



HAL
open science

Discontinuités spatiales, santé et mobilités. Analyse des POI Google et de Tweets pour caractériser les structures spatiales et les dynamiques d'attractivités de Bangkok (Thaïlande)

Alexandre Cebeillac, Éric Daudé, Alain Vaguet

► To cite this version:

Alexandre Cebeillac, Éric Daudé, Alain Vaguet. Discontinuités spatiales, santé et mobilités. Analyse des POI Google et de Tweets pour caractériser les structures spatiales et les dynamiques d'attractivités de Bangkok (Thaïlande). *Spatial Analysis and GEOmatics* 2017, INSA de rouen, Nov 2017, Rouen, France. hal-01649148

HAL Id: hal-01649148

<https://hal.science/hal-01649148v1>

Submitted on 27 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Discontinuités spatiales, santé et mobilités. Analyses et typologies des Google POI et des Tweets pour caractériser les structures spatiales et les dynamiques d'attractivités de Bangkok (Thaïlande)

Cebeillac Alexandre¹, Daudé Eric¹, Vaguet Alain¹

1. UMR 6266-IDEES, Université de Rouen ;
7 rue Thomas Becket , 76821 Mont Saint-Aignan, France
prenom.nom@univ-rouen.fr

RÉSUMÉ. L'objectif général de ce papier consiste à discuter des fortes discontinuités spatiales dans les métropoles asiatiques et de leurs effets, notamment en termes de santé des populations. Dans le même temps, l'exploration de jeux de données originaux permet désormais de mieux caractériser les structures et les dynamiques dans des villes où le chercheur ne dispose pas aisément de données spatialisées. La diffusion d'Internet, des smartphones et des applications ouvrant à la géolocalisation, a contribué à renouveler les opportunités de recueil de données pour les chercheurs. Nous proposons dans cet article d'explorer les données issues de Google POI (Point Of Interest) et du réseau social Twitter pour dresser une typologie des activités de la ville de Bangkok et pour mieux en mesurer les pulsations par les mobilités quotidiennes qui s'y exercent.

ABSTRACT. The general objective of this paper is to discuss the strong spatial discontinuities in Asian metropolises and their effects, particularly in terms of population health. At the same time, the exploration of original datasets now makes it possible to better characterize structures and dynamics in cities where the researcher does not have spatialized data. The spread of the Internet, smartphones and applications for geolocation has helped to renew the opportunities for researchers. In this article, we propose to explore data from Google POI (Point of Interest) and the Twitter social network in order to draw up a typology of the activities of the city of Bangkok and to better measure the pulsations by the daily mobility.

MOTS-CLÉS : Discontinuité, google POI, Tweets, dynamique urbaine

KEYWORDS: Discontinuity, google POI, Tweets, urban dynamic

1. Introduction

Beaucoup de recherches en sciences sociales ont tenté de mettre en évidence le rôle des lieux, de l'espace et des territoires sur la santé humaine. Ces travaux ont débuté depuis longtemps, en particulier chez les historiens et géographes, qui ont ainsi mis en évidence la géographie sociale sous-jacente aux épidémies du passé et du présent. Beaucoup de recherches récentes se sont également orientées sur les discontinuités d'état de santé dans les zones urbaines contemporaines pour lesquelles les diverses questions traitées pourraient s'énoncer ainsi : Les disparités de santé se confondent-elles avec les discontinuités urbaines et les mobilités ?

Les cartographies fines des maladies dans l'espace urbain montrent dans leurs grandes lignes que les disparités spatiales de santé restent essentiellement associées aux déterminants sociaux (favorisés/défavorisés) et aux types de quartiers (réguliers/irréguliers, autorisés/non autorisés). Mais dans le détail, il est toujours délicat d'expliquer comment les disparités urbaines jouent un rôle consistant dans les différences d'état de santé. Celles-ci résultent en effet de combinaisons locales diverses et originales à chaque lieu. D'ailleurs, il n'est pas sûr que les grandes disparités classiquement présentées (vieille ville / nouvelle ville, quartier régulier/quartier non contrôlé) fassent sens pour toutes les maladies et en particulier pour la dengue (Daudé, Vaguet, 2015). Cette dernière est souvent associée aux espaces centraux et le lien avec les espaces défavorisés reste discuté. En effet cette maladie, dont le moustique vecteur synanthropique vit essentiellement en milieu urbain, semble atypique dans le sens où elle paraît pouvoir facilement transgresser les limites sociales et par là être perçue comme plus risquée par les élites des métropoles (Telle et al., 2016). Ce que l'on sait du rôle des facteurs caractéristiques de l'urbain, comme les fortes densités, est-il opérationnel dans le cas de cette maladie émergente? Les lieux les plus peuplés dissolvent-ils le risque vectoriel? Ou, à l'inverse, créent-ils des biotopes favorables à une accélération des diffusions des virus du fait des interactions hôtes/vecteurs plus importantes? A ces questions populationnelles s'ajoutent des effets de proximité ou de relégation entre les groupes sociaux.

Comment dans ce contexte évolutif articuler les connaissances de la géographie urbaine aux faits épidémiologiques nouveaux dont témoignent les maladies émergentes comme la dengue? Le différentiel du risque de ces systèmes pathogènes complexes est créé par la conjonction, dans un lieu donné, à un moment donné, de 4 facteurs : le virus, l'hôte, le vecteur et l'environnement. Il convient alors de savoir si la rencontre des divers facteurs est fortuite ou conditionnée par des déterminismes. Par exemple les températures qui permettent la présence du moustique *Ae. Aegypti* (principal vecteur des virus de la dengue, du Zika et du Chikungunya) se situent dans entre 15 et 40°C, avec une variation de la compétence du moustique dans cet intervalle. La température et ses variations microlocales constituent ainsi un ensemble de barrières invisibles qui conditionnent la vie du vecteur en un lieu (Misslin et al., 2016), mais ce ne sont pas les seules. Il s'agit alors de repérer les principales discontinuités de ce système pour tâcher d'en comprendre les principaux déterminants, et leurs évolutions dans le temps.

