



HAL
open science

Développement logiciel d'une plateforme pour l'évaluation de l'étalement urbain à partir de données participatives

Lucas Rezakhanlou

► To cite this version:

Lucas Rezakhanlou. Développement logiciel d'une plateforme pour l'évaluation de l'étalement urbain à partir de données participatives. Modélisation et simulation. 2017. hal-01593252

HAL Id: hal-01593252

<https://inria.hal.science/hal-01593252>

Submitted on 26 Sep 2017

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

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



INRIA Grenoble Rhône-Alpes - Equipe STEEP

Rapport de stage ING4

Développement logiciel d'une plateforme pour l'évaluation de l'étalement urbain à partir de données participatives

Auteur :
Lucas REZAKHANLOU

Encadrants :
Peter STURM
Luciano GERVASONI
Serge FENET

18 septembre 2017

Remerciements

Je tiens à remercier tout particulièrement M. Peter STURM, mon maître de stage et directeur de recherche au sein de l'équipe "Sustainability Transition, Environment, Economy and local Policy" (STEEP), qui m'a accordé sa confiance pour la réalisation du projet et sa disponibilité durant toute la durée du stage.

Je remercie également M. Luciano GERVASONI, doctorant au sein de l'équipe, pour sa gentillesse et son accueil, ainsi que pour toute l'aide qu'il m'a donnée pour comprendre les enjeux et les applications du projet.

Je tiens aussi à remercier M. Serge FENET, chercheur au sein du Laboratoire d'Informatique en Image et Systèmes d'information (LIRIS) du CNRS, pour ses nombreux conseils et son partage d'expérience.

Enfin, je souhaite remercier l'ensemble de l'équipe STEEP de l'INRIA Grenoble Rhône-Alpes, pour l'accueil chaleureuse et le partage dans lequel j'ai pu évoluer durant cette expérience de quatre mois.

Table des matières

Introduction	4
1 Présentation et organisation des travaux	5
1.1 Présentation de la structure d'accueil	5
1.1.1 L'INRIA	5
1.1.2 Equipe projet STEEP	5
1.2 Cahier des charges et planning des travaux	6
2 Etat de l'art et définitions	8
2.1 Définitions des indices utilisés	8
2.1.1 L'indice de mixité d'usage des sols (Land Use Mix)	8
2.1.2 L'indice d'accessibilité	9
2.1.3 L'indice de dispersion	9
2.2 Réalisation du site	9
2.2.1 Lien avec l'outil <code>urbansprawl</code>	9
2.2.2 Choix de l'outil d'extraction des données géographiques	10
3 La plateforme USAT	12
3.1 Méthode de réalisation	12
3.1.1 Conception générale du site	12
3.1.2 Modèle de la base de données	13
3.1.3 Processus détaillé du fonctionnement de la plateforme	14
3.2 Organisation de la plateforme	15
3.3 Développement du serveur	16
3.4 Organisation des tests	16
3.5 Résultats obtenus	18
3.5.1 Lancement des analyses	18
3.5.2 Résultats des calculs des indices géographiques	19
4 Analyse des travaux	21

Table des matières	3
<hr/>	
4.1 Analyse des résultats	21
4.2 Difficultés rencontrées	22
4.3 Perspectives	22
Conclusion	23
A Captures d'écran de la plateforme	25
B Algorithme de représentation géographique des résultats	28

Introduction

La population vivant dans des centres urbains a augmenté de façon considérable depuis les années 50, passant de 746 millions à 3,9 milliards en 2014[6]. En parallèle avec un léger déclin de la population rurale globale prévue dans les 35 prochaines années, l'augmentation de l'urbanisation devrait ajouter près de 2.5 milliards de personnes dans les centres urbains d'ici 2050. Cette situation impose de nouveaux défis liés à la conception de villes capables d'héberger de telles populations de manière soutenable et nous force à penser les villes du futur sous des aspects économiques, sociaux et environnementaux. Afin de participer au processus d'organisation de ces espaces, nous nous focalisons ici sur l'exploitation d'indices mesurant de manière globale l'étalement urbain, afin d'analyser les évolutions d'une zone urbaine selon différents aspects, en lien avec l'urbanisation et la péri-urbanisation des villes. L'intérêt d'avoir une mesure globale de ces phénomènes est de fournir un support universel aux planificateurs urbains et acteurs politiques du monde entier.

Dans le cadre de ses travaux, l'équipe STEEP de l'INRIA Grenoble Rhône-Alpes a proposé de développer une plateforme Web permettant d'effectuer ces mesures, rendant ainsi disponible à un utilisateur l'analyse de n'importe quelle zone du monde, respectant l'idée de globalité souhaitée. Cet outil d'aide à la décision est construit à partir de données libres d'utilisation dans la logique de l'Open Source, et la construction d'indicateurs multiples est issue de l'état de l'art du domaine. Nous allons dans un premier temps expliquer le cadre de ce travail en présentant la structure d'accueil du stage, les travaux de l'équipe, puis l'organisation respectée pour le projet. Dans un second temps, une reprise de l'état de l'art sera donnée afin d'expliquer les indices permettant d'analyser l'étalement urbain de manière globale. Cette présentation sera suivie d'une explication détaillée de la plateforme Web réalisée. Enfin, une analyse et une conclusion seront données sur le travail effectué et cette expérience de quatre mois.

Chapitre 1

Présentation et organisation des travaux

1.1 Présentation de la structure d'accueil

1.1.1 L'INRIA

L'Institut National de Recherche en Informatique et en Automatique (INRIA) est un des organismes de recherche public français, rattaché à trois ministères : le Ministère de l'Education Nationale, celui de l'Enseignement Supérieur de la Recherche, et celui de l'Industrie. Il se caractérise par ses travaux à caractères scientifiques et technologiques, particulièrement en sciences mathématiques et informatiques.

L'Institut est composé de huit centres de recherches et son siège est situé à Rocquencourt dans les Yvelines. Les centres de recherches en France sont à taille variables, mais l'organisation interne est similaire : ils sont composés de plusieurs équipes de recherches, nommées "équipes-projets", ayant des domaines et thèmes d'études différents. En moyenne, chaque équipe est formée de 10 à 20 personnes. Au total, l'INRIA emploie plus de 2600 collaborateurs de nombreuses nationalités et diverses universités qui fournissent des travaux pouvant être exploités par différentes entreprises et industries. De manière générale, les projets qui y sont encadrés ont une visée pluridisciplinaire et portés vers la transition numérique, comme l'atteste son slogan "Inventeurs du monde numérique". Tous les types d'industries sont touchés concernées : santé, réseaux informatiques et sécurité, transports, ou énergie. On peut ainsi citer en exemple la création du logiciel "Scilab" permettant d'effectuer du calcul scientifique, ou encore le logiciel "SOFA", utilisé pour effectuer des simulations physiques notamment appliquées au domaine médical.

1.1.2 Equipe projet STEEP

STEPP, sigle pour "Sustainability Transition, Environment, Economy and local Policy", est une équipe-projet du centre INRIA Grenoble Rhône-Alpes (GRA). Les travaux de recherches qui y sont menés sont principalement liés à la modélisation et la simulation des interactions entre les facteurs environnementaux, sociétaux et économiques à l'échelle locale (une région ou une métropole par exemple). L'analyse de ces interactions dans un contexte de transition vers un développement durable local accompagne la décision des acteurs locaux, qui peuvent anticiper l'impact qu'aurait une décision sur la biodiversité, ou sur l'organisation d'une ville, d'un territoire,

voire d'une filière industrielle. L'objectif de l'équipe est d'apporter des outils mathématiques et informatiques qui seront utiles pour faire face aux challenges actuels et à venir, autour de la transformation des villes et du changement climatique.

Pour mener à bien ces études et travaux, STEEP s'entoure de compétences au travers de partenariats, avec par exemple le Laboratoire d'Ecologie Alpine, le Laboratoire d'Economie des Transports, ou encore le Laboratoire d'Informatique en Image et Systèmes d'Information. L'ensemble des partenaires participent à l'élaboration des travaux à travers des financements de thèses de doctorat, ou différentes études qui ont lieu entre les chercheurs pour l'exploration de nouveaux thèmes de recherche. Ces sujets de recherches sont sans cesse en évolution, suivant les objectifs de l'INRIA ou les nouvelles problématiques qu'ils produisent. Concernant l'équipe STEEP, avant d'obtenir le statut d'équipe-projet en 2015, elle avait celui d'action exploratoire (à partir de 2010). Ce statut permet aux chercheurs de l'équipe de tester de nouvelles thématiques de recherche, parfois risquées et en rupture avec les approches traditionnelles de l'INRIA. Fin 2018, l'équipe atteindra la fin d'un cycle de recherche et va devoir présenter les nouvelles thématiques qu'elle souhaite explorer dans ses travaux. Il s'agira à ce moment-là de possiblement modifier les problématiques que cherche à étudier l'équipe et ses membres.

