

Université du Québec
Institut National de la Recherche Scientifique
Centre Eau Terre Environnement

Approches de modélisation des extrêmes dans l'étude des relations entre la santé et la météo

Par

Yohann Moanahere Chiu

Thèse présentée pour l'obtention
du grade de *Philosophiæ doctor* (Ph.D.)
en sciences de l'eau

Jury d'évaluation

Président du jury et
examineur interne

Taha B.M.J. Ouarda
Institut National de la Recherche Scientifique

Examineur externe

Francis Zwiers
Pacific Climate Impacts Consortium

Examineur externe

Eric Lavigne
Santé Canada

Codirecteur de recherche

Belkacem Abdous
Université Laval

Directeur de recherche

Fateh Chebana
Institut National de la Recherche Scientifique

REMERCIEMENTS

Fiu ! Je n'ai jamais couru de marathon de ma vie, car je n'ai pas une bonne endurance. Pourtant, me voici en fin de doctorat... alors que le doctorat est un vrai marathon, tant physique que mental ! La ligne d'arrivée m'a semblé inaccessible, parfois. J'ai eu du mal à garder confiance tout le long. Au final, je crois qu'il est normal de trébucher... mais l'essentiel est de se relever. Pour finir la course, de la persévérance et du courage sont nécessaires. En ce qui me concerne, j'en dois beaucoup à mon entourage. Je suis donc reconnaissant envers un grand nombre de personnes (j'ai d'ailleurs délibérément réduit la taille de l'interligne pour cette section... le doctorat rend très malin). À commencer par mon directeur, Fateh Chebana. Tu as été (plus que) patient avec moi et m'as bien formé pour un début de carrière en recherche. Merci pour ton temps, ton énergie, tes conseils. Je pense avoir été à bonne école et je garde pour la suite de mon parcours tes nombreuses métaphores sur la nourriture ainsi que ton esprit critique. Indispensables, en recherche ! Je remercie aussi mon co-directeur, Belkacem Abdous. Tu t'es montré disponible malgré les kilomètres et travailler avec toi a été un vrai plaisir. J'espère avoir l'occasion de collaborer avec vous deux par la suite. Par ailleurs, je ne pourrais jamais assez remercier Diane Bélanger (ma chère « maman de la recherche ») et Pierre Gosselin (MD, MPH, BC : blagueur chronique). J'ai été extrêmement chanceux de pouvoir travailler avec vous. Votre bonne humeur, vos suggestions et votre humanité ont grandement aidé à la réalisation de cette thèse. Un merci chaleureux à Taha Ouarda, André St-Hilaire, Sophie Dabo-Niang et Denis Talbot. Sans oublier Jean-Xavier Giroux, maître des données. Finalement, j'aimerais remercier Eric Lavigne et Francis Zwiers pour avoir accepté d'évaluer mon travail. Vos commentaires et suggestions ont permis une réelle bonification de ma thèse.

Que serait une vie de recherche dénuée d'amis, de famille et d'amour ? Impensable ! En tant que statisticien qui se respecte, j'ai dressé un très beau tableau récapitulatif (voir ci-dessous (je me suis abstenu de faire un graphique sous R)).

J'ai une tendre pensée pour ma popoho, partie se reposer. Ton *tchoun* était le meilleur du monde. Kounkoug doit être arrivé aussi, je vous souhaite une bonne continuation à tous les deux.

À la fin de ma thèse, j'ai reçu le cadeau LE PLUS EXTRAORDINAIRE POSSIBLE ! Merci à tim, Mel, Clarisse et Guillaume, Topher, Pons, Ben et Jenna, Chou, Benjex et Math, Nahems,

Riri, Meliki, Anjenny, Martin, Anaïs et Jeff, Ilid, Damien, Lisa, Caro, Daf, Mir et Gio, Enrica, Moux, Peichou, Camille, Tom et LN, Flo, Inadmissible, Sarah et PY.

Variable	Description
$\pi\mathbb{R}^2$	Mon collègue, mon ami, Pierrinounet. Il y a bien longtemps que toi et moi avons mis les pieds à l'INRS... et nous voilà docteurs. MERCI pour toutes les aventures que l'on a vécues ensemble !
Didictatrice et Pizza (aïe !)	À mes copines d'étage : merci de m'avoir laissé trainer dans votre bureau pour placoter. Accessoirement, d'excellentes statisticiennes (!). Dai dai !
Cécilienne et leurs enfants	Namaste Etienne pour les joies du Bikram et de la construction d'arbre à chat et merci à toi Cécile de ne pas être un déchet à tous les jeux (enfin...).
Véro, Kim et Charlotte	Mon amie épidémiologiste, mon amie asiatique parisienne et mon amie italienne qui fait de bons tiramisus. On a fait plein de trucs ensemble (avant que vous ne nous abandonniez toutes !) et c'était bien.
François et Marc	L'homme présent dans tout l'INRS (littéralement) et l'homme à tout siffler dans la vie (littéralement).
Isa et Ana	Isa c'est chouette de jaser biologie en tout temps avec toi. Ana, a mi me gusta dormir y comer los churros. Longue vie à Copito !
Laura, Broch', Shoni et Max	Merci aux BFF pour plein de soirées délirantes ! Promis, le prochain Photoshop sera meilleur. Peut-être. Un PhD en arts graphiques ?
Gerem et Pauline	Vous et êtes repartis trop tôt ! En espérant vraiment vous recroiser.
Will, Nono, Antoine, Vivien	« la gang des géologiques ». Pour les cailloux et pour la découverte d'un monde de cailloux dont je ne soupçonnais même pas l'existence.
Alexandrine et la famille Bourke	Vous avez une place toute particulière dans mon cœur. Merci de m'avoir si bien intégré – Limoilou style ! J'ai vécu des moments fantastiques avec vous, ma famille v2.39.
Mamoune, Babynou, Kekex et papa ours	Papa ours, mes sœurs préférées et la grande prêtresse, tellement loin mais toujours présents. Un merci infini à ma mamoune. Sans ton soutien indéfectible de mère aimante, <i>rien</i> n'aurait été possible. Rien.

Enfin, on dit (du moins, je l'ai entendu une fois) que derrière tout grand homme il y a une grande femme. Je ne suis pas très grand (1m72), mais la femme qui est derrière moi est importante à un point que je ne pourrais jamais décrire. Pour tout le réconfort, les petits plats, les jeux avec les chats (coucou à poupoune et pouche), les encouragements, les voyages (les petits comme les grands), pour m'avoir porté et supporté... merci namoureuse. Je suis heureux et reconnaissant de t'avoir eue à mes côtés tout au long de cette aventure. On l'a franchie, la ligne d'arrivée.

AVANT-PROPOS

Ce document de thèse présente mes travaux de recherche effectués au cours des dernières années.

La première partie de la thèse est constituée d'une synthèse générale qui a pour objectif de

1) présenter une revue de la littérature concernant les problématiques abordées durant la thèse,

2) survoler les principaux résultats et de discuter de leur portée scientifique et 3) de présenter une

conclusion et une discussion générales. La deuxième partie de la thèse contient quatre articles et

rapports de recherches publiés, soumis ou sur le point d'être soumis à des revues internationales

avec comité de lecture. Chaque article ou rapport correspond à un chapitre de la thèse.

RÉSUMÉ

L'impact de la météo sur la santé humaine est reconnu et largement étudié. Toutefois, la grande majorité des travaux qui s'y consacrent font usage de l'ensemble des observations, en se concentrant sur le caractère global de l'impact qui se résume à travers la moyenne. Cela a pour effet de ne pas prendre en compte spécifiquement les phénomènes extrêmes tels que les pics sanitaires. Ces derniers représentent les extrêmes des variables sanitaires et surviennent à une fréquence rare, avec un impact fort sur le système de la santé publique. Leurs caractéristiques sont donc très différentes des phénomènes ordinaires. Ils nécessitent par conséquent des outils et examens spécialisés, ce qu'une modélisation de l'ensemble des observations ne fournit habituellement pas. De la même façon, les extrêmes météorologiques ne sont pas étudiés exclusivement en relation avec la santé dans un cadre de modélisation. Cette thèse a ainsi pour but principal l'étude des valeurs extrêmes dans les relations entre la santé et les conditions météorologiques à l'aide d'outils statistiques appropriés. En premier lieu, la théorie des valeurs extrêmes couplée à des modèles de régression non-paramétriques permettent d'étudier en détail les *pics sanitaires* ainsi que leurs relations avec la météo (pas nécessairement extrême). Les distributions extrêmes des pics sanitaires ainsi que les niveaux de retour associés sont en particulier estimés. Ensuite, la régression quantile offre une vision alternative de ces relations, grâce à un portrait complet évalué dans l'ensemble des quantiles de la distribution de la variable sanitaire. Elle met en lumière les impacts hétérogènes que peuvent avoir les variables météorologiques (pas nécessairement extrême) selon le quantile étudié. Finalement, la problématique inverse est abordée, celle des extrêmes météorologiques. Ils peuvent survenir dans un contexte univarié (p. ex. température) ou multivarié (p. ex. température et humidité). Leurs effets sur la santé sont enfin analysés. Les résultats obtenus dans cette thèse montrent que les

ABSTRACT

The influence of meteorological conditions on human health is largely documented. Most of the studies in this context use all events, summarizing them via the mean. This is necessary in order to understand global trends and everyday situations. However, the mean measures are not appropriate to comprehend specific events such as extreme events. For example, extremes of a health variable (sanitary peaks) very rarely occur but have an important impact on the public health system. Their characteristics are thus different from those of the mean events. This implies that sanitary peaks require specialized tools and insights, which are not provided by a modeling of all the events. Likewise, meteorological extremes are not exclusively examined in relation to health in terms of modeling. Therefore, this thesis aims at studying extreme events in health and meteorological conditions relationships using suitable statistical tools. The case studies are cardiovascular deaths and hospitalizations in Montréal and Québec (Canada). First, the extreme value theory paired with non-parametric regression techniques allow to study in detail *sanitary peaks* and their relationships with meteorological conditions (which are not necessarily extreme). The sanitary peaks extreme distributions and the associated return levels are fitted. Then, quantile regression presents an alternative point view of these relationships through a different but more complete description. Indeed, quantile regression can detect heterogeneities in meteorological effects on the sanitary variable, depending on the studied quantile. Finally, the inverse problem is looked at, i.e. the meteorological extremes. They can occur in a univariate (e.g. temperature) or multivariate (e.g. temperature and humidity) context, therefore a correct identification is needed before analyzing their impacts on health. The obtained results of this thesis show that models using all the observations and those using only extreme events (sanitary or meteorological) differ

in many ways. Hence, specific studies on extremes and their associated statistical techniques are required, especially in regression in a public health context.

Keywords : Sanitary peaks; Meteorological extremes; Extreme value theory; Quantile regression; Generalized additive models; Multivariate extremes; Cardiovascular health.

CONTRIBUTIONS

Articles

- [A1] **Chiu Y.**, Chebana F., Abdous B., Bélanger, D. et Gosselin, P. (2016). Mortality and morbidity peaks modeling: An extreme value theory approach. *Stat Methods Med Res* 10.1177/0962280216662494:0962280216662494.
- [A2] **Chiu Y.**, Chebana, F., Abdous, B., Bélanger, D. et Gosselin, P. (2017a). Study of health peaks and meteorological conditions. À soumettre.
- [A3] **Chiu Y.**, Chebana, F., Abdous, B., Bélanger, D. et Gosselin, P. (2017b). Use of quantile regression to study health peaks. À soumettre.
- [A4] **Chiu Y.**, Chebana, F., Abdous, B., Bélanger, D. et Gosselin, P. (2017c). Impact of multivariate meteorological extremes on health. À soumettre.

Rapports de recherche

- [R1] **Chiu Y.**, Chebana F., Abdous B., Bélanger D. et Gosselin P. (2015). Modélisation des pics de mortalité et de morbidité hospitalière pour cause de maladies cardiovasculaires à Québec et Montréal (Québec) : Une approche par la théorie des valeurs extrêmes. Institut national de la recherche scientifique, R1593, p 129.
- [R2] **Chiu Y.**, Chebana, F., Abdous, B., Bélanger, D. et Gosselin, P. (2016). Modélisation des pics sanitaires de maladies cardiovasculaires en fonction de la météo dans les communautés métropolitaines de Québec et Montréal. Institut national de la recherche scientifique, R1598, p 128.
- [R3] **Chiu Y.**, Chebana, F., Abdous, B., Bélanger, D. et Gosselin, P. (2017a). Étude des relations entre les pics de maladies cardiovasculaires et les conditions météorologiques par la régression quantile, dans les communautés métropolitaines de Québec et de Montréal (Québec). Institut national de la recherche scientifique, R1723, p 77.
- [R4] **Chiu Y.**, Chebana, F., Abdous, B., Bélanger, D. et Gosselin, P. (2017b). Sélection des extrêmes météorologiques et étude de leurs relations avec la santé. Institut national de la recherche scientifique, p 105.

Le premier chapitre de cette thèse est constitué d'un article [A1] tandis que les trois chapitres suivants sont constitués de trois rapports scientifiques [R2-R4]. Ces rapports, rédigés à des fins de transfert des connaissances, détaillent les méthodologies complètes et contiennent les applications exhaustives de ces dernières. Les articles [A1-A4] en reprennent les fondements mais ne présentent que quelques résultats sélectionnés.

L'article [A1] étudie les pics sanitaires exclusivement grâce à la théorie des valeurs extrêmes. En particulier, les distributions extrêmes des pics sanitaires ainsi que leurs niveaux de retour sont obtenus. Les résultats exhaustifs sont décrits dans le rapport [R1] disponible en ligne.

Le rapport [R2] propose d'expliquer les pics sanitaires obtenus dans [A1] à l'aide des variables météorologiques. Il permet ainsi d'analyser l'effet de ces dernières sur la survenue des pics sanitaires.

Le rapport [R3] propose l'utilisation de la régression quantile pour étudier les relations entre les pics sanitaires et les variables météorologiques. Il propose une vision complémentaire de [R2], tout en s'inscrivant dans la même optique.

Le dernier rapport [R4] s'intéresse à la problématique inverse de [R1-R3], à savoir les extrêmes météorologiques (au lieu des pics sanitaires). Diverses définitions statistiques de ces derniers sont considérées puis leur impact sur la santé est évalué.

L'ensemble des travaux a été mené par Y. Chiu, sous la supervision de F. Chebana et la co-supervision de B. Abdous. De plus, D. Bélanger et P. Gosselin ont fourni une aide inestimable en participant régulièrement à l'élaboration des méthodologies, l'interprétation des résultats et la rédaction des articles.

TABLES DES MATIERES

REMERCIEMENTS	ii
AVANT-PROPOS	iv
RÉSUMÉ.....	v
ABSTRACT	vii
CONTRIBUTIONS.....	ix
Articles	ix
Rapports de recherche	ix
TABLES DES MATIERES	xi
LISTE DES FIGURES	xviii
LISTE DES TABLEAUX.....	xxvi
LISTE DES ACRONYMES ET ABRÉVIATIONS.....	xxx
SYNTHÈSE	1
1. INTRODUCTION.....	2
1.1. Contexte	2
1.2. Problématique	4
1.3. Objectifs et réalisations.....	6
1.4. Organisation du document	9
2. REVUE DE LA LITTÉRATURE.....	10
2.1. État de l'art sur les MCV et la météo.....	10

2.2.	État de l'art sur les méthodes statistiques en santé publique	15
2.3.	Étude des valeurs extrêmes.....	20
3.	DONNÉES	26
3.1.	Données sanitaires	27
3.2.	Données météorologiques.....	30
4.	MÉTHODES PROPOSÉES.....	31
4.1.	Études des pics sanitaires par la théorie des valeurs extrêmes	31
4.2.	Étude des relations entre la météo et les pics sanitaires.....	34
4.3.	Étude des pics sanitaires et de la météo par la régression quantile.....	37
4.4.	Extrêmes météorologiques et influence sur la santé.....	39
5.	PRINCIPAUX RÉSULTATS	41
5.1.	Études des pics sanitaires par la théorie des valeurs extrêmes	42
5.2.	Étude des relations entre la météo et les pics sanitaires.....	44
5.3.	Étude des pics sanitaires et de la météo par la régression quantile.....	46
5.4.	Extrêmes météorologiques et impact sur la santé.....	48
6.	CONCLUSIONS ET PERSPECTIVES.....	52
6.1.	Conclusions.....	52
6.2.	Perspectives de recherche	57
	ARTICLES ET RAPPORTS DE RECHERCHE.....	60

I. MORTALITY AND MORBIDITY PEAKS MODELING: AN EXTREME VALUE	
THEORY APPROACH	61
Résumé	62
Abstract	63
Keywords	64
1 Background	65
2 Data and methods	67
2.1 Data	67
2.2 Methods	71
3 Application	78
3.1 Hospitalisation peaks in Montreal	79
3.2 Death peaks in Quebec	83
4 Discussion	88
5 Conclusion	91
Acknowledgments	92
II. MODÉLISATION DES PICS SANITAIRES POUR MALADIES CARDIOVASCULAIRES EN FONCTION DE LA MÉTÉO DANS LES COMMUNAUTÉS MÉTROPOLITAINES DE QUÉBEC ET MONTRÉAL	93
Résumé	94
1. Introduction	95

2.	Analyse statistique.....	99
3	Résultats	105
3.1	Analyse descriptive	106
3.2	Modélisations des pics d’hospitalisations en fonction de la météo à Montréal	109
3.3	Modélisations des pics d’hospitalisations en fonction de la météo à Québec	113
3.4	Modélisation des pics de décès en fonction de la météo à Montréal	118
3.5	Modélisation des pics de décès en fonction de la météo à Québec	123
3.6	Résumé des principaux résultats.....	125
4	Discussion	126
4.1	Température	126
4.2	Humidité	127
4.3	Pression atmosphérique	128
4.4	Précipitations.....	129
4.5	Quantité de neige au sol.....	129
4.6	Considérations statistiques.....	130
5	Conclusion.....	133
6	Limitations	134
	Figures et Tableaux	136
	Figures.....	136
	Tableaux	170

Annexe 1 : Figure illustrative.....	183
Annexe 2 : Modèle additif généralisé	184
III. ÉTUDE DES RELATIONS ENTRE LES PICS DE MALADIES	
CARDIOVASCULAIRES ET LES CONDITIONS MÉTÉOROLOGIQUES PAR LA	
RÉGRESSION QUANTILE	187
Résumé	188
1. Introduction	189
2. Analyse statistique.....	194
3. Résultats	199
3.1. Hospitalisations à Montréal	199
3.2. Hospitalisations à Québec	201
3.3. Décès à Montréal	203
3.4. Décès à Québec	204
3.5. Résumé des résultats	206
4. Discussion	209
4.1. Température	210
4.2. Pression atmosphérique	212
4.3. Humidité relative	213
4.4. Quantité de neige au sol	214
4.5. Précipitations.....	215

4.6. Considérations statistiques.....	215
5. Conclusion.....	219
6. Limitations	220
Figures.....	222
Annexes.....	233
Régression quantile : estimation	234
Résultats avec les variables météorologiques minimales.....	235
IV. IDENTIFICATION DES EXTRÊMES MÉTÉOROLOGIQUES ET ÉTUDE DE LEURS RELATIONS AVEC LA SANTÉ.....	239
Résumé.....	240
1. Introduction	241
2. Analyse statistique.....	246
2.1. Extrêmes univariés.....	246
2.2. Extrêmes multivariés	249
2.3. Régression.....	252
3. Résultats	254
3.1. Extrêmes univariés.....	254
3.2. Extrêmes multivariés	257
4. Discussion	261
5. Conclusion.....	267

6. Limitations	268
Tableaux et figures	270
Tableaux	270
Figures	272
Annexes	293
A- Fonctions de profondeur	293
B- Indices météorologiques	295
RÉFÉRENCES	297

LISTE DES FIGURES

Figure 0-1 : Représentation des différents liens entre les chapitres de la thèse. Les problématiques visées par la thèse sont encadrées en rouge. L'étoile indique le caractère extrême de la variable. .9	
Figure 0-2: Changements possibles et implications dans la distribution de la température et des précipitations sous les effets du changement climatique (Zwiers <i>et al.</i> , 2013).....	14
Figure 0-3 : Résumé des différentes approches utilisées dans l'étude des relations entre la santé et la météo. Les approches utilisées dans cette thèse sont en rouge.	15
Figure 0-4: Illustration de la convergence d'une loi uniforme. Les moyennes convergent vers une loi normale, les maxima vers une loi de Weibull.	21
Figure 0-5 : a) Extrêmes de température par bloc, b) extrêmes d'humidité par bloc, c) extrêmes de température par seuil et d) extrêmes d'humidité par seuil, à Montréal.	26
Figure 0-6: Communautés métropolitaines de Montréal et de Québec.....	27
Figure 0-7 : Hospitalisations et décès journaliers à Montréal et Québec.....	29
Figure 0-8 : QQ plot (gauche) et estimation des niveaux de retour (droite) pour les pics d'hospitalisations à Montréal. Les observations sont en bleu, la première bissectrice (QQ-plot) et les intervalles de confiance (niveaux de retour) sont en rouge.	43
Figure 0-9 : Fonction de la température au jour même pour les pics de décès (haut) et pour l'ensemble des hospitalisations (bas) à Québec.	45
Figure 0-10 : Résultats sélectionnés de la QR et de la régression par moindres carrés ordinaires sur l'effet de la température sur les hospitalisations à Montréal.	47
Figure 0-11 : Profondeur de la température et de l'humidité (les pointillés rouges représentent le seuil à 1 %), extrêmes bivariés et hospitalisations correspondants (en rouge) à Montréal.	49
Figure I-1 : Montreal and Quebec Metropolitan communities.	68

Figure I-2: Number of hospitalisations in Montreal (top) and deaths in Quebec (bottom). 70

Figure I-3: Block (up) and POT (bottom) methods illustration on hospitalisations, in Montreal. The block method uses a block size of 180 days while the POT method uses the 90 % quantile. Peaks are in red. 75

Figure I-4: Diagram illustrating the different study steps. WW, MK and WX: Wald-Wolfowitz, Mann-Kendall and Wilcoxon tests. KS and AD: Kolmogorov-Smirnoff and Anderson-Darling tests. RMSE: root mean squared errors. 78

Figure I-5: Hospitalisations peaks histograms in Montreal, depending on various thresholds expressed in quantiles..... 81

Figure I-6: Number of hospitalisations (blue) and selected peaks (red) in Montreal. The red dashed line is the selected threshold (95 % quantile)..... 82

Figure I-7: QQ plot for hospitalisations peaks in Montreal (left). Observations are in blue while the first bisector is in red. Return level for hospitalisations à Montreal (right). Confidence intervals are in red. 83

Figure I-8: Deaths peaks histograms in Quebec, depending on various block sizes expressed in days..... 85

Figure I-9: Number of deaths (blue) and selected peaks (red) in Quebec..... 86

Figure I-10: QQ plot for deaths peaks in Quebec (left). Observations are in blue while the first bisector is in red. Return level for deaths in Quebec (right). Confidence intervals are in red. 87

Figure II-1 : Méthodologie appliquée pour la modélisation de la relation santé-météo. 136

Figure II-2 : Histogrammes des pics sanitaires, à Montréal (gauche) et à Québec (droite)..... 137

Figure II-3: Histogrammes des variables météorologiques considérées à Montréal (gauche) et Québec (droite)..... 138

Figure II-4 : Relations entre les variables météorologiques considérées à Montréal..... 139

Figure II-5 : Relations entre les variables météorologiques considérées à Québec.	140
Figure II-6 : Conditions météorologiques associées aux pics d'hospitalisations (HP) à Montréal.	141
Figure II-7 : Fonction lissée de Tmax (lag de 5 jours) pour les pics d'hospitalisations à Montréal, avec et sans la fonction du temps.	142
Figure II-8 : Fonction lissée de Tmax (lag de 5 jours) pour l'ensemble des hospitalisations à Montréal, avec et sans la fonction du temps.	143
Figure II-9 : Conditions météorologiques associées au pic d'hospitalisations le plus élevé (220 hospitalisations) à Montréal : a) les variables brutes et b) les variables moyennées sur 5 jours.	144
Figure II-10 : Conditions météorologiques associées aux pics d'hospitalisations (HP) à Québec (jours de semaine).	145
Figure II-11 : Fonction lissée de Prec (lag de 7 jours) pour les pics d'hospitalisations à Québec sur les jours de semaine (jds), avec et sans la fonction du temps.	146
Figure II-12 : Fonction lissée de Prec (lag de 7 jours) pour l'ensemble des hospitalisations à Québec sur les jours de semaine (jds), avec et sans la fonction du temps.	147
Figure II-13 : Conditions météorologiques associées au pic d'hospitalisations le plus élevé (67 hospitalisations) à Québec en jours de semaine : a) les variables brutes et b) les variables moyennées sur 7 jours.	148
Figure II-14 : Conditions météorologiques associées aux pics d'hospitalisations (HP) à Québec (fins de semaine).	149
Figure II-15 : Fonction lissée de Tmax (lag de 3 jours) pour les pics d'hospitalisations à Québec en fins de semaine (fds), avec et sans la fonction du temps.	150

Figure II-16 : Fonction lissée de Tmax (lag de 3 jours) pour l'ensemble des hospitalisations à Québec en fins de semaine (fds), avec et sans la fonction du temps. Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.	151
Figure II-17 : Conditions météorologiques associées aux pics de décès (DC) à Montréal (jours de semaine).	152
Figure II-18 : Fonction lissée de Prec (jour même) pour les pics de décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (premier modèle).	153
Figure II-19 : Fonction lissée de Prec (jour même) pour l'ensemble des décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (premier modèle).	154
Figure II-20 : Fonction lissée de Hmax (lag de 7 jours) pour les pics de décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (deuxième modèle).	155
Figure II-21 : Fonction lissée de Hmax (lag de 7 jours) pour l'ensemble des décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (deuxième modèle).	156
Figure II-22 : Conditions météorologiques associées aux pics de décès (DC) à Montréal (fins de semaine).	157
Figure II-23 : Fonction lissée de Tmax (lag de 2 jours) pour les pics des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (premier modèle).	158
Figure II-24 : Fonction lissée de Tmax (lag de 2 jours) pour l'ensemble des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (premier modèle).	159
Figure II-25 : Fonction lissée de Tmax (lag de 2 jours) pour les pics des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (deuxième modèle).	160
Figure II-26 : Fonction lissée de Tmax (lag de 2 jours) pour l'ensemble des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (premier modèle).	161

Figure II-27 : Conditions météorologiques associées au pic de décès le plus élevé (53 décès) à Montréal en fins de semaine : a) les variables brutes et b) les variables moyennées sur 2 jours.	162
Figure II-28 : Conditions météorologiques associées aux pics de décès (DC) à Québec.	163
Figure II-29 : Fonction lissée de Tmax (jour même) pour les pics de décès à Québec, avec et sans la fonction du temps (premier modèle).	164
Figure II-30 : Fonction lissée de Tmax (jour même) pour l'ensemble des décès à Québec, avec et sans la fonction du temps (premier modèle).	165
Figure II-31 : Fonction lissée de Tmax (lag de 1 jour) pour les pics de décès à Québec, avec et sans la fonction du temps (deuxième modèle).	166
Figure II-32 : Fonction lissée de Tmax (lag de 1 jour) pour l'ensemble des décès à Québec, avec et sans la fonction du temps (deuxième modèle).	167
Figure II-33 : Fonction lissée de Snow (lag de 4 jours) pour les pics de décès à Québec, avec et sans la fonction du temps.	168
Figure II-34 : Fonction lissée de Snow (lag de 4 jours) pour l'ensemble des décès à Québec, avec et sans la fonction du temps.	169
Figure II-35 : Comparaison des fonctions lissées de Tmax (lag de 2 jours) pour les pics de décès à Montréal en fins de semaine, avec et sans la fonction du temps : a) approche de Kim et Gu (2004), b) approche sans paramètre de correction.	183
Figure III-1 : Illustration de l'effet moyen de la température maximale et de la température minimale sur les hospitalisations (graphiques du haut) et sur les décès (graphiques du bas) à Montréal.	222
Figure III-2 : Exemple de résultats de la régression quantile. a) Différentes droites de régression quantile, b) Valeurs des différents coefficients estimés des droites de régression quantile.....	223

Figure III-3 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les hospitalisations à Montréal.	224
Figure III-4 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les hospitalisations à Québec.....	225
Figure III-5 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les décès à Montréal.	226
Figure III-6 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les décès à Québec.....	227
Figure III-7 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des hospitalisations à Montréal.....	228
Figure III-8 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des hospitalisations à Québec.....	229
Figure III-9 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des décès à Montréal.....	230
Figure III-10 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des décès à Québec.....	231
Figure III-11 : Comparaison des fonctions de régression obtenues par régression quantile (haut) et avec l’approche par modèles additifs généralisés de Chiu <i>et al.</i> (2016b) (bas) sur les hospitalisations à Montréal.....	232
Figure IV-1: Exemples d’extrêmes de température a) par bloc et b) par seuil, à Montréal. Les blocs et le seuil sont représentés en pointillés.....	272
Figure IV-2 : a) Extrêmes de température par bloc, b) extrêmes d’humidité par bloc, c) extrêmes de température par seuil et d) extrêmes d’humidité par seuil, à Montréal (période de 1996 à 2000 inclusivement).	273

Figure IV-3 : Exemple de fonction de profondeur entre la température et l'humidité à Montréal.	274
Figure IV-4 : Figure illustrant l'approche de Chebana et Ouarda (2011b). Les observations sont en gris tandis que les extrêmes sélectionnés sont en bleu. Le paramètre u contrôle l'orientation selon laquelle ces derniers sont choisis.	275
Figure IV-5 : Extrêmes météorologiques univariés de température (seuil à 99 %, en pointillés rouges) pour les hospitalisations à Montréal.	276
Figure IV-6 : a) Hospitalisations et température à Montréal et b) Droite de régression des extrêmes de température par seuil sur les hospitalisations.	277
Figure IV-7 : Extrêmes météorologiques univariés de température (blocs de 60 jours) pour les hospitalisations à Montréal.	278
Figure IV-8 : a) Hospitalisations et température à Montréal et b) Fonction de régression des extrêmes de température par blocs sur les hospitalisations.	279
Figure IV-9 : a) Hospitalisations et variations de l'humidité à Montréal et b) Droite de régression des extrêmes des variations de l'humidité sur les hospitalisations.	280
Figure IV-10 : Profondeur de la température et de l'humidité (seuil à 1 %) à Montréal.	281
Figure IV-11 : Extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %) pour les hospitalisations à Montréal.	282
Figure IV-12 : Association des hospitalisations correspondantes aux extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %). Les hospitalisations vont du jour même à 8 jours après les extrêmes.	283
Figure IV-13 : Plans de régression des extrêmes multivariés extraits par fonction de profondeur sur les hospitalisations à Montréal. a) Tmax et PM et b) Tmax et Hmax.	284

Figure IV-14: Résumé des conditions météorologiques survenues avant le pic du 12 janvier 2004 à Montréal. Les pointillés rouges représentent le jour du pic de 220 hospitalisations.....	285
Figure IV-15 : Sélection des extrêmes météorologiques avec la méthodologie de Chebana et Ouarda (2011b).	286
Figure IV-16 : Extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %) pour les hospitalisations à Montréal. Les extrêmes sont extraits selon la méthodologie de Chebana et Ouarda (2011b).....	287
Figure IV-17 : Association des hospitalisations correspondantes aux extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %). Les hospitalisations vont du jour même à 8 jours après les extrêmes. Les extrêmes sont extraits selon la méthodologie de Chebana et Ouarda (2011b).....	288
Figure IV-18 : Extrêmes météorologiques univariés de l'humidex (seuil à 99 %) pour les hospitalisations à Montréal.....	289
Figure IV-19 : Extrêmes météorologiques univariés de l'indice de chaleur (seuil à 99 %) pour les hospitalisations à Montréal.....	290
Figure IV-20 : a) Hospitalisations et indice de chaleur à Montréal et b) Droite de régression des extrêmes de l'indice de chaleur sur les hospitalisations.....	291
Figure IV-21 : Position des pics étudiés par Chiu <i>et al.</i> (2016b) dans l'espace température-humidité, jusqu'à 7 jours avant l'apparition des pics.....	292

LISTE DES TABLEAUX

Tableau 0-1 : Définitions des MCV considérées dans le Programme de recherche MCV, incluant leurs codes CIM-9 et CIM-10.	28
Tableau 0-2 : Statistiques descriptives des données sanitaires brutes à Montréal et à Québec.	30
Tableau 0-3 : Statistiques descriptives des données météorologiques brutes à Montréal et à Québec.....	31
Tableau 0-4 : Résultats de modélisation des pics d’hospitalisations à Montréal. Le seuil sélectionné est en gras. Pour chaque seuil, la taille d’échantillon (n) et les p-values pour les tests de MK, WW et WX sont données. Les paramètres de forme et d’échelle correspondants ($\hat{\xi}$ et $\hat{\sigma}$) sont renseignés, avec les p-values des tests d’adéquation pour les distributions GPD, EXP, LNO et GAM. Finalement, les critères RMSE sont renseignés.	42
Tableau 0-5 : Résumé des différents modèles retenus pour expliquer les pics d’hospitalisations et de décès à Montréal et à Québec.	46
Tableau 0-6 : Résultats retenus de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques univariés et multivariés (profondeur) extraits par seuil. Lorsqu’une droite de régression est estimée, le coefficient de régression est renseigné, tandis que dans le cas d’une fonction non-linéaire il s’agit du degré de liberté (ddl).	50
Tableau 0-7 : Résumé des principales différences entre les études classiques et les études sur les pics sanitaires et les extrêmes météorologiques.	56
Tableau II-1 : Résumé des modèles sélectionnés pour les séries de pics d’hospitalisations (HP) et de décès (DC) à Montréal et Québec (Chiu <i>et al.</i> , 2015).	170
Tableau II-2 : Corrélations entre les différentes déclinaisons de la température, de l’humidité et de la pression atmosphérique.	170

Tableau II-3 : Statistiques descriptives des données météorologiques brutes utilisées, pour Montréal et Québec.	171
Tableau II-4 : Résumé numérique des conditions météorologiques associées aux pics d'hospitalisations à Montréal.	171
Tableau II-5 : Résultats du modèle retenu pour expliquer la relation entre les pics d'hospitalisations et la météo à Montréal.....	172
Tableau II-6 : Résumé numérique des conditions météorologiques associées aux pics d'hospitalisations à Québec (jours de semaine).	172
Tableau II-7 : Résultats du modèle retenu pour expliquer la relation entre les pics d'hospitalisations et la météo à Québec (jours de semaine).....	173
Tableau II-8 : Résumé numérique des conditions météorologiques associées aux pics d'hospitalisations à Québec (fins de semaine).	174
Tableau II-9 : Résultats du modèle retenu pour expliquer la relation entre les pics d'hospitalisations et la météo à Québec (fins de semaine).....	174
Tableau II-10 : Résumé numérique des conditions météorologiques associées aux pics de décès à Montréal (jours de semaine).....	175
Tableau II-11 : Résultats du premier modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (jours de semaine).	175
Tableau II-12 : Résultats du deuxième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (jours de semaine).	176
Tableau II-13 : Résumé numérique des conditions météorologiques associées aux pics de décès à Montréal (fins de semaine).....	176
Tableau II-14 : Résultats du premier modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (fins de semaine).	177

Tableau II-15 : Résultats du deuxième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (fins de semaine).	178
Tableau II-16 : Résumé numérique des conditions météorologiques associées aux pics de décès à Québec.....	179
Tableau II-17 : Résultats du premier modèle retenu pour expliquer la relation entre les pics de décès et la météo à Québec.	179
Tableau II-18 : Résultats du deuxième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Québec.	180
Tableau II-19 : Résultats du troisième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Québec.	181
Tableau II-20 : Résumé des différents modèles retenus pour expliquer les pics d'hospitalisations de janvier 1996 à mars 2007 et de décès de janvier 1981 à décembre 2011, à Montréal et à Québec.....	182
Tableau IV-1 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques univariés extraits par seuil. Lorsqu'une droite de régression est estimée, le coefficient de régression est renseigné, tandis que dans le cas d'une fonction non-linéaire il s'agit du degré de liberté (ddl).	270
Tableau IV-2 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques univariés (variations).	270
Tableau IV-3 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques multivariés extraits par fonctions de profondeur.	271
Tableau IV-4 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques multivariés (indices météorologiques). Lorsqu'une droite de	

régression est estimée, le coefficient de régression est renseigné, tandis que dans le cas d'une
fonction non-linéaire il s'agit du degré de liberté (ddl).271

LISTE DES ACRONYMES ET ABRÉVIATIONS

AIC : *Critère d'information d'Akaike*

CC : Changements climatiques

CIM : Classification internationale des maladies

CMM : Communauté métropolitaine de Montréal

CMQ : Communauté métropolitaine de Québec

DC : Décès

DEX : Déviance expliquée

edf : *Degré de liberté équivalent*

EME : Evènement météorologique extrême

EVT : *Théorie des valeurs extrêmes*

fds : Fins de semaine

GAM : *Modèle additif généralisé*

GCV : *Validation croisée généralisée*

GEV : *Distribution des valeurs extrêmes généralisée*

GPD : *Distribution de Pareto généralisée*

HP : Hospitalisations

IC : Indice de chaleur

IC : Intervalles de confiance

jds : Jours de semaine

LNO : *Distribution lognormale*

MCV : Maladies cardiovasculaires

PM : *Particules fines*

QR : *Régression quantile*

SYNTHÈSE

1. INTRODUCTION

Cette section a pour but de décrire en premier lieu le contexte général dans lequel la thèse s'inscrit. La problématique et les objectifs y sont ensuite décrits. Enfin, l'organisation du document de synthèse est présentée.

1.1. Contexte

L'influence des conditions météorologiques sur la santé humaine est reconnue. Les études qui s'intéressent au lien entre la santé et la météo sont nombreuses car il s'agit avant tout d'un enjeu de santé publique. Par exemple, le taux de propagation de certaines maladies infectieuses comme la dengue réagit fortement aux températures et précipitations (Patz *et al.*, 2005), tandis que la mortalité et la morbidité hospitalière (nommée morbidité ci-après) pour toutes causes fluctuent avec la température (Kovats et Hajat, 2008) et que les variations de la pression atmosphérique influent sur l'infarctus du myocarde (Danet *et al.*, 1999).

La mortalité et la morbidité sont traditionnellement analysées à l'aide de l'ensemble des données issues des fichiers administratifs. De ce fait, l'information est résumée sous sa forme moyenne dans les analyses statistiques et ainsi l'un des objets d'étude est la compréhension du caractère *moyen* de la variable sanitaire. Cette méthodologie permet ainsi de se prémunir face aux situations *ordinaires* (voir par exemple Soyiri *et al.*, 2013), ce qui est vital pour les décisions à court terme (par exemple, la gestion quotidienne d'un hôpital). C'est une pratique commune, aussi bien en santé publique que dans d'autres domaines. Par exemple, la plupart des analyses de changement de température et de précipitations se concentrent sur les changements des valeurs moyennes (Alexander *et al.*, 2006). Cependant, il existe des épisodes accrus où de fortes valeurs de la variable sanitaire sont observées, appelées pics sanitaires. Ces derniers s'éloignent des

valeurs moyennes et surviennent à une fréquence faible. Autrement dit, ils ne surviennent que rarement mais impactent fortement le système de la santé. Une étude avec des modèles statistiques classiques, dont l'attention est portée sur la moyenne et les événements *ordinaires*, ne permet pas d'appréhender correctement ces événements *inhabituels*. Les pics sont alors mal représentés et les modèles associés pourraient mener à des prédictions erronées. Ces caractéristiques des valeurs extrêmes sont aussi problématiques dans d'autres domaines, tels que l'hydrologie (Katz *et al.*, 2002) ou la finance (Longin, 2005).

Parmi les maladies les plus graves affectées par la météo, on trouve les maladies cardiovasculaires (MCV). Elles représentent presque 1/3 des décès totaux au Québec. À l'échelle du Canada, les MCV constituent d'ailleurs la famille de maladies la plus coûteuse (Daigle, 2007). Les conditions météorologiques ont maintes fois été considérées comme indicateurs de risque du caractère moyen de certains événements par MCV (p. ex. dans Portier *et al.*, 2010). Ainsi, à titre d'exemple, les variations de la pression atmosphérique ont été associées à l'infarctus du myocarde, tout comme la quantité de neige au sol (Houck *et al.*, 2005; Hopstock *et al.*, 2012). De même, l'effet de la température moyenne sur la mortalité par MCV a été documenté (Goldberg *et al.*, 2011).

Par ailleurs, le Canada n'est pas épargné par les changements climatiques en cours. Au Québec en particulier, on peut citer l'augmentation des températures journalières moyennes dans le sud et l'augmentation des précipitations dans le nord de la province (Ouranos, 2015). Ainsi, étant donné la gravité des MCV et les relations qui existent entre ces dernières et la météo, il est important d'analyser ces relations en détail, afin de permettre une adaptation et une mitigation des changements climatiques.

1.2. Problématique

Malgré l'importance socioéconomique des pics sanitaires, il n'existe pas de définition générale pour ces derniers. En réalité, ils sont analysés de façon ponctuelle. Par exemple, plusieurs études tentent d'expliquer les raisons de la canicule européenne de 2003 (Black *et al.*, 2004; Stott *et al.*, 2004). D'autres études s'attardent sur ses conséquences en santé publique, d'un point de vue descriptif. Cet épisode qui a marqué un record sanitaire n'est cependant pas incorporé dans une définition générale de pics. L'accent de ces études est d'analyser les causes et conséquences du phénomène en question afin de donner des recommandations sanitaires. Ce genre d'études est ainsi crucial puisqu'elles permettent d'instaurer plusieurs mesures pour tenter d'améliorer la réactivité du réseau de la santé. Les effets de la canicule de 2006 ont en ce sens été amoindris par rapport à ceux de 2003 (Fouillet *et al.*, 2008). Pourtant, une méthodologie générale dans un contexte de modélisation des pics sanitaires manque encore. De nombreuses études s'intéressent effectivement soit à des événements précis, ce qui ne permet pas une vision globale, soit à l'ensemble des observations en considérant les pics sanitaires comme des cas particuliers, ce qui dilue l'information liée à ces derniers. Aucune étude n'aborde en fait exclusivement les pics sanitaires dans une optique de modélisation. Pour cela, il faudrait pouvoir les sélectionner afin d'analyser leurs spécificités, chose qui n'est pas faite jusqu'à présent. À plus forte raison, les pics sanitaires ne sont pas intégrés dans un contexte de régression (dans des modèles explicatifs). Pourtant, il apparaît nécessaire de comprendre les conditions de survenue de ces derniers afin d'améliorer la surveillance en santé publique (par exemple, en termes d'alerte sanitaire).

Au vu des relations entre la santé par MCV et les conditions météorologiques, il est pertinent de s'intéresser aux *extrêmes météorologiques*, événements caractérisés par des valeurs inhabituellement éloignées de la moyenne et une fréquence d'apparition rare (Ouranos, 2015). Ils

entraînent des dégâts sociaux-économiques importants, comme l'ont fait les inondations du Richelieu en Richelieu en 2011, qui constituent la plus longue catastrophe naturelle de la province de Québec (Environnement et Changement climatique Canada, 2013). Les extrêmes météorologiques ont un impact reconnu sur les MCV (LeTertre *et al.*, 2006; Sharma *et al.*, 2008; Zanobetti *et al.*, 2013). Tout comme les pics sanitaires, ils possèdent des caractéristiques très différentes des phénomènes météorologiques moyens et leur analyse nécessite donc des outils statistiques adaptés.

Pour mener une analyse appropriée des extrêmes météorologiques, il est essentiel de bien les définir en premier lieu. Pour ce faire, il existe plusieurs façons. Les définitions dépendent toutefois des organismes de recherche concernés, de la variable météorologique étudiée ou encore de la zone géographique en question. Par exemple, l'Institut royal météorologique de Belgique définit une vague de chaleur par une température de plus de 25 °C persistante pendant au moins cinq jours de suite, dont au moins trois jours avec 30 °C ou plus (Institut Royal Météorologique, 2017). Au Canada, on parle de façon équivalente de chaleur accablante, définie par une température de 30 °C ou plus et un indice humidex de plus de 40 (Ministère de la Santé et des Services sociaux du Québec, 2016). Si la mesure de la température des deux définitions est similaire (30 °C), l'échelle temporelle considérée est en revanche différente (cinq jours contre un jour, respectivement). En plus du manque de définition officielle des extrêmes météorologiques, il existe un problème d'ordre statistique intrinsèque à ces derniers. En effet, établir un pic sanitaire ne dépend que d'une seule variable (la variable sanitaire en question). Or, un extrême météorologique peut être établi selon une variable (p. ex. intensité des températures), mais il est souvent plus réaliste de le définir selon plusieurs variables (p. ex. intensité, durée et rareté des températures ou encore humidité et température). Il se trouve que passer d'un contexte

univarié à un contexte bivarié (voire multivarié) entraîne plusieurs complications, dont le classement des observations. S'il est aisé d'ordonner des observations qui ne comprennent qu'une composante (univariées), il est plus difficile de le faire pour des observations à plusieurs composantes (multivariées). Ainsi, il y a deux problèmes à la définition des extrêmes météorologiques : un problème d'ordre statistique (extrêmes multivariés) et un problème d'ordre pratique (différences des effets de la météo selon la zone étudiée). Par ailleurs, aucune étude ne se consacre à la régression des extrêmes météorologiques (dans l'optique d'expliquer une variable réponse par une variable explicative) et aucune direction n'est donc indiquée pour ce genre d'étude, que ce soit en statistique ou en santé publique.

1.3. Objectifs et réalisations

L'objectif principal de cette thèse est d'étudier spécifiquement le caractère extrême dans le cadre de l'épidémiologie environnementale. Plus précisément, il s'agit de proposer et d'appliquer des méthodes de modélisation adaptées à l'étude des valeurs extrêmes dans les relations entre la santé et la météo. Ces valeurs extrêmes peuvent survenir d'un point de vue sanitaire ou météorologique. L'objectif principal se décline en plusieurs sous-objectifs :

- 1- Le développement d'une méthodologie générale pour la sélection et la modélisation des pics sanitaires à l'aide d'outils statistiques adaptés ;
- 2- La modélisation des relations entre les pics sanitaires et les conditions météorologiques ;
- 3- L'application de plusieurs définitions des extrêmes météorologiques et la compréhension de leurs conséquences sur les variables sanitaires ;
- 4- Le développement d'une méthodologie statistique générale adaptée aux pics sanitaires et extrêmes météorologiques en santé publique, en comparant avec les études classiques (utilisant toutes les observations).

En termes de réalisation de ces objectifs, les méthodes sont appliquées aux MCV, à Québec et Montréal (Canada). Les travaux effectués en vue de répondre à ces objectifs ont permis l'écriture de quatre rapports de recherche, desquels ont été extraits quatre articles scientifiques.

Dans le premier chapitre, les pics d'hospitalisations et de décès à Montréal et à Québec sont modélisés. Les pics sanitaires sont extraits à l'aide de deux méthodes issues de la théorie des valeurs extrêmes (EVT), soit la méthode du seuil et celle des blocs. L'EVT n'a pas été utilisée en santé publique, malgré sa capacité à analyser les pics. Une méthodologie statistique globale et rigoureuse est appliquée sur les variables sanitaires. Aucune variable météorologique n'est incluse à cette étape car il est important d'examiner en premier lieu les séries des pics ainsi que leurs caractéristiques statistiques. Cette étape est souvent omise par les études des relations entre la météo et les MCV. En effet, ces études s'intéressent directement à la modélisation des relations, la variable sanitaire ne fait habituellement pas l'objet d'une étude à part entière.

Une fois que les pics sanitaires sont extraits et qu'une distribution extrême adéquate est déterminée, les relations entre ces derniers et les conditions météorologiques sont analysées dans le deuxième chapitre. Pour cela, les modèles additifs généralisés sont mis à contribution, implémentés pour s'adapter aux pics sanitaires. Ce sont des modèles non-paramétriques qui permettent une plus grande flexibilité dans l'exploration des relations entre les données. Cette application sur les pics sanitaires est la première en santé environnementale car ils ne sont généralement pas étudiés *exclusivement*, dans les modèles de régression.

Les premier et deuxième chapitres constituent deux parties d'un même objectif, l'étude des pics sanitaires et leurs relations avec la météo (non nécessairement extrême). Le travail a cependant dû être séparé en deux car la quantité de résultats générés était jugée trop importante pour un seul chapitre (bien que les méthodes du premier chapitre puissent être utilisées indépendamment du

deuxième chapitre). De plus, les deux articles proposent plusieurs sous-étapes qui rendent le travail final complexe. Dans le même but que le deuxième chapitre, la régression quantile (QR) est considérée et évaluée dans le troisième chapitre. Il s'agit d'une approche plus directe car elle nécessite moins de sous-étapes, en particulier celle de sélection des pics et d'ajustement d'une distribution extrême à ces derniers. La régression quantile permet d'obtenir des résultats différents (mais complémentaires) en examinant plusieurs quantiles de la distribution conditionnelle des variables sanitaires (les pics sont considérés comme des quantiles élevés) et aussi sur plusieurs lags. Les relations entre les pics sanitaires et les conditions météorologiques sont ainsi analysées dans un niveau de détail poussé, ce qui n'a pas été fait en santé publique.

Finalement, dans le quatrième chapitre, les *extrêmes météorologiques* sont étudiés, de sorte à obtenir un portrait complet qui n'a pas encore été décrit dans la littérature actuelle. Puisqu'aucune définition générale des extrêmes n'existe, plusieurs définitions statistiques pour ces derniers sont évaluées dans le quatrième chapitre. En utilisant des méthodes statistiques, on s'assure ainsi une certaine objectivité dans la sélection des extrêmes à modéliser. Ceci est important car les extrêmes impactent directement les modèles, les formes des relations et les décisions qui en découleront. Toutefois, il n'est pas simple de proposer une définition générale car les événements extrêmes météorologiques dépendent de plusieurs variables (température, humidité, etc.). En effet, le cas bivarié ne permet pas un classement des observations simple, c'est pourquoi en plus des méthodes univariées, certaines méthodes multivariées comme les fonctions de profondeur et les indices météorologiques sont appliquées. Par ailleurs, un cadre théorique fait encore défaut concernant la régression des extrêmes météorologiques car aucune étude en statistique, en santé publique ou en météorologie ne l'a encore explorée. Les complications qu'elle entraîne sont très

similaires à celles pour les modèles des pics sanitaires, mais avec une difficulté supplémentaire liée aux aspects multivariés des événements.

La Figure 0-1 schématise les liens entre les différents chapitres de la thèse.

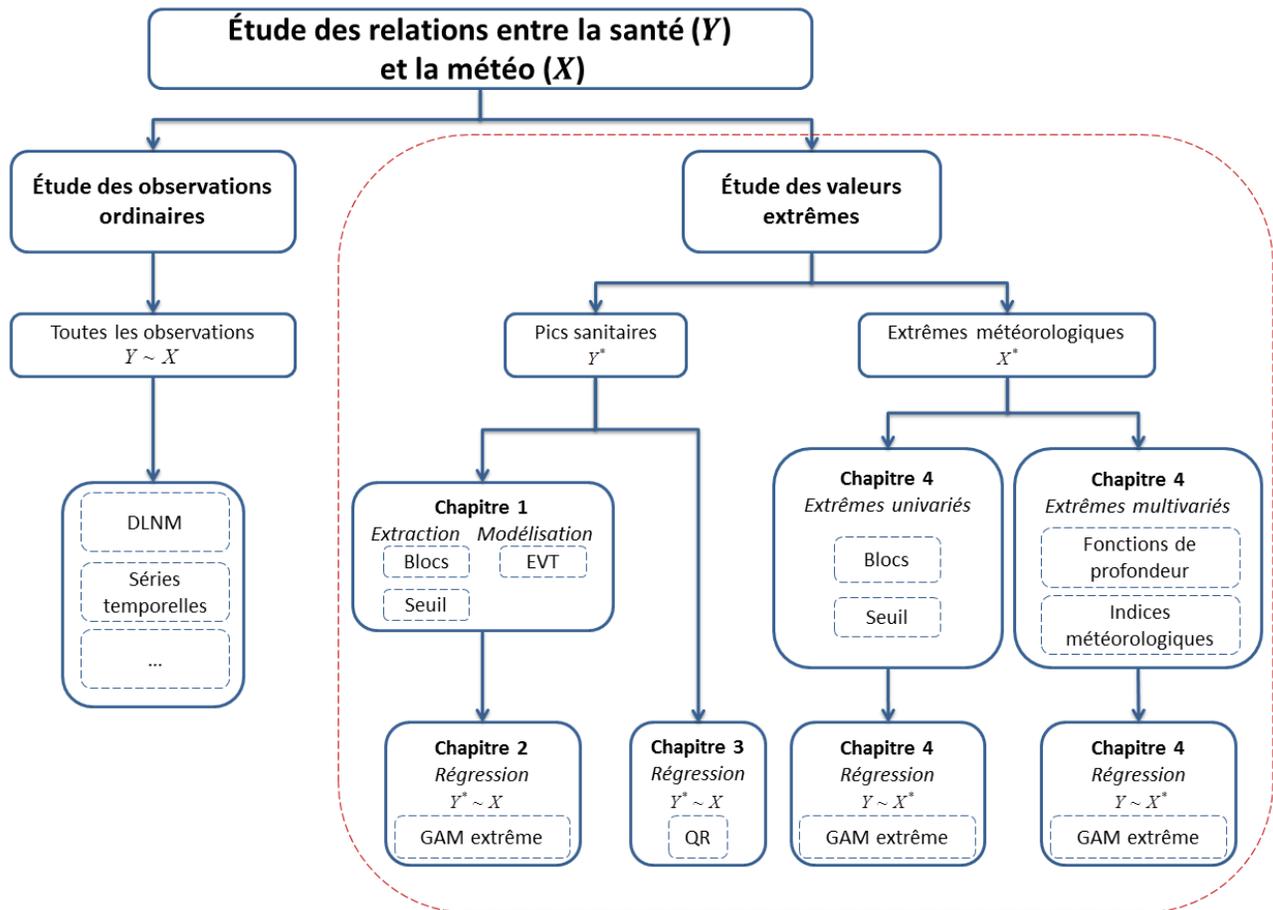


Figure 0-1 : Représentation des différents liens entre les chapitres de la thèse. Les problématiques visées par la thèse sont encadrées en rouge. L'étoile indique le caractère extrême de la variable.

DNLM : modèles non linéaires à effets retardés distribués; EVT : Théorie des valeurs extrêmes; GAM : Modèles additifs généralisés; QR : Régression quantile.

1.4. Organisation du document

Le reste de la synthèse est organisé de la façon suivante. Dans la section 2, une revue de la littérature des relations entre les MCV et les conditions météorologiques, des méthodes statistiques utilisées en santé publique puis des méthodes appliquées aux pics et extrêmes est présentée. Dans la section 3, les méthodologies qui ont donné lieu aux articles sont présentées

tandis que la section 4 résume les principaux résultats. Finalement, la section 5 contient les conclusions et des perspectives de recherche dans le domaine de la santé environnementale, lorsque des variables (sanitaires ou météorologiques) de nature extrême sont impliquées.

2. REVUE DE LA LITTÉRATURE

Cette section présente une revue de littérature concernant les différents aspects de la thèse. Un portrait général des MCV ainsi que de leurs relations avec la météo est d'abord dressé. Ensuite, les différents modèles statistiques utilisés dans la littérature sont présentés. Enfin, les différentes techniques d'études des extrêmes sont décrites (tous domaines confondus et en santé publique).

2.1. État de l'art sur les MCV et la météo

L'Organisation mondiale de la Santé définit les MCV comme la famille des maladies qui touchent le cœur et les vaisseaux sanguins (Organisation mondiale de la santé, 2015). On peut y trouver des maladies telles que l'infarctus du myocarde ou l'accident vasculaire cérébral. Ce sont donc des maladies relativement graves, qui peuvent mener à l'hospitalisation ou au décès. La gravité de ces maladies se traduit par des coûts élevés : au Canada, les MCV sont la première cause d'hospitalisation et la deuxième cause de décès, toutes causes confondues (Daigle, 2007; Girard, 2013). Elles représentent de plus la famille de maladies qui pèse le plus sur le système de la santé, avec des coûts directs estimés à 11,7 milliards de dollars par année, selon le dernier rapport de l'agence de santé publique du Canada (Agence de la santé publique du Canada, 2013). Le Canada n'est pas le seul pays dans cette situation puisqu'aux États-Unis, entre 2010 et 2030, les coûts directs liés aux MCV sont prévus de tripler, de 272 milliards à 818 milliards de dollars (Heidenreich *et al.*, 2011). En Europe, les coûts totaux sont estimés à 169 milliards d'euros par année (Leal *et al.*, 2006). En fait, la situation se généralise au monde avec une

estimation de 17,5 millions de personnes de décès par année, soit la première cause mondiale de décès (31 % des décès totaux, Organisation mondiale de la santé, 2015).

Le nombre de décès dû aux MCV est toutefois en baisse depuis les années 1960 (Agence de la santé publique du Canada, 2009). Les progrès de la médecine et des technologies médicales ont participé à cette diminution (par exemple par l'administration des statines ou l'implantation d'un stent, Law *et al.*, 2003; Wholey *et al.*, 2003). Si les décès de l'ensemble des MCV semblent diminuer, chaque maladie possède des taux différents (Tu *et al.*, 2009). Les hospitalisations quant à elles semblent être en augmentation. Plusieurs explications sont proposées par la littérature. Entre autres, l'obtention de meilleurs outils de diagnostic (Jin *et al.*, 2009; Nazerian *et al.*, 2010) permet de diagnostiquer plus de MCV ainsi qu'une prise en charge hospitalière plus rapide. Le nombre de chirurgies semble aussi s'être accru (Lloyd-Jones *et al.*, 2010), ce qui augmenterait le risque lié aux opérations et donc le nombre d'hospitalisations. Ces caractéristiques des hospitalisations et des décès sont observables sur les données utilisées dans cette thèse (voir la section 3).

Il existe principalement 4 classes de facteurs de risque reconnus : les facteurs génétiques (Kathiresan et Srivastava, 2012), les facteurs psychosociaux (tels que le stress ou l'anxiété; Aldana *et al.*, 2006), les habitudes de vie (en particulier une mauvaise alimentation, le tabagisme, le manque d'activité physique et l'usage nocif de l'alcool; Daigle, 2007) et les conditions environnementales (Portier *et al.*, 2010). Étant donné l'importance des MCV, chaque classe est grandement étudiée. En particulier, la dernière classe est d'un grand intérêt puisque les changements climatiques risquent de modifier les relations avec les MCV connues actuellement.

La température influence plusieurs paramètres physiologiques du corps humain (tels que la viscosité sanguine ou les battements cardiaques; Keatinge *et al.*, 1986). Il est ainsi

compréhensible qu'elle soit omniprésente dans la littérature et qu'elle entraîne des décès et hospitalisations de toutes causes (Kovats et Hajat, 2008). Les MCV ne font pas exception, et l'effet des températures chaudes ou froides est régulièrement rapporté, au Canada (Goldberg *et al.*, 2011; Lavigne *et al.*, 2014; Chen *et al.*, 2016), aux États-Unis (Barnett *et al.*, 2012), en France (Laaidi *et al.*, 2006), en Chine (Guo *et al.*, 2013) ou encore en Espagne (Linares *et al.*, 2015). Il est toutefois à noter que l'influence des températures est différente selon le type de climat. Par exemple, Goldberg *et al.* (2011) observent une influence accrue des températures élevées sur les décès par MCV à Montréal (Canada), tandis que Khanjani et Bahrampour (2013) trouvent une influence accrue des températures froides à Kerman (Iran). Cette différence est probablement due à l'adaptation physiologique des populations. La plupart des études indiquent une relation non-linéaire entre la température et la santé, en forme de U, de V ou de J (Martens, 1998; Gouveia *et al.*, 2003; Doyon *et al.*, 2008). En effet, le corps humain possède une zone de confort thermal en dehors de laquelle il lui est plus difficile de fonctionner normalement. L'influence des températures froides et des températures chaudes se traduit ainsi par des influences non-linéaires. De plus, Gasparrini *et al.* (2015) et Rytty *et al.* (2016) notent que les températures froides ont un plus grand effet que les températures chaudes. En ce qui concerne les MCV, les relations semblent non-linéaires sans nécessairement présenter de formes particulières (Bhaskaran *et al.*, 2010).

D'autres variables météorologiques telles que l'humidité relative (Abrignani *et al.*, 2012), la pression atmosphérique (Danet *et al.*, 1999) et la quantité de neige au sol (Modarres *et al.*, 2012) ont montré une influence sur les MCV. Les variables de pollution de l'air influent aussi les MCV (Zanobetti *et al.*, 2000). Ce ne sont pas des variables météorologiques à proprement parler, mais elles sont directement liées aux conditions météorologiques et sont souvent étudiées

conjointement avec celles-ci (Ren *et al.*, 2011; Mustafic *et al.*, 2012). En fait, l'effet des variables de pollution est potentiellement majeur et a motivé la création d'une nouvelle sous-discipline en santé publique, la cardiologie environnementale (Bhatnagar, 2006).

Considérant l'effet des variables météorologiques, il est important de prendre en compte les changements climatiques observés depuis plusieurs décennies. En effet, ces derniers touchent l'ensemble du globe, et le Canada n'y échappe pas. Au Québec en particulier, les températures journalières moyennes n'ont cessé d'augmenter depuis 1950 (entre 1 et 3 °C selon la zone étudiée du Québec). Ces changements vont vraisemblablement persévérer et même s'accroître dans un futur à court et moyen termes (IPCC, 2014). Autant de variables météorologiques qui vont ainsi être modifiées (voir la Figure 0-2), avec pour conséquences des relations avec la santé humaine différentes de celles d'aujourd'hui. Dans ce contexte, encore bien des questions restent ouvertes quant aux effets précis du changement climatique sur la santé (Patz *et al.*, 2005; Knowlton *et al.*, 2007; Watts *et al.*, 2015) et plus particulièrement sur les MCV (Costello *et al.*, 2009; Hajat *et al.*, 2014; Watts *et al.*, 2015), qui, rappelons-le, sont la première cause de décès dans le monde. Entre autres, l'effet exact de l'augmentation de la fréquence et de la durée des canicules sur les MCV n'est pas connu. En effet, si plusieurs études sur les canicules passées ont déjà été menées en rapport avec les MCV, l'incertitude liée aux changements climatiques ne permet pas de conclure quant à l'évolution des MCV durant les canicules futures. Par exemple, il est connu que les températures chaudes sont associées à de plus hauts taux de pollution de l'air, qui en retour influencent l'apparition de MCV (Spickett *et al.*, 2011). Cependant, malgré le réchauffement prévu, il est possible que le nombre des décès par MCV durant l'hiver ne diminue pas nécessairement. Ebi et Mills (2013) indiquent que la mortalité globale (dont par MCV) due à des températures plus douces durant les hivers n'aura pas pour conséquence de réduire les décès pour

équilibrer la balance avec les étés. Ainsi, afin de pouvoir prévoir les relations futures entre les MCV et les variables météorologiques, il est crucial de les modéliser de façon adéquate.

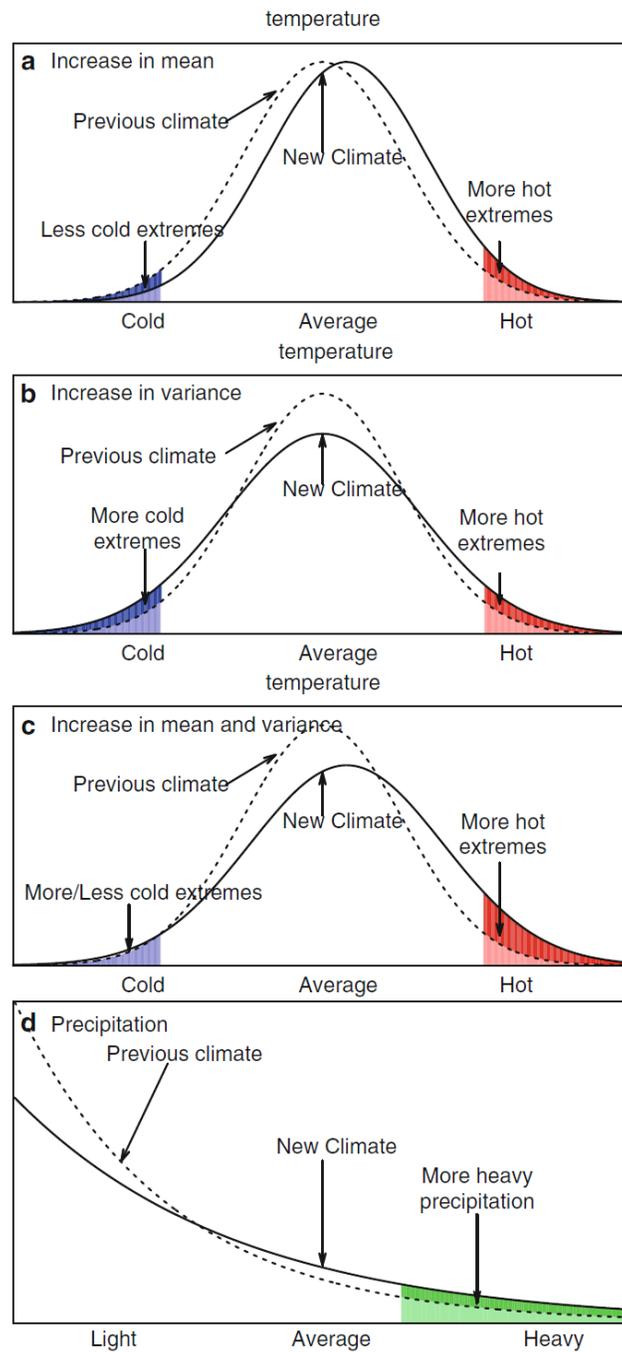


Figure 0-2: Changements possibles et implications dans la distribution de la température et des précipitations sous les effets du changement climatique (Zwiers *et al.*, 2013).

2.2. État de l'art sur les méthodes statistiques en santé publique

Puisqu'il est reconnu que les conditions météorologiques influencent les MCV, il est nécessaire d'en quantifier l'effet exact. Dans ce sens, plusieurs méthodes statistiques existent. Toutes les études en santé publique n'ont pas nécessairement pour objet les MCV, c'est pourquoi certaines méthodes citées ici sont utilisées pour d'autres causes de mortalité ou de morbidité. La Figure 0-3 résume les différentes approches mentionnées dans cette section, qui sont habituellement utilisées de façon indépendante. Elles ont toutes pour but d'étudier une issue sanitaire avec les conditions météorologiques. Dans cette thèse, on s'intéresse particulièrement aux différentes approches de régression, bien que certaines méthodes tirées des autres approches soient aussi abordées. De plus, des méthodes liées aux valeurs extrêmes sont aussi mises en place (prochaine sous-section), mais étant donné qu'elles ne sont pas utilisées en santé publique, elles ne sont pas présentées à la Figure 0-3.

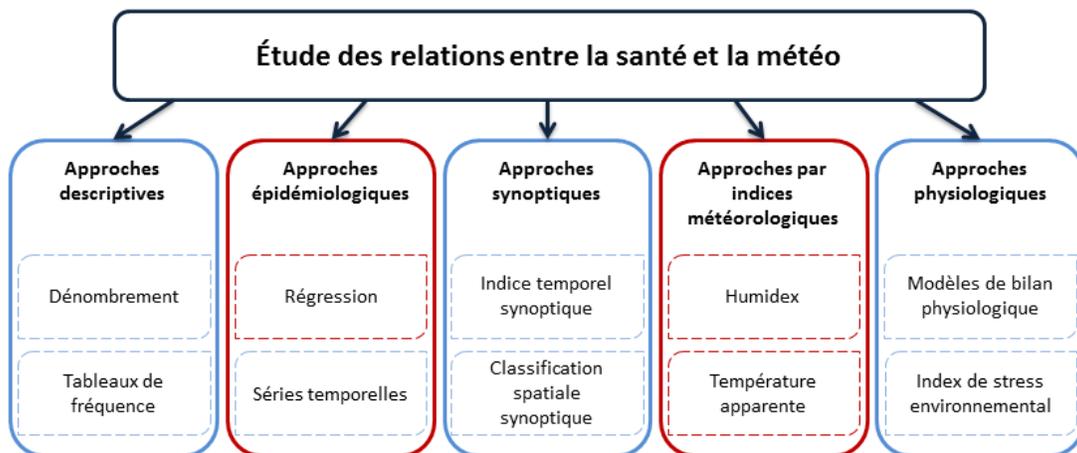


Figure 0-3 : Résumé des différentes approches utilisées dans l'étude des relations entre la santé et la météo. Les approches utilisées dans cette thèse sont en rouge.

La première étape avant la modélisation entre une variable à expliquer et des variables explicatives est l'analyse descriptive. Dans ce cadre, il est possible de citer le dénombrement (Greenough *et al.*, 2001) ou les tableaux de fréquence (Curriero *et al.*, 2001). Ces approches

mettent ainsi en évidence des liens qui existent potentiellement et les quantifient (par exemple, les coefficients de corrélation ont une interprétation directe sur la force de la relation). Cependant, la corrélation n'entraîne pas la causalité et elles ne permettent pas d'expliquer l'influence d'une variable explicative sur une variable d'intérêt.

La régression est un moyen naturel d'évaluer l'impact d'une variable explicative (variable météorologique ici) sur une variable à expliquer (variable sanitaire ici). Le modèle de base, la régression linéaire, qui présume d'une relation linéaire entre la variable à expliquer et les variables explicatives, semble toujours jouir d'une certaine popularité (Lumley *et al.*, 2002; Krantz *et al.*, 2013; Wang *et al.*, 2014). En effet, elle permet une interprétation immédiate des résultats et toutes les relations en santé publique ne sont pas nécessairement non-linéaires. Partant de cette constatation, il n'est alors pas nécessaire d'introduire une fonction non-linéaire dans les modèles, qui ajouterait une couche de complexité (et parfois une non généralisation du modèle).

Le modèle de régression linéaire suppose une distribution normale de la variable à expliquer. Cette hypothèse étant très restrictive en pratique, il est possible d'étendre la régression à une famille de distributions dite « famille exponentielle » (à ne pas confondre avec la distribution exponentielle) grâce aux modèles linéaires généralisés (*generalized linear models* ou GLM, Nelder et Wedderburn, 1972). Les GLM sont régulièrement appliqués dans l'étude des MCV (par exemple Cao *et al.*, 2012) ou en pollution de l'air (par exemple Mostofsky *et al.*, 2012). Ils ont par ailleurs déjà été appliqués au Québec (Doyon *et al.*, 2008) pour analyser l'effet du climat sur la mortalité de toutes causes.

Il existe aussi des relations non-linéaires entre la variable sanitaire et les variables environnementales qu'un modèle de base ne parviendrait pas à capturer de telles relations. Pour intégrer des relations non-linéaires dans un modèle de régression, les GLM sont étendus aux

modèles additifs généralisés (*generalized additive models* ou GAM, Hastie et Tibshirani, 1986). Ces derniers font usage de méthodes non paramétriques telles que les splines de régression. Une droite de régression étant un cas particulier d'une fonction de régression, on ne perd pas en généralité. Ils sont appliqués en majorité pour des études concernant les relations entre la santé (y compris les MCV) et les conditions météorologiques et/ou la pollution de l'air (Johnston *et al.*, 2011; Caussin *et al.*, 2015; Zhang *et al.*, 2016). Ces modèles sont de plus régulièrement cités pour leur flexibilité et leur capacité d'adaptation aux données dans le domaine de la santé publique (Dominici *et al.*, 2002; Jbilou et El Adlouni, 2012). Dans de tels modèles, il est coutumier de rajouter une fonction du temps ou même un indicateur du jour de la semaine, afin de contrôler les différentes tendances ou saisonnalités présentes dans les données.

En plus de l'aspect non-linéaire, il est utile de prendre en compte les retards (ou *lags* en anglais) de certaines variables. À cet effet, les modèles à effets retardés distribués peuvent être utilisés (DLNM, Gasparrini *et al.*, 2010). Ils sont populaires dans l'étude des relations entre la santé et le climat (Goldberg *et al.*, 2011; Lavigne *et al.*, 2014; Phung *et al.*, 2016a), mais ils nécessitent l'estimation d'un nombre important de coefficients. En effet, la caractéristique principale d'un DLNM est sa fonction bidimensionnelle pour chaque variable explicative (une dimension pour la variable explicative et une dimension pour les lags appliqués sur cette dernière). Ces modèles demandent donc un nombre important de données pour être estimés correctement. Les études de Lavigne *et al.* (2014) ou de Gasparrini *et al.* (2015) font par exemple usage du DLNM dans des études des températures extrêmes sur la santé, mais les échantillons sont largement supérieurs ($n > 200000$) à ceux présents dans cette étude ($n < 12000$ pour les décès et $n < 4100$ pour les hospitalisations), leur conférant ainsi plus de puissance statistique. Ils ne se concentrent toutefois

pas exclusivement sur ces extrêmes de température comme il est proposé ici et n'étudient pas les pics sanitaires.

La régression quantile (QR, Koenker, 2005) est une méthode de modélisation très proche de la régression, dans le sens où elle permet d'expliquer une variable dépendante en fonction de d'une ou plusieurs variables explicatives. L'interprétation des paramètres obtenus par la QR est d'ailleurs similaire à celle de la régression classique. La différence réside dans le fait que la QR est capable d'estimer l'influence des variables explicatives dans n'importe quel quantile de la variable à expliquer, contrairement à une régression classique, où seule la moyenne conditionnelle est analysée. Cet apport majeur de la QR permet une représentation plus complète des relations. En effet, même s'il y a une faible (ou aucune) relation entre la moyenne conditionnelle de la variable réponse et les variables explicatives, il est possible que d'autres relations significatives existent dans le reste de la distribution conditionnelle (Cade et Noon, 2003). La QR est reconnue dans plusieurs problématiques de la santé publique, en particulier pour des indicateurs susceptibles de varier dans la distribution de la variable sanitaire. À titre d'exemple, Bind *et al.* (2015) et Bind *et al.* (2016) étudient la façon dont les particules fines et le charbon influencent certains biomarqueurs connus des MCV. Leurs résultats leur permettent de conclure que les effets de la pollution peuvent être prononcés parmi les individus déjà fortement à risque, autrement dit dans les pics. La QR est une approche flexible car elle ne présume pas de distribution particulière pour la variable dépendante et il est possible de la combiner avec des méthodes non-paramétriques (Fenske *et al.*, 2013) ou des séries temporelles (Soyiri et Reidpath, 2013).

Les séries temporelles représentent une autre grande famille de méthodes statistiques en santé publique (Zeger *et al.*, 2006; Modarres *et al.*, 2012; Krall *et al.*, 2013). Elles sont centrées sur

l'aspect temporel et la dépendance dans le temps des observations sanitaires. Autrement dit, il est crucial d'avoir à disposition des observations consécutives ordonnées par le temps. Or, les pics sanitaires ou extrêmes météorologiques ne surviennent pas les uns à la suite les autres. Les indices temporels sont donc interrompus et les séries temporelles ne sont pas adaptées à ce contexte de modélisation. Le cas des « séries temporelles interrompues » est toutefois un sujet ouvert, même en santé publique (Bernal *et al.*, 2016). En particulier, un cadre théorique reste à établir.

Les approches introduites jusqu'à présent considèrent les variables météorologiques séparément dans la régression, ou au mieux en font interagir deux. Elles contrastent avec les approches synoptiques qui sont des approches « globales » (Sheridan, 2002), largement répandues aux États-Unis mais moins au Canada (Hondula *et al.*, 2014). Dans le sens d'approche globale, leur philosophie ressemble à celle des indices météorologiques et physiologiques. Eux aussi combinent plusieurs variables (souvent la température et l'humidité, Smoyer-Tomic et Rainham, 2001; Blazejczyk *et al.*, 2012) pour en construire de nouvelles, jugées plus représentatives de l'effet de la météo sur le corps humain. Dans une comparaison des approches synoptiques, des méthodes épidémiologiques (GAM) et des indices météorologiques et physiologiques, Hajat *et al.* (2010) parviennent à deux conclusions générales : 1) L'efficacité des seuils d'alerte de vagues de chaleur dépend de la méthode implémentée car les résultats entre ces différentes méthodes sont discordants dans l'identification des jours à risque et 2) Les méthodes basées sur les approches synoptiques et épidémiologiques performant généralement mieux que les deux autres.

2.3. Étude des valeurs extrêmes

Un exemple simple de modèle de régression impliquant une variable sanitaire Y (variable dépendante) et des variables météorologiques T , H et P (température, humidité et pression comme variables explicatives) s'écrit

$$Y = \beta_0 + \beta_1 T + \beta_2 H + \beta_3 P + \varepsilon \quad (1)$$

avec ε l'erreur du modèle. Un cas particulièrement problématique dans l'étude des conditions météorologiques sur la santé (contexte de régression) est celui des valeurs extrêmes, qui peuvent intervenir dans l'une ou l'autre des variables. En effet, elles peuvent survenir d'un point de vue sanitaire sur Y (les *pics sanitaires*, côté gauche de l'équation (1)) ou d'un point de vue météorologique sur T , H ou P (les *extrêmes météorologiques*, côté droit de l'équation (1)). Dans un cas comme dans l'autre, les valeurs extrêmes sont caractérisées par une fréquence d'occurrence plus rare et un impact plus important que ceux des autres observations. Il est ainsi inapproprié de se tourner vers les modèles classiques, qui utilisent l'ensemble des observations et s'intéressent de ce fait à la moyenne de la distribution de la variable dépendante. Ces modèles sont évidemment cruciaux pour la prévision et l'adaptation aux événements communs, mais ne sont pas aptes à prendre en compte les valeurs extrêmes. Pour mener une étude des pics sanitaires ou des extrêmes météo, il est donc nécessaire d'utiliser des outils statistiques spécifiques.

La Figure 0-4 illustre un exemple de différence fondamentale entre la moyenne et les valeurs extrêmes. Il est connu que les sommes (ou de façon équivalente, les moyennes) de plusieurs variables aléatoires indépendantes et identiquement distribuées convergent vers une loi normale, même si ces dernières ne sont pas distribuées de façon normale. Il s'agit d'un résultat régulièrement pris en compte en santé publique (Lumley *et al.*, 2002). Par exemple, dans le cadre de la modélisation du caractère moyen d'une variable, il permet d'établir les intervalles de

confiance des paramètres (Lejeune, 2011). Cependant, ce résultat asymptotique n'est plus vérifié dans le cas du maximum ou du minimum (au lieu de la moyenne). La Figure 0-4 représente la convergence de 200 moyennes de variables qui suivent une loi uniforme. Or, il est clair que les maxima ne suivent pas une loi normale. Pour étudier correctement les maxima et leurs propriétés, il est essentiel de mettre en œuvre des outils de sélection et de modélisation adaptés. Il apparaît ainsi naturel de se tourner vers la théorie des valeurs extrêmes (*extreme value theory* ou EVT, Fisher et Tippett, 1928; Balkema et Dehaan, 1974). Son objet d'étude est justement l'analyse des valeurs extrêmes et non plus de la moyenne. Dans l'exemple des 200 variables uniformes, l'EVT montre que les maxima suivent plutôt une distribution de Weibull.

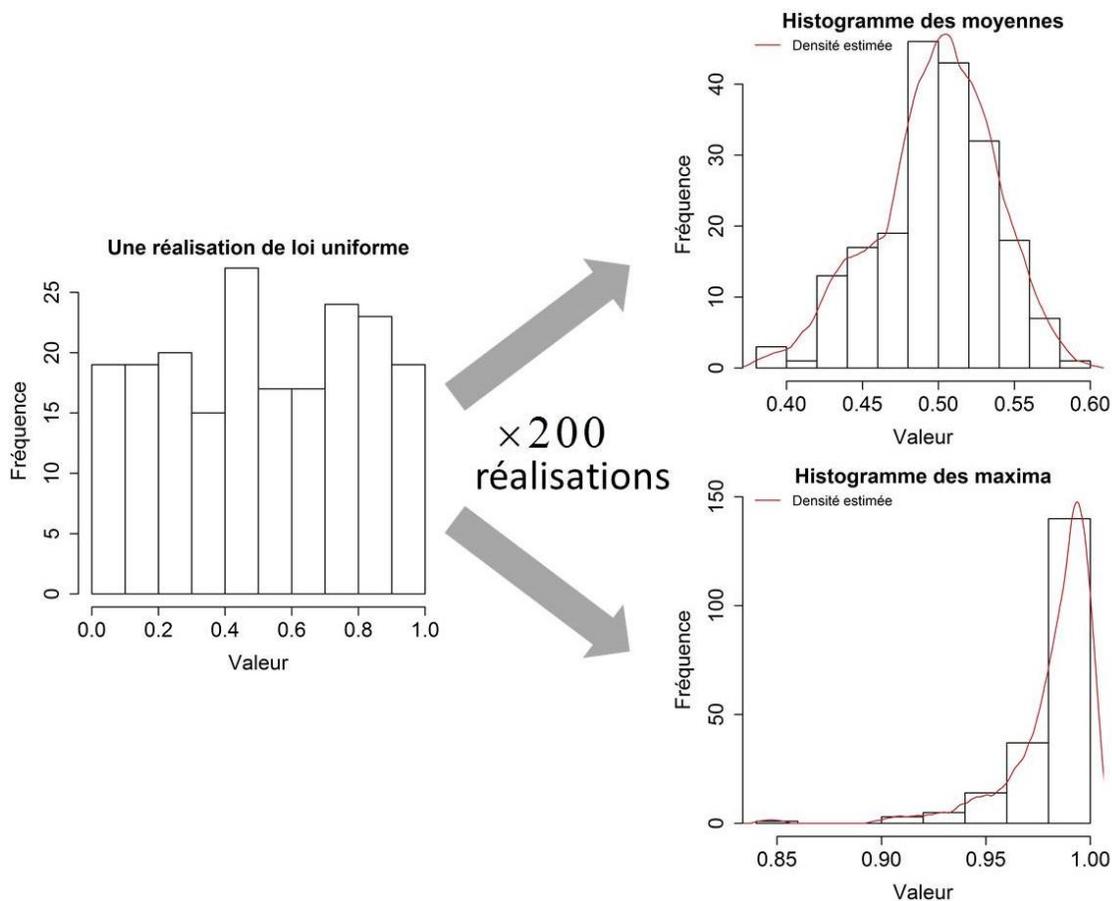


Figure 0-4: Illustration de la convergence d'une loi uniforme. Les moyennes convergent vers une loi normale, les maxima vers une loi de Weibull.

Le *niveau de retour* constitue l'un des résultats importants en l'EVT. Il s'agit du niveau z_p qui sera dépassé en moyenne une fois toutes les $1/p$ années (Coles, 2001). Le niveau de retour est donc automatiquement associé à une période de retour $1/p$ et une probabilité p . Puisqu'il s'agit d'une occurrence moyenne, le niveau de retour peut être atteint plusieurs fois, une fois, voire aucune fois sur la période considérée. Pour pouvoir le calculer, il est donc important d'associer une distribution extrême adéquate aux minima ou au maxima.

En santé publique, la littérature concernant les pics sanitaires est importante. En revanche, les études qui se consacrent à la modélisation des pics sanitaires sont rares. Les méthodes statistiques de l'EVT y sont à plus forte raison très peu exploitées, contrairement à des domaines comme l'hydrologie (El Adlouni *et al.*, 2008) ou la finance (Gilli, 2006). Quelques récentes études marginales qui en font usage tirent toutefois des conclusions intéressantes. Par exemple, De Zea Bermudez et Mendes (2012) s'intéressent au niveau de cholestérol dans différentes régions du Portugal. En appliquant la méthodologie du seuil associée à la distribution généralisée de Pareto (GPD), ils modélisent les niveaux élevés de cholestérol. De leur côté, Guillou *et al.* (2014) examinent les occurrences de différentes souches de *Salmonella* en France. L'intérêt de leur étude étant la surveillance, les auteurs s'attardent principalement sur les niveaux de retour plutôt que sur la modélisation en tant que telle. Chen *et al.* (2015) se servent de plus de la même méthodologie pour prédire les niveaux de retour de l'influenza en Chine. Enfin, Thomas *et al.* (2016) modélisent les maxima annuels (mensuels) de la mortalité (morbidité) par la méthode des blocs et la distribution des valeurs extrêmes généralisée (GEV). Ils estiment ensuite les niveaux de retour associés. Ces quatre études indiquent que l'EVT est toute indiquée pour l'étude des pics sanitaires, en particulier à des fins de surveillance. Cependant, aucune d'elles n'utilisent de méthodologie complète qui permette de comparer la méthode du seuil et celle des blocs. De plus,

elles ne vérifient pas les hypothèses nécessaires à la modélisation en EVT. Elles n'intègrent enfin pas les pics sanitaires extraits dans une optique explicative grâce à des modèles de régression.

En santé publique, les extrêmes météorologiques sont aussi grandement étudiés en raison de leurs conséquences sur la santé. En effet, l'apparition d'un extrême météorologique peut causer un stress intense sur le système de la santé (par exemple, les inondations ou les vagues de chaleur, Ashley et Ashley, 2008; Knowlton *et al.*, 2009). Un problème majeur est qu'il n'existe pas de définition globale et objective d'un extrême météorologique. Une règle de base proposée est de définir les extrêmes météorologiques comme les observations supérieures au quantile à 90 % de la variable en question (IPCC, 2013). Toutefois, elle est loin d'être générale et la majorité des études ont recouru à des définitions variables. Par exemple, les températures extrêmes varient d'une zone géographique à l'autre, entraînant des définitions des canicules différentes (Ryti *et al.*, 2016). Ebi et Schmier (2005) précisent d'ailleurs que les définitions des événements extrêmes devraient être établies en collaboration entre les chercheurs et les agences de santé publique. De nombreuses études s'intéressant aux effets des extrêmes météorologiques sur la santé utilisent l'ensemble des observations et incluent une variable indicatrice (fixée à un certain seuil élevé) pour quantifier l'effet de ces extrêmes (Anonyme, !!! INVALID CITATION !!! Kovats *et al.*, 2004; Hajat *et al.*, 2007; Li *et al.*, 2012). Néanmoins, considérer l'ensemble des observations peut avoir pour conséquence de diluer l'information liée aux extrêmes météorologiques et peut ne pas permettre d'apprécier pleinement les caractéristiques de ces derniers.

Dans d'autres contextes que la santé publique, les valeurs extrêmes sont étudiées dans la régression grâce à des modèles spécifiques. Ces modèles ont été principalement développés pour intégrer le caractère non-stationnaire d'une distribution. Le but est alors de modéliser les paramètres de la distribution extrême en fonction de covariables (p. ex. indices climatiques ou

latitude et longitude, Chavez-Demoulin et Davison, 2005; Padoan *et al.*, 2010). Les valeurs extrêmes des variables météorologiques sont bien modélisées, mais l'objectif est différent de celui de la présente étude. En effet, dans cette thèse, on tente d'expliquer une variable sanitaire (ou des pics sanitaires) en fonction de variables météorologiques (ou des extrêmes météorologiques). Par exemple, les modèles des études citées ne permettent pas d'obtenir l'influence directe d'une variable météorologique sur les pics d'hospitalisations.

L'EVT dans le cas univarié a fait l'objet de plusieurs de travaux (Beirlant *et al.*, 2012; Gomes et Guillou, 2015). Toutefois, les extrêmes multivariés existent aussi et sont nombreux (par exemple, une crue peut être caractérisée par sa hauteur, son débit et sa durée). La méthode du bloc et du seuil dans un contexte à plus d'une dimension consiste à extraire les extrêmes de chaque variable puis à les associer entre eux (Coles *et al.*, 1999; Salvadori, 2007), créant ainsi parfois des combinaisons de valeurs inexistantes dans la série originale (Coles, 2001). Par exemple, si un extrême de température t_i^* survient au jour i et un extrême d'humidité h_j^* à un autre jour j , une observation simple s'écrirait

$$y_\gamma = \beta_0 + \beta_1 t_i^* + \beta_2 h_j^* + \varepsilon_\gamma \quad (2)$$

avec y la variable sanitaire et l'étoile indiquant le caractère extrême de la variable. Il est alors impossible d'associer un indice cohérent à la variable sanitaire, d'où l'indice « ? ». Les méthodes classiques de sélection des extrêmes rencontrent ainsi des difficultés dans un contexte multivarié. De plus, ordonner un échantillon univarié est direct puisqu'il se fait selon la variable d'étude. Il n'est en revanche plus aussi simple d'obtenir un classement des observations multivariées puisqu'elles sont définies selon plusieurs variables. Sans classement simple, il est donc complexe de définir des extrêmes. Il faut noter que les méthodes de sélection d'extrêmes multivariés ne sont pas encore entièrement développées, que ce soit dans les études théoriques ou appliquées.

Habituellement, dans un contexte multivarié, les études se concentrent directement sur la modélisation de la dépendance entre les extrêmes (Renard et Lang, 2007). Une des études qui s'intéressent à l'étape de sélection plutôt que de modélisation est celle de Chebana et Ouarda (2011b). Ils font appel aux fonctions de profondeur (Zuo et Serfling, 2000) pour définir des extrêmes multivariés.

La Figure 0-5 illustre la problématique des extrêmes météorologiques multivariés sélectionnés par bloc et par seuil pour la température et l'humidité. Il est clair qu'un extrême de température et un extrême d'humidité ne surviennent pas nécessairement le même jour, même si cela peut arriver. Avec deux variables météorologiques, les deux méthodes se soldent donc par des séries dont les valeurs extrêmes ne coïncident pas au niveau des dates.

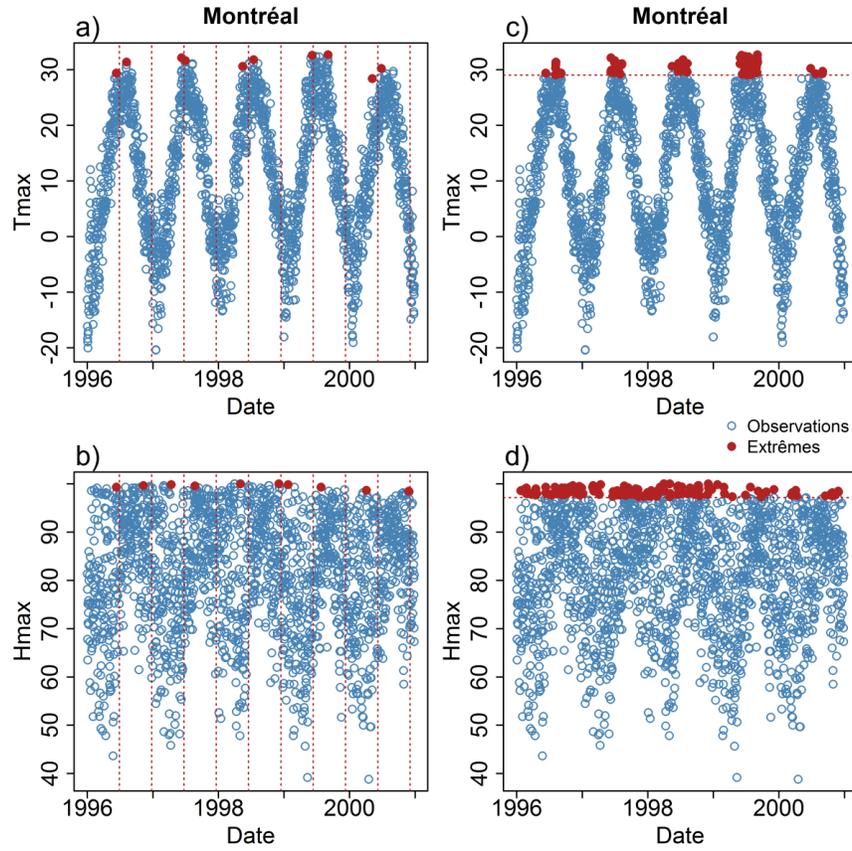


Figure 0-5 : a) Extrêmes de température par bloc, b) extrêmes d'humidité par bloc, c) extrêmes de température par seuil et d) extrêmes d'humidité par seuil, à Montréal.

3. DONNÉES

Cette section décrit les données utilisées. Deux types de données sont utilisés dans ce projet : des données météorologiques et des données sanitaires. Les données sont elles-mêmes divisées en données brutes et données extrêmes (pics sanitaires pour les variables sanitaires et extrêmes météorologiques pour les variables météorologiques). Les pics sanitaires et extrêmes météorologiques sont extraits des données brutes à partir des méthodes décrites dans la prochaine section.

Les zones étudiées sont les régions de Montréal et de Québec, représentées au travers de leurs communautés métropolitaines (CMM et CMQ respectivement, voir la Figure 0-6). Dans cette

thèse, pour simplifier la lecture, les termes CMM et Montréal, ou CMQ et Québec, sont synonymes. En 2016, Montréal et Québec possèdent des populations de 3 921 233 et de 797 219 habitants respectivement (Ministère des Affaires municipales et de l'Occupation du territoire, 2016; Communauté métropolitaine de Montréal, 2017).

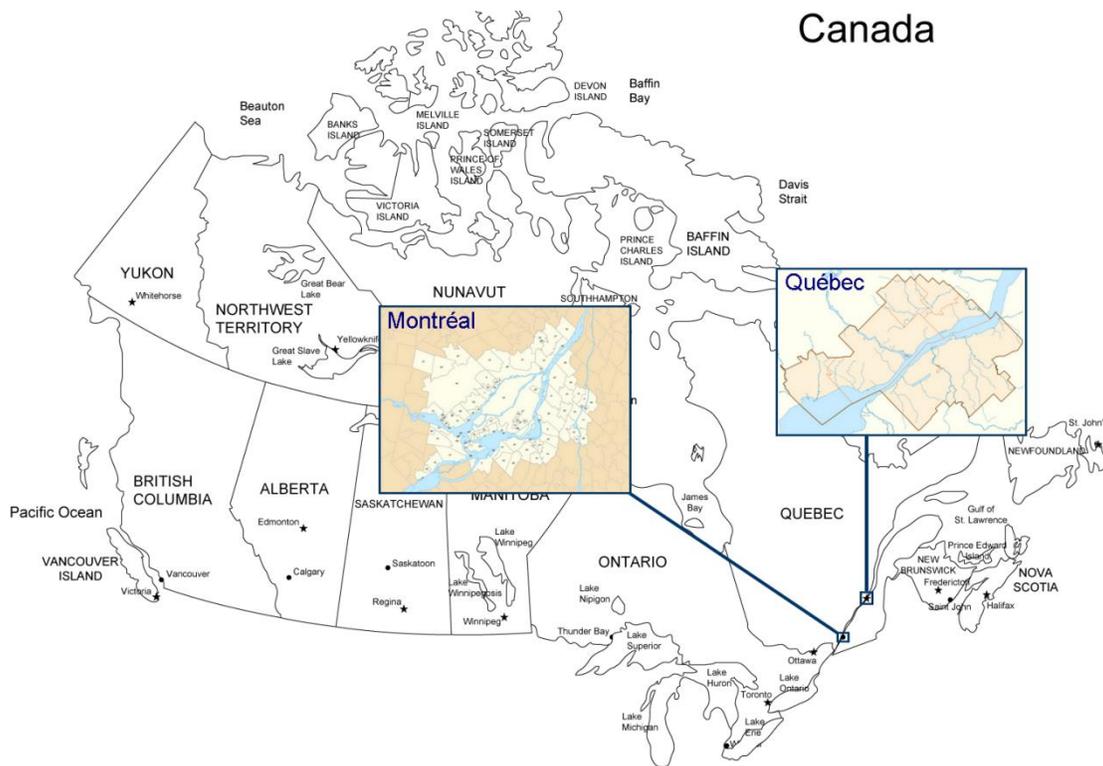


Figure 0-6: Communautés métropolitaines de Montréal et de Québec.