En effet au-delà des moustiques et des hommes, finalement relativement ubiquistes même s'ils présentent des gradients de densité variable, les virus se trouvent vraisemblablement brassés par les mobilités des populations contribuant à leur diffusion dans les groupes socio-spatiaux (Reiner *et al.*, 2014). Cette dynamique de déplacement du virus liée à celle des populations est elle-même canalisée selon les foyers d'immunités acquises des populations lors des épidémies précédentes ainsi que par les principales discontinuités spatiales urbaines, et de leurs spécificités. C'est donc à travers une modélisation à base d'agent que nous tentons d'explorer la complexité de ce système pathogène. Nous développons pour cela un cadre méthodologique et un modèle (Daudé *et al.* 2015) dans lequel le moustique (Maneerat, Daudé, 2016-2017), l'hôte (Cebeillac *et al.*, 2017) et l'environnement (Misslin, Daudé, 2016) sont décrits puis les interactions simulées afin d'explorer de nombreuses dynamiques. Celles du moustique d'abord, et de ses préférences : climat favorable, lieux de ponte et de repos sont les atouts majeurs d'un environnement propice au développement des populations de moustique. Celle des hommes ensuite, et de leurs multiples espaces de vie dans la ville. Celle de l'environnement, celui construit et produit par les sociétés qui ont tant contribué à favoriser l'implantation de moustiques synanthropiques, et celui naturel et évolutif des conditions météorologiques. Celle des épidémies enfin, dont les cycles sont liés aux précédentes dynamiques et aux trajectoires passées.

Dans ce contexte, l'objectif de cet article est de présenter les données et méthodes employées pour caractériser l'environnement urbain discontinu de nos simulations et pour calibrer le modèle de mobilités. Notre travail se déroule essentiellement dans la métropole de Bangkok, capitale de la Thaïlande, qui comptait 9.2 millions en 2015, ce qui la classe au 36ème rang mondial des plus grandes métropoles, et au 3ème rang parmi les pays de l'ASEAN. C'est une ville très étendue, d'une superficie d'environ 1500 km², divisée en 50 districts (Khets) et 169 sous-districts, les khwaengs – l'équivalent des arrondissements Français. La dengue y est endémique depuis les années 1960 et des dizaines de milliers de cas y sont enregistrés chaque année.

Nous décrivons dans les sections qui suivent (1) les discontinuités des structures spatiales par l'entremise des activités qui sont renseignées dans les Google POI (Point Of Interest) et (2) les discontinuités spatio-temporelles par l'intermédiaire des mobilités quotidiennes qui y sont observées à partir des données géolocalisées issues du réseau social Twitter.

2. Structuration spatiale des activités à Bangkok

L'objectif de cette première section est de décrire, à une résolution fine, la ville de Bangkok au regard des lieux et des activités qui y sont localisées. La fréquentation d'un lieu est liée à sa situation géographique dans la ville et au nombre, à la diversité et à la spécificité des sites qu'il accueille : école, résidence, banque, aéroport, marché etc. Construire une typologie fonctionnelle de la ville doit

permettre, en relation avec les mobilités, de spécifier des potentiels d'attractions différenciés spatialement.

2.1 Données

Dans leurs travaux sur la caractérisation des zones commerciales à Paris (Fleury et al. 2012) ont mobilisé la base de donnée commerciale de la ville agrégée à l'échelle des IRIS. Il en ressort une analyse des différents niveaux de centralité, que cela soit en termes de concentration des commerces que des niveaux de spécialisations des quartiers. Reproduire une telle étude dans d'autres villes n'était par le passé pas aisée car ces types de bases de données n'étaient pas forcément existantes ou accessibles, notamment à Bangkok. Avec l'augmentation des services liés à la géolocalisation, qu'ils soient libres et open sources (Open Street Map : OSM) ou propriétaire (Google), il devient aujourd'hui possible d'accéder à d'importantes bases de données géographiques concernant les types d'établissements qui jalonnent les grandes villes.

Ainsi le site participatif Open Street Map¹, permet à quiconque de créer ou de modifier des points (POI), des lignes (routes) ou des polygones (bâtiments, rivières, etc.), pour peu qu'il respecte les nomenclatures et les conventions. D'autre part, Maps, le service cartographique de Google, affiche en plus des couches d'occupation du sol, des points d'intérêts (POI) géolocalisés. Ces derniers représentent en général un établissement et contiennent des attributs spécifiant la nature de ce dernier (école, commerce, ou autre) et parfois leur nom. L'entreprise acquiert ces POI de deux manières : en collectant des bases de données officielles ou privées, et en mobilisant les utilisateurs du services, qui peuvent ajouter des POI en renseignant le nom et la catégorie, potentiellement vérifié par les équipes de Google. De plus le projet « Ground Truth » de Google vise entre autre la validation et la création automatique de POI en se basant sur des méthodes de machine learning qui permettent l'analyse des images obtenues par son service « Street View »². Il s'agit là de détecter automatiquement les enseignes des commerces et de les localiser sur une carte (Wojna et al., 2017).

