Building types in France

Clustering building morphometrics using national spatial data

Alessandro Araldi[1], **David Emsellem**[2], **Giovanni Fusco**[1], **Andrea Tettamanzi**[3], **Denis Overal**[2]

1. *ESPACE, Université Côte d'Azur, CNRS*
   98, bd Herriot, BP 3209, 06200 Nice, France
   {alessandro.araldi, giovanni.fusco@univ-cotedazur.fr
2. *KCITYLABS - Kinaxia Group*
   80, route des Lucioles, 06560 Valbonne, France
   {david.emsellem, denis.overal]@kcitylabs.fr
3. *I3S, Inria, CNRS, Université Côte d'Azur*
   2000, route des Lucioles, 06900 Sophia Antipolis, France
   andrea.tettamanzi@univ-cotedazur.fr

ABSTRACT. The identification and description of building typologies play a fundamental role in the understanding of the overall built-up form. A growing body of research is developing and implementing sophisticated, computer-aided protocols for the identification of building typologies. This paper shares the same goal. An innovative data-driven procedure for the unsupervised identification and description of building types and organization is here presented. After a specific pre-processing procedure, we develop an unsupervised clustering combining a new algorithm of Naive Bayes inference and hierarchical ascendant approaches relying on six morphometric features of buildings. This protocol allows us to identify groups of buildings sharing specific similar morphological characteristics and their overall structure at different aggregation levels. The proposed methodology is implemented and evaluated on the overall ordinary (e.g. not-specialized) building stock of France.

RÉSUMÉ. L'identification et la description des typologies de bâtiments jouent un rôle fondamental dans la compréhension de la forme de l'espace bâti. Un nombre croissant de travaux développe et implémente de nouveaux protocoles sophistiqués de géomatique pour l'identification des typologies de bâtiments et leur organisation. Cet article présente une procédure innovante, basée sur l'analyse quantitative des données, avec comme objectif l'identification et la description non *supervisée des types de bâtiments. Après une procédure de pré-traitement spécifiquement adaptée à notre donnée source, nous développons un protocole de* clustering *non supervisé combinant un algorithme novateur d'inférence bayésienne Naïve avec des approches ascendantes hiérarchiques ; le tout, reposant sur six caractéristiques morphométriques intrinsèques de chaque bâtiment. Ce protocole permet d'identifier des groupes de bâtiments partageant des caractéristiques morphologiques similaires spécifiques ainsi que leur structure globale à différents niveaux d'agrégation. La méthodologie proposée est implémentée et évaluée pour l'ensemble du parc immobilier ordinaire (non spécialisé) de France.*

KEYWORDS: building, typology, bayesian naive clustering, INBIAC, HCA, France.

MOTS-CLÉS : bâtiments, typologie, clustering bayésien naïf, INBIAC, CAH, France.

DOI:10.3166/RIG. DOI:10.3166/RIG.31.265-302© 2022

1. Introduction: the interest of building typologies on large geographical extents

An interdisciplinary and growing body of research investigates the relationship occurring between the form and the functioning of cities (Carmona, 2019). In the typo-morphological tradition, the properties of the urban form are defined by specific spatial combinations of its constitutive elements: streets, plots and buildings (Moudon, 1997). While streets and plots show a higher inertia over time, buildings change at a faster pace depending on the specific historical, socioeconomic and technological context. Moreover, while some buildings (and building types) are easily torn down and replaced, others can endure for several decades or centuries showing higher resilience to urban transformation.

The availability of national databases describing building stocks also opened the way to quantitative analysis of buildings and building types. Different aspects are presently being studied in the context of the environmental transition. First of all, energy-related characteristics of buildings (energy consumption for heating, cooling, construction and recycling). Secondly, the densification potential of building types, as well as their assemblage in specific urban forms, for policies of compact city development. Thirdly, the connection with urban functions (potential of buildings to be adapted to different usages), including issues of adaptation to different housing needs (creation of smaller/bigger dwellings within given building types). Of course, questions of social preferences (and even acceptability) and of differential economic valorization/devalorization between building types are also important in economic and sociological research. National policies are thus creating a demand for scientific research on nation-wide building types, with different disciplinary focus.

An unavoidable passage in the analysis of the enormous diversity of observable buildings in any given geographic context, is the creation of general categories, i.e.

building types. As defined by Case-Scheer (2017, p. 171) "*a building type is an abstraction, a pattern, where we observe formal similarities between one building and another even though the buildings may have different architectural expressions.* […] *buildings share many common formal characteristics, but are very different in color, materials, style, and expressiveness*." Building types can result from a given time period, in a specific regional context with certain easily recognizable stylistic patterns (e.g. the English terraced house; the American bungalow house; the Parisian Haussmann apartment-buildings). They could also be more overarching, grouping together similar buildings produced in different time periods, but sharing consistent common features. Different information at different description levels can be used for the identification and description of building types. The combination of specific features can thus result in a comprehensive building typology, *i.e.* an organization of building types having a given logic and inner coherence. Three levels of details can be distinguished: i) aesthetical and stylistic features (such as façades materials and composition, colors, etc.), ii) the internal organization of the building (including some structural considerations) iii) the overall external hull of a building (shape, footprint area, height etc.). Nonetheless, there are rarely any formal definitions of building types, categories and structures describing the whole building stock of a large urban region (Orford and Readcliffe, 2007). Similarly, there is no agreement about the definition of which combination of formal characteristics is required for the identification of building types and their differentiation, since different sets of features underlie the definition of each building type. When the goal of the study is the identification of building types over large urban regions, databases encompass few features, often limited to the simple geometrical form. The study of the external building hulls allows typifying what might be called the *skeletal form* of the building.

This paper proposes a quantitative computer-based development of a building typology for France based precisely on building envelopes and thus consistent with the newly available national database BD TOPO by IGN (2020). By focusing on forms, we see building envelope geometry as a key factor of the material culture of a given society (as already pointed out by the Italian school of typo-morphology, see Caniggia and Maffei, 1979). The resulting typology will thus be relevant from an urban planning and design point of view, when analyzing a local context in order to propose planning options. It will also concern national policies, when conceiving prescriptions for the building stock. In the latter case, knowing the magnitude of the different building types throughout the national space will be of paramount importance to assess the regional impact of policies.

The novelty of the present work is twofold. Firstly, in order to address the challenge of inductively producing a building typology from millions of records, a new Bayesian clustering protocol has been specifically developed and implemented: INBIAC (Iterative Naive Bayesian Inference Agglomerative Clustering). Secondly, for the first time a form-based typology has been produced for the whole building stock of France. The paper is organized as follows. Section 2 presents related works highlighting the specificities of this paper. Section 3 describes the proposed clustering protocol. In Section 4, the protocol is implemented to the French building stock, and its results are first presented numerically. A specific sub-section (4.3) proposes a general building typology in France by interpreting and projecting in geographic space the clustering results. Section 5 closes the paper with a discussion on limitations of the protocol, improvements and future perspectives.

2. Related works

Traditionally, expert-based typologies of buildings have been proposed by urban geographers, morphologists and architects with a specific focus on stylistic and morphological aspects (Deffontaines 1972; Caniggia and Maffei 1979; Bonillo *et al.* 1988, etc.). These *knowledge-based approaches* are based on the study of small but exhaustive datasets including building features at each of the three detail levels mentioned in the introduction. They are highly supervised, expert-driven and labor-intensive qualitative methods, where experts identify relevant building types and features in a given study area. These works are limited in the capacity to systematically capture the heterogeneity of large datasets, with a consequent limitation in the scalability and reproducibility of the protocol (Fleishman *et al.,* 2022). *Computer-aided approaches* have thus been developed more recently to take advantage of available databases, algorithms and computing power, and produce more general typologies. The spatial extent of the proposed typologies is also much wider. Since the 2000s, an important research effort has been put on the analysis of the German building stock (Meinel *et al.,* 2009; Hecht *et al.,* 2015; Hartmann *et al.,* 2016, etc.). These works focus particularly on the morphometric descriptors of the built-up space. Nonetheless, while studying the building typologies, these works often integrate context-related variables within their quantitative analysis (such as local built-up density): resulting in a conceptual shift from the study of the building typology to the study of the urban fabric, which are two connected but distinct aspects of urban form in the tradition of urban morphology. In the UK, Brown and Steadman (1991a, 1991b) had already analyzed the forms of British housing using dimensional, functional and topological characteristics. Steadman *et al.* (2000) further extended the research by extracting “primary forms” through a decomposition process of the city blocks, followed by a classification performed on the primary forms. Steadman (2014) proposes a summary of these different works, combining history of building types and building form classification. More recently, a large number of energy-related studies investigated the relationship between energy consumption (or other energy-related variables) and building morphological properties (Garbasevschi *et al*., 2021; Evans *et al*. 2019). Within this group of studies, building typologies are identified by expert-based classification (LSE Cities, 2014) or through clustering approaches (Maiullari *et al*., 2021). Nonetheless, building types identified by these works result in energy-oriented typologies rather than building types as defined in urban typo-morphology and architecture.

In France, works proposing building typologies at the national level are thus mostly interested in energy issues. Bonhomme (2013) produces a 7-class building typology for Paris and Toulouse, generalized for the whole of France. APUR (2007) classifies buildings by construction periods characterized by different normative and construction techniques along urban history (9 building types). The European project TABULA, where different EU countries produced specific expert-based building classifications at the national level (in France, Rochard *et al*., 2015), uses two parameters: residential type (single-family, terraced-house, multi-family, apartment block) and construction period (10 periods). The combination of these two parameters gives a matrix of 40 building types. Another example is the PACTE project (2015), proposing a building typology based on three parameters: single-family/collective, construction period and localization (rural/urban). Their combination results in a 26-group typology. More recently Haffner (2022) proposed an 8-class building typology to be applied to the whole of France. All these works are expert-based, classification-based, defined by the combination of two/three energy and construction indicators. Even at the local scale, several studies have focused on characterizing the energy behavior of a housing stock (APUR, 2007; A'urba, 2011; IAU, 2010; Rochard *et al*., 2015). While these works are all energy-related typologies, Perez *et al*. (2020) propose a typology based on the skeletal form of buildings, still limited to the metropolitan region of Marseille in a comparative analysis with Osaka (Japan).

This rapid review of related works must be completed by some considerations on the methodologies used to produce building typologies. We already noticed that traditional *knowledge-based approaches* have more recently been superseded by computer-aided protocols. The first of these protocols had been developed since the 1960s, especially by the Centre for Land Use and Built Form Studies (LUBFS) at Cambridge University (Steadman, 2016). Thanks to more recent algorithmic data analysis developments, together with the increasing computing power availability, more sophisticated protocols have been developed. *Data-based approaches* have thus been proposed especially from digital cartography, with the goal of cartographic generalization or for the identification of urban structures, building detection and building pattern recognition. While the study of building typologies with expert-based protocols and based on highly detailed and historical datasets find its origin in the urban typo-morphology tradition since the 1950s, the identification of building types from footprint-based data is a more recent field of study (Hecht *et al.*, 2015). Two subgroups might be further specified in data-based approaches. *Supervised protocols* (similarly to knowledge-based analysis) require prior knowledge of the target groups we want to identify within the dataset: features are attributed to each group based on similarity rules. Different algorithmic protocols can be chosen. Examples of classificatory approaches can be found in Orford and Radcliffe (2007), Hecht *et al.* (2015), Hartmann *et al.* (2016), Meinel *et al.* (2009), Wurm *et al.* (2009), Smith and Crooks (2010), Sester (2000), Steiniger *et al.* (2008), Römer and Plumer (2010), Henn *et al.* (2012). Supervised protocols can be assessed by their accuracy level to predict the correct predefined classification: Hecht *et al.* (2015) propose a comparative study of 16 machine learning classifiers. In this work the authors assess the superiority of random forest algorithms in terms of generalization and computational efficiency/scalability. *Unsupervised approaches* encompass clustering protocols where the identification of groups is based on algorithms looking for internal similarity among features and without prior knowledge of the target groups or user intervention. Clustering protocols automatically determine natural partitions (clusters) arising from the specific data structure of the inputs without imposing a predefined identification of the classes. Examples can be found in the works of Neidhart and Sester (2004), Werder *et al*. (2010), Schirmer and Axhausen (2016), Perez *et al.* (2020), etc. In fact, expert-based knowledge is never completely absent. In supervised protocols, it is required at the beginning of the analysis to define the target groups, their numerosity and their overall organization. In clustering approaches, it is needed for the interpretation of outcomes. The former allows the analyst to better identify specific predefined building typologies, while the latter allows a more exploratory analysis where natural groups emerge from the data structure and are later interpreted and related to the specific characteristics of the study region. When focusing on clustering approaches, group identification is also influenced by the underlying algorithmic rules. As discussed in Fusco and Perez (2019) most of the traditional approaches (such as K-means) impose the sphericity of clusters (*i.e.* intra-cluster homogeneity) on all the descriptive variables, which could not always be coherent with the complexity of the context under study. Bayesian Clustering (BC) allows us to overcome these limitations. Still, as for most of the clustering approaches, BC, even when using Naive clustering models, imposes other kinds of constraints and can be particularly time-consuming when exploring possible solutions in parameter space. Another relevant aspect that should be considered when implementing computer-aided protocols is the number and the nature of the descriptors used for building classification/clustering. As mentioned in the introduction, several descriptors and proxy variables might be used to describe the geometrical, topological and semantic aspects of the three levels of detail in the description of a building. Some of the aforementioned works also include variables such as geographical location (address/coordinates), land use, urban block geometry, distance between buildings, describing the morphological and functional context of the building (the urban fabric) rather than the building itself. A high number of variables can benefit the algorithm accuracy (for instance, Hecht *et al.*, 2015 use between 72 and 87 features); yet the strong correlation among redundant features can arise issues of biases in data modeling and outcome interpretation. In this case, additional protocols of dimensionality reduction are required (such as PCA in Hecht *et al*., 2015; Maiullari *et al.*, 2021, etc.), each one coming with a further cost of lower interpretability of the intermediary variables, and strong assumptions about the underlying data structure (for instance the absence of outliers and the linear relationship among variables required by the PCA protocol). Finally, the choice of the variables underlying building clustering/classification analysis, depends on the thematic and methodological goal of the work.