L'équipe est organisée de manière similaire aux autres équipes du centre INRIA GRA. Elle est composée de chargés de recherche, personnel administratif, post-doctorants, doctorants, ingénieurs, stagiaires. Les chargés de recherche encadrent les études des doctorants ou stagiaires, et sont principalement à l'origine des voies d'exploration vers de nouveaux travaux.

1.2 Cahier des charges et planning des travaux

L'objectif de ce stage, comme présenté en introduction, est de s'appuyer sur le travail auparavant effectué par l'équipe sur l'étalement urbain et la définition de trois indices permettant de l'évaluer, afin de fournir une plateforme web servant à les analyser pour un territoire donné. Les trois indices qui ont été définis sont : l'indice de mixité d'usage des sols (Land Use Mix), l'indice d'accessibilité, et l'indice de dispersion. Ils seront expliqués de manière plus détaillée dans le chapitre suivant. Cette plateforme doit être graphique, facile d'utilisation, et doit permettre la comparaison des résultats entre eux. L'analyse d'une zone doit pouvoir se faire de manière intuitive et des paramètres relatifs aux calculs d'un indice doivent pouvoir être modifiés. Une utilisation en "batch" du site doit être aussi disponible, cela signifie qu'un utilisateur peut soumettre plusieurs fichiers contenant les paramètres d'une analyse et obtenir les résultats pour plusieurs zones en une seule fois. Cette fonctionnalité est tournée vers les utilisateurs souhaitant obtenir rapidement des résultats personnalisés.

Pour calculer les indices, il est nécessaire d'utiliser une base de données contenant les informations d'une ville. Parmi ces informations, on peut trouver par exemple l'emplacement des zones d'activité, des zones industrielles, de boutiques, et des différentes commodités. Il est important d'utiliser une source cartographique ouverte et libre qui nous permettra de proposer librement notre travail dans une logique OpenSource. Pour cela nous préférons utiliser les données géographiques de la base OpenStreetMap (OSM) au lieu d'une source payante, telle que Google Maps. Cette base de données est remplie de manière collaborative par des utilisateurs du monde entier qui référencent des informations géographiques. Cela peut aller d'une salle cinéma au banc situé au bord d'un carrefour. Par ailleurs, les résultats du calcul des indices doivent pouvoir être visualisés de manière dynamique et interactive par l'utilisateur. Il doit être également possible de les exporter dans un format exploitable par des acteurs venant de différents domaines : le Shapefile. Ce format vectoriel est largement adopté dans le secteur des Systèmes d'Informations Géographiques ("GIS" en anglais), et visualisable au travers de nombreux logiciels, comme QGIS par exemple.

Enfin, le serveur hébergeant la plateforme doit être configuré à travers une machine virtuelle et prêt pour le lancement en production. Il doit ainsi contenir l'ensemble des bibliothèques logicielles exploitées pour le calcul des indices, puis des scripts permettant sa mise-à-jour automatique afin de respecter les limites de stockage autorisées. Le serveur va être situé dans le centre de calcul CC-IN2P3 du CNRS, ainsi il est nécessaire de fabriquer une machine virtuelle qui soit rapide à configurer et à lier avec une machine physique.

Nous nous sommes donnés un jalon dans ce projet : la date de soumission d'un article court pour la participation à la conférence PFIA 2017 (Plate-forme Intelligence Artificielle). Il sera nécessaire de démontrer une preuve de concept et concevoir tous les aspects techniques avant la date limite de soumission de l'article, le 29 mai 2017. Ensuite, le stage se finissant le 18 août 2017, la plateforme Web et le serveur doivent être configurés et prêts au déploiement à cette date.

Chapitre 2

Etat de l'art et définitions

2.1 Définitions des indices utilisés

Avant de présenter chaque indice individuellement, il est nécessaire de rappeler la définition de l'étalement urbain. De manière simple, il s'agit du développement et de l'expansion des surfaces urbanisées autour d'une ville, au détriment de surfaces agricoles ou d'espaces ruraux. Il n'y a toutefois pas de définition consensuelle concernant sa définition formelle, et les caractéristiques qui différencient une ville étalée d'une ville non étalée [3, 4, 7]. Cependant, on sait que ce phénomène est principalement causé par l'augmentation démographique, et les mêmes conséquences se retrouvent dans les villes qui respecteraient une des définitions : baisse de la qualité de l'air et de l'eau, augmentation de la dépendance à la voiture et embouteillages systématiques, perte de surfaces agricoles, effets sur la santé des citoyens...

Afin d'évaluer l'étalement urbain il est nécessaire d'obtenir de nombreuses données sur la zone d'étude. Ces données doivent contenir la localisation des usages résidentiels et des activités (toutes activités confondues), ainsi que le réseau routier et urbain pour évaluer l'accès entre les différentes commodités. A partir de ces données nous pouvons ainsi expliquer les trois indices définis dans le papier [5].

2.1.1 L'indice de mixité d'usage des sols (Land Use Mix)

La mixité d'usage des sols est le rapport entre les surfaces dédiées aux usages résidentiels et celles dédiées aux activités pour une zone donnée. Les activités sont par exemple des industries, des boutiques, des bâtiments administratifs, des hôpitaux, etc. On peut ainsi montrer qu'une faible mixité d'usage des sols induit une plus forte utilisation des transports. Nous pouvons ainsi supposer que pour le centre-ville d'une grande métropole, en fonction du nombre de données disponibles, l'usage résidentiel et le nombre d'activités sont élevés : on y estime alors une forte mixité d'usage des sols. A l'inverse, dans une zone industrielle, considérant que peu d'usage résidentiels s'y situent, la mixité serait alors basse. Cet indice permet d'identifier les zones de faible mixité et de localiser les lieux où il serait possible de l'améliorer.

Il existe dans la littérature plusieurs manières de calculer cet indice [8], la méthode de calcul utilisée pour les analyses effectuée dans la plateforme est celle qui a été choisie dans [5].

2.1.2 L'indice d'accessibilité

L'indice d'accessibilité permet pour une zone donnée et une distance fixée de calculer le nombre de services accessibles. Imaginons ainsi un rayon de 250 mètres et votre lieu de résidence comme point de départ, si l'indice d'accessibilité vaut 200, cela signifie que vous avez accès à un très grand nombre d'activités. Plus l'indice est élevé, moins l'usage des transports est nécessaire. A l'inverse, si la valeur de l'indice est très faible, alors l'utilisation d'une voiture ou de transports en commun est plus probable. Pour mesurer cet indice, le réseau routier est extrait de la base OSM à partir de la zone définie par l'utilisateur. Une grille de points d'analyse, espacés d'une distance définie également par l'utilisateur (par défaut 100 mètres), est fabriquée. Pour chaque point on mesure le nombre d'activités disponibles dans un rayon défini. L'ensemble des étapes est également détaillé dans [5].

2.1.3 L'indice de dispersion

La dispersion est liée à la densité d'une zone donnée. On peut la définir comme l'inverse de la compacité. Dans une territoire donné, si une zone urbaine est construite et que les bâtiments sont très dispersés (autrement dit espacés), alors l'étalement urbain y est élevé. Un lien a été prouvé entre des espaces dispersés, la perte de territoires ruraux et la faible existence de transports publics. Ainsi, la dispersion permet de montrer à des planificateurs que des espaces peuvent être exploités autrement, permettant un usage plus rationnel des sols. Toutefois il est important de considérer les espaces verts de manière spécifique pour le calcul de cet indice puisqu'ils jouent un rôle important dans la qualité de vie des citoyens. L'indice de dispersion est mesuré en extrayant les données associées à des bâtiments depuis la base OSM, et en calculant la distance entre chaque bâtiment suivant la procédure définie dans [5].

