A geoprocessing framework to compute urban indicators: The MApUCE tools chain

Erwan Bocher^{a,*}, Gwendall Petit^b, Jérémy Bernard^a, Sylvain Palominos^a

^a CNRS, Lab-STICC laboratory UMR 6285, Vannes (France)

^b Université de Bretagne Sud, Lab-STICC laboratory UMR 6285, Vannes (France)

Abstract

A growing demand from urban planning services and various research thematics concerns urban fabric characterization. Several projects (such as WU-DAPT) are currently lead in the urban climate field to answer this demand. However there is currently a need to propose standardized methods to calculate urban indicators and to automatically classify the urban fabric for any city in the world as well as to propose platforms to share these methods and the associated results. Our contribution answers partially to this challenge. A total of 64 standardized urban morphological indicators are calculated for three scales of analysis : building, block and a reference spatial unit (RSU). A supervised classification is performed for the building and the RSU scales using a regression trees model based on these indicators and on 10 urban fabric typological classes defined by urbanists and architects. A processing chain is proposed to realize indicator calculation and urban fabric classification for any french municipality according to reference data provided by the French National Geographical Institute (IGN). Spatial reasoning and morphological indicators description are formalized with SQL language and statistical analysis is carried out with R language. Finally a geoprocessing framework based on free and open source softwares, conform to the Open Geospatial Consortium (OGC) standards and ready to serve open data is built. Indicators values and classification results for 6% of the french municipalities (corresponding to 41% of all french buildings) are available through a web cartographic portal by any person interested in such analysis.

*Corresponding author

Email address: `firstname.lastname@univ-ubs.fr` (Erwan Bocher)

Keywords: Morphological indicators, Local Climate Zone, GIS, Urban Heat Island, Open source

1 Introduction

2 According to the Intergovernmental Panel on Climate Change (IPCC)
3 projections, global surface temperature will increase during the XXI^{st} cen-
4 tury. In the meantime, the world population living in cities is expected to
5 grow (5,058 millions by 2030 against 4,250 millions in 2018 - [1]). Two factors
6 explain this number: the population of existing cities will grow, and new cities
7 will appear. Urbanization often implies urban temperature rise due to land
8 cover change (pervious to impervious ground [2, 3]) and morphology change
9 (new buildings mean more short and long-wave radiation trapping as well as
10 wind speed decreasing [4]). Without urbanization control, this phenomenon
11 called Urban Heat Island (UHI) may become more intense since tempera-
12 ture differences between an urban area and its surrounding is proportional
13 to the logarithm of its population [5, 6]. The combination of climate change
14 and UHI may lead to higher heat related death occurrence [7, 8] and higher
15 energy consumption related to cooling use [9, 10]. Therefore, the reduction
16 of the urban heat island phenomenon may contribute both to attenuate the
17 climate change (by reducing urban greenhouse emissions) and to mitigate
18 its impacts. Several levers have proved their efficiency to lower urban air
19 temperature such as surface painting to modify the albedo, planting trees or
20 covering roofs and facades with low vegetation, decreasing energy consump-
21 tions, *etc.* [11, 12]. Santamouris et al. [11] showed that their performance and
22 surface application potential differ greatly depending on the urban environ-
23 ment where they are applied in. To study the influence of urban morphology
24 and urban land-cover on urban air temperature as well as the efficiency of
25 each UHI counter-measure, urban climate models have been developed [13]
26 and urban classification have been proposed [14]. Urban climate models are
27 applied to a grid of urban mesh. For each mesh, several urban parameters
28 are needed such as mean building height, aspect ratio, etc. Concerning the
29 urban classification, the territory is also split into elementary units which
30 are then classified according to a Local Climate Zone (LCZ) definition based
31 on urban parameters describing urban morphology, urban land-cover, urban
32 land-use and material properties [15]. Research has been made to identify
33 LCZ within urban areas from geographical data. Most of the methods are

34 based on a process using three steps. First the territory is split according to
 35 a certain grid. Second, urban parameters are calculated within each mesh
 36 from vector database or satellite images. Third, rules are created to allocate
 37 each mesh to a LCZ. Each of these steps may be manually [16] or auto-
 38 matically [17, 18] performed. The limitations of those works is their lack of
 39 reproducibility. The manual classification is time-consuming and based on
 40 expert analysis. The automatic classifications proposed by Lelovics et al.
 41 [17], Zheng et al. [18] rely on local data and on their own urban indicator
 42 definition. Thus simulation and classification approaches are very sensitive
 43 to data and methodology used to calculate urban indicators (characteristic
 44 of the morphology and the land cover of the urban fabric) [19]. To obtain
 45 comparable indicators at world scale, there is a need :

- 46 • to standardize data and methodology used for urban indicators calcu-
 47 lation [20],
- 48 • to propose collaborative and open tools to allow any user to calculate
 49 urban indicators for the city of its choice, thus allowing to share and
 50 reuse results from any calculation.

51 In this spirit, a collaborative project called *World Urban Database and*
 52 *Access Portal Tools* (WUDAPT¹) gathers a community of researchers to clas-
 53 sify the urban fabric by climate properties from homogeneous and available
 54 data at world scale. The objective is to identify Local Climate Zones as
 55 defined by Stewart and Oke [15]. The first step of the project have been
 56 applied. The LCZ of several urban areas have been identified according to
 57 supervised machine learning method using Landsat images (30 m resolution)
 58 as input and LCZ identified by climate expert from Google Earth software
 59 as desired output [21]. However, WUDAPT is open to improvements:

- 60 • the need to install locally several softwares (Google Earth², SAGA³)
 61 on its computer may be a break to collaborative contribution,
- 62 • it is now necessary to provide data and urban indicators at finer scale
 63 [20]. Plenty of indicators exist but they have several definitions and

¹<http://www.wudapt.org/> accessed in July 2017

²<https://www.google.com/earth/> accessed in July 2017

³<http://www.saga-gis.org/> accessed in July 2017

64 they are implemented within different softwares using numerous lan-
 65 guages and methods. Thus comparing the value of such undefined
 66 indicator throughout the world or along time is impossible [22].

67 Our contribution consists in the production of standardized urban mor-
 68 phological indicators dedicated for urban climate and useful for any other ur-
 69 ban planning purpose. It is a component of a french research project called
 70 MApUCE⁴ and is encompassed in a task of urban tissue characterization,
 71 illustrated in Figure 1.

Figure 1: The main steps of the method

72 Input data are produced from reference data. They are used to pro-
 73 duced both morphological and socio-economic indicators, that will be used
 74 to classify the urban fabric into typological classes. In this article, we will fo-
 75 cus on the morphological indicators production and we will described briefly
 76 the classification step. Further details concerning the input data production
 77 and the socio-economic indicators production are available in Plumejeaud-
 78 Perreau et al. [23] whereas the classification process is further described in
 79 Faraut et al. [24], Masson et al. [25]. Because one of the objective is that
 80 the overall process be reproducible simply without any software requirement,
 81 this paper proposes an open geoprocessing framework based on free and open
 82 source softwares, conform to the Open Geospatial Consortium standards and
 83 ready to serve open data.

⁴<http://www.umn-cnrm.fr/ville.climat/spip.php?rubrique120> accessed in July 2017

1. Data

1.1. Scale definition

Whereas streets may be considered as more durable than blocks and buildings [26], building is the elementary object structuring the territory [27] and also the object of interest when focusing on urban climate application [28]. However, building scale is not appropriate when dealing with issues at city scale. For this reason, Berghauser-Pont and Haupt [29] proposed five scales to analyze urban areas : buildings, lots, island, fabric and district. The first described only building properties whereas the others described the building properties and their surrounding environment. Lots are defined by the legal boundaries specified in the cadastral map. Islands include several lots limited by road boundaries. Fabrics include several islands as well as the road network whereas districts gather several fabrics and include public parks and water surfaces. All these scales are the result of arbitrary objects aggregation, except building and lot. In this context, we define the building as the elementary scale. A second scale is chosen: the building block, defined by Berghauser-Pont and Haupt [29] as an aggregation of buildings that are in contact. This scale is particularly adapted when dealing with building energy or urban climate issues [30]. By simplification, it will be called block in this paper. To consider all the components of the urban context, the legal boundaries specified in the cadastral map are also utilized (such as the lot defined by Berghauser Pont). This scale offers the advantage to have a size close to the one usually recommended in the urban climate literature (several hundred meters wide - [28]). They are slightly modified to include public spaces such as road surfaces, public parks and water surfaces. The generic name of Reference Spatial Unit (RSU) is set for the resulting feature. Any other well defined geographical entity may be used as an RSU, such as the urban block defined by the road network [31]. Finally, three scales of analysis will be considered: building, blocks and RSU.

