Nº 18-001-X au catalogue
ISBN 978-0-660-75055-2

Rapports sur les projets spéciaux sur les entreprises

Cartographie de la localisation et de la colocalisation des industries au niveau du quartier : une approche de densité du noyau spatial

par Jérôme Blanchet, Robert Oikle, Dennis Huynh et Alessandro Alasia

Date de diffusion : le 10 octobre 2025

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 infostats@statcan.gc.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

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.

Table des matières

Tableaux

Figures

Cartes

Remerciements

Le Laboratoire d'exploration et d'intégration des données (LEID) et le Laboratoire de données urbaines (LDU) du Centre de projets spéciaux sur les entreprises (CPSE) remercient Statistique Canada, en particulier Christian Wolfe, Shujaat Ansari et Serge Godbout, pour leur connaissance du Registre des entreprises, le Dr Mahamat Hamit-Haggar pour la coordination du processus éditorial et des révisions internes, Chris Li pour la révision institutionnelle, le Dr Bjenk Ellefsen pour l'élaboration de la vision et le soutien à la faisabilité du projet, le Dr Ala'a Al-Habashna pour le lancement du traitement des données, et Zheng Yu et Wafa Ashraf pour la révision technique finale du document. Nous remercions également le Dr Stephen Tapp et Patrick Gill de la Chambre de commerce du Canada (CCC) pour leur révision, leurs commentaires et leurs suggestions.

Renseignements pour l'utilisateur

Symboles

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

.	indisponible pour toute période de référence
..	indisponible pour une période de référence précise
...	n'ayant pas lieu de figurer
0	zéro absolu ou valeur arrondie à zéro
0^s	valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
^{p}	provisoire
^{r}	révisé
x	supprimé en raison des dispositions de la *Loi sur la statistique* en matière de confidentialité
^{E}	à utiliser avec prudence
F	trop peu fiable pour être publié
*	valeur significativement différente de l'estimation pour la catégorie de référence ($p<0,05$)

Les termes mathématiques suivants sont utilisés dans cet article:

$RMR()$: identifiants des régions métropolitaines de recensement (plus information de base)

$ID()$: identifiants des îlots de diffusion (plus information de base)

$AD()$: identifiants des aires de diffusion (plus information de base)

KDE : estimation de la densité du noyau (kernel density estimation)

$PKDF()$: fonction de densité de noyau polynomiale (polynomial kernel density function)

$G()$: grille de cellules ou tuiles de sortie (grid of output cells or tiles)

$G(mDB)$: grille des ID médians (grid of median DBs)

$\Phi()$: identifiants des cellules de sortie de la grille (plus information de base)

Φc : centroïde des cellules de sortie de la grille

$\psi()$: taille de la bande passante de densité du noyau

$m()$: minimum de certains éléments (minimum of some items)

$X()$: répartition de l'échantillon (sample distribution)

$s()$: taille de l'échantillon (sample size)

$\sigma()$: distance spatiale standard par rapport au centre moyen

$\mu()$: distance spatiale médiane par rapport au centre moyen

$IQR()$: intervalle interquartile 75 %-25 % (interquartile range)

$L1()$: coordonnée numérique de latitude (latitude numerical coordinate)

$L2()$: coordonnée numérique de longitude (longitude numerical coordinate)

$L1C()$: coordonnée numérique de latitude du centre moyen de la RMR

$L2C()$: coordonnée numérique de longitude du centre moyen de la RMR

$\sqrt{A}()$: longueur latérale d'un ID médian (side length of a median DB)

$\pi\psi^2()$: superficie spatiale totale couverte par la bande passante autour du centroïde des cellules de sortie de la grille

i : indice des éléments de X inclus dans la superficie spatiale couverte par la bande passante autour du centroïde des cellules de sortie

$n()$: nombre total d'éléments de X inclus dans la superficie spatiale couverte par la bande passante autour du centroïde des cellules

$JW()$: poids unitaire des emplois (job unity weight)

$JL()$: emplacement spatial aléatoire des emplois (job random spatial location)

$\Theta\ ()$: distance entre l'emplacement aléatoire des emplois et le centroïde ciblé des cellules de sortie de la grille

$f = \frac{3}{\pi} r\ ()$: rapport principal de la fonction de densité de noyau polynomiale (contribution de la densité individuelle)

$\sim U(DBP)$: distribution spatiale uniforme sur un polygone d'îlot de diffusion (uniform spatial distribution over a dissemination block polygon)

i,i,d : indépendant et identiquement distribué

RP : processus aléatoire multinomial de la méthodologie (multinomial random process for the methodology)

$E(RP)$: événement attendu pour le processus aléatoire multinomial de la méthodologie (expected event for multinomial random process)

$T(RP)$: événement exceptionnel pour le processus aléatoire multinomial de la méthodologie (tail rare event for multinomial random process)

BRP : processus aléatoire binomial de l'exemple analogique (binomial random process for the analogical example)

$E(BRP)$: événement attendu pour le processus aléatoire binomial de l'exemple analogique (expected event for the binomial process)

$T(BRP)$: événement exceptionnel pour le processus aléatoire binomial de l'exemple analogique (tail rare event for binomial random process)

b : grand nombre de petits îlots de diffusion pour le processus aléatoire binomial de l'exemple analogique

v : grand nombre pair d'emplois pour le processus aléatoire binomial de l'exemple analogique

CLT : théorème de la limite centrale (central limit theorem)

$(->N())$: convergence vers la distribution normale (convergence toward normal distribution)

$\sum_{j=1}^{v}$: somme d'un nombre fini de v éléments

$C(v,j)$: nombre de façons de combiner les j éléments parmi les v éléments sans ordre

! : factoriel

$J()$: ensemble de dimensions n x 2 pour $JW()$ et $JL()$ sur $n()$

y : valeur de densité totale issue de la somme finie de la contribution de la densité individuelle (total density value)

x : rapport de la distance à la taille de la bande passante $\Theta()/\psi()$ (ratio of distance)

df/dx ou f' : première dérivée de f

d^2f/dx^2 ou f'' : deuxième dérivée de f

$\in(,)\in[,]$: élément inclus dans l'intervalle ouvert et borné

$-\infty,+\infty$: infini négatif et positif

MCLT, Théorème central limite multivarié

S, Grand nombre fini de points spatiaux disponibles dans un ID

p, Vecteur de S probabilités de dimension S dont la somme est égale à 1

vp, Vecteur de S entrées représentant le nombre attendu d'emplois alloués à chacun des S points spatiaux de l'ID

vM, Matrice de variance-covariance de dimension finit S x S de la distribution normale multivariée de notre processus aléatoire multinomial

$\left(\mathrm{vM}\right)^{-1}$, Inverse de la matrice vM

I, Matrice identité de dimension SxS

P, Matrice diagonale dont les éléments diagonaux sont les éléments du vecteur p

$\mathrm{SVN}\left(\mathrm{x}';\mathrm{vp},\mathrm{vM}\right)$, Distribution normale multivariée ou S-variée

$\mathrm{det}\left(\mathrm{vM}\right)$ or $\left|\mathrm{vM}\right|$, Déterminant de la matrice vM

k, Facteur de réduction de la densité des points spatiaux uniformes sur un plan

$\mathrm{g}\left(\mathrm{k}\right)=\mathrm{k}^{1/2}$, Expansion en distance moyenne des points spatiaux uniformes sur un plan

Δ, Faible variation positive du facteur k

$\approx$, Approximation

Cartographie de la localisation et de la colocalisation des industries au niveau du quartier : une approche de densité du noyau spatial

par **Jérôme Blanchet, Robert Oikle, Dennis Huynh** et **Alessandro Alasia**

Sommaire

Quand les entreprises décident où exercer leurs activités, elles ne se contentent pas de choisir une ville, elles choisissent un quartier[1]. Aussi, le regroupement d'entreprises dans un quartier donnée a une incidence sur les possibilités économiques et la qualité de vie dans la région. Alors que le rôle de ces dynamiques à l'échelle locale sont documentées dans la littérature, les recherches sur le regroupement d'entreprises au Canada se sont principalement concentrées sur une échelle régionale ou métropolitaine. Ceci a pour effet de limiter les applications possibles de l'analyse des grappes d'entreprises pour l'aménagement urbain, la création d'infrastructures et le développement local par les acteurs qui mettent en œuvre des programmes à l'échelle locale.

L'amélioration constante de la géolocalisation des données d'entreprise offre de nouvelles possibilités d'analyse. La présente analyse décrit une méthode qui permet de définir les grappes d'entreprises à un niveau sous-métropolitain détaillé. En utilisant les données du Registre des entreprises (RE) de Statistique Canada pour certaines industries, l'emplacement des emplois à l'échelle des établissements est réparti spatialement dans les îlots de diffusion (ID) respectifs (îlots dans les zones urbaines et rurales). Une méthode d'estimation de la densité du noyau (KDE, pour *kernel density estimation*) spatiale est effectuée sur ces emplacements d'emploi pour définir les limites des grappes d'entreprises. Une nouvelle méthode pour définir la taille de la bande passante du noyau est détaillée, car la méthode traditionnelle de la règle de Silverman échoue dans le cas de nos applications à reconnaître directement la configuration de la structure d'ID dans les villes ciblées. Les résultats sont élaborés pour trois secteurs (fabrication, commerce de détail et services d'hébergement et de restauration) ainsi que pour certaines grappes industrielles définies par Delgado et coll. (2014) pour quatre grandes régions métropolitaines (Montréal, Toronto, Winnipeg et Vancouver).

Les résultats sont cartographiés pour chaque type de grappe et de région métropolitaine montrant différentes configurations spatiales pour différents secteurs d'activité. Comme prévu, les grappes du commerce de détail et des services d'hébergement et de restauration sont relativement plus dispersées dans les régions métropolitaines que dans la grappe du secteur de la fabrication. Cependant, des statistiques simples sur les établissements et les emplois montrent que les limites géographiques des grappes à l'échelle des quartiers générées par cette analyse capturent la plupart des établissements et emplois situés dans la RMR de référence et associés à la grappe industrielle. Par exemple, la grappe du secteur de la fabrication regroupe 89,7 % (Montréal), 94,4 % (Toronto), 91,9 % (Winnipeg) et 90,8 % (Vancouver) de l'ensemble des emplois dans le secteur de la fabrication dans les RMR correspondantes. Les résultats indiquent également une colocalisation accrue de certains types d'entreprises, tels que le commerce de détail et les services d'hébergement et de restauration. Cette analyse préliminaire semble prometteuse pour révéler les schémas de colocalisation des entreprises dans les zones de quartier définies.

Les méthodes utilisées pour définir ces grappes à l'échelle des quartiers ouvrent de nouvelles possibilités d'analyse en temps opportun des conditions économiques locales ainsi que des analyses plus larges à l'échelle des quartiers (par exemple, les disparités sociales et la qualité de vie) en tenant compte de la composition des entreprises de la région. L'utilisation de fichiers de limites géographiques de grappes d'entreprises précises, comme outil de géorepérage, peut servir à surveiller la performance et les tendances des entreprises locales, en les combinant à d'autres fonds de données de Statistique Canada ou à d'autres sources de données, telles que les flux de mobilité.

Ce projet propose une méthodologie expérimentale et un ensemble de grappes industrielles expérimentales. Vos commentaires sont appréciés. Vous pouvez communiquer vos commentaires et suggestions à l'auteur principale Jérôme Blanchet (819-576-5502), chef d'unité au Laboratoire d'exploration et d'intégration des données (LEID), et Laboratoire de données urbaines (LDU) , du Centre des projets spéciaux sur les entreprise (CPSE), Statistique Canada.

1. Voir: Bridges, Winter 2005-2006 (Federal Reserve Bank of St. Louis)

Introduction

Le débat universitaire et politique sur les grappes d'entreprises s'étend sur plus de trois décennies. Dans la plupart de la littérature et des documents de politique qui en découlent, les grappes d'entreprises sont définies comme une concentration géographique d'entreprises, d'organisations et d'institutions interconnectées au sein d'une industrie ou d'un secteur particulier (Wolfe et Gertler, 2004). La théorie et les preuves suggèrent que la proximité spatiale et l'agglomération facilitent les liens de collaboration, l'échange de ressources et d'autres synergies. Ainsi, les politiques de soutien aux grappes reposent sur l'idée que le regroupement d'entreprises et d'organisations de soutien dans une zone géographique précise stimule les synergies, l'innovation et les avantages concurrentiels (Bekar et Lipsey, 2001).

La littérature concernant les grappes au Canada a étudié l'effet des grappes sur la performance des entreprises, les emplois et les salaires (Lucas et coll., 2009; Niosi et Bas, 2001; Steiner et Ali, 2011; Spencer et coll., 2010). Elle a également analysé les politiques de soutien aux grappes ainsi que leur impact et leur efficacité (Niosi et Bas, 2001) et exploré des méthodologies pour identifier et cartographier les grappes dans l'espace (Spencer, 2014). Dans l'ensemble, des documents mettent en évidence que le regroupement spatial des entreprises crée un environnement propice à l'innovation, à l'efficacité des ressources, à la collaboration et au développement économique global. Ces grappes contribuent à la croissance et au succès de chaque entreprise tout en améliorant la compétitivité et la résilience d'une région ou d'une région métropolitaine.

La plupart des documents canadiens se sont concentrés à l'échelle régionale ou sur des régions métropolitaines particulières. Pour de nombreuses applications et à des fins politiques, une analyse à l'échelle régionale (ville, région métropolitaine ou marché du travail) restera adéquate. Néanmoins, un nombre croissant d'applications nécessitent une analyse à l'échelle des quartiers et offrent des perspectives uniques tant pour les acteurs locaux (municipalités et autres organisations économiques locales) que pour les acteurs provinciaux ou fédéraux. La demande de données sur la situation et les tendances des entreprises, à des niveaux géographiques détaillés, ne cesse d'augmenter, tant de la part des organismes fédéraux que des parties prenantes locales, dont des municipalités, des associations économiques et le milieu des affaires.

La présente analyse amène l'analyse des grappes d'entreprises à un niveau géographique plus détaillé, grâce à une méthodologie mise au point pour identifier les grappes d'entreprises à l'échelle des quartiers. La méthode proposée identifie les grappes d'entreprises à l'échelle des ID, qui est l'une des unités spatiales d'analyse les plus détaillées définies par Statistique Canada[2]. Cette méthode est appliquée à quatre régions métropolitaines de recensement (RMR) de tailles différentes et pour diverses spécifications de grappes industrielles, y compris des regroupements simples de codes à deux chiffres du Système de classification des industries de l'Amérique du Nord (SCIAN), ainsi que des grappes industrielles résultant de regroupements de codes du SCIAN, tels qu'ils ont été définis par Delgado et coll. (2014).

La précision croissante de la géolocalisation pour les établissements d'entreprises dans le RE de Statistique Canada rend cette analyse possible, tandis que les possibilités offertes par le couplage de microdonnées d'entreprises ouvrent la voie à l'exploration de multiples dimensions de la performance des entreprises[3]. Dans ce contexte, l'un des principaux défis de ce type d'analyse consiste à préserver la confidentialité des renseignements confidentiels des entreprises tout en fournissant des renseignements utiles au milieu des affaires et aux décideurs. Par conséquent, cet article se veut une première exploration à l'intersection du niveau de détail le plus précis et de la protection de la confidentialité des renseignements des entreprises.

Cet article est organisé en cinq grandes sections. La section qui suit donne un aperçu des travaux existants d'analyse des grappes au Canada, mettant en lumière le manque d'information à l'échelle des quartiers ainsi que les principaux aspects motivant cette recherche, tout en proposant des exemples d'analyses à l'échelle des quartiers dans d'autres pays. Elle est suivie d'une présentation détaillée des données et de la méthodologie appliquées dans cette analyse, y compris une nouvelle approche pour le calcul de la bande passante du noyau. La section suivante présente certains résultats et valide les conclusions, avant de discuter des développements futurs et des applications possibles des délimitations de grappes. Enfin, l'annexe contient un grand ensemble complémentaire de 23 cartes de grappes à haute résolution.

2. Voir : Dictionnaire, Recensement de la population, 2021 – Îlot de diffusion (ID) (statcan.gc.ca)
3. Voir, par exemple, les possibilités offertes par l'environnement de fichiers couplables des entreprises : Environnement de fichiers couplables – Entreprises (statcan.gc.ca)

Pourquoi intégrer la dimension de quartier dans l'analyse des grappes?

La plupart des études sur les grappes d'entreprises au Canada utilisent des régions ou des villes comme unités géographiques d'analyse. Pour de nombreuses applications et analyses de politiques, cette échelle géographique offre un niveau de détail adéquat. Les indicateurs de performance des entreprises sont relativement abondants à l'échelle municipale ou régionale. Au sein des villes ou des régions, on suppose que la proximité des entreprises est suffisante pour permettre les interactions qui sous-tendent le concept même de grappes et les avantages qui en découlent. Certaines des grappes bien connues, comme la Silicon Valley et le Research Triangle Park en Caroline du Nord, s'étendent sur plusieurs municipalités, et une échelle régionale semble appropriée pour étudier leur dynamique et leur développement.

Néanmoins, des preuves montrent que certaines grappes, telles que les grappes des secteurs financiers, culturels, du commerce de détail ou de la fabrication, se concentrent dans des quartiers précis au sein d'une région métropolitaine. De même, il est établi que les disparités spatiales et les variations de performance des entreprises peuvent être aussi marquées à l'échelle des quartiers qu'à celle des régions (Wheeler, 2006; OCDE, 2018). Un grand nombre de parties prenantes et de décideurs, y compris des organisations économiques locales et des municipalités, adoptent une perspective axée sur les quartiers et élaborent ou mettent en œuvre des politiques ayant une incidence sur les entreprises dans des zones précises au sein d'une municipalité. Par conséquent, ces intervenants ont besoin d'analyses des grappes et de leur performance à des niveaux géographiques plus détaillés.

Pour répondre à ces besoins d'information, différents courants de recherche ont analysé des grappes d'entreprises à l'échelle des quartiers en examinant les choix d'emplacement dans les zones métropolitaines, l'impact des réglementations et des politiques municipales sur la formation et la croissance des grappes, le rôle des associations locales et l'impact ainsi que les retombées économiques et les effets de diffusion des grappes sur les quartiers environnants. Ces recherches s'inscrivent dans diverses perspectives disciplinaires, allant des analyses plus traditionnelles des grappes d'entreprises à des études sur l'aménagement urbain, en passant par la recherche appliquée et à l'analyse menée par des associations locales, des chambres de commerce et des services de planification municipale. Le reste de cette section fournit un aperçu de ces documents, en mettant en évidence les principaux enseignements et en soulignant les lacunes actuelles en matière d'information dans le contexte canadien.

Les disparités spatiales au sein des villes, à l'échelle des quartiers, constituent un enjeu bien connu et étudié (OCDE, 2018). Les choix des entreprises relativement à leur emplacement et leur regroupement dans certains quartiers sont influencés par ces disparités, et ils contribuent à les renforcer. Par exemple, Wheeler (2006) montre que la croissance positive des entreprises dans la région métropolitaine de St. Louis aux États-Unis, découle en réalité d'une croissance substantielle dans un quartier, combinée à un déclin dans un autre. Ces dynamiques sont motivées à la fois par les caractéristiques des quartiers et des entreprises. Ces observations sont pertinentes tant du point de vue des entreprises (pour leurs choix d'emplacement) que du point de vue de la municipalité (pour l'élaboration de politiques visant à soutenir le regroupement d'entreprises et à réduire les disparités socioéconomiques entre quartiers). L'importance d'une approche axée sur les quartiers dans l'analyse des grappes spatiales est également soulignée par Gabaix (2011), qui utilise des données d'imagerie satellitaire pour définir des grappes de densité de population. Ces résultats suggèrent que les données issues de méthodes basées sur la densité des grappes, pour délimiter les frontières urbaines, s'intègrent plus facilement dans un modèle spatial que les régions définies directement par des limites administratives.

Les analyses des grappes d'entreprises à l'échelle des quartiers apportent des éclairages sur les choix d'emplacement des entreprises. Wheeler (2006) souligne que les investisseurs potentiels ne choisissent pas seulement une région ou une ville, mais aussi un quartier, ce qui peut, dans certains cas, correspondre à un parc industriel[4] ou un quartier d'affaires précis. Par conséquent, la compréhension de la dynamique des parcs industriels au sein d'une région métropolitaine et de leurs environs revêt une importance particulière. Des conclusions similaires ressortent des travaux d'Arauzo-Carod (2021), qui examine les choix d'emplacement des entreprises de haute technologie à l'échelle des quartiers à Barcelone. Cette analyse montre que les

4. Parc industriel — Wikipédia

caractéristiques et les installations des quartiers influencent les décisions d'emplacement des entreprises et que les retombées spatiales jouent un rôle essentiel pour certaines industries de haute technologie,

Les grappes d'entreprises ne servent pas uniquement à établir la composition économique et la disponibilité des emplois dans un quartier. Plusieurs études soulignent qu'elles peuvent également avoir une incidence sur la qualité de vie dans les quartiers et leur potentiel d'attraction à des fins résidentielles (Shybalkina, 2022; Stern et Seifert, 2010). Des travaux de recherche ont porté sur l'incidence de la présence de certaines grappes sur les quartiers, en particulier les grappes artistiques et culturelles dans les zones métropolitaines. Grodach et coll. (2014) ont identifié des grappes artistiques à l'échelle des régions et des quartiers en utilisant des codes postaux pour des zones métropolitaines américaines de tailles variées. Leurs résultats montrent que les industries artistiques présentent des schémas d'emplacement distincts à l'échelle métropolitaine et des quartiers. Ils révèlent que, bien que de nombreuses caractéristiques des grappes artistiques soient propres à un lieu, les arts sont associés à des indicateurs généraux d'innovation et de développement local, ce qui suggère que ces grappes d'entreprises peuvent jouer un rôle plus large dans le développement économique des régions métropolitaines.

Un autre courant de littérature sur les grappes d'entreprises intra-urbaines se concentre sur la délimitation et le rôle des centres d'affaires. Cette littérature est souvent regroupée sous la rubrique de l'analyse des secteurs du centre des affaires (SCDA) (Meltzer 2012; Yu et coll., 2015). Elle présente à la fois une pertinence méthodologique et politique, avec une partie axée sur les aspects de modélisation et une autre explorant le rôle et la dynamique des organisations impliquées dans la gestion, le développement et la promotion de ces centres d'affaires. Les méthodes utilisées pour délimiter les SCDA vont de l'utilisation des données de télédétection (Taubenböck et coll., 2013) aux données de recensement sur la densité des emplois (Yu et coll., 2015).

L'approche adoptée dans la présente analyse s'inspire de cette littérature et, en particulier, de l'analyse de Sergerie et coll. (2021) qui visait à élaborer une méthode pour identifier les limites géographiques des quartiers centraux de villes canadiennes. Sergerie et coll. (2021) utilisent une KDE spatiale pour calculer une surface de densité des données sur l'emplacement des emplois à l'échelle des ID, rendant possible la comparaison de ces zones partout au Canada.

La formation de grappes d'entreprises à l'échelle des quartiers, tout comme à l'échelle des régions, n'est pas simplement un processus spontané ou le résultat cumulatif d'accidents historiques. Autrement dit, ce n'est pas un phénomène aléatoire et il peut être expliqué. Les municipalités jouent un rôle clé dans le façonnement, le développement et le soutien du regroupement des entreprises dans des domaines précis (Zhang, 2019). Cela se fait généralement par l'intermédiaire du zonage, une méthode réglementaire utilisée par les municipalités ou les gouvernements locaux pour établir des règles définissant les activités et les constructions autorisées à un emplacement donné. Ainsi, les municipalités fournissent l'espace, les infrastructures et les services nécessaires au développement de parcs industriels en tant que zones dédiées à des usages industriels.

Comme les municipalités, les associations professionnelles locales sont des acteurs clés dans le soutien du développement de grappes d'entreprises locales (Dhamo et coll., 2023). À Toronto, par exemple, la Toronto Board of Trade (2021) a publié une étude cartographiant cinq types de secteurs dans la région métropolitaine élargie qui comprennent un centre métropolitain, des zones de production et de distribution de biens, des zones de services et à usage mixte, des centres régionaux et des centres de création de connaissances. Parallèlement, cette municipalité compte 84 zones d'amélioration des affaires (ZAA). Ces associations professionnelles locales visent à soutenir des zones commerciales compétitives et attrayantes pour les consommateurs et les nouvelles entreprises.

Le rôle proactif que peuvent jouer les acteurs locaux dans le développement économique et la prospérité de leur région explique pourquoi l'analyse à l'échelle des quartiers devient de plus en plus pertinente. L'analyse des grappes locales ou des concentrations d'entreprises au sein d'un quartier peut renseigner sur le comportement des consommateurs, l'emploi local et la santé économique globale du quartier. Ces acteurs locaux opèrent dans des écosystèmes géographiquement définis pour lesquels des données peuvent être générées à partir de sources locales ou analysées à l'échelle locale. Ce qui manque, particulièrement dans le contexte canadien, c'est un cadre élargi et comparatif qui permettrait l'analyse des grappes à l'échelle des quartiers à l'échelle nationale, avec des définitions normalisées dans l'ensemble des administrations. C'est ce déficit d'information que cette analyse vise à combler. Les progrès dans le géoréférencement des micro données d'entreprises et dans l'analyse spatiale de grandes bases de données facilitent de plus en plus la réalisation d'analyses détaillées à l'échelle des quartiers.

Méthodologie proposée

L'approche générale adoptée aux fins de la présente analyse s'appuie sur les travaux de Sergerie et coll. (2021) portant sur la définition des centres-villes des régions métropolitaines du Canada. Dans le cadre de ces travaux, les KDE spatiales sont appliquées à la géolocalisation du total des emplois, obtenue de la variable du lieu de travail du Recensement de la population, et l'unité géographique d'analyse utilisée est l'aire de diffusion (AD).

Dans la présente, les grappes d'entreprises sont définies à l'aide d'une méthode analogue, utilisant des KDE spatiales appliquées à la géolocalisation des emplois enregistrés à l'échelle des établissements. Les données sur les établissements sont extraites du RE pour certains secteurs et l'unité d'analyse est l'ID, une unité plus précise que l'AD. Compte tenu du niveau de détail nettement plus précis (tant pour les industries sélectionnées que pour la géographie), la méthodologie utilisée afin de définir les grappes d'entreprises comporte plusieurs étapes supplémentaires par rapport au flux de travail décrit par Sergerie et coll. (2021). Les sections suivantes décrivent en détail la méthodologie proposée.

Zones d'étude

Quatre zones d'étude, représentant différents degrés de densité urbaine au Canada, ont été sélectionnées pour mettre au point la méthodologie. Il s'agit des RMR de Montréal, de Toronto, de Winnipeg et de Vancouver. Chaque RMR comprend un nombre différent de municipalités (subdivisions de recensement), avec des quartiers présentant des densités de population et d'emploi, ainsi que des degrés d'urbanisation, sensiblement différents.

L'unité géographique d'analyse utilisée pour la géolocalisation des entreprises est l'ID. Un ID est un territoire dont tous les côtés sont délimités par des routes et/ou par des limites; dans les zones urbaines, c'est ce que l'on appelle communément un « îlot ». Les ID font partie des régions géographiques normalisées de Statistique Canada aux fins de diffusion et ils sont la plus petite région géographique pour laquelle les chiffres de population et des logements sont diffusés. Les îlots de diffusion couvrent tout le territoire du Canada[5].

Registre des entreprises

Les données utilisées dans l'analyse proviennent du RE de Statistique Canada, qui est le répertoire central de données de base sur les entreprises et les établissements ayant des activités au Canada. Il est continuellement mis à jour. Pour cette analyse, les données se rapportent à la période de référence de décembre 2023.

Dans le RE, les secteurs d'activité sont définis selon les codes du SCIAN. L'utilisation des données du RE présente des avantages par rapport à d'autres sources de données possibles, comme les données sur le lieu de travail du Recensement de la population. Les données à l'échelle des établissements du RE sont classées selon des codes du SCIAN plus détaillés (à six chiffres), ce qui permet la création de grappes personnalisées. Les données sur l'emploi dans le RE sont également mises à jour plus fréquemment. Bien qu'elles soient moins précises que les statistiques sur l'emploi spécialisées, elles constituent une solution de rechange plus opportune aux données du recensement.

Les codes du SCIAN inclus dans cette analyse représentent six grappes industrielles différentes. Trois d'entre elles sont composées de codes du SCIAN à deux chiffres et les autres sont définies selon les travaux de Delgado et coll. (2014). Ces grappes industrielles et leurs codes du SCIAN correspondants sont résumés dans le tableau 1.

Les grappes industrielles générées à l'échelle des codes du SCIAN à deux chiffres comprennent le secteur de la fabrication (SCIAN 31, 32 et 33), le secteur du commerce de détail (SCIAN 44 et 45) et le secteur des services d'hébergement et de restauration (SCIAN 72).

Les trois grappes industrielles définies selon Delgado et coll. (2014) sont les suivantes : premièrement, la distribution et le commerce électronique (grappe 10). Cette grappe comprend principalement les grossistes traditionnels, les maisons de vente[6] par correspondance et les marchands électroniques. Les entreprises de cette grappe achètent, stockent et distribuent principalement une large gamme de produits, tels que des vêtements, des aliments, des produits chimiques, du gaz, des minéraux, des matériaux agricoles, des machines et d'autres

5. Voir : Glossaire illustré - Îlot de diffusion (ID) (statcan.gc.ca)
6. Mail Order Houses: Meaning, Advantages, and Disadvantages | GeeksforGeeks

marchandises. La grappe comprend également des entreprises qui soutiennent les opérations de distribution et de commerce électronique, y compris l'emballage, l'étiquetage et la location de matériel. La deuxième grappe est celle des services financiers (grappe 16). Elle comprend des établissements qui aident à la transaction et à la croissance d'actifs financiers pour les entreprises et les particuliers. Ces sociétés comprennent des courtiers en valeurs mobilières, des négociants et des bourses, des institutions de crédit et des services de soutien à l'investissement financier. Les sociétés d'assurance sont regroupées dans une autre grappe, celle des services d'assurance. Enfin, le secteur de l'hôtellerie et du tourisme (grappe 22) comprend des établissements liés aux services et aux lieux de l'hôtellerie et du tourisme. Cela inclut des sites sportifs, des casinos, des musées et d'autres attractions, ainsi que des hôtels et autres types d'hébergement, des transports, et des services liés aux voyages récréatifs, tels que les services de réservation et les voyagistes.

La structure hiérarchique entre les 3 grappes d'industries générées au niveau du SCIAN à 2 chiffres et la grappe à 3 industries définie selon Delgado et coll. (2014) n'est pas simple. C'est-à-dire que les différents niveaux à 4 chiffres du SCIAN utilisés selon Delgado et coll. (2014) ne sont pas nécessairement composés des premiers niveaux à 2 chiffres 31, 32, 33, 44, 45 et 72. De plus, pour des raisons de commodité et de simplicité, la récurrence des 6 grappes industrielles n'est pas nécessairement la même dans le présent document. C'est-à-dire que certaines étapes de l'analyse des résultats et de la méthodologie se concentrent sur les 3 grappes industrielles générées au niveau à 2 chiffres du SCIAN seulement, tandis que d'autres se concentrent sur les 6 grappes industrielles.

Tableau 1
Grappes industrielles avec codes du SCIAN connexes

Grappe industrielle	Codes du SCIAN inclus
Secteur de la fabrication	31, 32, 33
Secteur du commerce de détail	44, 45
Secteur des services d'hébergement et de restauration	72
Distribution et commerce électronique (grappe 10)	4111, 4121, 4131, 4132, 4133, 4141, 4142, 4143, 4144, 4145, 4161, 4162, 4163, 4171, 4172, 4173, 4179, 4182, 4183, 4184, 4189, 4191, 4232, 4234, 4235, 4236, 4238, 4239, 4241, 4242, 4243, 4244, 4245, 4246, 4247, 4248, 4249, 4251, 4541, 4931, 5324, 5614, 5619
Services financiers (grappe 16)	5211, 5221, 5222, 5223, 5231, 5232, 5239, 5259, 5269, 5614
Hôtellerie et tourisme (grappe 22)	1142, 4539, 4871, 4872, 4879, 5322, 5615, 7112, 7121, 7131, 7132, 7139, 7211, 7212, 7213

Source : calculs des auteurs à partir de la base de données sur le RE.

Estimation de la densité

La concentration spatiale des industries a été déterminée à l'aide d'une méthode de KDE spatiale. Il s'agit d'une technique non paramétrique qui estime la fonction de densité de probabilité d'une variable aléatoire sur un domaine spatial. Pour chaque région métropolitaine, des grappes ont été identifiées à l'aide des résultats de la méthode de KDE en agrégeant les ID adjacents ayant une densité minimale d'emploi dans chaque secteur ou combinaison de secteurs.

Les données sur l'emploi provenant des données des établissements du RE ont été géocodées en fichiers de limites spatiales des ID. Le nombre total d'employés par ID a ensuite été calculé. Les lieux des emplois représentant chaque employé ont été distribués de façon aléatoire et uniforme dans la limite de l'ID[7]. Cette étape de traitement doit être reconnue, non pas comme un manque de précision et une faiblesse des données, mais comme un moyen de faire progresser la méthodologie. En d'autres mots, cette méthode permet de rendre plus homogène la distribution spatiale relativement clairsemée des emplois dans l'ID et facilite le processus de KDE en générant une distribution relativement plus continue[8]. L'argument de l'uniformité a été avancé pour ne privilégier aucune sous-région de l'ID pendant le processus de randomisation. Cela signifie que la concentration de la densité des établissements dans certains points précis de l'ID n'influence pas le processus de randomisation

7. L'annexe 2 de cet article propose une méthode aléatoire alternative dans laquelle les emplois de l'ID sont réparties aléatoirement au sein de l'ID, de l'AD et de l'aire de diffusion agrégé (ADA).
8. La limite de cette méthode s'applique aux grands ID qui contiennent un seul établissement ou à quelques établissements comptant un effectif important situé dans une partie restreinte de l'ID. Dans ces cas, cette configuration spatiale serait remplacée par la distribution aléatoire des emplois au sein de l'ID. Cependant, l'utilisation des ID plutôt que des AD atténue ces défis, car les ID sont une géographie administrative plus précise que les AD et se rapprochent davantage du concept de quartier.

des emplois. Certaine section ci-bas de ce rapport décrive en plus grand détails les processus de distribution de données aléatoires.

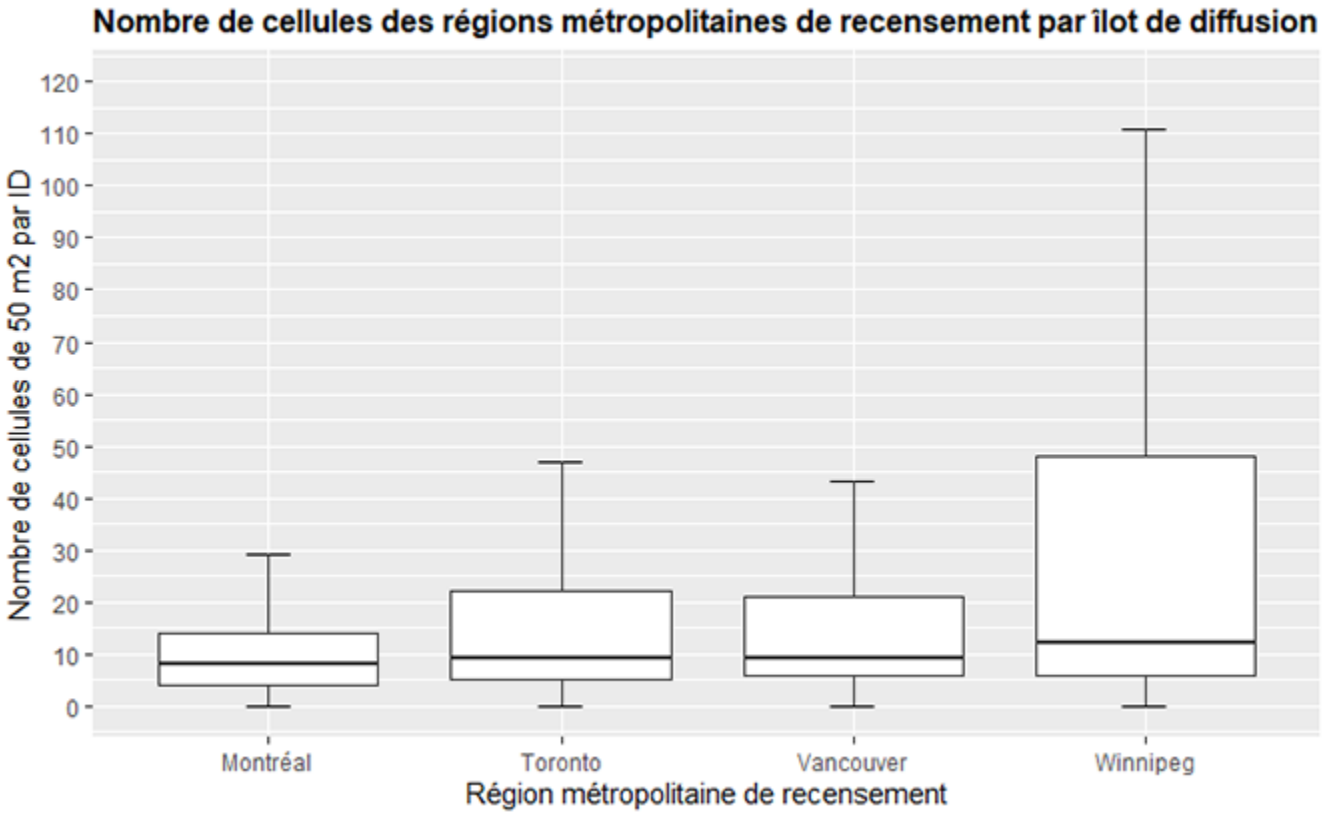
La section d'estimation de la densité de cet article est composée de 2 sous-sections. Tout d'abord, nous documentons des informations partielles sur la fonction de densité du noyau polynomiale. Juste assez d'arguments de fonction pour comprendre l'utilité de la taille de la bande passante du noyau. Nous expliquons également pourquoi l'approche de la bande passante du noyau Silverman n'est pas adaptée à notre application, puis nous décrivons une nouvelle méthode pour le calcul de la bande passante. Deuxièmement, nous décrivons toutes les informations sur la fonction de densité du noyau polynomiale. C'est-à-dire tous les arguments restants de la fonction non décrits jusqu'à présent. Nous documentons également les processus aléatoires utilisés.

