



**HAL**  
open science

# Identification of district archetypes for energy policy design and urban planning through a two-step clustering method

Victor Marty-Jourjon, Thomas Berthou, Bruno Duplessis

## ► To cite this version:

Victor Marty-Jourjon, Thomas Berthou, Bruno Duplessis. Identification of district archetypes for energy policy design and urban planning through a two-step clustering method. 2022. hal-04225528v2

**HAL Id: hal-04225528**

**<https://hal.science/hal-04225528v2>**

Preprint submitted on 15 Nov 2023

**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 district archetypes for energy policy design and urban planning through a two-step clustering method

Victor MARTY-JOURJON(\*), Thomas BERTHOU, Bruno DUPLESSIS  
Center for Energy Efficiency of Systems  
Mines Paris, PSL University  
60 Boulevard Saint-Michel, 75006 Paris, France  
victor.marty.jourjon@gmail.com  
(\* ) corresponding author

This manuscript is currently under peer review.

**Abstract:** Districts are now a privileged scale for local authorities in their energy transition policies. A variety of techniques and models have been developed in recent years to analyse energy strategies at this scale. In this context, this study introduces an original two-step clustering method to identify district archetypes. These representative districts can be used as case studies for energy policy investigation and urban planning and allow to extrapolate district-level results to a larger scale. The approach relies on the k-means algorithm initialized with a priori knowledge based on the literature and in line with architectural and urban planning considerations. The proposed method is applied on the French region Ile De France. A reduced panel of 19 district archetypes is identified. Various open databases containing information on buildings, households, and urban fabrics are used. This method can be adapted to investigate various district level issues by selecting appropriate parameters.

## Highlights :

- A representative panel of district archetypes is identified to support energy policy studies and urban planning
- A two-step district clustering method, considering both urban fabric and residential heating system parameters, is proposed
- The method identifies, on a French region, 19 archetypes representative of the territory
- This method can be adapted to investigate various district level issues by selecting appropriate parameters

## 1. Introduction

Districts are a privileged scale for local authorities in their energy transition policies. At this scale, local energy sources can be used through district heating system, strategies of demand side management (DSM) can be set-up, global action of retrofiting can be conducted, and urban renewals for mitigating heat island can be achieved.

Several methods and modelling approaches are proposed in the literature to investigate energy policies at district scale. (Allegrini et al. 2015; Ferrando et al. 2020; Johari et al. 2020; Kavagic et al. 2010; Keirstead, Jennings, and Sivakumar 2012). The type of district (e.g : urban center, a rural city ) influences the choice of the energy strategies. In (Quan et al. 2014; Vartholomaios 2017), the authors show the influence of the urban fabric on the energy consumption of the districts. Thus, most of the time, to illustrate their work, the authors select case studies, that are representative of a large territory, and highlight the capacities of the proposed methods and models. For instance, in (Braulio-Gonzalo et al. 2016), a neighbourhood of the Spanish city Castellon de la Plana is used as a case study to assess a method for estimating energy demand and indoor thermal comfort in buildings. This district is supposed to be representative of the whole city. Similarly, in (Fichera et al. 2016), a neighbourhood 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 neighbourhoods. This neighbourhood is characterized by a high diversity of land uses and buildings (e.g., ages, morphologies). Such characteristics allow to highlight the capacities of the presented model. Besides, the district is described as representative of many Italian cities. Finally, in (Fahed et al. 2020), the authors present the impact of some UHI (urban heat island) mitigation strategies through numerical microclimate simulations of an existing Lebanese district of Dora in Beirut. The selected case study is described as a dense urban morphology that is repeated in most areas of the region.