1.2. Reference data

The reference data sets of the MApUCE tool chain are provided from the french national databases which are freely available for research and academic purpose. They have been produced following identical rules since many years and they offer a complete spatial coverage of the French territory. It concerns two types of data: spatial (Table 1) and statistical (Table 2).

data set	Description
BD Topo®	Topographic data, in vector format, provided by the French National Geographical Institute (IGN) (see http://professionnels.ign.fr/bdtopo). The data are classified in ten topics. Each topic contains a set of layers distributed in a GIS file format. <i>e.g</i> "BUILDING" theme includes undefined, industrial or remarkable building layers, ...
Parcels	Cadastral parcels, in vector format, provided by IGN (see http://professionnels.ign.fr/bdparcellaire)
Gridded population	This data set depicts the distribution of human population across the french territory. The data is distributed by the French National Institute for Statistics and Economic Studies (INSEE) (see https://www.insee.fr/en/accueil)
IRIS contouring	The IRIS contouring contains a set of polygons that represents an area of 2,000 grouping inhabitants. The median area is about 740 ha, and maximal size is of 36,700 ha. This data set is provided by IGN.

Table 1: Input spatial data sets used by the MApUCE tool chain

data set	Description
Households survey	The french households survey is provided by the National Institute of Statistics and Economics and Studies (INSEE). This survey is linked to the IRIS contouring thanks to a key index.

Table 2: Input statistical data set used by MApUCE tool chain

119 1.3. Data pre-processing

120 The above data sets were used to derive 3 spatial layers computed in 3
121 main steps [23].

122 1.3.1. Step 1: Data cleaning and structuring

123 The quality aspects of the spatial data sets are inspected using quality
124 control metrics and assessment procedures. They are implemented using

the Structured Query Language (SQL) extended with spatial functions. The PostGreSQL-PostGIS database has been selected for this purpose. Five types of geometry inconsistencies are checked : redundancy (same geometry, same geometry with different attributes), overlapping (geometries having a surface in common), invalid, null, size (geometry area or length greater than a threshold). They are corrected using a rules based system. The following pseudo code illustrates the principle (Table 3)

```

1  if the geometry is null
2      then delete
3  else if the geometry is invalid
4      then correct
5  if the geometry overlaps another geometry
6      then remove the part of the geometry that have
           lowest overlapping area
7  ...

```

Table 3: Pseudo code to control and fix the geometry quality

The data quality processes are chained with a data structuring task used to organize the input data sets into main tables. This is especially the case for the BD Topo® data set that are grouped in two layers : *BUILDING* and *ROADS*. *e.g* the *BUILDING* table contains all features from the three vector layers undefined, industrial or remarkable building theme.

1.3.2. Step 2: RSU computing

A new partitioning of the urban territory is computed. Based on the dual of a Delaunay triangulation, its boundaries correspond to the medial axis of negative area of the union of the cadastral parcels. We call it "Reference Spatial Unit" (RSU).

The properties of the Voronoï tessellation are used to create the RSU (Figure 2). First, contiguous parcels are unified (1). Then the seeding points of the Voronoi tessellation are prepared. Those points are used to compute the Voronoi polygons (3). Finally, RSU are generated (4) and smoothed (5).

The RSU geometries are stored into one table. They are computed municipality by municipality. Each RSU is related to one and only one municipality using a national index named in the data.

Figure 2: The RSU generation process (adapted from Plumejeaud-Perreau et al. [23])

1.3.3. Step 3: Data enriching

Data enriching is the final step of the data preprocessing stage. It involves to integrate new variables on the two tables *BUILDING* and *RSU*. The integration is solved by chaining spatial analysis methods and aggregating processes. The tables 4 and 5 list the final variables computed. Note that these variables include the socio-economic indicators computed by [23].

The final result of this data pre-processing task is a set of 4 tables stored in a PostGIS database. They are the main entries for the MAppUCE tool chain (implying the computation of the required urban indicators). The Table 6 gives some statistics about the number of features yet processed.

2. Method

2.1. Morphological indicators

The calculation method of a wide range of indicators is presented in this section. The list of indicators has been established based on the results of a literature review (mixing urban climate and geography issues but also geographical issues only), discussions with architects and urban planners and needed indicators as input in urban climate models. Some of them are specific to the calculation scale: they are called "primary indicators". The others are

Variable	Description	Method	Step
pk_build	Building unique identifier	Incremental value (Primary Key)	1
the_geom	Building geometry	Geometry of the building	1
insee_code	Id of the commune that contains the building	Unique key value that refers to a municipality	3
pk_rsu	Id of the RSU that contains the building	Unique key value that refers to a RSU geometry. A spatial join process is used with a constrained area. e.g. If a building overlaps two RSU, the affected pk_rsu is the one corresponding to the maximal intersected area	3
h	Building height available in the BD Topo	-	1
h_fixed	Corrected height (<i>calculated from an iterative process using indicators computed in the section 2</i>)	<p>If $h = 0$ or <i>Null</i> then {if $\frac{h_std_{rsu}}{h_mean_{rsu}} < 0.5$ then $h_fixed = round(h_std_{rsu} - h_mean_{rsu})$ else $h_fixed = round(h_mean_{rsu})$ else $\{h_fixed = h\}$</p>	3
nb_level	Number of level deduced from $h_correct$	<p>For building with the <i>indifferencie</i> theme, if $h_fixed \geq 3$ $nb_level = round((h_fixed - \frac{4}{3}) + 1)$ else $nb_level = 1$</p>	3
insee_inhab	Number of inhabitants	Derived from INSEE 200m gridded cells	3
theme	Name of BDTopo theme	Building theme from BD Topo : <i>industriel</i> (industrial), <i>remarquable</i> (remarkable) or <i>indifferencie</i> (undistinguished)	3

Table 4: List of variables in the BUILDING table

Variable	Description	Method	Step
pk_rsu	RSU unique identifier	Incremental value (Primary Key)	1
the_geom	RSUs geometry	Geometry of the RSU	2
veget_surface	Total vegetation surface	Area of vegetation intersecting the RSU	3
road_surface	Total road surface	Area of roads intersecting the RSU. This area is determined thanks to a width attribute included in the road layer. Spatial processes, using buffer and intersection are done to compute this area.	3
road_length	Total road length	Length of roads intersecting the RSU	3
sidewalk_length	Total length of sidewalk	Perimeter of the geometry resulting from the union of contiguous parcels	3
hydro_surface	Total hydrographic surface	Area of hydrological objects (based on <i>RESERVOIR_EAU</i> and <i>SURFACE_EAU</i> layers from BD Topo) intersecting the RSU	3
hydro_length	Total hydrographic length	Length of hydrological objects (based on <i>TRONCONCOURS_EAU</i> layer from BD Topo) intersecting the RSU	3
insee.inhabit	Number of inhabitants	*	3
insee.hh	Number of household	Number of households having a principal residence. *	3
insee.hh_coll	Number of households in collective dwellings	Number of households living in collective housing. *	3
insee.men_surf	Cumulative Surfaces of Main Residences in square meters	Cumulated area of housings for households having a principal residence computed in square meter. *	3
insee.surf_col	Estimation of the area of collective housing	Estimation of collective housing from INSEE indicators. *	3
insee.code	French municipality unique identifier	Transferring the municipality identifier from the municipality layer to the RSU geometry using a spatial join.	3

* Derived from INSEE 200m gridded cells

Table 5: List of variables in the RSU table

Data set	Description	Number of features
BUILDINGS	French buildings	8 942 135
ROADS	French road network	17 043 575
RSU	Reference Spatial Units	454 308
MUNICIPALITIES	French municipalities	36 553

Table 6: Number of features available after the pre-processing task

168 aggregated from primary indicators calculated at lower level: they are called
169 "derived indicators".