Certes, ces bases ne sont pas aussi précises et exhaustives dans les grandes métropoles du Sud que dans celles du Nord, mais elles permettent de réduire progressivement les incertitudes sur les réalités de l'utilisation du sol, et ne sont pas agrégées. Bien qu'il soit délicat de quantifier leur niveau d'exhaustivité (Senaratne et al., 2017), il est probable que la quantité et la qualité des données soit meilleure chez Google, notamment lorsqu'il s'agit de commerces. Nous mobilisons toutefois ces données afin de faire ressortir différents niveaux d'attraction potentielle et de spécialisation.

Pour le cas de la métropole de Bangkok, nous avons élaboré un code sous python qui utilise les outils disponible dans l'API de Google Place³ pour récupérer les

¹ www.openstreetmap.com

² <https://research.googleblog.com/2017/05/Updating-Google-Maps-with-deep-learning.html>

³ <https://developers.google.com/places/web-service/search?hl=en>

différents POI dans une fenêtre spatiale mobile. Le nombre de requêtes quotidienne est limité à 2000 par clé fournies par le service, et après plusieurs jours 257924 POI ont été collecté en juillet 2015. Parmi ces POI, 43% d'entre eux ont un type d'établissement spécifié, dans l'une des 99 catégories renseignées - qu'il s'agisse d'une école, d'un spa, ou d'un centre commercial. Nous avons également téléchargé la base de donnée géographique du service libre et participatif OpenStreetMap (OSM). Nous avons enfin utilisé une image satellite Landsat 8 afin d'estimer le bassin de population.

2.2 Méthodes

2.2.1 Cartographie dasymétrique et ISODATA pour estimer la répartition de la population

Cartographier les densités de population à des échelles fines représente un enjeu lorsque l'information statistique fournie par les organismes du recensement ne correspond pas à l'échelle souhaitée. C'est par exemple le cas du recensement thaïlandais (2013) qui fournit l'information à l'échelle du khwaeng. Nous avons désagrégé cette information en utilisant la cartographie dasymétrique (Yuan, 1997). Cette méthode est principalement utilisée pour affiner des données démographiques en les distribuant d'une échelle large à une échelle plus fine. Schématiquement la méthode revient à ventiler les données de populations (ici le nombre d'habitants par arrondissement) dans les surfaces bâties, estimées par une classification non supervisée ISODATA à partir d'images Landsat 8.

2.2.2 Kmeans et CAH pour la typologie des assemblages de commerces

Peu d'études ont à notre connaissance utilisé des POI issus d'Internet pour réaliser des cartes d'utilisation du sol. Nous pouvons néanmoins citer les travaux de (Phithakkitnukoon et al., 2010) qui ont appliqué une Kmeans sur des POI de Yahoo regroupés dans des mailles de 500m, ou encore (Zhan et al., 2014) qui ont utilisé le même algorithme mais sur des POI issus de Foursquare dans des cellules de 200m. (Hu et al., 2016) ont réalisé une classification supervisée sur des données OSM en créant au préalable des couches de densités pour les différentes catégories. Dans un premier temps, nous regroupons les POI de Google orienté vers les activités commerciales et de services, soit environ 66800 points en 6 classes :

- les lieux d'hébergement (hôtel, guest house)
- les lieux de restauration et de loisirs bien définis (restaurant, café, pub, discothèques)
- les lieux de restauration moins bien définis (nourriture)
- les magasins non définis, ou seul l'attribut 'shop' figurait
- les magasins liés à l'apparat (salon de coiffure, magasins de vêtements, vendeur d'appareils électroniques, magasins pour la maison etc.)
- les entreprises diverses (réparation de voiture, entreprise spécialisée dans les appareils électriques ou électroniques etc.).

Ces POI sont ensuite agrégés dans un carroyage de cellule de 180m de côté. Ces dimensions ont été choisies car elles sont proportionnelles à l'unité de base de notre étude⁴ - un pixel d'une image landsat 8 de 30 m de côté - tout en constituant une nouvelle entité spatiale de la taille d'un pâté de maison.

La typologie des fonctions commerciales et de services est ensuite réalisée en deux étapes : (i) calcul des Kmeans et (ii) classification ascendante hiérarchique (CAH). La réalisation d'une classification non supervisée suivant l'algorithme des nuées dynamiques (Kmeans) en 500 classes (2000 itérations, 100 échantillons aléatoires) permet de faire un premier regroupement des mailles. Nous réduisons ainsi le nombre d'individus afin d'accélérer les traitements ultérieurs, et le nombre élevé de classes permet d'optimiser le regroupement sans perdre trop d'information. En nous inspirant de Fleury et al, 2012, nous effectuons ensuite une CAH selon la méthode de la variance minimum de Ward (Ward, 1963) (figure 1), en utilisant la distance du diamètre maximum afin de maximiser la distance entre deux individus appartenant à deux classes distinctes. Compte tenu du fait que dans la plupart des cas, les branches du dendrogramme ne sont pas équilibrées en termes de nombre de classes regroupées, nous avons décidé d'adapter ce dernier. L'objectif est de conserver la structure hiérarchique et d'associer les branches en fonction d'un nombre de maille maximum par cluster. La première étape est une homogénéisation des niveaux, avec un maintien de la structure hiérarchique globale. Les différents niveaux sont ici gardés à titre indicatif, mais ne reflètent plus la distance réelle entre les différents groupes. La seconde étape est un regroupement des branches de manière à ce qu'elles aient un nombre de cluster maximum, ce qui permet d'éviter qu'une « coupe » du dendrogramme entraîne la création de classes contenant un nombre trop faible ou trop grand d'individus (figure 1).

