

Rapport de Stage de Travail de Fin d'Etudes : Analyse des liens entre cultures et formes urbaines par utilisation de la télédétection

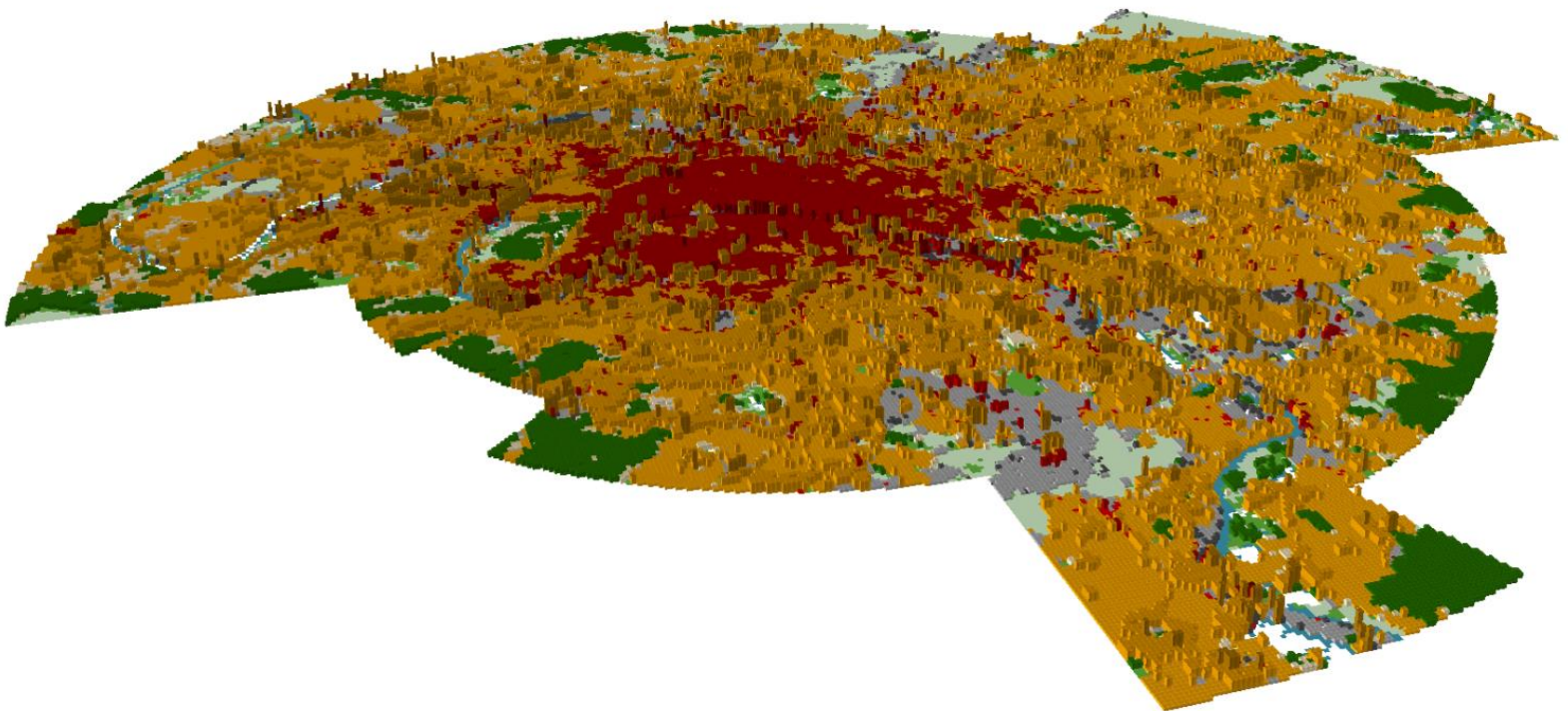


Figure 1: Morphologie de Paris selon une classification en Local Climate Zones.

Table des matières

I.	Introduction.....	3
A.	Abstract/Résumé.....	3
B.	Acknowledgments/Remerciements.....	4
II.	Déroulement du stage.....	5
A.	Organisme d'accueil.....	5
B.	Intégration dans l'équipe et dans le monde de la recherche.....	5
C.	Présentation de la mission.....	6
III.	Contexte d'étude.....	7
A.	Motivation de l'étude.....	7
B.	Etat de l'art.....	7
1.	Région culturelle.....	7
2.	Aire urbaine.....	8
3.	Analyse spatiale appliquée à la morphologie urbaine.....	9
4.	Influence de la culture sur la forme d'une ville.....	9
C.	Outils utilisés.....	11
IV.	Jeux de données.....	11
A.	Classification « Local Climate Zone » d'une centaine de villes.....	11
1.	Local Climate Zone.....	12
2.	Production de la classification.....	13
3.	Limites du produit.....	14
B.	Frontières morphologiques des villes.....	14
1.	Principes et avantages.....	15
2.	Méthodologie de production.....	15
3.	Limites.....	17
C.	Innovation proposée par l'étude.....	19
V.	Méthodologie.....	20
A.	Première approche.....	20
1.	Préparation des données.....	20

2.	Première analyse statistique	20
3.	Premières conclusions	21
B.	Méthodes et algorithmes	21
1.	Présentation des algorithmes de partitionnement de données	21
2.	Présentation des différentes approches statistiques.....	27
3.	Mode opératoire des expériences	29
VI.	Expériences et résultats	30
A.	Méthode d'évaluation des résultats	30
B.	Résultats des expériences	30
1.	Approche statistique par composantes principales.....	30
2.	Approche statistique par paramètres sémantiques.....	32
3.	Approche statistique par distribution géographique de la densité	34
4.	Approche statistique par distribution géographique des paramètres sémantiques	35
C.	Discussion	36
VII.	Conclusion	38
	Glossaire	39
	Références.....	39
	Liste des figures et tableaux.....	43
	Annexes	45

I. Introduction

A. Abstract/Résumé

D'un pays à l'autre, d'une ville à l'autre, les impressions sensibles du milieu urbain que nous avons différent. Les formes, les couleurs changent. Mais ces différences n'apparaissent pas aussi grandes entre deux villes proches dans le monde qu'entre deux villes sur deux continents différents.

La télédétection et ses produits permettent aujourd'hui de mener des études sur les villes à l'échelle mondiale plus facilement. Le Centre Aérospatiale Allemand (*DLR*) a récemment créé en interne deux jeux de données à l'échelle mondiale décrivant certains aspects de la morphologie urbaine. Ces produits sont une classification de l'occupation du sol urbain et une approximation de l'étalement des villes.

Ce rapport présente les résultats de mon stage de fin d'étude, clôturant mon cursus en écoles d'ingénieur, qui s'est déroulé au sein du Centre d'Observations Terrestres (*EOC*) du *DLR* sous la direction du Dr. Hannes Taubenböck. Mon objectif a été de proposer et effectuer une méthodologie de traitement de ces données pour quantifier l'influence qu'a la culture sur la forme urbaine.

Nous avons mené notre étude sur 110 villes à travers toutes les régions du globe. Nous proposons une approche basée sur un prétraitement par Systèmes d'Information Géographique (*SIG*) automatisé, une analyse par algorithmes de classification non supervisés, une qualification des résultats. Les algorithmes de *K-means* et *EM* se sont révélés être les plus satisfaisants et proposaient un partitionnement des villes du monde proche des régions culturelles proposées par les ethnologues et historiens. Les résultats de cette approche apportent des informations sur les types de ville aux échelles globale et locale.

Thesaurus

Morphologie urbaine – Sphère d'influence culturelle – Local Climate Zones – Morphological Urban Areas – Télédétection – Machine Learning

From one country to the next, from a city to the next, the sensible feeling of the urban atmosphere differs. Shape, forms, shadows, and colors change. But the closer the distance between two cities, the smaller the differences are.

Nowadays, remote-sensing and earth observations products allow easy city analysis at a global scale. The German Aerospace Center (*DLR*) recently and internally produced two global datasets on urban morphology: an urban land cover classification and a set of boundaries for cities sprawl.

This report presents the outcomes of my final internship at the Earth Observation Center of the *DLR* supervised by Dr. Hannes Taubenböck. This internship is the ending milestone of my engineer's curriculum. My main goal during this internship was to develop and apply a framework to process those

data in order to analyze the influence of culture on the morphological structure of cities.

We led this survey on 110 cities all across the world. We propose here an approach based on an automated GIS preprocessing and an unsupervised clustering of the dataset. *K-means* and *EM* proved themselves to be better suited for this task. The clustering they defined came close to the dividing of cultural spheres proposed by ethnologist and historian. Those results bring a new scope on the typologies of cities at both global and local scales.

Keywords

Urban morphology – Kulturkreis – Local Climate Zones – Morphologic Urban Areas – Remote-sensing – Machine learning

B. Acknowledgments/Remerciements

J'aimerais tout d'abord remercier Hannes Taubenböck pour son encadrement très inspirant et pour sa confiance de me confier ce projet. J'aimerais remercier Jeroen Staab pour son accompagnement administratif et pour son courage d'avoir fait de perpétuels allers-retours entre les systèmes administratifs allemand et français. Je souhaite remercier Patrick Aravena Pelizari et Christian Knopp pour le travail formateur effectué ensemble sur leurs projets. Je tiens tout particulièrement à remercier Monika Kuffer et mes autres anciens collègues de l'ITC, Divyani Kohli, Caroline Gaevert, Karin Pfeffer, pour toutes leurs recommandations qui m'ont permis d'effectuer mon stage et bientôt ma thèse au DLR, pour leur coaching combien enrichissant et pour leur sympathie permanente. Je remercie aussi Cécile Haeck pour son accompagnement bienveillant à travers ce stage. Enfin j'aimerais remercier les autres stagiaires pour leur esprit de groupe et leur volonté de m'intégrer au cours de mon stage.

First, I would like to thank Hannes Taubenböck for his inspiring mentorship and for the trust he put in me with this project. I would like to thank Jeroen Staab for the management of my administrative case and for the dedication he showed doing all the back-and-forth between the French administrative system and the German one. I also wish to thank Patrick Aravena Pelizari and Christian Knopp for the educational work accomplished together on their projects. Very especially, I would like to thank Monika Kuffer and my other former colleagues at the ITC: Divyani Kohli, Caroline Gaevert and Karin Pfeffer, for all their nice recommendations which got me in the DLR for my internship and soon my Ph.D. thesis, for their rewarding coaching and their constant sympathy. I also thank a lot Cécile Haeck for her truly kind educational support. Finally, I would like to thank all the other interns for their team spirit and their wish to integrate me along with during this internship.

II. Déroulement du stage

A. Organisme d'accueil

Le DLR (*Centre allemand pour l'aéronautique et l'astronautique*) est l'équivalent allemand du CNES et de l'ONERA français et de la NASA américaine. Répartie en 14 sites à travers le pays, l'agence couvre des missions allant de la recherche fondamentale à l'exploitation sur les domaines de l'aéronautique, de l'énergie et du transport. Ses deux plus grands sites sont : Cologne (Centre d'entraînement des astronautes européens et centre administratif), Oberpfaffenhofen-Munich (Centres de télédétection, Centre de robotique et mécatronique, Centre des opérations du module européen de l'ISS – *International Space Station*, ...) (1).

Le DLR regroupe 8 000 employés dans 33 instituts spécialisés. Son budget de recherche interne, en grande partie dédié à l'agence spatiale européenne, s'élevait en 2017 à 1,536 milliards d'euros (2).

Le stage que j'ai effectué se déroulait au sein de l'institut « *Deutsche Fernerkundungsdatenzentrum* » (DFD) – *Centre national allemand des données de la télédétection*, à Oberpfaffenhofen, dans la banlieue sud-est de Munich (Bavière). Avec l'institut de télédétection, ils forment le Centre des observations de la Terre. Ses missions sont orientées sur l'utilisation des données acquises par satellite dans : Le développement de méthodes et processus de cartographie de régions en crise ; Développement et déploiement de produits tels que les cartes d'occupation du sol, données de variables géophysiques ou biophysiques ; Analyse des sociétés et des villes, de leurs écosystèmes et des risques auxquels elles sont soumises (3).

B. Intégration dans l'équipe et dans le monde de la recherche

C'est dans ce dernier segment que j'ai pu travailler, dans l'équipe « *Villes et sociétés* » sous la direction du Dr. H. Taubenböck. L'équipe se compose actuellement d'une vingtaine de membres et de quelques stagiaires et étudiants en rédaction de thèse de Master et de Licence. La synergie dans laquelle s'inscrit le travail des stagiaires est axée sur le développement de nouvelles compétences et la production de données utiles au reste de l'équipe. Cet objectif d'enseignement se reflète dans la politique de l'équipe de reprendre les stagiaires année après année et de les accompagner dans leurs rédactions de thèse de Master puis de Licence. Les projets et articles sur lesquels l'équipe travaille sont réalisés par groupes de 2 à 5 personnes généralement, les groupes changeants en fonction du projet en question. Cela permet à l'ensemble de l'équipe, stagiaires compris, d'être brassés et de travailler dans un bon esprit. L'équipe étant en majorité allemande mais avec des éléments internationaux, les langues de communication principale pour moi était l'Anglais pour tous les sujets techniques et l'Allemand pour le reste de la communication interpersonnelle.

Cette équipe a ses axes de travail équilibrés entre la production et la recherche. Ayant pour but de continuer mes études en doctorat, c'est une chance pour moi d'avoir pu travailler principalement sur la composante recherche. Les attentes que nous avons, la direction de l'équipe et moi-même, est la

publication d'un article sur mon projet dont je serais le premier auteur. Au cours de mon stage, j'ai eu le plaisir de participer à la conférence Joint Urban Remote Sensing Event – *JURSE*. Cette conférence biennale réunit la communauté internationale des spécialistes de la télédétection travaillant sur les thèmes liés à la Ville (4). Ma participation à cette conférence était motivée par ma présentation d'un article que j'ai écrit avec M. Kuffer (Dr. Ass. Prof.), C. Persello (Dr. Ass. Prof.), C. Klaufus (Dr. Ass. Prof.) et K. Pfeiffer (Full Prof. Dr.) l'année précédente au cours de mon stage étude et recherche à l'ITC (University of Twente, Pays-Bas) (5). Par ailleurs mon maître de stage, H. Taubenböck, est membre du comité de pilotage de cette conférence et m'a permis de me mettre en contact avec de nombreux membres de cette communauté.

L'intégration dans mon équipe et dans le monde de mon domaine de recherche s'est très bien passée et se concrétisera en Novembre 2019 par le début de ma thèse de doctorat dans mon équipe actuelle.

C. Présentation de la mission

A partir de 2 jeux de données concernant la morphologie urbaine récemment développés par le DLR, ma mission était de proposer une méthodologie permettant de voir si l'on pouvait en extraire des informations sur la culture des villes. Ce projet est un projet de recherche qui avait pour but pour le DLR de trouver des applications de test pour leurs jeux de données. Le thème du projet avait été proposé par mon maître de stage, le Dr. Hannes Taubenböck, chef de l'équipe Villes et Sociétés. Nous avons défini l'axe de recherche ensemble. Par la suite, j'ai réalisé cette mission de manière autonome, incluant présentations des résultats intermédiaires, proposition d'améliorations et de nouvelles pistes. Ce projet a été terminé dans le temps de mon stage et sera suivi pour moi de la corédaction d'un papier scientifique sur le sujet sous la direction de H. Taubenböck.

En parallèle, d'autres tâches m'ont été confiées, en relation avec les activités de l'équipe : préparation de présentations des résultats de l'équipe dans des projets ; production d'estimation de la distribution de la population à l'échelle des bâtiments pour l'ensemble de la Bavière en vue d'une estimation de la sensibilité des populations aux inondations ; classification de plus de 60 000 images Google Street View de la ville de Santiago de Chile pour entraîner un réseau de neurone à classifier les bâtiments en fonction des matériaux de construction et de leur structure.

III. Contexte d'étude

Ce projet de recherche vise à déterminer si la structure morphologique d'une ville est liée à la zone culturelle dans laquelle se trouve cette ville. Si la morphologie urbaine peut être décrite de manière géométrique relativement simplement, la « culture » de la ville et *a fortiori* sa « culture urbaine » restent plus difficilement accessibles. A partir de données d'occupation du sol, comment pourrait-on extraire ce type d'information ? Ici, nous présenterons tout d'abord l'état de l'art et les différents outils mis à notre disposition.

A. Motivation de l'étude

Les aires urbaines abritent, aujourd'hui, plus que le seuil symbolique des 50% de la population mondiale et accueilleront en 2050 plus des deux tiers de cette population (6). Les préoccupations portant sur ces milieux sont au cœur des débats sur le développement durable. Il s'agit d'un des 17 thèmes avancés dans ce domaine par l'ONU, et de nombreux thèmes lui sont rattachés (7). La ville est un pivot sur lequel il est important d'agir dans cette perspective de développement durable.

Il a été montré que la morphologie des villes avait de grands impacts sur différents aspects de la vie de ses habitants et plus particulièrement sur la santé publique (8) (9) (10) : activité physique (11), exposition à la pollution (12), mais aussi santé mentale et sociale (13) (14) (15). Les études sur ces derniers points montrent l'importance de la prise en compte de la psyché des habitants d'une ville et de sa relation avec les formes des villes.

Il est donc important de comprendre la morphologie urbaine sous tous ces aspects dans le processus de planification urbaine auquel nous sommes confrontés dans ce contexte démographique urbain. Nous pensons que c'est dans l'étude de la morphologie actuelle et passée que doit se forger la compréhension de ses impacts sur notre mode de vie.

La forme d'une ville est impactée par plusieurs facteurs (16) (17) (18) et nous pensons qu'il n'existe pas de solution globale de planification urbaine. Cette planification se doit d'être menée en accord avec la psyché de ses habitants. Un des paramètres importants déterminant de cette psyché est la culture à laquelle l'habitant appartient. C'est pourquoi, il nous semble être pertinent de mener une étude sur le lien entre cette culture et la morphologie urbaine qu'elle produit afin de comprendre plus globalement l'importance de ce paramètre local dans le contexte de la planification urbaine.

B. Etat de l'art

1. Région culturelle

Comme écrit plus haut, nous nous intéressons à l'impact culturel régional sur la morphologie urbaine. Ici, par « région culturelle », nous entendons une portion de la surface de la Terre, d'échelle

subcontinentale, peuplée par des populations partageant cette surface, sa géographie et son climat, un ensemble de traits culturels distinctifs et une histoire commune (19) (20) (21). Nous nous appuyons sur les travaux de A. Toynbee « *A Study of History* » (22), de S. Huntington « *The Clash of Civilizations and the Remaking of World Order* » (21) notamment de leur proposition de distribution des cultures existant aujourd'hui. Nous adopterons cependant la dénomination « régions culturelles » pour la distribution que nous proposons, dans un souci de différenciation avec les civilisations proposées dans (21) et (22), et la raison que cette dénomination correspond plus à l'idée d'aire géographique que nous souhaitons exprimer. Voici la répartition des régions culturelles que nous avons choisi d'utiliser, cf. Figure 2

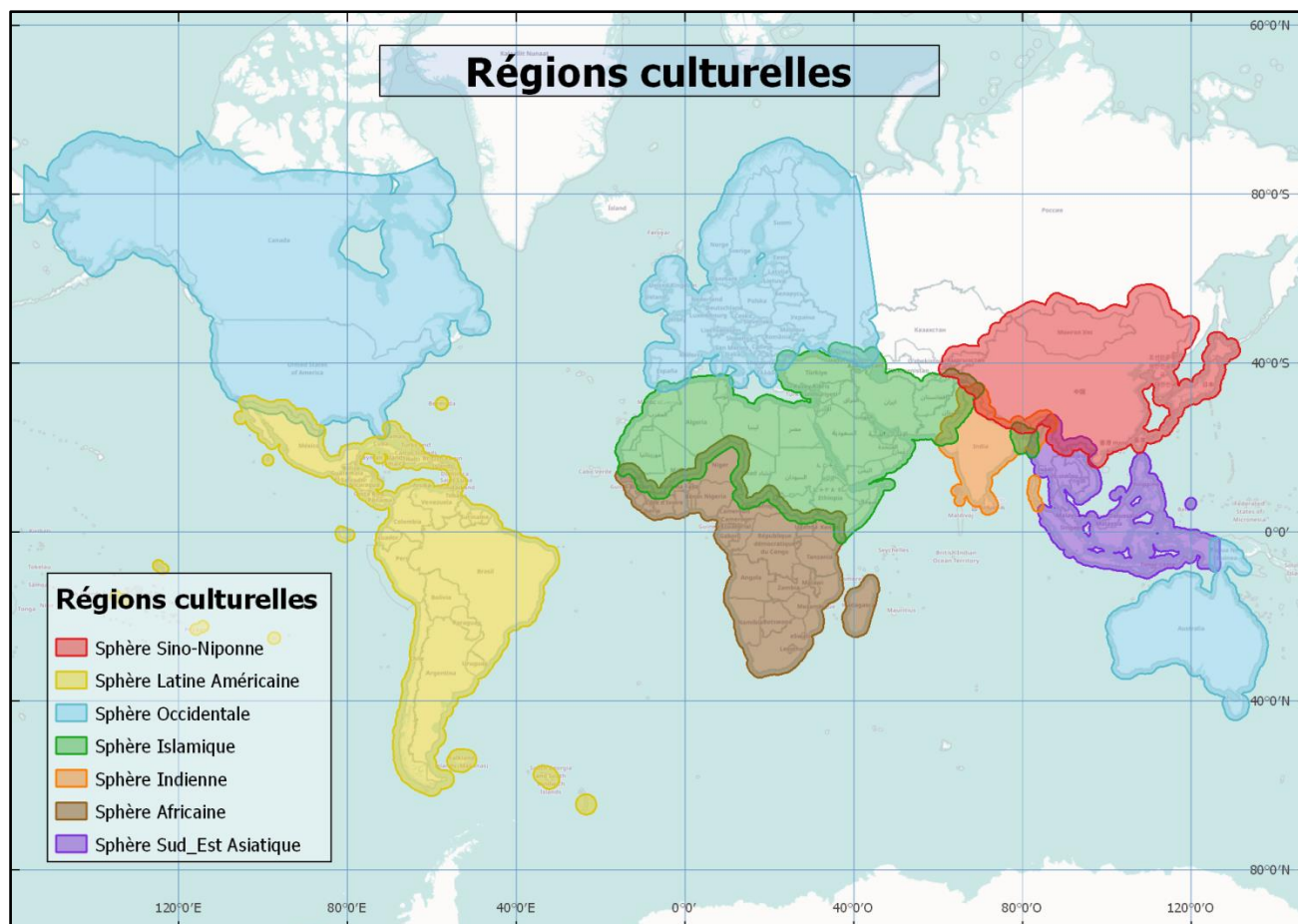


Figure 2: Répartition des sphères culturelles considérées

2. Aire urbaine

Si nous nous intéressons aux villes à travers le monde, nous pouvons nous rendre compte que la conception administrative de la ville dépend fortement suivant le pays. En ce qui concerne l'étude de la morphologie d'une ville, se limiter à ces limites administratives appauvrirait notre étude car cette morphologie est tributaire de celle des villes voisines. Nous choisissons donc de porter notre étude à l'échelle de la zone urbaine. Or ici encore nous nous retrouvons face à la diversité de définitions de l'aire urbaine suivant le pays : continuité du bâti, agglomération administrative de villes, seuil de population, ...

