

Modélisation de la pauvreté urbaine dans la région métropolitaine de Montréal entre 1986 et 2016

Apport des régressions spatiales par panel

Modelling urban poverty in the Montreal metropolitan region between 1986 and 2016

The contribution of spatial panel data models

Modelización de la pobreza urbana en la región metropolitana de la ciudad de Montreal, entre 1986 y 2016

Aporte de regresiones por panel

Maxime Gaboriault-Boudreau, Philippe Apparicio and Cédric Brunelle

Volume 63, Number 179-180, September–December 2019

URI: <https://id.erudit.org/iderudit/1084230ar>

DOI: <https://doi.org/10.7202/1084230ar>

[See table of contents](#)

Publisher(s)

Département de géographie de l'Université Laval

ISSN

0007-9766 (print)

1708-8968 (digital)

[Explore this journal](#)

Cite this article

Gaboriault-Boudreau, M., Apparicio, P. & Brunelle, C. (2019). Modélisation de la pauvreté urbaine dans la région métropolitaine de Montréal entre 1986 et 2016 : apport des régressions spatiales par panel. *Cahiers de géographie du Québec*, 63(179-180), 165–182. <https://doi.org/10.7202/1084230ar>

Article abstract

There is a growing interest for spatial econometric models in geography. Recently, extensions of the cross-sectional spatial autoregressive models have been proposed to consider the temporal dimension of panel data (N locations at T time periods). Currently, these models are still little known to geographers. Therefore, the aim of this paper is to describe these models and present their advantages in analyzing the linear relation between a variable of interest and explanatory variables in the case of urban panel data. We first summarize panel models and their specifications. Then, we describe a methodological routine to identify the most appropriate model for a spatial panel dataset among various spatial panel models. Our empirical application of this method on poverty in the region of Montreal (at census tract level) from 1986 to 2016 shows its robustness to control for spatial dependency, but also its ability to identify several spatial effects ignored by classical aspatial models (spillover effects of the dependent and independent variables and feedback effect).

Modélisation de la pauvreté urbaine dans la région métropolitaine de Montréal entre 1986 et 2016

Apport des régressions spatiales par panel

Maxime GABORIAULT-BOUDREAU

Institut national de la recherche scientifique
Centre Urbanisation Culture Société
maxime.gaboriault.boudreau@gmail.com

Résumé

Les modèles d'économétrie spatiale sont de plus en plus utilisés en géographie. Récemment, des extensions des modèles spatiaux autorégressifs ont été proposées afin d'analyser des données par panel, comprenant N entités spatiales pour lesquelles un ensemble de variables sont observées à T dates. À l'heure actuelle, ces modèles de régression par panel sont très peu connus des géographes. Par conséquent, notre objectif, dans cet article, est de décrire ces modèles et d'illustrer leurs avantages pour l'élaboration de diagnostics urbains longitudinaux. Pour ce faire, nous proposons un retour sur les régressions par panel. Puis nous décrivons une routine méthodologique permettant de déterminer le modèle le plus approprié au jeu de données à l'étude, parmi plusieurs modèles de régression par panel. L'application empirique, portant sur la modélisation de la pauvreté à Montréal de 1986 à 2016 par secteur de recensement, démontre que cette méthode est très robuste pour contrôler la variable de dépendance spatiale, mais aussi pour établir plusieurs types d'effets (directs, de débordement et de renvoi).

Mots-clés

Pauvreté urbaine, analyse spatiale, régression spatiale, régression par panel, Montréal.

Philippe APPARICIO

Institut national de la recherche scientifique
Centre Urbanisation Culture Société
philippe.apparicio@inrs.ca

Modelling urban poverty in the Montreal metropolitan region between 1986 and 2016: The contribution of spatial panel data models

Abstract

There is a growing interest for spatial econometric models in geography. Recently, extensions of the cross-sectional spatial autoregressive models have been proposed to consider the temporal dimension of panel data (N locations at T time periods). Currently, these models are still little known to geographers. Therefore, the aim of this paper is to describe these models and present their advantages in analyzing the linear relation between a variable of interest and explanatory variables in the case of urban panel data. We first summarize panel models and their specifications. Then, we describe a methodological routine to identify the most appropriate model for a spatial panel dataset among various spatial panel models. Our empirical application of this method on poverty in the region of Montreal (at census tract level) from 1986 to 2016 shows its robustness to control for spatial dependency, but also its ability to identify several spatial effects ignored by classical spatial models (spillover effects of the dependent and independent variables and feedback effect).

Keywords

Urban poverty, spatial analysis, spatial regression, spatial panel regression, Montreal.

Cédric BRUNELLE

Institut national de la recherche scientifique
Centre Urbanisation Culture Société
cedric.brunelle@inrs.ca

Modelización de la pobreza urbana en la región metropolitana de la ciudad de Montreal, entre 1986 y 2016: aporte de regresiones por panel

Resumen

En Geografía, los modelos de econometría espacial son utilizados con mayor frecuencia. Últimamente, extensiones de modelos espaciales auto regresivos han sido propuestos para analizar datos por panel (muestra de grupo permanente) con N entidades espaciales, por las que se observan un conjunto de variables en T data. Actualmente, los modelos de regresión por panel son poco conocidos por los geógrafos. Nuestro objetivo, en el presente artículo, es de describirlos e ilustrar sus ventajas para la elaboración de diagnósticos urbanos longitudinales. Para ello, proponemos una retrospectiva sobre las regresiones por panel. Luego, describimos una rutina metodológica que permita determinar el modelo apropiado a la serie de datos estudiados, entre los varios modelos de regresión por panel. La aplicación empírica, que concierne la modelización de la pobreza en Montreal de 1986 a 2016 por sector de empadronamiento, demuestra la robustez del método, tanto para controlar la variable de dependencia espacial, como para establecer varios tipos de efectos (directos, de exceso y de exclusión).

Palabras claves

Pobreza urbana, análisis espacial, regresión espacial, regresión por panel, Montreal.

Introduction

Depuis une trentaine d'années, économètres, épidémiologistes et géographes développent et utilisent abondamment des méthodes de modélisation intégrant l'espace, et ce, afin de contrôler la variable de dépendance spatiale et l'hétérogénéité spatiale des modèles classiques. Parmi ces modèles, citons notamment les modèles spatiaux autorégressifs (SAR-Lag, SAR-Error, CAR-Error, SAR lagged-mixed) (Anselin, 1988; Keitt *et al.*, 2002; Haining, 2003), les modèles de régression géographiquement pondérée (Fotheringham *et al.*, 2003), les modèles *Spatial Eigenvector Mapping* (SEVM) (Griffith et Peres-Neto, 2006; Griffith, 2013), les modèles additifs généralisés avec une fonction pour les coordonnées géographiques (*Trend surface GAM*) (Zuur *et al.*, 2013), les modèles GAM avec *Random Markov Field* (RMF) (Rue et Held, 2005; Wood, 2017) ou encore les modèles des moindres carrés généralisés (*Generalized Least Square Model*) avec une structure d'autocorrélation dans le terme d'erreur (Cressie, 2015). Pour une présentation et une comparaison succincte de ces méthodes qui prennent en compte l'autocorrélation spatiale, on pourra consulter la revue de littérature de Dormann *et al.* (2007).

Parmi ces méthodes, les modèles autorégressifs proposés par Anselin (1988) sont sans aucun doute les plus utilisés par les géographes. Au Québec, ils ont été surtout mobilisés en géographie urbaine. En guise d'exemple, pour la région de Montréal, les régressions spatiales ont été utilisées dans le cadre de travaux rattachés à des thématiques très diverses afin de modéliser les distributions de la végétation urbaine (Pham *et al.*, 2012 et 2013), du bruit routier (Carrier *et al.*, 2016 a et b), de la pollution atmosphérique (Carrier *et al.*, 2014), de plusieurs nuisances urbaines simultanément (Carrier *et al.*, 2016c), des emplois de la nouvelle économie liée aux nouvelles technologies (Duvivier *et al.*, 2018) ou encore l'exposition des cyclistes au bruit et à la pollution (Apparicio *et al.*, 2016). D'autres auteurs tirent profit des régressions spatiales. Dubé et Polèse (2015) analysent ainsi les impacts de la récession de 2008 au Canada à l'aide de modèles spatiaux, mais c'est surtout dans le champ de la modélisation hédonique des valeurs résidentielles que ces modèles sont utilisés (Kestens *et al.*, 2006). La dimension temporelle des données est même intégrée dans ce champ, notamment dans des modèles spatiaux de différence-de-différence (SDID) (Dubé *et al.*, 2014; Devaux *et al.*, 2017) et dans les modèles spatiotemporels autorégressifs (STAR) (Dubé et Legros, 2014a; Thanos *et al.*, 2016).

Récemment, des extensions des modèles autorégressifs ont été proposées afin de traiter de cette nature spatiotemporelle, mais cette fois en considérant des données par panel (Elhorst, 2009; Lesage, 2014). Ces modèles spatiaux par panel constituent ainsi des outils économétriques puissants pour la modélisation de données spatiales longitudinales (Kopczewska *et al.*, 2017), en particulier celles extraites de plusieurs recensements de population ancrées à un découpage spatial spécifique (secteur, subdivision ou division de recensement, par exemple). Ces méthodes sont employées dans plusieurs études empiriques récentes en géographie économique et en économie régionale. Notamment, Parajuli et Haynes (2017) étudient la formation de nouvelles entreprises dans les comtés de Nouvelle-Angleterre (NA) entre 1999 et 2009 alors que Qian et Zhao (2018) le font pour le cas particulier des entreprises de la nouvelle économie, de 1999 à 2012, et comparent la NA et la Californie. Quant à Yu *et al.* (2013) et Tong *et al.* (2013), ils s'intéressent aux impacts des infrastructures de transport: les premiers sur le produit intérieur brut (PIB) réel de 29 provinces chinoises, de 1978 à 2009, et les seconds, sur la valeur de la production agricole dans 44 États américains durant la période 1981-2004. Les régressions spatiales par panel sont aussi utilisées pour l'étude de la pollution. Ge *et al.* (2018) s'intéressent à l'influence de facteurs économiques comme la croissance du PIB ou l'urbanisation sur les émissions d'oxydes de carbone dans 30 provinces chinoises, entre 2010 et 2015, alors que Burnett *et al.* (2013) estiment l'impact des prix des différentes sources d'énergie sur les émissions de CO² dans 48 États américains contigus, de 1970 à 2009.

Il n'en demeure pas moins que, à l'heure actuelle, ces modèles spatiaux par panel sont très peu connus et utilisés par les géographes. Par conséquent, notre objectif, dans cet article, est avant tout méthodologique: décrire en détail cette méthode et illustrer ses avantages pour l'élaboration de diagnostics urbains longitudinaux, et ce, en utilisant un jeu de données sur la pauvreté urbaine dans la région métropolitaine de Montréal de 1986 à 2016.

