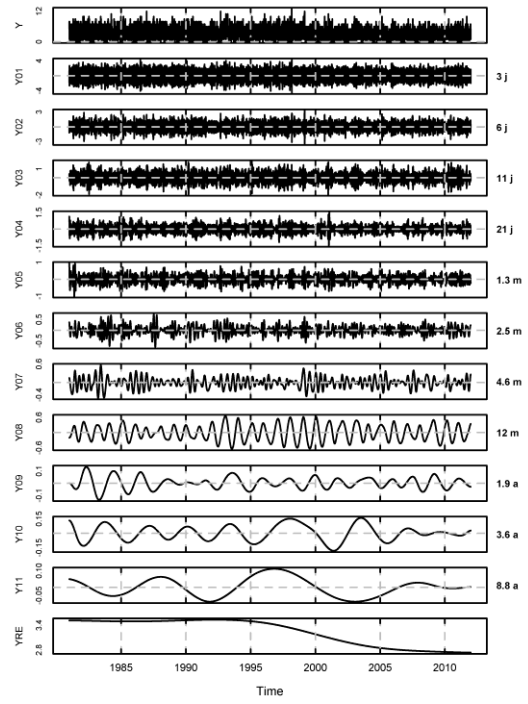
g) Hmin



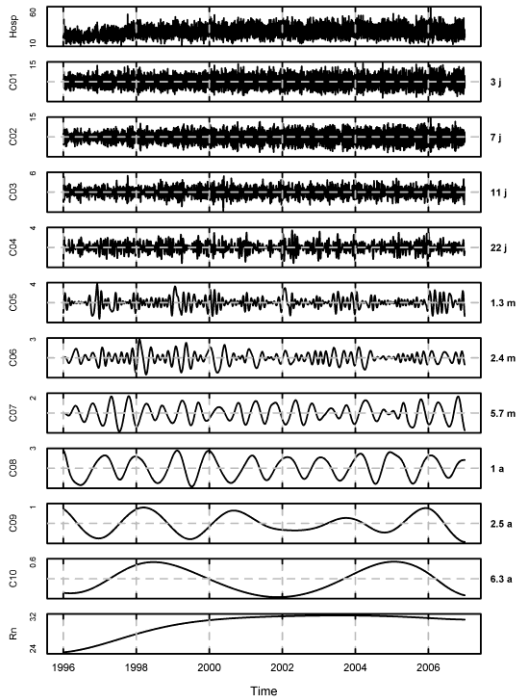
h) Hmoy



i) *Hmax*

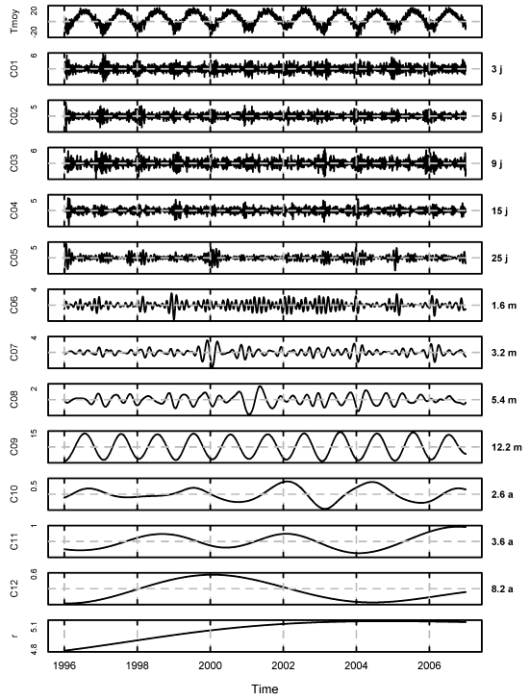


j) *Deces*

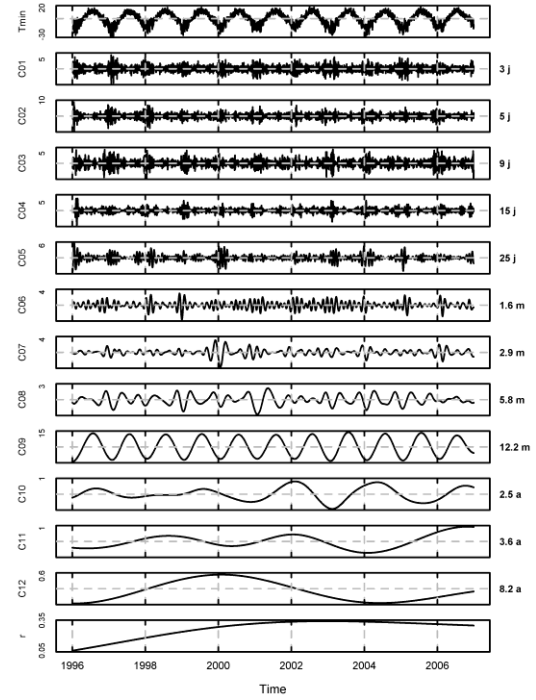


k) *Hosp*

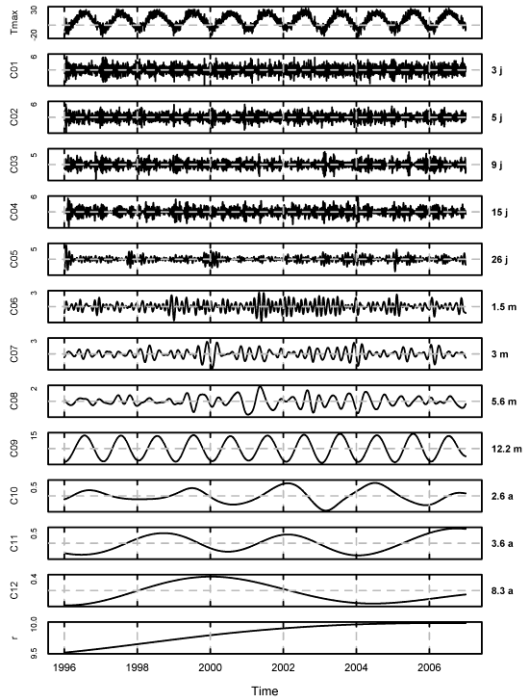
Figure B. 3 : IMF issus de l'EEMD pour la CMQ.



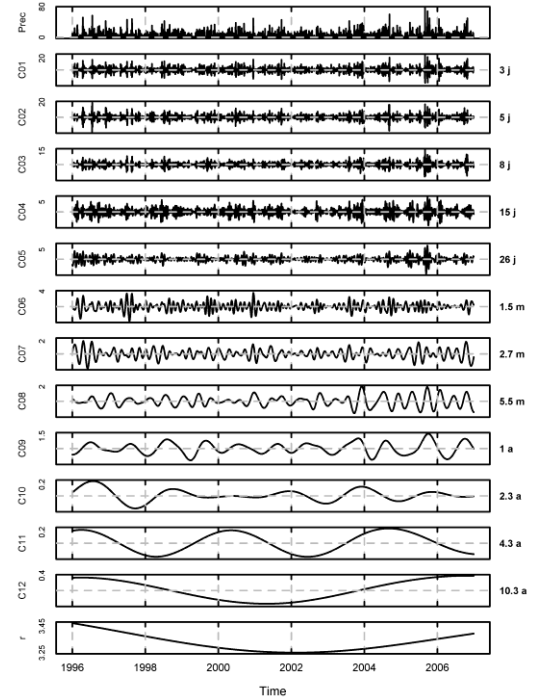
a) T_{moy}



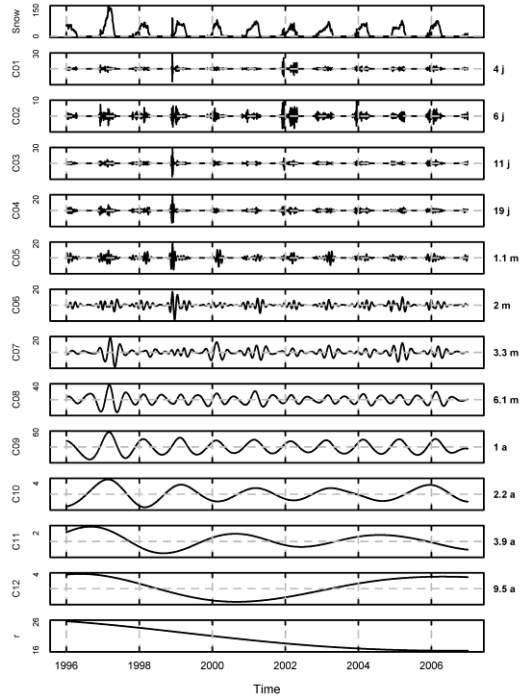
b) T_{min}



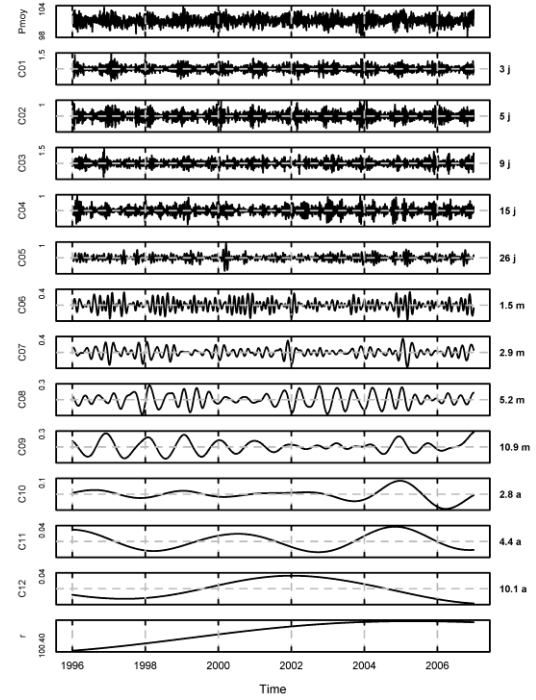
c) T_{max}



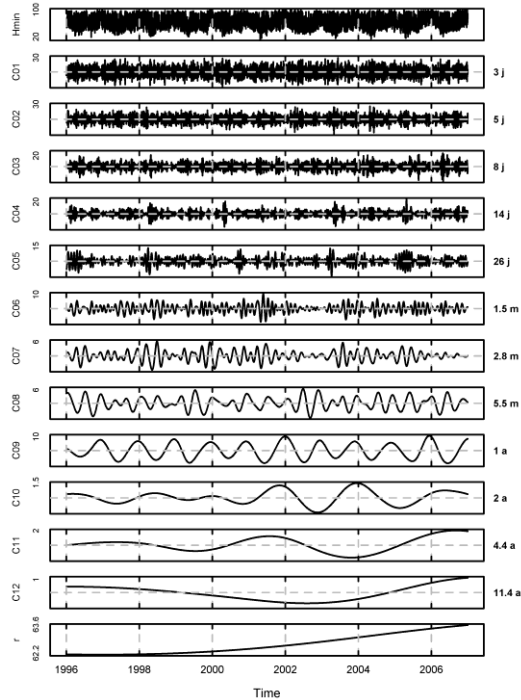
d) $Prec$



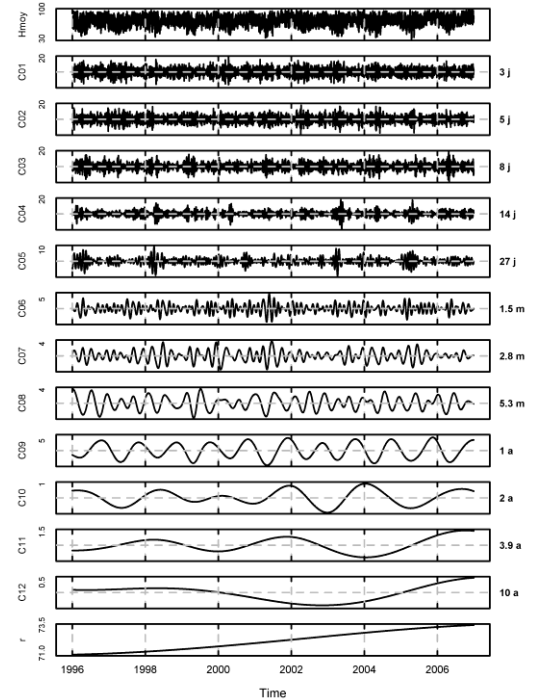
e) Snow



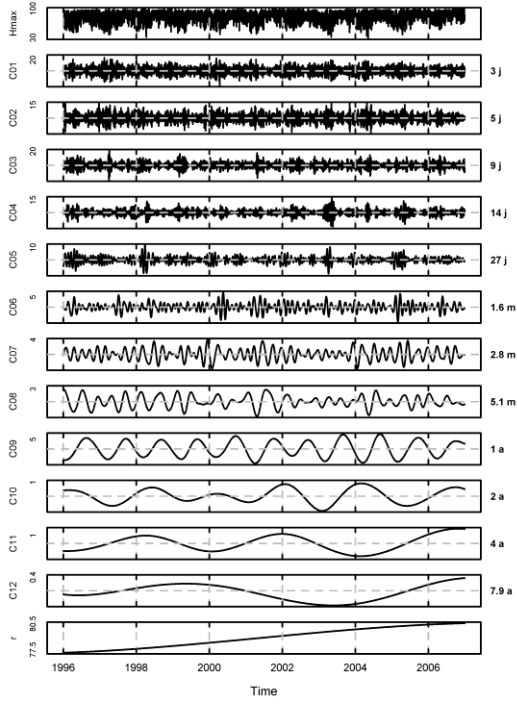
f) Pmoy



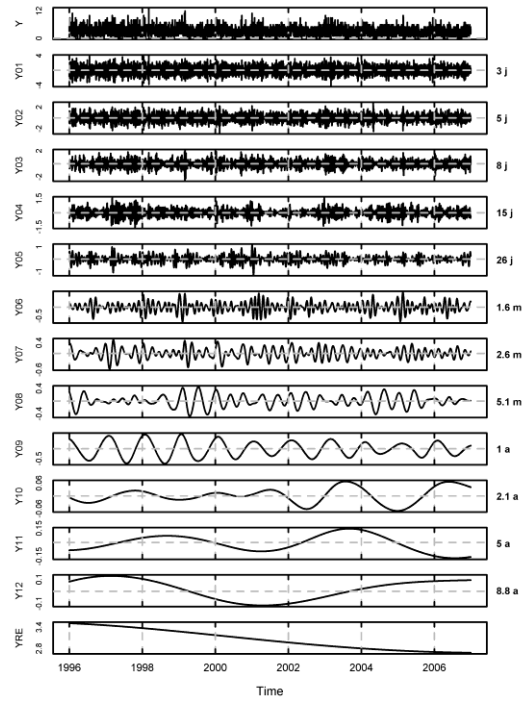
g) Hmin



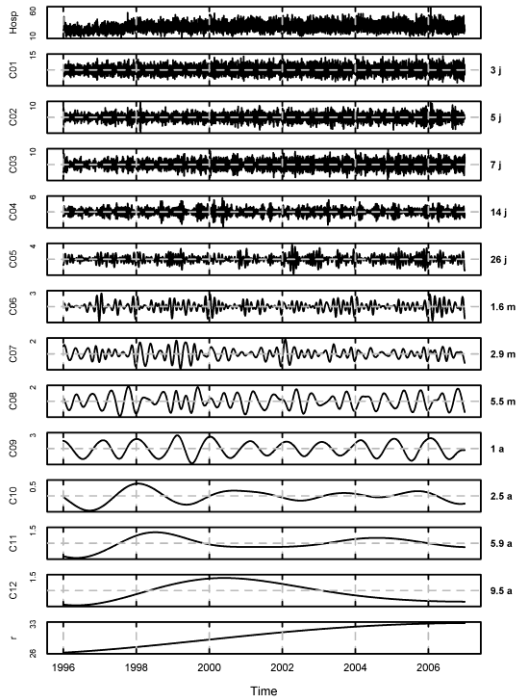
h) Hmoy



i) *Hmax*



j) *Deces*



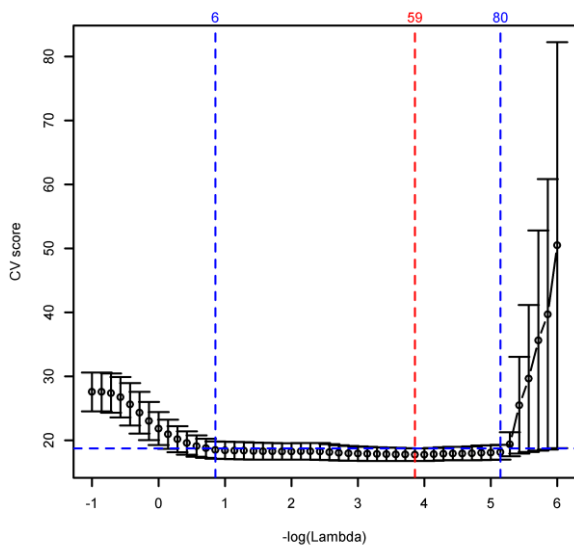
k) *Hosp*

Figure B. 4 : IMF issus du NA-MEMD pour la CMQ.

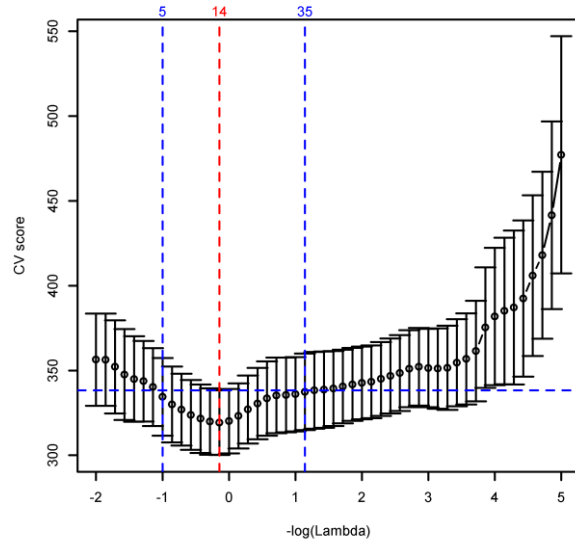
C. Figures supplémentaires pour la R-EMD

Cette annexe présente les figures détaillant les résultats de la régression EMD, en commençant par la modélisation R-EMD1 puis R-EMD2.