3.1. Données sanitaires

Les données sanitaires à disposition sont les hospitalisations (diagnostic principal ou secondaire) et les décès (diagnostic principal) pour cause de MCV les plus mortelles (Tableau 0-1, dans lequel les fichiers de rémunération à l'acte n'ont en réalité pas été utilisés dans la présente étude). Elles proviennent du fichier MED-ÉCHO et ont été fournies au centre Eau Terre Environnement

de l'Institut national de recherche scientifique par l'Institut national de santé publique du Québec. Les hospitalisations et les décès ont été agrégés de sorte à obtenir des indicateurs journaliers sommés par communauté métropolitaine. La Figure 0-7 montre les données d'hospitalisations et de décès utilisées dans cette thèse.

Tableau 0-1 : Définitions des MCV considérées dans le Programme de recherche MCV, incluant leurs codes CIM-9 et CIM-10.

MCV	Définitions (un ou l'autre des deux critères)	CIM-9	CIM-10
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
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) OU un diagnostic (principal ou secondaire) d'IC enregistré au fichier MED-ÉCHO (≥ 40 ans).	428	I50
Maladies vasculaires cérébrales (MVC) et ischémie cérébrale transitoire. (Définition du 29 avril 2013, Agence de la santé publique du Canada)	Deux diagnostics ou plus de MVC 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 MVC 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
		Codes dans les services médicaux 436, 432, 435	

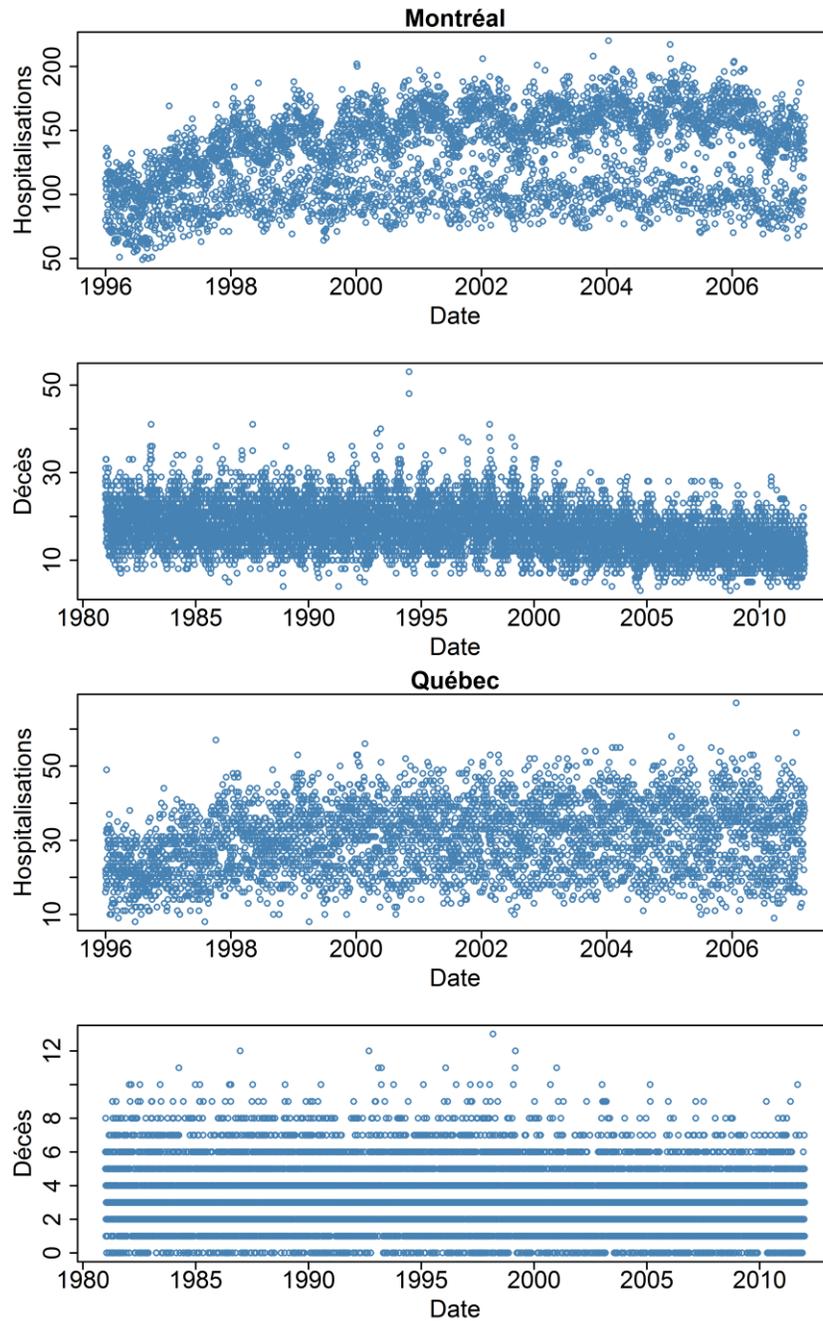


Figure 0-7 : Hospitalisations et décès journaliers à Montréal et Québec.

La période couverte est de janvier 1996 à mars 2007 inclusivement pour les hospitalisations et de janvier 1981 à décembre 2011 inclusivement pour les décès. On dispose d'indicateurs journaliers pour les hospitalisations et décès, soit un total de 4 077 observations pour les hospitalisations et 11 322 observations pour les décès, pour chacune des deux communautés métropolitaines. Les

statistiques descriptives sont données au Tableau 0-2. La population de Montréal étant plus importante, les nombres d'hospitalisations et de décès sont plus élevés dans cette dernière. Par ailleurs, les hospitalisations sont plus nombreuses que les décès dans les deux villes.

Tableau 0-2 : Statistiques descriptives des données sanitaires brutes à Montréal et à Québec.

	Montréal		Québec	
	Décès	Hospitalisations	Décès	Hospitalisations
Minimum	3	49	0	8
Maximum	53	220	13	67
Moyenne	17	131	3	30
Médiane	17	136	3	30
Q75	20	158	4	37
Q90	24	172	6	42

3.2. Données météorologiques

Les données météorologiques sont issues du portail de téléchargement DAI d'Environnement Canada (Données Accès et Intégration). Les variables météorologiques utilisées sont la température, la pression atmosphérique, l'humidité relative (valeurs minimales, moyennes et maximales pour les trois variables), les précipitations totales et l'épaisseur de neige au sol (degrés Celsius, kilopascals, pourcentage, millimètres et centimètres, respectivement). Elles sont mesurées à partir de 27 et 18 stations différentes pour Montréal et Québec, respectivement. Ces données sont journalières et disponibles pour la période 1981-2011, en cohérence avec les décès et les hospitalisations par MCV. De plus, pour le quatrième chapitre, une variable de pollution de l'air a été ajoutée (PM_{2,5}).

Les statistiques descriptives des données brutes sont résumées au Tableau 0-3 (pour les variantes maximales de la température, humidité et pression). Elles montrent qu'il fait plus frais (selon les statistiques minimale et moyenne) à Québec, tandis que l'humidité est plus élevée à Montréal. La

différence la plus notable est la quantité de neige : Québec en reçoit nettement plus (en moyenne, 18 cm journaliers en hiver, contre 6 cm pour Montréal).

Tableau 0-3 : Statistiques descriptives des données météorologiques brutes à Montréal et à Québec.

	Québec					Montréal				
	Tmax	Hmax	Pmax	Prec	Snow	Tmax	Hmax	Pmax	Prec	Snow
Min	-26,46	34,00	99,06	0,00	0,00	-28,51	29,25	98,86	0,00	0,00
Max	35,05	100,00	105,2	89,77	79,23	34,84	100,00	105,28	77,67	170,5
Moy	11,42	81,59	101,96	2,83	6,78	9,67	78,26	101,88	3,34	17,67
Méd	12,23	84,00	101,92	0,32	0,00	9,88	80,00	101,85	0,45	0,00
Q75	22,37	91,83	102,47	3,00	8,88	20,41	89,00	102,42	3,77	31,87
Q90	26,85	96,00	103,00	8,87	27,41	25,4	94,25	102,95	10,62	63,2

Min : minimum, Max : maximum, Moy : moyenne, Méd: médiane, Q75: quantile à 75 %, Q90: quantile à 90 %

4. MÉTHODES PROPOSÉES

Cette section décrit les méthodes utilisées. Elle traite de la théorie des valeurs extrêmes, des modèles de régression non-paramétriques, de la régression quantile et des extrêmes multivariés.

4.1. Études des pics sanitaires par la théorie des valeurs extrêmes

Le premier chapitre traite de la sélection et la modélisation des pics sanitaires via l'application de l'EVT. L'objectif final est de mettre en relation les variables météorologiques avec les pics sanitaires, mais il est crucial de bien sélectionner ces derniers en premier lieu. Cette étape de sélection et d'inférence des variables à expliquer est souvent omise dans les études en santé publique, qui se concentrent plutôt sur des modèles explicatifs. Comme expliqué dans la revue de littérature, la théorie des valeurs extrêmes n'a été que très rarement appliquée dans un contexte de santé publique. Les études le faisant n'ont par ailleurs pas utilisé les deux méthodes d'extraction et n'ont ainsi pas pu effectuer de comparaison sur des données sanitaires. De plus, aucune étude (tous domaines confondus) n'a appliqué de vérification concernant les hypothèses associées à une

modélisation dans un cadre de l'EVT, l'adéquation à une distribution extrême et la comparaison avec d'autres distributions extrêmes. Enfin, l'EVT n'a jamais été considérée sur les pics de MCV. Chacun de ces aspects est traité dans la méthodologie développée ici.

Afin de sélectionner les pics sanitaires, la méthode des blocs associée à la distribution généralisée des valeurs extrêmes (GEV) et la méthode du seuil associée à la distribution généralisée de Pareto (GPD), méconnues dans le domaine de la santé, sont utilisées. Le principe de la méthode des blocs est le suivant: les données brutes sont séparées en plusieurs blocs, dans lesquels on extrait l'observation maximale. Toutes les observations maximales constituent la série de pics, qui peut être modélisée par une GEV :

$$G(y; \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \text{ pour } \xi \neq 0 \quad (3)$$

avec $\left\{ y : 1 + \xi \left(\frac{y - \mu}{\sigma} \right) > 0 \right\}$, $-\infty < \mu < \infty$ le paramètre de position, $\sigma > 0$ le paramètre d'échelle

et $-\infty < \xi < \infty$ le paramètre de forme. Il y a ainsi autant de pics sanitaires que de blocs. Fisher et Tippett (1928) suggèrent que la loi limite plus appropriée à un échantillon de pics extraits par blocs sera nécessairement une GEV, peu importe la distribution des données originales. La méthode du seuil repose quant à elle sur l'établissement d'un seuil élevé, relativement aux données brutes (par exemple, le quantile à 90 %). Une fois ce seuil fixé, toutes les observations supérieures sont considérées comme des pics sanitaires, qui peuvent alors être modélisés grâce à une GPD (Balkema et Dehaan, 1974), peu importe la distribution des données originales :

$$H(y; \sigma, \xi) = 1 - \left(1 + \frac{\xi y}{\sigma} \right)^{-1/\xi} \text{ pour } \xi \neq 0 \quad (4)$$

avec $\{y: y > 0 \text{ et } (1 + \xi y / \sigma) > 0\}$, $\sigma > 0$ le paramètre d'échelle et $-\infty < \xi < \infty$ le paramètre de forme.

Un bon choix du seuil et du nombre de blocs apporte un équilibre entre le biais et la variance des paramètres estimés (Reiss et Thomas, 2007). En effet, un faible nombre de blocs (ou un seuil trop élevé) entraînera une courte série de pics sanitaires tandis qu'un nombre de blocs trop élevé (ou un seuil trop bas) aboutira à une série longue, mais risquant de sélectionner des événements qui ne sont pas des pics sanitaires. Le premier cas donnera lieu à une forte variance, le deuxième à un biais important. En l'absence de méthode de détermination de blocs et de publications dans le domaine de la santé, les blocs explorés sont ceux de 7, 14, 21, 30, 60, 90, 120 et 180 jours. Concernant le seuil, il est possible de se baser sur les études de De Zea Bermudez et Mendes (2012) ou Watts *et al.* (2006), qui considèrent les quantiles élevés à 80, 85, 90 et 95 %. En se basant sur ces études, les quantiles explorés ici sont ceux à 75, 80, 85, 90, 92,5, 95, 97,5 et 99 %. Ces auteurs précisent tout de même la difficulté pratique de sélectionner un seuil approprié.

Pour respecter les hypothèses qui permettent une modélisation par la GEV ou la GPD (Coles, 2001), plusieurs sous-étapes sont nécessaires. En particulier, il est nécessaire de tester l'indépendance, l'homogénéité et la présence d'une tendance dans la série des pics sanitaires extraits. Ces conditions sont respectivement vérifiées grâce aux tests de Wald-Wolfowitz, Wilcoxon et Mann-Kendall. Ces tests sont aussi utilisés par Ouarda *et al.* (2006) pour vérifier les mêmes hypothèses dans un contexte hydrologique. De plus, plusieurs séparations des données brutes sont explorées afin d'assurer l'homogénéité des pics (fins de semaine et jours de semaine, deux saisons, quatre saisons) et pour la méthode du seuil, une procédure de déclusterisation (Smith, 1989) est appliquée. Cette procédure a pour but de s'assurer de l'indépendance des pics en ne sélectionnant pas des clusters entiers (p. ex., ceux qui surviennent lors des vagues de

chaleur). Dans le cas où une ou plusieurs hypothèses sont violées, il existe toutefois des méthodes qui permettent d'y pallier, voir par exemple le résumé de Khaliq *et al.* (2006). En revanche, ces extensions demandent habituellement plus de données et entraînent des interprétations moins directes.

Une fois que les pics sanitaires sont sélectionnés, les paramètres de (3) et (4) sont estimés. L'estimation se fait par le maximum de vraisemblance (*maximum likelihood*, ML) et par les L-moments (LMOM, Hosking et Wallis, 1987). Il convient ensuite de tester l'adéquation des distributions extrêmes (dont la GEV et de GPD) aux séries des pics sanitaires. L'évaluation se fait par les tests d'adéquation de Kolmogorov-Smirnoff et d'Anderson-Darling (Davison et Smith, 1990; Laio, 2004; Bezak et al., 2014). Enfin, pour choisir la distribution finale parmi celles jugées adéquates, le critère de la racine carrée de la moyenne des erreurs quadratiques est calculé (Hosking et Wallis, 1987; Gomes et Guillou, 2015). Enfin, le choix d'une distribution finale permet de calculer un niveau de retour associé, qui est le niveau que l'on s'attend à atteindre ou à dépasser en moyenne une fois au cours d'une période donnée (Coles, 2001). La méthodologie générale est résumée à la Figure I-4 du premier chapitre.

4.2. Étude des relations entre la météo et les pics sanitaires

Une fois que les pics sanitaires sont analysés et que leurs distributions extrêmes sont sélectionnées, il est possible de se pencher sur leur modélisation en fonction de variables explicatives (ici les variables météorologiques). Ainsi, le deuxième chapitre est dans la continuité directe du premier puisqu'il en reprend les pics sanitaires extraits pour les intégrer comme variables à expliquer dans des modèles de régression. Des GAM sont mis en œuvre afin d'étudier la relation entre les pics sanitaires et la météo. Toutefois, aucune étude ne s'est intéressée à la sélection de pics sanitaires, à plus forte raison à la régression de ces derniers. Utiliser comme

variable à expliquer des pics sanitaires entraîne plusieurs différences avec le cas classique (toutes les observations). On peut citer la taille réduite des échantillons, les distributions de la variable dépendante (et par conséquent du terme d'erreur) ou encore le nombre de variable explicatives. Ces problématiques sont traitées grâce à la méthodologie générale développée ici.

Les GAM présentent une grande flexibilité dans le cadre d'une modélisation non-paramétrique. Par ailleurs, cette flexibilité est un aspect attrayant pour cette étude puisqu'aucune connaissance *a priori* n'est disponible sur la forme des relations entre les pics sanitaires et la météo. Dans le cas des pics sanitaires, un GAM s'exprime sous la forme (Wood, 2006)

$$g(y_i^*) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_p(x_{pi}) + f_{p+1}(\text{temps}_i) \quad (5)$$

où les y_i^* sont les pics sanitaires (l'étoile indiquant le caractère extrême de la variable), g est la fonction de lien, les x_1, \dots, x_p sont les variables météorologiques et les f_1, \dots, f_{p+1} sont des splines cubiques de régression (Hastie et Tibshirani, 1990), avec $f_{p+1}(\text{temps}_i)$ qui permet de contrôler pour le temps. L'un des avantages est de disposer maintenant d'une influence continue, contrairement à une influence sous la forme d'un paramètre ponctuel. La contrepartie réside dans l'estimation du GAM, plus complexe que pour un modèle linéaire classique qui estime la solution par moindres carrés pondérés. En effet, la procédure d'estimation du GAM passe par des approximations itératives des moindres carrés pondérés pénalisés (P-IRLS, Wood, 2006), impliquant plus d'étapes de calcul et certains critères de convergence. Hastie et Tibshirani (1990) proposent quant à eux comme méthode d'estimation le *backfitting*, qui est une méthode plus générale puisqu'elle permet d'estimer n'importe quel modèle additif. Toutefois, la sélection du paramètre de lissage des fonctions f_1, \dots, f_{p+1} ne fait pas partie de l'estimation (rendant le *backfitting* plus subjectif contrairement aux P-IRLS qui l'estiment automatiquement) et peut

mener à des biais (Dominici *et al.*, 2002). Étant donné le désir d'automatisation de la méthodologie, les P-IRLS sont choisis ici.

Il est important de noter que dans l'équation (5), l'indice i indique les jours de survenue de pics, à la différence du cas classique où l'indice serait journalier et consécutif. Étant donné que les pics surviennent à plusieurs jours d'intervalle (dû entre autres à la procédure de déclusterisation), l'indice i ne représente donc pas des indices consécutifs. Par ailleurs, les séries des pics sanitaires possèdent des distributions particulières (GEV, GPD, lognormale, etc.). La variable réponse dans un GAM peut suivre une variété de distributions appartenant à la famille exponentielle (Hastie et Tibshirani, 1986), dont la GEV et la GPD ne font pas partie. Elles étaient toutes les deux cruciales pour l'établissement des niveaux de retour les plus précis possibles (voir le premier chapitre). La distribution lognormale est donc ici plus appropriée dans la mesure où elle représente une approximation satisfaisante parmi celles sélectionnées alternativement pour les pics sanitaires. Elle fait aussi partie des distributions les plus utilisées pour modéliser les séries des valeurs extrêmes, particulièrement en hydrologie (p. ex. Singh, 1998) et en finance (p. ex. Embrechts et Schmidli, 1994). C'est donc cette dernière qui sera utilisée dans les modèles ici.

Les splines et autres fonctions de lissage impliquent plus de paramètres à estimer qu'une régression linéaire, donc plus de données sont nécessaires. Le degré de complexité associé à une fonction $f(\cdot)$ est mesuré à l'aide du degré de liberté équivalent (edf). Afin d'éviter un éventuel biais, l'edf est choisi de façon optimale et automatique par validation croisée généralisée (GCV). Cette dernière est modifiée selon l'approche de Kim et Gu (2004) pour éviter les situations de surajustement, fréquentes en petits échantillons. D'ailleurs, cette solution a été utilisée par Rocklov et Forsberg (2008) pour étudier les effets de la température sur la mortalité de toutes causes.

Dans le présent projet, des lags allant jusqu'à 7 jours sont considérés pour les variables météorologiques, le lag optimal étant sélectionné par le critère d'information d'Akaike (AIC). De plus, les variables météorologiques sont moyennées les lags. Certaines études montrent que la moyenne de plusieurs jours avant l'évènement sanitaire par MCV joue un rôle significatif sur ce dernier, que ce soit une hospitalisation ou un décès (p. ex. Curriero *et al.*, 2002).

La significativité des fonctions est jugée à l'aide d'un test statistique. Il s'agit en réalité d'un test de nullité des paramètres des splines (Wood, 2006). Enfin, pour limiter les situations de multicollinéarité, les variables météorologiques maximales seulement sont utilisées dans les modèles de régression. En effet, Barnett *et al.* (2010) indiquent que les températures maximales ont le même pouvoir prédictif que les variantes minimales et moyennes. La méthodologie générale est résumée à la Figure II-1 du deuxième chapitre.

4.3. Étude des pics sanitaires et de la météo par la régression quantile

Les pics sanitaires ont été sélectionnés et intégrés en tant que variable à expliquer dans un modèle de régression dans les deux premiers chapitres respectivement. Plusieurs sous-étapes et vérifications ont été nécessaires, engendrant une quantité importante de résultats. Le but de ce chapitre est d'évaluer la QR pour l'étude des pics sanitaires. Cette dernière est effectivement adaptée car elle ne requiert pas autant de sous-étapes et permet une modélisation des pics sanitaires en fonction des variables météorologiques à partir de l'échantillon complet des observations. Son avantage principal repose dans sa capacité à décrire les relations dans un niveau de détails avancé puisqu'elle s'intéresse à un ou plusieurs quantiles de la distribution conditionnelle de la variable sanitaire. Cependant, aucune étude n'a utilisé cette capacité en l'appliquant sur plusieurs variables explicatives, sur plusieurs quantiles et avec plusieurs lags,

comme il est fait ici. La méthodologie développée permet ainsi de reconstituer l'entière des relations et de s'intéresser en particulier aux quantiles élevés, à savoir les pics sanitaires.

Pour un quantile $0 < q < 1$, la QR est définie par :

$$Q_{y_i} [q | x_{1i}, x_{2i}, \dots, x_{pi}] = \beta_0(q) + \beta_1(q)x_{1i} + \beta_2(q)x_{2i} + \dots + \beta_p(q)x_{pi} \quad (6)$$

avec Y la variable à expliquer et X_1, \dots, X_p les variables explicatives. Contrairement à la régression linéaire classique, les paramètres $\beta_0(q), \beta_1(q), \dots, \beta_p(q)$ dépendent du quantile dans lequel l'estimation a lieu.

L'interprétation des paramètres estimés de la QR est semblable à celle d'une régression linéaire ordinaire, c'est-à-dire qu'une augmentation d'une unité de X_1 augmente Y de $\beta_1(q)$ unités, toutes choses égales par ailleurs. Il y a cependant une différence fondamentale pour la QR : étant donné qu'une estimation est obtenue pour chaque quantile considéré, l'interprétation se fait donc dans chaque quantile séparément (contrairement à une estimation moyenne pour la régression linéaire ordinaire). Il est de plus important de noter que l'estimation des paramètres de la QR passe par des méthodes de calcul bien distinctes des moindres carrés ordinaires. Pour ces derniers, la somme des carrés résiduels est minimisée, tandis que Koenker et Bassett (1978) proposent de minimiser une somme des écarts absolus pondérés pour la QR.

La QR présente plusieurs avantages par rapport aux modèles du deuxième chapitre. D'abord, elle ne présume pas de distribution particulière quant aux erreurs de la régression (contrairement à la régression classique). Ensuite, elle permet de sauter l'étape de sélection des pics sanitaires (et les possibles erreurs associées) ou de vérification de loi théorique comme dans le premier chapitre. Enfin, c'est une régression robuste, dans le sens où elle n'est pas influencée par les valeurs aberrantes.

Dans ce chapitre, seules des régressions linéaires sont implémentées car le problème du croisement des quantiles s'accroît fortement dans un cadre non-linéaire (Yu *et al.*, 2003). Plusieurs méthodes sont proposées pour pallier cet inconvénient (Dette et Volgushev, 2008; Bondell *et al.*, 2010) mais aucun cadre théorique général n'existe encore, la décision est donc prise de ne pas explorer la QR non-paramétrique. De plus, il serait complexe de présenter une fonction non-linéaire pour chaque quantile exploré. Le fait de n'estimer qu'un paramètre par quantile facilite la présentation visuelle et l'interprétation.

Les lags considérés dans ce chapitre sont fixés à 0 (jour même), 3, 7 et 14 jours sur plusieurs quantiles. Les intervalles de confiance (IC) associés aux paramètres estimés de la QR, obtenus grâce à la méthode d'inversion des rangs (Koenker, 2005), sont aussi calculés.

4.4. Extrêmes météorologiques et influence sur la santé

Les trois premiers chapitres de la thèse s'intéressaient aux pics sanitaires et à leurs relations avec les variables météorologiques. Toutefois, il n'y a pas d'équivalence dans le sens où un pic sanitaire n'est pas nécessairement provoqué par un extrême météorologique et un extrême météorologique ne mène pas forcément à un pic sanitaire. Il est donc maintenant intéressant dans ce dernier chapitre de s'intéresser à la problématique inversée des trois premiers, c'est-à-dire les extrêmes météorologiques. Ils ne sont habituellement pas analysés exclusivement, que ce soit d'un point de vue des définitions ou de la modélisation. La plupart des études en santé publique qui s'y intéressent les considèrent uniquement comme cas particulier de l'ensemble des observations. Cette façon de faire ne permet pas d'apprécier pleinement le caractère extrême des événements météorologiques correspondants. Dans ce chapitre, certaines définitions sont évaluées afin de disposer d'extrêmes météorologiques justifiés d'un point de vue statistique, avec pour but de les intégrer comme variables explicatives dans un modèle de régression. Aucune méthodologie

n'existe pour le moment concernant cet aspect de régression. En fait, les études dont l'intérêt repose sur les valeurs extrêmes en plus d'une dimension se concentrent habituellement sur des méthodes de modélisation de la dépendance (p. ex. les copules).

Deux types d'extrêmes météorologiques sont étudiés dans ce dernier chapitre, les extrêmes univariés et les extrêmes multivariés. Les premiers sont sélectionnés de la même façon que les pics sanitaires, à savoir via la méthode des blocs et celle du seuil. Dans le premier chapitre, la distribution des pics sanitaires était nécessaire car ils représentaient les variables à expliquer. Hors, dans ce chapitre, les extrêmes météorologiques représentent les variables explicatives et il n'est donc pas nécessaire d'ajuster des distributions à ces derniers. La méthode des blocs et du seuil est appliquée sur les observations brutes, mais aussi sur les variations journalières des variables météorologiques. Certaines études montrent effectivement un effet plus important des variations plutôt que des valeurs absolues (Laschewski et Jendritzky, 2002; Houck *et al.*, 2005).

Les fonctions de profondeur (Zuo et Serfling, 2000) sont une solution envisagée pour définir les extrêmes multivariés. Il n'est pas aisé d'attribuer un « ordre » à des observations au-delà d'une dimension, ce à quoi les fonctions de profondeur permettent de répondre. Elles étendent directement la notion d'ordre à des données multivariées en leur assignant une valeur de « profondeur », tenant compte de leur valeur dans chaque dimension. En classant les observations multivariées, elles établissent donc un ordre. Le classement obtenu est dépendant de la distance utilisée. La distance de Mahalanobis est utilisée ici, car ses propriétés permettent d'établir naturellement un « point central » multivarié. Ce dernier possède la valeur de profondeur maximale de l'échantillon. Plus les observations s'en éloignent, plus leurs valeurs de profondeur diminuent. Ainsi, il est possible de définir des extrêmes météorologiques dépendant de plusieurs variables comme les observations les plus éloignées de ce point central. Les seuils considérés

pour définir les extrêmes dans ce cas sont les quantiles à 1, 5 et 10 % de la profondeur (les profondeurs les plus faibles correspondent aux points les plus éloignés du point central). Les fonctions de profondeur ne sont pour le moment pas utilisées en santé publique ni en météorologie, à plus forte raison en régression (à ne pas confondre avec la *depth regression*, qui est une toute autre technique non applicable ici, Rousseeuw et Hubert, 1999).

Les indices météorologiques constituent une autre façon de ramener le cadre multivarié à un cadre univarié, étant donné qu'ils sont construits à l'aide de plusieurs variables météorologiques (souvent la température et l'humidité). Ils considèrent les interactions (via les combinaisons) que les différentes variables intégrées peuvent avoir et leurs effets sur l'organisme. Il est alors possible d'attribuer une intensité numérique associée à chaque jour, ce qui permet de définir des extrêmes météorologiques. Les deux indices principaux utilisés dans les études de santé environnementale sont l'indice de chaleur et l'humidex. À nouveau, la méthode des blocs et du seuil sont utilisés pour extraire les valeurs extrêmes, une fois que les indices météorologiques sont calculés sur tout l'échantillon. De rares études incluent les indices météorologiques dans un modèle de régression (p. ex. Stieb *et al.*, 1996). Ici, le principal avantage de ces indices est de contourner l'aspect multivarié ainsi que ses complications.

5. PRINCIPAUX RÉSULTATS

L'application des méthodes présentées dans la section précédente a donné lieu à une quantité importante de résultats, décrits dans la deuxième partie de la thèse dans chacun des chapitres. Des résultats sélectionnés sont présentés dans cette section.

5.1. Études des pics sanitaires par la théorie des valeurs extrêmes

La méthodologie, se voulant complète, a généré une quantité importante de résultats (deux méthodes de sélection des valeurs extrêmes, deux villes étudiées, trois séparations des données, plusieurs seuils et blocs évalués). Un exemple de résultats ainsi obtenus est présenté au Tableau 0-4, pour la méthode du seuil appliqué aux hospitalisations à Montréal (l'ensemble des résultats est disponible dans le premier rapport). Puisqu'il s'agit de pics sanitaires, il est attendu que les tailles d'échantillon diminuent rapidement, jusqu'à 33 pics sanitaires (comparativement à la taille d'origine qui est de 4 077 observations). Il est possible de constater que toutes les estimations du paramètre de forme de la GPD sont négatives, impliquant des distributions extrêmes à queues légères et bornées.

Tableau 0-4 : Résultats de modélisation des pics d'hospitalisations à Montréal. Le seuil sélectionné est en gras. Pour chaque seuil, la taille d'échantillon (n) et les p-values pour les tests de MK, WW et WX sont données. Les paramètres de forme et d'échelle correspondants ($\hat{\xi}$ et $\hat{\sigma}$) sont renseignés, avec les p-values des tests d'adéquation pour les distributions GPD, EXP, LNO et GAM. Finalement, les critères RMSE sont renseignés.

Quantile (%)	GPD n	Tests d'hypothèses (p-value)			Paramètres estimés		Adéquation (p-value)		Distributions alternatives (p-value)			Comparaison (RMSE)			
		MK	WW	WX	$\hat{\xi}$	$\hat{\sigma}$	KS	AD	EXP	LNO	GAM	GPD	EXP	LNO	GAM
0.90	218	0.03	0.14	0.01	-0.3	13.55	0.21	0.31	0	0	0	1.14	172.11	2.62	2.73
0.925	187	0	0.08	0.03	-0.31	13.25	0.26	0.31	0	0.01	0	1.16	173.8	2.54	2.64
0.95	133	0.13	0.17	0.24	-0.37	13.14	0.45	0.35	0	0.03	0.02	1.48	177.04	2.49	2.57
0.975	77	0.31	0.57	0.04	-0.32	10.89	0.58	0.36	0	0.04	0.04	1.61	180.58	2.62	2.7
0.99	33	0.29	0.95	0.93	-0.23	8.89	0.45	0.33	0	0.05	0.05	1.78	181.85	2.91	3

Dans ce cas, les pics sanitaires supérieurs au seuil à 95 % sont sélectionnés, puis modélisés par une GPD. Le niveau de retour correspondant indique que 200 hospitalisations journalières seront atteintes ou dépassées une fois sur 1 an, tandis que sur 10 ans ce niveau monte à

214 hospitalisations. Ces niveaux sont à mettre en perspective avec le niveau moyen des hospitalisations de 131.

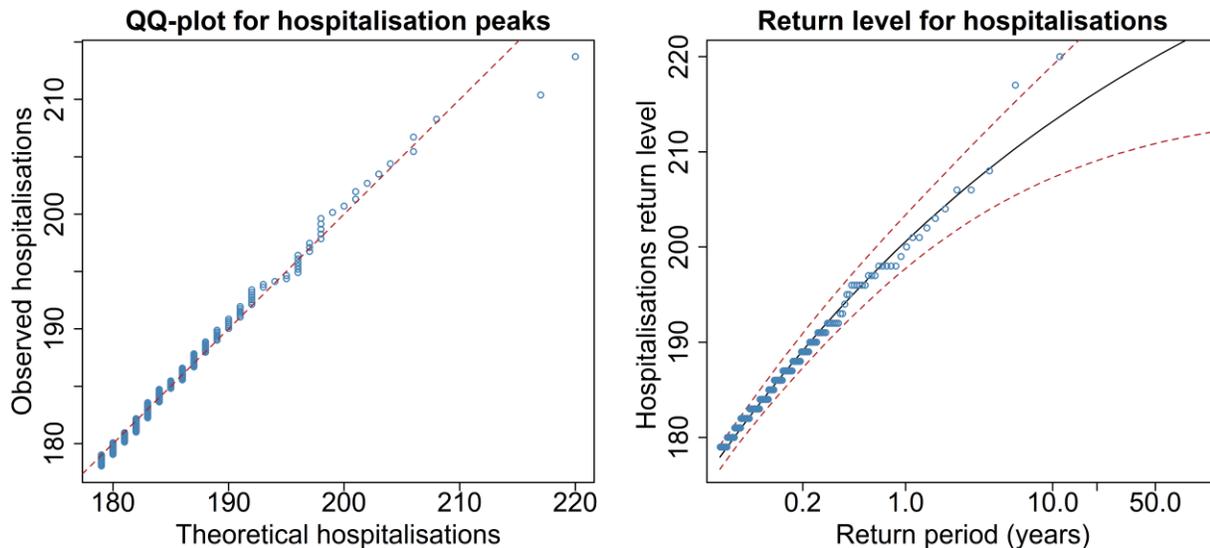


Figure 0-8 : QQ plot (gauche) et estimation des niveaux de retour (droite) pour les pics d'hospitalisations à Montréal. Les observations sont en bleu, la première bissectrice (QQ-plot) et les intervalles de confiance (niveaux de retour) sont en rouge.

Il s'agit de la première application de l'EVT en santé publique avec une méthodologie aussi complète. À plus forte raison, au meilleur de la connaissance des auteurs, c'est la première fois que l'EVT est appliquée aux MCV. Il reste possible de comparer les résultats obtenus avec les autres (rares) études qui s'intéressent aux pics sanitaires. Par exemple, les études qui appliquent l'EVT à l'espérance de vie (Aarssen et De Haan, 1994; Watts *et al.*, 2006) estiment aussi des paramètres de forme négatifs, impliquant une espérance de vie bornée. La littérature en santé publique n'a pas encore établi de règle sur le paramètre de forme, dont le signe semble dépendre de l'issue sanitaire étudiée. En effet, Chen *et al.* (2015) obtiennent une GPD avec un paramètre de forme positif pour les pics d'hospitalisations pour cause influenza, tandis que Thomas *et al.* (2016) obtiennent une GEV avec un paramètre de forme nul pour les pics de décès pour cause de pneumonie et d'influenza. Par ailleurs, De Zea Bermudez et Mendes (2012) modélisent plusieurs GPD et obtiennent les trois cas (majoritairement des paramètres de forme estimés négatifs).

Les niveaux de retour obtenus pourraient permettre une meilleure planification dans la gestion des hôpitaux ou pour les besoins des établissements de santé, en particulier en termes de personnel ou de capacité d'accueil. Par exemple, les résultats obtenus indiquent que sur 1 an, 200 hospitalisations journalières sont susceptibles d'être atteintes ou dépassées en moyenne une fois, sachant que le niveau moyen est de 131 hospitalisations. Ces niveaux sont d'autant plus pertinents en période de stress externe, comme les vagues de chaleur ou les épisodes de pollution. Ils peuvent aussi servir d'information complémentaire lors de la construction ou rénovation d'hôpitaux. Une analogie peut être faite avec les barrages hydrauliques, qui sont construits en fonction de certains niveaux de retour, afin d'être suffisamment résistants pour une période donnée. Sans prendre en compte cette information, le barrage en question pourrait s'avérer inefficace lors d'une crue qui n'a pas été intégrée à la conception.

5.2. Étude des relations entre la météo et les pics sanitaires

La Figure 0-9 illustre un exemple de résultats obtenus sur les pics de décès à Québec (l'ensemble des résultats est disponible dans le deuxième chapitre). Deux graphiques sont comparés pour montrer l'influence de la température sur les pics sanitaires et sur l'ensemble des hospitalisations. Le lag optimal sélectionné est ici 0 (soit le jour même). Ainsi, les pics de décès semble être influencés de façon légèrement plus importante par des températures au jour même autour de -10°C et 25°C , indiquant qu'une régression linéaire n'aurait pas été adaptée dans ce cas. Il s'avère que la majorité des relations dans le cas des pics sanitaires sont non-linéaires (en forme de « S » aplati ou de « J » inversé). Une zone de « moindre influence » a aussi été observée pour les pics sanitaires. Elle se situe autour de 10°C ($\pm 5^{\circ}\text{C}$) et a été associée à une influence moins marquée, que ce soit pour les hospitalisations ou les décès.

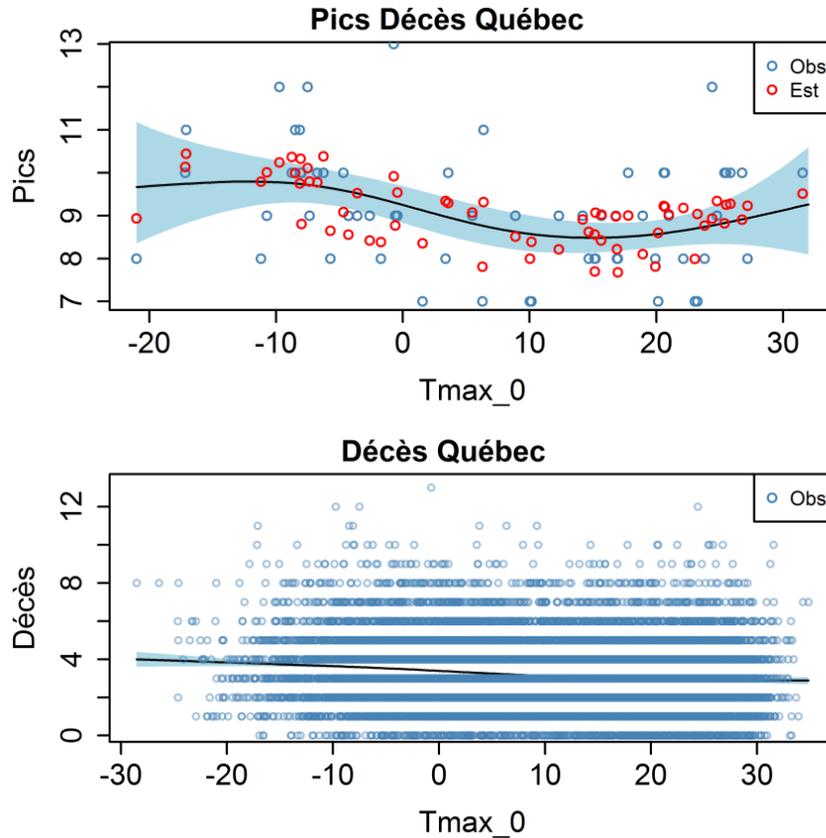


Figure 0-9 : Fonction de la température au jour même pour les pics de décès (haut) et pour l'ensemble des hospitalisations (bas) à Québec.

Le Tableau 0-5 résume les différents modèles significatifs obtenus au cours de l'étude. La température est présente dans la majorité des cas et possède une déviance expliquée forte. Toutes les déviances expliquées sont par ailleurs relativement importantes (comparativement aux modèles classiques, qui ont donné des déviances expliquées allant jusqu'à l'ordre de huit fois moins importantes). Cela peut s'expliquer par le fait que considérer toutes les observations peut « diluer » l'effet des variables météorologiques sur les variables sanitaires. En revanche, se focaliser directement sur le sous-échantillon des pics sanitaires met en lumière le fait qu'elles jouent un rôle bien plus important lorsque l'on se concentre sur les pics sanitaires. De plus, les formes des relations sont différentes des classiques « U » ou « V ». Un point commun entre les modèles des pics sanitaires et ceux de la littérature classique est l'effet majeur de la température.

En effet, dans les modèles obtenus, la variable de la température est quasiment omniprésente et possède un pouvoir explicatif important.

Tableau 0-5 : Résumé des différents modèles retenus pour expliquer les pics d'hospitalisations et de décès à Montréal et à Québec.

Ville	Variable sanitaire	Séparation	Modèle retenu des pics (variable _{lag} ; degré de liberté)	Déviante expliquée (%) Pics sanitaires / Cas classique
Montréal	Hospitalisations	Aucune	$f(T_{max_5}; 4,1)$	39,6 / 21,0
		Décès	Jours de semaine	$f(Prec_0; 1,6)$
	Fins de semaine		$f(H_{max_7}; 2,3)$	30,4 / 22,2
			$f(T_{max_2}; 3,0) + Prec_1$	59,8 / 27,4
	$f(T_{max_2}; 3,1)$	53,6 / 27,2		
Québec	Hospitalisations	Jours de semaine	$f(Prec_7; 1,0) + Snow_0$	38,7 / 27,6
		Fins de semaine	$f(T_{max_3}; 2,5)$	19,1 / 9,2
	Décès	Aucune	$f(T_{max_0}; 2,8)$	40,5 / 4,9
			$f(T_{max_1}; 2,2)$	35,8 / 5,0
			$f(Snow_4; 1,9)$	37,5 / 5,1

5.3. Étude des pics sanitaires et de la météo par la régression quantile

La Figure 0-10 montre des résultats de la QR obtenus pour les hospitalisations à Montréal (l'ensemble des résultats est disponible dans le troisième chapitre). La variable météorologique présentée est la température. Les résultats sont comparés à ceux estimés par une régression linéaire classique, afin d'illustrer les différences qui existent entre les deux méthodes. Les paramètres estimés des droites de régression sont tracés en fonction des différents quantiles. Des courbes de l'évolution des paramètres sont ainsi obtenues, indiquant l'intensité de l'influence des

variables météorologiques selon le quantile de la variable sanitaire dans lequel l'analyse des relations se fait. On a donc la possibilité de comparer les relations entre les pics sanitaires et la météo avec celles autour de la moyenne.

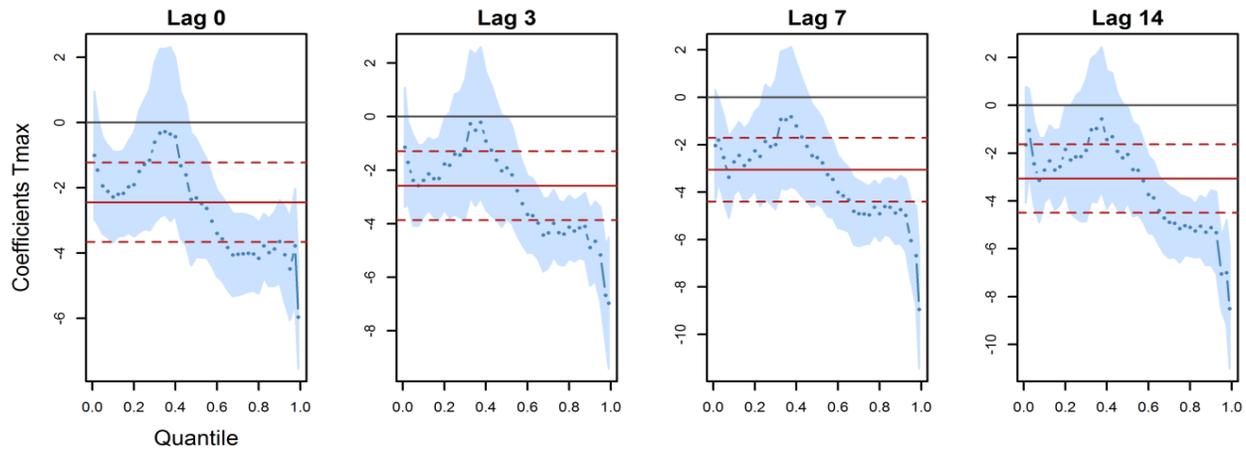


Figure 0-10 : Résultats sélectionnés de la QR et de la régression par moindres carrés ordinaires sur l'effet de la température sur les hospitalisations à Montréal.

● coefficients estimés par QR ; ■ IC de la QR ; — coefficients estimés par moindres carrés ordinaires ; - - - IC des moindres carrés ordinaires; — $y = 0$

Les résultats indiquent que les relations entre les variables sanitaires et les variables météorologiques sont très hétérogènes au long des quantiles conditionnels. Comparativement à un impact autour de la moyenne ou de la médiane, l'impact des températures est différent dans les pics sanitaires (tels que définis par les observations supérieures au quantile à 90 %). Les coefficients estimés sont négatifs et ont de plus tendance à s'accroître fortement dans les pics sanitaires. Les valeurs des coefficients y sont de deux à trois fois plus importantes, comparées à celles dans la médiane. Cette accentuation dans les quantiles élevés est présente pour les autres variables météorologiques, même si l'humidité et la pression atmosphérique se révèlent la plupart du temps non-significatives. Ces résultats appuieraient donc la nécessité de traiter les pics sanitaires à part car ces derniers entraînent des relations différentes avec les conditions

météorologiques, comparativement aux événements sanitaires moyens. Par ailleurs, les coefficients estimés de la température sont majoritairement négatifs, impliquant une diminution des hospitalisations et décès en fonction d'une augmentation des températures. Ce résultat pourrait s'expliquer par le fait que la plupart des pics sanitaires surviennent en hiver. Ainsi, une augmentation de la température correspondrait à un temps moins froid.

Comme mentionné dans la méthodologie, seules des modèles linéaires sont implémentés dans ce chapitre. Néanmoins, afin de pouvoir comparer avec les modèles du deuxième chapitre, un modèle non-linéaire de QR est estimé sur les hospitalisations à Montréal. Les pics sanitaires considérés dans les cas comparés ne sont pas exactement les mêmes (en raison entre autres de la déclusterisation avant l'application du GAM). Les résultats s'avèrent pourtant similaires en termes de forme d'effet, impliquant que les deux méthodes peuvent être utilisées de façon complémentaire sans contradiction majeure dans les résultats.

5.4. Extrêmes météorologiques et impact sur la santé

Étant donné que ce chapitre possède des problématiques similaires aux trois premiers et en ajoutent de nouvelles liées aux extrêmes multivariés, il est susceptible de générer une quantité importante de résultats. Afin de privilégier le développement méthodologique, la décision est prise de n'étudier que les hospitalisations à Montréal dans ce chapitre. Ce cas particulier permet tout de même de produire des résultats intéressants.

La Figure 0-11 illustre un exemple d'extrêmes météorologiques bivariés sélectionnés grâce aux fonctions de profondeur. Il s'agit des profondeurs calculées pour la température et l'humidité, pour les hospitalisations à Montréal. L'appariement se fait ensuite naturellement, en utilisant les indices (c'est-à-dire les jours) des profondeurs extrêmes. Une fois que les hospitalisations et les

variables météorologiques extrêmes sont sélectionnées, la régression indique l'impact qu'ont ces dernières.

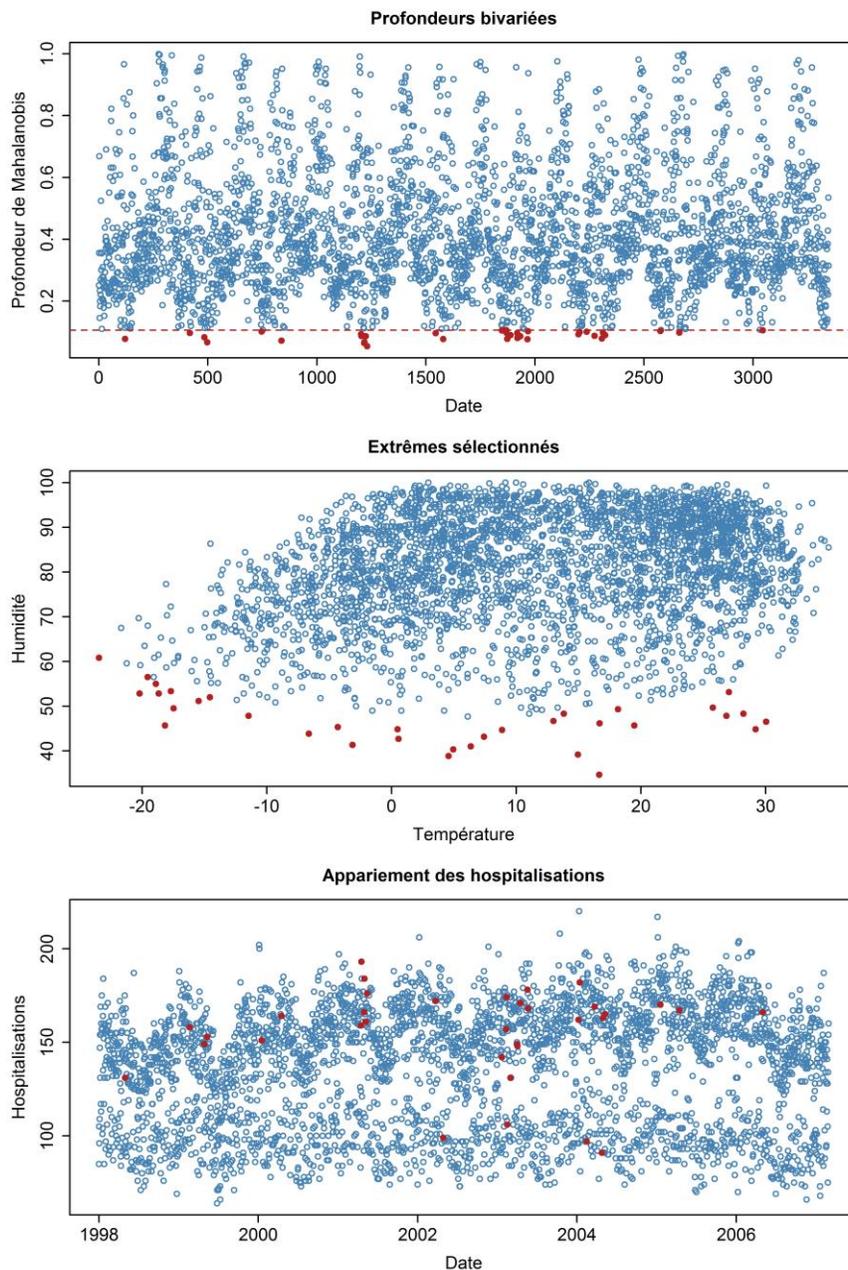


Figure 0-11 : Profondeur de la température et de l'humidité (les pointillés rouges représentent le seuil à 1 %), extrêmes bivariés et hospitalisations correspondants (en rouge) à Montréal.

Les extrêmes météorologiques n'ont pas révélé de correspondance systématique avec les pics sanitaires, c'est-à-dire qu'un extrême météorologique n'a pas nécessairement donné lieu à un

nombre d'hospitalisations élevé (jusqu'à une semaine après la survenue de l'extrême météorologique).

Les résultats montrent un taux de mortalité réduit après un extrême de température, ou *harvesting effect* (voir les résultats sélectionnés du Tableau 0-6). En effet, on constate une évolution différente des hospitalisations durant les sept jours caniculaires. Les signes des coefficients des droites de régression estimés s'inversent trois jours après la survenue de l'extrême de température. Un coefficient positif est d'abord observé, puis un coefficient négatif. Ce changement de signe pourrait s'expliquer par le fait que les personnes les plus à risque sont les premières hospitalisées après la survenue de l'extrême de température, entraînant une augmentation de la morbidité. Toutefois, des mesures préventives sont mises en place rapidement et on assiste par la suite à une diminution de la morbidité, donc à un coefficient de régression négatif. Cet effet est documenté dans d'autres villes (p. ex. Hajat *et al.*, 2005), mais pas dans un contexte de modélisation exclusive des extrêmes météorologiques.

Tableau 0-6 : Résultats retenus de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques univariés et multivariés (profondeur) extraits par seuil. Lorsqu'une droite de régression est estimée, le coefficient de régression est renseigné, tandis que dans le cas d'une fonction non-linéaire il s'agit du degré de liberté (ddl).

Extrême	Seuil (quantile) Lag (jours)	Déviance expliquée (%)	Test de significativité (p-value)	Coefficient de régression	n
Univarié : température	99 % 3	13,3	0,03	-12,31	34
	99 % 4	21,6	0,01	-15,83	
	99 % 5	13,1	0,03	13,32	
	99 % 6	24,4	0,01	18,51	
Multivarié : température et humidité	1 % 6	27,6	0,01	2,00 (ddl)	34
	1 % 7	27,2	0,02	2,32 (ddl)	

Il s'avère que globalement, les déviations expliquées sont relativement importantes (jusqu'à 27 % pour deux variables explicatives), considérant que les modèles sont linéaires. Une explication similaire au cas des pics sanitaires peut aider à comprendre ce résultat. En effet, dans la littérature classique, considérer toutes les observations a souvent pour effet de diluer l'effet des observations extrêmes. Dans ce cas, il s'agit des extrêmes météorologiques. Étant donné la rareté de ces derniers, il est attendu que leurs effets soient difficilement discernables dans les modèles utilisant l'ensemble des observations, d'où entre autres l'intérêt de développer des méthodes les concernant exclusivement. Ceci n'est habituellement pas fait alors que la méthodologie développée permettrait de mieux comprendre les conséquences des extrêmes météorologiques.

Les extrêmes météorologiques multivariés sont considérés pour la première fois, tant dans une optique de sélection que de régression. Plusieurs fonctions de profondeur ont été évaluées et les extrêmes météorologiques sélectionnés par la profondeur de Mahalanobis ont donné de meilleurs résultats dans les modèles de régression. De plus, l'inclusion des extrêmes météorologiques dans la régression a donné lieu à des influences majoritairement linéaires. Le pic sanitaire le plus élevé de toutes les observations est par ailleurs associé uniquement à un extrême météorologique sélectionné par les fonctions de profondeur, impliquant que ces dernières peuvent être particulièrement adaptées pour les pics sanitaires les plus importants. De plus, environ un extrême météorologique multivarié sur quatre (26 %) est associé à un pic sanitaire le jour même. Si cette proportion ne représente pas la majorité des extrêmes météorologiques, elle reste tout de même non-négligeable.

6. CONCLUSIONS ET PERSPECTIVES

Cette section présente les conclusions générales de l'ensemble des travaux réalisés durant la thèse, ainsi que leurs apports à la recherche et à la santé publique respectifs. Enfin, certaines avenues de recherche possibles clôturent cette synthèse.

6.1. Conclusions

Les conditions météorologiques ont un rôle non-négligeable sur la santé humaine. Dans un contexte de changements climatiques, il est important de comprendre les relations qui unissent les variables météorologiques aux variables sanitaires. En effet, les enjeux socioéconomiques derrière la compréhension de ces relations sont nombreux, ce qui se traduit par des études massives se concentrant sur l'ensemble des observations sanitaires et météorologiques.

Dans l'étude de ces relations, des valeurs extrêmes peuvent survenir. Ces extrêmes peuvent concerner les variables sanitaires (pics sanitaires) ou les variables météorologiques (extrêmes météorologiques). Dans les deux cas, ils sont problématiques du fait de leur rareté et de leur impact fort sur le système de la santé publique. Ils sont mal expliqués par les études classiques et ne peuvent être analysés adéquatement qu'à l'aide d'outils statistiques appropriés. Dans cette thèse, on propose une vision essentielle et complémentaire aux résultats habituels en santé publique, en se concentrant sur les valeurs extrêmes, très peu étudiées jusqu'à présent.

Le premier chapitre a pour objectif la sélection et la modélisation des pics sanitaires grâce à l'EVT. Ces étapes préliminaires à tout modèle de régression sont souvent omises dans les études de santé publique, qui se focalisent sur l'explication d'une variable sanitaire par des variables explicatives. L'EVT fournit des outils qui permettent de s'intéresser uniquement aux pics sanitaires. Une méthodologie rigoureuse et complète a permis de sélectionner des pics sanitaires

pour mieux les caractériser, en particulier grâce à leurs distributions extrêmes et leurs niveaux de retour. Ces derniers constituent un outil de gestion intéressant et pourraient permettre d'obtenir des alertes sanitaires plus précises. Cela aurait pour effet une meilleure préparation des hôpitaux en termes d'organisation du personnel durant les périodes où les pics sanitaires sont susceptibles de survenir, par exemple. L'application d'une méthodologie basée sur l'EVT permettrait aussi d'associer des facteurs externes aux pics sanitaires lors de leur survenue.

Le deuxième chapitre intègre les pics sanitaires dans des modèles de régression. Au meilleur de notre connaissance, aucune étude n'a effectué de sélection de valeurs extrêmes pour les insérer dans un contexte de régression. Ce travail intègre ainsi plusieurs problématiques statistiques, à savoir celles liées aux valeurs extrêmes, à la régression ainsi qu'au lissage. De plus, plusieurs différences avec les cas classiques (qui utilisent l'ensemble des observations disponibles) sont observées, comme la forme des influences ou le pouvoir explicatif plus élevé dans le cas des pics sanitaires. Dans notre cas, les pics sanitaires suivent des distributions extrêmes et les échantillons sont réduits par rapport aux échantillons d'origine. Ces faibles échantillons peuvent entraîner des surajustements. Toutefois, la méthode d'estimation couplée à une pondération de la GCV a permis d'éviter tant que possible ces situations problématiques. La méthodologie est ainsi automatisée et a donné principalement lieu à des relations non-linéaires. En termes d'application, l'explication des pics sanitaires par la météo permet de mieux comprendre les conditions météorologiques qui les engendrent. Ces modèles donnent par ailleurs la possibilité d'identifier précisément les conditions météorologiques liées aux pics de MCV. La température a entre autres montré un rôle majeur dans l'apparition des pics sanitaires. Ainsi, ces modèles pourraient aider dans la prévision et la gestion de certains pics sanitaires.

Le troisième chapitre a un but similaire au précédent, à savoir l'étude des pics sanitaires en fonction de la météo, mais à l'aide d'une approche statistique différente (la QR). Toutefois, alors que le deuxième chapitre se base sur les pics sanitaires extraits par l'EVT, la QR a été évaluée comme approche plus directe. Avec plusieurs sous-étapes en moins, la QR représente un gain de temps considérable. De plus, elle donne la possibilité d'évaluer non seulement les relations entre la météo et les pics sanitaires (les quantiles élevés de la distribution de la variable à expliquer), mais aussi avec tous les autres quantiles. La méthodologie développée apporte ainsi une vision complète de ces relations, ce qui a permis ici de découvrir des hétérogénéités importantes. Ainsi, l'effet des variables météorologiques est souvent largement plus marqué lorsque les pics sanitaires sont étudiés. Ces informations sont donc complémentaires à celles obtenues dans le deuxième chapitre, où l'effet précis des variables météorologiques a été analysé dans les pics sanitaires. Les résultats ont non seulement mis en valeur l'effet majeur de la température dans les pics sanitaires, mais ont en plus détaillé à quel point les relations variaient selon le quantile étudié. Ainsi, ils pourraient servir à ajuster le personnel déployé dans un établissement de santé. Par exemple, si l'impact estimé de la température sur les pics sanitaires est trois fois plus important que sur les quantiles médians, un hôpital pourrait déployer en conséquence trois fois plus de personnel durant les périodes susceptibles de voir survenir des pics sanitaires (canicule ou grand froid).

Le quatrième chapitre s'intéresse à la situation inverse des trois premiers chapitres : que se passe-t-il d'un point de vue sanitaire lorsque surviennent des extrêmes météorologiques? La première étape consiste à les définir de façon statistique. À nouveau, les études en santé publique ne s'intéressent habituellement pas à la sélection des extrêmes météorologiques et considèrent plutôt directement la modélisation par la régression. Il est cependant important d'établir les modèles à

l'aide des observations pertinentes (les extrêmes) afin de ne pas biaiser l'analyse, d'où l'importance de les définir dans une étape à part entière. Un cadre théorique et pratique manque pourtant à cet égard. Toutefois, les techniques développées dans ce chapitre ont permis de disposer d'extrêmes météorologiques définis selon des critères statistiques et d'étudier leurs impacts sur la santé par MCV. Si la survenue d'un extrême météorologique ne donne pas nécessairement lieu à un pic sanitaire, il existe tout de même une relation avec les MCV survenues après l'extrême météorologique. Ces relations s'avèrent pour la plupart linéaires, avec des déviations expliquées relativement élevées (allant jusqu'à 27 %), considérant les restrictions sur le nombre de variables explicatives, les régressions linéaires employées et le contexte de santé publique. Il faut aussi prendre en compte qu'il s'agit de la première intégration d'extrêmes univariés et multivariés dans des modèles de régression. Les résultats obtenus pourraient améliorer les programmes de surveillance, afin de mieux préparer la population à risque lors de la survenue des extrêmes météorologiques. Aussi, dans un contexte de changements climatiques, ils permettraient de mieux comprendre l'effet que les extrêmes météorologiques pourraient avoir dans le futur, sachant qu'ils seront probablement plus nombreux et plus intenses.

D'un point de vue statistique, cette thèse souligne les différences fondamentales entre les modèles avec des valeurs extrêmes et ceux avec l'ensemble des observations (voir le Tableau 0-7) et essaye d'y répondre de façon adéquate. En particulier, les modèles de régression implémentés ont dû être optimisés pour les petits échantillons (p. ex., l'estimation par P-IRLS et pondération de la GCV), en plus de tenir compte des distributions extrêmes. Dans ces modèles de régression, des pics sanitaires et des extrêmes météorologiques ont été inclus pour la première fois (comme variables dépendantes et variables explicatives, respectivement). Enfin, l'aspect

multivarié des extrêmes météorologiques abordé dans le dernier chapitre n'a jamais été traité en santé publique ou en météorologie.

Tableau 0-7 : Résumé des principales différences entre les études classiques et les études sur les pics sanitaires et les extrêmes météorologiques.

Caractéristique	Étude classique (contexte de régression)	Étude des pics sanitaires (contexte de régression)	Étude des extrêmes météorologiques (contexte de régression)
Échantillon	Grand nombre d'observations	Nombre d'observations restreint	Nombre d'observations restreint
Sélection des observations	Pas de sélection	Théorie des valeurs extrêmes	- Théorie des valeurs extrêmes - Indices météorologiques - Fonctions de profondeur
Distribution de la variable dépendante	- Distribution Normale - Distribution de Poisson	Distributions extrêmes	Distribution de Poisson
Littérature	Importante	Limitée	- Aucune sur la modélisation des variables sanitaires - Limitée sur la modélisation des paramètres de distribution
Déviante expliquée	Faible (<10 %)	Très élevée (de 20 à plus de 40 %)	Élevée (de 10 à plus de 25 %)
Forme des relations	- Majoritairement non-linéaires - Parfois linéaires	Majoritairement non-linéaires (mais formes différentes du cas classique)	Majoritairement linéaires

Dans une perspective plus globale, il est nécessaire que les autorités de la santé travaillent en partenariat avec les chercheurs de sorte à reconnaître les manques méthodologiques concernant les valeurs extrêmes. En effet, cette thèse montre entre autres que les séries des valeurs extrêmes (pics sanitaires et les extrêmes météorologiques) entraînent des relations très différentes du cas classique qui considère toutes les observations. Il est crucial de continuer à analyser les phénomènes communs dans une optique de gestion quotidienne pour les établissements de santé

publique. Néanmoins, ces différences motiveraient à compléter les études classiques qui portent sur tous les phénomènes par des études sur les valeurs extrêmes (pics sanitaires et/ou extrêmes météorologiques). Ce faisant, un portrait plus complet des relations qui unissent la météo à la santé pourrait être obtenu pour mener éventuellement à une meilleure adaptation de la population aux changements climatiques.

Enfin, il est important de noter que les méthodologies ont été développées de la manière la plus globale et objective possible, de sorte à ce qu'elles puissent être étendues à l'étude d'autres maladies chroniques ou de d'autres zones géographiques.

6.2. Perspectives de recherche

Ce projet ouvre différentes avenues de recherche pour de futurs travaux dans le domaine de la santé environnementale, que ce soit d'un point de vue pratique ou d'un point de vue méthodologique.

Des développements théoriques sont possibles concernant les valeurs extrêmes incluses dans la régression. Par exemple, les propriétés des intervalles de confiance (ou des tests de significativité) ne sont pas connues si la variable à expliquer est distribuée selon une GEV ou une GPD. Dans cette thèse, les distributions alternatives faisant partie de la famille exponentielle ont été utilisées, afin de permettre des modélisations par GLM ou GAM. Or, la GEV ou la GPD ne font pas partie de cette famille. Un début de réponse pour les intégrer pourrait être les modèles linéaires généralisés et additifs vectoriels (VGAM, Yee, 2015). Ces modèles ont pour principe la modélisation des paramètres d'une distribution. Ils acceptent un grand nombre de distributions pour la variable à expliquer (dont les distributions extrêmes). De plus, cette classe de modèles permet de prendre en compte les éventuelles tendances et saisonnalités présentes dans les données. Il faut toutefois prendre en compte que l'interprétation n'est plus aussi directe qu'avec

les modèles présentés dans cette thèse. Dans le cas des VGAM, elle se fait au travers des paramètres modélisés et permet donc une analyse de la distribution plutôt qu'une influence directe des variables explicatives.

La QR est un outil approprié pour l'étude des pics sanitaires puisqu'elle permet de se concentrer sur les relations dans différents quantiles de la variable sanitaire. Elle est flexible dans le sens où elle n'impose pas de distribution sur la variable dépendante et peut se combiner avec plusieurs autres techniques statistiques telles que les méthodes de lissage. Toutefois, son cadre théorique force les relations à être monotones d'un quantile à l'autre, sans quoi le croisement des quantiles est inévitable (il ne peut pas y avoir de coefficients positif dans un quantile et négatif dans le suivant p. ex.). Ce dernier cas de figure, s'il se produit, contredit les hypothèses de base du modèle concernant la définition des quantiles. Des travaux pourraient donc mêler le problème de croisement des quantiles dans un contexte de QR extrême (Beirlant *et al.*, 2004b).

Les extrêmes multivariés gagneraient à être étudiés plus en détail, que ce soit d'un point de vue théorique ou pratique. Les méthodes multivariées telles que les fonctions de profondeur ou l'EVT multivariée ne sont pas encore développées en santé publique. Pourtant, elles permettraient l'étude générale des phénomènes extrêmes, météorologiques ou non. Leur insertion en tant que variables explicatives dans la régression ne bénéficie d'ailleurs pas d'un cadre théorique à l'heure actuelle. Ainsi, les méthodes déployées sont appliquées ici au MCV mais pourraient se généraliser à des problématiques plus larges.

Une autre perspective intéressante serait l'évaluation spécifique des valeurs extrêmes dans l'étude du lien entre les conditions météorologiques et les personnes atteintes d'une ou plusieurs conditions de comorbidité. En effet, il est reconnu que certaines conditions comme l'hypertension, le diabète ou la maladie d'Alzheimer augmentent les risques de MCV (Arnett *et*

al., 2014). Dans une population vieillissante comme celle du Québec (Daigle, 2007), ces conditions de comorbidité pourraient devenir de plus en plus fréquentes et l'étude des pics sanitaires ou des extrêmes météorologiques pourrait alors s'avérer nécessaire pour de meilleures prévention des MCV et gestion des établissements de santé publique. De telles études permettraient de mettre en avant l'effet des conditions météorologiques sur une population fortement à risque et de distinguer le risque selon chaque condition de comorbidité. Elles nécessiteraient toutefois l'accès aux données individuelles (qui n'étaient pas disponibles dans la présente étude) et soulèveraient le défi technique des grosses bases de données.

**ARTICLES ET RAPPORTS DE
RECHERCHE**

I. MORTALITY AND MORBIDITY PEAKS MODELING: AN EXTREME VALUE THEORY APPROACH

Y. Chiu ^{a*}, F. Chebana ^a, B. Abdous ^b, D. Bélanger ^{a, c}, P. Gosselin ^{a, d}

^a *INRS-ETE, 490 rue de la Couronne, Québec (QC), G1K 9A9, Canada*

^b *Université Laval, Département de médecine sociale et préventive, 2325 Rue de l'Université, Québec (QC), G1K 7P4, Canada*

^c *Centre de recherche du centre hospitalier universitaire de Québec, 2875 boulevard Laurier, Québec (QC), G1V 2M2, Canada*

^d *Institut national de santé publique du Québec, 945 avenue Wolfe, Québec (QC), G1V 5B3, Canada*

* Corresponding author: yohann.chiu@ete.inrs.ca

Soumis le 6 avril 2016, accepté le 10 juillet 2016, publié le 1^{er} septembre 2016 dans *Statistical Methods in Medical Research*

Résumé

Les décès et hospitalisations figurent parmi les variables sanitaires les plus étudiées en santé publique. La distribution de ces variables est habituellement analysée au travers d'une modélisation de l'ensemble des données. Cette méthode est toutefois inadaptée pour appréhender les pics sanitaires, évènements « inhabituels » ayant de fortes répercussions sur le système de services et soins de santé. Cet article a pour but de modéliser les variables sanitaires non plus au travers de l'ensemble des données, mais par les pics, ce qui est très rarement fait. Pour cela, l'utilisation de la théorie des valeurs extrêmes (EVT) s'avère adéquate, avec une extraction des pics par la méthode des blocs ou du seuil élevé. La sélection et l'analyse des pics en tirant parti de l'EVT n'ont pas encore été considérées en santé publique, attribuant entre autres un caractère exploratoire à l'étude. La modélisation est précédée de plusieurs étapes de prétraitement des données. L'application est faite sur les pics d'hospitalisations et de décès des maladies cardiovasculaires, dans les communautés métropolitaines de Montréal et de Québec (Canada). Le développement d'une méthodologie rigoureuse et générale permet cependant l'application à d'autres issues sanitaires ou d'autres zones d'étude. Une des finalités des modèles obtenus est l'utilisation des niveaux de retour comme outil utile pour la gestion ou l'aménagement d'un hôpital existant, ou la conception d'un nouvel établissement de santé par exemple.

Abstract

Hospitalisations and deaths belong to the most studied health variables in public health. Those variables are usually analyzed through mean events and trends, based on the whole dataset. However, this approach is not appropriate to comprehend health outcome peaks which are unusual events that strongly impact the health care network (e.g. overflow in hospital emergency rooms). Peaks can also be of interest in etiological research, for instance when analyzing relationships with extreme exposures (meteorological conditions, air pollution, social stress, etc.). Therefore, this paper aims at modeling health variables exclusively through the peaks, which is rarely done except over short periods. Establishing a rigorous and general methodology to identify peaks is another goal of this study. To this end, the extreme value theory (EVT) appears adequate with statistical tools for selecting and modeling peaks. Selection and analysis for deaths and hospitalisations peaks using EVT has not been applied in public health yet. Therefore, this study also has an exploratory goal. A declustering procedure is applied to the raw data in order to meet EVT requirements. The application is done on hospitalisation and death peaks for cardiovascular diseases, in the Montreal and Quebec metropolitan communities (Canada) for the period 1981-2011. The peak return levels are obtained from the modeling and can be useful in hospital management or planning future capacity needs for health care facilities, for example. This paper focuses on one class of diseases in two cities, but the methodology can be applied to any other health peaks series anywhere, as it is data driven.

Keywords

Extreme Value Theory; Cardiovascular diseases; Return levels; Quebec; Public health;
Generalized extreme value; Generalized Pareto distribution

1 Background

Mortality and morbidity (hospitalisation) variables are usually analyzed through the whole dataset (e.g. Goldberg *et al.*, 2011 or Lim *et al.*, 2012). The corresponding methods focus on mean events. Therefore, these studies aim at understanding and predicting the mean features in order to be prepared for usual circumstances (e.g. weekly hospitalisations). This is an important step in public health surveillance and planning. There are nevertheless extreme events, known as “sanitary peaks” (or “peaks”), which demonstrate high values for the sanitary variable. Using all the data series leads to a focus on overall results which miss specific events, especially the extremes. The latter are far from the mean events and less frequent, they thus have to be dealt with proper statistical techniques. Among these techniques, the extreme value theory (EVT) is largely ignored in public health.

Because of their extreme nature, peaks cause significant stress on human and material public health services. Their modeling will eventually allow for more enlightened decisions, regarding peaks. For example, emergency services could be better prepared and hospital planning could also integrate needs related to such sanitary peaks. The underlying causes for peaks could also be different from the usual ones highlighted through analysis of the whole dataset using means. Hence, studying them can help clarify specific relationships with temperature, precipitations, air pollution, etc.

Cardiovascular diseases (CVDs) are a socioeconomic burden on the Quebec province as they account for a quarter of the total deaths count and represent the second cause of death (Girard, 2013). They are otherwise the first cause of hospitalisation and the disease group that costs the most to the Quebec health system (Daigle, 2007). All of these factors place CVDs on the list of

priority diseases to monitor. Furthermore, the EVT has never been applied to a CVDs context, to our knowledge.

This study thus focuses on sanitary peaks and aims at extracting and modeling them specifically. Since the EVT application on CVDs is new, this study is partly exploratory. Though EVT is well established in statistics, few studies in public health or medical sciences make use of it. Among these rare studies, De Zea Bermudez et Mendes (2012) were interested in cholesterol levels. Using the Generalized Pareto Distribution (GPD), they modelled these high levels in Portugal. Watts *et al.* (2006) were able to estimate the life expectancy upper limit of Canadian and Japanese populations, although not adjusted for compartmental or biological factors. Chen *et al.* (2015) predicted return levels for influenza in China. Finally, Guillou *et al.* (2014) examined *Salmonella* occurrences in France for use in surveillance. These four studies point out that EVT tools are suitable for health peaks study. However, none of them analyses CVDs nor uses the pretreatment steps required by some EVT hypotheses, nor do they use the Generalized Extreme Value (GEV) distribution as one of the important distributions in EVT.

One of the main outputs of EVT is the return period notion. It indicates the mean time before a random variable reaches or exceeds a certain level over a long period, thus specifying how “extreme” the event is. The return period is common in hydrology as a modeling tool to help dam design (Salvadori *et al.*, 2011) or in finance to evaluate profitability in time of financial shares (Longin, 2000). For example, in hydrology, a flood event associated to a return period of 100 years indicates that this event is expected to appear in average once every 100 years, over a long period. Its importance for design is then vital. In a symmetrical way, one can interpret the return level, closely related to the return period: it is the quantile that will be reached or exceeded once in average, over a long period. This piece of information might be useful for a hospital manager,

for example. The manager would like to know with more accuracy the bed capacity that should possess the hospital over a certain time period. Besides, peaks study could help public health surveillance to set up critical thresholds in protection interventions.

The purpose of the present paper is to explore EVT tools for extracting and modeling CVDs peaks, in a real-life environment. A data pretreatment is required in order to meet certain conditions before modeling peaks. Published studies often neglect these steps. This is not the case here as these steps are deemed essential for understanding the analysis, especially as peaks study is new in the realm of public health.

The rest of this paper is organized as follows. Section 2 describes data, followed by the developed and applied methods. The application is presented in section 3. Finally, discussion and conclusion are developed respectively in sections 4 and 5.

2 Data and methods

In this study, administrative health data are used. They are described in this section, along with statistical methods.

2.1 Data

The metropolitan communities of Montreal (hereinafter Montreal) and Quebec (hereinafter Quebec) were studied (Figure I-1). According to the 2011 census, 3 709 045 and 751 990 people lived in Montreal and Quebec respectively.

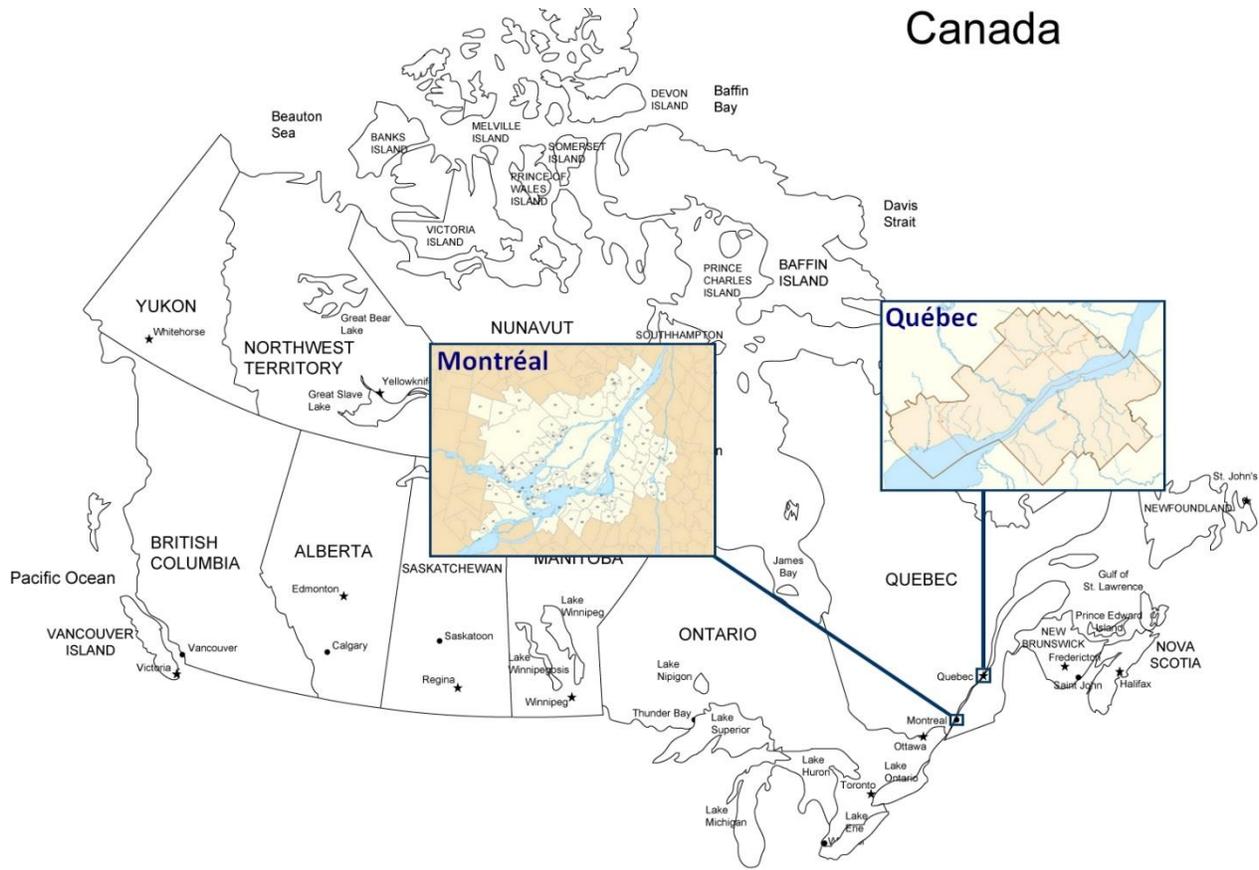


Figure I-1 : Montreal and Quebec Metropolitan communities.

Health outcome data consist of CVD hospitalisations in Montreal and CVD deaths in Quebec. Daily raw data were provided by the *Institut national de santé publique du Québec* [INSPQ, National Institute of Public Health of Quebec]. Versions 9 and 10 of the International Classification of Diseases were used to classify hospitalisation and death causes in order to select CVDs (Table I-1). Hospitalisations and deaths were combined in Montréal and Québec respectively so as to obtain daily numbers. The peak series are then extracted from these daily indicators.

Table I-1: ICD-9 and ICD-10 of considered diseases.

MOST DEADLY CVDs	ICD-9 ¹	ICD-10
Ischemic heart diseases	410-414	I20-I25
Heart failure	428	I50
Cerebrovascular diseases Transient cerebral ischemia	362.3	G45.x (excluding G45.4)
	430	H34.0
	431	H34.1
	434.x	I60.x
	435.x	I61.x
	436	I63.x (excluding I63.6)
		I64

Hospitalisation data supplied by the INSPQ ranged from January 1996 to March 2007, and from January 1981 to December 2011 for deaths. This leads to a total of $n = 4077$ days of hospitalisations in Montreal and $n = 11322$ days of deaths in Quebec. There were more deaths observations because they were spread over 31 years whereas the hospitalisations were spread over 11 years. Descriptive statistics are summarized in Table I-2. Due to the different sizes of the two metropolitan communities, the hospitalisations have a wider numerical scale than the one for deaths. Moreover, deaths are less frequent. The data are plotted in Figure I-2. An increasing trend can be seen for hospitalisations in Montreal before the year 2000 while it seems linearly decreasing for deaths in Quebec overall. Moreover, the hospitalisations in Montréal seem to be bimodal. Indeed, a further investigation reveals that hospitalisations during weekdays are mainly superior to those observed during weekends. It is a known effect regarding cardiovascular diseases (e.g. Gruska *et al.*, 2005; Mathes *et al.*, 2011). One possible explanation is the stress due to going back to work on Mondays. Thus, the bimodality may be considered as real. Moreover, there are more human and material resources in Québec than in Montréal *per capita*, which means that it would be easier to deal with the higher number of hospitalisations on Mondays in

¹ In the Quebec province, the transition from ICD-9 to ICD-10 occurred in April 2006 for the hospitalisation files and in January 2000 for the death files.

Québec. This in turn leads to data that are not necessarily bimodal in Québec as opposed to what is observed in Montréal, as illustrated in Figure I-2.

Table I-2: Descriptive statistics for hospitalisations and deaths, in Montreal and Quebec.

	Hospitalisations in Montréal	Deaths in Quebec
Minimum	49	0
Maximum	220	13
Mean	131	3
Median	136	3
Q75	158	4
Q90	172	6

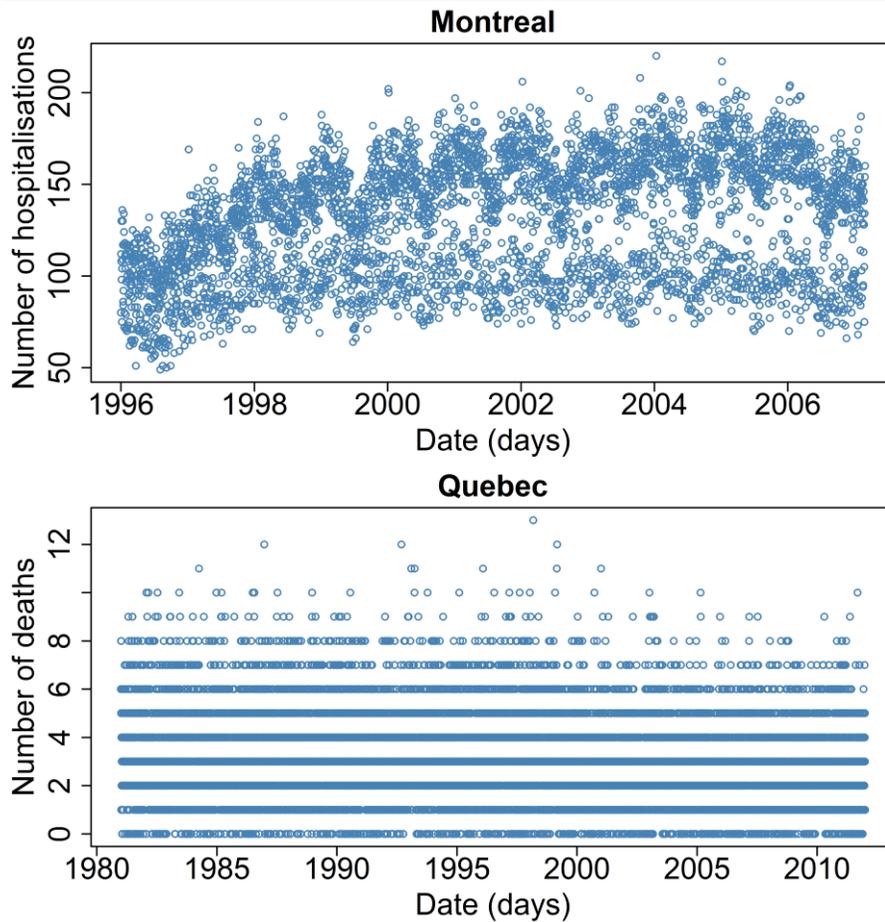


Figure I-2: Number of hospitalisations in Montreal (top) and deaths in Quebec (bottom).

2.2 Methods

The methodology applied here is based on tools from the EVT, introduced in this section. Two main methods are used to select peaks: the block and the peak-over-threshold methods. The first one extracts peaks from blocks (i.e., a vertical division of the data according to time intervals) while the second one extracts peaks greater than a given threshold (i.e., a horizontal division of the data according to their values). They respectively lead to peak series fitted by the generalized extreme value distribution and the generalized Pareto distribution (Coles, 2001). Once the extreme distributions have been fitted, statistical properties such as return levels are studied.

A summary is available in Figure I-4. For more details, the reader is referred to Coles (2001) or Reiss et Thomas (2007).

2.2.1 The generalized extreme value distribution

The generalized extreme value distribution (GEV, Fisher et Tippett, 1928) fits maxima extracted by the block method. In this paper, morbidity and mortality peaks are the maxima of an unknown distribution. Raw data are divided in blocks over time, in which the maximum observation is selected. The maxima series are then considered as the peaks to be analyzed and their limiting distribution is the GEV. Each block has the same size and the number of blocks determines the number of peaks.

Blocks may be defined as seasons, set to a certain size or set according to the needs of the concerned authorities. There is no global and objective methodology to select block size, in spite of their importance. Indeed, a too small number of blocks will give a small peak series whereas a too large one will give a long peak series with the risk of selecting non extreme events. The first case will lead to a high variance while the second case will lead to a high bias. McNeil et Frey

(2000) and Gilli (2006) pointed out that the block size is important because the GEV characteristics depend on it, but the practical choice is difficult.

Let X be the extracted peak series, the GEV probability distribution function is defined as:

$$G(x; \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \text{ for } \xi \neq 0 \quad (7)$$

for $\left\{ x : 1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0 \right\}$, $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter

and $-\infty < \xi < \infty$ is the shape parameter. In the special case $\xi = 0$, (7) reduces to

$G(x; \mu, \sigma, \xi) = \exp \left\{ - \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\}$. The Weibull, Frechet or Gumbel distribution is obtained

whether $\xi < 0$, $\xi > 0$ or $\xi = 0$ respectively. It should be noted that some authors favor the

notation $\xi^* = -\xi$ (e.g. El Adlouni *et al.*, 2009).

Whatever the original distribution may be, any series of maxima is “attracted” to a specific extreme value distribution. This is related to the “domain of attraction” notion. For example, maxima from a uniform or a Beta distribution are in the Weibull domain. A full characterization of the domains of attraction is given by Beirlant *et al.* (2004a) (chapter 2).

Note that some distributions are predominant depending on the study field. For instance, extreme wind speed is often fitted with a Weibull distribution (e.g. Ouarda *et al.*, 2015) while extreme precipitation is often fitted with a Gumbel distribution (e.g. Koutsoyiannis, 2004). Details about CVDs peaks are missing as there is currently no study involving EVT.

Once the appropriate distribution has been selected and the corresponding parameters estimated, the theoretical return level z_p , associated with a return period T (equivalently to a quantile order p where $T = 1/p$, Coles, 2001, chapter 3), is defined by:

$$\begin{aligned} z_p &= \mu - \frac{\sigma}{\xi} \left\{ 1 - \left[-\ln(1-p) \right]^{-\xi} \right\} \text{ for } \xi \neq 0 \\ &= \mu - \sigma \ln \left[-\ln(1-p) \right] \text{ for } \xi = 0. \end{aligned} \quad (8)$$

Confidence intervals can be computed for return levels using the delta method (Coles, 2001, chapter 3). One has to keep in mind that z_p is a mean occurrence, thus it may be reached more than once or not at all during the considered time period. Given that no study using the GEV in public health exists, the explored blocks are 7, 14, 21, 30, 60, 90, 120 and 180 days. The blocks are set to begin at the 1st of January, for each new year. Thus, the number of blocks is constant throughout the years and different cycles are kept, depending on the blocks length. This ensures that the number of blocks stays the same throughout the years, although this may leave some days out at the end of the year.

2.2.2 The generalized Pareto distribution

The generalized Pareto distribution (GPD) was introduced by Smith (1987). The GPD fits peaks extracted from the peak-over-threshold (POT) method. The latter consists in establishing a high threshold, relative to the raw data, where observations above this threshold are then considered as peaks.

Let X be the raw data variable, the exceedances of X over a threshold u are then expressed as $Y = X - u$. The GPD distribution function is defined by:

$$H(y; \sigma, \xi) = 1 - \left(1 + \frac{\xi y}{\sigma} \right)^{-1/\xi} \text{ for } \xi \neq 0 \quad (9)$$

on the interval $\{y: y > 0 \text{ and } (1 + \xi y / \sigma) > 0\}$ where $\sigma > 0$ is the scale parameter and $-\infty < \xi < \infty$ is the shape parameter. In the special case $\xi = 0$, (9) reduces to $H(y; \sigma, \xi) = 1 - \exp(-\frac{y}{\sigma})$. Depending on the sign of ξ , the GPD takes three different forms. $\xi > 0$ implies distribution with no upper limit (Pareto distribution), $\xi < 0$ implies an upper bounded distribution (Beta type distribution) and $\xi = 0$ is the unbounded exponential distribution (AghaKouchak *et al.*, 2012, chapter 2). This notation for the shape parameter is conventional in statistical literature, but other fields may express it as $\xi^* = -\xi$ (for instance in hydrology Katz *et al.*, 2002).

An appropriate threshold u balances between bias and variance for the estimated parameters (Reiss et Thomas, 2007, chapter 5). A too high threshold will select too few peaks whereas a too low threshold will select events that may not be peaks, leading to a high variance and bias respectively. Several thresholds can be suggested using statistical tools (e.g. Lang *et al.*, 1999), although there is no objective method to definitely select the most appropriate one. To practical use, De Zea Bermudez et Mendes (2012) recommends using the 90 % quantile as a standard threshold. As for Watts *et al.* (2006), they deal with the 80, 85, 90 et 95 % quantiles. The explored quantiles here are the 75, 80, 85, 90, 92.5, 95, 97.5 and 99 % quantiles.

The theoretical GPD return level z_p is expressed by inverting equation (9) :

$$\begin{aligned} z_p &= \frac{\sigma}{\xi} \left\{ (1-p)^{-\xi} - 1 \right\} \text{ for } \xi \neq 0 \\ &= -\sigma \ln(1-p) \quad \text{for } \xi = 0. \end{aligned} \tag{10}$$

When plotting return levels against time, the time abscissa is usually indicated on a logarithmic scale. Confidence intervals can be computed for return levels using the delta method (Coles, 2001, chapters 3 and 4).

Both the block and POT methods are illustrated in Figure I-3 for hospitalisations, in Montreal.

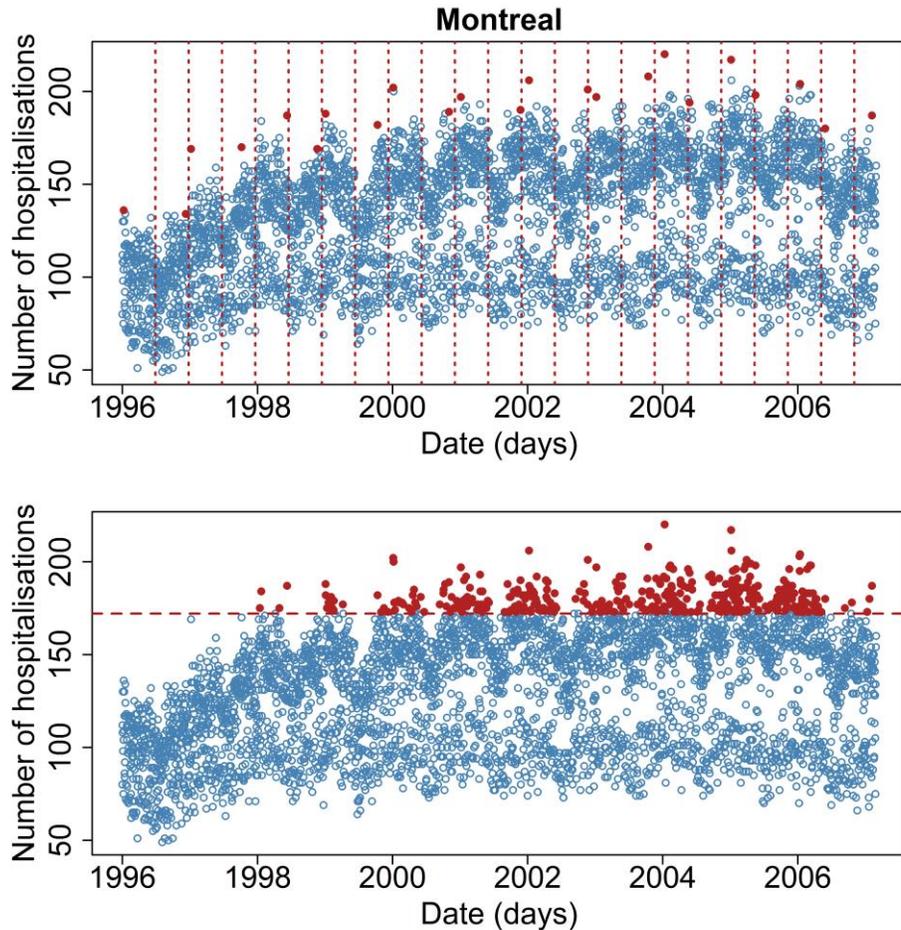


Figure I-3: Block (up) and POT (bottom) methods illustration on hospitalisations, in Montreal. The block method uses a block size of 180 days while the POT method uses the 90 % quantile. Peaks are in red.

2.2.3 Data pretreatment and statistical tests

In the classical approach of EVT, independence and identical distribution assumptions (i.i.d.) are required of the peak series before modeling with an extreme distribution (Coles, 2001, chapters 3 and 4). Equivalently, peak series stationarity, independence and homogeneity have to be tested.

