

Comparaison de méthodes d'agrégation spatiale de valeurs météorologiques à des fins de surveillance en santé publique

SURVEILLANCE DES MALADIES CHRONIQUES

Numéro 22

PRINCIPAUX CONSTATS

+ Les études sur la santé, la météorologie et le climat présentent souvent des modèles de régression utilisant la moyenne spatiale de plusieurs stations météo à proximité de l'évènement de santé, mais sans tenir compte de leurs caractéristiques géographiques ou humaines.

Cette étude comparait des techniques par krigeage à diverses techniques de pondération, dont la moyenne spatiale, afin d'en évaluer l'effet sur les résultats.

Le krigeage améliore les résultats de la modélisation, mais la différence n'est pas suffisante pour justifier sa complexité et son temps de calcul. Les techniques de pondération donnent des résultats similaires à la moyenne spatiale.

Depuis la réalisation de ce projet, des données météorologiques krigées exhaustives et d'une grande qualité (soit Daymet et ANUSPLIN) sont disponibles en Amérique du Nord et leur utilisation est recommandée.

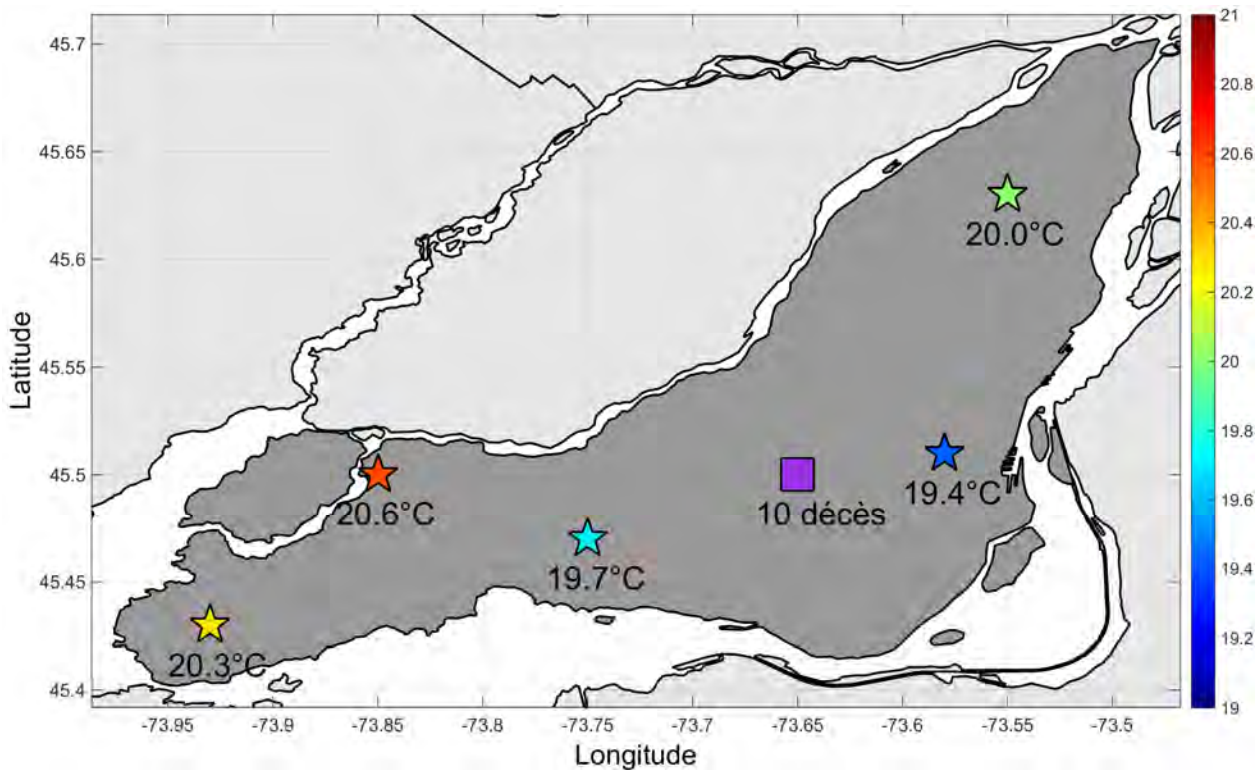
Mise en contexte

Selon le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC, 2014), le nombre de publications portant à la fois sur le changement climatique et ses impacts n'a cessé de croître depuis quelques décennies. Ce nombre est passé de 365 pour la période 1981 à 1990 à 3 596 de 1991 à 2000, et à 21 599 de 2001 à 2010, lorsque toutes les langues sont considérées pour les publications. Les impacts étudiés couvrent notamment la mortalité et la morbidité des populations humaines.

L'une des méthodes statistiques les plus utilisées pour étudier les relations entre les impacts sanitaires et la météorologie ou le climat est la régression. La régression regroupe un ensemble de modèles (ex. : linéaire simple, linéaire multiple, généralisé additif) qui permettent d'expliquer une variable dépendante (ex. : nombre de décès par jour dans une ville) à l'aide d'une ou de plusieurs variables météorologiques indépendantes (ex. : température journalière).

Les modèles de régression, sous leur forme la plus simple, nécessitent une valeur unique à chaque pas de temps (ex. : chaque jour) pour chaque variable (dépendante et indépendante). Cependant, les sources de données multiples ont des répartitions spatiales différentes selon la nature de la variable. La figure 1 donne une illustration de cette problématique pour la mortalité et la température moyenne, un jour i , sur l'île de Montréal.

Figure 1 Illustration de la problématique d'agrégation, un jour *i*, sur l'île de Montréal



Note : les étoiles de couleur représentent les différentes stations météorologiques avec la température moyenne observée ; le carré mauve, les décès pour une maladie donnée.

En général, les séries de données de la variable sanitaire fournissent une valeur par jour (ex. : nombre de décès pour cause de maladies cardiovasculaires le 21 octobre 2016), par région d'étude (ex. : région sociosanitaire ou communauté métropolitaine de Montréal). Parallèlement, les nombreuses valeurs de séries météorologiques proviennent de multiples stations météorologiques situées sur le territoire de la région visée. Pour que ces données soient analysées pour leur relation avec la variable sanitaire d'intérêt à travers un modèle de régression, elles doivent être agrégées spatialement et à chaque pas de temps (ici un jour) en une valeur unique représentative de la région. Pour ce faire, les approches utilisées dans la littérature scientifique sont généralement simples. Elles consistent à prendre les valeurs météorologiques de la station la plus proche de la région à l'étude ou la moyenne des valeurs provenant des stations sur l'ensemble de cette région (ex. : Abrignani et collab., 2009 ; Kriszbacher et collab., 2009). Toutefois, ces approches ne tiennent compte ni des caractéristiques géographiques ni des caractéristiques humaines des régions étudiées.

Dès lors, la principale question qui se pose est comment agréger les multiples valeurs d'une variable météorologique d'une région en une valeur unique représentative de la région ? L'objectif de ce projet est d'évaluer si le krigeage et certaines techniques de pondération sont préférables à la moyenne spatiale directe, généralement considérée dans les études.

L'ensemble de cette démarche s'inscrit au Programme de recherche en santé cardiovasculaire et changements climatiques 2011- 2016, réalisé dans le cadre d'une entente entre l'Institut national de santé publique du Québec (INSPQ) et le Centre de recherche Eau Terre Environnement de l'Institut national de recherche scientifique du Québec. Pour plus de détails, voir le document source (Giroux et collab., 2013).

Qu'est-ce que le krigeage ?

Le krigeage est une technique géostatistique spatiale permettant d'interpoler la valeur d'une variable aléatoire à un emplacement non observé à partir des observations des points situés à proximité de cet emplacement (annexe 1).

Le krigeage s'avère donc particulièrement utile pour les données météorologiques sur grille, comme les Daymet et les ANUSPLIN (annexe 2), qui peuvent offrir une couverture insuffisante dans certaines régions. Dans le contexte présent, l'attrait pour cette technique est d'obtenir une estimation d'une valeur météorologique à n'importe quel emplacement de la région d'intérêt.