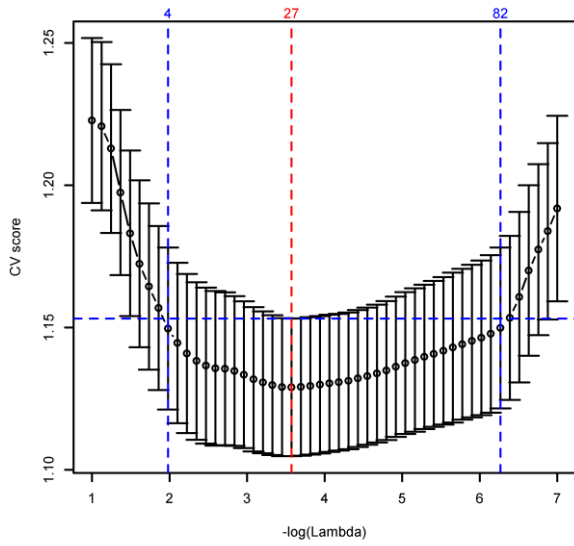
C.1 R-EMD1



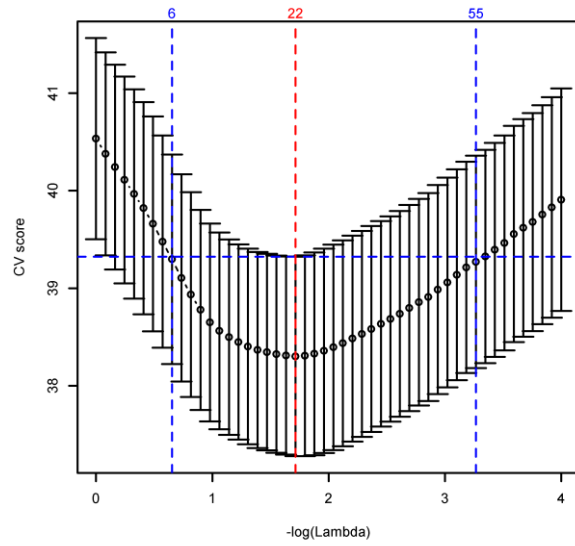
a) Décès de la CMM



b) Hospitalisations de la CMM

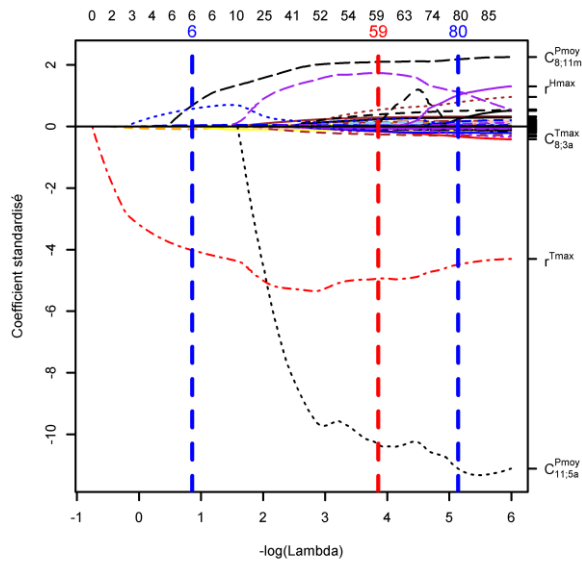


c) Décès de la CMQ

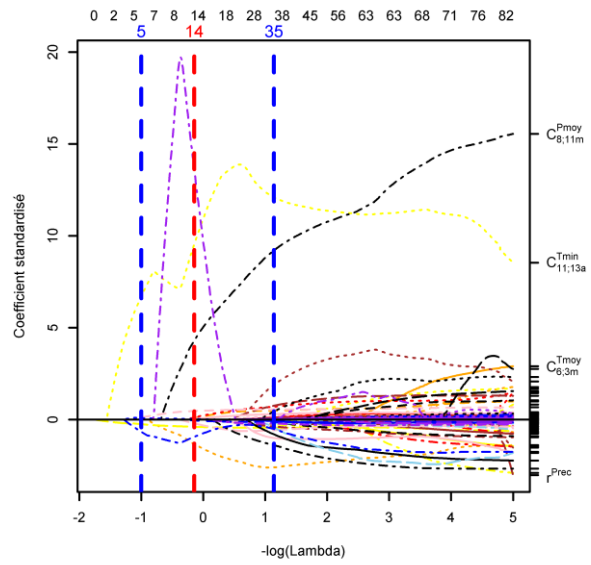


d) Hospitalisations de la CMQ

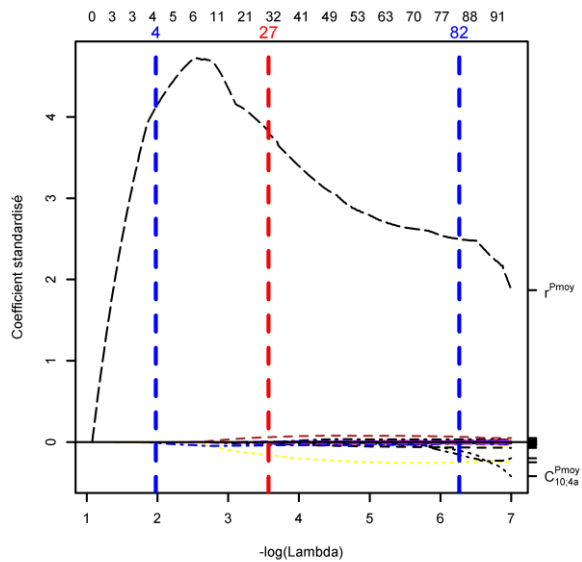
Figure C. 1 : Erreur de prédiction estimée par CV en fonction du paramètre de pénalisation λ du Lasso pour la modélisation R-EMD1. L'axe des abscisses correspond à $-\log(\lambda)$ pour que le début du graphe corresponde au modèle le plus simple et la fin au modèle le plus compliqué. Les barres verticales noires indiquent l'écart-type de la valeur minimale de CV obtenue. La ligne pointillée rouge indique le minimum de la fonction de CV. La ligne pointillée bleue horizontale indique la valeur minimum de CV en ajoutant son erreur type. Les lignes pointillées bleues verticales délimitent l'intervalle pendant lequel la CV est au sein de l'erreur type du minimum, ce que l'on peut interpréter comme la zone des valeurs acceptables. Le nombre de variables présentes dans le modèle pour les valeurs de $-\log(\lambda)$ sont indiquées au haut du graphe.



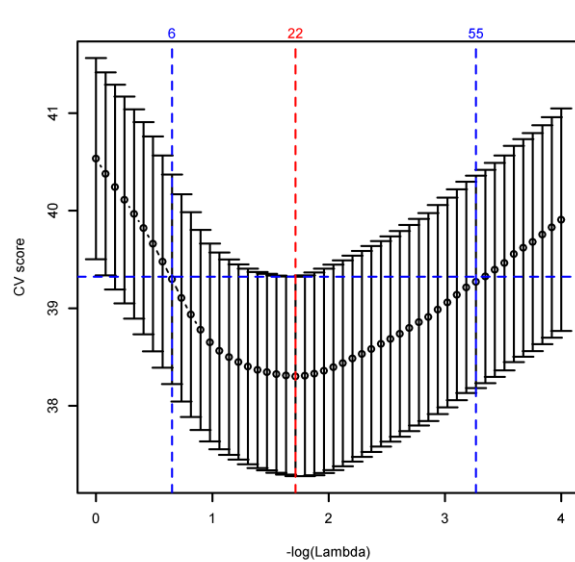
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ



d) Hospitalisations de la CMQ

Figure C. 2 : Chemins de pénalisation de la modélisation R-EMD1. L'axe des abscisses correspond à $-\log(\lambda)$ pour que le début du graphe corresponde au modèle le plus simple et la fin au modèle le plus compliqué. Ici, les coefficients sont standardisés, c'est-à-dire qu'ils ne dépendent pas de l'ordre de grandeur des variables auxquelles ils sont associés et peuvent donc être directement comparés. La marge de droite indique les IMF représentés par chaque ligne. La marge supérieure indique le nombre de variables présentes dans le modèle à certaines étapes. Les lignes pointillées verticales indiquent la valeur optimale (en rouge) et les limites des valeurs acceptables de λ (en bleu).

C.2 R-EMD2

À cause de la quantité de résultats à montrer pour cette modélisation, toutes les courbes de CV et chemins de pénalisations ne sont pas rapportés ici. Seul le détail des coefficients obtenus pour chaque modèle est présenté.

Tableau C. 1 : Coefficients obtenus pour les modèles toutes saisons de la R-EMD2 des décès de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	-0,01	-	0,04	-	-	0,09	-	-	-	-
$\tilde{C}_{2;5j}^{Deces}$	-	0,01	0,03	-	0,03	-0,01	-	-	-	-
$\tilde{C}_{3;9j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{4;15j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{5;1m}^{Deces}$	-	-	-0,01	0,02	-	-0,04	-	-0,02	-	-
$\tilde{C}_{6;2m}^{Deces}$	-0,01	-	-	-0,03	-	-0,03	-	-0,01	-0,02	-
$\tilde{C}_{7;3m}^{Deces}$	-	-	-0,12	-	-	-	-	-0,01	-	-
$\tilde{C}_{8;5m}^{Deces}$	-	-	-0,08	-	-	-	-	0,01	-	-
$\tilde{C}_{9;1a}^{Deces}$	-0,04	-0,09	-0,01	-	-	-	1,40	-	-	-
$\tilde{C}_{10;2a}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{11;5a}^{Deces}$	0,01	-	-0,04	-0,12	-0,61	0,01	2,20	0,01	-	0,35
$\tilde{C}_{12;7a}^{Deces}$	-	-	0,28	-0,57	-0,79	0,04	-7,75	0,03	-0,02	-
$\tilde{C}_{13;13a}^{Deces}$	-0,02	-	0,07	-	-	0,09	-4,32	-	-0,03	-
$\tilde{C}_{14;26a}^{Deces}$	-0,04	-	-0,31	-	6,00	-0,86	-	-	-0,40	-
\tilde{f}^{Deces}	-1404,10	-3,80	-1,31	-	-	-	14,25	-	-	-

Tableau C. 2 : Coefficients obtenus pour les modèles estivaux de la R-EMD2 des décès de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	-0,01	-	-	-	-	-	-	-	-	-
$\tilde{C}_{2;5j}^{Deces}$	-	0,43	0,13	-0,18	0,03	-0,09	0,40	-0,02	0,04	-0,01
$\tilde{C}_{3;9j}^{Deces}$	-	-	0,07	-	-	-	-	-	-0,01	-
$\tilde{C}_{4;15j}^{Deces}$	-	0,10	-	-	-	-	-	-	-	-
$\tilde{C}_{5;1m}^{Deces}$	-	0,06	-	-	-0,14	-	0,03	-	-0,03	-
$\tilde{C}_{6;2m}^{Deces}$	-0,01	-	-	-	-	-	-	-0,02	-	-
$\tilde{C}_{7;3m}^{Deces}$	-0,03	-	-	-	-	-	-	-	-	-

Tableau C. 3 : Coefficients obtenus pour les modèles hivernaux de la R-EMD2 des décès de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	-0,01	-	-	-	0,03	-0,03	-0,25	-	-	-
$\tilde{C}_{2;5j}^{Deces}$	-0,01	-	-	-	0,02	-	-	-	-	-
$\tilde{C}_{3;9j}^{Deces}$	-	-	0,08	-0,08	0,03	-	-0,32	-0,01	-0,02	-0,04
$\tilde{C}_{4;15j}^{Deces}$	-	-	0,03	-0,08	0,05	-0,03	0,41	0,01	0,01	0,04
$\tilde{C}_{5;1m}^{Deces}$	-	-	-0,07	0,04	0,06	-0,02	-0,28	-	-	-0,04
$\tilde{C}_{6;2m}^{Deces}$	-0,01	-	-	-	-	-0,01	-	-	-	-
$\tilde{C}_{7;3m}^{Deces}$	-0,02	-4,39	2,13	2,23	0,16	0,01	0,21	0,17	-0,18	-0,03

Tableau C. 4 : Coefficients obtenus pour les modèles toutes saisons de la R-EMD2 des hospitalisations de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	-0,33	-	-	-	-	-	-	-	-	-
$\tilde{C}_{2;5j}^{Hosp}$	0,29	-	-	-	-	-	-	-	-	-
$\tilde{C}_{3;7j}^{Hosp}$	-0,04	-	-	-	-	-	-	-	-	-
$\tilde{C}_{4;15j}^{Hosp}$	-0,03	-1,34	-0,13	1,41	-0,14	0,34	-2,18	-0,07	-0,10	0,12
$\tilde{C}_{5;26j}^{Hosp}$	-0,01	0,61	-	-	-0,29	-0,44	1,64	0,01	0,08	-
$\tilde{C}_{6;2m}^{Hosp}$	-	-	-	-0,14	-	0,16	-0,21	-	-	-
$\tilde{C}_{7;3m}^{Hosp}$	-0,01	-	-0,13	-	-0,29	-	1,41	-	-	-0,50
$\tilde{C}_{8;6m}^{Hosp}$	0,05	-	-	-	0,72	0,20	4,46	-0,16	-	0,34
$\tilde{C}_{9;1a}^{Hosp}$	0,15	-	-0,31	-	-	-	-	-	0,22	-
$\tilde{C}_{10;2a}^{Hosp}$	-0,03	-0,48	-0,34	-	-	0,39	-11,70	-	-	0,21
$\tilde{C}_{11;4a}^{Hosp}$	0,04	-	-3,19	-	-10,78	0,18	-31,17	-	-	2,45
$\tilde{C}_{12;5a}^{Hosp}$	0,46	-	-	-0,01	23,25	1,94	-	-4,42	-	-
\tilde{r}^{Hosp}	255,38	-	-	9,01	-11,05	-8,23	-	-	-1,85	-

Tableau C. 5 : Coefficients obtenus pour les modèles estivaux de la R-EMD2 des hospitalisations de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	-0,31	-	-	-	-	-	-	-	-	-
$\tilde{C}_{2;5j}^{Hosp}$	0,26	3,10	1,61	2,00	0,07	-12,14	-3,92	0,21	-1,00	0,83
$\tilde{C}_{3;7j}^{Hosp}$	-0,04	-	-	0,48	-	-	-	-	-	0,02
$\tilde{C}_{4;15j}^{Hosp}$	-0,02	-0,32	-0,55	-	-	-	-	-	-	0,06
$\tilde{C}_{5;26j}^{Hosp}$	-0,03	-0,46	-	-	-0,07	-	-	-	-	-0,06
$\tilde{C}_{6;2m}^{Hosp}$	0,22	-	-	-0,33	-	-	-	-	-	-
$\tilde{C}_{7;3m}^{Hosp}$	0,06	-	-	-0,32	-0,64	-0,15	-	-	-	-

Tableau C. 6 : Coefficients obtenus pour les modèles hivernaux de la R-EMD2 des hospitalisations de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	-0,35	-	-	-	-	-	-	-	-	-
$\tilde{C}_{2;5j}^{Hosp}$	0,30	-	-	-	-0,06	-	-	-	-	-
$\tilde{C}_{3;7j}^{Hosp}$	-0,03	0,20	-	-	-	-	-2,14	-	-	-
$\tilde{C}_{4;15j}^{Hosp}$	-0,05	-	-	-	-	0,11	-	-	-	-
$\tilde{C}_{5;26j}^{Hosp}$	-0,04	-	0,32	-	-	0,06	1,80	-	-	-
$\tilde{C}_{6;2m}^{Hosp}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{7;3m}^{Hosp}$	-0,30	-	-	-	-	-	0,87	-	-	-0,44

Tableau C. 7 : Coefficients obtenus pour les modèles toutes saisons de la R-EMD2 des décès de la CMQ. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	-0,01	-	0,01	-	-	0,03	-0,12	-	-0,01	0,01
$\tilde{C}_{2;5j}^{Deces}$	-	-	-	-	-	-	-0,08	-	-	-
$\tilde{C}_{3;9j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{4;15j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{5;27j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{6;2m}^{Deces}$	-	-	-	-0,02	0,01	-0,01	-	0,01	-	0,01
$\tilde{C}_{7;3m}^{Deces}$	-	-	-0,02	-	-	0,01	-	-	-	-
$\tilde{C}_{8;5m}^{Deces}$	-	-	-0,03	-	-	-	-0,18	0,02	-	-
$\tilde{C}_{9;1a}^{Deces}$	-0,01	-0,02	-	-	-	-	-	-	-	-
$\tilde{C}_{10;2a}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{11;4a}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{12;8a}^{Deces}$	-	-	-0,19	-	-0,16	-	-0,33	0,12	-	-0,09
$\tilde{C}_{13;14a}^{Deces}$	-	-	-	-	0,15	-	-0,09	0,01	-	-
$\tilde{C}_{14;25a}^{Deces}$	0,01	-	-0,15	-	0,65	0,02	-	-	-	-
\tilde{r}^{Deces}	3,96	-	-0,64	-	-0,24	-	-	-	-	-

Tableau C. 8 : Coefficients obtenus pour les modèles estivaux de la R-EMD2 des décès de la CMQ. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	-0,01	-	-	0,01	-	0,58	-0,18	-	-0,01	0,01
$\tilde{C}_{2;5j}^{Deces}$	-	0,01	0,01	-	-	-	-	-	-	-
$\tilde{C}_{3;9j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{4;15j}^{Deces}$	-	-	-	-	-	-	-0,06	-	-	-
$\tilde{C}_{5;27j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{6;2m}^{Deces}$	-	-	-	-0,02	-0,01	-	-	-	-	-
$\tilde{C}_{7;3m}^{Deces}$	0,01	-	-	-	-	-	-	-	-	-

Tableau C. 9 : Coefficients obtenus pour les modèles hivernaux de la R-EMD2 des décès de la CMQ. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	-0,01	-	-	-	-	0,02	-0,12	-	-	-
$\tilde{C}_{2;5j}^{Deces}$	-	-	-	-	-	-	-0,08	-	-	-
$\tilde{C}_{3;9j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{4;15j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{5;27j}^{Deces}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{6;2m}^{Deces}$	0,01	-	-	-	-	-	-	-	-	-
$\tilde{C}_{7;3m}^{Deces}$	-0,01	-	-	-0,01	-	-	-	-	-	-

Tableau C. 10 : Coefficients obtenus pour les modèles toutes saisons de la R-EMD2 des hospitalisations de la CMQ. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	-0,07	-	-	-	-	-	-	-	-	-
$\tilde{C}_{2;5j}^{Hosp}$	0,03	0,09	-	0,02	-	-	-	-	-	-
$\tilde{C}_{3;7j}^{Hosp}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{4;14j}^{Hosp}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{5;26j}^{Hosp}$	-	-	-	0,11	-	-0,02	-0,30	-	-	-
$\tilde{C}_{6;2m}^{Hosp}$	-	-0,05	-	-0,05	-	0,04	-0,96	-	-0,08	-
$\tilde{C}_{7;3m}^{Hosp}$	-	-	-	-	0,05	-0,01	-	-	-0,14	-
$\tilde{C}_{8;6m}^{Hosp}$	0,01	-	-	-	-	0,02	-	-	-	-
$\tilde{C}_{9;1a}^{Hosp}$	0,03	-	-0,10	-	-0,01	-	-	0,03	-	-
$\tilde{C}_{10;3a}^{Hosp}$	0,01	-	-	-	-1,82	-	-1,75	-	-	-0,11
$\tilde{C}_{11;6a}^{Hosp}$	0,01	-	-2,75	2,95	0,59	0,02	-12,24	0,49	-4,09	2,87
$\tilde{C}_{12;8a}^{Hosp}$	0,05	-	-	0,18	-	-	-30,22	-	-	-
$\tilde{\gamma}^{Hosp}$	-6,36	-	-	-	-	-0,53	-	-	-	0,60

Tableau C. 11 : Coefficients obtenus pour les modèles estivaux de la R-EMD2 des hospitalisations de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

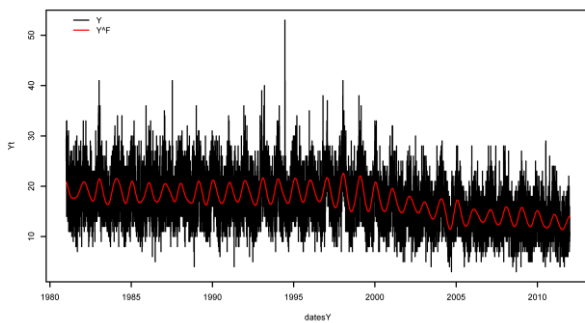
	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	-0,05	-	-	-	-	-	-	-	-	-
$\tilde{C}_{2;5j}^{Hosp}$	0,01	0,20	-	-	0,06	-	0,28	-	-0,02	-
$\tilde{C}_{3;7j}^{Hosp}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{4;14j}^{Hosp}$	-0,02	-	-	-	-	-	-	-	-	-
$\tilde{C}_{5;26j}^{Hosp}$	-0,03	-	-	-	-	-	-	-	-0,03	-0,01
$\tilde{C}_{6;2m}^{Hosp}$	0,02	-	-	-	-	-	-	-	-	0,04
$\tilde{C}_{7;3m}^{Hosp}$	0,04	-	-0,06	0,03	0,06	-0,02	-	-	-0,06	-

Tableau C. 12 : Coefficients obtenus pour les modèles hivernaux de la R-EMD2 des hospitalisations de la CMM. Le caractère « - » indique que l'IMF n'est pas présent dans le modèle final.