In all these examples, a panel of district archetypes, statistically defined, and representative of a large territory, can be a great asset for case studies regarding energy policy investigation and urban planning. The case study can be selected according to the statistical description of the districts. The results, evaluated locally (e.g : the potential of energy saving resulting from a district retrofiting), can be extrapolated at larger scales, with the statistical weight of the archetypes. At last, a set of energy policies could be associated to each of the representative district's archetypes and used by stakeholder as guidelines for urban 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. Literature review

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 (Cionco and Ellefsen 1998; Pauleit and Duhme 2000; Theurer 1999). 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 (Chaillou 1996). Three parameters are calculated for a sample of neighbours: the height, the contiguity, and the density of buildings. 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 number of parameters and to achieve more detailed classifications (Long and Kergomard 2005), (Stewart, Oke, and Krayenhoff 2014). In 2005, Long & Kergomard proposed a 10-parameter automatic classification of the districts of Marseille by using the k-means method (Long and Kergomard 2005). The city is divided into 6300 squares with sides of 200 m. 9 classes of neighbours are finally identified. This classification has been realized 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 considering the historical process of construction (Bonhomme 2013; Bonhomme, Ait Haddou, and Adolphe 2012). 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 neighbourhoods (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 are not fully automated. The classification outputs are thereafter used to model future urban expansions.

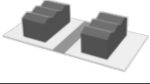
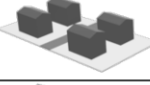




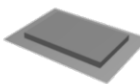
Continuous pavilion	
Discontinuous pavilion	
Continuous block	
Discontinuous block	
High-rise tower	
Ancient center	
Industrial building	

Figure 1: The 7 typical blocks identified by (Bonhomme et al. 2012)

### 3. Method

According to the 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 statistical classification, the interpretation of the results, and the communication with the stakeholders of urban planning. For this first step, the present article extends beyond the precedent works by identifying these districts using the K-means automatic clustering algorithm of (Hartigan and Wong 1979). (Ali et al. 2018, 2019; Bonhomme et al. 2012; De Jaeger et al. 2020) have demonstrated the adequacy of clustering methods for archetype identification, especially the K-means methods which are easy to use and have low computation time. The K-means process is guided by manually defining the districts used for initialization. In the second step, an original statistical sub-classification is 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 according to the study undertaken. Here, this second classification is achieved with parameters describing the residential energy systems. Such parameters are relevant for conducting analysis with Urban Building Energy Modelling platforms such as SMART-E (Berthou et al. 2015), DIMOSIM (Riederer et al. 2015), UMI (Buckley et al. 2021) or City Energy Analyst (Fonseca et al. 2016). The method is applied on the large French region Ile-de-France. This area knows a homogeneous climate and covers a large panel of territories, from small villages to old and dense urban centres.

### 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 territorial 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 using aerial photographs.

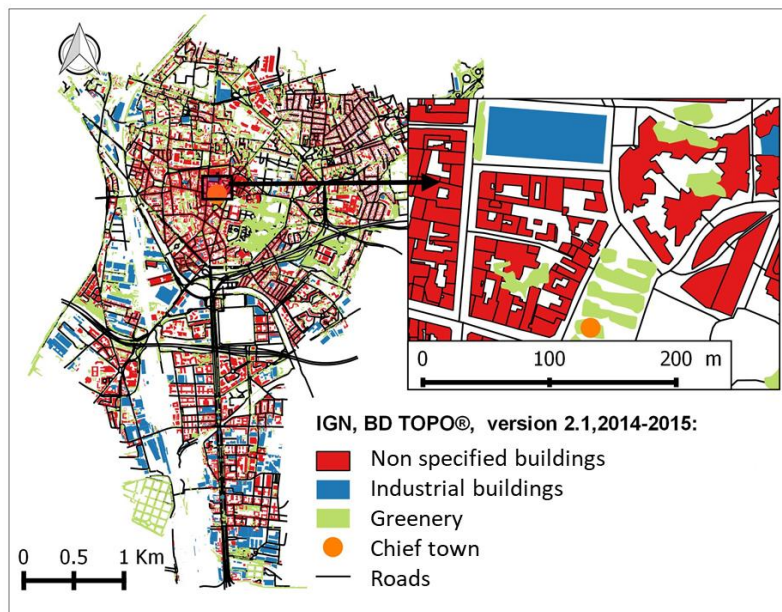


Figure 2: Illustration of the data BDTOPO® on Saint-Denis city (93200 - France)

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 housings...	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,	Chief town localisations.	7

	2014-2015		
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 other attributes such as the main heating fuel category.	8

Table 1: Data used in this article

### 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 outline of the IRIS (Figure 3) is a result of a collaboration between the INSEE and the IGN: it respects geographic and demographic criteria and shows clearly identifiable and stable demarcations. 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. Hence, the IRIS division has been chosen for the classification approaches proposed below.

