



HAL
open science

Identification of a reduced panel of district archetypes for energy policies design and urban planning

Victor Marty-Jourjon, Thomas Berthou, Bruno Duplessis

► To cite this version:

Victor Marty-Jourjon, Thomas Berthou, Bruno Duplessis. Identification of a reduced panel of district archetypes for energy policies design and urban planning. 2022. hal-04225528v1

HAL Id: hal-04225528

<https://hal.science/hal-04225528v1>

Preprint submitted on 2 Oct 2023 (v1), last revised 15 Nov 2023 (v2)

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Identification of a reduced panel of district archetypes for energy policies design and urban planning

Victor MARTY-JOURJON, Thomas BERTHOU, Bruno DUPLESSIS (*)
Center for Energy Efficiency of Systems Mines Paris, PSL University
60 Boulevard Saint-Michel, 75006 Paris, FRANCE
bruno.duplessis@minesparis.psl.eu
(*) corresponding author

Abstract: Districts are now a privileged scale for local authorities in their energy transition policy. Hence, in the last years, diverse methods and models were developed to study energy strategies at this scale. To help the realization of energy studies, urban planners or energy managers often use building or district archetypes. In this paper, in order to identify district archetypes, an original two-step clustering approach is introduced. The approach relies on the k-means algorithm initialized with a priori knowledge from the literature and in line with architectural and urban planning considerations. Different open databases describing buildings, households and urban fabrics are used. A reduced panel of 19 district archetypes, representative of the large territory of Paris area, is identified. These representative districts can be used as case studies for energy policy investigation and allow to extrapolate district-level results to a larger scale. The proposed method is applied on the French region Ile De France and is readily replicable to any other region and where similar data are available.

Highlights :

- A representative panel of archetypal districts is identified for the purpose to support to the energy policies studies
- A two steps district clustering method, considering both urban fabric and residential heating system parameters, is proposed
- The method is applied on the French region Ile-De-France, and identify up to 19 clusters and representative archetypes of districts
- The method is readily replicable to any other French region and other countries where similar data are available

1. Introduction

To mitigate the global warming, the “Paris Agreement” encourages the reduction of energy consumption and the development of local renewable energy sources. Districts are a privileged scale for local authorities in their energy transition policies. At this scale, shared heating energy system such as geothermal heat pump can be used, strategies of demand side management (DSM) can be set-up, global action of retrofitting can be conducted, and urban renewals for mitigating heat island can be achieved.

Several methods and modelling approaches are proposed in the literature in order to investigate energy policies at district scale [1], [2], [3], [4] or [5]). Choices in energy strategies to be implemented vary whether the territory is an urban center, a rural city, or another urban fabric. The urban fabric has been shown to significantly influence the energy consumptions [6] [7]. Thus, most of the time, in order to illustrate their work, the authors select case studies representative of the thermal and energy characteristics of the studied territories and that highlight the capabilities of the tools that have been developed.

For instance, in [8], a neighborhood of the Spanish city Castellon de la Plana, supposed to be representative of the layout in the city is used as a case study to assess a method for estimating energy demand and indoor thermal comfort in buildings. Similarly, in [9], a neighborhood of the Italian city Catania has been selected as a case of study to illustrate a model mapping the energy consumption of buildings, transports

and outdoor lighting of neighborhoods. This neighborhood was chosen because it is characterized by a high diversity of land uses and building characteristics (e.g., ages, morphologies), and is described as representative of many Italian cities. Such characteristics allow to highlight the methodological and scientific advances presented in the article.

Thus, a statically defined panel of representative districts can be used as case studies for energy policy investigation. Besides it can allow to extrapolate at larger scales, consequences of energy policies, evaluated locally with Urban Building Energy Modelling platforms such as SMART-E [10], DIMOSIM [11], UMI [12] or City Energy Analyst [13]. For example, the potential of energy saving resulting from a retrofitting strategy could be investigated at district scale and estimated for a large region, thanks to the statistical weight of the representative district. At last, a set of energy management strategies could be associated to the representative districts and used by stakeholder as guidelines for energy policies.

In this paper, a methodology is proposed for identifying a reduced panel of districts representative of a larger territory to support energy policies investigations. In section 2, through the review of the state-of-the-art of district classification, the most appropriate approach is selected. Then, the selected method is detailed while applying in Ile-de-France region in section 3. Finally, the results are presented in section 4.

2. Review of literature

The identification of a reduced panel of districts representative of a larger territory requires selecting a classification method. In the literature, several non-automatic district classifications have been proposed [14], [15], [16]. In 1999, Theurer has conducted a classification of the urban fabrics of 3 German cities and studied the relations between these urban fabrics and the air quality. The method is based on photos and street investigations giving information about the height and width of the buildings, their arrangement, the spacings between them and the width of the streets.

An automatic urban classification has been proposed in 1996 by [17]. Three parameters are calculated for a sample of neighbors: the height, the contiguity, and the buildings density. For each parameter, thresholds are fixed, and the resulting classes are combined to obtain the urban classification. This method appears to be inadequate when the number of parameters increases.

In the following years, the availability of geographic databases allows to gather higher amount of parameters and to achieve more detailed classifications [18], [19]. In 2005, Long & Kergomard proposed a 10 parameters automatic classification of the districts of Marseille by using the k-means method [18]. The city is divided into 6300 squares with sides of 200 m. 9 classes of neighbors are finally identified. This classification has been realized in order to study the influence of urban fabric on heat islands.

Despite the development of databases and automatic classification tools, it has been highlighted in the literature that an urban classification cannot be achieved without taking into account the historical process of construction [20], [21]. The authors proposed a theoretical panel of 7 types of building blocks for describing the French urban fabric, based on various urban analysis and interviews with urban planners and architects. These 7 typical blocks are identified among the neighborhoods (200 m side squares) of Paris and its suburbs. Three methods of statistical classification have been compared: K-means, self-organizing map, and principal component analysis (PCA). The K-mean method features the best statistic indicators; the clusters are homogenous and more clearly separated from each other. However, the method does not allow the identification of the 7 a priori defined reference urban blocks used by urban planners and architects. Therefore, the classification was carried out using a PCA method. The results are more in line with urban planning but is not fully automated. The classification outputs are thereafter used to model future urban expansions.

