

Universal patterns of intra-urban morphology: Defining a global typology of the urban fabric using unsupervised clustering

Henri Debray^{a,b,c,*}, Matthias Gassilloud^d, Richard Lemoine-Rodríguez^{b,e}, Michael Wurm^b, Xiaoxiang Zhu^{c,f}, Hannes Taubenböck^{a,b}

^a Institute of Geography and Geology, Department of Global Urbanization and Remote Sensing, University of Würzburg 97074 Würzburg, Germany

^b German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany

^c Data Science in Earth Observation, Technical University of Munich (TUM), 80333 Munich, Germany

^d Chair of Remote Sensing and Landscape Information Systems, University of Freiburg 79106 Freiburg, Germany

^e Geolinguistic Studies Team, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

^f Munich Center for Machine Learning, 80333 Munich, Germany

ABSTRACT

The physical dimension of cities and its spatial patterns play a crucial role in shaping society and urban dynamics. Understanding the complexity of urban systems requires a detailed assessment of their physical structure. Urban geography has long focused on framing typologies to represent common patterns in the urban fabric using various methodologies. However, only recent advancements in computational methods and global land cover data have enabled to comprehensively identify typologies of urban patterns at the city scale through new unsupervised approaches. Nevertheless, typologies of finer-grained patterns at intra-urban scale have not yet been explored comprehensively at a global level. In this paper, building upon these advances, we explore the intra-urban patterns of more than 1500 cities across the globe. We rely on a Local Climate Zone land cover classification to represent the multidimensional variabilities of intra-urban morphology. Adapting a deep learning based unsupervised clustering approach, we find a typology of 138 intra-urban patterns. Analyzing the results of this data-driven approach, we prove that each pattern identified is unique, i.e. statistically different, in its composition and configuration. With this study summarizing the global diversity of the urban fabric, we reveal that any city of the world can be described as a specific assemblage of a fraction of these 138 universal patterns. These universal patterns reveal a predominance at a global scale of built-up forms of low density in the intra-urban fabric.

1. Introduction

From one street to the next, from one neighborhood to the next, the urban tapestry endlessly changes in an almost seamless manner, continuously presenting ever-new spectacles of arrangements of shapes, weaving of volumes, materials and more at each turn of a corner. Such is the remarkable morphological diversity that a single city can exude on its own.

Now, looking beyond the confines of one city, we see that this morphological diversity does intensify manifold. Beholding a neighboring city, a city in another country, or even a city in another continent, chances are that the urban fabric of this other city could then feel astonishingly alien and singular. So much so that, paradoxically, the fabric of one city in comparison to another one, would show clear signs of sameness in contrast to its internal diversity.

The segments of the urban fabric in these two cities, although rich of

their own internal diversities, once juxtaposed, have this own internal diversity sometime fade in light of the differences between them. If we then apply this juxtaposition to the pieces of the urban fabric of many cities, between their internal and external diversities what should be expected? Beneath all the imaginable tweaks of the fabric, big and small, making pieces of urban fabric seem similar or different, would we observe re-occurring patterns being revealed? And if so, could we capture the essences of the templates that underlie these re-occurring patterns? In every city, all at once? In other words, we pose here the question of whether we can draw a universal typology of the urban fabric.

Despite the seemingly purely theoretical character of these introductory questions, we see this at the core of very pragmatic issues in the field of urban planning. For instance, over the last decades, the world itself has become impressively more urban (UN, 2018). Thus, the study of urban phenomena, be they economic, environmental, political or

* Corresponding author at: Institute of Geography and Geology, Department of Global Urbanization and Remote Sensing, University of Würzburg, 97074 Würzburg, Germany.

E-mail addresses: henri.debray@uni-wuerzburg.de (H. Debray), matthias.gassilloud@felis.uni-freiburg.de (M. Gassilloud), richard.lemoine-rodriguez@uni-wuerzburg.de (R. Lemoine-Rodríguez), michael.wurm@dlr.de (M. Wurm), xiaoxiang.zhu@tum.de (X. Zhu), hannes.taubenboeck@dlr.de (H. Taubenböck).

<https://doi.org/10.1016/j.jag.2025.104610>

Received 19 January 2025; Received in revised form 23 April 2025; Accepted 18 May 2025

Available online 29 May 2025

1569-8432/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

social, becomes increasingly relevant. Correspondingly, the relevance of urban morphology rises as the field of research studying the spatial expression of cities, i.e. the backdrop and backbone structuring and shaping these urban phenomena (Batty, 2009; Kostof, 1991; Lefebvre, 1974; Lynch, 1992; Wentz et al., 2018). Ultimately, there is a rising need to explore, catalog and study the intra-urban patterns of the fabric of cities at a global scale to develop the understanding of fine-grained phenomena of the main human habitat: cities.

While this need has recently become more pressing, the field of urban morphology has not waited for recent times to concern itself with this topic. Since the 19th century (Andrews, 1993; Bertyák, 2021; Durand, 1805, 1825) and more prominently hereafter over the last 60 years (Alexander et al., 1977; Bertyák, 2021; Castex et al., 1997; Cataldi et al., 2002; Fleischmann et al., 2021a; Kostof, 1991, 1992; Lynch, 1992; Muratori, 1960; Panerai et al., 1999; Taubenböck et al., 2018) a large fraction of the research on urban morphology was focused on finding common patterns in the urban fabric and to derive typologies.

Over the last decade, to meet the need for systematic knowledge about the spatial dimension of cities, the field of urban morphology has been operating a shift towards more quantitative techniques of analysis, as identified in (Bertyák, 2021; Dibble et al., 2017; Fleischmann et al., 2021b). This shift was vastly supported and embodied by the production of and access to: 1) increasing volume and coverage of systematic geographical datasets, e.g.: (Angel et al., 2016; Arsanjani et al., 2015; Boeing, 2017; Demuzere et al., 2022; Herfort et al., 2021; Microsoft, 2020; Zhu et al., 2021, 2022) and 2) scalable computational tools (Biljecki and Chow, 2022; Boeing, 2017; Bosch, 2019; Fleischmann, 2019) combined into adequate methodologies tackling diverse geographical data types (Arribas-Bel and Fleischmann, 2022; Boeing, 2019; Fleischmann et al., 2021a; Fleischmann and Arribas-Bel, 2022; Pont and Olsson, 2018; Taubenböck et al., 2020; Wang et al., 2023). These two aspects of the new quantitative turn of the field – massive quantitative data and scalable tools to leverage them – compounded neatly to open new horizons for the development of typologies of the urban fabric.

The access to quantitative data permitted the field to notably turn away from methodologies centered on the almost *ex-ante* creation of typologies, e.g.: (Alexander et al., 1977; Braunfels, 1976; L. Krier, 1998; R. Krier and Rowe, 1991; Panerai et al., 1999) which tend to focus on the more noteworthy historical types and, as a result, forego unspectacular, ordinary, or culturally less important types (Bertyák, 2021; Kostof, 1991). While these precursor works laid the foundation of the field, this approach was prone to potential culturally biased preconception of what types exist or are worthy of being described and what their characteristics should be. Against these shortcomings, the quantitative turn of the field translated itself into more data-driven empirical approaches, and notably, toward unsupervised clustering methodologies, e.g.: (Boeing, 2019; Fleischmann et al., 2021a; Gil et al., 2012; Lemoine-Rodríguez et al., 2020; Pont and Olsson, 2018; Taubenböck et al., 2020). Such methodologies, indeed, focus on finding clusters (i.e. groups, or categories) of data samples from a dataset sharing similar traits, without having to rely on prior constraints of which samples should be grouped together or what the expected categories should be. The field of urban morphology, until now was primarily using traditional statistical techniques such as k-means (e.g.: (Taubenböck et al., 2020)), or Ward hierarchical clustering (e.g.: (Boeing, 2019; Fleischmann et al., 2021a)). While these methods are well suited for these tasks, they scale poorly with the amount of data or of dimensionality (de Amorim, 2015; Jin and Han, 2010). In the meantime, deep learning applied to computer vision have now seen the development of more scalable methods of high accuracy performance on popular computer vision benchmarks. (Van Gansbeke et al., 2020), for example developed the SCAN method that reached an accuracy of 88.3 % on the benchmark dataset CIFAR-10 (Krizhevsky, 2009) in an unsupervised setting. Following, (S. Park et al., 2021) developed the SCAN + RUC framework that reached 90.1 % in the same context. These two approaches have now been recently

bested by the TURTLE approach (Gadetsky et al., 2024) with a 99.5 % accuracy. While research applying the latter type of methods in the field of urban morphology remains rare, forays in this research directions are already promising (e.g.: (Wang et al., 2024)).

Although these purely data-driven, unsupervised clustering methodologies bypass the mentioned pitfalls of more classical approaches, they often do not suffice in themselves. For this reason, they are often complemented by further analysis based on expert knowledge for interpreting the clustering results and framing a typology of cities, for which the clusters are representative. Combining in this way a data-driven approach and expert knowledge has been identified to enable the formulation of better informed and more objective typologies (Bertyák, 2021; Fleischmann et al., 2021a).

In recent years, a critical mass of diverse data (e.g.: vector, land-use/land-cover classification) concerning urban morphology was reached, such that they present global coverage (Herfort et al., 2021; Microsoft, 2020; Zhou et al., 2022; Zhu et al., 2022). Their global coverage and their *completeness* (in the data sense (Dempster et al., 1977; Eiben et al., 2021; Rubin, 1976)) now enable to develop data-driven global and as complete as possible typologies. While any typology, even on fragmentary empirical data is informative, a complete (to the extent of current capacities) and global typology would bring new grounds for more *universal* comparisons across the globe and across types.

Managing to leverage the global dimension of these new datasets the field of urban morphology produced a few global typologies of cities based on their urban fabric, e.g.: (Boeing, 2019; Chakraborty et al., 2024; W. Chen et al., 2024; Lemoine-Rodríguez et al., 2020; Taubenböck et al., 2020; Zhou et al., 2022). We see these studies as pioneering work in the use of global datasets on urban fabric and deriving typologies at a macro-scale (Wu et al., 2025).

Literature shows that several urban issues have direct impacts at the urban *meso*-scale (Wu et al., 2025), for example urban heat islands (Lemoine-Rodríguez et al., 2022), air pollution (Zeng et al., 2024), noise pollution (Staab et al., 2023), traffic (Ulvi et al., 2024), walkability (Droin et al., 2024), social segregation (Tammaru et al., 2021), among others. This emphasizes the need of global research at this scale to support urban planning. Yet, to the best of our knowledge, no studies have proposed global and complete (i.e. universal) typology of intra-urban patterns at the *meso*-scale. To help bridge this gap, we aim in this study at developing a data-driven methodology for framing a universal typology of the intra-urban fabric.

Towards this end, we propose in this study to leverage the recent advances in term of comprehensive dataset by using a globally consistent land cover classification (Zhu et al., 2022) that relies on the Local Climate Zones scheme developed by (Stewart and Oke, 2012), as a fine-grained proxy of the intra-urban landscape (see section Data). We apply the best accuracy yielding unsupervised deep-learning framework (at the time of this study) combining the Semantic Clustering by Adopting Nearest neighbors (SCAN) framework (Van Gansbeke et al., 2020) and the Robust learning for Unsupervised Clustering (RUC) framework (S. Park et al., 2021) to find the existing clusters of intra-urban patterns in a data-driven way (see section Methodology). With this, we identify clusters with unique composition and configuration characteristics (see section Results) capturing in a consistent way the universal elementary bricks constituting each and every city of our world.

2. Study design

In this study, we propose to frame a typology of the internal elements of the *urban fabric* derived from a data-driven approach. Fig. 1 illustrates these concepts and the data (further described in Data section) used as proxy.

In this study, the *urban fabric* is considered as the physical aspect of the city in its spatialized extent. We consider the *urban fabric* to be a highly heterogeneous spatial arrangement of voids and solids of different natures (built and/or natural landscapes). In that sense we are

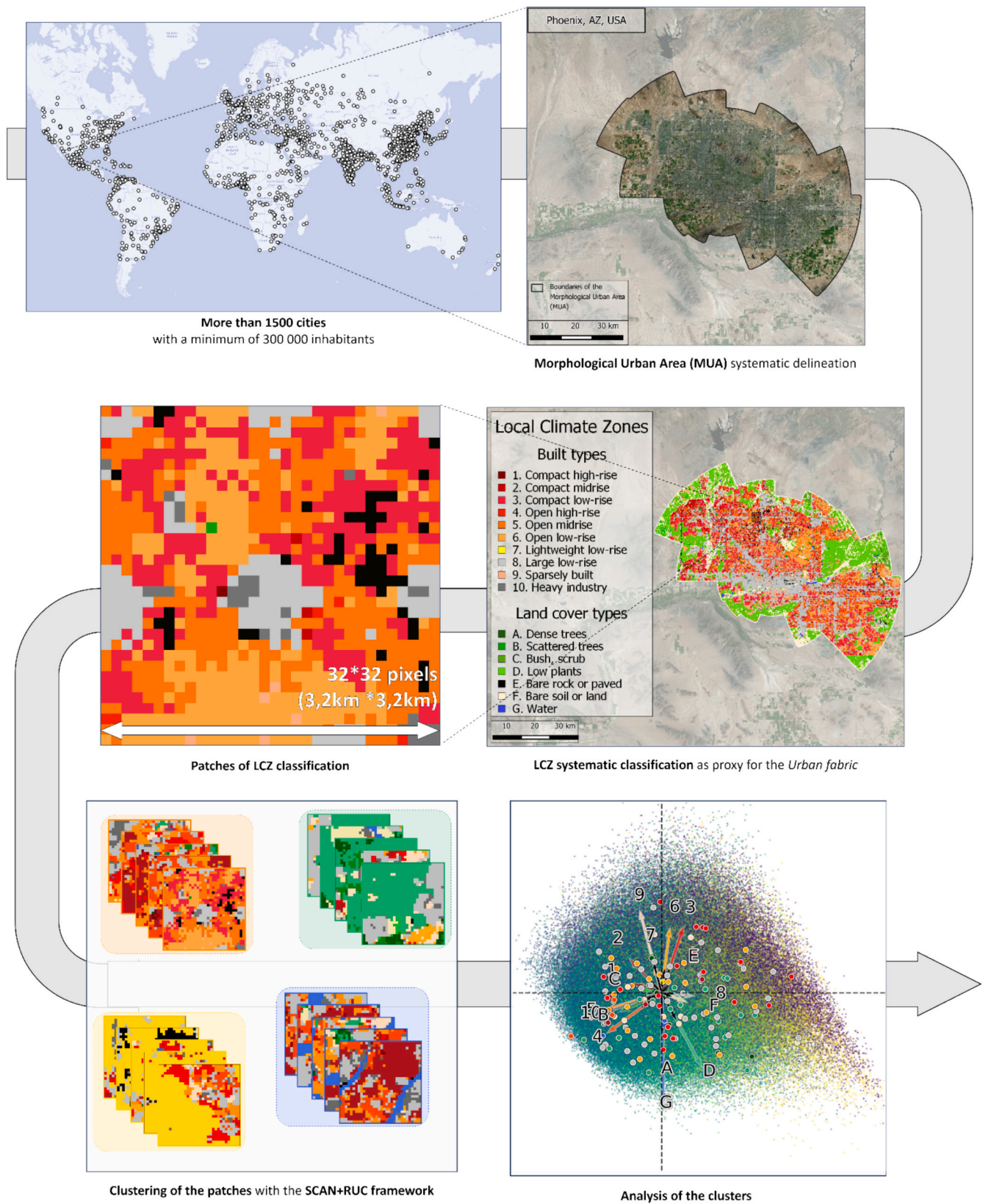


Fig. 1. Design of the study.

focusing here on the composition and the configuration of both the man-made and the natural elements of a city following (Wentz et al., 2018). The land use types are not considered here, as we focus solely on the physical aspect of the city. A global multidimensional dataset of urban

land cover classification (LCZ) is used as a proxy of the *urban fabric*. To frame a typology of the internal elements of the *urban fabric*, we pursue the identification of the recurring spatial *patterns* within our global urban land cover classification dataset at an intra-urban scale. To do so,

we employ the SCAN + RUC framework on sub-segments (later referred to as *patches*) of the LCZ classification. Considerations on urban dimensions, resolution of the urban land cover classification and computational necessities of the SCAN + RUC framework drove the final dimensions of the *patches* (see section Methodology). The *patches* sharing similar *patterns* are clustered together through the SCAN + RUC framework. Finally, we analyze the typical compositions and configurations of the urban landscapes of the *clusters* identified and of their *patches* and draw our conclusions on the universal typology of intra-urban patterns drawn from this.