Territoire d'étude, unités spatiales et variables retenues

Afin de caractériser l'effet des facteurs associés à la pauvreté dans la région métropolitaine de recensement (RMR) de Montréal sur une période de 30 ans (1986 à 2016), nous avons recours à une approche par panel. À l'instar d'autres travaux menés à Montréal sur des données

longitudinales (Ades *et al.*, 2012; Séguin *et al.*, 2012 et 2016), cela nécessite deux traitements géographiques préalables puisque la délimitation de plusieurs secteurs de recensement varie d'un recensement à l'autre, tout comme celle de la région métropolitaine, qui prend habituellement de l'expansion avec le temps.

Premièrement, nous avons retenu comme délimitation de la région d'étude les frontières de la RMR de 1986, soit l'année de départ. Deuxièmement, étant donné que le nombre et la géographie des secteurs de recensement (SR – unité géographique comprenant habituellement de 2 500 à 8 000 habitants) varie légèrement d'un recensement à l'autre, certains SR adjacents ont été agrégés. On s'assure ainsi d'avoir le même nombre et la même délimitation des SR pour chacune des sept années de recensement (1986, 1991, 1996, 2001, 2006, 2011 et 2016). En d'autres termes, ces deux traitements géographiques permettent d'obtenir une homogénéité spatiale et temporelle des unités statistiques, condition indispensable à des données par panel. Au final, le jeu de données est composé de 672 individus (SR avec $N = 672$) et sept années des recensements ($T = 7$).

Comme variable dépendante, nous avons retenu le pourcentage de personnes dans les ménages privés à faible revenu¹ dans l'ensemble de la population. Tel que montré par des études longitudinales récentes (Ades *et al.*, 2012; Séguin *et al.*, 2012), la cartographie de ce pourcentage pour les sept années de recensement (1986 à 2016) souligne la concentration de la pauvreté sur l'île de Montréal, mais aussi la gentrification des quartiers centraux et une certaine suburbanisation de la pauvreté durant la période d'étude (figure 1). Concernant les facteurs prédictifs, nous avons retenu les facteurs classiques associés à la pauvreté, à savoir le chômage, la monoparentalité, la faible scolarité, le fait de vivre seul, l'immigration récente et le fait d'avoir 65 ans ou plus (Lee, 2000; Heisz et McLeod, 2004; Groulx, 2011; Gouvernement du Canada, 2016). Par conséquent, les pourcentages de ces différents groupes constituent nos variables indépendantes (tableau 1), telles que sélectionnées dans les études récentes sur la modélisation de la pauvreté à Montréal (Apparicio *et al.*, 2007; Séguin *et al.*, 2012). Il aurait aussi été judicieux d'inclure d'autres variables dans nos modèles, comme le pourcentage de minorités visibles, soit un groupe de population particulièrement vulnérable à la pauvreté (Leloup, 2007; Groulx, 2011). Toutefois, cette variable n'est disponible qu'à partir du recensement de 1996, et non pour toute la période d'étude.

TABLEAU 1 Statistiques descriptives pour les différentes variables des modèles par panel ($N = 672$) | Conception: Gaboriault-Boudreau, Apparicio et Brunelle, 2019

	Y_FR	Chomag	FaibSc	FaMono	ImgRec	Menag1	P65
Moyenne	23,29	10,05	45,32	19,15	4,24	31,99	13,86
Écart-type	13,85	4,80	14,12	7,23	4,56	14,05	6,90
Minimum	0,00	0,00	5,10	0,00	0,00	2,09	1,20
Quartile 1	12,27	6,57	35,21	13,90	0,87	20,43	9,07
Médiane	20,90	9,15	46,08	18,45	2,77	31,99	12,98
Quartile 3	31,56	12,39	55,61	23,53	5,95	42,79	17,37
Maximum	90,99	51,11	83,07	56,76	31,65	78,79	62,11

Y_FR : pourcentage de personnes à faible revenu dans la population totale / **Chomag** : taux de chômage chez les personnes de 15 ans et plus / **FaibSc** : pourcentage des personnes faiblement scolarisées dans la population de 15 à 64 ans, correspondant aux études secondaires ou moins / **FaMono** : pourcentage des familles monoparentales dans l'ensemble des familles / **ImgRec** : pourcentage des immigrants récents (établis au Canada depuis moins de cinq ans) dans la population totale. Notons que, pour les recensements de 1986 et 1991, le nombre d'années retenu depuis l'installation est de trois ans, et non de cinq comme pour les recensements subséquents / **Menag1** : pourcentage des ménages d'une personne dans l'ensemble des ménages / **P65** : pourcentage de personnes de 65 ans et plus dans la population totale

1 Selon l'approche des seuils de faible revenu de Statistique Canada, une personne ou une famille est considérée comme ayant un faible revenu si son revenu familial réel se situe en deçà d'un seuil établi à partir de la structure des dépenses des familles canadiennes. Les seuils de faible revenu après impôt sont fixés à partir des données de l'Enquête sur les dépenses des familles canadiennes. Ces dépenses sont exprimées en termes de pourcentage du revenu avant impôt. Les seuils de faible revenu de l'année sont établis à partir des familles qui consacrent 20 points de pourcentage de plus que l'ensemble des ménages canadiens pour trois postes de dépenses : le logement, l'alimentation et l'habillement.

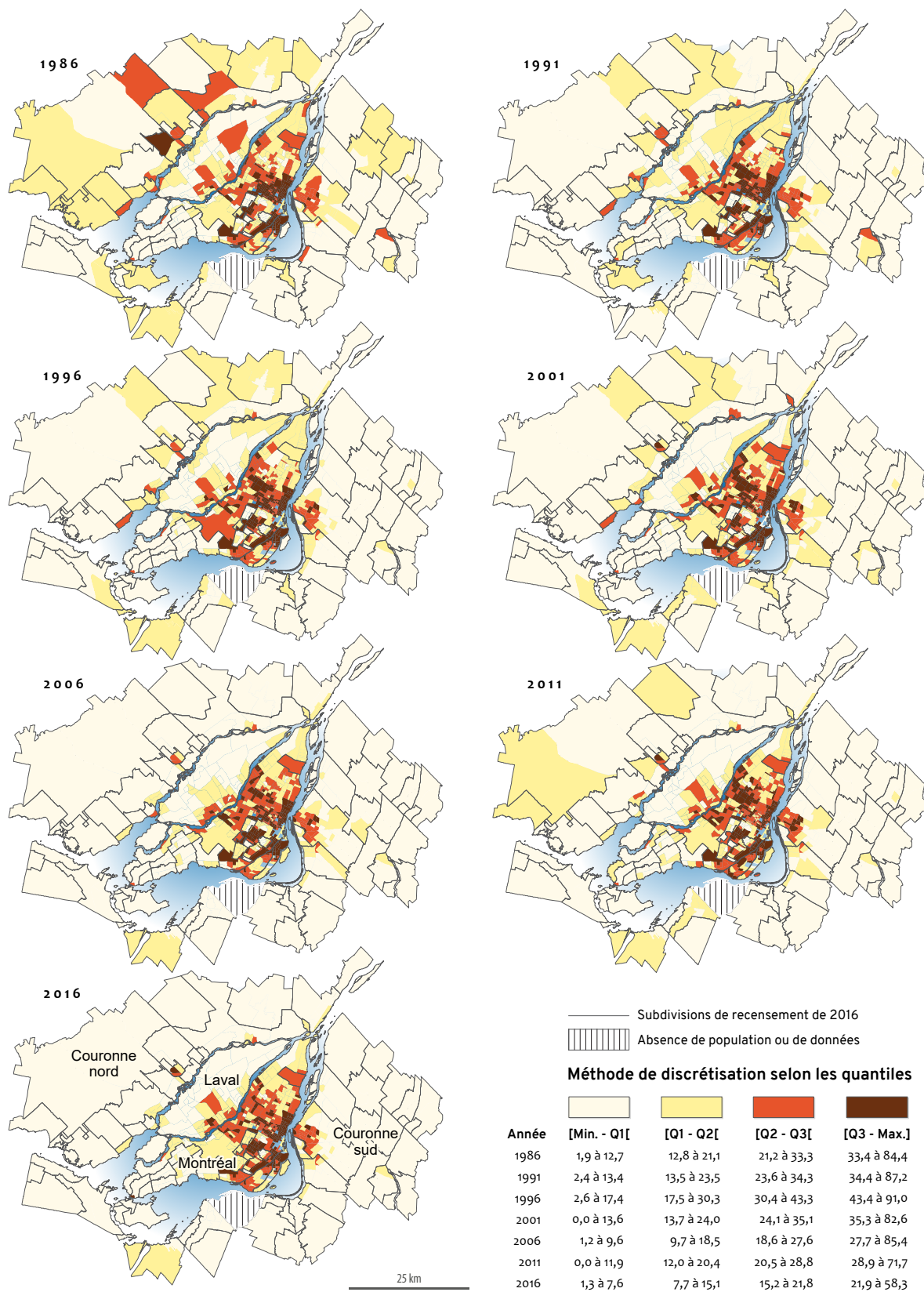


FIGURE 1 Distribution spatiale de la pauvreté dans la région métropolitaine de Montréal, de 1986 à 2016 | Conception: Gaboriault-Boudreau, Apparicio et Brunelle, 2019 | Source: Statistique Canada, 2010 et 2018

Un bref retour sur les modèles par panel standards (non spatiaux)

Notre objectif, dans cette section, est de décrire brièvement les différents modèles non spatiaux qui permettent d'analyser des jeux de données par panel, composés de N individus pour lesquels un ensemble de variables sont observées à T dates (Baltagi, 2013). Nous nous limiterons ici aux modèles linéaires. En d'autres termes, nous n'aborderons pas les modèles de régression par panel logistique (Wooldridge, 2002: 482-497), les régressions catégorielles ordinales par panel (Pfarr *et al.*, 2010) ou encore les processus de comptage par panel (Hausman *et al.*, 1984).

Les différents types de modèles par panel

La régression linéaire par panel standard consiste à modéliser une variable dépendante Y sur la base de K variables indépendantes à partir de l'équation suivante (Baltagi, 2013):

$$y_{it} = \alpha + X_{it}^T \beta + u_{it} \quad [1]$$

$$\text{où } u_{it} = \mu_i + \lambda_t + v_{it} \quad (i = 1, \dots, N; t = 1, \dots, T)$$

avec X_{it} le vecteur des valeurs des K variables indépendantes de l'individu i (SR) au temps t (année de recensement), β le vecteur de coefficient associé et α représente la constante. Le terme d'erreur u_{it} est ici défini dans sa forme la plus spécifique où μ_i représente les effets individuels qui ne changent pas pour un même individu au cours du temps et λ_t les effets temporels, qui affectent tous les individus de la même façon pour une période t donnée. Le terme restant v_{it} est le résidu usuel de la régression tel que $v_{it} \sim iid(0, \sigma_v^2)$. En cas d'absence d'effets, soit quand $u_{it} = v_{it}$ dans l'équation 1, on a alors un modèle dit groupé (*pooled OLS*). Par contre, si l'on a des effets individuels μ_i et/ou temporels λ_t , il convient alors de déterminer leur nature, qui peut être soit aléatoire, soit fixe (Baltagi, 2013). Dans le cas d'effets aléatoires (RE - *random effects*) mixtes (effets individuels et temporels), on a alors $\mu_i \sim iid(0, \sigma_\mu^2)$ et $\lambda_t \sim iid(0, \sigma_\lambda^2)$ tel que μ_i , λ_t et v_{it} sont tous indépendants entre eux et indépendants de X_{it} pour tout i et tout t (Baltagi, 2013: 42). Dans le cas d'effets fixes (FE - *fixed effects*), μ_i et λ_t ne sont plus tenus d'être indépendants avec le vecteur de variables indépendantes X_{it} (Wooldridge, 2002: 251-252).