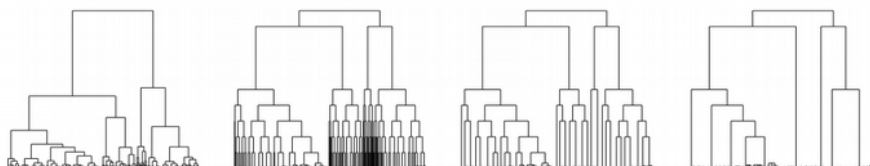


Figure 1 : dendrogramme original (gauche), et dendrogrammes avec niveaux homogénéisés avec au maximum respectivement 1, 20 et 100 clusters par branches finales

Cette méthode, même si elle fait perdre la notion d'inertie et donc la part d'information résumée, permet de garder la structure des classes de niveau inférieur, et donc d'éviter d'obtenir des classes contenant un faible nombre d'individus d'un côté, et un regroupement de classes présentant un grand nombre d'individus de l'autre. On s'affranchit en quelque sorte des distances inter-classes trop faible, en gardant la structure en grappe originale. Nous avons ensuite défini de manière

⁴ Ce travail a été réalisé dans le cadre du programme européen DENFREE.

empirique un critère d'agrégation (taille minimum du cluster, ici 100) et un nombre de classe désiré (ici 11). Les 11 classes obtenues sont ensuite regroupées en 6 grandes classes caractérisées respectivement par une sur-représentation (i) des lieux d'hébergements, (ii) des lieux de loisirs, shopping, (iii) des lieux de shopping, (iv) des commerces indifférenciés, (v) des activités du secteur secondaire (activités liées aux voitures, entreprise d'électronique etc.) et (vi) des zones indifférenciées contenant peu de commerces. La validation a été effectuée en confrontant les résultats de la classification à la connaissance du terrain. Cette méthode, qualifiable de semi-supervisée est applicable à d'autres types de données matricielles (e.g images satellites).

2.2.3 Fusion des Google POI et d'OSM

Alors que les commerces sont d'étendue spatiale en général assez limitée, en général inférieur à notre maille de 180m, ce n'est pas le cas pour les lieux d'éducatifs, de santé (hôpitaux), de loisir (parc d'attractions) ou encore religieux (grands temples) et administratif. Ainsi, un de ces types d'utilisation du sol peut occuper plusieurs mailles, ce que ne reflètent pas les POI. Nous avons donc réalisé une fusion entre les POI de Google et les polygones d'OpenStreetMap regroupés selon la typologie d'activité évoquée ci-dessus. La fusion des bases de Google et d'OSM fût ensuite effectuée en deux temps. Tout d'abord nous avons intersecté les polygones d'OSM avec les cellules de notre grille puis nous avons ajouté les POI de Google. Au final chaque cellule contient une valeur booléenne en fonction de la présence ou de l'absence d'un type d'utilisation du sol et certaines mailles peuvent contenir plusieurs catégories.

2.3 Résultats

La cartographie dasymétrique offre l'avantage de mettre en évidence les grandes disparités des densités de population à Bangkok à une échelle très fine de 30*30 mètres. La figure 2 révèle à la fois un centre urbain densément peuplé et une périphérie plus rurale irriguée par des axes de transport le long desquels les implantations humaines apparaissent clairement, notamment dans l'est de la municipalité. Sans surprise cette carte des densités des populations est le négatif de la carte des espaces végétalisés (Misslin, Daudé, 2016).

La carte des activités économiques est réalisée en agrégeant les POIs dans des mailles, ici de 180 m de côté (figure 3). Cette répartition des commerces et services dans la ville, tant du point de vue de leur nombre que de leur diversité, fait également ressortir les disparités entre le centre et la périphérie, et le rôle de la proximité avec un axe de communication dans leur répartition.

La typologie des activités fonctionnelles (figure 5) montre l'organisation de Bangkok sous forme d'assemblages, diffus en périphérie et plus dense et spécialisé dans le centre ville. Globalement, les zones contenant quelques commerces forment une sorte de bruit de fond de la ville, que nous pouvons interpréter comme des commerces de proximités, auxquels se juxtaposent sporadiquement quelques zones

ou les activités commerciales sont plus importantes. Mais dans le grand centre apparaît des niveaux de spécialisation nettement plus marqués, qu'il s'agisse des lieux d'hébergements caractéristiques des lieux touristiques (Khao San, Sukhumvit) ou d'affaire (Silom), des zones où les commerces sont extrêmement nombreux et variés, comme dans les malls de Siam Paragon ou de Central World, ou dans l'un des plus grand marché asiatique qu'est le marché du week end de Chatuchak. Le quartier chinois de Yaowarat est quant à lui divisé en deux, une partie contenant des commerces indifférenciés, et l'autre une représentations des ateliers et petites entreprises, ce qui reflète la réalité du terrain.

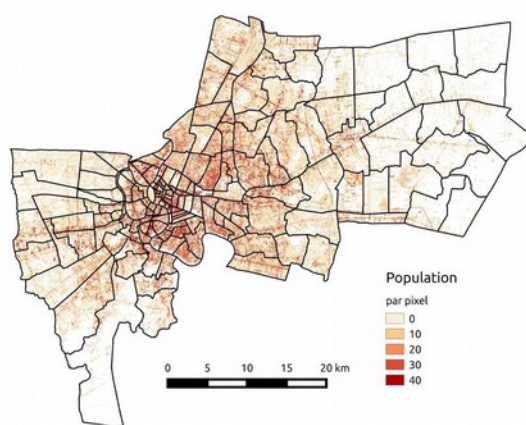


Figure 2: Estimation des densités de population à Bangkok par pixel de $30m^2$ (source: Recensement, méthode: cartographie dasymétrique, d'après Misslin, Daudé, 2016)