170

171 Three scales are considered for the morphological indicators production:
172 building, blocks and RSU (Figure 3). A block is an aggregation of buildings
173 that have at least one point in common when intersected.

Figure 3: The three scales of analysis

174 2.1.1. For buildings

175 27 indicators are computed at the building scale (Table 7).

176

Name	Description	Method	Biblio
area	Building area	Area of the building geometry (footprint)	[27]
floor_area	Building floor area	$area \cdot nb_level$	
vol	Building volume	$area \cdot h_fixed$	

Table 7: List of primary building indicators

Name	Description	Method	Biblio
perimeter	Building perimeter	Perimeter of the building geometry	
perimeter_cvx	Building convexhull perimeter	-	
form_factor	Building form factor	$\frac{area}{perimeter^2}$	[32]
nb_neighbor	Building number of neighbor	Number of buildings that are in contact (at least one point) with the building of interest	[27]
b_wall_area	Total area of building walls (including holes)	Sum of the linear of facades multiplied by the building height	
p_wall_long	Total length of common (party) walls	Sum of the linear of facades that are in contact with other buildings	
p_wall_area	Total area of common (party) walls	When the building has a common linear of facade with another one, the common wall area is the linear of facades multiplied by the height of the smallest building. Then the sum of these areas is realized for each neighbors in contact with the building.	
free_ext_area	Area of free external facades, that are in contact with the air	$\sum b_wall_area - \sum p_wall_area + \sum area$	
concavity	Building concavity	Building area divided by its convex hulls area	[33, 34]
contiguity	Building contiguity	$\frac{p_wall_area}{b_wall_area}$	[35]
compactness_r	Building raw compactness	$\frac{b_wall_area + area}{volume^{\frac{2}{3}}}$	[29]
compactness_n	Building net compactness	$\frac{free_ext_area}{volume^{\frac{2}{3}}}$	
compactness	Building compactness	Ratio between the building perimeter and the perimeter of a circle having the same area	[36, 37]

Table 7: List of primary building indicators

Name	Description	Method	Biblio
main_dir	Building main direction (in degree)	The main direction is defined as the direction given by the longest side of the building Smallest Minimum Bounding Rectangle (SMBR)	[38, 39]
p_vol_ratio	Building passive volume ratio. This ratio can be expressed as the building portion that can be naturally lit and ventilated.	Area taken up to 6 m from a free facade inside the building, then divided by the building area	[40, 41, 36]
fractal_dim	Building fractal dimension	$2 \cdot \frac{\log(perimeter)}{\log(area)}$	[42, 43]
min_dist	Distance between the building of interest and the closest building which is in the same RSU	Minimum distance between the building of interest and the other ones in the same RSU	
max_dist	Distance between the building of interest and the furthest building which is in the same RSU	Maximum distance between the building of interest and the other ones in the same RSU	
mean_dist	Mean distance between the building of interest and the other buildings which are in the same RSU	-	[44]
std_dist	Population standard deviation distance between the building of interest and the other buildings which are in the same RSU	-	

Table 7: List of primary building indicators

Name	Description	Method	Biblio
num_points	Building number of points	Count the building number of points after removing duplicate (e.g startpoint and endpoint are counted once)	[27]
l_3m	Linear of building walls next to road	For each building, total length of walls that are less than 3m far from the road	
l_ratio	Part of building walls next to road	$\frac{l_{3m}}{perimeter}$	
l_ratio_cvx	Ratio between linear of building walls next to road and the building convex-hull perimeter	$\frac{l_{3m}}{perimeter_{cvx}}$	

Table 7: List of primary building indicators

177 2.1.2. For blocks

178 A total of 9 indicators are computed at the block scale (tables 8 and 9).

179

Name	Description	Method	Biblio
area	Building area composing the block	Footprint area	
compacity	Block net compacity	$\frac{\sum free_ext_area}{Sumvol}^{\frac{2}{3}}$	
main_dir	Block main direction	The main direction is defined as the direction given by the longest side of the blocks Smallest Minimum Bounding Rectangle (SMBR)	
holes_area	Area of holes in a block	-	[27]
holes_ratio	Ratio between the holes area and the blocks area	$\frac{holes_area}{area+holes_area}$	

Table 8: List of primary block indicators

Name	Description	Aggregation method	Biblio
floor_area	Block floor area	$\sum floor_area$	
vol	Block volume	$\sum vol$	
h_mean	Block mean height	$\frac{\sum area \cdot h_fixed}{\sum area}$	[35]
h_std	Block standard deviation height	Block population standard deviation building height	

Table 9: List of derived block indicators

180 2.1.3. For RSU

181 A total of 9 indicators are computed at RSU scale (tables 10 and 11).

182

Name	Description	Formula	Biblio
build_numb	Number of buildings in the RSU	-	
dist_to_center	Distance to the city center	Distance between RSU centroid and the city center	

Table 10: List of primary RSU indicators

Name	Description	Aggregation method	Biblio
area	Building area in the RSU	$\sum area_{build}$	
floor_area	Building floor area in the RSU	$\sum floor_area_{build}$	
floor_ratio	Building floor area ratio	$\frac{\sum floor_area_{build}}{rsu_area}$	[29, 35, 45, 46, 47]
vol	Building volume	$\sum vol_{build}$	
vol_m	Building mean volume	$\frac{\sum vol_{build}}{build_numb}$	
ext_env_area	Building external area	$\sum free_ext_area_{build}$	
compac_m_w	Building weighted mean compacity	$\frac{\sum compacity_n_{build} \cdot area_{build}}{\sum area_{build}}$	

Table 11: List of derived RSU indicators

Name	Description	Aggregation method	Biblio
compac_m	Building non-weighted mean capacity	$\frac{\sum capacity_{n_{build}}}{build_numb}$	
contig_m	Building mean contiguity	$\frac{\sum contiguity_{build}}{build_numb}$	
contig_std	Building standard deviation contiguity	Population standard deviation contiguity of buildings	
main_dir_std	Main direction of buildings standard deviation	Population standard deviation main direction of buildings	
h_mean	Building mean height	$\frac{\sum area_{build} \cdot h_{fixed_{build}}}{\sum area_{build}}$	
h_std	Building standard deviation height	Population standard deviation height of buildings	
p_vol_ratio_m	Building mean passive volume ratios	$\frac{\sum floor_area_{build} \cdot p_vol_ratio_{build}}{\sum floor_area_{build}}$	
min_m_dist	Mean of the minimum distance between buildings that are in the same RSU	$\frac{\sum min_dist_{build}}{build_numb}$	
mean_m_dist	Mean of the mean distance between buildings that are in the same RSU	$\frac{\sum mean_dist_{build}}{build_numb}$	
mean_std_dist	Standard deviation of the mean distance between buildings that are in the same RSU	Population standard deviation of the mean distance between buildings that are in the same RSU	
bl_hole_area_m	Mean courtyard ratio of blocks within an RSU	$\frac{\sum holes_ratio_{block} \cdot area_{block}}{\sum area_{block}}$	

Table 11: List of derived RSU indicators

Name	Description	Aggregation method	Biblio
bl_std_h_mean	Mean of the standard deviation height of buildings, computed at the blocks scale within a RSU.	$\frac{\sum h_std_{block} \cdot area_{block}}{\sum area_{block}}$	
bl_m_nw_comp	Block non weighted mean compacity	$\frac{\sum compacity_n_{block}}{block_numb}$	
bl_m_w_comp	Block weighted mean compacity	$\frac{\sum compacity_n_{block} \cdot area_{block}}{\sum area_{block}}$	
bl_std_comp	Blocks standard deviation compacity	Population standard deviation of block compacities	
build_density	Building density in the RSU (based on the RSU area called "rsu_area", computed on the fly)	$\frac{\sum area_{build}}{rsu_area}$	[29, 27, 46, 45, 47, 35]
hydro_density	Hydrographic areas density in the RSU	$\frac{hydro_surface}{rsu_area}$	[48, 49, 50]
veget_density	Vegetation areas density in the RSU	$\frac{veget_surface}{rsu_area}$	[48, 49, 50]
road_density	Road areas density in the RSU	$\frac{road_surface}{rsu_area}$	[48, 49, 50]

Table 11: List of derived RSU indicators

183 2.2. Urban fabric typology

184 Energy consumption and urban climate issues differ greatly throughout
185 a city depending on the urban structure, the building use and the socio-
186 economic profile of the inhabitants. Ten french types of urban fabric have
187 been identified using a review of technical literature combined with the result
188 of a survey addressed to urbanists [51] (Figure 4).

