

# Méthodologie de classification des bâtiments à l'échelle de la ville au regard de leur potentiel de flexibilité énergétique

## Methodology for classifying buildings at the city scale regarding their energy flexibility potential

Astrid GANSWINDT<sup>123\*</sup>, Auline RODLER<sup>1</sup>, Sihem GUERNOUTI<sup>1</sup>, Julien RAMOUSSE<sup>2</sup>, David DA SILVA<sup>3</sup>, Jean-Baptiste VIDEAU<sup>3</sup>, Valentin GAVAN<sup>3</sup>

<sup>1</sup>Cerema, Equipe Bâtiments Performants dans leur Environnement (BPE)  
Cité Administrative de Nantes, 12, bd Vincent Gâche - 44200 Nantes

<sup>2</sup>Laboratoire Procédés Energie Bâtiment (LOCIE), CNRS UMR 5271, Université Savoie Mont Blanc, Polytech Annecy-Chambéry  
Campus scientifique, Savoie Technolac - 73376 Le Bourget-Du-Lac

<sup>3</sup>ENGIE Lab Crigen, Lab Ville et Bâtiments de demain  
4, rue Joséphine Baker - 93240 Stains

\*(auteur correspondant : astrid.ganswindt@cerema.fr)

**Résumé** - Cet article présente une méthodologie d'identification d'archétypes de bâtiments représentatifs des différents gisements de flexibilité à l'échelle d'une ville. Elle repose sur des techniques de clustering, qui répartissent les bâtiments en fonction de leurs usages, géométrie, morphologie et de leur contexte urbain proche, suivant l'influence que ces caractéristiques peuvent avoir sur le potentiel de flexibilité des bâtiments. La méthode développée est illustrée par un exemple d'application sur la ville de Nantes, mais se veut répliquable à d'autres villes, en fonction des données disponibles.

**Abstract** - This article presents a methodology for identifying building archetypes representing the different energy flexibility potentials at the city scale. This methodology employs clustering techniques to categorize buildings based on their usage, geometry, morphology, and urban surrounding, focusing on the characteristics that influence their flexibility potential. The method has been developed for the French city of Nantes, but is also intended to be generalized to other cities, depending on the available data.

### 1. Introduction

L'utilisation d'énergies renouvelables intermittentes et par conséquent, non pilotables (i.e. solaire, éolien) nécessite que les bâtiments adaptent leur consommation aux contraintes du réseau et aux conditions climatiques locales, tout en respectant les besoins des usagers : il s'agit de la flexibilité de la consommation ou « demand-side management » [1].

À l'échelle du bâtiment, le potentiel de stockage dans la masse thermique du bâtiment constitue un gisement de flexibilité conséquent [2]. Si l'exploitation des gisements de flexibilité des bâtiments à l'échelle du quartier permet d'augmenter le volume de stockage, elle permet aussi de développer des stratégies de maximisation du potentiel de flexibilité des bâtiments, avec notamment du foisonnement entre usages, c'est-à-dire un lissage de la demande énergétique d'un quartier du fait du caractère asynchrone des besoins des différents bâtiments [3], en particulier dans les secteurs résidentiel et tertiaire.

L'objectif de ces travaux de recherche est d'évaluer les gisements de flexibilité de différents types de bâtiments (résidentiels, bureaux, commerces, anciens, neufs, en tissu urbain dense, etc.) notamment via leur inertie, vis-à-vis de la production de chauffage et de refroidissement, dans le but de définir des configurations permettant de valoriser ces potentiels à l'échelle urbaine. Il s'agit de définir quels types de bâtiments mettre en relation au vu de leur inertie et de leur localisation les uns par rapport aux autres, ainsi que de déterminer à quelle distance de la source de production d'énergie du réseau de chaleur et de froid urbain (RCU) l'inertie des bâtiments peut-elle encore être valorisée et, dans le cas où la configuration bâtiment/RCU n'est pas adaptée, quelle autre source d'énergie est à favoriser (électricité, gaz, EnR).

