

Centre Eau Terre Environnement

IMPACTS SANITAIRES ET ÉCONOMIQUES DE LA CHALEUR AU QUÉBEC AVEC APPROCHES DE LA SCIENCE DES DONNÉES

Par
Jérémie Boudreault

Thèse présentée pour l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en santé environnementale et science des données

Jury d'évaluation

Président du jury et examineur interne	Amadou Diogo Barry Institut national de la recherche scientifique
Examinatrice externe	Audrey Smargiassi École de santé publique de l'Université de Montréal (ESPUM)
Examineur externe	Karim Barigou Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain
Directeur de recherche	Fateh Chebana Institut national de la recherche scientifique
Codirectrice de recherche	Céline Campagna Ouranos, INRS et Université Laval

REMERCIEMENTS

D'abord, je tiens à remercier mon directeur de recherche, Fateh Chebana. Cela fait bientôt dix ans que tu es un mentor d'exception pour moi et ma carrière. Au-delà de m'avoir montré la voie pour devenir un chercheur rigoureux travaillant en multidisciplinarité, tu as surtout fait de moi une personne meilleure, plus à l'écoute des autres. Je me sens choyé d'avoir pu compter sur tes sages conseils pendant aussi longtemps. Je nous souhaite encore de belles années de collaborations et d'amitiés.

Je tiens aussi à remercier sincèrement ma codirectrice, Céline Campagna. Tu m'as ouvert toutes grandes les portes de la santé publique du Québec, m'as initié à ce merveilleux monde et m'as fait rencontrer les bonnes personnes à un moment crucial de ma jeune carrière. Merci de m'avoir fait confiance dès le début, alors qu'on se connaissait à peine. J'espère que nos collaborations, mais aussi nos discussions autour d'un café à Limoilou, se poursuivront encore longtemps.

En plus de mon équipe officielle de supervision, j'ai pu compter pour ce projet sur plusieurs autres mentors dans les organismes de santé à travers le Canada. À l'Institut national de santé publique du Québec (INSPQ), je remercie spécialement Félix Lamothe d'avoir toujours vu la pertinence de mes travaux et les avoir supportés de toutes les façons possibles. Mon seul regret est de ne pas être allé à Montréal plus souvent afin de t'y voir. Je souligne aussi l'apport essentiel à ce projet d'Éric Lavigne de Santé Canada, particulièrement tes connaissances poussées en épidémiologie. J'espère vivement que l'on continuera de collaborer et de se voir lors de tes sorties en ski à Québec. Finalement, je remercie Sarah Henderson et Kathleen McLean de m'avoir accueilli lors d'un stage au *BC Center for Disease Control* et à l'Université de Colombie-Britannique. J'ai beaucoup appris grâce à vous deux et j'espère retourner bientôt à Vancouver afin de poursuivre nos collaborations, mais aussi afin d'y faire encore plus de randonnées.

Je tiens à remercier tous mes collègues et amis à l'INRS et à l'INSPQ qui ont pimenté le travail d'activités sociales qui étaient plus que nécessaires. Par le fait même, je veux souligner le travail de tous les employés du Centre Eau Terre Environnement et de l'administration de l'INRS d'avoir supporté mon projet de doctorat sur mesure, ainsi que toutes mes autres initiatives sociales, scientifiques et environnementales. Merci aussi aux stagiaires qui ont travaillé avec moi au cours des dernières années et qui ont contribué au succès de ces recherches.

Je me dois de souligner l'apport essentiel du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), des Instituts de recherche en santé du Canada (IRSC), d'Ouranos

et de l'INSPQ pour le support financier tout au long du doctorat. J'espère que vous serez satisfaits par les retombées de ces travaux. Je remercie grandement Amadou Barry, Audrey Smargiassi et Karim Barigou d'avoir accepté de réviser les quelque 200 pages que constitue cette thèse.

Finalement, je remercie mes amis et ma famille pour tout le support au cours des dernières années. Mon frère, Simon, tu es mon modèle d'excellence en recherche depuis le début de ma jeune carrière, et tu continueras de l'être encore de nombreuses années. Mes parents, Gaétan et Marlène, vous m'avez offert un environnement familial exceptionnel où toutes les conditions étaient réunies pour que je devienne la personne que je suis aujourd'hui. Ma conjointe, meilleure amie et mère de notre enfant, Andréa, tu as toujours été présente pour moi et m'as soutenu sans rien demander en retour. Tu mérites autant que moi de partager les fruits de cet accomplissement. Louis, sache que j'ai mis toutes mes énergies dans ce projet en pensant à toi, alors que tu n'étais encore qu'au stade de projet. J'espère avoir pu contribuer, grâce à mes recherches, à t'offrir un avenir un tant soit peu meilleur. Pour terminer, je souligne l'effort de notre chatte Cannelle d'avoir tenté de me supporter, à sa manière, durant toutes ces années.

RÉSUMÉ

Les changements climatiques augmentent l'intensité, la fréquence et la durée des épisodes de chaleur. Ces événements ont des impacts importants sur la santé de la population au Québec comme ailleurs dans le monde. Des méthodes de la science des données peuvent être mises à profit afin de mieux modéliser les impacts sanitaires de la chaleur, ainsi que pour projeter les coûts sanitaires qui y sont associés. Dans le premier objectif de cette thèse, des modèles d'apprentissage automatique et profond sont utilisés afin de prévoir les impacts à court terme de la chaleur sur la santé. Ces modèles sont comparés avec des approches statistiques couramment utilisées dans le domaine. Le second objectif de la thèse vise à calculer le fardeau sanitaire de la chaleur au Québec pour plusieurs issues sanitaires incluant la mortalité et plusieurs variables de morbidité. Des modèles statistiques de type *Distributed Lag Non-Linear Models* (DLNM) sont mis à profit à cette fin, comme ils sont les plus utilisés pour ce faire dans la littérature en épidémiologie environnementale. Les DLNM développés sont ensuite réutilisés pour le troisième objectif de la thèse visant à quantifier les coûts sanitaires historiques et futurs de la chaleur au Québec. Les résultats du premier objectif montrent que, selon le contexte, les approches d'apprentissage automatique de type *gradient boosting*, les réseaux de neurones profonds ainsi que les modèles statistiques non linéaires sont tous pertinents pour modéliser les impacts sanitaires de la chaleur. Au second objectif, les DLNM permettent de calculer un fardeau sanitaire attribuable à la chaleur de 470 décès, 225 hospitalisations, 36 000 visites à l'urgence, 7200 transports ambulanciers et 15 000 appels à Info-Santé chaque été en moyenne au Québec. Finalement, au troisième objectif, les coûts totaux historiques de la chaleur ont été estimés à 3.6 milliards de dollars annuellement. À l'horizon 2050, les coûts totaux devraient être multipliés par 3 selon un scénario climatique et démographique médian, mais par 5 selon un scénario plus pessimiste, le tout en ne considérant aucune adaptation supplémentaire à la chaleur. En somme, cette thèse a permis de démontrer le potentiel des méthodes de la science des données, incluant l'apprentissage automatique et l'épidémiologie environnementale, pour développer des outils et générer des données probantes pour s'adapter à la chaleur dans un contexte de changements climatiques. Les autorités gouvernementales et sanitaires pourront utiliser ces résultats pour accélérer l'action climatique et aussi les étendre à d'autres problématiques environnementales ou sanitaires.

Mots-clés : température, vague de chaleur, chaleur extrême, gradient boosting, long short-term memory, distributed lag non-linear model, mortalité, morbidité, évaluation économique, changement climatique.

ABSTRACT

Climate change increases the intensity, frequency and duration of heat episodes. These events have significant impacts on population health in Quebec and around the world. Data science methods can be leveraged to better model the heat-related health impacts and to project the associated health costs. For the first objective of this thesis, machine and deep learning models are introduced to predict the short-term health impacts of heat. These models are compared to statistical approaches commonly used in the field. The second objective aims to calculate the heat-related health burden in the province of Quebec for multiple health outcomes, including mortality and several morbidity variables. Distributed Lag Non-Linear Models (DLNM) are leveraged as they are the most commonly used approach in the environmental epidemiology literature to that end. The developed DLNMs are then reused in the third objective aimed at quantifying the historical and future heat-related health costs in Quebec. The results for the first objective show that, depending on the context, machine learning approaches such as gradient boosting, deep neural networks and non-linear statistical models are all relevant for modelling the health impacts of heat. In the second objective, DLNMs allow calculating a health burden attributable to heat of 470 deaths, 225 hospitalizations, 36 000 emergency department visits, 7200 ambulance transports and 15 000 calls to the Info-Santé health hotline each summer in average in Quebec. Finally, in the third objective, the total historical heat-related health costs were estimated at \$3.6 billion annually. By 2050, total costs are expected to increase 3-fold according to a median climate and demographic scenario, and 5-fold according to a more pessimistic scenario, but without considering any additional adaptation to heat. In summary, this thesis demonstrated the potential of data science methods, including machine learning and environmental epidemiology models, to develop cutting-edge tools and generate new evidence for heat adaptation in the context of climate change. Health and governmental authorities can use these results to accelerate climate action, as well as leverage them for other environmental or health issues.

Keywords : temperature, heatwave, extreme heat, gradient boosting, long short-term memory, distributed lag non-linear model, mortality, morbidity, economic evaluation, climate change.

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ.....	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES FIGURES	XIII
LISTE DES TABLEAUX.....	XV
LISTE DES ABRÉVIATIONS	XVII
LISTE DES CONTRIBUTIONS	XIX
1 INTRODUCTION	3
1.1 MISE EN CONTEXTE	3
1.2 REVUE DE LITTÉRATURE	6
1.2.1 <i>Effets sanitaires de la chaleur</i>	6
1.2.2 <i>Méthodes statistiques utilisées en épidémiologie environnementale</i>	7
1.2.3 <i>Méthodes d'apprentissage automatique</i>	8
1.2.4 <i>Fardeau sanitaire de la chaleur</i>	10
1.2.5 <i>Coûts sanitaires de la chaleur</i>	12
1.3 PROBLÉMATIQUES.....	14
1.3.1 <i>Limites des applications actuelles de l'apprentissage automatique</i>	14
1.3.2 <i>Limites des études sur le fardeau sanitaire de la chaleur</i>	15
1.3.3 <i>Limites des études sur les coûts de santé de la chaleur</i>	16
1.4 OBJECTIFS DE LA THÈSE ET RÉSUMÉ DE LA MÉTHODOLOGIE	18
1.4.1 <i>Objectif 1 : Modélisation des impacts à la santé de la chaleur avec apprentissage automatique et profond</i>	19
1.4.2 <i>Objectif 2 : Quantification du fardeau sanitaire de la chaleur</i>	24
1.4.3 <i>Objectif 3 : Estimation et projection des coûts sanitaires de la chaleur</i>	25
1.5 STRUCTURE DE LA THÈSE ET LIENS ENTRE LES CHAPITRES	27
2 MACHINE AND DEEP LEARNING FOR MODELLING HEAT-HEALTH RELATIONSHIPS.....	33
2.1 INTRODUCTION	37
2.2 MATERIALS AND METHODS	39
2.2.1 <i>Data sources</i>	39
2.2.2 <i>Tree-based methods</i>	42
2.2.3 <i>Neural networks</i>	43
2.2.4 <i>Statistical models</i>	44

2.2.5	<i>Models' evaluation</i>	45
2.3	RESULTS	45
2.3.1	<i>Tree-based methods</i>	45
2.3.2	<i>Neural networks</i>	47
2.3.3	<i>Statistical models</i>	49
2.3.4	<i>Models' comparison</i>	49
2.3.5	<i>Models' performance</i>	51
2.4	DISCUSSION	53
2.5	CONCLUSION	56
3	REVISITING THE IMPORTANCE OF TEMPERATURE, WEATHER AND AIR POLLUTION VARIABLES IN HEAT- MORTALITY RELATIONSHIPS WITH MACHINE LEARNING	63
3.1	INTRODUCTION	66
3.2	MATERIALS AND METHODS	67
3.2.1	<i>Mortality data</i>	67
3.2.2	<i>Temperature, weather and air pollution data</i>	68
3.2.3	<i>Machine learning (ML) and statistical models</i>	70
3.2.4	<i>Models' and predictors' evaluation</i>	72
3.3	RESULTS	73
3.3.1	<i>Models using temperature indices only</i>	73
3.3.2	<i>Models using temperature and other heat-related predictors</i>	75
3.3.3	<i>Feature importance in models with all predictors</i>	76
3.3.4	<i>Results summary and comparison</i>	77
3.4	DISCUSSION	78
3.5	CONCLUSION	81
4	MULTI-REGION MODELS BUILT WITH MACHINE AND DEEP LEARNING FOR PREDICTING SEVERAL HEAT-RELATED HEALTH OUTCOMES	87
4.1	INTRODUCTION	90
4.2	MATERIAL AND METHODS	91
4.2.1	<i>Study design and health outcomes</i>	91
4.2.2	<i>Weather, regional and temporal predictors</i>	93
4.2.3	<i>Machine learning models</i>	96
4.2.4	<i>Models' performance and explainability</i>	97
4.3	RESULTS	98
4.3.1	<i>Performance across all regions</i>	99
4.3.2	<i>Performance by region</i>	100
4.3.3	<i>Validation during heatwaves</i>	101
4.3.4	<i>Models' explainability</i>	103

4.4	DISCUSSION	105
4.5	CONCLUSION	109
5	ESTIMATING THE HEAT-RELATED MORTALITY AND MORBIDITY BURDEN IN THE PROVINCE OF QUEBEC, CANADA	115
5.1	INTRODUCTION	119
5.2	MATERIAL AND METHODS.....	120
5.2.1	<i>Study design and health outcomes</i>	<i>120</i>
5.2.2	<i>Weather, air pollution and regional data</i>	<i>121</i>
5.2.3	<i>Statistical analysis</i>	<i>123</i>
5.2.4	<i>Heat burden quantification.....</i>	<i>124</i>
5.3	RESULTS	125
5.3.1	<i>Descriptive statistics</i>	<i>125</i>
5.3.2	<i>Regional and pooled cumulative exposure-response curves</i>	<i>127</i>
5.3.3	<i>Heat burden in the province of Quebec.....</i>	<i>130</i>
5.3.4	<i>Temporal and spatial variability of the heat burden.....</i>	<i>131</i>
5.4	DISCUSSION	134
5.5	CONCLUSION	138
6	PROJECTING THE OVERALL HEAT-RELATED HEALTH BURDEN AND ASSOCIATED ECONOMIC COSTS IN A CLIMATE CHANGE CONTEXT IN QUEBEC, CANADA	143
6.1	INTRODUCTION	147
6.2	MATERIAL AND METHODS.....	148
6.2.1	<i>Study area</i>	<i>148</i>
6.2.2	<i>Statistical analysis of the association between heat and health outcomes</i>	<i>148</i>
6.2.3	<i>Climate and demographic data.....</i>	<i>149</i>
6.2.4	<i>Attributable numbers to heat and number of extreme heatwaves</i>	<i>150</i>
6.2.5	<i>Heat-related health costs.....</i>	<i>150</i>
6.3	RESULTS	151
6.3.1	<i>Climate and demographic evolution</i>	<i>151</i>
6.3.2	<i>Heat-related numbers.....</i>	<i>152</i>
6.3.3	<i>Health costs of heat exposure</i>	<i>153</i>
6.3.4	<i>Uncertainties in estimated costs.....</i>	<i>154</i>
6.3.5	<i>Breakdown of health costs</i>	<i>155</i>
6.4	DISCUSSION	158
6.5	CONCLUSION	161
7	CONCLUSIONS ET PERSPECTIVES	169
7.1	PRINCIPALES CONCLUSIONS	169
7.2	PRINCIPALES PERSPECTIVES	172

8 BIBLIOGRAPHIE	179
ANNEXE : DESCRIPTION DES MODÈLES UTILISÉS.....	203
A.1. MODÈLES STATISTIQUES UTILISÉS EN ÉPIDÉMIOLOGIE ENVIRONNEMENTALE	203
A.2. MODÈLES D'APPRENTISSAGE AUTOMATIQUE BASÉS SUR LES ARBRES	204
A.3 MODÈLES D'APPRENTISSAGE AUTOMATIQUE ET PROFOND BASÉ SUR LES RÉSEAUX DE NEURONES.....	207

LISTE DES FIGURES

FIGURE 1.1	RÉSUMÉ DES APPROCHES DE LA SCIENCE DES DONNÉES CONSIDÉRÉES DANS LE CADRE DE CETTE THÈSE.	5
FIGURE 1.2	APERÇU DES CHAPITRES 2 À 6 DE LA THÈSE ET DES LIENS ENTRE LES CHAPITRES	29
FIGURE 2.1	GRAPHICAL ABSTRACT.	36
FIGURE 2.2	WEATHER AND AIR POLLUTION STATIONS WITHIN THE CENSUS METROPOLITAN AREA (CMA) OF MONTREAL.	40
FIGURE 2.3	RESULTING PRUNED DECISION TREE (DT).....	46
FIGURE 2.4	FEATURE IMPORTANCE (FI) METRICS FOR A) RANDOM FOREST (RF) AND B) GRADIENT BOOSTING MACHINE (GBM)	47
FIGURE 2.5	FEATURE IMPORTANCE (FI) METRICS FOR A) SINGLE-LAYER PERCEPTRON (SLP), B) MULTI-LAYER PERCEPTRON (MLP) AND C) LONG SHORT-TERM MEMORY (LSTM)	48
FIGURE 2.6	PERFORMANCE OF ALL MODELS ON THE TEST SET (2014–2019) GIVEN BY A) ROOT MEAN SQUARE ERROR (RMSE), B) MEAN ABSOLUTE ERROR (MAE) AND C) COEFFICIENT OF DETERMINATION (R^2).....	51
FIGURE 2.7	WEEKLY MORTALITY DEVIATION PREDICTIONS FROM ALL MODELS AND OBSERVATIONS FOR THE 2010–2019 PERIOD.	53
FIGURE 3.1	WEATHER AND AIR POLLUTION STATIONS WITHIN THE TWO STUDIED CENSUS METROPOLITAN AREAS (CMA) OF A) MONTREAL AND B) QUEBEC.....	68
FIGURE 3.2	OVERVIEW OF THE FITTED MACHINE LEARNING (ML) AND STATISTICAL MODELS IN THIS STUDY. 72	
FIGURE 3.3	PERFORMANCE METRICS ON TEST SET BASED ON R^2 (Y-AXIS) AND DECREASE IN MEAN ABSOLUTE ERROR (dMAE) (X-AXIS) OF ALL MODELS USING DIFFERENT LAGGED TEMPERATURE INDICES IN THE CMAS OF A) MONTREAL AND B) QUEBEC.....	74
FIGURE 3.4	PERFORMANCE METRICS ON TEST SET BASED ON R^2 (Y-AXIS) AND DECREASE IN MEAN ABSOLUTE ERROR (dMAE) (X-AXIS) OF ALL MODELS USING MEAN TEMPERATURE AND OTHER HEAT-RELATED PREDICTORS IN THE CMAS OF A) MONTREAL AND B) QUEBEC	76
FIGURE 4.1	OVERVIEW OF THE METHODOLOGY.....	92
FIGURE 4.2	REGIONAL PERFORMANCE IN TERMS OF R^2 IN THE 15 REGIONS FOR ALL MODELS IN THE GENERAL POPULATION FOR A) MORTALITY, B) HOSPITALIZATIONS C) EMERGENCY DEPARTMENT VISITS, D) AMBULANCE TRANSPORTS AND E) 811 CALLS.....	101
FIGURE 4.3	OBSERVED (SOLID BLACK LINE) AND PREDICTED (DASHED COLOURED LINES) DAILY COUNTS OF A) MORTALITY, B) HOSPITALIZATIONS C) EMERGENCY DEPARTMENT VISITS, D) AMBULANCE TRANSPORTS AND E) 811 CALLS IN ALL STUDIED REGIONS FOR THE GENERAL POPULATION DURING I) 2018 AND II) 2021 HEATWAVES.....	103
FIGURE 5.1	MAP OF THE STUDIED HEALTH REGIONS (HR) IN THE PROVINCE OF QUEBEC (CANADA) AND LOCATION OF THE CONSIDERED AIR POLLUTION STATIONS OF THE NATIONAL AIR POLLUTION SURVEILLANCE (NAPS) PROGRAM.....	121
FIGURE 5.2	CUMULATIVE EXPOSURE-RESPONSE FUNCTIONS OVER 8 DAYS IN EACH HEALTH REGION (HR) FROM REGIONAL DLNM FOR A) MORTALITY, B) HOSPITALIZATIONS, C) ED VISITS, D) AMBULANCE TRANSPORTS AND E) 811 CALLS.	128
FIGURE 5.3	POOLED CUMULATIVE EXPOSURE-RESPONSE FUNCTIONS OVER 8 DAYS IN THE PROVINCE OF QUEBEC FOR A) MORTALITY, B) HOSPITALIZATIONS, C) ED VISITS, D) AMBULANCE TRANSPORTS AND E) 811 CALLS.....	129

FIGURE 5.4	ATTRIBUTABLE NUMBERS (AN) TO ALL HEAT AND EXTREME HEAT EXPOSURES DURING 1996–2019 BASED ON REGIONAL BLUP ACROSS QUEBEC.....	132
FIGURE 5.5	ATTRIBUTABLE FRACTION (AF) IN EACH HEALTH REGION (HR) COMPUTED FROM REGIONAL BLUP FOR A) ALL HEAT AND B) EXTREME HEAT EXPOSURES	133
FIGURE 6.1	GRAPHICAL ABSTRACT	146
FIGURE 6.2	EVOLUTION OF A) MEAN SUMMER TEMPERATURE (MAY TO SEPTEMBER) AND B) DEMOGRAPHY ACROSS QUEBEC BASED ON SSP2–4.5 AND SSP5–8.5 SCENARIOS OVER THE 1990–2070 PERIOD.....	152
FIGURE 6.3	RESULTS OF MONTE CARLO SIMULATIONS FOR HISTORICAL (1990–2019) AND PROJECTED (2040–2069) YEARLY HEAT-RELATED HEALTH COSTS FOR DIRECT, INDIRECT, INTANGIBLE AND TOTAL COMPONENTS FOR A) ALL HEAT AND B) EXTREME HEAT.....	155
FIGURE 6.4	HEALTH COSTS BREAKDOWN OF ALL HEAT BY A) DIRECT, B) INDIRECT, C) INTANGIBLE AND D) TOTAL COSTS DURING THE HISTORICAL (1990–2019) AND PROJECTED (2040–2069) PERIODS BASED ON SSP2–4.5 AND SSP5–8.5 CLIMATE AND DEMOGRAPHIC PROJECTIONS.	157
FIGURE 7.1	PERSPECTIVES POUR LES TRAVAUX FUTURS DE LA THÈSE.	172
FIGURE A.1	ILLUSTRATION D'UN PERCEPTRON À UNE SEULE COUCHE.	207
FIGURE A.2	ILLUSTRATION D'UN PERCEPTRON MULTICOUCHE.	208
FIGURE A.3	ILLUSTRATION D'UN RÉSEAU DE NEURONES RÉCURRENT.....	208
FIGURE A.4	ILLUSTRATION D'UNE CELLULE D'UN <i>LONG SHORT-TERM MEMORY</i>	209

LISTE DES TABLEAUX

TABLEAU 1.1	RÉSUMÉ DES TROIS SOUS-OBJECTIFS 1A, 1B ET 1C DE L'OBJECTIF 1 DE LA THÈSE, AINSI QUE L'IDENTIFICATION DES CONTRIBUTIONS ET ORIGINALITÉS PRINCIPALES ET SECONDAIRES.....	21
TABLEAU 1.2	COMPARAISON DES APPROCHES SOUS LA GRANDE FAMILLE DE LA SCIENCE DES DONNÉES UTILISÉES DANS LE CADRE DE CETTE THÈSE.....	27
TABLEAU 2.1	OVERVIEW OF THE DATA FOR MODELLING THE HEAT-MORTALITY RELATIONSHIP IN MONTREAL CMA (1998–2019, MAY–SEPTEMBER).....	41
TABLEAU 2.2	FIVE MOST IMPORTANT PREDICTORS IN ALL CONSIDERED MODELS.....	50
TABLEAU 3.1	MORTALITY, TEMPERATURE, WEATHER AND AIR POLLUTION DATA IN THE CMAs OF A) MONTREAL (1998–2019) AND B) QUEBEC (2003–2019).....	70
TABLEAU 3.2	TOP 5 MOST IMPORTANT PREDICTORS IN MODELS USING ALL PREDICTORS IN THE CMAs OF A) MONTREAL AND B) QUEBEC.....	77
TABLEAU 3.3	RESULTS SUMMARY IN TERMS OF PREDICTORS, MODELS AND PERFORMANCE IN THE CMAs OF A) MONTREAL AND B) QUEBEC.....	78
TABLEAU 4.1	OVERVIEW OF THE STUDIED HEALTH OUTCOMES.	93
TABLEAU 4.2	WEATHER, REGIONAL AND TEMPORAL PREDICTORS CONSIDERED IN THIS STUDY.	95
TABLEAU 4.3	PERFORMANCE METRICS DURING OUT-OF-SAMPLE TEST YEARS IN TERMS OF R^2 (LEFT SIDE) AND DECREASE IN MEAN ABSOLUTE ERROR (dMAE) (RIGHT SIDE) ACROSS ALL REGIONS.	99
TABLEAU 4.4	TOP 15 MOST IMPORTANT PREDICTORS FOR ALL MODELS IN THE GENERAL POPULATION FOR A) MORTALITY, B) HOSPITALIZATIONS, C) EMERGENCY DEPARTMENT VISITS, D) AMBULANCE TRANSPORTS AND E) 811 CALLS.	105
TABLEAU 5.1	DESCRIPTIVE STATISTICS OF THE STUDIED HEALTH OUTCOMES FOR EACH HEALTH REGION (HR) AND THE TOTAL FOR THE PROVINCE OF QUEBEC.	126
TABLEAU 5.2	DESCRIPTIVE STATISTICS OF WEATHER AND AIR POLLUTION VARIABLES FOR EACH HEALTH REGION (HR).....	127
TABLEAU 5.3	ATTRIBUTABLE NUMBER (AN) AND FRACTION (AF) IN THE PROVINCE OF QUEBEC COMPUTED FROM REGIONAL BLUP FOR ALL HEAT AND EXTREME HEAT EXPOSURES.	130
TABLEAU 6.1	ATTRIBUTABLE NUMBERS (AN) TO ALL AND EXTREME HEAT EVERY SUMMER IN QUEBEC FOR MORTALITY (MOR), HOSPITAL ADMISSIONS (HOS), ED VISITS (EDV), AMBULANCE TRANSPORTS (AMB) AND CALLS TO THE 811-HEALTH HOTLINE (811), AS WELL AS NUMBER OF EXTREME HEATWAVES (HW) ACROSS THE 15 HEALTH REGIONS OF QUEBEC AND PERSON-DAYS EXPOSED TO HW DURING THE HISTORICAL (1990–2019) AND PROJECTED (2040–2069) PERIODS UNDER SSP2–4.5.	153
TABLEAU 6.2	YEARLY DIRECT, INDIRECT, INTANGIBLE AND TOTAL HEALTH COSTS (IN CANADIAN M\$2019) ATTRIBUTABLE TO ALL AND EXTREME HEAT IN THE HISTORICAL (1990–2019) AND PROJECTED (2040–2069) PERIODS UNDER SSP2–4.5.	154
TABLEAU 6.3	OVERVIEW OF THE ECONOMIC METRICS USED IN THIS STUDY AND CORRESPONDING REFERENCES, ALONG WITH THEIR UNCERTAINTY AND DISTRIBUTIONS USED TO COMPUTE MONTE CARLO SIMULATIONS.	164

LISTE DES ABRÉVIATIONS

DLNM	Modèle non-linéaire à délais distribués (<i>Distributed Lag Non-Linear Model</i>)
DT	Arbre de décision (<i>Decision Tree</i>)
DTR	Écart diurne de température (<i>Diurnal temperature range</i>)
ECCC	Environnement et Changement Climatique Canada
GAM	Modèle additif généralisé (<i>Generalized Additive Model</i>)
GBM	<i>Gradient Boosting Model</i>
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
GLM	Modèle linéaire généralisé (<i>Generalized Linear Model</i>)
INRS	Institut national de la recherche scientifique
INSPQ	Institut national de santé publique du Québec
Lasso	<i>Least Absolute Shrinkage and Selection Operator</i>
LGBM	<i>Light Gradient Boosting Model</i>
LSTM	<i>Long Short-Term Memory</i>
MAE	Erreur absolue moyenne (<i>Mean absolute error</i>)
MLP	Perceptron à multicouche (<i>Multi-Layer Perceptron</i>)
MMT	Température de mortalité/morbidité minimale (<i>Minimum mortality/morbidity temperature</i>)
MTC	Changement moyen de température (<i>Mean temperature change</i>)
RF	Forêt aléatoire (<i>Random Forest</i>)
RMR	Région métropolitaine de recensement
RMSE	Racine de l'erreur quadratique moyenne (<i>Root mean square error</i>)
RSS	Région sociosanitaire
SLP	Perceptron à une seule couche (<i>Single-Layer Perceptron</i>)
SSP	Trajectoire socioéconomique partagé (<i>Shared socioeconomic pathway</i>)

LISTE DES CONTRIBUTIONS

Cette thèse contient cinq articles qui sont présentés aux chapitres 2 à 6. L'originalité et la contribution de chacun d'entre eux sont décrites à la section 1.4, alors que le lien entre chaque article est présenté à la section 1.5. L'apport de chacun des auteurs est présenté au début des chapitres 2 à 6.

[A1] J **Boudreault***, C Campagna, F Chebana (2023). [Machine and deep learning for modelling heat-health relationships](#). *Science of the Total Environment*. 682: 164660.

[A2] J **Boudreault***, C Campagna, F Chebana (2024). [Revisiting the importance of temperature, weather and pollution variables in heat-mortality relationships with machine learning](#). *Environmental Science and Pollution Research*. 31 (9): 14059-14070.

[A3] J **Boudreault***, A Ruf, C Campagna, F Chebana (2024). [Multi-region models built with machine and deep learning for predicting several heat-related health outcomes](#). *Sustainable Cities and Society*. 115: 105785.

[A4] J **Boudreault***, E Lavigne, C Campagna, F Chebana (2024). [Estimating the heat-related mortality and morbidity burden in the province of Quebec, Canada](#). *Environmental Research*. 257: 119347.

[A5] J **Boudreault***, C Campagna, E Lavigne, F Chebana (2025). [Projecting the overall heat-related health burden and associated economic costs in a climate change context in Quebec, Canada](#). *Science of the Total Environment*. 958: 178022.

CHAPITRE 1 : INTRODUCTION

1 INTRODUCTION

Dans ce chapitre, nous ferons d'abord une brève mise en contexte du sujet (section 1.1). Puis, nous présenterons une revue de la littérature spécifique à celui-ci (1.2). Les problématiques (1.3) et les objectifs (1.4) seront ensuite abordés, avant de conclure sur la structure de la thèse et le lien entre les différents chapitres (1.5).

1.1 Mise en contexte

Le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) indique dans son dernier rapport que les changements climatiques amènent une augmentation dans la fréquence, l'intensité et la durée de plusieurs événements météorologiques extrêmes (IPCC, 2021). Parmi ceux-ci, les augmentations prévues dans les épisodes de chaleur sont bien documentées depuis plusieurs décennies déjà (Jeong et al., 2016; Meehl & Tebaldi, 2004). Ces épisodes de chaleur entraînent des impacts importants sur la santé des populations au niveau mondial, notamment une mortalité et une morbidité en hausse (Basu, 2009; Campbell et al., 2018; Li et al., 2015). Au Québec, par exemple, plusieurs épisodes de chaleur affectant la santé ont eu lieu dans les 15 dernières années, notamment en 2010, 2018, 2020, 2021 et 2024 (Lamothe et al., 2025). Les changements climatiques et sociodémographiques viendront exacerber ces impacts déjà importants si des mesures d'atténuation et d'adaptation ne sont pas mises en place rapidement (Cole et al., 2023; Huang et al., 2011; Sanderson et al., 2017). Il est donc primordial que les autorités sanitaires aient en main les meilleurs outils et données disponibles pour réduire le fardeau actuel et à venir de la chaleur (Austin et al., 2019; Huang et al., 2013).

Malgré les récents développements en apprentissage automatique, ces modèles demeurent peu utilisés en santé publique et environnementale (Fisher & Rosella, 2022; Morgenstern et al., 2020; Schmidt, 2020), avec de très faibles utilisations dans le contexte de la chaleur et de la santé comparativement à d'autres domaines (Ssebyala et al., 2024; Yin et al., 2024). Ces modèles sont intéressants, car ils permettent d'employer des ensembles de données plus importants (*big data*), de modéliser des relations complexes et non linéaires, et ne nécessitent pas d'hypothèses fortes sur la distribution des données (Bi et al., 2019), ce qui peut mener à un pouvoir prédictif accru par rapport aux approches plus classiques. En appliquant ces modèles pour la prévision des impacts sanitaires de la chaleur en contexte québécois, ces derniers pourraient améliorer notre préparation face aux épisodes de chaleur en s'intégrant à la vigie sanitaire des événements météorologiques extrêmes déjà en place (Toutant et al., 2011). À la clé, l'apprentissage

automatique donnerait aux autorités des outils solides, validés et vérifiés pour mieux se préparer aux conséquences sanitaires des épisodes de chaleur à venir.

En plus de la faible utilisation de l'apprentissage automatique dans le domaine, on manque encore de connaissances sur les coûts associés à la chaleur (Schmitt et al. 2016; Wondmagegn et al. 2019). Cela peut s'expliquer, d'une part, par le manque de données facilement accessibles sur les coûts sanitaires de la chaleur, comparativement aux pertes assurables lors d'autres catastrophes naturelles comme les inondations ou les feux de forêt (Kousky, 2014). D'autre part, le fait que la chaleur soit souvent reconnue comme un « tueur silencieux » rend ses effets moins visibles, quoique bien réels (Luber & McGeehin, 2008). Or, les plus récents modèles statistiques utilisés en épidémiologie environnementale permettent d'attribuer les conséquences sanitaires à la chaleur et, donc, de calculer le fardeau sanitaire en période historique (Gasparrini et al. 2014; 2015) et future (Vicedo-Cabrera et al. 2019). Le calcul du fardeau sanitaire combiné à une approche coût de la maladie permettrait d'estimer le fardeau économique associé à la chaleur en contexte québécois (Paudel et al., 2020). Ces nouvelles données exposeraient aux gouvernements et autorités de santé l'ampleur du fardeau sanitaire et économique de la chaleur, actuel et à venir, leur permettant ainsi de mettre en place les mesures d'adaptation les plus efficaces afin de le réduire.

Dans cette thèse, deux grandes familles d'approches de modélisation sont considérées, c'est-à-dire 1) l'apprentissage automatique et profond, afin de prédire les impacts à court terme de la chaleur sur la santé, et 2) l'épidémiologie environnementale, afin de calculer le fardeau sanitaire et économique de la chaleur, actuel et projeté. Ces deux familles d'approches mises à profit dans le cadre de cette thèse sont présentées sous le parapluie de la science des données, une science émergente qui met à profit les relations inhérentes aux données comme on les retrouve dans ce projet (Figure 1.1). La science des données permet donc de revisiter les impacts sanitaires et économiques de la chaleur et de générer des résultats probants et novateurs qui seront d'intérêt pour les acteurs du milieu de la santé au Québec, comme ailleurs dans le monde.

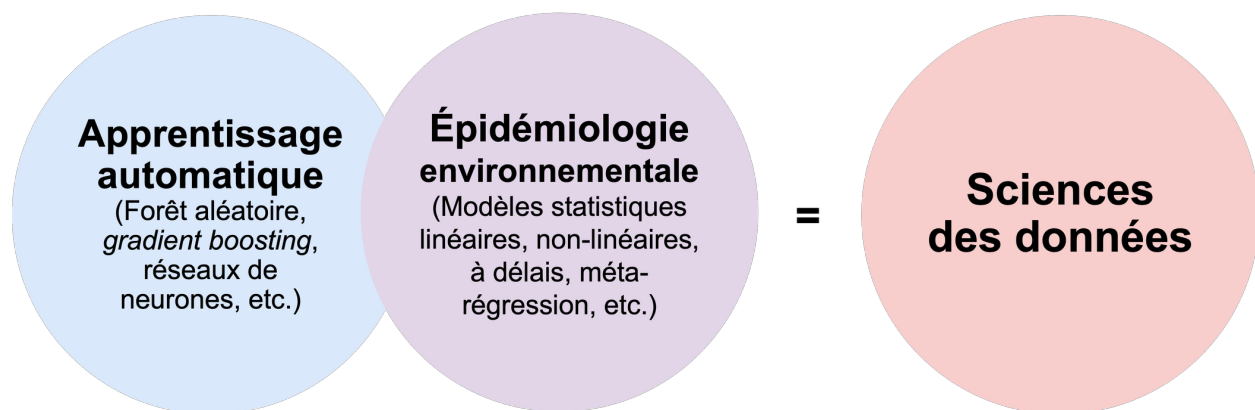


Figure 1.1 Résumé des approches de la science des données considérées dans le cadre de cette thèse.

1.2 Revue de littérature

La revue de littérature pour cette thèse se divise en cinq parties. D'abord, les effets sanitaires de la chaleur sont revus (section 1.2.1). Puis, les méthodes statistiques (1.2.2) et d'apprentissage automatique (1.2.3) utilisées pour modéliser les impacts sanitaires de la chaleur sont présentées. Finalement, les calculs du fardeau sanitaire (1.2.4) et des coûts de santé (1.2.5) associés à la chaleur concluent cette section.

1.2.1 Effets sanitaires de la chaleur

Dans le cadre de cette thèse, nous définissons par **chaleur** toute température au-delà d'un seuil sécuritaire pour la santé, que cette température soit extrême ou pas (Basu, 2009; Gasparrini et al., 2015). Ainsi, cette définition de la chaleur diffère de celle d'une vague de chaleur, pour laquelle il existe de multiples définitions en fonction de l'intensité, la durée, la métrique de température choisie, etc. (Kent et al., 2014; Xu et al., 2016). Par exemple, au Québec, les *vagues de chaleur extrême* sont définies par l'Institut national de santé publique (INSPQ) comme des périodes de trois journées consécutives où les températures minimales et maximales sont très élevées (Lamothe et al., 2025). Ainsi, notre définition plus flexible de la chaleur permet de laisser les données elles-mêmes caractériser le phénomène de chaleur affectant la santé de la population, plutôt que d'utiliser une définition basée sur un excès de mortalité (Chebana et al., 2013).

Les effets de la chaleur sur la santé sont nombreux et bien documentés. Des revues de littérature ont déjà démontré des effets sur la mortalité (Basu, 2009; Gosling et al., 2009) ainsi que sur la morbidité (Li et al., 2015; Xu et al., 2023; Ye et al., 2012). Dans le cadre de cette thèse, nous faisons référence à la morbidité pour les impacts sanitaires autres que la mortalité, soit les admissions à l'hôpital, les visites à l'urgence, les transports ambulanciers et les appels à une ligne d'information de santé (comme Info-Santé 811 au Québec). À ce jour, les effets sur la mortalité sont beaucoup plus étudiés que ceux sur la morbidité (Campbell et al., 2018; Cheng et al., 2019). Bien que les indicateurs de mortalité et de morbidité toutes causes confondues soient souvent utilisés, des causes spécifiques de maladies sont affectées par la chaleur, notamment les maladies cardiovasculaires, respiratoires, cérébrovasculaires et génito-urinaires, le diabète, l'hypertension, les troubles de santé mentale, la déshydratation et les autres troubles liés à la chaleur comme le coup de chaleur (Basu, 2009; Ebi et al., 2021).

De nombreuses personnes peuvent être plus vulnérables à la chaleur, par exemple les jeunes enfants, les personnes âgées de 65 ans et plus, les travailleurs extérieurs, les personnes avec des maladies chroniques, celles qui prennent des médicaments ou des drogues, celles vivant

seules ou ne pouvant pas prendre soin d'elles-mêmes, etc. (Ebi et al., 2021; Health Canada, 2022; Luber & McGeehin, 2008; Martiello & Giacchi, 2010). Au niveau populationnel, les personnes vivant dans les quartiers les plus défavorisés socialement et/ou matériellement ou dans les quartiers affichant une faible canopée et un fort phénomène d'îlot de chaleur urbain peuvent aussi être particulièrement affectées (Ebi et al., 2021; Health Canada, 2022; Luber & McGeehin, 2008; Martiello & Giacchi, 2010). Bien que des causes de maladies et populations spécifiques soient plus affectées par les effets de la chaleur, nous ne traiterons dans le cadre de cette thèse que des indicateurs de mortalité et de morbidité toutes causes confondues et pour l'ensemble de la population québécoise, comme il s'agit des premiers travaux de ce genre à être effectués au Québec. Ces indicateurs de santé sont les mêmes que ceux présentés dans les bilans des vagues de chaleur extrême effectués par l'INSPQ depuis les 15 dernières années (e.g., Bustinza et al., 2013; Lamothe et al., 2025; Lebel et al., 2019) ainsi que ceux intégrés au système de surveillance des effets sanitaires des événements météorologiques extrêmes au Québec (Toutant et al., 2011).

1.2.2 Méthodes statistiques utilisées en épidémiologie environnementale

Le devis statistique le plus utilisé à ce jour pour modéliser les effets sanitaires de la chaleur est la régression de séries chronologiques (Gosling et al., 2009; Graffy et al., 2024). Brièvement, une régression entre la mortalité ou la morbidité quotidienne et une variable d'exposition comme la température moyenne quotidienne est effectuée sur une période de plusieurs années, en ajustant le modèle pour les facteurs temporels et autres covariables comme l'humidité relative ou la pollution de l'air (Bhaskaran et al., 2013). Ces modèles peuvent être linéaires ou non, et incluent souvent des délais (*lags*) de la variable d'exposition avant l'observation de la mortalité ou de la morbidité. En effet, les variables de température à délais sont très importantes pour caractériser la relation entre la température et la santé, car il peut s'écouler plusieurs jours avant d'observer l'effet à la santé (Basu, 2009; Son et al., 2019).

Différentes variables de température peuvent être considérées pour l'exposition à la chaleur, comme la température moyenne, maximale ou minimale quotidienne, ou des indices composites représentant la température ressentie qui incluent l'humidité ou la vitesse des vents comme l'humidex au Canada ou le *heat index* aux États-Unis (Song et al., 2017). Généralement, les études épidémiologiques qui ont comparé différents indicateurs de température ont montré que la température moyenne était équivalente à utiliser d'autres indices (e.g., Barnett et al. 2010; Vaneckova et al. 2011). Les variations de températures à différentes échelles temporelles

peuvent aussi affecter la santé (e.g., Cheng et al. 2014a; Guo et al. 2014; Pan et al. 2022), bien que celles-ci ne soient pas nécessairement liées à l'exposition à la chaleur.

Au cours de la dernière décennie, un modèle statistique s'est imposé en épidémiologie environnementale pour analyser les relations entre la chaleur et la santé : le modèle non-linéaire à délais distribués (*Distributed Lag Non-Linear Model*, DLNM) (Armstrong 2006; Gasparrini et al., 2010). Le DLNM a l'avantage de considérer à la fois l'effet retardé (*lag*) de l'exposition aux températures sur la santé et son effet non-linéaire. Le DLNM est la méthode la plus utilisée à ce jour pour analyser les relations entre la chaleur et la santé (Graffy et al., 2024). Ce modèle a d'ailleurs été étendu afin de permettre le calcul du fardeau sanitaire associé à la chaleur (Gasparrini & Leone, 2014), effectuer une méta-régression pour combiner les résultats de DLNM régionaux (Sera et al., 2019) et faire des projections des effets à la santé en contexte de changements climatiques (Vicedo-Cabrera et al. 2019).

Au niveau des études menées au Québec, des modèles de régression avec des splines ont d'abord été utilisés pour modéliser la mortalité toutes causes dans plusieurs régions du Québec (Doyon et al., 2006; 2008). Un modèle Poisson a aussi été ajusté à Montréal afin d'analyser l'effet d'îlot de chaleur urbain (Smargiassi et al., 2009) et pour prévoir l'effet des changements climatiques (Benmarhnia et al., 2014) sur la mortalité. Le DLNM a été utilisé pour la mortalité à Montréal (Goldberg et al., 2011), puis dans 15 villes canadiennes incluant Québec et Montréal (Martin et al., 2012), et finalement pour toutes les régions sociosanitaires (RSS) du Canada (Hebborn et al., 2023). Des études internationales sur la mortalité incluent souvent plusieurs villes au Canada, mais seulement Montréal au Québec (Gasparrini et al., 2015). D'autres études au Québec touchent plus particulièrement les troubles de santé mentale (e.g., Tupinier Martin et al., 2024; Vida et al., 2012), les maladies cardiovasculaires (e.g., Chiu, 2017; 2021; Masselot et al., 2018; 2022), les accidents de travail (Adam-Poupard et al., 2014; 2021), la santé périnatale (e.g., Auger et al., 2014; 2017; He et al., 2018) ou la modélisation de seuils pour la chaleur extrême (e.g., Chebana et al., 2013; Issa et al., 2021; Masselot et al., 2021).

1.2.3 Méthodes d'apprentissage automatique

Les modèles d'apprentissage automatique incluent une variété d'approches, dont les deux principales familles d'apprentissage de type supervisé seront abordées dans cette thèse : 1) les modèles basés sur les arbres et 2) les réseaux de neurones. La première famille se base sur l'arbre de décision (*Decision Tree*, DT), qui permet de séparer l'espace des prédicteurs en différentes régions ayant la forme d'un arbre inversé (Quinlan, 1986). Les méthodes d'ensemble

tirent avantages de plusieurs arbres, comme c'est le cas pour la forêt aléatoire (*Random Forest*, RF) (Breiman, 2001) et les modèles de *boosting* de gradient (*Gradient Boosting Model*, GBM) (Friedman, 2001; Ke et al., 2017). La seconde famille inclut les plus simples réseaux de neurones, à savoir le perceptron à une couche (*Single-Layer Perceptron*, SLP) et à plusieurs couches (*Multi-Layer Perceptron*, MLP) (Chapitre 6 de Goodfellow et al. 2016). De plus, le *Long Short-Term Memory* (LSTM), un réseau de type récurrent est aussi considéré dans le cadre de cette thèse, pour prendre en compte la structure temporelle des données (Hochreiter & Schmidhuber, 1997). Ces réseaux de neurones font partie de la famille de l'apprentissage profond lorsqu'ils contiennent plusieurs couches cachées. Pour plus d'informations sur ces différentes approches, le lecteur est invité à consulter la description des modèles présentée en annexe de la thèse ainsi que les références ci-dessus.

Les modèles d'apprentissage automatique commencent à être utilisés en santé publique, alors que leur utilisation est beaucoup plus répandue en santé clinique (Morgenstern et al., 2020). En santé environnementale, leur utilisation est souvent associée à l'estimation de la pollution de l'air (Bellinger et al., 2017). Ces modèles n'ont été que peu utilisés pour prédire les effets sanitaires de la chaleur. En effet, une revue de la littérature effectuée jusqu'en 2022 n'a relevé que 3 articles utilisant l'apprentissage automatique à cette fin (Ssebyala et al., 2024). Une autre revue portant spécifiquement sur les coups de chaleur n'a identifié que 5 articles (Yin et al., 2024). La première application recensée des modèles d'apprentissage automatique pour la relation entre la chaleur et la santé remonte à l'utilisation du DT en Australie en 2008 (Hu et al., 2008) puis du RF aux États-Unis en 2014 (Zhang et al. 2014). Une nouvelle revue de la littérature effectuée dans le cadre de la rédaction de cette thèse a démontré que le RF est le modèle le plus utilisé dans la littérature (Jian et al. 2023; Ke et al. 2023; Kim et Kim 2022; Nishimura et al. 2021; Ogata et al. 2021; Oka et al. 2024; Oka et Hijioka 2021; Park et al. 2020; Wang et al. 2019; Wertis et al. 2023; Xu et al. 2024), suivi par le modèle GBM (Ke et al. 2023; Ogata et al. 2021; Ohashi et al. 2023; Wertis et al. 2023; Xu et al. 2024) et les réseaux de neurones de type MLP ou LSTM (Nishimura et al. 2021; Wang et al. 2024; Xu et al. 2024). Cela démontre que le choix des approches retenues pour cette thèse, c'est-à-dire les modèles basés sur les arbres et les réseaux de neurones, est adéquat car il permet des comparaisons avec les études antérieures de la littérature.

Les approches d'apprentissage automatique utilisées afin de prévoir les effets sanitaires de la chaleur sont souvent comparées aux modèles statistiques plus traditionnels comme le modèle linéaire généralisé (*Generalised Linear Model*, GLM), le modèle additif généralisé (*Generalized Additive Model*, GAM) ou une régression pénalisée comme le *Least Absolute Shrinkage and*

Selection Operator (Lasso) (Hu et al. 2008; Jian et al. 2023; Ogata et al. 2021; Oka et Hijioka 2021; Park et al. 2020; Wertis et al. 2023; Xu et al. 2024). À notre connaissance, il n'y a pas eu de comparaison formelle de l'apprentissage automatique avec le DLNM, bien qu'il soit le modèle le plus utilisé pour la chaleur en épidémiologie environnementale tel que vu à la section 1.2.2. Ce dernier a par contre été utilisé comme étape de pré- ou de post-traitement dans certaines études utilisant l'apprentissage automatique (Ohashi et al. 2023; Wertis et al. 2023; Xu et al. 2024).

En début de thèse, une seule étude avait été effectuée au Québec utilisant l'apprentissage automatique afin de modéliser des seuils de chaleur extrême liés à la mortalité à Montréal (Masselot et al., 2021). Comme cette étude ne visait pas la prévision des impacts sanitaires mais plutôt l'identification de "sauts" dans la relation, des approches différentes de celles proposées dans le cadre de cette thèse avaient été utilisées, notamment le *Multivariate Adaptive Regression Spline*. Depuis, deux autres études ont vu le jour portant sur la prévision de la mortalité toutes causes dans trois régions du Québec (Côté et al. 2024a) et dans la région de Montréal (Côté et al. 2024b). Ces études, qui utilisaient un modèle *AutoGluon* (un ensemble de différents modèles) et des processus gaussiens, ont été publiées après les travaux issus de cette thèse (chapitres 2, 3 et 4). Aucune autre étude n'a été recensée ailleurs au Canada utilisant l'apprentissage automatique pour prédire les impacts sanitaires de la chaleur.

Bien que les méthodes utilisées en épidémiologie environnementale et en apprentissage automatique aient été revues en profondeur ici, il est important de noter que d'autres domaines s'intéressent aussi à la relation chaleur-santé, notamment en sciences actuarielles (e.g., Seklecka et al. 2017; Li & Tang 2022; Robben et al. 2025) et en sciences économiques (e.g., Deschênes and Greenstone, 2014; Dell et al. 2014; Diaz & Moore 2017). Cependant, ces méthodes dépassent le cadre de cette thèse qui s'intéresse plutôt aux familles utilisées en épidémiologie environnementale et apprentissage automatique (Figure 1.1).

1.2.4 Fardeau sanitaire de la chaleur

Le fardeau sanitaire (*burden of disease* en anglais) permet de mesurer l'effet d'un problème de santé ou d'une exposition environnementale à une échelle populationnelle (Paudel et al., 2020). Dans le cadre de cette thèse, le fardeau sanitaire est représenté par les nombres et fractions des issues sanitaires (mortalité et morbidité) attribuables à la chaleur. Grâce aux développements récents, les modèles DLNM, présentés à la section 1.2.2, permettent de calculer directement le fardeau sanitaire attribuable à la chaleur. Cependant, aucun des modèles d'apprentissage automatique (revus à la section 1.2.3) n'a été développé à cette fin dans la littérature actuelle.

Ainsi, seul le DLNM sera présenté et utilisé pour le calcul du fardeau sanitaire dans le cadre de la thèse, comme il s'agit de l'approche la plus pertinente et la plus utilisée dans la littérature (Gasparrini et al., 2015; Zhao et al., 2021).

Brièvement, les fonctions dose-réponse obtenues à l'aide du DLNM donnent le risque relatif de mortalité ou de morbidité pour chaque valeur de température par rapport à une température de référence pour laquelle le risque est nul. Cette température est appelée la température de mortalité ou de morbidité minimale (*minimum mortality/morbidity temperature*, MMT). Les températures plus élevées que la MMT sont associées à l'effet de la chaleur sur la santé. Les risques relatifs permettent de calculer la fraction et le nombre attribuable (p. ex. de décès) à la chaleur, qui sont des mesures du fardeau sanitaire (Gasparrini & Leone, 2014). À l'échelle mondiale, on estime que 500 000 décès étaient attribuables à la chaleur chaque année en moyenne entre 2000 et 2019 (Zhao et al., 2021).

En plus de quantifier le fardeau historique de la chaleur, les modèles DLNM peuvent être également utilisés avec des simulations climatiques pour évaluer l'évolution du fardeau sanitaire selon différents scénarios du GIEC (Vicedo-Cabrera et al., 2019). Les trajectoires socioéconomiques (*shared socioeconomic pathways*; SSP) souvent considérées sont le scénario SSP2–4.5, soit un scénario médian d'émissions de gaz à effet de serre menant à +2.7°C d'ici la fin du siècle (par rapport à l'ère préindustrielle), et le scénario SSP5–8.5, soit un scénario beaucoup plus pessimiste menant à +4.4°C d'ici la fin du siècle (IPCC, 2021). À l'échelle mondiale, le fardeau de la chaleur sur la mortalité pourrait être multiplié par 6 à 12 fois d'ici la fin du siècle (Gasparrini et al., 2017; Lee et al., 2019). Des projections démographiques sont souvent incluses dans les projections du fardeau sanitaire de la chaleur afin de tenir compte d'une population plus grande et potentiellement plus vulnérable à la chaleur dans le futur (Cole et al., 2023).

Les études internationales sur le fardeau sanitaire historique de la chaleur ont déjà été revues (Cheng et al., 2019). Au niveau du Québec, l'étude de Gasparrini et al. (2015) étudiait le fardeau de la chaleur sur la mortalité pour 384 villes, dont Montréal. Ils ont trouvé que 72 décès pouvaient être attribuables à la chaleur par année entre 1986 et 2009. Benmarhnia et al. (2014) ont trouvé des nombres similaires, avec 62 décès annuels attribuables à la chaleur pour les mois de juin à août. Une étude canadienne a reporté une fraction attribuable à la chaleur de 0.4% des décès totaux au Canada pour la période 2010 à 2019 (Hebborn et al., 2023). Les études internationales sur le fardeau sanitaire futur de la chaleur ont aussi été revues (Cole et al., 2023; Huang et al., 2011; Sanderson et al., 2017). Au Québec, une étude sur 3 villes rapportait des pourcentages de

mortalité attribuables à la chaleur de 0.5–2.0% en période historique et de 3–10% en 2080 (Doyon et al., 2008). Une autre étude sur plusieurs villes canadiennes a prédit un changement de +18 décès par 100 000 habitants d'ici la fin du siècle pour Montréal et de +1.5 décès par 100 000 habitants pour la ville de Québec (Martin et al., 2012). Finalement, l'étude canadienne de Hebborn et al. (2023) prévoyait une fraction attribuable des décès à la chaleur de 10.7% d'ici la fin du siècle au Canada selon le scénario de réchauffement SSP5–8.5 en incluant le vieillissement de la population.

1.2.5 Coûts sanitaires de la chaleur

Les événements météorologiques extrêmes amènent des coûts de santé importants (Kousky, 2014; Schmitt et al., 2016). Une revue de littérature portant spécifiquement sur les coûts des soins de santé de la chaleur effectuée en 2019 recensait seulement 10 études (Wondmagegn et al., 2019). Comme cette revue n'incluait que les coûts associés à la morbidité, une nouvelle revue a été effectuée en début de thèse et a permis d'identifier 33 études, dont 7 sur la mortalité, 16 sur la morbidité et 10 combinant la mortalité et la morbidité (Boudreault et al., 2022). Parmi les études identifiées, 17 portaient sur les coûts historiques de la chaleur et 16 incluaient des projections de coûts. Une mise à jour de cette revue dans le cadre de la rédaction de cette thèse n'a pas recensé de différences majeures dans le type et la qualité des études publiées. La majorité des études utilise l'approche du coût de la maladie (*cost of illness* en anglais) qui permet de quantifier économiquement les impacts directs, indirects et intangibles d'une maladie (Paudel et al., 2020). La maladie ici n'est pas définie comme une seule maladie, mais bien par tous les problèmes de santé qui découlent de l'exposition à la chaleur.

Une majorité d'études se base sur l'estimation du fardeau sanitaire avec une étude épidémiologique en premier lieu (tel que revu à la section 1.2.4), puis une conversion de ce fardeau en coûts sanitaires en utilisant différentes métriques comme la valeur statistique de la vie humaine pour la mortalité, le coût moyen d'un soin de santé (p. ex. une hospitalisation) pour la morbidité, la durée des soins de santé et le salaire médian pour les pertes de productivité associées, la perte de bien-être pendant l'exposition aux vagues de chaleur extrême, etc. (Boudreault et al., 2022; Wondmagegn et al., 2019). C'est cette approche de type épidémiologique qui sera considérée dans le cadre de cette thèse, excluant les approches de la famille macro-économique qui permettent de quantifier les impacts économiques des changements climatiques, mais qui n'incluent que certaines composantes sanitaires comme la mortalité (Dell et al., 2014; Diaz & Moore, 2017). Notons aussi que certaines études modélisent

directement la série chronologique des coûts de santé en fonction de la température (Tong et al., 2021abc; Wang et al., 2021; Wondmagegn et al., 2021a; Yu et al., 2024), ce qui nécessite l'accès à des séries temporelles de coûts de santé. Or, de telles données ne sont actuellement pas disponibles au Québec. Finalement, d'autres études n'analysent que les coûts de certains troubles liés à la chaleur comme les coups de chaleur (Merrill et al., 2008; Noe et al., 2012; Schmeltz et al., 2016), ce qui représente qu'une faible fraction des coûts sanitaires totaux associés à la chaleur.

Au Canada, un rapport sur les coûts de mortalité de la chaleur dans 4 villes canadiennes estimait les coûts à 0.6–0.8 milliard de dollars en 2020 et de 1.4–2.8 en 2050 pour la ville de Montréal (NRTEE, 2011). Ouranos a estimé les coûts sanitaires de la chaleur induits par les changements climatiques au Québec à 32.4 milliards de dollars (pour la mortalité) et à 372 millions (pour la morbidité) pour la période 2015 à 2065 avec un taux d'actualisation de 4% (Larivée et al., 2015). Un rapport de l'Institut climatique du Canada a estimé des coûts sanitaires annuels dans l'ensemble du Canada de 3.0–3.9 milliards de dollars en 2050 pour la mortalité et de 138–167 millions de dollars pour la morbidité (Clark et al., 2021). Les coûts du dôme de chaleur de 2021 en Colombie-Britannique ont été estimés à 5.5 milliards de dollars en termes de vies humaines perdues et à 12 millions de dollars en termes de soins de santé dans un récent rapport (Beugin et al., 2023). À l'exception des rapports cités ci-dessus et de deux études en Ontario (Bai et al., 2018) et en Alberta (Tran & Warkentin, 2025), au meilleur de notre connaissance, il n'existe pas d'autres études sur les coûts sanitaires de la chaleur portant spécifiquement sur le Québec ou le Canada.

1.3 Problématiques

La revue de littérature a mis en évidence plusieurs lacunes dans la littérature actuelle des impacts sanitaires et économiques de la chaleur. En général, un manque a été identifié au niveau de la mise à profit des modèles d'apprentissage automatique et profond dans le domaine, ainsi qu'au niveau de la quantification du fardeau sanitaire et économique de la chaleur, actuel et futur. Ces faiblesses freinent la mise en place de mesures d'adaptation permettant de mieux protéger la population face aux effets grandissants de la chaleur. Les problématiques spécifiques sont décrites ci-dessous, à savoir celles liées à l'apprentissage automatique (1.3.1), au calcul du fardeau sanitaire (1.3.2) et aux coûts de santé associés (1.3.3).

1.3.1 Limites des applications actuelles de l'apprentissage automatique

Les revues de littérature passées sur l'utilisation de l'apprentissage automatique pour modéliser les impacts sanitaires de la chaleur n'ont relevé que quelques études (Ssebyala et al., 2024; Yin et al., 2024), en plus de celles identifiées dans le cadre de la rédaction de cette thèse. Ces études présentent diverses lacunes. Dans plusieurs de celles-ci, un seul modèle d'apprentissage automatique a été utilisé (Kim & Kim, 2022; Ohashi et al., 2023; Wang et al., 2024; Wang et al., 2019; Zhang et al., 2014). Cela ne permet donc pas de conclure sur le potentiel accru de l'apprentissage automatique par rapport aux approches plus traditionnellement utilisées en épidémiologie environnementale. De plus, la majorité des études se basent sur de petits jeux de données de moins de 6 années (Kim & Kim, 2022; Nishimura et al., 2021; Ogata et al., 2021; Oka et al., 2024; Oka & Hijioka, 2021; Park & Kim, 2018; Park et al., 2020) ou sur une seule région d'étude (Hu et al., 2008; Ikeda & Kusaka, 2021; Jiang et al., 2021; Masselot et al., 2021; Ohashi et al., 2023; Xu et al., 2024; Zhang et al., 2014). Cela génère de jeux de données de taille potentiellement insuffisante pour calibrer adéquatement les modèles d'apprentissage automatique, qui nécessitent beaucoup de données pour bien performer. D'ailleurs, la plupart des travaux passés n'ont étudié que la mortalité (Côté et al., 2024ab; Hu et al., 2008; Kim & Kim, 2022; Masselot et al., 2021; Ohashi et al., 2023; Wang et al., 2024; Zhang et al., 2014) ou les troubles liés à la chaleur comme les coups de chaleur (Ikeda & Kusaka, 2021; Ke et al., 2023; Nishimura et al., 2021; Ogata et al., 2021; Oka et al., 2024; Oka & Hijioka, 2021; Park et al., 2020; Wang et al., 2019; Xu et al., 2024), brossant un portrait incomplet des nombreux effets de la chaleur sur la santé, tels que revus à la section 1.2.1.