189 It is important to note that a simplified typology has been recently pro-
190 posed by Tornay et al. [52] to make these classes fit with the LCZ classes
191 (Table 12).

Figure 4: Typological classes used to classify the urban fabric

192 These classes have been used to automatically classify the urban fabric of
 193 any french municipality into each of this urban type. For this purpose, a su-
 194 pervised classification method has been used. First, a sample of 27,096 build-
 195 ings from 7 french conurbations have been manually classified from satellite
 196 images according to a predefined typological identification procedure (Fig-
 197 ure 5). Second, a classification algorithm is established to automatically
 198 allocate a building to one of these typological classes from its 78 morpholog-
 199 ical and 6 socioeconomic indicators values. For this purpose, 6 supervised
 200 classification methods are tested, based on 70 % of the buildings total sam-
 201 ple. Finally, each of these methods are evaluated from the last third of the
 202 building sample (30% of the total sample). The classification obtained using
 203 the regression tree analysis is finally selected since it has the lowest predic-
 204 tion error (11.06%). Any building from any French municipality may then
 205 be classified according to the corresponding algorithm as well as the mor-
 206 phological and socioeconomic indicators. The dominant building typological
 207 class within a RSU is finally selected to characterize the RSU scale. Further
 208 details regarding the methodology is available in Faraut et al. [24], Masson
 209 et al. [25].

Proposed urban typology	Simplified typology	Corresponding LCZ
Detached house	Low-rise	Sparsely built
Semi-detached house		Open low-rise
Row house on open island		Compact low-rise
Row house on closed island		
Detached building	Mid-rise	Open mid-rise
Linear building on open urban island		Compact mid-rise
Linear building on closed urban island		
High-rise building	High-rise building	Open high-rise
		Compact high-rise
Industrial building	Extended low-rise	Large low-rise
Informal building		Heavy industry
		Lightweight low-rise

Table 12: Link between the proposed urban typology and the LCZ typology through a simplified typology used for defining the architectural parameters in the simulation tool Town Energy Balance (TEB). Adapted from Tornay et al. [52]

210 3. Implementation

211 The development of standards for data description and data exchange
212 (interoperability) as well as the arrival of the concept of Spatial Data Infras-
213 tructure (SDI) facilitate the interconnection of systems and the implementa-
214 tion of systemic approaches [53, 54]. Several issues have been solved by the
215 Geographical Information Sciences (GIS) community in order to unify sys-
216 tems and tools and to organize the knowledge in the fields of spatial analysis
217 and cartography. The MApUCE geoprocessing framework takes advantage of
218 these trends. Based on open source tools, open standards and ready for open
219 data, it relies on full transparency and explicit references to both methods
220 and data to target: verifiability, cross-disciplinary studies, re-use, compati-
221 bility [55, 56].

222 3.1. Languages

223 To develop an open processing framework, two languages have been se-
224 lected: SQL and R. The first one has been used to formalize spatial reasoning
225 and to describe the morphological indicators. The second one has been cho-
226 sen to carry out statistical analysis.

Figure 5: Production of the sample data set

227 3.1.1. From indicators to SQL scripts

228 Various approaches have been proposed to manipulate spatial data and
 229 formalize spatial analysis [57, 58]. From the Map Algebra language [59] to
 230 GeoScript⁵ or GeoSPARQL [60], developers and scientists have shown great
 231 imagination and originality to propose extensions or new syntaxes and opera-
 232 tors to query geographical informations, including 3D, temporal, topological
 233 features, *etc.* Nevertheless, the Structured Query Language (SQL) extended
 234 with spatial capabilities remains the heart of many GIS applications. SQL
 235 spatial offers several advantages :

- 236 • the preservation of SQL concepts such as the ability to *Create* a new
 237 entry, as well as *Read*, *Update* and *Delete* existing entries in a data set
 238 containing geometries (CRUD),
- 239 • the incorporation of spatial operations and relationships normalized
 240 by the Open Geospatial Consortium specified in the OpenGIS Simple
 241 Features Specification for SQL [61, 62],
- 242 • a comprehensible and human readable language.

243 The developments of the open-source relational database PostgreSQL⁶
 244 with the spatial extender PostGIS⁷ are a key of this success. PostGIS offers

⁵<http://geoscript.org/> accessed July 2017

⁶<https://www.postgresql.org/> accessed in july 2017

⁷<http://postgis.net/> accessed in july 2017

245 a flexible analytical tool to organize spatial analysis allowing overlay, spa-
 246 tial joining and spatial summaries. Despite the NOSQL trend, the use of
 247 SQL spatial grows in the last years due to the development of new spatial
 248 databases like SpatialLite⁸ or H2GIS⁹ [63]. Therefore, to facilitate the reuse
 249 of the morphological indicators available in the MApUCE processing chain,
 250 each indicator has been described in SQL spatial.

251

252 An illustration with the form factor indicator (FF_{build}), calculated at the
 253 building level is given below.

$$FF_{build} = \frac{S_{build}}{L_{build}^2} \quad (1)$$

254 Where

- 255 - S_{build} is the building area
- 256 - L_{build} is the building length (perimeter)

257

258 Translated into a SQL script, the form factor is computed using two spa-
 259 tial operators "ST_Area" and "ST_Length" and one mathematical function
 260 ("Power") (Table 13).

261

```

1  -- Drop the table if it already exists
2  DROP TABLE IF EXISTS BUILD_FORM_FACTOR;
3  -- Create the table and compute the form factor value
4  CREATE TABLE BUILD_FORM_FACTOR (PK integer primary key,
   FORM_FACTOR double)
5  AS SELECT PK, ST_AREA( THE_GEOM ) / POWER( ST_LENGTH(
   THE_GEOM ), 2 ) AS FORM_FACTOR
6  FROM BUILDINGS;
```

Table 13: SQL script to compute the building form factor

262 This kind of approach allows to describe in a generic way a set of indi-
 263 cators that will be applied in any Relational DataBase Management System
 264 (RDBMS) that supports the SQL spatial standard.

⁸<http://www.gaia-gis.it/gaia-sins/> accessed in july 2017

⁹<http://www.h2gis.org/> accessed in july 2017

265 3.1.2. *R language to build the urban fabric classification*

266 R¹⁰ is one of the most famous statistical analysis tool. Using R provides
267 a broad range of advantages. It incorporates a great number of the standard
268 statistical methods and it is a comprehensive language for managing and
269 manipulating data. The R interpreted language permits to easily and quickly
270 create new computational methods. Moreover, R is driven by an important
271 community that provides an impressive list of packages that do everything:
272 data loading, manipulation, visualization and modelling as well as results
273 reporting in various application fields such as finance, biology or any time
274 series or spatial application, *etc.*

275 To compute the typology of the urban fabric, two R scripts are written.
276 The first one is used to elaborate the decision trees model (Table A.15) and
277 the second one is executed to predict the typology classes of each buildings
278 (Table A.16). The scripts take advantage of the two packages called *random-*
279 *Forest* [64] and *RPostgreSQL* [65].

280 To extract the first and second main type of urban fabric, the result of
281 the typology prediction at building scale is aggregated at RSU level based
282 on the percentage of the floor area.