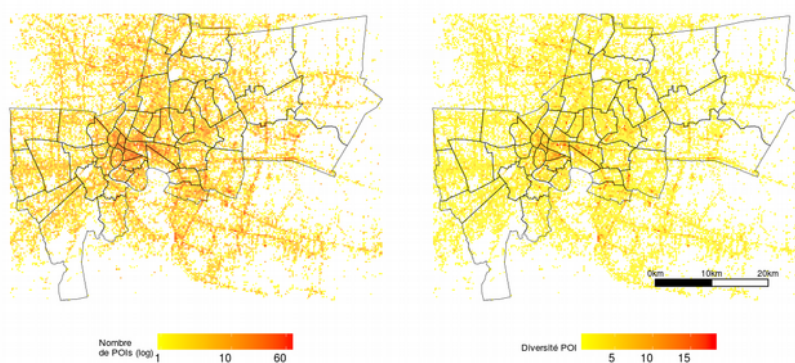


Figure 3: Distribution (a) du stock des activités économiques et (b) de leur diversité à Bangkok.

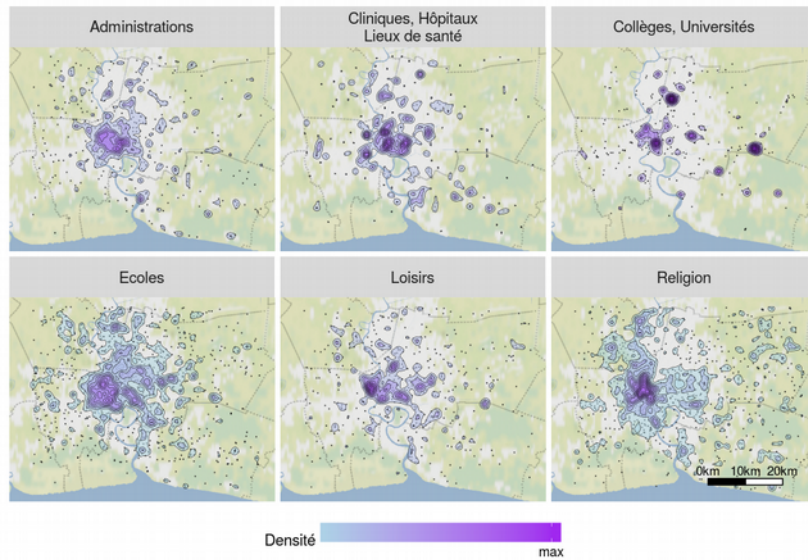


Figure 4: Exemples des densités de lieux d'activités à Bangkok. Le passage d'une isoligne à une autre correspond à une progression de +/- 10% (source : Google POI et OSM). Réalisé sous R avec ggplot2, ggmap et ggsn.

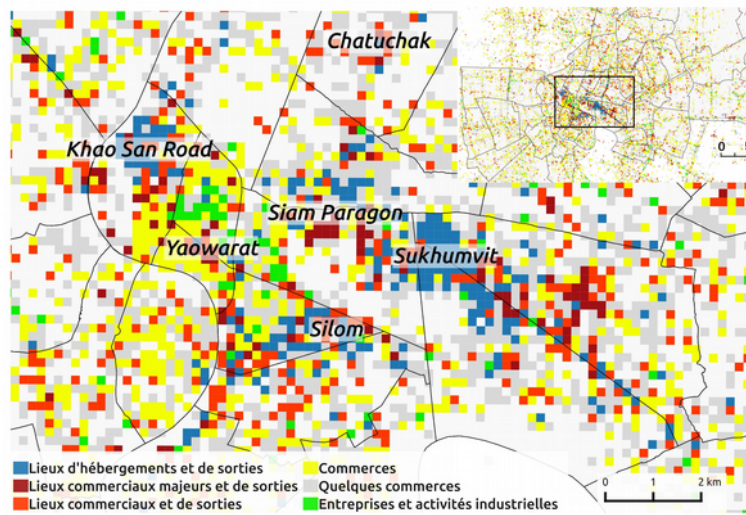


Figure 5: Typologie des activités économiques à Bangkok et dans le centre ville

Cette organisation spatiale des activités économiques (figure 5), culturelles et sociales (figures 3 & 4) suggère donc des potentiels, des motivations et des profils de déplacement bien différents pour la population (figure 2), que nous allons aborder à partir des traces numériques laissées par les habitants de Bangkok.

3. Dynamiques spatio-temporelles des mobilités quotidiennes à Bangkok

Les espaces urbains sont caractérisés par des zones d'émissions et de réceptions de population, et les flux domicile-travail en sont une illustration. Une zone attractive la journée peut devenir une zone déserte la nuit selon son niveau de spécialisation dans une activité donnée. Ainsi les zones « dortoirs » se vident la journée au profit des zones commerciales et industrielle, tendance qui s'inversent la nuit. La ville est donc traversée de « couloirs » de circulation pour ces flux qui pourraient favoriser le déplacement de virus en période épidémique.

3.1 Données

Dans l'objectif de quantifier ces flux, nous utiliserons ici des données géolocalisées issues de site de micro-blogging Twitter a déjà permis par le passer de quantifier les déplacements à l'échelle nationale et internationale (Hawelka et al., 2014; Jurdak et al., 2015) et urbaine (Cebeillac et al., 2017; Khan et al., 2017; Luo et al., 2016). Ces données, qui offrent l'avantage d'être assez massive et facilement accessible⁵ ont cependant des limites, notamment en terme de représentativité vis-à-vis de la population (Cebeillac et al., 2017; Khan et al., 2017) et de l'individu (sont-elles conformes à l'espace réellement vécu par la personne?).

