`LUM_OSM` : une plateforme pour l'évaluation de la mixité urbaine à partir de données participatives

Luciano GERVASONI[*,**], Martì BOSCH[***]
Serge FENET[*,****] Peter STURM[*,**]

[*]Inria Grenoble – Rhône-Alpes, France
[**]Univ. Grenoble Alpes, Lab. Jean Kuntzmann, Grenoble, France
[***]École Polytechnique Fédérale de Lausanne, Lausanne, Suisse
[****]Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France

Résumé. L'augmentation de plus en plus rapide des populations urbaines depuis les dernières décennies, en conjonction avec l'accessibilité croissante de données urbaines massives, motivent de plus en plus de recherches portant sur la soutenabilité du développement urbain.

Nous présentons dans cet article une plateforme d'analyse et de construction de cartes de mixité urbaine afin d'aider les planificateurs urbains à évaluer des indices de mixité. Dans ce premier travail, nous nous focalisons sur les usages résidentiels et d'activités, qui sont extraits des données collaboratives d'OpenStreetMap. Ils servent à la construction d'une représentation continue de l'indice d'entropie, mais peuvent être généralisés à tout indice. La méthode est appliquée à plusieurs villes, et les premiers résultats sont présentés.

1 Introduction

La population vivant dans des centres urbains a augmenté de façon considérable depuis les années 50, passant de 746 millions à 3,9 milliards en 2014 (United Nations (2014)). Les projections actuelles prévoient que plus de 66% de la population mondiale vivra dans des villes d'ici 2050, contre 30% en 1950. En parallèle avec un léger déclin de la population rurale globale dans les 35 prochaines années, l'augmentation de l'urbanisation devrait ajouter près de 2,5 milliards de personnes dans les centres urbains d'ici 2050. Cette situation, qui impose de nouveaux défis liés à la conception de villes capables d'héberger de telles populations de manière soutenable, nous force à penser les villes du futur sous des aspects économiques, sociaux et environnementaux.

Nous nous concentrons dans cet article sur la formalisation de la mesure de développement mixte et de mixité urbaine (LUM pour *land use mix*) dans les villes, afin notamment d'étudier son impact sur la soutenabilité des villes. La mixité urbaine décrit la coexistence dans un rayon localisé de différents usages du sol urbain (habitations, lieux de travail, lieux de consommation et de loisirs, etc.). Elle est donc associée à une certaine compacité, mais aussi à des aménagements rendant la vie dans les villes plus agréables et permettant de limiter la consommation de carburants fossiles (de par la présence de parcs, de jardins ouvriers ou de pistes cyclables, par exemple).

La soutenabilité urbaine décrit la capacité de la structure urbaine à maintenir son intégrité dans le temps sans épuiser les ressources qui lui permettent d'exister, tout en assurant un service de qualité aux habitants. Il a déjà été prouvé qu'une forte mixité urbaine apporte de nombreux bénéfices en termes de soutenabilité, et les structures urbaines denses et compactes ont été largement étudiées dans la littérature en terme de soutenabilité de développement (Gordon et Richardson (1997); Breheny (1995); Burgess (2000); Neuman (2005)). Bien que le degré nécessaire de compacité reste une question ouverte, un accord large existe sur les impacts positifs de la mixité urbaine, non seulement en termes de soutenabilité de développement, mais aussi en termes sociétaux, de santé publique et d'infrastructure de transports publics.

Afin d'aider les planificateurs urbains, nous proposons une plateforme logicielle permettant de capturer l'état courant de mixité urbaine d'une ville à l'aide de données collaboratives publiques. Cet outil permet de construire une carte de type SIG (système d'information géographique) contenant le degré de mixité de tout point de la ville, avec une résolution arbitraire. L'usage des sols associé à chaque polygone est d'abord extrait des données OpenStreetMap en entrée. Ensuite une représentation spatiale de chaque usage est calculée à l'aide de noyaux statistiques. Enfin, ces représentations génériques sont utilisées pour calculer l'indice final. Même s'il est souvent désirable d'atteindre un degré de mixité le plus élevé possible, nous n'effectuons pas ici d'interprétation numérique de la valeur optimale nécessaire à un développement soutenable. Nous nous focalisons plutôt sur la méthode de calcul d'une représentation spatialisée continue de la carte de mixité, et permettant de calculer n'importe quel indice (nous l'appliquons ici à l'indice d'entropie) pour peu qu'il soit basé sur des données présentes dans la base OpenStreetMap.

Cet article est organisé de la manière suivante. Nous présentons dans la section 2 une vue d'ensemble du contexte de ces travaux. La section 3 présente une rapide analyse des indicateurs de mixité existants. Nous argumentons dans la section 4 l'importance des données collaboratives urbaines dans le cadre des études de mixité, et nous présentons notre plateforme dans la section 5. Finalement, nous illustrons dans la section 6 son application à plusieurs villes, avant de conclure en présentant les travaux futurs envisagés.

2 Contexte

2.1 Motivation

Un motif migratoire évident des zones rurales vers des centres urbains de taille croissante peut être observé dans l'histoire récente, et ce phénomène devient de plus en plus important depuis la fin du 19ème siècle. Ce processus, présent de plus dans tous les pays, a impliqué une augmentation constante de la taille des populations urbaines : plus de la moitié de la population humaine vit maintenant dans les villes, et ces dernières font face à des problèmes croissants pour continuer à assurer aux habitants les services essentiels et une bonne qualité de vie. Cette concentration impose donc de nouvelles façons de concevoir les villes, ces systèmes complexes dynamiques avec de nombreux niveaux d'interaction, de manière à assurer à la fois leur soutenabilité et une consommation efficace des ressources.

Cependant, plus récemment, l'accessibilité grandissante de données urbaines parfois massives rend possible l'analyse des interactions entre les habitants et l'infrastructure de leur ville. Ainsi, la compréhension des boucles de rétroaction complexes liant les citoyens et le

développement de leur cité pourrait nous permettre de résoudre plus efficacement les questions de soutenabilité future. En conséquence, des efforts considérables et récents visent à comprendre les conditions de la soutenabilité urbaine en se basant sur ces nouvelles sources de données.

Dans ce contexte, le présent travail a été motivé par trois observations :

– Tout d'abord, et alors que les questions de soutenabilité deviennent un sujet central dans plusieurs domaines scientifiques, nous pensons que les outils numériques, et plus spécifiquement la construction de modèles numériques à partir de processus de fouille de données issues du monde réel, peuvent aider à mieux comprendre la dynamique des systèmes socio-économiques complexes. Nous espérons que ces outils permettront de rendre plus efficaces les décisions prises pour améliorer la soutenabilité et la résilience des villes du future.
– Ensuite, plusieurs indicateurs pour mesurer la mixité urbaine existent (voir Song et al. (2013)). Ils peuvent tous se catégoriser en mesures divisionnelles ou intégrales ; les premières étant plus expressives que les secondes en délivrant une valeur pour chaque zone considérée. Cependant, ils sont tous sensibles à l'échelle d'agrégation utilisée pour calculer l'indicateur. Nous pensons qu'un bon indicateur devrait être indépendant de l'échelle de l'analyse.
– Enfin, même si les données participatives voient leur quantité et leur qualité croître régulièrement, il y aura toujours des données manquantes pouvant impacter les mesures sus-citées. En conséquence, il est important d'utiliser des outils statistiques permettant de les prendre en compte afin de mieux capturer les structures urbaines sous-jacentes.