Afin de répondre à ces questionnements, une méthodologie pour l'ensemble des travaux de recherche est mise en place pour modéliser le potentiel de flexibilité des bâtiments à l'échelle urbaine. Celle-ci se déroule en trois temps. Il s'agit tout d'abord de définir des archétypes de bâtiments représentatifs des différents potentiels de flexibilité de la demande du quartier, à partir de techniques de clustering. Ensuite, le potentiel de flexibilité de chacun de ces archétypes de bâtiments est évalué par simulation thermique dynamique (STD). Enfin, une dernière étape d'agrégation des bâtiments permet d'évaluer le potentiel de flexibilité à l'échelle urbaine.

Cet article présente la démarche correspondant à la première étape de ce travail de recherche. L'enjeu de cette étape de classification des bâtiments est de trouver un compromis pour définir un nombre suffisants d'archétypes de bâtiments représentant au mieux les différents gisements de flexibilité de la ville de Nantes, tout en limitant ce nombre d'archétypes afin de réduire les temps de calculs liés aux STD.

## **2. Méthode**

### **2.1. État de l'art**

#### ***2.1.1. Outils et méthodes de modélisation des bâtiments à l'échelle de la ville***

L'enjeu du choix d'une méthode de modélisation des bâtiments à l'échelle urbaine est de trouver un compromis entre le nombre de bâtiments modélisés et leur niveau de détail. Ici, le nombre de bâtiments à modéliser est de l'ordre de plusieurs dizaines de milliers. Un outil comme URBANopt [5], qui modélise jusqu'à quelques dizaines de bâtiments n'est donc pas adapté à cette étude, dont les objectifs en terme d'échelle s'apparentent davantage à ceux des logiciels DIMOSIM [4] et CityBES [5]. Si ces outils possèdent des versions disponibles gratuitement, il reste cependant difficile d'en changer le paramétrage pour une utilisation plus approfondie. C'est pourquoi ces outils modélisant l'ensemble des bâtiments ont été écartés au profit des techniques de clustering, qui ont pour but de définir des groupes de bâtiments (clusters), représentés par un bâtiment archétype. Ce dernier est ensuite modélisé par STD avec un niveau de détail élevé.

#### ***2.1.2. Méthodes de classification des bâtiments***

Différentes méthodes de classification des bâtiments ont été envisagées pour cette étude, comme la méthode TABULA [6], qui répertorie les bâtiments résidentiels à l'échelle nationale, la classification en Local Climate Zones (LCZ) [7] qui qualifie l'environnement urbain à une maille de 200m, ou la classification GENIUS [8] qui définit des archétypes de bâtiments en fonction de leur contexte urbain et de leur besoins énergétiques. La particularité de la méthode

proposée ici est qu'elle prend en compte les spécificités d'une ville en particulier, avec une définition à la maille bâtiment pour le résidentiel et le tertiaire, en fonction du potentiel de flexibilité de ces bâtiments.

### 2.1.3. État de l'art des caractéristiques influentes pour la flexibilité

Les caractéristiques influençant le potentiel de flexibilité à l'échelle du bâtiment sont la qualité de son enveloppe pour le chauffage, et pour le refroidissement, il s'agit du contexte morphologique urbain et de la période d'activation de flexibilité, selon une analyse de sensibilité de Vivian et al.[9]. À l'échelle du quartier (trentaine de bâtiments), Martinez et al.[10] soulignent également l'importance des modalités d'activation de flexibilité (durée, différence de température), et ajoutent également les usages (scénarios d'occupation) comme facteur influent du potentiel de flexibilité. Dans cette étude, la morphologie urbaine a un impact secondaire sur ce potentiel, mais cela peut être dû au fait que le quartier étudié est relativement homogène. L'attention portée à la morphologie urbaine comme facteur influant sur le potentiel de flexibilité, notamment pour le refroidissement, constitue une originalité de l'approche proposée dans cet article.