Faire du krigeage requiert toutefois l'utilisation de divers paramètres pouvant être ajustés. Dans notre projet, 12 types de krigeage différents ont ainsi été considérés afin d'en obtenir les paramètres optimaux. Ils réfèrent à 12 combinaisons entre les variogrammes (Cauchy, gaussien ou sphérique), l'hétérogénéité géographique (isotrope ou anisotrope) et le nombre limité (ou pas) de stations météorologiques utilisées.

Quelles techniques de pondération utiliser ?

Les techniques de pondération déterminent la pondération à utiliser pour résumer les données météorologiques des stations. Ainsi, chaque technique mène à une pondération spécifique.

Dans notre projet, outre la moyenne spatiale directe en tant que mesure étalon, les techniques de pondération considérées sont la moyenne spatiale pondérée, l'analyse en composantes principales et la régression par station (annexe 3).

La « moyenne spatiale directe » correspond à la moyenne arithmétique de valeurs ayant des positions spatiales différentes. Ses principaux avantages sont : simple, directe et applicable aux stations réelles et aux valeurs par points de grille.

La « moyenne spatiale pondérée » consiste à attribuer le poids associé à chaque station météorologique en fonction d'un certain facteur. Dans notre projet, cette méthode a seulement été utilisée avec les données obtenues par krigeage pour les territoires de Centres locaux de services communautaires (CLSC). Le facteur de pondération utilisé est la population du recensement de 2001 (soit le point milieu de la période étudiée) pour les différents CLSC.

L'analyse en composantes principales transforme des variables corrélées en nouvelles variables non corrélées les unes aux autres (Jolliffe, 2002). Les composantes principales sont des combinaisons linéaires des variables de départ. Généralement, les premières composantes regroupent la majorité de la variance expliquée. Dans la présente étude, les stations jouent le rôle des variables de départ et la composante première principale obtenue représente la nouvelle série pondérée.

Enfin, la technique de régression par station consiste d'abord à faire un modèle de régression pour chaque station météorologique. Ensuite, la pondération des stations est définie en utilisant les valeurs de déviance expliquée ou du critère d'information d'Akaike (ou AIC, Akaike, 1973) de ces modèles.

Comment comparer l'effet de ces techniques sur les données ?

Afin de comparer les résultats obtenus avec la moyenne spatiale directe aux résultats découlant du krigeage et d'autres techniques de pondération, nous avons modélisé les relations santé-météorologie à l'aide du modèle additif généralisé (GAM) (annexe 4). Le principal attrait de ce modèle est qu'il permet de déceler des effets non linéaires entre l'issue sanitaire à expliquer et la météorologie.

Dans notre projet, l'issue sanitaire considérée est le nombre quotidien de décès pour cause principale de maladies cardiovasculaires (MCV) dans les régions sociosanitaires (RSS) de Montréal (RSS-06), de la Capitale-Nationale (RSS-03) et du Saguenay-Lac-Saint-Jean (RSS-02), de 1996 à 2007. Par MCV, on entend les maladies coronariennes, les maladies vasculaires cérébrales et l'insuffisance cardiaque (voir l'annexe 5 pour les statistiques descriptives des décès). Quant aux variables météorologiques, elles incluent : la température (minimale, moyenne et maximale), l'humidité relative (minimale, moyenne et maximale), la pression ajustée au niveau de la mer (minimale, moyenne et maximale), les précipitations totales et l'épaisseur de la neige au sol. Les données sanitaires proviennent du Système intégré de surveillance des maladies chroniques du Québec (SISMACQ) de l'INSPQ¹. Les données météorologiques sont tirées du portail de téléchargement DAI d'Environnement Canada².

Des modèles GAM ont d'abord été appliqués sur les données issues des différentes techniques de pondération. Ensuite, les valeurs de déviance expliquée³

de ces modèles ont été comparées. Enfin, lorsque les valeurs de déviance expliquée n'étaient pas suffisamment différentes pour faire une comparaison adéquate entre les modèles, deux approches inspirées de Wood (2006) ont été considérées :

- une agrégation temporelle sur 7 jours pour la somme des décès, la moyenne de la température, de l'humidité et de la pression, et le cumul des précipitations et de neige au sol ;
- l'ajout de deux variables liées au « temps » ; l'une pour le jour ordinal de la série (1 à N, avec N étant le nombre de jours dans la série), l'autre pour le jour ordinal de l'année (1 à 365).

Comment s'imbriquent toutes ces étapes ?

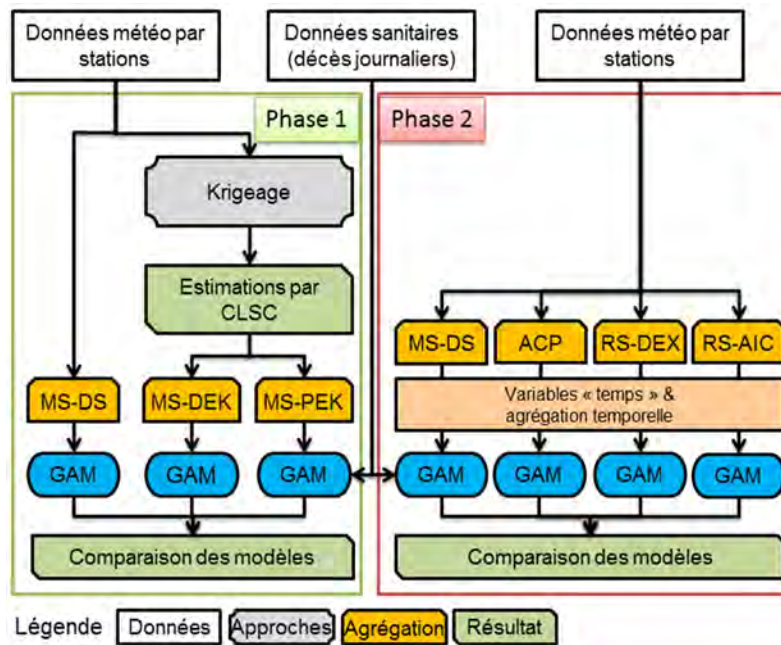
Comme illustrée à la figure 2, notre démarche méthodologique compte deux phases :

Dans la première phase, deux techniques de pondération des estimations obtenues par krigeage, selon les territoires de CLSC, ont été comparées à la moyenne spatiale directe des stations météorologiques (MS-DS, observations réelles). Il s'agit de : 1) la moyenne spatiale directe des estimations par krigeage (MS-DEK) et 2) la moyenne spatiale pondérée des estimations par krigeage (MS-PEK). De plus, comme déjà rapportés, 12 types de krigeage ont été considérés afin d'obtenir les paramètres optimaux du krigeage.

¹ Le Système intégré de surveillance des maladies chroniques du Québec (SISMACQ) est une banque de fichiers médico-administratifs jumelés obtenus avec l'accord de la Commission d'accès à l'information du Québec (CAI) le 27 mars 2008 (# 11 09 45 (08 19 12, 07 00 76)) dans le cadre de l'Étude de la faisabilité de développer la surveillance des MCV au Québec (1996-2007).

² Les données [DAI](#) sont des données d'observations disponibles aux emplacements des stations de mesure. Elles ont un pas temporel journalier pour la température moyenne, les précipitations totales et l'épaisseur de neige au sol, et une résolution temporelle horaire pour l'humidité relative et la pression ajustée au niveau de la mer.

³ La déviance expliquée (DEX) peut être interprétée de façon similaire au coefficient de détermination R^2 , mais dans un cadre non linéaire.

Figure 2 Algorithme de la méthodologie utilisée dans le projet


Dans la seconde phase, trois techniques de pondération ont été comparées à la moyenne spatiale directe des stations météorologiques, soit 1) la pondération basée sur l'analyse en composantes principales (ACP), 2) la pondération basée sur la déviance expliquée des régressions individuelles par station (RS-DEX), et 3) la pondération basée sur le critère d'information d'Akaike des régressions individuelles par station (RS-AIC).