In our research agenda, the study of the building types is the first step of a wider research project aiming at the detailed analysis of the urban fabric of French cities. In this work we focus on the study of the morphological typology of ordinary buildings of all of France, through a limited set of variables describing their convex hull shape.

3. Method

3.1. Data preprocessing

The protocol presented in this paper is implemented on the building stock of Metropolitan France (Section 3.1, not including the outermost regions outside of Europe). As for many western countries, France provides several authoritative datasets about the national building stock through its National Geographic Institute (IGN): the BD TOPO® is an exhaustive dataset of metrical precision providing the information about building footprint and height, retrieved from satellite or aerial imagery. Since April 2019, a new version of this dataset has been released (BD TOPO®, V3.0): its main novelty consists in the combination of the original data with information from the national computerized cadastral plan (MAJIC)[1]. Several new features are made available such as the number of dwellings, the age of construction and the number of floors. Nonetheless, this dataset has some limitations, among which three are of concern for our research, requiring specific pre-processing protocols.

The first issue is related to the building footprint definition. Part of the building dataset, enriched by the cadastral plan information, have a detailed definition of their footprint: while one (or more) polygon(s) corresponds to the main built-up body(ies), a number of extensions are separately modeled as adjoining polygons. These extensions are identified by a specific feature attribute, namely *light structures*, defined as structures "*not attached to the ground by a foundation, or a building or part of a*

1 Two points should be highlighted: firstly, not all the IGN building polygons have been coupled with the cadastral data MAJIC for several reasons, mainly missing data, size and nature of the building (further explained in IGN 2020, p. 67). Secondly, in four departments, Marne, Meuse, Ardennes and Yonne the combination with the MAJIC was not complete in 2019. In our protocol these four departments have been temporarily removed and the final clustering has been successively projected on the fully-combined 2021 database.

building open on at least one side" (IGN 2020) such as terraces, loggias, porches, etc. The same attribute is also associated with independent structures such as greenhouses, garages, small and large industrial sheds. Another share of buildings is defined from satellite-/aerial-based methods: in this case, the delineation of the footprint is defined with imagery detection algorithms where the footprint corresponds to the external demarcation of the overall built-up structure, therefore including both the main building body and all extensions. Thus, a harmonization protocol has been specifically developed to re-aggregate buildings made up of several constituent parts (Figure 1).

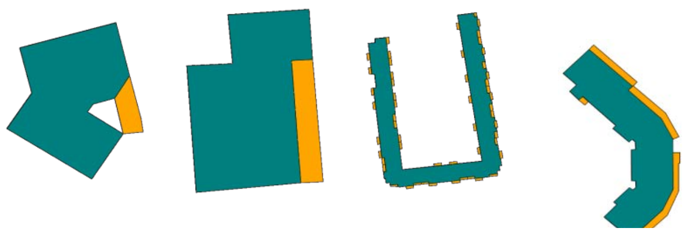


Figure 1. Harmonization of the building volumes: examples of building footprints with their relative light structures in orange (a, b: porches; c, d: balconies), reaggregated into individual polygons

The second issue is related to the definition of the functional specialization attribute. The feature "Building Nature" provides the distinction between several specialized structures (e.g. religious, industrial, agricultural). When the "*overall architecture or aspect of a building does not reveal with exactitude its function*" (*ibid*.) the building is classified as undifferentiated. This corresponds well to our goal of producing a typology of ordinary buildings, *i.e.* non specialized buildings including dwellings, whose serial production in the course of time allows an easier determination of types. Unfortunately, after a manual assessment of this field on a subspace encompassing about 15 thousand buildings, we observed that even if the specialization attribute is always correctly assigned, the overall accuracy of functional specialized buildings is limited to 55.7% because of a large number of false negatives. Thus, this feature is enriched through another IGN BD TOPO layer, namely "Activity Zones" (*zones d'activités*) where specialized buildings are retrieved and collected from other national authoritative sources (polygons and POI layers). A set of specific rules and filters are defined and implemented to associate the specialized function to the original building dataset: on the same 15-thousand features subset, the resulting enriched definition of the field shows 86.9% accuracy in identifying specialized buildings (and 93.3% when building surface is considered). This allows us to filter out these buildings and implement our clustering protocols only on ordinary residential or mixed function buildings. Indeed, as discussed in the typo-morphological literature, specialized buildings (industrial, commercial, religious, administrative, etc.) have often specific and extreme morphological properties (Maffei and Maffei, 2018) which would introduce important outliers in our dataset, biasing the final outcomes.

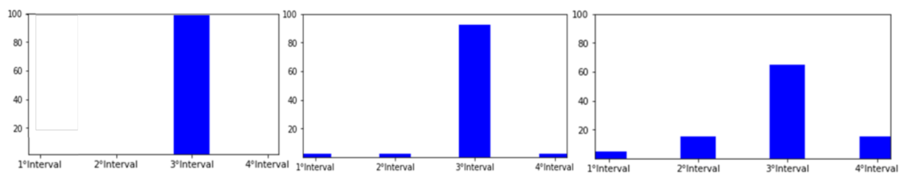
Table 1. Building morphological descriptors and their discretization

Indicator	**Discretization**					
Build. Footprint Surface (S) [m^2]	(0 : 75)	[75 : 150)	[150 : 300)	[300 : 600)	[600 : 1200)	[1200 : ∞]
Build. Topological Contiguity (TC) [n°]	1	2-3	4-8	9+		
Build. Convexity (C) [-]	[0 : 0.8)	[0.8 : 0.9)	[0.9 : 0.96)	[0.96 : 0.99)	[0.99 : 1]	
Build. Elongation (E) [-]	[1 : 1.15)	[1.15 : 1.2)	[1.2 : 1.3)	[1.3 : 1.5)	[1.5 : ∞]	
N° of Dwellings (NbD) [-]	1	2-8	9-24	25+		
N° of Floors (NbF) [-]	1	2	3	4-5	6-7	8+

Once the subset of non-specialized buildings is redefined, morphometric descriptors can be implemented. Three indicators are directly computed from building footprint: Surface (S); Elongation (E), defined as the ratio between the building perimeter and the perimeter of a circle of equivalent surface; Convexity (C), defined as the ratio between the building footprint surface and the area of the minimal convex hull. One indicator, Topological Contiguity (TC), is defined as the number of neighbors within a continuous built-up unit of adjoining buildings. Finally, the building number of floors (NbF) and the number of dwellings in the buildings (NbD) are provided by the original dataset. These six indicators represent a set of minimal descriptors of the building form obtained by a simplified 3D dataset (LOD0+ of the CityGML data model in Biljecki *et al.*, 2016). Nonetheless, the number of floors and the number of dwellings in the buildings (as for several new features of the latest BD Topo V0.3) can be unknown. As we will see in the next section, the clustering protocol developed in this work is specifically conceived to deal with partial missing information. As previously introduced, BC algorithms are based on a probabilistic framework. The fastest implementation of BC algorithms requires qualitative or categorical data. A discretization of the five morphometric descriptors is therefore necessary: for our case study the discretization was obtained through a mix of univariate data analysis (enough numerosity in every bin, possible existence of natural breaks) and expert knowledge (considering domain-relevant classes as single-floor buildings and single-family houses), as in Table 1. This last pre-processing step produces a first reduction of the overall complexity of the original data: in our specific case study, for example, it allows us to pass from about 28.8 million buildings to 2.2 thousand building descriptor tuples, each one corresponding to a specific combination of our bins of feature values.

3.2. Iterative Naive Bayesian Inference Agglomerative Clustering

Bayesian inference is a powerful probabilistic option for quantitative and qualitative multivariate data clustering using simple model architectures as the Naive Bayesian Classifier, where the cluster variable is conceived as the common parent of all the other variables, and conditional independence among them is assumed knowing the cluster variable (Duda and Hart, 1973). The expectation-maximization (EM) algorithm is normally used to identify an optimal clustering solution in terms of log-likelihood, for a given number of clusters (Dempster *et al*., 1977). Exploring solutions with varying number of clusters can be done with a random walk in solution space, using a clustering score combining log-likelihood and a penalization for model complexity, e.g. number of clusters (like in Fusco and Perez 2019). McCaffrey (2013) offers an interesting alternative to the EM algorithm for BC: Iterative Naive Bayesian Inference Agglomerative Clustering (INBIAC). So far, the implementation of the INBIAC algorithm can be only found in Carneiro *et al*. (2015) for credit card fraud detection. INBIAC is a much faster algorithm than EM as it replaces recursive batch inference of cluster assignments for all records to an iterative assignment of individual records which are randomly extracted from the database and assigned to the highest likelihood cluster at that given moment of the clustering procedure. The higher speed of the INBIAC algorithm can be used to perform a higher number of clustering solutions. Just like EM, INBIAC results are sensitive to the clustering initialization. In EM, initialization implies randomly assigning all records to clusters. In INBIAC, a k-cluster solution needs the use of k records as initial seeds of the clusters. McCaffrey (2013) proposes a preliminary phase of seed initialization, randomly choosing k seeds and finally keeping the set of seeds with maximum Hamming distance. In our algorithm we improved McCaffrey's protocol in several respects. Firstly, in order to better represent the ordinal structure of our data, we used Manhattan distance instead of Hamming distance, after normalizing for the cardinality of the ordinal values. Secondly, the usual Laplacian smoothing in the initialization of conditional probability tables, is reinforced by a further smoothing on the ordinal values which are contiguous to those of the seeds (Figure 2). Laplacian smoothing is a classical procedure of Bayesian reasoning and allows for non-null likelihoods when calculating posterior probabilities of cluster assignment of new records. Distributing some of the probability mass from the exact value of the selected seed to its neighboring values allows for taking into account the ordinal nature of our variables (which were obtained from the discretization of continuous variables). By doing so, the algorithm favors the assignment of new records to seeds which have close enough, even if not exactly the same, values on some of the clustering variables.



Seed with all probability mass on the exact value

Seed with homogeneous Laplacian smoothing

Seed with smoothing reinforced on contiguous ordinal values (20% of the original probability mass on each of the two neighboring values, 30% in one neighbor only)

Figure 2. Probability smoothing in the initialization of the conditional probability table

After seed selection, all remaining records are assigned one by one to the cluster having maximum posterior probability using constant and equal priors for every cluster (initialization phase). The cluster conditional probability tables are updated after each record assignment, but not cluster priors. The iteration phase can then begin. Individual records are randomly selected and removed from their current cluster to be re-evaluated and reassigned to a better cluster (the one having maximum posterior probability, and hence log-likelihood). This operation changes cluster composition incrementally (which is the main difference with the Expectation-Maximization algorithm). During iterations, cluster priors are updated in the process and used in the Bayesian calculus. Given the Naive model architecture and the resulting additive formula of model log-likelihood, the local optimization of log-likelihood in table record assignment to clusters produces a global log-likelihood improvement for the whole model.

