



TYPOLOGIES DE VILLES FRANÇAISES SUR LA BASE DES ZONES CLIMATIQUES LOCALES ET DES DÉTERMINANTS DE LA SURCHAUFFE URBAINE

Mémoire de stage de Besnard Pauline
Master SIGAT – Université Rennes 2
Année universitaire 2023/2024

Sous la direction technique de Mr ALGLAVE Baptiste
Sous la direction académique de Mr DEMORAES Florent

Soutenance le 25 octobre 2024



Remerciements

Je tiens à exprimer ma profonde gratitude à l'ensemble des personnes qui ont contribué à la réussite de ce stage et à l'obtention de mon Master 2 Systèmes d'Information Géographique et Analyse des Territoires (SIGAT).

Je tenais tout d'abord à remercier mon encadrant de stage Baptiste ALGLAVE pour son soutien, sa confiance et l'opportunité qu'il m'a offerte de mener à bien ce projet dans les meilleures conditions.

Je souhaiterais remercier le Lab-STICC de m'avoir accueilli pendant 6 mois. Et je voulais particulièrement exprimer ma reconnaissance à toute l'équipe : Elisabeth LE SAUX WIEDERHOLD, Erwan BOCHER, Matthieu GOUSSEFF, François LECONTE, Jeremy BERNARD, Lise BELLANGER et Jessica PIC, pour leurs conseils et leur bienveillance.

Je remercie également Virginie GUILLET pour la gestion efficace des aspects administratifs, facilitant ainsi le bon déroulement de mon stage.

Par ailleurs, les deux années de formation au sein du master SIGAT m'ont permis d'aborder ce stage avec confiance, en m'apportant des bases solides dans le domaine des systèmes d'information géographique. Je tiens à remercier tout particulièrement Florent DEMORAES pour son suivi pédagogique attentif, ainsi que Boris MERICKSKAY, François LEPRINCE, Erwann QUESSEVEUR ainsi que les autres intervenants, qui se sont investis pleinement pour offrir une formation de qualité.

Je souhaiterais ajouter des remerciements à mes camarades de classe qui ont rendu ces deux années instructives et conviviales.

Je vous remercie tous d'avoir apporté votre gentillesse et votre savoir pour améliorer mes compétences.

Glossaire

AFDM : Analyse Factorielle des Données Mixtes

CNRS : Centre National de Recherche Scientifique

DMID : Données, Modèle, Informations et Décisions

ICU : îlots de Chaleur Urbains

Lab-STICC : Laboratoire des Sciences et Techniques de l'Information, de la Communication et de la Connaissance

LERMAB : Laboratoire d'Études et de Recherche sur le Matériaux Bois

MApUCE : Modélisation Appliquée au droit de l'Urbanisme : Climat urbain et Énergie

MNT : Modèle Numérique de Terrain

OSM : OpenStreetMap

PÆNDORA : Planification, Adaptation et Énergie : Données territoriales et accompagnement

RSU : Unités Spatiales de Référence

SRTM : Shuttle Radar Topography Mission

SIGAT : Systèmes d'Information Géographique et Analyse des Territoires

TEB : Town Energy Balance

TRI : Indice de Rugosité Topographique

WUDAPT : World Urban Database and Access Portal Tools

ZCL : Zones Climatiques Locales

Table des matières

| | |
|---|-----------|
| Remerciements..... | 2 |
| Glossaire..... | 3 |
| Table des matières..... | 4 |
| Introduction..... | 6 |
| 1. Cahier des charges..... | 7 |
| 1.1 La mission..... | 7 |
| 1.2 Les besoins..... | 7 |
| 1.3 L'analyse des risques..... | 7 |
| 1.4 Planification du travail..... | 8 |
| 2. Présentation de l'équipe de recherche et de l'espace de travail..... | 9 |
| 2.1 La structure..... | 9 |
| 2.2 L'équipe..... | 9 |
| 2.3 L'espace de travail..... | 10 |
| 2.4 Des rencontres professionnelles : séminaire et rencontres R..... | 10 |
| 3. Etat de l'art..... | 11 |
| 3.1 Définition des phénomènes de surchauffe et des îlots de chaleur urbains..... | 11 |
| 3.2 Définition du concept de Zone Climatique Locale (ZCL) et lien avec la surchauffe..... | 12 |
| 3.3 Méthodologie de cartographie des ZCL : Geoclimate / WUDAPT..... | 13 |
| 3.4 Les déterminants du microclimat sur le territoire..... | 14 |
| 4. Matériels et méthodes..... | 17 |
| 4.1 Choix du domaine d'étude..... | 17 |
| 4.2 Construction des indicateurs du phénomène de surchauffe urbaine..... | 18 |
| 4.2.1 Climat et topographie..... | 19 |
| 4.2.1.1 Eloignement à la mer..... | 19 |
| 4.2.1.2 Indice de Rugosité Topographique..... | 19 |
| 4.2.1.3 Climat..... | 19 |
| 4.2.2 Occupation du sol..... | 19 |
| 4.2.2.1 Présence, distribution, et importance de la végétation..... | 19 |
| 4.2.2.2 Perméabilité des sols..... | 20 |
| 4.2.3 Morphologie urbaine..... | 20 |
| 4.2.3.1 Obstacles urbains..... | 20 |
| 4.2.3.2 Densité bâtiments..... | 21 |
| 4.2.3.3 Typologie de la structure urbaine..... | 21 |
| 4.3 Méthode de classification..... | 21 |

| | |
|--|-----------|
| 4.4 Métriques utilisées pour comparer les groupes de villes et les données d'ICU | 22 |
| 4.5 Automatisation du processus..... | 23 |
| 5. Résultats..... | 25 |
| 5.1 Sur l'ensemble des villes françaises..... | 25 |
| 5.1.1 Construction des axes..... | 25 |
| 5.1.2 Regroupement des individus..... | 25 |
| 5.1.3 Typologie des villes pour l'ensemble des villes françaises..... | 26 |
| 5.2 Sur les grandes villes françaises..... | 30 |
| 5.2.1 Construction des axes..... | 30 |
| 5.2.2 Regroupement des individus..... | 31 |
| 5.2.3 Typologie des grandes villes françaises..... | 32 |
| 5.2.4 Comparaison avec les données ICU..... | 34 |
| 6. Discussions..... | 36 |
| 6.1 Synthèse des principaux résultats..... | 36 |
| 6.2 Limites de la méthode et pistes d'amélioration..... | 36 |
| 6.4 Point de vue prospectif..... | 37 |
| Conclusion..... | 38 |
| Bibliographie..... | 39 |
| Tableau des tables et des figures..... | 42 |
| Annexes..... | 44 |
| Fiche Synthèse..... | 68 |

Introduction

La **surchauffe des territoires** est un **enjeu majeur dans le contexte du réchauffement climatique**. À l'échelle mondiale, la température terrestre de la dernière décennie a augmenté de 1,1 degré depuis l'ère industrielle (1850 - 1900, période de référence prise par l'Accord de Paris (GIEC, 2021)). L'augmentation des températures moyennes annuelles en France est rapide: 13,8 °C en 2014, 13,9 °C en 2018, 14,1 °C en 2020 et 14,5 °C en 2022. Sur cette base, l'année 2022 a été la plus chaude jamais enregistrée sur le territoire français métropolitain (Ministère de la Transition Écologique et de la Cohésion des territoires, 2023). Dans ce contexte, les phénomènes de surchauffe urbaine sont de plus en plus intenses et étendus dans le temps et dans l'espace. Les experts du GIEC prévoient une intensification de ces phénomènes au cours du XXI^e siècle (Pachauri, 2014, Cerema, 2018). À titre d'exemple, Météo France prévoit à Rennes 18 jours de vagues de chaleur¹ pour la période 2041-2070 alors que sur la période 1971-2005, 5 jours de vague de chaleur ont été recensés (Haut Conseil du Climat. (2021) / Reghezza-Zitt, (2023)) (cf annexe 1).

Un certain nombre de déterminants favorisent la surchauffe en ville : l'occupation du sol, les matériaux de construction, la morphologie urbaine, la topographie et les activités humaines. Ainsi, sur la base d'indicateurs décrivant les déterminants du climat urbain, **l'objectif de ce stage est de constituer une typologie de villes homogènes au regard de leur microclimat et de déterminer leur sensibilité au phénomène de surchauffe.**

¹ Jour où la température maximum est supérieure à 5 degrés à la normale pendant 5 jours consécutifs

1.Cahier des charges

1.1 La mission

La mission est de **classifier les villes** françaises en se basant sur des indicateurs du microclimat urbain dans l'objectif de **constituer une typologie de villes homogènes** et de **déterminer leur sensibilité au phénomène de surchauffe**. Mes missions consistaient en :

- **L'étude bibliographique** exhaustive des déterminants de la surchauffe, pour déterminer les indicateurs à mettre en place.
- **La construction des indicateurs** du climat urbain pour réaliser la classification.
- La mise en place d'une méthode de **classification non supervisée**, adaptée aux descripteurs (données quantitatives et qualitatives).
- **L'analyse et la validation** des résultats notamment à l'aide de représentations géographiques.
- **Rédaction** d'un rapport et mise à disposition du script automatisé.

1.2 Les besoins

Le besoin à travers la mission de ce stage est de **documenter les territoires**, en se concentrant sur la classification des territoires urbains sur la base d'indicateurs du microclimat urbain. Cette classification doit avoir une interprétation physique dans le sens où certains groupes de villes doivent être plus ou moins reliés à des températures élevées. L'objectif est de mieux comprendre et caractériser ces phénomènes en analysant les données disponibles. Cette analyse fournira des informations utiles pour les collectivités locales pour la mise en place de leur stratégie d'adaptation au réchauffement climatique. Cette analyse enrichit également les **recherches** en cours, notamment celles sur la prédition des ICU et qui fera l'objet d'un sujet de thèse qui sera étudié par Lenaig LE GROGNEC.

1.3 L'analyse des risques

Certains facteurs ont dû être pris en compte pour assurer le succès de ce projet. En premier lieu, bien que la problématique de la surchauffe urbaine soit un sujet d'actualité, mes connaissances sur le sujet n'étaient pas exhaustives. J'ai dû m'approprier la **bibliographie** afin d'avoir des repères et développer une culture propre à ce champ de recherche. De plus, une phase d'**adaptation** de l'environnement de travail était nécessaire. Par exemple, mes collègues utilisent le système d'exploitation Linux et j'ai dû apprendre à l'utiliser. Enfin, j'ai dû manipuler deux outils pour conduire les analyses : l'outil Geoclimate, des codes R pour extraire les données de Geoclimate et pour réaliser la classification. Je n'avais pas utilisé R depuis janvier et j'ai dû me remettre à niveau pour être opérationnelle et mener à

bien ce stage. Ces éléments ont été intégrés dans le **planning** pour **anticiper** le temps de formation et la mise à niveau que cela impliquait.

1.4 Planification du travail

Un planning prévisionnel a été défini pour ce stage. Dans un premier temps, j'ai **contextualisé le stage** en identifiant les besoins de l'unité et en réalisant des lectures préliminaires générales sur le sujet. L'**analyse bibliographique** m'a permis de mieux connaître le champ de recherche lié à l'analyse du climat urbain. Cette étape m'a permis d'identifier les grands déterminants de la surchauffe urbaine et de **sélectionner les indicateurs** les plus pertinents pour réaliser la classification. Une fois les indicateurs choisis, je me suis **familiarisée avec mon environnement de travail** (linux, Geoclimate) et les codes déjà existants et que j'ai adapté, complété et corrigé pour répondre à la problématique de mon stage. En particulier, j'ai ajouté **les indicateurs climatiques** qui n'étaient pas présents à l'origine, en mobilisant d'autres sources de données que Geoclimate quand c'était nécessaire. Une fois l'ensemble du processus finalisé, j'ai réalisé les **analyses** de classification que j'ai **validé** à l'aide de données d'ICU disponibles grâce au projet MApUCE. La **rédaction du rapport** a été réalisée de façon continue sur le stage. Le planning effectif est très proche du planning prévisionnel (figure 1) pour cette raison il a été mis en annexe (cf annexe 2 : *Planning prévisionnel*).

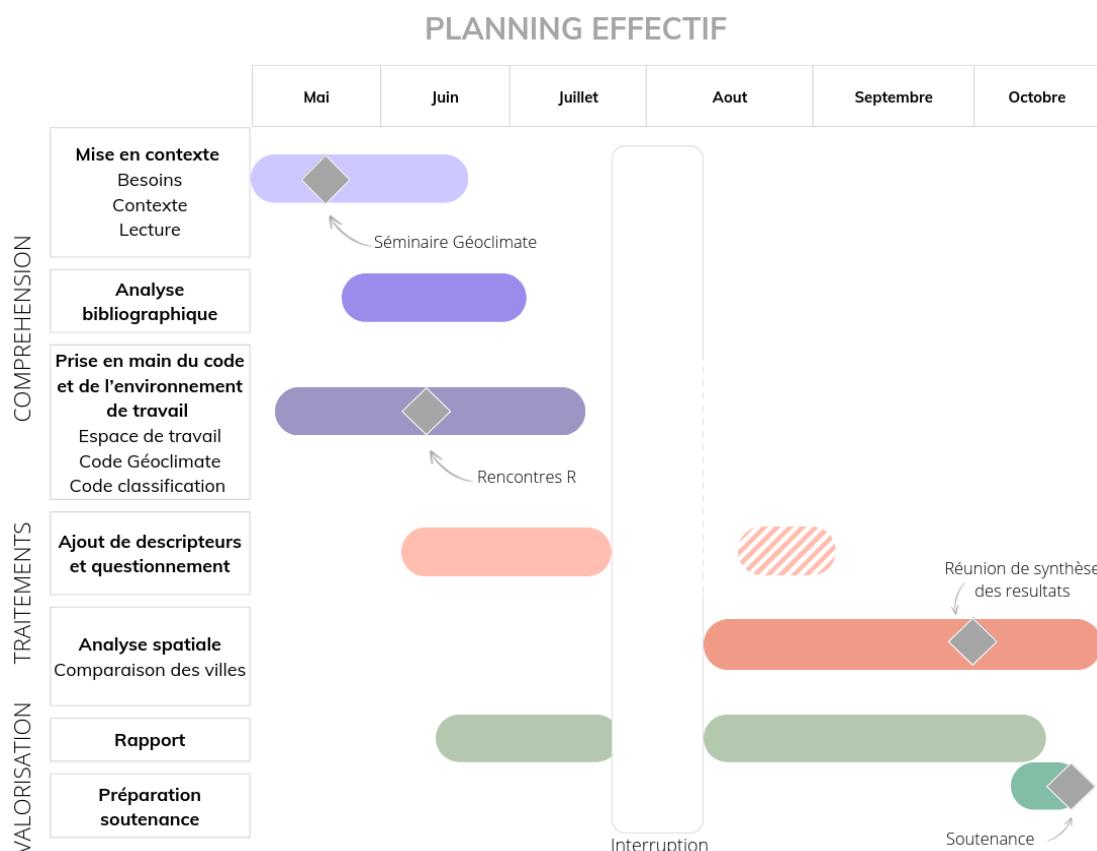


Figure 1 : Planning effectif

2. Présentation de l'équipe de recherche et de l'espace de travail

2.1 La structure

Le Laboratoire des Sciences et Techniques de l'Information, de la Communication et de la Connaissance (Lab-STICC) est un laboratoire de recherche rattaché au Centre National de Recherche Scientifique (CNRS). Ce laboratoire est une unité de recherche spécialisée dans les Sciences et Technologies de l'Information et de la Communication et qui s'étend sur plusieurs champs disciplinaires comme les sciences des données, l'intelligence artificielle, la cyber sécurité, l'analyse du signal, les drones. Il est séparé en 9 pôles de recherche, subdivisés en 26 équipes qui se situent à Brest, Lorient, Quimper et Vannes.

L'équipe DECIDE à laquelle je suis rattachée, fait partie du pôle : Données, Modèle, Informations et Décisions (DMID). Cette équipe mène un travail multidisciplinaire puisqu'elle combine statistique, géographie, analyse spatiale et informatique.

Au sein de l'équipe DECIDE, le groupe de SIG s'intéresse notamment aux phénomènes de surchauffe en ville et au lien entre aménagement du territoire et climat urbain. L'équipe a notamment développé des outils SIG permettant de cartographier des indicateurs du climat urbain comme les Zones Climatiques Locales (ZCL). Ils ont participé au projet [MApUCE](#) (Modélisation Appliquée au droit de l'Urbanisme : Climat urbain et Énergie) et [PÆNDORA](#) (Planification, Adaptation et Énergie : Données territoriales et accompagnement).

Le projet MApUCE (2014-2018) visait à collecter des données sur le microclimat urbain à l'échelle nationale pour aider les acteurs politiques à mieux intégrer les enjeux climatiques à l'aménagement des villes. Sur 42 aires urbaines, une simulation des ICU pour deux configurations climatiques a été réalisée.

Lors du projet PÆNDORA (2017-2020), l'équipe du Lab-STICC a développé l'outil Geoclimate, qui permet d'intégrer un ensemble d'indicateurs géographiques permettant de cartographier les ZCL à l'échelle mondiale. PÆNDORA a également contribué à rendre les résultats de MApUCE opérationnels.

2.2 L'équipe

Lors de mon arrivée, l'équipe du Lab-STICC de Vannes se composait de quatre membres. Pendant mon stage elle s'est agrandie avec l'arrivée de deux enseignantes-chercheuses et d'une doctorante (figure 2).

ORGANIGRAMME ÉQUIPE DECIDE PÔLE SIG

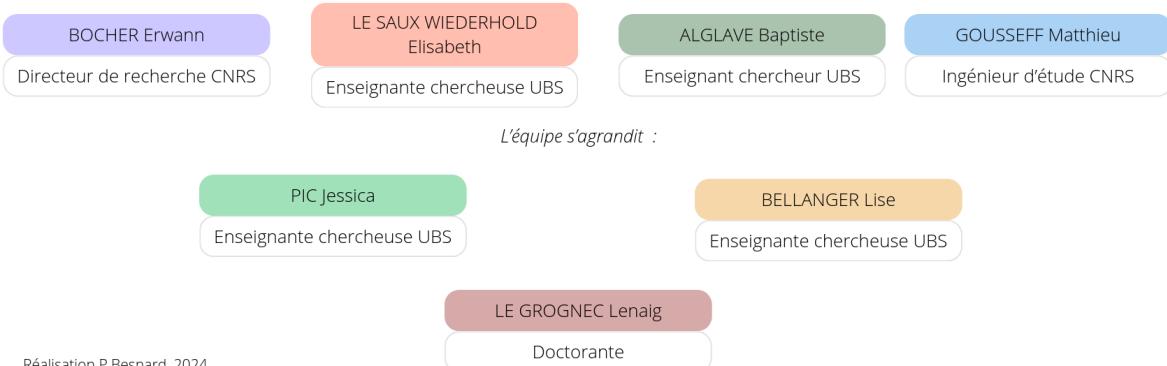


Figure 2 : Organigramme de l'équipe de Vannes

2.3 L'espace de travail

Lors de ce projet j'ai utilisé différents outils (figure 3) :

- GNU Linux comme système d'exploitation,
- IntelliJ / Rstudio comme environnement de développement,
- Obsidian pour la prise de note,
- Github pour le partage et archivage du code,
- Google Drive pour le partage du rapport et des présentations,
- QGIS pour exploration des données spatiales de sorties,
- Signal pour la communication avec l'équipe,
- R pour toute la partie extraction et analyse des données.



Figure 3 : Outils utilisés pendant le stage

2.4 Des rencontres professionnelles : séminaire et rencontres R

Du 21 au 24 mai, un séminaire à eu lieu au sein du Lab-STICC. Lors de cette rencontre l'équipe DECIDE était présente ainsi que Jérémy BERNARD, chercheur associé au Lab-STICC et François LECONTE Maître de conférence au Laboratoire d'Études et de Recherche sur le Matériaux Bois (LERMAB) de l'Université de Lorraine. L'objectif principal de cette rencontre était la construction d'un article scientifique comparant les outils de construction des ZCL.

De plus, lors de mon stage, j'ai assisté aux conférences R. Cette conférence annuelle permet de rassembler des acteurs privés et des universitaires utilisant R autour de sujets comme la reproductibilité, le packaging et l'enseignement avec R. J'ai notamment pu participer à un tutoriel sur le package Torch présenté par Tristan Mary-Huard, chercheur à INRAE. Ce package offre un environnement dédié à l'apprentissage des réseaux de neurones.

3. Etat de l'art

3.1 Définition des phénomènes de surchauffe et des îlots de chaleur urbains

Une grande partie de la littérature liée à la surchauffe urbaine se focalise sur les phénomènes d'ICU (Se Woong Kim, R.D Brown, 2021). Un ICU correspond à un "phénomène d'amplification des fortes chaleurs en ville" (ADEME). Il a été mis en évidence par Luke Howard en 1833 (Oke, 1982). L'ICU peut être défini comme un **dôme thermique** où la température est plus élevée (Cerema, 2018). C'est un phénomène **physique qui impacte principalement les villes la nuit pendant l'été**, la dynamique de l'ICU est variable dans le temps et dans l'espace.

Les ICU concernent essentiellement les espaces urbains, puisque l'ensemble du bâti absorbe l'énergie solaire le jour et libère l'énergie stockée la nuit. Dans les espaces ruraux, cette énergie solaire est consommée par l'évapotranspiration. Ainsi, il est possible d'observer une différence de température entre le milieu urbain et rural d'environ 5 à 10 degrés (Renard et Alonso, 2023) (figure 4).

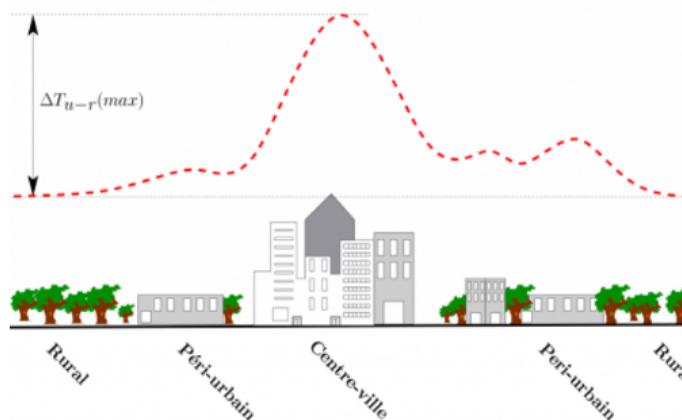


Figure 4 : Schématisation du phénomène d'îlot de chaleur urbain (Cerema)

Avec l'**augmentation de l'intensité** et de la fréquence des vagues de chaleur liée au réchauffement climatique, les ICU sont de plus en plus fréquents et ils recouvrent des périodes de plus en plus longues (Météo France, 2023). Avant 1989, la France enregistrait en moyenne 1,7 jour de vagues de chaleur par an. Depuis 2000, ce nombre a considérablement augmenté pour atteindre environ 8 jours par an. De plus, sur la décennie 2010-2020, il a été recensé 22 vagues de chaleur à travers le territoire français, montrant l'augmentation significative de la fréquence et de l'intensité de ces événements climatiques extrêmes (Météo France, 2023). Ceux-ci ont un impact direct sur la santé publique, le bien-être des habitants, l'attractivité des centres-villes et la consommation d'énergie (utilisation de la climatisation).

L'intensité des ICU dépend d'un grand nombre de **déterminants** tels que l'occupation du sol, la morphologie urbaine, le climat et la topographie de la ville et

enfin les activités humaines. Ceux-ci seront décrits plus en détail par la suite (cf 3.4 *Les déterminants du microclimat sur le territoire*).

3.2 Définition du concept de Zone Climatique Locale (ZCL) et lien avec la surchauffe

Le concept des **zones climatiques locales** (ZCL) a été introduit par les auteurs Stewart et Oke en 2012. C'est une classification qui fait référence dans la communauté scientifique à l'international. L'objectif des ZCL est de découper le territoire en **zones uniformes** suivant l'occupation du sol et la morphologie urbaine. Chaque ZCL correspond à une combinaison de facteurs caractérisant un **comportement climatique** homogène. En d'autres termes, chaque ZCL réagit de manière similaire aux facteurs extérieurs, tels que le rayonnement solaire. Dix-sept classes de ZCL permettent de caractériser le territoire mondial, 10 classes concernent le bâti et 7 concernent les espaces naturels (cf annexe 3 : Description des ZCL par Oke).