Identification de la présence et de la nature des effets dans les modèles par panel

Les spécifications des trois modèles décrits précédemment – groupé, FE et RE – font non seulement varier considérablement les estimations des coefficients non spatiaux, mais aussi les dépendances spatiales à inclure dans le modèle final (Elhorst, 2012). Il est donc primordial, dans une étude spatiale par panel, de bien déterminer, au préalable, le modèle non spatial le mieux adapté aux données. Pour identifier la présence, le type et la nature d'effets dans un modèle panel non spatial, nous proposons une méthode en trois étapes schématisées à la figure 2, sur la base de tests présentés dans Baltagi (2013).

À l'étape 1, à partir des statistiques du LM de Breusch et Pagan, on vérifie s'il est préférable d'utiliser un modèle avec effets aléatoires temporels (avec l'hypothèse $H_0^1: \sigma_\lambda^2 = 0$), individuels ($H_0^2: \sigma_\mu^2 = 0$) ou mixtes ($H_0^3: \sigma_\lambda^2 = \sigma_\mu^2 = 0$) plutôt qu'un modèle groupé (Baltagi, 2013: 68). Simultanément, à partir des statistiques F, on teste si l'inclusion d'effets fixes dans le modèle groupé est pertinente selon trois hypothèses (Baltagi, 2013: 40): l'absence d'effets fixes temporels en présence d'effets fixes individuels ($H_0^1: \mu_i \neq 0, \lambda_t = 0, t = 1, \dots, T - 1$), l'absence d'effets fixes individuels en présence d'effets fixes temporels ($H_0^2: \mu_i = 0, \lambda_t \neq 0, i = 1, \dots, N - 1$), l'absence à la fois d'effets fixes individuels et temporels ($H_0^3: \mu_i = \lambda_t = 0, i = 1, \dots, N - 1; t = 1, \dots, T - 1$). Notons que si H_0^1 et H_0^2 sont rejetées ou si H_0^3 est rejetée, on penchera pour les effets mixtes, peu importe les valeurs respectives des tests.

Si aucun des tests LM ou des statistiques F n'est significatif, on conserve le modèle groupé sans effets (figure 2.a). Si une seule des séries de tests entre les LM et les statistiques F donne des résultats significatifs, on conserve le modèle associé avec des effets fixes ou aléatoires de nature individuelle, temporelle ou mixte (figure 2.b et c). Finalement, si les tests LM et F sont tous deux significatifs, il convient de passer à l'étape 2, qui consiste à choisir l'un des deux modèles (figure 2.d). Pour ce faire, on a recours au test d'Hausman qui, lorsque significatif, signale qu'il est préférable d'utiliser un modèle à FE. Il est à noter que ce modèle est généralement approprié quand l'échantillon utilisé est exhaustif par rapport à la population représentée (Bouayad-Agha *et al.*, 2018), particulièrement dans le cas de données spatiales où l'ensemble des entités spatiales forme la population totale d'une région donnée (Huang et Chand, 2015 ; Parajuli et Haynes, 2017; Qian et Zhao, 2018). C'est le cas ici puisque les N secteurs de

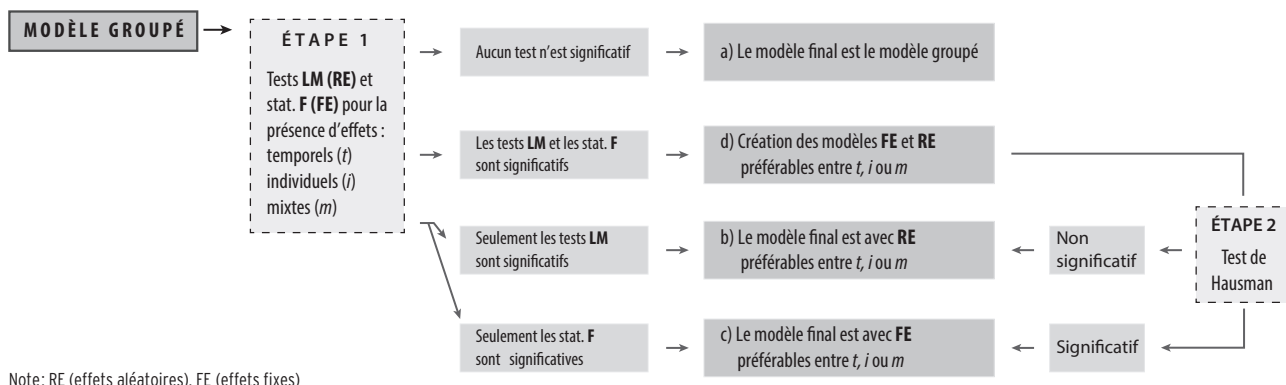


FIGURE 2 Arbre décisionnel pour le choix du modèle panel non spatial | Conception: Gaboriault-Boudreau, Apparicio et Brunelle, 2019

recensement représentent la population de l'ensemble de la RMR de Montréal; on s'attend ainsi à obtenir un modèle non spatial à FE.

Les modèles par panel spatiaux

Dépendance spatiale et matrices de pondération spatiale

Dans un modèle, les résidus sont la différence entre la valeur observée (y_i) et la valeur prédite (\hat{y}_i). Parmi les conditions d'application bien connues d'une régression, il faut que les résidus soient normalement distribués, homoscédastiques et indépendants entre eux. Appliquée à une régression sur des données spatiales, cette dernière condition nécessite alors que les résidus soient distribués aléatoirement dans l'espace: si les résidus sont autocorrélés spatialement (positivement ou négativement), il y a alors un problème de dépendance spatiale du modèle, qui fait que les coefficients obtenus sont biaisés (Lesage et Pace, 2009). Si le modèle est bien spécifié – c'est-à-dire qu'on n'a pas omis une variable importante qui permettrait de corriger le problème d'autocorrélation spatiale des résidus –, Dubé et Legros (2014b: 119-131) démontrent clairement que plusieurs raisons peuvent justifier l'utilisation d'une régression spatiale: la présence d'externalités, d'effets d'entraînement (*spillover effects*), d'hétérogénéité spatiale, d'effets mixtes. Pour pallier la dépendance spatiale, on peut alors introduire une structure spatiale dans le modèle avec des variables spatialement décalées construites à partir d'une matrice de pondération spatiale.

Ces matrices de pondération spatiale sont bien connues et largement utilisées en géographie, notamment pour calculer les indices d'autocorrélation spatiale (le I de Moran, entre autres). Pour une description détaillée de

ces matrices, on pourra notamment consulter Dubé et Legros (2014a: 41-68) et Dubin (2008). Une matrice de pondération spatiale W ($N \times N$, avec N étant le nombre d'entités spatiales) résume les relations que partagent des entités spatiales entre elles (Dubin, 2008), et ce, selon la contiguïté, l'interaction, le nombre de plus proches voisins ou la distance. Les valeurs d'une matrice de contiguïté et de voisinage sont binaires, avec: W_{ij} est égal à 1 lorsque i et j sont adjacents selon le partage d'une frontière commune (*queen contiguity*) ou d'un nœud (*rook contiguity*) ou lorsque j est le plus proche voisin de i ; et $W_{ij} = 0$ lorsque ce n'est pas le cas. Ces matrices de contiguïté sont particulièrement bien adaptées pour des entités spatiales de forme et de taille identiques comme les pixels d'une image satellitaire. Si ce n'est pas le cas, l'interaction entre deux polygones adjacents varie en fonction de la longueur de la frontière commune (l_{ij}) et la distance séparant leurs centroïdes (d_{ij}) (Can, 1996):

$$w_{ij} = (l_{ij} / \sum_{j \in J} l_{ij}) / d_{ij}, \quad i=1, \dots, N \text{ et } j=1, \dots, J \quad [2]$$

où $\sum_{j \in J} l_{ij}$ est la longueur de toutes les frontières communes pour l'entité spatiale i .

Quant aux matrices de proximité, elles sont le plus souvent basées sur l'inverse de la distance ($1/d_{ij}$) ou l'inverse de la distance au carré ($1/d_{ij}^2$). Il est à noter que ces matrices sont habituellement standardisées en ligne: la somme de chaque ligne est alors égale à 1 et la somme de l'ensemble de la matrice est égale au nombre d'observations (N), ce qui permet de comparer des matrices entre elles. Afin de neutraliser la dépendance spatiale du modèle – nous le verrons plus en détail dans la section suivante –, on peut multiplier la matrice de pondération spatiale W standardisée en ligne avec soit la variable dépendante (Y), soit les variables indépendantes (X) pour créer une variable dépendante ou des variables indépendantes spatialement

découlées (WY , WX). Cette opération peut aussi s'appliquer sur le terme d'erreur (Wu). Tel que souligné par Plummer (2009), le choix du type de matrice de pondération est un objet de débat en soi puisque celle-ci fait varier les résultats du modèle estimé. Dans le cadre de cette étude, bien que le choix d'une matrice d'interaction eût été plus judicieux (SR de taille et de forme différentes), nous avons recours à une matrice de contiguïté selon le partage d'une frontière commune d'ordre 1 (*Queen*) qui est plus largement utilisée. Nous n'avons pas retenu des matrices de distance ou de plus proches voisins pour deux raisons : les SR sont de taille très variable ; les SR proches les uns des autres ne communiquent pas nécessairement (exemple : cours d'eau les séparant, notamment entre Montréal et Laval et les couronnes nord et sud).

Formulation des différents modèles spatiaux par panel

Bien que les effets individuels inclus dans le modèle de l'équation 1 puissent neutraliser une partie de l'hétérogénéité spatiale (en captant les spécificités non observées des unités spatiales), il est possible que ce modèle ne soit pas encore assez spécifié (Burnett *et al.*, 2013 ; Parajuli et Haynes, 2017). En effet, les modèles classiques présentés dans la section précédente supposent l'indépendance des individus entre eux, une hypothèse non vérifiée en cas de dépendance spatiale (Bouayad-Agha *et al.*, 2018 ; Qian et Zhao, 2018). Comme pour les modèles spatiaux autorégressifs sur données transversales (Anselin, 1988), il est alors possible d'introduire l'autocorrélation spatiale au niveau de la variable dépendante, des variables indépendantes, du terme d'erreur, ou encore une combinaison de ces effets. Tel que signalé par Huang et Xia (2016), le modèle le plus spécifique est alors :

$$y_{it} = \alpha + \lambda \sum_{j=1}^N w_{ij} y_{jt} + X_{it}^T \beta + \sum_{j=1}^N w_{ij} X_{jt}^T \theta + u_{it} \quad [3]$$

$$\text{où } u_{it} = \rho \sum_{j=1}^N w_{ij} u_{jt} + \mu_i + \lambda_t + v_{it}$$

où w_{ij} est le poids associé à j pour i selon une matrice de voisinage ou de proximité W telle que définie précédemment. Ainsi, λ et ρ sont respectivement les coefficients pour la variable dépendante et le terme d'erreur spatialement décalés. Quant à θ , il représente le vecteur des coefficients des variables indépendantes spatialement décalées.

