

Université de Montréal

**Développement et évaluation d'approches géostatistiques à  
l'échelle urbaine pour l'estimation de l'exposition aux  
particules fines et à l'ozone troposphérique**

par

Yuddy Ramos

Département de géographie

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)  
en géographie

Décembre, 2016

© Yuddy Ramos, 2016.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

Développement et évaluation d'approches géostatistiques à l'échelle urbaine pour l'estimation  
de l'exposition aux particules fines et à l'ozone troposphérique

Présentée par :

Yuddy Ramos

a été évaluée par un jury composé des personnes suivantes :

Liliana Pérez, présidente-rapporteuse

Benoît St-Onge, directeur de recherche

Audrey Smargiassi, codirectrice de recherche

Nathalie Barrette, membre du jury

Marie-Soleil Cloutier, examinatrice externe

Pierre Bergeron, représentante du doyen de la FESP



# Résumé

La pollution atmosphérique constitue un risque environnemental majeur dont les effets néfastes sur la santé et sur l'environnement sont déjà clairement démontrés. Toutefois, la mesure d'exposition des populations aux polluants tels que les particules fines et l'ozone troposphérique demeure approximative en raison de la faible densité des stations d'échantillonnage de ces polluants. Peu d'études ont considéré la variation spatiale intra-urbaine dans la modélisation spatiale des concentrations de polluants. Certaines études ont cependant combiné interpolation spatiale et corrélation avec des facteurs locaux. De plus, l'effet du régime météorologique (par exemple l'occurrence d'une inversion de température) sur l'amplitude de ces corrélations n'est pas pris en compte. Cette thèse a donc pour objectif d'évaluer de nouvelles manières de caractériser la distribution spatiale et temporelle des particules fines ( $PM_{2.5}$ ) et de l'ozone troposphérique ( $O_3$ ) à l'échelle intra-urbaine. Plus particulièrement, nous avons développé un modèle de géostatistique multivariable appelée krigeage avec dérive externe (KED, kriging with external drift) basé sur l'intégration de variables auxiliaires dans le processus d'estimation journalière des  $PM_{2.5}$  et de l' $O_3$ . Le krigeage constitue une forme d'interpolation spatiale des données de stations de mesures éparées, alors que la dérive externe mise sur des corrélations entre des conditions locales (axes de transport routier, espaces verts, etc.) et la concentration des polluants atmosphériques. Afin de prendre en compte les variations temporelles, notamment celles liées aux conditions météorologiques, ces modèles ont été développés par groupes basés sur des conditions synoptiques et six classes d'états établies selon la température, le vent, l'humidité relative et les précipitations, d'après des données météorologiques journalières.

Les résultats montrent que l'intégration des variables auxiliaires telles que la densité de la végétation et les zones des activités industrielles locales dans le KED expliquent en partie les variations intra-urbaines des  $PM_{2.5}$  de l'île de Montréal, mais que cet apport est variable selon la classe météorologique. Ainsi, lorsque les corrélations sont très faibles, une interpolation spatiale simple, comme la méthode IDW (inverse distance weighting) est plus exacte que l'interpolation multivariable, alors que pour d'autres conditions synoptiques le KED produit les prédictions les plus certaines. Nous avons pour cette raison proposé un modèle d'interpolation hybride (KED-IDW) s'adaptant aux conditions météorologiques. Nous avons également montré, particulièrement dans le cas de l' $O_3$ , que le krigeage avec dérive externe améliore les résultats obtenus par krigeage ordinaire (sans variables auxiliaires).

Cette thèse a aussi permis d'évaluer l'apport d'un modèle spatio-temporel (BME, bayesian maximum entropy) dans l'estimation de l'effet à court terme de l'exposition à l' $O_3$  sur les décès à Montréal. Les résultats suggèrent que ce modèle spatio-temporel dans les conditions développées (par ex. basé sur les données de 12 stations de mesures, pour un territoire de 1 310 km<sup>2</sup>) n'apporte pas de gains significatifs dans l'estimation de l'effet de l'exposition.

Dans l'ensemble, cette thèse contribue au progrès de modélisation spatiale empirique des polluants atmosphériques en se fondant notamment sur l'adaptation aux conditions météorologiques et par l'ajout de certains facteurs météorologiques comme prédicteurs. Dans ce contexte, cette thèse ouvre une voie prometteuse pour l'amélioration des estimations de

polluants atmosphériques à l'échelle intra-urbaine et de la capacité à évaluer les risques à la santé de la population par une meilleure caractérisation de l'exposition.

**Mots-clés** : pollution de l'air, particules fines, ozone troposphérique, santé, géostatistique, krigeage avec dérive externe, environnement urbain.

## Abstract

Air pollution is a major environmental hazard with clearly demonstrable adverse effects on health and the environment. However, the measurement of populations' exposure to pollutants such as particulate matter and ground-level ozone remains approximate due to the low density of sampling stations for these pollutants. Moreover, intra-urban spatial variation in the spatial modeling of pollutant concentrations has received little research attention. If anything, some studies have combined spatial interpolation and correlation with local factors; however, they do so without examining the effect of the weather regime (e.g., a temperature inversion) on the magnitude of these correlations. In order to overcome these shortcomings, this dissertation aims to evaluate new ways of characterizing the spatial and temporal distribution of fine particles ( $PM_{2.5}$ ) and ground-level ozone ( $O_3$ ) at the intra-urban scale. In particular, we developed a multivariable geostatistical model called "kriging with external drift" (KED) based on the integration of auxiliary variables into the process of estimating daily  $PM_{2.5}$  and  $O_3$  concentrations. Kriging is a form of spatial interpolation of data from measurement stations that are dispersed, while external drift is based on correlations between local conditions (road transport arteries, green spaces, etc.) and the concentration of atmospheric pollutants. In order to take account of temporal variations, especially those related to weather conditions, we designed these models around six synoptic weather classes based on daily meteorological data (such as temperature, wind, relative humidity and precipitation).

The results show that the integration of auxiliary variables (such as vegetation density and local industrial activity areas) in KED partly explains the intra-urban variations of PM<sub>2.5</sub> on the island of Montreal, but that this contribution is variable depending on the weather conditions. Thus, when the correlations are very low, a simple spatial interpolation (such as the inverse distance weighting (IDW) method) is more accurate than multivariable interpolation, whereas for other synoptic conditions KED produces the most certain predictions. For this reason, we proposed a hybrid interpolation model (KED-IDW) that can adapt to different weather conditions. We have also shown, particularly in the case of O<sub>3</sub>, that KED improves the results obtained from ordinary kriging (without auxiliary variables).

This dissertation also allowed to evaluate the contribution of a spatial-temporal model—BME (bayesian maximum entropy)—in the estimation of the short-term effect of exposure to O<sub>3</sub> on deaths in Montreal. The results suggest that this spatio-temporal model under the determined conditions (e.g., based on data from 12 measurement stations, for a territory of 1 310 km<sup>2</sup>) does not offer significant improvements to the estimation of the effect of exposure.

Overall, this dissertation contributes to the advancement of the empirical spatial modeling of air pollutants, namely by taking into account the adaptation to weather conditions as well as certain predictive meteorological factors. In this context, the dissertation opens up a promising path for improving the estimation of air pollution at the intra-urban scale and the capacity to assess population health risks through better characterization of exposure.



**Keywords** : air pollution, fine particles, ground-level ozone, health, geostatistics, kriging with external drift (KED), urban environment.

# Table des matières

Résumé.....	i
Abstract.....	iv
Table des matières.....	vii
Liste des tableaux.....	xi
Liste des figures.....	xiii
Liste des sigles et abréviations.....	xvi
Remerciements.....	xxi
Chapitre 1 : Mise en contexte.....	23
1    Introduction générale.....	25
1.1    La pollution de l'air.....	26
1.2    Particules fines (PM <sub>2,5</sub> ).....	27
1.2.1    Sources d'émission.....	29
1.2.2    Facteurs météorologiques.....	30
1.3    Ozone troposphérique (O <sub>3</sub> ).....	31
1.3.1    Sources d'émission.....	32
1.3.2    Facteurs météorologiques.....	33
1.4    Effets sur la santé des PM <sub>2,5</sub> et de l'O <sub>3</sub> .....	34
1.4.1    Mortalité reliée à l'exposition aigüe.....	34
1.4.2    Effets respiratoires aigus.....	35
1.4.3    Effets cardiovasculaires aigus.....	35

1.5	Mesures et modélisation de l'exposition.....	36
1.5.1	Mesures.....	36
1.5.2	Les méthodes d'interpolation spatiales.....	38
1.5.3	Méthodes de modélisation utilisées en épidémiologie.....	41
2	Hypothèses de recherche.....	46
3	Objectifs de recherche.....	47
3.1	Objectif général.....	47
3.2	Objectifs spécifiques.....	47
Chapitre 2 : Méthodologie .....		51
1.	KED ( <i>Kriging with external drift</i> ) .....	53
1.1	Données sur la qualité de l'air et météorologiques.....	53
1.2	Groupes météorologiques synoptiques .....	54
1.3	Variables externes.....	54
1.3.1	Variables météorologiques spatio-temporelles.....	55
1.3.2	Variables environnementales et géographiques.....	55
1.4	Modélisation spatio-temporelle .....	57
1.5	Validation.....	59
2.	BME ( <i>Bayesian maximum entropy</i> ).....	61
Chapitre 3 : Spatio-temporal models to estimate daily concentrations of fine particulate matter in Montreal: Kriging with external drift and inverse distance weighted approaches – Article 1 .....		64
2	Abstract.....	68
3	Introduction.....	69

4	Methods.....	71
5	Results.....	82
6	Discussion.....	85
7	Conclusions.....	88
8	Conflict of interest .....	88
9	Acknowledgements.....	89
10	References.....	90
11	Tables and figures.....	98
Chapitre 4 : Spatio-temporal modeling of daily concentrations of ground-level ozone in Montreal, Canada – Article 2..... 109		
1	Abstract.....	113
2	Introduction.....	114
3	Methods.....	117
4	Results.....	130
5	Discussion and conclusions .....	133
6	Acknowledgements.....	137
7	References.....	138
8	Tables and figures.....	148
Chapitre 5 : Association of mortality with short-term exposures to ground-level ozone in Montreal, Canada: A comparison of risk estimation from non-spatial and spatio-temporal exposure approaches – Article 3..... 157		
1	Abstract.....	161
2	Introduction.....	163

3	Methods.....	166
4	Results.....	171
5	Discussion.....	172
6	Conclusions.....	175
7	Conflict of interest .....	175
8	References.....	176
9	Tables and figures.....	184
	Chapitre 6 : Discussion et conclusions .....	190
	Bibliographie.....	203
	Annexe 1: Approbation éthique.....	220

# Liste des tableaux

## Chapitre 1

Table 1. Sources primaires des polluants (PM <sub>2.5</sub> et O <sub>3</sub> ) et la relation avec variables externes potentielles (variables géographiques-environnementales).....	45
--	----

## Chapitre 3

Table 1. Descriptive statistics for meteorological variables and PM <sub>2.5</sub> for 2008–2010* .....	98
Table 2. Pearson’s correlation coefficients between PM <sub>2.5</sub> concentrations and independent geographic variables by synoptic weather class for the year 2010.....	99
Table 3. Pearson’s correlation coefficients and coefficients of variation based on daily measurements at NAPS station locations between PM <sub>2.5</sub> concentrations and spatio-temporal meteorological variables for the year 2010.....	100
Table 4. Leave-one-station-out cross-validation results for the PM <sub>2.5</sub> prediction for the year 2010 from KED, IDW and KED–IDW models.....	101

## Chapitre 4

Table 1. Descriptive statistics of the six synoptic weather classes for O <sub>3</sub> and meteorological variables for 2008–2010. ....	148
Table 2. Median values of daily Pearson’s correlation coefficients between O <sub>3</sub> concentrations and independent geographic and meteorological covariates for the year 2010.....	149

Table 3. Leave-one-station-out cross-validation results for the O <sub>3</sub> prediction for the year 2010 from the KED with vegetation and road density as covariates, OK, Daily mean and BME model (Adam-Poupart <i>et al.</i> , 2014). .....	150
--	-----

## Chapitre 5

Table 1. Descriptive statistics of causes of deaths, ground-level ozone and temperature in the Montreal region for summer months (May through September) from 1996 to 2007 (N cases and controls = 435 520 and N days of measurements = 1 734). .....	184
---	-----

Table 2. Risk of non-accidental mortality for increments of IQR of daily O <sub>3</sub> levels evaluated at lag 0 and lag 1 day, in Montreal region for summer months (May through September) from 1996 to 2007. ....	185
---	-----

Table 3. Summary of spatial and temporal variation in the O <sub>3</sub> dataset: Spatial variation relate to the average of O <sub>3</sub> (ppb) daily standard deviation over all monitoring sites and the temporal variation relate to standard deviation of O <sub>3</sub> (ppb) daily means over all days for the study period (May through September, 1996 to 2007). .....	186
--	-----

# Liste des figures

## Chapitres 1 et 2

Figure 1. Sources d'émission des particules fines (PM <sub>2.5</sub> ) en 2013 : (A) Québec, (B) Canada. .....	30
Figure 2. Site d'étude et localisation des stations météorologiques et du réseau national de surveillance de la pollution atmosphérique (RNSPA). .....	38
Figure 3. Méthodes d'estimation des concentrations de polluants atmosphériques. MODIS et CALIPSO sont des satellites; MERRAero est un modèle de la NASA ( <i>National Aeronautics and Space Administration</i> ) et GEM-MACH est un modèle d'Environnement Canada. ....	41
Figure 4. Diagramme méthodologique pour l'application du krigeage avec dérive externe (KED). .....	60

## Chapitre 3

Figure 1. Geographical location of PM <sub>2.5</sub> stations and meteorological stations in the study region. ....	102
Figure 2. PM <sub>2.5</sub> mapping error estimates from cross validation [where error = predicted – observed PM <sub>2.5</sub> concentration (in µg/m <sup>3</sup> ) at each monitoring station] based in KED (mean ± SD; -0.22 ± 2.78 µg/m <sup>3</sup> ), IDW (mean ± SD; 0.42 ± 2.21 µg/m <sup>3</sup> ) and KED-IDW (mean ± SD; 0.24 ± 2.13 µg/m <sup>3</sup> ) for days of the year 2010. ....	103



Figure 3. Daily temporal PM<sub>2.5</sub> error estimates (RMSE) based on leave-one-station cross-validation for KED, IDW (A) and KED-IDW models (B)..... 104

Figure 4. PM<sub>2.5</sub> daily levels predicted with the hybrid model KED-IDW vs. PM<sub>2.5</sub> concentrations measured at NAPS stations for the year 2010 by synoptic weather class. .... 105

Figure 5. Spatial distribution of mean PM<sub>2.5</sub> error estimates (RMSE) of the year 2010 based on cross-validation for KED (A), IDW (B) and KED-IDW (C)..... 106

Figure 6. Predicted PM<sub>2.5</sub> concentrations for April 22nd, 2010: (A) KED (covariates: wind speed and greenness) and (B) IDW. Predicted PM<sub>2.5</sub> concentrations for July 16, 2010: (C) KED (covariates: wind speed and industrial area) and (D) IDW. .... 107

#### **Chapitre 4**

Figure 1. Geographical location of O<sub>3</sub> and meteorological stations in the study region. .... 151

Figure 2. Wind roses (based on 8 meteorological stations) by synoptic weather classes (6) generated using hourly wind speed and direction data of the year 2010. .... 152

Figure 3. O<sub>3</sub> mapping error estimates from cross validation [where error = predicted – observed O<sub>3</sub> concentration (in ppb) at each monitoring station] based on the KED (mean ± SD; 0.29 ± 3.19 ppb), OK (mean ± SD; -0.13 ± 4.30 ppb), Daily mean (mean ± SD; 0.00 ± 4.58 ppb) and BME (Adam-Poupart *et al.*, 2014) (mean ± SD; 0.63 ± 7.04 ppb) models and for the summer months of the year 2010..... 153

Figure 4. Daily temporal O<sub>3</sub> error estimates (RMSE) based on leave-one-station cross-validation for the KED, OK, Daily mean and BME (Adam-Poupart *et al.*, 2014) models for May through September of 2010. .... 154

Figure 5. Mean O<sub>3</sub> levels (0900–1700 hours) predicted (A), standard error - SE (B) and relative prediction error (C) from the KED model at a 100-m grid across the Montreal Census Metropolitan Area (Canada) for the year 2010. .... 155

## **Chapitre 5**

Figure 1. Study area: Montreal census metropolitan area, Québec, Canada. .... 187

Figure 2. Daily number of all non-accidental deaths in Montreal CMA during summer months (May through September) from the years 1996 to 2007. .... 188

# Liste des sigles et abréviations

## En français :

BC : Noir de carbone

CH<sub>4</sub> : Méthane

CO : Monoxyde de carbone

COV : Composés organiques volatils

EC : Carbone élémentaire

INRP : Inventaire national des rejets de polluants

INSPQ : Institut national de santé publique du Québec

RNSPA : Réseau national de surveillance de la pollution atmosphérique

Na<sup>+</sup> : Sodium

NH<sub>4</sub><sup>+</sup> : Ammonium

NH<sub>4</sub>NO<sub>3</sub> : Nitrate d'ammonium

NO : Monoxyde d'azote

NO<sub>2</sub> : Dioxyde d'azote

NO<sub>3</sub><sup>-</sup> : Nitrate

NO<sub>x</sub> : Oxydes d'azote

O<sub>3</sub> : Ozone troposphérique

OC : Carbone organique

OMS : Organisation mondiale de la santé

PM<sub>2.5</sub> : Particules fines

PM<sub>10</sub> : Particules d'un diamètre égal ou inférieur à 10 µm

ppb : partie par milliard

Si : Silicium

SO<sub>2</sub> : Dioxyde de soufre

SO<sub>4</sub><sup>2-</sup> : Sulfate

SO<sub>x</sub> : Oxydes de soufre

µg : microgramme

**En anglais:**

AIC : Akaike's Information Criterion

BC : Black carbon

BME : Bayesian maximum entropy

CI : Confidence interval

CH<sub>4</sub> : Methane

CO : Carbon monoxide

COPD : Chronic obstructive pulmonary disease

EC : Elemental carbon

ED : Emergency department

HCA : Hierarchical cluster analysis

IDW : Inverse distance weighting

IQR : Interquartile range

KED : Kriging with external drift

NAPS : National air pollution surveillance

NPRI : National pollutant release inventory

Na<sup>+</sup> : Sodium

NH<sub>4</sub><sup>+</sup> : Ammonium

NH<sub>4</sub>NO<sub>3</sub> : Ammonium nitrate

NO : Nitrogen monoxide

NO<sub>2</sub> : Nitrogen dioxide

NO<sub>3</sub><sup>-</sup> : Nitrate

NO<sub>x</sub> : Nitrogen oxides

O<sub>3</sub> : Ground-level ozone

OC : Organic carbon

OR : Odds ratio

OK : Ordinary kriging

PCA : Principal Component Analysis

PM<sub>2.5</sub> : Fine particulate matter

PM<sub>10</sub> : Particulate matter that are 10 μm or less in diameter

ppb : parts per billions

RMSE : Root mean squared error

SE : Standard error

Si : Silicon

SO<sub>2</sub> : Sulfur dioxide

SO<sub>4</sub><sup>2-</sup> : Sulfate

SO<sub>x</sub> : Sulfur oxides

U.S. EPA : United States Environmental protection agency

μg : microgram

VOC : Volatile organic compounds

WHO : World health organization

*Je dédie cette thèse :*

*À mes parents pour leur amour et pour m'avoir appris la  
persévérance.*

*À mon mari Éric pour son amour et son soutien.*

*À mon petit Edward à qui je souhaite un monde sans pollution.*

## Remerciements

Cette recherche n'aurait pas été possible sans le soutien et la confiance de plusieurs personnes. Les mots ne suffisent pas pour exprimer toute ma gratitude.

Je tiens à remercier mon directeur de recherche, Benoît St-Onge, pour avoir accepté de m'encadrer, pour ses conseils précieux et pour sa confiance indéfectible. Je tiens également à remercier ma codirectrice, Audrey Smargiassi, de m'avoir fait une place dans son laboratoire ainsi que sa disponibilité, ses conseils et son soutien financier.

Un grand merci à Jean-Pierre Blanchet pour sa disponibilité, son partage de connaissance et l'intérêt qu'il a porté à mes travaux de recherche. Sans oublier ma reconnaissance à Yan Kestens pour sa confiance et ses conseils.

Je remercie aussi les différents instituts qui ont contribué d'une façon ou autre à la continuité de mes études : à l'Institut national de santé publique du Québec (INSPQ) pour son soutien financier dans la première partie de ma recherche et à l'Institut de recherche en santé publique (IRSPUM) pour m'avoir adoptée comme étudiante. Aussi je tiens à remercier les départements qui m'ont accueillie temporairement dans leurs laboratoires : le département de géographie de l'Université du Québec à Montréal (UQÀM) et le département de santé environnementale et santé au travail de l'Université de Montréal.



Un merci spécial à mes amis qui se reconnaîtront pour avoir toujours été à l'écoute et pour leurs encouragements.

Je tiens à remercier particulièrement ma famille pour son soutien moral et ses encouragements malgré la distance qui nous sépare.

Finalement, un remerciement très spécial à mon mari, pour son encouragement, sa sagesse, sa patience et sa confiance sans faille tout au long de mon doctorat. Sa présence et son soutien affectif ont été pour moi d'un immense réconfort. Sans oublier son aide en statistiques.

## **Chapitre 1 : Mise en contexte**



# 1 Introduction générale

La communauté scientifique est unanime, la pollution atmosphérique constitue un risque majeur pour la santé humaine partout dans le monde. Plusieurs recherches épidémiologiques ont rapporté des associations entre l'exposition aux divers polluants de l'air et des effets néfastes pour la santé, particulièrement l'occurrence et l'exacerbation des maladies pulmonaires et cardiovasculaires (Fuks *et al.*, 2016; Pope III *et al.*, 2002; Von Klot *et al.*, 2005; Zanobetti *et al.*, 2006, 2009). Les polluants dont les effets néfastes pour la santé sont les mieux documentés incluent les particules fines ( $PM_{2.5}$ ) et l'ozone troposphérique ( $O_3$ ) (Brauer *et al.*, 2016). Selon le rapport de l'Organisation mondiale de la santé (OMS), en 2012, la pollution de l'air ambiant a causé 3.7 millions de décès prématurés dans le monde (WHO, 2014). L'étude sur la charge mondiale de morbidité de 2013 (GBD 2013, Global Burden of Disease) estimait à 2.9 millions les décès prématurés dans le monde entraînés par l'exposition aux particules fines ( $PM_{2.5}$ ) et 217 000 décès dus à l'exposition à long terme à l'ozone troposphérique ( $O_3$ ) (Forouzanfar *et al.*, 2015).

Tout récemment, l'OMS dans son rapport intitulé, « Réduire les risques pour la santé mondiale en limitant les polluants climatiques à courte durée » (Reducing global health risks through mitigation of short-lived climate pollutants) souligne l'urgence de réduire ce type d'émissions, c.-à-d. le BC (noir de carbone ou carbone suie, un des composants les plus dangereuses pour la santé des  $PM_{2.5}$ ), l' $O_3$  (ozone troposphérique) et le  $CH_4$  (méthane). Ces trois polluants entraînent en effet un réchauffement climatique important et contribuent de manière substantielle à la charge de morbidité due à pollution atmosphérique (WHO, 2015).

Dans ce contexte, des estimations les plus exactes possibles des concentrations de polluants de l'air ambiant à l'échelle intra-urbaine sont requises afin de mieux connaître les niveaux de pollution auxquels la population est exposée. Depuis quelques années, plusieurs modèles d'interpolation spatiale sont utilisés pour estimer les concentrations de polluants aux endroits non mesurés par les stations fixes de surveillance de la qualité de l'air. Toutefois, la qualité des interpolations univariées peut être limitée par le nombre et de la distribution spatiale des stations de surveillance (Singh *et al.*, 2011). Ce n'est que récemment que l'intégration des variables prédictives dans les modèles d'estimation de la pollution de l'air (Beelen *et al.*, 2009; Hart *et al.*, 2009; Singh *et al.*, 2011) a mis en lumière l'importance des variations spatiales intra-urbaines des polluants dans l'évaluation des effets sur la santé (Brauer *et al.*, 2008; Jerrett *et al.*, 2005). À cet égard, nous proposons dans la présente thèse de nouvelles manières de caractériser la distribution spatiale des polluants. Plus particulièrement, nous proposons un modèle d'interpolation multivariée avec l'intégration de facteurs secondaires corrélés spatialement aux mesures des concentrations de polluants réalisées aux stations de surveillance de qualité de l'air.

## **1.1 La pollution de l'air**

La qualité de l'air se détériore en raison d'un ensemble de polluants atmosphériques qui diffèrent grandement dans leur composition chimique, réactivité, sources d'émission, persistance dans l'environnement, mobilité et répercussion. Ces polluants peuvent être classés en: 1) polluants primaires, d'origine naturelle ou anthropique, qui comprennent les substances émises directement de la source vers l'atmosphère, et 2) polluants secondaires formés dans

l'air suite à l'interaction entre polluants primaires et d'autres composés présents dans l'atmosphère (Colls, 2002). Les  $PM_{2.5}$  sont à la fois d'origine primaire et secondaire alors que l' $O_3$  est d'origine secondaire (Seinfeld *et al.*, 2006).

De plus, la variabilité spatiale des polluants de l'air ambiant peut être influencée par des facteurs météorologiques (Camalier *et al.*, 2007; Cox *et al.*, 1996; Dawson *et al.*, 2007); des facteurs géographiques tels que la topographie, ainsi que l'environnement naturel et bâti (Pinto *et al.*, 2004; U.S. EPA, 2013). Des facteurs comme l'utilisation du sol, la densité locale de la population, celle des routes ou de la végétation sont aussi associés de façon directe ou indirecte aux concentrations des polluants (Matthes *et al.*, 2007; Nowak, 2002; Pearce *et al.*, 2009).

Par ailleurs, on s'attend à ce que les changements climatiques entraînent une détérioration de la qualité de l'air dans le monde (IPCC, 2014). Une hausse des températures pourrait notamment favoriser une augmentation de la pollution atmosphérique causée par l'exacerbation des concentrations de polluants comme l' $O_3$  (OMS, 2010). De plus, cette augmentation de température pourrait aussi modifier la concentration des  $PM_{2.5}$  à travers l'évolution des émissions, les incendies de forêt plus fréquents et la volatilité des aérosols de nitrate (Tai *et al.*, 2012).

## **1.2 Particules fines ( $PM_{2.5}$ )**

Les matières particulaires (PM) font partie de la famille des aérosols et constituent un ensemble très hétérogène; leur nature physique (solide, liquide), chimique (organique,

minérale) ou biologique est en effet très variable selon leurs sources d'émission (Colls, 2002; Hewitt *et al.*, 2009). Dans le cadre de leur surveillance environnementale, les PM sont classées selon leur taille granulométrique en : particules totales (PT) d'un diamètre maximal d'environ 100 µm, particules atmosphériques d'un diamètre médian inférieur à 10 µm (PM<sub>10</sub>), et particules atmosphériques fines d'un diamètre médian inférieur à 2.5 µm (PM<sub>2.5</sub>), incluant les particules ultrafines (nanoparticules) (Environment Canada, 2013a). Les particules fines et ultrafines pénétreraient plus profondément dans les poumons pour atteindre la circulation sanguine et occasionner des effets au-delà de ceux touchant strictement la santé pulmonaire (U.S. EPA, 2009).

La masse totale des PM<sub>2.5</sub> est composée principalement de : sulfate (SO<sub>4</sub><sup>2-</sup>), nitrate (NO<sub>3</sub><sup>-</sup>), ammonium (NH<sub>4</sub><sup>+</sup>), carbone organique (OC), carbone élémentaire (EC) ou noir de carbone (BC), silicium (Si) et sodium (Na<sup>+</sup>) (Bell *et al.*, 2007). Les PM<sub>2.5</sub> primaires sont directement émises dans l'air et les PM<sub>2.5</sub> secondaires sont formées par des réactions physico-chimiques à partir des gaz précurseurs tels que les oxydes de soufre (SO<sub>x</sub>), les oxydes d'azote (NO<sub>x</sub>) et les composés organiques volatils (COV) (U.S. EPA, 2009). Généralement les PM<sub>2.5</sub> sont dominées par les particules secondaires (Pinto *et al.*, 2004). Par ailleurs, plus les particules sont fines, plus elles restent longtemps en suspension dans l'atmosphère, elles voyagent donc sur les longues distances, contribuant à l'augmentation des niveaux de polluants loin des sources d'émissions.

De plus, la nature de la formation des particules secondaires et la durée de vie des PM<sub>2.5</sub> (plusieurs jours) assurent un certain degré d'uniformité spatiale (homogénéité) dans

certaines zones urbaines (U.S. EPA, 2004). Au contraire, les particules primaires émises directement dans l'atmosphère par des sources locales présentent une grande variabilité spatiale (hétérogénéité) (Brauer *et al.*, 2014; Ito *et al.*, 2004; Sororian *et al.*, 2014). En se basant sur des études réalisées dans différentes villes du monde, Brauer *et al.* (2014) ont démontré que la variation spatiale intra-urbaine de la masse totale des PM<sub>2.5</sub> évaluée d'après le coefficient de variation est inférieure à 20% et celles des PM<sub>2.5</sub> primaires est supérieure à 200%. Cependant, la variation spatiale de la masse totale des PM<sub>2.5</sub> peut être supérieure à 20% dans certaines villes hautement polluées par des sources locales comme Detroit, États-Unis (43%), ou Windsor, Canada (40%) (L. Miller *et al.*, 2010).

Par ailleurs, la variabilité spatiale des PM<sub>2.5</sub> dans les zones urbaines peut être influencée (en plus du rôle joué par les facteurs météorologiques et topographiques) par la répartition spatiale des sources locales de particules primaires, les événements d'émissions temporaires (incendies, etc.), et les différences dans le comportement (mobilité, transport et dépôt) de composés semi-volatils (Pinto *et al.*, 2004).

### **1.2.1 Sources d'émission**

Les principales sources locales anthropiques des particules fines primaires sont: la combustion du bois de chauffage, et la combustion de combustibles fossiles (secteur industriel, résidentiel et du transport automobile) (Brauer *et al.*, 2014; U.S. EPA, 2004).

Selon les données de l'Inventaire national des rejets de polluants (INRP), au Québec en 2013, les poussières soulevées sur les routes (92 247 tonnes) constituent la plus grande source



d'émission des  $PM_{2.5}$ , suivis de la combustion de bois de chauffage (72 267 tonnes), des activités de construction (25 831 tonnes), des sources industrielles (12 885 tonnes), du transport (9 192 tonnes) et de l'agriculture (3 633 tonnes). Au Canada pour la même année, les poussières soulevées sur les routes (623 036 tonnes) forment aussi la plus grande source d'émission des  $PM_{2.5}$ , suivis des activités de construction (504 920 tonnes), de la combustion de bois de chauffage (164 707 tonnes), des sources industrielles (66 452 tonnes), du transport (57 020 tonnes) et de l'agriculture (44 359 tonnes) (Environment Canada, 2013b) Figure 1.

Pour ce qui est des sources d'émission des précurseurs de particules secondaires, soit de  $SO_x$ ,  $NO_x$  et  $COV$ , elles sont surtout industrielles pour le premier et associées aux combustibles fossiles pour les deuxième et troisième (Paradis, 2011).

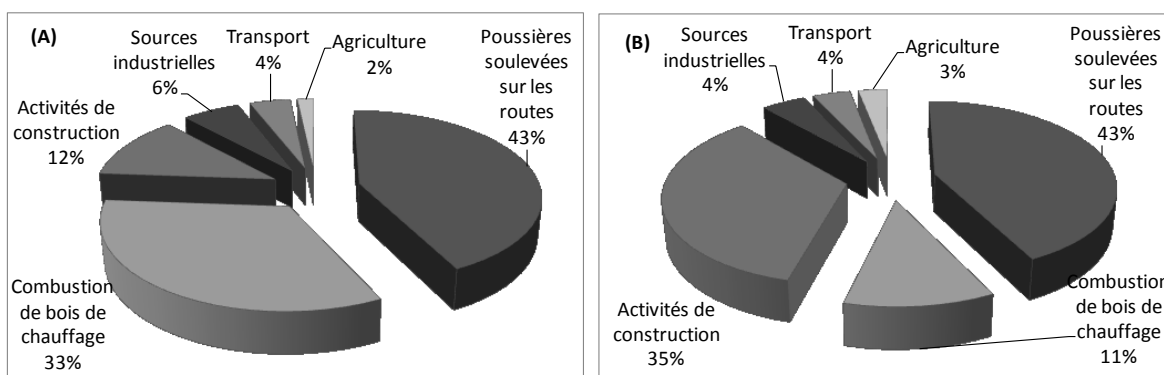


Figure 1. Sources d'émission des particules fines ( $PM_{2.5}$ ) en 2013 : (A) Québec, (B) Canada.

## 1.2.2 Facteurs météorologiques

Les facteurs météorologiques régissent en grande partie la formation, la dispersion, la diffusion, le transport, la transformation physique et chimique et le dépôt par voie sèche et par

voie humide des polluants. Parmi ces facteurs, les plus importants en ce qui a trait aux  $PM_{2.5}$  sont la température, l'humidité relative, les précipitations, le vent et la pression atmosphérique (Tai *et al.*, 2010). Par exemple, la formation du nitrate d'ammonium (composant principal semi-volatil des  $PM_{2.5}$ ) est favorisée par une humidité relative élevée (Tai *et al.*, 2010). Toutefois, une augmentation des précipitations entraîne une déposition humide des  $PM_{2.5}$ , contribuant à une diminution des concentrations par le processus de lessivage (Dawson *et al.*, 2007; Feng *et al.*, 2012; Tai *et al.*, 2010). Par ailleurs, des études réalisées dans le corridor Windsor-Québec qui comprend la région de Montréal suggèrent que les concentrations des  $PM_{2.5}$  dans ce corridor sont éliminées de l'atmosphère principalement par lessivage (dépôt humide) (McMurry *et al.*, 2004).

### **1.3 Ozone troposphérique ( $O_3$ )**

L'ozone troposphérique ( $O_3$ ) est un gaz (polluant secondaire) de la basse atmosphère qui provient des sources anthropiques et naturelles. L' $O_3$  est produit par des réactions photochimiques entre certains polluants dits primaires comme, les oxydes d'azote ( $NO_x$ ) et les composés organiques volatils (COV) en présence du rayonnement solaire (McKendry, 1993; U.S. EPA, 2013). Les  $NO_x$  sont les formes oxydées de l'azote et comprennent principalement, le monoxyde d'azote (NO) et le dioxyde d'azote ( $NO_2$ ).

L' $O_3$  est produit par la photodissociation de  $NO_2$  en présence du rayonnement solaire. Le NO formé à partir de cette réaction détruit l' $O_3$  pour redonner  $NO_2$  (Seinfeld, 1989). Ceci explique pourquoi, les maxima de concentration d' $O_3$  ne se trouvent généralement pas au sein des agglomérations, malgré qu'on y trouve les sources les plus intenses des émissions de NO.

L'O<sub>3</sub> dont la durée de vie est de quelques semaines (U.S. EPA, 2013) est l'un des principaux constituants du smog (INSPQ, 2012).

La variation spatiale de l'O<sub>3</sub> est généralement assez homogène et présente donc une autocorrélation spatiale élevée (Sororian *et al.*, 2014). En se basant sur des études réalisées dans différentes villes du monde, Brauer *et al.* (2014) a aussi démontré que la variation spatiale intra-urbaine de l'O<sub>3</sub> (< 65%) évaluée d'après le coefficient de variation est plus grande en comparaison avec la variation spatiale de la masse totale des PM<sub>2.5</sub> (< 20%), mais beaucoup plus faible que la variation spatiale des PM<sub>2.5</sub> primaires (> 200%) émises directement issues des sources de pollution.

En se basant sur la moyenne de l'écart-type journalier de l'O<sub>3</sub> sur tous les sites de surveillance (variabilité spatiale) et l'écart-type des moyennes journalières de l'O<sub>3</sub> sur tous les jours (variabilité temporelle), D. Lee *et al.* (2010) pour la région métropolitaine de Londres rapportent quant à eux une variabilité spatiale égale à 9.80 ppb et une variabilité temporelle égale à 19.40 ppb. Ceci signifie que dans ce cas, la variabilité spatiale de l'O<sub>3</sub> est non négligeable et représente environ la moitié de sa variabilité temporelle.

### **1.3.1 Sources d'émission**

Les NO<sub>x</sub> sont émis à partir de sources à combustibles fossiles, tels que les centrales électriques et le transport automobiles, et réagissent avec les COV provenant de sources telles que les voitures, les industries et certaines espèces végétales (COV biogènes naturellement rejetés dans l'atmosphère), dont des arbres émetteurs d'isoprène et de monoterpènes (U.S.

EPA, 2013). Les émissions naturelles d'isoprènes sont les principaux contributeurs du COV biogène (U.S. EPA, 2013). Parmi les principaux genres d'arbres émetteurs d'isoprène se trouvent le peuplier (*Populus* spp. - poplar), le chêne (*Quercus* spp. - oak), le robinier faux-acacia (*Robinia* spp. - black locust) et le saule (*Salix* spp. - willow) (Nowak, 2002). Du côté des arbres émetteurs de monoterpènes, on trouve principalement le pin (*Pinus* spp. - pine). Selon la base de données d'arbres publics, ces genres d'arbres se retrouvent notamment sur le territoire de la Ville de Montréal (Ville de Montréal, 2016).

Les émissions d'O<sub>3</sub> au Québec sont en partie attribuables aux polluants précurseurs provenant des régions fortement industrialisées localisées le long des rives des Grands Lacs et qui sont transportées sur de grandes distances le long du corridor Windsor-Québec. Les émissions d'O<sub>3</sub> sont aussi attribuables aux polluants précurseurs locaux des grandes villes du Québec (McKendry, 1993).

### **1.3.2 Facteurs météorologiques**

Les concentrations d'O<sub>3</sub> dans l'air ambiant sont conditionnées par les facteurs météorologiques tels que la température, la radiation solaire, la vitesse du vent, et la humidité relative (Camalier *et al.*, 2007; Cox *et al.*, 1996; U.S. EPA, 2013; Wise *et al.*, 2005). Les épisodes aigus d'O<sub>3</sub> sont souvent associés à des conditions anticycloniques (zones de haute pression) caractérisées par des temps ensoleillés, secs sans nuages avec des conditions de stagnation qui empêchent la dispersion et la dissolution des polluants. Ces conditions anticycloniques dans l'est de l'Amérique du Nord et du Canada facilitent le transport sur de longues distances de l'O<sub>3</sub> et de ses précurseurs lorsque les vents en direction ouest et sud-ouest

sont prédominants (McKendry, 1993). D'ailleurs, les précipitations entraînent une déposition des précurseurs solubles, contribuant à une réduction des concentrations d'O<sub>3</sub> (Hou *et al.*, 2015).

