A generic algorithm to automatically classify urban fabric according to the Local Climate Zone system: implementation in GeoClimate 0.0.1 and application to French cities

Jérémy Bernard[1,2,3], Erwan Bocher[3], Matthieu Gousseff[3,4], François Leconte[5], and Elisabeth Le Saux Wiederhold[4]

[1]University of Gothenburg, Department of Earth Sciences, Sweden
[2]University of Savoie Mont-Blanc, LOCIE, UMR 5271, France
[3]CNRS, Lab-STICC, UMR 6285, Vannes, France
[4]Université Bretagne Sud, Lab-STICC, UMR 6285, Vannes, France
[5]Université de Lorraine, INRAE, LERMaB, F88000, Epinal, France

Correspondence: Jérémy Bernard (jeremy.bernard@zaclys.net)

Abstract. Geographical features may have a considerable effect on local climate. The Local Climate Zone (LCZ) system proposed by Stewart and Oke (2012) is nowadays seen as a standard ~~referential~~ approach to classify any zone according to a set of urban canopy parameters. While many methods already exist to map the LCZ, only few tools are openly and freely available. This manuscript presents the algorithm implemented in the GeoClimate software to identify the LCZ of any place in the world based on vector data. Seven types of information are needed as input: building footprint, road and rail networks, water, vegetation and impervious surfaces. First the territory is partitioned into Reference Spatial Units (RSU) using the road and rail network as well as the boundaries of large vegetation and water patches. Then 14 urban canopy parameters are calculated for each RSU. Their values are used to classify each unit to a given LCZ type according to a set of rules. GeoClimate can automatically prepare the inputs and calculate the LCZ for two datasets: OpenStreetMap (OSM - available worldwide) and the BD Topo v2.2 (BDT - a French dataset produced by the national mapping agency). The LCZ are calculated for 22 French communes using these two datasets in order to evaluate the effect of the dataset on the results. About 55% of all areas has obtained the same LCZ type with large differences when differentiating this result by city (from 30% to 82%). The agreement is good for large patches of forest and water as well as for *compact mid-rise* and *open low-rise* LCZ types. It is lower for *open mid-rise*, *open high-rise* mainly due to height underestimation for OSM buildings located in open areas. By its simplicity of use, Geoclimate has a great potential for new collaboration in the LCZ field. The software (and its source code) used to produce the LCZ data is freely available at https://zenodo.org/record/6372337, the scripts and data used for the purpose of this manuscript can be freely accessed at https://zenodo.org/record/7687911 and are based on the R package available at https://zenodo.org/record/7646866.

1 Introduction

In its sixth assessment report, the Intergovernmental Panel on Climate Change (IPCC) underlines that cities demonstrate a two ways interaction with the climate system (IPCC et al., 2007). While they impact the climate locally (modification of the energy and mass balances) and globally (GHG emissions), urban areas are also vulnerable to meteorological hazards (Baklanov et al., 2020). Cities are very likely to face extreme climate events such as heatwaves more frequently in the coming decades (IPCC et al., 2007). The United Nations (UN) has identified urban resilience as a key challenge via its Sustainable Development Goals (SDG) (i.e. SDG11 Sustainable Cities and Communities) (Grimmond et al., 2020). Climate change attenuation and adaptation strategies are currently designed and implemented in cities. The efficiency of these strategies relies on our knowledge of the urban environment and our understanding of the urban climate.

The description of the urban fabric is essential for urban climate research, both for observation and modeling purpose. Regarding observation, measurements of physical variables (such as air temperature, surface temperature, relative humidity) are analyzed relatively to local scale urban features (e.g. mean size of street canyons) and micro scale urban features (e.g. distance between the measurement point and a given wall) while models require information about building morphology, land use, materials and anthropogenic fluxes as input data. However, urban data collection has been identified as a challenging task (Masson et al., 2020).

In this context, many classification systems have been promoted to standardize the study of urban climate. In the last decade, the Local Climate Zone (LCZ) classification (Stewart and Oke, 2012) has encountered a growing interest amongst urban climate researchers. A Local Climate Zone is an area that demonstrates particular urban characteristics in terms of morphology, land use, materials, anthropogenic heat release leading to a distinct thermal behavior under given weather conditions. Its size is approximately 400 m - 1000 m wide. The LCZ classification system is organized into 17 LCZ types (ten urbanized, seven non-urbanized). It requires the calculation of ten Urban Canopy Parameters (UCPs), namely sky view factor, aspect ratio, height of roughness element, terrain roughness class, surface fractions (built, impervious, pervious), surface admittance, surface albedo, anthropogenic heat output. Each LCZ type is associated with particular values of these ten UCPs (e.g. LCZ 1 called *compact high-rise* has sky view factor between 0.2 and 0.4, aspect ratio higher than 2, height or roughness class equal to 8, etc.).

The construction of LCZ maps for an area of interest is very time consuming if not automated. In their review, Quan and Bansal have identified two main research streams to build LCZ maps automatically (Quan and Bansal, 2021). A first stream uses remote sensing images as main input data. A prominent initiative of this remote sensing -approach is the World Urban Database and Access Portal Tool (WUDAPT) (Ching et al., 2018). Within this project, a LCZ generator tool has been released (Demuzere et al., 2021). It is based on a Random Forest model trained with areas that have been classified by ~~expert~~ experts to a given LCZ type. The corresponding workflow has been applied to several continents (Demuzere et al., 2019, 2020). A second stream involves detailed geographical information (often presented as vector data) processed by Geographic Information System (GIS). It is organized into six main steps (Quan and Bansal, 2021):

1. Collection of geographical data

2. Partitioning of the territory using Reference Spatial Unit (RSU - called basic spatial units by (Quan and Bansal, 2021)), defined as the smallest spatial unit where calculations are performed

3. Calculation of several UCPs within each RSU

4. Assignation of a LCZ type to each RSU based on UCP values

5. Post-processing, e.g. merge adjacent RSU for simplification and sizing purpose

6. Evaluation of the LCZ map

In the first stream, experts may not use the UCP values to attribute a LCZ type to a given area, thus leading to subjective decision while an objective LCZ referential has been proposed. However, due to the growing accessibility to high quality satellite images, it is applicable to most countries in the world. The second stream is spatially limited to the territory where a specific geographical dataset is available (often a city or a country). Its main advantage is that the UCP used in the LCZ referential are often calculated, inducing a potentially more objective method than the first stream. Another advantage is that geographical information can be crowd-sourced, which is probably less energy greedy than the use of, for instance, aerial photography.

For each of the six steps of this second stream, (Quan and Bansal, 2021) have observed a great diversity of methods. In step 1, the data collection concerns mostly vector data such as building footprint, building height or road cover. However, to determine surface fractions, satellite or airborne images are often used. The definition of the RSU (step 2) varies significantly amongst studies. It can be estimated by local knowledge (Leconte et al, 2015). It can also corresponds to lot area polygons (i.e. influential area surrounding each building) (Skarbit et al., 2017), urban ~~bloc~~ block (i.e. urban unit naturally bounded by streets) (Quan et al., 2017) or regular grid (Geletič et al., 2016). The vast majority of the methods does not calculate all ten UCPs included in the LCZ framework (step 3). The most calculated UCPs are the surface fractions and the height of roughness elements, followed by the sky view factor and the aspect ratio. However, additional UCPs are often proposed, such as building density, population density, areal number density, and additional surface fractions. Concerning step 4, previous studies adopted mainly three workflow types to assign a LCZ type to a given RSU: the standard, the modified and the fuzzy rule-based approaches. The standard rule-based approach associates a RSU with a given LCZ type only if all UCP values fall within the UCPs value ranges of the LCZ type (the other RSUs are set as unclassified). The modified standard rule-based approach uses often less than ten UCPs, adds new UCPs and proposes new rules for the LCZ type assignment (such as decision trees). The fuzzy rule-based approach calculates a degree of membership based on UCP values. For a given RSU, it selects the LCZ type with the highest degree of membership. Step 5 consists in merging RSU to simplify overcomplicated LCZ maps and to meet the size requirement of the LCZ scheme (LCZ larger than 400 meters). This stage also aims to smooth LCZ boundaries and partially reduce the amount of unclassified areas. Finally, LCZ maps are sometimes evaluated against expert knowledge, temperature measurements or other LCZ maps provided by remote sensing methods (step 6).

LCZ maps ~~building~~ workflows usually face two key issues, namely a lack of input data and a partially described methodology (Quan and Bansal, 2021). This study presents a new automated workflow for LCZ maps construction that address these issues by the following contributions:

- the workflow has been designed to be generic: it accepts all datasets as soon as the input data is structured following a well-described guideline (Bocher et al., 2021) and can be run with any type of RSU. The GeoClimate workflow is already designed to work with two input data types, namely BD Topo 2.2 and OpenStreetMap. The first is a French government dataset while the latter is an open source project that provides geographical information worldwide, and therefore tackle the data scarcity issue.
- the workflow, described in details within this article, is integrated in GeoClimate, a free and open source software and thus the code is fully open and available online[1] (Bocher et al., 2021).

This article presents the GeoClimate methodology to produce LCZ maps. Additionally, it is applied to 22 French cities using the worldwide available OpenStreetMap data and compared to the results of using the same method with the French reference dataset BD Topo 2.2. This comparison is useful to observe what are the respective advantages and shortcomings of each dataset and how they could be combined together for application to the French territory.

2 Method and data

2.1 Geoclimate library

GeoClimate is a free and open source toolbox developed in Groovy language that allows ~~to compute geographical indicators~~ geographical indicators calculation from vector-based data. It consists ~~in~~ of a preprocessing workflow where input data are processed to following a generic data structure according to well-described guidelines (Bocher et al., 2021) and in a processing workflow where all indicators are calculated based on this generic input data.

The preprocessing workflow of the GeoClimate version (0.0.1) currently supports two data source providers: the BDTopo database (version 2.2, later referred as BDT) produced by the French national mapping agency (IGNF[2]) and the community database OpenStreetMap (later referred as OSM[3]). Once the input data source is selected, GeoClimate applies a set of rules to extract and format the required spatial descriptors and then build seven GIS layers: building footprint, road and rail network, water, vegetation and impervious surfaces. GeoClimate performs all analysis on an extended zone (1,000 m larger than the original study area) and then the results are cropped to the initial zone of interest (limiting the edge effects at the boundary of the study area). Each layer follows a set of specifications to ensure the completeness and logical consistency of the input data and avoid potential geometry inconsistencies. For examples, only a single geometry is allowed to describe a feature (mutipolygon or multipolyline are exploded), the values used to qualify a building type or a vegetation type are restricted to

[1] https://github.com/orbisgis/geoclimate/wiki (last access: 7 February 2023)
[2] https://geoservices.ign.fr/ (last access: 7 February 2023)
[3] https://www.openstreetmap.org (last access: 7 February 2023)

the ones available in a dictionary provided by GeoClimate, some numeric attributes such as the width of a road are bounded by extreme values. Any datasource can then be connected to the GeoClimate LCZ processing chain (and thus for any place in the world), as long as the connection to the GeoClimate input data model is performed respecting the set of specifications described in Bocher et al. (2021).

The next step concerns the construction of two new spatial units : ~~block~~ building blocks and RSU.

1. a building block is defined as an aggregation of buildings that are in contact,

2. a RSU (Reference Spatial Unit) is generated according to several geographic features which could have an impact on environmental and climate effects and which structure the study area: the road and rail networks as well as the vegetation and water surfaces. The construction of the RSU is a key process in GeoClimate. First, a planar graph is built using all input geometries. The planar graph is then traversed to generate new polygons. Only two dimensional elements are considered for partitioning, therefore underground elements (such as tunnels), or overground (such as bridges) are excluded from the input. Water and vegetation surfaces are also excluded from the input data when they are smaller than a given threshold, set by default to 2'500 m^2 for water and 10'000 m^2 for vegetation. This behavior is visible on Fig. 1: on the Northern part of the river, many small vegetation patches are not considered for RSU creation while they are when they get bigger than 10'000 m^2 (along the river on its Southern part or on the West part of the area)

Figure 1. Illustration of the method to produce the Reference Spatial Units using BDT data

The geographical indicators are then computed using the seven GIS layers to characterize geometric properties and location of spatial features regarding three scales : building, building block and RSU. At building scale, GeoClimate measures the distance to nearest road, the number of building neighbors, area and shape indices (concavity, compactness). At building block scale, the volume, the main orientation, the total areas of the courtyard are computed while at RSU scale, the building and

building block indicators are aggregated (average number of levels per building, density of building floor areas) plus land type fractions and specific climate-oriented indicators as aspect ratio, mean sky view factor.

At the end, more than 100 indicators are available. Those indicators are used to describe the land fabric, to feed parametric climate models such as the Town Energy Balance (TEB) or the Weather Research and Forecasting (WRF) models, to perform classifications such as the Local Climate Zones.

2.2 Indicators used

The 14 indicators needed for the LCZ classification procedure are described Tab. 1. Note that the land cover fraction used in this work are calculated considering that high vegetation may be above any other land cover (including buildings). Thus if the high vegetation ~~superimposed~~ is superimposed on an other land cover, the sum of all land cover may be higher than 1.

Table 1: Description of the calculation method for each indicator used

Indicator abbreviation	Indicator name in Geoclimate library (or combination of Geoclimate indicators)	Calculation method*	Defined in Stewart et Oke (2012)
SVF	GROUND_SKY_VIEW_FACTOR	The calculation is based on the [ST_SVF](http://www.h2gis.org/docs/dev/ST_SVF/) function of H2GIS using only buildings as obstacles and with the following parameters: ray length = 100, number of directions = 60. Using a uniform grid mesh of 10 m resolution, the SVF obtained has a standard deviation of the estimate of 0.03 when compared with the most accurate method (according to Bernard et al. (2018)). Using a grid of regular points, the density of points used for the calculation actually depends on building density (higher the building density, lower the density of points). To avoid this phenomenon and have the same density of points per free ground surface, we use an H2GIS function to distribute randomly points within free surfaces (ST_GeneratePoints). This density of points is set by default to 0.008 (based on the median of Bernard et al. (2018) dataset).	Yes
H/W	ASPECT_RATIO	A simple approach based on the street canyons assumption is used for the calculation. $\frac{\sum_i^n A_{BF_i}}{A_{RSU}-\sum_i^n A_{B_i}}$ where n the number of buildings in the RSU A_{BF_i} the facade area of the building i A_{B_i} the area of building i A_{RSU} the area of the RSU	Yes
F_B	BUILDING_FRACTION_LCZ	$\frac{\sum_i^n A_{B_i}}{A_{RSU}}$ where n the number of building in the RSU A_{B_i} the area of building i A_{RSU} the area of the RSU	Yes
F_I	IMPERVIOUS_FRACTION_LCZ	$\frac{\sum_i^n A_{I_i}}{A_{RSU}}$ where n the number of impervious patches in the RSU (geometries from road and impervious tables) A_{I_i} the area of impervious patch i A_{RSU} the area of the RSU	Yes
F_P	PERVIOUS_FRACTION_LCZ	$\frac{\sum_i^n A_{P_i}}{A_{RSU}}$ where n the number of pervious patches in the RSU (geometries from water and vegetation tables, not intersecting with any geometry from impervious, road or building tables) A_{P_i} the area of pervious patch i A_{RSU} the area of the RSU	Yes
Hr	GEOM_AVG_HEIGHT_ROOF**	$\exp(\frac{\sum_i^n \log(H_{B_I})}{n})$ where n the number of buildings in the RSU H_{B_i} the height of building i	Yes
z_0	EFFECTIVE_TERRAIN_ROUGHNESS_LENGTH***	Based on the Hanna and Britter (2010) procedure (see equation (17) and examples of calculation p. 156 in the corresponding reference). If $\lambda_f \leq 0.15 : \lambda_f \cdot H_r$ If $\lambda_f > 0.15 : 0.15 \cdot H_r$ where Hr the geometric average height roof within the RSU $\lambda_f = \frac{\sum_j^{ndir} \sum_i^n A_{PF_{i,j}}}{ndir} \cdot \frac{1}{A_{RSU}}$ $A_{PF_{i,j}}$ the area of building facade i when projected in the j direction $ndir$ the number of direction used to calculate the projected facade area (default 12) A_{RSU} the area of the RSU Warning: the calculation is only performed for angles included in the range [0, 180[° (considered as equal for a given orientation independently of the direction). When the RSU do not split buildings this calculation is right but can slightly overestimate the results when buildings are split (there is actually overestimation only in one direction).	Yes
F_{AV}	HIGH_VEGETATION_FRACTION_LCZ + LOW_VEGETATION_FRACTION_LCZ	$\frac{\sum_i^n A_{V_i}}{A_{RSU}}$ where n the number of all vegetation (low and high) patches in the RSU A_{V_i} the area of vegetation patch i A_{RSU} the area of the RSU	No

Table 1: Description of the calculation method for each indicator used

Indicator abbreviation	Indicator name in Geoclimate library (or combination of Geoclimate indicators)	Calculation method*	Defined in Stewart et Oke (2012)
F_W	WATER_FRACTION_LCZ	$\frac{\sum_i^n A_{W_i}}{A_{RSU}}$ where n the number of water patches in the RSU A_{W_i} the area of water patch i A_{RSU} the area of the RSU	No
$F_{HV/AV}$	HIGH_VEGETATION_FRACTION_LCZ / (HIGH_VEGETATION_FRACTION_LCZ + LOW_VEGETATION_FRACTION_LCZ)	$\frac{\sum_i^{n_{hv}} A_{HV_i}}{\sum_i^n A_{V_i}}$ where n_{hv} the number of high vegetation patches in the RSU A_{HV_i} the area of high vegetation patch i n the number of all vegetation (low and high) patches in the RSU A_{V_i} the area of vegetation patch i	No
$F_{IND/B}$	AREA_FRACTION_HEAVY_INDUSTRY / BUILDING_FRACTION	$\frac{\sum_i^{n_{bind}} A_{BIND_i}}{\sum_i^n A_{B_i}}$ where n_{bind} the number of industrial buildings in the RSU (buildings having TYPE = 'HEAVY INDUSTRY') A_{BIND_i} the area of the industrial building i n the number of buildings in the RSU A_{B_i} the area of the building i	No
$F_{LLR/B}$	(AREA_FRACTION_COMMERCIAL) + AREA_FRACTION_LIGHT_INDUSTRY) / BUILDING_FRACTION	$\frac{\sum_i^{n_{bllr}} A_{BLLR_i}}{\sum_i^n A_{B_i}}$ where n_{bllr} the number of large low-rise buildings in the RSU (buildings having TYPE = 'COMMERCIAL' or TYPE = 'LIGHT INDUSTRY') A_{BLLR_i} the area of the large low-rise building i n the number of buildings in the RSU A_{B_i} the area of the building i	No
$F_{RES/B}$	AREA_FRACTION_RESIDENTIAL / BUILDING_FRACTION	$\frac{\sum_i^{n_{bres}} A_{BRES_i}}{\sum_i^n A_{B_i}}$ where n_{bres} the number of residential buildings in the RSU (buildings having TYPE = RESIDENTIAL) A_{BRES_i} the area of the residential building i n the number of buildings in the RSU A_{B_i} the area of the building i	No
N_{lev}	AVG_NB_LEV_AREA_WEIGHTED	$\frac{\sum_i^n A_{B_i} \cdot Nlev_{B_i}}{\sum_i^n A_{B_i}}$ where n the number of buildings in the RSU A_{B_i} the area of the building i $Nlev_{B_i}$ the number of level of building i	No