Devant cette diversité d'échelles que cette définition implique, nous proposons de suivre la définition de la continuité du bâti qui, nous le pensons, nous permettra de mieux correspondre à l'étude morphologique urbaine que nous voulons mener.

3. Analyse spatiale appliquée à la morphologie urbaine

On peut faire remonter l'apparition de l'étude des villes et de leur organisation spatiale à Burgess et la description qu'il fait de la manière dont Chicago s'est développée dans l'après Première Guerre Mondiale (*Growth of the City : An Introduction to a Research Project*) (23). Cette œuvre apportant une des premières théories sociales de la morphologie urbaine sera fondatrice de « l'école de Chicago ». À sa suite, naîtront différentes écoles (17) : L'école anglo-germanique, sous l'impulsion de M.R.G. Cozen, de sa visée de compréhension « morphogénétique » de la ville et de son approche multi-échelle introduite dans son analyse de la ville de Alnwick (Angleterre) (24), plantant les germes de l'ISUF (*International Seminar on Urban Form*); L'école italienne, portée par S. Muratori et son approche plus architecturale axée sur la « typomorphologie » historique des bâtiments. Les travaux de A. Rossi, notamment *L'Architettura della città* (18), s'inscriront dans cette lignée de théories basées sur les traditions d'édification de la ville (17) ; Enfin viendra dans les années 60 l'école française, incarnée par J. Castex et P. Panerai, qui s'intéresse particulièrement à la distinction à apporter entre la théorie d'un design urbain et à son application pratique ainsi qu'à la manière dont ce design va rapidement échapper aux conditions de sa création (25) .

Ces quatre écoles, bien que portées par des objectifs différents (explicatif, applicatif, prédictif) ont permis la mise en place d'une méthodologie de l'analyse de la morphologie urbaine reposant sur trois principes (17) :

- La définition de la forme urbaine par au moins 1 des 3 éléments physiques (le bâti, le lot et les rues)
- L'échelle à laquelle se place l'analyse (de l'échelle du bâtiment à celle de la ville)
- La période considérée

Ainsi, suivant cette méthodologie, notre analyse portera sur les formes d'unités de la taille d'un lot, à l'échelle de la ville et pour la période contemporaine à cette étude (2018-2019).

L'analyse spatiale est finalement assez peu développée en tant qu'outil d'analyse de la morphologie urbaine, bien qu'elle soit mature dans l'analyse géographique en général (par exemple : (26) (27) (28) (29) sur des propositions d'indices de covariances spatiales, de diversité, de dimension fractale, etc.).

Nous proposons ici seulement une analyse spatiale relativement simple, prenant en compte les spécificités de nos données caractérisant la morphologie urbaine des villes étudiées.

4. Influence de la culture sur la forme d'une ville

Avant de mener notre étude, il convient d'admettre que notre hypothèse reste ... une hypothèse. Le fait

que la culture ait une influence sur la morphologie des villes a été déjà discuté mais les preuves à l'échelle mondiale n'existent pas pour l'instant (il convient cependant de souligner les travaux menés par G. Boeing sur les réseaux viars (30)). Même si un voyageur, un minimum attentif, identifiera sans difficulté les différences entre la morphologie de la Médina de Fès, du quartier de Battery Park sur l'île de Manhattan, de la ville forte de Ping Yao ou de Vannes, cf. Figure 4, comment savoir factuellement si ces différences dépassent le scope discret de quelques villes ?



Figure 4: Images Google Street View actuelles d'anciennes villes fortifiées (sans le sens de lecture : la Medina de Fès, la vieille ville de Vannes, Lower Manhattan, la Cité de PyongYang)

C'est justement cette problématique que nous souhaitons adresser dans cette étude. Jusqu'ici, les travaux liés à l'influence de la culture sur la morphologie urbaine dressaient des comparaisons entre un nombre relativement réduit de villes (16) (31). Ces études ont sacrifié leur représentativité mondiale pour des descriptions très précises sur les villes traitées.

La compilation de telles recherches permet à Rossi de présenter la ville comme un objet structuré par l'histoire, et plus largement, par la civilisation à laquelle elle appartient (18). Quand Rossi parle de cette structure, il entend plus « architecture » que « plan urbain », cependant il insiste sur le fait que l'architecture (ou du moins, l'agrégation de production architecturale) est l'ossature, si ce n'est plus, de la forme urbaine. Bien que, à notre sens et selon l'école Cozenienne, cette vision de l'architecture ne représente qu'un aspect de la morphologie urbaine, l'idée proposée par Rossi reste valide. Rossi établit

qu'une ville possède sa singularité qui est issue de son histoire et des civilisations qui l'ont habitée. Ainsi, si l'on élargit cette idée à un ensemble de villes partageant une histoire commune et habitée par les membres d'une même culture, cet ensemble de ville devrait partager des traits morphologiques similaires.

L'organisation des réseaux viaires dans la ville est un des éléments physiques de l'étude morphologique que nous avons soulignés précédemment. Le travail récent de G. Boeing (30) dans ce domaine est pionnier. Dans cet article, Boeing mène une analyse sur la distribution des orientations des rues dans 100 villes à travers le monde. Ses résultats sont regroupés par région d'une manière plus administrative que culturelle et la représentativité de certaines de ces régions est discutable. Cependant, nous pouvons nous servir de ces résultats pour apprécier de manière quantitative une tendance non négligeable de la répartition des types de réseaux viaires à travers les régions culturelles que nous avons choisi de considérer.

Ainsi, plusieurs études ont été menées, à diverses échelles, pouvant nous aider dans la compréhension du lien entre sphère culturelle et morphologie des villes. Ces études mettent en avant une influence non négligeable de la culture sur la ville. Cependant, l'ensemble de ces études semble laisser une portion de la question en suspens, laquelle nous tenterons de combler.

C. Outils utilisés

Lors de ce projet, les données en entrée étaient de type raster et fichier de formes. L'emploi de logiciels SIG était donc quotidien (QGIS et suite ArcGIS). Le nombre d'éléments à traiter de la même manière était relativement élevé et impliquait des traitements complexes. J'ai donc fait le choix d'automatiser la majorité des traitements. Pour ce faire j'ai utilisé principalement le langage R dans Rstudio sous Windows. Ce choix a été motivé par d'un côté, la présence de la bibliothèque GDAL/OGR qui permet de manipuler des données géo spatiales, et d'un autre côté, la richesse de ses bibliothèques d'analyse de données.

Ont été mis à ma disposition : un poste de travail relativement récent, un espace de stockage en réseau et un espace d'échange de données en réseau.

IV. Jeux de données

A. Classification « Local Climate Zone » d'une centaine de villes

Récemment, en janvier 2019, l'équipe du professeur X. X. Zhu de l'institut de méthodologie de la télédétection du DLR (*IMF*) a réalisé une classification de l'usage du sol à l'échelle mondiale. Cette production se base sur le travail de C. Qiu (32). Dans cet article, Qiu propose une méthodologie basée sur la classification d'images multi-saisonnières de Sentinel-2 par réseau de neurones résiduels récurrents. Cette classification s'inscrit dans la taxonomie des « Local Climate Zones » introduit par I.D. Stewart et T.R. Oke (33)

1. Local Climate Zone

La taxonomie des Local Climate Zones (LCZ) a été, dans un premier temps, définie dans l'optique de servir de grammaire commune dans la communication sur les îlots de chaleur urbains (32) (34). Cette taxonomie prend en considération le type de surface (artificielle, végétale, aqueuse), la structure 3D (coefficient d'occupation du sol, hauteurs des bâtiments et de la végétation).

Concept	Signification commune	Signification dans ce contexte
Local	Caractéristiques de, ou associées à, un certain lieu	Entre 100m et 10km alentours
Climate	Caractéristiques durables dans le temps d'un lieu	Température de l'air à 1-2m au-dessus du sol (33)
Zone	Un lieu possédant des caractéristiques distinctives	Un lieu distingué par la morphologie, l'occupation du sol et sa température à 1-2m

Tableau 1: Définition du concept de Local Climate Zone (d'après Stewart et Oke (33))

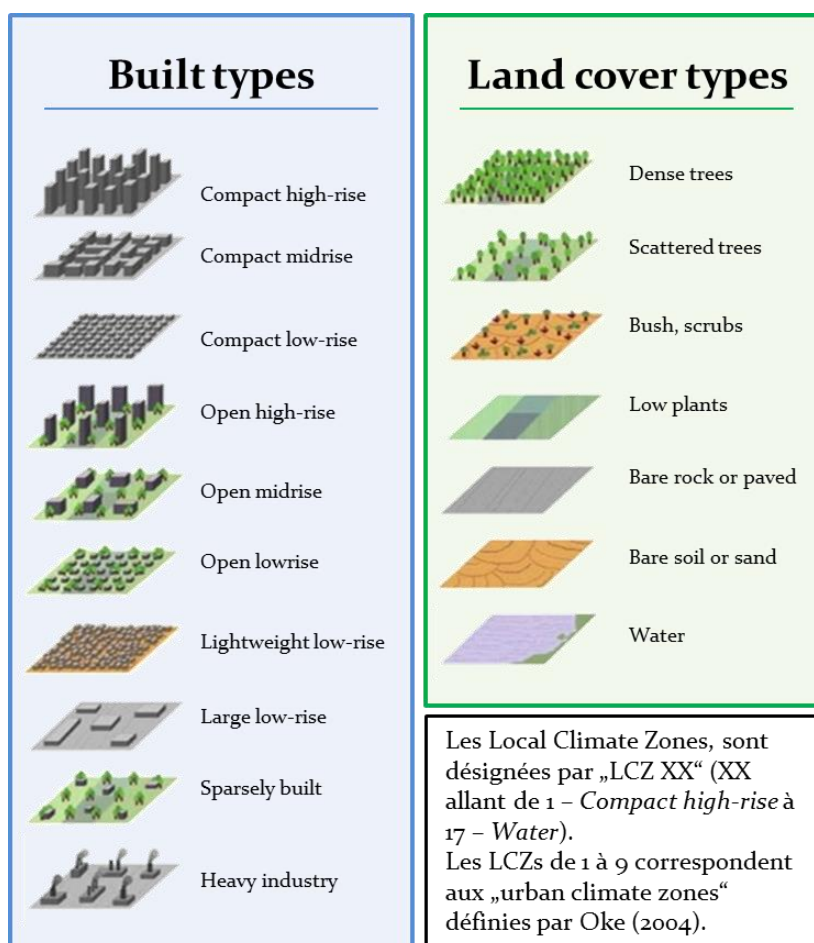


Figure 5: Les Local Climate Zones d'après Stewart et Oke (33) (35)

Cette taxonomie propose 10 classes d'occupation du sol bâties et 7 « naturelles » (32) (33) (cf. Figure 5). Cette taxonomie contient donc des informations très riches du point de vue de la morphologie urbaine (36). Dans cette étude, nous utiliserons, lors de certaines phases, les informations concernant le COS et la hauteur des éléments de rugosité. Le reste du temps, nous considérerons le type de LCZ comme étant notre paramètre étudié, ce paramètre étant encore plus riche en information. En effet, le type de LCZ donne aussi des informations sur : le facteur de forme ciel/sol, rapport largeur/hauteur, coefficient d'imperméabilité du sol et la classe de rugosité du sol (33) .

De plus, cette taxonomie a l'avantage de pouvoir être appliquée à tous types de villes à travers le monde, indépendamment des cultures (36), ce qui la rend très intéressante dans le cadre de notre étude et quasiment toute désignée.

2. Production de la classification

Comme nous l'avons montré, bien que créée originellement pour servir de support aux études des îlots de chaleurs urbains, la taxonomie des LCZs permet de mieux comprendre les dynamiques spatiales des structures urbaines (32). Partant de ce constat, le DLR a confié la tâche de produire cette classification à Qiu et al. . Plusieurs approches avaient jusque-là été utilisées pour extraire cette classification avec des exactitudes globales (*Overall Accuracy*) allant de 60% à 90% (37) (38) (36). Ces approches s'étaient limitées à quelques cas d'étude et étaient très dépendantes des connaissances de la morphologie locale.

L'approche proposée dans (32) adresse une classification globale. La classification a été produite à l'aide d'un réseau de neurones résiduels récurrents (*Re-ResNet* (39)). Ce réseau a été entraîné sur les images Sentinel 2 de Paris, Amsterdam, Cologne, Munich, Milan, London, et Berlin de 2016 à 2017 sur les 4 saisons. La particularité de cette approche est d'utiliser la variabilité saisonnière des données spectrales des différents types de LCZs et d'entraîner un réseau de neurones capable d'intégrer cette dynamique temporelle (32) (*Long Short-Time Memory - LSTM* (39)). L'exactitude globale moyenne des échantillons de validation est de 79,8%.

La classification produite est donnée sous forme de raster de résolution de 100m (cf. Figure 6), ce qui correspond aux définitions des LCZs dans (33). La classification a été menée sur plus d'une centaine de villes au début de mon stage (d'ici fin septembre, l'objectif est d'avoir réalisé la classification de toutes les villes de plus de 300 000 habitants). Ces villes ont été choisies à travers les régions du monde définies par l'ONU, pour que leurs populations soient proportionnelles à celle de leurs régions au regard de la population mondiale (40). Finalement, 110 villes ont été retenues et classifiées pour cette étude (cf. Annexe 1)

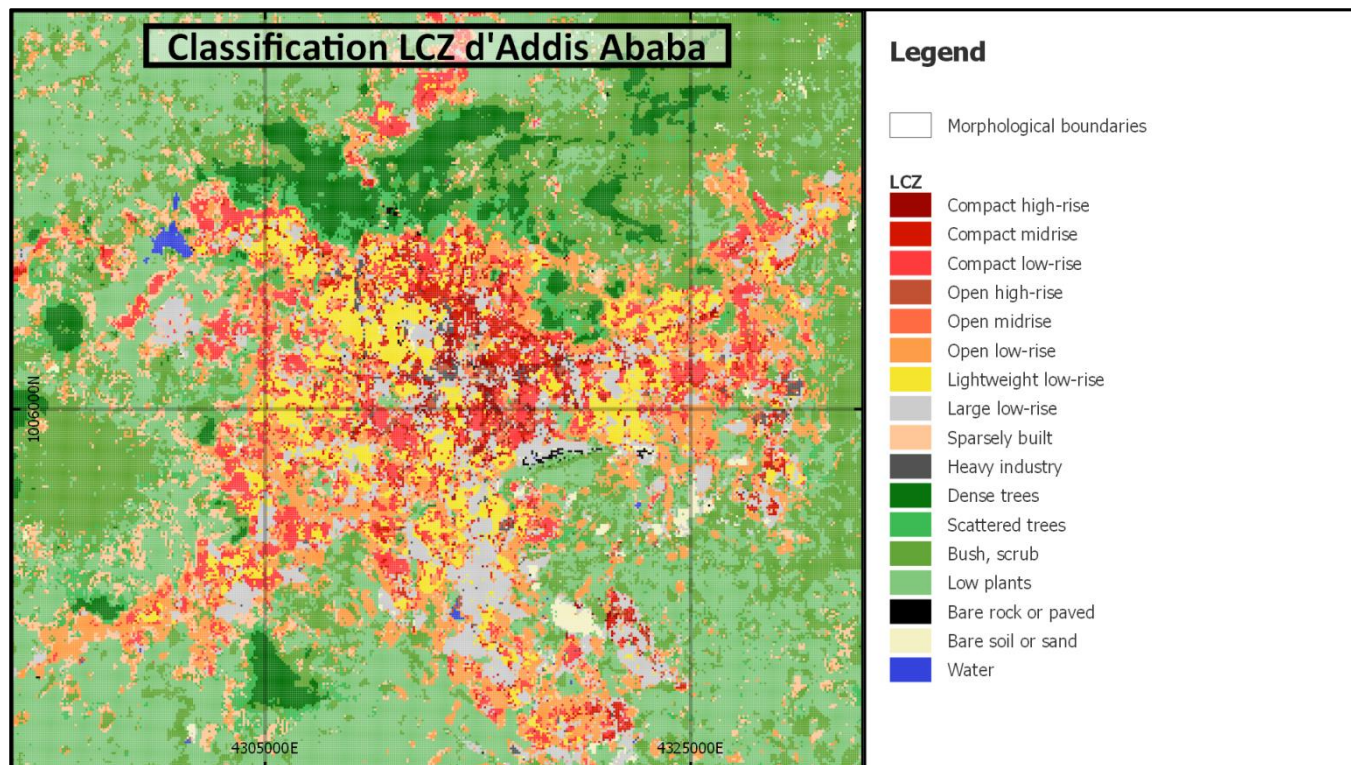


Figure 6: Exemple de classification en Local Climate Zones. Ici, Addis Ababa (Éthiopie).

3. Limites du produit

La taxonomie des LCZs, nous l'avons vu, est très intéressante dans le cadre de notre étude. Cependant, certaines limites apparaissent dans la classification que nous utilisons. La première est que les frontières entre les différents types de LCZs peuvent être compliquées. En effet, les centroïdes des classes sont bien définis dans (33) mais les étendues des classes dans l'espace des attributs ont tendance à se chevaucher (36). Par exemple les classes LCZ 4 et LCZ 5 (*Open high-rise* et *Open midrise*) se recoupent très largement. Cette proximité sémantique induit donc des confusions dans la classification et la variabilité de classes sémantiquement proches doit être considérée avec prudence. Une seconde limite est le nombre de sites d'entraînement et de validation. Les villes qui ont servi à l'entraînement des modèles sont des Villes d'Europe centrale, avec des spectrographies distinctes mais relativement proches. Le modèle n'a pas été entraîné sur des villes avec des albédos aussi élevés qu'en Afrique. C'est pourquoi on retrouve certaines classes de LCZs, par exemple LCZ 14 et 16 - *Low plants* et *Bare soil or sand*, en surreprésentation (selon nous) dans les villes aux sols sableux ou secs.

B. Frontières morphologiques des villes

Un projet non encore publié du DLR, sous la direction du Dr. Hannes Taubenböck (41) s'est intéressé à la définition des unités d'études dans le cadre de l'analyse urbaine. Ce projet part du constat que les unités administratives varient trop entre les villes de différents pays pour servir de base d'études à une échelle

mondiale. Taubenböck et. al. proposent dans cet article une définition de l'unité d'étude des villes basée sur leurs morphologies : les Morphological Urban Areas (MUAs – *Aires Urbaines Morphologiques*).

1. Principes et avantages

Comme nous l'avons évoqué précédemment, l'analyse de la morphologie urbaine se fait de plus en plus systématiquement en utilisant des SIG. Les méthodologies reposant sur les SIG nécessitent une définition très précise du cadre géographique de l'étude. Communément le cadre d'étude utilisé est celui des frontières administratives d'une ville. Or cette définition retranscrit peu la réalité morphologique des villes. On peut par exemple penser à Paris et la petite couronne qui forment un ensemble bâti continu qui dépasse les frontières de Paris. Au contraire dans certains pays, les limites administratives d'une ville dépassent de loin son étalement urbain (cf. Figures 7).



Figures 7: Limites administratives de Paris (France) et étalement de sa tâche urbaine ; Limites administratives de Zagreb (Croatie) et étalement de sa tâche urbaine

Il a été montré que, entre pays, ces différences sont très loin d'être négligeables (41).

Le but de ce projet a donc été de produire, pour toutes les villes de plus de 300 000 habitants, un shapefile des frontières basées sur l'étalement et la morphologie de la ville. Ce projet s'appuie sur la Global Urban Footprint (GUF – *Empreinte Urbaine Mondiale*) (42) qui proposait une couverture mondiale du masque urbain avec une résolution d'environ 12m.

De par sa nature issue de l'analyse de la morphologie des villes, nous argumentons que ce produit, à l'échelle mondiale, nous permet d'étudier, au plus près de son essence, la morphologie des villes, sans être influencé par des définitions administratives versatiles des villes.

2. Méthodologie de production

Comme dit précédemment, ce produit [Frontières Morphologiques Urbaines] s'appuie sur la GUF du DLR et de l'ESA (European Space Agency – *Agence Spatiale Européenne*). Les dernières versions de la GUF

(non commercialisées) proposaient, en plus de la couche « masque », une couche de densité.

Pour chaque ville de plus de 300 000 habitants dans le monde, le centre -ville a été extrait.

Autour de ce point, une distance maximum de 100 km a été fixée comme limite supérieure arbitraire du rayon d'une ville (dans les faits, aucune ville n'a dépassé les 60km de rayon).

Cette aire de 100km de rayon a été divisée en 16 secteurs radialement. Ces secteurs ont été ensuite divisés en 100 zones concentriquement (une zone a une longueur radiale de 1km).

Pour chacune de ces zones, la densité moyenne est calculée, ainsi que la densité médiane.

Pour chaque secteur, le gradient de la densité est calculé en pondérant avec les 2 secteurs voisins (les secteurs voisins comptent pour $\frac{1}{4}$ chacun).

Pour chaque secteur, la frontière est fixée là où le gradient de la densité devient plus faible que le gradient moyen.

Ces frontières concentriques interviennent donc à des distances au centre différentes pour chaque secteur (cf. Figure 8)



Figure 8 : Limites et emprise de Québec (Canada) par le produit des frontières morphologiques.

La GUF dispose dans ses masques, de 3 valeurs :

- La valeur « habitat »
- La valeur « surface non habitée »

- La valeur « étendue d'eau »

Ici, seule la valeur « habitat » a été retenue, masquant de ce fait les fleuves et étendues d'eau (cf. Figure 8).

Sur la Figure 9, on peut voir la différence flagrante entre les limites administratives de Québec et celles proposées par Taubenböck et. al. . Bien qu'imparfait, ce produit a le mérite de proposer une alternative interculturelle aux découpages administratifs et donc, dans le cadre de notre étude, de ne pas superposer une variable culturelle indépendante de la morphologie des villes.