Quant aux prédicteurs utilisés dans ces études, le choix de l'indice de température à utiliser demeure incertain. Plusieurs indices sont utilisés conjointement, ce qui ne permet pas de conclure

sur le meilleur à utiliser. De plus, il n'existe aucune étude spécifique à ce sujet en contexte québécois. Concernant la pollution de l'air, cette dernière n'a été que rarement considérée dans les applications actuelles de l'apprentissage automatique pour prédire les impacts sanitaires de la chaleur (Hu et al., 2008; Jian et al., 2023). Or, on sait que celle-ci pourrait avoir un effet synergique avec la chaleur en période estivale (Analitis et al., 2018; Kovats & Hajat, 2008; Pascal et al., 2021b; Stafoggia et al., 2023). Des variables comme les variations de températures et d'autres variables météorologiques pourraient aussi être d'intérêt, mais n'ont été que rarement testées dans les modèles (Ke et al., 2023; Ogata et al., 2021; Ohashi et al., 2023; Oka et al., 2024; Oka & Hijioka, 2021; Wertis et al., 2023). Finalement, une revue récente en santé publique a relevé que la majorité des études utilisant l'apprentissage automatique manque de transparence quant aux hyperparamètres choisis et à leur optimisation (Morgenstern et al., 2020), ce qui est aussi observé dans plusieurs études de la littérature chaleur-santé présentant ces mêmes lacunes (Côté et al. 2024a; Hu et al., 2008; Ikeda & Kusaka, 2021; Jian et al., 2023; Nishimura et al., 2021; Park & Kim, 2018; Wang et al., 2024). Cela impacte directement les résultats et les conclusions pouvant être tirées de ces études.

Les points soulevés ci-dessus s'appliquent aussi au contexte québécois ou canadien, où seules trois études ont été recensées. La première visait spécifiquement la modélisation de seuils de chaleur extrême pour la mortalité à Montréal (Masselot et al., 2021), la seconde ne s'intéressait qu'à la mortalité pour 3 régions du Québec (Côté et al., 2024a) et la troisième n'utilisait qu'un processus gaussien présenté comme un modèle d'apprentissage automatique (Côté et al., 2024b). Ainsi, il manque manifestement d'études comparant différentes approches d'apprentissage automatique avec des modèles statistiques, appliquées autant à la mortalité qu'à la morbidité liées à la chaleur, qui sont calibrées et validées adéquatement avec de grandes bases de données sur plusieurs années et régions, incluant plusieurs prédicteurs atmosphériques, dont la pollution de l'air, et ce, en contexte québécois comme international.

1.3.2 Limites des études sur le fardeau sanitaire de la chaleur

Bien qu'une multitude d'études aient analysé la relation entre la chaleur et la santé, il n'existe qu'un faible pourcentage de celles-ci qui ont fait l'étape supplémentaire de quantifier le fardeau sanitaire associé (Cheng et al., 2019). La quantification du fardeau est une étape importante, surtout en vue de calculer les coûts sanitaires associés (Boudreault et al., 2022). De plus, une majorité d'études sur le fardeau sanitaire s'est concentrée sur la mortalité, ce qui limite les conclusions pouvant être tirées concernant la morbidité liée à la chaleur (Cheng et al., 2019).

Ainsi, bien que les effets de la chaleur sur la morbidité soient bien documentés (Campbell et al., 2018; Li et al., 2015; Ye et al., 2012), des travaux sont nécessaires afin de calculer le fardeau sanitaire que ces indicateurs de santé représentent, autant en période historique que future. Ces travaux permettraient de couvrir l'ensemble des aspects du réseau de santé, et donc, d'être plus pertinents pour les autorités sanitaires et gouvernementales. D'ailleurs, plusieurs études n'ont quantifié le fardeau sanitaire que lors de vagues de chaleur historiques récentes (Fouillet et al., 2006; Guo et al., 2018; Yin et al., 2018; Zhao et al., 2024), ce qui ne couvre qu'une partie du fardeau de la chaleur. En effet, les impacts à la santé peuvent avoir lieu sous les seuils de vagues de chaleur et hors des journées consécutives de chaleur (Campbell et al., 2018; Song et al., 2017). Cela est autant plus problématique sachant qu'il existe une multitude de façons de définir les vagues de chaleur, ce qui mènera à des estimations toujours différentes (e.g., Kanti et al., 2022; Kent et al., 2014; Xu et al., 2016).

Ces problématiques quant à la littérature internationale s'appliquent aussi au contexte québécois, où la majorité des études effectuées au Québec ont soit seulement modélisé la relation sans calculer le fardeau associé (e.g., Chebana et al., 2013; Smargiassi et al., 2009; Tupinier Martin et al., 2024), ont calculé le fardeau sanitaire que pour la mortalité (e.g., Benmarhnia et al., 2014; Doyon et al., 2008; Hebborn et al., 2023; Martin et al., 2012) ou ne se sont intéressées qu'aux vagues de chaleur extrême (e.g., Bustinza et al., 2013; Lamothe et al., 2025). Ainsi, la littérature est limitée concernant la modélisation du fardeau sanitaire de plusieurs issues sanitaires simultanément, incluant à la fois la mortalité et la morbidité, pour la chaleur dans son ensemble et pas seulement lors de vagues de chaleur extrême, au Québec et à l'international.

1.3.3 Limites des études sur les coûts de santé de la chaleur

Un très faible nombre d'études existe dans la littérature internationale quant aux coûts sanitaires de la chaleur (Boudreault et al., 2022; Wondmagegn et al., 2019). Cette littérature limitée présente plusieurs problèmes. D'abord, plusieurs études ont touché seulement aux événements historiques de chaleur extrême (Adélaïde et al., 2021; Knowlton et al., 2011; Limaye et al., 2019; Wondmagegn et al. 2021a; Yan et al., 2022). Ces résultats ne représentent donc qu'une fraction des coûts sanitaires associés à la chaleur, tel qu'aussi noté à la section 1.3.2 pour le fardeau sanitaire. En outre, plusieurs études ne se sont intéressées qu'aux troubles directement liées à la chaleur, par exemple le coup de chaleur (Jagai et al., 2017; Limaye et al., 2019; Merrill et al., 2008; Noe et al., 2012), ce qui est problématique étant donné la multitude d'autres maladies pouvant être affectées par la chaleur et la codification souvent déficiente des problèmes de santé

liés à la chaleur. Quant aux études effectuant des projections, celles existantes n'ont projeté les coûts de santé que pour une seule issue sanitaire, par exemple la mortalité (Cheng et al., 2018; Díaz et al., 2019), les hospitalisations (Hübler et al., 2008; Lin et al., 2012; Tong et al. 2021ab; Wondmagegn et al. 2021b), les admissions à l'urgence (Lay et al., 2018; Toloo et al., 2015; Tong et al. 2021c) ou les transports ambulanciers (Campbell et al., 2023; Oka et al., 2025). Ainsi, ces études présentent un portrait incomplet des coûts de santé de la chaleur et sous-estiment vraisemblablement ses impacts économiques totaux à venir.

Ces limites quant à la littérature internationale s'appliquent aussi dans un contexte québécois et canadien où les études y sont limitées. Par exemple, le premier rapport sur le sujet ne considère que la mortalité pour la ville de Montréal (NRTEE, 2011). Un second rapport a étudié l'ensemble du Québec, mais n'a modélisé que les accidents de travail et les coûts reliés au déploiement des équipes d'urgence pour ce qui est de la morbidité (Larrivée et al., 2015). Finalement, bien qu'un rapport canadien récent ait documenté les coûts de la mortalité et des hospitalisations liées à la chaleur, les modèles utilisés pour les hospitalisations étaient basés sur des données de l'Ontario puis appliquées à l'ensemble du Canada, ce qui pourrait limiter leur applicabilité au contexte du Québec (Clark et al., 2021). Ainsi, il est évident que de nouvelles études sur les coûts actuels totaux de la chaleur sont nécessaires, incluant à la fois les coûts de la mortalité et de la morbidité, et qui ne sont pas restreints qu'aux vagues de chaleur historiques. De plus, des projections de ces coûts sanitaires sont aussi nécessaires afin d'avoir un portrait global de la situation à venir et faciliter la mise en place des bonnes mesures d'atténuation et d'adaptation aux changements climatiques.

1.4 Objectifs de la thèse et résumé de la méthodologie

Cette thèse propose d'utiliser des méthodes novatrices adaptées au contexte de la santé environnementale afin de combler les lacunes identifiées dans la littérature à la section 1.3. Pour ce faire, des modèles de la famille de l'apprentissage automatique et profond, ainsi que ceux statistiques couramment utilisés en épidémiologie environnementale, seront utilisés (Figure 1.1). L'objectif général ainsi que les objectifs spécifiques de la thèse sont présentés dans les encadrés ci-dessous :

Objectif général de la thèse

- Mettre à profit les approches de la science des données les plus appropriées afin de mieux modéliser les impacts sanitaires et économiques de la chaleur au Québec.

Objectifs spécifiques

- **Objectif 1** : Introduire puis comparer différents modèles d'apprentissage automatique et profond afin de modéliser les impacts à la santé de la chaleur.

Cet objectif est réalisé en trois sous-objectifs :

- **Objectif 1a** : Explorer une multitude de modèles d'apprentissage automatique, d'apprentissage profond et statistiques pour modéliser la mortalité liée à la chaleur dans la région de Montréal
- **Objectif 1b** : Analyser une grande base de prédicteurs météorologiques et atmosphériques avec l'apprentissage automatique afin de mieux comprendre leur impact sur la mortalité estivale dans deux régions du Québec
- **Objectif 1c** : Établir des modèles d'apprentissage automatique et profond pour modéliser plusieurs issues sanitaires liées à la chaleur dans un ensemble de régions simultanément
- **Objectif 2** : Quantifier le fardeau sanitaire de la chaleur au Québec pour différentes issues sanitaires, incluant la mortalité et plusieurs variables de morbidité
- **Objectif 3** : Estimer les coûts sanitaires associés à la chaleur au Québec, actuels et futurs, en contexte de changements climatiques

Les objectifs spécifiques 1 à 3 sont décrits plus en détail aux sections 1.4.1 à 1.4.3 ci-dessous, accompagnés de leur originalité, leurs contributions ainsi qu'un résumé de leur méthodologie. Les liens entre les objectifs sont expliqués à la section 1.5 ainsi qu'au début de chaque chapitre (article) de la thèse.

1.4.1 Objectif 1 : Modélisation des impacts à la santé de la chaleur avec apprentissage automatique et profond

Originalité et contributions

Le premier objectif se situe dans un domaine de recherche encore peu exploré, où seules quelques études (revues à la section 1.2.3) se sont intéressées à comparer différentes approches d'apprentissage automatique, profond et statistiques pour les modéliser les impacts sanitaires de la chaleur. Notre hypothèse est que les modèles d'apprentissage automatique et profond performeront mieux que les modèles statistiques plus traditionnellement utilisés en épidémiologie environnementale. Comme ce domaine est très récent, cet objectif est réalisé de manière progressive sous la forme de trois sous-objectifs (un chapitre par sous-objectif) qui couvrent des problématiques différentes de la littérature soulevées à la section 1.3.1. Les deux premiers sous-objectifs (1a et 1b) sont de nature préparatoire, alors que le troisième sous-objectif (1c) est la contribution principale pour l'objectif 1 de la thèse. Le sous-objectif 1a permet de mieux comprendre les différents modèles (un ensemble et pas un seul) en vue des sous-objectifs 1b et 1c, alors que le sous-objectif 1b permet d'analyser les prédicteurs météorologiques et atmosphériques les plus pertinents en vue du sous-objectif 1c. Bien que présentés comme une suite logique, ces trois sous-objectifs demeurent des études de cas distinctes au niveau des issues sanitaires étudiées, des prédicteurs utilisés, des modèles retenus et des régions à l'étude. Cela fait en sorte que les résultats obtenus pour chacun des sous-objectifs ne sont pas directement comparables entre eux. L'originalité et les contributions principales des trois sous-objectifs sont décrites dans les paragraphes ci-dessous et résumées au Tableau 1.1.

Au premier sous-objectif (1a), neuf approches d'apprentissage automatique, profond et statistiques sont considérées et comparées entre elles, la littérature actuelle ne considérant souvent qu'un seul ou peu de modèles. Les données utilisées sont de longues séries temporelles de 20 à 40 années de données journalières incluant la mortalité, la température et la pollution de l'air, un jeu de données plus intéressant pour l'apprentissage automatique et profond que dans les études passées. De plus, l'intégration de la pollution de l'air permet de tester pour un effet potentiellement synergique avec la température, alors que celle-ci n'a été que peu utilisée lors

des récentes applications de l'apprentissage automatique. La calibration des modèles est effectuée de manière transparente et robuste, en décrivant les hyperparamètres testés et retenus, ainsi que la procédure d'entraînement et de validation, ce qui n'est pas toujours le cas dans la littérature. Afin de limiter l'effet boîte noire des modèles d'apprentissage automatique et profond, leur évaluation se fait par plusieurs critères de performance ainsi que par l'importance des prédicteurs. Cela permet d'améliorer la transparence de l'utilisation de l'apprentissage automatique dans ce domaine pour les non-experts. Cette étude est la première dans la littérature à comparer une telle variété d'approches pour modéliser les impacts à la santé de la chaleur.

Le second sous-objectif (1b) étend l'analyse des prédicteurs aux valeurs minimales, moyennes et maximales de la température quotidienne ainsi que des indices composites comme l'humidex, en plus de tester des variables de variations de température et d'autres variables météorologiques. Cela se distingue du sous-objectif 1a et des autres études dans la littérature qui ne considèrent souvent qu'un seul indice de température, en permettant de mieux comprendre comment la température affecte la santé. Encore une fois, la pollution de l'air est intégrée aux modèles. Afin d'identifier les prédicteurs les plus pertinents, les résultats de 5 modèles d'apprentissage automatique, profond et statistiques sont comparés entre eux, alors que ce genre d'exploration dans la littérature est souvent effectuée avec un seul modèle. Une calibration itérative des modèles est d'ailleurs introduite, c'est-à-dire en testant les modèles séparément avec différents sous-groupes de prédicteurs, ce qui permet de démontrer les prédicteurs les plus importants en termes de performances et guider la sélection de modèles plus parcimonieux en vue de nouveaux systèmes d'alerte basés sur l'apprentissage automatique. Finalement, ce sous-objectif étend la région d'étude à deux RMR (Québec et Montréal). Ces travaux sont les premiers à prendre en compte un éventail aussi large de variables environnementales et de modèles afin de cibler les meilleurs prédicteurs pour prédire les impacts sanitaires de la chaleur dans deux régions.

Le troisième et dernier sous-objectif (1c) vise à inclure plusieurs régions simultanément au sein d'un même modèle d'apprentissage automatique ou profond, en incluant des variables régionales afin de représenter les différences démographiques, socioéconomiques, d'environnement bâti et climatologiques entre les régions. Ainsi, ces modèles peuvent tirer de l'information des données de diverses régions afin d'améliorer potentiellement leur performance. Cette recherche se démarque des études passées et de celles des sous-objectifs 1a et 1b où une seule région est étudiée ou différentes régions sont étudiées de manière indépendante (c'est-à-dire un modèle par région). Elle est aussi originale au niveau des indicateurs de santé analysés, c'est-à-dire 5

issues sanitaires incluant la mortalité et plusieurs variables de morbidité, soit les mêmes issues que celles utilisés par l'INSPQ dans les bilans des vagues de chaleur extrême et dans le système de surveillance des événements météorologiques extrêmes. D'ailleurs, en plus de modéliser les impacts à la santé de la chaleur dans la population générale, ce sous-objectif propose aussi l'étude d'une population plus vulnérable à la chaleur, soit les personnes âgées de 65 ans et plus. Comme aux sous-objectifs 1a et 1b, nous considérons différents modèles d'apprentissage automatique, profond et statistiques. Leur calibration et leur évaluation se font encore une fois de manière transparente afin de se démarquer des études de la littérature. Il s'agit des premiers travaux au niveau international à modéliser plusieurs issues de santé liées à la chaleur dans de multiples régions simultanément, et de comparer différents modèles d'apprentissage automatique et profond à cette fin.

Tableau 1.1 Résumé des trois sous-objectifs (1a, 1b et 1c) de l'objectif 1 de la thèse, ainsi que l'identification des contributions et originalités principales et secondaires.

	Région	Issue sanitaire	Population	Modèles	Prédicteurs
Sous-objectif 1a (modèles)	RMR de Montréal	Mortalité (Toutes causes)	Générale	Stat. : GLM, GAM, DLNM Arbres : DT, RF, GBM Réseaux : SLP, MLP, LSTM	Tmoy, humidité et vent Pollution de l'air
Sous-objectif 1b (prédicteurs)	RMR de Montréal et Québec	Mortalité (Toutes causes)	Générale	Stat. : GAM Arbres : RF, GBM Réseaux : MLP, LSTM	Température (Tmoy, Tmax, Tmin, Humidex, point de rosée) Δ Température Autres variables météo. Pollution de l'air
Sous-objectif 1c (issues et régions)	Toutes les RSS du Québec méridional (n=15)	Mortalité Hospitalisations Visites à l'urgence Transports ambulanciers Appels au 811 (Toutes causes)	Générale Personnes âgées de 65 ans et plus	Stat. : Lasso Arbres : RF, LGBM Réseaux : MLP, LSTM	Météorologiques (températures, humidité et Δ température) Régionales (démographie, socio-économie, environnement bâti, climatologie) Temporelles (jours de la semaine et fériés, mois, année)

	Contribution et originalité principales du sous-objectif
	Contribution et originalité secondaires du sous-objectif

DLNM = Modèle non-linéaire à délais distribués. DT = Arbre de décision. GAM = Modèle additif généralisé. GLM = Modèle linéaire généralisé. GBM = *Gradient Boosting Model*. Lasso = *Least Absolute Shrinkage and Selection Operator*. LGBM = *Light Gradient Boosting Model*. LSTM = *Long Short-Term Memory*. MLP = Perceptron Multicouche. RF = Forêt aléatoire. RMR = Région métropolitaine de recensement. RSS = Région sociosanitaire. SLP = Perceptron à une couche. Tmax = Température maximale. Tmin = Température minimale. Tmoy = Température moyenne.

Résumé de la méthodologie

Pour l'objectif 1, différents modèles d'apprentissage automatique, profond et statistiques sont utilisés, à savoir les méthodes basées sur les arbres comme l'arbre de décision (DT), la forêt aléatoire (RF) et le *gradient boosting model* (GBM), les réseaux de neurones comme le

perceptron à une couche (SLP), le perceptron multicouche (MLP) et *Long Short-Term Memory* (LSTM), en plus d'approches statistiques traditionnellement utilisées en épidémiologie environnementale comme le GLM, le GAM ou le DLNM. Ces modèles sont décrits plus en détail à l'annexe de cette thèse. Ils sont utilisés afin de modéliser en période estivale (c'est-à-dire durant les mois de mai à septembre) diverses issues sanitaires incluant la mortalité ou la morbidité, toutes causes confondues, à l'aide de différents prédicteurs météorologiques et de pollution de l'air, selon leur disponibilité. Les mois d'hiver ont été exclus de la période d'analyse comme cette thèse porte spécifiquement sur la chaleur (et non le froid). L'inclusion des mois froids aurait complexifié et pu venir biaiser l'analyse subséquente des résultats, par exemple en termes de performances et de variables les plus importantes obtenues. Des délais (*lags*) jusqu'à 7 jours des prédicteurs sont considérés dans tous les modèles. Les hyperparamètres optimaux des modèles sont trouvés via une recherche en grille avec des données de validation. Après inspection visuelle des distributions des variables réponses (résultat non présenté), la calibration des modèles a pu s'effectuer en utilisant une distribution normale. La performance est calculée sur un jeu de données de test (hors échantillon) avec des critères comme le coefficient de détermination (R^2), la racine de l'erreur quadratique moyenne (*root mean square error*, RMSE) et l'erreur absolue moyenne (*mean absolute error*, MAE). Finalement, l'importance des prédicteurs est calculée avec la méthode de la permutation des prédicteurs ou de la pureté des arbres. Les détails méthodologiques spécifiques à chaque sous-objectif sont décrits ci-dessous et résumés au Tableau 1.1.

Le sous-objectif 1a vise spécifiquement à modéliser la déviation de mortalité (c'est-à-dire la mortalité au-delà ou au-dessous de la valeur attendue pour chaque journée de l'été) dans la RMR de Montréal pour la période 1998 à 2019 comme analyse principale et la période 1981 à 2019 en analyse de sensibilité. La déviation de mortalité est calculée en suivant la méthode de Zhang et al. (2012; 2014), c'est-à-dire en modélisant la mortalité avec une *spline* naturelle cubique en fonction du jour et de l'année au sein d'un GLM Poisson, et en calculant le résidu après avoir ajusté ce modèle. Neuf approches de modélisation sont utilisées et comparées, à savoir le DT, le RF, le GBM, le SLP, le MLP, le LSTM, le GLM, le GAM et le DLNM. Les prédicteurs utilisés sont la température moyenne quotidienne, l'humidité relative, la vitesse des vents, l'ozone (O_3), les particules fines ($PM_{2.5}$), le dioxyde d'azote (NO_2), le dioxyde de soufre (SO_2) et le monoxyde de carbone (CO). Les données météorologiques et de pollution de l'air proviennent des stations d'Environnement et Changement Climatique Canada (ECCC). Afin d'évaluer les différentes approches, les critères de performance et l'importance des variables sont calculés, en plus d'analyser visuellement les prévisions des modèles de 2010 à 2019 lors des périodes de chaleur

extrême amenant de fortes surmortalités documentées par l'INSPQ (p. ex. 2010 et 2018). Cette validation supplémentaire est produite en agrégeant les prévisions quotidiennes à l'échelle hebdomadaire afin de mieux distinguer les pics de mortalité associés aux vagues de chaleur qui durent souvent près d'une semaine. La méthodologie détaillée est disponible à la section 2.2.

Le sous-objectif 1b s'intéresse aussi à la modélisation de la déviation de mortalité (calculée comme au sous-objectif 1a), mais pour les RMR de Montréal et de Québec. Différentes variables de température sont considérées de manière univariée (c'est-à-dire une variable à la fois), comme la température de l'air, l'indice humidex et le point de rosée, avec des agrégations par la moyenne, la valeur minimale ou maximale quotidienne. Des changements de température comme l'écart diurne de température (*diurnal temperature range*, DTR) et le changement moyen de températures sur quelques jours (*mean temperature change*, MTC) sont aussi inclus. D'autres variables météorologiques comme les précipitations, la visibilité et la pression atmosphérique sont aussi incluses, en plus de l'humidité, de la vitesse des vents et de la pollution de l'air comme au sous-objectif 1a. Les données des stations d'ECCC sont aussi utilisées pour ce sous-objectif. Au total, 17 sous-groupes de prédictors sont créés et testés avec les 5 modèles d'apprentissage automatique, profond et statistiques les plus pertinents identifiés au sous-objectif 1a, à savoir le GAM, le RF, le GBM, le MLP et le LSTM. La performance hors échantillon et l'importance des prédictors permet de conclure sur les meilleurs modèles et variables pour les deux RMR à l'étude. La méthodologie détaillée est disponible à la section 3.2.

Le sous-objectif 1c porte sur chacune des 15 RSS du Québec méridional intégrée au sein d'un même modèle. Cinq issues sanitaires quotidiennes par RSS (mortalité, hospitalisations, visites à l'urgence, transports ambulanciers et appels à Info-Santé 811), toutes causes confondues, chez deux populations (générale et 65 ans et plus), sont modélisées. Les taux bruts de mortalité ou de morbidité (par 10 000 habitants) sont modélisés, et non les déviations comme aux sous-objectifs 1a et 1b. Cela permet de tester le potentiel des modèles d'apprentissage automatique et profond pour modéliser directement les tendances temporelles contenues dans ces issues sanitaires, tout en limitant les étapes de préparation des données sanitaires. Les prédictors météorologiques les plus pertinents identifiés au sous-objectif 1b, selon leur disponibilité, sont utilisés dans les modèles de manière multivariée (c'est-à-dire en les considérant tous simultanément). Afin d'avoir des données météorologiques les plus complètes possibles pour toutes les RSS du Québec, les données quotidiennes de Daymet de la NASA sont utilisées plutôt que celles d'ECCC aux sous-objectifs 1a et 1b. De plus, la pollution de l'air n'a pas pu être incluse car les données n'étaient complètes que pour quelques RSS. Des variables régionales (démographiques,

socioéconomiques, d'environnement bâti et climatologiques) et temporelles (p. ex. jour de la semaine, mois, année) sont aussi incluses afin de tenir compte des tendances spatio-temporelles dans les taux bruts des issues sanitaires modélisées. Les modèles considérés sont le RF, le *Light Gradient Boosting Model* (LGBM), plus rapide et efficient que la version GBM utilisée précédemment, le MLP et le LSTM. Ces modèles sont comparés à une régression pénalisée de type Lasso, permettant de faire une sélection des variables parmi toutes celles considérées. La performance hors échantillon est analysée sur l'ensemble des RSS, par RSS ainsi que visuellement pendant les journées des vagues de chaleur de 2018 et de 2021. Finalement, l'importance des variables permet d'identifier les associations d'intérêt entre les différentes issues sanitaires et les prédicteurs chez deux populations. La méthodologie détaillée est disponible à la section 4.2.

1.4.2 Objectif 2 : Quantification du fardeau sanitaire de la chaleur

Originalité et contributions

Ce second objectif émerge dans le cadre d'une littérature limitée sur le fardeau sanitaire de la chaleur, c'est-à-dire la quantification de l'ensemble des impacts à la santé incluant la mortalité et la morbidité. Pour combler cette lacune, nous mettons à profit le modèle DLNM, qui est le modèle le plus reconnu à ce jour pour modéliser les relations température-santé en épidémiologie environnementale, afin de calculer ce fardeau dans la province de Québec. Ainsi, bien que le premier objectif de la thèse ait exploré en détail les modèles d'apprentissage automatique et profond, ces modèles ne sont pas réutilisés pour la suite de la thèse, car les objectifs sont différents (i.e., faire des prévisions à l'objectif 1 vs quantifier le fardeau sanitaire et les coûts aux objectifs 2 et 3). Le DLNM avait d'ailleurs été utilisé et comparé au sous-objectif 1a, mais n'avait pas été retenu aux sous-objectifs 1b et 1c, car il n'était pas sensiblement différent du GAM en mode prévisionnel. Ici, le DLNM est plus pertinent à utiliser que le GAM car il permet de modéliser explicitement la relation à délai (ainsi que non linéaire comme pour le GAM) entre les températures estivales et les différentes issues sanitaires grâce au terme croisé exposition-délai.

Ce second objectif étend donc l'application du DLNM au-delà de la mortalité en étudiant aussi le fardeau sanitaire de différents indicateurs de morbidité, la littérature internationale s'étant présentement concentrée sur le fardeau de la mortalité (voir la section 1.3.2). Le DLNM est donc appliqué à 5 issues sanitaires pour les 15 RSS du Québec méridional. Les résultats des DLNM régionaux sont mis en commun au sein d'une méta-régression multivariée, incluant différents prédicteurs régionaux tels qu'utilisés au sous-objectif 1c. Cette analyse du fardeau sanitaire pour

autant d'indicateurs de santé incluant la mortalité et la morbidité est une première au Québec, ainsi qu'une des premières dans la littérature internationale. Notre hypothèse est que toutes les issues sanitaires étudiées augmenteront à des températures plus élevées, représentant ainsi un fardeau sanitaire important pour le Québec.

Résumé de la méthodologie

Les données sanitaires utilisées sont les mêmes qu'à l'objectif 1c, c'est-à-dire les issues de mortalité et de morbidité quotidiennes toutes causes dans la population générale par RSS telles que fournies par l'INSPQ. La variable d'exposition utilisée est la température moyenne quotidienne avec un délai (*lag*) jusqu'à 7 jours. Une seule variable d'exposition est utilisée (cadre univarié), car inclure plusieurs métriques de température fortement corrélées apporterait des problèmes de multicollinéarité dans le modèle. Le DLNM contrôle pour l'humidité relative, les tendances quotidiennes, saisonnières et à long terme. Afin d'avoir des données complètes pour toutes les RSS, les données météorologiques utilisées proviennent de Daymet tout comme au sous-objectif 1c. L'analyse se fait en 3 étapes pour chaque issue sanitaire, c'est-à-dire 1) l'estimation du modèle DLNM de la relation bidimensionnelle (i.e., non-linéaire et à délai) entre l'indicateur de santé et la température pour chaque RSS, puis la réduction de ces fonctions en une seule dimension en intégrant sur l'espace des délais, 2) la méta-régression multivariée pour mettre en commun les résultats des DLNM régionaux, en modélisant les coefficients associés aux fonctions réduites (et leur structure de variance-covariance) à l'aide des mêmes prédicteurs régionaux qu'au sous-objectif 1c, 3) le calcul des nombres et fractions attribuables à la chaleur (toutes températures au-delà de la MMT) et à la chaleur extrême (au-delà du quantile 95% de température) à l'aide des fonctions obtenues après la méta-régression multivariée. Les résultats sont présentés sur l'ensemble des RSS et de la période, mais aussi par RSS et par année. Des analyses de sensibilité sont effectuées avec les régions qui ont des données de pollution de l'air, pour d'autres variables d'exposition comme la température minimale ou maximale, l'indice humidex et le point de rosée, et pour les mois utilisés pour définir la période estivale, entre autres. La méthodologie détaillée est disponible à la section 5.2.

1.4.3 Objectif 3 : Estimation et projection des coûts sanitaires de la chaleur

Originalité et contributions

Cet objectif s'inscrit dans un contexte où il manque d'études estimant les coûts sanitaires totaux, actuels et futurs, associés à la chaleur au Québec comme à l'international. Pour combler cette lacune, nous proposons un cadre afin d'analyser les coûts sanitaires de la chaleur en climat

historique et projeté, mettant à profit les fonctions dose-réponse obtenues à l'aide du DLNM et de la méta-régression à l'objectif 2. Les coûts sanitaires de la chaleur analysés sont liés à la mortalité, la morbidité, l'absentéisme et la perte de bien être lors des vagues de chaleur extrême. Cette quantification est donc plus exhaustive que dans les études précédentes qui n'ont souvent touché qu'à une seule issue sanitaire. De plus, notre étude ne se limite pas seulement aux vagues de chaleur extrême, qui ne représentent d'une fraction du fardeau de la chaleur, mais plutôt à l'ensemble des coûts de la chaleur tout au long de l'été comme à l'objectif 2. En outre, nous analysons aussi les coûts futurs (projetés) pour ces mêmes catégories de coûts avec un ensemble de 15 modèles climatiques mis à l'échelle pour l'Amérique du Nord, ainsi que des projections démographiques par RSS correspondant aux scénarios du GIEC. Il s'agit de l'étude la plus exhaustive effectuée à ce jour sur les coûts sanitaires de la chaleur en climats historique et futur. Nous nous attendons à ce que les coûts sanitaires de la chaleur soient déjà importants à l'heure actuelle, et que ceux-ci augmentent en contexte de changements climatiques et démographiques.

Résumé de la méthodologie

Les fonctions dose-réponse de la relation entre la chaleur et la santé sont reprises depuis les résultats de l'objectif 2 mettant à profit le DLNM et la méta-régression. Celles-ci sont combinées à des projections climatiques provenant du consortium Ouranos et des projections démographiques préparées par Statistique Canada permettant de calculer le fardeau sanitaire futur lié à la chaleur au Québec jusqu'en 2070. Pour compléter l'analyse des effets sanitaires futurs, nous calculons aussi le nombre de vagues de chaleur extrême, telles que définies par l'INSPQ, et le nombre de personnes exposées à ces vagues, en périodes historique et future. Les coûts sanitaires incluent les coûts directs de soins de santé (i.e., hospitalisations, urgences, ambulances, appels 811 et équipes d'urgence lors des vagues de chaleur extrême), les coûts indirects d'absentéisme pour recevoir les soins médicaux et les coûts intangibles liés aux pertes de vies humaines et de bien être lors des vagues de chaleur extrême. Des métriques économiques de sources officielles et de la littérature sont utilisées pour faire la conversion du fardeau sanitaire en coûts sanitaires, par exemple le coût moyen et la durée moyenne d'une hospitalisation au Québec, la valeur statistique de la vie au Canada, le salaire médian horaire au Québec, etc. Finalement, des simulations de type Monte Carlo permettent de voir l'évolution de la variabilité des coûts estimés, incluant l'incertitude dans les fonctions dose-réponse, la variabilité des modèles climatiques et les intervalles de coûts de santé possibles. La méthodologie détaillée est disponible à la section 6.2.

1.5 Structure de la thèse et liens entre les chapitres

Cette thèse se divise en 7 chapitres. Le chapitre 1 contient l'introduction, alors que le chapitre 7 contient la conclusion et les perspectives. Les chapitres 2 à 6 contiennent les cinq articles qui constituent le corps de cette thèse (un article par chapitre) et répondent respectivement aux objectifs et sous-objectifs 1a, 1b, 1c, 2 et 3 décrits ci-dessus. De surcroît, ces différents chapitres permettent de répondre à l'objectif général de la thèse qui est de mettre à profit les approches les plus appropriées de la science des données afin de mieux modéliser les impacts sanitaires et économiques de la chaleur au Québec. Selon l'objectif poursuivi, la famille d'approches et les modèles sous-jacents les plus adaptés sont donc utilisés. Cette thèse est la première à comparer les modèles d'apprentissage automatique à ceux statistiques couramment utilisés en épidémiologie environnementale afin de modéliser les impacts à la santé de la chaleur. Les différences principales entre ces deux grandes familles d'approches sont détaillées au Tableau 1.2.

Tableau 1.2 Comparaison des approches de la grande famille de la science des données utilisées dans le cadre de cette thèse.

	Science des données	
	Apprentissage automatique	Épidémiologie environnementale
Résultat attendu	Modèle de prévision	Fonctions dose-réponse
Exemples d'approches	DT, RF, GBM, MLP, LSTM	GLM*, GAM*, DLNM*, méta-régression
Méthodes d'évaluation	Critères de performance, importance des variables	Risque relatif, intervalles de confiance, AIC
Exemples d'utilisation en santé publique	Prévision des impacts à court terme (vigie), identification des variables et interactions d'intérêt	Calcul du fardeau sanitaire (surveillance), projection du fardeau et des coûts associés
Principaux chapitres de la thèse	Chapitres 2, 3 et 4	Chapitres 5 et 6
*Les modèles GLM, GAM et DLNM sont aussi utilisés aux chapitres 2, 3 ou 4 afin de faire la prévision des impacts sanitaires de la chaleur à des fins de comparaison avec les modèles d'apprentissage automatique.		

AIC = Critère d'information d'Akaike. DLNM = Modèle non-linéaire à délais distribués. DT = Arbre de décision. GAM = Modèle additif généralisé. GBM = *Gradient Boosting Model*. GLM = Modèle linéaire généralisé. LSTM = *Long Short-Term Memory*. MLP = Perceptron Multicouche. RF = Forêt aléatoire.

La Figure 1.2 décrit les liens entre les différents chapitres de la thèse, que ce soit au niveau des données mises à profit ou des méthodes de la science des données utilisées. Les chapitres 2, 3 et 4 de la thèse portent principalement sur les modèles de la famille de l'apprentissage automatique et profond, alors que les chapitres 5 et 6 s'intéressent plutôt aux modèles statistiques couramment utilisés en épidémiologie environnementale. Brièvement, le chapitre 2 permet d'introduire différentes approches d'apprentissage automatique et profond pour modéliser les impacts à la santé de la chaleur et de les comparer aux modèles statistiques plus traditionnels, avec l'exemple de la mortalité dans la RMR de Montréal. Les approches les plus pertinentes sont ensuite reprises aux chapitres 3 et 4. Le chapitre 3 étend le chapitre 2 avec une plus grande base de données de prédicteurs météorologiques et l'application de l'apprentissage automatique aux RMR de Montréal et de Québec. Il permet de conclure sur les prédicteurs les plus intéressants à utiliser, permettant ainsi de guider les chapitres 4 et 5. Le chapitre 4 met à profit les connaissances générées aux chapitres 2 et 3 en apprentissage automatique et profond afin de proposer des modèles s'adaptant à l'ensemble des 15 RSS du Québec méridional et à cinq issues sanitaires incluant la mortalité et plusieurs variables de morbidité.

Les connaissances développées aux chapitres 2, 3 et 4 quant aux variables sanitaires et environnementales ainsi que sur les impacts à la santé de la chaleur guident les chapitres 5 et 6 qui suivent, bien que les approches utilisées diffèrent. La base de données utilisées au chapitre 4 est reprise intégralement au chapitre 5, ce qui permet de mettre en évidence les différences entre les deux familles d'approches au niveau des buts visés et des résultats attendues avec une base de données commune (revoir le Tableau 1.2). Au chapitre 5, le DLNM et la méta-régression sont mis à profit, utilisés traditionnellement pour le calcul du fardeau sanitaire que les modèles d'apprentissage automatique des chapitres 2, 3 et 4. Finalement, le chapitre 6 reprend les fonctions dose-réponse développées au chapitre 5 à l'aide du DLNM et de la méta-régression, et les utilise avec des projections climatiques et démographiques ainsi que des métriques économiques, afin de calculer les coûts sanitaires historiques et futurs associés à la chaleur dans le contexte québécois.

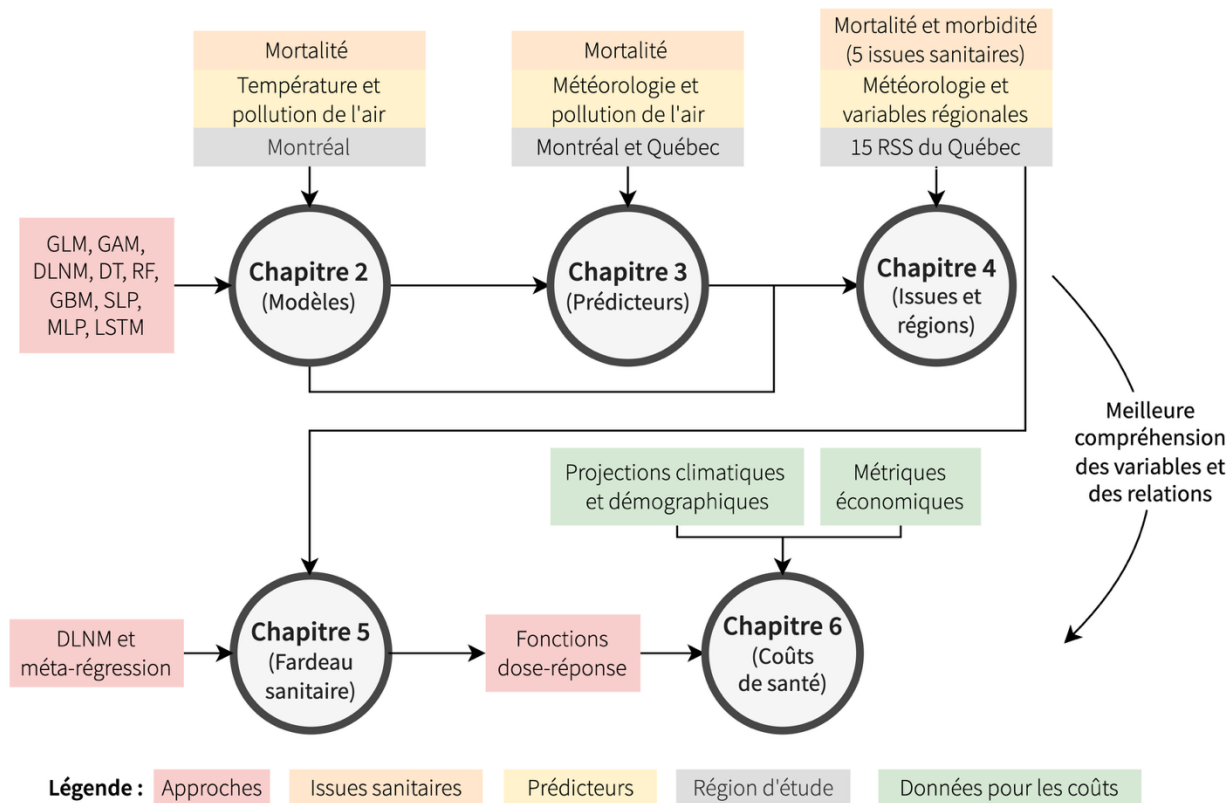


Figure 1.2 **Aperçu des chapitres 2 à 6 de la thèse et des liens entre les chapitres.** Les approches de la science des données utilisées sont en rouge (voir le Tableau 1.2 pour la comparaison entre les différentes familles de modèles), les issues sanitaires en orange, les prédicteurs en jaune, la région d'étude en gris, et les données pour la projection des coûts en vert.

DLNM = Modèle non-linéaire à délais distribués. DT = Arbre de décision. GAM = Modèle additif généralisé. GLM = Modèle linéaire généralisé. GBM = *Gradient Boosting Model*. LSTM = *Long Short-Term Memory*. MLP = Perceptron multicouche. RF = Forêt aléatoire. RSS = Région sociosanitaire. SLP = Perceptron à une couche.

CHAPITRE 2 :
MACHINE AND DEEP LEARNING FOR MODELLING
HEAT-HEALTH RELATIONSHIPS

2 MACHINE AND DEEP LEARNING FOR MODELLING HEAT-HEALTH RELATIONSHIPS

Modèles d'apprentissage automatique et profond pour modéliser les relations entre la chaleur et la santé

Auteurs : Jérémie Boudreault^{1,2}, Céline Campagna^{1,2}, Fateh Chebana¹

¹ Centre Eau Terre Environnement, Institut national de la recherche scientifique (INRS), 490 de la Couronne, Québec (QC) Canada, G1K 9A9.

² Direction de la santé environnementale, au travail et de la toxicologie, Institut national de santé publique du Québec (INSPQ), 945 avenue Wolfe, Québec (QC) Canada, G1V 5B3.

Titre de la revue : *Science of the Total Environment*

Publié le 5 juin 2023

<https://doi.org/10.1016/j.scitotenv.2023.164660>

Contribution des auteurs :

J. Boudreault a été le principal contributeur de cet article, avec la conceptualisation de l'étude, la méthodologie, l'analyse des données, la modélisation, la production des figures et tableaux, et l'écriture du manuscrit. C. Campagna et F. Chebana ont aidé à conceptualiser l'étude et définir la méthodologie. Ils ont aussi validé les résultats, révisé l'article et approuvé sa version finale.

Abstract

Extreme heat events pose a significant threat to population health that is amplified by climate change. Traditionally, statistical models have been used to model heat-health relationships, but they do not consider potential interactions between temperature-related and air pollution predictors. Artificial intelligence (AI) methods, which have gained popularity for health applications in recent years, can account for these complex and non-linear interactions, but have been underutilized in modelling heat-related health impacts. In this paper, six machine and deep learning models were considered to model the heat-mortality relationship in Montreal (Canada) and compared to three statistical models commonly used in the field. Decision Tree (DT), Random Forest (RF), Gradient Boosting Machines (GBM), Single- and Multi-Layer Perceptron (SLP and MLP), Long Short-Term Memory (LSTM), Generalized Linear and Additive models (GLM and GAM), and Distributed Lag Non-Linear Model (DLNM) were employed. Heat exposure was characterized by air temperature, relative humidity and wind speed, while air pollution was also included in the models using five pollutants. The results confirmed that air temperature at lags of up to 3 days was the most important variable for the heat-mortality relationship in all models. NO₂ concentration and relative humidity (at lags 1 to 3 days) were also particularly important. Ensemble tree-based methods (GBM and RF) outperformed other approaches to model daily mortality during summer months based on three performance criteria. However, a partial validation during two recent major heatwaves highlighted that non-linear statistical models (GAM and DLNM) and simpler decision tree may more closely reproduce the spike of mortality observed during such events. Hence, both machine learning and statistical models are relevant for modelling heat-health relationships depending on the end user goal. Such extensive comparative analysis should be extended to other health outcomes and regions.

Keywords: mortality, temperature, humidity, air pollution, machine learning, deep learning.

Highlights

- Heat-health relationship modelled with 9 machine/deep/statistical learning models
- Interactions considered between temperature-related and air pollution variables
- Air temperature, NO₂ and relative humidity were the most important predictors
- Ensemble tree-based models outperformed neural networks and statistical models
- Non-linear statistical models may better represent peak mortality during heatwaves

Graphical abstract

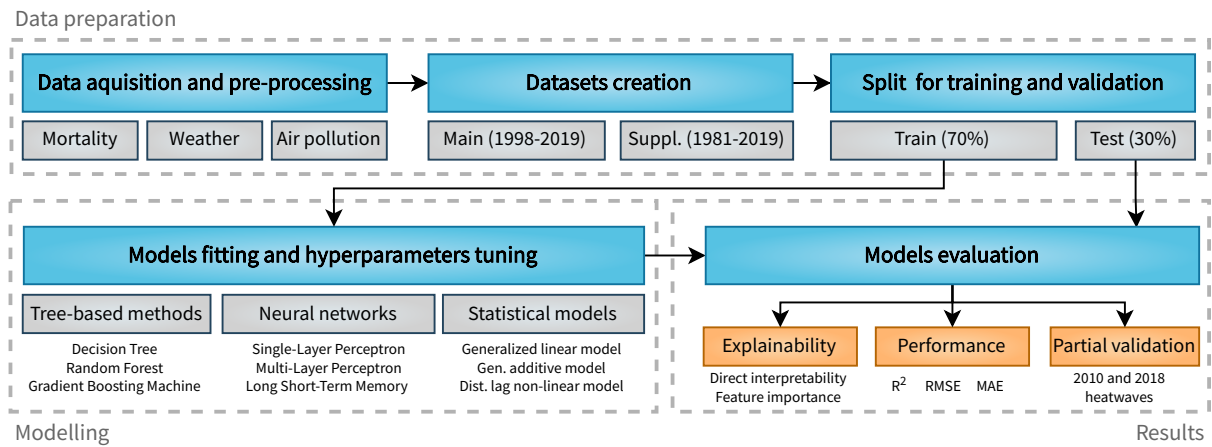


Figure 2.1 Graphical abstract.

2.1 Introduction

In its latest report (IPCC, 2021), the Intergovernmental Panel on Climate Change (IPCC) reaffirmed that climate change increases the frequency, intensity, length and spatial extent of many weather events such as extreme heat (Casati et al. 2013; Jeong et al. 2016; Meehl et Tebaldi 2004). Extreme heat events pose a significant threat to population health because of their impact on both mortality (Basu 2009; Basu et Samet 2002; Gosling et al. 2009; Xu et al. 2016) and morbidity (Li et al., 2015; Ye et al., 2012), as well as important economic consequences (Wondmagegn et al., 2019).

The heat-health relationship is commonly studied using an over-dispersed Poisson regression statistical model (Gosling et al., 2009). Non-linear approaches using either splines (e.g., Doyon et al., 2008; Ishigami et al., 2008) or Generalized Additive Models (GAM) (e.g., Bayentin et al., 2010; Lin et al., 2012; Pascal et al., 2013) are usually preferred to linear ones i.e., Generalized Linear Models (GLM) (e.g., Basu et al., 2012; Schwartz et al., 2004). To describe both the lag structure and the non-linear effect of the exposure, the Distributed Lag Non-Linear Model (DLNM) was proposed (Armstrong, 2006; Gasparrini et al., 2010) and became very popular in the last decade (e.g., Gasparrini et al., 2015, 2017; Pascal et al., 2018, 2021a; Vicedo-Cabrera et al., 2018, 2021).

The exposure to heat is characterized by a temperature variable at various lags, mainly the daily mean temperature (Son et al., 2019). Minimum or maximum temperature, as well as composite temperature indices (e.g., Humidex, Heat Index, Apparent Temperature, Wet Bulb Globe Temperature, etc.) can also be used (Barnett et al. 2010; Kovats et Hajat 2008; Tong et Kan 2011; Vaneckova et al. 2011; Zhang et al. 2014). It is still unclear if air pollution changes the heat-health relationship or not (Huang et al. 2011; Son et al. 2019). While some heat-health studies do not include air pollution (e.g., Barnett et al., 2010; Doyon et al., 2008), others include it only in their sensitivity analyses (e.g., Gasparrini et al., 2015). Because their levels are generally elevated during extreme heat events, air pollutants should be considered when studying the health effects of heat (Huang et al., 2011; Kovats & Hajat, 2008).

In the studies cited above, the temperature-related predictors and, when considered, air pollution variables, were always treated separately. Indeed, no interactions were included between these variables because of the difficulty of easily considering them in statistical models traditionally used. In contrast, machine and deep learning models, a branch of artificial intelligence (AI), can easily take these interactions into account. However, these models were only seldom used for modelling heat-health relationships (e.g., Lin et al., 2021; Masselot et al., 2021; Nishimura et al.,

2021; Ogata et al., 2021; Park & Kim, 2018; Park et al., 2020; Wang et al., 2019). In most applications, a single model was considered e.g., Random Forest (Wang et al., 2019; Zhang et al., 2014) or Multi-Layer Perceptron (Khatri & Tamil, 2017). Only a few studies have so far compared more than two approaches (Marien et al. 2022; Nishimura et al. 2021; Ogata et al. 2021; Park et al. 2020; Qiu et al. 2020). No studies have yet compared the results of machine and deep learning models with the widely used DLNM. Finally, recurrent neural networks such as the Long Short-Term Memory (LSTM) have only been seldom used (Lin et al. 2021; Nishimura et al. 2021).

In addition, the calibration of machine and deep learning models in above studies can be questioned. Models have been mostly calibrated on small datasets of <5 years (e.g., Khatri & Tamil, 2017; Ogata et al., 2021; Park et al., 2020; Qiu et al., 2020; Wang et al., 2019) or 5–10 years of data (e.g., Kassomenos et al., 2011; Lin et al., 2021; Nishimura et al., 2021; Zhang et al., 2014). Furthermore, a recent review of machine learning in public health reported that most studies failed to report their hyperparameters (Morgenstern et al., 2020). In some studies cited above, no hyperparameters optimization is performed at all (e.g., Zhang et al., 2014). Hence, subjective or unjustified choices of those hyperparameters can be expected, leading to a suboptimal fit of these models. Finally, air pollution was absent from most heat-health studies using machine and deep learning that primarily focused on temperature-related variables (e.g., Mora et al., 2017; Nishimura et al., 2021; Ogata et al., 2021; Park et al., 2020; Wang et al., 2019; Zhang et al., 2014).

By allowing potential complex interactions between temperature-related variables and air pollution, machine and deep learning can lead to better performance for modelling the heat-health relationship. In this study, six machine and deep learning models such as tree-based methods, feedforward and recurrent neural networks were compared to statistical models commonly used in the field. A transparent calibration procedure (i.e., hyperparameters optimization) with a long-enough training dataset and an adequate selection of heat-related predictors including air pollution was used for the fitting of the models.

The paper is organized as follows. Section 2.2 contains the material and methods. The results are presented in section 2.3. Section 2.4 discusses the obtained results while Section 2.5 concludes this paper.

2.2 Materials and Methods

This project received ethics approval from the Human Research Ethics Committee of the National Institute of Scientific Research (CER-22-693).

2.2.1 Data sources

The case study for analyzing the heat-health relationship with machine and deep learning focussed on the Census Metropolitan Area (CMA) of Montreal (Figure 2.2). The CMA of Montreal had a population of around 4.3 million inhabitants in 2021, which corresponds to approximately half the population of the province of Quebec, Canada (Statistics Canada, 2022).

The studied health outcome was the all-cause mortality. Mortality data from 1981 to 2019 (prior to the COVID-19 pandemic) in the CMA of Montreal was provided by the *Institut national de santé publique du Québec* (INSPQ). Seasonal and long-term trends in mortality time series were removed prior to modelling using a natural cubic spline of day of the year with 5 degrees of freedom for the seasonal trend, a linear function of year for the long-term trend and indicators for weekdays and holidays (Zhang et al., 2012; 2014). Other mortality trends were also tested (e.g., Gasparrini et al., 2010, 2015), but were not selected for further analyses as they did not differ significantly. The adjusted seasonal and long-term trends were then subtracted from the crude daily mortality to obtain the response variable of interest for the models, namely the daily mortality deviation, i.e. the over- and under-mortality relative to the expected seasonal and long-term value (Zhang et al., 2012, 2014).

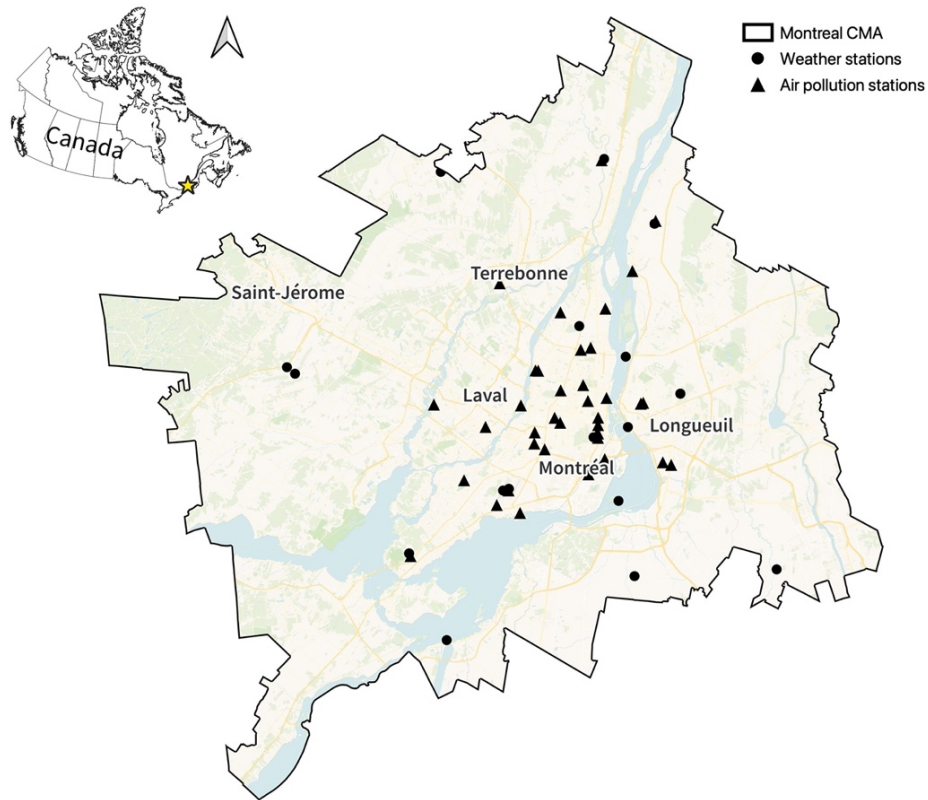


Figure 2.2 Weather and air pollution stations within the Census Metropolitan Area (CMA) of Montreal.

Weather and air pollution data were provided by Environment and Climate Change Canada (ECCC) and the National Air Pollution Surveillance (NAPS) program, respectively. Hourly data from weather and air pollution stations within Montreal CMA (Figure 2.2) were aggregated to mean daily values for the variables of interest. Temperature-related variables included air temperature, relative humidity and wind speed. These three variables are the ones commonly used to define composite temperature indices such as humidex or apparent temperature (e.g., Barnett et al., 2010; Vaneckova et al., 2011). For air pollution, the two most studied variables in heat-mortality relationships were used (Son et al., 2019), namely ozone (O_3) and particulate matter of 2.5 microns or less in diameter ($PM_{2.5}$) as PM_{10} was not available. Other air pollutants more broadly used in weather-health studies such as nitrogen dioxide (NO_2), sulfur dioxide (SO_2) and carbon monoxide (CO) were also considered (e.g., Basu et al., 2012; Goldberg et al., 2011; Lavigne et al., 2014; Wang et al., 2014).

Daily temperature-related and air pollution data at various stations were combined using a spatial mean to obtain a single time series for each variable for the entire studied region as done previously (e.g., Chiu et al., 2021; Masselot et al., 2018). Because the lagged effect of

temperature/air pollution variables on health is important (Son et al. 2019), lags up to 7 days for all predictors were considered, with the following aggregates: value at lag 0, mean of values at lags 1 to 3 days and mean of values at lags 4 to 7 days. The period May 1st to September 30th was studied for the heat-mortality relationship. The overlapping period between mortality, weather and air pollution data was 1998–2019 (Table 2.1). This led to 3366 observations for that 22-year period. As PM_{2.5} was the only variable missing for the 1981–1997 period (for which all other variables were available), another dataset, referred to as “supplementary dataset”, consisting of years 1981–2019, but without PM_{2.5} variable, was also created (Table S1). This supplementary dataset had 39 years of data for a total of 5967 observations.

Tableau 2.1 Overview of the data for modelling the heat-mortality relationship in Montreal CMA (1998–2019, May–September). All data are daily values.

	Min	Q10	Q25	Median	Mean	Q75	Q90	Max	Source
Mortality variables									
Crude mortality	37.00	56.00	61.00	68.00	68.37	75.00	81.00	135.00	INSPQ
Mortality deviation	-29.32	-12.07	-7.01	-1.06	-0.77	5.23	10.75	65.66	INSPQ
Predictors variables									
Air temperature (°C)	3.21	11.79	15.24	18.64	18.13	21.31	23.54	28.97	ECCC
Relative humidity (%)	30.88	54.78	63.44	71.16	70.45	78.44	85.64	97.35	ECCC
Wind speed (km/h)	3.23	6.45	8.01	10.33	10.79	12.97	15.76	29.70	ECCC
O ₃ concentration (ppb)	3.77	13.97	18.11	23.76	24.53	29.52	35.48	68.58	NAPS
NO ₂ concentration (ppb)	1.37	4.84	6.42	9.05	10.04	12.64	16.76	39.52	NAPS
SO ₂ concentration (ppb)	0.00	0.26	0.57	1.23	1.68	2.33	3.69	9.87	NAPS
CO concentration (ppm)	0.09	0.17	0.20	0.26	0.28	0.33	0.41	0.83	NAPS
PM _{2.5} concentration (µg/m ³)	0.33	3.47	5.15	7.66	9.08	11.51	16.25	59.08	NAPS

The datasets were split into two distinct periods following the hold-out method for time series using 70% of the first years of data for the calibration (training) of the models and the remaining 30% for the validation (test) of the models. The training datasets had respectively the years 1998–2013 and 1981–2009 for the main (with all predictors) and the supplementary (without PM_{2.5} predictor) datasets. The validation sets consisted of years 2014–2019 and 2009–2019 for the main and supplementary datasets, respectively.

2.2.2 Tree-based methods

Tree-based methods, also called classification and regression trees, come from both statistical and machine learning fields (James et al., 2013). A single tree is called a Decision Tree (DT) because of its reversed tree shape (Quinlan, 1986). DT is easy to interpret and explain, but often lack prediction accuracy (James et al., 2013). Hence, ensemble methods that consist of several trees have been proposed. Random Forest (RF) is a model in which a forest of fully grown trees is built using bootstrapped datasets of observations as well as a subset of predictors for each node in the underlying DTs (Breiman, 2001). Gradient Boosting Machines (GBM) are also based on DT but differ from RF in their construction (Friedman, 2001). Trees with fewer leaves (e.g., 1 to 5) called weak learners are grown using a sequential fitting method instead of bigger trees built in parallel in RF. The next tree in GBM is fitted on the residual of the last tree(s) given a shrinkage parameter λ also called learning rate. As in RF, GBM also uses bootstrapped datasets as well as a subset of the available predictors for the underlying trees.

DT, RF and GBM were fitted to model daily mortality deviation as a function of (lagged) temperature-related and air pollution variables. In DT, a fully grown tree was first fitted. Then, a pruning procedure was applied to decrease the generalization error (James et al., 2013). The optimal amount of pruning (i.e., the number of leaves) was found using a 5-fold cross-validation on the training set. Because we had no past evidence on the optimal tuning of RF and GBM in the context of heat-health relationships, an extensive grid search procedure was performed. This method allowed us to test various combinations of hyperparameters and draw conclusions for further analysis, even though it can be less efficient than other hyperparameter optimization methods such as random sampling (Bergstra & Bengio, 2012). DT, RF and GBM were fitted using *tree*, *randomForest* (Liaw & Wiener, 2002) and *gbm* (Greenwell et al. 2019) packages in R.

The unique hyperparameter for DT was tree depth. The hyperparameters for RF included the number of trees (500, 1000, 2500 and 5000) and the fraction of predictors considered at each tree split (square root of the number of predictors, 1/4, 1/3, 1/2, 3/4, and all predictors). The hyperparameters for GBM were learning rate (0.001, 0.01, and 0.1), tree depth (1, 3 and 5), number of trees (1000, 2500 and 5000) and the fraction of predictors at each split (1/3, 1/2 and 3/4). The hyperparameters grid search was performed using a 5-fold cross-validation on the training dataset to minimize out-of-sample root mean square error (RMSE) (Chapter 5 in Goodfellow et al., 2016).

To explain RF and GBM models, which are built using a large amount of DT (up to 5000 in our case), feature importance (FI) metrics were computed. FI is a standard method in machine

learning to know which variables contribute the more to the prediction success (Chapters 15 and 16 in Hastie et al., 2009). For RF, two FI metrics were computed using 1) node purity, based on the mean decrease in mean square error (MSE) for each predictor in all underlying trees, and 2) permutation, based on the increase in out-of-bag prediction error when one predictor is randomly shuffled. For GBM, FI was computed from the decrease in MSE for each predictor (i.e., node purity). All FI metrics were scaled to a maximum value of 1 to allow comparison between different models/metrics.