3. Data

3.1. Study sites – Morphological urban areas

The focus of this study is to frame a typology of intra-urban patterns that is as comprehensive and global as possible. We used the Morphological Urban Areas (MUAs) of these cities as provided by (Taubenböck et al., 2019). This delineation of urban areas was made in a data-driven way for all major cities with more than 300,000 inhabitants across all continents as of in 2019 by distinguishing the urban areas from distant rural hinterlands. It does so by considering the gradient of decreasing built-up density from the urban center to the periphery and finding the specific threshold of density adequate for each urban area. While this dataset primarily focuses on major cities, its definition permeates cities administrative boundaries and allows to capture the smaller cities in their vicinity, representing the diversity of city types and scales. This mask of urban areas according to the boundaries of cities' main morphological body ensures a consistent definition of the spatial unit at which we analyzed the *urban fabric* of our study sites (cf. Fig. 1).

The final list of study sites comprised 1523 MUAs (cf. Fig. 1 and covered an area of roughly 360,000 km².

3.2. Urban fabric – Local Climate Zones classification

As a proxy for the *urban fabric* we used a land cover classification following the classification scheme of the Local Climate Zones (LCZ) introduced in (Stewart and Oke, 2012). The LCZ scheme is comprised of seventeen classes, ten of which are of built types and seven of them are natural or non-built types. Each class of the LCZ scheme can be primarily characterized by its type (built, vegetation, surface) as well as its density and its height, representing multiple dimensions of the urban fabric. In addition, each class within the LCZ scheme is characterized by other secondary features such as sky view factors, aspect ratios, impervious surface fractions, terrain roughness classes and more making it a richly detailed classification scheme (Stewart and Oke, 2012). Therefore, although this scheme of classification was primarily developed to accommodate for urban climate analysis, its classes based on morphological features of the *urban fabric* make it valuable for the study of urban morphology (Bechtel et al., 2015; Debray et al., 2021; Demuzere et al., 2021; Lemoine-Rodríguez et al., 2022; Taubenböck et al., 2020; von Szombathely et al., 2017; Wentz et al., 2018).

We used a LCZ classification at a resolution of 100 m * 100 m introduced in (Zhu et al., 2021). The global LCZ classification was produced based on scenes of Sentinel-1 and -2 satellite imagery. This dataset covers various extents surrounding the urbanized parts of the cities, therefore, as mentioned above, we curated the data to our set of study sites. Beyond the relevance of the LCZ classification discussed above, this dataset is chosen because of the high accuracy reported for this product as well as its global coverage (Zhu et al., 2021, 2022), which was a key factor of the present study.

4. Methodology

4.1. Disaggregation of LCZ classification into patches

The disaggregation of the LCZ classification data was operationalized following by three considerations.

First, it is identified (e.g. in (Wu et al., 2025)) that urban patterns can be analyzed across distinct scales that can be summarized into three categories: the *micro-scale* (corresponding to the scale of analysis of e.g. individual street blocks, land parcels or singular buildings), the *meso-scale* (corresponding to the scale of analysis of e.g. district, neighborhoods or street layout), and the *macro-scale* (corresponding to the scale of analysis of e.g. a city agglomeration or a metropolitan area) (Wu et al., 2025). In the present study we are aiming at drawing a typology of intra-urban patterns at a *meso-scale*. While numerous studies refer to the *meso-scale* (e.g.: (Boeing, 2020; W. Chen et al., 2024; Lim et al., 2017; Mu et al., 2020; Ramiaramanana et al., 2025; Schirmer and Axhausen, 2016; Sharifi, 2019; Wu et al., 2025)) no metric dimensions are univocally identified for this scale. Based on these works we suggest a range of metric dimensions to analyze the *meso-scale* of the *urban fabric* as being comprised between 500 m to 5 km depending on the domain of application.

Second, the SCAN + RUC framework (further detailed in the next sections) relies on the use of a ResNet-18 convolutional neural network (He et al., 2015) which input square patches of data of systematic dimensions. The SCAN + RUC framework was developed and tested for *patches* of size 32 * 32 pixels.

Third and last, we consider the Modifiable Area Unit Problem (MAUP) (cf. (Openshaw, 1984)) in the disaggregation of our data. The MAUP is a bias that can appear when the selection of the extent of a study area (here meaning the disaggregation into *patches*) can affect directly the measure. To reduce this bias at a set dimension of *patches*, one strategy is to enforce an overlap between neighboring *patches* (Weigand et al., 2023).

As a result, we operationalized the subsetting of the *urban fabric* through the systematic disaggregation of the LCZ classification data into square tiles (*patches*) of 32 pixels by 32 of the LCZ classification. With the 100 m resolution of the LCZ classification this corresponds to regions with a side length of 3.2 km (cf. Fig. 1) which falls into dimensions of *meso-scale* areas, and which is large enough to reflect coherent urban patterns and their heterogeneity yet small enough to allow meaningful comparisons within and across cities of different scales. We extracted these *patches* in each study site with a 2/3 overlapping ratio and encoded them on 17 binary layers accounting for the presence of each of the 17 LCZ classes at each pixel position of the *patch*. Doing so, we produced roughly 315,000 *patches*, accounting for a cumulated area of roughly 360,150 km² without overlap (resp. roughly 3,225,600 km² with overlap).

4.2. Framework for unsupervised clustering

Recent progress in the field of unsupervised clustering using deep learning showcased newly developed capacities to identify and cluster similar spatial patterns in raster based input with state-of-the-art accuracy in large amounts of data (Han et al., 2020; Ji et al., 2019; S. Park et al., 2021; Qian et al., 2022; Van Gansbeke et al., 2020). From this pool of new methods, we selected the one yielding the best accuracy to our knowledge at the time of the study, namely the SCAN + RUC framework (S. Park et al., 2021).

This approach relies on the successive use of three consecutive frameworks on their own: the SimCLR (Simple framework for Contrastive Learning of visual Representations) (T. Chen et al., 2020), the SCAN framework (Van Gansbeke et al., 2020) and the RUC framework (S. Park et al., 2021). Further description of these individual frameworks is given in the following sections. The combination of these three frameworks demonstrated high accuracies of 90.1 % (resp. 86.6 %) on major

computer vision benchmark image datasets such as the Canadian Institute For Advanced Research 10 dataset (CIFAR-10) introduced in (Krizhevsky, 2009) (resp. Self-Taught Learning 10 dataset (STL-10) introduced in (Coates et al., 2011)), among others (S. Park et al., 2021).

The resulting SCAN + RUC framework can be understood as comprised of four sub-tasks (see Fig. 2): 1) the learning of discriminative features from *patches* using contrastive learning; 2) the pooling of similar *patches* based on learned features; 3) the training of the model based on highly confident *prototypes*; and 4) the refining of the model around clean samples only.

For this study, the RUC + SCAN framework had to be slightly adapted to the context of our research hypothesis and data. In the next sections, we give a synthesized explanation of each of the four sub-tasks of the framework and explain the adaptations made. For the purpose of consistency and clarity, we adapted the original technical vocabularies employed in (T. Chen et al., 2020; S. Park et al., 2021; Van Gansbeke et al., 2020) to a consistent vocabulary adapted to the concepts of this study.

4.3. SimCLR: Learning discriminant features

The first step of our method comprised the use of the SimCLR framework (T. Chen et al., 2020). The main goal of this step of the framework was for a neural network to learn semantically meaningful features in our set of roughly 315,000 *patches* that served as the base for clustering done in further steps of the framework.

Following (T. Chen et al., 2020), for every epoch of training, each *patch* was augmented (i.e. modified slightly in its composition or orientation – see (T. Chen et al., 2020) –) into two different versions. A ResNet-18 neural network (He et al., 2015) was trained to find features that associated the two augmented versions of the *patches* and discriminated them from the other *patches*' augmented versions. We kept technical aspects of this step as close to the one proposed in (T. Chen et al., 2020) as possible. We specifically kept the same augmentation operations adapted for the 17-channels encoding of the *patches*, albeit for the color distortions that we discarded as here they would have temper with the categorical nature of the LCZ classification. Due to limitations in computational capacities, the network was trained with batches of 256 *patches*. Therefore, to increase accuracy as indicated in (T. Chen et al., 2020), the network was trained for 500 epochs.

4.4. SCAN: Self-labelling unlabeled data

This second step was an adapted application of the SCAN framework (Van Gansbeke et al., 2020). The aim was to create relevant pseudo-labels for our unlabeled data. This step was comprised of two phases: First, we assigned temporary pseudo-labels for our *patches* by selecting their nearest neighbors in the latent space created by the SimCLR step. Through the SimCLR step, *patches* that were neighbors in this latent space shared similar features. Second, we refined these pseudo-labels in a self-supervised way, focusing on *patches* with highly confident pseudo-labels (i.e. *patches* with high values for the prediction probability of their

respective assigned pseudo-label).

Based on the results in (Van Gansbeke et al., 2020), we selected the 20 nearest neighbors of each *patch* in the latent space. From these 20 neighbors, one other *patch* was randomly selected. Both *patches* are independently augmented in the same manner as in (Van Gansbeke et al., 2020), albeit, as explained above, for the color distortion. On these bases, a second network with a ResNet-18 backbone was initialized with the weights of the previous one trained within the SimCLR framework. Within this step, this second network was trained with a loss function which, both at the same time, rewarded the assignment of the same pseudo-label to these two *patches* and penalized the assignment of the same pseudo-labels to *all patches* of the same batch (Van Gansbeke et al., 2020). Furthermore, we made the choice of using for this network a cluster head allocating 1024 possible pseudo-labels.

The choice of having 1024 pseudo-labels at this stage of the methodology was motivated by two reasons: The first reason was tied to the fact that these pseudo-labels correspond to groups of *patches* that are precursors to the final clusters. As we did not know *a priori* how many clusters we would find in our dataset (contrary to the seminal study proposing the SCAN framework (Van Gansbeke et al., 2020)), we didn't restrict the clustering to a small number of groups. We therefore chose to deliberately risk, at this stage, over-clustering our data, meaning creating more groups than needed, over the risk of under-clustering, i.e. degenerating the clustering by having too few clusters with too many *patches* within them, losing the homogeneity of the clusters. This risk of over-clustering was accepted based on the knowledge that our modifications of the further steps of the SCAN framework would allow us to correct it by aggregating pseudo-labels (see below). As for the second argument for this number of pseudo-labels, we empirically found that the method is insensitive to the change of numbers of pseudo-labels beyond 1024 and only induced more dimensionality for the computation of subsequent steps. Therefore, the choice of 1024 pseudo-labels was a fitting trade-off for our present application. For this first phase of the SCAN step, the network was trained for 50 epochs on batches of 384 *patches*. Here, the distribution of the number of *patches* across the pseudo-labels was uniform.

The second phase of the SCAN step focused on *patches* with high confidence in their pseudo-labelling to reduce the influence of inevitable false positive pseudo-labelling made during the first phase. This step was proposed by (Van Gansbeke et al., 2020) to retrain the neural network around high confidence samples called "*prototypes*" that are considered less prone to be false positives. We followed here the same logic as in (Van Gansbeke et al., 2020) but differed in the way we characterize *prototypes*.

(Van Gansbeke et al., 2020) proposed to define the *prototypes* as samples with a confidence score straightforwardly derived from the SoftMax value of the top-1 prediction of the pseudo-labelling above a threshold of 99.9 %. This threshold was defined in the frame of their seminal work with clearly defined classes of benchmark datasets such as STL-10 or CIFAR-10 on which the number of labels was known in advance. With our purposeful overshooting of the pseudo-label number, no sample reached such confidence, as the SoftMax score of each *patch*

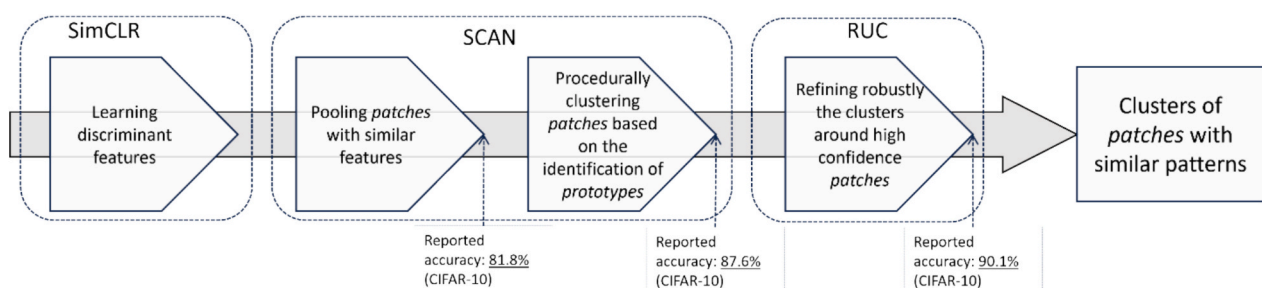


Fig. 2. Steps of the SCAN + RUC workflow adapted to our study.

was spread across multiple pseudo-labels representing smaller pseudo-clusters closer in the latent space. We therefore adapted the definition of *prototypes* in our context as follows: the confidence score \widehat{p}_x of a *patch* x was defined as the sum of the SoftMax values p_{x_i} of its top-3 predicted pseudo-labels i (see Eq. (1)); the confidence threshold τ_i for a pseudo-label i was set as the standard deviation of confidence values $\sigma(p)_i$ for the pseudo-label subtracted from the maximum confidence $\max(p)_i$ found for the pseudo-label (see Eq. (2)).

$$\widehat{p}_x = \sum_i \max_3(p_{x_i}) \quad (1)$$

$$\tau_i = \max(p)_i - \sigma(p)_i \quad (2)$$

With these, we defined as *prototypes* of a pseudo-label i any *patch* x for which $\widehat{p}_x \geq \tau_i$.

We used the sum of multiple top confidence values to account for the over-clustering of the *patches*. Our definition of the confidence score as the sum of the triplet of maximum-SoftMax-value *patches* with similar “confidence profiles” (i.e. similar triplets of maximum-SoftMax-values per pseudo-label) was made to account for the over-clustering of the *patches*. In our context of over-clustering, where multiple pseudo-labels were semantically overlapping this confidence score introduced the semantic connections between pseudo-labels frequently associated in their “confidence profile”. As the training progressed, to minimize the cross-entropy loss function, the confidence profiles of prototypes converged incrementally toward an increasingly dominant maximum-SoftMax-value for only one of their pseudo-labels. In turn, this meant that the other associated pseudo-labels were progressively de-allocated, and it effectively induced a merging of the associated pseudo-labels. Therefore, the use of the cross-entropy loss function combined with our specific definition of the confidence score allowed to correct the over-clustering induced in the previous phase of the SCAN step while retaining the semantic connections between the original pseudo-labels.