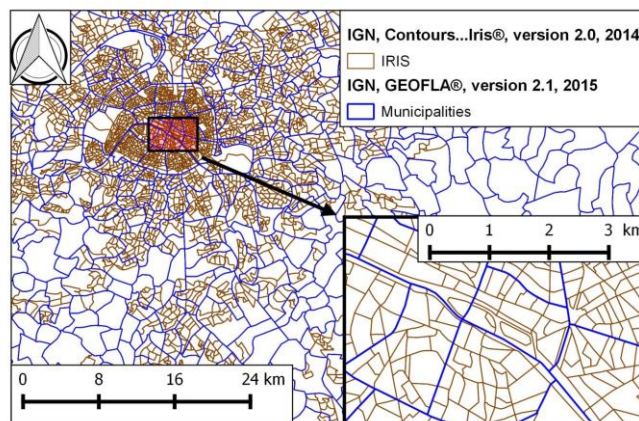


Figure 3: Territorial outlines: municipalities (blue) and IRIS (brown) of Ile-de-France Region (France)

### 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®.

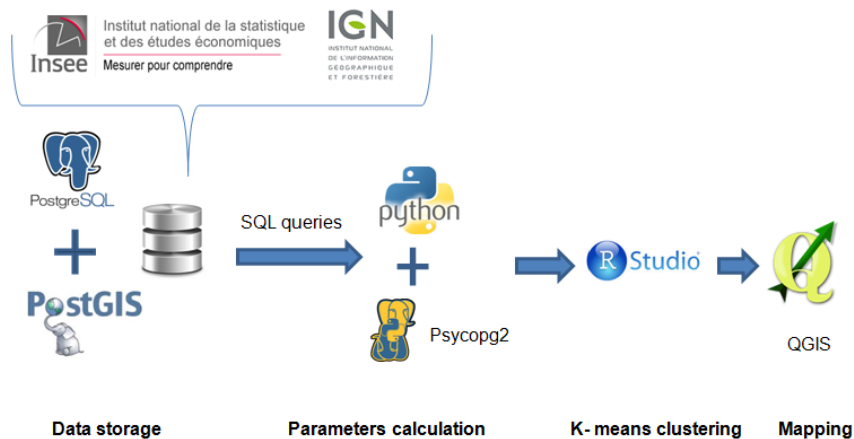


Figure 4: IT tools used in the study

### 3.4 First classification for identifying urban district archetypes

From the typical blocks proposed by (Bonhomme et al. 2012), seven categories of districts are suggested (Table 2).

District types (IRIS scale)	Main characteristics	Concordance with the typical blocs of (Bonhomme et al. 2012)
<b>Old urban centre</b>	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 centre
<b>Dense urban district</b>	This district type refers to a dense urban fabric, mixing the urban fabric “Old urban centre” and “medium building”. Located between the urban centres and the suburbs, this urban fabric is usually named as the “Faubourgs” and mainly made of continuous blocks	Continuous block
<b>Rural area</b>	This district type refers to villages mainly composed of individual housings.	Discontinuous pavilion
<b>Large housings</b>	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 centre.	Discontinuous block



<b>Urban periphery</b>	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
<b>High-rise buildings</b>	These districts are composed of the high-rise office or housing buildings (mainly built in the 70's).	High-rise tower
<b>Industrial area</b>	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 the databases presented in section 3.1, urban fabrics characteristics have been calculated for each IRIS (Figure 3): characteristics of buildings (e.g., orientation, age, blocks compactness and density, ground surface, continuity height), 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 (Bonhomme et al. 2012).