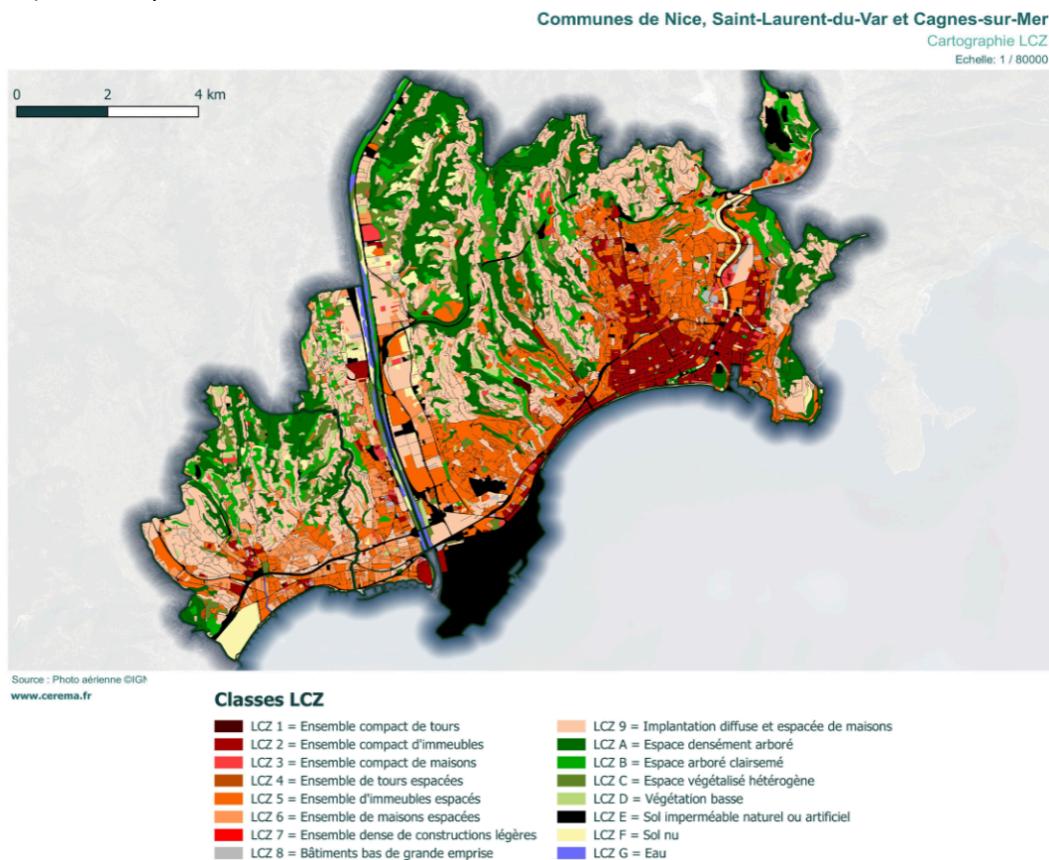


Figure 3 : Exemple de cartographie des ZCL sur la ville de Nice (Cerema)

Les auteurs Delidais et al. (2022) mettent en évidence la relation entre la surchauffe urbaine et le type de ZCL à partir de relevés de température de l'air. Par exemple, les zones à forte concentration de bâtiments bas dédiés au secteur tertiaire, telles que les zones industrielles (ZCL 8), présentent des anomalies de

température plus marquées que les zones végétalisées. Durant la période étudiée, le pic d'ICU observé pour cette ZCL a atteint 6,5 degrés.

3.3 Méthodologie de cartographie des ZCL : Geoclimate / WUDAPT

Plusieurs méthodes existent pour cartographier les ZCL. La méthode World Urban Database and Access Portal Tools (**WUDAPT**), a été formalisée en 2018 par Ching, J et ses co-auteurs. Elle vise à produire des cartes des ZCL d'une zone d'étude à l'échelle mondiale. Cette méthode s'appuie sur une classification supervisée se basant sur des données **satellitaires**. C'est une méthode qui fait référence aujourd'hui (Wang et al., 2024) mais elle présente une résolution spatiale de 100 mètres.

L'outil SIG **Geoclimate** cherche à produire différentes bases de données **vectorielles** dont les ZCL à une échelle locale plus fine (figure 5). Il est codé dans le langage informatique groovy, il est maintenu et une amélioration constante est faite par le laboratoire (Bocher et al., 2021). Cet outil est [open source](#).

Cet outil permet :

- L'**extraction des ZCL à l'échelle mondiale**, via le calcul d'indicateurs géographiques (cf annexe 4 : *Les 14 indicateurs pour le calcul des ZCL, et la description de la méthode de calcul pour chaque indicateur*). La création de données pour les modèles climatiques urbains comme le modèle Town Energy Balance (TEB). TEB calcule les échanges d'énergies et d'eau entre les villes et l'atmosphère.
- Une classification de l'occupation du sol basée sur la méthode Urban Typology by Random Forest décrite par [Tornay et al. \(2017\)](#).
- Des indicateurs pour évaluer l'impact du bruit.

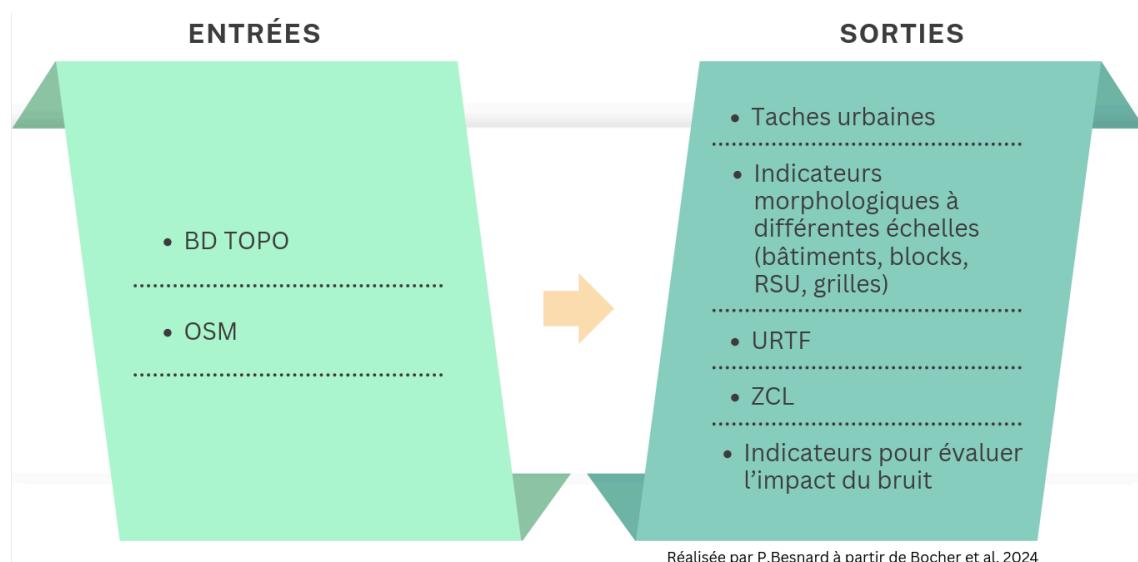


Figure 5 : Entrées et des sorties de Geoclimate

Concernant la construction des ZCL, celles-ci sont calculées à l'échelle de des unités spatiales de référence (RSU). Une RSU correspond à une zone homogène, subdivisée par des éléments structurants du territoire comme les routes, les chemins de fer, ... (cf annexe 5 : Exemple de calcul d'unité spatiale topographique).

Geoclimate s'appuie sur les données vecteur d'OpenStreetMap (OSM) ou de la BD-Topo (Bocher et al., 2024). Le choix entre OSM, disponible sur l'entièreté du globe et BD-Topo, mis à disposition pour la France, s'effectue suivant les besoins. Par exemple, la hauteur des bâtiments est souvent manquante dans la plupart des données OSM (Bernard et al., 2022). Le territoire est plus fragmenté lors de l'utilisation de la BD-Topo par rapport aux données OSM (le nombre de RSU est presque doublé). Un plugin R, *Iczexplore*, a été mis en place par le laboratoire principalement par Matthieu GOUSSEF pour comparer les différentes méthodes de production de ZCL (Bocher et al., 2024). Ainsi, Geoclimate permet, grâce à divers traitements (figure 6), de générer entre autres des cartes des ZCL d'un territoire à l'échelle mondiale.

Chaine de traitement de Geoclimate pour les ZCL

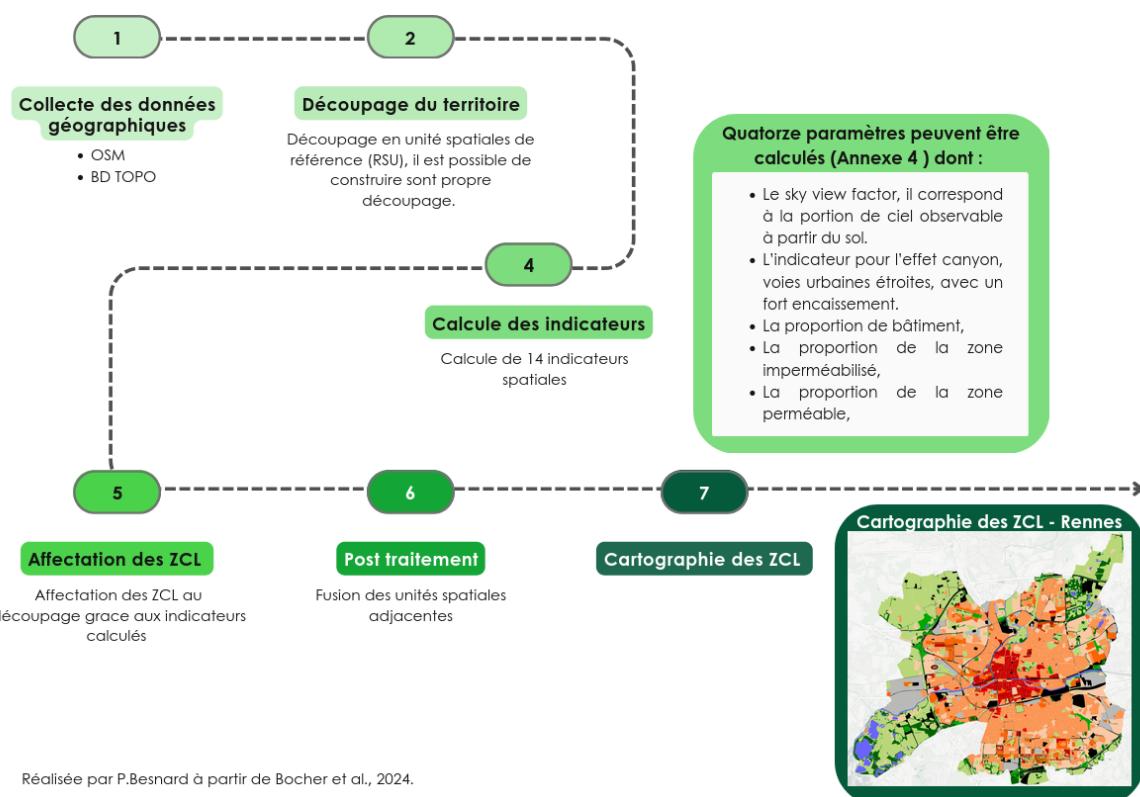


Figure 6 : Chaîne de traitements simplifié de Geoclimate pour la construction des ZCL

3.4 Les déterminants du microclimat sur le territoire

Les phénomènes de surchauffe et d'ICU sont influencés par un certain nombre de déterminants. L'analyse bibliographique a permis de mettre en

évidence onze déterminants de la surchauffe urbaine (figure 7), ceux-ci sont détaillés ci-dessous.

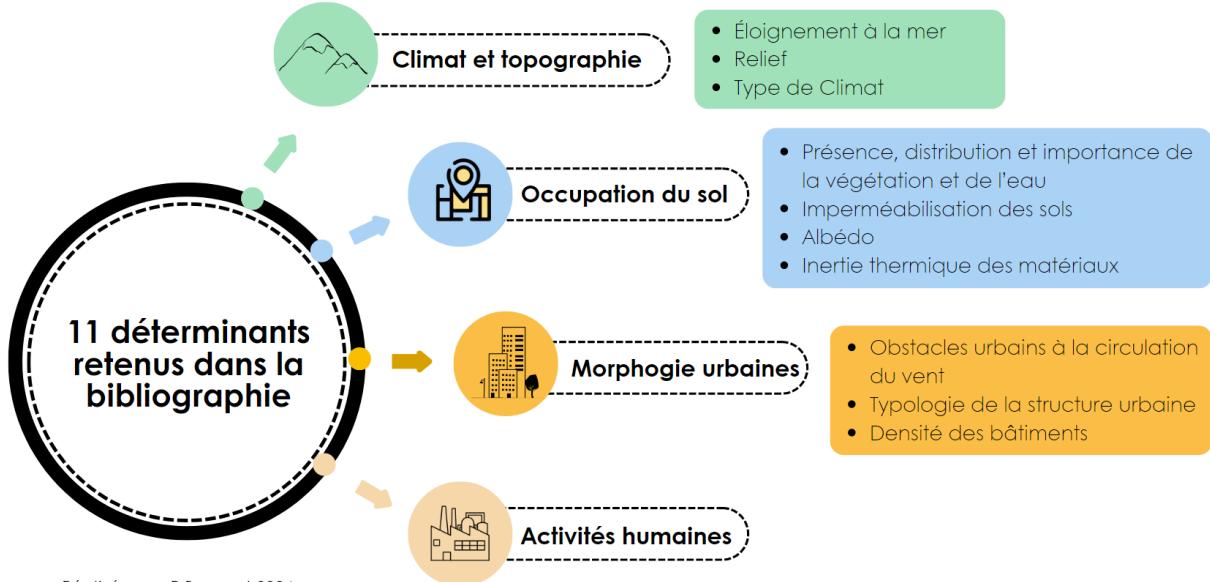


Figure 7 : Les déterminants renforçant le phénomène d'ICU

Le **rayonnement solaire** est la principale source de température sur la Terre. Cette énergie est soit absorbée ou réfléchie par la surface terrestre. L'**absorption ou la réflexion de l'énergie** par les matériaux est quantifiée par l'**albédo**. En effet, un matériau ayant un albédo² faible aura une grande capacité d'absorption. Les matériaux ayant un albédo fort vont fortement réfléchir l'énergie solaire (Phelan et al., 2015). Par exemple, la couleur blanche renvoie très bien l'énergie. La neige fraîche à un albédo 0,87, c'est-à-dire qu'elle réfléchit 87 % du rayonnement solaire. En ville, il a été estimé que plus de 40 % de la superficie avait un albédo faible (ADEME). De plus, l'**inertie thermique**³ de certains matériaux permet de restituer la chaleur emmagasinée le jour sur un temps long. Ainsi, certains matériaux comme le béton allient une forte inertie thermique et un albédo faible. Ils tendent à emmagasiner la chaleur et à relâcher la chaleur sur des intervalles de temps long. D'autre part, le phénomène de **convection thermique** fait circuler la chaleur. Lorsque l'air est chaud, alors il s'élève car il est plus léger que l'air froid, engendrant un dôme de chaleur.

Le climat est également impacté par des **sources de chaleur d'origine humaine**. Par exemple, le trafic routier, le chauffage, la climatisation, les industries, etc, augmentent la température (De Munk et al., 2013). Aussi, la **configuration de la ville** joue un rôle important dans l'ampleur de l'ICU. Les villes polycentriques semblent avoir une intensité plus faible d'ICU que les villes dites monocentriques (Wue Lui et al. 2022, Huimin Liu et al, 2021).

² Part des rayonnements solaires qui sont renvoyés vers l'atmosphère.

³ Capacité physique d'un matériau à conserver sa température.

La **position géographique**, le **climat** et la **topographie du territoire** jouent également un rôle important sur les phénomènes de surchauffe. Effectivement, le **climat** diffère suivant la position du territoire sur le globe. Par ailleurs, le **relief** bloque certaines masses d'air favorisant le maintien de la chaleur. En montagne où la température diminue avec l'altitude, le gradient thermique adiabatique occasionne une perte d'environ 0,6 °C tous les 100 mètres gagnés en altitude.

Le **vent** favorise la circulation de la chaleur, il joue un rôle essentiel dans le **rafraîchissement** de l'air. En ville, la **circulation du vent** est soumise à des **obstacles**, comme les immeubles hauts, les rues étroites, créant ainsi, un effet "canyon" emprisonnant la chaleur (Oke, 1982).

La **présence de végétation et d'eau** occasionne de l'évaporation et de l'évapotranspiration rafraîchissant l'air ambiant (Bowler al., 2010). À l'inverse, l'**imperméabilisation** des sols réduit le potentiel de stockage en l'eau et empêche l'évaporation. Enfin, la présence de la **mer** joue un rôle dans le potentiel de refroidissement de l'air avec notamment des vents plus forts et la présence de l'eau qui est susceptible d'emmageriser la chaleur (Guo, 2023). Une ville près de la mer sera plus fraîche alors qu'une ville continentale aura tendance à être moins aérée et avoir des températures plus chaudes en été et plus froides en hiver.

4. Matériels et méthodes

Pour rappel, le projet vise à créer une typologie de villes sur la base d'indicateurs du climat urbain dans l'**objectif de regrouper des profils de villes sensibles aux phénomènes de surchauffe**.

4.1 Choix du domaine d'étude

Notre étude porte sur les **villes françaises**. En particulier, les analyses se sont concentrées sur la **tache urbaine** de chaque ville. Une **tache urbaine concentre le bâti sans discontinuité**. Pour atténuer l'effet de bord lié au calcul de la tache urbaine de chaque ville, une **zone tampon** négative de 10 mètres a été mise en place. Elle permet de réduire l'impact des territoires adjacents à la zone d'étude. La possibilité de conduire les analyses à l'échelle des aires urbaines au lieu de l'échelle des villes est discuté dans la section 6.

J'ai constitué **deux échantillons** de villes sur lesquels ont été menées **deux classifications distinctes**.

Le **premier échantillon** est tiré dans les villes ayant plus de **2 000 habitants**. Cet échantillon est représentatif de l'**ensemble des villes françaises**. Les villes ont été choisies de manière **aléatoire** suivant un plan d'échantillonnage stratifié. La stratification est réalisée sur la population. La population a été divisée en 10 strates correspondant aux 10 quantiles de population des villes françaises. Quarante observations ont été échantillonnées de façon aléatoire dans chaque strate, soit 400 villes en tout. Les données de population sont issues du recensement de l'**INSEE** de 2021. En réalisant une classification sur cet échantillon, on répond à la question : quelle est la typologie des villes françaises suivant leur microclimat ?

Le **deuxième échantillon** est constitué des villes de plus de **50 000 habitants**. Ce seuil est utilisé car au-delà il est possible d'observer une augmentation de l'intensité de l'ICU (L.W.A. van Hove et al. 2011, P J. Marcotullio, 2021). La France compte 130 villes avec plus de 50 000 habitants et toutes les villes ont été prises dans l'échantillon. En réalisant une classification sur cet échantillon, on répond à la question : quels sont les différents profils climatiques des villes françaises sensibles à la surchauffe urbaine ?

Pour évaluer la pertinence de la classification au regard des phénomènes de surchauffe, les données de terrain de 42 aires urbaines issues du projet **MApUCE** ont été utilisées. Elles ont permis de **comparer** les résultats de la classification avec des données modélisées d'ICU, et ainsi montrer un lien entre la classification obtenue et les ICU.

4.2 Construction des indicateurs du phénomène de surchauffe urbaine

À l'aide de la recherche bibliographique présentée dans la section précédente, j'ai proposé différents indicateurs visant à décrire le profil climatique des villes et leur comportement face à la surchauffe (figure 8).

Figure 8 : Tableau des indicateurs à mettre en place

| | Facteurs | indicateurs calculés | Données | Echelle |
|-----------------------|---|---|------------------------------|------------------------|
| Climat et topographie | Eloignement à la mer | Distance minimale à la mer | NaturalEarth | Monde |
| | Topographie | Moyenne de l'Indice de Rugosité Topographique | SRTM de la NASA, (MNT (30m)) | 80 % du Monde (+ 60 °) |
| | Climat | Climat pour le centroïde de la ville | Classification de Koppen | Geoclimate |
| Occupation du sol | Présence, distribution et importance de la végétation / eau | Moyenne des plus petites distances entre les taches végétales | | |
| | Perméabilité des sols | Proportion de surface perméable | | |
| Morphologie urbaine | Obstacles urbains | Proportion de bâtiments hauts Proportion de bâtiments de hauteur moyenne | Geoclimate | Monde |
| | Densité des bâtiments | Proportion de surfaces compactes Proportion de surfaces non-compactes | | |
| | Typologie de la structure urbaine | Nombre de taches urbaines dans l'emprise de la commune | | |
| | | Surface moyenne des taches urbaines | | |
| Population | Population | Nombre d'habitants | INSEE 2021 | France |

Certains éléments de la bibliographie, comme l'inertie des matériaux, n'ont pas été retenus car il n'y a pas de données disponibles fiables pour constituer un indicateur pour l'ensemble de villes retenues.

Concernant les données mobilisées pour construire les indicateurs, je me suis appuyée sur Geoclimate (Bocher, 2021) pour calculer les indicateurs se basant sur les ZCL, la tache urbaine et la morphologie du bâti. Pour le tracé des lignes de côte, j'ai utilisé les données de **NaturalEarth**, le **Shuttle Radar Topography Mission** (SRTM) de la NASA pour récupérer, via le package `elevtr`, le Modèle Numérique de Terrain (MNT). Enfin, les données du climat selon la classification de **Köppen-Geiger** sont collectées avec le package `kgc`. La classification climatique de Köppen-Geiger divise le globe en plusieurs familles de climats. Elle s'appuie sur les moyennes annuelles et mensuelles des températures et des précipitations, elle a été publiée en 1884. La classification de Köppen-Geiger sépare les climats en cinq grandes

catégories : A : climats équatoriaux, B : climats arides et semi-arides, C : climats tempérés sans hiver froid, D : climats tempérés à hiver froid, E : climats polaire. Puis la typologie est complétée et affinée avec des informations mensuelles et annuelles, sur les précipitations et la température.

4.2.1 Climat et topographie

4.2.1.1 Eloignement à la mer

L'objectif de cet indicateur est de classer les villes suivant leur **éloignement à la mer**, puisque la mer joue un rôle sur le rafraîchissement de la ville.

En entrée, le script télécharge via le package *rnaturalhearth* (Massicotte, South, 2024) les données des lignes de côtes. Avec ce package le traitement est plus rapide qu'en requêtant OSM. Pour calculer la **distance**, à vol d'oiseau, le choix a été fait de prendre le centroïde de la limite communale de la ville en forçant le point dans le polygone.

4.2.1.2 Indice de Rugosité Topographique

L'indice de rugosité topographique (TRI) permet de caractériser l'**aspérité du relief**. Cet indicateur exprime la différence d'élévation entre les cellules adjacentes d'un MNT. Pour cet indicateur, la moyenne sur la surface totale de l'emprise a été choisie. Ainsi, plus la moyenne de TRI est faible, plus le terrain est plat et plus la circulation de l'air est élevée.

Concrètement, le script fait une boucle sur toutes les taches urbaines pour calculer le TRI. Il commence par créer une table en dehors de la boucle. Ensuite, il rentre dans la boucle et vérifie pour chaque ville si des données spatiales existent. Puis le code récupère via le package *elevatr* les données du MNT pour l'emprise, puis il calcule le TRI et sa moyenne. Le package *elevatr* va rechercher les données issues de la NASA et plus particulièrement le SRTM qui recouvre 80 % du monde avec une résolution spatiale de 30 mètres.

4.2.1.3 Climat

À l'aide du package *kgc* il est possible de déterminer pour un point sa classification de Köppen-Geiger. Ainsi, une table est créée avec la classification de Köppen-Geiger pour le centroïde du territoire.

4.2.2 Occupation du sol

4.2.2.1 Présence, distribution, et importance de la végétation

Pour étudier la distribution des espaces végétaux deux indicateurs ont été calculés : la **distance minimale entre les tâches végétales** et la **proportion de tâches végétales** sur le territoire. Ces indicateurs ont été calculés à l'aide de *Geoclimate*. La **tache végétale** a été définie avec les ZCL, elle correspond aux ZCL suivantes :

Espace dense arboré, Espace peu arboré, Espace végétalisé hétérogène et Végétation basse.

Avant de **mesurer la distance**, j'ai regroupé puis découpé spatialement les taches végétales. J'ai ensuite calculé pour chaque tache végétale la distance à la tache végétale la plus proche. La moyenne des distances sert d'indicateur à l'échelle de la ville.

Concernant la **proportion de taches végétales**, j'ai calculé le nombre de taches végétales sur le territoire que j'ai rapporté sur la superficie de celui-ci. L'objectif est de mesurer la **dispersion** des taches végétales qui permettent un rafraîchissement de l'air ambiant. Pour éviter le biais lié à la superficie du territoire, j'ai calculé cet indicateur en rapportant le nombre de taches végétales sur la surface totale. Cette méthode permet d'obtenir une proportion, assurant une interprétation claire des résultats. Effectivement, un territoire plus vaste pourrait avoir un nombre élevé de taches végétales sans pour autant refléter une meilleure répartition de ces espaces.

4.2.2.2 Perméabilité des sols

Le calcul de la perméabilité des sols permet de quantifier les espaces perméables qui rafraîchissent la ville. Pour ce faire, j'ai calculé la **proportion de surfaces perméables** sur la surface totale de la tache urbaine. Pour déterminer la perméabilité des sols, je me suis appuyée sur les ZCL suivantes : Espace dense arboré, Espace peu arboré, Espace végétalisé hétérogène, Végétation basse, Sol nu, Eau, Maisons diffuses. Le choix d'inclure les maisons diffuses se justifie par la végétalisation des jardins. Ces espaces verts favorisent l'infiltration naturelle de l'eau. Aussi, la présence d'espaces verts est significative, elle contribue, notamment, à réguler la température locale en atténuant les effets de la surchauffe. Ainsi, les maisons diffuses, bien que partiellement bâties, jouent un rôle physique dans la réduction de la chaleur en ville (Bowler et al. 2010).