2.2 Brefs rappels historiques

Avant le 20ème siècle, les usages urbains mixtes étaient une tendance naturelle étant données les contraintes de proximité géographique des activités journalières imposées par la rareté des possibilités de transport. À partir du début du 20ème siècle, d'abord aux États-Unis, les pratiques de zonage commencèrent à assigner des usages uniques aux différentes parcelles, induisant une ségrégation nouvelle entre les zones d'habitation et les zones d'activités, et imposant un usage intensif de la voiture pour se déplacer. Cette tendance est devenue particulièrement forte des années 1910 aux années 1950, où les zones mixtes étaient devenues rares, et où le développement ségrégé était la norme. Quelques années plus tard, à partir des années 1960 à 1970, et après avoir été négligés pendant des décennies, les usages mixtes réapparurent, apportant de nouveau leurs avantages dans plusieurs aspects de la société.

Plus récemment, depuis le début des années 1990, le développement mixte a ré-émergé en tant que concept essentiel de la planification urbaine. À titre d'exemple, le *Congress for the New Urbanism* (Leccese et McCormick (2000), `http://www.cnu.org`) milite pour des voisinages "pedestrian-friendly, and mixed-use" et le *Smart Growth Network* (Smart Growth Network (2006)) inclut les usages urbains mixtes dans ses 10 principes. Le concept de *Transit-Oriented Development* et plusieurs grandes agences de transport supportent de leur côté les usages mixtes (Cervero et al. (2002)). Afin d'atteindre de hauts niveaux de développements mixtes, tous ces acteurs recommandent à la fois l'intensité et la diversité des usages des sols urbains, ainsi que l'intégration des zones auparavant ségrégées.

2.3 Les impacts des usages urbains mixtes

Dans cette section, nous décrivons rapidement les principaux effets liés à la soutenabilité des villes en présence des différents type de développements. Les relations entre soutenabilité et mixité des usages, qu'elles soient directes ou indirectes, ont été étudiées de manière intensive dans la littérature :

- *Étalement urbain :* Le concept d'étalement urbain fait référence au processus d'extension spatiale de la population s'éloignant des zones centrales pour former des communautés satellites éparses, mono-fonctionnelles, et souvent très dépendantes de la voiture. La sub-urbanisation, c'est-à-dire la création de zones de banlieues éloignées des centre-villes, s'est accélérée durant tout le dernier siècle. En conséquence, une faible mixité urbaine est une des caractéristiques distinctives associées au phénomène d'étalement urbain, qu'il soit résidentiel ou industriel.
  Dans la littérature, l'étalement urbain a été largement associé à des effets négatifs en termes d'environnement, de santé, de relations sociales et d'économie (Squires (2002)). Plus récemment, et en conjonction avec l'apparition du concept de développement soutenable, les planificateurs urbains ont commencé à prendre de plus en plus en compte les conséquences de l'étalement urbain (Johnson (2001)), même si ces conséquences négatives avaient déjà été identifiées au début des années 1970.
  Le manque de planification coordonnée de l'usage de sols et ses conséquences négatives ont commencé à clairement apparaître comme des composants clés de l'étalement urbain (Nelson et Duncan (1995); Ewing (1995); Sierra Club (1998); United States Department of Housing and Urban Development (1999)).
- *Transport :* En termes de transports publics, les zones étalées ne sont pas soutenables lorsqu'elles sont associées à des indices de mixité faibles (Cervero et al. (2002)). En effet, des zones étalées et largement dispersées imposent de se déplacer sur de grandes distances pour des faibles demandes, abaissant l'efficacité des infrastructures de transport.
- *Santé :* Il a été montré que la mixité d'usages a un impact fort sur les activités physiques des habitants (Christian et al. (2011)). Cela implique que la conception *a priori* de la structure urbaine peut directement impacter l'activité et le bien-être physique des habitants, ce qui constitue actuellement un point important lié à la santé des habitants urbains : un taux de mixité élevé est associé à des déplacements à pied plus importants (Christian et al. (2011)) et implique une meilleure santé et une baisse de l'obésité (Frank et al. (2004)).
- *Dépendance à la voiture :* Le type de déplacement, en considérant le mode et la distance à parcourir, a été fortement corrélé au type de mixité urbaine dans Cervero (1996); Kockelman (1997). En effet, la présence de zones à usages commerciaux proches a été associée à la fois à des temps de parcours raccourcis et à une faible nécessité de posséder un véhicule personnel (Cervero (1996)), ce qui implique un impact environnemental positif en termes d'émission de polluants. Les études de Newman et Kenworthy (2006) sur la conception de structures urbaines réduisant la dépendance à la voiture ont déterminé qu'une dépendance plus basse implique d'avoir une valeur de densité urbaine (i.e. nombre d'emplois et de résidences par hectare) au delà d'un seuil minimal, ce qui est directement relié à la notion de mixité urbaine élevée.

3 Les limites des indicateurs existants

Les propositions de méthodes pour mesurer la mixité urbaine abondent dans la littérature. Une revue presque exhaustive (Song et al. (2013)) examine notamment quelles sont les forces et les faiblesses de chacune. La pertinence d'une mesure de mixité donnée repose directement sur deux aspects. D'abord, elle dépend du contexte d'application, dans lequel l'usage prévu pilote fortement les caractéristiques de la mesure. Ensuite, les données utilisées pour la calculer conditionnent directement sa performance : la présence ou l'absence de bruit, le biais impliqué par la source de données, leur quantité, leur ancienneté, sont autant d'aspects à prendre en compte.

Les mesures de mixité urbaine sont largement inspirées des métriques d'écologie du paysage et des statistiques spatiales (Raines (2002)). Plusieurs indices, tels que l'indice d'Atkinson, l'indice de partitionnement, l'indice de dissimilarité, l'indice d'exposition ou l'indice de Gini ont été proposés. Par ailleurs, plusieurs métriques sont basées sur la définition de l'entropie de Shannon.

Les données utilisées pour calculer ces indices sont rarement discutées. Elles sont généralement obtenues lors de l'étude d'une certaine région d'intérêt, et sont donc souvent fortement biaisées par la disponibilité des données relatives à cette région. D'un côté, de nombreux jeux de données sont privés et réduisent fortement la reproductibilité des résultats. D'un autre, certains jeux de données spécifiques existent, mais sont bien souvent restreints à certaines zones géographiques [1]. Il est bien connu que le résultat de toute méthode numérique est fonction des données en entrée, et l'utilisation de données hétérogènes liées aux multiples zones d'application ont pu favoriser le développement de mesures *ad hoc* qui exhibent un comportement désiré pour le jeu de données en entrée, tout en limitant l'application possible à cette seule région d'analyse.

De nombreux travaux de la littérature mettent en avant le problème de l'unité de surface modifiable ("modifiable areal unit problem" ou MAUP), qui se pose naturellement lorsqu'on agrège des données spatiales. Les sorties de la mesure dépendent du niveau d'agrégation et de l'unité de division géographique choisie. La sensibilité des résultats à ces choix constitue un gros inconvénient, indésirable dans toutes les applications, et se révèle être un vrai trouble lorsqu'il s'agit de comparer plusieurs métriques. Dans notre approche, l'utilisation de données géo-localisées, couplées à l'estimation d'une surface continue en utilisant une estimation de noyau de densité permet de contourner le problème du MAUP (Carlos et al. (2010)).

Le présent travail ne cherche pas a définir un nouvel indicateur qui serait meilleur que tous les indicateurs existants, mais propose une méthode et une plateforme pour en calculer un automatiquement, à une résolution spatiale quelconque, et à partir de données publiques. Nous évaluons notre outil en calculant une représentation SIG de l'indice d'entropie. Cela nous permet notamment de montrer qu'il est possible de dépasser la limitation séparant les indices intégraux et divisionnels (nous calculons la version géographiquement continue d'un indicateur divisionnel). Muni d'un tel outil, nous pourrons dans un second temps nous concentrer sur le calcul d'un indicateur optimal à l'aide de méthodes d'apprentissage supervisé, et comparer ce dernier aux indicateurs existants.

1. Bien que l'émergence des données nationales ouvertes (Open Data) soit en train de changer cela, tout au moins en Europe.