Quels résultats découlent de la phase 1 ?

Dans la phase 1, les modèles de régression GAM utilisés pour expliquer la mortalité par MCV dans les RSS-03, 06 et 12, de 1996 à 2007, incluent les variables météorologiques statistiquement significatives avec un seuil (α) < 5 % (parmi les 11 variables disponibles). Des sous-analyses ont été faites selon le genre et pour l'année, l'hiver (novembre à mars) et l'été (mai à septembre).

Dans son ensemble, le krigeage a légèrement amélioré la déviance expliquée des modèles de régression (annexe 6). La plus grande amélioration est observée avec les modèles estivaux, alors que la plus faible résulte des modèles annuels et en particulier, lorsqu'ils portaient

sur l'ensemble de la population de la région (plutôt qu'un genre spécifique).

Par ailleurs, des 12 types de krigeage retenus, celui défini par un variogramme de type Cauchy, une hétérogénéité géographique isotrope et un nombre limité de stations météorologiques utilisées (CIL) a conduit à la meilleure déviance expliquée (la plus élevée). À l'opposé, le type de krigeage défini par un variogramme de type gaussien, une hétérogénéité géographique isotrope et un nombre limité de stations météorologiques utilisées (GIL) a conduit à la déviance expliquée la moins élevée.

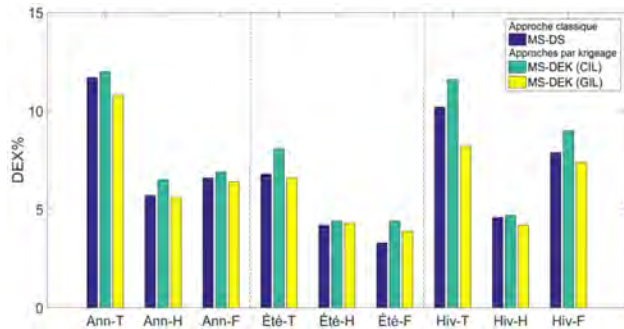
Toutefois, bien que le type CIL semble prometteur, on n'a pu distinguer de paramètres de krigeage optimaux, garantissant ainsi une amélioration suffisante des modèles de régression par rapport à la moyenne spatiale directe des stations météorologiques et permettant de justifier l'important temps de calcul nécessaire pour effectuer ces analyses.

Les résultats pour le meilleur (CIL) et le moins bon (GIL) types de krigeage sont comparés visuellement à la figure 3 pour les trois régions étudiées.

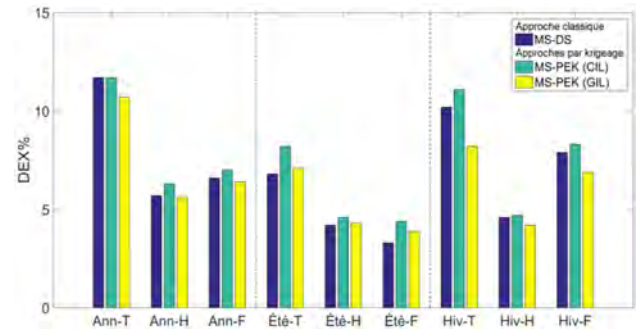
Figure 3 Comparaisons des moyennes spatiales des estimations par krigeage direct : (MS-DEK) ou pondéré (MS-PEK) par rapport à la moyenne spatiale directe des stations météorologiques - méthode classique (MS-DS), pour les krigeages CIL et GIL présentés pour les RSS de Montréal, de Québec et du Saguenay-Lac-Saint-Jean.

RSS de Montréal

a) MS-DEK par rapport à MS-DS

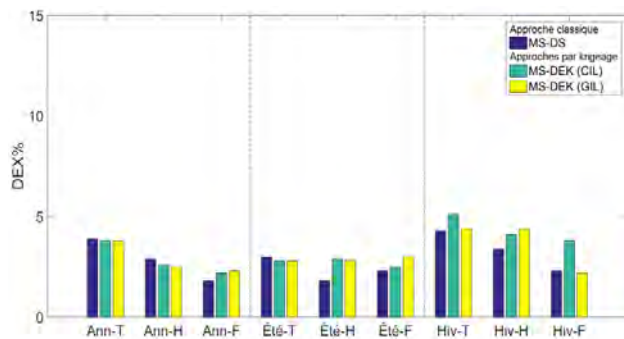


b) MS-PEK par rapport à MS-DS

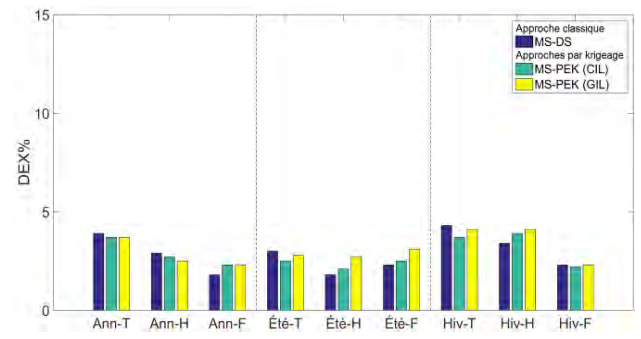


RSS de Québec

a) MS-DEK par rapport à MS-DS

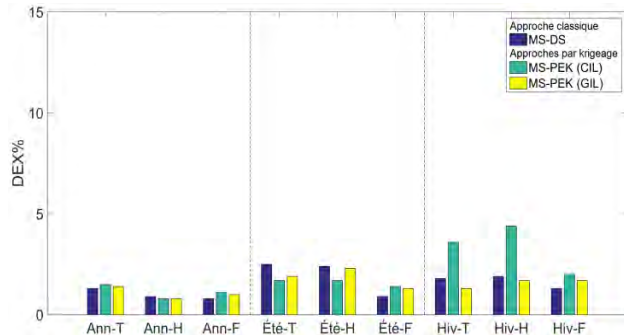


b) MS-PEK par rapport à MS-DS

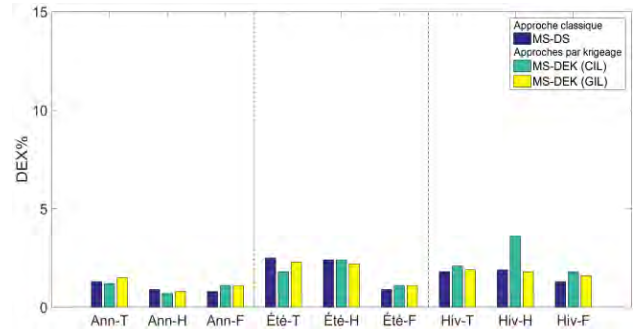


RSS du Saguenay-Lac-Saint-Jean

a) MS-DEK par rapport à MS-DS



b) MS-PEK par rapport à MS-DS



Légende : Ann : données annuelles -T : chez les hommes et femmes, -H : chez les hommes, -F : chez les femmes.
 Été : données estivales -T : chez les hommes et femmes, -H : chez les hommes, -F : chez les femmes.
 Hiv : données hivernales -T : chez les hommes et femmes, -H : chez les hommes, -F : chez les femmes.

Quels résultats découlent de la phase 2 ?