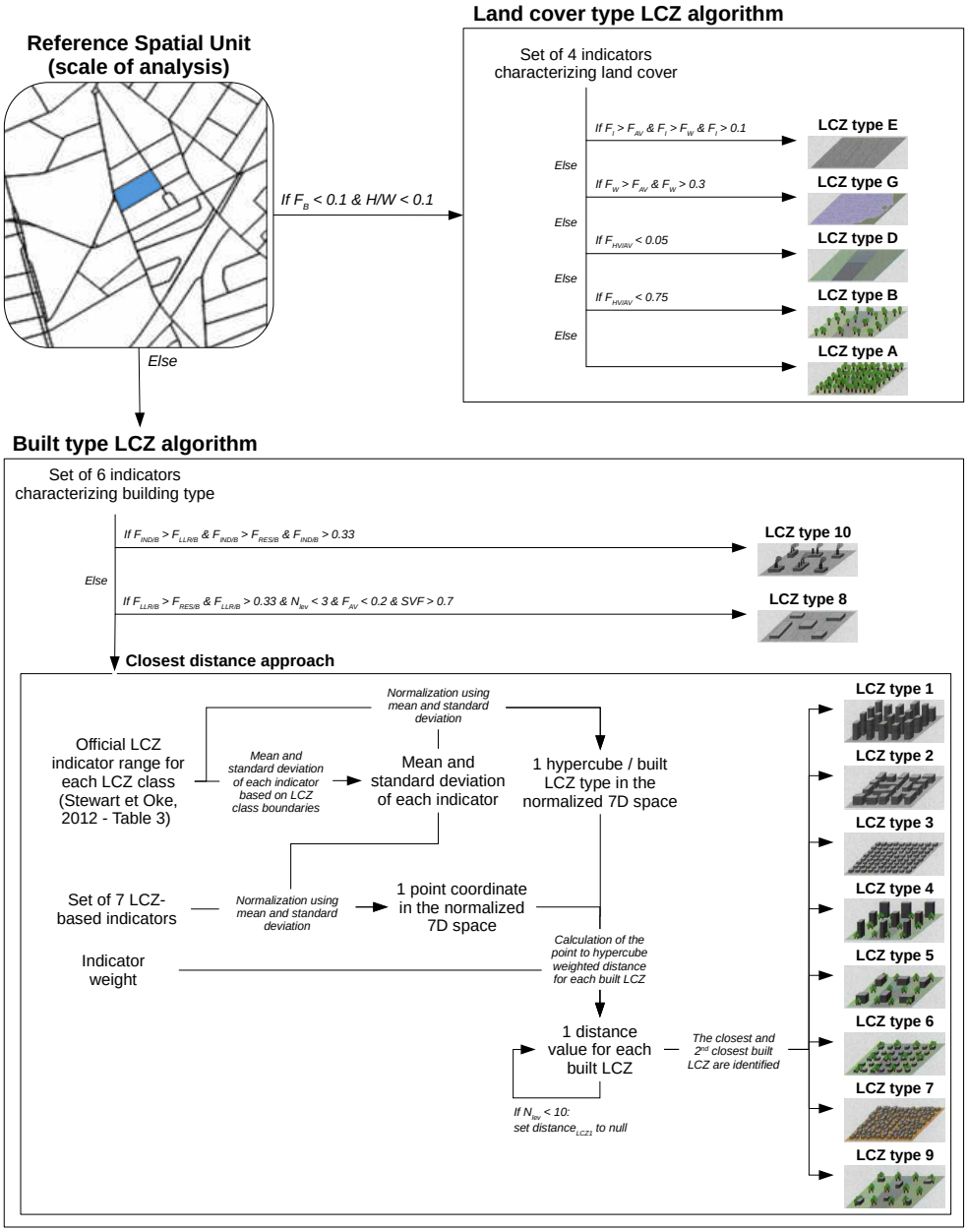
* This calculation is generic for any dataset since Geoclimate is based on a generic model described by Bocher et al. (2021)

** Stewart and Oke (2012) calls this "Height of roughness elements" and consider also the tree height in the calculation

*** In Stewart and Oke (2012), terrain roughness classes are actually used, which are defined according to roughness length ranges of values. We prefer to keep the length which is a continuous information (while the class would be a categorical variable)

2.3 Algorithm

In the GeoClimate code, the LCZ algorithm that assigns a LCZ to each RSU is called “identifyLczType”. It can work for any RSU definition as long as all needed indicators have been previously calculated for these units. The method, based on the Stewart and Oke (2012) ~~referential~~approach, is illustrated Fig. 2.



F_B: building fraction
H/W: aspect ratio
F_I: impervious fraction
F_{AV}: all vegetation fraction
F_W: water fraction
$F_{HV/AV}$: fraction of all vegetation being high vegetation

$F_{IND/B}$: fraction of buildings being industrial
$F_{LLR/B}$: fraction of buildings being large low rise
$F_{RES/B}$: fraction of buildings being residential
N_{liv}: area weighted average number of building level within a RSU
SVF: ground sky view factor

Figure 2. General procedure used to classify a RSU to a LCZ

A RSU is treated differently regarding the number and the size of the buildings it contains. If the building fraction and the aspect ratio are both lower than 0.1, the RSU is considered only as a land cover type and its LCZ will be affected using the "land cover type LCZ algorithm". Otherwise the built type LCZ algorithm will be used. The threshold of 0.1 is set according to ~~the Stewart et Oke referential~~Stewart and Oke (2012): building fraction and aspect ratio are higher than 0.1 for all LCZ built types while they are lower for all LCZ land cover types (if trees are excluded from the aspect ratio calculation, which is our case).

The land cover type algorithm currently works for five of the seven LCZ land cover types: LCZ types A, B and D (respectively *dense trees*, *scattered trees* and *low plants*), LCZ type E (*bare rock or paved*) and LCZ type G (*water*). The LCZ type F (*bare soil or sand*) and the LCZ type C (*bush, scrubs*) can not be identified since we currently consider only five land cover types: buildings, impervious, water, low vegetation and high vegetation. We first consider that our raw data are continuous anywhere on the planet, thus any piece of land should be covered by one of the five selected land cover types. This assumption can induce some errors in case there is missing information in the datasource. After considering the building fraction (which is lower than 0.1 for non urbanized LCZ types), if the impervious fraction is higher than the water fraction and higher than the vegetation fraction (low + high), thus the RSU can be set to LCZ E (impervious) since impervious areas correspond to the major fraction. In reality, the sum of our land fractions rarely reaches 100%. Depending on the data we use, we may have non identified areas. The position and the size of buildings, roads (at least the center line) and water are often accurately known. However, this is not the case for vegetation (especially in urban private lands) and impervious areas (such as sidewalk, parking lot, etc). Thus a land cover fraction should be higher than a given threshold in addition to represent the major fraction of a RSU. The higher the probability that a given land cover would be missing in the data, the lower the corresponding threshold. Based on empirical observations using OSM and BDT, we set these thresholds to 0, 0.1 and 0.3 respectively for vegetation, impervious and water. Hence, a RSU must satisfy the following condition : having vegetation fraction higher than 0.3 to be a water LCZ, impervious fraction higher than 0.1 to be an paved LCZ and vegetation fraction higher than 0. Results from simple examples may be presented from Fig. 3. If impervious and water fractions are below their threshold (respectively 0.1 and 0.3), the land cover type is set to vegetation. Whenever impervious or water fractions exceed their threshold, they should also be larger than any other land cover fractions. If the land cover is vegetation, the LCZ type is then set according to the high vegetation fraction to all vegetation (low and high) fraction ratio. The threshold values to distinguish *low plants* from *scattered trees* and *scattered trees* from *dense trees* have been roughly set to 0.05 and 0.75 arbitrarily from the drawing of each class and empirical observation using our data.

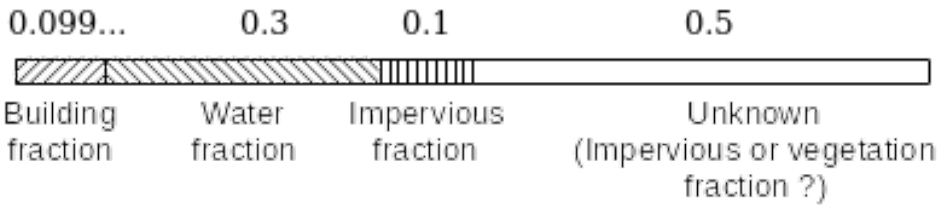


Figure 3. Representation of the land cover repartition with their default thresholds

The built type algorithm works for all LCZ built types. It is mainly based on a "closest distance approach" (called fuzzy rule-based approach in Quan and Bansal (2021)). Stewart and Oke (2012) proposed a set of seven UCPs to characterize the morphology and the land cover properties of a given area. In Table 3 of their manuscript, they set to each of the LCZ type a range of possible values taken by each of these UCPs. In a formal approach, a LCZ type is defined as an hypercube in a seven dimension space. In this space, a given RSU is defined by a point, its coordinates being the value of the seven UCPs. Then the "closest distance approach" consists in identifying the LCZ type hypercube the closest to our RSU point. Note that the distance is clearly influenced by the dimension having the largest variability: the building height which can vary from zero (no building) to several hundred meters for the highest buildings will have much more impact on the distance than the building fraction which is included in a [0, 1] range. Thus each dimension is normalized using the mean and the standard deviation of all LCZ type boundary values (note that we have replaced the initial "terrain roughness class" indicator by the more continuous information "effective terrain roughness length" using the Davenport et al. (2000) conversion table). However, we leave the opportunity to give more importance to some of the UCPs adding a given weight to each dimension. This can be useful for several reasons:

- the data used as input may not represent well the reality, thus we may decrease the weight of all impacted ~~axis~~ axes (e.g. pervious fraction if the vegetation is rarely identified),
- the method used to calculate some UCPs might not be in total agreement with the ~~Stewart et Oke~~ Stewart and Oke (2012) definition (SVF does not take into account vegetation nor elevation), a reason to decrease its corresponding weight,
- the user may simply think that the seven UCPs should not have equal weight in the classification.

Actually, three LCZ built types have a specific behavior:

- LCZ 1 (*compact high-rise*): the "closest distance approach" is used but the distance of a RSU to the LCZ 1 hypercube is set to null wherever the mean number of building level in the RSU is lower than 10 (threshold set according to the Stewart (2011) description of LCZ 1). Without this constraint, we have obtained LCZ 1 in numerous European cities where most of the urban researchers would not set any (Demuzere et al., 2019).
- LCZ 10 (heavy industry): the shape and the land cover of these zones results directly from the building use. Thus we decide to exclude this LCZ type from the "closest distance approach": a RSU is set as heavy industry when its fraction

of heavy industry among the building exceeds those of residential building and large low rise buildings and when it is higher than 0.33.

- LCZ 8 (large low-rise): as for LCZ 10, we set a RSU as large low rise when its building fraction exceeds those of industrial and residential and is higher than 0.33. However, even though the building use allows to simply identify those areas, we may have mid or high-rise mall in our sample. Thus the maximum average number of building levels should be lower than three, the SVF should be higher than 0.7 and the vegetation fraction lower than 0.2 (these thresholds come from Stewart and Oke (2012) and Stewart (2011)).

At the end of the "closest distance algorithm", two LCZ types are conserved: the closest (called LCZ Primary) and the second closest (LCZ2 Secondary). For all the RSU set with an other approach, only LCZ Primary has a value (LCZ2 Secondary is set to null).

2.4 Indicators of uncertainty

Once the LCZ type of each RSU is set, we can wonder how accurate is this information. If the LCZ type has been set according to the "closest distance approach", the distance to the closest LCZ is stored in the "MIN_DISTANCE" field. If the distance is 0, it means that the point is within the hypercube. Then higher is this value, worse is the classification. However, a point may have a relatively low "MIN_DISTANCE" but be at almost equal distance to several hypercubes. Then we calculate the "LCZ_UNIQUENESS_VALUE" defined by Eq. 1. Its value ranges between 0 and 1: the higher it is, the more relevant the LCZ type set to the RSU.

$$LCZ\_UNIQUENESS\_VALUE = \frac{|d_{closestLCZ} - d_{2^{nd}closestLCZ}|}{d_{closestLCZ} + d_{2^{nd}closestLCZ}} \quad (1)$$

where $d_{closestLCZ}$ is the distance from the RSU point to the closest LCZ hypercube $d_{2^{nd}closestLCZ}$ is the distance from the RSU point to the second closest LCZ hypercube

In all other cases (when the LCZ type is not set using the "closest distance approach"), "MIN_DISTANCE" and "LCZ_UNIQUENESS_VALUE" are by default set to null.

2.5 Sensitivity analysis: influence of the input data

The LCZ procedure proposed in Geoclimate is generic in the sense that it works with a fixed input tables set. However, we want to investigate the impact of input data modification (BDT or OSM) on the resulting LCZ map. According to preliminary observations, we have identified two major differences between these data: the building height is not filled for most buildings in OSM while land cover coverage seems better in OSM than in BDT.

~~Concerning building height lack, it~~ Lack of building height data is a major concern for urban climate studies. For this reason, the missing heights have been estimated using a Random Forest method based on geographical indicators characterizing the building environment (Bernard et al., 2022). Although the RSU mean building height could be quite far from the truth for

certain areas, other indicators such as the ground sky view factor are less impacted by the quality of the building height estimation. Then we expect that the LCZ classification using this modified OSM data would lead to comparable results when using BDT data.

Concerning land cover coverage, we expect OSM results to have a lower fraction of undefined land (for each RSU this fraction is calculated and called UNDEFINED_FRACTION). To verify these expectations, we have run Geoclimate on the 22 French communes that have previously been used in Bernard et al (2022). More information about these territories is given Tab. 2.

Table 2. Information and statistics about the communes used as study areas

Commune type	Commune name	INSEE code	Inhabitants (2019)	Population density (nb inhabitants/km^2)
Main urban areas	Paris	75056	2,165,423	20,545
	Toulouse	31555	493,465	4,171
	Nantes	44109	318,808	4,890
	Rennes	35238	220,488	4,376
	Dijon	21231	158,002	3,910
	Annecy	74010	130,721	1,953
	Avignon	84007	91,143	1,404
	La Rochelle	17300	77,205	2,716
	Nanterre	92050	96,277	7,898
	Meudon	92048	45,818	4,628
Small urban or rural areas	Blagnac	31069	25,525	1,512
	Charnay-lès-Mâcon	71105	7,742	616
	La Haie Fouassière	44070	4,691	397
	Gratentour	31230	4,387	1,073
	Staffelfelden	68321	4,046	545
	Allaire	56001	3,886	93
	Saint-Nicolas de Redon	44185	3,245	145
	Pont-de-Veyle	1306	1,641	846
	Bourgneuf	17059	1,375	513
	Corbonod	1118	1,278	41
	La Thuile	73294	338	19
	Saint-Ganton	35268	429	31

The GeoClimate LCZ algorithm is partially based on relative weights indicators (cf. Sect. 2.3). Default weights have been used for both OSM and BDT: SVF=4, H/W=3, FB=8, FI=0, FP=0, Hr=6, z0 = 0.5. Those weights are only used in the

"closest distance approach" for LCZ built types. Two main characteristics differentiate LCZ types: building compacity (mainly described by FB, SVF and H/W) and building height (mainly described by Hr, H/W and SVF). FP and FI weights are set to zero since they are secondary characteristics and pervious and impervious data often lack of accuracy in urban areas (at least it is the case for our input datasets). FB and Hr indicators are simply defined and are based on very few input variables, thus do not propagate uncertainties. However, building height is less certain than building footprint (especially for OSM data), thus FB has the highest weight. The SVF and H/W have lower weights since they do not consider all variables they should (vegetation and terrain level variations are excluded from the calculations). Moreover, H/W calculation method assumes that all LCZ built type are street canyon, thus we set its weight slightly lower than SVF which is calculated considering the real building setting. We may have had decreased the Hr weight for the OSM data but preliminary tests showed that decreasing its value down to 2 does not affect much the results.

3 Results

3.1 Scale used for comparison

The spatial units generated with OSM differ from those created with BDT since territory segmentation is performed using topographical data coming from two different sources. BDT partitioning results in about twice the number of RSU than OSM partitioning for most of the territories. Thus the mean RSU area is more than twice ~~bigger~~ as large in OSM than in BDT data (18,514 against 8,636 m^2). The median RSU area are much smaller than the mean (3,013 for OSM and 1167 m^2 for BDT), revealing the influence of some bigger RSU. However, the ratio between OSM median RSU size and BDT median RSU size remains higher than 2. Three main reasons explain this observation. First, many forest patches outside city centers present in BDT do not exist in OSM. Second, some roads which are considered of secondary importance in OSM are conserved for segmentation in BDT. Third, the rules to edit data in OSM are more restrictive than BDT due to relational data model used by OSM to store and describe a geometry. In BDT the geometries are stored in layers independently of each other (vegetation, water...), consequently there is a higher probability to find overlaps and gaps between the layers (e.g a surface of impervious that covers a surface of water), than in OSM. Indeed, in OSM all geometries are defined in three tables (nodes, ways and relations), therefore the snapping between geometries are more consistent, resulting in a lower number of very small RSU (Fig. 4). To compare the LCZ at RSU scale, a first step is then to create as many units as there are RSU intersections between the two datasets, thus leading to an increase of small size units.

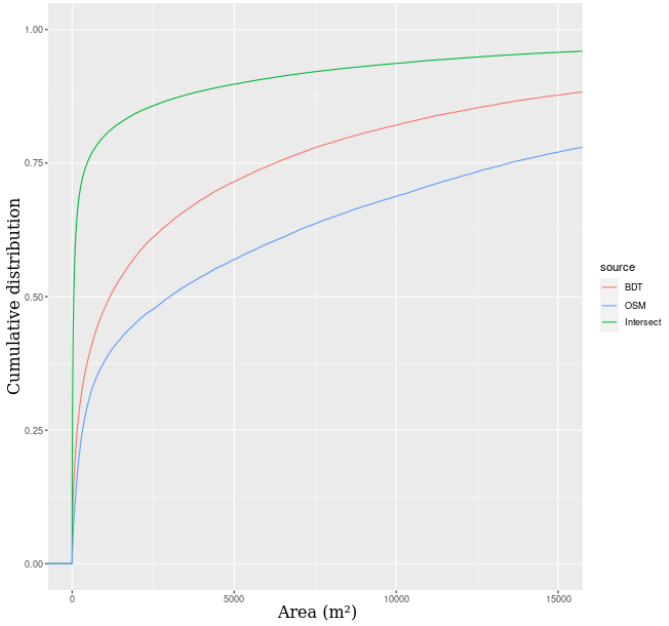


Figure 4. Cumulative distribution function of RSU areas with OSM input, BDT input, and their intersection

3.2 General agreement

For each LCZ, the areas of geometries who received the same LCZ value using BDT or OSM data as inputs are summed to compute a weighted agreement per LCZ value and a general agreement for the whole commune.

A figure comparing the OSM and BDT results is automatically generated for each city using the lczexplore R package[4]. It contains the LCZ map created for both datasets, a bi-colored map (red, green) that shows the spatial distribution of the zones in agreement and a confusion table based on percentage of area in agreement for each LCZ type. The Fig. 5 and 6 show a comparison of the LCZ obtained for the city of Dijon and for the rural commune of Saint-Nicolas de Redon, respectively. They illustrate well the main similarities and differences observed for major cities and rural territories.

Concerning Dijon, the city center (compact LCZ built-types) and the urban ring (open LCZ built-types) are well identified in BDT and OSM, even though the city center seems slightly bigger in OSM than in BDT. The built-up zone is visually more homogeneous in OSM than in BDT. The first reason is that the model used to predict OSM building height smooths its spatial variations (Bernard et al., 2022). The second reason is that the territory is more fragmented using BDT than OSM data (the number of RSU is 4,548 and 2,819 respectively for the whole Dijon commune). In the rural areas outside the city, the LCZ vegetation type is more diverse in BDT than in OSM. However, there is a rather good spatial agreement (81%) between zones covered by vegetation when we do not consider LCZ vegetation types (when we merge all vegetation). The agreement matrix can be used to identify the main misclassifications: for instance 84 percent of the area set to *compact low-rise* using BDT data

[4]https://github.com/MGousseff/lczexplore (last access: 7 February 2023)

is set to *compact mid-rise* using OSM data. However, *compact low-rise* type covers a negligible fraction of the Dijon territory (0.51%) as it is shown in the legend of Fig. 5 in the upper left map.