4 Données utilisées

OpenStreetMap [2] (OSM) est un projet collaboratif portant sur la création d'une carte éditable et ouverte du monde ; il s'agit d'un exemple de "volunteered geographic information" (VGI). OSM fournit, entre autres, des cartes routières générées par ses utilisateurs (Haklay et Weber (2008)). Des bénévoles du monde entier partagent des informations géographiques au sein d'OSM et sont parfois considérés comme des "capteurs intelligents" (Goodchild (2007)).

Depuis sa création, ce projet a été utilisé de manière de plus en plus intense pour une grande variété de tâches. Plusieurs métriques portant sur la qualité de ses données ont été proposées dans Forghani et Delavar (2014); Barron et al. (2014); Mooney et al. (2010); Fan et al. (2014), ainsi que des études sur la qualité par pays. Ainsi, OSM a été jugé comme "plutôt précis" pour l'Angleterre (Haklay (2010)) et comme supérieur aux données officielles pour le périmètre "Meridian 2" du Royaume-Uni. Ces études ont été étendues au cas de la France (Girres et Touya (2010)). En ce qui concerne l'Allemagne, le réseau routier d'OSM est considéré comme plus complet que pour des alternatives commerciales (Neis et al. (2011)). Par exemple, 99,8% du réseau routier de Hambourg sont couverts par OSM, selon le bureau de planification de la ville (Over et al. (2010)). Cette étude soulève également que "Outre le réseau routier, le réel avantage de cette base de données est la disponibilité de divers points d'intérêt". Ces derniers, contenant par exemple les arrêts de transports en communs ou les stations d'essence, permettent une meilleure compréhension des dynamiques urbaines, grâce à leur géo-localisation et aux diverses informations qui leur sont associées. En Chine, le volume des points d'intérêt a augmenté d'un facteur 9 au cours de la période 2007-2013 (Liu et Long (2016)).

Dans ce travail, nous utilisons des données OSM pour analyser la mixité des usages des sols au sein de villes. Utiliser cette base de données comporte plusieurs avantages. Premièrement, sa couverture est mondiale et elle permet déjà la comparaison de nombreuses villes sur la base d'une source de données homogène. Deuxièmement, OSM est mise à jour continuellement sur le principe de "crowd-sourcing", ce qui rend possible l'adaptation des informations aux changements rapides que connaissent les villes de nos jours. De plus, la communauté très active de ses utilisateurs améliore constamment la précision et la complétude des données. Troisièmement, il s'agit de données ouvertes. Ceci permet de corriger les erreurs de saisie de manière croisée, et surtout d'établir une cohérence entre différentes contributions en planification urbaine. Ainsi, les résultats de recherche deviennent reproductibles pour tout un chacun, ce qui est un facteur important pour l'évaluation et l'amélioration des outils proposés. Finalement, la présence de données géo-localisées permet une analyse spatiale plus fine qu'avec des données localisées sur des grilles, par exemple.

Actuellement, la principale faiblesse d'OSM réside dans la couverture encore imparfaite de quelques régions du monde. Cependant, surtout pour les grandes villes, ce problème est réduit grâce à la présence de nombreux contributeurs. En comparaison, des données de type "LULC" (Land Use Land Cover), typiquement utilisées dans ce genre d'études, ont plusieurs inconvénients. Tout d'abord, leur disponibilité est en général restreinte à des pays individuels, voire à certaines zones géographiques. Ensuite, même si leur résolution a été considérablement accrue, ces données ne contiennent pas d'informations sur l'intensité des usages des sols répertoriés (activités, résidentiel, etc.). De plus, la classification des usages des sols dans une seule classe d'usage par entité spatiale, ainsi que le fait que les catégories d'usages sont typique-

2. http://www.openstreetmap.org/

ment très agrégées, font que ces données sont insuffisantes pour analyser la mixité des usages des sols dans des villes (Christian et al. (2011)). Enfin, les bases de données LULC peuvent périmer rapidement à cause du rythme intense des dynamiques urbaines.

5 Approche proposée

Dans cette section, nous décrivons la chaîne de traitements qui nous permet de calculer une représentation spatialement continue de la mixité des usages des sols, à partir de données OSM brutes.

5.1 Extraction de données

Nous expliquons ici comment extraire les différents usages des sols qui seront ensuite analysés pour déterminer leur degré de mixité. Les données OSM d'un territoire donné, défini par ses contours, sont récupérées au format *shapefile* à travers le site *Mapzen Metro Extracts* [3]. Ces fichiers sont le résultat du processus *osm2pgsql* qui convertit des données OSM en bases de données PostgreSQL/PostGIS [4]. Des polygones et des points d'intérêt sont extraits de ces fichiers et utilisés comme entrées pour la suite des traitements.

Nous procédons d'abord à une classification des points d'intérêt (POI) et des polygones en usages de type "résidentiel" et "activité", en suivant le Wiki d'OSM [5]. Les points et polygones sont associés à des usages des sols à travers les informations qu'ils contiennent (clé, valeur), selon le Tableau 1.

Dans la base OSM, un grand nombre de polygones sont actuellement étiquetés comme "building", sans information supplémentaire. Dans ce cas, l'usage des sols d'un polygone est estimé selon la procédure détaillée dans le Tableau 2. D'abord, les polygones contenant une clé associée sans ambiguïté à un type d'usage des sols (résidentiel, activité ou autre, par exemple forêt, eau), sont classifiés en conséquence. Soient P_R, P_A et P_O les ensembles des polygones associés à ces trois types d'usage. Quant un polygone P possède "building" comme unique clé, l'usage est estimé comme suit :

$$LU(P) = LU(p) \quad if\ \exists p \text{ with } \min_{p \in P_R \cup P_A \cup P_O} A(p)\ ,\ P \subseteq p$$

où $A(\cdot)$ désigne l'aire d'un polygone et $LU(\cdot)$ son usage des sols (*land use*). Remarquons que des polygones associés à différents usages, peuvent se superposer. Nous utilisons alors l'information contenue dans le plus petit polygone qui englobe P et qui a un usage des sols connu.

Il arrive parfois que P ne soit englobé par aucun polygone dont l'usage des sols est connu. Dans ce cas, nous faisons l'hypothèse que P est de type "résidentiel". Ce choix *a priori* purement *ad hoc* est adopté pour contrebalancer le fait qu'OSM contient actuellement plus d'annotations de type "activité"" que de type "résidentiel". En effet, lorsque des utilisateurs définissent des polygones qui contiennent une activité, cette dernière est habituellement annotée, ce qui n'est pas forcément le cas avec les usages résidentiels. Néanmoins, une meilleure complétude

3. http://mapzen.com/data/metro-extracts/
4. http://postgis.net
5. http://wiki.openstreetmap.org/