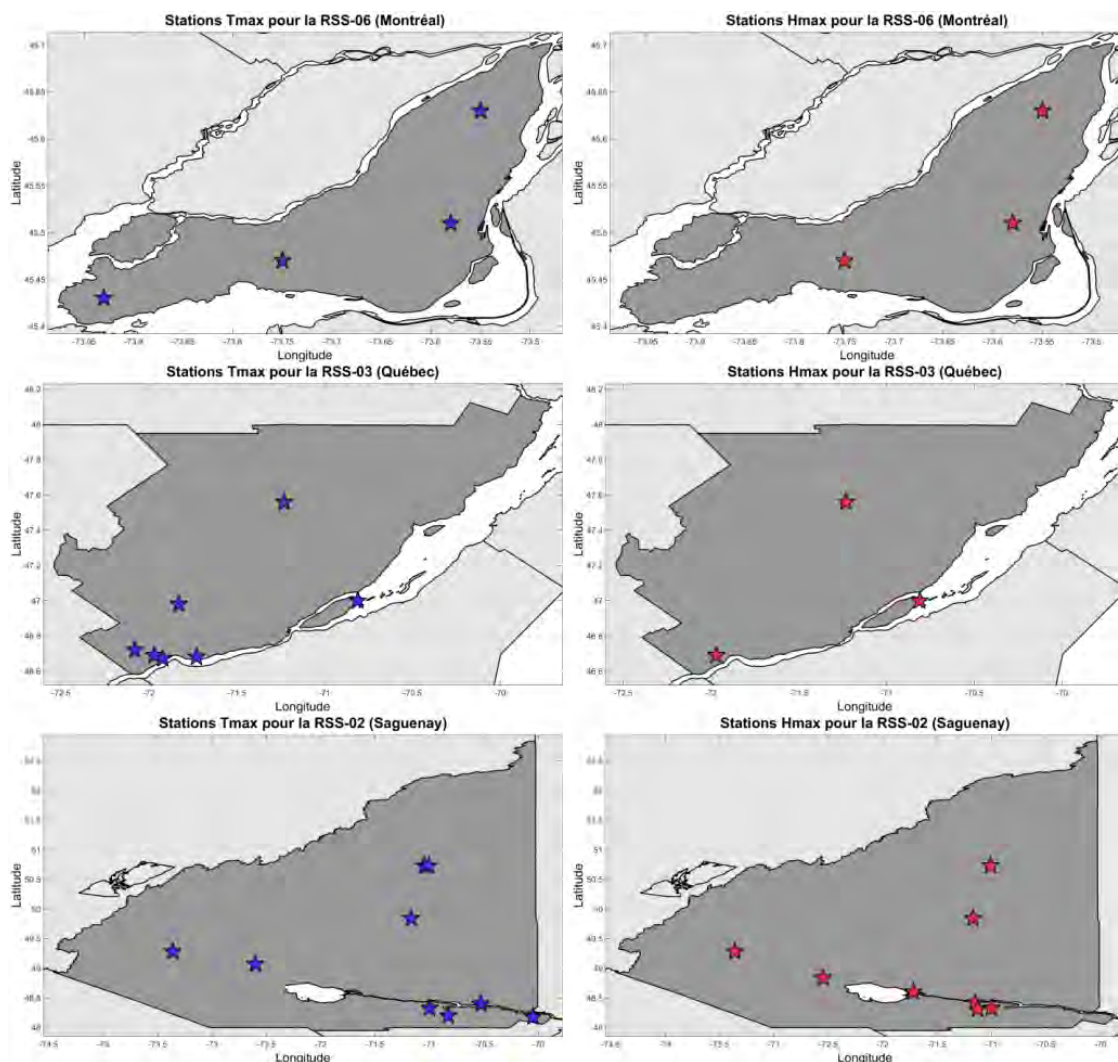
Dans la phase 2, des modèles de régression GAM ont aussi été produits afin d'estimer l'impact des techniques de pondération par rapport à la méthode spatiale directe des stations météorologiques (MS-DS, approche classique). Toutefois, contrairement à la phase 1, l'interpolation par krigeage n'a pas été utilisée.

Ainsi, les techniques de pondération ont été directement appliquées sur les séries de données observées des stations. Rappelons que ces techniques de pondération sont l'analyse en composantes principales (ACP), la

déviance expliquée des régressions individuelles par station (RS-DEK), et le critère d'Akaike des régressions individuelles par station (RS-AIC).

Comme dans la phase 1, les modèles de régression GAM utilisés pour expliquer la mortalité par MCV dans les trois régions sociosanitaires (RSS) étudiées, de 1996 à 2007, incluent les variables météorologiques statistiquement significatives avec un seuil (α) < 5 % (parmi les 11 variables disponibles). Ici, ces variables sont la température maximale et l'humidité maximale. L'agrégation temporelle et l'ajout de variables « temps » ont également été analysés (annexe 7). La figure 4 montre l'emplacement des stations météorologiques.

Figure 4 Emplacement des stations ayant les températures maximales (Tmax, figures de gauche) et l'humidité maximales (Hmax, figures de droite) pour les RSS étudiées



La figure 5 présente visuellement la comparaison des poids donnés à chaque station (hauteur dans la barre) par les différentes techniques de pondération pour les trois RSS considérées. On y remarque des résultats très différents entre les deux variables météorologiques retenues dans les modèles.

Dans le cas de la température maximale, les poids donnés aux stations sont uniformes (hauteurs des barres identiques), peu importe la technique de pondération employée. On note une très légère variation des poids pour la technique RS-DEX (pondération basée sur la déviance expliquée des régressions individuelles par station), mais cela ne se reflète pas dans la déviance expliquée des modèles GAM sous-jacents (annexe 7).

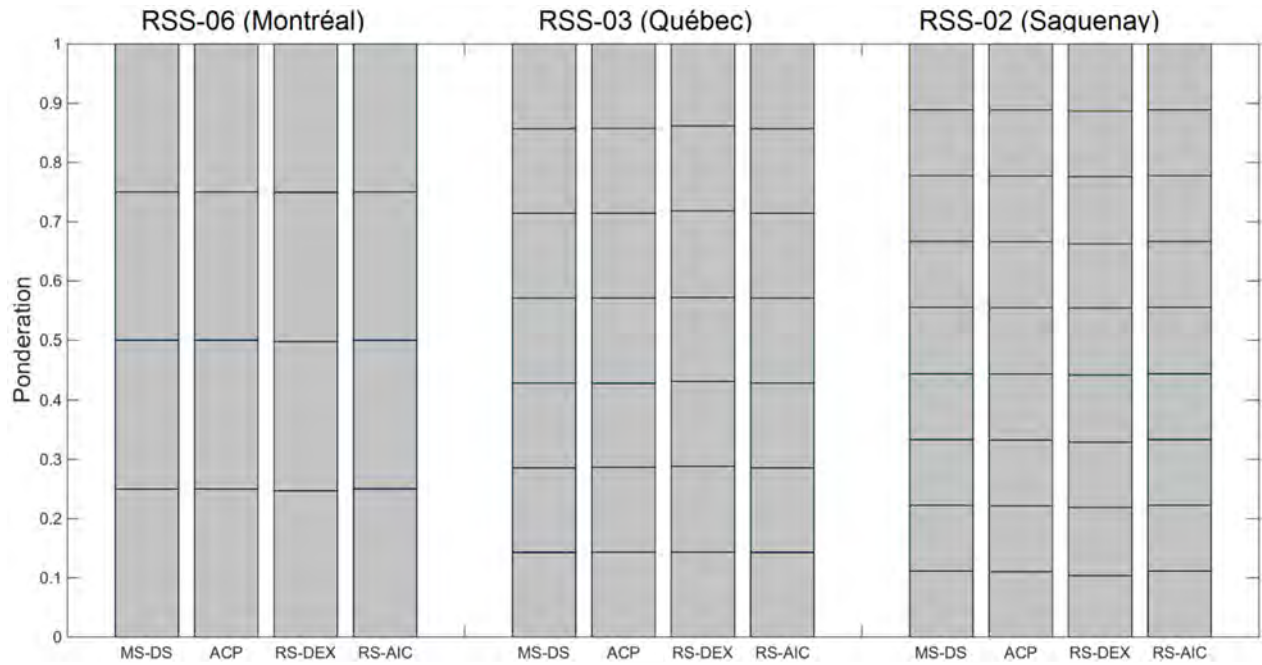
Dans le cas de l'humidité maximale :

- Les poids donnés aux stations ont moins d'uniformité. La technique RS-DEX produit les pondérations les plus différentes et cela est corroboré par une augmentation de la déviance expliquée des modèles GAM sous-jacents (annexe 7).
- Par ailleurs, les techniques ACP (pondération basée sur l'analyse en composantes principales) et RS-AIC (pondération basée sur le critère d'information d'Akaike des régressions individuelles par station) donnent de légères variations de pondération, sans toutefois qu'on puisse observer un changement de la déviance expliquée des modèles GAM (annexe 7).