Méthodologie de bande passante de densité du noyau et autres paramètres

Avant de décrire la version entière de la méthode KDE, nous détaillons partiellement la fonction de densité de noyau polynomiale $\mathrm{PKDF}(\Phi,\Phi c,\psi)$ qui constitue l'instrument principal de la méthode KDE. La fonction nécessite un minimum de trois arguments : l'identifiant de la cellule de sortie de la grille Φ, le centroïde de la cellule de sortie de la grille Φc, et enfin la bande passante ou le rayon ψ. La forme fonctionnelle polynomiale a été retenue selon Sergerie et coll. (2021), bien que des formes normale et uniforme aient été disponibles. Φc a été définie comme géométrique, sans pondération selon la concentration de la densité spatiale des établissements dans l'ID et non pondérée en fonction de la concentration de densité spatiale de l'établissement de l'ID. Plusieurs spécifications et essais pour Φ et ψ ont été envisagés afin de garantir l'interprétabilité économique des résultats et l'efficacité de calcul. Ceux-ci sont brièvement documentés ci-dessous.

La dimension d'une cellule (ou tuile) pour une grille G est un paramètre de choix. Aux fins de cette analyse, une grille carrée dont la dimension des cellules est de 50 x 50 mètres a été adoptée. Après l'essai de différentes spécifications (p. ex. une plage de 10 x 10 à 100 x 100 mètres), la grille de 50 x 50 mètres a été adoptée pour trouver un équilibre entre le détail spatial et l'efficacité de calcul. La Figure 1 ci-dessous illustre la distribution du nombre de cellules de 50 x 50 mètres par l'ID dans les quatre régions métropolitaines. La Figure confirme qu'un nombre minimal et raisonnable de cellules sont disponibles dans la majeure partie de l'ID de chaque RMR. En outre, le nombre médian de cellules par ID varie de 7,5 à 12,5. Enfin, la Figure montre également les variations entre les RMR quant au nombre total de cellules pouvant s'intégrer dans un seul ID. Cette variation est particulièrement importante, car elle représente les caractéristiques des grandes et des petites RMR au Canada. Les petites RMR présentent généralement une gamme plus étendue de dimensions d'ID, tandis que les grandes RMR de l'étude ont une gamme plus restreinte et plus petite de dimensions d'ID. Il est intéressant de noter que Montréal est la seule RMR à afficher une distribution symétrique des ID, contrairement aux trois autres RMR où la distribution est plus asymétrique.

Figure 1
Nombre total de cellules de sortie par ID pour les quatre zones d'étude



Source : calculs des auteurs à partir de la base de données sur le RE.

De même que le choix de la dimension des cellules de la grille, le choix de la bande passante du noyau, ψ, a des implications sur les résultats. On peut comparer cela à un histogramme simple : une bande passante trop grande (équivalente à un histogramme avec très peu de barres) masque la distribution sous-jacente. À l'inverse, une bande passante trop petite peut entraîner une fréquence de 1 unité pour chaque résultat, rendant difficile la compréhension de la distribution des données avec un histogramme trop aplati.

Pour définir ψ, diverses spécifications ont été mises à l'essai, notamment la célèbre règle empirique de Silverman (1986), appliquée dans l'étude de Sergerie et coll. (2021). Compte tenu de la taille des cellules de la grille et de la configuration des ID utilisées dans notre analyse, la bande passante calculée par Silverman n'a pas permis d'inclure suffisamment de données dans son environnement local. Cela a entraîné une distribution spatiale insuffisamment lissée des lieux des emplois, rendant la méthode KDE inefficace et générant des grappes fragmentées. Le paragraphe suivant propose une interprétation des raisons pour lesquelles le processus de Silverman n'a pas fourni une taille de bande passante satisfaisante pour nos applications dans cette recherche.

L'équation de Silverman saisit la dispersion des lieux des emplois autour d'un point de référence dans la RMR. De la section précédente, nous savons que les lieux des emplois sont initialement géocodés à la localisation spatiale fixe de leurs établissements respectifs, puis distribués de manière aléatoire et uniforme dans les limites de leurs ID respectifs, sans donner la priorité au lieu de l'établissement ou à une sous-région précise de l'ID. Par conséquent, nous pouvons affirmer que l'équation de Silverman encapsule des renseignements partiels sur la distribution de la dimension des ID dans la RMR. Cependant, l'équation ne prend pas nécessairement en compte la dimension d'un ID type, qu'il s'agisse d'un ID moyen ou médian. Par conséquent, dans les exemples précis de nos applications, la bande passante de Silverman ne parvient pas à agréger les données entre les ID et propose uniquement une transformation des données du noyau dans les limites de l'ID de référence, laissant les résultats finaux à l'échelle des ID inchangés par rapport aux données d'origine du RE.

Nous proposons rapidement une interprétation de ce problème expliquant pourquoi la bande passante est trop petite. La bande passante de Silverman est définie par $0{,}9\mathrm{m}/s^{1/5}$, où s est la taille de l'échantillon et $\mathrm{m} = \min\left(\sigma, \left(1/\ln(2)\right)^{1/2} * \mu\right)^{1/2}$. Ici, σ est la dispersion standard de la distribution des points de données spatiales d'emploi X par rapport au point central moyen unique de la RMR et μ est le moment statistique de dispersion médiane (50e centile) dans la distribution de la dispersion de X au centre moyen. Il convient de noter qu'une autre version de l'équation de Silverman remplace la médiane μ par l'intervalle interquartile $\mathrm{IQR}()$, où $\mathrm{IQR}() = 75\%$ centile moins 25 % centile.

Autrefois, l'équation non pondérée pour la distance standard était,

$$\sigma = \sqrt{\left(\sum_{j=1}^{s} \frac{\left(L1(j) - L1C\right)^2}{s} + \sum_{j=1}^{s} \frac{\left(L2(j) - L2C\right)^2}{s}\right)}$$

où $L1$ et $L2$ sont les coordonnées numériques de longitude et de latitude des éléments de X, respectivement, et $L1C$ et $L2C$ sont les coordonnées numériques de longitude et de latitude du centre moyen de la RMR respectivement. Intuitivement, σ et μ ne peuvent pas être contenus dans un ID type ou médian de la RMR parce que les lieux des emplois sont largement disponibles spatialement sur l'ensemble des superficies de la RMR. Par conséquent, nous supposons que le minimum (min) des deux quantités est raisonnablement assez grand et ne peut pas expliquer une taille de bande passante Silverman trop petite. D'autre part, la taille de l'échantillon s Figure au dénominateur de la formule de Silverman et s peut fournir une très petite bande passante si s est trop grand. La distribution du nombre d'employés par établissement dans le RE peut être fortement asymétrique à droite, comprenant des valeurs aberrantes positives atteignant des valeurs très élevées, ce qui contribue à une valeur s très grande en raison de l'épaisseur de l'extrémité droite de la distribution. Par conséquent, si s est assez grand pour maintenir $s^{1/5}$ assez élevé, alors la bande passante sera petite, et irréaliste. Cette étude ne s'attarde pas sur le processus d'élimination des valeurs aberrantes des données du RE, mais cela pourrait faire l'objet d'une étude future. Plus précisément, l'application des travaux publiés par Statistique Canada, tels que Outliers in Sample Surveys de Lee et coll. (1992), A Cautionary Note on Clark Winsorization de Mulry et coll. (2016), A Method of Determining the Winsorization Threshold with an Application to Domain Estimation de Martinoz et coll. (2015), et On Searls' Winsorized Mean for Skewed Populations de Rivest et coll. (1995), est pertinente. Il convient également de noter que l'équation de Silverman est adaptée aux valeurs de densité du noyau normalement distribuées, ce qui n'est pas le cas pour notre distribution de densité asymétrique extrêmement droite présentée ci-dessus.

Enfin, il faut souligner que Mathematica utilise une modification de l'équation de Silverman, $0{,}09 * \sigma$, pour de grands échantillons s dépassant 100 000 observations. Cependant, nous n'avons pas utilisé le logiciel analytique Mathematica pour cette recherche et avons décidé de ne pas tirer parti de cette version de l'équation de Silverman. Intuitivement, la suppression du grand échantillon s du RE au dénominateur du rapport de Silverman contribuerait à résoudre le problème d'un dénominateur trop grand et d'une bande passante trop petite, si le remplacement d'un coefficient de 0,9 par 0,09 (réduction d'environ 90 %) ne réduit pas trop la bande passante.

Pour résoudre le problème lié à la règle empirique de Silverman dans nos applications, cette recherche développe sa propre méthodologie de bande passante qui se concentre sur la distribution des superficies d'ID fournies par les données. En d'autres termes, notre bande passante personnalisée tient compte de la dimension de l'ID médian dans la RMR de référence[9]. Cette nouvelle approche garantit que la bande passante est suffisante pour couvrir à la fois un ID type et les ID des quartiers. Cette condition est fondamentale, car le produit final de ce projet se situe à l'échelle de l'ID. Par conséquent, si la densité de chaque cellule de la grille représente uniquement l'agrégation des renseignements situés dans les limites de leur ID respectif, aucune information spatiale exclusive ne serait générée par cette recherche.

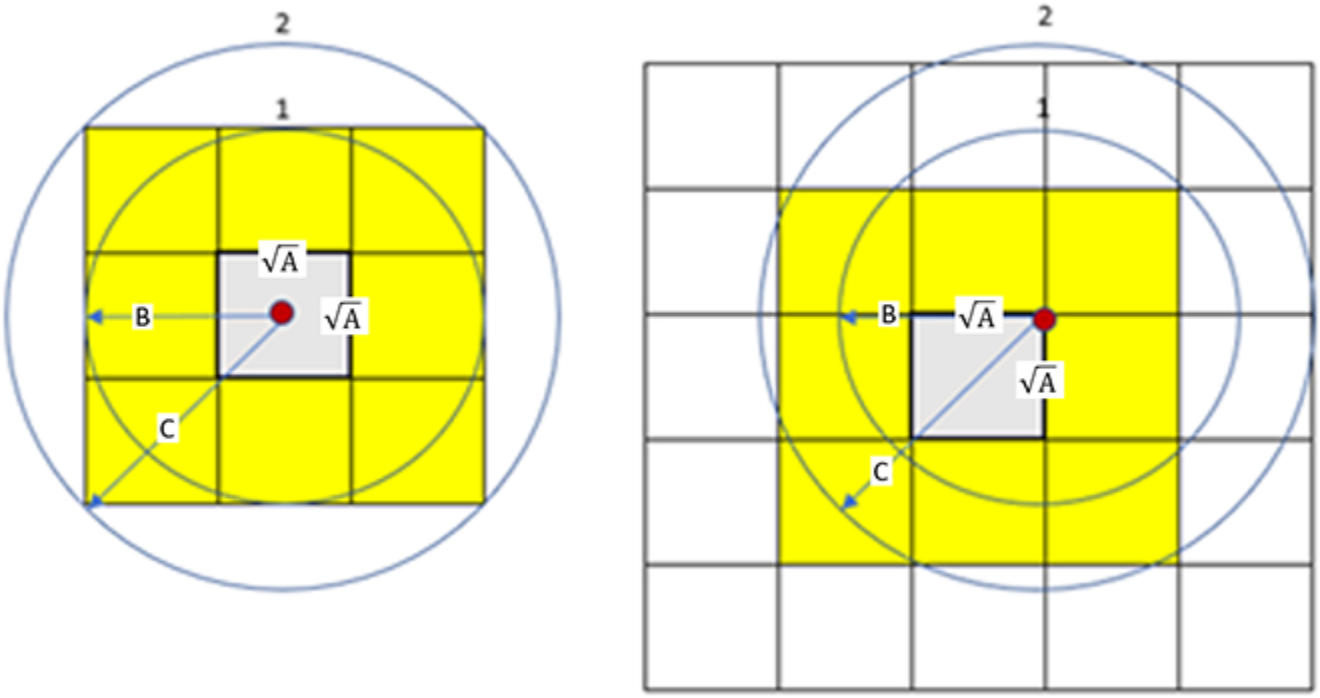
9. Étant donné que la distribution des dimensions des ID est généralement asymétrique, l'ID médian de la distribution a été utilisé à la place de l'ID moyen pour être plus robuste face aux valeurs aberrantes.

La Figure 2 illustre la méthode de la bande passante élaborée pour la présente analyse. La grille $G(\mathrm{mDB})$ remplie d'ID, à ne pas confondre avec la grille des cellules de sortie G susmentionnée, suppose un environnement local avec des ID au carré égaux dont les dimensions sont celles de l'ID médian empirique de la RMR de référence. L'ID médian est dérivée d'une zone projetée dans le système de référence de coordonnées (CRS) EPSG:3347. $\sqrt{A}$ est la longueur latérale d'un ID médian. Par conséquent, $\sqrt{A} * \sqrt{A}$ est le volume ou les superficies d'un ID médian. Les points rouges sont les centroïdes géométriques des cellules de la grille dans un ID médian de la grille $G(\mathrm{mDB})$ et ne doivent pas être confondue avec le centroïde géométrique d'un ID.

En se basant sur l'exemple à gauche, la bande passante nº 2, représentée par un cercle, parvient à couvrir l'intégralité des renseignements demandés, mis en évidence en gris et en jaune. C'est-à-dire que la bande passante couvre l'intégralité de l'ID de référence de la cellule et de ses huit ID voisins directs, respectivement. Il convient de noter qu'il s'agit de la bande passante la plus petite possible pour satisfaire à ces conditions. Plus précisément, cela couvre $\left(14{,}137 * A\right) / \left(9 * A\right) = 1{,}57$ fois la superficie demandée. Le calcul de la longueur de la bande passante repose sur une logique de distance euclidienne simple; $B = \sqrt{A} * 1{,}5$ par conséquent $C = \psi = \sqrt{\left(B^2 + B^2\right)}$. Cette logique garantit qu'un centroïde d'une cellule de sortie de la grille qui se superpose au centroïde géométrique de l'ID pourra capter toutes les données nécessaires. En outre, en se basant sur l'exemple de droite de la Figure 2, la cellule de sortie de la grille se situe maintenant dans le coin supérieur droit du même ID de référence. La bande passante nº 2 parvient à couvrir tous les ID demandés, à l'exception d'un ID situé en bas à gauche, dont la superficie est accessible à 50 %. Cette zone inaccessible est considérée comme acceptable dans notre méthodologie, car les nouveaux ID deviennent maintenant accessibles en haut à droite, même s'ils ne sont pas un voisin direct de l'ID de référence.

Les cartes thermiques des grappes produites à l'aide de cette nouvelle méthode de bande passante ont donné une surface lisse tout en préservant les détails des quartiers (voir les détails en annexe). Pour plus de clarté, la présente analyse reconnaît la notation de la bande passante sous la forme $\psi\left(\mathrm{mDB}\left(\mathrm{RMR}\right)\right)$, puisque le rayon personnalisé dépend maintenant d'un seul argument : la configuration de l'ID médian (IDM), et l'IDM est lui-même adapté à chaque RMR. C'est-à-dire que la bande passante varie d'une RMR à l'autre, mais reste fixe au sein de chaque RMR, quelle que soit la variance de la concentration des lieux des emplois dans les quartiers de la RMR. Cependant, la notation ψ restera en usage pour des raisons d'efficacité spatiale. Enfin, il convient de noter que notre bande passante personnalisée ψ ne dépend pas de la dispersion des lieux des emplois. Par conséquent, si les RMR A et B ont le même IDM et que la RMR B a deux fois plus de dispersion des lieux des emplois que la RMR A, alors les deux RMR obtiendront la même ψ, ce qui est inapproprié parce que la dispersion des lieux des emplois a son importance. D'autre part, notre bande passante personnalisée est robuste aux grandes valeurs aberrantes dans la distribution du nombre d'employés par établissement du RE. Une amélioration de notre bande passante comprendrait à la fois une dépendance à l'IDM et la dispersion des lieux des emplois.

Figure 2
Calcul de la largeur de bande de la densité du noyau pour la compréhension de la distribution des dimensions de l'ID



Légende : $\text{Longueur latérale d'un ID médian} = \sqrt{A}$, $A = \text{Superficie médiane de l'ID} = \sqrt{A} * \sqrt{A}$, $B = \sqrt{A} * 1{,}5$, et $C = \psi = \sqrt{(B^2 + B^2)}$

Source : calculs et méthodologie des auteurs.

Processus aléatoires multinomiaux uniformes et fonction de densité de noyau polynomiale

À la suite de la spécification des arguments Φ , Φc , et ψ , cette section documente le reste des détails de la fonction de densité de noyau polynomiale spatial (PKDF) appliqué dans l'analyse[10]. Ce modèle de densité de noyaux est en partie inspiré par Sergerie et al. (2021) concernant la façon d'appliquer un modèle de densité de noyaux à une population granulaire d'entreprises au Canada, Cependant, il prend la structure fondamentale des statisticiens Emanuel Parzen (1962) et Murray Rosenblatt (1956) qui ont créé indépendamment la forme théorique de densité du noyau. La notation utilisée est la nôtre. Cet article porte sur une contribution appliquée, avec la génération des cartes thermiques des clusters. Cet article n'a aucune contribution théorique, en dehors de la conception de la bande passante du noyau (présentée ci-dessus). Cet article utilise plusieurs résultats existants de la littérature pour l'explication des termes et concepts mathématiques théoriques de la densité du noyau, et pour l'explication des processus aléatoires appliqués avant le modèle de densité du noyau (expliqués ci-dessous), dans le cadre d'une population spatial de polygone d'ID et de localisation des emplois. Les termes restants à expliquer ci-dessous sont : i , $n()$, $JW()$, $JL()$, J , $\Theta()$ et $r()$.

$n(\Phi, \Phi c, \psi)$ représente le nombre fini non négatif des emplacements d'emplois disponibles dans l'environnement circulaire et symétrique $\pi\psi^2$ de la cellule ciblée de sortie de la grille Φ et inclus dans le calcul de la densité totale de la cellule. Par conséquent, $n()$ dépend de l'identifiant de la cellule de sortie de la grille évalué Φ , de l'emplacement du centroïde géométrique de la cellule Φc et de la longueur de la bande passante du noyau ψ . Par conséquent, le nombre d'ID qu'il couvre variera en fonction de l'endroit où Φ se situe. Il peut couvrir plusieurs

10. Une distribution normale pour l'approche de la densité du noyau pourrait être explorée dans des recherches futures. Cependant, pour ce projet, nous conservons le paramètre d'origine de Sergerie et al. (2021), car il définit déjà correctement la contribution de chaque lieu de travail à la densité finale. En outre, comme il sera décrit plus loin dans cet article, la forme polynomiale de la densité spatiale de noyaux se comporte de manière similaire à une approximation de série de Taylor pour une distribution normal et reproduit assez bien la forme d'une distribution normale pour le domaine d'intérêt.

ID ou moins d'un ID. i représente l'indice des emplois uniques inclus dans $\pi\psi^2$ autour de la cellule ciblée de sortie de la grille Φ et ne peut aller que de 1 à $\mathrm{n}()$.

Les pages suivantes documentent les processus aléatoires utilisés dans notre méthodologie. Cette documentation est importante pour comprendre la structure de ces processus. $\mathrm{JL}(\mathrm{i}) \sim \mathrm{U}(\mathrm{DBP})$ représente le lieu géographique de l'emploi i, qui, dans le cadre de l'estimation de la densité du noyau, est distribué de manière uniforme et aléatoire $(\mathrm{U}())$ sur la superficie du polygone de l'ID. Il ne se trouve pas nécessairement à l'emplacement de l'établissement d'origine auquel l'emploi i appartient. De plus, le lieu de l'emploi ne peut pas être distribué de manière aléatoire en dehors des limites de son propre ID, ni même à l'intérieur des limites de sa propre AD. En outre, tous les lieux des emplois i compris entre $\mathrm{i}=1$ et $\mathrm{i}=\mathrm{n}$ et appartenant au même ID sont i,i,d, c'est-à-dire indépendants et identiquement distribués à partir de la même distribution spatiale. Nous nommons ce processus aléatoire RP.

$\mathrm{JL}()$ est i.i.d. en ce sens qu'un emplacement de travail unique peut être attribué spatialement et aléatoirement indépendamment de l'emplacement aléatoire spatial d'un autre emploi unique. Cependant, comme nous l'expliquerons plus loin dans cet article, le processus aléatoire doit également être considéré comme le nombre d'emplois uniques attribuées au hasard sur un point spatial spécifique de l'ID. Dans cette perspective, plus le nombre de d'emplois uniques attribuées au hasard sur un point spatial spécifique est important, moins il est probable qu'un autre point spatial obtienne de nombreux emplois, car le nombre d'emplois uniques d'un ID est fini et non infini. Pour cette raison, l'idée clé est de visualiser les phénomènes comme des points de données (un grand ensemble d'emplois uniques) attribués aléatoirement aux multi-catégories (plusieurs points spatiaux sur un ID) d'une variable aléatoire multinomiale (un ID).

Quelques perspectives concernant le processus aléatoire (RP) méritent d'être mentionnées. Bien qu'il soit uniforme, le RP ne génère pas toujours une couverture uniforme pour un ID, certains quartiers de l'ID pouvant être plus couverts que d'autres. Plus précisément, l'événement le plus probable de notre processus RP est une allocation spatialement uniforme de points de données au sein de l'ID, où chaque point spatial reçoit le même nombre d'emploi. C'est notre événement attendu $\mathrm{E}(\mathrm{RP})$. D'autre part, l'événement le moins probable, mais toujours possible, se produit lorsque tous les points de données distincts se superposent à un endroit unique de l'ID. Il s'agit de notre événement exceptionnel $\mathrm{T}(\mathrm{RP})$.

Exemple pour ID de petite taille avec un processus aléatoire binomial uniforme

Pour mieux comprendre le processus aléatoire RP de notre méthodologie, une analogie peut être fournie avec un processus aléatoire binomial uniforme simple BRP avec b unique très petit ID de 2 espace spatial et v emploi unique par ID, où v est un grand nombre pair et b est un grand nombre. Pour construire l'intuition de manière simple, nous analysons principalement l'événement attendu et le plus rare de la distribution, et quelques autres événements entourant l'événement attendu et le plus rare, et non d'autres détails de la distribution. Analogiquement, une pièce de monnaie non biaisée propose un processus spatial uniforme composé de 2 points puisque les deux côtés de la pièce ont une probabilité égale et qu'il n'y a pas de priorité pour l'un des 2 côtés. Cependant, il n'y a aucune garantie que l'attribution pour pile et face sera uniforme, tout le temps. L'événement le plus probable est un mélange parfait de pile et de face. C'est-à-dire que l'événement attendu $\mathrm{E}(\mathrm{BRP})$ est une situation où chaque côté de la pièce reçoit une allocation égale de $\mathrm{v}/2$ emplois uniques. Cette situation arrivera à une très grande proportion des b ID. D'un autre côté, l'événement le plus improbable, mais toujours possible, est d'obtenir tous les v emplois d'un seul côté de la pièce (soit face, soit pile). C'est-à-dire que tous les emplois uniques sont attribués au hasard sur le même des 2 points spatiaux disponibles. Cette situation est notre événement $\mathrm{T}(\mathrm{BRP})$ et cela arrivera à une très petite proportion des b ID. Intuitivement, toutes les combinaisons possibles de v emplois attribuées sur l'espace spatial pile et face ont la même probabilité de réalisation, mais puisque l'événement de v pile ou v face est de 1 combinaison possible, respectivement, et que l'événement d'un mélange parfait de pile et de face est un très grand nombre de combinaisons distinctes, L'événement mixte parfait est alors plus susceptible de devenir l'événement attendu.

Formellement, nous avons la structure de nombre de possibilités suivante,

$$\sum_{j=0}^{v} C(v,j) = \sum_{j=0}^{v} \frac{v!}{(j)!(v-j)!} > Ec(BRP) = C\left(v,\frac{v}{2}\right) = \frac{v!}{\left(\frac{v}{2}\right)!\left(v-\frac{v}{2}\right)!}$$

$$> Tc(BRP) = C(v,0) = \frac{v!}{(0)!(v-0)!} = C(v,v) = \frac{v!}{(v)!(v-v)!} = 1$$

Où $C(v,j) = \binom{v}{j}$, est le nombre de façons de choisir j élément parmi v élément sans ordre[11] , l'inégalité de droite ($>$) est pour les événements les plus rares, l'inégalité du milieu ($>$) est pour l'événement attendu, l'égalité de gauche ($>$) est pour le nombre total de combinaisons possibles, et l'indice j de valeur 0 et de valeur v est pour v pile et v face, respectivement. À titre d'analyse supplémentaire, notons que $C\left(v,\frac{v}{2}-1\right) = C\left(v,\frac{v}{2}+1\right)$ représentent le nombre de combinaisons pour des événements qui sont à un pas d'une allocation spatiale parfaitement équilibrée, respectivement, et qui sont tout aussi moins susceptibles de se produire que l'événement attendu lié au nombre de combinaisons $C\left(v,\frac{v}{2}\right)$ mais toujours très susceptibles de se produire en probabilité et assez proches de la probabilité d'événement attendu. De la même manière, $C(v,0+1=1) = C(v,v-1)$ peut être considéré comme le nombre de combinaisons pour des événements qui sont un pas en avant d'une allocation spatiale plus uniforme, respectivement, et encore très peu probables de se produire, mais beaucoup plus susceptibles de se produire que le $T(BRP)$ (Grimaldi, 2003).

Nous devons également visualiser l'analyse en termes de somme de variables aléatoires de même pondération (moyenne non pondérée). C'est-à-dire que si nous étiquetons chacun des 2 points spatiaux de l'ID avec les valeurs numériques +1 et -1, alors pour l'événement attendu E(BRP), le nombre de +1 est égal au nombre de -1 et la moyenne est,

$$\left((+1) * (v/2) + (-1) * (v/2)\right)/v = 0$$

Si nous nous éloignons d'un pas d'une allocation spatiale parfaitement équilibrée, nous avons une moyenne de,

$$\left((+1) * \left((v/2)-1\right) + (-1) * \left((v/2)+1\right)\right) / v \text{ et } \left((+1) * \left((v/2)+1\right) + (-1) * \left((v/2)-1\right)\right) / v$$

qui sont différentes d'une moyenne de 0, respectivement, mais toujours proches de 0, respectivement.

Pour T(BRP), nous obtenons soit la valeur numérique -1, v fois d'affilée, soit la valeur numérique +1, v fois d'affilée, et la moyenne est respectivement,

$$-1 = (-1*v) / v \text{ et } +1 = (+1*v)/v$$

Si nous avançons d'un pas vers une allocation spatiale plus uniforme, nous avons une moyenne de,

$$\left((-1) * (v-1) + (+1) * (1)\right) / v \text{ et } \left((+1) * (v-1) + (-1) * (1)\right) / v$$

qui sont différentes d'une moyenne de -1 et +1 respectivement, mais toujours proches de -1 et +1, respectivement.

11. Cet article utilise le terme combinaison combinatoire et non permutation combinatoire. Le premier n'utilise pas la notion d'ordre, mais le second le fait. En effet, nous nous concentrons sur l'allocation spatiale finale des emplois et non sur les détails séquentiels de la façon dont l'allocation aléatoire a été réalisée. De plus, cette analyse porte sur le nombre d'occurrences (ou combinaisons) liées à un événement d'intérêt (l'évènement attendue et l'événement le plus rare) et non sur la probabilité de chaque occurrence. C'est-à-dire que, dans le cas de notre application, la probabilité de chaque occurrence est la même car les processus aléatoires sont uniformes. Par conséquent, la notion de probabilité de chaque occurrence n'est pas pertinente pour notre analyse comparative et le rapport entre la somme du nombre d'occurrences liées à un événement spécifique d'intérêt et le nombre total d'occurrences est suffisant pour obtenir une perspective de probabilité exacte (voir Annex 1 pour démonstration mathématique). Formellement, la probabilité d'une occurrence de l'événement le plus rare et d'une occurrence de l'événement moyen est la même en raison de l'uniformité. Ce qui veux dire que $\left(\frac{1}{2}\right)^{v} = \left(\frac{1}{2}\right)^{\frac{v}{2}} * \left(\frac{1}{2}\right)^{v-\frac{v}{2}}$.

Nous comprenons maintenant que la distribution d'ID ne concerne pas seulement une allocation spatiale parfaitement uniforme pour chaque ID de la RMR et qu'une dégradation symétrique et lisse de l'uniformité est présente des deux côtés (gauche et droite) de l'événement moyen, si le nombre d'emplois dans chaque ID est grand et si le nombre d'ID dans la RMR est grand. En répétant l'exercice pour l'ensemble du support de la distribution (et pas seulement pour les événements moyens et les plus rares), on obtient la forme d'une distribution approximative de la courbe en cloche. Il est maintenant temps de relier le théorème central limite univarié existant pour la distribution binomiale et sa variance finie à notre explication[12]. Autrement dit, asymptotiquement, si $v \to \infty$, alors BRP converge en distribution vers la distribution normale univariée ($BRP \to N()$)[13]. Dans un contexte fini, si v est suffisamment grand, nous pouvons approximer BRP[14] de telle sorte que[15],

$$BRP \approx N\left(v * 0.5, v * 0.5 * 0.5\right)$$

Où le terme 0.5 représente l'égalité des chances d'attribuer un emploi sur l'un des 2 points spatiaux de l'ID, le terme $v*0,5$ représente le nombre prévu d'emploi à attribuer sur l'un des 2 point spatial, ce qui représente la moitié du nombre total d'emploi uniques disponibles pour tous les établissements de l'ID, et où $v*0,5*0,5$ représente la variance du nombre d'emploi allouées sur l'un des 2 points spatial de l'ID (Severini, 2005). Enfin, dans un contexte fini, nous pouvons nous attendre à observer une distribution normale approximative réalisée à travers l'ensemble des ID d'une grande RMR si b est suffisamment grand dans la RMR et v est suffisamment grand dans chaque ID de la RMR, ou, dans un contexte théorique, une distribution normale si $b \to \infty$ dans la RMR et $v \to \infty$ dans chaque ID de la RMR. Ce paragraphe conclut notre exemple de processus aléatoire binomial.

Processus aléatoires multinomiaux uniformes pour un ID de grande taille

Les calculs mathématiques pour notre processus aléatoire principal RP sont une idée similaire, mais utilisent un processus multinomial uniforme (avec variance finit) au lieu d'un processus binomial uniforme (avec variance finit) car le nombre de points spatiaux par ID est fini[16] et grand et pas seulement égal à 2. De plus, d'après le théorème central limite multivarié (TCLM) existant pour le processus multinomial, si $v \to \infty$, alors la convergence en distribution du processus multinomial uniforme est une distribution normale multivariée plutôt qu'une distribution normale univariée (Severini, 2005) ($RP \to N()$). Dans un contexte fini, si v est très grand, le processus RP peut être approximé de la manière suivante :

$$RP \approx N\left(vp, vM\right)$$

où v est notre grand nombre habituel de notation de points spatiaux, et p est un vecteur de dimension S[17] et fait la somme de S probabilités égale à 1 pour le grand nombre de points S d'un ID typique d'une RMR. Pour des raisons de commodité, et sans perdre trop de généralité, v est également divisible par S pour permettre une allocation spatiale uniforme exacte[18]. Dans le cas particulier de nos applications[19],

$$p = \langle \frac{1}{S}, \frac{1}{S}, \ldots, \frac{1}{S}, \frac{1}{S} \rangle$$

et a une probabilité égale pour chacun des éléments S du vecteur car le processus RP est un processus aléatoire multinomial spatialement uniforme et ne donne la priorité à aucun point spatial parmi l'ensemble complet des

12. La manière dont nous relions le théorème central limite à l'explication est similaire à celle de Rao, 1981 (page 3, section 2.2 distribution asymptotique, paragraphe 1)
13. Toutes les distributions normales asymptotiques de cet article considèrent des distributions normales fixes, correctement mises à l'échelle et centrées. Par souci de simplification, la moyenne et la covariance ne sont pas affichées dans les expressions afin de ne pas focaliser l'analyse sur un vecteur espéré populé entièrement de valeur nulle. De plus, par souci de simplification, les processus aléatoires pour les contextes finis et asymptotiques conservent la même notation, même s'ils ne sont pas mis à l'échelle et centrés de la même manière.
14. Cet article utilise la notation RP comme un processus aléatoire car, techniquement, l'allocation aléatoire des tâches est séquentielle et comporte une dimension temporelle. Cependant, par souci de simplification, l'analyse se concentre uniquement sur l'allocation finale, sans tenir compte de la séquence.
15. Toutes les approximations de distribution normale dans cet article sont présentées comme non mises à l'échelle et non centrées.
16. Nous supposons un nombre fini de points spatiaux disponibles dans tout polygone d'ID puisque le nombre de pixels est fini et non infini.
17. Toutes les dimensions vectorielles de ce document utilisent initialement la notation à 1 rangée (ou vecteur horizontal) pour des raisons de commodité dans le texte. Par conséquent, l'utilisation de la transposition vectorielle se réfère à un vecteur à 1 colonne (ou vecteur vertical).
18. v est divisible par la quantité fixe S aux fins de simplifier l'analyse; cependant, nous ignorons cette simplification dans le contexte de la convergence $(v \to \infty)$.
19. Il convient de noter que si le processus aléatoire n'était pas exactement uniforme et que les voisinages des centroïdes des l'établissements étaient relativement plus prioritaires, alors le vecteur p élèverait les termes liés à ces points spatiaux dans ces voisinages ($> 1/S$) et rabaisserais tous les autres termes hors voisinage ($< 1/S$), de sorte que la somme des termes soit maintenue à 1.

points uniques de l'ID dans la RMR. Par conséquent, vp est le vecteur de S entrée représentant le nombre attendu d'emploi allouées à chacun des points spatiaux S de l'ID, qui est $\mathrm{v}*\left(1/S\right)$ pour chaque point de l'ID, car $\left(\mathrm{v}*1/S\right)*S=\mathrm{v}$. En d'autres termes,

$$\mathrm{vp}=\langle\frac{\mathrm{v}}{S},\frac{\mathrm{v}}{S},\ldots,\frac{\mathrm{v}}{S},\frac{\mathrm{v}}{S}\rangle$$

vM est la matrice de variance-covariance de dimension $S\mathrm{x}S$ de la distribution normale multivariée de notre processus aléatoire multinomial uniforme convergent. Elle s'exprime dans un contexte fini comme,

$$\mathrm{vM}=\begin{bmatrix} \mathrm{v}*\frac{1}{S}*\frac{S-1}{S} & (-1)*\mathrm{v}*\frac{1}{S}*\frac{1}{S} & \cdots & (-1)*\mathrm{v}*\frac{1}{S}*\frac{1}{S} \\ (-1)*\mathrm{v}*\frac{1}{S}*\frac{1}{S} & \mathrm{v}*\frac{1}{S}*\frac{S-1}{S} & \cdots & (-1)*\mathrm{v}*\frac{1}{S}*\frac{1}{S} \\ \vdots & \vdots & & \vdots \\ (-1)*\mathrm{v}*\frac{1}{S}*\frac{1}{S} & (-1)*\mathrm{v}*\frac{1}{S}*\frac{1}{S} & \cdots & \mathrm{v}*\frac{1}{S}*\frac{S-1}{S} \end{bmatrix}$$

$$=\begin{bmatrix} \frac{\mathrm{v}(S-1)}{S^2} & \frac{-\mathrm{v}}{S^2} & \cdots & \frac{-\mathrm{v}}{S^2} \\ \frac{-\mathrm{v}}{S^2} & \frac{\mathrm{v}(S-1)}{S^2} & \cdots & \frac{-\mathrm{v}}{S^2} \\ \vdots & \vdots & & \vdots \\ \frac{-\mathrm{v}}{S^2} & \frac{-\mathrm{v}}{S^2} & \cdots & \frac{\mathrm{v}(S-1)}{S^2} \end{bmatrix}$$

Et elle n'inclut que 2 composantes uniques parmi les $S\mathrm{x}S$ composantes non nul en raison de la simplification liée à l'uniformité. C'est-à-dire, la composante sur la diagonale et celle hors de la diagonale. La notation matricielle de vM suit la notation d'Ericson, 1969 (Annexe, équation A2) pour la manière appropriée de noter une matrice avec des termes égaux sur la diagonale, des termes égaux hors diagonale et lorsque le nombre de lignes et de colonnes est pair et grand. $\mathrm{M}=\mathrm{P}-\mathrm{pp}^{\mathrm{T}}$, où $\mathrm{P}=\mathrm{I}\mathrm{p}^{\mathrm{T}}$, et où I est la matrice d'identité de dimension $S\mathrm{x}S$. C'est-à-dire que P est une matrice diagonale dont les éléments diagonaux sont les éléments du vecteur p .

$$\mathrm{P}=\mathrm{I}\,\mathrm{p}^{\mathrm{T}}=\begin{bmatrix} 1/S & 0 & \cdots & 0 \\ 0 & 1/S & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/S \end{bmatrix}$$

Également,

$$\mathrm{pp}^{\mathrm{T}}=\frac{1}{S*S}+\frac{1}{S*S}+\ldots+\frac{1}{S*S}+\frac{1}{S*S}=\frac{S}{S*S}=\frac{1}{S}$$

La matrice variance-covariance vM informe que la variance de l'attribution d'emploi à n'importe quel point spatial est $\mathrm{v}*\frac{1}{S}*\frac{S-1}{S}$ et que la covariance entre n'importe quel point spatial distinct de l'ID est $(-1)*\mathrm{v}*\frac{1}{S}*\frac{1}{S}$.
Le terme négatif de la covariance est dû à la probabilité moindre d'obtenir plus de nombre d'emplois sur l'un des points spatiaux à mesure que le nombre d'emplois de l'autre point spatial augmente (Aitkin, 2022).