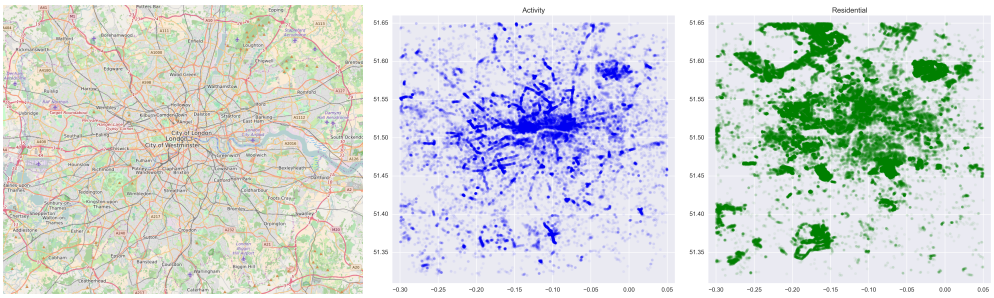
Clé	Valeurs
	(a) Activités
amenity	bar, pub, restaurant, biergarten, cafe, fast_food, food_court, ice_cream, pub, restaurant, college, kindergarten, library, public_bookcase, school, music_school, driving_school, language_school, university, fuel, bicycle_rental, bus_station, car_rental, taxi, car_wash, ferry_terminal, atm, bank, bureau_de_change, baby_hatch, clinic, dentist, doctors, hospital, nursing_home, pharmacy, social_facility, veterinary, arts_centre, brothel, casino, cinema, community_centre, fountain, gambling, nightclub, planetarium, social_centre, stripclub, studio, swingerclub, theatre, animal_boarding, animal_shelter, courthouse, coworking_space, crematorium, dive_centre, dojo, embassy, fire_station, gym, internet_cafe, marketplace, police, post_office, townhall
shop	*
building	commercial, office, industrial, retail, warehouse, cathedral, chapel, church, mosque, temple, synagogue, shrine, civic, hospital, school, stadium, train_station, transportation, university, public, kiosk, garage, garages, hangar, stable, cowshed, digester
leisure	adult_gaming_centre, amusement_arcade, beach_resort, dance, hackerspace, ice_rink, pitch, sports_centre, stadium, summer_camp, swimming_area, water_park
landuse	commercial, industrial, retail, port, quarry, salt_pond, construction, military, garages
	(b) Résidentiel
building	hotel, farm, apartment, apartments, dormitory, house, residential, retirement_home, terrace, houseboat, bungalow, static_caravan, detached
	(c) Autres
landuse	cemetery, landfill, railway, water, reservoir, basin, allotments, conservation, farmland, farmyard, forest, grass, greenfield, greenhouse_horticulture, meadow, orchard, pasture, peat_cutting, plant_nursery, recreation_ground, village_green, vineyard

TAB. 1: Classification des usages des sols en activité, résidentiel et autre, selon les données clé et valeur fournies par OSM.

des annotations résidentielles permettra de simplement abandonner cette hypothèse à l'avenir

Paires Clé=*Valeur*			Classification
leisure = *Activity*			Activité
amenity = *Activity*			Activité
shop = *Activity*			Activité
building = *Activity*			Activité
building = *Residential*			Résidentiel
building = *Yes*	*et*	landuse = *Activity*	Activité
	et	landuse = *Residential*	Résidentiel
	et	landuse = *Null*	à inférer
landuse = *Residential*	*et*	building = *Null*	Inférence : Résidentiel
landuse = *Activity*	*et*	building = *Null*	Inférence : Activité
landuse = *Other*	*et*	building = *Null*	Inférence : nul

TAB. 2: *Look up table* de la procédure d'assignation à une classe d'usage des sols à partir de données OSM.



(a) (b) (c)

FIG. 1: (a) OpenStreetMap de Londres, Angleterre. (b) et (c) montrent les lieux des usages des sols extraits pour, respectivement, les activités et le résidentiel.

si la qualité des annotations continue à augmenter.

Tous les polygones résidentiels avec une aire inférieure à $12m^2$ sont éliminés, comme il s'agit probablement de faux positifs plutôt que de bâtiments résidentiels.

Finalement, les polygones retenus sont convertis en points d'intérêt (POI) en utilisant comme position géographique le lieu de leurs centroïdes. L'objectif est d'obtenir une couverture de la ville par des points géo-référencés contenant des informations d'usage des sols.

Les résultats de cette procédure sont illustrés sur la figure 1 pour le cas de Londres, Angleterre, en utilisant de la transparence partielle pour permettre de distinguer clairement l'information dans les zones hautement concentrées.

5.2 Estimation de la densité par noyau

Nous décrivons dans cette section le processus d'estimation de la densité locale pour les deux usages d'habitation et d'activité. Nous utilisons pour cela le processus statistique d'estimation par noyau (KDE).

Comme montré dans Silverman (1986), le KDE est une technique reconnue pour l'analyse et la représentation de données complexes à des fins de représentations graphiques claires pour des non-mathématiciens.

Le KDE infère une fonction de densité f, sachant un échantillonnage indépendant et identiquement distribué $x_1, x_2, ..., x_n$ de la distribution de probabilité correspondante. L'estimateur du noyau de densité est défini par

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{\|x - x_i\|}{h}\right), \tag{1}$$

$K(\bullet)$ étant le noyau, et h un paramètre de lissage dénommé bande passante. Dans notre cas, les x_i sont des vecteurs bi-dimentionnels contenant les coordonnées de longitude et latitude des points d'intérêt.

Étant donné un ensemble de points, le KDE interpole une surface continue en utilisant le noyau défini (e.g. une fonction Gaussienne/normale). Cette procédure présente les deux avantages principaux de lissage et d'interpolation spatiale, qui sont des caractéristiques pertinentes dans notre contexte où des données peuvent être manquantes. Ainsi, les centres d'activités et les centres résidentiels ponctuels servent à construire une surface continue permettant de remplacer de manière réaliste les données manquantes.

Comme décrit ci-dessus, le KDE est contrôlé par la fonction de noyau K et la bande passante h. Pour le noyau, nous pré-supposons dans ce travail que les usage résidentiels et d'activités ont tous une influence spatiale normalement distribuée. Ainsi, l'impact d'une activité est maximale à sa localisation d'origine et réduit avec la distance en suivant une distribution normale.

La bande passante est choisie pour représenter une "distance d'influence moyenne". Il faut mentionner ici que plusieurs méthodes de calcul automatique de bande passante existent dans la littérature, mais leur but est de réduire la complexité durant la phase d'estimation aux dépens de la précision. Dans notre cas, étant donné que nous cherchons à capturer l'influence des usages d'habitation et d'activité au sein de leur voisinage géographique, nous identifions la bande passante à la distance moyenne de déplacement piéton. Il est montré dans Atash (1994) que la distance moyenne à laquelle un Américain préférera prendre sa voiture plutôt que de marcher est de $400m$. Selon Aultman-Hall et al. (1997), cette distance est aussi considérée comme la plus grande distance moyenne qu'un voyageur en transit entre des stations de métro ou de bus aura à parcourir. De manière similaire (Western Australian Planning Commission (2007)) propose une distance comprise entre $400m$ et $500m$. Certaines études proposent cependant des distances moyennes de déplacement à pied plus grandes, comme par exemple dans McCormack et al. (2008), qui étudient l'association entre activité physique et le mélange de destinations comprises entre $400m$ et $1500m$ des maisons des résidents. Dans notre contexte, et dans le but de favoriser des voisinages avec une grande mixité d'usages, nous avons décidé d'adopter la valeur suggérée de $400m$, qui sera donc la bande passante de notre procédure. Il faut toutefois garder à l'esprit que cette valeur devrait si possible être adaptée aux pays ou villes étudiés si des travaux pré-existants permettent de la préciser plus avant.

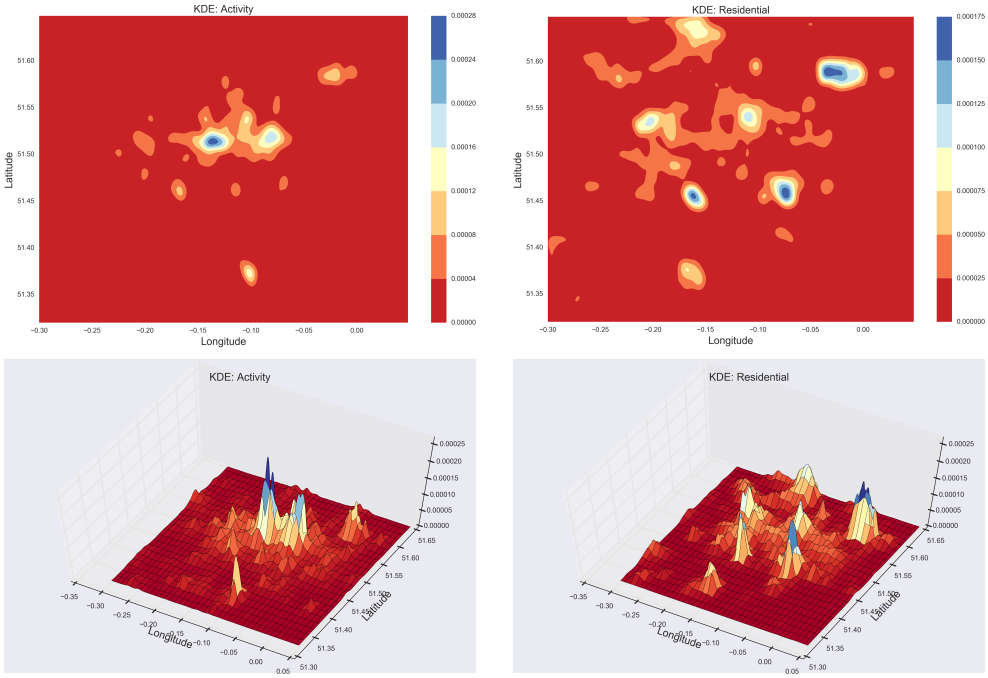


FIG. 2: KDE estimés pour Londres. A gauche, activités et à droite, résidentiel.