4.2.3 Morphologie urbaine

4.2.3.1 Obstacles urbains

J'ai défini **deux indicateurs** pour étudier les obstacles urbains : les bâtiments de hauteurs moyennes, les bâtiments de hauteur élevée. L'objectif de l'indicateur est d'étudier la **circulation du vent** puisque les obstacles urbains, définis par la hauteur des bâtiments, jouent un rôle sur le refroidissement de la ville. Les bâtiments de hauteurs élevées sont définis par les classes : Tours compactes et Tours espacées. Les bâtiments de hauteurs intermédiaires sont définis par : Immeubles compacts, Immeubles espacés. Les tours sont définies par une nombre d'étages supérieur à dix et les immeubles ont un nombre d'étage entre trois et neuf (cf annexe 3 : Description des ZCL par Oke) (Oke, 2012).

4.2.3.2 Densité bâtiments

La densité de l'urbain permet de quantifier **l'inertie thermique du bâti**. Une surface compacte est un espace avec une forte concentration de bâti élevé et rapproché. Elle aura tendance à beaucoup retenir la chaleur. Pour quantifier cet indicateur, la démarche s'appuie sur les classes de ZCL dites **compactes** : Tours compactes, Immeubles compacts, Maisons compactes. D'autre part, la proportion des surfaces **non-compactes** a été calculée à partir des ZCL suivantes : Tours espacées, Immeubles espacés, Maisons espacées, Bâtiments industriels, Maisons diffuses. Ainsi, la **proportion** de surface compacte et non-compact sur la ville a été calculée.

4.2.3.3 Typologie de la structure urbaine

Décrire la typologie de la structure urbaine permet de quantifier **l'étendue** et la **structure de la tache urbaine** qui ont un impact sur l'ICU (Wue Lui et al. 2022, Huimin Liu et al, 2021). Une ville, à taille équivalente, constituée d'une seule tache urbaine emmagasinera plus de chaleur qu'une ville constituée de plusieurs taches urbaines. Pour cela, je mesure le nombre de taches urbaines. Le script mesure la **surface moyenne** des taches urbaines en hectares et calcule le **nombre de taches urbaines**. Une tache urbaine concentre le bâti sans discontinuité.

4.3 Méthode de classification

L'étude comporte des **données quantitatives** et **qualitatives**. L'**analyse factorielle des données mixtes** (AFDM) permet de faire de la **classification non supervisée** sur ce type de données. L'objectif de cette démarche est de réduire la dimension du jeu de données étudié et de projeter les données dans un espace de faible dimension. Les données brutes positifs continus ont été **log-transformées** pour **limiter l'effet des valeurs extrêmes** et la présence d'artefacts dans les plans de l'AFDM. Les données de proportion ont été traitées suivant les méthodes d'imputation non-paramétrique décrites par Martin-Fernandez et ses coauteurs (2003) pour pouvoir gérer la zéro inflation. J'ai fait le choix de baser la classification sur la méthode des **kmeans**. C'est une méthode robuste, en comparaison avec la méthode de classification ascendante hiérarchique. Le nombre de groupes dans la classification a été déterminé à l'aide de la variance intra-classe de chaque groupe (figure 14, 21).

4.4 Métriques utilisées pour comparer les groupes de villes et les données d'ICU

Pour comparer les groupes issues de la classification avec les données de MApUCE. J'ai comparé la moyenne de l'intensité de l'ICU des villes disponibles à son groupe identifié par la classification. Le *figure 9* localise les aires urbaines étudiées par MApUCE et la *figure 10* illustre un ICU pour la ville de Clermont-Ferrand.

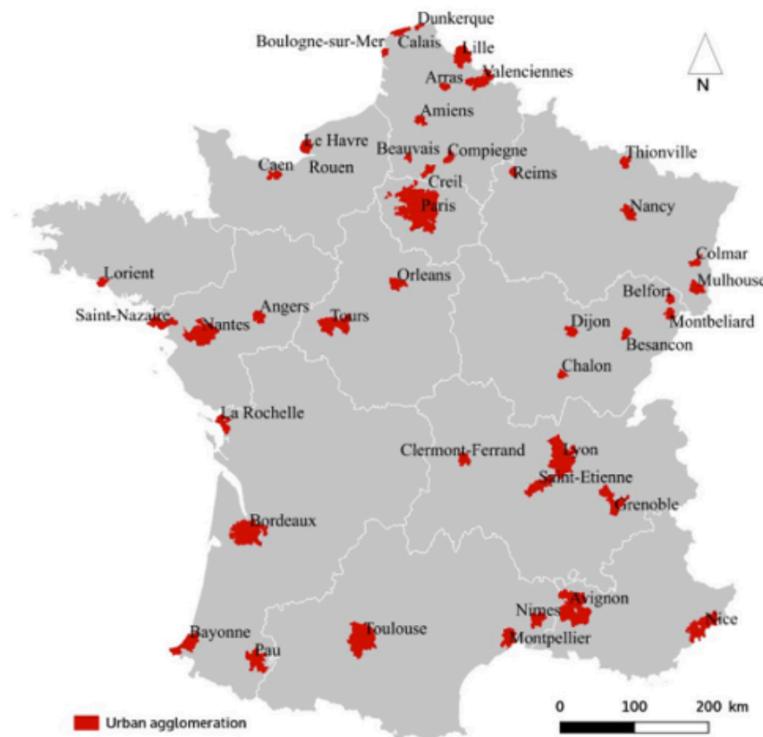


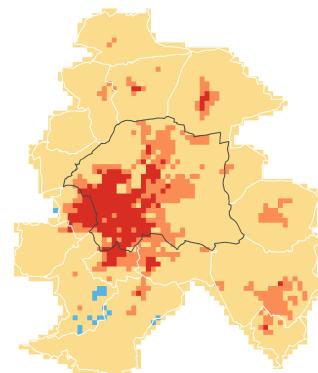
Figure 9 : Aires urbaines étudiées par MApUCE

Figure 10 : Exemple de représentation de l'ICU pour l'aire urbaine de Clermont-Ferrand (FranceInfo, 2021)

L'îlot de chaleur à Clermont-Ferrand

Par intensité

| | |
|---------------------|-------------------|
| ■ Ilot de fraîcheur | ■ Non négligeable |
| ■ Négligeable | ■ Fort |



Source : MApUCE. Crédits : franceinfo

4.5 Automatisation du processus

À la fin, j'ai constitué un ensemble de scripts R, pour automatiser la chaîne de traitement des données (figure 11). L'objectif était de scinder le script en plusieurs étapes et de clarifier le code, puisque certains scripts comportaient un nombre important de lignes, ce qui ne facilitait pas la manipulation du code.

Automatisation du processus

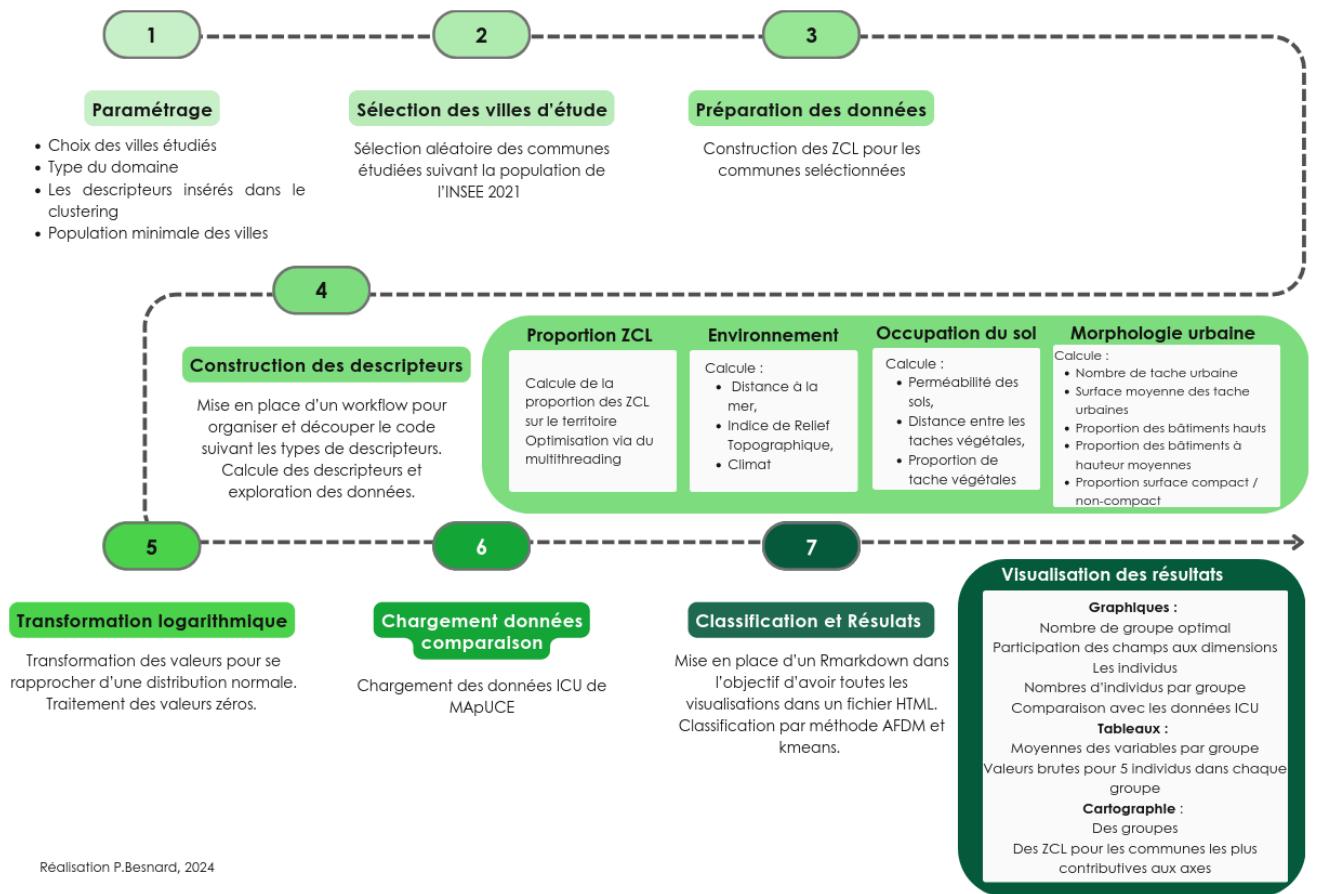


Figure 11 : Chaîne de traitement pour la construction et l'analyse des classifications des villes

La chaîne de traitement des données permet de découper les différentes étapes de l'analyse. En sortie, les données géographiques sont au format flat geo buffer. Ce fichier unique est plus léger que le format geojson ou shapefile puisqu'il utilise le format binaire. Pour chaque indicateur, j'ai construit des graphiques et des tableaux pour **explorer les données**. La figure 12 donne un exemple de graphique interactif réalisé avec le package ggplotly.

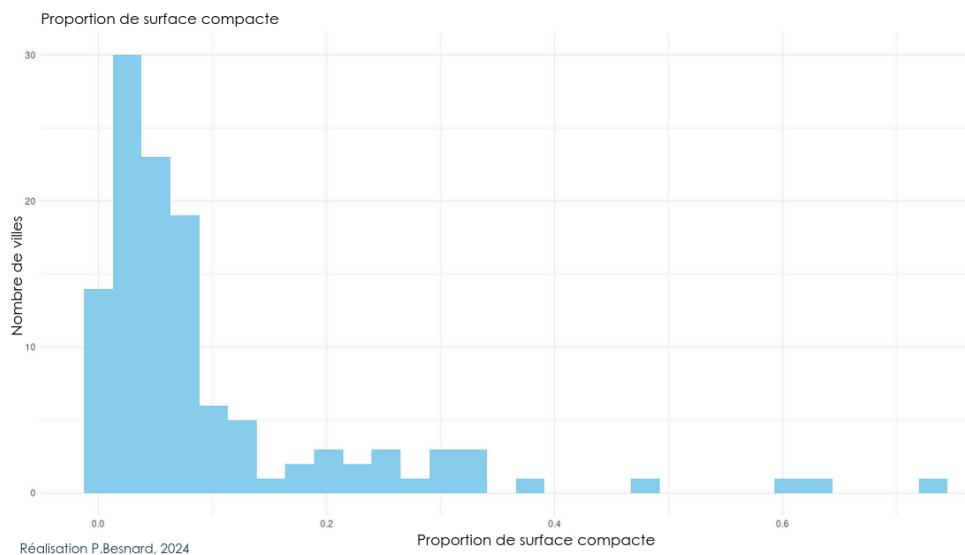


Figure 12 : Exemple de graphique permettant l'exploration de données Proportion de surface compacte pour les grandes villes françaises.

À noter que pour le calcul des indicateurs issus de Geoclimate, j'ai mis en place du multi-threading pour réduire le temps d'exécution du code. Le multi-threading permet à un programme d'exécuter plusieurs tâches en même temps. Il existe le package *doParallel* et *foreach* pour exécuter des itérations en parallèle.

En vue d'automatiser la visualisation des résultats. J'ai mis en place un Rmardown pour représenter les résultats de l'analyse. J'ai privilégié le format HTML car il permet la visualisation interactives de cartes et de graphiques. De nombreux graphiques, tableaux et cartes sont générés pour explorer au mieux les résultats. L'annexe 6 est un document HTML présentant les résultats de la classification pour les principales villes françaises (cf annexe 6 : *Résultats sous forme de fichier Rmardown pour les grandes villes françaises*).

5. Résultats

5.1 Sur l'ensemble des villes françaises

5.1.1 Construction des axes

La dimension 1 contribue à 20,3 % de la variance l'information et la dimension 2 à 11,6 %. Il est possible d'identifier deux **groupes de variables** (figure 13) : les variables favorables au refroidissement (végétalisation) et les variables favorisant la surchauffe en ville (compacté de l'urbain). En effet, plus la "proportion de taches végétales" est élevée, plus la "distance entre les taches végétales diminue" (corrélation négative de -0,38 selon le test de corrélation linéaire de Pearson). De plus, les variables : "proportion de compact", "proportion de bâtiments avec des hauteurs moyennes" sont corrélés positivement.

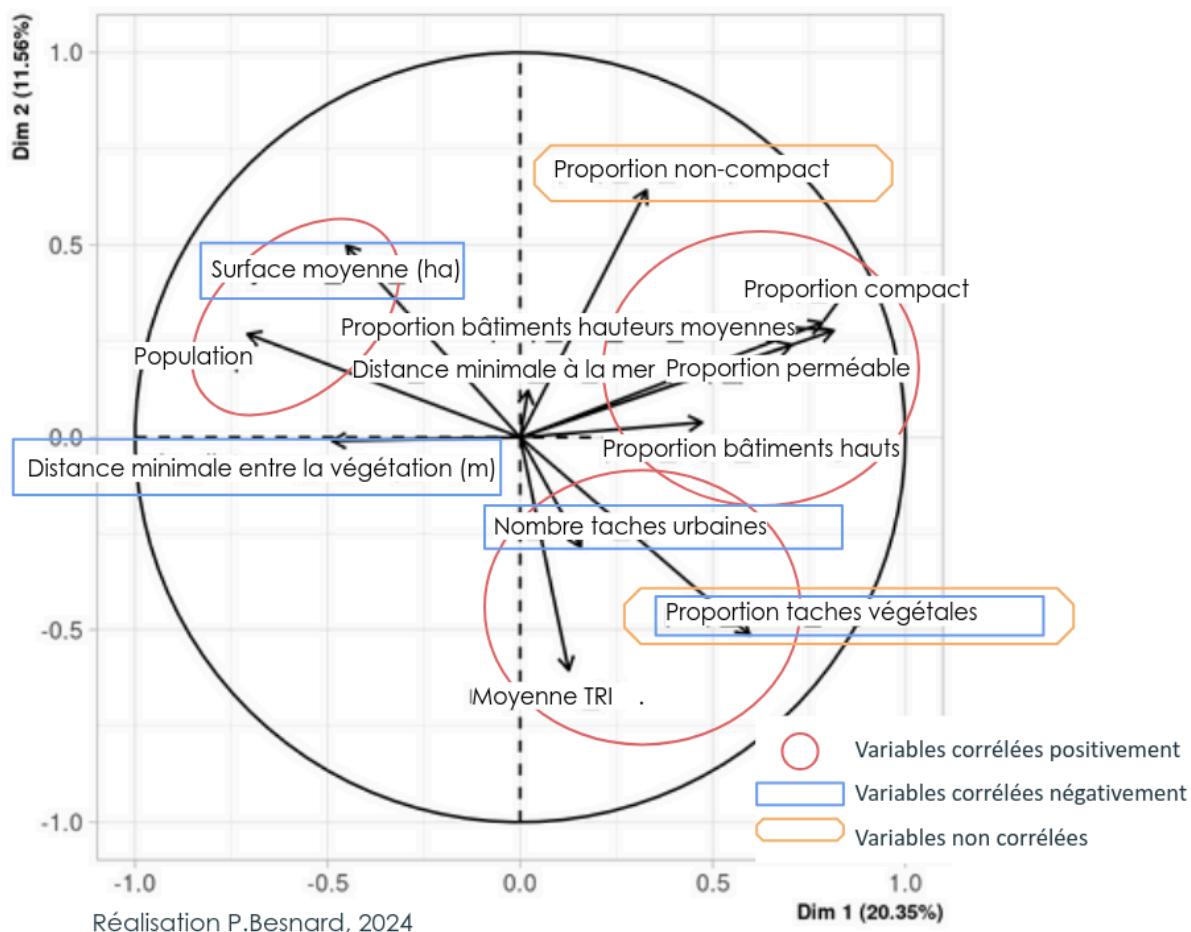


Figure 13 : Visualisation des variables quantitatives dans les deux premières dimensions de l'AFDM

5.1.2 Regroupement des individus

Le nombre de groupes de la classification a été déterminé à l'aide du graphique de l'inertie intra-classe (figure 14). Il représente la somme des carrés

intra-cluster, c'est-à-dire la distance entre chaque point et le centre du groupe. Cette approche permet de sélectionner un nombre de groupes pour lequel la variance intra-classe est faible garantissant ainsi une homogénéité des groupes. J'ai étudié une classification en 6 groupes puisqu'au-delà l'amélioration, l'inertie intra-classe est faible.

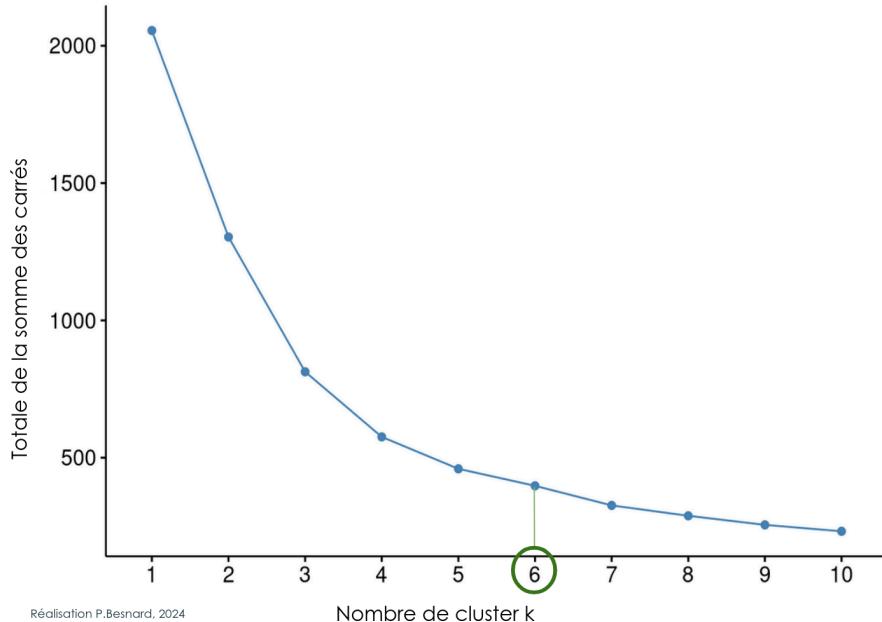


Figure 14 : Choix du nombre de groupes

5.1.3 Typologie des villes pour l'ensemble des villes françaises

À partir d'un tableau représentant les moyennes des valeurs des variables pour chaque groupe (tableau 1), il est possible de déterminer le profil de villes. En comparant ces profils moyens à l'analyse bibliographique, il est possible de caractériser les différents groupes de villes au regard de leur micro-climat et de mettre en évidence un **gradient de la sensibilité à la surchauffe urbaine**. Le nombre de villes est plus faible pour les groupes situés aux extrémités du gradient de sensibilité (figure 15, 16).

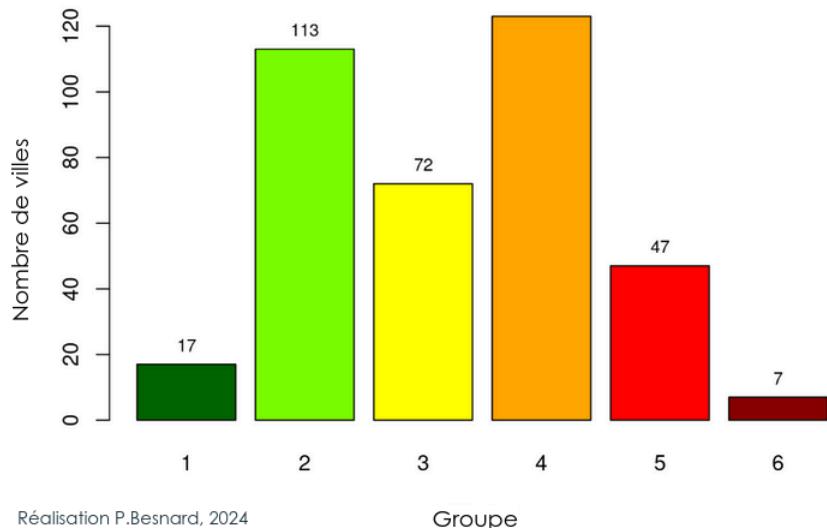


Figure 15 : Nombre d'individus par groupes allant du moins au plus sensible

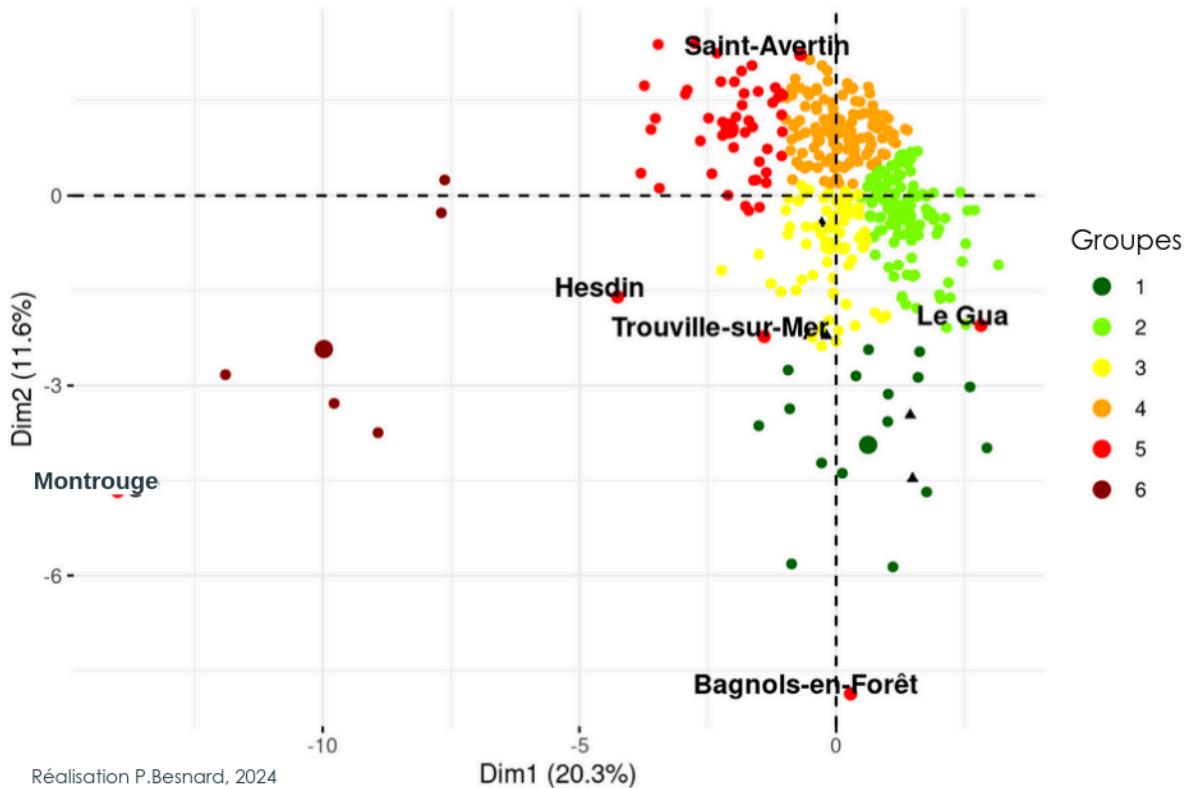


Figure 16 : Visualisation des individus sur les 2 premières dimensions de l'AFDM, le nom des communes représente les individus les plus contributifs.

