

Rapports sur la santé

Précision de localisation du géocodage à l'aide des codes postaux résidentiels comparativement à celle du géocodage fondé sur les adresses municipales complètes

par Saeeda Khan, Lauren Pinault, Michael Tjepkema
et Russell Wilkins

Date de diffusion : le 21 février 2018



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

Programme des services de dépôt

- | | |
|-----------------------------|----------------|
| • Service de renseignements | 1-800-635-7943 |
| • Télécopieur | 1-800-565-7757 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2018

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Précision de localisation du géocodage à l'aide des codes postaux résidentiels comparativement à celle du géocodage fondé sur les adresses municipales complètes

par Saeeda Khan, Lauren Pinault, Michael Tjepkema et Russell Wilkins

Résumé

Contexte : Les codes postaux sont souvent les seuls identificateurs géographiques dont nous disposons pour attribuer des renseignements contextuels ou environnementaux à une population à l'étude. La présente analyse évalue l'influence de trois facteurs : le type de mode de livraison (mode de livraison postale), le type de point représentatif (source des coordonnées de latitude et de longitude) et la taille de la collectivité, sur la précision de l'attribution spatiale à partir des codes postaux.

Données et méthodes : Le Fichier de conversion des codes postaux plus (FCCP+) a été utilisé pour attribuer un type de mode de livraison, un type de point représentatif et une taille de collectivité à tous les répondants du Recensement du Canada de 2011. Un échantillon (n = 1 004) a été sélectionné au hasard et comportait un minimum de 90 observations pour chacune des catégories de ces trois facteurs. D'après les adresses des personnes comprises dans l'échantillon, des mesures de la précision de localisation du géocodage à l'aide des codes postaux résidentiels (FCCP+), par opposition au géocodage fondé sur des emplacements de référence déterminés en fonction d'adresses municipales complètes (Google Maps), ont été établies à l'aide d'un système d'information géographique. La précision a été mesurée d'après la distance qui sépare la position géocodée et l'adresse municipale complète.

Résultats : La précision de localisation était principalement liée au type de mode de livraison. Les modes de livraison ruraux et mixtes (partiellement urbains, partiellement ruraux) présentaient un nombre beaucoup plus élevé d'erreurs de géocodage que les modes urbains. Les régions rurales et les petites villes du Canada ainsi que les latitudes et longitudes fondées sur les centroïdes d'aires de diffusion présentaient un faible niveau de précision, principalement parce qu'elles étaient étroitement liées aux modes de livraison ruraux et mixtes.

Interprétation : La précision du géocodage à l'aide des codes postaux peut varier. L'imprécision du géocodage peut entraîner une classification erronée, selon la résolution spatiale des mesures environnementales et contextuelles. La résolution spatiale nécessaire pour mener une étude aide à déterminer les sous-populations qui doivent être exclues en raison d'un niveau de précision de localisation inadéquat.

Mots clés : Taille de la collectivité, type de mode de livraison, classification erronée, type de point représentatif, adresse municipale

Le géocodage fondé sur une adresse municipale complète est une façon très précise d'attribuer des coordonnées géographiques au lieu de résidence d'une personne. Or, le plus souvent, les adresses ne sont pas disponibles. En outre, pour bon nombre de recherches sur la santé de la population, un tel niveau de précision n'est pas nécessaire.

La précision de localisation nécessaire pour mener une étude dépend du contexte de l'étude. Lorsqu'il s'agit d'attribuer des caractéristiques, telles que des variables socioéconomiques de quartiers, ou d'obtenir des données de recensement servant de dénominateur pour des populations à risque définies géographiquement, un niveau de précision raisonnable est généralement suffisant. En revanche, pour certaines recherches sur l'hygiène de l'environnement, telles que l'attribution de scores de potentiel piétonnier pour les résidents de centres urbains, une résolution spatiale très précise peut être nécessaire, car même une erreur d'à peine quelques centaines de mètres peut entraîner une classification fortement erronée de l'exposition¹⁻⁵.

Au Canada, les codes postaux sont souvent les seuls identificateurs géographiques dont nous disposons pour attribuer des renseignements contextuels ou environnementaux à des données administratives ou à des données d'enquête. Le Fichier de conversion des codes postaux (FCCP)⁶ et le Fichier de conversion des codes postaux plus (FCCP+)⁷ fournissent les moyens

requis pour géocoder des ensembles de données uniquement à l'aide des codes postaux. À ce jour, des centaines d'études ayant fait appel à ces outils ont été publiées⁸.

La fonction première des outils du FCCP est d'attribuer la hiérarchie complète de la géographie du recensement (îlot de diffusion, aire de diffusion, secteur de recensement, région métropolitaine de recensement ou agglomération de recensement, subdivision de recensement, division de recensement et province)⁹. Ces produits ont également pour fonction secondaire de permettre l'attribution de coordonnées de latitude et de longitude fondées sur les centroïdes de côtés d'îlot, d'îlots de diffusion ou d'aires de diffusion.

La présente étude compare la précision de localisation du géocodage à l'aide des codes postaux à celle du géocodage fondé sur des centroïdes représentatifs des toits établis à partir d'adresses municipales complètes (« emplacements de référence »). L'analyse repose sur des données sur les adresses autodéclarées tirées du Recensement de la population de 2011. Tous les codes postaux déclarés ont été traités à l'aide de la version 6C du FCCP+ afin d'obtenir trois variables qui influent sur la précision du géocodage à l'aide des codes postaux : le type de mode de livraison (TML), le type de point représentatif et la taille de la collectivité. La précision de localisation est mesurée d'après la distance qui sépare la position géocodée des coordonnées de latitude et de longitude de l'emplacement de référence.