To this end, these statistical tests are used:

- Mann-Kendall (MK) for H_0 : data have no trend,
- Wald-Wolfowitz (WW) for H_0 : data are independent,
- Wilcoxon (WX) for H_0 : data are homogeneous.

These are also employed, for instance, by Ouarda *et al.* (2006) to confirm the same hypotheses, in a hydrological context. Refinements of the classical approach exist in order to overcome any of these hypotheses violation (Coles, 2001, chapter 6). However, they usually require covariates, which would increase the complexity of the models in terms of computation, interpretation and choices to make. Moreover, the “exploratory” goal of this study means that, in the first place, the models have to be well-understood and adapted to the issue of peaks.

Peaks tend to occur in clusters especially in the POT method. This can compromise the independence hypothesis. This situation usually does not occur with the block method except for very short blocks such as daily ones. However, daily blocks would be in contradiction with the study’s object. Declustering the peak series is a common practice (Beirlant *et al.*, 2004a, chapter 10) in order to ensure peaks independence. It is based on a fixed parameter r which establishes the minimal number of observations separating two clusters. If the number of observations between two peaks is lower than r , then these peaks are considered as belonging to the same cluster. In this case, only the maximum over the two observations is selected in the cluster. On the contrary, if they are separated by a number of observations greater than r then they belong to two different clusters and are both selected. In this study, the parameter is chosen as $r = 3$. This assures independence in most of the main CVDs events by taking into account heatwaves, cold spells and holidays at the same time. Results for $r = 1, 2, 4, 5$ (not presented) showed no significant differences.

Following the GEV and GPD parameter estimations, the peak series adequacy to extreme distributions are tested with the Anderson-Darling (AD) and Kolmogorov-Smirnov (KS) tests. The null hypothesis H_0 is: the peaks series follow a GPD (or GEV) with associated estimated parameters. These tests are based on the distances between empirical and theoretical distributions. They are among the most used statistical tests in the context of extremes, e.g. Stephens (1977) or Laio (2004). Significant results for the p-values are set at the $\alpha = 5\%$ level.

The GEV and GPD are asymptotical distributions, therefore the peak series adequacy to three other extreme distributions is also considered: exponential distribution (EXP), lognormal distribution (LNO) and gamma distribution (GAM). These three distributions are often used in hydrology and finance, e.g. Engeland *et al.* (2004).

To compare the different fitted distributions, the root mean squared errors are computed (RMSE

criterion, Gomes et Guillou, 2015) :
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}$$
, where \hat{y} are the estimated peaks

and y are the observed peaks.

For a block size or threshold to be considered as valid in this study, the associated peak series must not reject any statistical tests: MK, WW and WX to check the i.i.d. assumptions and KS and AD to check the distributional fitting (Figure I-4). Then, among the block sizes and thresholds deemed valid, the one which possesses the smallest RMSE is finally selected for modeling the peak series.

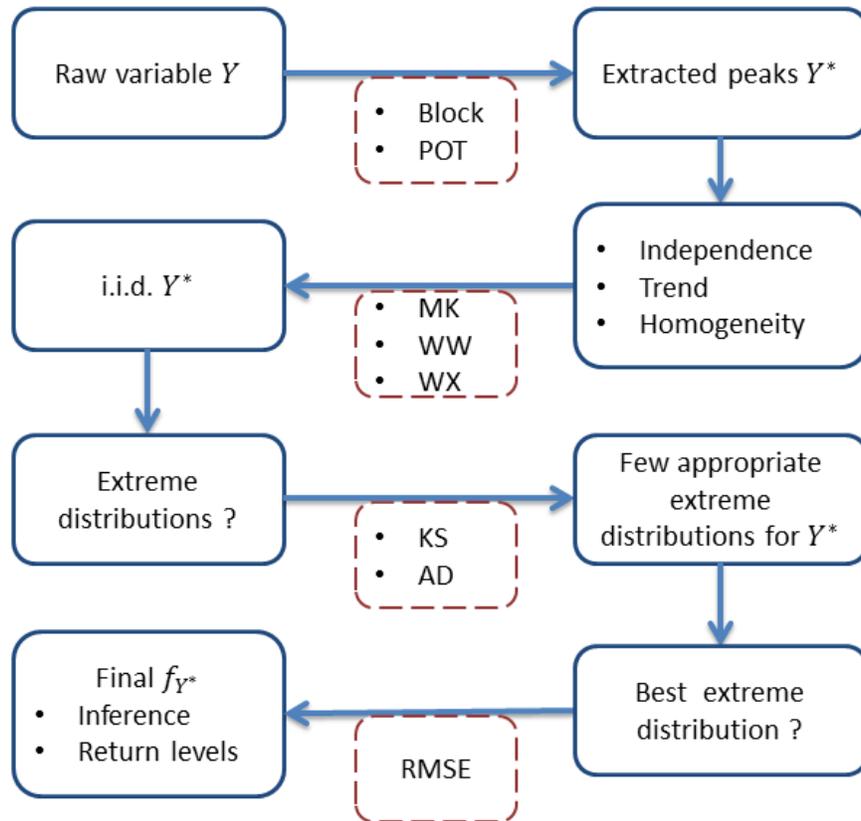


Figure I-4: Diagram illustrating the different study steps. WW, MK and WX: Wald-Wolfowitz, Mann-Kendall and Wilcoxon tests. KS and AD: Kolmogorov-Smirnoff and Anderson-Darling tests. RMSE: root mean squared errors.

All statistical analyses were carried out using the free software R (The Comprehensive R Archive Network: <http://cran.r-project.org>).

3 Application

Results are described here for CVD hospitalisations in Montreal and for CVD deaths in Quebec. Since this is the first application of EVT in a CVDs context, there is an exploratory concern and applications are chosen so as to diversify results (two metropolitan communities and two health variables). Hence, the POT method is applied to hospitalisations in Montreal whereas the block one is applied to deaths in Quebec. The full report (Chiu *et al.*, 2015) presents extensive results and conclusions, including the two cases treated in this paper.

3.1 Hospitalisation peaks in Montreal

The hospitalisation peaks in Montreal are extracted with the POT method with thresholds fixed as indicated in section 2.2.2. Table I-3 shows that below the 92.5 % threshold, some MK and WW tests (trend and independence) reject the null hypothesis for the corresponding extracted peak series. This could indicate that the threshold is not high enough. The trend before the year 2000 could be retained when the threshold is too low, which would lead to the rejection of the hypotheses.

Table I-3: Results for the hospitalisation peaks fitting in Montreal. The selected threshold is in bold characters. For each threshold, the sample size (n) and the p-values for the MK, WW and WX tests are given. The corresponding estimated shape and scale parameters ($\hat{\xi}$ and $\hat{\sigma}$) are shown, along with the p-values of the adequacy tests for the GPD, EXP, LNO and GAM distributions. Finally, the RMSE are shown.

MONTREAL															
GPD		Hypothesis (p-value)			Estimated parameters		Adequacy (p-value)		Alternative distributions (p-value)			Comparison (RMSE)			
Threshold (%)	n	MK	WW	WX	$\hat{\xi}$	$\hat{\sigma}$	KS	AD	EXP	LNO	GAM	GPD	EXP	LNO	GAM
0.75	269	0	0	0.25	-0.96	36.12	0.02	0.08	0	0.12	0.1	2.95	164.09	1.86	2.01
0.80	280	0	0	0.04	-0.82	29.04	0.05	0.11	0	0.01	0.01	2.56	166.49	2.03	2.16
0.85	257	0	0	0.04	-0.57	20.34	0.33	0.24	0	0.01	0.01	1.9	169.2	2.37	2.48
0.90	218	0.03	0.14	0.01	-0.3	13.55	0.21	0.31	0	0	0	1.14	172.11	2.62	2.73
0.925	187	0	0.08	0.03	-0.31	13.25	0.26	0.31	0	0.01	0	1.16	173.8	2.54	2.64
0.95	133	0.13	0.17	0.24	-0.37	13.14	0.45	0.35	0	0.03	0.02	1.48	177.04	2.49	2.57
0.975	77	0.31	0.57	0.04	-0.32	10.89	0.58	0.36	0	0.04	0.04	1.61	180.58	2.62	2.7
0.99	33	0.29	0.95	0.93	-0.23	8.89	0.45	0.33	0	0.05	0.05	1.78	181.85	2.91	3

For a threshold to be considered valid, it must not be rejected by any statistical test (section 2.2.3). Hence, three thresholds meet this condition: 95, 97.5 and 99 %. The sample sizes are respectively 133, 77 and 33 days with hospitalisation peaks. The threshold with the smallest RMSE is not necessarily valid if it is rejected by one of the above tests. For example, in this case, the one which minimizes the RMSE is the 90 % quantile (RMSE = 1.14) but the MK and WX tests reject the null hypothesis. Among the above three thresholds, the 95 % one minimizes the

RMSE (1.48) and is therefore chosen for the rest of the analysis, with the associated GPD. The adequacy tests do not reject the null hypothesis for the GPD, although the lognormal, gamma and exponential distributions are rejected. The estimated parameters for the GPD are $\hat{\xi} = -0.37$ and $\hat{\sigma} = 13.14$.

The GPD shape parameter $\hat{\xi}$ importance in EVT motivates further analysis. Its estimations are negative, spread out between -0.96 and -0.23 for the different thresholds. In accordance with the GPD theoretical properties (section 2.2.2), the corresponding distributions have an upper bound. Moreover, the shape parameter is increasing with the threshold whereas the scale parameter ($\hat{\sigma}$) is decreasing, implying that the numerical scale is lower in the highest quantiles. This could be explained by the fact that as the threshold is raised, only the more extreme peaks are selected, thereby reducing the numerical dispersion. This phenomenon is illustrated in Figure I-5.

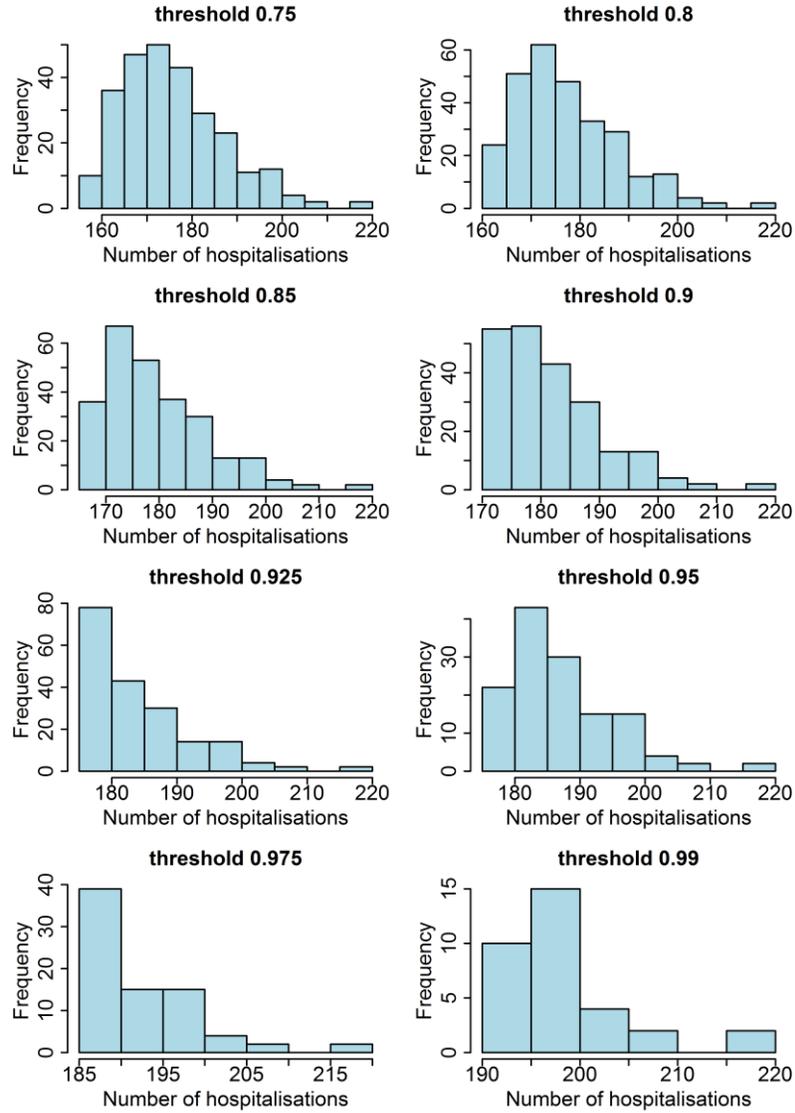


Figure I-5: Hospitalisations peaks histograms in Montreal, depending on various thresholds expressed in quantiles.

The peak series distribution changes with the increasing threshold, while the lower bound is raised as a consequence (Figure I-5). The selected peaks (95 % threshold equivalent to 133 peaks) are shown in Figure I-6. The peak selection begins after the 21 of January 1998. This implies that if a trend is present in the raw data, it will not be kept in the peak series, given that the threshold is fixed high enough. The quantile-quantile plot (Figure I-7) shows that most of the points lie on a $y = x$ line, indicating an adequacy between the GPD empirical and theoretical

quantiles. The two largest observations 217 and 220 are plotted away from the line and the model seems to underestimate them as 206 and 209 respectively.

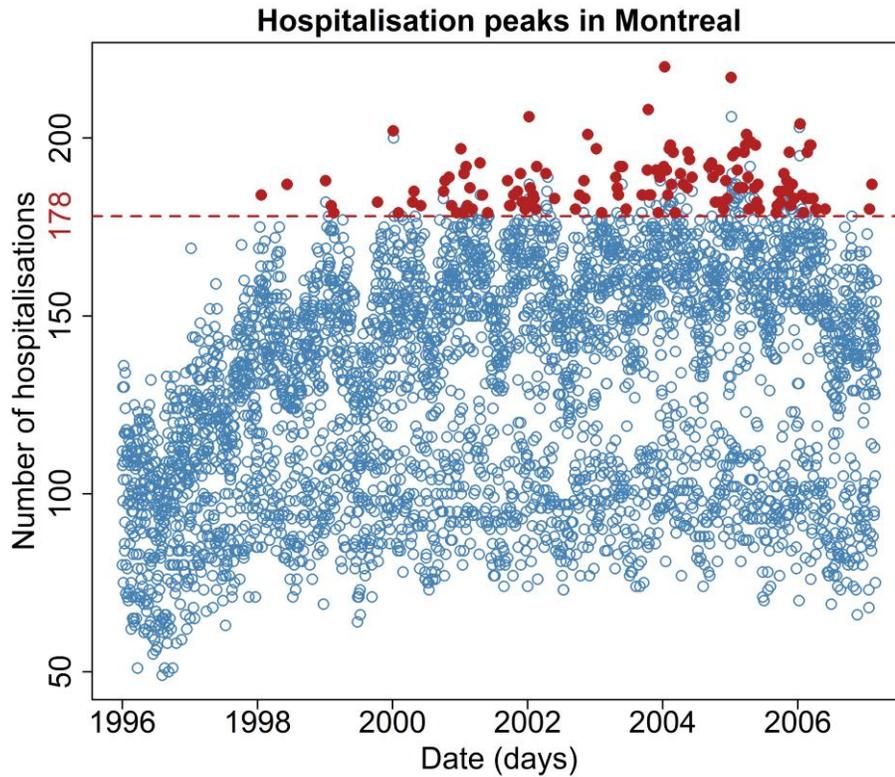


Figure I-6: Number of hospitalisations (blue) and selected peaks (red) in Montreal. The red dashed line is the selected threshold (95 % quantile).

The return level for the peaks corresponding to the selected threshold is shown in Figure I-7. Over a long period, it indicates that 200 hospitalisations per day are expected to be exceeded once in 1 year, while 214 hospitalisations per day are expected to be exceeded once in 10 years (the mean number of hospitalisations is 131). This latter level is a 7 % increase compared to the 1 year return level. This is significant since it probably indicates a 7 % increase in the number of patients at the emergency room, for instance. As with any prediction method, the uncertainty grows with time, conveyed by the rapidly growing confidence intervals. This point is essential when considering return levels for hospital management, emphasising short and medium term

decisions for example. Indeed, over 10 years, the confidence interval is spread out between 205 and 220 hospitalisations. This could imply different costs and organizational consequences.

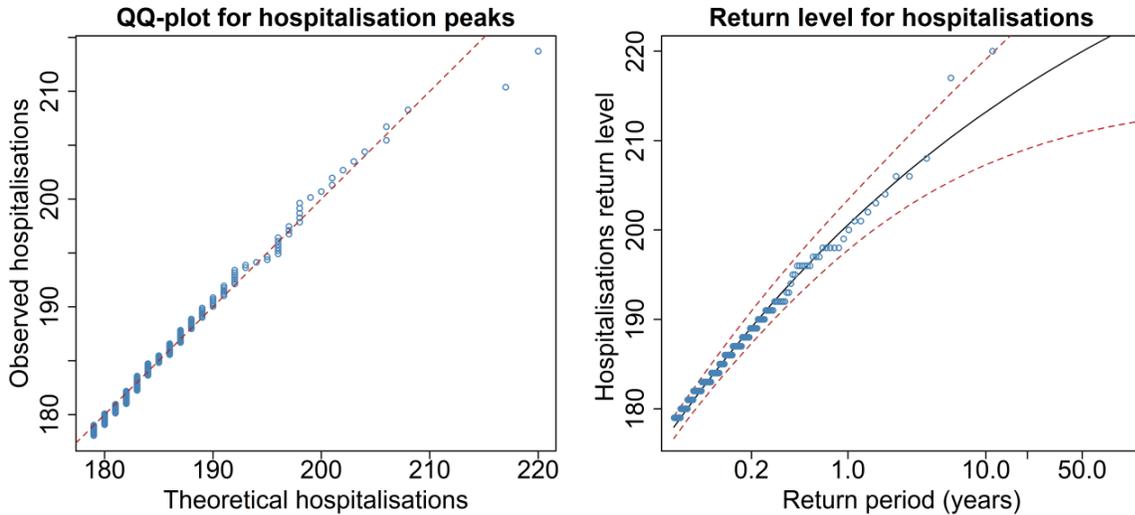


Figure I-7: QQ plot for hospitalisations peaks in Montreal (left). Observations are in blue while the first bisector is in red. Return level for hospitalisations à Montreal (right). Confidence intervals are in red.

3.2 Death peaks in Quebec

The death peaks in Quebec (Table I-4) are extracted from the raw data by the block method. Selecting an adequate block size for the deaths in Québec is more complex than the threshold selection for the hospitalisations in Montreal. Indeed, Table I-4 indicates that the MK test systematically rejects the stationarity hypothesis. This could be explained by the fact that the raw data is linearly decreasing (Figure I-2). The block method divides and selects peaks over the whole dataset, unlike the POT method. The trend is consequently kept in the peak series, leading to the MK test rejection of the null hypothesis. Besides, judging by the WW and WX tests results, no block size seems to be suitable because both reject the null hypothesis in most cases. The only exception is the block size of 180 days. A linear regression of the corresponding peak series over time results in an estimated slope of $-1.23e^{-4}$. Since this value is close to 0, the 180 days block size is finally chosen for the rest of the analysis, with the corresponding GEV distribution. The

adequacy tests do not reject the null hypothesis for the GEV, lognormal and gamma distributions whereas the adequacy for the exponential distribution is rejected. The associated estimated parameters for the GEV are $\hat{\xi} = -0.14$, $\hat{\mu} = 8.57$ and $\hat{\sigma} = 1.23$.

Table I-4: Results for the death peaks fitting in Quebec. The selected block size is in bold characters. For each block size, the sample size (n) and the p-values for the MK, WW and WX tests are given. The corresponding estimated shape, location and scale parameters ($\hat{\xi}$, $\hat{\mu}$ and $\hat{\sigma}$) are shown, along with the p-values of the adequacy tests for the GEV, EXP, LNO and GAM distributions. Finally, the RMSE are shown.

QUEBEC																
GEV		Hypothesis (p-value)			Estimated parameters			Adequacy (p-value)		Alternative distributions (p-value)			Comparison (RMSE)			
Block	n	MK	WW	WX	$\hat{\xi}$	$\hat{\mu}$	$\hat{\sigma}$	KS	AD	EXP	LNO	GAM	GEV	EXP	LNO	GAM
7	1618	0	0	0	-0.1	5.13	1.35	0	0	0	0	0	0.29	4.33	0.31	0.29
14	809	0	0	0	-0.12	5.94	1.33	0	0	0	0	0	0.29	5.14	0.3	0.3
21	540	0	0	0.62	-0.12	6.41	1.29	0	0	0	0	0	0.3	5.62	0.3	0.3
30	378	0	0	0	-0.15	6.74	1.33	0	0	0	0	0	0.3	5.92	0.3	0.3
60	189	0	0	0.74	-0.14	7.41	1.31	0	0.01	0	0	0	0.3	6.56	0.29	0.3
90	126	0	0.12	0	-0.07	7.81	1.16	0.01	0.02	0	0	0	0.29	6.98	0.31	0.32
120	95	0	0.15	0.02	-0.14	8.07	1.31	0.03	0.08	0	0.03	0.02	0.29	7.14	0.29	0.3
180	63	0.01	0.07	0.64	-0.14	8.57	1.23	0.12	0.13	0	0.15	0.15	0.31	7.6	0.31	0.32

The GEV shape parameter is estimated between -0.10 and -0.14, equivalent to Weibull distributions (section 2.2.1). It possesses a theoretical upper bound, which would be closer to reality than a death level indefinitely increasing. As for the GEV location parameter, it is increasing with the block size: the bigger the block, the lesser the probability of selecting the “small” peaks (in numerical terms). This shifts the distribution toward the right and as a consequence increases the location parameter. Lastly, the scale parameter is decreasing with the size blocks. The main reason could be that when choosing larger block sizes, the temporal window for selecting the peaks is also raised and possibilities for selecting “small” peaks are decreased. This contributes to the numerical dispersion reduction in favor of larger peaks, as illustrated in Figure I-8.

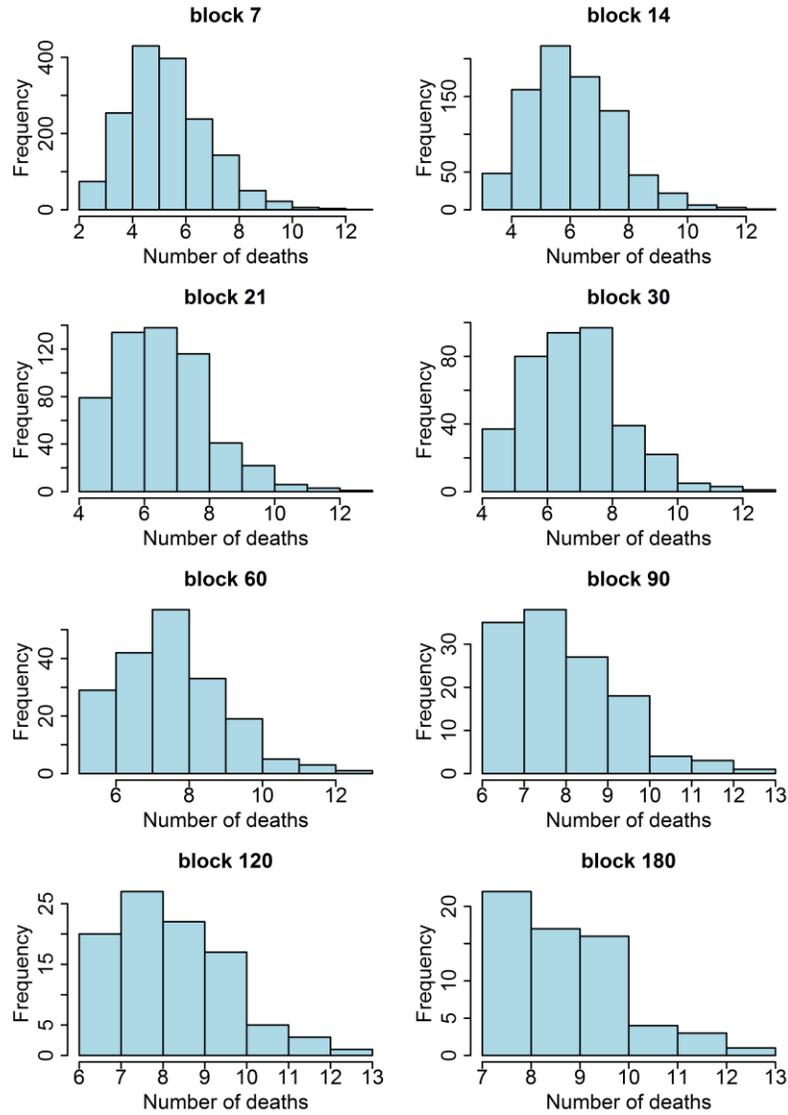


Figure I-8: Deaths peaks histograms in Quebec, depending on various block sizes expressed in days.

Histograms of the extracted peaks are shown in Figure I-8, depending on the block size. In the same manner as the hospitalisations in Montreal, the distribution changes with the increase of the block size. The lower bound goes up as raising the block size diminishes the probability of selecting small peaks (similar explanation to the scale parameter). The selected block size is 180 days (equivalent to 63 peaks) and corresponds to bi-annual peaks (Figure I-9). Unlike hospitalisation peaks in Montreal, the selection is performed on the whole dataset. This could be an advantage seeing that all the past information is kept. To verify the adequacy, the quantile-

quantile plot is shown in Figure I-10a. Overall, the GEV theoretical quantiles are aligned with the empirical ones, confirming a good fit.

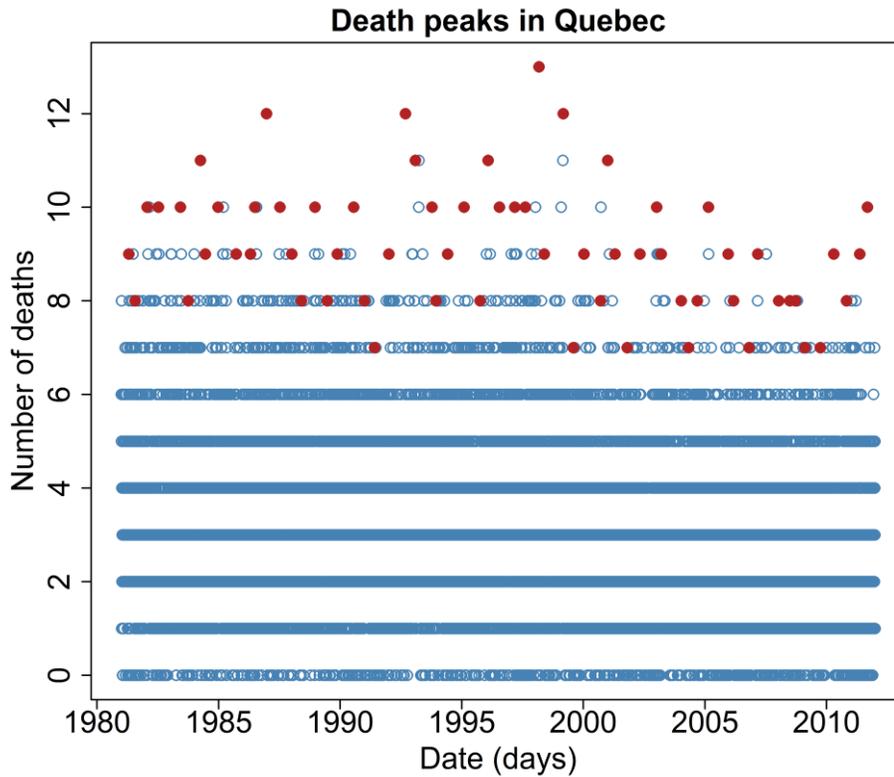


Figure I-9: Number of deaths (blue) and selected peaks (red) in Quebec.

The return level (Figure I-10b) indicates that over a long time period, 8 deaths will be reached or exceeded, on average once a year (the mean number of deaths is 3). When extending the period at 10 years, this level is 11 deaths. A hospital manager has at disposal a more complete description when considering those return levels along with the hospitalisation ones. Short term decisions could be made in accordance with the modeling results, before renovating hospital emergency rooms for instance, implying financial costs. It is however important to notice that confidence intervals grow with time, increasing the prediction uncertainty. Though the uncertainty is less obvious than in the case of hospitalisations in Montreal, it still reaches 11 to 15 deaths over 50 years.

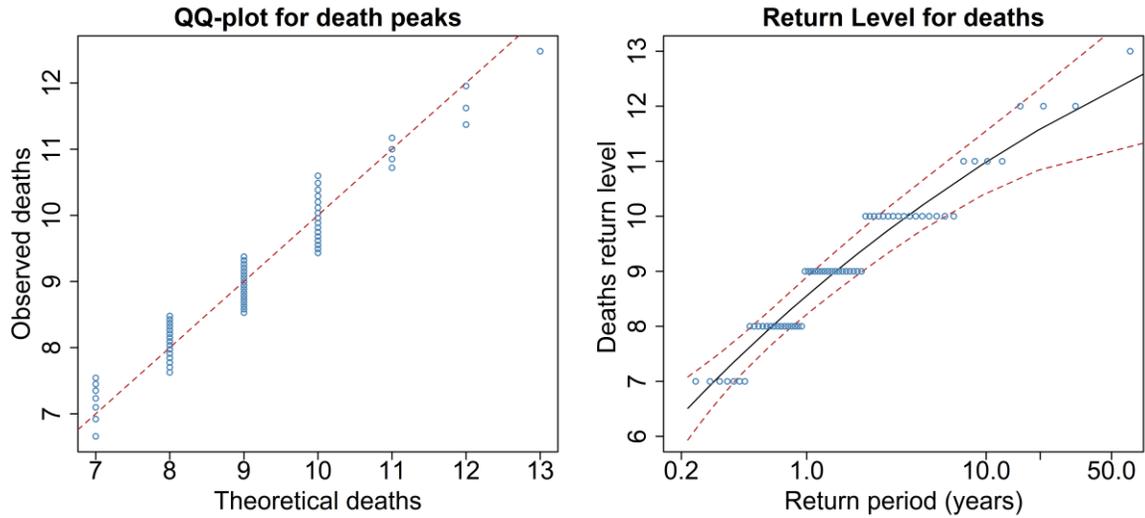


Figure I-10: QQ plot for deaths peaks in Quebec (left). Observations are in blue while the first bisector is in red. Return level for deaths in Quebec (right). Confidence intervals are in red.

For the death peaks in Quebec, it is also possible to select alternative distributions. This could be interesting for regression, for example. Since the GEV does not belong to the exponential family, its use in a regression model is more complex. On the contrary, lognormal and gamma distributions, which are not rejected by the adequacy tests, are members of the exponential family. Moreover, the RMSE of lognormal distribution (0.31) is equal to the GEV one, encouraging its use in a regression model. As for the exponential distribution (special case for a null shape parameter), its adequacy is rejected and its RMSE is the highest (7.6).

Table I-5 summarizes the main characteristics for the block and POT methods in this study. In particular, a trend in the raw data is to take into account when choosing one method or the other.

Table I-5: Main characteristics for the block and POT methods used in this study.

Method	POT	Block
Sample size	More important (threshold: 95 %)	Limited (block size: 180 days)
Peaks selection	No selection before 1998	Whole dataset
Trend	Not kept	Kept if present in raw data
Extreme distributions	GPD	GEV, lognormal and gamma
Bounds	Upper bound	Upper bound

4 Discussion

A public health manager would have an interesting tool at disposal with the return level. For instance, it allows a better preparation for periods during which peaks are likely to occur. In a hydrological context, Coles *et al.* (2003) explained that the return level is easier to interpret rather than the model parameters. Moreover, this piece of information brings a better financial prediction, as an increase in a hospital emergency room relates directly to certain costs. It is important to bear in mind that the return level indicates a level reached or exceeded once on average, over a long period of time. Since it is a mean measure, it is possible for this level to be reached more than once or none at all. It is recommended to take into account the associated confidence intervals. As a conclusion, McNeil (1998) explained that EVT supplies some of the best tools to quantify extreme events, although an uncertainty is inherent with small samples. In spite of this uncertainty, return levels still provide useful information during peak periods. Thus, the health care network can be more prepared given the associated confidence intervals, even if the additional precision is only a few hospitalisations or deaths. It has been showed (Robine *et al.*, 2008) that a few additional deaths here and there easily went by unnoticed during the 2003 European heatwave, with considerable cumulative consequences. It is however crucial to select

appropriate block sizes and thresholds since the estimated parameters and return levels depend on them.

From a practical point of view, the block method seems to be less appropriate than the POT method for peak extraction when there is a trend in the raw data. This property results in the MK and adequacy tests often rejecting their respective null hypotheses. As for the POT method, it does not seem affected by this issue since it only selects peaks in the recent years, in case of a raw data trend. This could be an advantage for the POT method as it allows meeting EVT requirements. In this study, 2 methods have been tested on 2 cities in order to make use of the most of EVT possibilities and explore results on different health variables. The results presented in this paper are drawn from Chiu *et al.* (2015), where exhaustive results and conclusions are presented. In particular, comparisons are also done on the same city or the same health variable. To complete the results shown in this paper, the POT method applied to the deaths in Quebec selects peaks in every year (as did as the block method applied on the same dataset in this study) because there is no trend. The block method applied to the hospitalisations in Montreal also selects peaks in all years, unlike the POT method applied on the same dataset in this study. Both methods give similar results to those presented in the paper in terms of the estimated shape parameters.

To the best of the authors' knowledge, this exploratory and modeling work is the first concerning CVDs. It displays new insights regarding peak extreme distributions. Particularly, the estimated shape parameter $\hat{\xi}$, crucial in EVT, is mainly negative. For the GEV distribution and GPD, this is equivalent to upper bounded distributions which means that the number of hospitalisations or deaths cannot grow indefinitely. This is in accordance with the study of De Zea Bermudez et Mendes (2012) on high cholesterol levels where GPD shape parameters are also mainly negative.

This result is strengthened by the work of Watts *et al.* (2006) who fitted GEV distribution to older populations in Canada and Japan. They obtain similar results regarding the shape parameter. This falls in line with other natural phenomena such as athletic performances (Henriques-Rodrigues *et al.*, 2011) that do show similar distributions.

Real datasets do not always respect i.i.d. conditions. For example, in this study, an increasing trend can be seen in the hospitalisations. A constant threshold is interesting in itself as it allows ruling out peaks that are considered common observations today. Nevertheless, if one needs to take into account those past peaks, non-stationary methods are more appropriate, such as models with parameters varying with covariates, see for example Chavez-Demoulin *et al.* (2015). Those complex methods are however not considered in the present study since it is the first time that health issues are dealt with this methodology in an extreme analysis framework. Moreover, including covariates means estimating more parameters. This is not always feasible given the short lengths of the extreme data series and the corresponding results could be unreliable. In addition, the complexity and interpretation of the return levels in a non-stationary case is to be taken into account (e.g. AghaKouchak *et al.*, 2012). Another option, which avoids using covariates, would be to use standardized ratios. Although in this case, the population standardized values have also been used (not shown) with no significant changes.

The deaths number may be treated as a discrete variable, because of its limited possible values. Although most of the studied variables in EVT are continuous, some applications have also been done on discrete variables, since they often occur in real-life situations (e.g. Nadarajah et Mitov, 2004; Mitov et Nadarajah, 2006). Anderson (1970) gives some theoretical properties for extremes of discrete variables. Most of the time, it is not physically possible to measure a real-life variable in a continuous fashion (this would require an infinite number of measurements); therefore using

the EVT on hospitalisations number (or streamflow in hydrology or returns in finance) is in fact a discrete approximation of a continuous variable. This is important to keep in mind when interpreting results since this approximation implies some uncertainties, especially concerning the goodness-of-fit. For example, Hitz *et al.* (2017) developed two discrete models based on the GPD model. They obtained better results in terms of goodness-of-fit with their refined models than with the classic GPD one. In the present study, this uncertainty was partly controlled by testing for various extreme distributions instead of fitting the GPD or GEV only. Rather than developing new models, another way to take into account the discrete nature of a variable in EVT would be to consider the arrival process. This leads to a Poisson process. However, this approach was not considered in this study because it is easier to compute the confidence intervals for the GEV and GPD return levels (Coles, 2001). There is a direct equivalence between a Poisson process and the GEV and GPD modeling concerning the model parameters and using one approach or the other should result in similar parameters.

5 Conclusion

Health peaks study is a serious challenge in public health. They are poorly understood by “classical” techniques based on mean events. Furthermore, they are scarcely studied per se leading to difficulties in predicting them. Identification and modeling steps for peaks are nonetheless imperative in order to fully characterize health events. This work aims at specifically analyzing health peaks with adequate statistical tools, applied to CVDs in the Province of Quebec. Despite the fact that CVDs are an economic and social burden, EVT had not been applied in this context yet.

Peaks modeling leads to hospitalisation and death return levels. Although uncertainty grows with the prediction horizon, return level is a useful tool that can allow for better decisions regarding

health care organisations. An analogy can be done with hydraulic works: it is essential to take into account return levels for a dam conception, in order for it to be built resilient enough for a certain time period. During a hospital refurbishment or construction, those details can be critical.

This study can also be of interest when searching for external factors (meteorological conditions, social stress, air pollution, etc.) possibly correlated to peaks. This can be the case in health research or surveillance, in particular if specific diagnoses are used instead of all CVDs. The previous arguments in favor of peak study do not mean that “classical” studies should not be conducted, rather that both types should be conducted in complementarity. This work focuses on CVDs in Quebec and Montreal, but the methodology can be applied to any other health peak series, i.e. to any other disease or region.

Acknowledgments

This work was funded by the *Fonds vert*, in the context of Action 21 of the 2006-2012 Quebec Action Plan on Climate Change. The authors thank the *Institut national de la santé publique du Québec* for access to the health data. The authors also wish to thank Jean-Xavier Giroux (INRS-ETE) for his expertise on the data and Pierre Masselot (INRS-ETE) and two anonymous reviewers whose comments significantly improved the quality of this paper.

II. MODÉLISATION DES PICS SANITAIRES POUR MALADIES CARDIOVASCULAIRES EN FONCTION DE LA MÉTÉO DANS LES COMMUNAUTÉS MÉTROPOLITAINES DE QUÉBEC ET MONTRÉAL

Y. Chiu ^a, F. Chebana ^a, B. Abdous ^b, D. Bélanger ^{a,c}, P. Gosselin ^{a,d}

^a *Institut national de la recherche scientifique, Centre Eau Terre Environnement, 490 rue de la Couronne, Québec (QC), G1K 9A9, Canada*

^b *Université Laval, Département de médecine sociale et préventive, 2325 Rue de l'Université, Québec (QC), G1K 7P4, Canada*

^c *Centre de recherche du Centre hospitalier universitaire de Québec, 2875 boulevard Laurier, Québec (QC), G1V 2M2, Canada*

^d *Institut national de santé publique du Québec, 945 avenue Wolfe, Québec (QC), G1V 5B3, Canada*

Rapport accepté à l'INSPQ le 19 février 2016

Résumé

Les décès et hospitalisations figurent parmi les variables sanitaires les plus étudiées en santé publique. Ces variables sont habituellement analysées au travers de l'ensemble des données disponibles puis mises en relation avec des variables explicatives (e.g. des variables météorologiques). Ces méthodes sont inadaptées pour appréhender les *pics* sanitaires, évènements « inhabituels » ayant de fortes répercussions sur le système de services et soins de santé. Leurs relations avec les variables explicatives peuvent ainsi être mal interprétées. Cette étude a ainsi pour but de modéliser et d'étudier les relations entre les pics sanitaires et la météo. Elle présente aussi un but exploratoire, puisque les pics ne sont habituellement pas extraits et étudiés en tant que tels. L'application est considérée sur les pics d'hospitalisations et de décès des maladies cardiovasculaires en relation avec les variables météorologiques (humidité, précipitations, pression, quantité de neige au sol, température), dans les communautés métropolitaines de Montréal et de Québec (Canada). L'étude se base ainsi sur les résultats de l'étude de Chiu *et al.* (2015), où ces mêmes pics ont été extraits et modélisés à l'aide d'outils statistiques adaptés aux pics. Les résultats montrent des différences et des similarités entre le cas classique des données agrégées et le cas des pics. Parmi les différences, les fortes déviations expliquées des modèles des pics peuvent être citées. Quant aux similarités avec le cas classique, il est possible de constater la forte influence de la variable de la température, entre autres. Les résultats peuvent être exploités afin d'améliorer certaines alertes sanitaires et météorologiques, dans l'établissement des seuils d'intervention basés sur le risque, par exemple.

1. Introduction

Les maladies cardiovasculaires (MCV) constituent la famille de maladies qui touchent le cœur ou la circulation sanguine, comme l'infarctus du myocarde ou l'accident vasculaire cérébral. Ce sont donc de graves maladies qui peuvent mener au décès ou à l'hospitalisation. Seulement au Québec, les MCV sont responsables de 19,5 % des décès totaux (deuxième cause de décès; Girard, 2013) et figurent à la première place des causes d'hospitalisation (Daigle, 2007). Cela se traduit par un poids socioéconomique élevé en services médicaux, en coûts d'hospitalisation, en perte de salaire, en perte de productivité et en nombre d'années de vie perdue. À l'échelle du Canada, les MCV constituent d'ailleurs la famille de maladies la plus coûteuse, avec des coûts directs et indirects estimés à 18,5 milliards de dollars par année (Daigle, 2007; Tu *et al.*, 2009).

Le Groupe d'experts intergouvernemental sur l'évolution du climat (IPCC, 2014, chapitre 9) avance avec un degré de confiance très élevé que le changement climatique, d'ici le milieu du XXI^e siècle, influera sur la santé humaine (surtout en exacerbant les problèmes de santé existants) ainsi que sur les coûts inhérents. Les MCV font partie de ces problèmes, car les températures extrêmes, tant l'été (Koppe *et al.*, 2004) que l'hiver (Beaudreau *et al.*, 2004), ajoutent un stress à des organismes déjà stressés par des maladies qui mettent à rude épreuve la réactivité vasculaire et la stabilité cardiovasculaire. Or, au Québec, des hausses significatives dans le nombre de nuits et de jours chauds ainsi que dans la durée des vagues de chaleur ont été observées de 1951 à 2010, et il est attendu que ces températures extrêmes maximales augmenteront plus que les températures moyennes estivales dans un futur rapproché. Par ailleurs, même si les températures extrêmes minimales en hiver augmenteront aussi plus que les températures moyennes hivernales, les projections pour le Québec montrent seulement une légère diminution dans la durée des vagues de froid (Ouranos, 2015).

Expliquer les MCV en fonction de la météo demeure donc d'un grand intérêt. Nombre d'études ont d'ailleurs été publiées à ce sujet (voir p. ex. Vanos *et al.*, 2014; Stewart *et al.*, 2016; Vanasse *et al.*, 2016). La méthodologie utilisée dans la présente étude vise toutefois à s'en distinguer de diverses manières.

Les études classiques analysant les variables de décès ou d'hospitalisations par MCV font habituellement usage de l'ensemble des données (voir p. ex. Siontis *et al.*, 2012 et Yang *et al.*, 2012 pour des applications à la prédiction du risque et aux modèles de survie, respectivement). Très nombreuses, ces études permettent de représenter la tendance, puis de comprendre et de prédire le caractère *moyen* de la variable sanitaire. Il s'agit d'une étape importante puisqu'il est nécessaire de se prémunir relativement aux situations *habituelles*, pour la gestion hebdomadaire d'un hôpital par exemple.

Cela dit, il existe aussi des événements *inhabituels*, appelés *pics sanitaires* (dénotés indifféremment « pics » ou « pics sanitaires » dans cette étude). Leur fréquence d'apparition est plus faible, mais leur impact plus marquant puisqu'il s'agit d'épisodes comptant des nombres de décès ou d'hospitalisations plus élevés qu'à l'habitude. Or, ces pics sont souvent mal captés par les études classiques, car la tendance générale ne présente qu'un portrait incomplet des données et des relations inhérentes. En d'autres mots, l'étude de la tendance dilue les informations propres aux pics. D'où la nécessité de traiter les pics avec des techniques statistiques appropriées afin que le réseau de la santé puisse parer non seulement aux événements moyens (études classiques), mais aussi aux pics sanitaires.

Diverses techniques adaptées aux pics ont souvent été appliquées avec succès dans d'autres sphères d'activité que la santé publique, notamment dans les domaines de l'hydrologie (Singh,

1998) et de la finance (Embrechts et Schmidli, 1994). Parmi ces techniques figure la théorie des valeurs extrêmes (EVT), très rarement exploitée dans le domaine de la santé publique.

En fait, seulement De Zea Bermudez et Mendes (2012), Guillou *et al.* (2014), Chen *et al.* (2015), Thomas *et al.* (2016) et Chiu *et al.* (2016a) se sont intéressés aux variables sanitaires extrêmes. Ces auteurs font usage de l'EVT pour modéliser les extrêmes de certaines variables sanitaires (p. ex., niveaux de cholestérol) par une loi adaptée à ce type de données, pour ensuite en étudier les propriétés. De leurs avis, les modèles basés sur l'EVT apparaissent d'ailleurs tout indiqués pour analyser les pics sanitaires.

Par rapport aux autres études, celle de Chiu *et al.* (2016a) est toutefois particulière. *Primo*, elle est la seule à s'être penchée sur les pics de MCV et plus spécifiquement sur les pics de décès et d'hospitalisations par MCV dans les communautés métropolitaines de Québec et de Montréal. *Secundo*, elle tente d'établir une méthodologie générale pour extraire les pics. *Tertio*, elle a introduit diverses séparations de données (p. ex., fins de semaine) selon plusieurs modèles et testé les hypothèses de base de l'EVT. Un des résultats ainsi obtenus est le niveau de retour associé à une période de retour (qui indique le temps moyen que mettra un événement avant d'atteindre ou de dépasser un certain niveau), outil intéressant p. ex. pour un gestionnaire d'hôpital. Les résultats de Chiu *et al.* (2016a) doivent donc être considérés comme une première étape à la présente analyse, puisqu'il est essentiel d'étudier en détail la variable à expliquer avant de modéliser les liens qui l'unissent à des variables explicatives.

Les conditions météorologiques ont maintes fois été considérées comme indicateurs de risque du caractère *moyen* de certains événements par MCV (p. ex., Portier *et al.*, 2010). Ainsi, à titre d'exemple, les variations de la pression atmosphérique ont été associées à l'infarctus du myocarde, tout comme la quantité de neige au sol (Houck *et al.*, 2005; Hopstock *et al.*, 2012). De

même, la relation entre la température moyenne et la mortalité par MCV a été documentée (Goldberg *et al.*, 2011). Enfin, il a été mis en lumière que le pic de mortalité survient très rapidement dès le début d'une canicule (Curriero *et al.*, 2002), voire sur une période allant jusqu'à trois jours (Ledrans et Isnard, 2003).

Par contre, aucune étude n'a encore mis en relation les pics sanitaires avec les données météorologiques; comme déjà rapportés, ces pics ont très rarement fait l'objet d'analyses spécifiques (p. ex. en utilisant des outils statistiques pour leur extraction, Chiu *et al.*, 2015). Conséquemment, il n'existe pas d'éléments de réponse quant aux liens existants entre les pics et la météo. Les études de la littérature classique (autrement dit, autres que l'étude des pics) utilisent généralement des méthodes de régression, paramétriques ou non (voir p. ex. Ren *et al.*, 2006; Abrignani *et al.*, 2012). Par rapport aux premières, les méthodes non paramétriques ont l'avantage d'une plus grande flexibilité dans la modélisation des relations. Cet aspect est important dans le cas présent, puisque la présente étude est la première qui s'intéresse uniquement aux pics et qu'il est donc difficile de se faire une idée *a priori* de la forme de la relation entre les pics MCV et les variables météorologiques. L'utilisation des méthodes non paramétriques est par ailleurs reconnue et bien établie dans les études de l'effet de la pollution de l'air ou de la météo sur les MCV dans un contexte classique (voir p. ex. les études de Dominici *et al.*, 2002; Goggins *et al.*, 2012).

Cette étude se propose donc d'étudier la relation entre les pics d'hospitalisations et de décès pour cause de MCV extraits dans Chiu *et al.* (2015) et les conditions météorologiques dans un cadre flexible (non paramétrique), afin de soutenir une prise de décisions plus éclairées à leur sujet. Un but exploratoire est aussi associé à cette étude, dans la mesure où les pics ne sont habituellement pas reliés à la météo de cette façon.

Le reste du chapitre est organisé comme suit. La méthodologie de l'étude est décrite dans la Section 2. La Section 3 présente l'application et les résultats obtenus. Enfin, la Section 4 contient la discussion, la Section 5 la conclusion et la Section 6 les limitations de l'étude.

2. Analyse statistique

Les pics sanitaires constituent les variables de sortie (variable d'intérêt, dépendante ou à expliquer). En premier lieu, avant de modéliser les effets de la météo sur les pics sanitaires, il est nécessaire d'extraire ces derniers d'une façon adéquate, c'est-à-dire de s'assurer que les évènements sélectionnés sont bien des pics sanitaires et non des évènements « communs ». Pour toutes ces étapes, l'EVT fournit des outils statistiques adaptés, mis à contribution par Chiu *et al.* (2015). Les auteurs ont extrait les pics de décès et d'hospitalisations provenant des mêmes données sanitaires utilisées ici. L'extraction s'est faite à l'aide des méthodes par blocs et par seuil, qui ont donné lieu à une modélisation des différentes séries de pics par des distributions extrêmes. En EVT, les conditions d'indépendance et d'homogénéité des pics sanitaires sont requises. Pour vérifier ces conditions, une étape de prétraitement des données (« déclusterisation ») a précédé l'extraction. De plus, plusieurs séparations ont été explorées (fins de semaine/jours de semaine et été/hiver). L'indépendance et l'homogénéité ont été ainsi respectées. Des tests d'hypothèse ont permis de vérifier ces conditions sur les séries extraites. Les séries de pics sanitaires ont finalement été modélisées selon plusieurs distributions extrêmes et le meilleur ajustement a été retenu. Cela a permis, entre autres, d'obtenir des périodes et des niveaux de retour.

Une fois les pics extraits et modélisés, les modèles additifs généralisés (GAM; Hastie et Tibshirani, 1986) sont utilisés afin d'étudier la relation entre les pics sanitaires et la météo. Les

GAM présentent une grande flexibilité dans le cadre d'une modélisation non paramétrique, car ils laissent les données décider de la forme de la relation. Cette caractéristique des GAM permet une compréhension plus détaillée de l'influence des événements météorologiques sur la mortalité/morbidité par MCV. Par ailleurs, cette flexibilité est un aspect attrayant pour cette étude puisqu'aucune connaissance *a priori* sur la forme des relations entre les pics et la météo n'est disponible. En effet, les GAM reposent sur la modélisation de la réponse par une somme de fonctions non linéaires qui s'adaptent aux données. Dans le contexte ordinaire (pas des pics), Braga *et al.* (2002) et Doyon *et al.* (2008) précisent justement que la relation entre la mortalité et la météo peut être en forme de « J », de « U » ou de « V ». Ceci signifie que les décès sont nombreux lors de canicules et de périodes de grands froids, mais une « zone de moindre effet » existe entre les deux. Un résultat similaire pour les hospitalisations par maladies coronariennes est détecté par Bayentin *et al.* (2010) au Québec : les auteurs rapportent une forme de « U » ou de « V » pour la relation entre les hospitalisations et les variables météorologiques. Enfin, la partie *additive* des GAM signifie qu'il est possible d'examiner séparément l'influence de chaque variable, p. ex. au travers de graphiques, compte tenu des autres variables fixées à leurs valeurs moyennes. Ces qualités des GAM ont contribué à leur utilisation dans des études de santé publique concernant l'impact de la météorologie ou de la pollution de l'air sur la santé (voir p. ex. Dominici *et al.*, 2002 ou Curriero *et al.*, 2002). À noter que les études sur la santé environnementale impliquant des modèles GAM sont de plus en plus nombreuses, voir Jbilou et El Adlouni (2012) pour une revue de la littérature. Pour plus de détails, voir l'Annexe 2 : Modèle additif généralisé.

La variable réponse dans un GAM peut suivre une variété de distributions appartenant à la famille exponentielle (Hastie et Tibshirani, 1986), dont la loi lognormale. Cette dernière figure

parmi les distributions les plus utilisées pour modéliser les séries d'extrêmes, particulièrement en hydrologie et en finance, qui sont deux domaines où l'étude des extrêmes est répandue. En particulier, elle a été déterminée comme distribution appropriée pour la majeure partie des séries de pics dans l'étude de Chiu *et al.* (2015). Deux cas particuliers ont été modifiés : les hospitalisations à Montréal et à Québec, en jours de semaine. En effet, les seuils choisis pour ces deux cas par Chiu *et al.* (2015) amènent à des séries modélisables par la distribution de Pareto généralisée, mais pas par la distribution lognormale. Étant donné que l'intérêt ici n'est plus l'ajustement de distribution mais bien de modéliser des liens entre une variable dépendante et des variables explicatives, il est possible de modifier les seuils sélectionnés (aussi longtemps que les choix sont valides d'un point de vue statistique). En les sélectionnant à 99 % et à 85 % respectivement, toutes les conditions de l'EVT sont toujours respectées et la distribution lognormale est adéquate pour modéliser les pics. Cette adéquation est un point non négligeable car toutes les propriétés théoriques du GAM sont ainsi conservées.

La forme générale du GAM est donnée par l'équation suivante (Wood, 2006) :

$$g(y_i^*) = \beta_0 + f_1(x_i) + f_2(\text{temps}_i) \quad (11)$$

où les y_i^* sont les pics sanitaires, g est la fonction de lien, les x_i sont les variables météorologiques et les f sont des fonctions de lissage. Pour ces dernières, les splines cubiques sont choisies (Hastie et Tibshirani, 1990, chapitre 2). Ainsi, chaque variable explicative n'est plus nécessairement reliée de façon linéaire à la variable réponse, et le modèle gagne en flexibilité en n'imposant aucune contrainte sur la forme des relations. L'interprétation se fait au travers des fonctions f_i (p. ex. à l'aide de graphiques). L'avantage est de disposer maintenant d'une influence continue sous la forme d'une fonction lisse, contrairement à une influence sous la forme d'un paramètre. Il est important de noter que dans l'équation (11), l'indice i indique les

jours de survenue de pics, à la différence du cas classique où l'indice serait journalier p. ex. Étant donné que les pics surviennent à plusieurs jours d'intervalle (dû entre autres à la procédure de « déclusterisation » appliquée par Chiu *et al.*, 2015), l'indice i ne représente donc pas des indices consécutifs.

Dans cette étude, il est difficile d'inclure plus de deux variables lissées par modèle, la contrainte étant que les pics sanitaires sont par définitions rares (et les modèles peu puissants en raison de petits effectifs). Il en résulte donc des séries courtes (relativement aux données brutes). Ce problème ne se pose habituellement pas, car les études classiques font usage de l'ensemble des observations contrairement au cas présent. Les splines et autres fonctions de lissage impliquent plus de paramètres à estimer qu'une régression linéaire, donc plus de données disponibles. Le degré de complexité associé à une spline est mesuré à l'aide du degré de liberté équivalent (ou *edf* en anglais, voir Wood, 2006, chapitre 4). Ce dernier mesure la flexibilité du modèle au niveau de la variable explicative associée. Un *edf* égal à 1 correspond à une droite tandis qu'un *edf* plus grand indique que la courbe s'adapte plus aux données et devient en contrepartie plus sinueuse. Afin d'éviter un éventuel biais, l'*edf* est choisi de façon optimale et automatique par validation croisée généralisée (ou *GCV* en anglais). Cependant, le nombre d'observations dans les séries des pics sanitaires est souvent insuffisant pour l'estimation de plus de deux fonctions lissées. Aussi, le choix est fait de n'étudier qu'une variable explicative lissée à la fois, la deuxième variable lissée étant la spline du temps, incluse par défaut dans tous les modèles. L'inclusion du temps est une pratique commune afin de capter les divers effets de la tendance et de la saisonnalité (p. ex. Dominici *et al.*, 2002 ou Doyon *et al.*, 2006). Il est à noter que dans les études entre la santé et la météo, les lisseurs LOESS sont aussi utilisés (Schwartz, 1999).

Dans un cadre non paramétrique, un cas de figure problématique est celui où la fonction lissée tente de trop s'adapter aux données (ou *surajustement*). Cela est particulièrement vrai dans le cas de petits échantillons, comme dans cette étude où les pics sont par définition rares. Les fonctions ainsi obtenues avec un grand edf peuvent être inexploitable car insuffisamment lissées. Une solution qui corrige grandement ce problème sans compromettre l'estimation est d'introduire un paramètre de pondération dans le calcul de la GCV (Kim et Gu, 2004; Wood, 2006). Dans cette étude, c'est cette approche qui est utilisée, afin d'éviter des courbes trop sinueuses.

Par ailleurs, il est intéressant de considérer les variables météorologiques décalées d'un certain retard (ou *lag*, voir p. ex. Ren *et al.*, 2006). Il est en effet reconnu que l'action des variables météorologiques sur la santé s'étend sur plusieurs jours. Ainsi, l'évènement météorologique peut survenir un ou plusieurs jours avant l'évènement sanitaire (Braga *et al.*, 2002; Goldberg *et al.*, 2011; Breitner *et al.*, 2014). Les variables météorologiques sont moyennées sur le nombre de jour correspondant au lag optimal k . Ce dernier est toutefois propre à chaque variable météorologique et à chaque maladie. Plusieurs études montrent que la somme ou la moyenne de plusieurs jours avant l'évènement sanitaire par MCV joue un rôle significatif sur ce dernier, que ce soit une hospitalisation ou un décès (Curriero *et al.*, 2002; Schwartz *et al.*, 2004; Medina-Ramon et Schwartz, 2007). Cela permet en effet de prendre en compte l'information contenue dans les jours précédents, plutôt que de s'intéresser à des évènements ponctuels.

Dans le présent projet, des lags allant jusqu'à 7 jours sont considérés pour les hospitalisations et les décès par MCV. Au meilleur de la connaissance des auteurs, il n'existe pas de littérature sur laquelle se baser pour sélectionner les lags appropriés pour la relation entre la météo et les *pics* par MCV. Cependant, en se basant sur le cas classique, Curriero *et al.* (2002) mentionnent que les lags les plus intéressants pour les décès par MCV vont jusqu'à 3 jours précédant la survenue de

l'événement. Concernant les hospitalisations par MCV, Ebi *et al.* (2004) considèrent des lags jusqu'à 7 jours.

Plusieurs modèles explicatifs sont évalués pour 1) les hospitalisations à Montréal, 2) les hospitalisations à Québec, 3) les décès à Montréal et 4) les décès à Québec. Pour chacun des cas, les deux étapes suivantes sont considérées :

- 1) Chaque variable météorologique est introduite dans le modèle de façon linéaire (sans passer par des splines) au lag optimal choisi à l'aide du critère d'information Akaike (AIC). Ce lag va de 0 à 7 jours, c'est-à-dire du jour même à une semaine avant le pic.
- 2) Chaque variable météorologique est introduite dans le modèle sous forme de fonction lissée au lag optimal, allant de 0 à 7 jours, choisi par AIC.

Les variables météorologiques sont moyennées du lag optimal jusqu'au pic sanitaire, de sorte à moyenner l'effet des conditions météorologiques sur plusieurs jours plutôt que d'utiliser des valeurs ponctuelles. En effet, il est reconnu que la moyenne des conditions météorologiques sur plusieurs jours avant l'évènement sanitaire par MCV peut jouer un rôle significatif sur ce dernier, que ce soit une hospitalisation ou un décès (Curriero *et al.*, 2002; Vanasse *et al.*, 2017). D'autres études utilisent la somme plutôt que la moyenne (Schwartz *et al.*, 2004; Medina-Ramon et Schwartz, 2007). Les études classiques ne s'intéressent cependant pas exclusivement aux pics sanitaires et les lags utilisés dans ces dernières peuvent ne pas être comparables à ceux adaptés à l'étude des pics sanitaires. Ainsi, la décision est prise de limiter les lags à 7 jours. De plus, les étapes 1) et 2) incluent une variable météorologique accompagnée de la fonction du temps, qui est incluse dans tous les modèles pour « capter » l'influence du temps comme les tendances et saisonnalité éventuelles. Ceci est une pratique courante dans les études des relations santé-météo (voir p. ex. Schwartz *et al.*, 2004; Lim *et al.*, 2012). En plus de la fonction du temps, les autres

variables météorologiques sont aussi introduites de façon linéaire lorsque la taille d'échantillon le permet. Celles qui sont jugées non significatives sont retirées. La significativité des fonctions lissées est aussi jugée à l'aide d'un test statistique. Il s'agit en réalité d'un test de nullité des paramètres des splines (Wood, 2006, chapitre 4). Dans le cas où plusieurs modèles significatifs sont retenus, la comparaison est faite en se basant sur la minimisation des critères AIC ou GCV. Dans le cas de modèles emboîtés, l'équivalent d'une ANOVA pour GAM peut être implémenté pour décider du meilleur modèle en termes explicatifs (Wood, 2006, chapitre 4).

Les étapes sont résumées à la Figure II-1. Tous les résultats des tests statistiques sont présentés au seuil de significativité $\alpha = 5\%$.

3 Résultats

Cette section contient l'analyse descriptive des données sanitaires et météorologiques, ainsi que les résultats des estimations des GAM pour les hospitalisations et les décès, à Montréal et à Québec. La méthode d'extraction basée sur l'EVT ainsi que la distribution adéquate des pics sont présentées dans Chiu *et al.* (2015); chaque cas (hospitalisations ou décès, Montréal ou Québec) rappelle la méthode d'extraction utilisée et les échantillons retenus. Les résultats sont obtenus à l'aide du logiciel libre R et du package *mgcv* (v1.8-2, Wood, 2011). Un résumé de tous les résultats obtenus est présenté à la fin de cette section.

Force est de rappeler l'impossibilité d'inclure plus de deux variables explicatives lissées dans un modèle des pics sanitaires, en raison de leurs petits nombres (voir la section 2). La variable du temps étant incluse par défaut, cela laisse place à une seule variable météorologique.

3.1 Analyse descriptive

Avant de modéliser les relations entre les pics sanitaires et la météo, une brève analyse descriptive des pics sanitaires et des données météorologiques est effectuée afin d'avoir une meilleure idée du comportement des variables concernées.

3.1.1 Données sanitaires

Un portrait complet des données sanitaires brutes a été dressé dans le rapport de Chiu *et al.* (2015). Dans le présent chapitre, l'accent est donc porté sur les pics sanitaires extraits. Les différents éléments en lien avec l'extraction des pics sanitaires par Chiu *et al.* (2015) sont résumés au Tableau II-1. Les échantillons sont de tailles 33, 46 et 115 respectivement pour les hospitalisations à Montréal et à Québec (jours de semaine, jds, et fins de semaine, fds), et de 42, 32 et 63 pour les décès à Montréal (jds et fds) et Québec. La plus grosse taille d'échantillon (115) est due au seuil relativement bas (85 %) par rapport aux autres (97,5 % et 99 %). Plusieurs modèles ont été étudiés par Chiu *et al.* (2015) afin d'extraire les pics. Les modèles retenus ici sont ceux jugés les meilleurs selon les divers tests et critères statistiques. En particulier, les auteurs ont vérifié les hypothèses de stationnarité, d'homogénéité et d'indépendance ainsi que l'adéquation aux distributions extrêmes des différentes séries de pics. Puisqu'un même seuil (ou taille de blocs) ne convient pas nécessairement à chaque cas, il en résulte des tailles d'échantillons différentes.

Pour comparer la distribution des pics retenus par Chiu *et al.* (2015), les histogrammes sont tracés à la Figure II-2. Toutes les distributions sont asymétriques, avec une queue plus longue à droite. Cette remarque est compatible avec le fait que seuls les pics sont retenus. En effet, les méthodes du seuil et des blocs reviennent à n'observer qu'une partie des distributions d'origines,

celle située dans la queue à droite. Puisque les pics sont considérés, les échelles des histogrammes sont supérieures à celles des observations brutes. P. ex., les pics d'hospitalisations journaliers à Montréal s'étendent de 190 à 220 tandis que dans le cas des données brutes, les hospitalisations journalières vont de 30 à 220. De même, les pics de décès journaliers à Québec s'étendent de 7 à 13 tandis que pour les données journalières brutes, la variable est comprise entre 0 et 13.

3.1.2 Données météorologiques

Les variables de température et de pression atmosphérique (« pression » dans la suite du texte) sont des relevés horaires. De ces derniers sont extraites les valeurs minimales et maximales journalières. Concernant l'humidité relative (« humidité » dans la suite du texte), les mesures à 6 h et 15 h sont utilisées respectivement pour les valeurs maximales et minimales (Hmax et Hmin). Il est possible que ces valeurs ne correspondent pas aux vraies valeurs maximales et minimales sur 24 h, mais en raison d'une quantité importante de données manquantes, ces définitions ont été adoptées selon les recommandations d'Environnement Canada. Les données de précipitations totales (Prec) et d'épaisseur de neige au sol (Snow) sont, quant à elles, journalières.

Les histogrammes des variables météorologiques sont tracés afin d'observer les différentes distributions. Ils sont présentés à la Figure II-3. Seules les variables maximales sont retenues, comme expliqué dans les prochains paragraphes. La distribution de la température est bimodale, avec des modes autour de 0 °C et 25 °C. Cela indiquerait que le climat est principalement marqué par deux saisons aussi bien à Montréal qu'à Québec. La pression atmosphérique est distribuée presque normalement, tandis que l'humidité relative (ou « humidité » dans la suite du texte), les

précipitations et la quantité de neige au sol (ou « neige » dans la suite du texte) sont asymétriques. L'asymétrie est marquée de façon prononcée pour les deux dernières variables, ce qui semble compatible avec le fait que ces variables prennent des valeurs non nulles uniquement dans le cas de pluies ou de chutes de neige.

Pour étudier plus en profondeur les variables météorologiques, les différentes relations entre elles sont considérées (Figure II-4 pour Montréal et Figure II-5 pour Québec). Au vu de ces figures, il n'y a généralement pas de relation claire entre les variables météorologiques. Il reste possible de distinguer certaines tendances, comme la température qui est croissante en fonction de l'humidité.

La Figure II-4 et la Figure II-5 ne présentent que les variables maximales de l'humidité, la pression atmosphérique et la température. La raison est illustrée par l'exemple suivant concernant la température (le raisonnement est similaire pour l'humidité et la pression atmosphérique) : les variables minimales, moyennes et maximales de la température sont fortement corrélées entre elles, avec un coefficient de Pearson de l'ordre de 90 %, comme indiqué dans le Tableau II-2. Un fort biais serait donc introduit dans les modèles en incluant les trois versions de la variable température en même temps. Puisque les corrélations sont élevées, il est probable que les différents déclinaisons (minimum, moyenne, maximale) d'une même variable apportent la même information. En effet, les forts coefficients de corrélation indiquent qu'une variable météorologique maximale varie de façon similaire à une variable minimale ou moyenne, à un facteur de translation près. La seule exception est la corrélation entre l'humidité minimale et maximale (44 %). Le choix est ainsi fait de conserver les variables maximales pour expliquer les pics sanitaires dans cette étude. Il est possible de raffiner les modèles décrits ici en sélectionnant les variantes minimales ou moyennes, mais cela ne modifierait pas de façon significative les

interprétations. De plus, dans la visée exploratoire de cette étude, l'intérêt est de développer une méthodologie objective et généralisable dans un premier temps. Sélectionner différentes variantes d'une variable météorologique au cas par cas ne permettrait pas d'atteindre cet objectif.

Les statistiques descriptives des variables météorologiques ainsi retenues sont données au Tableau II-3. Toutes les variables sont positives, excepté la température maximale. A noter qu'il est impossible dans des conditions normales d'obtenir une pression atmosphérique ou une humidité relative négative. Les variables sont du même ordre de grandeur entre Montréal et Québec, avec quelques différences notables comme les températures maximales plus basses à Québec ou l'humidité maximale plus élevée à Montréal. Par ailleurs, Québec reçoit beaucoup plus de neige que Montréal.

3.2 Modélisations des pics d'hospitalisations en fonction de la météo à Montréal

Les pics d'hospitalisations à Montréal sont extraits par la méthode du seuil élevé à 99 % (Chiu *et al.*, 2015). Cette série de pics extraits est expliquée grâce à un GAM, avec pour variables explicatives les différents variables météorologiques lissées. Ces dernières sont sélectionnées selon la significativité de leur p-value et selon le lag choisi par AIC (section 2).

Avant de se lancer dans la modélisation, il est possible d'effectuer une analyse graphique et descriptive des relations entre les pics et les conditions météorologiques correspondantes (Figure II-6 et Tableau II-4 respectivement). La Figure II-6 révèle un effet plus prononcé des températures chaudes, confirmé par le Tableau II-4, puisque la moyenne se situe à 4,54 °C. De plus, les températures minimales et maximales associées aux pics sont à -11,36 °C et à 25,41 °C respectivement, indiquant une plus grande vulnérabilité (du point de vue des pics) dans les températures chaudes. À noter, il semble il y a voir moins de pics autour de 10 °C. Les autres

variables météorologiques ne montrent pas de relation particulière et une modélisation non paramétrique est alors nécessaire.

Les résultats de la modélisation sont résumés au Tableau II-5. Les pics sont en nombre restreint : $n_{pics} = 33$ comparativement aux données brutes avec $n_{total} = 4\ 077$. Le seul modèle retenu pour les pics d'hospitalisations à Montréal est celui avec la température, moyennée sur 5 jours. Il est retenu sur la base de la méthodologie de la section 2. Considérant qu'il s'agit d'un modèle avec une seule variable explicative, la DEX est relativement élevée, de l'ordre de 39,6 %. Cela est à comparer au cas classique avec toutes les observations (toujours avec les fonctions de la température moyennée sur 5 jours et du temps), où la déviance expliquée atteint 21 % soit quasiment deux fois moins que dans le cas des pics. Cela peut s'expliquer par le fait que cette étude soit placée dans un contexte de *pics* et non plus de l'*ensemble* des données. En se restreignant uniquement aux pics d'hospitalisations, l'accent est mis sur une plus petite partie de la variabilité de l'ensemble des hospitalisations. Autrement dit, l'effet de l'exposition (ici la température) sur l'ensemble des cas MCV est *dilué*. De plus, cette variabilité spécifique des pics d'hospitalisations semble être expliquée en plus grande partie par la température comparativement au cas classique, où d'autres facteurs non météorologiques sont plus susceptibles d'intervenir. Enfin, la température étant reconnue comme influente dans le contexte classique, il est intuitif qu'elle le soit toujours dans le cas des pics, mais son importance dans ce dernier cas était jusqu'alors méconnue.

Il existe une autre différence importante entre le cas des pics et le cas classique : le modèle sans le temps dans le premier cas possède une déviance expliquée de 40,1 %, tandis que celle du modèle dans le deuxième cas est de 2,87 %. Les déviances des modèles des pics avec et sans le temps sont proches, alors que celles des modèles de l'ensemble des observations diffèrent de

18 %. Cela peut renforcer l'idée que la température est responsable d'une partie des pics plus importante que dans le cas de toutes les observations. Ainsi, dans ce dernier cas, la fonction du temps (effet de la saison, fins de semaine, etc.) capte la majeure partie de la variabilité des hospitalisations. L'edf (7,95) est d'ailleurs nettement plus élevé que celui du modèle avec le temps (3,40). Conséquemment, retirer ce bruit de fond du modèle résulte en une chute de la déviance expliquée (2,87 %).

Les courbes des effets de la température avec le temps inclus dans le modèle et de la température sans le temps possèdent respectivement un edf de 4,11 et de 4,26. Elles sont représentées à la Figure II-7 et ont sensiblement la même forme (alors que le temps n'a presque pas d'effet), présentant un « J » inversé. Cette forme est cependant due à un pic isolé de 220 hospitalisations qui mérite une analyse plus poussée (prochain paragraphe). Les fonctions sont globalement décroissantes en fonction de la température (moyennée sur 5 jours). Cela indique une forte influence des températures froides sur les pics d'hospitalisations par MCV. De plus, deux autres zones d'influence (correspondant aux légères remontées de la courbe) sont observables autour de 0 °C et de 20 °C. À l'opposé, autour de -5 °C et de 15 °C se situent deux zones d'inflexion, indiquant possiblement des températures de moindre influence. La fonction du temps est une droite n'indiquant pas de tendance particulière. La Figure II-8 montre les fonctions de la température cumulée sur 5 jours sur l'ensemble des hospitalisations à Montréal, avec et sans le temps (les estimations ne sont pas présentées comme à la Figure II-7, car l'intérêt est de comparer les fonctions, non les estimations). L'allure des fonctions est différente selon que le temps soit inclus ou non, avec un edf passant respectivement de 3,40 à 7,95 (Tableau II-5). Cette différence peut provenir du fait que le temps permet de capter la saisonnalité et la tendance avec un edf

relativement élevé de 8,49. Sans le temps, la température doit s'adapter aux données, avec visiblement plus d'irrégularités puisque l'edf a plus que doublé (de 3,40 à 7,95).

Sur la Figure II-7, l'interprétation de la courbe des effets doit cependant se faire avec précaution car un pic est isolé à gauche et donne à la courbe sa forme de « J » inversé. Il s'agit du pic le plus élevé de la série (220 hospitalisations). Son estimation s'avère assez précise, même si elle est légèrement sous-estimée par le modèle à 217. Il est pertinent de l'analyser en détail, car ce pic pourrait être retiré de l'analyse si son influence sur la courbe n'est pas justifiée (une erreur de mesure, p. ex.). Dans ce cas, la courbe n'aurait probablement pas la même forme. La Figure II-9 présente les conditions météorologiques (non moyennées et moyennées au lag choisi, 5 jours) ayant précédé le pic en question, survenu le 12 janvier 2004. Même si seule la température maximale est significative dans le modèle sélectionné, toutes les variables météorologiques sont présentées car le but est de tenter d'expliquer ce pic en particulier. Il est possible qu'elles aient toutes été impliquées dans la création de ce pic, mais pas nécessairement dans l'ensemble des pics. Les températures brutes maximales chutent de façon abrupte, du 3 au 9 janvier, soit 3 jours avant le pic : de 4,62 °C à -20,2 °C en 6 jours. Ebi *et al.* (2004) expliquent que des variations de température influent grandement sur les hospitalisations par MCV. Les basses températures pourraient aussi avoir provoqué une augmentation de la quantité de neige au sol, comme indiqué sur la Figure II-9. De plus, l'humidité et la pression ont atteint un minimum et un maximum local (8 janvier et 9 janvier), respectivement. Cette combinaison de variables météorologiques peut avoir eu des répercussions physiologiques sur la santé cardiovasculaire et dans ce cas, le pic observé le 12 janvier pourrait avoir réellement été observé. Ces conditions météorologiques particulières permettent donc d'appuyer la présence d'un pic et non d'une valeur aberrante, d'où l'intérêt de le conserver dans l'analyse. L'effet est plus remarquable à la Figure II-9 avec les

variables moyennées sur 5 jours : les températures, humidité et précipitations moyennées indiquent un minimum local le jour du pic tandis que la pression est à un maximum local. La quantité de neige cumulée est quant à elle croissante et le pic survient une semaine après le minimum local.

3.3 Modélisations des pics d'hospitalisations en fonction de la météo à Québec

La méthode d'extraction retenue pour les pics d'hospitalisations à Québec est celle basé sur le seuil élevé (Chiu *et al.*, 2015). Afin de respecter les hypothèses de modélisation de l'EVT, la séparation en jours de semaine et fins de semaine a aussi été nécessaire (de lundi à jeudi pour les jours de semaine et de vendredi à dimanche pour les fins de semaine). Les seuils sont respectivement de 97,5 % et de 85 %, basés sur une série de tests statistiques et sur minimisation de l'AIC. Cette séparation donne donc lieu à deux modèles de régression complémentaires d'un point de vue temporel. Les résultats sont décrits ci-après.

3.3.1 Pics d'hospitalisations à Québec – jours de semaine

On dispose de 46 pics d'hospitalisations pour les jours de semaine. La Figure II-10 montre la façon dont les pics sont répartis en fonction de la météo et le Tableau II-6 résume les statistiques descriptives. Les températures chaudes semblent être plus influentes puisqu'elles sont en moyenne de 7 °C et vont jusqu'à 28,63 °C, tandis que le minimum est à -16,26 °C. À noter, la zone de moindre effet pour la température autour de 10-15 °C est visible (Figure II-10). Moins de pics semblent survenir dans cet intervalle de températures, comme dans le cas des hospitalisations à Montréal. La pression atmosphérique possède aussi une telle zone, autour de 102 kPa (dans une atmosphère normale, elle varie autour de 101 kPa). La quantité de neige au sol influe lors de ses valeurs élevées et nulles (il ne neige qu'une partie de l'année, expliquant ses

nombreuses valeurs nulles). L'humidité et les précipitations ne révèlent pas de relation particulière (nuages de points).

Les estimations des GAM sont résumées au Tableau II-7. Après évaluation des variables significatives et choix du lag, un modèle est retenu : celui avec la fonction des précipitations moyennées sur 7 jours ainsi que la quantité de neige au sol du jour même (introduite de façon linéaire). Même si la Figure II-10 ne montrait pas de lien avec les précipitations, elles sont ici moyennées sur 7 jours, ce qui peut mettre en lumière des relations non visibles sur un simple graphique. Un peu plus d'un tiers de la déviance est expliquée par ce modèle ($DEX = 38,7\%$). Cela n'est pas nécessairement dû au temps car le modèle sans le temps dispose d'une déviance expliquée de 30 %. En revanche, dans le cas classique, la différence des déviances expliquées est beaucoup plus importante puisqu'elle passe de 27,60 % à 1,35 %. Dans ce cas, il est d'ailleurs clair que le temps capte la majorité de la déviance expliquée des hospitalisations puisque son edf est de 7,72, ce qui indique une courbe complexe qui tend à s'adapter aux données. Ainsi, les précipitations sur 7 jours ne semblent pas avoir d'influence notable sur l'ensemble des hospitalisations tandis qu'elles peuvent en avoir dans le cas des pics.

La Figure II-11 présente la courbe des effets des précipitations moyennées sur 7 jours avec et sans le temps, ainsi que les pics observés et les estimations du modèle. La courbe est croissante monotone, indiquant une augmentation (légère) des pics en fonction de la moyenne des précipitations sur la dernière semaine. Son allure est moins complexe que celle du cas des hospitalisations à Montréal (Figure II-7), conséquence d'un edf égal à 1.08. Lorsque le temps n'est pas inclus, la fonction est complètement linéaire (edf de 1), bien que toujours croissante. La Figure II-12 montre les mêmes courbes, sur l'ensemble des observations cette fois-ci. Le même phénomène se produit : la fonction est croissante et devient linéaire lorsque la variable du temps

est retirée. La tendance est cependant inversée (ou décroissante) dans le cas classique comparativement au cas des pics, même si les deux coefficients sont proches de 0 (coefficient directeur de la droite de $-5e-3$ et $8e-3$, respectivement).

À la Figure II-11, le pic le plus élevé (67 hospitalisations) est présent à droite et influence la courbe en augmentant l'intervalle de confiance, puisque peu d'observations sont présentes dans cette partie du graphique. Il est aussi associé à de fortes moyennes de précipitations et son estimation est la moins bonne du modèle. Comme pour le cas précédent à Montréal, il est légitime de s'intéresser à cette observation. Ce pic est analysé plus en détail à l'aide des conditions météorologiques qui ont eu lieu dans les jours précédents. Ces dernières sont présentées à la Figure II-13, brutes et moyennées sur 7 jours. Le pic de 67 hospitalisations a lieu le 25 janvier 2006. La température maximale a connu une chute importante, puis une hausse au niveau initial trois jours avant (de $-0,3$ °C à $-9,45$ °C puis à $-0,93$ °C). Même si la température n'a pas été retenue comme variable explicative dans le modèle (pour la série de pics), elle peut être significative pour le pic en question. Par ailleurs, le 25 janvier, l'humidité a atteint un maximum local tandis que la pression, après une chute de 103,8 kPa à 100,94 kPa, a atteint un minimum local. La température et l'humidité moyennées sur 7 jours montrent des niveaux importants quelques jours avant le pic. Concernant les précipitations, l'une des variables retenues dans le modèle, le pic survient un jour où ces dernières sont importantes : 4,66 mm (équivalent au quantile à 78 % des précipitations). Par ailleurs, 4 et 7 jours avant le pic, le 18 et 21 janvier, avaient lieu des précipitations plus élevées encore (16,16 mm et 47,75 mm respectivement, Figure II-13). Ainsi, la quantité de neige au sol (l'autre variable explicative retenue dans le modèle) a augmenté graduellement durant les 10 jours précédant le pic, comme cela est visible sur les deux graphiques (quantité de neige brute et moyennée). Comme dans le cas des

hospitalisations à Montréal, cette combinaison unique des variables météorologiques a probablement eu un effet physiologique sur la santé cardiovasculaire. En particulier, les variations de la température ont un effet reconnu sur les MCV. Le pic de 67 hospitalisations est donc conservé dans l'analyse.

3.3.2 Pics d'hospitalisations à Québec – fins de semaine

La taille d'échantillon des pics d'hospitalisations en fins de semaine est de 115. Plus de pics sont ainsi à disposition comparativement à ceux disponibles pour les jours de semaine, conséquence du plus bas seuil de 85 % sélectionné par Chiu *et al.* (2015).

Les conditions météorologiques associées aux pics retenus sont tracées à la Figure II-14 et résumées de façon numérique au Tableau II-8. Les relations graphiques ne sont pas claires. À noter, aucun pic ne survient entre 0 et 19 cm de neige au sol exclusivement et la plupart ont lieu dans les valeurs élevées. Au contraire, les faibles précipitations (inférieures à 10 mm) semblent entraîner plus de pics que les fortes (supérieures à 10 mm). Pour les deux variables (quantité de neige au sol et précipitations), il faut garder en tête que beaucoup de pics surviennent aux valeurs nulles. La pression atmosphérique semble posséder une zone de moindre effet aux alentours de 101 kPa.

Un GAM est retenu, comprenant la température lissée et moyennée sur 3 jours. Les estimations sont résumées au Tableau II-9. Le modèle des pics avec le temps possède une DEX de 19,1%, pour un edf associé de 2,49 (Tableau II-9). Cela reste relativement élevé pour un modèle contenant une seule variable météorologique, bien que la déviance diminue à 13,70 % lorsque le temps est retiré. La fonction de la température essaye alors de s'adapter légèrement plus aux données avec un edf de 2,73 cette fois. L'AIC est néanmoins plus faible pour le modèle avec le

temps. Le critère de GCV est lui supérieur, mais la différence est minime, de l'ordre de $5e-4$. Le cas de l'ensemble des données est différent : le modèle avec le temps passe d'une déviance de 9,25 % à 1,69 % pour le modèle sans le temps. Une telle différence est prévisible dans la mesure où, comme expliqué dans la section 3.2, beaucoup d'autres facteurs (la pollution de l'air et les facteurs individuels entre autres) peuvent influencer la survenue de l'hospitalisation lorsque l'intérêt est porté sur toutes les observations. De plus, dans ce cas, l'effet devient quasiment linéaire (l'edf passe de 1,20 à 1,09) contrairement au cas des pics (2,49 à 2,73).

La fonction des effets de la température maximale moyennée sur 3 jours possède un edf de 2,49 (Figure II-15). Elle présente une forme de sigmoïde inversée et aplatie. Puisqu'elle est décroissante, cela indique un effet plus prononcé des températures froides sur les pics de MCV. Chiu *et al.* (2015) ont par ailleurs considéré une séparation en été et hiver afin d'homogénéiser les pics (la séparation en jours de semaines et fins de semaine est cependant jugée la meilleure et utilisée ici). La fonction de la température dans le modèle des pics sans le temps possède une allure similaire et donc une interprétation similaire. Ces fonctions sont à mettre en comparaison avec celles de la Figure II-16 sur l'ensemble des observations, toujours avec les fonctions de la température et du temps. Elles sont toujours décroissantes, mais quasiment linéaires comme noté précédemment. Une différence est tout de même notable : entre 0 °C et 15 °C, l'effet de la température moyennée sur 3 jours est constant sur les pics (avec et sans la fonction du temps) et diminue après 15 °C. En revanche, dans le cas de l'ensemble des observations, la relation est décroissante, dû à la forme de sigmoïde inversée. Cette zone de moindre effet après 15 °C a aussi été observée dans le cas des hospitalisations à Montréal.

3.4 Modélisation des pics de décès en fonction de la météo à Montréal

L'extraction des pics de décès à Montréal s'est faite via la méthode du seuil, fixé à 99% (Chiu *et al.*, 2015). Comme pour le cas des pics d'hospitalisations à Québec, la séparation en jours de semaine et fins de semaine a été nécessaire pour respecter les hypothèses de modélisation de l'EVT (lundi à jeudi pour les jours de semaine contre vendredi à dimanche pour les fins de semaine). Deux GAM complémentaires sont donc obtenus. Les résultats sont décrits ci-après.

3.4.1 Pics de décès à Montréal – jours de semaine

La série des pics de décès est constituée de 42 pics. La Figure II-17 présente les conditions météorologiques associées à ces pics tandis que le Tableau II-10 résume l'information numérique. Contrairement aux cas des hospitalisations (Montréal et Québec), ici les températures froides jouent un rôle plus important. La majeure partie des pics est en effet répartie sous 0 °C (Figure II-17) et la moyenne est à -3 °C (Tableau II-10). Par ailleurs, les valeurs de la température ne sont pas nécessairement extrêmes. De même, peu de pics surviennent entre 0 et 10 cm de neige, autour de 70 % d'humidité et après 15 mm de précipitations.

Deux modèles de régression sont retenus : 1) l'un implique les précipitations lissées du jour même; et 2) l'autre comprend l'humidité lissée moyennée sur 7 jours. Ces deux variables ne montraient pas de liens avec les décès à la Figure II-17. Les estimations sont résumées au Tableau II-11 et au Tableau II-12, pour les décès à Montréal en jours de semaine. La déviance expliquée reste relativement élevée pour des modèles contenant une variable explicative : 38 % et 30,40 % respectivement pour le modèles 1 et le modèle 2. Ces déviances ne sont probablement pas dues à l'inclusion de la variable temporelle, puisque sans elle, les déviances passent

respectivement à 36,80 % et 30,10 %, qui sont très similaires. Par contre, les résultats montrent que dans le cas classique (mêmes variables explicatives), la variable du temps capte la majeure partie de la déviance : les déviances expliquées passent respectivement de 18,30 % à $8e-7$ % et de 22,20 % à 3,65 %.

La fonction des précipitations pour le premier modèle est présentée à la Figure II-18. Elle est croissante et monotone avec un edf de 1,63. Il est difficile de se fier aux précipitations puisque presque toutes les observations semblent comprises entre 0 et 10 mm. Deux observations se situent au-delà, se traduisant par un élargissement visible de l'intervalle de confiance. Cet élargissement pointe vers des précautions à prendre sur la forme de la fonction. Dans le cas classique, les fonctions avec et sans le temps deviennent des droites (Figure II-19). De plus, la fonction du temps possède un edf de 7,15, indiquant une courbe complexe qui capte la majorité de la déviance comme noté précédemment. Pour le modèle 2, la fonction de l'humidité moyennée sur 7 jours (edf de 2,26) est présentée à la Figure II-20. Elle présente une forme de « J » aplati. L'effet est constant jusqu'à 78 % d'humidité, puis l'augmentation des pics semblent avoir lieu dans les valeurs tout de suite supérieures. Une forte humidité peut être un signe d'orage en été ou de tempête verglaçante en hiver, entre autres. Les mêmes courbes concernant l'ensemble des observations (Figure II-21) semblent plus « lisses » du fait de l'étendue temporelle beaucoup plus importante que celle des pics, mais elles possèdent des edf plus élevés (5,34 avec le temps et 5,13 sans le temps). La fonction du temps capte la tendance générale (et l'essentiel de la déviance expliquée, comme noté précédemment), avec une légère augmentation suivie d'une légère diminution autour de l'année 1996.

L'AIC semble favoriser le premier modèle avec les précipitations (AIC de -106,05 contre -99,94 pour le deuxième modèle). Cependant, il est difficile d'accorder de la crédibilité à la fonction de

lissage associée aux précipitations puisqu'elle est influencée par deux observations. Il est alors possible de se tourner vers le deuxième modèle, d'autant plus qu'il est préféré par le critère GCV ($4,64e-3$ pour le premier modèle, Tableau II-11, contre $5,46e-3$ pour le deuxième modèle, Tableau II-12).

3.4.2 Pics de décès à Montréal – fins de semaine

La taille d'échantillons est de 32 pics de décès. La météo associée aux jours de survenue des décès est présentée à la Figure II-22 et le Tableau II-13 présente un résumé numérique. Comme pour le modèle complémentaire en jours de semaine, l'effet des températures froides est constaté et seulement 4 pics ont lieu au-dessus de 5°C . Lorsque la température est élevée, il y a tout de même des conséquences puisque le pic le plus élevé est observé le 19 juin 1994, avec une température de 27°C (pour rappel, le quantile à 90 % est de $25,4^{\circ}\text{C}$, voir Tableau II-3). Concernant les précipitations, beaucoup de pics ont lieu dans les valeurs basses. Les autres variables ne présentent pas de relation particulière.

Après sélection des variables significatives et des lags associés, deux modèles sont retenus :

- 1) La température lissée et moyennée sur 2 jours;
- 2) La température lissée et moyennée sur 2 jours accompagnée des précipitations moyennées sur 1 jour précédant l'évènement.

Les résultats de l'estimation sont présentés au Tableau II-14 et au Tableau II-15. Les déviations expliquées du modèle 1 et du modèle 2 sont respectivement de 53,6% et de 59,80 %. Elles augmentent à 66 % et 69,5 % lorsque la fonction du temps est supprimée. Cette augmentation est à prendre avec précaution et n'implique pas que les modèles sont meilleurs. En effet, les edf augmentent aussi, passant respectivement de 3,11 à 6,17 et de 3,06 à 5,60. Cela indique que la

fonction de la température s'adapte plus aux données lorsque le temps est retiré, expliquant l'augmentation de DEX. Dans ce cas précis, il est possible que le paramètre de pondération de la GCV (section 2) n'ait pas été suffisant pour éviter un surajustement. Il n'y a toutefois pas d'augmentation dans les modèles concernant l'ensemble des données, les déviations expliquées sont moindres sans le temps. Elles restent malgré tout relativement élevées, de l'ordre de 10 % (autour de 27 % avec le temps), confirmant le rôle important de la température autant dans le cas des pics que dans le cas classique, pour les décès à Montréal en fins de semaine.

La Figure II-23 présente la fonction de la température du modèle 1, fonction en forme de « J » (edf de 3,06). La courbe présente un creux autour de 10 °C. Elle remonte ensuite pour atteindre son maximum. Cela indiquerait une sensibilité aux températures très froides et une plus grande sensibilité encore aux températures très chaudes, avec une zone de moindre effet entre les deux. Cette zone atteindrait son maximum (donc le minimum de l'effet de la température sur les pics) autour de la température déjà mentionnée de 10 °C. Lorsque le temps est retiré, la fonction est moins lissée (comme mentionné, l'edf passe de 3,11 à 6,17) mais l'allure générale reste la même. La Figure II-24 montre la fonction des températures avec et sans le temps pour l'ensemble des observations. La courbe avec le temps est similaire à celle des pics, à la différence que le creux se situe plutôt autour de 25 °C. La Figure II-25 présente la fonction de la température du modèle 2. Elle est similaire à la Figure II-23 et l'interprétation est donc semblable. Les mêmes courbes sont présentes à la Figure II-26 pour l'ensemble des données. À nouveau, les courbes et les interprétations sont très semblables à ce qui est illustré sur la Figure II-24.

Dans les deux modèles, un pic est présent à l'extrémité droite de la courbe et semble la tirer vers le haut, lui donnant son allure de « J ». Les conditions météorologiques qui ont mené au pic sont analysées grâce à la Figure II-27. Ce dernier a eu lieu à la date mentionnée plus tôt, le 19 juin

1994 (53 décès). La veille, la température a atteint un maximum local après une augmentation progressive sur 11 jours : de 19,35 °C à 34 °C, cette dernière valeur étant une température élevée pour la ville de Montréal (voir Figure II-3). Les précipitations et l'humidité étaient aussi relativement élevées les jours précédents. La quantité de neige était nulle sur la période car la température du mois de juin ne le permet habituellement pas. Il s'agit en réalité d'un pic survenu à la suite d'une canicule. Au sens d'Environnement Canada, une canicule se définit comme une succession de trois jours ou plus où la température égale ou dépasse 32 °C, ce qui a été le cas pour le 16, 17 et 18 juin 1994, alors qu'elle était sous la limite le 15 juin (29,32 °C). Cela va dans le sens du lag optimal choisi pour ce modèle (moyenne de 2 jours avant plus le jour même). Les accumulations de température, d'humidité et de pression sont visibles à la Figure II-27 avec les variables moyennées sur 2 jours. En particulier, le pic survient le lendemain du maximum des températures moyennées et quelques jours après le maximum de l'humidité et de la pression, toutes deux moyennées. Les précipitations ne présentent pas de conditions particulières, qu'elles soient brutes ou moyennées. Le pic est ainsi conservé dans l'analyse.

Parmi les deux modèles, la comparaison des critères AIC et GCV n'est pas unanime. L'AIC recommande le modèle 2 tandis que la GCV recommande le modèle 1. Le modèle 2 étant emboîté dans le modèle 1 (seule l'inclusion des précipitations les distingue), il est possible de passer par une procédure d'ANOVA (voir la section 2) pour déterminer si les précipitations sont nécessaires ou non. Le test du χ^2 donne une p-value légèrement supérieure à 0,05, ce qui signifie que la différence entre les deux modèles n'est pas significative. Le choix final se porte donc sur le modèle 1, sans les précipitations.

3.5 Modélisation des pics de décès en fonction de la météo à Québec

Les résultats concernant l'extraction des pics de décès à Québec présentés dans Chiu *et al.* (2015) permettent de retenir la méthode des blocs, avec une taille de 180 jours. Il en résulte une taille d'échantillon de 63 pics de décès. Les statistiques descriptives des conditions météorologiques associées à ces pics sont résumées au Tableau II-16 et présentées à la Figure II-28. Le nombre limité de pics autour de 10 °C, de 70 % et de 10 cm semble indiquer une zone de moindre effet pour la température, l'humidité et la quantité de neige au sol, respectivement. De plus, la plupart des pics surviennent dans des valeurs non nécessairement extrêmes de ces variables. Il y a quelques exceptions, comme la quantité importante de neige à l'hiver en 1997 et ou les températures froides de l'hiver en 20014 qui sont accompagnées d'un pic de décès.

Trois GAM sont retenus pour les décès à Québec. Les résultats sont présentés au Tableau II-17, Tableau II-18 et Tableau II-19 respectivement pour :

- 1) Le modèle avec la température du jour même lissée;
- 2) Le modèle avec la température lissée et moyennée sur 1 jour précédant l'évènement;
- 3) Le modèle avec la quantité de neige au sol lissée et moyennée sur 4 jours.