283 3.2. *The MApUCE tools chain*

284 The MApUCE tools chain implements the methodology and algorithms
285 described previously to compute indicators and urban fabric classification. It
286 is established around the concept of SDI in order to overcome inconsistencies
287 in data structure as well as in data querying and to break the barriers to share
288 and re-use spatial processing or results. The SDI includes several components
289 (Figure 6):

- 290 • a "Web Processing Service"¹¹ system to execute treatments in a doc-
291 umented and standardized way, available as a service using H2GIS,
292 Renjin¹² and managed from the OrbisGIS¹³ GIS platform [66],
- 293 • a spatial database management system to store all data (reference,
294 input and results), using the PostGreSQL and PostGIS applications,

¹⁰<https://www.r-project.org/> accessed in july 2017

¹¹<http://www.opengeospatial.org/standards/wps> accessed in july 2017

¹²<http://www.renjin.org/> accessed in july 2017

¹³<http://orbisgis.org/> accessed in july 2017

- a cartographic server, named GeoServer¹⁴, to publish maps within a standardized image stream, based on the "Web Map Service¹⁵" specification,
- a web cartographic portal to reconstitute in a user friendly way the results of the geoprocessing tools chain.

Figure 6: Components of the MAPUCE SDI tools chain

3.2.1. The Web Processing Service

The OrbisServer is the main piece of the SDI. It has been developed on top of the H2GIS database [63] and the Apache Groovy¹⁶ programming language. OrbisServer implements the version 2.0 of the Web Processing Service standard approved by the Open Geospatial Consortium. The WPS defines a standardized interface to facilitate the publishing of geospatial processes and to discover and execute those processes by a client. A WPS implementation allows to establish geospatial service chains in a distributed way.

¹⁴<http://geoserver.org/> accessed in july 2017

¹⁵<http://www.opengeospatial.org/standards/wms> accessed in july 2017

¹⁶<http://groovy-lang.org/> accessed in july 2017

309 In practice, the SQL and the R scripts used to process the data in the MA-
 310 pUCE framework are exposed as web processes. A web process is described
 311 from a groovy script that contains: a list of input data, a run method to exe-
 312 cute the process and a list of output data. The Table B.17 gives an example
 313 for the building form factor indicator. The script is defined with one input
 314 data corresponding to the buildings table (in the database) and one output
 315 data being a message warning the user that the processing method has been
 316 run.

317 To execute the R scripts, OrbisServer integrates the Renjin engine. Ren-
 318 jin is a JVM-based interpreter dedicated to the R language (for statistical
 319 computing [67]). Aside its capabilities, the biggest advantage of Renjin is
 320 that the R interpreter itself is a Java module which can be seamlessly inte-
 321 grated into any Java application (this is the case for the MApUCE SDI).

322
 323 The WPS scripts are managed from the OrbisGIS user interface. A tool-
 324 box plugin lists the 6 processes offered by the OrbisServer application (de-
 325 scribed in section 4).

326 3.2.2. The spatial database

327 As explained previously, the input data, the morphological indicators and
 328 the urban fabric classification results are stored in a PostgreSQL-PostGIS
 329 database. The database communicates with the OrbisServer to serve data
 330 to the scripts. The results are stored in 6 tables (Figure 7). These tables
 331 are suffixed with the name "METROPOLE" to isolate the geographical data
 332 that cover the french metropolitan territory. Each feature is linked to a
 333 municipality area using a key index.

Figure 7: Data model of tables finally produced

334 3.2.3. *The Cartographic server*

335 The Geoserver application is used to publish the data available in the
336 spatial database. Geoserver is advantageous because it is open source and
337 conform to the main OGC standards such as Web Map Service (WMS) or
338 Web Feature Service (WFS). It is stable since few years and it offers a well
339 suited user interface to control spatial data access. From GeoServer, the
340 MApUCE SDI delivers a set of WMS layers styled with the Style Layer
341 Descriptor¹⁷ (SLD) standard. These layers represent the morphological in-
342 dicators at RSU scale. Their rendering is automatically updated after each
343 change in the database.

344 3.2.4. *The web cartographic portal*

345 The web cartographic portal is based on the javascript framework mviewer¹⁸.
346 mviewer is a responsive template to build simple and elegant web mapping
347 applications organized around one configuration file. The configuration file
348 contains informations:

- 349 • to customize the *look and feel* of the portal,
- 350 • to build a thematic sidebar that lists a set of layers (WMS or geojson
351 file),
- 352 • to add tools on the map such as zooming and distance measurement,
353 map sharing from a permalink, ...

354 4. Results

355 Three types of results have been obtained. The first concerns the indi-
356 cators and the classification produced from the MApUCE data. The second
357 and the third are the description of respectively the MApUCE toolbox and
358 the MApUCE web cartographic portal.

359 4.1. *MApUCE data*

360 Currently, 80 of the main french urban areas have been processed, rep-
361 resenting 2,238 municipalities (6% of the total number in France), 3,726,108
362 buildings (41%) and 181,752 RSU (40%). Those computed areas are repre-
363 sented in Figure 8.

¹⁷<http://www.opengeospatial.org/standards/sld> accessed in july 2017

¹⁸<https://github.com/geobretagne/mviewer> accessed in july 2017

Figure 8: Cartography of the processed french urban areas (red polygons) on top of the OpenStreetMap layer

364 An overview of the results obtained for the three scales is presented Ta-
 365 ble 14. Two types of maps are provided. The first one uses a *unique symbol*
 366 *representation* to display the raw geometries. The second one shows the dis-
 367 tribution of two variables using a *choropleth technique* : the building height
 368 and the compacity.

369
 370 The result of the classification at both building and RSU scales is illus-
 371 trated in Figure 9 : the distribution of urban tissue is shown using a *unique*
 372 *values classification*.

373 The distribution of building height and compacity (Figure 14) as well
 374 as building typology (Figure 9) is not homogeneous. The building scale is
 375 interesting when we focus on a very restricted area but it is inappropriate
 376 for a larger scale analysis. Depending on the scale of interest (neighborhood,
 377 city, conurbation,etc.), the information should be aggregated. The block

Table 14: Maps for the three scales, zoomed in the city center of Nantes

378 and RSU scales are more suited for such scales (Table 14) : some clusters
 379 can be observed at Blocks and RSU scales, which reinforces the interest of
 380 geographic aggregation of data to classify a territory.

Figure 9: Comparison between the satellite image (top left) and the urban fabric typology classification at the building level (top right) and RSU level (bottom right) in the French municipality of Toulouse

381 4.2. The MAPUCE toolbox: An interface to execute the complete chain

382 As described in the section 3 ("Implementation"), a dedicated user in-
 383 terface called MAPUCE toolbox has been developed to execute the complete
 384 chain (Figure 1) through the open-source GIS software OrbisGIS (Figure 10).
 385 It takes advantage of the GIS capabilities to navigate, represent and query
 386 the data. This MAPUCE interface allows non-expert users to execute pro-
 387 cesses and to obtain data depending on their study area (defined by at least
 388 one municipality).

389 This dedicated interface is provided as a free plugin, making available a
 390 set of 6 scripts in the OrbisGIS Toolbox panel (top right red rectangle in
 391 Figure 10 and zoomed in Figure 11). For each script, a user interface is
 392 generated on the fly, offering to the user to choose some options and to set
 393 parameter values for the computation.

394 Two needs are answered by the plugin: either to get the final and the
 395 intermediate results using the step-by-step processing scripts, either to obtain
 396 only the final results (indicators and classification).

Figure 10: OrbisGIS UI

Figure 11: Zoom on the OrbisGISs Toolbox and the MAPuCE plugin

397 The performance of the algorithms have been evaluated for a middle size
 398 French city called Vannes. Its overall footprint is $32,30 \text{ km}^2$, it includes
 399 16'448 buildings, 775 RSU and 7'043 roads. To compute all the indicators,
 400 the hardware¹⁹ and software²⁰ used needed 130s whereas it tooked 40 s for
 401 the classification. The entire process is then run in less than three minutes.

¹⁹Processor : Intel W3520 - Cores/Threads : 4c/8t - Frequency : 2.66GHz / 2.93GHz -
 RAM : 32GB DDR3 1333 MHz - Disks : SoftRaid 2x2TB SATA

²⁰OS : Ubuntu server 14.10 - JAVA 8

402 4.2.1. Step-by-step processing

403 **1- Import commune areas**

404

405 This script has to be executed at first to return the list of available mu-
406 nicipalities which are ready to be processed. The user is invited to enter his
407 login and password²¹ and then to press the green arrow icon to execute the
408 script (Figure 12).