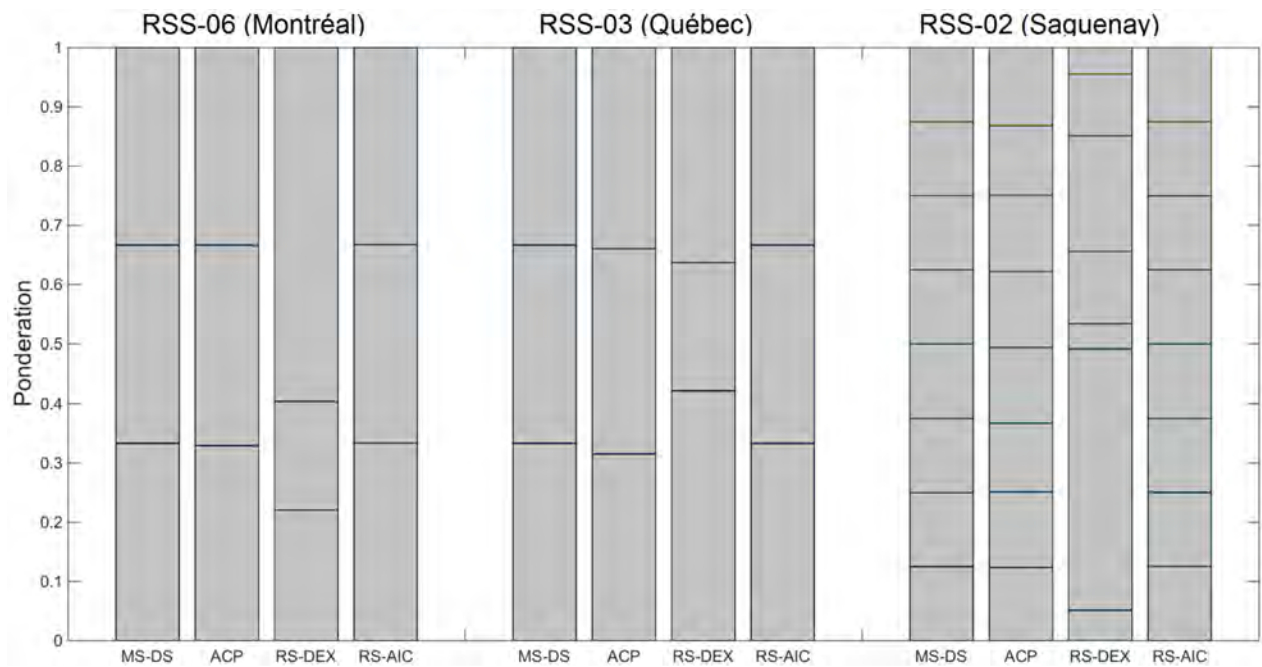
Enfin, pour les trois régions, l'agrégation temporelle sur 7 jours et l'ajout des variables temps conduisent à des augmentations importantes de la déviance expliquée de tous les modèles GAM (annexe 7). En outre, les modèles GAM avec les deux variables « temps » ont des valeurs de déviance similaires, suggérant qu'ils sont des facteurs explicatifs d'une importance comparable sur le plan statistique. Les résultats de l'agrégation temporelle montrent aussi une augmentation plus importante de la déviance expliquée pour Saguenay, puis pour Québec et enfin Montréal, ce qui apparaît lié au nombre de décès par jour (annexe 5).

Figure 5 Comparaison des poids donnés à chaque station (hauteur dans la barre) par les différentes techniques de pondération (a) pour la température maximale et (b) pour l'humidité maximale dans les trois RSS considérées

a) Température maximale



b) Humidité maximale



Conclusion

L'objectif principal de ce projet était de déterminer la méthode la plus appropriée pour agréger spatialement des multiples valeurs météorologiques en une valeur unique représentative de leur région et qui tient compte de la variabilité géographique et populationnelle, à des fins de surveillance en santé et changements climatiques.

La méthodologie utilisée contient deux principales composantes, soient l'interpolation spatiale (le krigeage) et la pondération des stations réelles ou virtuelles. Différents paramètres de krigeage ont été considérés afin de déterminer une combinaison optimale de paramètres. Une pondération par population de CLSC (MS-PEK) a également été testée. Étant donné que le krigeage nécessite d'importantes capacités computationnelles, des techniques de pondération déterminant directement la pondération des stations météorologiques ont été considérées. Le krigeage et les techniques de pondération ont été appliqués à trois régions sociosanitaires de la province de Québec ayant des caractéristiques géographiques et démographiques différentes.

Les résultats obtenus dans le cas d'étude montrent que, dans son ensemble, le krigeage a amélioré le pouvoir explicatif des modèles de régression. Toutefois, on n'a pas distingué de paramètres de krigeage optimaux menant à une amélioration systématique et suffisante des modèles de régression par rapport à l'approche classique utilisant la moyenne spatiale directe des stations météorologiques. Parmi les raisons pouvant étoffer cette observation, on relève : 1) la qualité et la quantité des données utilisées, notamment la distribution spatiale de stations et le faible nombre de décès pour la région sociosanitaire du Saguenay–Lac-Saint-Jean, 2) la faible part explicative du climat sur la santé cardiovasculaire, en général, par rapport à des facteurs individuels (comme le diabète) et possiblement 3) l'utilisation de modèles de régression non linéaires complexes (qui est utile dans ce genre d'études, selon la littérature). Une autre raison, d'ordre pratique, est que 4) le krigeage nécessite des choix et des temps de calcul importants, ce qui a limité le nombre d'essais possibles parmi toutes les combinaisons envisageables.

Les résultats obtenus ne permettent pas, pour les régions et les variables considérées dans le cas d'étude, de distinguer une technique de pondération spécifique qui améliore systématiquement et suffisamment les modèles de régression par rapport à la moyenne spatiale des stations. Néanmoins, les techniques de pondération considérées sont prometteuses, particulièrement en raison de leur faible complexité et de leurs temps de calcul raisonnable par rapport au krigeage. Pour certaines techniques de pondération (ACP en particulier), il pourrait être intéressant de modifier la formule de pondération avec une fonction de transfert pour amplifier les différences entre les poids des stations.

En conclusion, les résultats de ce projet n'ont pas permis de déterminer des paramètres de krigeage ou une technique de pondération conduisant systématiquement à de meilleurs résultats que la moyenne spatiale directe, suggérant que cette dernière reste une méthode d'agrégation simple et viable pour les cas étudiés et avec les séries de données actuellement disponibles. L'algorithme proposé et les techniques utilisées restent d'ordre général et peuvent donner des résultats plus distinctifs pour d'autres régions et d'autres maladies, notamment.

Références

Abriani, M. G., S. Corrao, G. B. Biondo, N. Renda, A. Braschi, G. Novo, A. D. Girolamo, G. B. Braschi, and S. Novo (2009). Influence of climatic variables on acute myocardial infarction hospital admissions, *International journal of cardiology*, 137(2), 123-129.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, pp. 267-281.

Eum, H.-I., P. Gachon, R. Laprise, and T. B. M. J. Ouarda (2012). Evaluation of regional climate model simulations versus gridded observed and regional reanalysis products using a combined weighting scheme, *Climate Dynamics* 38, Numbers 7-8, 1433-1457, DOI 10.1007/s00382-011-1149.

GIEC (2014) Impacts, adaptation, and vulnerability. Part A : Global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change: Cambridge and New York: Cambridge University Press.

Giroux, J. X., F. Chebana, D. Bélanger, E. Gloaguen, T. B. M. J. Ouarda, and A. St-Hilaire (2013). Comparaison de l'utilisation des moyennes spatiales à celle du krigeage, appliquée à la relation mortalité par MCV - météorologie, au Québec, de 1996 à 2007 (Projet M1). *INRS Eau Terre Environnement. Rapport de recherche no R1438*. 73 p.