La distribution normale multivariée de notre processus aléatoire multinomial spatialement uniforme convergent peut être exprimée dans un contexte fini de la manière suivante,

$$\mathrm{y}'=\mathrm{SVN}\left(\mathrm{x}';\mathrm{vp},\mathrm{vM}\right)=\left(2\pi\right)^{-\frac{S}{2}}*\det\left(\mathrm{vM}\right)^{-\frac{1}{2}}*\exp\left(-\frac{1}{2}\left(\mathrm{x}'-\mathrm{vp}\right)\left(\mathrm{vM}\right)^{-1}\left(\mathrm{x}'-\mathrm{vp}\right)^{\mathrm{T}}\right)\text{ et }\mathrm{x}'\approx\mathrm{N}\left(\mathrm{vp}^{\mathrm{T}},\mathrm{vM}\right)$$

Tous les termes de la distribution normale S-variée[20] $SVN(x';vp,vM)$ sont déjà définis ci-dessus, à l'exception de x'[21], qui est un vecteur ligne de dimension S, et composé de S entrée pour le nombre respectif d'allocation d'emploi pour les S point spatial unique de notre ID. En dépit d'être de nature multivariée et composé de vecteurs et de matrices, $SVN(x';\ vp,vM)$ produit une valeur de sortie de type scalaire y' et prend la forme d'une distribution normal univariée (Aitkin, 2022). $SVN(x';\ vp,vM)$ atteint sa densité[22] maximale lorsque $x' = vp$, qui est le vecteur d'allocation des emplois attendu présenté ci-dessus. C'est-à-dire, lorsque $x' = \left(\frac{v}{S}, \frac{v}{S}, \ldots, \frac{v}{S}, \frac{v}{S}\right)$.

Cette affirmation a été soulignée par Thompson et al, 2022 (page 78 (8) équation 2.9) pour le cas général d'une distribution normale multivariée avec transformation logarithmique. De plus, $SVN(x';\ vp,vM)$ atteint sa densité minimale pour $x' = (v,\ 0,\ \ldots\ ,\ 0,\ 0)$, ce qui est une situation où le total des emplois uniques des établissements de l'ID sont attribué exclusivement au premier point spatial. La même densité minimale sera atteinte pour tout autre point spatial unique de l'ID. Enfin, $det(vM)$ ou $|vM|$ représente le déterminant de la matrice variance-covariance vM. $|vM|$ a une valeur numérique scalaire différente de zéro, car la matrice vM est non-singulière et la matrice inverse de vM, $(vM)^{-1}$, existe. Ce paragraphe complète notre explication de la stabilité des processus aléatoires impliqués dans notre méthodologie. En d'autres termes, nous avons documenté l'idée que nos processus aléatoires se rapportent à la distribution normale et que l'utilisation de variables aléatoires convergent vers un résultat bien centré et ne sont pas synonymes d'imprécision.

Il convient de noter que, puisque le processus RP pour l'attribution des tâches est un processus aléatoire, toute autre mesure générée à partir de ce prétraitement est donc également une variable aléatoire. Cela inclut essentiellement toutes les étapes du programme informatique du KDE. La quantification de l'incertitude due aux processus aléatoires de notre méthode KDE n'est pas l'objet du présent article, mais pourrait constituer une recherche future. Cette rubrique est pertinente, car la précision spatiale des quartiers locaux est importante et pourrait varier uniquement en fonction de certaines réalisations aléatoires d'événements exceptionnels ou d'autres événements moins uniformes. De autre côté, si le nombre d'emplois est grand, alors l'ampleur de l'incertitude devrait également être raisonnable. De plus, comme documenté ci-dessus théoretiquement, si les processus aléatoires se rapportent à la distribution normale (c'est-à-dire, la loi de Student avec degré de liberté $\to \infty$), alors la plupart des événements aléatoires seront assez centrés vers l'événement attendu parfaitement équilibré et les queues[23] bilatérales de la distribution seront de densité mince pour les événements rares et extrêmement déséquilibrés, ce qui est l'opposé de ce que serait une distribution de Cauchy (c'est-à-dire, loi de Student avec 1 degré de liberté) (Fisher, 1925) and (Hurst, 2010).

Fonction de densité de noyau polynomiale

Nous poursuivons maintenant avec le reste des paramètres de la méthodologie. $JW(i)$ représente la variable de poids des emplois pour l'emploi i, où i est un entier positif inférieur ou égal à $n()$, de plus, JW est égal à l'unité, c'est-à-dire qu'il prend la valeur de 1 pour tous les emplois uniques. L'option de remplacement aurait été le scénario extrême, qui consiste à allouer spatialement un nombre de points d'emploi égal au nombre d'établissements uniques dans l'ID et de sélectionner un poids égal au nombre d'emplois uniques dans l'établissement respectif pour chaque emplacement de travail. Ce scénario exploite mieux la variable de poids $JW(i)$, car elle n'est pas égale à l'unité. Cependant, il ne favorise pas le prétraitement et le nettoyage des données pour l'application de la méthode KDE, car il rend la distribution spatiale discrète des emplois plus clairsemée

20. Toutes les dimensions des vecteurs et des matrices sont techniquement basées sur $S-1$, et non sur S. Nous conservons S par souci de simplification dans la notation. Par définition, la dernière catégorie d'une variable aléatoire multinomiale est une information redondante, car la somme des nombres par catégorie est une contrainte totale fixe, et le nombre d'emplois alloués à la dernière catégorie est automatiquement déduit de la somme de toutes les autres catégories. Par conséquent, la matrice de variance-covariance est déficiente en rang et non en rang complet. Pour éviter la singularité, l'impossibilité d'inversion de matrice et un déterminant matriciel incorrect, la dimension de la distribution normale multivariée doit être réduite à un sous-espace de S, qui est simplement $S-1$.
21. Cet article utilise les variables x, x' et x''. Il ne s'agit pas d'une variation de la même variable. Ce sont des variables distinctes utilisées pour des raisons de commodité. x est un rapport d'entrée dans la fonction de densité du noyau, x est également utilisé comme symbole mathématique pour la description de la dimension des matrices, x' est un vecteur d'entrée pour la distribution normale multivariée et x'' est la variable d'entrée d'une série de Taylor. y, y' et y'' font partie des même équations, respectivement, en se qui concerne les composants de sortie.
22. Cet article utilise le terme « densité » car la distribution est une approximation normale. Le terme « masse » est évité par souci de simplification, même si le vecteur d'entrée de la normale est fini et constitué des catégories du multinomial.
23. L'utilisation du théorème central limite et des approximations normales pour les multinomiaux sera indéniablement plus fiable pour l'inférence statistique autour de la moyenne de la distribution. Néanmoins, nous prenons le temps ici de mentionner les caractéristiques des queues, car elles sont plus pratiques que celles de la distribution de Cauchy.

et plus difficile à lisser. Le paramètre $JW(i)=1$ est privilégié, car il repose sur un très grand nombre d'entités pondérées de manière égale réparties dans l'espace, permettant de lisser au préalable la distribution spatiale d'origine des emplois avant d'appliquer la méthode KDE.

Pour une notation plus compacte des derniers termes introduits dans les paragraphes précédents, nous définissons l'ensemble J de la dimension $n \times 2$[24] comme étant $\{(JW(i), JL(i)); i \in (1,2,\ldots,n-1,n)\}$ et nous l'incluons comme un argument dépendant de la $PKDF()$. En outre, $\Theta(JL(i),\Phi c,\psi)$ est la distance euclidienne entre le centroïde géométrique de la cellule de sortie de la grille et le lieu de l'emploi aléatoire i. Il dépend de $JL(i)$, Φc, et de la taille de la bande passante ψ. Le rapport $1/\psi^2$ est un simple terme de normalisation en dehors de la sommation, constant pour toutes les cellules de sortie de la grille à l'intérieur d'une RMR et qui ne varie qu'entre les différentes RMR. Enfin, si nous remplaçons les deux exposants au carré par une répétition de leurs termes, le rapport

$$r(\Theta,\psi,i) = \left(1-\left(\frac{\Theta(JL(i),\Phi c,\psi())}{\psi()}\right) * \left(\frac{\Theta(JL(i),\Phi c,\psi())}{\psi()}\right)\right) * \left(1-\left(\frac{\Theta(JL(i),\Phi c,\psi())}{\psi()}\right) * \left(\frac{\Theta(JL(i),\Phi c,\psi())}{\psi()}\right)\right)$$

devient le terme clé de la fonction de densité du noyau. Le rapport $r()$ alloue correctement un emplacement d'emploi i plus proche du centroïde géométrique de la cellule de sortie de la grille avec une plus grande contribution individuelle à la densité totale de la cellule de référence. Pour sa part, un emplacement d'emploi i plus éloigné du centroïde géométrique de la cellule reçoit une contribution individuelle moindre. Formellement, la fonction de densité de noyau polynomiale génère la valeur de densité totale y et peut être notée comme suit :

$$y = PKDF(X,\Phi,\Phi c,\psi(),J(JW(),JL())) = \frac{1}{\psi()^2} * \sum_{i=1}^{n(\Phi,\Phi c,\psi)} \left(\frac{3}{\pi} * JW(i) * \left(1-\left(\frac{\Theta(JL(i),\Phi c,\psi())}{\psi()}\right)^2\right)^2\right)$$

$PKDF()$ dépend de X, l'a distribution spatiale complète des emplois de la RMR, et non seulement du sous-ensemble des éléments d'emploi de X indexés sous i pour une cellule de grille d'intérêt Φ. Dans cet article, afin de simplifier la méthodologie, X est considérée comme une distribution fixe, non soumise à l'incertitude liée à un modèle de super population générant une réalisation aléatoire complète d'une population de RMR, ni à un échantillon aléatoire ou non-aléatoire visant à représenter cette population.

Pour des raisons de simplicité, nous notons le rapport $\frac{\Theta()}{\psi()}$ inclus dans le rapport $r()$ sous la forme du rapport x. Nous notons également $\frac{3}{\pi} * JW(i) * r() = \frac{3}{\pi} r = f(x)$. Une étude rapide de la fonction est essentielle pour comprendre la forme de $PKDF()$, $f(x)$ qui peut être exprimée sous la forme $\frac{3}{\pi} - \frac{6}{\pi}x^2 + \frac{3}{\pi}x^4 \approx 0.95 - 1.91x^2 + 0.95x^4$. La première dérivée de $f(x)$ est $\frac{df(x)}{dx} = f'(x) = -\frac{12}{\pi}x + \frac{12}{\pi}x^3$ et $f'(x)$ est égale à zéro pour $x \in \{-1,0,+1\}$. La deuxième dérivée de $f(x)$ est $\frac{d^2f(x)}{dx^2} = f''(x) = -\frac{12}{\pi} + \frac{36}{\pi}x^2$ et $f''(x)$ gal à zéro pour $x \in \{-\left(\frac{1}{3}\right)^{1/2} = -0{,}577, +\left(\frac{1}{3}\right)^{1/2} = +0{,}577\}$. Par conséquent, $f(x)$ est une fonction polynomiale décroissante pour $x \in \{(-\infty,-1) \cup (0,+1)\}$, c'est-à-dire $df(x)/dx < 0$. $f(x)$ est une fonction

24. Cet article utilise rarement x pour définir les dimensions des éléments. Cependant, x est principalement utilisé dans un contexte mathématique pour les variables, les vecteurs et les processus.

polynomiale croissante pour $x \in \{(-1,0) \cup (+1,+\infty)\}$, c'est-à-dire $df(x) / dx > 0$. $f(x)$ est concave vers le haut pour $x \in \{(-\infty,0,577) \cup (+0,577,+\infty)$, c'est-à-dire $d^2f(x) / dx^2 > 0$. $f(x)$ est concave vers le bas pour $x \in (-0,577,+0,577)$, c'est-à-dire $d^2f(x) / dx^2 < 0$. En outre, le domaine d'origine de $f(x)$ n'est pas limité à un sous-ensemble précis des nombres réels R, c'est-à-dire $x \in (-\infty,+\infty)$. L'image d'origine de $f(x)$ est l'ensemble des nombres réels non négatifs R^+, c'est-à-dire $f(x) \in [0,+\infty)$. Enfin, même si elle est polynomiale, la forme de $f(x)$ est similaire à une distribution normale pour le domaine $x \in [-1,+1]$. Autrement dit, $f(x)$ est similaire à une approximation en série de Taylor (somme infinie de termes polynomiales) pour une distribution normale univarié dans un voisinage centré à zéro. Formellement,

$y''(x'') = \frac{1}{\sqrt{2\pi * \sqrt{1}}} e^{\frac{-(x''-0)^2}{2*1}}$ et $x'' \sim N(0,1)$, par définition d'une distribution normale univariée de moyenne 0 et de variance 1, et $y''(x'') = \frac{1}{\sqrt{2\pi}} e^{\frac{-x''^2}{2}}$

$= \frac{1}{\sqrt{2\pi}} * \left(\sum_{w=0}^{+\infty} (-1)^w * (x'')^{2w} / 2^w * w! \right)$, par égalité exacte du résultat expérimental de la série de Taylor[25], et

$= \frac{1}{\sqrt{2\pi}} * \left(1 - \left(\frac{1}{2*1!}\right) * x''^2 + \left(\frac{1}{4*2!}\right) * x''^4 - \ldots\right)$, par définition d'une somme infinie, et

$= \frac{1}{\sqrt{2\pi}} - \left(\frac{1}{\sqrt{2\pi}} * \frac{1}{2*1!}\right) * x''^2 + \left(\frac{1}{\sqrt{2\pi}} * \frac{1}{4*2!}\right) * x''^4 - \ldots$, par simple propriété de distributivité, et

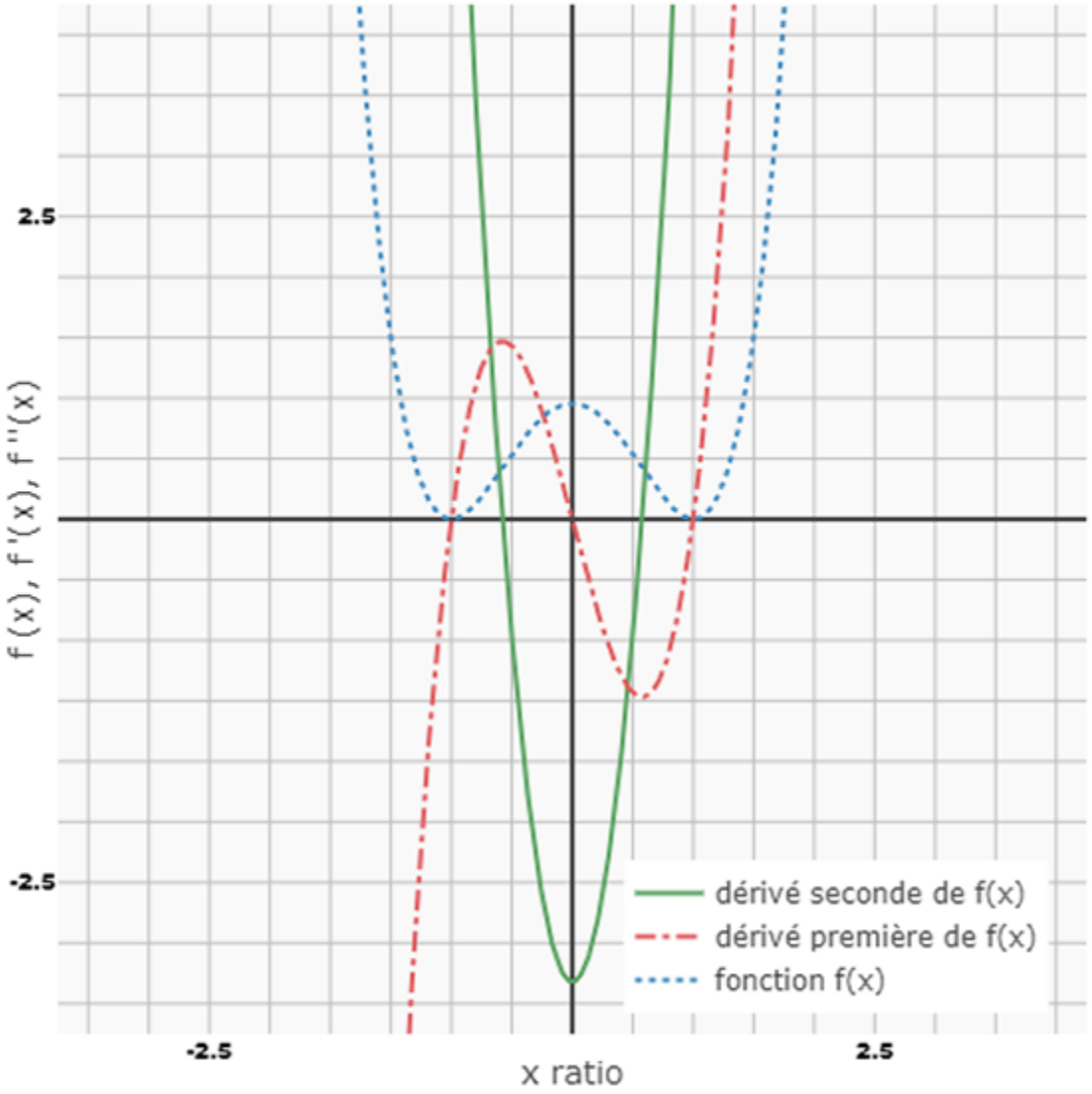
$\approx \frac{1}{\sqrt{2\pi}} - \left(\frac{1}{\sqrt{2\pi}} * \frac{1}{2*1!}\right) x''^2 + \left(\frac{1}{\sqrt{2\pi}} * \frac{1}{4*2!}\right) x''^4$, après avoir supprimé une infinité de termes de grandeur respective négligeable en comparaison des 3 premiers termes, et

$\approx 0.40 - 0.20x''^2 + 0.05x''^4$, après simplification arithmétique et un arrondissement difficile des termes π

La dernière approximation ci-dessus, $0.40 - 0.20x''^2 + 0.05x''^4$, est suffisamment[26] similaire à l'expression de $f(x) \approx 0.95 - 1.91x^2 + 0.95x^4$ et pour 2 raisons; 1) la structure signe +/- est la même, 2) la structure des termes exponentiels est la même. Cependant, la structure de l'amplitude des coefficients n'est pas la même. C'est-à-dire que $f(x)$ est assez stable tandis que l'autre équation est fortement décroissante, ce qui est typique de la série de Taylor. De plus, le troisième coefficient de magnitude 0,05 est suffisant pour comprendre la négligence respective relative de l'infinité restante de termes supprimés par rapport aux 3 premiers. Les similitudes entre $y''(x'')$ et $f(x)$ sont pertinentes pour comprendre le lien théorique de $f(x)$ avec la distribution normale, d'autant plus que la normale était un paramètre optionnel dans Sergerie et al., 2021. La Figure 3 présente les trois courbes d'intérêts pour comprendre l'image complète de $f(x)$, à savoir la fonction $f(x)$ en bleu, sa fonction de dérivée première en rouge et sa fonction de dérivée seconde en vert. L'axe des x représente le rapport x (Bartle and Sherbert, 2011).

25. Taylor Series Approximation for the Normal Distribution (Remarque : cette source est expérimentale mais utile pour comprendre notre modèle de densité de noyau)
26. La comparaison est effectuée à l'aide des expressions mises à l'échelle, et non des expressions non mises à l'échelle.

Figure 3
Fonction de densité du noyau polynomiale (zoom arrière pour f(x), f'(x) et f"(x))

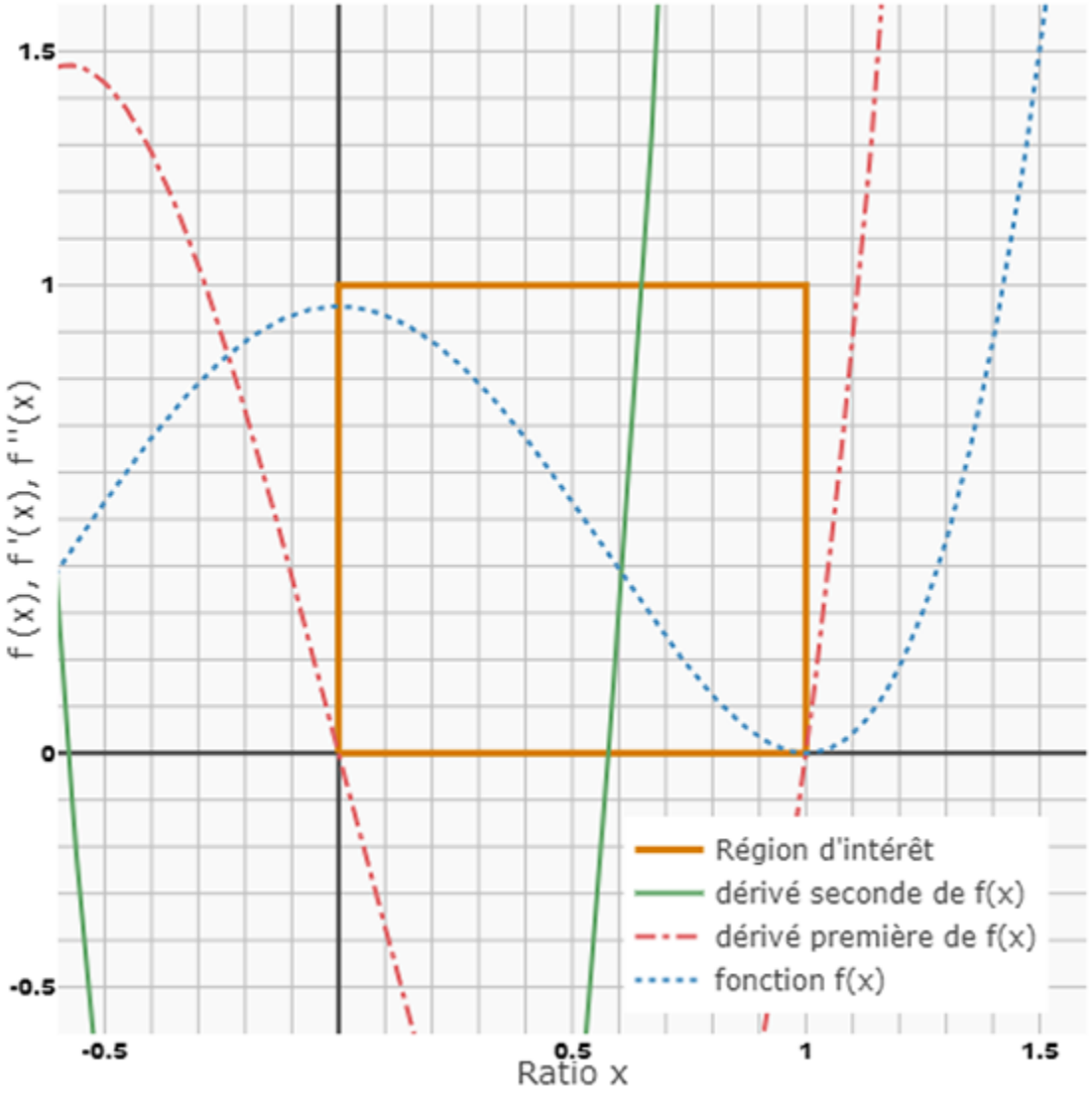


Source : calculs des auteurs à partir de la fonction d'estimation de la densité du noyau de Parzen (1962) et Rosenblatt (1956).

Maintenant que la fonction de densité du noyau polynomiale est bien décrite mathématiquement, nous nous concentrons sur le sous-domaine pertinent de la fonction $f(x)$ qui est $[0,1]$ et non $(-\infty\ ,\ +\infty)$. En effet, le rapport x n'a aucune utilité en dessous de zéro ni au-delà de 1 dans le contexte d'un exercice d'estimation de la densité, car Θ ne peut pas être inférieur à zéro (c'est-à-dire qu'une distance négative du point d'intérêt spatial n'est pas pertinente) et ne peut pas dépasser ψ (dépasser la zone circulaire ciblée n'est également pas pertinent). Par conséquent, l'image pertinente de $f(x)$ est $[0,\frac{3}{\pi}]$, parce que $f\ (0)\ =\ \frac{3}{\pi}$ et $f\ (1)\ =\ 0$, et f est une fonction décroissante entre 0 et 1. Plus important encore, la contribution distincte d'un emplacement d'emploi i dans le rayon ciblé ψ est approximativement normalisé entre 0 et 1 [c'est-à-dire entre les valeurs 0 et $(3\ /\ 3{,}14159)\ =\ 0{,}95)$]. De plus, $\mathrm{PKDF}()$ est une somme finie de termes approximativement normalisés puisque le nombre de lieux des emplois dans l'environnement circulaire et symétrique $\pi\psi^2$ est fini. La Figure 4 est identique à la Figure 3 précédente. Cependant, elle se concentre sur le domaine et l'image pertinents de $f(x)$ et

permet donc de mieux visualiser la valeur de $f(x)$ lorsque $x=0$, qui n'est pas 1, mais 0,95. La contribution du lieu de travail i diminue lorsque x augmente entre $[0,+1]$. La diminution est initialement moins forte autour de zéro, mais devient ensuite plus marquée puis moins forte une fois de plus juste avant d'atteindre un rapport x égal à 1. Cette structure est due au fait que la dérivée seconde (courbe verte) atteint une valeur nulle lorsque x atteint +0,577. Un carré surligné en jaune délimite la région d'intérêt (Bartle and Sherbert, 2011).

Figure 4
Fonction de densité du noyau polynomiale (zoom avant pour le domaine [0,1] de f(x), f'(x) et f"(x))



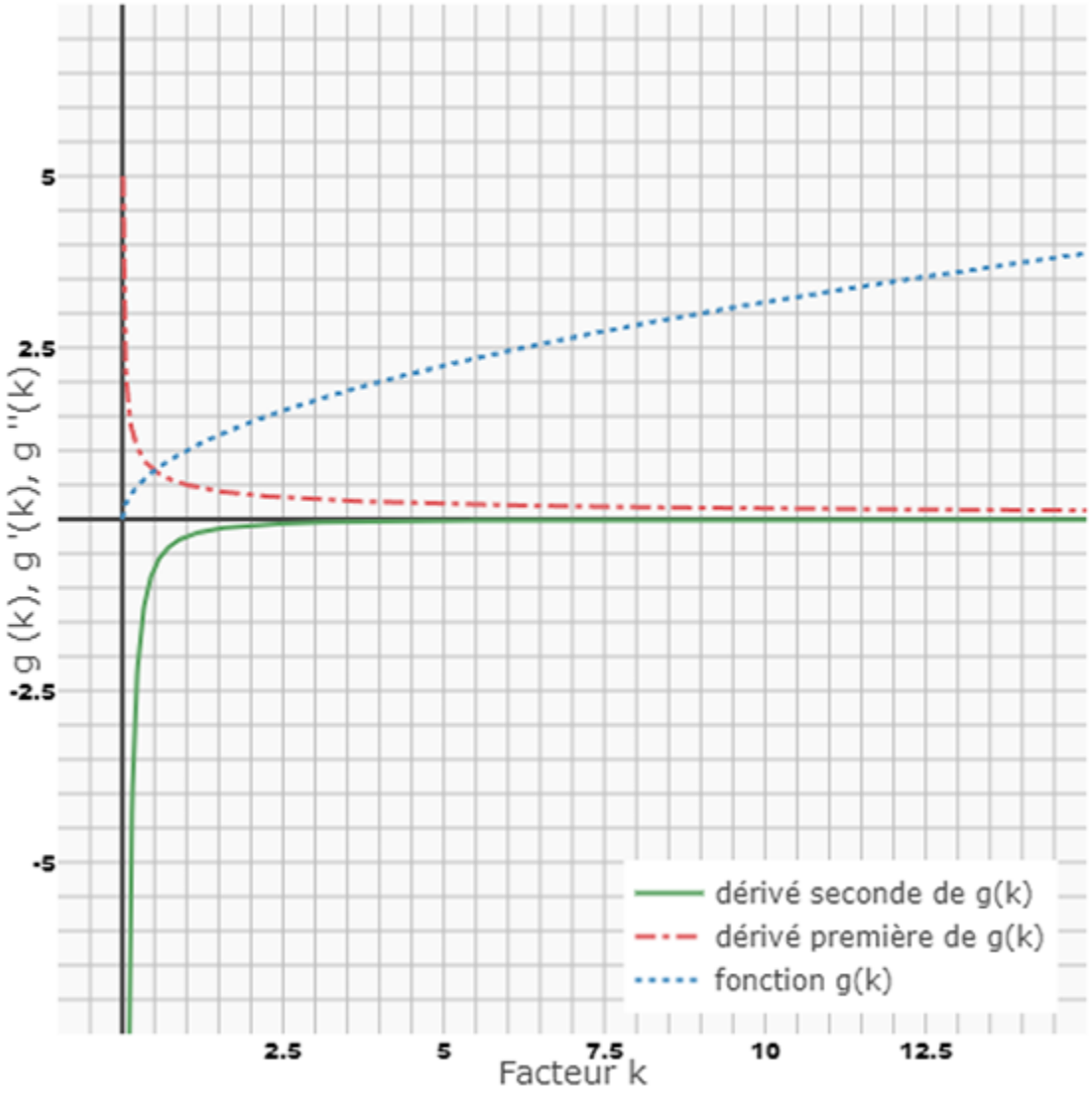
Source : calculs des auteurs à partir de la fonction d'estimation de la densité du noyau de Parzen (1962) et Rosenblatt (1956).

Connexion spatiale entre notre étude et Sergerie et al. (2021)

Avant de passer à la dernière partie de la méthodologie principale de cet article où les étapes de traitement de données sont effectuées sur la base de la distribution de densité du noyau estimée pour les RMR et les SCIAN d'intérêt, nous fournissons rapidement une explication intuitive de la dynamique entre la densité spatiale des emplacements d'emploi et la distance spatiale des emplacements d'emploi. Cette explication est essentielle pour bien comprendre la nature même de nos données et ses limites. Cette explication fournit également des liens entre notre étude et Sergerie et al. (2021) en termes de cartographie spatiale. Un article du Bureau du recensement des États-Unis (Biemer et Stokes, 1984), souligne la phrase suivante : « Justification : La distance moyenne entre des points répartis aléatoirement dans un plan est augmentée de $k^{1/2}$ lorsque la densité de ces points est diminuée d'un facteur k ». Cette explication est simple et intuitive, cependant, pour mieux comprendre leur idée, nous générons la fonction $g(k) = k^{1/2}$ et effectuons une étude rapide de ses propriétés de dérivées premières et secondes, de son domaine et de son image. Dans la Figure 5, l'axe horizontal est l'axe k et un nombre positif k représente le facteur de diminution de la densité spatiale. La courbe bleue est la fonction $g(k) = k^{1/2}$ et elle représente l'augmentation de la distance moyenne entre les points spatiaux. Le domaine et l'image de la fonction $g(k) = k^{1/2}$ sont l'ensemble des nombres réels non négatifs R^+. En d'autres mots, k et $k^{-1/2} \in [0,+\infty)$. La courbe rouge est la première dérivée $g'(k) = \frac{d}{dk}k^{\frac{1}{2}}$ et elle est égale à $\frac{1}{2} * k^{-\frac{1}{2}}$ et il n'y a pas de valeur de k telle qu'elle soit égale à zéro même si elle converge vers zéro, à mesure que k tend vers l'infinie $\left(\frac{1}{2} * k^{-\frac{1}{2}} \to 0\right)$. Son domaine et son image sont tous deux $(0, +\infty)$ car aucun zéro ne peut être au dénominateur. La courbe verte est la dérivée seconde $g''(k)=\frac{d^2}{dk^2}k^{1/2}$ et elle est égale à $-\frac{1}{4} * k^{-\frac{3}{2}}$ et encore une fois, il n'y a pas de valeur de k telle qu'elle soit égale à zéro même si elle converge vers zéro, à mesure que k tend vers l'infinie $\left(-\frac{1}{4} * k^{-\frac{3}{2}} \to 0\right)$, et à un rythme de convergence plus rapide que la fonction dérivée première en raison de la plus grande amplitude absolue des paramètres au dénominateur ($g'(k) = O\left(k^{-\frac{1}{2}}\right)$, $g''(k) = O(k^{-3/2})$, et $k^{\frac{1}{2}} < k^{3/2} \Leftrightarrow 1/k^{3/2} < 1/k^{\frac{1}{2}}$). Son domaine et son image sont respectivement $(0, +\infty)$ et $(-\infty, 0)$ (Bartle et Sherbert, 2011). Par conséquent, nous expliquons la justification de (Biemer et Stokes, 1984) de la manière suivante pour nos propres applications concernant une grande population de lieux de travail répartis spatialement au sein d'une population d'ID dans une RMR: pour un nombre grand et fixe d'emplacements d'emploi uniformes dans l'ID d'une RMR, alors que les frontières du polygone de l'ID tendent vers l'infini $(DBP \to \infty)$, et que la densité spatiale des emplacements d'emploi tend vers l'infini dans l'ID $(k \to \infty)$ [27], et que la distance spatiale moyenne entre les lieux de travail tend vers infiniment ($k^{1/2} \to \infty$), et que l'uniformité spatiale aléatoire de l'emplacement du travail est préservé dans l'ID, la variation de la distance spatiale moyenne entre l'emplacement du travail converge en probabilité vers 0 entre les incréments. Dans un context finit, pour 2 expansions larges et distinctes, k and $k+\Delta$, avec $\Delta > 0$, et chaque expansion est implémentée indépendamment à partir des frontières initiales du polygone de l'ID, alors $(k+\Delta)^{1/2} - k^{1/2} \approx 0$. Les informations fournies dans ce paragraphe établissent un lien théorique entre notre étude et celle de Sergerie et al. (2021). Autrement dit, il s'agit des éléments permettant de comprendre le comportement des données d'un ID si la surface aléatoire s'étendait d'un ID à une aire de diffusion (AD) ou même à une aire de diffusion agrégée (ADA).

27. Dans cet article, nous avons supposé un nombre fini de points spatiaux disponibles par ID, car le nombre de pixels par ID est fini et non infini. Cela était nécessaire pour simplifier la compréhension d'un processus aléatoire multinomial uniforme. Cependant, pour les besoins théoriques de l'explication ci-dessus, il est préférable de supposer une perspective infinitésimale pour le nombre de points spatiaux disponibles car les emplois doivent pouvoir se déplacer spatialement de manière continue, différentiable et non discrète. En d'autres mots, le nombre de points spatiaux est infinie et pas nécessairement dénombrable.