## 2.2. Démarche

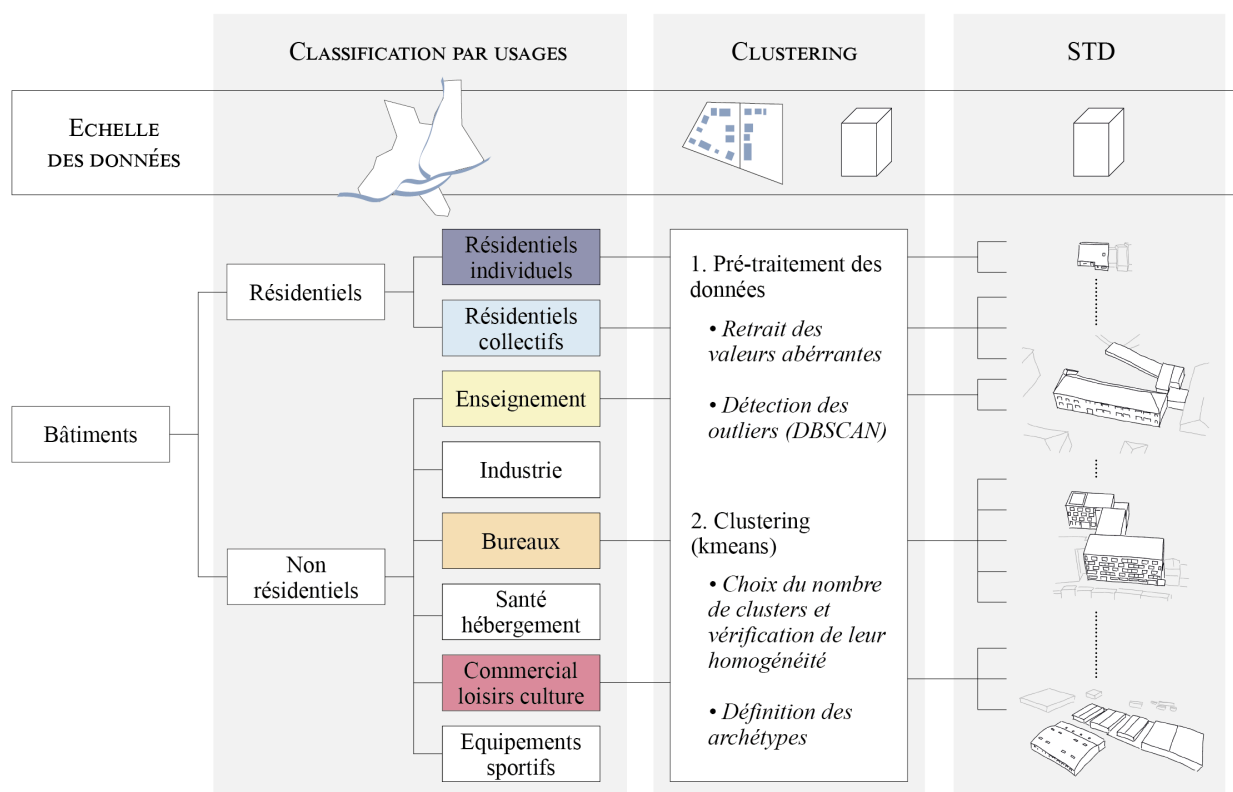


Figure 1 : Démarche générale de définition des archétypes

Comme représenté sur la Figure 1 les archétypes sont définis suivant deux étapes : une première étape de répartition des bâtiments par usage à l'échelle de la ville, suivie d'une seconde étape de clustering prenant en compte la géométrie du bâtiment et sa morphologie urbaine proche. Ce séquençage permet de hiérarchiser les données d'entrée en privilégiant une répartition par usages, qui ont une plus grande influence sur le potentiel de flexibilité que la

géométrie des bâtiments. Il permet aussi traiter spécifiquement les bâtiments tertiaires, dont la proportion restreinte par rapport au parc (15%) ne permet pas à ces bâtiments d'être suffisamment représentatifs à l'échelle de la ville, pour être classifiés dans des clusters spécifiques au tertiaire. Les paramètres les plus influents pour la flexibilité, à savoir les conditions d'activation de flexibilité et les caractéristiques thermiques des bâtiments, sont traités en dernière étape par simulation STD sur les archétypes de bâtiments, afin notamment de prendre en compte des variantes de rénovation.