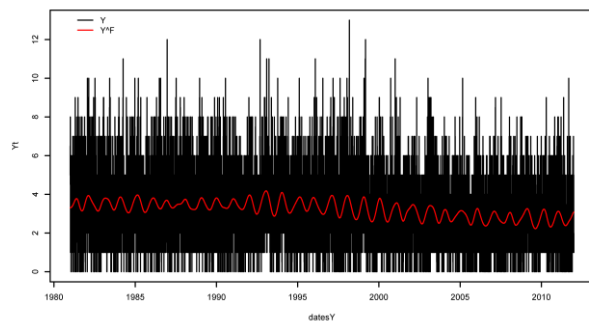
	<i>Intercept</i>	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	-0,08	-	-	-	-	-	0,07	-	-	-
$\tilde{C}_{2;5j}^{Hosp}$	0,06	-	-	0,15	-0,03	-	-	-0,01	-	-
$\tilde{C}_{3;7j}^{Hosp}$	0,02	-	-	0,02	-	-	-0,52	-	-	-
$\tilde{C}_{4;14j}^{Hosp}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{5;26j}^{Hosp}$	-0,01	-	-	-	-	-	-	-	-	-
$\tilde{C}_{6;2m}^{Hosp}$	-	-	-	-	-	-	-	-	-	-
$\tilde{C}_{7;3m}^{Hosp}$	-0,04	-	-	-	-	-	-	-	-	-

D. Filtrage des séries

Cette annexe montre les séries filtrées, qui seront ensuite utilisées dans les trois modèles dont les résultats sont montrés dans les annexes C, D et E. Le filtrage d'une série se réalise en décomposant cette série par EMD puis en déterminant les IMF porteurs d'information par le test de Wu et Huang (2004). Ce test consiste à comparer l'énergie (somme des carrés des valeurs) des IMF de la série avec l'énergie d'IMF qui seraient issus d'un bruit blanc pur. Afin de déterminer le niveau du bruit du signal, Wu et Huang (2004) utilisent l'énergie du premier IMF qui est automatiquement considéré comme du bruit. Le problème est que cette approche rend significativement informatifs quasiment tous les IMF. Ici, le test est donc légèrement modifié en standardisant les IMF par l'écart-type de la série dont ils sont issus. Cela permet de considérer les IMF informatifs uniquement lorsque leur énergie se détache des autres IMF. Les IMF porteurs d'information sont ensuite sommés pour obtenir la série filtrée.

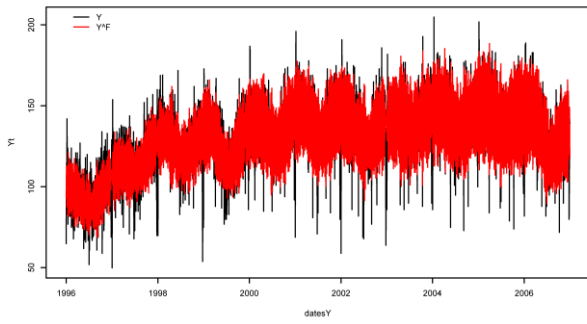


a) CMM

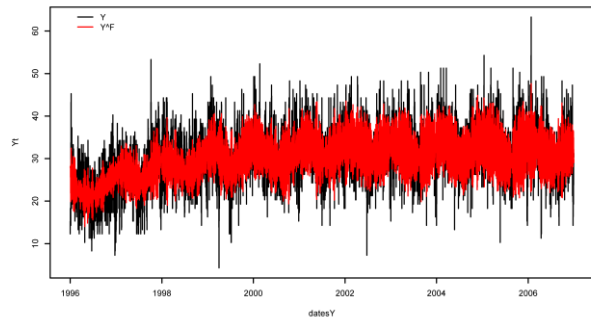


b) CMQ

Figure D. 1 : Décès par MCV filtrés. La ligne noire correspond à la série d'origine et la ligne rouge à la série filtrée. Le cycle annuel modulé ainsi que la tendance sont principalement conservés.

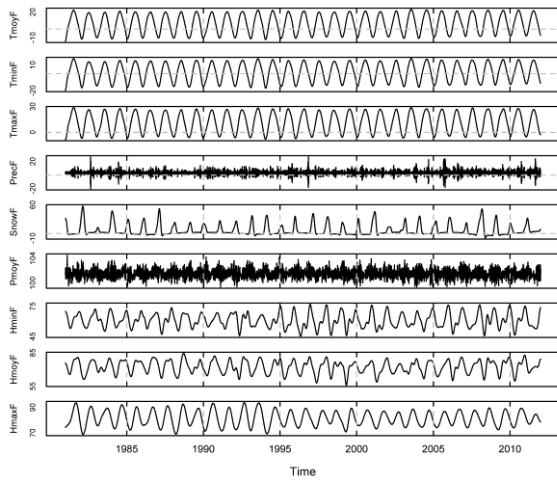


a) CMM

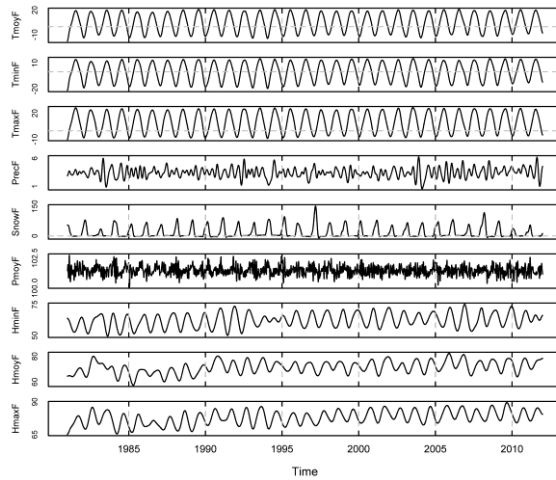


b) CMQ

Figure D. 2 : Hospitalisations filtrées (en rouge) et hospitalisations d'origine (en noir). En plus de la tendance et du cycle annuel, l'effet « fin de semaine » passe le test et correspond donc aux hautes fréquences visibles en rouge.



a) CMM

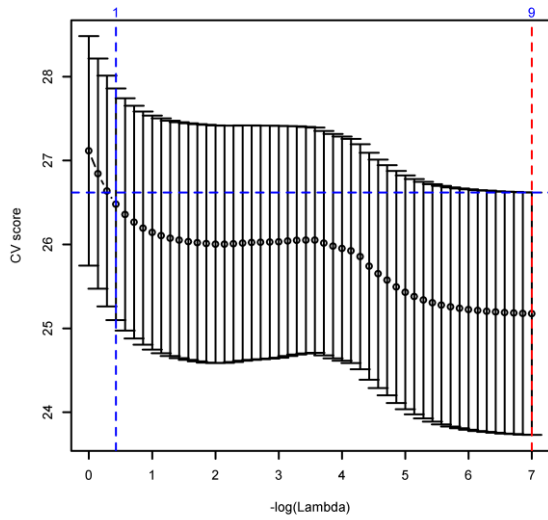


b) CMQ

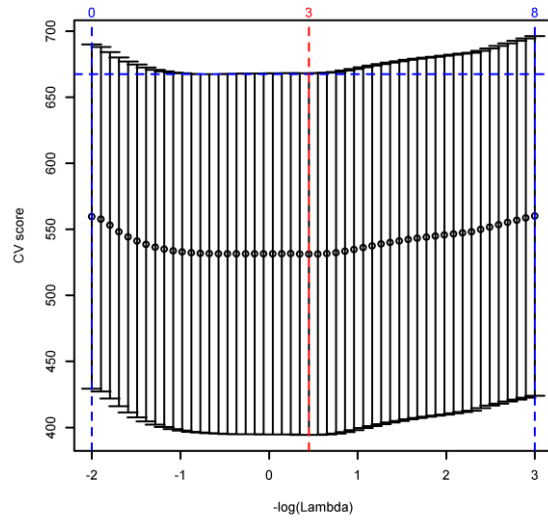
Figure D. 3 : Séries météorologiques filtrées. Les cycles conservés sont principalement le cycle annuel et la tendance. Seules les variables de précipitations et de pression contiennent de l'information dans les hautes fréquences, d'après le test.

E. Résultats R-EMDF1

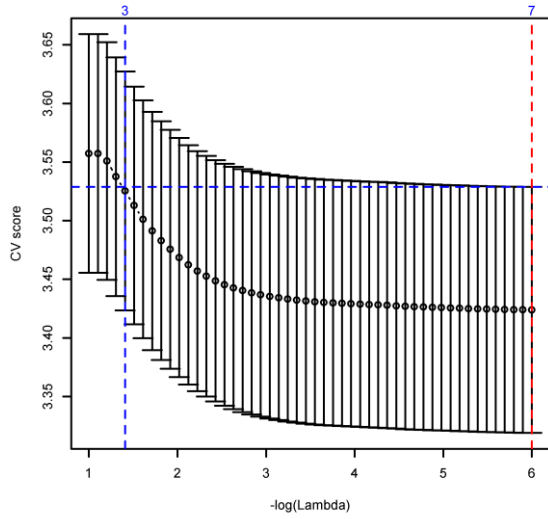
Cette modélisation prédit la réponse sanitaire d'origine (comme dans le R-EMD1) en fonction des variables météorologiques filtrées de la Figure D. 3. Cette modélisation n'apporte pas d'information supplémentaire par rapport à la R-EMD1, car seuls les cycles annuels sont prédits ici aussi. De plus, les variables mises en évidence sont les températures et l'humidité, exactement comme pour la R-EMD1. Les prédictions montrent que cette prédiction est même moins bonne que celle de la R-EMD1. En effet, les différentes échelles temporelles n'étant pas séparées, ce modèle est moins précis que la R-EMD1.



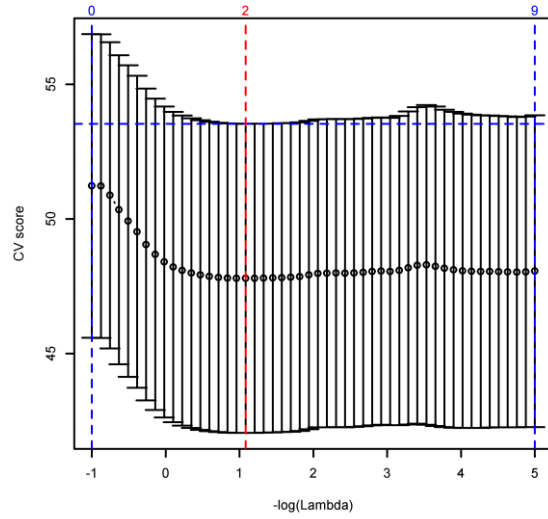
a) Décès de la CMM



b) Hospitalisations de la CMM

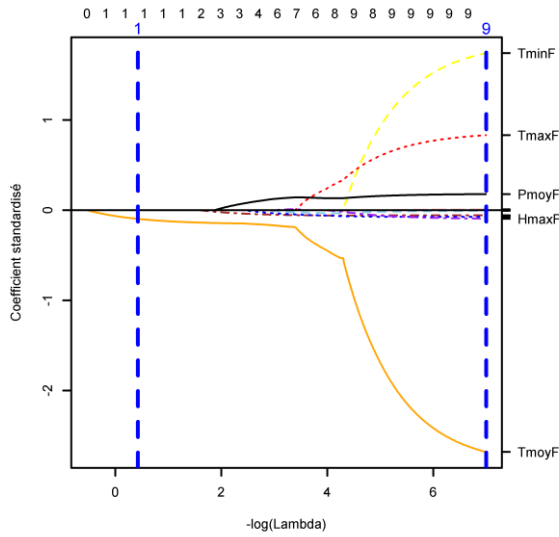


c) Décès de la CMQ

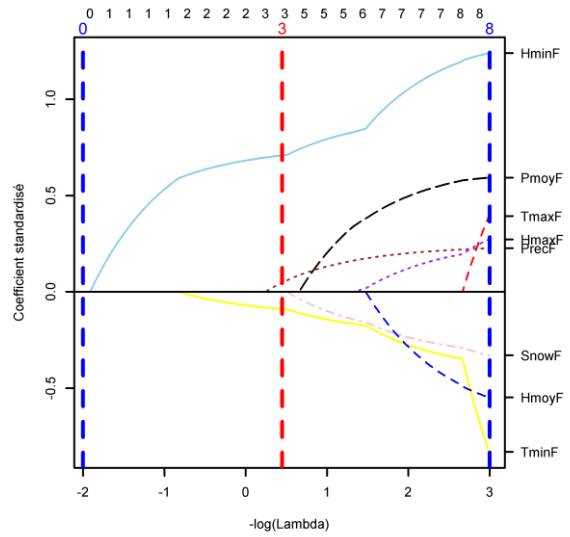


d) Hospitalisations de la CMQ

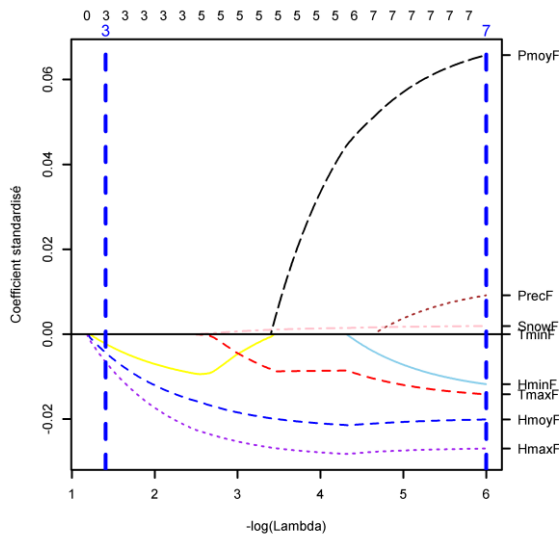
Figure E. 1 : Évolution du score CV avec son erreur type en fonction du paramètre de pénalisation λ . La ligne pointillée rouge indique la valeur minimale de CV et les lignes bleues indiquent les valeurs extrêmes de $-\log(\lambda)$ ayant un CV au sein de l'erreur type du minimum. L'axe supérieur indique le nombre de variables présentes dans les modèles repérés par les lignes pointillées verticales. Le modèle choisi est le plus parcimonieux avec une CV au sein de l'erreur type du minimum, c'est-à-dire la ligne pointillée bleue de gauche.



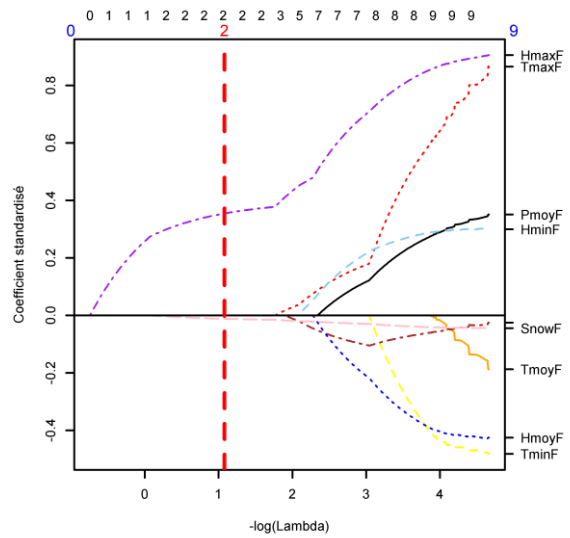
a) Décès de la CMM



b) Hospitalisations de la CMM

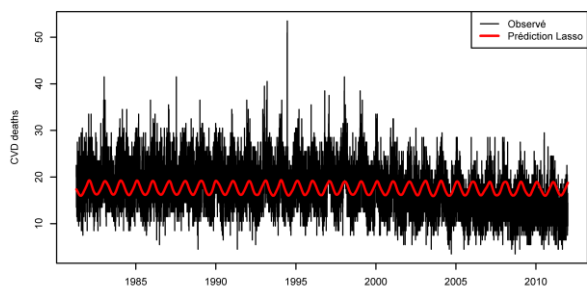


c) Décès de la CMQ

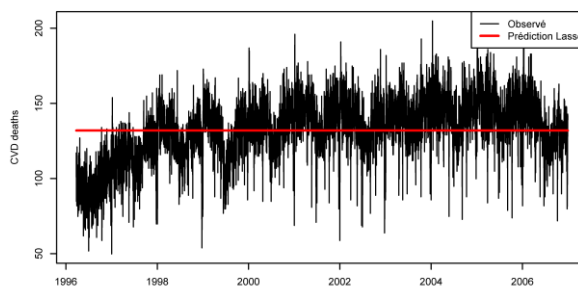


d) Hospitalisations de la CMQ

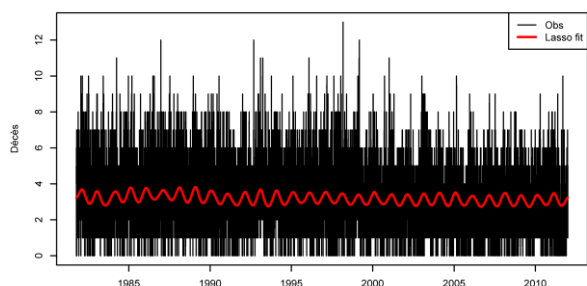
Figure E. 2 : Chemin de pénalisation. L'axe des abscisses correspond à $-\log(\lambda)$ pour que le début du graphe corresponde au modèle le plus simple et la fin au modèle le plus compliqué. Ici, les coefficients sont standardisés, c'est-à-dire qu'ils ne dépendent pas de l'ordre de grandeur des variables auxquelles ils sont associés et peuvent donc être directement comparés. La marge de droite indique les IMF représentés par chaque ligne. La marge supérieure indique le nombre de variables présentes dans le modèle à certaines étapes. Les lignes pointillées verticales indiquent la valeur optimale (en rouge) et les limites des valeurs acceptables de λ (en bleu).



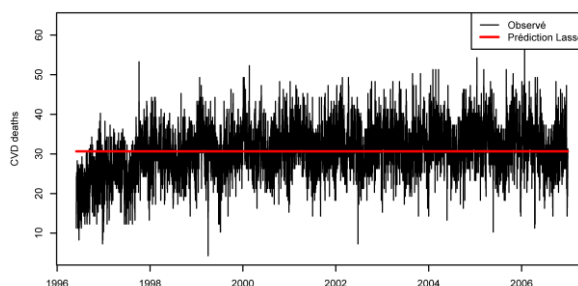
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ



d) Hospitalisations de la CMQ

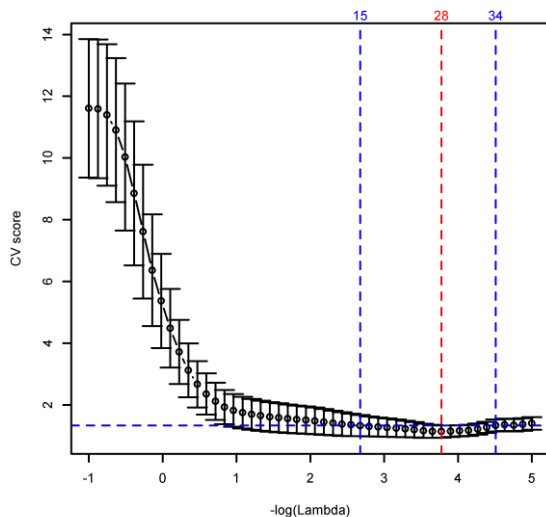
Figure E. 3 : Prédications de la modélisation (en rouge). Les réponses d'origine sont en noir. Pour les décès, la modélisation est capable de prédire uniquement le cycle annuel. Pour les hospitalisations, cette modélisation ne trouve pas d'effet significatif de la météorologie filtrée.