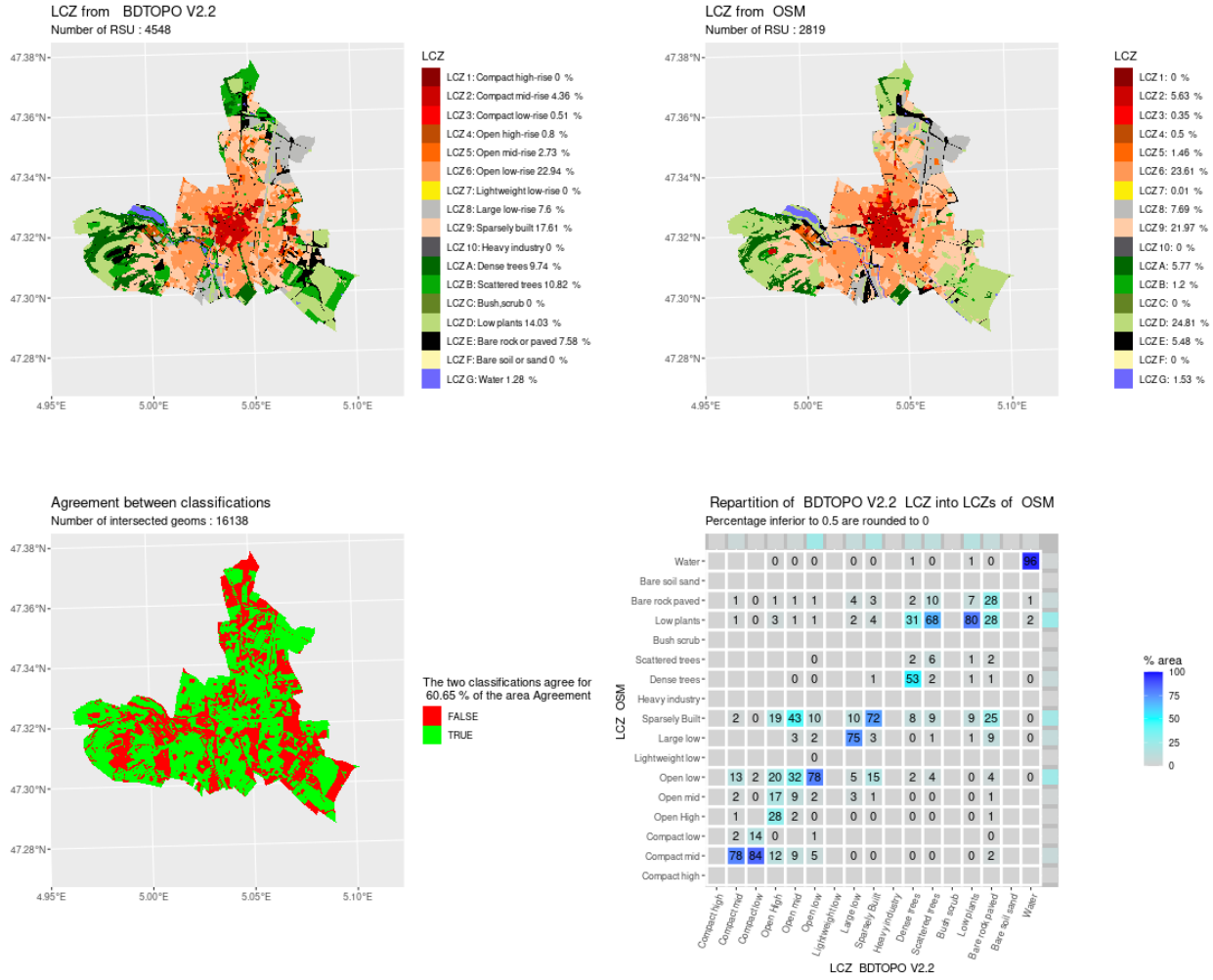


Figure 5. Comparison of LCZ generated for the city of Dijon by the GeoClimate method using BDT and OSM datasets

Concerning Saint-Nicolas de Redon, the observations made for Dijon remain valid: the center of the village is well identified using both OSM and BDT data. The vegetation is more heterogeneous in BDT than in OSM, resulting in a rather low agreement fraction (44%). For such territory having very low LCZ built type, the vegetation type has a considerable impact since once the vegetation is merged, the general agreement fraction is very high (94%).

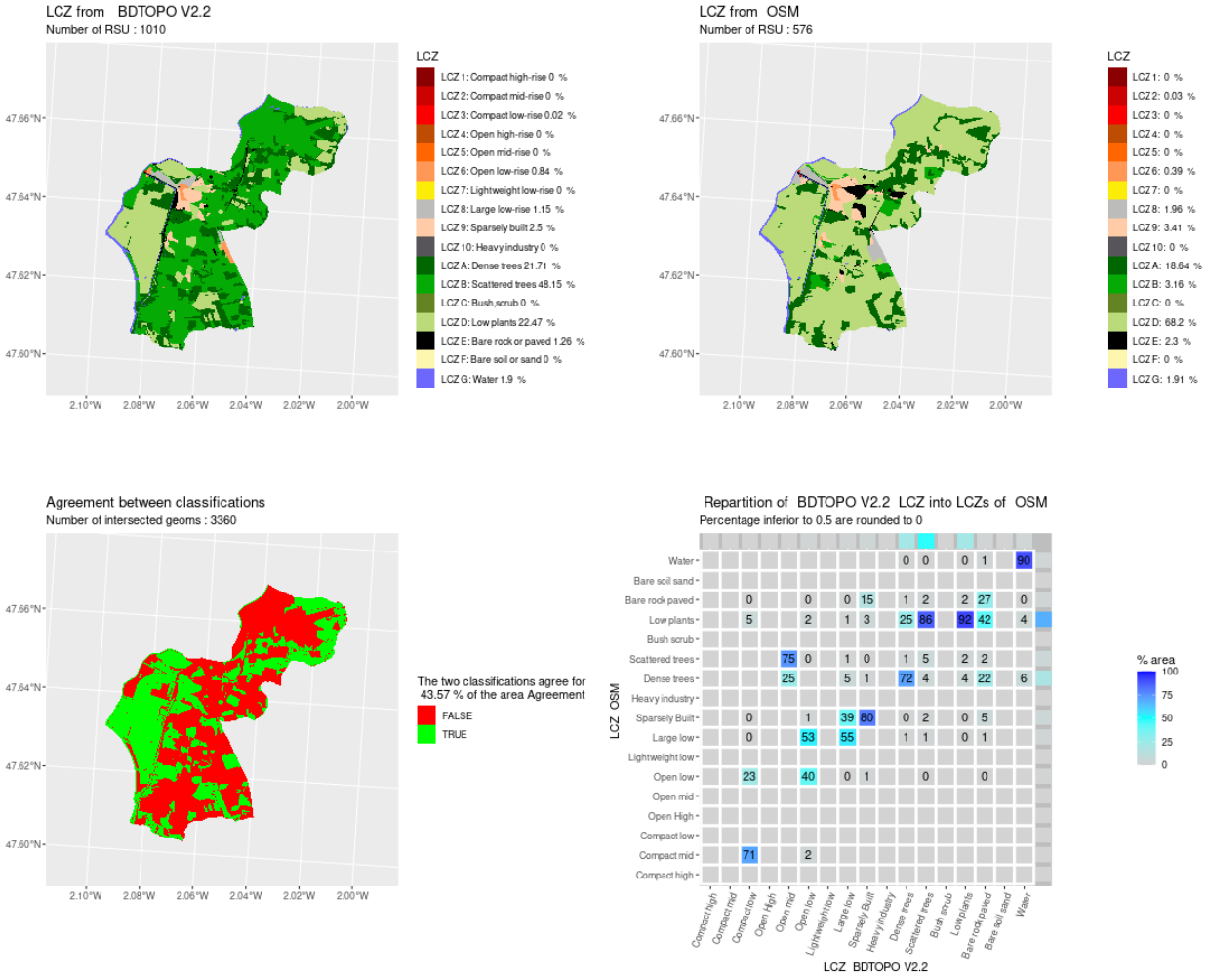


Figure 6. Comparison of LCZ generated for the city of Saint-Nicolas de Redon by the GeoClimate method using BDT and OSM datasets

~~Over~~ If we gather LCZ types into urban (all LCZ built type plus the bare rock and paved class) and rural classes (the rest of the LCZ land types) and compare the results obtained using OSM to the ones obtained using BDT, the degree of agreement is 87% on average when considering all cities. If we consider the agreement LCZ type by LCZ type, over 55% of all territories area has the same LCZ classification between OSM and BDT. This statistic may differ a lot by territory but also by LCZ type. The best agreements are found for *open low-rise* and *compact mid-rise* (which are the more common LCZ built types) as well as *water* with 79%, 81% and 91%, respectively (Fig. 7). The worst agreements are found for *lightweight low-rise* (1%), *scattered trees* (7%), *compact high-rise* (8%), *open mid-rise* (19%), *open high-rise* (21%) and *paved* (29%) but only *scattered trees* and *paved* represent a non-negligible area (15 and 8% with BDT input, none of the other above 2% of the total area). A majority of the *scattered trees* in BDT RSU has been classified as *low plant* in OSM. This difference is clearly attributable to the spatial resolution of the datasets since many small patches of forest are identified only in the BDT. Respectively 68% and

73% of *open mid-rise* and *open high-rise* in BDT are classified to lower rise types in OSM. This outcome is probably due to the model used to estimate OSM building height which often produces underestimations of mid and high-rise buildings in open areas (Bernard et al., 2022).

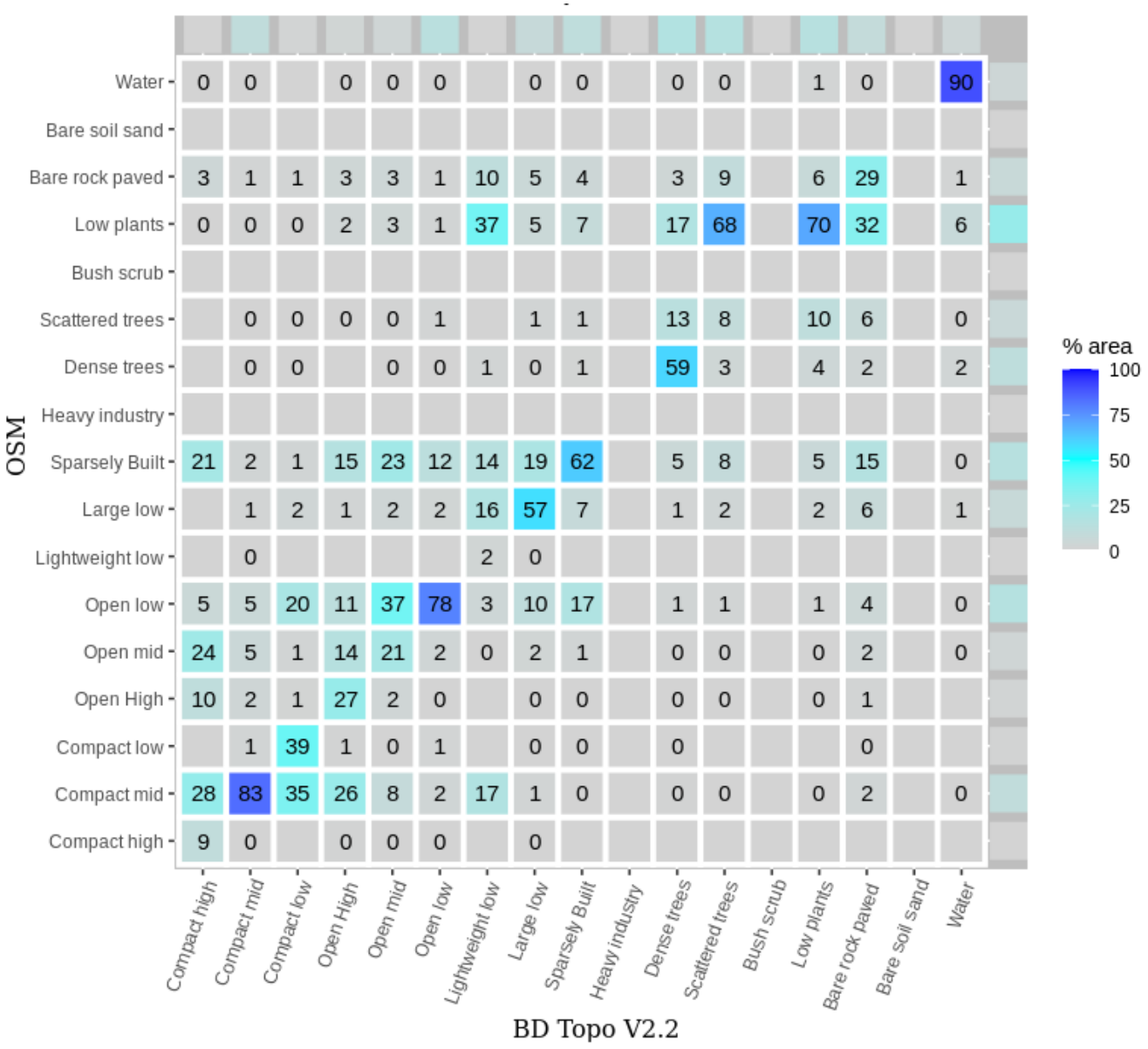


Figure 7. Repartition of BDT LCZ into OSM LCZ

3.3 Uncertainty indicator

For all urban classes except large low-rise and heavy industry, an uncertainty indicator called uniqueness value has been calculated (cf. Eq. 1). The higher the uniqueness value, the more certain the LCZ type attributed to a RSU. A value of 25% seems to be a reasonable threshold to filter out uncertain values. Above this threshold (more than 40% of the RSU), the agreement between OSM and BDT LCZ is higher than 90% while the agreement below the threshold is about 50% (Fig. 8). Note that this result is only valid for RSU having a uniqueness value. For those that do not have one, the average agreement is about 45%.

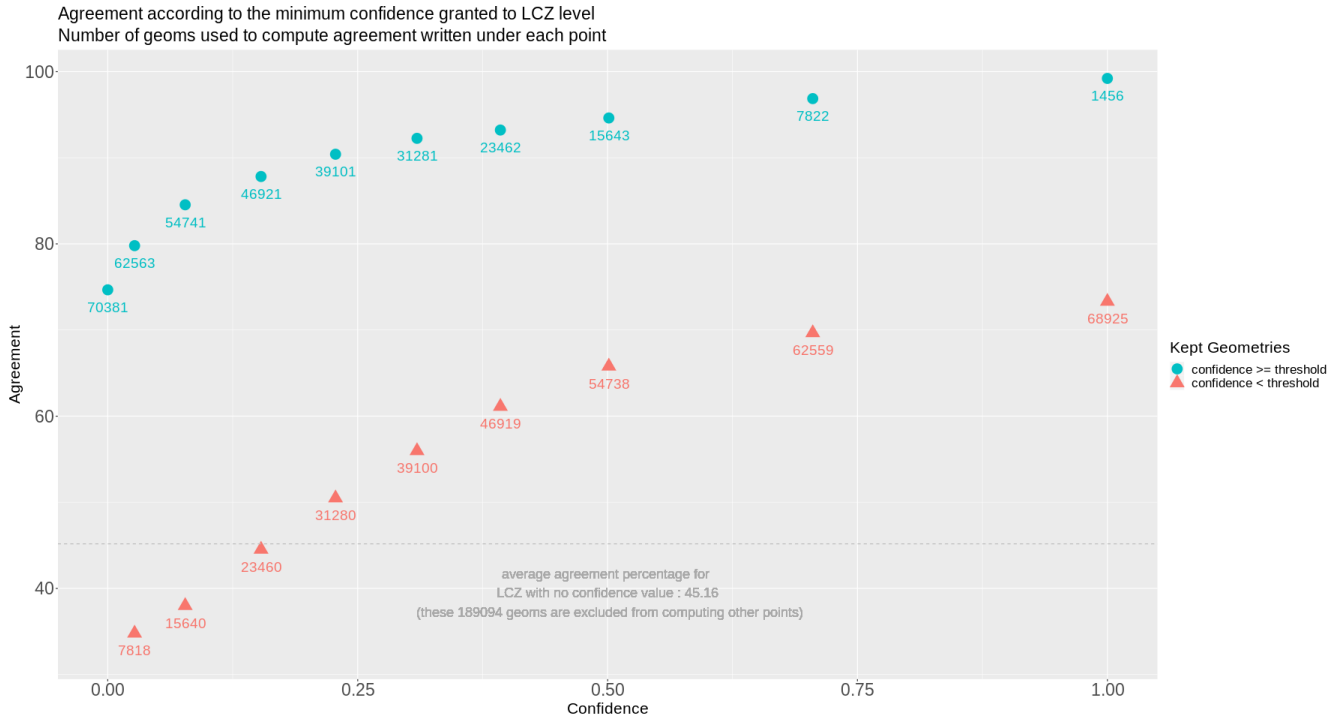


Figure 8. General agreement according to the minimum confidence granted to OSM or BDT LCZ attribution

The confidence threshold does not have the same impact for all LCZ types (Fig. 8). For those having already a quite good agreement such as *compact mid-rise*, the agreement fraction cannot increase much with the confidence threshold. Setting a threshold to 0.25 increases the agreement between OSM and BDT from about 90% to 98% but on the other hand, about 75% of the RSU that will be filtered out (50% of the total) show agreement between BDT and OSM LCZ. On the contrary, it has a positive impact for LCZ types having a low agreement (such as *open high-rise*): setting a threshold to 0.25 will increase the agreement from 30% to more than 60%. This filter will remove more than 70% of the RSU set as *open high-rise*, with only 25% of them having agreement between BDT and OSM.

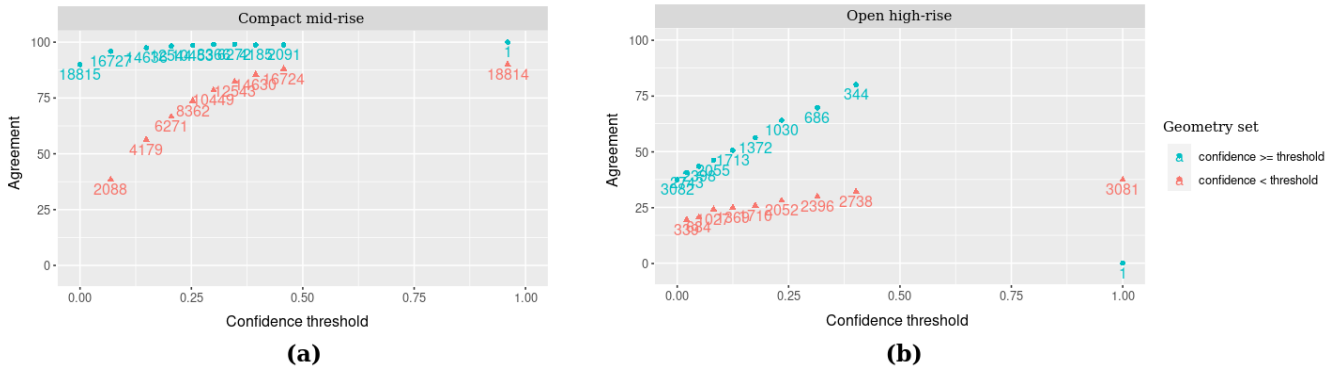


Figure 9. Agreement according to the minimum confidence granted to OSM or BDT LCZ attribution (a) when LCZ is *compact mid-rise* and (b) when LCZ is *open high-rise*

Concerning non-urbanized LCZ types, the major indicator of confidence is the fraction of non defined land. This fraction is calculated at RSU scale and is called "UNDEFINED_FRACTION". As expected from preliminary investigations, the OSM data has a higher land coverage: in average, for the 22 territories used as study areas, the fraction of undefined land is 37% in OSM against 55% in BDT.