Elhorst (2014 : 9) définit alors une famille de sept modèles spatiaux selon l'inclusion de certains termes

de dépendance, représentée à la figure 3. Comme Kopczewska *et al.* (2017), on peut être tenté de prendre le modèle le plus spécifique comme point de départ, soit celui incluant tous les termes de dépendance dans l'équation 3 (au niveau de la variable dépendante, $\lambda W y$, des variables indépendantes, $WX\theta$, et du terme d'erreur, ρWu) (figure 3.a). Par contre, Elhorst (2010) signale que ce modèle a tendance à être surspécifié, c'est-à-dire que l'inclusion des trois termes simultanément ne permet pas toujours de bien distinguer l'effet particulier de chaque terme de dépendance. En effet, un tel modèle pourrait présenter de la multicolinéarité due au fait qu'en présence d'autant de termes de dépendances spatiales, Y_i est en partie expliquée par la variable dépendante du voisinage Y_j qui dépend elle-même de tous les déterminants d' Y_i , soit Y_i dans WY , X_i dans WX et u_i dans Wu . Surviennent alors des biais dans les estimations des effets précis de chaque type de dépendance. Par conséquent, il convient de débiter avec un modèle à seulement deux termes de dépendance spatiale (figure 3.b, c ou d). Le *spatial Durbin panel data model* (SDPDM – figure 3.c) apparaît comme meilleur point de départ puisqu'il inclut les dépendances spatiales des variables dépendante et indépendantes et ne peut donc pas souffrir de biais de variable omise, en plus de pouvoir être simplifié à tous les autres modèles (figure 3.e à h) (Elhorst, 2010). Ainsi, en se basant sur la routine méthodologique proposée par Elhorst (2012) et sur d'autres travaux (Burnett *et al.*, 2013 ; Yu *et al.*, 2013 ; Liu *et al.*, 2016 ; Parajuli et Haynes, 2017 ; Ge *et al.*, 2018), parmi ces sept modèles spatiaux, on considère habituellement trois modèles principaux qui représentent le mieux les données à l'étude :

- le *spatial lag panel data model* (SLPDM) avec l'introduction de l'autocorrélation spatiale uniquement au niveau de la variable dépendante ($\lambda W y$, figure 3.e) ;
- le *spatial error panel data model* (SEPDM) avec l'introduction de l'autocorrélation spatiale uniquement au niveau du terme d'erreur (ρWu , figure 3.g) ;
- le *spatial Durbin panel data model* (SDPDM) avec l'introduction de l'autocorrélation spatiale à la fois pour les variables dépendante et indépendantes ($\lambda W y$ et $WX\theta$, figure 3.c).

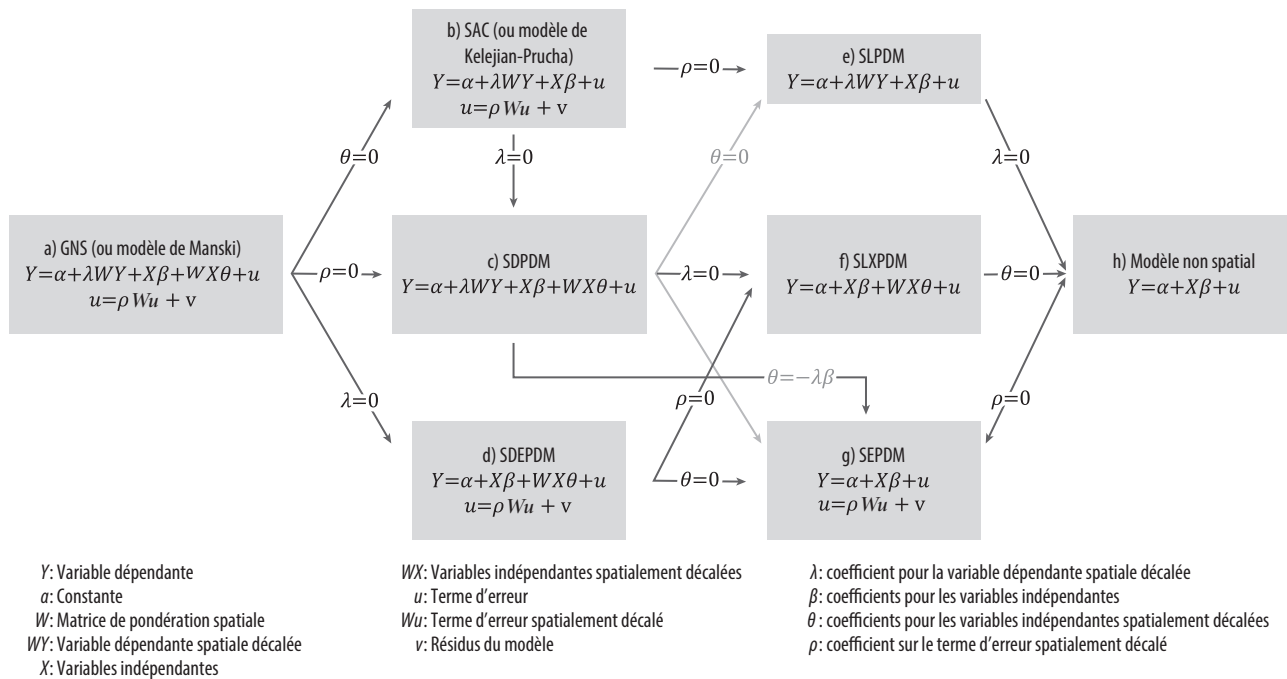


FIGURE 3 Dérivations des différents modèles économétriques spatiaux par panel | Source: Adaptée de Elhorst, 2010; 2014

Sélection du modèle spatial par panel le plus approprié

Afin de sélectionner le modèle spatial le plus approprié (SLPDM, SEPDM ou SDPDM), Elhorst (2012) propose une démarche schématisée à la figure 4 et largement utilisée dans des études empiriques récentes (Burnett *et al.*, 2013; Tong *et al.*, 2013; Ge *et al.*, 2018; Qian et Zhao, 2018). À la première étape, les tests LM permettent de vérifier si l'on rejette l'absence de dépendance spatiale au niveau de la variable dépendante (LM-lag) et l'absence de dépendance spatiale au niveau de l'erreur (LM-error). D'emblée, si aucun des tests LM n'est significatif, on retiendra le modèle non spatial comme modèle final (figure 4.a). Par contre, si un seul des deux tests LM est significatif, on calcule la version robuste du même test (RLM-lag à l'étape 2.i ou RLM-error en 2.ii) qui, lorsque significatif, nous fera retenir soit le SLPDM (figure 4.b), soit le SEPDM (figure 4.c).

En revanche, toujours à l'étape 2.i ou 2.ii, si le test robuste est non significatif, on estime alors les trois modèles (SEPDM, SLPDM, SDPDM; étape 3). Par ailleurs, si les deux tests LM sont significatifs à l'étape 1, on calcule les deux tests RLM à l'étape 2.iii. S'ils sont tous deux significatifs, on conserve le SDPDM comme modèle spatial final (figure 4.d). Cependant, si aucun ou un seul test RLM est significatif à l'étape 2.iii, on passe aussi à l'étape 3 où, une fois les trois modèles créés, on a recours aux tests LR pour

vérifier si le SDPDM peut être réduit au SLPDM ($H_0^1 : \theta = 0$) ou au SEPDM ($H_0^2 : \theta = -\lambda\beta$). Si les résultats des tests LR et LM sont cohérents et identifient un même modèle, c'est-à-dire qu'on rejette LM-lag (LM-error) et qu'on ne rejette pas H_0^1 (H_0^2), on choisit le SLPDM (SEPDM). Si l'on ne rejette ni H_0^1 ni H_0^2 ou qu'on rejette les deux hypothèses, ou encore si les résultats des tests LR et LM sont contradictoires (c.-à-d. qu'ils identifient chacun un modèle différent), alors on garde le SDPDM (figure 4.d). Cette logique de conserver le SDPDM en cas d'incertitude vient du fait qu'il peut être décomposé comme une somme pondérée des SLPDM et SEPDM lorsqu'on n'a pas la certitude qu'un seul de ces deux modèles représente mieux les données (Lesage et Pace, 2009: 30-31). Il est ainsi le modèle le plus inclusif (Elhorst, 2012) et ne peut pas être biaisé en raison de l'omission de WY ou WX dans l'un des deux autres modèles (Lesage, 2014). Finalement, une fois le bon modèle spatial déterminé entre SEPDM, SLPDM et SDPDM, nous pourrions vérifier si les effets individuels et/ou temporels inclus sont bien de nature fixe ou aléatoire, à l'aide du test de Hausman appliqué aux modèles spatiaux (Elhorst, 2009; Bouayad-Agha *et al.*, 2018).

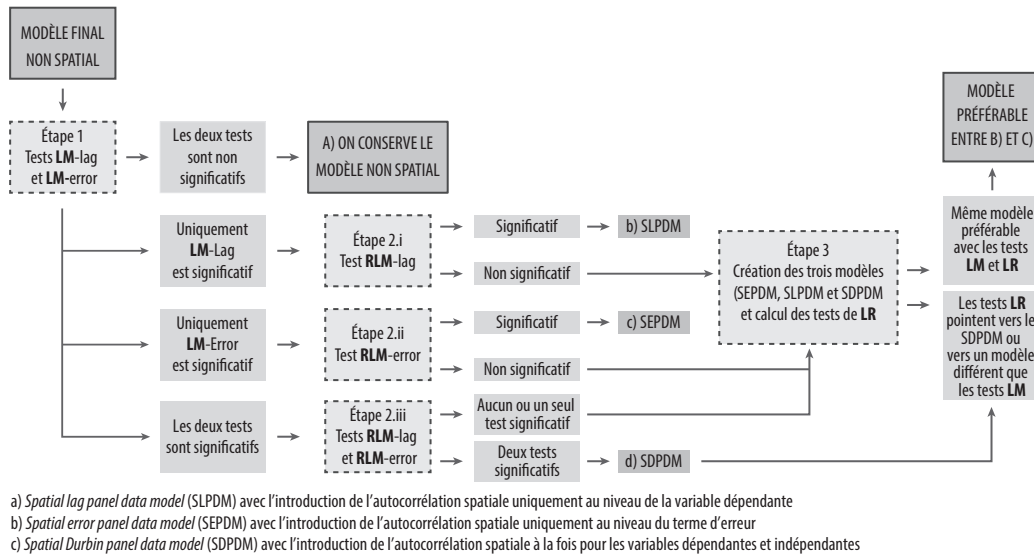


FIGURE 4 Arbre décisionnel pour le choix du modèle de régression spatial par panel | Source: Adaptée de Elhorst, 2012

Évaluation des impacts directs et indirects

Dans le cas des modèles spatiaux avec dépendance dans les variables dépendante (Wy) ou indépendantes (WX), il ne faut pas se limiter à la lecture du coefficient d'une variable indépendante pour analyser son effet sur la variable dépendante (Bivand *et al.*, 2013: 308; Elhorst, 2014: 20). Pour ces modèles, Elhorst (2012) décompose la variation totale de la valeur espérée de la variable dépendante au temps t due à la variation de la k -ième variable indépendante comme suit:

$$\left[\frac{\partial E(Y)}{\partial X_k} \right]_t = \begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & \dots & \frac{\partial E(y_1)}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_N)}{\partial x_{1k}} & \dots & \frac{\partial E(y_N)}{\partial x_{Nk}} \end{bmatrix} \quad [4]$$

$$= (I_N - \lambda W)^{-1} \begin{bmatrix} \beta_k & w_{12}\theta_k & \dots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \dots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \dots & \beta_k \end{bmatrix}$$

où w_{ij} , W et λ sont tels que définis plus haut, I_N est la matrice identité de dimension N et θ_k est le coefficient de la variable indépendante k spatialement décalée (WX_k).