Hastie, T., and R. Tibshirani (1986), *Generalized Additive Models*, *Statistical Science*, 1(3), 297-310.

Isaaks, E. H., and R. M. Srivastava (1989). *Applied geostatistics* (Vol. 2): Oxford University Press New York.

Jolliffe, I. T. (2002). Principal component analysis and factor analysis. *Principal component analysis*, 150-166.

Kriszbacher, I., J. Bódis, I. Csoboth, and I. Boncz (2009), The occurrence of acute myocardial infarction in relation to weather conditions, *International journal of cardiology*, 135(1), 136-138.

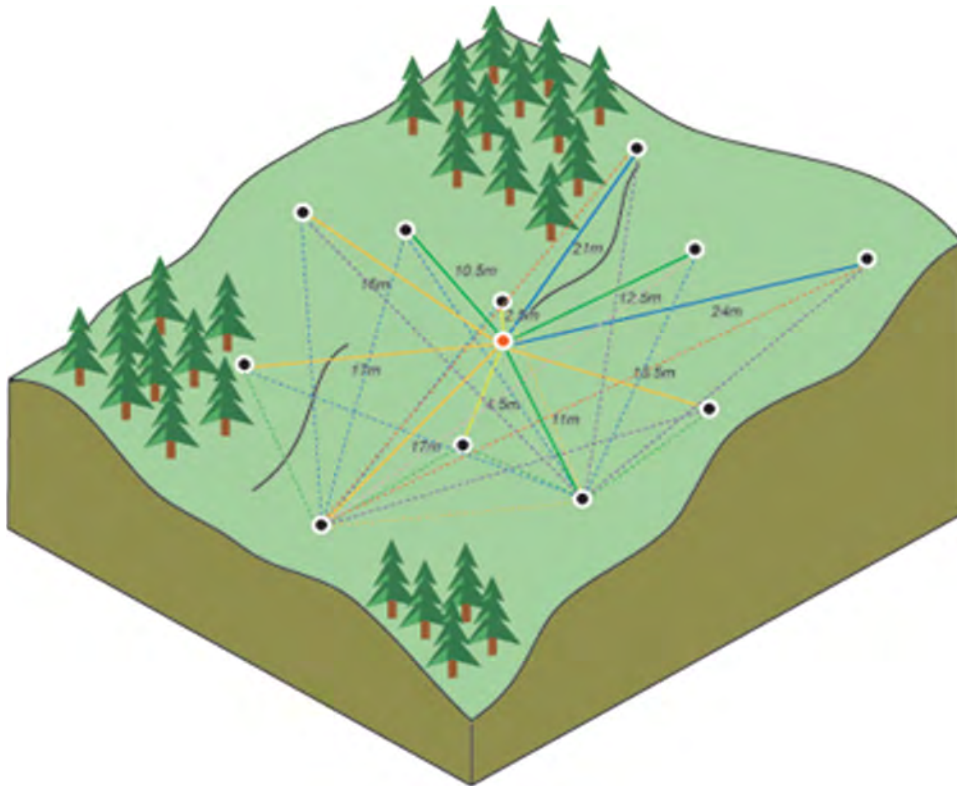
Wood, S. (2006). *Generalized additive models: an introduction with R*: CRC press.

Annexe 1 Krigeage

Le krigeage est une technique géostatistique conçue pour étudier des variables spatialement autocorrélées et continues dans l'espace (Isaaks & Srivastava, 1989). Il permet d'interpoler la valeur d'une variable aléatoire à un emplacement non observé à partir des observations des points à proximité. Il tient compte de la distribution et de la structure de corrélation spatiale de la variable à travers des fonctions de structure appelées variogrammes.

L'estimation à un point est obtenue par la combinaison linéaire (avec pondération) des valeurs à proximité. Cela signifie qu'on attribue un poids plus élevé aux valeurs les plus proches géographiquement parce qu'elles sont plus susceptibles d'être semblables à la valeur inconnue. Le fonctionnement du krigeage est illustré à la figure 1A. La valeur au point rouge est estimée à partir des valeurs aux points noirs, après avoir regroupé les valeurs équivalentes (ex. : 20-25 m : lignes bleues).

Figure 1A Illustration du fonctionnement du krigeage



Annexe 2 Données météorologiques ANUSPLIN et Daymet

Lorsque ce projet a été réalisé en 2012-2013, les [données ANUSPLIN](#) (grille de 10 km x 10 km) étaient disponibles de 1961 à 2003 pour les variables suivantes : température minimale, température maximale et précipitations. Puisque ces données étaient non disponibles pour la moitié de la durée de la présente étude, elles n'ont pas été utilisées.

Pour les [données Daymet](#), elles n'étaient pas encore disponibles au moment de l'étude. Si elles l'avaient été, elles auraient probablement été utilisées plutôt que de faire le krigeage par nous-mêmes. En effet, ces données fournissent des estimations maillées de paramètres météorologiques quotidiens pour l'ensemble de l'Amérique du Nord⁴ et elles sont interpolées sur une grille de 1 km x 1 km. Les variables disponibles sont : température (minimale et maximale), radiation solaire, pression de vapeur, équivalent eau-neige, précipitation et durée du jour.

Maintenant que les données Daymet Version 3 sont disponibles, il existe un jeu de données « gratuit », qui couvre l'ensemble du Canada de 1981 à 2015. Malgré que ces données soient volumineuses (~100 Go par variable pour l'Amérique du Nord, ~7 Go par variable pour le Québec seulement), leur grande qualité rend peu indiqué l'effort d'un krigeage maison. Enfin, si la variable d'intérêt ou la période considérée s'avère manquante dans Daymet, les données ANUSPLIN demeurent disponibles et d'une excellente résolution (Eum et collab., 2012).

En bref, les données Daymet et ANUSPLIN sont des données interpolées/extrapolées à partir de stations du Canada et des États-Unis utilisant des algorithmes complexes et plus avancés que ceux que nous pourrions faire par nous-mêmes. Nous suggérons donc leur utilisation pour les études en santé et changements climatiques au Québec et au Canada.

⁴ Données Daymet : version 2 disponible à partir de mai 2014, avec estimations jusqu'au 52° N. Version 3 : disponible à partir de mai 2014, avec estimations jusqu'au 52° N. Disponibles à partir de juillet 2016 ; estimations jusqu'au 84° N.

Annexe 3 Techniques de pondération

L'équation générale pour la pondération est :

$$\text{Valeur pondérée} = \sum_{i=1}^N w_i \cdot Z_i, \quad (1)$$

Où N est le nombre de stations météorologiques ; w_i , le poids de la station i ; et Z_i , la valeur de la variable météorologique à la station i .

Plus spécifiquement dans notre étude et selon la phase où ils ont été utilisés, les poids utilisés pour chaque technique de pondération sont :

Technique (phases)	MS-DS (1 et 2)	MS-P (1)	ACP (2)	RS (2)
Poids w_i	$\frac{1}{N}$	$\frac{Pop_i}{Pop_T}$	$\frac{\alpha_i}{\sum \alpha_i}$	$\frac{DEX_i}{\sum DEX_i}$ ou $\frac{1/AIC_i}{\sum 1/AIC_i}$

Légende : **MS-DS** : moyenne spatiale directe des stations. **MS-P** : moyenne spatiale pondérée. **ACP** : analyse par composantes principales. **RS** : Régression par stations. **N** : nombre de stations. **Pop_i** : population du CLSC i . **Pop_T** : population totale de la région. **α_i** : coefficients de l'ACP. $\sum \alpha_i$: sommation des α_i . **DEX_i** : déviance expliquée du modèle de régression i . $\sum DEX_i$: sommation des DEX_i. **AIC_i** : critère d'information Akaike du modèle de régression i . $\sum AIC_i$: sommation des AIC_i.