Type de mode de livraison (TML)

Le TML est une caractéristique des codes postaux qui indique le mode de livraison du courrier. Les TML peuvent être regroupés en trois catégories principales : urbains (A, B, E, G, M), ruraux (W) et mixtes (H, J, K, T, X). Les TML urbains peuvent être résidentiel (A ou B) ou commercial et institutionnel (E, G et M). Les codes postaux retirés (Z) ont été exclus de l'étude, tout comme les codes postaux desservis par les TML J (livraison générale à un bureau de poste urbain) ou X (itinéraires motorisés dans les zones industrielles), car leur présence dans le recensement était rare.

TML urbains Le TML A est le code associé aux ménages urbains ordinaires et comprend un service de livraison par facteur directement au ménage ou dans des boîtes postales communautaires. Le TML B est celui associé aux grands immeubles d'habitation urbains; le TML E est associé aux entreprises urbaines; le TML G est associé aux grands établissements commerciaux et institutionnels urbains qui comportent un service de livraison par facteur; et le TML M, est associé aux grands établissements commerciaux et institutionnels urbains qui récupèrent leur courrier dans une case postale ou un sac à un bureau de poste urbain. La plupart des codes postaux urbains sont liés à des points représentatifs (centroïdes) de côtés d'îlot ou d'îlots de diffusion, habituellement par un seul lien, mais parfois jusqu'à quatre.

TML ruraux Les codes postaux dont le deuxième caractère est un « 0 » (TML W) désignent les bureaux de poste ruraux, lesquels peuvent desservir une vaste région, comprenant de multiples aires de diffusion et subdivisions de recensement (cantons ou villages), et tous les types d'utilisateurs, y compris des immeubles d'habitation, des entreprises et des établissements institutionnels. Le TML W dessert également les régions éloignées et comprend les bureaux du Service aérien omnibus¹⁰, lesquels ont été exclus de l'échantillon. La plupart des codes postaux ruraux sont liés à un grand nombre de points repré-

sentatifs (centroïdes de multiples aires de diffusion).

TML mixtes Les TML H, K et T sont mixtes (partiellement urbains, partiellement ruraux). Les TML H et T sont associés aux routes rurales et suburbaines qui partent d'un bureau de poste urbain et se prolongent jusque dans les régions rurales. Le TML K (petites cases postales) peut desservir des clients urbains ou ruraux, mais à un bureau de poste urbain. La plupart des codes postaux mixtes sont liés à de multiples centroïdes d'aires de diffusion.

Type de point représentatif

Le type de point représentatif désigne la source des coordonnées de latitude et de longitude attribuées, par ordre décroissant de précision : côté d'îlot, îlot de diffusion, aire de diffusion ou subdivision de recensement. Les points représentatifs de subdivisions de recensement sont attribués à titre de mesure provisoire dans l'attente de plus amples renseignements sur les adresses correspondant aux codes postaux et sont peu fréquents; les codes postaux associés à ce type de point représentatif ont été exclus.

Taille de la collectivité

La taille de la collectivité correspond à des catégories de la population du Recensement de 2011 : 1 500 000 personnes ou plus (régions métropolitaines de recensement [RMR] de Toronto, Montréal et Vancouver); 500 000 à 1 499 999 personnes (RMR d'Ottawa-Gatineau, Edmonton, Calgary, Québec, Winnipeg et Hamilton); 100 000 à 499 999 personnes (18 autres RMR et 7 grandes agglomérations de recensement [AC]); 10 000 à 99 999 personnes (autres AC); moins de 10 000 personnes (régions rurales et petites villes du Canada, c'est-à-dire toutes les régions à l'exception des RMR et des AC).

Sélection de l'échantillon

L'échantillon d'analyse (n = 1 004) a été sélectionné au hasard à partir de l'ensemble de données du recensement, et vise à retenir 100 observations pour chaque TML, type de point représentatif et catégorie de taille de collectivité (tableau 1). Un effort a été fait pour que chaque région (provinces de l'Atlantique, Ontario, Québec, Prairies, Colombie-Britannique et territoires du Nord) soit représentée en parts égales dans l'échantillon. Chaque fois que cela a été possible, les mêmes observations ont été réutilisées d'un TML à l'autre, d'un type de point représentatif à l'autre et d'une catégorie de taille de collectivité à l'autre.

Dans le cadre de la présente étude, les codes postaux desservis par les modes ruraux et mixtes, les types de points représentatifs d'aires de diffusion moins précises et les petites agglomérations de recensement ainsi que les régions rurales et petites villes du Canada (population de moins de 10 000 personnes) ont intentionnellement fait l'objet d'un suréchantillonnage.

Géocodage à partir d'adresses municipales complètes

Pour déterminer les emplacements de référence, toutes les adresses municipales de l'échantillon ont fait l'objet d'une recherche manuelle dans Google Maps à l'aide de la fonction d'imagerie par satellite. La latitude et la longitude du centroïde de l'emplacement de l'immeuble ont ainsi été obtenues. Chaque fois que cela a été possible, l'adresse de l'immeuble a été visuellement confirmée à l'aide de Google Street View.

Les adresses n'ayant pas été localisées pendant le géocodage manuel soit parce que l'imagerie par satellite était médiocre, soit parce que l'adresse municipale était manquante et qu'aucune vue de la rue n'était disponible, ont été exclues et de nouvelles adresses ont été rééchantillonnées (n = 121). Les enregistrements qui ont nécessité un réé-

Précision de localisation du géocodage à l'aide des codes postaux résidentiels comparativement à celle du géocodage fondé sur les adresses municipales complètes • Coup d'œil méthodologique

chantillonnage tendaient à être associés à des régions rurales ou éloignées ou à des réserves indiennes.