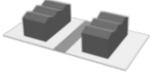
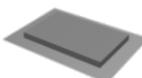
Continuous pavilion	
Discontinuous pavilion	
Continuous block	
Discontinuous block	
High-rise tower	
Ancient center	
Industrial building	

Figure 1: The 7 typical blocks identified by [21]

According to this literature review, the approach of using a statistical classification method to identify well-known categories of districts has been selected. Including a priori knowledge concerning the historical process of construction eases the interpretation of the results and the communication with the stakeholders of urban planning as well as it may ease the initialization step of the statistical classification process. The present article extends beyond the precedent works by identifying these districts using the K-means automatic clustering algorithm. [22], [23] or [24] have demonstrated the adequacy of clustering methods for archetype identification, especially the k-means methods which are easy to use and have low computation times. In a second step, an original statistical sub-classification will be proposed and conducted on each cluster resulting from the first classification. The objective of this second classification is to portray more precisely the specificities of the districts in terms of energy supply and identifying district archetypes for energy studies purposes.

3. Method

In the first instance, well-known categories of districts are identified through a statistical clustering algorithm. Then, a second classification is conducted on the resulting clusters considering heating systems parameters. The French region Ile-de-France has been chosen as the study area. This area knows a homogeneous climate and covers a large panel of territories, from small villages to old and dense urban centers. For both steps, the classifications are conducted with the unsupervised K-means clustering algorithm of Hartigan-Wong [25]. The Euclidian distance is used since this distance is appropriated for the quantitative features [26].

In this section, after introducing the databases, the territorial division, and the IT_tools, the two steps of the classification method are presented.

3.1 Databases

The data used in the study are listed in Table 1. They are issued from the French National Institute of Statistics and Economic Studies (INSEE), and the National Geographic Institute (IGN), providing territorials data at national scale. The data issued from the INSEE, come from the national census, fed by local surveys conducted among the population covering all municipal territories over a five-year period. The surveys deal with socio-economic information regarding households as well as their dwelling. The IGN publishes the BDTOPO® database illustrated in Figure 2 which contains geolocalised geometric shapes organized by themes:

buildings, road networks, vegetation zones. The data are updated every 3 to 5 years, according to the aerial photographs.

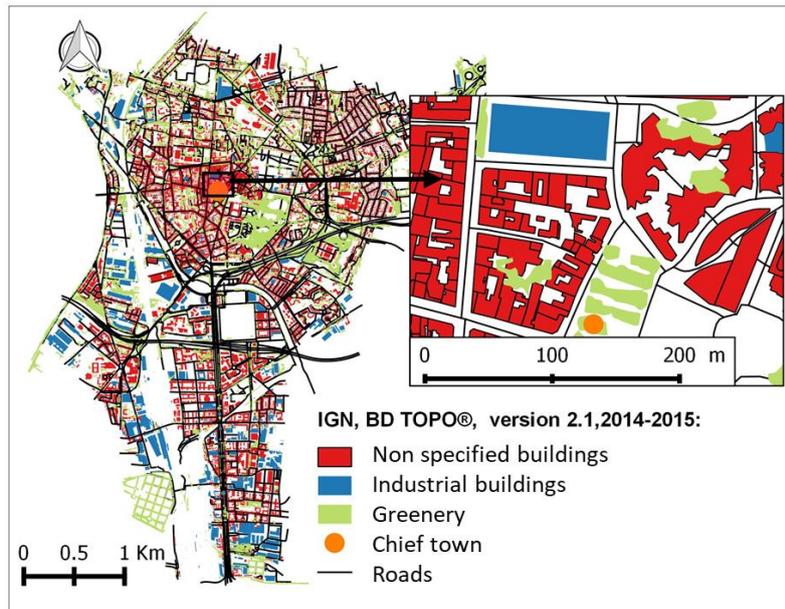


Figure 2: Illustration of the data BDTOPO® on Saint-Denis city (93200 - France)

Databases	Specific issues	Description / data	Ref
Population census	INSEE, « Population en 2013 - IRIS - France hors Mayotte, Recensement de la population - Base infracommunale (IRIS) », 2016	Structure of the population: sex, age, nationality, socio-professional category...	1
Population census	INSEE, « Logements en 2013 - IRIS - France hors Mayotte, Recensement de la population - Base infracommunale (IRIS) », 2016	Housing stock and households: building construction date, surface, move-in date, possession of a personal car, number of collective housing...	2
BD TOPO	IGN, « Contours...Iris® », version 2.0, 2014	Outline of the IRIS (Figure 3)	3
BD TOPO	IGN, BATI_INDIFFERENCIE, BD TOPO®, version 2.1, 2014-2015	Non-specified buildings, defined by a ground surface and a height.	4a
BD TOPO	IGN, BATI_INDUSTRIEL, BD TOPO®, version 2.1, 2014-2015	Industrial buildings, defined by a ground surface and a height.	4b
BD TOPO	IGN, ROUTE, BD TOPO®, version 2.1, 2014-2015	Roads, decomposed in a multitude of segments characterized by a width and a type.	5
BD TOPO	IGN, ZONE_VEGETATION, BD TOPO®, version 2.1, 2014-2015	Surfaces of greenery.	6
BD TOPO	IGN, CHEF_LIEU, BD TOPO®, version 2.1, 2014-2015	Chief town localisations.	7
Population census	INSEE Logement, Fichier détail Recensement de la population, 2013	Detailed results of the population census where each line refers to a housing characterized by the weight and others attributes such as the main heating fuel category.	8

Table 1: Data used in the study

3.2 The territorial division: IRIS

The information given by the census, concerning both the structure of the population and the housings, is supplied at a territorial division named IRIS, introduced by the INSEE. The territorial outline of the IRIS (Figure 3) results of a collaboration between the INSEE and the IGN: it respects geographic and demographic criteria and shows clearly identifiable and stable demarcations in the long term INSEE differentiates the IRIS in 3 different types:

- the “residential” type: population generally falls between 1,800 and 5,000. The unit is homogeneous in terms of living environment and the boundaries of the unit are based on the major dividing lines provided by the urban fabric (main roads, railways, bodies of water etc.),
- the “business” type: containing more than 1,000 employees, with at least twice as many employees as other residents,
- the “miscellaneous” type: specific large zones which are sparsely inhabited and have a large surface area (leisure parks, ports, forests etc.).

By featuring coherent urban fabrics, the IRIS division eases the identification of the district archetypes. Besides, this territorial division is familiar to the urban planning stakeholders and covers the whole French territory.