Annexe 4 Modèle additif généralisé (GAM)

Le modèle additif généralisé (GAM) a été introduit par Hastie et Tibshirani (1986). Sa forme générale est donnée par l'équation suivante (Wood, 2006) :

$$g[E(Y | X)] = g(\mu) = \eta = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m) + \varepsilon, \quad (2)$$

Où Y suit une distribution de la famille exponentielle, g est la fonction de lien et f_j est une fonction lisse représentée par une somme de fonctions de base $b_i(x)$ telles que :

$$f_j(x_j) = \sum_{i=1}^k b_i(x_j) \beta_i \quad (3)$$

Les coefficients β_i sont des paramètres inconnus à estimer.

Annexe 5 Statistiques descriptives des décès

Les statistiques descriptives des décès quotidiens pour cause principale de maladies cardiovasculaires (MCV) dans les régions sociosanitaires (RSS) de Montréal, de la Capitale-Nationale et du Saguenay–Lac-Saint-Jean, de 1996 à 2007 :

RSS	Minimum	Moyenne	Maximum	Écart-type	1 ^{er} quartile	Médiane	3 ^e quartile
Montréal	1	10,22	27	3,62	8	10	13
Québec	1	3,22	13	1,73	2	3	4
Saguenay	1	1,68	6	0,90	1	1	2

Légende : les codes CIM suivants ont été utilisés, selon les années respectives d'utilisation. CIM-9 : 410-414 ; 362,3 ; 430-438 ; 428. CIM-10 : 120-125 ; I60-I69 ; I50.

Annexe 6 Phase 1 de l'analyse

Les déviations expliquées des modèles GAM visant à expliquer la mortalité quotidienne par la météorologie dans les RSS de Montréal (RSS-06), de la Capitale-Nationale (RSS-03) et du Saguenay-Lac-Saint-Jean (RSS-02), de 1996 à 2007 sont présentées au tableau suivant.

Les paramètres météorologiques sont agrégés à l'aide de la moyenne spatiale directe des stations (MS-DS, modèle de référence) et de la moyenne spatiale directe

des estimations par krigeage (MS-DEK) ou de la moyenne spatiale pondérée des estimations par krigeage (MS-DPEK), selon 12 types de krigeage, pour l'ensemble de la population (T) ou selon le genre (H pour hommes, F pour femmes) dans chacune des régions sociosanitaires (RSS) étudiées (06, 03, 02).

Les modèles ayant une déviance expliquée plus élevée que le modèle de référence y sont identifiés en gris et ceux avec la déviance expliquée la plus élevée en vert.

RSS-06

MS-DEK vs MS-DS		MS-DEK ^C selon les 12 types de krigeage ^D													
S ^A	P ^B	MS-DS ^C	CIN	SIN	GIN	CAN	SAN	GAN	CIL	SIL	GIL	CAL	SAL	GAL	
A	T	11,7	11,7	10,9	11,1	11,1	11,7	10,9	12,0	11,1	10,8	11,7	11,7	11,7	
	H	5,7	6,2	5,8	5,8	5,8	6,1	5,7	6,5	5,9	5,6	6,2	6,2	6,2	
	F	6,6	6,9	6,3	6,4	6,8	6,4	6,2	6,9	6,2	6,4	6,9	6,9	6,9	
É	T	6,8	7,6	7,5	6,4	8,5	8,0	6,3	8,1	7,5	6,6	7,6	7,6	7,6	
	H	4,2	4,3	4,5	4,5	4,3	4,1	4,4	4,4	4,4	4,3	4,3	4,3	4,3	
	F	3,3	4,7	3,9	4,2	4,3	4,2	4,2	4,4	3,8	3,9	4,7	4,7	4,7	
H	T	10,2	10,4	9,6	8,9	8,8	11,1	9,5	11,6	9,0	8,2	10,4	10,4	10,4	
	H	4,6	5,2	4,0	3,8	5,5	5,5	6,8	4,7	4,3	4,2	5,2	5,2	5,2	
	F	7,9	8,1	7,3	7,4	6,5	8,0	6,7	9,0	7,3	7,4	8,1	8,1	8,1	

MS-PEK vs MS-DS		MS-PEK ^C selon les 12 types de krigeage ^D													
S ^A	P ^B	MS-DS ^C	CIN	SIN	GIN	CAN	SAN	GAN	CIL	SIL	GIL	CAL	SAL	GAL	
A	T	11,7	11,7	11,0	11,0	11,0	11,3	10,9	11,7	11,3	10,7	11,7	11,7	11,7	
	H	5,7	6,2	5,9	5,8	5,6	6,2	5,5	6,3	5,9	5,6	6,2	6,2	6,2	
	F	6,6	6,8	6,6	6,5	6,7	6,6	6,3	7,0	6,6	6,4	6,8	6,8	6,8	
É	T	6,8	7,8	7,8	6,2	8,3	7,5	6,4	8,2	7,7	7,1	7,8	7,8	7,8	
	H	4,2	4,2	4,5	4,4	4,4	4,3	4,5	4,6	4,4	4,3	4,2	4,2	4,2	
	F	3,3	4,7	4,0	4,2	4,3	4,2	4,2	4,4	3,9	3,9	4,7	4,7	4,7	
H	T	10,2	10,4	9,0	8,9	9,2	10,6	9,7	11,1	8,3	8,2	10,4	10,4	10,4	
	H	4,6	5,2	4,6	3,6	5,7	5,4	7,0	4,7	4,5	4,2	5,2	5,2	5,2	
	F	7,9	8,2	7,1	7,1	6,4	7,0	6,6	8,3	7,1	6,9	8,2	8,2	8,2	

RSS-03

MS-DEK vs MS-DS

S ^A	P ^B	MS-DS ^C	MS-DEK ^C selon les 12 types de krigeage ^D												
			CIN	SIN	GIN	CAN	SAN	GAN	CIL	SIL	GIL	CAL	SAL	GAL	
A	T	3,9	3,7	3,6	3,7	3,7	3,7	3,7	3,7	3,8	3,6	3,8	3,7	3,7	3,7
	H	2,9	2,7	2,7	2,8	2,8	2,7	2,7	2,7	2,6	2,7	2,5	2,7	2,7	2,7
	F	1,8	2,4	2,2	2,3	2,3	2,2	2,4	2,2	2,2	2,2	2,3	2,4	2,4	2,4
É	T	3,0	2,7	3,0	3,2	3,0	3,1	3,0	2,8	3,0	2,8	2,7	2,7	2,7	2,7
	H	1,8	2,9	2,8	3,4	3,0	2,8	2,6	2,9	2,8	2,8	2,9	2,9	2,9	2,9
	F	2,3	2,6	2,8	3,3	2,4	2,9	2,8	2,5	2,8	3,0	2,6	2,6	2,6	2,6
H	T	4,3	4,2	4,0	5,5	5,0	4,4	5,6	5,1	4,0	4,4	4,2	4,2	4,2	4,2
	H	3,4	4,1	4,6	4,4	4,6	4,7	4,4	4,1	4,7	4,4	4,1	4,1	4,1	4,1
	F	2,3	2,7	1,9	2,7	2,6	2,1	2,5	3,8	2,1	2,2	2,7	2,7	2,7	2,7

MS-PEK vs MS-DS

S ^A	P ^B	MS-DS ^C	MS-PEK ^C selon les 12 types de krigeage ^D												
			CIN	SIN	GIN	CAN	SAN	GAN	CIL	SIL	GIL	CAL	SAL	GAL	
A	T	3,9	3,7	3,7	3,8	3,6	3,7	3,7	3,7	3,7	3,7	3,7	3,7	3,7	3,7
	H	2,9	2,7	2,8	2,8	2,6	2,8	2,8	2,7	2,8	2,5	2,7	2,7	2,7	2,7
	F	1,8	2,4	2,2	2,3	2,2	2,2	2,3	2,3	2,2	2,3	2,4	2,4	2,4	2,4
É	T	3,0	2,4	2,9	3,1	2,8	3,0	3,5	2,5	3,0	2,8	2,4	2,4	2,4	2,4
	H	1,8	2,5	2,3	2,6	2,9	2,1	1,8	2,1	2,4	2,7	2,5	2,5	2,5	2,5
	F	2,3	2,6	2,7	3,2	2,4	2,8	2,9	2,5	2,7	3,1	2,6	2,6	2,6	2,6
H	T	4,3	4,1	3,6	4,2	4,5	3,9	3,7	3,7	3,6	4,1	4,1	4,1	4,1	4,1
	H	3,4	4,1	4,7	4,4	4,5	4,7	4,6	3,9	4,7	4,1	4,1	4,1	4,1	4,1
	F	2,3	3,1	1,9	2,4	2,6	2,1	2,4	2,2	1,9	2,3	3,1	3,1	3,1	3,1