Figure 5
Dynamique entre la valeur de la densité spatiale et la distance spatiale entre les points (g(k), g'(k) et g"(k))



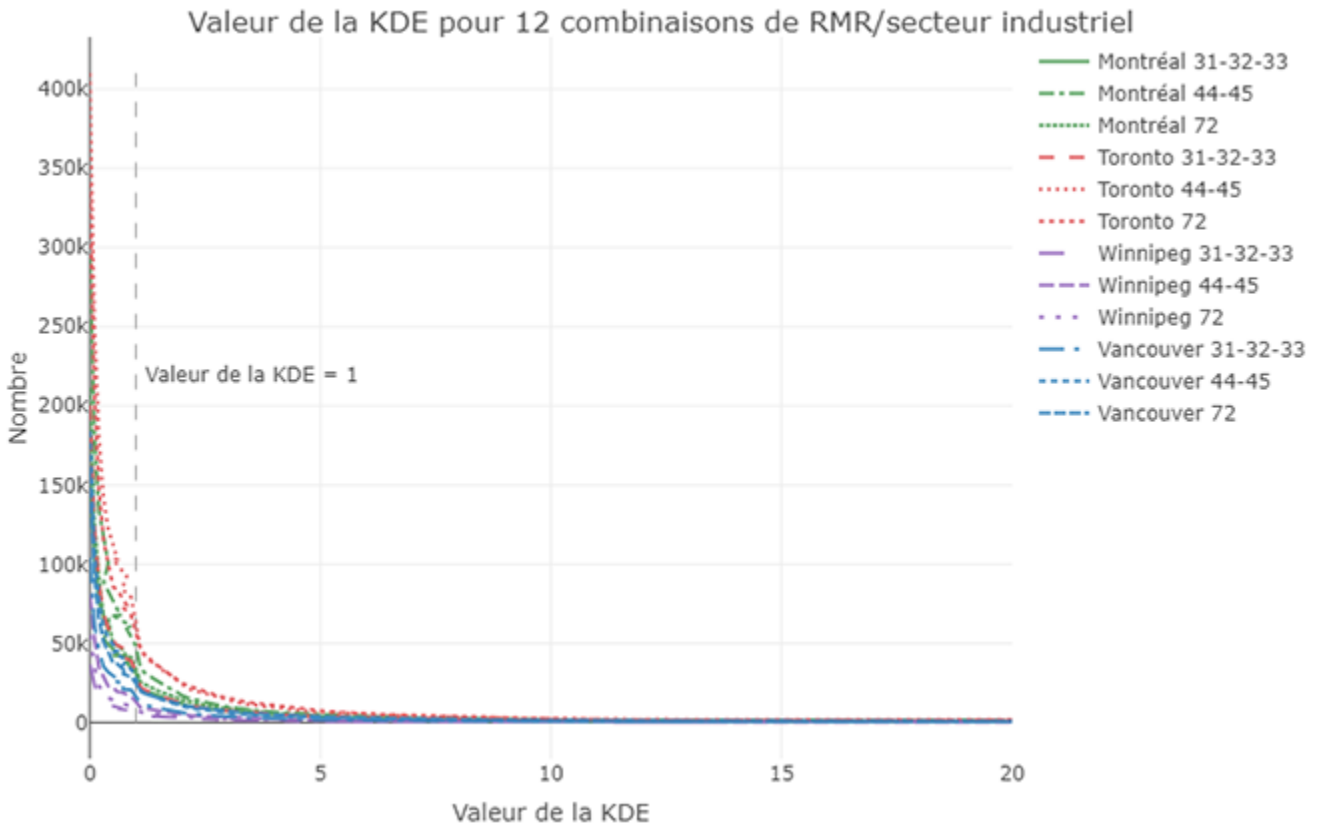
Source: calculs des auteurs basés sur une justification textuelle fournie dans Biemer et Stokes, 1984.

Identification des seuils de densité du noyau

Passons maintenant à la partie restante de la méthodologie principale de cet article. À l'aide de la fonction de densité polynomiale du noyau présentée ci-dessus, une valeur de densité pour chaque identifiant de cellule Φ de la grille est estimée, ce qui permet la génération d'une distribution de densité du noyau approximativement continue, comme le montre la Figure 6 pour trois grappes différentes dans quatre RMR. À la Figure 6, les valeurs de KDE s'affichent sur l'axe horizontal et le nombre de fréquences exprimé en milliers (nombre non pondéré de cellules issu de la grille complète de la RMR) sur l'axe vertical. Plus la valeur de KDE est élevée, plus la densité de la cellule est grande (c'est-à-dire plus le nombre de lieux des emplois disponibles dans le quartier de la cellule est important). Comme prévu, la distribution de la densité de KDE est fortement asymétrique. Il est intéressant de noter que la distribution empirique de la densité du noyau ci-dessous est de forme similaire à la fonction théorique de la dérivée première de la dynamique entre la densité spatiale et la distance spatiale entre les points, décrite dans la page précédente ci-dessus (Figure 5). Pour toutes les RMR et grappes, le nombre de cellules diminue considérablement lorsque la valeur de KDE dépasse 1 (ligne verticale en pointillés sur la Figure 6). Des baisses

importantes sont également observées pour des valeurs de KDE encore plus faibles, mais nous n'en tenons pas compte, car les valeurs de KDE se rapprochent d'une densité nulle. Étant donné le modèle de la Figure 6, un seuil de 1 a été utilisé comme étape de prétraitement pour filtrer le segment inutile de la distribution des cellules de la grille. Cependant, l'analyse a été affinée avec un deuxième ensemble de seuils comme l'indique la section ci-après.

Figure 6
Distributions des valeurs de KDE dans différentes RMR et grappes industrielles



Source : calculs des auteurs à partir de la base de données sur le RE.

À l'aide des distributions de densité présentées à la Figure 6, un ensemble de valeurs seuil de KDE a été mis à l'essai pour chaque combinaison de RMR et de secteur industriel en calculant le nombre d'employés et d'établissements des ID maintenus dans les grappes à chaque niveau de seuil de KDE. L'objectif était d'augmenter de façon itérative la valeur seuil et d'éliminer les ID de faible densité, puis de s'arrêter avant qu'une baisse significative de l'emploi total au sein de la grappe ne soit observée. En fin de compte, les valeurs seuil de KDE qui ont conservé un minimum d'environ 80 % du nombre total d'employés des RMR ont été appliquées pour la plupart des combinaisons emplacement/secteur, ce qui permet d'éliminer de nombreux ID peuplés de petites entreprises. Les valeurs seuil de KDE, principalement comprises entre 1 et 3 sont présentées dans le tableau 2 pour chaque combinaison de RMR et de grappe.

Comme nous pouvons le voir dans le tableau 2, une grande proportion de seuils préserve la valeur initiale de 1 (Figure 6) et n'a pas besoin d'un ajustement. Cela est particulièrement vrai pour les grappes industrielles définies par Delgado et coll. (2014) à une exception près. Cela indique que le seuil initial est utile et robuste. Ce résultat suggère également que la méthode KDE utilisée dans cette recherche parvient à saisir la notion de quartier des petites entreprises au Canada, sans avoir été alimentée au préalable par un système formel de classification en trois catégories, à savoir petites, moyennes et grandes entreprises. En effet, la distribution du RE des points de données spatiales X est continue et non multinomiale. Cependant, 7 des 24 seuils ont été fixés à une valeur seuil de KDE plus élevée, soit 3. Une analyse plus approfondie de ces grappes suggère que les valeurs plus élevées

étaient mieux adaptées pour représenter les grappes dans les RMR ayant une proportion plus élevée de petites entreprises, ce qui conduit à une distribution spatiale de l'emploi plus dispersée. En revanche, la valeur seuil pour le secteur de l'hôtellerie et du tourisme à Winnipeg a été fixée à 0,1, ce qui indique une très faible concentration de petites entreprises dans cette RMR (tableau 2). Enfin, il convient de noter que le paramétrage initial des seuils (autour de la valeur 1) est purement basé sur des statistiques et repose uniquement sur la forme des distributions statistiques de la KDE, tandis que le deuxième ensemble de seuils (tableau 2) est guidé par des considérations économiques et la possibilité de générer des indicateurs économiques significatifs à partir des limites des grappes. Par conséquent, ce paramétrage composite des seuils bénéficie d'une perspective multidimensionnelle et est plus susceptible d'être stable dans le temps et l'espace. Les trois sections ci-après décrivent les étapes restantes de la méthodologie.

Tableau 2
Seuils de KDE appliqués aux combinaisons RMR/grappes industrielles

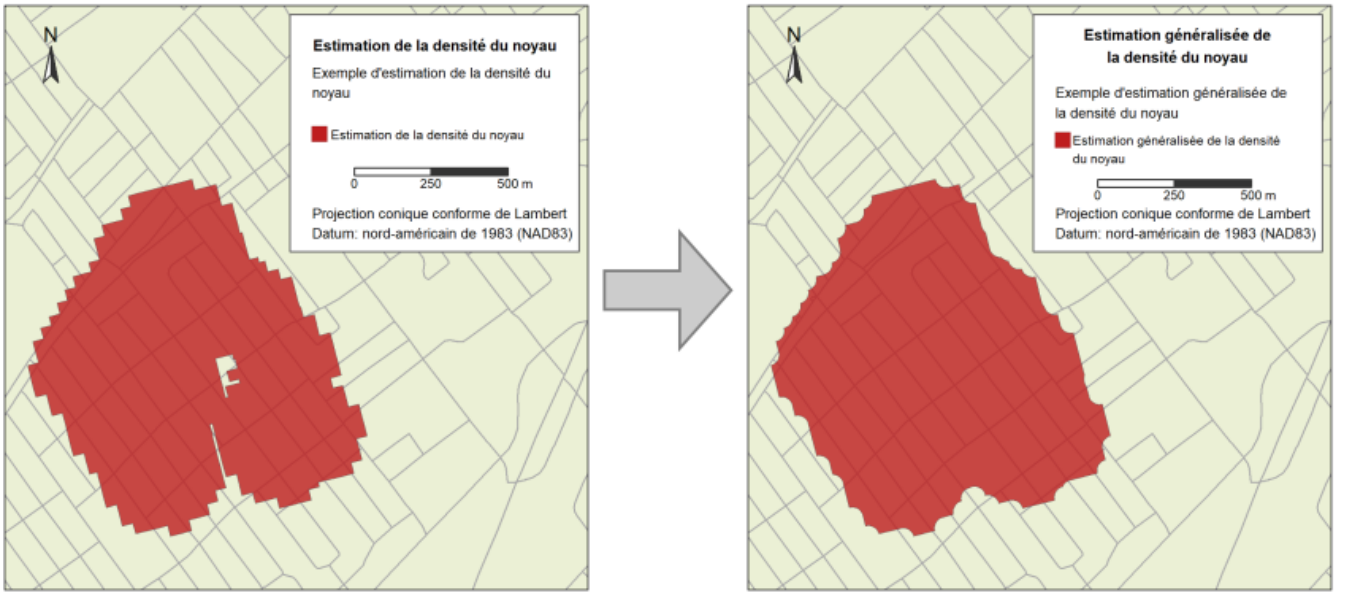
	Valeur seuil de la KDE			
Grappe industrielle	**Montreal**	**Toronto**	**Winnipeg**	**Vancouver**
Secteur de la fabrication	3	1	1	1
Secteur du commerce de détail	3	3	1	3
Secteur des services d'hébergement et de restauration	3	3	1	3
Distribution et commerce électronique (grappe 10)	1	1	1	1
Services financiers (grappe 16)	1	1	1	1
Hôtellerie et tourisme (grappe 22)	1	1	0.1	1

Source : calculs des auteurs à partir de la base de données sur le RE.

Généralisation de la KDE

Par définition, les résultats de la KDE identifient les zones au sein d'une RMR qui contiennent la majorité d'un type d'industrie particulier. Bien que les résultats filtrés fournissent une indication des endroits où un secteur précis est dominant dans une zone d'étude, ils ne sont pas uniformes, et de petites lacunes ou de petits trous peuvent apparaître dans les zones à haute densité identifiées. Pour lisser ces résultats, une étape de généralisation est appliquée aux sorties filtrées de la KDE.

Carte 1
Changements après l'étape de généralisation



Sources : Statistique Canada, Recensement de 2021 - Îlots de diffusion fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites et calculs des auteurs.

L'étape de généralisation commence par l'union de toutes les cellules de la grille résultant de la KDE en grandes entités polygonales. Une fois que les cellules des résultats de la KDE sont combinées, les polygones sont

généralisés en appliquant une technique consistant à créer un tampon de 50 mètres autour des résultats de l'union, puis à réduire ce tampon de la même valeur. Ce processus supprime les petites lacunes et les petits trous dans les polygones, produisant ainsi un résultat final plus propre pour la fusion. Un exemple de ce processus est présenté dans la Carte 1 ci-dessus.

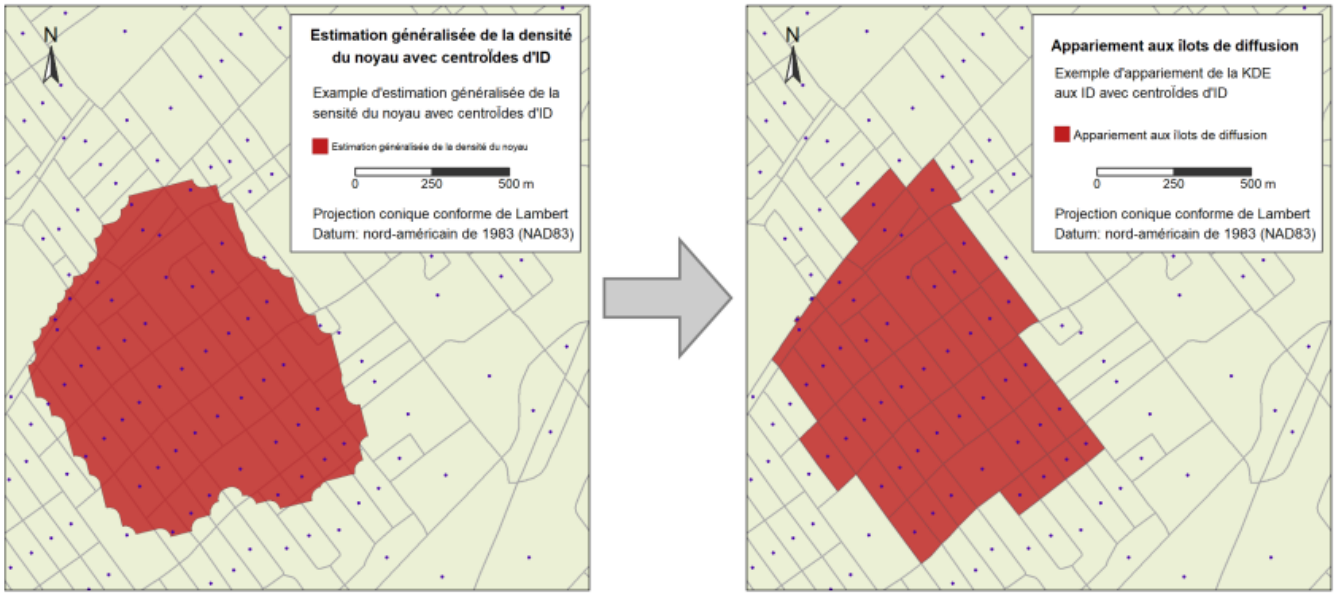
Appariement des résultats de la KDE et des ID

Les résultats obtenus par la KDE et la sélection des seuils de densité ont permis d'indiquer où se trouvent les concentrations de grappes industrielles au sein de chaque RMR. Cependant, ces résultats ont été générés au niveau des cellules de sortie de la grille et n'ont donc pas été associés aux fichiers de limites établis utilisés par Statistique Canada. En réassociant les résultats aux limites des ID, une analyse plus complète des grappes industrielles peut être réalisée, allant au-delà de la simple identification de la présence ou de l'absence d'un secteur.

L'appariement des résultats de la KDE a été réalisé en recoupant les centroïdes des ID d'une RMR avec les résultats de la KDE généralisée. Tous les ID dont les centroïdes recoupaient les résultats de la KDE ont été conservés pour représenter les grappes industrielles. En d'autres mots, la cellule de grille qui se superpose au centroïde de l'ID définit la densité représentative de cet ID. Un exemple du résultat de ce processus (Carte 2) illustre les polygones des ID associés à une grappe industrielle par l'intersection de centroïdes.

Il convient de noter qu'un centroïde pondéré des ID basé sur l'emplacement spatial de l'établissement du RE au sein de l'ID aurait pu être utilisé à la place d'un centroïde géométrique standard. C'est le cas, par exemple, pour la base de données ouverte sur les mesures spatiales de l'accès conçue au Laboratoire d'exploration et d'intégration des données (LEID) du Centre des projets spéciaux sur les entreprises de Statistique Canada. Cependant, la méthodologie utilisée pour la présente analyse repose sur un processus de randomisation uniforme pour les points de données sur les lieux des emplois au sein de l'ID, ce qui écarte la pertinence de préserver les emplacements des établissements du RE dans l'ID et valide l'utilisation d'un simple centroïde géométrique. Comme on l'explique plus tôt, l'espérance E(RP) du processus RP est une allocation spatialement uniforme des points de données au sein de l'ID où chaque point spatial reçoit le même nombre de points de données. Par conséquent, il n'est pas pertinent de privilégier un centroïde pondéré là où les lieux des emplois sont susceptibles de se trouver. En effet, même si la précision des quartiers est importante, elle doit également être partiellement compromise, car l'objectif reste de lisser la distribution spatiale discrète des lieux des emplois et de faciliter la contribution apportée par l'estimation de la densité du noyau.

Carte 2
Résultats de la généralisation convertis en résultats appariés par un processus d'intersection des centroïdes



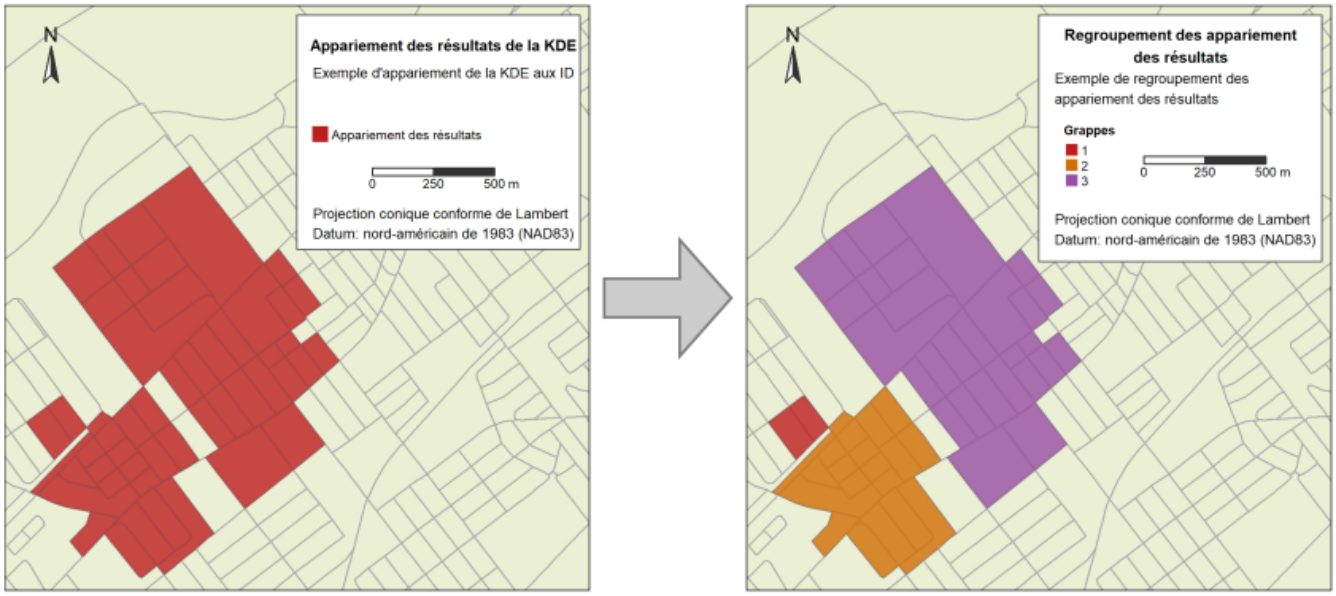
Sources : Statistique Canada, Recensement de 2021 - Îlots de diffusion fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites et calculs des auteurs.

Regroupement et filtrage des résultats appariés

Les résultats appariés aux limites spatiales des ID illustrent quels ID sont associés à un secteur. Cependant, en raison de la distribution spatiale de certains secteurs et des formes irrégulières des polygones des ID, les résultats de cet appariement peut générer de petits regroupements d'ID qui peuvent être associés à seulement quelques établissements industriels. Pour garantir que les résultats finaux se concentrent sur les principales concentrations d'entreprises et préservent la confidentialité dans une analyse plus approfondie, les extrants du processus d'appariement ont été regroupés en grappes de polygones d'ID connectés et des statistiques sommaires ont été calculées pour chaque polygone.

Le regroupement des polygones d'ID appariés a été réalisé en combinant tous les ID ayant une arête commune. Les polygones qui ne se touchent qu'à un coin n'ont pas été considérés comme faisant partie de la grappe, comme le montre l'exemple de regroupement de la Carte 3. Cette règle a été appliquée pour limiter la création de vastes grappes industrielles qui sont dispersées dans une RMR. Une fois les grappes d'ID identifiées, le nombre d'employés, d'établissements et de polygones des ID a été comptabilisé pour chaque grappe afin d'évaluer si celuici devait être conservé.

Carte 3
Exemple de résultats appariés de grappes industrielles divisées en grappes de polygones d'ID ayant une arête commune



Sources : Statistique Canada, Recensement de 2021 - Îlots de diffusion fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites et calculs des auteurs.

En utilisant les résultats des grappes tabulées à l'étape précédente, les grappes présentant un risque de confidentialité en raison de leur domination par un seul établissement ont été éliminées. Cela a été réalisé en supprimant toutes les grappes comptant trop peu d'établissements (moins de 5) ou où un seul établissement embauchait plus de 80 % des employés de la grappe. Ce processus souligne l'idée que, même isolée des plus grandes grappes de sa RMR, une petite grappe n'est pas automatiquement perçue comme un risque de confidentialité, à condition que les deux critères de confidentialité soient remplis.

Récapitulation des étapes du point de vue d'une cellule de grille $\Phi(z)$

Avant de passer à la section suivante de cet article, et pour rester cohérent avec notre notation, nous récapitulons les quelques étapes récentes du point de vue d'une cellule de grille d'intérêt $\Phi(z)$. $\Phi(z)$ est initialement apparié à une valeur numérique de densité estimée $u(z)$ incluse dans les nombres réels non négatifs R^+ (Figure 6). Si cette valeur $u(z)$ est inférieure à son seuil de RMR et de secteur d'activité (soit une valeur seuil de 1 ou de 3 selon le tableau 2), alors $\Phi(z)$ est exclu du reste de la méthodologie. Si $u(z)$ est égal ou supérieur à son seuil respectif, alors $\Phi(z)$ est conservé pour l'étape suivante. De plus, à ce stade, la représentation R^+ de $u(z)$ n'a plus d'importance, et $u(z)$ prend une valeur arbitraire u égale à la même valeur arbitraire de toutes les autres cellules de grille $\Phi()$ incluses dans le processus de la RMR jusqu'à présent. Pour l'étape de généralisation, si $\Phi(z)$ de valeur u est inclus dans le processus jusqu'à présent, alors la généralisation préserve également l'existence de $\Phi(z)$ et ne peut pas l'exclure. En d'autres termes, l'étape de mise en mémoire tampon et de désambiguïsation lisse les frontières du cluster, mais ne peut pas supprimer la cellule existante d'une frontière. Cependant, les cellules de grille qui ont été exclues jusqu'à présent peuvent maintenant être incluses, car l'étape de généralisation consiste à supprimer les petits espaces et les trous dans la forme du cluster. Pour l'amalgame avec les IDs, $\Phi(z)$ chevauche ou non le centroïde de son ID respectif. S'il ne se chevauche pas, alors $\Phi(z)$ n'a plus de raison d'être dans le processus. En cas de chevauchement, l'objectif de $\Phi(z)$ est d'accepter son ID respectif dans la représentation au niveau ID de la cartographie thermique du cluster industriel. Notez que nous notons maintenant l'ID de $\Phi(z)$ comme $ID(z)$ parce que cette explication est du point de vue de $\Phi(z)$ et que l'ID de $\Phi(z)$ est inclus dans le processus jusqu'à présent en raison de l'existence de $\Phi(z)$. À ce stade, le rôle de la cellule de grille $\Phi(z)$ est terminé, et le reste des décisions sont basées sur des clusters, et non sur des cellules ou des IDs. Pour l'étape finale de filtrage, la double condition de confidentialité préservera ou supprimera le cluster industriel où $ID(z)$ et la cellule de grille $\Phi(z)$ sont jusqu'à présent inclus. Si le cluster est supprimé, la contribution de la cellule de grille

$\Phi(z)$, quelle que soit sa signification dans les étapes précédentes, devient désormais nulle car $ID(z)$ ne fait plus partie d'un cluster. Si le cluster est conservé, alors la cellule de grille $\Phi(z)$ préserve sa contribution car $\Phi(z)$ est la cellule dont le rôle était d'ajouter $ID(z)$ dans une partie de cluster de la carte thermique finale.

Résumé non technique de la méthodologie

Cette section sur la méthodologie se termine par un résumé non technique des étapes de la génération des grappes industrielles. Nous décomposons la méthode en quatre étapes principales : 1) données 2) noyau 3) seuils et 4) traitement post-noyau.

Données :

- Extraire les données à l'échelle de l'établissement de la base de données interne du RE de Statistique Canada.
- Définir le lieu de chaque emploi unique faisant partie d'un établissement (répartir spatialement de manière aléatoire tous les emplois d'un ID à l'intérieur de ses propres limites si l'établissement se trouve dans l'ID).
- Définir le poids de chaque emploi (dans notre cas, un poids de 1 pour chaque emploi).

Noyau :

- Définir la forme et la dimension des cellules de la grille pour le noyau.
- Définir la forme fonctionnelle du noyau.
- Définir la longueur de bande passante pour le noyau.
- Définir le centroïde géométrique de chaque cellule de la grille dans la RMR.
- Calculer une valeur de densité unique pour chaque cellule de la grille dans la RMR en utilisant la distance entre le centroïde géométrique de la cellule et le lieu du travail.

Seuils :

- Calculer les seuils en fonction de la distribution statistique de la KDE pour chaque RMR et secteur.
- Calculer les seuils en fonction de la méthode du ratio de rétention pour chaque RMR et secteur.
- Privilégier les seuils en fonction de la méthode du ratio de rétention s'ils diffèrent des seuils basés sur la distribution statistique de la KDE.

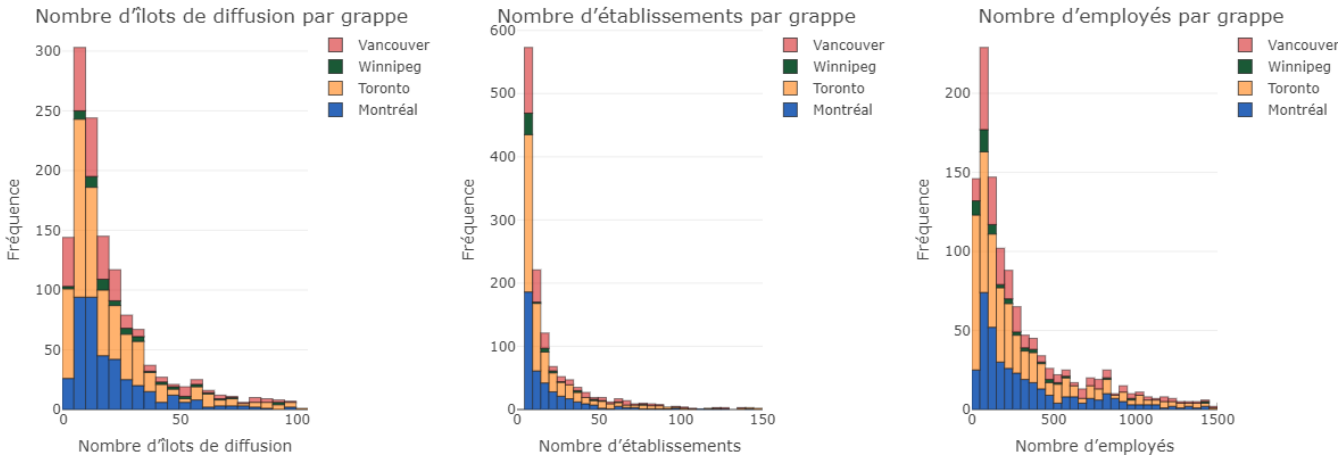
Traitement post-noyau :

- Généraliser la KDE en lissant les limites des grappes et en comblant leurs lacunes internes.
- Apparier les résultats de la KDE avec les ID pour obtenir une représentation finale des grappes à l'échelle des ID.
- Filtrer les grappes si elles ne répondent pas à la double condition de confidentialité.

Résultats

La Figure 7 présente un aperçu des statistiques descriptives pour les résultats de la grappe. Le nombre d'établissements des ID et d'employés par grappe est disponible pour chacune des quatre zones d'étude de la RMR. L'axe des ordonnées (x) représente le nombre d'établissements des ID et d'employés, respectivement. L'axe des abscisses (y) correspond à la fréquence ou au nombre de grappes distinctes. Les distributions statistiques présentent une forme similaire à celle de la distribution de la méthode KDE fortement asymétrique vers la droite de la Figure 6. Autrement dit, elles sont semblables à une distribution suivant une loi de puissance, ce qui concorde avec l'étude de Gabaix (1999) selon laquelle la taille des villes suit une loi de puissance. De plus, pour les trois histogrammes de la Figure 7, la fréquence est plus modérée pour Montréal (histogrammes surlignés en bleu), qui se trouve être la RMR ayant les superficies d'ID les plus petites et avec une portée réduite (Figure 1) et l'une des distributions de la méthode KDE les plus faibles (Figure 6) pour le secteur de la fabrication (codes du SCIAN 31, 32 et 33). Sur la base de ces observations empiriques, le nombre de grappes par RMR semble être corrélé à la structure de la population des ID de la RMR. Il s'agit d'un résultat intuitif. C'est-à-dire que les RMR avec une forte proportion de grands ID tendent à générer des segments plus dispersés de cellules de sortie de la grille à haute densité, ce qui produit par conséquent un plus grand nombre de grappes spatiales distinctes. À l'inverse, les RMR largement dominées par de petits ID carrés auraient tendance à regrouper plusieurs grappes en une seule grande grappe lors du processus de généralisation, de mise en tampon et suppression de ce processus, ainsi que d'appariement, en raison de la proximité des ID les uns avec les autres. Cela soulève des questions de recherche intéressantes, et nous prenons le temps de décrire deux d'entre elles ici. 1) Est-ce que la configuration des ID d'une RMR peut expliquer une proportion importante de la population des grappes? 2) Y a-t-il d'autres facteurs plus importants qui influent sur la façon dont les grappes émergent et prennent forme à partir des données spatiales?

Figure 7
Distribution du nombre d'ID, d'établissements et d'employés pour chaque zone d'étude de la RMR



Source : .

Les principaux résultats des grappes spatiales pour Montréal, Toronto, Winnipeg et Vancouver, ainsi que six spécifications de grappes industrielles, sont résumés dans le tableau 3, tandis que la cartographie de chaque grappe est présentée en annexe. Dans chaque carte, les grappes du secteur industriel de référence sont surlignées en rouge.

Comme prévu, la distribution spatiale des grappes varie considérablement d'une industrie à l'autre, les secteurs du commerce de détail et des services d'hébergement et de restauration couvrant généralement les plus grandes proportions d'ID dans chaque région métropolitaine. Cela est représenté également dans le nombre d'ID appartenant à chaque grappe (tableau 3). Pour toutes les RMR, les secteurs du commerce de détail et des services d'hébergement et de restauration représentent le plus grand nombre d'ID. De même, le ratio de rétention des ID, c'est-à-dire le pourcentage d'ID qui ont été inclus dans la grappe correspondante dans chaque RMR, varie de 36 % à 55 % et est généralement plus élevé (autour ou au-dessus de 40 %) pour les secteurs du commerce de détail et des services d'hébergement et de restauration (tableau 3).

de 36 % à 55 % et est généralement plus élevé (autour ou au-dessus de 40 %) pour les secteurs du commerce de détail et des services d'hébergement et de restauration (tableau 3).

Les grappes englobent la plupart des établissements et, plus important encore, la majorité des emplois dans les secteurs correspondants (tableau 3). Comme il a été décrit dans les sections précédentes, il s'agissait d'un critère clé pour déterminer le seuil de valeur de densité pour l'inclusion des ID dans la grappe. Le ratio de rétention des établissements représente la part des établissements au sein des ID de la grappe industrielle par rapport au nombre total d'établissements dans ce secteur dans la RMR. De même, le ratio du personnel représente la part de l'emploi généré par les entreprises situées dans les zones de la grappe industrielle par rapport à l'emploi total de ce secteur dans la RMR.

Tableau 3
Résultats : Statistiques sommaires des grappes, 2023

	ID	Établissement	ID	Emploi	Établissement
RMR Grappe	nombre		ratio de rétention		
Montréal					
Secteur de la fabrication	3 775	4 136	26,1	89,7	70,9
Secteur du commerce de détail	9 861	10 431	38,9	92,0	76,7
Secteur des services d'hébergement et de restauration	7 888	6 806	36,5	85,4	77,6
Distribution et commerce électronique (grappe 10)	5 660	5 406	34,4	94,9	74,6
Services financiers (grappe 16)	2 902	1 512	26,2	88,5	55,4
Hôtellerie et tourisme (grappe 22)	2 472	860	23,6	60,3	47,1
Toronto					
Secteur de la fabrication	5 732	7 594	34,8	94,4	81,6
Secteur du commerce de détail	9 716	15 075	35,5	97,1	76,0
Secteur des services d'hébergement et de restauration	9 946	10 085	38,5	89,2	77,2
Distribution et commerce électronique (grappe 10)	7 045	10 291	36,2	97,8	81,9
Services financiers (grappe 16)	5 972	4 856	35,6	95,7	72,0
Hôtellerie et tourisme (grappe 22)	2 842	1 329	25,3	71,2	50,9
Winnipeg					
Secteur de la fabrication	1 054	616	44,2	91,9	78,5
Secteur du commerce de détail	2 700	2 130	55,3	94,6	87,2
Secteur des services d'hébergement et de restauration	2 217	1 313	54,9	91,2	87,7
Distribution et commerce électronique (grappe 10)	855	850	33,1	93,2	73,5
Services financiers (grappe 16)	666	349	28,4	82,7	55,2
Hôtellerie et tourisme (grappe 22)	1 363	337	60,3	81,3	75,2
Vancouver					
Secteur de la fabrication	2 432	3 236	35,9	90,8	79,1
Secteur du commerce de détail	3 981	7 260	36,7	81,3	77,5
Secteur des services d'hébergement et de restauration	4 686	5 624	43,7	91,6	83,0
Distribution et commerce électronique (grappe 10)	2 662	4 610	32,0	93,7	77,1
Services financiers (grappe 16)	2 166	2 130	32,3	91,2	68,2
Hôtellerie et tourisme (grappe 22)	1 709	1 041	31,1	79,8	60,1

Source : calculs des auteurs à partir de la base de données sur le RE.

À l'exception du secteur de l'hôtellerie et du tourisme (grappe 22), toutes les autres grappes industrielles capturent bien plus de 80 % de l'emploi de ce secteur au sein de leurs secteurs respectifs dans les RMR de référence, certaines grappes atteignant même 95 % ou plus des emplois (tableau 3). Par exemple, la grappe du secteur de la fabrication regroupe 89,7 % de l'emploi total du secteur à Montréal, 94,4 % à Toronto, 91,9 % à Winnipeg et 90,8 % à Vancouver, par rapport à l'emploi total du secteur de la fabrication dans les RMR respectives.

Des pourcentages similaires sont calculés pour le nombre d'entreprises situées dans les zones des grappes. Bien que ces pourcentages soient légèrement inférieurs à ceux pour l'emploi, ils restent autour de 80 % pour la plupart des grappes, à l'exception des secteurs de l'hôtellerie et du tourisme (grappe 22) et des services financiers (grappe 16). Par exemple, les grappes du secteur de la fabrication contiennent 70,9 % à Montréal, 81,6 % à Toronto, 78,5 % à Winnipeg et 79,1 % à Vancouver du total des établissements manufacturiers dans les RMR respectives. Cette répartition des entreprises au sein de la grappe par rapport à celles opérant hors de celle-ci pourrait également être utilisée pour surveiller les tendances de l'évolution des grappes métropolitaines.

Il convient de noter que pour certaines grappes, le nombre d'établissements est beaucoup plus faible que le nombre d'ID qui constituent la grappe. Par exemple, la grappe Hôtellerie et tourisme (grappe 22) à Montréal

comprend 860 établissements et 2 472 ID. Ce résultat est dû aux méthodes de mise en tampon et d'appariement utilisées dans l'analyse ainsi qu'à la concentration des entreprises dans des zones comprenant de petits ID. L'approche méthodologique mise au point dans la présente analyse est conçue pour fournir une représentation à l'échelle des quartiers par opposition à une représentation des ID individuels. Ainsi, les entreprises situées dans les ID proches les uns des autres, mais toujours séparées par d'autres ID, sont regroupées, y compris les ID qui sont entourés par de telles entreprises, mais qui n'en contiennent aucune à l'intérieur de leurs limites. La situation inverse est également possible. C'est-à-dire que l'ID comprend les lieux des établissements et des emplois, mais celui-ci ne fait pas partie de la grappe finale. Cela peut se produire si la cellule de la grille se chevauchant avec notre centroïde de l'ID n'est pas assez dense et est filtrée lors du processus de seuil en deux étapes expliqué précédemment. De tels centroïdes d'ID se trouvent généralement à la limite d'une grande grappe et sont rejetés par le processus en raison de la bande passante qui agrège une trop grande proportion d'ID sans lieu d'établissement. Une analogie rapide serait d'ajuster un modèle de régression non linéaire sur un ensemble de points de données. Le modèle ajusté se trouve parfois au-dessus ou en dessous des points de données réels, mais dans l'ensemble, il contribue à fournir une approximation adéquate et continue des phénomènes de données et une représentation anonymisée de la distribution spatiale réelle des lieux des établissements et des emplois. La courbe ajustée permet à l'analyste d'effectuer des recherches sur les tendances et les modèles sans observer directement les données confidentielles.