2.2.3 Neural networks

Neural networks are machine and deep learning methods inspired by the human brain. One of the simplest neural networks is the feedforward Single-Layer Perceptron (SLP). It contains three layers: an input one, a hidden one and an output one. SLP is generalized into the Multi-Layer Perceptron (MLP) (Chapter 6 in Goodfellow et al., 2016). MLP allows for more than one hidden layer and belongs, in this context, to the family of deep learning models. SLP and MLP were kept separated in this study for comparison purposes. While feedforward neural networks consider observations independently, recurrent neural networks, such as the Long Short-Term Memory (LSTM), transfer information from one cell to the other and are particularly adapted for sequential data such as time series (Chapter 10 in Goodfellow et al., 2016). LSTM extends classical recurrent networks by including a memory function and correcting the vanishing gradient problem (Hochreiter & Schmidhuber, 1997).

As for RF and GBM, no past evidence was found about the optimal tuning of neural networks in the specific context of heat-health relationships. Hence, an extensive grid search was performed. For SLP and MLP, hyperparameters included learning rate (0.0001, 0.001 and 0.01), activation function (Rectified Linear activation Unit (ReLU), hyperbolic tangent (tanh) and logistic) and scaling function for the predictors (“std” with mean and standard deviation, “robust” with median and interquartile range, and “minmax” with values scaled from 0 to 1). For SLP, the optimal number of neurons (5, 10, 20, 30 and 40) was also included in the grid search. For MLP, the number of hidden layers and neurons were both included simultaneously in the grid search with six possible combinations : 1) two (hidden) layers of 15 and 10 neurons, 2) two layers of 30 and 20 neurons, 3) three layers of 20, 15 and 10 neurons, 4) three layers of 40, 30 and 20 neurons, 5) four layers of 20, 20, 15 and 10 neurons and, 6) four layers of 40, 30, 20 and 20 neurons. Optimal hyperparameters were found by a 5-fold cross-validation on the training set to minimize

out-of-sample RMSE. SLP and MLP were fitted using *scikit-learn* in Python (Pedregosa et al., 2011).

For LSTM, the hyperparameters tuning included the learning rate (0.0001, 0.001 and 0.01), the activation function (ReLU and tanh), the addition of a dropout layer of 20% to avoid overfitting, the number of epochs (up to 10 000, with early stopping at 500, 1000, 2500 and 5000 epochs) and the number of cells/layers with five combinations : 1) one layer of 5 cells, 2) 10 cells, 3) 15 cells, 4) two layers of 10 and 5 cells and 5) 15 and 10 cells. Optimal hyperparameters were found to minimize the RMSE on a validation dataset that consisted of 30% of the last years of the training dataset. LSTM was fitted using *keras* in Python (Gulli & Pal, 2017).

FI for neural networks (SLP, MLP and LSTM) were all computed using a permutation-based metric. The mean decrease in RMSE when a predictor was randomly shuffled was computed for each predictor and repeated 200 times. The more the RMSE decreased, the more the predictor was important to the model. FI was computed on both the training and test datasets and scaled to a maximum value of 1.

2.2.4 Statistical models

Machine and deep learning methods were compared to three statistical models also used in the field: Generalized Linear Model (GLM), Generalized Additive Model (GAM) and Distributed Lag Non-Linear Model (DLNM). GLM considers linear relations between a transformation of the response variable and predictors, while allowing for non-gaussian residuals (Nelder & Wedderburn, 1972). GAM extends GLM with smooth non-linear transformations of the predictors, while keeping the property of additivity of each individual effect (Hastie & Tibshirani, 2017). DLNM describes the exposure-response relationship with a non-linear cross-basis function of lags and exposure variable (Armstrong 2006; Gasparrini et al., 2010). In DLNM, a cross-basis function of mean temperature up to 7 days was considered with the same settings as in Gasparrini et al. (2010). Other predictors were also included as non-linear effects as in GAM. All predictors in the statistical models were considered independently (i.e., without interactions), as our goal was to compare models that easily account for potential interactions (machine and deep learning) with models that generally do not (statistical). The statistical models were fitted in R using packages *mgcv* (Wood, 2015) for GLM and GAM and *dlnm* (Gasparrini, 2011) for DLNM. To compare FI metrics of machine/deep learning models with results of statistical models, FI for statistical models was also quantified by the increase of the residual sum of square when one predictor was removed from the model.

2.2.5 Models' evaluation

Models' performance was assessed using the test dataset (i.e., 30% of observations that were removed before hyperparameters selection and models fitting). Three performance metrics were computed: coefficient of determination (R^2), mean absolute error (MAE) and root mean square error (RMSE). A high value of R^2 is preferred while low values of MAE and RMSE are desired. Note that because R^2 is computed for out-of-sample (test) predictions, its value can be lower than 0, meaning that the model performed worse than a non-informative model (i.e., a model containing only an intercept). In addition to classical performance metrics, a partial validation was also performed. Predictions were compared visually to the observed mortality for the last 10 years, during which two major heatwaves occurred in the CMA of Montreal in 2010 (Bustinza et al., 2013) and 2018 (Lebel et al., 2019). This validation was added to see how the models performed during extreme heat events of interest lasting several days. Hence, daily observations and predictions were converted to weekly values, which also reduced the noise in the data for this visual assessment. Performance of the models was assessed for models fitted on the main dataset (22 years of data with all predictors), as well for models fitted on the supplementary dataset (39 years of data, without $PM_{2.5}$).

2.3 Results

The results are only presented for the main dataset (1998–2019, with all predictors) in sections 2.3.1, 2.3.2 and 2.3.3. Then, models' comparison and performance are shown for the models fitted on both datasets (main and supplementary) in sections 2.3.4 and 2.3.5.

2.3.1 Tree-based methods

For DT, the optimal tree depth was found to be 5 in cross-validation (Figure S1). The resulting pruned DT used only three predictors: mean temperature at lag 0, mean NO_2 concentration at lags 1–3 days and mean temperature at lags 1–3 days (Figure 2.3). Higher temperature and higher NO_2 concentration were both associated with increased summer mortality. For example, temperature above 22°C (at lag 0) and 26.9°C (mean of values at lags 1 to 3 days) were found as breakpoints in the heat-mortality relationship (Figure 2.3).

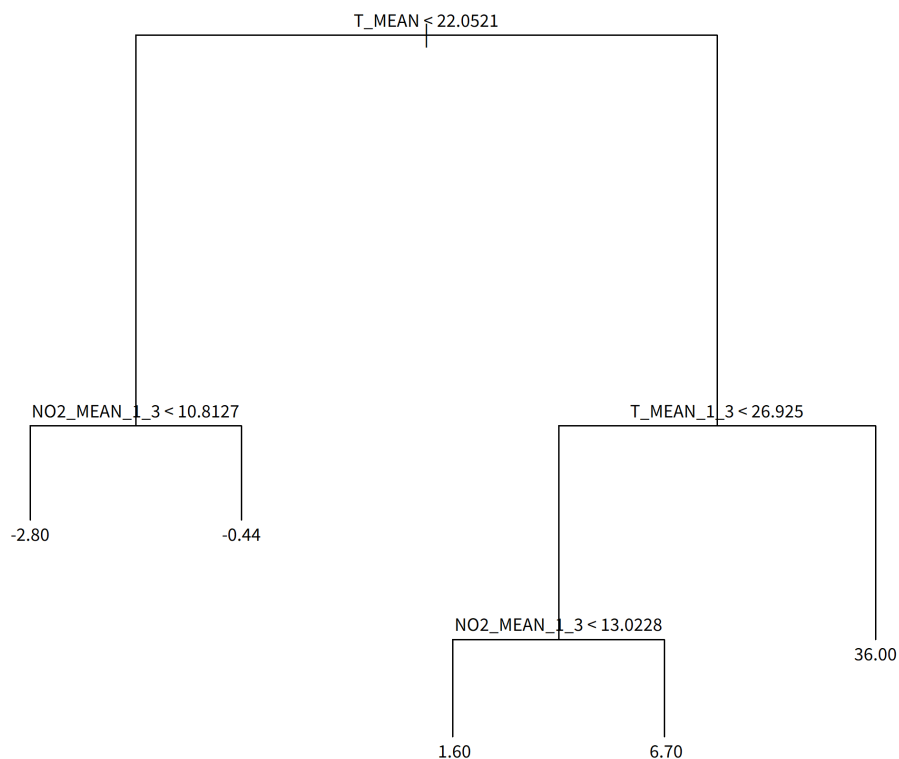


Figure 2.3 Resulting pruned Decision Tree (DT). The figure shows all predictors' splits (e.g., mean temperature <22.05°C, top of the figure) as well as the predicted daily mortality deviation at each terminal node (e.g., daily mortality deviation = -2.80 when mean temperature is <22.05 and mean NO₂ at lags 1-3 days is <10.81, bottom left of the figure).

For RF, the grid search of hyperparameters led to an optimal RF using 5000 trees and the square root of the number of predictors for each tree split (Figure S2). Interestingly, all combinations of hyperparameters for RF led to very close out-of-sample RMSE values of 8.74–8.78 (Figure S2). The two Feature Importance (FI) metrics for RF, based on node purity and permutation, showed that mean temperature, mean temperature at lags 1–3 days and NO₂ concentration at lags 1–3 days were the top 3 predictors (Figure 2.4a). The fourth most important predictor was either SO₂ or PM_{2.5} concentration (at lags 1–3 days), depending on the FI metric examined.

Finally, hyperparameters for GBM were also found using the grid search method. The best GBM had a tree depth of 5, 2500 trees, 1/3 of the predictors used at each split and a learning rate of 0.001 (Figure S3). FI showed that the three most important variables were mean temperature, mean temperature at lags 1–3 days and NO₂ concentration at lags 1–3 days (Figure 2.4b). These most important predictors were the same as in RF, but much further away from the others compared to RF. In GBM, the fourth most important predictor was relative humidity at lags 1–3 days.

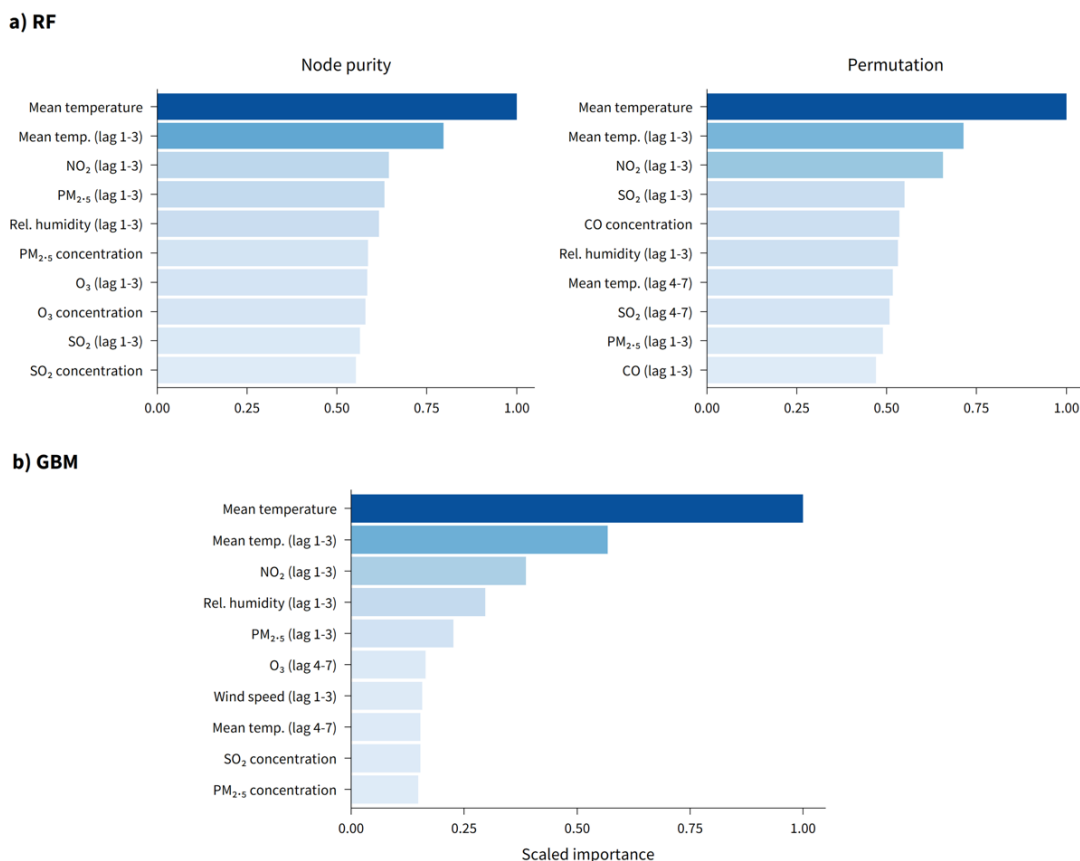


Figure 2.4 Feature importance (FI) metrics for a) Random Forest (RF) and b) Gradient Booting Machine (GBM). Only the 10 most important predictors are shown for each model.

2.3.2 Neural networks

For SLP, all models performed better when the “minmax” scaling function was applied to the predictors, compared to the two other methods tested i.e., standard and robust (Figure S4). The SLP with the smallest number of neurons (5) was selected as the best model, using ReLU as an activation function and trained with a learning rate of 0.001 (Figure S4). The FI metrics showed that only one variable seemed to contribute to the predictive power: mean temperature at lag 0 day (Figure 2.5a). Other important variables, such as relative humidity (at lags 1–3 days) and NO₂ (at lags 1–3 days), had much lower importance in the model. This can be explained by the fact that the optimal SLP found by cross-validation had only 5 neurons, limiting the amount of information that could be learned in that model.

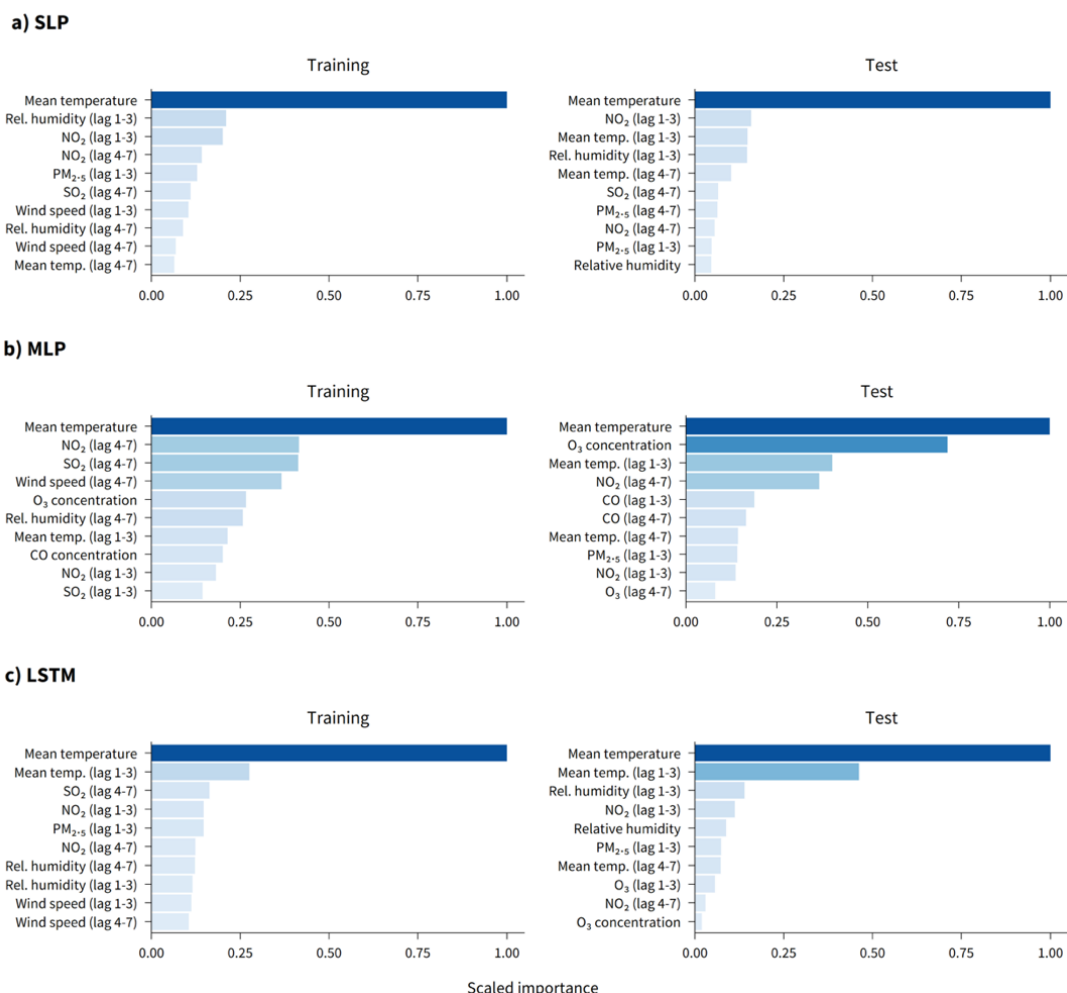


Figure 2.5 Feature importance (FI) metrics for a) Single-Layer Perceptron (SLP), b) Multi-Layer Perceptron (MLP) and c) Long Short-Term Memory (LSTM). FI metrics are computed for both the training (left) and test (right) datasets. Only the 10 most important predictors are shown for each model.

Like SLP, MLP also performed better when the predictors were scaled using the “minmax” scaler (Figure S5). When two or more hidden layers were considered in MLP, a relatively bigger neural network than SLP was found to be the best by the grid search method. This optimal MLP consisted of three hidden layers with 40 neurons in the first one, 30 in the second one and 20 in the third one. The activation function was logistic and the learning rate was 0.01 (Figure S5). FI metrics showed that, again, mean temperature at lag 0 day was the most important variable (Figure 2.5b). Results for the other most important predictors differ depending on whether FI was calculated on the training or test dataset. Among the most important variables in MLP, NO_2 and SO_2 (at lags 4–

7 days), O₃ concentration (at lag 0 day) and mean temperature (at lags 1–3 days) were found to be relevant contributors.

For the LSTM recurrent neural network, the best model was found among 450 potential candidates (i.e., all combinations of tested hyperparameters). This LSTM consisted of 15 cells in the first hidden layer and 10 cells in the second one and was trained with 10 000 epochs, a ReLU activation function, a learning rate of 0.0001 and no dropout layer (Figure S6). FI metrics for LSTM showed that mean temperature (at lag 0 day) was the most important variable, followed by mean temperature (at lags 1–3 days) for both metrics (Figure 2.5c). The third most important predictor was either relative humidity (at lags 1–3 days) or SO₂ concentration (at lags 4–7 days) depending on the training or test dataset.

2.3.3 Statistical models

Top five most important regression coefficients from GLM, GAM and DLNM were extracted and ranked in order of feature importance from left to right (Figure S7). For GLM, the results showed that mean temperature (at lag 0 day) was the most significant variable with a positive association with mortality deviation (Figure S7a). Relative humidity (at lags 1–3 days) was the second most important predictor and had a negative relation to mortality deviation. The third and fourth variables were wind speed (at lags 1–3 days) and mean temperature (at lags 1–3 days). For GAM, the two most important predictors were mean temperature at lags 0 and 1–3 days (Figure S7b). These two temperature variables exhibited the classical U/J shape of temperature-health relationships, which could not be obtained by the FI metrics of tree-based or neural networks models. The third and fourth most important predictors were SO₂ concentration (at lags 4–7 days), that had a negative association with mortality, and NO₂ (at lags 1–3 days), that had a positive one. For DLNM, the bidimensional cross-basis function of air temperature and lags showed a strong positive relationship at lag 0, as well as at lags 1 and 2 days, specifically for high temperature values (Figure S7c). This function that describes the whole temperature-lags-mortality relationship was the most important predictor in the model. The second and third most important predictors were NO₂ (at lags 1–3 days) and SO₂ (at lags 4–7 days) that had similar effects than the ones noted above for GAM.

2.3.4 Models' comparison

When models using the main dataset were compared in terms of FI, all models agreed that mean temperature (at lag 0 day) was the most important variable to explain mortality deviation

(Table 2.2). In the second rank, mean temperature (at lags 1–3 days) was the most important predictor for all tree-based methods (DT, RF and GBM), LSTM and GAM. Most models also identified NO₂ has a key variable for modelling mortality deviation during summer months, especially at lags 1 to 3 days (for most models) and lags 4 to 7 days (for MLP). Relative humidity also appeared in the second, third or fourth rank for four models (RF, GBM, SLP and GLM) and in the fifth rank for two models (LSTM and GAM). Wind speed was rarely in the top 5 of the most important features except in GLM and DLNM. Air pollutants other than NO₂, such as SO₂, PM_{2.5} and O₃, sometimes appeared in the third to fifth rank of most important predictors.

Tableau 2.2 Five most important predictors in all considered models. Lags (in days) are indicated in parentheses (no indication means lag 0). Temperature variables are in light orange, NO₂ in blue and relative humidity in yellow.

		Variable #1	Variable #2	Variable #3	Variable #4	Variable #5
Tree-based methods	DT	Mean temperature	Mean temp. (1-3)	NO ₂ (1-3)	-	-
	RF	Mean temperature	Mean temp. (1-3)	NO ₂ (1-3)	Relative hum. (1-3)	SO ₂ (1-3)
	GBM	Mean temperature	Mean temp. (1-3)	NO ₂ (1-3)	Relative hum. (1-3)	PM _{2.5} (1-3)
Neural networks	SLP	Mean temperature	NO ₂ (1-3)	Relative hum. (1-3)	NO ₂ (4-7)	SO ₂ (4-7)
	MLP	Mean temperature	NO ₂ (4-7)	O ₃ concentration	Mean temp. (1-3)	NO ₂ (1-3)
	LSTM	Mean temperature	Mean temp. (1-3)	NO ₂ (1-3)	PM _{2.5} (1-3)	Relative hum. (1-3)
Statistical models	GLM	Mean temperature	Relative hum. (1-3)	Wind speed (1-3)	Mean temp. (1-3)	NO ₂ (1-3)
	GAM	Mean temperature	Mean temp. (1-3)	SO ₂ (4-7)	NO ₂ (1-3)	Relative hum. (1-3)
	DLNM	Mean temp. CB	NO ₂ (1-3)	SO ₂ (4-7)	O ₃ (4-7)	Wind speed (1-3)

CB = Cross-basis function.

When FI were extracted for models fitted on the supplementary dataset (that is, the dataset of 39 years of data, but without PM_{2.5}), the main conclusions as noted above still held (Table S2). Mean temperature was again the most importance variable in all models. Mean temperature (at lags 1–3 days) was the second most important predictor for 7 out of 9 models. NO₂ (at lags 1–3 days) was consistently the second or third most important predictor, followed closely by relative humidity (at lags 1–3 days). For the other most important variables, the results differed slightly than what was obtained above on the main dataset. For example, SO₂ never appeared in the five most important predictor while wind speed and O₃ concentration came up more often. Finally, mean temperature at lags 4 to 7 days appeared in fifth position for 3 models (GBM, MLP and LSTM), which was not seen when models were trained on the main dataset.

2.3.5 Models' performance

The ability of the models to predict daily mortality deviation (i.e., over- and under- mortality when long-term and seasonal trends were first removed) on the out-of-sample test set was compared (Figure 2.6, Table S3a). R^2 did not reach 5% for all models considered and was even negative for some models. These results demonstrated that the studied environmental factors such as temperature-related variables and air pollution only explained a little proportion of the variation in daily mortality. Based on RMSE, MAE and R^2 , GBM was the most performing model, closely followed by RF in second rank, especially for the MAE criteria. Then, DT, GAM and DLNM performed equally to model mortality deviation. Only a little performance difference was noted between GAM and DLNM, even though DLNM was given the complete temperature values over the last 7 days, while GAM only used aggregated values of the predictors at lags 0, 1–3, and 4–7 days. For neural networks (SLP, MLP and LSTM), all models performed poorly to model mortality deviation during summer months compared to a non-informative model (i.e., R^2 values close to 0). Finally, GLM had the worst overall results on this dataset.

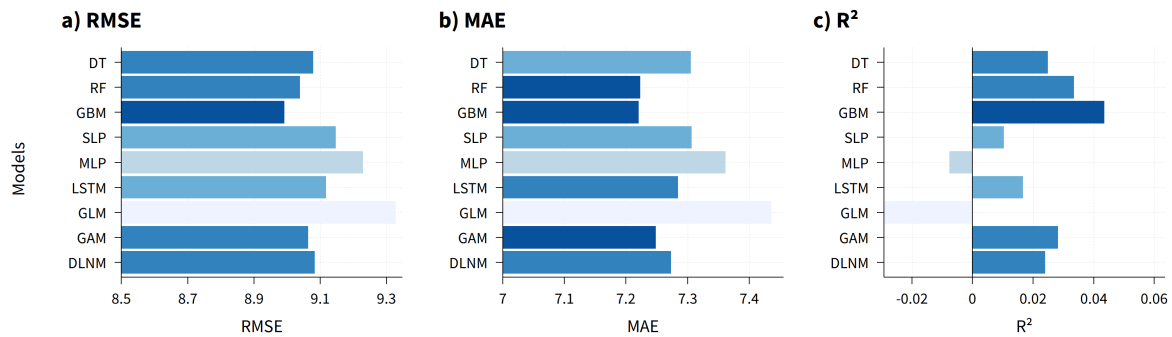


Figure 2.6 Performance of all models on the test set (2014–2019) given by a) Root Mean Square Error (RMSE), b) Mean Absolute Error (MAE) and c) Coefficient of determination (R^2). Low values of RMSE and MAE are desired, while high values are of R^2 are preferred. The best performance metrics are in dark blue, while the worst are in light blue.

When performance of models fitted using the supplementary dataset was compared, main conclusions held, but some differences were also noted (Figure S8, Table S3b). First, both GBM and RF were again the two best approaches to model mortality deviation. Overall, the performance metrics were better than when fitted on a smaller number of years (best out-of-sample R^2 values of 6.5%), but such comparison should be made with caution given that the test set was not the same (2014–2019 for the main dataset and 2009–2019 for the supplementary dataset). Both MLP and SLP obtained better results compared to when fitted with fewer data,

especially MLP, which could mean that the greater size of the dataset helped the MLP to be better calibrated. Overall, LSTM and DT were the worst models. As highlighted above, GAM outperformed DLNM again. For this dataset, GLM had better performance metrics than GAM and DLNM.

As a final partial validation of the models, weekly mortality predictions for the 2010-2019 period were compared to the observed values, especially for the 2010 and 2018 heatwaves (Figure 2.7). For the 2010 heatwave, 6 models predicted with great accuracy the over-mortality during this extreme heat event, namely DT, RF, MLP, LSTM, GAM and DLNM. However, since this heatwave was part of the training dataset, no conclusion can be drawn about the performance of the models. For the 2018 heatwave, which was in the out-of-sample test set, only 3 out of the 9 models correctly predicted the peak of mortality, namely DT, GAM and DLNM. These models are relatively simpler models than RF, GBM or neural networks. During more modest heatwaves (e.g., 2011 and 2013), the peak was over-estimated in 2011, but well predicted in 2013, for most models. For non-heatwave years (e.g., 2019), the weekly mortality deviation was generally not well predicted using only temperature-related and air pollution variables.

The same validation was performed with models calibrated using the supplementary dataset (calibration performed from 1991 to 2008), for which both 2010 and 2018 heatwaves were in the out-of-sample test set (Figure S9). Models that could reproduce the over-mortality in 2010 also reproduced appropriately the over-mortality in 2018. These models were the same as noted above for the 2018 heatwave (DT, GAM and DLNM), as well as two other models, RF and GBM.

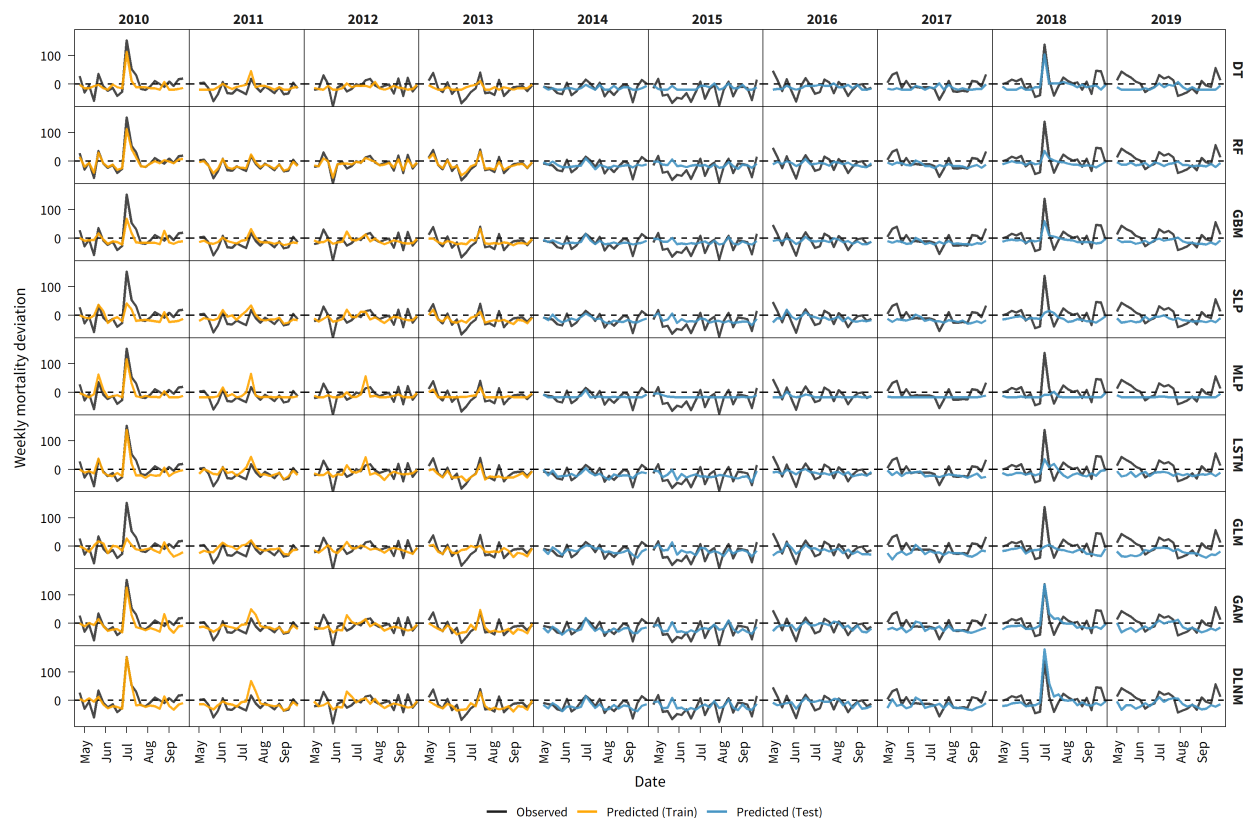


Figure 2.7 Weekly mortality deviation predictions from all models and observations for the 2010–2019 period.

2.4 Discussion

This study considered various machine, deep and statistical learning approaches to model the heat-mortality relationship in the metropolitan area of Montreal, Canada. Two datasets of ~20 and ~40 years of data were considered for the calibration of the models. These are much longer than datasets used in the recent literature of <10 years to calibrate such models (e.g., Khatri & Tamil, 2017; Ogata et al., 2021; Park et al., 2020; Qiu et al., 2020; Wang et al., 2019). The hyperparameters optimization results were thoroughly detailed. This contrasts with the existing literature in which hyperparameters are not always indicated (Morgenstern et al., 2020) or optimized (e.g., Zhang et al., 2014), which could lead to a “black box” feeling of the methodology. This new knowledge about the optimal and transparent tuning of machine and deep learning models can facilitate the adoption of these approaches in heat-health studies, as well as increase confidence in these methods by non-expert users.

To emphasize the importance of transparency in machine learning, we made sure that we could explain all models. Indeed, machine/deep learning models are often called black boxes because end users are not provided with information about the contribution of each predictor (Wiemken & Kelley, 2019). Feature Importance (FI) metrics were computed and revealed interesting information in addition to reaffirming established knowledge. The two most important predictors to model daily summer mortality were mean air temperature values at lags 0 and 1 to 3 days, while temperature at lags 4 to 7 days was less important in all models. These results are consistent with the literature and confirmed the relevance of short-term lagged mean temperature to model mortality (Son et al., 2019). Indeed, mean temperature was found to be a good compromise between various temperature indices for the heat-mortality relationship in comparative studies conducted in the United States (Barnett et al., 2010) and Australia (Vaneckova et al., 2011). NO₂ at lags 1–3 days was overall the third most important variable. This finding is interesting given that NO₂ is less often included in heat-mortality studies that focused more on other air pollutants such as O₃ or PM (Basu, 2009; Son et al., 2019), although it is a known driver of mortality (Wang et al., 2019). The fourth most important predictor was relative humidity at lags 1–3 days. This was expected given the additional effect of humidity on mortality during hot days (Gosling et al., 2009). Although relative humidity is the most commonly used humidity variable in weather-health studies and has been useful in our study for modelling summer mortality using machine and deep learning, it may not be the optimal variable to be considered (Davis et al., 2016). All the above findings were also confirmed when the models were calibrated on the supplementary dataset.

When the ability of the models to predict out-of-sample daily mortality deviation was compared using performance metrics (i.e., RMSE, MAE and R²), ensemble tree-based methods (GBM and RF) exceeded the performance of traditional statistical models (GLM, GAM and DLNM) and neural networks (SLP, MLP and LSTM). Non-linear statistical models (GAM and DLNM) performed better than neural networks (SLP, MLP and LSTM) when fitted on the 22-year dataset, but MLP had better performance (close to the one of RF) when fitted on the 39-year dataset. This means that a larger dataset size can lead to better performance for more complex models such as neural networks. That said, LSTM performed poorly with both datasets. This could be due to the higher number of parameters to be estimated compared to the other models. Results of GLM and DT were highly dependent on the dataset (i.e., main or supplementary), so no clear conclusion could be drawn for either model. The best models (GBM and RF) had out-of-sample R² values of 3–4% for the main dataset and 6–7% for the supplementary dataset. Such results are

not surprising given that R^2 (or explained deviance) below 10% are reported for in-sample predictions in weather-health studies (Table 07 in Chiu, 2017).

As an additional validation, performance of the models during two major heatwaves (i.e., 2010 and 2018) was visually evaluated. Five models (DT, GAM, DLNM, RF and GBM) correctly reproduced the excess mortality during out-of-sample heatwaves when fitted on the 39-year period, but only three (DT, GAM, DLNM) when fitted on the shorter period. Interestingly, models that correctly modelled peak mortality during heatwave were not necessarily the same as the ones having better performance metrics. This could be due in part to the J- or U- shaped relationships between temperature and mortality (Gosling et al., 2009), which were correctly modelled by non-linear statistical models (GAM and DLNM). These results raise the question of the type of validation that should be performed depending on the end user's objective: to predict daily mortality deviations or to reproduce mortality peaks during extreme heat events. In the latter case, partial validation might be more informative for the final model choice than classical performance metrics (e.g., RMSE, MAE, R^2).

To our knowledge, this is the first study to compare such a large variety of approaches to model heat-health relationships, from more easily explainable models to more complex ones. While simpler models can straightforwardly give an indication of the sign of the relationship (GLM, GAM, DLNM) or on potential interactions (DT), more advanced models that allow for complex interactions between predictors may perform better (e.g., GBM and RF). We considered 9 modelling approaches, whereas the few other comparative studies in the literature were limited to 3–5 models (e.g., Marien et al., 2022; Nishimura et al., 2021; Ogata et al., 2021; Park et al., 2020; Qiu et al., 2020). Also, it is one of the first applications of the LSTM in the field (e.g., Lin et al., 2021; Nishimura et al., 2021) and a first comparison of the DLNM with machine and deep learning models.

The few other comparative studies found in the literature differ significantly from ours in terms of the studied health outcome, the considered predictors and the models used. For example, Marien et al. (2022) used DT, RF, GBM, MLP and Ridge Regression to model the annual myocardial infarctions and found that MLP and Ridge Regression slightly outperform tree-based methods. Nishimura et al. (2021) modelled the daily number of heat-related-illness using non-linear regression equations, LSTM and RF. Both LSTM and non-linear regression equations outperformed RF. Ogata et al. (2021) compared GLM, GAM, RF and GBM to predict heatstroke and found that GAM was the best model for out-of-sample predictions. Park et al. (2020) found that RF was the most performing approach compared to GLM, DT and Support Vector Machine

(SVM) to model weekly morbidity due to heatwave. Qiu et al. (2020) modelled days with high hospital admissions for cardiovascular disease and found that RF and GBM performed better than SVM, GLM and DT. These divergent results highlight the need for more comparative studies that consider various modelling approaches and study different health variables (e.g., all-cause mortality in our case).

The main strengths of the study should be highlighted. First, it considered a wide variety of modelling techniques, from statistical models to tree-based methods, to feedforward and recurrent neural networks, thus allowing for complex interactions between temperature-related and air pollution predictors to be considered. Second, the calibration of the models (i.e., hyperparameters optimization) was transparent and used two long datasets of ~20 and ~40 years of data. Third, FI metrics were extracted and compared for all considered models. They allowed explaining machine and deep learning models and finding the most relevant predictors for heat-health relationships modelling. Finally, two evaluations of models' predictions were performed using three performance metrics and a partial validation based on recent heatwaves.

Some limitations of the study must also be noted. First, only one case study was presented, i.e., the all-cause mortality in the Montreal metropolitan area. Hence, results cannot be directly applied to other regions or health impacts (e.g., cause-specific mortality or morbidity). Second, all models considered the same predictors i.e., mean daily values of temperature-related and air pollution variables at fixed lags 0, 1–3 and 4–7 days. No other lags (except for temperature in DLNM), aggregation (e.g., minimum or maximum), nor temperature metrics (e.g., Humidex, Heat Index) were tested. Third, relative humidity was used in all models considered, but it may not be the best indicator of humidity for heat-health studies. Fourth, FI metrics were computed using either node purity or permutation that are only two of the many methods to explain machine learning models. Finally, even though 9 models were considered, other modelling approaches could have been explored.

2.5 Conclusion

Mean temperature at lags up to 3 days, as well as NO₂ concentration and relative humidity (both at lags 1–3 days) were the most important predictors for modelling summer mortality in Montreal, Canada, based on temperature-related and air pollution variables. Ensemble tree-based methods (GBM and RF) outperformed decision tree, statistical models and neural networks based on three performance metrics. However, a partial validation during recent heatwaves showed that these models may underestimate the mortality spike during these events if they are not calibrated with

enough data. Therefore, we conclude that both machine/deep learning and statistical models are relevant for modelling heat-health relationships depending on a myriad of factors : size of the dataset, available computing time and resource, information to be derived (e.g., shape of the relationship), end user goal with the fitted model, etc. In the context of increased extreme heat events due to climate change, these new results can support the implementation or improvement of heat adaptation measures (e.g., early alert system). Hence, it is suggested that such in-depth comparison of various modelling approaches be extended to other health indicators, predictors and regions.

Funding

The main author has received funding from the Natural Sciences and Engineering Research Council of Canada (Vanier Scholarship, #CGV-180821), the Canadian Institute of Health Research (Health System Impact Fellowship, #IF1-184093), Ouranos (Real-Décoste Excellence Scholarship, #RDX-317725) and the National Institute of Public Health of Quebec (no grant number).

CRedit authorship contribution statement

Jérémie Boudreault: Conceptualization, Methodology, Data Curation, Formal analysis, Visualization, Software, Writing - Original Draft, Review and Editing, Funding acquisition. **Céline Campagna:** Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **Fateh Chebana:** Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Data availability

The authors do not have permission to share health data. Weather and air pollution data are freely available from Environment and Climate Change Canada (ECCC).

Declaration of competing interest

The authors declare that they have no known competing interests or personal relationships that could have appeared to in work reported in this paper.

Acknowledgments

The authors would like to thank Denis Hamel and Louis Rochette (from INSPQ) for their help with the mortality data extraction, and Magalie Canuel, Ray Bustinza, Félix Lamothe, Nathalie Gravel and Yohann Chiu (also from INSPQ) for their comments on early versions of this work. The authors also thank the editor, the associate editor and two anonymous reviewers who helped improve the quality of this paper.

Supplementary data

Supplementary data to this article can be found online at:

<https://doi.org/10.1016/j.scitotenv.2023.164660>

CHAPITRE 3 :
REVISITING THE IMPORTANCE OF TEMPERATURE,
WEATHER AND AIR POLLUTION VARIABLES IN
HEAT-MORTALITY RELATIONSHIPS
WITH MACHINE LEARNING

3 REVISITING THE IMPORTANCE OF TEMPERATURE, WEATHER AND AIR POLLUTION VARIABLES IN HEAT-MORTALITY RELATIONSHIPS WITH MACHINE LEARNING

Revisiter l'importance des variables de température, météorologiques et de pollution de l'air dans les relations chaleur-mortalité avec apprentissage automatique

Auteurs : Jérémie Boudreault¹², Céline Campagna¹², Fateh Chebana¹

¹ Centre Eau Terre Environnement, Institut national de la recherche scientifique (INRS), 490 de la Couronne, Québec (QC) Canada, G1K 9A9.

² Direction de la santé environnementale, au travail et de la toxicologie, Institut national de santé publique du Québec (INSPQ), 945 avenue Wolfe, Québec (QC) Canada, G1V 5B3.

Titre de la revue: *Environmental Science and Pollution Research*

Publié le 25 janvier 2024

<https://doi.org/10.1007/s11356-024-31969-z>

Contribution des auteurs :

J. Boudreault a été le principal contributeur de cet article, avec la conceptualisation de l'étude, la méthodologie, l'analyse des données, la modélisation, la production des figures et tableaux, et l'écriture du manuscrit. C. Campagna et F. Chebana ont aidé à conceptualiser l'étude et définir la méthodologie. Ils ont aussi validé les résultats, révisé l'article et approuvé sa version finale.

Lien entre cet article et le précédent :

Le premier article (chapitre 2) a permis d'identifier les modèles les plus intéressants parmi 9 approches différentes d'apprentissage automatique, profond et statistiques, pour modéliser la relation entre la chaleur et la mortalité à Montréal. Les 5 modèles les plus pertinents sont repris ici pour deux régions d'étude, c'est-à-dire Montréal et Québec. La base de données des prédicteurs est étendue à d'autres indices de température (p. ex. température maximale,

température minimale, humidex, etc.), des variables de changements de température, d'autres variables météorologiques (p. ex. les précipitations, la pression atmosphérique), tout en conservant l'humidité, la vitesse de vents et la pollution de l'air. Cet article permettra de cibler les meilleures variables et modèles pour prédire la mortalité estivale et de comparer les différences entre les deux régions à l'étude.

Abstract

Extreme heat events have significant health impacts that need to be adequately quantified in the context of climate change. Traditionally, heat-health association methods have relied on statistical models using a single air temperature index, without considering other heat-related variables that may influence the relationship and their potentially complex interactions. This study aims to introduce and compare different machine learning (ML) models, which naturally consider interactions between predictors and non-linearities, to re-examine the importance of temperature, weather and air pollution predictors in modelling the heat-mortality relationship. ML approaches based on tree ensembles and neural networks, as well as non-linear statistical models, were used to model the heat-mortality relationship in the two most populated metropolitan areas of the province of Quebec, Canada. The models were calibrated using a comprehensive database of heat-related predictors including various lagged temperature indices, temperature variations, meteorological and air pollution variables. Performance was evaluated based on out-of-sample summer mortality predictions. For the two studied regions, models relying only on lagged temperature indices performed better, or equally well, than models considering more heat-related predictors such as temperature variations, weather and air pollution variables. The temperature index with the best performance differed by region, but both mean temperature and humidex were among the best indices. In terms of modelling approaches, non-linear statistical models were as competent as more advanced ML models for predicting out-of-sample summer mortality. This research validated the current use of non-linear statistical models with the appropriate lagged temperature index to model the heat-mortality relationship. Although ML models have not improved the performance of all-cause mortality modelling, these approaches should continue to be explored, particularly for other health effects that may be more directly linked to heat exposure and, in the future, when more data become available.

Keywords: mortality; temperature; temperature variations; weather; air pollution; machine learning.

3.1 Introduction

Extreme heat has significant health effects (Basu, 2009; Ye et al., 2012) that need to be adequately modelled as the occurrence, frequency and intensity of heatwaves are increasing globally due to climate change (Meehl & Tebaldi, 2004). This may be particularly important in northern regions that are warming more rapidly than elsewhere (IPCC, 2021). Recent events in Canada, for example, have demonstrated the unexpected health effects that extreme heat can have in these regions. Indeed, the 2021 heat dome in British Columbia resulted in a death toll of nearly 700 people (Lapointe, 2021), while the 2010 heatwave in Quebec caused 280 deaths and 3400 additional emergency departments visits (Bustanza et al., 2013).

Traditionally, heat-health association methods have relied on mean, maximum or minimum temperatures, or on a perceived temperature index that also includes humidity and/or wind speed, such as humidex, dew point, apparent temperature or heat index, among others (Huang et al., 2011; Kovats & Hajat, 2008; Son et al., 2019). The few comparative studies between different temperature indices performed in the United States (Barnett et al., 2010) or Australia (Vaneckova et al., 2011) have shown that the best index differed by region and health outcomes. To date, no such studies have been conducted in Canada.

Besides humidity and wind speed that can be incorporated into temperature indices, other meteorological variables such as atmospheric pressure and lack of precipitation can also be used to characterize heat events. However, these variables have only been seldom considered in weather-health studies, with little attention paid to heat (e.g., Chiu et al. 2021; Karlsson and Ziebarth 2018; Kolb et al. 2007). In addition, temperature variations, such as Diurnal Temperature Range (DTR) and temperature change between days, can also contribute to the observed excess mortality, but were less commonly studied in the literature (e.g., Cheng et al. 2014a; 2014b; Lee et al. 2020; Vicedo-Cabrera et al. 2016). Finally, air pollution also has adverse effects on mortality (Basu, 2009; Manisalidis et al., 2020; Atkinson et al., 2014) and a potential synergistic effect with temperature during heatwaves (Kovats & Hajat, 2008; Pascal et al., 2021b; Analitis et al., 2018; 2014). However, air pollutants were not always considered in heat-health studies (e.g., Barnett et al., 2010; Zhang et al., 2014).

In modelling heat-health relationships, statistical models that consider each predictor independently fail to account for complex and potentially unknown interactions between different heat-related predictors, such as between temperature, humidity and wind speed as a measure of perceived temperature, or between temperature and air pollution for synergistic effects. To this end, machine learning (ML) models such as ensemble tree-based methods or neural networks

may be better suited for considering multiple atmospheric predictors in heat-health relationships as they naturally account for interactions and non-linearities by construction (Bi et al., 2019). However, ML approaches have only been seldom used for modelling heat-health relationships (e.g., Boudreault et al., 2023; Jian et al., 2023; Lee et al., 2022; Nishimura et al., 2021; Ogata et al., 2021; Park & Kim, 2018; Park et al., 2020; Wang et al., 2019; Włodarczyk et al., 2022; Zhang et al., 2014).

ML could reveal potentially unknown and new interactions between the aforementioned heat-related predictors and lead to better performance to predict the adverse effect of extreme heat. In this study, a comprehensive database of atmospheric predictors, including multiple temperature indices, temperature variations, other weather and air pollution variables, was used to train various ML and statistical models to predict out-of-sample summer mortality in the two most populated metropolitan areas of the province of Quebec, Canada. An iterative calibration of the models with different subsets of the predictors was employed and allowed us to conclude on the best predictors and models to characterize the heat-mortality relationship in these two regions. This study is one of the first applications of ML in the field to take such a wide range of environmental exposure predictors into account.

3.2 Materials and methods

This project received ethics approval from the Human Research Ethics Committee of the National Institute of Scientific Research (CER-22-693).

3.2.1 Mortality data

The two most populated Census Metropolitan Areas (CMA) in the Province of Quebec, Canada, were considered for this study: Montreal CMA and Quebec CMA (referred to as Montreal and Quebec for the rest of the paper). Population was 4.3 million inhabitants in Montreal and 0.8 million in Quebec (in 2021), of which, respectively, 18.0% and 21.7% were aged 65 and over (Statistics Canada, 2023). Population density was 919 hab./km² in Montreal and 240 hab./km² in Quebec. In terms of wealth, the median annual income among recipients in 2020 was respectively 40 800\$ and 46 000\$ in Montreal and Quebec (Statistics Canada, 2023).

All-cause daily mortality data was provided by the *Institut national de santé publique du Québec* (INSPQ) from 1981 to 2019 (prior to the COVID-19 pandemic). Because mortality data was spatially aggregated by Forward Sortation Area (FSA) (i.e., the first three digits of the postal code), only the FSA that were fully within the CMA of Montreal and Quebec were considered for the

analysis (Figure 3.1). Since Montreal and Quebec have different population sizes, crude daily mortality was first normalized by the total population in each CMA to obtain mortality rates by 100 000 inhabitants. Then, seasonal and long-term trends in mortality time series were adjusted to the mortality rate following the method described in Zhang et al. (2012, 2014). The adjusted trend was then subtracted from daily mortality rate to obtain mortality rate deviation i.e., the over (or under) mortality above (or below) the expected seasonal and long-term value for the two CMAs (Table 3.1).

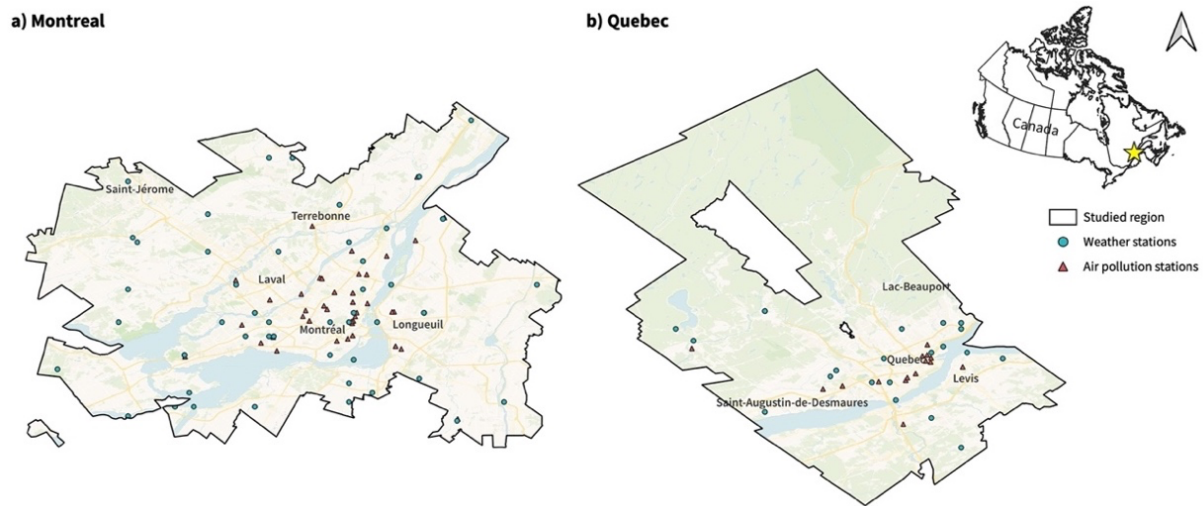


Figure 3.1 Weather and air pollution stations within the two studied Census Metropolitan Areas (CMA) of a) Montreal and b) Quebec.

3.2.2 Temperature, weather and air pollution data

Weather stations of Environment and Climate Change Canada (ECCC) and air pollution stations of the National Air Pollution Surveillance (NAPS) program of ECCC were used to obtain hourly weather and air pollution data, respectively (Figure 3.1). Hourly data was first converted to daily values (i.e., mean, maximum and minimum values). Then, weather and air pollution daily data at various stations within each CMA were combined using a spatial mean to obtain a single time series for each predictor (Chiu et al., 2021; Masselot et al., 2018). Considered temperature indices were air temperature, dew point and humidex. These indices were chosen because they were readily available from ECCC data and are well-known temperature metrics in Canada. Daily mean, minimum and maximum values were considered, leading to 9 temperature indices (Table 3.1). Lags up to 7 days prior to observed mortality were considered using the following aggregates: 0 day, 1–3 days and 4–7 days. A sensitivity analysis also considered the aggregation of temperature at lags 0–1, 2–3 and 4–7 days.

Temperature variations were computed from temperature indices. Mean temperature change (MTC) over 1, 3 and 7 days (i.e., the current mean temperature minus the mean temperature 1, 3 or 7 days before), as well as diurnal temperature range (DTR), computed from maximum temperature minus minimum temperature at lags 0, 1, 2 and 3 days, were considered (Table 3.1). Other weather variables included daily mean relative humidity, wind speed, visibility, atmospheric pressure and total precipitation. Available air pollution variables were concentration of O₃, PM_{2.5}, NO₂, SO₂ and CO in Montreal, but only O₃ and PM_{2.5} were available in Quebec. Daily mean values of air pollution concentration were considered, as well as maximum values in a sensitivity analysis. For weather and air pollution variables, same lags as for temperature indices were considered. Predictors were fully available from 1998 to 2019 (21 years) in Montreal and from 2003 to 2019 (17 years) in Quebec. The period from May to September was considered for studying the heat-mortality relationship.

Minimum, mean and maximum temperature were approximately 2°C higher in Montreal than in Quebec (Table 3.1). The dew point was also higher in Montreal of approximately 1°C. Humidex was generally 2.5 units higher in Montreal than in Quebec. Temperature change from 1 to 7 days were in the same range for the two CMAs, from approximately -17°C to +17°C. DTR values were also similar, but reaches higher values in Quebec. Weather variables were comparable in the two CMAs, but mean relative humidity was 2% higher in Quebec than in Montreal, mean visibility was 5 km lower in Quebec than in Montreal and total precipitation was slightly higher in Quebec. Finally, mean daily O₃ and PM_{2.5} concentration were higher in Montreal than in Quebec. These slight atmospheric differences between the two CMAs can be attributed to geographical location (Montreal being located further south than Quebec), as well as socioeconomic characteristics and industrial activities (for air pollution).

Tableau 3.1 Mortality, temperature, weather and air pollution data in the CMAs of a) Montreal (1998–2019) and b) Quebec (2003–2019). All data are daily values from May to September. Lagged variables were also included in the models, but not shown in the table (i.e., temperature indices, weather and air pollution variables at lags 1–3 and 4–7 days, DTR at lags 1, 2 and 3 days).

	a) Montreal				b) Quebec			
	Min	Mean	SD	Max	Min	Mean	SD	Max
Mortality-related variables								
Crude daily mortality	37.0	68.4	±10.0	135.0	4.0	13.7	±3.8	32
Mortality rate (/100 000 hab.)	1.1	1.8	±0.2	3.5	0.5	1.9	±0.5	4.4
Mortality rate deviation	-0.8	0.0	±0.2	1.7	-1.4	0.0	±0.5	2.4
Temperature indices								
Mean temperature (°C)	3.2	18.1	±4.4	29.0	2.2	16.3	±4.4	28.0
Maximum temperature (°C)	4.9	23.0	±4.9	35.1	4.1	21.3	±5.2	34.9
Minimum temperature (°C)	-1.1	13.1	±4.6	24.6	-3.1	11.1	±4.6	23.1
Mean dew point (°C)	-7.8	12.0	±5.2	23.5	-9.7	10.8	±5.3	23.3
Maximum dew point (°C)	-4.9	14.8	±5.0	26.8	-7.6	13.7	±5.1	25.1
Minimum dew point (°C)	-12.4	9.1	±5.6	22.1	-14	7.8	±5.7	22.4
Mean humidex	0.1	24.8	±7.6	43.0	-1.4	22.1	±7.8	40.9
Maximum humidex	2.9	29.8	±7.4	49.3	2.1	27.4	±7.7	46.9
Minimum humidex	-7.8	18.6	±8.5	39.4	-12.3	15.5	±8.7	38.8
Temperature variations (ΔT)								
MTC (0–1 days) (°C)	-12.7	0.0	±2.6	11.3	-10.3	0.0	±2.5	12.1
MTC (0–3 days) (°C)	-14.1	0.1	±4.2	15	-15.1	0.1	±3.9	13.8
MTC (0–7 days) (°C)	-17.1	0.2	±4.7	17.7	-15.3	0.3	±4.3	16.6
DTR (°C)	1.7	9.9	±3.2	21.9	1.4	10.2	±3.9	26.5
Other weather variables								
Relative humidity (%)	30.9	70.4	±11.7	97.3	31.6	73.2	±12.1	97.9
Wind speed (km/h)	3.2	10.8	±3.7	29.7	2.7	10.9	±4.5	38.3
Atmospheric pressure (kPa)	98.6	100.9	±0.6	102.9	97.6	100.2	±0.6	102.6
Visibility (km)	5.3	27.7	±7.8	50.6	3.4	23.4	±6.9	48.6
Total precipitation (mm)	0.0	3.1	±6.3	69.8	0.0	3.7	±7.3	76.2
Air pollution variables								
O ₃ concentration (ppb)	3.8	24.7	±8.9	68.2	3.3	21.9	±7.5	53.4
PM _{2.5} concentration (µg/m ³)	0.3	8.9	±5.9	60.8	0.7	7.9	±4.9	58.2
NO ₂ concentration (ppb)	1.3	9.7	±4.7	38.2	-	-	-	-
SO ₂ concentration (ppb)	0.0	1.7	±1.5	9.9	-	-	-	-
CO concentration (ppm)	0.1	0.3	±0.1	0.8	-	-	-	-

SD = Standard Deviation. MTC = Mean Temperature Change. DTR = Diurnal Temperature Range.

3.2.3 Machine learning (ML) and statistical models

Predicting daily summer mortality based on atmospheric predictors is a regression-type supervised learning problem. Therefore, five statistical and machine learning (ML) models adapted to this context were considered: Generalized Additive Model (GAM), Random Forest (RF), Gradient Boosting Machine (GBM), Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM). The choice of the models was based on their recent use in heat-health studies in order to have a wide range of different approaches while limiting the total number models

considered (Boudreault et al., 2023; Jian et al., 2023; Lee et al., 2022; Nishimura et al., 2021; Ogata et al., 2021; Park & Kim, 2018; Park et al., 2020; Wang et al., 2019; Włodarczyk et al., 2022; Zhang et al., 2014). GAM is a statistical model allowing for non-linear transformations of the predictors, while maintaining the property of additivity of each individual effect (Hastie & Tibshirani, 2017). GAM showed similar performance to the well-known Distributed Lag Non-Linear Model (DLNM) for out-of-sample mortality predictions in a recent study (Boudreault et al., 2023), which explains why only the former was considered in the current study. RF and GBM are both ensemble tree-based methods, but differ in the way they are built (Chapter 8 in James et al., 2013). In RF, fully grown decision trees are built in parallel (Breiman, 2001), while in GBM, trees with fewer leaves are built sequentially based on the residual of the last fitted tree(s) (Friedman, 2001). MLP is the vanilla feedforward artificial neural network inspired by the human brain. It can have many hidden layers and thus belong to the family of deep learning (Chapter 6 in Goodfellow et al., 2016). LSTM is part of recurrent neural networks (RNN), allowing for a sequential structure of the data such as in time series and extending classical RNN with a memory function (Chapter 10 in Goodfellow et al., 2016; Hochreiter & Schmidhuber, 1997). More details about these models can be found in the above references.

Instead of including all the predictors presented in Table 3.1 (and their lagged values) directly in the models, as it is generally the case in ML studies cited above, different subsets of predictors were first created based on four atmospheric groups: temperature indices, temperature variations, other weather variables and air pollution (Figure 3.2). These subsets of predictors enabled us to assess the additional predictive power of different atmospheric variables, as well as to test different temperature indices separately to model the heat-mortality relationship. Hence, models presented above (GAM, RF, GBM, MLP and LSTM) were first fitted using solely each of the 9 lagged temperature indices (i.e., a single (lagged) temperature index was included in the model(s) each time). Then, models were also fitted iteratively with temperature and other groups of heat-related predictors, namely 1) temperature variations (ΔT), 2) other weather variables and 3) air pollution, leading to 8 other subsets of predictors tested with the models (Figure 3.2). These other groups of predictors used the lagged mean temperature as their temperature index. Indeed, mean temperature is the most commonly used temperature index in heat-mortality relationships (Son et al., 2019) and was highly correlated with other temperature indices ($p > 0.90$). To summarize, the 5 modelling approaches were fitted using 9 lagged temperature indices and 8 subsets of mean temperature and other heat-related predictors for the two CMAs, resulting in a total 170 fits (Figure 3.2). Finally, sensitivity analyses were performed for considered lags in models with

temperature indices only and with maximum concentration of pollution (instead of mean) for models including air pollution.

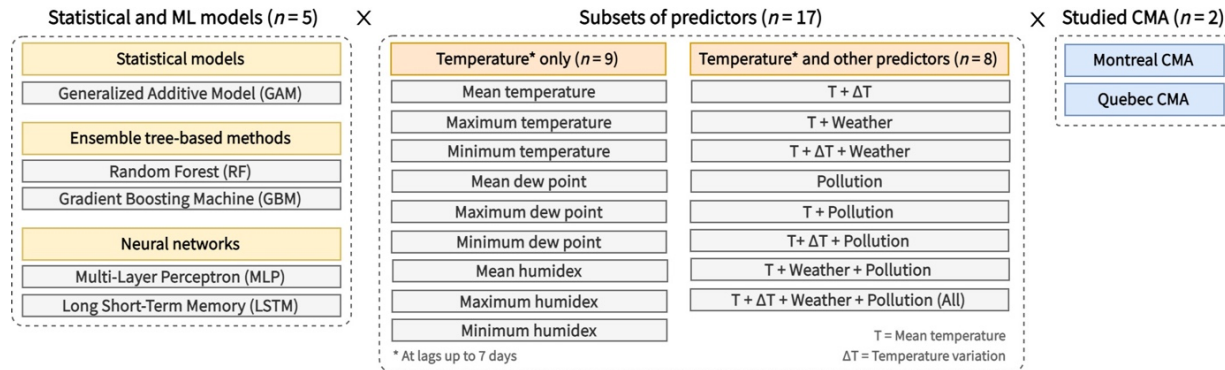


Figure 3.2 Overview of the fitted machine learning (ML) and statistical models in this study.