Ainsi, en présence de dépendance spatiale au niveau des variables ($\theta_k \neq 0$ et/ou $\lambda \neq 0$), la variation de $x_{it,k}$ cause non seulement une variation locale de y_{it} (impact direct), mais aussi une variation dans l'ensemble des autres localisations y_{jt} (impact indirect ou effet de débordement) (Elhorst, 2014: 21). À juste titre, Elhorst (2012) souligne que

la matrice à droite dans l'équation 4 est indépendante de t et donc que les calculs d'impacts définis pour des données uniquement transversales par Lesage et Pace (2009) sont valides dans le cas panels avec poids w_{ij} constants dans le temps. Pour simplifier la présentation des résultats, Lesage et Pace (2009: 36-37) calculent l'impact direct comme la moyenne des éléments diagonaux de la matrice à droite dans l'équation 4 et l'impact indirect comme la moyenne des sommes par rangées (ou colonnes) des éléments non diagonaux de cette même matrice. Elhorst (2014: 23) définit l'effet de renvoi (*feedback effect*) comme l'impact supplémentaire en i dû à un premier changement en i . Plus précisément, il est possible qu'une variation de $x_{it,k}$ cause une première variation de y_{it} , puis une autre variation de y_{jt} pour i et j voisins (si $\theta_k \neq 0$) qui, à son tour, génère une seconde variation de y_{it} (si $\lambda \neq 0$). Ainsi, l'analyse des impacts directs et indirects contient plus d'informations sur la variation de Y due à celle de X_k que le simple coefficient β_k (Kopczewska *et al.*, 2017). C'est d'ailleurs l'une des forces de l'économétrie spatiale d'arriver à définir et distinguer ces impacts (Lesage et Pace, 2009: 33).

Les résultats: application empirique des modèles spatiaux par panel

Avant d'analyser les résultats, il convient de signaler l'absence de multicollinéarité excessive entre les variables indépendantes introduites dans les différents modèles. En effet, les valeurs du facteur d'inflation de la variance (VIF), non reportées ici, sont toutes inférieures à 5

(valeur maximale de 2,10), soit le seuil habituellement retenu par plusieurs auteurs (Chatterjee et Hadi, 2015; Studenmund et Cassidy, 2016). Aussi, la mise en œuvre des modèles a été réalisée dans R (R Core Team, 2017) avec les paquets *plm* (Croissant et Millo, 2008) et *splm* (Millo et Piras, 2012), respectivement pour les modèles panel standards et spatiaux.

Détermination du modèle le mieux adapté aux données

Tel que décrit dans la section méthodologique et illustré à la figure 2, avant de mettre en œuvre les modèles spatiaux, il convient de déterminer quel est le modèle standard par panel optimal. Pour ce faire, on vérifie s’il y a présence d’effets temporels, individuels ou mixtes à partir des tests LM (pour les RE) et des statistiques F (pour les FE). À la lecture des valeurs reportées au tableau 2, comme on rejette l’absence d’effets mixtes dans les deux cas, on en conclut qu’il est préférable d’utiliser les modèles FE et RE avec des effets temporels et individuels plutôt qu’un modèle groupé. Par la suite, la valeur du test d’Hausman ($\chi^2_{12} = 1\,201$; $p < 0,001$) permet de conclure qu’il est préférable de retenir un modèle à effets fixes (FE) mixtes.

TABLEAU 2 Tests pour déterminer les types d’effets présents dans le modèle | Conception: Gaboriault-Boudreau,

Effets	Tests LM (H ₀ : absence de RE)	Tests F (H ₀ : absence de FE)
Temporels	119 996***	339,9***
Individuels	301,53***	3,61***
Mixtes (individuels et temporels)	120 297***	13,65***

Significativité: ***p<0,001

On peut maintenant s’attarder à la détermination des termes de dépendance spatiale à inclure dans le modèle. Suivant l’approche méthodologique illustrée à la figure 4, nous effectuons les tests LM et RLM sur les résidus du modèle standard par panel à FE mixtes. Les deux tests LM sont significatifs (LM-error = 26,527, $p < 0,001$; LM-lag = 45,078, $p < 0,001$) alors que seule la dépendance au niveau de la variable dépendante est détectée par les tests robustes (RLM-error = 0,567, $p = 0,452$; RLM-lag = 19,118, $p < 0,001$). On calcule donc les trois modèles spatiaux (SEPD, SLPD et SDDP), dont les résultats sont présentés au tableau 3 avec ceux du modèle standard FE mixtes à des fins de comparaison.

On effectue ensuite les tests de LR pour vérifier si le SDPDM peut être réduit à un SLPDM ou à un SEPD. Ces deux hypothèses sont rejetées à un seuil inférieur à 0,1%: nous conservons donc le SDPDM comme modèle spatial. Finalement, la statistique d’Hausman spatiale ($\chi^2_{18} = 1\,771$; $p < 0,001$) nous amène encore une fois à préférer les FE aux RE. Le modèle final spatial le plus approprié à nos données est donc le SDPDM à FE mixtes, soit le quatrième modèle dans le tableau 3. Le SDPDM ressort d’ailleurs comme le modèle le plus explicatif par comparaison des valeurs de R² et pseudo-R².

Comparons les résultats des différents modèles (tableau 3). D’abord, quand on considère le modèle non spatial, on remarque que tous les coefficients des variables indépendantes sont en absolu plus élevés que ceux des trois modèles spatiaux. On peut y voir qu’une partie de la dépendance spatiale non modélisée par le premier modèle est alors distribuée dans les β_k , qui sont alors surestimés. Pour ce qui est des variables muettes pour les années, les effets estimés sont très similaires entre le modèle non spatial et le SEPD. Ils sont dans la plupart des cas plus élevés que ce qui est observé pour les SLPD et SDPDM. Encore une fois, on peut penser que cette dissolution des effets fixes temporels est liée à l’ajout de termes de dépendance spatiale dans ces deux modèles: Wy pour le SLPD, Wy et WX pour le SDPDM. Ainsi, contrôler les dépendances spatiales permet de mieux estimer les effets des variables d’intérêt. Dans notre meilleur modèle (SDPDM), toutes les années sauf 2016 présentent des coefficients positifs significatifs comparativement à l’année de référence (1986). En guise d’exemple, en 1996 et 2001, l’effet apparaît plus marqué, toutes choses étant égales par ailleurs, avec une hausse de 5,8 % du faible revenu par rapport à 1986. Pour ce qui est du paramètre spatial λ , son estimateur très significatif signale qu’une augmentation d’un point de pourcentage du faible revenu dans le voisinage d’un secteur de recensement fait augmenter de 0,326 le pourcentage de personnes à faible revenu dans ce même SR. Nous n’analysons pas ici les valeurs des coefficients des variables indépendantes du SDPDM, qui ne représentent pas leur influence réelle sur le faible revenu (Parajuli et Haynes, 2017; Ge *et al.*, 2018). Nous nous concentrons plutôt sur leurs impacts, dans la section suivante.

Analyse des impacts

Les résultats des simulations des impacts sont présentés au tableau 4. Rappelons que, pour notre modèle, l’impact

TABLEAU 3 Résultats pour les modèles non spatial et spatiaux | Conception: Gaboriault-Boudreau, Apparicio et Brunelle, 2019

FE mixtes (non spatial)		SEPDM ^a		SLPDM ^b		SDPDM ^c		
Coefficients spatiaux								
W_u			0,348***	(16,65)				
W_y					0,333***	(18,38)	0,326***	(15,39)
Variables indépendantes								
Chomag	0,453***	(18,63)	0,392***	(16,71)	0,390***	(16,75)	0,384***	(16,29)
FaibSc	0,305***	(20,53)	0,292***	(18,66)	0,244***	(16,70)	0,243***	(14,28)
FaMono	0,193***	(9,86)	0,176***	(8,67)	0,127***	(6,71)	0,124***	(5,77)
ImgRec	0,412***	(14,86)	0,409***	(14,07)	0,347***	(13,00)	0,386***	(12,75)
Menag1	-0,056*	(-2,43)	-0,035	(-1,47)	-0,018	(-0,80)	0,007	(0,30)
P65	0,118***	(5,15)	0,109***	(4,49)	0,055*	(2,52)	0,070**	(2,60)
Variables muettes pour les années de recensement								
1986	Réf.		Réf.		Réf.		Réf.	
1991	1,648***	(7,36)	1,620***	(5,15)	1,139***	(5,30)	1,257***	(5,30)
1996	8,051***	(29,18)	7,944***	(21,98)	5,384***	(17,91)	5,843***	(16,05)
2001	6,513***	(19,46)	6,009***	(14,37)	5,033***	(15,33)	5,856***	(13,00)
2006	-0,313	(-0,89)	-0,786	(-1,82)	0,331	(0,99)	1,569***	(3,46)
2011	1,533***	(3,97)	1,038*	(2,21)	1,547***	(4,22)	2,639***	(5,36)
2016	-3,143***	(-7,65)	-3,675***	(-7,44)	-1,378***	(-3,44)	-0,411	(-0,78)
Variables indépendantes spatialement décalées								
W x Chomag							0,183***	(3,63)
W x FaibSc							-0,044	(-1,60)
W x FaMono							-0,020	(-0,55)
W x ImgRec							-0,173***	(-3,49)
W x Menag1							-0,083	(-1,93)
W x P65							-0,005	(-0,13)
Statistiques sur la qualité d'ajustement du modèle								
R ² ou pseudo-R ²	0,7107		0,9399		0,9458		0,9461	
-2 Log-vraisemblance	24 897		24 917		24 427		24 401	
AIC	26 267		26 289		25 799		25 785	

Valeurs de T des coefficients entre parenthèses. Significativité: ***p<0,001, **p<0,01, *p<0,05

^a Spatial error panel data model. ^b Spatial lag panel data model. ^c Spatial Durbin panel data model.

direct (local) d'une variable indépendante (le chômage, par exemple) est le changement qu'elle occasionne sur le pourcentage de personnes à faible revenu dans le secteur de recensement, alors que son impact indirect (effet de débordement) est le changement qu'elle cause dans l'ensemble des autres SR. Aussi, l'impact total est-il la somme des deux impacts et représente le changement moyen de faible revenu induit par cette variable indépendante dans toute la RMR.