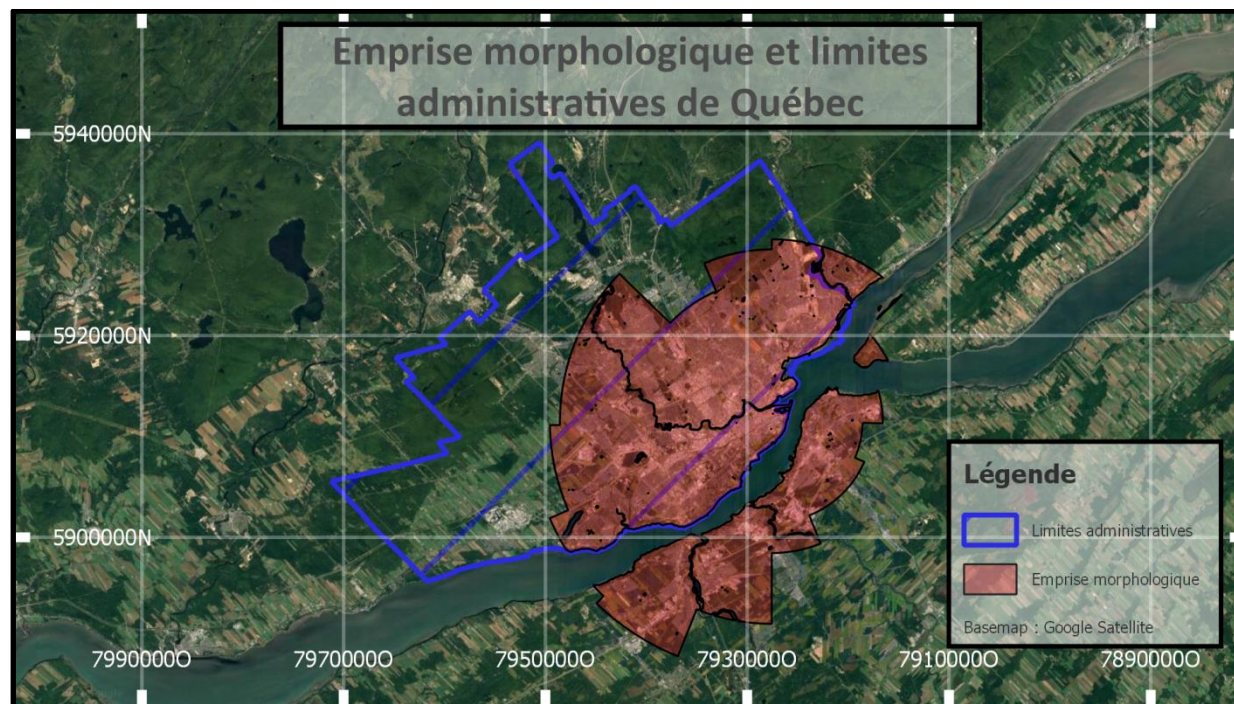


Figure 9 : Comparaison des limites administratives et de l'emprise morphologique de la ville de Québec (Canada).

3. Limites

Bien que ce produit propose des avancées intéressantes, il n'est pas sans défauts qu'il nous faut prendre en compte.

Premièrement, ces limites sont une approximation. La séparation en 16 secteurs proposée dans la méthodologie donne un zonage relativement fin près du centre mais ce zonage devient très large lorsque l'on s'éloigne. De plus, cette méthode prend comme hypothèse une forme urbaine approximable par des portions de disques, ce qui n'est pas tout le temps le cas en raison de divers facteurs (topologie, présence d'axes de passage, etc...).

Deuxièmement, le masquage des étendues d'eau engendre certaines incohérences. Ainsi, les villes se faisant face de part et d'autre d'une large étendue d'eau mais relativement proches sont regroupées en

une seule agglomération (cf. Figure 10 gauche). La présence d'un trop grand nombre de canaux peut créer des artefacts dans le produit final (cf. Figure 10 droite). Enfin, le masque des étendues d'eau ayant une précision encore relative, différents cours d'eau de même gabarit peuvent être ou ne pas être référencés (cf. **Figure 11**). Ces incohérences peuvent créer des problèmes au niveau de la prise en compte de la classe LCZ-17 (*Water*).

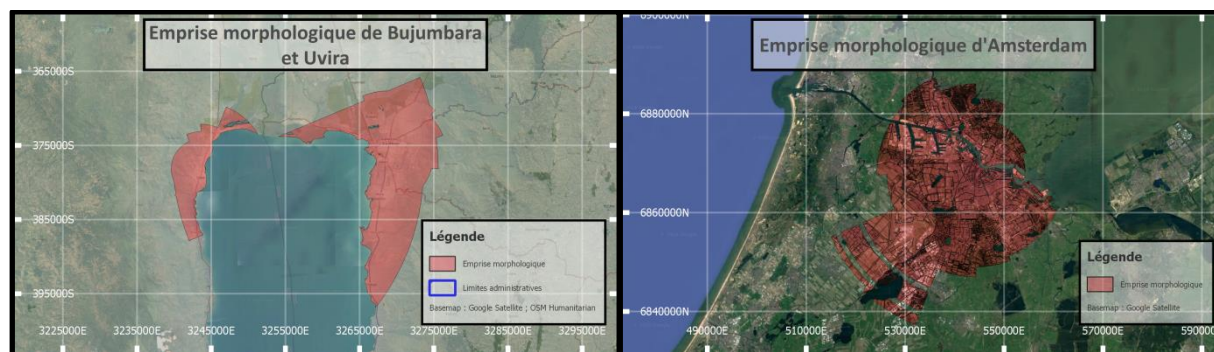


Figure 10: Bujumbara (Burundi) et Uvira (RDC) considérés comme une entité sans connexion terrestre ; Artefacts au Sud-Ouest d'Amsterdam (Pays-Bas)

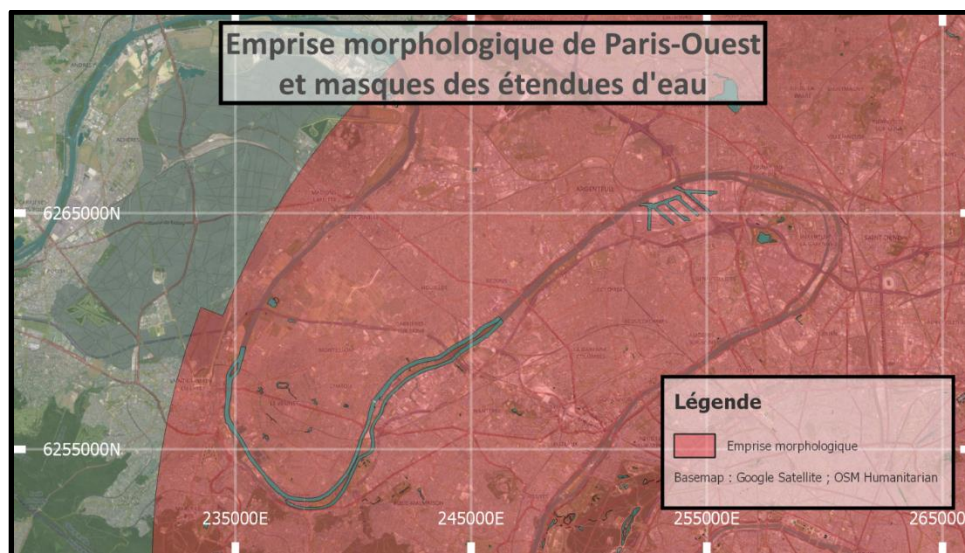


Figure 11: Segments de la Seine masqués et non masqués

Un dernier problème est l'agrégation des villes dans un contexte de proximité de villes proches (par exemple la métropole de Los Angeles ou celle de la Ruhr – cf. Figure 12). Ce problème intervient plutôt au niveau de l'analyse automatique proposée dans notre méthodologie. En effet, même si au sens morphologique cette emprise a du sens, une métropole pourra être comptabilisée dans l'étude plusieurs fois (une fois par centre-ville présent dans l'emprise morphologique de la métropole).

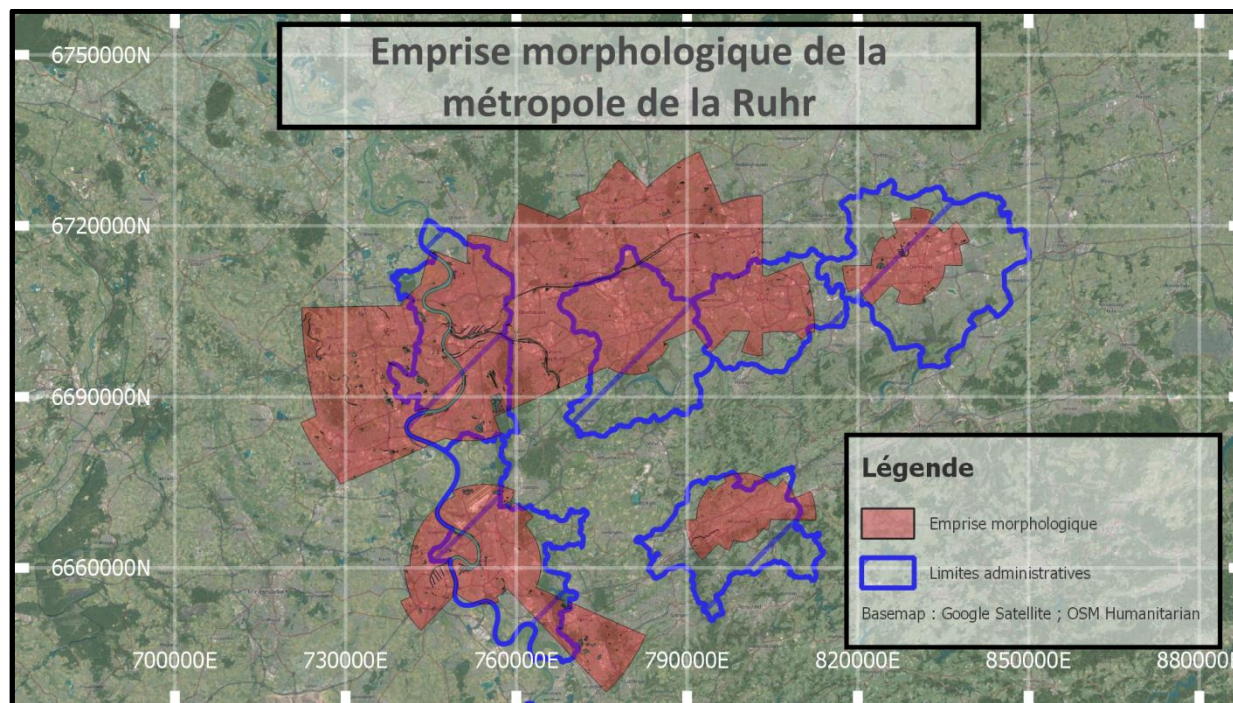


Figure 12: Emprise morphologique de la métropole de la Ruhr par aggrégation de ville voisines au sens morphologique.

C. Innovation proposée par l'étude

Ainsi, comme nous l'avons évoqué plus haut, cette étude est pionnière dans l'analyse morphologique des villes à l'échelle mondiale. Par les jeux de données utilisés (Classification LCZ et MUA) à très hauts degrés d'informations sur la morphologie urbaine et de par l'échelle mondiale (110 villes réparties dans le monde), ce projet se trouve paré à pouvoir s'adresser à la question de la morphologie urbaine à travers les cultures. De plus, nous argumentons que la méthodologie développée, basée sur de l'apprentissage non supervisé, adresse de manière pertinente car objective un sujet (la culture) qui peut très facilement faire appel au subjectif.

V. Méthodologie

Nous présenterons dans les paragraphes suivants, les différentes étapes du développement de notre méthodologie pour finir par la version finale à laquelle nous avons abouti.

A. Première approche

Dans un premier temps, nous présenterons la préparation des données et les premières analyses statistiques qui ont permis à notre réflexion de développer les premières esquisses de notre démarche méthodologique.

1. Préparation des données

La classification LCZ nous a été transmise en premier lieu sous format raster. Nous avons tout d'abord restreint l'ensemble de ces rasters à l'étendue des MUAs (cf. Figure 13).

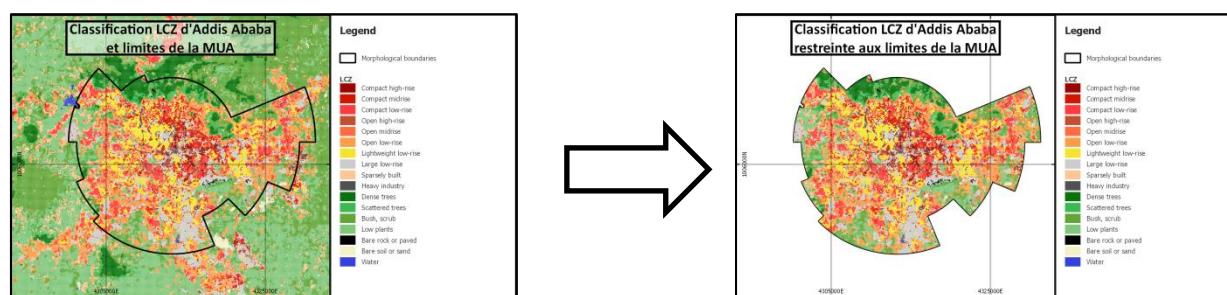


Figure 13: "Crop" de la classification LCZ par les frontières de la MUA

Une fois ces rasters cropés, nous avons extrait pour chaque ville la répartition statistique de chacune des classes. Nous avons regroupé les villes suivant les régions fournies par l'ONU (40) et avons observé les distributions suivantes en pourcentages (cf. Annexe 6).

2. Première analyse statistique

Ces premiers résultats montrent clairement des patterns de distributions des classes pour les villes de ces régions. Certaines régions montrent des modes de distributions avec peu d'écart (Amérique du Sud, Amérique du Nord, Europe de l'Ouest, ...) et certaines semblent présenter en leur sein plusieurs modes de répartition (Afrique de l'Ouest, Asie du Sud). Cette diversité apparente nous a donc conduits à ne pas suivre la définition des régions proposée par l'ONU.

Nous avons dès lors fait le choix de ne pas « forcer » des classes mais plutôt d'aborder le problème de cette répartition d'un point de vue indépendant de la position géographique. Les régions proposées dans

la Figure 2 ne serviront que dans la partie d'évaluation à la fin de notre étude. À noter que ces régions de références ont été développées pour palier la rigidité des régions de l'ONU et avoir une définition plus appropriée à l'étude des villes suivant un critère culturel.

Par ailleurs, la distribution des types de LCZ nous semble être un bon paramètre d'étude. Cependant, la confusion probable entre deux classes sémantiquement proches nous a poussés à essayer d'extraire un paramètre ordonné afin de mener notre étude. Ce paramètre est le volume (par unité de surface) calculé comme :

$$Volume_{cell} = Building\ surface\ fraction * Height\ of\ roughness\ elements$$

Avec les valeurs de "building surface fraction" et "height of roughness elements" proposées dans (33). Par la suite, ce paramètre sera utilisé parallèlement à la distribution des types de LCZs.

Enfin, lors de cette première approche, seule la proportion des classes de LCZs comptait. Or, selon nous, cette donnée perdait en signification si la taille de la ville n'était pas prise en compte. Aussi nous avons choisi d'intégrer ce paramètre à notre étude.

3. Premières conclusions

Nous avons donc vu que pour mener notre étude de l'influence de la culture sur la morphologie urbaine, la rigidité et la subjectivité de régions préétablies serait un handicap. Par ailleurs, nous avons vu que l'approche consistant à décrire les villes étudiées par un vecteur de valeurs prises par des paramètres adéquats (dans un premier temps la part de chaque type de LCZs dans une ville) semblait pertinente.

Ainsi nous avons redéfini notre problématique en s'axant sur l'espace des paramètres décrivant chacune des villes. La problématique d'un point de vue d'analyse de données devenait donc : Est-il possible d'obtenir un partitionnement statistique des 110 villes étudiées, dans l'espace des paramètres morphologiques sélectionnés, tel que ce partitionnement corresponde à une répartition géographique s'approchant de régions culturelles de référence ?

B. Méthodes et algorithmes

Dans ce paragraphe nous décrivons les différents algorithmes de partitionnement sélectionnés, ainsi que les différents espaces de paramètres constitués.

1. Présentation des algorithmes de partitionnement de données

Comme nous l'avons présenté plus haut, notre problématique peut être reformulée en un problème de partitionnement de données. Le partitionnement de données (*Clustering*) consiste en la classification (généralement – et dans le cas présent – non supervisée) de sous-ensembles d'un jeu de données, tel que les différences inter sous-ensemble soient maximisées et les différences intra sous-ensembles

minimisés (43) (44). Ce regroupement en sous-ensembles s'opère par la comparaison quantitative des objets les constituant. Cette comparaison se fait entre des éléments, la plupart du temps numériques, de ces objets que nous appellerons « paramètres ». Ces paramètres numériques sont projetables dans un espace à n-dimensions (n étant le nombre de paramètres considérés). Le partitionnement de données (*clustering*) a pour but, basé sur une analyse de cet espace des paramètres, de découvrir par lui-même le groupement naturel des objets (43) (exemple cf. Figure 14).

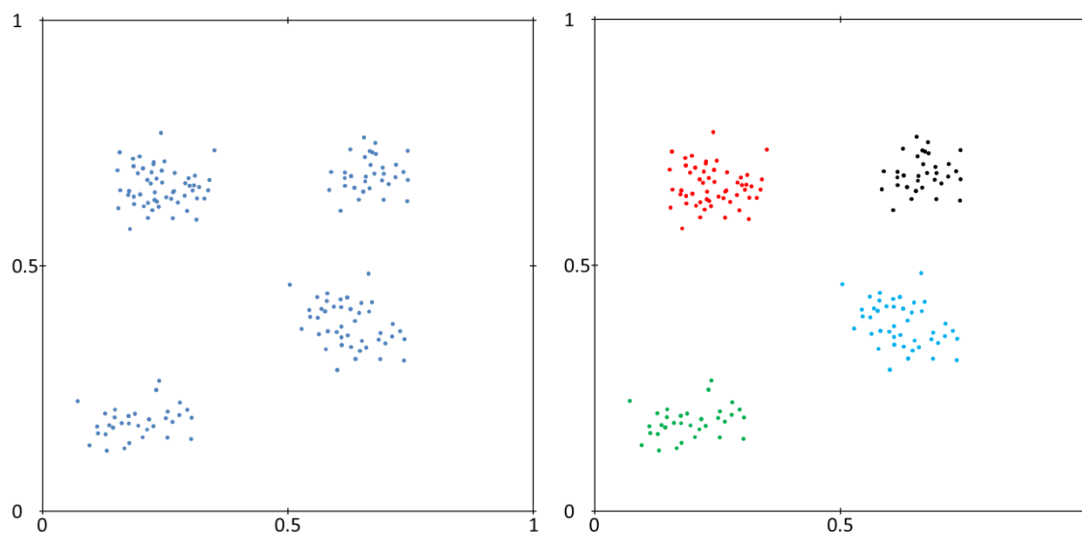


Figure 14: Exemple de distribution de données dans un espace à 2 dimensions (gauche) et partitionnement escompté (droite).

Les données, en fonction de l'espace des paramètres choisi, peuvent se présenter dans différentes configurations de distributions (cf. Figure 15), de gauche à droite, de haut en bas : distribution en patches denses, distribution annulaire concentrique, distribution elliptique, distribution en spirales). En observant ces distributions, on peut se rendre compte de l'éventuelle complexité de la tâche de partitionnement. Les différentes méthodes de partitionnement ne fonctionnent pas avec la même efficacité sur tous les types de distribution. La question se pose donc sur le but du partitionnement. Selon quels critères veut-on produire un partitionnement de nos données ? Cette question est très étroitement liée à l'ensemble de paramètres considérés, i.e. l'espace de représentation des données (43). Nous présenterons plus loin les espaces des paramètres choisis.

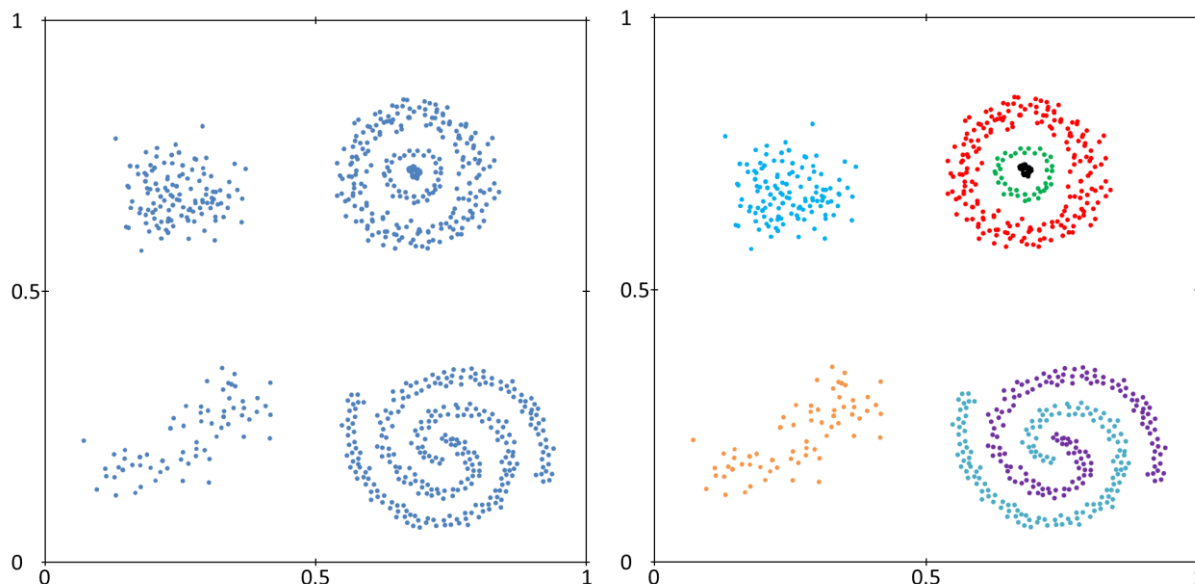


Figure 15: Exemple de différentes topologies de distribution de données visualisées dans un espace à 2 dimensions (gauche) et partitionnement escompté (droite).

Outre la distribution des données, un élément majeur de l'analyse par clusters est l'algorithme employé. La littérature nous propose plusieurs types d'algorithmes (43) (45) (46) que l'on peut regrouper en plusieurs familles : hiérarchiques, par centroïdes, par analyse de distribution, par analyse de densité. Dans ces familles, nous avons sélectionné 3 algorithmes (K-means, EM, OPTICS) nous permettant de ne pas présupposer trop d'informations sur les classes finales, autre que le nombre de clusters. Ces différents algorithmes ont été sélectionnés à des fins de tests exploratoires sur la structure des distributions de nos données.

La recherche du nombre de clusters est un problème relativement complexe de manière analytique : comment savoir combien de clusters seront présent dans le partitionnement final des données, sans savoir quel serait ce partitionnement ? Souvent, la détermination du nombre idéal de sous-groupes se fait par des tests systématiques du nombre de clusters ou par connaissance *a priori*. Dans notre étude pourtant, nous ne voulons pas biaiser nos résultats par des découpages culturels pré-proposés comme dans (21). De même, nous ne pouvons pas nous fier à une heuristique basée sur un tâtonnement vers une répartition qui nous semblerait être la plus fidèle à la représentation que nous avons des cultures.

Pour déterminer le nombre de clusters, nous avons utilisé la technique du « *Gap Statistic* » (47). Nous l'avons appliqué depuis R en utilisant la fonction `clusGap()` de la bibliothèque `cluster` (48) avec une population maximale de 20 clusters, testé sur 1000 batchs, avec une distance euclidienne et la méthode « *firstSEmax* » pour trouver le premier maximum local. Cette méthode dépendant de l'initialisation aléatoire, nous avons lancé cette méthode sur 1000 seeds indépendantes.