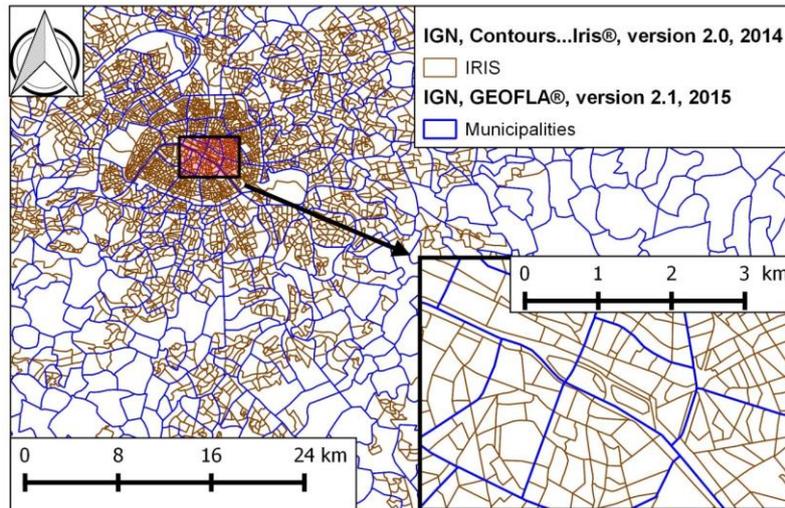


Figure 3: Territorial outlines: municipalities (blue) and IRIS (black) of Ile-de-France Region (France)

For all these reasons, the IRIS territorial division has been chosen as the basic individual for the classification approaches proposed below.

3.3 IT tools

The method has been processed with open-source tools (Figure 4). The geographic and statistical data are stored on the database management system (DBMS) PostgreSQL. The extension PostGIS allows to process geographic data. The data are called and processed with the programming language Python. The library Psycopg2 is used to interface with the DBMS. Then the statistic classification is realized with the free and open-source integrated development environment (IDE) for R: RStudio®. At last, the results are mapped with the open-source software Quantum GIS (QGIS). The extension Openlayer is used to display the background satellite map issued by Google®.

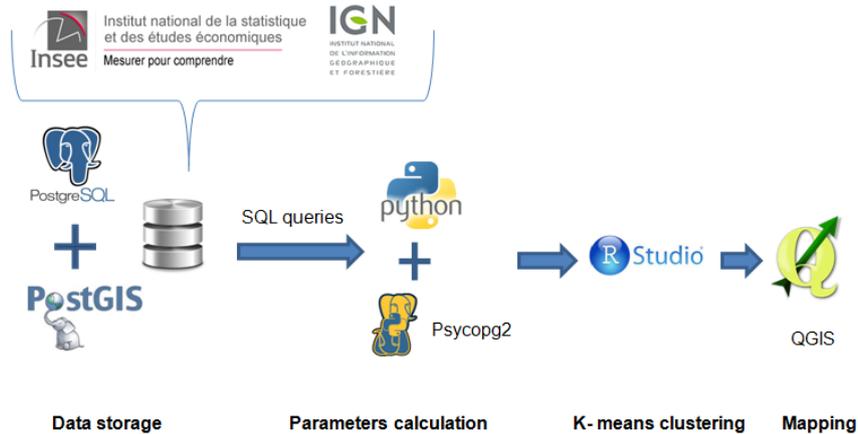


Figure 4: IT tools used in the study

3.4 First classification for identifying urban district archetypes

From the typical blocks proposed by [21], seven categories of districts are suggested (Table 2).

District types (IRIS scale)	Main characteristics	Concordance with the typical blocs of [21]
<u>Old urban center</u>	Very characteristic of the architecture of Hausmann, the urban fabric features a very high density of buildings. The buildings have around 7 floors and many courtyards.	Ancient center
<u>Dense urban district</u>	This district type refers to a dense urban fabric, mixing the urban fabric “Old urban center” and “medium building”. Located between the urban centers and the suburbs, this urban fabric is usually named as the “Faubourgs” and mainly made of continuous blocks	Continuous block
<u>Rural area</u>	This district type refers to villages mainly composed of individual housings.	Discontinuous pavilion
<u>Large housings</u>	These districts are specific to the urban fabric of large cities. It is composed of large bars and towers of housings, built around 1960 and 1970. They are usually located at the periphery of the city center.	Discontinuous block
<u>Urban periphery</u>	These districts are mainly composed of individual housings developed between the large cities and the rural areas. It is mainly composed of individual housings and features collective buildings.	Continuous pavilions Discontinuous pavilions Discontinuous block
<u>High-rise buildings</u>	These districts are composed of the high-rise office or housing buildings (mainly built in the 70's for housing).	High-rise tower
<u>Industrial area</u>	These districts are composed of industrial, commercial, and even agricultural buildings. They are mainly located at the periphery of cities.	Industrial building

Table 2: A priori defined urban district archetypes for Ile-de-France region

From databases presented in section 3.1, urban fabrics characteristics have been extracted for each IRIS(see Figure 3): characteristics of buildings (e.g., orientation, age, blocks compacity and density, ground surface, continuity height), of roads (e.g., density, lengths, width) and surroundings (e.g., vegetation, industry rate, courtyard rate). All these parameters have been adapted from the comprehensive inventory of parameters used to characterize urban fabrics as in [20].