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 <sup>2</sup> )	Average floor surface of buildings.	3, 4a
Block floor surface (m <sup>2</sup> )	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	Average of number of buildings by block.	3, 4a

(unitless)		
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 <sup>-1</sup> )	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 <sup>-1</sup> )	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 <sup>2</sup> )	Number of inhabitants divided by the sum of the floor surfaces.	3, 4a, 1
Gross inhabitant density (hab/m <sup>2</sup> )	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 m<sup>2</sup> 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 centred (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.

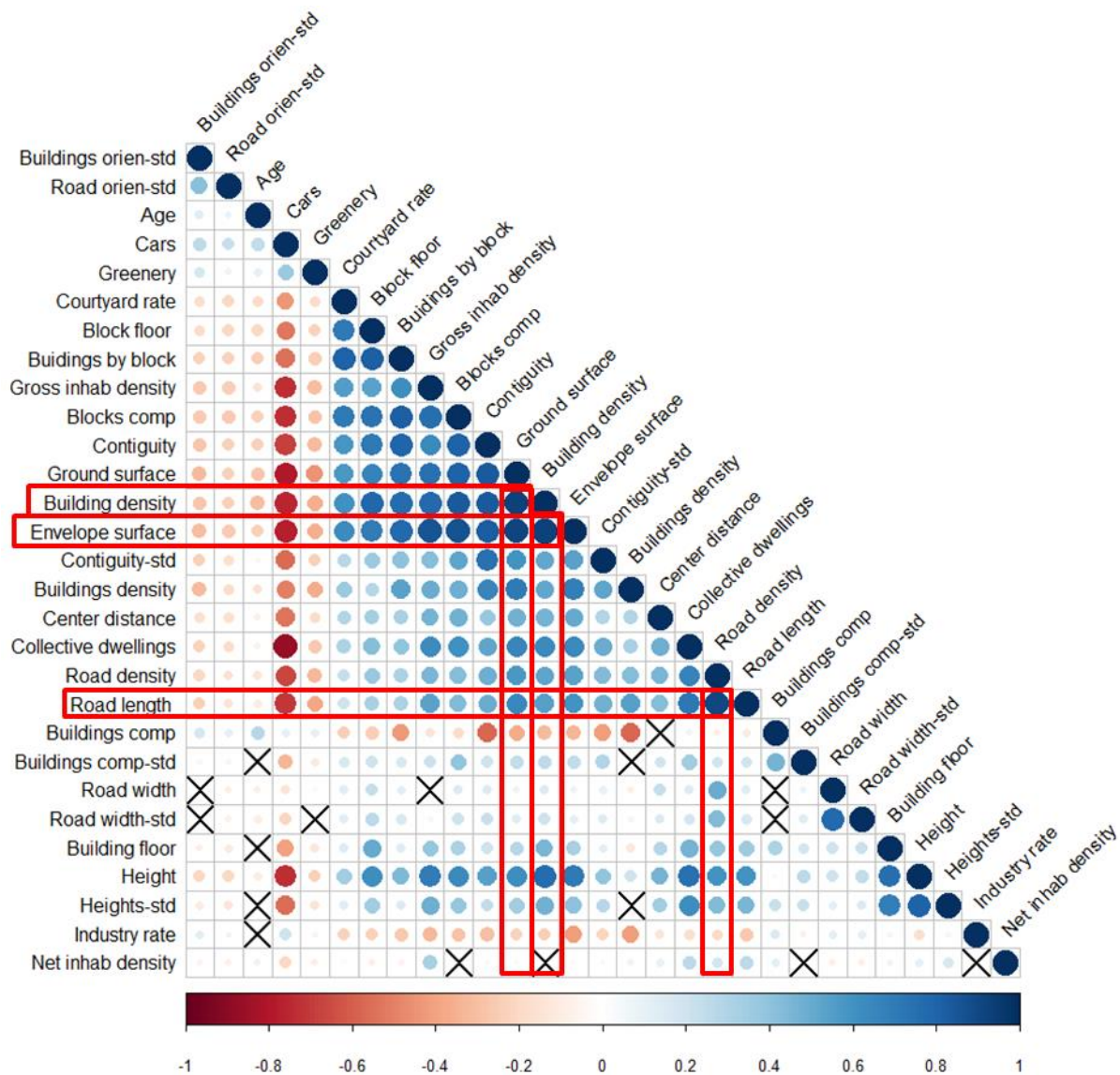


Figure 5: Matrix of Pearson correlation coefficients

### 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; they describe the main heating source of energy used in residential housings and are calculated from the national census database (reference (Braulio-Gonzalo et al. 2016) 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 analysing 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 number of clusters increases. 