Le **groupe 1** fait référence aux villes rurales **peu peuplées** (tableau 1). Les **taches urbaines sont disjointes** et **peu étendues**. La végétation y joue un rôle important. Ce sont des villes influencées par le **relief** et par le **climat**. Ce groupe se caractérise par un relief **accidenté**, cela est visible spatialement, puisque les communes de ce groupe se localisent principalement dans les massifs montagneux et en particuliers

dans les Alpes (figure 17). Exemple : Bagnols-en-Forêt.

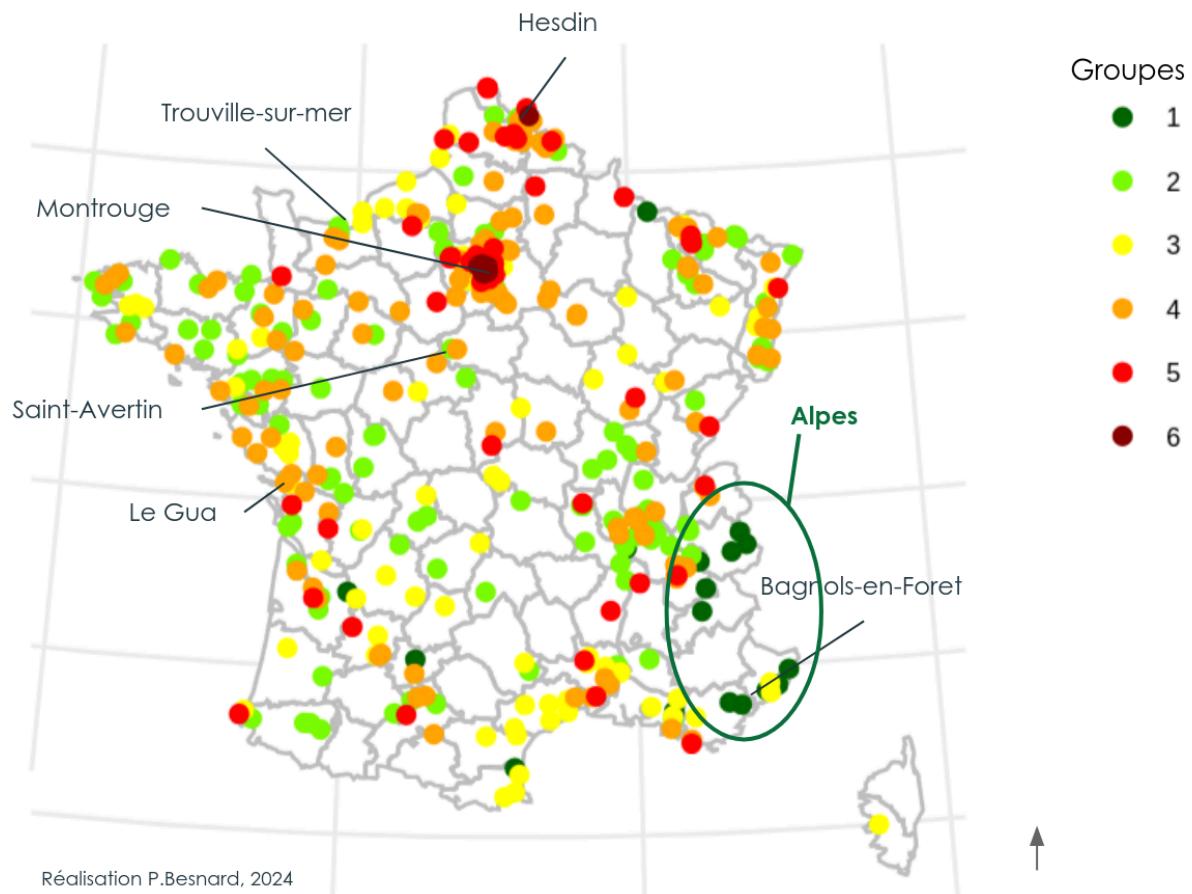


Figure 17 : Spatialisation des profil de villes, focus sur les villes du groupe 1 impactées par le relief

Le **groupe 2** correspond aux villes rurales avec des **taches urbaines disjointes peu étendues**. Il y a une forte prédominance de la **végétation** avec une **distance entre les taches végétales faible**. Les surfaces sont essentiellement **perméables**. Exemple : Le Gua.

Le **groupe 3** est caractérisé par des **zones urbaines disjointes peu étendues**. La distance entre les taches végétales est moyenne, et leur proportion diminue par rapport aux deux groupes précédents. Exemple : Trouville-sur-Mer.

Les villes du **groupe 4** sont constituées par **une seule tache urbaine**. Les surfaces sont principalement **perméables**, cependant des bâtiments de **hauteurs moyennes** commencent à apparaître. Exemple : Saint-Avertin.

Les villes du **groupe 5**, sont semblables au groupe 4 avec une surface plus étendue. La surface perméable diminue. Exemple : Hesdin.

Enfin, le **groupe 6** se compose de villes constituées d'**une seule tache urbaine**, les bâtiments sont de **hauteurs moyennes** et s'étendent sur presque 50 % de la superficie. La part de la végétation est plus restreinte, la distance entre les taches végétales est plus grande. Exemple : Montrouge. Ce profil de ville se concentre autour de la capitale (figure 18).

Au final, la proportion de taches végétales est une variable qui **diminue** le long du

gradient de sensibilité, là où la variable de population augmente (tableau 1). La figure 19 illustre ce gradient en gardant les villes qui caractérisent chaque groupe.

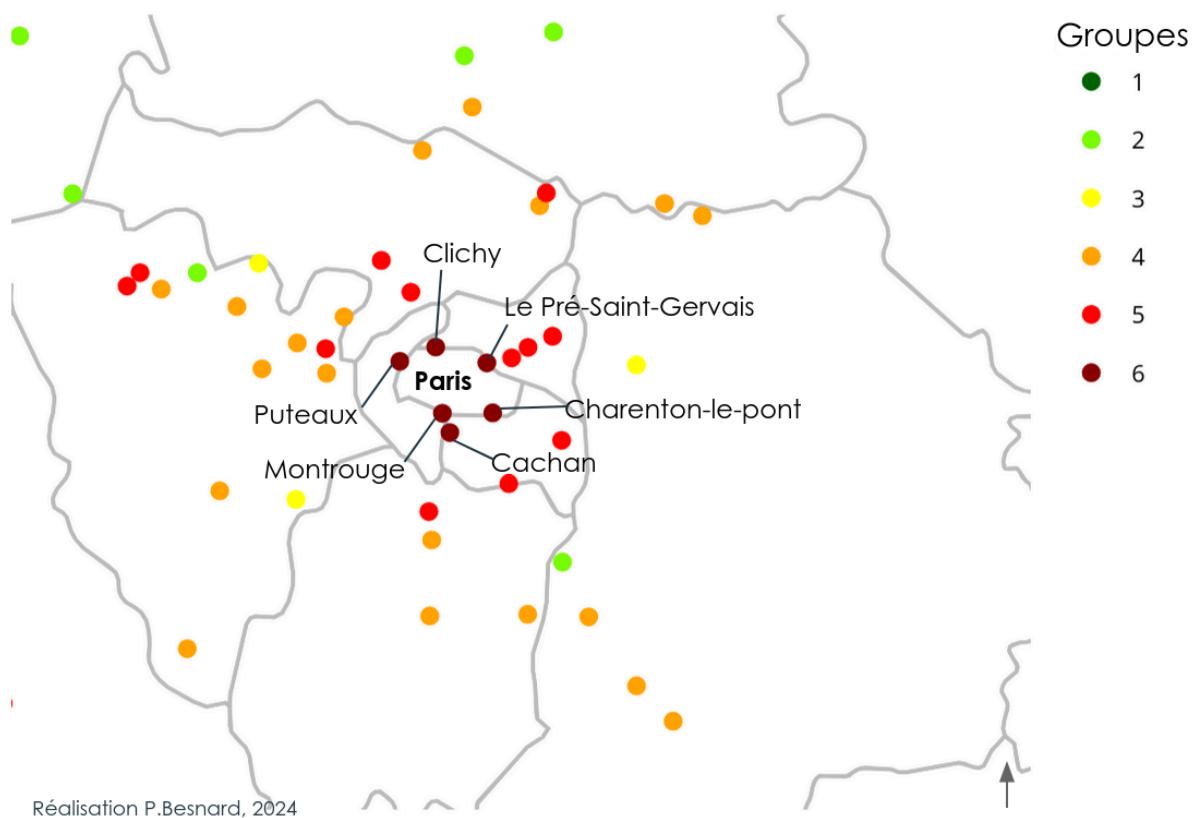


Figure 18 : Localisation des villes appartenant au profil 6

Tableau 1 : Tableau des moyennes pour les six groupes avec des variables sélectionnées et arrondies

| | Groupe 1 | Groupe 2 | Groupe 3 | Groupe 4 | Groupe 5 | Groupe 6 |
|---|----------|----------|----------|----------|----------|----------|
| Population | 3 000 | 3 500 | 5 300 | 6 000 | 25 000 | 47 000 |
| Moyenne indice TRI | 1.25 | 0.51 | 0.56 | 0.34 | 0.34 | 0.43 |
| Surface perméable | 49 % | 78 % | 56 % | 70 % | 44 % | 10 % |
| Moyenne des distances minimales entre les taches végétales | 75 | 55 | 65 | 70 | 90 | 175 |
| Proportion taches végétales | 0.4 | 0.38 | 0.23 | 0.19 | 0.1 | 0.06 |
| Surface de bâtiment à hauteurs moyennes | 1,8 % | 0,4 % | 2 % | 7 % | 4 % | 48 % |
| Superficie moyenne des tache urbaines (hectares) | 40 | 70 | 130 | 225 | 625 | 405 |

| | | | | | | |
|----------------------------------|---|---|---|-----|-----|---|
| Nombre de taches urbaines | 2 | 2 | 2 | 1.3 | 1.4 | 1 |
|----------------------------------|---|---|---|-----|-----|---|

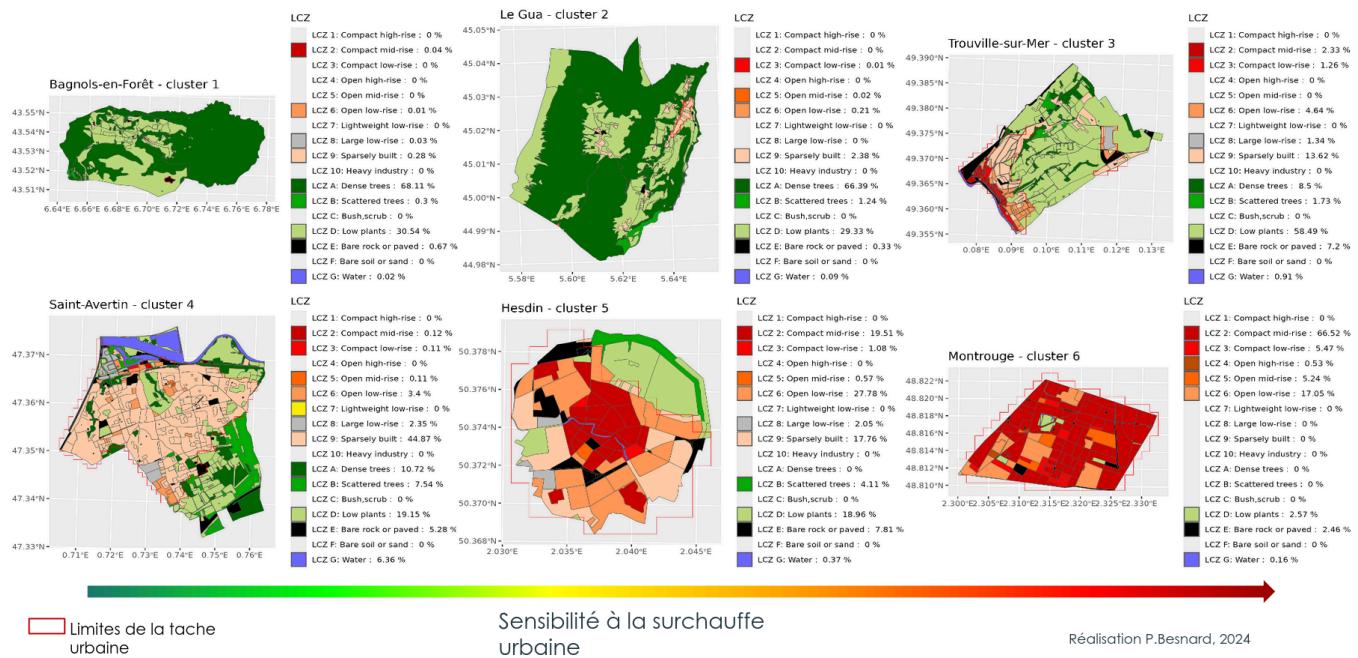


Figure 19 : Cartographie des ZCL suivant leur sensibilité

5.2 Sur les grandes villes françaises

5.2.1 Construction des axes

Les deux premières dimensions rassemblent 43,9 % de la variance. Ici, la visualisation de **deux groupes de variables opposés** est plus nette (figure 20). En effet, les variables “proportion de compact” et “proportion de bâtiment à hauteurs moyennes” sont fortement corrélées (test de corrélation linéaire de Pearson de 0,98). De l'autre côté, les variables “nombre de tache urbaine” et “proportion de perméable” sont corrélées avec un test de corrélation linéaire de Pearson de 0,53.

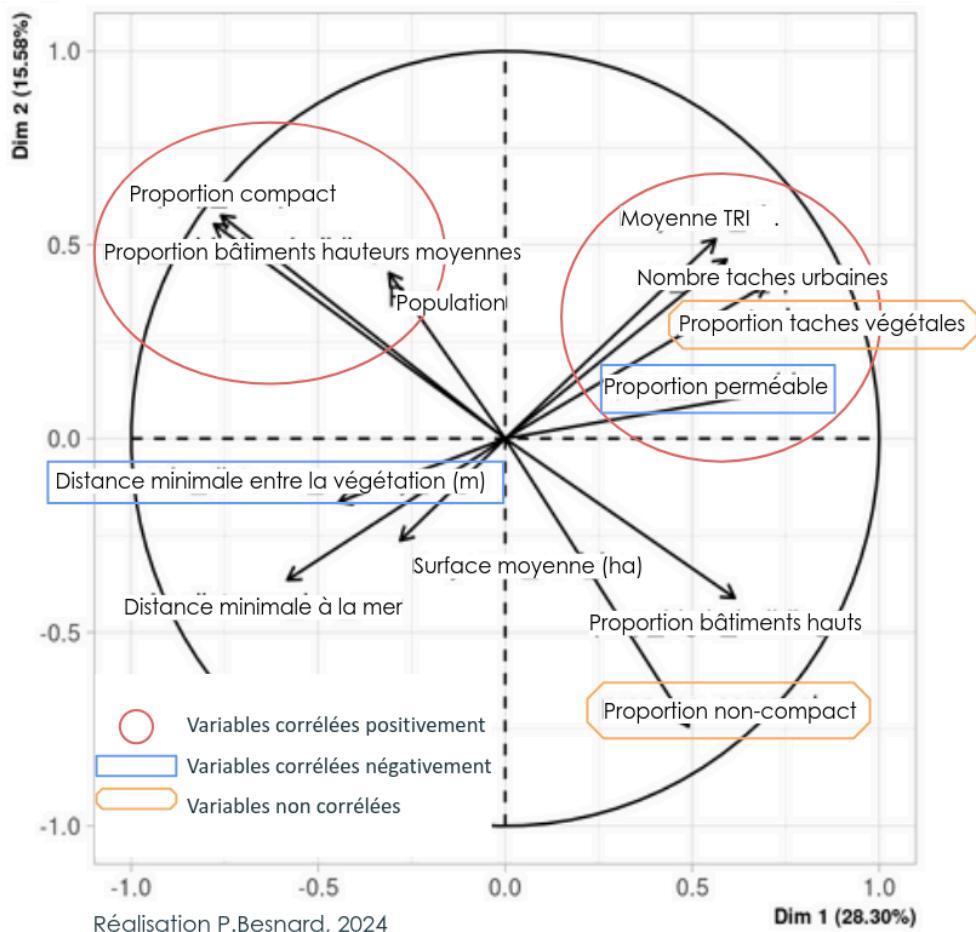


Figure 20 : Visualisation des variables quantitatives qui contribuent aux dimensions

5.2.2 Regroupement des individus

Pour la classification, j'ai retenu cinq groupes, puisque plus de groupes ne permettait pas de constituer des groupes plus homogènes (figure 21).

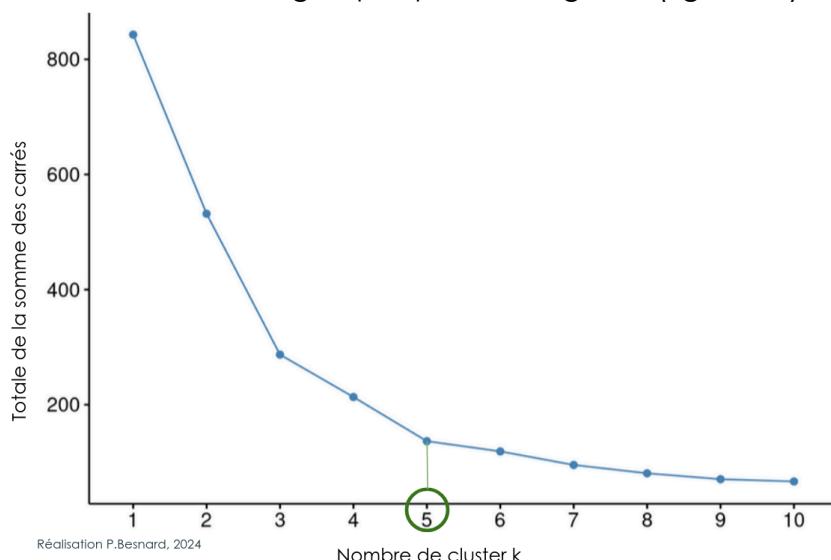


Figure 21 : Choix du nombre de groupes

5.2.3 Typologie des grandes villes françaises

L'étude des groupes issus des **grandes villes françaises** montre bien un gradient à la sensibilité pour la surchauffe urbaine. Les groupes situés aux extrémités du gradient comptent moins de villes (figure 22, 23).

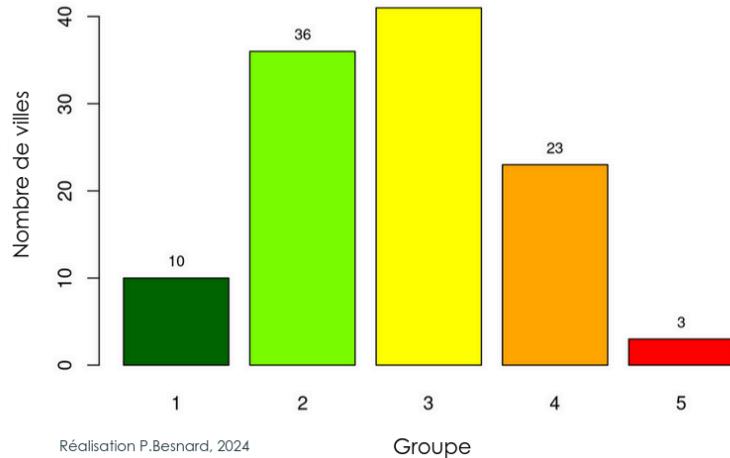


Figure 22 : Nombre d'individus par groupes allant du moins au plus sensible

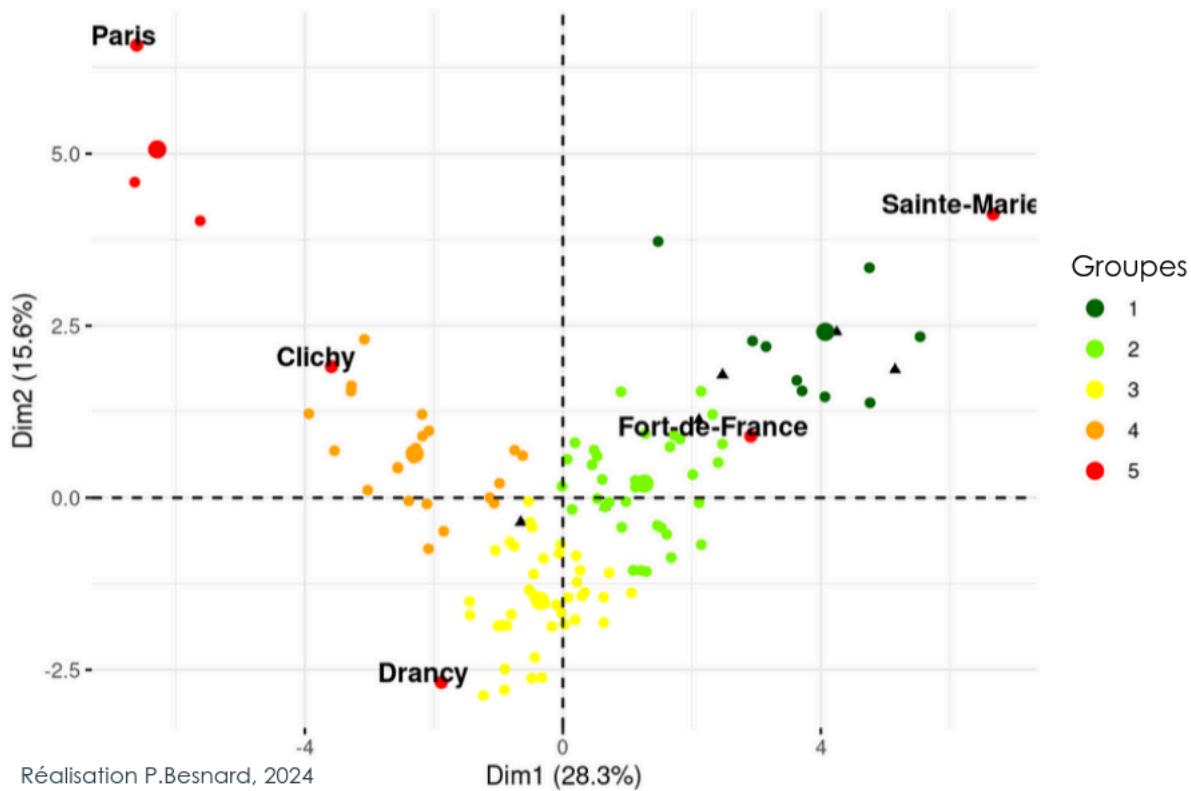


Figure 23 : Visualisation des individus sur les 2 premières dimensions de l'AFDM, le nom des communes représente les individus les plus contributifs.

Le **groupe 1** correspond aux villes **côtières** avec **nombreuses tâches urbaines disjointes** (figure 24, tableau 2). Les territoires ont un taux de recouvrement **perméable** assez fort. Exemple : Sainte-Marie (Martinique)

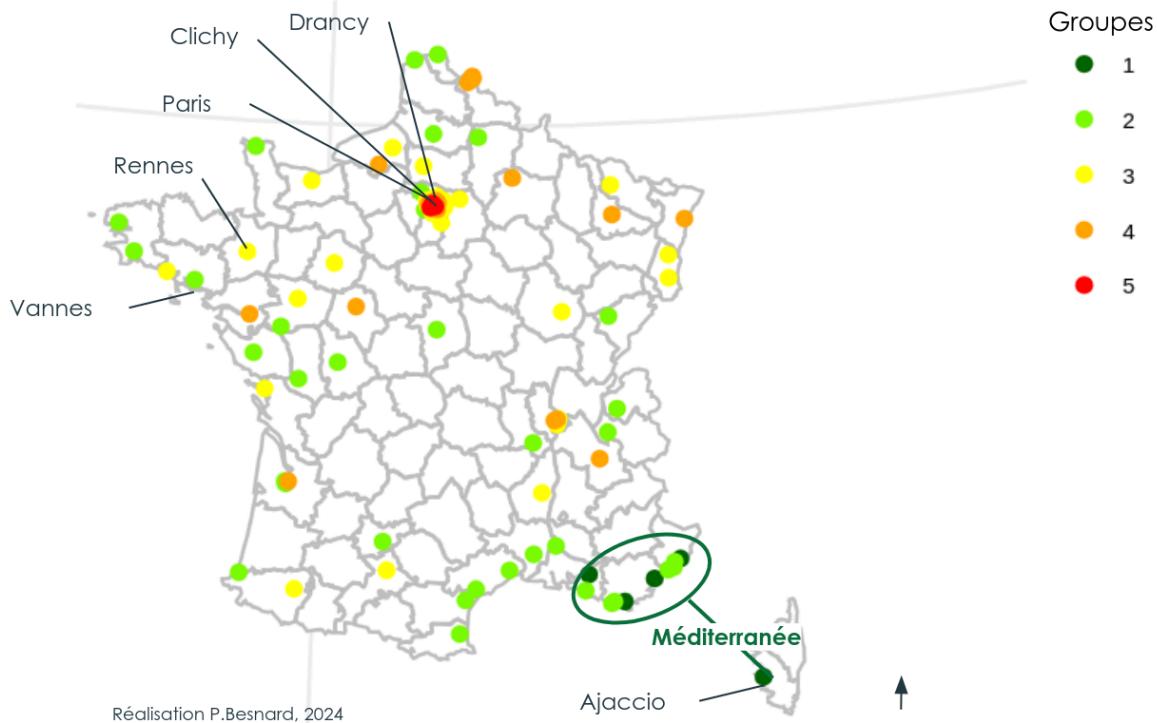


Figure 24 : Spatialisation des différents profils de villes, spatialisation marquée pour les villes du groupe 1

Le **groupe 2** correspond à des villes plus **étendues** que le **groupe 1**. La place de la **végétation** est importante avec une distance entre les taches végétales faibles. Exemple : Fort-de-France (Martinique).