La Figure E. 2 montre le chemin de pénalisation de cette modélisation, pour chacun des quatre cas étudiés. Le modèle choisi est le plus parcimonieux avec un CV au sein de l'erreur type du minimum (délimité par les pointillés bleus les plus à gauche). Cette modélisation met en évidence les différences entre les décès et les hospitalisations. D'un côté, seules les variations annuelles des décès sont expliquées, et ce, notamment par les températures (moyennes pour la CMM et minimales pour la CMQ). La présence des variables d'humidité moyenne et d'humidité maximale dans le modèle de la CMQ permet de moduler les variations annuelles (Figure E. 3). Ainsi l'on retrouve des conclusions similaires qu'avec la R-EMD1. Dans le cas des hospitalisations, le

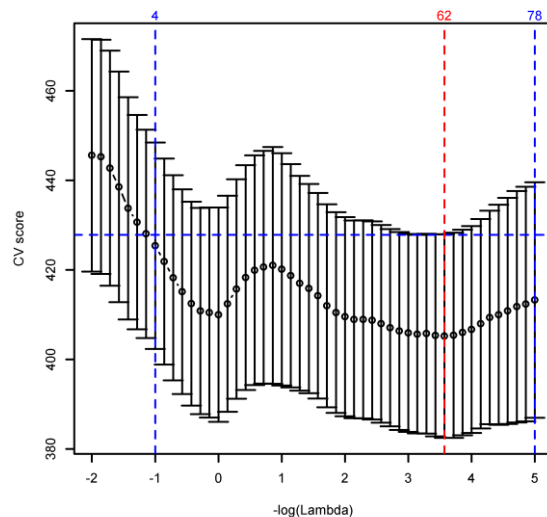
modèle choisi est vide pour la CMM et la CMQ. Ceci suggère un effet de la météorologie difficile à mettre en évidence car masqué par les biais humains. Ainsi, la précision de la R-EMD1 et de la R-EMD2 est nécessaire pour mettre en évidence l'effet de la météorologie trouvé dans les résultats présentés dans la section 4.

F. Résultats R-EMDF2

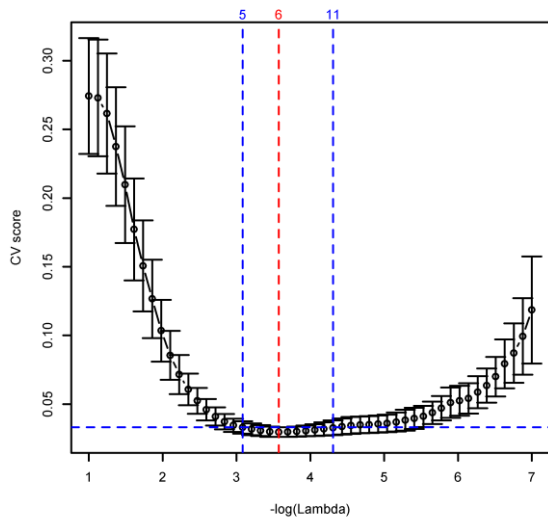
Cette section présente les résultats de la modélisation des réponses sanitaires filtrées en fonction des IMF météorologiques. Comme l'indiquaient la Figure D. 1 et la Figure D. 2, le filtrage des séries sanitaires consiste essentiellement à conserver le motif annuel et la tendance. L'effet « fin de semaine » est également conservé dans le cas des hospitalisations. Il est ainsi également redondant avec la R-EMD1, dans la mesure où les IMF obtenus sont globalement les mêmes que dans la R-EMD1 (Figure F. 3). Comme le cycle annuel est le motif le mieux prédit par la météorologie, ce modèle obtient de très bonnes prédictions (Figure F. 4).



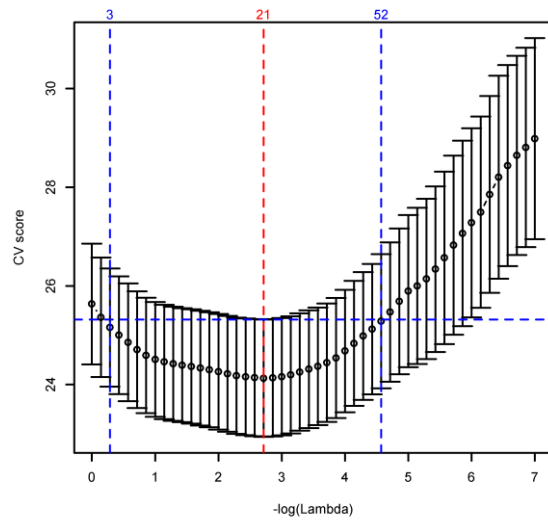
a) Décès de la CMM



b) Hospitalisations de la CMM

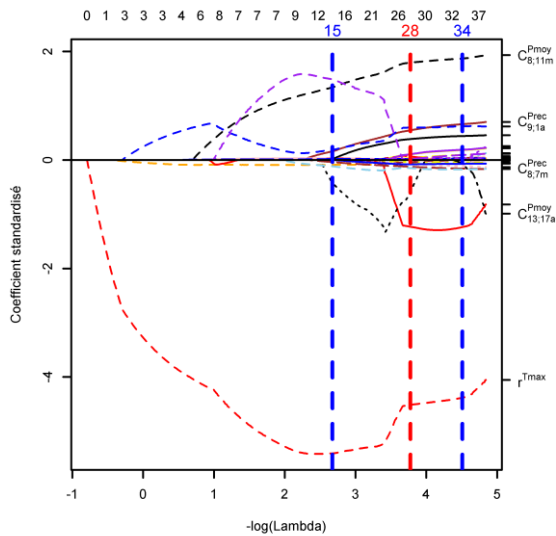


c) Décès de la CMQ

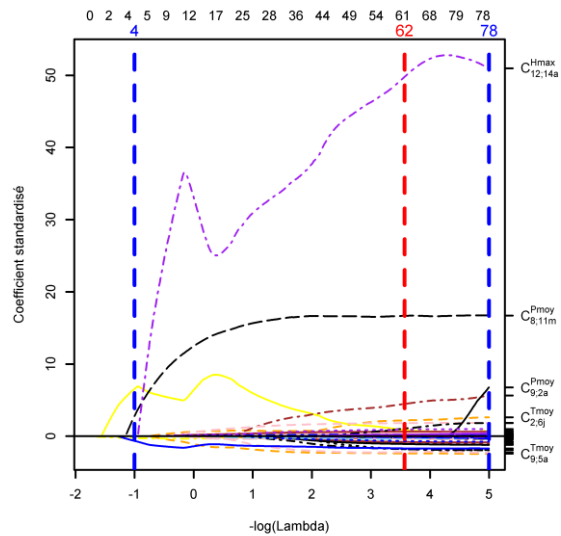


d) Hospitalisations de la CMQ

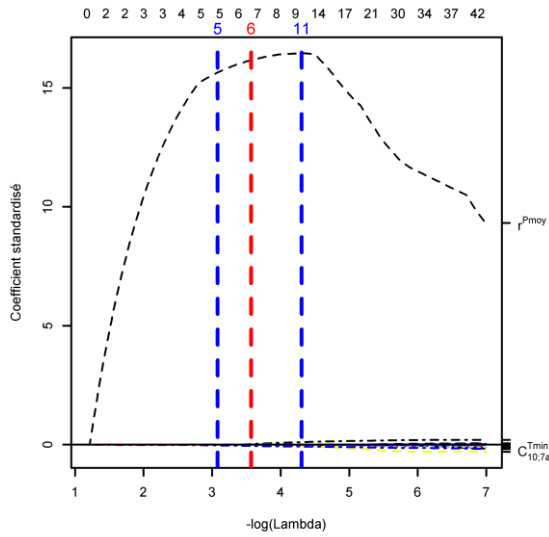
Figure F. 1 : Évolution du score CV avec son erreur type en fonction du paramètre de pénalisation λ . La ligne pointillée rouge indique la valeur minimale de CV et les lignes bleues indiquent les valeurs extrêmes de $-\log(\lambda)$ ayant un CV au sein de l'erreur type du minimum. L'axe supérieur indique le nombre de variables présentes dans les modèles repérés par les lignes pointillées verticales. Le modèle choisi est le plus parcimonieux avec une CV au sein de l'erreur type du minimum, c'est-à-dire la ligne pointillée bleue de gauche.



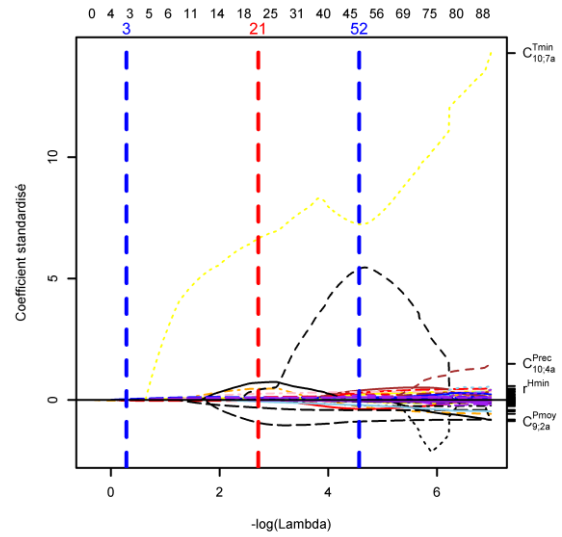
a) Décès de la CMM



b) Hospitalisations de la CMM

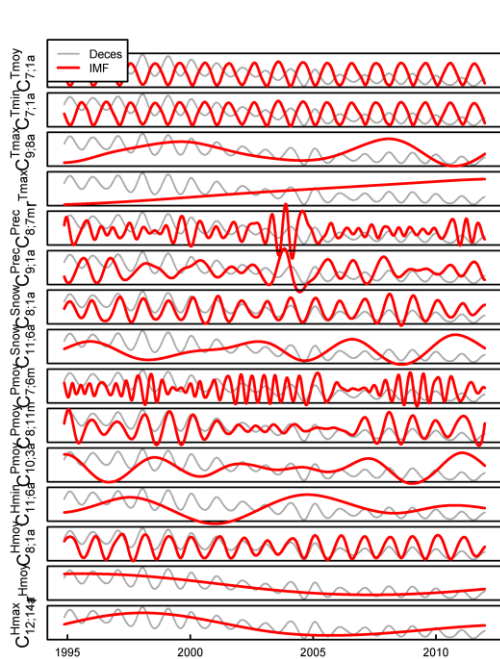


c) Décès de la CMQ

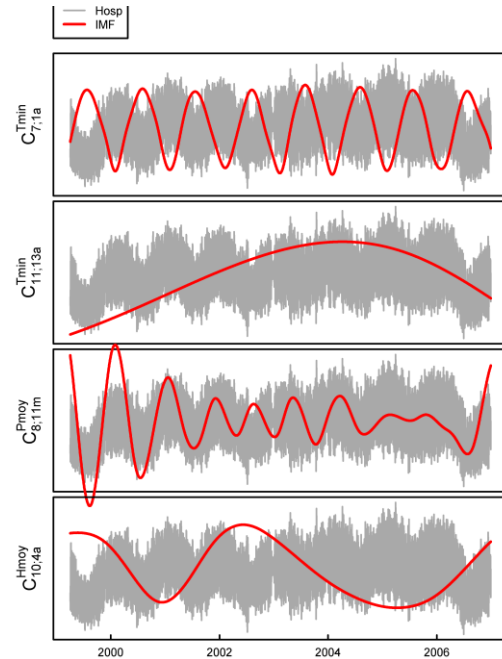


d) Hospitalisations de la CMQ

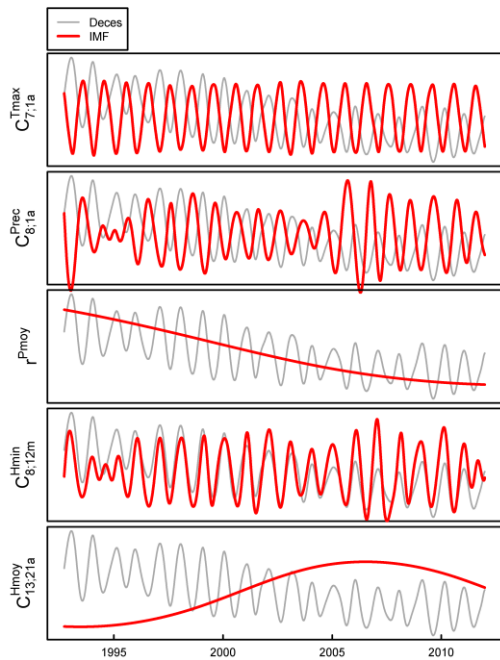
Figure F. 2 : Chemin de pénalisation. L'axe des abscisses correspond à $-\log(\lambda)$ pour que le début du graphe corresponde au modèle le plus simple et la fin au modèle le plus compliqué. Ici, les coefficients sont standardisés, c'est-à-dire qu'ils ne dépendent pas de l'ordre de grandeur des variables auxquelles ils sont associés et peuvent donc être directement comparés. La marge de droite indique les IMF représentés par chaque ligne. La marge supérieure indique le nombre de variables présentes dans le modèle à certaines étapes. Les lignes pointillées verticales indiquent la valeur optimale (en rouge) et les limites des valeurs acceptables de λ (en bleu).



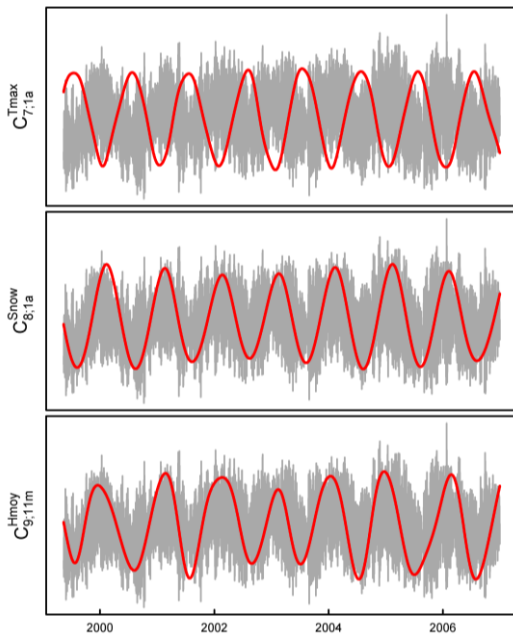
a) Décès de la CMM



b) Hospitalisations de la CMM

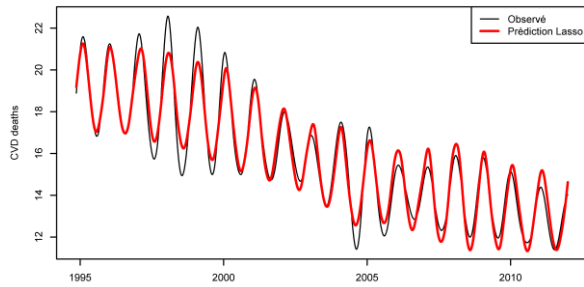


c) Décès de la CMQ

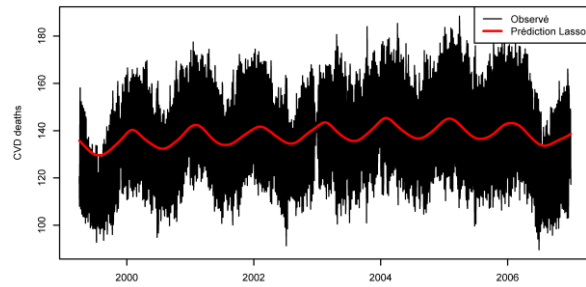


d) Hospitalisations de la CMQ

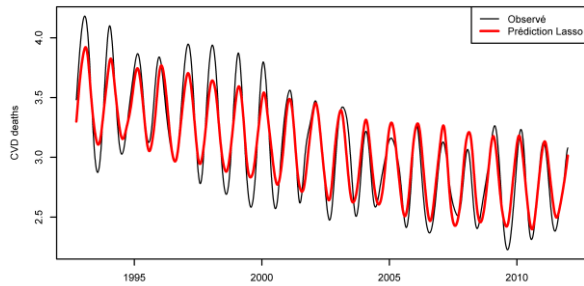
Figure F. 3 : IMF explicatifs présents dans la modélisation (en rouge) et série réponse filtrée (en gris). La série réponse ainsi que les IMF tracés sont standardisés afin de pouvoir les comparer visuellement. Les IMF sont décalés en fonction de leur lag obtenu dans la modélisation afin d'être en phase avec les variations qu'ils expliquent.



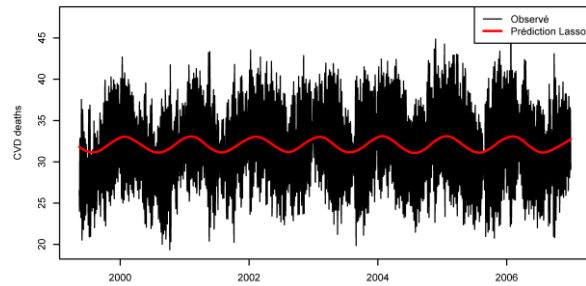
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ

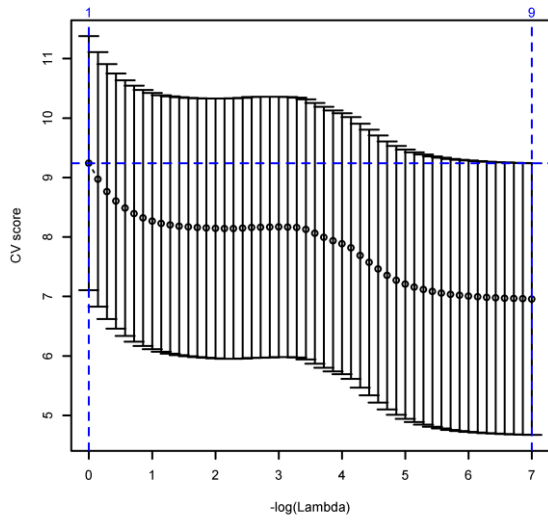


d) Hospitalisations de la CMQ

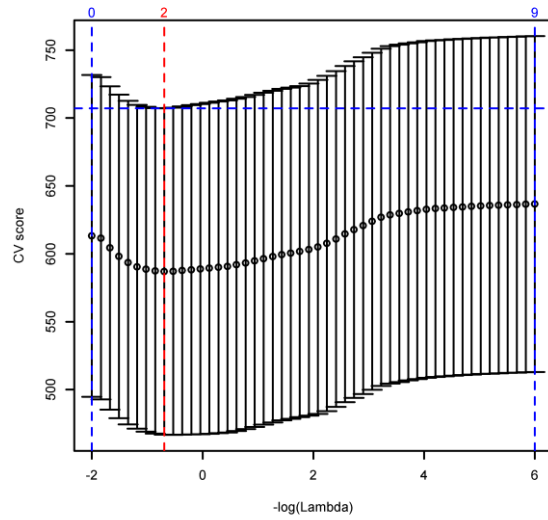
Figure F. 4 : Prédictions de la modélisation (en rouge). Les réponses d'origine sont en noir.

G. Résultats R-EMDF3

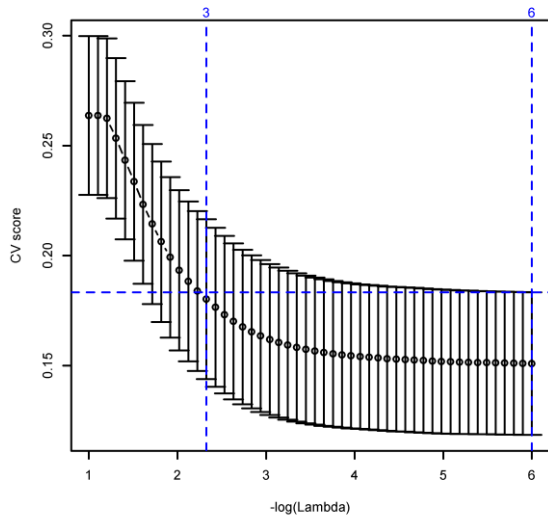
Cette section présente la modélisation des séries sanitaires filtrées (Figure D. 1 et Figure D. 2) en fonction des séries météorologiques filtrées (Figure D. 3). Cette modélisation met uniquement en avant le rôle de la température et de l'humidité dans le cycle annuel des décès et hospitalisations par MCV (Figure G. 2). Ce cycle annuel est donc la seule prédiction apportée par ce modèle (Figure G. 3). Ainsi, ce modèle est une fois encore redondant par rapport à la R-EMD1.