## **1.4 Effets sur la santé des PM<sub>2.5</sub> et de l'O<sub>3</sub>**

Les effets de la pollution de l'air sur la santé ont été analysés à travers de nombreuses études. Ces effets peuvent être aigus pour une exposition à court terme ou chroniques pour une exposition à long terme. Pour ce qui est de l'exposition à court terme (ce qui fait l'objet de la présente recherche doctorale), plusieurs études (Stieb *et al.*, 2009; Strickland *et al.*, 2010; U.S. EPA, 2009, 2013) suggèrent que les niveaux journaliers des PM<sub>2.5</sub> et d'O<sub>3</sub> sont associés à des symptômes respiratoires, des visites aux urgences, des hospitalisations et à la mortalité causée par de troubles cardio-respiratoires. De plus, certaines populations sont identifiées comme étant plus vulnérables à l'exposition aux polluants de l'air, notamment les personnes souffrant des maladies chroniques, les personnes âgées et les enfants (U.S. EPA, 2009, 2013).

### **1.4.1 Mortalité reliée à l'exposition aigüe**

Les relations entre les concentrations journalières de PM<sub>2.5</sub> et O<sub>3</sub> et la mortalité sont bien documentées dans la littérature (Franklin *et al.*, 2007; Goldberg *et al.*, 2001; Jerrett *et al.*, 2009; Zanobetti *et al.*, 2009). Dans le cas des PM<sub>2.5</sub>, des études ont montré qu'une exposition de courte durée aux PM<sub>2.5</sub> a des effets sur la mortalité entraînée par des problèmes cardiovasculaires (U.S. EPA, 2009). Toutefois et selon l'U.S. EPA (2009), dans le cas de la mortalité découlant de problèmes respiratoires une relation serait plausible, mais elle ne serait

pas nécessairement causale comme c'est le cas pour la mortalité due à des problèmes cardiovasculaires.

Concernant l'O<sub>3</sub>, plusieurs études épidémiologiques suggèrent que l'exposition de courte durée à ce polluant a des effets sur la mortalité non accidentelle et cardio-pulmonaire (Bell *et al.*, 2005; U.S. EPA, 2013; Zanobetti *et al.*, 2008).

#### **1.4.2 Effets respiratoires aigus**

L'exposition à court terme aux PM<sub>2.5</sub> et à l'O<sub>3</sub> a de plus été associée à une augmentation des admissions à l'hôpital ainsi qu'à des visites aux urgences pour cause de maladies pulmonaires obstructives chroniques ou d'asthme. L'exposition à court terme a aussi été associée à des symptômes respiratoires (c.-à-d. irritation des voies respiratoires et difficultés respiratoires), et à une inflammation pulmonaire chez des enfants souffrant d'asthme (U.S. EPA, 2009, 2013).

#### **1.4.3 Effets cardiovasculaires aigus**

Finalement, plusieurs études (Bell *et al.*, 2016; INSPQ, 2012; U.S. EPA, 2009) ont souligné des associations entre l'exposition à court terme aux PM<sub>2.5</sub> et l'admission à l'hôpital, ainsi qu'avec des visites aux urgences pour problèmes cardiovasculaires, principalement en raison de maladies comme la cardiopathie ischémique et l'insuffisance cardiaque congestive. Par ailleurs, selon l'U.S. EPA (2009), les PM<sub>2.5</sub> induisent des changements dans la fonction

cardiovasculaire chez les personnes en bonne santé en plus des personnes dont la santé est déjà compromise.

## **1.5 Mesures et modélisation de l'exposition**

En modélisation de la pollution atmosphérique, on peut distinguer deux principales approches: déterministes et mathématiques-statistiques.

L'approche déterministe (dont nous ne traiterons pas ici) calcule à partir d'une description précise du territoire les phénomènes physiques et chimiques pouvant influencer la pollution atmosphérique. En raison de la lourdeur de sa mise en œuvre (accès aux données et logiciels, etc.) et pour s'assurer de la reproduction réaliste des phénomènes physico-chimiques, cette approche est généralement utilisée à une échelle régionale.

Quant aux approches mathématiques-statistiques (i.e. l'interpolation géométrique et géostatistique), nous les expliquons en détail dans les sections suivantes.

### **1.5.1 Mesures**

Toutes les grandes villes du monde possèdent des réseaux de surveillance de la qualité de l'air. Au Canada, le Réseau national de surveillance de la pollution atmosphérique (RNSPA) fournit des données à long terme depuis 1969 afin de surveiller et d'évaluer la qualité de l'air ambiant à partir d'environ 300 stations situées dans les centres urbains du Canada (Environment Canada, 2013c). Le RNSPA effectue une surveillance horaire des

polluants tels que  $\text{SO}_2$ ,  $\text{O}_3$ ,  $\text{CO}$ ,  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ,  $\text{NO}_x$ ,  $\text{NO}_2$ ,  $\text{NO}$ , et  $\text{COV}$ . Généralement, la distance entre stations du RNSPA est de l'ordre de 6 à 8 km dans les zones urbaines (CCME, 2011). Quatorze stations du RNSPA sont localisées dans l'île de Montréal (Ville de Montréal, 2010), parmi elles dix stations mesurent les  $\text{PM}_{2.5}$  et neuf stations mesurent l' $\text{O}_3$  (

Figure 2).

Une meilleure connaissance des effets des polluants atmosphériques sur la santé nécessite des données sur les variations intra-urbaines de l'exposition. Cependant, la limite la plus importante à cet égard réside dans le fait que les concentrations ne sont mesurées qu'à l'endroit de quelques stations de mesures au sol. C'est notamment le cas des stations du RNSPA de Montréal. En d'autres termes, quelles variations dans les niveaux de concentration se produisent sur des distances pouvant aller jusqu'à 8 km entre stations du RNSPA? Le faible nombre de stations de mesures et les variations spatiales importantes de polluants comme les  $\text{PM}_{2.5}$  et l' $\text{O}_3$  constituent un problème qui nous amène à concevoir des méthodes de modélisation afin d'estimer le plus exactement possible les variations intra-urbaines



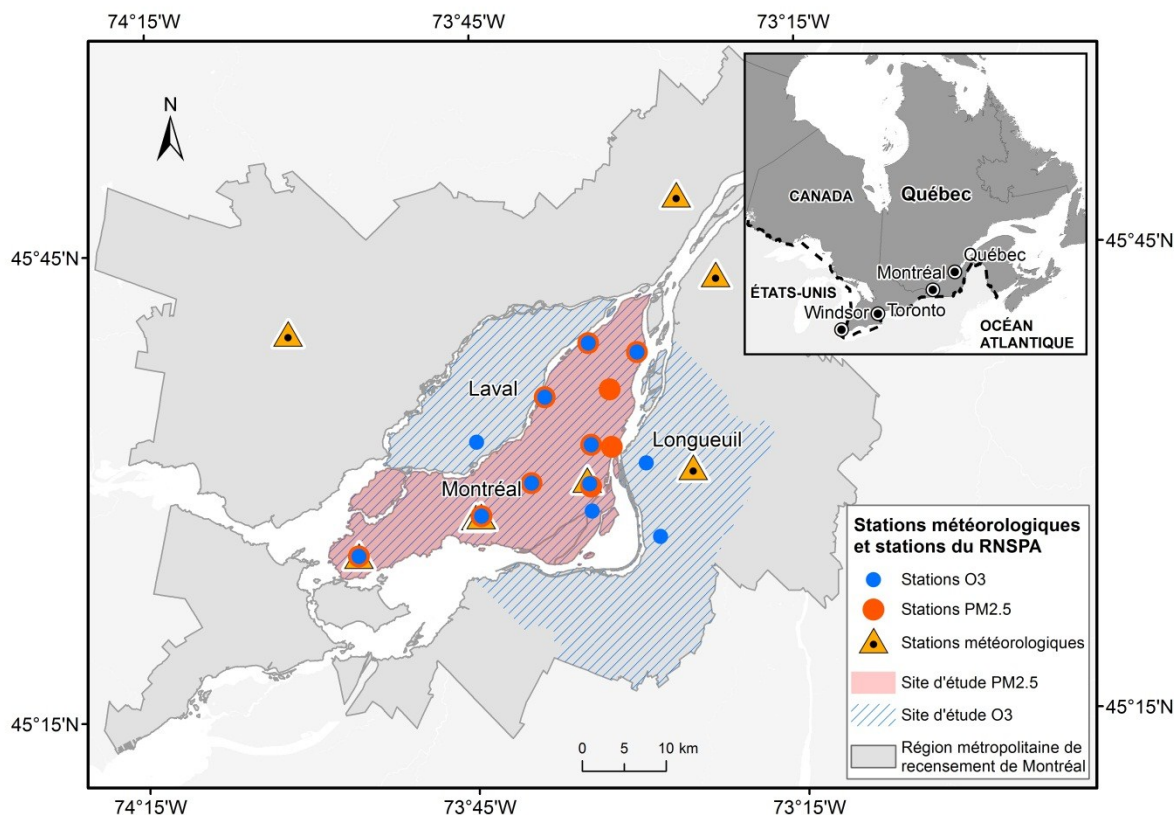


Figure 2. Site d'étude et localisation des stations météorologiques et du réseau national de surveillance de la pollution atmosphérique (RNSPA).

### 1.5.2 Les méthodes d'interpolation spatiales

L'interpolation spatiale permet de prédire la valeur d'une variable en un point pour lequel il n'existe aucune mesure à partir d'un nombre limité de valeurs en points échantillonnés (p.ex. aux stations du RNSPA). Il existe de nombreuses méthodes d'interpolation spatiales qui se départagent en deux types: 1) les méthodes géométriques se basant sur une fonction purement mathématique sans prendre en compte la structure de

dépendance spatiale des données (p.ex. l'interpolation basée sur la pondération inverse par rapport à la distance, ou IDW - *inverse distance weighting*) et; 2) les méthodes géostatistiques (postulant des processus stochastiques) dite de « krigeage ».

Les méthodes de krigeage s'appuient sur le principe de la dépendance spatiale, exprimée notamment par l'autocorrélation spatiale (Jerrett *et al.*, 2005) et évoquée par la première loi de la géographie : « *Everything is related to everything else, but near things are more related* » (Tobler, 1970, p. 236). Ces méthodes existent depuis plus de 50 ans et sont souvent utilisées en sciences géologiques et minières. Il existe plusieurs variantes du krigeage, tel que le krigeage ordinaire (OK – *ordinary kriging*), le krigeage universel (UK – *universal kriging*), ou le krigeage avec dérive externe (KED – *kriging with external drift*).

Précisons que l'OK est la plus commune des méthodes de krigeage et que le KED était originalement appelé UK (Matheron, 1969). L'UK et le KED sont tous deux des modèles avec tendance spatiale. Le terme UK a finalement été retenu, pour le cas où la tendance spatiale dépend directement des coordonnées, et le terme KED pour le cas général où la variation de la variable indépendante est définie comme une fonction linéaire des variables auxiliaires (Chiles, 1991; Wackernagel, 2003).

En modélisation environnementale de la pollution atmosphérique, les méthodes d'interpolation sont utilisées depuis environ deux décennies. Par exemple, l'OK a été appliqué aux données des concentrations de polluants pour la cartographie d'O<sub>3</sub> sur un vaste territoire pour une année spécifique (Beelen *et al.*, 2009). En région urbaine, l>IDW et l'OK ont été

utilisés pour l'estimation des concentrations d'O<sub>3</sub>; le premier pour 16 jours spécifiques (Varns *et al.*, 2001) et le deuxième pour une année spécifique (Liu *et al.*, 1996). À notre connaissance, seule une étude utilisant le KED pour l'estimation des concentrations des PM<sub>2.5</sub> en région urbaine et pour une journée spécifique a été publiée (Pearce *et al.*, 2009).

La Figure 3 montre les principales méthodes d'estimation des concentrations de polluants atmosphériques.

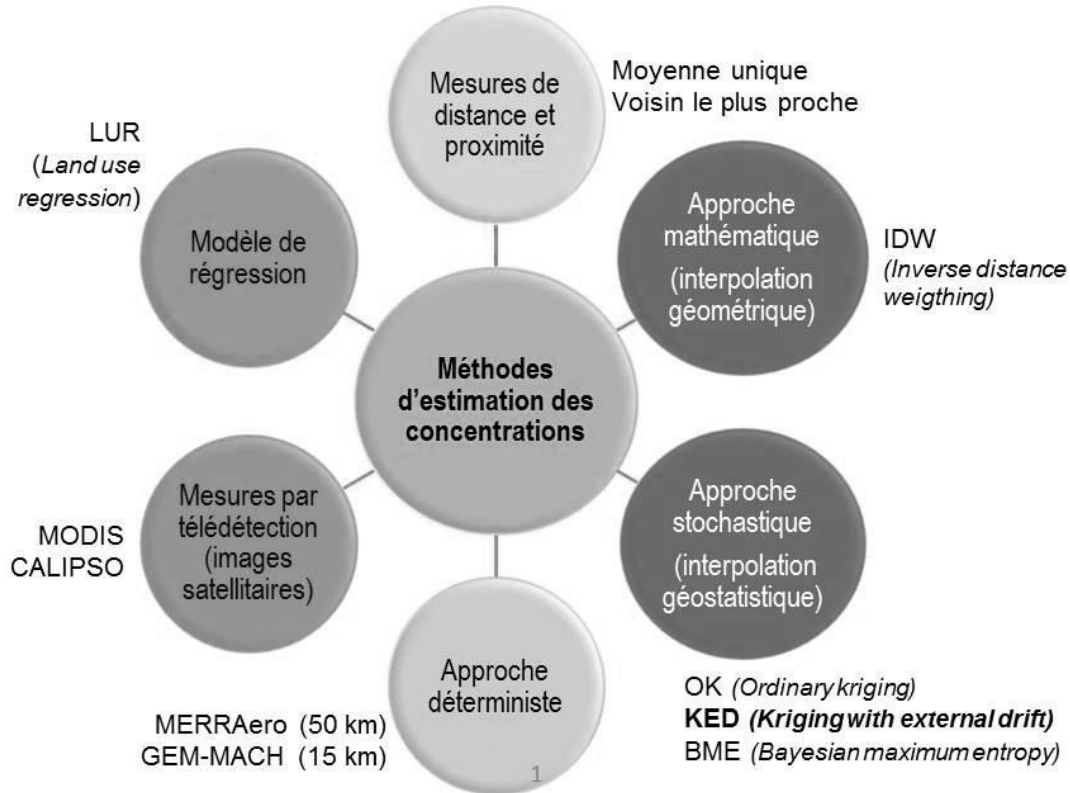


Figure 3. Méthodes d'estimation des concentrations de polluants atmosphériques. MODIS et CALIPSO sont des satellites; MERRAero est un modèle de la NASA (*National Aeronautics and Space Administration*) et GEM-MACH est un modèle d'Environnement Canada.

### 1.5.3 Méthodes de modélisation utilisées en épidémiologie

Plusieurs études épidémiologiques utilisent des méthodes simples pour l'estimation des niveaux d'exposition journaliers des individus (Gulliver *et al.*, 2011; Jerrett *et al.*, 2009; K. A. Miller *et al.*, 2007; Pope III *et al.*, 2004). Par exemple, certaines études ont employé une valeur moyenne unique des mesures réalisées sur l'ensemble des stations de surveillance (Dominici *et al.*, 2002). Cette approche est fréquemment mise en œuvre dans l'analyse des

séries chronologiques. D'autres études ont plutôt considéré que l'exposition journalière correspondait aux concentrations de polluants observées à la station de mesure la plus proche de la résidence des individus. Étant donné que ces méthodes reposent sur de simples approximations, elles créent une incertitude quant à l'estimation de l'exposition réelle aux polluants (L. Chen *et al.*, 2007; D. Lee *et al.*, 2010; Strickland *et al.*, 2015). Bien que la méthode basée sur la proximité des stations puisse prendre en compte dans une certaine mesure la variation spatiale dans les données, celle-ci est considérée comme un piètre indicateur de l'exposition aux polluants dans les environnements urbains complexes (Gulliver *et al.*, 2011). Par ailleurs, la nature de la formation des polluants (aussi associée à une distribution hétérogène pour les polluants primaires et homogène pour les polluants secondaires) détermine le degré de variabilité spatiale, donc leur prise en compte peut influencer les niveaux d'incertitudes dans les estimations des risques pour la santé (Strickland *et al.*, 2011; Strickland *et al.*, 2015).

Depuis environ deux décennies, des études de modélisation environnementale (avec application en épidémiologie) ont eu recours à des méthodes d'interpolation spatiale (section 1.5.2) afin de mieux représenter les variations spatiales de polluants atmosphériques et d'ainsi estimer l'exposition journalière des individus d'une manière plus exacte (Briggs, 2005; Hoek *et al.*, 2002; Jerrett *et al.*, 2005). La plupart de ces méthodes d'interpolation sont construites à partir des concentrations moyennes annuelles. Néanmoins, l'estimation de l'exposition à court terme nécessite des estimations journalières des niveaux de pollution de l'air.

Par ailleurs, parmi les méthodes basées sur des variables auxiliaires utilisées comme prédicteurs, la régression basée sur l'utilisation du sol (LUR - *Land use regression*) (Hoek *et al.*, 2008) est la méthode d'estimation spatialisée de la pollution atmosphérique la plus couramment utilisée. Toutefois, cette approche ne considère pas l'autocorrélation spatiale de la variable dépendante, ce que le KED fait tout en considérant des informations supplémentaires dérivées des variables auxiliaires. La LUR est une méthode de régression linéaire développée pour estimer les concentrations des polluants aux sites non échantillonnés par les stations de mesures et basés sur des caractéristiques d'utilisation du sol locales (variables prédictives) (Crouse *et al.*, 2009). Cette méthode est souvent construite par régression des moyennes mensuelles ou annuelles en négligeant les variations temporelles de plus courtes durées (c.-à-d. journalières) (C. Chen *et al.*, 2012). Certaines études épidémiologiques ont utilisé cette méthode pour estimer une exposition à long terme à la pollution atmosphérique (Gehring *et al.*, 2013; Stafoggia *et al.*, 2014).

D'autre part, des modèles spatio-temporels ont été aussi développés, enrichissant ainsi les approches basées sur les statistiques spatiales en incluant la dimension temporelle dans les modèles de prédiction (Adam-Poupart *et al.*, 2014; Bogaert *et al.*, 2009; Yang *et al.*, 2015). Ces modèles dits spatio-temporels prennent en compte simultanément la dépendance spatiale et temporelle générale des données afin de prédire les concentrations de polluants de l'air ambiant dans l'espace et dans le temps. Par exemple, l'approche spatiotemporelle BME (*bayesian maximum entropy*), a été utilisée pour prédire les concentrations de l'ozone troposphérique sur le territoire québécois (Adam-Poupart *et al.*, 2014). Elles ont notamment comme avantage de fournir des prédictions pour des jours où les données sont manquantes.

Les estimations de ce modèle ont été utilisées pour évaluer les effets d'expositions aiguës à la pollution extérieure chez les travailleurs (Adam-Poupart *et al.*, 2015).

Depuis le début des années 2000, les approches basées sur la télédétection ont été aussi utilisées pour estimer les concentrations de polluants atmosphériques sur de vastes territoires (Chu *et al.*, 2003; S.-J. Lee *et al.*, 2012; Van Donkelaar *et al.*, 2006). Néanmoins, un nombre limité d'études ont utilisé ces mesures pour évaluer l'exposition à la pollution atmosphérique.

Dans une perspective de santé publique, l'estimation des concentrations de polluants en considérant les variations intra-urbaines avec l'inclusion de variables auxiliaires comme facteurs prédictifs des PM<sub>2.5</sub> et de l'O<sub>3</sub> (**Erreur ! Source du renvoi introuvable.**), peut contribuer à mieux évaluer les risques sanitaires associés à la variation journalière de la pollution atmosphérique et à développer les stratégies préventives les plus appropriées. Ceci nous amène à proposer l'utilisation du krigeage avec dérive externe (*kriging with external drift*, KED, (Chiles, 1991)) et à comparer les prédictions spatialisées journalières des concentrations des PM<sub>2.5</sub> et de l'O<sub>3</sub> avec des méthodes déjà répandues. Le KED, par l'ajout des variables externes utilisées comme prédicteurs peut contribuer à améliorer l'interpolation spatiale des polluants atmosphériques visés. De plus, l'utilisation d'un modèle spatio-temporel comme BME (Adam-Poupart *et al.*, 2014) peut mieux expliquer les niveaux d'exposition des individus que les méthodes simples d'estimation de l'exposition. Nous présentons nos hypothèses spécifiques dans la section suivante.

Polluant	Formation	État physique	Sources Primaires
COV (composés organiques volatiles)	primaire	gaz	végétation: <b>densité de la végétation</b> industries: <b>utilisation industrielle des sols</b> véhicules: <b>densité des routes</b>
PM (différentes tailles, dont les PM <sub>2.5</sub> )	primaire et secondaire (à partir de NOx et SOx)	solide, liquide	véhicules: <b>densité des routes</b> combustion de bois de chauffage: <b>densité de la population</b> industries: <b>utilisation industrielle des sols, émission industrielle des PM<sub>2.5</sub></b>
NOx (oxydes d'azote)	primaire	gaz	véhicules: <b>densité des routes</b> industries: <b>utilisation industrielle des sols</b>
SOx (oxydes de soufre)	primaire	gaz	véhicules: <b>densité des routes</b> industries: <b>utilisation industrielle des sols</b>
O <sub>3</sub> (ozone troposphérique)	secondaire (à partir des NOx et des COV)	gaz	photo-oxydation

Table 1. Sources primaires des polluants (PM<sub>2.5</sub> et O<sub>3</sub>) et la relation avec les variables externes potentielles (variables géographiques-environnementales).



## 2 Hypothèses de recherche

Nous avons formulé les hypothèses suivantes, en nous basant sur la littérature scientifique ainsi que les principes d'émission et de formation des  $PM_{2.5}$  et d' $O_3$ , afin d'orienter notre travail de recherche, et en avons testé la validité :

- 1) Des niveaux élevés des  $PM_{2.5}$  sont associés à des températures froides, des vents légers, une pression atmosphérique élevée, des activités industrielles locales, et aux secteurs densément peuplés.
- 2) Des niveaux élevés d' $O_3$  sont associés à des températures hautes, des vents légers, une faible humidité relative, à la présence de végétation dense et une faible densité de routes.
- 3) Les erreurs d'estimation des méthodes basées sur la moyenne des stations, sur la station la plus proche, ou alors sur l'interpolation univariable (IDW ou OK) sont plus grandes que celle d'une méthode d'interpolation multivariable (KED).
- 4) Les estimations des concentrations de polluants provenant d'un modèle spatio-temporel expliquent le mieux les effets sur la santé en comparaison avec la méthode d'estimation basée sur la moyenne des stations ou sur la station la plus proche.
- 5) La relation entre les variables auxiliaires et les  $PM_{2.5}$  et l' $O_3$  change selon le régime climatique.

## **3 Objectifs de recherche**

### **3.1 Objectif général**

L'objectif général de cette thèse est de développer une méthode d'intégration de données basée sur la géostatistique, les systèmes d'information géographique et la télédétection permettant d'estimer spatialement et temporellement; les concentrations journalières des particules fines ( $PM_{2.5}$ ) et de l'ozone troposphérique ( $O_3$ ) entre les stations de surveillance de qualité de l'air à Montréal, Canada.

### **3.2 Objectifs spécifiques**

Le premier objectif de cette recherche vise à développer et comparer des modèles d'interpolation dans le but d'estimer les concentrations journalières des  $PM_{2.5}$  et d' $O_3$ .

Le deuxième objectif vise à comparer l'estimation du risque de décès à partir de différentes méthodes d'estimation des niveaux journaliers d' $O_3$ .

Le premier objectif s'est traduit sous la forme de deux articles (chapitres 3 et 4). Ces articles sont axés plus particulièrement sur le développement d'un modèle de la géostatistique multivariante (KED) pour estimer les concentrations journalières de  $PM_{2.5}$  et d' $O_3$ . Le modèle dit hybride (intégration de l'interpolation univariante et multivariante), KED-IDW a été développé pour le cas particulier des  $PM_{2.5}$  et le modèle KED pour le cas de l' $O_3$ . Les estimations obtenues dans les deux cas ont été comparées avec des estimations obtenues à

partir des méthodes d'interpolation univariable (IDW pour les PM<sub>2.5</sub> et OK pour l'O<sub>3</sub>) et d'un modèle spatio-temporel (BME) développé dans une étude précédente pour les concentrations d'O<sub>3</sub>. Nous avons aussi comparé les résultats à la moyenne journalière des concentrations dans toutes les stations de surveillance.

Le deuxième objectif s'est traduit sous la forme d'un article (chapitre 5). Dans cet article, nous avons évalué l'apport du modèle BME par rapport aux méthodes simples (basées sur la moyenne des stations et sur la station la plus proche de surveillance de qualité de l'air) dans l'estimation du risque de décès chez les personnes ayant habité dans la RMR de Montréal.

Ces articles sont présentés sous la forme d'articles soit publiés, soumis pour publication ou prêts à être soumis (Chapitres 3, 4 et 5).

**ARTICLE 1** : *Spatio-temporal models to estimate daily concentrations of fine particulate matter in Montreal: Kriging with external drift and inverse distance-weighted approaches*; publié dans *Journal of Exposure Science and Environmental Epidemiology* (2016) 26, 405–414.

**ARTICLE 2** : *Spatio-temporal modeling of daily concentrations of ground-level ozone in Montreal, Canada*; soumis à *Atmospheric Environment* en date du 4 décembre 2015 (des commentaires mineurs de trois réviseurs ont été reçus en date du 26 janvier 2016, les trois réviseurs ayant donné un avis favorable pour la publication de l'article).

**ARTICLE 3** : *Association of mortality with short-term exposures to ground-level ozone in Montreal, Canada : A comparison of risk estimation from non-spatial and spatio-temporal exposure approaches*; version finale du manuscrit prêt à être soumis.



## **Chapitre 2 : Méthodologie**



Dans ce chapitre, nous ferons une brève description de la méthodologie du KED (Figure 4**Erreur ! Source du renvoi introuvable.**) ainsi que des sources de données. Toutefois, chacun des articles inclus dans cette thèse (chapitres 3 et 4) comporte une description détaillée de la méthodologie adaptée selon le cas, aux estimations des concentrations des  $PM_{2.5}$  ou d' $O_3$ . Par exemple pour le cas des estimations des  $PM_{2.5}$ , nous avons utilisé, une méthode combinée d'interpolation univariable (IDW) et multivariable (KED) dit, KED-IDW. Nous ferons aussi dans le présent chapitre une courte description du modèle BME utilisé dans les articles 2 et 3 (chapitre 4 et 5).

La mise en œuvre de la méthodologie décrite dans cette thèse a été réalisée principalement à l'aide du logiciel R version 3.1.1 (*R Project for Statistical Computing*; <http://www.r-project.org>).

## **1. KED (*Kriging with external drift*)**

### **1.1 Données sur la qualité de l'air et météorologiques**

Des données horaires des  $PM_{2.5}$  (10 stations) et d' $O_3$  (12 stations) ont été obtenues du RNSPA, alors que des données météorologiques horaires (8 stations) ont été obtenues des archives nationales d'information et de données climatologiques d'Environnement Canada. La Figure 2**Erreur ! Source du renvoi introuvable.** montre la localisation des stations du RNSPA qui mesurent les  $PM_{2.5}$  et l' $O_3$ , ainsi que les stations météorologiques selon le site d'étude. Pour chaque station, les moyennes journalières des  $PM_{2.5}$ , d' $O_3$ , du vent, de la température et de l'humidité relative ont été calculées pour toutes les données qui ne



comportent pas plus de deux heures de valeurs manquantes. Des données sur la quantité totale de précipitations journalières ont été aussi utilisées.

## **1.2 Groupes météorologiques synoptiques**

Bien que les conditions météorologiques ne soient jamais similaires, certaines régularités sont observées au cours de l'année (Hufty, 1982). En faisant l'hypothèse que les variations spatiale et temporelle des  $PM_{2.5}$  et de l' $O_3$  sont corrélées aux conditions météorologiques, nous avons utilisé, les moyennes journalières de toutes les stations afin d'identifier des groupes ayant des régimes météorologiques similaires. Deux étapes ont été suivies pour identifier ces régimes: dans une première étape, nous avons réalisé une analyse en composantes principales (PCA - *Principal component analysis*) de 949 valeurs journalières de quatre variables météorologiques (la vitesse du vent, la température, l'humidité relative et la précipitation). Ceci a permis, la réduction de l'existence de corrélation entre les variables de départ à un nombre plus petit de variables non corrélées tout en conservant un maximum d'information. Dans une deuxième étape, nous avons réalisé une classification hiérarchique (HCA - *Hierarchical cluster analysis*) afin d'identifier les groupes avec régimes météorologiques relativement similaires.

## **1.3 Variables externes**

Deux groupes de variables externes (indépendantes) potentielles pour le développement du KED ont été considérées: les variables météorologiques spatio-temporelles et les variables environnementales et géographiques.

### 1.3.1 Variables météorologiques spatio-temporelles

Des surfaces météorologiques pour chacune de variables météorologiques (vitesse du vent, température, humidité relative et précipitations) et pour tous les jours de l'année 2010 ont été créées en utilisant la méthode d'interpolation IDW.

### 1.3.2 Variables environnementales et géographiques

#### Densité de la végétation

La végétation est associée aux sources primaires des  $PM_{2.5}$  et d' $O_3$  (Markevych *et al.*, 2014; J. Su *et al.*, 2010). Pour prendre en compte ce facteur, la méthode *Tasseled-cap* a été utilisée afin d'extraire la composante de verdure à partir de l'image Landsat TM. *Tasseled-cap* transforme les bandes (1-5 et 7) de l'image et les décompose en trois composantes orthogonales appelées; la verdure (*greenness*), l'humidité (*wetness*), et la brillance (*brightness*). La composante de verdure est corrélée à la densité chlorophyllienne; des valeurs élevées sont associées aux fortes densités de végétation et de faibles valeurs sont associées aux surfaces du sol. Pour chaque pixel de 100 mètres, la moyenne des valeurs de la composante de la verdure a été calculée pour différents rayons d'influence.

La source de données satellitaires est l'image Landsat TM (*Land satellite Thematic Mapper*) prise le 27 juillet 2010 de la RMR de Montréal (*path/row 014/028*), orthorectifiée et fournie par l'USGS (*U.S. Geological Survey*).

## **L'utilisation du sol**

Trois classes d'utilisation du sol (résidentiel, commercial et industriel) ont été utilisées comme mesures de substitution des émissions résidentielles, commerciales et industrielles liées aux  $PM_{2.5}$  et à l' $O_3$ . Les polygones d'utilisation du sol ont été convertis au format matriciel dans le but d'agréger chaque classe d'utilisation du sol se trouvant à l'intérieur des différents rayons (300, 500 et 750 mètres). Plusieurs études ont montré des associations avec les  $PM_{2.5}$  et l' $O_3$  dans ces rayons (Brauer *et al.*, 2008; Henderson *et al.*, 2007; Ross *et al.*, 2007). La fonction de statistiques de voisinage (*Neighborhood Statistics function*) de l'extension *Spatial Analyst* du logiciel ArcGIS a été utilisée pour obtenir une valeur pour chaque cellule de 100 mètres à l'intérieur du rayon de recherche défini.

La source de données d'utilisation du sol est *2010 Land Use Layer (CanMap® RouteLogistics version 2010.3)*, une base de données cartographique compilée par DMTI (*Digital Mapping Technology Inc.*).

## **La densité du réseau routier**

Les oxydes d'azote ( $NO_x$ ) impliqués dans la formation et la destruction de l' $O_3$  sont composés principalement de monoxyde d'azote (NO) et de dioxyde d'azote ( $NO_2$ ). En zones urbaines, la principale source d'émission de NO et de  $NO_2$  est le trafic routier. Pour prendre en compte ces émissions, nous avons utilisé une approche de pondération de la densité routière (WDR - *Weighted road density approach*) développée par Rose *et al.* (2009). En raison d'une

relation log-linéaire avec l'O<sub>3</sub> révélée par une analyse graphique de données, nous avons calculé le logarithme de WDR.

Deux étapes ont été suivies pour l'application de WDR : dans la première, des valeurs de pondération en fonction des niveaux hiérarchiques du système routier ont été assignées. Cette pondération a été établie en supposant une relation similaire avec les NOx et la densité basée sur le comptage de circulation (Hansell *et al.*, 2014; Rose *et al.*, 2009). Dans une deuxième étape, WDR a été utilisé pour calculer la densité à l'intérieur des différents rayons. Ces rayons ont été établis à l'intérieur du seuil d'influence du trafic routier sur les niveaux de NOx (1 500 m), rapporté par (J. G. Su *et al.*, 2009).

La source de données du réseau routier est *2010 Road Layer Dataset (CanMap® RouteLogistics version 2010.3)*, une base de données cartographique compilée par *Digital Mapping Technology Inc. (DMTI)*.

## **1.4 Modélisation spatio-temporelle**

Le KED a été utilisé pour estimer les concentrations des PM<sub>2.5</sub> et d'O<sub>3</sub> pour tous les jours de l'année 2010. Le KED modélise la tendance (dérive) dans les données (mesure des concentrations de polluants aux stations fixes) comme une fonction linéaire de variables externes (indépendantes) (Wackernagel, 2003). La sélection de variables externes était basée sur une analyse graphique et sur les niveaux de corrélation (*Pearson's correlation coefficient*) avec les concentrations des PM<sub>2.5</sub> et d'O<sub>3</sub>. Le coefficient de corrélation a été calculé pour

chaque jour et par groupe synoptique. La relation entre une variable indépendante  $Y(u)$  et la variable dépendante  $Z(u)$  peut s'exprimer comme suit :

$$Z(u) = c_1 + c_2 Y(u) + Y_R(u) \quad (1)$$

où  $c_1$  et  $c_2$  sont constantes,  $Y(u)$  est connue en tout point de la grille d'interpolation et représente la valeur de la variable indépendante à l'emplacement  $u$ , et  $Y_R(u)$  est le résidu. Une fois la dérive définie (selon une fonction linéaire de variables externes), un modèle de variogramme est ajusté aux résidus  $Y_R(u)$ .

Le niveau de concentration des  $PM_{2.5}$  et d' $O_3$  à un point d'intérêt donné  $u$  est estimé selon :

$$Z_{KED}^*(u) = \sum_{\alpha=1}^n \lambda_{\alpha}(u) Z(u_{\alpha}) \quad (2)$$

où,  $\lambda_{\alpha}(u)$  sont les poids du krigeage aux  $n$  échantillons à la localisation  $u$  obtenue par la solution du système linéaire développé par (Goovaerts, 1997) impliquant la connaissance du variogramme. Les estimations des  $PM_{2.5}$  et de l' $O_3$  sont générées par tranche de temps (c.-à-d. jour par jour) et basées sur le calcul des variogrammes des données journalières.

## 1.5 Validation

La précision des estimations obtenues avec KED a été évaluée par la méthode de validation croisée « *leave-one-out* » (LOOCV - *Leave-one-out cross-validation*). En utilisant cette méthode, la concentration réelle mesurée en chaque station à un jour donné a été retirée des données (une à la fois), les données restantes ont été utilisées pour estimer la concentration du polluant (PM<sub>2.5</sub> ou O<sub>3</sub> selon le cas) à l'emplacement de la station exclue. Par la suite, les estimations obtenues et les concentrations réelles mesurées aux stations exclues ont été comparées afin de déterminer l'erreur de la prédiction.

La Figure 4 résume schématiquement la méthodologie du KED menant à l'estimation des concentrations journalières des PM<sub>2.5</sub> et d'O<sub>3</sub>.

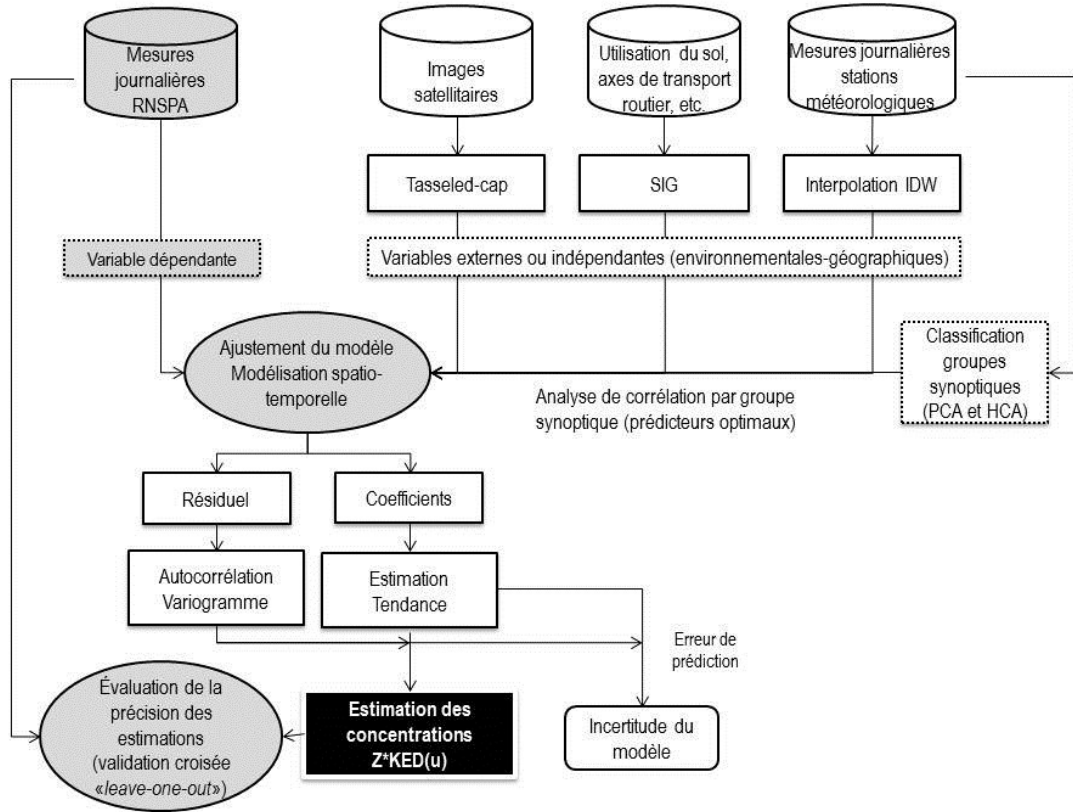


Figure 4. Diagramme méthodologique pour l'application du krigeage avec dérive externe (KED)

## 2. BME (*Bayesian maximum entropy*)

Dans le but de comparer les résultats du KED (article 2, chapitre 4) avec un modèle déjà répandu, nous avons utilisé des estimations des concentrations d'O<sub>3</sub> obtenues avec le modèle BME et développé par Adam-Poupart *et al.* (2014). Ce modèle a été développé pour les concentrations d'O<sub>3</sub> pendant les mois d'été (de mai à septembre de 1996 à 2009). Nous avons aussi utilisé ces estimations pour mesurer l'effet à court terme (~10 ans) de l'exposition à l'O<sub>3</sub> sur les décès à Montréal (article 3, chapitre 5).