Cette méthode propose une heuristique statistique de détermination du nombre optimal de sous-

groupes pour le clustering. L'idée de base est d'observer l'évolution de la différence entre le logarithme de la somme des inter-distances d'un clusters et l'espérance du logarithme de la somme des clusters dans le cas d'une distribution gaussienne des données, le tout en faisant varier le nombre de clusters. Cette méthode fonctionne avec tous types d'algorithmes de clustering et tous types de distances. Dans notre cas, nous avons utilisé la distance euclidienne.

R. Tibshirani et. al. proposent :

En considérant les données $\{x_{i,j}\}, i = 1, 2, \dots, n, j = 1, 2, \dots, p$, n étant le nombre d'observations indépendantes, p étant le nombre de paramètres. En considérant $d_{ii'}$ la distance entre l'observation i et i' (ici euclidienne). En supposant un partitionnement des données en C_1, C_2, \dots, C_k , avec C_r les observations du cluster r et $n_r = |C_r|$ le nombre d'éléments de C_r .

La somme des inter-distances d'un cluster est donnée par :

$$D_r = \sum_{i,i' \in C_r} d_{ii'}$$

On définit W_k :

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r$$

Enfin, le « Gap » est défini par :

$$Gap_n(k) = E_n^*\{\log(W_k)\} - \log(W_k)$$

A noter que ce Gap est dépendant du nombre d'observations (dans la définition de l'espérance E_n^* sur un échantillon de n observations suivant une distribution gaussienne). Or notre nombre d'observations est relativement faible (110). C'est pourquoi l'algorithme a été lancé sur 1000 distributions aléatoires gaussiennes indépendantes, afin d'augmenter la fiabilité de l'expérience.

L'algorithme fait varier le nombre de cluster (ici de 0 à 20) et identifie le nombre de clusters pour lequel le Gap est maximisé (par exemple 3 dans la **Figure 16**)

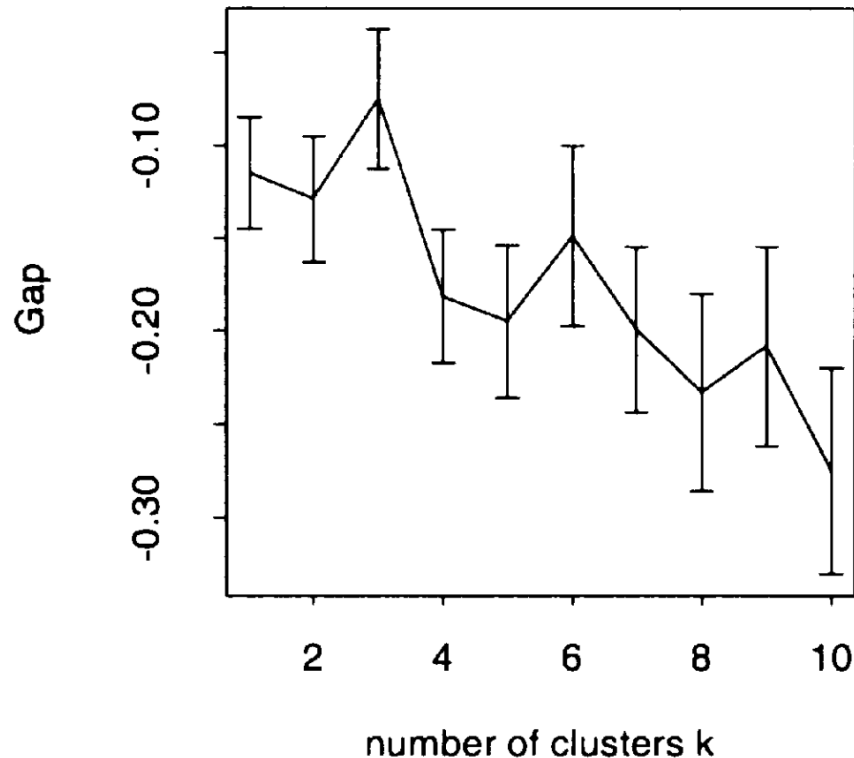


Figure 16: Valeur du "Gap" en fonction du nombre de clusters proposés. (Figure d'après Tibshinari et al. [37])

Le type de distribution de nos données dans les espaces des paramètres que nous présentons plus loin nous étant inconnu, nous avons choisis d'utiliser 3 algorithmes de type différents. Un s'intéresse plus particulièrement aux centres des clusters (K-means), un autre est plus adapté pour décrire les clusters aux densités différentes (OPTICS) et le dernier s'intéresse aux frontières des clusters (EM).

a) *K-means*

La méthode des K-means (*K-moyennes*) est une méthode itérative de partitionnement reposant sur la composition d'un cluster par son centre. L'algorithme que nous utilisons est celui proposé par Hartigan et Wong (49). L'algorithme compose une phase d'initialisation de K centres (ici initialisation aléatoire) dans l'espace des paramètres. Suit l'attribution aux points de données le cluster dont le centre est le plus proche. Cette étape est suivie par une nouvelle estimation du centre du cluster en effectuant le calcul du barycentre des points composant le cluster. Ces deux dernières étapes sont répétées jusqu'à convergence de la somme des inter-distances au sein des clusters.

Nous avons utilisé la fonction `kmeans()` du package natif de RStudio `stat`.

Cet algorithme a l'avantage d'être facile à utiliser dans un cas de connaissance du nombre de clusters et converge. Cependant, il est à noter que la convergence vers le minimum global n'est pas garantie et est tributaire de l'initialisation.

Cette méthode est particulièrement efficace pour trouver des sous-ensembles non-mélangés comme dans la Figure 14, avec des centres résumant fidèlement l'ensemble du cluster.

b) EM

La méthode EM (*Espérance-Maximisation*) est une méthode de partitionnement par estimation des paramètres de distribution des données (50). Cette méthode se décompose en deux étapes itératives. En partant du postulat de la connaissance de la famille du type de distribution présente (famille unique), est calculé pour l'initialisation, un premier vecteur des paramètres. On procède ensuite à l'*estimation* : le partitionnement des données en prenant le cluster le plus proche en termes d'espérance. Ensuite on procède à la *maximisation* : le calcul d'un nouveau set de paramètres qui expliqueront mieux la répartition des données dans les clusters. A chaque itération est calculée le logarithme de l'espérance des données d'appartenir aux clusters d'après le set de paramètres de la distribution. Il s'agit de la valeur que cet algorithme va maximiser :

$$L(\theta) = \log g(y|\Phi)$$

Avec g la famille de distributions, y les données, Φ le set de paramètres.

Dans cette étude, en utilisant la fonction *Mclust()* du package *mclust* (51) nous avons procédé à une sélection de la famille de distribution considérée en choisissant celle qui optimiserait le critère d'information bayésien (*Bayesian Information Criterion - BIC*) pour 7 clusters. Il s'agit du type "EII" (mélanges gaussiens sphériques, de volumes égaux, de formes semblables).

Comme les K-means, les valeurs initiales ont un impact non négligeable sur la maximisation de l'espérance et ne garantissent pas le caractère global du maximum trouvé.

Cet algorithme est utile en présence de groupes partiellement mélangés avec des paramètres de distributions très différents, permettant ainsi de mettre en avant les patterns des données.

c) OPTICS

L'algorithme OPTICS (*Ordering Points to Identify the Clustering Structure*) (52) n'est pas un algorithme de clustering à proprement parler. Il s'agit d'un algorithme qui trouve l'ordre des points de données basé sur leur distance d'accessibilité défini pour l'algorithme de clustering DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) (53). DBSCAN offre la possibilité de trouver les structures de données à partir de la densité des regroupements. A partir d'un voisinage ε et d'un nombre de points minimum par cluster *MinPts*, cet algorithme propage le voisinage de points contenu dans un voisinage d'un point central, aux points atteignables depuis les points appartenant au cluster. Cette méthode de partitionnement est très efficace dans un contexte de frontière non linéaire dans l'espace des paramètres considérés (exemple donné : la répartition concentrique et la répartition en spirales de la Figure 15) si l'on connaît ε et *MinPts* et que la densité des clusters reste la même. Si la densité change d'un cluster à l'autre, les paramètres choisis ne sont plus valables (particulièrement ε).

OPTICS permet de visualiser la distance d'accessibilité des points à partir d'un point central par un graphique en « vallées » et « collines » (cf. Figure 17). Ce graphique ordonné des points montre en

abscisse l'indice (ordonné) du point et en ordonnées sa distance d'accessibilité au point précédent. Les vallées représentent une distance d'accessibilité au point précédent très faible et donc très probablement des points appartenant au même cluster. Les collines représentent une distance au point précédent très grande et donc probablement un changement de cluster. Contrairement à DBSCAN, cette méthode permet de repérer les clusters même lorsqu'ils sont de densités différentes. Les vallées n'auront pas la même profondeur collines seront toujours là pour marquer la frontière d'un cluster (exemple clusters vert et rouge Figure 17).

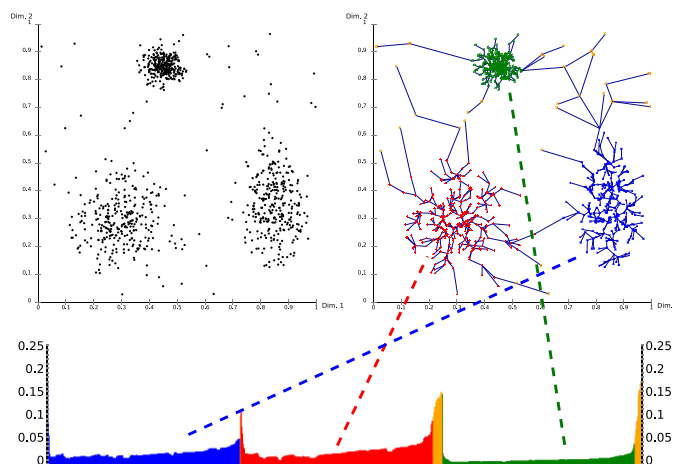


Figure 17: Graphe des distances d'accessibilité (partie basse) pour un jeu de données comportant des clusters de différentes tailles et différentes densités (représentés partie haute). Par Chire — Travail personnel, Domaine public, <https://commons.wikimedia.org/w/index.php?curid=10293701>

Pour la production du graphe des distances d'accessibilité, comme pour DBSCAN deux paramètres sont requis : ε et $MinPts$. La fonction *optics()* du package *dbscan* (54) offre la possibilité de n'utiliser que $MinPts$, ε étant considéré comme infini. En contrepartie, l'utilisation du gradient χ est requis. Nous avons choisi cette solution pour la détermination des clusters car elle nous permettait de produire moins de valeurs aberrantes. Nous avons fixé $MinPts = 4$ qui nous semblait une condition nécessaire pour pouvoir considérer un groupe de villes comme un cluster. Nous avons fait varier χ pour nos différentes expériences afin d'obtenir 7 clusters et respecter ce que l'analyse du Gap nous avait donné.

Les résultats obtenus par cette méthode ont été décevants et semblent indiquer plus de raffinement de notre méthodologie d'utilisation d'OPTICS.

2. Présentation des différentes approches statistiques

Comme expliqué en section IV.A.3., pour nos expériences de partitionnement, nous avons décrit chaque ville par un vecteur dans un espace de paramètres. Nous avons fait varier le nombre de dimensions de cet espace pour évaluer la séparabilité des données en fonction des dimensions considérées pour la description des villes.

a) 18 paramètres sémantiques

La première approche a été de décrire les villes, comme vu précédemment, avec le pourcentage de chaque classe de LCZs et en considérant la superficie de la ville.

Ville	LCZ 1 (%)	LCZ 2 (%)	LCZ 16 (%)	LCZ 17 (%)	Superficie (nombre de pixels)
Munich	4,5	15,2	0,3	1,1	31043

Tableau 2: Description d'une ville avec 18 paramètres

Cette description, étant donné le data set, est très intuitive, facile à implémenter et permet une bonne séparabilité des données. Le paramètre de la superficie a été ajouté aux proportions des classes de LCZs car, à notre sens, il renseigne sur le type de ville considérée et sur l'emprise du MUA. Pour procéder au clustering, nous avons normalisé les données. Les 17 paramètres des LCZs ont été normalisés d'un côté et la superficie a été normalisée indépendamment.

b) 3 composantes principales

La deuxième approche visait à réduire le nombre de paramètres en jeu afin d'éventuellement espérer une visualisation en 3 dimensions des résultats du clustering. Pour ce faire, nous avons procédé à l'extraction des 3 premières composantes principales, par ACP, de l'espace à 18 paramètres. Ces trois composantes principales ne représentaient que 56,2% des données. Nous avons décidé de tout de même mener l'expérience sur ce jeu de données.

Ville	CP1	CP2	CP3
Munich	0,250	0,724	0,004

Tableau 3: Description d'une ville avec 3 composantes principales

c) 101 paramètres de la distribution géographique de la densité

Comme vu en section IV. A. 2., nous avons pensé à un paramètre morphologique unique qui résumerait en partie les LCZ. Il s'agit du volume des LCZs. La troisième approche visait à retranscrire la distribution géographique de ce paramètre morphologique.

Chaque ville a été découpée en anneau concentrique comptant chacun pour 1% du rayon maximum de la ville. Pour chacun de ces 100 anneaux, le volume moyen par LCZ est calculé. Le vecteur des paramètres des villes, dans cette troisième approche se compose donc de ces 100 valeurs, auquel on rajoute celui de la superficie encore une fois.

Ville	Anneau 1 (m^3)	Anneau 2 (m^3)	Anneau 99 (m^3)	Anneau 100 (m^3)	Superficie (Nombre de pixels)
Munich	1 500 000	14500	200	0	31043

Tableau 4: Description d'une ville avec 101 paramètres

Ici encore, une normalisation est menée d'un côté sur les paramètres de densité et d'un autre sur la superficie.

d) 1701 paramètres de la distribution géographique des paramètres sémantiques

Nous avons développé la dernière approche en combinant les 17 paramètres des LCZs et la distribution géographique de ceux-ci.

Chaque ville a été découpée en anneau concentrique comptant chacun pour 1% du rayon maximum de la ville. Pour chacun de ces 100 anneaux, nous avons calculé le pourcentage de chaque classe de LCZs. Le dernier paramètre reste la superficie.

Ville	Anneau 1 LCZ 1 (%)	Anneau 1 LCZ 2 (%)	Anneau 100 LCZ 16 (%)	Anneau 100 LCZ 17 (%)	Superficie (nombre de pixels)
Munich	20,7	19,3	5,2	0	31043

Tableau 5: Description d'une ville avec 1701 paramètres

Encore une fois la normalisation pour le clustering se fait avec les 17*100 premiers paramètres d'un côté et la superficie de l'autre.

3. Mode opératoire des expériences

Nous avons souhaité effectuer un test systématique des méthodes de clustering sélectionnées sur les espaces de paramètres considérés. Pour chaque expérience, nous avons testé un algorithme sur un jeu de données. Avec nos 4 types de données et nos 3 algorithmes, nous avons produits 12 clusterings. Nous avons ensuite ordonné les 7 clusters de la manière suivante : le barycentre des données était calculé, le barycentre de chaque cluster était calculé, le barycentre de cluster le plus proche du barycentre des données se voyait attribuer le rang 1, de voisin en voisin, le barycentre du cluster le plus proche du barycentre de rang n se voit attribuer le rang $n + 1$. Nous avons ensuite cartographié les partitionnements obtenus, en spécifiant l'ordre de voisinage des clusters.

VI. Expériences et résultats

A. Méthode d'évaluation des résultats

A l'issu de ces expériences, nous avons donc 12 partitionnements des villes en clusters.

Ces partitionnements ne sont pas égaux et, pour certains, sont totalement différents. Nous avons voulu évaluer ces différents partitionnements et leur capacité à décrire des régions culturelles. Pour cela, nous avons choisi de prendre les régions culturelles proposées Figure 2 comme référence. A chaque expérience, nous avons calculé le Simpson Evenness Index (SIEI -*indice de régularité de Simpson*) donné, pour la région culturelle i par :

$$SIEI_{Région\ culturelle_i} = \frac{1 - \sum_{k=1}^7 \left(\frac{\text{nombre de villes de } i \text{ dans le cluster } k}{\text{nombre de villes dans } i} \right)^2}{1 - \left(\frac{1}{7} \right)}$$

Nous avons ensuite calculé la moyenne de cet indice sur l'ensemble des régions culturelles. Enfin pour l'ensemble des expériences, nous avons réécalonné les valeurs entre 0 et 1. 0 représente la régularité la plus basse, i.e. une répartition où un type de cluster est dominant dans la région culturelle. Le 1 représente la valeur la plus haute, i.e. une répartition régulière de la présence des clusters dans une région. Ainsi, nous argumentons que les meilleures combinaisons de jeux de données et d'algorithme de clustering sont données par les valeurs les plus basses dans le tableau suivant :

	em	kmeans	optics
3 PCA	0,37138116	0,21534295	0,74267774
18f	0,23096535	0,11421632	0,92008946
101f	0,52839873	0,69212382	1
1701f	0,11897809	0	0,22386154

Tableau 6: SIEI moyen par expérience.

B. Résultats des expériences

Dans ce paragraphe nous présenterons et discuterons les résultats de nos partitionnements. Nous ne présenterons pas les résultats des expériences réalisées avec la méthode OPTICS mais les cartes produites seront présentes en annexes. (Annexes 2 à 4)

1. Approche statistique par composantes principales

Sur les deux figures suivantes, nous pouvons voir les résultats du clustering sur les 3 CPs par EM et K-means. On peut retrouver des semblants de regroupements par régions culturelles : le monde islamique se dessine, l'Amérique Latine et l'Asie du Sud Est aussi ; l'Afrique semble se partager en 3 clusters cohérents ; l'Europe commence à montrer une relative tendance ; le Japon semble plus appartenir au même cluster que l'Asie du Sud Est plutôt qu'à celui de la Chine.

L'ordre des clusters reste globalement le même, d'autant plus sur les clusters centraux en termes de voisinage. On peut noter que l'Asie du Sud Est et la région Sino-Nipponne sont sémantiquement proche ; que l'Amérique du Nord et l'Europe semblent plutôt proches et forment une entité globalement cohérente ; que deux des clusters de l'Afrique sont très proches sémantiquement.

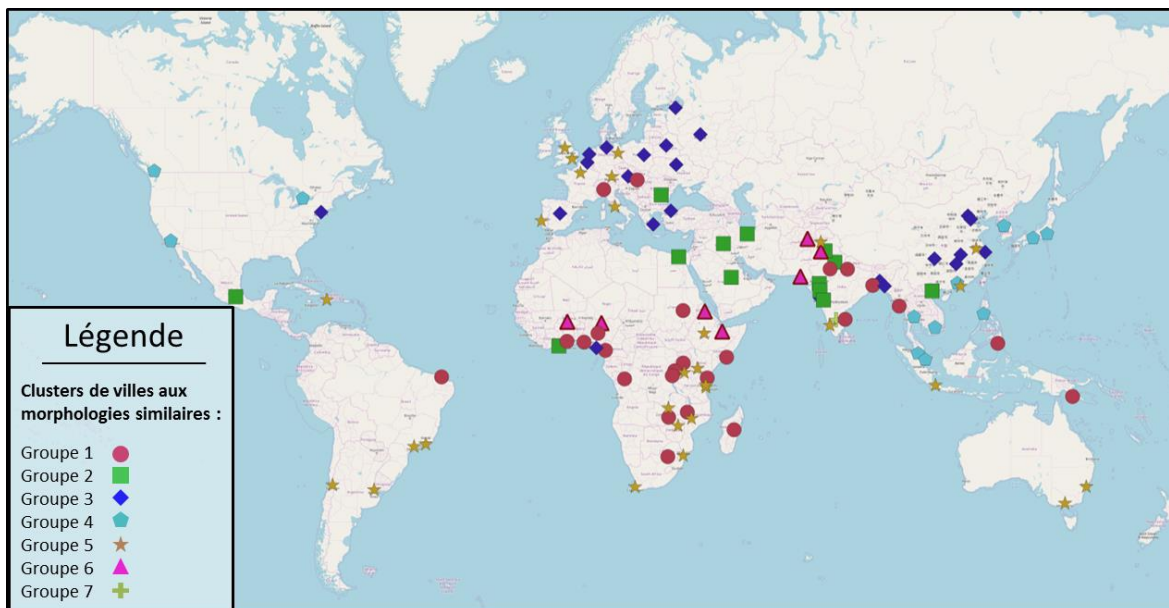


Figure 18: Résultat du clustering sur les 3 CPS par EM

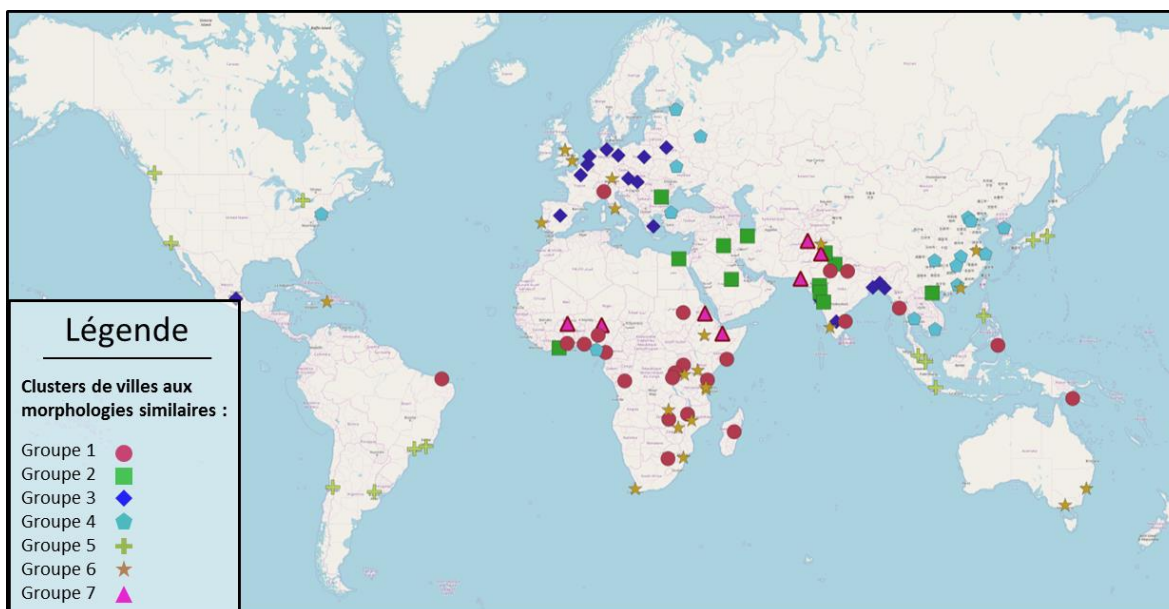


Figure 19: Résultat du clustering sur les 3 CPS par K-means

L'approximation nous semble plutôt fidèle à la référence de la Figure 2, plus fidèle que ce que nous attendions avec seulement 3 CPs.

2. Approche statistique par paramètres sémantiques

Sur les deux figures suivantes, nous pouvons voir les résultats du clustering sur 18 paramètres par EM et K-means. Le diagramme en bas à gauche est une méthode originale de visualisation des 18 paramètres en 2 dimensions. Nous nous sommes basés sur une idée proposée dans (55). Pour chacun des clusters, nous avons calculé la moyenne et l'écart-type pour chaque paramètre.