Names	Parameters description	References to Table 1
Height (m)	Average height of buildings, weighted by the ground surfaces.	3, 4a
Heights-std (m)	Standard deviation of building heights.	3, 4a
Industry rate (unitless)	Sum of the ground surfaces of industrial buildings divided by the sum of the ground surfaces of all buildings.	3, 4a, 4b
Building density (unitless)	Sum of the floor surfaces divided by the IRIS surface. <i>Building floor surface = Building ground surface * (Building height/3)</i>	3, 4a, 4b
Building floor surface (m ²)	Average floor surface of buildings.	3, 4a
Block floor surface (m ²)	Average floor surface of blocks. Juxtaposed buildings are defined as a block <i>Block floor surface=sum(Building floor surfaces)</i>	3, 4a
Envelope surface (unitless)	Sum of envelope surfaces divided by the IRIS surface. <i>Building envelope surface= Vertical surface–Shared surface</i>	3, 4a
Building compactness (unitless)	Average compactness of buildings, weighted by the ground surfaces. <i>Building compactness = (Building envelope surface+ Building ground surface)/(Building ground surface* Building height)^(2/3)</i>	3, 4a
Building compactness-std (unitless)	Standard deviation of the compactness of buildings, weighted by the ground surfaces.	3, 4a
Block compactness (unitless)	Average compactness of blocks, weighted by the ground surfaces.	3, 4a
Courtyard rate (unitless)	Sum of the Courtyard surfaces divided by the ground surfaces. The Courtyard surfaces are the ground surfaces surrounded by buildings.	3, 4a
Contiguity (unitless)	Average contiguity, weighted by the ground surfaces. <i>Building contiguity= Shared surface/Vertical surface</i>	3, 4a
Contiguity-std (unitless)	Standard deviation of Building contiguities, weighted by the ground surfaces.	3, 4a
Building orientation-std (°)	Standard deviation of orientations of buildings. The orientation is calculated as the average of the orientations of every segment composing the ground surface, weighted by the length of segments.	3, 4a
Buidings by block (unitless)	Average of number of buildings by block.	3, 4a
Age (unitless)	In the data base, for every IRIS, the number of housings is detailed for 6 sections of year of construction. In every section the distribution of housings is supposed to be uniform. The age parameter is calculated as the year where the cumulative percentage of the building distribution reaches 50%.	2
Collective dwellings (unitless)	Number of collective dwellings divided by the total number of housings.	2
Cars (unitless)	Number of cars divided by the number of households.	2
Road density (unitless)	Sum of the surface of roads divided by the IRIS surface.	3, 5
Road orientation-std (°)	Standard deviation of the road orientations, weighted by the road lengths. The road orientation is calculated as the average of the orientations of every segments of the road weighted by the length of the segments.	3, 5
Road length density (m ⁻¹)	Sum of the length of roads divided by the IRIS surface.	3, 5
Road width (m)	Average width of roads, weighted by the lengths.	3, 5
Road width-std (m)	Standard deviation of road widths, weighted by the lengths.	3, 5
Building density (m ⁻¹)	Number of buildings divided by the IRIS surface.	3, 4a
Ground surface (unitless)	Sum of the Building ground surfaces divided by the IRIS surface.	3, 4a
Greenery (unitless)	Sum of the surfaces of greeneries divided by the IRIS surface.	3, 6
Center distance (m)	Average distance between buildings and the closest chief town of a municipality.	3, 4a, 7

Net inhabitant density (hab/m ²)	Number of inhabitants divided by the sum of the floor surfaces.	3, 4a, 1
Gross inhabitant density (hab/m ²)	Number of inhabitants divided by the IRIS surface.	3, 1

Table 3 : Parameters selected for the first classification, describing the urban fabrics of IRIS

Before processing the classification, redundant parameters are removed through the analysis of the matrix of Pearson correlation coefficients (PCC) (Figure 5): the parameters “Ground surface”, “Road length” and “Envelope surface” presenting PCCs higher than 0.9, have been removed. The districts having no residential housing and an industrial surface area below 1000 m² at IRIS scale have been also removed.

Thus, the K-means algorithm is applied to the 26 columns matrix of the IRIS of Ile-de-France. The parameters are centered (average = 0) and scaled (standard deviation=1) to have identical weights. Besides, 7 districts have been manually selected for the initialization of the clustering algorithm: this eases the identification of the urban district archetypes, especially for category “High-rise buildings”, which gather few individuals at IRIS scale and is thus more difficult to be identified by the algorithm.

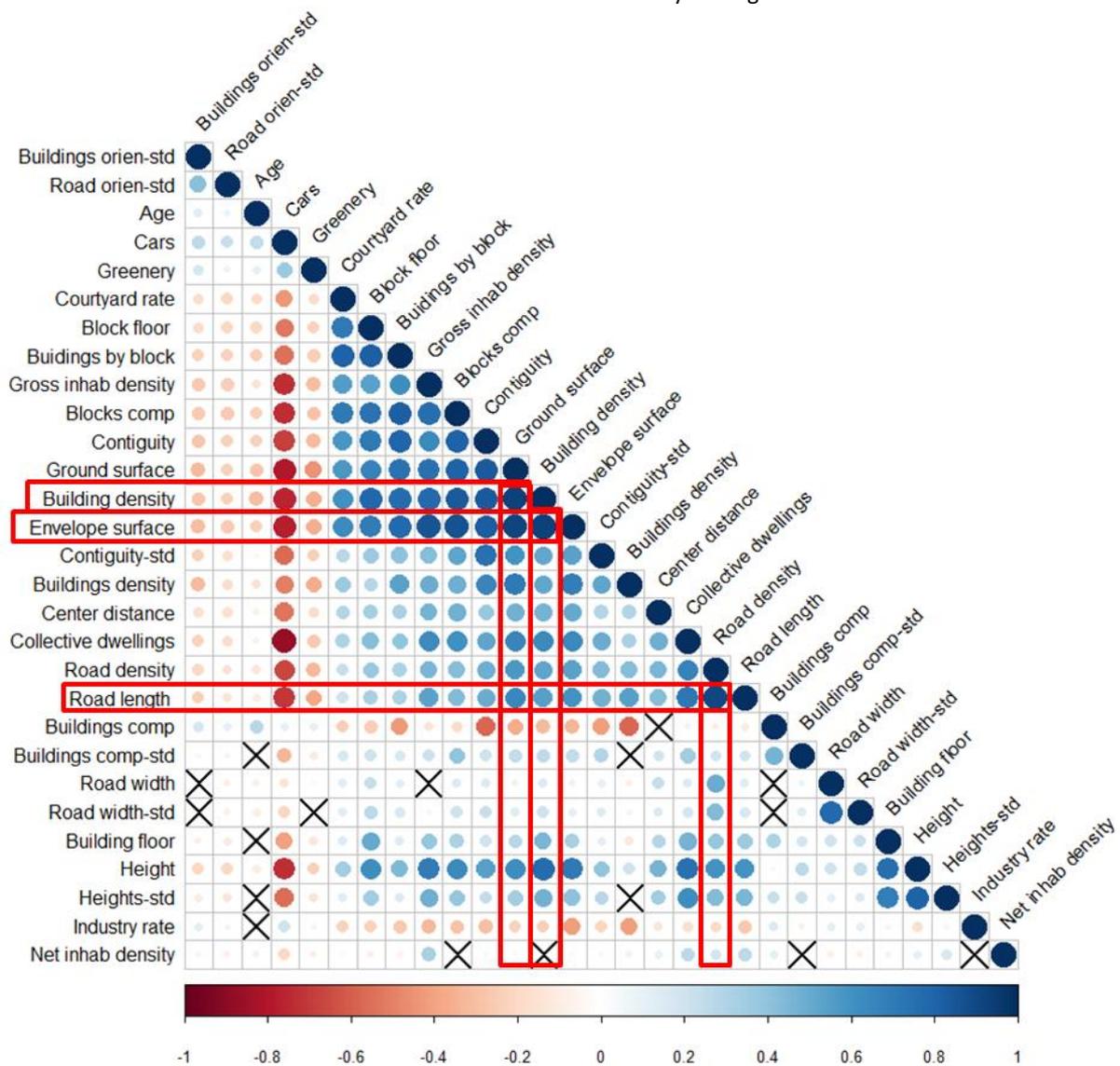


Figure 5: Matrix of Pearson correlation coefficients

3.5 Sub-classification of urban district archetypes based on parameters describing heating energy sources

A second step of classification is conducted on each cluster resulting from the first application of the K-means algorithm. The parameters considered in this second classification are presented in Table 4: Parameters selected for second classification; they describe the main heating source of energy used in residential housings and are calculated from the national census database (reference [8] in Table 1). Because of the small number of dwellings, the group "Industrial area" is not considered, similarly, the category "High-rise buildings" is discarded because it contains too few individuals to allow a reliable classification.