Le BME est un modèle spatio-temporel basé sur l'estimation des densités de probabilités à l'aide du maximum d'entropie. Le niveau de concentration à un point donné, tout comme les autres méthodes de la géostatistique, est estimé selon un facteur de pondération. Dans le cas de BME, ces facteurs de pondération sont calculés à l'aide des mesures physiques obtenues des stations fixes du RNSPA (données observées « *hard data* ») et des informations supplémentaires telles que les estimations de régression à variables multiples (données estimées « *soft data* »). Ces informations sont mises en commun afin de calculer le facteur de pondération qui sera associée à chaque concentration mesurée aux stations fixes les plus près à un point d'intérêt donné (Adam-Poupart *et al.*, 2013). Les niveaux des concentrations en O<sub>3</sub> estimés (« *soft data* ») ont été obtenus à partir de la régression basée sur l'utilisation du sol (LUR) à effets mixtes avec l'intégration de variables. Les variables telles que la température, la précipitation, le jour de l'année, l'année, la densité des routes dans une zone tampon de 1km et la latitude ont été utilisés dans le modèle de régression. C'est-à-dire



que la concentration moyenne journalière en O<sub>3</sub> sur 8-h pour chaque station-jour observée a été estimée comme suit (Adam-Poupart *et al.*, 2014):

$$\begin{aligned} O_3 = & \beta_0 + \beta_1 X_{low\_temperature} + \beta_2 X_{high\_temperature} + \beta_3 X_{precipitation} \\ & + \beta_4 X_{dayofyear} + \beta_5 X_{year} + \beta_6 X_{low\_road} + \beta_7 X_{high\_road} \\ & + \beta_8 X_{low\_latitude} + \beta_9 X_{high\_latitude} + u_{station} + \varepsilon, \end{aligned} \quad (3)$$

où  $X$  est la valeur de la variable pour la station-jour,  $\beta$  est le coefficient pour cette variable,  $u$  est l'effet aléatoire associé à cette station, et  $\varepsilon$  est l'erreur restante de la station-jour.



**Chapitre 3 : Spatio-temporal models to estimate daily concentrations of fine particulate matter in Montreal: Kriging with external drift and inverse distance weighted approaches – Article 1**



**Title:** Spatio-temporal models to estimate daily concentrations of fine particulate matter in Montreal : Kriging with external drift and inverse distance weighted approaches

*Article published in the Journal of Exposure Science and Environmental Epidemiology*

**Authors :** Yuddy Ramos<sup>1</sup>, Benoît St-Onge<sup>2</sup>, Jean-Pierre Blanchet<sup>3</sup>, Audrey Smargiassi<sup>4,5</sup>

<sup>1</sup>Département de géographie, Université de Montréal, Montréal, Québec, Canada

<sup>2</sup>Département de géographie, Université du Québec à Montréal (UQAM), Montréal, Québec, Canada

<sup>3</sup>Département des sciences de la Terre et de l'atmosphère, Université du Québec à Montréal (UQAM), Montréal, Québec, Canada

<sup>4</sup>Département de santé environnementale et de santé au travail, Université de Montréal, Montréal, Québec, Canada

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

### **Contribution de l'auteure principale et des coauteurs :**

L'article présenté dans ce chapitre a été réalisé par Yuddy Ramos sous la supervision de Benoît St-Onge, la cosupervision d'Audrey Smargiassi, et la collaboration de Jean-Pierre Blanchet.

Yuddy Ramos a conçu l'approche méthodologique, effectué tous les traitements de données, analysé les résultats, rédigé la première version de l'article et effectué toutes les modifications selon les suggestions des coauteurs. Elle a aussi produit la version révisée et répondu aux évaluateurs de l'article jusqu'à sa publication.

Benoît St-Onge à titre de directeur de recherche a supervisé la conception méthodologique et contrôlé l'analyse des résultats. Il a révisé la version initiale et toutes les versions avant publication. Il a supervisé le travail de révision de l'article et les réponses aux évaluateurs suite au processus d'évaluation par les pairs.

Audrey Smargiassi a cosupervisé la conception méthodologique. Elle a fourni les données de qualité de l'air et météorologiques, ainsi que le soutien financier. Elle a révisé la version initiale et toutes les versions avant publication et cosupervisé la révision de l'article et les réponses aux évaluateurs suite au processus d'évaluation par les pairs.

Jean-Pierre Blanchet a fourni son avis scientifique d'ensemble et a été consulté tout au long du développement méthodologique et de l'analyse des résultats. Il a aussi relu les versions avancées de l'article et a proposé des modifications.

## 2 Abstract

Air pollution is a major environmental and health problem, especially in urban agglomerations. Estimating personal exposure to fine particulate matter (PM<sub>2.5</sub>) remains a great challenge because it requires numerous point measurements to explain the daily spatial variation in pollutant levels. Furthermore, meteorological variables have considerable effects on the dispersion and distribution of pollutants, which also depends on spatio-temporal emission patterns. In this study we developed a hybrid interpolation technique that combined the Inverse Distance Weighted (IDW) method with Kriging with External Drift (KED), and applied it to daily PM<sub>2.5</sub> levels observed at 10 monitoring stations. This provided us with downscaled high resolution maps of PM<sub>2.5</sub> for the Island of Montreal. For the KED interpolation, we used spatio-temporal daily meteorological estimates and spatial covariates as land use and vegetation density. Different KED and IDW daily estimation models for the year 2010 were developed for each of the six synoptic weather classes. These clusters were developed using principal component analysis and unsupervised hierarchical classification. The results of the interpolation models were assessed with a leave-one-station-out cross-validation. The performance of the hybrid model was better than that of the KED or the IDW alone for all six synoptic weather classes (the daily estimate for R<sup>2</sup> was 0.66 to 0.93 and for RMSE 2.54 to 1.89 µg/m<sup>3</sup>).

**Keywords** : fine particulate matter, air pollution, interpolation, kriging with external drift, inverse distance weighted

### 3 Introduction

Fine particulate matter with a median diameter of 2.5  $\mu\text{m}$  or less ( $\text{PM}_{2.5}$ ) is a main component of smog (Environment Canada, 2012).  $\text{PM}_{2.5}$  can be emitted directly in the atmosphere or generated through the gas-to-particle process (secondary) from sulfur oxides, nitrogen oxides, ammonia and volatile organic compounds (U.S. EPA, 2009). Many studies have shown associations between daily  $\text{PM}_{2.5}$  levels and respiratory or cardiovascular morbidity and mortality (Bell *et al.*, 2008; Ito *et al.*, 2011; U.S. EPA, 2009; Zanobetti *et al.*, 2009). One major limitation of many studies on the health effects of short term air pollutant exposures is that they fail to consider spatial variation in pollutant levels (Brauer *et al.*, 2003; Goldman *et al.*, 2012). This may introduce uncertainty in the estimation of the exposure of individuals in epidemiologic studies.

The spatial variation of pollutant levels such as  $\text{PM}_{2.5}$  is influenced by the location of the emission sources, meteorological conditions and site topography (Pinto *et al.*, 2012). Meteorological conditions such as precipitation, wind speed, temperature, humidity and atmospheric pressure determine the diffusion and transport of  $\text{PM}_{2.5}$ . For example, an increase in the precipitation rate has a negative effect on the levels of  $\text{PM}_{2.5}$  (Dawson *et al.*, 2007), local air circulation controls the dispersion and transport in the region (Munoz-Alpizar *et al.*, 2003), high winds disperse and transport pollutants over larger areas (Dawson *et al.*, 2007; Tai *et al.*, 2010), whereas temperature inversion (increases of temperature with height) traps pollutants close to the ground, thereby increasing their levels (Malek *et al.*, 2006).



To estimate the spatial variation in pollutant levels over extensive urban environments, a number of methods have been commonly used, among them univariate models such as the inverse distance weighted (IDW) (Lipsett *et al.*, 2011) model and multivariate models such as land use regression (LUR) (Briggs *et al.*, 1997; Henderson *et al.*, 2007) and Bayesian Maximum Entropy (BME) models (Adam-Poupart *et al.*, 2014; Beckerman *et al.*, 2013). The multivariate models rely on some form of spatial dependence between covariates to create a continuous surface of exposure over time for a given area of interest (Jerrett *et al.*, 2003). Nevertheless, such multivariate models are typically based on regressing annual averages of pollutant concentrations using geographic covariate invariants in time (Chen *et al.*, 2012).

The main objective of this work was to develop and compare interpolation models aimed at mapping daily ground-level PM<sub>2.5</sub> concentrations at high resolution on the Island of Montreal. To pursue this objective, we explored the use of KED in combination with IDW, to predict daily PM<sub>2.5</sub> concentrations. The approach developed takes into account both broad spatial trends and local variations (Pearce *et al.*, 2009). To account for spatial dependence on meteorological variables, these models were calibrated separately for each synoptic weather class.

## 4 Methods

### Study area

The study area is the Island of Montreal, which is mainly occupied by the City of Montreal (the second largest city in Canada) with a population of 1.89 million people. The city is bordered by the Rivière des Prairies and the Saint Lawrence River and is situated between 45.40°–45.72°N and 73.98°–73.45°W (Figure 1). The island is mostly flat, with the exception of a moderate hill, the Mount Royal, culminating at 234 meters above sea level. The climate is classified as humid continental which, with severe winters and hot, humid summers, is subject to tremendous meteorological variations. The predominant wind direction is south-west, followed by north-east winds along the Saint Lawrence River. Poor air quality days on the Island of Montreal are associated with high concentrations of PM<sub>2.5</sub>, which are attributed to residential wood burning, road traffic, industrial activities, fireworks, forest fires occurring hundreds of kilometers north of the island (Ville de Montréal, 2010), and long range transport from industrial regions south of the border and around the Great Lakes. Also, meteorological conditions play an important role in determining pollution levels at proximity to sources. The study period for which interpolation models were developed spans from January to December 2010.

### PM<sub>2.5</sub> and meteorological data

For each day of the study period, hourly data of PM<sub>2.5</sub> concentrations were obtained from the National Air Pollution Surveillance (NAPS) network of Environment Canada

(Environment Canada, 2013a). The NAPS network uses different equipment to measure ambient  $PM_{2.5}$  concentration. Currently, the most widely used methods in Canada are the Tapered Element Oscillating Microbalance (TEOM<sup>TM</sup>) method, which is based on inertial properties, and the BAM-1020 beta attenuation monitor (CCME, 2011). On the Island of Montreal, fine particulate matter is measured with 10 NAPS monitors using TEOM technology with a Filter Dynamic Measurement System (FDMS) (Figure 1).  $PM_{2.5}$  monitoring stations provide a spatially sparse sample of particulate data with a good continuity in time. Meteorological measurements were obtained from the National Climatic Data and Information Archive (DAI) of Environment Canada (Environment Canada, 2013b). There are eight meteorological monitoring stations across the Montreal Census Metropolitan Area (Figure 1). We calculated daily averages of  $PM_{2.5}$  levels, whereby wind speed, temperature, relative humidity and precipitation for all available monitoring data had no more than two hours of missing values on any given day. In addition, extreme  $PM_{2.5}$  levels (outliers) that occurred during major forest fire events were excluded in the preprocessing stage. To identify these outliers, we used a list of major smog episodes from Environment Canada (Environment Canada, 2014).

## **Synoptic weather analysis**

Although weather conditions in the study area are very variable, some general patterns were observed over the year (Hufty, 1982). Assuming that the temporal variation and spatial distribution of  $PM_{2.5}$  concentrations are correlated with meteorological conditions, we used data from 2008–2010 to identify clusters of days having similar meteorological patterns, in

other words, synoptic groups. A time period longer than the study period (i.e., beyond 2010) was used to classify days into synoptic groups, allowing for a robust classification based on a larger number of observations. To identify the synoptic groups, we proceeded in two stages. First we used principal component analysis (PCA) to reduce the existing correlation between the mean daily values of four meteorological variables (wind speed, temperature, relative humidity and precipitation) (949 observations) while retaining most of the information contained in the data. Then, to identify groups that contain similar meteorological patterns, we applied a hierarchical cluster analysis (HCA) to the variables resulting from the PCA. HCA is a method based on a hierarchical divisive and agglomerative approach where each data subset is associated to levels of the tree diagram (dendrogram).

## **External variables for KED**

The following were considered as potential external (independent) variables for the KED: spatio-temporal daily meteorological estimates, land use, population density, proximity to industry (weighted emissions), weighted road density and greenness tasseled-cap.

### ***Spatio-temporal meteorological variables***

We performed an IDW interpolation to estimate meteorological surfaces (with cells sizes of 100 m) of all meteorological variables (wind speed, temperature, relative humidity and precipitation) for all days of the study period (2010). This operation was carried out using the R software version 3.1.1 (R Project for Statistical Computing; <http://www.r-project.org>).

### ***Geographic variables***

Weighted road density, industrial emission-weighted proximity, land use (urban, industrial and commercial area), population density and the greenness component of the tasseled-cap image transformation were calculated using circular buffers of different radii around the geographical features using the procedures described below. Throughout the design of geographical layers, the radii were selected based on a priori knowledge of the range of influence of each variable represented in the layer. A compromise was made between investigating several possible ranges of influences by using different radii and by limiting the total number of layers to be generated and analyzed. All output variables were created in raster format (uniform grids) with cells of 100 x 100 meters using ArcGIS v.10.2 (ESRI).

Weighted road density: Road traffic is a source of particles and their precursors. We obtained road network data from the 2010 Road Layer Dataset (CanMap® RouteLogistics version 2010.3) compiled by Digital Mapping Technology Inc. (DMTI). We extracted the classes “highway” (expressway, principal and secondary highway), “major road” and “local road” from the complete road dataset and arbitrarily assigned weight value 3 to “highway”, 2 to “major road” and 1 to “local road” to represent their relative contribution to local concentrations of particles. Weights based on the hierarchical level of roads showed a similar accuracy as the density of traffic counts, as showed in previous studies (Hansell *et al.*, 2014; Rose *et al.*, 2009). The weighted road density (WRD) (Rose *et al.*, 2009) variable is used as a proxy for exposure to traffic-related pollution. A maximum distance of influence related to the traffic of 1 500 meters has been reported (Su *et al.*, 2009). Here we calculated the total kilometers of weighted road within 100, 150, 250 and 1 000 meter buffers around each cell

(100 x 100 meters) representing a measurement site. WRD was expressed in meters per square kilometer and is based on the following equation:

$$WRD = \sum_{i=1}^n L_i V_i / \pi r^2 \quad (1)$$

where  $L_i$  is the length of roads within the buffer,  $V_i$  the arbitrary weights applied to each road class, and  $r$  the radial size of the buffer.

#### PM<sub>2.5</sub> industrial emission-weighted proximity:

Industries are important emitters of PM<sub>2.5</sub>. PM<sub>2.5</sub> industrial emission data for the year 2010 was obtained from the National Pollutant Release Inventory (NPRI) of Environment Canada. They consist of point locations of the annual emissions volume of each industrial source. The exposure risk is defined as a ratio of the exposure risk value for the receptor to the average exposure risk value for all the receptors present within the study area (Zou *et al.*, 2009). The proximity model used (the emission-weighted proximity model, EWPM) assumes that exposure decreases with distance and increases with emissions (Zou *et al.*, 2009). EWPM is a simple approach used to estimate spatially the PM<sub>2.5</sub> industrial emissions in order to facilitate the implementation of the proposed method. The calculation of EWPM is based on: First, to simulate exposure at given points in space, we converted a uniform grid with a cell size of 100 meters to points (centroids of each cell to represent the receptors). Second, the

calculation of exposure intensity value (REWPM) at each point was based on the following EWPM equation:

$$R_{EWPMi} = n \left( \sum_{j=1}^m E_{i,j} T_{i,j} / D_{i,j} \right) / \sum_{i=1}^n \sum_{j=1}^m (E_{i,j} T_{i,j} / D_{i,j}) \quad (2)$$

where  $E_{i,j}$  and  $T_{i,j}$  are the emission rate and emission year of  $j^{\text{th}}$  emission source,  $i$  the index of receptors;  $D_{i,j}$  the distance separating  $i$  and  $j$ ; and  $n$  and  $m$  the number of exposure points and the number of emission sources respectively.

#### Land use data:

The land use types (residential, commercial and industrial) were extracted from DMTI Spatial data for the year 2010. These classes are used as surrogate measures for residential, commercial and industrial emissions related to fine particulate matter. The original land use polygons were converted to a raster grid in order to allow aggregating the area of each land use class within search radii of 300, 500 and 750 meters. We used these buffer sizes because these scales showed associations with  $PM_{2.5}$  in many studies (Brauer *et al.*, 2008; Henderson *et al.*, 2007; Ross *et al.*, 2007). We used neighborhood statistics (a function of the ArcGIS Spatial Analyst extension) to compute the sum of the area of each land use type within circular buffers. This calculation resulted in three layers (one for each land use type) for each circle's radius.

### Population density:

Households may emit  $PM_{2.5}$  from heating activities, proportionally to population density. To model this, we used data of population per dissemination area from the 2006 census from Statistics Canada. A dissemination area is a small census unit with a population of 400 to 700 persons (Statistic Canada, 2009). The dissemination area polygons were converted to centroids and the population count of each polygon was assigned to its centroid. We used kernel density to estimate the values within radii of 1 000 and 2 500 meters (Henderson *et al.*, 2007) Again, these buffer sizes were used because these sizes showed correlations with  $PM_{2.5}$  in similar studies (Brauer *et al.*, 2003; Henderson *et al.*, 2007; Hoek *et al.*, 2008) and because they better reflected the population distribution of larger dissemination areas that also include low population land uses, such as of commercial and industrial zones.

### Remote sensing greenness tasseled-cap transformation:

Vegetation can reduce levels of ambient  $PM_{2.5}$  and improve air quality (Nowak *et al.*, 2013). We extracted the greenness component from a tasseled-cap transformation of a Landsat Thematic Mapper (TM) image. The selected TM scene (path/row 014/028) for Montreal was captured on July 27, 2010. This image was orthorectified and provided by the U.S. Geological Survey (USGS). The tasseled-cap is an orthogonal transformation of reflective bands which isolates greenness, wetness and soil brightness (Kauth *et al.*, 1976). The greenness component exhibits high values for surfaces with high vegetation density and low values for soil surfaces. For each grid of 100 meters, the mean of the greenness component was calculated in a 500-



meter buffer, whereby this size was used because other studies demonstrated an association with PM<sub>2.5</sub> and other related pollutants in that buffer (Markevych *et al.*, 2014; Su *et al.*, 2010).

## **Spatio-temporal modeling**

To map spatial and temporal variation of PM<sub>2.5</sub> levels, we developed interpolation models based on mathematical and geostatistical approaches to interpolate values using the R software version 3.1.1 (R Project for Statistical Computing; <http://www.r-project.org>). The mathematical interpolation approach does not incorporate a statistical probability theory but uses mathematic formulas to interpolate values (e.g., IDW). The geostatistical approach uses statistical probability (randomness concept) within the interpolation process (e.g., KED). In this study, three approaches were used to estimate PM<sub>2.5</sub> concentrations in the City of Montreal for each day of the study period: kriging with external drift (KED), a simple inverse distance weighted (IDW) interpolation, and a hybridization of those two models (KED-IDW). To demonstrate the gain brought about by the three spatial interpolation models, we compared their results to the simple daily mean of PM<sub>2.5</sub> concentrations at the 10 monitoring stations.

KED is a non-stationary geostatistical method also referred to as universal kriging (UK) by some authors (Stein *et al.*, 1999; Wackernagel, 2003). The term “universal kriging” is reserved for the case where the trend is modeled as a function of coordinates (Matheron, 1969) and the expression “kriging with external drift” (Chiles, 1991) is employed for the specific case where the trend is modeled as a linear function of independent external variables. This linear function is used to derive the local mean (trend) of the dependent variable. Besides, this method requires relevant independent variables (in a physical sense) (Goovaerts, 1997). The

independent variable that is physically related to the dependent variable and measured at a high spatial density improves the estimation to create maps at the resolution of the independent variable.

KED uses a first (broad spatial trends) and second order effect (local variation) of spatial dependence (Pearce *et al.*, 2009), in this case the spatial dependence between the independent variables and PM<sub>2.5</sub> concentrations. Differing from a regression model where the model errors are assumed to be independent, KED uses variogram models under a dependence assumption of model errors. Assuming that the expected value of the dependent variable  $Z(u)$  (PM<sub>2.5</sub> concentrations) is a linear function of another regionalized variable  $Y(u)$  (independent meteorological and geographical variables), the random function of  $Z(u)$  can be expressed as follows:

$$Z(u) = c_1 + c_2 Y(u) + Y_R(u) \quad (3)$$

where  $c_1$  and  $c_2$  are constants,  $Y(u)$  is known everywhere and represents the value of the independent variable at location  $u$ , and  $Y_R(u)$  is the residual. Because of the limited number of ground stations available ( $n = 10$ ), we selected only two covariates to develop KED models in the present study. Independent variables  $Y(u)$  were selected based on the size of the Pearson correlation and their potential for reflecting the spatial variations of PM<sub>2.5</sub> concentrations. More precisely, we first selected the meteorological variable presenting the greatest correlation and coefficient of variation with daily PM<sub>2.5</sub> measures. We then selected the geographical variable that showed the greatest correlation with PM<sub>2.5</sub>.

In the KED model, the spatial dependence is computed using an empirical semi-variogram  $\hat{\gamma}(h)$ , a function calculated using the variability between data pairs of various distances and defined as:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(u_{\alpha}) - Z(u_{\alpha} + h)]^2 \quad (4)$$

where  $N(h)$  is the number of pairs of sampling locations at distance  $h$  from one another, and  $Z(u_{\alpha})$  is the observed value at the sampling site.

The concentration of  $PM_{2.5}$  at a given point of interest  $u$  is estimated according to:

$$Z_{KED}^*(u) = \sum_{\alpha=1}^n \lambda_{\alpha}(u) Z(u_{\alpha}) \quad (5)$$

where,  $\lambda_{\alpha}(u)$  are the weights corresponding to the  $n$  samples at location  $u$  and obtained by solving a system of linear equations developed by Goovaerts (1997) involving knowledge of the variogram.  $PM_{2.5}$  predictions are generated for time slices and based on the daily estimated variograms. Temporal interactions were not incorporated in this study.

The second model, IDW, is a simple spatial interpolation approach commonly used in air pollution epidemiology (Al-Hamdan *et al.*, 2009; Lipsett *et al.*, 2011). IDW calculates the values as a distance-weighted average of the observation values without any dependence on external variables. It can be expressed as:

$$Z_{IDW}^*(u) = \frac{\sum_{\alpha=1}^n \frac{1}{d^\lambda} Z(u_\alpha)}{\sum_{\alpha=1}^n \frac{1}{d^\lambda}} \quad (6)$$

where  $d^\lambda$  are the IDW weight distances and  $Z(u_\alpha)$  the observed values at the  $n$  measured sites. We integrated the KED and IDW models by selecting for each day the estimate that provided the best performance, namely by performing cross-validations.

## Cross-validation

A leave-one-out cross-validation approach was used for assessing the performance of each model and to find the best predictive model. In this approach, each station observation on a given day is removed from the data one at a time and the remaining data is used to predict the  $PM_{2.5}$  concentration at the location of the excluded station. The observations and predictions at the excluded stations are then compared. In order to compare interpolation models to simple daily mean  $PM_{2.5}$  concentrations, the same form of cross-validation was also applied to the daily mean. In other words, the performance of this approach was evaluated by calculating the difference between the  $PM_{2.5}$  concentration at a given station on a given day and the mean of the remaining ones.

We computed  $PM_{2.5}$  estimation errors as predicted values minus observed values across interpolation methods. The root mean square error (RMSE) was used to estimate the total magnitude of errors for each day, across monitoring stations and for all station-days. Finally, to further evaluate the performance of the hybrid model prediction, we used cross-validation results to create scatter plots of predicted versus observed values and to evaluate the

proportion of the variance explained by the model as well as the consistency and the model bias.

## 5 Results

Four factors were derived from the PCA, whereby the three first explained 87% of the total variance. We therefore performed the hierarchical cluster analysis using the first three principal components to classify each day of the study period into groups with similar meteorological conditions. Six synoptic weather classes with similar patterns were obtained.

Descriptive statistics of  $PM_{2.5}$  concentrations and meteorological variables by synoptic weather classes are shown in Table 1. The results for all synoptic classes show that  $PM_{2.5}$  concentrations decrease when wind speed increases. Days of the synoptic class 1 are spread over the year and correspond to conditions of high relative humidity and high precipitation; days of class 2 and 3 are spread over the winter season; days of class 4 and 6 are spread over the summer and fall seasons; and days of class 5 are found mostly in the spring season. As shown in Table 1, the synoptic class 3 has the typical characteristics of temperature inversion: low winter temperatures, high atmospheric pressure and low wind speed, to which are associated high levels of  $PM_{2.5}$ .

The Pearson correlation coefficients for geographic variables and coefficients of variation for the spatio-temporal meteorological variables that were used to select the KED

covariables are presented in Table 2 and 3. The wind speed variable shows high correlations and explains the spatial variability of PM<sub>2.5</sub> concentrations for all synoptic classes (Table 3).

The following geographic predictors were selected for the KED model: greenness for synoptic class 1 to 4 and industrial area for synoptic class 5 and 6; wind speed was the meteorological variable selected for all synoptic classes.

Table 4 describes the cross-validation results for the three models for the days of the year 2010 at the 10 stations (n = 3 344). Overall, the hybrid model KED-IDW performed better for all synoptic classes compared to the IDW and KED models taken separately, although only marginally compared to IDW. The two performance metrics, R<sup>2</sup> and RMSE, show different behaviors when comparing the models. The proportion of explained variance is almost always the lowest in the case of the mean of the stations, making it the worst approach. However, the highest RMSEs were in all cases but one observed for KED. The improvements brought about by a spatial interpolation approach are very significant for some synoptic classes (2, 4, 5), but rather small for other classes (3 and 6).

The distribution of daily PM<sub>2.5</sub> errors also shows that the IDW and KED-IDW models overestimate PM<sub>2.5</sub> levels, while KED underestimates them (errors respectively of 0.42, 0.24 and -0.22 µg/m<sup>3</sup>) (Figure 2).

Figure 3 shows the individual prediction RMSEs of the IDW and KED models (Figure 3A), and of the hybrid KED-IDW model (Figure 3B). For the two individual models, the error

is quite variable, oscillating mostly between 1 and 4  $\mu\text{g}/\text{m}^3$ , with a few large spikes. The RMSEs of the KED-IDW model show much less variation, being generally confined between 1 and 3  $\mu\text{g}/\text{m}^3$ , with only one spike. In the majority of cases, the RMSE of IDW is lower than that of KED (Figure 3A), but KED outperformed IDW for several days, featuring an RMSE that is sometimes 1  $\mu\text{g}/\text{m}^3$  lower than IDW and being often in the same range of error. Figure 4 presents the scatter plots of predicted values modeled with KED-IDW versus monitoring site values ( $\text{PM}_{2.5}$  daily levels observed) for the six synoptic classes. Scatter around the 1:1 line is rather moderate and quite similar between one class and the next. Lower  $R^2$  values occurred when the range of observed values was itself smaller. Figure 5 shows that the RMSE of the hybrid model in space (at all stations) was closest to zero in comparison with the KED and IDW models and more uniform in space.

It also reveals that errors of the KED-IDW model are higher when stations are scarce (e.g., southwest of the Island of Montreal). In Figure 6, we present two examples of 100 m  $\text{PM}_{2.5}$  resolution maps for a spring day (April 22, 2010) and a summer day (July 16, 2010) generated with the KED and IDW models. In these examples, for the selected spring day, IDW produced better accuracy (RMSE = 1.32 and  $R^2 = 0.60$ ) than KED (RMSE = 1.53 and  $R^2 = 0.47$ ); and for the selected summer day, KED produced better accuracy (RMSE = 1.26 and  $R^2 = 0.72$ ) than IDW (RMSE = 1.74 and  $R^2 = 0.46$ ), likely due to correlations with covariates. However, it is clear that KED produces a higher resolution map of  $\text{PM}_{2.5}$  level variations, as evidenced by the more complex patterns visible on the KED maps.

## 6 Discussion

In this paper we show that it is possible to reduce the interpolation errors in estimating  $PM_{2.5}$  levels with the combination of the commonly used IDW approach and the KED model. Our results show that the estimates of  $PM_{2.5}$  across monitoring sites with the hybrid model were on average 11% higher in  $R^2$  and on average  $0.60 \mu\text{g}/\text{m}^3$  lower in RMSE compared to KED. Compared to the IDW model, the hybrid model was only slightly higher, on average 3% in  $R^2$  and slightly lower, on average  $0.22 \mu\text{g}/\text{m}^3$ , in RMSE. IDW also performed better than KED, but the hybrid model was best. Nonetheless, the KED model produced better accuracy for specific days when  $PM_{2.5}$  observations showed strong associations and spatial dependency upon certain environmental and meteorological conditions.

We also found that the decomposition of the temporal distribution of meteorological factors in synoptic groups using principal components and hierarchical cluster analyses demonstrated an effective classification showing an important relationship between fine particulate matter and meteorological conditions. This approach allowed us to classify the temporal variation in  $PM_{2.5}$  due to meteorological effects while selecting covariates according to their effect level. As a result, for all models, the accuracy ( $R^2$  and RMSE) varies with the meteorological conditions captured by each synoptic class.

Univariate approaches such as IDW (Hystad *et al.*, 2012) or ordinary kriging (OK) (Chen *et al.*, 2012; Pearce *et al.*, 2009) and multivariate approaches such as KED (Pearce *et al.*, 2009) and LUR (Chen *et al.*, 2012; Hystad *et al.*, 2012; Pearce *et al.*, 2009) have been used



to estimate ambient particulate levels. The errors of these approaches vary widely from one study to another. For example, using KED to estimate PM<sub>2.5</sub> daily levels, Pearce *et al.* (2009) reported RMSE values of 0.49 µg/m<sup>3</sup>, while Beelen *et al.* (2009) reported RMSE of 5.19 µg/m<sup>3</sup> for particulate levels with a median diameter of 10 µg/m<sup>3</sup>. This probably depends on the density of stations (number of stations per km<sup>2</sup>) and on the scale of the study (regional, rural or urban). It is therefore difficult to compare our findings to other studies.

Like others have also observed Yu *et al.* (2009) we found that the error estimates from the models were higher in areas where the monitoring stations were sparse, such as in the South-West of the study area (Figure 5).

To our knowledge, KED and IDW have not been compared before. Although the results of this research show an overall better performance for IDW compared to KED, further investigation is needed to confirm this performance. This stems from the very concept of KED, in which an association between the dependent variable and external variables is expected. If, however, the correlation between these variables and the PM<sub>2.5</sub> levels is low, noise is introduced in the interpolation, thus creating inaccurate predictions. It is therefore important to seek the best correlated external variables, and to evaluate their usefulness for each synoptic weather class.

The approach that we used to spatialize the meteorological observations from the stations over the study area is too simple to capture the full complexity of the phenomena. We recognize that parameters such as wind speed do not necessarily vary linearly between

stations. The 3D configuration of built areas for example has an important influence on the spatial behavior of this parameter. Again, we have chosen methods that allow a rather straightforward implementation and focused on studying the benefits of KED.

As mentioned previously, our model was based on only 10 monitoring sites and a limited number of covariates (the selection of predictors is limited to two). The small number of monitoring stations may have produced instability in the variogram estimation for the KED model. Another constraint in applying the proposed hybrid model approach is the substantial computational time required to create daily maps at a 100-meter resolution.

Regardless of the limitations mentioned above, the hybrid KED-IDW model may improve the estimation of exposure of populations to estimate acute effects of  $PM_{2.5}$  because it generates higher resolution maps with smaller errors in space and time than other approaches often used in the field (e.g. closest monitor or only IDW). The predictive power of the hybrid model thus could reduce the exposure misclassification in epidemiological studies.

For implementation in future health studies, the hybrid interpolation model might be improved by expanding the number of monitoring sites and the inclusion of additional covariates. Furthermore, the choice of variables for a daily prediction could be improved by an automatic procedure depending of the degree of fit with daily  $PM_{2.5}$  measures.

Ambient levels of  $PM_{2.5}$  as interpolated in the present study provide imperfect estimates of true exposure, as individuals are exposed to different levels of pollution

depending on their locations across time. As suggested by others (Kwan, 2009, 2012), results from intra-urban interpolation models could be combined in future work, with information on time-activity patterns, in order to estimate individual exposure in future health studies

## **7 Conclusions**

The main objective of this study was to map the spatial and temporal variation of fine particulate matter in order to improve  $PM_{2.5}$  exposure predictions within an urban environment. The hybrid KED-IDW model was more effective for exposure prediction and outperformed the KED and IDW models. This study demonstrated the improvements brought about by a hybrid KED-IDW model for mapping  $PM_{2.5}$  concentrations over time and space in the Montreal area.

Our study also shows that it is possible to develop high resolution maps with a limited number of  $PM_{2.5}$  monitors under specific environmental and meteorological conditions. The integration of covariates in KED intends to allow capturing the intra-urban variation of  $PM_{2.5}$  levels; however, the model could perform better if there were more monitoring stations on the Island of Montreal.

## **8 Conflict of interest**

The authors declare no conflict to interest.

## 9 Acknowledgements

The authors would like to acknowledge Allan Brand for providing a compilation of meteorological and ground-level PM<sub>2.5</sub> data for this study. We also thank Éric de Montigny for comments on an earlier version of this paper. This project was financially supported by the Quebec Institute of Public Health.

## 10 References

- Adam-Poupart, A., Brand, A., Fournier, M., Jerrett, M., and Smargiassi, A. (2014). Spatiotemporal modeling of ozone levels in Quebec (Canada): a comparison of kriging, land-use regression (LUR), and combined bayesian maximum entropy–LUR approaches. *Environmental Health Perspectives*, 122(9), 970–976.
- Al-Hamdan, M. Z., Crosson, W. L., Limaye, A. S., Rickman, D. L., Quattrochi, D. A., Estes, M. G., Qualters, J. R., Sinclair, A. H., Tolsma, D. D., Adeniyi, K. A., and Niskar, A. S. (2009). Methods for characterizing fine particulate matter using ground observations and remotely sensed data: potential use for environmental public health surveillance. *Journal of the Air & Waste Management Association*, 59, 865–881.
- Beckerman, B., Jerrett, M., Serre, M., Martin, R., Lee, S., Van Donkelaar, A., Ross, Z., Su, J., and Burnett, R. (2013). A hybrid approach to estimating national scale spatiotemporal variability of PM<sub>2.5</sub> in the contiguous United States. *Environmental Science & Technology*, 7233–7241.
- Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., De Hooghd, K., and Briggs, D. J. (2009). Mapping of background air pollution at a fine spatial scale across the European Union. *Science of the Total Environment*, 407, 1852–1867.

- Bell, M., Ebisu, K., Peng, R., Walker, J., Samet, J., Zeger, S., and Dominici, F. (2008). Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999–2005. *American Journal of Epidemiology*, 168(11), 1301–1310.
- Brauer, M., Hoek, G., van Vliet, P., Meliefste, K., Fischer, P., Gehring, U., Heinrich, J., Cyrys, J., Bellander, T., Lewne, M., and Brunekreef, B. (2003). Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology*, 14(2), 228–239.
- Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., and Karr, C. (2008). A cohort study of traffic-related air pollution impacts on birth outcomes. *Environmental Health Perspectives*, 116(5), 680–686.
- Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pyl, K., Van Reeuwijk, H., Smallbone, K., and Van Der Veen, A. (1997). Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science of the Total Environment*, 11(7), 699–718.
- CCME. (2011). Ambient air monitoring protocol for PM<sub>2.5</sub> and ozone - Canada-wide standards for particulate matter and ozone. Canadian Council of Ministers of the Environment Winnipeg, MB, 978-1-896997-99-5 PDF.
- Chen, C., Wu, C., Yu, H., Chan, C., and Cheng, T. (2012). Spatiotemporal modeling with temporal-invariant variogram subgroups to estimate fine particulate matter PM<sub>2.5</sub> concentrations. *Atmospheric Environment*, 54, 1–8.

- Chiles, J. P. (1991). Application du krigeage avec dérive externe à l'implantation d'un réseau de surveillance piézométrique. *Sciences de la Terre*, 30, 131–147.
- Dawson, J. P., Adams, P. J., and Pandis, S. N. (2007). Sensitivity of PM<sub>2.5</sub> to climate in the Eastern US: a modeling case study. *Atmospheric Chemistry and Physics*, 7, 4295–4309.
- Environment Canada. (2012). Canadian smog science assessment. Gouvernement of Canada, En88-5/2011E-PDF.
- Environment Canada. (2013a). National Air Pollution Surveillance Network (NAPS) Data Products. Retrieved 10 March 2013, from <http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx>
- Environment Canada. (2013b). Canada National Climate Data and Information Archive. Retrieved 5 March 2013, from [http://climate.weather.gc.ca/index\\_e.html](http://climate.weather.gc.ca/index_e.html)
- Environment Canada. (2014). Major smog episodes. Retrieved 20 January 2014, from <http://www.ec.gc.ca/info-smog/default.asp?lang=En&n=669E620B-1>
- Goldman, G., Mulholland, J., Russell, A., Gass, K., Strickland, M., and Tolbert, P. (2012). Characterization of ambient air pollution measurement error in a time-series health study using a geostatistical simulation approach. *Atmospheric Environment*, 57, 101–108.

- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York, USA: Oxford University Press.
- Hansell, A. L., Rose, N., Cowie, C. T., Belousova, E. G., Bakolis, I., Ng, K., Toelle, B. G., and Marks, G. B. (2014). Weighted Road Density and Allergic Disease in Children at High Risk of Developing Asthma. *PLoS One*, 9(6), 1-9.
- Henderson, S., Beckerman, B., Jerrett, M., and Brauer, M. (2007). Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science & Technology*, 41, 2422–2428.
- Hoek, G., Beelen, R., De Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., and Briggs, D. J. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42, 7561–7578.
- Hufty, A. (1982). Analyse en composantes principales des situations synoptiques au Québec. *Géographie physique et Quaternaire*, XXXVI(3), 307–314.
- Hystad, P., Demers, P., Johnson, K., Brook, J., Van Donkelaar, A., Lamsal, L., Martin, R., and Brauer, M. (2012). Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. *Environmental Health*, 11, 1–13.
- Ito, K., Mathes, R., Ross, Z., Nádas, A., Thurston, G., and Matte, T. (2011). Fine particulate matter constituents associated with cardiovascular hospitalizations and mortality in New York City. *Environmental Health Perspectives*, 119(4), 467–473.



- Jerrett, M., Burnett, R., Goldberg, M., Sears, M., Krewski, D., Catalan, R., Kanaroglou, P., Giovis, C., and Finkelstein, N. (2003). Spatial analysis for environmental health research: concepts, methods, and examples. *Journal of Toxicology and Environmental Health, Part A*, 66, 1783–1810.
- Kauth, R., and Thomas, G. (1976). The tasseled Cap - A graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. Paper presented at the Symposium on Machine Processing of Remote Sensed Data, West Lafayette, Indiana.
- Kwan, M.-P. (2009). From place-based to people-based exposure measures. *Social Science & Medicine*, 69, 1311–1313.
- Kwan, M.-P. (2012). The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers*, 102(5), 958–968.
- Lipsett, M. J., Ostro, B. D., Reynolds, P., Goldberg, D., Hertz, A., Jerrett, M., Smith, D. F., Garcia, C., Chang, E., and Bernstein, L. (2011). Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study Cohort. *American Journal of Respiratory and Critical Care Medicine*, 184, 828–835.
- Malek, E., Davis, T., Martin, R., and Silva, P. (2006). Meteorological and environmental aspects of one of the worst national air pollution episodes (January, 2004) in Logan, Cache Valley, Utah, USA. *Atmospheric Research*, 79, 108–112.