At the same time, we had to avoid enforcing an under-clustering by assuming that too many pseudo-labels overlapped. Henceforth, we limited the sum to the top three values. We argue that this pseudo-label-individualized threshold allowed to account for the variability of the overall confidence scores between pseudo-labels.

Beside this adapted definition of prototypes, this phase of the SCAN remained the same as in (Van Gansbeke et al., 2020). The neural network was trained for 200 additional epochs with batches of 800 *patches* using the *prototypes* and their pseudo-labels in a self-supervised way. Like in the previous steps, the *prototypes* were augmented to avoid overfitting. The loss function used for the training of the network in this phase followed a regular cross-entropy formula that, contrary to the first step, did not force an equal distribution among the 1024 pseudo-labels.

4.5. RUC: Improving the self-labeling robustly

The last step of our framework was a straightforward application of the “hybrid strategy” of the RUC framework proposed in (S. Park et al., 2021). The aim of this step was to improve the pseudo-labeling of the SCAN step by revising potentially mislabeled samples and adjusting the confidence of the results. The RUC was developed to be used as an add-on step on top of unsupervised frameworks such as SCAN. The original study presenting the RUC framework showed great benefits as an add-on to SCAN reaching high-end accuracies (S. Park et al., 2021).

For further details on the technical details of the RUC framework we refer to (S. Park et al., 2021). The “hybrid strategy” of the RUC framework was applied on the pseudo-labels obtained with the SCAN step. The training was operated over 200 epochs with minibatches of 110 samples. It is to be noted that the RUC step did not alter the number of pseudo-labels and therefore proposed as many *clusters* as the second phase of our SCAN adaptation. This step allowed us to obtain the final *clusters* on which we draw our universal typology of intra-urban patterns.

4.6. Analysis of results

As mentioned above, the accuracy of the SCAN + RUC framework has been evaluated on standard computer vision benchmarks with good results in a reproducible way (S. Park et al., 2021; Van Gansbeke et al., 2020). Although transferability of the high accuracy was expected between standard benchmarks and our dataset, we chose to test the results of this framework in the context of our research by analyzing the clusters obtained through this framework and the statistical differences between them.

In that purpose, we characterized the patterns of the *patches* using landscape metrics, drawing on previous studies assessing urban form for a wide range of scales (Huang et al., 2007; Inostroza et al., 2013; Lemoine-Rodríguez et al., 2020; Taubenböck et al., 2009).

Using the Python library Pylandstats (Bosch, 2019), we computed a set of 91 typical landscape metrics inspired by the commonly used software FRAGSTATS (Bosch, 2019; McGarigal et al., 2023) on each of the *patches* of our dataset. The aim of using a large amount of landscape metrics was to characterize in a multi dimension fashion the configurations and compositions of each single *patch*. This set comprised metrics that allowed to describe the characteristics of intra-urban composition and configuration.

To test if the clusters identified were different in terms of their *patches’* composition and configuration, these quantitative metrics were utilized as basis for two series of statistical tests. Since our dataset is large (ca. 315,000 *patches*), we discarded the use of a typical Null Hypothesis Significance Testing (NHST) as it was expected to lack power to inform us on the differences between *clusters* and would trivially be invalidated (Khalilzadeh and Tasci, 2017; Lakens, 2013; Nakagawa and Cuthill, 2007). First, a Mann-Whitney (Wilcoxon, 1945) post-hoc test was performed to evaluate which pairs of *clusters* differed significantly. This non-parametric test is suited for contexts where the characteristics of the statistical distribution of the metrics are *a priori* unknown and where cluster sizes might differ (Fay and Proschan, 2010; J. Liu et al., 2022). Second, we performed a series of tests on the effect size of the differences between the *clusters*. For the effect size measure, for each landscape metric we first computed the overall effect size of all *clusters* and second the effect sizes between *clusters* pairs. For categorical metrics, Cramer’s V measure (Cramér, 1991) was computed and for continuous metrics, we computed the effect size using the η^2 , ϵ^2 and ω^2 (Carroll and Nordholm, 1975; Cohen, 1973; Hays, 1973; Lakens, 2013). For these different measure of effect sizes we report below the categories of effect size they are associated with (no effect, small effect size, medium effect size, large effect size) as per in (Khalilzadeh and Tasci, 2017). Following recommendations to tailor the interpretations of these results to the domain of application (Cohen, 1988; Sawilowsky, 2009), we suggest interpretations for these four categories. We interpret “no effect” as corresponding to the fact that the difference between the measured metric of any pair of samples across the two different groups is not enough that we could expect any differences in the landscapes along the consideration of this particular metric. We interpret a small effect size as the fact that the difference between the measured metric of any pair of samples across the two different groups is enough to be considered different although the landscape themselves might still share similarly looking features along the consideration of this particular metric. Correspondingly, a medium effect size corresponds to the fact that the difference between the measured metric of any pair of samples across the two different groups is enough to be considered different although the landscape themselves have different features along the consideration of this particular metric. Last, a large effect size is here interpreted as the fact that the difference between the measured metric of any pair of samples across the two different groups is enough to be considered different although the landscapes themselves have very different features, to the extent that, along the metric considered, they look nothing alike. In each case, we report further below the amount and percentage of landscape metrics and the percentage of pairs pertaining

to each category of effect size. As this could foster skewed results coming from the correlations between landscape metrics, we filtered from the original set of landscape metrics, a subset where no landscape metrics share any strong correlations (at a threshold of $p < 0.7$). A summary of the 91 metrics formulas, along with their correlations and the subset finally selected can be found in [Appendix A-B](#).

After this statistical evaluation, we explored the geographical significance of the compositions and configurations of the landscapes of the *patches* and the *clusters*. We did so by performing a principal components analysis of the landscape metrics and reporting on the general distributions and co-linearity of key explicative metrics being the share of the different LCZ classes across the *patches* as well as eight landscape metrics pertaining to the configuration and composition of the *patches*. We namely used the **Entropy** as a measure of the diversity of composition within a *patch*; the median area of segments within the *patch* (**Med. Area**), as a measure of the typical size of “pockets” of the same LCZ class within a *patch*; the Relative Mutual Information (**RMI**) as a measure of the tendency of neighboring pixels of a *patch* to be of the same LCZ class; the median Shape Index (**Med. Shape Index**), as a measure of the typical shape complexity of the homogeneous sub-regions of the same LCZ class in a *patch*; the share of pixels belonging to the most dominant LCZ class in the *patch* (**Top-1 %**), as a measure of dominance of its landscape; the share of pixels belonging to the second most dominant LCZ class in the *patch* (**Top-2 %**), as an indicator of the hegemonic or multi-faceted composition of the its landscape; the coefficient of variation of the area (**CV of Area**) of segments within the *patch*, as a measure of the variation of the side of homogeneous sub-regions of the same LCZ class; the coefficient of variation of the Shape Index (**CV of Shape Index**) of the segments, as a measure of the diversity of the shapes of homogeneous sub-regions of the same LCZ class.

5. Results

5.1. Typology of intra-urban patterns

Applying the SCAN + RUC framework on the more than 315, 000 *patches* of our global dataset, we obtained 138 *clusters*. Each *patch* considered in our approach was classified into one of these 138 *clusters*, meaning that 138 types of intra-urban patterns are sufficient to define the *urban fabric* of any city in the world. We found that the 138 *clusters* presented overall a rich breadth of distinctive visual features and internal consistency. We organized the identified intra-urban pattern types by order of their shares of built-up LCZ classes computed at the *cluster* level. A summary of exemplary *patches*, morphological descriptions and semantic interpretations of the typical landscapes of the 138 identified intra-urban pattern types can be found in [Appendix C](#). A subset of this summary for 10 pattern types along the gradient of the 138 ordered *clusters* is presented in [Fig. 3](#).

This richness of patterns was of course not without a few similarities between *clusters* as can be visually understood in [Figs. 4A-B](#). We further found that the *clusters* vary in representativity as they featured between 92 and 5, 689 *patches* with a median of 2, 313 ([Fig. 5](#)).

The characteristic landscapes of each cluster was found to be qualitatively different between each of them while internal diversity was also observed at the individual *cluster* level (cf. [Figs. 4A-B](#) and [Appendix C](#)). Overall, we observed a large diversity of landscape compositions and configurations across the *clusters* identified ([Appendix C](#)). Some *clusters* were largely dominated by vegetation (e.g.: *clusters* 1 and 15), some others were mainly comprised of dense built-up (e.g.: *clusters* 107, 121 and 138), while some others were more heterogeneous, exhibiting similar proportions of vegetation and built-up (e.g.: *clusters* 30, 48 and 63). Conversely, the *clusters* differed in configurations, i.e. some of them featured homogeneous areas (e.g.: *clusters* 1 and 138), some were more divided (e.g.: *cluster* 59, 77, 128), some were interspersed (e.g.: *clusters* 10, 22, 56, 100, 125), or some presented multiple small scale homogeneous sub-regions (e.g.: *clusters* 28, 55, 116).

We found quantitatively that the share of the built-up type ‘LCZ-8: large low-rise buildings’ was dominant in 45 of the identified *clusters* while representing 18.9 % of the global urban fabric being the most represented LCZ class. Second on this rank was ‘LCZ-D: low plants that presented similar shares of the global urban fabric (18.6 %) and was dominant in 22 *clusters*. Overall, we found that 65.6 % of the global urban fabric is built-up and 34.4 % is not while 118 *clusters* were on average dominantly built-up against 20 that were on average dominantly non-built-up ([Fig. 5](#)).

Interestingly, we further observed that built-up forms of low density (LCZ-4: Open high-rise; LCZ-5: Open midrise; LCZ-6: Open low-rise; LCZ-8: Large low-rise; LCZ-9: Sparsely built) together represented 48.6 % of the global urban fabric and accounted for more than half of the surface in 52.9 % of all *patches* and in 71 out of 138 types of the *clusters*. Exposed soil and low vegetation non-built-up cover types (LCZ-D: Low plants; LCZ-E: Bare rock or paved; LCZ-F: Bare soil or sand) accounted for 22.9 % of the LCZ share of the global urban fabric, were sharing above half of the surface in 9.2 % of all *patches* and in 11 out of 138 *clusters*. Against this, we found that dense built-up forms (LCZ-1: Compact high-rise; LCZ-2: Compact midrise; LCZ-3: Compact low-rise; LCZ-7: Lightweight low-rise) accounted for 13.8 % of the LCZ share of the global urban fabric, were sharing above half of the surface in 8.1 % of all *patches* and in 8 out of 138 *clusters*, while higher vegetation (LCZ-A: Dense trees; LCZ-B: Scattered trees; LCZ-C: Bush, scrub) accounted for 9.4 % of the LCZ share of the global urban fabric, were sharing above half of the surface in 2.2 % of all *patches* and in 1 out of 138 *clusters* ([Fig. 5](#)).

5.2. Statistical tests

The Mann-Whitney post-hoc test, resulted in the observation that across the set of 49 non-strongly correlated metrics, on average 84.2 % of the *cluster* pairs presented significant differences (at a threshold of $p < 0.05$). At the minimum, one landscape metric presented 31.7 % of the *cluster* pairs to be significantly different, and at the maximum, one metric presented 94.5 % of the pairs to be significantly different.

Analyzing the different effect size measures for the continuous landscape metrics, we observed only marginal differences of a range of 10^{-4} units between the values of η^2 , ϵ^2 and ω^2 which was to be expected thanks to the large sample sizes in our study. In no case did these marginal differences change the effect size category for any landscape metrics. We found that a large majority of 34 metrics out of 49 presented overall large effect sizes. 10 metrics showed medium effect sizes, and 5 metrics presented small effect sizes with no metric presenting an insignificant effect size. These effect sizes results represent the fact that for the overall dataset grouped into the 138 *clusters* identified, the *patches* are significantly different at least to a small extent in term of features of their landscapes.

Analyzing the effect sizes between pairs of *clusters* for each of these 49 metrics, we found that on average for the landscape metrics, a large majority of 77.2 % of the pairs of *clusters* presented a significant effect size (of which, respectively, 28.5 % presented a small effect size, 21.4 % a medium effect size, 27.6 % a large effect size). These effect sizes results represent the fact that each pair of the 138 *clusters* identified are on average significantly different at least to a small extent in term of features of their landscapes.

Moreover, we identified that the pairs of *clusters* that presented non-significant effect sizes were not the same across different landscape metrics. This indicated that the *clusters* identified by our approach might have shared some features but differed on a multi-dimensional level considering multiple landscape metrics.

To summarize, we observed that our approach identified *clusters* that were quantitatively different in terms of the morphology of their *patches*, represented by measures of their compositions and configurations. These quantitative statistical results coincided with the qualitative

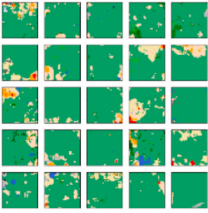
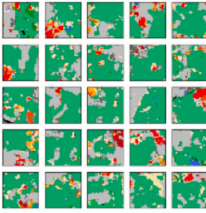
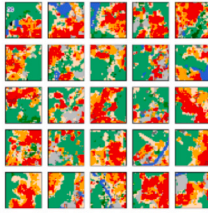
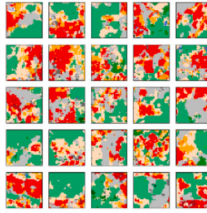
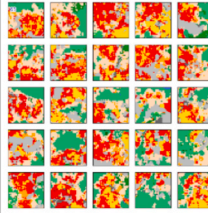




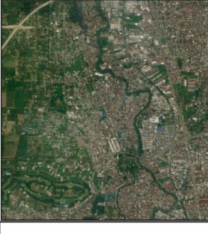
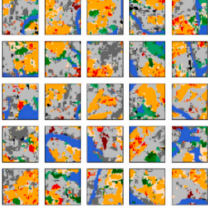
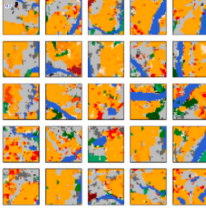
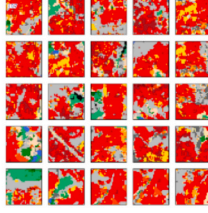
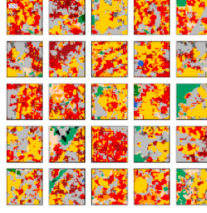
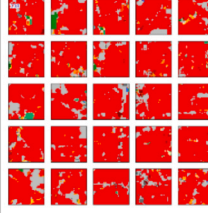





I-UPT label	1	15	30	48	63
Exemplary patches (each patch is 3.2km×3.2km)					
Morphological description	Predominantly low plants (LCZ-D) with encroachments of bare soil or sand (LCZ-F)	Extents of low plants (LCZ-D) with agglomerations of large low-rise buildings (LCZ-8) pocketed with compact low-rise buildings (LCZ-3)	Extents of low plants (LCZ-D) with agglomerations of compact low-rise buildings (LCZ-3) bordered by open low-rise buildings (LCZ-6) and edges of sparsely built buildings (LCZ-9), and pockets of large low-rise buildings (LCZ-8)	Extents of low plants (LCZ-D) fragmented by agglomerations of compact low-rise buildings (LCZ-3) and open low-rise buildings (LCZ-6) with edges of sparsely built buildings (LCZ-9) and bordered by large low-rise buildings (LCZ-8)	Agglomerations of mixed compact low-rise buildings (LCZ-3) and lightweight low-rise buildings (LCZ-7) with edges of sparsely built buildings (LCZ-9) and pockets of large low-rise buildings (LCZ-8), bordered by extents of low plants (LCZ-D)
Exemplary satellite picture (each picture is 3.2km×3.2km)					
Semantic label	Agricultural land in arid context	Agricultural lands in vicinity of commercial or industrial parks with some compact-low habitations	Farm communities with logistic centers	Farmland at edges of compact to open settlements with industrial areas	Compact low settlements with a part of makeshift habitations around industrial areas
I-UPT label	78	93	107	121	138
Exemplary patches (each patch is 3.2km×3.2km)					
Morphological description	Thin rivers of Water (LCZ-G) with agglomerations of large low-rise buildings (LCZ-8) and heavy industry areas (LCZ-10) on the shores and core pockets of open low-rise buildings (LCZ-6)	Thin rivers and bodies of Water (LCZ-G) bordered by agglomerations of open low-rise buildings (LCZ-6) and agglomerations of large low-rise buildings (LCZ-8) on the shores	Agglomerations of compact low-rise buildings (LCZ-3) fragmented by stretches and agglomerations of large low-rise buildings (LCZ-8)	Agglomerations and pockets of lightweight low-rise buildings (LCZ-7) interlocked with compact low-rise buildings (LCZ-3), bordered by agglomerations of large low-rise buildings (LCZ-8)	Extents of compact low-rise buildings (LCZ-3) with pockets of large low-rise buildings (LCZ-8)
Exemplary satellite picture (each picture is 3.2km×3.2km)					
Semantic label	Fluvial industrial harbors with low density residential settlements	Low density residential neighborhood on riversides with malls	Compact mix layout of midrise buildings with inner industrial areas	Very dense mixed layout of low buildings around industrial facilities	Compact-low residential neighborhood with commercial zone
<div> <div>Compact high-rise</div> <div>Compact low-rise</div> <div>Open midrise</div> <div>Lightweight low-rise</div> <div>Sparsely built</div> <div>Dense trees</div> <div>Bush, scrubs</div> <div>Bare rock or paved</div> <div>Water</div> <div>Compact midrise</div> <div>Open high-rise</div> <div>Open low-rise</div> <div>Large low-rise</div> <div>Heavy industry</div> <div>Scattered trees</div> <div>Low plants</div> <div>Bare soil or sand</div> </div>					

Fig. 3. Exemplary patches, morphological descriptions, exemplary satellite picture (credit: Bing Aerial Imagery) and semantic interpretation of 10 intra-urban pattern types (I-UPTs) along the gradient of the ordered 138 I-UPTs.