Name	Definition	Source (Table 1)
district heating	District heating heated housing surface divided by the total housing surface	8
natural gas (network)	Natural gas heated housing surface divided by the total housing surface	8
electricity	Electricity heated housing surface divided by the total housing surface	8
other	Fuel oil + bottled gas + other heated housing surface divided by the total housing surface	8

Table 4: Parameters selected for second classification, describing the heating energy sources

Unlike the previous classification step, where 7 categories of urban districts are a priori defined and statistically identified, this second classification step does not presuppose any number of categories. For each sub-clustering, the numbers of clusters set in the K-means algorithm is chosen by analyzing the indicator WSS: within group sum of squares:

$$WSS = \sum_{i=1}^k \sum_{x \in C_i} \sum_{j=1}^n (x_j - m_{i,j})^2 \quad (1)$$

k: number of clusters

C_i: cluster *i*

x_j: value of the parameters *j* for district *x*

m_{i,j}: barycentre of the cluster *i* for the parameter *j*

n: number of parameters: 4

A small WSS value reflects good homogeneity of the districts within the clusters, which naturally gets better when the clusters number increase. For each sub-classification, the analysis of WSS as a function of the number of clusters allows for a trade-off between increasing homogeneity within clusters and the need to keep a reduced panel of district archetypes. For instance, for the urban cluster "Old urban center", a slope discontinuity in the WSS curve is observed for a 4 clusters classification (Figure 6).

In order to obtain an optimal result, for each sub-classification, the K-means algorithm is applied 150 times with randomly initialized centers; the solutions which minimize the WSS are chosen.

Through this process, 19 archetypal districts have been identified. The results are presented in the following paragraph.

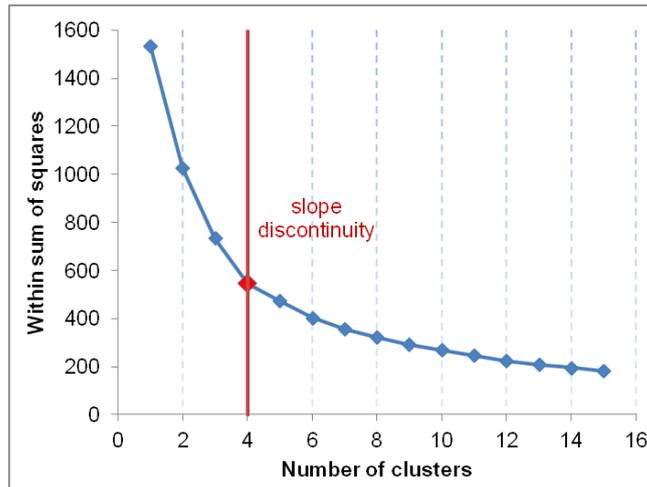


Figure 6: Curve of the WSS indicator, in function of the number of clusters considered the sub-classification of the districts “Old urban center”

4. Results

The first classification allows to characterize the seven urban district archetypes a priori defined (Table 2). The 7 resulting clusters are presented in Table 5. They are characterized by the mean values of the urban fabric parameters (column 3). The statistical weights of each archetype correspond to the number of IRSI in the cluster, given in column 2. Each archetype is pictured with the closest IRIS to the mean center in column 4. The geographic distribution of the district archetypes is mapped in Figure 7.

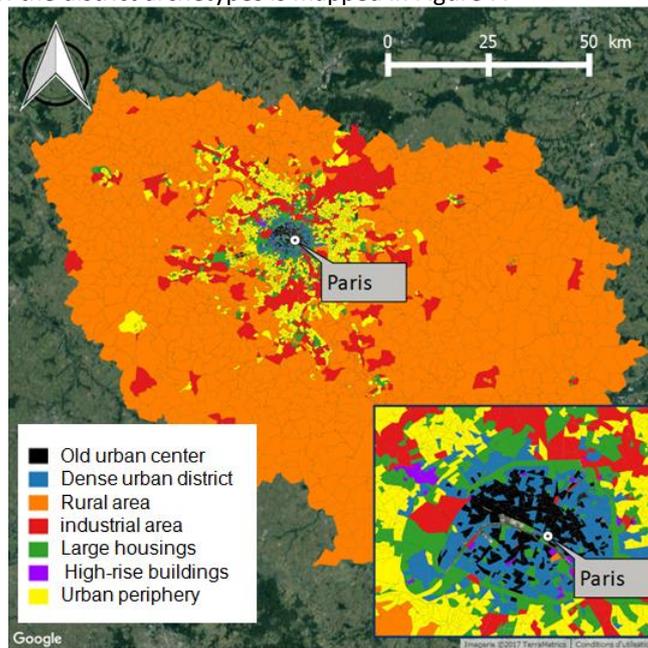


Figure 7: Map of the 7 urban district archetypes identified in the Ile-de-France region at the IRIS scale

The “Old urban center” archetype presents a very high building density and is composed with more than 90% of collective housings, “pre-modern” buildings built before 1940. Buildings include 5 to 7 floors and are characterized by many courtyards and high contiguity. This cluster is portrayed by a district in the center of Paris.

In the cluster “Dense urban district” the old buildings are mixed with buildings achieved after the urban renovation plan of 1975. The density, the contiguity and especially the number of buildings by blocks are significantly lower than for the cluster “Old urban center”.

The cluster “High-rise buildings” gathers only 38 individuals (less than 1% of the IRIS population in the Ile-de-France region). This archetype is very singular: the average height of buildings is 41 m and the building density is the highest among the district archetypes. The parameter “age” has a value of 1976. The construction of these high buildings follows the new safety regulations adapted to the high-rise buildings, set up in 1967. Most of these districts are in the business center of Paris - La Défense.