A further and last improvement has been the treatment of missing values, which was not foreseen in McCaffrey's original INBIAC algorithm. Under the Missing at Random (MAR) assumption, likelihoods and posterior probabilities of cluster assignments are calculated only on the observed values, but missing value imputation is later performed based on the most probable values within the assigned cluster. Imputed values are iteratively erased and re-imputed within the INBIAC procedure, and the final log-likelihood of the clustering solution includes the contribution of imputed values. The clustering iterations within INBIAC stop when no record can be reassigned to a different cluster. Finally, buildings are weighted by their footprint surface in the clustering algorithm, giving the same importance to each square meter of built-up surface (the clustering solution would otherwise be biased by the overrepresentation of small buildings). The INBIAC clustering process is described in Figure 3. For each number of cluster k, several competing models can be produced with a different initialization of the

INBIAC algorithm. The optimal clustering solution can be then selected as the one scoring the minimal log-likelihood loss score. The implementation of the protocol for several numbers of clusters k, ranging in a user-defined interval, allows to explore different clustering solutions.

3.3. Hierarchical clustering analysis

The implementation of the INBIAC protocol, allows us to identify one (or more) optimal solution(s) based on the optimization of the log-likelihood of the corresponding clustering model. Nonetheless, the selected clustering(s) solution(s) would always provide a specific partition of the original dataset defined for a given number of groups k. It is thus interesting to study the variation of building clustering across k. Buildings constantly grouped in the same clusters could reveal stronger structural patterns within the data. We thus implement an agglomerative Hierarchical Clustering Analysis (HCA) using as input the INBIAC best outcomes for each k clustering solution in the interval explored. The rationale underlying this methodological choice is the following: the subset of n best clustering solutions can be used to partition our original dataset (or similarly, the 2.2 thousand tuples) in smaller subgroups of elements (kernels) always being clustered together independently of the number of clusters k. These kernels represent the finest partition for which the highest inter-level consensus is observed: no cluster at any level further divides the element in finer groups. Within our specific context, these kernels correspond to a highly detailed meta-cluster solution of specific building sub-types; few kernels gather most of the buildings (more precisely of the built-up surface, given our weighting scheme), and vice versa a large number of kernels encompass less built-up surface with less recurrent shapes. HCA will be implemented with Gower's dissimilarity metric (Gower, 1971) for cluster distance and centroid-linkage agglomerative principle among clusters. Implementing an HCA allows us to produce hierarchically nested groupings based on the similarities within this elementary kernel partition. Kernels are arranged in a hierarchical manner. Thus, the combination of the INBIAC and HCA protocol combines the advantages of the two protocols. On the one side, we keep the ability of probabilistic Bayesian inference to select non-spherical clusters defined with a maximum log-likelihood approach on subgroups of features. On the other, an overall hierarchical structure allows the analyst to observe the overall data clustering organization similarly to knowledge- and ontological-based classifications. Moreover, the main advantage compared to regular HCA applied on raw data, is that the outcome variability produced by the high sensitivity to the initial clustering settings is strongly reduced. This approach shares the same underlying hypothesis of consensus clustering protocols (Monti *et al.,* 2003), where several cluster solutions are combined in order to achieve a more robust solution.

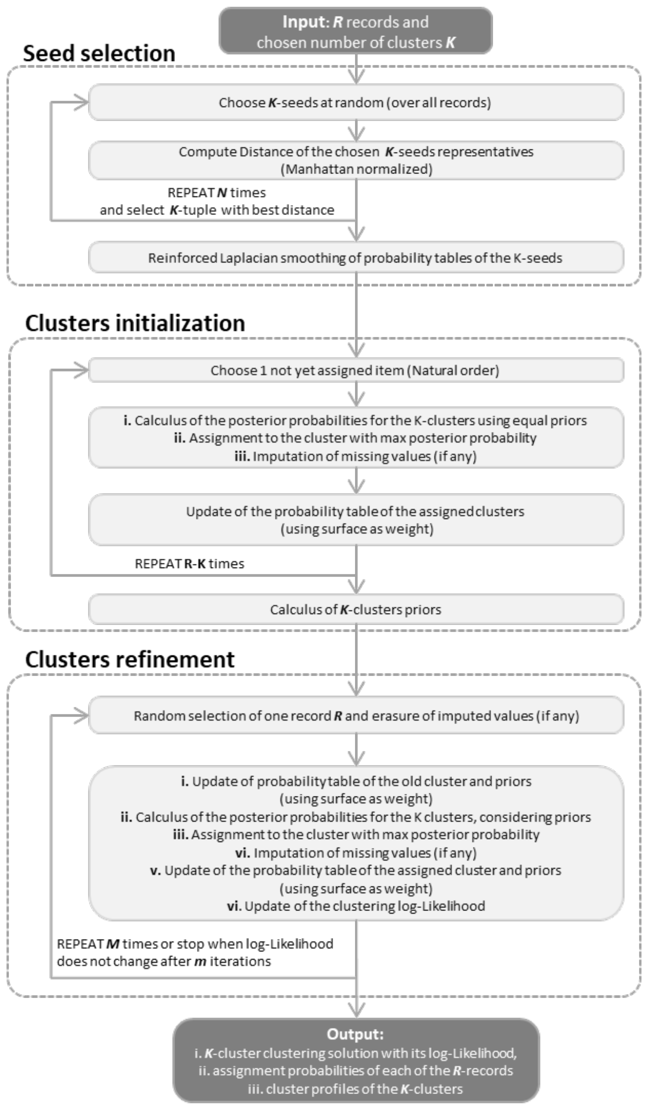


Figure 3. Flowchart of the INBIAC algorithm

Nonetheless while consensus clustering looks for similarities within a larger number of clustering solutions for a given number of clusters k, in our case we use a more “controlled” subset of n optimal clustering solutions at different levels k. Moreover, instead of implementing the same clustering protocol at two different stages of the analysis, we combine a non-hierarchical and a hierarchical protocol in the first and second stage of the analysis, respectively. Finally, our protocol provides a profile for each cluster as probability distributions of its members over the values of each clustering variable. These profiles are later used for cluster interpretation (section 3.2).

4. Application

4.1. Study area

The protocol presented in this work is implemented on the building stock of Metropolitan France (not including the extra-European regions). The IGN layer modeling the overall building stock of France (as of January 2020), includes about 47,27 million polygons (39,33 million after implementing the harmonization protocol, explained in Section 2). Within this paper, the 96 departments of Metropolitan France will be used as convenient spatial units to observe distribution of building types within the country[2]. Established during the French Revolution and seldom modified afterwards, they are a major administrative and functional fact in France. Departments will be further grouped within wider regions whenever appropriate in our text, using either the 13 present-day administrative regions or the more detailed 22 regions which were used before the 2015 reform. Beyond the well-known urban/rural built-up density differences, wider geographical regionalization can also be observed when mapping building morphological features at the department level (Figure 4). The building coverage ratio of ordinary buildings (Figure 4a) shows higher values for both departments containing large metropolitan areas (Paris, Marseille, Lyon, Lille, Nice, Toulouse, Nantes, Bordeaux, Strasbourg) and those located along the Mediterranean and the Atlantic coasts; on the contrary low values are found along the northeast-southwest diagonal (commonly known as the empty diagonal, *diagonale du vide*). Its spatial pattern shows strong similarities to the demographic and urbanization trends of the last 50 years (Oliveau and Doignon, 2016). Specialized building coverage ratio (Figure 4b) shows a relatively similar pattern. However, its values are also higher over the northwestern regions, explained by the large presence of farming facilities mainly in the Normandie, Bretagne, and Pays-de-la-Loire regions.

2. With the exception of Paris Metropolitan region (Ile-de-France) where 8 departments have been merged.

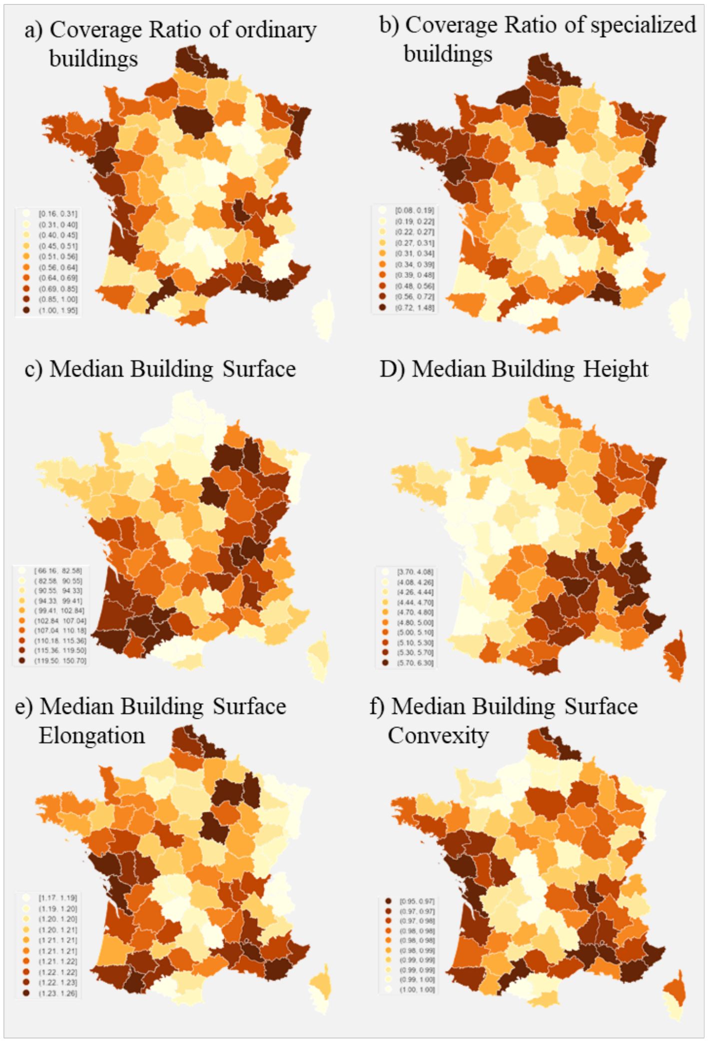


Figure 4. Explorative data analysis of the main building stock features at the departemental level. Decile analysis of six indicators: a) Coverage Ratio (ratio between the built-up and overall surface). b) Specialized Building Ratio (ratio between specialized and overall built-up surface). c) Median building footprint surface. d) Median building height. e)Median building footprint elongation. e)Median building footprint convexity

Locally high values are also observed in departments with large industrial and logistic areas such as Aude, Haute-Marne, Loire, Bouche-du-Rhône, Haut-Rhin. The median surface areas of ordinary buildings (Figure 4c) is higher in the northeastern and at the southeastern regions. The median building height (Figure 4d), with the exception of the Ile-de-France Region, shows a northwest-southeast trend with higher values in mountainous regions (Massif Central and Alpes). Median values of building footprint Elongation and Convexity (Figure 4e-4f) describe the presence of a main trend of increasing building complexity in departments with large metropolitan areas (Marseille, Lille, Bordeaux, Toulouse, Nice and the 8 departments of the Paris region). High values are also observed in some departments of the Mediterranean and Atlantic regions without metropolitan areas.

4.2. Results – Overview of clustering outcome

From the original 47,27 million polygons, the data pre-processing protocol allows us to identify about 39,33 million buildings: 26,3% are specialized and the remaining 73,3% have a residential or mixed use. Our clustering protocol will be applied here only to this larger stock of ordinary buildings. The segmentation of the six features (S, C, E, TC, NbF, NbD) further reduces the variability of our dataset resulting to 2,200 observed combinations, over the 14,400 possible ones of the discretized variables. Mutual information correlation is then used to measure how much of the information is shared between any two variables. This allowed us to verify that at the building level, just like at the department level, the two indicators of geometrical shape of the building footprint, namely convexity and elongation, are highly negatively correlated. They are two slightly different ways of accounting for building footprint compacity. We decided to keep both measures in the clustering analysis but to weight each of them with half of the weight of the other variables, in order not to bias the clustering results towards a single dimension of the analysis.