Le modèle 1 possède une déviance expliquée de 40,5 %. Celle-ci retombe cependant à 15,90 % sans la fonction du temps. Cette DEX reste malgré tout une valeur relativement élevée, d'autant plus que le modèle ne contient que la température du jour. Les déviances sont largement moindres dans le cas de l'ensemble des observations, passant de 4,95 % à 2,10 %. La courbe d'effet de la température sur les pics de décès est présentée à la Figure II-29, avec un edf de 2,80. Elle présente une allure de sigmoïde inversée et aplatie. Ceci indique que les températures froides ont légèrement plus d'influence que les températures chaudes sur les décès, avec un maximum

autour de $-10\text{ }^{\circ}\text{C}$. La fonction conserve sa forme lorsque la variable temps est retirée (edf égal, 2,80). À noter, la présence de la zone de moindre effet aux alentours de $10\text{-}15\text{ }^{\circ}\text{C}$. La Figure II-30 présente les mêmes courbes pour l'ensemble des observations. Bien que plus aplatie, elle est d'allure globalement similaire à celle du cas des pics, impliquant les mêmes interprétations. Les edf sont d'ailleurs comparables aux deux précédents : 3,02 et 2,49 avec et sans le temps, respectivement. La zone de moindre effet se situe également autour de $10\text{-}15\text{ }^{\circ}\text{C}$.

Le modèle 2 (fonction de la température moyennée sur 1 jour avant l'évènement) possède une déviance expliquée légèrement inférieure au modèle 1, de 35,8 % tombant à 14,7 % sans le temps, phénomène similaire au modèle 1. Dans le cas de l'ensemble des observations, la déviance passe de 5,03 à 2,16 %. La fonction de la température est présentée à la Figure II-31 (edf de 2,2). Similairement à la Figure II-29, elle présente une allure de sigmoïdale inversée. L'interprétation est donc semblable, à la différence que la courbe remonte moins dans les plus hautes températures. Dans le cas précédent, la température au jour même avait donc une influence quasiment équivalente qu'elle soit froide ou chaude. En revanche, dans le cas des températures moyennées sur 1 jour (avant l'évènement sanitaire), il semble que les températures froides influent légèrement plus sur les pics de décès. La Figure II-32 présente quant à elle les courbes de la température pour l'ensemble des observations. Les fonctions sont une fois de plus semblables à celles de la Figure II-30 (l'ensemble des observations pour le modèle 1).

Le modèle 3 (fonction de la quantité de neige au sol moyennée sur 4 jours) possède une DEX de 37,5 %, qui retombe à 19,10 % sans la variable du temps. Lorsque toutes les observations sont considérées, elle passe de 5,12 % à 2,32 %. Les edf sont néanmoins élevés dans ce derniers cas (7,28 et 6,82, la fonction du temps possédant un edf de 5,08). Même si ce modèle fait preuve d'une déviance expliquée supérieure à celle des modèles 1 et 2, il n'est sans doute pas à

considérer car il n'est adapté qu'à une partie de l'année. En effet, la seule variable explicative étant la neige, les prédictions sont impossibles en dehors de l'hiver. Beaucoup d'observations sont d'ailleurs associées à 0 cm de neige, comme le montre la Figure II-33 (edf de 1,93). Une seule observation est présente après 100 cm de neige, ce qui a pour effet de grandement élargir l'intervalle de confiance. En dehors de ce pic de décès, la relation semble être croissante. Cela peut se traduire par le fait qu'il n'y a aucun stress particulier d'un point de vue de la santé par MCV lorsque la neige ne tombe pas. En revanche, lorsque la neige commence à tomber dans une certaine quantité, les efforts physiques augmentent pour se déplacer, entraînant un stress physiologique. Il est à noter que la fonction du modèle sans le temps présente une diminution après 60 cm, contrairement au modèle avec le temps. Les mêmes courbes sont visibles à la Figure II-34, pour l'ensemble des observations. Leurs allures sont semblables à celles du modèle des pics sans le temps. Ainsi, l'effet maximal de la quantité de neige au sol est atteint en deux points, autour de 70 cm et 150 cm.

3.6 Résumé des principaux résultats

Les principaux résultats sont résumés au Tableau II-20. Les déviances expliquées sont relativement élevées (de 19,10 % à 59,80 %) par rapport aux cas classiques, en particulier pour des modèles contenant une seule variable explicative, indépendamment du type de sélection pour extraire les pics. Tous les effets sont non linéaires, ce qui contraste avec le cas classique où beaucoup d'études considèrent des relations linéaires. La température est la variable la plus présente pour les hospitalisations à Montréal et à Québec (fins de semaine) ainsi que pour les décès à Montréal (fins de semaine) et à Québec. Les lags retenus pour cette variable sont 1, 2, 3 et 5 jours, soit moins d'une semaine. Au contraire, l'humidité semble avoir un effet plus long dans le temps puisqu'elle est présente dans le modèle des décès à Montréal en jours de semaine

au lag 7. L'influence des précipitations dans le temps varie puisqu'elles sont retenues pour les décès à Montréal en jours et fins de semaine au jour même et au lag 1 respectivement, ainsi que pour les hospitalisations à Québec en jours de semaine au lag 7. La quantité de neige apparaît dans les modèles d'hospitalisations en jours de semaine et pour les décès à Québec (linéaire au jour même et non linéaire au lag 4, respectivement). Par ailleurs, la température et la pression atmosphérique semblent présenter une zone de moindre effet pour les pics d'hospitalisations ainsi que de décès autour de 10-15 °C et de 101 kPa (la pression d'une atmosphère sans dépression ni anticyclone), respectivement.

4 Discussion

Force est de rappeler que dans cette section, il est difficile de se comparer aux études déjà effectuées sur les relations entre la santé et la météo, dans la mesure où elles se basent sur l'ensemble des observations et non juste sur les pics. Cependant, cette étude étant la première à se baser sur des pics sanitaires extraits de façon statistique, elle forme une étape nécessaire mais la comparaison avec les études classiques doit se faire avec précaution.

4.1 Température

La variable météorologique la plus souvent significative dans les modèles est la température, bien qu'elle ne soit pas forcément présente dans tous les modèles (6 modèles sur 10). Cela semble concorder avec les multiples études dans le cas classique qui indiquent que la température est un facteur majeur dans les hospitalisations et décès pour cause de MCV (voir p. ex. Schwartz *et al.*, 2004; Knowlton *et al.*, 2009). Par ailleurs, Martens (1998) montre la forme de « V » ou de « J » de la relation entre la température et la mortalité par MCV et maladies respiratoires. Ce genre de résultats est obtenu dans le cas classique mais le cas des pics est plus complexe car si l'aspect non

linéaire est toujours présent, les formes diffèrent d'un modèle à l'autre. P. ex., pour les décès à Montréal (fins de semaine) la relation est en forme de « J », tandis que pour les hospitalisations à Montréal la fonction est plutôt en forme de « J » inversé. Pour les décès à Québec, la relation prend une forme de sigmoïde (« S ») inversée et aplatie. Soit autant de relations non linéaires qui auraient été mal estimées par une régression linéaire, lorsque la série des données est considérée globalement.

La température est en particulier présente dans les modèles de fins de semaine. Elle pourrait en partie justifier la séparation des modèles entre jours et fins de semaine, retenue par Chiu *et al.* (2015). En effet, il y a une différence significative de cette variable entre les jours de semaine et les fins de semaine (Beaney et Gough, 2002). Les résultats obtenus ici montrent qu'elle pourrait avoir plus d'influence les fins de semaine, comparativement aux jours de semaine.

Dans cette étude, la zone de moindre effet pour les décès et les hospitalisations a été plusieurs fois observée, aux alentours de 10-15 °C. De plus, les hospitalisations semblent davantage influencées par les températures chaudes, tandis que les décès apparaissent plus intimement liés aux températures froides. Par exemple, pour les décès à Québec, la température d'influence maximale se situe autour de -10 °C (moyenne du jour même et de la veille).

4.2 Humidité

La variable d'humidité a été significative dans un seul modèle, celui des décès à Montréal en jours de semaine. Il semble y avoir une relation croissante entre les pics et l'humidité moyennée sur une semaine. Lee *et al.* (2014) expliquent que l'humidité joue potentiellement un rôle sur les décès, dans 148 villes aux États-Unis. Le rôle de l'humidité serait lié à celui de la température dans l'adaptation aux changements climatiques. Ce résultat diffère de la conclusion de l'étude de

Braga *et al.* (2002), qui, dans une analyse temporelle de 12 villes situées aux États-Unis, n'ont pas réussi à montrer d'effet clair de l'humidité sur les décès par MCV. De la même façon, Schwartz *et al.* (2004) n'observent pas de relation entre l'humidité et les hospitalisations par MCV dans les 12 mêmes villes. Il est cependant difficile de tirer de comparaison définitive puisque dans la présente étude, un seul modèle comprend l'humidité et encore une fois, les trois études américaines ne se situent pas dans un contexte de pics. Cela appuierait la nécessité de considérer les pics en tant que tels dans d'autres études afin de clarifier cette relation.

Il est utile de mentionner que l'humidité à 6 h (maximale) semble plus représentative du risque de MCV. En effet, les maladies coronaires et la pression artérielle ont tendance à survenir plus souvent en matinée, entre 6 h et midi (Weber et Fodera, 2004; Carter *et al.*, 2016). Ainsi, la mesure de l'humidité à 6 h est plus intéressante pour expliquer la survenue d'une MCV en matinée, par rapport à la mesure à 15 h. Cependant, une humidité élevée à 15 h et qui le reste jusqu'au lendemain conjointement à une température élevée pourrait effectivement aider à expliquer la survenue d'une MCV. Toutefois, les données présentes ne permettent pas d'explorer ce niveau de détails.

4.3 Pression atmosphérique

La pression atmosphérique a été retenue dans un seul modèle, celui des hospitalisations à Montréal. Elle est de plus introduite de façon linéaire. Par contre, Abrignani *et al.* (2012) n'observent pas de relation entre la pression atmosphérique et l'infarctus du myocarde en Sicile. Danet *et al.* (1999) montrent quant à eux une relation entre la pression atmosphérique et l'infarctus du myocarde dans le nord de la France, cette fois-ci en forme de « V ». Le climat dans ces deux études est néanmoins différent de celui à Montréal ou à Québec (continental humide),

puisque'il s'agit dans le premier cas d'un climat tempéré océanique, dans le deuxième cas d'un climat méditerranéen, deux types de climats plus doux. Dans cette étude, la pression semble posséder une zone de moindre effet autour de 101 kPa, qui est la valeur de pression d'une atmosphère normale (pas de front météorologique particulier).

4.4 Précipitations

Les précipitations ont été significatives dans deux modèles en jours de semaine indépendamment du type d'extraction utilisée, soit les décès à Montréal et les hospitalisations à Québec. La relation avec les pics est croissante dans les deux cas. Peu d'études montrent le rôle des précipitations dans les décès et les hospitalisations par MCV dans le cas classique. Baker-Blocker (1982) indique que les précipitations sont un facteur complexe à mettre en relation avec les MCV, dans une étude réalisée à Minneapolis (États-Unis) entre la mortalité par MCV et les précipitations, la température et la quantité de neige. Cette auteure précise tout de même que les précipitations contribuent à la mortalité par MCV, en particulier en hiver et combiné à de la neige car elles pourraient être un indicateur de fortes tempêtes (verglaçantes notamment) qui amènent des conditions extérieures difficiles.

4.5 Quantité de neige au sol

La quantité de neige au sol est présente dans deux modèles à Québec, pour les hospitalisations en jours de semaine et les décès. Ces modèles sont toutefois difficiles à interpréter. En effet, la quantité de neige au sol n'est crédible que pour la saison hivernale. Selon Baker-Blocker (1982), une combinaison de pluie et de neige peut grandement augmenter les chances de mortalité par MCV. Cela concorde avec le fait qu'à Québec (en jours de semaine), ce sont justement les deux variables qui ont été retenues dans le modèle, à la différence qu'il s'agit cette fois des

hospitalisations par MCV. Une des raisons pour laquelle la neige est présente uniquement dans les modèles à Québec pourrait être qu'il y a plus de neige qu'à Montréal (voir la section 3.1.2). Par ailleurs, la relation entre la quantité de neige au sol et les décès semble être croissante, à Québec. Cela indiquerait son influence dans le cas de chutes de neige importantes. P. ex., le pelletage en hiver est un effort qui a un effet sur les MCV, voir Franklin *et al.* (1995). Pour les pics de décès à Québec, une décroissance de l'effet est observé après un certain niveau élevé. Cela pourrait s'expliquer par une grande quantité de neige au sol qui équivaut à des chutes de neige importantes et une visibilité réduite, ce qui découragerait éventuellement les gens à sortir et à s'exposer. Martel *et al.* (2010), dans une étude sur la fracture de la hanche à Québec, concluent que les personnes âgées de 75 ans en particulier et plus auraient tendance à éviter le mauvais temps. Barnett *et al.* (2012) suggèrent la même hypothèse, en analysant les décès lors des canicules et des épisodes de grand froid dans 99 villes aux États-Unis.

4.6 Considérations statistiques

La sélection des lags s'est faite sur la minimisation de l'AIC, comme dans les études de Curriero *et al.* (2002) et de Lanzinger *et al.* (2014). Ceci est motivé par le fait que la majorité des effets de la météo peut être captée en tenant compte de l'information cumulative des jours précédents. Par exemple, Hong *et al.* (2003) estiment que la réponse la plus forte par rapport à la température pour les accidents cérébraux vasculaires à Incheon (Corée du Sud) se fait sur 1 à 2 jours. Cela concorde avec Madrigano *et al.* (2013) qui montrent que sur la mortalité par infarctus du myocarde à Worcester (Angleterre), la température d'il y a 2 jours est significative. Certains auteurs vont plus loin avant l'évènement sanitaire comme Goldberg *et al.* (2011) qui montrent que pour la mortalité toutes causes à Montréal (Canada), 5 jours sont nécessaires. De plus, l'effet découvert par les auteurs est non linéaire. Dans la présente étude, les lags sont conservés

majoritairement entre 0 et 4 jours (deux cas d'un lag à 7 jours), ce qui concorde avec les autres études des cas classiques. Il est donc intéressant de constater que du point de vue du temps de réponse, les pics par MCV ne sont pas différents des cas classiques.

L'effet de la météo sur les MCV est reconnu. Toutefois, toutes les conditions météorologiques associées aux pics ne sont pas nécessairement « extrêmes ». Ceci est vrai pour les pics sélectionnés dans cette étude. En particulier les variations des températures ou de la pression atmosphérique peuvent déclencher des pics (Ebi *et al.*, 2004; Houck *et al.*, 2005). L'étude de Turner *et al.* (2007) montre qu'une combinaison de température minimale, d'humidité relative maximale et de taux de particules fines journalières (PM₁₀) a engendré un pic de morbidité par MCV à Sydney (Australie), en 2005. Cela indique un intérêt pour une étude symétrique à celle présente, c'est-à-dire centrée sur les conditions météorologiques extrêmes de sorte à analyser leurs relations avec les pics par MCV. Dans cette étude, alors qu'une valeur extrême d'une variable météorologique ne semble pas avoir systématiquement de conséquence d'un point de vue des pics, des combinaisons particulières de plusieurs variables météorologiques semblent expliquer les pics les plus grands (hospitalisations ou décès). Un raffinement possible serait alors de considérer des indices climatiques, qui sont construits en fonction de plusieurs variables météorologiques. Cependant, les indices climatiques comportent une part de subjectivité. De plus, force est de rappeler que cette étude a aussi un but exploratoire et qu'il convient d'analyser en premier lieu les relations avec les variables météorologiques seules.

Les situations de surajustement des fonctions des variables explicatives ont été évitées en suivant l'approche de Kim et Gu (2004). Dans le cas contraire, sans paramètre de correction, des edf supérieurs à 8 ont été constatés (non présentés) mais les fonctions étaient complexes car trop adaptées aux données. Dans les modèles retenus, aucun edf ne dépasse 4,11, ce qui a permis

d'éviter le surajustement – un aspect important à retenir dans une optique de prédiction (Hawkins, 2004). Une fonction trop complexe ne permettra pas d'obtenir des prédictions fiables, car trop adaptée aux données présentes. Pour illustrer ce point, le modèle des décès à Montréal en fins de semaine avec la température lissée et moyennée sur 2 jours est présenté à la Figure II-35, avec (comme conseillé par Kim et Gu, 2004) et sans paramètre de correction. L'edf est immédiatement moins limité et passe de 3,06 (avec) à 7,26 (sans). En conséquence, la fonction devient moins lisse (présence de variations supplémentaires) et s'adapte inutilement un peu plus aux données.

La plus grande différence entre le cas classique et les pics réside dans la déviance expliquée des modèles retenus. Dans une revue de littérature, Lipfert (1993) indique qu'une étude sur les causes météorologiques et de pollution pour les admissions hospitalières en Ontario (Canada) obtient typiquement des valeurs de variance expliquée de 1 à 5 %, ce qui est aussi obtenu dans les modèles sur l'ensemble des données dans cette étude. En revanche, les DEX des modèles sur les pics sont largement plus élevées. Une explication possible est que les pics forment un sous-ensemble des données sanitaires brutes, liées à un contexte plus spécifique, réduisant ainsi le nombre de causes impliquées. L'intérêt est donc porté sur une sous-partie de la déviance totale, où la météo prend une part importante. Il s'agit d'une information pertinente pour la recherche sur les MCV puisque cela indiquerait un besoin d'approfondir les connaissances sur les relations entre les pics et la météo, de sorte à mieux gérer ou prévenir la survenue de certains pics (p. ex., une meilleure précision dans les alertes sanitaires).

Finalement, les modèles retenus ici se basent sur les séries de pics extraites par Chiu *et al.* (2015). Puisque ces derniers sélectionnent les seuils et les tailles des blocs de manière automatique (en optimisant des critères statistiques), cela a pour conséquence des techniques d'extraction et des

tailles d'échantillons différentes. Un autre raffinement possible serait alors d'homogénéiser la technique d'extraction et de sélectionner des seuils ou tailles de blocs identiques pour les modèles divisés (comme les jours de semaine et fins de semaine). Cela nécessiterait cependant d'examiner les résultats de Chiu *et al.* (2015) au cas par cas, ce qui demanderait beaucoup plus de temps. De plus, des résultats préliminaires sur les pics d'hospitalisations à Montréal (non présentés) ne montrent pas de différences significatives en utilisant des seuils autres que celui retenu ici.

5 Conclusion

Cette étude a pour but de modéliser la relation entre les pics sanitaires (hospitalisations et décès à Québec et Montréal) et les variables météorologiques. Elle diffère des études classiques dans le sens où ce n'est pas l'ensemble des observations qui nous intéresse, mais uniquement les pics (sélectionnés par des outils statistiques appropriés). Ces derniers sont d'un intérêt particulier car ils provoquent un stress sur l'offre de services de santé. Il est ainsi important de comprendre leurs conditions de survenue et de mieux les caractériser.

Des modèles non paramétriques (GAM) ont été utilisés pour tenter d'expliquer les séries de pics en fonction des conditions météorologiques. Les séries de pics ont été obtenues et étudiées par Chiu *et al.* (2015). La méthodologie utilisée a pour but de ne pas imposer de forme linéaire aux relations, puisqu'il s'agit de la première étude tentant d'expliquer des pics sanitaires par la météo.

Les modèles explicatifs sur les pics sanitaires diffèrent en plusieurs points de ceux sur l'ensemble des données. P. ex., le nombre d'observations est très restreint dans le cas des pics (par définition de ces derniers), en plus de ne pas être à intervalles fixes. D'un point de vue statistique, les pics ne suivent pas les mêmes distributions que l'ensemble des observations, d'où leurs estimations erronées par les méthodes basées sur la moyenne.

Les résultats montrent que la température joue toujours un rôle important, comme dans le cas classique. D'autres variables sont aussi présentes, de façon plus marginale, comme les précipitations ou la quantité de neige au sol. Les relations sont majoritairement non linéaires (elles auraient ainsi été mal estimées par des régressions linéaires). Les déviations expliquées sont par ailleurs élevées par rapport aux déviations expliquées par l'ensemble des données, d'autant plus qu'il s'agit de modèles avec une seule variable météorologique de façon générale. Se restreindre ainsi à la sous-partie des observations brutes que sont les pics permet de diminuer le nombre d'indicateurs de risque possibles pour l'évènement sanitaire. L'ordre de la déviance expliquée montre que la météo joue un rôle important dans cette sous-partie, ce qui était alors méconnu ou peu investigué.

Cela dit, les deux types d'études (pics et classiques) devraient être réalisés de façon complémentaire afin de mieux préparer le système de la santé aux diverses éventualités engendrées par les changements climatiques. P. ex., les alertes sanitaires ciblées peuvent se baser sur les résultats obtenus sur les pics les plus importants qui sont survenus suite à une combinaison particulière des différentes variables météorologiques, plutôt que sur une variable. Il est aussi nécessaire de rester vigilant lors des périodes de « sensibilité », comme lorsque la température des 2 derniers jours est de $-10\text{ }^{\circ}\text{C}$ pour les décès à Québec.

6 Limitations

Puisque l'intérêt est porté sur les pics sanitaires dans cette étude, seuls de petits échantillons sont disponibles par définition. Cela implique immédiatement des restrictions sur les modèles. La plus importante est que le cadre non paramétrique (modèle additif généralisé) impose de ne considérer qu'une seule variable explicative lissée, en plus de la variable de temps. Les splines cubiques étant des fonctions qui nécessitent plusieurs paramètres à estimer par rapport à une droite, leur

utilisation ne permet pas l'inclusion de toutes les variables explicatives sous forme de fonctions, pour cause de problème d'identifiabilité du modèle (plus de paramètres à estimer que d'observations). La taille des blocs et le niveau des seuils sont choisis à l'aide d'outils statistiques par Chiu *et al.* (2015). De sorte à obtenir des échantillons plus importants, il est cependant possible de diminuer la taille des blocs ou le niveau des seuils, impliquant une balance à obtenir entre la taille des échantillons et la flexibilité du modèle (ou principe de parcimonie).

Puisque les données sanitaires sont un total d'évènements journaliers, il est impossible d'utiliser des variables individuelles et comportementales significativement associées au risque de MCV dans la littérature. Parmi ces variables, il est possible de citer la pratique d'un sport régulier et la consommation de tabac ou d'alcool. Le statut socioéconomique ou l'âge joue aussi un rôle prépondérant (Center for Disease Control Prevention, 2004).

Pour tenter d'expliquer les pics, seules des variables météorologiques ont été utilisées. La littérature classique leur applique souvent des opérations de lag et/ou de moyenne. Bien que la température soit dominante, l'utilisation de variables de pollution atmosphérique pourrait améliorer les modèles. Morris (2001) indique que la part de la pollution de l'air sur les décès par MCV est potentiellement majeure. En particulier, l'ozone joue un rôle non négligeable, en particulier lors des canicules, comme rapporté par Filleul *et al.* (2006), dans une étude sur la relation entre l'ozone, la température et la mortalité en France.

Figures et Tableaux

Figures

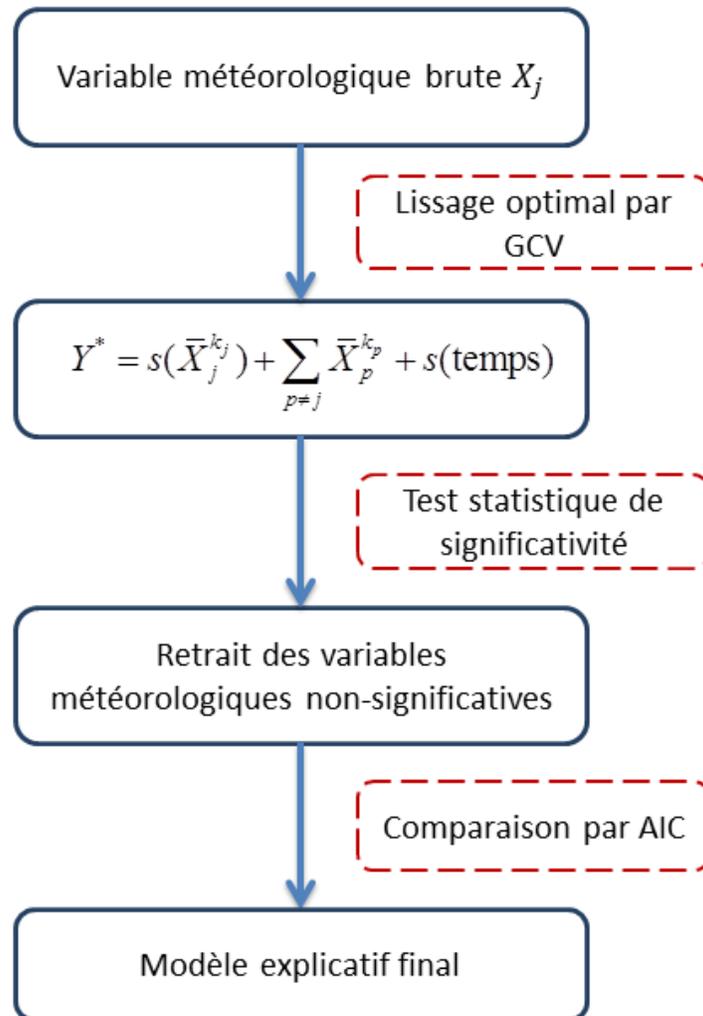


Figure II-1 : Méthodologie appliquée pour la modélisation de la relation santé-météo.

$\bar{X}_p^{(k)}$ représente la moyenne de la variable X_p sur les k derniers jours tandis que k_p représente le lag k optimal pour la variable X_p .

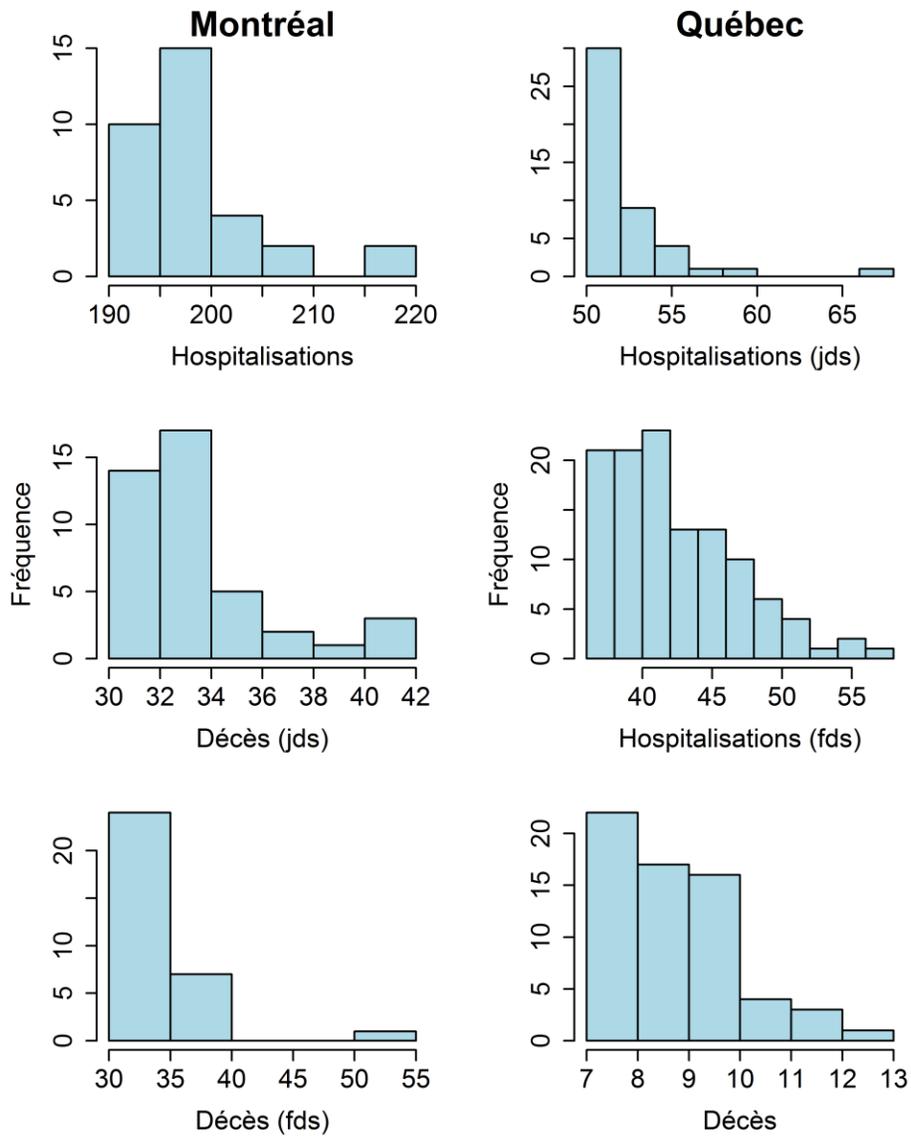


Figure II-2 : Histogrammes des pics sanitaires, à Montréal (gauche) et à Québec (droite).

Fds = fins de semaine, jds = jours de semaine

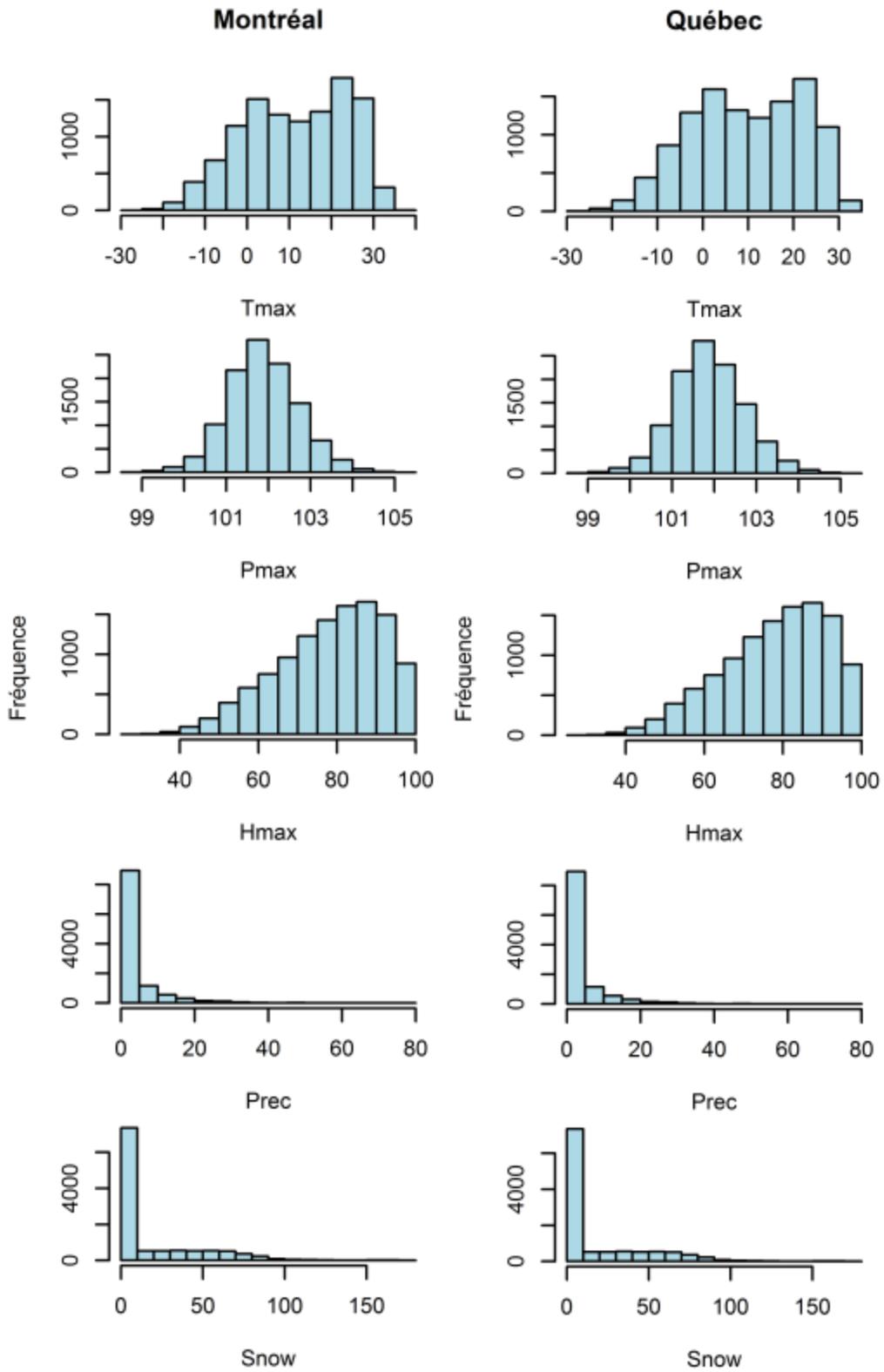


Figure II-3: Histogrammes des variables météorologiques considérées à Montréal (gauche) et Québec (droite).

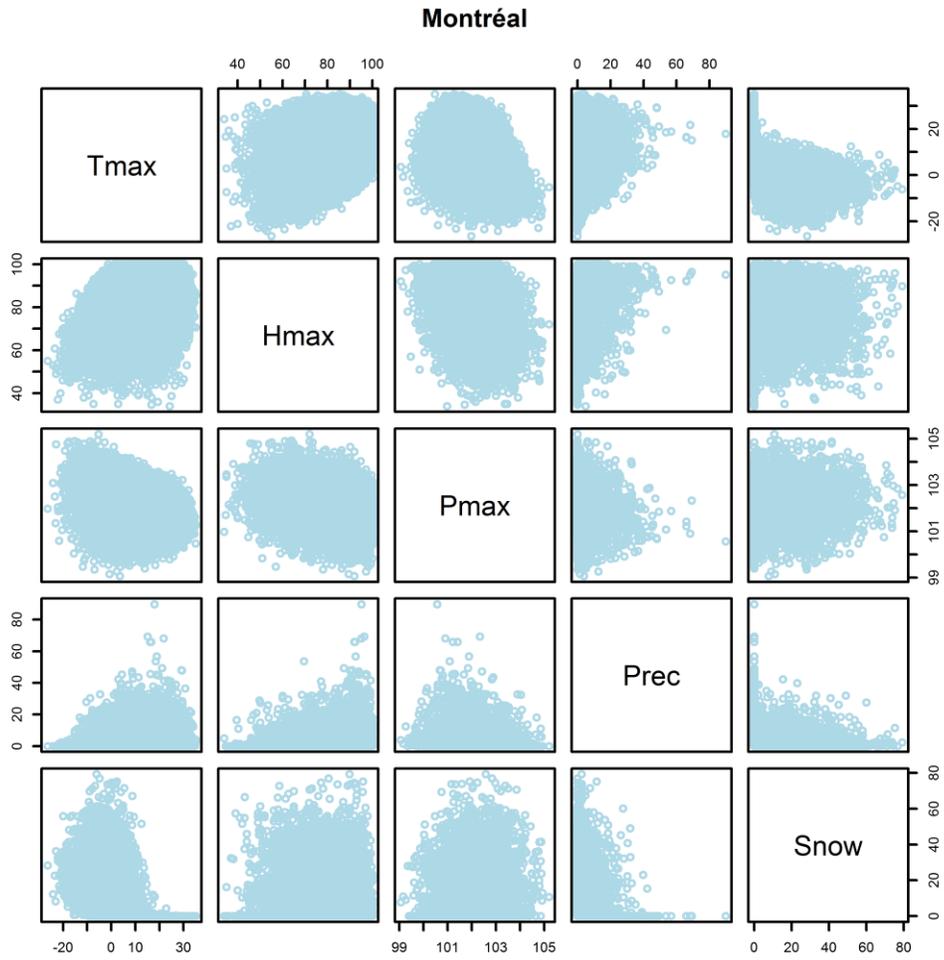


Figure II-4 : Relations entre les variables météorologiques considérées à Montréal.

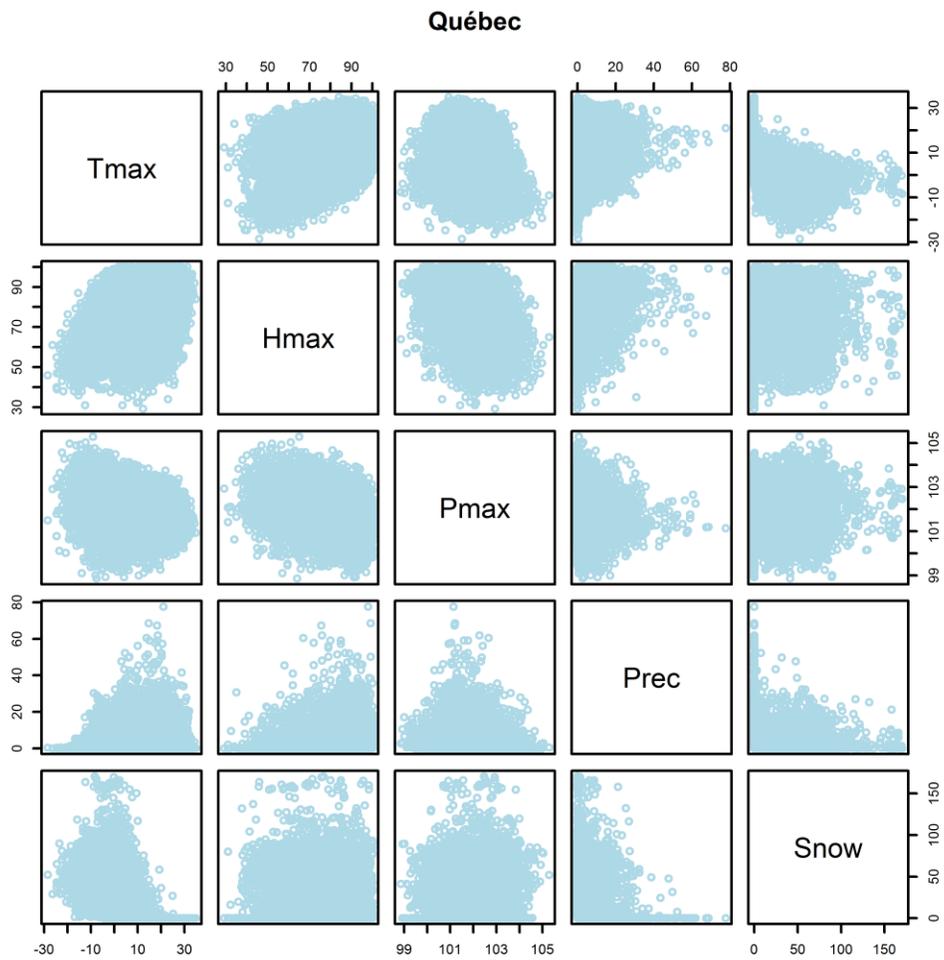


Figure II-5 : Relations entre les variables météorologiques considérées à Québec.

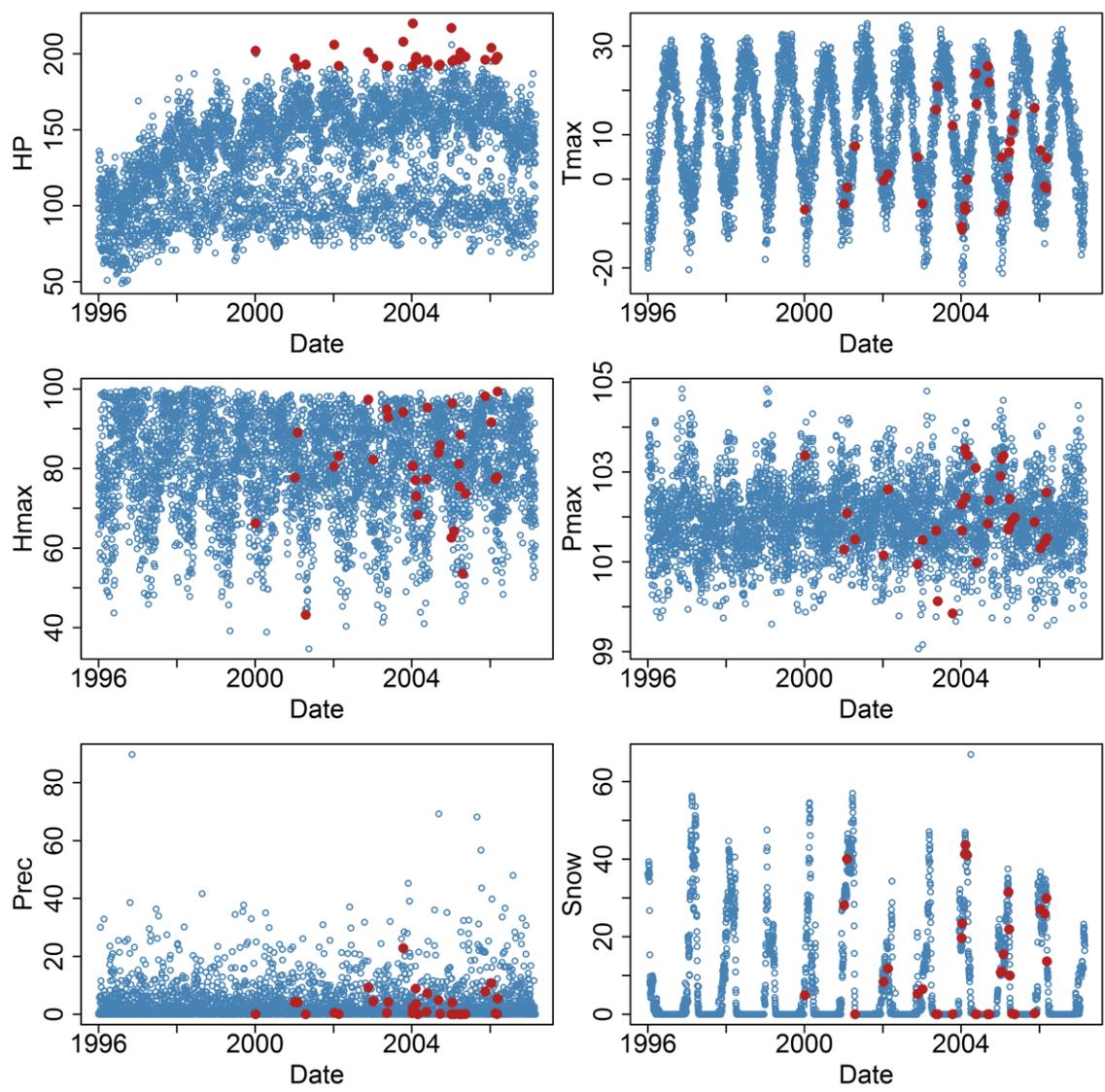


Figure II-6 : Conditions météorologiques associées aux pics d'hospitalisations (HP) à Montréal.

Les observations sont en bleu tandis que les pics sélectionnés sont en rouge.

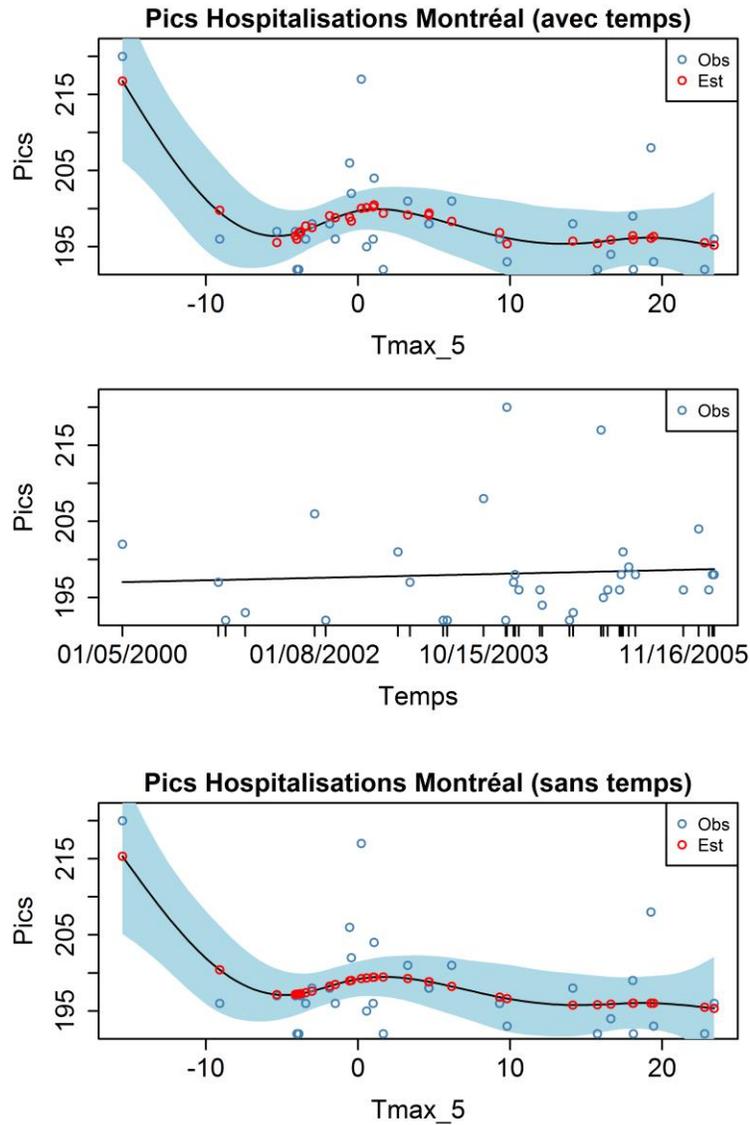


Figure II-7 : Fonction lissée de Tmax (lag de 5 jours) pour les pics d'hospitalisations à Montréal, avec et sans la fonction du temps.

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

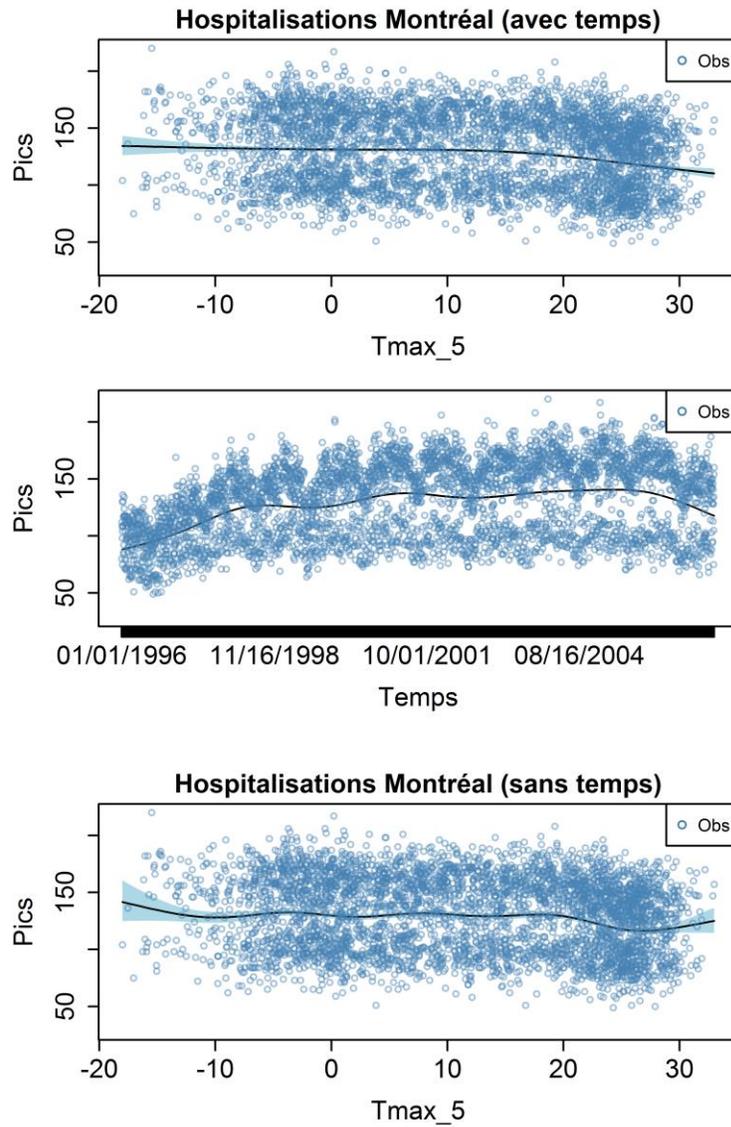


Figure II-8 : Fonction lissée de Tmax (lag de 5 jours) pour l'ensemble des hospitalisations à Montréal, avec et sans la fonction du temps.

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

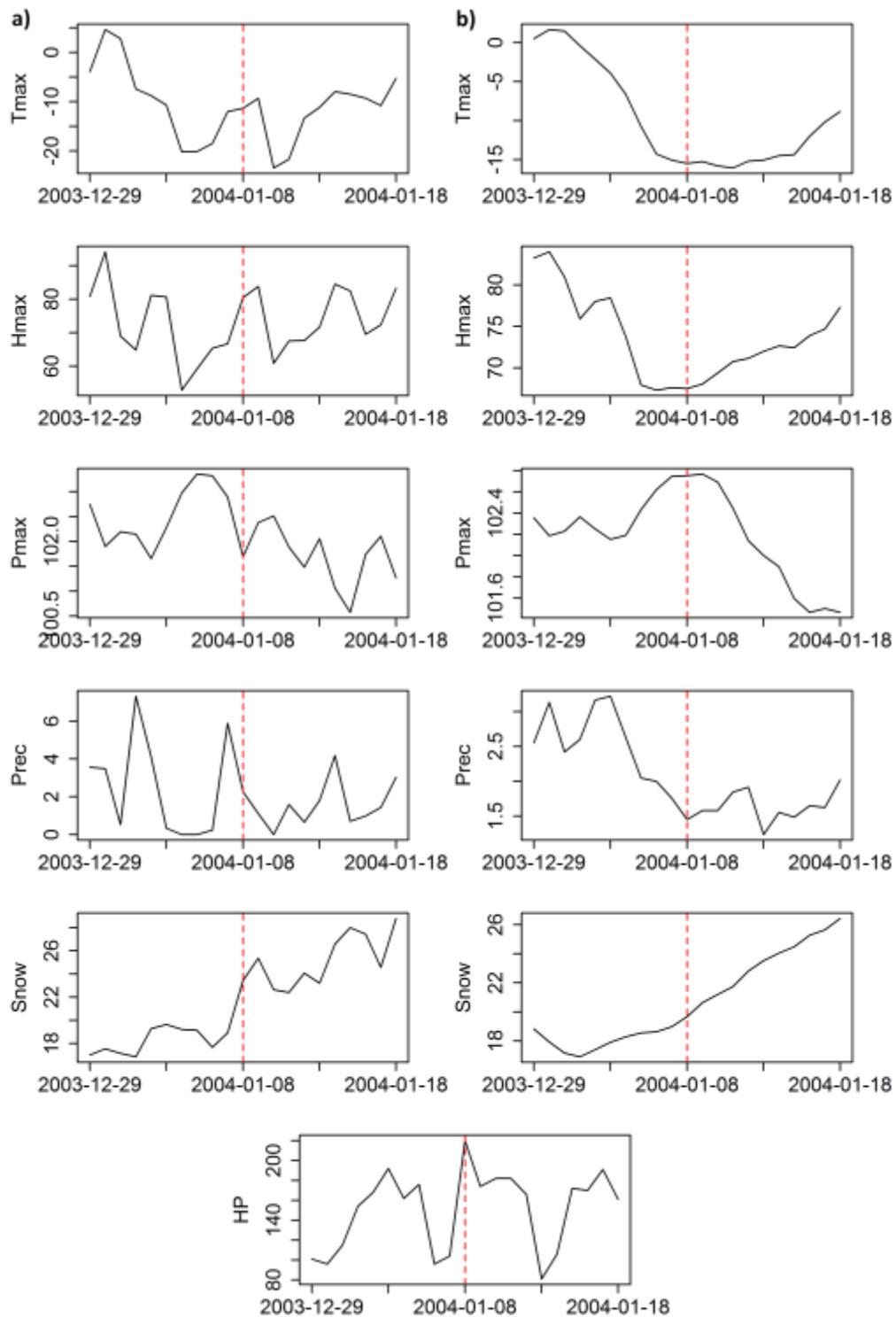


Figure II-9 : Conditions météorologiques associées au pic d'hospitalisations le plus élevé (220 hospitalisations) à Montréal :
a) les variables brutes et b) les variables moyennées sur 5 jours.

Les pointillés rouges représentent le jour du pic (12 janvier 2004).

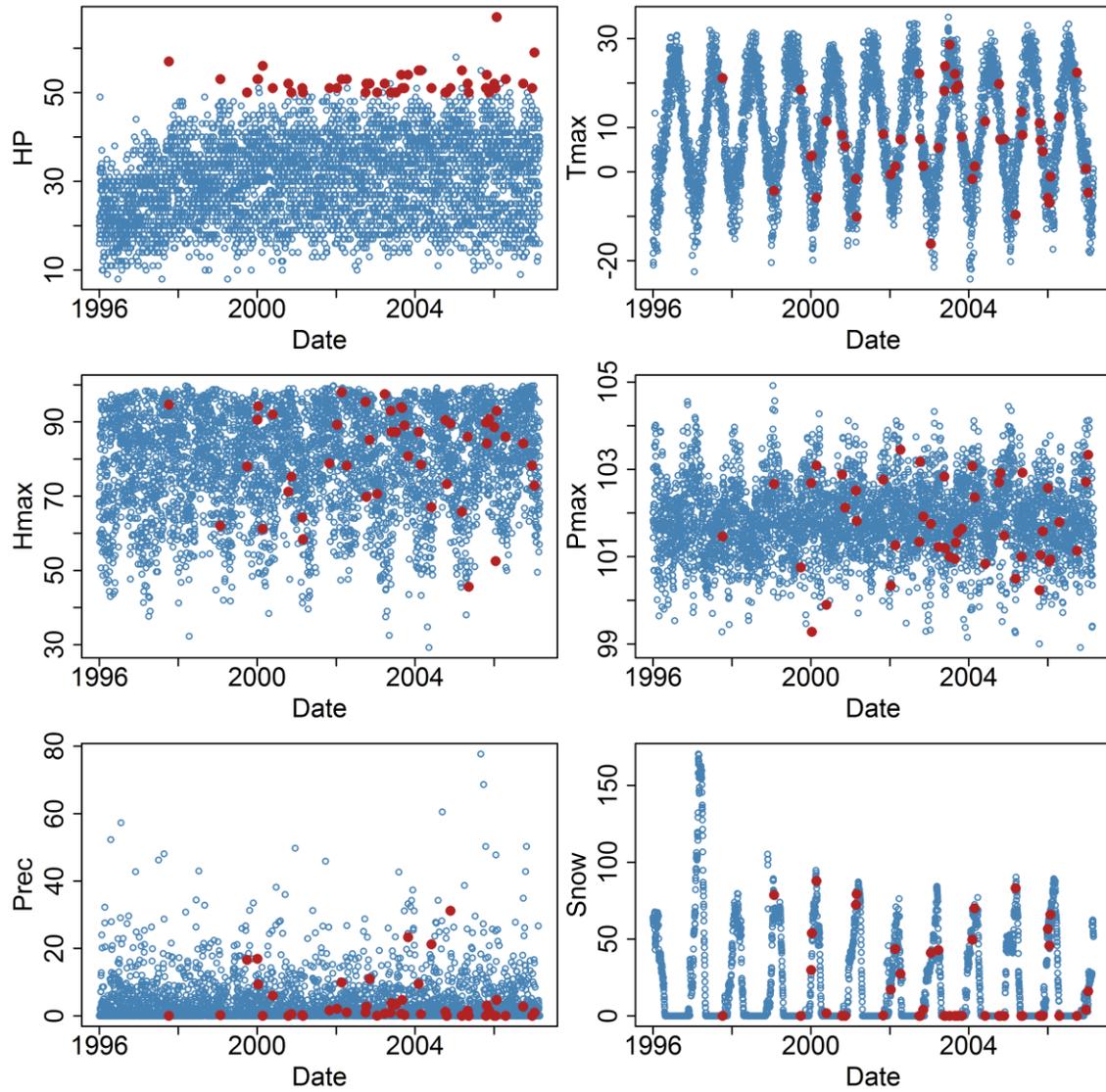


Figure II-10 : Conditions météorologiques associées aux pics d'hospitalisations (HP) à Québec (jours de semaine).

Les observations sont en bleu tandis que les pics sélectionnés sont en rouge.

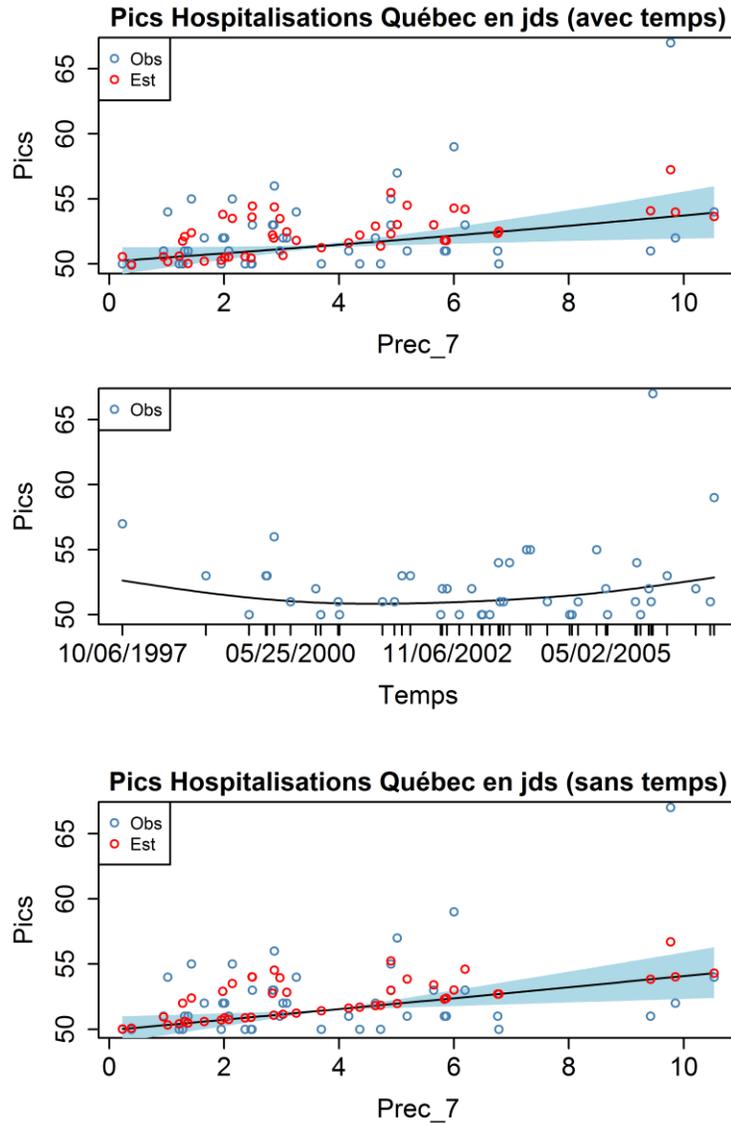


Figure II-11 : Fonction lissée de Prec (lag de 7 jours) pour les pics d'hospitalisations à Québec sur les jours de semaine (jds), avec et sans la fonction du temps.

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

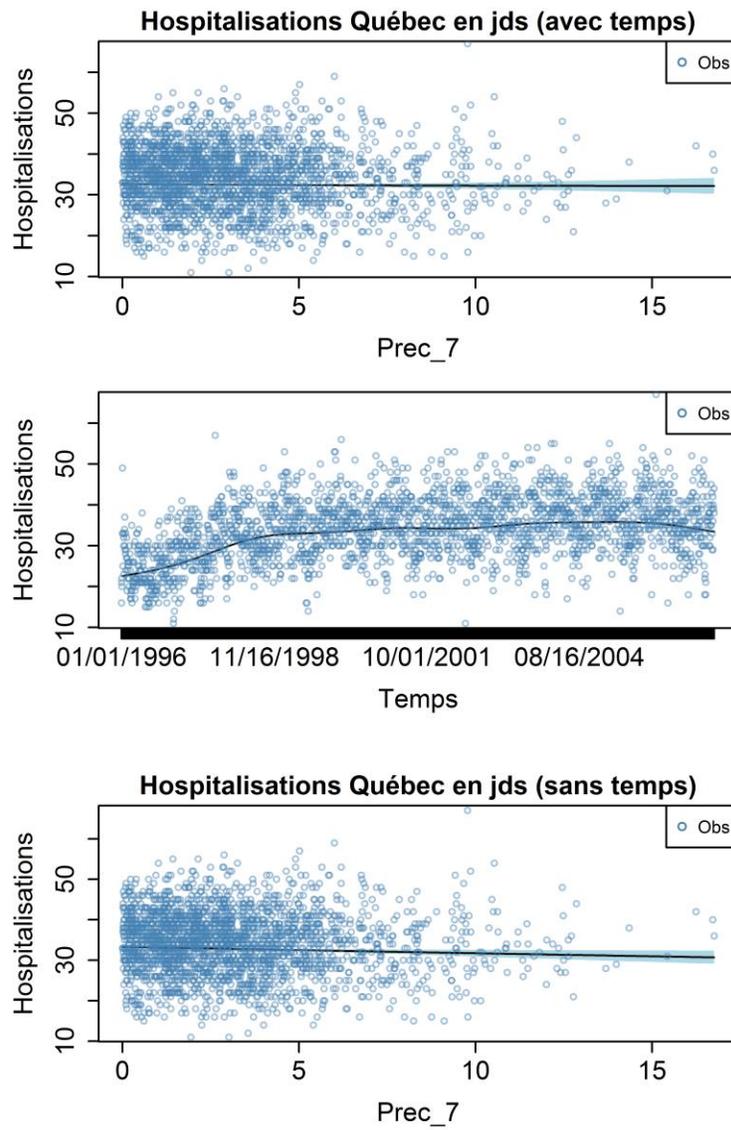


Figure II-12 : Fonction lissée de Prec (lag de 7 jours) pour l'ensemble des hospitalisations à Québec sur les jours de semaine (jds), avec et sans la fonction du temps.

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

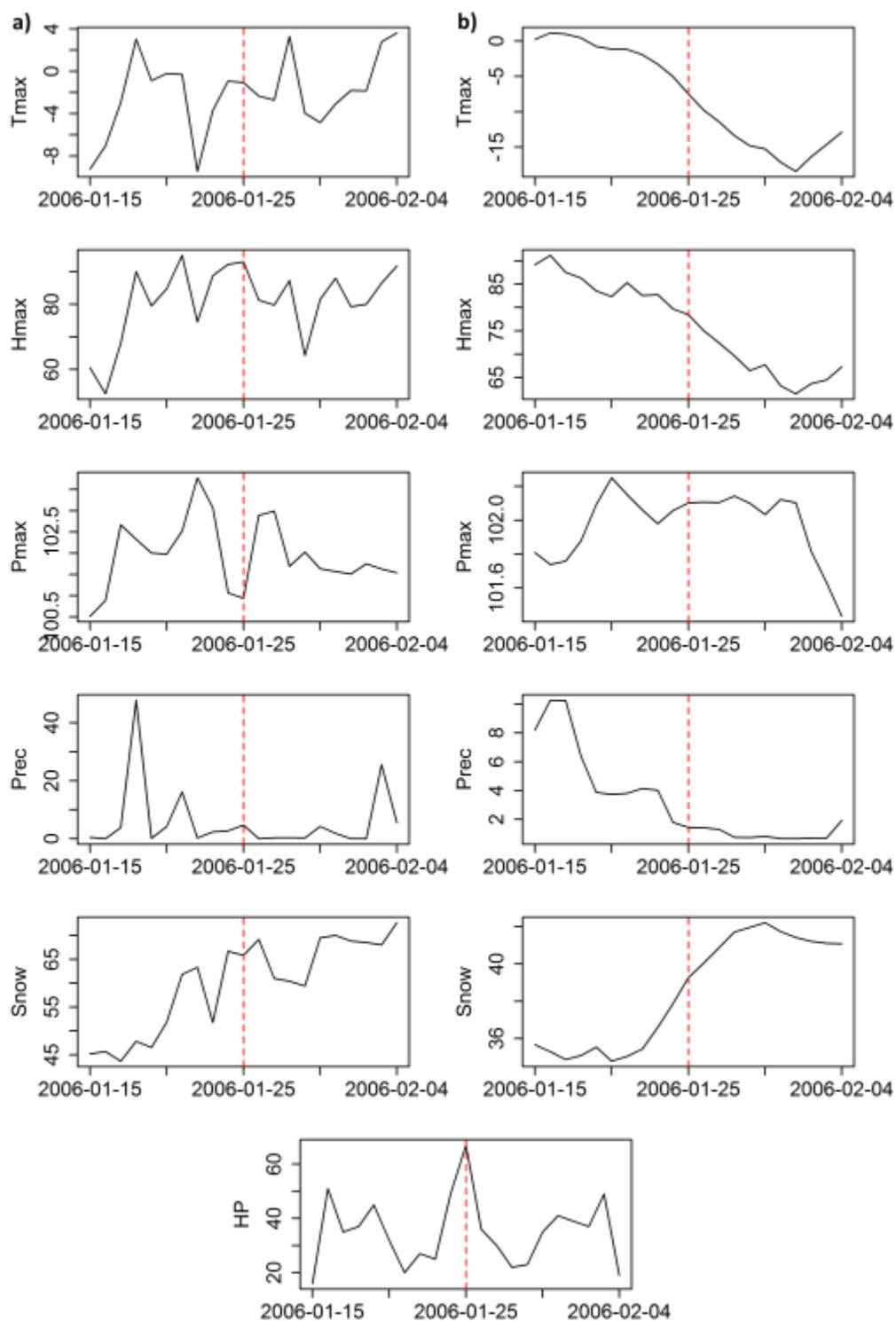


Figure II-13 : Conditions météorologiques associées au pic d'hospitalisations le plus élevé (67 hospitalisations) à Québec en jours de semaine : a) les variables brutes et b) les variables moyennées sur 7 jours.

Les pointillés rouges représentent le jour du pic (25 janvier 2006).

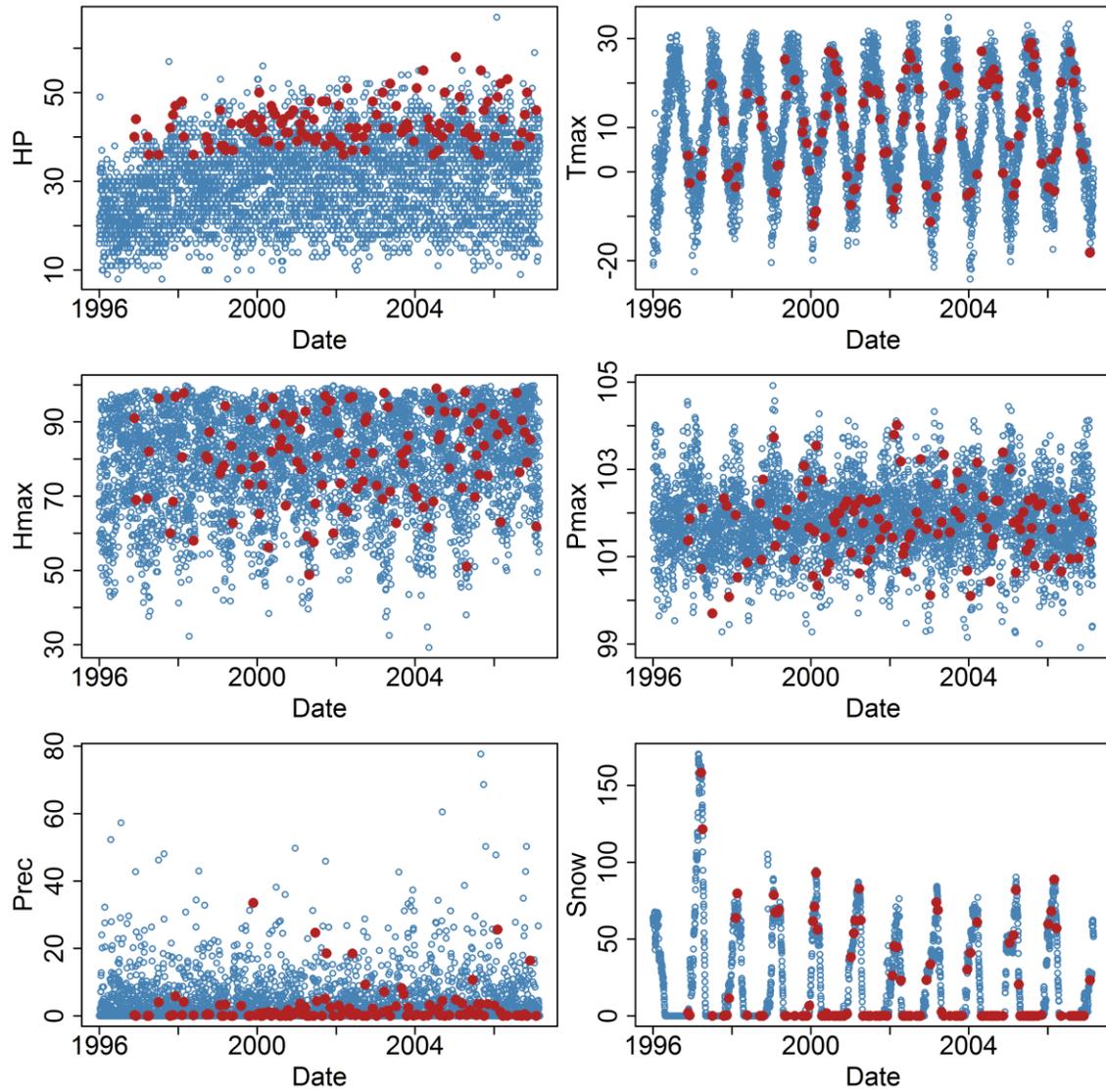


Figure II-14 : Conditions météorologiques associées aux pics d'hospitalisations (HP) à Québec (fins de semaine).

Les observations sont en bleu tandis que les pics sélectionnés sont en rouge.

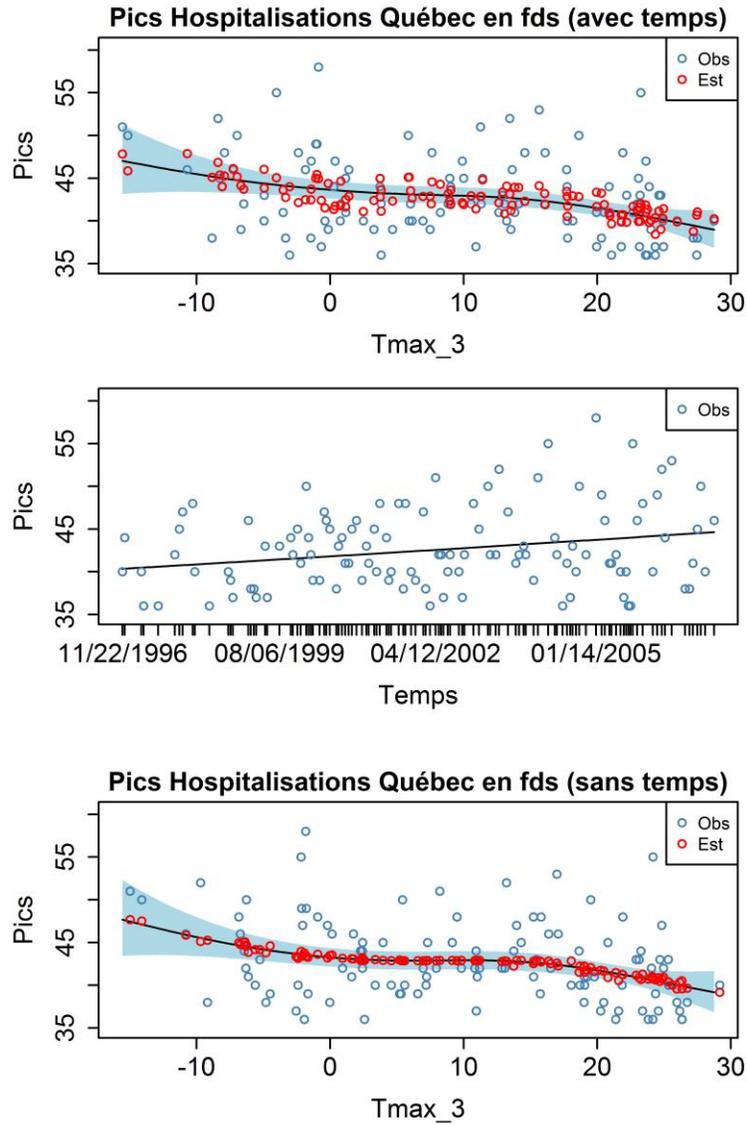


Figure II-15 : Fonction lissée de Tmax (lag de 3 jours) pour les pics d'hospitalisations à Québec en fins de semaine (fds), avec et sans la fonction du temps.

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

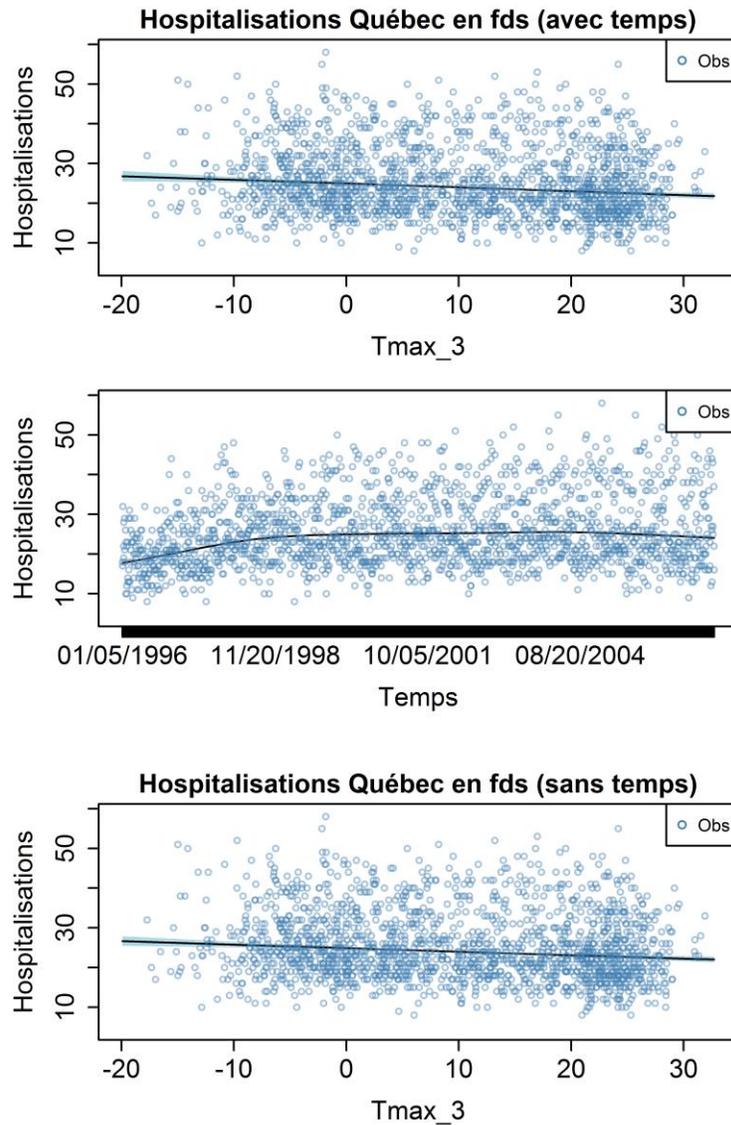


Figure II-16 : Fonction lissée de Tmax (lag de 3 jours) pour l'ensemble des hospitalisations à Québec en fins de semaine (fds), avec et sans la fonction du temps. Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

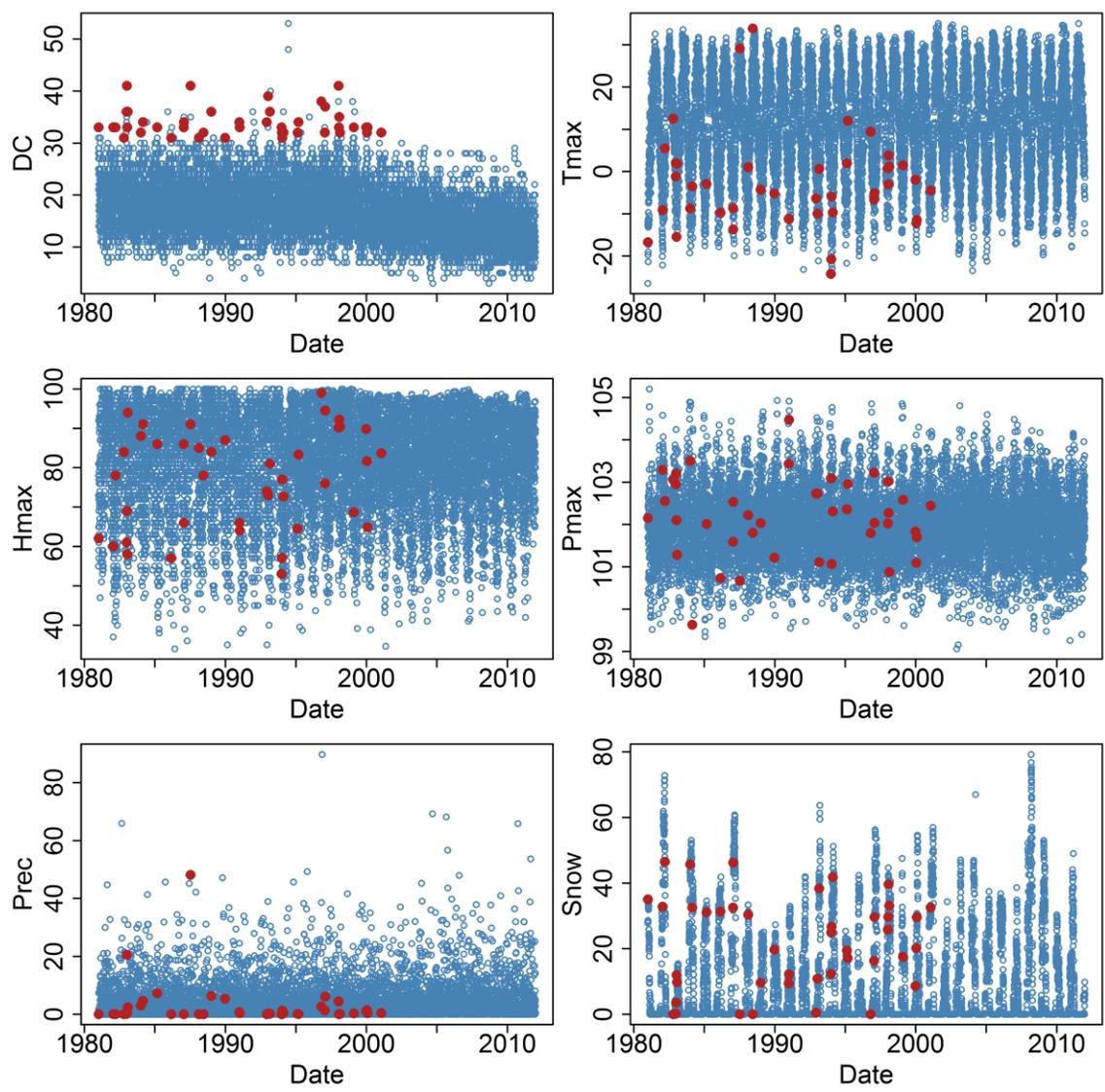


Figure II-17 : Conditions météorologiques associées aux pics de décès (DC) à Montréal (jours de semaine).

Les observations sont en bleu tandis que les pics sélectionnés sont en rouge.

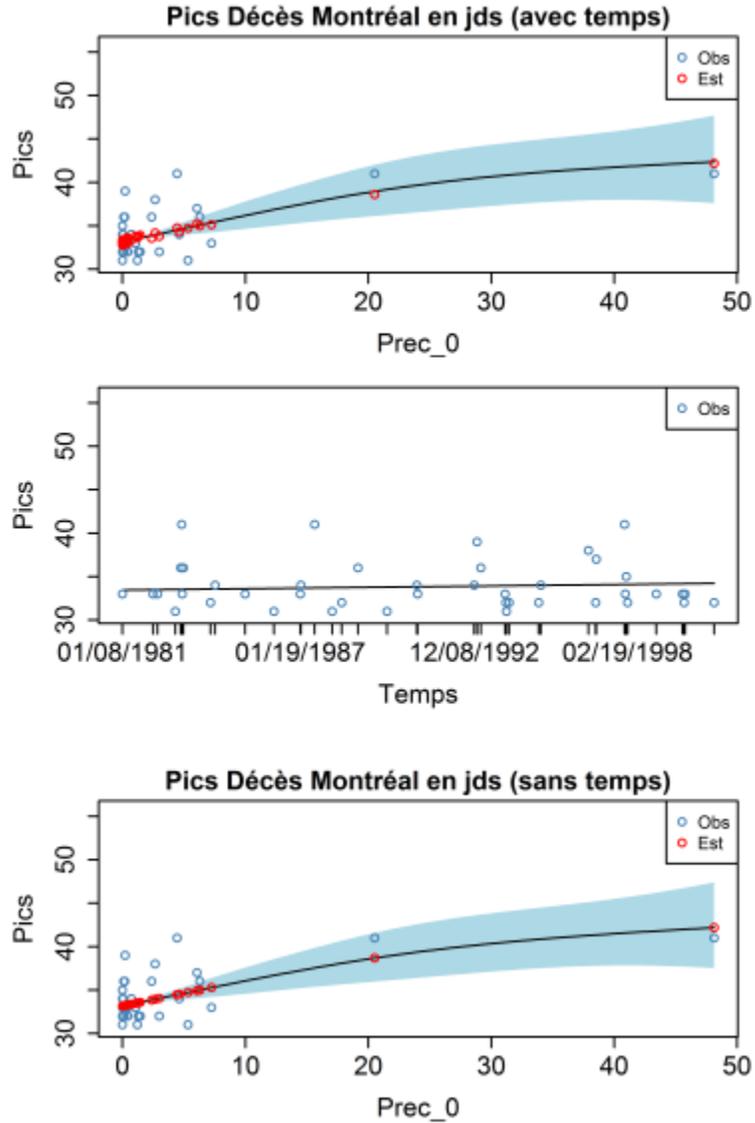


Figure II-18 : Fonction lissée de Prec (jour même) pour les pics de décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (premier modèle).

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

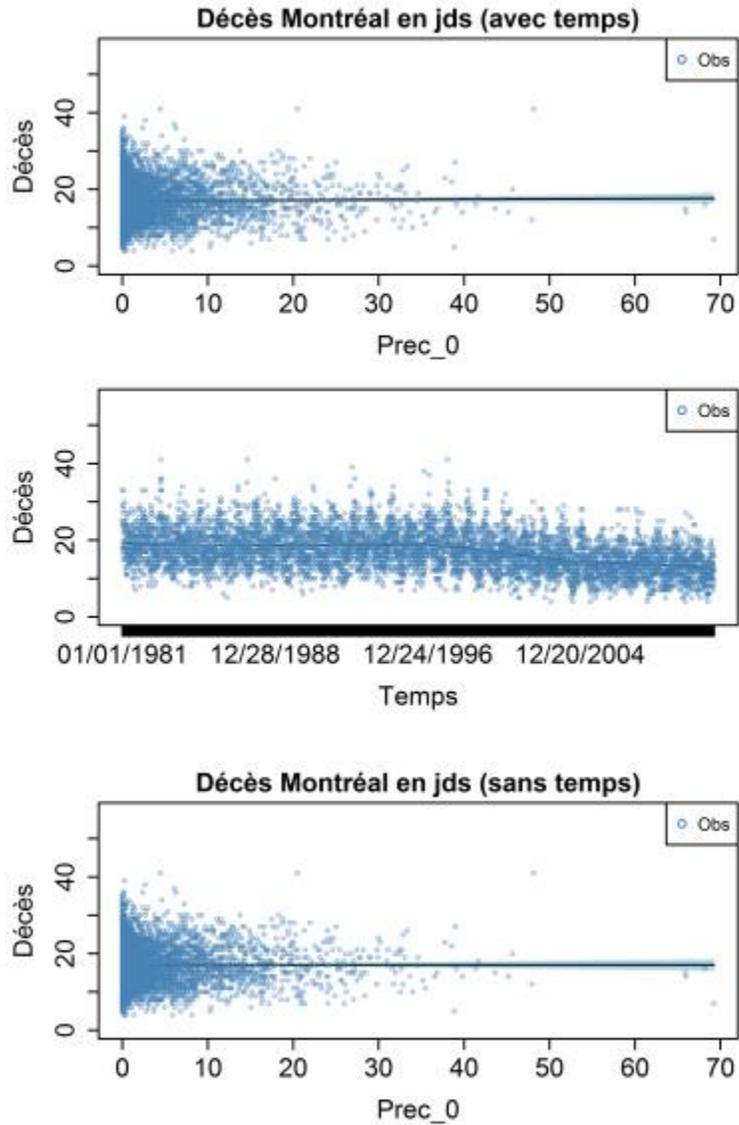


Figure II-19 : Fonction lissée de Prec (jour même) pour l'ensemble des décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (premier modèle).

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

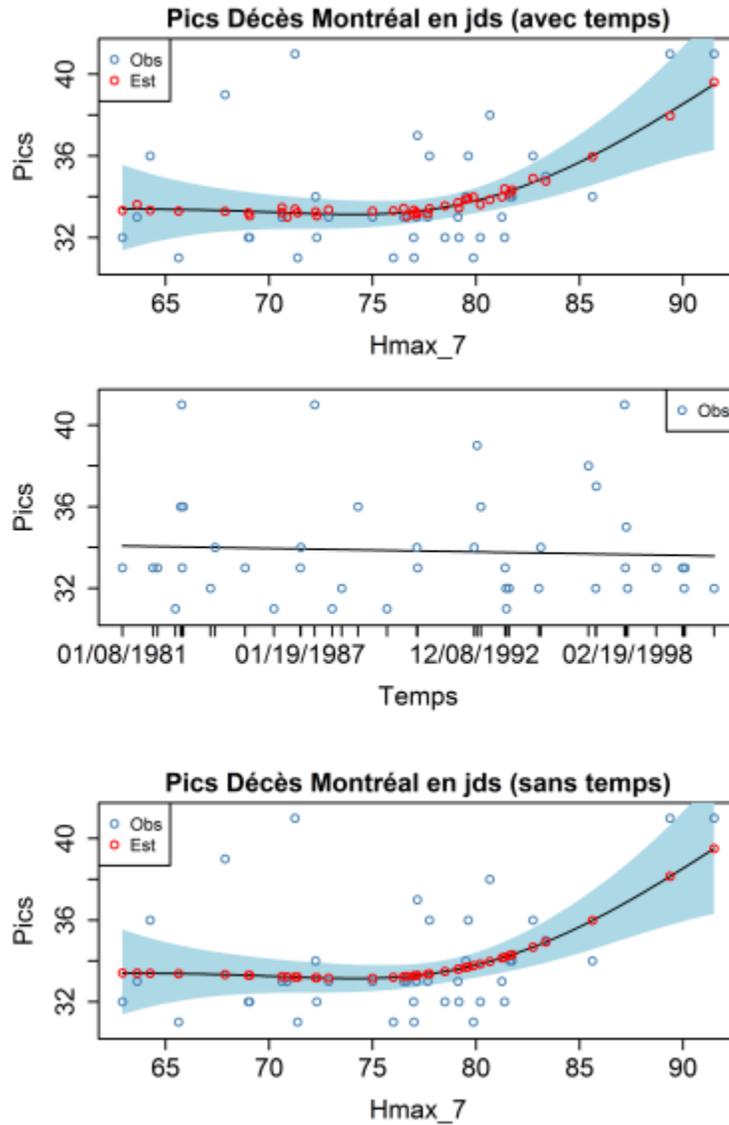


Figure II-20 : Fonction lissée de Hmax (lag de 7 jours) pour les pics de décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (deuxième modèle).

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

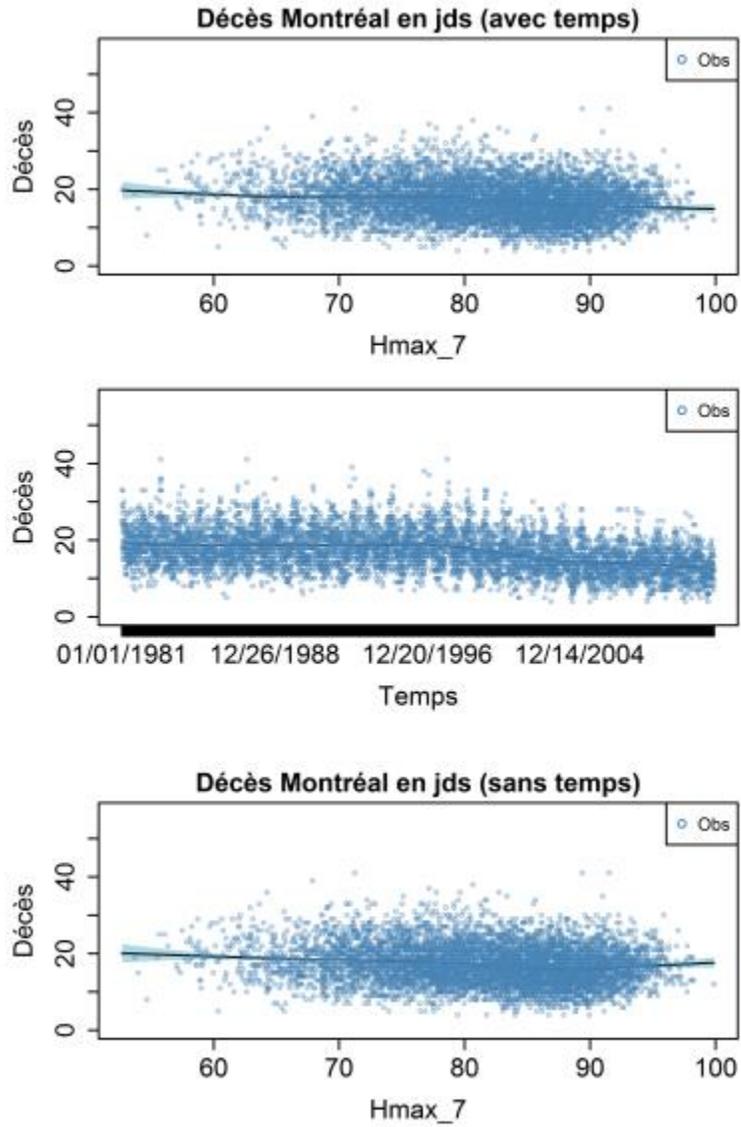


Figure II-21 : Fonction lissée de Hmax (lag de 7 jours) pour l'ensemble des décès à Montréal sur les jours de semaine (jds), avec et sans la fonction du temps (deuxième modèle).

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

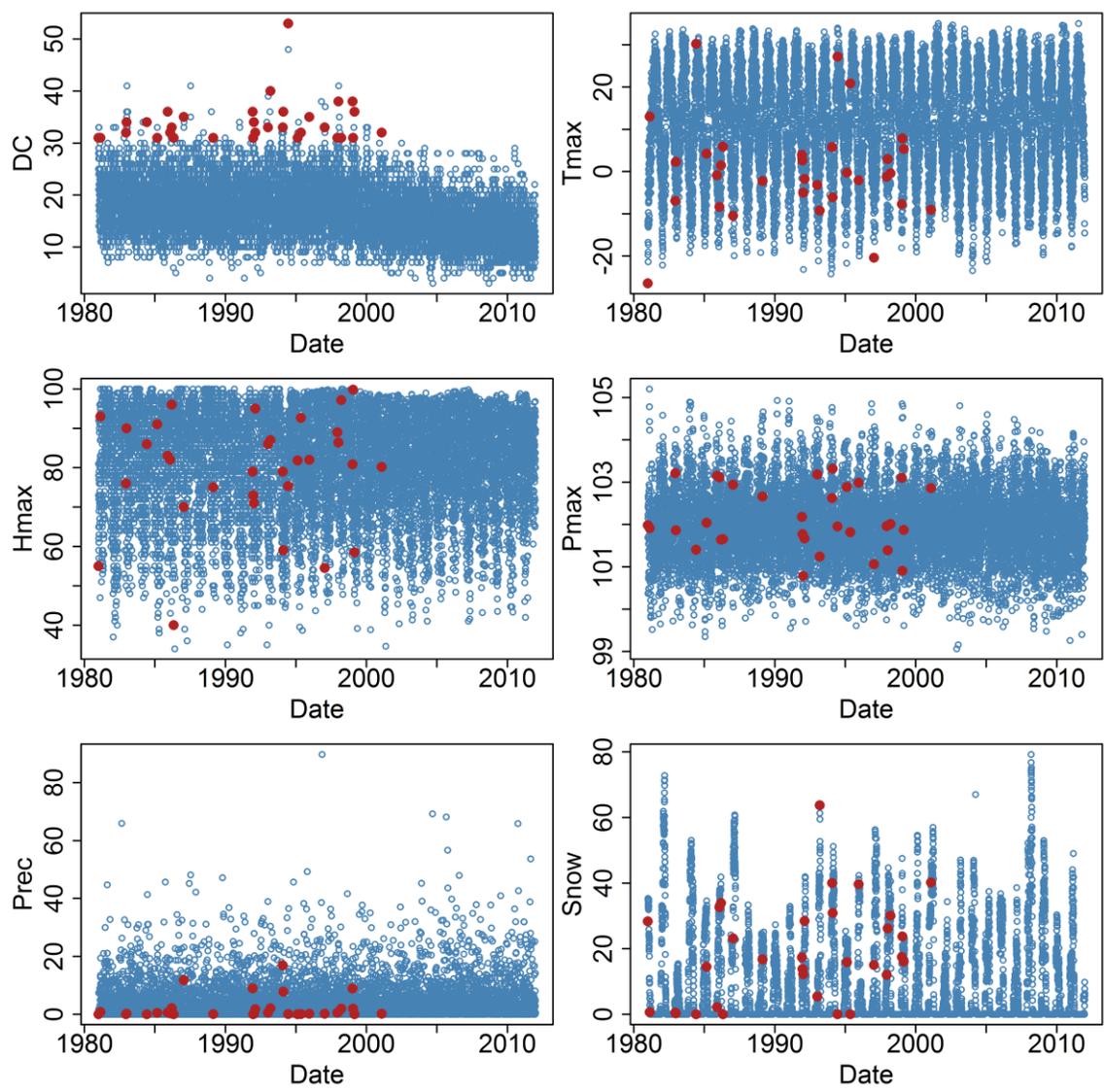


Figure II-22 : Conditions météorologiques associées aux pics de décès (DC) à Montréal (fins de semaine).

Les observations sont en bleu tandis que les pics sélectionnés sont en rouge.