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).

To obtain an optimal result, for each sub-classification, the K-means algorithm is applied 150 times with randomly initialized centres; 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.

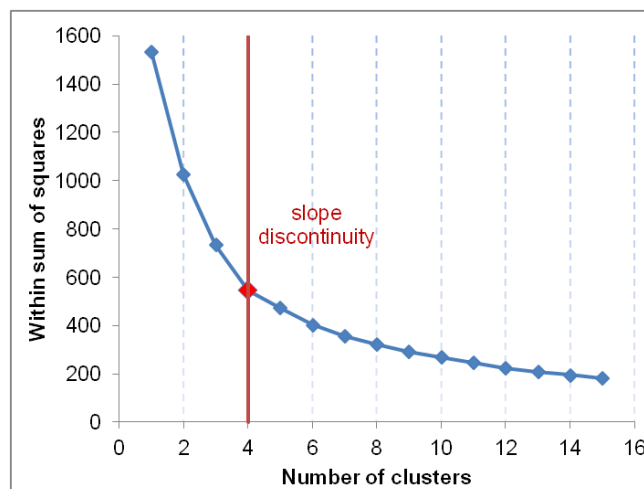


Figure 6: Curve of the WSS indicator, in function of the number of clusters considered the

## sub-classification of the districts “Old urban centre”

### 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 weight of each archetype is given in column 2; it corresponds to the number of IRIS in the cluster. In column 4, the archetypes are pictured with the closest IRIS to the mean centre. The geographic distribution of the district archetypes is mapped in Figure 7.

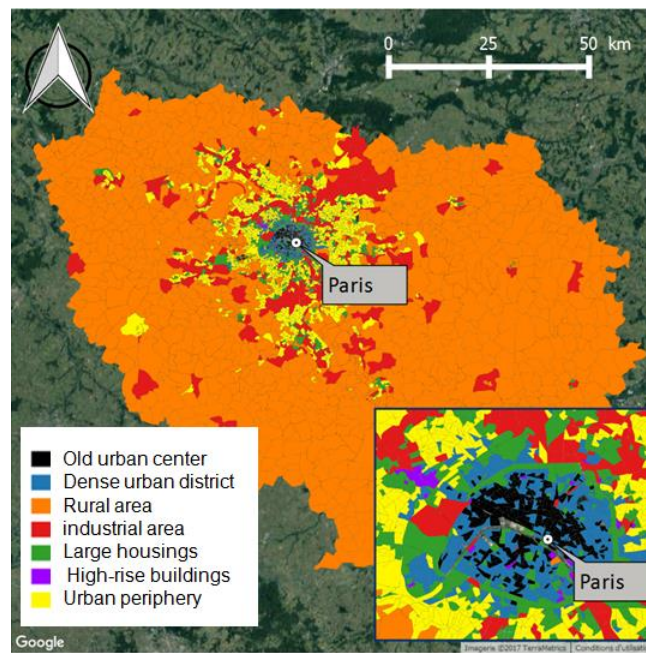


Figure 7: Map of the 7 urban district archetypes identified in the Ile-de-France region at the IRIS scale

The “Old urban centre” 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 centre of Paris.

In the cluster “Dense urban district”, the old buildings are mixed with those built 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 centre”.