Le **groupe 3** est constitué de villes formant une **seule tache urbaine**. Ce groupe est caractérisé par une **forte superficie** de la tache urbaine. Exemple : Drancy.

Le **groupe 4** regroupe des villes formant une **seule tache urbaine**, elles sont fortement **étendues**, avec des **bâtiments à hauteurs moyennes**. La surface perméable reste élevée. Exemple : Clichy

Le **groupe 5** regroupe des villes formant seule tache urbaine, avec de nombreux **obstacles urbains**. Exemple : Paris.

La surface perméable est une variable qui diminue constamment le long du gradient de sensibilité, de même pour la proportion de taches végétales. À l'inverse, la surface des bâtiments de hauteurs moyennes augmente continuellement (tableau 2, figure 25).

Tableau 2 : Tableau des moyennes pour les cinq groupes avec des variables sélectionnées et arrondies

| | Groupe 1 | Groupe 2 | Groupe 3 | Groupe 4 | Groupe 5 |
|---|----------|----------|----------|----------|----------|
| Pop | 102 000 | 113 000 | 90 000 | 157 000 | 774 000 |
| Distance minimum à la mer (kilomètres) | 6 | 70 | 162 | 151 | 147 |

| Surface perméable | 60 % | 45 % | 30 % | 20 % | 10 % |
|---|------|------|------|------|------|
| Moyenne des distances minimales entre les taches végétales | 71 | 70 | 103 | 105 | 143 |
| Proportion taches végétales | 0.16 | 0.1 | 0.07 | 0.06 | 0.04 |
| Surface de batiment à hauteurs moyennes | 3 % | 6 % | 6 % | 27% | 70 % |
| Superficie moyenne des tache urbaines (hectares) | 365 | 1320 | 1645 | 1640 | 855 |
| Nombre de taches urbaines | 6.8 | 2.8 | 1.1 | 1.2 | 2.3 |
| Surface compact | 2 % | 5 % | 4 % | 24 % | 66 % |

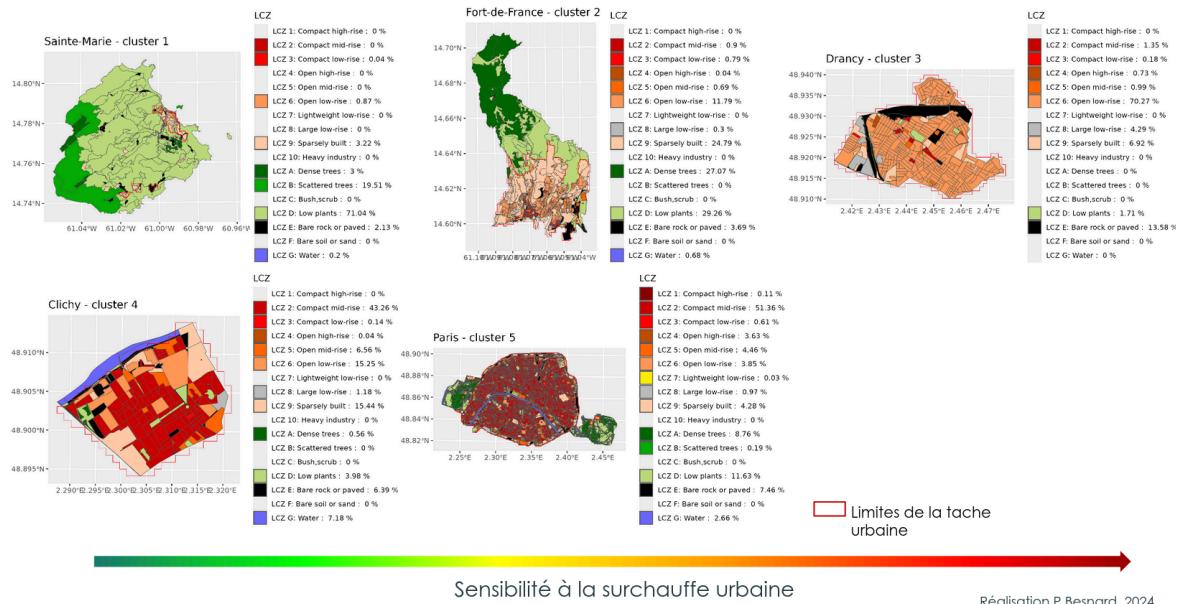


Figure 25 : Représentation des ZCL pour les villes d'exemple suivant leur sensibilité

5.2.4 Comparaison avec les données ICU

J'ai cherché à valider le **gradient de sensibilité à la surchauffe urbaine**. identifié par la classification, en comparant la classification aux **données d'ICU** modélisées par MApUCE. Le graphique ci-dessous (figure 26) représente l'**intensité des ICU** en fonction du groupe identifié dans la classification. Il permet d'identifier bien de visualiser un gradient d'intensité d'ICU qui corrobore la classification.

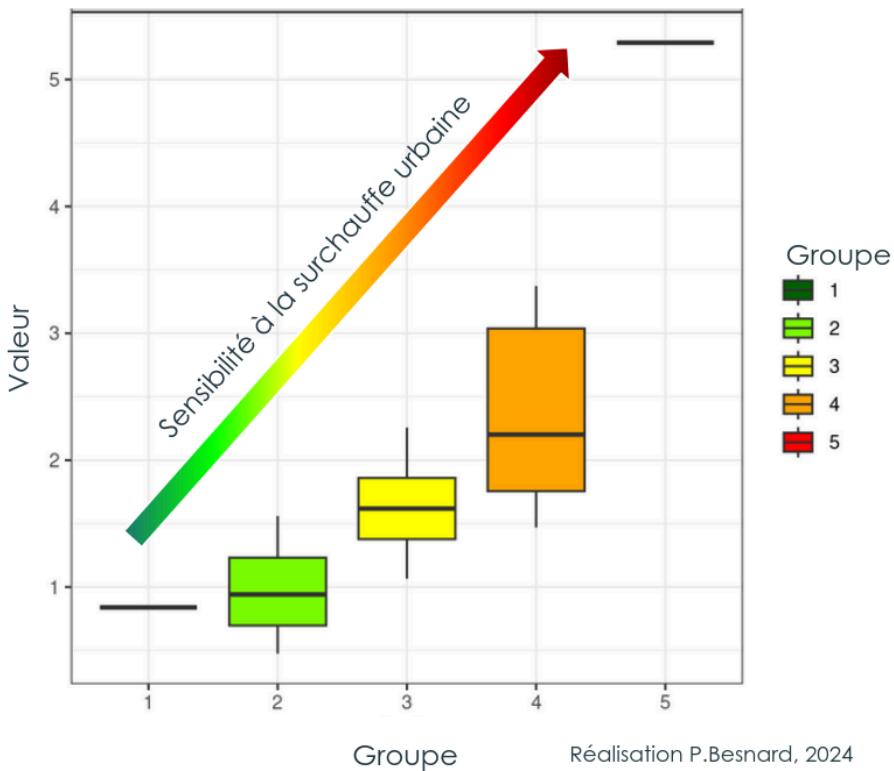


Figure 26 : Graphique de la moyenne de l'intensité des ICU suivant les données de MApUCE

Sur les dix villes, seule Nice est une ville où des données MApUCE sont disponibles pour le **groupe 1**, l'intensité de l'ICU est de 0,8 °C. Sept villes de MApUCE sont contenues dans le **groupe 2**, la **médiane du groupe est légèrement supérieure** à celle du **groupe 1** (0,9 °C). L'écart s'agrandit avec le **groupe 3**, puisque la médiane augmente de 0,6 °C par rapport au **groupe 2**. Le **groupe 4** comporte 9 villes, les données d'intensité à l'ICU sont **plus dispersées** que pour le **groupe 3** qui comporte 10 villes. La médiane reste supérieure à celle du **groupe 3**, 2,2 °C. Cependant, le groupe se démarquant le plus est le **groupe 5** (Paris), se démarque le plus des autres villes avec une intensité d'ICU de 5,2 °C.

6.Discussions

6.1 Synthèse des principaux résultats

L'objectif de cette étude était de constituer une typologie des villes françaises au regard de leur microclimat et d'identifier des profils de villes sensibles au phénomène de surchauffe. J'ai identifié un gradient à la sensibilité aux phénomènes de surchauffe. Ce gradient est corrélé aux indicateurs du climat définis dans l'étude. Plus particulièrement deux groupes de variables s'opposent très clairement pour les grandes villes françaises (végétation/urbain - figure 27).

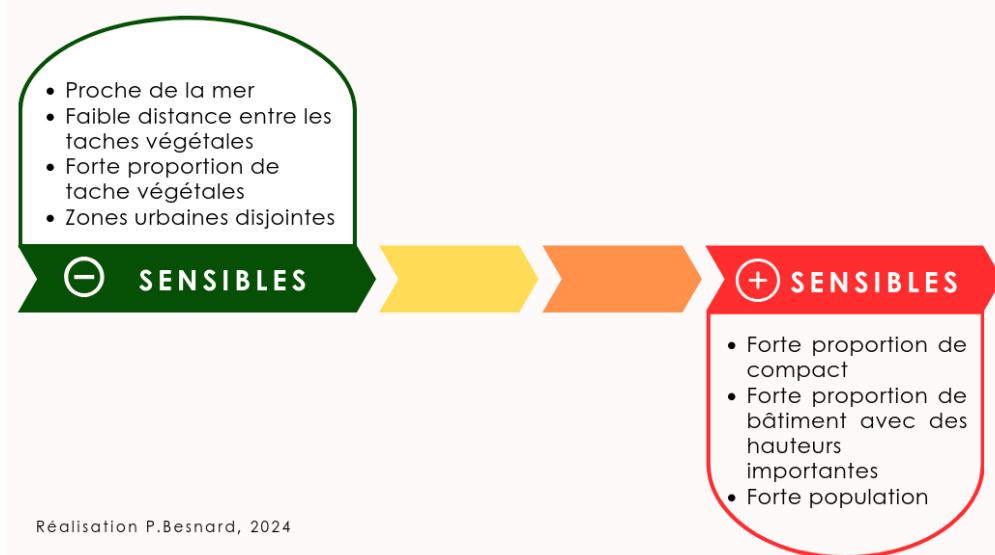


Figure 27 : Gradient de la sensibilité des villes à la surchauffe urbaines suivant les indicateurs mis en place.

6.2 Limites de la méthode et pistes d'amélioration

La méthode comporte certaines limites, effectivement, malgré la volonté de limiter l'**effet de bord**, celui-ci est potentiellement encore présent pour certains indicateurs comme le relief. En outre, la définition de la tache urbaine élaborée par Geoclimate via un algorithme est améliorable, il a encore tendance à empiéter sur la végétation autour de la tache urbaine.

De plus, la méthode ne prend pas en compte certains **indicateurs** comme l'albédo et l'inertie thermique par manque de données fiables et complètes. Pour étudier l'albédo, il serait possible d'étudier des données satellitaires. Une autre limite est le nombre de villes disponibles dans l'étude MApUCE pour la comparaison entre la classification et la modélisation d'ICU. Des **sources de données additionnelles** pour être inclus dans l'analyse pour valider les résultats de la classification. Par exemple, la métropole de Dijon a implanté un réseau de 67 capteurs de température : MUSTARDijon. La métropole de Rennes a aussi construit son propre réseau de capteurs dès 2003 dans le cadre du projet Rennes Urban Network (RUN), en 2020 le réseau comptait 92 capteurs.

Par ailleurs, il serait intéressant de **changer d'échelle** d'étude et d'élargir l'analyse à d'autres pays. Il serait potentiellement possible de mener une étude à l'échelle de l'Europe en étant capable de comparer avec une modélisation d'ICU à l'échelle européenne. En ce sens, le projet Urb Clim modélise les ICU sur 100 villes européennes. L'analyse d'autres pays où la morphologie urbaine est complètement différente permettrait d'étudier **d'autres profils microclimatiques** et **d'autres schémas urbains**. Par exemple, les pays au Moyen-Orient font face au phénomène de chaleur et ils ont une morphologie urbaine et un microclimat différents.

De même, il serait pertinent de passer à l'échelle des aires urbaines pour voir au-delà des limites de la ville, pour mieux appréhender les continuités urbaines en dehors de la limite administrative.

6.4 Point de vue prospectif

Pour continuer cette étude, une doctorante, Lenaig LE GROGNEC, a été recrutée au sein du laboratoire pour débuter sa thèse en novembre. Son travail a pour objectif de **compléter l'approche du stage**. Cette **thèse** vise à mettre en place de **modèle prédictif** d'ICU pour évaluer **des scénarios d'aménagement de l'espace**.

Conclusion

Pour conclure, ce stage m'a permis de **classifier les villes françaises** selon des **indicateurs du microclimat**, établis à partir d'une **analyse bibliographique**. Cette classification **automatisée** a permis d'identifier **un gradient de sensibilité à la surchauffe urbaine**.

Ce stage a été l'occasion de **mettre en pratique** les **compétences** acquises tout au long de mon parcours scolaire et en particulier celles développées au sein du master SIGAT. J'ai pu approfondir mes connaissances en **statistiques, programmation, gestion de projet**, tout en apportant mon expertise en **analyse spatiale et en géographie**. Par ailleurs, j'ai eu l'**opportunité** de découvrir l'univers de la recherche sous un angle opérationnel.

Enfin, je voudrais finir ce rapport en **remerciant** de nouveaux toutes les personnes qui m'ont accompagnées lors de ce stage : Florent DEMORAES, Baptiste ALGLAVE, Elisabeth LE SAUX WIEDERHOLD, Erwan BOCHER, Matthieu GOUSSEFF, François LECONTE, Jeremy BERNARD, Lise BELLANGER et Jessica PIC.

Bibliographie

Bernard, J., Bocher, E., Gousseff, M., Leconte, F., & Le Saux Wiederhold, E. (2024). A generic algorithm to automatically classify fabrics according to the local climate zone system: Implementation in GeoClimate 0.0.1 and application to French cities. *Geoscientific Model Development*, 17, 2077-2116. <https://doi.org/10.5194/gmd-17-2077-2024>

Bernard, J., Bocher, E., Le Saux Wiederhold, E., Leconte, F., and Masson, V.: Estimation of missing building height in OpenStreetMap data: a French case study using GeoClimate 0.0.1, *Geosci. Model Dev.*, 15, 7505–7532, <https://doi.org/10.5194/gmd-15-7505-2022>, 2022.

Bernard, J., Musy, M., Calmet, I., Bocher, E., & Keravec, P. (2017). Urban heat island temporal and spatial variations: Empirical modeling from geographical and meteorological data. *Building and Environment*, 125, 423-438. <https://doi.org/10.1016/j.buildenv.2017.08.009>

Bocher, E., Bernard, J., Le Saux Wiederhold, E., Leconte, F., Petit, G., et al. (2021). GeoClimate: A Geospatial processing toolbox for environmental and climate studies. *Journal of Open Source Software*, 6(65), 3541. <https://doi.org/10.21105/joss.03541>

Bocher, E., Gousseff, M., Bernard, J., Le Saux Wiederhold, E., Alglave, B., & Kerjouan, E. (2024). Comparison of different methods to produce local climate zone maps using the LczExplore tool. *EGU General Assembly 2024*, Vienna, Austria, 14–19 Apr 2024, EGU24-15984. <https://doi.org/10.5194/egusphere-egu24-15984>

Bowler, D. E., Buyung-Ali, L., Knight, T. M., & Pullin, A. S. (2010). Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning*, 97(3), 147-155.

Crétat, J., Richard, Y., Planchon, O., Emery, J., Poupelin, M., Rega, M., Pergaud, J., Joly, D., Diallo-Dudek, J., Roy, D., Granjon, L., & Pohl, B. (2023). Impact de la topographie et de la circulation atmosphérique sur l'îlot de chaleur urbain en situation de canicule (Dijon, France). *Climatologie*, 20. <https://doi.org/10.1051/climat/202320010>

FranceInfo. (2021). Paris, Lille, Toulouse, Lyon, Grenoble, Bordeaux... Visualisez l'îlot de chaleur dans votre zone urbaine. Disponible sur : https://www.francetvinfo.fr/meteo/canicule/carte-paris-lille-toulouse-lyon-grenoble-bordeaux-visualisez-l-ilot-de-chaleur-dans-votre-zone-urbaine_4719121.html

Hove, L.W.A. van, Steeneveld, G.J., Jacobs, C.M.J., Heusinkveld, B.G., Elbers, J.A., Moors, E.J., & Holtslag, A.A.M. (2011). Exploring the urban heat island intensity of

Dutch cities: Assessment based on a literature review, recent meteorological observation, and datasets provided by hobby meteorologists. Alterra Report 2170, Wageningen: Alterra. <https://library.wur.nl/WebQuery/wurpubs/406819>

Institut Paris Région. (2010, Novembre). Les îlots de chaleur urbains. Répertoire de fiches connaissances. Disponible sur: https://www.institutparisregion.fr/fileadmin/NewEtudes/Etude_774/Les_ilots_de_chaleur_urbains_REPERTOIRE.pdf

Kim, S. W., & Brown, R. D. (2021). Urban heat island. *Renewable and Sustainable Energy Reviews*, 148, 111256. <https://doi.org/10.1016/j.rser.2021.111256>

Lab-STICC. (n.d.). Projet Lab-STICC : Activités et projets. <https://labsticc.fr/fr/equipes/decide/activites/projets>

Lauwaet, D., Berckmans, J., Hooyberghs, H., Wouters, H., Driesen, G., Lefebre, F., & De Ridder, K. (2024). High resolution modelling of the urban heat island of 100 European cities. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2024.106209>

Magali Reghezza-Zitt. (2023, Avril). Sociétés humaines et territoires dans un climat qui change. Du réchauffement climatique global aux politiques climatiques. Géoconfluences.

Marcotullio, P. J., Keßler, C., & Fekete, B. M. (2021). The future urban heat-wave challenge in Africa: Exploratory analysis. *Global Environmental Change*, 66, 102190. <https://doi.org/10.1016/j.gloenvcha.2020.102190>

Massicotte, P., & South, A. (2024). rnaturalearth: World Map Data from Natural Earth. R package version 1.0.1.9000, <https://github.com/ropensci/rnaturalearth>, <https://docs.ropensci.org/rnaturalearthhires/>, <https://docs.ropensci.org/rnaturalearth/>.

Masson, V., Hidalgo, J., Bocher, E., Bretagne, G., Touchard, O., Haouès-Jouve, S., Bernard, J., Pilot, L., Gardes, T., Goret, M., Machado, T., Mhedhbi, Z., Petit, G., Schoetter, R., & Touati, N. (2020). PÆNDORA Planification, Adaptation et Énergie : Données territoriales et accompagnement. Rapport scientifique. http://www.umr-cnrm.fr/ville.climat/IMG/pdf/rapport_final_paendora_v1.pdf

Météo France. (2023). Vagues de chaleur et changement climatique. Disponible sur : <https://meteofrance.com/changement-climatique/observer/changement-climatique-et-vagues-de-chaleur>

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108(455), 1-24.

Oke, T. R. (2006). Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites. World Meteorological Organization (WMO/TD-No. 1250).

Pachauri, R. K., & Meyer, L. (Eds.). (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC.

Phelan, P. E., Kaloush, K., Miner, M., Golden, J., Phelan, B., Silva III, H., & Taylor, R. A. (2015). Urban Heat Island: Mechanisms, Implications, and Possible Remedies. *Annual Review of Environment and Resources*, 40, 285-307.
<https://doi.org/10.1146/annurev-environ-102014-021155>

Rennes Climate Urban Network. (2022). Suivi en temps réel de l'îlot de chaleur urbain, projet RUN. Disponible sur: <https://run.letg.cnrs.fr/le-projet.html>

Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Science*, 5(1-4).
http://download.osgeo.org/qgis/doc/reference-docs/Terrain_Ruggedness_Index.pdf

Santamouris, M. (2014). Cooling the cities: A review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments. *Solar Energy*, 103, 682-703.
<https://doi.org/10.1016/j.solener.2012.07.003>

Stewart, I. D., & Oke, T. R. (2012). Local Climate Zones for Urban Temperature Studies. *Bulletin of the American Meteorological Society*, 93, 1879–1900.
<https://doi.org/10.1175/BAMS-D-11-00019.1>

Wang, R., Wang, M., Ren, C., Chen, G., Mills, G., & Ching, J. (2024). Mapping local climate zones and its applications at the global scale: A systematic review of the last decade of progress and trend. *Urban Climate*, 57, 102129.

Xue, L., Ming, Y., Liu, Y., Yue, W., & Han, G. (2022). Influences of landform and urban form factors on urban heat island: Comparative case study between Chengdu and Chongqing. *Science of the Total Environment*.
<https://doi.org/10.1016/j.scitotenv.2022.106209>

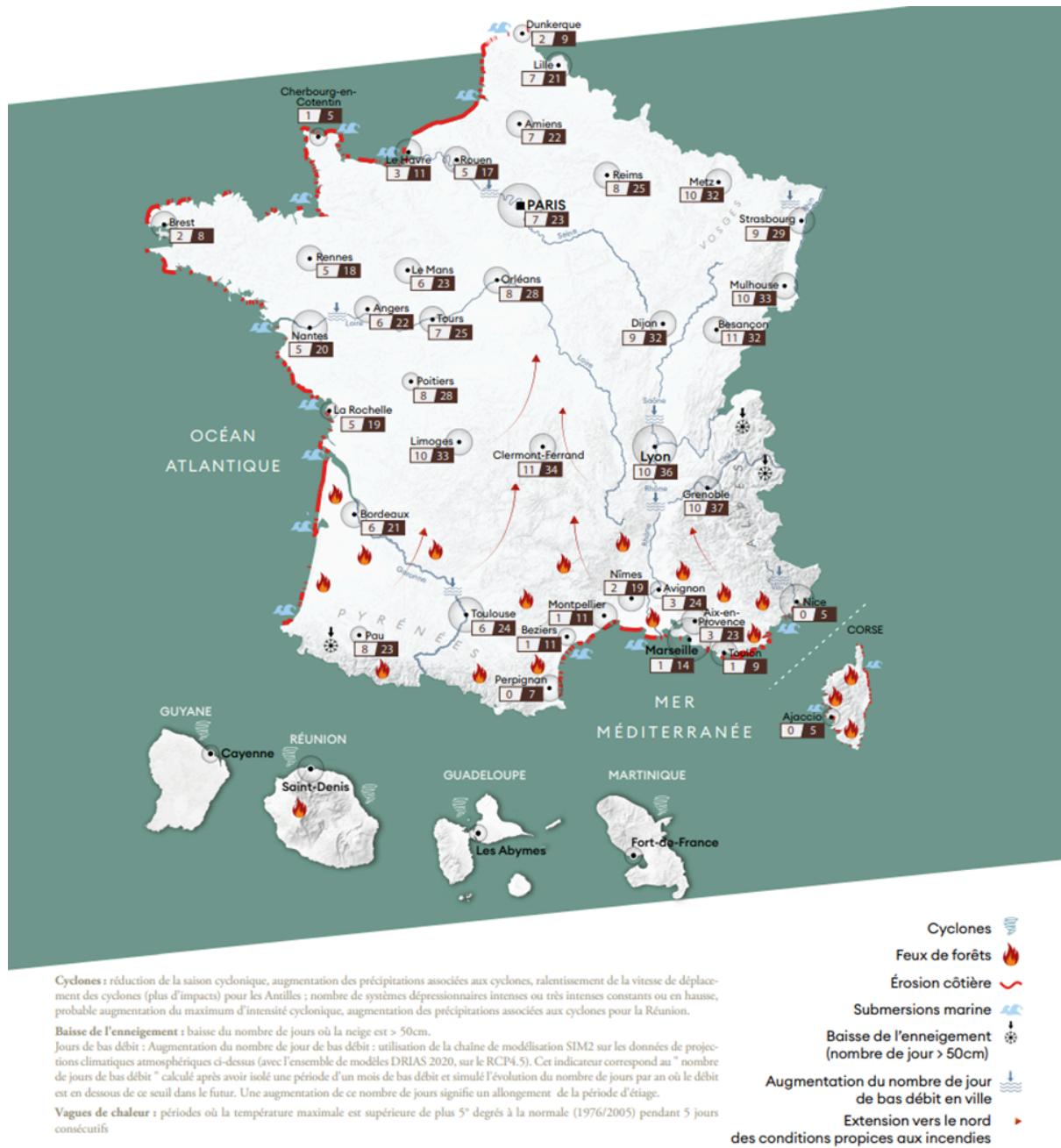
Tableau des tables et des figures

| | |
|---|----|
| Figure 1 : Planning effectif..... | 8 |
| Figure 2 : Organigramme de l'équipe de Vannes..... | 10 |
| Figure 3 : Outils utilisés pendant le stage..... | 10 |
| Figure 4 : Schématisation du phénomène d'îlot de chaleur urbain (Cerema)..... | 11 |
| Figure 5 : Entrées et des sorties de Geoclimate..... | 13 |
| Figure 6 : Chaîne de traitements simplifié de Geoclimate pour la construction des ZCL..... | 14 |
| Figure 7 : Les déterminants renforçant le phénomène d'ICU..... | 15 |
| Figure 8 : Tableau des indicateurs à mettre en place..... | 18 |
| Figure 9 : Aires urbaines étudiées par MApUCE..... | 22 |
| Figure 10 : Exemple de représentation de l'ICU pour l'aire urbaine de Clermont-Ferrand (FranceInfo, 2021)..... | 22 |
| Figure 11 : Chaîne de traitement pour la construction et l'analyse des classifications des villes..... | 23 |
| Figure 12 : Exemple de graphique permettant l'exploration de données..... | 24 |
| Figure 13 : Visualisation des variables quantitatives dans les deux premières dimensions de l'AFDM..... | 25 |
| Figure 14 : Choix du nombre de groupes..... | 26 |
| Figure 15 : Nombre d'individus par groupes allant du moins au plus sensible..... | 27 |
| Figure 16 : Visualisation des individus sur les 2 premières dimensions de l'AFDM, le nom des communes représente les individus les plus contributifs..... | 27 |
| Figure 17 : Spatialisation des profil de villes, focus sur les villes du groupe 1 impactées par le relief..... | 28 |
| Figure 18 : Localisation des villes appartenant au profil 6..... | 29 |
| Tableau 1 : Tableau des moyennes pour les six groupes avec des variables sélectionnées et arrondies..... | 29 |
| Figure 19 : Cartographie des ZCL suivant leur sensibilité..... | 30 |
| Figure 20 : Visualisation des variables quantitatives qui contribuent aux dimensions. | 31 |
| Figure 21 : Choix du nombre de groupes..... | 31 |
| Figure 22 : Nombre d'individus par groupes allant du moins au plus sensible..... | 32 |
| Figure 23 : Visualisation des individus sur les 2 premières dimensions de l'AFDM, le nom des communes représente les individus les plus contributifs..... | 32 |
| Figure 24 : Spatialisation des différents profils de villes, spatialisation marquée pour les villes du groupe 1..... | 33 |
| Tableau 2 : Tableau des moyennes pour les cinq groupes avec des variables sélectionnées et arrondies..... | 33 |

| | |
|--|----|
| Figure 25 : Représentation des ZCL pour les villes d'exemple suivant leur sensibilité... | 34 |
| Figure 26 : Graphique de la moyenne de l'intensité des ICU suivant les données de MApUCE..... | 35 |
| Figure 27 : Gradient de la sensibilité des villes à la surchauffe urbaines suivant les indicateurs mis en place..... | 36 |