3.4 City specificities

The agreement fraction by city is in average 55% and varies from 30% (Gratentour) to 82% (Corbonod - cf. Tab. 3 and 4). Most of the very low scores for rural or small urban areas are found for territories having small patches of high vegetation. As previously discussed in Sect. 3.2, only BDT demonstrates this level of details. Territories containing small patches of vegetation (most of them being agricultural lands) are then identified as *dense trees* or *scattered trees* in BDT instead of *low plants* in OSM. This is the case for Allaire having only 23% agreement for areas covered by dense trees in BDT while it represents 20% of its territory area, but also for Gratentour and Pont-de-Veyle having both no agreement for area covered by *scattered trees* in BDT while it represents respectively 53% and 43% of their territory area. For territories having large and homogeneous vegetation types the general agreement is much better. This is the case for forested areas such as Corbonod (or La Thuile): 65% (62%) of its territory is covered by dense trees in BDT and it has 94% (87%) agreement with OSM data for this specific land type.

Concerning main urban areas, the agreement between OSM and BDT is partially correlated to the fraction of their natural land and its corresponding type (Tab. 3 and 4). Annecy and Meudon show the highest agreement fraction (64% and 70%) and have respectively 31% and 45% of their territory covered by a large and homogeneous patch of dense trees (thus having 84% and 90% agreement fraction). However, their agreement fraction for most of the mid and high-rise LCZ types is very low. As described by Bernard et al. (2022), Annecy and Meudon are two cities where the model used to estimate OSM building height fails quite a lot. This is clearly highlighted by their very low agreement for *compact mid-rise* LCZ (47% for Annecy and 40% for Meudon while the average is 81%). On the contrary, Paris is the city having the highest agreement for almost all

mid-rise and high-rise LCZ types. The reason is twofold: many Paris building have a height tag in OSM and most of the Paris urban fabric has a quite homogeneous structure (Haussmannian architecture - large blocks of buildings of regular height with internal courtyards) which has been well caught by the building height model. However, if low-rise buildings are in general overestimated by the model, it is particularly the case for Paris where agreement fraction for compact low-rise (LCZ3), open low-rise (LCZ6) and large low-rise (LCZ8) are respectively 23%, 40% and 40% lower in Paris than the average for these classes.

Table 3. Percentage of agreement using BDT and OSM data for urban LCZ types

Territory types	Commune name	all LCZ types	LCZ1	LCZ2	LCZ3	LCZ4	LCZ5	LCZ6	LCZ7	LCZ8	LCZ9	LCZ10
Rural or small urban areas	Allaire	46	-	-	-	-	75	49	-	81	64	-
	Blagnac	37	-	18	-	-	2	79	-	28	49	-
	Bourgneuf	77	-	-	-	-	-	98	-	-	90	-
	Charnay-lès-Mâcon	64	-	26	-	-	-	91	-	18	85	-
	Corbonod	**82**	-	-	-	-	-	-		-	85	-
	Gratentour	**30**	-	-	-	-	-	20	-	91	72	-
	La-Haie-Fouassière	60	-	-	-	-	-	72	-	89	74	-
	La Thuile	63	-	-	-	-	-	-	-	97	53	-
	Pont-de-Veyle	34	-	-	-	-	-	66	-	-	44	-
	Saint-Ganton	46	-	-	-	-		83	-	-	76	-
	Saint-Nicolas de Redon	44	-	-	0	-	-	40	-	55	80	-
	Staffelfelden	73	-	-	-	-	-	-	-	20	92	-
Main urban areas	Annecy	64	-	**47**	9	2	9	56	-	71	71	-
	Avignon	40	-	77	52	0	15	75	-	64	51	-
	Dijon	61	-	78	14	28	9	78	-	75	72	-
	La Rochelle	59	-	69	44	4	21	84	-	89	42	-
	Meudon	**70**	0	**40**	-	40	7	81	-	-	56	-
	Nanterre	39	0	17	65	16	27	88	-	11	38	-
	Nantes	60	0	68	36	19	21	84	-	62	59	-
	Paris	63	12	89	**23**	39	41	**39**	4	**14**	61	-
	Rennes	47	0	65	25	14	21	82	-	64	47	-
	Toulouse	50	0	70	51	3	15	80	-	50	65	-
All types	All cities	-	8	**81**	**46**	21	19	79	1	**54**	63	-

Table 4. Percentage of agreement using BDT and OSM data for rural LCZ types

Territory types	Commune name	all LCZ types	LCZA	LCZB	LCZC	LCZD	LCZE	LCZF
Rural or small urban areas	Allaire	46	**23**	26	-	64	18	-
	Blagnac	37	65	1	-	69	10	-
	Bourgneuf	77	64	7	-	85	39	-
	Charnay-lès-Mâcon	64	70	15	-	75	43	-
	Corbonod	**82**	**94**	-	-	81	10	-
	Gratentour	**30**	14	**0**	-	75	41	-
	La-Haie-Fouassière	60	70	17	-	88	14	-
	La Thuile	63	**87**	-	-	86	23	-
	Pont-de-Veyle	34	81	**0**	-	74	11	-
	Saint-Ganton	46	86	-	-	98	6	-
	Saint-Nicolas de Redon	44	72	5	-	92	27	-
	Staffelfelden	73	87	-	-	84	14	-
Main urban areas	Annecy	64	**84**	2	-	74	30	-
	Avignon	40	17	6	-	79	36	-
	Dijon	61	53	6	-	80	28	-
	La Rochelle	59	21	2	-	62	20	-
	Meudon	**70**	**90**	11	-	39	32	-
	Nanterre	39	5	4	-	14	35	-
	Nantes	60	23	17	-	43	44	-
	Paris	63	26	26	-	18	35	-
	Rennes	47	42	9	-	53	25	-
	Toulouse	50	29	4	-	65	29	-
All types	All cities	-	54	7	-	69	29	-

4 Conclusion

According to Quan and Bansal (2021), two main streams coexist to identify Local Climate Zones. The first, based on images and supervised training, is applicable everywhere. However, the training is performed using expert classification that might be quite subjective inducing a potential bias between the resulting classification and the LCZ referential such as defined by Stewart and Oke (2012). The second stream is based on UCP values calculated using geographical data, resulting in a classification where the link with the LCZ referential is easier to make. The main shortcomings of the work presented in the literature for this second stream concern (i) datasets which are limited to a certain area (often a commune or a country) and (ii) methodologies that are often partially described and thus limited regarding their reproducibility. We presented a new method belonging to this stream that tries to address these limitations. The method described in the article can be reproduced by anyone since it is integrated in the free and open source software GeoClimate. It can be applied anywhere using OpenStreetMap data, available worldwide.

GeoClimate is designed to work with any dataset. Currently, it can automatically calculate LCZ using OSM and BDT data (a French national dataset). After a detailed description of the algorithm implemented in GeoClimate, it is applied to 22 French communes to compare the LCZ produced using OSM and BDT data. About 55% of the area of all studied territories has obtained the same LCZ type. The agreement fraction between OSM and BDT classification varies greatly between the communes (from 30 to 82%). Large patches of forest and water are well indexed in both datasources, thus leading to a good agreement for territories containing a large share of these land types. Concerning built LCZ types, the agreement is high for *compact mid-rise* and *open low-rise* (83% and 78% respectively) which are the main LCZ built types. However, a large part of the RSU classified into *open mid-rise* and *open high-rise* using BDT are set to *open low-rise* using OSM. This difference is attributed to the height underestimation for OSM buildings located in open areas (cf. (Bernard et al., 2022)).

Whenever a LCZ built type (except LCZ8 and LCZ10) is attributed to a RSU, a confidence indicator called "LCZ_UNIQUENESS_VALUE" is calculated. The agreement between OSM and BDT increases ~~quite much~~ substantially when we consider only RSU having a larger confidence value. Using a confidence threshold of 0.25 when mapping LCZ with the GeoClimate method is a good way to ensure that the LCZ type attributed to a RSU is reasonably accurate while minimizing the number of RSU removed from the analysis. This threshold only applied to some LCZ built types. There is currently no confidence indicator for LCZ8 or LCZ10, which might be a source of improvements for future versions of this work. Concerning non-urbanized LCZ types, a good indicator of confidence is the fraction of undefined land. This information is produced by GeoClimate for each RSU under the name "UNDEFINED_FRACTION". Above 50% of "UNDEFINED_FRACTION", we may assume that the attribution of a given LCZ is quite random. Moreover, some land types are currently not defined in GeoClimate (bare soil, sand, bush, scrubs) which induces missing LCZ types (C and F). These should soon integrated to a future GeoClimate version.

Quan and Bansal (2021) have identified six steps that are classically used in vector-based approaches for LCZ classification. The four first steps used in the GeoClimate methodology have been presented and the limitations corresponding to each dataset used as input identified. Potential future work could be to propose a methodology to aggregate small RSU into bigger ones

to reach the minimal LCZ unit size of 400 m wide (step 5) and then compare the resulting LCZ to climate data provided by observations or models (step 6).

The GeoClimate software has a great potential for new collaborations. Along with the lczexplore R package, it can be used to efficiently compare the LCZ produced with GeoClimate to any other method. The influence of each step of the LCZ creation can be investigated separately: impact of the dataset (such as presented in this paper), of the unit of analysis (RSU), impact of the method used for UCPs calculation and impact of the algorithm used to assign a LCZ. GeoClimate also has the potential to interact with the current WUDAPT approach. While GeoClimate may be used to train the WUDAPT model on areas where the results are quite confident, WUDAPT can in return be used on areas where OSM data are still quite poor. To confirm the complementarity between these two workflows, a more in-depth study of their differences on similar locations needs to be performed.

Code and data availability. The LCZ calculation is performed using the GeoClimate 0.0.1 software available at https://zenodo.org/record/6372337 while the figures used in the manuscript are created using the lczexplore R package available at https://zenodo.org/record/7646866. All the work performed in this manuscript can be reproduced following the Readme of the following Zenodo repository: https://zenodo.org/record/7687911.

Appendix A: Comparison of LCZ produced using BDT and OSM datasets

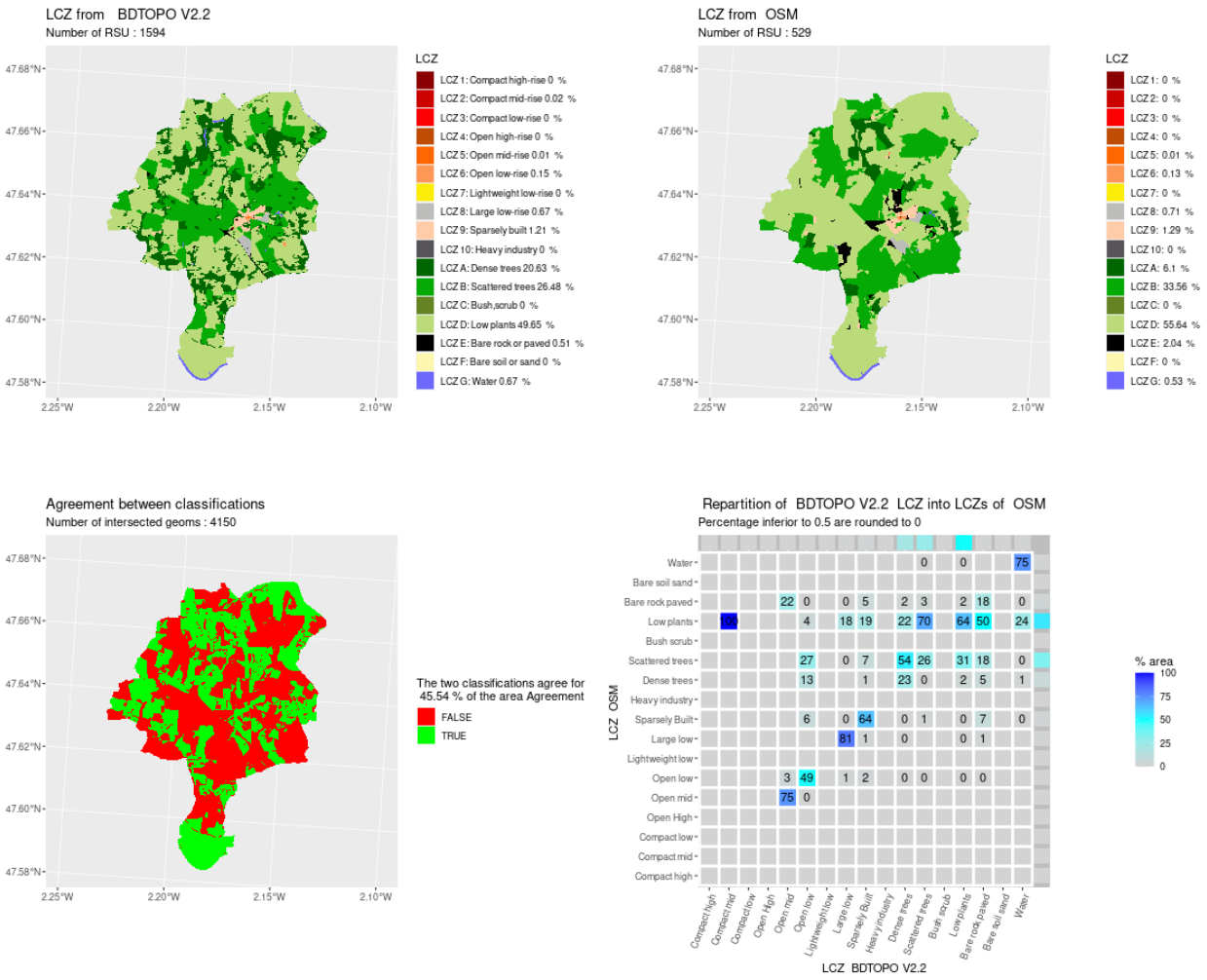


Figure A1. Comparison of LCZ generated for the city of Allaire by the GeoClimate method using BDT and OSM datasets

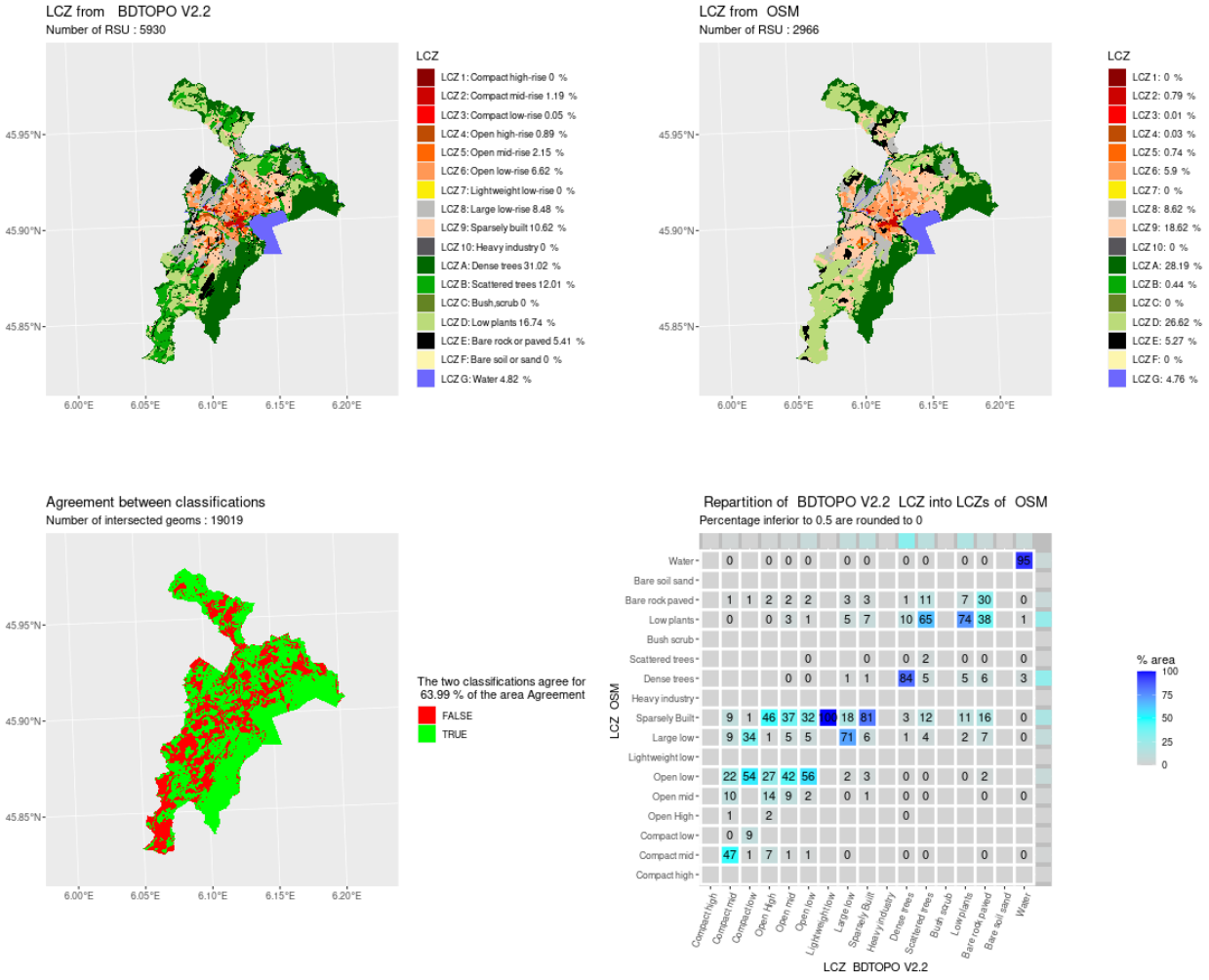


Figure A2. Comparison of LCZ generated for the city of Annecy by the GeoClimate method using BDT and OSM datasets

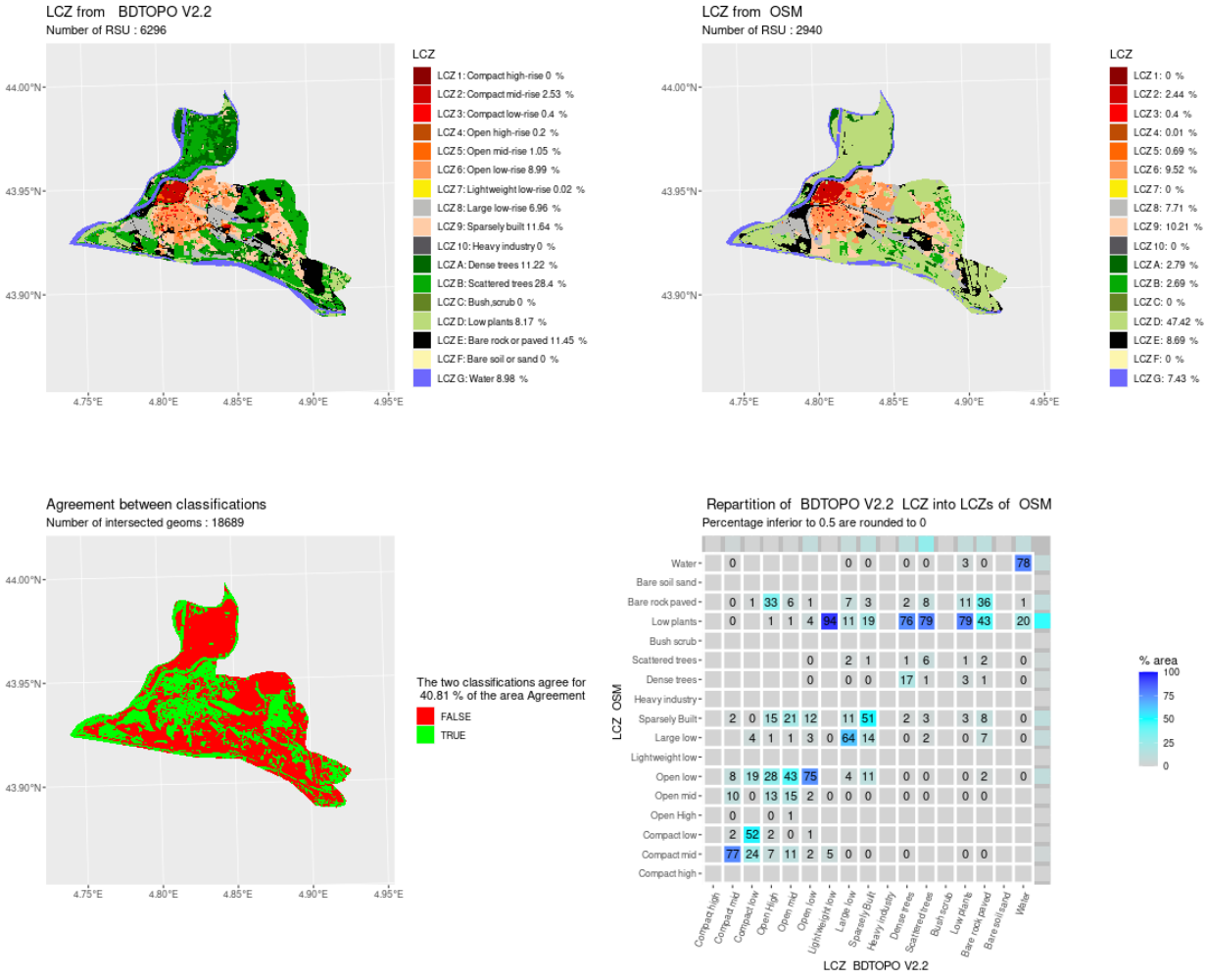


Figure A3. Comparison of LCZ generated for the city of Avignon by the GeoClimate method using BDT and OSM datasets