2.2 Réalisation du site

2.2.1 Lien avec l'outil `urbansprawl`

Pour la réalisation technique de la plateforme, plusieurs langages de programmation doivent être utilisés, et des outils disponibles autour de ces langages sont disponibles afin de concevoir et déployer le site dans les meilleures conditions. Parmi les contraintes techniques du cahier des charges, la réutilisation des fonctions développées en langage Python permettant le calcul des indices est une étape centrale du projet. Il est ainsi nécessaire de faire le lien logiciel dans le serveur entre le projet `urbansprawl` précédemment développé¹, et le site développé au cours de ce stage. Pour cela, une étude sur les différentes possibilités techniques a été réalisée. Il existe des frameworks (traduit littéralement par "cadres de travail") pour le langage Python, listés dans la documentation², et permettant le développement d'une application web en utilisant ce langage de programmation. Pour rappel, la manière classique de développer des sites web se fait à travers le langage PHP. Les pages web visibles depuis un navigateur sont quant à elles développées généralement dans les langages HTML, CSS et JavaScript.

Certains frameworks disposent d'une communauté plus importante, et sont très fréquemment mis-à-jour afin de proposer des versions stables. Les deux les plus utilisés, Flask et Django, sont très largement cités pour leur simplicité d'usage et facilité de déploiement[1]. Des différences sont toutefois notables, elles sont présentées dans le tableau suivant.

1. <https://gitlab.inria.fr/gervason/urbansprawl>

2. <https://wiki.python.org/moin/WebFrameworks>

Caractéristiques	Plus grande communauté	Simplicité d'utilisation	Nombres de fonctionnalités
Framework	Django	Flask	Django

TABLE 2.1 – Comparaison synthétique des frameworks pour la réalisation du site.

Pour un utilisateur ayant de l'expérience en développement web, Django est une solution plus naturelle puisqu'une grande communauté permet de rapidement trouver les erreurs qui peuvent survenir. En revanche, Flask dispose de nombreuses fonctionnalités mais il est nécessaire de les ajouter manuellement, alors que Django en dispose dès la création d'un projet. C'est pour ces raisons, que le framework Django sera utilisé pour le projet.

2.2.2 Choix de l'outil d'extraction des données géographiques

Pour calculer les indices d'une zone définie, il est nécessaire d'extraire les données situées à l'intérieur des limites géographiques données, d'une manière rapide et dynamique. Plusieurs logiciels permettent de réaliser cette extraction, il est toutefois possible de télécharger ces données depuis des serveurs proposés par différents fournisseurs OpenSource ou non.

Outil	Méthode	Utilisation automatique possible	Utilisation illimitée	Temps d'extraction
OSMConvert ³	Commande d'extraction en spécifiant les coordonnées des limites géographiques	Oui	Oui	~25 minutes
Osmosis ⁴	Commande d'extraction en spécifiant les coordonnées des limites géographiques	Oui	Oui	~75 minutes
Mapzen Metro Extract ⁵	Téléchargement de zones géographiques spécifiques depuis le serveur	Non	Avec un compte utilisateur	Dépend de la taille de la zone
Overpass ⁶	Téléchargement des données depuis le serveur OpenStreetMap	Oui	Oui	Dépend de la taille de la zone
Geofabrik ⁷	Téléchargement de zones géographiques spécifiques depuis le serveur	Non	Oui	Dépend de la taille de la zone

TABLE 2.2 – Outils disponibles pour l'extraction de données géographiques

Il convient de noter une contrainte importante liée à l'utilisation des logiciels d'extraction, il est nécessaire d'avoir téléchargé au préalable l'ensemble des données mondiales géographiques

3. <http://wiki.openstreetmap.org/wiki/Osmconvert>

4. <http://wiki.openstreetmap.org/wiki/Osmosis>

5. <https://mapzen.com/documentation/metro-extracts/>

6. http://wiki.openstreetmap.org/wiki/Overpass_API

7. <http://www.geofabrik.de/data/download.html>

d'OpenStreetMap. Ces données sont regroupées dans un fichier compressé nommé "PlanetOSM", disponible au téléchargement libre⁸, mais d'une taille supérieure à 37Go. De plus, ce fichier étant mis-à-jour hebdomadairement du fait de l'ajout continu de nouvelles données, la taille est en constante augmentation. Pour donner un ordre de grandeur, ce fichier a grossi d'environ 2Go entre le début du stage et aujourd'hui, soit en 4 mois.

Le tableau précédent résume les principales solutions et leurs caractéristiques pour l'extraction des données géographiques pour des limites définies. D'après l'ensemble des caractéristiques, on peut déduire que OSMConvert offre la plus grande souplesse d'utilisation, mais nécessite tout de même le téléchargement préalable de PlanetOSM. Ce programme est disponible à l'installation pour Linux via le package "osmctools".

8. <https://planet.osm.org/>

Chapitre 3

La plateforme USAT

3.1 Méthode de réalisation

3.1.1 Conception générale du site

La plateforme, nommée "Urban Sprawl Analysis Toolkit" (USAT), a été réalisée suivant les limites imposées par le framework utilisé, Django, et les contraintes du cahier des charges. L'organisation du projet suit les principes du développement web, qui implique la création d'un serveur web dans lequel les pages seront stockées, et des scripts qui vont effectuer les actions que l'utilisateur aura choisies.

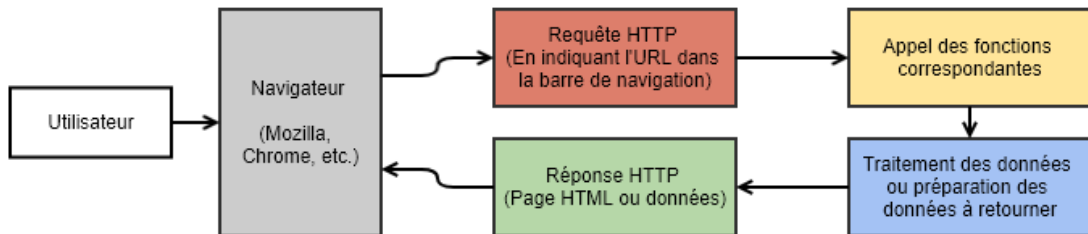


FIGURE 3.1 – Schéma du processus de fonctionnement d'un site web

A travers la figure précédente, on observe que les deux actions représentées par les blocs en rouge et vert sont faites après l'appel de l'utilisateur à une page ou à une action, et avant le retour de la page demandée ou du résultat de cette action. Ces blocs sont des traitements effectués au sein du serveur web, que l'on peut plus simplement nommer "traitement côté serveur". A l'inverse, les requêtes et des réponses sont envoyées et reçues via le navigateur par l'utilisateur. On nomme ces actions "traitement côté client". A partir de ces définitions il est possible de séparer les tâches du cahier des charges en deux groupes : les fonctions qui seront réalisées par le client (l'utilisateur), et celles qui seront réalisées dans le serveur. Logiquement, le rendu graphique des pages, la sélection des zones et des indices à analyser ainsi que l'affichage des résultats doit se faire sur un navigateur. A l'inverse, le calcul de ces indices, l'enregistrement des résultats et de l'ensemble des zones préalablement analysées doit être effectué sur le serveur. On y ajoutera aussi des scripts logiciels permettant la mise-à-jour automatique du serveur et la destruction de résultats trop anciens.

3.1.2 Modèle de la base de données

Il est important de considérer que des mêmes zones géographiques peuvent être analysées par plusieurs utilisateurs. Ainsi l'historique des analyses, avec leurs paramètres, va être stocké dans une base de données pour permettre la visualisation des résultats à ceux qui pourraient à nouveaux analyser ces zones. Une utilisation intensive et inutile du serveur est ainsi évitée. De plus, avant de lancer les calculs pour une région donnée, une vérification géographique est faite en comparant les coordonnées géographiques de cette région avec celles enregistrées dans la base de données. Si l'utilisateur n'a pas vérifié que les résultats étaient déjà disponibles il sera alors averti avant le lancement de son analyse.