Annexes

Annexe 1 : Augmentation du nombre de vagues de chaleur



Haut Conseil du Climat. (2021) / Reghezza-Zitt, (2023)

Annexe 2 : Planning prévisionnel

PLANNING PRÉVISIONNEL



Annexe 3 : Description des ZCL par Oke

TABLE 2. Abridged definitions for local climate zones (see electronic supplement for photographs, surface property values, and full definitions). LCZs 1–9 correspond to Oke's (2004) urban climate zones.

| Built types | Definition | Land cover types | Definition |
|-------------------------|--|---------------------------------------|--|
| 1. Compact high-rise | Dense mix of tall buildings to tens of stories. Few or no trees. Land cover mostly paved. Concrete, steel, stone, and glass construction materials. | A. Dense trees | Heavily wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park. |
| 2. Compact midrise | Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials. | B. Scattered trees | Lightly wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park. |
| 3. Compact low-rise | Dense mix of low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials. | C. Bush, scrub | Open arrangement of bushes, shrubs and short, woody trees. Land cover mostly pervious (bare soil or sand). Zone function is natural scrubland or agriculture. |
| 4. Open high-rise | Open arrangement of tall buildings to tens of stories. Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials. | D. Low plants | Featureless landscape of grass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park. |
| 5. Open midrise | Open arrangement of midrise buildings (3–9 stories). Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials. | E. Bare rock or paved | Featureless landscape of rock or paved cover. Few or no trees or plants. Zone function is natural desert (rock) or urban transportation. |
| 6. Open low-rise | Open arrangement of low-rise buildings (1–3 stories). Abundance of pervious land cover (low plants, scattered trees). Wood, brick, stone, tile, and concrete construction materials. | F. Bare soil or sand | Featureless landscape of soil or sand cover. Few or no trees or plants. Zone function is natural desert or agriculture. |
| 7. Lightweight low-rise | Dense mix of single-story buildings. Few or no trees. Land cover mostly hard-packed. Lightweight construction materials (e.g., wood, thatch, corrugated metal). | G. Water | Large, open water bodies such as seas and lakes, or small bodies such as rivers, reservoirs, and lagoons. |
| 8. Large low-rise | Open arrangement of large low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Steel, concrete, metal, and stone construction materials. | VARIABLE LAND COVER PROPERTIES | |
| 9. Sparsely built | Sparse arrangement of small or medium-sized buildings in a natural setting. Abundance of pervious land cover (low plants, scattered trees). | b. bare trees | Leafless deciduous trees (e.g., winter). Increased sky view factor. Reduced albedo. |
| 10. Heavy industry | Low-rise and midrise industrial structures (towers, tanks, stacks). Few or no trees. Land cover mostly paved or hard-packed. Metal, steel, and concrete construction materials. | s. snow cover | Snow cover >10 cm in depth. Low admittance. High albedo. |
| | | d. dry ground | Parched soil. Low admittance. Large Bowen ratio. Increased albedo. |
| | | w. wet ground | Waterlogged soil. High admittance. Small Bowen ratio. Reduced albedo. |

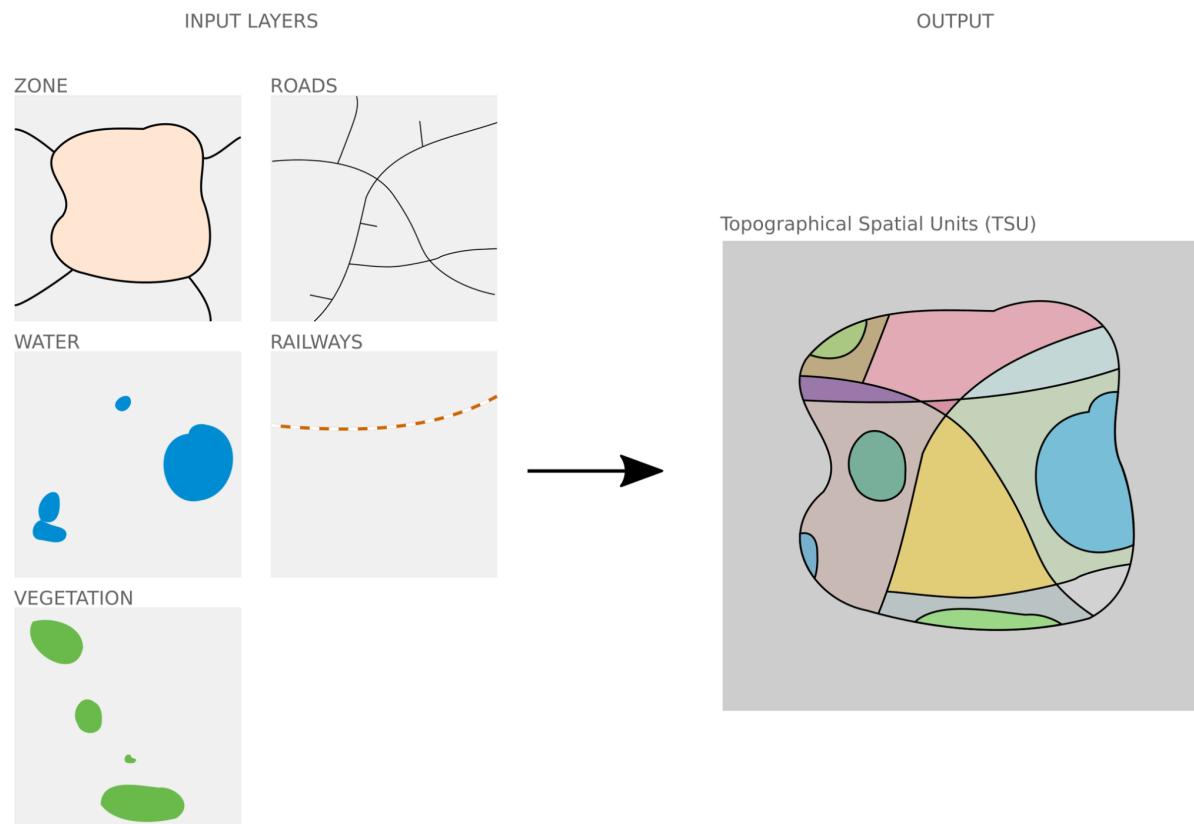
Stewart, I. D., and Oke, T. R. (2012)

Annexe 4 : Les 14 indicateurs pour le calcul des ZCL, et la description de la méthode de calcul pour chaque indicateur.

| Indicator abbreviation | Indicator name in GeoClimate library (or combination of GeoClimate indicators) | Calculation method ^a | Defined in Stewart and Oke (2012) |
|------------------------|---|---|-----------------------------------|
| SVF | GROUND_SKY_VIEW_FACTOR | The calculation is based on the [ST_SVF] (http://www.h2gis.org/docs/dev/ST_SVF/ , last access: 11 March 2024) function of H2GIS using only buildings as obstacles and with the following parameters: ray length = 100 and number of directions = 60. Using a uniform grid mesh of 10 m resolution, the SVF obtained has a standard deviation of the estimate of 0.03 when compared with the most accurate method (according to Bernard et al., 2018). Using a grid with regular points, the density of points used for the calculation actually depends on building density (the higher the building density, the lower the density of points). To avoid this phenomenon and have the same density of points per free ground surface, we use a H2GIS function to randomly distribute points within free surfaces (ST_GeneratePoints). This density of points is set by default to 0.008 (based on the median of Bernard et al., 2018, dataset). | Yes |
| H/W | ASPECT_RATIO | A simple approach based on the street canyon assumption is used for the calculation. $\frac{\sum_i A_{B_i}}{A_{RSU} - \sum_i A_{B_i}}$, where n is the number of buildings in the RSU, A_{B_i} is the facade area of the building i , A_{B_i} is the area of building i , and A_{RSU} is the area of the RSU. | Yes |
| F _B | BUILDING_FRACTION_LCZ | $\frac{\sum_i A_{B_i}}{A_{RSU}}$, where n is the number of buildings in the RSU, A_{B_i} is the area of building i , and A_{RSU} is the area of the RSU. | Yes |
| F _I | IMPERVIOUS_FRACTION_LCZ | $\frac{\sum_i A_{I_i}}{A_{RSU}}$, where n is the number of impervious patches in the RSU (geometries from road and impervious tables), A_{I_i} is the area of impervious patch i , and A_{RSU} is the area of the RSU. | Yes |
| F _P | PERVIOUS_FRACTION_LCZ | $\frac{\sum_i A_{P_i}}{A_{RSU}}$, where n is the number of pervious patches in the RSU (geometries from water and vegetation tables not intersecting with any geometry from impervious, road, or building tables), A_{P_i} is the area of pervious patch i , and A_{RSU} is the area of the RSU. | Yes |
| H _r | GEOM_AVG_HEIGHT_ROOF ^b | $\exp\left(\frac{\sum_i \log(H_{B_i})}{n}\right)$, where n is the number of buildings in the RSU, and H_{B_i} is the height of building i . | Yes |
| z_0 | EFFECTIVE_TERRAIN_ROUGHNESS_LENGTH ^c | Based on the Hanna and Britter (2010) procedure (see Eq. (17) and examples of calculation on p. 156 of the corresponding reference). If $\lambda_f \leq 0.15 : \lambda_f \cdot H_r$, If $\lambda_f > 0.15 : 0.15 \cdot H_r$, where H_r is the geometric average height roof within the RSU, $\lambda_f = \frac{\sum_j \text{ndir} \sum_i A_{B_{i,j}}}{\text{ndir} \cdot A_{RSU}}$, $A_{B_{i,j}}$ is the area of building facade i when projected in the j direction, ndir is the number of directions used to calculate the projected facade area (default 12), and A_{RSU} is the area of the RSU. Warning: the calculation is only performed for angles included in the range [0, 180°] (considered equal for a given orientation independent of the direction). When the RSUs do not split buildings, then this calculation is right but can slightly overestimate the results when buildings are split (there is actually an overestimation only in one direction). | Yes |
| Indicator abbreviation | Indicator name in GeoClimate library (or combination of GeoClimate indicators) | Calculation method ^d | Defined in Stewart and Oke (2012) |
| F _{AV} | HIGH_VEGETATION_FRACTION_LCZ + LOW_VEGETATION_FRACTION_LCZ | $\frac{\sum_i A_{V_i}}{A_{RSU}}$, where n is the number of all vegetation (low and high) patches in the RSU, A_{V_i} is the area of vegetation patch i , and A_{RSU} is the area of the RSU. | No |
| F _W | WATER_FRACTION_LCZ | $\frac{\sum_i A_{W_i}}{A_{RSU}}$, where n is the number of water patches in the RSU, A_{W_i} is the area of water patch i , and A_{RSU} is the area of the RSU. | No |
| F _{HV/AV} | HIGH_VEGETATION_FRACTION_LCZ / (HIGH_VEGETATION_FRACTION_LCZ + LOW_VEGETATION_FRACTION_LCZ) | $\frac{\sum_i A_{HV_i}}{\sum_i A_{V_i}}$, where n_{HV} is the number of high vegetation patches in the RSU, A_{HV_i} is the area of high vegetation patch i , n is the number of all vegetation (low and high) patches in the RSU, and A_{V_i} is the area of vegetation patch i . | No |
| F _{IND/B} | AREA_FRACTION_HEAVY_INDUSTRY / BUILDING_FRACTION | $\frac{\sum_i n_{BIND} A_{BIND_i}}{\sum_i A_{B_i}}$, where n_{BIND} is the number of industrial buildings in the RSU (buildings having TYPE = "HEAVY INDUSTRY"), A_{BIND_i} is the area of the industrial building i , n is the number of buildings in the RSU, and A_{B_i} is the area of the building i . | No |
| F _{LLR/B} | (AREA_FRACTION_COMMERCIAL) + AREA_FRACTION_LIGHT_INDUSTRY / BUILDING_FRACTION | $\frac{\sum_i n_{BLLR} A_{BLLR_i}}{\sum_i A_{B_i}}$, where n_{BLLR} is the number of large low-rise buildings in the RSU (buildings having TYPE = "COMMERCIAL" or TYPE = "LIGHT INDUSTRY"), A_{BLLR_i} is the area of the large low-rise building i , n is the number of buildings in the RSU, and A_{B_i} is the area of the building i . | No |
| F _{RES/B} | AREA_FRACTION_RESIDENTIAL / BUILDING_FRACTION | $\frac{\sum_i n_{BRES} A_{BRES_i}}{\sum_i A_{B_i}}$, where n_{BRES} is the number of residential buildings in the RSU (buildings having TYPE = RESIDENTIAL), A_{BRES_i} is the area of the residential building i , n is the number of buildings in the RSU, and A_{B_i} is the area of the building i . | No |
| N _{lev} | AVG_NB_LEV_AREA_WEIGHTED | $\frac{\sum_i A_{B_i} \cdot Nlev_{B_i}}{\sum_i A_{B_i}}$, where n is the number of buildings in the RSU, A_{B_i} is the area of the building i , and $Nlev_{B_i}$ is the number of levels of building i . | No |

Bernard et al. 2024.

Annexe 5 : Exemple de calcul d'unité spatiale topographique.



Bernard et al. 2022

Annexe 6 : Résultats sous forme de fichier Rmardown pour les grandes villes françaises

AFDM Clustering Analysis - Large French cities

Pauline Besnard - Alglaive Baptiste

2024-10-02

Initialize the environment

```
#Library
library(cowplot)
library(dplyr)
library(FactoMineR)
library(factoextra)
library(ggplot2)
library(ggspatial)
library(kableExtra)
library(knitr)
library(lczexplore)
library(maps)
library(plotly)
library(rnaturalearth)
library(sf)
library(tidygeocoder)
library(tidyr)

# Settings
selected_cities <- "insee_sup"
min_pop <- 50000
nb_class <- 5

# Load the data
if (selected_cities == "pandora") {
  file_name <- paste0("res/build_analysis/analysis_df_clean/analysis_df_famd_",
                      selected_cities, ".csv")
} else {
  file_name <- paste0("res/build_analysis/analysis_df_clean/analysis_df_famd_",
                      min_pop, "_", selected_cities, ".csv")
}
analysis_df_famd <- read.csv(file_name)
analysis_df_famdcity <- read.csv(file_name)
# Delete row who is cities probleme
analysis_df_famd <- analysis_df_famd %>%
  filter(city != "Montreuil" & city != "Chelles" & city != "Évry"
        & city != "Saint-Louis" & city != "Saint-Paul"
        & city != "Saint-Joseph" & city != "Valence")

# Add row names as a column in each table and delete the columns
rownames(analysis_df_famd) <- analysis_df_famd$city
analysis_df_famd <- analysis_df_famd %>% dplyr::select(-city)

# Colors for the clusters
colors <- c("#006400",
            "#7CFC00",
            "#FFFF00",
            "#FFA500",
            "#FF0000",
            "#8B0000")
```

Factorial Analysis of Mixed Data (FAMD)

```

famd <- FAMD(analysis_df_famd, graph = FALSE, ncp = 5)

# Save
dir.create("res/clustering_famd", showWarnings = FALSE, recursive = TRUE)
file_name <- paste0("res/clustering_famd/analysis_famd_summary_",
                     min_pop, selected_cities, ".txt")
sink(file_name)
summary(famd)
sink()

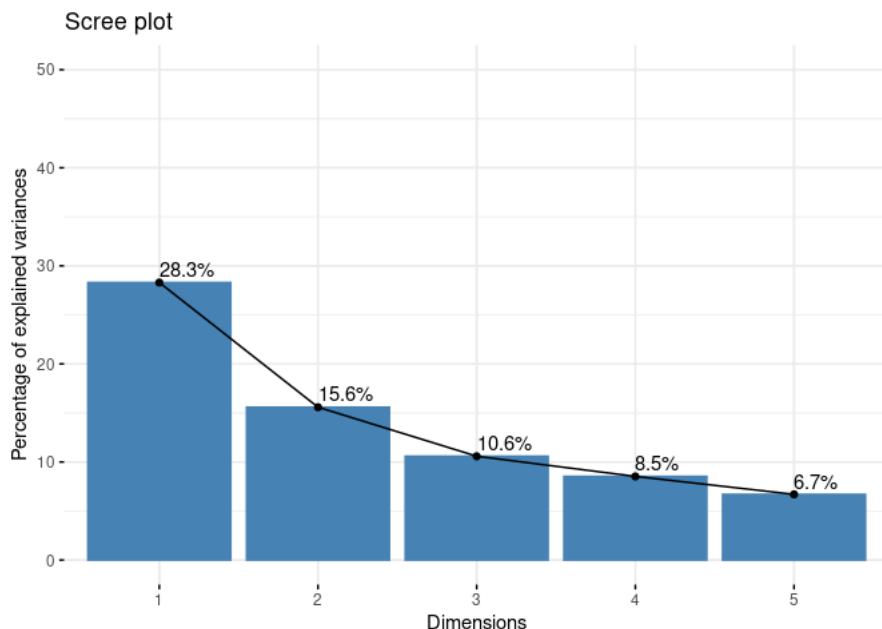
```

View and explore the results

```

# Visualization of variance and contribution of variables
fviz_screeplot(famd, addlabels = TRUE, ylim = c(0, 50))

```

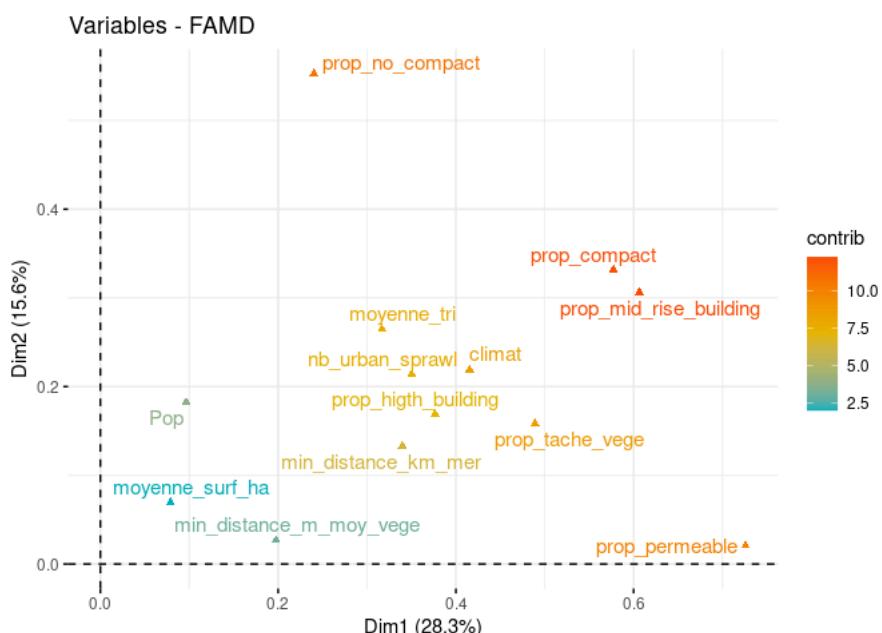


```

# Choose the number of components
ncp <- 2
famd <- FAMD(analysis_df_famd, graph = FALSE, ncp = ncp)

fviz_famd_var(famd, repel = TRUE, col.var = "contrib",
               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))

```

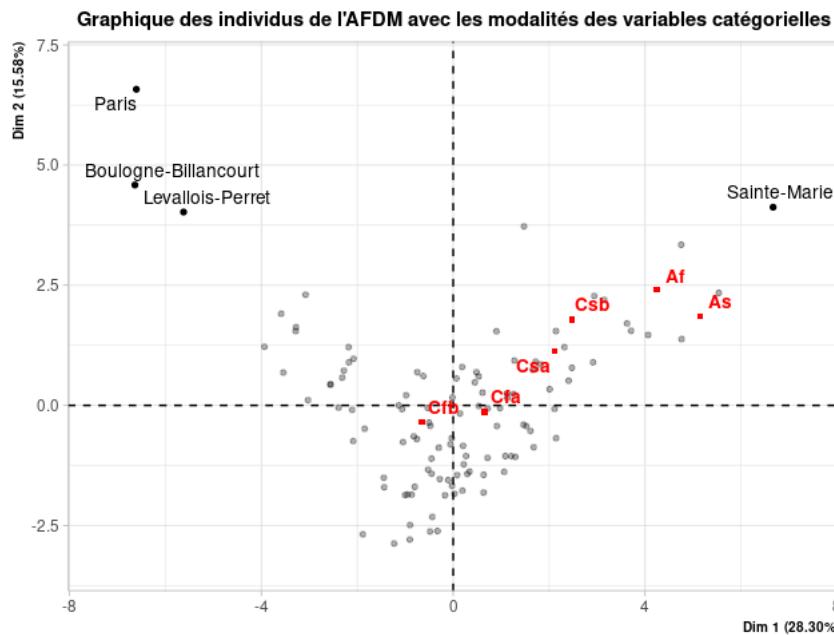


```

# Plot 3% of the most contributive individuals
contrib <- transform(data.frame(famd$ind$contrib), contribPlan1 = Dim.1 + Dim.2)
selected_individuals <- rownames(contrib)[contrib$contribPlan1 >= quantile(contrib$contribP

# Graph of AFDM individuals with the modalities of the categorical variables
plot.FAMD(famd,
           title = "Graphique des individus de l'AFDM avec les modalités des variables catég",
           cex = 0.9, cex.main = 1,
           graph.type = "ggplot",
           cex.axis = 0.75,
           label = c('ind','quali.var'),
           select = selected_individuals)

```

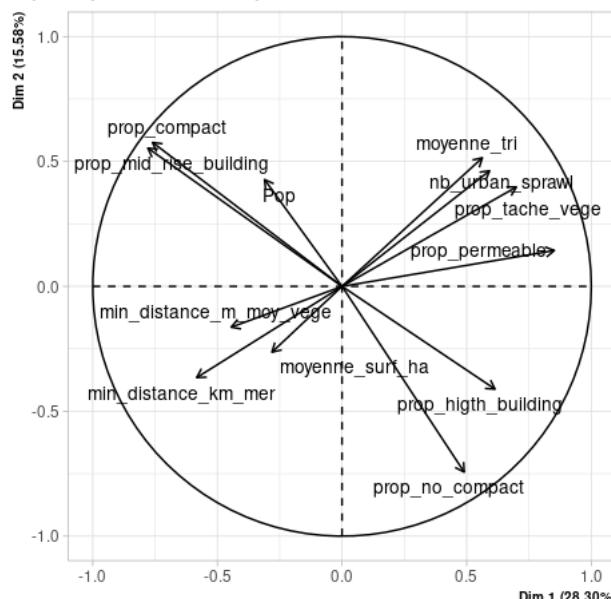


```

# Graphs of quantitative and categorical modalities
plot.FAMD(famd, choix = "quanti",
           title = "Graphique de la participations des champs de l'AFDM avec les modalités d",
           cex = 0.9, cex.main = 1,
           graph.type = "ggplot",
           cex.axis = 0.75)