RSS-02
MS-DEK vs MS-DS

S ^A	P ^B	MS-DS ^C	MS-DEK ^C selon les 12 types de krigeage ^D												
			CIN	SIN	GIN	CAN	SAN	GAN	CIL	SIL	GIL	CAL	SAL	GAL	
A	T	1,3	1,1	1,3	1,3	1,3	1,3	1,2	1,2	1,4	1,5	1,1	1,1	1,1	1,1
	H	0,9	0,7	0,9	0,8	0,9	0,9	0,7	0,7	0,8	0,8	0,7	0,7	0,7	0,7
	F	0,8	1,1	1,1	1,0	1,0	1,3	1,3	1,1	1,0	1,1	1,1	1,1	1,1	1,1
É	T	2,5	2,6	2,7	2,4	2,5	2,4	2,5	1,8	2,8	2,3	2,6	2,6	2,6	2,6
	H	2,4	1,9	1,3	1,5	1,8	1,4	1,9	2,4	1,5	2,2	1,9	1,9	1,9	1,9
	F	0,9	1,3	1,6	2,0	1,9	2,5	2,3	1,1	1,9	1,1	1,3	1,3	1,3	1,3
H	T	1,8	1,8	1,7	1,5	1,2	1,4	1,3	2,1	2,0	1,9	1,8	1,8	1,8	1,8
	H	1,9	3,4	1,9	1,7	1,4	2,7	2,1	3,6	1,9	1,8	3,4	3,4	3,4	3,4
	F	1,3	1,7	1,6	1,7	2,0	1,6	1,6	1,8	1,6	1,6	1,7	1,7	1,7	1,7

MS-PEK vs MS-DS														
S ^A	P ^B	MS-DS ^C	MS-PEK ^C selon les 12 types de krigeage ^D											
			CIN	SIN	GIN	CAN	SAN	GAN	CIL	SIL	GIL	CAL	SAL	GAL
A	T	1,3	1,1	1,3	1,4	1,3	1,3	1,2	1,5	1,4	1,4	1,1	1,1	1,1
	H	0,9	0,8	1,0	0,7	0,9	0,8	0,7	0,8	0,8	0,8	0,8	0,8	0,8
	F	0,8	1,1	1,0	1,0	1,0	1,3	1,2	1,1	1,0	1,0	1,1	1,1	1,1
É	T	2,5	1,8	2,5	2,3	2,2	2,4	2,1	1,7	2,5	1,9	1,8	1,8	1,8
	H	2,4	1,8	1,7	1,3	1,8	1,4	2,2	1,7	1,6	2,3	1,8	1,8	1,8
	F	0,9	1,3	1,5	1,4	2,0	2,2	1,9	1,4	1,5	1,3	1,3	1,3	1,3
H	T	1,8	3,1	2,6	1,8	1,2	2,4	1,4	3,6	2,2	1,3	3,1	3,1	3,1
	H	1,9	3,6	2,9	1,6	1,3	3,1	1,8	4,4	3,0	1,7	3,6	3,6	3,6
	F	1,3	3,0	1,6	1,9	1,9	1,6	1,5	2,0	1,6	1,7	3,0	3,0	3,0

^A S : saisons étudiées, A pour annuel, É pour été et H pour hiver.

^B P : populations, T pour total, H pour hommes et F pour femmes.

^C MS-DS : moyenne spatiale directe des stations. MS-DEK : moyenne spatiale « directe » des estimations par krigeage. MS-PEK : moyenne spatiale « pondérée » des estimations par krigeage.

^D Les acronymes correspondants aux types de krigeage sont obtenus en combinant la 1^{re} lettre de chaque paramètre : 1^{re} lettre : Cauchy, Gaussien ou Sphérique ; 2^e lettre : Isotrope ou Anisotrope ; 3^e lettre : Non limité ou Limité.

Annexe 7 Phase 2 de l'analyse

Les résultats des modèles incluant la température maximale ou l'humidité maximale sont présentés au tableau suivant pour les régions sociosanitaires (RSS) étudiées (06, 03, 02).

Les modèles utilisant les techniques de pondération ont été comparés à la moyenne spatiale directe des stations (MS-DS, modèle de référence). Ces techniques de pondération sont l'analyse en composantes principales (ACP), la déviance expliquée des régressions individuelles par station (RS-DEX), et le critère Akaike des régressions individuelles par station (RS-AIC).

Les modèles plus satisfaisants que le modèle de référence sont indiqués en gris et ceux avec la déviance expliquée la plus élevée en vert.

Température maximale

T ^A	AT ^B	RSS-06				RSS-03				RSS-02			
		MS-DS	ACP	RS-DEX	RS-AIC	MS-DS	ACP	RS-DEX	RS-AIC	MS-DS	ACP	RS-DEX	RS-AIC
T0	AT1	8,7	8,7	8,7	8,7	2,8	2,8	2,8	2,8	0,7	0,7	0,7	0,7
T0	AT7	23,3	23,3	23,3	23,3	15,4	15,4	15,4	15,4	5,2	5,2	5,2	5,2
T2	AT1	22,0	22,0	22,0	22,0	5,2	5,2	5,2	5,2	1,5	1,5	1,5	1,5
T2	AT7	62,1	62,1	62,1	62,1	29,9	29,9	29,9	29,9	12,4	12,4	12,4	12,4

Humidité maximale

TA	ATB	RSS-06				RSS-03				RSS-02			
		MS-DS	ACP	RS-DEX	RS-AIC	MS-DS	ACP	RS-DEX	RS-AIC	MS-DS	ACP	RS-DEX	RS-AIC
T0	AT1	1,4	1,4	1,6	1,4	0,9	0,9	1,0	0,9	0,1	0,1	0,3	0,1
T0	AT7	4,7	4,7	5,3	4,7	6,2	6,2	6,0	6,2	1,6	1,8	2,5	1,6
T2	AT1	21,1	21,1	21,1	21,1	5,2	5,2	5,2	5,2	1,5	1,5	1,5	1,5
T2	AT7	61,5	61,5	61,5	61,5	29,5	29,5	29,5	29,5	12,2	12,6	12,5	12,3

^A **T0** : Aucune variable « temps » dans le modèle GAM. **T2** : les deux variables « temps » sont dans le modèle GAM.

^B **AT1** : Aucune agrégation temporelle n'a été faite. **AT7** : l'agrégation temporelle sur 7 jours a été faite.

Comparaison de méthodes d'agrégation spatiale de valeurs météorologiques à des fins de surveillance en santé publique

AUTEURS

Jean-Xavier Giroux
Fateh Chebana
Diane Bélanger
Erwan Gloaguen
Taha Ouarda
André St-Hilaire

Institut de la recherche scientifique – Centre Eau Terre Environnement, Québec
Centre de recherche du Centre hospitalier universitaire de Québec, Québec

RELECTURE DU DOCUMENT

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

MISE EN PAGE

Nabila Haddouche
Bureau d'information et d'études en santé des populations,
Institut national de santé publique du Québec

ORGANISME SUBVENTIONNAIRE

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

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

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

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

Dépôt légal – 2^e trimestre 2019
Bibliothèque et Archives nationales du Québec
ISBN : 978-2-550-83904-0 (PDF)

© Gouvernement du Québec (2019)

N° de publication : 2539