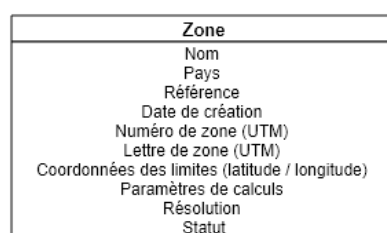


FIGURE 3.2 – Schéma du modèle de la table associée à une zone

Dans la figure précédente, on observe que l'ensemble des données stockées permettent d'identifier individuellement chaque zone, et de les projeter géographiquement selon différentes références. Pour cette projection, nous utilisons ici deux types de systèmes de référence de coordonnées, UTM et WGS84 (Latitude/longitude). UTM est le système le plus précis des deux puisqu'il divise la planète en plusieurs zones (référéncées par une lettre et un nombre). A l'intérieur d'une zone, on peut précisément localiser un point à partir de coordonnées "Easting/Northing", qui sont des valeurs métriques. Pour le système WGS84, un point est localisé à partir d'angles (la latitude 0 étant l'équateur, et la longitude 0 le méridien de Greenwich). Les angles sont représentés par des valeurs décimales, il est nécessaire d'utiliser une formule mathématique (formule d'Haversine) afin de convertir une distance métrique en angles décimaux. Les deux projections sont nécessaires pour nos travaux, d'une part afin de simplifier les calculs des indices en utilisant la projection UTM avec des valeurs métriques, et d'autre part pour projeter sur la carte de visualisation des résultats en associant des latitudes et longitudes (système de projection le plus utilisé).

Pour une utilisation en batch, des données simples liées à la soumission des fichiers seront demandées, un nom servant de référence pour que l'utilisateur puisse suivre l'avancée des calculs, ainsi qu'un email pour l'avertir lorsqu'ils seront terminés. Deux dates clés sont aussi enregistrées : la date de soumission et la date d'effacement. Un mois après la soumission d'une analyse, l'ensemble des fichiers et résultats liés à cette utilisation en batch sont supprimés pour éviter la saturation de la mémoire du serveur.

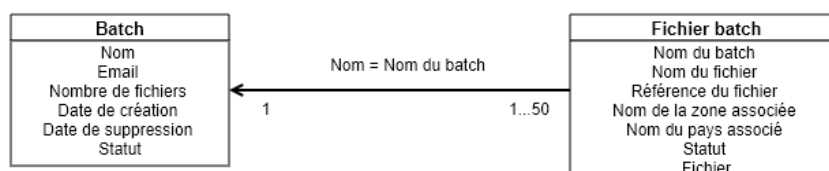


FIGURE 3.3 – Schéma du modèle des tables associée à une soumission batch

L'utilisateur peut soumettre de 1 à 50 fichiers à analyser. Ils sont tous enregistrés en utilisant le nom indiqué lors de la soumission pour les identifier individuellement.

3.1.3 Processus détaillé du fonctionnement de la plateforme

Le processus d'analyse des zones et du calcul des indices est résumé dans la figure suivante.

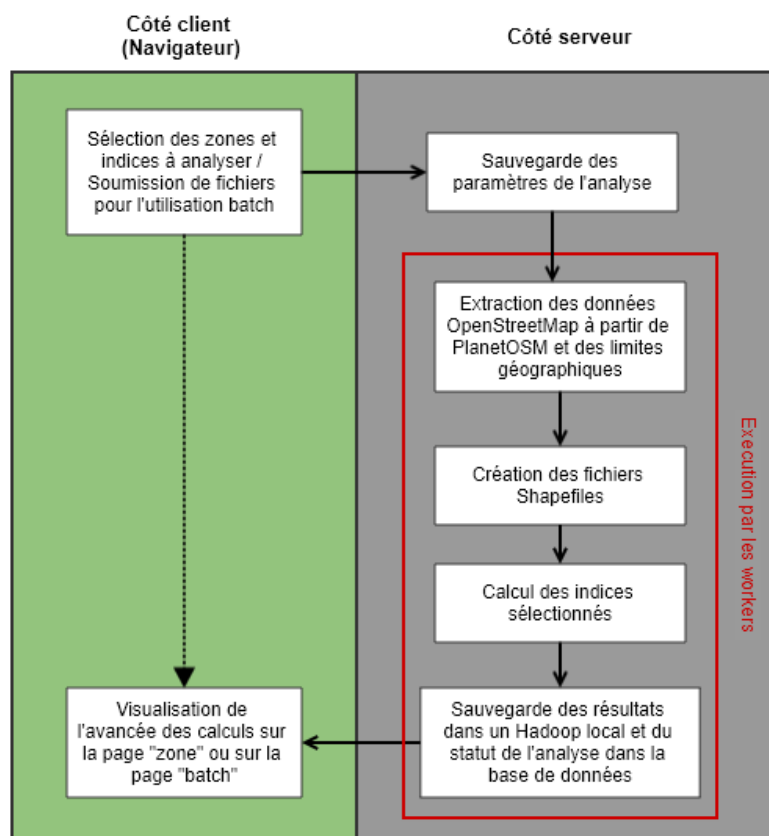


FIGURE 3.4 – Schéma du processus de fonctionnement d'USAT

L'étape "Création des fichiers Shapefiles" est essentielle puisque les indices peuvent être calculés essentiellement à partir de ce format. Le format Shapefile est mondialement utilisé dans les Systèmes d'Informations Géographiques, il réutilise les principes du format vectoriel en organisant les données sous forme de points, lignes et polygones. Dans notre utilisation, les données sont par exemple un bâtiment représenté par un polygone, ou une rue symbolisée par une ligne. Le bloc rouge de la figure précédente indique que les étapes situées à l'intérieur seront réalisées par des workers du CC-IN2P3. Les workers sont des machines distantes utilisées uniquement pour effectuer des calculs pouvant être long et consommant beaucoup de ressources. Cette étape est cruciale puisqu'elle évite d'encombrer le serveur par de lourds calculs. Dans le cas contraire, il serait quasiment impossible d'utiliser le site web puisque les pages seraient servies aux utilisateurs dans un délai très long si beaucoup de calculs y sont effectués. Les workers peuvent accéder au même système de fichiers que le serveur, afin d'y enregistrer convenablement les résultats et de mettre à jour le statut des analyses.

Au-delà de la visualisation du statut des analyses, le téléchargement des résultats est une fonctionnalité du site permettant aux utilisateurs, que ce soit via la liste des résultats disponibles ou en batch, de récupérer les données qu'ils pourront ensuite réutiliser dans leurs études ou recherches personnelles. Pour les résultats des analyses batch, un script exécuté en tâche de fond vérifie si l'ensemble des fichiers a été traité. Lorsque tous les fichiers ont été traités, alors une archive contenant les résultats est créée et est disponible au téléchargement pour l'utilisateur ayant soumis

la demande d'analyse.

Pour l'affichage des résultats sur le navigateur, la récupération des données depuis le serveur doit se faire rapidement, et l'utilisateur doit pouvoir naviguer géographiquement de manière fluide. Ces résultats sont affichés en surcouche d'une carte, et sont échelonnés avec une palette de couleur dépendante de la résolution choisie par l'utilisateur. Pour cela un algorithme de création de polygones associés à une couleur et à une région a été conçu. Des explications complémentaires sur l'algorithme se situent en annexe B.

3.2 Organisation de la plateforme

Pour rappel, l'utilisateur doit pouvoir lancer un calcul (en batch ou graphiquement), visualiser les résultats, et voir la liste des zones préalablement calculées. Ces conditions permettent d'organiser le site en trois pages :

- La page principale d'où l'utilisateur soumet des analyses ;
- La page listant les résultats disponibles ;
- La page permettant de visualiser les résultats.

De plus, nous ajoutons à cette liste une quatrième page spécifique à une utilisation en batch pour un utilisateur ayant soumis des fichiers personnalisés. Cette page permettra uniquement de voir le statut de l'analyse, et de télécharger les résultats lorsqu'ils seront disponibles.

Le plan du site est organisé de la manière suivante :

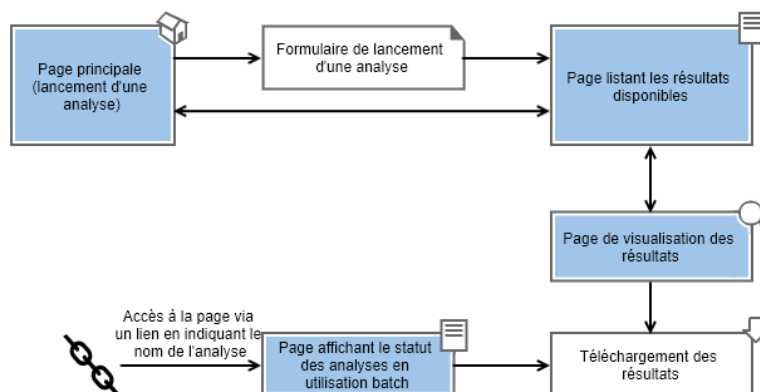


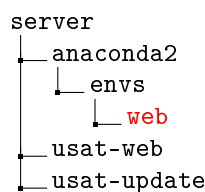
FIGURE 3.5 – Schéma de l'organisation des pages et leur accès

A partir de la page principale, nous pouvons accéder à la liste des zones calculées directement, ou après avoir démarré une analyse (graphique ou en batch). Depuis la page listant les zones, on accède à celle permettant de visualiser les résultats pour une zone. Une fois sur cette page, on peut ainsi télécharger ces résultats à travers une archive dynamiquement créée et contenant les fichiers Shapefiles. Comme expliqué auparavant, les résultats pour une utilisation en batch sont disponibles uniquement au téléchargement via une page accessible par un lien.