```

Graphique de la participations des champs de l'AFDM avec les modalités des variables quantita

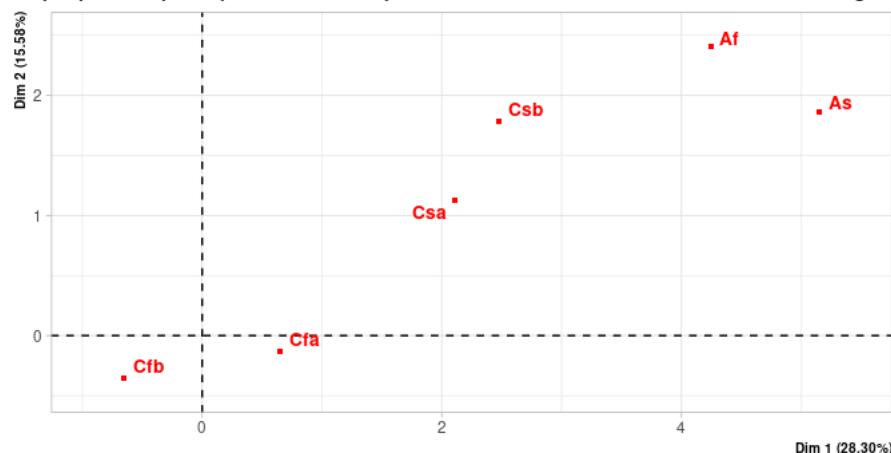


```

plot.FAMD(famd, choix = "quali",
          title = "Graphique de la participations des champs de l'AFDM avec les modalités des variables catégorielles",
          cex = 0.9, cex.main = 1,
          graph.type = "ggplot",
          cex.axis = 0.75)

```

Graphique de la participations des champs de l'AFDM avec les modalités des variables catégorielles

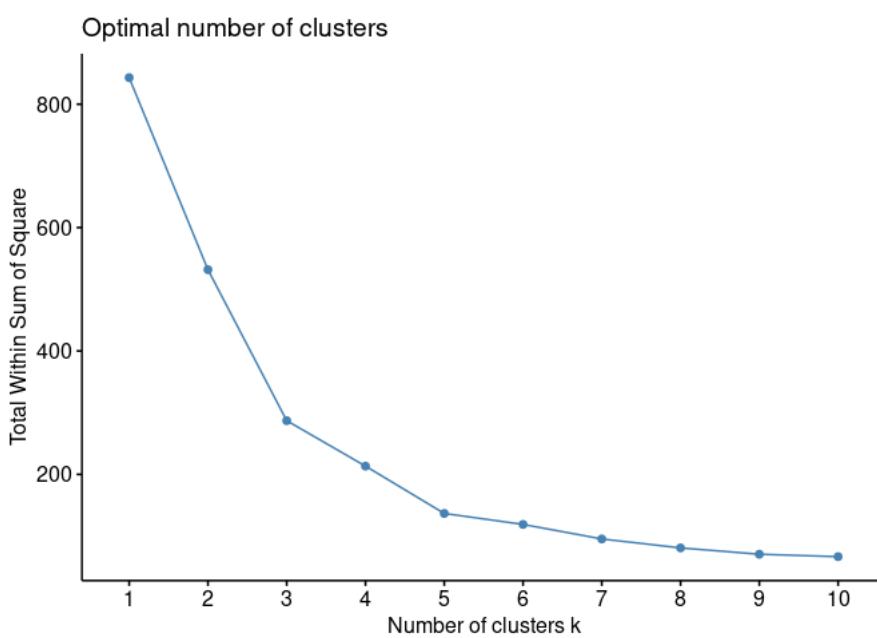


K-means clustering

```

ind_coord <- famd$ind$coord
fviz_nbclust(ind_coord, kmeans, method = "wss")

```



```

# Choose the number of clusters
# nb_class <- 6
km.res <- kmeans(ind_coord, centers = nb_class, nstart = 50)
# Add the cluster column to the data frame
analysis_df_famd$cluster <- as.factor(km.res$cluster)

# Reattribute the clusters to the cities
recode_clusters <- c('3' = '1', '4' = '2', '5' = '3', '1' = '4', '2' = '5')

# Apply this correspondence to the 'cluster' column in km.res

```

```

km.res$cluster <- recode(as.character(km.res$cluster), !!!recode_clusters)

# Convert the re-assigned clusters to an ordered factor
km.res$cluster <- factor(km.res$cluster, levels = c('1', '2', '3', '4', '5'), ordered = TRUE)

# Apply this correspondence to the 'cluster' column in analysis_df_fam
analysis_df_fam$cluster<- recode(analysis_df_fam$cluster, !!!recode_clusters)

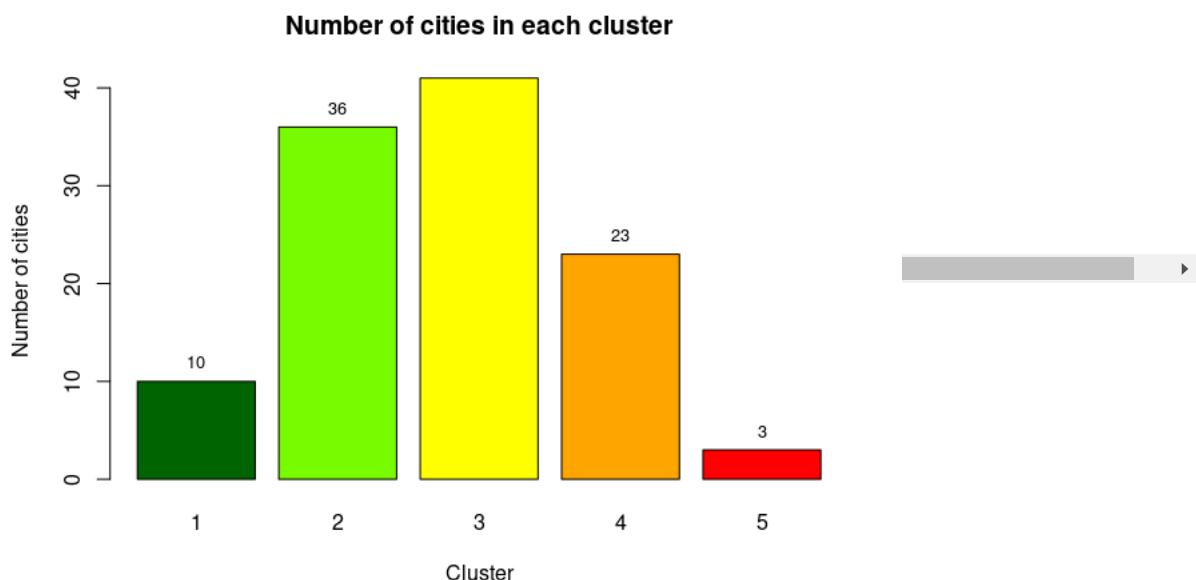
# Convert the re-assigned clusters to an ordered factor
analysis_df_fam$cluster <- factor(analysis_df_fam$cluster,
                                    levels = c('1', '2', '3', '4', '5'),
                                    ordered = TRUE)

# How many cities in each cluster ?
cities_in_cluster <- table(analysis_df_fam$cluster)

cities_in_cluster_df <- as.data.frame(cities_in_cluster)
colnames(cities_in_cluster_df) <- c("Cluster", "Frequency")

# Barplot of the number of cities in each cluster
bar_positions <- barplot(cities_in_cluster,
                           col = colors,
                           main = "Number of cities in each cluster",
                           xlab = "Cluster",
                           ylab = "Number of cities")
text(x = bar_positions,
      y = cities_in_cluster,
      labels = cities_in_cluster,
      pos = 3,
      cex = 0.8,
      col = "black")

```



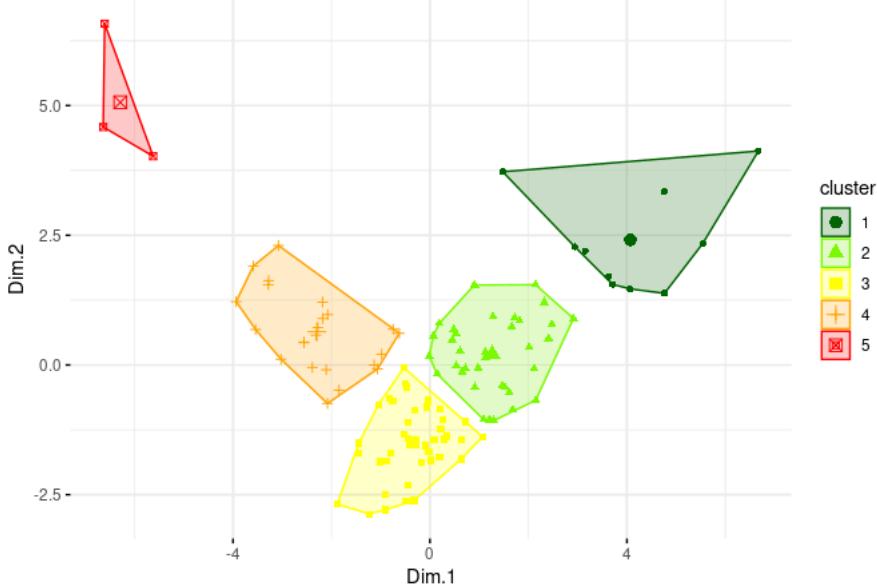
Visualize clusters

```

fviz_cluster(km.res, data = ind_coord,
             geom = "point",
             stand = FALSE,
             ellipse.type = "convex",
             palette = colors,
             ggtheme = theme_minimal()
)

```

Cluster plot



```

if (selected_cities == "pandora") {
  file_name <- paste0("res/build_analysis/analysis_df_clean/analysis_df_famdsans_log_", se
} else {
  file_name <- paste0("res/build_analysis/analysis_df_clean/analysis_df_famdsans_log_", mi
}
analysis_df_var_clust_moy <- read.csv(file_name)

# add city
analysis_df_var_clust_moy$city <- analysis_df_famdcity$city

# Filter the cities with problems
analysis_df_var_clust_moy <- analysis_df_var_clust_moy %>%
  filter(city != "Montreuil" & city != "Chelles" & city != "Évry" & city != "Saint-Louis" &

## add contribution to the cluster
contribution <- get_famd_ind(famd)$contrib

# Add the contribution to the first two axes
analysis_df_var_clust_moy$contrib_to_axis1 <- contrib[, 1] # Axe 1 contribution
analysis_df_var_clust_moy$contrib_to_axis2 <- contrib[, 2]

# Add the cluster column
analysis_df_var_clust_moy$cluster <- as.factor(km.res$cluster)
# rownames by city
rownames(analysis_df_var_clust_moy) <- analysis_df_var_clust_moy$city

# Table des 5 villes les plus représentatives (contributives) dans chaque cluster
tableau <- data.frame()
for (i in 1:nb_class) {
  # Trier par contribution décroissante et sélectionner les 5 premières
  cluster_data <- analysis_df_var_clust_moy %>%
    filter(cluster == i) %>%
    arrange(desc(contrib_to_axis1 + contrib_to_axis2)) # Trier par la somme des contributions

  tableau <- rbind(tableau, head(cluster_data, 5))
}

# Formate and color the table
tableau_with_cluster <- tableau %>%
  mutate(cluster = factor(cluster)) %>%
  mutate(across(where(is.numeric), round, 3))

kable_table <- kable(tableau_with_cluster, format = "html", row.names = TRUE) %>%
  row_spec(which(tableau_with_cluster$cluster == 1), background = colors[1]) %>%

```

```

row_spec(which(tableau_with_cluster$cluster == 2), background = colors[2]) %>%
row_spec(which(tableau_with_cluster$cluster == 3), background = colors[3]) %>%
row_spec(which(tableau_with_cluster$cluster == 4), background = colors[4]) %>%
row_spec(which(tableau_with_cluster$cluster == 5), background = colors[5]) %>%
row_spec(which(tableau_with_cluster$cluster == 6), background = colors[6]) %>%
row_spec(which(tableau_with_cluster$cluster == 7), background = colors[7]) %>%
row_spec(which(tableau_with_cluster$cluster == 8), background = colors[8])

```

kable_table

| | Pop | min_distance_km_mer | climat | moyenne_tri | prop_permeable | min_distance |
|-----------------------|---------|---------------------|--------|-------------|----------------|--------------|
| Sainte-Marie | 57443 | 3.262 | Af | 1.841 | 0.671 | |
| Ajaccio | 73822 | 1.143 | Csb | 1.153 | 0.589 | |
| Le Tampon | 81943 | 13.443 | As | 1.593 | 0.841 | |
| Nice | 348085 | 1.490 | Csb | 1.467 | 0.427 | |
| Saint-André | 61465 | 3.608 | As | 0.668 | 0.761 | |
| Fort-de-France | 74921 | 3.788 | Af | 1.986 | 0.601 | |
| Béziers | 80341 | 14.542 | Csa | 0.511 | 0.307 | |
| Narbonne | 56395 | 15.044 | Csa | 0.434 | 0.395 | |
| Cagnes-sur-Mer | 52580 | 2.815 | Csb | 1.078 | 0.509 | |
| Quimper | 63642 | 16.960 | Cfb | 0.709 | 0.536 | |
| Drancy | 71363 | 147.445 | Cfb | 0.162 | 0.086 | |
| Saint-Maur-des-Fossés | 75441 | 161.208 | Cfb | 0.207 | 0.111 | |
| Le Blanc-Mesnil | 58257 | 147.588 | Cfb | 0.178 | 0.189 | |
| Bondy | 52905 | 151.274 | Cfb | 0.186 | 0.191 | |
| Sartrouville | 51220 | 135.764 | Cfb | 0.238 | 0.285 | |
| Clichy | 64849 | 143.844 | Cfb | 0.291 | 0.218 | |
| Lyon | 522250 | 248.292 | Cfb | 0.672 | 0.292 | |
| Neuilly-sur-Seine | 59267 | 144.168 | Cfb | 0.340 | 0.061 | |
| Issy-les-Moulineaux | 68580 | 150.005 | Cfb | 0.533 | 0.100 | |
| Courbevoie | 81516 | 142.435 | Cfb | 0.456 | 0.088 | |
| Paris | 2133111 | 148.738 | Cfb | 0.452 | 0.129 | |
| Boulogne-Billancourt | 119808 | 147.767 | Cfb | 0.335 | 0.091 | |
| Levallois-Perret | 68009 | 144.143 | Cfb | 0.323 | 0.098 | |

```

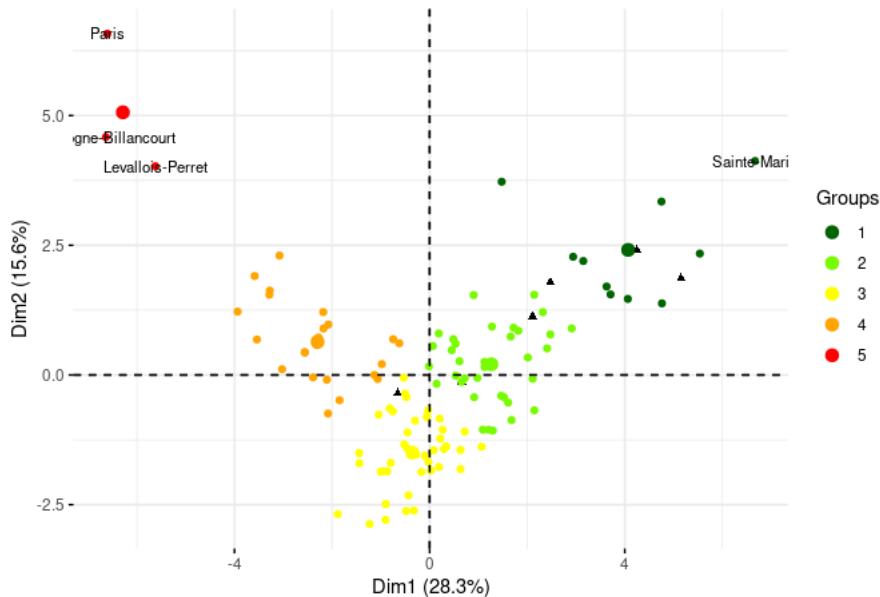
# Visualize individuals in the first factorial design
# Extract the coordinates of the individuals
coord <- as.data.frame(famd$ind$coord)

# Select the individuals to display
coord_selected <- coord[rownames(coord) %in% selected_individuals, ]

# Visualize the individuals of the FAMD with the modalities of the categorical variables
fviz_famd_ind(famd,
               habillage = analysis_df_famd$cluster,
               repel = TRUE,
               label = "none",
               palette = colors
) +
  geom_text(data = coord_selected, # Use filtered coordinates
            aes(x = Dim.1, y = Dim.2, label = rownames(coord_selected)),
            size = 3, color = "black", check_overlap = TRUE)

```

Individuals - FAMD



```

# save the plot
file_name <- paste0("res/clustering_famd/analysis_famd_plot_ind_", min_pop, selected_cities)
ggsave(file_name)

```

Define and visualize parangons

```

# Initialize df to store the parangons
parangons <- data.frame()

for (i in 1:nb_class) {
  # Filter data to select the paragon by cluster
  cluster_data <- analysis_df_var_clust_moy %>%
    filter(cluster == i) %>%
    arrange(desc(contrib_to_axis1 + contrib_to_axis2)) # Sort data

  # Select paragon
  paragon <- head(cluster_data, 1)

  # Add paragon to the dataframe
  parangons <- rbind(parangons, paragon)
}

# Extract the coordinates of the parangons
coord <- as.data.frame(famd$ind$coord)

```

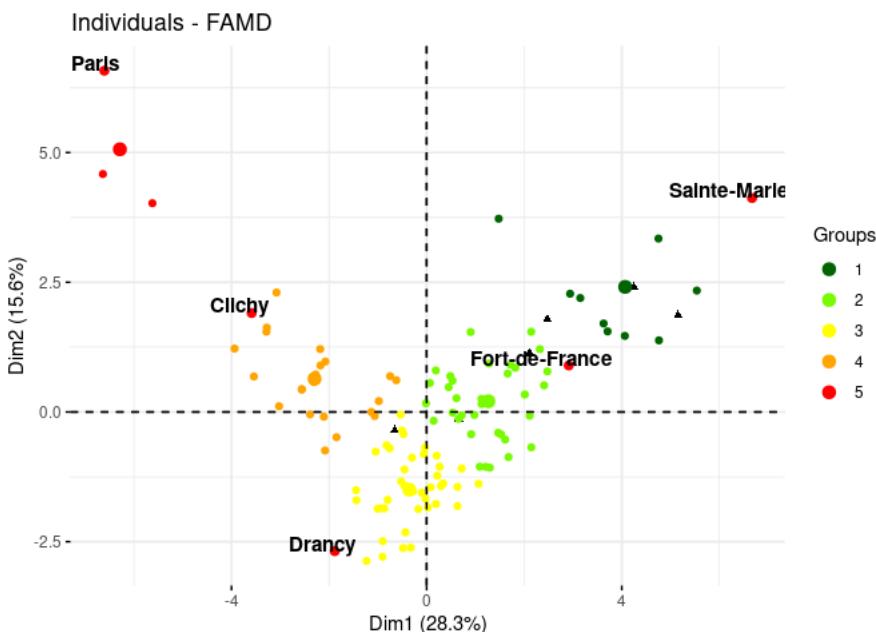
```

coord_parangons <- coord[rownames(coord) %in% parangons$city, ]

plot_famd <- fviz_famd_ind(famd,
                             habillage = analysis_df_famd$cluster,
                             repel = TRUE,
                             label = "none",
                             palette = colors
)

# Visualize individuals
plot_famd +
  geom_point(data = coord_parangons,
             aes(x = Dim.1, y = Dim.2),
             color = "red", size = 2, shape = 19) +
  geom_text(data = coord_parangons,
            aes(x = Dim.1, y = Dim.2, label = rownames(coord_parangons),
                hjust = 0.7, vjust = 0),
            size = 4, color = "black", fontface = "bold", check_overlap = TRUE)

```



```

# Save
file_name <- paste0("res/clustering_famd/analysis_famd_plot_ind_paragon_",
                     min_pop, selected)
ggsave(file_name)

```

Saving 7 x 5 in image

Table of means of variables by cluster whitout log transformation

```

# Summary Table of means of variables by cluster
cluster_means <- aggregate(analysis_df_var_clust_moy,
                            by = list(cluster = analysis_df_var_clust_moy$cluster), mean)
# Save
file_name <- paste0("res/clustering_famd/analysis_famd_cluster_moy_var_",
                     min_pop, selected)
sink(file_name)
sink()

# Table with conditional formatting
cluster_means2 <- cluster_means %>%
  dplyr::select(-cluster, -climat)

names(cluster_means2) <- make.unique(names(cluster_means2))

```

```

# Function to color the cells
highlight_cells <- function(x) {
  cell_spec(x, "html",
            color = ifelse(is.na(x), "black",
                           ifelse(x == max(x, na.rm = TRUE), "red",
                                              ifelse(x == min(x, na.rm = TRUE), "blue", "black"))))
}

cluster_means_formatted <- cluster_means2 %>%
  mutate(across(where(is.numeric), ~ round(., 3))) %>%
  mutate(across(where(is.numeric), highlight_cells))

# Create the table with `kable` and add style with `kableExtra`
cluster_means_formatted %>%
  kable("html", escape = FALSE, caption = "Averages of variables per cluster") %>%
  kable_styling(full_width = FALSE, bootstrap_options = c("striped", "hover", "condensed",

```

| Pop | min_distance_km_mer | moyenne_tri | prop_permeable | min_distance_m_moy_vege | prop |
|------------|---------------------|-------------|----------------|-------------------------|------|
| 102448.2 | 5.952 | 1.084 | 0.601 | 70.492 | 0.16 |
| 113374.083 | 69.924 | 0.548 | 0.456 | 70.254 | 0.09 |
| 90327.805 | 162.471 | 0.313 | 0.301 | 102.646 | 0.06 |
| 157763.696 | 150.844 | 0.346 | 0.203 | 105.484 | 0.06 |
| 773642.667 | 146.883 | 0.37 | 0.106 | 143.406 | 0.04 |

Frequency of climates by cluster

```

# Calculate the frequency of climates by cluster
climat_freq <- table(analysis_df_famd$cluster, analysis_df_famd$climat)
# Proportion by cluster
climat_prop <- prop.table(climat_freq, 1) # Calcul des proportions par Ligne (cluster)
# print(climat_prop)
climat_prop <- round(climat_prop, 3)
rownames(climat_prop) <- paste0("Cluster", 1:nb_class)

# Save
file_name <- paste0("res/clustering_famd/analysis_famd_climat_cluster_freq_", min_pop, sele
sink(file_name) # redirect output to the file
kable(climat_prop, rownames = "Cluster", caption = "Frequency of climates by cluster") # p

```

Frequency of climates by cluster

| Cluster | Af | As | Cfa | Cfb | Csa | Csb |
|----------|-------|-----|-------|-------|-------|-------|
| Cluster1 | 0.200 | 0.2 | 0.000 | 0.100 | 0.300 | 0.200 |
| Cluster2 | 0.028 | 0.0 | 0.028 | 0.639 | 0.222 | 0.083 |
| Cluster3 | 0.000 | 0.0 | 0.000 | 1.000 | 0.000 | 0.000 |
| Cluster4 | 0.000 | 0.0 | 0.000 | 1.000 | 0.000 | 0.000 |
| Cluster5 | 0.000 | 0.0 | 0.000 | 1.000 | 0.000 | 0.000 |

```

sink() # stop redirecting output to the file

# Save data with cluster
analysis_df_famd$city <- analysis_df_famd$city[match(rownames(analysis_df_famd), rownames(a

```

```

write.csv(analysis_df_fam,
          file = paste0("res/clustering_famd/analysis_famd_cluster_ind_", min_pop, selected)

```

Help to interpret the clusters

Profile for Cluster 1

| Summary of Statistics for Cluster 1 | |
|---|----------|
| Statistic | Value |
| Number of Individuals | 10 |
| Average Population | 102448.2 |
| Average Distance to the Sea (km) | 5.95 |
| Dominant Climate | Csa |
| Average Recycling Rate | 1.08 |
| Permeable Proportion | 0.6 |
| Average Distance to Vegetation (m) | 70.49 |
| Vegetation Patch Proportion | 0.17 |
| Average Surface Area (ha) | 363.29 |
| Number of Urban Sprawl Areas | 6.8 |
| Compact Buildings Proportion | 0.03 |
| High Buildings Proportion | 0 |

Profile for Cluster 2

| Summary of Statistics for Cluster 2 | |
|---|-----------|
| Statistic | Value |
| Number of Individuals | 36 |
| Average Population | 113374.08 |
| Average Distance to the Sea (km) | 69.92 |
| Dominant Climate | Cfb |
| Average Recycling Rate | 0.55 |
| Permeable Proportion | 0.46 |
| Average Distance to Vegetation (m) | 70.25 |
| Vegetation Patch Proportion | 0.1 |
| Average Surface Area (ha) | 1320.14 |
| Number of Urban Sprawl Areas | 2.83 |
| Compact Buildings Proportion | 0.06 |
| High Buildings Proportion | 0 |