Nous avons enregistré ici les données publiques et géolocalisées des utilisateurs de Twitter en Thaïlande, en limitant l'extraction des données au pseudonyme, l'horodatage et les coordonnées précises générées lors de l'envoi d'un tweet. L'intérêt de l'étude ne portant pas sur une typologie des utilisateurs de Twitter, nous assurons un niveau d'anonymat des utilisateurs en chiffrant les pseudonymes et en ajoutant un flou temporel et géographique au tweet enregistré (Cebeillac et al., 2017). Ce travail de collecte, de sauvegarde et de traitement de données est en cours d'enregistrement au registre de la CNIL. La période d'enregistrement des données s'étale du 25 juin 2014 au 4 décembre 2015 pour un total de 527 jours. L'emprise spatiale est la zone élargie de Bangkok. Au terme de la période, 20 126 426 tweets pour 308 808 utilisateurs ont été récoltés, pour la zone de Bangkok.

3.2 Méthodes

Nous avons dans un premier temps nettoyé la base de donnée en filtrant parmi les tweets envoyés ceux qui proviennent de bots (automatisation d'envois de tweets par des algorithmes) et ceux qui proviennent d'utilisateurs peu réguliers (qui ont

⁵ <https://dev.twitter.com/streaming/overview>

tweeté moins de n fois sur la période). Nous avons ensuite analysé pour chaque utilisateur son espace d'activité, c'est-à-dire l'ensemble des lieux qu'il fréquente, à différents moments de la journée, et à différents jours de la semaine. Ce travail nous a permis de caractériser pour environ 35 000 utilisateurs le lieu d'habitation (présence régulière au même endroit entre 20:00 et 07:00), et leur espaces d'activités.

Nous avons ensuite estimé les interactions entre les différents arrondissements de Bangkok via la probabilité de présences par khwaeng, définie ici comme le rapport entre le nombre de jours où une personne a tweeté dans un arrondissement et le nombre total de jours où elle a été active sur le réseau social. Ces pourcentages sont ensuite sommés par zone d'origine formant une matrice de probabilité d'interaction, similaire à une matrice origine-destination. Nous avons ensuite cartographié pour chaque arrondissement les 3 flux principaux, sans pondérer par le nombre d'utilisateur de Twitter vivant dans la zone afin de garder les valeurs absolues démographiques (figure 7). Les différentes couches d'utilisation du sol ont été intersectées avec les tweets datés et géoréférencés afin de faire ressortir les horaires de fréquentation de chacune des différentes classes (figure 8).

3.3 Résultats

La figure 6 illustre les principaux flux de mobilité des utilisateurs de Twitter un mardi type. Elle met en évidence globalement la structure centre-périphérie de la ville identifiée dans les cartes de la figure 3, notamment celles des administrations ou des écoles, et localement permet d'identifier des zones d'attractivités plus éloignées du centre, telles que les universités ou les lieux de loisirs.

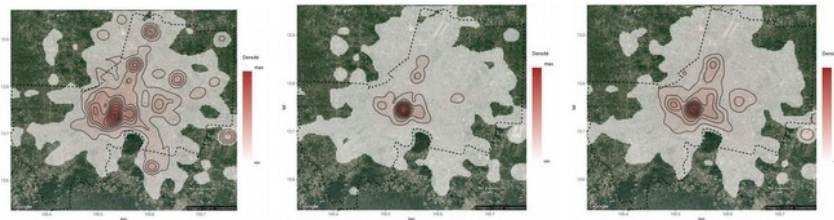


Figure 6: Densités moyenne des utilisateurs de Twitter le mardi entre 6:00-8:00, 14:00-16:00 et après 22:00

On retrouve cette forte attractivité du centre dans la représentation des flux à partir du khwaeng du domicile vers d'autres khwaeng (figure 7). Globalement, la structure étoilée montre de fortes interactions spatiales entre la périphérie et le centre. Cette carte montre également que le quart nord-est de la première couronne de Bangkok est concurrent en termes d'attractivité au centre. En effet le poids des mobilités locales, c'est-à-dire qui sont internes à la zone de résidence, y est beaucoup plus élevé que dans les autres zones (cf. taille des cercles dans la figure 7). Au-delà de cette structure étoilée centre-périphérie apparaissent le poids des

interactions locales, de proximité entre khwaeng contiguës. On observe également un nombre de khet visités assez faible pour les districts situés dans le centre et à l'est. Ceci pourrait s'expliquer pour les premiers par une concentration des activités, ce qui réduit la nécessité de déplacement, et pour les seconds par un relatif enclavement.

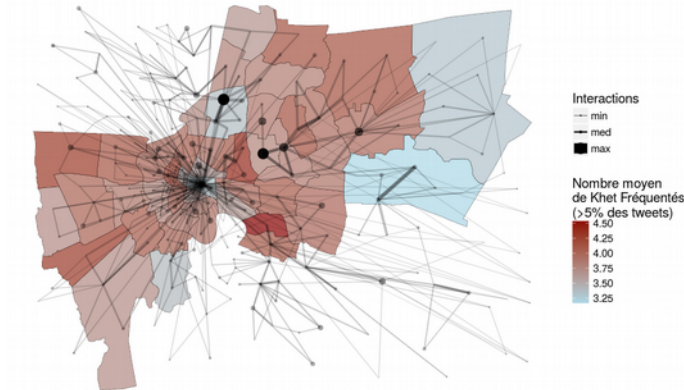


figure 7 : les 3 principales interactions pour chaque arrondissement, en valeur absolue. La taille des cercles symbolise l'importance des flux locaux.