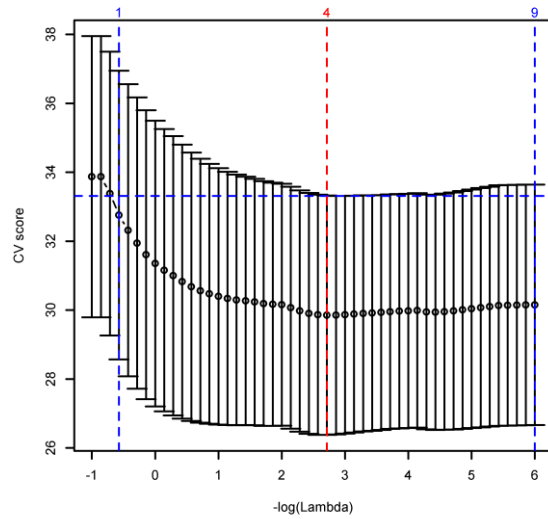
a) Décès de la CMM



b) Hospitalisations de la CMM

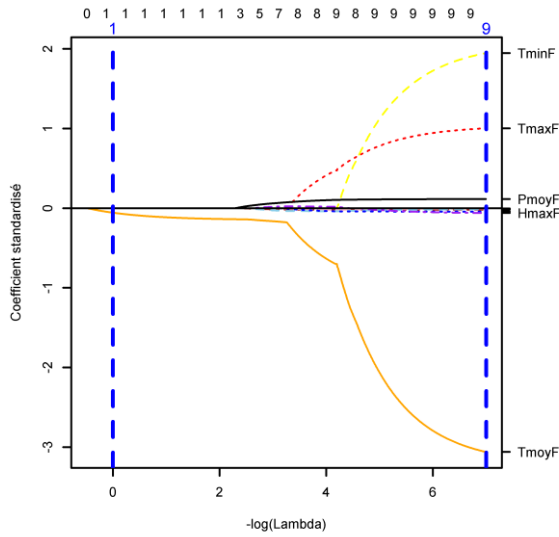


c) Décès de la CMQ

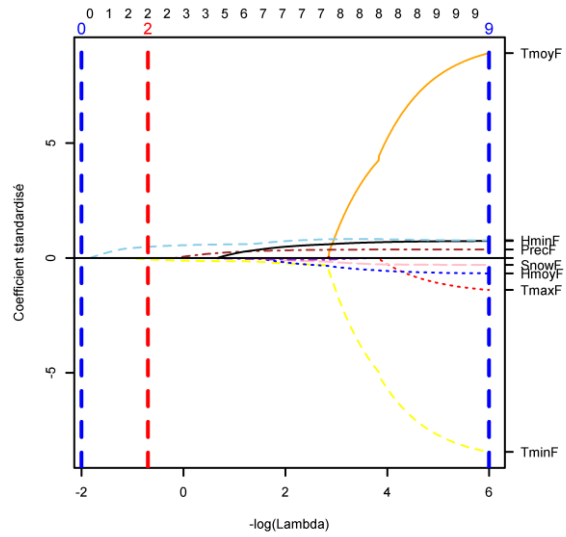


d) Hospitalisations de la CMQ

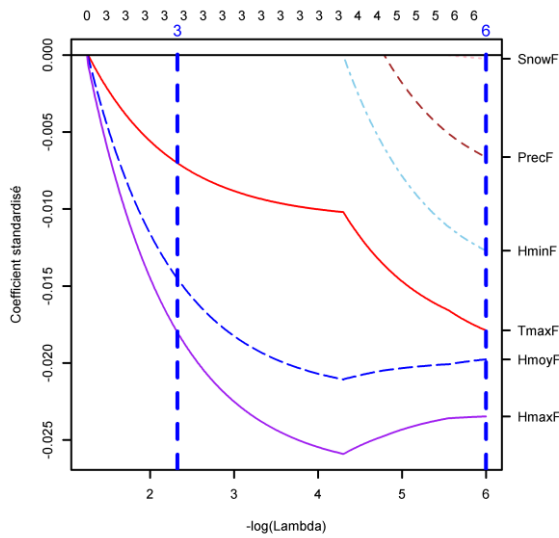
Figure G. 1 : Évolution du score CV avec son erreur type en fonction du paramètre de pénalisation λ . La ligne pointillée rouge indique la valeur minimale de CV et les lignes bleues indiquent les valeurs extrêmes de $-\log(\lambda)$ ayant un CV au sein de l'erreur type du minimum. L'axe supérieur indique le nombre de variables présentes dans les modèles repérés par les lignes pointillées verticales. Le modèle choisi est le plus parcimonieux avec une CV au sein de l'erreur type du minimum, c'est-à-dire la ligne pointillée bleue de gauche.



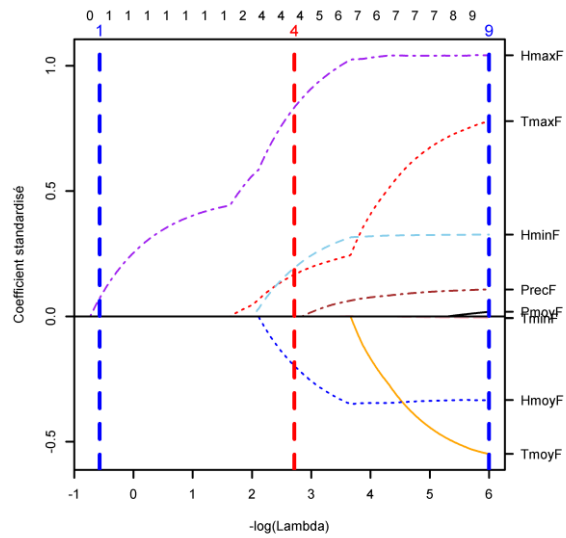
a) Décès de la CMM



b) Hospitalisations de la CMM

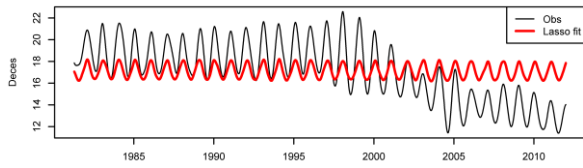


c) Décès de la CMQ

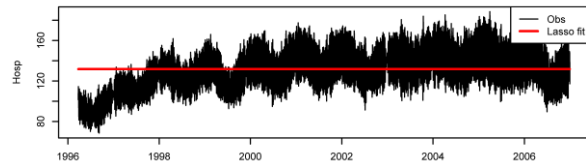


d) Hospitalisations de la CMQ

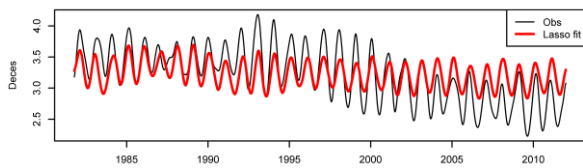
Figure G. 2 : Chemin de pénalisation. L'axe des abscisses correspond à $-\log(\lambda)$ pour que le début du graphe corresponde au modèle le plus simple et la fin au modèle le plus compliqué. Ici, les coefficients sont standardisés, c'est-à-dire qu'ils ne dépendent pas de l'ordre de grandeur des variables auxquelles ils sont associés et peuvent donc être directement comparés. La marge de droite indique les IMF représentés par chaque ligne. La marge supérieure indique le nombre de variables présentes dans le modèle à certaines étapes. Les lignes pointillées verticales indiquent la valeur optimale (en rouge) et les limites des valeurs acceptables de λ (en bleu).



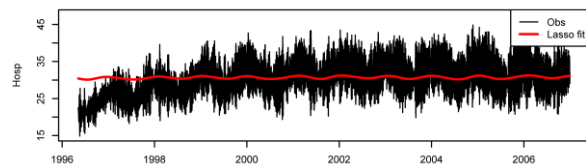
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ



d) Hospitalisations de la CMQ

Figure G. 3 : Prédications de la modélisation (en rouge). Les réponses d'origine sont en noir.

H. Lags utilisés

Comme expliqué dans la section 3.2.3.2, les variables explicatives des modèles sont décalées afin de considérer le retard entre une exposition et la réponse. Cette annexe présente tous les lags des différents modèles utilisés.

Tableau H. 1 : Lags utilisés dans la R-EMD1 des décès de la CMM.

$C^{Tmoy}_{1;3j}$	0	$C^{Tmin}_{1;3j}$	0	$C^{Tmax}_{1;3j}$	1	$C^{Prec}_{1;4j}$	1	$C^{Snow}_{1;3j}$	0	$C^{Pmoy}_{1;4j}$	0	$C^{Hmin}_{1;3j}$	1	$C^{Hmoy}_{1;3j}$	0	$C^{Hmax}_{1;3j}$	0
$C^{Tmoy}_{2;6j}$	1	$C^{Tmin}_{2;6j}$	0	$C^{Tmax}_{2;6j}$	1	$C^{Prec}_{2;6j}$	0	$C^{Snow}_{2;7j}$	2	$C^{Pmoy}_{2;8j}$	0	$C^{Hmin}_{2;6j}$	0	$C^{Hmoy}_{2;6j}$	0	$C^{Hmax}_{2;6j}$	2
$C^{Tmoy}_{3;12j}$	1	$C^{Tmin}_{3;12j}$	0	$C^{Tmax}_{3;12j}$	2	$C^{Prec}_{3;12j}$	0	$C^{Snow}_{3;13j}$	0	$C^{Pmoy}_{3;15j}$	0	$C^{Hmin}_{3;12j}$	3	$C^{Hmoy}_{3;12j}$	0	$C^{Hmax}_{3;12j}$	10
$C^{Tmoy}_{4;22j}$	0	$C^{Tmin}_{4;22j}$	8	$C^{Tmax}_{4;22j}$	1	$C^{Prec}_{4;21j}$	0	$C^{Snow}_{4;26j}$	10	$C^{Pmoy}_{4;29j}$	1	$C^{Hmin}_{4;22j}$	5	$C^{Hmoy}_{4;23j}$	5	$C^{Hmax}_{4;22j}$	6
$C^{Tmoy}_{5;1m}$	9	$C^{Tmin}_{5;1m}$	7	$C^{Tmax}_{5;2m}$	13	$C^{Prec}_{5;1m}$	6	$C^{Snow}_{5;2m}$	15	$C^{Pmoy}_{5;2m}$	21	$C^{Hmin}_{5;1m}$	3	$C^{Hmoy}_{5;1m}$	4	$C^{Hmax}_{5;1m}$	4
$C^{Tmoy}_{6;3m}$	2	$C^{Tmin}_{6;3m}$	2	$C^{Tmax}_{6;3m}$	0	$C^{Prec}_{6;2m}$	0	$C^{Snow}_{6;5m}$	46	$C^{Pmoy}_{6;3m}$	93	$C^{Hmin}_{6;3m}$	0	$C^{Hmoy}_{6;3m}$	0	$C^{Hmax}_{6;2m}$	0
$C^{Tmoy}_{7;1a}$	7	$C^{Tmin}_{7;1a}$	5	$C^{Tmax}_{7;1a}$	12	$C^{Prec}_{7;4m}$	43	$C^{Snow}_{7;6m}$	0	$C^{Pmoy}_{7;6m}$	112	$C^{Hmin}_{7;5m}$	13	$C^{Hmoy}_{7;5m}$	0	$C^{Hmax}_{7;5m}$	24
$C^{Tmoy}_{8;2a}$	132	$C^{Tmin}_{8;2a}$	146	$C^{Tmax}_{8;3a}$	0	$C^{Prec}_{8;7m}$	0	$C^{Snow}_{8;1a}$	0	$C^{Pmoy}_{8;1a}$	40	$C^{Hmin}_{8;1a}$	60	$C^{Hmoy}_{8;1a}$	103	$C^{Hmax}_{8;1a}$	0
$C^{Tmoy}_{9;5a}$	470	$C^{Tmin}_{9;3a}$	841	$C^{Tmax}_{9;8a}$	569	$C^{Prec}_{9;1a}$	0	$C^{Snow}_{9;2a}$	0	$C^{Pmoy}_{9;2a}$	53	$C^{Hmin}_{9;2a}$	0	$C^{Hmoy}_{9;2a}$	0	$C^{Hmax}_{9;2a}$	541
$C^{Tmoy}_{10;13a}$	781	$C^{Tmin}_{10;7a}$	1422	$C^{Tmax}_{10;14a}$	1963	$C^{Prec}_{10;2a}$	141	$C^{Snow}_{10;3a}$	541	$C^{Pmoy}_{10;3a}$	1044	$C^{Hmin}_{10;3a}$	298	$C^{Hmoy}_{10;3a}$	253	$C^{Hmax}_{10;4a}$	233
r^{Tmoy}	0	$C^{Tmin}_{11;13a}$	781	r^{Tmax}	0	$C^{Prec}_{11;5a}$	642	$C^{Snow}_{11;6a}$	781	$C^{Pmoy}_{11;5a}$	2631	$C^{Hmin}_{11;6a}$	1899	$C^{Hmoy}_{11;8a}$	106	$C^{Hmax}_{11;7a}$	331
		r^{Tmin}	0			$C^{Prec}_{12;10a}$	2150	$C^{Snow}_{12;9a}$	541	$C^{Pmoy}_{12;8a}$	298	$C^{Hmin}_{12;10a}$	2022	r^{Hmoy}	2903	$C^{Hmax}_{12;14a}$	3375
						r^{Prec}	0	r^{Snow}	5056	$C^{Pmoy}_{13;17a}$	713	r^{Hmin}	3			r^{Hmax}	2464
										r^{Pmoy}	8						

Tableau H. 2 : Lags utilisés dans la R-EMD2 des décès de la CMM (modèles non saisonniers). Chaque ligne correspond à un modèle, désigné par l'IMF sanitaire réponse. Il y a un IMF de chaque variable météorologique dans chaque modèle.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	1	0	1	0	0	1	0	1	0
$\tilde{C}_{2;5j}^{Deces}$	0	0	0	0	1	1	2	0	2
$\tilde{C}_{3;9j}^{Deces}$	0	0	2	0	1	2	2	1	0
$\tilde{C}_{4;15j}^{Deces}$	0	0	1	0	0	4	5	5	0
$\tilde{C}_{5;1m}^{Deces}$	7	6	0	4	9	0	3	3	3
$\tilde{C}_{6;2m}^{Deces}$	9	7	13	6	12	0	2	3	4
$\tilde{C}_{7;3m}^{Deces}$	0	0	0	32	31	3	0	0	0
$\tilde{C}_{8;5m}^{Deces}$	0	0	0	39	0	47	11	2	35
$\tilde{C}_{9;1a}^{Deces}$	10	8	11	0	0	46	60	103	0
$\tilde{C}_{10;2a}^{Deces}$	0	0	0	85	0	65	0	0	46
$\tilde{C}_{11;5a}^{Deces}$	219	168	256	0	64	16	0	403	430
$\tilde{C}_{12;7a}^{Deces}$	724	730	720	180	685	480	634	0	0
$\tilde{C}_{13;13a}^{Deces}$	0	0	1341	872	117	121	664	270	0
$\tilde{C}_{14;26a}^{Deces}$	1379	962	1759	2117	821	883	0	0	0
\tilde{r}^{Deces}	0	0	0	0	441	0	0	0	0

Tableau H. 3 : Lags utilisés dans les modèles estivaux de la R-EMD2 des décès de la CMM.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	0	0	0	0	1	1	1	1	0
$\tilde{C}_{2;5j}^{Deces}$	0	2	0	1	0	1	1	1	2
$\tilde{C}_{3;9j}^{Deces}$	3	0	3	2	0	0	2	2	2
$\tilde{C}_{4;15j}^{Deces}$	3	2	4	0	0	5	4	4	3
$\tilde{C}_{5;1m}^{Deces}$	5	5	6	0	0	7	10	8	7
$\tilde{C}_{6;2m}^{Deces}$	10	6	12	11	9	6	0	4	10
$\tilde{C}_{7;3m}^{Deces}$	2	0	4	32	20	10	23	27	17

Tableau H. 4 : Lags utilisés dans les modèles hivernaux de la R-EMD2 des décès de la CMM.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	0	0	0	1	1	0	1	1	0
$\tilde{C}_{2;5j}^{Deces}$	1	0	1	1	0	1	1	0	1
$\tilde{C}_{3;9j}^{Deces}$	1	1	1	0	0	0	0	0	2
$\tilde{C}_{4;15j}^{Deces}$	0	0	0	0	6	4	10	7	0
$\tilde{C}_{5;1m}^{Deces}$	6	5	8	6	3	3	6	6	7
$\tilde{C}_{6;2m}^{Deces}$	15	17	12	0	10	0	0	0	0
$\tilde{C}_{7;3m}^{Deces}$	17	17	18	0	0	44	46	47	5

Tableau H. 5 : Lags utilisés dans la R-EMD1 des hospitalisations de la CMM.

$C^{Tmoy}_{1;3j}$	1	$C^{Tmin}_{1;3j}$	1	$C^{Tmax}_{1;3j}$	1	$C^{Prec}_{1;4j}$	0	$C^{Snow}_{1;3j}$	1	$C^{Pmoy}_{1;4j}$	1	$C^{Hmin}_{1;3j}$	0	$C^{Hmoy}_{1;3j}$	0	$C^{Hmax}_{1;3j}$	0
$C^{Tmoy}_{2;6j}$	1	$C^{Tmin}_{2;6j}$	1	$C^{Tmax}_{2;6j}$	0	$C^{Prec}_{2;6j}$	1	$C^{Snow}_{2;7j}$	2	$C^{Pmoy}_{2;8j}$	14	$C^{Hmin}_{2;6j}$	2	$C^{Hmoy}_{2;6j}$	2	$C^{Hmax}_{2;6j}$	1
$C^{Tmoy}_{3;12j}$	0	$C^{Tmin}_{3;12j}$	0	$C^{Tmax}_{3;12j}$	0	$C^{Prec}_{3;12j}$	5	$C^{Snow}_{3;13j}$	3	$C^{Pmoy}_{3;15j}$	0	$C^{Hmin}_{3;12j}$	2	$C^{Hmoy}_{3;12j}$	2	$C^{Hmax}_{3;12j}$	1
$C^{Tmoy}_{4;22j}$	0	$C^{Tmin}_{4;22j}$	0	$C^{Tmax}_{4;22j}$	0	$C^{Prec}_{4;21j}$	11	$C^{Snow}_{4;26j}$	11	$C^{Pmoy}_{4;29j}$	7	$C^{Hmin}_{4;22j}$	0	$C^{Hmoy}_{4;23j}$	0	$C^{Hmax}_{4;22j}$	0
$C^{Tmoy}_{5;1m}$	10	$C^{Tmin}_{5;1m}$	10	$C^{Tmax}_{5;2m}$	10	$C^{Prec}_{5;1m}$	1	$C^{Snow}_{5;2m}$	45	$C^{Pmoy}_{5;2m}$	9	$C^{Hmin}_{5;1m}$	1	$C^{Hmoy}_{5;1m}$	4	$C^{Hmax}_{5;1m}$	9
$C^{Tmoy}_{6;3m}$	12	$C^{Tmin}_{6;3m}$	10	$C^{Tmax}_{6;3m}$	60	$C^{Prec}_{6;2m}$	6	$C^{Snow}_{6;5m}$	93	$C^{Pmoy}_{6;3m}$	1	$C^{Hmin}_{6;3m}$	6	$C^{Hmoy}_{6;3m}$	12	$C^{Hmax}_{6;2m}$	7
$C^{Tmoy}_{7;1a}$	6	$C^{Tmin}_{7;1a}$	6	$C^{Tmax}_{7;1a}$	6	$C^{Prec}_{7;4m}$	15	$C^{Snow}_{7;6m}$	65	$C^{Pmoy}_{7;6m}$	26	$C^{Hmin}_{7;5m}$	8	$C^{Hmoy}_{7;5m}$	6	$C^{Hmax}_{7;5m}$	57
$C^{Tmoy}_{8;2a}$	0	$C^{Tmin}_{8;2a}$	0	$C^{Tmax}_{8;3a}$	0	$C^{Prec}_{8;7m}$	158	$C^{Snow}_{8;1a}$	1	$C^{Pmoy}_{8;1a}$	26	$C^{Hmin}_{8;1a}$	61	$C^{Hmoy}_{8;1a}$	88	$C^{Hmax}_{8;1a}$	0
$C^{Tmoy}_{9;5a}$	399	$C^{Tmin}_{9;3a}$	0	$C^{Tmax}_{9;8a}$	0	$C^{Prec}_{9;1a}$	0	$C^{Snow}_{9;2a}$	168	$C^{Pmoy}_{9;2a}$	1	$C^{Hmin}_{9;2a}$	205	$C^{Hmoy}_{9;2a}$	203	$C^{Hmax}_{9;2a}$	539
$C^{Tmoy}_{10;13a}$	0	$C^{Tmin}_{10;7a}$	0	$C^{Tmax}_{10;14a}$	28	$C^{Prec}_{10;2a}$	1	$C^{Snow}_{10;3a}$	588	$C^{Pmoy}_{10;3a}$	21	$C^{Hmin}_{10;3a}$	733	$C^{Hmoy}_{10;3a}$	599	$C^{Hmax}_{10;4a}$	0
r^{Tmoy}	0	$C^{Tmin}_{11;13a}$	1175	r^{Tmax}	0	$C^{Prec}_{11;5a}$	829	$C^{Snow}_{11;6a}$	168	$C^{Pmoy}_{11;5a}$	28	$C^{Hmin}_{11;6a}$	42	$C^{Hmoy}_{11;8a}$	28	$C^{Hmax}_{11;7a}$	0
		r^{Tmin}	0			$C^{Prec}_{12;10a}$	42	$C^{Snow}_{12;9a}$	0	$C^{Pmoy}_{12;8a}$	0	$C^{Hmin}_{12;10a}$	0	r^{Hmoy}	0	$C^{Hmax}_{12;14a}$	0
						r^{Prec}	0	r^{Snow}	0	$C^{Pmoy}_{13;17a}$	0	r^{Hmin}	0			r^{Hmax}	155
										r^{Pmoy}	0						

Tableau H. 6 : Lags utilisés dans la R-EMD2 des hospitalisations de la CMM (modèles non saisonniers).