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”, defined in Table 3, 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 centre 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 dwellings (more than 80%). The closest district to the statistic centre 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 are 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
<b>Old urban center</b>	<b>393 (7.5%)</b>	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 
<b>Dense urban district</b>	<b>653 (12.5%)</b>	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 








<b>High-rise buildings</b>	<b>38 (0.7%)</b>	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 
<b>Urban periphery</b>	<b>1579 (30%)</b>	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 
<b>Large housings</b>	<b>857 (16.4%)</b>	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 
<b>Rural area</b>	<b>1310 (25%)</b>	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 
<b>Industrial area</b>	<b>393 (7.5%)</b>	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. An exhaustive list of sub-cluster parameters is presented in detail in the appendix.

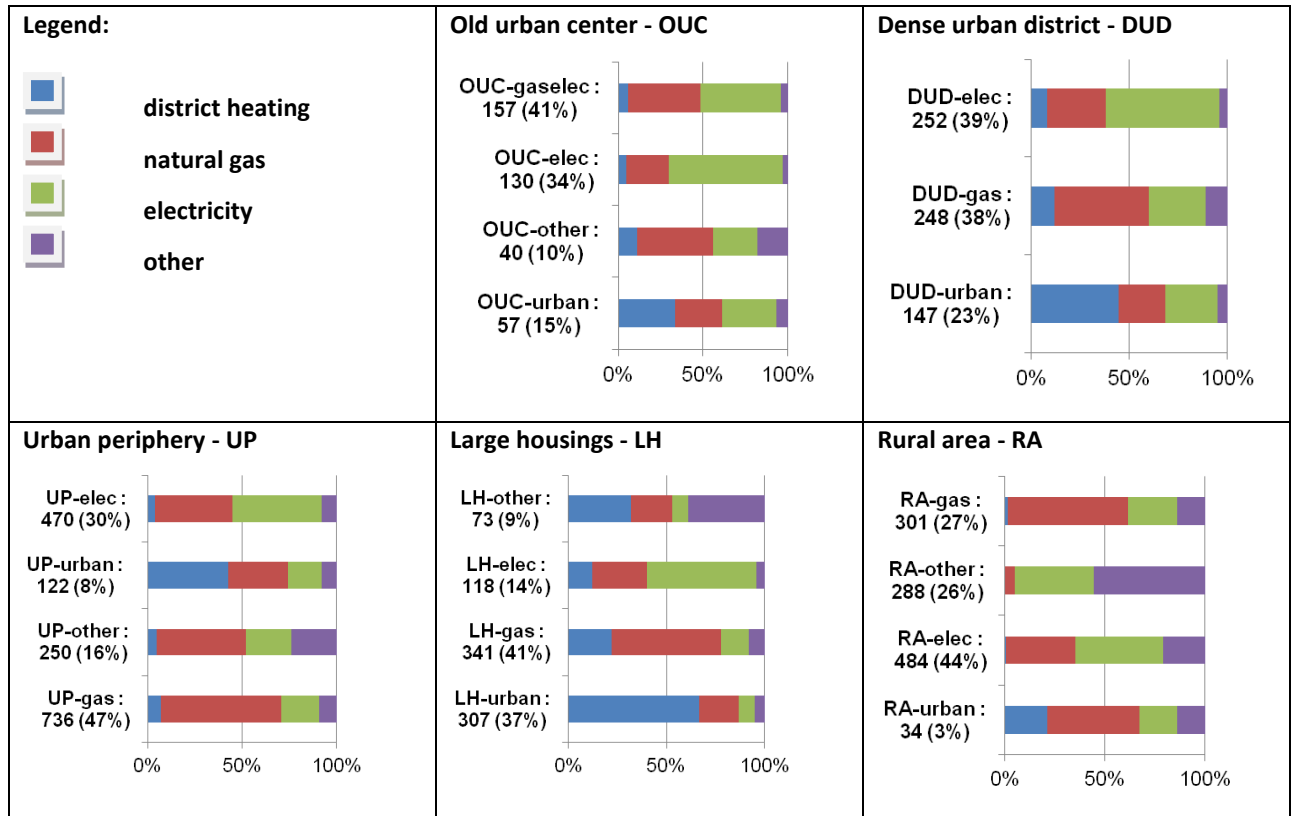


Table 6: Clusters resulting from the sub-classifications

The “Old urban centre” archetype is sub-divided in 4 groups, 2 of which gather 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 source; this system is strongly represented in each sub-groups of this urban archetype. The collective housings are more recent than in the urban centres, but the building density is not high enough for district heating.