Profile for Cluster 3

| Summary of Statistics for Cluster 3 | |
|-------------------------------------|-------|
| Statistic | Value |

| Statistic | Value |
|---|---------|
| Number of Individuals | 41 |
| Average Population | 90327.8 |
| Average Distance to the Sea (km) | 162.47 |
| Dominant Climate | Cfb |
| Average Recycling Rate | 0.31 |
| Permeable Proportion | 0.3 |
| Average Distance to Vegetation (m) | 102.65 |
| Vegetation Patch Proportion | 0.07 |
| Average Surface Area (ha) | 1646.84 |
| Number of Urban Sprawl Areas | 1.12 |
| Compact Buildings Proportion | 0.04 |
| High Buildings Proportion | 0 |

Profile for Cluster 4

Summary of Statistics for Cluster 4

| Statistic | Value |
|---|----------|
| Number of Individuals | 23 |
| Average Population | 157763.7 |
| Average Distance to the Sea (km) | 150.84 |
| Dominant Climate | Cfb |
| Average Recycling Rate | 0.35 |
| Permeable Proportion | 0.2 |
| Average Distance to Vegetation (m) | 105.48 |
| Vegetation Patch Proportion | 0.06 |
| Average Surface Area (ha) | 1638.72 |
| Number of Urban Sprawl Areas | 1.22 |
| Compact Buildings Proportion | 0.23 |
| High Buildings Proportion | 0.01 |

Profile for Cluster 5

Summary of Statistics for Cluster 5

| Statistic | Value |
|---|-----------|
| Number of Individuals | 3 |
| Average Population | 773642.67 |
| Average Distance to the Sea (km) | 146.88 |
| Dominant Climate | Cfb |
| Average Recycling Rate | 0.37 |

| Statistic | Value |
|------------------------------------|--------|
| Permeable Proportion | 0.11 |
| Average Distance to Vegetation (m) | 143.41 |
| Vegetation Patch Proportion | 0.04 |
| Average Surface Area (ha) | 856.13 |
| Number of Urban Sprawl Areas | 2.33 |
| Compact Buildings Proportion | 0.66 |
| High Buildings Proportion | 0.03 |

Table and view ZCL of Parangons

```

# Initialize the dataframe to store the parangons
parangons <- data.frame()

for (i in 1:nb_class) {
  # Filter data to select the parangon by cluster
  cluster_data <- analysis_df_var_clust_moy %>%
    filter(cluster == i) %>%
    arrange(desc(contrib_to_axis1 + contrib_to_axis2))

  # Select the parangon
  parangon <- head(cluster_data, 1)

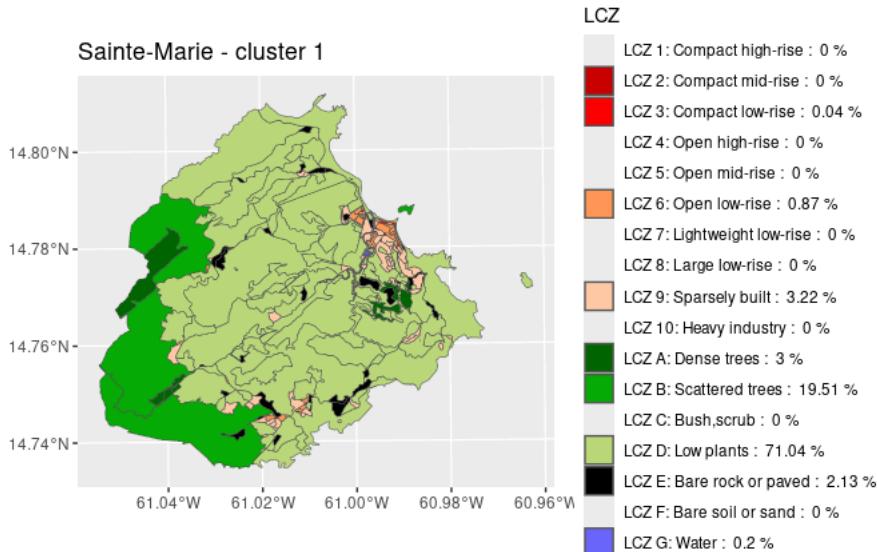
  # Add the parangon to the dataframe
  parangons <- rbind(parangons, parangon)
}

# Add the city
parangon_vec <- parangons$city
vec_to_plot <- parangon_vec

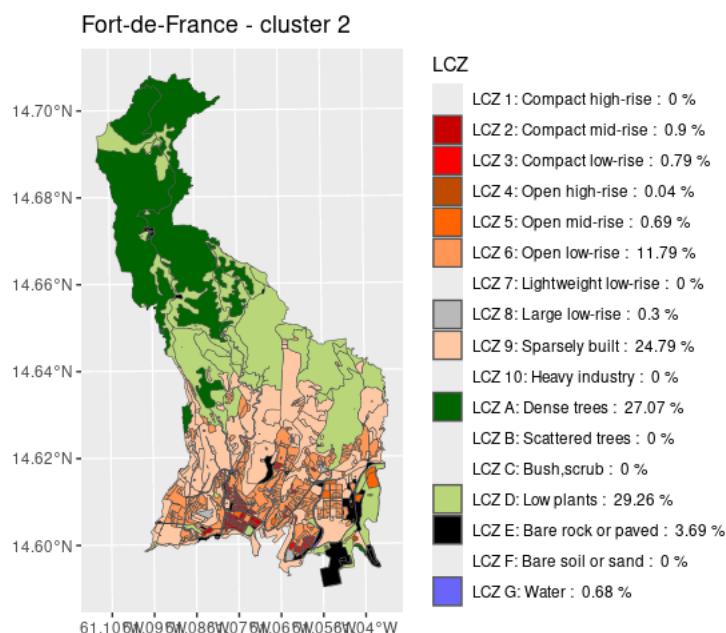
data_file <- "res/cities"

# Cluster 1
city_to_plot <- vec_to_plot[1]
city_data <- importLCZvect(dirPath = paste0(data_file, "/osm_", city_to_plot), file="rsu_lczi.
geomID="ID_RSU", confid="LCZ_UNIQUENESS_VALUE")
urb_sprawl <- st_read(paste0(data_file, "/osm_", city_to_plot, "/sprawl_areas.fgb"), quiet = T
city_data2 <- st_intersection(urb_sprawl,city_data)
clust1_city <- showLCZ(city_data,title = paste0(city_to_plot," - cluster 1")) +
  geom_sf(data=urb_sprawl,col="red",size=10,alpha=0)

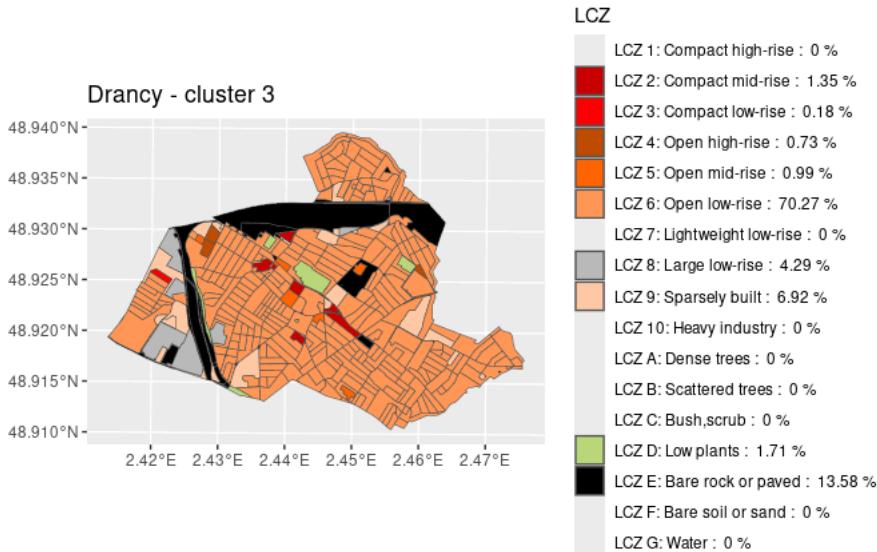
```



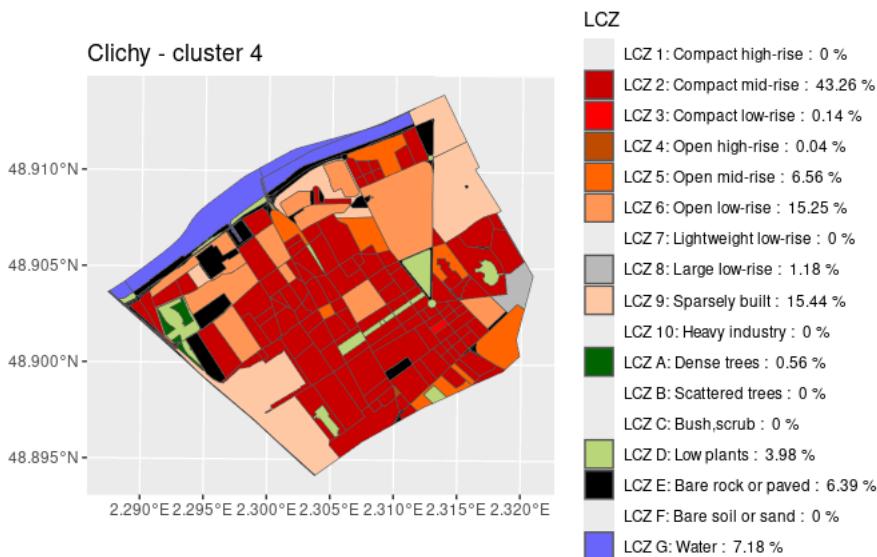
```
# Cluster 2
city_to_plot <- vec_to_plot[2]
city_data <- importLCZvect(dirPath = paste0(data_file,"/osm_",city_to_plot), file="rsu_lczi.
geomID="ID_RSU", confid="LCZ_UNIQUENESS_VALUE")
urb sprawl <- st_read(paste0(data_file,"/osm_",city_to_plot,"/sprawl_areas.fgb"), quiet = T
city_data2 <- st_intersection(urb sprawl,city_data)
clust2_city <- showLCZ(city_data,title = paste0(city_to_plot," - cluster 2")) +
  geom_sf(data=urb sprawl,col="red",size=10,alpha=0)
```



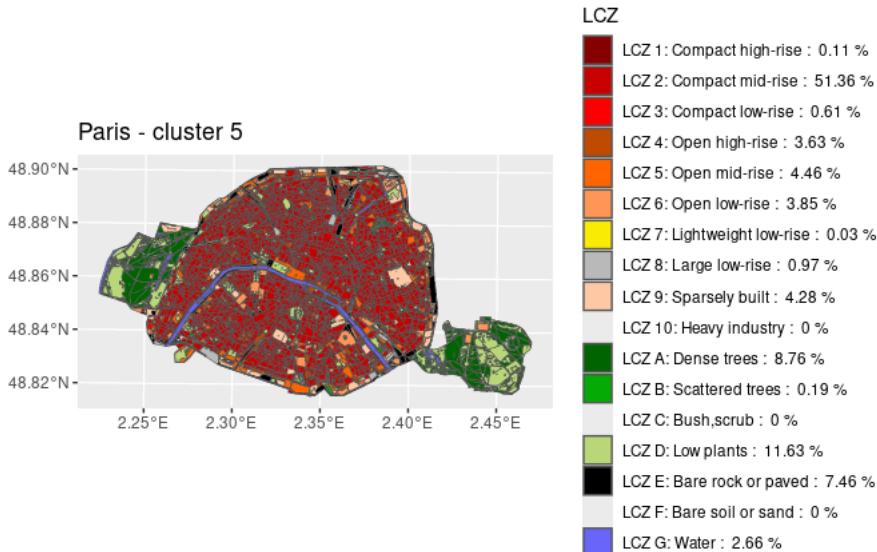
```
# Cluster 3
city_to_plot <- vec_to_plot[3] # Berck, Avallon, Rambouillet, Chatou
city_data <- importLCZvect(dirPath = paste0(data_file,"/osm_",city_to_plot), file="rsu_lczi.
geomID="ID_RSU", confid="LCZ_UNIQUENESS_VALUE")
urb sprawl <- st_read(paste0(data_file,"/osm_",city_to_plot,"/sprawl_areas.fgb"), quiet = T
city_data2 <- st_intersection(urb sprawl,city_data)
clust3_city <- showLCZ(city_data,title = paste0(city_to_plot," - cluster 3")) +
  geom_sf(data=urb sprawl,col="red",size=15,alpha=0)
```



```
# Cluster 4
city_to_plot <- vec_to_plot[4]
city_data <- importLCZvect(dirPath = paste0(data_file,"/osm_",city_to_plot), file="rsu_lczi.geomID="ID_RSU", confid="LCZ_UNIQUENESS_VALUE")
urb sprawl <- st_read(paste0(data_file,"/osm_",city_to_plot,"/sprawl_areas.fgb"), quiet = T)
city_data2 <- st_intersection(urb sprawl,city_data)
clust4_city <- showLCZ(city_data,title = paste0(city_to_plot," - cluster 4")) +
  geom_sf(data=urb sprawl,col="red",size=10,alpha=0)
```



```
# Cluster 5
city_to_plot <- vec_to_plot[5]
city_data <- importLCZvect(dirPath = paste0(data_file,"/osm_",city_to_plot), file="rsu_lczi.geomID="ID_RSU", confid="LCZ_UNIQUENESS_VALUE")
urb sprawl <- st_read(paste0(data_file,"/osm_",city_to_plot,"/sprawl_areas.fgb"), quiet = T)
city_data2 <- st_intersection(urb sprawl,city_data)
clust5_city <- showLCZ(city_data,title = paste0(city_to_plot," - cluster 5")) +
  geom_sf(data=urb sprawl,col="red",size=15,alpha=0)
```



```
full_plot <- plot_grid(clust1_city,clust2_city,clust3_city, clust4_city,clust5_city,
                       ncol=3)

file_name <- paste0("res/clustering_famd/analysis_famd_plot_zcl_parangons_", min_pop, selec
                     ggsave(filename = file_name ,width = 20,height = 9)
```

Map of the cities

```
analysis_df_famd$city <- rownames(analysis_df_famd)

analysis_df_famd <- analysis_df_famd %>%
  mutate(city = case_when(
    city == "Sainte-Marie" ~ "Sainte-Marie, Martinique",
    city == "Saint-Pierre" ~ "Saint-Pierre, La Réunion",
    city == "Saint-André" ~ "Saint-André, La Réunion",
    TRUE ~ city # Utiliser uniquement la ville pour les autres cas
  ))


# Géocoded cities with Nominatim
data <- analysis_df_famd %>%
  geocode(city, method = 'osm', lat = latitude, long = longitude)

# Converted to Lambert 93 coordinates to have a map of France
data_sf <- st_as_sf(data, coords = c("longitude", "latitude"), crs = 4326) %>%
  st_transform(crs = 2154) # Lambert 93

# Download the map of France
france_map <- ne_states(geounit = "France", returnclass = "sf")

# Convert the coordinates of the map of France to Lambert 93
france_sf <- st_as_sf(france_map, coords = c("long", "lat"), crs = 4326) %>%
  st_transform(crs = 2154) # Lambert 93

# Create the map
map <- ggplot() +
  geom_sf(data = france_sf, fill = "white", color = "grey", size = 0.2) +
  geom_sf(data = data_sf, aes(color = as.factor(cluster),
                               text = paste("Ville: ", city, "<br>Cluster: ", cluster)),
          size = 2) +
  theme_minimal() +
  scale_color_manual(values = colors, name = "Cluster") +
```

```

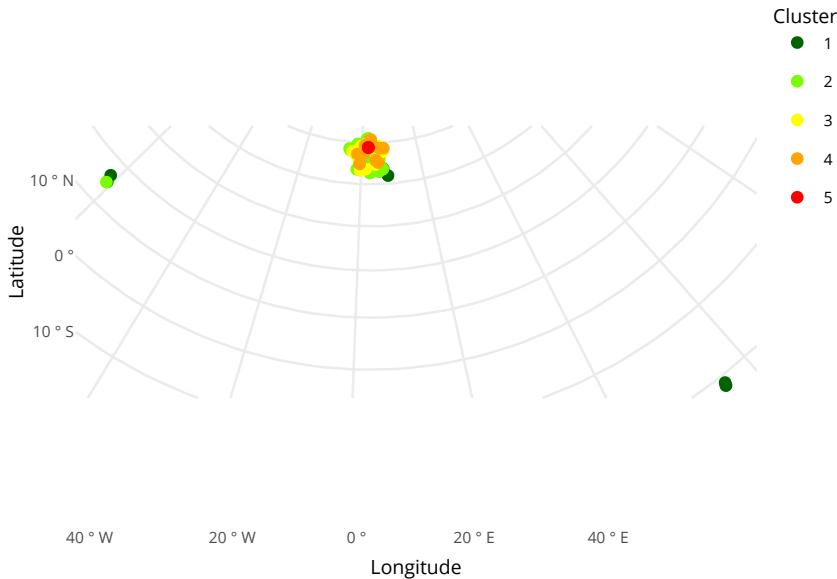
  labs(title = "Map of cities by cluster",
       x = "Longitude",
       y = "Latitude")

  # INteractive map with plotly
  map_interactive <- ggplotly(map, tooltip = "text")

  # Print the interactive map
  map_interactive

```

Map of cities by cluster



ICU Comparison

```

# Download the data
indic_icu_df_full <- read.csv("res/ICU_Comparison/indic_icu_df_full.csv")

# Add the city as a row name
analysis_df_famd$cluster <- as.factor(km.res$cluster)
analysis_df_famd$city <- rownames(analysis_df_famd)

# Join
clust_icu_df <- inner_join(analysis_df_famd, indic_icu_df_full)

# List of ICU indicators
vec_indicator <- c("mean_icu", "max_icu", "sd_icu", "cv_icu", "max5_mean_icu", "p_tmax_10", "p_tm"

# Initialize the R^2
R2_full <- c()

# Verify if there are enough data to perform the regression
if(length(unique(clust_icu_df$cluster)) > 1) {
  # Loop over the indicators
  for(i in 1:length(vec_indicator)){
    formula_string <- paste0(vec_indicator[i], " ~ cluster")

    mod1 <- lm(as.formula(formula_string), data = clust_icu_df)
    sum_mod1 <- summary(mod1)

    if(i == 1) R2_full <- sum_mod1$r.squared
    else R2_full <- c(R2_full, sum_mod1$r.squared)
  }

} else {
  cat("Pas assez de données pour effectuer des régressions. Données seulement pour les clus"

```

```

R2_full <- rep(NA, length(vec_indicator))
}

# Add the names of the indicators
names(R2_full) <- vec_indicator

# Filter the indicators with a R2 > 0.5
R2_full[which(R2_full > 0.5)]

```

| | | | | |
|----|-----------|-----------|-----------------------------|---------------|
| ## | mean_icu | max_icu | cv_icu | max5_mean_icu |
| ## | 0.8852406 | 0.8191229 | 0.5796201 | 0.8205832 |
| ## | p_tmax_10 | p_tmax_30 | p_tmax_50 | gini |
| ## | 0.7326609 | 0.8569378 | 0.7470346 | 0.5802154 |
| ## | p_tsup_2 | p_tsup_3 | p_tinf_1 p_tsup_meansd_plus | |
| ## | 0.7018484 | 0.8300904 | 0.6566183 | 0.5864986 |

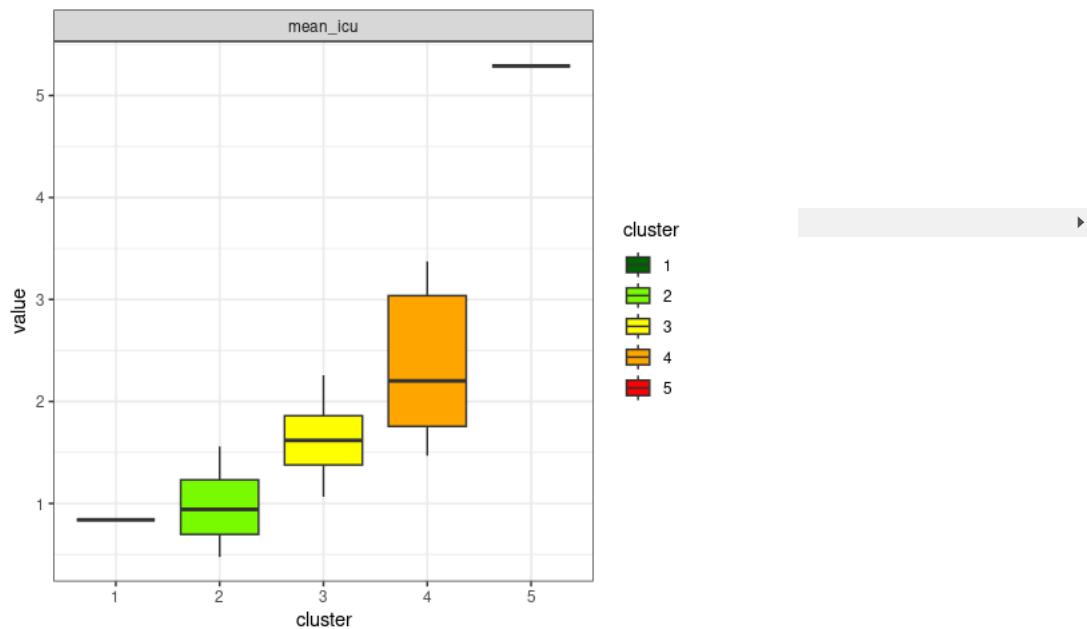
```

# Graph of the indicators
clust_icu_df_2 <- clust_icu_df[,c("cluster","city","mean_icu","max_icu")] %>%
  inner_join(analysis_df_famd) %>%
  pivot_longer(mean_icu) %>%
  distinct(cluster, city, name, value)

# Plot
clust_icu_plot <- ggplot(clust_icu_df_2) +
  geom_boxplot(aes(x = cluster, y = value, fill = cluster)) +
  theme_bw() +
  theme(aspect.ratio = 1) +
  facet_wrap(~name, scales = "free_y") +
  scale_fill_manual(values = colors, drop = FALSE)

print(clust_icu_plot)

```



```

# Number of cities in each cluster
city_count_by_cluster <- clust_icu_df_2 %>%
  group_by(cluster) %>%
  summarise(city_count = n_distinct(city))
kable(city_count_by_cluster, caption = "Number of cities in each cluster")

```

Number of cities in each cluster

| cluster | city_count |
|---------|------------|
|---------|------------|

| cluster | city_count |
|---------|------------|
| 1 | 1 |
| 2 | 7 |
| 3 | 10 |
| 4 | 9 |
| 5 | 1 |

Fiche Synthèse

Typologies de villes françaises sur la base des zones climatiques locales (ZCL) et des déterminants de la surchauffe urbaine

Étudiante : Besnard Pauline

Date stage : 13 mai 2024 / 1 novembre 2024

Date de soutenance : 25 octobre 2024

La surchauffe des villes est un **enjeu majeur dans le contexte du réchauffement climatique**. À l'échelle mondiale la température terrestre de la dernière décennie a augmenté de 1.1 degré depuis l'ère industrielle⁴. Cette hausse des températures est plus importante en ville avec des conséquences sur la santé des populations. En particulier, les phénomènes d'**îlots de chaleur urbain** (ICU) sont de plus en plus intenses et de plus en plus étendus dans le temps et dans l'espace.

Les phénomènes de surchauffe sont étroitement liés à l'occupation du sol, aux matériaux de construction, à la morphologie urbaine, au climat, à la topographie et aux activités humaines. La notion de zone climatique locale permet de synthétiser l'ensemble de ces déterminants en indicateur synthétique du microclimat urbain.

Dans ce stage, on cherche à **construire une classification des villes françaises en se basant sur les zones climatiques locales et d'autres indicateurs conditionnant le climat urbain** afin d'identifier des profils de villes sensibles aux phénomènes de surchauffe.

L'outil SIG Geoclimate ainsi que d'autres sources de données auxiliaires (modèles numériques de terrain) ont été utilisés pour produire les indicateurs du climat urbain. Une première étape bibliographique a permis de formaliser des indicateurs du climat urbain suivant quatre catégories : l'occupation du sol, la morphologie urbaine, le climat et la topographie et enfin les activités anthropiques. Du fait de la présence de groupes de variables et de variables qualitatives et quantitatives, les indicateurs ont été analysés par analyse factorielle des données mixtes (AFDM) dont les résultats ont fait l'objet d'une classification. Plusieurs groupes de villes ressortent : (1) des villes avec une densité de zones compactes importantes, (2) des villes, situées en périphérie de gros centres urbains, caractérisées par une prédominance de bâtiments non-compacts ainsi que peu de végétation et enfin (3) des villes se distinguant par la présence de végétation à l'intérieur ou en périphérie de la ville. L'étude confirme, comme attendu, que le premier groupe connaît des phénomènes d'ICU significativement plus intenses que les villes végétalisées. Ce travail pose les bases pour une étude mondiale de la classification des villes en fonction de leur exposition aux phénomènes d'ICU en vue de l'adaptation des villes au réchauffement climatique.

Mots-clefs : Statistiques, Réchauffement climatique, Classification non-supervisée, Analyse spatiale, Surchauffe urbaine

Formation

Tuteur universitaire : Demoraes Florent

Master SIGAT – Université Rennes 2

Année Universitaire 2023 / 2024

Structure d'accueil

Tuteur professionnel : Alglave Baptiste

Structure d'accueil : Lab-STICC, IUT de Vannes, Morbihan, France

⁴ 1850 - 1900, période de référence prise par l'Accord de Paris (GIEC, 2021)