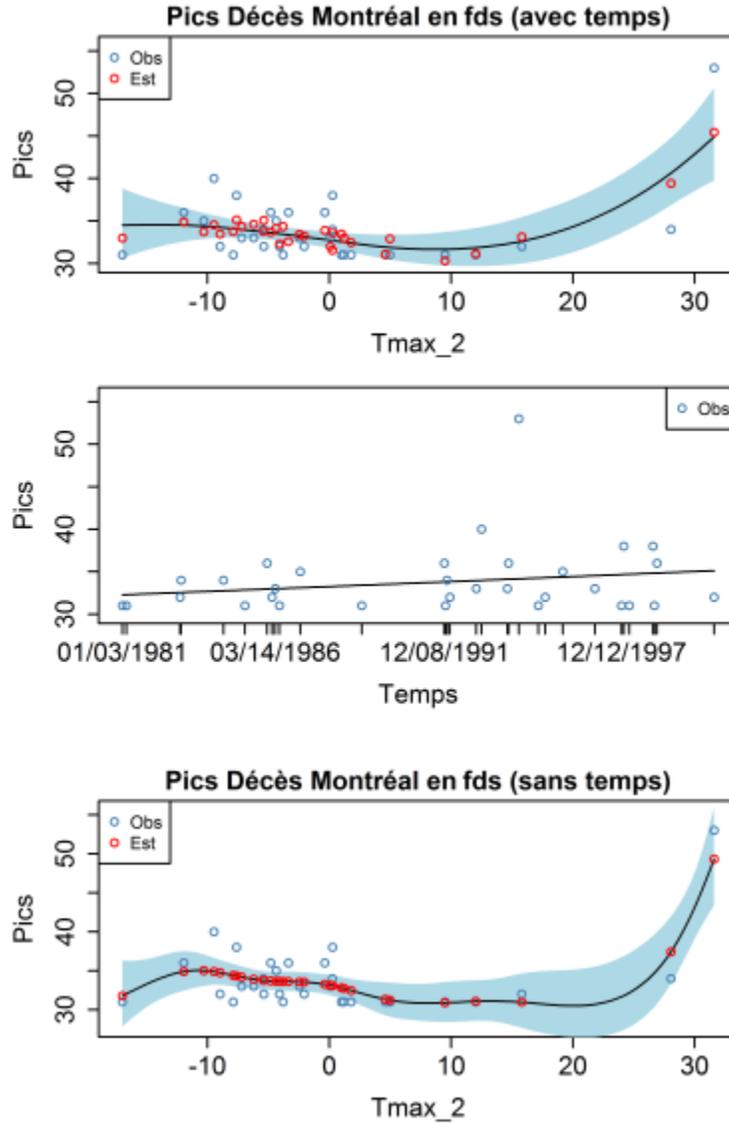


Figure II-23 : Fonction lissée de $Tmax$ (lag de 2 jours) pour les pics des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (premier modèle).

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

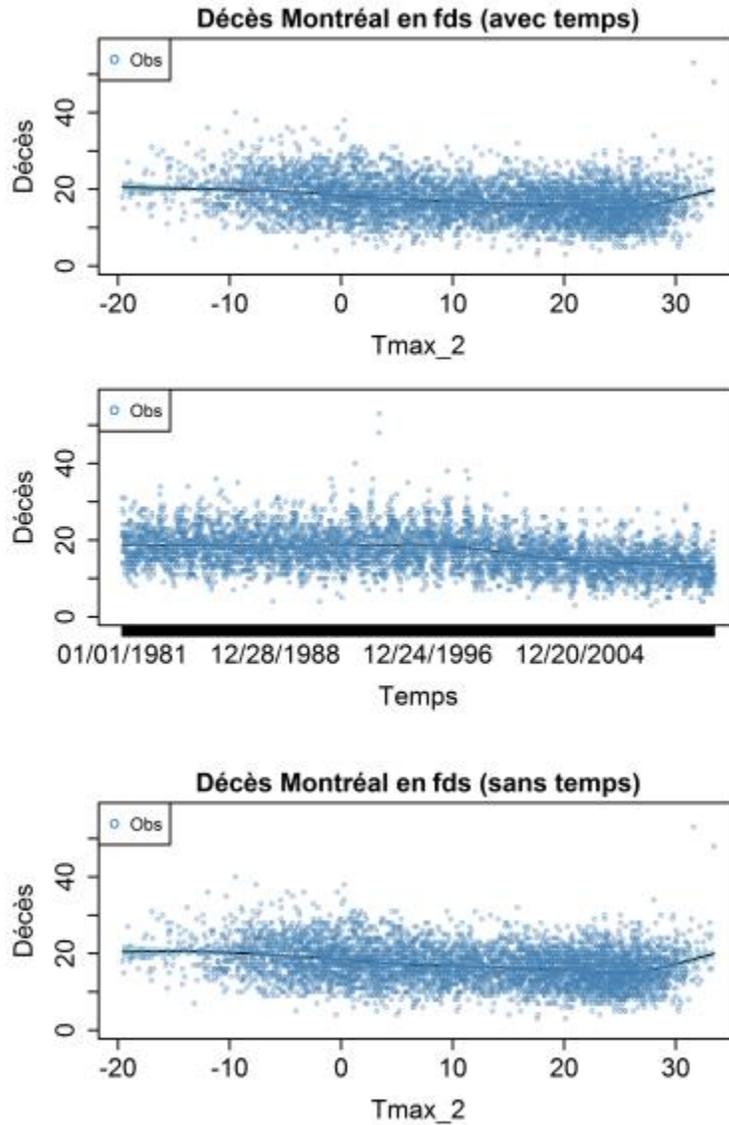


Figure II-24 : Fonction lissée de Tmax (lag de 2 jours) pour l'ensemble des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (premier modèle).

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

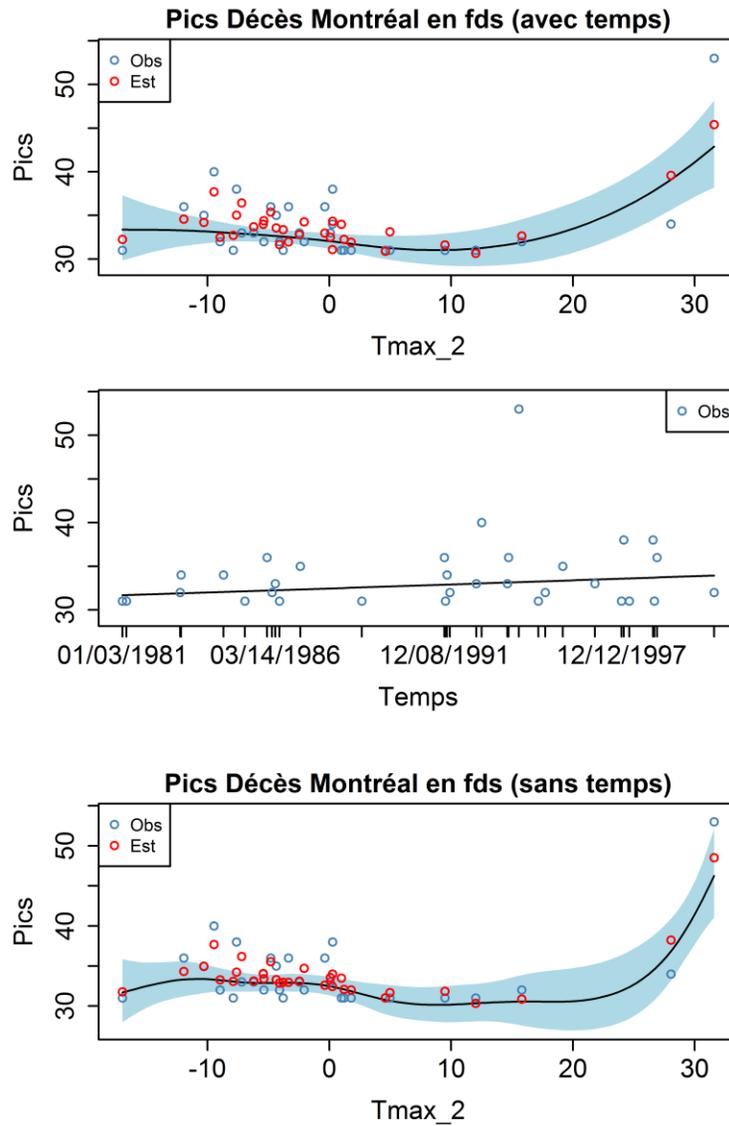


Figure II-25 : Fonction lissée de Tmax (lag de 2 jours) pour les pics des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (deuxième modèle).

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

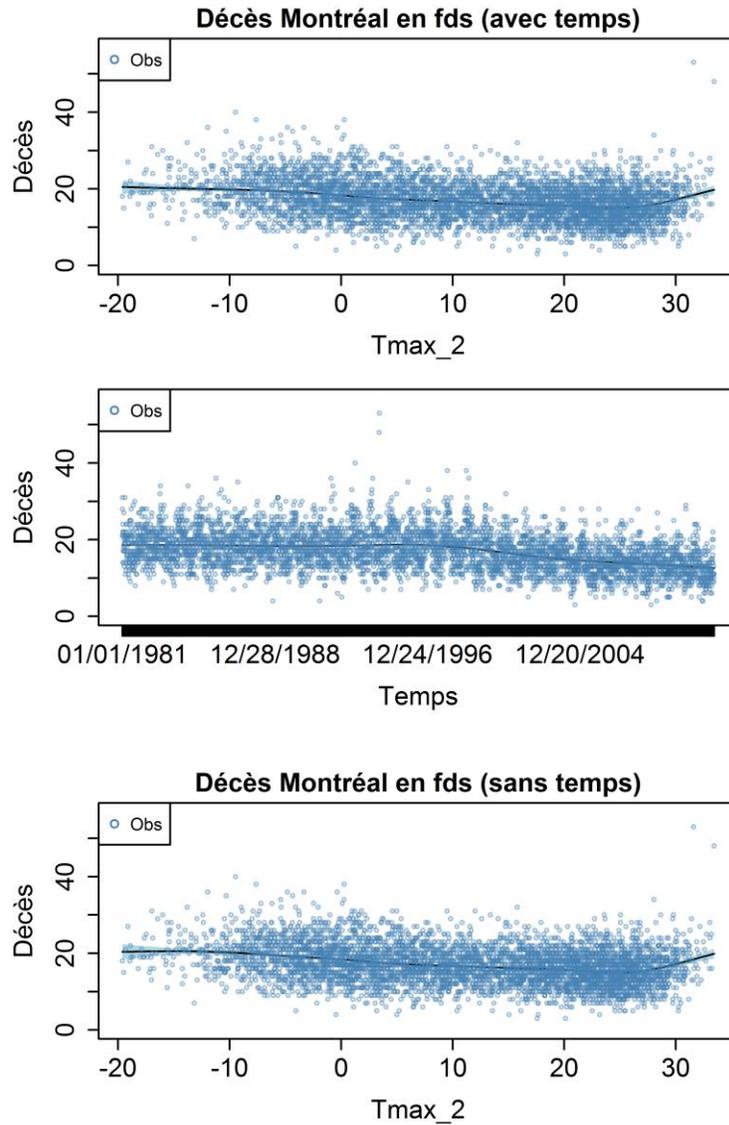


Figure II-26 : Fonction lissée de Tmax (lag de 2 jours) pour l'ensemble des décès à Montréal en fins de semaine (fds), avec et sans la fonction du temps (premier modèle).

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

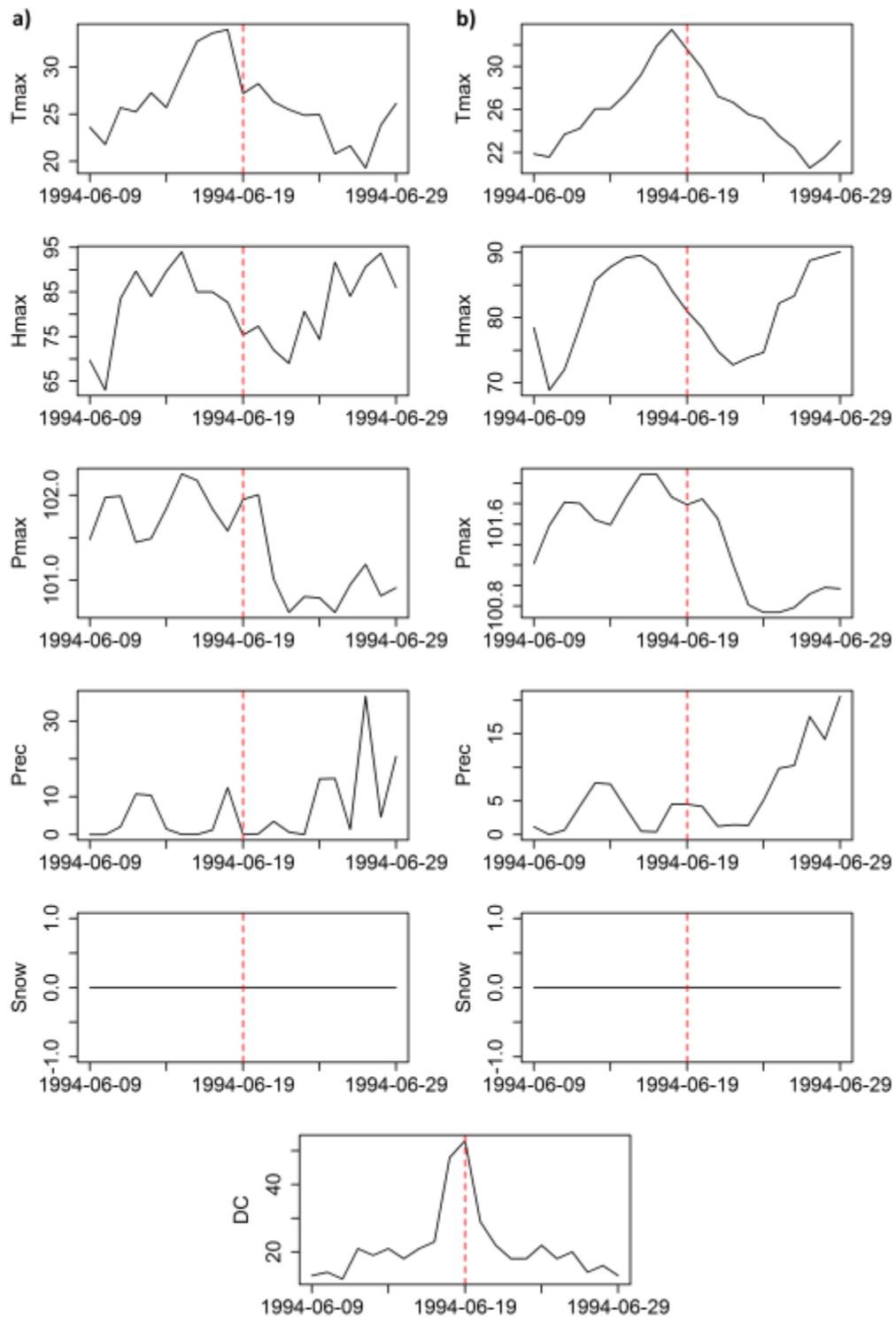


Figure II-27 : Conditions météorologiques associées au pic de décès le plus élevé (53 décès) à Montréal en fins de semaine :
a) les variables brutes et b) les variables moyennées sur 2 jours.

Les pointillés rouges représentent le jour du pic (19 juin 1994).

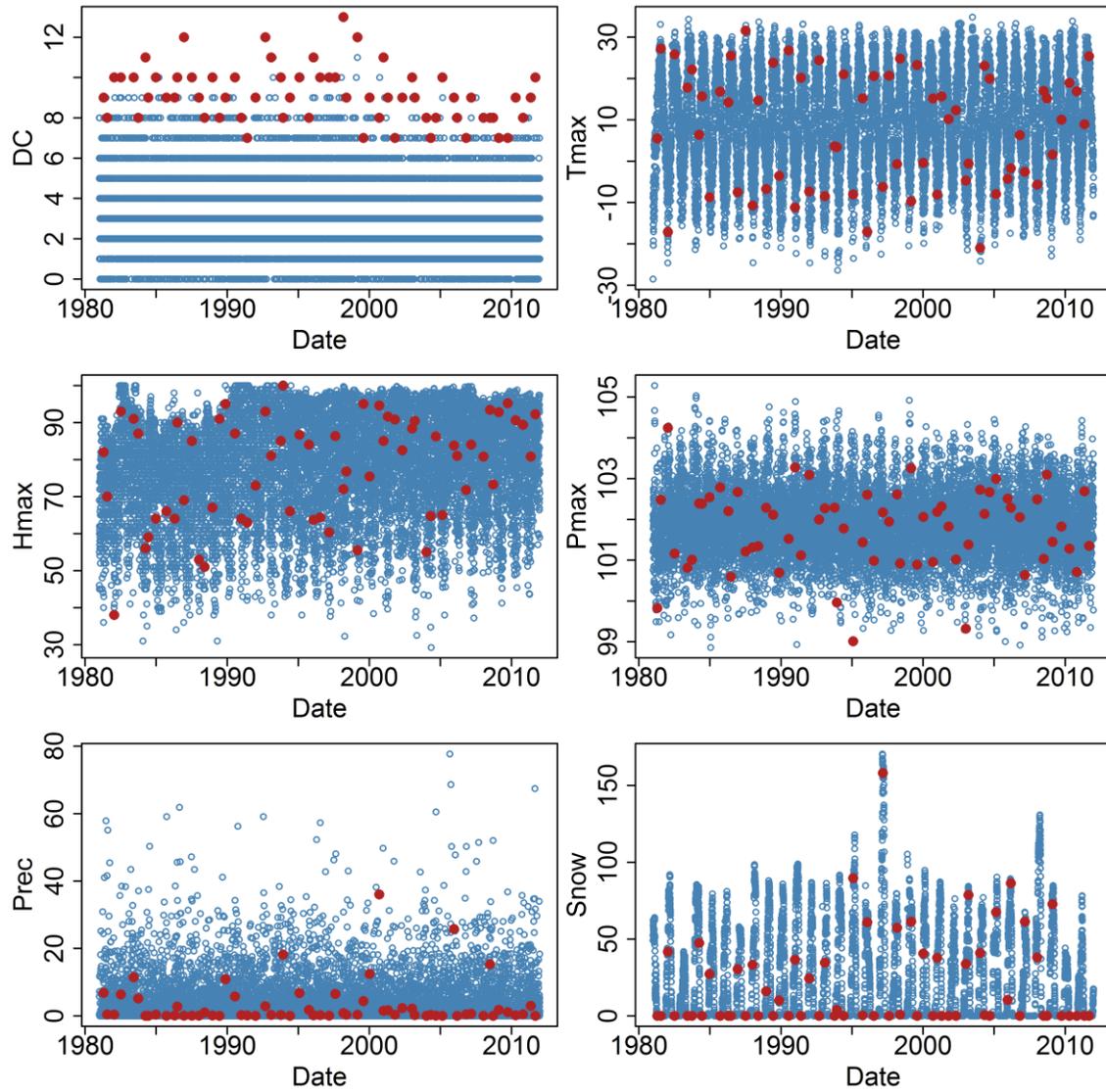


Figure II-28 : Conditions météorologiques associées aux pics de décès (DC) à Québec.

Les observations sont en bleu tandis que les pics sélectionnés sont en rouge.

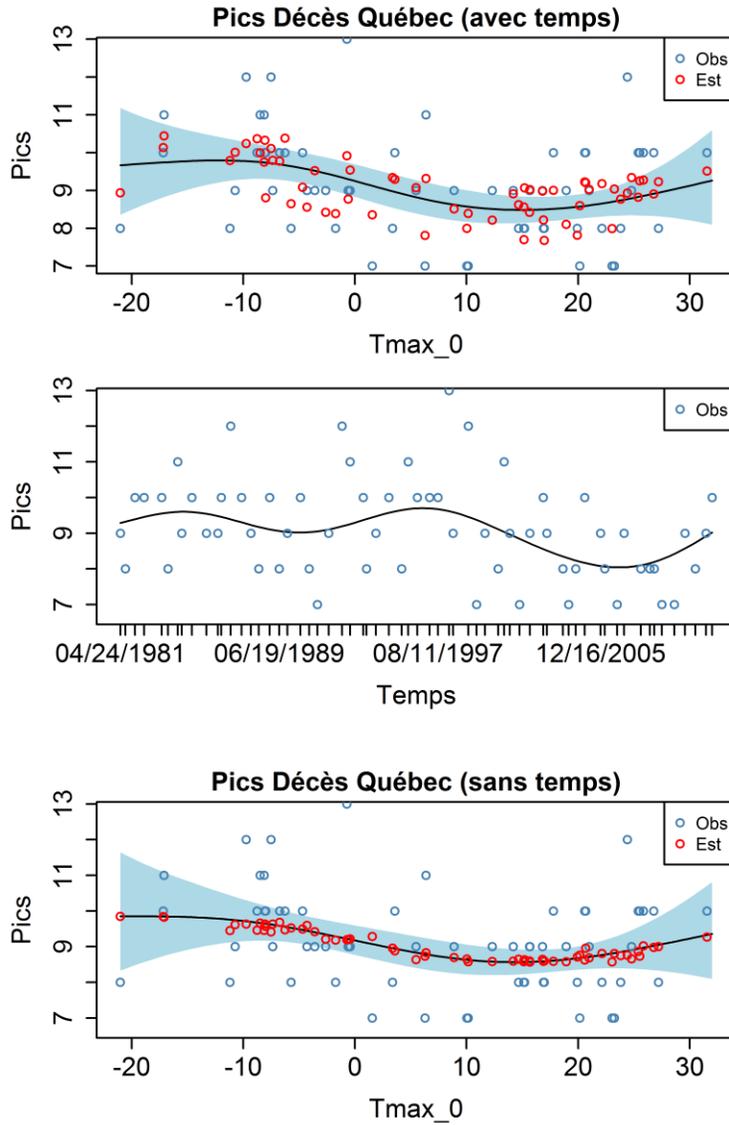


Figure II-29 : Fonction lissée de Tmax (jour même) pour les pics de décès à Québec, avec et sans la fonction du temps (premier modèle).

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

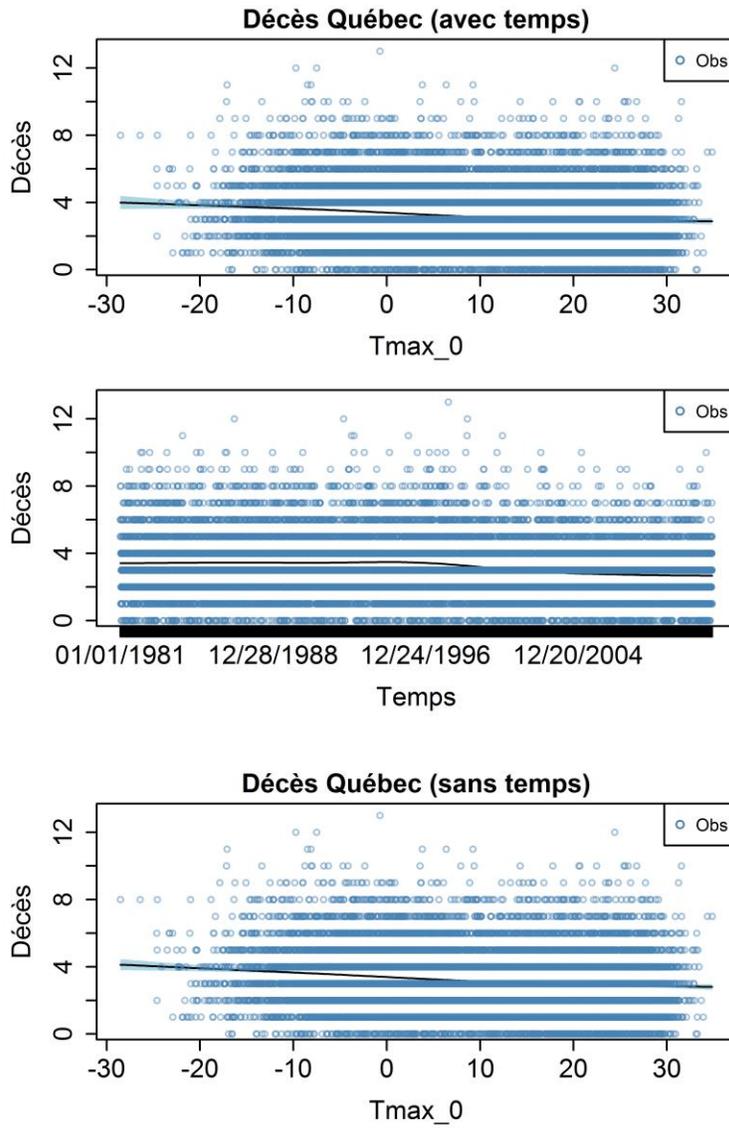


Figure II-30 : Fonction lissée de Tmax (jour même) pour l'ensemble des décès à Québec, avec et sans la fonction du temps (premier modèle).

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

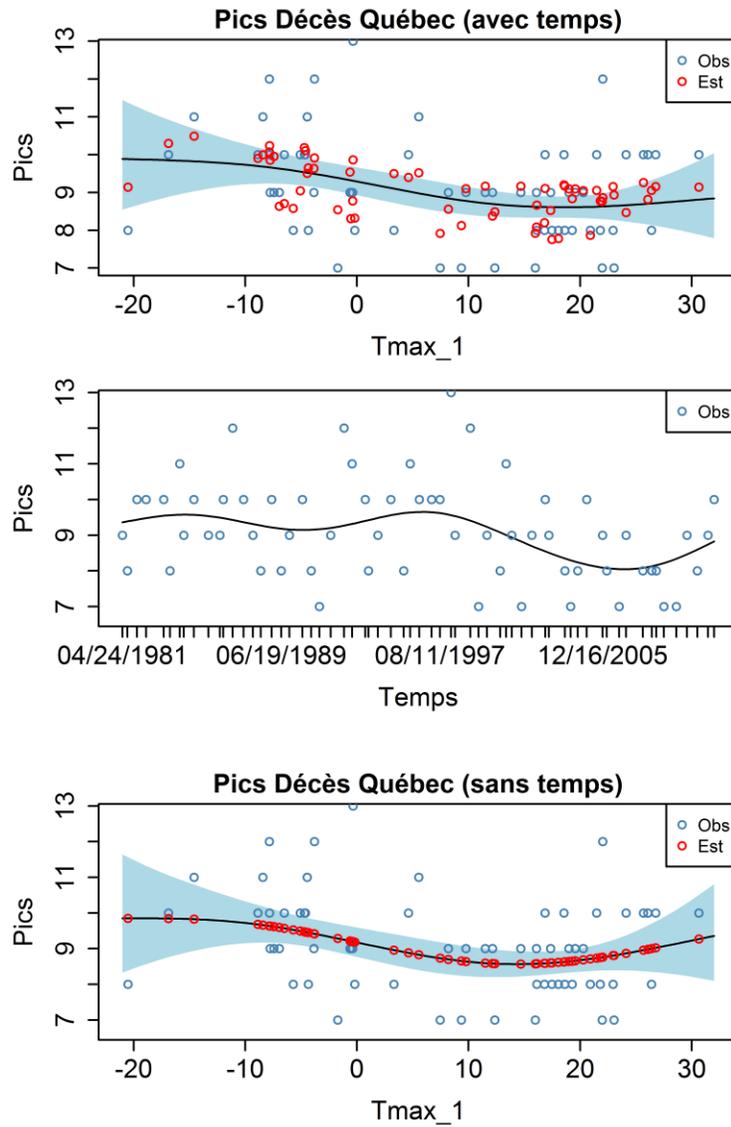


Figure II-31 : Fonction lissée de $Tmax$ (lag de 1 jour) pour les pics de décès à Québec, avec et sans la fonction du temps (deuxième modèle).

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

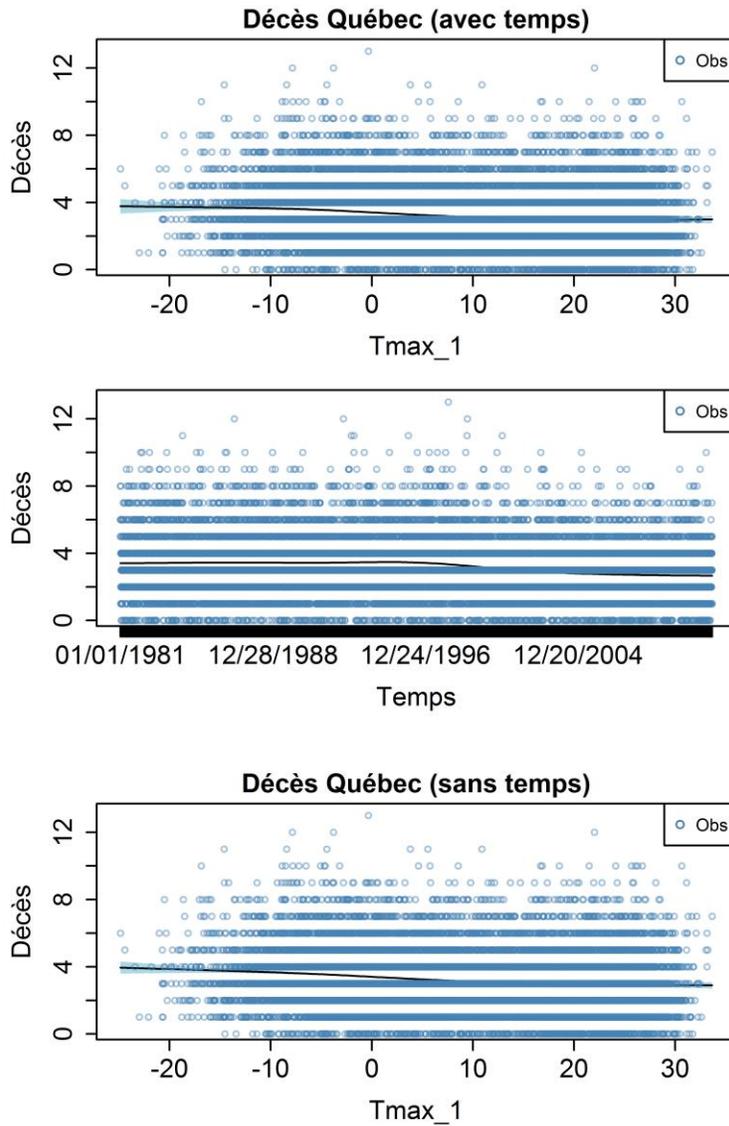


Figure II-32 : Fonction lissée de Tmax (lag de 1 jour) pour l'ensemble des décès à Québec, avec et sans la fonction du temps (deuxième modèle).

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

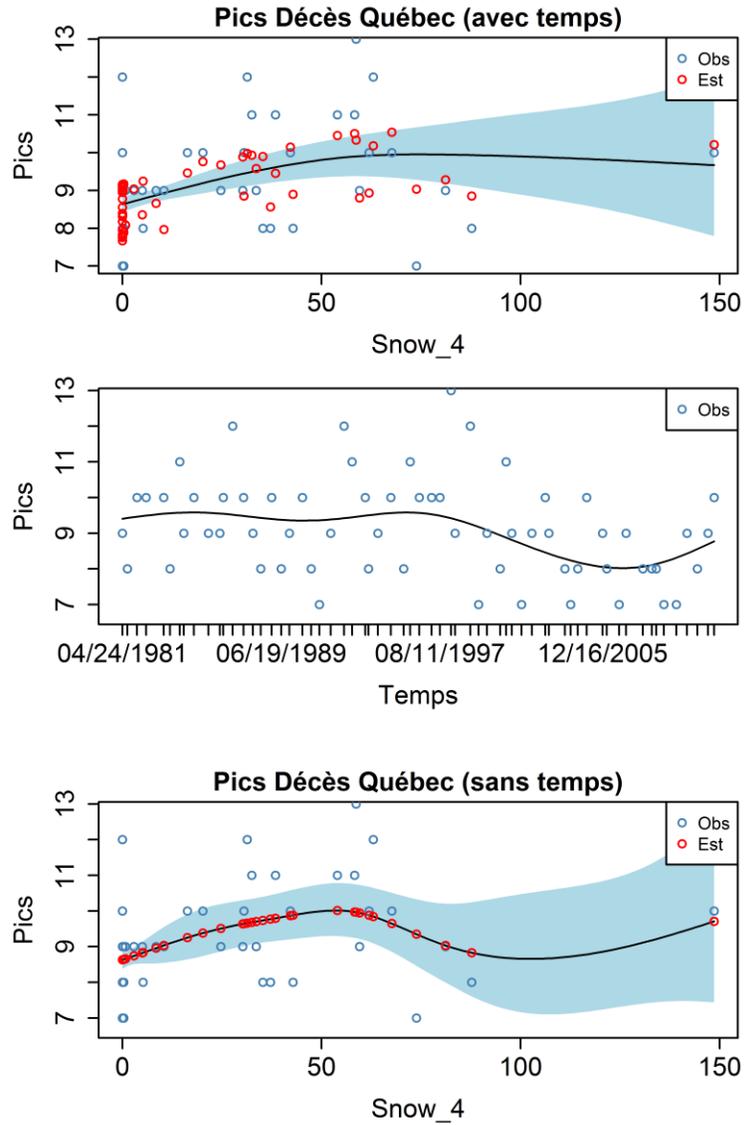


Figure II-33 : Fonction lissée de Snow (lag de 4 jours) pour les pics de décès à Québec, avec et sans la fonction du temps.

Les observations sont en bleu tandis que les estimations des pics par le modèle sont en rouge. Les intervalles de confiance à 95 % sont aussi présentés.

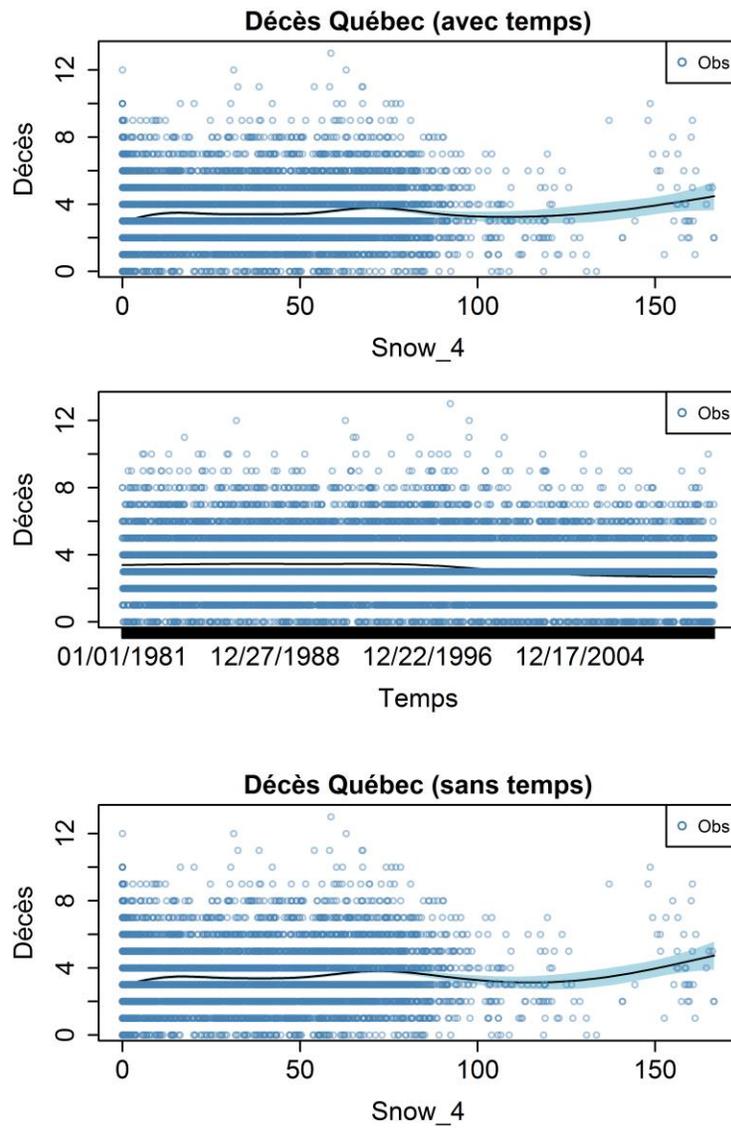


Figure II-34 : Fonction lissée de Snow (lag de 4 jours) pour l'ensemble des décès à Québec, avec et sans la fonction du temps.

Les observations sont en bleu. Les intervalles de confiance à 95 % sont aussi présentés.

Tableaux

Tableau II-1 : Résumé des modèles sélectionnés pour les séries de pics d'hospitalisations (HP) et de décès (DC) à Montréal et Québec (Chiu *et al.*, 2015).

		MODÈLE À MONTRÉAL			MODÈLE À QUÉBEC		
		Extraction par <i>seuil</i>	Seuil en %	n	Extraction par <i>seuil</i>	Seuil en %	n
HP	Global	99	33	Jours de semaine	97,5	46	
				Fins de semaine	85	115	
		MODÈLE À MONTRÉAL			MODÈLE À QUÉBEC		
		Extraction par <i>seuil</i>	Seuil en %	n	Extraction par <i>bloc</i>	Bloc en jours	n
DC	Jours de semaine	99	42	Global	180	63	
	Fins de semaine	99	32				

Tableau II-2 : Corrélations entre les différentes déclinaisons de la température, de l'humidité et de la pression atmosphérique.

Variable	Déclinaisons	Minimale	Moyenne	Maximale
Température	Minimale	1	0,99	0,95
	Moyenne	0,99	1	0,99
	Maximale	0,95	0,99	1
Humidité	Minimale	1	0,88	0,44
	Moyenne	0,88	1	0,76
	Maximale	0,44	0,76	1
Pression	Minimale	1	0,94	0,79
	Moyenne	0,94	1	0,94
	Maximale	0,79	0,94	1

Tableau II-3 : Statistiques descriptives des données météorologiques brutes utilisées, pour Montréal et Québec.

	MONTRÉAL					QUÉBEC				
	<i>Tmax</i>	<i>Hmax</i>	<i>Pmax</i>	<i>Prec</i>	<i>Snow</i>	<i>Tmax</i>	<i>Hmax</i>	<i>Pmax</i>	<i>Prec</i>	<i>Snow</i>
Minimum	-26,46	34,00	99,06	0,00	0,00	-28,52	29,25	98,86	0,00	0,00
Maximum	35,05	100,00	105,2	89,77	79,23	34,84	100,00	105,28	77,67	170,5
Moyenne	11,43	81,59	101,96	2,83	6,79	9,68	78,26	101,88	3,34	17,67
Médiane	12,23	84,00	101,92	0,32	0,00	9,88	80,00	101,85	0,46	0,00
Q75	22,37	91,83	102,47	3,00	8,88	20,41	89,00	102,42	3,78	31,87
Q90	26,85	96,00	103,00	8,87	27,41	25,4	94,25	102,95	10,62	63,2

Les températures, humidité relative, pression atmosphérique, précipitations et quantité de neige au sol sont en degrés Celsius, pourcentage, kilo Pascals, millimètres et centimètres

Tableau II-4 : Résumé numérique des conditions météorologiques associées aux pics d'hospitalisations à Montréal.

	HP	Tmax	Hmax	Pmax	Prec	Snow
Minimum	192	-11.36	43.17	99.85	0	0
1er quantile	194	-5.51	75.33	101.5	0	0
Médiane	197	4.71	80.67	101.9	0.48	10.75
Moyenne	198.2	4.54	80.68	102	3.22	14.28
3ème quantile	199	12.04	91.57	102.5	4.43	25.88
Maximum	220	25.41	99.33	103.5	22.86	43.64

Tableau II-5 : Résultats du modèle retenu pour expliquer la relation entre les pics d'hospitalisations et la météo à Montréal.

Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
PICS D'HOSPITALISATIONS AVEC LE TEMPS					
Intercept		5,29 (*)	39,60	-134,13	1,12e-3
s(Tmax)	5	4,11			
s(Time)		1,00			
PICS D'HOSPITALISATIONS SANS LE TEMPS					
Intercept		5,29 (*)	40,10	-136,13	1,01e-3
s(Tmax)	5	4,26			
TOUTES LES HOSPITALISATIONS AVEC LE TEMPS					
Intercept		4,84 (*)	21,00	5,66e-2	-142,88
s(Tmax)	5	3,40			
s(Time)		8,49			
TOUTES LES HOSPITALISATIONS SANS LE TEMPS					
Intercept		4,84 (*)	2,87	6,94e-2	689,54
s(Tmax)	5	7,95			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-6 : Résumé numérique des conditions météorologiques associées aux pics d'hospitalisations à Québec (jours de semaine).

	HP	Tmax	Hmax	Pmax	Prec	Snow
Minimum	50	-16.26	45.6	99.28	0	0
1er quantile	50.25	-0.29	72.91	101	0.01	0
Médiane	51.5	7.22	85.6	101.6	0.99	0
Moyenne	52.37	6.99	81.15	101.8	4.22	21.08
3ème quantile	53	13.21	90.57	102.7	4.44	43.33
Maximum	67	28.63	98	103.4	31.15	87.75

Tableau II-7 : Résultats du modèle retenu pour expliquer la relation entre les pics d'hospitalisations et la météo à Québec (jours de semaine).

Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
PICS D'HOSPITALISATIONS AVEC LE TEMPS					
Intercept		3,94 (*)	38,70	-147,29	2,45e-3
s(Prec)	7	1,08			
Snow	0	7,74e-4 (*)			
s(Time)		2,04			
PICS D'HOSPITALISATIONS SANS LE TEMPS					
Intercept		3,94 (*)	30,00	-145,39	2,42e-3
s(Prec)	7	1,00			
Snow	0	7,46e-4 (*)			
TOUTES LES HOSPITALISATIONS AVEC LE TEMPS					
Intercept		3,48 (*)	27,60	-798,14	4,16e-2
s(Prec)	7	1,21			
Snow	0	1,27e-3 (*)			
s(Time)		7,72			
TOUTES LES HOSPITALISATIONS SANS LE TEMPS					
Intercept		3,49 (*)	1,35	-93,61	5,62e-2
s(Prec)	7	1,00			
Snow	0	8e-4 (*)			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-8 : Résumé numérique des conditions météorologiques associées aux pics d'hospitalisations à Québec (fins de semaine).

	HP	Tmax	Hmax	Pmax	Prec	Snow
Minimum	36	-18.24	48.8	99.7	0	0
1er quantile	40	-0.08	72.12	101.2	0	0
Médiane	42	10.31	81.25	101.8	0.43	0
Moyenne	42.83	9.78	80.42	101.8	2.53	19.85
3ème quantile	46	19.33	90.5	102.2	2.9	39.62
Maximum	58	28.97	99	104	33.47	158.2

Tableau II-9 : Résultats du modèle retenu pour expliquer la relation entre les pics d'hospitalisations et la météo à Québec (fins de semaine).

Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
PICS D'HOSPITALISATIONS AVEC LE TEMPS					
Intercept		3,75 (*)	19,10	-200,27	1,02e-2
s(Tmax)	3	2,49			
s(Time)		1,00			
PICS D'HOSPITALISATIONS SANS LE TEMPS					
Intercept		3,75 (*)	13,70	-194,40	1,07e-2
s(Tmax)	3	2,73			
TOUTES LES HOSPITALISATIONS AVEC LE TEMPS					
Intercept		3,18 (*)	9,25	977,73	0,10
s(Tmax)	3	1,20			
s(Time)		4,46			
TOUTES LES HOSPITALISATIONS SANS LE TEMPS					
Intercept		3,18 (*)	1,69	1108,23	0,11
s(Tmax)	3	1,09			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-10 : Résumé numérique des conditions météorologiques associées aux pics de décès à Montréal (jours de semaine).

	HP	Tmax	Hmax	Pmax	Prec	Snow
Minimum	31	-24.24	53	99.63	0	0
1er quantile	32	-9.73	66	101.7	0	10.03
Médiane	33	-4.42	79.5	102.2	0.29	22.54
Moyenne	33.93	-3.03	77.42	102.2	2.84	21.76
3ème quantile	34.75	1.38	87.75	103	2.14	32.5
Maximum	41	33.86	99	104.5	48.12	46.47

Tableau II-11 : Résultats du premier modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (jours de semaine).

Modèle	Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
Premier modèle	PICS DE DÉCÈS AVEC LE TEMPS					
	Intercept		3,52 (*)	38,00	-106,05	4,64e-3
	s(Prec)	0	1,63			
	s(Time)		1,00			
	PICS DE DÉCÈS SANS LE TEMPS					
	Intercept		3,52 (*)	36,80	-107,35	4,37-e-3
	s(Prec)	0	1,6			
	TOUS LES DÉCÈS AVEC LE TEMPS					
	Intercept		17,04 (*)	18,30	38177,14	21,41
	s(Prec)	0	1,00			
	s(Time)		7,15			
	TOUS LES DÉCÈS SANS LE TEMPS					
	Intercept		17,04 (*)	0	39467,28	26,12
	s(Prec)	0	1,00			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-12 : Résultats du deuxième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (jours de semaine).

Modèle	Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
Deuxième modèle	PICS DE DÉCÈS AVEC LE TEMPS					
	Intercept		3,52 (*)	30,40	-99,94	5,46e-3
	s(Hmax)	7	2,26			
	s(Time)		1,00			
	PICS DE DÉCÈS SANS LE TEMPS					
	Intercept		3,52 (*)	30,10	-101,72	5,09e-3
	s(Hmax)	7	2,28			
	TOUS LES DÉCÈS AVEC LE TEMPS					
	Intercept		17,03 (*)	22,20	37808,20	20,41
	s(Hmax)	7	5,34			
	s(Time)		7,49			
	TOUS LES DÉCÈS SANS LE TEMPS					
	Intercept		17,03 (*)	3,65	39170,67	25,17
	s(Hmax)	7	5,13			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-13 : Résumé numérique des conditions météorologiques associées aux pics de décès à Montréal (fins de semaine).

	HP	Tmax	Hmax	Pmax	Prec	Snow
Minimum	31	-26.46	40	100.8	0	0
1er quantile	31	-6.33	74.5	101.7	0.04	4.59
Médiane	33	-0.66	81.92	102	0.35	16.4
Moyenne	33.97	0.36	79.51	102.2	2.14	18.76
3ème quantile	35.25	4.5	89.25	102.9	1.88	28.82
Maximum	53	30.22	99.83	103.3	16.85	63.75

Tableau II-14 : Résultats du premier modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (fins de semaine).

Modèle	Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
Premier modèle	PICS DE DÉCÈS AVEC LE TEMPS					
	Intercept		3,52 (*)	53,60	-62,36	8,87e-3
	s(Tmax)	2	3,11			
	s(Time)		1,00			
	PICS DE DÉCÈS SANS LE TEMPS					
	Intercept		3,52 (*)	66,00	-68,06	8,36e-3
	s(Tmax)	2	6,17			
	TOUS LES DÉCÈS AVEC LE TEMPS					
	Intercept		17,02 (*)	27,20	27878,74	18,35
	s(Tmax)	2	7,80			
	s(Time)		6,36			
	TOUS LES DÉCÈS SANS LE TEMPS					
	Intercept		17,02 (*)	10,00	28896,12	22,6
	s(Tmax)	2	7,83			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-15 : Résultats du deuxième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Montréal (fins de semaine).

Modèle	Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
Deuxième modèle	PICS DE DÉCÈS AVEC LE TEMPS					
	Intercept		3,49 (*)	59,80	-65,07	8,57e-3
	s(Tmax)	2	3,06			
	Prec	1	7,67e-3 (*)			
	s(Time)		1,00			
	PICS DE DÉCÈS SANS LE TEMPS					
	Intercept		3,49 (*)	69,50	-70,81	7,89e-3
	s(Tmax)	2	5,60			
	Prec	1	7,65e-3 (*)			
	TOUS LES DÉCÈS AVEC LE TEMPS					
	Intercept		16,91 (*)	27,40	27873,06	18,33
	s(Tmax)	2	7,80			
	Prec	1	3,91e-2 (*)			
	s(Time)		6,33			
	TOUS LES DÉCÈS SANS LE TEMPS					
	Intercept		16,94 (*)	10,10	28895,20	22,61
	s(Tmax)	2	7,82			
	Prec	1	2,66e-2 (*)			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-16 : Résumé numérique des conditions météorologiques associées aux pics de décès à Québec.

	HP	Tmax	Hmax	Pmax	Prec	Snow
Minimum	7	-21.01	38	99.01	0	0
1er quantile	8	-5.21	65.5	101.1	0.05	0
Médiane	9	10.03	82	102	0.34	0
Moyenne	9.13	7.72	77.88	101.8	3.16	20.66
3ème quantile	10	20.04	90.2	102.5	2.79	37.12
Maximum	13	31.55	100	104.2	35.97	158

Tableau II-17 : Résultats du premier modèle retenu pour expliquer la relation entre les pics de décès et la météo à Québec.

Modèle	Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
Premier modèle	PICS DE DÉCÈS AVEC LE TEMPS					
	Intercept		2,20 (*)	40,50	-73,31	2e-2
	s(Tmax)	0	2,80			
	s(Time)		5,00			
	PICS DE DÉCÈS SANS LE TEMPS					
	Intercept		2,20 (*)	15,90	-61,64	2,20e-2
	s(Tmax)	0	2,80			
	TOUS LES DÉCÈS AVEC LE TEMPS					
	Intercept		3,18 (*)	4,95	45711,56	3,32
	s(Tmax)	0	3,02			
	s(Time)		5,76			
	TOUS LES DÉCÈS SANS LE TEMPS					
	Intercept		3,18 (*)	2,10	46033,85	3,41
	s(Tmax)	0	2,49			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-18 : Résultats du deuxième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Québec.

Modèle	Variable	Lag	Edf ou coefficient estimé	DEX	AIC	GCV
Deuxième modèle	PICS DE DÉCÈS AVEC LE TEMPS					
	Intercept		2,20 (*)	35,80	-70,38	2,06e-2
	s(Tmax)	1	2,20			
	s(Time)		4,69			
	PICS DE DÉCÈS SANS LE TEMPS					
	Intercept		2,20 (*)	14,70	-61.41	2,18e-2
	s(Tmax)	1	2,45			
	TOUS LES DÉCÈS AVEC LE TEMPS					
	Intercept		3,18 (*)	5,03	45699,19	3,31
	s(Tmax)	1	3,96			
	s(Time)		5,74			
	TOUS LES DÉCÈS SANS LE TEMPS					
	Intercept		3,18 (*)	2,16	46022,92	3,41
	s(Tmax)	1	3,23			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-19 : Résultats du troisième modèle retenu pour expliquer la relation entre les pics de décès et la météo à Québec.

Modèle	Variable	Lag	Edf ou coefficient estimé (*)	DEX	AIC	GCV
Troisième modèle	PICS DE DÉCÈS AVEC LE TEMPS					
	Intercept		2,20 (*)	37,50	-73,23	1,94e-2
	s(Snow)	4	1,93			
	s(Time)		4,40			
	PICS DE DÉCÈS SANS LE TEMPS					
	Intercept		2,20 (*)	19,10	-62,91	2,16e-2
	s(Snow)	4	3,39			
	TOUS LES DÉCÈS AVEC LE TEMPS					
	Intercept		3,18 (*)	5,12	45675,59	3,31
	s(Snow)	4	7,28			
	s(Time)		5,08			
	TOUS LES DÉCÈS SANS LE TEMPS					
	Intercept		3,18 (*)	2,32	45994,02	3,41
	s(Snow)	4	6,82			

Le lag optimal, l'edf de la fonction lissée, la déviance expliquée du modèle (DEX en %) et les critères AIC et GCV sont présentés. Lorsque la variable n'est pas introduite sous forme de spline, son edf est remplacé par son coefficient estimé (*) dans la régression.

Tableau II-20 : Résumé des différents modèles retenus pour expliquer les pics d'hospitalisations de janvier 1996 à mars 2007 et de décès de janvier 1981 à décembre 2011, à Montréal et à Québec.

Ville	Variable sanitaire	Séparation	Modèle retenu	DEX (%)
Montréal	Hospitalisations	Aucune	$s(Tmax_5)$	39,60
		Jours de semaine	$s(Prec)$	38,00
	Décès		$s(Hmax_7)$	30,40
		Fins de semaine	$s(Tmax_2) + Prec_1$	59,80
Québec	Hospitalisations	Jours de semaine	$s(Prec_7) + Snow$	38,70
		Fins de semaine	$s(Tmax_3)$	19,10
	Décès	Aucune	$s(Tmax)$	40,50
			$s(Tmax_1)$	35,80
			$s(Snow_4)$	37,50

L'indice numérique représente le lag sélectionné de la variable auquel il est associé.

Annexe 1 : Figure illustrative

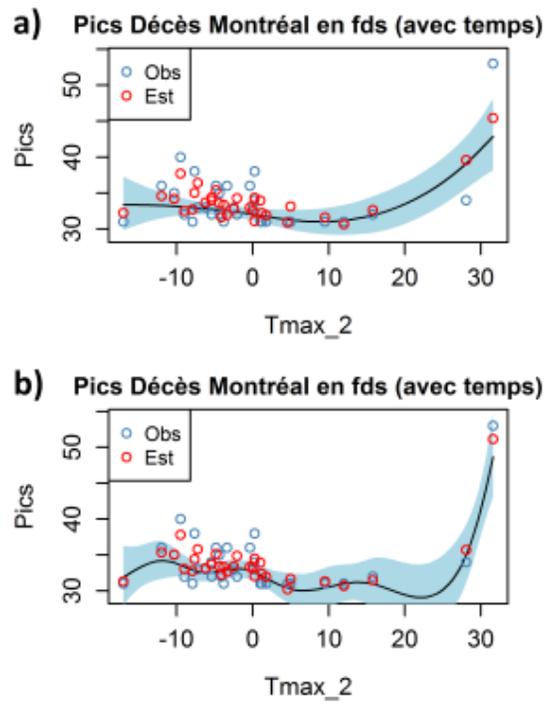


Figure II-35 : Comparaison des fonctions lissées de Tmax (lag de 2 jours) pour les pics de décès à Montréal en fins de semaine, avec et sans la fonction du temps : a) approche de Kim et Gu (2004), b) approche sans paramètre de correction.

Annexe 2 : Modèle additif généralisé

Le GAM est une extension du modèle linéaire généralisé (GLM), qui étend lui-même le modèle linéaire. Le modèle de régression linéaire permet d'expliquer une variable réponse Y^* (ici la variable sanitaire, l'étoile indique le caractère extrême), supposée être distribuée comme une loi normale, en fonction de la somme de variables explicatives X_1, \dots, X_n (ici les variables météorologiques) et d'une erreur normale ε de moyenne nulle et de variance σ^2 :

$$Y^* = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_n X_n + \varepsilon \quad (12)$$

L'espérance conditionnelle de Y^* par rapport à $X_1 \dots X_n$ est obtenue par :

$$E[Y^* | X] = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_n X_n \quad (13)$$

Dans un tel modèle, Y^* dépend linéairement de l'ensemble des $X_1 \dots X_n$. Cette dépendance peut être relâchée dans le GLM (Nelder et Wedderburn, 1972) grâce à l'introduction d'une fonction continue, monotone et deux fois différentiable g , appelée fonction de lien :

$$g(E[Y^* | X]) = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_n X_n \quad (14)$$

De plus, l'hypothèse de normalité de Y^* est remplacée par une hypothèse plus large, celle d'appartenance de la distribution de Y^* à la famille exponentielle (qui contient entre autres la loi normale). Le GAM étend par la suite le modèle GLM en utilisant des fonctions non linéaires pour représenter les variables explicatives.

Hastie et Tibshirani (1986) proposent un algorithme de *backfitting* pour l'estimation des fonctions f_j . Dans cette étude, l'estimation se fait plutôt par l'approche de Wood (2006, chapitre 4) grâce à un algorithme des moindres carrés pénalisés réitérés (ou *penalized iteratively re-weighted least squares*, P-IRLS, en anglais). Cela consiste à répéter

l'étape classique des moindres carrés, augmentés d'une pénalisation (qui contrôle l'équilibre de la fonction, entre lissage et adéquation aux données) et d'une pondération (qui permet de s'assurer que la variance des observations reste constante). L'avantage de l'algorithme de backfitting est qu'il s'adapte à n'importe quelle méthode de régression (comme les arbres de régression). Cependant, il ne permet pas d'estimer le degré de lissage associé à la fonction d'une variable explicative et cela peut s'avérer moins intéressant dans un contexte explicatif. En effet, il est souhaitable de pouvoir contrôler ce degré pour pouvoir obtenir des fonctions qui ne s'adaptent pas trop aux données, dans un contexte de petits échantillons où les fonctions sont plus susceptibles de varier plus rapidement.

III. ÉTUDE DES RELATIONS ENTRE LES PICS DE MALADIES CARDIOVASCULAIRES ET LES CONDITIONS MÉTÉOROLOGIQUES PAR LA RÉGRESSION QUANTILE

Y. Chiu ^a, F. Chebana ^a, B. Abdous ^b, D. Bélanger ^{a, c}, P. Gosselin ^{a, d}

^a *INRS-ETE, 490 rue de la Couronne, Québec (QC), G1K 9A9, Canada*

^b *Université Laval, Département de médecine sociale et préventive, 2325 Rue de l'Université, Québec (QC), G1K 7P4, Canada*

^c *Centre hospitalier universitaire de Québec, Centre de recherche, 2875 boulevard Laurier, Québec (QC), G1V 2M2, Canada*

^d *Institut national de santé publique du Québec, 945 avenue Wolfe, Québec (QC), G1V 5B3, Canada*

Rapport accepté à l'INSPQ le 20 mars 2017

Résumé

Les études qui s'intéressent aux décès et hospitalisations par maladies cardiovasculaires (MCV) en lien avec les conditions météorologiques se concentrent habituellement sur les événements moyens ou sur des événements ponctuels (la canicule européenne de 2003 p. ex.). Cependant, les pics sanitaires (aussi appelés « pics » dans la suite du texte), événements rares et éloignés de la moyenne, ne sont pas expliqués spécifiquement par ces analyses. En particulier, pour ces derniers, les relations avec les variables explicatives, comme les variables météorologiques, peuvent être mal interprétées. Les modèles utilisés dans la littérature considérant toute la série se basent sur la moyenne conditionnelle. Cette dernière n'est en effet qu'une mesure centrale et ne présente qu'un portrait incomplet des relations entre la variable sanitaire et les variables météorologiques. Cette étude a ainsi pour but d'étudier les pics de la variable sanitaire et ses relations avec la météo. Les hospitalisations et décès par MCV sont étudiés en relation avec les variables météorologiques dans différents quantiles (et plus spécifiquement au travers de quantiles élevés pour les pics) de la variable sanitaire grâce à la régression quantile dans les communautés métropolitaines de Montréal et de Québec (Canada). Les résultats montrent des effets non constants des variables météorologiques selon le quantile considéré. À titre d'exemple, l'effet de la température est souvent beaucoup plus fort dans les pics que dans les quantiles faibles ou médians de l'issue sanitaire. Ces effets sont habituellement cachés dans les résultats d'une régression classique. Les résultats obtenus confirment l'intérêt d'étudier les variables sanitaires non plus uniquement au travers des événements observés en moyenne, mais bien de les compléter par une analyse des pics aussi. Une analyse des pics peut être plus intéressante afin d'établir de meilleures alertes sanitaires et météorologiques aux fins de surveillance de la santé publique, par exemple.

1. Introduction

Les maladies cardiovasculaires (MCV) représentent un enjeu de santé publique majeur au Québec puisqu'elles y causent près d'un cinquième des décès totaux. Elles représentent ainsi la deuxième cause de décès au Québec (Girard, 2013) et figurent par ailleurs à la première place des causes d'hospitalisation (Daigle, 2007). La situation est généralisée puisque selon l'Organisation mondiale de la santé, les MCV sont la première cause de mortalité dans le monde en 2015 (Organisation mondiale de la santé, 2015). La compréhension, le traitement et la prévention des MCV constituent ainsi des défis importants et plusieurs d'études existent à ces sujets, voir p. ex. les revues de littérature de Taylor *et al.* (2011) ou Wang *et al.* (2014).

Différents facteurs peuvent influencer les MCV. En effet, les habitudes de vie telles que la consommation de tabac, l'alimentation ou la pratique régulière de sport (Dinas *et al.*, 2013; Afshin *et al.*, 2014; Vasconcellos *et al.*, 2014) sont reconnues pour jouer un rôle important. La pollution atmosphérique (ozone, particules fines) est aussi proportionnellement associée à la survenue de MCV (Hoek *et al.*, 2013). Morris (2001) indique que la part de la pollution de l'air sur les décès par MCV est potentiellement majeure. De façon plus marginale, le bruit est aussi mis en cause dans l'apparition des MCV (Van Kempen *et al.*, 2002).

Il existe aussi une littérature importante sur les liens entre les conditions météorologiques et les MCV (voir p. ex. la revue de littérature dans la discussion de Ishikawa *et al.*, 2012). En particulier, la température est reconnue pour avoir un effet sur la mortalité (Basu, 2009) et sur la morbidité (Phung *et al.*, 2016b). Les effets de l'exposition au froid, des vagues de chaleur et des variations dans les températures diurnes sont significatifs. Bayentin *et al.* (2010) montrent que les liens entre les conditions météorologiques et la santé par MCV sont présents au Québec. Or, le changement climatique, en modifiant la distribution des variables météorologiques, risque aussi

de changer la relation avec les MCV. Au Québec, une hausse significative des températures minimales et maximales est, entre autres, attendue d'ici 2050 (Ouranos, 2015). Cela aura pour effet d'augmenter fortement la durée des vagues de chaleur ainsi que la fréquence des nuits chaudes, des facteurs reliés à l'apparition de MCV.

Toutes les études précédemment citées utilisent des modèles qui étudient le caractère moyen des MCV, et donc basés sur l'ensemble des observations. Il est nécessaire de prévoir les moyennes ou les tendances, notamment pour la surveillance des MCV ou pour la gestion quotidienne des établissements de santé. Cependant, il existe des événements extrêmes appelés *pics sanitaires* (ou « pics » dans la suite du texte). Il s'agit d'événements rares et très éloignés des événements moyens. Ils ont donc une probabilité de survenue bien moindre que les événements moyens, mais ont en contrepartie un impact beaucoup plus fort sur le système de la santé (p. ex. engorgement des salles d'urgences dans les hôpitaux, Knowlton *et al.*, 2009).

Très peu d'études en santé publique s'intéressent au caractère extrême des variables sanitaires d'un point de vue de la modélisation. Il est possible de citer De Zea Bermudez et Mendes (2012) et Chiu *et al.* (2016a). Ainsi, malgré le fait que les pics aient de fortes répercussions, ils ne sont généralement pas étudiés de façon exclusive. Ils sont la plupart du temps sélectionnés de façon ponctuelle et descriptive, comme lors de la canicule européenne de 2003 (Robine *et al.*, 2008). De plus, dans un contexte de changement climatique, il est essentiel de bien quantifier la relation entre les pics et les conditions météorologiques de sorte à favoriser l'adaptation des populations et des systèmes de services et de soins médicaux. Le but de ce chapitre n'est ainsi plus d'étudier la réponse moyenne, mais plutôt la réponse des pics. Une option possible est d'extraire les pics puis de les modéliser, comme dans Chiu *et al.* (2015). Cette méthodologie utilise deux catégories de modèles classiques (régression et théorie des valeurs extrêmes). Elle nécessite cependant

plusieurs étapes de prétraitement et de tests statistiques. Une alternative plus directe pour étudier les pics en fonction de variables explicatives est la régression quantile (RQ, Koenker, 2005).

La régression quantile est similaire à une régression classique dans le sens où l'estimation d'un coefficient indique l'ampleur de l'effet de la variable explicative sur la variable à expliquer. La différence principale est que tous les quantiles de la réponse peuvent être étudiés directement grâce à la régression quantile, contrairement à seulement la moyenne conditionnelle dans une régression classique. Or, les pics peuvent être définis comme des quantiles d'ordre élevé q (p. ex. $q=0.90$ ou $q=0.95$, c.-à-d. des observations qui sont supérieures à 90 % ou à 95 % des observations ordonnées). La régression quantile permet donc une description plus complète de la distribution réponse conditionnelle. Cade et Noon (2003) montrent que s'il y a une faible (ou aucune) relation entre la moyenne conditionnelle de la variable réponse et les variables explicatives, il est tout de même possible que d'autres relations significatives existent dans le reste de la distribution conditionnelle. En fait, étudier uniquement la moyenne d'une distribution conditionnelle « dilue » l'information liée au reste de la distribution, en particulier des pics sanitaires. Dans un contexte de relations entre les conditions météorologiques et les MCV, il serait ainsi possible d'observer un effet significatif de la température dans les pics de la variable sanitaire, alors qu'il serait non significatif pour la médiane ($q=0.50$). Cette influence non homogène ne serait pas mise en évidence par une régression classique.

La régression quantile est reconnue dans plusieurs problématiques de la santé publique, en particulier pour des indicateurs susceptibles de varier dans la distribution de la variable sanitaire. À titre d'exemple, dans une étude sur les impacts du stress et de la qualité du sommeil sur l'indice de masse corporelle (indice), Yang *et al.* (2014) montrent que les deux variables n'ont pas un effet constant sur l'indice dans les différents quantiles. En particulier, l'effet est plus

marqué dans les quantiles élevés de la distribution de l'indice. Dans l'étude de Bergh *et al.* (2015), les auteurs indiquent que les participants déjà en surpoids (IMC élevé) ont tendance à gagner du poids plus rapidement que ceux avec un indice plus bas. Le lien entre la pollution de l'air et la santé est aussi particulièrement étudié à l'aide de la régression quantile. Par exemple, Bind *et al.* (2015) et Bind *et al.* (2016) étudient la façon dont les particules fines et le charbon influencent certains biomarqueurs connus des MCV. Leurs résultats leur permettent de conclure que les effets de la pollution peuvent être prononcés parmi les individus déjà fortement à risque, autrement dit dans les pics.

La RQ peut s'apparenter graphiquement à d'autres méthodes telles que les courbes de dose-réponse (Ahrens et Pigeot, 2014). Effectivement, ces courbes permettent de quantifier l'effet d'une variable d'exposition sur la réponse selon différentes « doses ». Elles peuvent en fait être considérées comme un cas particulier de régression spécifiquement adapté à l'épidémiologie ou la pharmacologie, où elles sont répandues. Toutefois, la RQ en diffère puisqu'elle n'estime pas une seule réponse, mais plutôt l'ensemble des réponses dans les différents quantiles de la variable réponse. La RQ et les courbes de dose-réponse présentent donc des résultats et des utilisations très différents.

La régression quantile est aussi utilisée en météorologie ou climatologie. Jagger et Elsner (2009) modélisent l'intensité des cyclones tropicaux dans plusieurs quantiles et en fonction de diverses variables météorologiques. Ils rapportent que l'influence de l'activité solaire est la plus forte près de la médiane de la distribution de l'intensité des cyclones, mais que cette relation change de signe dans les pics. Elsner *et al.* (2008) étudient la vitesse des vents des cyclones tropicaux. Leurs résultats montrent que s'il y a une tendance à la hausse dans l'intensité des cyclones, cette

dernière n'est pas homogène et semble plus forte dans les pics. À nouveau, dans ces deux études, une régression classique n'aurait pas pu accéder à ce type de résultats détaillés.

À noter, il existe aussi des études qui s'intéressent aux quantiles extrêmes de façon théorique dans un contexte de régression quantile, voir p. ex. Beirlant *et al.* (2004b) ou Daouia *et al.* (2013). Ces méthodologies ne sont pas applicables ici, mais il est intéressant de savoir que les quantiles extrêmes présentent des comportements suffisamment différents pour motiver de telles études.

La présente étude se différencie des études citées, principalement sous deux aspects. Premièrement, si beaucoup d'études font usage de la régression quantile en climatologie ou météorologie et en santé publique, aucune ne lie la santé aux conditions météorologiques, à plus forte raison les MCV. Deuxièmement, très peu d'études portent leur attention sur les pics sanitaires, encore moins sur les pics sanitaires dans des modèles de régression. Chiu *et al.* (2016b), qui utilisent des modèles de régression semi paramétriques, en font partie, mais une étape d'extraction des pics est d'abord nécessaire dans leur méthodologie, contrairement à la régression quantile. Enfin, aucune étude faisant usage de la régression quantile n'inspecte en détail les relations comme il est fait ici, soit dans l'ensemble des quantiles (évaluation sur 40 quantiles) et sur quatre effets retardés (nommés lags dans ce texte).

Le reste du chapitre est organisé comme suit. Les données disponibles ainsi que la méthodologie de l'étude sont décrites dans la Section 2. La Section 3 présente l'application et les résultats obtenus. Enfin, la Section 4 contient la discussion, la Section 5 la conclusion et la Section 6 les limitations de l'étude.

2. Analyse statistique

Les modèles statistiques de ce chapitre repose sur les techniques liées à la régression quantile, introduite par Koenker et Bassett (1978). Dans la suite du texte, q fait référence à un ordre de quantile (avec $0 < q < 1$). Pour plus de détails théoriques sur cette régression, le lecteur est référé à l'ouvrage de Koenker (2005).

La régression quantile est un modèle de régression linéaire dans le sens où elle permet d'expliquer de façon linéaire l'influence d'une ou plusieurs variables explicatives sur une variable réponse. Elle est cependant très différente de la régression linéaire classique. Pour illustrer l'intérêt et les caractéristiques de la régression quantile, un rappel sur la régression par moindres carrés ordinaires est utile. Les moindres carrés ordinaires minimisent la somme des carrés résiduels pour estimer la meilleure droite explicative (Dekking *et al.*, 2005, chapitre 22). La régression par moindres carrés ordinaires décrit la façon dont la moyenne conditionnelle d'une variable réponse varie en fonction d'un changement de la variable dépendante. Pour une variable à expliquer Y et des variables explicatives X_1, \dots, X_p , elle est définie généralement par :

$$E[y_i | x_{1i}, x_{2i}, \dots, x_{pi}] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} \quad (15)$$

L'interprétation du modèle est la suivante : une augmentation d'une unité de X_1 augmente Y de β_1 unités, toutes choses égales par ailleurs.

Il s'avère que les moindres carrés ordinaires sont non robustes, c.-à-d. que les estimations peuvent être affectées suite au non-respect d'une des hypothèses de base (Box, 1953). Les moindres carrés ordinaires sont ainsi facilement influencés par les données aberrantes. Cela vient du fait que la démarche est basée sur la moyenne conditionnelle. Un moyen de remédier à cela est de passer par les quantiles. En effet, la médiane est robuste comparativement à la moyenne,

laquelle est influencée par les valeurs éloignées (minimales ou maximales). L'exemple suivant l'illustre : considérant une série fictive d'observations de température {25;26;25;27;28;29;22} , la moyenne et la médiane sont toutes les deux égales à 26. En rajoutant la valeur éloignée -20 à la série, la moyenne tombe à 20,25 tandis que la médiane devient 25,5. Cela vient du fait que les *quantiles*, qui sont basés sur le rang des observations, sont moins sujets à changement que les *valeurs* des observations. Il est ainsi possible d'obtenir des moyennes non représentatives d'une série, dû à la présence de valeurs aberrantes. Par exemple, la moyenne de la quantité de neige au sol à Québec est de 17,67 mm tandis que la médiane est de 0 mm (Tableau 0-3). La robustesse de la régression quantile a déjà été mise à profit en santé publique pour l'étude des méfaits de la fumée secondaire (MacNaughton *et al.*, 2016).

Pour un quantile $0 < q < 1$, la régression quantile est définie par :

$$Q_{y_i} [q | x_{1i}, x_{2i}, \dots, x_{pi}] = \beta_0(q) + \beta_1(q)x_{1i} + \beta_2(q)x_{2i} + \dots + \beta_p(q)x_{pi} \quad (16)$$

avec une variable à expliquer Y et des variables explicatives X_1, \dots, X_p . Contrairement à l'équation (15), les paramètres $\beta_0(q), \beta_1(q), \dots, \beta_p(q)$ dépendent cette fois du quantile dans lequel l'estimation a lieu.

L'interprétation des paramètres estimés de la régression quantile est semblable à celle d'une régression par moindres carrés ordinaires. Il y a cependant une différence fondamentale entre les deux, à savoir la régression quantile doit se faire dans chaque quantile séparément, c'est-à-dire qu'une augmentation d'une unité de X_1 augmente Y de $\beta_1(q)$ unités dans le quantile q , toutes choses égales par ailleurs. Il est important de noter que l'estimation des paramètres de la régression quantile passe par des méthodes de calcul bien distinctes des moindres carrés ordinaires (optimisation linéaire et résolution d'équations linéaires, respectivement). Pour ces

derniers, la somme des carrés résiduels est minimisée, tandis que pour la régression quantile, Koenker et Bassett (1978) proposent de minimiser une somme des écarts absolus pondérés (voir l'annexe Régression quantile : estimation).

De plus, les deux régressions (lorsque la médiane est considérée pour la régression quantile) peuvent parfois mener à des résultats très similaires, surtout dans le cas où la distribution conditionnelle de la variable réponse est une distribution symétrique, la médiane étant égale à la moyenne (Marrie *et al.*, 2009).

La régression quantile présente certains avantages par rapport à une régression classique, qui justifient son usage dans cette étude. En effet, elle ne présume pas de distribution particulière quant aux erreurs de la régression (contrairement à la régression classique, p. ex. loi normale). De plus, elle permet de sauter les étapes de sélection des données (et les possibles erreurs associées) ou de vérification de loi théorique implémentées par Chiu *et al.* (2015). Elle est donc plus directe et permet de comparer les relations de la variable à expliquer avec les variables explicatives dans l'ensemble des quantiles, ce qu'aucune autre méthode n'est capable de faire. Ceci permet entre autres de déceler si les effets des variables explicatives sont constants ou non le long des quantiles. Enfin, c'est une régression robuste, dans le sens où elle n'est pas influencée par les valeurs aberrantes.

Contrairement à l'étude de Chiu *et al.* (2016b), les lags sur les variables météorologiques ne peuvent pas être différents dans les quantiles au sein d'un même modèle de régression. Par exemple, dans un modèle reliant la température aux hospitalisations, il n'est pas possible de sélectionner un lag de 3 jours au quantile $q = 0.50$ puis un lag de 5 jours au quantile $q = 0.90$. En effet, les coefficients estimés de la température ne seraient alors plus comparables entre eux. Pour ces raisons, les lags considérés dans la présente étude sont fixés à 0 (jour même), 3, 7 et

14 jours mais sur plusieurs quantiles. Ainsi, par exemple, la régression quantile appliquée dans le quantile $q = 0.90$ donnerait lieu à 4 modèles différents, c'est-à-dire un modèle pour chaque lag retenu.

En plus des variables météorologiques, les modèles de régression quantile dans cette étude incluent un coefficient du temps (variable *DATE* dans les graphiques) dans tous les modèles. Ce coefficient permet de « capter » l'influence du temps comme les tendances et la saisonnalité éventuelles. Ceci est une pratique courante dans les études des relations santé et météorologie (voir p. ex. Schwartz *et al.*, 2004 ou Lim *et al.*, 2012). Les intervalles de confiance (IC) associés aux paramètres estimés de la régression quantile, obtenus grâce à la méthode d'inversion des rangs (Koenker, 2005), sont aussi présentés. Ces IC sont utiles pour déterminer si l'effet d'une variable explicative dans un certain quantile est significatif ou non.

Plusieurs variables météorologiques sont utilisées dans un même modèle de régression, contrairement à l'étude de Chiu *et al.* (2016b), qui avait des échantillons très limités, réduisant ainsi les possibilités d'inclure plus d'une variable explicative. En raison de leurs unités de mesure très variées (voir Tableau 0-3), ces variables doivent être nécessairement standardisées pour être ramenées sur des échelles comparables.

La Figure III-2 illustre des exemples de résultats de la régression quantile sur les hospitalisations à Montréal, aux quantiles d'ordres 25, 50, 75, 90, 95, 97.5 et 99 %. La Figure III-2a montre les droites de régression dans les différents quantiles des hospitalisations en fonction de la température. Parallèlement, la Figure III-2b montre l'ordonnée à l'origine (*intercept*), qui est le croisement de la droite de régression en y au point $x=0$, ainsi que le coefficient directeur estimé de la température en fonction du quantile. Les coefficients varient selon les quantiles ce qui résulte en des droites non parallèles et indique un effet de la température différent selon le

quantile considéré. L'effet dans le quantile à 75 % est par exemple plus important que dans le quantile à 25 % (en valeur absolue, la valeur du coefficient directeur du premier est supérieure). Quant aux IC des paramètres estimés ainsi que la droite de régression par moindres carrés ordinaires et son IC, présentés par exemple sur la Figure 4, ils permettent de comparer avec les effets de la régression quantile.

Les quantiles considérés dans cette étude s'échelonnent entre $q=0.01$ et $q=0.99$, avec un pas de 2,5. Cela résulte en une exploration sur 40 quantiles de la variable sanitaire, menant à une analyse des relations dans un niveau de détail que seule la régression quantile permet. Contrairement à l'étude de Chiu *et al.* (2016b), il sera ainsi possible d'étudier l'effet des variables météorologiques non seulement dans les quantiles élevés mais aussi dans les quantiles médians ou faibles. Il n'existe pas de définition officielle des pics sanitaires, mais en se basant sur leurs équivalents météorologiques (IPCC, 2013), il est possible de les qualifier comme tels à partir du quantile à 90 %.

De la même façon que dans Chiu *et al.* (2016b), seuls les résultats avec les variables maximales sont décrits. En effet, les résultats avec les variables moyennes n'ont pas apporté de changements significatifs. Par ailleurs, les relations moyennes entre les variables sanitaires et les variables météorologiques maximales semblent de fait similaires à celles avec les variables météorologiques minimales (voir la Figure III-1). Les résultats montrent toutefois un impact différent concernant certaines variables minimales et lorsque c'est le cas, ils sont présentés dans la Section 3. L'application de la régression quantile permettra d'analyser ces relations hors de la moyenne et de dévoiler les différences entre variables maximales et minimales, s'il y a lieu. Les figures des modèles avec les variables minimales sont présentées en annexe.

Les résultats sont obtenus à l'aide du logiciel libre R et du package *quantreg* (v5.26, Koenker, 2016). Le niveau de significativité statistique est fixé à $\alpha = 5\%$.

3. Résultats

Cette section contient une description des résultats des estimations de la régression quantile pour les hospitalisations et les décès, pour Montréal et Québec (pour rappel, il s'agit des communautés métropolitaines). Un résumé des résultats obtenus est présenté à la fin de cette section. Les variables météorologiques sont standardisées tel que mentionné dans la section Analyse statistique.

3.1. Hospitalisations à Montréal

La méthodologie est appliquée sur les hospitalisations à Montréal. Les résultats sont présentés à la Figure III-3. Comme décrit dans l'exemple de la section 2, l'ensemble des coefficients estimés pour chaque variable météorologique est résumé sous forme de graphique. Il est important de rappeler que ce ne sont pas les droites de régression à proprement parler qui sont représentées sur les graphiques; il s'agit plutôt des coefficients directeurs de ces droites. Tracer l'ensemble des droites de régression rendrait les figures illisibles (présence de 40 quantiles, donc 40 droites de régression).

L'allure que prend la courbe des coefficients estimés de la température reste similaire au travers des différents lags (0, 3, 7 et 14 jours). La courbe commence par un coefficient négatif proche de 0, diminue jusqu'au quantile à 10 %, remonte jusqu'au quantile à 40 % puis redescend jusqu'aux quantiles les plus élevés. Les coefficients ne sont pas significatifs au quantile à 1 % et autour du quantile à 40 %. Dans les autres quantiles, une augmentation de la température est associée à une diminution des hospitalisations car les coefficients sont estimés négatifs. Il n'y a majoritairement

pas de différence significative entre les estimations par moindres carrés ordinaires et par régression quantile car les IC se croisent. Il y a toutefois une exception à partir du lag 7, dans les pics supérieurs au quantile à 95 %. Dans cette zone des pics, les différences sont significatives et une régression par moindres carrés ordinaires pourrait sous-estimer la véritable influence de la température. De plus, les pics possèdent les valeurs de coefficients estimés les plus élevées (de deux à trois fois plus élevées en valeur absolue que les estimations dans les quantiles médians).

La relation entre les hospitalisations et l'humidité est inverse à tous les lags (coefficients estimés négatifs), hormis dans quelques cas comme les pics aux lags 3, 7 et 14. Toutefois, les coefficients sont non significatifs la plupart du temps. Il existe quelques exceptions, comme aux lags 7 et 14, autour de la médiane. En particulier dans les pics (quantiles supérieurs à 95 %) au lag 7, l'effet de l'humidité est significatif et inverse. Dans l'ensemble, l'effet estimé par régression quantile n'est pas significativement différent de celui estimé par moindres carrés ordinaires.

L'effet de la pression ne semble pas significatif, ni différent d'une régression par moindres carrés ordinaires avant le lag 7. À partir du lag 7, les IC se croisent toujours mais les coefficients de la régression quantile deviennent significatifs dans les quantiles supérieurs à 40 %. Dans les pics, l'effet semble plus important (visible particulièrement au lag 7). Il est intéressant de noter que les coefficients estimés de la pression atmosphérique sont tous positifs, indiquant une hausse des hospitalisations avec une augmentation de la pression pour les lags de 3, 7 et 14 jours. Au jour 0, bien que les coefficients soient majoritairement négatifs, ils sont non significatifs.

Les coefficients estimés des précipitations sont majoritairement négatifs. L'effet de la variable des précipitations ne semble pas significatif aux lags et quantiles considérés, comme le témoignent les IC qui contiennent en majorité la ligne $y = 0$. Une partie des coefficients est

cependant significative dans tous les lags, autour du quantile à 80 %. Les IC des estimations par régression quantile et moindres carrés ordinaires se croisent pour toutes les estimations.

L'allure de la fonction des coefficients estimés de la neige est semblable à celle de la température. Les coefficients sont élevés et positifs autour des quantiles à 1 % et à 40 %. Ils sont majoritairement significativement différents de 0, sauf dans les pics (à partir du quantile à 95 %). Aux lags 7 et 14, la plupart des coefficients deviennent non significatifs alors que ceux des pics le sont, avec une relation négative (coefficients estimés négatifs) et un effet plus important que sur le reste des quantiles (hormis le quantile à 40 %, qui possède un coefficient similaire en valeur absolue mais positif). De plus, dans les pics, les effets de la régression quantile y sont différents de ceux des moindres carrés ordinaires.

3.2. Hospitalisations à Québec

La régression quantile est appliquée aux hospitalisations à Québec. Les résultats sont présentés à la Figure III-4.

La courbe des estimations des coefficients de la température est décroissante jusqu'aux pics du quantile à 90 %, puis remonte de façon plus ou moins marquée selon le lag considéré. En effet, le coefficient estimé passe de -1.6 à -0.8 sans lag (diminution de 50 %) et de -2.2 à -1.8 au lag 14 (diminution de 20 %). Dans tous les cas, les estimations sont les plus fortes (en valeur absolue) dans le pic à 90 % et sont toutes négatives. Cela indique une relation inverse plus grande juste avant le quantile à 90 % et un effet moins marqué (mais toujours inverse) dans les pics supérieurs. Les estimations sont significativement différentes de 0, sauf aux quantiles à 30 % et 99 % du lag 0. Les coefficients ne sont cependant pas statistiquement différents de ceux estimés par moindres carrés ordinaires car les IC se croisent.

L'humidité est significative uniquement aux lags 0 et 3, autour du quantile à 60 %. Sauf quelques exceptions, la relation est majoritairement dans le même sens (coefficients estimés positifs). Aux lags 7 et 14, les pics à partir du quantile à 92.5 % exhibent de fortes valeurs positives. L'effet de l'humidité estimé par régression quantile n'est toutefois pas différent de celui estimé par moindres carrés ordinaires.

La courbe des estimations pour la pression est croissante (avec une première partie négative) jusqu'au quantile à 70 %, pour ensuite devenir décroissante, dans tous les lags. Elle n'est majoritairement pas significative (tous les IC incluent la valeur nulle) et non différente de la régression par moindres carrés ordinaires. Il y a quelques exceptions significatives dans les quantiles autour de 60 % au lag 3 (où au contraire, la régression par moindres carrés ordinaires est non significative). L'effet de la pression minimale est significatif au jour même, particulièrement dans les pics où les coefficients sont les plus forts (Figure annexe 2). La version minimale contraste en ce sens avec la version maximale.

Les précipitations ont un effet significatif aux lags 3 et 14, ainsi que de façon ponctuelle au lag 0 (quantile à 10 %) et au lag 7 (quantile à 40 %). De façon générale, une augmentation des précipitations est liée à une diminution des hospitalisations, hormis dans les pics (supérieurs au quantile à 97.5 %) où la relation s'inverse alors car les coefficients deviennent positifs. En revanche, dans ces quantiles élevés, l'effet ne semble pas significatif. Les estimations ne sont pas différentes entre la régression quantile et les moindres carrés ordinaires.

L'effet de la neige est significatif à deux reprises, au lag 0 (jour même) entre les quantiles à 20 % et 40 % ainsi qu'au lag 14 entre les pics à 80 % et 90 %, ce qui oppose les deux lags en termes de quantiles significatifs (bas quantiles et pics sanitaires). De plus, il est à noter qu'au jour même, les coefficients estimés sont positifs tandis qu'au lag 14, ils sont négatifs. Par ailleurs, les IC des

estimations par régression quantile et moindres carrés ordinaires se croisent, ce qui ne donne pas la possibilité de distinguer entre les deux estimations.

3.3. Décès à Montréal

Les coefficients estimés de la régression quantile sur les décès à Montréal sont présentés à la Figure III-5.

Une augmentation de la température semble entraîner une diminution des décès (coefficients négatifs). Cet effet est de plus significatif dans tous les quantiles et lags. Il est aussi plus important en se déplaçant dans les pics et dans les lags plus élevés. En effet, les coefficients de régression atteignent des valeurs de 1,5 à 2 fois plus fortes que dans les quantiles bas ou médians. Toutefois, cet effet n'est visible que dans les pics. En effet, le coefficient estimé dans le quantile à 1 % est autour de -1 pour tous les lags tandis que le coefficient dans le quantile à 99 % débute à -1.1 (au jour même) pour finir égal à -1.8 (lag 14). De plus, au lag 14, les coefficients estimés dans les quantiles bas et élevés sont significativement différents de ceux obtenus par moindres carrés ordinaires.

La pression n'est significative qu'au jour même, entre la médiane et le pic à 95 %. Les coefficients γ sont estimés négatifs. Aux autres lags, les coefficients deviennent positifs mais non significativement différents de 0. Les IC des coefficients par régression quantile et moindres carrés ordinaires se croisent. À noter, les coefficients estimés dans les pics à partir du quantile à 92.5 % possèdent les plus fortes valeurs. La pression minimale est différente ici, puisqu'elle est majoritairement significative au jour même (Figure annexe 3). En particulier, elle l'est dans les pics, avec des coefficients nettement plus bas que ceux dans les autres quantiles. Cela indique une relation inverse plus marquée entre la pression et les décès dans les pics.

La courbe des coefficients de l'humidité relative semble constante ou légèrement décroissante dans les quantiles. L'humidité contraste avec la pression en étant non significative au jour même mais significative aux lags de 3, 7 et 14 jours. La relation est inverse, avec une plus grande intensité dans les pics. De plus, le pic à 99 % possède un coefficient qui devient de plus en plus important du lag 0 au lag 7 (de -0.5 à -1.5). Les coefficients estimés par la régression quantile et moindres carrés ordinaires ne sont pas significativement différents.

La courbe des coefficients des précipitations est stable ou légèrement croissante. Les estimations oscillent entre valeurs négatives et positives, sans être significatives au long des quantiles et des lags. Seule une petite partie est jugée différente de 0, soit au lag 14 en dessous du quantile à 12.5 %. Les IC par régression quantile et moindres carrés ordinaires se croisent dans toutes les estimations.

La quantité de neige au sol est significative aux lags 0, 3 et 7 dans la plupart des quantiles, après les quantiles 1 % à 5 %. La neige est par ailleurs associée à plus de décès dans les pics, où les coefficients sont les plus élevés. L'effet n'est plus significatif après 2 semaines. De plus, au jour même, après le quantile à 90 %, les effets estimés dans les pics par régression quantile sont différents (sous-estimés) de ceux par moindres carrés ordinaires.

3.4. Décès à Québec

Les résultats des estimations par régression quantile sont présentés à la Figure III-6.

Les quantiles d'ordre 1, 10, 30 et 50 % sont remarquables dans le cas des décès à Québec. En effet, les estimations des paramètres sont toujours nulles, pour chaque variable explicative. Une des explications pourrait être la suivante. Comme indiqué dans le Tableau 0-2, les décès journaliers vont de 0 à 13, ce qui laisse une étendue de valeurs possibles beaucoup plus limitée

que dans les 3 autres cas (hospitalisations à Québec et Montréal et décès à Montréal). Ainsi, plusieurs quantiles proches peuvent être égaux (par exemple, les quantiles d'ordre 25, 27.5 et 30 %). Cela ne signifie pas pour autant que l'effet est nul, il peut être significatif mais puisque la valeur du quantile est la même, l'estimation ne se fait pas.

La courbe des estimations des coefficients de la température a la même allure pour les 4 lags. Hormis les quantiles jugés remarquables, la courbe est décroissante jusqu'au quantile à 90 %, pour lequel les estimations augmentent pour diminuer à nouveau. Cette remontée s'atténue au fil des lags et n'existe plus au lag 14. De plus, les estimations sont négatives, ce qui indique un effet inverse de la température sur les décès. L'effet s'accroît dans les pics avec les lags (l'estimation au dernier quantile passe de -0.26 le jour même à -0.41 au lag 14), ce qui signifie que plus un quantile et un lag élevés sont considérés, plus la température est associée à une diminution des décès. Même si l'effet est plus fort dans les pics, les estimations ne semblent pas significativement différentes de celles des moindres carrés ordinaires.

Les estimations des paramètres de la pression fluctuent autour de 0 avant de diminuer et de demeurer négatives dans les pics, ceci dans tous les lags. Elles ne sont cependant pas significatives, sauf quelques exceptions au jour même. Il en va de même pour les estimations par moindres carrés ordinaires, desquelles celles par régression quantile ne semblent pas différentes. La pression minimale au jour même est très différente (Figure annexe 4). Même si les coefficients estimés sont aussi négatifs, elle est significative dans tous les quantiles. Par ailleurs, les coefficients atteignent leurs plus fortes valeurs dans les pics. Comme pour les décès à Montréal, cela indique une relation inverse marquée en particulier dans les pics.

Les estimations des coefficients de l'humidité alternent entre négatives et positives, avec une majorité de positives. À partir de la médiane, il est possible d'observer une augmentation

constante avant une décroissance dans les pics. La plupart des coefficients dans les pics sont négatifs, sauf au lag 3. Comme pour la pression, elles sont non significatives sauf pour de rares quantiles au jour même, autour des pics à 80 %. Les estimations ne semblent pas différentes de celles par moindres carrés ordinaires.

Les estimations des coefficients des précipitations sont constantes et positives jusqu'au quantile à 80 %. Passé ce quantile, elles augmentent dans les pics aux lag 0 et 3 tandis qu'elles diminuent puis remontent aux lags 7 et 14. Durant ces fluctuations, les coefficients changent de signe. Le coefficient le plus important (0.025) est celui au lag à 14 jours, dans le pic à 97.5 %. Les effets des précipitations ne sont cependant jugés ni significatifs ni différents des moindres carrés ordinaires.

La courbe des estimations de la neige présente une allure similaire dans les quatre lags considérés, à savoir croissante et positive. Cela indique qu'une augmentation de la quantité de neige au sol est liée à une augmentation des décès à Québec. L'effet est plus prononcé dans le pic à 99 %, après 3 jours (coefficient de 0.35). De plus, il est différent de celui estimé par moindres carrés ordinaires uniquement autour du pic à 90 %, aux lags 0 et 3.

3.5. Résumé des résultats

Les droites de régression sont visibles pour les hospitalisations à Montréal et à Québec à la Figure III-7 et Figure III-8 et pour les décès à Montréal et à Québec à la Figure III-9 et Figure III-10. Un sous-échantillon des quantiles représentatif de l'ensemble de la distribution conditionnelle a été sélectionné avec un accent sur les pics (10, 25, 50, 75, 90, 95, 97,5 et 99 %) car autrement, les figures seraient indéchiffrables.

Les coefficients estimés de la température sont décroissants et négatifs pour les hospitalisations comme pour les décès, à Québec et Montréal. Les droites de régression sont donc décroissantes et cela indique que quel que soit le quantile de la distribution des hospitalisations des décès analysée, une augmentation de la température amène une diminution de la variable sanitaire. Autrement dit, plus la température augmente, plus le nombre d'hospitalisations baisse. Cette diminution est par ailleurs de plus en plus forte dans les pics, puisque les coefficients estimés y sont les plus forts. Ceci peut s'expliquer entre autres par le fait que les pics surviennent principalement en hiver (voir la discussion sur la température), une augmentation de température amenant ainsi un réchauffement durant une période froide. Il s'agit par ailleurs de la variable la plus souvent significative dans les résultats.

La pression atmosphérique est surtout significative le jour même de l'issue sanitaire dans les deux communautés métropolitaines. De plus, elle l'est plus souvent pour la pression minimale que maximale. Lorsqu'elle est significative avec les variables sanitaires, elle est reliée négativement avec la variable sanitaire, indiquant qu'une augmentation de pression amène moins de décès et hospitalisations, à plus forte raison dans les pics (puisque les coefficients y sont les plus forts). De plus, la relation a tendance à s'inverser puisque les coefficients estimés deviennent positifs à partir du lag 3. À ce niveau, il serait intéressant d'analyser les phénomènes climatiques associés à ces variations de pression atmosphérique pour comprendre en détail ces résultats.

Concernant l'humidité, les coefficients estimés sont globalement non significatifs, pour les deux villes. Il est donc impossible d'interpréter les coefficients obtenus. Les décès à Montréal constituent la seule exception où ils sont significatifs, particulièrement à partir du lag 3. De plus, ils sont négatifs, indiquant qu'une augmentation de l'humidité entraîne une diminution des décès. L'effet dans les pics est accentué puisque les coefficients y sont plus forts que dans le reste des

quantiles. Cet effet est notable puisqu'il n'est pas présent pour les décès à Québec, soulignant une différence des deux villes.

L'effet des précipitations n'est pas clair. En effet, les coefficients estimés alternent entre valeurs positives et négatives et beaucoup d'estimations sont jugées non significatives. Néanmoins, les pics présentent les valeurs de coefficients les plus fortes. Les coefficients sont majoritairement significatifs pour les hospitalisations à Québec, au lag 14. Toutefois, ils ne le sont pas dans les quantiles les plus élevés, n'indiquant pas d'effet particulier des précipitations sur les pics d'hospitalisations.