Dans la mesure du possible, les codes postaux déclarés ont été vérifiés sur le site Web de Postes Canada¹¹. Exception faite des codes postaux desservis par un mode de livraison mixte (TML H ou T), lorsque les trois premiers caractères d'un code postal ne correspondaient pas au code postal prévu, le code postal était considéré comme erroné et était exclu de l'ensemble de données d'analyse.

Toutes les adresses de l'échantillon ont été appariées à tous les points représentatifs de codes postaux possibles contenus dans le FCCP+. La plupart des adresses avaient de multiples points représentatifs possibles. Sur les 105 enregistrements pour lesquels une très grande distance (plus de 50 km) a été constatée entre le point représentatif tiré du FCCP+ et l'emplacement de l'adresse selon Google Maps, 51 ont été exclus, parce que le code postal ou l'adresse municipale semblaient contenir des erreurs typographiques.

Calcul de la distance

Pour chacun des enregistrements de l'échantillon, tous les points représentatifs possibles ont été extraits du FCCP+ 6C. Tous ont conservé leurs poids, en commençant par ceux auxquels on avait attribué des poids représentatifs de la population : les codes postaux ruraux (TML W), les codes postaux mixtes (TML H, K ou T), les codes postaux urbains (TML A, B, E, G ou M) présentant des appariements à quatre aires de diffusion ou plus, et les réserves indiennes partiellement dénombrees. Les codes postaux présentant un appariement à un seul point représentatif ont été attribués à ce point et ont reçu un poids de 1. Les codes postaux urbains présentant des appariements à de multiples points représentatifs ont reçu des poids équivalents. Cette stratégie de pondération est à l'image du processus d'attribution des emplacements de référence utilisé pour le FCCP+.

Les emplacements de référence déterminés manuellement en fonction

des adresses municipales complètes et des points représentatifs générés par le FCCP+ ont été mis en correspondance dans les systèmes d'information géographique (ArcGIS v.10, ESRI 2010). La distance euclidienne (en ligne droite) entre chaque emplacement de référence et chaque point représentatif possible a été mesurée.

Une distance moyenne a été calculée entre l'emplacement de référence et les points représentatifs tirés du FCCP+, en utilisant tous les points représentatifs possibles pour chaque code postal. Dans le cas des appariements uniques, un seul point représentatif était possible (figure 1A), mais pour les enregistrements urbains présentant de multiples appariements possibles, la probabilité d'être attribué était équivalente pour tous les points représentatifs (figure 1B). La figure 1C illustre le cas d'un code postal rural pour lequel trois points représentatifs tirés du FCCP+ étaient possibles et dans lequel les probabilités d'attribution d'un de ces trois points étaient de 70 %, 20 % et 10 %. Ces poids de probabilité ont été utilisés pour calculer une distance moyenne pondérée entre l'emplacement de référence et les trois représentatifs tirés du FCCP+. Une distance moyenne a été calculée pour chaque enregistrement à l'aide de cette méthode.

D'après la distance établie (ou la distance moyenne) entre l'emplacement de référence et l'emplacement géocodé à partir du code postal, des mesures d'erreur, telles que la distance médiane et l'intervalle interquartile (IIQ), ont été calculées selon le TML, le type de point représentatif et la taille de la collectivité. Les pourcentages d'enregistrements de l'échantillon géocodés à l'intérieur de diverses distances (0,5, 1, 5 et 10 km) par rapport à l'emplacement de référence ont également été calculés.

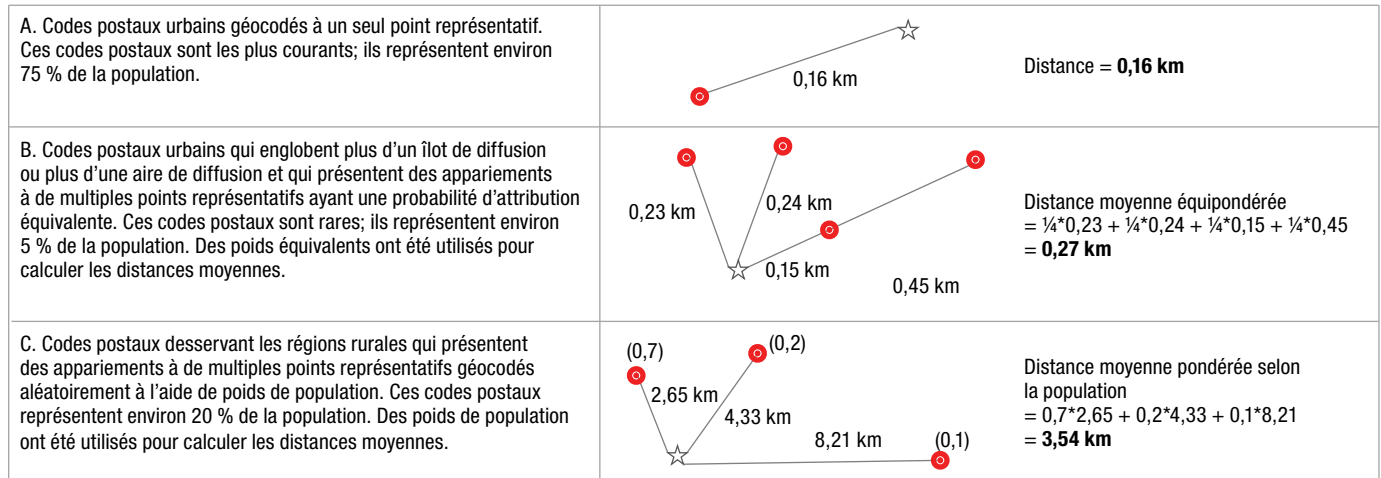
Sur les 4 401 appariements obtenus après la fusion avec les points représentatifs des codes postaux, les valeurs aberrantes associées à des distances invraisemblablement grandes (plus de 50 km) ont été examinées manuellement (178 appariements, représentant 105 enregistrements). Une part importante de ces valeurs ont été jugées