Il convient de noter qu'un ratio de rétention plus élevé (plus proche de 100 %) pour une RMR n'équivaut pas à un niveau de performance économique plus élevé de la RMR par rapport aux autres. Les ratios de rétention sont plutôt plus proches, du point de vue de leur signification, de la distribution de la taille des entreprises, où des ratios plus faibles impliquent une plus grande proportion de petites entreprises au sein de la RMR. La distribution de la taille des entreprises est accessible au public et n'est pas une information exclusive fournie par cette recherche, mais demeure une source de validation pour les résultats de cette recherche.

Les résultats obtenus avec la cartographie des grappes d'entreprises ont été validés visuellement en les comparant avec les cartes de zonage des terrains des municipalités incluses dans la RMR de la présente analyse. Cette validation était particulièrement évidente pour les grappes du secteur de la fabrication et les zones industrielles ou les parcs industriels. Par exemple, pour la municipalité de Toronto, le zonage industriel des parcelles de terrain correspond étroitement à l'emplacement des grappes du secteur de la fabrication. L'utilisation de données provenant de sources tierces, telles que OSM ou Google Maps, n'a pas été incluse dans la présente. Toutefois, elle pourrait être envisagée pour de futurs efforts de validation.

Voici quelques précisions supplémentaires pour comprendre ce qu'est le ratio de rétention. Le seuil de rétention repose sur un principe simple des gains marginaux. C'est-à-dire que nous considérons qu'il est pertinent de continuer à exclure les ID si cette exclusion est suffisamment supérieure à celle des emplois. En d'autres termes, la clarté visuelle des cartes thermiques des grappes est importante, mais seulement tant qu'elle ne compromet pas la performance économique. Pour la grappe 22 à Montréal, le ratio de rétention est de 60 % des emplois. Par conséquent, choisir un seuil de 80 % n'était pas pertinent, car une proportion importante d'ID continue de baisser au détriment d'une proportion suffisamment faible des lieux des emplois. Cependant, dépasser les 60 % d'emplois entraînerait non seulement une baisse des emplois nettement plus importante que celle des ID, mais atteindrait également un point où la diminution marginale des emplois deviendrait exponentielle, par exemple de 60 % à 30 %, ce qui serait inacceptable. Cette baisse exponentielle est intuitive et explicable compte tenu de la nature même de nos distributions statistiques du RE. Comme l'indiquent les sections précédentes du présent rapport, les données du RE sont fortement asymétriques avec des formes liées à une loi de puissance. Par conséquent, le segment non plat de la distribution comprend un segment chaotique où une très petite variation de l'information conduit à une baisse très importante. Les grappes définies sur une proportion relativement plus faible des établissements sont susceptibles d'être plus petites en taille de grappes et en nombre. En examinant la grappe 22 pour Montréal et Toronto dans l'annexe du présent rapport, on constate effectivement que les grappes sont plus petites, plus dispersées et moins nombreuses.

Tableau 4
Pourcentage de colocalisation entre deux types de grappes industrielles pour chaque zone d'étude de la RMR, 2023

	Pourcentage de colocalisation parmi les grappes industrielles		
	Secteur de la fabrication	Commerce de détail	Services d'hébergement et de restauration
RMR Grappe	pourcentage		
Montréal			
Secteur de la fabrication	-	37,3	20,5
Commerce de détail	30,6	-	43,6
Services d'hébergement et de restauration	24,6	63,6	-
Toronto			
Secteur de la fabrication	-	50,1	37,0
Commerce de détail	47,5	-	52,7
Services d'hébergement et de restauration	39,9	59,9	-
Winnipeg			
Secteur de la fabrication	-	47,5	37,4
Commerce de détail	24,6	-	46,0
Services d'hébergement et de restauration	25,4	60,4	-
Vancouver			
Secteur de la fabrication	-	39,5	30,8
Commerce de détail	46,6	-	61,1
Services d'hébergement et de restauration	30,3	51,0	-

Source : calculs des auteurs à partir de la base de données sur le RE.

Enfin, une analyse simple de colocalisation des grappes d'entreprises a été réalisée en superposant les fichiers des limites de deux grappes et en calculant le pourcentage de superficie d'une grappe (ligne) qui est également inclus dans une autre grappe (colonne). Le tableau 4 présente les résultats d'une évaluation de la colocalisation des secteurs de la fabrication, du commerce de détail, et des services d'hébergement et de restauration. Une visualisation de cette colocalisation entre deux ensembles de grappes est également fournie à la Carte 4.

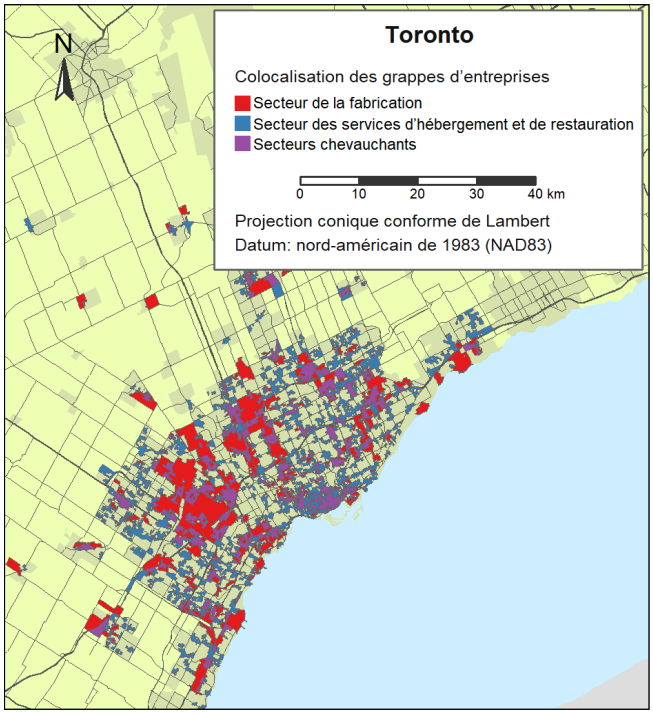
Deux éléments clés se dégagent de ces exemples. Premièrement, le pourcentage de colocalisation varie entre environ 20 % et 60 % de la superficie des grappes, selon la grappe en question. Cependant, la force de la colocalisation entre les grappes d'entreprises fournit des renseignements économiques significatifs. Plus précisément, dans toutes les RMR analysées, le chevauchement entre le commerce de détail et les services d'hébergement et de restauration est plus prononcé par rapport à ces deux secteurs et celui de la fabrication. Cette observation s'harmonise avec la perception générale selon laquelle ces types de services tendent à se regrouper dans les mêmes zones géographiques au sein d'une région métropolitaine.

Deuxièmement, le type de colocalisation offre des indications sur les différences potentielles qu'une même grappe industrielle peut présenter au sein de la région métropolitaine. Par exemple, la Carte 4 montre que les secteurs des services d'hébergement et de restauration sont concentrés dans des quartiers qui, d'une part, présentent également une forte concentration du commerce de détail et, d'autre part, des quartiers qui indiquent également une forte concertation du secteur de la fabrication. Il est probable que ces zones de chevauchement représentent des activités d'hébergement et de restauration s'adressant à des clientèles différentes et présentant des types variés de liens ou de dépendances sectoriels.

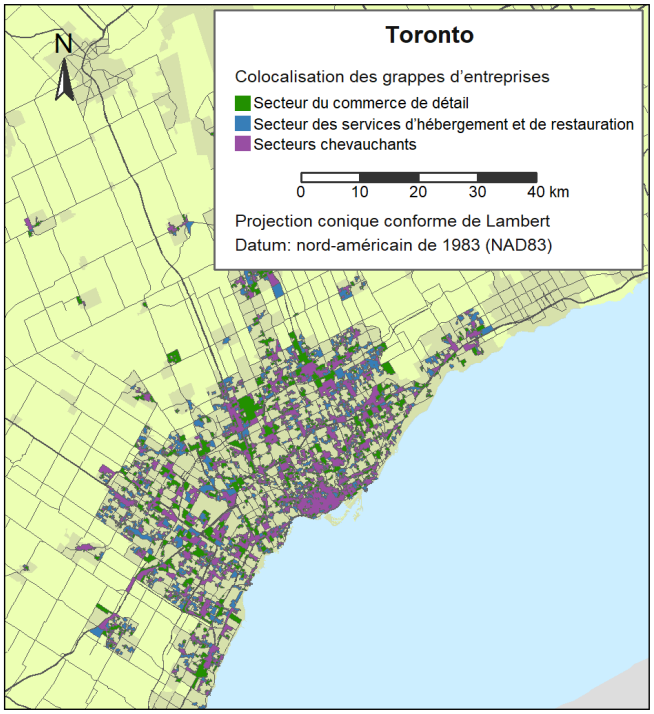
Enfin, il convient de noter que la concentration de certains types d'entreprises, comme celles du secteur de la fabrication, dans des zones situées à l'extérieur de ce qui semble être leurs zones municipales désignées, comme les zones industrielles, pourrait indiquer une concentration de fonctions opérationnelles particulières dans les zones commerciales. Par exemple, le regroupement d'entreprises manufacturières au centre-ville de Toronto, dans des zones désignées comme commerciales, suggère que les entreprises manufacturières de cette région pourraient être liées à des fonctions de sièges sociaux ou de bureaux, plutôt qu'à des établissements de production.

Carte 4
Exemple illustratif de colocalisation de grappes à Toronto

Panneau A : Secteur de la fabrication et secteur des services d'hébergement et de restauration



Panneau B : Secteur du commerce de détail et secteur des services d'hébergement et de restauration



Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Axes de recherche, d'analyse et d'applications futures

La méthodologie présentée dans cet article peut être mise au point et appliquée davantage pour générer des données sur les conditions des entreprises à l'échelle locale. Le développement le plus immédiat consiste à étendre le travail à toutes les régions métropolitaines du Canada. Une extension supplémentaire aux agglomérations de taille moyenne (agglomérations de recensement), ainsi qu'aux zones rurales et aux petites villes devrait également être envisagée.

L'augmentation de la couverture géographique peut se faire parallèlement à l'amélioration des regroupements du SCIAN qui définissent chaque grappe. Pour mettre au point la méthodologie, cet article a mis l'accent sur les codes du SCIAN simples à deux chiffres ou les regroupements préexistants du SCIAN, en se référant spécialement aux travaux de Delgado et coll. (2014) sur les grappes industrielles aux États-Unis. Bien qu'il soit pertinent de poursuivre cette approche, l'utilisation des microdonnées du RE permet de mettre en œuvre d'autres agrégations des codes du SCIAN à différents niveaux de chiffres. Des regroupements personnalisés pourraient être envisagés. Par exemple, une partie de la littérature existante s'est concentrée sur des grappes artistiques et culturelles et la vitalité des quartiers. Par conséquent, des définitions précises personnalisées de ces types de grappes pourraient être envisagées pour la mise en œuvre.

Une analyse plus approfondie devrait envisager d'établir le profil de la performance des grappes d'entreprises à l'aide de variables supplémentaires du RE. En plus du code à six chiffres du SCIAN, du nombre d'employés et de la position de latitude et de longitude pour chaque établissement commercial, le RE fournit plusieurs champs d'intérêt supplémentaires qui pourraient être intégrés dans l'analyse pour établir le profil et indexer la performance de l'entreprise. En particulier l'utilisation des revenus, des dépenses totales, des actifs totaux et de la date de création des entreprises pourrait être explorée pour générer des indices spatiaux agrégés. Il serait

également possible d'élaborer des ratios d'indicateurs de performance financière à partir des données du RE. Il en existe plusieurs : 1) ratio de rentabilité [(total des revenus moins totaux des dépenses) / total des actifs] 2) ratio de liquidité (actifs totaux courants/dettes totales courantes) 3) ratio de tangibilité (total des immobilisations/ total des actifs) 4) taux de croissance des ventes totales. Il est important de souligner que les données du RE ne sont pas sans difficultés techniques. Le niveau extrêmement précis de détail, à la fois en termes de géographie et de grappe industrielle, limiterait le type d'information qui pourrait être extraite. Cependant, l'utilisation d'indices, comme des cartes thermiques, ou de groupes catégoriques (par exemple, de simples classifications en valeurs élevées, moyennes et faibles) pourrait atténuer ces problèmes tout en fournissant des indications précieuses sur la performance et les tendances des grappes d'entreprises à l'échelle des quartiers. Cette classification impliquerait de regrouper les distributions statistiques du RE très asymétriques, où la plupart des observations se trouvent dans une petite plage de valeurs faibles.

D'autre part, le RE demeure un ensemble de données très important pour Statistique Canada et la population canadienne. Sur le site Web de l'organisme, on précise d'ailleurs, à propos du RE : « En tant que registre statistique, il fournit les listes d'unités et les attributs connexes nécessaires aux bases d'échantillonnage des enquêtes, à l'intégration des données, à la stratification et aux statistiques démographiques des entreprises. Le RE est un important élément du programme de la statistique économique de l'organisme, incluant le Recensement de l'agriculture. »

Le RE est mis à jour en continu avec l'ajout de nouvelles entreprises, tandis que les renseignements sur les emplois et les données financières des entreprises sont actualisées à intervalles réguliers. De nouvelles versions du RE, comprenant les nombres d'entreprises selon la taille des emplois, sont publiées deux fois par an; ainsi, certaines statistiques des cartes des grappes peuvent également être mises à jour à intervalles réguliers.

Enfin, les résultats obtenus avec des grappes à l'échelle des quartiers pourraient être combinés à d'autres sources de données, y compris à la fois les fonds de données de Statistique Canada et les sources de données de rechange provenant de fournisseurs externes. Par exemple, la cartographie des grappes à l'échelle des quartiers pourrait être superposée à des mesures de proximité des services et des installations et à des mesures d'accès spatial pour comprendre comment la présence d'installations ou l'accessibilité interagit avec le regroupement d'entreprises précises[28]. De même, les limites des grappes peuvent être superposées à des flux de mobilité ou à des flux de marchandises à une échelle géographique similaire. Ces renseignements peuvent fournir des indications sur le niveau d'activité économique au sein de chaque grappe. Plus précisément, la combinaison de fichiers de délimitation des grappes avec des données sur les flux de mobilité est un exemple d'intégration de données qui peut fournir des indications significatives sur les entreprises. Dans ce cas, les fichiers de délimitation seraient utilisés comme outils de géorepérage pour estimer la mobilité entrante et sortante à l'aide de données issues des appareils mobiles ou des flux de mobilité par le réseau routier des zones de grappes. Cette information pourrait servir à estimer ou surveiller les activités économiques au sein des grappes d'entreprises. Des modèles économétriques spatiaux seraient nécessaires pour reconnaître les dépendances spatiales entre ces ensembles de données.

En conclusion, nous avons observé une similarité des résultats entre les Figures 1, 6 et 10, ce qui semble suggérer que le nombre de grappes par RMR est corrélé à la structure de la population des ID de la RMR. Une question de recherche intéressante serait la suivante : dans quelle mesure la configuration des ID d'une RMR peut-elle expliquer l'émergence et la formation des grappes? Et quels seraient les autres facteurs expliquant ce phénomène? De plus, l'année actuelle du RE évaluée dans cette recherche pourrait être comparée aux années précédentes du RE disponibles dans la base de données de Statistique Canada. Cette analyse irait au-delà d'une simple analyse de colocalisation et exploiterait le signal spatio-temporel contenu dans le RE. Une régression spatiale et une autocorrélation (Moran' I) entre plusieurs versions du RE à l'échelle des quartiers permettraient de mesurer de manière robuste l'ampleur des changements dans la configuration des grappes au fil des ans. Plus précisément, cela permettrait d'identifier les quartiers d'une RMR où l'expansion industrielle actuelle en grappes s'explique par les bases historiques spatiales d'autres grappes industrielles. Il en va de même pour la stagnation et la contraction industrielles à l'échelle des quartiers. Si la comparaison de plusieurs années du RE est un exercice difficile, il faudrait alors officialiser une méthodologie suivant les étapes des méthodologies de Statistique

28. Voir : Base de données des mesures de proximité et Mesures d'accès spatial.

Canada pour les statistiques des flux bruts et des approches d'échantillonnage au fil du temps pour la création de statistiques non biaisées et cohérentes.

De plus, la bande passante actuelle du noyau reconnaît la distribution des ID et de la dimension médiane des ID de la RMR, mais ne prend pas en compte la dispersion des lieux des emplois. Par conséquent, une amélioration directe consisterait à créer une bande passante qui intègre ces deux renseignements dans le calcul. Il s'agirait d'une bande passante composite, pondérant deux composants : notre méthode et la méthode traditionnelle de Silverman.

Enfin, l'ingénierie de cartographie thermique spatiale granulaire, comme l'analyse de séries chronologiques, doit valider la présence de marches aléatoires et de corrélations parasites. Pour les séries chronologiques, la corrélation de 2 séries avec des moments statistiques non stationnaires, tels que la moyenne et la variance, peut conduire à des corrélations qui semblent fortes mais qui sont peu susceptibles de persister après une simple transformation comme une différenciation de décalage de premier pas de temps. Il en va de même pour les données spatiales. 2 cartes thermiques spatiales granulaires peuvent être non stationnaires, obtenir une très forte corrélation spatiale ou, dans certains cas, un très fort chevauchement ou une colocalisation de clusters, et ne pas préserver la forte corrélation après une simple transformation spatiale, telle que la différenciation spatiale par décalage. Des méthodes modernes pour identifier des racines unitaires spatiales et des corrélations spatiales robuste peuvent être trouvées dans Muller and Watson, 2023, and Hassan, 2012.

Conclusions

Dans la présente analyse, nous décrivons une méthode permettant de mettre au point des grappes d'entreprises à l'échelle des quartiers en utilisant les données à l'échelle des établissements du RE pour quatre RMR au Canada. Une nouvelle méthode pour définir la taille de la bande passante du noyau est détaillée, car la méthode traditionnelle de la règle de Silverman ne parvient pas, dans nos applications, à reconnaître directement la configuration de la structure des ID au sein des villes ciblées. Les grappes d'entreprises générées comblent une lacune de données sur l'analyse des grappes d'entreprises à un niveau géographique très détaillé, offrant un cadre pouvant soutenir la prise de décision et les politiques à l'échelle locale. Il propose également un cadre comparatif pour l'analyse des tendances dans les régions métropolitaines du Canada.

La littérature existante suggère que les choix d'emplacement des entreprises vont au-delà de la sélection des régions métropolitaines. Les caractéristiques des quartiers sont également des facteurs déterminants pertinents pour ces choix. En conséquence, les tendances des entreprises peuvent varier considérablement d'un quartier à l'autre au sein des mêmes régions métropolitaines. À son tour, le regroupement des entreprises dans un quartier peut influencer la prospérité économique globale et la qualité de vie dans ce quartier, contribuant ainsi soit à l'expansion, soit à la réduction des disparités spatiales dans l'ensemble de la région métropolitaine.

La méthodologie proposée dans la présente analyse s'inspire de la littérature existante sur l'identification des districts centraux. En termes simples, elle transforme la distribution spatiale discrète et fragmentée des lieux des emplois (basée sur la géolocalisation des établissements) en une représentation relativement plus lisse ou continue en utilisant une grille de niveau fin pour soutenir la méthode de KDE spatiale, puis fusionne les résultats à l'échelle des ID. Les données présentées proviennent des registres commerciaux du RE de Statistique Canada. Le modèle a introduit une nouvelle méthode de bande passante afin de reconnaître plus efficacement la configuration des ID au sein d'une RMR et d'être plus robuste aux valeurs aberrantes présentes dans la distribution très asymétrique du nombre d'emplois du RE par établissement. Les résultats générés par ce processus sont filtrés pour éliminer les ID uniques ou les petites agrégations d'ID qui ne répondraient pas aux seuils de confidentialité de base pour les données des entreprises. Les résultats montrent que la méthode proposée est efficace pour saisir la grande majorité, sinon la quasi-totalité, des emplois dans les industries respectives, tout en filtrant simultanément une grande proportion d'ID où la densité des lieux des emplois est relativement plus faible et moins pertinente. De plus, bien que cette présente soit axée sur la délimitation des frontières (les zones des grappes étant monochromes dans les cartes en annexe), l'application analytique future peut mettre en évidence différentes tendances avec les zones de grappes en utilisant, par exemple, des cartes thermiques.

Alors que la demande pour des renseignements spatiaux plus détaillés sur les entreprises ne cesse de croître, l'utilisation des grappes d'entreprises à l'échelle des quartiers, comme cadre de référence spatial, peut appuyer les travaux des agents économiques locaux et des responsables de l'élaboration de politiques. L'analyse à

l'échelle des quartiers peut servir aux associations d'entreprises locales désireuses de mieux comprendre et de suivre la situation des entreprises locales au sein d'un quartier précis. En outre, les limites des grappes locales peuvent être intégrées à d'autres mesures de la distribution spatiale à l'échelle des ID, comme la densité de population et la proximité des installations, pour générer des analyses plus complètes des conditions locales et des possibilités de développement.

Références

Aitkin, M. (2022). Introduction to Statistical Modelling and Inference. The multinomial distribution. 1st Edition. Chapman and Hall/CRC.

Bekar, C., & Lipsey, R. G. (2001). Clusters and economic policy. *Canadian Journal of Policy Research, 3*(1), 62–70.

Bartle, R. G, & Sherbert, D. R. (2001). Introduction to real analysis. Wiley; 4th edition.

Bathelt, H., & Li, P.-F. (2014). Global cluster networks: Foreign direct investment flows from Canada to China. *Journal of Economic Geography, 14*(1), 45–71. https://doi.org/10.1093/jeg/lbt020

Biemer, Paul. P., & Strokes, S, Lynne. (1984). An improved procedure for the estimating the components of response variance in complex surveys. Bureau of the Census. Statistical Research Division Report Series. SRD Research Report Number: CENSUS/SRD/RR-84/06. https://www.census.gov/content/dam/Census/library/working-papers/1984/adrm/rr84-06.pdf

Campaniaris, C., Hayes, S., Jeffrey, M., & Murray, R. (2011). The applicability of cluster theory to Canada's small and medium-sized apparel companies. *Journal of Fashion Marketing and Management: An International Journal, 15*(1), 8–26.

Christensen, P., McIntyre, N., & Pikholz, L. (2002). *Bridging community and economic development: A strategy for using industry clusters to link neighbourhoods to the regional economy*. Shorebank Enterprise Group.

Cottineau, C., & Arcaute, E. (2020). The nested structure of urban business clusters. *Applied Network Science, 5*(2). https://doi.org/10.1007/s41109-019-0246-9

Delgado, M., Porter, M. E., & Stern, S. (2014). *Defining clusters of related industries* (NBER Working Paper No. 20375). National Bureau of Economic Research. https://www.nber.org/papers/w20375

Dhamo, Z., Beleraj, I., & Kume, V. (2023). Business Improvement Districts: A comparative analysis of the legal framework and economic/social impact among different countries. In N. Persiani, I. E. Vannini, M. Giusti, A. Karasavvoglou, & P. Polychronidou (Eds.), *Global, regional and local perspectives on the economies of Southeastern Europe: EBEEC 2022*. Springer. https://doi.org/10.1007/978-3-031-34059-8_4

Director, H., & Raftery, A. (2022). Contour models for physical boundaries enclosing star-shaped and approximately star-shaped polygons. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 71*(5), 1688–1720. https://doi.org/10.1111/rssc.12592

Ericson, W. A., (1969). Subjectives Bayesian models in sampling finite populations. Journal of the statistical society. Vol. 31, No. 2, pp. 195-233

Fisher RA (1925). "Applications of 'Student's' distribution" (PDF). *Metron*. 5: 90–104. Archived from the original (PDF) on 5 March 2016.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics, 114*(3), 739–766.

Gabaix, X., & Ioannides, Y. (2004). The evolution of city size distributions. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2341–2378). Elsevier.

Gabaix, X., Lasry, J., Lions, P., & Moll, B. (2016). The dynamics of inequality. *Econometrica, 84*(6), 2071–2111.

Grimaldi, R, P,. (2003). Discret and combinatorial mathematics. 5th edition.

Grodach, C., Currid-Halkett, E., Foster, N., & Murdoch III, J. (2014). The location patterns of artistic clusters: A metro- and neighborhood-level analysis. *Urban Studies, 51*(13), 2822–2843.

Hassan, A. (2012). *Stationarity and unit roots in spatial auto-regressive models* (Doctoral dissertation, Universidad Nacional de Colombia). Retrieved from https://repositorio.unal.edu.co/bitstream/handle/unal/9882/Stationarity_and_unit_roots_in_spatial_autoregressive_models.pdf?sequence=1&isAllowed=y

Hurst, Simon. "The characteristic function of the Student *t* distribution". Financial Mathematics Research Report. Statistics Research Report No. SRR044-95. Archived from the original on February 18, 2010.

Li, C. H., & Calder, N. (2005). Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. https://doi.org/10.1111/j.1538-4632.2007.00708.x

Lucas, M., Sands, A., & Wolfe, D. A. (2009). Regional clusters in a global industry: ICT clusters in Canada. *European Planning Studies, 17*(2), 189–209. https://doi.org/10.1080/09654310802553415

Maoh, H., & Kanaroglou, P. (2007). Geographic clustering of firms and urban form: A multivariate analysis. *Journal of Geographical Systems, 9*(1), 29–52. https://doi.org/10.1007/s10109-006-0029-6

Meltzer, R. (2012). Understanding business improvement district formation: An analysis of neighborhoods and boundaries. *Journal of Urban Economics, 71*(1), 66–78. https://doi.org/10.1016/j.jue.2011.08.005

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1/2), 17–23.

Muller, U., & Watson, M. (2022). Spatial correlation robust inference. *Econometrica, 90*(6), 2901–2935.

Muller, U., & Watson, M. (2023). Spatial correlation robust inference in linear regression and panel models. *Journal of Business & Economic Statistics, 41*(4), 1050–1064. https://doi.org/10.1080/07350015.2022.2127737

Muller, U., & Watson, M. (2023). Spatial unit roots. Princeton University.

Niosi, J., & Bas, T. G. (2001). The competencies of regions – Canada's clusters in biotechnology. *Small Business Economics, 17*(1), 31–42. https://doi.org/10.1023/A:1011114220694

Na, J., Cao, X., Chen, J., & Chen, X. (2023). A new method for identifying the central business districts with nighttime light radiance and angular effects. *Remote Sensing, 15*(1), 239. https://doi.org/10.3390/rs15010239

OECD (2018). *Divided cities: Understanding intra-urban inequalities*. OECD Publishing. https://doi.org/10.1787/9789264300385-en

Patel, P. C. (2024). Nurturing neighborhoods, cultivating local businesses: The effects of amenities-to-infrastructure spending on new business licenses in Chicago's wards. *Journal of Business Venturing Insights, 21*, e00467.

E. Parzen. On the estimation of a probability density and the mode. Ann. Math. Statist., 33:1965–1976, 1962.

Rao, J.N.K., Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. Journal of American statistical association, Vol. 76, No. 374 (Jun., 1981), pp. 221-230

M. Rosenblatt. A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. USA, 42:43–47, 1956.

Rozenfeld, H., Rybski, D., Gabaix, X., & Makse, H. (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review, 101*(5), 2205–2225. http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.2205

Sergerie, F., Chastko, K., Saunders, D., & Charbonneau, P. (2021). Defining Canada's downtown neighbourhoods: 2016 boundaries. *Statistics Canada*, Issue 2021001. Retrieved from https://www150.statcan.gc.ca/n1/pub/91f0015m/91f0015m2021001-eng.pdf

Severini, A. T. (2005). Elements of distribution theory. Cambridge University Press. https://www.cambridge.org/core/books/elements-of-distribution-theory/1091612241C8BD626046E5F1E47D30A3

Shybalkina, I. (2022). Place-based small business support and its implications for neighborhood revitalization. *Economic Development Quarterly, 36*(4), 355–370. https://doi.org/10.1177/08912424221123501

Silverman, B. W. (1986). Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability. Available online at: https://ned.ipac.caltech.edu/level5/March02/Silverman/paper.pdf.

Spencer, G. M. (2014). *Cluster atlas of Canada*. Report, Munk School of Global Affairs, University of Toronto. Available online at: https://clustercollaboration.eu/sites/default/files/international_cooperation/cluster-atlas.pdf

Spencer, G. M., Vinodrai, T., Gertler, M. S., & Wolfe, D. A. (2010). Do clusters make a difference? Defining and assessing their economic performance. *Regional Studies, 44*(6), 697–715.

Steiner, B. E., & Ali, J. (2011). Government support for the development of regional food clusters: Evidence from Alberta, Canada. *International Journal of Innovation and Regional Development, 3*(2), 186–216. https://doi.org/10.1504/IJIRD.2011.038924

Stern, M. J., & Seifert, S. C. (2010). Cultural clusters: The implications of cultural assets agglomeration for neighborhood revitalization. *Journal of Planning Education and Research, 29*(3), 262–279. https://doi.org/10.1177/0739456X09358555

Taubenböck, H., Klotz, M., Wurm, J., Schmieder, B., Wooster, M., Esch, T., & Dech, S. (2013). Delineation of central business districts in mega city regions using remotely sensed data. *Remote Sensing of Environment, 136*, 386–401. https://doi.org/10.1016/j.rse.2013.05.019

Thompson, M.E., Sedransk, J., Fang, J. and Yi, G.Y. (2022). Bayesian inference for a variance component model using pairwise composite likelihood with survey data. *Survey Methodology*, Statistics Canada, Catalogue No. 12-001-X, Vol. 48, No. 1. Paper available at http://www.statcan.gc.ca/pub/12-001-x/2022001/article/00002-eng.htm.

Toronto Region Board of Trade (2021). Regional centres: A business district analysis. Report. Retrieved from https://bot.com/Resources/Resource-Library/Regional-Centres-A-Business-District-Analysis

Wang, B., & Wen, B. (2021). The spatial distribution of businesses and neighborhoods: What industries match or mismatch what neighborhoods? *Habitat International, 117*, 102440.

Wheeler, C. H. (2006). Businesses don't just choose a city, they choose a specific neighborhood. *Bridges*. Federal Reserve Bank of St. Louis. Retrieved https://fraser.stlouisfed.org/title/bridges-federal-reserve-bank-st-louis-6272/bridges-600970

Wolfe, D. A., & Gertler, M. S. (2004). Clusters from the inside and out: Local dynamics and global linkages. *Urban Studies, 41*(5-6), 1071–1093.

Yu, W., Ai, T., & Shao, S. (2015). The analysis and delimitation of central business districts using network kernel density estimation. *Journal of Transport Geography, 45*, 32–47. https://doi.org/10.1016/j.jtrangeo.2015.04.008

Zhang, X. (2019). Building effective clusters and industrial parks. In Célestin Monga & Justin Yifu Lin (Eds.), *The Oxford handbook of structural transformation*. Oxford Handbooks. https://doi.org/10.1093/oxfordhb/9780198793847.013.13

Annexe 1 : démonstration de la non-nécessité d'utiliser les probabilités de combinaison pour l'analyse comparative du processus binomiale

En admettant que $\Pr(h)$ est la probabilité de pile, l'expression complète est :

$$\left(c(v,v/2) * (\Pr(h))^{\frac{v}{2}} * (1-\Pr(h))^{v-\frac{v}{2}}\right) / \sum_{j=0}^{v} c(v,j) * (\Pr(h))^{j} * (1-\Pr(h))^{v-j}$$

$$>$$

$$\left(c(v,v) * (\Pr(h))^{v} * (1-\Pr(h))^{v-v}\right) / \sum_{j=0}^{v} c(v,j) * (\Pr(h))^{j} * (1-\Pr(h))^{v-j}$$

L'expression est alors équivalente à la suivante, si l'on remplace $\Pr(h)$ par l'uniformité de ½,

$$\left(c(v,v/2) * \left(\frac{1}{2}\right)^{\frac{v}{2}} * \left(1-\frac{1}{2}\right)^{v-\frac{v}{2}}\right) / \sum_{j=0}^{v} c(v,j) * \left(\frac{1}{2}\right)^{j} * \left(1-\frac{1}{2}\right)^{v-j}$$

$$>$$

$$\left(c(v,v) * \left(\frac{1}{2}\right)^{v} * \left(1-\frac{1}{2}\right)^{v-v}\right) / \sum_{j=0}^{v} c(v,j) * \left(\frac{1}{2}\right)^{j} * \left(1-\frac{1}{2}\right)^{v-j}$$

En simplifiant avec les soustractions des probabilités et des exposants, on obtient :

$$\left(\frac{v!}{\left(\frac{v}{2}\right)!\left(v-\frac{v}{2}\right)!} * \left(\frac{1}{2}\right)^{\frac{v}{2}} * \left(\frac{1}{2}\right)^{\frac{v}{2}}\right) / \sum_{j=0}^{v} \frac{v!}{(j)!(v-j)!} * \left(\frac{1}{2}\right)^{j} * \left(\frac{1}{2}\right)^{v-j}$$

$$>$$

$$\left(\frac{v!}{(v)!(v-v)!} * \left(\frac{1}{2}\right)^{v} * \left(\frac{1}{2}\right)^{0}\right) / \sum_{j=0}^{v} \frac{v!}{(j)!(v-j)!} * \left(\frac{1}{2}\right)^{j} * \left(\frac{1}{2}\right)^{v-j}$$

Maintenant, grâce à l'uniformité du processus binomial, on utilise la propriété additive de l'exposant de probabilité pour obtenir :

$$\left(\frac{v!}{\left(\frac{v}{2}\right)!\left(v-\frac{v}{2}\right)!} * \left(\frac{1}{2}\right)^{v} \right) / \sum_{j=0}^{v} \frac{v!}{(j)!(v-j)!} * \left(\frac{1}{2}\right)^{v}$$

$$>$$

$$\left(\frac{v!}{(v)!(v-v)!} * \left(\frac{1}{2}\right)^{v} \right) / \sum_{j=0}^{v} \frac{v!}{(j)!(v-j)!} * \left(\frac{1}{2}\right)^{v}$$

Enfin, en annulant les termes communs des deux côtés de l'inégalité et en admettant certaines égalités, on obtient:

$$\frac{v!}{\left(\frac{v}{2}\right)!\left(v-\frac{v}{2}\right)!} > \frac{v!}{(v)!(v-v)!} = \frac{v!}{(0)!(v-0)!} = 1$$

Ce qui explique pourquoi notre analyse comparative se concentre sur le nombre de combinaisons et non sur l'expression complète. ■

Cette preuve utilise un dénominateur égal à la somme de toutes les probabilités, de part et d'autre de l'inégalité. Ce dénominateur est techniquement redondant et égal à 1. Cependant, il reste utile pour visualiser les probabilités. Autrement dit, si l'on annule tous les termes de probabilité ($\Pr()$) en raison de l'uniformité du processus aléatoire, seuls les termes combinatoires subsistent au numérateur et au dénominateur. Le dénominateur devient la somme de toutes les combinaisons possibles et le numérateur le nombre de combinaisons étudiées. Cela fournit une probabilité en soi.

Appendix 2: Justification de l'application de l'allocation aléatoire des emplois au sein de l'ID plutôt que l'AD pendant le traitement pré-noyau

Sergerie et al. (2021) appliquent une répartition aléatoire uniforme des emplois au sein des limites de l'AD. Cette stratégie est excellente compte tenu du nombre important d'emplois au sein d'une AD et de la couverture d'un large ensemble de codes NAICS à deux chiffres. Cependant, dans le cas de nos applications, nous traitons un seul code NAICS à deux chiffres à la fois pour la génération des grappes. Par conséquent, le nombre d'emplois impliqués par AD est relativement plus limité. La théorie d'approximation de la normalité documentée dans cet article est conditionnelle à un grand nombre d'emplois. Autrement dit, si la surface géographique dédiée à l'allocation des emplois est grande par rapport au nombre d'emplois lui-même, il devient difficile d'obtenir une distribution normale précise. Cette annexe explique en détail la raison du traitement au sein de l'ID plutôt qu'au sein de l'AD de l'allocation aléatoire des emplois lors du traitement pré-noyau. Pour ce faire, nous décomposons l'allocation aléatoire globale de l'AD du point de vue des différents ID d'une même AD. Dans les pages suivantes de cette annexe, nous présenterons les équations y*, y** et y*** avant de présenter notre explication finale.