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	0	0	0	0	0	0	0	1	1
$\tilde{C}_{2;5j}^{Hosp}$	0	0	0	2	0	0	1	1	0
$\tilde{C}_{3;7j}^{Hosp}$	0	0	0	2	0	2	2	0	0
$\tilde{C}_{4;15j}^{Hosp}$	0	5	0	0	2	0	8	8	8
$\tilde{C}_{5;26j}^{Hosp}$	0	0	0	0	9	7	0	0	0
$\tilde{C}_{6;2m}^{Hosp}$	9	7	10	2	19	11	0	0	8
$\tilde{C}_{7;3m}^{Hosp}$	17	12	23	3	15	0	6	7	6
$\tilde{C}_{8;6m}^{Hosp}$	0	0	1	10	67	22	7	0	49
$\tilde{C}_{9;1a}^{Hosp}$	9	7	10	0	0	35	51	83	0
$\tilde{C}_{10;2a}^{Hosp}$	152	159	145	104	0	0	10	0	0
$\tilde{C}_{11;4a}^{Hosp}$	0	0	0	0	235	0	257	183	77
$\tilde{C}_{12;5a}^{Hosp}$	421	293	524	370	0	0	83	103	207
\tilde{r}^{Hosp}	0	0	0	811	0	213	0	481	263

Tableau H. 7 : Lags utilisés dans les modèles estivaux de la R-EMD2 des hospitalisations de la CMM.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	0	0	0	0	1	0	0	0	1
$\tilde{C}_{2;5j}^{Hosp}$	3	0	1	0	0	0	0	0	0
$\tilde{C}_{3;7j}^{Hosp}$	1	1	1	2	0	0	2	1	0
$\tilde{C}_{4;15j}^{Hosp}$	7	2	1	3	5	2	3	3	0
$\tilde{C}_{5;26j}^{Hosp}$	8	8	8	0	0	1	8	6	2
$\tilde{C}_{6;2m}^{Hosp}$	26	22	11	0	0	16	14	12	8
$\tilde{C}_{7;3m}^{Hosp}$	18	8	0	0	22	3	10	10	10

Tableau H. 8 : Lags utilisés dans les modèles hivernaux de la R-EMD2 des hospitalisations de la CMM.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	0	0	0	0	1	2	1	0	2
$\tilde{C}_{2;5j}^{Hosp}$	1	1	0	2	2	1	1	1	1
$\tilde{C}_{3;7j}^{Hosp}$	1	1	1	1	0	0	2	1	1
$\tilde{C}_{4;15j}^{Hosp}$	4	3	0	5	3	4	4	5	5
$\tilde{C}_{5;26j}^{Hosp}$	0	1	0	3	0	7	7	6	6
$\tilde{C}_{6;2m}^{Hosp}$	13	14	11	16	11	0	0	0	19
$\tilde{C}_{7;3m}^{Hosp}$	50	50	6	16	21	1	31	34	5

Tableau H. 9 : Lags utilisés dans la R-EMD1 des décès de la CMQ.

$C^{Tmoy}_{1;3j}$	0	$C^{Tmin}_{1;3j}$	0	$C^{Tmax}_{1;3j}$	1	$C^{Prec}_{1;3j}$	1	$C^{Snow}_{1;3j}$	0	$C^{Pmoy}_{1;4j}$	0	$C^{Hmin}_{1;3j}$	1	$C^{Hmoy}_{1;3j}$	0	$C^{Hmax}_{1;3j}$	0
$C^{Tmoy}_{2;6j}$	1	$C^{Tmin}_{2;6j}$	1	$C^{Tmax}_{2;6j}$	1	$C^{Prec}_{2;6j}$	1	$C^{Snow}_{2;6j}$	0	$C^{Pmoy}_{2;8j}$	0	$C^{Hmin}_{2;6j}$	3	$C^{Hmoy}_{2;6j}$	1	$C^{Hmax}_{2;6j}$	0
$C^{Tmoy}_{3;12j}$	3	$C^{Tmin}_{3;12j}$	3	$C^{Tmax}_{3;12j}$	3	$C^{Prec}_{3;12j}$	0	$C^{Snow}_{3;12j}$	4	$C^{Pmoy}_{3;15j}$	0	$C^{Hmin}_{3;12j}$	4	$C^{Hmoy}_{3;12j}$	3	$C^{Hmax}_{3;12j}$	0
$C^{Tmoy}_{4;22j}$	0	$C^{Tmin}_{4;22j}$	0	$C^{Tmax}_{4;22j}$	0	$C^{Prec}_{4;22j}$	0	$C^{Snow}_{4;24j}$	0	$C^{Pmoy}_{4;28j}$	0	$C^{Hmin}_{4;23j}$	6	$C^{Hmoy}_{4;22j}$	8	$C^{Hmax}_{4;22j}$	11
$C^{Tmoy}_{5;1m}$	1	$C^{Tmin}_{5;1m}$	0	$C^{Tmax}_{5;1m}$	10	$C^{Prec}_{5;1m}$	0	$C^{Snow}_{5;2m}$	0	$C^{Pmoy}_{5;2m}$	37	$C^{Hmin}_{5;1m}$	0	$C^{Hmoy}_{5;1m}$	8	$C^{Hmax}_{5;1m}$	7
$C^{Tmoy}_{6;3m}$	0	$C^{Tmin}_{6;3m}$	0	$C^{Tmax}_{6;3m}$	0	$C^{Prec}_{6;2m}$	6	$C^{Snow}_{6;5m}$	51	$C^{Pmoy}_{6;3m}$	3	$C^{Hmin}_{6;3m}$	19	$C^{Hmoy}_{6;2m}$	0	$C^{Hmax}_{6;3m}$	0
$C^{Tmoy}_{7;1a}$	7	$C^{Tmin}_{7;1a}$	2	$C^{Tmax}_{7;1a}$	13	$C^{Prec}_{7;5m}$	33	$C^{Snow}_{7;6m}$	0	$C^{Pmoy}_{7;6m}$	50	$C^{Hmin}_{7;5m}$	34	$C^{Hmoy}_{7;4m}$	0	$C^{Hmax}_{7;5m}$	0
$C^{Tmoy}_{8;2a}$	0	$C^{Tmin}_{8;2a}$	0	$C^{Tmax}_{8;2a}$	199	$C^{Prec}_{8;1a}$	0	$C^{Snow}_{8;1a}$	0	$C^{Pmoy}_{8;1a}$	51	$C^{Hmin}_{8;11m}$	51	$C^{Hmoy}_{8;7m}$	34	$C^{Hmax}_{8;1a}$	0
$C^{Tmoy}_{9;5a}$	903	$C^{Tmin}_{9;4a}$	915	$C^{Tmax}_{9;4a}$	0	$C^{Prec}_{9;2a}$	99	$C^{Snow}_{9;2a}$	339	$C^{Pmoy}_{9;2a}$	46	$C^{Hmin}_{9;2a}$	207	$C^{Hmoy}_{9;11m}$	100	$C^{Hmax}_{9;2a}$	481
$C^{Tmoy}_{10;10a}$	1839	$C^{Tmin}_{10;7a}$	1824	$C^{Tmax}_{10;7a}$	876	$C^{Prec}_{10;4a}$	0	$C^{Snow}_{10;3a}$	175	$C^{Pmoy}_{10;4a}$	175	$C^{Hmin}_{10;4a}$	1090	$C^{Hmoy}_{10;2a}$	278	$C^{Hmax}_{10;3a}$	199
r^{Tmoy}	0	$C^{Tmin}_{11;13a}$	903	$C^{Tmax}_{11;14a}$	481	$C^{Prec}_{11;6a}$	501	$C^{Snow}_{11;5a}$	1191	$C^{Pmoy}_{11;10a}$	0	$C^{Hmin}_{11;6a}$	481	$C^{Hmoy}_{11;4a}$	681	$C^{Hmax}_{11;4a}$	481
		r^{Tmin}	0	r^{Tmax}		$C^{Prec}_{12;12a}$	681	$C^{Snow}_{12;9a}$	903	$C^{Pmoy}_{12;16a}$	3376	$C^{Hmin}_{12;13a}$	1073	$C^{Hmoy}_{12;11a}$	1824	$C^{Hmax}_{12;7a}$	262
						r^{Prec}	0	r^{Snow}	0	r^{Pmoy}	0	r^{Hmin}	1824	$C^{Hmoy}_{13;20a}$	4254	$C^{Hmax}_{13;9a}$	1422
													r^{Hmoy}	355	$C^{Hmax}_{14;15a}$	5	
															r^{Hmax}	355	

Tableau H. 10 : Lags utilisés dans la R-EMD2 des décès de la CMQ (modèles non saisonniers).

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	0	0	1	0	0	0	1	0	0
$\tilde{C}_{2;5j}^{Deces}$	0	0	0	0	1	0	0	0	0
$\tilde{C}_{3;9j}^{Deces}$	0	0	0	0	1	0	0	0	0
$\tilde{C}_{4;15j}^{Deces}$	2	7	2	4	4	3	3	3	3
$\tilde{C}_{5;27j}^{Deces}$	0	0	0	7	0	0	6	7	7
$\tilde{C}_{6;2m}^{Deces}$	6	1	9	0	5	3	2	28	23
$\tilde{C}_{7;3m}^{Deces}$	0	0	0	17	0	0	24	0	0
$\tilde{C}_{8;5m}^{Deces}$	14	11	20	32	0	69	8	0	0
$\tilde{C}_{9;1a}^{Deces}$	8	6	10	0	0	43	54	102	0
$\tilde{C}_{10;2a}^{Deces}$	84	81	85	45	227	46	0	0	0
$\tilde{C}_{11;4a}^{Deces}$	0	0	0	560	0	0	1060	0	0
$\tilde{C}_{12;8a}^{Deces}$	563	528	609	295	0	176	822	835	845
$\tilde{C}_{13;14a}^{Deces}$	1310	1342	1281	640	390	0	1086	1286	1430
$\tilde{C}_{14;25a}^{Deces}$	963	717	1157	1698	0	0	0	0	0
\tilde{r}^{Deces}	0	0	0	0	3595	0	0	0	0

Tableau H. 11 : Lags utilisés dans les modèles estivaux de la R-EMD2 des décès de la CMQ.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	0	0	0	0	0	0	0	1	0
$\tilde{C}_{2;5j}^{Deces}$	0	0	0	1	0	0	3	0	0
$\tilde{C}_{3;9j}^{Deces}$	0	0	0	2	0	2	1	1	0
$\tilde{C}_{4;15j}^{Deces}$	0	5	1	4	4	0	4	5	0
$\tilde{C}_{5;27j}^{Deces}$	0	0	0	22	7	0	0	0	0
$\tilde{C}_{6;2m}^{Deces}$	15	0	13	10	20	26	6	4	2
$\tilde{C}_{7;3m}^{Deces}$	9	6	10	22	28	11	0	0	0

Tableau H. 12 : Lags utilisés dans les modèles hivernaux de la R-EMD2 des décès de la CMQ.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Deces}$	0	1	0	0	0	0	0	0	0
$\tilde{C}_{2;5j}^{Deces}$	0	0	0	0	1	0	0	1	0
$\tilde{C}_{3;9j}^{Deces}$	1	1	2	1	2	2	0	0	2
$\tilde{C}_{4;15j}^{Deces}$	1	1	1	0	7	0	4	1	2
$\tilde{C}_{5;27j}^{Deces}$	0	6	0	8	0	0	22	24	0
$\tilde{C}_{6;2m}^{Deces}$	3	4	2	22	5	19	10	10	19
$\tilde{C}_{7;3m}^{Deces}$	6	10	0	0	20	32	0	0	22

Tableau H. 13 : Lags utilisés dans la R-EMD1 des hospitalisations de la CMQ.

$C^{Tmoy}_{1;3j}$	1	$C^{Tmin}_{1;3j}$	1	$C^{Tmax}_{1;3j}$	1	$C^{Prec}_{1;3j}$	2	$C^{Snow}_{1;3j}$	1	$C^{Pmoy}_{1;4j}$	1	$C^{Hmin}_{1;3j}$	2	$C^{Hmoy}_{1;3j}$	1	$C^{Hmax}_{1;3j}$	1
$C^{Tmoy}_{2;6j}$	0	$C^{Tmin}_{2;6j}$	0	$C^{Tmax}_{2;6j}$	0	$C^{Prec}_{2;6j}$	0	$C^{Snow}_{2;6j}$	1	$C^{Pmoy}_{2;6j}$	0	$C^{Hmin}_{2;6j}$	1	$C^{Hmoy}_{2;6j}$	1	$C^{Hmax}_{2;6j}$	1
$C^{Tmoy}_{3;12j}$	0	$C^{Tmin}_{3;12j}$	0	$C^{Tmax}_{3;12j}$	0	$C^{Prec}_{3;12j}$	7	$C^{Snow}_{3;12j}$	3	$C^{Pmoy}_{3;15j}$	0	$C^{Hmin}_{3;12j}$	1	$C^{Hmoy}_{3;12j}$	1	$C^{Hmax}_{3;12j}$	1
$C^{Tmoy}_{4;22j}$	0	$C^{Tmin}_{4;22j}$	0	$C^{Tmax}_{4;22j}$	0	$C^{Prec}_{4;22j}$	5	$C^{Snow}_{4;24j}$	14	$C^{Pmoy}_{4;28j}$	7	$C^{Hmin}_{4;23j}$	3	$C^{Hmoy}_{4;22j}$	2	$C^{Hmax}_{4;22j}$	2
$C^{Tmoy}_{5;1m}$	11	$C^{Tmin}_{5;1m}$	9	$C^{Tmax}_{5;1m}$	11	$C^{Prec}_{5;1m}$	17	$C^{Snow}_{5;2m}$	4	$C^{Pmoy}_{5;2m}$	7	$C^{Hmin}_{5;1m}$	0	$C^{Hmoy}_{5;1m}$	0	$C^{Hmax}_{5;1m}$	0
$C^{Tmoy}_{6;3m}$	10	$C^{Tmin}_{6;3m}$	7	$C^{Tmax}_{6;3m}$	19	$C^{Prec}_{6;2m}$	6	$C^{Snow}_{6;5m}$	18	$C^{Pmoy}_{6;3m}$	1	$C^{Hmin}_{6;3m}$	7	$C^{Hmoy}_{6;2m}$	2	$C^{Hmax}_{6;3m}$	3
$C^{Tmoy}_{7;1a}$	5	$C^{Tmin}_{7;1a}$	5	$C^{Tmax}_{7;1a}$	5	$C^{Prec}_{7;5m}$	91	$C^{Snow}_{7;6m}$	53	$C^{Pmoy}_{7;6m}$	14	$C^{Hmin}_{7;5m}$	104	$C^{Hmoy}_{7;4m}$	16	$C^{Hmax}_{7;5m}$	0
$C^{Tmoy}_{8;2a}$	0	$C^{Tmin}_{8;2a}$	30	$C^{Tmax}_{8;2a}$	63	$C^{Prec}_{8;1a}$	0	$C^{Snow}_{8;1a}$	0	$C^{Pmoy}_{8;1a}$	45	$C^{Hmin}_{8;11m}$	52	$C^{Hmoy}_{8;7m}$	12	$C^{Hmax}_{8;1a}$	0
$C^{Tmoy}_{9;5a}$	161	$C^{Tmin}_{9;4a}$	0	$C^{Tmax}_{9;4a}$	350	$C^{Prec}_{9;2a}$	30	$C^{Snow}_{9;2a}$	352	$C^{Pmoy}_{9;2a}$	1	$C^{Hmin}_{9;2a}$	318	$C^{Hmoy}_{9;11m}$	96	$C^{Hmax}_{9;2a}$	485
$C^{Tmoy}_{10;10a}$	1633	$C^{Tmin}_{10;7a}$	0	$C^{Tmax}_{10;7a}$	0	$C^{Prec}_{10;4a}$	0	$C^{Snow}_{10;3a}$	169	$C^{Pmoy}_{10;4a}$	0	$C^{Hmin}_{10;4a}$	5	$C^{Hmoy}_{10;2a}$	0	$C^{Hmax}_{10;3a}$	346
r^{Tmoy}	0	$C^{Tmin}_{11;13a}$	0	$C^{Tmax}_{11;14a}$	1226	$C^{Prec}_{11;6a}$	0	$C^{Snow}_{11;5a}$	0	$C^{Pmoy}_{11;10a}$	0	$C^{Hmin}_{11;6a}$	704	$C^{Hmoy}_{11;4a}$	0	$C^{Hmax}_{11;4a}$	0
		r^{Tmin}	0	r^{Tmax}		$C^{Prec}_{12;12a}$	30	$C^{Snow}_{12;9a}$	32	$C^{Pmoy}_{12;16a}$	0	$C^{Hmin}_{12;13a}$	63	$C^{Hmoy}_{12;11a}$	0	$C^{Hmax}_{12;7a}$	24
						r^{Prec}	1158	r^{Snow}	0	r^{Pmoy}	0	r^{Hmin}	0	$C^{Hmoy}_{13;20a}$	0	$C^{Hmax}_{13;9a}$	1243
													r^{Hmoy}	0	$C^{Hmax}_{14;15a}$	0	
															r^{Hmax}	0	

Tableau H. 14 : Lags utilisés dans la R-EMD2 des hospitalisations de la CMQ (modèles non saisonniers).