Fig. 4A. Visualization of the 138 clusters defining the urban fabric of cities. Each cluster is represented by a line of 5 of its patches being at the 1st, 20th, 40th, 60th, 80th and 100th centiles of the closest neighbors of the cluster's centroid (i.e. the 5 patches describe a gradient from the closest to the centroid to the furthest from the centroid of the clusters, in the embedded space of the neural network). The figure is organized by clusters of increasing shares of built-up LCZ classes (LCZ 1 to 10), from top to bottom and left to right. (cluster 1 to 72).

assessment developed in Fig. 3 and Appendix C and allowed to relativize the fact that some *cluster* pairs visually seemed to be similar. Some pairs of *clusters* indeed were found to share some features of their landscapes, but our multi-dimensional approach still highlighted that their landscapes differed significantly.

5.3. Morphological characteristics

Plotting the result of the first two principal components (PC1 and PC2) of the PCA, we obtained the following biplots (Fig. 6). The first component PC1 explained 30.6 % of the variance and represented best the Entropy, the CV of Area and Top-1 % and Top-2 % for the landscape metrics. It also explained the best shares of 'LCZ-1: Compact high-rise',

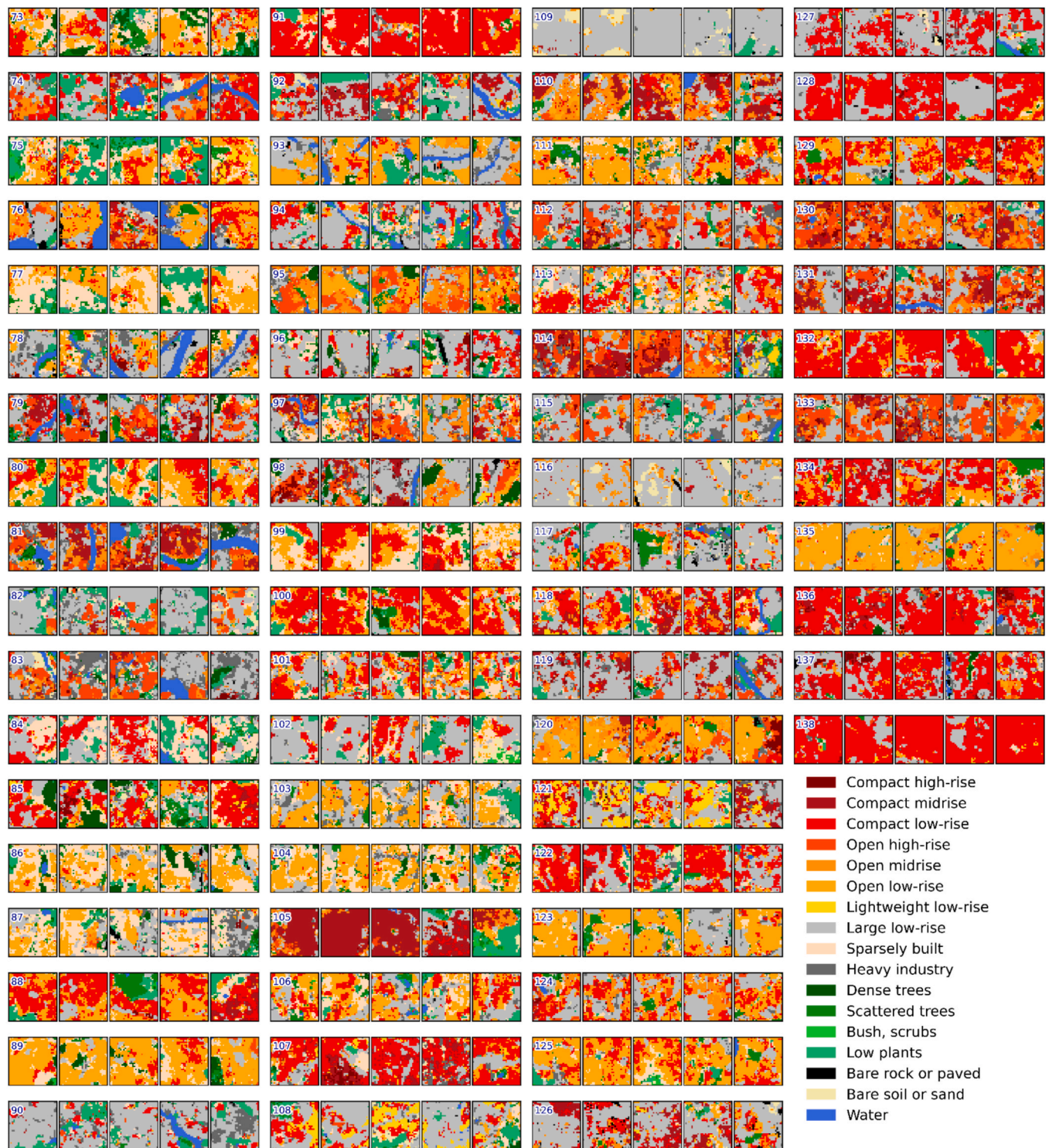


Fig. 4B. Visualization of the 138 clusters defining the urban fabric of cities. Each cluster is represented by a line of 5 of its patches being at the 1st, 20th, 40th, 60th, 80th and 100th centiles of the closest neighbors of the cluster's centroid (i.e. the 5 patches describe a gradient from the closest to the centroid to the furthest from the centroid of the clusters, in the embedded space of the neural network). The figure is organized by clusters of increasing shares of built-up LCZ classes (LCZ 1 to 10), from top to bottom and left to right. (cluster 73 to 138).

'LCZ-5: Open midrise', 'LCZ-8: Large low-rise', 'LCZ-10: Heavy industry', 'LCZ-B: Scattered trees', 'LCZ-C: Bush', 'scrubs', and 'LCZ-F: Bare soil or sand'. PC2 explained 12.1 % of the variance and represented best the CV of the Shape Index, Med. Shape Index, RMI and Med. Area for the landscape metrics. It represented best the shares of 'LCZ-2: Compact midrise', 'LCZ-3: Compact low-rise', 'LCZ-4: Open high-rise',

'LCZ-6: Open low-rise', 'LCZ-7: Lightweight low-rise', 'LCZ-9: Sparsely built', 'LCZ-A: Dense trees', 'LCZ-D: Low plants', 'LCZ-E: Bare rock or paved', and 'LCZ-G: Water'.

While we could observe a mixture of more diverse values of built-up share on the left part of the plot, on the right the points spread to the lower right part of the biplots where we observed more extreme ones

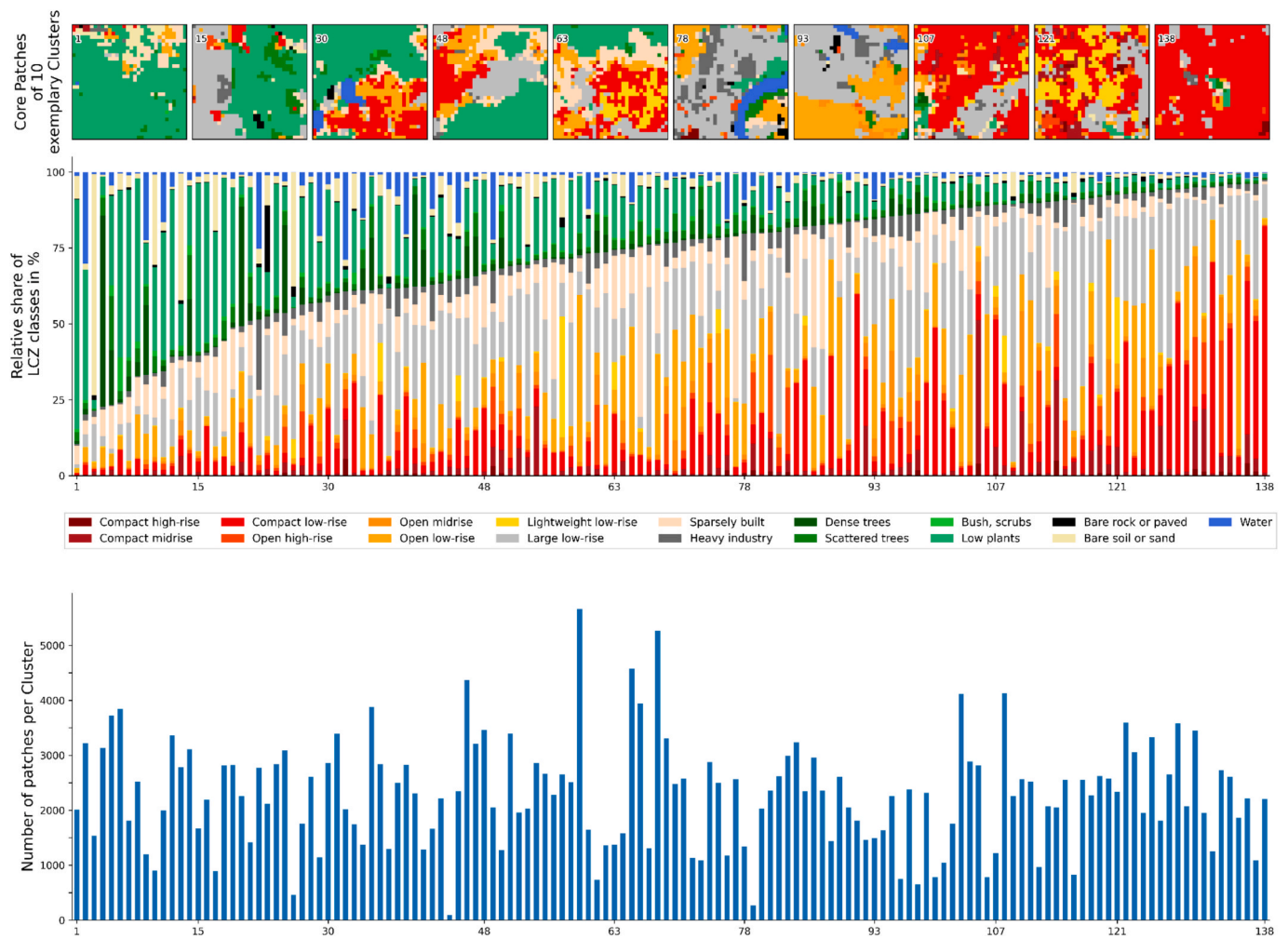


Fig. 5. LCZ compositions of the clusters, and number of patches per cluster. Clusters are ordered by increasing share of built-up LCZ classes.

with a strong divide between highly built-up and conversely highly non-built-up *clusters* and *patches*. Additionally, we observed lower built-up forms on the upper part of the plot and more compact ones on the left of the plot, while vegetation land cover seemed dominant in the lower part of the plot.

This biplots further helped us to identify groups of associated LCZ classes. For example, we found 'LCZ-3: Compact low-rise', 'LCZ-6: Open low-rise', 'LCZ-9: Sparsely built' and to a more minor effect 'LCZ-7: Lightweight low-rise' to be jointly grouped in *patches* of the urban fabric while being antagonistic to the group formed by 'LCZ-A: Dense trees', 'LCZ-D: Low plants' and 'LCZ-G: Water'. We identify LCZ classes 'LCZ-8: Large low-rise' and 'LCZ-F: Bare soil or sand' to be co-jointly present and negatively related to the presence of LCZ classes 'LCZ-1: Compact high-rise', 'LCZ-2: Compact midrise', and 'LCZ-C: Bush, scrub'. Lastly, we identified that 'LCZ-4: Open high-rise', 'LCZ-5: Open midrise', 'LCZ-10: Heavy industry' and 'LCZ-B: Scattered trees' were associated.

Based on this, we divided empirically these biplots in 4 quadrants: Upper right corresponding to mostly built-up *clusters* and *patches* of low-rise building types; Lower right corresponding to a high presence of non-built-up areas, among which mostly 'LCZ-D: Low plants'; Lower left corresponding to open built-up forms with dense to scattered trees; Upper left corresponding mostly to compact built-up forms with little vegetation.

With this, looking at the landscape metrics, we saw that the upper right quadrant is positively colinear to Top-1 % and CV of Area, indicating that these highly built-up *clusters* and *patches* were hegemonically covered by a single LCZ class with very small scattering of secondary

LCZ classes. The lower right quadrant showed some collinearity with Med. Area and Med. Shape Index, a strong one with RMI and some negative collinearity with Top-2 %, indicating these *clusters* and *patches* presented mostly landscapes of multiple medium sizes clearly defined segments of more than two different LCZ classes with probably a slightly dominant one. The lower left quadrant was strongly associated with Entropy and strongly negatively associated with CV of Area. This indicated *clusters* and *patches* with scattered landscapes of multiple segments of the same sizes with a balanced mix of LCZ classes such as LCZ-4, LCZ-5, LCZ-10, and LCZ-B. Last, the upper left quadrant presented collinearity with CV of Shape Index and Top-2 %, as well as strongly negative association with RMI indicating *clusters* and *patches* with speckled homogeneous sub-regions of different complexities with generally two LCZ classes sharing most of the landscape. In general, we observed that the right side of these biplots showed more homogeneous *clusters* and *patches* while the left side exhibited more heterogeneous *clusters* and *patches*.