3.3 Développement du serveur

Une préoccupation majeure du projet, comme tout projet informatique, est celle du stockage et de la performance. Afin d'utiliser le minimum de ressources, la machine virtuelle qui a été développée est basée sur le système d'exploitation Ubuntu Server 16.04 LTS (acronyme de "Long Term Support"). Cette version minimale d'Ubuntu, utilisable uniquement en ligne de commande contient toute les ressources nécessaires pour déployer la plateforme. Les bibliothèques Python exploitées par les codes développées sont quant à elles installées à travers un environnement virtuel, en d'autres termes elles ne sont pas installées directement sur la machine virtuelle mais dans un dossier créé sur l'espace de fichiers de l'utilisateur. Le logiciel Anaconda permet de gérer la création d'un environnement virtuel et facilite l'installation et la gestion de différentes bibliothèques.

Le schéma suivant indique les principaux dossiers utilisés dans le serveur. L'environnement "web" contient les bibliothèques installées, le dossier "usat-web" correspond à la plateforme développée, et le dossier "usat-update" contient les scripts de mise-à-jour automatique du serveur.



Cette mise-à-jour hebdomadaire, chaque dimanche à 21 heures, est essentielle pour s'assurer que la dernière version des codes développés est utilisée. Le processus est le suivant : nous mettons à jour le fichier PlanetOSM s'il n'est pas en cours d'utilisation, puis les codes sont automatiquement récupérés depuis un serveur GitLab. Dans un premier temps nous récupérons les scripts de calculs des indices depuis le premier dépôt¹, puis ceux correspondant à la plateforme développée². Si les codes sont en cours d'utilisation, cette mise-à-jour elle alors repoussée à la semaine suivante.

3.4 Organisation des tests

Pour tout nouveau développement d'une fonctionnalité ou modification de celles déjà existantes, une organisation méthodique des tests a été respectée afin de vérifier leur bon fonctionnement et s'assurer de l'obtention de résultats satisfaisants. Les tests ont été effectués suivant la deuxième branche du modèle "cycle en V", organisation classique pour des projet d'ingénierie électronique ou informatique, avec ce schéma de déroulement :

1. Tests unitaires ;
2. Tests d'intégration ;
3. Tests de validation.

Un test unitaire se déroule sur une fonction ou un algorithme développé, en vérifiant que la sortie obtenue est celle souhaitée. Pour cela plusieurs jeux de données sont utilisés pour tester plusieurs situations. En pratique, cela peut être simplement une fonction qui retourne une distance en kilomètres à partir de deux points géographiques.

Un test d'intégration nous permet de nous assurer que la fonction développée s'intègre dans une séquence complète d'algorithmes, et d'appels successifs de fonctions. Dans notre cas, nous

1. <https://gitlab.inria.fr/gervason/urbansprawl>

2. <https://gitlab.inria.fr/lrezakha/usat-web>

pouvons tester la récupération des surcouches correspondant aux résultats d'indices, à partir d'une référence d'une zone. Dans cette séquence plusieurs fonctions sont exploitées, nous vérifions ici que l'ensemble est en état de marche.

Enfin, les tests de validation sont les dernières étapes avant de finaliser la plateforme. Nous nous mettons dans la position d'un utilisateur externe visitant le site, et souhaitant effectuer différentes opérations. Il est essentiel que son expérience soit fluide et positive, et que la compréhension de la plateforme soit complète, pour augmenter les chances que le projet soit utilisé.

3.5 Résultats obtenus

3.5.1 Lancement des analyses

Nous présentons ici les pages du site développé, et montrons les résultats de certains indices calculés pour différentes zones. Plusieurs images du site se trouvent en annexe A afin de visualiser l'ensemble des fonctionnalités graphiques du projet. Les deux images de la figure suivante montrent la page principale à gauche (centrée sur la ville de Grenoble lors du chargement complet), et la même page lorsqu'on sélectionne une zone à analyser.

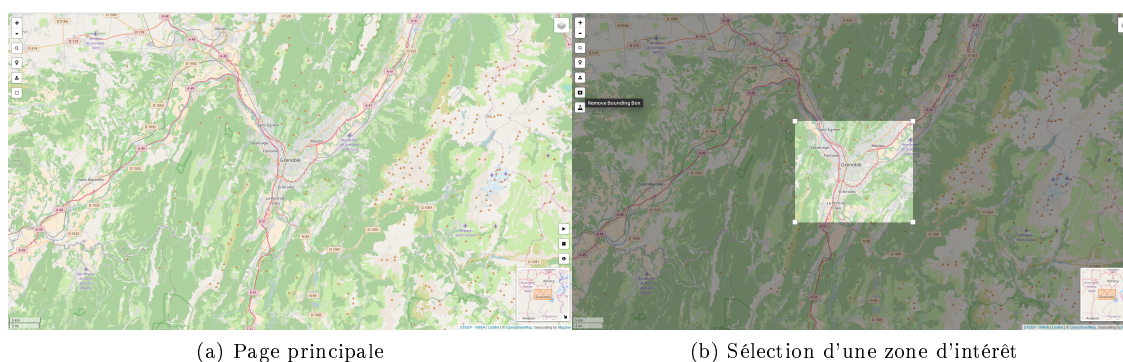


FIGURE 3.6 – Captures d'écran de la page principale

Pour ce lancement des calculs, un menu est disponible pour le choix des indices, la sélection de la méthode de calcul souhaitée et la modification des paramètres. Une fois les indices sélectionnés, les calculs sont lancés après la validation finale de l'utilisateur.

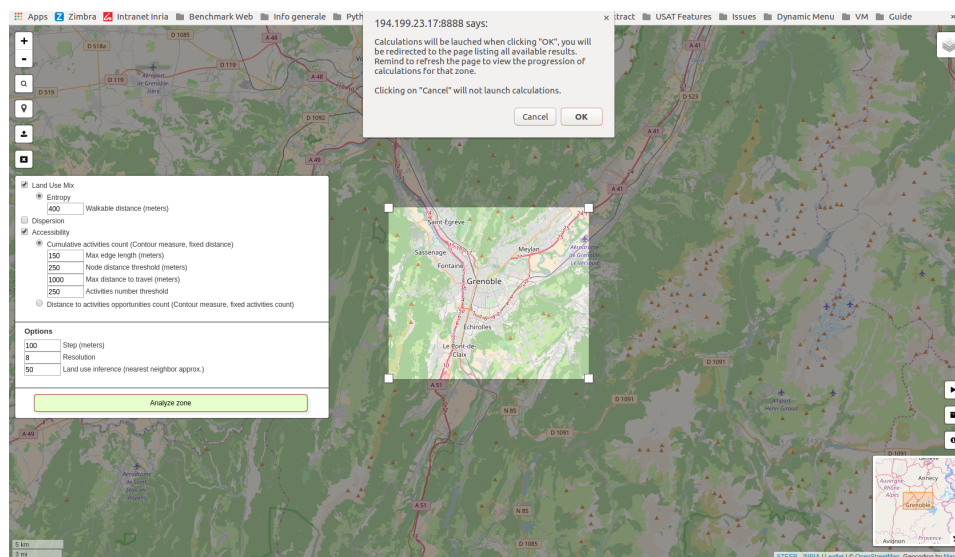


FIGURE 3.7 – Validation de l'analyse

3.5.2 Résultats des calculs des indices géographiques

Dans cette section, nous présentons deux résultats : la mixité de l'usage des sols de Grenoble et l'accessibilité pour Lyon. Nous allons procéder uniquement à l'observation et l'analyse des résultats, l'interprétation sera effectuée dans le chapitre suivant. Nous ne présentons pas la page listant les zones disponibles, elle sera visible en annexe A.