The building density is more favourable for district heating in the “Large housings” archetype. The sub-group LH-urban presents the highest rates of district heating: 37% of the IRIS 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-step method of classification has been proposed to identify a reduced panel of representative district for energy policy studies. This method is presented on the case study of the Ile-de-France region.

In the first step, 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 centre (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 remain 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 method gives robust results at the IRIS scale and thus, allows determining the distribution of the districts among the archetypes in an automatic way. 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 division 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 statistical centres 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 with Urban Building Energy Models (UBEM), 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 centres of the 19 sub-clusters resulting from the second classification

Old urban centre	OUC-urban	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	<b>0,060</b>	<b>0,059</b>	<b>0,052</b>

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

<b>Green surface rate</b>	0,17	0,19	0,15	0,19
<b>Industry rate</b>	0,151	0,140	0,166	0,149

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

## 7. References

- Ali, Usman, Mohammad Haris Shamsi, Cathal Hoare, Eleni Mangina, and James O'Donnell. 2019. "A Data-Driven Approach for Multi-Scale Building Archetypes Development." *Energy and Buildings* 202:109364. doi: 10.1016/j.enbuild.2019.109364.
- Ali, Usman, Mohammad Haris Shamsi, James O'Donnell, Fawaz Alshehri, and Eleni Mangina. 2018. "Comparative Analysis of Machine Learning Algorithms for Building Archetypes Development in Urban Energy Modeling."
- Allegrini, Jonas, Kristina Orehounig, Georgios Mavromatidis, Florian Ruesch, Viktor Dorer, and Ralph Evins. 2015. "A Review of Modelling Approaches and Tools for the Simulation of District-Scale Energy Systems." *Renewable and Sustainable Energy Reviews* 52:1391–1404. doi: 10.1016/j.rser.2015.07.123.
- Berthou, Thomas, Bruno Duplessis, Philippe Rivière, Pascal Stabat, Damien Casetta, and Dominique Marchio. 2015. "Smart-E : A Tool for Energy Demand Simulation And Optimization At the City Scale."
- Bonhomme, Marion. 2013. "Contribution to the Generation of Multiscalar and Evolutionary Databases for a Multidisciplinary Approach to Urban Energy." INSA Toulouse.
- Bonhomme, Marion, Hassan Ait Haddou, and L. Adolphe. 2012. "Genius: A Tool for Classifying and Modelling Evolution of Urban Typologies."
- Braulio-Gonzalo, Marta, María Dolores Bovea, María José Ruá, and Pablo Juan. 2016. "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* 139:646–65. doi: 10.1016/j.jclepro.2016.08.059.
- Buckley, Niall, Gerald Mills, Christoph Reinhart, and Zachary Michael Berzolla. 2021. "Using Urban Building Energy Modelling (UBEM) to Support the New European Union's Green Deal: Case Study of Dublin Ireland." *Energy and Buildings* 247:111115. doi: 10.1016/j.enbuild.2021.111115.
- Chaillou, F. 1996. *Classification Morphologique Sur Trois Critères de Tissus Urbains Résidentiels : Extraits Du Plan Référentiel de Nantes.*