Définissons l'ensemble d'espaces A, A' et A'' tel que le voisinage symétrique autour du centre de gravité de l'ID, l'ID et l'AD à laquelle appartient l'ID, respectivement. A est plus petit que l'ID et son rayon est proportionnel à la taille de l'ID. Définissons l'ensemble d'espace supplémentaire $A^{*'}$ et $A^{*''}$ comme la superficie soustrayant la superficie précédente plus petite. En d'autres mots, $A^{*'} = A' \setminus A$ and $A^{*''} = A'' \setminus A'$. Définissons S, S', S'' comme le nombre fini de points spatiaux dans A, A' et A'' disponible pour une attribution aléatoire d'emplois. Définissons également $S^{*'} = S' - S$ et $S^{*''} = S'' - S'$. Supposons que $A < A' < A''$, ce qui est par définition toujours le cas. Définissons un partage de probabilité normalisée telle que $0 < \Pr(A^{*''})$, $0 < \Pr(A^{*'})$, $0 < \Pr(A)$ et $\Pr(A^{*''}) + \Pr(A^{*'}) + \Pr(A) = 1$, qui définissent le niveau de lissage ou de fragmentation. Les 3 valeur de partage ne doivent pas être proportionnelles au volume relatif des ensembles spatiaux mais plutôt proportionnelles à leur importance relative. Définissons l'uniformité sur A, $A^{*'}$ et $A^{*''}$. Autrement dit, $\Pr(A) * (1/S)$, $\Pr(A^{*'}) * (1/S^{*'})$ et $\Pr(A^{*''}) * (1/S^{*''})$ sont les probabilités égales pour tous les points spatiaux situés respectivement dans A, $A^{*'}$ et $A^{*''}$. Il est trivial de montrer que la somme de toutes les probabilités est 1 dans le sens où l'espace des mesures de probabilités est bien défini. Autrement dit, $\left(\Pr(A) * (1/S) * S\right) + \left(\Pr(A^{*'}) * (1/S^{*'}) * S^{*'}\right) + \left(\Pr(A^{*''}) * (1/S^{*''}) * S^{*''}\right) = 1$. Si la somme des trois termes de $\Pr()$ est égale à 1, alors la somme de la dernière expression reste égale à 1, quelle que soit la taille de S, $S^{*'}$ et $S^{*''}$. Une application permettant une dispersion limitée au-delà de l'ID serait $\Pr(S) = 0{,}6$, $\Pr(S^{*'}) = 0{,}3$ et $\Pr(S^{*''}) = 0.1$. Pour des raisons de commodité, nous étiquetons les probabilités telles que $\check{Z} = \Pr(A) * (1/S)$, $\check{T} = \Pr(A^{*'}) * (1/S^{*'})$ et $\check{N} = \Pr(A^{*''}) * (1/S^{*''})$. De plus, $1 - \check{Z} = \check{Z}$-, $1 - \check{T} = \check{T}-$ et $1 - \check{N} = \check{N}-$. Un tel processus aléatoire RP^* n'est plus nécessairement uniforme et converge toujours en distribution vers la distribution normale multivariée. Autrement dit, $RP^* \rightarrow N()$, si $v \rightarrow \infty$, et dans un contexte finit, $RP^* \approx N(vp^{*T}, vM^*)$, si v est large. v[29] est le nombre total d'emplois de l'ID d'intérêt (ensemble spatial A') et vp^* est le nouveau vecteur attendu de dimension $(S + S^{*'} + S^{*''})$, autrement dit, $vp^* = \langle v\check{Z}, \ldots, v\check{Z}, v\check{T}, \ldots, v\check{T}, v\check{N}, \ldots, v\check{N} \rangle$, est un vecteur ligne et $M^* = P^* - p^* p^{*T}$, et $P^* = I^* p^{*T}$, et I^* est la matrice identité de dimension $(S + S^{*'} + S^{*''}) x (S + S^{*'} + S^{*''})$. Autrement dit, P^* est une matrice diagonale dont les éléments diagonaux sont les éléments du vecteur p^*. Autrement dit,

$$P^* = I^* p^{*T} =$$

29. v n'utilise pas la notation « * », car il s'agit du même v que celui présenté dans la méthodologie principale de cet article. Il s'agit du nombre d'emplois uniques v au sein d'un ID.

$$\begin{bmatrix}
\check{Z} & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
0 & \check{Z} & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & \check{Z} & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & \check{T} & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & \check{T} & \dots & 0 & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & \check{T} & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \check{N} & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \check{N} & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & \check{N}
\end{bmatrix}$$

$$\text{Et } p^* p^{*T} = \check{Z}\check{Z} + \dots + \check{Z}\check{Z} + \check{T}\check{T} + \dots + \check{T}\check{T} + \check{N}\check{N} + \dots + \check{N}\check{N}$$

De son côté, vM^* est la nouvelle matrice de variance-covariance, c'est-à-dire que la quantification complète de l'incertitude peut être représentée par $vM^* =$

$$\begin{bmatrix}
v\check{Z}(\check{Z}-) & -v\check{Z}\check{Z} & \cdots & -v\check{Z}\check{Z} & -v\check{Z}\check{T} & -v\check{Z}\check{T} & \cdots & -v\check{Z}\check{T} & -v\check{Z}\check{N} & -v\check{Z}\check{N} & \cdots & -v\check{Z}\check{N} \\
-v\check{Z}\check{Z} & v\check{Z}(\check{Z}-) & \cdots & -v\check{Z}\check{Z} & -v\check{Z}\check{T} & -v\check{Z}\check{T} & \cdots & -v\check{Z}\check{T} & -v\check{Z}\check{N} & -v\check{Z}\check{N} & \cdots & -v\check{Z}\check{N} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
-v\check{Z}\check{Z} & -v\check{Z}\check{Z} & \cdots & v\check{Z}(\check{Z}-) & -v\check{Z}\check{T} & -v\check{Z}\check{T} & \cdots & -v\check{Z}\check{T} & -v\check{Z}\check{N} & -v\check{Z}\check{N} & \cdots & -v\check{Z}\check{N} \\
. & . & \cdots & . & v\check{T}(\check{T}-) & -v\check{T}\check{T} & \cdots & -v\check{T}\check{T} & -v\check{T}\check{N} & -v\check{T}\check{N} & \cdots & -v\check{T}\check{N} \\
. & . & \cdots & . & -v\check{T}\check{T} & v\check{T}(\check{T}-) & \cdots & -v\check{T}\check{T} & -v\check{T}\check{N} & -v\check{T}\check{N} & \cdots & -v\check{T}\check{N} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
. & . & \cdots & . & -v\check{T}\check{T} & -v\check{T}\check{T} & \cdots & v\check{T}(\check{T}-) & -v\check{T}\check{N} & -v\check{T}\check{N} & \cdots & -v\check{T}\check{N} \\
. & . & \cdots & . & . & . & \cdots & . & v\check{N}(\check{N}-) & -v\check{N}\check{N} & \cdots & -v\check{N}\check{N} \\
. & . & \cdots & . & . & . & \cdots & . & -v\check{N}\check{N} & v\check{N}(\check{N}-) & \cdots & -v\check{N}\check{N} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
. & . & \cdots & . & . & . & \cdots & . & -v\check{N}\check{N} & -v\check{N}\check{N} & \cdots & v\check{N}(\check{N}-)
\end{bmatrix}$$

vM^* est une matrice par blocs finis dont la diagonale de bloc comprend 3 sous-matrices, une pour A, $A^{*'}$ et $A^{*''}$, respectivement. Les dimensions des 3 sous-matrices sont $S \times S$, $S^{*'} \times S^{*'}$ et $S^{*''} \times S^{*''}$, respectivement. Par conséquent, la matrice de variance-covariance vM^* est symétrique et a pour dimension $\left(S+S^{*'}+S^{*''}\right)x\left(S+S^{*'}+S^{*''}\right)$. vM^* est également une généralisation de vM, présentée dans cet article. Autrement dit, vM^* se réduit à vM et est égale aux deux premières des 3 sous-matrices de la diagonale de bloc de vM^*, si $Pr\left(A^{*''}\right)$ n'existe pas, et $Pr(A) + Pr\left(A^{*'}\right) = 1$ et $Pr(A) * (1/S) = Pr\left(A^{*'}\right) * \left(1/S^{*'}\right) = 1$. Les termes matriciels situés sous la diagonale de bloc sont omis pour plus de clarté et pour rendre les 3 sous-matrices de la diagonale de bloc plus évidentes. Un terme situé plus bas que la diagonale de block est égale au terme de la coordonnée inverse au-dessus de la diagonale de block. vM^* est légèrement plus sophistiquée que vM en raison de sa structure spatiale à trois niveaux impliquant A, $A^{*'}$ et $A^{*''}$. La nouvelle distribution normale multivariée de notre processus aléatoire multinomial convergent peut être exprimée[30] dans un contexte fini de la manière suivante,

30. Toutes les dimensions des vecteurs et des matrices sont techniquement basées sur $\left(S+S^{*'}+S^{*''}\right)-1$, et non $\left(S+S^{*'}+S^{*''}\right)$. Nous conservons $\left(S+S^{*'}+S^{*''}\right)$ pour simplifier la notation. Par définition, la dernière catégorie d'une variable aléatoire multinomiale est une information redondante car la somme des nombres par catégorie est une contrainte totale fixe, et le nombre d'emplois alloués à la dernière catégorie est automatiquement déduit de la somme de toutes les autres catégories. Par conséquent, la matrice de variance-covariance est déficiente en rang et non en rang complet. Pour éviter la singularité, l'impossibilité d'inversion de matrice et le déterminant de matrice incorrect, la dimension de la distribution normale multivariée doit s'effondrer dans un sous-espace de $\left(S+S^{*'}+S^{*''}\right)$, qui est simplement $\left(S+S^{*'}+S^{*''}\right)-1$.

$$y^* = \mathrm{MVN}\left(x^*; vp^*, vM^*\right) = \left(2\pi\right)^{-\frac{S+S^{*'}+S^{*''}}{2}} * \det\left(vM^*\right)^{-\frac{1}{2}} * \exp\left(-\frac{1}{2}\left(x^* - vp^*\right)\left(vM^*\right)^{-1}\left(x^* - vp^*\right)^{\mathrm{T}}\right) \cdot$$
$$\text{et } x^* \approx \mathrm{N}\left(vp^{*\mathrm{T}}, vM^*\right)$$

Définissons ADA comme l'ensemble de tous les ID de l'AD. Les éléments présentés jusqu'à présent dans cette annexe concernent l'allocation aléatoire d'une population de v emploie uniques d'un ID dans son ADA. Cependant, il convient de procéder de la même manière pour toutes les autres ID de la même ADA, car chaque ID de l'ADA allouera aléatoirement ses propres v emploie parmi les IDs de la même ADA. Par conséquent, nous devons maintenant considérer la distribution normale multivariée conjointe plutôt qu'une distribution normale multivariée unique. Fondamentalement, il s'agit d'un simple exercice de concaténation des éléments présentés précédemment dans cette annexe. Pour ce faire, il suffit de conserver la notation existante de cette annexe et d'introduire simplement l'indice de $\mathrm{l'ID}$ et de le faire varier entre $\mathrm{ID} = 1, 2, \ldots, \mathrm{q}\left(\mathrm{ADA}\right)$, où q(ADA) est le nombre total $\mathrm{d'ID}$ uniques dans l'ADA avec des emplois disponible. Les éléments précédents de cette annexe montrent que chaque processus aléatoire, un pour chaque ID, converge vers la distribution normale multivariée. Autrement dit,

$$\mathrm{RP}\left(1\right)^* \rightarrow \mathrm{N}(), \text{ si } \mathrm{v}\left(1\right) \rightarrow \infty ,$$

$$\mathrm{RP}\left(2\right)^* \rightarrow \mathrm{N}(), \text{ si } \mathrm{v}\left(2\right) \rightarrow \infty ,$$

...,

$$\mathrm{RP}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^* \rightarrow \mathrm{N}(), \text{ si } \mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right) \rightarrow \infty$$

Dans un contexte fini, cela équivaut à dire que,

$$\mathrm{RP}\left(1\right)^* \approx \mathrm{N}\left(\mathrm{v}\left(1\right)\mathrm{p}\left(1\right)^{*\mathrm{T}}, \mathrm{v}\left(1\right)\mathrm{M}\left(1\right)^*\right), \text{ si } \mathrm{v}\left(1\right) \text{ est grand,}$$
$$\mathrm{RP}\left(2\right)^* \approx \mathrm{N}\left(\mathrm{v}\left(2\right)\mathrm{p}\left(2\right)^{*\mathrm{T}}, \mathrm{v}\left(2\right)\mathrm{M}\left(2\right)^*\right), \text{ si } \mathrm{v}\left(2\right) \text{ est grand,}$$

...,

$$\mathrm{RP}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^* \approx \mathrm{N}\left(\mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)\mathrm{p}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^{*\mathrm{T}}, \mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)\mathrm{M}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^*\right), \text{ si } \mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right) \text{ est grand}$$

Cela équivaut également à dire que,

$$\mathrm{x}\left(1\right)^* \approx \mathrm{N}\left(\mathrm{v}\left(1\right)\mathrm{p}\left(1\right)^{*\mathrm{T}}, \mathrm{v}\left(1\right)\mathrm{M}\left(1\right)^*\right), \text{ si } \mathrm{v}\left(1\right) \text{ est grand,}$$
$$\mathrm{x}\left(2\right)^* \approx \mathrm{N}\left(\mathrm{v}\left(2\right)\mathrm{p}\left(2\right)^{*\mathrm{T}}, \mathrm{v}\left(2\right)\mathrm{M}\left(2\right)^*\right), \text{ si } \mathrm{v}\left(2\right) \text{ est grand,}$$

...,

$$\mathrm{x}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^* \approx \mathrm{N}\left(\mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)\mathrm{p}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^{*\mathrm{T}}, \mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)\mathrm{M}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^*\right), \text{ si } \mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right) \text{ est grand}$$

Chacune des distributions normales multivariées ci-dessus se réfère à un espace distinct de mesures de probabilités. Autrement dit, le vecteur de probabilité et l'ensemble des employés (emplois) sont distincts. En revanche, l'ensemble des points spatiaux est identique. En utilisant l'opérateur produit et en supposant l'indépendance des différentes distributions, les distributions normales multivariées peuvent être exprimées dans une seule équation et générer la distribution normale multivariée conjointe y^{**}, c'est-à-dire :

$$y^{**} = \mathrm{y}\left(1\right)^* * \mathrm{y}\left(2\right)^* * \ldots * \mathrm{y}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^*$$

$$= \mathrm{JMVN}\left(\mathrm{x}\left(1\right)^*, \mathrm{x}\left(2\right)^*, \ldots, \mathrm{x}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^*; \mathrm{v}\left(1\right)\mathrm{p}\left(1\right)^*, \mathrm{v}\left(2\right)\mathrm{p}\left(2\right)^*, \ldots, \mathrm{v}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)\mathrm{p}\left(\mathrm{q}\left(\mathrm{ADA}\right)\right)^*,\right.$$

$$v(1)M(1)^{*},v(2)M(2)^{*},\ldots,v(q(ADA))M(q(ADA))^{*}\Big)$$

$$=(2\pi)^{-\frac{S+S^{*'}+S^{*''}}{2}} * \det\left(v(1)M(1)^{*}\right)^{-\frac{1}{2}} * \exp\left(-\frac{1}{2}\left(x(1)^{*}-v(1)p(1)^{*}\right) * \left(v(1)M(1)^{*}\right)^{-1} * \left(x(1)^{*}-v(1)p(1)^{*}\right)^{T}\right)$$

$$* (2\pi)^{-\frac{S+S^{*'}+S^{*''}}{2}} * \det\left(v(2)M(2)^{*}\right)^{-\frac{1}{2}} * \exp\left(-\frac{1}{2}\left(x(2)^{*}-v(2)p(2)^{*}\right) * \left(v(2)M(2)^{*}\right)^{-1} * \left(x(2)^{*}-v(2)p(2)^{*}\right)^{T}\right)$$

$$*\ldots*$$

$$(2\pi)^{-\frac{S+S^{*'}+S^{*''}}{2}} * \det\left(v(q(ADA))M(q(ADA))^{*}\right)^{-\frac{1}{2}}$$

$$* \exp\left(-\frac{1}{2}\left(x(q(ADA))^{*}-v(q(ADA))p(q(ADA))^{*}\right) * \left(v(q(ADA))M(q(ADA))^{*}\right)^{-1} * \left(x(q(ADA))^{*}-v(q(ADA))p(q(ADA))^{*}\right)^{T}\right)$$

$$=(2\pi)^{-\frac{q(ADA)*(S+S^{*'}+S^{*''})}{2}} *\Big(\prod_{ID=1}^{ID=q(ADA)} \det\left(v(ID)M(ID)^{*}\right)^{-\frac{1}{2}}\Big)$$

$$*\left(\exp\left(\sum_{ID=1}^{ID=q(ADA)} -\frac{1}{2}\left(x(ID)^{*}-v(ID)p(ID)^{*}\right)*\left(v(ID)M(ID)^{*}\right)^{-1}*\left(x(ID)^{*}-v(ID)p(ID)^{*}\right)^{T}\right)\right)$$

En regardant la troisième égalité, nous pouvons remarquer que le terme $(S+S^{*'}+S^{*''})$ n'utilise pas l'indice ID. Autrement dit, il est utilisé de manière répétitive pour chaque équation normale multivariée car la valeur ne change pas à travers les ID. En regardant la quatrième égalité (la dernière égalité), la sommation est justifiée par la propriété de l'exposant sous le terme commun $\exp$ et est utilisée pour diminuer la longueur de l'expression complète. Il est à noter que la distribution conjointe y^{**} atteint sa densité maximale si chaque vecteur $x(1)^{*}, x(2)^{*},\ldots,x(q(ADA))^{*}$ est égal à leur vecteur moyen $v(1)p(1)^{*}, v(2)p(2)^{*},\ldots,v(q(ADA))p(q(ADA))^{*}$, respectivement. De la même manière, y^{**} atteint sa densité minimale si chaque vecteur $x(1)^{*}, x(2)^{*},\ldots,x(q(ADA))^{*}$ alloue la totalité de ses ressources à l'élément du vecteur de probabilité $p(1)^{*},p(2)^{*},\ldots,p(q(ADA))^{*}$ où la probabilité est la plus faible possible, respectivement. De plus, chaque variable d'entrée $x(1)^{*}, x(2)^{*},\ldots,x(q(ADA))^{*}$ de la distribution conjointe y^{**} est limitée à son nombre total d'emplois disponibles $v(1),v(2),\ldots,v(q(ADA))$, respectivement. Par conséquent, les distributions normales multivariées $y(1)^{*}, y(2)^{*},\ldots,y(q(ADA))^{*}$ sont totalement indépendantes les unes des autres, dans le sens où l'allocation aléatoire des emplois d'un ID au sein de l'ADA ne renseigne pas sur la façon dont une autre ID allouera ses propres ressources au sein de la même ADA. Autrement dit, nous pouvons écrire la matrice de variance-covariance M^{**} suivante pour résumer le contexte. En d'autres mots,

$$M^{**} = \begin{bmatrix} v(1)M(1)^{*} & 0 & \cdots & 0 \\ 0 & v(2)M(2)^{*} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & v\left(q(ADA)\right)M\left(q(ADA)\right)^{*} \end{bmatrix}$$

Où, M^{**} est la matrice de variance-covariance de la normale multivariée conjointe du vecteur de vecteurs $\left(x(1)^{*}, x(2)^{*},\ldots,x\left(q(ADA)\right)^{*}\right)$, ce qui veux dire que,

$$x^{**} = \left(x(1)^{*},x(2)^{*},\ldots,x\left(q(ADA)\right)^{*}\right) \approx N\left(\left(v(1)p(1)^{*T},v(2)p(2)^{*T},\ldots,v\left(q(ADA)\right)p\left(q(ADA)\right)^{*T}\right),M^{**}\right), \text{ si}$$

$$v(1),\ v(2),\ldots,v\left(q(ADA)\right) \text{ es large, respectivement.}$$

Chaque élément de la matrice M^{**} a été introduit précédemment. M^{**} n'est qu'un moyen de reformuler les informations existantes pour clarifier les idées. La dimension de M^{**} est $q(ADA)\left(S+S^{*'}+S^{*''}\right)*q(ADA)\left(S+S^{*'}+S^{*''}\right)$[31]. M^{**} doit être considérée comme une matrice de blocs imbriquée. Autrement dit, M^{**} est une matrice de blocs diagonale, où chaque élément sur la diagonale est lui-même une matrice de blocs de dimension $\left(S+S^{*'}+S^{*''}\right)x\left(S+S^{*'}+S^{*''}\right)$ représentant les interconnexions au sein d'une seule ID, et chaque élément hors diagonale est une matrice de blocs de dimension $\left(S+S^{*'}+S^{*''}\right)x\left(S+S^{*'}+S^{*''}\right)$ peuplée entièrement de zéros représentant l'absence d'interconnexion entre une paire de ID distinctes faisant partie de la même ADA avec données disponible. Cette dernière n'est rien de plus que la représentation matricielle de l'indépendance complète des différentes distributions normales multivariées justifiant l'utilisation de l'opérateur produit dans l'équation normale multivariée conjointe précédente.

Le matériel présenté sur la page précédente de cette annexe propose une distribution conjointe de $q(ADA)$ normales multivariées distinctes où chacune est dédiée à un ID spécifique de la même ADA. La variable d'entrée est un vecteur de vecteurs $\left(x(1)^{*},x(2)^{*}, \ldots ,x\left(q(ADA)\right)^{*}\right)$ et nécessite que l'analyste saisisse $q(ADA)*\left(S+S^{*'}+S^{*''}\right)$ valeur distincte. Ce nombre d'entrées à fournir est très grand et représente une charge considérable pour l'analyste. Pour cette raison, nous reformulons la même idée « conjointe » originale dans un modèle alternatif où la variable d'entrée est maintenant de dimension $\left(S+S^{*'}+S^{*''}\right)$ uniquement. Ce modèle alternatif utilise une seule distribution normale multivariée. Cependant, le vecteur moyen est la somme de tous les $q(ADA)$ vecteurs moyens et la matrice de variance-covariance est la somme de toutes les $q(ADA)$ matrices de variance-covariance des $q(ADA)$ distributions normales multivariées. Formellement, dans un contexte fini, nous avons :

$$x^{***} = x(1)^{*} + x(2)^{*} + \ldots + x\left(q(ADA)\right)^{*} \approx N\left(p^{**T},M^{***}\right),$$

$$\text{si } v(1),v(2),\ldots,v\left(q(ADA)\right) \text{ est grand, respectivement.}$$

Où, x^{***} est un vecteur ligne de dimension $\left(S+S^{*'}+S^{*''}\right)$, où la somme des termes de x^{***} est le scalaire VID, c'est-à-dire,

31. Toutes les dimensions des vecteurs et des matrices sont techniquement basées sur $\left(S+S^{*'}+S^{*''}\right)-1$, et non $\left(S+S^{*'}+S^{*''}\right)$. Nous conservons $\left(S+S^{*'}+S^{*''}\right)$ pour simplifier la notation. Par définition, la dernière catégorie d'une variable aléatoire multinomiale est une information redondante car la somme des nombres par catégorie est une contrainte totale fixe, et le nombre d'emplois alloués à la dernière catégorie est automatiquement déduit de la somme de toutes les autres catégories. Par conséquent, la matrice de variance-covariance est déficiente en rang et non en rang complet. Pour éviter la singularité, l'impossibilité d'inversion de matrice et le déterminant de matrice impropre, la dimension de la distribution normale multivariée doit s'effondrer dans un sous-espace de $\left(S+S^{*'}+S^{*''}\right)$, qui est simplement, $\left(S+S^{*'}+S^{*''}\right)-1$. Dans le cas spécifique de M^{**}, la dimension est basée sur $q(ADA)\left(\left(S+S^{*'}+S^{*''}\right)-1\right)$, et non $q(ADA)\left(S+S^{*'}+S^{*''}\right)$, car chacune des $q(ADA)$ matrices nécessite un ajustement dans la matrice de blocs M^{**}. Même chose pour le vecteur d'intrant et le vecteur espéré de la distribution multivariée jointe normale.

$$x^{***}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}=\left(1^{*}v(1)\right)+\left(1^{*}v(2)\right)+\ldots+\left(1^{*}v\left(q(ADA)\right)\right)=VID$$

Et où,

$$E(x^{***})=p^{**}=\sum_{ID=1}^{ID=q(ADA)} v(ID)p(ID)^{*}$$

Et où,

$$var\left(x^{***}\right)=M^{***}=\sum_{ID=1}^{ID=q(ADA)} v(ID)M(ID)^{*}$$

Sous l'indépendance des variables au niveau de l'ID, et où la nouvelle distribution normale multivariée peut être exprimée comme[32],

$$y^{***}=MVN\left(x^{***};p^{**},M^{***}\right)$$

$$=(2\pi)^{-\frac{S+S^{*'}+S^{*''}}{2}}*\det\left(M^{***}\right)^{-\frac{1}{2}}*\exp\left(-\frac{1}{2}\left(x^{***}-p^{**}\right)*\left(M^{***}\right)^{-1}*\left(x^{***}-p^{**}\right)^{T}\right)$$

y^{***} diffère significativement de y^{**} présenté précédemment dans cette annexe. La variable d'entrée x^{***} spécifie l'allocation de VID emplois uniques au sein d'un seul vecteur d'entrée de dimension $\left(S+S^{*'}+S^{*''}\right)$, au lieu de $q(ADA)$ vecteurs distincts de dimension $\left(S+S^{*'}+S^{*''}\right)$ chacun. x^{***} est approximé par une distribution normale multivariée puisque la distribution exacte serait une convolution de plusieurs distribution multinomiales non disponibles sous forme fermée.

Enfin, nous pouvons maintenant décrire le processus aléatoire de Sergerie et al, 2021 comme un cas particulier, où pour tous les ID d'un même AD avec données disponible, nous avons,

$$Pr\left(A(ID)\right)>0,\ Pr\left(A(ID)^{*'}\right)>0,\ Pr\left(A^{*''}(ID)^{*''}\right)>0,$$

$$Pr\left(A(ID)\right)+Pr\left(A(ID)^{*'}\right)+Pr\left(A(ID)^{*''}\right)=1,\text{ et}$$

$$Pr\left(A(ID)\right)*\left(1/S(ID)\right)=Pr\left(A(ID)^{*'}\right)*\left(1/S(ID)^{*'}\right)=Pr\left(A(ID)^{*''}\right)*\left(1/S(ID)^{*''}\right).$$

La dernière condition garantit que tous les emplois uniques de l'AD sont répartis aléatoirement et uniformément au sein de l'AD. De plus, la méthode aléatoire principale de cet article est un cas particulier si,

$$Pr\left(A^{*''}\right)\text{ n'existe pas, }Pr\left(A(ID)\right)>0,\ Pr\left(A(ID)^{*'}\right)>0$$

$$Pr\left(A(ID)\right)+Pr\left(A(ID)^{*'}\right)=1,\text{ et}$$

$$Pr\left(A(ID)\right)*\left(1/S(ID)\right)=Pr\left(A(ID)^{*'}\right)*\left(1/S(ID)^{*'}\right),$$

pour tous les ID de l'ADA avec données disponible. (ne s'applique pas pour y***)

32. Toutes les dimensions des vecteurs et des matrices sont techniquement basées sur $\left(S+S^{*'}+S^{*''}\right)-1$, et non $\left(S+S^{*'}+S^{*''}\right)$. Nous conservons $\left(S+S^{*'}+S^{*''}\right)$ pour simplifier la notation. Par définition, la dernière catégorie d'une variable aléatoire multinomiale est une information redondante car la somme des nombres par catégorie est une contrainte totale fixe, et le nombre d'emplois alloués à la dernière catégorie est automatiquement déduit de la somme de toutes les autres catégories. Par conséquent, la matrice de variance-covariance est déficiente en rang et non en rang complet. Pour éviter la singularité, l'impossibilité d'inversion de matrice et le déterminant de matrice impropre, la dimension de la distribution normale multivariée doit s'effondrer dans un sous-espace de $\left(S+S^{*'}+S^{*''}\right)$, qui est simplement, $\left(S+S^{*'}+S^{*''}\right)-1$. Dans le cas spécifique de la matrice M^{***}, la somme de $q(ADA)$ matrices de dimension $\left(S+S^{*'}+S^{*''}\right)-1$ est également de dimension $\left(S+S^{*'}+S^{*''}\right)-1$.

Maintenant que y*, y** et y*** ont été introduits en détail, nous pouvons présenter notre justification de la raison pour laquelle une allocation aléatoire de emplois au sein de l'ID est préférée pour le cas pratique de notre application.

Avec y**, sous une grande surface géospatiale telle qu'une AD, deux hypothèses sont à prendre en compte : 1) chaque distribution normale multivariée $\mathrm{y}(1)^{*},\mathrm{y}(2)^{*},\ldots,\mathrm{y}\left(\mathrm{q}(\mathrm{ADA})\right)^{*}$ est supposée être correctement spécifiée comme une distribution normale. Plus précisément, pour chaque distribution multivariée, le taux d'occurrence attendu λ de l'intervalle d'espace polygonal d'intérêt, à savoir un emplacement spatial ou un seul pixel dans l'AD, est plus que modéré. Autrement dit, les vecteurs de résultats attendus $\mathrm{v}(1)\mathrm{p}(1)^{*},\mathrm{v}(2)\mathrm{p}(2)^{*},\ldots,\mathrm{v}\left(\mathrm{q}(\mathrm{ADA})\right)\mathrm{p}\left(\mathrm{q}(\mathrm{ADA})\right)^{*}$ sont respectivement constitués de $(\mathrm{S}+\mathrm{S}^{*\prime}+\mathrm{S}^{*\prime\prime})$ éléments distincts égaux ou supérieurs à 10. Autrement dit, nous supposons ici qu'un nombre suffisant d'emplois v est présent dans chaque ID de l'ADA avec données disponible et que le nombre d'emplacement spatiales disponibles dans l'AD n'est pas suffisamment important pour abaisser le ratio du nombre moyen d'emplois par emplacement spatiale en dessous de 10. Sinon, une distribution de Poisson multivariée (une normale multivariée asymétrique à droite se comprimant vers sa coordonnée d'origine multivariée) serait plus adaptée à certaines des distributions composant la distribution jointe, et la forme de la normale jointe pourrait être modifiée. 2) Sur le même sujet, un vecteur de probabilité p pourrait potentiellement être composé de termes de probabilité proches de zéro en raison du grand nombre de catégories. Cela pourrait remettre en cause la bonne approximation de la distribution normale. Cependant, nous comptons sur la disponibilité d'un grand v par ID lorsque l'emplois est disponible pour compenser le problème de rareté et rétablir la qualité requise de l'approximation normale. L'intuition ici est similaire à celle d'un jeu de données composé de plusieurs variables, où le nombre d'observations est suffisamment important pour traiter les nombreuses dimensions de l'information.

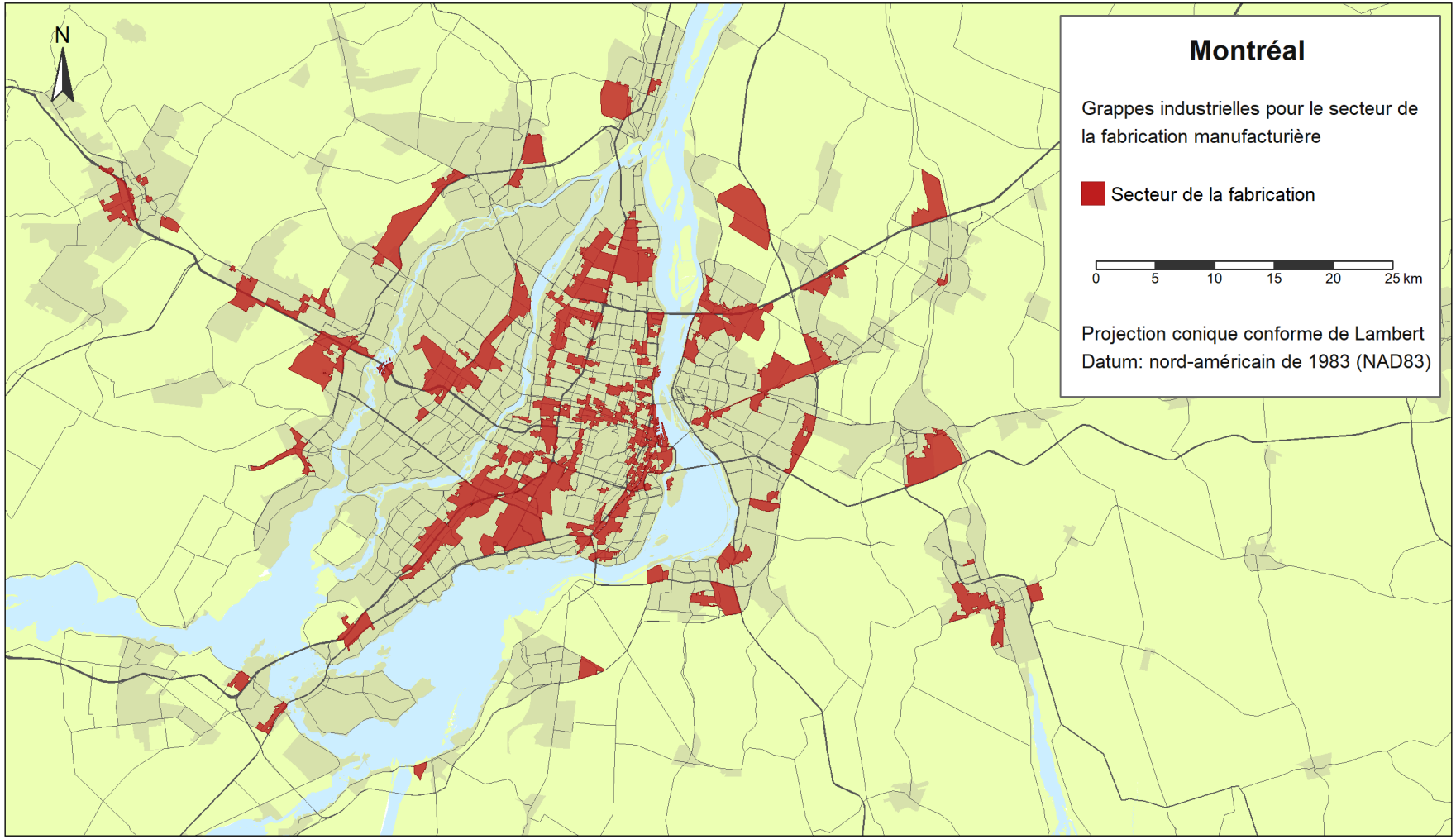
Dans le cas de notre propre application, si chaque ID de l'ADA avec données disponible alloue des points de données d'emplois provenant d'un seul code NAICS à deux chiffres au sein de l'AD, il est peu probable qu'au moins dix emplois parviennent à couvrir chaque zone spatiale de l'AD. Par conséquent, il est peu probable que des approximations de normales marginales précises soient produites. Il est également probable que la précision de la normale conjointe soit altérée, même en cas d'uniformité des probabilités pour l'ensemble de l'AD. Par conséquent, le résultat attendu ne sera pas une allocation uniforme des emplois au sein de l'AD. Il s'agira d'une allocation éparse, instable sur des réalisations aléatoires. La propriété de lissage des données sera perdue. Une bande passante de noyau plus importante sera nécessaire pour compenser l'instabilité et la rareté, et la précision locale des grappes sera perdue. Pour cette première raison, une allocation aléatoire au sein de l'ID semble préférable.

Une solution pourrait être y***. Avec y***, il est plus probable qu'une catégorie obtienne un plus grand nombre d'emplois attribués aléatoirement, car il s'agit de la somme des décomptes de tous les ID de l'ADA avec données disponible. Cependant, dans le cas de notre application, il est encore peu probable qu'un seul code NAICS à deux chiffres parvienne à atteindre 10 emplois par emplacement. Pour cette deuxième raison, une répartition aléatoire au sein de l'ID semble à nouveau préférable.