Then, the INBIAC protocol was implemented 100 times for each number of clusters *k* between 6 and 19, producing thus a total of 1.4 thousand models (Figure 5a). For each model *N*=1,000 sets of random seeds were drawn and the one maximizing the Manhattan distance was kept. After complete initialization of the *R*=2,200 records, a maximum of *M*=500,000 iterations of random selection and reassignment of records were carried out. Alternatively, the algorithm stops after *m*=50,000 iterations without log-likelihood improvement. The rapidity of the INBIAC algorithms allowed us to perform the calculus in acceptable time. We used 20 virtual machines working in parallel with a standard Core i3 processor at 2.0 Ghz and 10 Gb of available RAM for each calculus thread. This allowed us to perform the whole calculus in 24 hours. More precisely, computing times for each model ranged from a minimum of 13 minutes to a maximum of 6 hours, with median value being 2 hours.

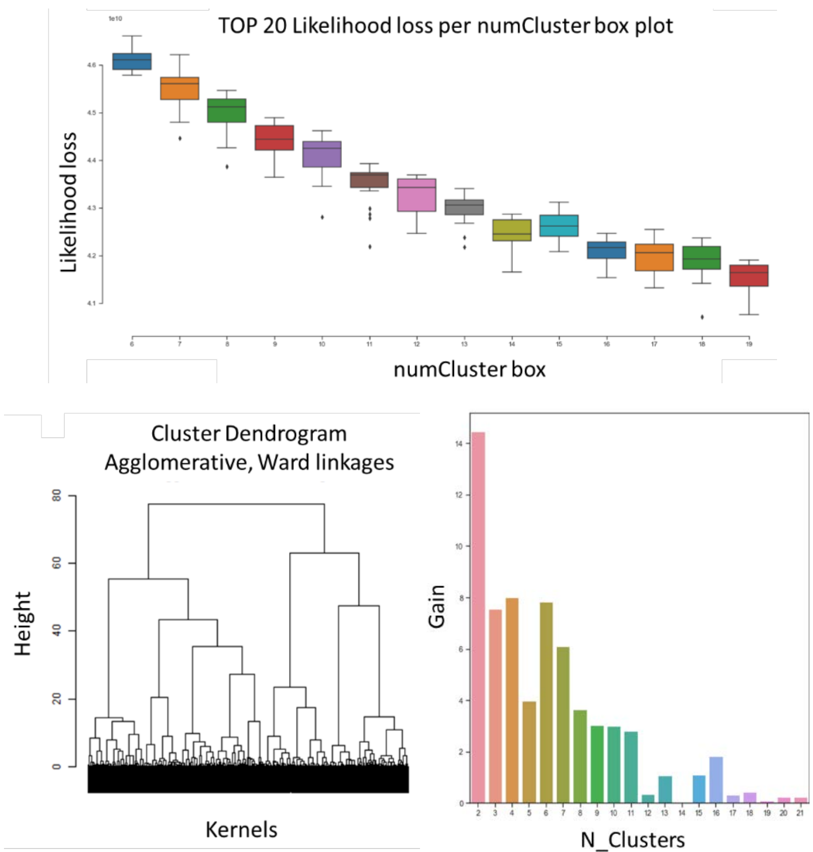


Figure 5. Outcomes of the building classification: a) Box-plots of the top 20 INBIAC models scoring the highest log likelihood-loss scores for each number of cluster k in the interval [6:19]; b) HCA dendrogram implemented on the outcomes of the 14 best scoring INBIAC models; c) identification of the partitions corresponding to the highest dendrogram depths

For each model, we can calculate the mutual information between the cluster variable and the original variable. This allows us to highlight the relative contribution of the morphometric variables to the clustering results. In 63.3% of models, the most important variable was the footprint surface, in 36.6% it was the building height calculated as the number of floors.

For each pre-selected number of clusters, the quality of the 100 models produced by the INBIAC algorithm can be accessed through the model log-likelihood. Figure 5 shows the boxplots of the log-likelihood of the 20 best models for each number of clusters (as usual in Bayesian modeling we actually plot -log-likelihood, which has thus to be minimized). As expected, the plotted values tend to decrease with models having an increasing number of clusters: more parameter-rich models tend to give better representations of observed values. Instead of selecting an optimum model using the elbow method (Lopez-Rubio *et al.*, 2018) or a score combining log-likelihood and a penalization for the number of parameters used (as in Akaike information criterion), we selected the single best model for each of the pre-selected number of clusters.

According to our approach, we will thus use the information derived from all the 14 best clustering solutions. Their combination further defines 1121 kernels, 7 of which encompassing, at least 2% of the overall built-up surface, for a total of 50.06%. The implementation of the agglomerative HCA on these kernels, allows us to identify and describe through a nested taxonomy the overall organization of the building types in France (Figure 5b). The dendrogram in Figure 5b shows the succession of cluster agglomerations along a distance axis, starting from the 1121 kernels (below) and arriving to the complete amalgamation of clusters (top). The length of the segment on the distance axis during which a given *k*-cluster solution is present is indicative of its importance in structuring the building typology in the study area. The first three solutions showing the longest segments on the distance axis of the dendrogram correspond to 2, 4 and 6 clusters. Moreover, when considering local peaks, 11, 13 and 16 clusters represent significant solutions (Figure 5c). Starting from the sixteen-cluster solution (the highest in number), these five solutions are described as follows.

4.3. A general typology of ordinary buildings in France

The combination of the six selected morphological descriptors of the building hull, produced consistent groups of building types. In this section the 16 building types are visualized in Figure 6 and described combining five different pieces of information: i) the quantitative distribution of the six morphological properties underlying the clustering results (Annexes A1); ii) the HCA dendrogram, allowing to observe similarities/dissimilarities between types and their overall taxonomical organization (Figure 5b); iii) scientific literature on French building typologies, offering some examples of well-known regional and/or historical building types which we can associate to our data-driven clusters; iv) the *ex-post* analysis of the construction year is used to characterize the *historical profile* of the 16 building types and describes their different deployment during the last century[3] (Figure 7). Finally, v) the building type prevalence, at the department level, allow us to identify the geographical distribution of types and the presence of specific regional patterns (Figure 8 and Annexes B1-6).

C1 Compact townhouses and **C2 very small buildings.** C1 and C2 clusters corresponds to low-rise buildings with a very small and compact footprint surface (S<150 m^2). These two building types mainly differ for two features. Firstly, their building height: 1 or 2 floors (probability 28 and 72%, respectively) for the former while 3 or 4-5 floors (probability 77 and 23% respectively) for the latter. Secondly, the number of dwellings: 93% of single-family houses for C1 and one or few dwellings for C2 (NbD 1 and 2-8 with probability 49 and 48%, respectively). In France, C1 and C2 account for 11.7% and 2.4% of the total building footprint surface, respectively.

Building types C1 and C2 have been built in different time periods, ranging from medieval to modern and contemporary ages (Figure 7). The most ancient types of these two classes corresponds to low-rise townhouses and small buildings from the medieval and early modern ages, still preserved in historical centers of large and small cities, faubourgs, villages and hamlets. Examples are the Provençal and more widely the Mediterranean old city centers combining both C1 and C2. Beyond these general types, different well-known specific building types belong to C1 and C2. Within the former group, regional-specific types are, for example, the *Echoppe Bordelaise* and the *Maison Toulousaine*, small ground-floor houses (or with, at most, one floor) originally built as winegrowers and market gardeners houses, respectively. Both types are from the XIX Century until mid-XX Century, and are found in the neighborhoods surrounding the historical center of Bordeaux and Toulouse (Barrère, 1956; Callais, 2018; Rewienska, 1937). Ground-floor and one-story adjoining houses are also the main building typology found in the typical working-class housing subdivisions, known as *Cités ouvrieres*, of the late XIX Century, of the industrialization age, until the interwar period. They are very common in north-eastern France (Haut-de-France, Alsace-Lorraine regions). Depending on their specific spatial arrangement relatively to the street and/or the courtyard, they are known by the specific names of, among others, *Cour*, *Courée*, *Choques*, *Corons*, *Forts*, C*arré Mulhousien* (Prouvost, 1969; Deyot, 1983; Guignet, 2008). Those from the interwar periods differ from the previous ones, by the lower number of contiguous houses and their spatial arrangement, organized as garden cities. While some of these types have been expressly designed as row houses, others have been originally planned as small individual or twin houses (C10) and successively

3. Two elements should be considered in the analysis of the building stock historical profile. Firstly, the construction age provided, cover only 54% of the overall building stock. Secondly, a survivorship bias might influence our observations since the construction age distribution is relatively to the existing building stock (in 2020). Finally, no information is provided about the historical modifications of buildings, accounting for possible type changes.

transformed by an infilling process with garages, extensions or other residential buildings. In the second afterwar, very large and homogeneous working-class subdivisions were built as semi-detached and row houses, common, again, in north-eastern departments. Spread in different cities in France we also find several *Cités de Castors*, working-class self-built housing projects from the 60s. Since the 80s, the model of the adjoining house become less common and employed only in contemporary horizontal housing projects, in different urban contexts, mostly developed by private contractors. Despite being composed by multiple dwellings, those detected from satellite images are often modeled in BD TOPO as a unique construction (thus clustered with C14). A very local-specific building types within the C2 group is, for example, the *"Trois-Fenêtres Marseillais"*, a specific building type which takes its name by the presence of three windows along the main façade and which was widely used in the planned urban extensions of Marseille between the end of the XVII and throughout the XIX Centuries. It is built over a rectangular narrow and deep plot, a common feature with the small buildings of the medieval old cities in Provence (Bonillo *et al.*, 1988). Another regional-specific example from northeast France, is the "*maison bel étages"*, a three-story townhouse. Both in its traditional and modernist forms, this type is characterized by functional spaces at the ground floor (typically garages for modernist version), living space at the first level and bedrooms at the second floor. While the original version was traditionally built by individual owners as adjoining houses in urban areas, the modern one was built in the form of row house subdivision by estate developers in the suburbs. The geographic distribution of the departmental share of C1 (A.1), shows a regional-specific overrepresentation in the northeastern regions (Haut-de-France, Alsace-Lorraine) with a peak of 27% share in the Nord department, followed by the southwestern Mediterranean departments (in Roussillon and Languedoc). The same geographic distribution, but with lower shares, is also observed for C2: highest shares at around 6-7% observed, again, in the northeastern departments. Beyond these specificities, a background general trend can also be observed, opposing departments of the plains and of the mountains, characterized by higher prevalence of C1 and C2, respectively, thus relating the building height to the morphological properties of the landscape.

The next three building types C3, C4 and C5 are three classes of articulated buildings and mid-sized townhouses. Indeed, they are both characterized by mid-sized footprint surface, mainly between 150 and 300m^2, and by high footprint complexity (high elongation and low convexity). Differences between these three types are made by the different number of floors and dwellings: C3 are thus defined as **mid-rise multifamily articulated buildings** characterized by 3-5 floors and accommodating a few dwellings (74% probability of NbD between 2 and 8). C4 corresponds to **large articulated townhouses**, adjoining buildings of two stories, with mainly one or sometimes a few dwellings (73 and 25% probability of NbD 1 and 2-8, respectively). C5 gathers **low**

articulated townhouses of one story, mainly single-family (probability 84%). They gather 3, 2.5 and 2.5% of France’s building footprint surface, respectively.

These three building types show a higher level of architectural, spatial and historical similarity. They are, indeed, combined together in the lower part of the dendrogram due to their stronger morphological similarity compared to the other classes. When observing their spatial distribution they are frequently combined in faubourgs and more historical parts of urban areas that have not known the substitution process of the late modern and contemporary ages. Their complex shape is often explained by their historical transformation and/or their functional-related origins. This is also confirmed by the similarity of their historical distribution, especially for C3 and C4. C3 gathers several more specific building types such as, large articulated townhouses, large mid-rise buildings, from the medieval to the XVIII Century, limited to three floors such as *Hôtels particuliers* (urban mansions) and *Maisons de Maître* (business-owner townhouses), originally conceived for single-family use and more recently transformed in multifamily residential houses. Other examples are *Maisons Ouvrières* (collective working-class townhouses), and more generally residential small buildings for workers, conceived from the origin for several dwellings. An urban-specific example is the wine-trader building, from the XVIII Century, located along the Garonne River, in the Chartrons neighborhood, Bordeaux. They have elongated shape (within a 10m large and 70-140m deep plot), and are composite buildings originally combining the noble part aligned on the street front, with apartments and offices of the business-owner, sometimes a second less ornate residential building behind the first, separated by a small courtyard. On the backside, they also include a working area with warehouses/cellars, also known as *chia* (Callais, 2015). C4 corresponds to small adjoining residential buildings, often born as mixed residential and commercial/handcraft related buildings. Regional specific type from the Mediterranean urban areas are, for example, the *Immeuble à logements à coursière*, or *courées*, in this case multifamily building. Within the one-story complex shape buildings of C5 group we find, for example, the various forms of lengthwise townhouses, known as *longeres*-type buildings in France. This is originally a typical built form of rural north-western regions but several building types refer to this name under the common definition of linear arrangement of farm buildings with one or more adjoining residential properties. The same name might also be associated to other urban-related functions located within denser urban fabrics; on the contrary, isolated lengthwise houses in suburbs are classified as C12.