- Markevych, I., Fuertes, E., Tiesler, C., Birk, M., Bauer, C., Koletzko, S., Von Berg, A., Berdel, D., and Heinrich, J. (2014). Surrounding greenness and birth weight: results from the GINIplus and LISApplus birth cohorts in Munich. *Health & Place*, 26, 39–46.
- Matheron, G. (1969). Le krigeage universel. In ENSMP (Ed.), *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau* (Vol. 1, pp. 83): École Nationale Supérieure des Mines de Paris.
- Munoz-Alpizar, R., and Blanchet, J. (2003). Application of the NARCM model to high-resolution aerosol simulations: Case study of Mexico City basin during the Investigacion sobre Materia Particulada y Atmosférico-Aerosol and Visibility Research measurements campaign. *Journal of Geophysical Research*, 108(D15), 1–14.
- Nowak, D., Hirabayashi, S., Bodine, A., and Hoehn, R. (2013). Modeled PM<sub>2.5</sub> removal by trees in ten U.S. cities and associated health effects. *Environmental Pollution*, 178, 395–402.
- Pearce, J. L., Rathbun, S. L., Aguilar-Villalobos, M., and Naeher, L. P. (2009). Characterizing the spatiotemporal variability of PM<sub>2.5</sub> in Cusco, Peru using kriging with external drift. *Atmospheric Environment*, 43, 2060–2069.
- Pinto, J., Lefohn, A., and Shadwick, D. (2012). Spatial variability of PM<sub>2.5</sub> in urban areas in the United States. *Journal of the Air & Waste Management Association*, 54, 440–449.

- Rose, N., Cowie, C., Gillett, R., and Marks, G. (2009). Weighted road density: a simple way of assigning traffic-related air pollution exposure. *Atmospheric Environment*, 43, 5009–5014.
- Ross, Z., Jerrett, M., Ito, K., Tempalski, B., and Thurston, G. (2007). A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment*, 41, 2255–2269.
- Statistic Canada. (2009). Dissemination area (DA). Retrieved 10 April 2012, from <http://www12.statcan.ca/census-recensement/2006/ref/dict/geo021-eng.cfm>
- Stein, A., Van der Meer, F., and Gorte, B. (1999). *Spatial Statistics for Remote Sensing* (Vol. 1, pp 284). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Su, J., Jerrett, M., and Beckerman, B. (2009). A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *Science of the Total Environment*, 407, 3890–3898.
- Su, J., Jerrett, M., Beckerman, B., Verma, D., Arain, M. A., Kanaroglou, P., Stieb, D., Finkelstein, M., and Brook, J. (2010). A land use regression model for predicting ambient volatile organic compound concentrations in Toronto, Canada. *Atmospheric Environment*, 44, 3529–3537.

- Tai, A. P. K., Mickley, L. J., and Jacob, D. J. (2010). Correlations between fine particulate matter (PM<sub>2.5</sub>) and meteorological variables in the United States: Implications for the sensitivity of PM<sub>2.5</sub> to climate change. *Atmospheric Environment*, 44, 3976–3984.
- U.S. EPA. (2009). Final Report: Integrated Science Assessment for Particulate Matter. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-08/139F, 2009.
- Ville de Montréal. (2010). Environmental assessment report - Air quality of Montreal (pp. 8): Réseau de surveillance de la qualité de l'air, Direction de l'environnement et du développement durable, Ville de Montréal.
- Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications* (3rd ed. pp 388): Springer Berlin Heidelberg.
- Yu, H., Chen, J., Christakos, G., and Jerrett, M. (2009). BME Estimation of Residential Exposure to Ambient PM<sub>10</sub> and Ozone at Multiple Time Scales. *Environmental Health Perspectives*, 110(4), 537–544.
- Zanobetti, A., and Schwartz, J. (2009). The Effect of Fine and Coarse Particulate Air Pollution on Mortality: A National Analysis. *Environmental Health Perspectives*, 117(6), 898–903.
- Zou, B., Wilson, J. G., Zhan, F. B., and Zeng, Y. (2009). An emission-weighted proximity model for air pollution exposure assessment. *Science of the Total Environment*, 407, 4939–4945.

## 11 Tables and figures

Table 1. Descriptive statistics for meteorological variables and PM<sub>2.5</sub> for 2008–2010\*

Synoptic weather class	N (days)	PM <sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )		Precipitation (mm)		Temperature ( $^{\circ}\text{C}$ )		Relative humidity (%)		Wind speed (km/h)		Atmospheric pressure (kPa)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	114	12.36	5.96	15.24	9.00	9.35	9.50	86.91	5.54	12.35	3.33	100.38	0.73
2	149	7.20	3.33	2.08	3.59	-2.85	7.12	69.35	11.32	18.50	3.71	100.65	0.94
3	250	13.74	8.54	0.84	1.54	-3.08	7.48	74.28	9.81	9.93	2.93	101.24	0.81
4	196	11.62	5.83	1.65	2.59	15.15	5.83	74.92	8.76	7.47	1.72	101.04	0.73
5	85	9.48	4.19	0.36	1.17	12.00	4.00	48.21	6.54	12.32	3.69	101.16	0.66
6	155	10.60	6.27	0.88	1.70	18.21	4.81	68.63	6.32	12.51	2.77	100.82	0.61
Total days:	949												

\* Each daily value used represents the MEAN for the 8-10 monitoring stations over the island of Montreal, of the daily mean of PM<sub>2.5</sub> levels or meteorological information

Table 2. Pearson's correlation coefficients between PM<sub>2.5</sub> concentrations and independent geographic variables by synoptic weather class for the year 2010.

Categories	Buffer (m)	Correlation coefficients (r)					
		1 (N=436)	2 (N=513)	3 (N=825)	4 (N=550)	5 (N=351)	6 (N=669)
Greenness tasseled-cap	500	-0.20**	-0.38**	-0.13**	-0.30**	-0.30**	-0.19**
Industrial (ha/πkm <sup>2</sup> )	300	0.17**	0.29**	0.06	0.29**	0.32**	0.18**
	500	0.18**	0.31**	0.06	0.29**	0.32**	0.19**
	750	0.17**	0.30**	0.06	0.29**	0.32**	0.19**
Commercial (ha/πkm <sup>2</sup> )	300	---	---	---	---	---	---
	500	0.06	0.07	0.03	0.10*	0.11*	0.09*
	750	0.14**	0.21**	0.07	0.20**	0.20**	0.15**
Residential (ha/πkm <sup>2</sup> )	300	-0.13**	-0.18**	-0.02	-0.13**	-0.10	-0.08*
	500	-0.17**	-0.24**	-0.05	-0.16**	-0.13*	-0.10**
	750	-0.17**	-0.24**	-0.06	-0.15**	-0.13*	-0.10**
PM <sub>2.5</sub> industrial emission-weighted proximity	---	0.01	0.07	0.02	0.07	0.04	0.06
Population density (hab/km <sup>2</sup> )	1000	0.04	0.13**	0.04	0.13**	0.15**	0.05
	2500	0.13**	0.26**	0.06	0.26**	0.27**	0.13**
Weighted road density (m/πkm <sup>2</sup> )	100	-0.01	-0.02	0.01	0.04	0.08	0.02
	150	0.03	0.04	0.03	0.08	0.11*	0.06
	250	0.01	0.02	0.03	0.06	0.09	0.05
	1000	0.05	0.10	0.03	0.14**	0.14**	0.09*

\*p < 0.05; \*\* p < 0.01.

Table 3. Pearson's correlation coefficients and coefficients of variation based on daily measurements at NAPS station locations between PM<sub>2.5</sub> concentrations and spatio-temporal meteorological variables for the year 2010.

<i>Synoptic weather class</i>	<i>Temperature</i>		<i>Relative humidity</i>		<i>Wind speed</i>	
	<i>Correlation coefficient (r)</i>	<i>Coefficient of variation (%)</i>	<i>Correlation coefficient (r)</i>	<i>Coefficient of variation (%)</i>	<i>Correlation coefficient (r)</i>	<i>Coefficient of variation (%)</i>
1	0.36	0.09	-0.20	1.95	-0.41	20.78
2	0.21	0.07	-0.07	2.59	-0.39	24.97
3	-0.06	0.14	0.29	3.23	-0.53	16.63
4	0.32	0.13	-0.11	4.23	-0.35	14.89
5	0.47	0.13	-0.04	7.15	-0.41	17.74
6	0.76	0.09	-0.05	3.43	-0.22	21.40

Table 4. Leave-one-station-out cross-validation results for the PM<sub>2.5</sub> prediction for the year 2010 from KED, IDW and KED-IDW models.

<i>Synoptic weather class</i>	<i>Covariables</i>	<i>N (days)</i>	<i>N (estimated points)</i>	<i>Model</i>	<i>R<sup>2</sup></i>	<i>RMSE (µg/m<sup>3</sup>)</i>
1	Wind speed and greenness (Tcap)	46	436	KED	0.77	2.73
				IDW	0.86	2.09
				KED-IDW	0.87	1.97
				Mean*	0.79	2.58
2	Wind speed and greenness (Tcap)	55	513	KED	0.46	2.36
				IDW	0.61	2.00
				KED-IDW	0.66	1.89
				Mean*	0.48	2.32
3	Wind speed and greenness (Tcap)	88	825	KED	0.86	3.34
				IDW	0.91	2.67
				KED-IDW	0.92	2.54
				Mean*	0.89	2.92
4	Wind speed and greenness (Tcap)	57	550	KED	0.66	2.75
				IDW	0.79	2.19
				KED-IDW	0.80	2.13
				Mean*	0.66	2.75
5	Wind speed and industrial (500m)	36	351	KED	0.69	2.45
				IDW	0.77	2.08
				KED-IDW	0.79	2.00
				Mean*	0.66	2.53
6	Wind speed and industrial (750m)	71	669	KED	0.89	2.55
				IDW	0.92	2.10
				KED-IDW	0.93	1.98
				Mean*	0.89	2.54

\* Daily mean leave-one-station-out cross-validation is computed



Figure 1. Geographical location of PM<sub>2.5</sub> stations and meteorological stations in the study region.

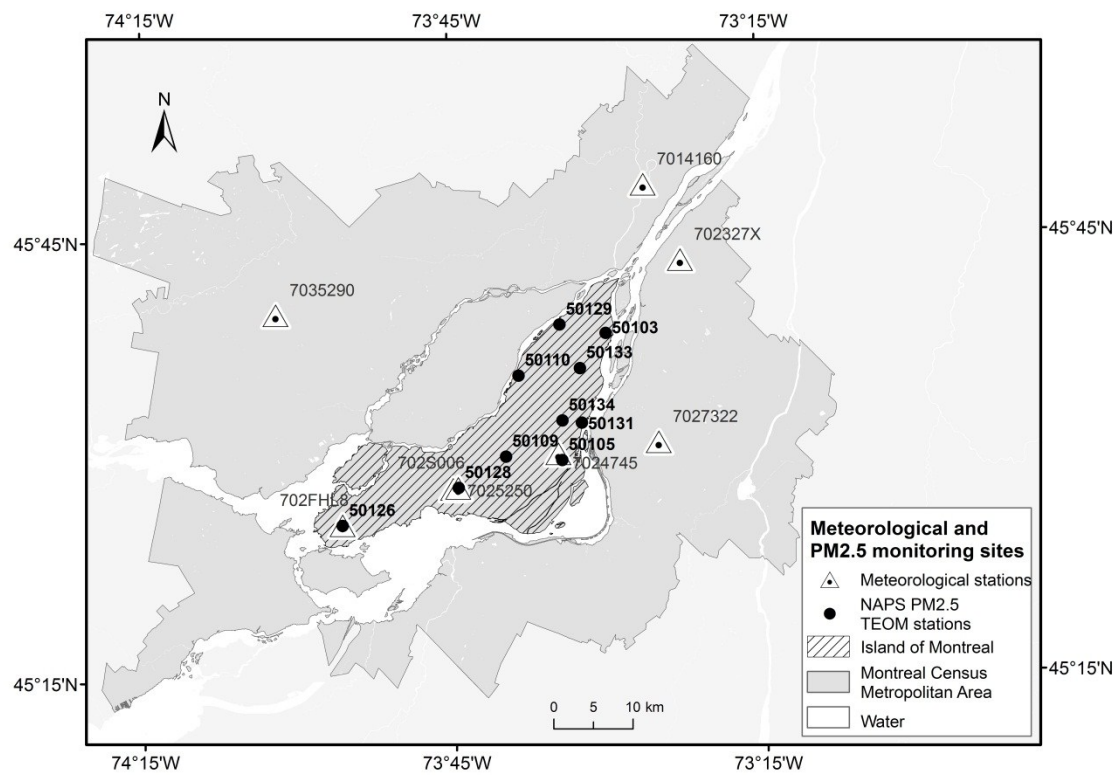


Figure 2. PM<sub>2.5</sub> mapping error estimates from cross validation [where error = predicted – observed PM<sub>2.5</sub> concentration (in μg/m<sup>3</sup>) at each monitoring station] based in KED (mean ± SD; -0.22 ± 2.78 μg/m<sup>3</sup>), IDW (mean ± SD; 0.42 ± 2.21 μg/m<sup>3</sup>) and KED-IDW (mean ± SD; 0.24 ± 2.13 μg/m<sup>3</sup>) for days of the year 2010.

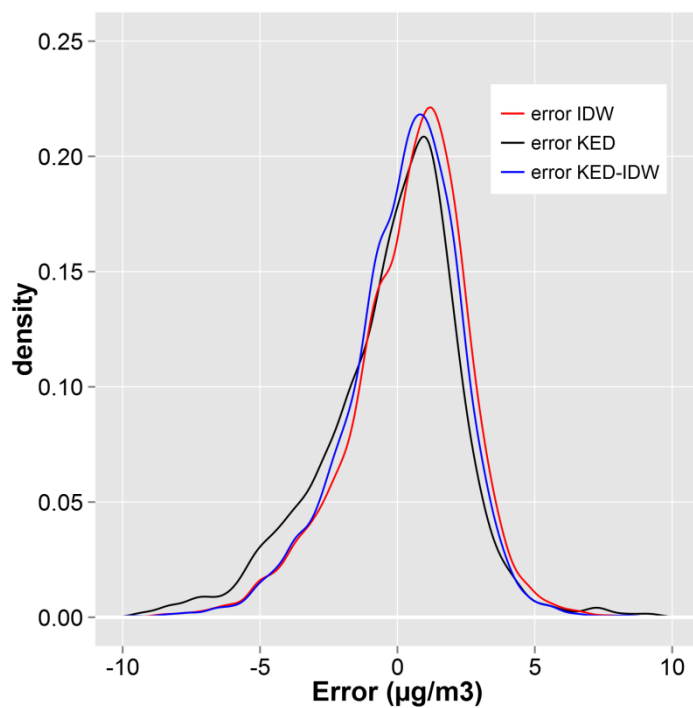


Figure 3. Daily temporal PM<sub>2.5</sub> error estimates (RMSE) based on leave-one-station cross-validation for KED, IDW (A) and KED-IDW models (B).

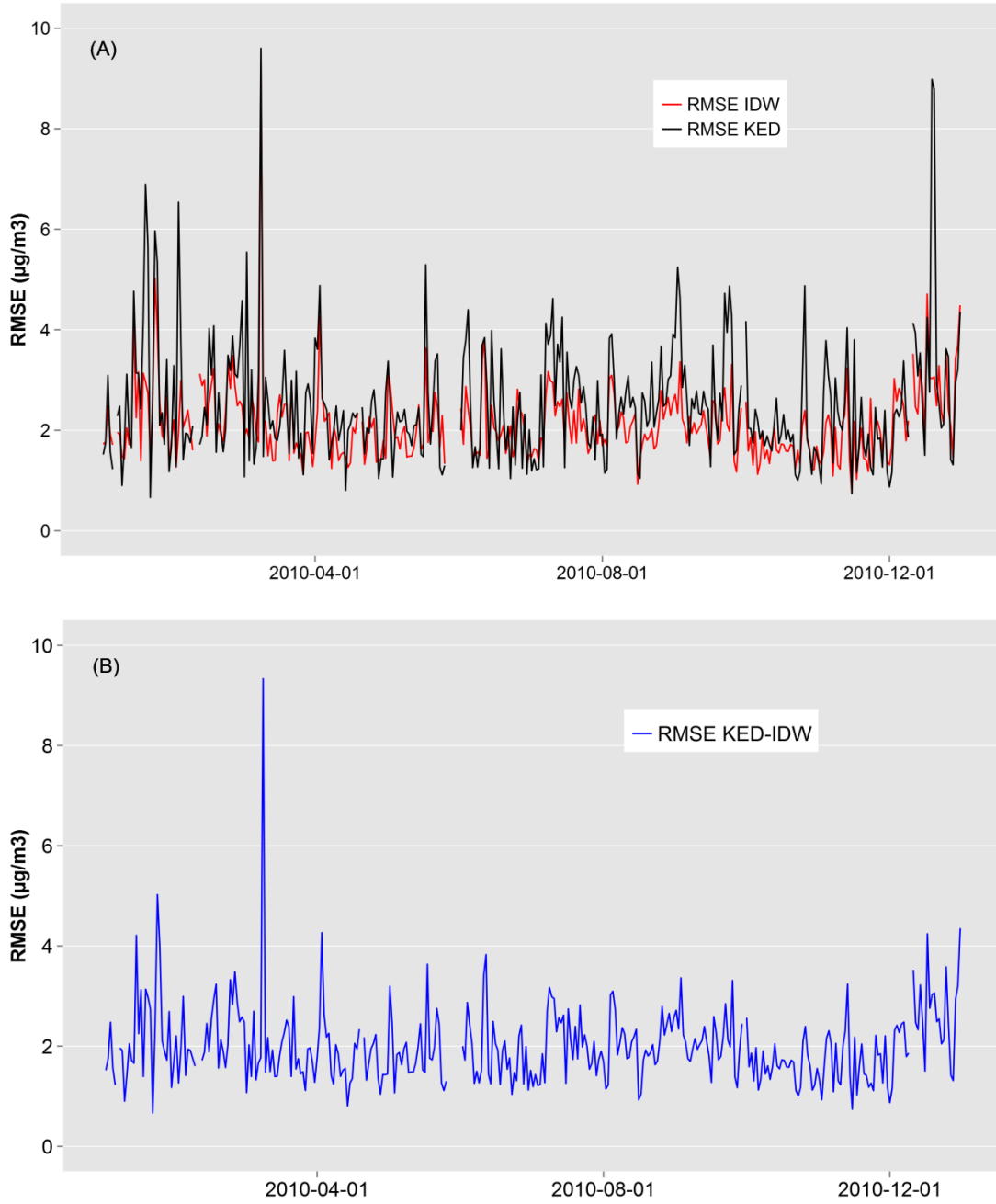


Figure 4. PM<sub>2.5</sub> daily levels predicted with the hybrid model KED-IDW vs. PM<sub>2.5</sub> concentrations measured at NAPS stations for the year 2010 by synoptic weather class.

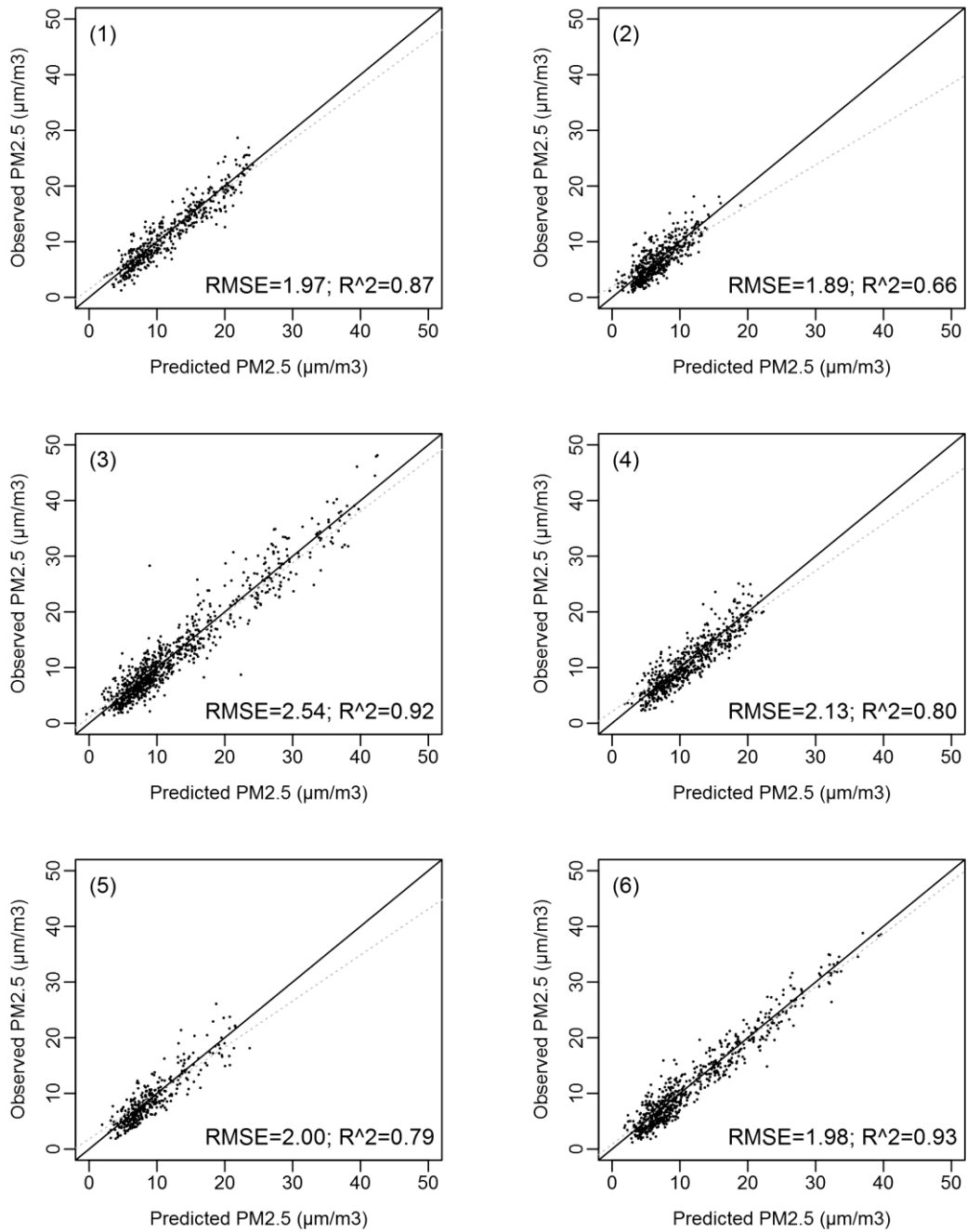


Figure 5. Spatial distribution of mean PM<sub>2.5</sub> error estimates (RMSE) of the year 2010 based on cross-validation for KED (A), IDW (B) and KED-IDW (C).

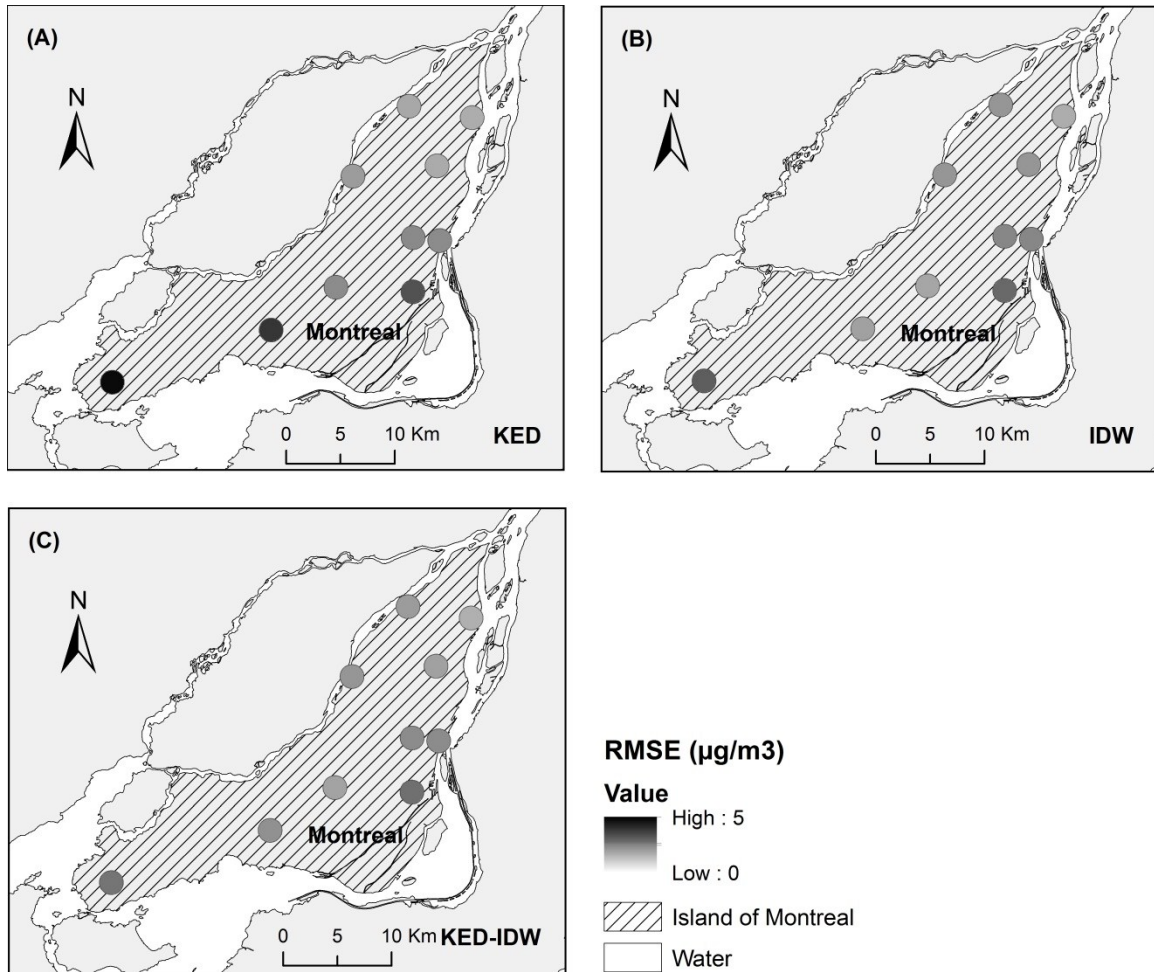
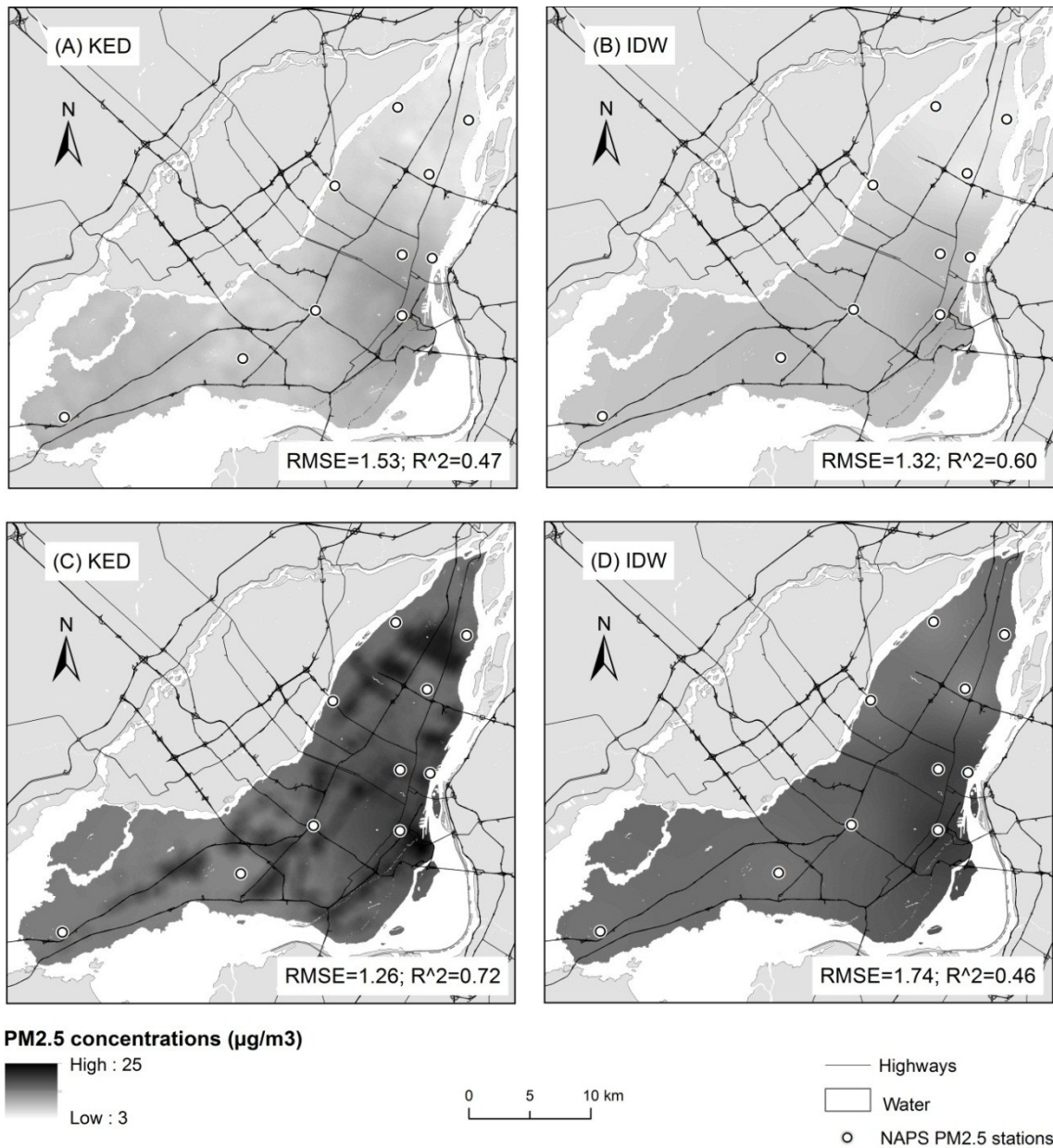


Figure 6. Predicted PM<sub>2.5</sub> concentrations for April 22nd, 2010 : (A) KED (covariates: wind speed and greenness) and (B) IDW. Predicted PM<sub>2.5</sub> concentrations for July 16, 2010: (C) KED (covariates: wind speed and industrial area) and (D) IDW.





**Chapitre 4 : Spatio-temporal modeling of daily  
concentrations of ground-level ozone in Montreal, Canada  
– Article 2**





**Title:** Spatio-temporal modeling of daily concentrations of ground-level ozone in Montreal, Canada

*\*Manuscript submitted to Atmospheric Environment*

**Authors :** Yuddy Ramos<sup>a</sup>, Benoît St-Onge<sup>b</sup>, Jean-Pierre Blanchet<sup>c</sup>, Yan Kestens<sup>d,e</sup>, Audrey Smargiassi<sup>f,g</sup>

<sup>a</sup>Département de géographie, Université de Montréal, Montréal, Québec, Canada

<sup>b</sup>Département de géographie, Université du Québec à Montréal (UQAM), Montréal, Québec, Canada

<sup>c</sup>Département des sciences de la Terre et de l'atmosphère, Université du Québec à Montréal (UQAM), Montréal, Québec, Canada

<sup>d</sup>Département de médecine sociale et préventive, Université de Montréal, Montréal, Québec, Canada

<sup>e</sup>Centre de recherche du Centre hospitalier de l'Université de Montréal (CHUM), Québec, Canada

<sup>f</sup>Département de santé environnementale et de santé au travail, Université de Montréal, Montréal, Québec, Canada

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

### **Contribution de l'auteure principale et des coauteurs :**

L'article présenté dans ce chapitre a été réalisé par Yuddy Ramos sous la supervision de Benoît St-Onge, la cosupervision d'Audrey Smargiassi, et la collaboration de Jean-Pierre Blanchet et Yan Kestens.

Yuddy Ramos a conçu l'approche méthodologique et réalisé les traitements et analyses de données. Elle a rédigé la première version de l'article et effectué toutes les modifications selon les suggestions des coauteurs. Elle a aussi produit la version révisée et répondu aux évaluateurs de l'article jusqu'à sa publication.

Benoît St-Onge a supervisé la conception méthodologique et l'analyse des résultats. Il a révisé toutes les versions de l'article avant soumission. Il a supervisé le travail de révision de l'article et les réponses aux évaluateurs suite au processus d'évaluation par les pairs.

Audrey Smargiassi a cosupervisé la conception méthodologique. Elle a fourni les données de qualité de l'air et météorologiques, ainsi que le soutien financier. Elle a révisé la version initiale et toutes les versions avant publication et cosupervisé la révision de l'article et les réponses aux évaluateurs suite au processus d'évaluation par les pairs.

Jean-Pierre Blanchet et Yan Kestens ont fourni leur avis scientifique d'ensemble. Ils ont été consultés lors du développement méthodologique et de l'analyse des résultats. Ils ont aussi relu les versions avancées de l'article et corrigé la version finale avant sa soumission.

# 1 Abstract

Ground-level ozone ( $O_3$ ) is a powerful oxidizing agent and a harmful pollutant to human health, forests and crops. Estimating  $O_3$  exposure is a challenge because it exhibits complex spatio-temporal patterns. The aim in this study was to provide high-resolution maps of ozone for the metropolitan area of Montreal, Canada. For this, we assessed the kriging with external drift (KED) model to estimate  $O_3$  concentration by synoptic weather classes for 2010. We compared these results with ordinary kriging (OK), and a simple average of 12 monitoring stations. We also compared the estimates obtained for the 2010 summer months with estimates from a Bayesian maximum entropy (BME) model (Adam-Poupart *et al.*, 2014). The KED model with road and vegetation density as covariates showed high performance for all six synoptic classes (daily  $R^2$  estimates ranging from 0.77 to 0.92 and RMSE from 2.79 to 3.37 ppb). For the summer months, the model using KED proved to be the best model ( $R^2 = 0.92$ ; RMSE = 3.14 ppb), followed by the OK model ( $R^2 = 0.85$ , RMSE = 4 ppb). Although errors appear to be substantially reduced with the KED model, future work integrating the temporal dependency in the data is needed to not overstate the performance of the KED model.

**Keywords:** air pollution, ground-level ozone, kriging with external drift, geostatistics, environment

## 2 Introduction

Ground-level ozone or tropospheric ozone is a secondary photochemical pollutant produced by anthropogenic and natural chemical precursors (Gibson *et al.*, 2009; U.S. EPA, 2013). Tropospheric ozone ( $O_3$ ) is formed by the interaction of nitrogen oxides ( $NO_x$ ) and volatile organic compounds (VOC) in the presence of sunlight (McKendry, 1993).  $NO_x$  is mainly composed of nitrogen monoxide (NO) and nitrogen dioxide ( $NO_2$ ). The ground-level  $O_3$  is produced by the photodissociation of  $NO_2$  in the presence of sunlight. In turn, the NO formed through this process catalytically destroys the  $O_3$  to reform  $NO_2$  (Seinfeld, 1989). In this complex chemical reaction, the principal role of VOC is the oxidation of NO to  $NO_2$  (Finlayson-Pitts *et al.*, 1993).

Motor vehicles and industrial emissions are the major sources of anthropogenic precursors of  $O_3$ . Upstream oil and gas production is also a large contributor in Canada (Environment Canada, 2013a). The  $O_3$  attributed to natural sources is formed by biogenic emissions (emission of isoprene from vegetation) and from abiotic emissions (biomass burning, lightning and geogenic sources) (U.S. EPA, 2013). In urban areas, both anthropogenic and natural  $NO_x$  and VOC are important precursors of  $O_3$  formation, unlike in non-urban areas, where the biogenic VOC emitted from vegetation is the most important precursor of  $O_3$  formation (U.S. EPA, 2013). High  $O_3$  levels usually occur during the summer months when heat and sunlight are more intense. However, high levels of  $O_3$  can also be observed at mid-latitude sites in the late winter and spring seasons (Ahmadov *et al.*, 2015; Schnell *et al.*, 2009). During the winter months, high levels of  $O_3$  can be caused by long-range

transport (i.e., transport of air pollutants in the atmosphere for a distance greater than 100 kilometers) (Leduc *et al.*, 2005) and buildup of O<sub>3</sub> precursors (Diem, 2004), while during the spring months, they can be caused by strong generation of isoprene and terpenes from vegetation (Liudchik *et al.*, 2013). High O<sub>3</sub> concentrations also occur in non-industrial areas and can be found in areas with low population densities and several hundred kilometers downwind from urban industrial areas (Canada-United States Air Quality Committee, 1999).

O<sub>3</sub> is a main component of smog and a strong oxidizing agent (Geddes *et al.*, 2009) and has harmful effects on human health (Gálvez, 2007; U.S. EPA, 2013), forests and crops (Diem, 2003; Felzer *et al.*, 2007; Zeng *et al.*, 2008). Many epidemiological studies have shown associations between O<sub>3</sub> levels and non-accidental and cardiopulmonary-related mortality (Goldberg *et al.*, 2001; Jerrett *et al.*, 2009; U.S. EPA, 2013). Studies also reported that O<sub>3</sub> levels have been associated with hospital admissions, and emergency department visits due to respiratory problems such as chronic obstructive pulmonary diseases and asthma (Stieb *et al.*, 2009; Yang *et al.*, 2003). The populations particularly susceptible to the effects of O<sub>3</sub> are young children, the elderly and people with pre-existing lung and heart diseases (U.S. EPA, 2013).

O<sub>3</sub> concentrations exhibit spatial variations, including intra-urban variations, in such a way that populations are exposed to different levels of O<sub>3</sub> (Liu *et al.*, 1996) depending on their residential or work location. The spatial variation of O<sub>3</sub> is influenced by the topography, the meteorological conditions and the long-range transport. For example, high levels of O<sub>3</sub> in Eastern Canada is associated with anticyclonic conditions when high temperatures and west to

south-west winds predominate, which facilitates the long-range transport of O<sub>3</sub> and its precursors (McKendry, 1993) from heavily industrial regions localized along the shorelines of the Great Lakes (Canada-U.S. border) where higher levels of O<sub>3</sub> occur (Garcia *et al.*, 2010). The intra-urban variations of O<sub>3</sub> levels are also associated with the geographic variation of sources of O<sub>3</sub> precursors and sources of oxidizing compounds such as road traffic-related NO<sub>x</sub> (Liu *et al.*, 1996).

The assessment of the health effects to O<sub>3</sub> exposure is a challenge because it exhibits complex spatio-temporal patterns (Adam-Poupart *et al.*, 2014). To improve the exposure assessment of populations, different air pollution modeling techniques have been used. Ordinary kriging (one version of kriging) is the most common geostatistical approach used in air pollution modeling (Jerrett *et al.*, 2005). More complex spatio-temporal models such as Bayesian maximum entropy (BME) have also been used to estimate exposure over large areas (Adam-Poupart *et al.*, 2014).

In this study, we assessed a multivariable kriging approach known as Kriging with external drift (KED) to predict daily O<sub>3</sub> concentrations at the local scale in the metropolitan area of Montreal, Canada. These results were compared with those of Ordinary kriging (OK) and with an approach based on the simple mean of ozone levels at fixed-site monitoring stations. We also compared summer estimates to those obtained from a BME model (Adam-Poupart *et al.*, 2014).