All models were trained with 70% of the first years of available data (*training* set) and validated with the 30% of remaining observations (*test* set) for the two CMAs, following the hold-out method for time series as in previous studies (e.g., Marien et al., 2022; Ogata et al., 2021; Park et al., 2020). The models GAM, RF, GBM, MLP and LSTM were fitted respectively using *mgcv* (Wood, 2015), *randomForest* (Liaw & Wiener, 2002) and *gbm* (Greenwell et al., 2019) in R, *scikit-learn* (Pedregosa et al., 2011) and *keras* (Gulli & Pal, 2017) in Python. The hyperparameters of ML models were optimized using a grid search with cross-validation using data only from the training set following the same procedure as described in Boudreault et al. (2023).

3.2.4 Models' and predictors' evaluation

Performance of the models was assessed in terms of the coefficient of determination (R^2) and percentage of decrease in Mean Absolute Error ($dMAE$) compared to a non-informative model (i.e., a model containing only an intercept). Higher values of these two metrics are desired. R^2 and $dMAE$ were computed on the test dataset (i.e., 30% of observations left out when training the models) to evaluate the models' performance on new data. Because R^2 and $dMAE$ were computed on the test set, potential negative values were capped at a lower value of 0 as an indication that the model does not outperform a non-informative model. The Root Mean Square Error (RMSE) was also considered and computed, but because it was proportional to the R^2 (i.e., only the constant denominator changes between these two metrics), it was not useful for models' comparison and not presented in the results section. These performance criteria were chosen on the basis of their use in previous studies, allowing for comparisons with the results of the latter

(e.g., Boudreault et al., 2023; Jian et al., 2023; Lee et al., 2022; Marien et al., 2022; Nishimura et al., 2021; Park et al., 2020).

To explain the contribution of each predictor in ML models, Feature Importance (FI) metrics were also computed. In tree-based models (RF and GBM), FI metrics were computed using node purity (i.e., total decrease in error for a given predictor) and permutation (i.e., increase in out-of-bag error when a predictor was randomly shuffled). In neural networks (MLP and LSTM), permutation was used to compute FI metrics. In GAM, FI was computed by the increase in the residual sum of squares when a predictor was removed from the model. Predictors were ranked in order of most important variable (i.e., highest FI) to the least important one (lowest FI) for each model. When two FI metrics were computed, the average rank between the two metrics was used. FI were only computed for the models including mean temperature and all other predictors.

3.3 Results

Models considering the different temperature indices are first presented (section 3.1), followed by the performance of models using temperature and other heat-related predictors (section 3.2). Then, feature importance metrics for models including all predictors are shown (section 3.3). This section concludes with a summary and comparison of the results (section 3.4).

3.3.1 Models using temperature indices only

Performance of GAM, RF, GBM, MLP and LSTM to model daily summer mortality in Montreal and Quebec using each of the 9 temperature indices (with lagged values up to 7 days) were compared. Performance metrics (R^2 and $dMAE$) on the test dataset are reported in Figure 3.3. In Montreal, the best performances were obtained using GAM, GBM and LSTM modelling approaches (illustrated respectively in yellow, red and blue on the top right-hand corner of Figure 3.3a). MLP, and especially RF, performed less well in Montreal. Temperature indices that modelled the best out-of-sample summer mortality in Montreal were maximum humidex, mean temperature, maximum dew point and mean humidex (Figure 3.3a). The best models had R^2 (on test set) values ranging from approximately 6% to 7% and $dMAE$ values of 3–3.5%.

In Quebec, results for the best models using lagged temperature indices were different from Montreal with the best indices being minimum temperature, minimum humidex, mean humidex and mean temperature (Figure 3.3b). The highest performance was generally obtained by GAM. LSTM and MLP were more competitive than GBM in Quebec. As in Montreal, RF performed poorly with R^2 and $dMAE$ values of 0 (or below). Generally, performance metrics in Quebec were lower

than in Montreal with, for example, the highest R^2 values on test dataset of 2–2.5%. Numerical results of R^2 and $dMAE$ for Montreal and Quebec can be found in Table S1.

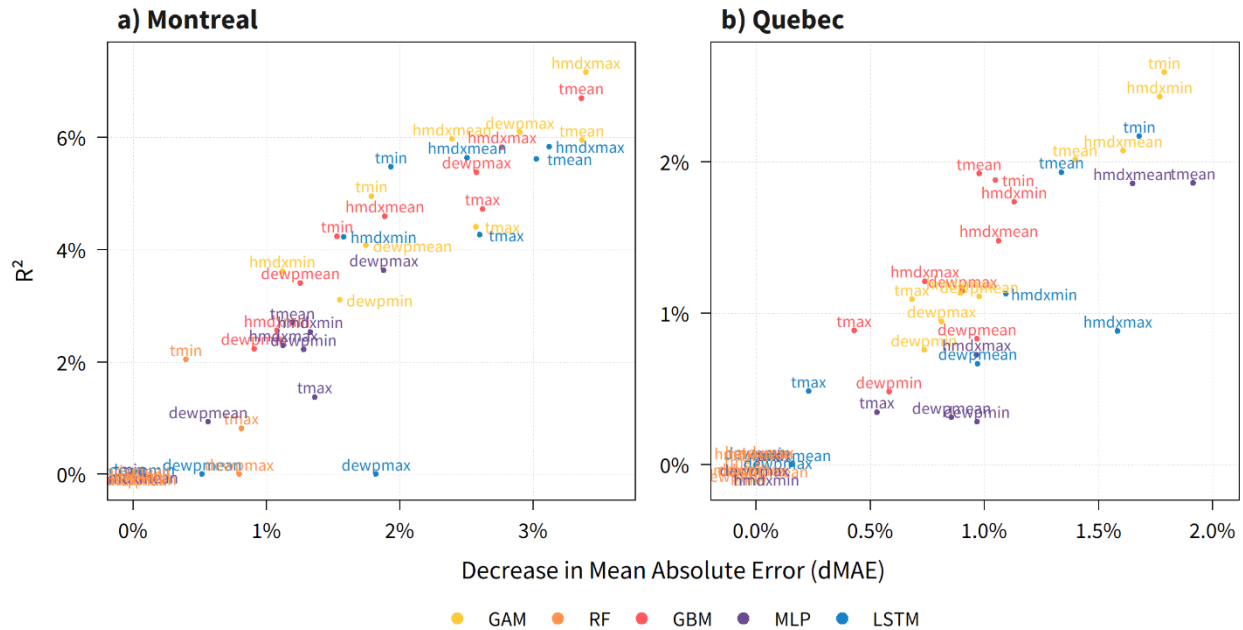


Figure 3.3 Performance metrics on test set based on R^2 (y-axis) and decrease in Mean Absolute Error (dMAE) (x-axis) of all models using different lagged temperature indices in the CMAs of a) Montreal and b) Quebec. Best combination of model (colour) and temperature index (text) are located in the top right-hand corner of each plot.

The sensitivity analysis showed that the performance decreased for the two CMAs when lags 0–1, 2–3 and 4–7 days of temperature were considered instead of 0, 1–3 and 4–7 days in the main analysis, particularly in Quebec (Figure S1). Overall, GAM, LSTM and GBM remained the best modelling approaches in Montreal, but an MLP with mean temperature was found to be the best model with this lag structure. The temperature indices performing the best remained generally the same in the sensitivity analysis in Montreal (i.e., mean and maximum temperature and humidex). However, in Quebec, minimum temperature and humidex were not anymore among the best temperature indices, which can be caused by the aggregation of temperature values at lags 0 and 1 days in the sensitivity analysis. GBM was more performant than MLP with this lag structure of temperature variables, while GAM and LSTM remained the most performing models in Quebec. For both CMAs, RF did not perform well again in the sensitivity analysis with performance metrics close to 0%.

3.3.2 Models using temperature and other heat-related predictors

All five models (GAM, RF, GBM, MLP, LSTM) were fitted iteratively with mean temperature, temperature variations, other weather and air pollution variables. Performance of all models on the test dataset is reported in Figure 3.4 in terms of R^2 and $dMAE$. In Montreal, the best models almost always included temperature, temperature variations and, sometimes, other weather and/or air pollution variables (Figure 3.4a). The best modelling approaches with temperature and other heat-related predictors were mostly RF, LSTM and GBM which had out-of-sample R^2 values of 5–6% and $dMAE$ values of 2.5–3.0%. These performance metrics were lower than the ones obtained by the best models using temperature indices only of section 3.3.1 (denoted by a black cross on Figure 3.4a).

In Quebec, models considering temperature and air pollution were the most performant, sometimes also including temperature variations and other weather variables (Figure 3.4b). Overall, GAM was the most performing approach, but MLP, LSTM and GBM also performed relatively well compared to RF that had the lowest performance in Quebec. R^2 values on test dataset of 2.0–2.5% were obtained for the best models with temperature and other predictors, while best $dMAE$ values ranged from 2% to 2.5%. R^2 values were slightly lower for Quebec when more heat-related predictors were considered compared to the models with temperature indices only of section 3.1, while $dMAE$ values were slightly higher. Numerical values of performance metrics for models with temperature and other heat-related predictors can be found in Table S2.

The sensitivity analysis with daily maximum pollutant concentrations showed that models had overall worse performance than when using daily mean concentrations (Figure S2). In Montreal, the best modelling approaches were still LSTM, RF and GBM (Figure S2a). In Quebec, GAM and MLP were the best approaches, but GBM was also often the best model with maximum air pollution predictors (Figure S2b). In both CMAs, the same subsets of predictors performed the best in the sensitivity analysis as in the main analysis, that is with temperature and its variations in Montreal and with temperature and air pollution variables in Quebec.

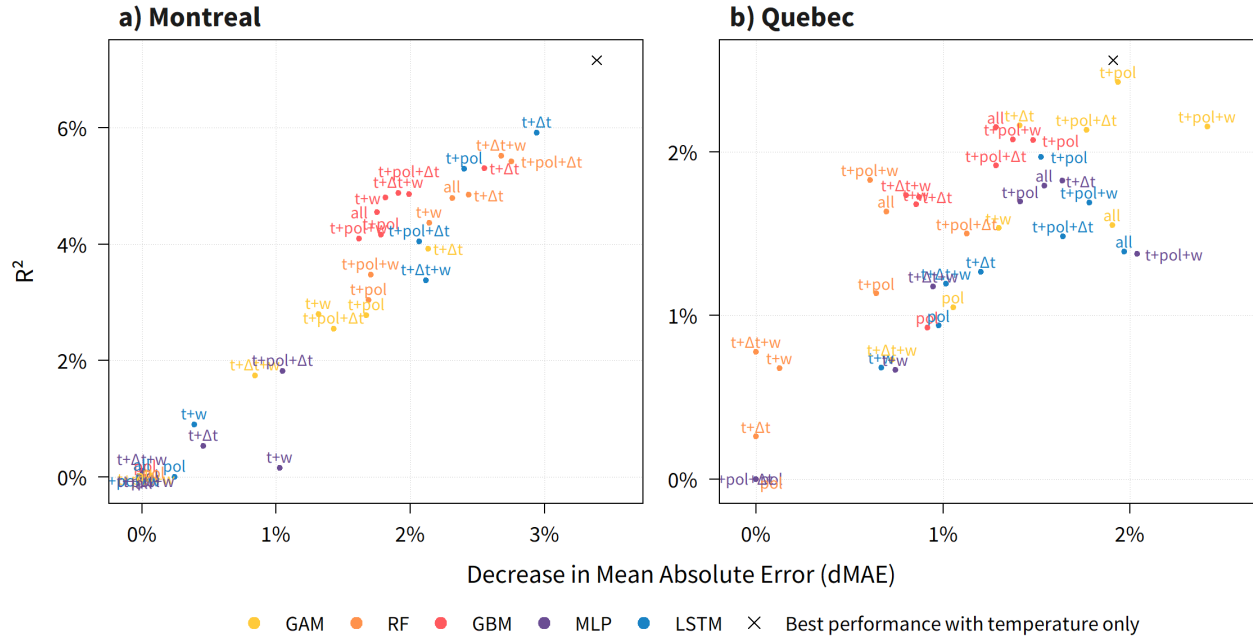


Figure 3.4 Performance metrics on test set based on R^2 (y-axis) and decrease in Mean Absolute Error (dMAE) (x-axis) of all models using mean temperature and other heat-related predictors in the CMAs of a) Montreal and b) Quebec. Best combination of model (colour) and predictors (text) are located in the top right-hand corner of each plot. t, Δt , w, pol and all stand for temperature, temperature variations, weather variables, air pollution and all variables, respectively. The black cross symbol shows the best performance achieved with models using temperature indices only of section 3.1.

3.3.3 Feature importance in models with all predictors

Feature importance (FI) metrics in models fitted using all predictors (but only using lagged mean temperature as the temperature index, recall section 3.2.3) were computed and used to rank predictors from most to least important. Top 5 most important predictors are shown in Table 3.2. In Montreal, mean temperature at lags 0 and/or 1–3 days were in the top 3 of the most important variables for all models considered (Table 3.2a). Also, mean NO_2 concentration (at lags 1–3 days) was either the second or third most important predictor for all models. For temperature variations, mean temperature change (MTC) over 3 days was the fourth most important variable in RF and GBM, and the most important one in MLP. Other important variables in Montreal were, depending on the model, wind speed, total precipitation, atmospheric pressure, mean concentration of $\text{PM}_{2.5}$ and SO_2 . Relative humidity was not in the top 5 of the most important predictors for any models.

In Quebec, temperature at either lag 0 day or 1–3 days was in top 3 of the most important predictors for four models out of five (Table 3.2b). In GAM, mean temperature was not in the top

5 of the most important predictors, which can be explained by the high number of predictors considered and potential correlations with other predictors. Temperature change over 3 days was the first or second most important predictor in RF and MLP. In both neural networks (MLP and LSTM), total precipitation (at lags 1–3 days) was the most important predictor. Mean PM_{2.5} was in the top 3 of the most important predictors for three models (GAM, RF and GBM), and the most important variable in GAM. Other important variables in Quebec were mean O₃ concentration, atmospheric pressure, wind speed, MTC over 1 or 7 days and Diurnal Temperature Range (DTR). Relative humidity was the second most important predictor in GAM, but not in the top 5 of the most important predictors in any other models.

Tableau 3.2 Top 5 most important predictors in models using all predictors in the CMAs of a) Montreal and b) Quebec. Variable #1 is the most important while variable #5 is less important. Mean temperature is in orange, temperature variations in yellow, other weather variables in grey and air pollution in blue. Lags in days are indicated in parentheses (no indication means lag 0).

	Variable #1	Variable #2	Variable #3	Variable #4	Variable #5
a) Montreal					
GAM	Mean temp. (1–3)	Mean NO ₂ (1–3)	Mean temperature	Wind speed (1–3)	Mean SO ₂ (4–7)
RF	Mean temperature	Mean temp. (1–3)	Mean NO ₂ (1–3)	MTC (0–3)	Mean SO ₂ (1–3)
GBM	Mean temperature	Mean temp. (1–3)	Mean NO ₂ (1–3)	MTC (0–3)	Mean PM _{2.5} (1–3)
MLP	MTC (0–3)	Mean NO ₂ (1–3)	Mean temperature	Total precipitation (1–3)	Mean NO ₂ (4–7)
LSTM	Mean temperature	Mean temp. (1–3)	Mean NO ₂ (1–3)	Atm. pressure (4–7)	Mean temp. (4–7)
b) Quebec					
GAM	Mean PM _{2.5} (1–3)	Relative humidity	Total precipitation (1–3)	Mean PM _{2.5} (4–7)	Atm. pressure (1–3)
RF	MTC (0–3)	Mean temperature	Mean PM _{2.5} (1–3)	MTC (0–7)	Mean O ₃
GBM	Mean temperature	Mean PM _{2.5}	Mean O ₃	MTC (0–1)	MTC (0–3)
MLP	Total precipitation (1–3)	MTC (0–3)	Mean temp. (1–3)	Mean O ₃	Wind speed
LSTM	Total precipitation (1–3)	Total precipitation (4–7)	Mean temp. (1–3)	Mean O ₃ (4–7)	DTR (3)

MTC = Mean Temperature Change. DTR = Diurnal Temperature Range.

3.3.4 Results summary and comparison

Summer mortality was best modelled using maximum humidex/dew point and mean temperature/humidex in Montreal, while minimum and mean temperature/humidex were more performant in Quebec (Table 3.3). GAM and LSTM were the best performing models using lagged temperature indices for both regions, but GBM (in Montreal) and MLP (in Quebec) also had competitive results. Performance metrics (R^2 and $dMAE$) computed on out-of-sample data were higher in Montreal than in Quebec for models with temperature indices only, meaning that the relationship between temperature and mortality was stronger in the former than the latter (Table 3.3).

Tableau 3.3 Results summary in terms of predictors, models and performance in the CMAs of a) Montreal and b) Quebec. Lags in days are indicated in parentheses (when relevant).

		a) Montreal	b) Quebec
Models with temperature indices only	Best temperature indices (with lags up to 7 days) among all models	Maximum humidex Mean temperature Maximum dew point Mean humidex	Minimum temperature Minimum humidex Mean humidex Mean Temperature
	Best models	GAM GBM LSTM	GAM MLP LSTM
	Best performance achieved	R^2 of 6.0–7.0% $dMAE$ of 3.0–3.5%	R^2 of 2.0–2.5% $dMAE$ of 1.5–2.0%
Models with temperature and other heat-related predictors	Best subsets of predictors among all models	T + ΔT T + ΔT + Weather T + ΔT + Pollution	T + Pollution T + Pollution + Weather T + Pollution + ΔT
	Best models	LSTM RF GBM	GAM
	Best performance achieved	R^2 of 5.5–6.0% $dMAE$ of 2.5–3.0%	R^2 of 2.0–2.5% $dMAE$ of 2.0–2.5%
	Most important predictors other than temperature when all considered	Mean NO ₂ (1–3) MTC (1–3)	Total precipitation (1–3) MTC (0–3) Mean PM _{2.5} (0 or 1–3)

T = Temperature. ΔT = Temperature variations. MTC = Mean temperature change. $dMAE$ = Decrease in MAE.

For models including temperature and other heat-related predictors, best subsets of predictors always included temperature in both CMAs, temperature variations in Montreal and air pollution in Quebec (Table 3.3). ML models (LSTM, RF, GBM) performed better with more predictors in Montreal. Those approaches differ slightly from the best models using temperature predictors only that were GAM, GBM and LSTM. In Quebec, GAM was still the best approach even with more predictors. Out-of-sample performance was not improved in Montreal with more heat-related predictors (compared to models with temperature indices only), and only marginally in terms of $dMAE$ in Quebec. Most important predictors (other than temperature) in models with all predictors were NO₂ at lags 1 to 3 days in Montreal, total precipitation and mean PM_{2.5} concentration (at lags 0 or 1–3 days) in Quebec, and MTC over 3 days in both CMAs (Table 3.3).

3.4 Discussion

This study considered machine learning (ML) and statistical models to revisit the role of various atmospheric variables for modelling heat-mortality relationships in two metropolitan areas in the Province of Quebec, Canada. ML allowed us to include by construction complex and potentially unknown interactions between the predictors in the models (Bi et al., 2019), while statistical models so far used in the literature have considered each atmospheric predictor independently. In addition, ML methods can be more flexible than statistical models, as they do not assume a

fixed shape for the predictor-response relationship or a prior statistical distribution of the error term. This can lead to better performance of ML over statistical approaches in various related fields such as public health (Fisher & Rosella, 2022), air pollution epidemiology (Bellinger et al., 2017) and environmental health (Schmidt, 2020). There are only a few other heat-health studies leveraging ML approaches in the literature (e.g., Boudreault et al., 2023; Jian et al., 2023; Lee et al., 2022; Nishimura et al., 2021; Ogata et al., 2021; Park & Kim, 2018; Park et al., 2020; Usmani et al., 2021; Wang et al., 2019; Wlodarczyk et al., 2022; Zhang et al., 2014). To the best of our knowledge, this study is the first to consider a wide range of modelling approaches (GAM, RF, GBM, MLP and LSTM), temperature indices (mean, maximum and minimum air temperature, dew point and humidex) and other heat-related predictors (temperature variations, weather and air pollution variables) to conclude on the best predictors and models to use to characterize the heat-mortality relationship.

Our results showed that among the models using lagged temperature indices only, the best performance was overall achieved by GAM and LSTM in both CMAs (Montreal and Quebec), but with different temperature indices (recall Table 3.3). Indeed, the mortality in summer was best modelled by mean/maximum metrics in Montreal and mean/minimum indices in Quebec. The difference in the most performing temperature index between the two CMAs can depend on a myriad of factors such as climatology, demography, adaptation to heat, etc. Explaining that difference goes, however, beyond the objective of the present study. These results were consistent with the common use of both maximum and minimum daily temperature values to define heatwave and heat alert thresholds in the Province of Quebec (Chebana et al., 2013) and around the world (Kent et al., 2014; Pascal et al., 2006, 2013; Tong et al., 2010; Xu et al., 2016). Mean temperature, which was also one of the best performing temperature indices for both CMAs, has been employed in most heat-mortality studies (Son et al., 2019) and was also found among the best temperature indices in other comparative studies using statistical approaches (Barnett et al., 2010; Vaneckova et al., 2011). Finally, our study also highlighted the good performance of the humidex to predict summer mortality, which is a well-known index of perceived temperature in Canada used by weather services nationally (Smoyer-Tomic & Rainham, 2001). Thus, the humidex could be used to further refine heat-health studies in Canada and, eventually, to improve heat alert systems, if humidex forecasts are as reliable as temperature ones.

For models considering temperature and other heat-related predictors, the most performant models were mostly RF, GBM and LSTM in Montreal and GAM in Quebec (recall Table 3.3). It is interesting to note that ML approaches only performed better in Montreal compared to Quebec.

This can be explained by the stronger mortality-temperature relationship found in this study for the former than the latter. That being said, even if more complex models were built using ML and a larger number of atmospheric predictors (such as temperature variations, weather and air pollution variables), these models did not perform better (in Montreal) and only marginally (in Quebec) than models using temperature indices only. These results are in the same vein as those of Pascal et al. (2021b) who studied the joint effects of temperature and air pollution in France with statistical models and concluded that it may not be necessary to consider air pollution when issuing heat alerts. By training our models iteratively with different subsets of predictors, our study allowed to quantify the increased (or decreased) predictive power of the different predictors considered and conclude on the optimal predictors to be used to model heat-mortality relationships in the two studied regions. These findings are usually not possible to obtain in classical ML applications in which all predictors are added in the model(s) at the same and unique time, leading to less parsimonious models than those obtained in our study.

Although less performing, models with more predictors indicated a pattern towards other important variables than air temperature in heat-mortality relationships such as temperature variations (in Montreal) and air pollution (in Quebec) (recall Table 3.3). Hence, air pollution may have a stronger short-term impact on mortality in Quebec than in Montreal, despite lower levels were found during the studied period compared to Montreal (Table 3.1). A more detailed analysis of the most important predictors (i.e., feature importance metrics in ML models) showed that NO₂ was the most important variable after temperature in Montreal. This is consistent with another study in the same city (Boudreault et al., 2023), but that used more models and fewer predictors than the current one. Unfortunately, NO₂ concentration was not available in Quebec. PM_{2.5} was particularly important in all models in this region, which is a known driver of mortality (Atkinson et al., 2014). Total precipitation was another important predictor, particularly in the neural network models. That said, interpretation of these results in Quebec must be made with caution due to the weaker relationship observed (out-of-sample R² of ~2%) compared to Montreal (~6–7%). In both CMAs, the mean temperature change (MTC) over 3 days was among the most important predictors and could be seen as an indicator of the adverse effect of a sudden temperature change when people are not yet used to such temperatures. This predictor was less studied in the literature than MTC over one day (e.g., Cheng et al., 2014b; Vicedo-Cabrera et al., 2016). Finally, relative humidity was not found to be of great importance, although relative and absolute humidity were important in other heat-health studies using ML approaches (Boudreault et al., 2023; Włodarczyk et al., 2022; Zhang et al., 2014). This might not be surprising given that relative

humidity is often described as a poor humidity predictor compared to other variables that were also considered in our study (e.g., dew point and humidex) (Davis et al., 2016).

Our study has several strengths. First, it considered five ML and statistical models to evaluate the best atmospheric predictors in heat-mortality relationships, thus allowing for potential complex interactions between predictors and non-linearities. Second, nine different temperature indices were tested based on air temperature, humidex and dew point at lags up to 7 days prior to mortality. Third, a comprehensive database of other heat-related predictors including temperature variations, meteorological variables and air pollution was considered to supplement temperature indices. Fourth, the iterative procedure of the models' fitting allowed assessing the predictive power of different subsets of predictors, leading to more parsimonious models. Finally, the study of two different metropolitan areas in the Province of Quebec, Canada, where no such comparative study has ever been carried out, provided different regional perspectives in terms of best predictors and modelling approaches.

Limitations should also be discussed. First, because the association between all-cause mortality and heat is generally weak (i.e., heat only explains a small variation of mortality), the performance of complex ML approaches can be limited. Cause-specific mortality (e.g., cardiovascular, respiratory) or morbidity impacts (e.g., heat-related illnesses) may have stronger relationships that could work better with ML models. Second, the models were built independently for the two CMAs with approximately 20 years of summer data. This may also limit the predictive power of ML approaches, which may require more data to achieve better performance. Third, temperature indices used in other contexts, such as the Wet-Bulb Globe Temperature (in occupational health) or Heat Index/Apparent Temperature (in the United States), could also have been considered, but were less appropriate to the Canadian public health context. Fourth, other modelling approaches could have been explored instead of the five considered, such as clustering of environmental conditions or classification techniques if the response variable was different. Finally, performance measures other than R^2 , Mean Absolute Error and Root Mean Square Error could have been used to evaluate models' performance.

3.5 Conclusion

Our study confirmed the current use of non-linear statistical models with the appropriate lagged temperature index to model heat-mortality relationships. Indeed, these models had comparable (or better) performance than more complex and less parsimonious ML models that included other heat-related predictors. In addition, this research highlighted the need to consider different

temperature indices, as the most performing index may vary regionally. Indeed, maximum humidex/dew point and mean temperature/humidex performed better in Montreal, while minimum or mean temperature/humidex yielded better results in Quebec. These new insights, derived from both ML and statistical models, may be particularly important for the development of new heat warning systems. Many temperature indices and other weather and air pollution variables may appear of interest at first sight for such systems, but these variables need to be carefully analyzed with the appropriate and most advanced modelling approaches prior to being included, as was done in this paper. The developed framework can easily be adapted and applied to other regions of the world also experiencing the adverse effects of extreme heat events. Finally, despite the similar performance of ML models obtained in this study compared to statistical models, we suggest that their potential continues to be further explored, particularly with specific health outcomes that may be more strongly associated with heat, in other regions and when more data will become available in the future.

Acknowledgments

The authors would like to thank Denis Hamel and Louis Rochette for the help with the mortality data extraction. They would also like to thank Yohann Chiu, Magalie Canuel, Ray Bustinza, Felix Lamothe, Natalie Gravel and Annabel Ruf for their comments on early versions of this work. The authors also thank the editor and two anonymous reviewers who helped improve the quality of this paper.

Supplementary Information

The online version contains supplementary material available at:

<https://doi.org/10.1007/s11356-024-31969-z>

Author contribution

Jérémie Boudreault: conceptualization, methodology, data curation, formal analysis, visualization, software, writing — original draft, review and editing, funding acquisition. **Céline Campagna:** conceptualization, writing — review and editing, supervision, project administration, funding acquisition. **Fateh Chebana:** conceptualization, writing — review and editing, supervision, project administration, funding acquisition.

Funding

The main author has received funding from the Natural Sciences and Engineering Research Council of Canada (Vanier Scholarship, #CGV-180 821), the Canadian Institute of Health Research (Health System Impact Fellowship, #IF1-184093), Ouranos (Real-Décoste Excellence Scholarship, #RDX-317725) and the National Institute of Public Health of Quebec (no grant number).

Data Availability

The authors do not have permission to share health data. Weather and air pollution data are freely available from Environment and Climate Change Canada (ECCC).

Ethical approval

This project received ethics approval from the Human Research Ethics Committee of the National Institute of Scientific Research (CER-22–693).

Consent to participate

Not applicable. Consent to participate was not required due to the nature of the data used (i.e., anonymized administrative health database).

Consent for publication

Not applicable. Consent to publish was not required due to the nature of the data used (i.e., anonymized administrative health database).

Competing interests

The authors declare no competing interests.

CHAPITRE 4 :
MULTI-REGION MODELS BUILT WITH MACHINE AND DEEP
LEARNING FOR PREDICTING SEVERAL HEAT-RELATED
HEALTH OUTCOMES

4 MULTI-REGION MODELS BUILT WITH MACHINE AND DEEP LEARNING FOR PREDICTING SEVERAL HEAT-RELATED HEALTH OUTCOMES

Modèles multi-région contruits avec l'apprentissage automatique et profond pour prédire plusieurs issues sanitaires liées à la chaleur

Auteurs : Jérémie Boudreault¹², Annabel Ruf¹, Céline Campagna¹²³, Fateh Chebana¹

¹ Centre Eau Terre Environnement, Institut national de la recherche scientifique (INRS), 490 de la Couronne, Québec (QC) Canada, G1K 9A9.

² Direction de la santé environnementale, au travail et de la toxicologie, Institut national de santé publique du Québec (INSPQ), 945 avenue Wolfe, Québec (QC) Canada, G1V 5B3.

³ Département de médecine sociale et préventive, Faculté de médecine, Université Laval, 2300 rue de la Terrasse, Québec (QC) Canada, G1V 0A6.

Titre de la revue : *Sustainable Cities and Society*

Publié le 10 septembre 2024

<https://doi.org/10.1016/j.scs.2024.105785>

Contribution des auteurs :

J. Boudreault a été le principal contributeur de cet article, avec la conceptualisation de l'étude, la méthodologie, l'analyse des données, la modélisation, la production des figures et tableaux, et l'écriture du manuscrit. A. Ruf a participé à l'écriture des codes, la modélisation, la production des figures et tableaux, et la relecture du manuscrit. C. Campagna et F. Chebana ont aidé à conceptualiser l'étude et la méthodologie, ont validé les résultats, révisé l'article et approuvé sa version finale.

Lien entre cet article et les précédents :

Les deux premiers articles (chapitres 2 et 3) ont permis de mieux comprendre les modèles d'apprentissage automatique et profond et les variables environnementales permettant de prédire

les impacts sanitaires de la chaleur dans deux villes, avec l'exemple de la mortalité. Ce troisième et dernier article de l'objectif 1 de la thèse vise à généraliser les résultats à l'ensemble des 15 RSS du Québec méridional et d'étendre l'étude de la mortalité à d'autres issues sanitaires de morbidité (hospitalisations, visites aux urgences, transports ambulanciers et appels à Info-Santé 811) chez deux populations (générale et 65 ans et plus). Il propose différents modèles multi-région tirant profit des données de plusieurs RSS au sein d'un même modèle pour prédire ces différentes issues sanitaires liées à la chaleur dans l'ensemble du Québec.

Abstract

As a result of climate change, populations worldwide will be exposed to more heat episodes. To ensure a sustainable future, cutting-edge tools must be developed to predict the health effects of heat and limit its consequences. However, current research has mainly focused on one health outcome in a single city/region, thus providing limited knowledge to improve society's resilience. In this study, a machine learning (ML) framework is introduced to predict several heat-related health outcomes in multiple regions simultaneously, using the province of Quebec (Canada) as a case study. Five ML models including penalized regression, ensemble tree-based models and deep neural networks were considered and compared. Models were trained to predict these health outcomes using various meteorological, regional and temporal predictors across all regions. Our results showed that deep learning models were the most promising, with out-of-sample R^2 of >60% for most of the studied health outcomes. However, ensemble tree-based approaches also had the best performance for some health outcomes, and were more sensitive to weather variables and to heatwaves. By introducing novel ML-based tools for predicting heat risks in several regions, this study can guide climate change adaptation and help cities and society to become more healthy, resilient and sustainable.

Keywords: temperature; extreme heat events; mortality; morbidity; ensemble tree-based methods; deep neural networks.

Highlights

- Daily summer mortality and morbidity predicted for general and elderly populations
- Machine/deep learning models trained with weather, regional and temporal predictors
- Deep learning performed better globally and across regions than other ML models
- Ensemble tree-based models were more sensitive to weather variables and heatwaves
- Models provided new insights on heat-related health impacts in the two populations

4.1 Introduction

Climate change is recognized as one of the greatest threats to human health of the current century (Watts et al., 2021). Among its consequences are heat episodes that will occur earlier, last longer, be more frequent and intense (Meehl & Tebaldi, 2004). Extreme heat already affects population health worldwide, both in terms of mortality (Basu, 2009; Bi et al. 2023; Gosling et al., 2009; Kotharkar et al, 2024; Son et al., 2019) and morbidity (Li et al., 2015; Lin et al., 2023; Seong et al., 2024; Ye et al., 2012), but its consequences will be even more devastating in the future (Curtis et al., 2017; Huang et al., 2011; 2013). Yet state-of-art tools still need to be developed (or improved) to better allocate health resources and adequately target interventions that will ultimately reduce heat effects and lead to more a sustainable future (Ebi & Schmier, 2005; Kotharkar & Ghosh, 2022; McGregor et al., 2015).

With this in mind, recent developments in artificial intelligence (AI) and, specifically, machine learning (ML), have opened the door to a wide variety of novel applications towards sustainability, for example to improve population health (Fisher & Rosella, 2022; Morgenstern et al., 2020; Wiemken & Kelley, 2019), optimize green infrastructures (Shaamala et al, 2024) and mitigate urban heat (Li et al., 2024), among others. Indeed, ML can exploit larger datasets than ever (i.e., big data), model complex interactions and non-linear relationships, learn from multiple data points simultaneously and does not require strong assumptions on data distribution, leading to enhanced performance over traditional methods (Bi et al., 2019). In the context of climate change and increasing extreme heat events, various ML-based models were proposed to predict heat-related health impacts (e.g., Boudreault et al., 2023; 2024a; Ke et al., 2023; Kim & Kim, 2022; Lee et al., 2022; Navares et al., 2018; Nishimura et al., 2021; Ogata et al., 2021; Park et al., 2020; Wang et al., 2019; Zhang et al., 2014). However, the aforementioned studies have shortcomings that need to be acknowledged.

First, most studies modelled a single health outcome such as mortality (Boudreault et al., 2023; Kim & Kim, 2022; Lee et al., 2022; Zhang et al., 2014) or heat-related illnesses (Ke et al., 2023; Nishimura et al., 2021; Ogata et al., 2021; Wang et al., 2019). Acknowledging that heat affects both mortality and morbidity, these models need to target multiple health variables to ensure a representative portrait of the heat-related burden for public health authorities. Second, existing studies have mainly focussed on a single city/region (Kim & Kim, 2022; Lee et al., 2022; Navares et al., 2018) or on multiple cities by (re)fitting the same model across locations (Boudreault et al., 2024a; Ogata et al., 2021; Zhang et al., 2014). This limits the potential use of big data in population health and ML's ability to take advantage of more data from multiple locations to improve its

performance. Third, the few multi-region studies found in the literature often lacked appropriate validations during particular events of interest such as heatwaves (Kim & Kim, 2022; Ogata et al., 2021; Park et al., 2020; Wang et al., 2019). This can lead to choosing a model that will underperform for periods during which the more important health effects are observed (i.e., heatwaves). Fourth, most studies have only considered one ML model, mainly ensemble tree-based methods (Ke et al., 2023; Kim & Kim, 2022; Wang et al., 2019; Zhang et al., 2014). This completely ignores the neural networks family (i.e., deep learning) that could be more effective. Moreover, the lack a benchmark/comparison model (e.g., linear regression) limits the ability to conclude on the gain of ML over more traditional approaches. Finally, some ML models in the literature were literally presented as “black boxes”, with no indications on models’ explainability (Lee et al., 2022; Navares et al., 2018) or on how to adequately tune these models (Ke et al., 2023; Park et al., 2020; Zhang et al., 2014)

These drawbacks can lead to an incomplete picture of heat-related health impacts, suboptimal predictions and a lack of confidence in these models by non-AI experts, thus limiting the ability of cities and society to become more resilient and sustainable in a climate change context. In this study, a ML-based framework is proposed to model both mortality and morbidity (5 health variables) in two populations (general and elderly) during summer in multiple regions simultaneously. To that end, five ML models based on penalized regression, ensemble tree-based models and deep neural networks were considered and applied to the province of Quebec, Canada. These models were calibrated with various weather, socioenvironmental and temporal predictors in a transparent way and were thoroughly validated. Finally, some explainability of the models was also provided, giving valuable insights for health authorities preparedness and surveillance activities of extreme heat events to foster climate adaptation and act towards more resilient cities.

4.2 Material and methods

4.2.1 Study design and health outcomes

An ML-based framework was proposed to predict several heat-related health outcomes in multiple regions simultaneously (hereinafter referred to “multi-region models”), instead of creating individual models for each region, thus leveraging information and data from all regions in this framework. The 15 southernmost health regions in the province of Quebec, Canada, were considered (top-right of Figure 4.1), excluding regions #10, #17 and #18 due to their different climatology and their smaller population compared to the other regions (Boudreault et al., 2024c).

The spatial scale of the health region (referred to simply as region for the rest of the paper) was chosen for this study as a trade-off between relevance for decision-makers (i.e., each region has its own public health team and director) and statistical power (i.e., having a large enough population in each region). These 15 regions represented over 99% of the Quebec population in 2021, i.e. around 8.5M inhabitants. The summer season was defined as May to September. This project received ethics approval from the Human Research Ethics Committee of the National Institute of Scientific Research (CER-22-693). Figure 4.1 presents an overview of the methodology.

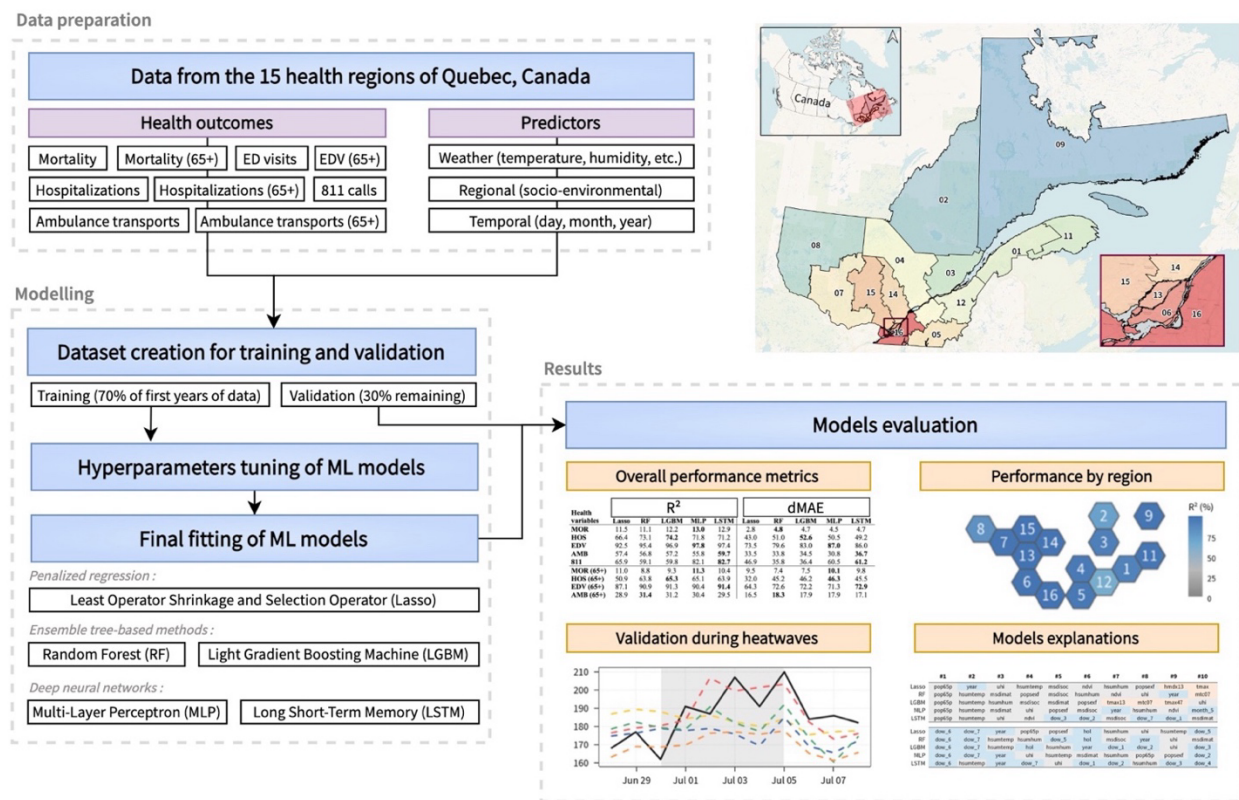


Figure 4.1 Overview of the methodology.

Health data was made available by the *Institut national de santé publique du Québec* (INSPQ). Studied health variables were the daily numbers in each region of all-cause: 1) deaths, 2) hospitalizations, 3) emergency department visits (EDV), 4) ambulance transports and 5) *Info-Santé* 811 calls (a free telephone consultation service for non-urgent health issues) (Table 4.1). All-cause health variables were chosen over cause-specific variables (e.g., cardiovascular diseases, heat-related illnesses) because they are more readily available in Quebec and already

tracked by public health authorities through the health surveillance system of extreme weather events called SUPREME (Toutant et al., 2011). In addition, all these health variables were shown to be heat-related in previous epidemiological studies performed in the same study area (Boudreault et al. 2024c; Lebel et al., 2019) and in international literature reviews (Campbell et al., 2018; Cheng et al., 2019; Ye et al., 2012). Health variables were provided for two populations: 1) the general population (of all ages), and 2) the elderly sub-population (aged 65 and over), known to be more vulnerable to the effects of heat (Benmarhnia et al., 2015). This led to 9 studied health outcomes for this study (5 health variables in two populations), as the data for 811 calls was only available for the general population (Table 4.1). Mortality and hospitalization data were available from 1996, EDV and ambulance transports from 2014 and 811 calls from 2008. All health outcomes were considered until 2019, prior to the COVID-19 pandemic.

Tableau 4.1 Overview of the studied health outcomes. Values are expressed in daily counts per 10⁴ inhabitants across all regions for the May to September period.

Health outcome	Age	Abbreviation	Mean $\pm$ SD	Range	Years
1 Mortality	All	MOR	0.2 $\pm$ 0.1	0.0–1.2	1996-2019
2 Hospitalizations	All	HOS	2.4 $\pm$ 0.7	0.6–6.4	1996-2019
3 EDV	All	EDV	16.0 $\pm$ 7.8	6.0–45.1	2014-2019
4 Ambulance transports	All	AMB	2.1 $\pm$ 0.5	0.4–5.1	2014-2019
5 811 calls	All	811	5.4 $\pm$ 1.3	1.3–10.9	2008-2019
6 Mortality (65+)	65+	MOR (65+)	1.0 $\pm$ 0.5	0.0–9.9	1996-2019
7 Hospitalizations (65+)	65+	HOS (65+)	6.0 $\pm$ 2.2	0.0–27.0	1996-2019
8 EDV (65+)	65+	EDV (65+)	20.9 $\pm$ 8.7	8.2–63.9	2014-2019
9 Ambulance transports (65+)	65+	AMB (65+)	6.1 $\pm$ 1.3	0.8–14.4	2014-2019

SD = Standard deviation. Range = Minimum and maximum daily values. EDV = Emergency department visits.
811 = Telephone consultation line for non-urgent health issue.

4.2.2 Weather, regional and temporal predictors

Three types of predictors were considered to model the 9 health outcomes of interest: 1) meteorological predictors, for the short-term effect of weather on health, 2) regional predictors, to account for the socioeconomic and environmental differences between regions, and 3) temporal predictors, to consider potential temporal patterns or trends in the modelled health outcomes (Table 4.2). For weather, gridded daily data at 1 km $\times$ 1 km over North America from 1980 to 2021 from NASA's Daymet database were used (Thornton et al., 2022). Pixel values of minimum (*tmin*) and maximum (*tmax*) temperature, as well as vapour pressure, were extracted and aggregated at the region level by weighting each pixel with the number of units located in each pixel using the *AQgéobâti* database (Adresses Québec, 2023). This led to population-weighted weather times series for each region. Mean temperature was computed from the average of daily maximum and minimum temperatures. Various humidity metrics related to

different health outcomes such as relative humidity, dew point and humidex (Davis et al., 2016) were computed from vapour pressure and temperature using *MetPy* in Python (May et al., 2022).

Lagged values of meteorological predictors up to 7 days were considered using the value at lag 0, the mean of values at lags 1 to 3 days and the mean of values at lags 4 to 7 days (Boudreault et al., 2024a). Temperature variation variables were also considered including Mean Temperature Change (MTC), computed from $(tmean_0 - tmean_d)$, where $tmean_0$ is the current mean temperature and $tmean_d$ is the mean temperature at $d = 1, 3$ and 7 previous days, and Diurnal Temperature Range (DTR), computed from $(tmax - tmin)$ at lags 0, 1, 2 and 3 days. Air pollution could not be included in this study because daily data was only available in a few regions throughout the study period. No explicit heatwave variables (e.g., intensity, duration, etc.) were included in the models. Indeed, ML models should be able to develop their own understanding of extreme heat events on the basis of the weather variables provided, as shown in a previous study (Boudreault et al., 2023).

Regional predictors included demographic, socioeconomic, environmental and climatological characteristics (Table 4.2). Total population (used as an offset in the models), proportion of elderly (i.e., people aged 65 years and more) and proportion of women were provided by the *Ministère de la Santé et des Services Sociaux* (MSSS) du Québec from 1996 to present (MSSS, 2022). The Material and Social Deprivation Index (MSDI) developed by INSPQ from data of the Canadian census at the dissemination area level was used (INSPQ, 2024). MSDI values for the social and material component of the index were aggregated at the region level by weighting the MSDI with the population in each dissemination area. Then, a linear interpolation was performed to obtain values between each census years (1996, 2001, 2006, 2011 and 2016) for the whole studied period.

Two built environment variables related to urban heat were considered: the Normalized Difference Vegetation Index (NDVI), a commonly known index of density of vegetation ranging from -1 to 1, and a measure of the Urban Heat Island (UHI), computed as the difference between surface temperature and a surrounding milder rural reference value, both provided by the *Centre d'enseignement et de recherche en foresterie de Sainte-Foy* (CERFO, 2022). NDVI and UHI values available at 15m x 15m resolution in 2013 were weighted by the number of units in each pixel (with a buffer of 15 m) to obtain NDVI and UHI population-weighted values in all regions (same as what was done for weighting meteorological variables above). The 2013 values of NDVI and UHI were applied to the whole studied period (i.e., 1996–2019) as done previously (Pascal et al., 2021a). Finally, 20-year averages of mean temperature and relative humidity by region

during summer (May to September) were considered as climatological predictors. They were calculated from daily Daymet values, centred on the years 1980, 1990, 2000, 2010 and 2020 (only 10 and 11 years were used respectively for 1980 and 2020, as Daymet data was only available from 1980 to 2021). Then, a linear interpolation was performed to get smooth yearly values between these 10-year periods. Mean values of regional predictors for each region are reproduced in Table S1.

Regarding temporal predictors, they included days of the week (one binary variable for each day), public holiday (0 or 1), month (one variable for each month from May to September) and year (as a numeric variable) (Table 4.2). In total, there were 47 predictors considered for the different models.

Tableau 4.2 Weather, regional and temporal predictors considered in this study. All data span from 1996 to 2019 across all regions for months May to September. Lagged variables were also used in the models, but not shown in the table (i.e., *tmean13* [at lags 1 to 3 days], *tmean47* [lags 4 to 7 days], *tmin13*, *tmin47*, *tmax13*, *tmax47*, *hmdx13*, *hmdx47*, *tdew13*, *tdew47*, *relh13*, *relh47*, *mtc03*, *mtc07*, *dtr1* [*dtr* at lag 1 day], *dtr2*, *dtr3*). Total population by region was used as an offset for the response variable (i.e., daily count / total population) rather than a predictor.

	Units	Abbreviation	Mean $\pm$ SD	Range	Source
Weather predictors					
Mean temperature	°C	<i>tmean</i>	16.0 $\pm$ 4.8	-2.6–29.6	Daymet
Minimum temperature	°C	<i>tmin</i>	10.5 $\pm$ 4.9	-6.5–24.3	Daymet
Maximum temperature	°C	<i>tmax</i>	21.5 $\pm$ 5.4	-0.9–35.9	Daymet
Mean humidex	-	<i>hmdx</i>	21.5 $\pm$ 8.1	-10.7–44.5	Daymet
Mean dew point	°C	<i>tdew</i>	10.3 $\pm$ 5.1	-10.6–24.2	Daymet
Relative humidity	%	<i>relh</i>	69.6 $\pm$ 9.8	19.8–95.7	Daymet
MTC over 1 day	°C	<i>mtc01</i>	0.0 $\pm$ 2.6	-15.3–11.8	Daymet
DTR	°C	<i>dtr</i>	11.1 $\pm$ 3.8	1.3–29.4	Daymet
Regional predictors					
Total population (<i>offset</i>)	10 ⁴ hab.	<i>poptot</i>	51.6 $\pm$ 46.4	9.0–206.6	MSSS
Elderly population (65+)	%	<i>pop65p</i>	15.2 $\pm$ 3.6	7.4–27.6	MSSS
Women population	%	<i>popsexf</i>	50.2 $\pm$ 0.6	48.7–52.0	MSSS
MSDI – Social component	-	<i>msdisoc</i>	-0.6 $\pm$ 1.0	-2.6–1.8	INSPQ
MSDI – Material component	-	<i>msdimat</i>	0.9 $\pm$ 1.8	-2.1–6.3	INSPQ
NDVI	-	<i>ndvi</i>	0.48 $\pm$ 0.06	0.36–0.60	CERFO
UHI measurement	°C	<i>uhi</i>	7.3 $\pm$ 1.9	3.3–10.5	CERFO
Historical summer temperature	°C	<i>hsumtemp</i>	16.0 $\pm$ 1.8	12.2–18.5	Daymet
Historical summer rel. humidity	%	<i>hsumhum</i>	69.5 $\pm$ 2.3	64.4–73.8	Daymet
Temporal predictors					
Day of week	-	<i>dow_1</i> to <i>dow_7</i>	-	0 or 1	-
Holiday	-	<i>hol</i>	-	0 or 1	-
Month	-	<i>month_5</i> to <i>month_9</i>	-	0 or 1	-
Year	-	<i>year</i>	Depends on health outcomes		-

SD = Standard deviation. Range = Minimum and maximum values. MTC = Mean temperature change. DTR = Diurnal Temperature Range. MSDI = Material and social deprivation index. NDVI = Normalized difference vegetation index. UHI = Urban heat island. MSSS = Ministère de la Santé et des Services Sociaux. INSPQ = Institut national de santé publique du Québec. CERFO = Centre d'enseignement et de recherche en foresterie de Sainte-Foy.

4.2.3 Machine learning models

Predicting each of the 9 health outcomes (Table 4.1) based on weather, regional and temporal predictors (Table 4.2) can be seen as supervised learning problem. Hence, five machine learning (ML) models adapted to that end were considered: Least Absolute Shrinkage and Selection Operators (Lasso), Random Forest (RF), Light Gradient Boosting Machine (LGBM), Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM). These models were selected to represent a diversity of architectures, required amounts of data, potential performance and novelty of the approach. In addition, most of these models were used in previous related studies, thus allowing for comparisons of our results with the latter (e.g., Boudreault et al., 2023; 2024a; Lee et al., 2019; Nishimura et al., 2021; Ogata et al., 2021; Park et al., 2020; Wang et al., 2019).

Lasso is a multiple regression model that automatically performs variable selection and regularization by penalizing the absolute size of the coefficients during model training (Tibshirani, 1996). RF is an ensemble tree-based method in which a forest of fully developed decision trees is built using bootstrapped datasets (Breiman, 2001). LGBM is also an ensemble tree-based method, but it uses trees with fewer leaves (called weak learners) that are sequentially added to the ensemble based on the residual of the last fitted tree(s) (Friedman, 2001; Ke et al., 2017). MLP and LSTM both belong to the family of deep learning models. MLP is the vanilla feedforward neural network that is characterized by multiple layers of fully connected nodes (neurons) inspired by the architecture of the human brain (Chapter 6 in Goodfellow et al., 2016). LSTM is a type of recurrent neural networks, which are naturally adapted for time series data, with the addition of memory cells that can retain information over long sequences (Hochreiter & Schmidhuber, 1997). More information about these models can be found in the above references or in Boudreault et al. (2023).

Following the holdout method for time series, ML models were calibrated using the first 70% of available years of data (*training*) and validated using the remaining 30% of years of data (*test*). Training years (and test years) were 1996–2012 (2013–2019) for mortality and hospitalizations, 2014–2018 (2019) for EDV and ambulance transports, and 2008–2016 (2017–2019) for 811 calls in both populations. Given this split of data, the models were thus validated in a close to reality context, i.e. with weather, regional and temporal predictors over years of data that the models have never seen in their training. It should be noted here that all regions were used in the training of the models. In other words, there is no external validation of the models on new regions. This was intended because the models developed here are meant to be used in Quebec only, and not in other provinces/countries where the administrative structure of healthcare may differ.

In Lasso, MLP and LSTM, predictors were first scaled to the 0–1 range as these models require scaled predictors. Then, given the low evidence found about optimal tuning of ML models in our specific context, an extensive grid search was performed. Tested hyperparameters (including models’ architectures) are reported in Table S2. The training dataset was used to find the hyperparameters that minimize the mean square error (MSE) in a 5-fold cross-validation grouped by years for Lasso, RF, LGBM and MLP and in a holdout validation of 30% of years of the training dataset for LSTM (because of the time series nature of this model). Models’ fitting was performed in Python using *scikit-learn* for Lasso, RF, LGBM and MLP (Pedregosa et al., 2011) and *keras* for LSTM (Gulli & Pal, 2017).

4.2.4 Models’ performance and explainability

Models’ performance was assessed in three different ways using the test dataset and based on two performance metrics: coefficient of determination (R^2) (eq. 4.1) and decrease in Mean Absolute Error ($dMAE$) compared to a non-informative model (i.e., a model containing only an intercept) (eq. 4.2) :

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (4.1)$$

$$dMAE = \frac{MAE_0 - MAE}{MAE_0} = 1 - \frac{\sum|y_i - \hat{y}_i|}{\sum|y_i - \bar{y}|} \quad (4.2)$$

where y_i is the observed response variable, $\hat{y}_i$ is the predicted response variable, MAE is the mean absolute error of the model, MAE_0 is the mean absolute error of a non-informative model and $\bar{y}$ is the mean response variable in the training dataset (equivalent to the intercept in a null model).

First, the overall model’s performance across all summer days of test years of data and regions was computed. Second, performance metrics were computed separately across the 15 regions to see how the performance was distributed among regions (again, on the test years of data only). These two first performance metrics correspond to the performance for the “heat-related” effect, as they are computed over all summer days. Third, the behaviour of the models during two well-known heatwaves in Quebec was visually inspected: the 2018 heatwave from June 30th to July 5th (Lebel et al., 2019) and the 2021 heatwave from June 6th to June 9th (Lamothe et al., 2023) (Figure S1). Note that the 2018 heatwave was part of the training set for EDV and ambulance transports, while the 2021 heatwave occurred during the COVID-19 pandemic that had an

important effect on reported numbers (Lamothe et al., 2023). These two heatwaves had to be selected due to the absence of heatwaves in 2019. Health and predictors data in 2021 were only used for this specific heatwave validation (i.e., they were not used in models' calibration nor in other types of evaluation). Validation during heatwaves was presented for daily mortality/morbidity values aggregated across the 15 regions (i.e., for the whole province of Quebec). This last validation of the models during specific extreme heat episodes was added to validate models' performance for the "heatwave" effect.

In addition to performance, models' explainability was also provided. Although many methods exist to that end and are still emerging (see Rasheed et al. (2022) for a review), feature importance (FI) was chosen for the current study as it can give a quick overview of the most important predictors for any type of model, thus allowing to compare results of various modelled health outcomes (9), machine learning models (5) and tested predictors (47) at a glance (e.g., Boudreault et al., 2023; Zhang et al., 2014). FI was computed based on permutation, which consists of randomly shuffling each predictor and quantifying the increase in mean square error (MSE) on the training dataset after shuffling. The more the MSE increases, the more the predictor was important to the model's predictive power. This process was repeated 100 times for each predictor and the average increase of MSE was taken. FI metrics were converted to ranks, rank #1 being the most important predictor and # p (i.e., the number of predictors) the least important one. Only the 15 most important variables were reported (out of a total of 47).

4.3 Results

The obtained results are described below, while their discussion and interpretation are presented in the Discussion (Section 4). Sections 4.3.1 and 4.3.2 present respectively the performance of the models across Quebec and by region through the summer. Overall, deep learning models (MLP and LSTM) seemed to outperform other approaches for most health outcomes in terms of out-of-sample R^2 and dMAE values. In section 4.3.3, performance during two major heatwaves (2018 and 2021) are reported. LGBM was the most sensible model to extreme heat events and better predicted the health effects during these two periods. Finally, section 4.3.4 highlights the most important predictors for all models and health outcomes. Overall, regional and temporal predictors were the most important variables, followed by weather factors such as temperature metrics. Optimal hyperparameters found for each ML model and health outcomes were reported in Supplementary Material (Table S3).

4.3.1 Performance across all regions

Models' performance in terms of R^2 and dMAE was computed on the test years of data across all regions in Table 4.3. All ML approaches performed well to predict daily morbidity variables (i.e., HOS, EDV, AMB and 811) with out-of-sample R^2 values of 60–95% for both general and elderly populations (Table 4.3, left side). However, mortality in both populations and ambulance transports in the elderly population were less well predicted with R^2 values of ~10% and ~30%, respectively, indicating weaker performance of all models to predict these health outcomes. Overall, ensemble tree-based methods (RF, LGBM) and neural networks (MLP and LSTM) outperformed Lasso. However, the difference in performance depended on the modelled health outcome. For example, there was a 10–15% increase in R^2 for hospitalizations and a 5% increase for EDV in both populations for RF, LGBM, MLP and LSTM compared to Lasso. The highest increase in R^2 was 15–20% for 811 calls in neural networks (MLP and LSTM) compared to Lasso. For mortality and ambulance transports, the difference between Lasso and other models was <5%. A few combinations of models and health outcomes had slightly worse performance compared to Lasso (e.g., RF, LGBM and LSTM for mortality in the elderly population; RF, LGBM and MLP for ambulance transports in the general population). These findings were similar when looking at the decrease in Mean Absolute Error (dMAE) performance criteria (Table 4.3, right side), although some slight differences were noted. For example, RF was the best model for mortality based on dMAE, while it was MLP based on R^2 . Across all health outcomes, LSTM and MLP were overall the most performing models for both performance criteria, but RF and LGBM were also sometimes the best approaches.

Tableau 4.3 Performance metrics during out-of-sample test years in terms of R^2 (left side) and decrease in Mean Absolute Error (dMAE) (right side) across all regions. Best performance metrics are in bold for each health outcome.

Health outcomes	R^2 (%)					dMAE (%)				
	Lasso	RF	LGBM	MLP	LSTM	Lasso	RF	LGBM	MLP	LSTM
MOR	11.5	11.1	12.2	13.0	12.9	2.8	4.8	4.7	4.5	4.7
HOS	66.4	73.1	74.2	71.8	71.2	43.0	51.0	52.6	50.5	49.2
EDV	92.5	95.4	96.9	97.8	97.4	73.5	79.6	83.0	87.0	86.0
AMB	57.4	56.8	57.2	55.8	59.7	33.5	33.8	34.5	30.8	36.7
811	65.9	59.1	59.8	82.1	82.7	46.9	35.8	36.4	60.5	61.2
MOR (65+)	11.0	8.8	9.3	11.3	10.4	9.5	7.4	7.5	10.1	9.8
HOS (65+)	50.9	63.8	65.3	65.1	63.9	32.0	45.2	46.2	46.3	45.5
EDV (65+)	87.1	90.9	91.3	90.4	91.4	64.3	72.6	72.2	71.3	72.9
AMB (65+)	28.9	31.4	31.2	30.4	29.5	16.5	18.3	17.9	17.9	17.1

4.3.2 Performance by region

The performance metrics were also reported separately for each of the 15 regions in terms of R^2 for the general population (Figure 4.2). dMAE values by region were also analyzed but were not shown because the same regional patterns were observed. Results for mortality (a) showed that R^2 reached up to 30–35% for regions #1 and #4 in all models. In other regions, R^2 was closer to the overall value of 10% found in section 3.1. For hospitalizations (b), only region #8 was less well predicted by all models. Hospitalizations in regions #6 and #13, respectively Montreal and Laval (which are highly populated regions, recall Table S1), were not well modelled by Lasso (R^2 of ~30%), but had a much better performance with the other ML models ($R^2 > 80\%$). The same behaviour was also observed for EDV (c) in regions #1 and #12 where performance was weaker for Lasso than for the other ML models. There was a high variability of performance in the modelled ambulance transports (d), with R^2 values ranging 0–80%. Finally, 811 calls (e) were worst modelled in regions #2, #3, #4 and #5, but regions #3 and #5 were improved with Lasso, MLP and LSTM models compared to RF and LGBM.