Pour les impacts totaux, le chômage est de loin la variable la plus importante (0,842; p<0,001), suivi par l'immigration récente (0,317; p<0,001) et la faible scolarité (0,295 p<0,001) qui sont très similaires. Par la suite, on distingue deux groupes de variables en fonction des impacts. Premièrement, le chômage est la seule variable

à avoir à la fois des impacts direct, indirect et total hautement significatifs. Son impact direct est déjà le plus élevé et significatif (0,405; p<0,001), mais son impact indirect est encore plus important (0,436; p<0,001) et vient donc plus que doubler l'impact total par rapport au coefficient du SDPDM (0,384; p<0,001). Ainsi, quand le chômage augmente dans un secteur, le faible revenu local est fortement affecté, mais cet impact se répercute aussi dans les secteurs voisins et crée un effet de débordement dans l'ensemble de la région.

Deuxièmement, les pourcentages d'immigrants récents, de personnes faiblement scolarisées et de familles monoparentales ont tous des impacts directs et totaux positifs et hautement significatifs, alors que leurs impacts indirects sont faibles et non significatifs. Concrètement,

on peut y voir que ces variables indépendantes ont un fort impact local sur le faible revenu, puis que cette première variation se répercute dans le faible revenu des voisins par le terme *Wy* hautement positif et significatif. Le même constat s’applique aussi au pourcentage de personnes âgées, mais dans une moindre mesure puisque les impacts sont plus faibles et moins significatifs. Notons que l’immigration récente voit son impact total réduit par rapport à son impact direct. Cela est dû à son impact indirect négatif non significatif quand il est considéré seul, mais qui amoindrit l’impact total quand il est évalué conjointement avec l’impact direct.

Aussi, on mesure l’effet de renvoi d’une variable comme la différence entre son impact direct (tableau 4) et son coefficient estimé avec le SDPDM (tableau 3). L’effet de renvoi le plus important est celui du chômage. S’il augmente localement de 10 points de pourcentage dans un secteur, le faible revenu y augmente initialement de 3,84 points de pourcentage (coefficient au tableau 3), toutes choses étant égales par ailleurs. Puis, les effets de débordement engendrent des augmentations du faible revenu dans le voisinage (impact indirect) qui débordent en retour sur le secteur initial. Finalement, on observe dans le secteur initial une hausse de 4,05 points de pourcentage du faible revenu, attribuable à l’augmentation du chômage (impact direct). Ainsi, l’effet de renvoi du chômage est de 0,021 point de pourcentage (0,405-0,384), soit 5,2 % de l’impact direct dans un secteur donné.

En résumé, le recours aux différents types d’impacts nous permet d’appréhender des réalités qu’on n’aurait pas pu déceler en se basant uniquement sur les résultats du SDPDM. L’identification de ces deux groupes de variables qui ont des effets différents démontre l’intérêt de recourir à l’analyse des coefficients des impacts (directs, indirects et totaux) comparativement à ceux du modèle SDPDM uniquement.

TABLEAU 4 Résultats des simulations des impacts pour le SDPDM | Conception: Gaboriault-Boudreau, Apparicio et Brunelle, 2019

	Direct		Indirect		Total	
Chomag	0,405***	(19,26)	0,436***	(6,53)	0,842***	(11,20)
FaibSc	0,245***	(16,91)	0,050	(1,48)	0,295***	(8,48)
FaMono	0,125***	(6,86)	0,029	(0,68)	0,154***	(3,43)
ImgRec	0,383***	(15,09)	-0,066	(-1,16)	0,317***	(5,47)
Menag1	0,002	(0,09)	-0,114*	(-2,28)	-0,112*	(-2,12)
P65	0,071**	(3,11)	0,025	(0,52)	0,095*	(1,97)

Valeurs de T des coefficients entre parenthèses. Significativité: ***p<0,001, **p<0,01, *p<0,05

Discussion

Apports des modèles spatiaux par panel

Les méthodes de détermination de modèles proposées dans cet article offrent une approche systématique de sélection permettant aux chercheurs de mieux spécifier leur modèle parmi un vaste ensemble de modèles spatiaux sur données de panel. Dans les paragraphes précédents, nous avons présenté les plus importantes spécificités en termes d’effets et de dépendances spatiales. Rappelons aussi que ces modèles sont relativement faciles à mettre en œuvre dans R (R Core Team, 2017) avec les paquets *plm* (Croissant et Millo, 2008) et *splm* (Millo et Piras, 2012). Pour plus de détail, le lecteur pourra télécharger le jeu de données utilisé pour cet article ainsi que le code documenté à l’adresse suivante: <http://laeq.ucs.inrs.ca/wp-content/uploads/2020/03/Spanel.zip>

L’application au cas de la pauvreté à Montréal de 1986 à 2016 nous a permis de comprendre l’importance de contrôler les effets de la dépendance spatiale dans les estimations des modèles sur données de panel. Toutefois, force est d’admettre que l’approche méthodologique de détermination du modèle proposée dans cet article reste générale, alors que plusieurs autres spécificités n’y sont pas directement abordées. D’abord, considérer uniquement les trois modèles (SLPDM, SEPDM, SDPDM) parmi les sept présentés à la figure 3 fait l’objet de débats à l’heure actuelle. Selon Lesage (2014), puisqu’il est impossible de trancher avec certitude entre un SLPDM ou un SEPDM, on devrait toujours retenir un modèle Durbin. Il faudrait alors seulement choisir entre le SDEPDM (figure 3.d) et le SDPDM (figure 3.c), le premier modèle prévalant si l’on soupçonne la présence d’effets de débordement locaux (entre voisins proches), et le second si les débordements attendus sont globaux et s’étendent à tout le territoire.

Quatre autres éléments spécifiques devraient aussi être considérés pour mieux adapter et étendre les modèles présentés aux besoins spécifiques des chercheurs, soit la méthode d'estimation, la dimension spatiale, la temporalité et les effets dynamiques. La méthode d'estimation reste un aspect important à évaluer pour adapter les modèles présentés à des contextes spécifiques. Nous avons appuyé notre approche sur les estimateurs par le maximum de vraisemblance, mais il est possible d'estimer ces modèles par la méthode des moments généralisée (MMG), elle aussi facile à mettre en œuvre dans R. Un aspect à considérer dans le choix entre ces deux estimateurs touche les hypothèses de normalité des résidus. Comme le soulignent Bouayad *et al.* (2018), les estimateurs par le maximum de vraisemblance sont généralement plus efficaces, mais reposent sur des conditions plus fortes quant à la distribution du terme d'erreurs. De fait, la MMG sera généralement privilégiée en présence d'hétéroscédasticité—cette dernière ne reposant pas sur l'hypothèse de normalité et offrant des estimateurs plus robustes dans ce contexte. L'utilisation de modèles bayésiens est également possible (Lesage, 2014).

Un second aspect qui devrait être considéré touche la nature des relations spatiales entre les entités étudiées et le choix de la matrice de pondération spatiale. Nous avons utilisé ici une matrice de contiguïté selon le partage d'une frontière commune de premier ordre. Il faut garder en tête que d'autres applications pourraient être mieux servies par des matrices de distance, de plus proches voisins ou d'interaction. Il est important de bien évaluer le type de données et la nature des effets de dépendances spatiales attendues lorsque ce choix est fait, étant donné que le choix du type de matrice de pondération fait varier les résultats du modèle estimé.

En outre, la temporalité est une composante fondamentale de l'hétérogénéité captée par les modèles en panel. Bien qu'on puisse vouloir se limiter à évacuer les effets conjoncturels de nos estimations, il semble important de se questionner sur la nature des dynamiques temporelles en jeu, notamment la présence d'effets de rétroaction ou d'autocorrélation temporelle. Ces derniers sont d'autant plus importants dans un contexte spatial, où les interactions entre entités voisines peuvent être dynamiques au fil du temps. Comme le mentionnent Kopczyńska *et al.* (2017), certaines décisions doivent ultimement être prises concernant la dimension temporelle. Outre les retards spatiaux, les modèles peuvent inclure des retards temporels et spatiotemporels d'une variable dépendante

et indépendante. Les phénomènes géographiques étant souvent persistants dans le temps et dans l'espace, il est crucial de séparer ces persistances de la variabilité actuelle due à d'autres facteurs (Elhorst, 2012). Afin d'intégrer ces effets, plusieurs extensions ont été plus récemment proposées aux modèles d'économétrie spatiale sur données de panel. Bouayad-Agha *et al.* (2018) mentionnent les modèles dynamiques spatiaux puisque « les valeurs pour une observation i à une période de temps t peuvent dépendre des valeurs prises par les observations voisines de i à la période précédente ». Ils mentionnent aussi les modèles multidimensionnels spatiaux qui traitent de données groupées ou hiérarchisées, ainsi que les modèles à facteurs communs qui permettent plus de flexibilité dans les interactions entre les unités dans le temps et l'espace.

Limites relatives à l'utilisation des modèles spatiaux par panel sur des données agrégées spatialement

Comme pour toute autre analyse statistique réalisée sur des données agrégées spatialement, l'échelle d'analyse et le découpage des entités spatiales génèrent des biais potentiels (*modifiable areal unit problem* [MAUP]) (Openshaw, 1977; Wong et Amrhein, 1996) dans les résultats des modèles spatiaux par panel. Concernant l'échelle d'analyse, il est bien connu que plus les unités spatiales sont grandes, moins la variation est importante (Jelinski et Wu, 1996) et donc plus la capacité à mesurer l'hétérogénéité sociale des milieux est réduite. Rappelons que les analyses ont été réalisées à partir de données au niveau de secteurs de recensement, de 1986 à 2016. Il aurait été très difficile de recourir à une échelle d'analyse plus fine telle que les aires de diffusion (AD), regroupant habituellement entre 400 et 700 habitants. En effet, les AD ont été créées en 2001 pour remplacer les secteurs de dénombrement (SD). Il en résulte que la création de fichiers de correspondance entre les AD et les SD pour sept années de recensement est beaucoup plus complexe et chronophage.

Concernant le problème de découpage, parmi les règles utilisées pour les délimiter, les SR doivent comprendre entre 2 500 et 10 000 habitants et doivent être le plus homogènes possible sur le plan socioéconomique (Statistique Canada, 2018). Or, nous avons agrégé certains SR adjacents afin d'obtenir le même nombre de SR pour les sept années de recensement. Cela cause deux biais potentiels. Premièrement, il est possible que ces secteurs agrégés ne soient plus si homogènes sur le plan socioéconomique pour les sept années de recensement.