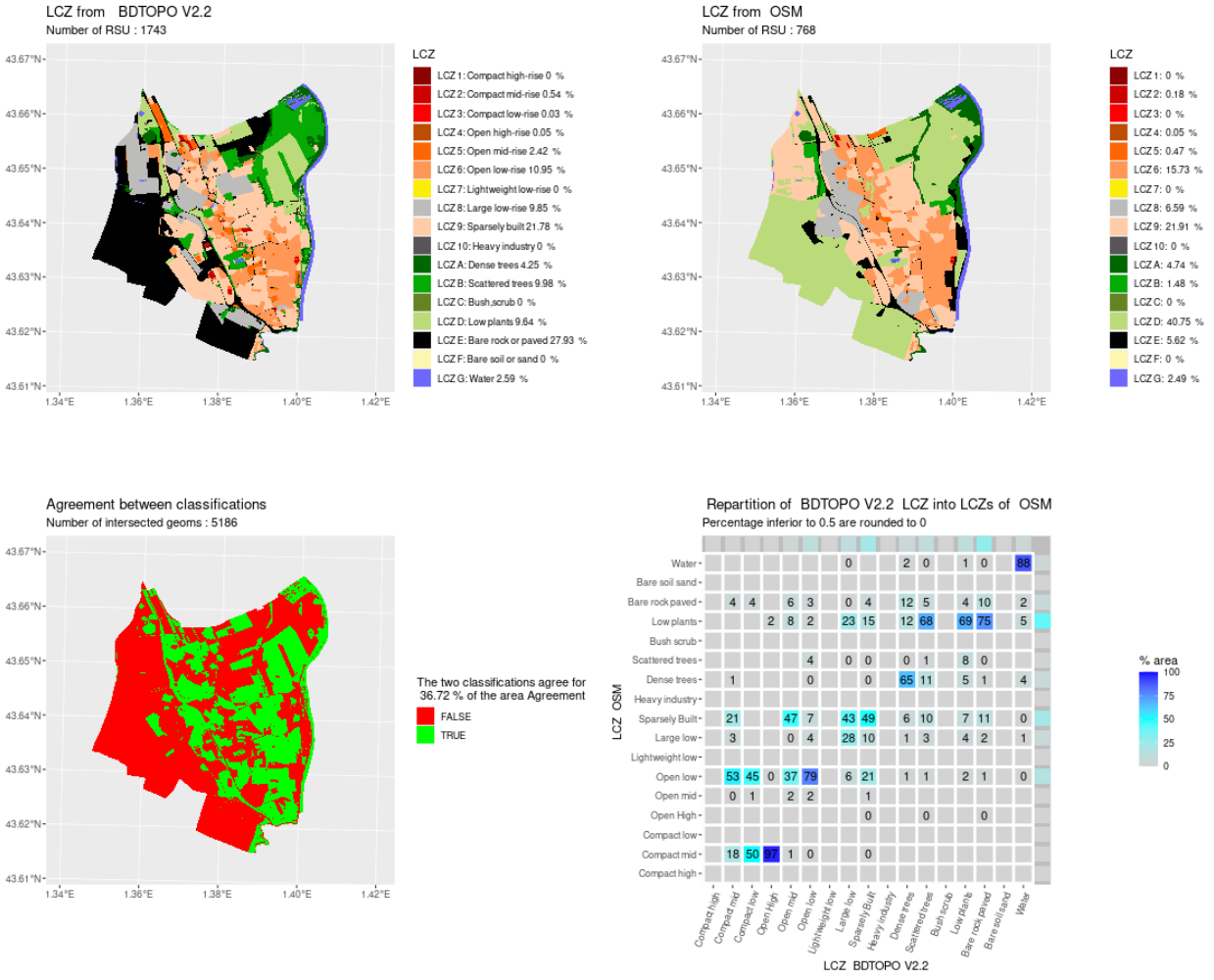


Figure A4. Comparison of LCZ generated for the city of Blagnac by the GeoClimate method using BDT and OSM datasets

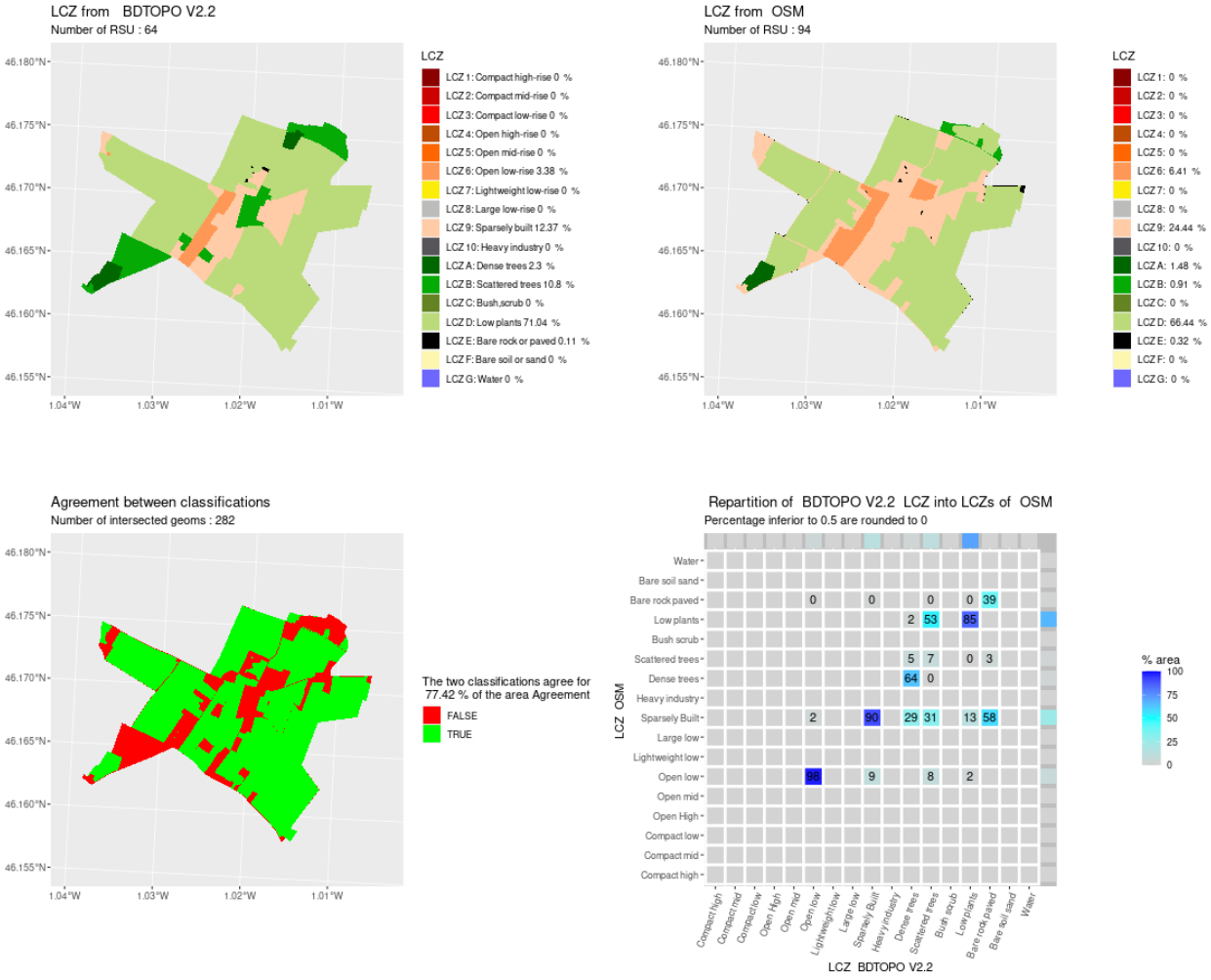


Figure A5. Comparison of LCZ generated for the city of Bourgneuf by the GeoClimate method using BDT and OSM datasets

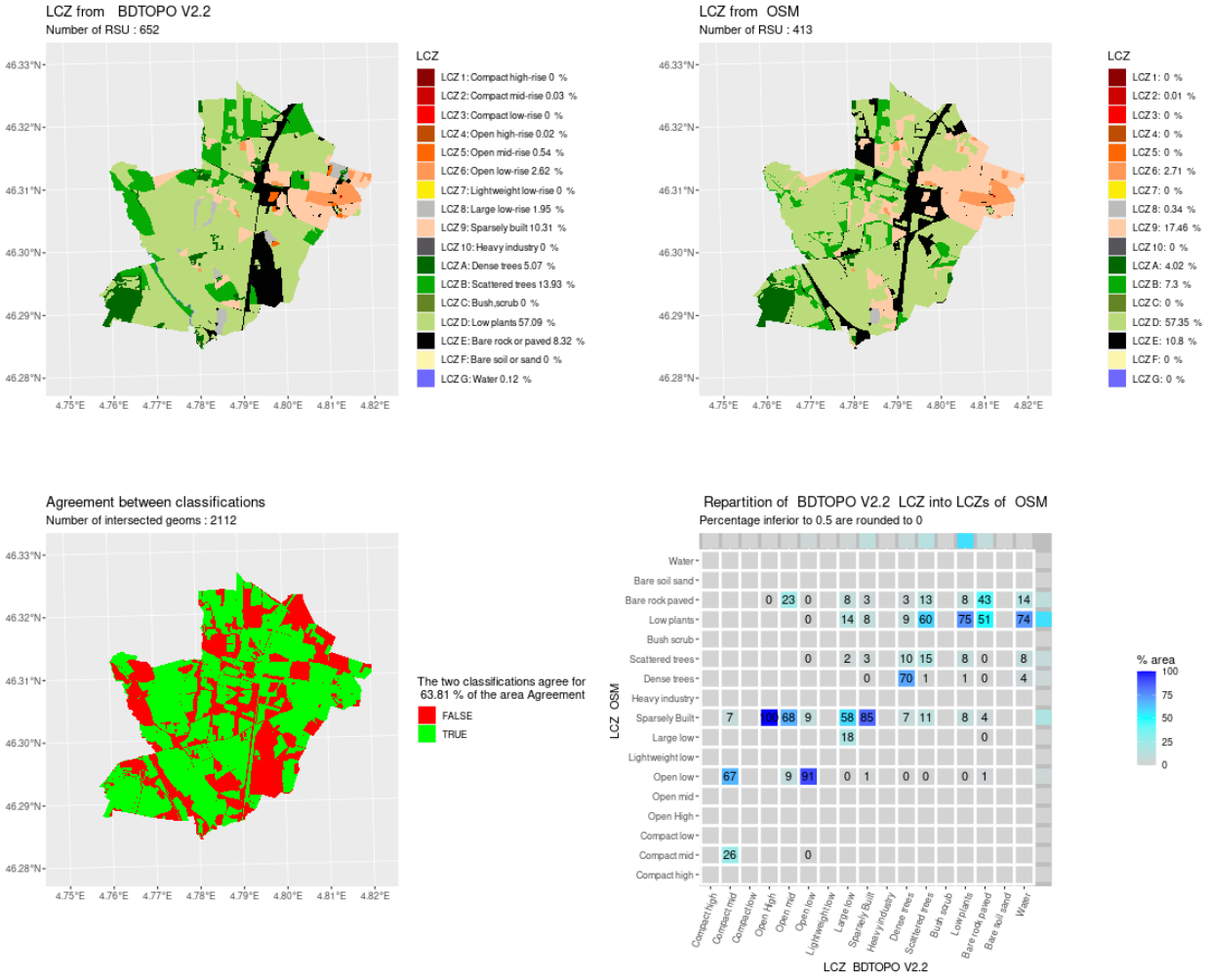


Figure A6. Comparison of LCZ generated for the city of Charnay-Lès-Mâcon by the GeoClimate method using BDT and OSM datasets

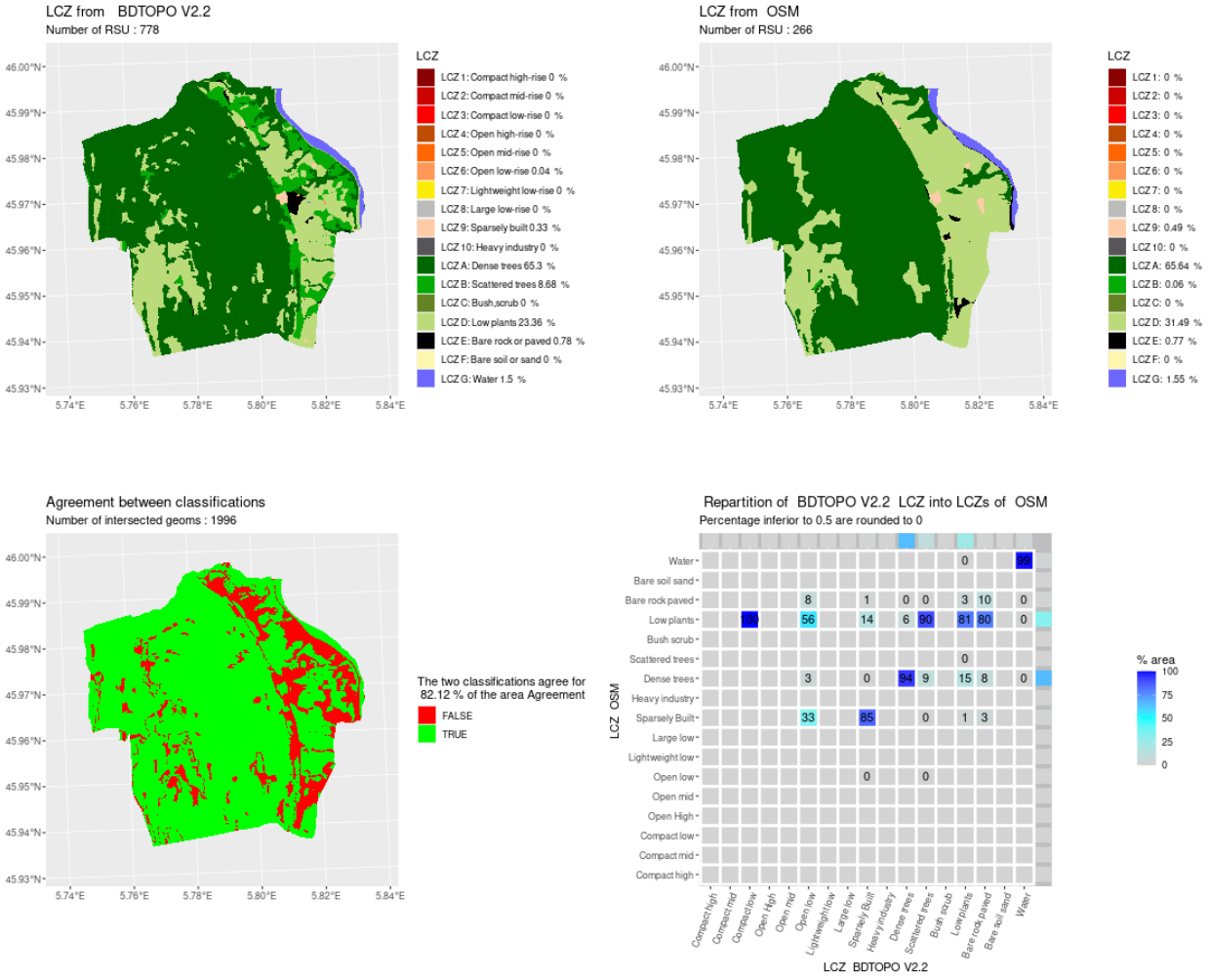


Figure A7. Comparison of LCZ generated for the city of Corbonod by the GeoClimate method using BDT and OSM datasets

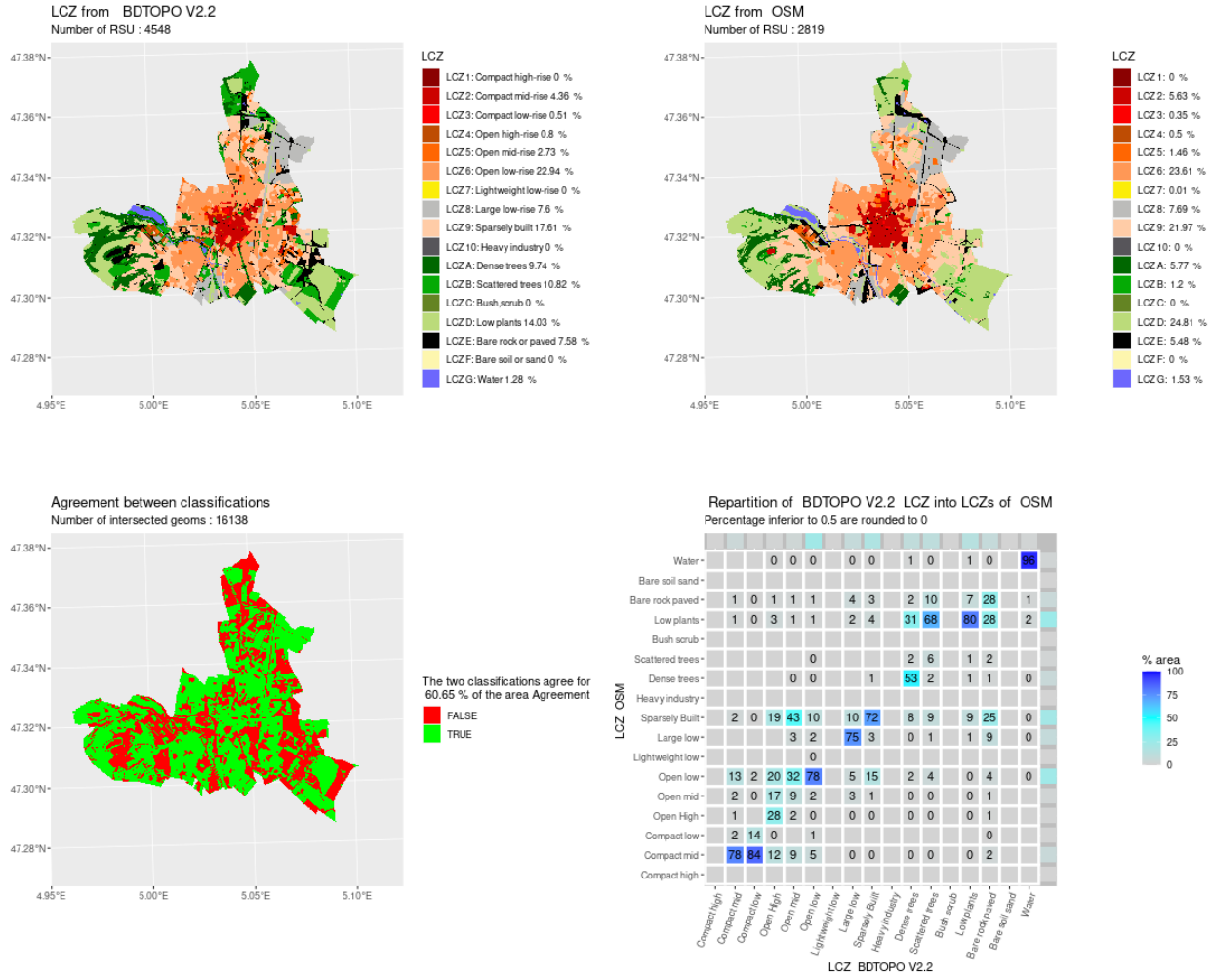


Figure A8. Comparison of LCZ generated for the city of Dijon by the GeoClimate method using BDT and OSM datasets