With a middle building density (more than 4 times less than the first three district archetypes), the “Urban periphery” archetype presents a balanced distribution between collective and individual dwellings. These districts are the consequence of the urban periphery development (see Figure 7). They are linked to the cities by important transport infrastructures.

With a building density of a similar order of magnitude, the "Large Buildings" cluster is mainly composed of collective housing (more than 90% of the housing stock) and located around the Paris ring road or in the close suburbs. The districts of this archetype kind are characterized by large towers and bars of collective dwellings and have been built between the 50's and the 90's.

With the lowest building density, the “Rural area” archetype is mainly composed of individual (more than 80%). The closest district to the statistic center is the municipality of Héricy – a 2500 inhabitants town.

At last, the cluster “Industrial area” is characterized by very high proportion of industrial buildings (0,65). These districts are mainly located around the cities and close to transport infrastructures.

District archetype name	Number of individuals (districts) in IdF Region	Values of the main parameters of the mean centers	Mean center Closest IRIS to the mean center
Old urban center	393 (7.5%)	Age : 1902 Density : 3,16 Contiguity : 0,38 Courtyard rate : 6,13E-03 Buildings by block : 19,3 Collective dwelling : 0,96 Road density : 0,12 Road width : 5,3 Green surface rate : 0,089 Industry rate : 0,019	Paris 6e IRIS : 751062307 
Dense urban district	653 (12.5%)	Age : 1948 Density : 2,17 Contiguity : 0,27 Courtyard rate : 1,60E-03 Buildings by block : 7,5 Collective dwelling : 0,95 Road density : 0,1 Road width : 4,6 Green surface rate : 0,13 Industry rate : 0,059	Paris 20e IRIS : 751207912 
High-rise buildings	38 (0.7%)	Age : 1976 Density : 3,27 Contiguity : 0,155 Courtyard rate : 3,16E-04 Buildings by block : 2,3 Collective dwelling : 0,93 Road density : 0,13 Road width : 5,2 Green surface rate : 0,13 Industry rate : 0,17	Courbevoie IRIS : 920260406 

Urban periphery	1579 (30%)	Age : 1968 Density : 0,49 Contiguity : 0,098 Courtyard rate : 1,37E-05 Buildings by block : 1,7 Collective dwelling : 0,61 Road density : 0,079 Road width : 4,4 Green surface rate : 0,16 Industry rate : 0,094	Antony IRIS : 920020102 
Large housings	857 (16.4%)	Age : 1966 Density : 0,87 Contiguity : 0,067 Courtyard rate : 1,93E-04 Buildings by block : 1,7 Collective dwelling : 0,94 Road density : 0,1 Road width : 4,9 Green surface rate : 0,18 Industry rate : 0,15	Bagneux IRIS : 920070103 
Rural area	1310 (25%)	Age : 1968 Density : 0,0 Contiguity : 0,023 Courtyard rate : 6,77E-06 Buildings by block : 1,1 Collective dwelling : 0,17 Road density : 0,02 Road width : 4,5 Green surface rate : 0,3 Industry rate : 0,19	Héricy IRIS : 772260000 
Industrial area	393 (7.5%)	Age : 1973 Density : 0,5 Contiguity : 0,065 Courtyard rate : 1,08E-05 Buildings by block : 1,4 Collective dwelling : 0,56 Road density : 0,07 Road width : 5,2 Green surface rate : 0,14 Industry rate : 0,65	Grigny 912860107 

Table 5 : Clusters resulting first classification (Ile-de-France region, France)

Regarding the energy systems, the results of the second classification are presented in Table 6. For each sub-cluster, the number of IRIS is given, and the repartition of the heating energy sources is plotted. The sub-clusters are named according to the main urban district archetype to which they belong, and the main heating energy source used in the sub-cluster. The exhaustive list of sub-cluster parameters is presented in detail in the appendix.



Table 6: Clusters resulting from the sub-classifications

The “Old urban center” archetype is sub-divided in 4 groups, 2 of which gathers 75% of the districts, and are characterized by high rates of electric heating. In most of cases, these ancient buildings rely on individual heating systems, mostly individual electric heaters but also individual gas boilers.

In contrast, the “Dense urban” district archetype is distributed in 3 sub-clusters, homogeneous in size, respectively dominated by district heating, natural gas, and electricity as heating energy.

In the “Urban periphery” archetype, the natural gas is the main heating energy; this system is strongly represented in each sub-groups of this urban archetype. The collective housings are more recent than in the urban centers, but the building density is not high enough for district heating.

The building density is more favorable for district heating in the “Large housings” archetype. The sub-group LH-urban presents the highest rates of district heating: 37% of the IRSI are mainly heated by district heating. The natural gas is the other main heating energy source for collective heating systems in large buildings.

Finally, the “Rural area” archetype presents the highest rate of housings heated by the system “other”: this can be explained by the low density of buildings, and long distances between the buildings and the energy infrastructures (especially for the natural gas).

5. Conclusion

In this paper a two-steps method of classification has been proposed to identify a reduce panel of representative district for energy policies studies and has been applied on the region Ile De France.

In a first instance, through the k-means clustering method applied on the 5000 IRIS of the region Ile-de-France, characterized by 26 urban fabric parameters, 7 clusters of urban archetypes have been identified: Old urban center (OUC), Dense urban district (DUD), Rural area (RA), Large housings (LH), Urban periphery (UP), High-rise buildings (HRB), Industrial area (IA). Although the selected urban fabric parameters can be discussed, the resulting clusters keep consistent by marginally adding or removing parameters and are congruent with the district archetypes a priori defined. Those archetypes have been defined based on the literature and are in line with architectural and urban planning considerations. Considering a priori knowledge concerning the historical process of construction eases the statistical classification, and the communication with the stakeholders of

urban planning. By opposition to previous works from the literature, the implementation of the k-means methods gives robust results at the IRIS scale and thus allows determining in an automatic way the distribution of the districts among the archetypes. The territorial division IRIS turned out to be a relevant scale for districts classification. This division allows an easier identification of the urban fabric than with a grid layout. Besides, on this mesh many geographical and statistical data are available on the entire French territory. The method can be easily replicated on any other French region to characterize a larger territory or in any country with similar data available.

The second step of classification allows to consider parameters that portray more precisely the specificities of the districts in terms of energy use. On the base of the 7 clusters identified with urban fabric parameters, a sub-clustering has been conducted considering 4 parameters describing the main heating energy source in the dwellings: 19 sub-clusters have been identified. The resulting sub clusters can be interpreted according to the energy supply constraints of each district archetype.