The geographic distribution of the departmental share of C3, C4 and C5 (A.2), shows different spatial patterns. C3 is characterized by higher presence in the southeastern regions with higher share values (about 3-4%) in the Alsace, Lorraine and Languedoc-Roussillon before-2015 regions. Underrepresentation of C3 is observed in western departments with minimal shares at less than 1%. C4 is much more common, mainly in the Grand-Est, Pays-de-la-Loire and Haut-de-France regions (between 10-18%). Lower values are observed in mountainous departments (in the Massif Central, Massif Armoricain and the Alpes). Finally, C5 is mainly found in the more planar regions, such as the departments along the Loire, Seine and Rhône rivers (6-8% of building surface share). Again, building height seems to be associate to the morphology of the landscape. Specific attention should be given to the departments of the Grand-Est region, where all these three typologies reach their highest values.

C6, C7, C8 are three different forms of multi-family mid-rise buildings. The main common feature of these types is the building height, mostly included between 4 and 7 floors. On the contrary, differences between these three types can be observed in the other features. **C7** are **small adjoining mid-rise buildings,** characterized by a compact small footprint (70-150 m^2 with probability 85%), building height mainly between 6 and 7 floors (84%), and a low number of dwellings (93% between 2 and 24). Compared to C7, **mid-rise mid-sized compact buildings C8** are characterized by more complex and larger surfaces (probabilities 39 and 61% of 150-300 and 300-600 m^2, respectively), accommodating a larger number of dwellings (NbD>8) and with a more variable building heights, even if always more than 4 floors. Historically, both C7 and C8 show a long-standing presence in France: while the former was relatively common until the first half of the XX Century, the latter shows a relative increase of its presence in the second afterwar. The higher similarity is also illustrated by an earlier convergence in the dendrogram, while the merge with C6 takes place relatively later. This last class encompasses **combined or isolated mid-sized compact buildings and small towers** (probability 40% for 4-5 floors and 52% for 6 or more floors). C6 shares with C8 the same characteristics of footprint size and number of dwellings while with C7, footprint complexity. Moreover, C6 is specifically characterized by a lower contiguity (probability 40% of being isolated and 60% of having 2-3 adjoining buildings), and for its historical profile. C6 is the only type with construction age concentrated in the two decades of the 60s and 70s. These three classes account for 0.2, 1.6 and 3.6% of France’s building surface, respectively.

Small adjoining mid-rise buildings C7 are found in old centers with high vertical development stimulated by natural constraints (site morphology) and/or anthropic forces (real estate market). In these ancient fabrics, townhouses and small adjoining buildings left the place to taller buildings, built over the same small underlying plot. Examples might be found in the historical center of Paris (although substituted by C8 mid-sized buildings), Nice, Grasse, Grenoble, Lyon, and its working-class Croix Rousse neighborhood (where taller buildings were also needed to host the weaving machines). C8 Adjoining mid-size and mid-rise in-line apartment buildings represents the most common building type characterizing larger city centers outside their historical core. Their building age span between XIX Century until today with a major production between mid-XIX and mid-XX Centuries. Despite their different aesthetical properties, ranging from belle epoque, art deco, regionalist, eclectic and more modern or contemporary stylistic features they all share the same morphometric properties. Specific examples are the *Immeubles de rapport* (tenement buildings), the Haussmann and the neoclassical style buildings, with strict architectural codes. These building types are prevalent in regularly planned urban extensions of the modern age such as the Parisian Haussmann renovation plan, the Lyon Morand and Perrache plans, and the Nice Consiglio d'Ornato plans, among others. C6 are mainly modernist (50s-70s) detached or semidetached mid-sized apartment buildings and free standing mid-to-high-rise towers. Towers are apartment buildings whose simple compact footprint is relatively small in comparison to their vertical development. Some more historical buildings also belong to C6. Some of them are mid-sized buildings, built within compact urban fabrics over small street blocks, thus with a resulting lower contiguity and higher ratio between surface footprint and building height. Similarly, C6 also encompasses some historical mid-sized buildings that have successively became isolated due to transformation of the street network (e.g. Haussmann transformation in Paris). Local-specific examples are detached palaces and hotels from the first half of the XX Century and buildings (Art Deco and Belle Époque style) built over the central hill of Cimiez in Nice, and today transformed into luxury condominiums.

The geographic distribution of C6, C7 and C8 is similar (A.3), despite the different magnitude of the observed proportions (varying between 0 and 7% for C8 while between 0 and 0.4/0.7% for C6 and C7). A local specific overrepresentation of these three types is observed in departments with larger metropolitan areas (Paris région, Marseille, Lyon and Nice departments having the highest shares), overlapping a northwest to southeast increasing trend. C7 shows higher presence in the southeastern departments (provençal and alpine old towns). C6 also has a overrepresentation located in southeastern coastal and mountain departments probably due to the natural morphology and the large presence of touristic and secondary homes (highest values of C6 are indeed observed in southern Corsica and Provence-Alpes-Côte d'Azur, before Lyon and Paris departments).

C9 groups together **big articulated adjoining mid-rise apartment buildings.** C9 is characterized by very distinctive features: large and very complex footprint (surface over $600m^2$), with few-to-some adjoining buildings (TC>4) and with varying building heights, always over 4 stories. These characteristics make C9 differ significantly from the other 15 types: indeed, it is only at the 4-class level of the dendrogram that C9 is merged with the group of large detached collective building forms (C14, C15, C16). Despite their different values of building contiguity, C9 shares with the modernist building forms a similar distribution of number of floors (NbS), footprint surface (S) and footprint complexity (E, C). This similarity is further confirmed by the similarity of their historical profiles. Most of these building types are found in some parts of historical/classic urban fabrics where modernization/transformation processes have been replacing groups of townhouses and small-/mid-sized buildings (C1-C8) or by infilling process of undeveloped urban blocks. Individual C9 building types are present in the center of metropolitan areas where higher real estate pressure has been transforming the urban landscape (*e.g.* Paris, Lyon and Nice, etc.). Groups of C9 buildings are also found where real estate operations have transformed entire neighborhoods such as the surroundings of the Montparnasse railways station in Paris, Saint George in Toulouse, Meriadeck in Bordeaux (a typical example of *urbanisme sur dalle*) etc. Some older buildings are also classified in C9 group such as originally specialized or residential large buildings with a monumental/ historical character. C9 encompass only 1.3% of the building stock surface; high geographical concentration of C9 is found in departments containing large metropolitan areas, Paris, Nice and Lyon being at the top three with over 3% of the departmental (or regional in the case of Paris) built-up surface.

The three building types C10, C11, C12 correspond to detached single-family houses and villas. **C10** groups together **small compact houses**: it is characterized, indeed, by very compact one-floor (49%) and duplex (49%) houses, with a footprint surface between 75 and 150m^2 (86,5%). **C11** and **C12** characterize **articulated villas**. Both groups combine large villas and some semidetached houses with a footprint surface mainly between 150 and 300m^2 (probability 100% for C11 while 75% for C12), with more complex footprint shapes than C10. What mainly differentiate these two types of villas is the floor number: one-story (99%) for the former (i.e. bungalow houses) and two-stories (95%) or more, for the latter. The strong similarity over the other five building hull descriptors makes these two types of buildings converge in the early part of the dendrogram. These three groups are particularly important: over 65% of the French ordinary building stock is associated with C10, C11 and C12, making the (semi)detached houses and the villas the most common building types in France. They account for 35.5, 14.3 and 15.7% of the overall built-up surface, respectively. A fourth type, accounting for only 0.2% of the building stock and also converging with the houses and villas is **C13**, **articulated composite large villas and small isolated residential blocks**. Despite not representing traditional houses and villas, this type is also characterized by detached (80%) or semidetached (20%), low height buildings (probability 85% of NbF equal to one) with few dwellings (NbD mainly one, sometimes 2-8, with probability 72 and 28% respectively). On the contrary, its distinctive features are the very large surface (over 300 m^2) and very complex footprint shape, with values beyond those observed for the previous three types.

As delineated by the historical profiles of C10, 11, 12, the widespread use of single-family houses and villas finds its origin in the late XIX Century, followed by a first diffusion in the interwar period, and a second wave in the second afterwar, when it became the prevalent residential form to be implemented (*ordinary* urbanism, Herrmann, 2017; Fourcaut, 2000; Wiel, 1999). While in the interwar period compact houses and duplex C10 represent the main built type within urban subdivisions, the second afterwar sees also the production of more articulated types of houses and villas C11 and C12 in suburban developments. The economic crisis of the 1990s profoundly slowed down the building industry, concerning houses and villas more than other building types. In the last two decades, the single-family house production rebounded almost at the same level of the 50s-70s but with a different composition: a higher proportion of complex large houses and villas C11 and C12, and a lower proportion of small compact houses C10. Due to the homogenization of the single-house morphologies of the last 70 years (Herrmann, 2017) specific national or regional types associated to these three classes are identified by their stylistic features, not included in this work, that differentiate between the different neo-regional styles (*e.g*. neo-breton/normand/basque /provençal) developed since the 70s. This stylistic diversification, correlated to an increase of C11/C12 types, is also a market evolution from the previous standardized and low-cost national models (like the *phénix* house, which definitely belongs to the C10 type). Nonetheless, we can still identify interesting geographical patterns when observing the share of the building type surface at the department level. C10 is characterized by a very marked spatial trend, with increasing shares from the south to the north/northwest of France. It varies between a minimum of about 15% in the southwestern departments (Pyrénées-Orientales, Aude, Gers) and it reaches the highest values at about 55% in northwestern France (Bretagne and Normandy regions). High values are also observed in those departments with a prevalent mountain landscape (in the Massif Central and in the Alpes). Similarly, the mountain morphology of the different departments seems also to influence the spatial distribution of C12 type, showing a higher presence below a diagonal going from northeast to southwest, varying between a minimal share of 8% observed in the Ile-de-France region, and 28% in the Mediterranean coastal department of Var. On the contrary, C11 shows a specific pattern that seems to follow the presence of the flat landscapes of the Loire, Garonne and Rhône valleys (exception made for the flat departments of the Haut-de-France region, where the townhouse type C1 prevails).

Finally, C14, C15, C16 are three groups of isolated mid-sized and large collective buildings. The main feature characterizing these three types, is the low contiguity (isolated with probability 81%, otherwise 1-2 adjoining buildings). **C15** corresponds to **short slabs and low-rise towers**, while **C16** to **long and/or articulated apartment blocks**. They are both isolated mid-to-high-rise apartment buildings, mainly between 4 and 6 floors (probability 54 and 55%, respectively) although sometimes higher. Both have a high number of dwellings: 78% of C15 has between 8-24 dwellings while 69% of C16 has more than 24 dwellings. These two types are differentiated mainly by their footprint surface and shape. The former is smaller, measuring between 300 and 600m^2 and with a more compact footprint shape. On the contrary, the latter has a very large (probability 57% between 600 and 1200m^2 while the remaining 42% for more than 1200m^2) and a very elongated/articulated footprint shape. As showed by the dendrogram, these two groups are the most similar among the 16 classes, the first to be merged of the overall taxonomy. **C14** gathers **large articulated low-to-mid-rise apartment blocks**. The large and complex footprint shape of C14 are common features with C16, whereas it is differentiated by C15 and C16 for its limited number of dwellings (mainly between 2 and 8, as for C12) and lower height (between 2 and 3 floors). C14, C15 and C16 account for 0.8, 2.3 and 2.4% of the overall building stock surface.