409

Figure 12: User Interface of the script called "1-Import commune areas"

410 **2- Import the USR (french translation for RSU), buildings and**
411 **roads**

412

413 The script number 2 will import all required data, related to the munic-
414 ipalities that users are invited to select in a dropdown list. The selection is
415 made through the "INSEE Code" which is the french unique identifier for
416 municipalities (Figure 13).

²¹For security reasons the remote database can only be accessed through personal accounts.

Figure 13: User Interface of the script called "2-Import the USR, buildings and roads"

3- Compute morphological indicators

Once data are imported, this script will automatically compute all the indicators. The user has nothing to do except pressing the "execute" button.

4- Classify the BUILDING and USR (RSU) features

Based on computed indicators, this script will perform the random forest classification. The user has nothing to do except pressing the "execute" button.

4.2.2. Direct final results obtention

It is possible to obtain directly the final tables (indicators and classification) using the two following scripts.

432 **Import computed indicators**

433

434 This script is used to download data that have already been computed
435 on the server side and are thus available in the spatial database. The user
436 must fill his login and password, choose the spatial unit scale (*commune*
437 (`CODE_INSEE`) or *urban area*) and select the corresponding identifiers.

438

439 **Complete geoprocessing chain**

440

441 This script is used to run the complete geoprocessing chain in a single step
442 (*i.e.* import the data, process the indicators and classify the urban fabric) in
443 the case where the municipalities have not been yet processed on the server
444 side. The user must fill his login and password, choose the spatial unit scale
445 ("*municipality*" (`CODE_INSEE`) or "*urban area*" (`UNITE_URBAINE`)) and
446 select the corresponding identifiers (Figure 14).

447 **4.3. The MApUCE web cartographic portal**

448 The result are available accessing the mapuce.orbisgis.org web carto-
449 graphic portal. People can navigate into the map and choose to display
450 a set of layers, grouped into three categories:

- 451 • spatial units : urban area, already computed municipalities and RSU
452 boundaries,
- 453 • RSU indicators : thematic analysis based on several morphological
454 indicators,
- 455 • typology : building classification at the RSU scale.

456 To comply with data licences, only RSU results are presented. Two
457 screenshots are presented in Figure 15 to illustrate the type of maps that
458 can be consulted by users.

459 **Conclusion & Prospect**

460 We have proposed an open geoprocessing framework to calculate stan-
461 dardized urban indicators useful for urban climate application and also for
462 planning purpose in some other fields.

463

Choose one (or many) municipality(ies)

Choose one (or many) urban area(s)

Figure 14: User interface for the script "Complete geoprocessing chain"

464 Morphological indicators have been computed at three different scales :
 465 building, block and the Reference Spatial Unit (RSU). The boundaries of
 466 the RSU are generated by a Voronoï tessellation from the legal boundaries
 467 specified in the cadastral map. This scale is appropriate to analyze the
 468 climate property heterogeneity of the urban fabric within an urban area. Such
 469 map can be used directly by the urban planners for planification purpose,
 470 by researchers to simulate the urban climate or by researchers to highlight
 471 the differences of development typology between several cities. However,
 472 geometric issues have been observed in certain RSU: some of them are too
 473 small (Figure 16 left), have weird shape (Figure 16 right) or they separate
 474 buildings in two parts. Investigations should be realized to overcome this

Figure 15: Screenshots of the mapuce.orbisgis.org portal, zoomed on the french city of Vannes (left: thematic analysis of the building mean compacity at the RSU scale / right: building classification by RSU)

475 issue, for example using the road network to slice the territory.

Figure 16: Two examples of geometric issues on RSU: too small (left), weird shape (right)

476 Indicator calculations are based on geographical databases which are
 477 available and homogeneous for the french territory. Preprocessing tasks have
 478 been performed by [23] to clean and structure those data, to create the RSU
 479 and to enrich building and RSU tables by database cross-feeding. Because

the french databases used are updated every year, the preprocessing task as well as the indicators calculation may be applied on annual data in order to make diachronic analysis. The use of open data such as OpenStreetMap²² may also be investigated to generate worldwide homogeneous information.

27 morphological indicators have been calculated at building scale, 9 at block scale and 28 for the RSU. These 64 indicators are finally affected to each building and are used (together with socioeconomic indicators calculated by Plumejeaud-Perreau et al. [23]) to classify buildings according to 10 typological classes defined by urbanists, architects and using technical literature. The supervised classification method used has a prediction error of about 11% [24]. The dominant building typological class within a RSU is selected to characterize the RSU scale. Indicator calculation and classification application may now be extended to the entire french territory. However, the list of the calculated indicators is not irrevocable. An overall reflection may be performed to both identify existing indicators that are redundant or not relevant and new indicators to improve the classification process.

The overall processing chain is uniquely composed of open-source tools and close to open standards: the OrbisGIS platform is used for morphological indicator calculation and Renjin is used for building and RSU classification. A free OrbisGIS plug-in called MAPuCE is available for any user interested in applying the processing chain or to analyze the results obtained for the municipality of its choice. It is also possible to access the results directly on the internet from a web portal dedicated to this work. Future work implies to give the opportunity to any user to produce the indicators through a full distributed service.

Finally, the results of this paper offers new opportunities to extend the WUDAPT database at a finest scale. Indeed, one idea would be to reuse the entire processing chain (territory segmentation, indicator calculation, classification model creation and application) but using as training data the LCZ typologies instead of the one presented in this article.

²²<https://www.openstreetmap.org/> accessed in july 2017

513 Acknowledgements

514 The research and developments presented in this paper have been made
515 within the MApUCE project, funded by the French National Research Agency
516 (ANR).

517 References

- 518 [1] U. Nations, World Urbanization Prospects: The 2014 Revision, 2015.
519 Available at <https://esa.un.org/unpd/wup/>.
- 520 [2] G. Levermore, J. Parkinson, K. Lee, P. Laycock, S. Lindley, The in-
521 creasing trend of the urban heat island intensity, *Urban Climate* (2017).
- 522 [3] J. He, J. Liu, D. Zhuang, W. Zhang, M. Liu, Assessing the effect of
523 land use/land cover change on the change of urban heat island intensity,
524 *Theoretical and Applied Climatology* 90 (2007) 217–226.
- 525 [4] L. Chen, E. Ng, X. An, C. Ren, M. Lee, U. Wang, Z. He, Sky view factor
526 analysis of street canyons and its implications for daytime intra-urban
527 air temperature differentials in high-rise, high-density urban areas of
528 Hong Kong: a GIS-based simulation approach, *International Journal of*
529 *Climatology* 32 (2012) 121–136.
- 530 [5] T. R. Oke, City size and the urban heat island, *Atmospheric Environ-*
531 *ment* 7 (1973) 769–779.
- 532 [6] H.-S. Park, Features of the heat island in Seoul and its surrounding
533 cities, *Atmospheric Environment* 20 (1986) 1859–1866.
- 534 [7] K. Laaidi, A. Zeghnoun, B. Dousset, P. Bretin, S. Vandentorren, E. Gi-
535 raudet, P. Beaudeau, The impact of heat islands on mortality in Paris
536 during the August 2003 heat wave, *Environmental health perspectives*
537 120 (2012) 254.
- 538 [8] S. Conti, P. Meli, G. Minelli, R. Solimini, V. Toccaceli, M. Vichi, C. Bel-
539 trano, L. Perini, Epidemiologic study of mortality during the summer
540 2003 heat wave in Italy, *Environmental research* 98 (2005) 390–399.
- 541 [9] Y. Hirano, Y. Yoshida, Assessing the effects of CO₂ reduction strategies
542 on heat islands in urban areas, *Sustainable Cities and Society* 26 (2016)
543 383–392.