- Cionco, Ronald M., and Richard Ellefsen. 1998. "High Resolution Urban Morphology Data for Urban Wind Flow Modeling." *Atmospheric Environment* 32(1):7–17. doi: 10.1016/S1352-2310(97)00274-4.
- Fahed, Jeff, Elias Kinab, Stephane Ginestet, and Luc Adolphe. 2020. "Impact of Urban Heat Island Mitigation Measures on Microclimate and Pedestrian Comfort in a Dense Urban District of Lebanon." *Sustainable Cities and Society* 61:102375. doi: 10.1016/j.scs.2020.102375.
- Ferrando, Martina, Francesco Causone, Tianzhen Hong, and Yixing Chen. 2020. "Urban Building Energy Modeling (UBEM) Tools: A State-of-the-Art Review of Bottom-up Physics-Based Approaches." *Sustainable Cities and Society* 62:102408. doi: 10.1016/j.scs.2020.102408.
- Fichera, Alberto, Giuseppe Inturri, Paolo La Greca, and Valentina Palermo. 2016. "A Model for Mapping the Energy Consumption of Buildings, Transport and Outdoor Lighting of Neighbourhoods." *Cities* 55:49–60. doi: 10.1016/j.cities.2016.03.011.
- Fonseca, Jimeno A., Thuy-An Nguyen, Arno Schlueter, and Francois Marechal. 2016. "City Energy Analyst (CEA): Integrated Framework for Analysis and Optimization of Building Energy Systems in Neighborhoods and City Districts." *Energy and Buildings* 113:202–26. doi: 10.1016/j.enbuild.2015.11.055.
- Hartigan, J. A., and M. A. Wong. 1979. "Algorithm AS 136: A K-Means Clustering Algorithm." *Applied Statistics* 28(1):100. doi: 10.2307/2346830.
- De Jaeger, Ina, Glenn Reynders, Chadija Callebaut, and Dirk Saelens. 2020. "A Building Clustering Approach for Urban Energy Simulations." *Energy and Buildings* 208:109671. doi: 10.1016/j.enbuild.2019.109671.
- Johari, F., G. Peronato, P. Sadeghian, X. Zhao, and J. Widén. 2020. "Urban Building Energy Modeling: State of the Art and Future Prospects." *Renewable and Sustainable Energy Reviews* 128:109902. doi: 10.1016/j.rser.2020.109902.
- Kavgic, M., A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic. 2010. "A Review of Bottom-up Building Stock Models for Energy Consumption in the Residential Sector." *Building and Environment* 45(7):1683–97. doi: 10.1016/j.buildenv.2010.01.021.
- Keirstead, James, Mark Jennings, and Aruna Sivakumar. 2012. "A Review of Urban Energy System Models: Approaches, Challenges and Opportunities." *Renewable and Sustainable Energy Reviews* 16(6):3847–66. doi: 10.1016/j.rser.2012.02.047.
- Long, Nathalie, and Claude Kergomard. 2005. "Classification Morphologique Du Tissu Urbain Pour Des Applications Climatologiques. Cas de Marseille." *Revue Internationale de Géomatique* 15(4):487–512. doi: 10.3166/ri.15.487-512.
- Pauleit, Stephan, and Friedrich Duhme. 2000. "Assessing the Environmental Performance of Land Cover Types for Urban Planning." *Landscape and Urban Planning* 52(1):1–20. doi: 10.1016/S0169-2046(00)00109-2.
- Quan, Steven Jige, Athanassions Economou, Thomas Grasl, and Perry Pei-Ju Yang. 2014. "Computing Energy Performance of Building Density, Shape and Typology in Urban Context." *Energy Procedia* 61:1602–5. doi: 10.1016/j.egypro.2014.12.181.
- Riederer, Peter, Vincent Partenay, Nicolas Perez, Christophe Nocito, Romain Trigance, and Thierry Guiot. 2015. "Development of a Simulation Platform for the Evaluation of District Energy System Performances."
- Stewart, Iain D., T. R. Oke, and E. Scott Krayenhoff. 2014. "Evaluation of the 'Local Climate Zone' Scheme Using Temperature Observations and Model Simulations." *International Journal of Climatology* 34(4):1062–80. doi: 10.1002/joc.3746.
- Theurer, W. 1999. "Typical Building Arrangements for Urban Air Pollution Modelling." *Atmospheric Environment* 33(24–25):4057–66. doi: 10.1016/S1352-2310(99)00147-8.
- Vartholomaios, Aristotelis. 2017. "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* 28:135–45. doi: 10.1016/j.scs.2016.09.006.