Tableau 1
Taille de l'échantillon et population représentée, selon le type de mode de livraison, le type de point représentatif et la taille de la collectivité, Canada, 2011

	Échantillon			Recensement de la population de 2011 [†]
	Nombre	%	Appariements uniques	%
Total	1 004	100,0	399	100,0
Type de mode de livraison (TML)				
A (ménages urbains ordinaires)	408	40,6	182	69,6
B (grands immeubles d'habitation urbains)	100	10,0	94	11,1
E, G, M (entreprises et établissements institutionnels urbains)	90	9,0	82	0,7
H, K, T (mixte : urbain et rural)	180	17,9	37	2,1
W (rural)	226	22,5	4	13,9
Type de point représentatif				
Côté d'îlot	474	47,2	289	75,7
Îlot de diffusion	101	10,1	51	4,1
Aire de diffusion	429	42,7	59	18,1
Taille de la collectivité				
1 500 000 ou plus	204	20,3	98	34,8
De 500 000 à 1 499 999	150	14,9	86	18,5
De 100 000 à 499 999	126	12,5	63	17,9
De 10 000 à 99 999	259	25,8	106	12,1
Moins de 10 000	265	26,4	46	15,3

[†] les chiffres ne totalisent pas 100 % en raison de l'exclusion de 723 960 adresses sans code postal et de 1 820 adresses ayant un code postal associé au TML J (livraison générale à un bureau de poste urbain) ou X (itinéraire motorisé en zone industrielle)

Source : Recensement du Canada de 2011.

Figure 1**Calculs appliqués à l'échantillon pour établir la distance moyenne entre l'emplacement de la résidence du répondant (déterminé par imagerie satellite) et les points représentatifs du code postal (tirés du FCCP+)**

● Point représentatif tiré du FCCP+ ☆ emplacement de la résidence du répondant — Distance euclidienne () poids de population attribué à l'AR

plausibles, car elles se trouvaient en région suburbaine ou rurale, et ont été géocodées selon une répartition pondérée selon la population (79 appariements, représentant 54 enregistrements de l'échantillon). Elles ont été conservées, car les résidents de certaines régions éloignées peuvent parcourir de grandes distances pour récupérer leur courrier. En outre, en raison du faible poids de population, les valeurs aberrantes n'ont pas pu fausser le calcul de la distance moyenne. Dans le Nord, trois appariements géographiques (deux à Dawson City, un à Yellowknife) représentant des distances extrêmement grandes (jusqu'à 1 068 km) ont été jugés improbables et supprimés. Les autres grandes distances étaient attribuables à des erreurs typographiques dans les codes postaux déclarés dans le cadre du recensement ou à des erreurs dans le géocodage de l'adresse municipale et ont été supprimées. L'échantillon final de l'étude était composé de 1 004 adresses.

Résultats

Le tableau 2 montre la distance médiane (et l'IIQ) entre les emplacements de référence et les points représentatifs moyens du FCCP+ pondérés selon la population, pour chaque TML, type de point représentatif et taille de collecti-

tivité. La précision était plus grande et la variabilité plus faible dans le cas des codes postaux urbains, comparativement aux codes postaux ruraux et mixtes. Le meilleur niveau de précision a été obtenu pour les grands immeubles d'habitation (TML B) comportant une distance médiane de 110 m (IIQ de 60 m à 200 m), lesquels étaient suivis des ménages ordinaires (TML A) dont la distance médiane était de 160 m (IIQ de 80 m à 320 m). Dans le cas des services commerciaux et institutionnels (TML E, G ou M), un niveau de précision relativement bon a été observé, leur distance médiane étant de 210 m, mais la variabilité était plus grande (IIQ de 80 m à 1,25 km). Les distances médianes les plus grandes ont été obtenues pour les services associés à des modes mixtes (services ruraux offerts par des bureaux de poste urbains; TML H, K ou T) : 5,87 km (IIQ de 3,81 km à 8,75 km) et les services ruraux (TML W) : 5,20 km (IIQ de 2,93 km à 9,24 km).

Plus l'échelle de la source de données pour les types de points représentatifs était précise, plus la précision était grande. La distance médiane pour les côtés d'îlot était de 130 m (IIQ de 60 m à 230 m), comparativement à 330 m (IIQ de 160 m à 750 m) pour les îlots de diffusion et à 5,53 km (IIQ de 2,89 km à

8,66 km) pour les aires de diffusion, qui comptaient des codes postaux mixtes et ruraux dans une proportion de 95 % (données non illustrées).

De façon générale, plus la taille de la collectivité était importante, meilleur était le niveau de précision. La plus grande distance médiane a été observée dans les régions rurales et les petites villes : 5,60 km (IIQ de 2,42 km à 9,48 km). Les distances médianes pour les autres groupes de tailles de collectivité variaient de 160 m (IIQ de 80 m à 830 m) dans le groupe des plus grandes collectivités urbaines à 330 m (IIQ de 130 m à 3,93 km) dans le groupe des plus petites collectivités.