Les coefficients estimés de la quantité de neige au sol sont croissants et positifs pour les décès, à Québec et Montréal. Ainsi, une augmentation de neige est liée à une augmentation des décès. Concernant les hospitalisations, les coefficients présentent une phrase croissante puis décroissante, passant ainsi des valeurs positives aux valeurs négatives. L'effet dans les pics est différent entre les deux villes car à Québec, les coefficients remontent tandis qu'ils continuent de décroître à Montréal. Ainsi, plus il y a de neige au sol, moins il y a de pics d'hospitalisations pour Montréal, et inversement pour Québec. Cela indique des effets non homogènes selon les quantiles et les zones étudiés.

Globalement, les versions minimales et maximales des variables météorologiques ont des effets similaires. Il y a cependant deux exceptions : la pression et l'humidité, qui exhibent des comportements différents en ce qui concerne la significativité selon que la variable soit minimale ou maximale (voir Figure annexe 1 à Figure annexe 4).

4. Discussion

Dans cette section, il est important de garder en tête que toutes les études citées en introduction portaient sur les phénomènes moyens. Il est ainsi difficile d'obtenir une comparaison juste avec la présente étude, qui se concentre sur plusieurs quantiles et en particulier sur les pics. Des considérations générales, des commentaires sur chaque variable météorologique puis des considérations statistiques sont présentées dans les paragraphes suivants.

De façon globale, la régression quantile présente des coefficients variant d'un quantile à l'autre et les pics (quantile le plus élevé) constituent les coefficients de régression les plus élevés (en valeur absolue). Il s'agit d'une hétérogénéité dans les coefficients qui ne serait pas mise à jour par une régression par moindres carrés ordinaires. C'est la même raison qui pousse Austin *et al.* (2005), Marrie *et al.* (2009) ou Siciliani *et al.* (2013) à utiliser la régression quantile dans un contexte de santé publique. Ils indiquent que la régression quantile, en permettant d'estimer des coefficients de régression différents dans la distribution conditionnelle de la variable à expliquer, amène une connaissance approfondie de la relation étudiée et aide ainsi à déceler des effets inconnus auparavant. Pour Siciliani *et al.* (2013), la régression quantile permet par ailleurs de mettre en avant un effet plus important de la variable explicative dans les pics, tel qu'illustré dans le présent chapitre. Quant à Marrie *et al.* (2009), ils observent des effets opposés entre une régression sur la médiane et une régression sur le quantile à 90 %. Cette inversion est parfois survenue dans les présents résultats, mais elle était la plupart du temps non significative (comme pour les précipitations dans les cas des décès des deux communautés). De plus, Cade et Noon (2003) signalent que les coefficients dans les quantiles plus extrêmes de la distribution peuvent être détectés comme significatifs plus souvent (impliquant plus de puissance statistique) que dans les estimations plus centrales comme la moyenne ou la médiane. Ainsi, la régression quantile est

utile pour l'étude des pics sanitaires en santé public car elle permet de se concentrer sur ces derniers, tout en comparant les estimations avec le reste de la distribution.

Dans des domaines autres que la santé publique, cet effet non homogène et particulièrement accentué des variables explicatives dans les quantiles élevés de la variable à expliquer a déjà été observé grâce à la régression quantile, comme en climatologie ou en hydrogéologie (Jagger et Elsner, 2009; Hirschi *et al.*, 2011). Les auteurs motivent d'ailleurs leur utilisation de cette régression avec sa capacité à décomposer la relation dans plusieurs quantiles.

Certains résultats montrent des différences dans les effets des variables météorologiques, entre Montréal et Québec (comme la quantité de neige au sol, voir plus loin dans cette section). Ceci est attendu dans la mesure où ces deux communautés diffèrent sur plusieurs points, comme une plus forte présence d'îlots de chaleur (Giguère, 2009) et une plus faible quantité de neige reçue à Montréal comparativement à Québec (Tableau 0-3). Certaines études ont trouvé aussi des différences d'influence des variables météorologiques, en étudiant deux ou plusieurs villes au Québec. À titre d'exemple, Bustinza *et al.* (2013) remarquent une hausse du taux de décès suite à la canicule de 2010 (comparativement aux taux de 2005 à 2009) à Montréal mais pas à Québec. Masselot *et al.* (2015) trouvent des R^2 plus élevés à Montréal comparativement à Québec, dans une application de la régression par décomposition modale empirique aux MCV. Ainsi, comme ces auteurs, les résultats obtenus ici corroborent les différences dans les deux communautés d'étude.

4.1. Température

La température s'est révélée être la variable la plus importante (en termes de significativité et de valeurs de coefficients), comme dans les études classiques. Étant donné que tous les coefficients ont été estimés négatifs, la température est liée de façon inverse aux variables sanitaires. Une

explication possible est que les pics de MCV surviennent majoritairement par temps froid (Mourtzoukou et Falagas, 2007; Gasparrini *et al.*, 2015; Yang *et al.*, 2015b). Ainsi, un réchauffement lors d'une vague de froid a un effet protecteur. Cela concorde avec Auliciems et Frost (1989) qui montrent qu'à Montréal, une augmentation de la température correspond à une diminution des décès par MCV. Leurs résultats diffèrent tout de même en termes méthodologiques car ils effectuent une régression polynomiale sur l'ensemble des observations. Dans le nord de la France, Danet *et al.* (1999) trouvent qu'une augmentation de la température fait décroître le nombre de décès par MCV. Enfin, Fong et Ma (2013), dans une étude sur les effets de la météo sur les hospitalisations par MCV (à l'aide d'une régression linéaire classique) à Honk Kong, rapportent que la température leur est associée de façon négative et linéaire. Force est de rappeler que les résultats de ces trois études sont basés sur les phénomènes moyens. Un effet notable de la température dans la présente étude est l'augmentation de cet effet dans les pics. Les estimations sont toujours négatives, mais avec de plus fortes valeurs dans les pics. En analysant spécifiquement les pics (et non divers quantiles dont les pics) en fonction de la température, Chiu *et al.* (2016b) ont obtenu des résultats similaires à ceux obtenus ici. Leurs résultats sont décrits avec plus de détails sur les relations non linéaires entre les pics et les variables météorologiques, car ils utilisent des méthodes non paramétriques, mais les courbes sont décroissantes (ce qui équivaut à une droite de régression avec un coefficient négatif). La Figure III-11 présente une régression quantile non linéaire (par l'utilisation de B-splines, deux paragraphes sont dédiés à cette approche dans la section 4.6), de sorte à pouvoir comparer avec les résultats de Chiu *et al.* (2016b). Les échelles de température ne sont pas les mêmes puisque leur approche mène à l'utilisation d'un sous-ensemble des données. Néanmoins, lorsqu'elles sont ramenées aux mêmes valeurs minimale et maximale, les courbes de régression sont similaires. Par exemple, les effets estimés des deux approches révèlent une influence légèrement supérieure

autour de 0 °C, qui diminue lorsque les températures augmentent. L'apport majeur de la présente étude est toutefois qu'elle a permis de dévoiler des effets très hétérogènes, en particulier dans les pics.

4.2. Pression atmosphérique

La pression atmosphérique a exhibé des résultats contradictoires. En effet, à Montréal comme à Québec, une augmentation de la pression entraîne une diminution des décès car les coefficients sont estimés négatifs lorsqu'ils sont significatifs (majoritairement au jour même). Cet effet est par ailleurs plus accentué dans les pics. Schwartz (2000) trouve aussi un effet protecteur de la pression dans 10 villes aux États-Unis, mais pour les décès toutes causes et avec une analyse de l'ensemble des événements. En revanche, Danet *et al.* (1999) montrent une relation entre la pression positive atmosphérique et les décès par infarctus du myocarde dans le nord de la France.

Néanmoins, cette association négative et significative de la pression le jour même n'est pas observé pour les hospitalisations à Montréal tandis qu'ils sont pour la plupart non significatifs à Québec. À partir du lag 7 à Montréal, des coefficients positifs sont observés. Dans l'étude de Fong et Ma (2013), les auteurs découvrent aussi une association positive linéaire entre les hospitalisations par MCV et la pression atmosphérique. Chiu *et al.* (2016b) n'ont par contre pas réussi à trouver de relation entre la pression et les hospitalisations (et les décès) par MCV. Il faut toutefois prendre en compte que leur définition des pics n'est pas la même que dans ce chapitre. En effet, ils utilisent les méthodes des blocs et du seuil (de la théorie des valeurs extrêmes) pour extraire les pics tandis qu'ici, il suffit de définir un quantile élevé. Cela ne donne donc pas nécessairement les mêmes séries de pics. Leur but est par ailleurs légèrement différent, puisqu'ils cherchent à expliquer les effets des variables météorologiques sur une série de pics en particulier,

tandis qu'ici l'intérêt est plutôt de montrer les différences (s'il elles existent) entre les effets dans les pics et dans les autres quantiles.

4.3. Humidité relative

Dans l'étude déjà citée de Fong et Ma (2013), les auteurs n'ont pas trouvé de lien particulier entre l'humidité relative et les hospitalisations. C'est aussi le cas pour Schwartz *et al.* (2004), aux États-Unis, concernant les hospitalisations pour cause de MCV. La présente étude corrobore ces résultats, dans le sens où l'humidité ne montre pas de schéma significatif. Le rôle de l'humidité sur les MCV n'est cependant pas encore clair, car certains auteurs établissent un lien. Il est possible de citer Abrignani *et al.* (2012) ou Masselot *et al.* (2015). En particulier, Masselot *et al.* (2015) travaillent sur les mêmes données que celles utilisées dans ce projet. Ils trouvent un lien entre les décès et hospitalisations par MCV et l'humidité sur le moyen et long termes (de l'échelle mensuelle à annuelle). Les lags considérés ici vont de 0 à 14 jours (donc moins d'un mois), ce qui pourrait expliquer en partie pourquoi la régression quantile n'a pas mis à jour cette relation. Si la contribution de l'humidité seule n'est pas encore définie, il est toutefois établi que l'humidité joue un rôle sur la température ressentie et ces deux variables sont souvent étudiées conjointement (Lin *et al.*, 2009 ou Schwartz *et al.*, 2004).

Le cas particulier des décès à Montréal indique une relation négative et particulièrement importante dans les pics. La seule autre étude ayant rapporté ce phénomène est celle de Chiu *et al.* (2016b), qui obtiennent le même résultat (effet significatif de l'humidité sur les décès à Montréal). La relation qu'ils trouvent est croissante, contrairement à celle obtenue ici. Cette contradiction peut s'expliquer par le fait qu'ils décomposent la relation en séparant les jours de semaine des fins de semaine, ce qui résulte en des séries de pics différentes de celles étudiées ici.

4.4. Quantité de neige au sol

Les estimations des coefficients de la quantité de neige au sol sont croissantes et positives dans les pics, pour les décès dans les deux communautés urbaines et pour les hospitalisations à Québec. Ainsi, plus les pics étudiés de la variable sanitaire sont dans des quantiles élevés, plus l'effet de la neige est associé à une augmentation de la variable sanitaire. Chiu *et al.* (2016b) ont aussi trouvé une relation croissante entre la neige et les pics de décès à Québec. L'effet est malgré tout différent pour les pics d'hospitalisations à Montréal, puisqu'il est associé à une diminution de la variable sanitaire. Aucune étude n'a rapporté cette différence entre Québec et Montréal jusqu'à présent, puisqu'aucune ne regarde les pics en particulier (et avec autant de quantiles détaillés).

Les courbes des coefficients de la neige ont tendance à suivre celles de la température et à décroître pour les hospitalisations tandis qu'elles évoluent de façon inverse pour les décès (décroissante pour la température et croissante pour la neige). Finalement, entre les deux communautés urbaines, il y a une différence au niveau des pics d'hospitalisations, à savoir qu'il y a une relation négative entre la quantité de neige au sol et les pics à Montréal (contrairement aux pics à Québec). Cela indique que lorsque les pics sont étudiés, plus la quantité de neige est importante, moins il y a d'hospitalisations (et inversement pour Québec). Dans le cas de Montréal, une hypothèse possible est que lors des pics hivernaux, une quantité de neige importante découragerait les gens à sortir et donc à s'exposer (Martel *et al.*, 2010; Barnett *et al.*, 2012). En revanche, de plus amples analyses sont nécessaires pour expliquer la relation positive observée à Québec. Par exemple, la population pourrait y être simplement plus adaptée à une quantité de neige abondante.

4.5. Précipitations

L'effet des précipitations sur les décès et les hospitalisations par MCV n'est pas significatif la plupart du temps. Quelques quantiles des hospitalisations à Montréal (tous les lags) et à Québec (tous les lags) révèlent des coefficients significatifs négatifs, ce qui signifie qu'une augmentation des précipitations est associée à une diminution des hospitalisations. L'effet est néanmoins non significatif dans les pics. Au meilleur de la connaissance des auteurs, il existe des études qui relient les pics de certaines conditions de santé aux précipitations (telles que les maladies hydriques, Curriero *et al.*, 2001), mais aucune sur les MCV. Comme pour l'humidité, les précipitations sont rarement étudiées seules en relation avec des variables sanitaires. Elles sont habituellement étudiées avec la température ou la neige. Par exemple, Baker-Blocker (1982) relie les décès par MCV aux précipitations à Minneapolis (États-Unis). Elle utilise aussi la température et la quantité de neige, car les précipitations seules expliquent difficilement les décès. Toutefois, les précipitations ont bien été intégrées avec ces variables dans les modèles présentés ici. Chiu *et al.* (2016b) ont quant à eux trouvé des relations significatives croissantes avec les décès à Montréal et les hospitalisations à Québec.

4.6. Considérations statistiques

Il pourrait être intéressant d'utiliser conjointement l'approche de Chiu *et al.* (2015) (théorie des valeurs extrêmes) et celle de la régression quantile présentée dans ce chapitre. Par exemple, la Figure III-11 présente une régression quantile non paramétrique. Une fois que la fonction de régression est établie, les pics supérieurs à cette fonction pourrait être modélisés par une distribution de Pareto généralisée, comme dans Chiu *et al.* (2015). Cela permettrait ainsi d'obtenir une vision des relations dans l'ensemble des quantiles, puis d'isoler et d'analyser en détail les pics. Mélanger les deux méthodologies n'est généralement pas appliqué en pratique. Un

rare exemple est celui de Jagger et Elsner (2009), qui étudient les cyclones tropicaux grâce à la régression quantile et la théorie des valeurs extrêmes (en particulier la méthode du seuil). Dans leur étude, ils signalent que cette théorie est appropriée pour estimer les quantiles les plus extrêmes, tandis que la régression quantile est plus indiquée pour estimer les changements dans les quantiles médians. Les méthodes liées à la théorie des valeurs extrêmes et à la régression quantile pourraient ainsi être complémentaires. Concernant l'exemple illustré à la Figure III-11, cette approche n'a pas été appliquée dans ce chapitre car elle nécessite plusieurs choix subjectifs et surtout, souffre de l'inconvénient du croisement des quantiles (expliqué dans les prochains paragraphes).

La significativité des coefficients estimés est jugée à l'aide des IC. Si les IC passent par la valeur nulle, alors les coefficients sont jugés non significatifs. De plus, si les IC se croisent, alors la différence entre les estimations par régression quantile et moindres carrés ordinaires est jugée non significative dans ce chapitre. Cet aspect non significatif a été présumé par manque de théorie sur ce sujet en particulier. En effet, Austin et Hux (2002), Cumming et Finch (2005) et Knezevic (2008) mentionnent qu'il est toujours vrai que si deux estimateurs ont des IC qui ne se croisent pas, alors ils sont statistiquement différents. Le contraire n'est néanmoins pas vérifié, deux estimateurs dont les IC se croisent ne sont pas nécessairement statistiquement différents. Une règle visuelle avancée par Cumming (2009) indique que si le croisement des IC de deux estimateurs différents s'étend jusqu'à la moitié d'un « bras » d'un IC, cela correspond approximativement à une différence significative au seuil statistique de 5 %. Cependant, comme mentionné plus tôt, le croisement des IC n'est pas un sujet encore exploré en régression quantile. Ainsi, même si les IC entre les estimations par régression quantile et moindres carrés ordinaires se croisent, il est possible que les coefficients soient différents. Dans l'étude présente, il a été

présupposé que les coefficients étaient différents pour suivre le principe de parcimonie. Si une régression linéaire classique est suffisante, alors il n'est pas utile de complexifier les modèles. Toutefois, une régression par moindres carrés ordinaires n'aurait sans doute pas été suffisante ici. Par exemple, les résultats sur la température et la neige révèlent des relations hétérogènes, que n'a pas obtenues la régression par moindres carrés ordinaires. Ainsi, même si une règle de précaution a été prise durant les interprétations, il y a clairement des différences et des avantages à utiliser la régression quantile dans l'étude des pics sanitaires.

Toutes les variables météorologiques disponibles ont été utilisées dans cette étude. Il existe néanmoins des outils de sélection des variables adaptées à la régression quantile. Il est possible de citer la pénalité *smoothly clipped absolute deviation* (SCAD, Fan et Li, 2001). Elle repose entre autres sur la sélection d'un paramètre de lissage, qui rend la pénalisation plus ou moins sévère. Fan et Tang (2013) proposent une sélection du paramètre de lissage par le critère d'information généralisé tandis que Sun *et al.* (2013) préconisent le choix du paramètre de lissage qui stabilise les paramètres estimés. La pénalité SCAD n'est cependant valide que pour un modèle dont le quantile et le lag sont fixés. Puisque dans la présente étude, un ensemble de quantiles et de lags sont étudiés en même temps, aucune méthode de sélection des variables n'est utilisée. En effet, il n'est pas justifié d'estimer un modèle par régression quantile avec une variable en particulier dans un quantile, puis de la supprimer pour un autre quantile. De plus, il serait alors impossible d'obtenir des graphiques comme celui de la Figure III-2, qui permettent une interprétation simple et rapide des résultats.

Une contrainte importante à satisfaire en régression quantile est d'éviter le croisement des quantiles (Chernozhukov *et al.*, 2010). En effet, en estimant plusieurs quantiles, il est possible que certaines droites de régression se croisent si les estimations des coefficients ne sont pas

monotones. Si par exemple la droite de régression du quantile à 90 % croise celle du quantile à 75 %, alors il ne s'agit plus de la définition des quantiles puisque la première droite passera en dessous de la seconde. Des méthodes pour pallier cette situation non désirable de quantiles croisés existent (Bondell *et al.*, 2010 ou Liu et Wu, 2011), mais elles sont développées pour répondre à certains cas de figure très spécifiques. Puisqu'aucune théorie générale n'existe encore, ces méthodes ne sont pas utilisées dans ce chapitre.

L'utilisation de méthodes non paramétriques, comme les *splines*, couplées à la régression quantile (Marrie *et al.*, 2009 ou Koenker, 2010) ajouterait une plus grande dimension exploratoire à l'étude. En effet, ces méthodes « laissent parler les données » en demandant en contrepartie des échantillons plus grands. Elles n'auraient cependant pas permis de graphiques explicites comme dans l'étude présente. En effet, ces graphiques ont pu être obtenus car la version de la régression quantile considérée estimait un paramètre par variable, par quantile. Des méthodes non paramétriques estiment plusieurs paramètres par variable et leur interprétation dans chaque quantile n'est pas aisée. De plus, le problème des quantiles croisés est encore plus prononcé lorsque des courbes non linéaires sont considérées (Bondell *et al.*, 2010). Il est tout de même possible de comparer des méthodes non paramétriques (comme les modèles additifs généralisés, GAM) avec la régression quantile, de façon indépendante. Par exemple, Pauli et Rizzi (2006) comparent l'utilisation d'un GAM à la régression quantile dans l'étude de l'effet de la température sur les admissions hospitalières toutes causes confondues, en Italie. Ils trouvent des résultats très similaires entre les deux méthodes. De la même façon, la présente étude et celle de Chiu *et al.* (2016b) révèlent des ressemblances au niveau de la forme des effets.

5. Conclusion

Cette étude a pour but d'explorer les relations entre les MCV et les conditions météorologiques à l'aide de méthodes reliées à la régression quantile. Elle diffère des études classiques sur plusieurs points : 1) en considérant plusieurs quantiles de la variable réponse et en particulier les pics, 2) en considérant plusieurs lags dans les variables météorologiques, et 3) en considérant plusieurs variables météorologiques en relation avec les variables sanitaires. De plus, il s'agit d'une approche directe qui ne nécessite pas de sélectionner un sous-ensemble de données.

Dans plusieurs études classiques ainsi que dans Chiu *et al.* (2016b) (dans le cas des pics sanitaires), la température a démontré un effet particulièrement remarquable. La présente étude ne fait pas exception et confirme une fois de plus l'importance de la température sur les MCV. Surtout, l'utilisation de la régression quantile a permis de mettre à jour l'effet non homogène dans les quantiles des variables sanitaires. Si les relations sont invariablement inverses (coefficients estimés négatifs), l'effet dans les pics est plus fort que dans le reste de la distribution.

Des estimations plus précises dans plusieurs quantiles montrent que les variables météorologiques influent différemment sur les variables sanitaires et dépendamment du quantile dans lequel l'estimation est faite. Ceci a été observé entre autres par des valeurs de coefficients estimés très différentes dans les pics. Cet effet est largement « dilué » par une régression par moindres carrés ordinaires ou toute autre technique basée sur la moyenne. En particulier, une régression par moindres carrés ordinaires pourrait passer à côté des quantiles dans lesquels certaines variables sont significatives tandis que d'autres, non. Ainsi, couplées à des prévisions météorologiques, ces résultats peuvent être utiles à la surveillance des pics sanitaires. Par exemple, ils permettraient d'ajuster le personnel disponible dans un établissement de santé lors

des situations les plus communes et lors des pics sanitaires, puisque la régression quantile estime l'effet des variables météorologiques dans ces différents quantiles. Finalement, cette étude renforce l'idée de compléter les études classiques entre la santé et la météo par des études sur les pics.

6. Limitations

L'agrégation journalière des hospitalisations et des décès n'a pas permis l'utilisation de variables individuelles. En outre, il n'a pas été possible de considérer les habitudes de vie, comme la consommation de tabac, l'alimentation riche en mauvais gras ou le manque d'exercice physique (Dinas *et al.*, 2013; Afshin *et al.*, 2014; Vasconcellos *et al.*, 2014) qui ont un rôle dans la survenue des MCV, car ces renseignements ne sont pas rapportés dans les fichiers médicaux administratifs. Par ailleurs, en plus des variables météorologiques, des variables de pollution atmosphérique pourraient aussi être incluses dans les futurs modèles (Gérin *et al.*, 2003).

La régression quantile permet de cibler des quantiles particuliers et de décomposer la relation entre une variable réponse et des variables explicatives. Dans le même état d'esprit, il serait intéressant de différencier et cibler certaines MCV, comme les plus coûteuses au système de la santé ou celles qui causent le plus de décès. En effet, les conditions météorologiques pourraient ne pas avoir le même effet que sur l'ensemble des MCV. Schuld *et al.* (2013), en étudiant l'impact des conditions météorologiques sur les hospitalisations pour cause de rupture d'anévrisme de l'aorte abdominale en Allemagne, n'ont pas trouvé d'impact clair. En contrepartie, Ebi *et al.* (2004) observent une relation significative entre les changements de température et les hospitalisations pour cause d'accident vasculaire cérébral. Il faut cependant prendre en compte que cibler certaines MCV aurait pour conséquence de réduire la taille des échantillons (et donc la

puissance de l'étude), ce qui peut ne pas être adapté pour l'étude des pics, notamment dans de petites populations comme les communautés urbaines de Montréal et de Québec.

Figures

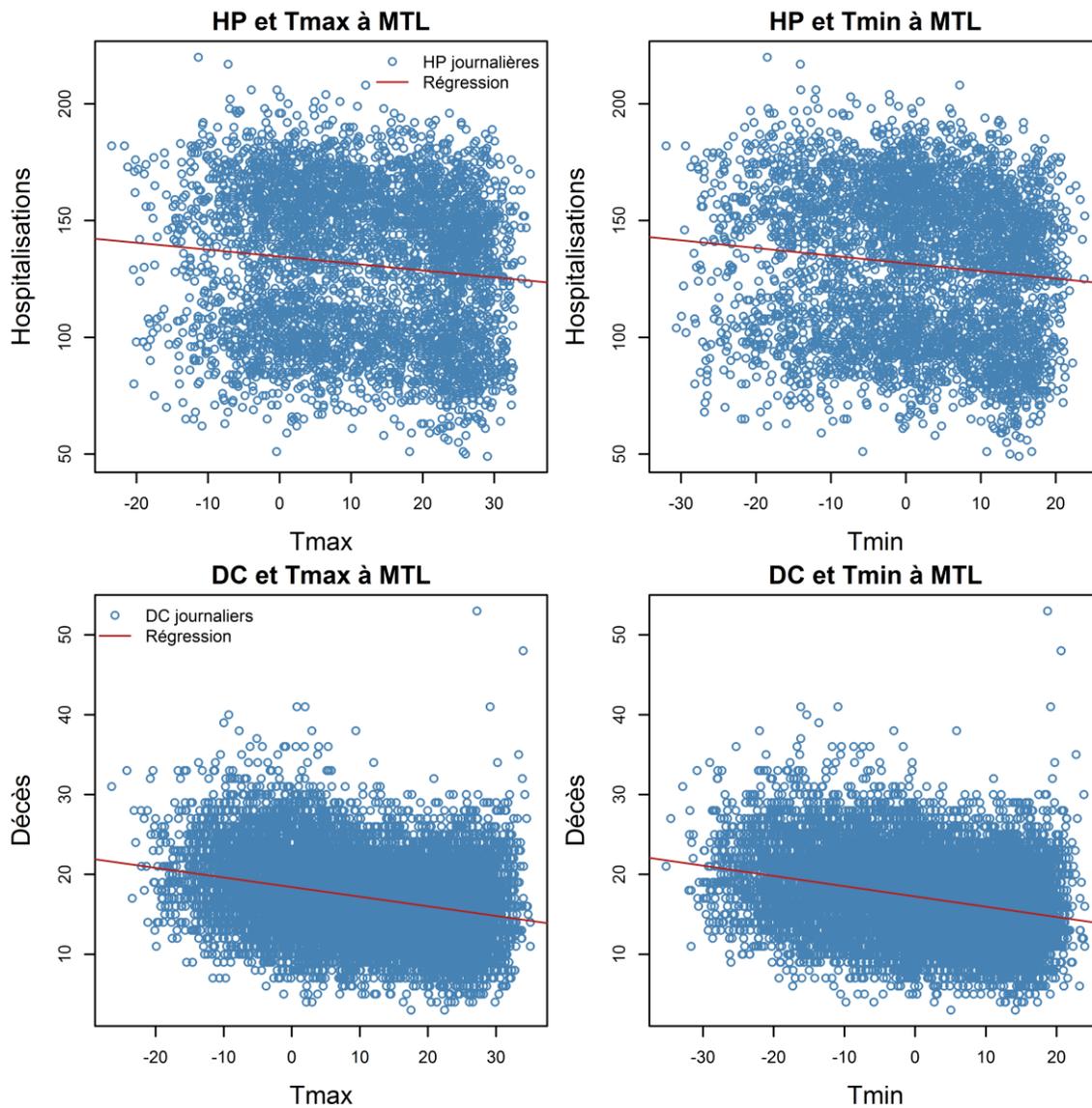


Figure III-1 : Illustration de l'effet moyen de la température maximale et de la température minimale sur les hospitalisations (graphiques du haut) et sur les décès (graphiques du bas) à Montréal.

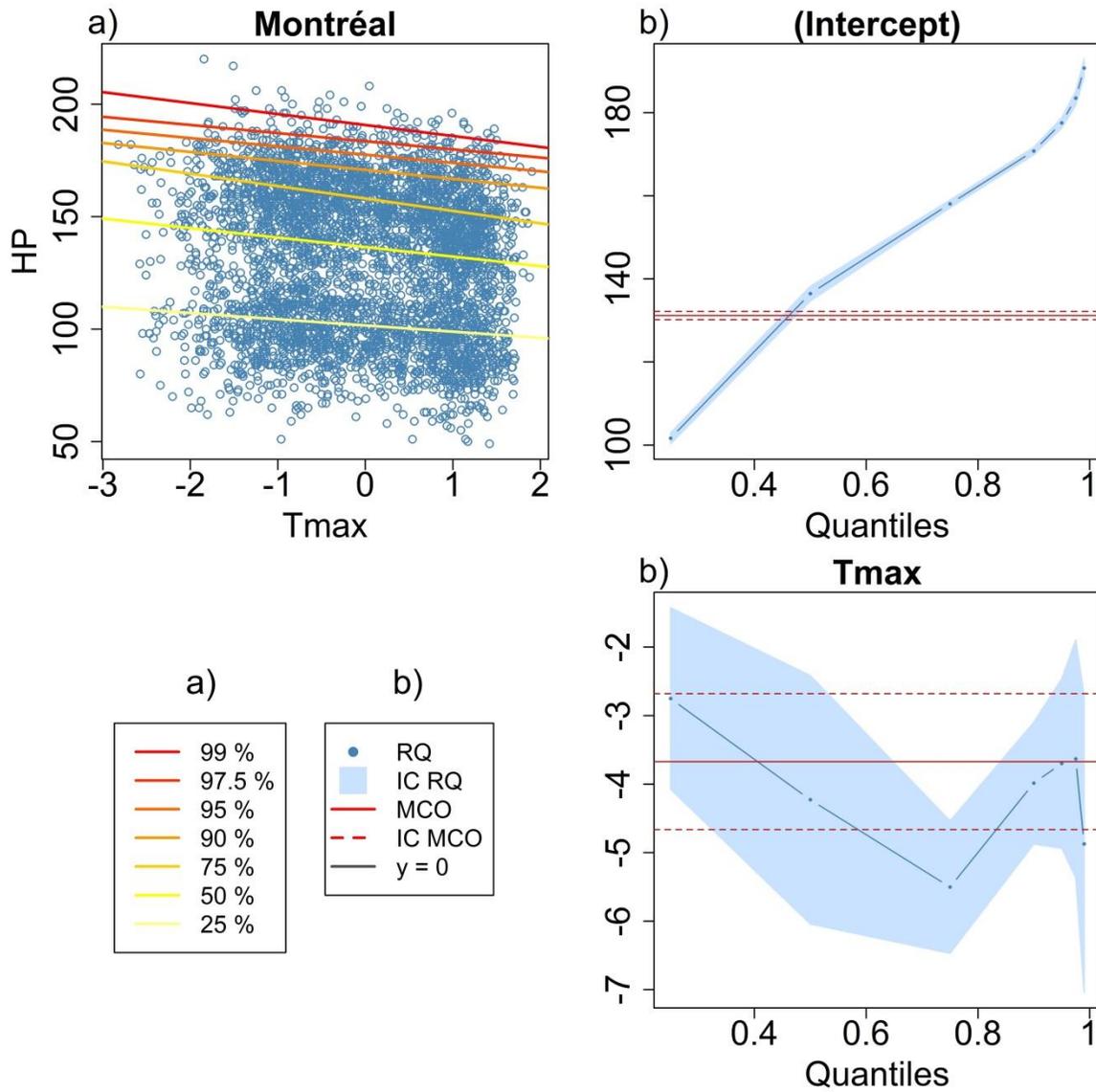


Figure III-2 : Exemple de résultats de la régression quantile. a) Différentes droites de régression quantile, b) Valeurs des différents coefficients estimés des droites de régression quantile.

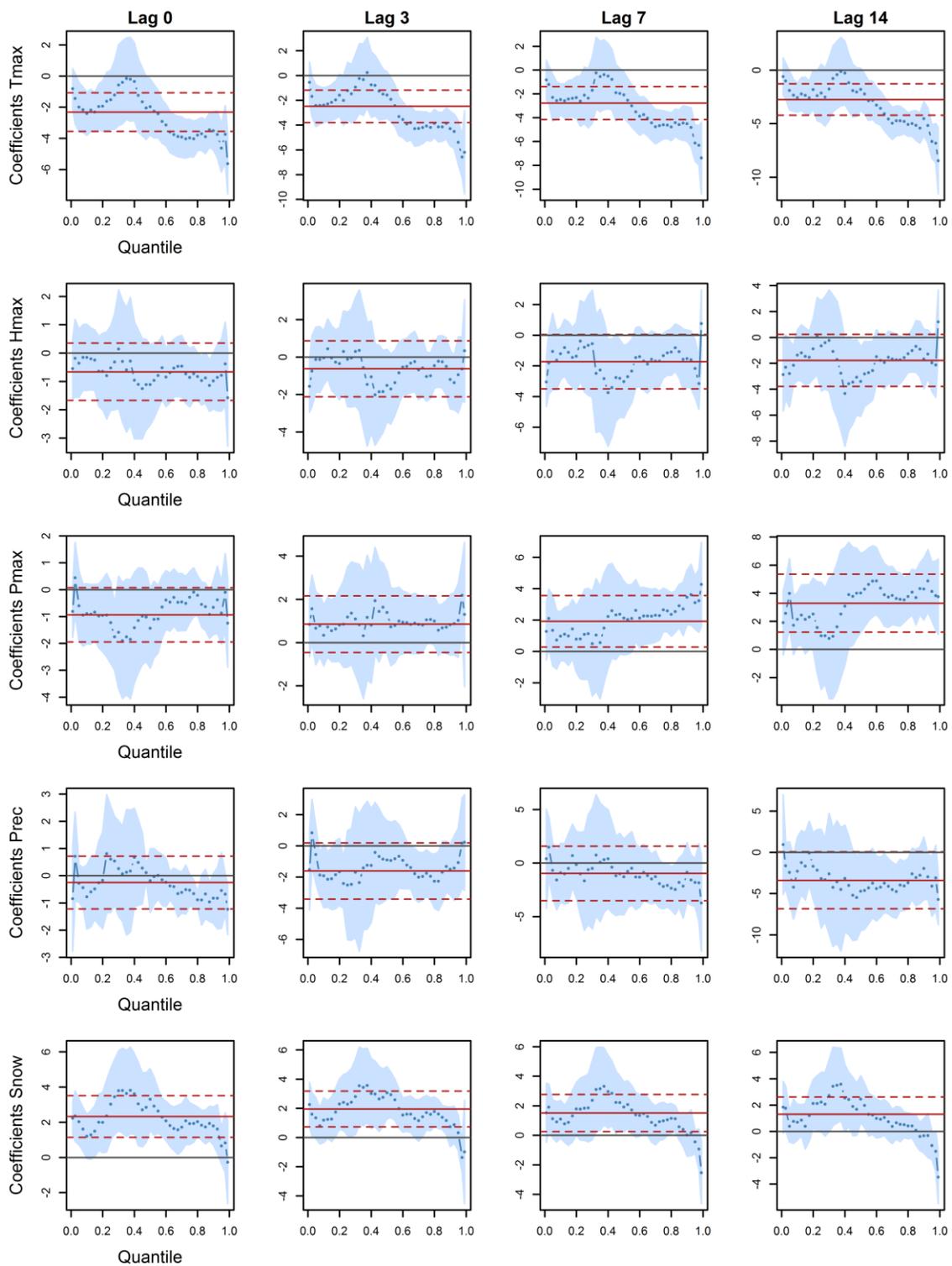


Figure III-3 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les hospitalisations à Montréal.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés

ordinaires ; - - - IC des moindres carrés ordinaires; — $y = 0$

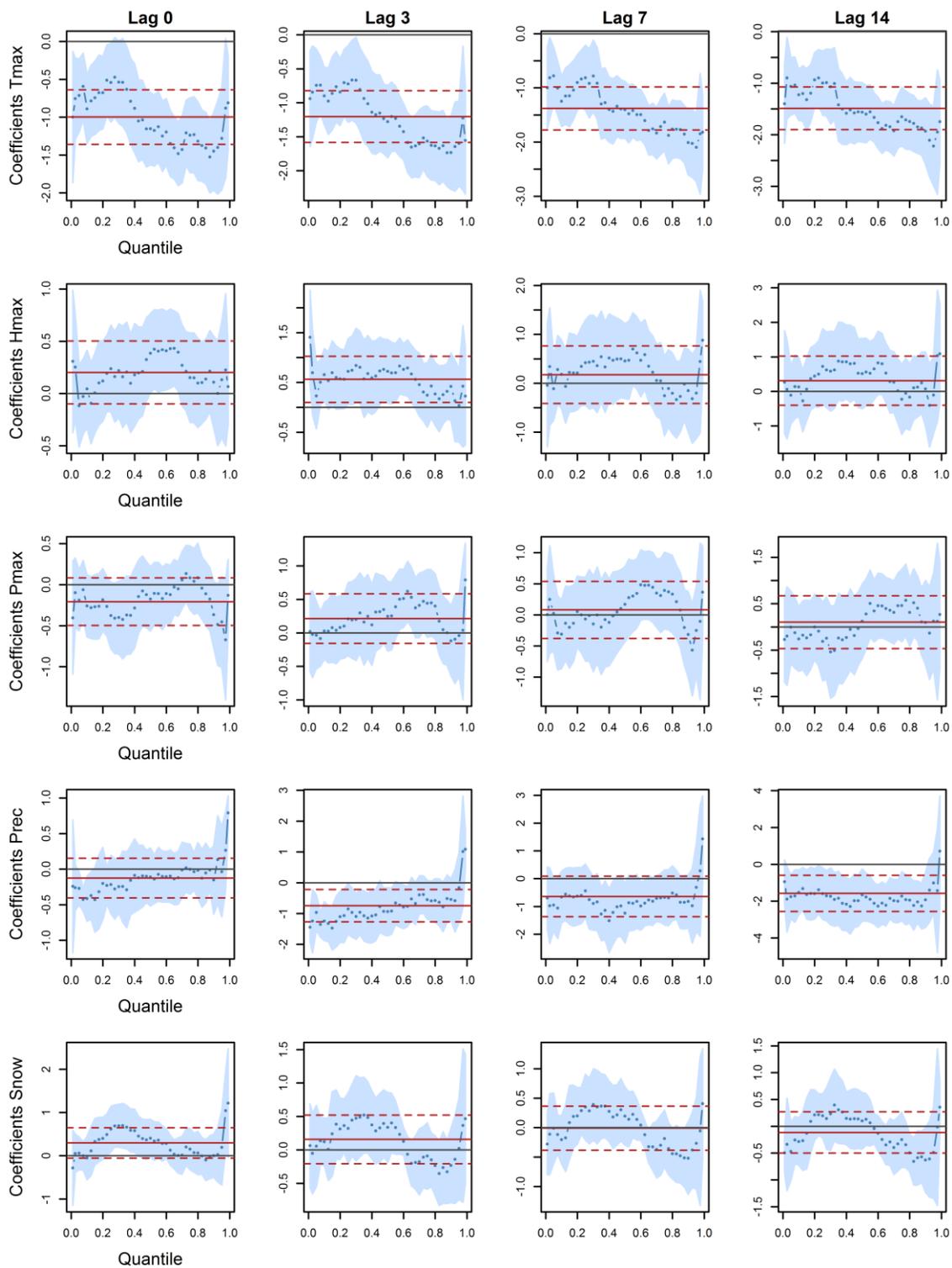


Figure III-4 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les hospitalisations à Québec.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés

ordinaires ; - - - IC des moindres carrés ordinaires; — $y = 0$

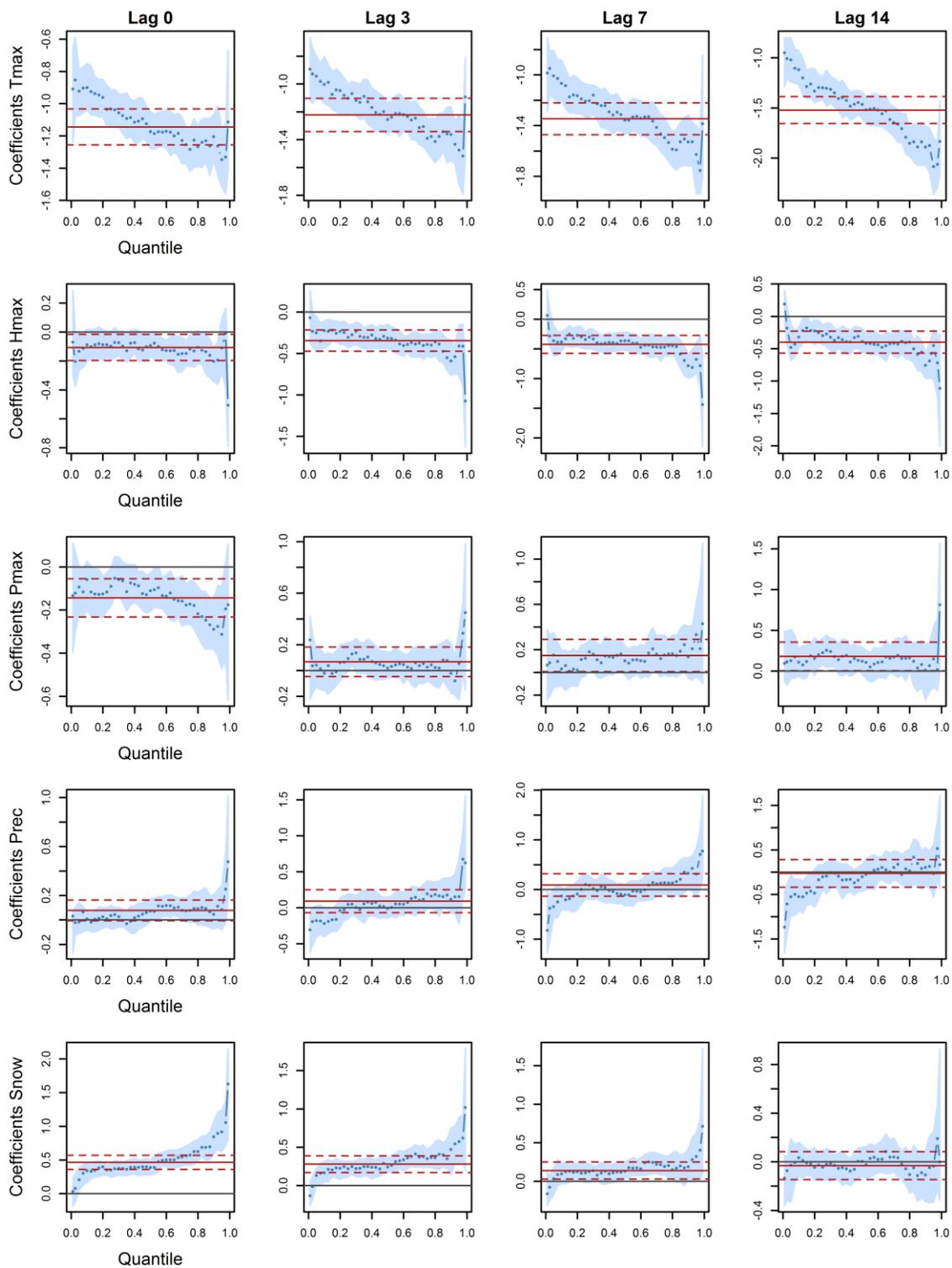


Figure III-5 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les décès à Montréal.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés

ordinaires ; - - - IC des moindres carrés ordinaires; — $y = 0$

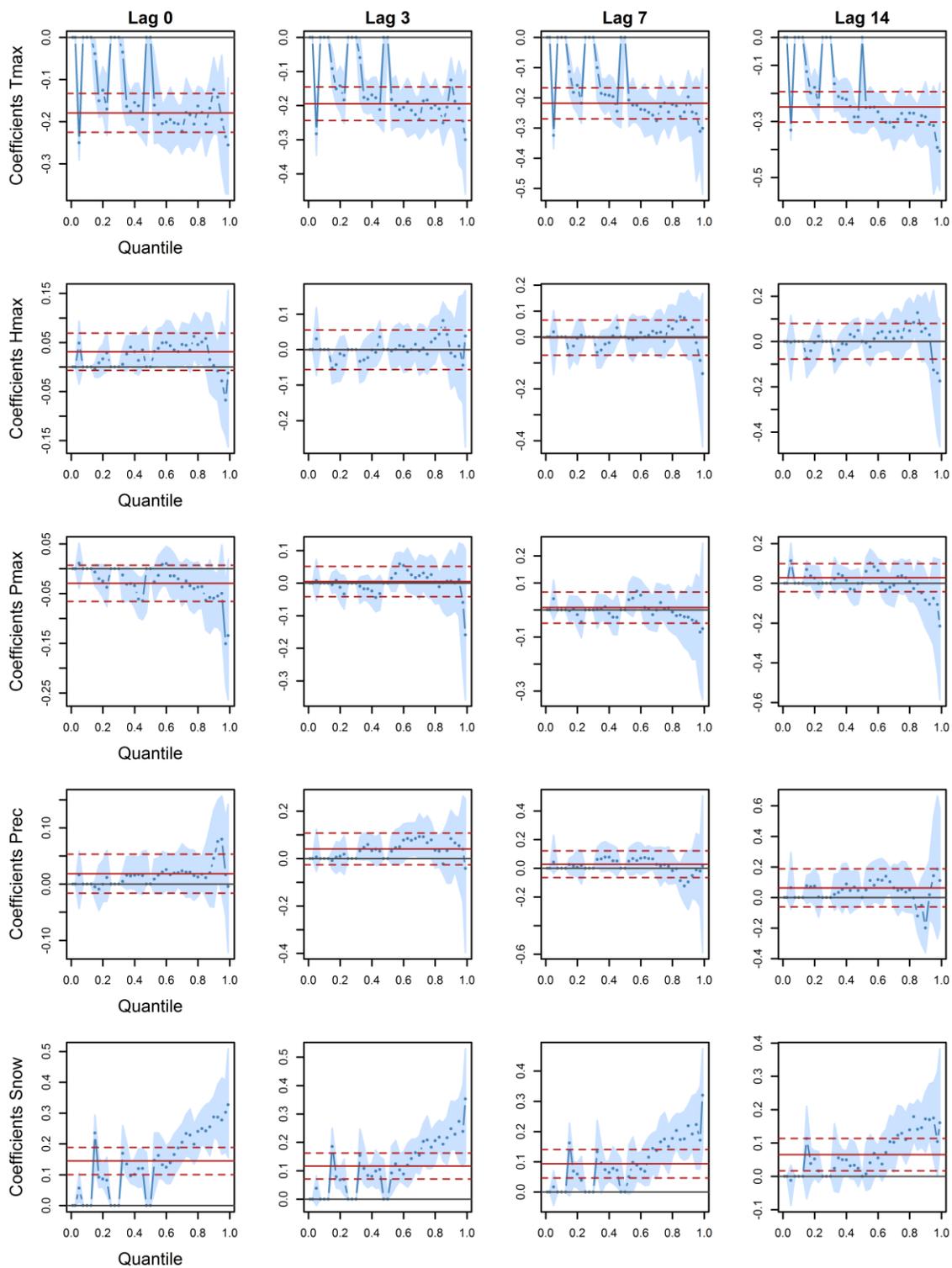


Figure III-6 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les décès à Québec.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés ordinaires ; - - - IC des moindres carrés ordinaires; — $y = 0$

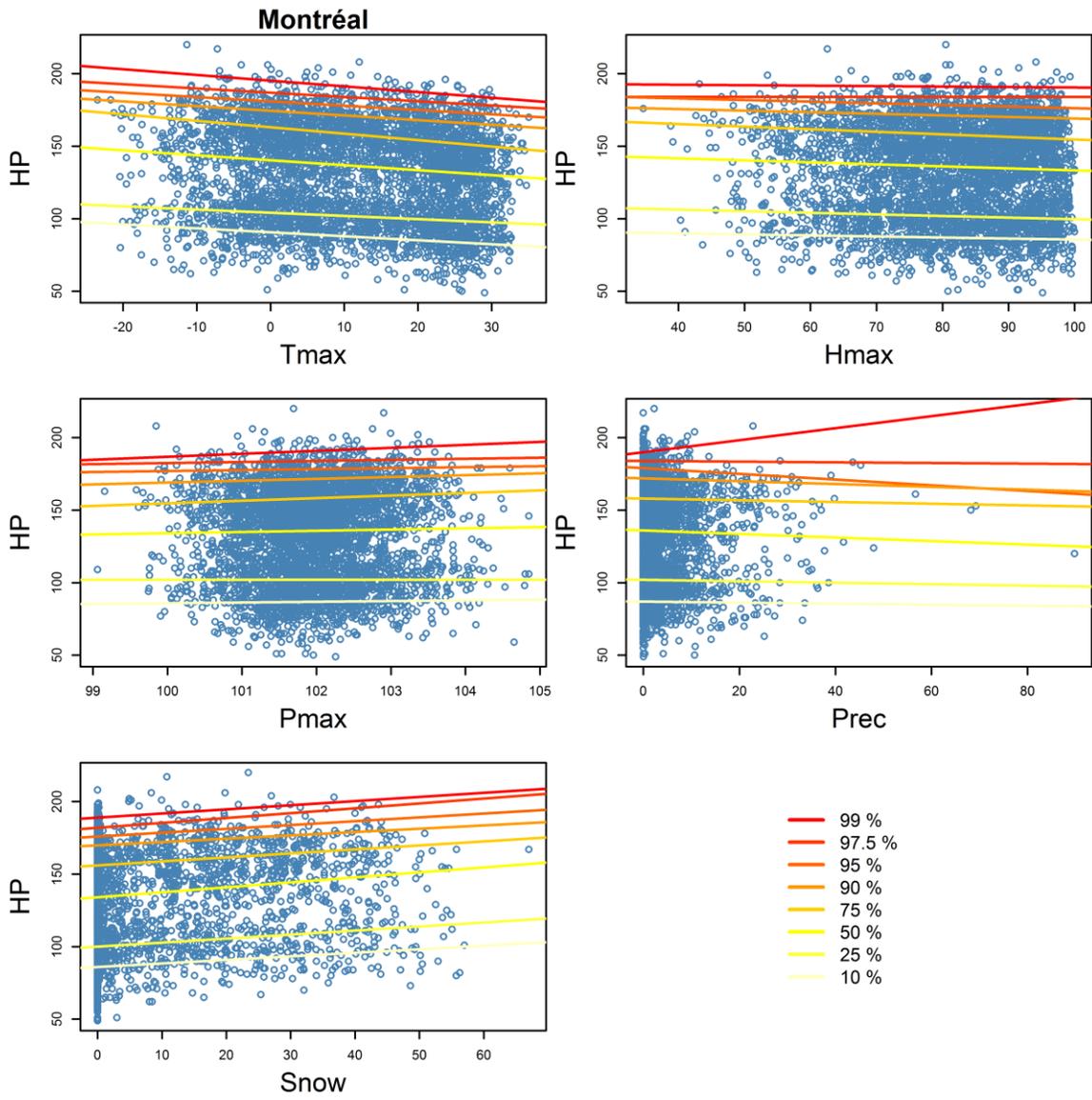


Figure III-7 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des hospitalisations à Montréal.

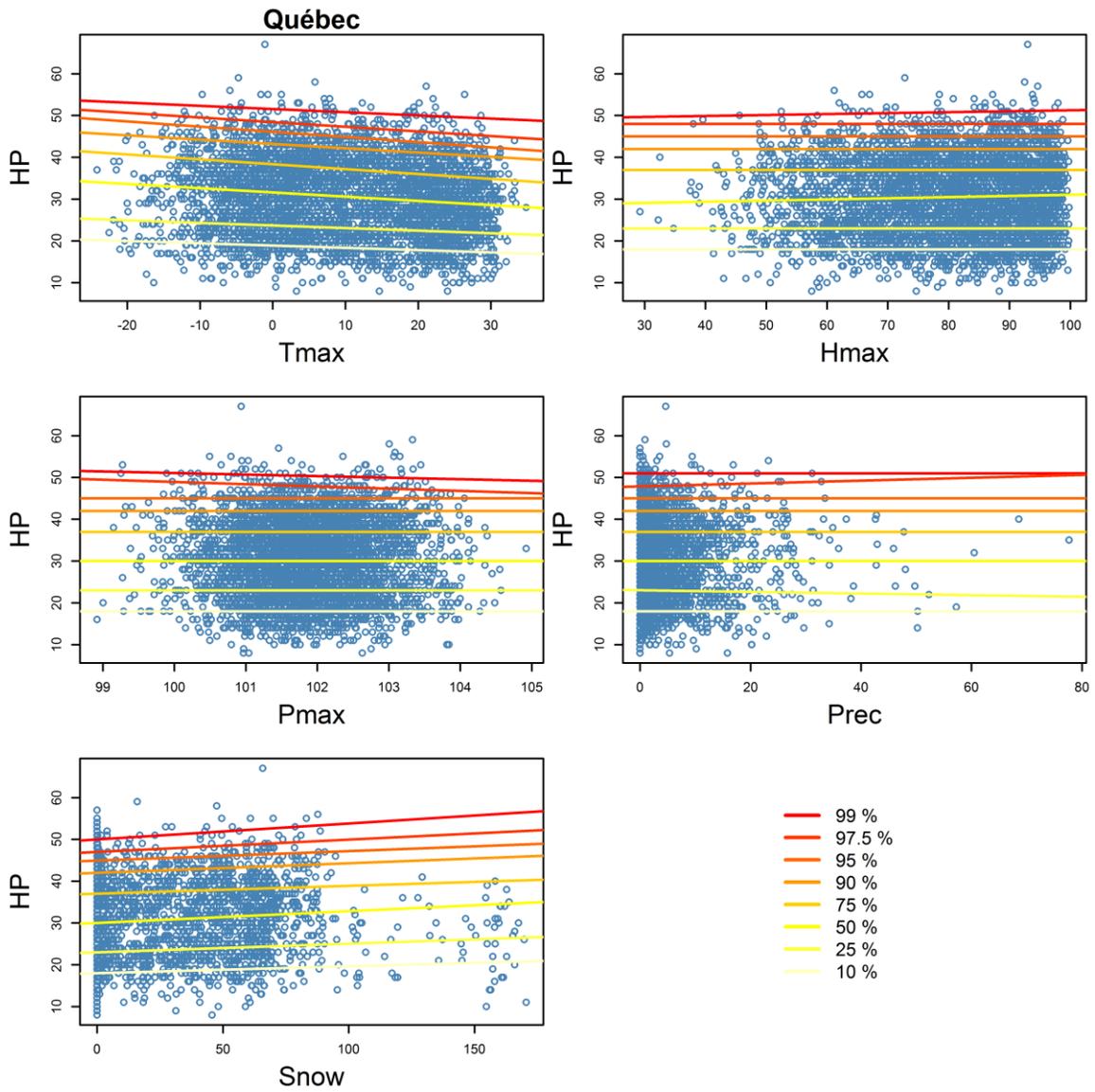


Figure III-8 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des hospitalisations à Québec.

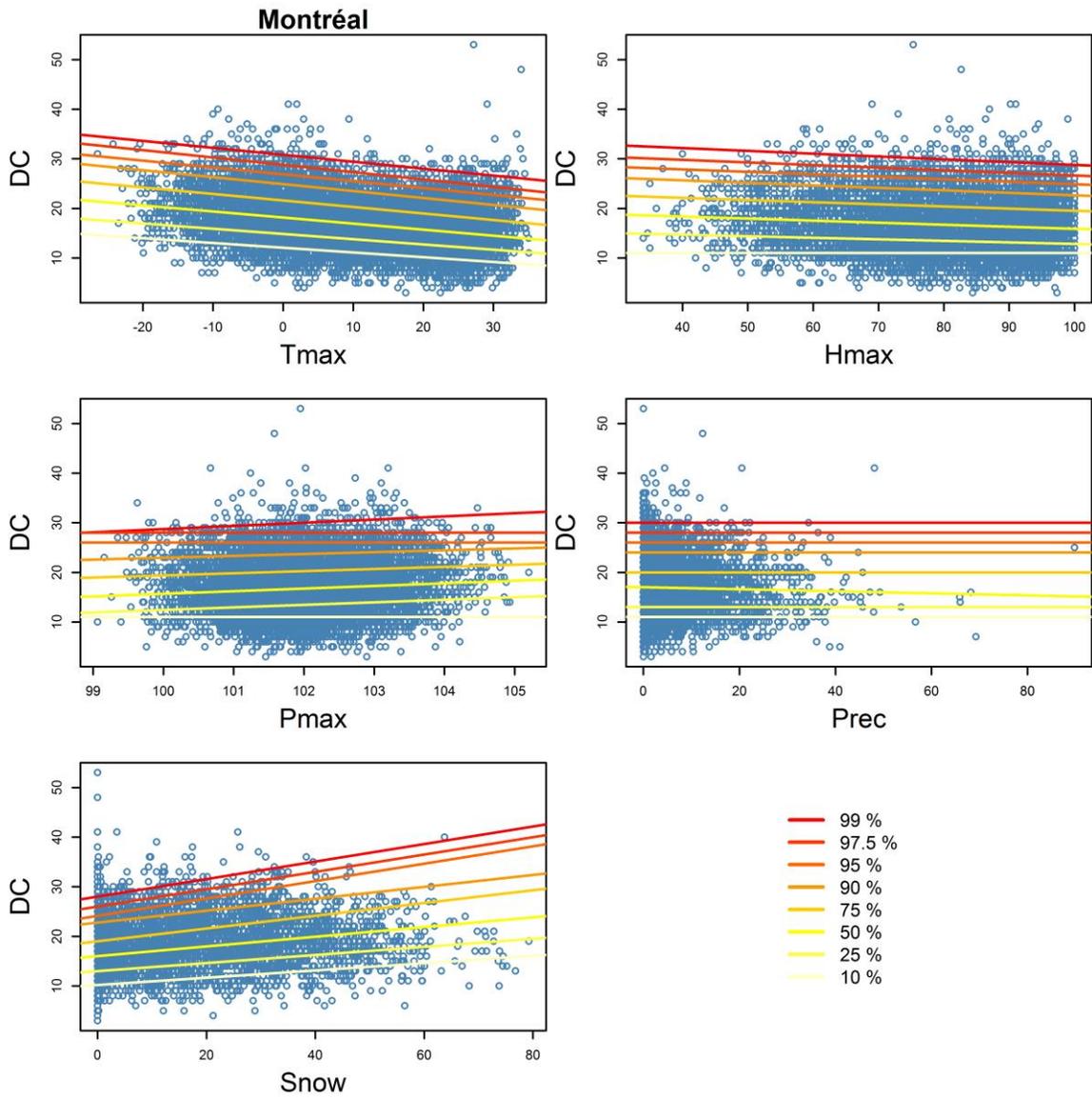


Figure III-9 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des décès à Montréal.

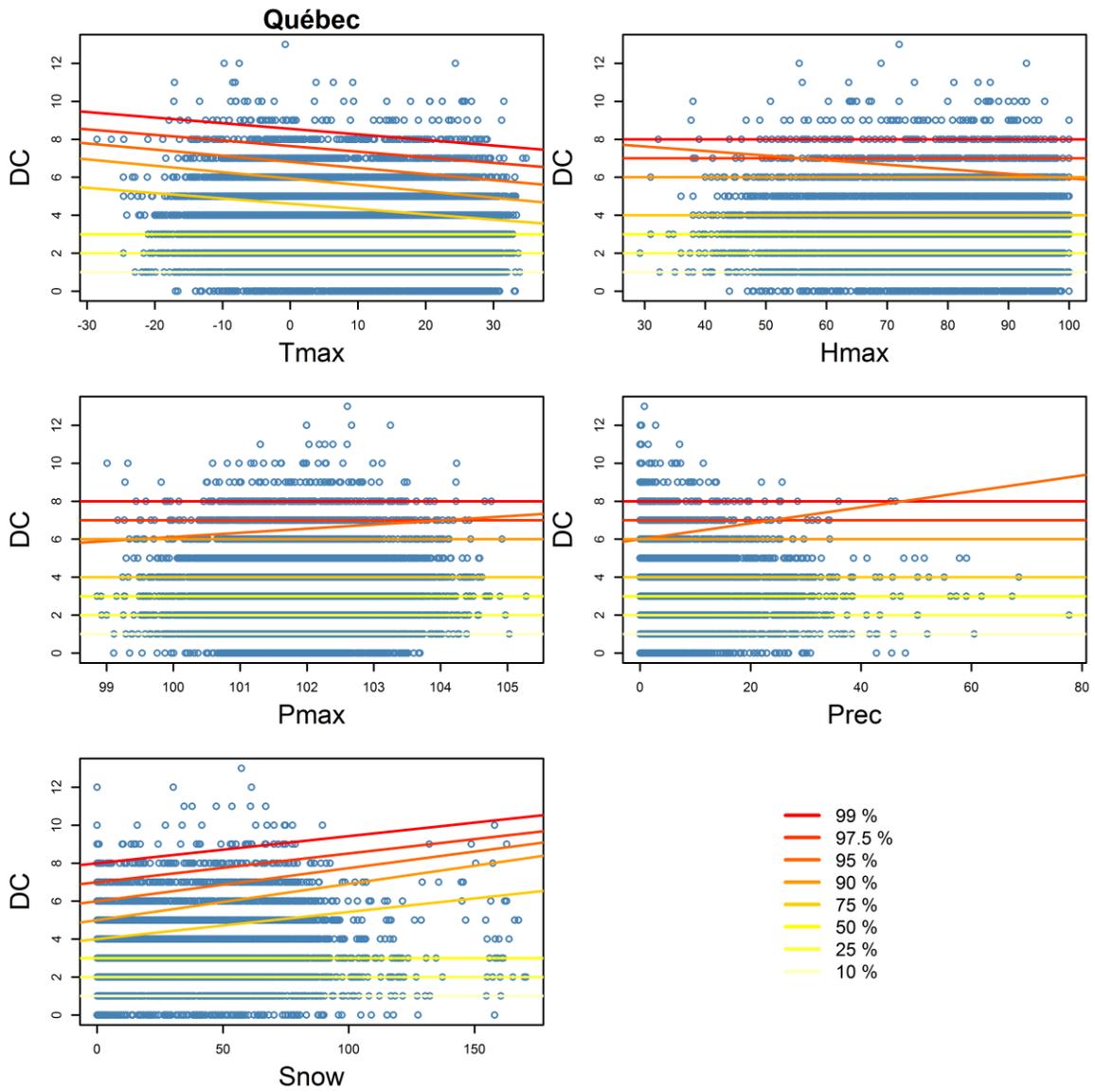


Figure III-10 : régression quantile dans les quantiles à 10, 25, 50, 75, 90, 95, 97,5 et 99 % des décès à Québec.

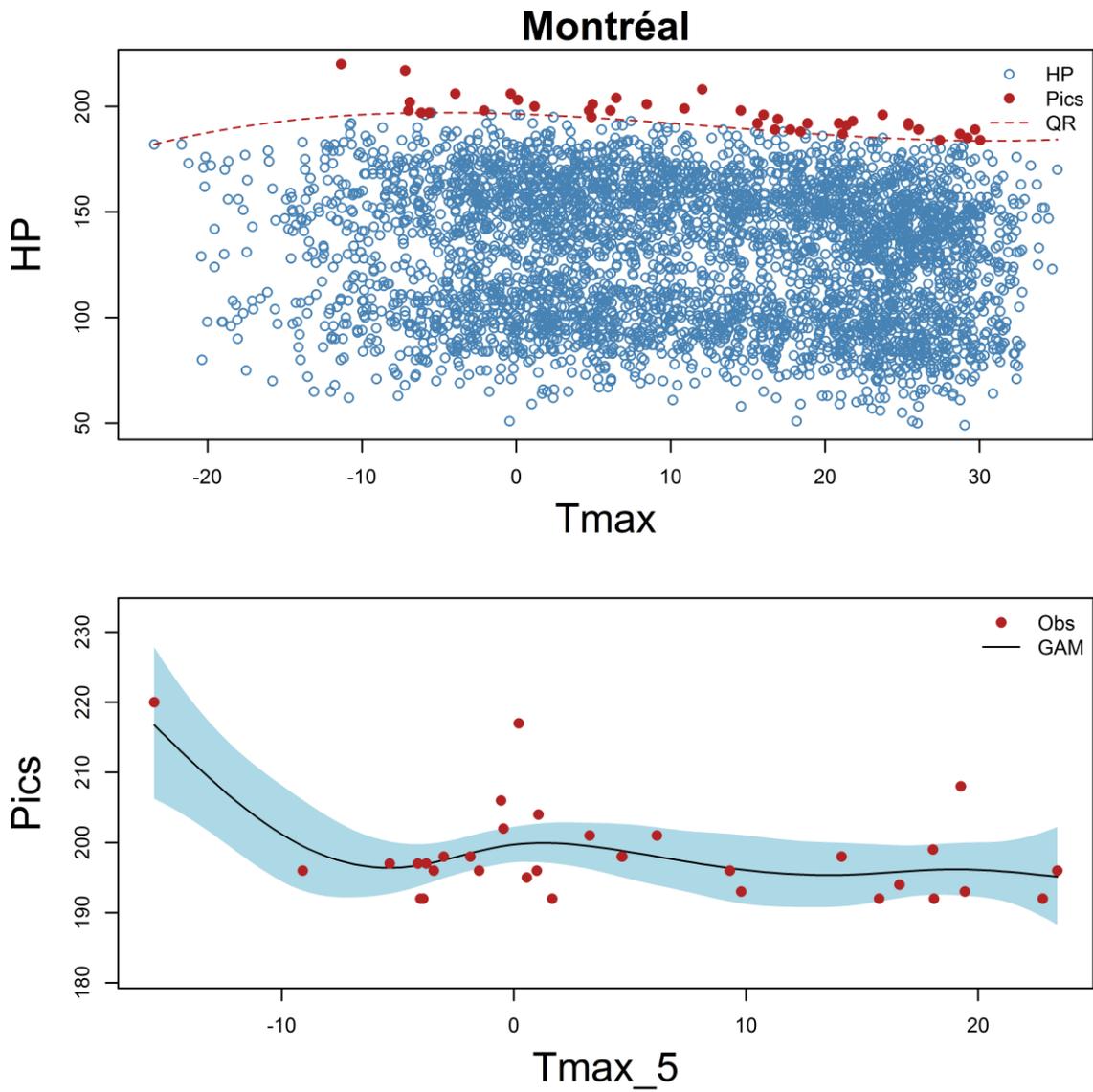


Figure III-11 : Comparaison des fonctions de régression obtenues par régression quantile (haut) et avec l'approche par modèles additifs généralisés de Chiu *et al.* (2016b) (bas) sur les hospitalisations à Montréal.

HP : Hospitalisations

GAM : Modèles additifs généralisés

QR : Régression quantile

Annexes

Cette section contient des détails supplémentaires concernant l'estimation de la régression quantile ainsi que les graphiques concernant les résultats de la régression quantile sur les variables météorologiques minimales.

Régression quantile : estimation

Formellement, la définition usuelle d'un quantile est la suivante : le quantile d'ordre q de la variable Y est la valeur y_q telle que $P(Y \leq y_q) = q$ ou, de même, $F_Y(y_q) = q$ avec F_Y la fonction de répartition associée à la variable Y . Koenker et Bassett (1978) montrent que le quantile d'ordre q peut aussi être défini par le programme de minimisation suivant :

$$\min_{b \in \mathbb{R}} \left[\sum_{t \in \{t: y \geq b\}} q |y_t - b| + \sum_{t \in \{t: y < b\}} (1-q) |y_t - b| \right] \quad (17)$$

pour une variable Y . Dans les deux cas, la médiane est obtenue par $q = 1/2$. Finalement, la régression quantile d'ordre q est définie par

$$\min_{b \in \mathbb{R}} \left[\sum_{t \in \{t: y \geq x_t b\}} q |y_t - x_t b| + \sum_{t \in \{t: y < x_t b\}} (1-q) |y_t - x_t b| \right] \quad (18)$$

avec X une variable explicative. Il s'agit simplement de l'équation (17) étendue pour y inclure des variables explicatives.

Ainsi, les résidus (représentés par $y_t - x_t b$) positifs sont pondérés plus fortement que les résidus négatifs lorsqu'un quantile supérieur à la médiane est étudié, et inversement pour un quantile inférieur à la médiane (Austin *et al.*, 2005). À noter, ce programme de minimisation n'a pas de solution analytique et il est nécessaire d'utiliser des techniques d'approximation numérique pour le résoudre, contrairement aux moindres carrés ordinaires.

Résultats avec les variables météorologiques minimales

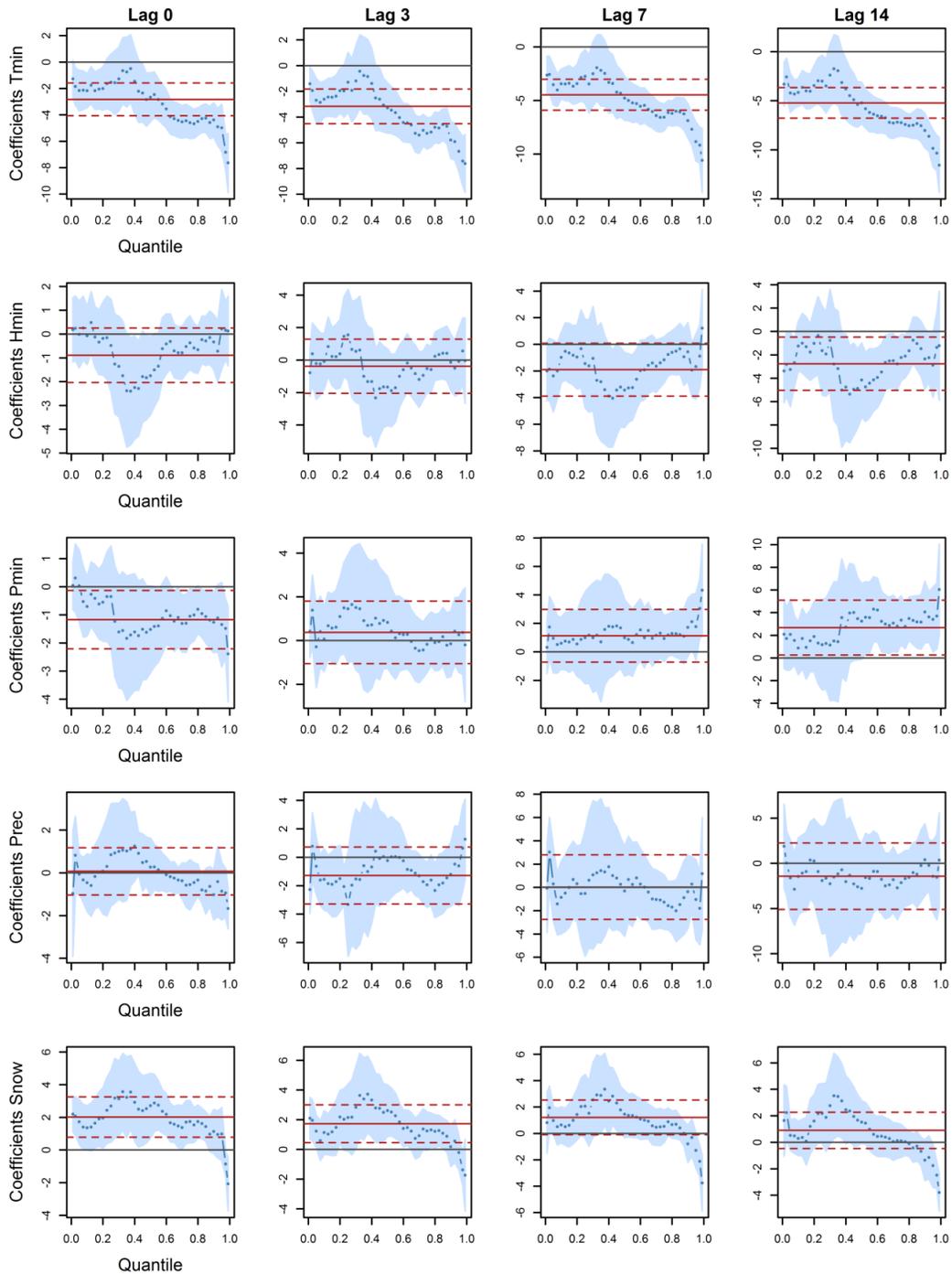


Figure annexe 1 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les hospitalisations à Montréal avec les variables météorologiques minimales.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés ordinaires ; - - - IC des moindres carrés ordinaires ; — $y = 0$

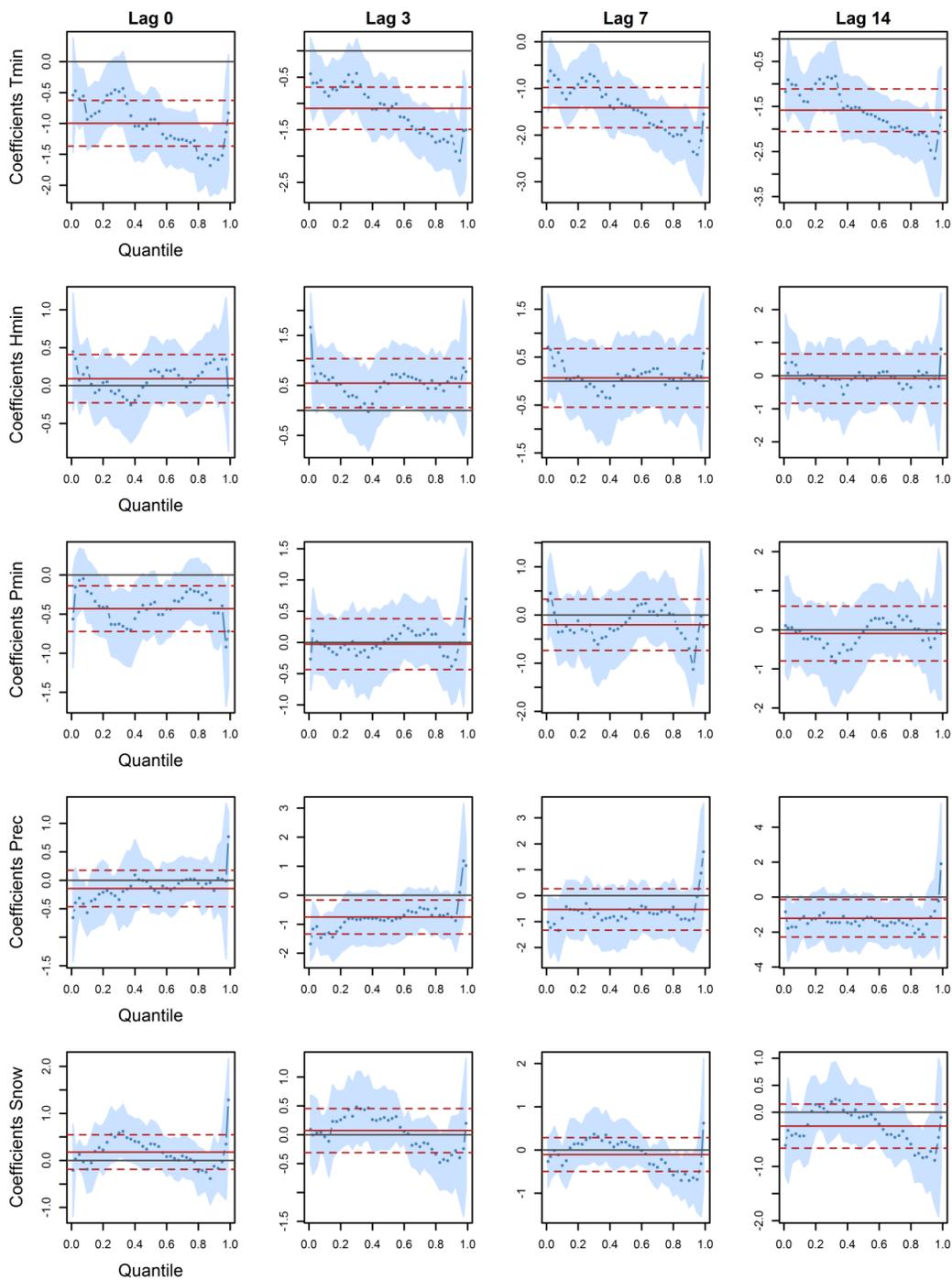


Figure annexe 2 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les hospitalisations à Québec avec les variables météorologiques minimales.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés ordinaires ; - - - IC des moindres carrés ordinaires ; — $y = 0$

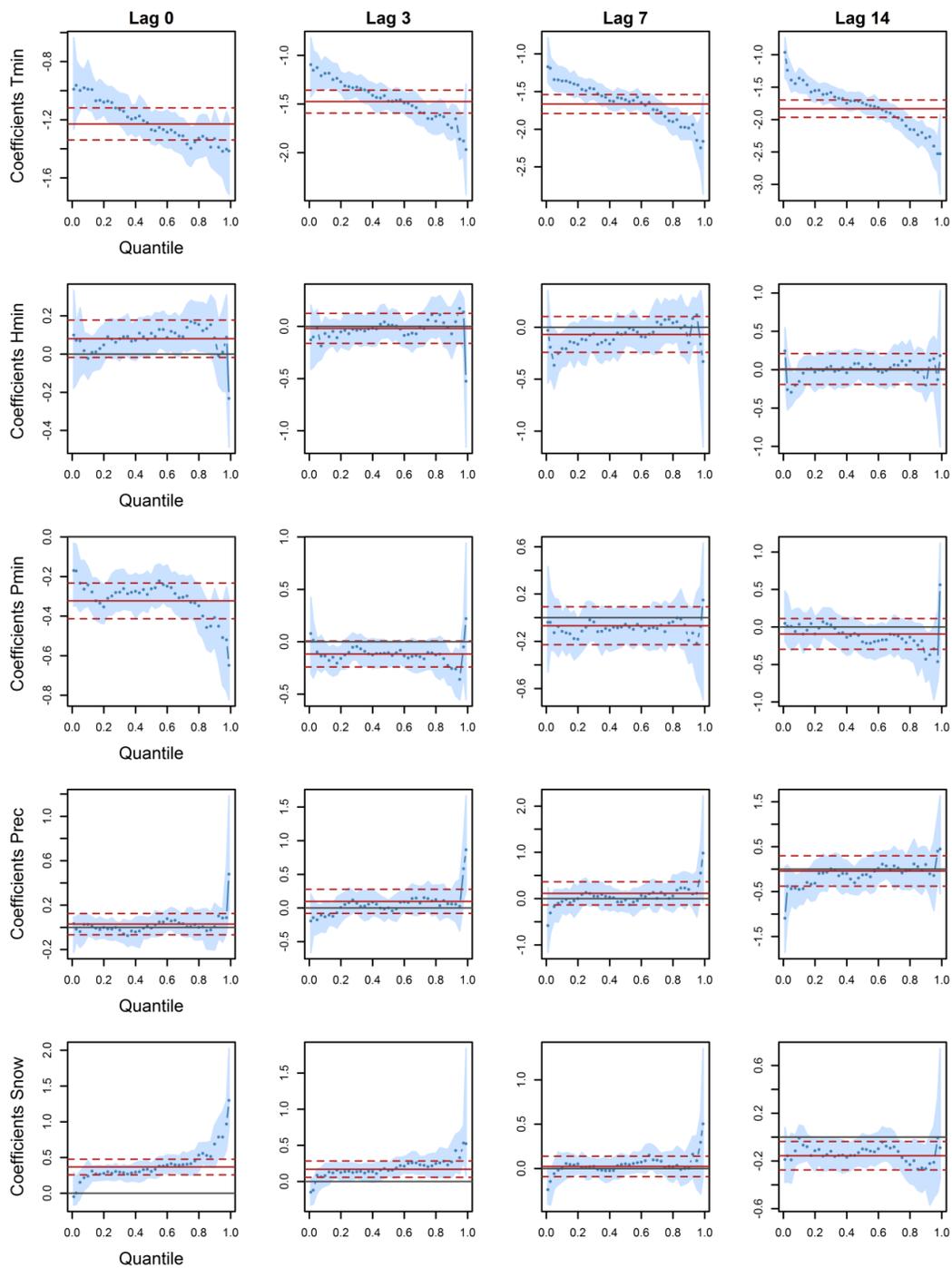


Figure annexe 3 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les décès à Montréal avec les variables météorologiques minimales.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés ordinaires ; - - - IC des moindres carrés ordinaires; — $y = 0$

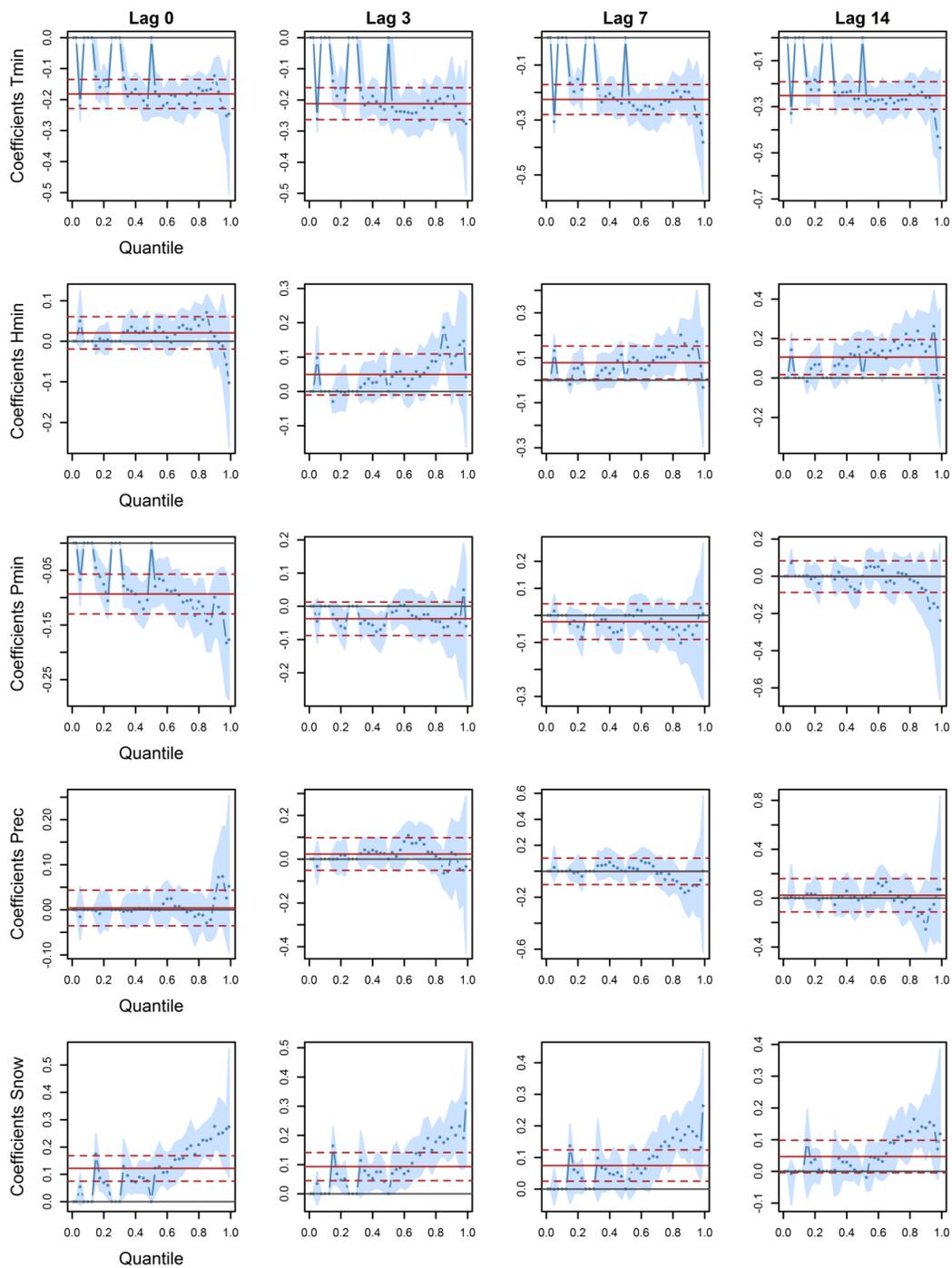


Figure annexe 4 : Résultats de la régression quantile et de la régression par moindres carrés ordinaires sur les décès à Québec avec les variables météorologiques minimales.

● estimations par régression quantile ; ■ IC de la régression quantile ; — estimation par moindres carrés ordinaires ; - - - IC des moindres carrés ordinaires; — $y = 0$

IV. IDENTIFICATION DES EXTRÊMES MÉTÉOROLOGIQUES ET ÉTUDE DE LEURS RELATIONS AVEC LA SANTÉ

Y. Chiu ^a, F. Chebana ^a, B. Abdous ^b, D. Bélanger ^{a,c}, P. Gosselin ^{a,d}

^a INRS-ETE, 490 rue de la Couronne, Québec (QC), G1K 9A9, Canada

^b Université Laval, Département de médecine sociale et préventive, 2325 Rue de l'Université, Québec (QC), G1K 7P4, Canada

^c Centre hospitalier universitaire de Québec, Centre de recherche, 2875 boulevard Laurier, Québec (QC), G1V 2M2, Canada

^d Institut national de santé publique du Québec, 945 avenue Wolfe, Québec (QC), G1V 5B3, Canada

Rapport accepté à l'INSPQ le 10 avril 2017

Résumé

Le Québec et le monde traversent des changements climatiques. Ces changements entraînent en particulier des modifications dans les queues de distributions des variables météorologiques, c'est-à-dire les *extrêmes*. Or, la santé des populations est influencée par les conditions météorologiques et à plus forte raison, par les extrêmes météorologiques. Un problème dans l'étude de ces extrêmes météorologiques est leur manque de définition statistique. De plus, les extrêmes météorologiques ne sont habituellement pas intégrés dans un contexte de régression. Pour combler cette lacune, cette étude évalue plusieurs définitions objectives et l'impact sur la santé des extrêmes météorologiques ainsi extraits est analysé. L'application est considérée sur les hospitalisations par maladies cardiovasculaires, famille de maladies qui influent lourdement sur le système de la santé québécois, en relation avec les variables météorologiques (humidité, température) et de pollution atmosphérique (particules ultra fines) dans les communautés métropolitaines de Montréal et Québec (Canada). Les résultats de cette étude indiquent que les extrêmes météorologiques n'entraînent pas nécessairement des pics sanitaires (valeurs élevées de la variable sanitaire), bien qu'ils soient responsables des pics les plus élevés. De plus, si dans le cas des extrêmes les déviations expliquées sont plus importantes que celles obtenues dans le cas classique (avec toutes les observations), les relations sont majoritairement linéaires. Ces résultats peuvent être utilisés pour améliorer les alertes de santé dans un contexte de surveillance des maladies cardiovasculaires. De plus, ils soulignent les différences qui existent avec le cas classique, indiquant la nécessité d'outils adaptés aux extrêmes météorologiques.

1. Introduction

Le climat de la province du Québec s'est transformé ces dernières décennies, notamment par l'augmentation des températures journalières moyennes dans le sud et des précipitations dans le nord de la province (Ouranos, 2015). De plus, ces changements vont probablement persévérer et même s'accroître dans un futur à court et moyen termes (IPCC, 2014). En particulier, les caractéristiques des *événements météorologiques extrêmes* (EME), soit des phénomènes qui se distinguent par des valeurs inhabituellement éloignées de la moyenne et une fréquence d'apparition rare (Ouranos, 2015), vont changer. Ces EME entraînent des coûts socio-économiques importants, comme illustrés par les inondations du Richelieu de la mi-avril à la fin juin 2011, qui constituent la plus longue catastrophe naturelle de la province du Québec (Environnement et Changement climatique Canada, 2013). Les EME possèdent des caractéristiques très différentes du reste des phénomènes, et surtout des phénomènes moyens, ils nécessitent donc des outils statistiques adaptés.

Les changements climatiques risquent de modifier les relations de la météo avec la santé humaine, par rapport à ce qui est connu aujourd'hui dans la littérature scientifique. En effet, les conditions météorologiques sont reconnues pour avoir un impact sur la santé de la population. Par exemple, l'augmentation des températures moyennes est synonyme d'une augmentation des décès par infarctus du myocarde (Bhaskaran *et al.*, 2009). Les précipitations, les chutes de neige et l'humidité relative moyennes ou totales ont aussi montré des liens significatifs avec la morbidité ou la mortalité, notamment sur le plan cardiovasculaire (Abrignani *et al.*, 2012), ostéo-articulaire (Modarres *et al.*, 2012) ou infectieux (Herrador *et al.*, 2015). Ainsi, en modifiant la distribution des variables météorologiques (donc leurs extrêmes), les changements climatiques risquent d'avoir des effets différents sur la santé humaine de ceux connus actuellement. Tenir

compte de l'influence des EME sur la santé des populations s'avère donc important, car elle diffère de l'influence des valeurs ordinaires. En été, l'effet des vagues de chaleur est à distinguer de l'effet de la température moyenne, par exemple. En témoigne la vague de chaleur de 2003 en France, marquée par de nombreux records de températures et de décès supplémentaires (par rapport à la mortalité habituelle l'été) estimés à 15 000 (Hémon et Jouglà, 2003). De la même façon, les vagues de froid amènent une mortalité plus élevée qu'attendu en moyenne (Kysely *et al.*, 2009).

Pour analyser les EME, il est essentiel de bien les définir en premier lieu. En effet, une mauvaise sélection des EME pourrait biaiser les résultats en incluant de faux EME ou, au contraire, en n'en incluant pas suffisamment. À ce jour, il existe plusieurs façons de les définir. Par exemple, l'Institut royal météorologique de Belgique définit une vague de chaleur par une température de plus de 25 °C persistante pendant au moins 5 jours de suite (Institut Royal Météorologique, 2017). Alors qu'au Québec, la chaleur extrême est définie, selon les régions et en se basant sur les mortalités historiques liées à ces températures, par une température entre 31 et 33 °C le jour et de 16 à 20 °C la nuit, pendant 2 ou 3 journées (Ministère de la Santé et des Services sociaux du Québec, 2016). Ainsi, les seuils de température et la durée sont différents.

Il n'existe donc pas de définition théorique consensuelle d'un EME, bien que certaines « règles » soient d'usage en pratique. À ce sujet, le Groupe d'experts intergouvernemental sur l'évolution du climat propose de sélectionner toutes les observations météorologiques supérieures au quantile à 90 % de la distribution comme des extrêmes (IPCC, 2013). Cette règle est régulièrement appliquée dans la littérature, comme dans Curriero *et al.* (2001), Beniston (2007) ou Sillmann *et al.* (2013). Elle ne fait toutefois pas l'objet d'un consensus et présente des inconvénients (Beniston et Stephenson, 2004), dont le fait que les 10 % de températures supérieures n'ont pas

les mêmes effets sur deux populations étudiées, en raison d'adaptations distinctes (par exemple, deux villes aux climats différents). Cette approche n'inclut donc pas nécessairement les effets sanitaires liés, comme le fait la notion de chaleur au Québec.

Même si la théorie n'est pas unifiée, en pratique, une quantité importante d'études portent sur les EME (canicule, vague de froid, inondations, etc.), étant donné leur importance sur la santé humaine. Cependant, la plupart de ces études ne réunissent pas toutes les étapes méthodologiques nécessaires à la réalisation du présent travail. Par exemple, même si l'étude d'Hales *et al.* (2003), commanditée par l'Organisation mondiale de la santé, fait mention explicite d'« extrêmes climatiques » et fournit un résumé d'impacts sanitaires (p. ex. maladies cardiovasculaires) engendrés par les extrêmes météorologiques (p. ex. sécheresses), une méthodologie statistique dans un cadre de régression y fait défaut. Il en va de même pour Greenough *et al.* (2001), qui étudient les extrêmes tels que les tempêtes de neige ou les inondations, sans donner de sens statistique à ces événements. En fait, de telles études ne considèrent pas nécessairement de modèles explicatifs, mais ont un aspect descriptif. Des définitions d'EME y sont aussi absentes.

Un EME est souvent considéré comme tel à partir du moment où son intensité, sa rareté ou son impact sur les populations est notable. Certains peuvent cumuler ces critères. La canicule européenne de 2003, étudiée de façon exhaustive, en constitue un exemple. Toutefois, les canicules, ou de façon équivalente la variable de température, sont étudiées de façon séparée la plupart du temps une fois de plus sans lien statistique (voir p. ex. la revue de littérature de Kovats et Hajat, 2008). Il en va de même pour les autres variables météorologiques. Ainsi, dans la revue de la littérature de De Sario *et al.* (2013), il est fait mention d'événements météorologiques extrêmes, mais aucune définition statistique n'est fournie. Cela ne signifie pas pour autant que les modèles statistiques ne sont pas présents dans l'étude des EME. Par exemple, Towler *et al.*

(2010) ou Waylen *et al.* (2012) utilisent la théorie des valeurs extrêmes, bien établie en statistique, dans au moins un outil avancé de modélisation, dans des contextes hydrologique et météorologique. En revanche, ils n'étudient pas l'effet des EME sur la santé des populations, bien que la deuxième étude fasse tout de même mention de recommandations sanitaires générales.

Parmi les études qui s'insèrent dans un cadre de modélisation, celle de Li *et al.* (2012) est réalisée dans une optique similaire à l'étude présente. Les auteurs analysent au travers d'un modèle de régression l'effet des températures sur la morbidité, en tentant d'établir des seuils extrêmes (quantiles élevés). Leur étude, cependant, ne répond pas entièrement aux problématiques étudiées dans ce chapitre, tant sur le plan des variables à expliquer (hospitalisations toutes causes) et explicatives retenues (l'ensemble des observations plutôt que la sélection des EME) que sur la population visée (différents groupes d'âge, à Milwaukee, Wisconsin).

Une des raisons pour lesquelles définir un EME n'est pas aisé est qu'il dépend habituellement de plus d'une variable. Par exemple, comme mentionnées plus tôt, les vagues de chaleur sont souvent définies en fonction de leur durée et de leur intensité. Il est aussi commun de caractériser les pluies intenses par leur intensité, leur durée et leur fréquence (*courbes idf*, voir par exemple Mailhot *et al.*, 2007). Or, passer d'une définition univariée à une définition bivariée et plus généralement multivariée entraîne plusieurs complications. En particulier, ordonner un échantillon non univarié n'est plus aussi direct. En l'absence d'un classement simple des observations, il est compréhensible que la définition d'un EME demande plus d'étapes. Pour pallier cette difficulté, il est possible d'utiliser des indices météorologiques (Kysely et Kim, 2009; Chebana *et al.*, 2013). Ceux-ci permettent de combiner plusieurs variables météorologiques en un indicateur, ce qui ramène des observations multivariées à un contexte univarié. D'un point de vue statistique, les fonctions de profondeur répondent aussi à cette problématique. En effet,

elles ont pour objectif d'ordonner un échantillon multivarié, tout en respectant certaines propriétés statistiques (Zuo et Serfling, 2000). Elles ne sont cependant pas exploitées en météorologie et santé publique. Il est tout de même intéressant de citer l'étude de Chebana et Ouarda (2011b) qui propose une méthodologie d'identification des observations extrêmes basée sur les fonctions de profondeur, avec une application en pollution de l'air.

La présente étude propose donc, dans un premier temps, d'évaluer certaines définitions statistiques pertinentes aux EME puis, dans un deuxième temps d'explorer les liens qui peuvent exister entre la survenue d'un EME et la santé des populations. Au meilleur de la connaissance des auteurs, aucune étude en santé publique ou en météorologie ne mène une méthodologie complète comme celle proposée ici. La plupart des études se concentrent sur une seule étape avec des objectifs pouvant différer. De plus, les outils avancés ici sont rarement utilisés dans l'étude des EME.