### **3 Methods**

#### **Study area**

The study area situated between 45.27°–45.72°N and 73.98°–73.28°W is composed of the following nine census consolidated subdivisions (CCS) of the Montreal census metropolitan area (CMA) : Montreal, Laval, Longueuil, Brossard, La Prairie, Saint-Philippe, Saint-Mathieu, Saint-Constant and Kahnawake, with a total area of 1 310 km<sup>2</sup> and a population of 2.48 million people. The Montreal CCS includes the City of Montreal, which is the second largest city in Canada, and is located in the Saint Lawrence Valley at the eastern end of the Windsor-Quebec corridor (located along the north shore of the Great Lakes and Saint Lawrence River), the most industrialized region of Canada (McKendry, 1993) (Figure 1).

In this study, the climate is categorized as humid continental area, with severe winters and with hot and humid summers. It is subject to large continental climate variations. The wind direction is predominantly from the south-west, alternating with northeasterlies along the axis of the Saint Lawrence Valley.

Our study was conducted over a period of one year, from January to December 2010. The OK, the KED model and the mean estimates were implemented for each of six synoptic weather classes to account for spatial and temporal dependencies on meteorological variables. In order to compare the results of these models with those of a BME model (developed for the



summer months), we also produced estimates for the months of May through September, 2010.

### **O<sub>3</sub> concentration and meteorological data**

Hourly data of O<sub>3</sub> concentrations were obtained from the National Air Pollution Surveillance (NAPS) network of Environment Canada (Environment Canada, 2013b). In the study area during 2010, ground-level ozone was measured by 12 NAPS monitors (Figure 1). These measurements are performed using gas analyzers operating on ultraviolet light absorption principles (Environment Canada, 2013c). O<sub>3</sub> fixed-site monitoring stations provide a spatially sparse sample (i.e., stations per km<sup>2</sup>) of ozone data with a good continuity in time.

Usually, research studies focus on the ground-level ozone prediction for the summer season (May through September). However, because high O<sub>3</sub> concentrations can also occur during winter and spring seasons, in this study we modeled O<sub>3</sub> concentrations throughout the year. For this reason, we calculated midday 8-hr averages (09:00-17:00 hours) of O<sub>3</sub> levels for 2010. We used midday 8-hr averages because the largest values of the 24-hr O<sub>3</sub> concentrations period for a given day are normally in this interval and because it corresponds to the Canada-wide standard for ground-level ozone estimation (CCME, 2000).

Meteorological measurements were obtained from the National Climatic Data and Information Archive (DAI) of Environment Canada (Environment Canada, 2013d). There were 8 meteorological monitoring stations across the Montreal CMA for the 2008–2010 period (Figure 1). To downscale spatio-temporal meteorological surfaces (for KED, see below), we calculated midday 8-hr averages for temperature and relative humidity. The same interval as for the ozone measurements was used because these meteorological variables influence ozone chemistry and concentration. Also, we used daily average precipitation since only daily total information was available. Precipitations are responsible for the reduction of ozone concentrations by wet removal of soluble ozone precursors (Hou *et al.*, 2015). We used daily average wind speed as representative for long-range transport associated with the distance traveled by the air mass within the 24-h period (Camalier *et al.*, 2007). Hourly measures of the wind speed and directions were also used to create wind roses (Figure 2). Each wind rose shows the frequency and speed of wind blowing from each direction (hourly measures) in a particular distribution (i.e., a synoptic group) for 2010.

O<sub>3</sub> concentration and meteorological data were calculated for all available monitoring data having no more than two hours of missing values on any given day.

### **Synoptic weather analysis**

Weather patterns evolve in time (Hufty, 1982) influencing the frequency and intensity of air pollutant concentrations including ground-level ozone (Ebi *et al.*, 2008). To account for the spatial and the temporal dependencies of the ozone amount on meteorological conditions,

we used data from the 2008–2010 period. Clusters of days with similar meteorological patterns (i.e., synoptic groups) are identified. A period longer than the study period (i.e., 2010) was used to ensure a robust classification with a larger number of observation (i.e., days). To identify the synoptic groups, we applied the 24-h daily average on each meteorological variable. We proceeded in two stages. First, we used principal component analysis (PCA) to transform correlated meteorological variables (i.e., wind speed, temperature, relative humidity and precipitation) into a new set of uncorrelated variables. A total of 1 087 observations, with the exclusion of 9 unavailable observations were used in this computation. Second, we identified clusters that contain similar patterns. For this we used the hierarchical cluster analysis (HCA) method to classify the uncorrelated values expressed on principal components axes. HCA is a method based on a hierarchical divisive or agglomerative approach where each partition of the data set is associated to levels of a tree diagram (dendrogram). The Ward's method based on the error sum of the squares criterion (Murtagh *et al.*, 2014) was used to define the distance between pairs of clusters within the classification.

### **Ordinary kriging (OK)**

Ordinary kriging (OK) is a form of univariate generalized least square (GLS) regression that assumes that the mean of observations is constant in local neighborhoods, but unknown. A weighted average based on the neighboring data is used to estimate unknown-values at a given location. We created daily maps of O<sub>3</sub> for each day of the year 2010 using OK at a resolution of 100 x 100 meters. The values of the grid cells at the locations of the 12

monitoring stations on each day of the study period were retrieved for cross-validation (see below).

### **Kriging with external drift (KED)**

A multivariable form of kriging, known as kriging with external drift (KED) (Goovaerts, 1997; Wackernagel, 2003) was applied with the following potential external (independent) variables: spatio-temporal daily meteorological estimates (meteorological surfaces), land use, population density, road density and vegetation density. These variables were chosen as they can directly and indirectly affect the formation, destruction and dispersion of ground-level ozone (Matthes *et al.*, 2007; Nowak *et al.*, 2013; Özbay *et al.*, 2011). All these external variables were represented in raster format (uniform grids) with cells of 100 x 100 meters using ArcGIS v.10.2 (ESRI). As with OK, we created daily maps of O<sub>3</sub> for each day of the year 2010 using KED at a resolution of 100 x 100 meters. The values of the grid cells at the locations of the 12 monitoring stations on each day of the study period were retrieved for cross-validation (see below).

### ***Spatio-temporal meteorological variables***

Meteorological variables introduced in section 2.2 were analyzed as potential covariates for the KED interpolation. The inverse distance weighted (IDW) interpolation method was used to estimate spatio-temporal meteorological surfaces of wind speed, temperature, relative humidity and precipitation for all days of the study period (2010). This

operation was carried out using the gstat and automap packages in R, version 3.1.1 (R Project for Statistical Computing; <http://www.r-project.org>).

### ***Geographic variables***

Weighted road density, land use (urban, industrial and commercial area), population density and the greenness component of the tasseled-cap image transformation were analyzed as potential covariates for KED interpolation. All covariate surfaces were computed within circular buffers of various radii to reflect local influence of emission sources (Zhang *et al.*, 2015). Different radii of influences were analyzed to identify covariates to be used in the interpolation.

### **Road density**

Involved in the production and destruction of O<sub>3</sub>, NO<sub>x</sub> are oxidized forms of nitrogen that are mainly composed of nitrogen monoxide (NO) and nitrogen dioxide (NO<sub>2</sub>). In urban areas, the principal source of NO and NO<sub>2</sub> emission is road traffic (Brook *et al.*, 2007). NO<sub>x</sub> concentrations decay with increasing distance from the emission source, producing an inverse pattern by the ozone-scavenging reaction (Beckerman *et al.*, 2008). To account for the association between O<sub>3</sub> and NO<sub>x</sub> emissions from vehicular traffic, we used the weighted road density approach (WDR) developed by Rose *et al.* (2009). We computed the logarithm of WDR because a graphical analysis revealed a log-linear relationship with O<sub>3</sub> concentration.

To compute WRD, the road network data used was the 2010 Road Layer Dataset (CanMap® RouteLogistics version 2010.3) compiled by Digital Mapping Technology Inc. (DMTI). Firstly, the classes highway (expressway, principal and secondary highway), major road and local road were extracted from the complete road dataset and assigned respective weight values of 3 for highway, 2 for major road and 1 for local road to represent relative contributions of the roads to local NO<sub>x</sub> concentrations. We used these weights based on the hierarchical levels of the road system assuming a similar relation with NO<sub>x</sub> as the density based on traffic counts, as suggested by previous studies (Hansell *et al.*, 2014; Rose *et al.*, 2009). Secondly, we used WRD to calculate the density of roads within radii of 100, 150, 250 and 1 000 meters for each cell size (100 x 100 m) of the output raster. These radii are below the threshold of the maximum distance (1 500 m) of influence on NO<sub>x</sub> levels related to the traffic source, reported by Su *et al.* (2009).

$$WRD = \sum_{i=1}^n L_i V_i / \pi r^2 \quad (1)$$

where  $L_i$  is the length of roads within the buffer,  $V_i$  the arbitrary weights applied to each road class, and  $r$  the radial size of the buffer.

#### Land use data:

Other primary sources of NO<sub>x</sub> and VOC besides road traffic are industrial, commercial and residential sources that burn fuels. Thus, the spatial variation of O<sub>3</sub> concentration can also

be linked to land use and population density (Beelen *et al.*, 2009). Three land use types (residential, commercial and industrial) were extracted from DMTI spatial data for the year 2010. The original land use polygons were converted to a raster grid in order to aggregate area of each land use class within search radii of 300, 500 and 750 meters. These buffer sizes have been used in previous studies to represent associations between land use variables and precursors of ground-level O<sub>3</sub> (Brauer *et al.*, 2008; Henderson *et al.*, 2007). The neighborhood statistics (a function of the ArcGIS Spatial Analyst extension) was used to compute the sum of the area of each land use type within circular buffers. Three layers were obtained (one for each land use type) for each radius size.

#### Population density:

Population density is associated with traffic-related NO<sub>x</sub> and VOC emissions (Henderson *et al.*, 2007). Total population per dissemination area from the 2006 census from Statistics Canada was used. A dissemination area is a small census unit with a population of 400 to 700 persons (Statistics Canada, 2009). The polygons of dissemination areas were converted to centroids and were assigned with the corresponding population to each polygon. We used kernel density to estimate the values within radii of 1 000 and 2 500 meters. In previous studies, these sizes showed correlations between population density and precursors of ground-level O<sub>3</sub> concentration (Gilbert *et al.*, 2005; Henderson *et al.*, 2007).

#### Vegetation density:

Vegetation is inversely associated with NO<sub>x</sub> emission sources (Su *et al.*, 2009). NO<sub>x</sub> concentrations from traffic emission sources decrease with distance to create an inverse pattern of scavenging reactions with O<sub>3</sub> (Beckerman *et al.*, 2008). This may produce a positive relationship between O<sub>3</sub> concentration and vegetation density. Green areas help in reducing the levels of some air pollutants from dry deposition through processes such as sedimentation (large particles) and turbulent diffusion (gases and small particles). However, biogenic VOC emissions from natural sources, such as plants and trees, can contribute to the formation of O<sub>3</sub> (Pun *et al.*, 2002; Wang *et al.*, 2008).

To model vegetation density, we extracted the greenness component from a tasseled-cap transformation of a Landsat Thematic Mapper (TM) image. This image (path/row 014/028) captured on July 27, 2010 over the Montreal CMA region was orthorectified and provided by the U.S. Geological Survey (USGS). The tasseled-cap in remote sensing is an orthogonal transformation of reflective bands which isolates greenness, wetness and soil brightness (Kauth *et al.*, 1976). The greenness component exhibits high values for surfaces with high vegetation density and low values for soil surfaces. For each grid of 100 meters, the mean of the greenness component (original resolution 30 meters) was calculated within radii of 100, 200, 300 and 500 meters. These buffer sizes have shown highest correlations between vegetation greenness and O<sub>3</sub> precursor's concentration in past studies (Su *et al.*, 2009; Su *et al.*, 2010).

***KED - geostatistical analyses***



KED—which is based on some form of spatial dependence—was applied to map the spatial and temporal variation of ground-level O<sub>3</sub>. For this we used three principal packages of R software version 3.1.1 (R Project for Statistical Computing; <http://www.r-project.org>): *gstat*—spatial and spatio-temporal geostatistical modeling, prediction and simulation; *automap*—automatic interpolation package and; *spacetime*—classes and methods for spatio-temporal data.

The spatial dependence effects include two broad categories; 1) first-order effects, which measure broad trends in the data across the study area, and 2) second-order effects, which measure local variations as a function of the distance between sample locations (Jerrett *et al.*, 2003). KED takes into account both aspects of spatial dependence (Pearce *et al.*, 2009), and replaces the local constant-mean of OK by the trend (drift) under the assumption of a linear relationship between dependent and independent variables (Goovaerts, 1997; Wackernagel, 2003). Thus, the independent variable  $Y_R(u)$  is considered as a drift that may follow the dependent variable  $Z(u)$  (O<sub>3</sub> concentrations) and can be expressed as:

$$Z(u) = c_1 + c_2 Y(u) + Y_R(u) \quad (2)$$

where  $c_1$  and  $c_2$  are constants,  $Y(u)$  is known everywhere and represents the value of the independent variable at location  $u$ , and  $Y_R(u)$  is the residual.

Since the geostatistic process is based on spatial dependence revealed by the measure of the degree of dissimilarity between observations, a semivariogram was defined for each

time slice. The experimental semivariogram  $\hat{\gamma}(h)$  of a random function  $Z(u)$  (a set of *dependent random variable*) quantifies the variability between data pairs of points at various distances and is defined as:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(u_{\alpha}) - Z(u_{\alpha} + h)]^2 \quad (3)$$

where  $N(h)$  is the number of pairs of sampling locations at distance  $h$  from one another, and  $Z$  are the observed values at sampling locations  $u_{\alpha}$  and  $u_{\alpha} + 1$ .

The selection of independent variables was based on the graphical analysis and on the level of Pearson's correlation coefficient with  $O_3$  concentrations. Since graphical analysis revealed log-linear relationship between road density (WDR - weighted road density) and  $O_3$  concentrations, we computed the logarithm of WDR. Daily correlation coefficients were calculated among samples collected at each monitoring station. They were then summarized per synoptic weather classes, or summer months, using the median of correlations (Table 2). Because of the limited number of monitoring stations available ( $n = 12$  for 2010) for developing the KED model, we restricted the selection of variables to only two to minimize the risk of overfitting.

The concentration of  $O_3$  at a given point of interest  $u$  is estimated according to:

$$Z_{KED}^*(u) = \sum_{\alpha=1}^n \lambda_{\alpha}(u) Z(u_{\alpha}) \quad (4)$$

where,  $\lambda_{\alpha}(u)$  are the weights corresponding to the  $n$  samples at location  $u$  and obtained by solving a system of  $(n(u) + 2)$  linear equations with two Lagrange multipliers (Goovaerts, 1997) and determined by the experimental variogram. We verified that fundamental assumptions of kriging systems were met.

In the case of OK and KED, temporal interactions were not incorporated, and O<sub>3</sub> predictions were generated for time slices using daily estimated variograms.

### **Bayesian Maximum Entropy (BME)**

BME is a nonlinear model for space-time data based on physical measures (“hard” data) and additional information such as the multivariate regression estimates (“soft” data) (Adam-Poupart *et al.*, 2014). The O<sub>3</sub> levels were estimated at the 12 monitoring stations from the BME model developed by Adam-Poupart *et al.* (2014) for the summer months (May through September).

### **Mean of daily O<sub>3</sub> concentrations**

The daily mean of O<sub>3</sub> concentrations at the 12 monitoring stations were calculated, using 11 stations for estimation at each location. These means were computed for all days of the study period.

## Model evaluation

The results of the KED, OK, mean and BME (Adam-Poupart *et al.*, 2014) models were evaluated using cross-validation, also referred to as “leave-one-out”. Using this approach, each station’s observation on a given day was removed from the data one at a time and the remaining data was used to predict the O<sub>3</sub> concentration at the location of the excluded station. The observations and predictions at the excluded stations were then compared. In order to compare interpolation models to the daily mean of O<sub>3</sub> concentrations, cross-validation was also applied to compute daily means. In other words, the performance of this approach was evaluated by calculating the difference between the O<sub>3</sub> concentrations at a given station on a given day and the mean of the remaining ones.

O<sub>3</sub> estimation errors were defined as predicted values minus observed values, and computed across interpolation methods for each station-day. We used the daily root mean-square errors (RMSE) based on cross-validation to estimate and compare the magnitude of errors of the three approaches. To measure each model’s performance, we also computed the coefficient of determination ( $R^2$ ) between measured and estimated O<sub>3</sub> values.

Finally, three maps were created, illustrating the predicted mean daily O<sub>3</sub> levels (09:00 – 17:00), the prediction standard error (SE) (square root of the kriging variance), and the relative prediction error (SE divided by the KED prediction \* 100) from KED at a 100-m grid across the Montreal area for the year 2010.

## 4 Results

In the analysis designed to identify synoptic weather classes, four factors with eigenvalues above 1 were obtained from the principal component analysis of meteorological variables. The three primary factors explained 87% of the total variance and were retained as input data for hierarchical clustering. The hierarchical classification produced a binary tree (dendrogram) with six distinct synoptic weather classes. An interactive 3D scatter plot of the three PCA factors also visually revealed a distinct separation between these classes, confirming the clustering results. The days forming synoptic class 1 were spread over the year; the days of classes 2 and 3 were spread over the winter season; the days of classes 4 and 6 were spread over the summer and fall seasons; and the days of class 5 were found mostly during the spring season.

Table 1 shows the descriptive statistics of O<sub>3</sub> concentrations, precipitation, temperature, relative humidity and wind speed for the days classified in each of the six synoptic classes. The results show that highest levels of O<sub>3</sub> concentrations occurred during days of the spring season (synoptic class 5) with low relative humidity and gentle breeze. High levels of O<sub>3</sub> also occurred during summer and fall days (synoptic class 6) with high temperatures and predominant southwesterly winds (Figure 2). The results also revealed high levels of O<sub>3</sub> during the winter season (synoptic class 2) when the winds were predominantly from the west and south-west directions (Figure 2). Finally, low levels of O<sub>3</sub> occurred in weather conditions with high relative humidity and high precipitation (synoptic class 1).

Table 2 shows median values of daily Pearson's correlation coefficients between O<sub>3</sub> and independent variables: for synoptic classes, total N = 355 days (10 days with more than two stations with missing values on any given day were excluded of the model) and summer months, total N = 153 days were used. This table shows that the daily mean of ground-level O<sub>3</sub> calculated for each monitoring station was highly correlated with vegetation density within 100 meters, either by synoptic class (median correlations between 0.74 and 0.81) or for the summer months period (median correlations of 0.75). It is important to note, however, that the positive correlation with vegetation density for synoptic classes 2 and 3 (within radius of 100 meters) corresponds to the winter months when there are no leaves on the deciduous trees, which form the majority of city trees. Also, the O<sub>3</sub> and road density (logarithm of WDR) within 250 meters were negatively correlated, both for each synoptic class (median correlations between -0.57 and -0.68) and during the summer months (median correlation of -0.64). In contrast, median correlations with other variables such as land use (industrial, commercial and residential), population density and all meteorological variables were lower. Associations were observed with wind speed (positive correlation) in particular for synoptic class 2, when winds were the strongest. In other words, the levels of O<sub>3</sub> increased with the prevailing wind, especially from the west and south-west directions (Johnson *et al.*, 2007; McKendry, 1993). In addition, a low positive correlation with temperature was observed for the synoptic classes 1, 4, 5 and 6 corresponding to warmer days and a low negative correlation with the synoptic classes 2 and 3 corresponding to the colder days. Also, a negative correlation with precipitation was observed for the synoptic class 1 (days spread over the year with precipitations and high relative humidity).

Table 3 describes the cross-validation results. First, comparisons are made between KED, OK, BME and the mean for the summer months of the year 2010 at the 12 stations; second, comparisons are made between KED, OK and the mean by synoptic classes for all days of the year 2010 at the 12 stations. For the summer months (May through September), the KED model performed better ( $R^2 = 0.92$ , RMSE = 3.14 ppb) than the OK ( $R^2 = 0.85$ ; RMSE = 4 ppb), the mean ( $R^2 = 0.83$ ; RMSE = 4.58 ppb) and the BME model ( $R^2 = 0.62$ ; RMSE = 6.37 ppb). In each synoptic class, the proportion of explained variance was always the lowest using the mean ( $R^2 = 0.52$  to 0.85 and RMSE = 4.27 to 5.12 ppb), the OK model was always second best ( $R^2 = 0.55$  to 0.88 and RMSE = 3.71 to 4.21 ppb), while the KED model always performed best ( $R^2 = 0.77$  to 0.92 and RMSE = 2.79 to 3.37 ppb).

The distribution of daily  $O_3$  errors (i.e., the difference between estimated and observed values) for the summer months based on the cross validation showed that the KED and BME models overestimated the mean  $O_3$  levels by 1% and 2% respectively (mean errors respectively of 0.29 ppb and 0.63 ppb), while the OK model underestimated them by 0.5% (mean error of -0.13 ppb) (Figure 3). Figure 4 shows the temporal prediction RMSEs of the KED, OK, mean and BME models for the summer months. For the four individual models, the errors were quite variable, oscillating mostly between 1 and 5 ppb for the KED, between 2 and 6 ppb for the OK and for the simple mean, and between 2 and 7 ppb for the BME—with a few large spikes in early June, July and September for all cases.

Figure 5 shows that high levels of  $O_3$  can be observed in areas that are farther away from the center of the island, in other words, areas that are generally suburban or rural (Figure

5A). Lowest concentrations of O<sub>3</sub> are observed in downtown Montreal (the center of the island); especially close to high traffic areas which could be explained for the NO<sub>x</sub> (road traffic-related emission) scavenging process. A greater difference between observed and estimated O<sub>3</sub> concentrations may be found in the northeast of the island, which is an industrial area, and in the southwest of the island (Figure 5B and 5C).

## 5 Discussion and conclusions

The goal of this study was to compare various methods to create prediction surfaces of ground-level ozone at a local resolution (e.g. 100-m grid). Our results demonstrate that the use of KED with covariates of vegetation density and road density (logarithm of WDR) can explain a large part of the variability in the spatial distribution of O<sub>3</sub> levels. The road density was negatively associated with O<sub>3</sub> levels caused by ozone destruction due to the NO<sub>x</sub>-scavenging process since road traffic is the principal source of NO<sub>x</sub> emissions. KED outperformed other methods, such as BME, OK or the simple mean method for determining ozone levels at monitoring stations across the region.

The time-point by time-point (Le *et al.*, 2006) KED provided a more accurate prediction than ordinary kriging because the inclusion of highly correlated covariates in the interpolation process improved the estimates. KED was also more accurate than the BME model, although we should stress that these two approaches are not entirely comparable. While the KED model predicted O<sub>3</sub> levels separately on each given day with new and specific daily information (one model per day), the BME model predicted any given day with a unique



model (based in all data). This may partly explain why the KED model outperformed the BME model. Furthermore, a disadvantage with KED is the limited number of fixed-site monitoring stations (12 station-measures per day) which may have produced instability in the variogram estimation for this model in a given day.

We also found that the decomposition of the temporal distribution of meteorological factors in synoptic classes using principal components and hierarchical cluster analyses demonstrate an effective classification showing an important relationship between ground-level O<sub>3</sub> and meteorological conditions. This allowed us to describe how O<sub>3</sub> concentrations are linked to daily meteorological conditions as well as observe differences in modeling performance of O<sub>3</sub> and variations in impact of environmental context on O<sub>3</sub> concentration. It also enabled us to develop KED models adapted to the conditions of each synoptic period. As observed in Table 1, high levels of O<sub>3</sub> can also occur in the winter season (synoptic class 2). This may be explained by the winter buildup of O<sub>3</sub> precursors and predominant winds and pollution advection from the west and the south-west directions (Figure 2) suggesting a long-range transport regime. This may also be explained by the temperature inversion, which is relatively frequent in the winter season, which traps pollution within the stagnant boundary layer (Pudasainee *et al.*, 2006).

High levels of O<sub>3</sub> during the summer and fall seasons (synoptic class 6) may be explained by the higher temperatures, solar radiation conducive to photochemical reactions, local O<sub>3</sub> production and predominant southwesterly winds (Figure 2), suggesting a long-range transport along the Windsor-Quebec corridor of the Saint Lawrence Valley (Figure 1). This is

consistent with that expressed by McKendry (1993) which indicated that high concentrations of O<sub>3</sub> in the Montreal region are influenced by the long-range transport of O<sub>3</sub> when the wind blows predominantly from the west and south-west from the Great Lakes (heavily industrial region at the Canada-U.S. border) along the St. Lawrence River Valley. The mean O<sub>3</sub> lifetime is long enough (30–35 days) (U.S. EPA, 2013) to allow transporting O<sub>3</sub> over great distances (typically 1 000–2 000 km) and its precursors over several thousands of kilometers (McKendry, 1993).

The results obtained with KED are consistent with those of Beelen *et al.* (2009), who estimated O<sub>3</sub> levels across the European Union applying Universal kriging (UK, or KED) on high-density residential areas, major roads, altitude, distance to sea, and a combination of meteorological covariates. They report an R<sup>2</sup> of 0.70 and RMSE of 7.69 µg/m<sup>3</sup> (~ 4 ppb). However, differences in the performance between two studies could be due to the difference in coverage: Beelen *et al.* (2009) with regional coverage (224 stations for 2 854 116 km<sup>2</sup>) versus our local coverage (12 stations for 1 310 km<sup>2</sup>).

The results obtained with OK can be compared to those of others. For example, Liu and Rossini, (1996) estimated O<sub>3</sub> levels with OK based on outdoor measurements at 40 home sites within in the Toronto metropolitan area (7 124 km<sup>2</sup>), Canada. They did not report the RMSE, but R<sup>2</sup> for daily estimates during the summer of 1992 were similar to ours (Liu and Rossini: R<sup>2</sup> = 0.82 for 12-h daytime; present study: R<sup>2</sup> = 0.85 for 8-h daytime). García *et al.* (2010) also estimated O<sub>3</sub> levels with OK based on 138 sampling locations (automatic portable analyzer) in the city of Badajoz, Spain. They did not report R<sup>2</sup> values, but RMSEs for each of

the summer months of 2007 (one estimation per month) were similar (RMSE = 2.14 ppb to RMSE = 4.24 ppb) to ours (RMSE = 4.00 ppb).

The maps of O<sub>3</sub> concentrations (Figure 5) illustrate the results of the scavenging effect (titration) produced by the reaction with NO (nitrogen monoxide) emissions (U.S. EPA, 2006). Unlike the suburban and rural areas and due to NO<sub>x</sub> scavenging, lowest concentrations of O<sub>3</sub> are observed in the high-density center of the city and near major highways, and were slightly higher in the eastern part of the Island of Montreal.

In this study, we demonstrated the possibility of developing high resolution maps with a limited number of O<sub>3</sub> monitoring stations under specific environmental and meteorological conditions using KED. The integration of covariates in KED aimed at capturing the intra-urban variation of O<sub>3</sub> concentrations; however, such a model could perform better if there were more monitoring stations across the region of study and more covariates used. Furthermore, the inclusion of more covariates could be improved by the reduction of the dimensionality (e.g., principal component analysis) while preserving variability.

Finally, the KED model adds value to the O<sub>3</sub> daily estimation because it generates daily maps over time and space with small errors across the Montreal region. This may increase our capacity of linking O<sub>3</sub> levels to health problems by means of improved assessments of personal exposure. Nonetheless, future work on integrating the temporal dependency in the data is needed to avoid overstating the performance of KED.

## **6 Acknowledgements**

The authors would like to acknowledge Allan Brand for providing a compilation of meteorological and ground-level O<sub>3</sub> data and estimates from the BME model for this study. We also thank Éric de Montigny for comments on an earlier version of this paper. This project was financially supported by the Quebec Institute of Public Health.

## 7 References

- Adam-Poupart, A., Brand, A., Fournier, M., Jerrett, M., and Smargiassi, A. (2014). Spatiotemporal modeling of ozone levels in Quebec (Canada): A comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy-LUR approaches. *Environmental Health Perspectives*, 122(9), 970–976. doi: 10.1289/ehp.1306566
- Ahmadov, R., McKeen, S., Trainer, M., Banta, R., Brewer, A., Brown, S., Edwards, P. M., de Gouw, J. A., Frost, G. J., Gilman, J., Helmig, D., Johnson, B., Karion, A., Koss, A., Langford, A., Lerner, B., Olson, J., Oltmans, S., Peischl, J., Pétron, G., Pichugina, Y., Roberts, J. M., Ryerson, T., Schnell, R., Senff, C., Sweeney, C., Thompson, C., Veres, P. R., Warneke, C., Wild, R., Williams, E. J., Yuan, B., and Zamora, R. (2015). Understanding high wintertime ozone pollution events in an oil- and natural gas-producing region of the western US. *Atmospheric Chemistry and Physics*, 15, 411–429.
- Beckerman, B., Jerrett, M., Brook, J. R., Verma, D. K., Arain, M. A., and Murray, M. F. (2008). Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. *Atmospheric Environment*, 42, 275–290.
- Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., De Hoogh, K., and Briggs, D. J. (2009). Mapping of background air pollution at a fine spatial scale across the European Union. *Science of the Total Environment*, 407, 1852–1867.

- Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., and Karr, C. (2008). A cohort study of traffic-related air pollution impacts on birth outcomes. *Environmental Health Perspectives*, 116(5).
- Brook, J. F., Burnett, R. T., Dann, T. F., Cakmak, S., M.S., G., Fan, X. F., and Wheeler, A. J. (2007). Further interpretation of the acute effect of nitrogen dioxide observed in Canadian time-series studies. *Journal of Exposure Science and Environmental Epidemiology*, 17, S36–S44.
- Camalier, L., Cox, W., and Dolwick, A. (2007). The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmospheric Environment*, 41, 7127–7137.
- Canada-United States Air Quality Committee. (1999). *Ground-Level Ozone: Occurrence and Transport in Eastern North America* (pp. 36).
- CCME. (2000). *Standards pancanadiens relatifs aux particules (PM) et à l’ozone*. Conseil Canadien des Ministres de l’Environnement.
- Diem, J. E. (2003). A critical examination of ozone mapping from a spatial-scale perspective. *Environmental Pollution*, 125, 369–383.
- Diem, J. E. (2004). Explanations for the spring peak in ground-level ozone in the southwestern United States. *Physical Geography*, 25(2), 105–129.

- Ebi, K. L., and McGregor, G. (2008). Climate change, tropospheric ozone and particulate matter, and health impacts. *Environmental Health Perspectives*, 116(11), 1449–1455.
- Environment Canada. (2013a). Ground-Level Ozone. Retrieved April, 2015, 2015, from <https://www.ec.gc.ca/air/default.asp?lang=En&n=590611CA-1>
- Environment Canada. (2013b). National Air Pollution Surveillance Network (NAPS) Data Products. Retrieved 10 March 2013, from <http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx>
- Environment Canada. (2013c). 10 years of Data from the National Air Pollution Surveillance (NAPS) Network, Data Summary from 1999 to 2008. Retrieved April, 2015, from <http://ec.gc.ca/Publications/default.asp?lang=En&xml=B14B2399-8661-4A71-90B3-6D0ACB521AA4>
- Environment Canada. (2013d). Canada National Climate Data and Information Archive. Retrieved 5 March 2013, from [http://climate.weather.gc.ca/index\\_e.html](http://climate.weather.gc.ca/index_e.html)
- Felzer, B. S., Cronin, T., Reilly, J. M., Melillo, J. M., and Wang, X. (2007). Impacts of ozone on trees and crops. *C.R. Geoscience*, 339, 784–798.
- Finlayson-Pitts, B. J., and Pitts, J. J. N. (1993). Atmospheric chemistry of tropospheric ozone formation: scientific and regulatory implications. *Air & Waste Management Association*, 43(8), 1091-1100.

- Gálvez, O. (2007). Synoptic-scale transport of ozone into Southern Ontario. *Atmospheric Environment*, 41, 8579–8595.
- Garcia, V. C., Foley, K. M., Gego, E., Holland, D. M., and Rao, S. T. (2010). A comparison of statistical techniques for combining modeled and observed concentrations to create high-resolution ozone air quality surfaces. *Journal of the Air & Waste Management Association*, 60, 586–595.
- Geddes, J. A., Murphy, J. G., and Wang, D. K. (2009). Long-term changes in nitrogen oxides and volatile organic compounds in Toronto and the challenges facing local ozone control. *Atmospheric Environment*, 43, 3407–3415.
- Gibson, M. D., Guernsey, J. R., Beauchamp, S., Waugh, D., Heal, M. R., Brook, J. R., Maher, R., Gagnon, G. A., McPherson, J. P., Bryden, B., Gould, R., and Terashima, M. (2009). Quantifying the spatial and temporal variation of ground-level ozone in the rural Annapolis Valley, Nova Scotia, Canada using nitrite-impregnated passive samplers. *Journal of the Air & Waste Management Association*, 59, 310–320.
- Gilbert, N. L., Goldberg, M. S., Beckerman, B., Brook, J. R., and Jerrett, M. (2005). Assessing spatial variability of ambient nitrogen dioxide in Montréal, Canada, with a land-use regression model. *Journal of the Air & Waste Management Association*, 55, 1059–1063.



- Goldberg, M. S., Burnett, R. T., Brook, J., Bailar III, J. C., Valois, M.-F., and Vincent, R. (2001). Associations between daily cause-specific mortality and concentrations of ground-level ozone in Montreal, Quebec. *Am J Epidemiol*, 154(9), 817–826.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.
- Hansell, A., Rose, N., Cowie, C. T., Belousova, E. G., Bakolis, I., Ng, K., Toelle, B. G., and Marks, G. B. (2014). Weighted road density and allergic disease in children at high risk of developing asthma. *PLoS One*, 9(6), 9.
- Henderson, S. B., Beckerman, B., Jerrett, M., and Brauer, M. (2007). Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science and Technology*, 41, 2422-2428.
- Hou, X., Zhu, B., Fei, D., and Wang, D. (2015). The impacts of summer monsoons on the ozone budget of the atmospheric boundary layer of the Asia-Pacific region. *Science of the Total Environment*, 502, 641–649.
- Hufty, A. (1982). Analyse en composantes principales des situations synoptiques au Québec. *Géographie physique et Quaternaire*, 36(3), 307–314. doi: 10.7202/032483ar
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., and Giovis, C. (2005). A review and evaluation of intraurban air pollution

- exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, 15(2), 185–204.
- Jerrett, M., Burnett, R., Goldberg, M., Sears, M., Krewski, D., Catalan, R., Kanaroglou, P., Giovis, C., and Finkelstein, N. (2003). Spatial analysis for environmental health research: concepts, methods, and examples. *Journal of Toxicology and Environmental Health, Part A*, 66, 1783–1810.
- Jerrett, M., Burnett, R. T., Pope III, C. A., Ito, K., Thurston, G., Krewski, D., Shi, Y., Calle, E., and Thun, M. (2009). Long-term ozone exposure and mortality. *The New England Journal of Medicine*, 360(11), 1085-1095.
- Johnson, D., Mignacca, D., Herod, D., Jutzi, D., and Miller, H. (2007). Characterization and identification of trends in average ambient ozone and fine particulate matter levels through trajectory cluster analysis in Eastern Canada. *Journal of the Air & Waste Management Association*, 57, 907–918.
- Kauth, R. G., and Thomas, G. S. (1976). The Tasseled Cap - A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. Paper presented at the Symposium on Machine Processing of Remote Sensed Data, West Lafayette, Indiana.
- Le, N. D., and Zidek, J. V. (2006). *Statistical Analysis of Environmental Space-Time Processes*. New York: Springer.

- Liu, L. J. S., and Rossini, A. J. (1996). Use of kriging models to predict 12-hour mean ozone concentrations in metropolitan Toronto - A pilot study. *Environment International*, 22(6), 677–692.
- Liudchik, A., Pakatashkin, V., Umreika, S., and Girgzdiene, R. (2013). Role of ozone deposition in the occurrence of the spring maximum. *Atmosphere-Ocean*, 53(1), 42–49.
- Matthes, S., Grewe, V., Sausen, R., and Roelofs, G.-J. (2007). Global impact of road traffic emissions on tropospheric ozone. *Atmospheric Chemistry and Physics*, 7, 1707–1718.
- McKendry, I. G. (1993). Ground-level ozone in Montreal, Canada. *Atmospheric Environment*, 27B(1), 93–103.
- Murtagh, F., and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, 31, 274–295.
- Nowak, D. J., Hirabayashi, S., Bodine, A., and Hoehn, R. (2013). Modeled PM<sub>2.5</sub> removal by trees in ten U.S. cities and associated health effects. *Environmental Pollution*, 178, 395–402. doi: 10.1016/j.envpol.2013.03.050
- Özbay, B., Keskin, G. A., Dođruparmak, S. C., and Ayberk, S. (2011). Multivariate methods for ground-level ozone modeling. *Atmospheric Research*, 102, 57–65.

- Pearce, J. L., Rathbun, S. L., Aguilar-Villalobos, M., and Naeher, L. P. (2009). Characterizing the spatiotemporal variability of PM<sub>2.5</sub> in Cusco, Peru using kriging with external drift. *Atmospheric Environment*, 43, 2060–2069.
- Pudasainee, D., Sapkota, B., Shrestha, M. L., Kaga, A., Kondo, A., and Inoue, Y. (2006). Ground level ozone concentrations and its association with NO<sub>x</sub> and meteorological parameters in Kathmandu valley, Nepal. *Atmospheric Environment*, 40, 8081–8087.
- Pun, B., Wu, S., and Seigneur, C. (2002). Contribution of biogenic emissions to the formation of ozone and particulate matter in the Eastern United States. *Environmental Science and Technology*, 36, 3586–3596.
- Rose, N., Cowie, C., Gillett, R., and Marks, G. (2009). Weighted road density: A simple way of assigning traffic-related air pollution exposure. *Atmospheric Environment*, 43, 5009–5014.
- Schnell, R. C., Oltmans, S. J., Neely, R. R., Endres, M. S., Molenaar, J. V., and White, A. B. (2009). Rapid photochemical production of ozone at high concentrations in a rural site during winter. *Nature geoscience*, 2, 120–122.
- Seinfeld, J. H. (1989). Urban air pollution: State of the science. *Science*, 243, 745–752.
- Statistics Canada. (2009). Dissemination Area (DA). Retrieved 10 April 2012, from <http://www12.statcan.ca/census-recensement/2006/ref/dict/geo021-eng.cfm>

- Stieb, D. M., Szyszkowicz, M., Rowe, B. H., and Leech, J. A. (2009). Air pollution and emergency department visits for cardiac and respiratory conditions: A multi-city time-series analysis. *Environmental Health*, 8(25), 1–13.
- Su, J. G., Jerrett, M., and Beckerman, B. (2009). A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *Science of the Total Environment*, 407(12), 3890–3898. doi: 10.1016/j.scitotenv.2009.01.061
- Su, J. G., Jerrett, M., Beckerman, B., Verma, D., Arain, M. A., Kanaroglou, P., Stieb, D., Finkelstein, M., and Brook, J. F. (2010). A land use regression model for predicting ambient volatile organic compound concentrations in Toronto, Canada. *Atmospheric Environment*, 44, 3529–3537.
- U.S. EPA. (2006). Air quality criteria for ozone and related photochemical oxidants (2006 Final). U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-05/004aF.
- U.S. EPA. (2013). Final Report: Integrated Science Assessment for Ozone and Related Photochemical Oxidants. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-10/076F, 2013.
- Wackernagel, H. (2003). *Multivariate Geostatistics : An Introduction with Applications*: Springer Berlin Heidelberg.