Comme mentionné ci-dessus, le KDE est un outil très efficace pour effectuer du lissage spatial et interpoler des données manquantes. Cependant, dans des cas limites tels que des frontières géographiques strictes (régions côtières, coupures franches, etc.) il impliquera un sur-lissage et l'estimation d'informations non existantes. Nous étudierons les limites de tels cas dans des travaux futurs.

Comme dans les travaux de Song et al. (2013), nous avons choisi de nous intéresser à deux types d'usages : les usages résidentiels et les usages d'activités (qui regroupent tous les usages non résidentiels tels que les magasins, les lieux publics, les lieux de travail, etc.). Par conséquent, tous les POI géo-localisés de la procédure ci-dessus sont utilisés pour calculer les représentations de ces deux catégories [6]. La fonction de densité de probabilité est ensuite évaluée le long d'une grille de points couvrant la région d'intérêt avec une résolution choisie par l'utilisateur. Dans les résultats présentés dans cet article, et étant donnée l'échelle des régions d'intérêt des villes analysées et le niveau de détail que nous recherchons dans les zones de centre ville, nous avons choisi une résolution de $100m$.

Enfin, pour terminer, les points calculés sont normalisés pour chaque KDE.

Les KDE résultants sont décrits dans la figure 2, appliqués à la ville Anglaise de Londres.

6. Il s'agit là d'une première étape de nos travaux, rien n'empêchant de détailler les usages dans des classes plus précises en fonction de la précision nécessaire pour le calcul de l'indice final.

5.3 Calcul de la mixité des usages des sols

Dans cette section, nous nous focalisons sur le calcul d'une mesure de la mixité des usages des sols, en partant des cartes de densité de type "activité"" et "résidentiel", estimées avec la méthode décrite ci-dessus. L'objectif est de décrire à quel point ces usages sont bien répartis dans une ville et de comprendre leur degré de co-localisation au sein de voisinages locaux. En règle générale, une bonne mixité locale est désirable, pour les raisons énoncées en début de l'article.

Selon Song et al. (2013), les mesures de mixité d'usages des sols doivent prendre en compte deux concepts sous-jacents : la *distance* et la *quantité*. La distance se réfère à la proximité entre des usages différents. Nous calculons cette proximité (nous parlons aussi de l'influence d'un usage sur un autre) en prenant en compte la distance usuelle de trajets pédestres (voir la section 5.2). En ce qui concerne la quantité, il s'agit pour nous de la sortie du calcul des densités tel que décrit dans la section 5.2.

En accord avec ces notions de distance et quantité, une mesure de mixité doit donc mettre en évidence des régions où différents usages sont présents en quantités comparables et à proximité les uns des autres.

Comme il a été mentionné plus haut, notre plateforme a vocation à être une aide pour des planificateurs urbains, et la possibilité de visualisation géo-spatiale de la mixité calculée est donc un point important. Dans ce contexte, il est également conçu pour mettre en évidence l'importance des différentes sous-régions au sein d'une ville : des grilles maillées calculées à partir des résultats d'estimation de densité par noyau permettent donc de déterminer l'importance de ces sous-régions en terme de l'intensité de leur usage des sols.

Nos premiers résultats sont visualisés à l'aide de "bubble plots", voir par exemple la figure 3. L'idée est de visualiser la LUM dans les régions d'une ville où l'intensité est élevée, soit au niveau activité, au niveau résidentiel, ou les deux. D'autres modes de visualisation ne permettent pas toujours de montrer cela.

Dans ce travail, nous utilisons l'indice d'entropie (EI – Entropy Index) comme mesure de mixité. Ce choix est essentiellement guidé par le fait qu'il s'agit d'une mesure fréquemment utilisée (voir par exemple Song et Knaap (2004)). Cependant, comme précisé section 3, notre but n'est pas ici d'évaluer la pertinence de cet indice ni d'en proposer un nouveau, mais de proposer une méthode qui permet de calculer directement une représentation spatiale continue de n'importe quel indice existant pour peu qu'il soit basé sur des données présentes dans la base OpenStreetMap.

Soit $P_j(i)$ la proportion d'usage de type j estimé pour le point i à l'aide des KDE de la section précédente. Soit K le nombre d'usages différents [7], alors :

$$EI(i) = \frac{-\sum_{j=1}^{K} P_j(i)\ ln(P_j(i))}{ln(K)} \tag{2}$$

L'EI varie entre 0 et 1, sa valeur croissant avec la mixité des usages.

La figure 4 présente l'usage des sols calculé pour la ville Anglaise de Londres. La figure 3 utilise un bubble plot pour présenter les intensités locales de mixité d'usage des sols. Une analyse de ces résultats est présentée section 6.

7. Actuellement, nous avons $K = 2$ (activités et résidentiel), mais dès qu'OSM devient plus complet et précis (annotations fines des usages, cf. tableau 1) un nombre plus grand de classes d'activités peut être traité.

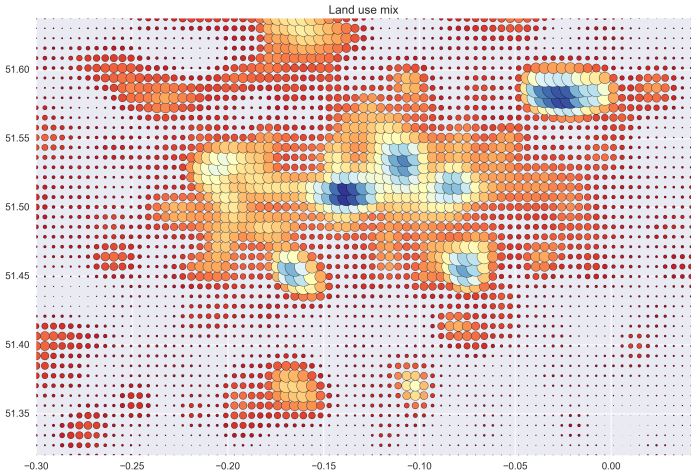


FIG. 3: “Bubble plot” de la LUM calculée pour Londres. La taille d’un point représente l’intensité locale de la mixité d’usage des sols.

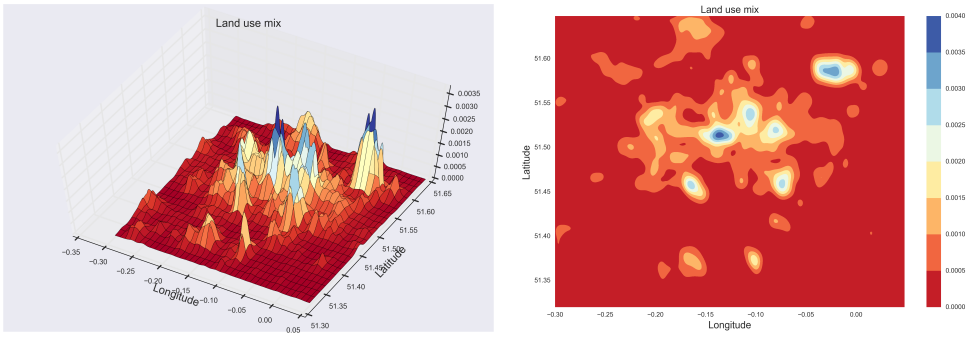


FIG. 4: LUM calculée pour Londres.