The statistical centers of the 19 resulting clusters can be used to study local energy policies such as the deployment of energy storage systems or new domestic heating systems. Through the respective weights of the representative district, consequences of energy policies, evaluated locally, can be extrapolated at the scale of the region Ile-de-France. A set of energy management strategies could be associated to the representative districts and used by stakeholder as guidelines for energy policies.

In further work, it could be interesting to consider additional parameters influencing the energy use such as the surface of office building and small businesses. This method can also be adapted, by considering in the sub-clustering phase, new parameters to study other issues of urban planning such as the heat islands.

6. Appendix: statistical centers of the 19 sub-clusters resulting from the second classification

Old urban center	OUC-urbain	OUC-other	OUC-elec	OUC-gaselec
Number of IRIS	57	40	130	157
% in the urban cluster	15%	10%	34%	41%
% of the total	1%	1%	3%	3%
Heating energy sources parameters				
District heating	0,34	0,11	0,05	0,06
Town gas	0,28	0,45	0,25	0,42
Electricity	0,32	0,26	0,67	0,47
Other	0,07	0,18	0,03	0,04
Urban fabric parameters				
Age	1906	1924	1913	1911
Density	3,00	3,05	3,23	3,16
Contiguity	0,34	0,35	0,42	0,38
Courtyard rate	6,5E-03	5,6E-03	6,5E-03	5,9E-03
Buildings by block	17,6	17,3	21,9	19,0
Collective dwelling	0,95	0,95	0,97	0,96
Road density	0,11	0,12	0,12	0,12
Road width	5,1	5,7	5,0	5,4
Green surface rate	0,11	0,13	0,07	0,09
Industry rate	0,039	0,004	0,007	0,012

Dense urban district	DUD-urban	DUD-gas	DUD-elec
Number of IRIS	147	248	252

% in the urban cluster	23%	38%	39%
% of the total	3%	5%	5%
Heating energy sources parameters			
District heating	0,45	0,12	0,08
Town gas	0,24	0,48	0,3
Electricity	0,27	0,29	0,58
Other	0,05	0,11	0,04
Urban fabric parameters			
Age	1961	1946	1942
Density	2,26	2,14	2,16
Contiguity	0,24	0,26	0,29
Courtyard rate	1,7E-03	1,4E-03	1,7E-03
Buildings by block	6,7	7,2	8,2
Collective dwelling	0,97	0,95	0,95
Road density	0,09	0,10	0,10
Road width	4,6	4,6	4,6
Green surface rate	0,14	0,14	0,11
Industry rate	0,060	0,059	0,052

Rural area	RA-urban	RA-elec	RA-other	RA-gas
Number of IRIS	34	484	288	301
% in the urban cluster	3%	44%	26%	27%
% of the total	1%	10%	6%	6%
Heating energy sources parameters				
District heating	0,21	0,01	0	0,02
Town gas	0,46	0,35	0,05	0,6
Electricity	0,19	0,44	0,39	0,24
Other	0,14	0,21	0,56	0,14
Urban fabric parameters				
Age	1970	1975	1963	1972
Density	0,15	0,07	0,01	0,14
Contiguity	0,02	0,03	0,01	0,03
Courtyard rate	1,5E-05	3,3E-06	7,0E-06	3,0E-06
Buildings by block	1,1	1,1	1,1	1,2
Collective dwelling	0,65	0,18	0,05	0,29
Road density	0,04	0,02	0,01	0,04
Road width	4,7	4,6	4,3	4,5
Green surface rate	0,46	0,30	0,25	0,37
Industry rate	0,165	0,191	0,199	0,136

Large housings	LH-urban	LH-gas	LH-elec	LH-other
Number of IRIS	307	341	118	73
% in the urban cluster	37%	41%	14%	9%
% of the total	6%	7%	2%	1%

Heating energy sources parameters				
District heating	0,66	0,22	0,12	0,32
Town gas	0,2	0,56	0,28	0,21
Electricity	0,08	0,14	0,56	0,08
Other	0,05	0,08	0,04	0,39
Urban fabric parameters				
Age	1968	1965	1971	1968
Density	0,88	0,82	1,04	0,73
Contiguity	0,06	0,07	0,09	0,05
Courtyard rate	1,9E-04	1,4E-04	2,3E-04	8,7E-05
Buildings by block	1,7	1,7	1,9	1,4
Collective dwelling	0,96	0,94	0,93	0,95
Road density	0,10	0,10	0,10	0,10
Road width	4,8	4,9	5,0	5,0
Green surface rate	0,17	0,19	0,15	0,19
Industry rate	0,151	0,140	0,166	0,149

Urban periphery	UP-gas	UP-other	UP-urban	UP-elec
Number of IRIS	736	250	122	470
% in the urban cluster	47%	16%	8%	30%
% of the total	15%	5%	2%	9%
Heating energy sources parameters				
District heating	0,07	0,05	0,43	0,04
Town gas	0,64	0,47	0,32	0,41
Electricity	0,2	0,24	0,18	0,47
Other	0,09	0,24	0,08	0,08
Urban fabric parameters				
Age	1965	1964	1974	1972
Density	0,46	0,50	0,49	0,54
Contiguity	0,10	0,08	0,09	0,11
Courtyard rate	7,8E-06	1,1E-05	1,2E-05	2,4E-05
Buildings by block	1,7	1,5	1,7	1,8
Collective dwelling	0,58	0,53	0,77	0,64
Road density	0,08	0,08	0,08	0,08
Road width	4,4	4,6	4,2	4,5
Green surface rate	0,16	0,17	0,15	0,16
Industry rate	0,090	0,087	0,105	0,101

7. References

- [1] J. Allegrini, K. Orehounig, G. Mavromatidis, F. Ruesch, V. Dorer, and R. Evins, 'A review of modelling approaches and tools for the simulation of district-scale energy systems', *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1391–1404, Dec. 2015, doi: 10.1016/j.rser.2015.07.123.