Nous avons divisé un disque en 7 anneaux concentriques, un pour chaque cluster. Nous avons ensuite divisé chacun de ces anneaux en 18 sections, une pour chaque paramètre. Ainsi nous avons obtenu 18*7 cellules représentant un paramètre pour un cluster.

Nous avons attribué à chaque cellule un niveau de gris qui représente la proportion de la LCZ du cluster ou la superficie des villes du cluster. Les valeurs hautes tirent vers le blanc, les valeurs basses, vers le noir.

Nous avons attribué à chaque cellule une pastille de couleur qui représente l'écart-type de la LCZ à travers le cluster ou l'écart-type de la superficie des villes du cluster. Les valeurs hautes sont rouges, les valeurs basses, bleues.

Les clusters sont ordonnés de la même manière que précédemment et le cluster de rang 1 est au centre du diagramme, le cluster de rang 7, à l'extérieur.

On peut aussi remarquer que les résultats deux approches sont très cohérents entre eux. Le monde islamique est encore une fois bien défini, l'Amérique latine un peu moins que précédemment, à moins de considérer la proximité des clusters dans l'approche par K-means. A noter que l'Europe prend réellement forme et par proximité des clusters, la sphère occidentale encore plus.

Ici aussi, l'ordre des clusters change peu et les deux diagrammes sont très similaires. Encore une fois, deux des clusters de l'Afrique sont très proches sémantiquement.

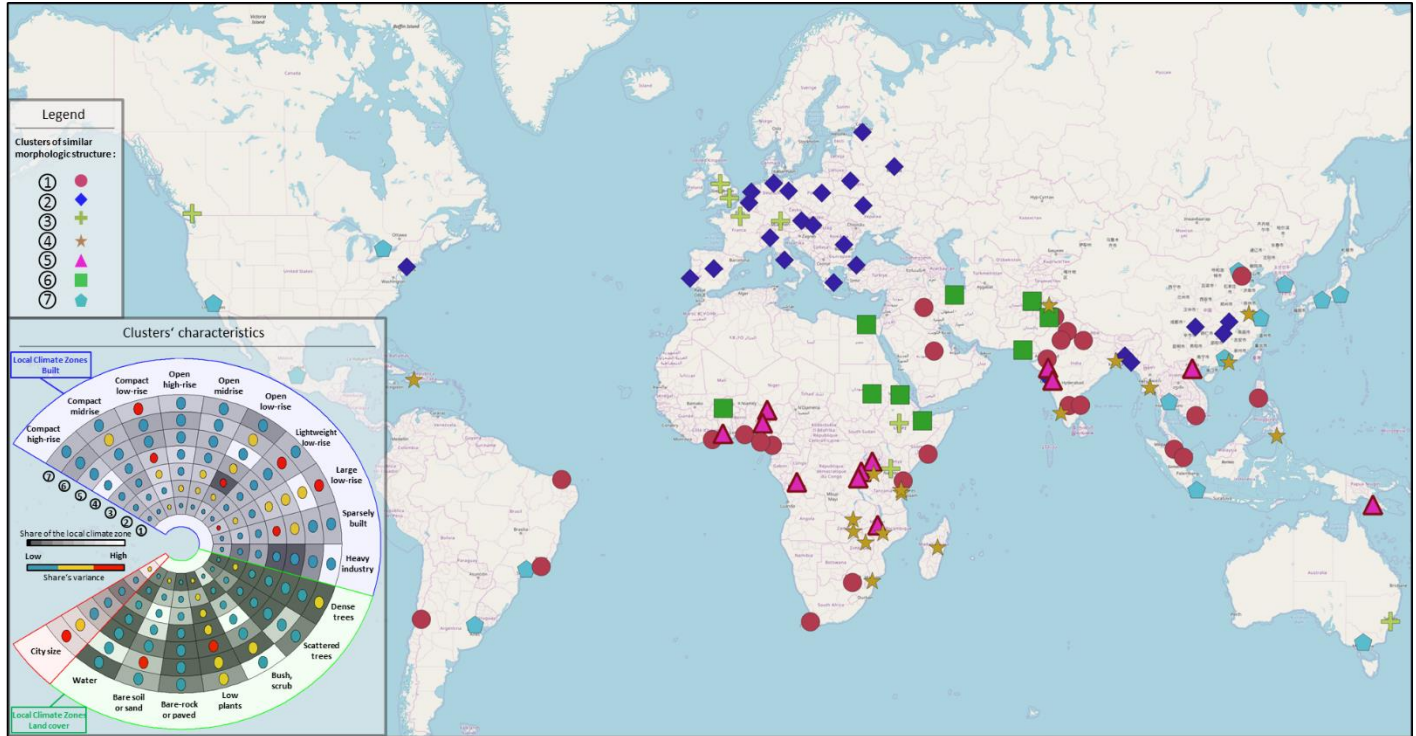


Figure 20: Résultat du clustering sur les 18 paramètres sémantiques par EM

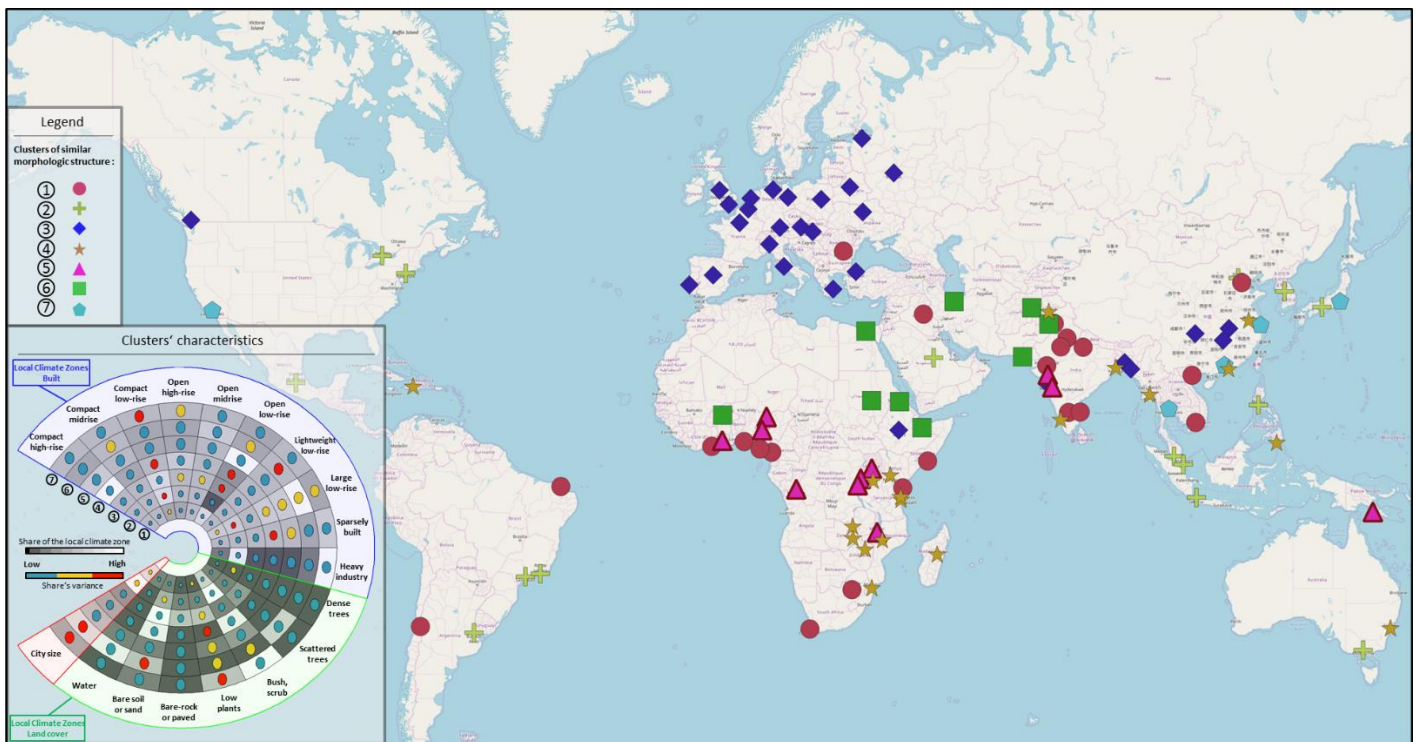


Figure 21: Résultat du clustering sur les 18 paramètres sémantiques par K-means

3. Approche statistique par distribution géographique de la densité

Sur les deux figures suivantes, nous pouvons voir les résultats du clustering sur 101 paramètres par EM et K-means. Comme on peut le remarquer, le clustering effectué ne semble pas être très pertinent en termes de régions culturelles, ce que semblent confirmer les hautes valeurs de SIEI pour ces approches.

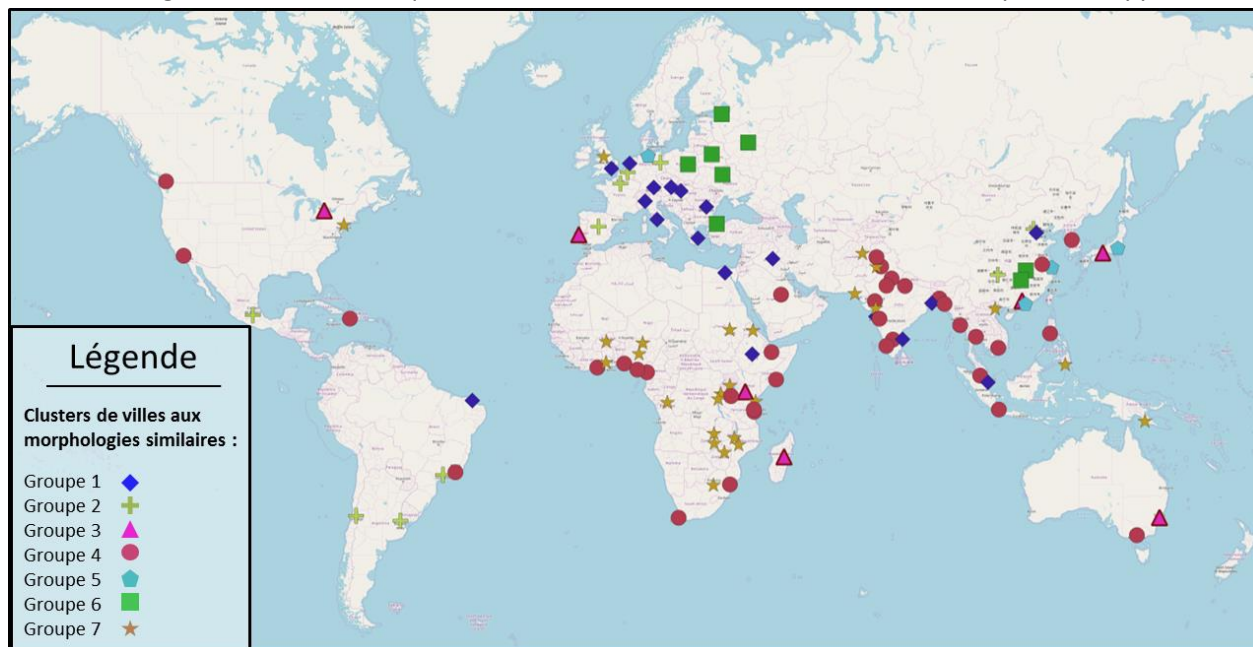


Figure 22: Résultat du clustering sur les 101 paramètres géographiques par EM

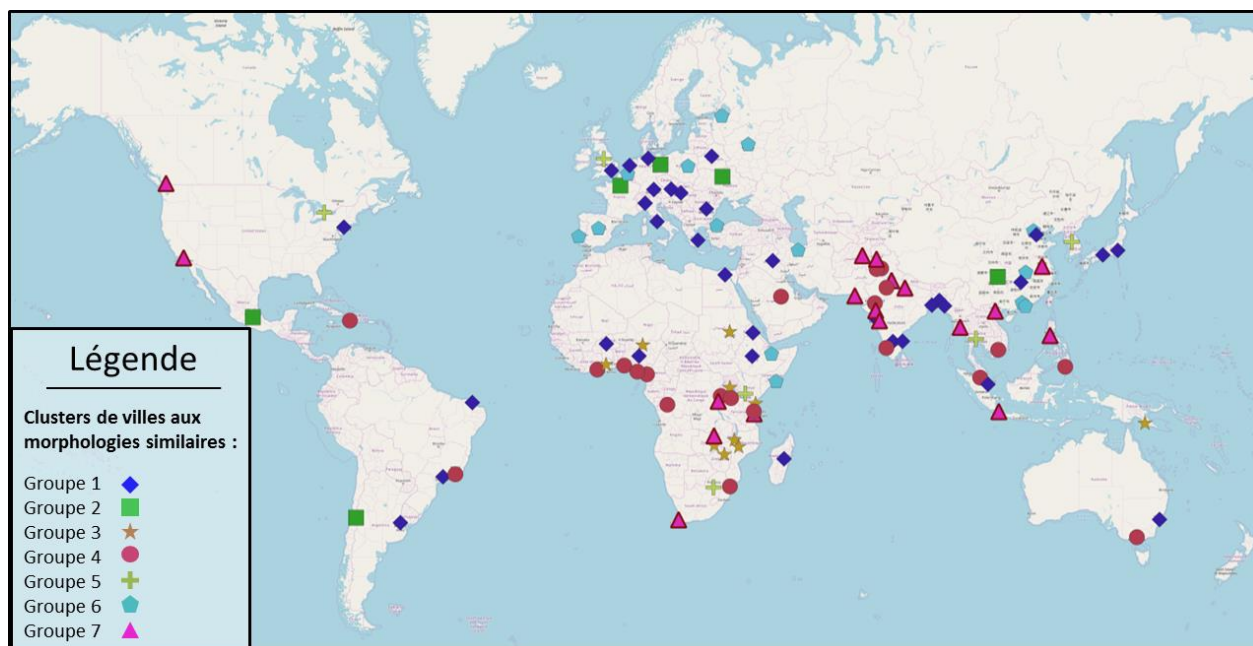


Figure 23: Résultat du clustering sur les 101 paramètres géographiques par K-means

4. Approche statistique par distribution géographique des paramètres sémantiques

Sur les deux figures suivantes, nous pouvons voir les résultats du clustering sur 1701 paramètres par EM et K-means. Ces deux partitionnements sont selon nous les plus aboutis. Particulièrement le clustering obtenu par EM. Les deux montrent des régions culturelles particulièrement dessinées. Mais l'approche par EM semble donner des divisions encore plus nettes au sein même des régions culturelles : Europe de l'Ouest et Europe de l'Est Chine continentale et Japon, Afrique de l'Ouest et Afrique de l'Est encore une fois, Sud et Nord de l'Inde...

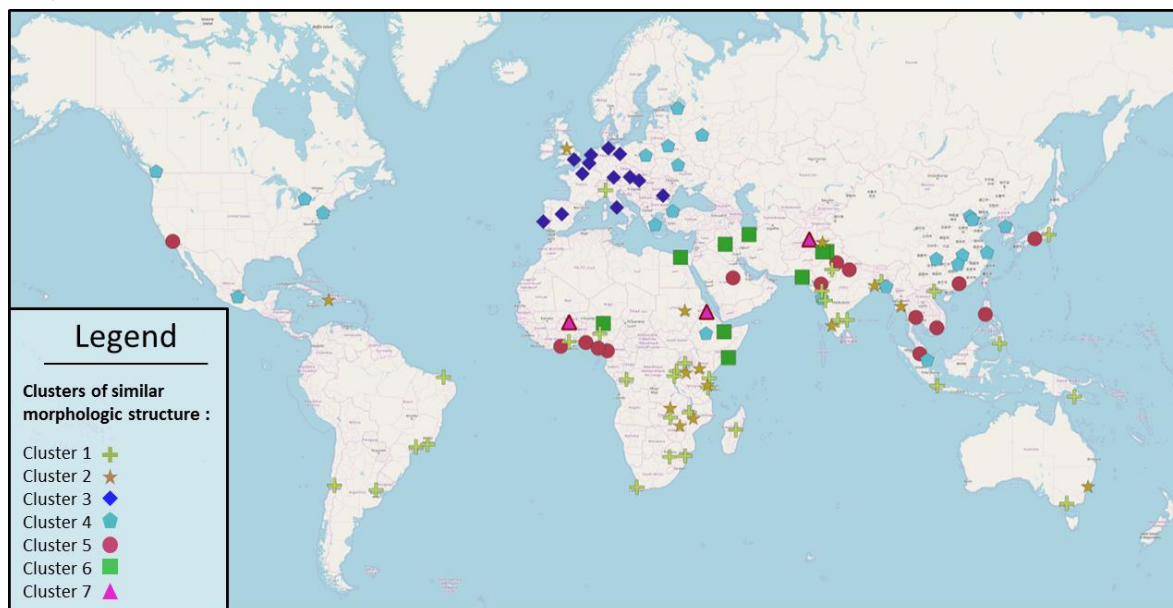


Figure 24: Résultat du clustering sur les 1701 paramètres sémantiques et géographiques par EM

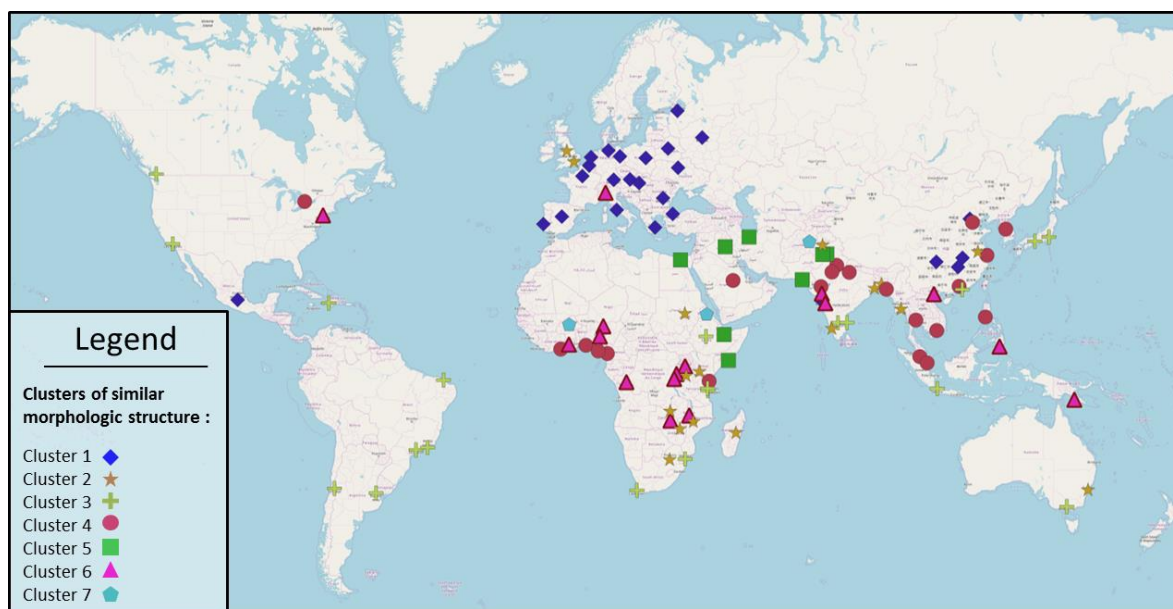


Figure 25: Résultat du clustering sur les 1701 paramètres sémantiques et géographiques par K-means

C. Discussion

Nous avons présenté les différents résultats de nos approches de clustering. L'index d'évaluation que nous avons mis en place (SIEI à travers les régions culturelles de référence) nous a permis de mettre en avant les expériences semblant créer les partitionnements les plus cohérents. Les trois approches avec les meilleurs scores sont : les k-means sur 1701 paramètres ; les k-means sur 18 paramètres ; EM sur 1701 paramètres. Les partitionnements issus de ces approches semblaient aussi, par interprétation géographique sur les cartes, les plus porteurs de sens.

Nous pensons que ces approches prouvent que la méthodologie globale utilisée ici est à même de procéder à un partitionnement des villes en différents groupes appartenant à des régions culturelles distinctes sur la base de critères morphologiques. Ainsi, nous défendons l'idée que la morphologie urbaine est belle et bien liée à la culture à laquelle la ville appartient.

Nous avons pu voir que les différents espaces de paramètres nous donnaient des résultats différents. Chacun des espaces présentait des avantages : simplicité dimensionnelle, simplicité de mise en place, indépendance de la proximité sémantique de certaines classes, caractère complet de la description. Nous pensons que dans le cadre de cette étude, les espaces à 18 et 1701 paramètres sont les plus pertinents. Cependant, nous ne réfutons pas l'idée que le clustering sur 101 paramètres puisse montrer un pattern que nous ne pouvons pas juger d'un point de vue géographique.

Nous avons testé 3 algorithmes de partitionnement de données. Il nous semble que les résultats donnés par la méthode OPTICS comportent trop de points aberrants pour pouvoir être conservés. Nous pensons que la spécificité technique de cet algorithme nous a partiellement échappée. Cette technique a été reportée dans ce rapport pour montrer l'ensemble de nos démarches mais ne sera pas utilisée dans nos travaux futurs tel quel. Les algorithmes EM et k-means nous ont donné des résultats similaires en dépit de leur mode de fonctionnement très différents. Nous voyons là une preuve de la bonne séparabilité de nos données et appuient notre conviction que l'on peut admettre un lien direct entre morphologie urbaine et culture.

Ainsi, le projet a été réalisé et nous a donné une réponse positive à notre question de recherche. Cependant, nous pensons qu'il y a matière à améliorer la méthode utilisée ici.

La définition de l'espace des paramètres géographique est pour l'instant établie pour des anneaux faisant 1% de la longueur du rayon maximale. Il est de notre avis que des anneaux d'une largeur fixée pour toutes les villes permettrait de se prémunir face aux très grandes différences de taille de villes que l'on peut rencontrer (74 km de différence entre les rayons maximums de la ville la plus grande et la ville la plus petite). De plus, une famille de paramètres à rajouter serait la superficie pour chacun des anneaux. En effet, cette donnée donnera des renseignements sur la forme globale de la ville - est-ce que tous les secteurs de la MUA sont pleins ou non ? -. Finalement, il pourrait être intéressant de trouver d'autres moyens de définir la distribution géographique des LCZs, comme par exemple par une distribution en secteurs ou encore par une segmentation préalable.

Par ailleurs, nous avons déjà discuté des limites des jeux de données de base i.e. la classification en Local

Climate Zones et les Morphological Urban Areas. Nous pensons qu'il s'agit d'excellents jeux de données mais il y a matière à les améliorer. Pour les MUAs, nous pourrions par exemple faire en sorte de ne pas masquer les étendues d'eaux telles que les lacs et les rivières, tout en continuant de masquer les mers et océans, tout en conservant les littoraux. Pour ce qui est des LCZs, une meilleure classification est à venir et étendue à plus de villes. Nous pensons que, comme pour l'approche par 1701 paramètres traités par EM, nous pourrions accéder à des niveaux de détails des régions culturelles plus locaux, l'analyse par statistique du Gap nous donnant certainement plus de clusters identifiables.