5.4 Implémentation

L’infrastructure logicielle de l’approche proposée, construite au-dessus d’une pile Python, est composée des modules suivants :

- `extract_uses.py` : étant donnés les points et polygones (*shapefiles*) *via* Mapzen Metro Extract, l’extraction des usages associés est effectuée selon la méthode décrite en section 5.1. La géométrie des polygones est traitée avec la librairie *Shapely* et la sortie finale est un *DataFrame Pandas* (McKinney (2010)) contenant les POIs géo-localisés et catégorisés, `pois_df`.
- `kde.py` : à partir des `pois_df`, ce module calcule les KDE pour les usages de type "activité" (`kde_act`) et "résidentiel" (`kde_res`) en utilisant la librairie *Statsmodels* (Seabold et Perktold (2010)).
- `loaders.py` : gère un store Hadoop distribué (HDFS, Hadoop Distributed File Store)

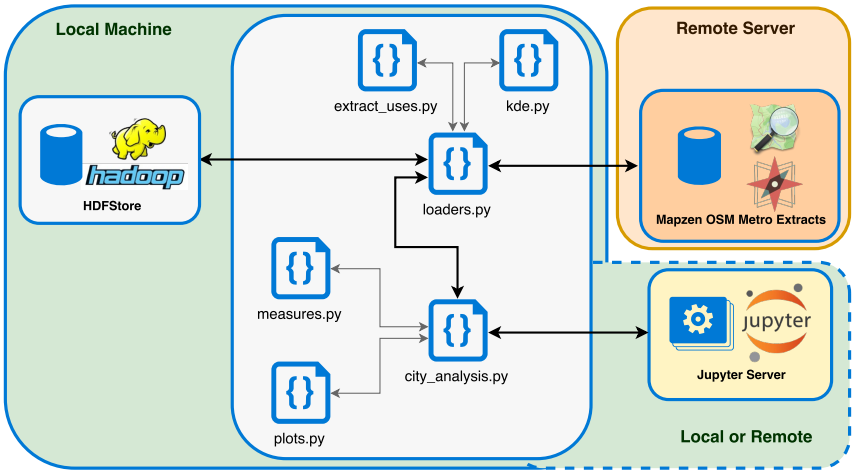


FIG. 5: Implémentation de la plateforme proposée.

à l'aide de la librairie décrite dans Alted et Fernández-Alonso (2003). Un tel HDFS stocke, pour chaque *ville* étudiée, les résultats des calculs potentiellement coûteux faits par les modules `extract_uses.py` et `kde.py` ci-dessus. Si, pour une *ville* donnée, ces informations ne sont pas stockées dans un HDFS, il est procédé comme suit :

1. Une requête pour les *shapefiles* de cette *ville* est envoyée à Mapzen [8]
2. Après réception des *shapefiles*, `extract_uses.py` et `kde.py` sont appelés afin de calculer `pois_df`, `kde_act` et `kde_res`
3. Les résultats sont stockés dans un HDFS local, sous une clé correspondant à la *ville*.

– `measures.py` : calcule les mesures de mixité qui sont implémentées (ici, l'indice d'entropie, cf. équation 2), à partir de `pois_df`, `kde_act` et `kde_res`, en utilisant des arrays *Numpy*, voir Van Der Walt et al. (2011).
– `plots.py` : ce module génère les sorties graphiques désirées à partir de `pois_df`, `kde_act` et `kde_res` (scatterplots, plots des KDE, etc.) à l'aide de la librairie *Matplotlib*, voir Hunter et al. (2007).
– `city_analysis.py` : c'est la classe Python principale, qui interagit avec le serveur Jupyter. Pour démarrer l'analyse pour une *ville*, elle demande les informations de région d'intérêt sur cette *ville* à `loaders.py`. Une fois celles-ci disponibles, la classe gère les appels aux méthodes `measures.py` et `plots.py` pour construire les sorties de manière interactive.

L'architecture de la plateforme est représentée sur la figure 5. Le code source est ouvert et disponible à l'URL `http://github.com/martibosch/landusemix` . Nous prévoyons d'implémenter une interface web permettant à des non-informaticiens d'utiliser notre outil en ligne.

8. `http://mapzen.com/data/metro-extracts/`

6 Applications

Cette section présente une première analyse de résultats sur deux villes Européennes aux structures assez différentes. Le but est d'abord de montrer que les données présentes dans la base OSM sont suffisamment précises pour reconstruire une carte continue de l'indice choisi à l'aide de la plateforme. Par ailleurs, cela permet aussi de constater que notre méthode permet de dépasser la limitation séparant les indices intégraux et divisionnels. En effet, nous calculons ici la version géographiquement continue d'un indicateur divisionnel, ce qui ouvre des perspectives sur la possibilité de comparer l'expressivité d'indices différents.

Cela dit, une validation plus poussée de la carte de LUM construite est nécessaire, et sera effectuée dans nos prochains travaux. Nous pourrons notamment comparer la carte LUM avec une vérité terrain pré-existante dans le cas de jeux de données suffisamment précis (et potentiellement propriétaires), reconstruire des indices intégraux connus sur des zones géographiques déjà analysées, et éliciter des avis d'experts du domaine sur les couches GIS calculées. Muni d'un tel outil, nous pourrons dans un second temps nous concentrer sur le calcul d'un indicateur optimal à l'aide de méthodes d'apprentissage supervisé, et comparer ce dernier aux indicateurs existants.

6.1 Grenoble, France

La ville Française de Grenoble est une ancienne cité Européenne de taille moyenne. Elle est localisée au pied des Alpes Françaises et possède une forme caractéristique de Y due aux massifs montagneux environnants de la Chartreuse, de Belledonne et du Vercors. Nous avons calculé la carte de LUM de cette agglomération de 665000 habitants. Un total de 33055 points et 118207 polygones ont été traités en $10min$; 11258 activités et 84445 points résidentiels ont été extraits. Les KDE de ces deux usages ont été calculés en $32min$, et la carte finale de mixité urbaine générée en $2s$.

Ces résultats sont présentés Fig. 6 et Fig. 7. On peut voir dans le KDE des activités une haute concentration autour du centre historique correspondant aux nombreuses activités commerciales qui y sont localisées. Ce motif est fréquent dans les villes Européennes dans lesquelles les centres contiennent souvent les plus grandes concentrations d'activités suite à une évolution régulière dans l'histoire.

La carte des usages résidentiels montre une forte concentration autour du centre historique, causée par le même phénomène. Les anciens habitats datant d'une époque antérieure à la voiture, la distance moyenne entre habitations est plus réduite, et leur concentration est donc plus élevée. Pour ce type de villes, les centres historiques présentent donc souvent de hauts niveaux de mixité urbaine.

Les résultats montrent de plus une caractéristique particulière de la ville de Grenoble. Les abords de la ville étant constitués de zones semi-montagneuses, plusieurs centres résidentiels peuvent être trouvés dans des zones suffisamment éloignées du centre pour fournir une bonne qualité de vie avec un accès rapide à la nature, tout en étant facilement accessible par transport public ou véhicule personnel. Les structures visibles aux coordonnées géographiques $5,58°E/45,13°N$ et $5,84°E/45,14°N$ représentent respectivement l'Est du plateau du Vercors et la vallée de Vaulnaveys. Il s'agit de deux zones résidentielles ayant beaucoup moins d'activités que le centre ville.