Le clustering présenté dans la suite est appliqué aux bâtiments de la ville de Nantes (56 000 bâtiments), dont 70% sont du résidentiel individuel, 15% du résidentiel collectif et 15% du tertiaire. Parmi ces deux dernières catégories, 450 bâtiments sont recensés par Nantes Métropole comme étant raccordés au RCU.

### **3. Clustering**

#### **3.1. Hypothèses et prétraitement des données**

##### **3.1.1. Données d'usages**

Les sous-catégories d'usages représentées en Figure 1 sont issues d'une analyse de scénarios d'occupation décrits dans la thèse de Nguyen [11]. Ceux-ci distinguent les occupations journalières (bureau), nocturnes (logements), en semaine et hors période estivale (bâtiments d'enseignement), etc.

La répartition des bâtiments en fonction de ces catégories nécessite un recoupement de plusieurs bases de données (BDD) à la maille bâtiment : BDTOPO [12], BDNB Expert [13], fichiers fonciers du Cerema [14], ainsi que des données recueillies pour le cas d'étude, à savoir les données de Nantes Métropole sur les bâtiments raccordés au réseau de chaleur. Ces dernières données spécifiques à Nantes représentent une centaine de bâtiments (0,2% du parc) et sont disponibles dans le cadre du partenariat de recherche. Le nombre de bâtiments recensés dans ces BDD y est généralement inversement proportionnel à la fiabilité des données. L'assignation finale des usages s'effectue en affectant pour chaque BDD une catégorie d'usages pour chaque bâtiment. Dans le cas où ces catégories entrent en contradiction, ce qui est le cas pour près de 80% des bâtiments, un ordre de priorité est donné en fonction de la fiabilité des BDD :

1. Données de Nantes Métropole sur les bâtiments raccordés au RCU.
2. Fichiers fonciers du Cerema issus de la BDNB Expert.
3. Fichiers fonciers du Cerema pour le tertiaire (typeact).
4. BDTOPO zone d'activité ou d'intérêt (ZOA), utilisée principalement pour le tertiaire.
5. Fichiers fonciers du Cerema pour le résidentiel et tertiaire (typeloc).
6. BDTOPO général (nature, usage 1 et usage 2).

Ainsi il est possible de répartir 90% des bâtiments en sous-catégories d'usages, à l'exception des bâtiments non renseignés dans les BDD ou avec des données trop imprécises.

Dans le cas de bâtiments multi-usages, tels que les logements collectifs sur socle commercial, un usage unique est affecté, avec priorité au résidentiel. Dans le cadre de cette étude, les locaux industriels ne sont pas traités dans le clustering car leurs gisements de flexibilité résident principalement dans leurs procédés industriels, qui sont très spécifiques et difficilement généralisables. Les équipements sportifs et les bâtiments de santé et d'hébergement (hôpitaux) sont également écartés de l'étude, car leur nombre est limité et car ils sont spécifiques à chaque ville. Il pourrait toutefois être intéressant de traiter ces usages spécifiques individuellement dans

le cadre de futures études.

### 3.1.2. Données morphologiques

Dans sa thèse, Bonhomme [8] présente les caractéristiques morphologiques urbaines qui impactent le plus les besoins de chaleur et de refroidissement, à savoir la densité (1), le coefficient d'emprise au sol (CES) (2) ainsi que la compacité (3), définis comme suit :

$$Densite = \frac{Surface_{batiment} \cdot Nb_{etages}}{Surface_{maille}} \quad (1)$$

$$CES = \frac{Surface_{batiment}}{Surface_{maille}} \quad (2)$$

$$Compacite = \frac{Surface_{verticale\ deperditive_{batiment}}}{Volume^{\frac{2}{3}}} \quad (3)$$

Ces indicateurs sont calculés sur Geoclimate à la maille RSU, ou Reference Spatial Unit [15], découpage généré à partir du tracé des voiries et du parcellaire. Cette maille présente l'avantage de ne pas dépendre du point d'ancrage du maillage (comme pour un maillage régulier par exemple), tout en associant chaque bâtiment à une seule maille, sans chevauchement pour la majorité d'entre eux.