Pour finir, ici nous avons traité de l'analyse urbaine à travers le prisme de la culture. Nous pensons que les mêmes partitionnements, ou d'autres encore plus poussés, peuvent et pourront servir à analyser, par la même méthodologie, d'autres aspects de la morphologie urbaine, que ce soit en rapport avec la géographie environnante, les seuils de population, la santé...

VII. Conclusion

Ce projet m'a permis d'appliquer de nombreux savoirs théoriques développés au cours de mon cursus tant à l'ENSG (Ecole Nationale des Sciences Géographiques) qu'à l'EIVP. En effet, ce projet a requis de nombreux savoir-faire de la géomatique et une grande part de culture et compréhension urbaine. J'ai pu me confronter au rôle de chercheur au sein d'un institut national et à la réalisation technique complète d'un projet.

Le sujet du projet ici présenté, s'inscrit dans une volonté de compréhension de la ville par l'analyse urbaine à grande échelle. Nous avons défini la question de recherche qui nous a guidé lors de ce projet ainsi : « Existe-t-il un lien entre culture et formes urbaines ? ». Les données mises à ma disposition nous renseignaient sur les frontières morphologiques et sur la composition morphologique des villes. L'échelle mondiale de ces jeux de données nous a permis de développer une méthode basée sur le partitionnement de données. En utilisant différents algorithmes, nous avons proposé plusieurs partitionnements des villes en clusters. Nous voulions savoir si ces clusters représentaient une vérité sur les grandes régions culturelles du monde. En évaluant la répartition des clusters dans ces régions, nous sommes arrivés à la conclusion que ces clusters représentaient bien, en partie, ces régions culturelles.

Nous avons donc réussi à répondre positivement à notre question de recherche. Il existe bien un lien entre culture et morphologie urbaine. Nous sommes conscients que cette étude ne dresse pas un profil type des villes des différentes régions culturelles. Ceci pourra être apporté par un prolongement de ce travail.

Au cours de ce stage, j'ai pu encore une fois être confronté au monde de la recherche. Les enjeux liés aux aires urbaines me semblent encore plus clairement un enjeu majeur de notre siècle. Mais nous ne pouvons plus compter sur des solutions disparates et sans réels fondement autres que celui de profiter à certains. La compréhension de cet environnement urbain, de la manière dont on peut ou non créer une ville durable et habitable doit s'accélérer drastiquement pour faire face aux challenges démographiques de demain. Et je sors de ce stage convaincu que nous avons besoin de plus de chercheurs dans le domaine du génie urbain, et notamment dans l'analyse urbaine. Nous devons comprendre la ville pour pouvoir la bâtir.

Mon parcours en écoles d'ingénieur m'a permis d'acquérir la capacité de voir le monde de manière systémique. La recherche a nourri ma curiosité de comprendre ces systèmes. C'est un environnement qui convient tout à fait à mes aspirations professionnelles. C'est pourquoi, au cours de ce stage, j'ai postulé pour un doctorat dans le laboratoire où je suis actuellement. Ce stage a permis à mon maître de stage de juger mes capacités et a accepté de m'accueillir pour ma thèse dans son laboratoire. Mon thème sera dans la lignée de ce stage (« Analyse de la morphologie urbaine par produits de télédétection ») et j'aurais l'occasion de travailler sur les mêmes types de données que pendant mon stage, inscrivant professionnellement mon doctorat dans la continuité de ce stage de fin d'études...d'ingénieur.

Glossaire

DLR : Deutsches Zentrum für Luft- und Raumfahrt (Centre Allemand pour l'aéronautique et l'astronautique)

EOC : Earth Observation Center

DFD: Deutsche Fernerkundungsdatenzentrum (Centre national Allemand des données de la télédétection)

IMF : Institut für Methodik der Fernerkundung (*Institut de méthodologie de la télédétection*)

LCZ : Local Climate Zone

MUA : Morphological Urban Areas

Références

1. DLR - Locations. [En ligne] 2019. <https://www.dlr.de/dlr/desktopdefault.aspx/tabid-10175/>.
2. DLR at a glance. [En ligne] 2019. https://www.dlr.de/dlr/desktopdefault.aspx/tabid-10443/637_read-251/#/gallery/8570.
3. DLR - EOC. [En ligne] 2019. https://www.dlr.de/eoc/de/desktopdefault.aspx/tabid-5278/8856_read-15911/.
4. JURSE. [En ligne] 01 07 2019. <http://jurse.org/>.
5. JURSE Special Session "Slums, informal areas, or invisible cities? Mapping the "missing spaces" " (Part B). [En ligne] 2019. <http://jurse2019.org/session-slums-informal-areas-or-invisible-cities-mapping-the-missing-spaces-part-b/>.
6. habitat, UN. UN habitat. [En ligne] [Citation : 10 08 2019.] <https://new.unhabitat.org>.
7. ONU. UN Sustainable development. [En ligne] [Citation : 10 08 2019.] <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>.
8. Organisation, World Health. World Health Organisation - Sustainable development. [En ligne] [Citation : 10 08 2019.] <https://www.who.int/sustainable-development/new/en/>.
9. *The Built Environment and Human Activity Patterns: Exploring the Impacts of Urban Form on Public Health*. Engelke, Lawrence D. Frank and Peter O. 202, 2001 , Journal of Planning Literature , Vol. 16, pp. 202-218.
10. McDonald, J., Wise, M., & Harris, P. *The Health Impacts of the Urban Form. A Review of Reviews*.

Sydney : UNSW Research Centre for Primary Health Care and Equity, 2008.

11. *Study of the Effects of Urban Morphology on Physical Activity*. Braçe, Olta. 1, 2016, Revista de Estudios Andaluces, Vol. 33, pp. 24-39.

12. *Assessing urban planning resilience to outdoor air pollution in Greater Paris*. Jean-Marie Cariolet, Morgane Colombert, Marc Vuillet. Bologna : s.n., 2015. PLEA.

13. *Urban form and mental wellbeing: scoping a theoretical framework for action*. Amir Hajrasoulih, Vicente del Rio, James Francis and Jessica Edmondson. 10, Journal of Urban Design and Mental Health; 2018:5;10, Vol. 5.

14. *An analysis of high-rise living in terms of provision of appropriate social spaces for children*. Modi, Suruchi. 4, 2018, Journal of Urban Design and Mental Health, Vol. 5.

15. *Urban Design and Mental Health Case Study in Montreal / Tiohtiá:ke, Canada*. St-Paul, Rose-Anne. Advance publication, 2019, Journal of Urban Design and Mental Health, Vol. 6.

16. Bourdin, Stephane. *La forme de la ville - De l'Antiquité à la Renaissance*. s.l. : Presse Universitaire de Rennes, 2015.

17. *Urban Morphology as an Emerging Interdisciplinary Field*. Moudon, A. Vernez. 1997, Urban Morphology, Vol. 1, pp. 3-10.

18. Rossi, A. *The Architecture of the City*. Cambridge, Massachusetts : The MIT Press, 1982.

19. Balasubramanian, A. *Basics of Cultural Geography*. 2018.

20. L. Crothers, C. Lockhart. *Culture and Politics*. 2000.

21. Huntington, S. *The Clash of Civilizations and the Remaking of World Order*. 1997.

22. Toynbee, A. *A Study of History* . 1934-1961.

23. *The Growth of the City: An Introduction to a Research Project*. W., Burgess Ernest. 1925, Publications of the American Sociological Society, Vol. XVIII, pp. 85-97.

24. *Alnwick, Northumberland: A Study in Town Plan Analysis* . Conzen, M.R.G. 1960, London: Institute of BritishGeographers, Vol. Publication no. 27, p. xi+122 pp. .

25. D. Mangin, P. Panerai. *Projet Urbain*. Marseille : Éditions Paranthèses, 2013.

26. *Notes on Continuous Stochastic Phenomena*. Moran, P. A. P. 1/2, 1950, Biometrika Trust, Vol. 37, pp. 17-23.

27. *The Contiguity Ratio and Statistical Mapping*. Geary, R C. 3, 1954, The Incorporated Statistician, Vol. 5, pp. 115-127+129-146.

28. *Local Indicators of Spatial Association-LISA*. Anselin, L. 2, 1995, *Geographical Analysis*, Vol. 27, pp. 93-115.
29. McGarigal, Cushman, Neel & Ene. *FRAGSATs software*. 2002.
30. *Urban Spatial Order: Street Network Orientation, Configuration, and Entropy*. Boeing, G. 2018, *SSRN Electronic Journal*.
31. Salat, S. *Les Villes et les Formes sur l'Urbanisme Durable*. s.l. : Hermann & CSTB, 2011.
32. *Local climate zone-based urban land cover classification from multi-seasonal*. Chunping Qiu, Lichao Mou, Michael Schmitt, Xiao Xiang Zhu. 2019, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 154, pp. 151-162.
33. *Local climate zones for urban temperature studies*. Stewart, I.D., Oke, T.R. 2012, *American Meteorology Society*, Vol. 93, pp. 1879-1900.
34. *Local climate zones: origins, development, and application to urban heat island studies*. Stewart, I.D. Seattle : s.n., 2011. Annual Meeting of the American Association of Geographers.
35. *Siting and Exposure of Meteorological Instruments at Urban Sites*. R.Oke, T. Banff : s.n., 2004. International Technical Meeting on Air Pollution Modelling and its Application.
36. *Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities*. Benjamin Bechtel, Paul J. Alexander, Jürgen Böhner, Jason Ching, Olaf Conrad, Johannes Feddema, Gerald Mills, Linda See and Iain Stewart. 2015, *ISPRS*, Vol. 4, pp. 199-219.
37. *The accuracy of LCZ maps generated by the world urban database and access portal tools (WUDAPT) method: A case study of Hong Kong*. Ren, C., Wang, R., Cai, M., Xu, Y., Zheng, Y., Ng, E. Singapore : s.n., 2016. 4th Int. Conf. Countermeasure Urban Heat Islands.
38. *An Introduction to the WUDAPT*. Mills, G., Ching, J., See, L., Bechtel, B., Foley, M. Toulouse : s.n., 2015. 9th International Conference on Urban Climate.
39. *Long-term Recurrent Convolutional Networks for Visual Recognition and Description*. Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, Trevor Darrell. 4, 2017, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, pp. 677-691.
40. *World Urbanization Prospects 2018*. [En ligne] ONU, 2018. [Citation : 23 07 2019.] <https://population.un.org/wup/Download/>.
41. *A new ranking of the world's largest cities - Do administrative units obscure morphological realities?* Taubenböck, Hannes, et al. 2019, *Remote Sensing of Environment*, Vol. 232.
42. *Urban Footprint Processor-Fully Automated Processing Chain Generating Settlements Masks From*

- Global Data of the TanDEM -X Mission*. T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Müller, and S. Dech. 2013, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, Vol. 10, pp. 1617-1621.
43. *Data clustering: 50 years beyond K-means*. Jain, Anil K. 2009, Pattern Recognition Letters, Vol. 31.
44. *Sur la division des corp matériels en parties*. Steinhaus, Hugo. 12, 1956, Bulletin de l'académie polonaise des sciences, Vol. 4, pp. 801-804.
45. Wilder, Tukey John. *Exploratory Data Analysis*. s.l. : Addison-Wesley, 1977.
46. *Data Clustering: a Review*. Anil K. Jain, M. N. Murty, P. J. Flynn. 3, 1999, ACM Computing Surveys, Vol. 31, pp. 264-323.
47. *Estimating the number of clusters in a data set via the gap statistic*. Robert Tibshinari, Günter Walther, Trevor Hastie. 2001, Journal of the Royal Statistical Society, Vol. 63, pp. 411-423.
48. Martin Maechler ORCID iD [aut, cre], Peter Rousseeuw ORCID iD [aut] (Fortran original), Anja Struyf [aut] (S original), Mia Hubert ORCID iD [aut] (S original), Kurt Hornik ORCID iD [trl, ctb] (port to R et maintenance(1999-2000)), Matthias Studer [ctb]. cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al. [En ligne] [Citation : 5 04 2019.] <https://cran.r-project.org/web/packages/cluster/index.html>.
49. *Algorithm AS 136: A K-Means Clustering Algorithm*. Wong, J. A. Hartigan and M. A. 1, 1979, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 28, pp. 100-108.
50. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Dempster, A. P., Laird, N. M. et Rubin, D. B. 1, 1977, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, pp. 1-38.
51. Chris Fraley [aut], Adrian E. Raftery ORCID iD [aut], Luca Scrucca ORCID iD [aut, cre], Thomas Brendan Murphy ORCID iD [ctb], Michael Fop ORCID iD [ctb]. mclust: Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density Estimation. [En ligne] [Citation : 1 06 2019.] <https://cran.r-project.org/web/packages/mclust/index.html>.
52. *OPTICS: Ordering Points To Identify the Clustering Structure*. Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, Jörg Sander. Philadelphia : s.n., 1999. CM SIGMOD'99 international conference on Management of data . pp. 49-60.
53. *A density-based algorithm for discovering clusters in large spatial databases with noise*. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. 1996. 2nd International Conference on Knowledge Discovery and Data mining.
54. Michael Hahsler [aut, cre, cph], Matthew Piekenbrock [aut, cph], Sunil Arya [ctb, cph], David Mount [ctb, cph]. dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. [En ligne] [Citation : 20 06 2019.] <https://cran.r-project.org/web/packages/dbscan/index.html>.

55. *Use of Circle-Segments as a Data Visualization Technique for Feature Selection in Pattern Classification*. Shir Li Wang, Chen Change Loy, Chee Peng Lim, Weng Kin Lai. 2007. ICONIP . pp. 625-634.

Liste des figures et tableaux

Figure 1: Morphologie de Paris selon une classification en Local Climate Zones.	0
Figure 2: Répartition des sphères culturelles considérées.....	8
Figure 3:.....	10
Figure 4: Images Google Street View actuelles d'anciennes villes fortifiées (sans le sens de lecture : la Medina de Fès, la vieille ville de Vannes, Lower Manhattan, la Cité de PyongYang)	10
Figure 5: Les Local Climate Zones d'après Stewart et Oke (33) (35)	12
Figure 6: Exemple de classification en Local Climate Zones. Ici, Addis Ababa (Éthiopie).	14
Figures 7: Limites administratives de Paris (France) et étalement de sa tâche urbaine ; Limites administratives de Zagreb (Croatie) et étalement de sa tâche urbaine	15
Figure 9 : Limites et emprise de Québec (Canada) par le produit des frontières morphologiques.....	16
Figure 10 : Comparaison des limites administratives et de l'emprise morphologique de la ville de Québec (Canada).	17
Figure 11: Bujumbara (Burundi) et Uvira (RDC) considérés comme une entité sans connexion terrestre ; Artefacts au Sud-Ouest d'Amsterdam (Pays-Bas)	18
Figure 13: Segments de la Seine masqués et non masqués.....	18
Figure 14: Emprise morphologique de la métropole de la Ruhr par agrégation de ville voisines au sens morphologique.	19
Figure 15: "Crop" de la classification LCZ par les frontières de la MUA.....	20
Figure 16: Exemple de distribution de données dans un espace à 2 dimensions (gauche) et partitionnement escompté (droite).	22
Figure 17: Exemple de différentes topologies de distribution de données visualisées dans un espace à 2 dimensions (gauche) et partitionnement escompté (droite).....	23
Figure 18: Valeur du "Gap" en fonction du nombre de clusters proposés. (Figure d'après Tibshinari et al. [37])	25
Figure 19: Graphe des distances d'accessibilité (partie basse) pour un jeu de données comportant des clusters de différentes tailles et différentes densités (représentés partie haute). Par Chire — Travail personnel, Domaine public, https://commons.wikimedia.org/w/index.php?curid=10293701	27
Figure 20: Résultat du clustering sur les 3 CPS par EM	31
Figure 21: Résultat du clustering sur les 3 CPS par K-means	31
Figure 23: Résultat du clustering sur les 18 paramètres sémantiques par EM.....	33
Figure 24: Résultat du clustering sur les 18 paramètres sémantiques par K-means	33
Figure 26: Résultat du clustering sur les 101 paramètres géographiques par EM	34
Figure 27: Résultat du clustering sur les 101 paramètres géographiques par K-means.....	34

Figure 29: Résultat du clustering sur les 1701 paramètres sémantiques et géographiques par EM.....	35
Figure 30: Résultat du clustering sur les 1701 paramètres sémantiques et géographiques par K-means .	35
Tableau 1: Définition du concept de Local Climate Zone (d'après Stewart et Oke (33))	12
Tableau 2: Description d'une ville avec 18 paramètres	28
Tableau 3: Description d'une ville avec 3 composantes principales	28
Tableau 4: Description d'une ville avec 101 paramètres	28
Tableau 5: Description d'une ville avec 1701 paramètres	29
Tableau 6: SIEI moyen par expérience.	30

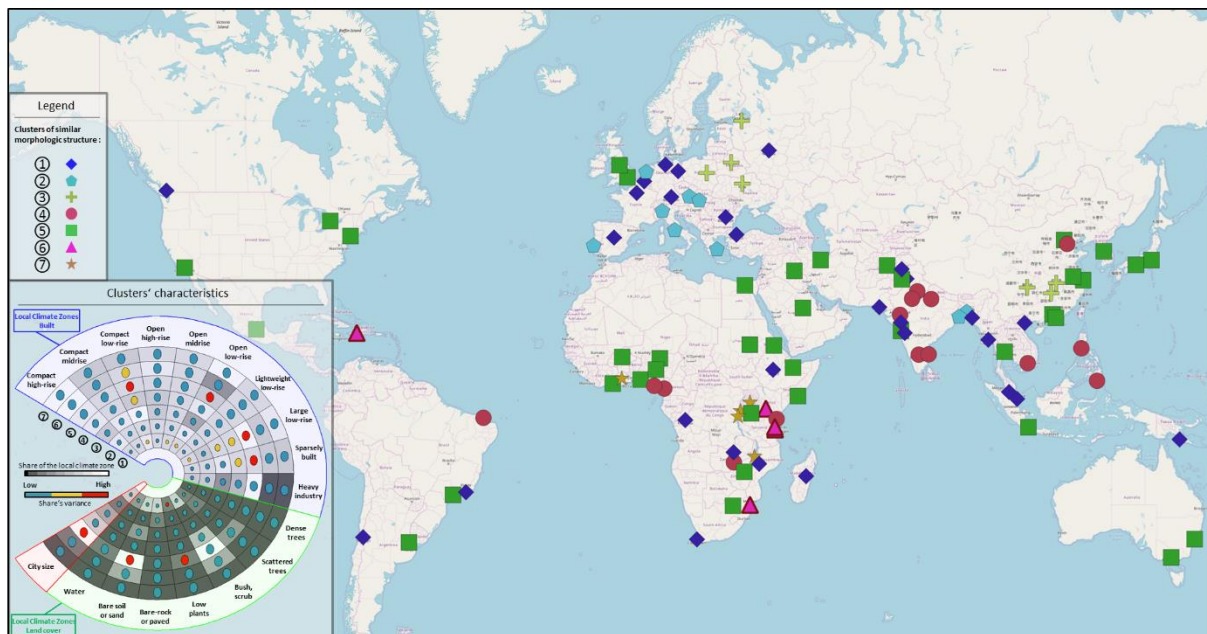
Annexes

Abidjan	Buenos Aires	Guangzhou	Kiev	Maputo	Poona	Tokyo
Abuja	Bujumbura	Ha Noi	Kigali	Melbourne	Port Au Prince	Toronto
Addis Ababa	Cairo	Hamburg	Kinshasa	Mexico City	Port Harcourt	Vancouver
Ahmadabad	Calcutta	Harare	Kitwe	Milan	Port Moresby	Vienna
Amsterdam	Calicut	Hargeysa	Kuala Lumpur	Minsk	Rio De Janeiro	Warsaw
Antananarivo	Cape Town	Ho Chi Minh City	Kumasi	Mogadishu	Riyadh	Wuhan
Asmara	Changsha	Hong Kong	Lagos	Mombasa	Rome	Yangon
Athens	Chennai	Islamabad	Lahore	Moscow	Saint Petersburg	Zanzibar
Baghdad	Chittagong	Istanbul	Lilongwe	Mumbai	Santiago	
Bangalore	Chongqing	Jaipur	Lisbon	Munich	Sao Paulo	
Bangkok	Dar Es Salaam	Jakarta	London	Mwanza	Seoul	
Beijing	Delhi	Johannesburg	Los Angeles	Nairobi	Shanghai	
Berlin	Dhaka	Kabul	Lucknow	Nanjing	Singapore	
Blantyre	Douala	Kampala	Lusaka	New York	Surat	
Brussel	Faisalabad	Kano	Madrid	Osaka	Sydney	
Bucharest	Fortaleza	Karachi	Manchester	Ouagadougou	Tehran	
Budapest	General Santos City	Khartoum	Manila	Paris	Tianjin	