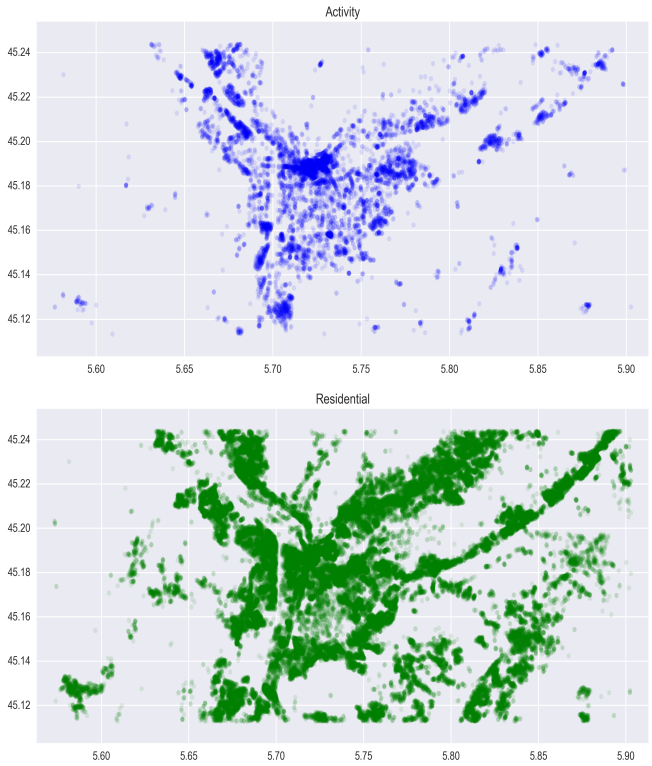


FIG. 6: Points d'intérêt extraits pour Grenoble, France (en haut les activités, en bas les zones résidentielles).

6.2 Londres, Angleterre

Nous avons évalué notre plateforme sur la cité de Londres. Un total de 80702 and 238626 points ont été respectivement extraits pour les usages de type "activité" et "résidentiel". Les résultats sont présentés Fig. 1 (usages), Fig. 2 (KDE) et Fig. 4 et 3 (LUM).

La cité Londonienne, qui a vécu de grands changements au cours de son histoire, possède actuellement $8,674$ millions d'habitants. On peut voir dans la figure 2 la City de Londres et le quartier de Soho comme ayant la plus grande densité d'activités. Du point de vue des usages résidentiels, East Dulwich, la colline de Higham, et la zone entre Wandsworth Common et Clapham Common sont les zones ayant le KDE le plus intense, ce qui correspond aux grandes concentrations d'habitation connues.

Croydon Vision 2020 est un programme de développement urbain qui vise à faire de Croydon un centre de vie, de commerce, de culture et d'activités dans le contexte du plan de développement local. Une importante concentration d'activités est donc présente dans cette zone, en conjonction avec une importante zone résidentielle un peu plus à l'Ouest. Nous devrions donc nous attendre à observer à cet endroit une forte mixité urbaine, d'autant plus que la volonté d'avoir un fort développement mixte est mise en avant par les concepteurs de ce pro-

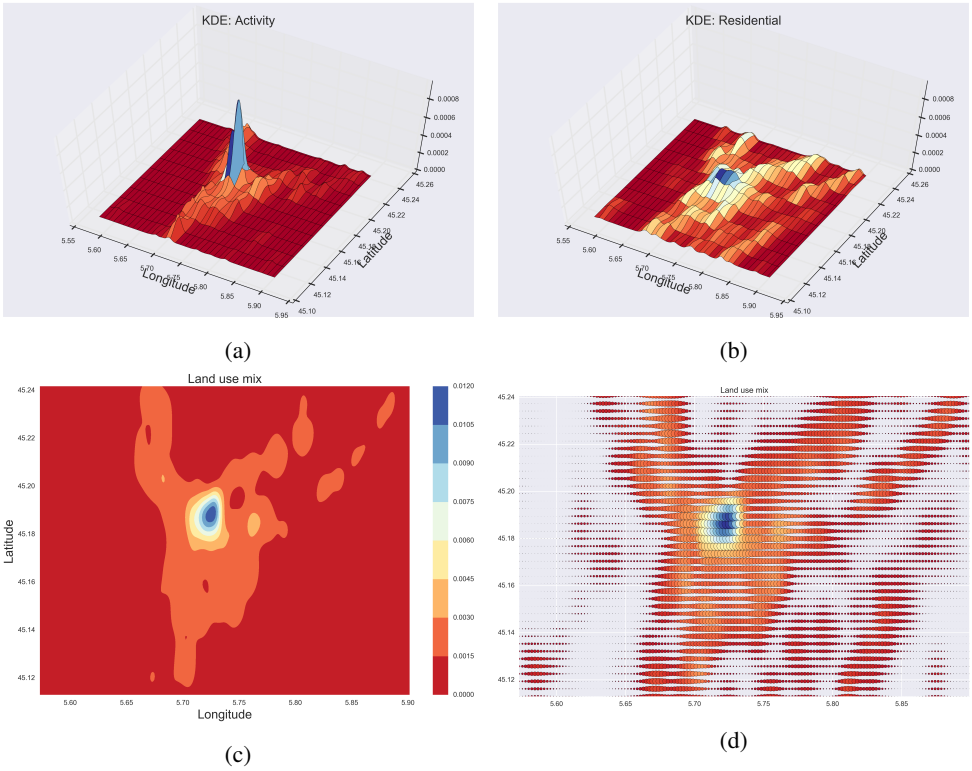


FIG. 7: Les résultats pour Grenoble, France. (a) et (b) montrent les KDE pour les activités et les zones résidentielles tandis que (c) et (d) montrent les LUM calculées selon deux modes différents.

gramme urbain. Et en effet, le résultat du calcul de l'indice de mixité montre qu'un haut niveau existe dans cette région, comme montré Fig.4, aux coordonnées géographiques -0.100594 et 51.376495. En surveillant l'évolution de cet indice dans cette zone, nous devrions par ailleurs constater une augmentation de l'intégration de différents usages dans le futur.

7 Conclusion et travaux futurs

Nous présentons dans cet article une plateforme pour capturer les informations de mixité d'usage des sols au sein des villes. L'outil logiciel présenté extrait des données collaboratives d'OpenStreetMap une représentation géo-localisée des usages. De par ses caractéristiques, cette source de données nous permet de traiter n'importe quelle zone géographique si suffisamment de données sont disponibles dans la base. Une estimation de densité par noyau permet de générer une distribution spatiale pour chaque usage. À partir de cette représentation, une mesure de mixité d'usage des sols est ensuite calculée. Dans le cas de ces travaux, nous utilisons l'indice d'entropie, mais n'importe quelle métrique peut être calculée avec cette

méthode. La couche SIG finale montre la zone géographique choisie enrichie par l'information de mixité urbaine dans le but d'aider les décideurs et les aménageurs.

Notre outil, disponible en ligne et dont l'implémentation est décrite, est ici appliqué aux villes de Londres et de Grenoble. Les estimations des usages résidentiel et d'activités sont pour l'instant validées en inspectant les voisinages locaux ayant une haute valeur d'usages, et en les re-contextualisant. Finalement, la distribution spatiale de la mixité urbaine est calculée, permettant une visualisation simple, et montrant les variations de la métrique sur l'ensemble de la zone d'intérêt.