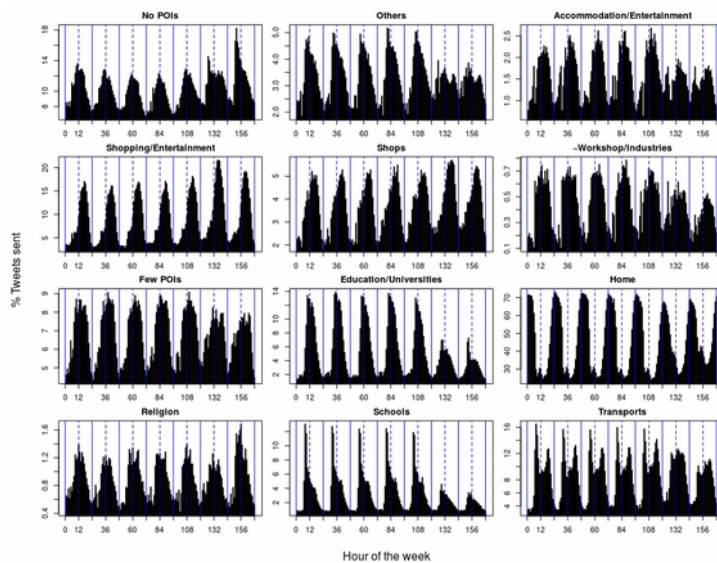


figure 8: Présence des utilisateurs de Twitter (fréquence de Tweets envoyés) dans différents lieux (encadrés) par tranche horaire (histogrammes en heure, les tirets représentent midi et le trait plein minuit)

Les motifs de ces mobilités peuvent être inférés en croisant cette base de données avec celles des Google POI et des données OSM décrites précédemment. En effet, en ayant qualifié localement le profil des activités exercées en un lieu, on pose l'hypothèse qu'un individu qui se rend régulièrement en ce lieu a une probabilité d'y exercer l'une ou l'autre de ces activités. Si la proximité de cette hypothèse avec une erreur écologique est forte, la force du concept peut cependant se lire dans les différents graphiques de la figure 8 : En soirée, une proportion élevée de Tweet (plus de 70%) est envoyé depuis le domicile; Le double pic des transports en commun illustre le départ (entre 5:00 et 8:00) et le retour (entre 16:00 et 19:00) du et au lieu de résidence; Les autres moments de la journée sont ensuite distribués, par vagues, entre différentes activités : certains vont à l'école, d'autres dans des zones commerciales ou dans des entreprises.

4. Discussion et Conclusion

L'apport d'informations sur les activités présentes dans le territoire de Bangkok et sur les mobilités quotidiennes de ses habitants ont permis d'éclairer les contextes des différents profils de mobilités : Quels types de lieux sont les plus fréquentés ? À quels endroits ? À quelles heures ? D'où viennent les individus qui fréquentent ces zones ?

La classification des bases de données géographiques POIs de google et OSM a permis de caractériser la densité et la diversité des services. Les données et méthodes employées ont révélé de fortes discontinuités dans la répartition et l'étendue géographique des divers assemblages de zones fonctionnelles, ce qui n'était pas évident en l'absence de données officielles adaptées. L'utilisation des tweets géolocalisés nous a permis d'associer à chaque individu son lieu de domicile hypothétique et les lieux qu'il fréquente dans la ville. Agrégés spatialement aux différentes couches d'utilisation du sol, nous obtenons des horaires de fréquentation moyennes pour chacun des types de lieux de notre typologie.

A ces discontinuités spatio-temporelle dans les fréquentations des différents arrondissements et services de la ville, peut également s'ajouter des discontinuités dans l'incidence de l'épidémie (figure 9). Ainsi, les zones ne sont pas affectées de la même manière dans le temps par la dengue, le grand centre étant globalement le plus touché, surtout entre 2011 et 2013. La comparaison avec les principales interactions (figure 7), suggèrent que l'afflux d'un grand nombre de personnes dans l'hyper centre pourrait être une explication, tout comme le niveau de connectivité des différents districts et l'importance des flux locaux pour les zones au nord est de la première couronne. Une analyse plus fine et poussée, notamment par l'intermédiaire de modèles à base d'agent orienté sur les activités pourrait potentiellement affiner ces rapides observations.

Ces premiers pas vers une meilleure connaissance des déplacements individuels et de l'attractivité différenciée des diverses zones de Bangkok nous permettent d'espérer prochainement dépasser une étape de la géographie de la santé, inaugurée par Mansell Prothero il y a 50 ans -à propos des liens entre malaria et mouvements

humains- qui s'est toujours trouvée limitée par le manque d'informations sur les flux de populations. Aujourd'hui de nouvelles perspectives s'ouvrent qui devraient permettre à terme de faire sauter des verrous de la recherche. En particulier, de nouvelles sources d'informations issues de grandes bases de données pourraient permettre d'éclairer d'un nouveau jour les liens unissant l'épidémiologie et les discontinuités.