- 544 [10] C. de Munck, G. Pigeon, V. Masson, F. Meunier, P. Bousquet,
545 B. Tréméac, M. Merchat, P. Poeuf, C. Marchadier, How much can
546 air conditioning increase air temperatures for a city like Paris, France?,
547 International Journal of Climatology 33 (2013) 210–227.
- 548 [11] M. Santamouris, L. Ding, F. Fiorito, P. Oldfield, P. Osmond, R. Paolini,
549 D. Prasad, A. Synnefa, Passive and active cooling for the outdoor built
550 environment - Analysis and assessment of the cooling potential of miti-
551 gation technologies using performance data from 220 large scale projects,
552 Solar Energy 154 (2017) 14 – 33.
- 553 [12] Y. Kikegawa, Y. Genchi, H. Kondo, K. Hanaki, Impacts of city-block-
554 scale countermeasures against urban heat-island phenomena upon a
555 buildings energy-consumption for air-conditioning, Applied Energy 83
556 (2006) 649–668.
- 557 [13] C. Grimmond, M. Blackett, M. Best, J. Barlow, J. Baik, S. Belcher,
558 S. Bohnenstengel, I. Calmet, F. Chen, A. Dandou, et al., The interna-
559 tional urban energy balance models comparison project: first results
560 from phase 1, Journal of Applied Meteorology and Climatology 49
561 (2010) 1268–1292.
- 562 [14] I. D. Stewart, Redefining the urban heat island, Ph.D. thesis, University
563 of British Columbia, 2011.
- 564 [15] I. D. Stewart, T. R. Oke, Local Climate Zones for urban temperature
565 studies, Bulletin of the American Meteorological Society 93 (2012) 1879–
566 1900.
- 567 [16] F. Leconte, J. Bouyer, R. Claverie, M. Pétrissans, Using local climate
568 zone scheme for uhi assessment: Evaluation of the method using mobile
569 measurements, Building and Environment 83 (2015) 39–49.
- 570 [17] E. Lelovics, J. Unger, T. Gál, Design of an urban monitoring network
571 based on local climate zone mapping and temperature pattern mod-
572 elling, Climate research 60 (2014) 51–62.
- 573 [18] Y. Zheng, C. Ren, Y. Xu, R. Wang, J. Ho, K. Lau, E. Ng, Gis-based
574 mapping of local climate zone in the high-density city of hong kong,
575 Urban Climate (2017).

- 576 [19] N. Tornay, R. Schoetter, M. Bonhomme, S. Faraut, V. Masson, Genius:
577 A methodology to define a detailed description of buildings for urban
578 climate and building energy consumption simulations, *Urban Climate*
579 20 (2017) 75–93.
- 580 [20] G. Mills, J. Ching, L. See, B. Bechtel, M. Foley, An introduction to the
581 WUDAPT project, in: 9th International Conference on Urban Climate
582 (ICUC9), Météo-France, Toulouse, France.
- 583 [21] B. Bechtel, M. Foley, G. Mills, J. Ching, L. See, P. Alexander, M. OCon-
584 nor, T. Albuquerque, M. de Fatima Andrade, M. Brovelli, et al., Census
585 of cities: LCZ classification of cities (Level 0)– Workflow and initial re-
586 sults from various cities, in: 9th International Conference on Urban
587 Climate (ICUC9), Météo-France, Toulouse, France.
- 588 [22] C. Böhringer, P. E. Jochem, Measuring the immeasurable - A survey of
589 sustainability indices, *Ecological Economics* 63 (2007) 1–8.
- 590 [23] C. Plumejeaud-Perreau, C. Poitevin, C. Pignon-Mussaud, N. Long,
591 Building Local Climate Zones by using socio-economic and topographic
592 vectorial databases, in: 9th International Conference on Urban Climate
593 (ICUC9), Météo-France, Toulouse, France.
- 594 [24] S. Faraut, M. Bonhomme, N. Tornay, A. Amossé, V. Masson, E. Bocher,
595 G. Petit, C. Plumejeaud, N. Long, G. Bretagne, R. Schoetter, Des
596 bases de données urbaines aux simulations énergétiques le projet
597 MApUCE, in: Séminaire de conception architecturale numérique
598 : Mètre et paramètre, mesure et démesure du projet, volume 1
599 of *SCAN'16 Toulouse*, École Nationale Supérieure d'Architecture de
600 Toulouse, Presses Universitaires de Nancy, Éditions Universitaires de
601 Lorraine, Toulouse, France, 2016.
- 602 [25] V. Masson, J. Hidalgo, A. Amossé, E. Bocher, M. Bonhomme, A. Bour-
603 geois, G. Bretagne, S. Caillerez, E. Cordeau, C. Demazeux, S. Faraut,
604 C. Gallato, S. Haoues-Jouve, M.-L. Lambert, A. Lemonsu, J.-P. Lévy,
605 N. Long, C.-X. Lopez, G. Petit, M. Pellegrino, C. Pignon-Mussaud, C. J.
606 Plumejeaud, V. Ruff, R. Schoetter, N. Tornay, D. D. Vye, Urban Cli-
607 mate, Human behavior & Energy consumption : from LCZ mapping to
608 simulation and urban planning(the MApUCE project), in: B. Beckers,

- 609 M. A. T. Pico, S. Jimenez (Eds.), First International Conference on Ur-
 610 ban Physics, volume 1 of *FICUP 2016 First International Conference*
 611 *on Urban Physics*, Pontifical Catholic University of Ecuador, UNDP
 612 Ecuador, Quito, Ecuador, 2016, pp. 155–167.
- 613 [26] V. Oliveira, The elements of urban form, in: *Urban Morphology*,
 614 Springer, 2016, pp. 7–30.
- 615 [27] S. Steiniger, T. Lange, D. Burghardt, R. Weibel, An approach for the
 616 classification of urban building structures based on discriminant analysis
 617 techniques, *Transactions in GIS* 12 (2008) 31–59.
- 618 [28] T. Oke, *Boundary layer climates*. 2nd, Methuen, 289p (1987).
- 619 [29] I. Berghauser-Pont, P. Haupt, The spacemate: Density and the Typo-
 620 morphology of the Urban Fabric, *Nordisk Arkitekturforskning (Nordic*
 621 *Journal of Architectural Research)* 4 (2005) 55–68.
- 622 [30] J. Bouyer, C. Inard, M. Musy, Microclimatic coupling as a solution to
 623 improve building energy simulation in an urban context, *Energy and*
 624 *Buildings* 43 (2011) 1549 – 1559.
- 625 [31] J. Lesbegueries, N. Lachiche, A. Braud, G. Skupinski, A. Puissant,
 626 J. Perret, A platform for spatial data labeling in an urban context,
 627 *Geospatial free and open source software in the 21st century* (2012)
 628 49–61.
- 629 [32] R. E. Horton, Drainage-basin characteristics, *Eos, Transactions Amer-*
 630 *ican Geophysical Union* 13 (1932) 350–361.
- 631 [33] L. Adolphe, A simplified model of urban morphology: Application to an
 632 analysis of the environmental performance of cities, *Environment and*
 633 *Planning B: Planning and Design* 28 (2001) 183–200.
- 634 [34] A. P. d’URbanisme (APUR), *Consommations d’énergie et émissions de*
 635 *gaz à effet de serre liées au chauffage des résidences principales parisi-*
 636 *ennes*, Technical Report, Atelier Parisien d’URbanisme (APUR), 2007.
- 637 [35] N. Gauthier, *Analyses morphologiques de formes urbaines et étude de*
 638 *l’impact des formes urbaines sur les gains énergétiques solaires.*, Ph.D.
 639 thesis, INSA de Strasbourg, 2014.