6. Discussion

The volumes and voids, the coverage, height and density of the built-up and vegetation vary greatly within the urban fabric. This renders the urban fabric extremely textured and richly heterogeneous. Previous studies showed the relevance to analyze types of urban fabric in their relations to urban functions and indirect physical socio-economic aspects (Arribas-Bel and Fleischmann, 2022; Batty and Longley, 1994; Wandl and Hausleitner, 2021; Wang et al., 2023; Whitehand, 2001),

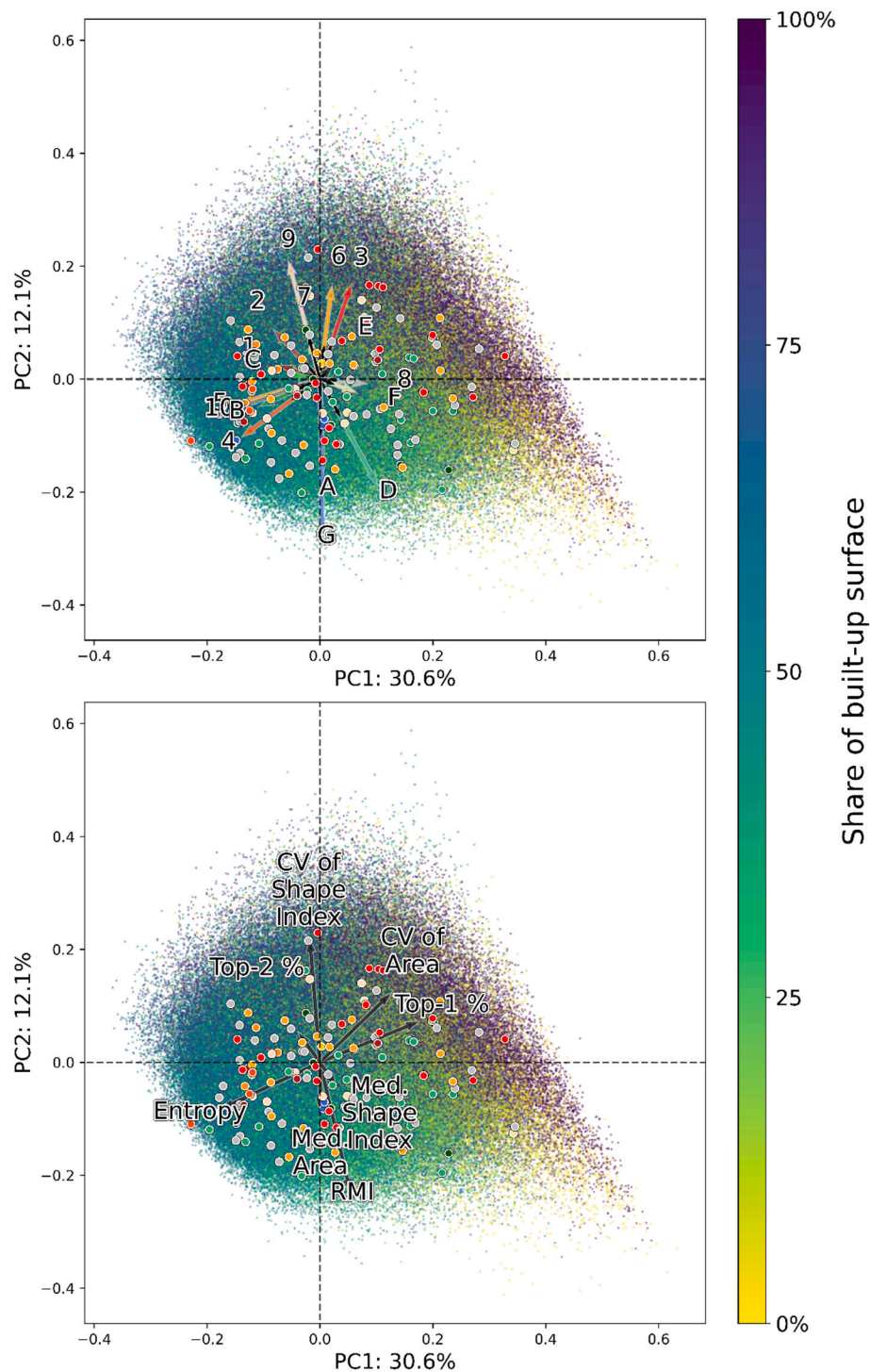


Fig. 6. Biplots of the patches and the centroids of the clusters identified according to the two first principal components of the landscape metrics feature space. The first biplot focuses on the share of the LCZ classes and the second one on the landscape metrics. The loadings (length of the arrows) of the shares of the different LCZ classes have been magnified by a factor 3 in the first biplot for ease of reading (inner small black arrows are to scale). Each background point in the biplot corresponds to a patch of which color signifies its share of built-up surface. Each foreground point corresponds to the patch closest to the centroid of its cluster, in the embedded space of the neural network. Its color corresponds to the predominant LCZ class in its cluster.

their historical processes (Cozzolino, 2020; Debray et al., 2023; Dovey, 2020; Fleischmann et al., 2021a) the inequal distribution of different types of urban fabric across the globe (Debray et al., 2021; Lemoine-Rodríguez et al., 2020; Taubenböck et al., 2020; Zhu et al., 2022), or within the cities themselves (Adams, 2005; Fleischmann et al., 2021a; R. E. Park et al., 1925). Yet, how rich is the urban fabric, how many faces does it present at a global level are questions that were not consistently

addressed until now and kept largely unsatisfactory unanswered. Without answering these questions, any attempt at global, consistent comparisons between cities on urban phenomena related to their intra-urban morphology are arduous and uncertain, if not impossible.

The emergence, over the last decade, of global datasets and powerful data-driven methods enable now to steer towards systematic and quantitative approaches allowing empirical analysis and proof over

observations on urban morphology. Leveraging these, we developed in this study a consistent method to derive a global and complete (i.e. universal) typology of the urban fabric across the globe by identifying *clusters* of similar intra-urban patterns among *patches* of urban fabric. We proved that the *clusters* identified are statistically different in terms of compositions and configurations of their *patches*, showing well defined types of intra-urban patterns as opposed to a continuum. With this, we prove that a universal typology of intra-urban patterns is attainable, and we provide the first insights into it.

We found that the diversity of patterns within the urban fabric can be summarized at a global level in 138 intra-urban pattern types. This result means that any large city on our globe can be, in an abstract view, seen as a complex jigsaw puzzle of which each and any pieces is drawn from this pool of only 138 types we identified. We find that different interpretations of this number of 138 intra-urban pattern types can make it seem very low or, conversely, high. In face of all combinations of compositions and configuration that we could imagine for the elementary unit *patches* at the base of our method (17^{322}), the fact that 138 types of intra-urban patterns are identified speaks for a great regularity of pattern types across the globe. At the same time, 138 types reflects a great richness of a nuanced typology as some of these types are accounting for as little as less than 0.05 % of the urban fabric of the world and the most prevalent type accounts for only 1.8 % of the global fabric. Further, through the analysis of the specific morphological characteristics of the typical landscapes of the *patches* associated to these types, we demonstrated the great variability existing between these 138 types. Therefore, the universal typology of intra-urban patterns here identified allows for a nuanced representation of the complexity and diversity of *meso*-scaled patterns existing throughout the global urban fabric.

Although diverse, we further found that most of this diversity is dominated by one specific built-up type: LCZ-8 corresponding to large low-rise buildings and one landcover type: LCZ-D corresponding to very low vegetation. ‘LCZ-8: large low-rise buildings’ are typical of industrial and commercial parks among others and are found to amount to 18.9 % of the global urban fabric accounting for more than a fourth of the built-up surface of the global urban fabric. ‘LCZ-D: low plants’ is associated with extents of grass in urban parks and with peripheral or interstitial fragmented herbaceous areas and we find it to amount to 18.6 % of the global urban fabric, making it account for more than the half of the non-built-up part of the global urban fabric. What’s more, we find these two LCZ classes to be correlated. On an even more generalized level we further observe that built-up forms of low-density cover around half of the global urban fabric and is predominant in half of the patterns we found. This information corroborates with previous studies (Hu et al., 2021; Taubenböck, 2021; Taubenböck et al., 2025) and clearly identifies that the current way our cities are being built, across the globe, is prevalently steered towards low density forms and low urban and *peri*-urban vegetation typical result from the urban sprawl. These over-present patterns of the current urbanization actively foster ineffective land consumption (Hu et al., 2021; Taubenböck et al., 2025), specific nefarious mobility patterns (Rode et al., 2017), lower ecosystem services (Chang et al., 2017) among many other issues.

The topology we developed is broadly applicable in urban planning, particularly at the *meso*-scale where many urban phenomena—such as urban heat islands, air pollution dispersion, and accessibility barriers—are structured. By identifying and characterizing 138 intra-urban pattern types, we offer a consistent ground for future research on several urban challenges. It is our aim that this typology can be used as the basis for cross-city and cross-type comparisons, supporting a better understanding of how different urban fabrics perform, therefore informing more effective planning practices.

We overall see the results of this study as proving the capacity of our methodology to find a universal typology of the urban fabric. The use of an unsupervised clustering approach combined with a large global dataset proves to be highly promising for identifying patterns of the urban fabric. With this, we hope to inspire new comparative global

investigations of the urban fabric. Nonetheless, against these positive results, we account for a few shortcomings that need to be considered in the interpretation of our result.

Despite their high accuracy, the LCZ classification and MUAs used in this study contain uncertainties that may impact the resulting typology. The LCZ classification presents uneven levels of confusion across categories (Zhu et al., 2022). Common misclassifications include LCZ-3: compact low-rise with LCZ-7: lightweight low-rise; LCZ-8: large low-rise and LCZ-10: heavy industry; and last LCZ-C: bush, scrubs with LCZ-D: low plants and LCZ-F: bare soil or sand. While these differences are semantically small, they may introduce bias in our input patches. While our method is robust to noise, a larger bias in the input data could have as consequence the creation of spurious cluster. This specific bias-induced error could not be quantified in this study. Comparison with other LCZ classifications (e.g. (Demuzere et al., 2022)) could potentially help quantify this error. Future studies using the ordinal values of the LCZ as opposed to categorical values could be, on their own, able to circumvent this potential error as the confused LCZ classes share these same ordinal characteristics. Further, the MUAs might have some influence on the typology we found. The MUAs include cities larger than 3000 000 inhabitants and smaller settlements in their direct vicinity. As smaller cities further away from major urban areas are not included in these datasets, patterns exclusive to small cities might be invisibilized in our study which cannot be here straightforwardly quantified. Further studies could investigate if some additional patterns are identified using alternative datasets (e.g.: (Corbane et al., 2019; Demuzere et al., 2022; Esch et al., 2017)). Further, it is to be acknowledged once more that the performance of the SCAN + RUC framework in the specific context of our study could not be quantified in term of accuracy. Our study, by design, could not be based on prior typologies and therefore other support was needed to test our results. Nevertheless, the statistical tests we led showed conclusive results that the clusters identified are indeed morphologically different and the qualitative analysis of the patterns further confirmed that the patterns identified are not only coherent but semantically different. On these grounds, the SCAN + RUC was well adapted to the task at hand and showcased the transferability of unsupervised computer vision frameworks in geographical analysis tasks as suggested in (W. Chen et al., 2024; P. Liu and Biljecki, 2022; Wang et al., 2024; Wang and Biljecki, 2022). Yet, these do not allow for a straightforwardly comparable evaluation of the performance of different methods. Therefore, in a frame of comparative research, we encourage future work to directly compare their typology obtained to the one here presented to better inform a baseline typology of intra-urban patterns.

In general, the typology framed by our unsupervised approach is not definitive in its number of types or in the details of its content. Therefore, the results of our study, at this stage, can be further investigated and improved.

7. Conclusion and outlook

It had been acknowledged for over a century that capturing the diversity of the urban fabric by drawing its typology is a relevant research direction. However, attempts at any universal typology have been prevented by a lack of comprehensive data and accordingly effective methods. Developments in data, methods and processing power open a new avenue for research in this domain.

Indeed, deriving empirically such a universal objective typology is a process paved with methodological pitfalls and technical hurdles. In recent years, advances in the fields of remote sensing and machine learning allowed to progressively circumvent these technical and methodological obstacles. Stepping on these advances, we adapted a deep learning unsupervised approach and applied it to comprehensive and systematic data used to proxy the urban fabric on a global scale. With this, we derived the first comprehensive universal typology of the intra-urban patterns produced empirically, i.e. based on more than 1,500 cities across the globe. Our data-driven approach identified 138

intra-urban pattern types representing characteristic landscape composition and configuration. Information gained from this typology allows to better understand the physical dimension of the current global urbanization. We believe that this result is one among many that a universal typology of intra-urban patterns can provide.

Going forward, we identify five main directions to develop the approach presented in this study. First, we see the semantic interpretation of each intra-urban pattern type proposed in this study as refinable. There is no claim here that the semantic interpretation we put forth for the 138 types of intra-urban patterns are fully representative of each of their instances across the globe. We encourage further refinement of the semantic description of each pattern type to present a unified pattern vocabulary as conceptualized for example in (Wu et al., 2025). (Arribas-Bel and Fleischmann, 2022; Fleischmann and Arribas-Bel, 2022) showcased that the range of patterns identified is only partially represented by supervised classification scheme of land use or urban functions. Therefore, we envision that a systematic approach should not only rely on land use categories or functional characteristics but should also be based in the location of the patterns in the cities (e.g., in the urban core or the urban fringe) and their direct spatial interactions with each other. As with the 138 types found here, the task of semantically describing each and every type identified is all the more arduous. In a fashion inspired by (Fleischmann et al., 2021a), a hierarchical approach could be used to summarize “families” of intra-urban pattern types at a generalized level and work from there toward finer levels of the semantic details of the typology, depending on research or planning goals. It is to be noted that hierarchical approaches are not as straightforward in the case of typologies derived from deep learning as for traditional feature spaces. Nonetheless recent advances in that field, specifically on hierarchical unsupervised clustering (Afchar et al., 2022), are promising and could be employed in the future.

Second, we believe that the comprehensiveness of the typology showcased in this study can be further extended. In our approach, we included cities with 300,000 inhabitants or more and all secondary cities or settlements in their direct vicinity. Although we are confident that our typology empirically captures most intra-urban pattern types, smaller urban areas are not captured in our data. These cities, due to their socio-economical or historical idiosyncrasies might exhibit pattern types that we did not identify in our study. Therefore, we encourage future studies to investigate the use of alternative datasets (e.g. (Corbane et al., 2019; Esch et al., 2017)) to include such cities.

Third, we are convinced and demonstrated that the explored SCAN + RUC framework proofed to be highly suited to detect patterns in our data. Yet we believe that future work can build upon our approach to propose even more performant methods. While the SCAN + RUC framework was the highest accuracy yielding method at the time of the study, development of new unsupervised approaches such as the TUR-TLE framework (Gadetsky et al., 2024) seem to significantly outperform it. If adequate technical adaptations would allow for the TUR-TLE to be operated in a fully unsupervised fashion (including without prior knowledge on the number of labels), we are convinced its results on the task developed in this study would achieve new unprecedented levels of confidence.

Fourth, while the results of this study represent a rich snapshot of the state of the global urban fabric, multitemporal LCZ data does not yet exist. Once available, our framework could be applied to assess the global evolution of intra-urban patterns over time. We encourage future research expanding on the present study to use arising multi-temporal

datasets to further develop understanding of the dynamics of the intra-urban morphology (Wentz et al., 2018) and of its diversity (Chakraborty et al., 2024; Lemoine-Rodríguez et al., 2020).

Fifth and finally, while the typology drawn in this study presents insights on the existing diversity of patterns in the urban fabric at a global scale, we envision this typology to be used in applied research. The typology here presented and the methodology to draw can be used as base input for domain specific applications. We encourage further studies to use the intra-urban pattern types identified as a scheme and to cross it with a wide range of spatial datasets (e.g.: socio-economic factors, housing capacities, urban heat islands, pollution, walkability, access to vegetation, access to amenities, health data, risk exposure...) to cross-examine how urban phenomena differ across different types of urban fabric. We believe that such groundwork could vastly improve our understanding of performances of certain urban planning practices in regard to current global challenges faced by cities (United Nations, 2015).