### 3.1.3. Variables retenues pour le clustering

Après analyse par information mutuelle des variables, qui mesure le degré de dépendance entre deux paramètres, une combinaison de six variables est retenue pour le clustering. À l'échelle du bâtiment, la hauteur, la surface et la compacité décrivent la géométrie des bâtiments, tandis que la surface verticale adjacente prend en compte leur mitoyenneté. À la maille RSU, la densité et le CES caractérisent la morphologie du contexte proche des bâtiments.

## 3.2. Détection des outliers

Le clustering est fortement sensible aux données isolées, qu'il convient donc d'analyser préalablement. Parmi ces valeurs se trouvent des bâtiments dont la répartition par usages est erronée, à l'exemple de résidentiels individuels de plus de 1000 m<sup>2</sup> ou de bureaux de moins de 50 m<sup>2</sup>, qui correspondent généralement à des annexes. Ceux-ci sont retirés des données en aval de l'application du clustering.

Les valeurs isolées concernent également des bâtiments dont l'usage est correct, mais dont les caractéristiques diffèrent de la majorité des bâtiments du cluster. Il s'agit de bâtiments qualifiés d'outliers, identifiés ici comme les grands centres commerciaux ou les barres d'immeubles de logements collectifs par exemple. Dans la perspective d'évaluer le potentiel de flexibilité du quartier, il peut être opportun de conserver ces types de bâtiments qui représentent un gisement de flexibilité conséquent, même si peu représentatif de l'ensemble du cluster. Ainsi, pour les catégories de bâtiments résidentiels collectifs, commerciaux, d'enseignement et de bureaux, les outliers sont conservés pour la suite du clustering. Pour les autres sous-usages, les outliers sont détectés par la méthode Density-Based Spatial Clustering of Application with Noise (DBSCAN), dont l'hyperparamétrage retire des bâtiments à hauteur de 5%, dans le but de conserver une représentativité suffisante de l'ensemble des bâtiments.

### 3.3. Choix de la méthode de clustering et identification des archétypes

Avant tout, soulignons le fait que le choix d'une méthode de clustering dépend principalement de la forme du jeu de données à traiter. Dans la méthode développée ici, un clustering est réalisé pour chaque catégorie d'usages, ce qui implique des jeux de données différents, notamment en terme de nombre de bâtiments. La catégorie enseignement par exemple comporte plusieurs centaines de bâtiments de typologies variées, tandis que le résidentiel individuel est composé de plusieurs dizaines de milliers de bâtiments présentant des typologies relativement homogènes. Cependant, pour des raisons de cohérence de la démarche générale, nous sélectionnons une seule méthode de clustering pouvant s'adapter à ces différents jeux de données.

Pour cette étude, les méthodes de clustering hiérarchique et spectral [16] ont été écartées car elles sont peu adaptées à des jeux de données de grandes dimensions, tel que celui des bâtiments résidentiels individuels. La méthode DBSCAN est également écartée du clustering car elle peut générer des clusters de forme concave, pour lesquels la détermination d'un centroid est inappropriée. Les méthodes kmeans et Gaussian Mixture Model (GMM) sont comparées afin de déterminer la méthode la plus adaptée. Pour cela, trois indicateurs d'homogénéité du clustering sont calculés pour deux à dix clusters : le coefficient de silhouette (à maximiser), l'indice de Davies Bouldin (à minimiser) et l'indice de Calinski Harabasz (à maximiser).