Le tableau 3 montre le pourcentage de l'échantillon de l'étude géocodé selon le FCCP+ à l'intérieur de diverses distances par rapport à l'adresse municipale complète. Parmi les enregistrements de l'échantillon ayant un code postal résidentiel urbain, 87 % des ménages ordinaires (TML A), 92 % des grands immeubles d'habitation (TML B) et 66 % des entreprises ou établissements institutionnels (TML E, G, M) ont été géocodés à l'intérieur d'une distance de 500 m de leur adresse municipale complète. En revanche, seulement 47 % des enregistrements ayant un code postal rural (TML W) et 39 % des enregistre-

Précision de localisation du géocodage à l'aide des codes postaux résidentiels comparativement à celle du géocodage fondé sur les adresses municipales complètes • Coup d'œil méthodologique

Tableau 2

Distance médiane et intervalle interquartile entre l'emplacement géocodé à partir de l'adresse municipale complète et l'emplacement géocodé à partir du code postal, selon le type de mode de livraison, le type de point représentatif et la taille de la collectivité, Canada, 2011

	Kilomètres		
	Distance médiane	Intervalle interquartile de à	
Total	0,46	0,12	4,89
Type de mode de livraison			
A (ménages urbains ordinaires)	0,16	0,08	0,32
B (grands immeubles d'habitation urbains)	0,11	0,06	0,20
E, G, M (entreprises et établissements institutionnels urbains)	0,21	0,08	1,25
H, K, T (mixte : urbain et rural)	5,87	3,81	8,75
W (rural)	5,20	2,93	9,24
Type de point représentatif			
Côté d'îlot	0,13	0,06	0,23
Îlot de diffusion	0,33	0,16	0,75
Aire de diffusion	5,53	2,89	8,66
Taille de la collectivité			
1 500 000 ou plus	0,16	0,08	0,83
De 500 000 à 1 499 999	0,22	0,09	0,52
De 100 000 à 499 999	0,20	0,08	2,53
De 10 000 à 99 999	0,33	0,13	3,93
Moins de 10 000	5,60	2,42	9,48

Source : Recensement du Canada de 2011.

Tableau 3

Pourcentage géocodé à partir du code postal à l'intérieur de diverses distances par rapport à l'emplacement géocodé à partir de l'adresse municipale complète, selon le type de mode de livraison, le type de point représentatif et la taille de la collectivité, Canada, 2011

	Distance			
	500 m ou moins	1 km ou moins	5 km ou moins	10 km ou moins
Type de mode de livraison				
A (ménages urbains ordinaires)	87,3	92,2	99,3	100,0
B (grands immeubles d'habitation urbains)	92,0	97,0	100,0	100,0
E, G, M (entreprises et établissements institutionnels urbains)	65,6	74,4	87,8	96,7
H, K, T (mixte : urbain et rural)	1,1	3,3	38,9	82,2
W (rural)	0,0	0,4	47,3	78,3
Type de point représentatif				
Côté d'îlot	91,4	94,7	98,1	99,8
Îlot de diffusion	65,4	78,2	98,0	99,0
Aire de diffusion	2,3	4,4	45,9	80,9
Taille de la collectivité				
1 500 000 ou plus	72,1	77,5	92,2	99,5
De 500 000 à 1 499 999	74,0	80,0	88,7	98,0
De 100 000 à 499 999	64,3	68,3	84,1	98,4
De 10 000 à 99 999 (toute RMR ou AR de moins de 100 000)	56,4	61,0	82,2	94,2
Moins de 10 000 (toute collectivité qui n'est pas une RMR ou une AR)	9,1	9,4	45,7	76,2

RMR = région métropolitaine de recensement

AR = agglomération de recensement

Source : Recensement du Canada de 2011.

ments ayant un code postal mixte urbain et rural (TML H, K, T) ont été géocodés à l'intérieur d'une distance de 5 km de leur résidence.

Selon le type de point représentatif, 91 % des enregistrements géocodés à un côté d'îlot et 65 % des enregistrements géocodés à un îlot de diffusion étaient à

500 m ou moins de leur adresse municipale complète. Seulement 46 % des enregistrements associés à des points représentatifs d'aires de diffusion ont été géocodés à l'intérieur d'une distance de 5 km de leur résidence.

Les pourcentages des enregistrements de l'échantillon géocodés à 500 m ou moins de leur adresse municipale complète étaient de 56 % dans le cas des collectivités de 10 000 à 99 999 personnes et de plus de 70 % dans le cas des collectivités d'au moins 500 000 personnes. Seulement 46 % des enregistrements de l'échantillon associés à une région rurale ou à une petite ville ont été géocodés à l'intérieur d'une distance de 5 km de leur résidence.

Discussion

La présente étude porte sur la précision de localisation du géocodage fondé sur le FCCP+ comparativement au géocodage effectué à partir des adresses municipales complètes.

Les résultats vont dans le même sens que les documents publiés sur le géocodage fondé sur le code postal; le taux d'erreur est beaucoup moins élevé dans les régions urbaines que dans les régions rurales^{12,13}. L'analyse démontre également l'importance que revêt le TML dans le niveau de précision du géocodage. Outre les problèmes déjà documentés liés au géocodage dans les régions rurales^{12,14}, la présente étude révèle que des taux d'erreur élevés sont associés aux TML mixtes. Ces TML comprennent les routes rurales et suburbaines qui partent de bureaux de poste urbains ainsi que la livraison générale et les services de cases postales dans des bureaux de poste urbains, lesquels desservent de nombreux résidents des régions rurales.

Même dans les régions urbaines, il est important de faire une distinction entre la livraison résidentielle ordinaire et la livraison commerciale ou institutionnelle. Les grands établissements commerciaux et institutionnels qui comptent plus d'un emplacement physique (campus, succursales, points de vente) peuvent recevoir tout leur courrier à des adresses associées à un même code postal (TML M, cour-

Ce que l'on sait déjà sur le sujet

- Au Canada, les codes postaux sont souvent les seuls identificateurs géographiques dont nous disposons pour attribuer des renseignements contextuels ou environnementaux à une population à l'étude.
- Le Fichier de conversion des codes postaux (FCCP) et le Fichier de conversion des codes postaux plus (FCCP+) fournissent les moyens requis pour géocoder des ensembles de données uniquement à l'aide des codes postaux.
- Les documents publiés sur le géocodage effectué à l'aide des codes postaux indiquent que le taux d'erreur est beaucoup moins élevé dans les régions urbaines que dans les régions rurales.