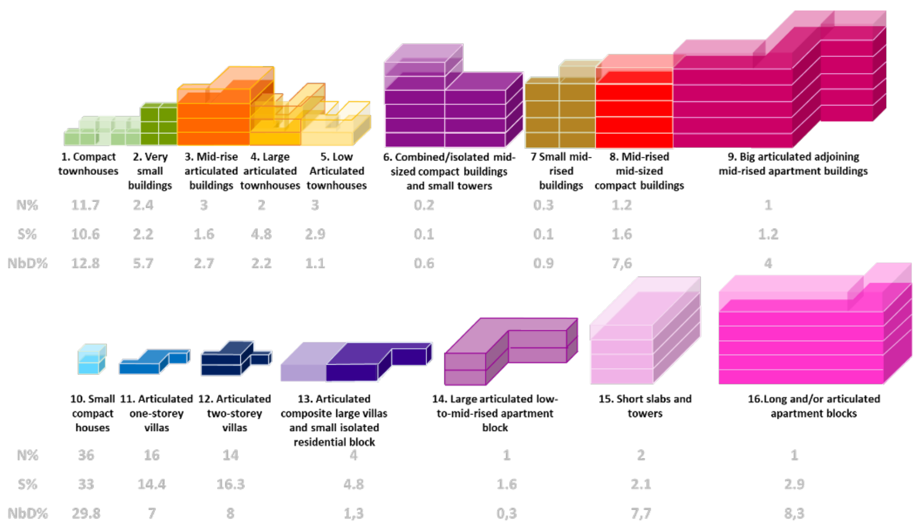


Figure 6. A general typology of residential buildings in France. A simplified graphical representation of building types based on 16-solution cluster profiles and their proportions in terms of building number (N), surface(S) and dwelling number (NbD)

Most of the C14, C15 and C16 housing production started from the 1960s and it is located in the close peripheries of large metropolitan areas. It encompasses both high standing apartment buildings and large housing projects developed by local and national social housing programs. Some of these buildings can also be found in more central and compact contexts (*i.e*. demolition and reconstruction of a large articulated building built over an entire block). Within the C15 and C16 groups we found some mid-sized collective building of the interwar periods while large collective buildings of the *HLM* (low-rent housing) *emergency* at the end of the 1950s and beginning of the 1960s are mainly classified as C15. Indeed, the historical profile of all the types C14, C15, C16 show an almost complete absence before the second world war, while a fast acceleration of their production during the economic boom of the *Trente Glorieuses* (50s, 60s and 70s), especially for C15 and C16. A slowdown during the 1980s saw an increased production of single-family suburban houses and villas (C10-C13). In the first decade of the XXI Century, we observe a rebound of apartment buildings production, this time privileging smaller and shorter types C14 and C15.

The geographic distribution of C15 and C16 follows the same spatial pattern: higher values of C16 are found in more urbanized departments of eastern and southern France with the highest shares found in the departments of Lyon, Marseille, Montpellier, Nice, Strasbourg and Paris region (over 7%) while C15 reaches its highest values in the alpine departments. Lower values of both C15 and C16 are found in the western part of France, especially in the Nouvelle Aquitaine region with a share of less than 1%. C14 on the contrary has a specific geographical pattern and is mainly found in northeastern departments of the Grand Est region and in the southwest[4].

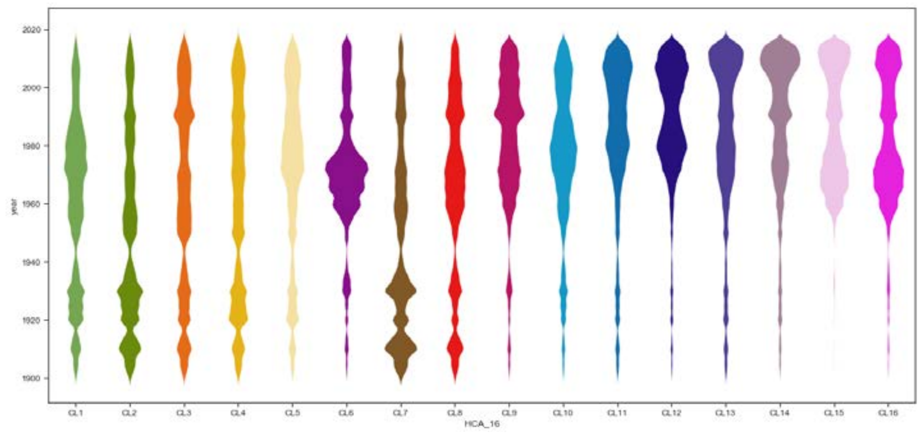


Figure 7. Historical profiles of building types. Violin plot of the building construction year (1900-2016) by types. Building age is provided for 54% of the ordinary building of France

4 As we discussed in the previous section, classes from 10 to 16, might present some misclassification error. Indeed, some building groups not present in the MAJIC cadastral database, have been modeled from satellite-based data, thus using their overall aerial footprint. Subsequently, some C11 and C12 articulated villas might instead corresponds to two or more adjoining single-family houses C1 and C10 respectively. Similarly, C15 short slabs might instead corresponds to few aligned adjoining townhouses of different shapes and size, while articulated apartment blocks C16 could encompass (heterogeneous) blocks of adjoining types of buildings. Nonetheless, these errors should be considered as a source data quality problem, rather than an error of the model. In order to reduce the amount of misclassification, future updated versions of the BD Topo with higher precision (for both footprint size and associated information) might be used to update the clustering assignment.

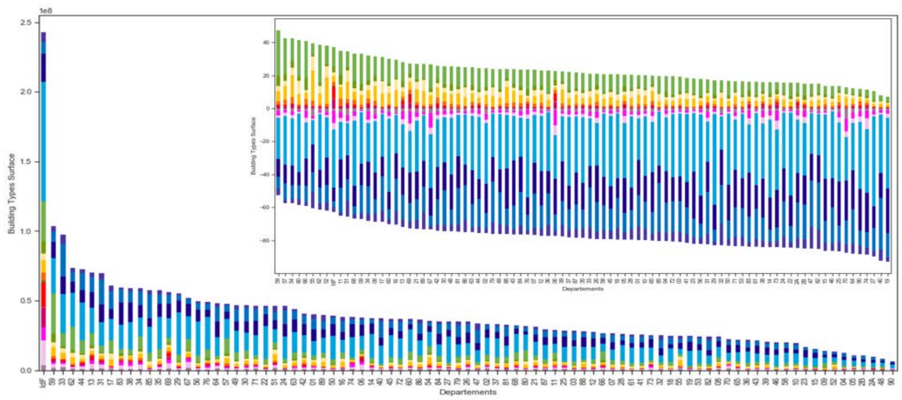


Figure 8. Building type by department: a) Building type surfaces and b) Share of the building type surface, divided in contiguous and isolated types (plotted as positive and negative values, respectively). The 8 departments of the Paris Metropolitan regions Ile de France have been merged.

5. Conclusion and perspectives

This paper presented an innovative protocol for the identification and description of nation-wide building types and their overall organization combining Naive Bayesian and Hierarchical clustering protocols. The outcomes of this systematic and quantitative analysis allow the data-driven derivation of a system of building types, hierarchically organized. The protocol is described and implemented for the real-world contemporary case study of the ordinary building stock of France. Sixteen building types have been identified and described. 65% of the ordinary building stock of France is made of three types of detached single-family houses. Half of them is made of small compact houses, a quarter of large bungalow houses and a quarter of large and articulated villas. Another 12% of the building stock is made of townhouses and rowhouses. The remaining 23% gathers eleven different types of collective housing: 5% takes the forms of isolated mid-sized and large collective buildings while 18% is composed by eight differently shaped types of adjoining buildings.

Several research perspectives can be outlined. From the methodological point of view, sensitivity analyses should be implemented to assess the robustness of the three main steps of the protocol presented in this paper. The first step considers the role of the variable discretization: the same protocol should be evaluated both with a general binning method and with segmentation approaches based on the specific statistical distribution of variables observed in the (sub)region under analysis. The second and third phases correspond to the two clustering protocols: both INBIAC and HCA can be assessed considering different distance measures and evaluating their validity under different parametric conditions. Independently of the specific parametric choices, a comparative analysis could also be carried out with more traditional approaches (e.g. k-means, DBSCAN) in order to further identify relative strengths and weaknesses of the protocol here proposed, both in the specific context of building type identification but also within other thematic fields.

From the thematic point of view of building typology, four major directions for further developments can be outlined. Firstly, this same protocol can be tested with an incremental number of descriptors of the building envelope aiming at testing and identifying the role played by individual morphometrics into the building typologies. This work might contribute to the debate on the definition of a reliable and universally accepted set of characters and variables for the identification of building envelope typologies. Moreover, internal layout and details of style, facade, roof coverage might also be included: the implementation of the same clustering approach with different levels and granularity of information can shed a new light on the relative role played by skeletal, internal and stylistic features in the identification and definition of building typologies. Secondly, future studies should implement cross-analysis combining building typologies with finer spatial and temporal descriptors in order to identify geographical patterns and their evolution over time. Similarly, the cross analysis with socioeconomic data might shed a new light on the relationship between urban form and its functioning. Nonetheless, building types represent only one of the main constituents of the urban form. The third research perspective is, indeed, the analysis of the spatial organization of building types, their relative cooccurrences and their spatial layout within the urban fabric, including the street network and the plot system. The study of the spatial relationships of building types and other urban form elements represents a key factor in the definition of streetscapes, urban fabrics and morphological regions. The work here described represents indeed only the first step of an undergoing larger research project aiming at understanding the urban fabrics of French contemporary cities.

Bibliography

A'urba (2011). *Rénovation thermique du parc bâti résidentiel de la CUB Définition d'une méthodologie.* Rapport technique, Agence d'urbanisme Bordeaux Aquitaine.

APUR (2007). *Consommations d'énergie et émissions de gaz à effet de serre liées au chauffage des résidences principales parisiennes*. Rapport technique, Atelier parisien d'Urbanisme.

Barrère P. (1956). Les quartiers de l'agglomération bordelaise (à suivre). *Revue géographique des Pyrénées et du Sud-Ouest. Sud-Ouest européen*, vol. 27, n° 2, p. 161-194.

Biljecki F., Ledoux H., Stoter J., Vosselman G. (2016). The variants of an LOD of a 3D building model and their influence on spatial analyses. *ISPRS Journal of Photogrammetry and Remote Sensing*, n° 116, p. 42-54.

Bonhomme M. (2013). Contribution to the generation of multiscalar and evolutionary databases for a multidisciplinary approach to urban energy, Doctoral dissertation, INSA Toulouse.

Bonillo J.-L. *et al.* (1988) *Atlas des formes urbaines de Marseille. Vol. 1 - Les types. INAMA*, Marseille.

Brown F. E., Steadman J. P. (1991a). The morphology of British housing: An empirical basis for policy and research. Part I: Functional and dimensional characteristics. Environment and Planning B: planning and design, vol. 18, n° 3, p. 277-299.

Brown F. E., Steadman J. P. (1991b). The morphology of British housing: an empirical basis for policy and research. Part 2: topological characteristics. Environment and Planning B: Planning and Design, vol. 18, n° 4, p. 385-415.

Callais C. (2018). *L'échoppe de Bordeaux*, éd. LA Geste, La Crèche.

Callais C. (2015). Interpréter les chais : le « grand îlot » des Chartrons à Bordeaux. In Situ. *Revue des patrimoines*, 26.

Caniggia G., Maffei G. (2017, original in Italian 1979). *Interpreting basic buildings.* Alinea, Firenze.

Carmona M. (2019). Place value: place quality and its impact on health, social, economic and environmental outcomes, *Journal of Urban Design*, vol. 24, n° 1, p. 1-48.

Carneiro EM, *et al.* (2015). Cluster analysis and artificial neural networks: A case study in credit card fraud detection. *12th International Conference on Information Technology-New Generations*. IEEE, 2015.

Deffontaines P (1972). *L'homme et sa maison*. Gallimard, Paris.

Dempster AP., Laird N.M., Rubin D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Series B., vol. 39, n° 1, p. 1-38

Deyon P. (1983). Roubaix dans la première moitié du xixe siècle. *Construire la ville : xviiie-xxe siècles*. Garden M. & Lequin Y. (Eds.), Presses universitaires de Lyon.

Duda RO., Hart PE. (1973). *Pattern classification and scene analysis*. NY: John Wiley & Sons.

Evans S., Liddiard R., Steadman P. (2019). Modelling a whole building stock: Domestic, non-domestic and mixed use. *Building Research & Information*, vol. 47, n° 2, p. 156-172.