- Wang, Q., Han, Z., Wang, T., and Zhang, R. (2008). Impacts of biogenic emissions of VOC and NO<sub>x</sub> on tropospheric ozone during summertime in eastern China. *Science of the Total Environment*, 395, 41–49.
- Yang, Q., Chen, Y., Shi, Y., Burnett, R. T., McGrail, K. M., and Krewskif, D. (2003). Association between ozone and respiratory admissions among children and the elderly in Vancouver, Canada. *Inhalation Toxicology*, 15, 1297–1308.
- Zeng, G., Pyle, J. A., and Young, P. J. (2008). Impact of climate change on tropospheric ozone and its global budgets. *Atmospheric Chemistry and Physics*, 8, 369–387.
- Zhang, J. J. Y., Sun, L., Barrett, O., Bertazzon, S., Underwood, F. E., and Johnson, M. (2015). Development of land-use regression models for metals associated with airborne particulate matter in a North American city. *Atmospheric Environment*, 106, 165–177.

## 8 Tables and figures

Table 1. Descriptive statistics of the six synoptic weather classes for O<sub>3</sub> and meteorological variables for 2008–2010.

Synoptic weather class	N (days) <sup>a</sup>	O <sub>3</sub> (ppb) <sup>b</sup>		Precipitation (mm)		Temperature (°C)		Relative humidity (%)		Wind speed (km/h)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	133	18.86	9.86	15.29	10.05	9.03	9.66	87.10	5.50	12.29	3.51
2	172	26.77	7.31	2.17	3.54	-2.70	7.15	69.24	11.45	18.60	3.77
3	279	20.46	8.00	0.83	1.52	-3.48	7.52	73.67	10.14	9.94	2.87
4	222	23.25	8.74	1.74	2.61	15.50	5.69	75.34	8.91	7.53	1.72
5	100	37.06	6.52	0.34	1.09	12.62	4.25	47.87	6.37	11.87	3.69
6	181	28.78	10.30	0.89	1.62	18.33	4.73	68.93	6.30	12.59	2.78
Total days <sup>c</sup> :	1,087	24.74	10.02	2.96	6.19	7.16	11.34	71.79	12.82	11.72	4.59

<sup>a</sup> Daily values were computed from the daily mean from 12 O<sub>3</sub> stations and 8 meteorological stations over the study area.

<sup>b</sup> O<sub>3</sub> mean values were computed from daily midday 8-hr (09:00-17:00 hours).

<sup>c</sup> Total days: 1,087 observations with exclusion of 9 not available observations.

Table 2. Median values of daily Pearson's correlation coefficients between O<sub>3</sub> concentrations and independent geographic and meteorological covariates for the year 2010.

Covariables	Buffer (m)	Correlation coefficients (r)						
		SUMMER MONTHS	SYNOPTIC CLASSES					
		(n days =153)	1 (n days =48)	2 (n days =53)	3 (n days =89)	4 (n days =56)	5 (n days =38)	6 (n days =71)
Vegetation density	100	0.75	0.74	0.81	0.81	0.77	0.80	0.77
	200	0.72	0.72	0.79	0.79	0.75	0.78	0.73
	300	0.65	0.66	0.77	0.75	0.68	0.70	0.66
	500	0.58	0.60	0.69	0.70	0.60	0.63	0.58
Industrial land use	300	-0.53	-0.43	-0.61	-0.58	-0.51	-0.48	-0.58
	500	-0.44	-0.37	-0.51	-0.48	-0.43	-0.39	-0.49
	750	-0.49	-0.42	-0.59	-0.53	-0.45	-0.43	-0.56
Commercial land use	500	-0.35	-0.36	-0.25	-0.31	-0.36	-0.30	-0.35
	750	-0.21	-0.20	-0.19	-0.22	-0.26	-0.24	-0.21
Residential land use	300	0.24	0.21	0.17	0.22	0.21	0.22	0.27
	500	0.20	0.20	0.14	0.16	0.19	0.16	0.18
	750	0.26	0.26	0.17	0.21	0.25	0.24	0.23
Population density	1,000	-0.13	-0.20	-0.22	-0.24	-0.20	-0.20	-0.10
	2,500	-0.26	-0.34	-0.37	-0.41	-0.32	-0.32	-0.24
Road density (log of WRD)	100	-0.60	-0.57	-0.66	-0.64	-0.62	-0.61	-0.60
	150	-0.61	-0.57	-0.68	-0.66	-0.62	-0.62	-0.63
	250	-0.64	-0.57	-0.68	-0.66	-0.64	-0.64	-0.68
	1,000	-0.50	-0.42	-0.53	-0.49	-0.49	-0.48	-0.52
Temperature (8-h)		0.10	0.12	-0.23	-0.28	0.08	0.19	0.08
Relative humidity (8-h)		0.31	0.32	0.56	0.57	0.35	0.36	0.35
Wind speed (24-h)		0.49	0.44	0.60	0.52	0.38	0.54	0.53
Precipitation (24-h)		---	-0.19	---	---	---	---	---

- Dashed line indicates no correlation



Table 3. Leave-one-station-out cross-validation results for the O<sub>3</sub> prediction for the year 2010 from the KED with vegetation and road density as covariates, OK, Daily mean and BME model (Adam-Poupart *et al.*, 2014).

Category	Synoptic classes	N (days)	N (estimated points)	Model	R <sup>2</sup>	RMSE (ppb)
SUMMER MONTHS	---	153	1,766	KED	0.92	3.14
	---			OK	0.85	4.00
	---			Mean*	0.83	4.58
	---			BME	0.62	6.37
SYNOPTIC CLASSES	1	48	556	KED	0.92	3.26
				OK	0.88	3.92
				Mean*	0.85	4.62
	2	54	595	KED	0.86	2.79
				OK	0.67	3.80
				Mean*	0.61	4.79
	3	89	1,012	KED	0.86	3.03
				OK	0.74	3.78
				Mean*	0.67	4.78
	4	56	651	KED	0.90	2.85
				OK	0.81	3.71
				Mean*	0.78	4.27
5	38	414	KED	0.77	3.37	
			OK	0.55	3.88	
			Mean*	0.52	5.12	
6	71	822	KED	0.92	3.22	
			OK	0.84	4.21	
			Mean*	0.82	4.89	

\* Daily mean leave-one-station-out cross-validation is computed.

Figure 1. Geographical location of O<sub>3</sub> and meteorological stations in the study region.

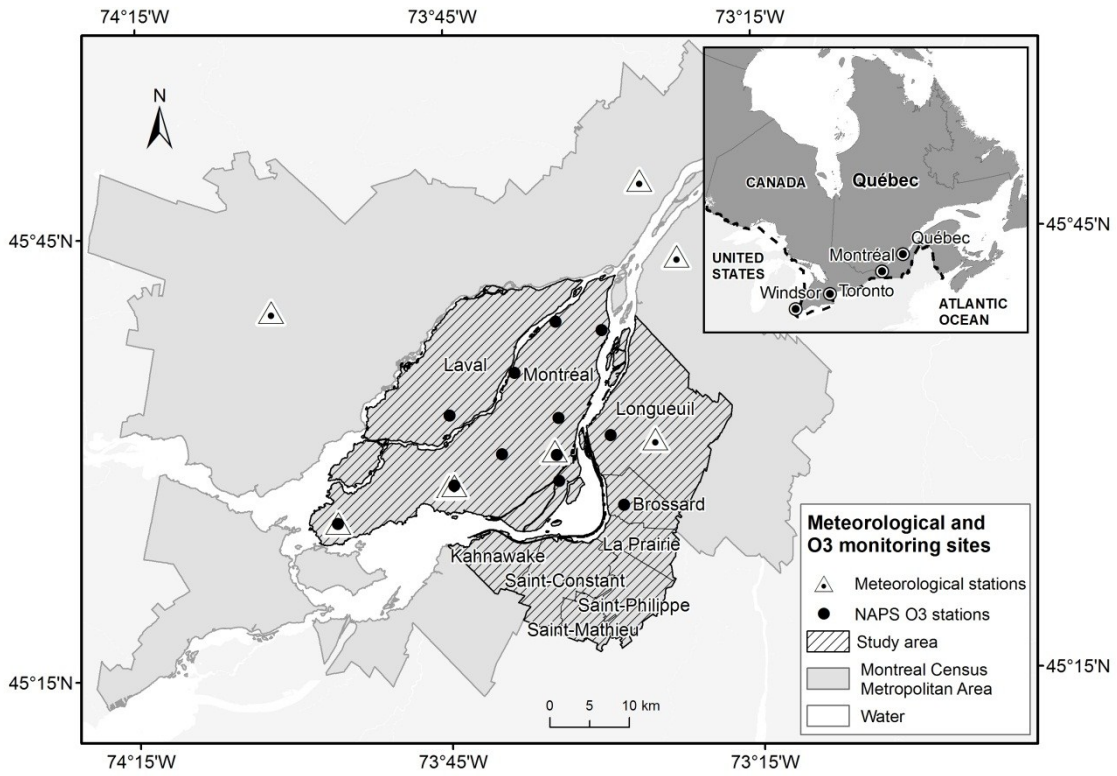


Figure 2. Wind roses (based on 8 meteorological stations) by synoptic weather classes (6) generated using hourly wind speed and direction data of the year 2010.

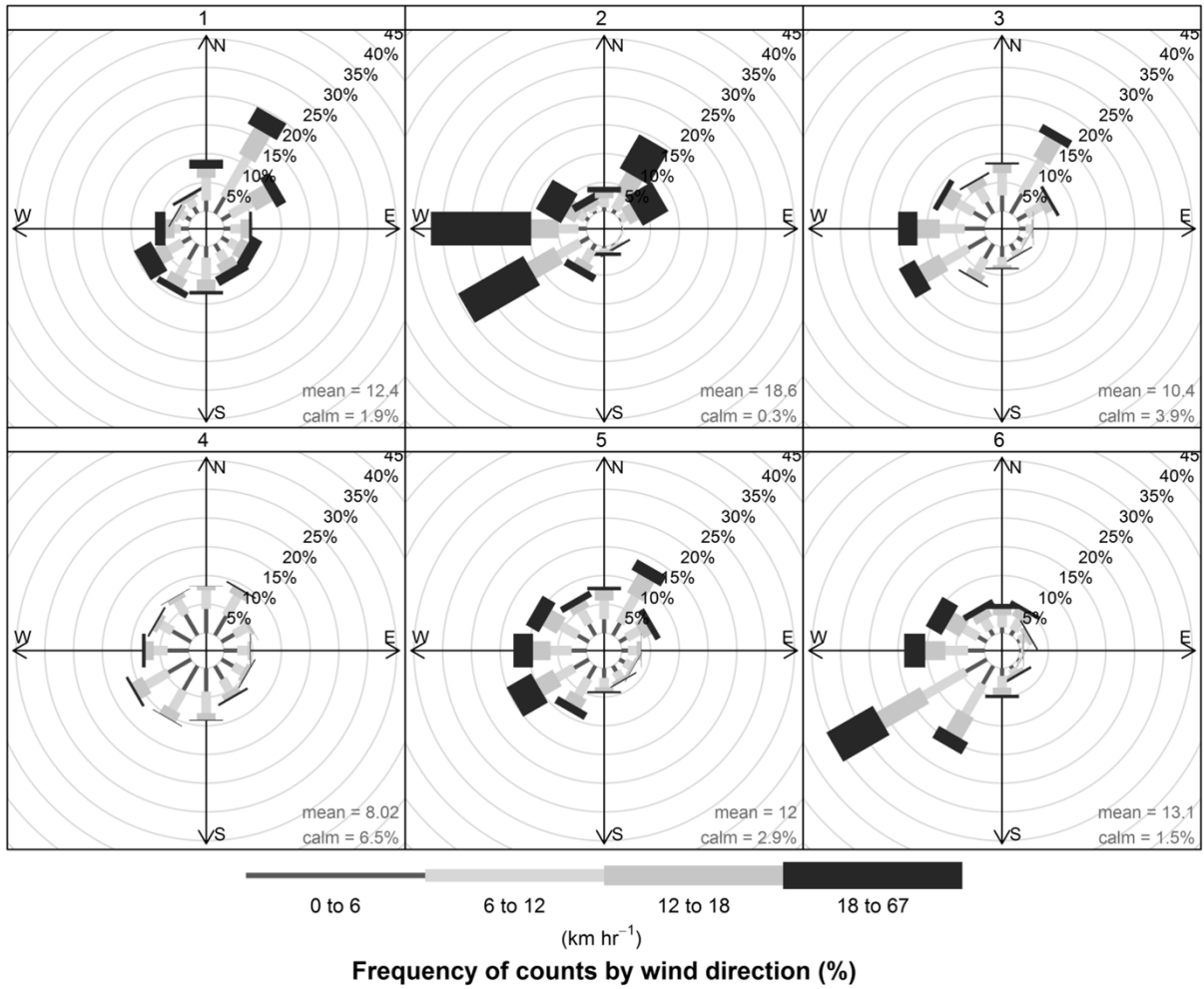


Figure 3. O<sub>3</sub> mapping error estimates from cross validation [where error = predicted – observed O<sub>3</sub> concentration (in ppb) at each monitoring station] based on the KED (mean ± SD; 0.29 ± 3.19 ppb), OK (mean ± SD; -0.13 ± 4.30 ppb), Daily mean (mean ± SD; 0.00 ± 4.58 ppb) and BME (Adam-Poupart *et al.*, 2014) (mean ± SD; 0.63 ± 7.04 ppb) models and for the summer months of the year 2010.

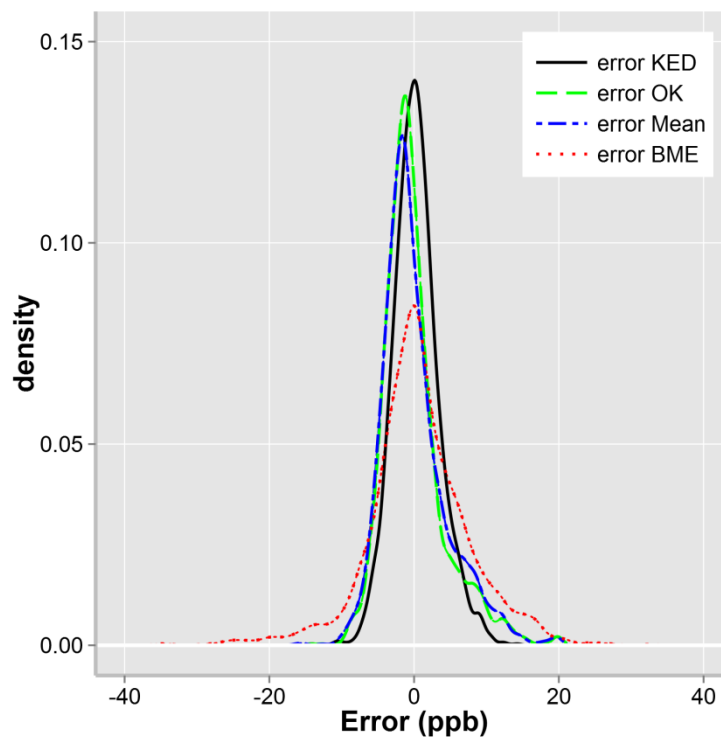


Figure 4. Daily temporal O<sub>3</sub> error estimates (RMSE) based on leave-one-station cross-validation for the KED, OK, Daily mean and BME (Adam-Poupart *et al.*, 2014) models for May through September of 2010.

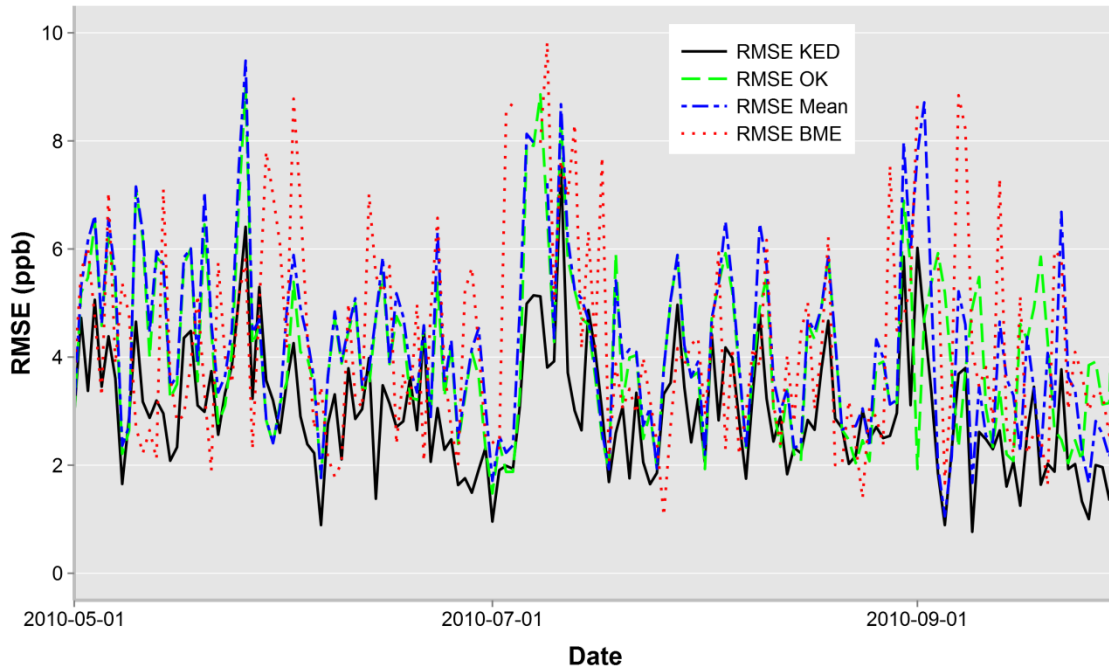
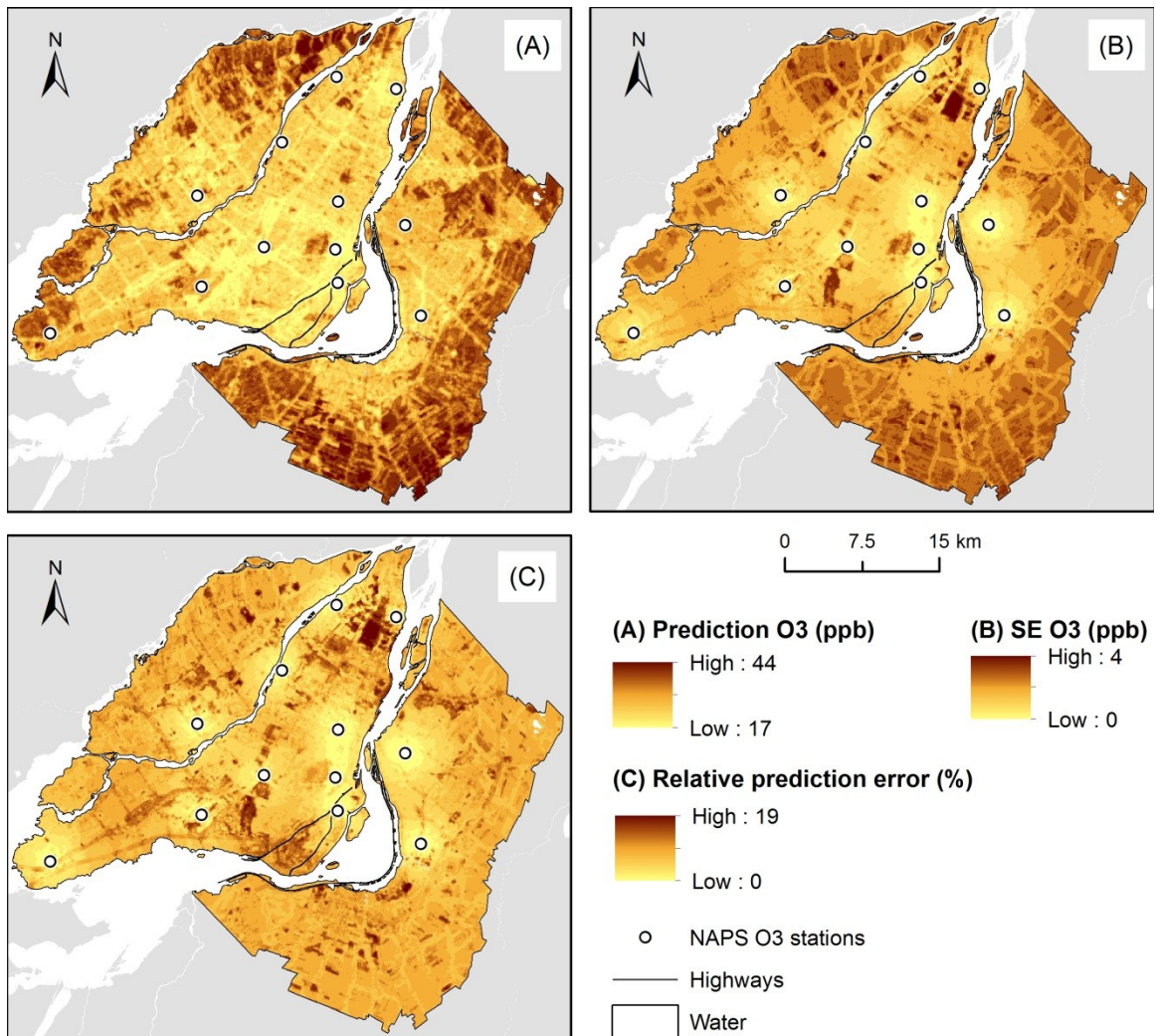


Figure 5. Mean O<sub>3</sub> levels (0900–1700 hours) predicted (A), standard error - SE (B) and relative prediction error (C) from the KED model at a 100-m grid across the Montreal Census Metropolitan Area (Canada) for the year 2010.





**Chapitre 5 : Association of mortality with short-term exposures to ground-level ozone in Montreal, Canada: A comparison of risk estimation from non-spatial and spatio-temporal exposure approaches – Article 3**





**Title:** Association of mortality with short-term exposures to ground-level ozone in Montreal, Canada: A comparison of risk estimation from non-spatial and spatio-temporal exposure approaches.

*Final version of the manuscript ready for submission*

**Authors :** Yuddy Ramos<sup>1</sup>, Benoît St-Onge<sup>2</sup>, Allan Brand<sup>3</sup>, Audrey Smargiassi<sup>3,4</sup>

<sup>1</sup>Département de géographie, Université de Montréal, Montréal, Québec, Canada

<sup>2</sup>Département de géographie, Université du Québec à Montréal (UQAM), Montréal, Québec, Canada

<sup>3</sup>Institut national de santé publique du Québec (INSPQ), Montréal, Québec, Canada

<sup>4</sup>Département de santé environnementale et de santé au travail, Université de Montréal, Montréal, Québec, Canada

### **Contribution de l'auteure principale et des coauteurs :**

L'article présenté dans ce chapitre a été réalisé par Yuddy Ramos sous la supervision de Benoît St-Onge, la cosupervision d'Audrey Smargiassi, et la collaboration d'Allan Brand.

Yuddy Ramos a réalisé les analyses de données, effectué la recherche bibliographique, rédigé l'article dans son entièreté, et produit tous les tableaux et figures. Elle a aussi effectué toutes les modifications selon les suggestions de ses directeurs de recherche jusqu'à la version définitive de l'article prête à être soumise pour publication.

Benoît St-Onge a supervisé les résultats et révisé toutes les versions de l'article avant d'avoir une version prête à être soumise pour publication.

Audrey Smargiassi a proposé la conception méthodologique et fourni toutes les données utilisées dans cet article. Elle a aussi supervisé les analyses, les résultats et révisé toutes les versions de l'article avant d'avoir une version définitive corrigée.

Allan Brand a programmé les méthodes d'évaluation de l'exposition et prétraité les données de santé.

# 1 Abstract

Ground-level ozone ( $O_3$ ) has been related to mortality, emergency department visits and hospital admissions due to respiratory, cardiovascular, and others diseases. To evaluate associations between  $O_3$  levels and mortality, several epidemiological studies estimate the average risk within a given region based on the mean  $O_3$  values of all fixed-site monitoring stations found therein. However, the spatial and temporal variations of air pollution daily levels are rarely considered in these studies. The use of a single mean value from multiple monitoring stations can create uncertainties in the estimation of population exposure. This study investigates the association between  $O_3$  concentrations and the all non-accidental and specific-cause mortality in Montreal, Canada during the summer months (May through September of 1996-2007), using three different approaches to obtain the  $O_3$  concentrations: (1) average at all monitoring stations, (2) levels measured at the nearest monitoring station and, (3) a spatio-temporal bayesian maximum entropy (BME). Data used included daily mortality and 8-hr midday  $O_3$  concentrations. A time-stratified case-crossover epidemiological approach adjusted for temperature was used to analyze the relationship between  $O_3$  concentrations and mortality on 0-day lag and 1-day lag.  $O_3$  was positively associated with the risk of deaths at lag 1 from all non-accidental, cardiovascular, and cardio-respiratory causes; the association and the fit were similar for the three exposure approaches. Our results support that no gain is obtained in the estimating the exposure effect on mortality with a spatio-temporal model for  $O_3$ .

**Keywords** : Air pollution, ground-level ozone, mortality, case-crossover, spatio-temporal model

## 2 Introduction

Ground-level ozone (O<sub>3</sub>) is a secondary pollutant formed by the reaction of nitrogen oxides (NO<sub>x</sub>) and volatile organic compounds (VOC) in the presence of sunlight (Diem, 2000; McKendry, 1993; U.S. EPA., 2013). The O<sub>3</sub> production is attributed to anthropogenic sources (motor vehicles and industrial emission) and natural sources that include biogenic emission (isoprene and terpenes from vegetation) and abiotic emission (biomass burning, lightning and geogenic sources) (U.S. EPA., 2013).

Ground-level ozone is harmful to human health (Goldberg *et al.*, 2001; Zeng *et al.*, 2008) and it is well established that short-term exposure to O<sub>3</sub> contributes to non-accidental and cardiopulmonary-related mortality (U.S. EPA., 2013; Zeng *et al.*, 2008). Stronger associations have been reported during the warm months when O<sub>3</sub> levels are higher (Bell *et al.*, 2005; Sheffield *et al.*, 2015; Stieb *et al.*, 2009). Mortality effects of O<sub>3</sub> were mainly seen at lag 0 (exposure on the same day as death occurred) and at lag 1 (exposure 1 day before the death occurred) (Bell *et al.*, 2005; Zanobetti *et al.*, 2008).

Young children, the elderly and persons with pre-existing respiratory diseases are particularly susceptible to O<sub>3</sub> effects (Pride *et al.*, 2015). Besides effects on mortality, short-term exposures to O<sub>3</sub> have been related to respiratory symptoms, increase in hospital admissions (HA) and emergency department (ED) visits for respiratory problems such as chronic obstructive pulmonary disease (COPD) and asthma (Burnett *et al.*, 1994; Stieb *et al.*, 2009; Strickland *et al.*, 2010; U.S. EPA., 2013; Villeneuve *et al.*, 2007; Q. Yang *et al.*, 2003).

Epidemiological studies on the effects of O<sub>3</sub> often use time-series analyses (Bell *et al.*, 2005; Goldman *et al.*, 2010; Strickland *et al.*, 2010), in which the daily O<sub>3</sub> exposure is computed as an average at all monitoring station across an entire area, thus ignoring spatial variations.

While the spatial variation of O<sub>3</sub> levels has usually been ignored in epidemiological studies, significant spatial variations of O<sub>3</sub> at different geographic scales have been noted, and tentatively attributed to the NO<sub>x</sub> scavenging process (Brauer *et al.*, 2012). It is also known that spatial and temporal variations of O<sub>3</sub> are influenced by meteorological conditions (Hidy, 2000).

Some epidemiological studies have to a limited extent taken into consideration the influence of the spatial variation in O<sub>3</sub> exposure levels by using the nearest monitoring site (proximity-based method) to estimate individual's exposure (Jerrett *et al.*, 2005; Nieuwenhuijsen, 2015). More sophisticated methods, such as spatial interpolation techniques have also been used (Bell, 2006). These served to model air pollution in a continuous way, assuming gradual transitions between measuring stations (Jerrett *et al.*, 2005), in order to obtain estimates of the pollutants concentrations at unsampled sites. These interpolation methods can be either in geometric in nature, and based on a single mathematical function (without taking spatial autocorrelation into account), or stochastic/geostatistical, and based on statistical modeling that includes spatial autocorrelation. In recent years, spatiotemporal or space-time models (Adam-Poupart *et al.*, 2014; Bogaert *et al.*, 2009; Y. Yang *et al.*, 2015) extended the spatial statistics approach by including time. Spatiotemporal modeling

simultaneously takes into account both the spatial and temporal dependence structures in the data to estimate air pollution over space and time within a single framework.

Omitting to consider the spatial variability of air pollutant levels, for example when using a single average of O<sub>3</sub> across an entire study area, can create uncertainties on the exposure of individuals of a population and, therefore uncertainties in health risk assessment (Barnett *et al.*, 2012; Butland *et al.*, 2013; Chen *et al.*; Goldman *et al.*, 2012; Strickland *et al.*, 2015). According to Strickland *et al.* (2011) the difference in uncertainty of health risks, by considering or not the spatial variation, will be small or large, depending of the spatial distribution of the air pollutants (homogeneous or heterogeneous). Also Goldman *et al.* (2012) demonstrated that geostatistical approach improves exposure prediction in the presence of spatial autocorrelation in the data, and that greater prediction errors can occur when no autocorrelation is present, thus reducing the statistical power of the epidemiologic analysis. To our knowledge, no studies exist which compared the effect of taking spatial variation of O<sub>3</sub> into consideration, or not, on the mortality risk assessment, using a single dataset.

In this context, our specific objective was to compare three approaches for estimating daily levels of exposure to O<sub>3</sub> concentrations in relation to mortality in Montreal (Canada). These approaches differed in the way they take or not into consideration the spatial or temporal aspect of the O<sub>3</sub> variations. The first approach consisted in simply computing daily mean concentrations over the study area and link mortality cases to the variations of this mean. It is devoid of any spatial or temporal modeling. The second approach linked the mortality cases to the daily concentrations observed at the closest O<sub>3</sub> measurements station, thus



implementing a very simple spatial model, subdividing the study area in proximal polygons (Voronoi cells). Finally, a third approach was used to predict levels of O<sub>3</sub> exposure based on spatiotemporal modeling of the O<sub>3</sub> fluctuations using a BME model. We surmise that taking into account the spatial and temporal variations, would provide better exposure estimates, and therefore increase our capacity to predict the risk of mortality.

We also aimed to explore if considering the spatial variation of ozone levels improves the estimation of the risk of death related to daily ozone levels, in a region with a temperate climate and when relatively few monitoring stations are available (~14 to 16 stations par 4 292 km<sup>2</sup>).

### **3 Methods**

This study obtained ethics approval from the Quebec Public Health Ethical Health Surveillance Committee and from CERES (Comité d'éthique de la recherche en santé) - Université de Montréal (certificate number: 15-014-CERES-D).

#### **Study population**

The study population is comprised of all residents of the Montreal census metropolitan area (CMA) who died from non-accidental causes during the summer months (May through September) of 1996-2007. Montreal CMA with a population of 3.82 million people (Statistics

Canada, 2015) includes the city of Montreal that is bordered by the Rivière des Prairies and the Saint Lawrence River (Figure 1).

### **Mortality data**

We obtained daily mortality data from the Quebec Institute of Public Health info-center. This data included information such as the age of each decedent, the six digit postal code of the residence at the time of death, and the cause of death according to the international classification of diseases (ICD). We included for analysis the following causes of deaths: all non-accidental (all deaths except ICD9 800 to 999 and ICD10 V to Z), pulmonary (ICD-9 460 - 519 and ICD10 J) and, cardiovascular (ICD9 390 - 459 and ICD10 I).

In order to geolocalize each decedent, the 6-digit postal code of the place of death was associated to the corresponding geographic coordinates obtained from the Postal Code Conversion File (PCCF) from Statistic Canada. For the year – 1996 we used the PCCF of 1996, for the years 1997-2000, the PCCF of 2001, for 2001 to 2005 the PCCF of 2006 and for 2006 and 2007, the PCCF of 2006. The 6-digit postal code unit represents an average of 35 households per unit.

### **Ground-level ozone data**

All three approaches used to estimate exposure to O<sub>3</sub> at the residential six digit postal code of the decedent on the date of deaths relied on hourly measurements of O<sub>3</sub> at fixed-site monitoring stations across the Montreal CMA (~14 to 16 stations per year) obtained from the

National Air Pollution Surveillance (NAPS) network. Daily mean 8-hr midday (0900-1700 hours) O<sub>3</sub> concentrations during the summer months (May through September) were computed per station for the study period.

The first (“daily mean”) approach consists in simply computing, for each observation day, the average of all stations 8-hr means of the study area. The second approach (“nearest monitoring station”) relied on a spatial linkage of daily levels of the closest monitoring station to the residence location of the decedent determined by the 6-digit postal code (zip code). In urban areas with highly urbanized areas as the city of Montreal, a residential postal code typically represents a small geographic area and in rural areas it may represent a larger area. On average a residential 6-digit postal code includes about 35 persons.

The third approach was based on O<sub>3</sub> estimations obtained with a spatiotemporal BME (Bayesian maximum entropy) model (Adam-Poupart *et al.*, 2014). A BME model is a nonlinear model for space-time data based on physical measures (“hard” data) and additional information such as the multivariate regression estimates (“soft” data) of Adam-Poupart *et al.*, (2014). The “soft data” refers to information used to compensating for the limited amount of measured. The “hard data” used by Adam-Poupart *et al.*, (2014) were the O<sub>3</sub> concentration data provided by the NAPS stations and the “soft data” were estimates from a land-use mixed effects regressions models that involved variables such as temperature, precipitation, day of year, year, road density in 1-km buffer areas, and latitude (Adam-Poupart *et al.*, 2014).

The O<sub>3</sub> concentrations at the geolocation of each decedent were estimated from the BME model developed over Québec by Adam-Poupart *et al.* (2014) for the summer months (May through September).

### **Temperature data**

Temperature data was used to adjust the conditional logistic regression models (see Statistical analyses below). Daily mean 8-hr (0900-1700 hours) temperature measured at Dorval station (Pierre Elliot Trudeau Airport) during summer months was computed from the National Climatic Data and Information Archive (DAI) of Environment Canada. ("Canada National Climate Data and Information Archive," 2013)

### **Analysis**

We investigated the association between daily O<sub>3</sub> concentrations and mortality using a case-crossover observational epidemiological design (Maclure, 1991) and in which control days for each death were selected by means of the time-stratified approach (Lumley *et al.*, 2000). Thus, the study period was divided into monthly strata and control days for each case were selected as the same day of the week in the month as when death occurred. In other words, each death served as his/her own control in the selected exposure window of O<sub>3</sub>.

The analysis of the relationship between O<sub>3</sub> and mortality was assessed using conditional logistic regression models with the R software (R Project for Statistical Computing; <http://www.r-project.org>). O<sub>3</sub> levels of control periods were compared with O<sub>3</sub>

levels during the hazard period, on the day of death (lag 0) and on the day before of death (lag 1). Model fit was assessed using Akaike's Information Criterion (AIC). Conditional logistic regression models were adjusted for temperature using cubic spline functions with 3 knots. Odds ratio (OR) and 95% confidence interval (CI) were reported for an increase in the concentration corresponding to an interquartile range (IQR) of O<sub>3</sub> concentrations.

Associations were performed for all non-accidental deaths as well as separately for respiratory, cardiovascular and cardio-respiratory deaths. The BME exposure prediction set was complete for all study days while there were days with missing information on average ozone levels and on levels at closest monitor. Thus case date with missing data for these two exposure measures were excluded, with their controls, in order to use the same "n" for models based on the three exposure variables.

To gain insight in the results, we also calculated the average of O<sub>3</sub> daily standard deviation over all monitoring sites (reflecting daily spatial variations), the standard deviation of O<sub>3</sub> daily means over all days (reflecting variations over time), and the average distance between monitoring stations and the estimated location of the residence of decedents, to assess the average spatial resolution of nearest monitoring station approach. Also, we computed the coefficient of variation (standard deviation/mean) based on daily measurements at monitoring station for comparing the degree of spatial variation with other studies.

## 4 Results

There were 114 870 non-accidental deaths in the Montreal CMA during the summer months of 1996 to 2007. As described previously, missing data (N = 102 days) were excluded in order to use the same “n” for the three dataset of O<sub>3</sub> (i.e. average over monitoring stations, nearest monitoring station and BME model (Adam-Poupart *et al.*, 2014)). Finally, a dataset with a total number of 98 682 of non-accidental deaths were used in this study. The daily variation of these numbers of deaths is shown in the Figure 2. Mortality due to respiratory causes represented 7.35%, and mortality due to cardiovascular cause represented 30.93% of all non-accidental deaths.

Table 1 shows descriptive statistics for each dataset of O<sub>3</sub> levels, for temperature and for daily number of deaths considered for the study period. The mean 8-hr (SD) of O<sub>3</sub> levels (cases and controls) for the dataset of the average over all monitoring stations was 30.12 ppb (12.27 ppb), for the nearest monitoring stations it was 29.56 ppb (13.61 ppb), and for the BME model (Adam-Poupart *et al.*, 2014) it was 30.81 ppb (8.47 ppb). Moreover, the average distance between the location of decedents and the nearest O<sub>3</sub> monitoring site was 8 km, distance within which a same level of O<sub>3</sub> is assumed in the case of the Voronoi cell tessellation.

Table 2 shows associations between daily 8-hr average O<sub>3</sub> levels and mortality adjusted for temperature for the three datasets; average at all monitoring stations, the nearest monitoring station, and the BME model estimates (Adam-Poupart *et al.*, 2014), for two different lag periods (lag 0 and lag 1) and for five causes of deaths grouping. The ORs are

shown per increments of IQR in daily O<sub>3</sub> levels. Overall, we found positive associations at lag 1 for all O<sub>3</sub> estimation approaches used and for non-accidental, cardiovascular and cardio-respiratory causes. According to the AIC the three ozone datasets produced similar fits. The OR estimated using the BME model (Adam-Poupart *et al.*, 2014) at lag 1 days, for an increase of ozone of 10.89 ppb, for all non-accidental causes was of 1.025 (95% CI: 1.009 to 1.041), the OR for cardiovascular causes was of 1.044 (95% CI: 1.015 to 1.073) and, the OR for cardio-respiratory causes was 1.050 (95% CI: 1.024 to 1.077).

Table 3 shows spatial and temporal variations for the study period (summer months from 1996 to 2007). The spatial variation for the study period was almost half of the variation over time.

Finally, the coefficient of variation (standard deviation/mean) for the study period (1996 to 2007) was in average 22% (ranged from 17% in 2005 and 2007, up to 28% in 1996).

## **5 Discussion**

We found small and positive associations between daily O<sub>3</sub> concentrations (from three O<sub>3</sub> datasets used in this study) and all non-accidental, cardiovascular and cardio-respiratory causes of deaths in the warm months. These results are in line with numerous studies that previously showed a relationship between O<sub>3</sub> concentrations and mortality (Goldberg *et al.*, 2001; Jerrett *et al.*, 2009).