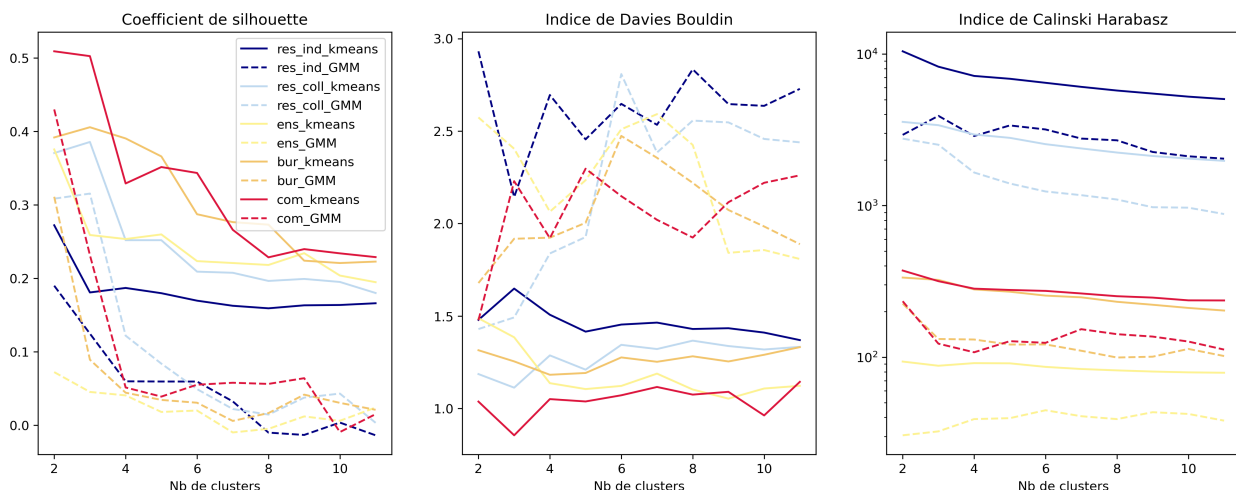


Figure 2 : Coefficients de performance du clustering par sous catégories d'usages

La Figure 2 montre que pour l'ensemble des sous-usages, la méthode kmeans (en trait plein) est préférable à GMM (en pointillés) et est donc la méthode retenue pour l'ensemble de la démarche. Les indicateurs de performance du clustering permettent également de déterminer le nombre de clusters pour chaque sous-usages, puis d'identifier le centroid de chaque cluster, défini comme le bâtiment le plus proche de la médiane. Quinze bâtiments archétypes, illustrés en Figure 3, sont ainsi identifiés pour représenter les 40 000 bâtiments traités dans le clustering. Il s'agit d'un compromis acceptable entre la représentativité des bâtiments de la ville de Nantes et le nombre de bâtiments à modéliser par STD.

Les archétypes obtenus ne sont pas une représentation statistique de la ville de Nantes, mais une représentation en terme de potentiel de flexibilité des bâtiments nantais, dont la répartition est représentée en Figure 4. Ainsi, plusieurs milliers de bâtiments résidentiels individuels sont représentés par un seul archétype de maison à deux étages, tandis qu'un autre archétype est utilisé pour représenter uniquement les trois grands centres commerciaux de la ville.