Fleischmann M., Feliciotti A., Romice O., Porta S. (2022). Methodological foundation of a numerical taxonomy of urban form. *Environment and Planning B: Urban Analytics and City Science*, vol. 49, n° 4, p. 1283-1299.

Fusco G., Perez J. (2019) Bayesian Clustering and SOM Neural Networks under the Test of Indian Districts. A comparison. Cybergeo, n°887, https://journals.openedition.org/cybergeo/31909

Fourcaut A. (2000). *La banlieue en morceaux. La crise des lotissements défectueux en France dans l'entre deux-guerres*. Grâne : Créaphis.

Garbasevschi O.M., Schmiedt J.E., Verma T., *et al.* (2021). Spatial factors influencing building age prediction and implications for urban residential energy modelling. *Computers, Environment and Urban Systems,* 88, p. 101637.

Guignet P. (2008). Cours, courées et corons. Contribution à un cadrage lexicographique, typologique et chronologique de types d'habitat collectif emblématiques de la France du Nord. *Revue du Nord*, vol. 90, n° 374, p. 29-47.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, p. 857-871.

Hartmann A., Meinel G., Hecht R., Behnisch M. (2016). A workflow for automatic quantification of structure and dynamic of the German building stock using official spatial data. *ISPRS International Journal of Geo-Information*, vol. 5, n° 8, p. 142.

Hecht R., Meinel G. Buchroithner M. (2015). Automatic identification of building types based on topographic databases – a comparison of different data sources. *International Journal of* Cartography, vol. 1, n° 1, p. 18-31

Henn A., Römer C., Gröger G. and Plümer L. (2012). Automatic classification of building types in 3D city models. GeoInformatica, vol. 16, n° 2, p. 281-306.

Herrmann L. (2017). Fabriquer la ville avec les lotissements. Une qualification possible de la production ordinaire des espaces urbains contemporains ? PhD Dissertation Université de Lausanne/Université Lumière Lyon 2, Suisse/France.

Haffner M. (2022) L'impact des formes urbaines dans la mise en place des politiques de transition énergétique : une approche par la modélisation. PhD Dissertation. École des Ponts ParisTech, 2022. France.

IAU (2010). Amélioration énergétique du parc résidentiel francilien: les enjeux socio-économiques. Rapport Technique, IAU Ile de France, 177p.

IGN (2020) BD TOPO®, Version 3.0. Descriptif de contenu. https://geoservices.ign.fr/ressources_documentaires/Espace_documentaire/BASES_VECTORIELLES/BDTOPO/DC_BDTOPO_3-0.pdf

LSE Cities (2014) Cities and energy: urban morphology and heat energy demand. LSE Cities, London School of Economics and Political Science, London, UK.

Lopez-Rubio E., Palomo E. J., and Ortega-Zamorano F. (2018). Unsupervised learning by cluster quality optimization. *Information Sciences*, 436, p. 31-55

Maffei G.L., Maffei M. (2018, original in Italian 2011) *Interpreting specialised buildings*, AltraLinea, Firenze.

Maiullari D., Pijpers-van Esch M., Van Timmeren A. (2021). A Quantitative Morphological Method for Mapping Local Climate Types. Urban Planning, vol. 6, n° 3, p. 240-257.

McCaffrey J. (2013) Data Clustering Using Naive Bayes Inference. http://msdn.microsoft.com/en-us/magazine/jj991980.aspx.

Meinel G., Hecht R., Herold H. (2009). Analyzing building stock using topographic maps and GIS. *Building Research & Information,* vol. 37, n° 5-6, p. 468-482.

Monti S., Tamayo P., Mesirov J., Golub T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, vol. 52, n° 1-2, p. 91-118.

Moudon A. V. (1997). Urban morphology as an emerging interdisciplinary field. Urban morphology, vol. 1, n° 1, p. 3-10.

Neidhart H., Sester M. (2004). Identifying building types and building clusters using 3-D laser scanning and GIS-data. Int Arch Photogramm Remote Sens Spatial Inf Sci, 35, p. 715-720.

Oliveau S, Doignon Y. (2016). La diagonale se vide? Analyse spatiale exploratoire des décroissances démographiques en France métropolitaine depuis 50 ans. *Cybergeo: European Journal of Geography.*

Orford S., Radcliffe J. (2007). Modelling UK residential dwelling types using OS Mastermap data: A comparison to the 2001 census. *Computers, Environment and Urban Systems*, vol. 31, n° 2, p. 206-227.

Perez J., Fusco G., Araldi, A., Fuse, T. (2020). Identifying building typologies and their spatial patterns in the metropolitan areas of Marseille and Osaka. *Asia-Pacific Journal of Regional Science*, 4, n° 1, p. 193-217.

Prouvost, J. (1969). Les courées à Roubaix. *Revue du Nord*, vol. 51, n° 201, p. 307-316.

Rewienska W. (1937). Quelques remarques sur la physionomie de la ville de Toulouse. Revue géographique des Pyrénées et du Sud-Ouest. Sud-Ouest Européen, vol. 8, n° 1, p.73-88.

Rochard U., Shanthirabalan S., Brejon C., Chateau Le Bras M. (2015). *Bâtiments résidentiels, Typologie du parc existant et solutions exemplaires.* Rapport Technique, Pouget-Consultants.

Römer C., Plümer L. (2010). Identifying architectural style in 3d city models with support vector machines. *Photogrammetrie-Fernerkundung-Geoinformation*, p. 371-384.

Case-Scheer, B. (2017). Urban morphology as a research method. *Planning Knowledge and Research,* Routledge, p. 167-181.

Schirmer P. M., Axhausen K. W. (2016). A multiscale classification of urban morphology. Journal of Transport and Land Use, 9, n° 1), p. 101-130.

Sester M. (2000). Knowledge acquisition for the automatic interpretation of spatial data. *International Journal of Geographical Information Science* vol. 14, n° 1, p. 1-24.

Smith D., Crooks A. (2010). *From buildings to cities: techniques for the multi-scale analysis of urban form and function.* London: Centre for Advanced Spatial Analysis, University College London: Working Paper 155.

Steadman P. (2014). *Building Types and Built Forms*. Troubador Publishing Ltd.

Steadman P. (2016). Research in architecture and urban studies at Cambridge in the 1960s and 1970s: what really happened. *The Journal of Architecture* vol. 21, n° 2, p. 291-306

Steadman P., Bruhns H. R., Holtier S., *et al.* (2000). A classification of built forms. *Environment and planning B: Planning and design*, vol. 27, n° 1, p. 73-91.

Steiniger S., Lange, T., Burghardt, D., Weibel R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS*, vol. 12, n° 1, p. 31-59.

Werder S., Kieler B., Sester M. (2010). Semi-automatic interpretation of buildings and settlement areas in user-generated spatial data. *18th SIGSPATIAL international conference on advances in geographic information systems,* p. 330-339.

Wiel M. (1999) *La transition urbaine ou le passage de la ville pédestre à la ville motorisée*. Mardaga, Liège.

Wurm M., Taubenbock H., Roth A., Dech S. (2009). Urban structuring using multisensoral remote sensing data: By the example of the German cities Cologne and Dresden. In *2009 Joint Urban Remote Sensing Event*, IEEE, p. 1-8.

Annexes A – Building cluster profiles: the probabilistic distribution of the six morphological features associated to each building class.

Annexes B – Building types examples and geographical distribution.

	Building Footprint Surface					
	0-75	75-150	150-300	300-600	600-1200	1200+
Cl1	30,5	69,5	0,0	0,0	0,0	0,0
Cl2	28,1	71,4	0,5	0,0	0,0	0,0
Cl3	0,0	0,0	69,2	30,8	0,0	0,0
Cl4	0,0	0,0	62,6	26,6	7,8	3,1
Cl5	0,0	0,0	63,1	25,3	7,7	3,8
Cl6	0,6	4,5	54,1	40,8	0,0	0,0
Cl7	11,4	88,6	0,0	0,0	0,0	0,0
Cl8	0,0	0,0	39,1	60,9	0,0	0,0
Cl9	0,0	0,0	0,0	0,0	59,7	40,3
Cl10	13,5	86,5	0,0	0,0	0,0	0,0
Cl11	0,0	0,0	100,0	0,0	0,0	0,0
Cl12	0,0	0,0	74,7	25,3	0,0	0,0
Cl13	0,0	0,0	0,0	78,6	14,8	6,6
Cl14	0,0	0,0	0,0	0,0	72,2	27,8
Cl15	0,0	0,0	9,9	90,1	0,0	0,0
Cl16	0,0	0,0	0,0	0,4	57,2	42,4

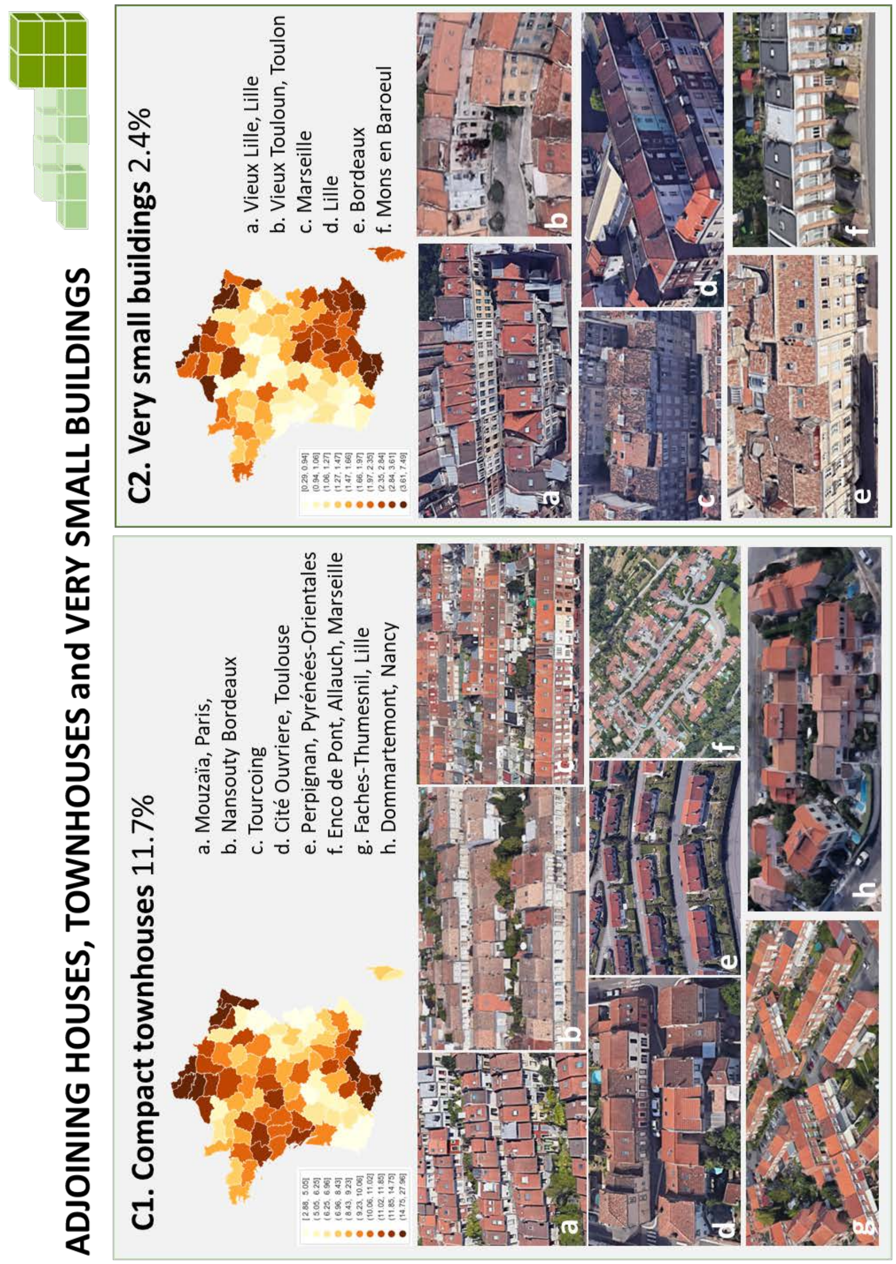
	Building Contiguity			
	1	2-3	4-8	8+
Cl1	0,00	26,76	46,19	27,05
Cl2	0,00	21,78	23,67	54,55
Cl3	0,00	15,04	31,08	53,88
Cl4	0,00	31,35	44,36	24,30
Cl5	0,00	38,62	45,29	16,10
Cl6	39,95	60,05	0,00	0,00
Cl7	0,00	2,58	12,53	84,89
Cl8	0,00	10,67	24,96	64,37
Cl9	0,00	20,86	36,16	42,98
Cl10	79,27	20,73	0,00	0,00
Cl11	85,10	14,90	0,00	0,00
Cl12	83,65	16,35	0,00	0,00
Cl13	80,80	19,20	0,00	0,00
Cl14	81,32	18,68	0,00	0,00
Cl15	83,87	16,13	0,00	0,00
Cl16	84,33	15,67	0,00	0,00