Annexe 1: Table des villes considérées dans l'étude



Annexe 2: Résultat du clustering sur les 3 PCs par OPTICs



Annexe 3: Résultat du clustering sur les 18 paramètres sémantiques par OPTICs

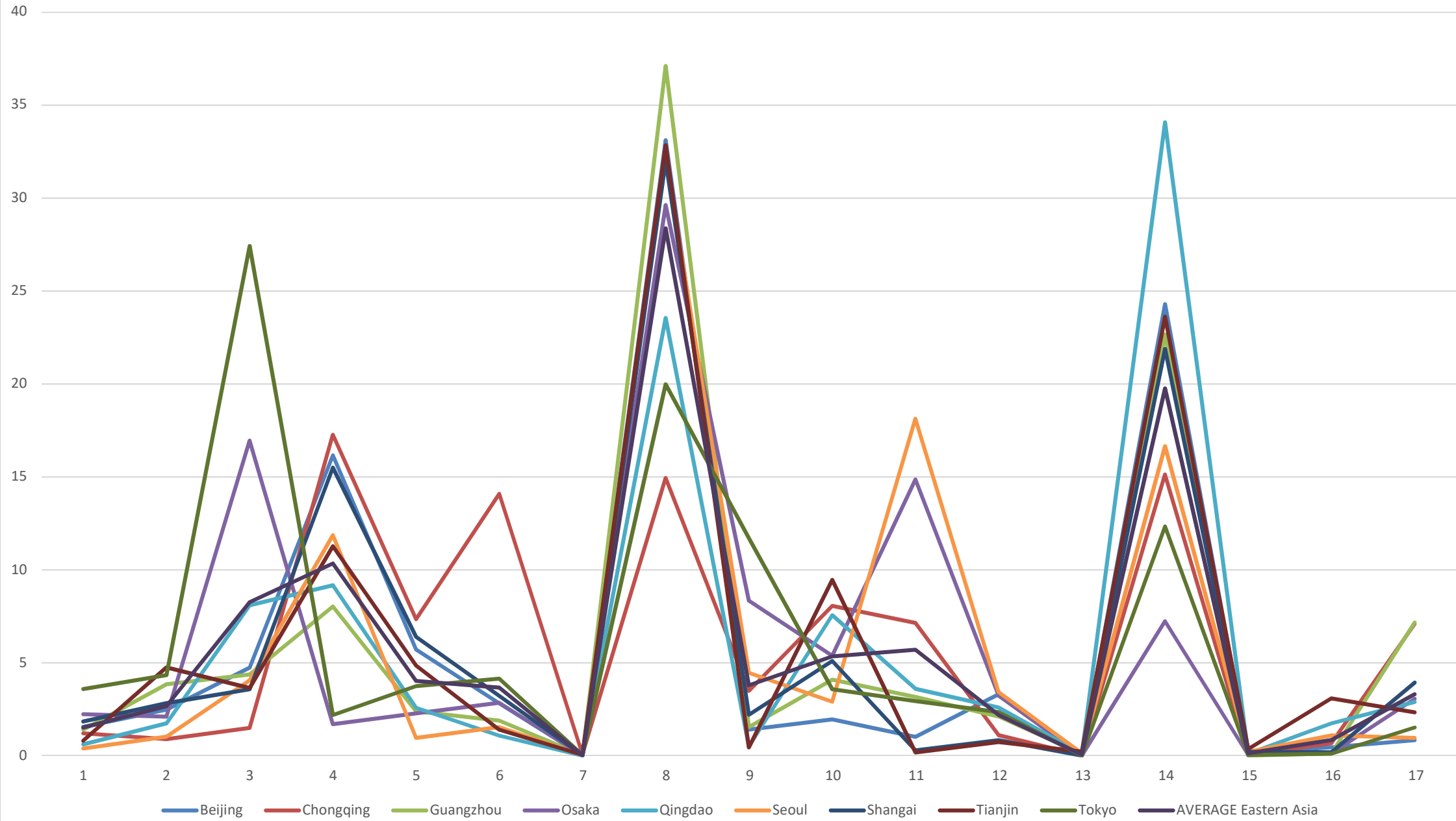


Annexe 4: Résultat du clustering sur les 101 paramètres géographiques par OPTICs

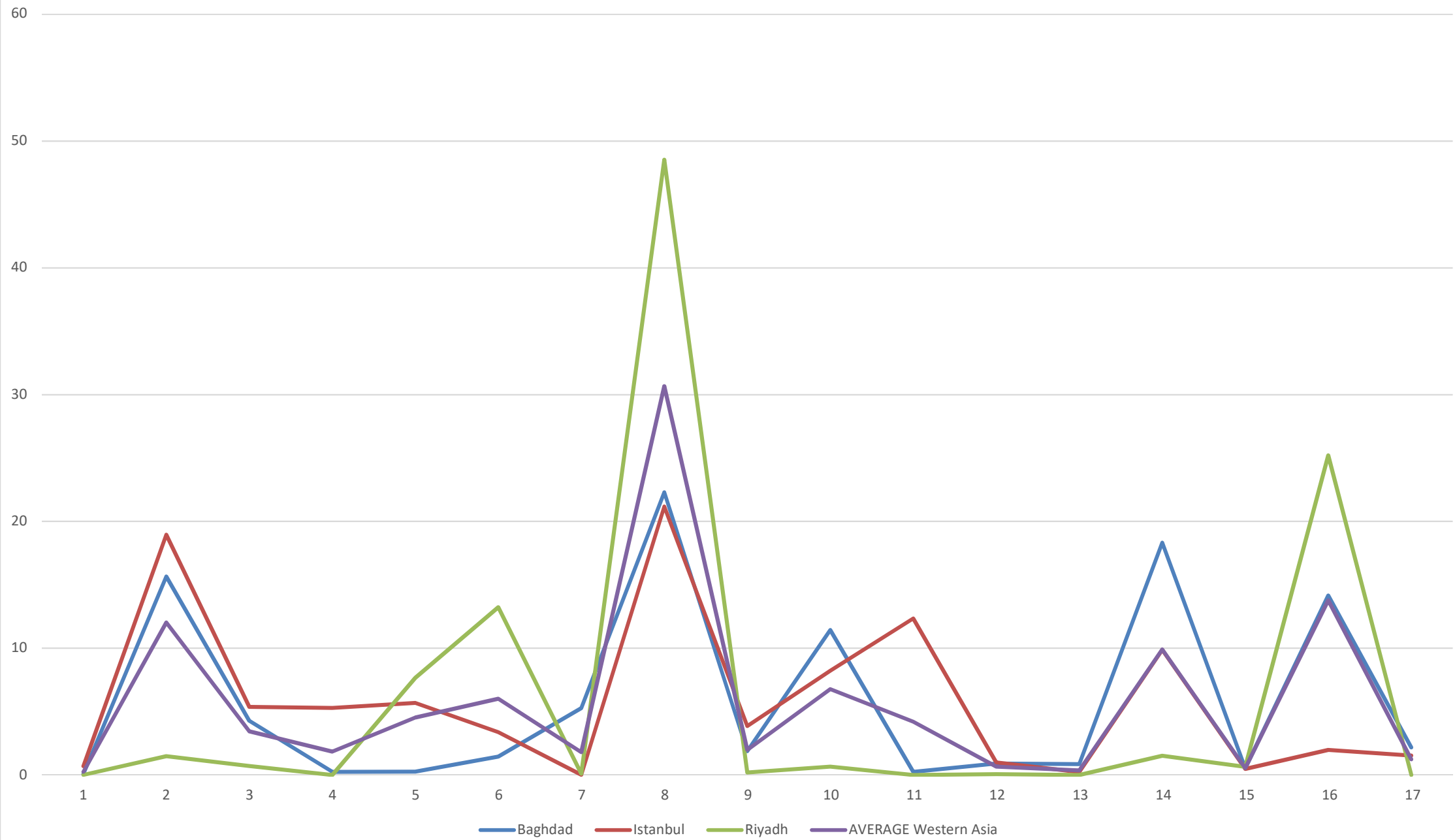


Annexe 5: Résultat du clustering sur les 1701 paramètres sémantiques et géographiques par OPTICs

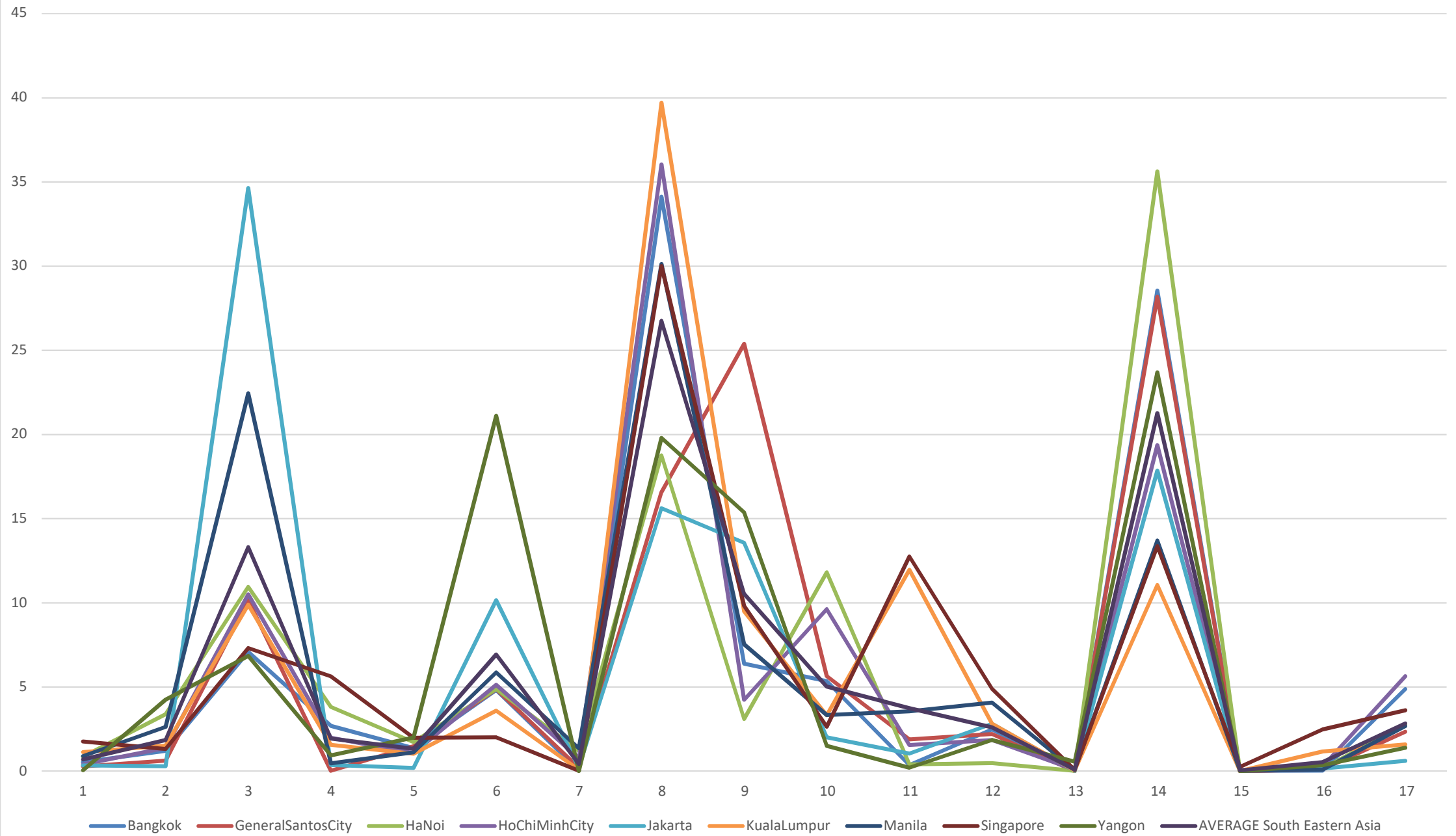
Eastern Asia



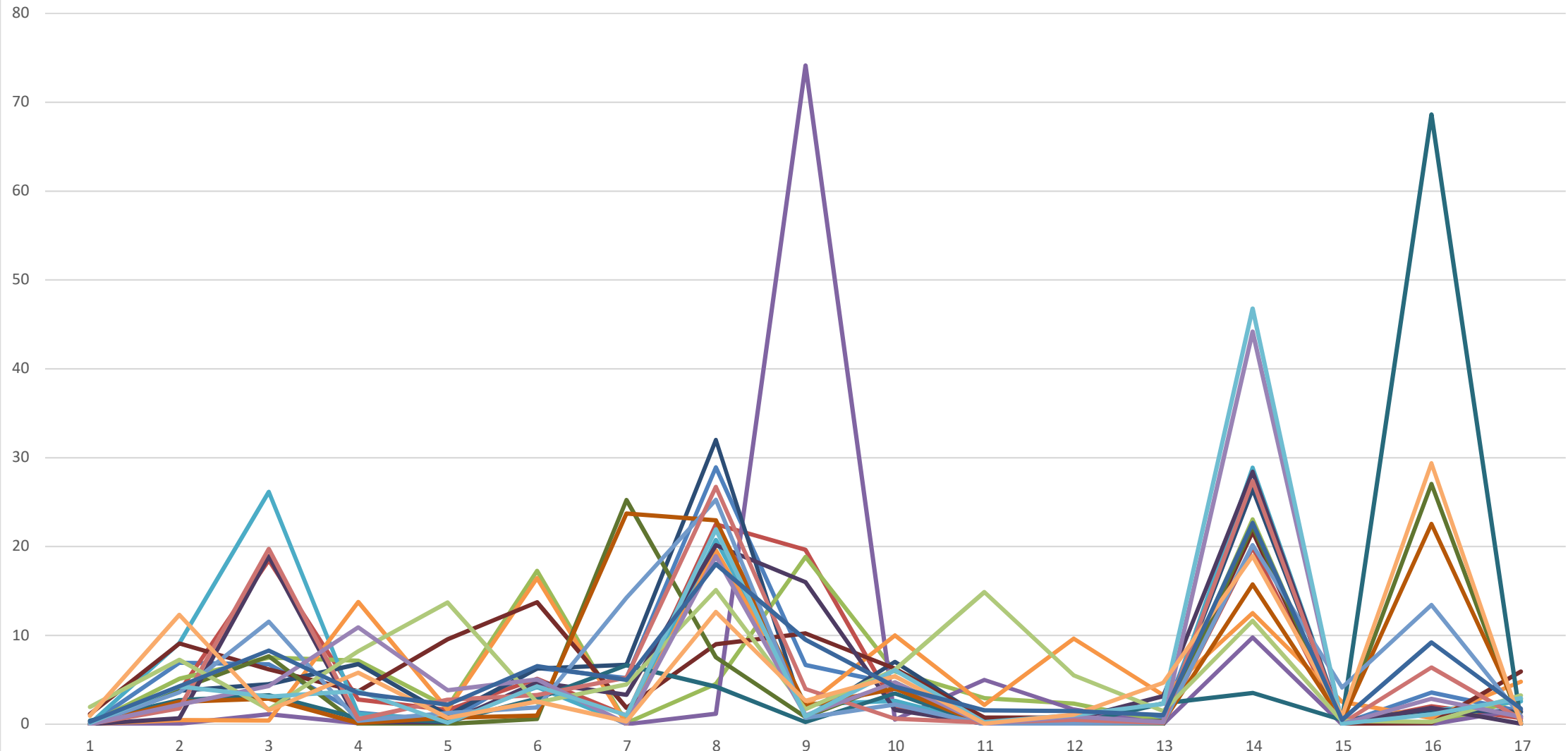
Western Asia



South Eastern Asia

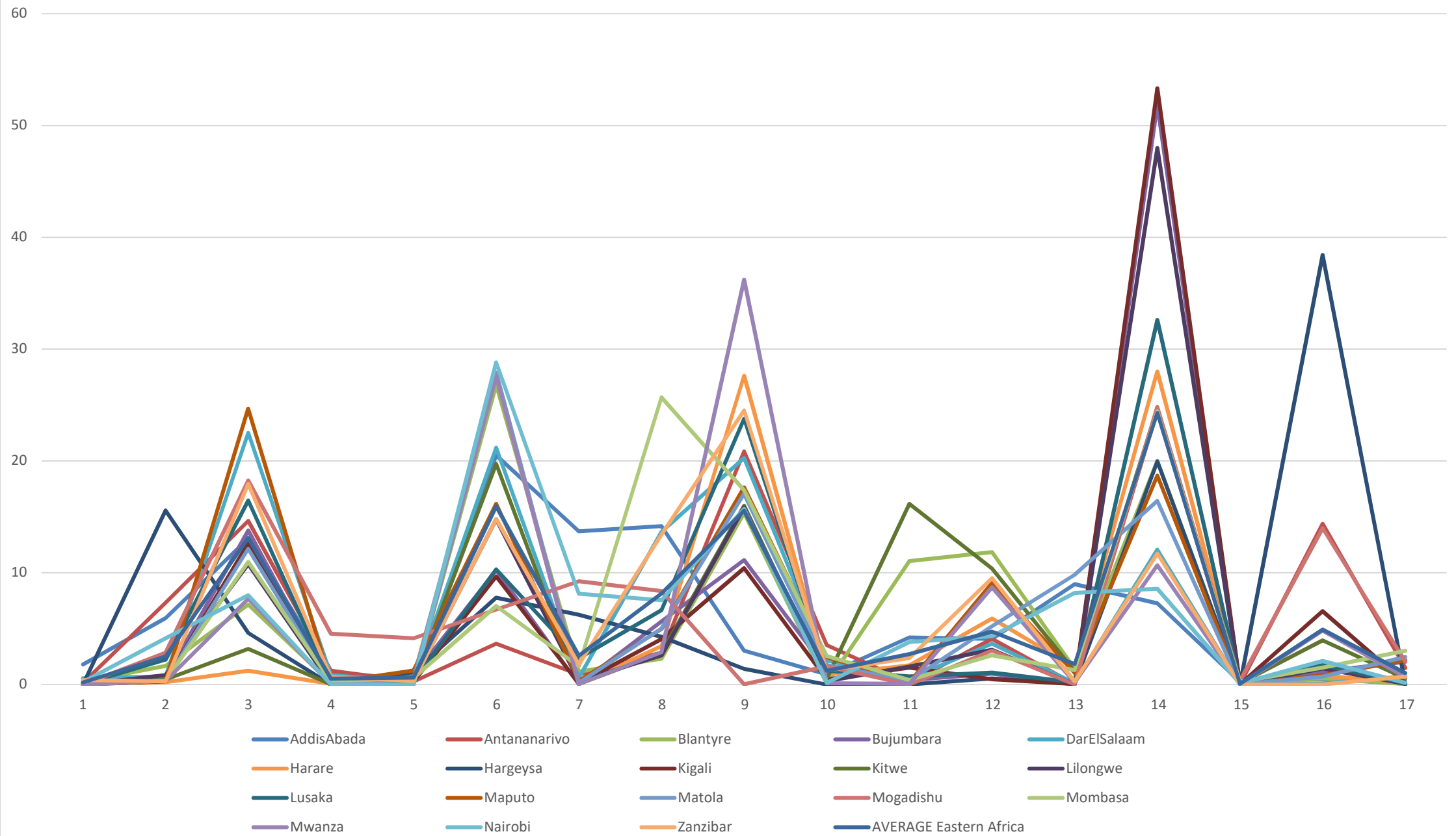


Southern Asia

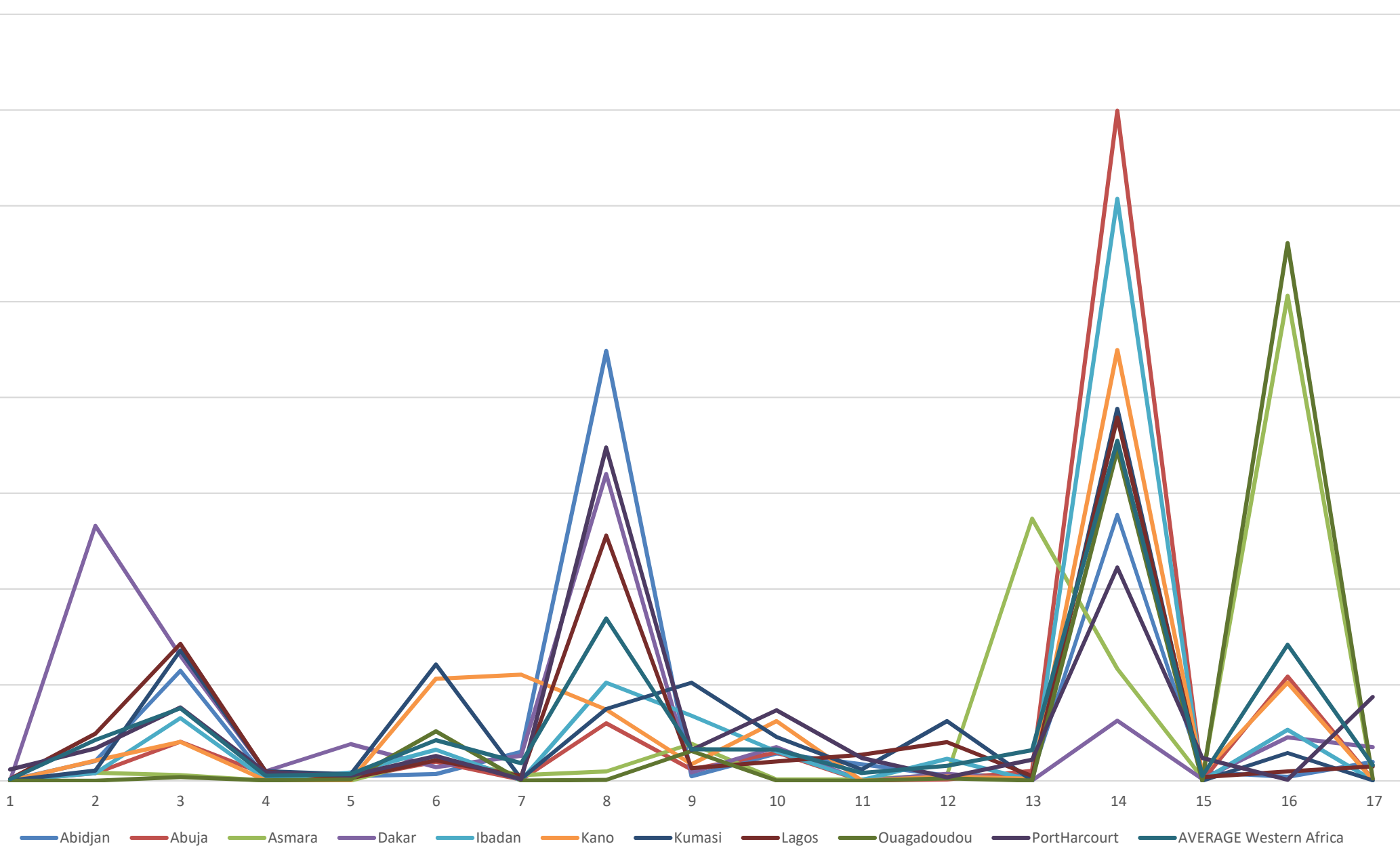


- | | | | | |
|--------------|-------------|------------|-------------------------|-----------|
| — Ahmadabad | — Bangalore | — Calcutta | — Calicut | — Chennai |
| — Chittagong | — Delhi | — Dhaka | — Faisalabad | — Jaipur |
| — Kabul | — Karachi | — Lahore | — Lucknow | — Mumbai |
| — Poona | — Surat | — Tehran | — AVERAGE Southern Asia | |