Results in the elderly population were comparable to the ones observed in the general population, with notable improvements in some regions by ensemble tree-based models and neural networks compared to Lasso (Figure S2). The greatest regional variability of performance was observed for mortality (a) and ambulance transports (d) in the elderly population, as also observed in the general population. In terms of hospitalizations (b) and EDV (c), these two health variables were slightly less well predicted in some regions for the elderly population compared to the general population.

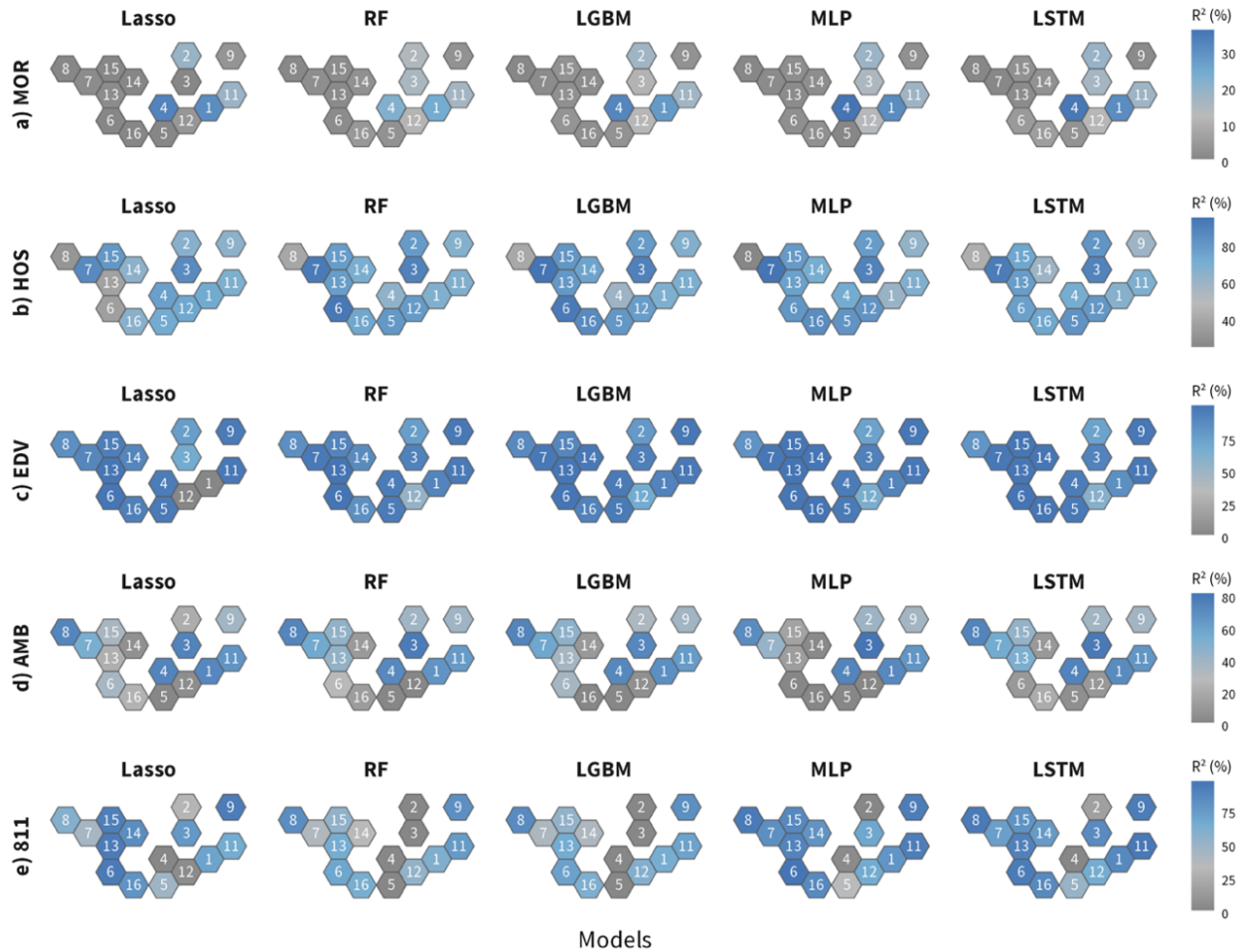


Figure 4.2 Regional performance in terms of R^2 in the 15 regions for all models in the general population for a) mortality, b) hospitalizations c) emergency department visits, d) ambulance transports and e) 811 calls. Each region is represented by a hexagon of equal size, irrespective of the actual area of the region. The geographical position of the regions has been preserved. The number in the hexagon indicates the region code as in Figure 4.1. Note that the scale of R^2 values is different across health outcomes as the goal is to compare regional performance for each health outcome, rather than to compare health outcomes with each other (refer to Table 4.3 to that end).

4.3.3 Validation during heatwaves

Observed and predicted values during 2018 and 2021 heatwaves across the province of Quebec were reproduced in Figure 4.3. The 2018 heatwave showed a high increase in observed daily mortality and ambulance transports, but no clear increasing trend in hospitalizations, EDV and 811 calls (solid black lines in Figure 4.3i). For mortality (i.a), LGBM predicted with the greatest accuracy the increased mortality, followed by MLP. For hospitalizations (i.b) and 811 calls (i.e), all models predicted well the trends in daily counts, but there was not an increased heatwave effect. For EDV (i.c), only RF predicted a peak in visits, while the observed data did not show any.

For ambulance transports (i.d), the two peaks also seemed better modelled with these two models (LGBM and MLP) compared to others (Lasso, RF and LSTM), even though the second peak was slightly underestimated by all models at different levels. During the 2018 heatwave, same results were generally found in the elderly population (Figure S3i) with some exceptions. For example, RF also modelled the peak of mortality well in the elderly population in 2018 along with LGBM (i.a). Also, only LGBM and LSTM reproduced the first increase in ambulance transports during 2018 heatwave in the elderly population, but all models underestimated the second peak later during that heatwave (i.d). There was no notable difference in hospitalizations (i.b) and EDV (i.c) in 2018 in the elderly population compared to the general population.

The 2021 heatwave was highly influenced by changes in mortality and morbidity due to the COVID-19 pandemic. Indeed, most models were no longer aligned with the observed values (Figure 4.3ii). Still, the patterns in the models during that heatwave could be analyzed. For example, the mortality peaked less than in 2018 and none of the models predicted an increase in mortality during that period (ii.a). It is worth recalling that this heatwave was not as hot and long as the 2018 one (Figure S1). The hospitalizations were still insignificantly impacted by the heatwave and all models reproduced that behaviour well (ii.b). The peak of EDV (ii.c) and ambulance transports (ii.d) observed in 2021 was well detected by most models, except RF for ambulance transports. Finally, 811 calls were predicted to increase by all models, but not such an increase was observed (ii.e). There was no noteworthy difference in the results for the elderly population during the 2021 heatwave compared to the general population (Figure S3ii).

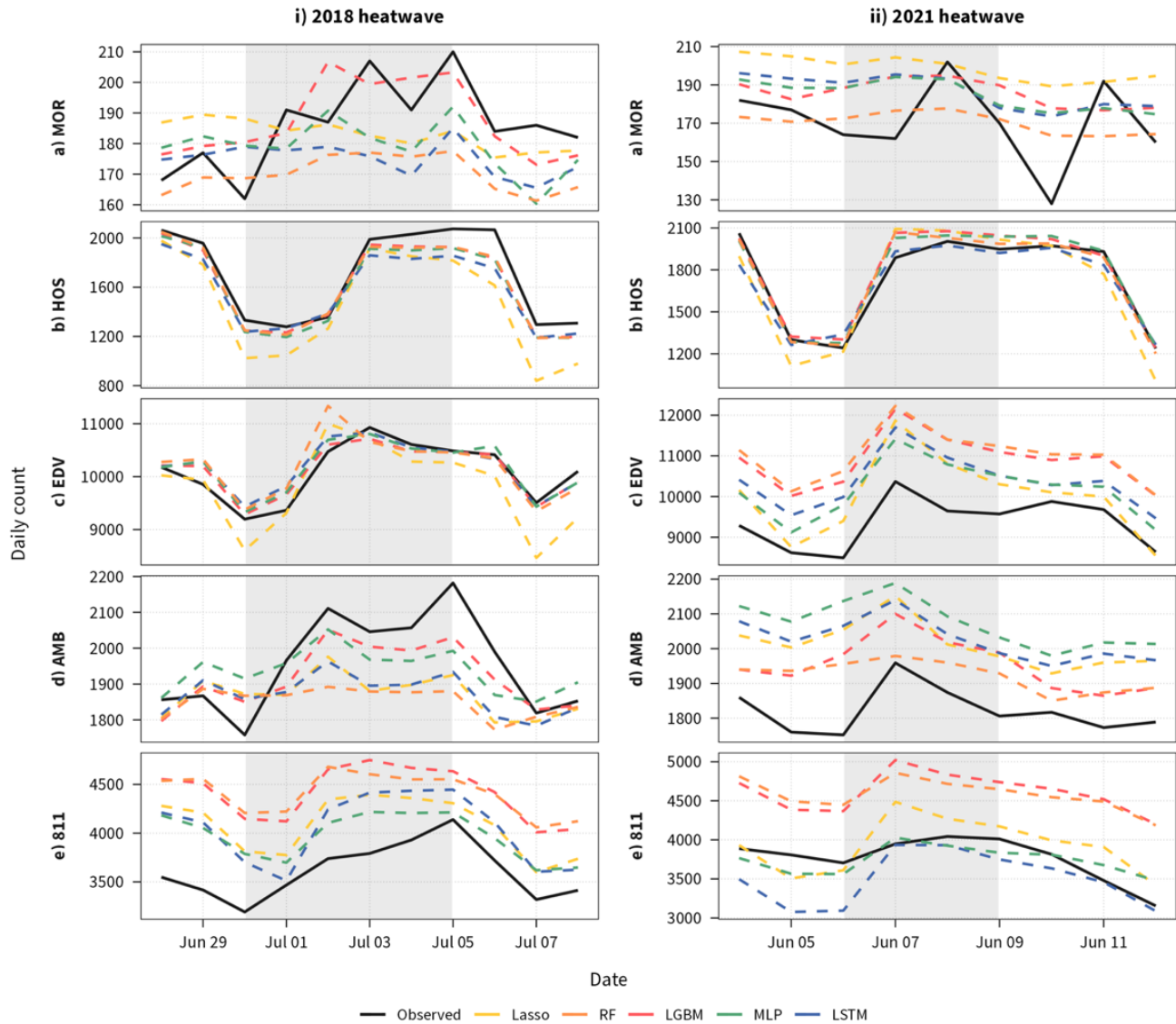


Figure 4.3 Observed (solid black line) and predicted (dashed coloured lines) daily counts of a) mortality, b) hospitalizations c) emergency department visits, d) ambulance transports and e) 811 calls in all studied regions for the general population during i) 2018 and ii) 2021 heatwaves. Heatwaves' periods are highlighted in grey.

4.3.4 Models' explainability

The 15 first most important predictors (out of 47) for all modelled health outcomes were extracted from all models for the general population (Table 4.4; refer to Table 4.1 for a description of the abbreviations). Overall, regional predictors, such as *pop65p* (percentage of population 65 and over), *hsumtemp* (historical summer temperature), *hsumhum* (historical summer humidity) and *msdisoc* (social deprivation index), appeared in the first 5 ranks for all health outcomes and

models. These variables can account for regional differences in the modelled health outcomes. Temporal predictors were also among the most important variables in ranks 1 to 15, depending on the health outcome and/or model. The most important temporal variables were associated with daily (*dow_*), yearly (*year*) and, less frequently, monthly (*month_*) patterns in the modelled health outcomes. It is worth noting that Saturday (*dow_6*) and Sunday (*dow_7*) were the two most important predictors in hospitalizations for all models (Table 4.4b).

Weather variables seemed less important (at ranks 7 and above) than regional and temporal predictors across all models (Table 4.4). However, major differences between models and health outcomes were noted. In terms of models, weather variables were generally more important in Lasso, RF and LGBM than in MLP and LSTM. In terms of health outcomes, weather predictors were more important for mortality (Table 4.4a) and ambulance transports (d), at ranks 7 to 10 approximately. They were respectively *tmax* (maximum temperature), *mtc07* (mean temperature change over 7 days), *tmean* (mean temperature) and *hmdx* (humidex) for mortality and *mtc07*, *tmax47* (maximum temperature at lags 4 to 7 days), *tmax13* and *hmdx13* for ambulance transports, according to Lasso, RF and LGBM. For hospitalizations (b), temperature exposures at 4 to 7 previous days appeared in ranks 13 to 15 in RF, but these variables did not seem important in the other models. For EDV (c), *tmax13*, *mtc07* and *tmax47* were the most important weather predictors (ranks 14–15) in RF and LGBM. For 811 calls (e), 4 to 7 days lagged humidex (*hmdx47*) and dew point (*tdew47*) exposure were among the most important predictors in Lasso and LGBM, while *tdew* and *tmin* (minimum temperature) were found in MLP.

In the elderly population, similar overall results as above were found, that is, regional and temporal variables were the most important predictors followed by weather variables (Table S4). Weather variables were again more important for mortality (a) and ambulance transports (d) in the elderly population, but with some notable differences. Change of temperature over 3 (*mtc03*) and 7 (*mtc07*) days seemed more important than direct temperature exposure in the elderly population compared to the general population for these two health outcomes. Even though weather variables were less important for hospitalizations (b), the 4 to 7 days lagged temperature exposure was consistent in the two populations. For EDV (c), higher lagged temperatures were found to be more important in the elderly compared to the general population.

Tableau 4.4 Top 15 most important predictors for all models in the general population for a) mortality, b) hospitalizations, c) emergency department visits, d) ambulance transports and e) 811 calls. Weather variables are in light orange, temporal predictors in blue and regional predictors in grey. Refer to Table 4.2 for a description of each abbreviation.

		Feature importance (FI) rank														
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
a) MOR	Lasso	pop65p	year	uhi	hsumtemp	msdisoc	ndvi	hsumhum	popsexf	hmdx13	tmax	msdimat	month_5	tdew	mtc07	hmdx47
	RF	pop65p	hsumtemp	msdimat	popsexf	msdisoc	hsumhum	ndvi	uhi	year	mtc07	tmax47	mtc03	tmean47	hmdx47	relh13
	LGBM	pop65p	hsumtemp	hsumhum	msdisoc	msdimat	popsexf	tmax13	mtc07	tmax47	uhi	relh13	mtc01	ndvi	tmax	relh47
	MLP	pop65p	hsumtemp	msdimat	uhi	popsexf	msdisoc	year	hsumhum	ndvi	month_5	month_6	month_9	dow_6	dow_3	dow_1
	LSTM	pop65p	hsumtemp	uhi	ndvi	dow_3	dow_2	msdisoc	dow_7	dow_1	msdimat	month_8	month_9	month_7	dow_4	dow_5
b) HOS	Lasso	dow_6	dow_7	year	pop65p	popsexf	hol	hsumhum	year	hsumtemp	dow_5	msdimat	month_7	month_8	month_9	ndvi
	RF	dow_6	dow_7	hsumtemp	hsumhum	dow_5	hol	msdisoc	uhi	msdimat	popsexf	pop65p	hmdx47	tdew47	tmean47	
	LGBM	dow_6	dow_7	hsumtemp	hol	hsumhum	year	dow_1	dow_2	uhi	dow_3	msdisoc	dow_4	msdimat	month_9	month_7
	MLP	dow_6	dow_7	year	uhi	hsumtemp	msdimat	hsumhum	pop65p	popsexf	dow_2	dow_1	dow_3	hol	dow_4	msdisoc
	LSTM	dow_6	hsumtemp	year	dow_7	uhi	dow_1	dow_2	hsumhum	dow_3	dow_4	pop65p	popsexf	month_9	month_5	msdimat
c) EDV	Lasso	hsumtemp	msdimat	hsumhum	uhi	ndvi	pop65p	msdisoc	dow_6	dow_1	dow_7	year	dow_2	hol	month_5	popsexf
	RF	msdisoc	hsumtemp	msdimat	pop65p	dow_6	hsumhum	dow_1	dow_7	uhi	year	dow_2	popsexf	dow_5	mtc07	tmax13
	LGBM	msdisoc	hsumtemp	msdimat	hsumhum	dow_6	pop65p	dow_1	dow_7	dow_2	popsexf	uhi	hol	month_5	tmax47	tmax13
	MLP	hsumtemp	popsexf	msdimat	msdisoc	uhi	ndvi	pop65p	dow_1	hsumhum	year	dow_6	month_5	month_9	dow_2	dow_7
	LSTM	hsumtemp	msdisoc	msdimat	uhi	popsexf	pop65p	hsumhum	dow_1	dow_6	year	ndvi	dow_7	dow_2	month_5	dow_4
d) AMB	Lasso	pop65p	hsumhum	tmean	hmdx	hmdx13	msdisoc	uhi	tmax	tmax13	tdew13	hmdx47	msdimat	tdew47	hsumtemp	tmin47
	RF	pop65p	hsumhum	hsumtemp	ndvi	uhi	msdimat	popsexf	msdisoc	mtc07	year	mtc03	dtr	mtc03	relh	tmean
	LGBM	pop65p	hsumhum	hsumtemp	popsexf	ndvi	msdisoc	tmax	msdimat	dow_1	mtc07	year	uhi	dtr	tmax13	mtc03
	MLP	pop65p	hsumhum	popsexf	hsumtemp	uhi	msdimat	msdisoc	ndvi	dow_2	dow_3	dow_6	month_8	month_5	month_9	dow_7
	LSTM	pop65p	hsumhum	msdisoc	ndvi	popsexf	uhi	hsumtemp	msdimat	dow_8	month_9	month_7	month_5	dow_2	dow_3	month_6
e) 811	Lasso	year	hsumtemp	uhi	popsexf	pop65p	hmdx47	tdew47	hmdx13	tdew13	msdimat	ndvi	dow_6	hmdx	dow_7	hsumhum
	RF	hsumhum	msdisoc	msdimat	year	dow_6	popsexf	hsumtemp	dow_7	uhi	ndvi	pop65p	dow_1	tdew47	hmdx47	relh47
	LGBM	hsumhum	msdisoc	year	dow_6	msdimat	dow_7	popsexf	uhi	dow_1	pop65p	hsumtemp	ndvi	month_5	month_6	month_8
	MLP	year	msdimat	uhi	pop65p	hsumhum	msdisoc	hsumtemp	popsexf	ndvi	dow_6	dow_7	dow_1	tdew	month_5	tmin
	LSTM	dow_1	year	dow_2	dow_3	msdimat	month_5	dow_4	dow_5	month_6	hsumhum	month_9	msdisoc	month_7	pop65p	month_8

Variable type: Weather Temporal Regional

4.4 Discussion

This study considered 5 machine learning (ML) models to predict several heat-related health outcomes in 15 regions in the province of Quebec (Canada). ML models trained with weather, regional and temporal predictors from multiple locations allowed to predict a health outcome in each of the regions simultaneously, without the need to (re)fit the same model to different locations (e.g., Ogata et al., 2021; Zhang et al., 2014; Boudreault et al., 2024a) or being limited to a single-location study (e.g., Boudreault et al., 2023; Lee et al., 2022; Navares et al., 2018). The models' calibration (initial and final hyperparameters, packages and software used), validation (overall, by regions and during heatwaves) and explainability (with feature importance) was transparent and comprehensive, which contrasts with the existing literature in population health (Morgenstern et al., 2020) and previous studies (e.g., Lee et al., 2022; Navares et al., 2018; Park et al., 2020; Wang et al., 2019; Zhang et al., 2014). To our knowledge, this is the first study to consider such a wide range of ML approaches, including deep learning, to model 9 health outcomes in several locations simultaneously.

In all the considered models, morbidity variables in both populations were well modelled with out-of-sample R^2 values of >60%, while mortality and ambulance transports in 65+ population were less accurately predicted with R^2 values of <30% (recall Table 4.3). This may be explained by different reasons, such as 1) the use of all-cause indicators instead of heat-related illnesses, 2) the presence of more noise/randomness in these health outcomes compared to the other ones, 3) a change of relationship in the test vs. the training period, and 4) a lack of predictors to correctly capture the relationship. Unfortunately, the focus on modelling a single health outcome in the literature limited the possibility to compare our results on different health outcomes with other studies. In terms of modelling approaches, deep neural networks (MLP and LSTM) were overall the best-performing approaches across the 9 health outcomes modelled, followed by ensemble tree-based approaches (RF and LGBM) when looking at the overall summer period. For some health outcomes, increases in out-of-sample R^2 of up to 20% were noted for neural networks compared to Lasso. Lasso was never the best modelling approach in our study.

In the literature, a few other studies compared the results of various ML models to predict heat-related health outcomes. For example, Lee et al. (2022) found that penalized regression and tree-based methods outperformed LSTM for all-cause and cause-specific mortality predictions in Seoul, South Korea. Boudreault et al. (2023) reported that ensemble tree-based models had better performance than neural networks and statistical models for mortality prediction during summer in Montreal, Canada. LSTM was found to be more efficient than RF and non-linear equations to model heat-related illnesses in the study of Nishimura et al. (2021) in Nagoya, Japan. Navares et al. (2018) found the best performance with the Auto-Regressive Integrated Moving Average (ARIMA) model and neural networks compared to tree-based methods for hospital admissions prediction in Madrid, Spain. These comparisons with international studies should be made with caution given the differences in study population, healthcare management, models used and studied health outcome. In addition, all the above studies were conducted with data of only one region/city and for a single health outcome. Indeed, other past multi-region studies mainly employed a single tree-based approach (e.g., Ke et al., 2023; Park et al., 2020; Wang et al., 2019) instead of various machine and deep learning models as in our study, an important novelty. Thus, this can explain the differences in performance observed in other studies and why deep learning was found to be superior in ours.

In addition to overall performance, further validations by region and during heatwaves presented a more complete picture of models' performance. First, the performance by regions showed that some regions were considerably better modelled by neural networks models compared to Lasso,

while some regions remained poorly modelled by all approaches (recall Figures 4.2 and S2). The 2018 heatwave showed sharp increases in mortality and ambulance transports that were well predicted by LGBM (for mortality), LGBM/RF (elderly mortality) and LGBM/MLP (ambulance transports) (recall Figures 4.3 and S3). These results were consistent during the 2021 heatwave, although COVID-19 had a strong influence on the reported numbers (e.g., mortality displacement, lower use of health services) and that heatwave was not as hot and long as 2018's (recall Figure S1). These additional validations, that have not been computed nor presented in previous studies, have highlighted some key information that overall performance metrics cannot reveal. They thus opened the door to future improvements to these multi-region models. For example, weighting observations when training models could ensure more accurate predictions in the most densely populated regions, or in those with the weakest performance. Furthermore, as the best model may vary depending on how performance is evaluated as seen above and in a previous study (Boudreault et al., 2023), stacked models could be considered in the future to “pool” the strengths of different approaches and result in a model likely to work better in various situations (e.g., Lu & Qiu, 2023; Navares et al., 2018).

To supplement these validations, some models' explainability was also provided, limiting the “black-box” feeling of the ML methodology when this information is not provided (e.g., Lee et al., 2022; Navares et al., 2018). Feature importance (FI) showed that regional and temporal predictors were more important than weather variables for all health variables, underlining the presence of strong regional and temporal patterns in the modelled health variables (recall Tables 4 and S4). This result was expected given that the raw daily counts of the different health outcomes (adjusted by population) were modelled, which is commonly done in the literature (e.g., Lee et al., 2022; Navares et al., 2018; Ogata et al., 2021; Park et al., 2020; Wang et al., 2019). Modelling raw daily data is an advantage as it limits the data preparation steps such as trends/patterns removal required in other studies (e.g., Boudreault et al. 2023, 2024a, Masselot et al. 2018, Zhang et al. 2014). The lower importance found for weather variables can also be explained by the inclusion of highly correlated temperature variables in the models. This may have prevented the detection of these variable with permutation-based FI metrics. For example, if *tmean* is no longer available to the model, but *tmax* and *tmin* are, the model can still perform relatively well, leading to a lower FI value for *tmean* (the same also occur for the other temperature variables following the same logic). However, considering all these temperature variables (even if highly correlated) was important as they can contribute to a better understanding of weather patterns and, consequently, improve the predictive power of heat-related health outcomes by the models.

Regarding weather, some key meteorological information for public health planning and surveillance was revealed. On one hand, mortality and ambulance transports in the general population were the most sensitive to short-term heat exposure (recall Tables 4 and S4) and heatwaves (recall Figures 4.4 and S3), supporting their use as early indicators of heat effects. On the other hand, hospitalizations, EDV and 811 calls were found to be associated with more delayed temperature effects. Finally, temperature changes could potentially be more important than direct temperature exposure in the elderly population. These divergent results highlight the need for differentiated analyses by population strata, which can lead to better-targeted heat-related prevention actions. These potentially new insights were made possible because of the introduced ML models combining weather information from multiple regions and by the large number of health outcomes studied. They could help health authorities to better plan the health burden during heatwaves and deliver appropriate messages to different vulnerable populations. However, explaining the pathways by which these weather variables may be linked to health effects is beyond the scope of the current study and will need to be confirmed by further research.

As the proposed models were trained with readily available predictors, they can be used to monitor health effects of heat over different time horizons in future applications, thus acting towards more resilient, healthier and sustainable cities. On the one hand, used in combination with weather forecasts, they can predict the short-term effects of heat on various health indicators and limit the heat effect on population through prevention. On the other hand, by leveraging climate and socioeconomic projections, these models can assess the long-term heat-related health burden in the coming decades, helping authorities to better plan climate change adaptation measures to reduce the heat burden. With these new tools in hand, public health authorities and decision-makers will have a better understanding of complex interactions between climate and health and a clearer portrait of the heat impacts in multiple locations and for several health outcomes. Ultimately, adopting these approaches will increase the resiliency of the society considering climate change ahead.

The strengths of the study must be highlighted. First, five ML models were considered and compared, including penalized regression (Lasso), ensemble tree-based methods (RF and LGBM) and deep learning (MLP and LSTM). Second, these models were trained using a comprehensive database of meteorological (e.g., temperature exposure and change, humidity), regional (e.g., demographic, socioeconomic, built environment, climatology) and temporal (e.g., day, month and year) predictors. Third, these models allowed for health outcomes predictions simultaneously in multiple locations without the need to create one model per location as it would

be the case for traditional statistical/ML models. Fourth, 9 health outcomes, including both mortality and morbidity (hospitalizations, EDV, ambulance transports and 811 calls) in two populations (general and the elderly), were modelled, providing a more complete picture of the heat effects than in other studies. Finally, our study was thoroughly transparent in terms of ML models' calibration, validation and explainability, which contrasted with existing literature.

Some drawbacks must also be noted. First, only all-cause health outcomes were modelled as they were more readily available information. This could explain the low performance obtained for some health outcomes. Studying cause-specific mortality or morbidity, such as heat-related illnesses that may be more closely related to weather variables, could lead to better performance and is left for future research. Second, the models were developed using data from all southernmost regions of Quebec, Canada. Developing models for each region individually could also be of interest to compare with the results obtained here, but was deemed out-of-scope for the current study. Third, the validation of the models was performed on out-of-sample years of data, but not on out-of-sample regions. This latter validation was deemed irrelevant in our study, since the developed models were intended to be used only in Quebec's regions. Fourth, not all predictors could be included in our models (such as air pollution due to data unavailability), which can explain the weaker performance obtained for some health outcomes. Finally, other models' explainability techniques, in addition to predictions during heatwaves and feature importance already presented, could have been explored to further explain ML models, such as the predicted responses for different temperature values (e.g., 10°C, 20°C or 30°C) or SHAP values.

4.5 Conclusion

Deep learning models (MLP and LSTM) were the most promising tools to predict several heat-related health impacts in multiple regions of Quebec (Canada) during summer, with out-of-sample R^2 of >60% for most of the health outcomes studied. However, tree-based methods (RF and LGBM) also performed best for some health outcomes, and LGBM was found to be more accurate during heatwaves. These new tools based on machine and deep learning can be used by decision-makers to better predict the heat burden during summer, plan resources and actions appropriately, and limit the consequences of increasing heat episodes in a climate change context. They therefore contribute directly to making cities more resilient and, more broadly, to create a more sustainable and healthier society. Finally, this general framework can also be used in other regions of the world experiencing the negative effects of heat, as well as extended to study the joint effect of multiple exposures such as extreme heat and air pollution.

Data availability

Authors do not have permission to share health data. Socioenvironmental data are freely available from the following organization: NASA for weather data, Ministère de la Santé et des Services Sociaux du Québec (MSSS) for demographic data and Institut national de la santé publique du Québec (INSPQ) for socioeconomic and built environment data.

CRedit authorship contribution statement

Jérémie Boudreault: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Annabel Ruf:** Writing – review & editing, Visualization, Software, Funding acquisition, Formal analysis. **Céline Campagna:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Fateh Chebana:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Denis Hamel and Louis Rochette from the *Bureau d'information et d'étude de la santé de la population* (BIESP) of INSPQ for the help with the health data extraction, Antoine Saint-Amand, Mathieu Tandonnet and Nathalie Gravel from the Geomatics team at INSPQ for their help with the geospatial data extraction, as well as Yohann Chiu, Magalie Canuel, Ray Bustinza, Felix Lamothe and Ernest Lo, also from INSPQ, for their comments on early versions of this work. The authors would like to thank the editor, the associate editor, the guest editor and two anonymous reviewers for their constructive comments that helped improve this manuscript.

Funding

The first author would like to acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (Vanier Scholarship #CGV-180821), the Canadian Institute of

Health Research (Health System Impact Fellowship #IF1-184093), Ouranos (Real-Décoste Excellence Scholarship #RDX-317725) and the National Institute of Public Health of Quebec (no grant number). Annabel Ruf received funding from the German Academic Exchange Service (DAAD) through the Mitacs Globalink internship program (#57679026).

Supplementary materials

Supplementary material associated with this article can be found in the online version at:

<https://doi.org/10.1016/j.scs.2024.105785>

CHAPITRE 5 :
ESTIMATING THE HEAT-RELATED MORTALITY
AND MORBIDITY BURDEN IN THE PROVINCE
OF QUEBEC, CANADA

5 ESTIMATING THE HEAT-RELATED MORTALITY AND MORBIDITY BURDEN IN THE PROVINCE OF QUEBEC, CANADA

Estimation de la mortalité et de la morbidité liées à la chaleur dans la province de Québec, Canada

Auteurs : Jérémie Boudreault¹², Éric Lavigne³⁴, Céline Campagna¹²⁵, Fateh Chebana¹

¹ Centre Eau Terre Environnement, Institut national de la recherche scientifique (INRS), 490 de la Couronne, Québec (QC) Canada, G1K 9A9.

² Direction de la santé environnementale, au travail et de la toxicologie, Institut national de santé publique du Québec (INSPQ), 945 avenue Wolfe, Québec (QC) Canada, G1V 5B3.

³ Environmental Health Science and Research Bureau, Health Canada, 251 Sir Frederick Banting Driveway, Ottawa (ON) Canada, K1A 0K9

⁴ School of Epidemiology & Public Health, University of Ottawa, 600 Peter Morand Crescent, Ottawa (ON) Canada, K1G 5Z3

⁵ Département de médecine sociale et préventive, Faculté de médecine, Université Laval, 2300 rue de la Terrasse, Québec (QC) Canada, G1V 0A6.

Titre de la revue : *Environmental Research*

Publié le 4 juin 2024

<https://doi.org/10.1016/j.envres.2024.119347>

Contribution des auteurs :

J. Boudreault a été le principal contributeur de cet article, avec la conceptualisation de l'étude, la méthodologie, l'analyse des données, la modélisation, la production des figures et tableaux, et l'écriture du manuscrit. É. Lavigne a participé à l'écriture des codes. É. Lavigne, C. Campagna et F. Chebana ont aidé à conceptualiser l'étude et la méthodologie, ont validé les résultats, révisé l'article et approuvé sa version finale.

Lien entre cet article et les précédents :

Les trois premiers articles (chapitres 2, 3 et 4) ont permis de comparer différents modèles d'apprentissage automatique et profond pour prédire la mortalité et la morbidité liées à la chaleur au Québec. Or, ces modèles de prédiction n'ont pas été développés afin d'obtenir des fonctions dose-réponse, et donc, d'estimer le fardeau sanitaire de la chaleur, ce qui est nécessaire pour le calcul subséquent des coûts sanitaires. Ainsi, ce quatrième article propose de mettre à profit l'approche la plus utilisée en épidémiologie environnementale, c'est-à-dire le DLNM et la méta-régression multivariée, afin de quantifier le fardeau sanitaire de la chaleur. Il réutilise les mêmes données qu'au chapitre précédent, c'est-à-dire cinq indicateurs de santé incluant la mortalité et la morbidité, toutes causes, dans l'ensemble des 15 RSS du Québec chez la population générale. En outre, il compare les résultats de différentes métriques de température comme aux deux articles précédents (p. ex. température moyenne, maximale, minimale, l'indice humidex et le point de rosée).

Abstract

Background: As climate change increases the frequency and intensity of extreme heat events, there is an urgent need to quantify the heat-related health burden. However, most past studies have focussed on a single health outcome (mainly mortality) or on specific heatwaves, thus providing limited knowledge of the total pressure heat exerts on health services.

Objectives: This study aims to quantify the heat-related mortality and morbidity burden for five different health outcomes including all-cause mortality, hospitalizations, emergency department (ED) visits, ambulance transports and calls to a health hotline, using the province of Quebec (Canada) as a case study.

Methods: A two-step statistical analysis was employed to estimate regional heat-health relationships using Distributed Lag Non-Linear Models (DLNM) and pooled estimates using a multivariate meta-regression. Heat burden was quantified by attributable fraction (AF) and attributable number (AN) for two temperature ranges: total heat (above the minimum mortality/morbidity temperature) and extreme heat (above the 95th percentile of temperature).

Results: Higher temperatures were associated with greater risk ratios for all the health outcomes studied, but at different levels. Significant AF ranging from 2–3% for the total heat effect and 0.4–1.0% for extreme heat effect were found for all health outcomes, except for hospitalizations that had an AF of 0.1% for both heat exposures. The estimated burden of all heat (and extreme heat) every summer across the province was 470 (200) deaths, 225 (170) hospitalizations, 36 000 (6 200) ED visits, 7 200 (1 500) ambulance transports and 15 000 (3 300) calls to a health hotline, all figures significant.

Discussion: This new knowledge on the total heat load will help public health authorities to target appropriate actions to reduce its burden now and in the future. The proposed state-of-the-art framework can easily be applied to other regions also experiencing the adverse effects of extreme heat.

Keywords: hot temperature, extreme heat events, health outcomes, distributed lag non-linear model, multivariate meta-regression, attributable fraction.

Highlights

- The heat-related burden of five health outcomes (HO) quantified in Quebec, Canada
- DLNM and meta-regression used to compute attributable fraction (AF) and number (AN)
- All HO showed increased risk ratios at higher temperatures, but at different levels
- AF to heat were 2–3% for all HO except for hospitalizations (0.1%), all significant
- Depending on the HO, AN to heat ranged from 100s to 10 000s each summer in Quebec

5.1 Introduction

It is well established that elevated ambient temperatures (i.e., heat) exacerbate health conditions, leading to increased mortality and morbidity (Basu, 2009; Campbell et al., 2018; Kovats & Hajat, 2008; Song et al., 2017). These heat-related health effects can translate into massive costs borne by the healthcare system and society in general (Callahan & Mankin, 2022; Wondmagegn et al., 2019). Because of climate change and population ageing, these impacts will become even more devastating in the future (Curtis et al., 2017; Huang et al., 2011; 2013). Therefore, there is an urgent need to understand and quantify the current heat load on health to reduce its effects now and in the future.

In the recent years, the quantification of the heat burden became increasingly important (e.g., Ballester et al., 2023; Burkart et al., 2021; Gasparrini et al., 2015; Vicedo-Cabrera et al., 2021; Zhao et al., 2021). However, previous studies on temperature-related health burden have focussed mainly on mortality, as recently reviewed by Cheng et al. (2019). The same emphasis on mortality has also been observed more generally in the heat-health literature (Campbell et al., 2018; Cole et al., 2023). Although heat-related mortality is a major concern, the burden on the healthcare system should also be measured in terms of morbidity indicators such as hospitalizations, emergency department (ED) visits and ambulance transports (Wondmagegn et al., 2019), among others. While studies estimating morbidity-health relationships exist (see Li et al. (2015) and Ye et al. (2012) for reviews), a more limited number of studies have taken a step further and quantified the morbidity burden associated with heat (e.g., Bai et al., 2016, 2018; Cheng et al., 2016; Lin et al., 2012; Liu et al., 2019; Wellenius et al., 2017; Wondmagegn, 2021a, 2021b; and Cheng et al. (2019) for a review). Finally, most of the aforementioned studies were limited to a single health outcome (i.e., either mortality or one morbidity variable), thus providing a limited portrait of the overall heat burden.

In addition, past research on the heat burden has sometimes focussed on specific episodes of extreme heat such as “official” heatwaves defined by public health authorities (e.g., Adélaïde et al., 2021; Bustinza et al., 2013; Fouillet et al., 2006; Knowlton et al., 2009; Limaye et al., 2019) or on the additional heatwave effect above average temperatures (e.g., Guo et al., 2018; Yin et al., 2018). While these studies are important, they only cover part of the total heat load. Indeed, heat can have adverse effects at temperatures below heatwave thresholds, as well as on singular days of hot temperature (i.e., non-consecutive) (Campbell et al., 2018; Song et al., 2017). Moreover, there exists many different definitions of heatwaves (e.g., using different temperature thresholds, single or multiple day average, etc.), in addition to the “official” heatwaves, that can all lead to

diverging results when quantifying the heat burden (e.g., Cheng et al., 2018; Kanti et al., 2022; Kent et al., 2014; Xu et al., 2016).

In this study, a detailed analysis of the total heat burden of 5 health outcomes, including both mortality and morbidity (i.e., hospitalizations, ED visits, ambulance transports and calls to a health hotline) was performed, using the province of Quebec (Canada) as a case study. Heat-health exposure-response functions were developed using Distributed Lag Non-Linear Models (DLNM) and multivariate meta-regressions. They were then used to compute attributable fractions and numbers to both total and extreme heat. This study provides, for one of the first time, a comprehensive portrait of the overall heat burden on various mortality and morbidity indicators using a unified and state-of-the-art methodology.

5.2 Material and methods

5.2.1 Study design and health outcomes

A time series ecological study in the 15 southernmost health regions (HR) in the province of Quebec, Canada, was performed (Figure 5.1). Quebec has a population of approximately 8.5M inhabitants and is the second most populous province of Canada, with 23% of the Canadian population. Given the diversity of socioeconomic and climate conditions across its HR, Quebec is a good candidate for this study. In addition, Quebec's health data was readily available for multiple health outcomes, which was not systematically the case for the other provinces of Canada. Finally, most heat burden studies performed in Quebec have focussed on heatwaves only (e.g., Bustinza et al., 2013; Lebel et al., 2017, 2019). This project received ethics approval from the Human Research Ethics Committee of the National Institute of Scientific Research (CER-22-693).

The HR spatial resolution was chosen because of its relevance for decision makers (i.e., each HR has its own public health team and director). Also, it represents a finer resolution of previously developed "climate regions" for heat-health studies in Quebec, that are combinations of multiple HR (e.g., Chebana et al., 2013; Tupinier Martin et al., 2024). As in previous studies, the northernmost HR #10, #17 and #18 were excluded due to their low population (i.e., <1% of Quebec population) and their colder climate. Associations between daily mortality or morbidity and ambient temperature in each HR during the months of May to September were evaluated.

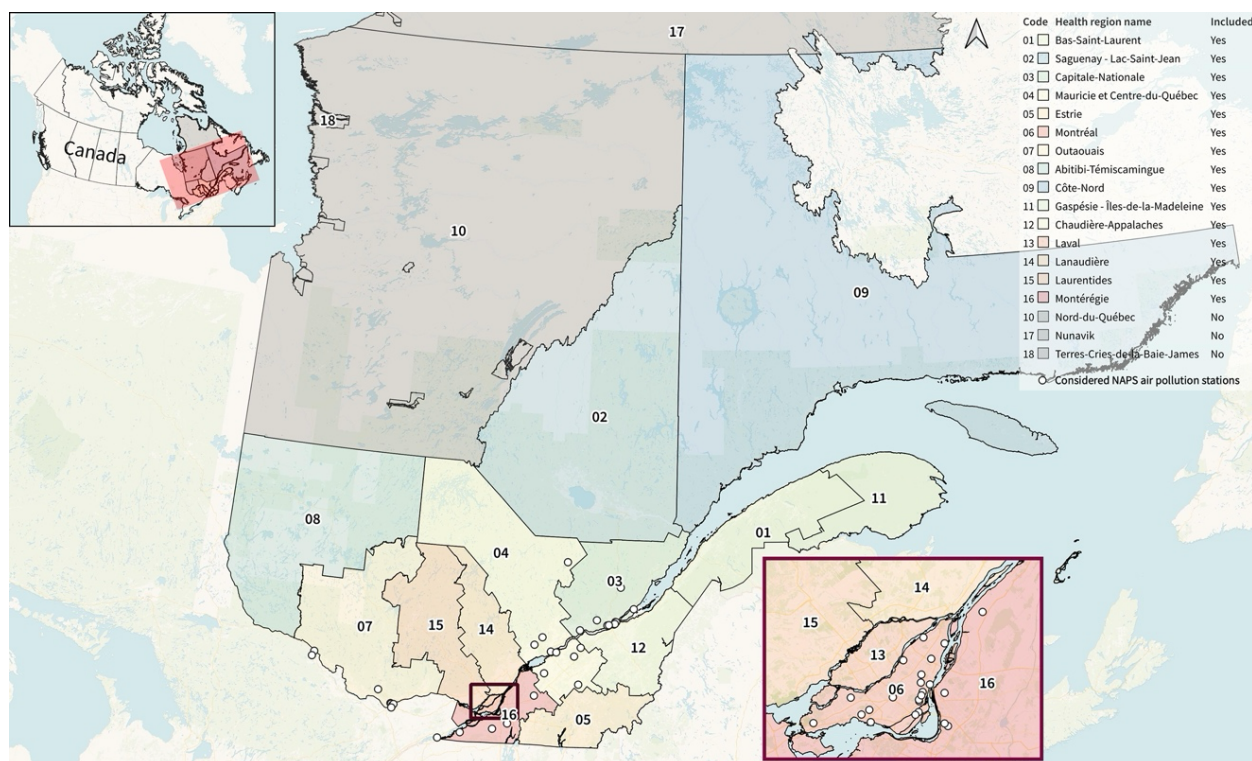


Figure 5.1 Map of the studied health regions (HR) in the province of Quebec (Canada) and location of the considered air pollution stations of the National Air Pollution Surveillance (NAPS) program. HR #10, #17 and #18 were excluded from the present study.

Health data was made available by the *Institut national de santé publique du Québec* (INSPQ) at the HR level, based on 2021 administrative boundaries for all years of data. Studied health outcomes were the daily cases of all-cause: 1) mortality, 2) hospitalizations, 3) ED visits, 4) ambulance transports and 5) *Info-Santé* 811 calls (a free hotline service for non-urgent health issues). All-cause indicators were selected instead of cause-specific (e.g., cardiovascular diseases) as they are more readily available indicators tracked by public health authorities in Québec through the health surveillance system of extreme weather events called *SUPREME* (Toutant et al., 2011), although they may include some non-heat-related conditions. Mortality and hospitalization data were available from 1996, ED visits and ambulance transports from 2014 and 811 calls from 2008. All health outcomes were considered until 2019, prior to the COVID-19 pandemic.

5.2.2 Weather, air pollution and regional data

Daily maximum and minimum temperature and water vapour pressure data at 1 km × 1 km over North America were extracted from NASA's Daymet database for the 1996–2019 period (Thornton et al., 2022). Mean temperature was computed from the average of daily maximum and

minimum temperatures. Relative humidity, dew point and humidex were computed from vapour pressure and mean temperature using *MetPy* in Python (May et al., 2022). Humidex is a well-known index of perceived temperature, derived from temperature and dew point, widely used for heat surveillance in Canada (Smoyer-Tomic & Rainham, 2001). Pixel values of weather data were aggregated at the HR level by weighting each pixel with the number of units located in each pixel using the *AQgéobâti* database (Adresses Québec, 2023), resulting in population-weighted weather times series for each HR. *AQgéobâti* provides the geocoded information of each residential address in Quebec, along with the number of units (e.g., single-family home = 1 unit, duplex = 2 units, etc.). The number of units was used as a proxy for the population, as population data at the address level was not available.

Hourly air pollution data was extracted at the stations of the National Air Pollution Surveillance (NAPS) program of Environment and Climate Change Canada (ECCC). All stations in the province of Quebec were included, in addition to those within 10 km of Quebec's border to supplement the database (e.g., Ottawa stations, located ~5 km from HR #7, were considered to be within this HR). NAPS hourly data was aggregated to daily mean values. Then, a spatial aggregation of daily data from all stations within the HR boundaries was performed to get a unique time series of air pollutants for each HR, as commonly done in similar studies (Boudreault et al., 2024a; Lavigne et al., 2023; Masselot et al., 2019). Resulting air pollution time series were complete (with less than 2% of daily missing values) for ozone concentration (O_3) in HR #3, #4, #6, #7 and #16 from 1996 to 2019 and particulate matter ($PM_{2.5}$) for the 5 same HR, but only from 2003 to 2019. There were approximately 5 to 15 stations used for the spatial aggregation in each of these five HR with complete O_3 and $PM_{2.5}$ data (Figure 5.1). Other air pollutants (PM_{10} , NO_2 , SO_2 and CO) were not available in sufficient regions/years to be included. Remaining missing values in O_3 and $PM_{2.5}$ time series in these 5 HR were linearly interpolated. While including air pollution as a confounding effect in heat-mortality relationships is debatable (Buckley et al., 2014; Reid et al., 2012), its inclusion was considered as a sensitivity analysis to see if the heat-health relationships change when air pollution is added in the models (e.g., Gasparrini et al., 2015; Guo et al., 2014).

HR-level information was considered to explain the heterogeneity that can exist across regions in the multivariate meta-regression (presented in section 2.3), including the percentage of elderly population (65+), the percentage of women, the material and social deprivation index (MSDI), the natural logarithm of population density, the normalized difference vegetation index (NDVI) and the historical summer temperature (Lavigne et al., 2023; Wang et al., 2021). Demographic data

was provided by the *Ministère de la Santé et des Services Sociaux* (MSSS) du Québec from 1996 to 2019 (MSSS, 2022). The MSDI for its social and material components was available every census year (1996, 2001, 2006, 2011, 2016 and 2021) from INSPQ (INSPQ, 2024). NDVI values during the hottest summer day without cloud interference in each region (i.e., varying spatially) were computed from 2013 Landsat-8 data by the *Centre d'enseignement et de recherche en foresterie de Sainte-Foy* and provided by INSPQ (CERFO, 2022) and provided by INSPQ. A single year of NDVI value (i.e., 2013) was used for the whole study period, shown to be equivalent than using multiple yearly values in such studies (Pascal et al., 2021b). Finally, historical temperature from the past 20 years during summer (May to September) by HR was computed from Daymet mean temperature values.

5.2.3 Statistical analysis

A two-stage statistical analysis was performed. First, a quasi-Poisson model was fitted in each HR to assess the association between daily fluctuations in temperature and daily mortality or morbidity. The Distributed Lag Non-Linear Model (DLNM) was employed to account for both the non-linear relationship between mean temperature and the health outcome, as well as the delayed (lagged) temperature effect (Armstrong, 2006; Gasparrini et al., 2010). Models were fitted with a lag period of 8 days (i.e., lag 0 to lag 7) before the observed mortality or morbidity (Boudreault et al., 2023; Wang et al., 2021), using a natural cubic spline with 2 knots placed on the log scale from 0 to 7 days. The non-linear heat-health association was accounted by using a natural spline with two internal knots placed at the 50th and 90th percentiles of each HR's temperature distribution (Vicedo-Cabrera et al., 2021). Day of week and holidays were considered as factors in the model. Seasonal and long-term trends were also adjusted with a natural spline of day of year with 4 degrees of freedom by year and a factor variable for each year (Vicedo-Cabrera et al., 2021). Daily relative humidity up to 7 days prior to the observed mortality/morbidity was also controlled with a linear effect.

Second, the region-specific estimates obtained from the first step were pooled using multivariate meta-regression models (Gasparrini & Armstrong, 2013; Sera et al., 2019). This pooling procedure allowed to decrease uncertainty around region-specific heat-health associations by deriving Best Linear Unbiased Predictions (BLUP). In this study, the Risk Ratio (RR) of heat-related health outcomes were reported to the Minimum Mortality/Morbidity Temperature (MMT), the temperature at which the lowest health effect is observed. This was done by scanning temperature values from the 25th and 98th percentiles of temperatures in each HR (Vicedo-

Cabrera et al., 2021), excluding potentially colder temperatures from the MMT value (i.e., if the function is decreasing for colder temperature, the MMT will refer to the 25th percentile of temperature). Sensitivity analyses were performed to validate the obtained MMT values (see last paragraph of this section).

Meta-regression was fitted with a forward stepwise approach using health region information. Briefly, the meta-regression was first fitted with no predictors. Then, each of the regional predictors (described in section 2.2) was tested in the model, but only the predictor minimizing the Akaike Information Criteria (AIC) was added to the model. This process was repeated until the AIC did not decrease anymore. Heterogeneity in the meta-regression was assessed using the I^2 statistic and Cochran's Q-test (Lavigne et al., 2023). All analyses were performed using the R software (version 4.3.0) with the packages *dlm* and *mixmeta*.

Apart from the main analysis, supplementary analyses included: 1) changing the definition of summer season to the three hottest months of summer (June–August) instead of May–September; 2) considering the 5th or the 10th percentile of temperature as the lower bound for searching the MMT value (instead of the 25th percentile); 3) changing the temperature exposure variable to minimum temperature, maximum temperature, mean humidex and mean dew point, as recommended for temperature-related burden studies (Cheng et al., 2019); and 4) adding controls for daily mean concentration of O₃ and PM_{2.5} (up to 7 days prior to observed mortality/morbidity) in regional estimates for which air pollution data was available for potential changes in the obtained heat-health relationships.

5.2.4 Heat burden quantification

To quantify the heat burden, attributable fraction (AF) and attributable number (AN) were computed using the BLUP of the reduced exposure-response functions obtained from meta-regression for each HR (Gasparrini et al., 2015; Vicedo-Cabrera et al., 2021). AF was computed from $AF_t = (RR_t - 1)/RR_t$, where RR_t is the risk ratio at a given temperature value t . AN was computed from $AN_t = AF_t \times n_t$, where n_t is the mortality/morbidity count at temperature t (Steenland & Armstrong, 2006). AF_t and AN_t for all temperature values were aggregated across the desired heat range (e.g., all heat or extreme heat only). To account for the uncertainty in the BLUP, 95% confidence intervals (CI) of AF/AN were derived from 1000 simulations of AN/AF computed by sampling the coefficients of the exposure-response curve, assuming a multivariate normal distribution (Vicedo-Cabreba et al., 2021). Two temperature ranges were considered to quantify the heat burden: 1) *all heat*, which refers to all the temperature effect above the MMT, and 2)

extreme heat, which refers to the effect of extremely hot temperature only, defined as the temperatures above the 95th percentile of the summer temperature distribution by HR during the 1996–2019 period (Liu et al., 2019; Tupinier Martin et al., 2024; Wang et al., 2021).

To supplement results obtained using the regional BLUP, two other attribution methods were also considered. First, the complete bidimensional functions from regional DLNM were also used (referred to as the “*regional DLNM*” method). The *attrdl* function described in Gasparrini & Leone (2014) was employed and, again, 95% CI were computed from 1000 simulations of the coefficients. Second, the pooled cumulative exposure-response function over the whole province (i.e., all HR) obtained from the meta-regression was also used (referred to as the “*pooled estimate*” method). The AF/AN computation from the pooled estimate (and its uncertainty) was similar as the one described above for the regional BLUP.

5.3 Results

Descriptive statistics of health outcomes, weather and air pollution data are first presented (section 5.3.1). Then, results of the regional and pooled exposure-response curves are shown (section 5.3.2). Finally, the heat-related health burden is presented for the whole province of Quebec (section 5.3.3), as well as its temporal and spatial variability (section 5.3.4).

5.3.1 Descriptive statistics

During May to September, there were every year approximately 23k deaths, 263k hospitalizations, 1.5M ED visits, 265k ambulance transports and 662k calls to the 811-health hotline in the province of Quebec (Table 5.1). HR #6 (Montréal), #16 (Montérégie) and #3 (Capitale nationale) had the highest observed health outcomes during the study period, with respectively 20–26%, 13–16% and 8–11% of the grand total across the province. HR #8 (Abitibi-Témiscamingue), #9 (Côte-Nord) and #11 (Gaspésie — Îles-de-la-Madeleine) were the HR with the fewest cases of health outcomes (i.e., <3%). Indeed, these more rural HR had less population and hence, experienced less of the total mortality and morbidity. Information about socioeconomic variables by HR (i.e., regional information) can be found in Table S1.

Tableau 5.1 Descriptive statistics of the studied health outcomes for each health region (HR) and the total for the province of Quebec. All values are the total count during May to September period, averaged across all available years of data. The percentage in parentheses is the proportion over the total for the whole province.

HR	Mortality	Hospitalizations	ED visits	Amb. transports	811 calls
1	735 (3.2%)	8 264 (3.1%)	60 782 (3.9%)	7 580 (2.9%)	16 523 (2.5%)
2	883 (3.9%)	11 887 (4.5%)	73 260 (4.7%)	9 133 (3.4%)	26 460 (4%)
3	2 110 (9.3%)	22 104 (8.4%)	162 564 (10.5%)	26 359 (9.9%)	60 008 (9.1%)
4	1 673 (7.4%)	18 517 (7.1%)	110 657 (7.2%)	18 581 (7%)	51 836 (7.8%)
5	1 320 (5.8%)	16 188 (6.2%)	94 105 (6.1%)	15 023 (5.7%)	45 488 (6.9%)
6	5 871 (25.9%)	59 807 (22.8%)	312 656 (20.2%)	68 247 (25.7%)	144 264 (21.8%)
7	938 (4.1%)	9 139 (3.5%)	71 055 (4.6%)	10 505 (4%)	34 328 (5.2%)
8	380 (1.7%)	5 066 (1.9%)	45 297 (2.9%)	3 142 (1.2%)	10 080 (1.5%)
9	275 (1.2%)	4 545 (1.7%)	47 447 (3.1%)	3 247 (1.2%)	6 147 (0.9%)
11	334 (1.5%)	4 482 (1.7%)	45 051 (2.9%)	3 948 (1.5%)	5 955 (0.9%)
12	1 236 (5.4%)	15 089 (5.7%)	110 144 (7.1%)	13 158 (5%)	32 986 (5%)
13	994 (4.4%)	12 332 (4.7%)	53 714 (3.5%)	12 381 (4.7%)	33 709 (5.1%)
14	1 157 (5.1%)	14 931 (5.7%)	71 666 (4.6%)	15 211 (5.7%)	42 677 (6.4%)
15	1 426 (6.3%)	18 542 (7.1%)	90 935 (5.9%)	16 962 (6.4%)	47 059 (7.1%)
16	3 352 (14.8%)	41 665 (15.9%)	197 621 (12.8%)	41 987 (15.8%)	104 448 (15.8%)
Total	22 684	262 558	1 546 954	265 464	661 968

ED = Emergency department.

Mean temperature values during summer were higher in southernmost HR #6 (Montréal), #13 (Laval) and #16 (Montréal) with a mean value during summer of ~18°C, while northernmost HR #1 (Saguenay-Lac-Saint-Jean), #9 (Côte-Nord) and #11 (Gaspésie — Îles-de-la-Madeleine) were colder with mean summer temperature below 14°C (Table 5.2). Generally, maximum temperature was 4–6°C higher than mean temperature, while minimum temperature and mean dew point were 4–6°C lower. Mean humidex was approximately 5–6 units higher than mean temperature. Its standard deviation of 7–8 units was higher than for other temperature metrics that had a standard deviation of 4–5°C. Mean relative humidity during summer ranged from 65% to 75% (with a standard deviation of approximately 10%) for all HR. Finally, mean daily O₃ and PM_{2.5} concentration during summer (in HR where air pollution data was available) ranged respectively from 22 to 27 ppb and from 6 to 9 µg/m³.

Tableau 5.2 Descriptive statistics of weather and air pollution variables for each health region (HR). All values are the mean daily values during May to September from 1996 to 2019, except for PM_{2.5} that was only available from 2003 to 2019. O₃ and PM_{2.5} were only available for five HR. Standard deviation is indicated in parentheses.

HR	Tmean °C	Tmin °C	Tmax °C	Dew point °C	Humidex	Rel. Hum. %	Mean O ₃ ppb	Mean PM _{2.5} µg/m ³
1	13.8 (4.2)	8.6 (4.0)	18.9 (4.9)	8.5 (4.1)	18.0 (7.0)	71.0 (8.2)	-	-
2	14.8 (4.7)	8.9 (4.9)	20.8 (5.5)	8.5 (5.2)	19.2 (7.9)	66.6 (11.4)	-	-
3	16.0 (4.5)	10.4 (4.7)	21.6 (5.2)	10.3 (4.7)	21.5 (7.6)	69.4 (9.8)	22.4 (7.8)	7.7 (4.8)
4	16.7 (4.5)	11.1 (4.8)	22.4 (5.0)	10.9 (5.0)	22.7 (7.7)	69.1 (9.7)	25.1 (8.4)	7.8 (5.0)
5	16.6 (4.6)	10.9 (4.9)	22.4 (5.0)	10.8 (5.0)	22.5 (7.8)	68.9 (9.5)	-	-
6	18.2 (4.6)	12.9 (4.7)	23.5 (5.0)	12.7 (4.7)	25.4 (7.7)	70.9 (8.7)	23.6 (8.9)	9.0 (5.5)
7	17.3 (4.6)	11.3 (4.9)	23.4 (5.2)	11.0 (5.1)	23.4 (7.7)	67.3 (10.7)	25.5 (9.2)	6.2 (4.6)
8	14.3 (5.1)	8.0 (5.2)	20.6 (5.8)	7.4 (5.7)	18.0 (8.4)	64.4 (12.0)	-	-
9	12.3 (4.1)	7.7 (4.1)	16.9 (4.5)	7.6 (4.1)	15.9 (7.0)	73.6 (7.8)	-	-
11	13.8 (4.5)	9.1 (4.5)	18.4 (5.0)	9.1 (4.5)	18.4 (7.7)	73.6 (7.9)	-	-
12	15.8 (4.6)	10.0 (4.8)	21.6 (5.2)	10.0 (4.9)	21.1 (7.7)	68.8 (9.9)	-	-
13	18.2 (4.6)	12.9 (4.7)	23.5 (5.0)	12.7 (4.8)	25.4 (7.7)	70.9 (8.9)	-	-
14	17.4 (4.6)	11.7 (4.8)	23.2 (5.1)	11.5 (4.9)	23.8 (7.7)	69.0 (9.8)	-	-
15	17.1 (4.6)	11.3 (4.9)	22.9 (5.1)	11.2 (4.9)	23.3 (7.7)	68.9 (9.7)	-	-
16	18.1 (4.5)	12.5 (4.7)	23.6 (5.0)	12.4 (4.8)	25.0 (7.7)	69.9 (9.2)	26.7 (9.1)	7.8 (5.0)

5.3.2 Regional and pooled cumulative exposure-response curves

Estimated cumulative exposure-response curves over 8 days from DLNM for the 15 HR and 5 health outcomes are presented in Figure 5.2. Overall, all health outcomes appeared to increase with higher temperatures, but there was considerable variability between HR and across the different health outcomes. Notably, curves in light blue (i.e., HR #11) were the most variables with generally a different shape than the others. An important point is that some of these relationships were not significant for some parts of the curves (as noted with dashed lines on the figure), particularly for hospitalizations.

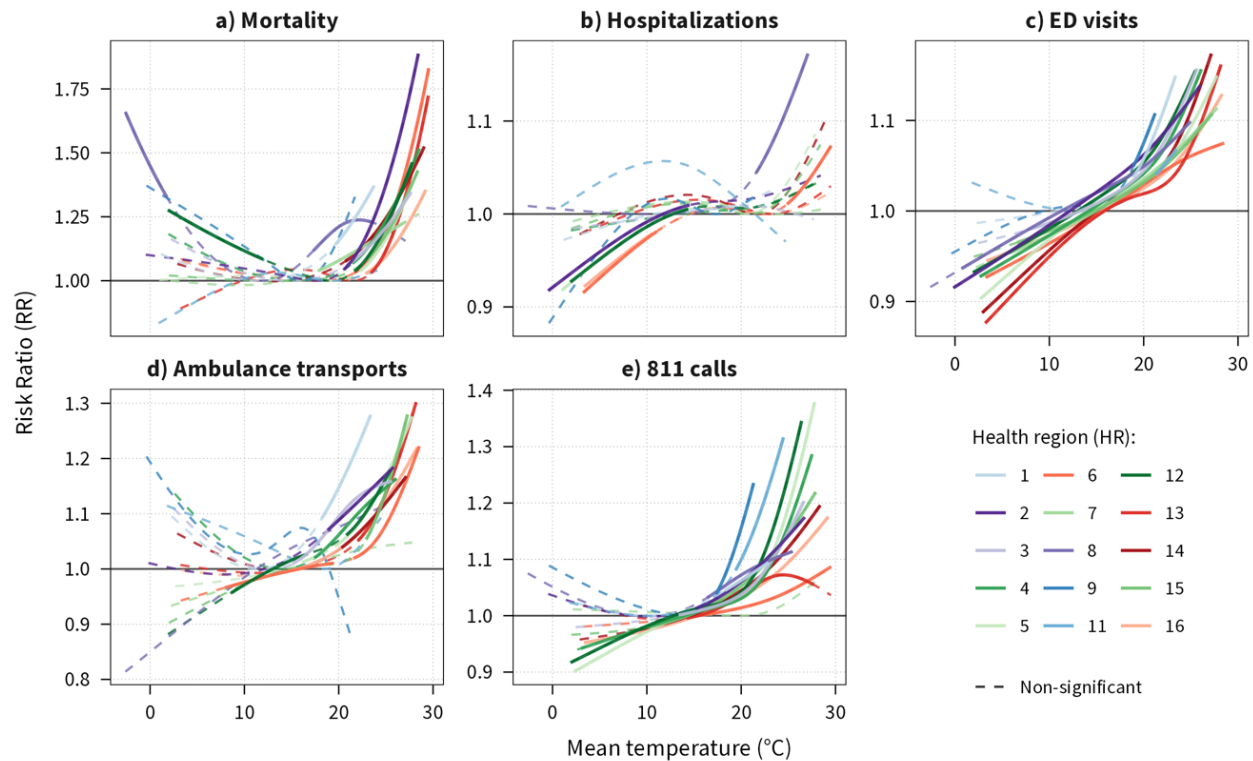


Figure 5.2 Cumulative exposure-response functions over 8 days in each health region (HR) from regional DLNM for a) mortality, b) hospitalizations, c) ED visits, d) ambulance transports and e) 811 calls. Dashed lines indicate non-significant parts of the curves, while solid lines indicate significant parts. Colours are set according to climate classification from red (warmer HRs) to green to purple to blue (colder HRs).