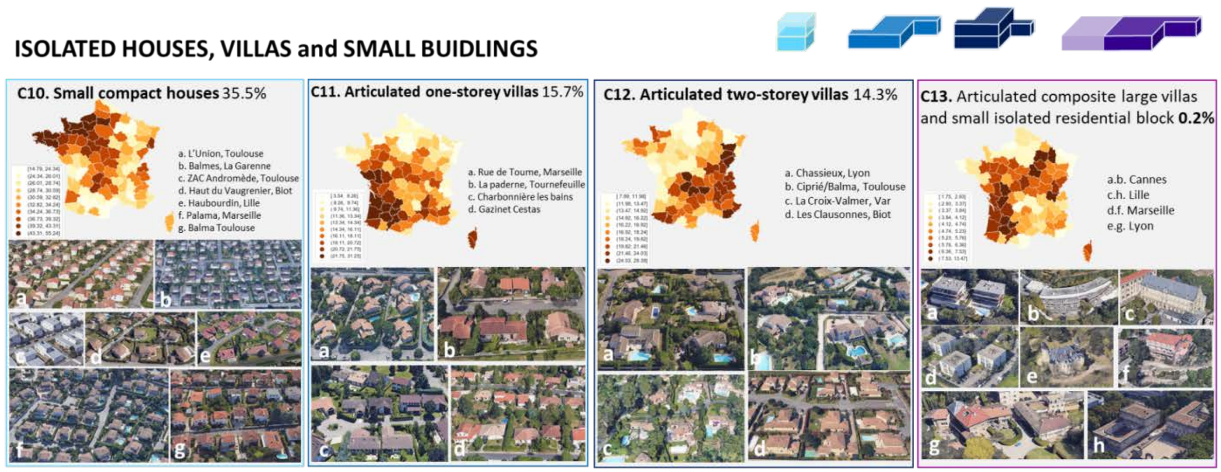
	Building Footprint Convexity				
	0-0,8	0,8-0,9	0,9-0,96	0,96-0,9999	0,9999+
Cl1	8,41	18,94	20,57	29,60	22,48
Cl2	6,56	16,93	19,86	31,24	25,42
Cl3	25,68	25,37	18,64	22,46	7,85
Cl4	35,90	25,22	15,90	16,93	6,05
Cl5	18,73	22,93	21,26	15,85	21,23
Cl6	8,39	17,93	23,29	26,87	23,53
Cl7	5,91	18,06	22,24	32,39	21,39
Cl8	18,50	26,70	21,73	25,22	7,86
Cl9	58,71	18,95	9,70	10,72	1,92
Cl10	3,61	15,27	23,98	21,08	36,07
Cl11	15,16	32,27	24,69	14,40	13,48
Cl12	22,64	30,19	21,30	12,94	12,93
Cl13	38,79	25,57	16,06	10,89	8,70
Cl14	60,07	18,76	10,14	6,48	4,55
Cl15	36,97	24,88	15,88	16,37	5,89
Cl16	58,89	17,23	9,06	7,02	7,81

	Building Number of Floors					
	1	2	3	4-5	6-7	8+
Cl1	27,95	72,05	0,00	0,00	0,00	0,00
Cl2	0,00	0,06	77,01	22,93	0,00	0,00
Cl3	0,00	0,00	67,96	32,04	0,00	0,00
Cl4	0,00	100,00	0,00	0,00	0,00	0,00
Cl5	96,53	0,00	3,47	0,00	0,00	0,00
Cl6	0,00	0,00	0,11	0,00	43,20	56,69
Cl7	0,00	0,00	0,00	2,60	85,45	11,95
Cl8	0,00	0,00	0,00	39,96	41,14	18,90
Cl9	0,00	0,00	7,67	39,28	28,60	24,45
Cl10	49,17	48,84	1,85	0,13	0,00	0,01
Cl11	98,87	0,00	1,13	0,00	0,00	0,00
Cl12	0,00	94,81	4,47	0,72	0,00	0,00
Cl13	84,49	0,00	10,48	4,75	0,20	0,08
Cl14	0,00	100,00	0,00	0,00	0,00	0,00
Cl15	0,00	6,65	22,70	54,88	8,96	6,81
Cl16	2,66	0,68	14,82	53,54	16,12	12,18

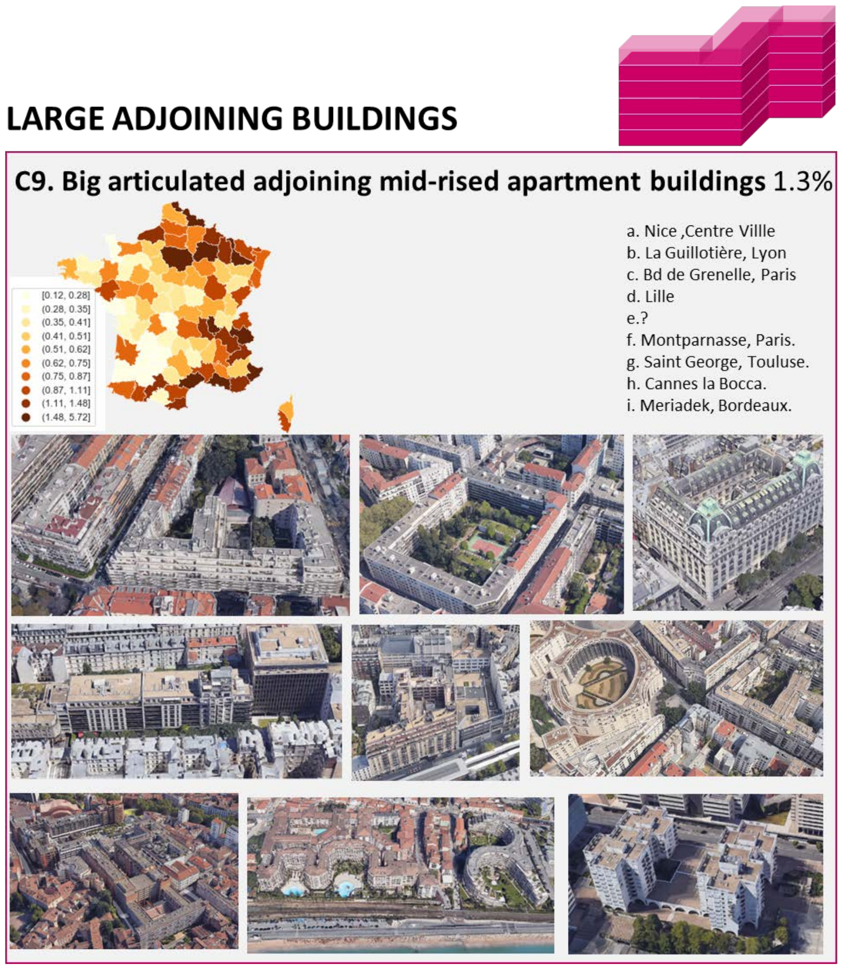
	Building Number of Dwellings			
	1	2-8	8-24	24+
Cl1	93,47	6,52	0,00	0,00
Cl2	49,44	47,67	2,89	0,00
Cl3	16,52	73,78	9,37	0,33
Cl4	73,32	25,19	1,32	0,17
Cl5	84,43	14,93	0,52	0,13
Cl6	7,43	8,47	44,44	39,66
Cl7	2,74	32,61	61,46	3,19
Cl8	0,51	5,36	71,37	22,76
Cl9	1,88	5,62	28,30	64,19
Cl10	96,57	3,40	0,03	0,00
Cl11	94,42	5,55	0,03	0,00
Cl12	80,61	19,39	0,00	0,00
Cl13	71,67	28,33	0,00	0,00
Cl14	43,39	36,97	15,40	4,24
Cl15	0,20	1,77	77,90	20,13
Cl16	0,00	0,00	30,72	69,28

	Building Footprint Elongation				
	1-1,15	1,15-1,2	1,2-1,3	1,3-1,5	1,5+
Cl1	20,90	21,23	26,63	22,52	8,72
Cl2	21,77	20,89	25,28	22,37	9,68
Cl3	8,81	12,05	20,34	27,42	31,37
Cl4	5,97	8,61	17,77	29,46	38,18
Cl5	8,94	9,14	20,21	32,87	28,85
Cl6	16,99	18,26	27,29	24,35	13,11
Cl7	18,87	19,65	25,14	24,30	12,04
Cl8	9,18	12,18	21,29	27,84	29,51
Cl9	1,67	2,98	6,75	17,57	71,03
Cl10	22,85	27,82	30,73	15,66	2,94
Cl11	6,10	12,24	31,67	36,39	13,60
Cl12	5,87	10,32	24,98	35,70	23,12
Cl13	3,55	5,76	14,48	29,59	46,62
Cl14	1,81	3,67	8,34	17,28	68,90
Cl15	5,32	8,06	17,58	29,96	39,08
Cl16	1,16	1,96	5,18	15,24	76,46


ADJOINING HOUSES, TOWNHOUSES and VERY SMALL BUILDINGS
C1. Compact townhouses 11.7%
a. Mouzaïa, Paris,
b. Nansouty Bordeaux
c. Tourcoing
d. Cité Ouvriere, Toulouse
e. Perpignan, Pyrénées-Orientales
f. Enco de Pont, Allauch, Marseille
g. Faches-Thumesnil, Lille
h. Dommartemont, Nancy
C2. Very small buildings 2.4%
a. Vieux Lille, Lille
b. Vieux Touloun, Toulon
c. Marseille
d. Lille
e. Bordeaux
f. Mons en Baroeul


ISOLATED HOUSES, VILLAS and SMALL BUIDLINGS
C10. Small compact houses 35.5%
a. L'Union, Toulouse
b. Balmes, La Garenne
c. ZAC Andromède, Toulouse
d. Haut du Vaugrenier, Biot
e. Haubourdin, Lille
f. Palama, Marseille
g. Balma Toulouse
C11. Articulated one-storey villas 15.7%
a. Rue de Toume, Marseille
b. La paderne, Tournefeuille
c. Charbonnière les bains
d. Gazinet Cestas
C12. Articulated two-storey villas 14.3%
a. Chassieux, Lyon
b. Ciprié/Balma, Toulouse
c. La Croix-Valmer, Var
d. Les Clausonnes, Biot
C13. Articulated composite large villas and small isolated residential block 0.2%
a.b. Cannes
c.h. Lille
d.f. Marseille
e.g. Lyon


ISOLATED LARGE COLLECTIVE BUILDINGS
C14: Large articulated low-to-mid-raised apartment block 0.8%
a. e. Marseille
b. c. Lille
d. g. Lyon
f. Bordeaux
C15. Short slabs and towers 2.3%
a. Chartrons, Bordeaux
b. Empalot, Toulouse
c. Cité Madrid, Toulouse
d. Haussmann, Bordeaux
C16. Long and/or articulated apartment blocks 2.4%
a. Bagatelle, Toulouse
b. Mirail, Reynerie, Toulouse
c. La Ceinture, Paris
d. La Rouvière, Marseille
e. Bastion Sant André Lille
f. Palmer, Cenon - Bordeaux


LARGE ADJOINING BUILDINGS
C9. Big articulated adjoining mid-rised apartment buildings 1.3%
[0.12, 0.28]
(0.28, 0.35]
(0.35, 0.41]
(0.41, 0.51]
(0.51, 0.62]
(0.62, 0.75]
(0.75, 0.87]
(0.87, 1.11]
(1.11, 1.48]
(1.48, 5.72]
a. Nice ,Centre Villle
b. La Guillotière, Lyon
c. Bd de Grenelle, Paris
d. Lille
e.?
f. Montparnasse, Paris.
g. Saint George, Touluse.
h. Cannes la Bocca.
i. Meriadek, Bordeaux.