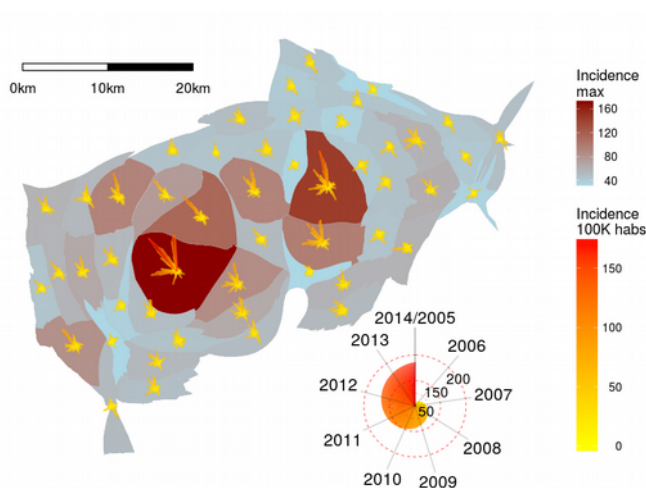


Figure 9 : Incidence de la dengue par district entre 2005 et 2014. Carte par anamorphose, où la taille de chaque district est proportionnel à l'incidence maximum. L'angle de chaque trait définit le mois dans la période et la longueur et la couleur correspondent à l'incidence

5. Références

- Cebeillac A., Daudé É., Huraux T. (à paraître), Where? When? And How Often? What Can We Learn on Daily Urban Mobilities From Twitter Data and Google Map in Bangkok (Thailand) and Which Perspectives For Dengue Studies? *Netcom*.
- Daudé É. (2017), Les Systèmes Pathogènes Complexes, Modélisation et Simulation. *Revue Francophone des Territoires et de la Santé*, 5 p., <https://halshs.archives-ouvertes.fr/halshs-01501488/document>
- Daudé É., Mazumdar S., Solanki V. (2017), Widespread Fear of Dengue Transmission but Poor Practices of Dengue Prevention: a Study in the Slums of Delhi, India. *PLoS ONE* 12(2): e0171543. doi:10.1371/journal.pone.0171543.
- Daudé É., Mazumdar S. (2016), *Combating Dengue in India: Challenges and Strategies*, EPW: Economic & Political Weekly, vol. 51, n. 8, pp.77-81, <http://www.epw.in/journal/2016/8/notes/combating-dengue-india.html>

- Daudé É., Vaguet A. (2015), *Surveillance, contrôle et épidémies de dengue en Inde : qui a échoué ?* l'Espace Politique, vol. 26, n. 2, <https://espacepolitique.revues.org/3485>, doi:10.4000/espacepolitique.3485.
- Fleury A., Mathian H., Saint-Julien T., (2012) « Définir les centralités commerciales au cœur d'une grande métropole : le cas de Paris intra-muros », *Cybergeo : European Journal of Geography* URL : <http://cybergeo.revues.org/25107> ; DOI : 10.4000/cybergeo.25107
- Gatrell T. (2011), *Mobilities and Health*, Ashgate, 220 p.
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Carlo, R., (2014). Geo-located Twitter as the proxy for global mobility patterns. *Cartography and Geographic Information Science*.
- Hu, T., Yang, J., Li, X., Gong, P., (2016). Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sensing* 8, 151. doi:10.3390/rs8020151
- Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D., (2015). Understanding human mobility from Twitter. *PloS one* 10, e0131469.
- Khan, S.F., Bergmann, N., Jurdak, R., Kusy, B., Cameron, M., (2017). Mobility in Cities: Comparative Analysis of Mobility Models Using Geo-tagged Tweets in Australia.
- Luo, F., Cao, G., Mulligan, K., Li, X., (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography* 70, 11–25. doi:10.1016/j.apgeog.2016.03.001
- Maneerat S., Daudé É. (2017), Étude par simulation à base d'agents des effets des discontinuités intra-urbaines à Delhi sur la dispersion des moustiques *Aedes aegypti*, vecteurs de la dengue, de la fièvre jaune, du chikungunya et du virus Zika. *Cybergéo : European Journal of Geography*, à paraître.
- Maneerat S., Daudé É. (2016), A spatial agent-based simulation model of the dengue vector *Aedes aegypti* to explore its population dynamics in urban areas. *Ecological Modelling*, vol. 333, pp.66-78, <http://dx.doi.org/10.1016/j.ecolmodel.2016.04.012>.
- Misslin R., Telle O., Daudé É., Vaguet A., Paul R. (2016), *Urban climate versus global climate change - what makes the difference for dengue?* Annals of the New York Academy of Sciences, 17p., doi: 10.1111/nyas.13084.
- Misslin R., Daudé É. (2016), Génération d'environnements artificiels pour la simulation spatiale d'arboviroses en milieu urbain : application à la dengue et au virus Zika. *SAGEO'16, 6-9 décembre 2016, Nice*.
- Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., Ratti, C., (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data, in: *International Workshop on Human Behavior Understanding*. Springer, pp. 14–25.
- Prothero M. (1965), *Migrants and malaria in Africa*. University of Pittsburgh Press.
- Reiner, R.C., Stoddard, S.T., Scott, T.W., (2014). Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal. *Epidemics* 6, 30–36. doi:10.1016/j.epidem.2013.12.003
- Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M. Muki, (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science* 31, 139–167. doi:10.1080/13658816.2016.1189556

- Telle O., Vaguet A., Yadav N.K., Lefebvre B, Cebeillac A., Nagpal B., Daudé É., Paul R. (2016), The spread of dengue in an endemic urban milieu – the case of Delhi, India. *PloS ONE*, 11(1): e0146539, p. 17, doi:10.1371/journal.pone.0146539
- Ward, J. H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, 58, 236–244
- Wojna, Z., Gorban, A., Lee, D.-S., Murphy, K., Yu, Q., Li, Y., Ibarz, J., (2017). Attention-based Extraction of Structured Information from Street View Imagery. *arXiv preprint arXiv:1704.03549*.
- Yuan, Y., R.M. Smith, and W.F. Limp (1997). Remodeling census population with spatial information from LANDSAT TM imagery. *Computers, Environment and Urban Systems* 21(3/4): 245-58
- Zhan, X., Ukkusuri, S.V., Zhu, F., (2014). Inferring Urban Land Use Using Large-Scale Social Media Check-in Data. *Networks and Spatial Economics* 14, 647–667. doi:10.1007/s11067-014-9264-4