These regional estimates from DLNM were then combined in a multivariate meta-regression model using regional information, resulting in BLUP for each HR (Figure S1) and pooled cumulative exposure-response curves for the whole province of Quebec (Figure 5.3). Mortality exhibited a U-shaped relationship, with an MMT at around 15°C. The exposure-response curve for hospitalizations had a wiggly shape (~), with a potential protective effect below 7°C. However, a significant increase in hospitalizations was noted after 22°C. ED visits, ambulance transports and 811 calls increased consistently with higher temperatures, with different inflection points. The MMT for these three health outcomes was thus assigned to ~13°C (i.e., corresponding the 25th percentile of the summer temperature distribution, recall section 2.3).

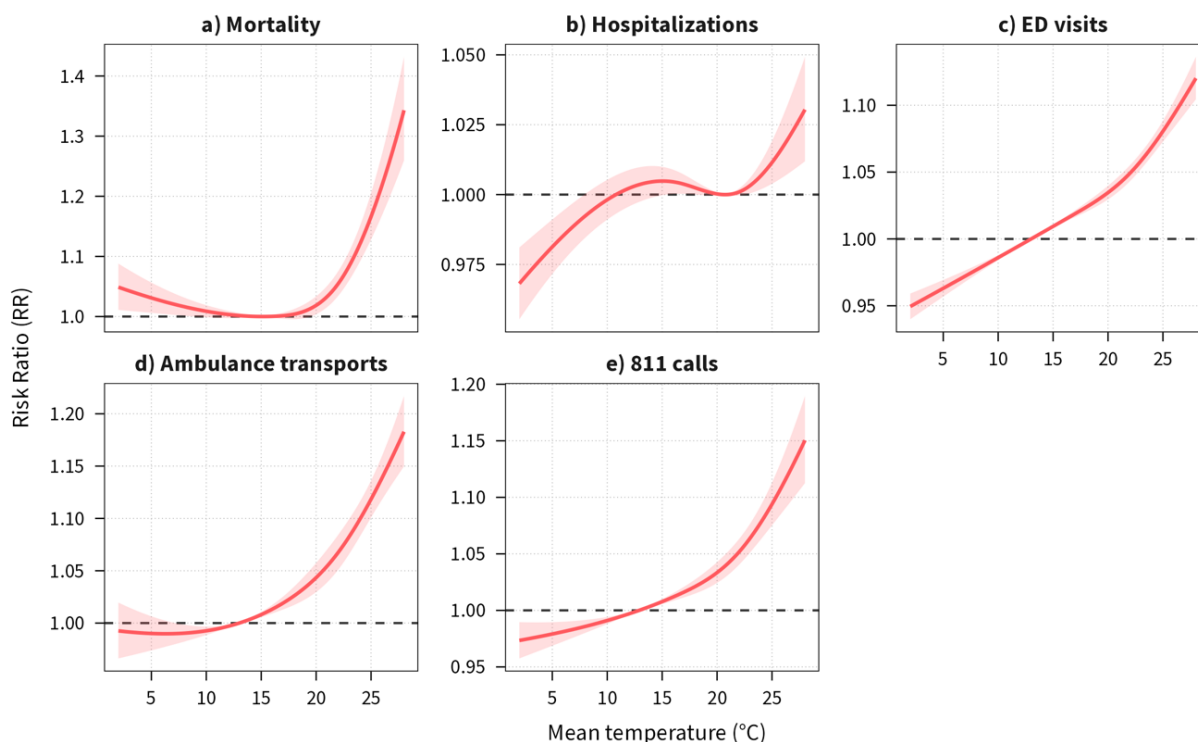


Figure 5.3 Pooled cumulative exposure-response functions over 8 days in the province of Quebec for a) mortality, b) hospitalizations, c) ED visits, d) ambulance transports and e) 811 calls. Light red region indicates the 95% confidence interval in the risk ratio.

As seen on Figure 5.2, there was some heterogeneity in the regional curves to develop these pooled effects, but also when looking at Q-Cochran test and I^2 statistics (Table S2). For example, the Q-Cochran test was only significant for 811 calls, while the I^2 statistic was $<5\%$ for hospitalizations, ED visits and ambulance transports. Thus, results from meta-regression should be interpreted with caution, especially for these three health outcomes. In addition, it was found that for all health outcomes, none of the meta-predictors significantly minimized the overall AIC in the meta-regression (Table S2). Hence, none of the meta-predictors was included in the model.

Supplementary analyses for the pooled exposure-response curves showed that considering summer to span only from June to August led to similar trends at high temperatures (Figure S2). For most of the studied health outcomes, the effect of colder temperature had wider CI. MMT values were consistent for mortality and hospitalizations, but slightly higher (i.e., $\sim 5^\circ\text{C}$ higher) for the other health outcomes (e.g., ED visits, ambulance transports and 811 calls) when using only June to August data. When testing lower percentiles for searching the MMT value (i.e., 5th and 10th), lower MMT values (i.e., $\sim 5^\circ\text{C}$ lower) were found for all health outcomes, except for mortality (Figure S3). Exposure-response curves with other temperature metrics showed similar shapes

than when using mean temperature (Figure S4). Finally, adjusting models for O₃ and PM_{2.5} (in HR for which air pollution data was available) did not change the shape of the pooled cumulative curve for most health outcomes, only slightly for hospitalizations when both O₃ and PM_{2.5} were controlled for (Figure S5).

5.3.3 Heat burden in the province of Quebec

AF and AN for all health outcomes derived from regional BLUP were computed across the whole province (Table 5.3). Supplementary analyses (e.g., with other temperature metrics, air pollution controls, etc.) were not pursued in this section as pooled exposure-response curves presented in section 3.2 did not differ significantly in these analyses. Approximately 2–3% of summer mortality, ED visits, ambulance transports and 811 calls were attributable to heat, but only 0.1% for hospitalization. Less than 1% of each health outcome was attributable to extreme heat (i.e., temperatures above the 95th percentile). AF/AN to heat and extreme heat were all statistically significant. In terms of AN, 471 and 203 deaths every summer were attributable respectively to all heat and extreme heat exposures, 226 and 170 hospitalizations, 36 273 and 6 228 ED visits, 7 177 and 1 517 ambulance transports, and 15 058 and 3 3306 calls to the 811-health hotline (Table 5.3). Differences between all heat and extreme heat exposures were greater for ED visits, ambulance transport and 811 calls (with factors of 4X–6X for all heat vs. extreme heat), while for deaths and hospitalizations, the difference was much smaller (1.5X–2.5X).

Tableau 5.3 Attributable number (AN) and fraction (AF) in the province of Quebec computed from regional BLUP for all heat and extreme heat exposures. . AN are presented as the yearly number of cases every summer across the studied period.

	All heat		Extreme heat	
	AN	AF	AN	AF
Mortality	471 (251 – 693)	2.1% (1.1–3.1)	203 (138 – 267)	0.9% (0.6–1.2)
Hospitalizations	226 (138 – 318)	0.1% (0.1–0.1)	170 (118 – 226)	0.1% (0.0–0.1)
ED visits	36 273 (32 328 – 40 060)	2.3% (2.1–2.6)	6 228 (5 589 – 6 870)	0.4% (0.4–0.4)
Ambulance transports	7 177 (4 594 – 9 592)	2.7% (1.7–3.6)	1 517 (1 326 – 1 700)	0.6% (0.5–0.6)
811 calls	15 058 (9 270 – 20 990)	2.3% (1.4–3.2)	3 306 (2 413 – 4 206)	0.5% (0.4–0.6)

ED = Emergency departement

Comparison of results using regional DLNM, regional BLUP and pooled estimates generally showed consistent values (within the same CI) for AN (Fig S6a) and AF (Fig S7a) for the total heat exposure. However, higher AN/AF for hospitalizations were found when using regional DLNM, but these values were not statistically significant anymore. For extreme heat exposure, AN (Figure S6b) and AF (Figure S7b) from pooled estimates yielded much lower values (1.5-2.0X smaller) than from regional DLNM and BLUP that were consistent. CI were generally wider with regional DLNM than with BLUP, except for mortality and 811 calls. Numerical values of AN/AF computed from the regional DLNM and pooled estimates are available in Table S3.

5.3.4 Temporal and spatial variability of the heat burden

AN were also computed separately for each year of observations (Figure 5.4 and Table S4 for numerical results of the last 10 years of data). Years with higher AN were respectively 2001, 2002, 2005, 2010 and 2018. For mortality, almost half the heat related AN was due to extreme heat (i.e., the 5% hottest days). This proportion was even higher for hospitalizations. For all health outcomes (except 811 calls), there were linear increasing trends in AN to total and extreme heat over the studied period. During summer of 2018 (during which all health outcomes data was available), approximately 480 deaths, 400 hospitalizations, 14 000 ED visits, 3 600 ambulance transports and 5 800 calls to the 811-health hotline were attributable to extreme heat (Table S4).

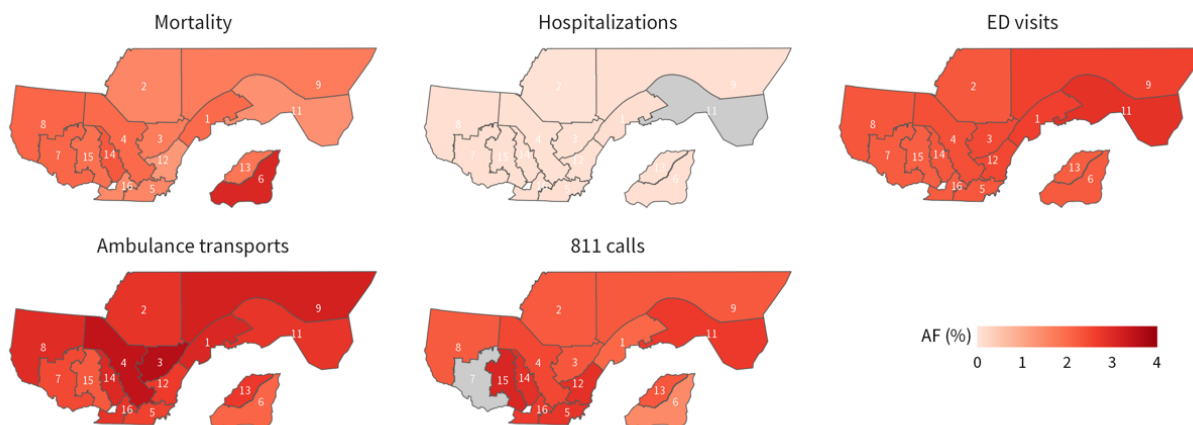
AF were also computed separately by HR (Figure 5.5 and Table S5 for numerical results of AF and AN). Significant AF due to all heat ranged from 1.5–3.0% for mortality, 0.1-0.1% for hospitalizations, 2.0–3.0% for ED visits, 2.0%–3.5% for ambulance transports and 1.5–3.0% for 811 calls (Figure 5.5). These AF were all <1.5% when looking only at extreme heat exposure. AF to all heat exposure in HR #11 for hospitalizations and in HR #7 for 811 calls was not significant. The example of the HR with the highest population (Montreal, #6) showed statistically significant AN of 180 (71) deaths, 53 (41) hospitalizations, 6 995 (1 171) ED visits, 1 415 (368) ambulance transports and 2 157 (446) calls to the 811 due to all heat (and extreme heat) every summer in average (Table S5).



Figure 5.4 Attributable numbers (AN) to all heat and extreme heat exposures during 1996–2019 based on regional BLUP across Quebec. Error bars represent the 95% CI. Dotted lines are linear trends over the studied period. Grey regions are years with major heatwaves. Data was not available for 1996–2013 for EDV and AMB, and 1996–2007 for 811 calls.

MOR = Mortality. HOS = Hospitalizations. EDV = Emergency department visits. AMB = Ambulance transports. 811 = Health hotline calls.

a) All heat



b) Extreme heat

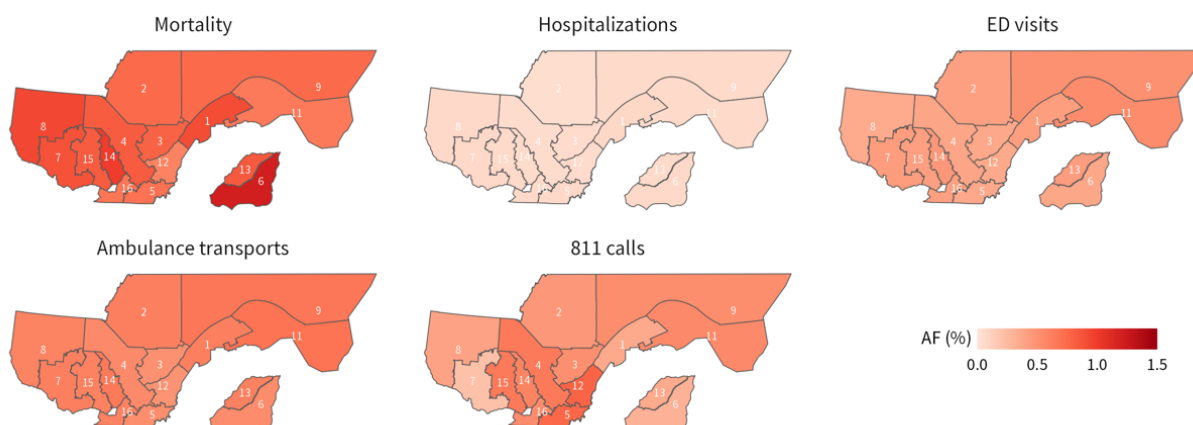


Figure 5.5 Attributable fraction (AF) in each health region (HR) computed from regional BLUP for a) all heat and b) extreme heat exposures. Non statistically significant AF are in grey. Refer to Table S5 for numeric results of AF (and AN).

ED = Emergency department.

5.4 Discussion

This study considered a two-stage statistical analysis with DLNM and multivariate meta-regression to estimate the heat burden for five health outcomes, including both mortality and morbidity, in the province of Quebec (Canada). While there was some notable variability and heterogeneity in the regional exposure-response functions (recall Figure 5.2 and Table S2), the pooled estimates showed that all studied health outcomes had higher RR at hotter temperatures, but at different levels. ED visits, ambulance transports and 811 calls did not exhibit the classical U- or J-shape relationship as observed with mortality, but rather monotonic increases in RR with higher temperatures. These less familiar relationships highlighted the need to correctly assign the MMT value, which will ultimately affect the calculation of the heat burden. MMT values were slightly higher when focussing on the three hottest months of summer (Figure S2), but lower when using lower percentiles of temperature (e.g., 5th or 10th) to extract MMT values (Figure S3). Thus, the reference MMT values used for these three health outcomes should be seen as conservative in our study. Further work is therefore recommended to that end for morbidity variables, as well as explaining the mechanisms between lower temperatures and these health impacts, that was out-of-scope for the current study focussing on heat. For mortality and hospitalizations, MMT values were consistent in the main and supplementary analyses.

Exposure-response functions obtained from BLUP were then used to compute the health burden associated with exposure to all heat (i.e., temperatures above the MMT) and extreme heat (i.e., temperatures above the 95th percentile) for the whole province, across years and regions. Results from BLUP were compared to regional DLNM (without meta-regression adjustment) and pooled estimates across the province. For the total heat exposure, the associated burden had consistent results for the three tested methods, except for hospitalizations when using regional DLNM (i.e., non-significant AN/AF). For extreme heat, both regional methods (BLUP and DLNM) had higher values (1.5-2X) than pooled estimates for the whole province, which could be due to a poorer representation of extreme temperatures in a unique pooled estimate for the whole province compared to regional estimates (recall Figure S6 and Figure S7). To the best of our knowledge, this study is one of the first to provide such a detailed portrait of the heat burden at a national scale combining multiple health outcomes, employing a robust and state-of-the-art methodology, as well as testing for different summer definitions, temperature exposures, air pollution adjustments and methods for the heat burden quantification.

In terms of mortality, the burden for the whole province was equivalent to AF of 2.1% of all summer mortality, representing AN of ~470 deaths each summer. When only the 5% hottest days of

summer were analyzed, AF and AN were respectively of 0.9% and ~200 deaths. In Montreal (Canada), Benmarhnia et al. (2014) found 62 deaths each summer attributable to heat during June to August months (years 1990–2007), while we found 180 deaths during the May–September period (years 1996–2019) for this location (recall Table S5). In Canada, Hebbert et al. (2023) found AF of 0.41% of annual mortality due to heat during the 2010–2019 period. In the multi-city multi-country study of Gasparrini et al (2015), the AF in Canada was 0.54% for all heat (0.68% in Montreal only) and 0.26% for extreme heat (above the 97.5th percentile), again reported over annual mortality. These last two results are consistent with ours, given that our AF were reported over summer mortality only.

During the heatwaves of 2010 and 2018 in Quebec, public health authorities reported respectively an excess of 280 (Bustinza et al., 2013) and 210 (Lebel et al., 2019) deaths across the province. We found that 462 and 480 deaths were associated to extreme heat during these two years, respectively (recall Figure 5.4 and Table S4). In addition, there was almost no significant impact on mortality during 2011–2015 official heatwaves in Quebec (Lebel et al., 2017), while we have found significant AN due to heat and extreme heat during all those years. These differences are mainly due to the definition of heatwaves used in the aforementioned studies (i.e., three consecutive days above thresholds), whereas extreme heat was defined as all days with temperatures above the 95th percentile in our study. Globally, the review of Cheng et al. (2019) reported that AF to heat were <2.0% for mortality (when only summer months were analyzed), which is consistent with our findings for the province of Quebec (i.e., AF of ~2.0%).

The AF for hospitalizations was found to be 0.1% (to both all heat and extreme heat exposures), representing AN of ~225 hospitalizations to heat and ~170 to extreme heat every summer across the province. In Ontario (Canada), Bai et al. (2016, 2018) found that 1.4% of hypertension, 11.2% of diabetes, 1.2% of coronary heart disease and 1.8% of stroke hospitalizations were due to heat. These AF are much higher than our results, but focussed on specific diseases that are known to be temperature-related (e.g., Liu et al., 2022; Phung et al., 2016). Indeed, focussing on all-cause hospitalizations could dilute the heat-hospitalization relationship, making the heat effect harder to detect. Studying cause-specific morbidity (or mortality) is left for future research. Furthermore, as our AF was computed over all hospitalizations, the obtained value will be lower than if it were calculated only for cause-specific hospitalizations as in Bai et al. (2016, 2018). Around the world, the study of Lin et al. (2012) found that ~100 hospitalizations for respiratory causes were attributable to extreme heat each year in the state of New York (~20M hab.). Wondmagegn et al. (2021a) found that ~500 hospitalizations were attributable to heat every year in Adelaide, an

Australian city of ~1.3M inhabitants. Finally, Liu et al. (2019) noted that 0.8% of hospitalizations in the 0–19 age group were related to moderate heat in the Minneapolis/St. Paul region (USA), while AF for other age groups were not statistically significant.

The heat burden on ED, ambulance services and health hotline was considerable in our study with AN of respectively 36 000, 7 000 and 15 000 cases due to heat every summer in Quebec, representing AF of 2.0–3.0%. During the 2010 heatwave in Quebec, Bustinza et al. (2013) noted an excess of 3 400 ED visits. While no ED data was available in 2010 in our study, we found ~14 000 ED visits associated with extreme heat in 2018, a comparable year to 2010 in terms of heatwave intensity. Kegel et al. (2021) studied the relationship between ED visits and summer temperatures in two hospitals of Montreal (Canada) and found that higher temperatures were associated with a higher number of patients in ED (same as in our study), but they did not report any AF nor AN. There was no other burden quantification study found in the literature for these morbidity variables in Quebec or Canada. Around the world, Wondmagegn et al. (2021a) found that ~290 ED visits were heatwave-related every year in Adelaide while Wellenius et al. (2017) estimated ~7 200 ED visits attributable to heat in 15 locations of New England, USA (2.7 M hab.). Cheng et al. (2016) reported AF of 2.2% for extreme heat-related ambulance transports in the city of Huainan (China) while Li et al. (2021) found AF of 11.7% to high temperatures for ambulance calls in 11 Chinese cities. In its review of temperature-related health burden, Cheng et al. (2019) reported that AF to heat was <1.4% for all-cause morbidity globally (though without differentiating by health outcomes), whereas higher figures were reported in our study (i.e., AF of 2–3% for ED visits, ambulance transports and calls to a health hotline).

Even though the above comparisons are interesting, they should be interpreted with caution, particularly for studies performed in other regions given the differences that exist in population size, demography, climate and health service management, among others (Gasparrini et al., 2015). In addition, in all studies cited above, there is not a single and universal temperature metrics used (e.g., mean, maximum, minimum temperature, heat index, etc.), no consistent definition of the total and extreme heat exposure (e.g., above different MMT values or percentiles such as 90th, 95th or 97.5th) and multiple heatwaves definitions (e.g., intensity, duration, temperature metrics), leading to even greater challenges in comparing the heat-related burden (Cheng et al., 2019). Thus, it highlights the need for more robust and comparable studies using a unified methodology that simultaneously quantify multiple health outcomes, as we have presented in this study and applied to Quebec, Canada.

By examining total and extreme heat exposures on a broader scaler than only during “official” heatwaves as was previously done in Quebec (e.g., Bustinza et al., 2013; Lebel et al., 2017; 2019), a new perspective on the heat load has been provided, which can be used to improve surveillance and protection activities by public health authorities. In addition, the variability of the heat burden between years (temporal) and health regions (spatial) was another insight that was revealed by our study. On the temporal side, years 2001, 2002, 2005, 2010 and 2018 were found to have the greatest health burden (recall Figure 5.4 and Table S4). The two last years (2010 and 2018) are also the ones during which recent major heatwaves occurred in Quebec. On the spatial side, the variability found in AF across health regions can be associated with different vulnerability factors such as demography, deprivation, built environment and adaptation to heat, among others (recall Figure 5.5 and Table S5). These results can therefore help prioritize the most effective actions to reduce the heat burden locally and in a timely manner. However, as meta-regressions did not detect any significant predictors (recall Table S2), future work is therefore recommended to explore the vulnerability factors associated with heat in Quebec at a finer scale, as well as to better understand the heterogeneity that was found between HR across the different studied health outcomes.

Our study has several strengths. First, it used a well-known and state-of-the-art framework combining DLNM and meta-regression to derive the heat-related health burden. Second, 5 health outcomes (including both mortality and morbidity) were included, most of them only seldom studied in Quebec and Canada, but also in the literature more generally. Third, our study was performed for the 15 southernmost health regions of Quebec (i.e., 99% of Quebec’s population of 8.5M inhabitants), a geographic scale representing a trade-off between statistical power and relevance for decision makers. Fourth, the computed AF and AN from regional BLUP were consistent compared to estimates from regional DLNM or pooled functions across the province, though some differences were also noted. Finally, supplementary analyses including a shorter definition of summer, different percentiles for the MMT value, four other temperature exposure variables and adjusting for air pollution, showed generally consistent results compared to the main analysis.

Some limitations must also be discussed. First, no stratification by sex/gender, age or cause of disease were performed. Our goal was to quantify the overall heat burden, not to assess the differences in heat-health relationships among population strata. This is left for further research, as well as the study of smaller geographical units that could be more suitable than the HR spatial scale for stratified analysis of some vulnerability factors such as built environment. Second, only

6 years of data was available for ED visits and ambulance transports (2014–2019), while longer time series (>10 years) were available for all other health outcomes. This limited the comparison of our results for ED visits and ambulance transports with the ones observed during the 2010 heatwave in Quebec. Third, the burden was only estimated for all heat (above the MMT) and extreme heat (above the 95th percentile) exposures, without studying, for example, the excess burden during heatwaves or the exposure to cold temperatures. Fourth, daily air pollution data was only fully available in 5 HR for the sensitivity analysis, while the low spatio-temporal resolution of air pollution data in Quebec prevented us from using population-weighted time series (as for weather data) in this analysis. Finally, while results were provided for other temperature exposure metrics, these analyses could be deepened in a future work.

5.5 Conclusion

By representing a state-of-the-art quantification of the heat burden, this study can serve as a reference for further studies in other parts of the world also experiencing the adverse effect of heat, drawing on the same detailed methodology and a comparable comprehensive database (i.e., several health outcomes in multiple regions combined with various temperature, air pollution and regional variables). Applied to the province of Quebec (Canada), all studied health outcomes had higher risk ratios at hotter temperatures, but at different levels. In terms of burden, heat translated into a massive load on the health system every summer in Quebec, with significant numbers of 470 (200) deaths, 225 (170) hospitalizations, 36 000 (6 200) ED visits, 7 200 (1 500) ambulance transports and 15 000 (3 300) calls to a health hotline due to heat (and extreme heat). Results also revealed a great deal of temporal (across years) and spatial (across health regions) variability. This new information will help public health authorities to target appropriate interventions depending on the health outcome, region and type of heat exposure, to reduce the heat-related health burden now and in the future.

Funding

The first author has received funding from the Natural Sciences and Engineering Research Council of Canada (Vanier Scholarship, #CGV-180821), the Canadian Institute of Health Research (Health System Impact Fellowship, #IF1-184093), Ouranos (Real-Décoste Excellence Scholarship, #RDX-317725) and the National Institute of Public Health of Quebec (no grant number).

Credit authorship contribution statement

Jérémie Boudreault: Conceptualization, Methodology, Data Curation, Formal analysis, Visualization, Software, Writing—Original Draft, Review and Editing, Funding acquisition. **Éric Lavigne:** Conceptualization, Methodology, Software, Writing—Review and Editing. **Céline Campagna:** Conceptualization, Writing—Review & Editing, Supervision, Project administration, Funding acquisition. **Fateh Chebana:** Conceptualization, Writing—Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Authors do not have permission to share health data. Socioenvironmental data are freely available from the following organization: NASA for weather data, Environment and Climate Change Canada (ECCC) for air pollution data, Ministère de la Santé et des Services Sociaux (MSSS) for demographic data and Institut national de la santé publique du Québec (INSPQ) for socioeconomic and built environment data. A code example to reproduce the conducted analyses with synthetic health data is available on the first author's Github page (https://github.com/jeremieboudreault/paper_heat_burden_qc).

Acknowledgments

The authors would like to thank Denis Hamel and Louis Rochette for their help with the health data extraction, Antoine Saint-Amand, Mathieu Tandonnet and Nathalie Gravel for their help with the spatial data extraction, Félix Lamothe, Magalie Canuel, Ray Bustinza and Annabel Ruf for their comments on this work, and Vaibhavi Mayya for her preliminary work on this project. The authors would also like to thank the editor, the associate editor and three anonymous reviewers for their comments that helped improve this manuscript.

Supplementary data

Supplementary data can be found online at: <https://doi.org/10.1016/j.envres.2024.119347>

CHAPITRE 6 :
PROJECTING THE OVERALL HEAT-RELATED HEALTH
BURDEN AND ASSOCIATED ECONOMIC COSTS IN A
CLIMATE CHANGE CONTEXT IN QUEBEC, CANADA

6 PROJECTING THE OVERALL HEAT-RELATED HEALTH BURDEN AND ASSOCIATED ECONOMIC COSTS IN A CLIMATE CHANGE CONTEXT IN QUEBEC, CANADA

Projection du fardeau sanitaire total lié à la chaleur et les coûts économiques associés dans un contexte de changements climatiques au Québec, Canada

Auteurs : Jérémie Boudreault¹², Céline Campagna¹²³, Éric Lavigne⁴⁵, Fateh Chebana¹

¹ Centre Eau Terre Environnement, Institut national de la recherche scientifique (INRS), 490 de la Couronne, Québec (QC) Canada, G1K 9A9.

² Direction de la santé environnementale, au travail et de la toxicologie, Institut national de santé publique du Québec (INSPQ), 945 avenue Wolfe, Québec (QC) Canada, G1V 5B3.

³ Département de médecine sociale et préventive, Faculté de médecine, Université Laval, 2300 rue de la Terrasse, Québec (QC) Canada, G1V 0A6.

⁴ Environmental Health Science and Research Bureau, Health Canada, 251 Sir Frederick Banting Driveway, Ottawa (ON) Canada, K1A 0K9

⁵ School of Epidemiology & Public Health, University of Ottawa, 600 Peter Morand Crescent, Ottawa (ON) Canada, K1G 5Z3

Titre de la revue : *Science of the Total Environment*

Publié le 1 janvier 2025

<https://doi.org/10.1016/j.scitotenv.2024.178022>

Contribution des auteurs :

J. Boudreault a été le principal contributeur de cet article, avec la conceptualisation de l'étude, la méthodologie, l'analyse des données, la modélisation, la production des figures et tableaux, et l'écriture du manuscrit. É. Lavigne, C. Campagna et F. Chebana ont aidé à conceptualiser l'étude et la méthodologie, ont validé les résultats, révisé l'article et approuvé sa version finale.

Lien entre cet article et les précédents :

Cet article est la suite du précédent (chapitre 5). Il réutilise les fonctions dose-réponse qui y ont été développées avec le modèle DLNM et la méta-régression, et les utilise pour projeter le fardeau sanitaire et les coûts de santé associés dans un contexte de changements climatiques et démographiques au Québec. Les issues de santé analysées sont les mêmes qu'aux deux derniers articles (chapitres 4 et 5), c'est-à-dire la mortalité et divers indicateurs de morbidité, toutes causes confondues, dans la population générale.

Abstract

Extreme heat represents a major health risk for the world's population, that is amplified by climate change. However, the health costs associated with these heat events have only been little studied. To stimulate the implementation of effective interventions against extreme heat, a more comprehensive economic valuation of these health impacts is crucial. In this study, a general framework for assessing historical and projected heat-related health costs is presented and then applied to the province of Quebec (Canada). First, heat-related mortality and morbidity, as well as the number of extreme heatwaves, were computed for a historical (~2000) and projected (~2050) period under two shared socioeconomic pathways (SSP). Then, these heat-related numbers were converted into 1) direct healthcare costs, 2) indirect productivity costs and 3) intangible societal costs, using the best available cost information. Results showed that historical heat-related health costs were respectively 15M\$, 5M\$ and 3.6G\$ (in 2019 Canadian dollars) annually for the direct, indirect and intangible components in Quebec, Canada. Under a middle-of-the-road scenario (SSP2–4.5), there was a 3-fold increase in total costs due to climate and population change (10.9G\$ annually), while under a pessimistic scenario (SSP5–8.5), the increase was 5-fold (17.4G\$). Total costs were mostly driven by intangible impacts, such as loss of life (~90–95%) and of well-being during heatwaves (~5–10%). Given that heat-related health costs are already significant, and likely to increase substantially in the future, this study has demonstrated the vital need to reduce its burden now and in the future by adopting more measures to mitigate climate change and adapt to heat.

Keywords: extreme heat, heatwaves, healthcare system, mortality, economic valuation, shared socioeconomic pathways.

Highlights

- Heat-related health costs computed for 1990–2019 and 2040–2069 in Quebec, Canada
- Considered costs were direct healthcare, indirect productivity and intangible costs
- Annual historical costs were 15M\$ (direct), 5M\$ (indirect) and 3.6G\$ (intangible)
- Total heat-related costs increased by 3X (SSP2–4.5) and 5X (SSP5–8.5) in 2050
- Total costs were mainly driven by loss of life (~90–95%) and of well-being (~5–10%)

Graphical abstract

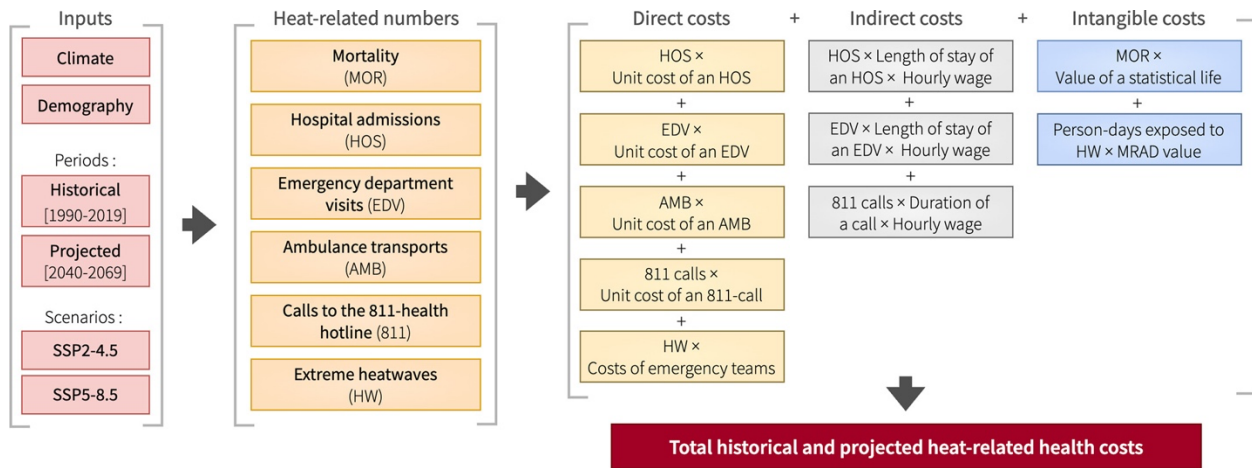


Figure 6.1 Graphical abstract

6.1 Introduction

Climate change is one of the greatest threats to human health in the 21st century (Haines et al., 2006; Watts et al., 2021). Among its numerous consequences, the increased frequency and intensity of extreme heat events have direct impacts on both mortality and morbidity (Campbell et al., 2018; Ebi et al., 2021; Song et al., 2017; Xu et al., 2016). Although many studies have projected the heat-related health burden under a changing climate (Cole et al., 2023; Huang et al., 2011; Sanderson et al., 2017), there is still only little research assessing the economic effects associated with these health impacts (Kousky, 2014; Schmitt et al., 2016; Wondmagegn et al., 2019). Estimating the costs of heat exposure is a key information that will help governmental and health authorities to better plan both mitigation and adaptation strategies to climate change (Austin et al., 2019; Huang et al., 2013).

The limited literature on heat-related health costs has some shortcomings that need to be acknowledged. On the one hand, many studies have only estimated the costs of historical extreme heat events of interest, leaving policymakers with no information on the future costs in the short-, medium- and long-term (Adélaïde et al., 2021; Beugin et al., 2023; Knowlton et al., 2011; Limaye et al., 2019; Wondmagegn et al., 2021a). On the other hand, most research projecting the health costs of heat exposure has analyzed a single health outcome such as mortality (Cheng et al., 2018; Díaz et al., 2019), hospital admissions (Crank et al., 2023; Hübler et al., 2008; Lin et al., 2012; Tong et al., 2021ab, 2022; Wondmagegn et al., 2021b), emergency department (ED) visits (Lay et al., 2018; Toloo et al., 2015; Tong et al., 2021c), or ambulance transports (Campbell et al., 2023), thus providing an incomplete and underestimated portrait of the overall expected heat-related health costs.

Economic valuation should include a wide variety of direct impacts including mortality and various morbidity outcomes, as well as indirect and intangible effects (Schmitt et al., 2016). In this study, we proposed a general framework for a comprehensive economic assessment of the historical and projected heat-related health costs. Given the difficulty of accessing reliable healthcare and cost data for such analysis (Wondmagegn et al., 2019), we applied the framework to Quebec, a Canadian province of 8.5M inhabitants (~22% of Canada) for which such data was readily available. In addition, we base our work on previously developed heat-health exposure-response functions (Boudreault et al., 2024c), an ensemble of downscaled and bias-corrected climate models from the latest Coupled Model Intercomparison Project 6 (CMIP6) (Lavoie et al., 2024a), and regional demographic projections (Hebbern et al., 2023).

Briefly, heat-related mortality, morbidity, as well as number of extreme heatwaves, were computed in a historical (1990–2019) and projected (2040–2069) period under two shared socioeconomic pathways (SSP). These heat-related numbers were then converted into 1) direct healthcare costs, 2) indirect productivity costs of absenteeism due to medical care, and 3) intangible societal costs for loss of life and well-being, using the best available economic data from official agencies. To our knowledge, this is the first study to provide such a comprehensive portrait of all heat-related health costs at a provincial level.

6.2 Material and methods

This study received ethics approval from the Human Research Ethics Committee of the National Institute of Scientific Research (CER-22-693). The graphical abstract provides an overview of the methodology.

6.2.1 Study area

This study took place in Quebec (Canada), a province of 8.5M inhabitants (in 2019), representing ~22% of the Canadian population. Quebec is divided into 18 health regions (HR), a geographical division of public healthcare services in Canada relevant for decision makers (Figure S1). The 3 northernmost HR (i.e., #10, #17, #18) were excluded from the current study due to their colder climate and small population (<1%) (Boudreault et al., 2024b). Due to provincial health jurisdictions, healthcare and cost data was not fully available in the other Canadian provinces, so this study only included Quebec. The summer period was defined as May to September, the months during which heat surveillance takes place in Quebec (Toutant et al., 2011).

6.2.2 Statistical analysis of the association between heat and health outcomes

Daily health data in each HR was provided by *Institut national de la santé publique du Québec* (INSPQ) for five all-cause health outcomes: mortality (1996–2019), hospital admissions (1996–2019), emergency department (ED) visits (2014–2019), ambulance transports (2014–2019) and calls to the 811-health hotline (2008–2019). Daily temperature data was obtained from Daymet at a 1 km x 1 km resolution (Thornton et al., 2022). Mean temperature was computed from the average of maximum and minimum daily temperatures, then aggregated in each HR by weighting temperature values with the number of residential units in each grid cell using geocoded addresses (Boudreault et al., 2024b). Distributed Lag Non-Linear Models (DLNM) were developed to capture the non-linear and lagged association over 8 days between mean temperature and

every health outcome during summer for each HR (Gasparrini et al., 2010). The non-linear heat-health association was accounted by using a natural cubic spline with two internal knots placed at the 50th and 90th percentiles of each region's temperature distribution and two degrees of freedom (Vicedo-Cabrera et al., 2021). The lag effect was modelled with a natural cubic spline with two knots placed on the log scale from 0 to 7 days and two degrees of freedom. Then, meta-regression models were fitted on the reduced cumulative effect over 8 days to obtain Best Linear Unbiased Predictions (BLUP) of each heat-health associations in all HR (Gasparrini & Armstrong, 2013; Sera et al., 2019). Risk ratios were computed relative to the minimum mortality/morbidity temperature (MMT) value, obtained by scanning the temperature values from the 25th to the 98th percentile of each HR temperature distribution (Vicedo-Cabrera et al., 2021). The readers are referred to Boudreault et al. (2024c) for additional details on the fitting of DLNM and meta-regression, sensitivity analyses and resulting graphs of exposure-response functions.

6.2.3 Climate and demographic data

The EPSO-G6-R2 ensemble of 15 downscaled and bias-corrected global circulation models (GCM) of CMIP6, developed by the Ouranos consortium on regional climatology, was employed (Lavoie et al., 2024a). Two shared socioeconomic pathways (SSP) were considered: SSP2–4.5, a middle-of-the-road scenario, and SSP5–8.5, a pessimistic scenario in which emissions continue to rise through the century (IPCC, 2021). Daily mean temperature during 1990–2069 in each HR was computed from the average of daily minimum and maximum temperatures, then bias-corrected using the Hempel method to adjust remaining bias between population-weighted temperature data (used to develop exposure-response functions) and climate models values (Hempel et al., 2013; Vicedo-Cabrera et al., 2021). Demographic data by HR was provided by *Ministère de la santé et des services sociaux du Québec* (MSSS) for the 1996–2019 period (MSSS, 2022). Data for 1990–1995 was linearly extrapolated from the HR to Quebec ratio of population trend during the 1996–2000 period. Projected population by HR from 2020 to 2069 was provided by Statistics Canada for two demographic scenarios associated with SSP2 and SSP5, previously used for forecasting future temperature-related mortality burden in Canada (Hebborn et al., 2023). These scenarios were in line with Quebec's population forecasts for the 2020–2041 period (MSSS, 2022).

6.2.4 Attributable numbers to heat and number of extreme heatwaves

Climate and demographic data were then used with BLUP of exposure-response functions in each HR to compute attributable numbers (AN) for the historical (1990–2019) and projected (2040–2069) periods. AN to all heat and extreme heat were computed respectively by summing the individual contributions of all temperatures above the MMT (i.e., *all heat*) and above the 95th percentile of historical summer temperature by HR (i.e., *extreme heat*) (Boudreault et al., 2024c; Gasparrini & Leone, 2014; Steenland & Armstrong, 2006). The mean historical daily count of each health outcome was used to compute AN for both periods (Gasparrini et al., 2017; Vicedo-Cabrera et al., 2019). To supplement AN to all and extreme heat, number of extreme heatwaves (HW) at each reference weather station's location of HR were also computed (Figure S1). HW were defined on the basis of official public health thresholds of 3-day weighted average values of maximum and minimum temperatures, respectively 33° and 20°C for HR #6, #13–#16, 31°C and 18°C for HR #2–#5, #7–#8 and #12, 31° and 16° for HR #1, #9 and #11 (Chebana et al., 2013; Lamothe et al., 2023). The person-days exposed to HW were computed by multiplying the duration of each heatwave by the population in the HR, in order to assess the loss of well-being due to extreme heat events (Adélaïde et al., 2021). To isolate the specific effect of climate and demographic changes, projections were made separately for 1) historical climate and demography (1990–2019), 2) projected climate (2040–2069), but with historical demography, and 3) both projected climate and demography in 2040–2069, as recommended in a recent review (Sanderson et al., 2017).

6.2.5 Heat-related health costs

Heat-related numbers were converted to health costs using the cost-of-illness approach and then summarized in three categories: 1) direct healthcare costs, 2) indirect productivity loss to seek medical care, and 3) intangible societal costs (Schmitt et al., 2016; Wondmagegn et al., 2019). To obtain comparable figures for the historical and projected periods, all costs were expressed in 2019 Canadian dollars (\$), without considering past healthcare costs or making assumptions on future healthcare expenditures (Lay et al., 2018; Limaye et al., 2019). When cost metrics were not in \$2019, they were converted using the consumer price index of the Bank of Canada (Bank of Canada, 2024). Direct costs were computed by multiplying AN to all heat (or extreme heat) with the unit cost (UC) of each health outcome (Knowlton et al., 2011; Limaye et al., 2019). In addition, the costs of emergency teams when HW thresholds are met were also considered (Larrivée et al., 2015). Indirect costs were computed by converting AN to the total duration using Length of

Stay (LoS), then multiplied by the mean hourly wage in Quebec (Knowlton et al., 2011; Limaye et al., 2019). Finally, intangible costs were estimated using the Value of a Statistical Life (VSL) multiplied by the AN of heat-related mortality (Limaye et al., 2019; Liu et al., 2019), and loss of well-being using a Minor Restricted Activity Day (MRAD) willingness-to-pay for each person-day exposed to HW (Adélaïde et al., 2021). All economic metrics used are listed in Table 6.3 in the appendix 6.A at the end of the article, along with the relevant references.

A Monte Carlo simulation was performed to assess the uncertainty in estimated costs. The simulation algorithm consisted of five steps: 1) selecting randomly 30 years in either the historical or projected period across all simulated years of the 15 GCMs (i.e., 450 years); 2) deriving AN during these years by sampling random coefficients of BLUP for each health outcome, assuming a multivariate-normal distribution of these coefficients as commonly done to assess the uncertainty of exposure-response functions for climate projections (e.g., Gasparrini et al., 2017; Vicedo-Cabrera et al., 2021); 3) computing the number of HW as well as the population exposed to HW during these 30 years; 4) simulating UC, LoS, VSL, MRAD and hourly wage based on uncertainties and distributions provided in Table 6.3 of appendix 6.A; and 5) aggregating the direct, indirect, intangible and total costs of all heat and extreme heat as an average annual cost over these 30 years. For the economic metrics with no uncertainties provided, a triangular distribution with $\pm 33\%$ around the central value was assumed (Adélaïde et al., 2021). This process was replicated 15 000 times for each period (historical and projected) and combination of climate and demographic scenarios (i.e., SSP2–4.5 and SSP5–8.5). Empirical 95% confidence intervals (eCI) were then computed based on these simulations. Dunn’s tests were performed to assess a statistically significance difference between the simulated costs in the different periods and scenarios, adjusted with the Bonferroni method for multiple comparisons.

6.3 Results

6.3.1 Climate and demographic evolution

The average summer temperature during May to September months in Quebec (Canada) was approximately 15.5°C during the historical period. Based on an ensemble of bias-corrected and downscaled climate models from CMIP6, mean temperature was projected to increase to ~18°C (+2.5°C) under SSP2–4.5 and to ~19°C (+3.5°C) under SSP5–8.5 over the 2040–2069 period (Figure 6.2a). Overall, the ensemble of climate models well represented the observed temperatures in Quebec (black line) during the historical period. In terms of demography, Quebec’s population grew from 7M inhabitants in 1990 to 8.5M in 2020 (Figure 6.2b). Under

SSP2, projections showed that it could reach 9.5M in 2050 and 10.3M in 2070. Under SSP5, for which a higher demographic growth is expected, Quebec's population could reach 10.5M in 2050 and 13M in 2070.

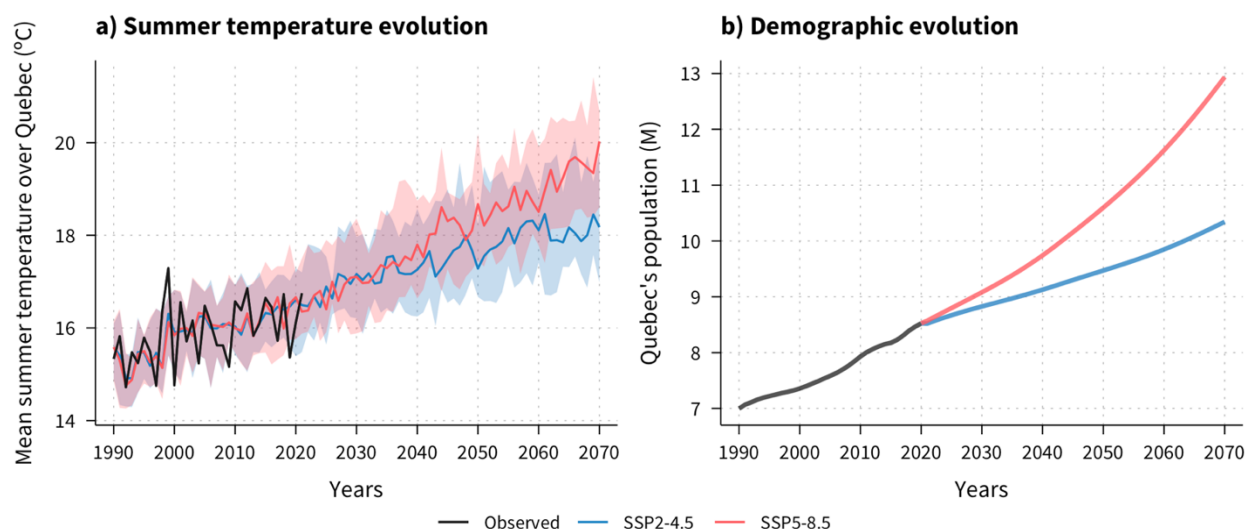


Figure 6.2 Evolution of a) mean summer temperature (May to September) and b) demography across Quebec based on SSP2-4.5 and SSP5-8.5 scenarios over the 1990–2070 period. The ribbon corresponds to the 10th and 90th percentile of the climate models ensemble.

6.3.2 Heat-related numbers

During the historical period (1990–2019), heat (i.e., all temperatures above MMT) resulted in AN of roughly 410 deaths, 200 hospital admissions, 31 900 ED visits, 6 100 ambulance transports and 13 300 calls to the 811-health hotline every summer in Quebec (Table 6.1). These AN were lower when considering only extreme heat (i.e., temperatures above the 95th percentile of historical summer temperature) with AN of approximately 160 deaths, 150 hospital admissions, 5 500 ED visits, 1 300 ambulance transports, 2 600 calls to the 811-health hotline. There were roughly 3.1 HW and 6.9M person-days exposed to HW every summer in Quebec during that period. Under SSP2-4.5, mortality increased by 118%, hospital admissions by 230%, and the other morbidity outcomes by 50–65% in the projected period (i.e., 2040–2069) compared to the historical period, only because of climate change (i.e., without considering any other socioeconomic change or adaptation). In addition, there was roughly 4X more HW and person-days exposed to these HW in the projected period. When also considering demographic change based on SSP, there was a 3-fold increase in heat-related mortality, 4-fold increase in hospital admissions, 2-fold increase in ED visits, ambulance transports and 811 calls, and 5-fold increase

in person-days exposed to HW. The increases in AN to extreme heat were even greater: 230–300% without considering demographic change, and 280–400% when considered. Under SSP5–8.5 (Table S1), increases in AN reached up to 7-fold for all heat and 9-fold for extreme heat.

Tableau 6.1 **Attributable numbers (AN) to all and extreme heat every summer in Quebec for mortality (MOR), hospital admissions (HOS), ED visits (EDV), ambulance transports (AMB) and calls to the 811-health hotline (811), as well as number of extreme heatwaves (HW) across the 15 health regions of Quebec and person-days exposed to HW during the historical (1990–2019) and projected (2040–2069) periods under SSP2–4.5.** The median of all climate models is presented. Refer to Table S1 for results under SSP5–8.5.

	Climate	Demography	All heat		Extreme heat	
			Number	Increase (%)	Number	Increase (%)
AN of MOR	Historical	Historical	414	Ref.	159	Ref.
	Projected	Historical	900	118%	548	246%
	Projected	Projected	1 140	176%	702	343%
AN of HOS	Historical	Historical	198	Ref.	146	Ref.
	Projected	Historical	654	230%	576	295%
	Projected	Projected	822	315%	728	399%
AN of EDV	Historical	Historical	31 895	Ref.	5 475	Ref.
	Projected	Historical	48 880	53%	17 997	229%
	Projected	Projected	56 972	79%	20 800	280%
AN of AMB	Historical	Historical	6 132	Ref.	1 256	Ref.
	Projected	Historical	10 029	64%	4 118	228%
	Projected	Projected	12 322	101%	5 144	310%
AN of 811	Historical	Historical	13 305	Ref.	2 631	Ref.
	Projected	Historical	21 413	61%	8 617	228%
	Projected	Projected	26 393	98%	10 465	298%
Number of HW	Historical	Historical	3.1	Ref.	Same as for all heat.	
	Projected	Historical	13.2	322%		
	Projected	Projected	13.2	322%		
Person-days (M) exposed to HW	Historical	Historical	6.9	Ref.	Same as for all heat.	
	Projected	Historical	27.4	299%		
	Projected	Projected	34.5	402%		

6.3.3 Health costs of heat exposure

The historical annual health costs attributable to all heat were estimated to be respectively 14.8M\$, 5.3M\$ and 3.6G\$ (in 2019 Canadian dollars) for the direct, indirect and intangible components, of which 5.3M\$, 1.5M\$ and 1.5G\$ were respectively attributable to extreme heat only (Table 6.2). Under SSP2–4.5, all costs doubled in 2050 compared to the historical period with increases of 85% to 135%, only because of climate change (i.e., keeping demography fixed as of 1990–2019). Higher increases of 250–290% were noted for health costs attributable to extreme heat. Demographic growth added an extra load on the health burden with increases of 120% to 200% for all heat, and of 320% to 400% for extreme heat. Under SSP5–8.5 (Table S2), health costs of all heat increased from 3.7G\$ in the historical period to 11.7G\$ in the projected (2040–2069) period only because of climate change, and to 17.4G\$ when considered both

evolving climate and demographic. All costs categories were 3X to 4X higher compared to the historical period. The total health costs attributable to extreme heat in 2050 was multiplied by more than 8X with both climate and demographic changes under SSP5–8.5, for a grand total of 12.6G\$. Roughly, SSP5–8.5 doubled the heat-related health costs in 2050 compared to SSP2–4.5.

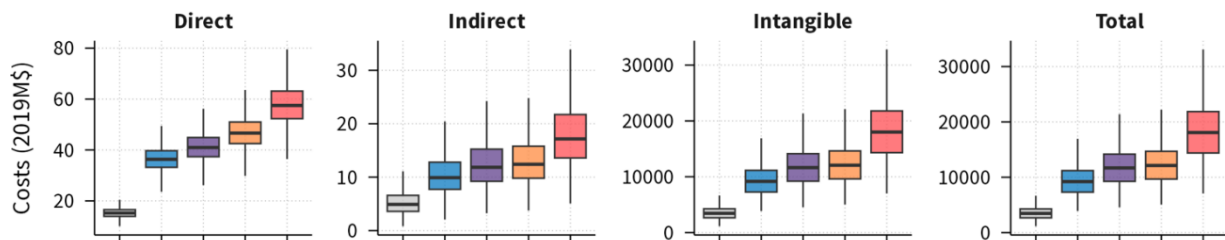
Tableau 6.2 **Yearly direct, indirect, intangible and total health costs (in Canadian M\$2019) attributable to all and extreme heat in the historical (1990–2019) and projected (2040–2069) periods under SSP2–4.5.** The median of all climate models is presented. Empirical 95% confidence intervals (eCI) based on Monte Carlo simulations are shown in parentheses. Refer to Table S2 for results under SSP5–8.5.

	Climate	Demography	All heat Cost in M\$2019 (95%eCI)	Increase %	Extreme heat Cost in M\$2019 (95%eCI)	Increase %
Direct healthcare costs	Historical	Historical	14.8 (11.6–19.2)	Ref.	5.3 (3.5–7.7)	Ref.
	Projected	Historical	32.1 (27.9–46.6)	116%	20.2 (17.5–34.0)	279%
	Projected	Projected	36.4 (31.2–52.9)	145%	22.3 (19.1–37.7)	319%
Indirect productivity costs	Historical	Historical	5.3 (2.0–11.1)	Ref.	1.5 (0.7–2.7)	Ref.
	Projected	Historical	9.8 (4.8–20.3)	84%	5.2 (3.2–10.7)	254%
	Projected	Projected	11.6 (5.9–24.0)	119%	6.3 (3.9–12.8)	331%
Intangible societal costs	Historical	Historical	3 589 (1 815–5 505)	Ref.	1 466 (744–2 372)	Ref.
	Projected	Historical	8 432 (5 076–14 115)	135%	5 687 (3 460–9 944)	288%
	Projected	Projected	10 837 (6 483–17 823)	202%	7 257 (4 388–12 529)	395%
Total costs	Historical	Historical	3 609 (1 836–5 527)	Ref.	1 473 (749–2 380)	Ref.
	Projected	Historical	8 474 (5 120–14 166)	135%	5 712 (3 485–9 981)	288%
	Projected	Projected	10 885 (6 538–17 878)	202%	7 286 (4 418–12 569)	395%

6.3.4 Uncertainties in estimated costs

Monte Carlo simulations demonstrated the uncertainties in the historical and projected health costs arising from the epidemiological models, economic metrics and climate projections, as illustrated by the heights of the boxplots in Figure 6.3. Uncertainty was smaller during the historical period, for which climate conditions are known, while climate scenarios added additional uncertainty during the projected period. This was particularly marked under the high-emission high-population scenario (i.e., SSP5–8.5) (Figure 6.3). These Monte Carlo simulations also allowed to derive 95% empirical confidence intervals (eCI) of estimated costs (Table 6.2 for SSP2–4.5 and Table S2 for SSP5–8.5). For example, the eCI of the total health costs of all heat was 1.8–5.3G\$ during the historical period, increasing during the projected period to 6.5–17.9G\$ under SSP2–4.5 and 10.1–27.8G\$ under SSP5–8.5, when considering both climate and demographic changes. Finally, the differences between all periods/scenarios were all statistically significant at a 5% level for all cost categories based on the Dunn’s test (Figure S2).

a) All heat



b) Extreme heat

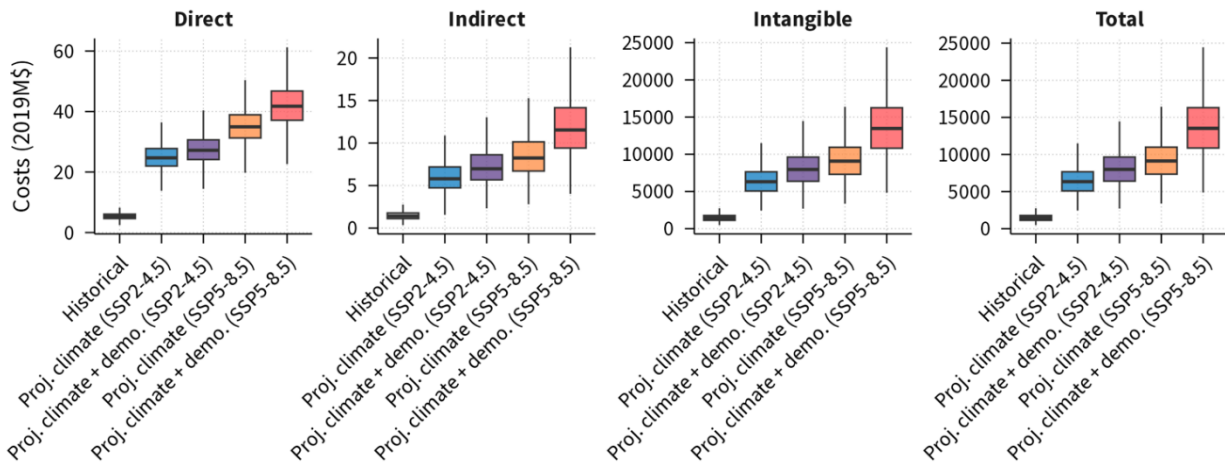


Figure 6.3 Results of Monte Carlo simulations for historical (1990–2019) and projected (2040–2069) yearly heat-related health costs for direct, indirect, intangible and total components for a) all heat and b) extreme heat. Refer to Table 6.2 (for SSP2–4.5) and Table S2 (SSP5–8.5) for numeric results of the 95% empirical confidence intervals.

6.3.5 Breakdown of health costs

Historical direct healthcare costs of all heat were mostly driven by ED visits, followed by hospital admissions, ambulance transports and emergency teams during HW (Figure 6.4). In all projected scenarios (i.e., SSP2–4.5 and SSP5–8.5), the share of emergency teams doubled from ~15% to ~30%, becoming the second most important direct cost. This can be linked to the higher number of heatwaves and more frequent deployments of emergency teams during the projected period. For the indirect costs, the share was ~75% for ED visits loss of productivity, followed by hospital admissions with ~25%. A higher share of loss of productivity due to hospital admissions was noted in the projected period (i.e., 2040–2069). For the intangible component, mortality costs were more important than loss of well-being in the historical period, but the share of well-being loss increased in 2050 with demographic growth, resulting in more people negatively affected by HW. Finally, the total costs were mostly driven by intangible loss of life (~90–95%) and of well-being (~5–10%)

costs, and less than 1% for other costs. For extreme heat (Figure S3), the portrait was a bit different with almost half the direct healthcare costs related to emergency teams, followed by ED visits and hospital admissions, in both historical and projected periods. The productivity costs were shared roughly by wage loss for ED visits (~50%) and hospital admissions (~50%). For intangible and total health costs, the same portrait as for all heat was noted. Refer to Table S3 (SSP2–4.5) and Table S4 (SSP5–8.5) for detailed heat-related health costs by outcome and their share over their category and the grand total.