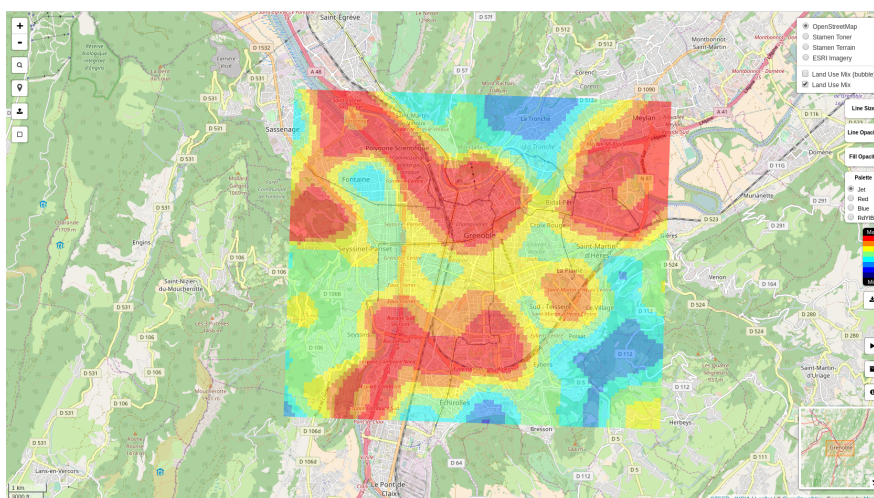


FIGURE 3.8 – Mixité de l'usage des sols pour Grenoble, France

Sur l'image précédente, les résultats sont organisés selon huit intervalles de couleur, représentés par la palette "Jet". Le nombre d'intervalles correspond à la résolution indiquée par l'utilisateur lors du lancement de l'analyse. Nous observons ici plusieurs choses, de grandes zones rouges signifiant une haute mixité d'usage des sols sont superposées à des zones urbaines de la ville de Grenoble, alors que les zones bleues et vertes, montrant une mauvaise mixité, correspondent en général à des zones rurales ou montagneuses. Certaines tâches bleues et vertes se retrouvent toutefois à l'intérieur de la ville. La zone est un peu "penchée", dû à la reprojection des coordonnées géographiques UTM vers des coordonnées Latitude/longitude.

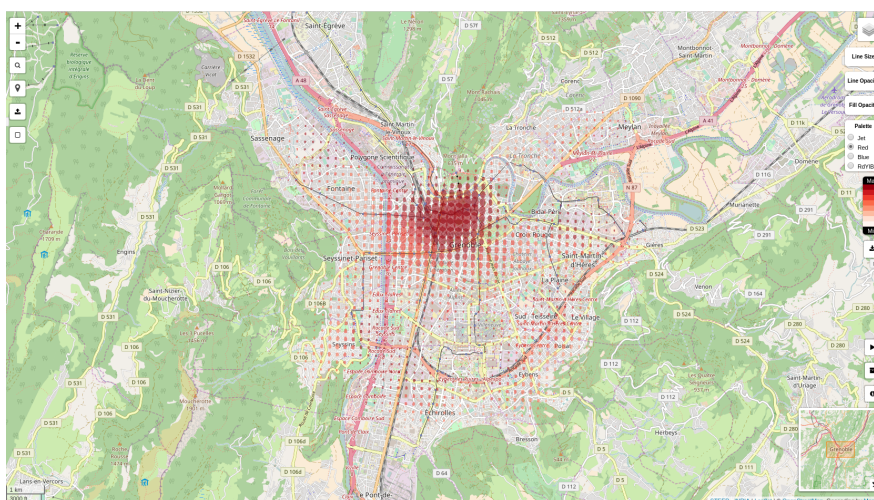


FIGURE 3.9 – Mixité de l'usage des sols en affichage "bubble plot" pour Grenoble, France

Le "bubble plot" nous permet de visualiser à la fois la valeur de la mixité d'usage des sols, ainsi

que le nombre de données disponibles pour un point. Cette affichage est cette fois-ci représenté sous la palette "Red", en nuances de rouge. On observe ici une haute mixité pour le centre-ville de Grenoble, ainsi que de larges bulles, montrant qu'un nombre important de données est disponible pour cette zone. A l'inverse, certaines régions ne sont pas représentées par des bulles puisque le nombre de données est insignifiant. Les bulles visibles sur la carte redessinent la forme urbaine de la ville et de sa périphérie.

La figure suivante affiche les résultats des calculs de l'accessibilité pour la ville de Lyon et l'ensemble de la métropole. La palette utilisée va du bleu au rouge, et l'opacité de la surcouche a été baissée.

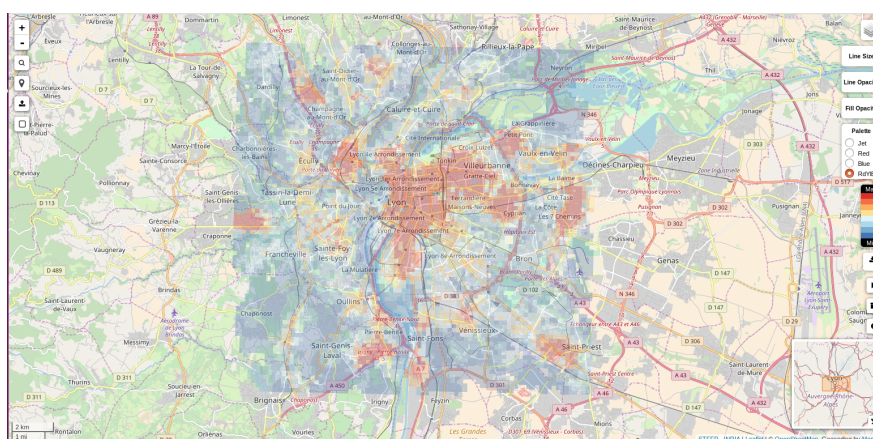


FIGURE 3.10 – Accessibilité pour Lyon, France

Nous observons sur l'image que les zones correspondant aux nuances de rouge, soit une valeur haute de l'indice d'accessibilité, se situent globalement dans les zones les plus urbanisées de la métropole : l'ensemble de la ville de Lyon et de Villeurbanne, le centre-ville des villes périphériques, et les abords des grandes routes. On peut également voir que les parcs ne sont pas superposés de couleurs, indiquant que l'indice n'a pas été calculé dû à l'absence de données.

Chapitre 4

Analyse des travaux

4.1 Analyse des résultats

En reprenant les résultats affichés dans le chapitre précédent, et en utilisant les observations effectuées, les valeurs des indices sont validées si l'on compare avec les données disponibles, et la connaissance que nous avons des villes de Grenoble et Lyon.

En reprenant méthodiquement l'exemple de Grenoble, nous visualisons des valeurs hautes de mixité urbaine pour des régions correspondant soit au centre-ville, soit des lieux où sont localisées à la fois des zones d'activités et des zones résidentiels, validant le modèle établi. Toutefois, cette méthode de visualisation supposerait que la mixité du centre-ville de Grenoble, où la présence d'activités et le zones résidentiels est très élevée, est à mettre au même niveau que celle d'une ville de sa proche banlieue. En analysant cette fois-ci les zones de faible mixité, par exemple la zone bleue située sur la ville de Fontaine, et celle située au sud-est de la carte, nous pouvons imaginer que ces zones ne sont pas comparables car la seconde correspond à une région de montagne, et non urbaine. En somme, cette méthode est très utile puisqu'elle permet à l'utilisateur d'appréhender les résultats de façon globale, mais il est compliqué de comparer l'ensemble des quartiers entre eux.

En utilisant la visualisation en bulles, indiquant le nombre de données disponibles par point d'analyse, il est cette fois-ci possible de comparer les quartiers. Logiquement, le centre-ville est représenté par de larges bulles d'une valeur élevée de mixité. Imaginons le scénario d'une zone également représentée par de larges bulles mais indiquant une très faible mixité, cela supposerait des résultats fiables, et montrant une forte présence d'activités et peu d'usages résidentiels, ou l'inverse. En prenant l'exemple de la ville de Fontaine, la couleur des bulles est assez claire, et leur largeur indique une bonne présence de données. Cette ville est globalement composée d'usages résidentiels, que ce soit des immeubles ou des pavillons, et peu d'activités. Cette organisation de ville est classique dans un schéma péri-urbain. Si l'on regarde les origines de l'expansion de la ville de Grenoble à partir de la moitié du XXème siècle, de nombreux plans d'urbanisme ont été conçus au moment de la préparation des Jeux Olympiques de 1968, et l'agrandissement des villes de la proche banlieue a été initié. Ce modèle valide ainsi un schéma d'étalement urbain qui s'est produit par le passé.