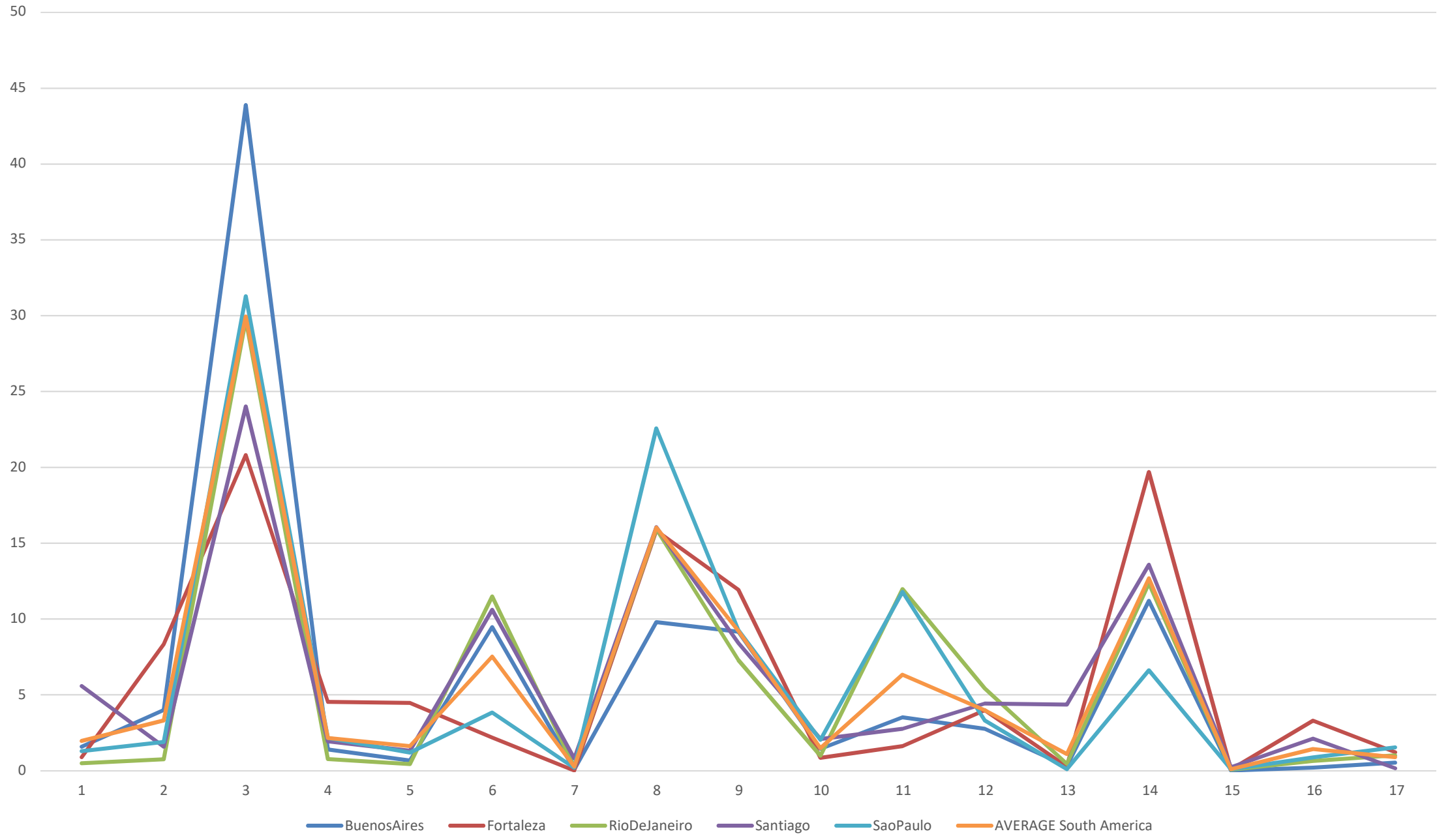
Eastern Africa



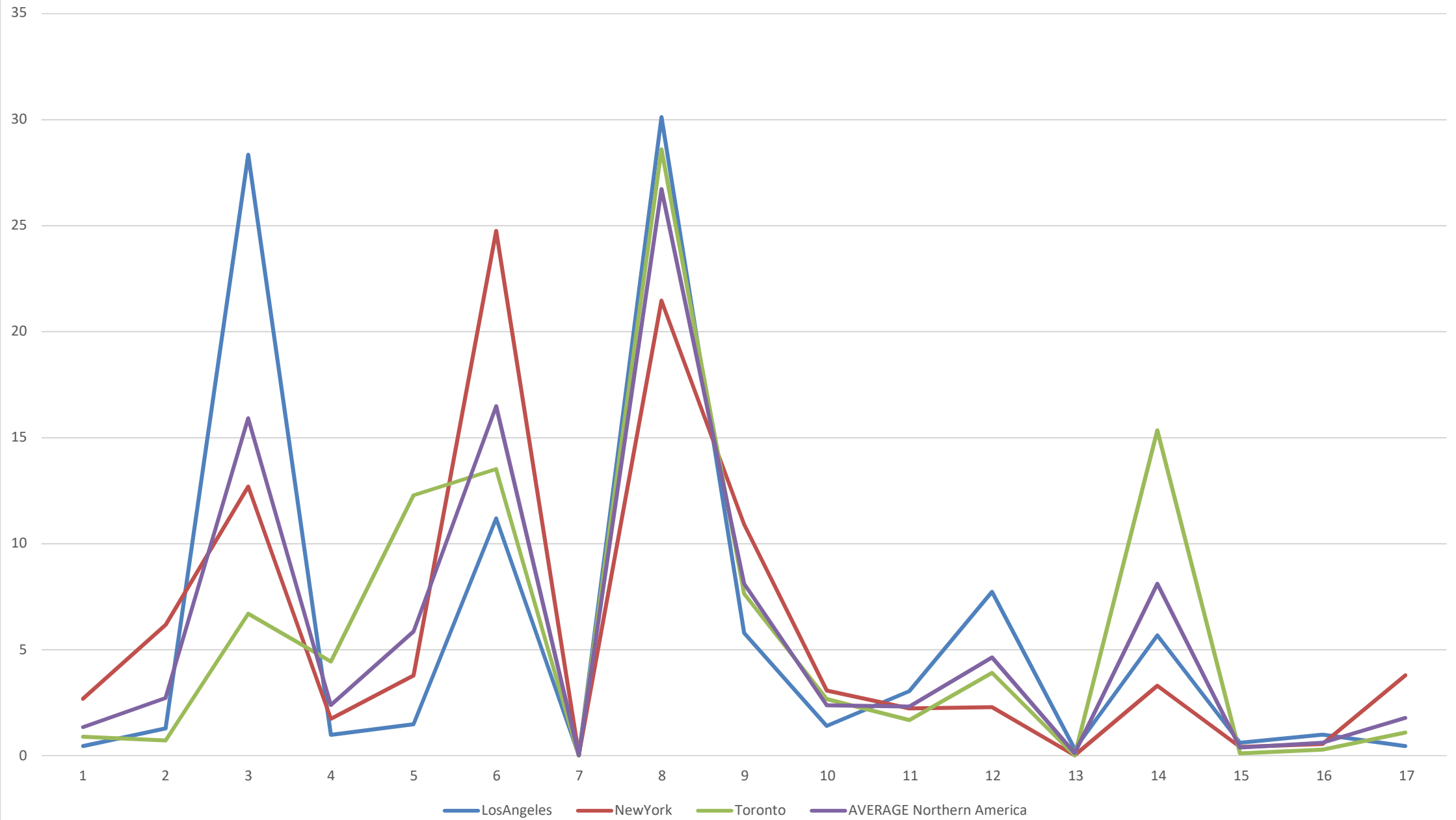
Western Africa



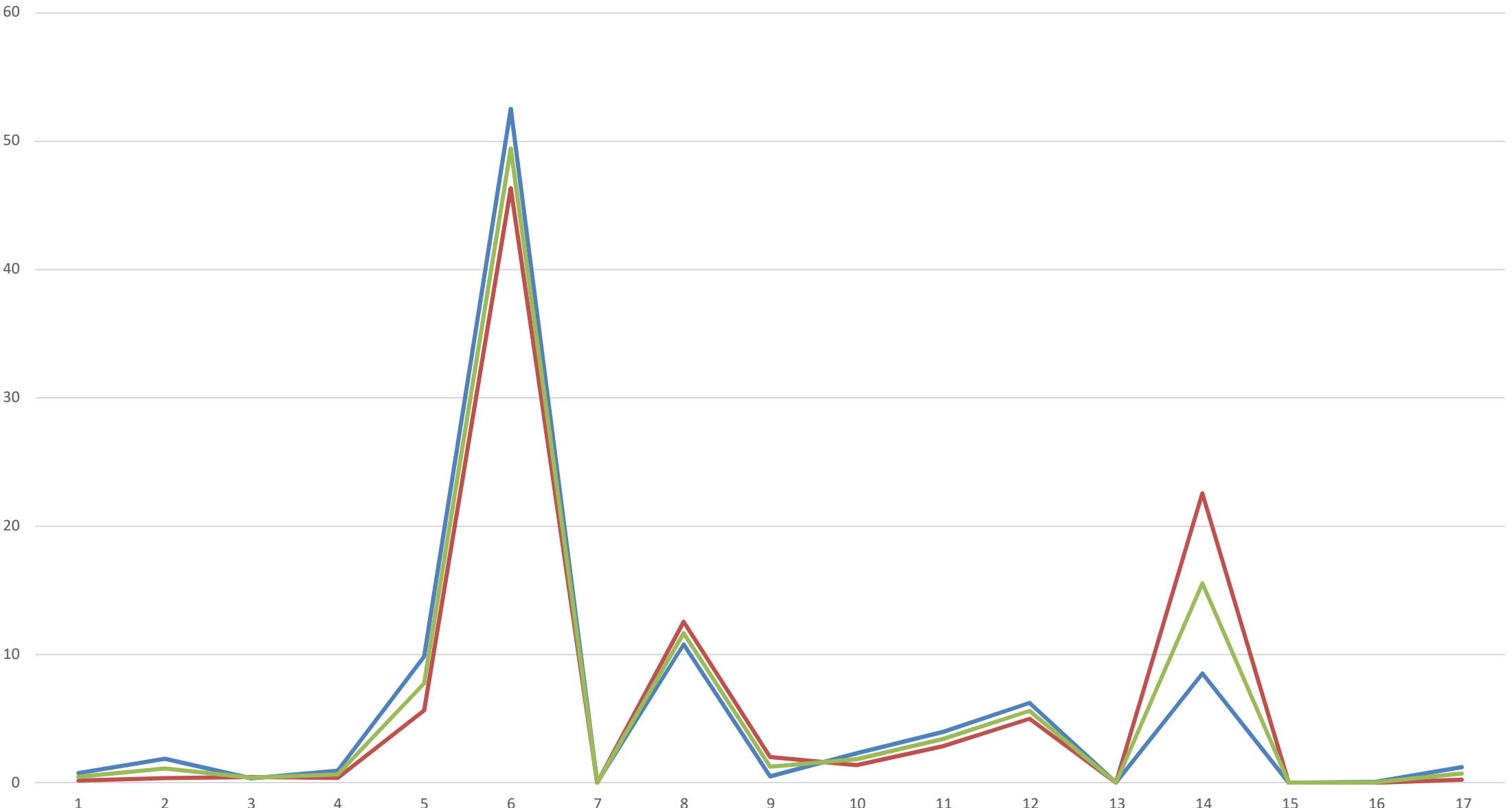
South America



North America

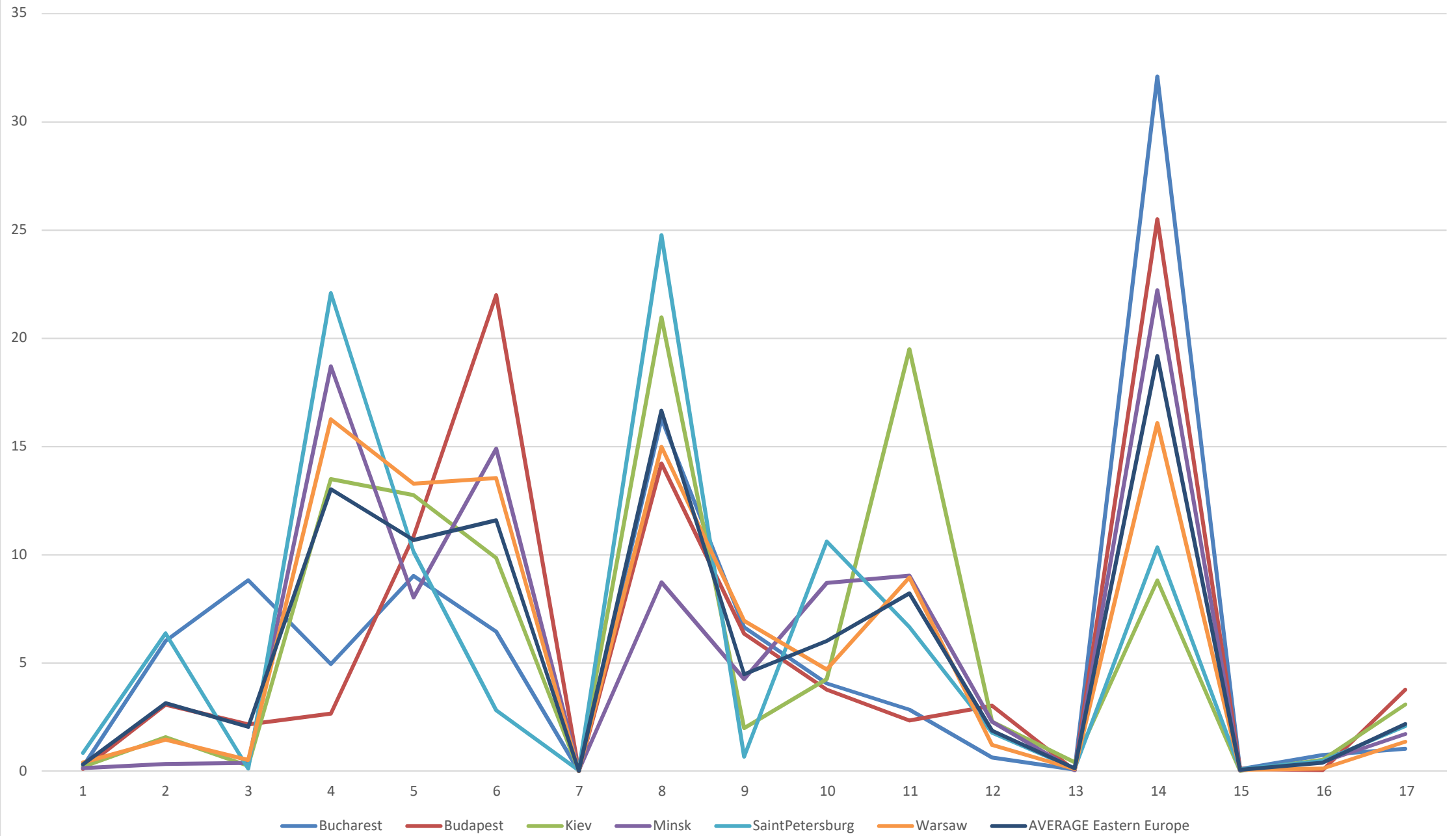


Northern Europe

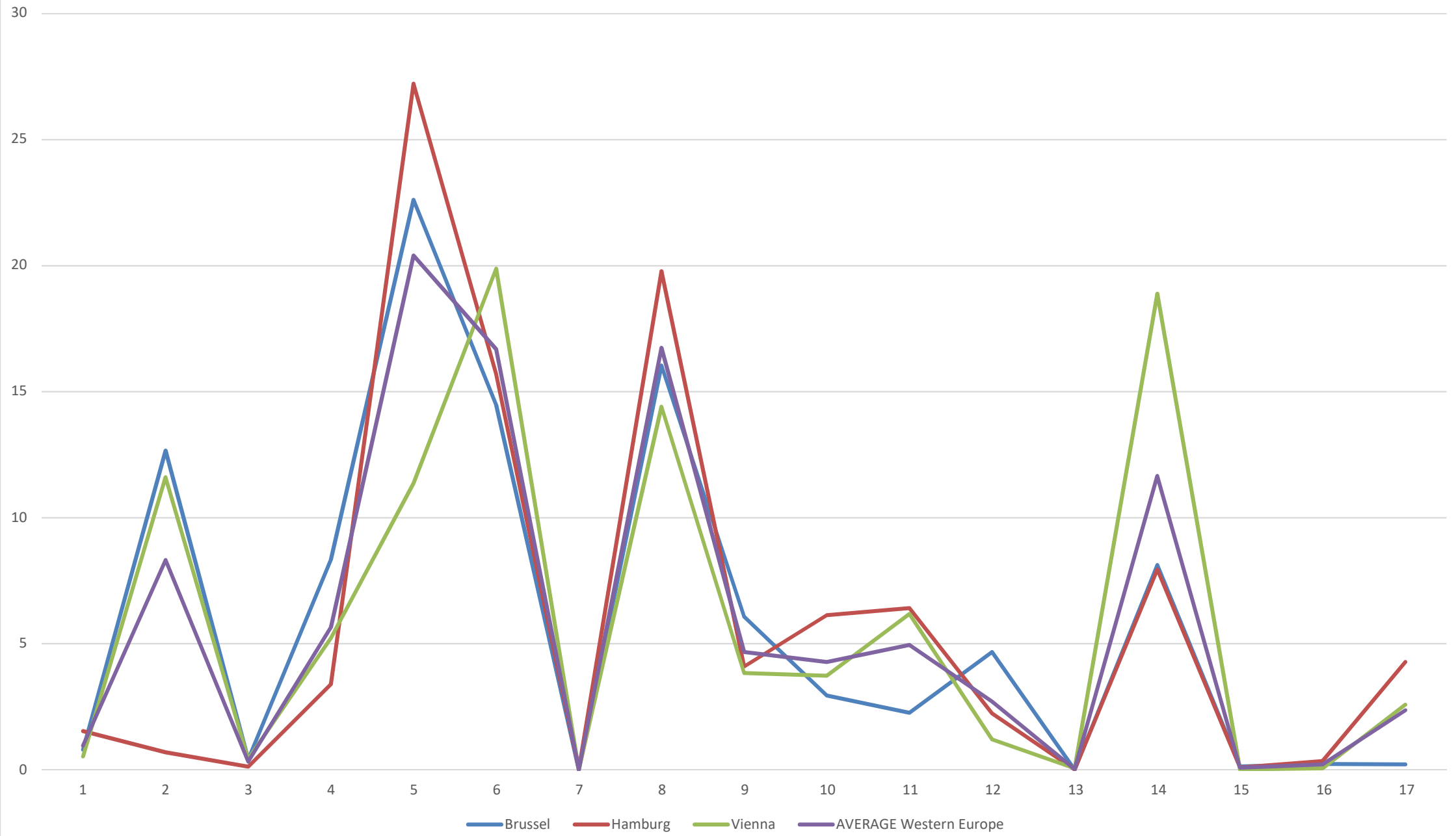


— London — Manchester — AVERAGE Northern Europe

Eastern Europe



Western Europe





Assessment Form TFE Internship

1. Practical Information.

Intern's last name and first name: Henri Debray

Name of the host organization: German Space Agency (DLR)

Internship Manager: Dr. Hannes Taubenböck Function: Leader of the team "City and Society"

Phone: +49 8153 28 2480 Email: hannes.taubenboeck@dlr.de

2. Internship evaluation.

For what reasons did you accept a student of the EIVP??

- To strengthen ties with the EIVP
- To reinforce an overworked team
- To pass on your knowledge
- To assess our students in the field with a view to future recruitment
- To introduce a new outlook with a view to the possibility of change

Yes	Partly	No
	X	
X		
	X	
X		
	X	

• Other (please specify):

Did the student occupy a specific post? Yes No

If yes, please specify? supporting us in an European Project

Was the student introduced to the rest of the team? Yes No

What responsibilities did you give him / her?

- At the start of the assignment He should get to know the topic of "urbanization" with a special focus on physical transformation of cities. He was supposed to analyze existing local Climate Zones classifications
- During the assignment: He was supposed to develop own ideas and methods to find patterns of city configurations across the globe; he had to support me and the team in other project work.

What follow-up do you foresee for the work carried out by the intern?

- | | | |
|---|---|--|
| • Interest limited to a written report for the time being | Yes <input type="checkbox"/> | No <input checked="" type="checkbox"/> |
| • A new intern will carry on the work | Yes <input type="checkbox"/> | No <input checked="" type="checkbox"/> |
| • Our departments will carry on the work | Yes <input checked="" type="checkbox"/> | No <input type="checkbox"/> |
| • A sub-contractor will carry on the work | Yes <input type="checkbox"/> | No <input checked="" type="checkbox"/> |

In what way can this apprenticeship in the field be reinforced by the EIVP?

We are very satisfied with Henri Debray, with his technical education level as well as with his social skills. We would be happy to host other students from EIVP in the future

Could the internship report be published? Yes No

because we work on a publication

3. Assessment of the intern.

= 5	= 4	= 3	= 2	= 1	→Note
Interest in the work					
Very interested in the work. Very enthusiastic. Makes it a point of honour to work well.	Above average interest and enthusiasm for the work	Satisfactory interest and motivation for the work	Mediocre interest and motivation	Little interest in the work	5
Initiative					
Autonomous. Asks for new tasks. Looks for work to do. Does not waste his/her time	Acts with some degree of autonomy on most jobs	Acts with some degree of autonomy on routine jobs	Relies on others. Often waits to be told what to do.	Always waits to be told what to do	5
Reliability					
Can always be relied on in every situation	Can generally be relied on in most situations	Can be relied on in routine situations	Not very reliable, needs above average supervision	Very unreliable	5
Ability to learn					
Brilliant and imaginative	Learns quickly	Average	Slow to learn	Very slow to learn	5
Creativity					
Constantly looks for new ways of carrying out his/her tasks; is extremely innovative	Often suggests new ways of carrying out his/her tasks; is very imaginative	Is of average imagination, has a reasonable number of new ideas	Occasionally has new ideas	Seldom has new ideas, is unimaginative	5
Necessary technical knowledge to carry out work					
Excellent	Very good	Average	Weak	Unsatisfactory	4
Analytical skills					
Excellent	Very good	Average	Weak	Unsatisfactory	4
Judgement					
Very good. Decisions based on a perfect understanding of the problems	Good. Usually makes the right decisions	Usually makes the right decisions in routine situations	Judgement is often unreliable	Poor judgement. Jumps to conclusions without adequate knowledge	5
Quality of work					
Very conscientious in carrying out tasks, makes very few mistakes	Usually conscientious. Good work, few mistakes	His/her work usually well thought of but contains some mistakes	Above average number of mistakes for an intern	Work carried out in a negligent manner with numerous mistakes	5
Quantity of work					
Very productive person	Productivity above expectations	Productivity corresponding to expectations	Productivity below expectations	Unsatisfactory	5
Written communication					
Always clear, well organized and easy to understand	Normally very clear, well organized and easy to understand	Usually clear and concise	Occasionally has difficulty writing clearly and concisely	Lacks clarity, which leads to confusion and damages productivity	4
Oral communication					
Always clear, well organized and easy to understand	Normally very clear and understandable	Usually clear and concise	Sometimes has difficulty expressing himself/herself clearly and concisely	Lacks clarity, which leads to confusion and damages productivity	4
Ability to manage his/her work					
Excellent	Very good	Acceptable	Weak	Unsatisfactory	5
Ability to adapt					
Excellent	Very good	Acceptable	Weak	Unsatisfactory	5

= 5	= 4	= 3	= 2	= 1	→ Note
Interpersonal qualities					
Excellent team member. Contributes to the harmony and efficiency of the group	Pleasant and helpful. A good team member	Gets on well with others	Occasionally ignores others and retreats into silence	Frequently at odds with others or withdrawn. Harms the group	5
Adherence to work regulations and procedures					
Is well-informed of the regulations in force and adheres to them diligently	Adheres well to the procedures	Generally adheres to the organization's regulations	Doesn't recognize the importance of sometimes critical procedures	negative attitude towards regulations and procedures	4
Accountability					
Excellent. Total trust between the intern, the team and the hierarchy	Able to share problems and results without getting lost in details	Tends not to know the limits between too much and too little	Weak. It is necessary to ask for essential elements	Unsatisfactory	5
Presentation					
	Always impeccable	Conforms to the company	Sometimes barely acceptable	Over-casual	4
Punctuality					
		Always on time	Regular	Irregular	3
Politeness					
		Courteous, respectful	Correct	Disrespectful, bad language	3
Intern's contribution					
What grade out of 5 would you give for the intern's contribution (human and /or technical) to your organization?					5
Total (out of 100)					97

If there were an opening, would you offer this intern a permanent position?

Yes



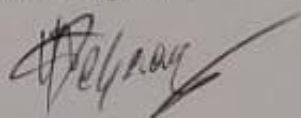
No



If not, please explain why:.....

Student's name Henri Debray.....

Student's signature



Mentor's signature