Les travaux futurs sont multiples. Nous allons tout d'abord approfondir le processus de validation de la carte de LUM construite, en l'utilisant d'une part pour reconstruire des indices intégraux connus sur des zones géographiques déjà analysées, et d'autre part en élicitant des avis d'experts du domaine. Nous envisageons ensuite d'appliquer notre outil à l'analyse du phénomène de l'étalement urbain, en effectuant une interprétation numérique des valeurs de mixité "désirables" dans ce contexte. Nous planifions également d'enrichir la représentation actuelle en intégrant un modèle de transport qui permettra une estimation de la mixité urbaine plus précise en prenant en compte, en plus de leur localisation, l'accessibilité des différents usages, en partie conditionnée par l'infrastructure de transport. Enfin, nous analyserons plus en détail les différents types d'activités qui sont actuellement toutes regroupées dans une seule classe. En effet, la richesse des informations de la base OpenStreetMap permet de détailler les activités extraites en nombreuses classes (activité commerciale, industrielle, de loisir, etc.). Notre modèle permet de créer un KDE pour chaque classe et, en faisant varier leur importance relative, pourrait converger vers une représentation plus réaliste.

Références

Alted, F. et M. Fernández-Alonso (2003). Pytables : processing and analyzing extremely large amounts of data in python. In *PyCon*, pp. 1–9.

Atash, F. (1994). Redesigning suburbia for walking and transit : Emerging concepts. *Journal of Urban Planning and Development 120*(1), 48–57.

Aultman-Hall, L., M. Roorda, et B. W. Baetz (1997). Using gis for evaluation of neighborhood pedestrian accessibility. *Journal of Urban Planning and Development 123*(1), 10–17.

Barron, C., P. Neis, et A. Zipf (2014). A comprehensive framework for intrinsic openstreetmap quality analysis. *Transactions in GIS 18*(6), 877–895.

Breheny, M. (1995). The compact city and transport energy consumption. *Transactions of the Institute of British Geographers 20*(1), 81–101.

Burgess, R. (2000). The compact city debate : A global perspective. In R. Burgess et M. Jenks (Eds.), *Compact cities : Sustainable urban forms for developing countries*, pp. 9–24. Spon Press : London, UK.

Carlos, H. A., X. Shi, J. Sargent, S. Tanski, et E. M. Berke (2010). Density estimation and adaptive bandwidths : a primer for public health practitioners. *International Journal of Health Geographics 9*(1), 1.

Cervero, R. (1996). Mixed land-uses and commuting : evidence from the american housing survey. *Transportation Research Part A : Policy and Practice 30*(5), 361–377.

Cervero, R., C. Ferrell, et S. Murphy (2002). Transit-oriented development and joint development in the united states : A literature review. Technical Report 52, Transportation Research Board.

Christian, H. E., F. C. Bull, N. J. Middleton, M. W. Knuiman, M. L. Divitini, P. Hooper, A. Amarasinghe, et B. Giles-Corti (2011). How important is the land use mix measure in understanding walking behaviour ? results from the reside study. *International Journal of Behavioral Nutrition and Physical Activity 8*(1), 1.

Ewing, R. H. (1995). Characteristics, causes, and effects of sprawl : A literature review. In *Urban Ecology*, pp. 519–535. Springer.

Fan, H., A. Zipf, Q. Fu, et P. Neis (2014). Quality assessment for building footprints data on openstreetmap. *International Journal of Geographical Information Science 28*(4), 700–719.

Forghani, M. et M. R. Delavar (2014). A quality study of the openstreetmap dataset for tehran. *ISPRS International Journal of Geo-Information 3*(2), 750–763.

Frank, L. D., M. A. Andresen, et T. L. Schmid (2004). Obesity relationships with community design, physical activity, and time spent in cars. *American journal of preventive medicine 27*(2), 87–96.

Girres, J.-F. et G. Touya (2010). Quality assessment of the french openstreetmap dataset. *Transactions in GIS 14*(4), 435–459.

Goodchild, M. F. (2007). Citizens as sensors : web 2.0 and the volunteering of geographic information. *GeoFocus 7*, 8–10.

Gordon, P. et H. W. Richardson (1997). Are compact cities a desirable planning goal ? *Journal of the American Planning Association 63*(1), 95–106.

Haklay, M. (2010). How good is volunteered geographical information ? a comparative study of openstreetmap and ordnance survey datasets. *Environment and planning B : Planning and design 37*(4), 682–703.

Haklay, M. et P. Weber (2008). Openstreetmap : User-generated street maps. *Pervasive Computing, IEEE 7*(4), 12–18.

Hunter, J. D. et al. (2007). Matplotlib : A 2D graphics environment. *Computing in science and engineering 9*(3), 90–95.

Johnson, M. P. (2001). Environmental impacts of urban sprawl : a survey of the literature and proposed research agenda. *Environment and planning A 33*(4), 717–735.

Kockelman, K. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance : evidence from san francisco bay area. *Transportation Research Record : Journal of the Transportation Research Board 1607*, 116–125.

Leccese, M. et K. McCormick (2000). *Charter of the new urbanism*. McGraw-Hill Professional.

Liu, X. et Y. Long (2016). Automated identification and characterization of parcels with openstreetmap and points of interest. *Environment and Planning B : Planning and Design 43*(2), 341–360.

McCormack, G. R., B. Giles-Corti, et M. Bulsara (2008). The relationship between destination proximity, destination mix and physical activity behaviors. *Preventive Medicine 46*(1), 33–

40.

McKinney, W. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, Volume 445, pp. 51–56.

Mooney, P., P. Corcoran, et A. C. Winstanley (2010). Towards quality metrics for openstreetmap. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pp. 514–517. ACM.

Neis, P., D. Zielstra, et A. Zipf (2011). The street network evolution of crowdsourced maps : Openstreetmap in germany 2007–2011. *Future Internet 4*(1), 1–21.

Nelson, A. et J. Duncan (1995). *Growth management principles and practices*. Planners Press, American Planning Association.

Neuman, M. (2005). The compact city fallacy. *Journal of Planning Education and Research 25*(1), 11–26.

Newman, P. et J. Kenworthy (2006). Urban design to reduce automobile dependence. *Opolis 2*(1), 35–52.

Over, M., A. Schilling, S. Neubauer, et A. Zipf (2010). Generating web-based 3d city models from openstreetmap : The current situation in germany. *Computers, Environment and Urban Systems 34*(6), 496–507.

Raines, G. L. (2002). Description and comparison of geologic maps with fragstats—a spatial statistics program. *Computers & Geosciences 28*(2), 169–177.

Seabold, S. et J. Perktold (2010). Statsmodels : Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, pp. 57–61.

Sierra Club (1998). Sprawl : The dark side of the American dream. Technical report.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Volume 26. CRC press.

Smart Growth Network (2006). This is smart growth. Technical report.

Song, Y. et G.-J. Knaap (2004). Measuring the effects of mixed land uses on housing values. *Regional Science and Urban Economics 34*(6), 663–680.

Song, Y., L. Merlin, et D. Rodriguez (2013). Comparing measures of urban land use mix. *Computers, Environment and Urban Systems 42*, 1–13.

Squires, G. D. (2002). *Urban sprawl : Causes, consequences, & policy responses*. The Urban Insitute.

United Nations (2014). World urbanization prospects : The 2014 revision. Technical Report ST/ESA/SER.A/352.

United States Department of Housing and Urban Development (1999). State of the cities – 1999. Technical report.

Van Der Walt, S., S. C. Colbert, et G. Varoquaux (2011). The numpy array : a structure for efficient numerical computation. *Computing in Science & Engineering 13*(2), 22–30.

Western Australian Planning Commission (2007). Liveable neighbourhoods : a Western Australian Government sustainable cities initiative.

Summary

Population in urban areas has been increasing at an alarming rate in the last decades. This evidence, together with the rising availability of massive data from cities, has motivated research on sustainable urban development.

In this paper we present a GIS-based land use mix analysis framework to help urban planners to compute indices for mixed uses development, which may be helpful towards developing sustainable cities. Residential and activities land uses are extracted using OpenStreetMap crowd-sourcing data. They are used to compute a continuous representation of the entropy index, but can be generalized to the building of any other index. The framework is applied to several cities and the first results are presented.