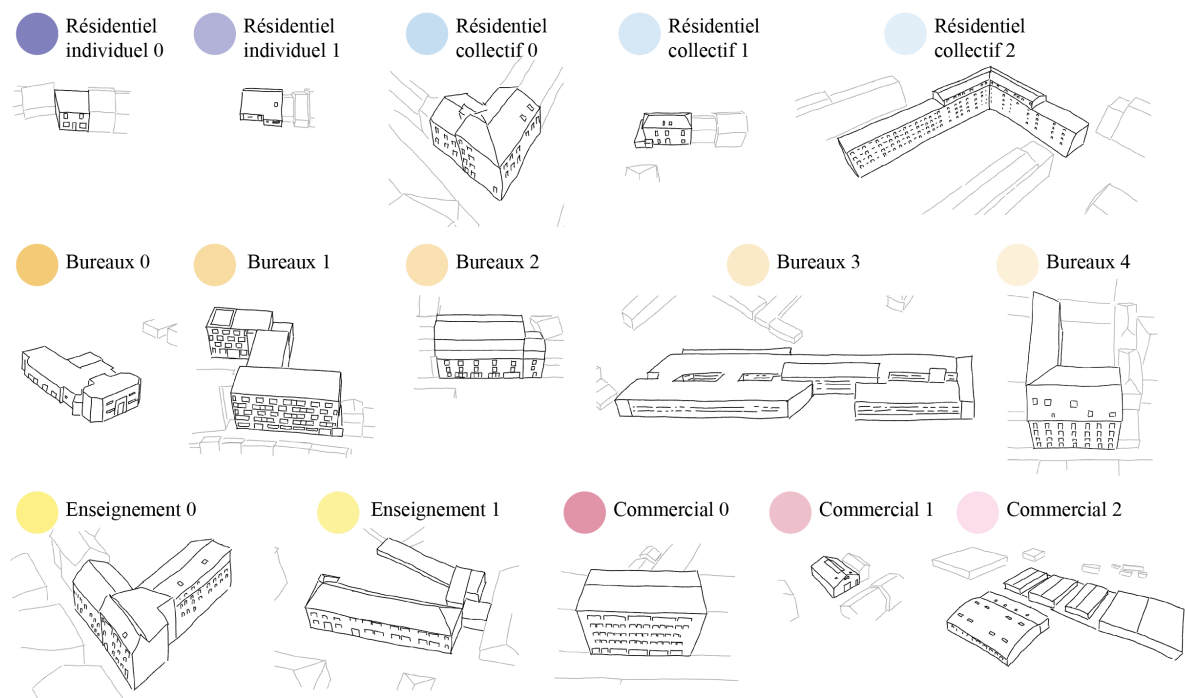


Figure 3 : Répartition des bâtiments de Nantes pour le clustering

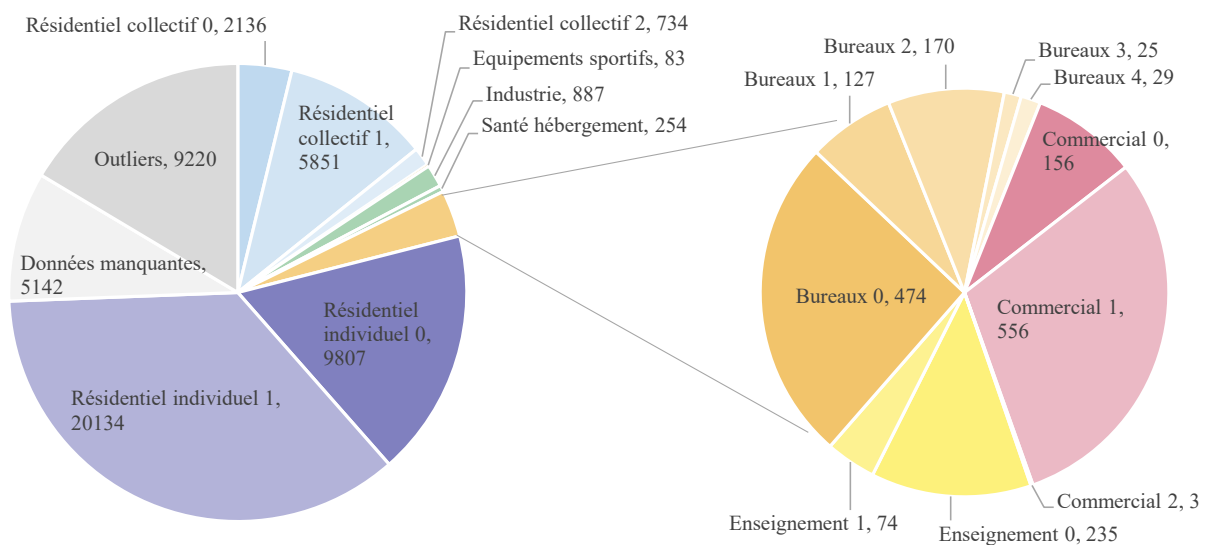


Figure 4 : Répartition des bâtiments de Nantes pour le clustering

## 4. Discussions et perspectives

La méthode développée présente une forte dépendance aux données disponibles, qu'il est nécessaire d'utiliser avec attention, d'autant plus que les techniques de clustering sont très sensibles aux données de base, qui doivent donc être les plus fiables possibles. Un travail conséquent de prétraitement des données est donc indispensable, notamment pour la jointure