Ce qu'apporte l'étude

- Fondée sur des données sur les adresses autodéclarées tirées du Recensement de la population de 2011, la présente étude a permis de comparer la précision de localisation du géocodage à l'aide des codes postaux avec celle du géocodage effectué à partir des adresses municipales complètes.
- Tous les codes postaux du recensement ont été traités à l'aide de la version 6C du FCCP+; la précision de localisation a été mesurée d'après la distance séparant la position géocodée et l'adresse municipale.
- La précision de localisation du géocodage fondé sur les codes postaux dépend principalement du type de mode de livraison.
- Le meilleur niveau de précision a été obtenu pour les grands immeubles d'habitation, lesquels étaient suivis des ménages ordinaires.
- De façon générale, plus la taille de la collectivité était importante, meilleur était le niveau de précision.
- Les codes postaux des régions rurales et des petites villes, ainsi que les cases postales, présentaient les niveaux d'erreurs les plus élevés.

rier récupéré dans une case postale ou un sac à un bureau de poste) et effectuer des livraisons vers des destinations finales à partir de cet emplacement. Il est également possible que certains des codes postaux déclarés dans le cadre du recensement qui sont desservis par un TML E, G ou M se rapportent à des lieux de travail plutôt qu'à des résidences. L'un ou l'autre de ces problèmes peut avoir contribué aux plus grandes distances observées dans le cas des codes postaux associés à des entreprises et des établissements institutionnels urbains (TML E, G, M), comparativement aux codes postaux résidentiels urbains (TML A, B).

La taille de la collectivité et le type de point représentatif ont influé sur la précision de localisation, principalement parce qu'ils sont liés au type de mode de livraison (TML). La plupart des résidents des régions rurales et des petites villes du Canada sont desservis par des modes de livraison ruraux ou mixtes et, dans le FCCP+, ces codes postaux sont presque exclusivement liés à des centroïdes d'aires de diffusion.

Malgré le risque d'imprécision de localisation et de classification erronée de la population associé aux diverses catégories de TML, le fait de supprimer un TML ou un groupe de taille de collectivité précis pourrait introduire un biais de sélection.

La précision de localisation nécessaire pour mener une étude dépend de la résolution spatiale des mesures environnementales ou contextuelles qui présentent un intérêt dans le cadre de l'étude^{2,3}. Une inadéquation entre les mesures et la précision de localisation du géocodage à l'aide des codes postaux peut, à elle seule, entraîner une classification erronée. Une comparaison des quintiles de revenu dans les secteurs de dénombrement attribués dans le cadre du Recensement de 1996 avec les quintiles de revenu attribués selon une version antérieure du FCCP+ a démontré que la classification erronée était de l'ordre de 3 % à 10 % dans les régions urbaines, mais de 39 % à 50 % dans les régions rurales¹⁴. Dans les régions urbaines, la classification erronée tendait à entraîner une attribution au quintile de revenu supérieur ou inférieur le plus proche;

dans les régions rurales, la classification erronée se traduisait par un écart de deux quintiles ou plus, dans une proportion de 15 % à 20 %.

Dans le cadre des études relatives à la pollution atmosphérique, des données d'une échelle plus précise tendent à être utilisées. À titre d'exemple, les estimations, par satellite, des taux de matières particulaires (MP_{2,5}) présentes dans l'air ambiant ont été réalisées à une résolution de 10 km²¹⁵, mais ont été ramenées à une résolution plus précise de 1 km²¹⁶. Une résolution spatiale plus précise est possible dans les centres de population urbains, mais pourrait poser problème dans les régions rurales, où la précision du géocodage à l'aide des codes postaux est nettement moindre.

La demande de données sur les caractéristiques urbaines (par exemple, potentiel piétonnier, densité, étalement urbain et verdure) s'est accrue, notamment aux fins d'études sur l'incidence de l'environnement bâti sur l'activité physique, l'embonpoint et l'obésité¹⁷⁻²⁰. Certaines de ces incidences peuvent être examinées à des échelles moins précises, comme la région métropolitaine de recensement et l'aire de diffusion^{18,21}, mais d'autres exigent des données environnementales propres à un îlot résidentiel²⁰. Le degré de résolution spatiale utilisé dans le cadre des études sur les caractéristiques urbaines varie généralement de 500 m (une marche de cinq minutes) et 1 km. Selon la présente analyse, de telles études devraient exclure les modes de livraison ruraux (TML W) et mixtes (TML H, J, K ou T), mais pourraient inclure les modes de livraison résidentielle urbains dans les collectivités de plus petite taille.

La résidence de la plupart des Canadiens n'est associée qu'à un seul code postal. Cependant, il peut arriver, en particulier dans les régions rurales, qu'un code postal supplémentaire, correspondant à une case postale ou à une route rurale partant d'un bureau de poste urbain, soit associé à une résidence. Les codes postaux qui correspondent à des cases postales et à des routes rurales partant d'un bureau de poste urbain sont courants dans les fichiers fondés sur la population; cependant, le géocodage effectué à partir de ces codes postaux

Précision de localisation du géocodage à l'aide des codes postaux résidentiels comparativement à celle du géocodage fondé sur les adresses municipales complètes • Coup d'œil méthodologique

desservis par un mode de livraison « mixte » présente le même niveau de précision que le géocodage à l'aide des codes postaux ruraux.