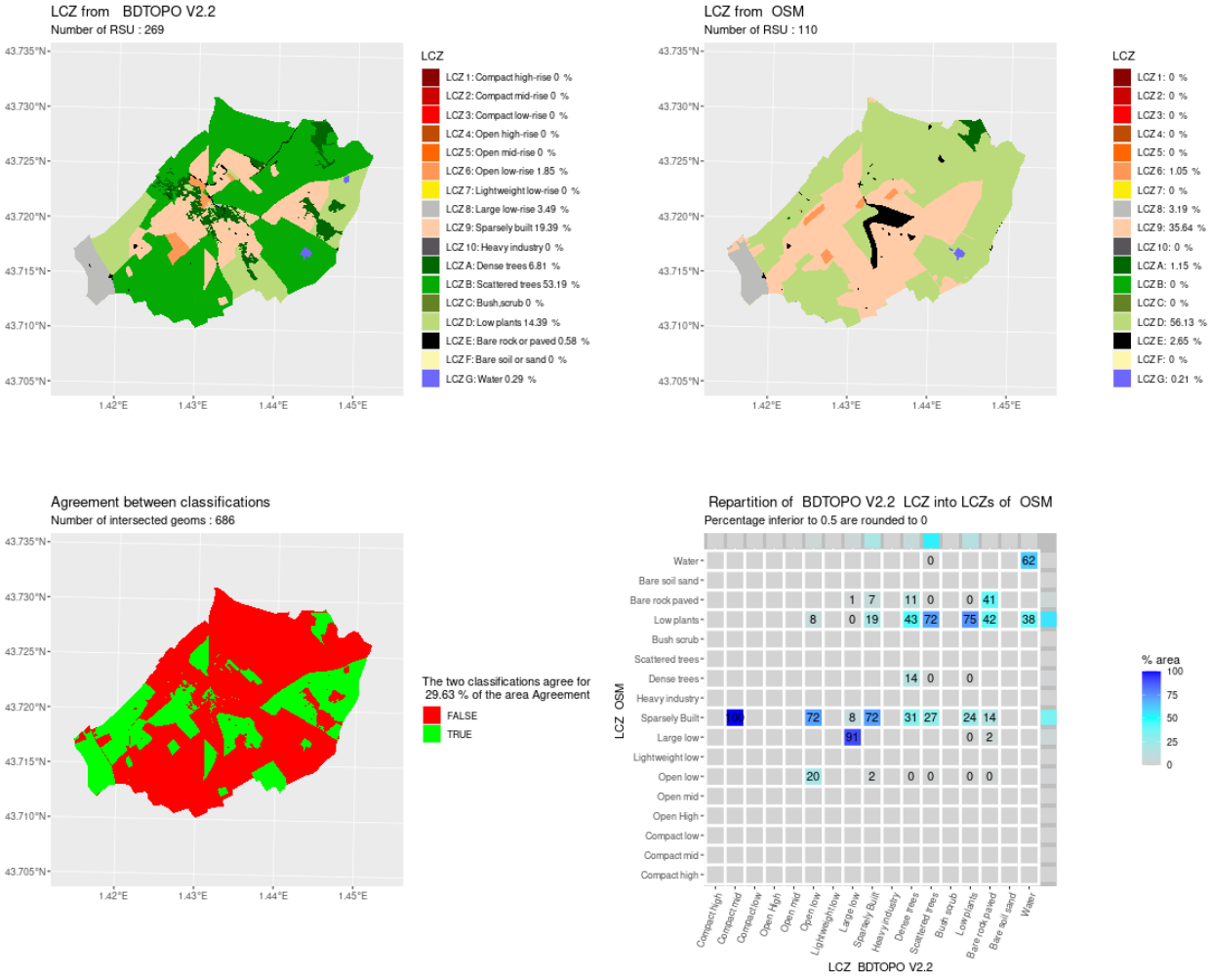


Figure A9. Comparison of LCZ generated for the city of Gratentour by the GeoClimate method using BDT and OSM datasets

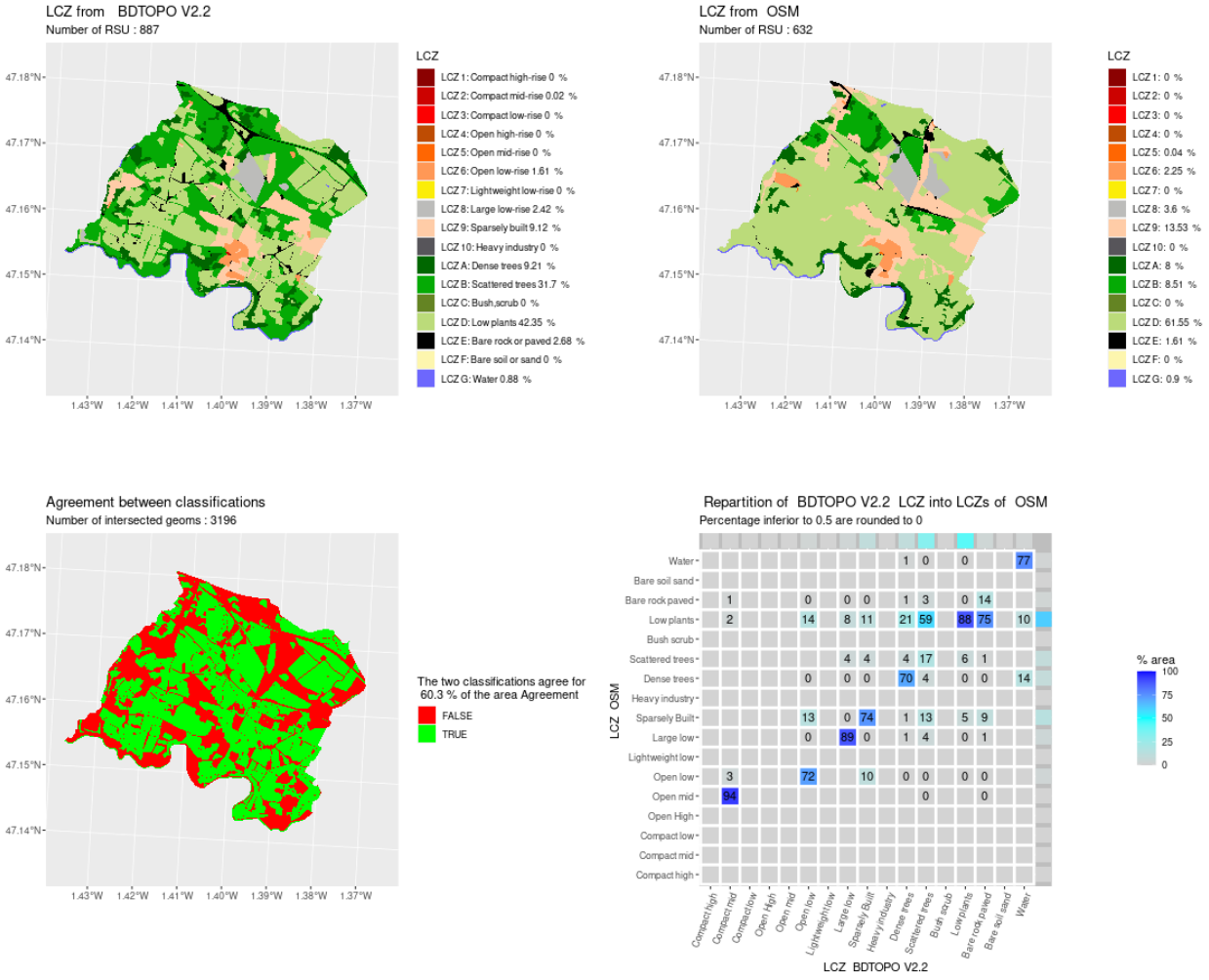


Figure A10. Comparison of LCZ generated for the city of La Haie-Fouassière by the GeoClimate method using BDT and OSM datasets

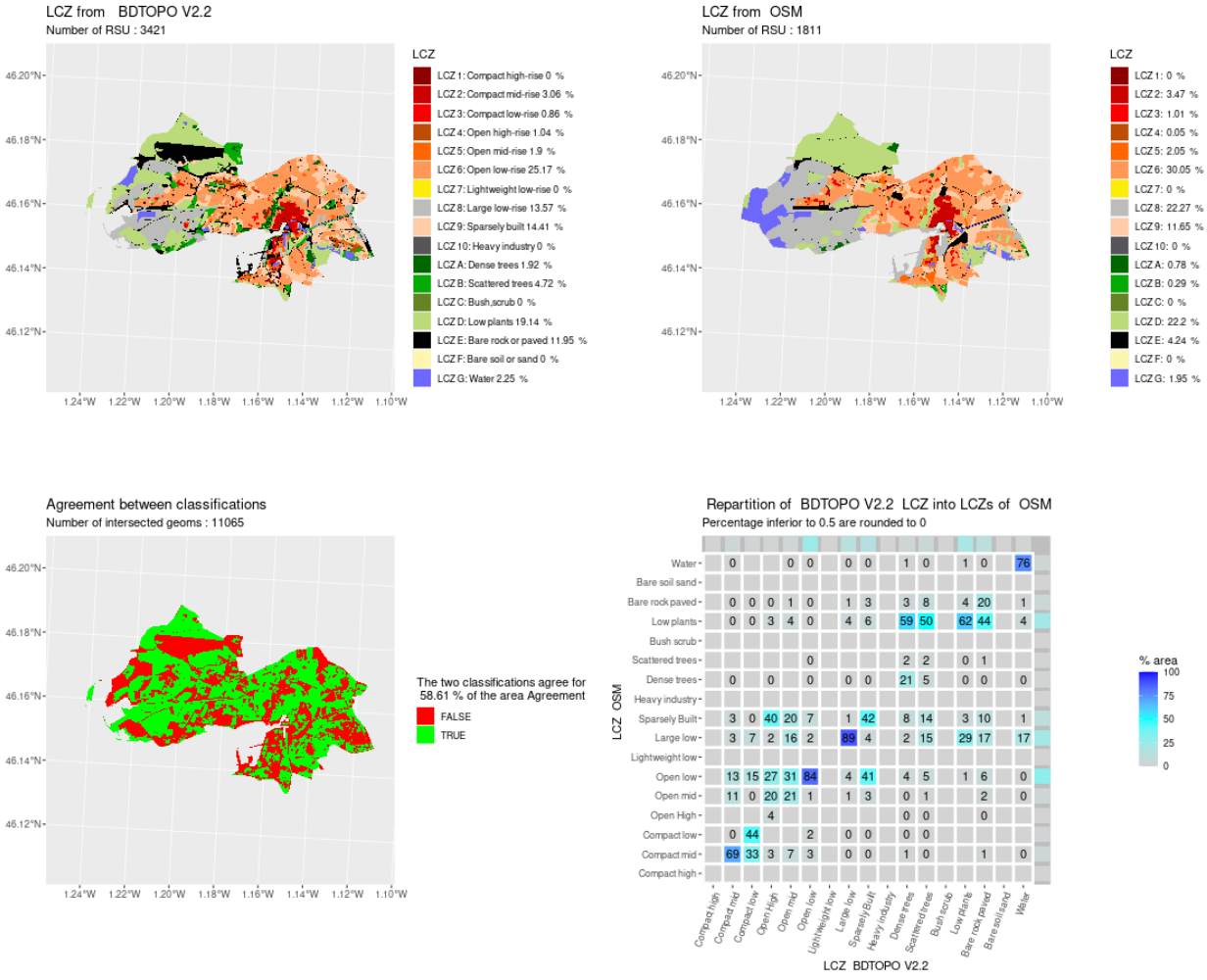


Figure A11. Comparison of LCZ generated for the city of La-Rochelle by the GeoClimate method using BDT and OSM datasets

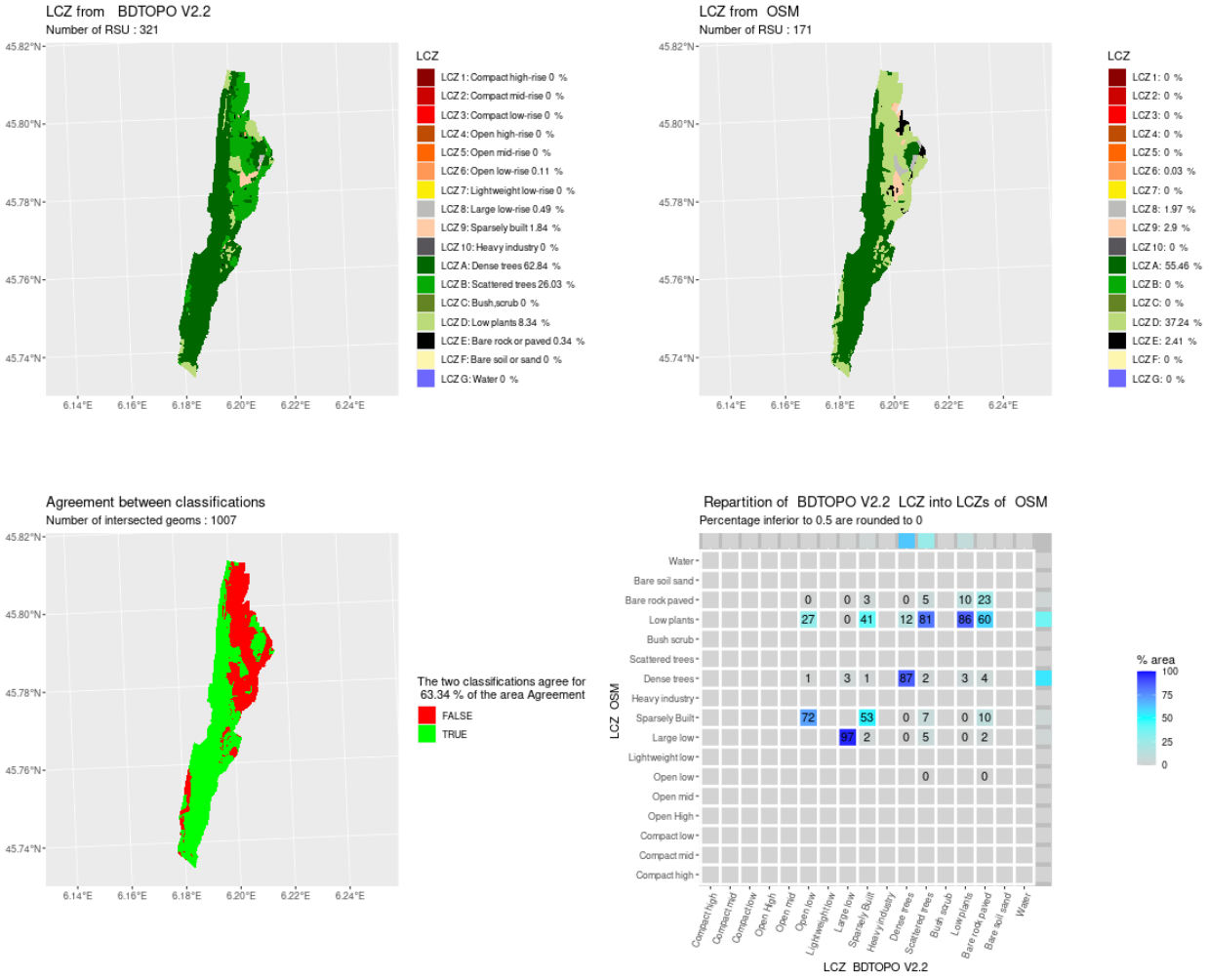


Figure A12. Comparison of LCZ generated for the city of Lathuile by the GeoClimate method using BDT and OSM datasets

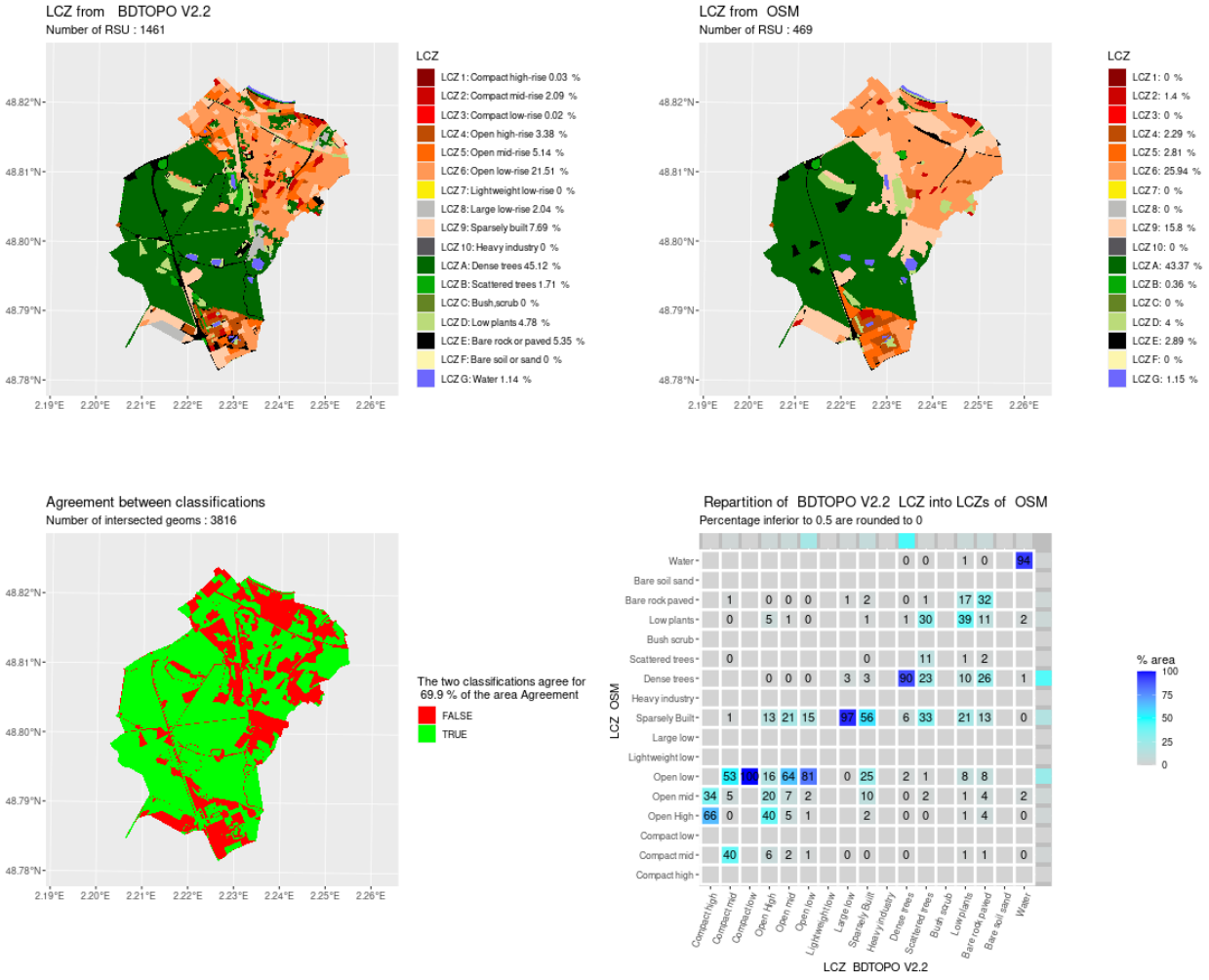


Figure A13. Comparison of LCZ generated for the city of Meudon by the GeoClimate method using BDT and OSM datasets