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	1	1	0	1	0	0	0	1	1
$\tilde{C}_{2;5j}^{Hosp}$	0	2	0	1	0	1	2	0	0
$\tilde{C}_{3;7j}^{Hosp}$	0	0	0	2	2	2	0	0	0
$\tilde{C}_{4;14j}^{Hosp}$	0	1	0	5	0	4	3	1	0
$\tilde{C}_{5;26j}^{Hosp}$	0	0	0	6	4	6	12	0	4
$\tilde{C}_{6;2m}^{Hosp}$	11	8	13	0	2	8	0	0	0
$\tilde{C}_{7;3m}^{Hosp}$	11	6	29	32	31	10	3	6	8
$\tilde{C}_{8;6m}^{Hosp}$	0	0	0	2	58	24	22	13	0
$\tilde{C}_{9;1a}^{Hosp}$	3	1	5	0	0	29	51	94	0
$\tilde{C}_{10;3a}^{Hosp}$	59	75	42	102	0	0	232	227	223
$\tilde{C}_{11;6a}^{Hosp}$	0	0	0	38	0	277	0	111	74
$\tilde{C}_{12;8a}^{Hosp}$	0	0	0	0	0	627	87	91	0
\tilde{r}^{Hosp}	0	0	0	796	0	0	0	0	0

Tableau H. 15 : Lags utilisés dans les modèles estivaux de la R-EMD2 des hospitalisations de la CMQ.

	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	0	1	1	0	1	0	0	0	0
$\tilde{C}_{2;5j}^{Hosp}$	1	1	0	0	4	3	0	3	3
$\tilde{C}_{3;7j}^{Hosp}$	2	4	2	0	6	0	1	4	2
$\tilde{C}_{4;14j}^{Hosp}$	2	1	3	0	2	0	0	0	1
$\tilde{C}_{5;26j}^{Hosp}$	1	1	1	9	5	5	0	1	2
$\tilde{C}_{6;2m}^{Hosp}$	4	0	7	17	7	35	13	14	18
$\tilde{C}_{7;3m}^{Hosp}$	17	8	24	26	15	0	8	10	9

Tableau H. 16 : Lags utilisés dans les modèles hivernaux de la R-EMD2 des hospitalisations de la CMQ.

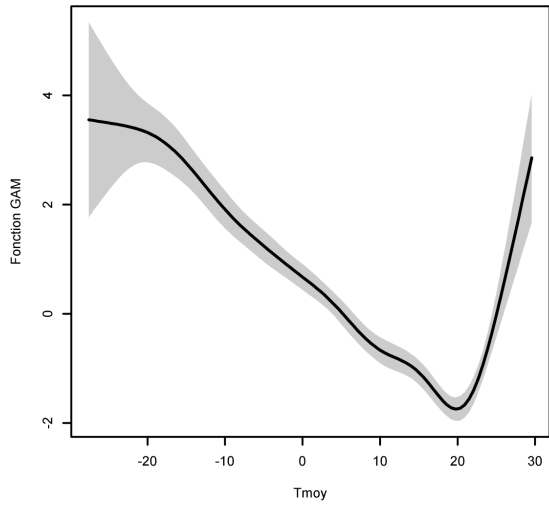
	<i>Tmoy</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Pmoy</i>	<i>Hmin</i>	<i>Hmoy</i>	<i>Hmax</i>
$\tilde{C}_{1;3j}^{Hosp}$	0	0	0	1	0	0	0	1	1
$\tilde{C}_{2;5j}^{Hosp}$	1	2	0	1	0	0	0	1	1
$\tilde{C}_{3;7j}^{Hosp}$	0	3	0	0	0	0	1	4	1
$\tilde{C}_{4;14j}^{Hosp}$	0	0	0	1	3	11	2	1	1
$\tilde{C}_{5;26j}^{Hosp}$	0	0	0	12	7	9	0	4	0
$\tilde{C}_{6;2m}^{Hosp}$	5	4	8	16	6	0	2	4	3
$\tilde{C}_{7;3m}^{Hosp}$	13	12	16	27	32	22	37	26	21

I. Résultats des modèles classiques

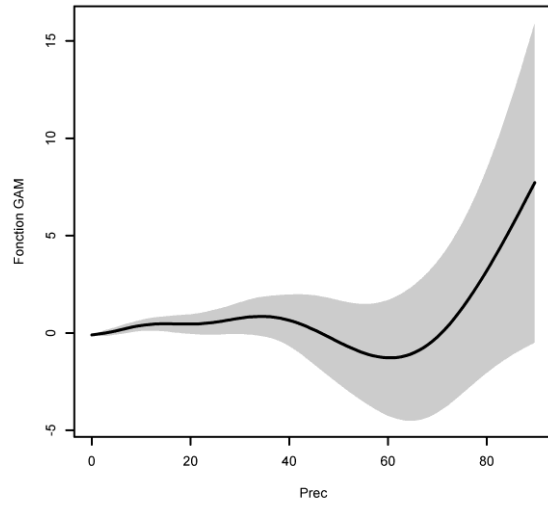
Cette annexe présente les résultats des modèles classiques sur les données de cette étude, à des fins de comparaison avec la régression EMD. Le terme « classique » réfère aux modèles les plus utilisés dans la littérature des études sur la santé et le climat. Ces modèles sont principalement les GAM (Hastie et Tibshirani 1986) et les DLNM (Gasparrini et al. 2010). Ils sont appliqués sur les variables sanitaires utilisées ici, en utilisant les mêmes variables météorologiques que pour la R-EMD, en retirant cependant les températures et humidité minimales et maximales, à cause de trop grandes colinéarités avec les températures et humidités moyennes. Une discussion de ces résultats par rapport aux résultats de la R-EMD se trouve dans la section 5.2.

I.1. GAM

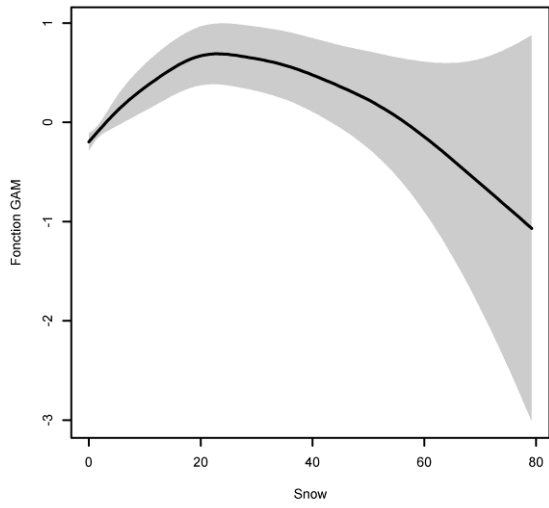
Les fonctions obtenues par GAM sont ici présentées suivies de l'ajustement que ce modèle a permis d'obtenir. Dans cette modélisation, un lag est pris en compte pour chaque variable, selon l'heuristique décrite dans la section 3.2.3.2.



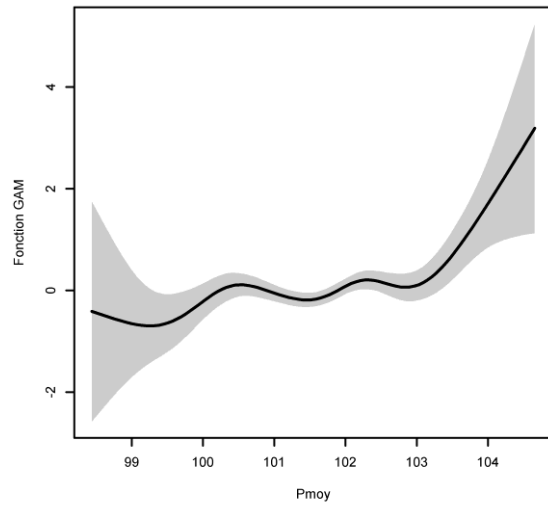
a) Tmoy



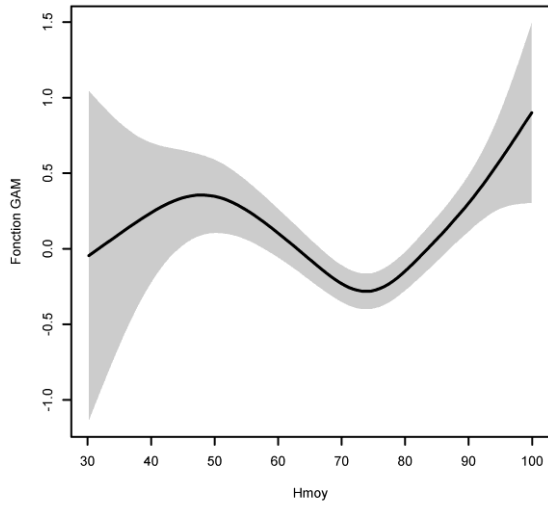
b) Prec



c) Snow

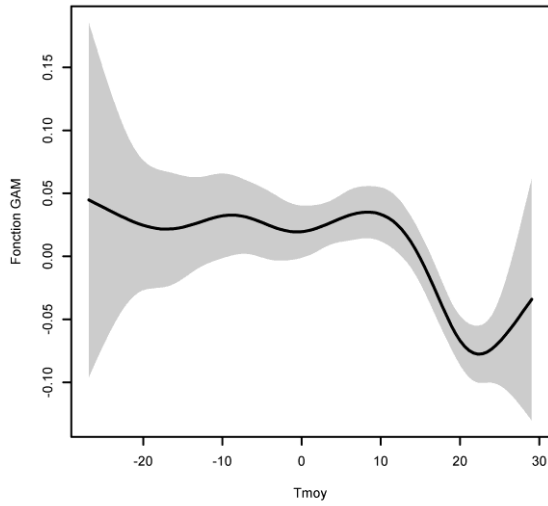


d) Pmoy

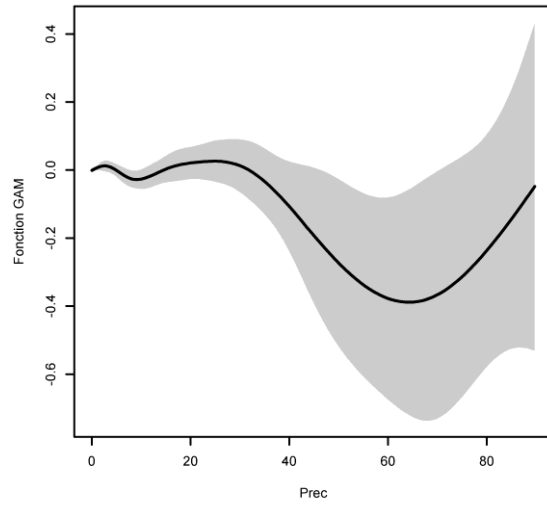


e) Hmoy

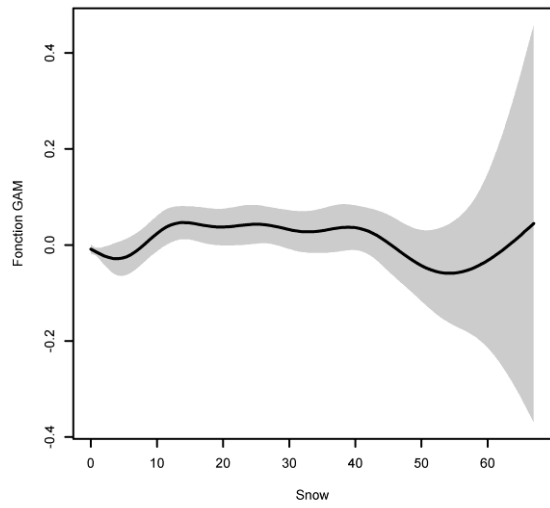
Figure I. 1 : Fonctions GAM obtenues dans l'explication des décès de la CMM.



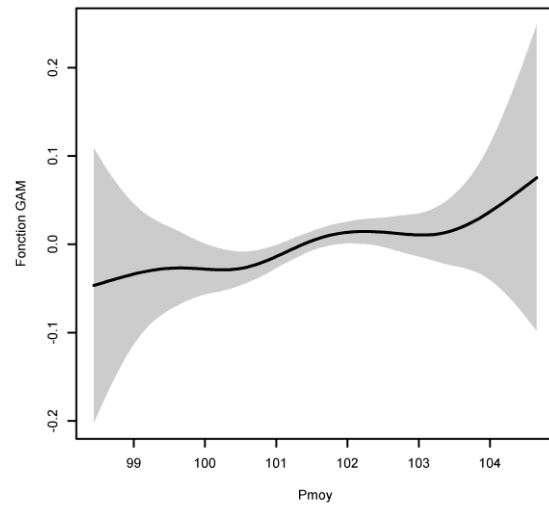
a) Tmoy



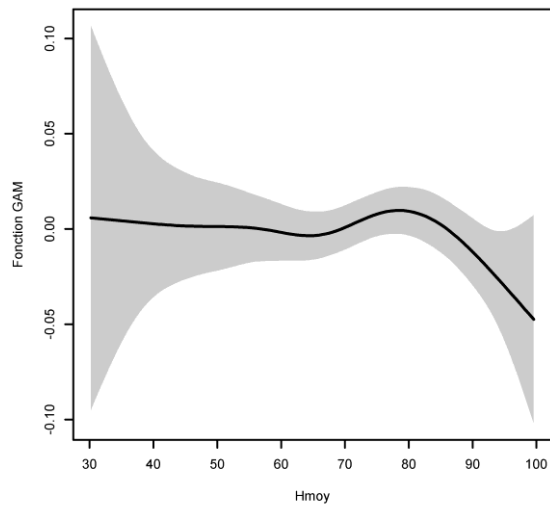
b) Prec



c) Snow

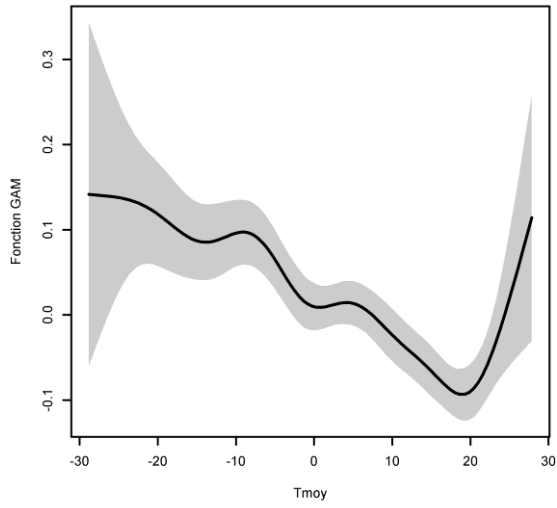


d) Pmoy

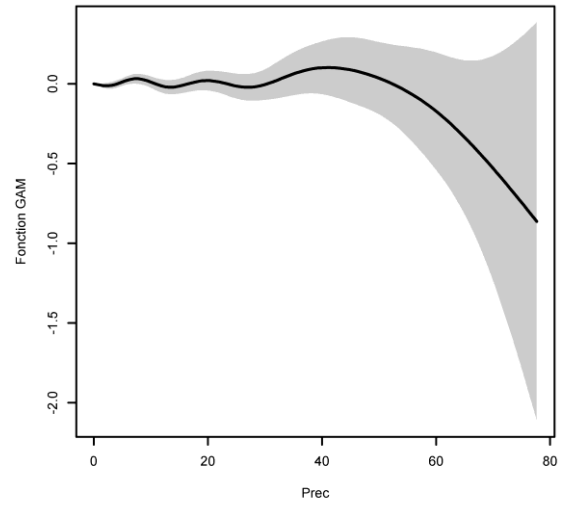


e) Hmoy

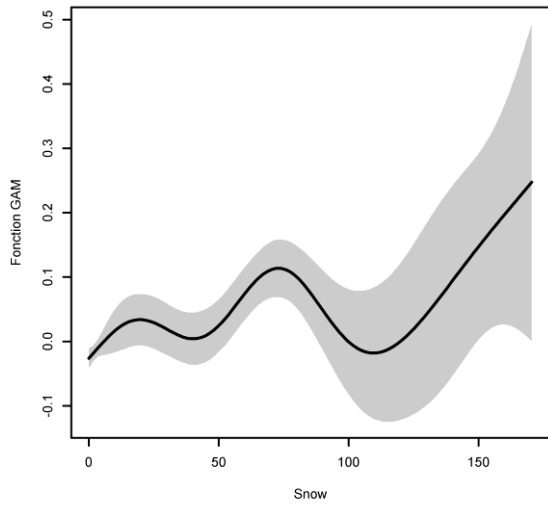
Figure I. 2 : Fonctions GAM obtenues dans l'explication des hospitalisations de la CMM.



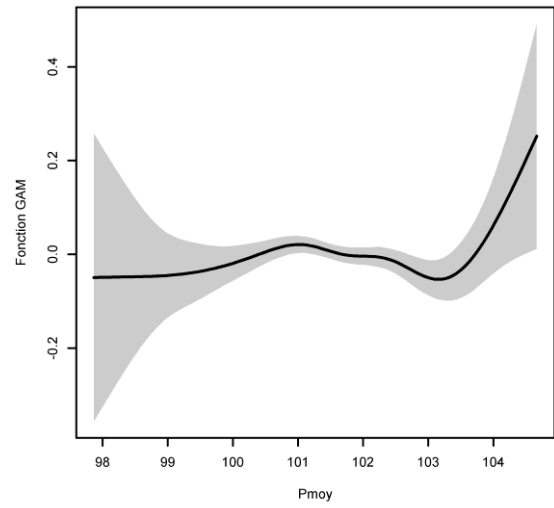
a) Tmoy



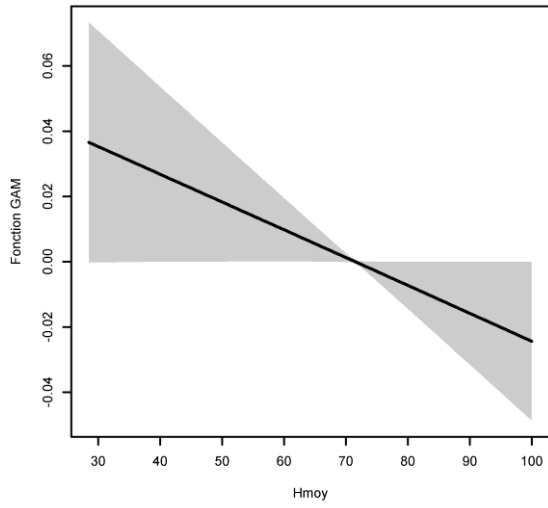
b) Prec



c) Snow

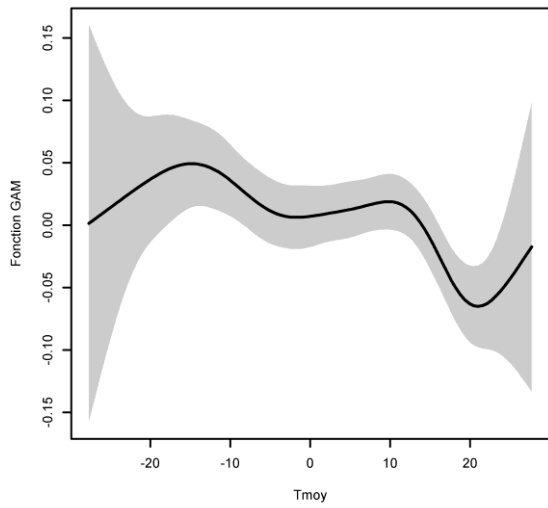


d) Pmoy

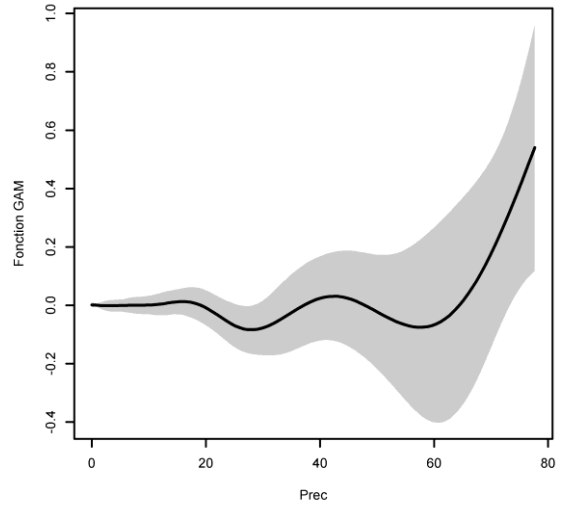


e) Hmoy

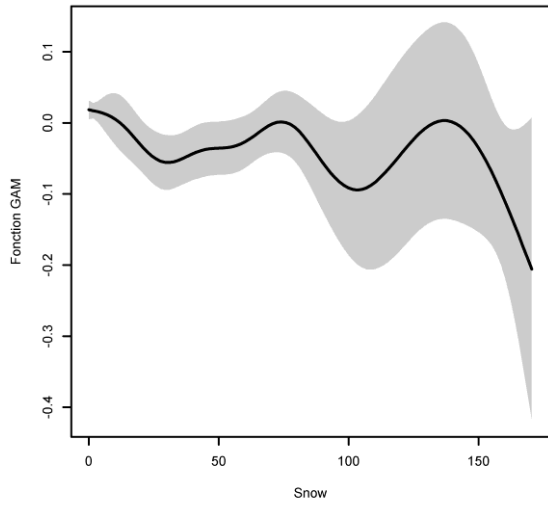
Figure I. 3 : Fonctions GAM obtenues dans l'explication des décès de la CMQ.