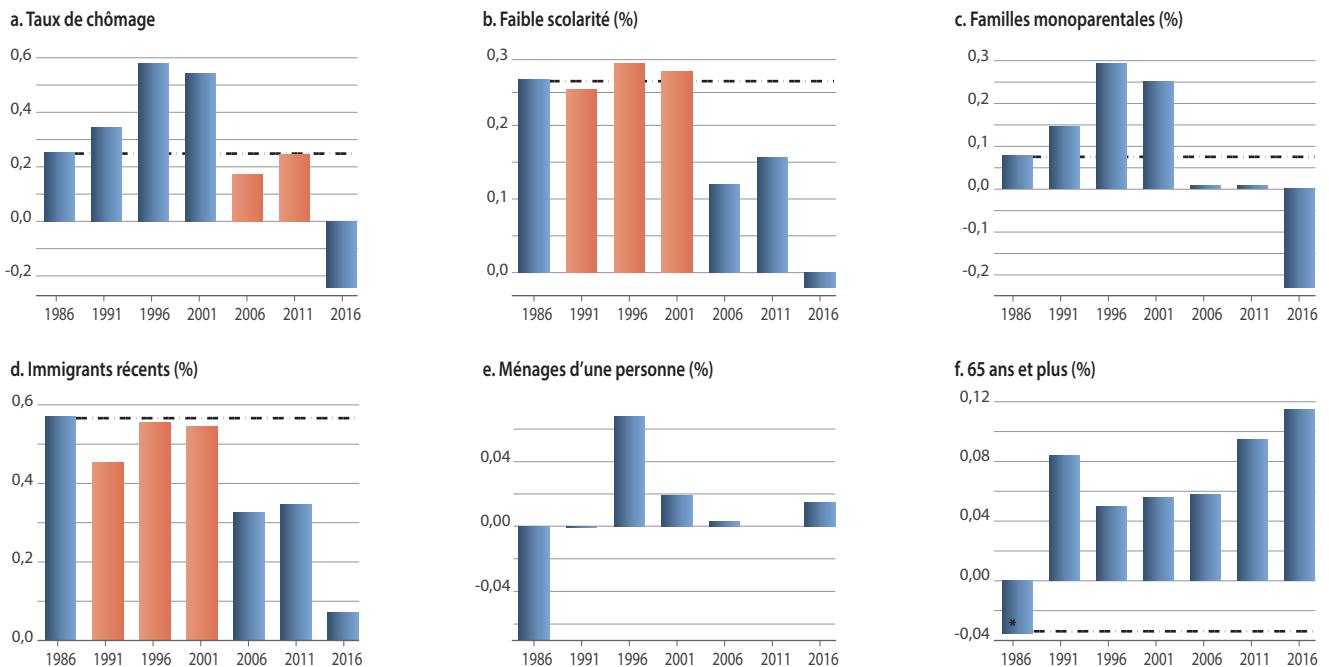
Deuxièmement, les effectifs de population de secteurs agrégés risquent de dépasser largement le seuil de 10 000 habitants. Pour y remédier, il conviendrait alors de construire des modèles de régression par panel en pondérant chaque entité spatiale par la population qu'elle comprend pour les différentes années de recensement.

Limites relatives à la formulation des modèles présentés

Il convient aussi de signaler que la dimension spatiotemporelle a été peu explorée dans les modèles présentés, notamment le modèle final (SDPDM, tableau 3). Par exemple, il est fort probable que l'effet du chômage (Chomag) et de sa variable spatialement décalée (W_Chomag) varie à travers le temps, et ce, d'autant plus que notre jeu de données couvre une période de 30 ans (1986 à 2016). Par conséquent, il est possible de bonifier le modèle SDPDM en ajoutant des interactions entre les années de recensement et respectivement chacune des variables indépendantes et de leurs versions spatiales décalées (par exemple, AN x Chomag et AN x WChomag). Cela a toutefois comme corolaire d'ajouter un nombre important de coefficients et de générer possiblement des problèmes de multicollinéarité excessive.

Afin d'explorer la dimension spatiotemporelle, nous proposons un exercice très simple: nous avons construit six nouveaux modèles – un pour chaque variable indépendante – dans lesquels nous avons ajouté, dans le modèle SDPDM final, une interaction entre la variable indépendante et l'année. Les résultats de ces interactions sont présentés graphiquement à la figure 5. Prenons l'exemple de la variable immigrants récents (figure 5.a). En 1986 (année de référence), son effet est de 0,294 ($P < 0,001$). La taille de cet effet reste relativement similaire pour les années 1991 à 2006 (les coefficients obtenus n'étant pas significativement différents au seuil de 5%), tandis qu'à partir de 2006, son effet diminue significativement. Ce constat s'applique aussi pour le pourcentage de personnes faiblement scolarisées (figure 5.b). Cela signifie que les pourcentages d'immigrants récents et de personnes faiblement scolarisées sont de moins en moins importants dans la prédiction du pourcentage de personnes à faible revenu à partir de 2006. Inversement, l'effet du pourcentage de personnes âgées de 65 ans et plus ne cesse d'augmenter d'année en année.

FIGURE 5 Exploration des effets spatiotemporels des variables indépendantes | Conception: Gaboriault-Boudreau, Apparicio et



Effet de la variable indépendante par année sur le pourcentage de personnes à faible revenu

■ Différence significative par rapport à l'année de référence (1986) ($P < 0,05$) ■ Différence non significative * Coefficient non significatif en 1986

Conclusion

Cet article contribue à la littérature émergente en géographie portant sur les méthodes d'économétrie spatiale sur données de panel. En appliquant cette méthode au cas de la pauvreté urbaine dans la région métropolitaine de Montréal, entre 1986 et 2016, nous obtenons des résultats qui soulignent le potentiel des données de panel pour l'élaboration de diagnostics urbains longitudinaux. Spécifiquement, nous proposons, dans cet article, une méthode générale de détermination et de sélection permettant aux chercheurs de mieux spécifier leur modèle parmi un vaste ensemble de modèles récemment développés. Si les avantages des données de panel relativement aux données en coupe transversale sont aujourd'hui bien connus, notamment le contrôle des effets conjoncturels et locaux inobservés et des estimations plus précises, ces derniers sont tout aussi nécessaires dans le contexte des modèles spatiaux.

De fait, les recherches futures en géographie pourraient fortement bénéficier des méthodes d'économétrie spatiale sur données de panel, alors que de nombreuses sources de données utilisées par les géographes – dont les données de recensement – se présentent sous forme de panel avec certaines entités statistiques (régions, villes, quartiers) répétées au fil du temps. En outre, la disponibilité et le développement accéléré de nouveaux paquets dans R permettront aux chercheurs d'appliquer les modèles présentés dans cet article à un vaste ensemble de problématiques et d'être mieux préparés aux développements accélérés des nouvelles méthodes dans ce domaine foisonnant. Ces méthodes pourraient être utilisées pour poser des diagnostics d'équité environnementale longitudinaux, notamment pour analyser la distribution de la végétation, ou encore de nuisances urbaines (pollution de l'air, bruit environnemental, etc.) sur plusieurs années. Également, cela ouvre la voie au développement de nouvelles méthodes qui permettront de poser des diagnostics intra-urbains longitudinaux, comme les méthodes de régression géographiquement pondérée par panel.

Remerciements

Les auteurs remercient Jérémy Gelb, les évaluateurs et évaluatrices pour leurs suggestions et commentaires judicieux sur la première version du manuscrit. Cette recherche a bénéficié du soutien financier de la Chaire de recherche du Canada sur l'équité environnementale et la Ville (950-230813).

Bibliographie

- ADES, Josefina, APPARICIO, Philippe et SÉGUIN, Anne Marie (2012) Are new patterns of low income distribution emerging in Canadian metropolitan areas? *The Canadian Geographer*, vol. 56, n°3, p. 339-361.
- ANSELIN, Luc (1988) *Spatial econometrics: Methods and models*. Dordrecht, Kluwer Academic Publishers.
- APPARICIO, Philippe, CARRIER, Mathieu, GELB, Jérémy, SÉGUIN, Anne-Marie et KINGHAM, Simon (2016) Cyclists' exposure to air pollution and road traffic noise in central city neighbourhoods of Montreal. *Journal of Transport Geography*, vol. 57, p. 63-69.
- APPARICIO, Philippe, SÉGUIN, Anne-Marie et LELOUP, Xavier (2007) Modélisation spatiale de la pauvreté à Montréal: apport méthodologique de la régression géographiquement pondérée. *Le Géographe canadien*, vol. 51, n°4, p. 412-427.
- BALTAGI, Badi H. (2013) *Econometric analysis of panel data*. Hoboken, John Wiley and Sons, Inc.
- BIVAND, Roger, PEBESMA, Edzer J. et GÓMEZ-RUBIO, Virgilio (2013) *Applied spatial data analysis with R*. New York, Springer.
- BOUAYAD-AGHA, Salima, LE GALLO, Julie et VEDRINE, Lionel (2018) Économétrie spatiale sur données de panel. Dans Vincent Loonis (dir.) *Manuel d'analyse spatiale. Théorie et mise en oeuvre pratique avec R*. Paris, Institut national de la statistique et des études économiques, p. 183-209.
- BURNETT, J. Wesley, BERGSTROM, John C. et DORFMAN, Jeffrey H. (2013) A spatial panel data approach to estimating US state-level energy emissions. *Energy Economics*, vol. 40, n°C, p. 396-404.
- CAN, Ayse (1996) Weight matrices and spatial autocorrelation statistics using a topological vector data model. *International Journal of Geographical Information Systems*, vol. 10, n°10, p. 1009-1017.