With this approach, we provide the field of urban morphology with new methodological ground and empirical bases for cross-city and cross-regional comparative analysis of the urban fabric and its pertaining phenomena. We see this study as a precursor for further work in the same direction with improved technical capacities, built-upon methodology, larger data coverage of higher accuracies or richer data basis. Further, with this work, we aim to trigger a range of geographical investigations that could provide new types of empirical insights on the sustainability of different types of urban fabric, on their geographical distributions at a global level or on the way cities are structured. In this perspective, we encourage future studies to employ this typology to investigate the urban fabric from the intra-urban scale to the global scale.

CRedit authorship contribution statement

Henri Debray: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Matthias Gassilloud:** Writing – review & editing, Methodology, Investigation, Data curation. **Richard Lemoine-Rodríguez:** Writing – review & editing, Visualization, Methodology. **Michael Wurm:** Writing – review & editing, Supervision. **Xiaoxiang Zhu:** Writing – review & editing, Supervision, Funding acquisition. **Hannes Taubenböck:** Writing – original draft, Visualization, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors want to thank Lukas Müller for his help in the exploration of the connection between intra-urban pattern types and landscape metrics.

This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No [714087]-So2Sat).

Appendix A:. List of metrics

Metric nr.	Metric name	Formula	Description	Type
General composition of the patches				
	Share of LCZ-1	a_{LCZ-1}/a_{patch} With a_{LCZ-1} being the area covered by the LCZ-1 class in the patch and a_{patch} being the area of the patch	The proportion of a patch that pixels of class LCZ-1, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-2	a_{LCZ-2}/a_{patch}	The proportion of a patch that pixels of class LCZ-2, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-3	a_{LCZ-3}/a_{patch}	The proportion of a patch that pixels of class LCZ-3, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-4	a_{LCZ-4}/a_{patch}	The proportion of a patch that pixels of class LCZ-4, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-5	a_{LCZ-5}/a_{patch}	The proportion of a patch that pixels of class LCZ-5, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-6	a_{LCZ-6}/a_{patch}	The proportion of a patch that pixels of class LCZ-6, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-7	a_{LCZ-7}/a_{patch}	The proportion of a patch that pixels of class LCZ-7, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-8	a_{LCZ-8}/a_{patch}	The proportion of a patch that pixels of class LCZ-8, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-9	a_{LCZ-9}/a_{patch}	The proportion of a patch that pixels of class LCZ-9, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-10	a_{LCZ-10}/a_{patch}	The proportion of a patch that pixels of class LCZ-10, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-A	a_{LCZ-A}/a_{patch}	The proportion of a patch that pixels of class LCZ-A, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-B	a_{LCZ-B}/a_{patch}	The proportion of a patch that pixels of class LCZ-B, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-C	a_{LCZ-C}/a_{patch}	The proportion of a patch that pixels of class LCZ-C, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-D	a_{LCZ-D}/a_{patch}	The proportion of a patch that pixels of class LCZ-D, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-E	a_{LCZ-E}/a_{patch}	The proportion of a patch that pixels of class LCZ-E, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-F	a_{LCZ-F}/a_{patch}	The proportion of a patch that pixels of class LCZ-F, between 0 (none) and 1 (fully).	Continuous
	Share of LCZ-G	a_{LCZ-G}/a_{patch}	The proportion of a patch that pixels of class LCZ-G, between 0 (none) and 1 (fully).	Continuous
Dominant LCZ classes in the patches				
	Top 1 LCZ class	$lcz_k \setminus \forall lcz_i, a_{lcz_i} \geq a_{lcz_k}$ With lcz_i being all the LCZ classes	The LCZ class with the highest share in the patch	Categorical
	Top 2 LCZ class	$lcz_l \setminus \forall lcz_i \setminus lcz_k, a_{lcz_i} \geq a_{lcz_l}$	The LCZ class with the second highest share in the patch	Categorical
	Top 3 LCZ class	$lcz_m \setminus \forall lcz_i \setminus lcz_k, l, a_{lcz_m} \geq a_{lcz_i}$	The LCZ class with the third highest share in the patch	Categorical
	Top 4 LCZ class	$lcz_n \setminus \forall lcz_i \setminus lcz_k, l, m, a_{lcz_n} \geq a_{lcz_i}$	The LCZ class with the fourth highest share in the patch	Categorical
	Top 5 LCZ class	$lcz_p \setminus \forall lcz_i \setminus lcz_k, l, m, n, a_{lcz_p} \geq a_{lcz_i}$	The LCZ class with the fifth highest share in the patch	Categorical
Importance of the dominant LCZ classes in the patches				
	Share of Top 1 LCZ class	$a_{Top1LCZ}/a_{patch}$	The proportion of a patch that pixels of the Top 1 LCZ class in the patch, between 0 (none) and 1 (fully).	Continuous
	Share of Top 2 LCZ class	$a_{Top2LCZ}/a_{patch}$	The proportion of a patch that pixels of the Top 2 LCZ class in the patch, between 0 (none) and 1 (fully).	Continuous
	Share of Top 3 LCZ class	$a_{Top3LCZ}/a_{patch}$	The proportion of a patch that pixels of the Top 3 LCZ class in the patch, between 0 (none) and 1 (fully).	Continuous
	Share of Top 4 LCZ class	$a_{Top4LCZ}/a_{patch}$	The proportion of a patch that pixels of the Top 4 LCZ class in the patch, between 0 (none) and 1 (fully).	Continuous
	Share of Top 5 LCZ class	$a_{Top5LCZ}/a_{patch}$	The proportion of a patch that pixels of the Top 5 LCZ class in the patch, between 0 (none) and 1 (fully).	Continuous
	Number of segments of Top 1 LCZ class	$ \{S_{Top1LCZ}\} $ With $\{S_{Top1LCZ}\}$ being the set of segments (i.e. contiguously homogeneous area of one specific class) of the Top 1 LCZ class	The number of segments belonging to the Top 1 LCZ class.	Continuous
	Number of segments of Top 2 LCZ class	$ \{S_{Top2LCZ}\} $	The number of segments belonging to the Top 2 LCZ class.	Continuous
	Number of segments of Top 3 LCZ class	$ \{S_{Top3LCZ}\} $	The number of segments belonging to the Top 3 LCZ class.	Continuous
	Number of segments of Top 4 LCZ class	$ \{S_{Top4LCZ}\} $	The number of segments belonging to the Top 4 LCZ class.	Continuous
	Number of segments of Top 5 LCZ class	$ \{S_{Top5LCZ}\} $	The number of segments belonging to the Top 5 LCZ class.	Continuous

(continued on next page)

(continued)

Metric nr.	Metric name	Formula	Description	Type
	Share of the largest segment of Top 1 LCZ class	$\max(\{a_{S_{Top1LCZ}}\})/a_{patch}$	The proportion of the patch comprised by the largest segment of the Top 1 LCZ class, between 0 (none) and 1 (fully).	Continuous
	Share of the largest segment of Top 2 LCZ class	$\max(\{a_{S_{Top2LCZ}}\})/a_{patch}$	The proportion of the patch comprised by the largest segment of the Top 2 LCZ class, between 0 (none) and 1 (fully).	Continuous
	Share of the largest segment of Top 3 LCZ class	$\max(\{a_{S_{Top3LCZ}}\})/a_{patch}$	The proportion of the patch comprised by the largest segment of the Top 3 LCZ class, between 0 (none) and 1 (fully).	Continuous
	Share of the largest segment of Top 4 LCZ class	$\max(\{a_{S_{Top4LCZ}}\})/a_{patch}$	The proportion of the patch comprised by the largest segment of the Top 4 LCZ class, between 0 (none) and 1 (fully).	Continuous
	Share of the largest segment of Top 5 LCZ class	$\max(\{a_{S_{Top5LCZ}}\})/a_{patch}$	The proportion of the patch comprised by the largest segment of the Top 5 LCZ class, between 0 (none) and 1 (fully).	Continuous
Morphologies of the dominant LCZ classes in the patches				
	Shape index of segments of Top 1 LCZ class	$\frac{\sum \{e_{S_{Top1LCZ}}\}}{4\sqrt{a_{patch}}}$ With $\{e_{S_{Top1LCZ}}\}$ being the set of all the edges length of segments of Top 1 LCZ class in the patch.	Standardized measure of the cumulated length of edges the segments of Top 1 LCZ class shares with the rest of the patch. It is equal to 1 when the entire patch is comprised of the Top 1 LCZ class and increases without limit as the segments of Top 1 LCZ class become more disaggregated.	Continuous
	Shape index of segments of Top 2 LCZ class	$\frac{\sum \{e_{S_{Top2LCZ}}\}}{4\sqrt{a_{patch}}}$	Standardized measure of the cumulated length of edges the segments of Top 2 LCZ class shares with the rest of the patch. It is equal to 1 when the entire patch is comprised of the Top 2 LCZ class and increases without limit as the segments of Top 2 LCZ class become more disaggregated.	Continuous
	Shape index of segments of Top 3 LCZ class	$\frac{\sum \{e_{S_{Top3LCZ}}\}}{4\sqrt{a_{patch}}}$	Standardized measure of the cumulated length of edges the segments of Top 3 LCZ class shares with the rest of the patch. It is equal to 1 when the entire patch is comprised of the Top 3 LCZ class and increases without limit as the segments of Top 3 LCZ class become more disaggregated.	Continuous
	Shape index of segments of Top 4 LCZ class	$\frac{\sum \{e_{S_{Top4LCZ}}\}}{4\sqrt{a_{patch}}}$	Standardized measure of the cumulated length of edges the segments of Top 4 LCZ class shares with the rest of the patch. It is equal to 1 when the entire patch is comprised of the Top 4 LCZ class and increases without limit as the segments of Top 4 LCZ class become more disaggregated.	Continuous
	Shape index of segments of Top 5 LCZ class	$\frac{\sum \{e_{S_{Top5LCZ}}\}}{4\sqrt{a_{patch}}}$	Standardized measure of the cumulated length of edges the segments of Top 5 LCZ class shares with the rest of the patch. It is equal to 1 when the entire patch is comprised of the Top 5 LCZ class and increases without limit as the segments of Top 5 LCZ class become more disaggregated.	Continuous
	Effective Mesh size of Top 1 LCZ class	$\frac{\sum \{a_{S_{Top1LCZ}}^2\}}{a_{patch}}$	Measure of the aggregation of the Top 1 LCZ class based on its cumulative segments size distribution in the patch.	Continuous
	Effective Mesh size of Top 2 LCZ class	$\frac{\sum \{a_{S_{Top2LCZ}}^2\}}{a_{patch}}$	Measure of the aggregation of the Top 2 LCZ class based on its cumulative segments size distribution in the patch.	Continuous
	Effective Mesh size of Top 3 LCZ class	$\frac{\sum \{a_{S_{Top3LCZ}}^2\}}{a_{patch}}$	Measure of the aggregation of the Top 3 LCZ class based on its cumulative segments size distribution in the patch.	Continuous
	Effective Mesh size of Top 4 LCZ class	$\frac{\sum \{a_{S_{Top4LCZ}}^2\}}{a_{patch}}$	Measure of the aggregation of the Top 4 LCZ class based on its cumulative segments size distribution in the patch.	Continuous
	Effective Mesh size of Top 5 LCZ class	$\frac{\sum \{a_{S_{Top5LCZ}}^2\}}{a_{patch}}$	Measure of the aggregation of the Top 5 LCZ class based on its cumulative segments size distribution in the patch.	Continuous
Diversity and aggregation of LCZ classes in the patches				
	Entropy	$-\sum_{LCZ_i} P_{LCZ_i} \log_b(P_{LCZ_i})$ With P_{LCZ_i} being the proportion of pixels in the landscape belonging to the LCZ class LCZ_i	Measures the LCZ composition diversity within the patch by considering the number of LCZ classes present and their relative abundance.	Continuous
	Shannon Diversity index	$-\sum_{LCZ_i} P_{LCZ_i} \ln(P_{LCZ_i})$	Measures the LCZ composition diversity within the patch by considering the number of LCZ classes present and their relative abundance.	Continuous

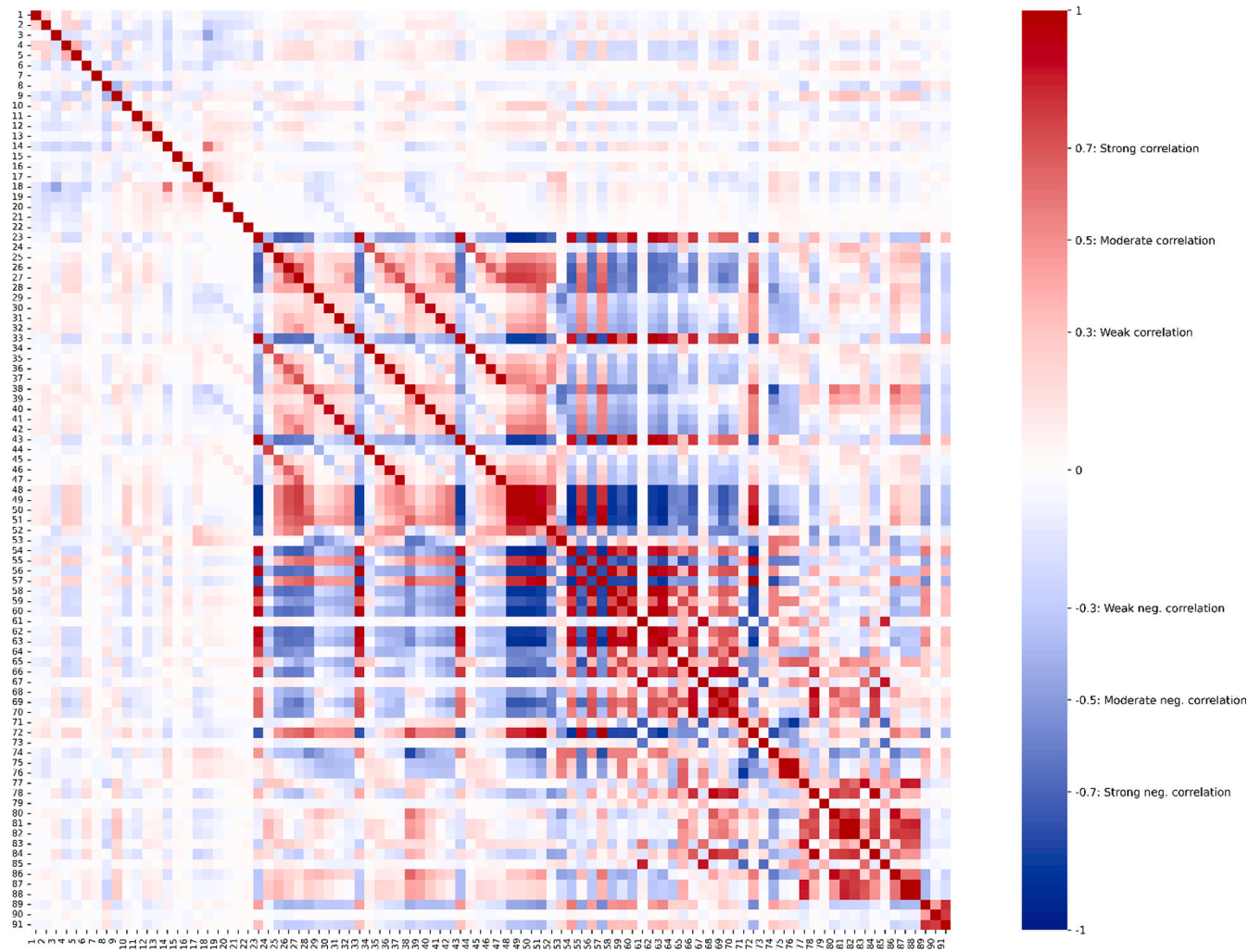
(continued on next page)