L'application de régression est faite sur les maladies cardiovasculaires (MCV). Au Québec, les MCV sont une catégorie de maladies à surveiller particulièrement, car elles constituent la deuxième cause de mortalité et comptaient pour le quart de la totalité des décès en 2008 (Girard, 2013). Les MCV représentent aussi la catégorie de maladies pesant financièrement le plus sur le système de santé québécois (Daigle, 2007) et figurent parmi les maladies affectées par les conditions météorologiques (Portier *et al.*, 2010). Par exemple, les effets des variations de la pression atmosphérique sur l'infarctus du myocarde ou de la température ambiante sur les accidents vasculaires cérébraux sont reconnus (Chang *et al.*, 2004; Houck *et al.*, 2005). Étudier les MCV dans un contexte de changement climatique est une problématique encore ouverte, voir par exemple Bobb *et al.* (2014) ou Costello *et al.* (2009). Il est donc important de comprendre l'effet des EME sur les MCV, d'autant plus que ces extrêmes peuvent avoir une action stressante

sur le réseau de la santé (par exemple en termes du nombre d'hospitalisations, Dolney et Sheridan, 2006; Knowlton *et al.*, 2009).

Le reste du chapitre est organisé comme suit : la section 2 décrit les outils statistiques utilisés dans cette étude et la section 3 contient les résultats. La discussion est présentée en section 4, suivie de la conclusion en section 5. Enfin, la section 6 indique les principales limitations de l'étude. À noter, les tableaux et figures utiles à la compréhension du texte suivent les références. D'autres figures sont toutefois placées en annexe.

2. Analyse statistique

L'analyse statistique compte deux étapes principales : 1) l'identification des extrêmes météorologiques et 2) la mise en relation de ces derniers avec les MCV dans un modèle de régression. La première étape est décrite dans les sous-sections 2.1 (extrêmes univariés) et 2.2 (extrêmes multivariés), tandis que la deuxième étape est décrite dans la dernière sous-section (2.3).

Les extrêmes utilisés dans cette étude peuvent être classés dans deux catégories : les extrêmes univariés (selon une seule variable météorologique) et les extrêmes multivariés (selon deux variables météorologiques ou plus). À noter, il est ici uniquement question d'extrêmes définis au sens statistique plutôt que d'extrêmes mesurés selon les impacts engendrés (p. ex., impacts socioéconomiques).

2.1. Extrêmes univariés

Les extrêmes univariés sont définis à l'aide d'une seule variable météorologique, puis les valeurs des autres variables météorologiques lui sont associées.

2.1.1. Méthode du seuil et des blocs

La théorie des valeurs extrêmes Coles (2001) permet l'identification et l'étude de valeurs extrêmes en météorologique. Par exemple, Blanchet *et al.* (2009) analysent les chutes extrêmes de quantité de neige dans les Alpes à l'aide d'outils de la théorie des valeurs extrêmes. Dans la présente étude, puisque la distribution des EME n'est pas nécessairement d'intérêt, seules les deux méthodes d'extraction des valeurs extrêmes sont nécessaires. La première, la méthode du seuil, consiste à fixer un seuil élevé (par exemple, à partir du quantile à 90 %) et à sélectionner toutes les observations supérieures comme des extrêmes. Quant à la seconde méthode, dite des blocs, elle consiste à établir des séparations verticales des données (selon les semaines, mois, années, etc.), puis à sélectionner la valeur maximale dans chaque bloc comme un extrême. Chiu *et al.* (2016a) se servent de la méthode du seuil et de celle des blocs dans une étude « symétrique » à la présente, c'est-à-dire qu'ils définissent les pics sanitaires (valeurs extrêmes des hospitalisations et des décès) grâce aux outils de la théorie des valeurs extrêmes, encore méconnue en santé publique.

Une fois que les extrêmes sont extraits, il faut leur associer les observations des autres variables météorologiques, qui ne sont pas nécessairement extrêmes. Un exemple simple d'observation associée à un extrême météorologique s'écrit ainsi

$$y_i = t_i^* + h_i + \varepsilon_i \quad (19)$$

avec y pour la variable sanitaire, t la température et h l'humidité, l'étoile indiquant le caractère extrême de la variable. Cette régression des extrêmes météorologiques extraits par les méthodes des blocs et celle du seuil constitue la première à la connaissance des auteurs, à plus forte raison dans un contexte de santé publique.

Dans cette étude sont explorés des blocs de 7, 30 et 60 jours et des seuils aux quantiles 90, 95 et 99 %. Des exemples d'extrêmes de température à Montréal extraits par les deux méthodes sont illustrés à la Figure IV-1.

2.1.2. Variations

Il peut être plus intéressant d'étudier les variations entre les observations météorologiques plutôt que d'utiliser les valeurs brutes. Par exemple, Jimenez-Conde *et al.* (2008) ont mis à jour une relation significative entre les changements de pression atmosphérique et l'accident vasculaire cérébral à Barcelone, en Espagne (voir aussi Houck *et al.* (2005), aux États-Unis). Puli *et al.* (2004) ont montré un effet significatif des variations de la température à Split (Croatie) sur la tachycardie ventriculaire tandis que Danet *et al.* (1999) ont illustré un effet des variations d'humidité relative et de la température sur l'infarctus du myocarde, à Lille (France). Ces quatre études se concentrent sur les variations entre le jour même de l'issue sanitaire et la veille (voir aussi Goldberg *et al.*, 2015). Il en sera donc fait de même lors de l'application de cette méthode ici.

Ainsi, un extrême météorologique n'est plus défini selon les variables météorologiques brutes, mais selon les valeurs des variations. Une observation de variation de température s'écrit alors

$$y_i = (t_i - t_{i-1}) + \varepsilon_i \quad (20)$$

et les méthodes précédentes des blocs et du seuil sont appliquées pour extraire les extrêmes de ces nouvelles observations. Étant donné que la sélection des extrêmes des variations par ces méthodes d'extraction n'a pas encore été appliquée en météorologie ou en santé publique, aucune référence n'existe concernant la taille des blocs ou les seuils à considérer. Ainsi, les mêmes blocs (7, 30 et 60 jours) et seuils (quantiles 90, 95 et 99 %) sont explorés.

2.2. Extrêmes multivariés

Considérer des extrêmes univariés peut ne pas être représentatif de la réalité. En effet, les extrêmes météorologiques font plutôt intervenir un ensemble de variables. Par exemple, une vague de chaleur est définie selon son intensité et selon sa durée (un certain seuil de température atteint sur plusieurs jours). Définir un extrême multivarié n'est cependant pas simple, comme expliqué dans le prochain paragraphe.

L'extension directe des méthodes du bloc et du seuil au cadre multivarié consiste à extraire les extrêmes de chaque variable puis de les associer entre eux (extrêmes *componentwise* en anglais, Coles *et al.*, 1999). Les études dans le cadre multivarié de la théorie des valeurs extrêmes s'attardent principalement à la modélisation. En effet, la théorie mathématique associée est élégante. En pratique, elle peut toutefois s'avérer de moindre usage, car il est souvent possible qu'aucune observation réelle ne corresponde au maximum de toutes les variables. Par exemple, la Figure IV-2 illustre des extrêmes sélectionnés par bloc et par seuil pour la température et l'humidité. Il est clair qu'un extrême de température et un extrême d'humidité ne surviennent pas le même jour. Avec deux variables météorologiques, les deux méthodes se soldent donc par des séries dont les extrêmes ne coïncident pas au niveau des dates. Considérant que le maximum pour la température survient au jour i et celui pour l'humidité au jour j , une observation bivariée dans cet exemple s'écrit

$$y_{\gamma} = t_i^* + h_j^* + \varepsilon_{\gamma} \quad (21)$$

En obtenant des dates différentes, il est donc impossible d'associer une date cohérente pour la variable sanitaire. Le problème s'accroît à mesure que le nombre de variables météorologiques augmente. Pour pallier cette difficulté, les fonctions de profondeur et les indices météorologiques sont utilisés.

2.2.1. Fonctions de profondeur

Les fonctions de profondeur (Zuo et Serfling, 2000) sont une solution envisagée pour définir les extrêmes multivariés. Il n'est pas aisé d'attribuer un « ordre » à des observations au-delà d'une dimension, ce à quoi les fonctions de profondeur permettent de répondre. Elles étendent directement la notion d'ordre à des données multivariées en leur assignant une valeur de « profondeur », tenant compte de leur valeur dans chaque dimension. En classant les observations multivariées, elles établissent un ordre dans les observations. Le classement obtenu est centré sur un point central multivarié, qui possède la valeur de profondeur maximale de l'échantillon. Plus les observations s'en éloignent, plus leurs valeurs de profondeur diminuent. Ainsi, il est possible de définir des extrêmes météorologiques dépendant de plusieurs variables comme les observations les plus éloignées de ce point central. La fonction de profondeur retenue ici est celle de Mahalanobis (voir l'Annexe A) et les seuils considérés pour définir les extrêmes dans ce cas sont de 1, 5 et 10 % (les profondeurs les plus faibles correspondent aux points les plus éloignés du point central). La Figure IV-3 illustre la fonction de profondeur de la température et de l'humidité relative. Le point dont la profondeur est maximale (en d'autres termes, la médiane bivariée) se situe autour d'une température de 10 °C et d'une humidité relative de 80 %.

Les fonctions de profondeur ne sont pas utilisées actuellement en météorologie ni en santé. Un exemple rare d'application en qualité de l'air est l'étude de Chebana et Ouarda (2011b), qui introduisent une méthodologie faisant usage des fonctions de profondeur pour sélectionner des événements extrêmes multivariés. Leur méthodologie implique de plus un paramètre u , qui contrôle « l'orientation » de l'espace dans laquelle sélectionner les extrêmes, et un paramètre λ , qui contrôle le nombre d'extrêmes à sélectionner (Figure IV-4). Ici, étant donné la dimension exploratoire de l'étude (autant d'un point de vue de l'extraction des extrêmes que de l'intégration

de ces derniers dans une régression), ces paramètres supplémentaires ne sont utilisés. Une comparaison entre nos résultats et l'approche de Chebana et Ouarda (2011b) est tout de même effectuée en fin de résultats, à titre illustratif.

2.2.2. Indices météorologiques

Les indices météorologiques permettent de considérer des extrêmes multivariés. En effet, ils sont construits à l'aide de plusieurs variables météorologiques, souvent la température et l'humidité. En considérant les interactions que les différentes variables intégrées peuvent avoir et leurs effets sur l'organisme, un ordre peut être établi pour ces indices. Il est ainsi possible d'attribuer une intensité numérique, ce qui permet de définir des extrêmes. Les deux indices principaux utilisés dans les études de santé environnementale sont l'indice de chaleur et l'humidex. L'indice de chaleur, créé par le *National Weather Service* des États-Unis, est utilisé pour évaluer la température ressentie par le corps humain, ajustée pour l'humidité (Weir, 2002). L'humidex, créé pour les besoins d'Environnement Canada, vise le même but que l'indice de chaleur (Weisskopf *et al.*, 2002). De plus, l'humidex est utilisé par Chebana *et al.* (2013) pour définir des seuils de chaleur au Québec, Canada. Ces deux indices sont aussi parfois utilisés de façon complémentaire, comme dans l'étude de Smoyer-Tomic et Rainham (2001) à Toronto (Canada).

Les indices météorologiques sont calculés pour chaque jour et représentent ainsi les nouvelles séries d'observations brutes (à la place des observations météorologiques). Des données brutes sont extraites les observations extrêmes, à l'aide des méthodes précédemment citées (seuil et blocs). En fait, cela revient à réduire le cas multivarié à un cas univarié en intégrant les différents variables météorologiques au sein d'une même entité. Il n'existe pas d'étude qui associe des méthodes d'extraction d'extrêmes avec les indices météorologiques, puis qui les insère dans un

contexte de régression, impliquant une optique exploratoire des résultats présentés ici. Pour plus de détails sur le calcul des deux indices, voir l'annexe B.

2.3. Régression

Une fois les EME sélectionnés, ils sont mis en relation avec les hospitalisations pour cause de MCV. Les variables explicatives sont lissées pour être intégrées dans des modèles additifs généralisés (Wood, 2006). Ce sont des modèles de régression dits « semi-paramétriques » qui permettent une grande flexibilité dans l'exploration des relations entre la santé et la météo (Jbilou et El Adlouni, 2012).

L'influence des variables météorologiques se juge au travers d'une fonction, qui est l'équivalent d'un coefficient de régression dans une régression classique. La complexité de la fonction est déterminée par son « degré de liberté ». Plus il est élevé, plus la fonction s'ajuste aux données et plus l'interprétation est difficile. Par exemple, un degré de liberté de 1 représente une droite tandis qu'un degré de liberté de 2 représente une fonction quadratique. Des splines de régression sont utilisées dans cette étude afin d'évaluer les fonctions d'influence. Un test statistique permet de juger de la significativité de ces dernières (Wood, 2006). La déviance expliquée (l'équivalent du R^2 dans la régression classique) des modèles est aussi calculée.

Dans le présent projet, des lags sont considérés, car l'impact des conditions météorologiques sur la santé peut durer plusieurs jours (Braga *et al.*, 2002; Goldberg *et al.*, 2011; Hajat *et al.*, 2016) et l'évènement sanitaire peut survenir un ou plusieurs jours après l'extrême météorologique. Les observations sanitaires sont donc sélectionnées un certain nombre de jours après l'apparition de l'extrême. Il n'y a toutefois pas de lag absolu, mais plutôt une étendue de lags spécifiques à chaque variable météorologique et à chaque maladie. De plus, au meilleur de la connaissance des auteurs, il n'existe pas de littérature sur laquelle se baser pour sélectionner les lags appropriés

pour la relation entre les *extrêmes météorologiques* et les MCV. Cependant, en se basant sur le cas classique, Curriero *et al.* (2002) mentionnent que les lags les plus intéressants pour les décès par MCV vont jusqu'à 3 jours précédant la survenue de l'événement. Concernant les hospitalisations par MCV, Ebi *et al.* (2004) considèrent des lags jusqu'à 7 jours. Certaines études vont plus loin, par exemple Doyon *et al.* (2006) incluent dans leur modèle des lags pour la température allant jusqu'à 20 jours, mais ils expliquent les décès toutes causes. Aussi, dans la présente étude, des lags jusqu'à 7 jours ont été retenus et sont évalués. À noter, les modèles à effets retardés distribués peuvent être utilisés (Gasparrini *et al.*, 2010) pour analyser plusieurs lags simultanément, en plus de prendre en compte l'effet non linéaire. Ils sont populaires dans l'étude des relations entre la santé et le climat (Goldberg *et al.*, 2011; Yang *et al.*, 2015a; Phung *et al.*, 2016a), mais ils nécessitent l'estimation d'un nombre important de coefficients (par rapport à un modèle additif généralisé). En effet, leur caractéristique principale est la fonction bidimensionnelle estimée pour chaque variable explicative (une dimension pour la variable explicative et une dimension pour les lags appliqués sur cette dernière). Ces modèles demandent donc un nombre important de données pour être estimés. Ils ont toutefois déjà été utilisés dans le cas de plus grands échantillons que ceux présents ici pour l'étude d'extrêmes météorologiques (Lavigne *et al.*, 2014 ; Gasparrini *et al.*, 2015).

Les variables météorologiques utilisées dans les modèles explicatifs de cette étude sont la température et l'humidité maximales. Les variables minimales et moyennes n'ont pas montré de différences significatives dans les résultats. De plus, Barnett *et al.* (2010) montrent que les températures maximales, minimales ou moyennes résument la même information et que leurs potentiels explicatifs sont très similaires. La pollution atmosphérique et les conditions météorologiques étant souvent fortement corrélées (Zanobetti et Peters, 2015), une variable de

pollution atmosphérique est aussi incluse dans les modèles, celle des particules ultrafines de 2,5 micromètres (PM_{2,5}).

Tous les résultats sont obtenus à l'aide du logiciel libre R (R Core Team, 2016) et le seuil pour les tests de significativité statistique est fixé à $\alpha = 5\%$.

3. Résultats

Étant donné la quantité importante de résultats générés, le choix est fait de ne présenter en détail que les résultats et interprétations du cas des hospitalisations à Montréal, en se concentrant sur deux variables explicatives : la température et l'humidité. Travailler dans une dimension supérieure à 2 augmenterait nettement la complexité des algorithmes, en particulier pour les fonctions de profondeur. Les autres variables météorologiques sont présentées dans les figures à titre informatif mais ne sont pas considérées pour la sélection d'EME.

3.1. Extrêmes univariés

Cette sous-section présente les résultats pour les extrêmes extraits par seuil et par blocs, puis les résultats des extrêmes des variations.

3.1.1. Extrêmes par seuils et blocs

Les résultats de la modélisation des extrêmes extraits par seuil et bloc sont résumés au Tableau IV-1. Par souci de concision, seuls les modèles avec des variables significatives sont présentés. Pour la méthode du seuil, les modèles contenant une variable significative au jour même sont ceux incluant la température au seuil à 99 % et l'humidité au seuil à 95 %. Les déviations expliquées associées sont de 14 % et 6,2 %, respectivement. Pour la méthode du bloc, un seul modèle possède une variable significative au jour même. Il s'agit du modèle avec les extrêmes de température pour des blocs de 7 jours. La déviance expliquée est toutefois faible (2,21 %) et il est

possible que la significativité ne soit due qu'à un effet de taille (plus la taille d'un échantillon est grande, plus de petits changements seront significatifs, selon Sullivan et Feinn, 2012). Les facteurs qui peuvent pointer vers un effet de taille sont : une faible déviance expliquée, un grand échantillon et une variable jugée significative dans le modèle.

Plusieurs lags de 3 à 7 jours sont retenus et tous les modèles résultent en des relations linéaires. De plus, les déviances expliquées avec lags sont supérieures au cas sans les lags, et vont jusqu'à 24 %, ce qui est relativement important dans un contexte de relations entre santé et la météo. Un phénomène particulier à relever est que la droite de régression pour les extrêmes de température extraits par seuil a une pente (coefficient directeur associé à la variable température) négative au lag 4 et positive au lag 5. Cela indiquerait un effet inverse des extrêmes de température sur les hospitalisations, entre les 4^{ème} et 5^{ème} jours suivant l'extrême, indiquant un possible effet de moisson après l'extrême de température. Enfin, la variable de pollution atmosphérique n'était pas significative dans les modèles sans lags, mais elle le devient une fois que les lags sont appliqués, signifiant un effet significatif uniquement dans les jours suivant l'extrême de particules ultrafines.

Les extrêmes de température sélectionnés par seuil, illustrés pour le seuil à 99 %, ainsi que les observations correspondantes des autres variables sont présentés à la Figure IV-5. Puisque le seuil supérieur des températures est évalué, les observations retenues surviennent naturellement en été. Seuls les étés les plus chauds sont retenus, c'est-à-dire ceux de 1999, 2001, 2002, 2003, 2005 et 2006. Sur le graphique, il n'apparaît pas de lien entre les extrêmes et les hospitalisations qui leur sont associées, car ces dernières ne sont pas nécessairement des pics sanitaires. Cependant, même si plusieurs hospitalisations non extrêmes sont associées aux extrêmes météorologiques retenus, les pics sanitaires sont aussi retenus dans chaque été considéré. La Figure IV-6 montre le nuage de points des hospitalisations en fonction des températures, ainsi

qu'avec les extrêmes sélectionnés au seuil à 99 %. La relation est linéaire et croissante en fonction des extrêmes de températures.

La Figure IV-7 montre les extrêmes de la température sélectionnés par la méthode des blocs (illustrée pour une taille de 60 jours). À nouveau, les hospitalisations ne constituent pas toutes des pics sanitaires. La principale différence avec la méthode du seuil est que les blocs sélectionnent des observations en été comme en hiver. La Figure IV-8 ne montre aucune association particulière entre les hospitalisations et la température.

Les températures froides ont aussi été considérées (seuil à 1, 5 et 10 %) afin de n'examiner que les hivers. Toutefois, ces ajouts n'ont apporté aucune amélioration significative des modèles, ni une plus grande sélection de pics sanitaires.

3.1.2. Extrêmes des variations

Un seul modèle est jugé significatif pour les extrêmes des variations (Tableau IV-2). Il s'agit du cas des extrêmes des variations d'humidité pour une taille de blocs de 7 jours. La déviance expliquée est malgré tout très faible (1,94 %).

La Figure IV-9 montre le nuage de points des extrêmes des variations d'humidité et des hospitalisations associées. Les échelles numériques sont modifiées du fait de la transformation des variables (variations journalières). La régression des extrêmes sur les hospitalisations est aussi présentée. Elle est linéaire et croissante en fonction des extrêmes des variations de l'humidité. L'application des lags n'a pas apporté de résultats significatifs dans le cas des variations. Ainsi, un seul modèle a montré une variable significative, mais il est possible qu'elle ne soit due à un effet de taille d'échantillon.

3.2. Extrêmes multivariés

Cette sous-section présente les résultats pour les extrêmes extraits par les fonctions de profondeur, puis par les indices météorologiques.

3.2.1. Extrêmes par fonctions de profondeur

La Figure IV-10 illustre les observations retenues pour le seuil à 1 % de la profondeur et les hospitalisations associées à ces extrêmes, tandis que la Figure IV-11 montre les observations associées pour les variables météorologiques. Il s'agit effectivement d'extrêmes météorologiques car les observations sélectionnées sont extrêmes soit par la température ou l'humidité, soit par les deux variables en même temps. À noter, les fonctions de profondeur sélectionnent des observations dans toutes les saisons. Cependant, en général, peu de pics sanitaires sont sélectionnés (sauf exception, comme en hiver 2001). Les hospitalisations avec lag (dans la semaine suivant les extrêmes) sont présentées à la Figure IV-12. À nouveau, les pics sanitaires sont peu nombreux.

Les résultats des modèles avec les extrêmes définis grâce aux fonctions de profondeur sont présentés au Tableau IV-3. Deux modèles sont significatifs au jour même, le premier avec les extrêmes de température et d'humidité dont la profondeur est inférieure au seuil à 10 %, le deuxième avec la température et les particules ultrafines dont la profondeur est inférieure au seuil à 1 %. Ce dernier modèle est particulièrement intéressant en matière de déviance expliquée et de significativité. En effet, la fonction des deux variables est significative, résultant en une déviance expliquée de 18,9 %. Puisqu'il s'agit d'une fonction de régression de deux variables et que le degré de liberté est de 2, la fonction s'interprète comme un plan (l'équivalent bivarié d'une droite de régression). Un plan de régression est utilisé ici car le nombre d'observations est restreint, une

surface non-linéaire aurait demandé plus d'observations. Le plan de régression est présenté à la Figure IV-13a et il est possible d'observer qu'il est décroissant en fonction des deux variables (lorsqu'une variable augmente, l'effet diminue). Ces extrêmes météorologiques sont ainsi représentatifs de la plus haute morbidité en hiver. En revanche, les modèles associés aux extrêmes de température et de particules ultrafines n'ont pas donné de résultat significatif à l'application des lags, contrairement à celui du jour même. Il faut toutefois noter qu'un extrême de température et de particules ultrafines sur quatre (26 %) est associé à un pic sanitaire le jour même, tel que défini par le quantile à 90 % des observations sanitaires. Ce nombre monte jusqu'à 29 % une semaine après.

Le modèle avec la température et l'humidité est significatif avec un lag aux 6^{ème} et 7^{ème} jours. Il s'agit à nouveau de deux plans de régression (les degrés de liberté sont égaux à 2) et les déviations expliquées sont relativement élevées (27 %). Les plans de régression sont très similaires et celui pour le lag de 7 jours est présenté à la Figure IV-13b. Il est croissant en fonction des deux variables : plus la température ou l'humidité augmente, plus le nombre d'hospitalisations est élevé. Ces extrêmes de température et d'humidité mettent donc plutôt en valeur les périodes de chaleur intense en été, contrairement aux extrêmes de température et de particules ultrafines. Environ 18 % des extrêmes de température et d'humidité sont associés à des pics sanitaires le jour même (21 % le jour suivant).

Il est intéressant de constater que le pic sanitaire le plus important (220 hospitalisations au 12 janvier 2004, Figure IV-12) est associé à un extrême météorologique extrait par profondeur. Aucune autre méthode proposée dans ce chapitre n'est parvenue à « capturer » ce pic sanitaire. La Figure IV-14 résume les conditions météorologiques le jour du pic sanitaire en question, ainsi que 7 jours avant sa survenue. Il est possible de constater que l'extrême sélectionné par la

fonction de profondeur (8 janvier) est aussi extrême d'un point de vue univarié, car la température et l'humidité sont des minima locaux. Il peut ainsi s'agir d'un mélange de conditions météorologiques qui a provoqué ce pic, mélange qui a été capté par la fonction de profondeur. Il peut toutefois s'agir aussi de variables externes différentes de la météo, comme l'influenza.

Enfin, la méthodologie de Chebana et Ouarda (2011b) est utilisée pour extraire les extrêmes météorologiques et les comparer avec ceux déjà obtenus. Ils sont présentés à la Figure IV-15. Le paramètre d'orientation u est choisi dans l'intervalle $[0;0,25]$, correspondant ainsi aux températures chaudes les plus extrêmes, et λ est choisi de sorte à obtenir un échantillon de taille comparable à ceux du Tableau IV-3. Cependant, plusieurs valeurs de l'humidité sont moyennes (centrées autour d'une valeur de 80 %, soit la moyenne de l'ensemble des observations d'humidité), comme visible à la Figure IV-16. La Figure IV-17 montre les hospitalisations correspondantes au jour même et jusqu'à 8 jours après les extrêmes. Aucun pic n'est visible et il semble même y avoir moins de pics que sans le paramètre d'orientation (Figure IV-12). De plus, les modèles explicatifs ne donnent pas de variables significatives. Plusieurs autres valeurs du paramètre d'orientation ont été considérées (en particulier pour les températures froides extrêmes), de même que d'autres fonctions de profondeur (Liu, Simplex et Tukey) et lags, mais sans obtention de résultats probants.

3.2.2. Extrêmes par indices météorologiques

La Figure IV-18 illustre les extrêmes définis selon l'humidex. Comme il peut être vu dans la construction de l'humidex (voir l'annexe B), ce dernier est principalement un indicateur de la température pondérée par l'humidité. Il accorde plus de poids à la température et ceci est visible à Figure IV-18. Les températures sélectionnées sont toutes extrêmes tandis que les valeurs d'humidité le sont moins souvent. De la même façon que pour les fonctions de profondeur, il

n'est pas nécessaire qu'une observation soit élevée dans ses deux variables de construction pour être considérée comme un extrême. Les hospitalisations correspondantes ne semblent pas nécessairement être des pics sanitaires, mais des pics saisonniers sont visibles (comme en été 2001 ou 2005).

La Figure IV-19 représente les extrêmes définis selon l'indice de chaleur. Étant donné sa construction plus complexe, les températures sont moins élevées que dans le cas de l'humidex, mais elles sont compensées par de plus fortes valeurs d'humidité. À nouveau, certains pics sanitaires sont associés aux extrêmes sélectionnés, mais il ne s'agit pas de la majorité des observations.

Les résultats des modèles avec les indices météorologiques sont présentés au Tableau IV-4. Trois modèles disposent d'une variable significative au jour même : un premier modèle avec les extrêmes d'indice de chaleur extraits par seuil (99 %), un deuxième avec les extrêmes d'indice de chaleur et d'humidex extraits par bloc (7 jours) et un troisième avec les extrêmes d'humidex extraits par bloc (7 jours). Ces deux derniers modèles peuvent subir un effet de taille à nouveau, étant donné les déviations expliquées faibles, les degrés de liberté élevés et les « grandes » tailles d'échantillon. En revanche, le premier modèle dispose d'une déviance expliquée de 13,2 % et d'une taille d'échantillon de 34. La Figure IV-20 montre les hospitalisations en fonction de l'indice de chaleur et de ses extrêmes. La droite de régression est croissante et il est à noter que les hospitalisations associées ne sont pas toutes des pics. Les résultats avec lags indiquent des relations linéaires et décroissantes (coefficients de régression tous négatifs). Plusieurs lags sont retenus (1, 4 et 7) et les déviations expliquées s'échelonnent entre 9 et 19 %.

4. Discussion

Les résultats n'ont pas montré d'impact significatif concernant les extrêmes des variations pour la température et l'humidité. Ceci contraste avec les études de Danet *et al.* (1999), Ebi *et al.* (2004) ou Wolf *et al.* (2009). Ces études portent sur la mortalité et la morbidité par MCV, à Lille (France), en Californie (États-Unis) et à Ausburg (Allemagne) respectivement. Toutefois, à la différence des modèles de la présente étude, elles incluent l'ensemble des observations, pas uniquement les extrêmes. Il s'agit d'ailleurs d'un point essentiel dans la comparaison avec d'autres études. Ce projet ne porte que sur un sous-ensemble des observations et en ce sens, il est difficile de comparer les résultats obtenus à d'autres résultats de la littérature classique.

Un cas particulier dans les modèles de régression est la variable de pollution de l'air (particules ultrafines ou $PM_{2,5}$). En effet, elle a un impact significatif le jour même qui devient au contraire non-significatif dans les jours suivants. Cela indique un effet immédiat des extrêmes de $PM_{2,5}$ sur les hospitalisations, concordant avec d'autres auteurs (p. ex. Dominici *et al.*, 2006; Bell *et al.*, 2009) qui trouvent eux aussi une relation le jour même entre les $PM_{2,5}$ et la morbidité par MCV. Des mécanismes immédiats pourraient impliquer l'asthme ou les infections respiratoires. Une fois encore, les auteurs utilisent toutes les observations. Le rapport de l'*American Heart Association* (Brook *et al.*, 2010) quant à lui pointe un effet des $PM_{2,5}$ qui peut aller de quelques heures à plusieurs mois. Par ailleurs, la revue de littérature de Pope et Dockery (2006) indique que les relations entre les $PM_{2,5}$ (et les PM_{10}) et la santé en général sont la plupart du temps linéaires ou quasiment linéaires. Il s'agit d'un point commun avec les relations observées dans la présente étude.

Chiu *et al.* (2016b) mènent une étude symétrique à la présente, c'est-à-dire la sélection et l'étude des pics sanitaires qui sont les valeurs extrêmes de variables sanitaires (la variable dépendante

dans la régression). Ils utilisent les outils issus de la théorie des valeurs extrêmes. Leurs résultats vont dans le sens de ceux produits ici, à savoir que les conditions météorologiques correspondantes aux pics sanitaires ne sont pas nécessairement extrêmes. Pour le visualiser, les pics d'hospitalisations obtenus à Montréal dans Chiu *et al.* (2016b) sont présentés à la Figure IV-21, mais dans l'espace de la température et de l'humidité. Hormis quelques exceptions (comme aux lags 3 et 4 lors de températures froides et d'humidité basse), les conditions météorologiques ne sont pas extrêmes, même jusqu'à une semaine avant l'apparition du pic. Ainsi, cela se répercute dans ce chapitre car de façon générale, les extrêmes échantillonnés n'ont pas donné lieu à des pics sanitaires (au jour même et jusqu'à une semaine après leur apparition). En fait, seuls quelques pics sanitaires sont observés (dont le plus élevé de la série des hospitalisations en particulier), mais il ne s'agit pas de la règle générale. Certains auteurs indiquent que les extrêmes les plus élevés n'entraînent effectivement pas nécessairement de pics, tels que Smoyer-Tomic *et al.* (2003) ou Beniston et Stephenson (2004). En contrepartie, il y a très peu de pics sanitaires qui surviennent avec des conditions météorologiques « ordinaires » (telles que définies par la fonction de profondeur, c'est-à-dire une humidité de 80 % et une température de 10 °C). Par ailleurs, les modèles classiques que Chiu *et al.* (2016b) ont établi pour comparaison (températures, humidité et hospitalisations) donnent des déviations expliquées inférieures à 5 % avec toutes les observations. Les modèles avec lags de l'étude présente ont ainsi des déviations expliquées largement supérieures aux modèles classiques (de 2 à 27 %).

Un point important est la forme des relations obtenues entre les extrêmes et les hospitalisations. Tandis que Chiu *et al.* (2016b) trouvent principalement des effets non-linéaires, les modèles de la présente étude donnent surtout des relations linéaires comme ceux de Morabito *et al.* (2005) mais pour l'étude des hospitalisations pour infarctus du myocarde et les températures extrêmes. Pour

Morabito *et al.* (2005), les variables explicatives sont les classes de jours durant lesquelles la température a dépassé un certain seuil. De plus, ces auteurs ont utilisé un niveau de détail avancé, en étudiant des sous-groupes (séparation selon l'âge et le sexe et observations horaires). Enfin, il est reconnu que la forme de la relation entre la mortalité ou la morbidité toutes causes et la météo est généralement en forme de « U » ou de « V » (Martens, 1998; Gouveia *et al.*, 2003; Doyon *et al.*, 2006). Ce n'est pas nécessairement le cas en considérant uniquement les MCV, mais les relations sont tout au moins non-linéaires (p. ex. Bhaskaran *et al.*, 2010). En étendant la régression aux interactions bivariées, des plans de régression sont obtenus dans ce chapitre (l'équivalent bivarié d'une droite de régression). Il s'agit encore d'une différence avec le contexte classique puisque Schwartz (1997) ou Braga *et al.* (2002) obtiennent des surfaces de régression (l'équivalent multivarié d'une fonction non-linéaire) en examinant la relation entre la météo et les MCV avec toutes les observations.

À notre connaissance, les fonctions de profondeur n'ont jamais été utilisées dans un contexte de santé publique ou de météorologie. Pourtant, elles permettent d'établir un ordre dans un échantillon multivarié et ainsi de définir des extrêmes. Leurs propriétés mathématiques en font des candidates idéales pour la recherche de valeurs extrêmes dans le cas multivarié (Dang et Serfling, 2010; Chebana et Ouarda, 2011b), mais cette pratique reste encore peu répandue de façon générale. De plus, l'insertion dans un modèle de régression des extrêmes ainsi identifiés n'est proposée par aucune étude, que ce soit dans un cadre purement statistique, de santé publique ou de météorologie. Une méthodologie claire manque donc encore à cet égard. Les résultats obtenus dans ce chapitre indiquent qu'une méthodologie incorporant les fonctions de profondeur permettrait de définir des extrêmes de façon efficace et objective d'un point de vue statistique. Par ailleurs, en disposant d'une connaissance préalable des processus physiologiques qui entrent

en jeu lors des réponses du corps humain aux différentes conditions météorologiques et en contrôlant la plage du paramètre d'orientation, il est possible de définir une zone restreinte où surviennent les extrêmes d'intérêt. Ces derniers sont ainsi choisis grâce à une connaissance du domaine plutôt que d'un point de vue purement statistique. Cette approche a été utilisée par Chebana et Ouarda (2011b) qui ont sélectionné un type d'événement météorologique en particulier. Ainsi, certaines maladies plus affectées par la chaleur ou par le froid peuvent être sélectionnées en modifiant cette plage, plutôt que de faire une sélection globale.

La principale difficulté méthodologique dans la présente étude repose sur la définition d'un extrême, en particulier lorsqu'il est multivarié. Comme l'appuient Beniston et Stephenson (2004) et Gosling *et al.* (2009), une définition commune est nécessaire, mais demeure très difficile à établir en pratique. Ceci est le cas pour diverses notions statistiques, dont la médiane multivariée et les quantiles multivariés. En particulier, ces auteurs conseillent des approches basées sur les quantiles comme dans la présente étude ainsi que celle de Chiu *et al.* (2015). En effet, les définitions qui reposent sur les valeurs brutes sont susceptibles de varier fortement selon la zone géographique (Ye *et al.*, 2012), tandis que les quantiles tiennent compte de la variabilité de chaque région. Il semble néanmoins que même l'utilisation d'un quantile en particulier (par exemple 90 % ou 95 %) ne soit jamais réellement justifiée, si bien que ce choix repose sur les auteurs. Dans cette étude, les approches ayant donné les meilleurs résultats en termes de déviations expliquées et de significativité statistique sont justement les approches basées sur les quantiles des variables météorologiques.

La littérature concernant les EME est principalement divisée en deux types d'études. Dans le premier type, les études portent sur un extrême en particulier (un événement ponctuel et pas une série) et non sélectionné, la plupart du temps, sur une base statistique. Dans le deuxième type, les

études considèrent toutes les observations météorologiques et établissent des seuils élevés pour définir un EME.

Les premières se basent habituellement sur les répercussions qu'a eues cet extrême ponctuel. Si ces répercussions sont suffisamment importantes, alors l'extrême en question sera analysé de façon plus importante par la suite. Un exemple récent de ce type d'analyse ponctuelle est la canicule européenne de 2003. Une surmortalité importante a mené à plusieurs études a posteriori (p. ex. Hémon et Jouglu, 2003; Vandentorren *et al.*, 2004; Bouchama *et al.*, 2007), en revanche aucune n'a tenté d'intégrer cette canicule dans un cadre général de modélisation. Les intérêts à étudier un événement en particulier sont multiples, par exemple en termes d'adaptation. Fouillet *et al.* (2008) indiquent que l'étude exhaustive de la canicule de 2003 a davantage bonifié la préparation du système de la santé (meilleures mesures préventives, réactivité du système de la santé, etc.) pour la canicule de 2006.

D'un autre côté, le deuxième type d'études, basées sur des méthodes statistiques pour définir les EME (p. ex. Li *et al.*, 2012; Lavigne *et al.*, 2014), permet d'obtenir une vue plus globale des EME. En effet, le fait qu'un phénomène météorologique n'entraîne pas de pics sanitaires ne signifie pas pour autant pas qu'il ne représente pas un EME. Par définition, les EME sont des phénomènes rares et pour mener une analyse globale, il convient de tous les étudier, plutôt que de procéder à une sélection ponctuelle. La présente étude s'inscrit dans cette optique, mais contrairement aux études qui ont déjà été réalisées, elle s'intéresse exclusivement aux EME. Cela a pour effet de ne pas diluer l'information dans l'ensemble des observations tout en essayant d'expliquer les conséquences sur la santé d'un point de vue général. Une modélisation qui n'est pas focalisée sur les extrêmes apporte une compréhension complémentaire aux études

ponctuelles, tout en fournissant des outils différents quant à la surveillance en santé publique (en particulier d'un point de vue des alertes sanitaires).

La médiatisation de certains événements météorologiques joue aussi parfois un rôle dans la classification de ces derniers comme « extrêmes » ou non. Il est possible de citer FitzGerald *et al.* (2010) qui analysent une base de données des inondations en Australie, les inondations étant la catastrophe naturelle qui cause le plus de décès sur ce continent. Les auteurs sélectionnent les extrêmes à retenir à partir d'études scientifiques et de la presse écrite. Ashley et Ashley (2008) élargissent cette approche à tous les médias, pour l'étude des inondations aux États-Unis. De façon plus générale, Knowlton *et al.* (2011) étudient six événements extrêmes reliés aux changements climatiques. Un extrême météorologique remarquable est sélectionné dans chaque catégorie (pollution de l'air, canicule, tornade, épidémie du virus du Nil, inondation et feu de forêt) sur les 20 dernières années. Cette sélection se fait sur la base du poids socioéconomique des extrêmes, comme les décès et hospitalisations et les coûts associés pour le gouvernement (voir aussi Greenough *et al.*, 2001). Chaque événement sélectionné a d'abord été rapporté par les médias, avant de faire l'objet de plusieurs études scientifiques. Cela est également parfois le cas au Québec. Ainsi, dans les résultats décrits dans ce chapitre, la canicule de l'été 2003 au Québec est incluse. Les méthodes qui y sont appliquées ont permis de sélectionner les températures extrêmes de cette période, laquelle a aussi été examinée spécifiquement par Smoyer-Tomic *et al.* (2003). Étant donné que la présente étude utilise une approche plus globale, l'échantillon d'extrêmes est plus grand et permet une modélisation plus générale. Il est justifié d'investiguer les causes et les conséquences d'un EME en particulier mais il est aussi nécessaire d'établir des définitions plus générales afin d'avoir un portrait global. Par ailleurs, un EME ne donne pas nécessairement lieu à un pic sanitaire, tel qu'il a été vu dans les résultats de chapitre. C'est aussi

le cas entre autres pour la canicule d'août 2010 à Québec, qui n'a donné lieu à aucun excès de mortalité (Bustinza *et al.*, 2013) par rapport à la même période de 2005 à 2009.

En matière de confondance, une variable de pollution atmosphérique (PM_{2,5}) a été ajoutée aux modèles. De plus, une variable d'ensoleillement a aussi été considérée, car il est connu que l'exposition au soleil a un rôle important dans la création de vitamine D, qui en retour influence la santé cardiovasculaire (Holick, 2004). Cette dernière variable météorologique n'a toutefois pas pu être incluse dans les modèles en raison d'un trop grand nombre de valeurs manquantes. Concernant le contrôle de la confondance, il est possible de le faire au niveau du design de l'étude et lors de l'analyse à proprement dite (Tilaki, 2012). La structure des données (données agrégées) n'a pas permis l'utilisation de méthodes telles que la randomisation ou la stratification lors du design. Toutefois, la majorité des modèles utilisés ici sont des modèles de régression multivariés, qui permettent de contrôler une partie de la confondance dans l'analyse (Tilaki, 2012). Ces modèles semblent plus appropriés pour contrôler la confondance que la stratification par exemple (Pourhoseingholi *et al.*, 2012; McNamee, 2005). Ils peuvent demander en revanche une connaissance poussée du domaine d'étude concernant les variables confondantes à inclure.

5. Conclusion

Cette étude a deux buts principaux, qui tentent de répondre aux deux problématiques principales dans l'étude des extrêmes météorologiques. La première problématique porte sur la définition d'un « extrême ». Dans ce chapitre, plusieurs définitions ont été considérées. Elles ont été proposées sur une base statistique (théorie des valeurs extrêmes, fonctions de profondeur, indices météorologiques) afin d'obtenir des résultats les plus objectifs et globaux possibles. Jusqu'à présent, aucune étude, que ce soit en littérature météorologique ou de santé publique, n'a tenté d'extraire plusieurs extrêmes météorologiques de façon objective. Les modèles basés sur les

quantiles ainsi que ceux faisant appel aux fonctions de profondeur ont montré de meilleurs résultats en termes de sélection d'EME. Il est probable que les EME fassent intervenir de fait plusieurs variables météorologiques, impliquant un contexte multivarié. Les fonctions de profondeur sont donc appropriées pour leur étude.

Une fois l'étape d'identification effectuée, la deuxième problématique porte sur l'évaluation de l'effet des EME sur la santé. Les modèles explicatifs n'ont pas montré d'association systématique entre un EME et un pic sanitaire. En revanche, beaucoup d'études sur les relations entre la santé et la météo obtiennent des déviations expliquées relativement faibles (moins de 5 %, Lipfert, 1993), du fait d'un échantillon qui comprend toutes les observations et qui dilue les relations. Ici, les déviations expliquées sont relativement élevées, atteignant jusqu'à 27 % pour les modèles significatifs. De plus, contrairement au cas classique, les relations observées sont majoritairement linéaires dans le cas des EME.

6. Limitations

Concernant les variables explicatives, des variables météorologiques et de pollution de l'air ont été utilisées dans cette étude pour se conformer aux résultats connus de la littérature en ce qui concerne la santé par MCV. Une variable non disponible est la vitesse du vent. Il a été montré qu'elle influe sur les MCV (Carder *et al.*, 2005; Morabito *et al.*, 2005; Urban et Kysely, 2014), en particulier lorsqu'elle est intégrée dans les indices météorologiques. De plus, Dominici *et al.* (2000) utilisent d'autres variables de pollution de l'air que les particules ultrafines, comme l'ozone. Les auteurs trouvent des relations significatives entre ces ajouts et les MCV, mais l'approche de modélisation est différente (modèles hiérarchiques bayésiens) et l'intérêt ne repose pas sur les extrêmes. Par ailleurs, une façon de bonifier l'étude présente (sans être nécessairement une limite) serait de se baser sur le coût engendré par les événements pour les qualifier

d'extrêmes. Certains auteurs utilisent cette approche, par exemple pour les tornades ou les inondations (Greenough *et al.*, 2001; Knowlton *et al.*, 2011). Il n'est pas possible d'utiliser cette approche ici, car les données de coût correspondantes ne sont pas disponibles dans les fichiers administratifs. Toutefois, si elles l'étaient, cela pourrait apporter des informations utiles pour le système de la santé. En effet, les estimations financières ont un impact concret pour la planification. Il arrive parfois que l'approche par coût ou l'approche par définition statistique soient toutefois équivalentes (Morss *et al.*, 2011).

Concernant les méthodes d'extraction des extrêmes, la méthode du seuil ne permet pas de prendre en compte les variations saisonnières. Cela résulte du fait que les seuils soient fixes, ce qui a l'avantage de ne considérer que l'été ou l'hiver. En revanche, la méthode des blocs permet d'intégrer les différentes saisons dans les extrêmes, étant donné que la séparation est cette fois-ci verticale. Cela peut être intéressant, car un extrême météorologique durant l'automne, par exemple, n'est pas nécessairement un extrême lorsque toute l'année est considérée, mais peut tout de même avoir des répercussions sur la santé par MCV. Cependant, en considérant l'ensemble des observations, il peut résulter des observations qui ne sont pas nécessairement des extrêmes, en particulier avec une taille des blocs plus courte que le cycle. Considérer une séparation en saisons comme dans l'étude de Chiu *et al.* (2015) pour ensuite y établir des blocs pourrait donc s'avérer pertinent dans de futurs travaux.

Tableaux et figures

Tableaux

Tableau IV-1 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques univariés extraits par seuil. Lorsqu'une droite de régression est estimée, le coefficient de régression est renseigné, tandis que dans le cas d'une fonction non-linéaire il s'agit du degré de liberté (ddl).

Variable	Méthode Lag	Déviance expliquée (%)	Coefficient ou ddl	n
Température	Seuil (99 %) 0	14	15	
	Seuil (99 %) 3	13,3	-12,31	
	Seuil (99 %) 4	21,6	-15,83	34
	Seuil (99 %) 5	13,1	13,32	
	Seuil (99 %) 6	24,4	18,51	
	Bloc (7 jours) 0	2,21	2,12 (ddl)	478
	Bloc (60 jours) 4	11,2	-1,08	55
Humidité	Seuil (95 %) 0	6,2	2,11	153
PM	Seuil (99 %) 7	13	-0,01	34
	Bloc (60 jours) 5	8,93	0,73	56

Tableau IV-2 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques univariés (variations).

Variable	Méthode Lag	Déviance expliquée (%)	Coefficient	n
Humidité	Bloc (7 jours) 0	1,94	1	478

Tableau IV-3 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques multivariés extraits par fonctions de profondeur.

Fonction	Méthode Lag	Déviance expliquée (%)	Degré de liberté	n
Température Humidité	Seuil (10 %) 0	2,59	2	335
	Seuil (1 %) 6	27,6	2	34
	Seuil (1 %) 7	27,2	2,32	
Température PM	1 % 0	18,9	2	34

Tableau IV-4 : Résultats de modélisation pour les hospitalisations à Montréal en fonction de différents extrêmes météorologiques multivariés (indices météorologiques). Lorsqu'une droite de régression est estimée, le coefficient de régression est renseigné, tandis que dans le cas d'une fonction non-linéaire il s'agit du degré de liberté (ddl).

Variable	Méthode Lag	Déviance expliquée (%)	Coefficient ou ddl	n
Humidex	Bloc (7 jours) 0	3,59	2,32 (ddl)	478
	Bloc (60 jours) 4	12	-0,6	56
	Seuil (99 %) 0	13,6	4,16	34
	Seuil (99 %) 4	11,6	-1,94	34
Indice de chaleur	Bloc (7 jours) 0	3,72	2,86 (ddl)	478
	Bloc (60 jours) 1	16,5	-0,75	56
	Bloc (60 jours) 4	19,6	-0,8	
	Bloc (60 jours) 7	9,21	-0,58	34
	Seuil (99 %) 0	13,2	2,22	
	Seuil (99 %) 4	11,6	-1,94	34

Figures

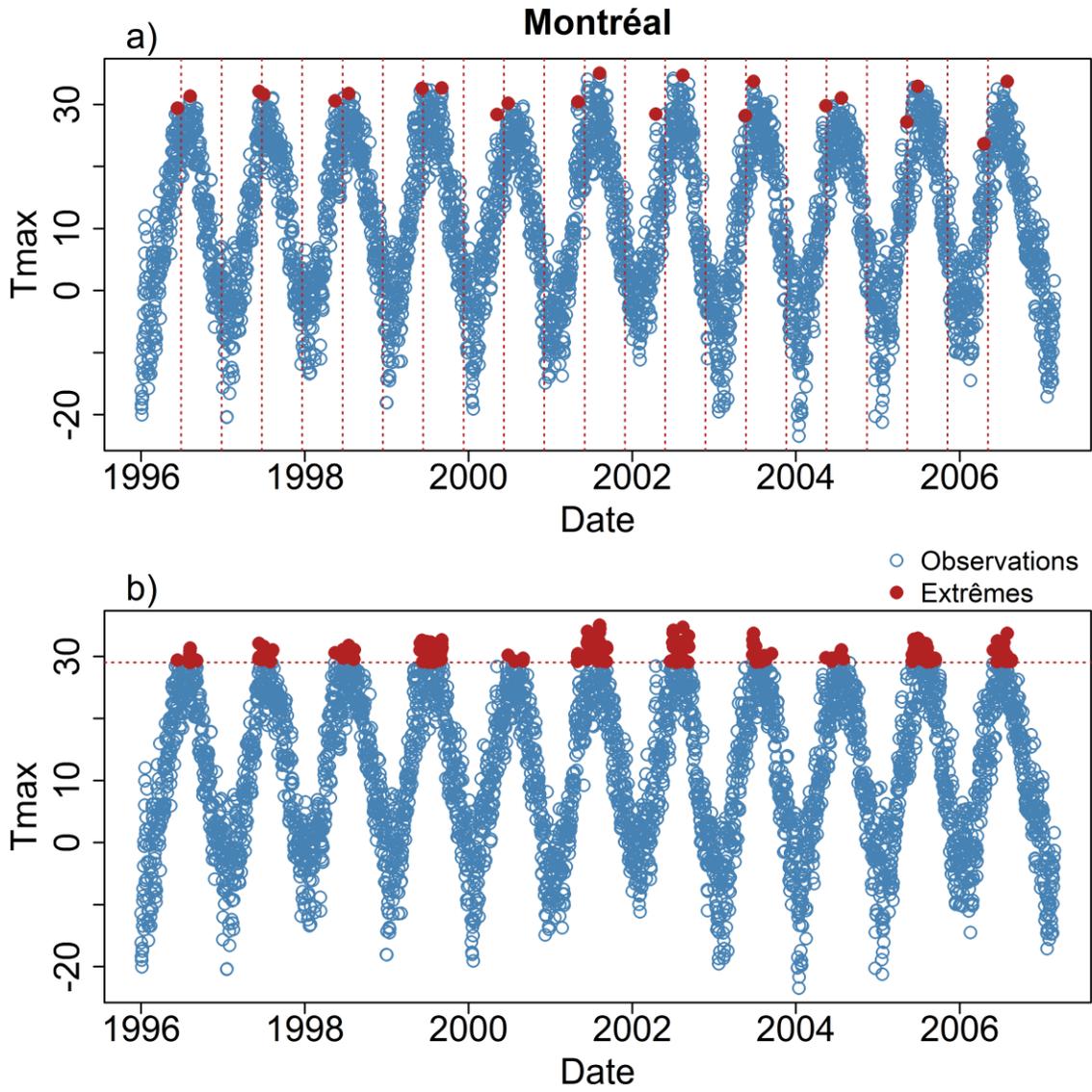


Figure IV-1: Exemples d'extrêmes de température a) par bloc et b) par seuil, à Montréal. Les blocs et le seuil sont représentés en pointillés.

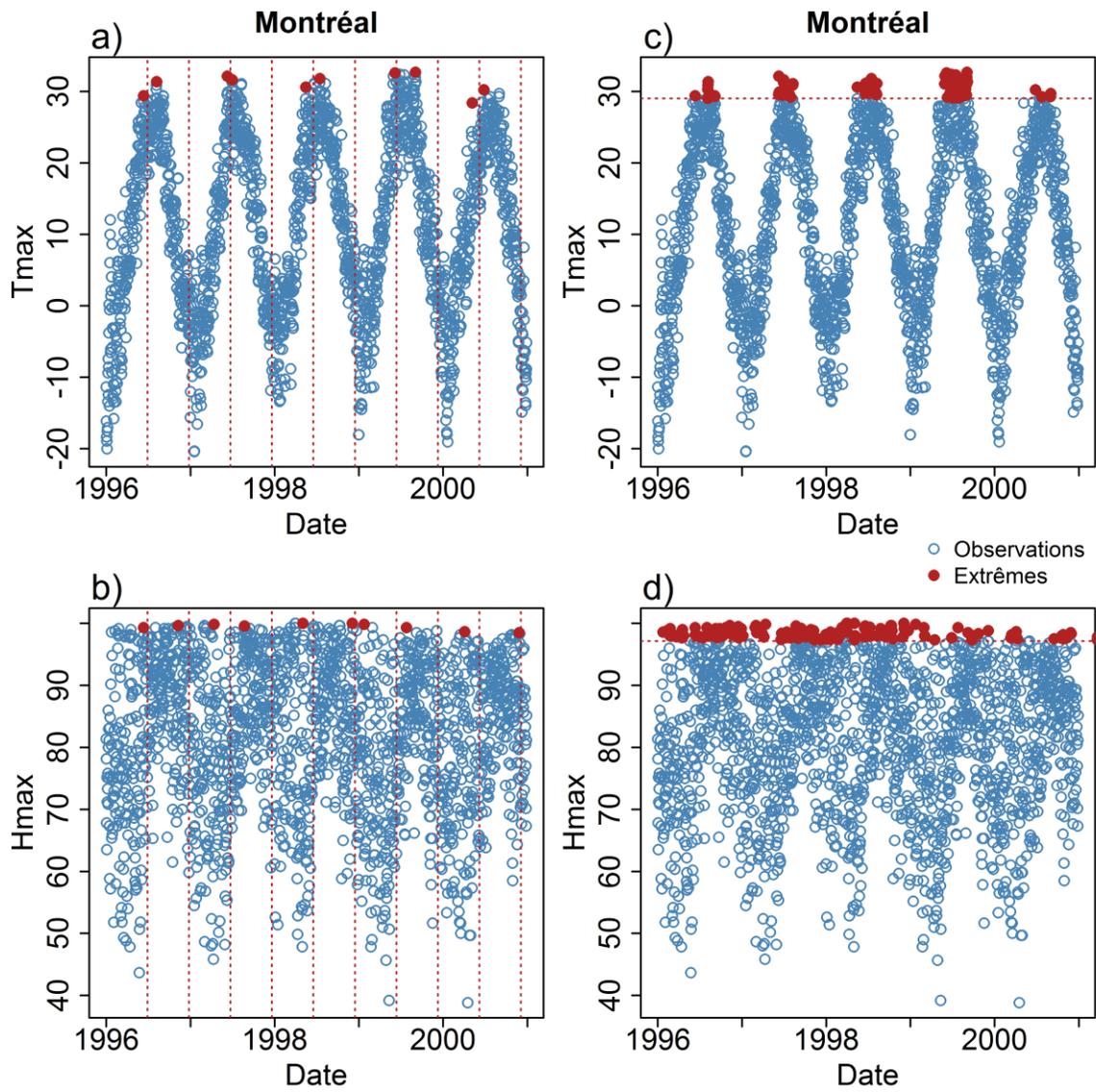


Figure IV-2 : a) Extrêmes de température par bloc, b) extrêmes d'humidité par bloc, c) extrêmes de température par seuil et d) extrêmes d'humidité par seuil, à Montréal (période de 1996 à 2000 inclusivement).

Montréal

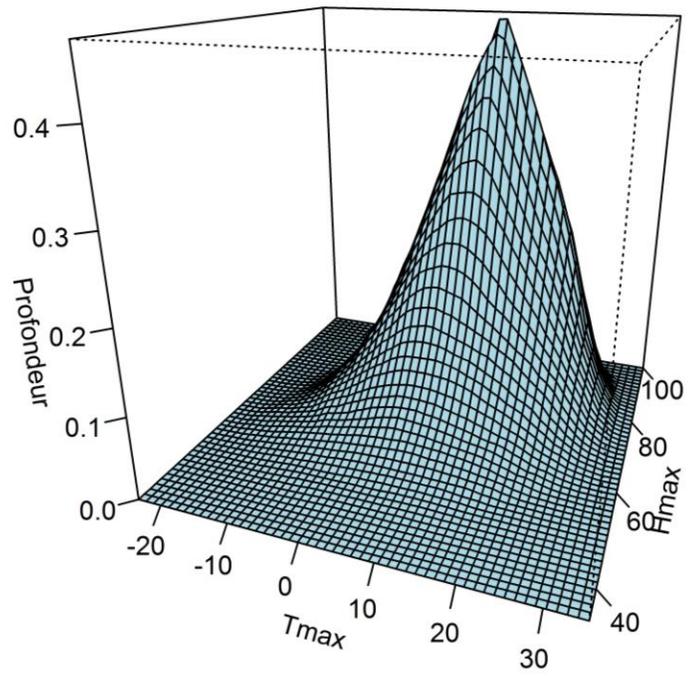


Figure IV-3 : Exemple de fonction de profondeur entre la température et l'humidité à Montréal.

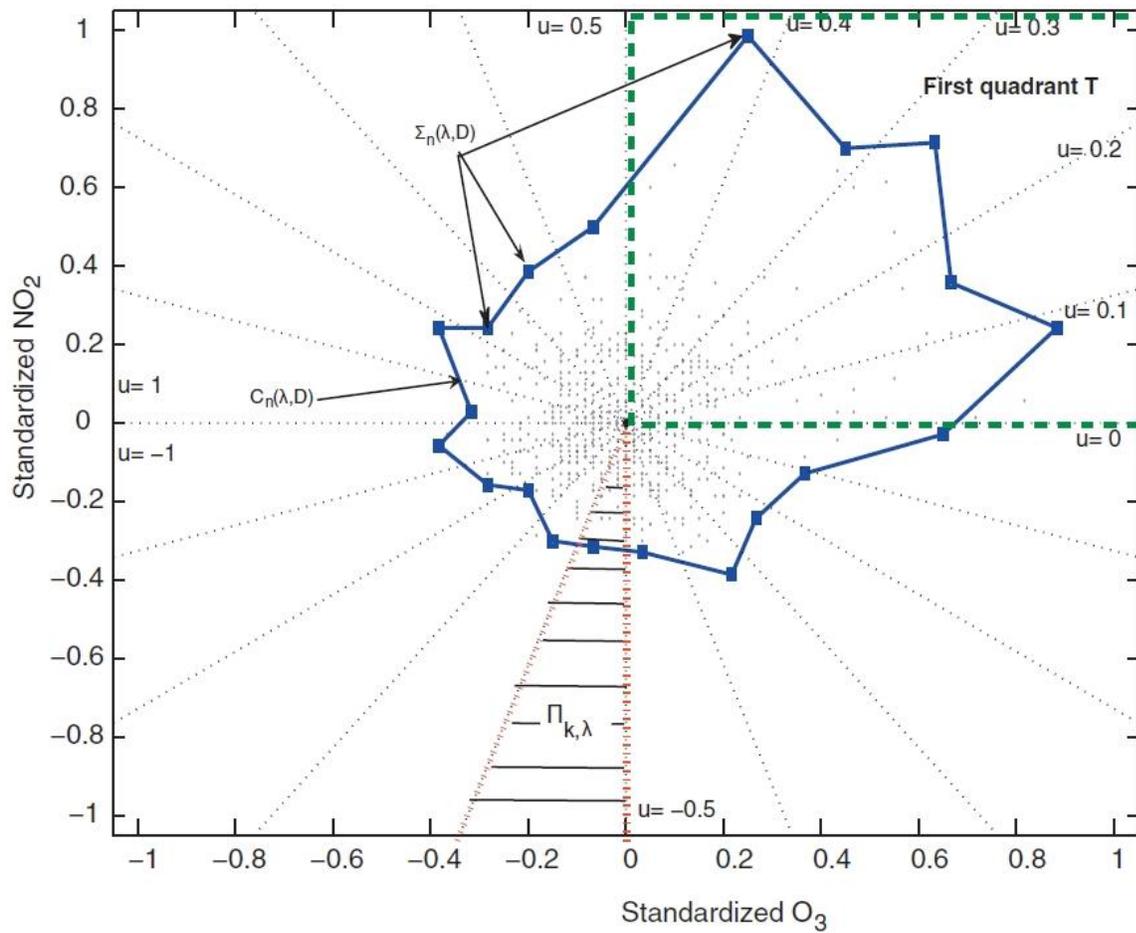


Figure IV-4 : Figure illustrant l'approche de Chebana et Ouarda (2011b). Les observations sont en gris tandis que les extrêmes sélectionnés sont en bleu. Le paramètre u contrôle l'orientation selon laquelle ces derniers sont choisis.

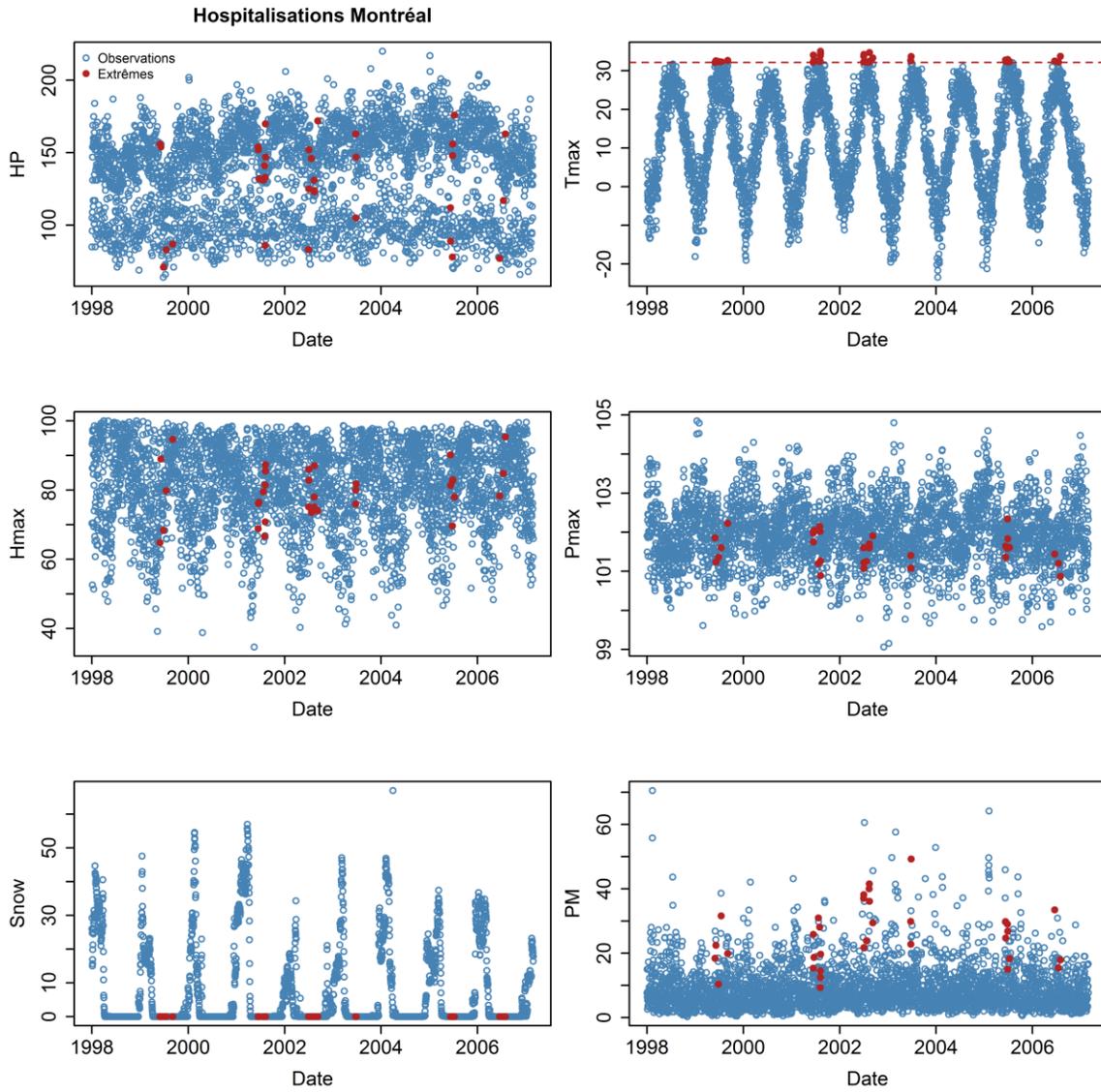


Figure IV-5 : Extrêmes météorologiques univariés de température (seuil à 99 %, en pointillés rouges) pour les hospitalisations à Montréal.

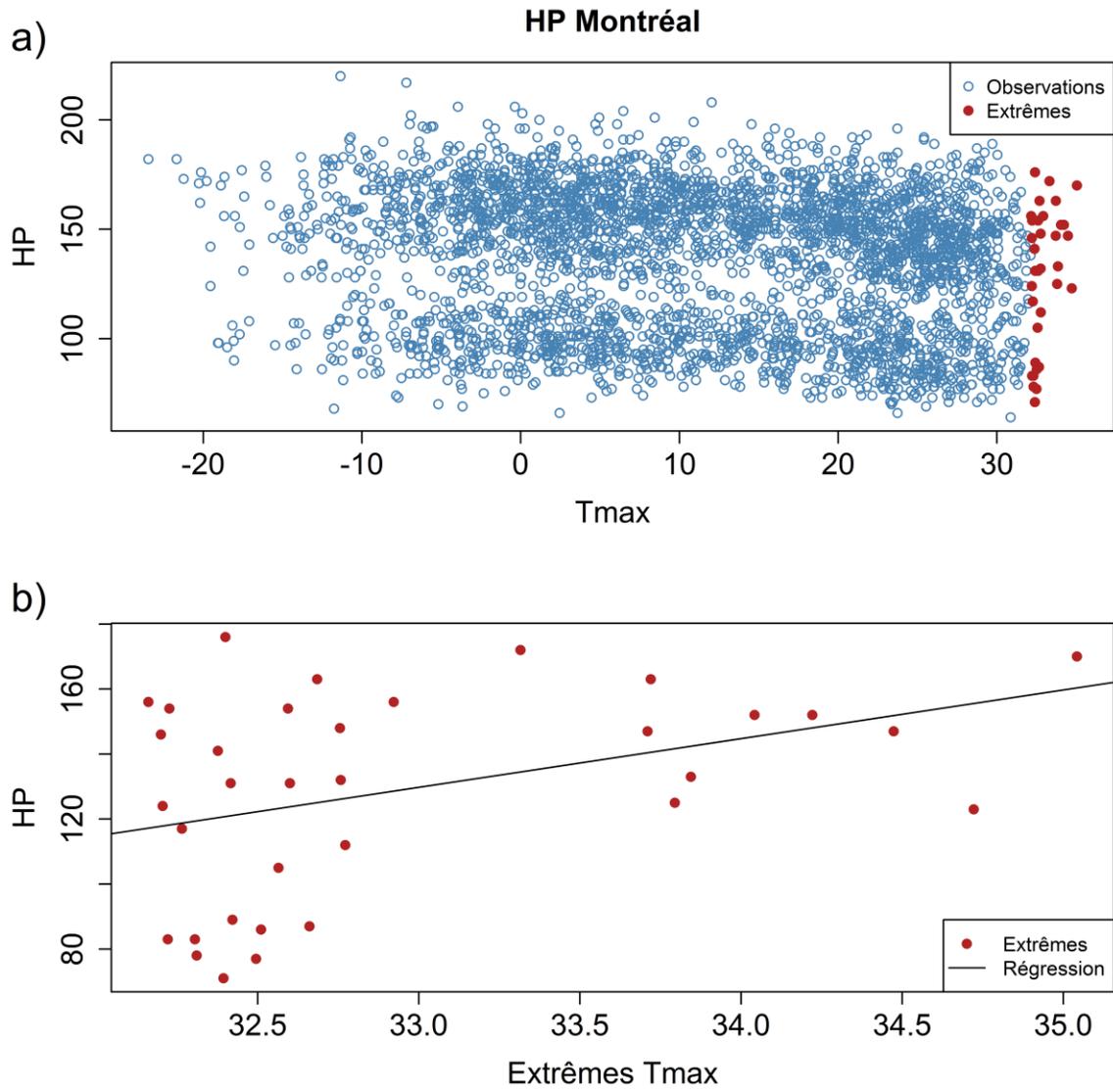


Figure IV-6 : a) Hospitalisations et température à Montréal et b) Droite de régression des extrêmes de température par seuil sur les hospitalisations.

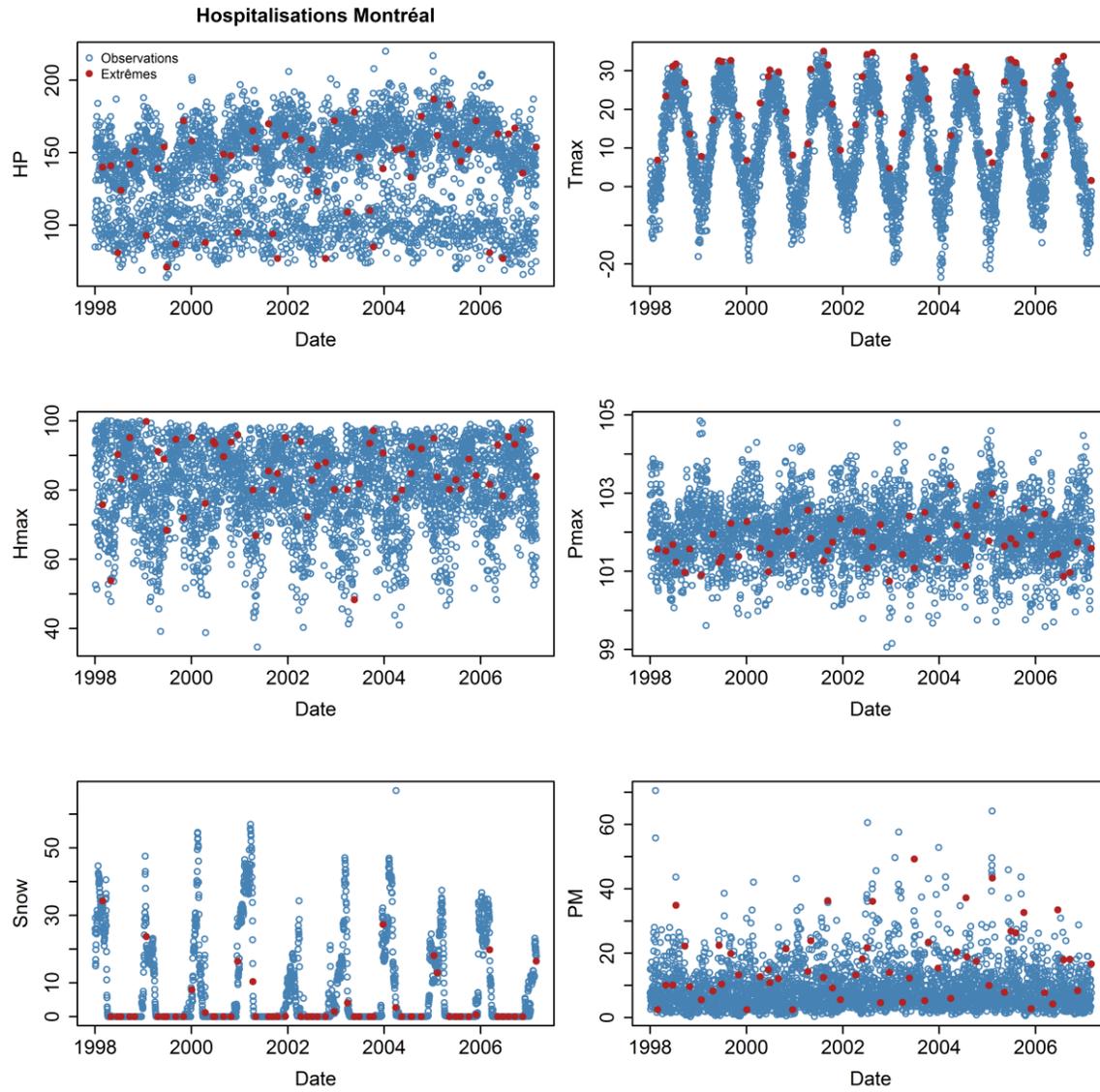


Figure IV-7 : Extrêmes météorologiques univariés de température (blocs de 60 jours) pour les hospitalisations à Montréal.

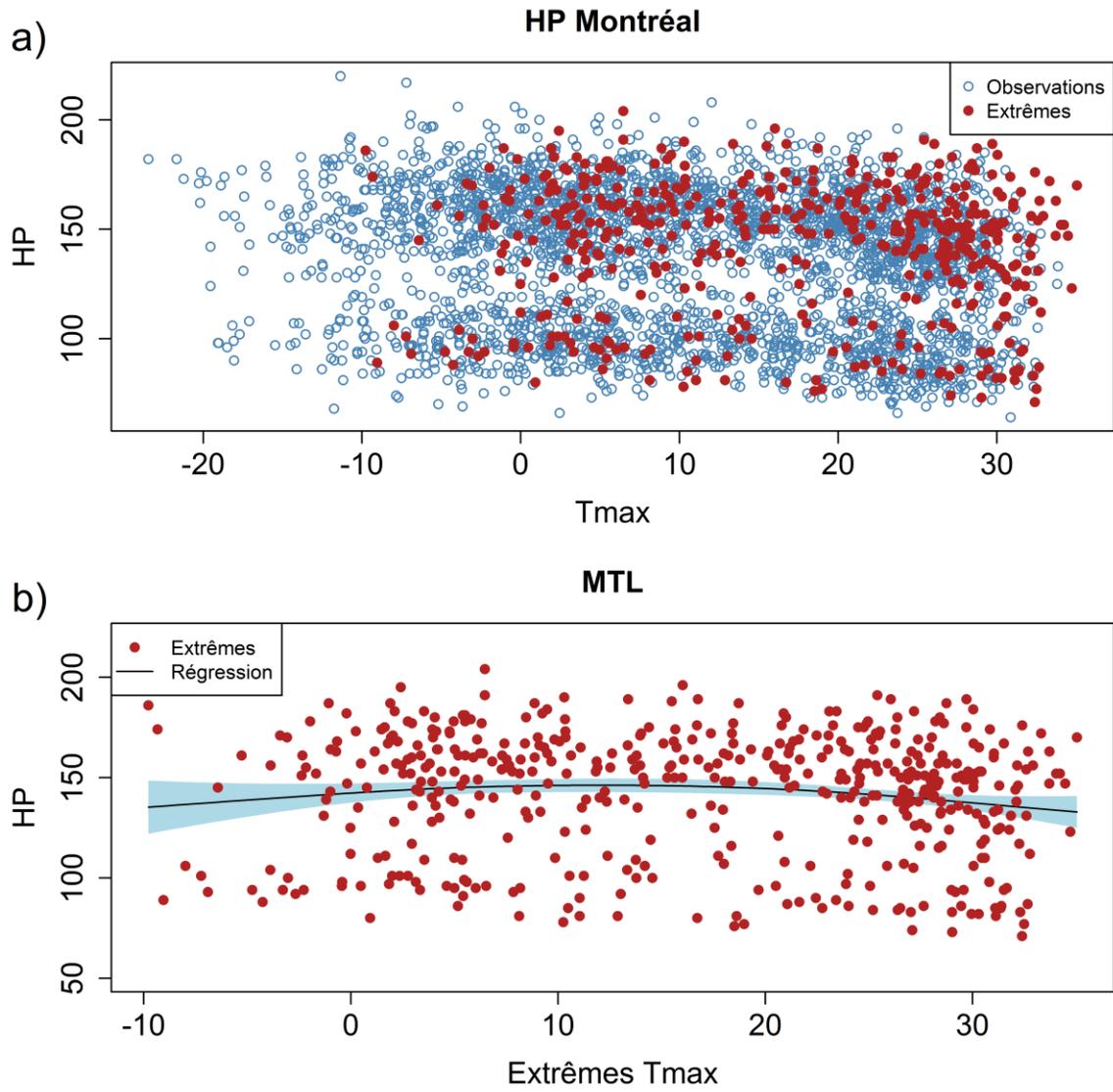


Figure IV-8 : a) Hospitalisations et température à Montréal et b) Fonction de régression des extrêmes de température par blocs sur les hospitalisations.

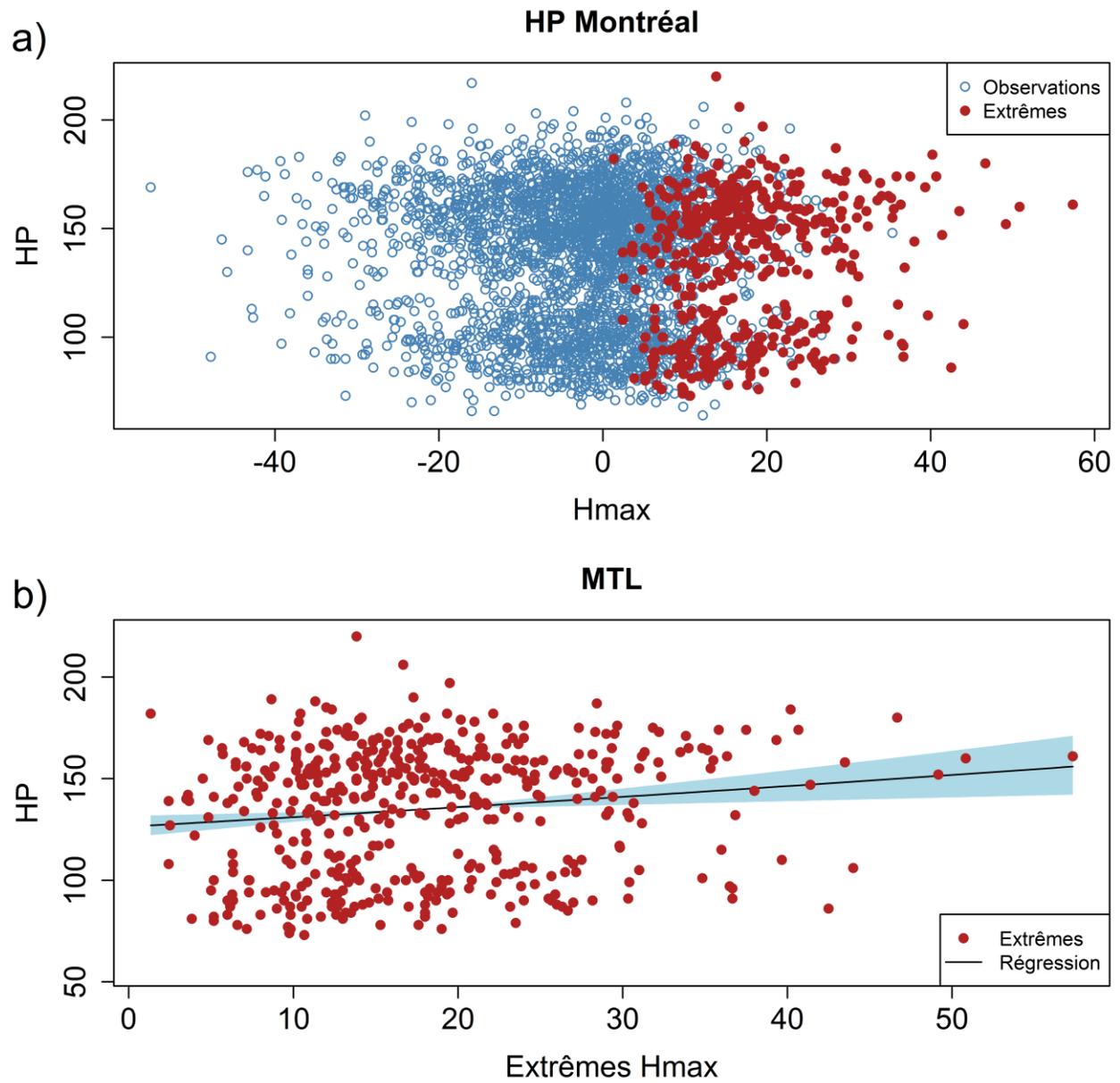


Figure IV-9 : a) Hospitalisations et variations de l'humidité à Montréal et b) Droite de régression des extrêmes des variations de l'humidité sur les hospitalisations.

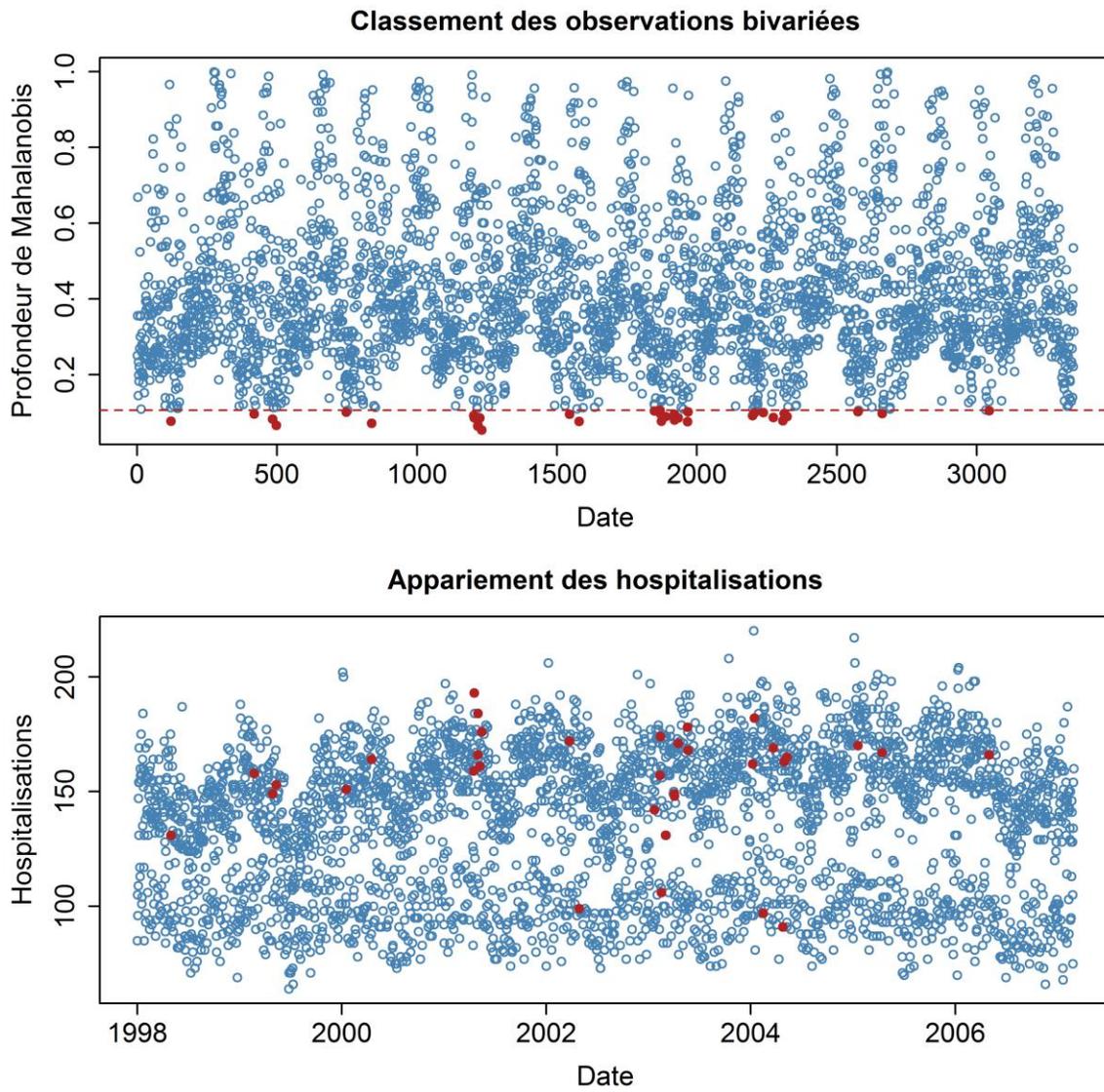


Figure IV-10 : Profondeur de la température et de l'humidité (seuil à 1 %) à Montréal.

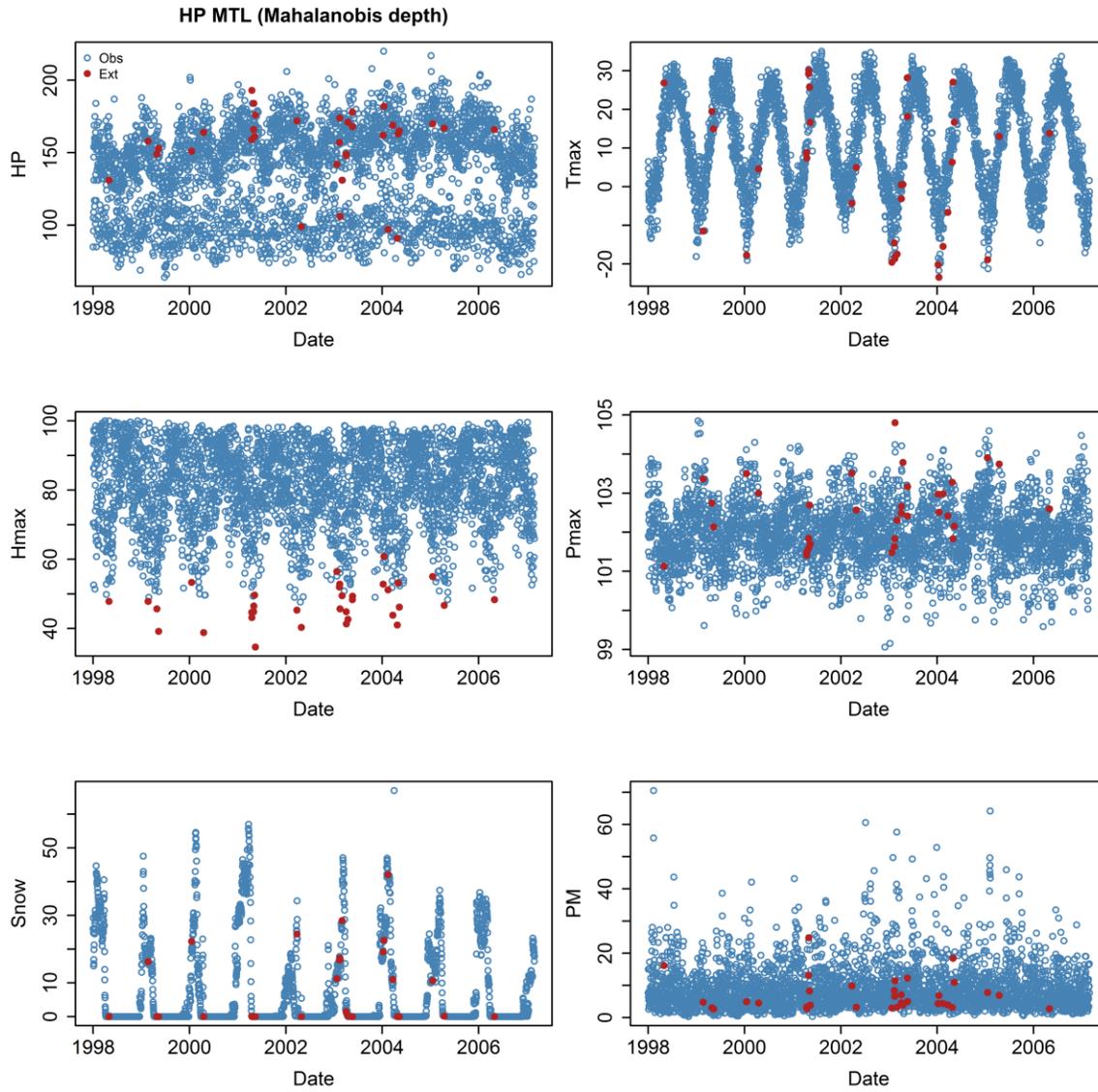


Figure IV-11 : Extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %) pour les hospitalisations à Montréal.

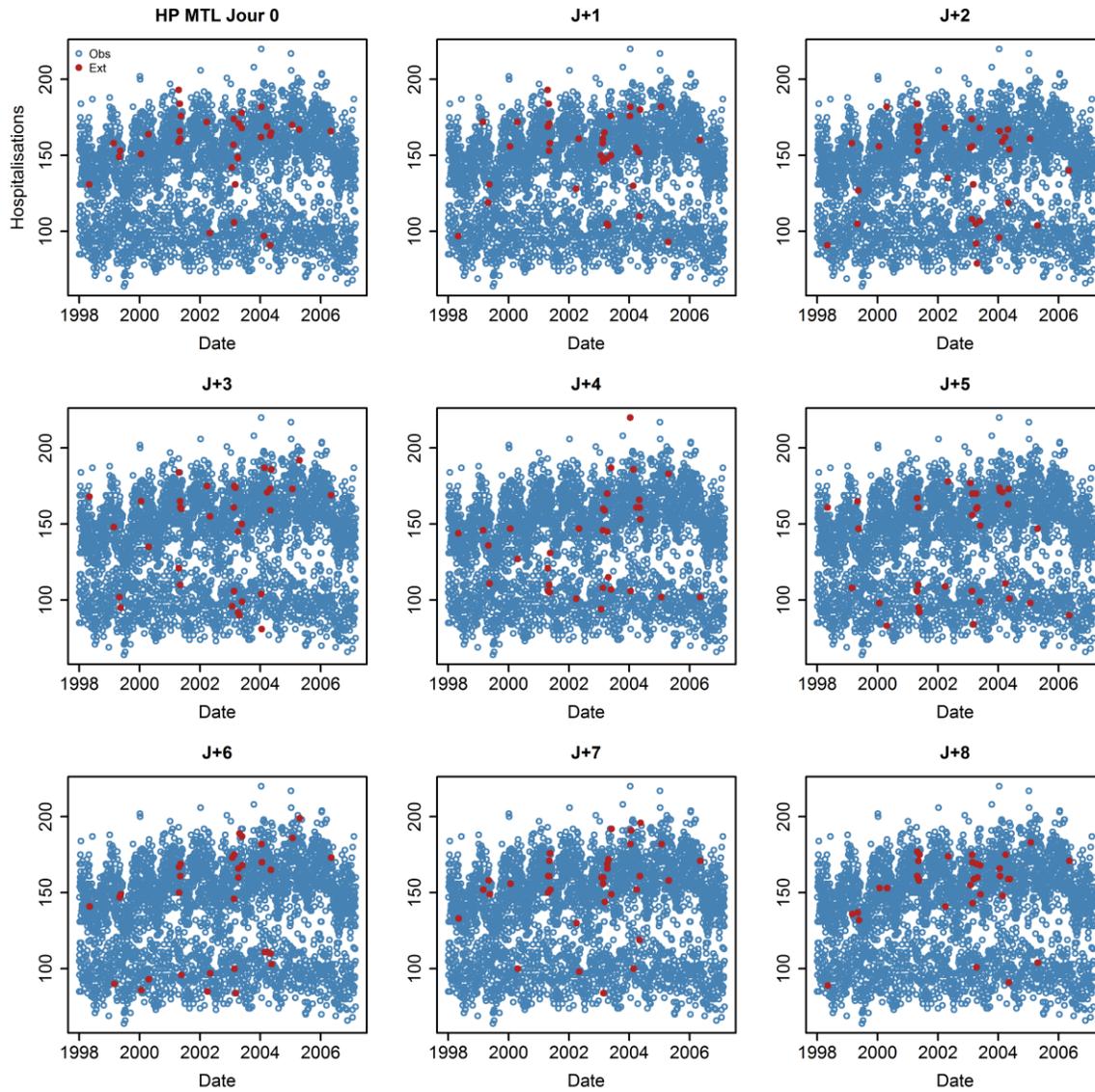


Figure IV-12 : Association des hospitalisations correspondantes aux extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %). Les hospitalisations vont du jour même à 8 jours après les extrêmes.

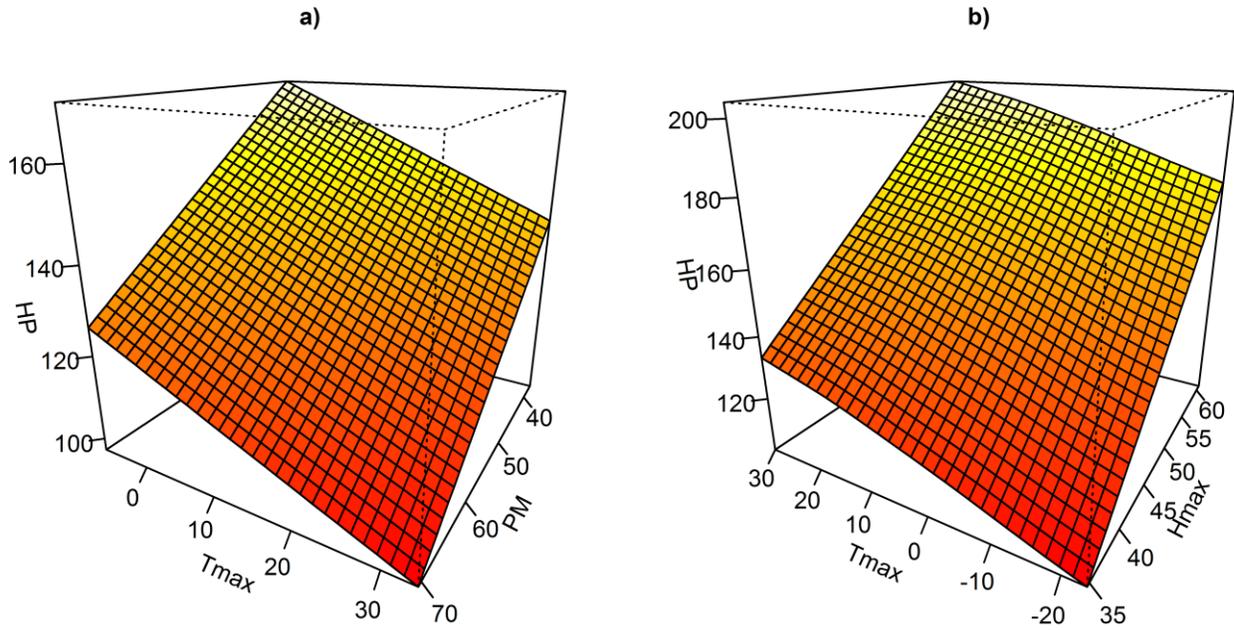


Figure IV-13 : Plans de régression des extrêmes multivariés extraits par fonction de profondeur sur les hospitalisations à Montréal. a) Tmax et PM et b) Tmax et Hmax.