The associations and model fits were similar with the three exposure methods; the association was not more accurate when the exposure to O<sub>3</sub> was based on a spatio-temporal model (i.e. the BME). This lack of difference between associations obtained with different exposure approaches was observed with a spatial variation (estimation based on ~14-16 stations per year) that was not negligible (5.85 ppb) and that was half the temporal variation (12.18 ppb) in O<sub>3</sub> levels (Table 3). These variations are similar with those of Lee *et al.* (2010) who based on 23 stations of O<sub>3</sub> localized across the Greater London (England), reported a level of spatial variation of 9.80 ppb and a level of temporal variation of 19.4 ppb for a 3 years study period (2003-2005).

The coefficient of variation (22%) for our study period was also similar to the one of Brauer *et al.* (2012) who reported spatial variations greater than 20% at within-city scale in O<sub>3</sub> and at all spatial scales due to the O<sub>3</sub> destruction by NO<sub>x</sub> scavenging process. However, Brauer *et al.* (2012) also indicated that the magnitude of O<sub>3</sub> spatial variation (< 65%) is relatively small compared to that of primary PM<sub>2.5</sub> (> 200%) considered as spatially heterogeneous. These are consistent with others studies, which indicate that primary pollutants largely from mobile source as carbon monoxide (CO), nitrogen dioxide (NO<sub>2</sub>), sulfur dioxide (SO<sub>2</sub>), and elemental carbon (EC) are spatially heterogeneous (greater spatial variation) and that secondary pollutants like O<sub>3</sub>, nitrate (NO<sub>3</sub><sup>-</sup>), sulfate (SO<sub>4</sub><sup>2-</sup>) are spatially more homogeneous (Darrow *et al.*, 2009; Goldman *et al.*, 2010; Sororian *et al.*, 2014). That is to say, the primary pollutants lack or have low spatial autocorrelation, while the secondary pollutants have higher spatial autocorrelation (Sororian *et al.*, 2014), and the pollutants of mixed origin like PM<sub>2.5</sub> total mass and organic carbon (OC) have intermediate levels of spatial



autocorrelation (Goldman *et al.*, 2012). High spatial autocorrelation of O<sub>3</sub> was also shown by Ito *et al.* (2005) who demonstrated similar on monitor-to-monitor temporal correlation (~0.6 to 0.8) for O<sub>3</sub> when the separation distance was smaller than 80 km.

The lack of difference between associations and model fits obtained with different exposure prediction methods is also corroborated with results of previous works. Our results are for example consistent with those of Lee *et al.* (2010) who used non-spatial (i.e. average at all monitoring stations) and spatiotemporal (Bayesian model) approaches to estimate short-term effects of air pollution on daily respiratory mortality in Greater London (England), and found that for secondary pollutants like O<sub>3</sub> both modeling approaches produce similar risk estimates. Also, our results are consistent with those of Sarnat *et al.* (2010) who examined the effects of exposure measurement errors from air pollutant spatial variability in time-series studies in Atlanta, Georgia (USA) and indicated that non-spatial estimation approaches show similar associations for O<sub>3</sub>, regardless of the monitoring site location (urban or rural). As suggested by Goldman *et al.* (2010), the lack of difference between associations and fits obtained with exposure approaches that consider or not spatial variability demonstrates that the amount of errors resulting from models ignoring spatial variations is smaller for secondary pollutants like O<sub>3</sub> (high spatial autocorrelation) compared to that of primary pollutants (lack or low spatial autocorrelation).

The lack of increase in fit quality when using BME instead of simpler modeling, can also be attributed to the fact that the BME approach fits a single modeling framework to numerous spatially distributed and possibly heterogeneous observations over several years.

Heterogeneity may arise from the fact that temporal and spatial variations may behave differently depending on the weather regime (e.g. normal atmosphere vs. inversion conditions), season, wind strength, etc. In previous studies, we showed that modeling of the spatial variations in the concentration of O<sub>3</sub> or PM<sub>2.5</sub> varied in accuracy between different classes of weather regimes (Ramos *et al.*, 2015a; Ramos *et al.*, 2015b). Beside this, Hidy (2000) emphasized the importance of meteorological conditions on all spatial and temporal scales for foreseeing extreme concentrations of O<sub>3</sub>. The full spatiotemporal modeling, per weather regime, over several years is certainly a complex task but could lead to improvements in the capacity to assess mortality risks.

## **6 Conclusions**

The main objective of this study was to compare the evaluation of risks of mortality with three O<sub>3</sub> predictions obtained with non-spatial and spatiotemporal approaches in Montreal CMA, Québec, Canada. No difference was noted when the estimation of the risk of death considered the spatiotemporal dependence in the estimation of ozone levels compared to the models ignoring spatial variations. This is likely due to larger the temporal variability in O<sub>3</sub> levels, compared to its spatial variability in the study area, and possibly, to imperfect modeling of the complex spatio-temporal variations.

## **7 Conflict of interest**

The authors declare no conflict to interest.

## 8 References

- Adam-Poupart, A., Brand, A., Fournier, M., Jerrett, M., and Smargiassi, A. (2014). Spatiotemporal modeling of ozone levels in Quebec (Canada): A comparison of kriging, land-use regression (LUR), and combined bayesian maximum entropy–LUR approaches. *Environmental Health Perspectives*, 122, 970–976.
- Barnett, A. G., Clements, A. C. A., and Vaneckova, P. (2012). Estimating the effects of environmental 428 exposures using a weighted mean of monitoring stations. *Spatial and Spatio-temporal Epidemiology*, 3, 225–234.
- Bell, M. L. (2006). The use of ambient air quality modeling to estimate individual and population exposure for human health research: A case study of ozone in the Northern Georgia Region of the United States. *Environment International*, 32, 586–593.
- Bell, M. L., Dominici, F., and Samet, J. M. (2005). A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity, mortality, and air pollution study. *Epidemiology*, 16(4), 436–445.
- Bogaert, P., Christakos, G., Jerrett, M., and Yu, H.-L. (2009). Spatiotemporal modelling of ozone distribution in the state of California. *Atmospheric Environment*, 43, 2471–2480.

- Brauer, M., Hystad, P., and Poplawski, K. (2012). Assessing the spatial representativeness of PM<sub>2.5</sub> and O<sub>3</sub> measurements from the national air pollutant surveillance system. Population and Public Health, School of (SPPH) Publications, U. o. B. Columbia.
- Burnett, R. T., Dales, R. E., Raizenne, M. E., Krewski, D., Summers, P. W., Roberts, G. R., Raad-Young, M., Dann, T., and Brook, J. (1994). Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admission to Ontario hospitals. Environmental Research, 65(172–194).
- Butland, B. K., Armstrong, B., Atkinson, R. W., Wilkinson, P., Heal, M. R., Doherty, R. M., and Vieno, M. (2013). Measurement error in time-series analysis: a simulation study comparing modelled and monitored data. Medical Research Methodology, 13 (126), 1–12.
- Canada National Climate Data and Information Archive. (2013). Retrieved 5 March 2013, from [http://climate.weather.gc.ca/index\\_e.html](http://climate.weather.gc.ca/index_e.html)
- Chen, L., Mengersen, K., and Tong, S. Spatiotemporal relationship between particle air pollution and respiratory emergency hospital admissions in Brisbane, Australia. Science of the Total Environment, 373, 57–67.
- Darrow, L. A., Klein, M., Flanders, W. D., Waller, L. A., Correa, A., Marcus, M., Mulholland, J. A., Russell, A. G., and Tolbert, P. E. (2009). Ambient air pollution and preterm birth: A time-series analysis. Epidemiology, 20(5), 689–698.

- Diem, J. E. (2000). Comparisons of weekday-weekend ozone: importance of biogenic volatile organic compound emissions in the semi-arid southwest USA. *Atmospheric Environment*, 34, 3445–3451.
- Goldberg, M. S., Burnett, R. T., Brook, J., Bailar III, J. C., Valois, M., and Vincent, R. (2001). Associations between daily cause-specific mortality and concentrations of ground-level ozone in Montreal, Quebec. *Am J Epidemiol*, 154(9), 817–826.
- Goldman, G. T., Mulholland, J. A., Russell, A. G., Gass, K., Strickland, M. J., and Tolbert, P. E. (2012). Characterization of ambient air pollution measurement error in a time-series health study using a geostatistical simulation approach. *Atmospheric Environment*, 57, 101–108.
- Goldman, G. T., Mulholland, J. A., Russell, A. G., Srivastava, A., Strickland, M. J., Klein, M., Waller, L. A., Tolbert, P. E., and Edgerton, E. S. (2010). Ambient air pollutant measurement error: characterization and impacts in a time-series epidemiologic study in Atlanta. *Environmental Science & Technology*, 44, 7692–7698.
- Hidy, G. M. (2000). Ozone process insights from field experiments Part I: Overview. *Atmospheric Environment*, 34, 2001–2022.
- Ito, K., De Leon, S., Thurston, G. D., Nádas, A., and Lippmann, M. (2005). Monitor-to-monitor temporal correlation of air pollution in the contiguous US. *Journal of Exposure Analysis and Environmental Epidemiology*, 15, 172–184.

- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., and Giovis, C. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Science and Environmental Epidemiology*, 15, 185–204.
- Jerrett, M., Burnett, R. T., Pope III, C. A., Ito, K., Thurston, G., Krewski, D., Shi, Y., Calle, E., and Thun, M. (2009). Long-term ozone exposure and mortality. *The New England Journal of Medicine*, 360, 1085–1095.
- Lee, D., and Shaddick, G. (2010). Spatial modeling of air pollution in studies of its short-term health effect. *Biometric*, 66, 1238–1246.
- Lumley, T., and Levy, D. (2000). Bias in the case-crossover design implications for studies of air pollution. *Environmetrics*, 11, 578–693.
- Maclure, M. (1991). The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, 133(2), 144–153.
- McKendry, I. G. (1993). Ground-level ozone in Montreal, Canada. *Atmospheric Environment*, 27B(1), 93–103.
- Nieuwenhuijsen, M. J. (2015). *Exposure Assessment in Environmental Epidemiology* (Second edition ed.). New York: Oxford University Press.

- Pride, K. R., Peel, J. L., Robinson, B. F., Busacker, A., Grandpre, J., Bisgard, K. M., Yip, F. Y., and Murphy, T. D. (2015). Association of short-term exposure to ground-level ozone and respiratory outpatient clinic visits in a rural location – Sublette County, Wyoming, 2008–2011. *Environmental Research*, 135, 1–7.
- Ramos, Y., St-Onge, B., Blanchet, J.-P., Kestens, Y., and Smargiassi, A. (2015a). Spatiotemporal modeling of daily concentrations of ground-level ozone in Montreal, Canada. *Atmospheric Environment*, Submitted for publication.
- Ramos, Y., St-Onge, B., Blanchet, J.-P., and Smargiassi, A. (2015b). Spatio-temporal models to estimate daily concentrations of fine particulate matter in Montreal: Kriging with external drift and inverse distance-weighted approaches. *Journal of Exposure Science and Environmental Epidemiology*, 1–10.
- Sarnat, S. E., Klein, M., Sarnat, J. A., Flanders, W. D., Waller, L. A., Mulholland, J. A., Russell, A. G., and Tolbert, P. E. (2010). An examination of exposure measurement error from air pollutant spatial variability in time-series studies. *Journal of Exposure Science and Environmental Epidemiology*, 20, 135–146.
- Sheffield, P. E., Zhou, J., Shmool, J. L., and Clougherty, J. E. (2015). Ambient ozone exposure and children’s acute asthma in New York City: a case-crossover analysis. *Environmental Health*, 14:25, 10.
- Sororian, S. A., Holmes, H. A., Friberg, M., Ivey, C., Hu, Y., Mulholland, J. A., Russell, A. G., and Strickland, M. J. (2014). Temporally and spatially resolved air pollution in Georgia

using fused ambient monitor data and chemical transport model results. In D. Steyn & R. Mathur (Eds.), *Air Pollution Modeling and its Application XXIII* (pp. 301–306). Switzerland: Springer International Publishing.

Statistics Canada. (2015). Census profile - Montreal CMA, Canada 2011 Census. Retrieved December, 2015, from <http://www12.statcan.ca/census-recensement/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMA&Code1=462&Geo2=PR&Code2=24&Data=Count&SearchText=montreal&SearchType=Begins&SearchPR=01&B1=All&Custom=&TABID=1>

Stieb, D. M., Szyszkowicz, M., Rowe, B. H., and Leech, J. A. (2009). Air pollution and emergency department visits for cardiac and respiratory conditions: a multi-city time-series analysis. *Environmental Health*, 8(25), 13.

Strickland, M. J., Darrow, L. A., Klein, M., Flanders, W. D., Sarnat, J. A., Waller, L. A., Sarnat, S. E., Mulholland, J. A., and Tolbert, P. E. (2010). Short-term associations between ambient air pollutants and pediatric asthma emergency department visits. *American Journal of Respiratory and Critical Care Medicine*, 182, 307–316.

Strickland, M. J., Darrow, L. A., Mulholland, J. A., Klein, M., Flanders, W. D., Winqvist, A., and Tolbert, P. E. (2011). Implications of different approaches for characterizing ambient air pollutant concentrations within the urban airshed for time-series studies and health benefits analyses. *Environmental Health*, 10(36), 1–9.



- Strickland, M. J., Gass, K. M., Goldman, G. T., and Mulholland, J. A. (2015). Effects of ambient air pollution measurement error on health effect estimates in time-series studies: a simulation-based analysis. *Journal of Exposure Science and Environmental Epidemiology*, 25, 160–166.
- U.S. EPA. (2013). Final Report: Integrated Science Assessment of Ozone and Related Photochemical Oxidants. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-10/076F, 2013.
- Villeneuve, P. J., Chen, L., Rowe, B. H., and Coates, F. (2007). Outdoor air pollution and emergency department visits for asthma among children and adults: A case-crossover study in northern Alberta, Canada. *Environmental Health*, 6(40), 1–15.
- Yang, Q., Chen, Y., Shi, Y., Burnett, R. T., McGrail, K. M., and Krewskif, D. (2003). Association between ozone and respiratory admissions among children and the elderly in Vancouver, Canada. *Inhalation Toxicology*, 15, 1297–1308.
- Yang, Y., and Christakos, G. (2015). Spatiotemporal characterization of ambient PM<sub>2.5</sub> concentrations in Shandong province (China). *Environ Science & Technology*, 49, 13431–13438.
- Zanobetti, A., and Schwartz, J. (2008). Mortality displacement in the association of ozone with mortality: an analysis of 48 cities in the United States. *American Journal Respiratory and Critical Care Medicine*, 177, 184–189.

Zeng, G., Pyle, J. A., and Young, P. J. (2008). Impact of climate change on tropospheric ozone and its global budgets. *Atmospheric Chemistry and Physics*, 8, 369–387.

## 9 Tables and figures

Table 1. Descriptive statistics of causes of deaths, ground-level ozone and temperature in the Montreal region for summer months (May through September) from 1996 to 2007 (N cases and controls = 435 520 and N days of measurements = 1 734).

	Mean	SD	Minimum	25th percentile	75th percentile	Maximum	IQR
Daily number of deaths							
All non-accidental	56.91	10.42	25	50	64	93	---
Respiratory	4.19	2.28	0	3	6	18	---
Cardiovascular	17.60	4.94	4	14	21	37	---
Cardio-respiratory	21.79	5.70	7	18	25	49	---
8h mean temperature (°C) for cases and controls	20.59	4.86	3.21	17.46	24.21	32.13	---
8h mean O <sub>3</sub> (ppb) for cases and controls							
Average over all monitoring station	30.12	12.27	2.84	21.62	36.70	79.15	15.08
Nearest monitoring station	29.56	13.61	0.00	20.12	37.25	104.60	17.13
BME model	30.81	8.47	2.13	25.40	36.29	58.79	10.89

Table 2. Risk of non-accidental mortality for increments of IQR of daily O<sub>3</sub> levels evaluated at lag 0 and lag 1 day, in Montreal region for summer months (May through September) from 1996 to 2007.

O <sub>3</sub> dataset	Cause of death	Lag days	OR (95% CI)	IQR	N cases	N cases and controls	AIC
Average at all monitoring station	All non-accidental	0	1.016 (1.001 - 1.030)	15.08	98 682	435 520	284 425
		1	1.024 (1.013 - 1.034)	15.08	98 682	435 520	284 411
	Respiratory	0	0.984 (0.935 - 1.037)	15.08	7 258	31 862	20 897
		1	1.040 (1.001 - 1.081)	15.08	7 258	31 862	20 893
	Cardiovascular	0	0.999 (0.974 - 1.025)	15.08	30 521	134 811	88 008
		1	1.033 (1.014 - 1.053)	15.08	30 521	134 811	87 997
	Cardio-respiratory	0	0.997 (0.974 - 1.020)	15.08	37 779	166 673	108 908
		1	1.035 (1.017 - 1.052)	15.08	37 779	166 673	108 893
Nearest monitoring station	All non-accidental	0	1.016 (1.001 - 1.031)	17.13	98 682	435 520	284 425
		1	1.022 (1.010 - 1.034)	17.13	98 682	435 520	284 415
	Respiratory	0	1.006 (0.954 - 1.062)	17.13	7 258	31 862	20 897
		1	1.037 (0.995 - 1.080)	17.13	7 258	31 862	20 894
	Cardiovascular	0	1.009 (0.983 - 1.037)	17.13	30 521	134 811	88 008
		1	1.037 (1.017 - 1.059)	17.13	30 521	134 811	87 996
	Cardio-respiratory	0	1.009 (0.985 - 1.033)	17.13	37 779	166 673	108 908
		1	1.037 (1.019 - 1.057)	17.13	37 779	166 673	108 893
BME model	All non-accidental	0	1.036 (0.996 - 1.078)	10.89	98 682	435 520	284 426
		1	1.025 (1.009 - 1.041)	10.89	98 682	435 520	284 420
	Respiratory	0	1.023 (0.886 - 1.182)	10.89	7 258	31 862	20 897
		1	1.077 (1.017 - 1.141)	10.89	7 258	31 862	20 891
	Cardiovascular	0	1.018 (0.948 - 1.093)	10.89	30 521	134 811	88 008
		1	1.044 (1.015 - 1.073)	10.89	30 521	134 811	87 999
	Cardio-respiratory	0	1.019 (0.956 - 1.086)	10.89	37 779	166 673	108 908
		1	1.050 (1.024 - 1.077)	10.89	37 779	166 673	108 894

Table 3. Summary of spatial and temporal variation in the O<sub>3</sub> dataset: Spatial variation relate to the average of O<sub>3</sub> (ppb) daily standard deviation over all monitoring sites and the temporal variation relate to standard deviation of O<sub>3</sub> (ppb) daily means over all days for the study period (May through September, 1996 to 2007).

O <sub>3</sub> dataset								
Spatial variation (ppb)				Temporal variation (ppb)				
<b>Lag 0</b>	<b>5.85</b>	Median of dailySD	Min. of daily SD	Max. of daily SD	<b>12.18</b>	Median of daily means	Min. of daily means	Max. of daily means
		5.47	1.56	15.14		28.50	2.84	79.15
<b>Lag 1</b>	<b>5.89</b>	Median of dailySD	Min. of daily SD	Max. of daily SD	<b>12.16</b>	Median of daily means	Min. of daily means	Max. of daily means
		5.59	1.30	15.38		28.77	2.84	79.15

Figure 1. Study area: Montreal census metropolitan area, Québec, Canada.

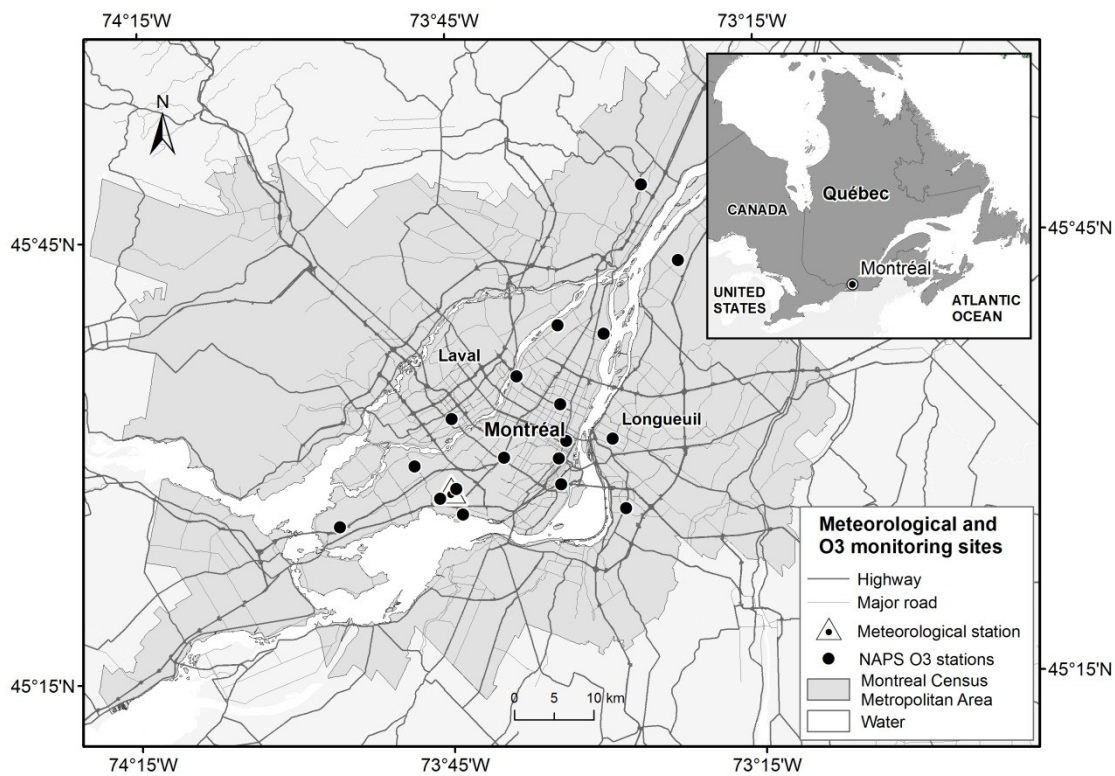
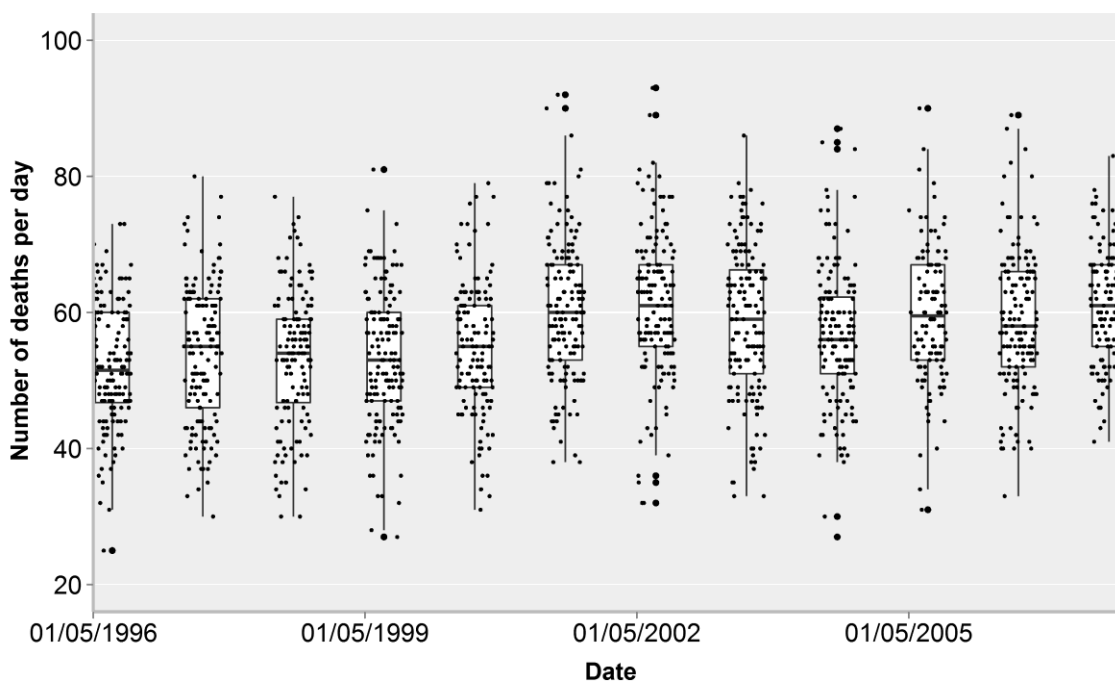


Figure 2. Daily number of all non-accidental deaths in Montreal CMA during summer months (May through September) from the years 1996 to 2007.







## **Chapitre 6 : Discussion et conclusions**



L'objectif de cette recherche était de développer et de valider un modèle spatio-temporel d'interpolation multivariable intégrant des variables auxiliaires (environnementales et/ou météorologiques) et permettant d'estimer spatialement et temporellement les concentrations journalières des  $PM_{2.5}$  et de l' $O_3$  entre les stations du RNSPA à Montréal, Canada. Ceci nous a amenés à proposer une méthodologie d'intégration de données basée sur la géostatistique, les SIG et la télédétection. Au meilleur de notre connaissance, ce genre d'approche n'avait été que peu utilisé auparavant pour la cartographie journalière des polluants atmosphériques tels que les  $PM_{2.5}$  et l' $O_3$  à l'échelle intra-urbaine. Pour mettre en relief les avantages de l'interpolation multivariable, une comparaison des résultats du modèle d'interpolation multivariable avec des modèles d'interpolation dits univariés, hybrides, et un modèle de prédiction spatio-temporel a été aussi réalisée. De plus, ce projet de recherche visait aussi à évaluer les associations entre les expositions journalières à court terme à l' $O_3$  et la mortalité basée sur différents modèles d'estimation des concentrations de polluants atmosphériques.

Dans le but de répondre aux objectifs spécifiques de cette thèse (chapitre 1, section 3.2), nous avons rédigé trois articles scientifiques; deux articles pour répondre au premier objectif (chapitres 3 et 4) et un article pour répondre au deuxième objectif (chapitre 5). Dans les prochains paragraphes, nous discutons de principaux résultats, de notre contribution à l'avancement des connaissances dans le domaine de la modélisation des polluants de l'air et des risques pour la santé, et nous présentons la conclusion générale de cette thèse.

Les modélisations réalisées dans le premier article (chapitre 3) ont permis d'estimer les concentrations journalières des  $PM_{2.5}$  à l'échelle intra-urbaine dans l'île de Montréal pour l'année 2010. L'approche d'interpolation hybride basée sur les modèles de krigeage avec dérive externe (KED) et de pondération inverse de la distance (IDW), appliquée de manière différenciée par groupes de conditions météorologiques synoptiques a amélioré les précisions des estimations journalières des  $PM_{2.5}$ . Ce modèle hybride, KED-IDW réduit l'erreur d'estimation des concentrations journalières des  $PM_{2.5}$  de  $2.61 \mu\text{g}/\text{m}^3$  à  $2.09 \mu\text{g}/\text{m}^3$  et améliore les  $R^2$  de 73% à 83% en comparaison avec la moyenne simple des concentrations des  $PM_{2.5}$  aux stations. En comparaison avec l'interpolation univariable (IDW), le KED-IDW réduit l'erreur d'estimation des concentrations journalières des  $PM_{2.5}$  de  $2.19 \mu\text{g}/\text{m}^3$  à  $2.09 \mu\text{g}/\text{m}^3$  et améliore les  $R^2$  de 81% à 83%. Cependant, ces résultats montrent que le gain apporté par KED-IDW est faible comparé avec l'interpolation univariable (IDW) dans le cas des  $PM_{2.5}$ . Ceci suggère que l'autocorrélation spatiale (dépendance spatiale) estimée à partir de quelques stations de mesures et la corrélation avec les variables auxiliaires utilisées n'ont pas été assez significatives pour tous les jours de la période d'étude. Dans un tel cas, une interpolation univariable ne tenant pas compte de l'autocorrélation spatiale comme IDW (interpolation géométrique basée sur une fonction mathématique) performe pratiquement aussi bien qu'une interpolation géostatistique multivariable.

Par ailleurs et d'après les coefficients de variation (CV) calculés à partir des mesures journalières des  $PM_{2.5}$  enregistrées aux stations du RNSPA de l'île de Montréal, la variation spatiale intra-urbaine par groupe synoptique se situe entre 23% et 36%. Le CV le plus élevé (36%) correspond à la classe synoptique 2 (jours répartis dans la saison d'hiver avec des vents

modérés à forts soufflant de l'ouest et du sud-ouest) et révèle des valeurs variables des concentrations des  $PM_{2.5}$ , c'est-à-dire avec faible autocorrélation spatiale. Ces valeurs variables des concentrations peuvent être dues aux  $PM_{2.5}$  d'origine primaire émise principalement par des sources locales et qui présentent une grande variabilité spatiale. Étant donné que le KED modélise la structure de la dépendance spatiale (par l'autocorrélation) de la variable dépendante en intégrant des informations de variables auxiliaires qui lui sont corrélées, une faible autocorrélation spatiale peut créer des incertitudes quant à la performance du modèle. Dans un tel cas, la performance de l'interpolation univariable (IDW) serait supérieure à celle de l'interpolation multivariable (KED). Malgré tout, se limiter à une interpolation univariable conduit à négliger la corrélation spatiale existante entre la variable dépendante et certaines variables auxiliaires analysées dans cette recherche. De plus, le coefficient de variation est faible (un certain degré d'homogénéité sur le plan spatial avec une autocorrélation élevée) lorsque des épisodes de forte concentration en  $PM_{2.5}$  caractérisent le groupe synoptique; CV = 23% (groupe synoptique 1) et CV = 26% (groupe synoptique 3).

Les épisodes de forte concentration de  $PM_{2.5}$  surviennent en hiver (classe synoptique 3) lors de conditions météorologiques montrant les caractéristiques typiques d'inversion de température (pression atmosphérique élevée et vents légers). Dans ce type de condition météorologique, les  $PM_{2.5}$  montrent une forte corrélation négative avec la vitesse du vent. Une corrélation négative signifie que de fortes concentrations des  $PM_{2.5}$  sont enregistrées lors de vents légers qui favorisent la stabilité atmosphérique. De forts épisodes de  $PM_{2.5}$  surviennent aussi lors de jours avec une humidité relative élevée (groupe synoptique 1) qui favorisent la formation du nitrate d'ammonium (composant principal semi-volatile des  $PM_{2.5}$ ). Les résultats

de ce premier article contredisent partialement la troisième hypothèse proposée au chapitre 1, section 2, i.e. les erreurs d'estimation d'une méthode d'interpolation multivariable sont moins grandes. On remarque cependant que les estimations produites par le modèle d'interpolation multivariable (KED) ont fourni les meilleures précisions pour la classe synoptique 3 comparativement aux autres classes, soit lors des jours où les  $PM_{2.5}$  montraient une structure de dépendance spatiale et de fortes corrélations avec la vitesse du vent (corrélation négative) et la densité de la végétation (corrélation négative). Cette performance particulière du modèle KED est cohérente avec les résultats des études réalisées par Pearce *et al.* (2009) et Beelen *et al.* (2009) qui ont obtenu de meilleurs résultats pour les estimations des  $PM_{2.5}$  avec le KED en comparaison avec les méthodes d'analyse de régression et de krigeage ordinaire (OK) pour des études réalisées à une échelle intra-urbaine et régionale respectivement. Ceci illustre bien l'importance que prend la corrélation avec les variables auxiliaires et la distribution du polluant atmosphérique ciblé. Par ailleurs, nous avons aussi montré que la relation entre les variables auxiliaires analysées dans cette étude et les  $PM_{2.5}$  change selon le régime climatique, comme l'hypothèse que nous avons émise. Ces résultats corroborent aussi la partie concernant les  $PM_{2.5}$  de la première hypothèse sur les relations avec les données météorologiques et autres variables de l'environnement soulevées au chapitre 1, section 2.

Les modélisations réalisées dans le deuxième article (chapitre 4) ont permis d'estimer les concentrations journalières d' $O_3$  à l'échelle intra-urbaine dans la région métropolitaine de Montréal pour l'année 2010. Le modèle d'interpolation KED développé dans cette recherche avec l'intégration de la densité de la végétation et la densité des routes comme variables prédictives a amélioré les précisions des estimations de l' $O_3$  pour les deux périodes

d'analyses : toute l'année par groupes de conditions météorologiques synoptiques et pour les mois d'été. Le KED réduit l'erreur d'estimation des concentrations journalières de l'O<sub>3</sub> de 4.75 ppb à 3.09 ppb et améliore les R<sup>2</sup> de 71% à 87% en comparaison avec la moyenne simple des concentrations d'O<sub>3</sub> aux stations. En comparaison avec la méthode d'interpolation univariable OK, le KED réduit l'erreur d'estimation des concentrations journalières de l'O<sub>3</sub> de 4.75 ppb à 3.88 ppb et améliore les R<sup>2</sup> de 75% à 87%. Ceci suggère que les erreurs d'estimation du modèle KED sont plus petites en comparaison avec la moyenne simple des concentrations d'O<sub>3</sub> aux stations et avec l'interpolation univariable (OK), apportant donc un gain dans les estimations de l'O<sub>3</sub>. Ces résultats sont cohérents avec les études réalisées par Beelen *et al.* (2009) qui ont obtenu de meilleurs résultats pour les estimations de l'O<sub>3</sub> avec le KED qu'avec le krigeage ordinaire (OK) pour des études réalisées à une échelle régionale. Ils ont obtenu un R<sup>2</sup> de 70% et une erreur d'estimation de 7.69 µg/m<sup>3</sup> (~ 4 ppb). La similitude des erreurs de leur KED avec celles de notre étude est surprenante, notamment en considérant la différence d'échelle entre les deux études: Beelen *et al.* (2009) pour une échelle régionale ont utilisé 224 stations pour une couverture de 2 854 116 km<sup>2</sup>, ce qui diffère fortement de l'échelle locale de notre étude avec ses 12 stations pour une couverture de 1 310 km<sup>2</sup>. À la vue de ces résultats, il semblerait pertinent de documenter davantage l'influence de l'échelle géographique sur la performance d'approches multivariées pour interpoler les niveaux d'ozone. Néanmoins, les résultats obtenus dans cette recherche confirment la troisième hypothèse formulée au chapitre 1, section 2 selon laquelle des erreurs d'estimation des méthodes basées sur la moyenne des stations, sur la station la plus proche, ou alors sur l'interpolation univariable (OK) sont plus grandes que celle de l'interpolation multivariable (KED).

De plus, les coefficients de variation (CV) calculés à partir des mesures journalières de l'O<sub>3</sub> enregistrées aux stations du RNSPA de l'île de Montréal montrent que la variation spatiale intra-urbaine de l'O<sub>3</sub> par groupe synoptique se situe entre 14% et 29%. Le CV le plus faible (autocorrélation élevée) correspond à la classe synoptique 5 (jours répartis principalement dans la saison du printemps) suivie par la classe synoptique 6 (jours répartis dans la saison d'été et d'automne) et révèle des concentrations élevées d'O<sub>3</sub>. Les variables auxiliaires telles que la densité de la végétation (corrélation positive) et la densité de routes (corrélation négative) montrent de fortes corrélations avec l'O<sub>3</sub> pour toutes les classes synoptiques. On constate toutefois que les corrélations varient peu entre classes météorologiques synoptiques. Cependant dans le cas de la densité de la végétation de la classe synoptique 1 (jours répartis dans l'année avec conditions d'humidité relative élevée et fortes précipitations), cette variation des corrélations est plus évidente et elle serait due au taux de précipitation et d'humidité relative élevée, deux facteurs météorologiques qui jouent un rôle important sur les niveaux des concentrations des COV biogéniques. Ces résultats confirment la deuxième et cinquième hypothèse formulée au chapitre 1, section 2, en lien avec les relations entre l'O<sub>3</sub> et différentes variables de l'environnement et avec la variation de ces relations selon les classes synoptiques.

Les gains de performance du KED relativement faibles rapportés dans les deux premiers articles (chapitres 3 et 4) sont expliqués en partie par la relative homogénéité sur le plan spatial des polluants d'origine secondaires (PM<sub>2.5</sub> et O<sub>3</sub>) et d'origine mixte (PM<sub>2.5</sub>). Par exemple, pour l'année 2010, à Montréal, la variation spatiale (la moyenne de l'écart-type journalier des concentrations à tous les sites de surveillance) des PM<sub>2.5</sub> a été de 2.36 µg/m<sup>3</sup> et



la variation temporelle (l'écart-type des moyennes journalières des concentrations sur tous les jours) de  $6.24 \mu\text{g}/\text{m}^3$  et pour l' $\text{O}_3$ , elle a été de 4.23 ppb et de 9.88 ppb respectivement. Ceci montre que tant les  $\text{PM}_{2.5}$  que l' $\text{O}_3$  ont des variations spatiales relativement faibles par rapport à leurs variations temporelles (d'environ la moitié de l'amplitude). Ceci montre aussi que ces polluants sont relativement homogènes sur le plan spatial (à travers l'île de Montréal pour les  $\text{PM}_{2.5}$  et à travers les neuf subdivisions de la Région métropolitaine de recensement de Montréal pour l' $\text{O}_3$ ) et qu'ils présentent donc de la dépendance spatiale. Cette relation entre la variabilité et la dépendance spatiale est en accord avec les résultats de l'étude de Sororian *et al.* (2014) qui ont analysé les données de qualité de l'air ambiant provenant des stations fixes et celles provenant d'un modèle de prévision déterministe en Georgie, (États-Unis). Sororian *et al.* (2014) ont constaté que les polluants d'origine secondaire comme les  $\text{PM}_{2.5}$  et  $\text{O}_3$  et d'origine mixte comme les  $\text{PM}_{2.5}$  sont relativement homogènes sur le plan spatial et présentent une autocorrélation élevée. Cette homogénéité sur le plan spatial de l' $\text{O}_3$  est aussi cohérente avec l'étude de D. Lee *et al.* (2010), qui en se basant sur des mesures de 23 stations fixes de la région métropolitaine de Londres ont rapporté une variabilité spatiale égale à 9.80 ppb et une variabilité temporelle égale à 19.40 ppb.

En général, les travaux de recherche rapportés dans les deux premiers articles de cette thèse (chapitres 3 et 4) en se basant sur des mesures météorologiques journalières de la RMR de Montréal (données enregistrées entre 2008 et 2010) ont permis d'identifier six classes synoptiques, mettant ainsi en lumière l'influence des principaux facteurs météorologiques dans les variations des  $\text{PM}_{2.5}$  et de l' $\text{O}_3$ . Cette influence sur les niveaux des polluants atmosphériques est bien connue (Dawson *et al.*, 2007; Gibson *et al.*, 2009; Malek *et al.*, 2006;

Tai *et al.*, 2010; Tai *et al.*, 2012), cependant, nous avons pu en particulier démontrer le rôle des conditions météorologiques synoptiques dans les variations spatiales intra-urbain et par conséquent, leur influence sur la performance des différents interpolateurs. Nous avons pu aussi démontrer que le niveau de corrélation entre les variables environnementales et les PM<sub>2.5</sub> et l'O<sub>3</sub> change selon les conditions météorologiques synoptiques, que l'exactitude des prédictions (R<sup>2</sup> et RMSE) des modèles varie selon le groupe synoptique, et qu'une meilleure prise en compte des facteurs météorologiques est donc importante lors de l'estimation journalière des concentrations des PM<sub>2.5</sub> et de l'O<sub>3</sub>. Puisque les conditions météorologiques synoptiques ont rarement été prises en compte dans le cadre de la modélisation pour l'estimation des polluants, nos travaux de recherche pourront contribuer à une amélioration stratégique de la modélisation des polluants atmosphériques, en considérant certains facteurs météorologiques comme prédictifs ou comme indicateurs de la meilleure méthode d'interpolation spatiale à appliquer. Ceci confirme la cinquième hypothèse formulée au chapitre 1, section 2 au sujet de l'influence du régime climatique dans le niveau de corrélation entre les variables auxiliaires et les concentrations des PM<sub>2.5</sub> et de l'O<sub>3</sub>.