Certains codes postaux desservis par un TML E, G ou M peuvent être associés à de véritables lieux de résidence (par exemple, les maisons de repos, les résidences universitaires, les prisons, les pièces situées dans un immeuble à vocation principalement commerciale), mais la majorité correspondent à des établissements commerciaux et institutionnels. Leur présence dans des dossiers administratifs est probablement due au fait que le lieu de travail a été déclaré à la place du lieu de résidence. Le FCCP+ signale ces enregistrements afin d'attirer l'attention sur la nature potentiellement non résidentielle de ces codes postaux.

Forces et limites

L'une des forces de la présente étude tient au fait que les données sur les adresses tirées du Recensement de la population, lequel constitue le relevé le plus complet des codes postaux au Canada²², puissent faire l'objet d'un échantillonnage aléatoire. Il a ainsi été possible de constituer un fichier analytique qui comprenait différents TML, différents types de points représentatifs et différentes tailles de collectivités représentant l'ensemble du pays. L'étude a également tiré profit du logiciel de cartographie de Google qui a permis de géocoder manuellement les adresses au moyen de la photographie satellite et de confirmer les adresses à l'aide de la fonction Google Street View. La photographie aérienne orthorectifiée et l'imagerie par satellite comptent parmi les méthodes de géocodage des résidences les plus précises^{1,3,23}.

Deux types d'erreurs peuvent se présenter lorsque des données sur les adresses autodéclarées, en particulier des codes postaux, sont utilisées aux fins d'une analyse spatiale : des erreurs de déclaration et des erreurs de géocodage, d'affectation géographique ou de localisation. Bien que la présente étude soit axée sur le second type, des tentatives ont été faites pour atténuer l'incidence des erreurs de déclaration survenues lors du recensement, en particulier dans les codes postaux. Les erreurs présentes

dans les trois premiers caractères (région de tri d'acheminement) génèrent des erreurs spatiales plus importantes que les erreurs présentes dans les trois derniers caractères (unité de distribution locale). Une évaluation de l'incidence des adresses erronées sur la surveillance de la santé publique à Montréal a révélé que, sur les 10 % d'enregistrements de l'ensemble de données qui contenaient une adresse erronée, l'adresse erronée était liée à un code postal inexact dans 80 % des cas. L'évaluation a aussi permis de constater que 20 % de ces codes postaux contenaient des erreurs dans la partie désignant la région de tri d'acheminement, 60 % contenaient des erreurs dans la partie désignant l'unité de distribution locale, et 20 % contenaient des erreurs dans les deux parties²⁴. Aux fins de la présente étude, chaque fois que cela été possible, les codes postaux ont été vérifiés sur le site Web de Postes Canada¹¹; les codes postaux du recensement (autres que ceux desservis par les TML H, K ou T) dont la région de tri d'acheminement ne correspondait pas à la région de tri d'acheminement générée à partir de l'adresse complète ont été exclus.

En ce qui concerne l'emplacement de référence utilisé dans le cadre de la présente étude, l'imagerie par satellite de Google Maps a été utilisée pour attribuer des coordonnées de latitude et de longitude à des centroïdes représentatifs des toits à partir d'adresses municipales complètes. Un certain degré d'erreur est toutefois associé à cette méthode^{25,26}. En outre, la qualité des images par satellite était variable et peut avoir entraîné des erreurs dans la détermination du centroïde des résidences. Cet effet a été atténué par l'exclusion de toutes les adresses pour lesquelles un centroïde n'a pas pu être sélectionné pour la résidence ou un petit regroupement de bâtiments adjacents (par exemple la maison de ferme et les étables ou les granges adjacentes). Le rééchantillonnage rendu nécessaire par la faible qualité de l'imagerie par satellite et l'inaccessibilité de Google Street View a pu introduire un biais, car ces régions sont plus éloignées et ont une incidence disproportionnée sur certains groupes, comme les résidents du Grand Nord ou des réserves indiennes.

Comme l'indique l'important IQ associé au TML A (80 m à 320 m), les TML urbains peuvent, eux aussi, être hétérogènes, en raison sans doute de la densité des logements et de la période de construction (centres-ville d'avant-guerre par rapport à l'aménagement suburbain ayant suivi la Seconde Guerre mondiale). Une analyse ultérieure pourrait permettre d'examiner les effets de ces caractéristiques sur la précision du géocodage à l'aide des codes postaux.

Mot de la fin

Fondée sur des adresses provenant d'un échantillon aléatoire de répondants au recensement, la présente étude a permis de comparer la précision de localisation du géocodage à l'aide des codes postaux avec celle du géocodage effectué à partir des adresses municipales complètes. Un lien a été établi entre la précision de localisation et le mode de livraison (urbain, rural ou mixte), la source des coordonnées de latitude et de longitude (centroïdes de côtés d'îlot, d'îlots de diffusion ou d'aires de diffusion) et la taille de la collectivité. Les différences observées entre les groupes de taille de collectivité et entre les types de points représentatifs étaient liées au mode de livraison du courrier.