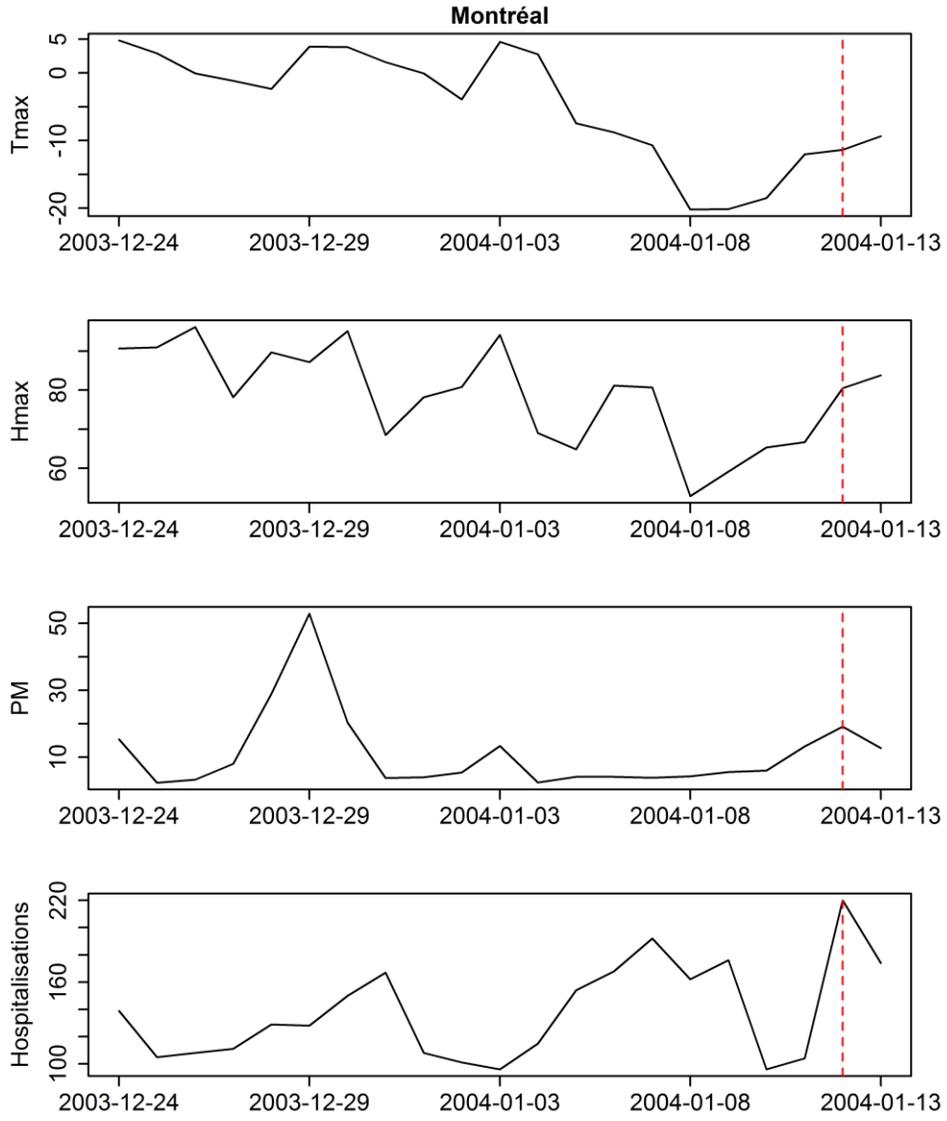


Figure IV-14: Résumé des conditions météorologiques survenues avant le pic du 12 janvier 2004 à Montréal. Les pointillés rouges représentent le jour du pic de 220 hospitalisations.

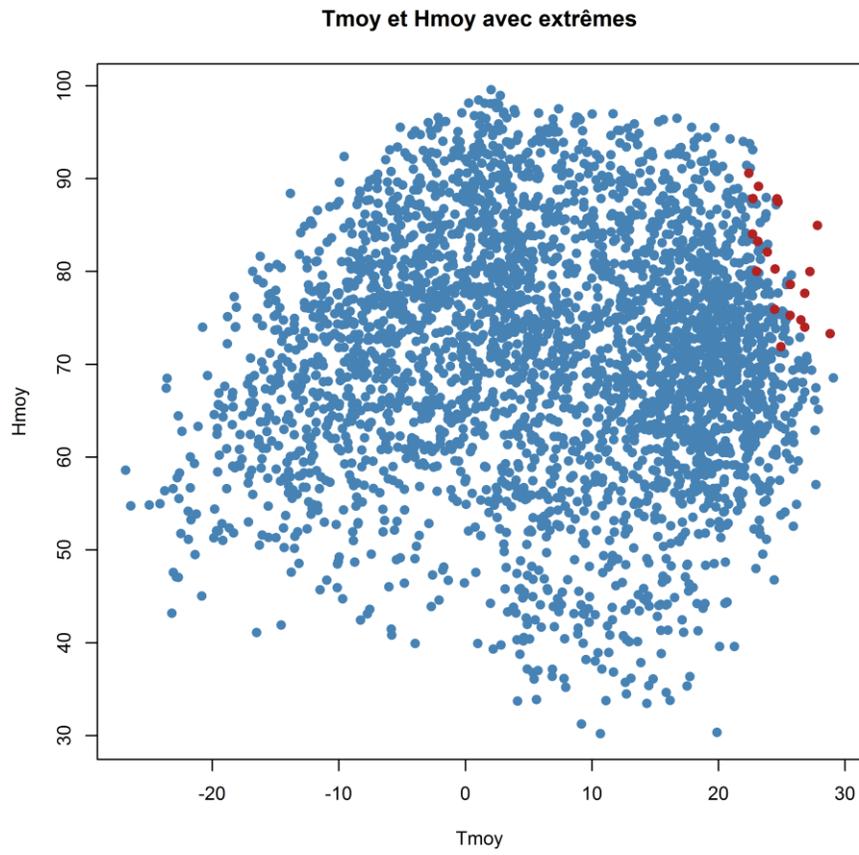


Figure IV-15 : Sélection des extrêmes météorologiques avec la méthodologie de Chebana et Ouarda (2011b).

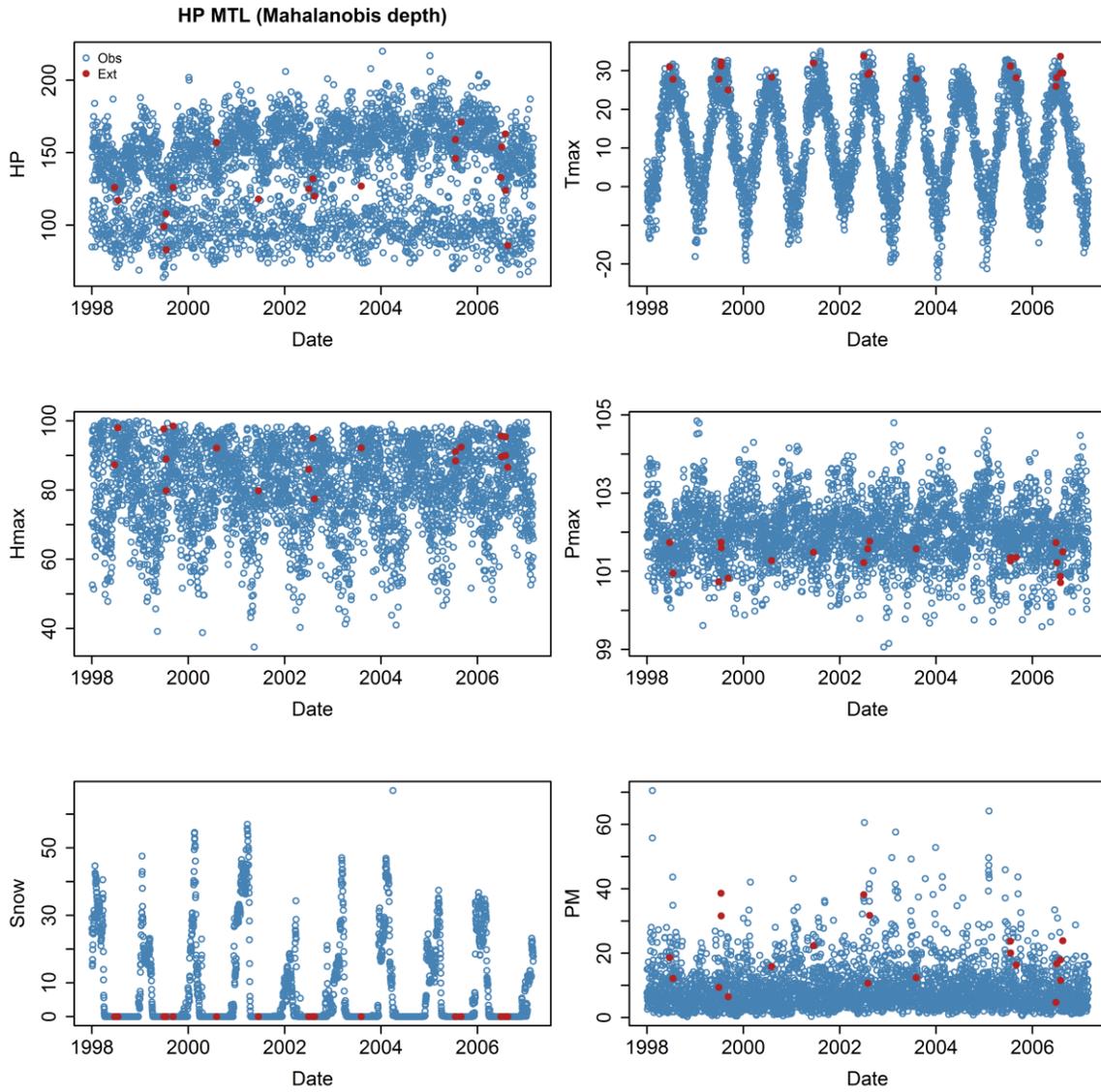


Figure IV-16 : Extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %) pour les hospitalisations à Montréal. Les extrêmes sont extraits selon la méthodologie de Chebana et Ouarda (2011b).

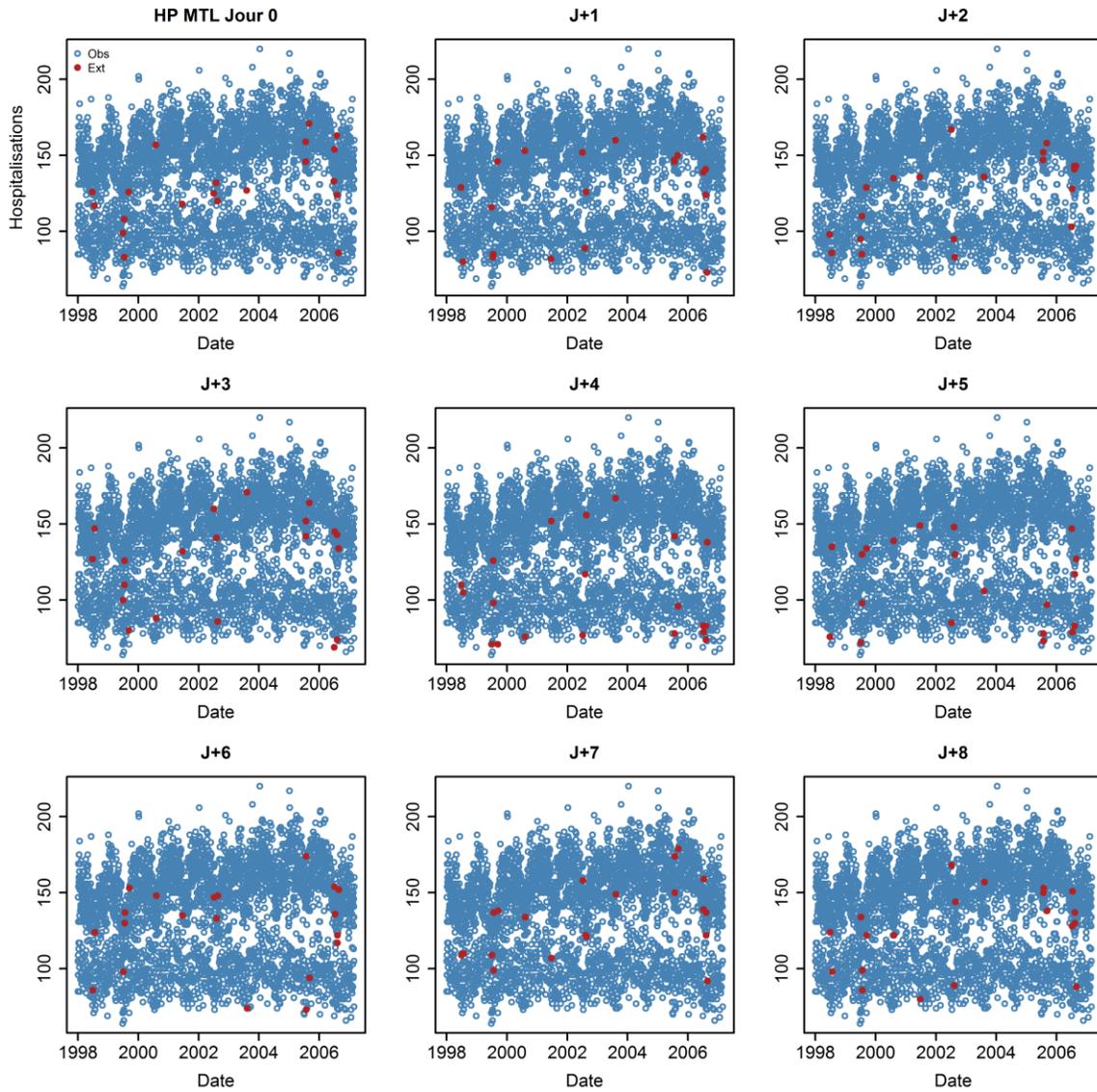


Figure IV-17 : Association des hospitalisations correspondantes aux extrêmes météorologiques multivariés de la température et de l'humidité (seuil à 1 %). Les hospitalisations vont du jour même à 8 jours après les extrêmes. Les extrêmes sont extraits selon la méthodologie de Chebana et Ouarda (2011b).

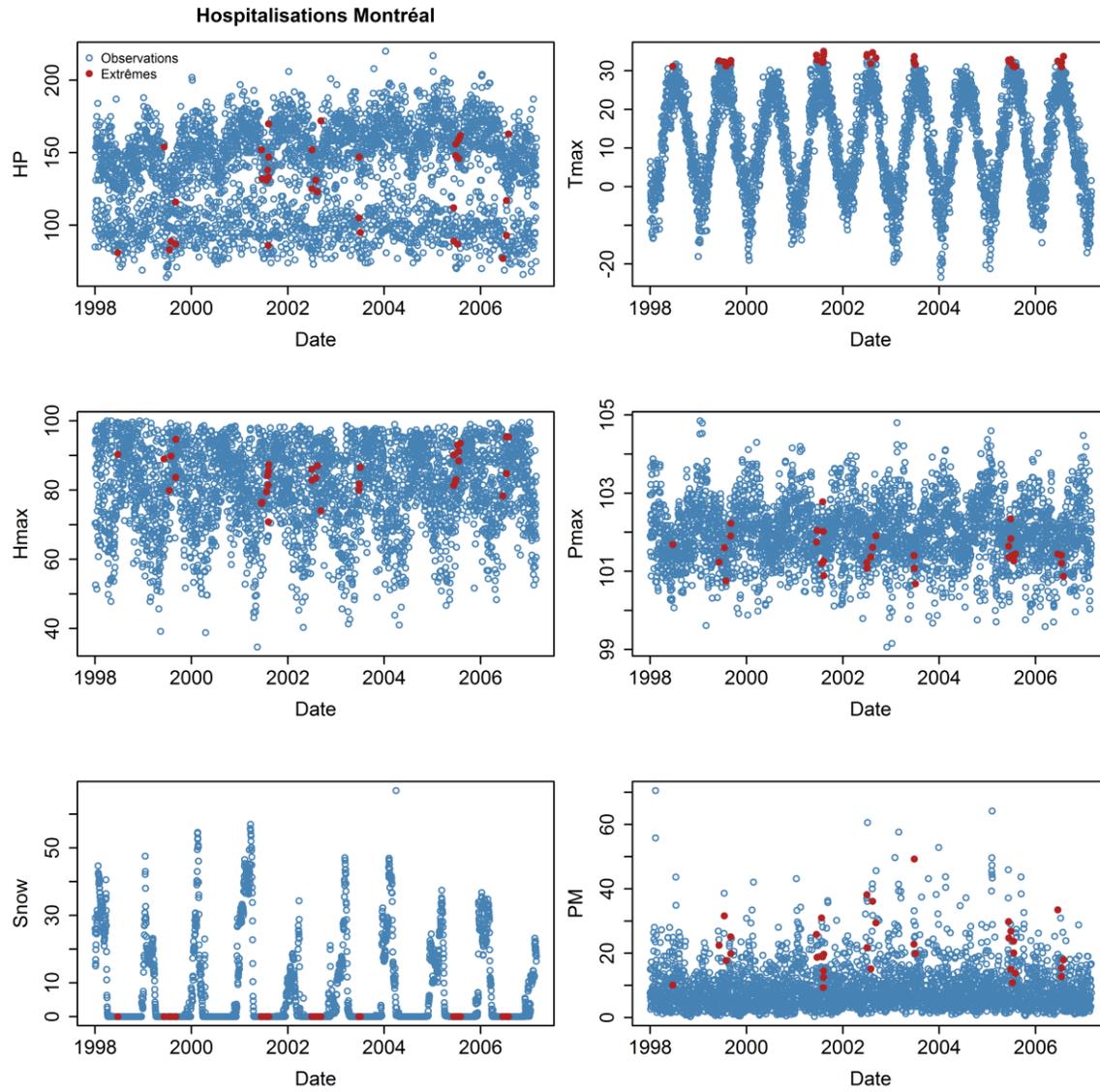


Figure IV-18 : Extrêmes météorologiques univariés de l'humidex (seuil à 99 %) pour les hospitalisations à Montréal.

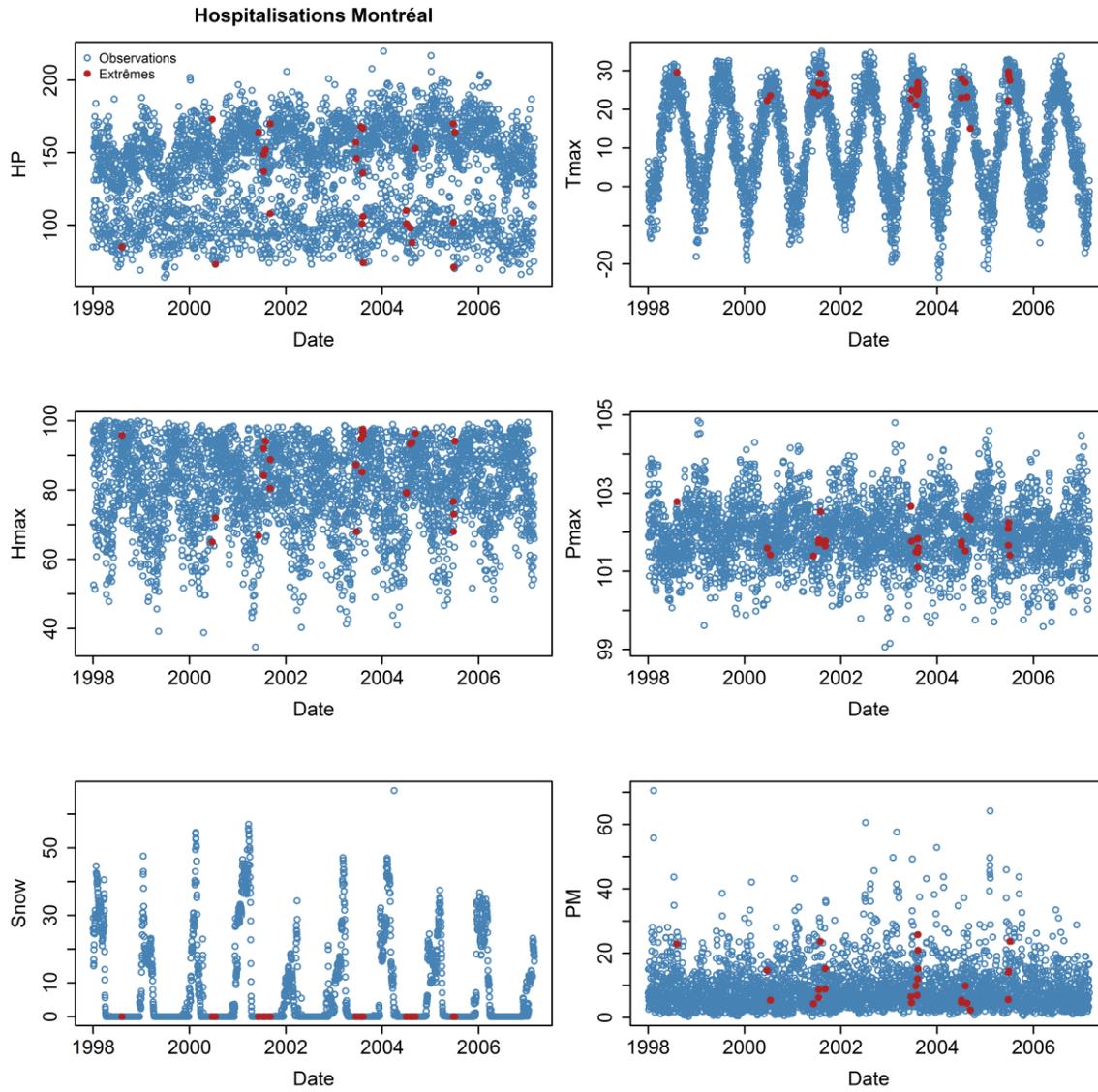


Figure IV-19 : Extrêmes météorologiques univariés de l'indice de chaleur (seuil à 99 %) pour les hospitalisations à Montréal.

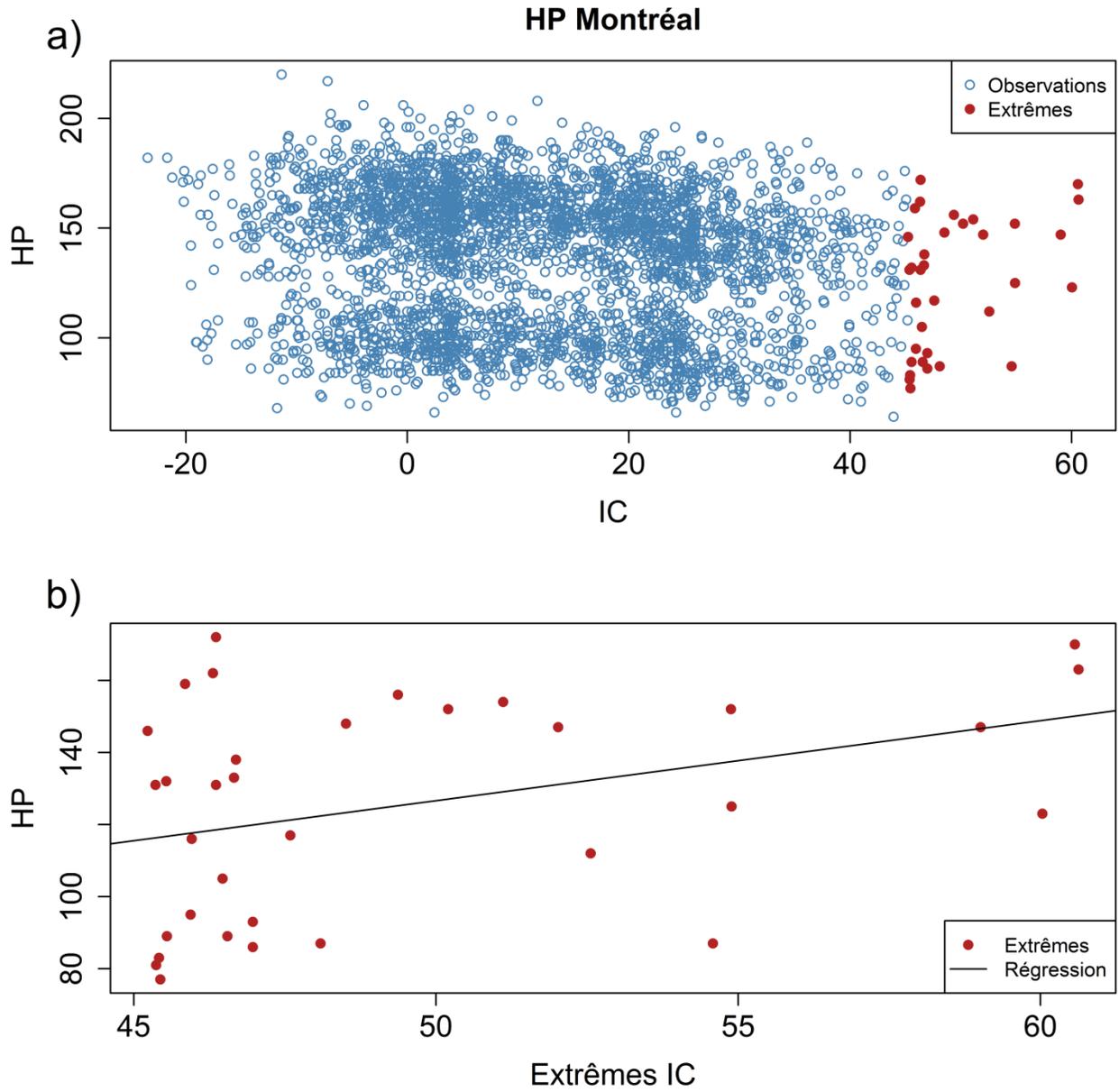


Figure IV-20 : a) Hospitalisations et indice de chaleur à Montréal et b) Droite de régression des extrêmes de l'indice de chaleur sur les hospitalisations.

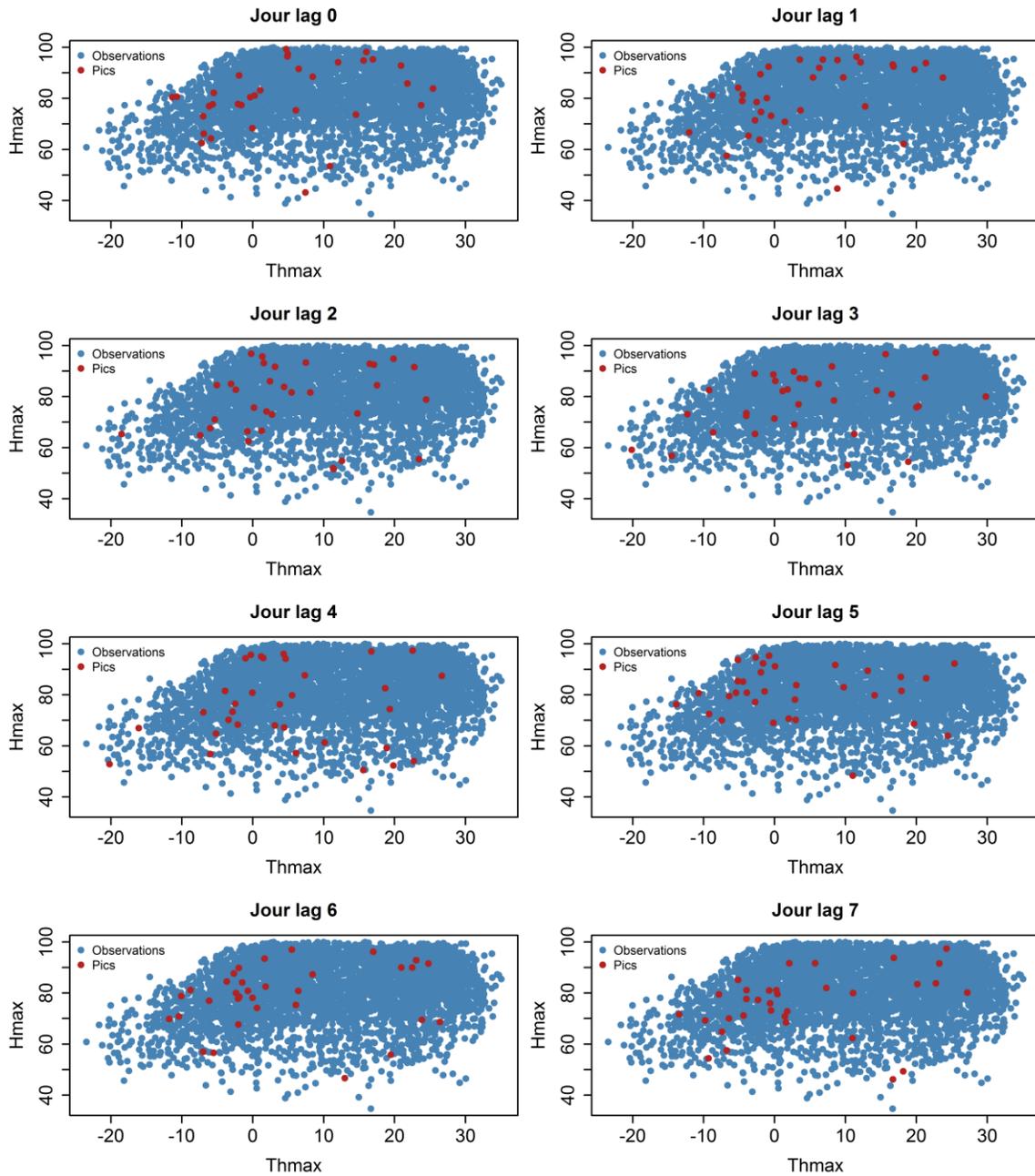


Figure IV-21 : Position des pics étudiés par Chiu *et al.* (2016b) dans l'espace température-humidité, jusqu'à 7 jours avant l'apparition des pics.

Annexes

A- Fonctions de profondeur

Le concept de fonction de profondeur a été introduit par Tukey (1975) pour pallier au manque de définition d'une médiane multivariée. Une fonction de profondeur $D(., F)$ est une fonction non négative et bornée, associée à une distribution multivariée F sur \mathbb{R}^d ($d \geq 1$), qui vérifie idéalement les quatre propriétés suivantes selon Zuo et Serfling (2000) :

- 1) *Invariance par transformation affine* : pour toute matrice A de taille $(d \times d)$ et pour tout vecteur $x, b \in \mathbb{R}^d$, $D(Ax + b; F_{AX+b}) = D(x; F_x)$. Ainsi, la profondeur ne dépend pas du système de coordonnées ;
- 2) *Maximalité au centre* : si θ est le centre de la distribution F , alors $D(\theta, F) = \max(D(x, F))$. Donc, D atteint sa valeur maximale en θ , qui est le point le plus profond de F ;
- 3) *Monotonie par rapport au point de profondeur maximale* : si θ est le centre de la distribution F , alors $D(x; F) \leq D(\theta + \alpha(x - \theta); F)$ pour tout $\alpha \in [0; 1]$, i.e. la profondeur est décroissante le long de toute demi-droite à partir de θ ;
- 4) *Valeur nulle à l'infini* : $D(x; F) \rightarrow 0$ quand $\|x\| \rightarrow \infty$. La fonction de profondeur tend vers 0 lorsque l'on s'éloigne du centre à l'infini.

La fonction de profondeur choisie $D(.,.): \mathbb{R}^d \times F \rightarrow \mathbb{R}$ associe à chaque observation de l'espace multivarié un nombre qui mesure sa centralité (ou son éloignement) pour une distribution donnée et permet ainsi d'établir un ordre pour un échantillon d'observations multivariées.

Diverses fonctions de profondeur sont définies et disponible dans la littérature comme les fonctions Halfspace, Projection et Simplex. Chebana et Ouarda (2011a) expliquent comment les fonctions de profondeur peuvent être utile en hydrologie en exprimant plusieurs caractéristiques d'une fonction (paramètre de position, d'échelle, de skewness, etc) en termes de profondeur. En particulier, la profondeur de Mahalanobis est utilisée par Chebana et Ouarda (2011b) dans une étude de pollution atmosphérique (dans un contexte bivarié), à Leeds (Angleterre). Elle est définie par :

$$MD(x; F) = \frac{1}{1 + d_A^2(x, \mu)} \quad (22)$$

où $d_A^2(x, y) = (x - y)A'(x - y)$ est la distance de Mahalanobis entre deux points $x, y \in \mathbb{R}^d$ avec A une matrice définie positive et F est une distribution associée aux mesures de position et de variance-covariance μ et A , respectivement. En utilisant un paramètre d'orientation u , il est possible de sélectionner une portion restreinte de l'espace multivarié (par exemple, uniquement les valeurs positives des variables). La Figure IV-4 illustre cette approche.

B- Indices météorologiques

L'indice de chaleur est défini par

$$IC = c_1 + c_2T + c_3H + c_4TH + c_5T^2 + c_6H^2 + c_7T^2H + c_8TH^2 + c_9T^2H^2 \quad (23)$$

avec T la température de l'air ($^{\circ}\text{F}$), H l'humidité relative (%), $c_1 = -42,38$, $c_2 = 2,05$, $c_3 = 10,14$, $c_4 = -0,22$, $c_5 = -6,84 \times 10^{-3}$, $c_6 = -5,48 \times 10^{-2}$, $c_7 = 1,23 \times 10^{-3}$, $c_8 = 8,53 \times 10^{-4}$ et $c_9 = -1,99 \times 10^{-6}$.

L'humidex est défini par

$$\text{Humidex} = T + 0,55 \times \left(6,11 \times \exp \left\{ 5417,75 \times \left(\frac{1}{273,16} - \frac{1}{T_d} \right) \right\} - 10 \right) \quad (24)$$

avec t_{air} la température de l'air ($^{\circ}\text{C}$) et T_d le point de rosée en kelvins. T_d est à son tour défini par

$$T_d = \left(\frac{H}{100} \right)^{1/8} \times (112 + 0.9T) + 0.1T - 112 \quad (25)$$

avec H l'humidité relative (%) et T la température de l'air ($^{\circ}\text{C}$).

RÉFÉRENCES

- Aarssen, K. et De Haan, L. (1994). On the maximal life span of humans. *Math Popul Stud* 4(4):259-281.
- Abrignani, M. G., Corrao, S., Biondo, G. B., Lombardo, R. M., Di Girolamo, P., Braschi, A., Di Girolamo, A. et Novo, S. (2012). Effects of ambient temperature, humidity, and other meteorological variables on hospital admissions for angina pectoris. *Eur J Prev Cardiol* 19(3):342-348.
- Afshin, A., Micha, R., Khatibzadeh, S. et Mozaffarian, D. (2014). Consumption of nuts and legumes and risk of incident ischemic heart disease, stroke, and diabetes: a systematic review and meta-analysis. *Am J Clin Nutr* 100(1):278-288.
- Agence de la santé publique du Canada (2009) *Suivi des maladies du cœur et des accidents vasculaires cérébraux au Canada*. Ottawa, Canada. 132 p
- Agence de la santé publique du Canada (2013) *Economic Burden of Illness in Canada, 2005-2008*. Centre for Public Health Information and Surveillance Strategy. 111 p
- AghaKouchak, A., Easterling, D. R., Hsu, K., Schubert, S. et Sorooshian, S. (2012) *Extremes in a changing climate*. Springer, Dordrecht
- Ahrens, W. et Pigeot, I. (2014) *Handbook of epidemiology*. Springer,
- Aldana, S. G., Whitmer, W. R., Greenlaw, R., Avins, A. L., Thomas, D., Salberg, A., Greenwell, A., Lipsenthal, L. et Fellingham, G. W. (2006). Effect of intense lifestyle modification and cardiac rehabilitation on psychosocial cardiovascular disease risk factors and quality of life. *Behav Modif* 30(4):507-525.
- Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B. et Rahimzadeh, F. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres* 111(D5).
- Anderson, C. W. (1970). Extreme Value Theory for a Class of Discrete Distributions with Applications to Some Stochastic Processes. *Journal of Applied Probability* 7(1):99-&.
- Anonyme (!!! INVALID CITATION !!! Kovats et al., 2004; Hajat et al., 2007; Li et al., 2012).
- Arnett, D. K., Goodman, R. A., Halperin, J. L., Anderson, J. L., Parekh, A. K. et Zoghbi, W. A. (2014). AHA/ACC/HHS strategies to enhance application of clinical practice guidelines in patients with cardiovascular disease and comorbid conditions. *Circulation* 130(18):1662-1667.
- Ashley, S. T. et Ashley, W. S. (2008). Flood fatalities in the United States. *Journal of Applied Meteorology and Climatology* 47(3):805-818.
- Auliciems, A. et Frost, D. (1989). Temperature and cardiovascular deaths in Montreal. *Int J Biometeorol* 33(3):151-156.
- Austin, P. C. et Hux, J. E. (2002). A brief note on overlapping confidence intervals. *J Vasc Surg* 36(1):194-195.

- Austin, P. C., Tu, J. V., Daly, P. A. et Alter, D. A. (2005). The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy. *Statistics in medicine* 24(5):791-816.
- Baker-Blocker, A. (1982). Winter weather and cardiovascular mortality in Minneapolis-St. Paul. *American journal of public health* 72(3):261-265.
- Balkema, A. A. et Dehaan, L. (1974). Residual Life Time at Great Age. *Ann Probab* 2(5):792-804.
- Barnett, A. G., Hajat, S., Gasparrini, A. et Rocklov, J. (2012). Cold and heat waves in the United States. *Environ Res* 112(0):218-224.
- Barnett, A. G., Tong, S. et Clements, A. C. A. (2010). What measure of temperature is the best predictor of mortality? *Environmental research* 110(6):604-611.
- Basu, R. (2009). High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environ Health* 8(1):40.
- Bayentin, L., El Adlouni, S., Ouarda, T. B., Gosselin, P., Doyon, B. et Chebana, F. (2010). Spatial variability of climate effects on ischemic heart disease hospitalization rates for the period 1989-2006 in Quebec, Canada. *Int J Health Geogr* 9(5):5.
- Beaney, G. et Gough, W. A. (2002). The influence of tropospheric ozone on the air temperature of the city of Toronto, Ontario, Canada. *Atmospheric environment* 36(14):2319-2325.
- Beaudreau, P., Besancenot, J., Caserio-Schönemann, C., Cohen, J., Dejour-Salamanca, D. et Empereur-Bissonnet, P. (2004) *Froid et santé: éléments de synthèse bibliographique et perspectives*. Institut de Veille Sanitaire,
- Beirlant, J., Caeiro, F. et Gomes, M. I. (2012). An overview and open research topics in statistics of univariate extremes. *Revstat* 10(1):1-31.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D. et Ferro, C. (2004a) *Statistics of Extremes: Theory and Applications*. Wiley, Chichester,
- Beirlant, J., Wet, T. D. et Goegebeur, Y. (2004b). Nonparametric estimation of extreme conditional quantiles. *Journal of statistical computation and simulation* 74(8):567-580.
- Bell, M. L., Ebisu, K., Peng, R. D., Samet, J. M. et Dominici, F. (2009). Hospital admissions and chemical composition of fine particle air pollution. *Am J Respir Crit Care Med* 179(12):1115-1120.
- Beniston, M. (2007). Linking extreme climate events and economic impacts: Examples from the Swiss Alps. *Energy Policy* 35(11):5384-5392.
- Beniston, M. et Stephenson, D. B. (2004). Extreme climatic events and their evolution under changing climatic conditions. *Global and Planetary Change* 44(1):1-9.
- Bergh, I. H., Skare, O., Aase, A., Klepp, K. I. et Lien, N. (2015). Weight development from age 13 to 30 years and adolescent socioeconomic status: The Norwegian Longitudinal Health Behaviour study. *Int J Public Health* 10.1007/s00038-015-0748-x:1-9.
- Bernal, J. L., Cummins, S. et Gasparrini, A. (2016). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International journal of epidemiology*.dyw098.

- Bezak, N., Brilly, M. et Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal* 59(5):959-977.
- Bhaskaran, K., Hajat, S., Haines, A., Herrett, E., Wilkinson, P. et Smeeth, L. (2009). The effects of ambient temperature on the incidence of myocardial infarction—A systematic review. *Heart*.
- Bhaskaran, K., Hajat, S., Haines, A., Herrett, E., Wilkinson, P. et Smeeth, L. (2010). Short term effects of temperature on risk of myocardial infarction in England and Wales: time series regression analysis of the Myocardial Ischaemia National Audit Project (MINAP) registry. *Bmj* 341:c3823.
- Bhatnagar, A. (2006). Environmental cardiology: studying mechanistic links between pollution and heart disease. *Circ Res* 99(7):692-705.
- Bind, M. A., Coull, B. A., Peters, A., Baccarelli, A. A., Tarantini, L., Cantone, L., Vokonas, P. S., Koutrakis, P. et Schwartz, J. D. (2015). Beyond the Mean: Quantile Regression to Explore the Association of Air Pollution with Gene-Specific Methylation in the Normative Aging Study. *Environ Health Perspect* 123(8):759-765.
- Bind, M. A., Peters, A., Koutrakis, P., Coull, B., Vokonas, P. et Schwartz, J. (2016). Quantile Regression Analysis of the Distributional Effects of Air Pollution on Blood Pressure, Heart Rate Variability, Blood Lipids, and Biomarkers of Inflammation in Elderly American Men: The Normative Aging Study. *Environ Health Perspect* 10.1289/ehp.1510044.
- Black, E., Blackburn, M., Harrison, G., Hoskins, B. et Methven, J. (2004). Factors contributing to the summer 2003 European heatwave. *Weather* 59(8):217-223.
- Blanchet, J., Marty, C. et Lehning, M. (2009). Extreme value statistics of snowfall in the Swiss Alpine region. *Water Resources Research* 45(5).
- Blazejczyk, K., Epstein, Y., Jendritzky, G., Staiger, H. et Tinz, B. (2012). Comparison of UTCI to selected thermal indices. *Int J Biometeorol* 56(3):515-535.
- Bobb, J. F., Peng, R. D., Bell, M. L. et Dominici, F. (2014). Heat-related mortality and adaptation to heat in the United States. *Environ Health Perspect* 122(8):811-816.
- Bondell, H. D., Reich, B. J. et Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika* 97(4):825-838.
- Bouchama, A., Dehbi, M., Mohamed, G., Matthies, F., Shoukri, M. et Menne, B. (2007). Prognostic factors in heat wave related deaths: a meta-analysis. *Arch Intern Med* 167(20):2170-2176.
- Box, G. (1953). Non-normality and tests on variances. *Biometrika* 40(3/4):318-335.
- Braga, A. L. F., Zanobetti, A. et Schwartz, J. (2002). The effect of weather on respiratory and cardiovascular deaths in 12 US cities. *Environmental Health Perspectives* 110(9):859-863.
- Breitner, S., Wolf, K., Devlin, R. B., Diaz-Sanchez, D., Peters, A. et Schneider, A. (2014). Short-term effects of air temperature on mortality and effect modification by air pollution in three cities of Bavaria, Germany: a time-series analysis. *The Science of the total environment* 485-486(0):49-61.

- Brook, R. D., Rajagopalan, S., Pope, C. A., 3rd, Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Jr., Whitsel, L., Kaufman, J. D., American Heart Association Council on, E., Prevention, C. o. t. K. i. C. D., Council on Nutrition, P. A. et Metabolism (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation* 121(21):2331-2378.
- Bustinza, R., Lebel, G., Gosselin, P., Belanger, D. et Chebana, F. (2013). Health impacts of the July 2010 heat wave in Quebec, Canada. *BMC Public Health* 13(1):56.
- Cade, B. S. et Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* 1(8):412-420.
- Cao, J., Xu, H., Xu, Q., Chen, B. et Kan, H. (2012). Fine particulate matter constituents and cardiopulmonary mortality in a heavily polluted Chinese city. *Environ Health Perspect* 120(3):373-378.
- Carder, M., McNamee, R., Beverland, I., Elton, R., Cohen, G. R., Boyd, J. et Agius, R. M. (2005). The lagged effect of cold temperature and wind chill on cardiorespiratory mortality in Scotland. *Occup Environ Med* 62(10):702-710.
- Carter, S. J., Durrington, H. J., Gibbs, J. E., Blaikley, J., Loudon, A. S., Ray, D. W. et Sabroe, I. (2016). A matter of time: study of circadian clocks and their role in inflammation. *Journal of leukocyte biology* 99(4):549-560.
- Caussin, C., Escolano, S., Mustafic, H., Bataille, S., Tafflet, M., Chatignoux, E., Lambert, Y., Benamer, H., Garot, P., Jabre, P., Delorme, L., Varenne, O., Teiger, E., Livarek, B., Empana, J. P., Spaulding, C., Jouven, X. et Investigators, C.-A. R. (2015). Short-term exposure to environmental parameters and onset of ST elevation myocardial infarction. The CARDIO-ARSIF registry. *Int J Cardiol* 183:17-23.
- Center for Disease Control Prevention (2004). Impact of heat waves on mortality--Rome, Italy, June-August 2003. *MMWR. Morbidity and mortality weekly report* 53(17):369.
- Chang, C. L., Shipley, M., Marmot, M. et Poulter, N. (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.
- Chavez-Demoulin, V., Embrechts, P. et Hofert, M. (2015). An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates. *Journal of Risk and Insurance* 10.1111/jori.12059.
- Chavez-Demoulin, V. et Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54(1):207-222.
- Chebana, F., Martel, B., Gosselin, P., Giroux, J. X. et Ouarda, T. B. (2013). A general and flexible methodology to define thresholds for heat health watch and warning systems, applied to the province of Quebec (Canada). *Int J Biometeorol* 57(4):631-644.
- Chebana, F. et Ouarda, T. B. M. J. (2011a). Depth-based multivariate descriptive statistics with hydrological applications. *J Geophys Res-Atmos* 116(D10).
- Chebana, F. et Ouarda, T. B. M. J. (2011b). Multivariate extreme value identification using depth functions. *Environmetrics* 22(3):441-455.

- Chen, H., Wang, J., Li, Q., Yagouti, A., Lavigne, E., Foty, R., Burnett, R. T., Villeneuve, P. J., Cakmak, S. et Copes, R. (2016). Assessment of the effect of cold and hot temperatures on mortality in Ontario, Canada: a population-based study. *CMAJ Open* 4(1):E48-58.
- Chen, J., Lei, X., Zhang, L. et Peng, B. (2015). Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China. *PLoS One* 10(2):e0118521.
- Chernozhukov, V., Fernández-Val, I. et Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica* 78(3):1093-1125.
- Chiu, Y., Chebana, F., Abdous, B., Belanger, D. et Gosselin, P. (2016a). Mortality and morbidity peaks modeling: An extreme value theory approach. *Stat Methods Med Res* 10.1177/0962280216662494:0962280216662494.
- Chiu, Y., Chebana, F., Abdous, B., Bélanger, D. et Gosselin, P. (2015) Modélisation des pics de mortalité et de morbidité hospitalière pour cause de maladies cardiovasculaires à Québec et Montréal (Québec) : Une approche par la théorie des valeurs extrêmes. (Institut national de la recherche scientifique, Québec), p 129.
- Chiu, Y., Chebana, F., Bélanger, D., Gosselin, P. et Abdous, B. (2016b) Modélisation des pics sanitaires de maladies cardiovasculaires en fonction de la météo dans les communautés métropolitaines de Québec et Montréal. (Institut national de la recherche scientifique), p 128.
- Coles, S. (2001) *An introduction to statistical modeling of extreme values*. Springer, London
- Coles, S., Heffernan, J. et Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* 2(4):339-365.
- Coles, S., Pericchi, L. R. et Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology* 273(1-4):35-50.
- Communauté métropolitaine de Montréal (2017) *Grand Montréal en statistiques*. <http://cmm.qc.ca/fr/donnees-et-territoire/observatoire-grand-montreal/produits-statistiques/grand-montreal-en-statistiques/?t=2&st=8&i=1221&p=2016&e=3> (Consulté le 19-01-2017)
- Costello, A., Abbas, M., Allen, A., Ball, S., Bell, S., Bellamy, R., Friel, S., Groce, N., Johnson, A., Kett, M., Lee, M., Levy, C., Maslin, M., McCoy, D., McGuire, B., Montgomery, H., Napier, D., Pagel, C., Patel, J., de Oliveira, J. A., Redclift, N., Rees, H., Rogger, D., Scott, J., Stephenson, J., Twigg, J., Wolff, J. et Patterson, C. (2009). Managing the health effects of climate change: Lancet and University College London Institute for Global Health Commission. *Lancet* 373(9676):1693-1733.
- Cumming, G. (2009). Inference by eye: reading the overlap of independent confidence intervals. *Stat Med* 28(2):205-220.
- Cumming, G. et Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *American Psychologist* 60(2):170.
- Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L. et Patz, J. A. (2002). Temperature and mortality in 11 cities of the eastern United States. *Am J Epidemiol* 155.

- Curriero, F. C., Patz, J. A., Rose, J. B. et Lele, S. (2001). The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. *American journal of public health* 91(8):1194-1199.
- Daigle, J. M. (2007) *Les maladies du coeur et les maladies vasculaires cérébrales: Prévalence, morbidité et mortalité au Québec*. Institut national de santé publique du Québec, Québec
- Danet, S., Richard, F., Montaye, M., Beauchant, S., Lemaire, B., Graux, C., Cottel, D., Marecaux, N. et Amouyel, P. (1999). Unhealthy effects of atmospheric temperature and pressure on the occurrence of myocardial infarction and coronary deaths. A 10-year survey: the Lille-World Health Organization MONICA project (Monitoring trends and determinants in cardiovascular disease). *Circulation* 100(1):E1-7.
- Dang, X. et Serfling, R. (2010). Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties. *Journal of Statistical Planning and Inference* 140(1):198-213.
- Daouia, A., Gardes, L. et Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* 19(5B):2557-2589.
- Davison, A. C. et Smith, R. L. (1990). Models for Exceedances over High Thresholds. *J Roy Stat Soc B Met* 52(3):393-442.
- De Sario, M., Katsouyanni, K. et Michelozzi, P. (2013). Climate change, extreme weather events, air pollution and respiratory health in Europe. *Eur Respir J* 42(3):826-843.
- De Zea Bermudez, P. et Mendes, Z. (2012). Extreme value theory in medical sciences: Modeling total high cholesterol levels. *Journal of Statistical Theory and Practice* 6(3):468-491.
- Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P. et Meester, L. E. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media,
- Dette, H. et Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(3):609-627.
- Dinas, P. C., Koutedakis, Y. et Flouris, A. D. (2013). Effects of active and passive tobacco cigarette smoking on heart rate variability. *Int J Cardiol* 163(2):109-115.
- Dolney, T. J. et Sheridan, S. C. (2006). The relationship between extreme heat and ambulance response calls for the city of Toronto, Ontario, Canada. *Environmental Research* 101(1):94-103.
- Dominici, F., McDermott, A., Zeger, S. L. et Samet, J. M. (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.
- Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L. et Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA* 295(10):1127-1134.
- Dominici, F., Samet, J. M. et Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 163(3):263-302.

- Doyon, B., Bélanger, D. et Gosselin, P. (2008). The potential impact of climate change on annual and seasonal mortality for three cities in Quebec, Canada. *Int J Health Geogr* 7(1):23.
- Doyon, B., Bélanger, D. et Gosselin, P. (2006) *Effets du climat sur la mortalité au Québec méridional de 1981 à 1999 et simulations pour des scénarios climatiques futures*. Direction risques biologiques, environnementaux et occupationnels, Institut national de santé publique Québec,
- Ebi, K. L., Exuzides, K. A., Lau, E., Kelsh, M. et Barnston, A. (2004). Weather changes associated with hospitalizations for cardiovascular diseases and stroke in California, 1983-1998. *Int J Biometeorol* 49(1):48-58.
- Ebi, K. L. et Mills, D. (2013). Winter mortality in a warming climate: a reassessment. *Wiley Interdisciplinary Reviews: Climate Change* 4(3):203-212.
- Ebi, K. L. et Schmier, J. K. (2005). A stitch in time: improving public health early warning systems for extreme weather events. *Epidemiol Rev* 27(1):115-121.
- El Adlouni, S., Bobée, B. et Ouarda, T. B. M. J. (2008). On the tails of extreme event distributions in hydrology. *Journal of Hydrology* 355(1):16-33.
- El Adlouni, S., Chebana, F. et Bobée, B. (2009). Generalized extreme value versus Halphen system: Exploratory study. *Journal of Hydrologic Engineering* 15(2):79-89.
- Elsner, J. B., Kossin, J. P. et Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* 455(7209):92-95.
- Embrechts, P. et Schmidli, H. (1994). Modelling of extremal events in insurance and finance. *Zeitschrift für Operations Research* 39(1):1-34.
- Engeland, K., Hisdal, H. et Frigessi, A. (2004). Practical extreme value modelling of hydrological floods and droughts: a case study. *Extremes* 7(1):5-30.
- Environnement et Changement climatique Canada (2013) *Les dix événements météorologiques les plus marquants au Canada en 2011*. <http://www.ec.gc.ca/meteo-weather/default.asp?lang=Fr&n=D7CF8BE3-1> (Consulté le 17-01-2017)
- Fan, J. Q. et Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96(456):1348-1360.
- Fan, Y. Y. et Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *J Roy Stat Soc B* 75(3):531-552.
- Fenske, N., Burns, J., Hothorn, T. et Rehfuess, E. A. (2013). Understanding child stunting in India: a comprehensive analysis of socio-economic, nutritional and environmental determinants using additive quantile regression. *PLoS One* 8(11):e78692.
- Filleul, L., Cassadou, S., Médina, S., Fabres, P., Lefranc, A., Eilstein, D., Le Tertre, A., Pascal, L., Chardon, B. et Blanchard, M. (2006). The relation between temperature, ozone, and mortality in nine French cities during the heat wave of 2003. *Environmental Health Perspectives*:1344-1347.
- Fisher, R. A. et Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24(02):180-190.

- FitzGerald, G., Du, W., Jamal, A., Clark, M. et Hou, X. Y. (2010). Flood fatalities in contemporary Australia (1997-2008). *Emerg Med Australas* 22(2):180-186.
- Fong, T. et Ma, E. (2013). Effects of meteorological parameters on hospital admission for respiratory and cardiovascular diseases. *Journal of Public Health* 21(2):175-182.
- Fouillet, A., Rey, G., Wagner, V., Laaidi, K., Empereur-Bissonnet, P., Le Tertre, A., Frayssinet, P., Bessemoulin, P., Laurent, F., De Crouy-Chanel, P., Jouglu, E. et Hemon, D. (2008). Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. *Int J Epidemiol* 37(2):309-317.
- Franklin, B. A., Hogan, P., Bonzheim, K., Bakalyar, D., Terrien, E., Gordon, S. et Timmis, G. C. (1995). Cardiac demands of heavy snow shoveling. *JAMA* 273(11):880-882.
- Gasparrini, A., Armstrong, B. et Kenward, M. G. (2010). Distributed lag non-linear models. *Stat Med* 29.
- Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklov, J., Forsberg, B., Leone, M., De Sario, M., Bell, M. L., Guo, Y. L., Wu, C. F., Kan, H., Yi, S. M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P. H., Honda, Y., Kim, H. et Armstrong, B. (2015). Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *Lancet* 386(9991):369-375.
- Gérin, M., Gosselin, P., Cordier, S., Viau, C., Quénel, P. et Dewailly, É. (2003) *Environnement et santé publique-Fondements et pratiques*. Édisem/Tec & Doc. 1062 p
- Giguère, M. (2009) *Mesures de lutte aux îlots de chaleur urbains*. Revue de littérature, Direction des risques biologiques, environnementaux et occupationnels, Institut National de Santé Publique, Gouvernement du Québec,
- Gilli, M. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics* 27(2-3):207-228.
- Girard, C. (2013) *Le bilan démographique du Québec*. Institut de la statistique du Québec, Québec
- Goggins, W. B., Woo, J., Ho, S., Chan, E. Y. et Chau, P. H. (2012). Weather, season, and daily stroke admissions in Hong Kong. *Int J Biometeorol* 56(5):865-872.
- Goldberg, M. S., Gasparrini, A., Armstrong, B. et Valois, M. F. (2011). The short-term influence of temperature on daily mortality in the temperate climate of Montreal, Canada. *Environ Res* 111(6):853-860.
- Goldberg, M. S., Wheeler, W. J., Burnett, R. T., Mayo, N. E., Valois, M.-F., Brophy, J. M. et Giannetti, N. (2015). Physiological and perceived health effects from daily changes in air pollution and weather among persons with heart failure: A panel study. *Journal of Exposure Science and Environmental Epidemiology* 25(2):187-199.
- Gomes, M. I. et Guillou, A. (2015). Extreme Value Theory and Statistics of Univariate Extremes: A Review. *Int Stat Rev* 83(2):263-292.
- Gosling, S. N., Lowe, J. A., McGregor, G. R., Pelling, M. et Malamud, B. D. (2009). Associations between elevated atmospheric temperature and human mortality: a critical review of the literature. *Climatic Change* 92(3-4):299-341.
- Gouveia, N., Hajat, S. et Armstrong, B. (2003). Socioeconomic differentials in the temperature-mortality relationship in Sao Paulo, Brazil. *Int J Epidemiol* 32(3):390-397.

- Greenough, G., McGeehin, M., Bernard, S. M., Trtanj, J., Riad, J. et Engelberg, D. (2001). The potential impacts of climate variability and change on health impacts of extreme weather events in the United States. *Environ Health Perspect* 109 Suppl 2(Suppl 2):191-198.
- Gruska, M., Gaul, G. B., Winkler, M., Levnaic, S., Reiter, C., Voracek, M. et Kaff, A. (2005). Increased occurrence of out-of-hospital cardiac arrest on Mondays in a community-based study. *Chronobiol Int* 22(1):107-120.
- Guillou, A., Kratz, M. et Le Strat, Y. (2014). An extreme value theory approach for the early detection of time clusters. A simulation-based assessment and an illustration to the surveillance of Salmonella. *Stat Med* 33(28):5015-5027.
- Guo, Y., Li, S., Zhang, Y., Armstrong, B., Jaakkola, J. J., Tong, S. et Pan, X. (2013). Extremely cold and hot temperatures increase the risk of ischaemic heart disease mortality: epidemiological evidence from China. *Heart* 99(3):195-203.
- Hajat, S., Armstrong, B. G., Gouveia, N. et Wilkinson, P. (2005). Mortality displacement of heat-related deaths: a comparison of Delhi, Sao Paulo, and London. *Epidemiology* 16.
- Hajat, S., Chalabi, Z., Wilkinson, P., Erens, B., Jones, L. et Mays, N. (2016). Public health vulnerability to wintertime weather: time-series regression and episode analyses of national mortality and morbidity databases to inform the Cold Weather Plan for England. *Public Health* 137:26-34.
- Hajat, S., Sheridan, S. C., Allen, M. J., Pascal, M., Laaidi, K., Yagouti, A., Bickis, U., Tobias, A., Bourque, D., Armstrong, B. G. et Kosatsky, T. (2010). Heat-health warning systems: a comparison of the predictive capacity of different approaches to identifying dangerously hot days. *American journal of public health* 100(6):1137-1144.
- Hajat, S., Vardoulakis, S., Heaviside, C. et Eggen, B. (2014). Climate change effects on human health: projections of temperature-related mortality for the UK during the 2020s, 2050s and 2080s. *J Epidemiol Community Health* 68(7):641-648.
- Hales, S., Edwards, S. J. et Kovats, R. S. (2003). Impacts on health of climate extremes. *Climate change and health: risks and responses*. Geneva, World Health Organization.
- Hastie, T. et Tibshirani, R. (1986). Generalized additive models. *Statistical science* 1(3):297-310.
- Hastie, T. J. et Tibshirani, R. J. (1990) *Generalized additive models*. CRC Press,
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences* 44(1):1-12.
- Heidenreich, P. A., Trogon, J. G., Khavjou, O. A., Butler, J., Dracup, K., Ezekowitz, M. D., Finkelstein, E. A., Hong, Y., Johnston, S. C., Khera, A., Lloyd-Jones, D. M., Nelson, S. A., Nichol, G., Orenstein, D., Wilson, P. W., Woo, Y. J., American Heart Association Advocacy Coordinating, C., Stroke, C., Council on Cardiovascular, R., Intervention, Council on Clinical, C., Council on, E., Prevention, Council on, A., Thrombosis, Vascular, B., Council on, C., Critical, C., Perioperative, Resuscitation, Council on Cardiovascular, N., Council on the Kidney in Cardiovascular, D., Council on Cardiovascular, S., Anesthesia, Interdisciplinary Council on Quality of, C. et Outcomes, R. (2011). Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation* 123(8):933-944.
- Hémon, D. et Jouglu, E. (2003). Surmortalité liée à la canicule d'août 2003: rapport d'étape.

- Henriques-Rodrigues, L., Gomes, M. I. et Pestana, D. (2011). Statistics of Extremes in Athletics. *Revstat-Stat J*9(2):127-153.
- Herrador, B. R. G., De Blasio, B. F., MacDonald, E., Nichols, G., Sudre, B., Vold, L., Semenza, J. C. et Nygård, K. (2015). Analytical studies assessing the association between extreme precipitation or temperature and drinking water-related waterborne infections: a review. *Environmental Health* 14(1):1.
- Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orłowsky, B. et Stepanek, P. (2011). Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience* 4(1):17-21.
- Hitz, A., Davis, R. et Samorodnitsky, G. (2017). Discrete Extremes. *arXiv preprint arXiv:1707.05033*.
- Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B. et Kaufman, J. D. (2013). Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environ Health* 12(1):43.
- Holick, M. F. (2004). Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *Am J Clin Nutr* 80(6 Suppl):1678S-1688S.
- Hondula, D. M., Vanos, J. K. et Gosling, S. N. (2014). The SSC: a decade of climate-health research and future directions. *Int J Biometeorol* 58(2):109-120.
- Hong, Y. C., Rha, J. H., Lee, J. T., Ha, E. H., Kwon, H. J. et Kim, H. (2003). Ischemic stroke associated with decrease in temperature. *Epidemiology* 14(4):473-478.
- Hopstock, L. A., Fors, A. S., Bonna, K. H., Mannsverk, J., Njolstad, I. et Wilsgaard, T. (2012). The effect of daily weather conditions on myocardial infarction incidence in a subarctic population: the Tromso Study 1974-2004. *J Epidemiol Community Health* 66(9):815-820.
- Hosking, J. R. M. et Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* 29(3):339-349.
- Houck, P. D., Lethen, J. E., Riggs, M. W., Gantt, D. S. et Dehmer, G. J. (2005). Relation of atmospheric pressure changes and the occurrences of acute myocardial infarction and stroke. *Am J Cardiol* 96(1):45-51.
- Institut Royal Météorologique (2017) *FAQ sur le temps*. <http://www.meteo.be/meteo/view/fr/68771-FAQ+sur+le+temps.html?view=3269224> (Consulté le 17-01-2017)
- IPCC (2013) *The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 1007 p
- IPCC (2014) *Climate change 2014: mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Ishikawa, K., Niwa, M. et Tanaka, T. (2012). Difference of intensity and disparity in impact of climate on several vascular diseases. *Heart and vessels* 27(1):1-9.
- Jagger, T. H. et Elsner, J. B. (2009). Modeling tropical cyclone intensity with quantile regression. *International Journal of Climatology* 29(10):1351-1361.

- Jbilou, J. et El Adlouni, S. (2012). Generalized additive models in environmental health: a literature review. *Novel Approaches and Their Applications in Risk Assessment*; Luo, Y., Ed.
- Jimenez-Conde, J., Ois, A., Gomis, M., Rodriguez-Campello, A., Cuadrado-Godia, E., Subirana, I. et Roquer, J. (2008). Weather as a trigger of stroke. *Cerebrovascular Diseases* 26(4):348-354.
- Jin, Z., Oresko, J., Huang, S. et Cheng, A. C. (2009) HeartToGo: a personalized medicine technology for cardiovascular disease prevention and detection. *Life Science Systems and Applications Workshop, 2009. LISSA 2009. IEEE/NIH. IEEE*, p 80-83.
- Johnston, F., Hanigan, I., Henderson, S., Morgan, G. et Bowman, D. (2011). Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994–2007. *Environmental research* 111(6):811-816.
- Kathiresan, S. et Srivastava, D. (2012). Genetics of human cardiovascular disease. *Cell* 148(6):1242-1257.
- Katz, R. W., Parlange, M. B. et Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources* 25(8-12):1287-1304.
- Keatinge, W. R., Coleshaw, S. R., Easton, J. C., Cotter, F., Mattock, M. B. et Chelliah, R. (1986). Increased platelet and red cell counts, blood viscosity, and plasma cholesterol levels during heat stress, and mortality from coronary and cerebral thrombosis. *Am J Med* 81(5):795-800.
- Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P. et Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology* 329(3-4):534-552.
- Khanjani, N. et Bahrampour, A. (2013). Temperature and cardiovascular and respiratory mortality in desert climate. A case study of Kerman, Iran. *Iranian J Environ Health Sci Eng* 10(1):11.
- Kim, Y. J. et Gu, C. (2004). Smoothing spline Gaussian regression: more scalable computation via efficient approximation. *J Roy Stat Soc B* 66(2):337-356.
- Knezevic, A. (2008). Overlapping confidence intervals and statistical significance. *StatNews: Cornell University Statistical Consulting Unit* 73.
- Knowlton, K., Lynn, B., Goldberg, R. A., Rosenzweig, C., Hogrefe, C., Rosenthal, J. K. et Kinney, P. L. (2007). Projecting heat-related mortality impacts under a changing climate in the New York City region. *American journal of public health* 97(11):2028-2034.
- Knowlton, K., Rotkin-Ellman, M., Geballe, L., Max, W. et Solomon, G. M. (2011). Six climate change-related events in the United States accounted for about \$14 billion in lost lives and health costs. *Health Affairs* 30(11):2167-2176.
- Knowlton, K., Rotkin-Ellman, M., King, G., Margolis, H. G., Smith, D., Solomon, G., Trent, R. et English, P. (2009). The 2006 California heat wave: impacts on hospitalizations and emergency department visits. *Environ Health Perspect* 117(1):61-67.
- Koenker, R. (2005) *Quantile regression*. Cambridge university press,
- Koenker, R. (2010) Additive models for quantile regression: An analysis of risk factors for malnutrition in india. *Advances in Social Science Research Using R*, Springer. p 23-33.

- Koenker, R. et Bassett, G. (1978). Regression Quantiles. *Econometrica* 46(1):33.
- Koppe, C., Kovats, S., Menne, B. et Jendritzky, G. (2004) *Heat-waves: risks and responses*. Organisation Mondiale de la Santé. 123 p
- Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrological sciences journal* 49(4):575–590.
- Kovats, R. S. et Hajat, S. (2008). Heat stress and public health: a critical review. *Annual review of public health* 29:41-55.
- Krall, J. R., Anderson, G. B., Dominici, F., Bell, M. L. et Peng, R. D. (2013). Short-term exposure to particulate matter constituents and mortality in a national study of US urban communities.
- Krantz, M. J., Coronel, S. M., Whitley, E. M., Dale, R., Yost, J. et Estacio, R. O. (2013). Effectiveness of a community health worker cardiovascular risk reduction program in public health and health care settings. *American journal of public health* 103(1):e19-27.
- Kysely, J. et Kim, J. (2009). Mortality during heat waves in South Korea, 1991 to 2005: how exceptional was the 1994 heat wave? *Climate research* 38(2):105-116.
- Kysely, J., Pokorna, L., Kyncl, J. et Kriz, B. (2009). Excess cardiovascular mortality associated with cold spells in the Czech Republic. *BMC Public Health* 9(1):19.
- Laaidi, M., Laaidi, K. et Besancenot, J.-P. (2006). Temperature-related mortality in France, a comparison between regions with different climates from the perspective of global warming. *International journal of biometeorology* 51(2):145-153.
- Laio, F. (2004). Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research* 40(9).
- Lang, M., Ouarda, T. B. M. J. et Bobee, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 225(3-4):103-117.
- Lanzinger, S., Breitner, S., Neas, L., Cascio, W., Diaz-Sanchez, D., Hinderliter, A., Peters, A., Devlin, R. B. et Schneider, A. (2014). The impact of decreases in air temperature and increases in ozone on markers of endothelial function in individuals having type-2 diabetes. *Environ Res* 134C(0):331-338.
- Laschewski, G. et Jendritzky, G. (2002). Effects of the thermal environment on human health: an investigation of 30 years of daily mortality data from SW Germany. *Climate Research* 21(1):91-103.
- Lavigne, E., Gasparrini, A., Wang, X., Chen, H., Yagouti, A., Fleury, M. D. et Cakmak, S. (2014). Extreme ambient temperatures and cardiorespiratory emergency room visits: assessing risk by comorbid health conditions in a time series study. *Environ Health* 13(1):5.
- Law, M. R., Wald, N. J. et Rudnicka, A. R. (2003). Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *Bmj* 326(7404):1423.
- Leal, J., Luengo-Fernandez, R., Gray, A., Petersen, S. et Rayner, M. (2006). Economic burden of cardiovascular diseases in the enlarged European Union. *Eur Heart J* 27(13):1610-1619.
- Ledrans, M. et Isnard, H. (2003) *Impact sanitaire de la vague de chaleur d'août 2003 en France: bilan et perspectives, octobre 2003*. InVS, Saint-Maurice, France

- Lee, M., Nordio, F., Zanobetti, A., Kinney, P., Vautard, R. et Schwartz, J. (2014). Acclimatization across space and time in the effects of temperature on mortality: a time-series analysis. *Environ Health* 13(1):89.
- Lejeune, M. (2011) *Statistique: La théorie et ses applications*:. Springer Science & Business Media. 446 p
- LeTertre, A., Lefranc, A., Eilstein, D., Declercq, C., Medina, S., Blanchard, M., Chardon, B., Fabre, P., Filleul, L. et Jusot, J. F. (2006). Impact of the 2003 heatwave on all-cause mortality in 9 French cities. *Epidemiology* 17.
- Li, B., Sain, S., Mearns, L. O., Anderson, H. A., Kovats, S., Ebi, K. L., Bekkedal, M. Y. V., Kanarek, M. S. et Patz, J. A. (2012). The impact of extreme heat on morbidity in Milwaukee, Wisconsin. *Climatic Change* 110(3-4):959-976.
- Lim, Y. H., Hong, Y. C. et Kim, H. (2012). Effects of diurnal temperature range on cardiovascular and respiratory hospital admissions in Korea. *The Science of the total environment* 417-418:55-60.
- Lin, S., Luo, M., Walker, R. J., Liu, X., Hwang, S. A. et Chinery, R. (2009). Extreme high temperatures and hospital admissions for respiratory and cardiovascular diseases. *Epidemiology* 20(5):738-746.
- Linares, C., Díaz, J., Tobías, A., Carmona, R. et Mirón, I. J. (2015). Impact of heat and cold waves on circulatory-cause and respiratory-cause mortality in Spain: 1975–2008. *Stochastic Environmental Research and Risk Assessment* 29(8):2037-2046.
- Lipfert, F. W. (1993). A critical review of studies of the association between demands for hospital services and air pollution. *Environmental Health Perspectives* 101(Suppl 2):229.
- Liu, Y. et Wu, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of nonparametric statistics* 23(2):415-437.
- Lloyd-Jones, D., Adams, R. J., Brown, T. M., Carnethon, M., Dai, S., De Simone, G., Ferguson, T. B., Ford, E., Furie, K., Gillespie, C., Go, A., Greenlund, K., Haase, N., Hailpern, S., Ho, P. M., Howard, V., Kissela, B., Kittner, S., Lackland, D., Lisabeth, L., Marelli, A., McDermott, M. M., Meigs, J., Mozaffarian, D., Mussolino, M., Nichol, G., Roger, V. L., Rosamond, W., Sacco, R., Sorlie, P., Roger, V. L., Thom, T., Wasserthiel-Smoller, S., Wong, N. D. et Wylie-Rosett, J. (2010). Heart disease and stroke statistics—2010 update: a report from the American Heart Association. *Circulation* 121(7):e46-e215.
- Longin, F. (2005). The choice of the distribution of asset returns: How extreme value theory can help? *J Bank Financ* 29(4):1017-1035.
- Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *J Bank Financ* 24(7):1097-1130.
- Lumley, T., Diehr, P., Emerson, S. et Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual review of public health* 23(1):151-169.
- MacNaughton, P., Adamkiewicz, G., Arku, R. E., Vallarino, J. et Levy, D. E. (2016). The impact of a smoke-free policy on environmental tobacco smoke exposure in public housing developments. *The Science of the total environment* 557-558:676-680.

- Madrigano, J., Mittleman, M. A., Baccarelli, A., Goldberg, R., Melly, S., von Klot, S. et Schwartz, J. (2013). Temperature, myocardial infarction, and mortality: effect modification by individual- and area-level characteristics. *Epidemiology* 24(3):439-446.
- Mailhot, A., Duchesne, S., Caya, D. et Talbot, G. (2007). Assessment of future change in intensity–duration–frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM). *Journal of hydrology* 347(1):197-210.
- Marrie, R. A., Dawson, N. V. et Garland, A. (2009). Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables. *Journal of clinical epidemiology* 62(5):511-517. e511.
- Martel, B., Ouarda, T. B. M. J., Vanasse, A., Chebana, F., Orzanco, M. G., Charron, C., Courteau, J., Asghari, S. et Gosselin, P. (2010) *Modélisation des fractures de la hanche en fonction de variables météorologiques à l'aide du modèle GAM dans deux régions métropolitaines du Québec, Canada*. INRS, Centre Eau, Terre et Environnement. 256 p
- Martens, W. J. (1998). Climate change, thermal stress and mortality changes. *Soc Sci Med* 46(3):331-344.
- Masselot, P., Chebana, F., Bélanger, D., St-Hilaire, A., Abdous, B., Gosselin, P. et Ouarda, T. B. M. J. (2015) *Régression EMD avec application à la relation entre les maladies cardiovasculaires et le climat*. INRS, Centre Eau Terre Environnement. 181 p
- Mathes, R. W., Ito, K. et Matte, T. (2011). Assessing syndromic surveillance of cardiovascular outcomes from emergency department chief complaint data in New York City. *PLoS One* 6(2):e14677.
- McNamee, R. (2005). Regression modelling and other methods to control confounding. *Occup Environ Med* 62(7):500-506, 472.
- McNeil, A. J. (1998) *Calculating quantile risk measures for financial return series using extreme value theory*. Department of Mathematics, ETH, Swiss Federal Technical University, Zurich
- McNeil, A. J. et Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance* 7(3):271-300.
- Medina-Ramon, M. et Schwartz, J. (2007). Temperature, temperature extremes, and mortality: a study of acclimatisation and effect modification in 50 US cities. *Occup Environ Med* 64(12):827-833.
- Ministère de la Santé et des Services sociaux du Québec (2016) *Prévenir les effets de la chaleur accablante et extrême*. <http://sante.gouv.qc.ca/conseils-et-prevention/prevenir-les-effets-de-la-chaleur-accablante-et-extreme/> (Consulté le 17-01-2017)
- Ministère des Affaires municipales et de l'Occupation du territoire (2016) *Communauté métropolitaine de Québec*. http://www.mamrot.gouv.qc.ca/pub/organisation_municipale/cartotheque/CMQuebec.pdf (Consulté le 20-01-2017)
- Mitov, K. et Nadarajah, S. (2006). Limit distributions for the bivariate geometric maxima. *Extremes* 8(4):357-370.

- Modarres, R., Ouarda, T. B., Vanasse, A., Orzanco, M. G. et Gosselin, P. (2012). Modeling seasonal variation of hip fracture in Montreal, Canada. *Bone* 50(4):909-916.
- Morabito, M., Modesti, P. A., Cecchi, L., Crisci, A., Orlandini, S., Maracchi, G. et Gensini, G. F. (2005). Relationships between weather and myocardial infarction: a biometeorological approach. *International journal of cardiology* 105(3):288-293.
- Morris, R. D. (2001). Airborne particulates and hospital admissions for cardiovascular disease: a quantitative review of the evidence. *Environ Health Perspect* 109 Suppl 4(Suppl 4):495-500.
- Morss, R. E., Wilhelmi, O. V., Meehl, G. A. et Dilling, L. (2011). Improving societal outcomes of extreme weather in a changing climate: an integrated perspective. *Annual Review of Environment and Resources* 36:1-25.
- Mostofsky, E., Schwartz, J., Coull, B. A., Koutrakis, P., Wellenius, G. A., Suh, H. H., Gold, D. R. et Mittleman, M. A. (2012). Modeling the association between particle constituents of air pollution and health outcomes. *American journal of epidemiology*.
- Mourtzoukou, E. G. et Falagas, M. E. (2007). Exposure to cold and respiratory tract infections [Review Article]. *The International Journal of Tuberculosis and Lung Disease* 11(9):938-943.
- Mustafic, H., Jabre, P., Caussin, C., Murad, M. H., Escolano, S., Tafflet, M., Perier, M. C., Marijon, E., Vernerey, D., Empana, J. P. et Jouven, X. (2012). Main air pollutants and myocardial infarction: a systematic review and meta-analysis. *JAMA* 307(7):713-721.
- Nadarajah, S. et Mitov, K. (2004). Extremal limit laws for discrete random variables. *Journal of Mathematical Sciences* 122(4):3404-3415.
- Nazerian, P., Vanni, S., Zanobetti, M., Polidori, G., Pepe, G., Federico, R., Cangioli, E. et Grifoni, S. (2010). Diagnostic accuracy of emergency Doppler echocardiography for identification of acute left ventricular heart failure in patients with acute dyspnea: comparison with Boston criteria and N-terminal prohormone brain natriuretic peptide. *Acad Emerg Med* 17(1):18-26.
- Nelder, J. A. et Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)* 135(3):370-384.
- Organisation mondiale de la santé (2015) *Maladies cardiovasculaires*. <http://www.who.int/mediacentre/factsheets/fs317/fr/> (Consulté le 17-01-2017)
- Ouarda, T. B. M. J., Charron, C., Shin, J. Y., Marpu, P. R., Al-Mandoos, A. H., Al-Tamimi, M. H., Ghedira, H. et Al Hosary, T. N. (2015). Probability distributions of wind speed in the UAE. *Energ Convers Manage* 93:414-434.
- Ouarda, T. B. M. J., Cunderlik, J. M., St-Hilaire, A., Barbet, M., Bruneau, P. et Bobee, B. (2006). Data-based comparison of seasonality-based regional flood frequency methods. *Journal of Hydrology* 330(1-2):329-339.
- Ouranos (2015) Vers l'adaptation. Synthèse des connaissances sur les changements climatiques au Québec. Édition 2015. (Ouranos, Montréal, Québec), p 415.
- Padoan, S. A., Ribatet, M. et Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* 105(489):263-277.

- Patz, J. A., Campbell-Lendrum, D., Holloway, T. et Foley, J. A. (2005). Impact of regional climate change on human health. *Nature* 438(7066):310-317.
- Pauli, F. et Rizzi, L. (2006). Statistical analysis of temperature impact on daily hospital admissions: analysis of data from Udine, Italy. *Environmetrics* 17(1):47-64.
- Phung, D., Guo, Y., Thai, P. K., Rutherford, S., Wang, X., Nguyen, M., Do, C. M., Nguyen, N. H., Alam, N. et Chu, C. (2016a). The effects of high temperature on cardiovascular admissions in the most populous tropical city in Vietnam. *Environmental Pollution* 208:33-39.
- Phung, D., Thai, P. K., Guo, Y., Morawska, L., Rutherford, S. et Chu, C. (2016b). Ambient temperature and risk of cardiovascular hospitalization: An updated systematic review and meta-analysis. *The Science of the total environment* 550:1084-1102.
- Pope, C. A., 3rd et Dockery, D. W. (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. *Journal of the Air & Waste Management Association* 56(6):709-742.
- Portier, C., Thigpen-Tart, K., Hess, J., Lubner, G., Maslak, T. et Radtke, M. (2010). A human health perspective on climate change. *Research Triangle Park, NC: Environmental Health Perspectives, National Institute of Environmental Health Sciences*.
- Pourhoseingholi, M. A., Baghestani, A. R. et Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology from bed to bench* 5(2):79.
- Puli, V., Eterovi, D., Miri, D., Giunio, L., Lukin, A. et Fabijani, D. (2004). Triggering of ventricular tachycardia by meteorologic and emotional stress: protective effect of β -blockers and anxiolytics in men and elderly. *American journal of epidemiology* 160(11):1047-1058.
- R Core Team (2016) R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria).
- Reiss, R. D. et Thomas, M. (2007) Statistical Analysis of Extreme Values. (Birkhauser, Basel).
- Ren, C., O'Neill, M. S., Park, S. K., Sparrow, D., Vokonas, P. et Schwartz, J. (2011). Ambient temperature, air pollution, and heart rate variability in an aging population. *Am J Epidemiol* 173(9):1013-1021.
- Ren, C., Williams, G. M. et Tong, S. (2006). Does particulate matter modify the association between temperature and cardiorespiratory diseases? *Environmental health perspectives* 114(11):1690.
- Renard, B. et Lang, M. (2007). Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Advances in Water Resources* 30(4):897-912.
- Robine, J. M., Cheung, S. L., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J. P. et Herrmann, F. R. (2008). Death toll exceeded 70,000 in Europe during the summer of 2003. *C R Biol* 331(2):171-178.
- Rocklov, J. et Forsberg, B. (2008). The effect of temperature on mortality in Stockholm 1998--2003: a study of lag structures and heatwave effects. *Scand J Public Health* 36(5):516-523.
- Rousseeuw, P. J. et Hubert, M. (1999). Regression Depth. *Journal of the American Statistical Association* 94(446):388-402.

- Ryti, N. R., Guo, Y. et Jaakkola, J. J. (2016). Global Association of Cold Spells and Adverse Health Effects: A Systematic Review and Meta-Analysis. *Environ Health Perspect* 124(1):12-22.
- Salvadori, G. (2007) *Extremes in nature: an approach using copulas*. Springer,
- Salvadori, G., De Michele, C. et Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrol Earth Syst Sc* 15(11):3293-3305.
- Schuld, J., Kollmar, O., Schuld, S., Schommer, K. et Richter, S. (2013). Impact of meteorological conditions on abdominal aortic aneurysm rupture: evaluation of an 18-year period and review of the literature. *Vasc Endovascular Surg* 47(7):524-531.
- Schwartz, J. (1997). Air pollution and hospital admissions for cardiovascular disease in Tucson. *Epidemiology* 8(4):371-377.
- Schwartz, J. (1999). Air pollution and hospital admissions for heart disease in eight US counties. *Epidemiology* 10(1):17-22.
- Schwartz, J. (2000). The Distributed Lag between Air Pollution and Daily Deaths. *Epidemiology* 11(3):320-326.
- Schwartz, J., Samet, J. M. et Patz, J. A. (2004). Hospital admissions for heart disease: the effects of temperature and humidity. *Epidemiology* 15(6):755-761.
- Sharma, A. J., Weiss, E. C., Young, S. L., Stephens, K., Ratard, R., Straif-Bourgeois, S., Sokol, T. M., Vranken, P. et Rubin, C. H. (2008). Chronic disease and related conditions at emergency treatment facilities in the New Orleans area after Hurricane Katrina. *Disaster medicine and public health preparedness* 2(01):27-32.
- Sheridan, S. C. (2002). The redevelopment of a weather-type classification scheme for North America. *International Journal of Climatology* 22(1):51-68.
- Siciliani, L., Sivey, P. et Street, A. (2013). Differences in Length of Stay for Hip Replacement between Public Hospitals, Specialised Treatment Centres and Private Providers: Selection or Efficiency? *Health Econ* 22(2):234-242.
- Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W. et Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres* 118(4):1716-1733.
- Singh, V. (1998) *Entropy-based parameter estimation in hydrology*. Springer Science & Business Media,
- Siontis, G. C., Tzoulaki, I., Siontis, K. C. et Ioannidis, J. P. (2012). Comparisons of established risk prediction models for cardiovascular disease: systematic review. *Bmj* 344:e3318.
- Smith, R. L. (1987). Estimating Tails of Probability-Distributions. *Annals of Statistics* 15(3):1174-1207.
- Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* 4(4):367-377.
- Smoyer-Tomic, K. E., Kuhn, R. et Hudson, A. (2003). Heat wave hazards: an overview of heat wave impacts in Canada. *Natural hazards* 28(2-3):465-486.

- Smoyer-Tomic, K. E. et Rainham, D. G. (2001). Beating the heat: development and evaluation of a Canadian hot weather health-response plan. *Environ Health Perspect* 109(12):1241-1248.
- Soyiri, I. N. et Reidpath, D. D. (2013). The Use of Quantile Regression to Forecast Higher Than Expected Respiratory Deaths in a Daily Time Series: A Study of New York City Data 1987-2000. *Plos One* 8(10):e78215.
- Soyiri, I. N., Reidpath, D. D. et Sarran, C. (2013). Forecasting asthma-related hospital admissions in London using negative binomial models. *Chronic respiratory disease* 10(2):85-94.
- Spickett, J. T., Brown, H. L. et Rumchev, K. (2011). Climate change and air quality: the potential impact on health. *Asia Pac J Public Health* 23(2 Suppl):37S-45.
- Stephens, M. A. (1977). Goodness of fit for the extreme value distribution. *Biometrika* 64(3):583-588.
- Stewart, S., Ahamed, Y., Wiley, J. F., McDermott, C. J., Ball, J., Keates, A. K., Løchen, M.-L. et Carrington, M. J. (2016) Seasonal variations in cardiovascular-related mortality but not hospitalization are modulated by temperature and not climate type: a systematic review and meta-analysis of 4.5 million events in 26 countries. (Am Heart Assoc).
- Stieb, D. M., Burnett, R. T., Beveridge, R. C. et Brook, J. R. (1996). Association between ozone and asthma emergency department visits in Saint John, New Brunswick, Canada. *Environmental Health Perspectives* 104(12):1354-1360.
- Stott, P. A., Stone, D. A. et Allen, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature* 432(7017):610-614.
- Sullivan, G. M. et Feinn, R. (2012). Using Effect Size-or Why the P Value Is Not Enough. *J Grad Med Educ* 4(3):279-282.
- Sun, W., Wang, J. H. et Fang, Y. X. (2013). Consistent Selection of Tuning Parameters via Variable Selection Stability. *J Mach Learn Res* 14(1):3419-3440.
- Taylor, F., Ward, K., Moore, T. H., Burke, M., Davey Smith, G., Casas, J. P. et Ebrahim, S. (2011). Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 10.1002/14651858.CD004816.pub4(1):CD004816.
- Thomas, M., Lemaitre, M., Wilson, M. L., Viboud, C., Yordanov, Y., Wackernagel, H. et Carrat, F. (2016). Applications of Extreme Value Theory in Public Health. *PLoS One* 11(7):e0159312.
- Tilaki, K. H. (2012). Methodological issues of confounding in analytical epidemiologic studies. *Caspian journal of internal medicine* 3(3):488.
- Towler, E., Rajagopalan, B., Gilleland, E., Summers, R. S., Yates, D. et Katz, R. W. (2010). Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research* 46(11):W11504.
- Tu, J. V., Nardi, L., Fang, J., Liu, J., Khalid, L., Johansen, H. et Canadian Cardiovascular Outcomes Research, T. (2009). National trends in rates of death and hospital admissions related to acute myocardial infarction, heart failure and stroke, 1994-2004. *CMAJ* 180(13):E118-125.

- Tukey, J. W. (1975) Mathematics and the picturing of data. *Proceedings of the international congress of mathematicians*. p 523-531.
- Turner, R. M., Muscatello, D. J., Zheng, W., Willmore, A. et Arendts, G. (2007). An outbreak of cardiovascular syndromes requiring urgent medical treatment and its association with environmental factors: an ecological study. *Environ Health* 6:37.
- Urban, A. et Kysely, J. (2014). Comparison of UTCI with other thermal indices in the assessment of heat and cold effects on cardiovascular mortality in the Czech Republic. *Int J Environ Res Public Health* 11(1):952-967.
- Van Kempen, E. E. M. M., Kruize, H., Boshuizen, H. C., Ameling, C. B., Staatsen, B. A. M. et de Hollander, A. E. M. (2002). The association between noise exposure and blood pressure and ischemic heart disease: a meta-analysis. *Environmental health perspectives* 110(3):307.
- Vanasse, A., Cohen, A., Courteau, J., Bergeron, P., Dault, R., Gosselin, P., Blais, C., Bélanger, D., Rochette, L. et Chebana, F. (2016). Association between Floods and Acute Cardiovascular Diseases: A Population-Based Cohort Study Using a Geographic Information System Approach. *International journal of environmental research and public health* 13(2):168.
- Vanasse, A., Talbot, D., Chebana, F., Belanger, D., Blais, C., Gamache, P., Giroux, J. X., Dault, R. et Gosselin, P. (2017). Effects of climate and fine particulate matter on hospitalizations and deaths for heart failure in elderly: A population-based cohort study. *Environ Int* 106:257-266.
- Vandentorren, S., Suzan, F., Medina, S., Pascal, M., Maulpoix, A., Cohen, J. C. et Ledrans, M. (2004). Mortality in 13 French cities during the August 2003 heat wave. *American journal of public health* 94(9):1518-1520.
- Vanos, J. K., Hebborn, C. et Cakmak, S. (2014). Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 Canadian cities. *Environmental Pollution* 185:322-332.
- Vasconcellos, F., Seabra, A., Katzmarzyk, P. T., Kraemer-Aguiar, L. G., Bouskela, E. et Farinatti, P. (2014). Physical activity in overweight and obese adolescents: systematic review of the effects on physical fitness components and cardiovascular risk factors. *Sports Med* 44(8):1139-1152.
- Wang, X., Ouyang, Y., Liu, J., Zhu, M., Zhao, G., Bao, W. et Hu, F. B. (2014). Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. *Bmj* 349:g4490.
- Watts, K. A., Dupuis, D. J. et Jones, B. L. (2006). An extreme value analysis of advanced age mortality data. *North American Actuarial Journal* 10(4):162-178.
- Watts, N., Adger, W. N., Agnozzi, P., Blackstock, J., Byass, P., Cai, W., Chaytor, S., Colbourn, T., Collins, M., Cooper, A., Cox, P. M., Depledge, J., Drummond, P., Ekins, P., Galaz, V., Grace, D., Graham, H., Grubb, M., Haines, A., Hamilton, I., Hunter, A., Jiang, X., Li, M., Kelman, I., Liang, L., Lott, M., Lowe, R., Luo, Y., Mace, G., Maslin, M., Nilsson, M., Oreszczyn, T., Pye, S., Quinn, T., Svendsdotter, M., Venevsky, S., Warner, K., Xu, B., Yang, J., Yin, Y., Yu, C., Zhang, Q., Gong, P., Montgomery, H. et Costello, A. (2015).

- Health and climate change: policy responses to protect public health. *Lancet* 386(10006):1861-1914.
- Waylen, P., Keellings, D. et Qiu, Y. (2012). Climate and health in Florida: Changes in risks of annual maximum temperatures in the second half of the twentieth century. *Applied Geography* 33:73-81.
- Weber, M. A. et Fodera, S. M. (2004). Circadian variations in cardiovascular disease: chronotherapeutic approaches to the management of hypertension. *Rev Cardiovasc Med* 5(3):148-155.
- Weir, E. (2002). Heat wave: first, protect the vulnerable. *CMAJ* 167(2):169.
- Weisskopf, M. G., Anderson, H. A., Foldy, S., Hanrahan, L. P., Blair, K., Torok, T. J. et Rumm, P. D. (2002). Heat wave morbidity and mortality, Milwaukee, Wis, 1999 vs 1995: an improved response? *American journal of public health* 92(5):830-833.
- Wholey, M. H., Al-Mubarek, N. et Wholey, M. H. (2003). Updated review of the global carotid artery stent registry. *Catheter Cardiovasc Interv* 60(2):259-266.
- Wolf, K., Schneider, A., Breitner, S., von Klot, S., Meisinger, C., Cyrys, J., Hymer, H., Wichmann, H. E., Peters, A. et Cooperative Health Research in the Region of Augsburg Study, G. (2009). Air temperature and the occurrence of myocardial infarction in Augsburg, Germany. *Circulation* 120(9):735-742.
- Wood, S. (2006) *Generalized additive models: an introduction with R*. CRC press,
- Yang, C., Meng, X., Chen, R., Cai, J., Zhao, Z., Wan, Y. et Kan, H. (2015a). Long-term variations in the association between ambient temperature and daily cardiovascular mortality in Shanghai, China. *Science of the Total Environment* 538:524-530.
- Yang, J., Yin, P., Zhou, M., Ou, C. Q., Guo, Y., Gasparrini, A., Liu, Y., Yue, Y., Gu, S., Sang, S., Luan, G., Sun, Q. et Liu, Q. (2015b). Cardiovascular mortality risk attributable to ambient temperature in China. *Heart* 101(24):1966-1972.
- Yang, Q., Cogswell, M. E., Flanders, W. D., Hong, Y., Zhang, Z., Loustalot, F., Gillespie, C., Merritt, R. et Hu, F. B. (2012). Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. *Jama* 307(12):1273-1283.
- Yang, T. C., Matthews, S. A. et Chen, V. Y. (2014). Stochastic variability in stress, sleep duration, and sleep quality across the distribution of body mass index: insights from quantile regression. *Int J Behav Med* 21(2):282-291.
- Ye, X., Wolff, R., Yu, W., Vaneckova, P., Pan, X. et Tong, S. (2012). Ambient Temperature and Morbidity: A Review of Epidemiological Evidence. *Environ Health Perspect* 120.
- Yee, T. W. (2015) *Vector Generalized Linear and Additive Models: With an Implementation in R*. Springer, New York
- Yu, K. M., Lu, Z. D. et Stander, J. (2003). Quantile regression: applications and current research areas. *J Roy Stat Soc D-Stat* 52(3):331-350.
- Zanobetti, A., O'Neill, M. S., Gronlund, C. J. et Schwartz, J. D. (2013). Susceptibility to mortality in weather extremes: effect modification by personal and small area characteristics in a multi-city case-only analysis. *Epidemiology (Cambridge, Mass.)* 24(6):809.

- Zanobetti, A. et Peters, A. (2015). Disentangling interactions between atmospheric pollution and weather. *Journal of epidemiology and community health* 69(7):613-615.
- Zanobetti, A., Schwartz, J. et Dockery, D. W. (2000). Airborne particles are a risk factor for hospital admissions for heart and lung disease. *Environ Health Perspect* 108(11):1071-1077.
- Zeger, S. L., Irizarry, R. et Peng, R. D. (2006). On time series analysis of public health and biomedical data. *Annual review of public health* 27:57-79.
- Zhang, X., Zhou, Y., Tian, Y. et Zhang, S. (2016). Analysis of the Influence of Heat Wave on Death among the Elderly in Nanjing City. *Journal of Geoscience and Environment Protection* 4(11):62.
- Zuo, Y. J. et Serfling, R. (2000). General notions of statistical depth function. *Annals of Statistics* 28(2):461-482.
- Zwiers, F. W., Alexander, L. V., Hegerl, G. C., Knutson, T. R., Kossin, J. P., Naveau, P., Nicholls, N., Schär, C., Seneviratne, S. I. et Zhang, X. (2013) Climate extremes: challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events. *Climate Science for Serving Society*, Springer. p 339-389.