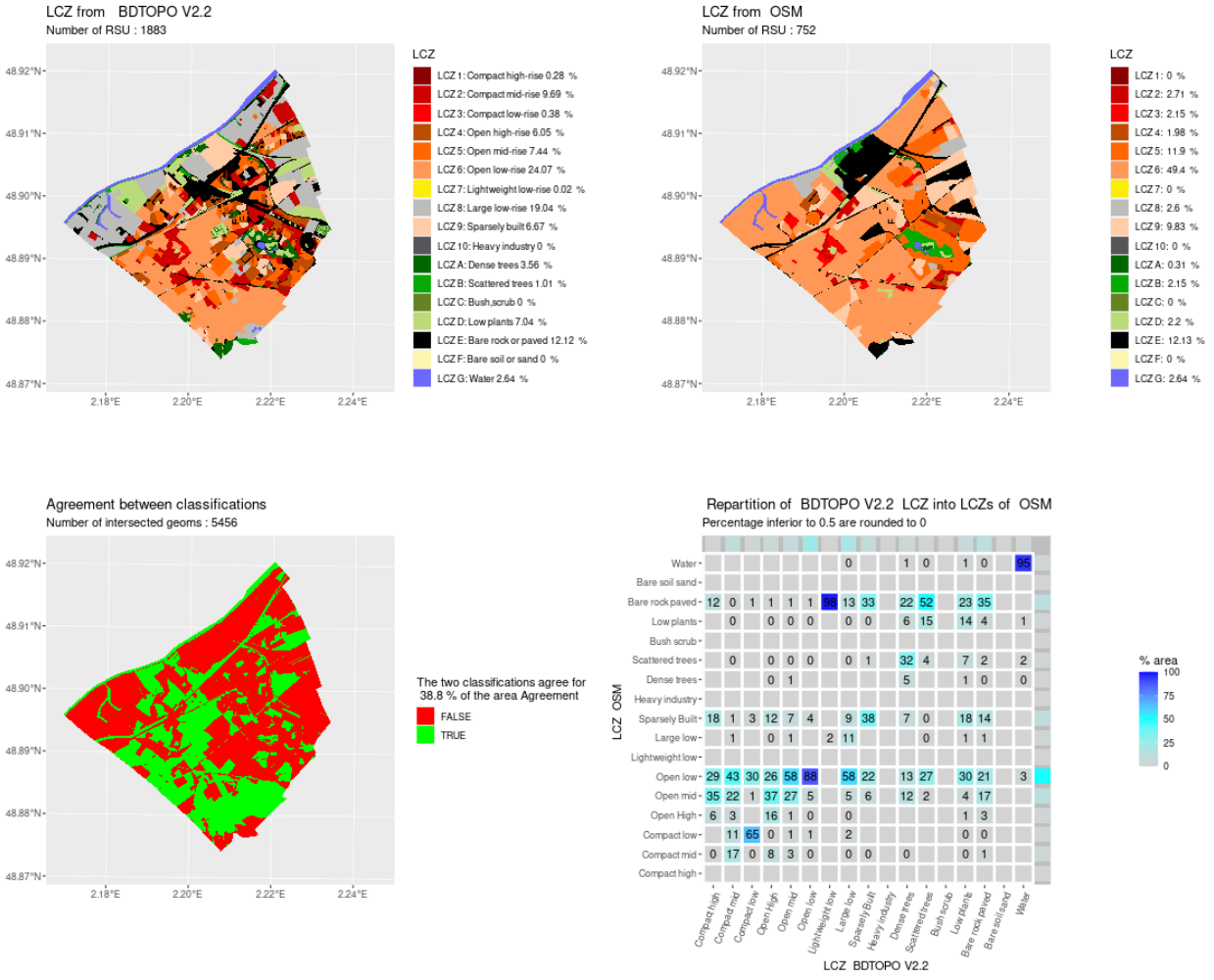


Figure A14. Comparison of LCZ generated for the city of Nanterre by the GeoClimate method using BDT and OSM datasets

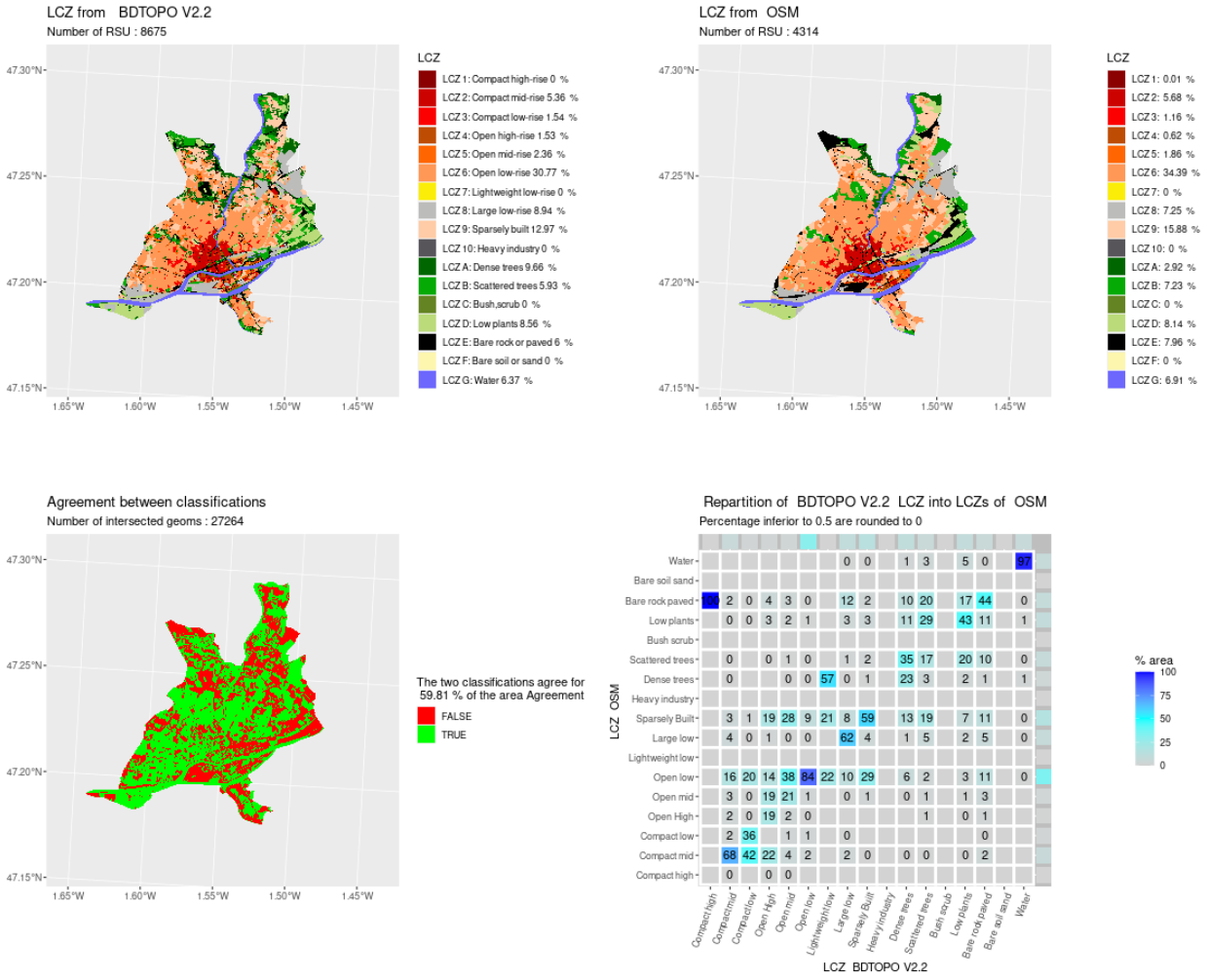


Figure A15. Comparison of LCZ generated for the city of Nantes by the GeoClimate method using BDT and OSM datasets

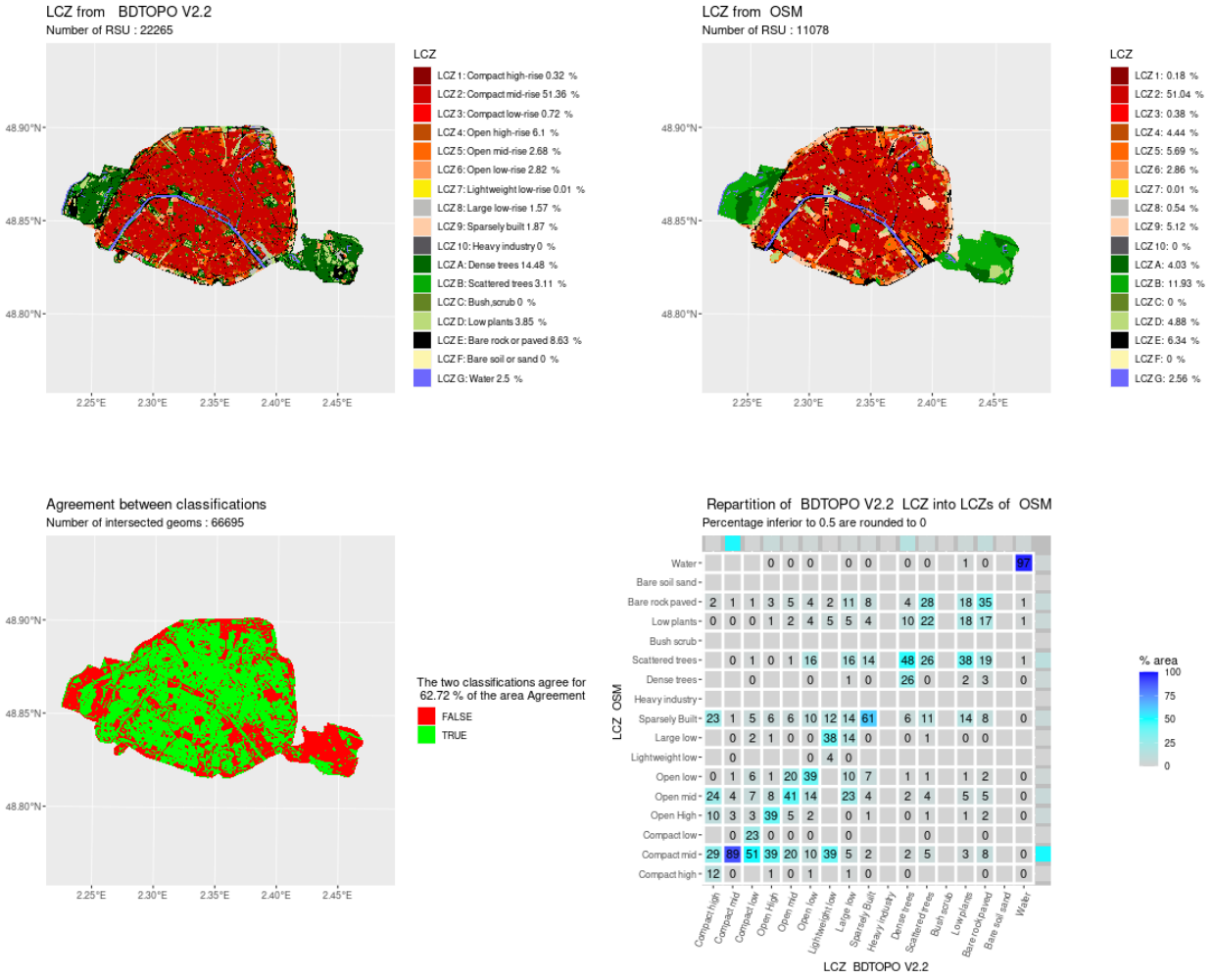


Figure A16. Comparison of LCZ generated for the city of Paris by the GeoClimate method using BDT and OSM datasets

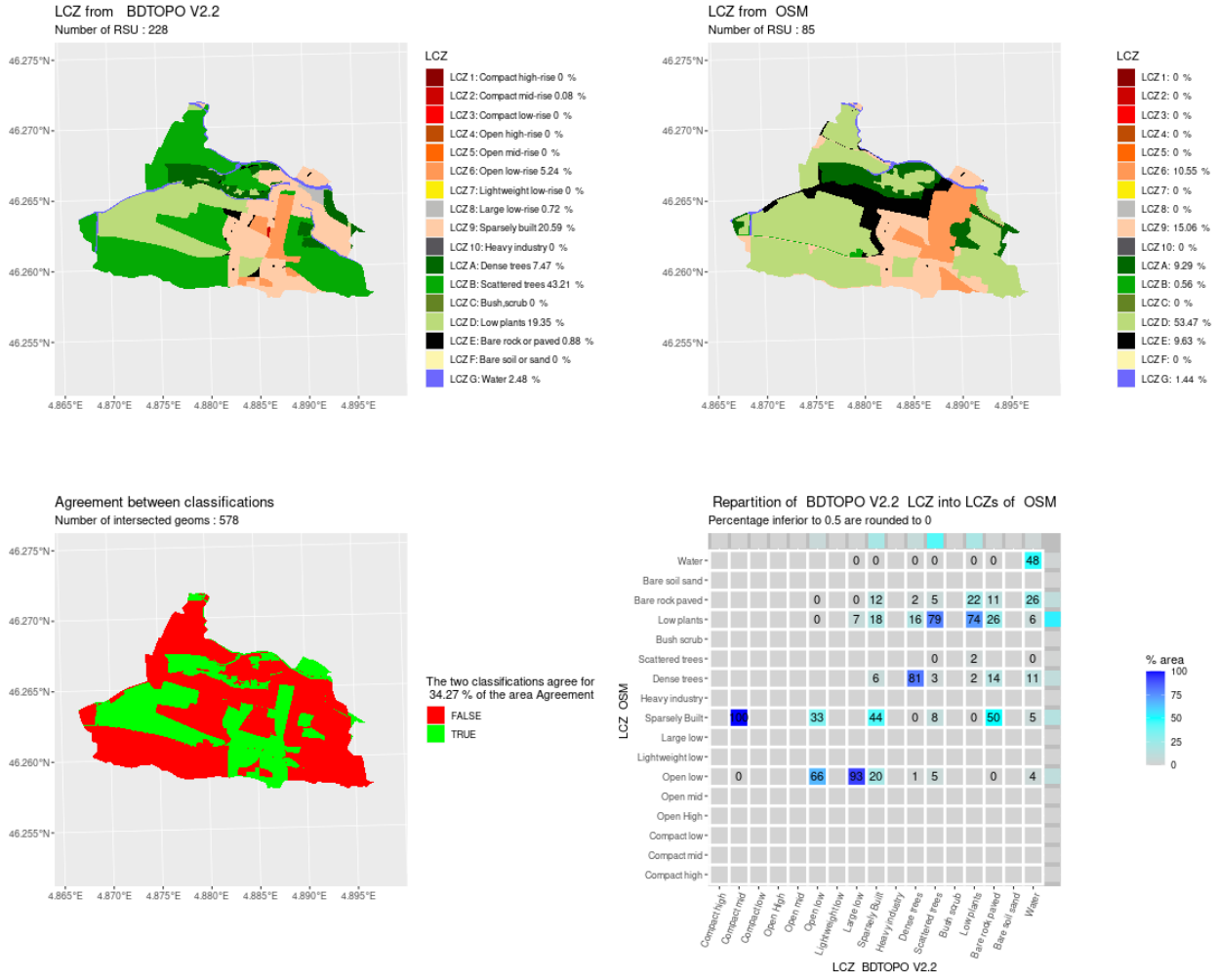


Figure A17. Comparison of LCZ generated for the city of Pont-de-Veyle by the GeoClimate method using BDT and OSM datasets

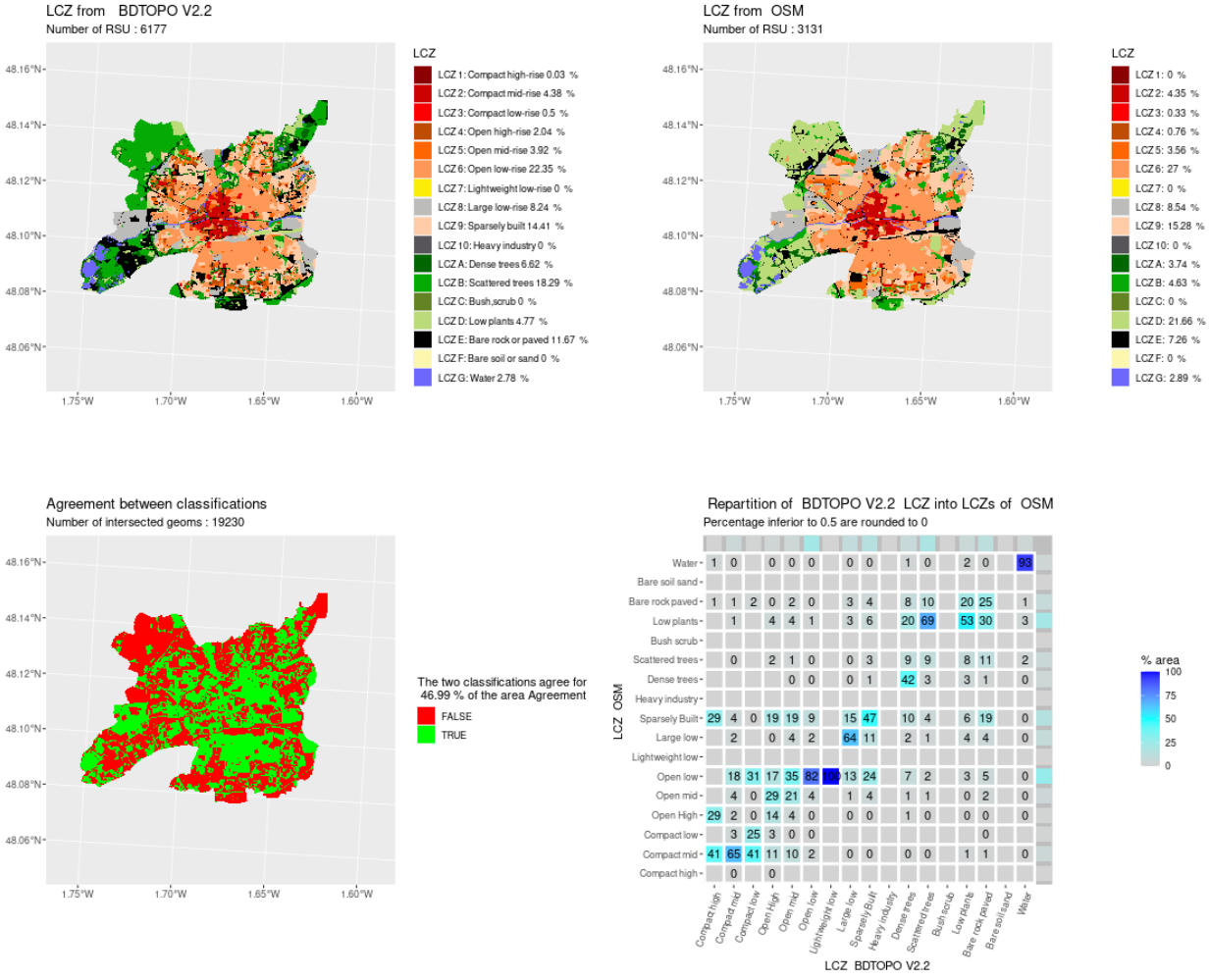


Figure A18. Comparison of LCZ generated for the city of Rennes by the GeoClimate method using BDT and OSM datasets