entre les différentes BDD utilisées. Ceci a également conduit au choix de répartition préliminaire en sous-usages, qui produit des archétypes plus pertinents vis-à-vis de la flexibilité, mais qui génère aussi une étape supplémentaire. La démarche de clustering se veut également répliquable à d'autres villes, d'où le choix de traiter à cette étape uniquement les caractéristiques géométriques et morphologiques des bâtiments résidentiels et tertiaires, dont les données sont disponibles et aisément accessibles à l'échelle nationale. Les données thermiques des bâtiments, en particulier tertiaires, sont en revanche plus difficiles d'accès, et l'ajout par exemple des années de construction au clustering conduirait à écarter de la démarche environ 30% supplémentaires des bâtiments. Des hypothèses supplémentaires pour les caractéristiques thermiques sont donc nécessaires pour la suite de ce travail pour évaluer les gisements de flexibilité par STD de chacun de ces archétypes (notamment l'inertie). Ceux-ci seront ensuite agrégés à l'échelle de la ville, à partir de modèles de substitution, ou surrogate modeling, qui permettent de prendre en compte les variations de typologies au sein d'un même cluster.

## Références

- [1] A. J. Marszal-Pomianowska, *Principles of energy flexible buildings*, Danish Technological Institute, 2019.
- [2] J. Le Dréau, P. Heiselberg, "Energy flexibility of residential buildings using short term heat storage in the thermal mass", *Energy*, 111, 991-1002, 2016.
- [3] Cerema, *Foisonnement*, Réseaux de chaleur et de froid, <https://reseaux-chaleur.cerema.fr/glossaire/f/foisonnement>, consulté le 20.01.2025, s.d.
- [4] L.Lefort, R. Bonabe de Rougé, P. Schetelat, T. Berthou, B. Duplessis, P. Riederer, E. Peirano, "Méthodologie de validation d'outils de simulation énergétique urbaine avec un cas d'application sur la Seine-et-Marne", *Proceedings IBPSA France 2020, Reims, France*, Nov., 2020.
- [5] M. Ferrando, F. Causone, T. Hong, Y. Chen, "Urban building energy modeling (UBEM) tools : A state-of-the-art review of bottom-up physics-based approaches", *Sustainable Cities and Society*, 62, 102408, 2020.
- [6] I. Ballarini, S. P. Corgnati, V. Corrado, "Use of reference buildings to assess the energy savings potential of the residential building stock : The experience of TABULA project", *Energy Policy*, 68, 273-284, 2014.
- [7] I. D. Stewart, T. Oke, "Local climate zones for urban temperature studies", *Bulletin of the American Meteorological Society*, 93, 1879-1900, 2012.
- [8] M. Bonhomme, *Contribution à la génération de bases de données multi-scalaires et évolutives pour une approche pluridisciplinaire de l'énergie urbaine*, "Thèse de doctorat", Université de Toulouse, 2013.
- [9] J. Vivian, U. Chiodarelli, G. Emmi, A. Zarrella, "A sensitivity analysis on the heating and cooling energy flexibility of residential buildings", *Sustainable Cities and Society*, 52, 101815, 2020.
- [10] S. Martinez, M. Vellei, J. Le Dréau, "Demand-side flexibility in a residential district : What are the main sources of uncertainty ?", *Energy and Buildings*, 255, 111595, 2022.
- [11] A. Nguyen, *Modélisation du parc de bâtiments du secteur tertiaire et simulation énergétique*, "Thèse de doctorat", Université Paris sciences et lettres, 2021.
- [12] IGN, *BD TOPO® — Géoservices*, <https://geoservices.ign.fr/bdtopo>, consulté le 16.01.2025, 2024.
- [13] CSTB, *La Base de Données Nationale des Bâtiments (BDNB)*, <https://bdnb.io/>, consulté le 02.01.2025, 2024.
- [14] Cerema, *Les Fichiers Fonciers*, <https://datafoncier.cerema.fr/fichiers-fonciers>, consulté le 20.01.2025, 2024.
- [15] E. Bocher, J. Bernard, E. Wiederhold, F. Leconte, G. Petit, S. Palominos, C. Noûs, "GeoClimate : a Geospatial processing toolbox for environmental and climate studies", *Journal of Open Source Software*, 6, 3541, 2021.
- [16] J. Singh, D. Singh, "A comprehensive review of clustering techniques in artificial intelligence for knowledge discovery : Taxonomy, challenges, applications and future prospects", 62, 107299, 2024.

## Remerciements

La thèse a été financée avec le soutien d'ENGIE R&I et de GRDF.