Les résultats font ressortir l'incidence du type de mode de livraison sur la précision de localisation. Les codes postaux ruraux et mixtes (partiellement urbains, partiellement ruraux) présentaient un nombre beaucoup plus élevé d'erreurs de géocodage que les codes postaux urbains. Ces résultats démontrent l'importance de comprendre que la précision de localisation du géocodage à l'aide des codes postaux diffère selon la nature des codes postaux. Puisque de telles différences peuvent modifier considérablement l'effet des variables environnementales et contextuelles, il est recommandé d'effectuer d'abord une stratification par type de mode de livraison ou par taille de collectivité. La résolution spatiale que requièrent les mesures environnementales ou contextuelles d'intérêt peut aider les analystes à déterminer les sous-populations qui doivent être exclues d'une étude. ■

Références

1. S. Mazumdar, G. Rushton, B.J. Smith *et al.*, « Geocoding accuracy and the recovery of relationships between environmental exposures and health », *International Journal of Health Geographics*, 7, 2008, p. 13.
2. P.F. DeLuca, P.S. Kanaroglou, « Effects of alternative point pattern geocoding procedures on first and second order statistical measures », *Spatial Science*, 53(1), 2008, p. 131–141.
3. M.R. Bonner, D. Han, J. Nie J *et al.*, « Positional accuracy of geocoded addresses in epidemiologic research », *Epidemiology*, 14(4), 2003, p. 408–412.
4. T. Burra, M. Jerrett, R.T. Burnett, M. Anderson, « Conceptual and practical issues in the detection of local disease clusters: a study of mortality in Hamilton, Ontario », *The Canadian Geographer*, 46(2), 2002, p. 160–171.
5. J.R. Guernsey, R. Dewar, S. Weerasinghe *et al.*, « Incidence of cancer in Sydney and Cape Breton County, Nova Scotia 1979-1997 », *Canadian Journal of Public Health*, 91(4), 2000, p. 285–292.
6. Statistique Canada. *Fichier de conversion des codes postaux^{MO} (FCCP)*, Guide de référence (n° 92-154-G au catalogue), Ottawa, Statistique Canada, 2016.
7. Statistique Canada. *Fichier de conversion des codes postaux^{MO} plus (FCCP+) version 6C*, Guide de référence (n° -82-F0086-XDB au catalogue), Ottawa, Ministère de l'Industrie, 2016.
8. P. Peller, *An Analysis of the Postal Code Conversion File's Use in Research*, Calgary, Alberta, University of Calgary, 2011.
9. Statistique Canada. *Dictionnaire de recensement—Année de recensement 2011* (n° 98-301-X2011001 au catalogue), disponible à l'adresse <http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/index-f.cfm>
10. Postes Canada, *Liste des Bureaux du Services aérien omnibus*, disponible à l'adresse <http://www.canadapost.ca/tools/pg/prices/RCRZ-f-AIR.pdf> (document consulté le 11 juillet 2016).
11. Postes Canada, *Outil Trouver un code postal*, disponible à l'adresse www.canadapost.ca/cpo/mc/personal/postalcode/fpc.jsf
12. M.A. Healy, J.A. Gilliland, « Quantifying the magnitude of environmental exposure misclassification when using imprecise address proxies in public health research », *Spatial and Spatio-temporal Epidemiology*, 3, 2012, p. 55–67.
13. E. Ng, R. Wilkins, A. Perras, « À quelle distance se trouve le plus proche hôpital? Le calcul des distances à l'aide du Fichier de conversion des codes postaux de Statistique Canada », *Rapports sur la santé*, 5(2), 1993, p. 179–188.
14. R. Wilkins, *Neighbourhood income quintiles derived from Canadian postal codes are apt to be misclassified in rural but not urban areas*, Ottawa, Statistique Canada, 2004, disponible à l'adresse <file=r3jc96paper.pdf> [disponible à l'adresse www.ResearchGate.net]
15. D.L. Crouse, P.A. Peters, P. Hystad *et al.*, « Ambient PM_{2.5}, O₃, and NO₂ exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC) », *Environmental Health Perspectives*, 123(11), 2015, p. 1180–1186.
16. L. Pinault, M. Tjepkema, D.L. Crouse *et al.*, « Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian Community Health Survey cohort », *Environmental Health*, 15(8), 2016, p. 1–15.
17. T. Poulou, S.J. Elliott, « Individual and socio-environmental determinants of overweight and obesity in urban Canada », *Health & Place*, 16, 2010, p. 389–398.
18. L. Seliske, W. Pickett, I. Janssen, « L'étalement urbain et sa relation avec le transport actif, l'activité physique et l'obésité chez les jeunes au Canada », *Rapports sur la santé*, 23(2), 2012, p. 19–28.
19. E. de Sa, C.I. Ardern, « Associations between the built environment, total, recreation, and transit-related physical activity », *BMC Public Health*, 14(693), 2014, p. 1–8.
20. Polsky JY, Moineddin R, Dunn JR *et al.*, « Absolute and relative densities of fast-food versus other restaurants in relation to weight status: Does restaurant mix matter? », *Preventive Medicine*, 82, 2016, p. 28–34.
21. M. Winter, R. Barnes, S. Venners *et al.*, « Older adults' outdoor walking and the built environment: Does income matter? », *BMC Public Health*, 15(876), 2015, p. 1–8.
22. K. Mechanda, H. Puderer, *Mise en correspondance des codes postaux et des régions géographiques*, Série de documents de travail de la géographie (n° 92F0138MIF – n° 001 au catalogue) Ottawa, Ministère de l'Industrie, 2007.
23. C.J.D. Bow, N.M. Water, P.D. Faris *et al.*, « Accuracy of city postal code coordinates as a proxy for location of residence. *International Journal of Health Geographics*, 3, 2004, p. 3: 1–9.
24. K. Zinszer, C. Jauvin, A. Verma *et al.*, « Residential address error in public health surveillance data: A description and analysis of the impact on geocoding », *Spatial and Spatio-temporal Epidemiology*, 1, 2010, p. 163–168.
25. T. Ubukawa, *An Evaluation of the Horizontal Positional Accuracy of Google and Bing Satellite Imagery and Three Road Data Sets Based on High Resolution Satellite Imagery*, New York, New York, Centre for International Earth Science Information Network, The Earth Institute at Columbia University, 2013, p. 1–16.
26. D. Potere, « Horizontal positional accuracy of Google Earth's high resolution imagery archive », *Sensors*, 8, 2009, p. 7973–7981.