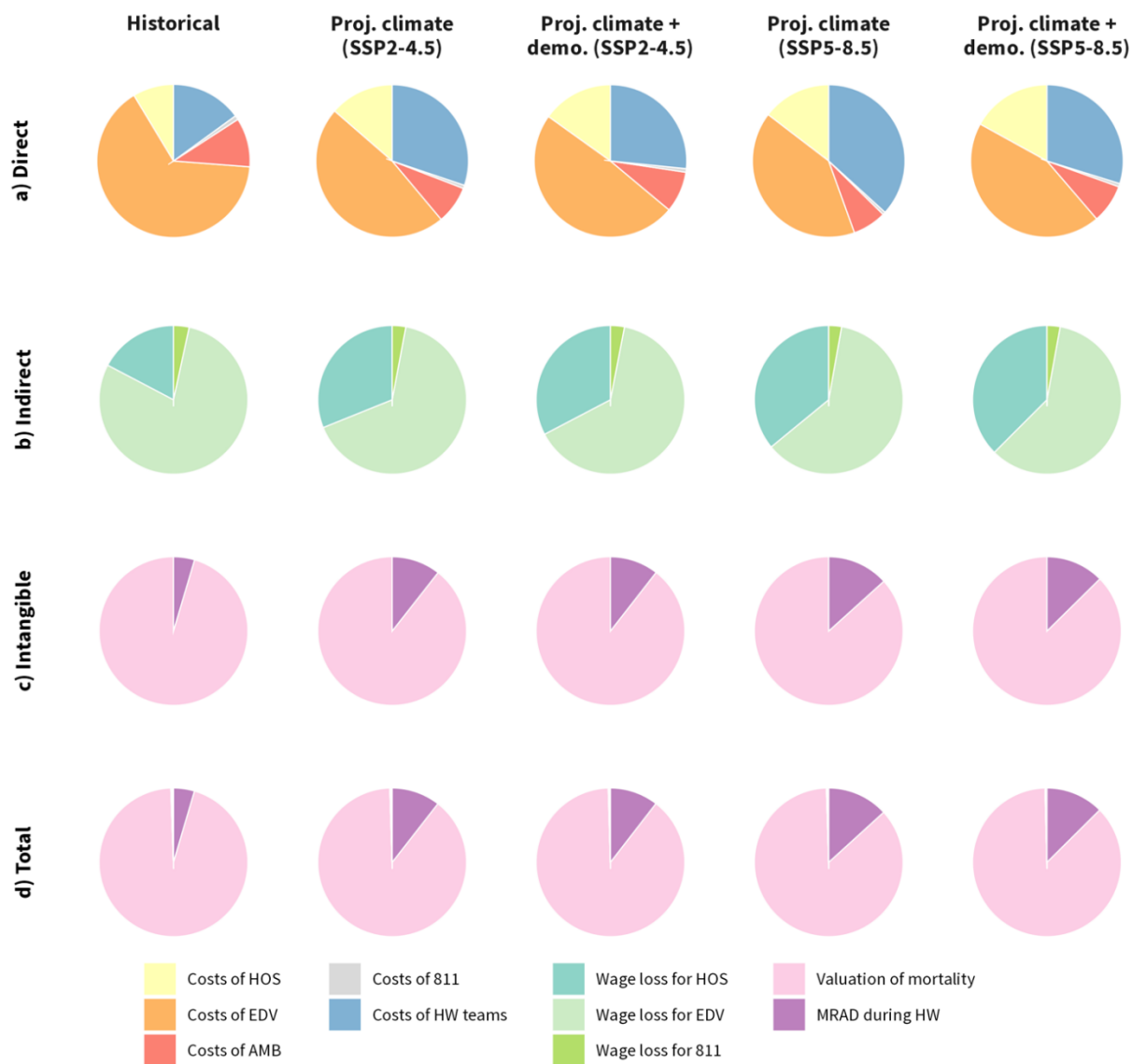


Figure 6.4 Health costs breakdown of all heat by a) direct, b) indirect, c) intangible and d) total costs during the historical (1990–2019) and projected (2040–2069) periods based on SSP2–4.5 and SSP5–8.5 climate and demographic projections. Refer to Table S3 (for SSP2–4.5) and Table S4 (SSP5–8.5) for numerical results.

AMB = Ambulance transports. EDV = Emergency department visits. HOS = Hospital admissions. HW = Extreme heatwaves. MRAD = Minor restricted activity day. 811 = Calls to the 811-health hotline.

6.4 Discussion

This study is among the first to evaluate the historical and projected heat-related health costs arising from multiple mortality and morbidity outcomes. Applied to the province of Quebec, Canada, total health costs attributable to all heat (i.e., temperatures above the MMT) were estimated at 3.6G\$ annually during the 1990–2019 period. Considering both climate and demographic changes, these costs were projected to increase to 10.9G\$ annually under SSP2–4.5 (a 3-fold increase) and to 17.4G\$ under SSP5–8.5 (a 5-fold increase) during the projected 2040–2069 period. Costs attributable to extreme heat only (i.e., temperatures above the 95th percentile of historical summer temperature) had even greater increases in the projected period: 5-fold under SSP2–4.5 and 8.5-fold under SSP5–8.5. As in other international studies (Gasparrini et al., 2017; Vicedo-Cabrera et al., 2018), these projected numbers represent a “hypothetical future”, as other changes will also occur such as biological acclimatization to heat and socioeconomic changes that may reduce the heat effect (Hondula et al., 2015). However, because hypotheses on adaptation are often simplistic and would add another layer of uncertainty to these projections (Sanderson et al., 2017), their inclusion was deemed out-of-scope for this initial attempt of assessing historical and projected heat-related health costs.

Our study reaffirmed the importance of mitigation strategies aimed at limiting greenhouse gases (GHG) emissions to further reduce the projected heat-related health burden (Romanello et al., 2023; Vicedo-Cabrera et al., 2018). Indeed, projected costs were approximately 2 times higher under SSP5–8.5 than SSP2–4.5. However, even under the more optimistic scenario, the heat-related health costs still significantly increased by 2050 and were amplified by a larger population exposed to heat. Therefore, this research also highlighted the need for better adaptation measures to heat such as increased greening, reduced heat urban islands or new alert systems, among others, to reduce the historical heat-related costs and the increasing costs in the future (Ebi et al., 2018; Huang et al., 2013). In addition, our results can be used to carry out cost-benefit analyses of such measures to stimulate their adoption, as already demonstrated for heat-health warning systems (e.g., Carmona et al., 2016; Chiabai et al., 2018; Ebi et al., 2004; Hunt et al., 2017; Williams et al., 2021).

Over the last decades, there have been a few reports in Canada that estimated some components of the heat-related health costs. NRTEE (2011) calculated heat-related mortality costs in Montreal of 0.6–0.8G\$ in 2020 and 1.4–2.8G\$ in 2050, which are comparable to our results of 3.4G\$ in 2000 and 7.4G\$ in 2050 for the whole province, given that Montreal is one fifth of Quebec’s population. Larrivée et al. (2015) estimated the health costs associated with climate changed-

induced heat at 32.4G\$ (for mortality) and 372M\$ (for morbidity) during the 2015–2064 period in Quebec using a discount rate of 4%. Performing a similar discounting (results not shown), our costs for the 2015–2064 period would be ~50G\$ for mortality and ~300M\$ for healthcare costs, which is slightly higher for mortality but comparable for morbidity (although the morbidity outcomes studied were different in the two studies). Clark et al. (2021) estimated heat-related costs at 3.0–3.9G\$ each year in 2050 for mortality and 138–167M\$ for morbidity in Canada. While our climate change-related morbidity costs of 23M\$ were comparable (assuming Quebec represents 22% of the Canadian population $\times$ 138M\$ = 30M\$), our mortality costs were much higher at 10.3G\$ in 2050 versus 0.7G\$ ($22\% \times 3\text{G\$} = 0.7\text{G\$}$) in their study.

Comparing our results with studies published internationally is difficult due to the differences in studied population, health outcomes, projections parameters and health systems, among others. More importantly, most studies have only assessed historical costs of specific heatwaves or projected a single health outcome (see references in the introduction), making comparisons even more challenging. That being said, projections of heat-related health costs in the literature obtained similar orders of magnitude that ours. For example, Lay et al. (2018) found that ED visit costs for hyperthermia could triple by 2050 in USA. Hübler et al. (2008) found a 6-fold increase in heat-related hospital admission costs in 2071–2100 in Germany compared to their historical period. Gronlund et al. (2019) projected increases of 6X to 7X in the costs associated with extreme heat-related mortality, hospital admissions and ED visits in Michigan (USA). A recent review on heat-related healthcare costs noted that heat already poses a significant burden on healthcare and that this burden will substantially increase in the future (Wondmagegn et al., 2019), which is consistent with our results. To our knowledge, our study is the first to consider such a wide range of health outcomes to provide a complete picture of both historical and projected heat-related health costs, using the province of Quebec as a case study.

Our results showed that the heat-related health costs were mainly driven by intangible components such as loss of life (90–95% of the grand total) and loss of well-being (5–10%). Some previous studies have also highlighted that mortality effects account for the bulk of the health costs when several health outcomes are simultaneously analyzed (Adélaïde et al., 2021; Beugin et al., 2023; Clark et al., 2021; Knowlton et al., 2011; Larrivée et al., 2015; Limaye et al., 2019). On the other hand, the loss of well-being during HW has been less studied, but it could account for a significant proportion of the heat load, as also demonstrated in France (Adélaïde et al., 2021). The strong importance of mortality costs raises the question of the optimal valuation method using either the Value of a Statistical Life (VSL), an age-adjusted VSL or the Value of a

Life Year (VoLY) (Chiabai et al., 2018). In our study, a central VSL of 8.2M\$ was considered with lower and upper bounds of respectively 4.1M\$ and 11.8M\$, as used in other climate change and air pollution impact assessments in Canada (Beugin et al., 2023; Clark et al., 2021; Health Canada, 2021). Given that the older people are more likely to die during extreme heat events (Benmarhnia et al., 2015), an age-adjusted VSL or VoLY metric to value mortality would lead to a lower estimate than what was obtained here. Comparing different mortality valuation methods is left for future research as this would require age-adjusted exposure-response functions that were not available at the time this study was conducted. Such a study could also help understand the role of aging in projecting heat-related health costs (Cole et al., 2023).

Although intangible costs were the most important for the grand total, direct and indirect costs still represented annual costs of ~15M\$ and ~5M\$ respectively in the historical period, expected to rise to ~36M\$ and ~12M\$ under SSP2–4.5 and to ~53M\$ and ~17M\$ under SSP5–8.5, respectively. Compared to intangible effects that are borne by the society as a whole (Limaye et al., 2019), direct costs are assumed by the government, individuals or insurers. In the case of Quebec, which has universal health insurance, these costs are covered by the Ministry of Health's limited budget, and will contribute to the ever-increasing costs of the healthcare system (Wondmagegn et al., 2019). Furthermore, direct healthcare costs were probably underestimated in this study because 1) not all heat-related impacts were considered due to the lack of data (e.g., general practitioner consultations, community nursing visits, 911 emergency calls) and 2) the relationship between all-cause hospital admissions and heat was found to be weak with an historical attributable fraction of only 0.1% in Quebec (Boudreault et al., 2024c). Studying cause-specific hospitalizations could lead to higher direct costs, as shown by Bai et al. (2018) with heat-related hospitalization costs for coronary heart disease and stroke alone totalling 14.5M\$ by year in Ontario, Canada. Regarding indirect costs, they only included the loss of productivity associated with medical consultations (i.e., absenteeism), while heat has much broader productivity and economic effects (Borg et al., 2021; Callahan & Mankin, 2022; García-León et al., 2021), that were out-of-scope for this study focussing on health impacts.

This study has several strengths. First, it uses state-of-the-art and locally developed exposure-response functions to compute historical (1990–2019) and projected (2040–2069) heat-attributable numbers for various mortality and morbidity outcomes. Second, it considered a variety of direct healthcare, indirect productivity and intangible societal costs, thus providing a comprehensive picture of the health costs associated with two heat exposures (i.e., all heat and extreme heat). Third, costs were calculated for both periods with combinations of demographic

and climate projections under two SSP, based on an ensemble of 15 bias-corrected and downscaled climate models from CMIP6. Finally, uncertainties in health costs were assessed using Monte Carlo simulations that considered the variability in exposure-response functions, economic metrics and climate models employed.

Limitations must also be acknowledged. First, all projections were made without socioeconomic changes other than demographic growth, excluding potential future adaptation to heat, rising healthcare costs and population aging, for example. Second, in the absence of stratified heat-health exposure-response functions, all economic metrics employed were averaged across all genders, age groups, causes of illness and regions, which may mask large differences in economic impacts (Merrill et al., 2008; Schmeltz et al., 2016), especially for mortality. Third, due to Canada's health jurisdictions, the studied health outcomes were only fully available in the province of Quebec. Therefore, the study region was limited to that province. Fourth, even though correction techniques were applied to the climate simulations, some residual bias may remain between observed and simulated data. However, such difference should not affect the users' needs to plan adaptation accordingly (Lavoie et al., 2024b). Finally, simulations (and resulting empirical 95% confidence intervals) were derived using a Monte Carlo technique, assuming for example a multivariate normal distribution of the parameters of the exposure-response functions as commonly done in environmental epidemiology studies. This method could be refined in a future study, as well as adapted to include the uncertainties of demographic projections that were not available at the time this study was conducted.

6.5 Conclusion

Heat-related health costs already represented a significant direct, indirect and intangible burden in Quebec, Canada. As a result of climate and demographic changes, but without considering other socioeconomic factors, this burden will significantly increase over the next 50 years by 3X under a middle-of-the-road scenario and by 5X under a pessimistic one. This new evidence underlines the importance of both climate change mitigation (i.e., reducing GHG emissions) and new measures to limit heat-related effects. In addition, this study can serve as a basis for investigating the role of adaptation in projecting health costs, as well as for eventual cost-benefit analyses of targeted adaptation interventions. Finally, the proposed robust and state-of-the-art methodology can be applied to other regions of the world that are also experiencing the devastating effects of extreme heat.

Credit authorship contribution statement

Jérémie Boudreault: Conceptualization, Methodology, Data Curation, Formal analysis, Visualization, Software, Writing—Original Draft, Review and Editing, Funding acquisition. **Céline Campagna:** Conceptualization, Writing—Review & Editing, Supervision, Project administration, Funding acquisition. **Éric Lavigne:** Conceptualization, Methodology, Software, Writing—Review and Editing. **Fateh Chebana:** Conceptualization, Writing—Review & Editing, Supervision, Project administration, Funding acquisition.

Funding

The first author has received funding from the Natural Sciences and Engineering Research Council of Canada (Vanier Scholarship, #CGV-180821), the Canadian Institute of Health Research (Health System Impact Fellowship, #IF1-184093), Ouranos (Real-Décoste Excellence Scholarship, #RDX-317725) and the National Institute of Public Health of Quebec (no grant number).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Denis Hamel and Louis Rochette from INSPQ for their help with the health data extraction, as well as the scenarios and climate services team at Ouranos for their help with the climate data. The authors are grateful to the editor and two anonymous reviewers who helped improved this paper.

Data availability

Authors do not have permission to share health data that was provided by *Institut national de santé publique du Québec* (INSPQ). Weather data was obtained from Daymet of the NASA. Climate data was obtained from Ouranos using the Power Analytics and Visualization for Climate Science (PAVICS) platform. PAVICS is funded through Ouranos, the Computer Research Institute of Montreal, Environment and Climate Change Canada, CANARIE, the Fonds Vert and

the Fonds d'électrification et de changements climatiques, the Canadian Foundation for Innovation, and the Fonds de Recherche du Québec. Demographic data was obtained from the *Ministère de la Santé et des Services Sociaux du Québec* (MSSS) for the historical period and from Statistics Canada for the projected period. Economic metrics were all publicly available from the Canadian Institute for Health Information (CIHI), MSSS, *Institut de la Statistique du Québec* (ISQ), Statistics Canada or previously published literature. The codes to reproduce the analyses are available on the first author's Github page: https://github.com/jeremieboudreault/paper_heat_health_costs_qc.

Supplementary data

Supplementary data to this article can be found online at:

<https://doi.org/10.1016/j.scitotenv.2024.178022>

Appendix 6.A. Economic metrics

Tableau 6.3 Overview of the economic metrics used in this study and corresponding references, along with their uncertainty and distributions used to compute Monte Carlo simulations. All values are expressed in 2019 Canadian dollars (\$), unless otherwise specified.

Metric	Central value	Uncertainty values	Distribution	References
Unit cost of an HOS	6 500 \$ ^a	Lower value of 1 750\$ used in Quebec ^b and higher value of 11 500\$ used in Ontario ^c	Uniform (lower to higher bounds)	^a CIHI (2024a) ^b Ripoche et al. (2023) ^c Bai et al (2018)
Unit costs of an EDV	303 \$ ^d	None found	Triangular (±33% around central value)	^d CIHI (2020)
Unit costs of an AMB	250 \$ ^e	Lower value of 125\$ in Quebec ^f and upper value of 500\$ in Canada ^e	Uniform (lower to higher bounds)	^e CIHI (2016) ^f Québec (2024a)
Unit costs of an 811-call	10 \$, the median nurse wage of 38.5\$ ^g × a mean duration of call of 0.25h ^h	None found	Triangular (±33% around central value)	^g Canada (2024) ^h Québec (2024b)
Costs of HW emergency teams	3.2 M\$ for a HW touching 4–5 HR ⁱ	None found	Triangular (±33% around central value)	ⁱ Larrivée et al. (2015)
LoS of an HOS	7.2 days ^j	None found	Triangular (±33% around central value)	^j CIHI (2023)
LoS of an EDV	4.9 hours ^k	90% quantile of 19.6 hours in Quebec ^k	Gamma (Mean = 4.9 and Q90 closest to 19.6)	^k CIHI (2024b)
Duration of an 811 call	0.5 hours, 0.25h for the call duration ^h + waiting time of 0.25h	Lower value of 0.25h ^h and higher value of 0.75h	Uniform (lower to higher bounds)	
Hourly wage	27 \$/h ^l	None found	Triangular (±33% around central value)	^l ISQ (2024)
VSL	8.2 M\$ ^m	A lower 4.1M\$ used in Quebec ⁱ and an upper value of 11.8M\$ ^m	Uniform (lower to higher bounds)	^m Boyd et al. (2020)
MRAD	33 \$/person-day ⁿ	A SD of 13\$ ⁿ	Normal (Mean = 33\$, SD = 13\$)	ⁿ Health Canada (2021)

AMB = Ambulance transport. EDV = Emergency department visit. HOS = Hospital admission. HR = Health region. HW = Extreme heatwave. LoS = Length of Stay. MRAD = Minor Restricted Activity Day. SD = Standard deviation. VSL = Value of a Statistical Life. 811 = Call to the 811-health hotline.

CHAPITRE 7 : CONCLUSIONS ET PERSPECTIVES

7 CONCLUSIONS ET PERSPECTIVES

Ce chapitre conclut la thèse en présentant une synthèse des travaux de recherche, les principaux résultats, ainsi que quelques éléments supplémentaires de discussion (section 7.1). Finalement, les perspectives pour les travaux futurs sont présentées (section 7.2).

7.1 Principales conclusions

Aux chapitres 2, 3 et 4, l'apprentissage automatique et profond a été utilisé pour une première fois au Québec afin de prédire les impacts de la chaleur sur la santé publique (Boudreault et al., 2023; 2024a; 2024b). Ces travaux se sont démarqués de ceux déjà publiés dans la littérature de diverses façons. D'abord, le nombre d'approches utilisées et comparées était notable, d'abord avec neuf modèles au chapitre 2, puis en réutilisant les cinq modèles les plus pertinents aux chapitres 3 et 4. Les résultats ont permis de conclure sur les forces et faiblesses des différentes approches selon le contexte. Au chapitre 2, les modèles d'ensemble d'arbres se sont démarqués, mais aussi les modèles statistiques non linéaires et l'arbre de décision lors des validations partielles pendant les vagues de chaleur extrême. Au chapitre 3, le modèle statistique non linéaire (GAM) a obtenu des performances similaires aux modèles plus complexes d'apprentissage automatique et profond après avoir effectué une sélection des prédicteurs les plus pertinents, concluant qu'une seule variable de température était suffisante pour prévoir la mortalité liée à la chaleur. Suivant cette sélection, les performances pour prédire la mortalité à Montréal étaient supérieures que celles obtenues au chapitre 2, ce qui expose un potentiel surapprentissage des modèles au chapitre 2 par rapport à ceux du chapitre 3. Finalement, au chapitre 4, les modèles de réseaux de neurones profonds semblaient les plus performants pour une majorité d'issues sanitaires testées. Cela démontre la capacité de ces modèles, avec leur structure de plusieurs couches cachées, de bien performer sur de grandes bases de données contenant à la fois des représentations temporelles et spatiales des issues sanitaires, ce qui n'était pas le cas aux chapitres 2 et 3. Cependant, la modélisation de ces tendances au sein même des modèles peut venir gonfler artificiellement les performances obtenues, par rapport à la modélisation de la déviation de l'issue de santé comme aux chapitres 2 et 3 (où les tendances sont retirées avant la modélisation). Au chapitre 4, le modèle de *gradient boosting* a obtenu les meilleures performances lors des vagues de chaleur extrême, ce qui démontre le besoin de valider ces modèles en contexte d'événements rares et extrêmes en vue de leur utilisation potentielle par les autorités de santé publique, comme aussi fait au chapitre 2. En conclusion, autant les modèles

statistiques non linéaires, les modèles d'ensemble de type *gradient boosting* et les réseaux de neurones profonds sont pertinents pour prévoir les impacts sanitaires de la chaleur. Cependant, il n'est pas possible de conclure qu'un modèle ou qu'un prédicteur est systématiquement plus performant qu'un autre, comme les trois premiers chapitres présentent des cas d'études qui ne sont pas directement comparables étant donné les différentes issues sanitaires, variables et régions considérées. De plus, les différences de R^2 ou de MAE sont parfois faibles, ce qui ne justifierait pas de choisir un modèle ou un prédicteur plutôt qu'un autre qui obtient des performances assez similaires.

Les travaux des chapitres 2, 3 et 4 se sont aussi démarqués par les données mises à profit, soit de longues séries temporelles, des données sur un ensemble de régions simultanément, et diverses issues sanitaires incluant à la fois la mortalité et plusieurs variables de morbidité. Cela a donc permis de tester les modèles d'apprentissage automatique avec des jeux de données cohérents avec leur gourmandise en données. De plus, la richesse des prédicteurs utilisés incluant une multitude de variables météorologiques et de pollution de l'air, ainsi que l'explicabilité fournie par les modèles, ont permis de conclure sur les meilleures variables à utiliser pour modéliser les effets sanitaires de la chaleur selon le contexte, plutôt que de présenter ces modèles comme des boîtes noires tel que noté lors de la revue de littérature. Finalement, la calibration transparente au niveau des choix initiaux et finaux des hyperparamètres des modèles et l'évaluation subséquente de leur performance et de leur explicabilité ont permis une compréhension et une reproductibilité inédite dans ce domaine comparativement aux applications passées. Ces travaux ont ainsi permis de générer de nouvelles connaissances sur les impacts à la santé de la chaleur, ainsi que sur les variables environnementales et sanitaires sous-jacentes.

Ces connaissances ont ensuite été mises à profit dans les chapitres subséquents de la thèse. Les chapitres 5 et 6 de cette thèse ont utilisé le *Distributed Lag Non-Linear Model (DLNM)* et la méta-régression multivariée afin de quantifier le fardeau de la chaleur pour plusieurs issues sanitaires et les coûts associés, ce qui était une première au Québec (Boudreault et al., 2024c; 2025a). Bien que le DLNM n'était pas le modèle le plus pertinent pour la prévision des décès liés à la chaleur au chapitre 2, ce modèle demeure le plus utilisé dans la littérature pour la quantification du fardeau sanitaire de la chaleur, d'où son utilisation pour la suite de la thèse. Il est important de rappeler ici que les modèles des chapitres 2, 3 et 4 ont été développés dans l'optique de faire la prévision des impacts sanitaires de la chaleur, et non pour le calcul de fonctions dose-réponse nécessaires au calcul du fardeau sanitaire (voir la perspective #2 pour la mise à profit de l'apprentissage automatique pour le développement de ces fonctions). De plus, l'utilisation

d'autres approches que le DLNM aurait produit de multiples estimations du fardeau sanitaire et économique, ce qui pourrait complexifier la prise de décision ultérieure. Ainsi, pour cette thèse, nous avons préféré estimer ces nouveaux résultats avec une seule approche, soit le DLNM, qui est l'approche la plus utilisée dans la littérature internationale pour ce faire.

Le calcul du fardeau sanitaire représente une contribution importante du chapitre 5 plutôt que d'étudier seulement la relation chaleur-santé comme dans plusieurs autres études passées dont celles des chapitres 2, 3 et 4. De plus, l'application du DLNM à plusieurs issues sanitaires incluant la mortalité et plusieurs variables de morbidité était aussi novatrice par rapport aux travaux passés dans la littérature. Appliqués à l'ensemble des régions du Québec, les modèles DLNM ont permis de calculer un fardeau attribuable à la chaleur de 470 décès, 225 hospitalisations, 36 000 visites à l'urgence, 7200 transports ambulanciers et 15 000 appels à Info-Santé 811 chaque été en moyenne au Québec durant la période historique. Ces résultats ont été validés par plusieurs analyses de sensibilité, notamment en comparant différents indices de température et en analysant l'effet de la pollution de l'air, comme cela avait aussi été fait aux chapitres 2, 3 et 4.

Les modèles développés au chapitre 5 ont été repris au chapitre 6 visant à quantifier les coûts historiques et futurs associés à la chaleur. Les coûts directs en termes de soins de santé, les coûts indirects en termes d'absentéisme et les coûts intangibles en termes de mortalité et d'activités réduites ont été estimés en combinant chacun des DLNM à des projections climatiques et démographiques, ainsi qu'à l'aide de différentes métriques économiques. Il s'agissait de la première étude aussi complète sur les coûts actuels et futurs de la chaleur à être effectuée dans la littérature internationale. En étudiant la chaleur dans l'ensemble de l'été, un portrait plus exhaustif des coûts sanitaires a pu être brossé par rapport aux études passées ne portant que sur les vagues de chaleur extrême. De surcroît, ces coûts ont aussi été projetés jusqu'en 2070 à l'aide de simulations climatiques et démographiques, illustrant le portrait de la situation pour les décennies à venir. Ces projections ont été faites sur la base de deux scénarios de réchauffement et de plusieurs modèles globaux du climat, ce qui a aussi permis d'analyser l'incertitude dans les projections à l'aide de simulations de type Monte Carlo. Les résultats ont montré que les coûts directs représentaient 15 millions de dollars annuellement dans la période historique, les coûts indirects 5 millions de dollars, et les coûts intangibles 3,6 milliards de dollars. Projetés pour la période 2040 à 2070, les coûts totaux seraient multipliés par 3 (10,9 milliards de dollars annuellement) selon un scénario climatique et démographique médian, mais par 5 (17,4 milliards de dollars annuellement) selon un scénario plus pessimiste. Ces projections sanitaires et

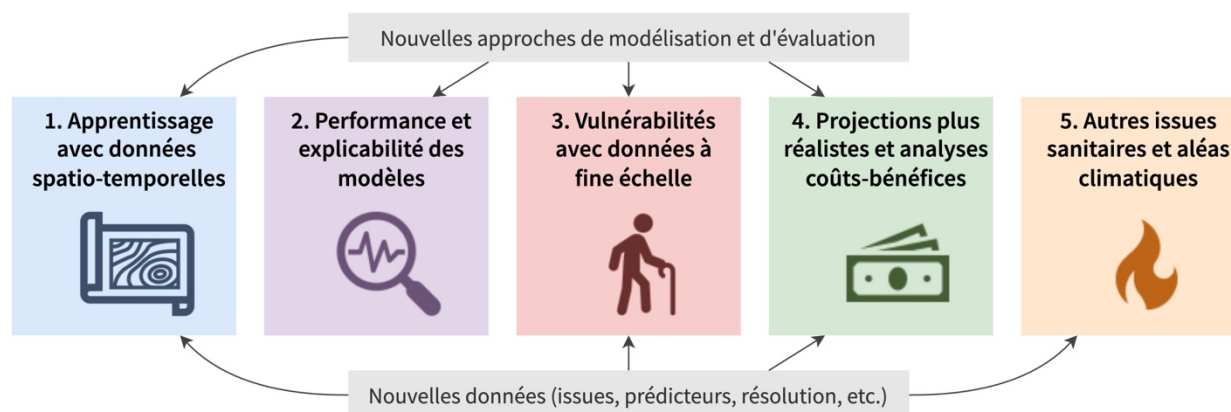
économiques étaient cohérentes avec celles effectuées précédemment au Canada ainsi qu'au niveau international.

Au meilleur de notre connaissance, cette thèse constitue **la première à comparer différentes approches de la science des données, soit celles d'apprentissage automatique et celles couramment utilisées en épidémiologie environnementale, afin de mieux modéliser les impacts sanitaires et économiques de la chaleur** dans le contexte de la province de Québec au Canada. Cette thèse a permis de démontrer le potentiel, les forces et les faiblesses de chacune de ces deux grandes familles d'approches selon les besoins, pour une même grande base de données contenant plusieurs issues sanitaires, prédicteurs météorologiques et régions d'étude. Les modèles développés, les connaissances générées ainsi que les données probantes produites sur les impacts à la santé et les coûts associés de la chaleur sont une clé importante pour la lutte aux changements climatiques sur le territoire québécois et canadien. Nous espérons que ces travaux seront mis à profit par les acteurs à tous les niveaux afin d'améliorer la vigie et la surveillance de la chaleur au Québec, ainsi que mieux prévoir les actions d'adaptation et d'atténuation afin de limiter la facture humaine et financière actuelle et à venir de la chaleur.

7.2 Principales perspectives

Les travaux poursuivis dans cette thèse ainsi que leurs limites ouvrent la porte à plusieurs perspectives. Cinq perspectives principales sont décrites ci-dessous et résumées à la Figure 7.1.

Figure 7.1 Perspectives pour les travaux futurs de la thèse.



Perspective #1. Modèles d'apprentissage profond avec données spatio-temporelles

Bien que les modèles d'apprentissage automatique et profond retenus pour cette thèse étaient ceux les plus utilisés et les plus pertinents dans le contexte, ils se sont limités à des modèles basés sur des arbres et des réseaux de neurones, tous utilisant des données tabulaires. En effet, les variables météorologiques ont été agrégées sur une durée (p. ex., 1 à 3 jours) et par région (p. ex., la région de Montréal) avant d'être utilisées pour entraîner les différents modèles. Or, il est connu que les modèles d'apprentissage profond peuvent sous-performer avec cette structure plus classique de données tabulaires (Shwartz-Ziv & Armon, 2022). Cela pourrait entre autres expliquer pourquoi la performance des modèles d'apprentissage profond n'était pas toujours supérieure à celle des modèles d'épidémiologie environnementale ou d'apprentissage automatique, comme aussi observé en santé clinique (Christodoulou et al., 2019). Ainsi, il est proposé comme première perspective de mettre à profit des modèles d'apprentissage profond combinant les réseaux à convolution et les réseaux récurrents afin de permettre l'utilisation de variables spatio-temporelles brutes comme prédicteurs (VoPham et al., 2018). Par exemple, la grille des températures quotidiennes pourrait être utilisée directement dans le modèle plutôt qu'une agrégation spatiale et temporelle de celle-ci, ce qui permettrait de mieux mesurer l'exposition à la température et, potentiellement, mieux prédire les effets à la santé. Cette approche permettrait de surmonter les limites des modèles statistiques plus classiques utilisés en épidémiologie environnementale, mais rendrait par le même coup plus difficiles les comparaisons avec les approches traditionnelles, d'où pourquoi elle n'a pas été considérée dans le cadre de la présente thèse. De surcroît, considérer les issues sanitaires simultanément, plutôt que séparément comme aux chapitres 4 ou 5, serait une autre piste intéressante telle que soulignée dans une revue de littérature récente sur le sujet (Boudreault et al., 2025c).

Perspective #2. Nouvelles méthodes d'évaluation des modèles

Les résultats de cette thèse ont montré que les conclusions sur les approches ou variables les plus performantes étaient influencées par la façon d'évaluer les modèles. Cette deuxième perspective vise donc l'utilisation de nouvelles méthodes plus robustes et complètes afin d'évaluer les modèles d'apprentissage automatique dans le contexte de la chaleur et de la santé. Les critères retenus pour cette thèse étaient ceux les plus utilisés dans la littérature comme le R^2 , le RMSE ou le MAE afin de permettre des comparaisons avec les études passées. Or, les résultats de ces critères ont montré des divergences de performance lorsque comparés avec les validations partielles effectuées lors des vagues de chaleur récentes. Ainsi, des critères spécifiques pour valider les modèles en période de chaleur extrême devraient être adaptés ou

développés à cette fin. Par exemple, un critère basé sur la distribution des données, comme la déviance d'une distribution de Poisson, pourrait être plus approprié qu'un critère basé sur une loi normale comme le R^2 . Un tel critère pourrait aussi être utilisé directement comme fonction de perte pour calibrer les différents modèles, plutôt que de supposer une loi normale qui n'est pas toujours adéquate en présence de données de comptage.

Au-delà de la performance, l'évaluation des modèles repose aussi sur l'explicabilité qui peut être fournie afin de limiter l'effet "boîte noire" des modèles d'apprentissage automatique. Or, seule l'importance des prédicteurs (*feature importance* en anglais) a été utilisée dans cette thèse afin d'expliquer les modèles d'apprentissage automatique. Ainsi, il est proposé de mettre à profit des méthodes plus complètes d'explicabilité comme les *SHapley Additive exPlanations* (aussi appelées valeurs SHAP) afin de visualiser l'association entre les prédicteurs et les impacts sanitaires, ainsi que les interactions potentielles (p. ex. entre la température et la pollution de l'air) (Rasheed et al., 2022). L'explicabilité des modèles d'apprentissage automatique a été identifiée comme une lacune dans la revue de littérature, d'où l'intérêt de présenter au moins l'importance de chaque prédicteur comme une première étape pour des modèles plus transparents. Avec une explicabilité accrue, on pourrait aussi penser à développer de fonctions dose-réponse directement avec les modèles d'apprentissage automatique (Cox Jr, 2023; Linden et al., 2016), comme cela a été fait aux chapitres 5 et 6 mais seulement avec le modèle DLNM. Il s'agit d'une avenue intéressante à explorer, bien qu'elle dépasse le cadre de cette thèse qui visait plutôt à utiliser l'apprentissage automatique pour prédire les impacts sanitaires de la chaleur, comme cela avait été fait historiquement dans la littérature limitée sur le sujet. Le développement de telles fonctions chaleur-santé avec l'apprentissage automatique nécessiterait des travaux de recherche supplémentaires afin de permettre, entre autres : 1) l'établissement d'un niveau de risque minimum, 2) une forme flexible de la relation permettant l'extrapolation, 3) la considération des délais, et 4) la quantification de l'incertitude dans ces relations.

Perspective #3. Étude des populations vulnérables avec données à fine échelle

Cette thèse s'est concentrée principalement sur les issues sanitaires toutes causes dans l'ensemble de la population québécoise à l'échelle des RMR ou des RSS. Comme il s'agissait des premiers travaux visant à comparer les méthodes de la science des données pour modéliser les impacts de la chaleur, ce choix a été fait afin de se coller aux indicateurs utilisés par l'INSPQ pour la vigie et la surveillance de la chaleur au Québec (Lamothe et al., 2025; Toutant et al., 2011). Or, l'étude de la population générale à une échelle géographique relativement large a pour conséquence de masquer des divergences potentielles des effets de la chaleur en fonction des

caractéristiques démographiques, sanitaires ou géographiques (Benmarhnia et al., 2015; Ebi et al., 2021). De surcroît, cela pourrait aussi masquer les différences qui existent en termes de coûts de santé pour différentes sous-populations, causes de maladie ou régions (Chiabai et al., 2018; Noe et al., 2012; Schmeltz et al., 2016). Ainsi, la troisième perspective suggère de revoir les impacts sanitaires et les coûts associés à la chaleur à l'égard de différentes stratifications comme l'âge, le sexe ou le genre et la cause de maladie. Il s'agit d'une piste intéressante, car de telles données stratifiées sur les coûts de santé permettraient de mieux cibler l'implémentation de mesures efficaces d'adaptation à la chaleur pour les populations les plus vulnérables. De surcroît, une échelle spatiale plus fine comme l'aire de diffusion permettrait d'analyser les facteurs géographiques de vulnérabilité à la chaleur comme la défavorisation matérielle et sociale ou l'environnement bâti, ce qui n'était pas possible à l'échelle utilisée. Qui plus est, des données individuelles de santé permettraient d'analyser d'autres facteurs de vulnérabilités comme les comorbidités ou la prise de médicaments (Boudreault et al., 2025b; McLean et al., 2024; Thompson et al., 2024; Tupinier Martin et al., 2024), bien que l'accès à de telles données nécessite de nouvelles approbations afin de respecter la confidentialité des renseignements de santé. Si de telles données à plus fine échelle sont disponibles, autant des approches d'apprentissage automatique que d'épidémiologie environnementale pourraient être utilisées et comparées entre elles, mettant notamment à profit le devis de type *Case Time Series* récemment introduit avec le DLNM (Gasparrini, 2021; 2022). Comme l'étude des vulnérabilités à la chaleur dépassait largement les objectifs de la thèse, l'utilisation de données agrégées par RMR ou RSS sur l'ensemble de la population québécoise demeurerait appropriée.

Perspective #4. Projections plus réalistes et analyses coûts-bénéfices de l'adaptation

Dans cette thèse, les projections des effets à la santé et des coûts associés ont été effectuées en incluant les changements climatiques et la croissance de la population. Ces projections permettent d'analyser le portrait futur des impacts de la chaleur avec l'hypothèse d'aucune nouvelle adaptation à la chaleur, comme c'est souvent le cas dans la littérature (Vicedo-Cabrera et al., 2019). Par conséquent, cette quatrième perspective recommande d'intégrer à ces projections d'autres hypothèses comme le vieillissement de la population en produisant d'abord des fonctions dose-réponse stratifiées pour l'âge (Chen et al., 2020; 2024), l'acclimatation aux températures plus élevées (Cordiner et al., 2024; Navas-Martín et al., 2024), et la hausse des coûts de soins de santé ou l'inflation (Orlov et al., 2024), entre autres. Des travaux initiaux sur les coûts sanitaires de la chaleur en climat futur étaient cependant nécessaires avant d'inclure toutes ces hypothèses qui présentent aussi leur lot d'incertitudes. Une autre façon de rendre plus

réalistes les projections serait de mettre à profit des données à plus fine échelle et d'inclure diverses stratifications, comme déjà proposé à la perspective #3. Bien que perfectible, cette première quantification des coûts sanitaires de la chaleur ouvre déjà la porte à la réalisation d'analyses coûts-bénéfices de diverses mesures d'adaptation à la chaleur dans le contexte québécois. De telles études ont déjà été effectuées dans d'autres juridictions, par exemple pour les systèmes d'alerte à la chaleur (Chiabai et al., 2018; Ebi et al., 2004; Hunt et al., 2017; Williams et al., 2021). Les politiques de verdissement et de lutte aux îlots de chaleur urbains sont d'autres exemples intéressants à traiter pour ce genre d'analyses, en cohérence avec les schémas d'aménagement des villes (Wu et al., 2025; Smargiassi et al., 2020). À notre connaissance, de telles études n'ont pas encore été effectuées dans la littérature concernant l'adaptation à la chaleur.

Perspective #5. Extension à d'autres aléas climatiques et issues sanitaires

Cette thèse s'est intéressée à l'impact de la chaleur sur la mortalité et diverses variables de morbidité durant l'ensemble de l'été. Ainsi, comme cinquième et dernière perspective, nous proposons d'étendre les présents travaux à d'autres issues sanitaires et aléas climatiques. Par exemple, des travaux futurs utilisant la science des données pourraient toucher, entre autres, aux accidents de travail (Adam-Poupart et al., 2014; Larrivée et al., 2015), aux consultations en médecine familiale (Adélaïde et al., 2021) ou aux appels au 911 (Bassil et al., 2009). Des démarches auprès d'autres organisations que l'INSPQ seraient nécessaires afin d'accéder à ces données, notamment la Commission des normes, de l'équité, de la santé et de la sécurité au travail (CNESST), la Régie de l'assurance maladie du Québec (RAMQ) ou les municipalités, en plus de déposer une nouvelle demande d'approbation éthique. Concernant les aléas climatiques, les vagues de chaleur extrême pourraient faire l'objet d'analyses similaires plutôt que la chaleur de manière plus générale, bien que cela soit déjà en partie fait par l'INSPQ pour quantifier les impacts sanitaires (Lamothe et al., 2025). La modélisation des issues sanitaires à une autre résolution temporelle (p. ex. hebdomadaire) pourrait être intéressante pour ces événements ayant des impacts sanitaires sur de plus longues durées. De surcroît, des études à l'international ont montré que les températures froides ont des impacts significatifs à la santé, souvent même plus importants que ceux des températures chaudes (Gasparrini et al., 2015; Zhao et al., 2021). Ainsi, les températures hivernales ainsi les vagues de froid extrême pourraient être analysées afin d'avoir une idée du fardeau sanitaire et économique associé à l'ensemble des gammes de températures au Québec, comme déjà effectué pour la chaleur dans le cadre de la thèse. Finalement, d'autres aléas climatiques comme les précipitations estivales ou hivernales, les

inondations ou la pollution liée aux feux de forêt pourraient faire l'objet d'analyses similaires avec les méthodes de la science des données pour en connaître les impacts sanitaires et les coûts qui y sont associés.

8 BIBLIOGRAPHIE

- Adam-Poupart, A., Nicolakakis, N., Anassour Laouan Sidi, E., Berry, P., Campagna, C., Chaumont, D., Hamel, D., Labrèche, F., Sassine, M.-P., Smargiassi, A., & Zayed, J. (2021). Climate change and heat vulnerabilities of Canadian workers : Focus on the Central and Western provinces of Canada. *Institut national de la santé publique du Québec*.
- Adam-Poupart, A., Smargiassi, A., Busque, M.-A., Duguay, P., Fournier, M., Zayed, J., & Labrèche, F. (2014). Summer outdoor temperature and occupational heat-related illnesses in Quebec (Canada). *Environmental research*, 134, 339-344.
- Adélaïde, L., Chanel, O., & Pascal, M. (2021). Health effects from heat waves in France : An economic evaluation. *The European Journal of Health Economics*, 23(1), 119-131.
- Adresses Québec. (2023). *AQgéobâti*. <https://adressesquebec.gouv.qc.ca/aqgeobati.asp>
- Analitis, A., De'Donato, F., Scortichini, M., Lanki, T., Basagana, X., Ballester, F., Astrom, C., Paldy, A., Pascal, M., & Gasparrini, A. (2018). Synergistic effects of ambient temperature and air pollution on health in Europe : Results from the PHASE project. *International journal of environmental research and public health*, 15(9), 1856.
- Analitis, A., Michelozzi, P., D'Ippoliti, D., De'Donato, F., Menne, B., Matthies, F., Atkinson, R. W., Iñiguez, C., Basagaña, X., & Schneider, A. (2014). Effects of heat waves on mortality : Effect modification and confounding by air pollutants. *Epidemiology*, 25(1), 15-22.
- Armstrong, B. (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, 17(6), 624-631.
- Atkinson, R. W., Kang, S., Anderson, H. R., Mills, I. C., & Walton, H. A. (2014). Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions : A systematic review and meta-analysis. *Thorax*, 69(7), 660-665.
- Auger, N., Fraser, W. D., Sauve, R., Bilodeau-Bertrand, M., & Kosatsky, T. (2017). Risk of congenital heart defects after ambient heat exposure early in pregnancy. *Environmental health perspectives*, 125(1), 8-14.
- Auger, N., Naimi, A. I., Smargiassi, A., Lo, E., & Kosatsky, T. (2014). Extreme heat and risk of early delivery among preterm and term pregnancies. *Epidemiology*, 25(3), 344-350.
- Austin, S. E., Ford, J. D., Berrang-Ford, L., Biesbroek, R., & Ross, N. A. (2019). Enabling local public health adaptation to climate change. *Social Science & Medicine*, 220, 236-244.
- Bai, L., Li, Q., Wang, J., Lavigne, E., Gasparrini, A., Copes, R., Yagouti, A., Burnett, R. T., Goldberg, M. S., & Cakmak, S. (2018). Increased coronary heart disease and stroke hospitalisations from ambient temperatures in Ontario. *Heart*, 104(8), 673-679.

- Bai, L., Li, Q., Wang, J., Lavigne, E., Gasparrini, A., Copes, R., Yagouti, A., Burnett, R. T., Goldberg, M. S., & Villeneuve, P. J. (2016). Hospitalizations from hypertensive diseases, diabetes, and arrhythmia in relation to low and high temperatures : Population-based study. *Scientific reports*, 6(1), 30283.
- Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R. F., Pegenaute, F., Herrmann, F. R., Robine, J. M., Basagaña, X., Tonne, C., Antó, J. M., & Achebak, H. (2023). Heat-related mortality in Europe during the summer of 2022. *Nature medicine*, 29(7), 1857-1866.
- Bank of Canada. (2024). Inflation Calculator. <https://www.bankofcanada.ca/rates/related/inflation-calculator/>
- Barnett, A., Tong, S., & Clements, A. C. (2010). What measure of temperature is the best predictor of mortality? *Environmental research*, 110(6), 604-611.
- Bassil, K. L., Cole, D. C., Moineddin, R., Craig, A. M., Lou, W. W., Schwartz, B., & Rea, E. (2009). Temporal and spatial variation of heat-related illness using 911 medical dispatch data. *Environmental research*, 109(5), 600-606.
- Basu, R. (2009). High ambient temperature and mortality: A review of epidemiologic studies from 2001 to 2008. *Environmental health*, 8(1), 1-13.
- Basu, R., Pearson, D., Malig, B., Broadwin, R., & Green, R. (2012). The effect of high ambient temperature on emergency room visits. *Epidemiology*, 23(6), 813-820.
- Basu, R., & Samet, J. M. (2002). Relation between elevated ambient temperature and mortality : A review of the epidemiologic evidence. *Epidemiologic reviews*, 24(2), 190-202.
- Bayentin, L., El Adlouni, S., Ouarda, T. B., Gosselin, P., Doyon, B., & Chebana, F. (2010). Spatial variability of climate effects on ischemic heart disease hospitalization rates for the period 1989-2006 in Quebec, Canada. *International journal of health geographics*, 9(1), 1-10.
- Bellinger, C., Mohamed Jabbar, M. S., Zaïane, O., & Osornio-Vargas, A. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. *BMC public health*, 17(1), 1-19.
- Benmarhnia, T., Deguen, S., Kaufman, J. S., & Smargiassi, A. (2015). Vulnerability to heat-related mortality. *Epidemiology*, 26(6), 781-793.
- Benmarhnia, T., Sottile, M.-F., Plante, C., Brand, A., Casati, B., Fournier, M., & Smargiassi, A. (2014). Variability in temperature-related mortality projections under climate change. *Environmental health perspectives*, 122(12), 1293-1298.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).
- Beugin, D., Clark, D., Miller, S., Ness, R., Pelai, R., & Wale, J. (2023). The case for adapting to extreme heat : Costs of the 2021 B.C. heat wave. *Canadian Climate Institute*.

- Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., & Armstrong, B. (2013). Time series regression studies in environmental epidemiology. *International journal of epidemiology*, 42(4), 1187-1195.
- Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning ? A primer for the epidemiologist. *American journal of epidemiology*, 188(12), 2222-2239.
- Bi, X., Wu, C., Wang, Y., Li, J., Wang, C., Hahs, A., Mavoa, S., Song, C., Konrad, C., & Emch, M. (2023). Changes in the associations between heatwaves and human mortality during two extreme hot summers in Shanghai, China. *Sustainable Cities and Society*, 95, 104581.
- Borg, M. A., Xiang, J., Anikeeva, O., Pisaniello, D., Hansen, A., Zander, K., Dear, K., Sim, M. R., & Peng, B. (2021). Occupational heat stress and economic burden : A review of global evidence. *Environmental research*, 195, 110781.
- Boudreault, J., Campagna, C., & Chebana, F. (2022). *Coûts de santé de la chaleur extrême en contexte de changements climatiques : État des connaissances et orientations futures*. 94 p. INRS, Centre Eau, Terre et Environnement. <https://espace.inrs.ca/id/eprint/13052/>
- Boudreault, J., Campagna, C., & Chebana, F. (2023). Machine and deep learning for modelling heat-health relationships. *Science of The Total Environment*, 892, 164660.
- Boudreault, J., Campagna, C., & Chebana, F. (2024a). Revisiting the importance of temperature, weather, and air pollution variables in heat-mortality relationships with machine learning. *Environmental Science and Pollution Research*, 31(9), 14059-14070.
- Boudreault, J., Ruf, A., Campagna, C., & Chebana, F. (2024b). Multi-region models built with machine and deep learning for predicting several heat-related health outcomes. *Sustainable Cities and Society*, 115, 105785.
- Boudreault, J., Lavigne, É., Campagna, C., & Chebana, F. (2024c). Estimating the heat-related mortality and morbidity burden in the province of Quebec, Canada. *Environmental Research*, 257, 119347.
- Boudreault, J., Campagna, C., Lavigne, É., & Chebana, F. (2025a). Projecting the overall heat-related health burden and associated economic costs in a climate change context in Quebec, Canada. *Science of The Total Environment*, 958, 178022.
- Boudreault, J., McLean, K. E., & Henderson, S. B. (2025b). Exploring the relationship between medications and heat-related community deaths during the 2021 heat dome: A hybrid approach using machine learning. *eBioMedicine*, 117, 105788.
- Boudreault, J., Lamothe, F., Campagna, C., & Chebana, F. (2025c). Machine learning for modelling the health impacts of extreme heat: A comprehensive literature review. *Environment International*, 206, 109965.
- Boyd, R., Eyzaguirre, J., Poulsen, F., Siegle, M., Thompson, A., Yamamoto, S., Osornio-Vargas, Erickson, A., & Urcelay, A. (2020). Costing Climate Change Impacts on Human Health Across Canada. *Prepared by ESSA Technologies Ltd. for the Canadian Institute for Climate Choices*.

- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Buckley, J. P., Samet, J. M., & Richardson, D. B. (2014). Commentary: Does air pollution confound studies of temperature? *Epidemiology*, 25(2), 242-245.
- Burkart, K. G., Brauer, M., Aravkin, A. Y., Godwin, W. W., Hay, S. I., He, J., Iannucci, V. C., Larson, S. L., Lim, S. S., & Liu, J. (2021). Estimating the cause-specific relative risks of non-optimal temperature on daily mortality: A two-part modelling approach applied to the Global Burden of Disease Study. *The Lancet*, 398(10301), 685-697.
- Bustinza, R., Lebel, G., Gosselin, P., Bélanger, D., & Chebana, F. (2013). Health impacts of the July 2010 heat wave in Quebec, Canada. *BMC public health*, 13(1), 1-7.
- Callahan, C. W., & Mankin, J. S. (2022). Globally unequal effect of extreme heat on economic growth. *Science Advances*, 8(43), eadd3726.
- Campbell, S. L., Remenyi, T., & Johnston, F. H. (2023). Methods of assessing health care costs in a changing climate: A case study of heatwaves and ambulance dispatches in Tasmania, Australia. *GeoHealth*, 7(10), e2023GH000914.
- Campbell, S., Remenyi, T. A., White, C. J., & Johnston, F. H. (2018). Heatwave and health impact research: A global review. *Health & place*, 53, 210-218.
- Carmona, R., Díaz, J., Mirón, I., Ortiz, C., Luna, M., & Linares, C. (2016). Mortality attributable to extreme temperatures in Spain: A comparative analysis by city. *Environment International*, 91, 22-28.
- Casati, B., Yagouti, A., & Chaumont, D. (2013). Regional climate projections of extreme heat events in nine pilot Canadian communities for public health planning. *Journal of applied meteorology and climatology*, 52(12), 2669-2698.
- CERFO. (2022). Cartographie des îlots de chaleur et de fraîcheur dans le Québec urbain à l'aide d'imagerie satellitaire Landsat-8 (2013-2014). *Centre d'enseignement et de recherche en foresterie de Sainte-Foy*. <https://www.donneesquebec.ca/recherche/dataset/ilots-de-chaleur-fraicheur-urbains-et-ecarts-de-temperature-relatifs-2013-2014/resource/a33969ba-143a-4524-88c3-8ec7485676b1>
- Chebana, F., Martel, B., Gosselin, P., Giroux, J.-X., & Ouarda, T. B. (2013). A general and flexible methodology to define thresholds for heat health watch and warning systems, applied to the province of Québec (Canada). *International journal of biometeorology*, 57(4), 631-644.
- Chen, K., De Schrijver, E., Sivaraj, S., Sera, F., Scovronick, N., Jiang, L., Roye, D., Lavigne, E., Kysely, J., & Urban, A. (2024). Impact of population aging on future temperature-related mortality at different global warming levels. *Nature communications*, 15(1), 1796.
- Chen, K., Vicedo-Cabrera, A. M., & Dubrow, R. (2020). Projections of ambient temperature-and air pollution-related mortality burden under combined climate change and population aging scenarios: A review. *Current environmental health reports*, 7, 243-255.

- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.
- Cheng, J., Xu, Z., Bambrick, H., Su, H., Tong, S., & Hu, W. (2018). Heatwave and elderly mortality: An evaluation of death burden and health costs considering short-term mortality displacement. *Environment international*, 115, 334-342.
- Cheng, J., Xu, Z., Bambrick, H., Su, H., Tong, S., & Hu, W. (2019). Impacts of exposure to ambient temperature on burden of disease: A systematic review of epidemiological evidence. *International journal of biometeorology*, 63, 1099-1115.
- Cheng, J., Xu, Z., Zhao, D., Xie, M., Zhang, H., Wang, S., & Su, H. (2016). The burden of extreme heat and heatwave on emergency ambulance dispatches: A time-series study in Huainan, China. *Science of the Total Environment*, 571, 27-33.
- Cheng, J., Xu, Z., Zhu, R., Wang, X., Jin, L., Song, J., & Su, H. (2014a). Impact of diurnal temperature range on human health: A systematic review. *International journal of biometeorology*, 58(9), 2011-2024.
- Cheng, J., Zhu, R., Xu, Z., Xu, X., Wang, X., Li, K., & Su, H. (2014b). Temperature variation between neighboring days and mortality: A distributed lag non-linear analysis. *International journal of public health*, 59, 923-931.
- Chiabai, A., Spadaro, J. V., & Neumann, M. B. (2018). Valuing deaths or years of life lost? Economic benefits of avoided mortality from early heat warning systems. *Mitigation and adaptation strategies for global change*, 23(7), 1159-1176.
- Chiu, Y. M. (2017). *Approches de modélisation des extrêmes dans l'étude des relations entre la santé et la météo*. Université du Québec, Institut national de la recherche scientifique.
- Chiu, Y. M., Chebana, F., Abdous, B., Bélanger, D., & Gosselin, P. (2021). Cardiovascular Health Peaks and Meteorological Conditions: A Quantile Regression Approach. *International Journal of Environmental Research and Public Health*, 18(24), 13277.
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, 110, 12-22.
- CIHI. (2016). Ambulance Use for Time-Sensitive Conditions: Stroke and Heart Attack—Data Tables, *Canadian Institute for Health Information*. <https://www.cihi.ca/sites/default/files/document/ambulance-use-data-tables-en.xlsx>
- CIHI. (2020). Hospital spending: Focus on the emergency department. *Canadian Institute for Health Information*. <https://www.cihi.ca/sites/default/files/document/hospital-spending-highlights-2020-en.pdf>
- CIHI. (2023). Hospital stays in Canada, 2021–2022. *Canadian Institute for Health Information*. <https://www.cihi.ca/en/hospital-stays-in-canada-2021-2022>

- CIHI. (2024a). Cost of a Standard Hospital Stay. *Canadian Institute for Health Information*. <https://www.cihi.ca/en/indicators/cost-of-a-standard-hospital-stay>
- CIHI. (2024b). NACRS emergency department visits and lengths of stay. *Canadian Institute for Health Information*. <https://www.cihi.ca/en/nacrs-emergency-department-visits-and-lengths-of-stay>
- Clark, D. G., Ness, R., Coffman, D., & Beugin, D. (2021). The Health Costs of Climate Change: How Canada Can Adapt, Prepare and Save Lives. *Canadian Institute for Climate Choices*.
- Cole, R., Hajat, S., Murage, P., Heaviside, C., Macintyre, H., Davies, M., & Wilkinson, P. (2023). The contribution of demographic changes to future heat-related health burdens under climate change scenarios. *Environment International*, 173, 107836.
- Cordiner, R., Wan, K., Hajat, S., & Macintyre, H. L. (2024). Accounting for adaptation when projecting climate change impacts on health: A review of temperature-related health impacts. *Environment international*, 188, 108761.
- Côté, J.-N., Germain, M., Levac, E., & Lavigne, E. (2024a). Vulnerability assessment of heat waves within a risk framework using artificial intelligence. *Science of The Total Environment*, 912, 169355.
- Côté, J.-N., Levac, E., Germain, M., & Lavigne, E. (2024b). Projected risk and vulnerability to heat waves for Montreal, Quebec, using Gaussian processes. *Sustainable Cities and Society*, 116, 105907.
- Cox Jr, L. A. (2023). What is an exposure-response curve? *Global Epidemiology*, 6, 100114.
- Crank, P. J., Hondula, D. M., & Sailor, D. J. (2023). Mental health and air temperature: Attributable risk analysis for schizophrenia hospital admissions in arid urban climates. *Science of The Total Environment*, 862, 160599.
- Curtis, S., Fair, A., Wistow, J., Val, D. V., & Oven, K. (2017). Impact of extreme weather events and climate change for health and social care systems. *Environmental Health*, 16(1), 23-32.
- Davis, R. E., McGregor, G. R., & Enfield, K. B. (2016). Humidity: A review and primer on atmospheric moisture and human health. *Environmental research*, 144, 106-116.
- Deschênes, O., & Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics*, 3(4), 152-185.
- Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, 52(3), 740-798.
- Diaz, D., & Moore, F. (2017). Quantifying the economic risks of climate change. *Nature Climate Change*, 7(11), 774-782.

- Díaz, J., Sáez, M., Carmona, R., Mirón, I., Barceló, M., Luna, M., & Linares, C. (2019). Mortality attributable to high temperatures over the 2021–2050 and 2051–2100 time horizons in Spain: Adaptation and economic estimate. *Environmental research*, 172, 475-485.
- Doyon, B., Bélanger, D., & Gosselin, P. (2006). *Effets du climat sur la mortalité au Québec méridional de 1981 à 1999 et simulations pour des scénarios climatiques futurs*. Direction risques biologiques, environnementaux et occupationnels, Institut national de la santé publique du Québec.
- Doyon, B., Bélanger, D., & Gosselin, P. (2008). The potential impact of climate change on annual and seasonal mortality for three cities in Québec, Canada. *International journal of health geographics*, 7(1), 1-12.
- Ebi, K. L., Boyer, C., Bowen, K. J., Frumkin, H., & Hess, J. (2018). Monitoring and evaluation indicators for climate change-related health impacts, risks, adaptation, and resilience. *International journal of environmental research and public health*, 15(9), 1943.
- Ebi, K. L., Capon, A., Berry, P., Broderick, C., de Dear, R., Havenith, G., Honda, Y., Kovats, R. S., Ma, W., & Malik, A. (2021). Hot weather and heat extremes: Health risks. *The lancet*, 398(10301), 698-708.
- Ebi, K. L., & Schmier, J. K. (2005). A stitch in time: Improving public health early warning systems for extreme weather events. *Epidemiologic reviews*, 27(1), 115-121.
- Ebi, K. L., Teisberg, T. J., Kalkstein, L. S., Robinson, L., & Weiher, R. F. (2004). Heat watch/warning systems save lives: Estimated costs and benefits for Philadelphia 1995–98. *Bulletin of the American Meteorological Society*, 85(8), 1067-1074.
- Fisher, S., & Rosella, L. C. (2022). Priorities for successful use of artificial intelligence by public health organizations: A literature review. *BMC Public Health*, 22(1), 2146.
- Fouillet, A., Rey, G., Laurent, F., Pavillon, G., Bellec, S., Guihenneuc-Jouyaux, C., Clavel, J., Jougl, E., & Hémon, D. (2006). Excess mortality related to the August 2003 heat wave in France. *International archives of occupational and environmental health*, 80(1), 16-24.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 29(5), 1189-1232.
- García-León, D., Casanueva, A., Standardi, G., Burgstall, A., Flouris, A. D., & Nybo, L. (2021). Current and projected regional economic impacts of heatwaves in Europe. *Nature communications*, 12(1), 5807.
- Gasparri, A. (2011). Distributed lag linear and non-linear models in R: the package dlnm. *Journal of statistical software*, 43, 1-20.
- Gasparri, A. (2021). The case time series design. *Epidemiology (Cambridge, Mass.)*, 32(6), 829.
- Gasparri, A., & Armstrong, B. (2013). Reducing and meta-analysing estimates from distributed lag non-linear models. *BMC medical research methodology*, 13(1), 1-10.

- Gasparrini, A., Armstrong, B., & Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in medicine*, 29(21), 2224-2234.
- Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., & Forsberg, B. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The lancet*, 386(9991), 369-375.
- Gasparrini, A., Guo, Y., Sera, F., Vicedo-Cabrera, A. M., Huber, V., Tong, S., Coelho, M. de S. Z. S., Saldiva, P. H. N., Lavigne, E., & Correa, P. M. (2017). Projections of temperature-related excess mortality under climate change scenarios. *The Lancet Planetary Health*, 1(9), e360-e367.
- Gasparrini, A., & Leone, M. (2014). Attributable risk from distributed lag models. *BMC medical research methodology*, 14(1), 1-8.
- Gasparrini, A., Masselot, P., Scortichini, M., Schneider, R., Mistry, M. N., Sera, F., Macintyre, H. L., Phalkey, R., & Vicedo-Cabrera, A. M. (2022). Small-area assessment of temperature-related mortality risks in England and Wales: A case time series analysis. *The Lancet Planetary Health*, 6(7), e557-e564.
- Goldberg, M. S., Gasparrini, A., Armstrong, B., & Valois, M.-F. (2011). The short-term influence of temperature on daily mortality in the temperate climate of Montreal, Canada. *Environmental research*, 111(6), 853-860.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Gosling, S. N., Lowe, J. A., McGregor, G. R., Pelling, M., & Malamud, B. D. (2009). Associations between elevated atmospheric temperature and human mortality: A critical review of the literature. *Climatic change*, 92(3), 299-341.
- Gouvernement of Canada. (2024). Labour market information: Registered Nurse in Québec. <https://www.jobbank.gc.ca/marketreport/wages-occupation/993/QC>
- Graffy, P. M., Sunderraj, A., Visa, M. A., Miller, C. H., Barrett, B. W., Rao, S., Camilleri, S. F., Harp, R. D., Li, C., & Brenneman, A. (2024). Methodological approaches for measuring the association between heat exposure and health outcomes: A comprehensive global scoping review. *GeoHealth*, 8(9), e2024GH001071.
- Greenwell, B., Boehmke, B., & Cunningham, J. (2019). Package 'gbm'. *R package version*, 2(5).
- Gronlund, C. J., Cameron, L., Shea, C., & O'Neill, M. S. (2019). Assessing the magnitude and uncertainties of the burden of selected diseases attributable to extreme heat and extreme precipitation under a climate change scenario in Michigan for the period 2041–2070. *Environmental Health*, 18(1), 1-17.
- Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.