- [2] J. Keirstead, M. Jennings, and A. Sivakumar, 'A review of urban energy system models: Approaches, challenges and opportunities', *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3847–3866, Aug. 2012, doi: 10.1016/j.rser.2012.02.047.
- [3] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, and J. Widén, 'Urban building energy modeling: State of the art and future prospects', *Renewable and Sustainable Energy Reviews*, vol. 128, p. 109902, Aug. 2020, doi: 10.1016/j.rser.2020.109902.
- [4] M. Ferrando, F. Causone, T. Hong, and Y. Chen, 'Urban Building Energy Modeling (UBEM) Tools: A State-of-the-Art Review of Bottom-up Physics-Based Approaches', *Sustainable Cities and Society*, p. 102408, Jul. 2020, doi: 10.1016/j.scs.2020.102408.
- [5] M. Kavagic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic, 'A review of bottom-up building stock models for energy consumption in the residential sector', *Building and Environment*, vol. 45, no. 7, pp. 1683–1697, Jul. 2010, doi: 10.1016/j.buildenv.2010.01.021.
- [6] A. Vartholomaios, 'A parametric sensitivity analysis of the influence of urban form on domestic energy consumption for heating and cooling in a Mediterranean city', *Sustainable Cities and Society*, vol. 28, pp. 135–145, Jan. 2017, doi: 10.1016/j.scs.2016.09.006.
- [7] S. J. Quan, A. Economou, T. Grasl, and P. P.-J. Yang, 'Computing Energy Performance of Building Density, Shape and Typology in Urban Context', *Energy Procedia*, vol. 61, pp. 1602–1605, 2014, doi: 10.1016/j.egypro.2014.12.181.
- [8] M. Braulio-Gonzalo, M. D. Bovea, M. J. Ruá, and P. Juan, 'A methodology for predicting the energy performance and indoor thermal comfort of residential stocks on the neighbourhood and city scales. A case study in Spain', *Journal of Cleaner Production*, vol. 139, pp. 646–665, Dec. 2016, doi: 10.1016/j.jclepro.2016.08.059.
- [9] A. Fichera, G. Inturri, P. La Greca, and V. Palermo, 'A model for mapping the energy consumption of buildings, transport and outdoor lighting of neighbourhoods', *Cities*, vol. 55, pp. 49–60, Jun. 2016, doi: 10.1016/j.cities.2016.03.011.
- [10] T. Berthou, B. Duplessis, P. Rivière, P. Stabat, D. Casetta, and D. Marchio, 'SMART-E: A TOOL FOR ENERGY DEMAND SIMULATION AND OPTIMIZATION AT THE CITY SCALE', p. 9, 2015.
- [11] P. Riederer, V. Partenay, N. Perez, C. Nocito, R. Trigance, and T. Guiot, 'DEVELOPMENT OF A SIMULATION PLATFORM FOR THE EVALUATION OF DISTRICT ENERGY SYSTEM PERFORMANCES', in *BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India*, 2015.
- [12] N. Buckley, G. Mills, C. Reinhart, and Z. Michael Berzolla, 'Using Urban Building Energy Modelling (UBEM) to support the new European Union's Green Deal: Case study of Dublin Ireland', *Energy and Buildings*, p. 111115, May 2021, doi: 10.1016/j.enbuild.2021.111115.
- [13] J. A. Fonseca, T.-A. Nguyen, A. Schlueter, and F. Marechal, 'City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts', *Energy and Buildings*, vol. 113, pp. 202–226, Feb. 2016, doi: 10.1016/j.enbuild.2015.11.055.
- [14] W. Theurer, 'Typical Building Arrangements for Urban Air Pollution Modelling', *Atmospheric Environment*, vol. 33, no. 24, pp. 4057–4066, Oct. 1999, doi: 10.1016/S1352-2310(99)00147-8.
- [15] R. M. Cionco and R. Ellefsen, 'High resolution urban morphology data for urban wind flow modeling', *Atmospheric Environment*, vol. 32, no. 1, pp. 7–17, Jan. 1998, doi: 10.1016/S1352-2310(97)00274-4.
- [16] S. Pauleit and F. Duhme, 'Assessing the environmental performance of land cover types for urban planning', *Landscape and Urban Planning*, vol. 52, no. 1, pp. 1–20, Nov. 2000, doi: 10.1016/S0169-2046(00)00109-2.
- [17] F. Chaillou, 'Classification morphologique sur trois critères de tissus urbains résidentiels : extraits du plan référentiel de Nantes: Rapport interne', CERMA UMR CNRS 1563, NANTES, 1996.
- [18] N. Long and C. Kergomard, 'Classification morphologique du tissu urbain pour des applications climatologiques. Cas de Marseille', *Revue internationale de géomatique*, vol. 15, no. 4, pp. 487–512, Dec. 2005, doi: 10.3166/ri.15.487-512.
- [19] I. D. Stewart, T. R. Oke, and E. S. Krayenhoff, 'Evaluation of the "local climate zone" scheme using temperature observations and model simulations: EVALUATION OF THE "LOCAL CLIMATE ZONE" SCHEME', *Int. J. Climatol.*, vol. 34, no. 4, pp. 1062–1080, Mar. 2014, doi: 10.1002/joc.3746.
- [20] M. Bonhomme, 'Contribution à la génération de bases de données multi-scalaires et évolutives pour une approche pluridisciplinaire de l'énergétique urbaine', Thèse de doctorat, INSA Toulouse, Université de Toulouse, 2013.

- [21] M. Bonhomme, H. A. Haddou, and L. Adolphe, 'A tool for classifying and modelling evolution of urban typologies', p. 6, 2012.
- [22] U. Ali, M. H. Shamsi, C. Hoare, E. Mangina, and J. O'Donnell, 'A data-driven approach for multi-scale building archetypes development', *Energy and Buildings*, vol. 202, p. 109364, Nov. 2019, doi: 10.1016/j.enbuild.2019.109364.
- [23] U. Ali, M. H. Shamsi, E. Mangina, and J. O'Donnell, 'Comparative Analysis of Machine Learning Algorithms for Building Archetypes Development in Urban Energy Modeling', presented at the 2018 Building Performance Modeling Conference and SimBuild co-organized by ASHRAE and IBPSA-USA, Chicago, 2018.
- [24] I. De Jaeger, G. Reynders, C. Callebaut, and D. Saelens, 'A building clustering approach for urban energy simulations', *Energy and Buildings*, vol. 208, p. 109671, Feb. 2020, doi: 10.1016/j.enbuild.2019.109671.
- [25] J. A. Hartigan and M. A. Wong, 'Algorithm AS 136: A K-Means Clustering Algorithm', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979, doi: 10.2307/2346830.
- [26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning - Data Mining, Inference, and Prediction, 2nd Edition*. Springer New York, NY, 2009. Accessed: Jun. 21, 2022. [Online]. Available: <https://doi.org/10.1007/978-0-387-84858-7>