De plus, les analyses des deux premiers articles (chapitres 3 et 4) ont aussi mis en évidence que les erreurs des modèles étaient plus élevées dans des endroits où les stations de mesures étaient plus éparées (Figure 5 des chapitres 3 et 4). Ceci est en accord avec l'étude de Yu *et al.* (2009) qui a montré des erreurs plus élevées dans les endroits avec des mesures éparées. Des améliorations dans la modélisation pourraient inclure l'installation temporaire de quelques stations de surveillance dans les zones faiblement couvertes.

Dans l'article 3 de cette thèse (chapitre 5) nous avons exploré les associations entre les décès et l'exposition à court terme à l'O<sub>3</sub> entre les années 1996 à 2007. Les résultats de cette analyse montrent des associations entre les décès (causes non accidentelle, cardio-vasculaire et cardio-respiratoire) au délai « *lag* » 1 (exposition ayant eu lieu un jour avant le décès) et l'exposition à l'O<sub>3</sub>. Les associations par incrément d'interquartile (IQR) selon la cause de décès sont similaires pour les trois approches (la moyenne sur l'ensemble des stations de surveillance, la station de mesure la plus proche et le modèle BME) d'estimations d'O<sub>3</sub> utilisées dans cette étude. Ces résultats montrent que le modèle spatio-temporel utilisé n'apporte pas de gain dans l'estimation de l'effet de l'exposition à l'O<sub>3</sub> par rapport aux deux autres méthodes d'estimation utilisées dans cette recherche. Ceci est probablement dû aux erreurs de prédictions importantes du BME, comme nous l'avons aussi montré à l'article 2 (chapitre 4). Les résultats de cet article contredisent partiellement la quatrième hypothèse formulée au chapitre 1, section 2 selon laquelle les estimations de polluants provenant d'un modèle spatio-temporel expliquent le mieux les effets sur la santé en comparaison avec la méthode d'estimation basée sur la moyenne des stations ou sur la station la plus proche. Par ailleurs, et à notre connaissance, dans la littérature scientifique, il n'existe pas d'études ayant exploré les associations entre les décès et l'exposition à l'O<sub>3</sub> avec différentes méthodes d'estimations des concentrations de polluants atmosphériques. Ceci pourrait justifier l'initiation de nouvelles recherches avec l'intégration de modèles d'interpolation multivariable comme le KED pour une période de plusieurs années et dans les endroits où il existe un nombre minime des mesures journalières par zone d'étude.

En conclusion, cette thèse a permis d'estimer les concentrations journalières des  $PM_{2.5}$  et de l' $O_3$  à l'échelle intra-urbaine en utilisant un nombre limité des stations de mesure fixe en conditions environnementales et météorologiques spécifiques. Les résultats montrent que la précision de la modélisation des concentrations journalières des  $PM_{2.5}$  et d' $O_3$  varie selon la condition météorologique synoptique.

Le modèle hybride KED-IDW développé dans cette recherche pour estimer les concentrations des  $PM_{2.5}$  fournit de meilleures estimations journalières par l'intégration de variables auxiliaires lorsqu'elles lui sont corrélées. Cependant, KED-IDW améliore peu les estimations des  $PM_{2.5}$  en comparaison avec KED, IDW et la moyenne sur l'ensemble des stations de surveillance.

Pour les estimations journalières de l' $O_3$ , c'est le modèle d'interpolation multivariable (KED) développé dans cette recherche avec l'intégration de variables auxiliaires qui a fourni les meilleures précisions en comparaison à l'interpolation univariable (OK), la moyenne sur l'ensemble des stations de surveillance et le modèle spatio-temporel de prédiction, BME.

Dans le contexte de l'estimation de l'exposition des populations, le modèle hybride KED-IDW (pour les  $PM_{2.5}$ ) et le modèle KED (pour l' $O_3$ ) avec l'intégration des variables prédictives (environnementales et météorologiques) identifiées dans cette thèse contribueront sans doute à produire de meilleures estimations journalières en  $PM_{2.5}$  et en  $O_3$ , dans des lieux se situant entre les stations du RNSPA.

Les études concernant les modèles spatio-temporels et l'intégration de variables prédictives (spatiales et spatio-temporels) dans les estimations de polluants atmosphériques sont encore limitées, et leur application dans l'estimation de l'effet de l'exposition aux polluants est encore plus rare. Dans ce contexte, cette thèse constitue un apport important dans le processus d'estimation des concentrations des  $PM_{2.5}$  et d' $O_3$  à l'échelle intra-urbaine et une option qui s'avère peu coûteuse avec l'utilisation des mesures journalières provenant des stations fixes du RNSPA. De plus, la comparaison et l'évaluation entre méthodes d'estimation constituent une référence pour la modélisation de polluants atmosphériques en zones urbaines et périurbaines. Le modèle d'interpolation multivariable développé dans cette recherche fournit des estimations des concentrations journalières de l' $O_3$  avec une précision suffisante et à l'échelle intra-urbaine, toutefois il peut être amélioré par l'intégration des variables géographiques-environnementales avec une meilleure résolution temporelle et des variables météorologiques avec une meilleure résolution spatiale. Enfin, cette recherche sera très utile à d'autres investigations futures concernant les modèles spatio-temporels avec l'ajout de variables auxiliaires spatialisés en voie de meilleures estimations des polluants atmosphériques. Ceci constitue une importante contribution non seulement pour une meilleure caractérisation de l'exposition et de l'évaluation des risques à la santé de la population, sinon aussi pour d'autres applications concernant l'environnement urbain.

## Bibliographie

- Adam-Poupart, A., Brand, A., Fournier, M., Jerrett, M., and Smargiassi, A. (2014). Spatiotemporal Modeling of Ozone Levels in Quebec (Canada): A Comparison of Kriging, Land-Use Regression (LUR), and Combined Bayesian Maximum Entropy–LUR Approaches. *Environmental Health Perspectives*, 122(970–976).
- Adam-Poupart, A., Brand, A., Fournier, M., and Smargiassi, A. (2013). *Estimation de l'exposition environnementale à l'ozone troposphérique: un exemple de modélisation pour la population québécoise*. Institut National de Santé Publique du Québec (INSPQ).
- Adam-Poupart, A., Labrèche, F., Busque, M.-A., Brand, A., Duguay, P., Fournier, M., Zayed, J., and Smargiassi, A. (2015). Association between outdoor ozone and compensated acute respiratory diseases among workers in Quebec (Canada). *Industrial Health*, 2015(53), 171–175.
- Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., De Hoogh, K., and Briggs, D. J. (2009). Mapping of background air pollution at a fine spatial scale across the European Union. *Science of the total environment*, 470, 1852–1867.

- Bell, M. L., Dominici, F., Ebisu, K., Zeger, S. L., and Samet, J. M. (2007). Spatial and temporal variation in PM<sub>2.5</sub> chemical composition in the United States for health effects studies. *Environmental Health Perspectives*, 115(9), 989–995.
- Bell, M. L., Dominici, F., and Samet, J. M. (2005). A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity, mortality, and air pollution study. *Epidemiology*, 16(4), 436–445.
- Bell, M. L., Son, J.-Y., Peng, R. D., Wang, Y., and Dominici, C. (2016). Ambient PM<sub>2.5</sub> and risk of hospital admissions: Do risks differ for men and women? *Epidemiology*, 26(4), 575–579.
- Bogaert, P., Christakos, G., Jerrett, M., and Yu, H.-L. (2009). Spatiotemporal modelling of ozone distribution in the State of California. *Atmospheric Environment*, 43, 2471–2480.
- Brauer, M., Freedman, G., Frostad, J., van Donkelaar, A., Martin, R., Dentener, F., van Dingenen, R., Estep, K. V., Amini, H., Apte, J. S., Balakrishnan, K., Barregard, L., Broday, D., Feigin, V., Ghosh, S., Hopke, P. K., Knibbs, L. D., Kokubo, Y., Liu, Y., Ma, S., Morawska, L., Texcalac, J. L., Shaddick, G., Anderson, H. R., Vos, T., Forouzanfar, M. H., Burnett, R. T., and Cohen, A. (2016). Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013. *Environmental Science and Technology*, 50, 79–88.

- Brauer, M., Hystad, P., and Poplawski, K. (2014). *Assessing the spatial representativeness of PM<sub>2.5</sub> and O<sub>3</sub> measurements from the National air pollutant surveillance system*. School of Environmental Health and School of Population and Public Health, University of British Columbia.
- Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., and Karr, C. (2008). A cohort study of traffic-related air pollution impacts on birth outcomes. *Environmental Health Perspectives*, 116(5), 680–686.
- Briggs, D. (2005). The role of GIS: Coping with space (and time) in air pollution exposure assessment. *Journal of Toxicology and Environmental Health, Part A*, 68, 1243–1261.
- Camalier, L., Cox, W., and Dolwick, P. (2007). The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmospheric Environment*, 41, 7127–7137.
- CCME. (2011). *Protocole de Surveillance de la Qualité de l’Air Ambiant Relatif aux PM<sub>2,5</sub> et à l’Ozone*. Conseil Canadien des Ministres de l’Environnement.
- Chen, C., Wu, C., Yu, H., Chan, C., and Cheng, T. (2012). Spatiotemporal modeling with temporal-invariant variogram subgroups to estimate fine particulate matter PM<sub>2.5</sub> concentrations. *Atmospheric Environment*, 54, 1–8.



- Chen, L., Mengersen, K., and Tong, S. (2007). Spatiotemporal relationship between particle air pollution and respiratory emergency hospital admissions in Brisbane, Australia. *Science of the total environment*, 373, 57–67.
- Chiles, J. P. (1991). Application du krigeage avec dérive externe à l'implantation d'un réseau de surveillance piézométrique. *Sciences de la Terre*, 30, 131–147.
- Chu, D. A., Kaufman, Y. J., Zibordi, G., Chern, J. D., Mao, J., Li, C., and Holben, B. N. (2003). Global monitoring of air pollution over land from the Earth Observing System -Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research*, 108(D21), 1–18.
- Colls, J. (2002). *Air Pollution* (2nd ed.). New York: Spon Press.
- Cox, W., and Chu, S.-H. (1996). Assessment of interannual ozone variation in urban areas from a climatological perspective. *Atmospheric Environment*, 30(14), 2615–2625.
- Crouse, D. L., Goldberg, M. S., and Ross, N. A. (2009). A prediction-based approach to modelling temporal and spatial variability of traffic-related air pollution in Montreal, Canada. *Atmospheric Environment*, 43, 5075–5084.
- Dawson, J. P., Adams, P. J., and Pandis, S. N. (2007). Sensitivity of PM<sub>2.5</sub> to climate in the Eastern US: a modeling case study. *Atmospheric Chemistry and Physics*, 7, 4295–4309.

- Dominici, F., Daniels, M., Zeger, S. L., and Samet, J. M. (2002). Air pollution and mortality: Estimating regional and national dose-response relationships. *Journal of the American Statistical Association*, 97(457), 100–111.
- Environment Canada. (2013a). Polluants atmosphériques. Retrieved Juin, 2013, from <http://www.ec.gc.ca/Air/default.asp?lang=Fr&n=BCC0B44A-1>
- Environment Canada. (2013b). Inventaire national des rejets de polluants - Sommaires nationaux des émissions des principaux polluants atmosphériques, 2013. Retrieved December, 2015, from <http://ec.gc.ca/inrp-npri/donnees-data/ap/index.cfm?lang=Fr>
- Environment Canada. (2013c). *10 years of data from the national air pollution surveillance (NAPS) network*. En49-2/7-40-PDF.
- Feng, X., and Wang, S. (2012). Influence of different weather events on concentrations of particulate matter with different sizes in Lanzhou, China. *Journal of Environmental Sciences*, 24(4), 665–674.
- Forouzanfar, M. H., Alexander, L., Anderson, H. R., Bachman, V., F.; Biryukov, S., Brauer, M., Burnett, R., Casey, D., Coates, M. M., and Cohen, A. (2015). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*.

- Franklin, M., Zeka, A., and Schwartz, J. (2007). Association between PM<sub>2.5</sub> and all-cause and specific-cause mortality in 27 US communities. *Journal of Exposure Science and Environmental Epidemiology*, 17, 279–287.
- Fuks, K. B., Weinmayr, G., Basagaña, X., Gruzieva, O., Hampel, R., Oftedal, B., Sørensen, M., Wolf, K., Aamodt, G., Aasvang, G. M., Aguilera, I., Becker, T., Beelen, R., Brunekreef, B., Caracciolo, B., Cyrus, J., Elosua, R., Eriksen, K. T., Foraster, M., Fratiglioni, L., Hilding, A., Houthuijs, D., Korek, M., Künzli, N., Marrugat, J., Nieuwenhuijsen, M., Östenson, C.-G., Penell, J., Pershagen, G., Raaschou-Nielsen, O., Swart, W. J., Peters, A., and Hoffmann, B. (2016). Long-term exposure to ambient air pollution and traffic noise and incident hypertension in seven cohorts of the European study of cohorts for air pollution effects (ESCAPE). *European Heart Journal*, 1–8.
- Gehring, U., Gruzieva, O., Agius, R. M., Beelen, R., Custovic, A., Cyrus, J., Eeftens, M., Flexeder, C., Fuertes, E., Heinrich, J., Hoffmann, B., De Jongste, J. C., Kerkhof, M., Klümper, C., Korek, M., Mölter, A., Schultz, E. S., Simpson, A., Sugiri, D., Svartengren, M., Von Berg, A., Wijga, A. H., Pershagen, G., and Brunekreef, B. (2013). Air pollution exposure and lung function in children: the ESCAPE project. *Environmental Health Perspectives*, 121(11–12), 1357-1364.
- Gibson, M. D., Guernsey, J. R., Beauchamp, S., Waugh, D., Heal, M. R., Brook, J. R., Maher, R., Gagnon, G. A., McPherson, J. P., Bryden, B., Gould, R., and Terashima, M. (2009). Quantifying the spatial and temporal variation of ground-level ozone in the rural

- Annapolis Valley, Nova Scotia, Canada using nitrite-impregnated passive samplers. *Journal of the Air & Waste Management Association*, 59, 310–320.
- Goldberg, M. S., Burnett, R. T., Brook, J., Bailar III, J. C., Valois, M.-F., and Renaud, V. (2001). Associations between daily cause-specific mortality and concentrations of ground-level ozone in Montreal, Quebec. *Am J Epidemiol*, 154(9), 817–826.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York, USA: Oxford University Press.
- Gulliver, J., De Hoogh, K., Fecht, D., Vienneau, D., and Briggs, D. (2011). Comparative assessment of GIS-based methods and metrics for estimating long-term exposures to air pollution. *Atmospheric Environment*, 45, 7072–7080.
- Hansell, A., Rose, N., Cowie, C. T., Belousova, E. G., Bakolis, I., Ng, K., Toelle, B. G., and Marks, G. B. (2014). Weighted road density and allergic disease in children at high risk of developing asthma. *PLoS One*, 9(6), 9.
- Hart, J. E., Yanosky, J. D., Puett, R. C., Ryan, L., Dockery, D. W., Smith, T. J., Garshick, E., and Laden, F. (2009). Spatial Modeling of PM10 and NO2 in the Continental United States, 1985–2000. *Environmental Health Perspectives*, 117(11).
- Henderson, S., Beckerman, B., Jerrett, M., and Brauer, M. (2007). Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science and Technology*, 41, 2422–2428.

- Hewitt, C. N., and Jackson, A. V. (2009). *Atmospheric Science for Environmental Scientists* (pp 314). Chichester: Wiley-Blackwell.
- Hoek, G., Beelen, R., De Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., and Briggs, D. J. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42, 7561–7578.
- Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P., and Van den Brandt, P. A. (2002). Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *The Lancet*, 360, 1203–1209.
- Hou, X., Zhu, B., Fei, D., and Wang, D. (2015). The impacts of summermonsoons on the ozone budget of the atmospheric boundary layer of the Asia-Pacific region. *Science of the total environment*, 502, 641–649.
- Hufty, A. (1982). Analyse en composantes principales des situations synoptiques au Québec. *Géographie physique et Quaternaire*, XXXVI(3), 307–314.
- INSPQ. (2012). *Bilan de la qualité de l'air au Québec en lien avec la santé, 1975-2009*. Institut National de Santé Publique du Québec - Ministère du Développement durable, de l'Environnement et des Parcs, Québec.
- IPCC. (2014). *Climate Change 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.* Intergovernmental Panel on Climate Change, C. U. Press, New York, USA.

Ito, K., Xue, N., and Thurston, G. (2004). Spatial variation of PM<sub>2.5</sub> chemical species and source-apportioned mass concentrations in New York City. *Atmospheric Environment*, 38, 5269–5282.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., and Giovis, C. (2005). A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol*, 15, 185–204.

Jerrett, M., Burnett, R. T., Pope III, C. A., Ito, K., Thurston, G., Krewski, D., Shi, Y., Calle, E., and Thun, M. (2009). Long-term ozone exposure and mortality. *The New England Journal of Medicine*, 360(11), 1085–1095.

Leduc, R., and Roy-Loubier, K. (2005). Ozone dans le Sud du Québec : origine et trajectoire. *Vecteur Environnement*(38), 44–54.

Lee, D., and Shaddick, G. (2010). Spatial modeling of air pollution in studies of its short-term health effect. *Biometrics*, 66, 1238–1246.

Lee, S.-J., Serre, M. L., Van Donkelaar, A., Martin, R. V., Burnett, R. T., and Jerrett, M. (2012). Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM<sub>2.5</sub> concentrations across the continental United States. *Environmental Health Perspectives*, 120(12), 1727–1732.

- Liu, L. J. S., and Rossini, A. J. (1996). Use of kriging models to predict 12-hour mean ozone concentrations in metropolitan Toronto - A pilot study. *Environment International*, 22(6), 677–692.
- Malek, E., Davis, T., Martin, R. S., and Silva, P. J. (2006). Meteorological and environmental aspects of one of the worst national air pollution episodes (January, 2004) in Logan, Cache Valley, Utah, USA. *Atmospheric Research*, 79, 108–122.
- Markevych, I., Fuertes, E., Tiesler, C., Birk, M., Bauer, C., Koletzko, S., Von Berg, A., Berdel, D., and Heinrich, J. (2014). Surrounding greenness and birth weight: results from the GINIplus and LISApplus birth cohorts in Munich. *Health Place*, 26, 39–46.
- Matheron, G. (1969). Le krigeage universel. In ENSMP (Ed.), *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau* (Vol. 1, pp. 83): École Nationale Supérieure des Mines de Paris.
- Matthes, S., Grewe, V., Sausen, R., and Roelofs, G.-J. (2007). Global impact of road traffic emissions on tropospheric ozone. *Atmospheric Chemistry and Physics*, 7, 1707–1718.
- McKendry, I. G. (1993). Ground-level ozone in Montreal, Canada. *Atmospheric Environment*, 27B, 93–103.
- McMurry, P. H., Shepherd, M. F., and Vickery, J. S. (2004). *Particulate Matter Science for Policy Makers: A NARSTO Assessment*. Cambridge University Press.

- Miller, K. A., Siscovick, D. S., Sheppard, L., Shepherd, K., Sullivan, J. H., Anderson, G. L., and Kaufman, J. D. (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *The New England Journal of Medicine*, 356(5), 447–458.
- Miller, L., Lemke, L. D., Xu, X., Molaroni, S. M., You, H., Wheeler, A. J., Booza, J., Grgicak-Mannion, A., Krajenta, R., Graniero, P., Krouse, H., Lamerato, L., Raymondh, D., Reiners Jr, J., and Weglicki, L. (2010). Intra-urban correlation and spatial variability of air toxics across an international airshed in Detroit, Michigan (USA) and Windsor, Ontario (Canada). *Atmospheric Environment*, 44, 1162–1174.
- Nowak, D. (2002). *The effects of urban trees on air quality*. USDA Forest Service, Syracuse, NY.
- OMS. (2010). *La Face Cachée des Villes : Mettre au Jour et Vaincre les Inégalités en Santé en Milieu Urbain*. Organisation Mondiale de la Santé, Genève.
- Paradis, J. (2011). *Inventaire des émissions des principaux contaminants atmosphériques au Québec en 2008 et évolution depuis 1990*. Ministère du Développement durable, de l'Environnement et des Parcs, Québec.
- Pearce, J. L., Rathbun, S. L., Aguilar-Villalobos, M., and Naeher, L. P. (2009). Characterizing the spatiotemporal variability of PM<sub>2.5</sub> in Cusco, Peru using kriging with external drift. *Atmospheric Environment*, 43, 2060–2069.



- Pinto, J. P., Lefohn, A. S., and Shadwick, D. S. (2004). Spatial variability of PM<sub>2.5</sub> in urban areas in the United States. *Air & Waste Management Association*, 54, 440–449.
- Pope III, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association*, 287, 1132–1141.
- Pope III, C. A., Burnett, R. T., Thurston, G. D., Thun, M. J., Calle, E. E., Krewski, D., and Godleski, J. J. (2004). Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. *Circulation*, 2109, 71–77.
- Rose, N., Cowie, C., Gillett, R., and Marks, G. (2009). Weighted road density: a simple way of assigning traffic-related air pollution exposure. *Atmospheric Environment*, 43, 5009–5014.
- Ross, Z., Jerrett, M., Ito, K., Tempalski, B., and Thurston, G. (2007). A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment*, 41, 2255–2269.
- Seinfeld, J. H. (1989). Urban air pollution: State of the science. *Science*, 243(4892), 745–752.
- Seinfeld, J. H., and Pandis, S. N. (2006). *Atmospheric chemistry and physics: from air pollution to climate change* (pp 1232). New York, US: Wiley-Blackwell.

- Singh, V., Carnevale, C., Finzi, G., Pisoni, E., and Volta, M. (2011). A cokriging based approach to reconstruct air pollution maps, processing measurement station concentrations and deterministic model simulations. *Environmental Modelling & Software*, 26, 778e786.
- Sororian, S. A., Holmes, H. A., Friberg, M., Ivey, C., Hu, Y., Mulholland, J. A., Russell, A. G., and Strickland, M. J. (2014). Temporally and spatially resolved air pollution in Georgia using fused ambient monitor data and chemical transport model results. In D. Steyn & R. Mathur (Eds.), *Air Pollution Modeling and its Application XXIII* (pp. 301–306). Switzerland: Springer International Publishing.
- Stafoggia, M., G., C., Peters, A., Andersen, Z. J., Badaloni, C., Beelen, R., Caracciolo, B., Cyrus, J., De Faire, U., De Hoogh, K., Eriksen, K. T., Fratiglioni, L., Galassi, C., Gigante, B., Havulinna, A. S., Hennig, F., Hilding, A., Hoek, G., Hoffmann, B., Houthuijs, D., Korek, M., Lanki, T., Leander, K., Magnusson, P. K., Meisinger, C., Migliore, E., Overvad, K., Östenson, C.-G., Pedersen, N. L., Pekkanen, J., Penell, J., Pershagen, G., Pundt, N., Pyko, A., Raaschou-Nielsen, O., Ranzi, A., Ricceri, F., Sacerdote, C., Swart, W. J. R., Turunen, A. W., Vineis, P., Weimar, C., Weinmayr, G., Wolf, K., Brunekreef, B., and Forastiere, F. (2014). Long-term exposure to ambient air pollution and incidence of cerebrovascular events: results from 11 european cohorts within the ESCAPE project. *Environmental Health Perspectives*, 122(9), 919–925.

- Stieb, D. M., Szyszkowicz, M., Rowe, B. H., and Leech, J. A. (2009). Air pollution and emergency department visits for cardiac and respiratory conditions: a multi-city time-series analysis. *Environmental Health*, 8(25), 1–13.
- Strickland, M. J., Darrow, L. A., Klein, M., Flanders, W. D., Sarnat, J. A., Waller, L. A., Sarnat, S. E., Mulholland, J. A., and Tolbert, P. E. (2010). Short-term associations between ambient air pollutants and pediatric asthma emergency department visits. *Am J Respir Crit Care Med*, 182, 307–316.
- Strickland, M. J., Darrow, L. A., Mulholland, J. A., Klein, M., Flanders, W. D., Winqvist, A., and Tolbert, P. E. (2011). Implications of different approaches for characterizing ambient air pollutant concentrations within the urban airshed for time-series studies and health benefits analyses. *Environmental Health*, 10(36), 1–9.
- Strickland, M. J., Gass, K. M., Goldman, G. T., and Mulholland, J. A. (2015). Effects of ambient air pollution measurement error on health effect estimates in time-series studies: a simulation-based analysis. *Journal of Exposure Science and Environmental Epidemiology*, 25, 160–166.
- Su, J., Jerrett, M., Beckerman, B., Verma, D., Arain, M. A., Kanaroglou, P., Stieb, D., Finkelstein, M., and Brook, J. (2010). A land use regression model for predicting ambient volatile organic compound concentrations in Toronto, Canada. *Atmospheric Environment*, 44(3529–3537).

- Su, J. G., Jerrett, M., and Beckerman, B. (2009). A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *Science of the total environment*, 407(12), 3890–3898.
- Tai, A. P. K., Mickley, L. J., and Jacob, D. J. (2010). Correlations between fine particulate matter (PM<sub>2.5</sub>) and meteorological variables in the United States: Implications for the sensitivity of PM<sub>2.5</sub> to climate change. *Atmospheric Environment*, 44, 3976–3984.
- Tai, A. P. K., Mickley, L. J., Jacob, D. J., Leibensperger, E. M., Zhang, L., Fisher, J. A., and Pye, H. O. T. (2012). Meteorological modes of variability for fine particulate matter (PM<sub>2.5</sub>) air quality in the United States: implications for PM<sub>2.5</sub> sensitivity to climate change. *Atmospheric Chemistry and Physics*, 12, 3131–3145.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (supplement), 234–240.
- U.S. EPA. (2004). *Air Quality Criteria for Particulate Matter (Final Report)* U.S. Environmental Protection Agency, Washington, D.C., EPA 600/P-99/002aF-bF, 2004.
- U.S. EPA. (2009). *Final Report: Integrated Science Assessment for Particulate Matter*. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-08/139F, 2009.
- U.S. EPA. (2013). *Final Report: Integrated Science Assessment for Ozone and Related Photochemical Oxidants*. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-10/076F, 2013.

- Van Donkelaar, A., Martin, R. V., and Park, R. J. (2006). Estimating ground-level PM<sub>2.5</sub> using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research*, *111*(D21), 1–10.
- Varns, J., Sather, M., Glen, G., Smith, L., and Stallings, C. (2001). Passive ozone network of Dallas: a modeling opportunity with community involvement. *Environmental Science and Technology*, *35*, 845–855.
- Ville de Montréal. (2010). *Bilan environnemental 2010: Qualité de l'air à Montréal*. Ville de Montréal, Montréal.
- Ville de Montréal. (2016). *Arbres publics sur le territoire de la Ville Montréal*, Octobre.
- Von Klot, S., Peters, A., Aalto, P., Bellander, T., Berglind, T., D'Ippoliti, D., Elosua, R., Hörmann, A., Kulmala, M., Lanki, T., Löwel, H., Pekkanen, J., Picciotto, S., Sunyer, J., and Forastiere, F. (2005). Ambient air pollution is associated with increased risk of hospital cardiac readmissions of myocardial infarction survivors in five European cities. *Circulation Research*, *112*(3073–3079).
- Wackernagel, H. (2003). *Multivariate geostatistics : an introduction with applications* (Springer-Verlag Ed.).
- WHO. (2014). *Burden of Disease from Household Air Pollution for 2012*. World Health Organization, Geneva.

- WHO. (2015). *Reducing global health risk through mitigation of short-lived climate pollutants*. World Health Organization, Geneva.
- Wise, E. K., and Comrie, A. C. (2005). Meteorologically adjusted urban air quality trends in the Southwestern United States. *Atmospheric Environment*, 29, 2969–2980.
- Yang, Y., and Christakos, G. (2015). Spatiotemporal Characterization of Ambient PM<sub>2.5</sub> Concentrations in Shandong Province (China). *Environmental Science and Technology*, 49, 13431–13438.
- Yu, H.-L., Chen, J.-C., Christakos, G., and Jerrett, M. (2009). BME estimation of residential exposure to ambient PM<sub>10</sub> and ozone at multiple time scales. *Environmental Health Perspectives*, 117(4), 537–544.
- Zanobetti, A., and Schwartz, J. (2006). Air pollution and emergency admissions in Boston, MA. *Journal of Epidemiology and Community Health*, 60, 890–895.
- Zanobetti, A., and Schwartz, J. (2008). Mortality displacement in the association of ozone with mortality: An analysis of 48 cities in the United States. *Am J Respir Crit Care Med*, 177, 184–189.
- Zanobetti, A., and Schwartz, J. (2009). The Effect of Fine and Coarse Particulate Air Pollution on Mortality: A National Analysis. *Environmental Health Perspectives*, 117(6), 898–903.

## **Annexe 1: Approbation éthique**

24 février 2015

Objet: Approbation éthique – « Mise en relation de la pollution atmosphérique et des épisodes d'asthme par modélisation spatiale: Développement d'approches géostatistiques à l'échelle urbaine pour l'estimation de l'exposition aux particules fines et à l'ozone troposphérique »

Mme Yuddy Ramos Yngaroca,

Le Comité d'éthique de la recherche en santé (CERES) a étudié le projet de recherche susmentionné et a délivré le certificat d'éthique demandé suite à la satisfaction des exigences précédemment émises. Vous trouverez ci-joint une copie numérisée de votre certificat; copie également envoyée à votre directeur/directrice de recherche et à la technicienne en gestion de dossiers étudiants (TGDE) de votre département.

Notez qu'il y apparaît une mention relative à un suivi annuel et que le certificat comporte une date de fin de validité. En effet, afin de répondre aux exigences éthiques en vigueur au Canada et à l'Université de Montréal, nous devons exercer un suivi annuel auprès des chercheurs et étudiants-chercheurs.

De manière à rendre ce processus le plus simple possible et afin d'en tirer pour tous le plus grand profit, nous avons élaboré un court questionnaire qui vous permettra à la fois de satisfaire aux exigences du suivi et de nous faire part de vos commentaires et de vos besoins en matière d'éthique en cours de recherche. Ce questionnaire de suivi devra être rempli annuellement jusqu'à la fin du projet et pourra nous être retourné par courriel. La validité de l'approbation éthique est conditionnelle à ce suivi. Sur réception du dernier rapport de suivi en fin de projet, votre dossier sera clos.

Il est entendu que cela ne modifie en rien l'obligation pour le chercheur, tel qu'indiqué sur le certificat d'éthique, de signaler au CERES tout incident grave dès qu'il survient ou de lui faire part de tout changement anticipé au protocole de recherche.

Nous vous prions d'agréer, Madame, l'expression de nos sentiments les meilleurs,

Dominique Langelier, présidente  
Comité d'éthique de la recherche en santé (CERES)  
Université de Montréal

DL/GP/gp

c.c. Gestion des certificats, BRDV

Benoît St-Onge, professeur associé, Faculté des arts et Sciences - Département de géographie

Audrey Smargiassi, professeure agrégée, École de santé publique - Département de santé environnementale et santé au travail

Fanny Duval (Géographie)

p.j. Certificat #15-014-CERES-D

**adresse postale**

C.P. 6128, succ. Centre-ville  
Montréal QC H3C 3J7

3744 Jean-Brillant  
4e étage, bur. 430-11  
Montréal QC H3T 1P1

Téléphone : 514-343-6111 poste 2604  
ceres@umontreal.ca  
www.ceres.umontreal.ca



Comité d'éthique de la recherche en santé

### CERTIFICAT D'APPROBATION ÉTHIQUE

*Le Comité d'éthique de la recherche en santé (CERES), selon les procédures en vigueur, en vertu des documents qui lui ont été fournis, a examiné le projet de recherche suivant et conclu qu'il respecte les règles d'éthique énoncées dans la Politique sur la recherche avec des êtres humains de l'Université de Montréal.*

Projet	
<b>Titre du projet</b>	<b>Mise en relation de la pollution atmosphérique et des épisodes d'asthme par modélisation spatiale: Développement d'approches géostatistiques à l'échelle urbaine pour l'estimation de l'exposition aux particules fines et à l'ozone troposphérique</b>
<b>Étudiante requérante</b>	<b>Yuddy Ramos Yngaroca</b> [redacted] Candidate au Ph.D. en géographie, Faculté des arts et sciences - Département de géographie
<b>Sous la direction de</b>	Benoît St-Onge, professeur associé, Faculté des arts et Sciences - Département de géographie, Université de Montréal & Audrey Smargiassi, professeure agrégée, École de santé publique - Département de santé environnementale et santé au travail, Université de Montréal.

Financement	
<b>Organisme</b>	Non financé
<b>Programme</b>	
<b>Titre de l'octroi si différent</b>	
<b>Numéro d'octroi</b>	
<b>Chercheur principal</b>	
<b>No de compte</b>	

#### MODALITÉS D'APPLICATION

Tout changement anticipé au protocole de recherche doit être communiqué au CERES qui en évaluera l'impact au chapitre de l'éthique.

Toute interruption prématurée du projet ou tout incident grave doit être immédiatement signalé au CERES

Selon les règles universitaires en vigueur, un suivi annuel est minimalement exigé pour maintenir la validité de la présente approbation éthique, et ce, jusqu'à la fin du projet. Le questionnaire de suivi est disponible sur la page web du CERES.

[redacted]  
Dominique Langelier, présidente  
Comité d'éthique de la recherche en santé  
Université de Montréal

**24 février 2015**  
Date de délivrance

**1er avril 2016**  
Date de fin de validité

adresse postale  
C.P. 6128, succ. Centre-ville  
Montréal QC H3C 3J7

3744 Jean-Brillant  
4e étage, bur. 430-11  
Montréal QC H3T 1P1

Téléphone : 514-343-6111 poste 2604  
ceres@umontreal.ca  
www.ceres.umontreal.ca

25 avril 2016

Objet: Certificat d'approbation éthique - 1er renouvellement - « Mise en relation de la pollution atmosphérique et des épisodes d'asthme par modélisation spatiale: Développement d'approches géostatistiques à l'échelle urbaine pour l'estimation de l'exposition aux particules fines et à l'ozone troposphérique »

Mme Yuddy Ramos Yngaroca,

Le Comité d'éthique de la recherche en santé (CERES) a étudié votre demande de renouvellement pour le projet de recherche susmentionné et a délivré le certificat d'éthique demandé suite à la satisfaction des exigences qui prévalent. Vous trouverez ci-joint une copie numérisée de votre certificat; copie également envoyée à votre directeur/directrice de recherche et à la technicienne en gestion de dossiers étudiants (TGDE) de votre département.

Notez qu'il y apparaît une mention relative à un suivi annuel et que le certificat comporte une date de fin de validité. En effet, afin de répondre aux exigences éthiques en vigueur au Canada et à l'Université de Montréal, nous devons exercer un suivi annuel auprès des chercheurs et étudiants-chercheurs.

De manière à rendre ce processus le plus simple possible et afin d'en tirer pour tous le plus grand profit, nous avons élaboré un court questionnaire qui vous permettra à la fois de satisfaire aux exigences du suivi et de nous faire part de vos commentaires et de vos besoins en matière d'éthique en cours de recherche. Ce questionnaire de suivi devra être rempli annuellement jusqu'à la fin du projet et pourra nous être retourné par courriel. La validité de l'approbation éthique est conditionnelle à ce suivi. Sur réception du dernier rapport de suivi en fin de projet, votre dossier sera clos.

Il est entendu que cela ne modifie en rien l'obligation pour le chercheur, tel qu'indiqué sur le certificat d'éthique, de signaler au CERES tout incident grave dès qu'il survient ou de lui faire part de tout changement anticipé au protocole de recherche.

Nous vous prions d'agréer, Madame, l'expression de nos sentiments les meilleurs,

Guillaume Paré

Conseiller en éthique de la recherche.

Comité d'éthique de la recherche en santé (CERES)

Université de Montréal

c.c. Gestion des certificats, BRDV

Benoît St-Onge, professeur associé, Faculté des arts et Sciences - Département de géographie

Audrey Smargiassi, professeure agrégée, École de santé publique - Département de santé environnementale et santé au travail

Fanny Duval (Géographie)

p.j. Certificat #15-014-CERES-D

**adresse postale**

C.P. 6128, succ. Centre-ville  
Montréal QC H3C 3J7

3744 Jean-Brillant  
4<sup>e</sup> étage, bur. 430-11  
Montréal QC H3T 1P1

Téléphone : 514-343-6111 poste 2604  
ceres@umontreal.ca  
www.ceres.umontreal.ca

**CERTIFICAT D'APPROBATION ÉTHIQUE**  
**- 1er renouvellement -**

*Le Comité d'éthique de la recherche en santé (CERES), selon les procédures en vigueur et en vertu des documents relatifs au suivi qui lui a été fournis conclu qu'il respecte les règles d'éthique énoncées dans la Politique sur la recherche avec des êtres humains de l'Université de Montréal*

Projet	
<b>Titre du projet</b>	<b>Mise en relation de la pollution atmosphérique et des épisodes d'asthme par modélisation spatiale: Développement d'approches géostatistiques à l'échelle urbaine pour l'estimation de l'exposition aux particules fines et à l'ozone troposphérique</b>
<b>Étudiante requérante</b>	<b>Yuddy Ramos Yngaroca</b> [redacted] Candidate au Ph.D. en géographie, Faculté des arts et sciences - Département de géographie
<b>Sous la direction de</b>	Benoît St-Onge, professeur associé, Faculté des arts et Sciences - Département de géographie, Université de Montréal & Audrey Smargiassi, professeure agrégée, École de santé publique - Département de santé environnementale et santé au travail, Université de Montréal.

Financement	
<b>Organisme</b>	Non financé
<b>Programme</b>	
<b>Titre de l'octroi si différent</b>	
<b>Numéro d'octroi</b>	
<b>Chercheur principal</b>	
<b>No de compte</b>	

**MODALITÉS D'APPLICATION**

Tout changement anticipé au protocole de recherche doit être communiqué au CERES qui en évaluera l'impact au chapitre de l'éthique. Toute interruption prématurée du projet ou tout incident grave doit être immédiatement signalé au CERES.

Selon les règles universitaires en vigueur, un suivi annuel est minimalement exigé pour maintenir la validité de la présente approbation éthique, et ce, jusqu'à la fin du projet. Le questionnaire de suivi est disponible sur la page web du CERES.

[redacted]  
Guillaume Paré  
Conseiller en éthique de la recherche.  
Comité d'éthique de la recherche en santé  
Université de Montréal

<b>25 avril 2016</b> Date de délivrance du renouvellement ou de la réémission*	<b>1er mai 2017</b> Date du prochain suivi
<b>24 février 2015</b> Date du certificat initial	<b>1er mai 2017</b> Date de fin de validité

\*Le présent renouvellement est en continuité avec le précédent certificat