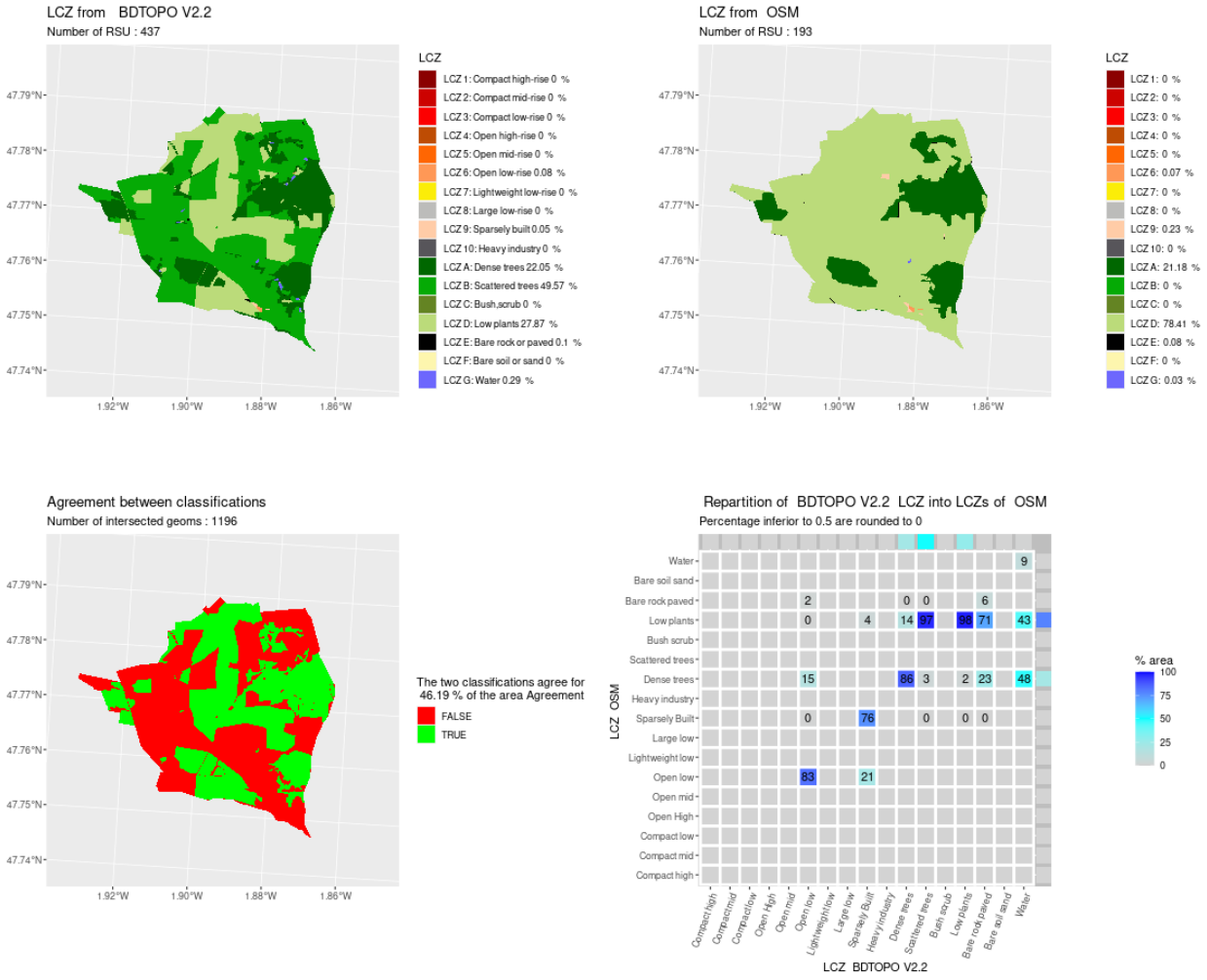


Figure A19. Comparison of LCZ generated for the city of Saint-Ganton by the GeoClimate method using BDT and OSM datasets

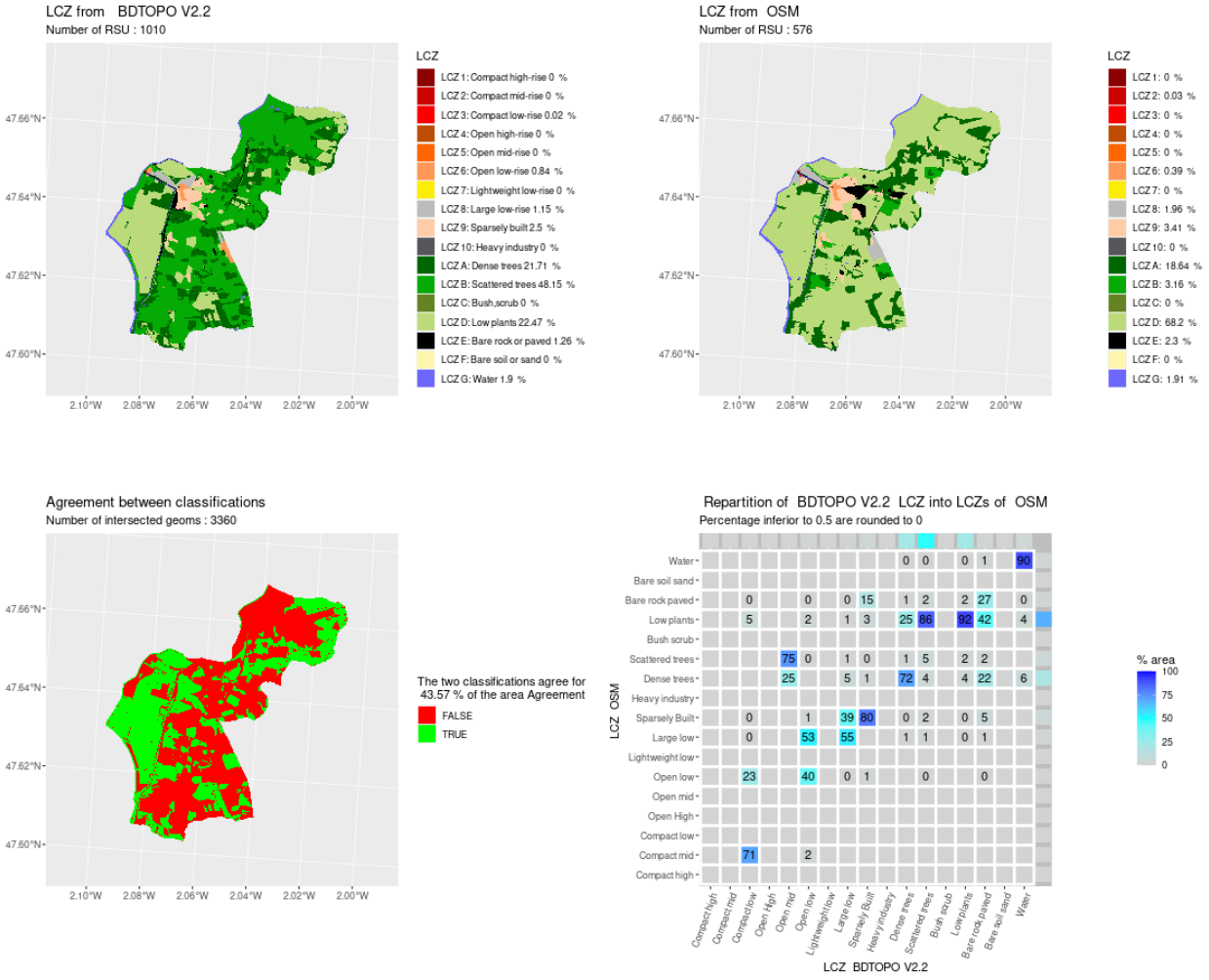


Figure A20. Comparison of LCZ generated for the city of Saint-Nicolas-de-Redon by the GeoClimate method using BDT and OSM datasets

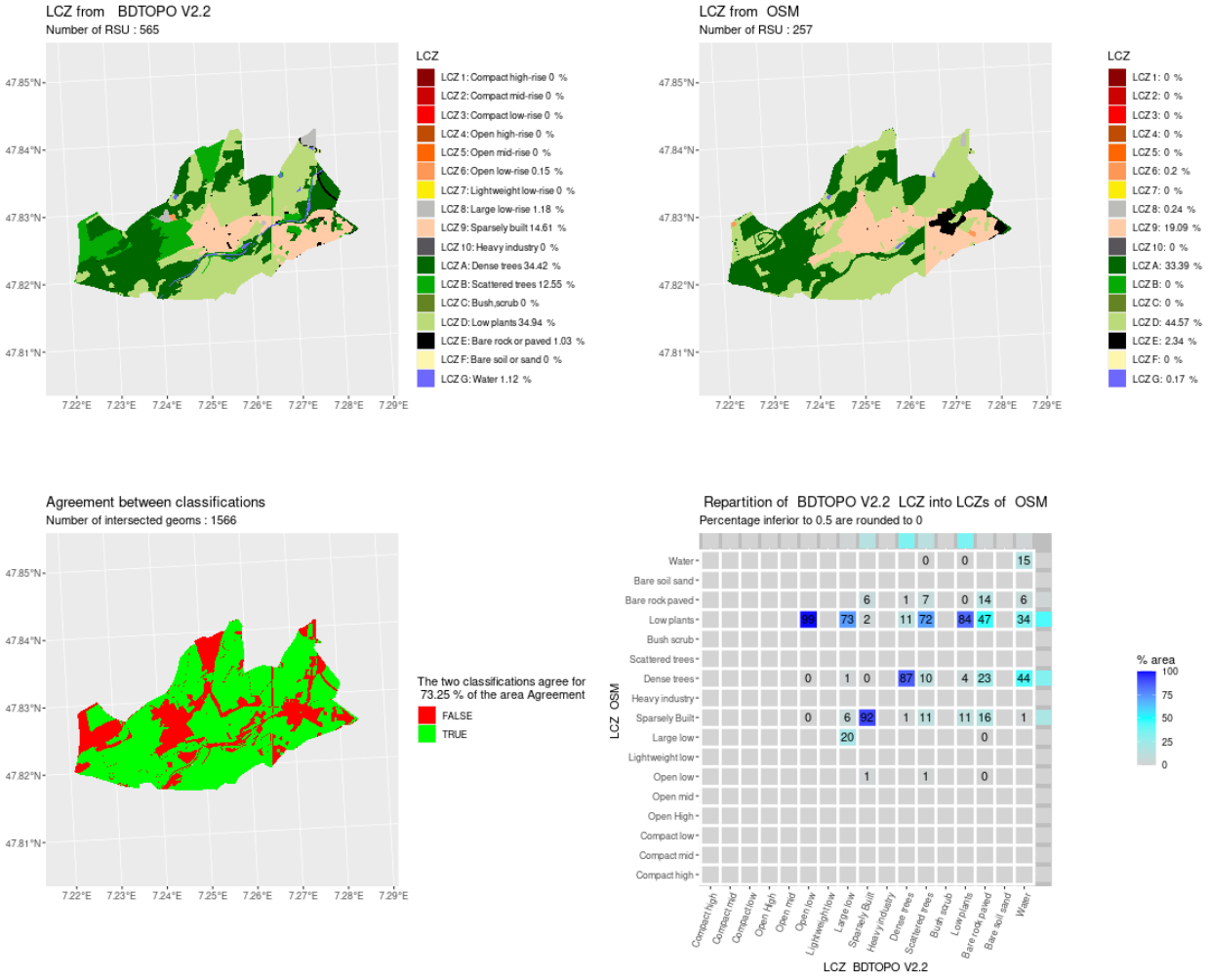


Figure A21. Comparison of LCZ generated for the city of Staffelfelden by the GeoClimate method using BDT and OSM datasets

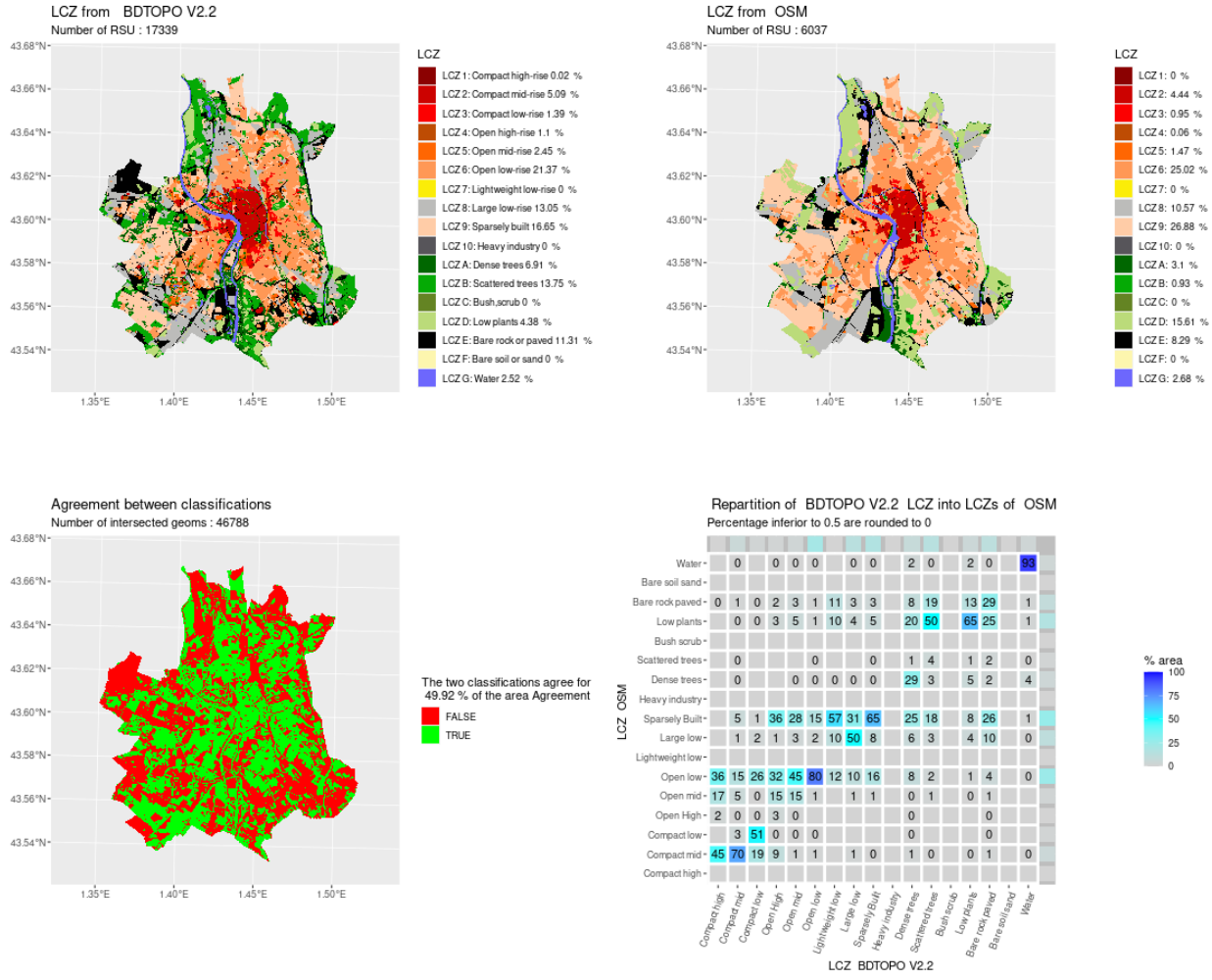


Figure A22. Comparison of LCZ generated for the city of Toulouse by the GeoClimate method using BDT and OSM datasets

Author contributions. Conceptualization: JB, EB, FL / Data curation: JB, EB, MG, FL / Formal analysis: MG, JB, EB, FL / Funding acquisition: EB, JB / Investigation: JB, EB, MG, FL / Methodology: JB, EB / Project administration: JB, EB / Resources: EB / Software: EB, JB, MG, FL, ELS / Supervision: JB, EB / Validation: EB, JB, MG, FL, ELS / Visualization: MG, JB, EB, FL / Writing - original draft preparation: JB, FL, EB / Writing - review and editing: FL, MG, JB, EB

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The method presented in this paper has been integrated in the GeoClimate tool and developed within the following research projects:

- Urb-TWin (2020-2023): European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 896069
- PAENDORA2 (Pour la gestion du confort estival : Données, Outils et Recherche-Action) (2022-2025): funded by the French Agency for Ecological Transition (ADEME) within the PACT^{2}e 2021 research call

References

Baklanov, A., Cárdenas, B., Lee, T.-c., Leroyer, S., Masson, V., Molina, L. T., Müller, T., Ren, C., Vogel, F. R., and Voogt, J. A.: Integrated urban services: Experience from four cities on different continents, Urban Climate, 32, 100 610, 2020.

Bernard, J., Bocher, E., Petit, G., and Palominos, S.: Sky view factor calculation in urban context: computational performance and accuracy analysis of two open and free GIS tools, Climate, 6, 60, 2018.

Bernard, J., Bocher, E., Le Saux Wiederhold, E., Leconte, F., and Masson, V.: Estimation of missing building height in OpenStreetMap data: a French case study using GeoClimate 0.0. 1, Geoscientific Model Development, 15, 7505–7532, 2022.

Bocher, E., Bernard, J., Wiederhold, E. L. S., Leconte, F., Petit, G., Palominos, S., and Noûs, C.: GeoClimate: a Geospatial processing toolbox for environmental and climate studies, Journal of Open Source Software, 6, 3541, 2021.

Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X., Ren, C., Brousse, O., Martilli, A., Neophytou, M., et al.: WUDAPT: An urban weather, climate, and environmental modeling infrastructure for the anthropocene, Bulletin of the American Meteorological Society, 99, 1907–1924, 2018.

Davenport, A. G., Grimmond, C. S. B., Oke, T. R., and Wieringa, J.: Estimating the roughness of cities and sheltered country, in: Preprints, 12th Conf. on Applied Climatology, Asheville, NC, Amer. Meteor. Soc, vol. 96, p. 99, 2000.

Demuzere, M., Bechtel, B., Middel, A., and Mills, G.: Mapping Europe into local climate zones, PloS one, 14, e0214 474, 2019.

Demuzere, M., Hankey, S., Mills, G., Zhang, W., Lu, T., and Bechtel, B.: Combining expert and crowd-sourced training data to map urban form and functions for the continental US, Scientific data, 7, 264, 2020.

Demuzere, M., Kittner, J., and Bechtel, B.: LCZ Generator: a web application to create Local Climate Zone maps, Frontiers in Environmental Science, 9, 637 455, 2021.

Geletič, J., Lehnert, M., and Dobrovolnỳ, P.: Land surface temperature differences within local climate zones, based on two central European cities, Remote Sensing, 8, 788, 2016.

Grimmond, S., Bouchet, V., Molina, L. T., Baklanov, A., Tan, J., Schlünzen, K. H., Mills, G., Golding, B., Masson, V., Ren, C., et al.: Integrated urban hydrometeorological, climate and environmental services: Concept, methodology and key messages, Urban Climate, 33, 100 623, 2020.

Hanna, S. R. and Britter, R. E.: Wind flow and vapor cloud dispersion at industrial and urban sites, John Wiley & Sons, 2010.

IPCC, C. C. et al.: The physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996, 113–119, 2007.

Masson, V., Heldens, W., Bocher, E., Bonhomme, M., Bucher, B., Burmeister, C., de Munck, C., Esch, T., Hidalgo, J., Kanani-Sühring, F., et al.: City-descriptive input data for urban climate models: Model requirements, data sources and challenges, Urban Climate, 31, 100 536, 2020.

Quan, S. J. and Bansal, P.: A systematic review of GIS-based local climate zone mapping studies, Building and Environment, 196, 107 791, 2021.

Quan, S. J., Dutt, F., Woodworth, E., Yamagata, Y., and Yang, P. P.-J.: Local climate zone mapping for energy resilience: a fine-grained and 3D approach, Energy Procedia, 105, 3777–3783, 2017.

Skarbit, N., Stewart, I. D., Unger, J., and Gál, T.: Employing an urban meteorological network to monitor air temperature conditions in the 'local climate zones' of Szeged, Hungary, International Journal of Climatology, 37, 582–596, 2017.

Stewart, I. D.: Redefining the urban heat island, Ph.D. thesis, University of British Columbia, 2011.

435 Stewart, I. D. and Oke, T. R.: Local climate zones for urban temperature studies, Bulletin of the American Meteorological Society, 93, 1879–1900, 2012.