Enfin, en analysant les résultats de l'indice d'accessibilité pour la ville de Lyon, on retrouve classiquement une large zone rouge au centre-ville, puisque dans le schéma de cette ville, de nombreuses activités y sont regroupées. Les autres tâches rouges visibles correspondent généralement à des zones d'activité ou zones industrielles (par exemple la zone industrielle de la Mouche à Irigny, au sud de la carte). Pour ces zones il est logique que pour chaque point d'analyse on retrouve

de nombreuses activités dans un rayon faible, toutes activités confondues. Le schéma montré par ces résultats est par ailleurs un schéma urbain classique que l'on retrouve dans de nombreuses villes françaises voire européennes : un centre très dense et des zones d'activités ou industrielles éparpillées en périphérie[2]. De plus, il est tout à fait probable que dans les zones périphériques correspondant aux nuances de bleu, l'usage de transports personnels est plus important au détriment de moyens de transport non polluants.

4.2 Difficultés rencontrées

La réalisation de ce projet a été séparée en plusieurs grandes sections, différentes en temps de compréhension et de préparation, et en temps de développement. Ce projet était théoriquement destiné à des personnes ayant une formation en systèmes d'informations, et idéalement en systèmes géographiques et en urbanisme. La principale difficulté du projet réside dans la compréhension des enjeux et dans la mise en situation en tant que planificateur urbain. Ainsi, comprendre les indices développés, et la façon dont ils sont formés a été la principale difficulté du projet puisqu'elle nécessitait une lecture importante de la littérature. La personne ayant développée ces indices, M. Gervasoni, m'a apporté une aide importante dans la compréhension à la fois globale et également détaillée de la formation des indices, et de la manière dont ils sont exploitables par des acteurs ayant un pouvoir de décision. Toutefois, il a été compliqué de s'impliquer dans l'élaboration de nouveaux indices, ou de nouvelles méthodes de calculs, par le fait qu'il est très difficile d'avoir une connaissance importante de la littérature en une période de quatre mois.

De plus, en ce qui concerne le développement de la plateforme, certaines étapes devant être réalisées méthodiquement et permettant d'obtenir des résultats efficacement furent techniquement difficiles. Pour certains algorithmes nécessitant un temps d'élaboration important, il a été possible de le surmonter en partageant ces difficultés avec mes encadrants. Grâce à leur expertise et leur expérience, ils ont pu m'indiquer plusieurs voies d'exploration me permettant d'avancer plus rapidement vers le résultat désiré.

4.3 Perspectives

La plateforme n'étant pas encore accessible sur le Web, la première perspective est de la déposer dans le centre CC-IN2P3 du CNRS. Le serveur étant complètement configuré et prêt au déploiement, cette étape sera réalisée très prochainement. Elle n'a pas pu être réalisée dû à l'impossibilité d'accéder aux serveurs du centre durant le mois d'août.

Une autre perspective principale est l'ajout de nouveaux indices géographiques permettant d'analyser d'autres aspects de l'étalement urbain pour une zone donnée. Plusieurs indices étant disponibles dans la littérature, et le développement de la plateforme étant pensé pour l'ajout rapide de nouveaux indices, cette étape peut être réalisée rapidement. De plus, la possibilité d'ajouter une validation terrain des indices par des acteurs locaux via la plateforme renforcerait le projet en lui donnant plus de références.

Enfin, une étape importante pour que le projet soit utilisé, la communication autour de la plateforme doit être développée. Cela se fait au travers de conférences, par exemple le projet est candidat à la participation au colloque international SAGEO 2017¹ qui aura lieu en novembre prochain, ou via la communication vers des acteurs décisionnaires, tel que des urbanistes ou des acteurs politiques.

1. <https://sites.google.com/site/sageodemo/home>

Conclusion

La réalisation d'une plateforme d'analyse d'étalement urbain va dans le sens du développement durable et soutenable des villes du futurs. Elle permettra aux acteurs du monde entier de voir la structure urbaine des villes où une rapide expansion urbaine s'est potentiellement déroulée. A partir de cette compréhension, des discussions et réflexions doivent être engagées par rapport à la réorganisation d'une ville et la planification urbaine des structures résidentielles et des zones d'activité. Il est important de considérer que la plateforme est utile en tant que support, elle doit être combinée avec d'autres informations essentielles aux planificateurs urbains et acteurs politiques, telles que l'organisation des réseaux de distribution d'énergie ou la fabrication de nouveaux réseaux de transport. Cela s'expliquant par la singularité de l'organisation de chaque ville, et l'ensemble des paramètres qu'il faut prendre en compte.

Plus largement, la question de fournir de nouveaux outils technologiques et scientifiques permettant la réorganisation des villes et des espaces urbains est très largement posée dans le monde industriel et celui de la recherche. En la traitant, l'équipe STEEP se démarque de l'approche traditionnelle de l'INRIA, mais fournit une expertise ajoutant une nouvelle vision dans le domaine de l'urbanisme et la planification urbaine.

Bibliographie

- [1] Ryan Brown. Django vs flask vs pyramid : Choosing a python web framework. <https://www.airpair.com/python/posts/django-flask-pyramid>.
- [2] Christel Alvergne et William J. Coffey. "similitudes et différences des formes urbaines en europe et en Amérique du nord. *Cahiers de géographie du Québec*, 2000.
- [3] Reid Ewin, Rolf Pendall, and Don Chen. Measuring sprawl and its transportation impacts. *Transportation Research Record : Journal of the Transportation Research Board*, 2003.
- [4] Jochen AG Jaeger and Christian Schwick. Improving the measurement of urban sprawl : Weighted urban proliferation (wup) and its application to switzerland. *Ecological Indicators*, 2014.
- [5] Gervasoni Luciano, Bosch Marti, Sturm Peter, and Fenet Serge. Calculating spatial urban sprawl indices using open data. *The 15th International Conference on Computers in Urban Planning and Urban Management, Adelaide, Australia*, July 2017.
- [6] United Nations. World urbanization prospects. <https://esa.un.org/unpd/wup/publications/files/wup2014-highlights.Pdf>, 2014.
- [7] Paul M Torrens. A toolkit for measuring sprawl. *Applied Spatial Analysis and Policy*, 2008.
- [8] Song Yan, Merlin Louis, and Rodriguez Daniel. Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*, 2013.

Urban Sprawl Analysis Toolkit

Precalculated zones

You will find here all the history of zones on which users have chosen to perform analysis on. For each zone, its location is shown on the map, and you can see its properties by clicking on the name.

You can perform your custom analysis by choosing a zone on our [Urban Sprawl Analysis Toolkit \(USAT\)](#). All the process needs time to be done (one hour or more depending on the zone size and data), the calculation progress for each zone is written next to its reference.

List of all zones by name and country [Refresh](#)

- Canale, Italia
- Bayern, Deutschland
- Grenoble, France
- Les Sables-d'Olonne, France
- Stoke-on-Trent, UK
- Bern, Switzerland
- Guangxi, China
- Mays Chapel, United States of America
- Community of Madrid, Spain
- Central Serbia, Serbia
- City of Zagreb, Croatia
- Arizona, United States of America

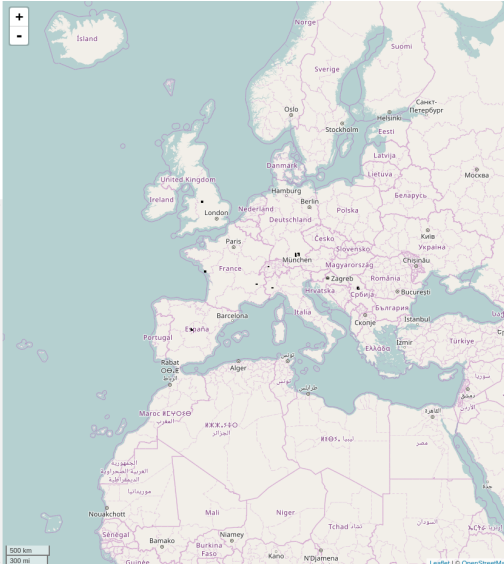


FIGURE A.3 – Page listant les résultats publics disponibles

Urban Sprawl Analysis Toolkit

Precalculated zones

You will find here all the history of zones on which users have chosen to perform analysis on. For each zone, its location is shown on the map, and you can see its properties by clicking on the name.