- 640 [36] W. E. Dramstad, Spatial metrics—useful indicators for society or
641 mainly fun tools for landscape ecologists?, *Norsk Geografisk Tidsskrift-*
642 *Norwegian Journal of Geography* 63 (2009) 246–254.
- 643 [37] H. Gravelius, *Grundriß der gesamten Gewässerkunde: in vier Bänden,*
644 *vol.1, Göschen, 1914.*
- 645 [38] D. Rainsford, W. Mackaness, *Template Matching in Support of Gener-*
646 *alisation of Rural Buildings*, Springer, Berlin, Heidelberg, pp. 137–151.
- 647 [39] C. Duchêne, S. Bard, X. Barillot, A. Ruas, J. Trévisan, F. Holzapfel,
648 Quantitative and qualitative description of building orientation, in: 6th
649 ICA Workshop on Generalisation and Multiple Representation, 28-30
650 April, Paris (France).
- 651 [40] S. Salat, *Les villes et les formes: sur l’urbanisme durable*, Hermann
652 Editeurs, 2011.
- 653 [41] C. Ratti, N. Baker, K. Steemers, Energy consumption and urban tex-
654 ture, *Energy and Buildings* 37 (2005) 762 – 776.
- 655 [42] M. Herold, J. Scepan, K. C. Clarke, The use of remote sensing and
656 landscape metrics to describe structures and changes in urban land uses,
657 *Environment and Planning A* 34 (2002) 1443–1458.
- 658 [43] K. McGarigal, B. J. Marks, *Fragstats: spatial pattern analysis program*
659 *for quantifying landscape structure.*, Gen. Tech. Rep. PNW-GTR-351.
660 Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific
661 Northwest Research Station. 122 p. (1995).
- 662 [44] N. Colaninno, J. R. Cladera, K. Pfeffer, An automatic classification of
663 urban texture: form and compactness of morphological homogeneous
664 structures in Barcelona, in: 51st European Congress of the Regional
665 Science Association International, pp. 1–20.
- 666 [45] E. R. Alexander, Density measures: A review and analysis, *Journal of*
667 *Architectural and Planning Research* 10 (1993) 181–202.
- 668 [46] M. Berghauser-Pont, P. Haupt, *Space, density and urban form*, Ph.D.
669 thesis, TU Delft, Delft University of Technology, 2009.

- [47] C. S.-L. Chan, Measuring Physical Density: Implications on the Use of Different Measures on Land Use Policy in Singapore, Ph.D. thesis, Massachusetts Institute of Technology, Department of Urban Studies and Planning, 1999.
- [48] N. Long, C. Kergomard, Classification morphologique du tissu urbain pour des applications climatologiques. Cas de Marseille, *Revue Internationale de Géomatique* 15 (2006) 487–512.
- [49] J. Tratalos, R. A. Fuller, P. H. Warren, R. G. Davies, K. J. Gaston, Urban form, biodiversity potential and ecosystem services, *Landscape and Urban Planning* 83 (2007) 308–317.
- [50] N. Schwarz, Urban form revisited-selecting indicators for characterising european cities, *Landscape and Urban Planning* 96 (2010) 29–47.
- [51] N. Tornay, M. Bonhomme, S. Faraut, GENIUS, a methodology to integer building scale data into urban microclimate and energy consumption modelling, in: 9th International Conference on Urban Climate (ICUC9), Météo-France, Toulouse, France.
- [52] N. Tornay, R. Schoetter, M. Bonhomme, S. Faraut, V. Masson, GENIUS: A methodology to define a detailed description of buildings for urban climate and building energy consumption simulations, *Urban Climate* 20 (2017) 75 – 93.
- [53] S. Steiniger, A. J. S. Hunter, Free and Open Source GIS Software for Building a Spatial Data Infrastructure, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 247–261.
- [54] G. S. D. I. Association, et al., The Spatial Data Infrastructure Cookbook, GSDI/Nebert (2009). Available at <http://www.gsdi.org/gsdicookbookindex>.
- [55] M. A. Parsons, R. Duerr, J.-B. Minster, Data Citation and Peer Review, *Eos, Transactions American Geophysical Union* 91 (2010) 297.
- [56] S. Steiniger, E. Bocher, An overview on current free and open source desktop GIS developments, *International Journal of Geographical Information Science* 23 (2009) 1345–1370.

- 701 [57] E. Bocher, T. Leduc, G. Moreau, F. G. Cortès, GDMS: an abstraction layer to enhance spatial data infrastructures usability, in: 11th
702 AGILE International Conference on Geographic Information Science-
703 AGILE'2008.
704
- 705 [58] T. Leduc, E. Bocher, F. G. Cortès, G. Moreau, GDMS-R: A mixed
706 SQL to manage raster and vector data, in: J. Horák, L. Halounová,
707 D. Kusendová, P. Rapant, V. Vozenílek (Eds.), Advances in Geoinformation Technologies, VŠB - Technical University of Ostrava, 2009, pp.
708 43–56.
709
- 710 [59] C. D. Tomlin, Map algebra: one perspective, Landscape and Urban
711 Planning 30 (1994) 3–12.
- 712 [60] R. Battle, D. Kolas, Geosparql: enabling a geospatial semantic web,
713 Semantic Web Journal 3 (2011) 355–370.
- 714 [61] J. Herring, Implementation specification for geographic information-
715 simple feature access-part 1: Common architecture, Open Geospatial
716 Consortium Inc 95 (2006).
- 717 [62] J. R. Herring, Opendgis implementation specification for geographic
718 information-simple feature access-part 2: Sql option, Open Geospatial
719 Consortium Inc (2006).
- 720 [63] E. Bocher, G. Petit, N. Fortin, S. Palominos, H2GIS a spatial database
721 to feed urban climate issues, in: 9th International Conference on Urban
722 Climate (ICUC9), Météo-France, Toulouse, France.
- 723 [64] A. Liaw, M. Wiener, Breiman and Cutler's Random Forests for Classification and Regression, Technical Report, University of California, Berkeley, 2015. Version : 4.6-12, Available at <https://www.stat.berkeley.edu/~breiman/RandomForests/>.
724
725
726
- 727 [65] J. Conway, D. Eddelbuettel, T. Nishiyama, S. K. Prayaga, N. Tiffin, R
728 Interface to the 'PostgreSQL' Database System, Technical Report, PostgreSQL Global Development Group and The Regents of the University of California, 2017.
729
730
- 731 [66] E. Bocher, G. Petit, OrbisGIS: Geographical Information System Designed by and for Research, John Wiley & Sons, Inc., pp. 23–66.
732

733 [67] M.-J. Kallen, H. Mhleisen, Latest developments around Renjin,
734 in: R Summit & Workshop, Centrum Wiskunde & Informatica,
735 Copenhagen, Denmark. Available at [http://www.renjin.org/blog/](http://www.renjin.org/blog/2015-06-28-renjin-at-rsummit-2015.html)
736 2015-06-28-renjin-at-rsummit-2015.html.

737 **Appendix A. R scripts**

```

1 ## Load packages
2 library(randomForest)
3 library(RPostgreSQL)
4 ## Import the training data from the database
5 ## con object is the connection to the database
6 training_data= dbGetQuery(con, "SELECT * FROM building_
   training")
7 ## Build model
8 treesModel=randomForest(i_typo~.,data=training_data,
   ntree=500,mtry=7,replace=TRUE)
9 ## Save the model
10 save(treesModel,file="mapuce_model.RData")

```

Table A.15: Pseudo-R script to create the decision trees model

```

1 ## Load packages
2 library(randomForest)
3 library(RPostgreSQL)
4 ### Load the model based on the morphological train data
5 treesModel=get(load(model_path))
6 ## Get the data to predict from the database
7 ## The buildings_to_predict is a temporary table created
   on the fly with a SQL command. It contains all the
   indicators at building, block and RSU scales
8 data_to_predict = dbGetQuery(con, "SELECT * FROM
   buildings_to_predict")
9 ## Apply the predict function to compute the typological
   class for each building
10 typology=predict(treesModel,data_to_predict,type="class"
   )

```

Table A.16: Pseudo-R script to predict the urban fabric typological class

738 Appendix B. WPS Script

```

1  /** String input of the process */
2  @LiteralDataInput(
3      title="Buildings▯table",
4      description="Name▯of▯the▯buildings▯table")
5  String buildingsTable
6
7  /** SQL code to execute with some metadata */
8  @Process(title = "Building▯form▯factor",
9      description = "Compute▯the▯building▯form▯factor"
10     )
11  def processing() {
12      -- Drop the table if it already exists
13      DROP TABLE IF EXISTS BUILD_FORM_FACTOR;
14      -- Create the table and compute the form factor value
15      CREATE TABLE BUILD_FORM_FACTOR (PK integer primary key,
16          FORM_FACTOR double)
17      AS SELECT PK, ST_AREA( THE_GEOM ) / POWER( ST_LENGTH(
18          THE_GEOM ), 2 ) AS FORM_FACTOR)
19      FROM $buildingsTable; //The input table name
20
21  literalOutput =    The form factor indicator has been
22                    computed }
23
24  /** String output of the process */
25  @LiteralDataOutput(
26      title="Output▯message",
27      description="The▯output▯message")
28  String literalOutput

```

Table B.17: Example of a WPS script