- CARRIER, Mathieu, APPARICIO, Philippe, SÉGUIN, Anne-Marie et CROUSE, Dan (2014) The application of three methods to measure the statistical association between different social groups and the concentration of air pollutants in Montreal: A case of environmental equity. *Transportation Research Part D: Transport and Environment*, vol. 30, p. 38-52.
- CARRIER, Mathieu, APPARICIO, Philippe et SÉGUIN, Anne-Marie (2016a) Road traffic noise in Montreal and environmental equity: What is the situation for the most vulnerable population groups? *Journal of Transport Geography*, vol. 51, p. 1-8.
- CARRIER, Mathieu, APPARICIO, Philippe et SÉGUIN, Anne-Marie (2016b) Road traffic noise geography during the night in Montreal: An environmental equity assessment. *The Canadian Geographer*, vol. 60, n°3, p. 394-405.
- CARRIER, Mathieu, APPARICIO, Philippe, KESTENS, Yan, SÉGUIN, Anne-Marie, PHAM, Hien *et al.* (2016c) Application of a Global Environmental Equity Index in Montreal: Diagnostic and further implications. *Annals of the American Association of Geographers*, vol. 106, n°6, p. 1268-1285.
- CHATTERJEE, Samprit et HADI, Ali (2015) *Regression analysis by example*. Hoboken, John Wiley & Sons, Inc.
- CRESSIE, Noel (2015) *Statistics for spatial data*. Hoboken, John Wiley & Sons, Inc, [1993].
- CROISSANT, Yves et MILLO, Giovanni (2008) Panel data econometrics in R: The plm package. *Journal of Statistical Software*, vol. 27, n°2, p. 1-43.
- DEVAUX, Nicolas, DUBÉ, Jean et APPARICIO, Philippe (2017) Anticipation and post-construction impact of a metro extension on residential values: The case of Laval (Canada), 1995–2013. *Journal of Transport Geography*, vol. 62, p. 8-19.
- DORMANN, Carsten, MCPHERSON, Jana, ARAÚJO, Miguel, BIVAND, Roger, BOLLIGER, Janine *et al.* (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data : A review. *Ecography*, vol. 30, n°5, p. 609-628.
- DUBÉ, Jean et LEGROS, Diègo (2014a) Spatial econometrics and the hedonic pricing model: What about the temporal dimension? *Journal of Property Research*, vol. 31, n°4, p. 333-359.
- DUBÉ, Jean et LEGROS, Diègo (2014b) *Économétrie spatiale appliquée des microdonnées*. Londres, ISTE Group.
- DUBÉ, Jean, LEGROS, Diègo, THÉRIAULT, Marius et DES ROSIERS, François (2014) A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices. *Transportation Research Part B: Methodological*, vol. 64, p. 24-40.
- DUBÉ, Jean et POLÈSE, Mario (2015) The view from a lucky country: Explaining the localised unemployment impacts of the Great Recession in Canada. *Cambridge Journal of Regions, Economy and Society*, vol. 9, n°1, p. 235-253.
- DUBIN, Robin (2008) Spatial weights. Dans A. Stewart Fotheringham et Peter A. Rogerson (dir.) *The SAGE handbook of spatial analysis*. Thousand Oaks, SAGE Publications, p. 125-158.
- DUVIVIER, Chloé, POLÈSE, Mario et APPARICIO, Philippe (2018) The location of information technology-led new economy jobs in cities: Office parks or cool neighbourhoods? *Regional Studies*, vol. 52, n°6, p. 756-767.
- ELHORST, J. Paul (2009) Spatial panel data models. Dans Manfred M. Fischer et Arthur Getis (dir.) *Handbook of applied spatial analysis: Software tools, methods and applications*. Berlin, Springer Science & Business Media, p. 377-407.
- ELHORST, J. Paul (2010) Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, vol. 5, n°1, p. 9-28.
- ELHORST, J. Paul (2012) Matlab software for spatial panels. *International Regional Science Review*, vol. 37, n°3, p. 389-405.
- ELHORST, J. Paul (2014) *Spatial econometrics: From cross-sectional data to spatial panels*. New York, Springer Publishing.
- JELINSKI, Dennis E. et WU, Jianguo (1996) The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, vol. 11, n°3, p. 129-140.
- FOTHERINGHAM, Stewart, BRUNSDON, Chris et CHARLTON, Martin (2003) *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, John Wiley & Sons, Inc.
- GE, Xiangyu, ZHOU, Zhimin, ZHOU, Yanli, YE, Xinyue et LIU, Songlin (2018) A spatial panel data analysis of economic growth, urbanization, and NOx emissions in China. *International Journal of Environmental Research and Public Health*, vol. 15, n°4, p. 725-745.
- GOUVERNEMENT DU CANADA (2016) *Document d'information sur la pauvreté au Canada*. Ottawa, Gouvernement du Canada.
- GRIFFITH, Daniel (2013) *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Berlin, Springer Science & Business Media.
- GRIFFITH, Daniel et PERES-NETO, Pedro (2006) Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. *Ecology*, vol. 87, n°10, p. 2603-2613.

- GROULX, Lionel-Henri (2011) *Les facteurs engendrant l'exclusion au Canada: survol de la littérature multidisciplinaire*. Québec, Centre d'étude sur la pauvreté et l'exclusion.
- HAINING, Robert (2003) *Spatial data analysis: Theory and practice*. Cambridge (Angleterre), Cambridge University Press.
- HAUSMAN, Jerry, HALL, Bronwyn H. et GRILICHES, Zvi (1984) *Econometric models for count data with an application to the patents-R&D relationship*. Technical working paper, n°17, Cambridge (États-Unis), National Bureau of Economic Research.
- HEISZ, Andrew et MCLEOD, Logan (2004) *Faible revenu dans les régions métropolitaines de recensement, 1980 à 2000*. Ottawa, Statistique Canada.
- HUANG, Qiong et CHAND, Satish (2015) Spatial spillovers of regional wages: Evidence from Chinese provinces. *China Economic Review*, vol. 32, p. 97-109.
- HUANG, Jianhuan et XIA, Jiejun (2016) Regional competition, heterogeneous factors and pollution intensity in China: A spatial econometric analysis. *Sustainability*, vol. 8, n°171, p. 1-26.
- KEITT, Timothy H., BJØRNSTAD, Ottar, DIXON, Philip M. et CITRON POUSTY, Steve (2002) Accounting for spatial pattern when modeling organism environment interactions. *Ecography*, vol. 25, n°5, p. 616-625.
- KESTENS, Yan, THÉRIAULT, Marius et DES ROSIERS, François (2006) Heterogeneity in hedonic modelling of house prices: Looking at buyers' household profiles. *Journal of Geographical Systems*, vol. 8, n°1, p. 61-96.
- KOPCZEWSKA, Katarzyna, KUDŁA, Janusz et WALCZYK, Konrad (2017) Strategy of spatial panel estimation: Spatial spillovers between taxation and economic growth. *Applied Spatial Analysis and Policy*, vol. 10, n°1, p. 77-102.
- LEE, Kevin Kaalip (2000) *Urban poverty in Canada: A statistical profile*. Ottawa, Canadian Council on Social.
- LELOUP, Xavier (2007) Vers la ville pluraliste ? Distribution et localisation des minorités visibles à Montréal, Toronto et Vancouver en 2001. *Canadian Journal of Regional Science*, vol. 30, n°2, p. 263-292.
- LESAGE, James (2014) Spatial econometric panel data model specification: A Bayesian approach. *Spatial Statistics*, vol. 9, p. 122-145.
- LESAGE, James et PACE, Kelly (2009) *Introduction to spatial econometrics*. Boca Raton, Chemical Rubber Company Press.
- LIU, Yu, XIAO, Hongwei et ZHANG, Ning (2016) Industrial carbon emissions of China's regions: A spatial econometric analysis. *Sustainability*, vol. 8, n°3, p. 210-224.
- MILLO, Giovanni et PIRAS, Gianfranco (2012) splm: Spatial panel data models in R. *Journal of Statistical Software*, vol. 47, n°1, p. 1-38.
- OPENSHAW, Stan (1977) Optimal zoning systems for spatial interaction models. *Environment and planning A*, vol. 9, n°2, p. 169-184.
- PARAJULI, Jitendra et HAYNES, Kingsley (2017) Panel data models of new firm formation in New England. *Region: The Journal of ERSA*, vol. 4, n°3, p. 65-76.
- PFARR, Christian, SCHMID, Andreas et SCHNEIDER, Udo (2010) *Estimating ordered categorical variables using panel data: A generalized ordered probit model with an autofit procedure*. Paper n°24181, Munich, Munich Personal RePEc Archive.
- PHAM, Thi Thanh Hiên, APPARICIO, Philippe, SÉGUIN, Anne-Marie, LANDRY, Shawn et GAGNON, Martin (2012) Spatial distribution of vegetation in Montreal: An uneven distribution or environmental inequity? *Landscape and Urban Planning*, vol. 107, n°3, p. 214-224.
- PHAM, Thi Thanh Hiên, APPARICIO, Philippe, LANDRY, Shawn, SÉGUIN, Anne-Marie et GAGNON, Martin (2013) Predictors of the distribution of street and backyard vegetation in Montreal, Canada. *Urban forestry & Urban Greening*, vol. 12, n°1, p. 18-27.
- PLUMMER, Lawrence A. (2009) Spatial dependence in entrepreneurship research: Challenges and methods. *Organizational Research Methods*, vol. 13, n°1, p. 146-175.
- QIAN, Haifeng et ZHAO, Chang (2018) Space-time analysis of high technology entrepreneurship: A comparison of California and New England. *Applied Geography*, vol. 95, n°1, p. 111-119.
- RUE, Havard et HELD, Leonhard (2005) *Gaussian Markov random fields: Theory and applications*. Londres, Chapman and Hall/CRC Press.
- SÉGUIN, Anne-Marie, APPARICIO, Philippe et RIVA, Mylène (2012) Identifying, mapping and modelling trajectories of poverty at the neighbourhood level: The case of Montréal, 1986-2006. *Applied Geography*, vol. 35, n°1-2, p. 265-274.
- SÉGUIN, Anne-Marie, APPARICIO, Philippe, RIVA, Mylène et NEGRON-POBLETE, Paula (2016) The changing spatial distribution of Montreal seniors at the neighbourhood level: A trajectory analysis. *Housing Studies*, vol. 31, n°1, p. 61-80.
- STATISTIQUE CANADA (2010) *Recensement de la population, 1986*. N°97-550-XWF2006001 au catalogue, Ottawa, Gouvernement du Canada.
- STATISTIQUE CANADA (2018) *Dictionnaire. Recensement de la population, 2016*. N°98-301-X2016001 au catalogue, Ottawa, Gouvernement du Canada.

- STUDENMUND, Arnold et CASSIDY, Henry (2016) *Using econometrics: A practical guide*. Londres, Pearson.
- R CORE TEAM (2017) *R: A language and environment for statistical computing*. Vienne, R Foundation for Statistical Computing.
- THANOS, Sotirios, DUBÉ, Jean et LEGROS, Diègo (2016) Putting time into space: The temporal coherence of spatial applications in the housing market. *Regional Science and Urban Economics*, vol. 58, p. 78-88.
- TONG, Tingting, YU, Tun-Hsiang Edward, CHO, Seong-Hoon, JENSEN, Kimberly et DE LA TORRE UGARTE, Daniel (2013) Evaluating the spatial spillover effects of transportation infrastructure on agricultural output across the United States. *Journal of Transport Geography*, vol. 30, p. 47-55.
- WONG, David et AMRHEIN, Carl (1996) Research on the MAUP: Old wine in a new bottle or real breakthrough? *Journal of Geographical Systems*, vol. 3, n°2, p. 73-76.
- WOOD, Simon N. (2017) *Generalized additive models: An introduction with R*. Londres, Chapman and Hall/CRC.
- WOOLDRIDGE, Jeffrey M. (2002) *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts Institute of Technology Press.
- YU, Nannan, DE JONG, Martin, STORM, Servaas et MI, Jianing (2013) Spatial spillover effects of transport infrastructure: Evidence from Chinese regions. *Journal of Transport Geography*, vol. 28, p. 56-66.
- ZUUR, Alain F., HILBE, Joseph M. et IENO, Elena N. (2013) *A beginner's guide to GLM and GLMM with R: A frequentist and Bayesian perspective for ecologists*. Newburg, Highland Statistics Limited.