You can perform your custom analysis by choosing a zone on our [Urban Sprawl Analysis Toolkit \(USAT\)](#). All the process needs time to be done (one hour or more depending on the zone size and data), the calculation progress for each zone is written next to its reference.

List of all zones by name and country [Refresh](#)

- Canale, Italia
- Bayern, Deutschland
- Grenoble, France

Land Use Mix calculated: Yes
 Used Method: Entropy
 Available distance (meters): 400

Dispersion calculated: No

Accessibility calculated: No

Options:
 Steps: 100
 Land use inference (nearest neighbor approx.): 50
 Resolution: 8

Created at: July 18, 2017, 10:52 a.m.
[View this zone on USAT](#)

- Les Sables-d'Olonne, France
- Stoke-on-Trent, UK
- Bern, Switzerland
- Guangxi, China
- Mays Chapel, United States of America
- Community of Madrid, Spain
- Central Serbia, Serbia
- City of Zagreb, Croatia
- Arizona, United States of America




FIGURE A.4 – Affichage des paramètres d'une zone lors d'un clic sur son nom

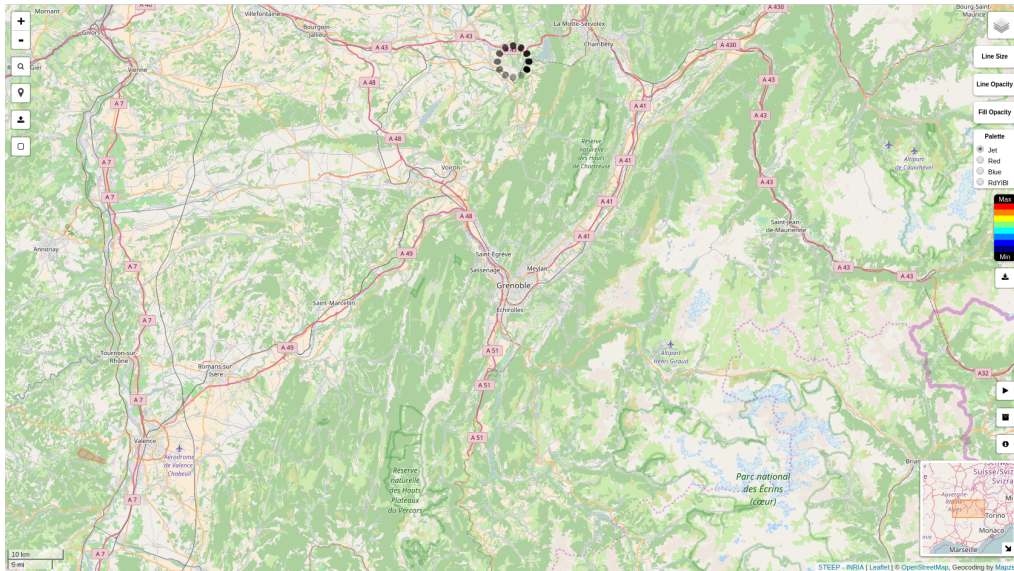


FIGURE A.5 – Page affichant les résultats d'une zone - un "spinner" s'affiche pendant le chargement des résultats

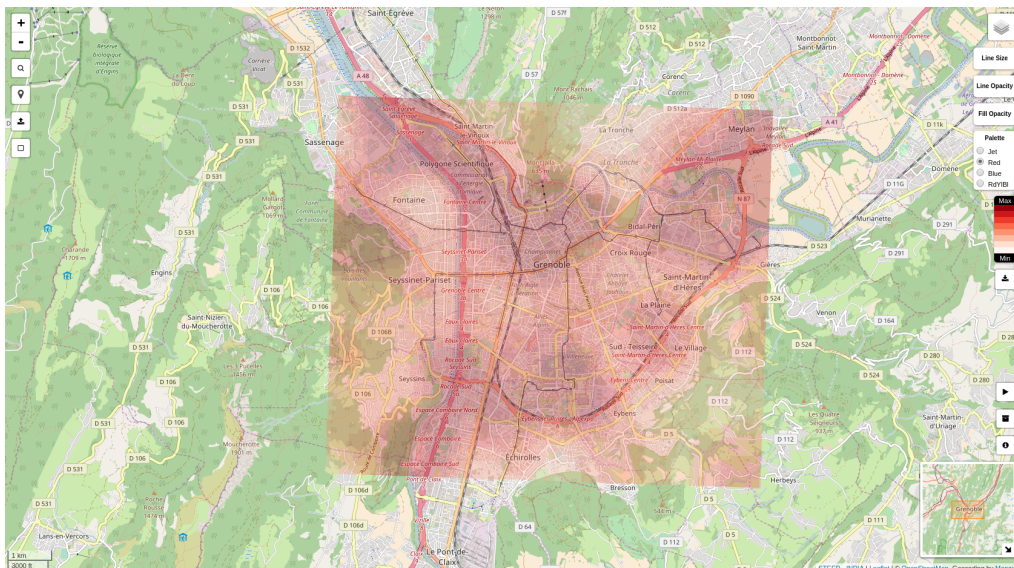


FIGURE A.6 – Affichage des résultats de mixité d'usage des sols pour Grenoble avec une palette en nuances de rouge

Annexe B

Algorithme de représentation géographique des résultats

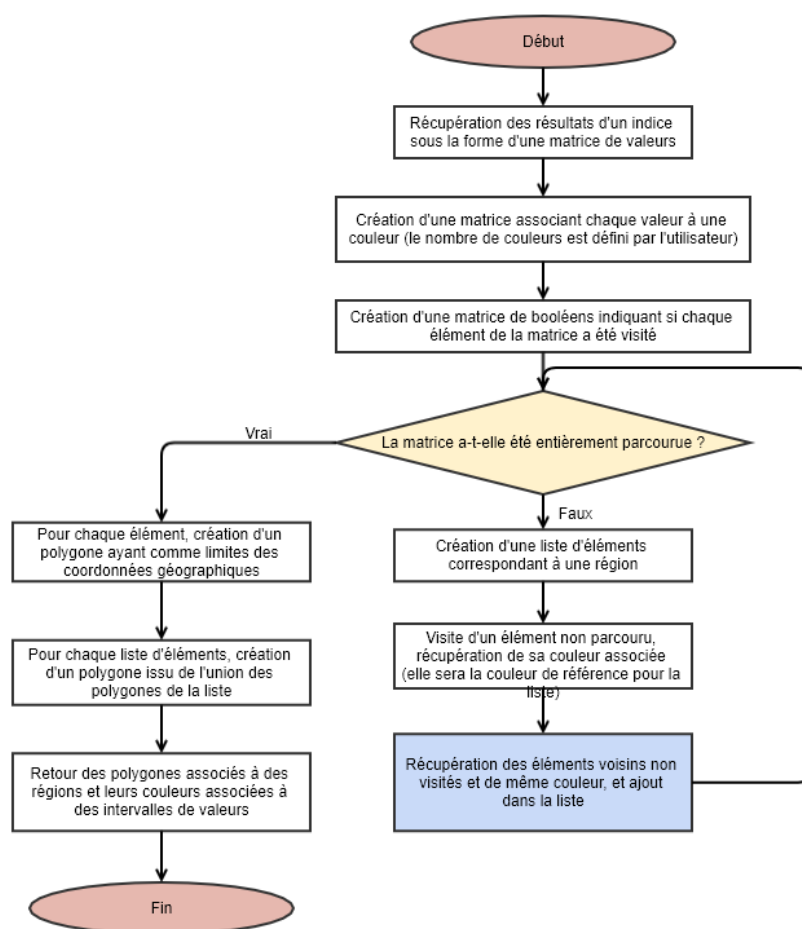


FIGURE B.1 – Organigramme de l’algorithme permettant d’associer les résultats d’un indice pour une zone à des régions géographiques

Le bloc ayant un fond bleu est réalisé grâce à un algorithme récursif souvent utilisé pour le parcours d'une matrice, nommé en français "Algorithme de remplissage par diffusion" ou "Flood fill" en anglais. Pour chaque voisin d'un élément, s'il n'a pas été visité et qu'il est de la même couleur, on visite alors ses voisins puis on l'ajoute à la liste d'éléments. S'il a été visité, qu'il est d'une couleur différente ou que les coordonnées sont au-delà des limites de la matrice, alors cet élément n'est pas parcouru.