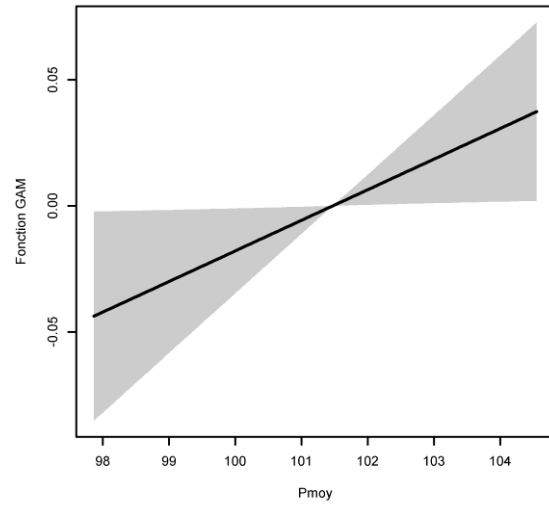
a) Tmoy



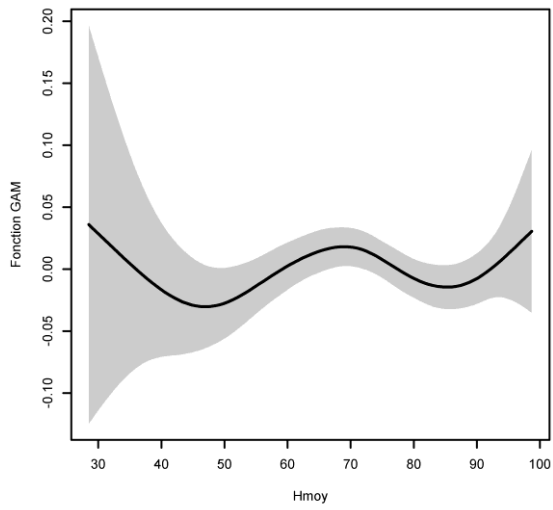
b) Prec



c) Snow

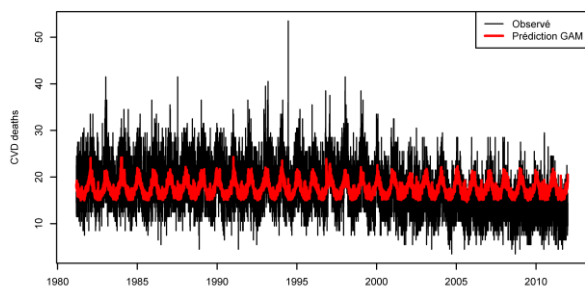


d) Pmoy

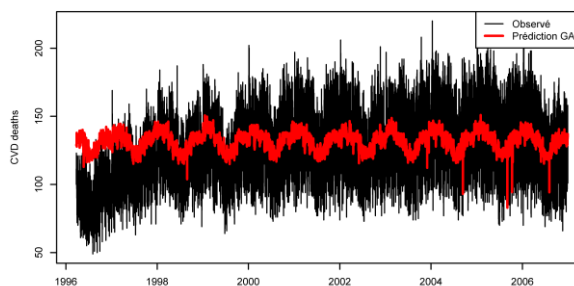


e) Hmoy

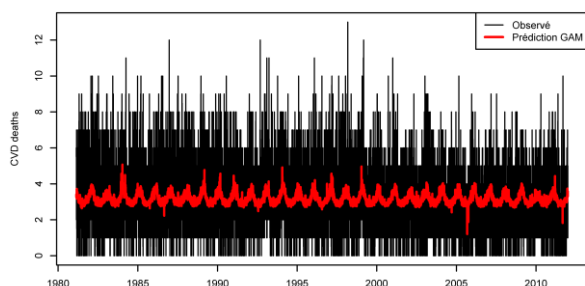
Figure I. 4 : Fonctions GAM obtenues dans l'explication des hospitalisations de la CMQ.



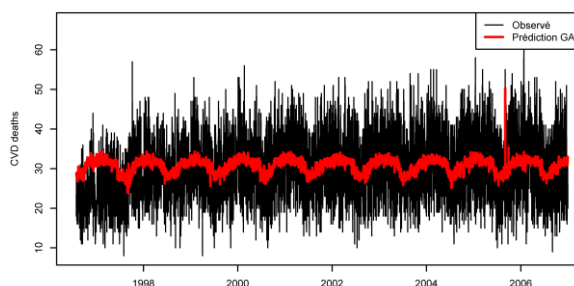
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ



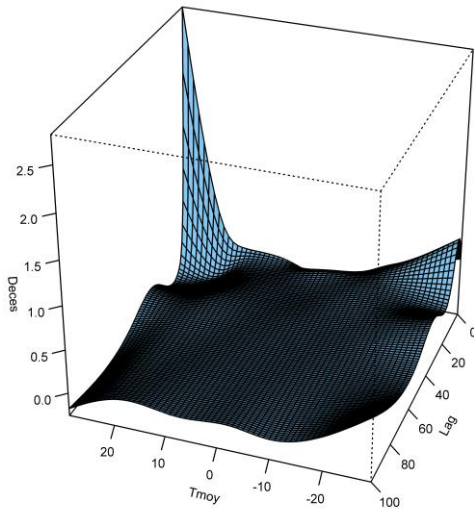
d) Hospitalisations de la CMQ

Figure I. 5 : Séries sanitaires prédites avec les GAM ajustés.

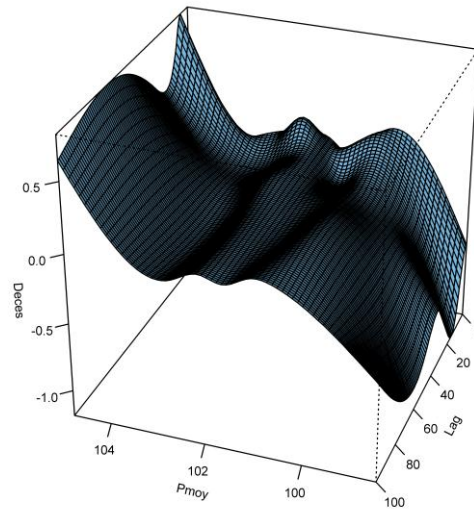
Sur les résultats des GAM, se retrouve notamment la relation en forme de « J » entre les décès et la température. Les GAM suggèrent également une relation linéaire entre l'humidité et les décès par MCV. Les prédictions montrent que cette modélisation est incapable d'expliquer les tendances des séries sanitaires. C'est pour ceci qu'est généralement ajouté un terme de temps dans les modèles afin de capter cette tendance. Cependant, les résultats de la R-EMD présentés en section 4 suggèrent que cette tendance peut être (au moins en partie) liée aux changements météorologiques à *long terme*.

I.2. DLNM

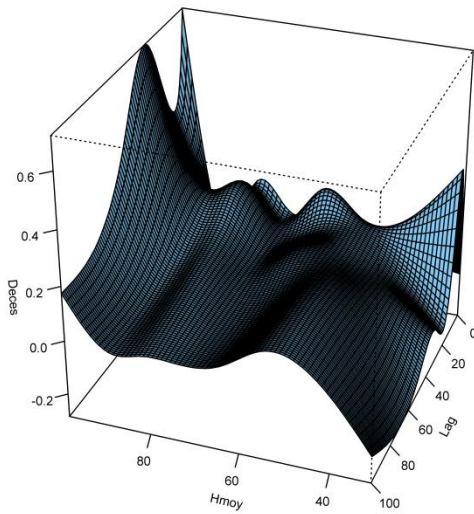
Cette partie présente les résultats des DLNM. Seules les fonctions de l'influence des variables Tmoy, Pmoy et Hmoy sont montrées, car le programme n'a pu ajuster des fonctions pour Prec et Snow à cause des colinéarités. Ceci donne un avantage à la R-EMD sur les DLNM malgré des performances statistiques moindres (cf. section 5.2).



a) *Tmoy*

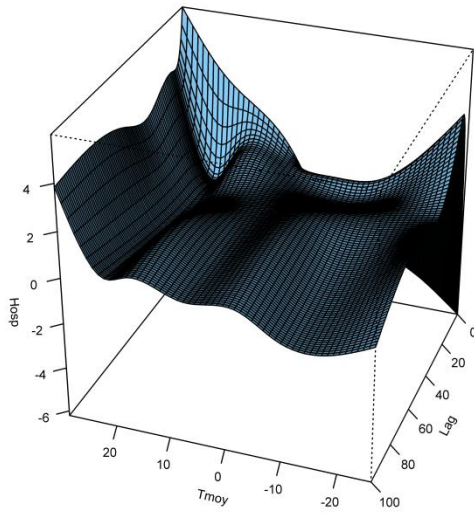


b) *Pmoy*

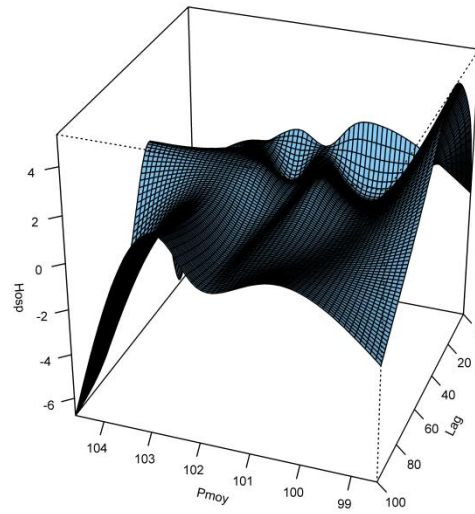


c) *Hmoy*

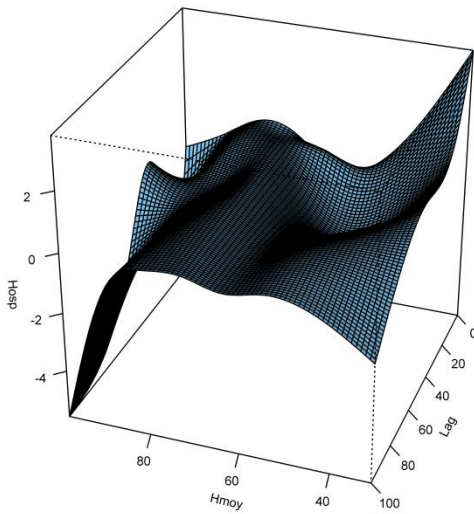
Figure I. 6 : Fonctions DLNM pour la régression des décès de la CMM.



a) *Tmoy*

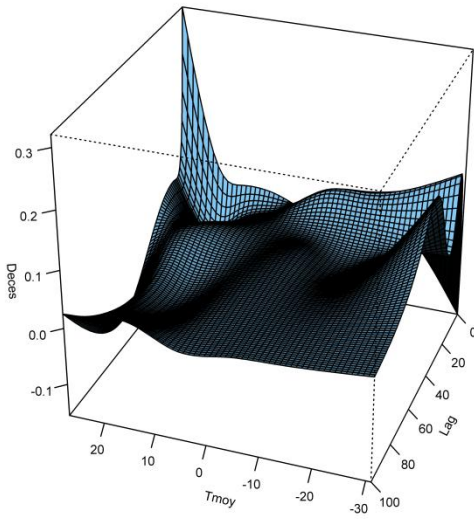


b) *Pmoy*

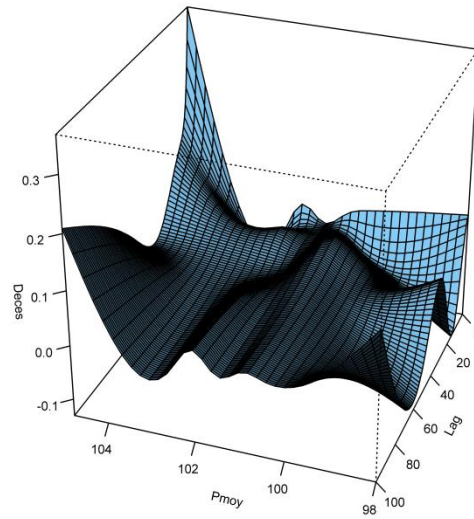


c) *Hmoy*

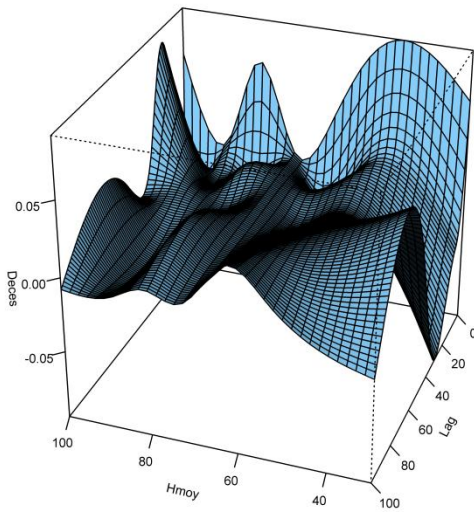
Figure I. 7 : Fonctions DLNM pour la régression des hospitalisations de la CMM.



a) *Tmoy*

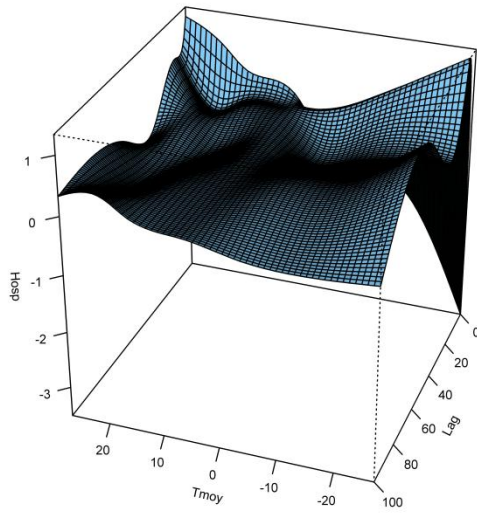


b) *Pmoy*

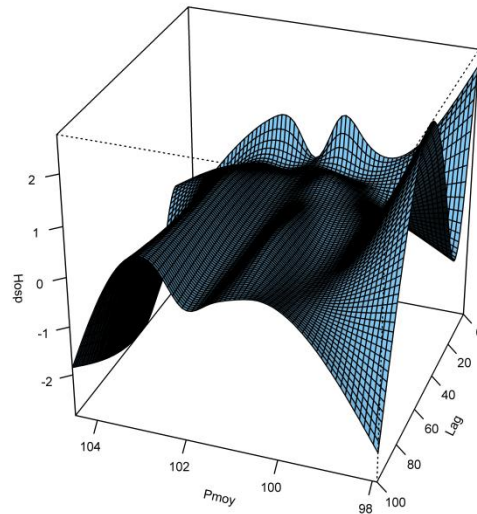


c) *Hmoy*

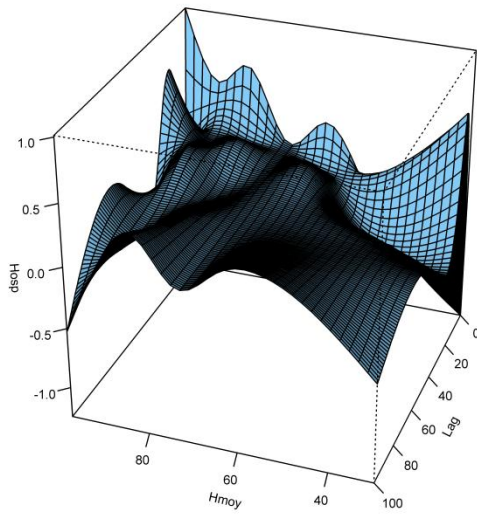
Figure I. 8 : Fonctions DLNM pour la régression des décès de la CMQ.



a) *Tmoy*

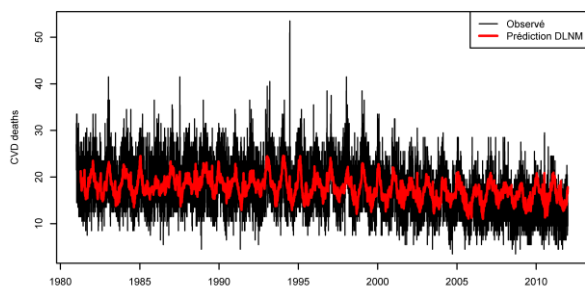


b) *Pmoy*

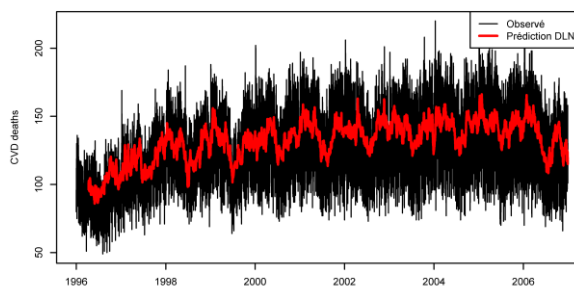


c) *Hmoy*

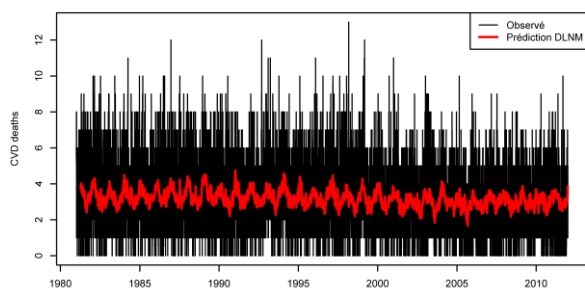
Figure I. 9 : Fonctions DLNM pour la régression des hospitalisations de la CMQ.



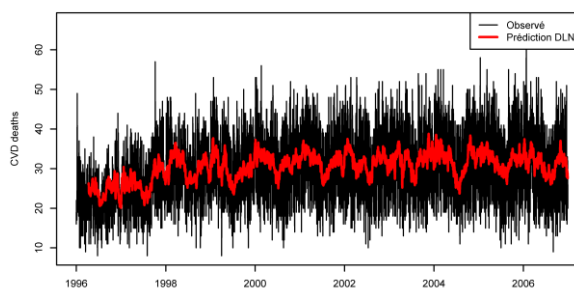
a) Décès de la CMM



b) Hospitalisations de la CMM



c) Décès de la CMQ



d) Hospitalisations de la CMQ

Figure I. 10 : Séries sanitaires prédites avec les GAM ajustés.

La complexité des fonctions bidimensionnelles des DLNM (une dimension du prédicteur et une dimension des lags) permet des ajustements précis. Notamment, le fait de considérer des lags distribués semble améliorer les performances et apporter du pouvoir explicatif. Cet aspect est discuté dans la section 5.2.