Une dernière solution alternative pourrait consister à agréger uniformément les zones spatiales de l'AD et à réduire le nombre de catégories de manière à ce que le nombre total d'emplois disponibles dans l'AD, divisé par le nombre de catégories agrégées, soit désormais égal à 10. Cependant, dans le cas de notre application, avec une industrie NAICS unique à deux chiffres, le nombre de catégories agrégées ne devrait pas réussir à être intéressant. Une approximation normale sera fiable et le résultat attendu sera une allocation uniforme et stable des emplois au sein de l'AD. Cependant, une bande passante de noyau plus importante sera nécessaire pour compenser le manque de haute résolution, et la précision locale des grappes sera perdue. Pour cette troisième raison, une allocation aléatoire au sein de l'ID semble préférable. Une approximation normale précise n'est pas toujours facile à obtenir. Cependant, dans le cas de nos applications, une allocation aléatoire au sein de l'ID reste notre meilleure option. Il est essentiel de documenter l'approximation normale dans cet article, car en cas de déviation, la meilleure façon de comprendre cette déviation est de comprendre d'abord de quel équilibre l'on s'écarte. ■

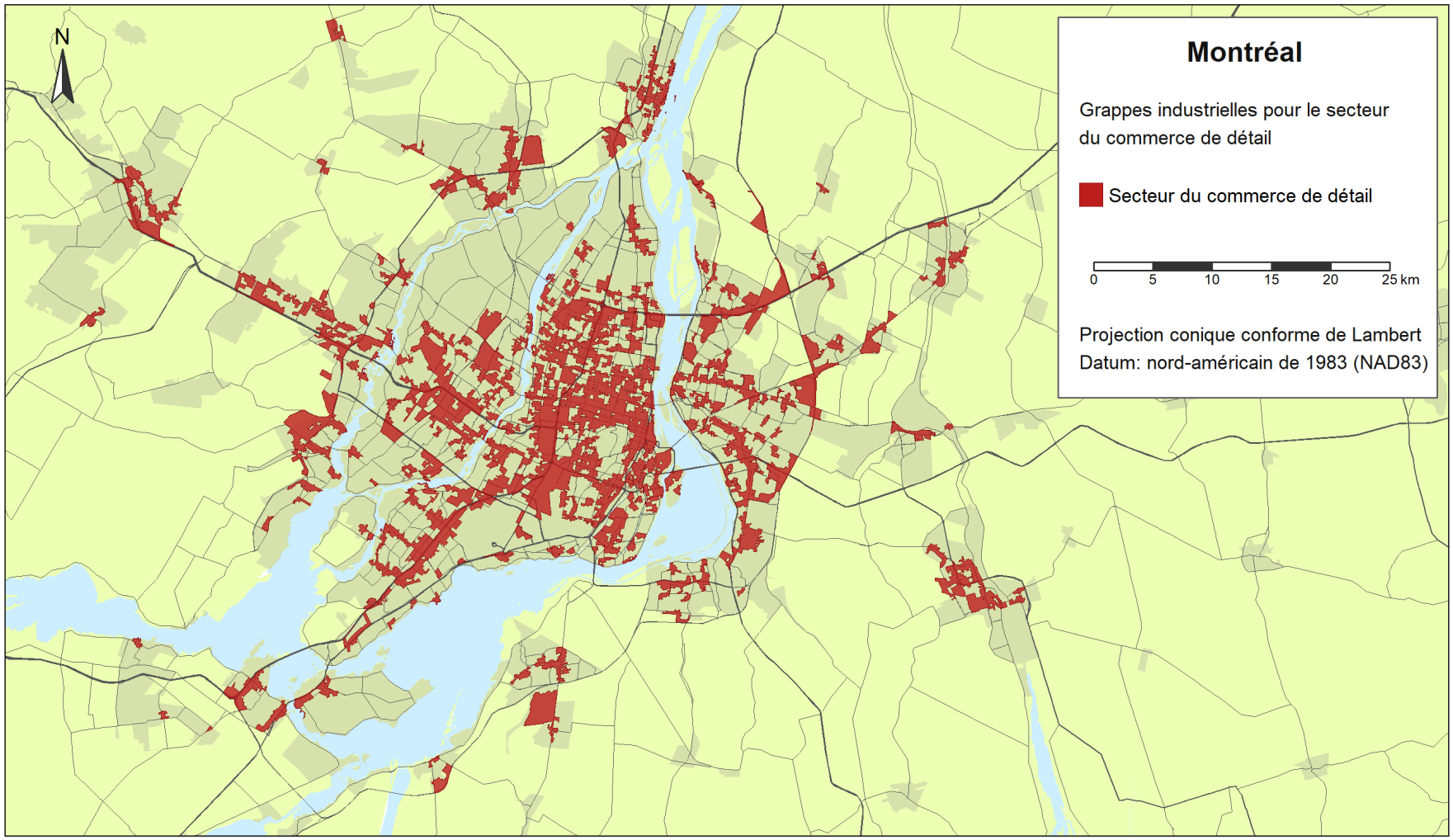
Annexe 3: Résultats de la cartographie des grappes

Carte 5
Secteur de la fabrication, Montréal



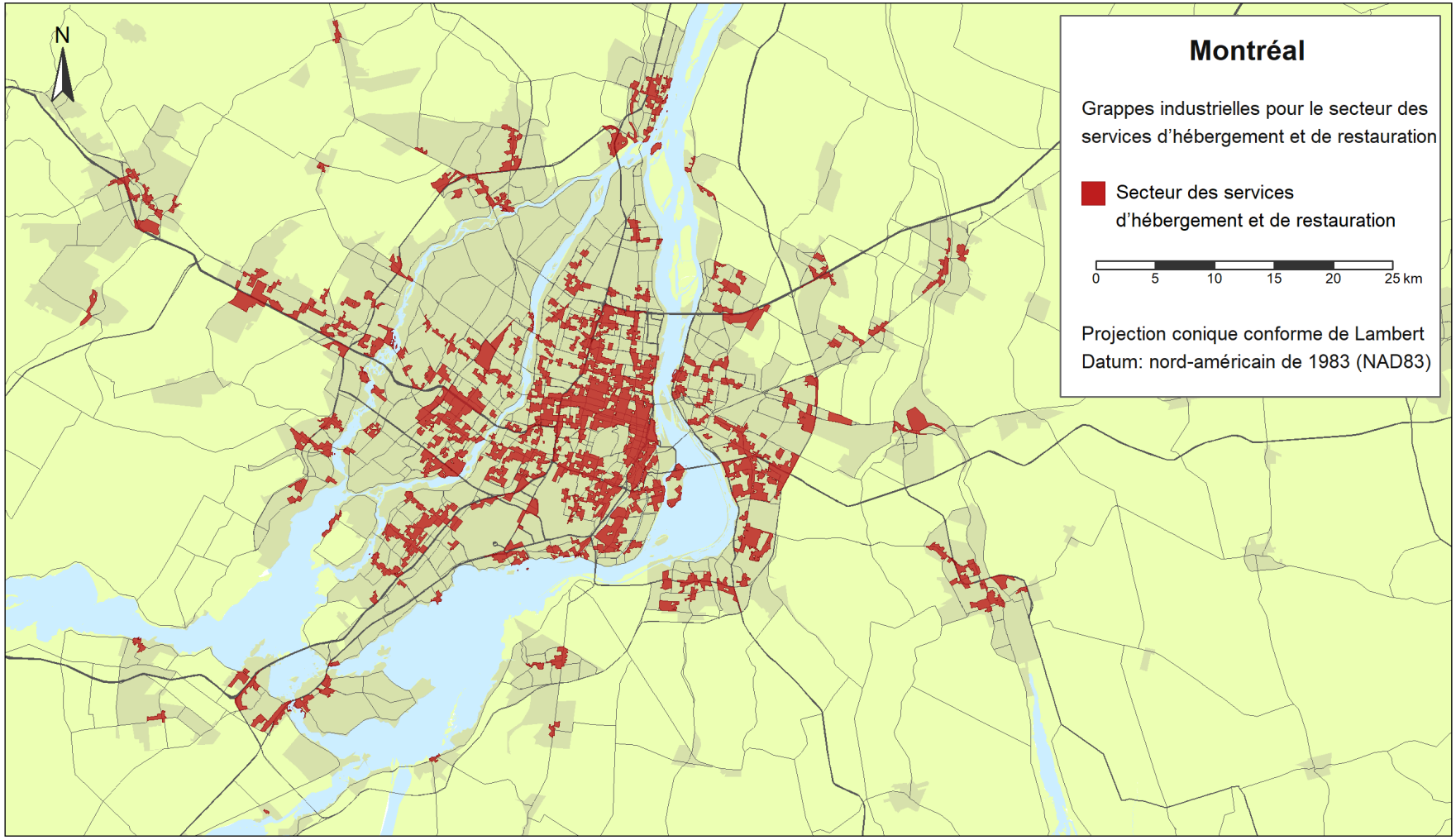
Note : 26 % des ID dans la RMR, contenant 90 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 6
Secteur du commerce de détail, Montréal



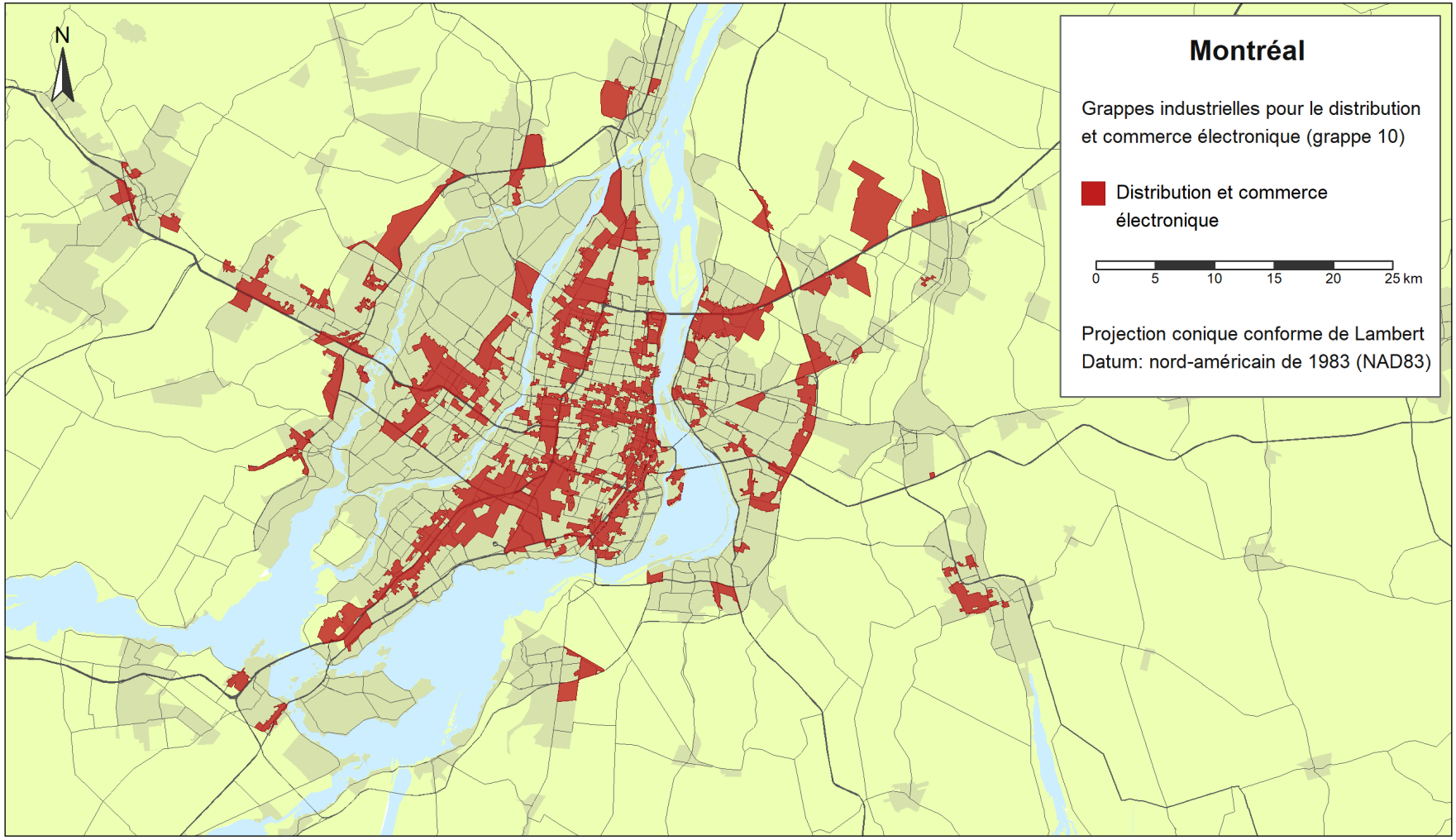
Note : 38 % des ID dans la RMR, contenant 92 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 7
Secteur des services d'hébergement et de restauration, Montréal



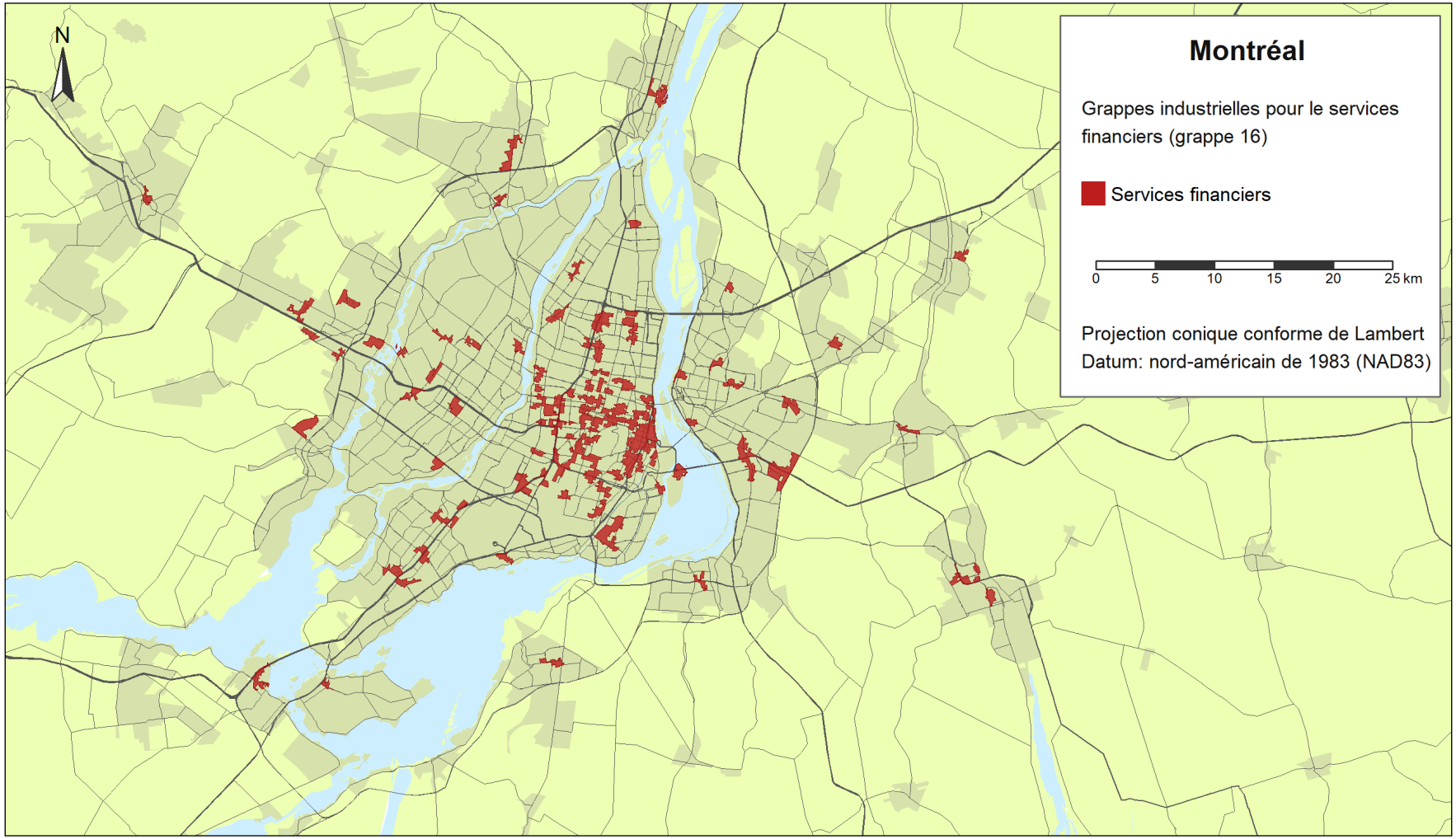
Note : 36 % des ID dans la RMR, contenant 85 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 8
Distribution et commerce électronique (grappe 10), Montréal



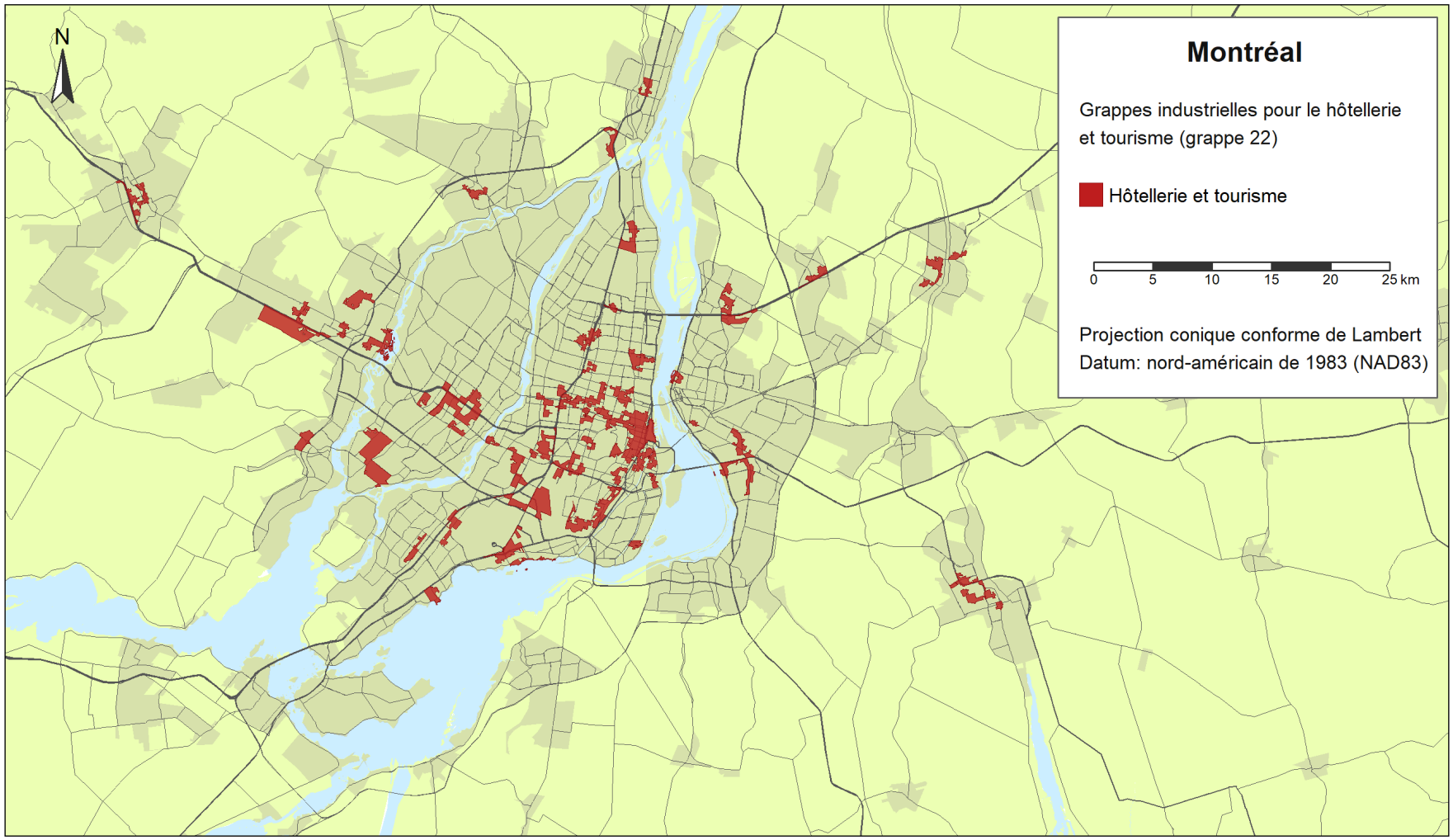
Note : 34 % des ID dans la RMR, contenant 95 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 9
Services financiers (grappe 16), Montréal



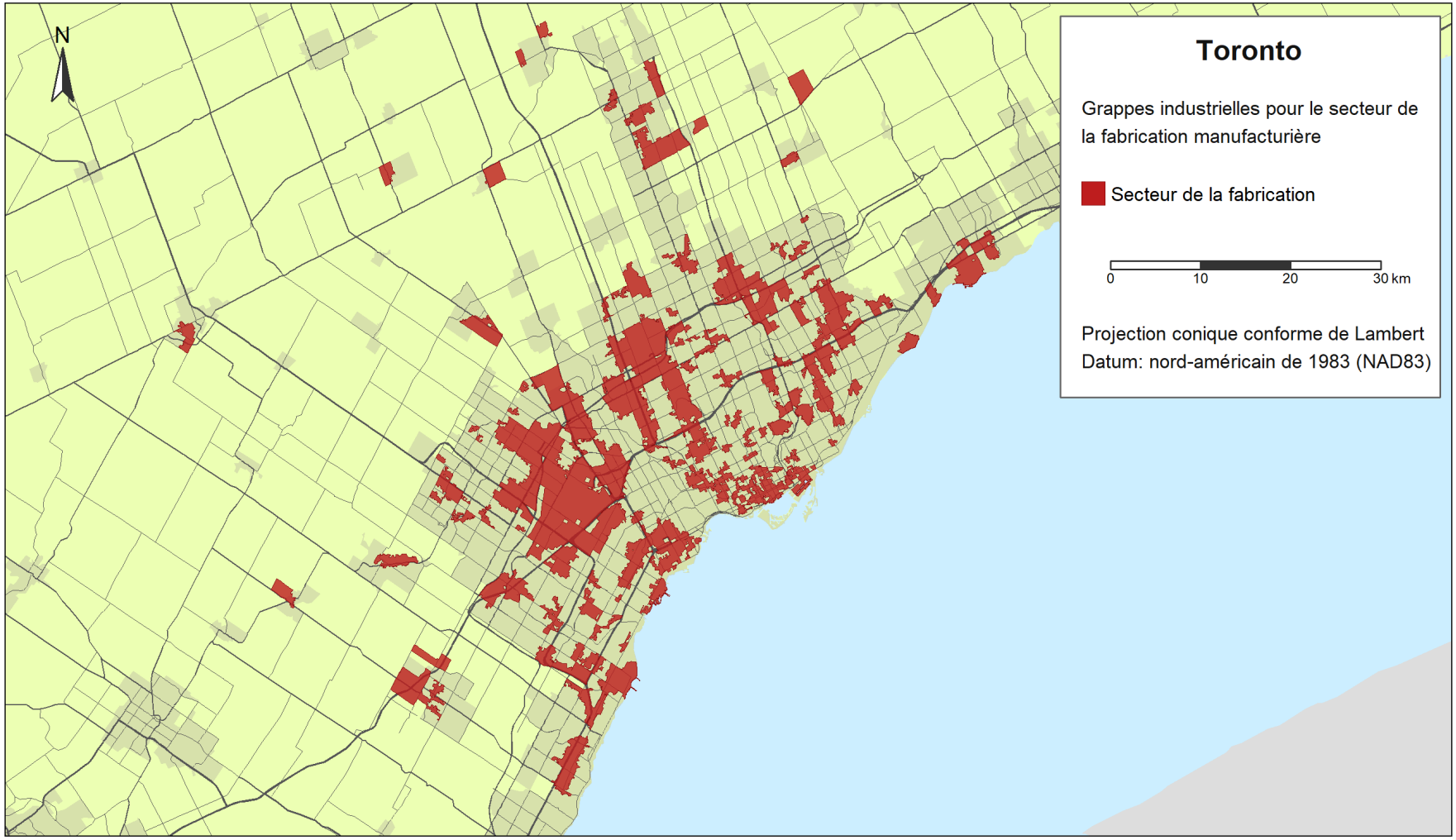
Note : 26 % des ID dans la RMR, contenant 89 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 10
Hôtellerie et tourisme (grappe 22), Montréal



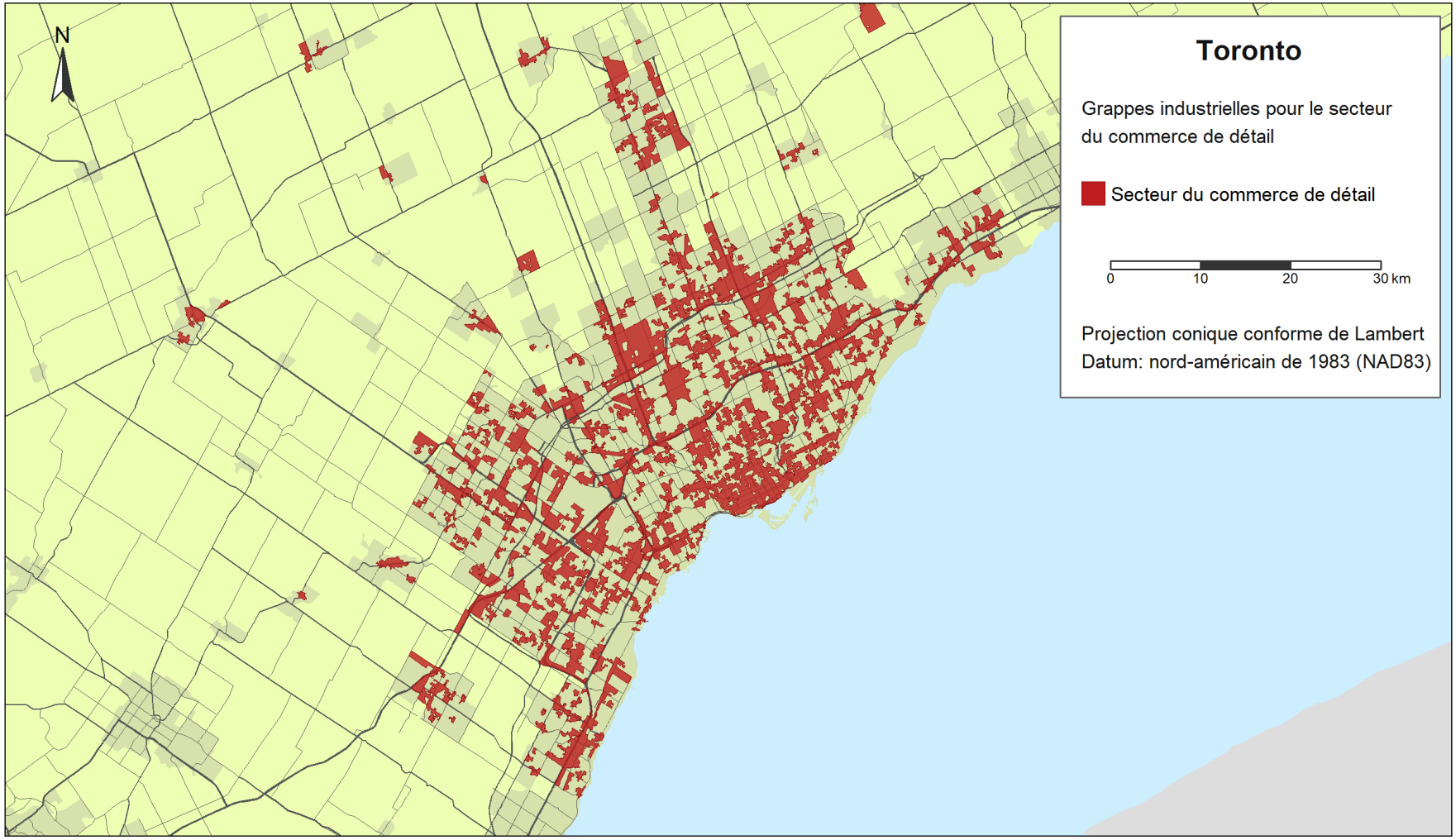
Note : 24 % des ID dans la RMR, contenant 60 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 11
Secteur de la fabrication, Toronto



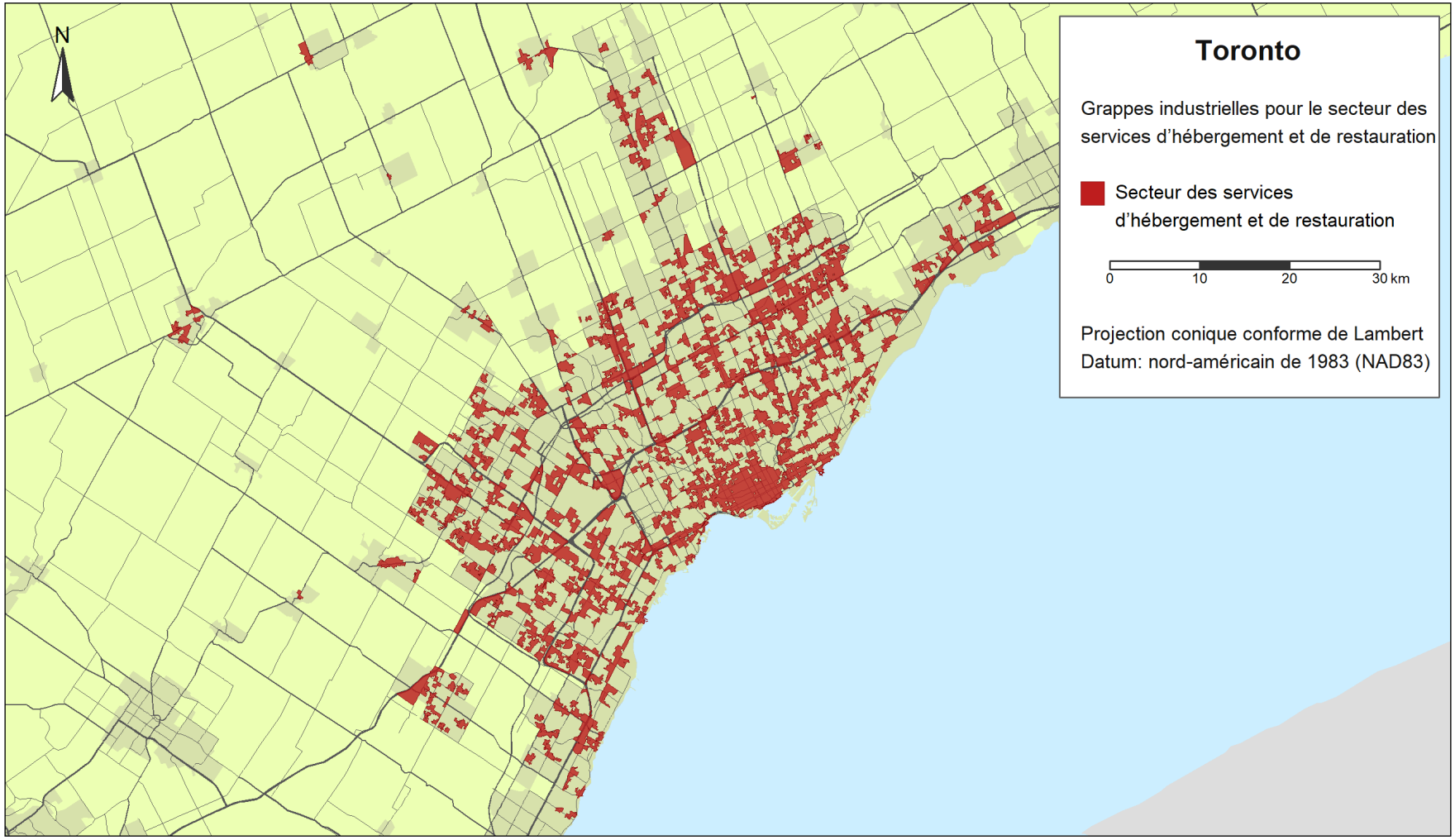
Note : 34 % des ID dans la RMR, contenant 94 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 12
Secteur du commerce de détail, Toronto



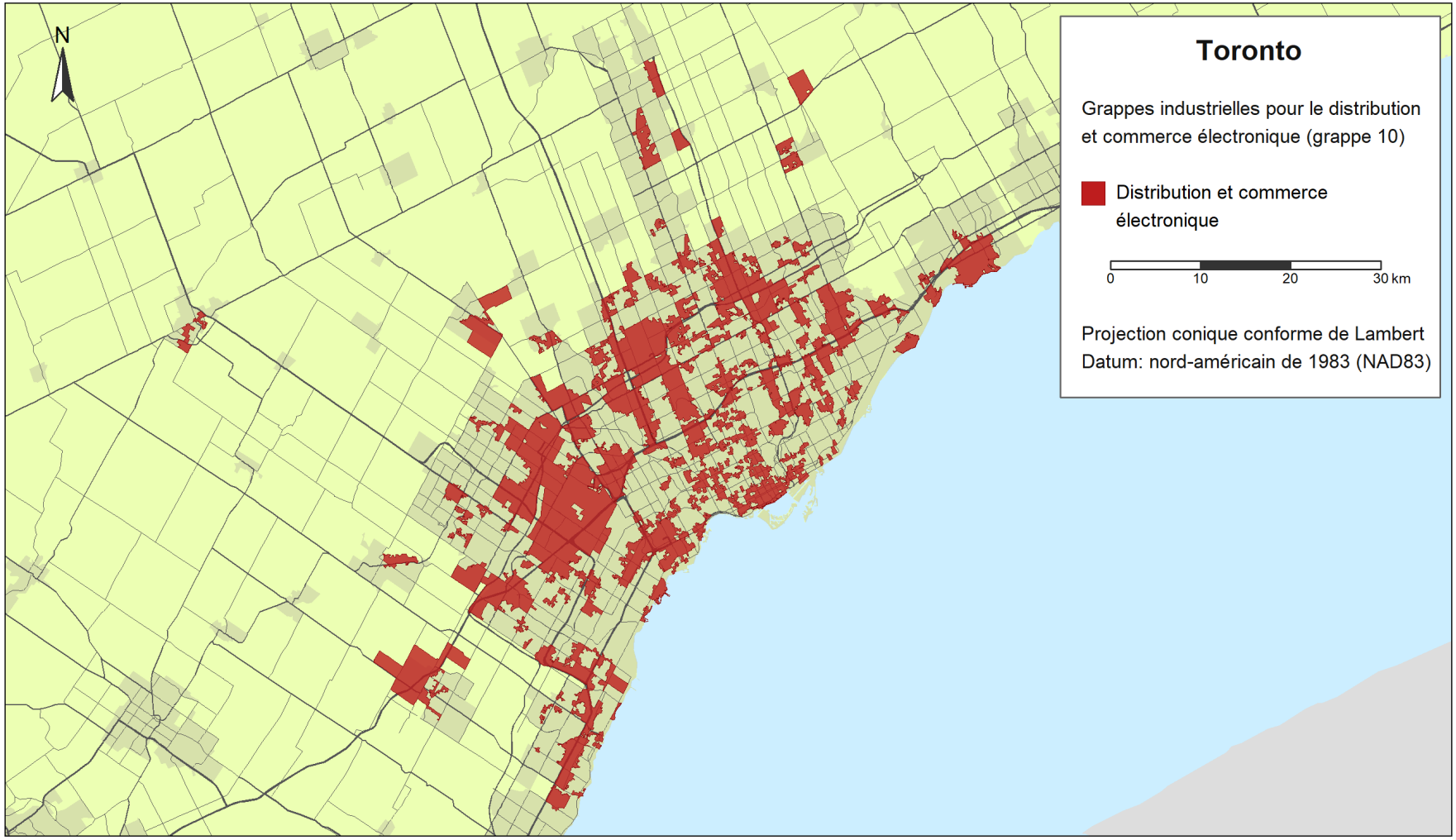
Note : 35 % des ID dans la RMR, contenant 97 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 13
Secteur des services d'hébergement et de restauration, Toronto



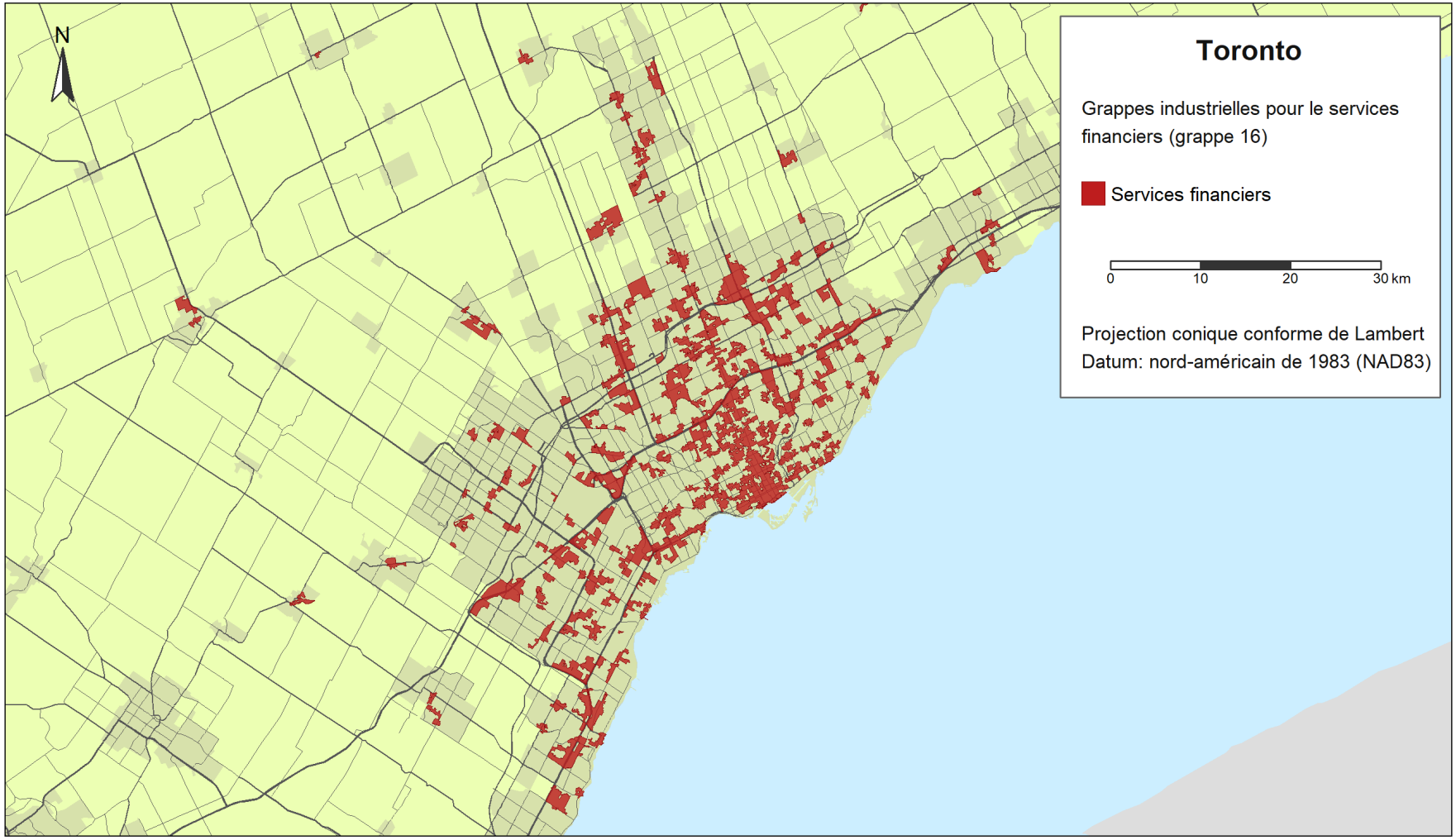
Note : 38 % des ID dans la RMR, contenant 89 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 14
Distribution et commerce électronique (grappe 10), Toronto