- Guo, Y., Gasparrini, A., Armstrong, B., Li, S., Tawatsupa, B., Tobias, A., Lavigne, E., Coelho, M. de S. Z. S., Leone, M., & Pan, X. (2014). Global variation in the effects of ambient temperature on mortality: A systematic evaluation. *Epidemiology*, 25(6), 781-789.
- Guo, Y., Gasparrini, A., Li, S., Sera, F., Vicedo-Cabrera, A. M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P. H. N., Lavigne, E., Tawatsupa, B., & Punnasiri, K. (2018). Quantifying excess deaths related to heatwaves under climate change scenarios: A multicountry time series modelling study. *PLoS medicine*, 15(7), e1002629.
- Haines, A., Kovats, R. S., Campbell-Lendrum, D., & Corvalán, C. (2006). Climate change and human health: Impacts, vulnerability and public health. *Public health*, 120(7), 585-596.
- Hastie, T., & Tibshirani, R. (2017). *Generalized additive models*. Routledge.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.
- He, S., Kosatsky, T., Smargiassi, A., Bilodeau-Bertrand, M., & Auger, N. (2018). Heat and pregnancy-related emergencies: Risk of placental abruption during hot weather. *Environment international*, 111, 295-300.
- Health Canada. (2021). Health Impacts of Air Pollution in Canada—Estimates of premature deaths and nonfatal outcomes. <https://www.canada.ca/en/health-canada/services/publications/healthy-living/health-impacts-air-pollution-2021.html>
- Health Canada. (2022). Health of Canadians in a Changing Climate: Advancing our Knowledge for Action. <https://changingclimate.ca/health-in-a-changing-climate/>
- Hebborn, C., Gosselin, P., Chen, K., Chen, H., Cakmak, S., MacDonald, M., Chagnon, J., Dion, P., Martel, L., & Lavigne, E. (2023). Future temperature-related excess mortality under climate change and population aging scenarios in Canada. *Canadian Journal of Public Health*, 114(5), 726-736.
- Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013). A trend-preserving bias correction—the ISI-MIP approach. *Earth System Dynamics*, 4(2), 219-236.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Hondula, D. M., Balling, R. C., Vanos, J. K., & Georgescu, M. (2015). Rising temperatures, human health, and the role of adaptation. *Current Climate Change Reports*, 1(3), 144-154.
- Hu, W., Mengersen, K., McMichael, A., & Tong, S. (2008). Temperature, air pollution and total mortality during summers in Sydney, 1994–2004. *International journal of biometeorology*, 52, 689-696.
- Huang, C., Barnett, A. G., Wang, X., Vaneckova, P., FitzGerald, G., & Tong, S. (2011). Projecting future heat-related mortality under climate change scenarios: A systematic review. *Environmental health perspectives*, 119(12), 1681-1690.

- Huang, C., Barnett, A. G., Xu, Z., Chu, C., Wang, X., Turner, L. R., & Tong, S. (2013). Managing the health effects of temperature in response to climate change: Challenges ahead. *Environmental health perspectives*, 121(4), 415-419.
- Hübler, M., Klepper, G., & Peterson, S. (2008). Costs of climate change: The effects of rising temperatures on health and productivity in Germany. *Ecological Economics*, 68(1-2), 381-393.
- Hunt, A., Ferguson, J., Baccini, M., Watkiss, P., & Kendrovski, V. (2017). Climate and weather service provision: Economic appraisal of adaptation to health impacts. *Climate services*, 7, 78-86.
- Ikeda, T., & Kusaka, H. (2021). Development of models for predicting the number of patients with heatstroke on the next day considering heat acclimatization. *Journal of the Meteorological Society of Japan. Ser. II*, 99(6), 1395-1412.
- INSPQ. (2024). Index of material and social deprivation compiled by the Bureau d'information et d'études en santé des populations (BIESP) from 1991, 1996, 2001, 2006, 2011, 2016 and 2021 Canadian Census data. *Institut national de santé publique du Québec*. <https://www.inspq.qc.ca/en/deprivation/material-and-social-deprivation-index>
- IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Intergovernmental Panel on Climate Change*. <https://www.ipcc.ch/report/ar6/wg1/>
- Ishigami, A., Hajat, S., Kovats, R. S., Bisanti, L., Rognoni, M., Russo, A., & Paldy, A. (2008). An ecological time-series study of heat-related mortality in three European cities. *Environmental Health*, 7(1), 1-7.
- ISQ. (2024). Rémunération horaire. *Institut de la Statistique du Québec*. <https://statistique.quebec.ca/vitrine/egalite/dimensions-egalite/revenu/remuneration-horaire>
- Jagai, J. S., Grossman, E., Navon, L., Sambanis, A., & Dorevitch, S. (2017). Hospitalizations for heat-stress illness varies between rural and urban areas: An analysis of Illinois data, 1987–2014. *Environmental Health*, 16(1), 1-10.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
- Jeong, D. I., Sushama, L., Diro, G. T., Khaliq, M. N., Beltrami, H., & Caya, D. (2016). Projected changes to high temperature events for Canada based on a regional climate model ensemble. *Climate Dynamics*, 46(9), 3163-3180.
- Jian, L., Patel, D., Xiao, J., Jansz, J., Yun, G., Lin, T., & Robertson, A. (2023). Can we use a machine learning approach to predict the impact of heatwaves on emergency department attendance? *Environmental Research Communications*, 5(4), 045005.
- Jiang, S., Warren, J. L., Scovronick, N., Moss, S. E., Darrow, L. A., Strickland, M. J., Newman, A. J., Chen, Y., Ebelt, S. T., & Chang, H. H. (2021). Using logic regression to characterize extreme heat exposures and their health associations: A time-series study of emergency department visits in Atlanta. *BMC Medical Research Methodology*, 21(1), 87.

- Kanti, F. S., Alari, A., Chaix, B., & Benmarhnia, T. (2022). Comparison of various heat waves definitions and the burden of heat-related mortality in France: Implications for existing early warning systems. *Environmental research*, 215, 114359.
- Karlsson, M., & Ziebarth, N. R. (2018). Population health effects and health-related costs of extreme temperatures: Comprehensive evidence from Germany. *Journal of Environmental Economics and Management*, 91, 93-117.
- Kassomenos, P., Petrakis, M., Sarigiannis, D., Gotti, A., & Karakitsios, S. (2011). Identifying the contribution of physical and chemical stressors to the daily number of hospital admissions implementing an artificial neural network model. *Air Quality, Atmosphere & Health*, 4(3), 263-272.
- Ke, D., Takahashi, K., Takakura, J., Takara, K., & Kamranzad, B. (2023). Effects of heatwave features on machine-learning-based heat-related ambulance calls prediction models in Japan. *Science of the total environment*, 873, 162283.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Kegel, F., Luo, O. D., & Richer, S. (2021). The impact of extreme heat events on emergency departments in Canadian hospitals. *Wilderness & environmental medicine*, 32(4), 433-440.
- Kent, S. T., McClure, L. A., Zaitchik, B. F., Smith, T. T., & Gohlke, J. M. (2014). Heat waves and health outcomes in Alabama (USA): The importance of heat wave definition. *Environmental health perspectives*, 122(2), 151-158.
- Khatri, K. L., & Tamil, L. S. (2017). Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. *IEEE journal of biomedical and health informatics*, 22(1), 285-290.
- Kim, Y., & Kim, Y. (2022). Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustainable Cities and Society*, 79, 103677.
- Knowlton, K., Rotkin-Ellman, M., Geballe, L., Max, W., & 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., & English, P. (2009). The 2006 California heat wave: Impacts on hospitalizations and emergency department visits. *Environmental health perspectives*, 117(1), 61-67.
- Kolb, S., Radon, K., Valois, M.-F., Héguy, L., & Goldberg, M. S. (2007). The short-term influence of weather on daily mortality in congestive heart failure. *Archives of environmental & occupational health*, 62(4), 169-176.

- Kotharkar, R., Dongarsane, P., Ghosh, A., & Kotharkar, V. (2024). Numerical analysis of extreme heat in Nagpur city using heat stress indices, all-cause mortality and local climate zone classification. *Sustainable Cities and Society*, 101, 105099.
- Kotharkar, R., & Ghosh, A. (2022). Progress in extreme heat management and warning systems : A systematic review of heat-health action plans (1995-2020). *Sustainable Cities and Society*, 76, 103487.
- Kousky, C. (2014). Informing climate adaptation: A review of the economic costs of natural disasters. *Energy economics*, 46, 576-592.
- Kovats, R. S., & Hajat, S. (2008). Heat stress and public health: A critical review. *Annu. Rev. Public Health*, 29, 41-55.
- Lamothe, F., Dubé, M., Boudreault, J., & Bustinza, R. (2025). Vagues de chaleur extrême et chaleur estivale au Québec : Impacts sur la santé de 2010 à 2024. *Institut national de santé publique du Québec*. <https://www.inspq.qc.ca/sites/default/files/publications/3662-vagues-chaleur-extreme-estivale-QC-2010-2024.pdf>
- Lamothe, F., Dubé, M., & Bustinza, R. (2023). Surveillance des impacts sanitaires des vagues de chaleur extrême au Québec – Bilan de la saison estivale 2021. *Institut national de santé publique du Québec*. <https://www.inspq.qc.ca/sites/default/files/2023-12/3396-impacts-sanitaires-vagues-chaleur-extreme-2021.pdf>
- Lapointe, L. (2021). Chief coroner's statement on public safety during high temperatures. https://www2.gov.bc.ca/assets/gov/birth-adoption-death-marriage-and-divorce/deaths/coroners-service/news/2021/chief_coroner_statement_-_heat_related_deaths.pdf.
- Larivée, C., Sinclair-Desgagné, N., Da Silva, L., Revéret, J., & Desjarlais, C. (2015). Évaluation des impacts des changements climatiques et de leurs coûts pour le Québec et l'État québécois. *Ouranos*. <https://www.environnement.gouv.qc.ca/changementsclimatiques/evaluation-impacts-cc-couts-qc-etat.pdf>
- Lavigne, E., Gasparrini, A., Wang, X., Chen, H., Yagouti, A., Fleury, M. D., & Cakmak, S. (2014). Extreme ambient temperatures and cardiorespiratory emergency room visits: Assessing risk by comorbid health conditions in a time series study. *Environmental health*, 13(1), 1-8.
- Lavigne, E., Maltby, A., Côté, J.-N., Weinberger, K. R., Hebborn, C., Vicedo-Cabrera, A. M., & Wilk, P. (2023). The effect modification of extreme temperatures on mental and behavior disorders by environmental factors and individual-level characteristics in Canada. *Environmental research*, 219, 114999.
- Lavoie, J., Bourgault, P., Smith, T. J., Logan, T., Leduc, M., Caron, L.-P., Gammon, S., & Braun, M. (2024a). An ensemble of bias-adjusted CMIP6 climate simulations based on a high-resolution North American reanalysis. *Scientific Data*, 11(1), 64.
- Lavoie, J., Caron, L.-P., Logan, T., & Barrow, E. (2024b). Canadian climate data portals: A comparative analysis from a user perspective. *Climate Services*, 34, 100471.

- Lay, C., Mills, D., Belova, A., Sarofim, M., Kinney, P., Vaidyanathan, A., Jones, R., Hall, R., & Saha, S. (2018). Emergency department visits and ambient temperature: Evaluating the connection and projecting future outcomes. *GeoHealth*, 2(6), 182-194.
- Lebel, G., Bustinza, R., & Dubé, M. (2017). Analyse des impacts des vagues régionales de chaleur extrême sur la santé au Québec de 2010 à 2015. *Institut national de la santé publique du Québec*. https://www.inspq.qc.ca/sites/default/files/publications/2221_impacts_vagues_chaleur_extreme_sante.pdf
- Lebel, G., Dubé, M., & Bustinza, R. (2019). Surveillance des impacts des vagues de chaleur extrême sur la santé au Québec à l'été 2018. *Bulletin d'information en santé environnementale*, 1, 1-10.
- Lee, J. Y., Kim, H., Gasparrini, A., Armstrong, B., Bell, M. L., Sera, F., Lavigne, E., Abrutzky, R., Tong, S., & Coelho, M. de S. Z. S. (2019). Predicted temperature-increase-induced global health burden and its regional variability. *Environment international*, 131, 105027.
- Lee, W., Kim, Y., Sera, F., Gasparrini, A., Park, R., Choi, H. M., Prifti, K., Bell, M. L., Abrutzky, R., & Guo, Y. (2020). Projections of excess mortality related to diurnal temperature range under climate change scenarios: A multi-country modelling study. *The Lancet Planetary Health*, 4(11), e512-e521.
- Lee, W., Lim, Y.-H., Ha, E., Kim, Y., & Lee, W. K. (2022). Forecasting of non-accidental, cardiovascular, and respiratory mortality with environmental exposures adopting machine learning approaches. *Environmental Science and Pollution Research*, 29(58), 88318-88329.
- Li, F., Yigitcanlar, T., Nepal, M., Nguyen, K., & Dur, F. (2023). Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework. *Sustainable Cities and Society*, 96, 104653.
- Li, M., Gu, S., Bi, P., Yang, J., & Liu, Q. (2015). Heat waves and morbidity: Current knowledge and further direction-a comprehensive literature review. *International journal of environmental research and public health*, 12(5), 5256-5283.
- Li, H., & Tang, Q. (2022). Joint extremes in temperature and mortality: A bivariate POT approach. *North American Actuarial Journal*, 26(1), 43-63.
- Li, Y.-H., Ye, D.-X., Liu, Y., Li, N., Meng, C.-S., Wang, Y., Wang, Y., Jin, X., Bi, P., & Tong, S. (2021). Association of heat exposure and emergency ambulance calls: A multi-city study. *Advances in Climate Change Research*, 12(5), 619-627.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Limaye, V. S., Max, W., Constible, J., & Knowlton, K. (2019). Estimating the health-related costs of 10 climate-sensitive US events during 2012. *GeoHealth*, 3(9), 245-265.
- Lin, S., Hsu, W.-H., Van Zutphen, A. R., Saha, S., Lubet, G., & Hwang, S.-A. (2012). Excessive heat and respiratory hospitalizations in New York State: Estimating current and future public health burden related to climate change. *Environmental health perspectives*, 120(11), 1571-1577.

- Lin, X., Tian, T., Shi, C., Wang, P., Chen, S., Guo, T., Li, Z., Liang, B., Zhang, W., & Qin, P. (2023). What are the individual and joint impacts of key meteorological factors on the risk of unintentional injuries? A case-crossover study of over 147,800 cases from a sentinel-based surveillance system. *Sustainable Cities and Society*, 91, 104413.
- Lin, Y.-C., Tsai, C.-H., Hsu, H.-T., & Lin, C.-H. (2021). Using Machine Learning to Analyze and Predict the Relations Between Cardiovascular Disease Incidence, Extreme Temperature and Air Pollution. *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability*, 234-237.
- Linden, A., Yarnold, P. R., & Nallamothu, B. K. (2016). Using machine learning to model dose–response relationships. *Journal of Evaluation in Clinical Practice*, 22(6), 860-867.
- Liu, J., Varghese, B. M., Hansen, A., Zhang, Y., Driscoll, T., Morgan, G., Dear, K., Gourley, M., Capon, A., & Bi, P. (2022). Heat exposure and cardiovascular health outcomes: A systematic review and meta-analysis. *The Lancet Planetary Health*, 6(6), e484-e495.
- Liu, Y., Saha, S., Hoppe, B. O., & Convertino, M. (2019). Degrees and dollars—Health costs associated with suboptimal ambient temperature exposure. *Science of The Total Environment*, 678, 702-711.
- Lu, X., & Qiu, H. (2023). Explainable prediction of daily hospitalizations for cerebrovascular disease using stacked ensemble learning. *BMC Medical Informatics and Decision Making*, 23(1), 1-13.
- Luber, G., & McGeehin, M. (2008). Climate change and extreme heat events. *American journal of preventive medicine*, 35(5), 429-435.
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: A review. *Frontiers in public health*, 8, 14.
- Marien, L., Valizadeh, M., zu Castell, W., Nam, C., Rechid, D., Schneider, A., Meisinger, C., Linseisen, J., Wolf, K., & Bouwer, L. (2022). Machine learning models to predict myocardial infarctions from past climatic and environmental conditions. *Natural Hazards and Earth System Sciences*, 22(9), 3015-3039.
- Martiello, M. A., & Giacchi, M. V. (2010). High temperatures and health outcomes: A review of the literature. *Scandinavian journal of public health*, 38(8), 826-837.
- Martin, S. L., Cakmak, S., Hebborn, C. A., Avramescu, M.-L., & Tremblay, N. (2012). Climate change and future temperature-related mortality in 15 Canadian cities. *International journal of biometeorology*, 56(4), 605-619.
- Masselot, P., Chebana, F., Bélanger, D., St-Hilaire, A., Abdous, B., Gosselin, P., & Ouarda, T. B. (2018). Aggregating the response in time series regression models, applied to weather-related cardiovascular mortality. *Science of The Total Environment*, 628, 217-225.
- Masselot, P., Chebana, F., Campagna, C., Lavigne, É., Ouarda, T. B., & Gosselin, P. (2021). Machine learning approaches to identify thresholds in a heat-health warning system context. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 184(4), 1326-1346.

- Masselot, P., Chebana, F., Lavigne, É., Campagna, C., Gosselin, P., & Ouarda, T. B. (2019). Toward an improved air pollution warning system in Quebec. *International journal of environmental research and public health*, 16(12), 2095.
- Masselot, P., Chebana, F., Ouarda, T. B., Bélanger, D., & Gosselin, P. (2022). Data-enhancement strategies in weather-related health studies. *International Journal of Environmental Research and Public Health*, 19(2), 906.
- May, R. M., Goebbert, K. H., Thielen, J. E., Leeman, J. R., Camron, M. D., Bruick, Z., Bruning, E. C., Manser, R. P., Arms, S. C., & Marsh, P. T. (2022). MetPy: A meteorological Python library for data analysis and visualization. *Bulletin of the American Meteorological Society*, 103(10), E2273-E2284.
- McGregor, G. R., Bessmoulin, P., Ebi, K., & Menne, B. (2015). *Heatwaves and health: Guidance on warning-system development*. World Meteorological Organization and World Health Organization.
- McLean, K. E., Lee, M. J., Coker, E. S., & Henderson, S. B. (2024). A population-based case-control analysis of risk factors associated with mortality during the 2021 western North American heat dome: Focus on chronic conditions and social vulnerability. *Environmental Research: Health*, 2(3), 035010.
- Meehl, G. A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, 305(5686), 994-997.
- Merrill, C. T., Miller, M., & Steiner, C. (2008). *Hospital Stays Resulting from Excessive Heat and Cold Exposure Due to Weather Conditions in US Community Hospitals, 2005*. HCUP Statistical Brief #55.
- Mora, C., Dousset, B., Caldwell, I. R., Powell, F. E., Geronimo, R. C., Bielecki, C. R., Counsell, C. W., Dietrich, B. S., Johnston, E. T., & Louis, L. V. (2017). Global risk of deadly heat. *Nature climate change*, 7(7), 501-506.
- Morgenstern, J. D., Buajitti, E., O'Neill, M., Piggott, T., Goel, V., Fridman, D., Kornas, K., & Rosella, L. C. (2020). Predicting population health with machine learning: A scoping review. *BMJ open*, 10(10), e037860.
- MSSS. (2022). Estimations et projections de population par territoire sociosanitaire. *Ministère de la Santé et des Services sociaux du Québec*. <https://publications.msss.gouv.qc.ca/msss/document-001617/>
- Navares, R., Díaz, J., Linares, C., & Aznarte, J. L. (2018). Comparing ARIMA and computational intelligence methods to forecast daily hospital admissions due to circulatory and respiratory causes in Madrid. *Stochastic environmental research and risk assessment*, 32, 2849-2859.
- Navas-Martín, M. Á., Ovalle-Perandones, M.-A., López-Bueno, J. A., Díaz, J., Linares, C., & Sánchez-Martínez, G. (2024). Population adaptation to heat as seen through the temperature-mortality relationship, in the context of the impact of global warming on health: A scoping review. *Science of the Total Environment*, 908, 168441.

- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370-384.
- Nishimura, T., Rashed, E. A., Kodera, S., Shirakami, H., Kawaguchi, R., Watanabe, K., Nemoto, M., & Hirata, A. (2021). Social implementation and intervention with estimated morbidity of heat-related illnesses from weather data: A case study from Nagoya City, Japan. *Sustainable Cities and Society*, 74, 103203.
- Noe, R. S., Jin, J. O., & Wolkin, A. F. (2012). Exposure to natural cold and heat: Hypothermia and hyperthermia Medicare claims, United States, 2004–2005. *American journal of public health*, 102(4), e11-e18.
- NRTEE. (2011). Paying the price: The economic impacts of climate change for Canada. *National Round Table on the Environment and the Economy*. <https://publications.gc.ca/site/eng/9.571245/publication.html>
- Ogata, S., Takegami, M., Ozaki, T., Nakashima, T., Onozuka, D., Murata, S., Nakaoku, Y., Suzuki, K., Hagihara, A., & Noguchi, T. (2021). Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. *Nature communications*, 12(1), 1-11.
- Ohashi, Y., Ihara, T., Oka, K., Takane, Y., & Kikegawa, Y. (2023). Machine learning analysis and risk prediction of weather-sensitive mortality related to cardiovascular disease during summer in Tokyo, Japan. *Scientific reports*, 13, 17020.
- Oka, K., He, J., Honda, Y., & Hijioka, Y. (2024). Random forest analysis of the relative importance of meteorological indicators for heatstroke cases in Japan based on the degree of severity and place of occurrence. *Environmental Research*, 263, 120066.
- Oka, K., He, J., Honda, Y., & Hijioka, Y. (2025). Economic burden of ambulance transports for patients with heatstroke under climate change: Comparison of the effects of gradual temperature increases and extreme heat events. *Environmental Research Communications*, 7(4), 041009.
- Oka, K., & Hijioka, Y. (2021). Prediction of the number of heatstroke patients transported by ambulance in Japan's 47 prefectures: Proposal of heat acclimatization consideration. *Environmental Research Communications*, 3(12), 125002.
- Orlov, A., Schleyden, J., Aunan, K., Sillmann, J., Gasparri, A., & Mistry, M. N. (2024). A better integration of health and economic impact assessments of climate change. *Environmental research letters*. 19(3), 031004.
- Pan, R., Yao, Z., Yi, W., Wei, Q., He, Y., Tang, C., Liu, X., Son, S., Ji, Y., & Song, J. (2022). Temporal trends of the association between temperature variation and hospitalizations for schizophrenia in Hefei, China from 2005 to 2019: A time-varying distribution lag nonlinear model. *Environmental Science and Pollution Research*, 29(4), 5184-5193.
- Park, J., & Kim, J. (2018). Defining heatwave thresholds using an inductive machine learning approach. *Plos one*, 13(11), e0206872.

- Park, M., Jung, D., Lee, S., & Park, S. (2020). Heatwave Damage Prediction Using Random Forest Model in Korea. *Applied Sciences*, 10(22), 8237.
- Pascal, M., Laaidi, K., Ledrans, M., Baffert, E., Caserio-Schönemann, C., Le Tertre, A., Manach, J., Medina, S., Rudant, J., & Empereur-Bissonnet, P. (2006). France's heat health watch warning system. *International Journal of Biometeorology*, 50(3), 144-153.
- Pascal, M., Gorla, S., Wagner, V., Sabastia, M., Guillet, A., Cordeau, E., Maucclair, C., & Host, S. (2021a). Greening is a promising but likely insufficient adaptation strategy to limit the health impacts of extreme heat. *Environment international*, 151, 106441.
- Pascal, M., Wagner, V., Alari, A., Corso, M., & Le Tertre, A. (2021b). Extreme heat and acute air pollution episodes: A need for joint public health warnings? *Atmospheric Environment*, 249, 118249.
- Pascal, M., Wagner, V., Corso, M., Laaidi, K., Ung, A., & Beaudreau, P. (2018). Heat and cold related-mortality in 18 French cities. *Environment international*, 121, 189-198.
- Pascal, M., Wagner, V., Le Tertre, A., Laaidi, K., Honoré, C., Bénichou, F., & Beaudreau, P. (2013). Definition of temperature thresholds: The example of the French heat wave warning system. *International Journal of Biometeorology*, 57(1), 21-29.
- Paudel, U., Adhikari, S. R., & Pant, K. P. (2020). Economics of environmental effects on health: A methodological review based on epidemiological information. *Environmental and Sustainability Indicators*, 5, 100020.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine learning research*, 12, 2825-2830.
- Phung, D., Thai, P. K., Guo, Y., Morawska, L., Rutherford, S., & Chu, C. (2016). Ambient temperature and risk of cardiovascular hospitalization: An updated systematic review and meta-analysis. *Science of The Total Environment*, 550, 1084-1102.
- Qiu, H., Luo, L., Su, Z., Zhou, L., Wang, L., & Chen, Y. (2020). Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. *BMC Medical Informatics and Decision Making*, 20(1), 1-11.
- Québec. (2024). *Cost of ambulance transportation*. Gouvernement du Québec. <https://www.quebec.ca/en/health/health-system-and-services/pre-hospital-emergency-services/cost-ambulance-transportation>
- Québec. (2024b). *Info-Santé 811*. Gouvernement du Québec. <https://www.quebec.ca/en/health/finding-a-resource/info-sante-811>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.

- Rasheed, K., Qayyum, A., Ghaly, M., Al-Fuqaha, A., Razi, A., & Qadir, J. (2022). Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Computers in Biology and Medicine*, 149, 106043.
- Reid, C. E., Snowden, J. M., Kontgis, C., & Tager, I. B. (2012). The role of ambient ozone in epidemiologic studies of heat-related mortality. *Environmental health perspectives*, 120(12), 1627-1630.
- Ripoche, M., Irace-Cima, A., Adam-Poupart, A., Baron, G., Bouchard, C., Carignan, A., Milord, F., Ouhoummane, N., Pilon, P. A., & Thivierge, K. (2023). Current and future burden from Lyme disease in Québec as a result of climate change. *Canada Communicable Disease Report*, 49(10), 446.
- Robben, J., Antonio, K., & Kleinow, T. (2025). The short-term association between environmental variables and mortality: evidence from Europe. *Journal of the Royal Statistical Society Series A: Statistics in Society*, qnaf052.
- Romanello, M., Di Napoli, C., Green, C., Kennard, H., Lampard, P., Scamman, D., Walawender, M., Ali, Z., Ameli, N., & Ayeb-Karlsson, S. (2023). The 2023 report of the Lancet Countdown on health and climate change: The imperative for a health-centred response in a world facing irreversible harms. *The Lancet*, 402(10419), 2346-2394.
- Sanderson, M., Arbuthnott, K., Kovats, S., Hajat, S., & Falloon, P. (2017). The use of climate information to estimate future mortality from high ambient temperature: A systematic literature review. *PLOS one*, 12(7), e0180369.
- Schmeltz, M. T., Petkova, E. P., & Gamble, J. L. (2016). Economic burden of hospitalizations for heat-related illnesses in the United States, 2001–2010. *International journal of environmental research and public health*, 13(9), 894.
- Schmidt, C. W. (2020). Into the Black Box: What Can Machine Learning Offer Environmental Health Research? *Environmental Health Perspectives*, 128(2), 022001.
- Schmitt, L. H., Graham, H. M., & White, P. C. (2016). Economic evaluations of the health impacts of weather-related extreme events: A scoping review. *International journal of environmental research and public health*, 13(11), 1105.
- Schwartz, J., Samet, J. M., & Patz, J. A. (2004). Hospital admissions for heart disease: The effects of temperature and humidity. *Epidemiology*, 15(6), 755-761.
- Seklecka, M., Pantelous, A. A., & O'Hare, C. (2017). Mortality effects of temperature changes in the United Kingdom. *Journal of Forecasting*, 36(7), 824-841.
- Seong, K., Jiao, J., Mandalapu, A., & Niyogi, D. (2024). Spatio-temporal patterns of heat index and heat-related Emergency Medical Services (EMS). *Sustainable Cities and Society*, 111, 105562.
- Sera, F., Armstrong, B., Blangiardo, M., & Gasparrini, A. (2019). An extended mixed-effects framework for meta-analysis. *Statistics in Medicine*, 38(29), 5429-5444.

- Shaamala, A., Yigitcanlar, T., Nili, A., & Nyandega, D. (2024). Algorithmic green infrastructure optimisation: Review of artificial intelligence driven approaches for tackling climate change. *Sustainable Cities and Society*, 105182.
- Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84-90.
- Smargiassi, A., Goldberg, M. S., Plante, C., Fournier, M., Baudouin, Y., & Kosatsky, T. (2009). Variation of daily warm season mortality as a function of micro-urban heat islands. *Journal of Epidemiology & Community Health*, 63(8), 659-664.
- Smargiassi, A., Plante, C., Morency, P., Hatzopoulou, M., Morency, C., Eluru, N., Tétreault, L.-F., Goudreau, S., Bourbonnais, P. L., & Bhowmik, T. (2020). Environmental and health impacts of transportation and land use scenarios in 2061. *Environmental research*, 187, 109622.
- Smoyer-Tomic, K. E., & Rainham, D. G. (2001). Beating the heat: Development and evaluation of a Canadian hot weather health-response plan. *Environmental health perspectives*, 109(12), 1241-1248.
- Son, J.-Y., Liu, J. C., & Bell, M. L. (2019). Temperature-related mortality: A systematic review and investigation of effect modifiers. *Environmental Research Letters*, 14(7), 073004.
- Song, X., Wang, S., Hu, Y., Yue, M., Zhang, T., Liu, Y., Tian, J., & Shang, K. (2017). Impact of ambient temperature on morbidity and mortality: An overview of reviews. *Science of the Total Environment*, 586, 241-254.
- Ssebyala, S. N., Kintu, T. M., Muganzi, D. J., Dresser, C., Demetres, M. R., Lai, Y., Mercy, K., Li, C., Wang, F., & Setoguchi, S. (2024). Use of machine learning tools to predict health risks from climate-sensitive extreme weather events: A scoping review. *PLOS Climate*, 3(1), e0000338.
- Stafoggia, M., Michelozzi, P., Schneider, A., Armstrong, B., Scortichini, M., Rai, M., Achilleos, S., Alahmad, B., Analitis, A., & Åström, C. (2023). Joint effect of heat and air pollution on mortality in 620 cities of 36 countries. *Environment international*, 181, 108258.
- Statistics Canada. (2022). *Census metropolitan area (CMA) and census agglomeration (CA)*. <https://www150.statcan.gc.ca/n1/pub/92-195-x/2021001/geo/cma-rmr/cma-rmr-eng.htm>
- Steenland, K., & Armstrong, B. (2006). An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology*, 512-519.
- Thompson, R., Kovats, S., Hajat, S., Macintyre, H., & O'Connell, E. (2024). Identification of individual-level clinical factors associated with increased risk of death during heatwaves: A time-stratified case-crossover study using national primary care records in England. *BMJ Public Health*, 2(1).
- Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S., & Wilson, B. E. (2022). Daymet: Daily surface weather data on a 1-km grid for North America, Version 4. *ORNL DAAC, Oak Ridge, Tennessee, USA*.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.
- Toloo, G. S., Hu, W., FitzGerald, G., Aitken, P., & Tong, S. (2015). Projecting excess emergency department visits and associated costs in Brisbane, Australia, under population growth and climate change scenarios. *Scientific reports*, 5(1), 1-9.
- Tong, M. X., Wondmagegn, B. Y., Williams, S., Hansen, A., Keith, D., Pisaniello, D., Xiang, J., Jianguo, X., Le, J., & Scalley, B. (2021a). Hospital healthcare costs attributable to heat and future estimations in the context of climate change in Perth, Western Australia. *Advances in Climate Change Research*, 12(5), 638-648.
- Tong, M. X., Wondmagegn, B. Y., Xiang, J., Williams, S., Hansen, A., Dear, K., Pisaniello, D., Varghese, B. M., Xiao, J., & Jian, L. (2021b). Heat-attributable hospitalisation costs in Sydney: Current estimations and future projections in the context of climate change. *Urban Climate*, 40, 101028.
- Tong, M. X., Wondmagegn, B. Y., Xiang, J., Williams, S., Hansen, A., Dear, K., Pisaniello, D., Xiao, J., Jian, L., & Scalley, B. (2021c). Emergency department visits and associated healthcare costs attributable to increasing temperature in the context of climate change in Perth, Western Australia, 2012-2019. *Environmental Research Letters*, 16(6), 065011.
- Tong, M. X., Wondmagegn, B., Xiang, J., Hansen, A., Dear, K., Pisaniello, D., Varghese, B., Xiao, J., Jian, L., & Scalley, B. (2022). Hospitalization costs of respiratory diseases attributable to temperature in Australia and projections for future costs in the 2030s and 2050s under climate change. *International Journal of Environmental Research and Public Health*, 19(15), 9706.
- Tong, S., & Kan, H. (2011). Heatwaves: What is in a definition? *Maturitas*, 69(1), 5-6.
- Tong, S., Wang, X. Y., & Barnett, A. G. (2010). Assessment of heat-related health impacts in Brisbane, Australia: Comparison of different heatwave definitions. *PloS one*, 5(8), e12155.
- Toutant, S., Gosselin, P., Bélanger, D., Bustinza, R., & Rivest, S. (2011). An open source web application for the surveillance and prevention of the impacts on public health of extreme meteorological events: The SUPREME system. *International journal of health geographics*, 10(1), 1-11.
- Tran, D. T., & Warkentin, L. M. (2025). Healthcare resource use and costs associated with extreme heat in Alberta, Canada. *Discover Health Systems*, 4(1), 1-11.
- Tupinier Martin, F., Boudreault, J., Campagna, C., Lavigne, É., Gamache, P., Tandonnet, M., Généreux, M., Trottier, S., & Goupil-Sormany, I. (2024). The relationship between hot temperatures and hospital admissions for psychosis in adults diagnosed with schizophrenia: A case-crossover study in Quebec, Canada. *Environmental Research*, 246, 118225.
- Usmani, R. S. A., Pillai, T. R., Hashem, I. A. T., Marjani, M., Shaharudin, R., & Latif, M. T. (2021). Air pollution and cardiorespiratory hospitalization, predictive modeling, and analysis using artificial intelligence techniques. *Environmental Science and Pollution Research*, 28(40), 56759-56771.

- Vaneckova, P., Neville, G., Tippet, V., Aitken, P., FitzGerald, G., & Tong, S. (2011). Do biometeorological indices improve modeling outcomes of heat-related mortality? *Journal of Applied Meteorology and Climatology*, 50(6), 1165-1176.
- Vicedo-Cabrera, A. M., Forsberg, B., Tobias, A., Zanobetti, A., Schwartz, J., Armstrong, B., & Gasparrini, A. (2016). Associations of inter-and intraday temperature change with mortality. *American journal of epidemiology*, 183(4), 286-293.
- Vicedo-Cabrera, A. M., Guo, Y., Sera, F., Huber, V., Schleussner, C.-F., Mitchell, D., Tong, S., Coelho, M. de S. Z. S., Saldiva, P. H. N., & Lavigne, E. (2018). Temperature-related mortality impacts under and beyond Paris Agreement climate change scenarios. *Climatic change*, 150, 391-402.
- Vicedo-Cabrera, A. M., Scovronick, N., Sera, F., Royé, D., Schneider, R., Tobias, A., Astrom, C., Guo, Y., Honda, Y., & Hondula, D. M. (2021). The burden of heat-related mortality attributable to recent human-induced climate change. *Nature climate change*, 11(6), 492-500.
- Vicedo-Cabrera, A. M., Sera, F., & Gasparrini, A. (2019). Hands-on tutorial on a modeling framework for projections of climate change impacts on health. *Epidemiology (Cambridge, Mass.)*, 30(3), 321.
- Vida, S., Durocher, M., Ouarda, T. B., & Gosselin, P. (2012). Relationship between ambient temperature and humidity and visits to mental health emergency departments in Québec. *Psychiatric Services*, 63(11), 1150-1153.
- VoPham, T., Hart, J. E., Laden, F., & Chiang, Y.-Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): Potential applications for environmental epidemiology. *Environmental Health*, 17, 1-6.
- Wang, J., Nikolaou, N., an der Heiden, M., & Irgang, C. (2024). High-resolution modeling and projection of heat-related mortality in Germany under climate change. *Communications Medicine*, 4(1), 206.
- Wang, X., Lavigne, E., Ouellette-kuntz, H., & Chen, B. E. (2014). Acute impacts of extreme temperature exposure on emergency room admissions related to mental and behavior disorders in Toronto, Canada. *Journal of affective disorders*, 155, 154-161.
- Wang, Y., Liu, Y., Ye, D., Li, N., Bi, P., Tong, S., Wang, Y., Cheng, Y., Li, Y., & Yao, X. (2021). Temperatures and health costs of emergency department visits: A multisite time series study in China. *Environmental Research*, 197, 111023.
- Wang, Y., Song, Q., Du, Y., Wang, J., Zhou, J., Du, Z., & Li, T. (2019). A random forest model to predict heatstroke occurrence for heatwave in China. *Science of the Total Environment*, 650, 3048-3053.
- Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Beagley, J., Belesova, K., Boykoff, M., Byass, P., Cai, W., & Campbell-Lendrum, D. (2021). The 2020 report of the Lancet Countdown on health and climate change: Responding to converging crises. *The Lancet*, 397(10269), 129-170.

- Wellenius, G. A., Eliot, M. N., Bush, K. F., Holt, D., Lincoln, R. A., Smith, A. E., & Gold, J. (2017). Heat-related morbidity and mortality in New England: Evidence for local policy. *Environmental research*, 156, 845-853.
- Wertis, L., Sugg, M. M., Runkle, J. D., & Rao, D. (2023). Socio-Environmental Determinants of Mental and Behavioral Disorders in Youth: A Machine Learning Approach. *GeoHealth*, 7(9), e2023GH000839.
- Wiemken, T. L., & Kelley, R. R. (2019). Machine Learning in Epidemiology and Health Outcomes Research. *Annual review of public health*, 41, 21-36.
- Williams, S., Nitschke, M., Wondmagegn, B. Y., Tong, M., Xiang, J., Hansen, A., Nairn, J., Karnon, J., & Bi, P. (2021). Evaluating cost benefits from a heat health warning system in Adelaide, South Australia. *Australian and New Zealand Journal of Public Health*, 46(2), 149-154.
- Wlodarczyk, A., Molek, P., Bochenek, B., Wypych, A., Nessler, J., & Zalewski, J. (2022). Machine learning analyzed weather conditions as an effective means in the predicting of acute coronary syndrome prevalence. *Frontiers in Cardiovascular Medicine*, 9, 830823.
- Wondmagegn, B. Y., Xiang, J., Dear, K., Williams, S., Hansen, A., Pisaniello, D., Nitschke, M., Nairn, J., Scalley, B., & Varghese, B. M. (2021a). Impact of heatwave intensity using excess heat factor on emergency department presentations and related healthcare costs in Adelaide, South Australia. *Science of The Total Environment*, 781, 146815.
- Wondmagegn, B. Y., Xiang, J., Dear, K., Williams, S., Hansen, A., Pisaniello, D., Nitschke, M., Nairn, J., Scalley, B., & Xiao, A. (2021b). Increasing impacts of temperature on hospital admissions, length of stay, and related healthcare costs in the context of climate change in Adelaide, South Australia. *Science of The Total Environment*, 773, 145656.
- Wondmagegn, B. Y., Xiang, J., Williams, S., Pisaniello, D., & Bi, P. (2019). What do we know about the healthcare costs of extreme heat exposure? A comprehensive literature review. *Science of the total environment*, 657, 608-618.
- Wood, S. (2015). Package 'mgcv'. *R package version*, 1(29), 729.
- Wu, Y., Wen, B., Ye, T., Huang, W., Liu, Y., Gasparrini, A., ... & Li, S. (2025). Estimating the urban heat-related mortality burden due to greenness: a global modelling study. *The Lancet Planetary Health*, 9(7), 101235.
- Xu, H., Guo, S., Shi, X., Wu, Y., Pan, J., Gao, H., Tang, Y., & Han, A. (2024). Machine learning-based analysis and prediction of meteorological factors and urban heatstroke diseases. *Frontiers in Public Health*, 12, 1420608.
- Xu, Z., FitzGerald, G., Guo, Y., Jalaludin, B., & Tong, S. (2016). Impact of heatwave on mortality under different heatwave definitions: A systematic review and meta-analysis. *Environment international*, 89, 193-203.

- Xu, Z., Watzek, J. T., Phung, D., Oberai, M., Rutherford, S., & Bach, A. J. (2023). Heat, heatwaves, and ambulance service use: A systematic review and meta-analysis of epidemiological evidence. *International Journal of Biometeorology*, 67(10), 1523-1542.
- Yan, M., Xie, Y., Zhu, H., Ban, J., Gong, J., & Li, T. (2022). The exceptional heatwaves of 2017 and all-cause mortality: An assessment of nationwide health and economic impacts in China. *Science of the Total Environment*, 812, 152371.
- Ye, X., Wolff, R., Yu, W., Vaneckova, P., Pan, X., & Tong, S. (2012). Ambient temperature and morbidity: A review of epidemiological evidence. *Environmental health perspectives*, 120(1), 19-28.
- Yin, L. K., Yogarayan, S., Razak, A., Fatimah, S., Bukar, U. A., & Sayeed, M. S. (2024). Heat stroke prediction: A perspective from the internet of things and machine learning approach. *International Journal of Electrical & Computer Engineering*, 14, 3427-3433.
- Yin, P., Chen, R., Wang, L., Liu, C., Niu, Y., Wang, W., Jiang, Y., Liu, Y., Liu, J., & Qi, J. (2018). The added effects of heatwaves on cause-specific mortality: A nationwide analysis in 272 Chinese cities. *Environment international*, 121, 898-905.
- Yu, Y.-Y., Liang, Q., Hou, J., Fujii, M., Qian, T., He, Z., & Wang, H. (2024). Health cost impacts of extreme temperature on older adults based on city-level data from 28 provinces in China. *Environmental Research Letters*, 19(4), 044017.
- Zhang, K., Li, Y., & Schwartz, J. D. (2014). What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods. *Environmental research*, 132, 350-359.
- Zhang, K., Rood, R. B., Michailidis, G., Oswald, E. M., Schwartz, J. D., Zanobetti, A., Ebi, K. L., & O'Neill, M. S. (2012). Comparing exposure metrics for classifying 'dangerous heat' in heat wave and health warning systems. *Environment international*, 46, 23-29.
- Zhao, Q., Guo, Y., Ye, T., Gasparrini, A., Tong, S., Overcenco, A., Urban, A., Schneider, A., Entezari, A., & Vicedo-Cabrera, A. M. (2021). Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019 : A three-stage modelling study. *The Lancet Planetary Health*, 5(7), e415-e425.
- Zhao, Q., Li, S., Ye, T., Wu, Y., Gasparrini, A., Tong, S., Urban, A., Vicedo-Cabrera, A. M., Tobias, A., & Armstrong, B. (2024). Global, regional, and national burden of heatwave-related mortality from 1990 to 2019: A three-stage modelling study. *PLoS medicine*, 21(5), e1004364.

ANNEXE : DESCRIPTION DES MODÈLES UTILISÉS

A.1. Modèles statistiques utilisés en épidémiologie environnementale

Modèle linéaire généralisé (Generalized Linear Model; GLM)

Le GLM est une extension du modèle de régression linéaire multiple qui assouplit l'hypothèse de normalité et qui permet une transformation (non linéaire) de la variable réponse (Nelder & Wedderburn, 1972). Le GLM avec p variables explicatives ou prédicteurs est défini comme suit :

$$g(E[y_i]) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (\text{A.1})$$

où $y_1, \dots, y_n$ sont les observations indépendantes et identiquement distribuées de la variable réponse avec une loi de la famille exponentielle, $g(\cdot)$ est la fonction de lien, les β_j sont les paramètres à estimer du modèle et les x_{ij} sont les variables explicatives aussi appelées prédicteurs.

Least absolute shrinkage and selection operator (Lasso)

Afin de sélectionner les variables explicatives les plus pertinentes pour le modèle GLM de l'équation A.1, plusieurs méthodes de sélection de variables existent. La régression pénalisée permet d'ajouter une pénalité à la fonction de perte afin d'éviter d'avoir trop de variables dans le modèle final. Le Lasso a l'avantage de réduire la taille de certains paramètres du modèle afin qu'ils prennent la valeur stricte de 0, ce qui permet donc de faire la sélection des variables (Tibshirani, 1996). Pour ce faire, une pénalité est ajoutée à la log-vraisemblance, qui est la somme des valeurs absolues des paramètres. Le niveau de pénalité est défini par un hyperparamètre λ à optimiser en validation croisée dans l'équation suivante :

$$l(\beta_0, \beta_1, \dots, \beta_p, y) - \lambda \sum |\beta_j| \quad (\text{A.2})$$

Modèle additif généralisé (Generalized Additive Model; GAM)

Le GAM est une généralisation du GLM de l'équation (A.1) qui permet des transformations non linéaires de chaque prédicteur, tout en maintenant le concept d'additivité (Hastie & Tibshirani, 2017). Il est défini comme suit :

$$g(E[y_i]) = \beta_0 + \sum_{j=1}^p f_j(x_{ij}) \quad (\text{A.3})$$

où les $f_j(\cdot)$ sont des fonctions non linéaires lisses appliquées à chaque prédicteur qui remplacent les β_j de l'équation (A.1). Ces fonctions peuvent être de différentes natures comme des splines cubiques naturelles ou de lissage (James et al., 2013). Le modèle GAM permet aussi d'inclure des effets linéaires comme dans le GLM.

Modèle non-linéaire à délais distribués (Distributed Lag Non-Linear Model; DLNM)

En épidémiologie environnementale, la variable d'exposition à différents délais (*lags*) doit souvent être considérée. Le DLNM a été introduit pour considérer à la fois les effets non linéaires (comme dans le GAM) et les différents délais de la variable d'exposition (Armstrong, 2006; Gasparrini et al., 2010). Le DLNM utilise des fonctions de bases bidimensionnelles (*cross-basis functions*) qui décrivent la forme de la relation dans l'espace du prédicteur et dans l'espace des délais. De manière simplifiée, le DLNM s'écrit :

$$g(E[y_t]) = \alpha + CB(x_t; d) \quad (\text{A.4})$$

où $CB(\cdot; \cdot)$ est la fonction de base bidimensionnelle pour la variable d'exposition x , une surface en deux dimensions où le premier axe correspond aux valeurs de variable d'exposition x et le second axe aux différents délais jusqu'à d . Le DLNM de l'équation peut aussi comprendre d'autres prédicteurs, notamment pour contrôler l'effet de certaines variables et les tendances saisonnières et à long terme. Ces effets sont définis de la manière que dans le GAM de l'équation (A.3), c'est-à-dire avec des fonctions $f_j(\cdot)$ qui peuvent être linéaires ou non, ou avec des fonctions de base bidimensionnelles comme dans l'équation (A.4).

A.2. Modèles d'apprentissage automatique basés sur les arbres

Arbre de décision (Decision Tree; DT)

Le DT a la forme d'un arbre inversé (Quinlan, 1986). Il est construit en séparant l'espace des prédicteurs en J régions $R_1, \dots, R_J$ sans chevauchement qui prennent la forme de boîtes (James et al., 2013). Pour un arbre de régression, la valeur prédite $\hat{y}_{R_j}$ pour chaque région correspond à la moyenne des observations de cette région. Pour entraîner un tel modèle, on veut trouver les régions qui minimisent la somme des carrés résiduels (SCR) :

$$SCR = \sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \quad (\text{A.5})$$

Considérer toutes les régions possibles serait computationnellement impossible. Ainsi, on utilise plutôt une approche *top-down* qui utilise la division binaire récursive (James et al., 2013). On commence en haut de l'arbre (où toutes les observations appartiennent à une même région) et on cherche la meilleure séparation de l'espace des prédicteurs. Parmi toutes les séparations possibles des prédicteurs, on retient la séparation pour le prédicteur qui réduit le plus la SCR, créant ainsi un nœud interne. On répète cette opération au sein des régions nouvellement créées pour trouver la prochaine séparation de l'espace des prédicteurs qui réduira le plus la SCR. Cette procédure est répétée jusqu'à ce qu'un critère d'arrêt soit atteint (p. ex. un nombre inférieur à 5 observations au sein des nœuds terminaux ou feuilles). Un arbre trop gros (c.-à-d. avec un trop grand nombre de régions J) pourrait être surentraîné, menant à des performances faibles sur de nouvelles observations. La procédure d'élagage permet de sélectionner un sous-arbre plus petit que l'arbre complet qui aura de meilleures performances en généralisation.

Forêt aléatoire (*Random Forest*; RF)

Étant donné la grande variance des prédictions de l'arbre de décision, les modèles d'ensembles ont été proposés (James et al., 2013). Le *bagging* (*bootstrap aggregation*) est une procédure qui consiste à entraîner B modèles sur B jeux de données différents obtenus en tirant des échantillons avec remise (*bootstrap*) du jeu de données d'entraînement. Le *bagging* est particulièrement utilisé avec les arbres de décisions comme modèles sous-jacents. La prédiction est donnée par la moyenne des prédictions des B modèles :

$$\hat{f}(\mathbf{X}) = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(\mathbf{X}) \quad (\text{A.6})$$

L'idée ici est que la prédiction moyenne de plusieurs modèles calibrés sur des données différents aura une variance réduite par rapport à la prédiction d'un seul modèle. Le RF est une extension du *bagging* qui vise aussi à décorréler les modèles sous-jacents (Breiman, 2001). Pour ce faire, le RF ne permet que l'utilisation d'une sous-sélection aléatoire de m prédicteurs (parmi les p disponibles, $m \approx \sqrt{p}$) comme candidats pour chaque séparation dans les DT sous-jacents. En faisant ainsi, on évite qu'un prédicteur très explicatif se retrouve dans tous les modèles sous-jacents. La variance des prédictions en sera ainsi davantage réduite par rapport à la moyenne de

plusieurs modèles fortement corrélés. Il faut noter que bien que le RF gagne en prédictibilité par rapport au DT, il perd en interprétabilité étant donné la multitude d'arbres sous-jacents impossibles à analyser individuellement (B variant entre 100 et 1000). Il existe cependant une méthode assez simple pour déterminer l'importance de chacun des prédicteurs qui consiste à calculer la réduction de SCR pour chaque prédicteur au sein de tous les DT entraînés : une large réduction correspondant à une plus grande importance dans le RF (James et al., 2013).

Gradient Boosting Model (GBM)

Dans le RF présenté ci-dessus, l'entraînement des arbres se fait de manière parallèle et indépendante entre les arbres. L'idée du GBM est d'utiliser le résultat des arbres précédents pour faire l'entraînement des arbres successifs (Friedman, 2001). En gros, un premier arbre est entraîné sur les données. Habituellement, cet arbre n'est pas très profond (i.e., quelques feuilles seulement), ce qu'on appelle un *weak learner*. Une fois un premier arbre entraîné, le résidu est calculé, c'est-à-dire qu'on retire de la variable réponse y la valeur prédite par l'arbre multipliée par le taux d'apprentissage λ (p. ex. 0.01 ou 0.001). Le deuxième arbre est donc entraîné sur le résidu obtenu, et non sur la valeur initiale de y . Cette procédure est répétée pour tous les arbres suivants. Ainsi, l'apprentissage à chaque itération se fait sur les observations où les arbres précédents n'ont pas bien performé. Cet apprentissage séquentiel plutôt qu'en parallèle est donc intéressant afin de mieux calibrer le modèle et éviter le surapprentissage. Dans le GBM, on peut aussi utiliser des échantillons *bootstrap* pour chaque nouvel arbre ainsi qu'une sous-sélection des prédicteurs pour chaque noeud interne comme dans le RF. D'autres hyperparamètres sont aussi à optimiser comme la profondeur des arbres, le niveau de sous-échantillonnage, le taux d'apprentissage, le nombre d'arbres, etc. De nos jours, de nouveaux algorithmes ont été développés sur l'idée initialement proposée du GBM, qui s'adaptant à différentes fonctions de pertes et qui sont plus rapides que ce dernier. Les plus récentes implémentations sont le *eXtreme Gradient Boosting ou XGBoost* (Chen & Guestrin, 2016), suivi par le *Light Gradient Boosting Model* (LGBM) (Ke et al., 2017).

A.3 Modèles d'apprentissage automatique et profond basé sur les réseaux de neurones

Perceptron à une seule couche (Single-Layer Perceptron; SLP)

Les réseaux de neurones ont été développés sur la base du cerveau humain, où les unités (en bleu) sont des neurones et où les connexions (flèches) sont les synapses (Figure A.1). Le SLP est le modèle le plus simple de réseau de neurones à propagation avant (*feedforward*) (Hastie et al., 2009). Les prédicteurs $x_1, \dots, x_p$ de la couche d'entrée sont combinés au sein de la couche cachée. Chaque neurone $z_1, \dots, z_q$ est défini par la combinaison linéaire des valeurs de la couche d'entrée avec un vecteur de poids, à laquelle un biais est ajouté. Une fonction d'activation $g(\cdot)$ est ensuite appliquée (p. ex. une fonction sigmoïde) à chacun des neurones. Les résultats de chacun des neurones sont ensuite combinés (avec poids et biais) avant la couche de sortie et la fonction de transformation finale pour faire la prévision.

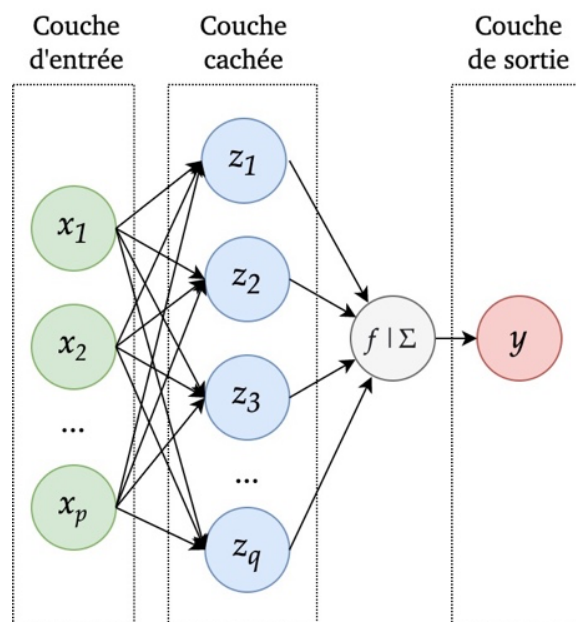


Figure A.1 Illustration d'un perceptron à une seule couche.

Perceptron multicouche (Multi-Layer Perceptron; MLP)

Le MLP généralise l'idée du SLP avec plusieurs couches successives cachées de neurones (Figure A.2). Dans le MLP, comme on suppose au moins deux couches cachées, on parle ici d'apprentissage profond. Tout comme dans le SLP, les résultats de la couche précédente sont combinés avec des poids, des biais et des fonctions d'activation avant d'être envoyés à la couche cachée suivante ou dans la couche de sortie. Les poids et biais de chacune des couches sont les

paramètres à optimiser du réseau qui sont trouvés en minimisant la fonction de perte (p. ex. la SCR) avec la méthode de la descente du gradient, aussi appelée la rétropropagation de l'erreur dans ce contexte (Hastie et al., 2009). En pratique, on applique une régularisation (pénalité ou *early stopping*) pour éviter un MLP surentraîné. L'architecture du MLP, c'est-à-dire le nombre de couches et de neurones à chacune des couches, est un hyperparamètre à optimiser du modèle.

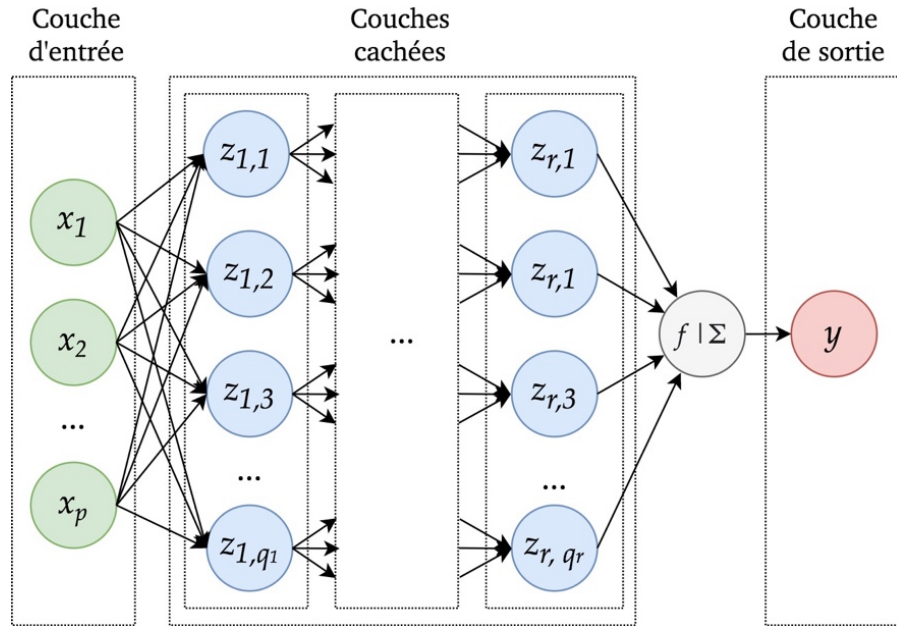


Figure A.2 Illustration d'un perceptron multicouche.

Long Short-Term Memory (LSTM)

Les réseaux de neurones récurrents sont particulièrement adaptés au traitement de données séquentielles. À l'inverse des réseaux à propagation avant comme le MLP, la cellule du réseau récurrent transfère de l'information au temps $t-1$ vers la cellule au temps t (Figure A.3) (Goodfellow et al., 2016).

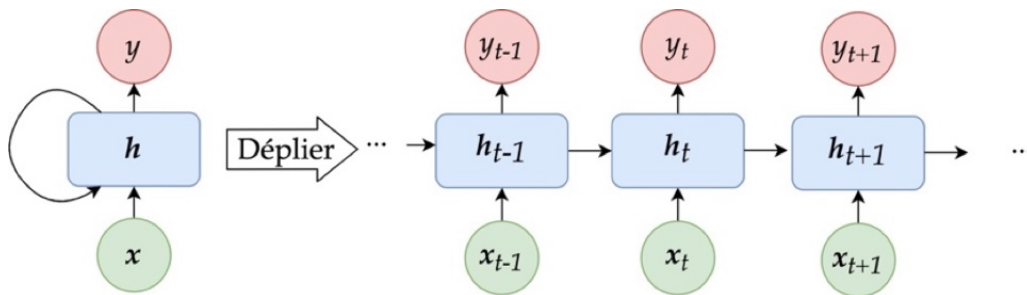


Figure A.3 Illustration d'un réseau de neurones récurrent.

Le LSTM est une extension des réseaux de neurones récurrents classique qui possède une mémoire et qui corrige certains problèmes liés à l'optimisation de ces réseaux avec la méthode du gradient (Hochreiter & Schmidhuber, 1997). Les cellules du LSTM comportent désormais une mémoire à long terme c_t qui est transmise aux cellules suivantes (Figure A.4). Cette mémoire peut oublier de l'information grâce à la porte de sortie f_t ou en emmagasiner grâce à porte d'entrée i_t . La porte de sortie o_t contrôle quant à elle l'information transmise à la prochaine cellule h_t et permet de faire la prédiction y_t . Le nombre optimal de cellules et de couches dans le LSTM sont encore une fois des hyperparamètre à optimiser, tout comme le taux d'apprentissage, le nombre d'époques, l'ajout de couches de régularisation, etc.

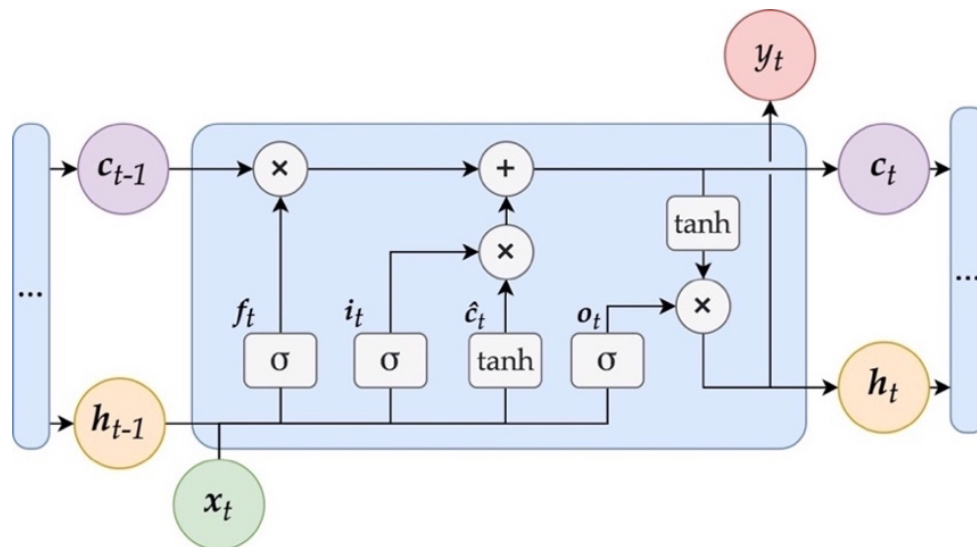


Figure A.4 Illustration d'une cellule d'un *Long Short-Term Memory*.