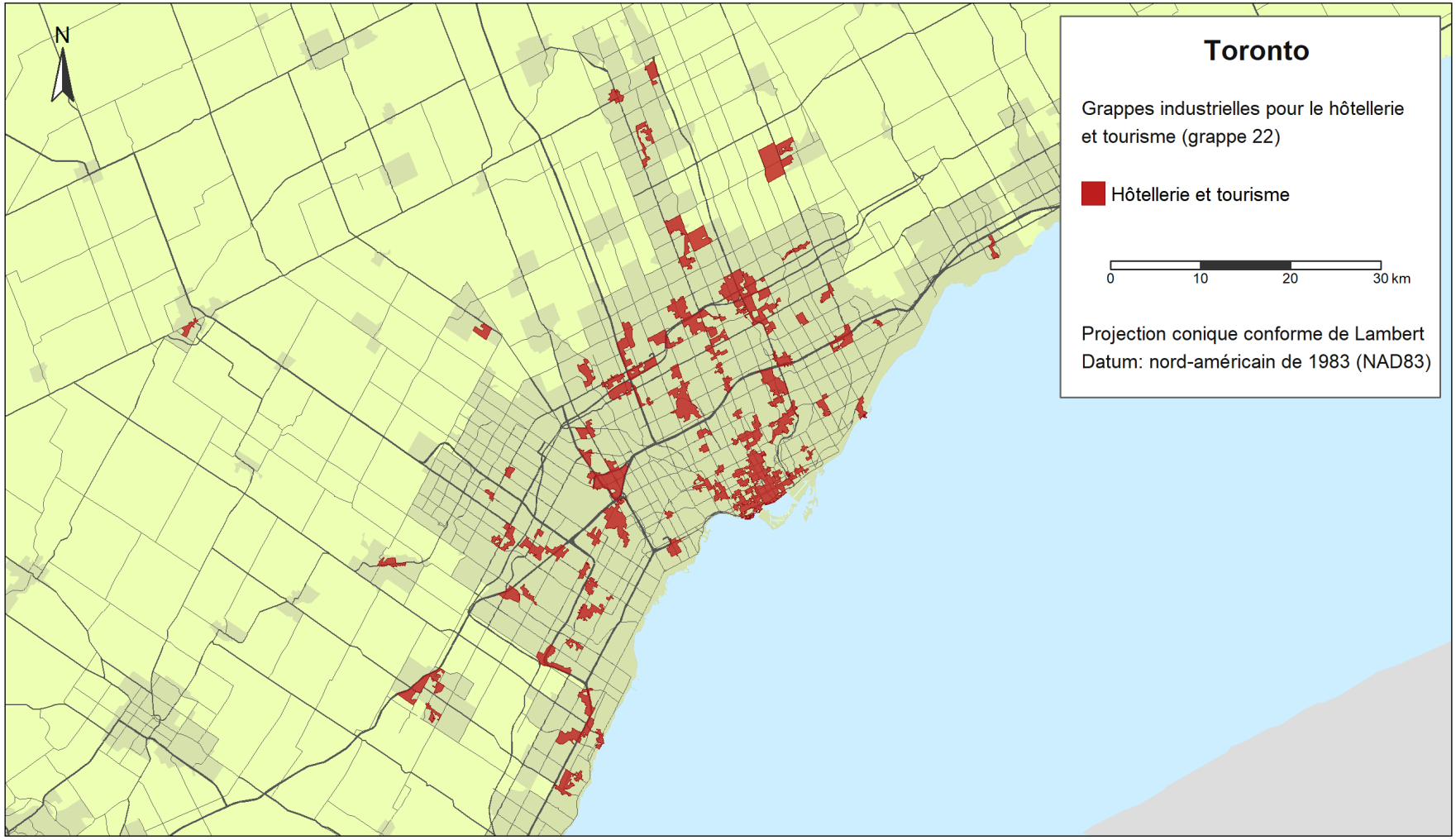
Note : 36 % des ID dans la RMR, contenant 98 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 15
Services financiers (grappe 16), Toronto



Note : 36 % des ID dans la RMR, contenant 96 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 16
Hôtellerie et tourisme (grappe 22), Toronto

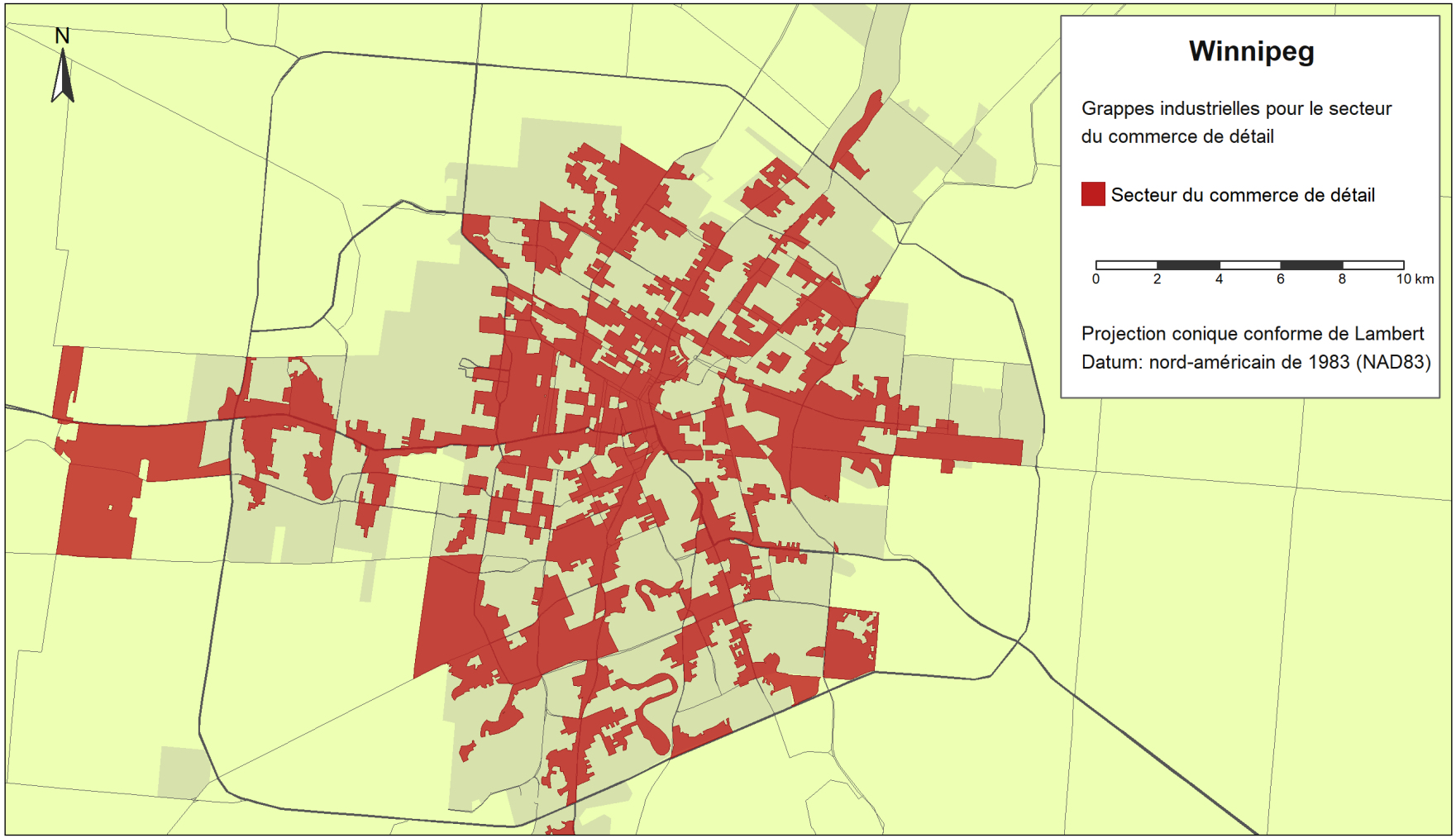


Note : 25 % des ID dans la RMR, contenant 71 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 17
Secteur de la fabrication, Winnipeg

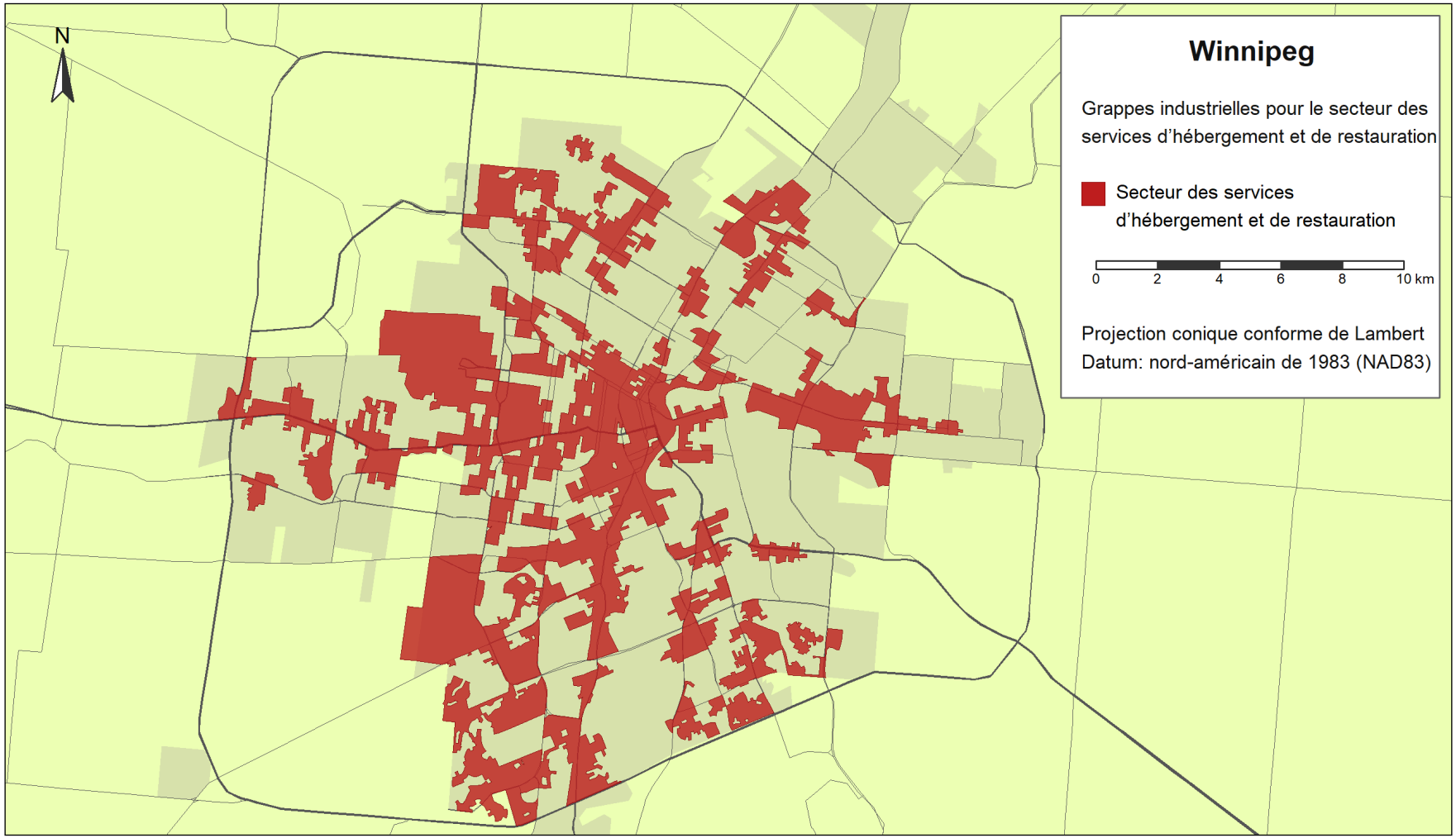
Note : 44 % des ID dans la RMR, contenant 91 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 18
Secteur du commerce de détail, Winnipeg



Note : 55 % des ID dans la RMR, contenant 94 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 19
Secteur des services d'hébergement et de restauration, Winnipeg



Note : 54 % des ID dans la RMR, contenant 91 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 20
Distribution et commerce électronique (grappe 10), Winnipeg

Note : 33 % des ID dans la RMR, contenant 93 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

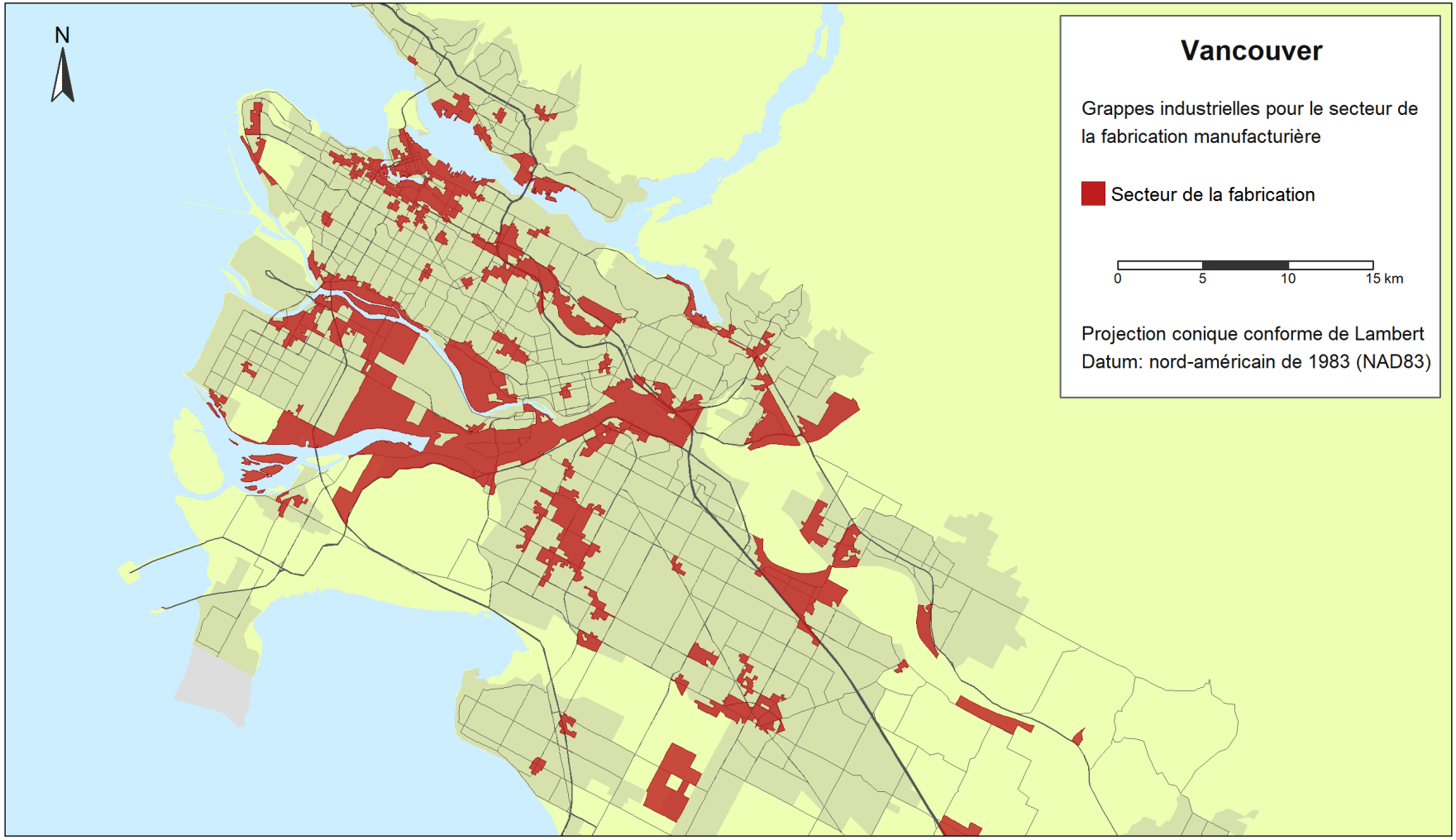
Carte 21
Services financiers (grappe 16), Winnipeg

Note : 28 % des ID dans la RMR, contenant 83 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 22
Hôtellerie et tourisme (grappe 22), Winnipeg

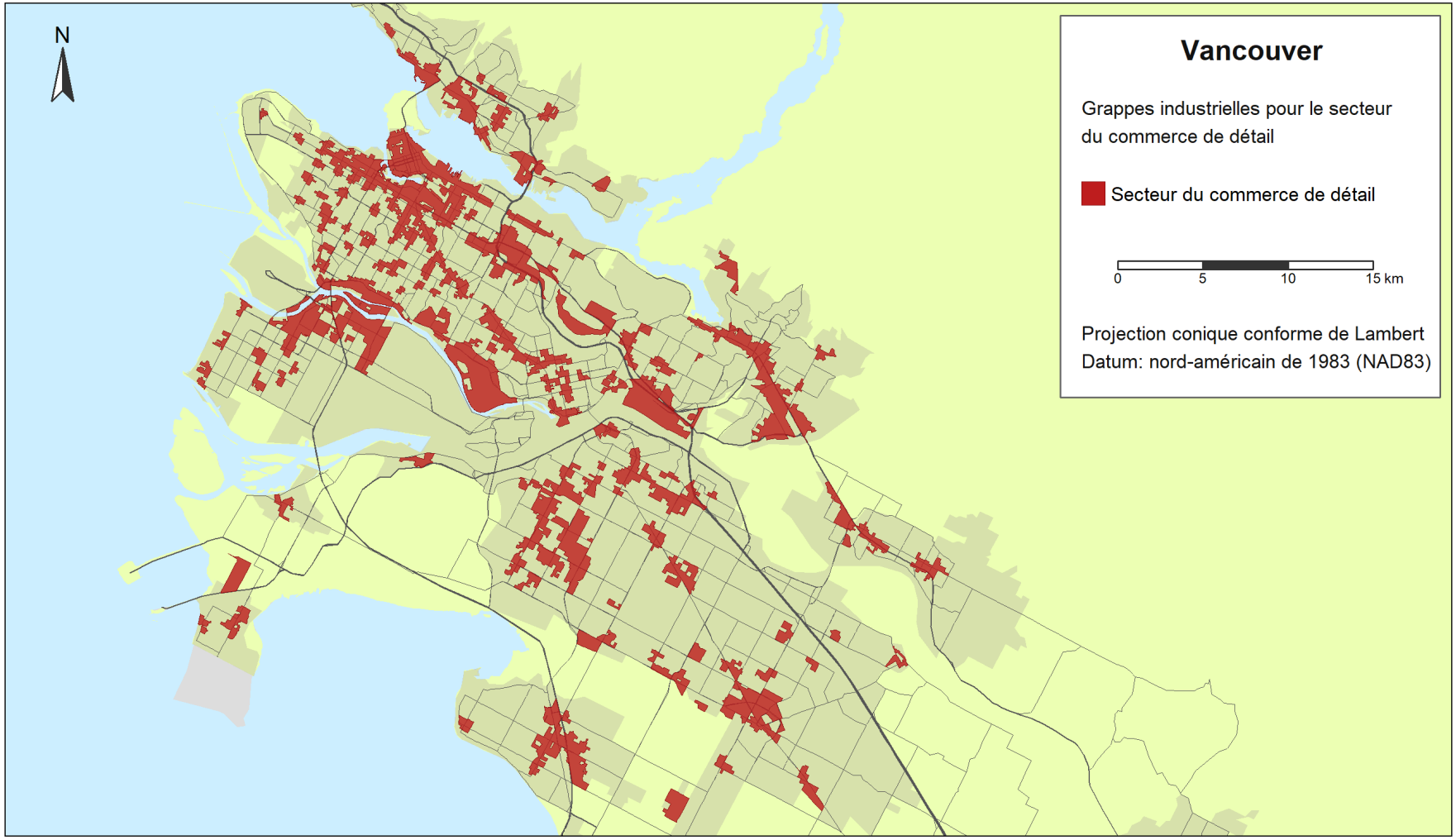
Note : 27 % des ID dans la RMR, contenant 60 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 23
Secteur de la fabrication, Vancouver



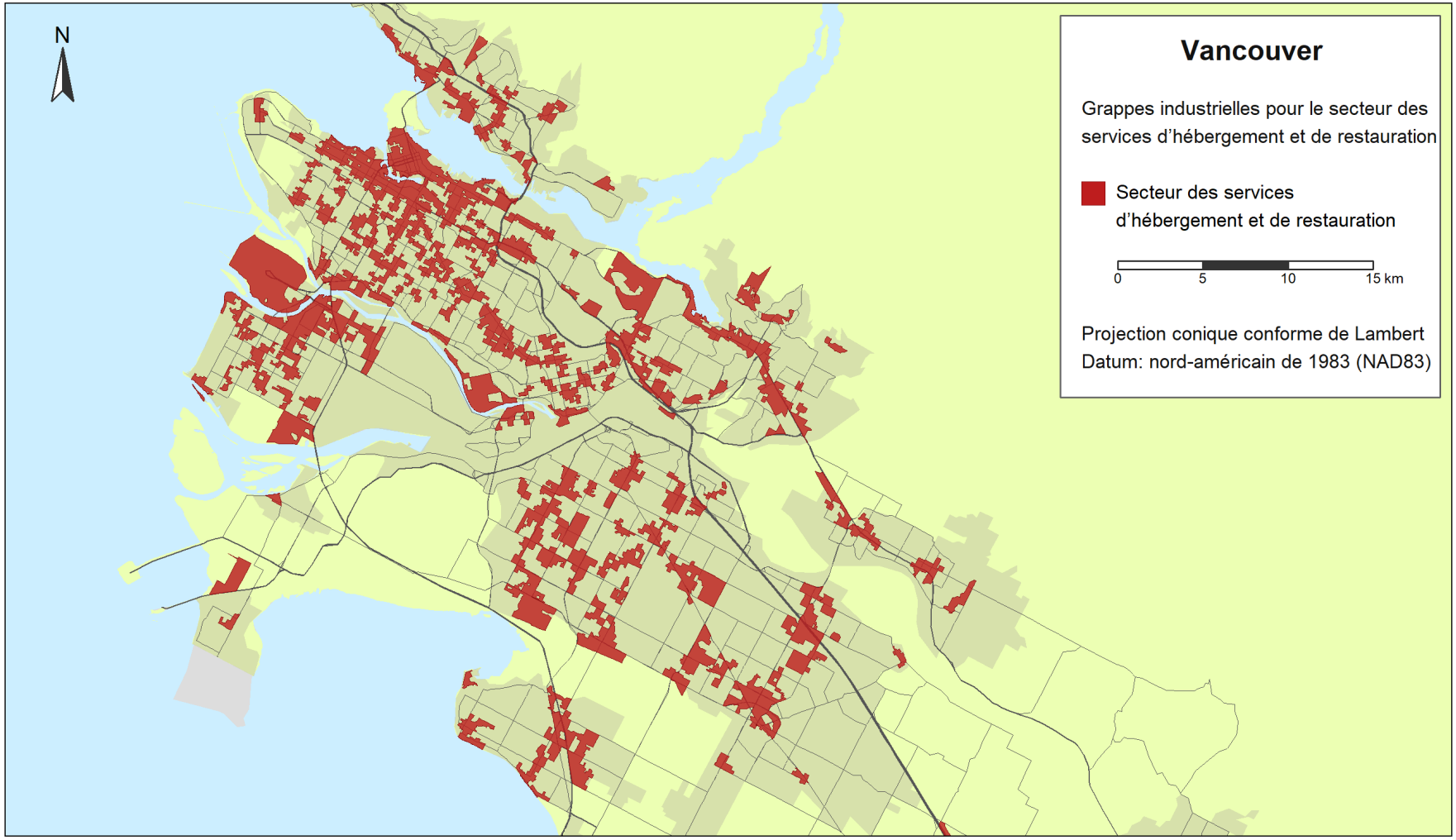
Note : 35 % des ID dans la RMR, contenant 90 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 24
Secteur du commerce de détail, Vancouver



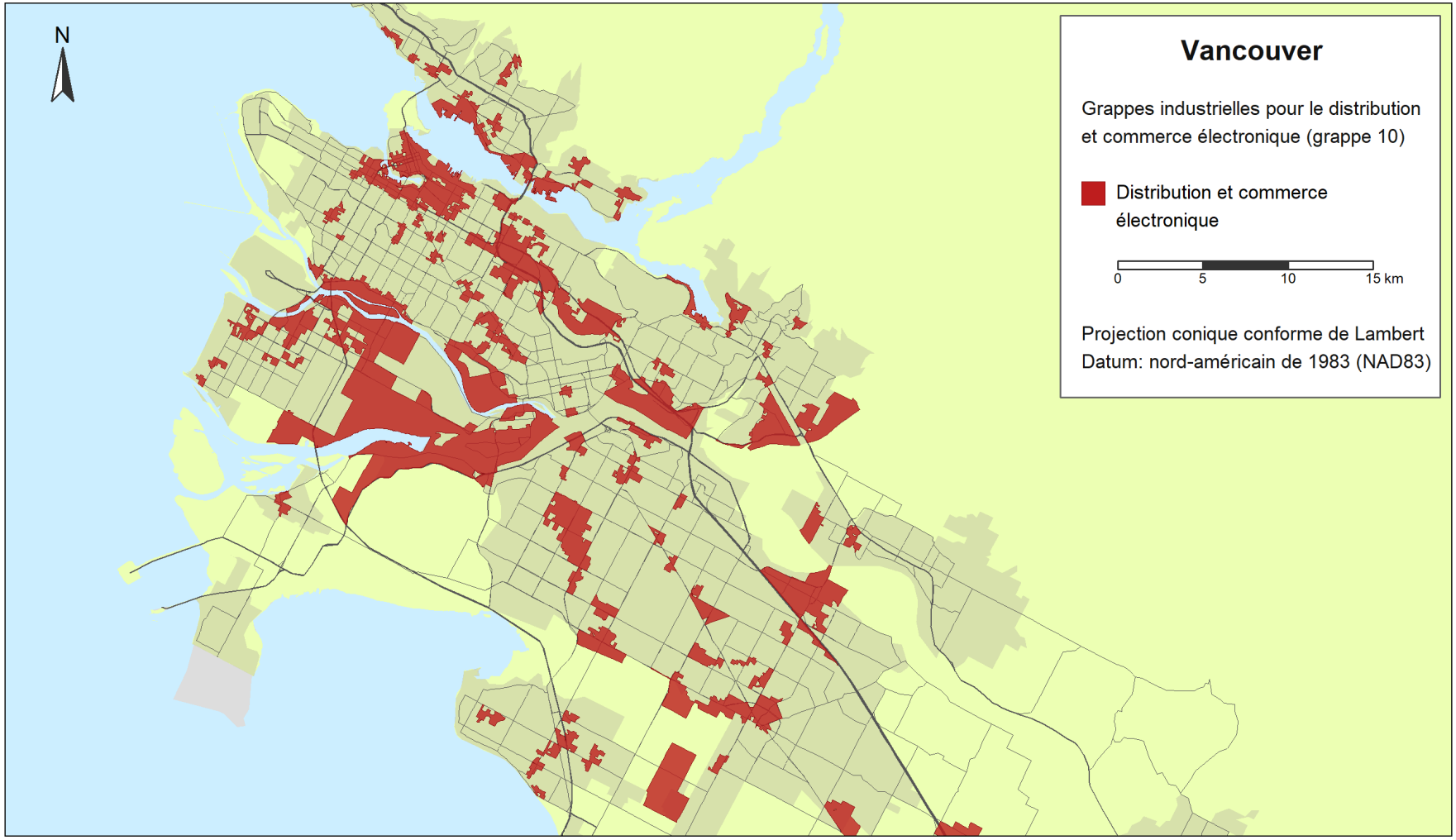
Note : 36 % des ID dans la RMR, contenant 81 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 25
Secteur des services d'hébergement et de restauration, Vancouver



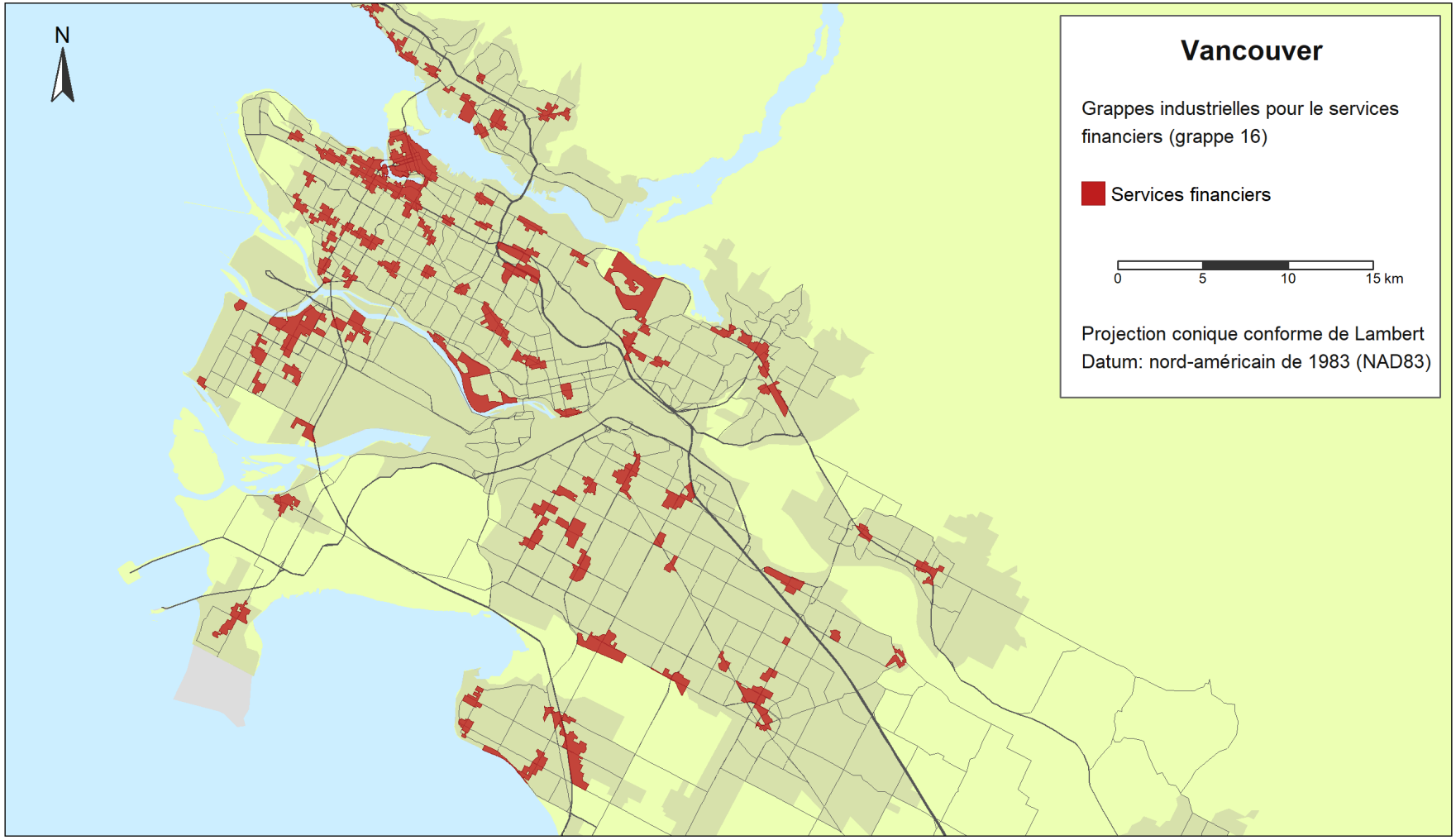
Note : 43 % des ID dans la RMR, contenant 91 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 26
Distribution et commerce électronique (grappe 10), Vancouver



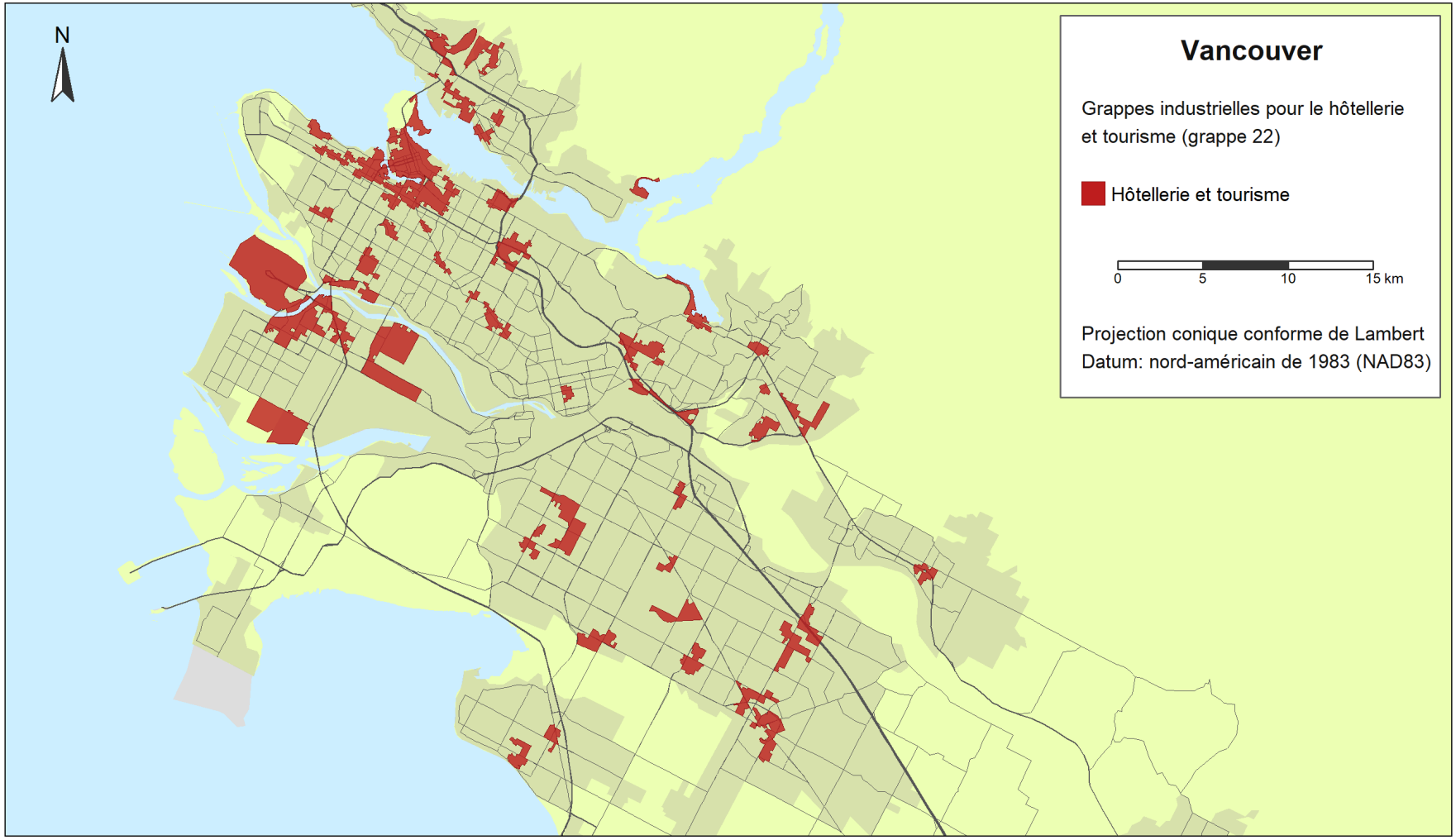
Note : 32 % des ID dans la RMR, contenant 94 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 27
Services financiers (grappe 16), Vancouver



Note : 32 % des ID dans la RMR, contenant 91 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.

Carte 28
Hôtellerie et tourisme (grappe 22), Vancouver



Note : 31 % des ID dans la RMR, contenant 80 % des employés de ce secteur d'activité
Sources : Statistique Canada, Recensement de 2021 - Provinces / territoires fichier des limites, Recensement de 2021 - Centres de population territoires fichier des limites, Recensement de 2021 – Fichier du réseau routier, Recensement de 2016 - Littoraux territoires fichier des limites, Recensement de 2016 - Lacs et rivières territoires fichier des limites et calculs des auteurs à partir de la base de données sur le RE.