(continued)

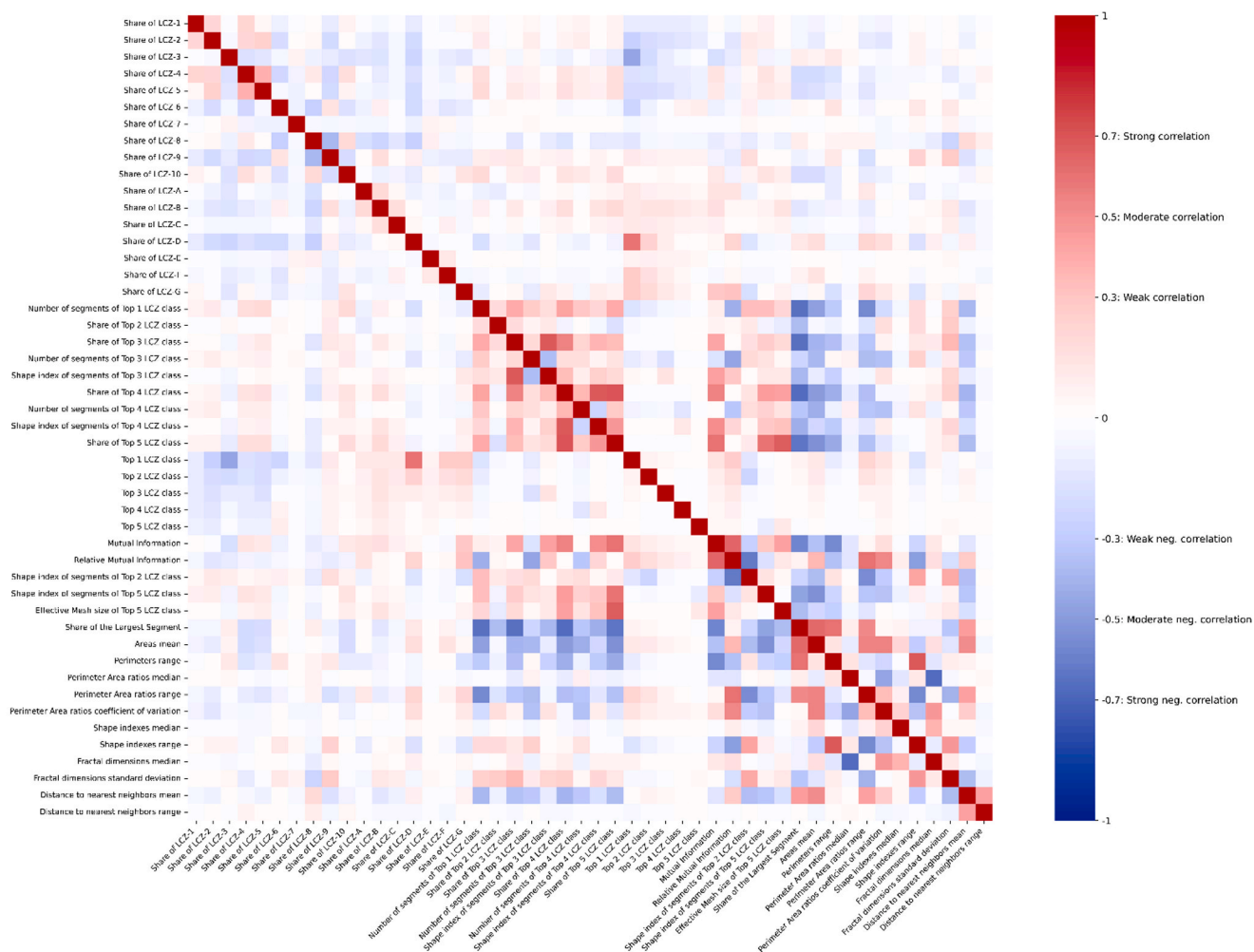
Metric nr.	Metric name	Formula	Description	Type	
	Joint Entropy	$-\sum_{lczi \in lcZ} \sum_{lczk \in lcZ} \left[P_{lczi} \frac{g_{lczi,k}}{\sum_{lczk \in lcZ} g_{lczi,k}} \right] \left[\log_b \left(P_{lczi} \frac{g_{lczi,k}}{\sum_{lczk \in lcZ} g_{lczi,k}} \right) \right]$ <p>With $g_{lczi,k}$ being the number of pixels of class lcz_k neighboring pixels of LCZ class $lczi$</p>	Measures the composition and configuration complexity of the patch by computing the frequency to which two adjacent pixels belong to the same LCZ class.	Continuous	
	Conditional Entropy	$-\sum_{lczi \in lcZ} \sum_{lczk \in lcZ} \left[P_{lczi} \frac{g_{lczi,k}}{\sum_{lczk \in lcZ} g_{lczi,k}} \right] \left[\log_b \left(\frac{g_{lczi,k}}{\sum_{lczk \in lcZ} g_{lczi,k}} \right) \right]$	Measures the configuration complexity of the patch in reflecting the spatial interspersions of the LCZ classes.	Continuous	
	Mutual Information	Entropy – ConditionalEntropy	Measures the aggregation in the patch by computing the difference between the composition complexity and the configuration complexity.	Continuous	
	Relative Mutual Information	$\frac{MutualInformation}{Entropy}$	Measures the aggregation by adjusting the Mutual information for the number of LCZ classes present in the patch.	Continuous	
	Contagion	$1 + \frac{\sum_{lczi} \sum_{lczk} \left[P_{lczi} \frac{g_{lczi,k}}{\sum_{lczk \in lcZ} g_{lczi,k}} \right] \left[\ln \left(P_{lczi} \frac{g_{lczi,k}}{\sum_{lczk \in lcZ} g_{lczi,k}} \right) \right]}{2 \ln(m)}$ <p>With m being the number of LCZ classes present in the patch</p>	Measures the aggregation by computing the frequency to which two adjacent pixels belong to the same LCZ class.	Continuous	
Landscape morphology of patches					
	Number of Segments	$ \{S\} $ With $\{S\}$ being the set of all segments in the patch	The number of segments in the patch.	Continuous	
	Share of the Largest Segment	$\max(\{a_s\})/a_{patch}$ With $\{a_s\}$ being the set of the area values all segments in the patch	The proportion of the patch comprised by its largest segment, between 0 (none) and 1 (fully).	Continuous	
	Shape Index	$\frac{\sum \{e_s\}}{4\sqrt{a_{patch}}}$ With $\{e_s\}$ being the set of the edges length counted once between all segments in the patch	Standardized measure of the cumulated length of edges the segments of the patch. It is equal to 1 when the entire patch is comprised of the Top 5 LCZ class and increases without limit as the segments of Top 5 LCZ class become more disaggregated.	Continuous	
	Effective Mesh Size	$\frac{\sum \{a_s^2\}}{a_{patch}}$	Measure of the aggregation of the Top 1 LCZ class based on its cumulative segments size distribution in the patch.	Continuous	
Morphology of the segments within patches					
(59–64)	Areas	$\{a_s\}$	On which values are computed: The mean (mn) The area-weighted mean (am) The median (md) The range (ra) The standard deviation (sd) The coefficient of variation (cv)	Areas values of each segment of the patch.	Continuous
(65–70)	Perimeters	$\{e_s\}$	(1) mn (2) am (3) md (4) ra (5) sd (6) cv	Edges length of each segment of the patch.	Continuous
(71–76)	Perimeter Area ratios	$\{e_s/a_s\}$	(7) mn (8) am (9) md (10) ra (11) sd (12) cv	Perimeter Area ratios of each segment of the patch as a measure of shape complexity.	Continuous
(77–82)	Shape indexes	$\left\{ \frac{e_s}{4\sqrt{a_{patch}}} \right\}$	(13) mn (14) am (15) md (16) ra (17) sd (18) cv	Shape indexes of each segment of the patch as a standardized measure of shape complexity.	Continuous
(83–88)	Fractal dimensions	$\left\{ 2 \frac{\ln(\frac{e_s}{4})}{\ln(a_s)} \right\}$	(19) mn (20) am (21) md (22) ra (23) sd (24) cv	Fractal dimensions values of each segment of the patch as a measure of shape complexity.	Continuous
(89–91)	Scattering of the LCZ classes Distance to nearest neighbors	$\{d_s\}$	(25) mn (26) ra (27) sd	Distances to the nearest segment of the same LCZ class for each LCZ class as a measure of their scatterings.	Continuous

Appendix B:. Correlation matrices

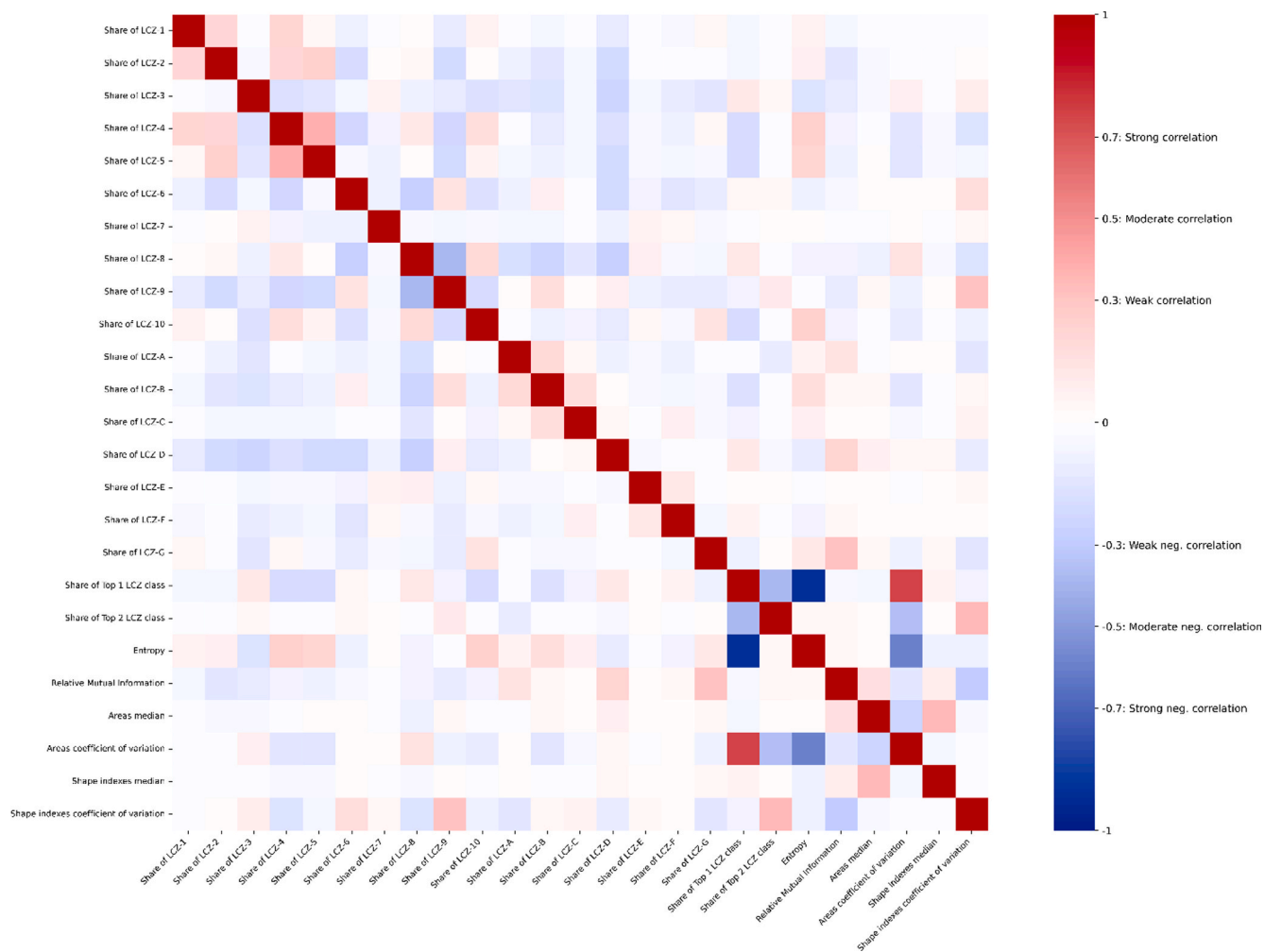
1) Correlation matrix of all landscape metrics:



2) Correlation matrix of not strongly correlated landscape metrics:









































































3) Correlation matrix of explicative landscape metrics:

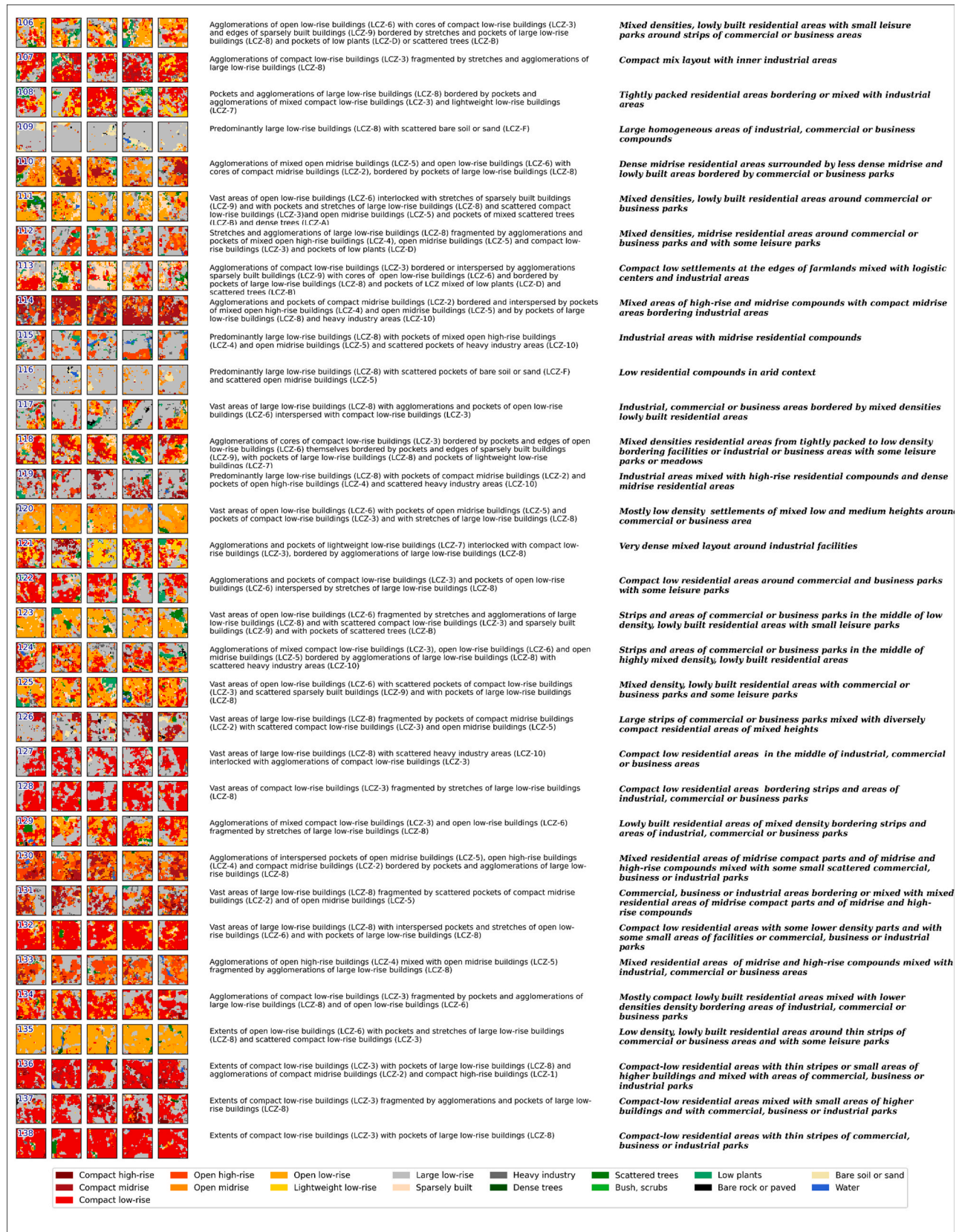


Appendix C.: Qualitative analyses of the 138 intra-urban pattern types

	Predominantly low plants (LCZ-D) with encroachments of bare soil or sand (LCZ-F)	<i>Agricultural land in arid context</i>
	Extents of low plants (LCZ-D) around rivers of water (LCZ-G) with some agglomerations of compact low-rise buildings (LCZ-3), open low-rise buildings (LCZ-6) or large low-rise buildings (LCZ-8)	<i>Plains or agricultural lands on river shores with few scattered small settlements</i>
	Predominantly bare soil or sand (LCZ-F) with some agglomerations of large low-rise buildings (LCZ-8)	<i>Desert plains with industrial encroachments</i>
	Predominantly forests of dense trees (LCZ-A) bordered by rims of bush or scrub areas (LCZ-C) and low plants (LCZ-D) and agglomeration of open low-rise buildings (LCZ-6)	<i>Dense forests bordering settlements</i>
	Extents of low plants (LCZ-D) with agglomerations of dense trees (LCZ-A) bordered by scattered trees (LCZ-B) and agglomerations of sparsely built buildings (LCZ-9)	<i>Scattered woods encroached by very low density settlements</i>
	Extents of low plants (LCZ-D) with pockets of compact low-rise buildings (LCZ-3)	<i>Agricultural extents with small villages</i>
	Extents of low plants (LCZ-D) with agglomerations of large low-rise buildings (LCZ-8) and sparsely built buildings (LCZ-9)	<i>Plains or agricultural lands next to industrial areas or commercial and business parks</i>
	Extents of low plants (LCZ-D) with agglomerations of compact low-rise buildings (LCZ-3) surrounded by open low-rise buildings (LCZ-6) and sparsely built buildings (LCZ-9)	<i>Plains or agricultural lands at the edge of lowly built settlements</i>
	Rivers and extents of water (LCZ-G) surrounded by forests of dense trees (LCZ-A) and Extents of LCZ of low plants (LCZ-D) with agglomerations of open low-rise buildings (LCZ-6) and sparsely built buildings (LCZ-9) on the shores	<i>Small settlements on the shores of rivers at the edge of forests</i>
	Fragmented extents of low plants (LCZ-D) mixed with bare soil or sand (LCZ-F) with some agglomerations of sparsely built buildings (LCZ-9)	<i>Fragmented agricultural lands or plains in arid context and at the edge of settlements</i>
	Thin rivers of water (LCZ-G) surrounded by Extents of low plants (LCZ-D) and agglomerations of large low-rise buildings (LCZ-8) on and across shores	<i>Industrial or commercial parks on the shores of small rivers, surrounded by plains or agricultural land</i>
	Extents of low plants (LCZ-D) with pockets of sparsely built buildings (LCZ-9) scattered and few agglomerations of large low-rise buildings (LCZ-8)	<i>Scattered clusters of low density villages in agricultural lands with some business parks or industrial areas</i>
	Extents of bare soil or sand (LCZ-F) with scattered low plants (LCZ-D) with agglomerations of large low-rise buildings (LCZ-8) bordering agglomerations of compact midrise buildings (LCZ-2), compact low-rise buildings (LCZ-3), open midrise buildings (LCZ-5) and open low-rise buildings (LCZ-6)	<i>Arid areas at the edges of settlements</i>
	Forests of dense trees (LCZ-A) bordered with scattered trees (LCZ-B) and low plants (LCZ-D) bordered with agglomerations of large low-rise buildings (LCZ-8) and of sparsely built buildings (LCZ-9) and compact low-rise buildings (LCZ-3)	<i>Forests bordered by industrial or commercial part of settlements</i>
	Extents of low plants (LCZ-D) with agglomerations of large low-rise buildings (LCZ-8) with pockets of compact low-rise buildings (LCZ-3)	<i>Agricultural lands in vicinity of business parks or industrial areas with some compact-low habitations</i>
	Extents of low plants (LCZ-D) with agglomerations of large low-rise buildings (LCZ-8) mixed with agglomerations and pockets of compact low-rise buildings (LCZ-3) surrounded by sparsely built buildings (LCZ-9)	<i>Scattered lowly built villages and small commercial areas surrounded by plains or agricultural land</i>
	Mix of low plants (LCZ-D) and bare soil or sand (LCZ-F) with agglomerations of large low-rise buildings (LCZ-8) and pockets of sparsely built buildings (LCZ-9)	<i>Industrial or commercial parks at the edge of agricultural or undeveloped land</i>
	Extents of low plants (LCZ-D) with agglomerations and pockets of sparsely built buildings (LCZ-9) with cores of compact low-rise buildings (LCZ-3) and open low-rise buildings (LCZ-6)	<i>Edges of low density settlements and farmlands</i>
	Extents of low plants (LCZ-D) with agglomerations of LCZ 6 surrounded by sparsely built buildings (LCZ-9)	<i>Low density villages surrounded by farmlands</i>
	Forests of dense trees (LCZ-A) bordered with agglomerations of open low-rise buildings (LCZ-6) with cores of compact low-rise buildings (LCZ-3) or compact midrise buildings (LCZ-2)	<i>Industrial/commercial and residential lowly built part of settlements at the edge of dense forests</i>
	Extents of low plants (LCZ-D) fragmented with agglomerations and pockets of sparsely built buildings (LCZ-9) with cores of open low-rise buildings (LCZ-6) and inner core of compact low-rise buildings (LCZ-3)	<i>Lowly built edge of small settlements encroaching on unbuilt surrounding</i>
	Rivers and extents of water (LCZ-G) surrounded with agglomerations of heavy industry areas (LCZ-10) and large low-rise buildings (LCZ-8) on the shores with Extents of low plants (LCZ-D)	<i>Light and heavy industrial fluvial harbors</i>
	Stretches of bare rock or paved (LCZ-E) in large low-rise buildings (LCZ-8), low plants (LCZ-D) or bare soil or sand (LCZ-F)	<i>Major infrastructures in commercial, business or industrial areas</i>
	Extents of low plants (LCZ-D) fragmented with agglomerations of large low-rise buildings (LCZ-8) and pockets of sparsely built buildings (LCZ-9) with cores of open low-rise buildings (LCZ-6)	<i>Small clusters of low density habitations bordering industrial or commercial parks at the edge of meadows or fragmented agricultural lands</i>
	Forests of dense trees (LCZ-A) bordered by agglomerations of large low-rise buildings (LCZ-8) and agglomerations of open high-rise buildings (LCZ-4)	<i>Mixed zones of highly built housing estates with commercial, industrial areas at the edge of forests</i>
	Extents of bare soil or sand (LCZ-F) with agglomerations of large low-rise buildings (LCZ-8) and scattered fragmented pockets of lightweight low-rise buildings (LCZ-7)	<i>Industrial or commercial areas bordering arid or low productivity agricultural lands</i>
	Extents of low plants (LCZ-D) around bodies of water (LCZ-G) with agglomerations of mixed open high-rise buildings (LCZ-4), open midrise buildings (LCZ-5) and open low-rise buildings (LCZ-6) neighboring agglomerations of large low-rise buildings (LCZ-8)	<i>Mixed zones of large housing estates with commercial, industrial areas at the edge of meadows or agricultural lands on the side of rivers or lakes</i>
	Extents of low plants (LCZ-D) fragmented by agglomerations of large low-rise buildings (LCZ-8) with pockets of compact low-rise buildings (LCZ-3)	<i>Small residential settlements at the edge of large industrial areas</i>
	Rivers of water (LCZ-G) bordered by agglomerations of large low-rise buildings (LCZ-8), Extents of low plants (LCZ-D) and agglomerations of open low-rise buildings (LCZ-6) with sparsely built buildings (LCZ-9) edges	<i>Low density residential areas on the edge to commercial, business or industrial areas on the shores of rivers or lakes with meadows or farmlands attached</i>
	Extents of low plants (LCZ-D) with agglomerations of compact low-rise buildings (LCZ-3) bordered by open low-rise buildings (LCZ-6) and edges of sparsely built buildings (LCZ-9), and pockets of large low-rise buildings (LCZ-8)	<i>Farm communities with logistic centers</i>
	Extents of low plants (LCZ-D) fragmented by agglomerations of large low-rise buildings (LCZ-8) with pockets of compact low-rise buildings (LCZ-3)	<i>Small compact, lowly built settlements surrounded by commercial, business or industrial area at the edge of meadows or agricultural lands</i>
	Extents of water (LCZ-G) bordered by large low-rise buildings (LCZ-8) at the edge of agglomerations of mixed compact midrise buildings (LCZ-2), compact high-rise buildings (LCZ-1) and open high-rise buildings (LCZ-4)	<i>Business and administrative districts or mixed highly built residential and commercial areas on shores</i>
	Extents of low plants (LCZ-D) with scattered bare soil or sand (LCZ-F) with agglomerations of compact low-rise buildings (LCZ-3) bordered by large low-rise buildings (LCZ-8)	<i>Lowly built residential areas next to business parks or industrial areas at the edge of undeveloped lands</i>
	Thin rivers of water (LCZ-G) surrounded by agglomerations of sparsely built buildings (LCZ-9) with cores of open low-rise buildings (LCZ-6)	<i>Low density residential areas on the edge to commercial areas or industrial harbors on the shores of rivers or lakes</i>
	Extents of sparsely built buildings (LCZ-9) with scattered open low-rise buildings (LCZ-6) and fragmented by forests of mixed dense trees (LCZ-A) and scattered trees (LCZ-B) and low plants (LCZ-D) clearings	<i>Sparsely built residential developments integrated in the fringe of woodlands</i>

																																			<p>Agglomerations of compact low-rise buildings (LCZ-3) mixed with lightweight low-rise buildings (LCZ-7) bordered by low plants (LCZ-D)</p> <p>Agglomerations of large low-rise buildings (LCZ-8) fragmented by extents of bare soil or sand (LCZ-F) and some water (LCZ-G) bodies</p> <p>Extents of low plants (LCZ-D) fragmented by thin rivers of water (LCZ-G) and scattered agglomerations of mix of large low-rise buildings (LCZ-8), sparsely built buildings (LCZ-9), compact low-rise buildings (LCZ-3) and open high-rise buildings (LCZ-4)</p> <p>Rivers of water (LCZ-G) bordered by agglomerations of compact low-rise buildings (LCZ-3) mixed with open low-rise buildings (LCZ-6) and bordered by agglomerations of large low-rise buildings (LCZ-8) and extents of low plants (LCZ-D)</p> <p>Forests of dense trees (LCZ-A) bordered by agglomerations of large low-rise buildings (LCZ-8) and agglomerations of mixed open high-rise buildings (LCZ-4), open low-rise buildings (LCZ-6), compact low-rise buildings (LCZ-3), open midrise buildings (LCZ-5) and compact midrise buildings (LCZ-2)</p> <p>Extents of low plants (LCZ-D) with forests of dense trees (LCZ-A), fragmented by agglomerations of open low-rise buildings (LCZ-6) with edges of sparsely built buildings (LCZ-9) and agglomerations of large low-rise buildings (LCZ-8)</p> <p>Thin rivers of water (LCZ-G) bordered by agglomerations of large low-rise buildings (LCZ-8) and fragmented extents of low plants (LCZ-D)</p> <p>Extents of low plants (LCZ-D) with agglomerations of large low-rise buildings (LCZ-8), agglomerations of open low-rise buildings (LCZ-6) with cores of compact low-rise buildings (LCZ-3) and scattered bare soil or sand (LCZ-F)</p> <p>Agglomerations of large low-rise buildings (LCZ-8) fragmented by extents of bare soil or sand (LCZ-F) and pockets of low plants (LCZ-D), some water (LCZ-G) bodies and scattered open midrise buildings (LCZ-5)</p> <p>Rivers and extents of water (LCZ-G) with agglomerations of mix of large low-rise buildings (LCZ-8) and compact low-rise buildings (LCZ-3) with scattered lightweight low-rise buildings (LCZ-7)</p> <p>Agglomerations of large low-rise buildings (LCZ-8) fragmented by mix of low plants (LCZ-D) and dense trees (LCZ-A) and pockets of low plants (LCZ-D)</p> <p>Mix of Vast areas of compact low-rise buildings (LCZ-3), large low-rise buildings (LCZ-8) and sparsely built buildings (LCZ-9) and pockets dense trees (LCZ-A) and low plants (LCZ-D)</p> <p>Extents of low plants (LCZ-D) fragmented by agglomerations of compact low-rise buildings (LCZ-3) and open low-rise buildings (LCZ-6) with edges of sparsely built buildings (LCZ-9) and bordered by large low-rise buildings (LCZ-8)</p> <p>Rivers of water (LCZ-G) bordered by agglomerations of mixed open high-rise buildings (LCZ-4), heavy industry areas (LCZ-10), compact midrise buildings (LCZ-2) and open low-rise buildings (LCZ-6) pocketed with large low-rise buildings (LCZ-8)</p> <p>Forests of dense trees (LCZ-A) with edges of scattered trees (LCZ-B) bordered by agglomerations of mixed compact low-rise buildings (LCZ-3) and compact midrise buildings (LCZ-2), pocketed with large low-rise buildings (LCZ-8) and with edges of open low-rise buildings (LCZ-6) and sparsely built buildings (LCZ-9)</p> <p>Agglomerations of large low-rise buildings (LCZ-8) with pockets of open high-rise buildings (LCZ-4) and compact low-rise buildings (LCZ-3) and pockets and extents of low plants (LCZ-D)</p> <p>Extents of low plants (LCZ-D) fragmented by agglomerations of large low-rise buildings (LCZ-8) and scattered sparsely built buildings (LCZ-9) and agglomerations of mixed open midrise buildings (LCZ-5) and open high-rise buildings (LCZ-4)</p> <p>Agglomerations of large low-rise buildings (LCZ-8) with Vast areas of bare soil or sand (LCZ-F) and scattered sparsely built buildings (LCZ-9), compact midrise buildings (LCZ-2) and open low-rise buildings (LCZ-6)</p> <p>Interweaving of Vast areas of compact midrise buildings (LCZ-2) and large low-rise buildings (LCZ-8) bordered by extents of low plants (LCZ-D)</p> <p>Agglomerations of large low-rise buildings (LCZ-8) with fragmented extents of low plants (LCZ-D) and scattered agglomerations of mix of open low-rise buildings (LCZ-6) and compact low-rise buildings (LCZ-3) with edges of sparsely built buildings (LCZ-9)</p> <p>Agglomerations and pockets of sparsely built buildings (LCZ-9) interlocked with fragmented pockets of low plants (LCZ-D) and scattered pockets of mix of open low-rise buildings (LCZ-6) and compact low-rise buildings (LCZ-3) and scattered stretches of large low-rise buildings (LCZ-8)</p> <p>Agglomerations of lightweight low-rise buildings (LCZ-7) and scattered compact low-rise buildings (LCZ-3), with scattered pockets of large low-rise buildings (LCZ-8) and bordered by bare soil or sand (LCZ-F) with scattered low plants (LCZ-D)</p> <p>Vast areas of large low-rise buildings (LCZ-8) bordered by extents of low plants (LCZ-D) fragmented by sparsely built buildings (LCZ-9)</p> <p>Agglomerations of open low-rise buildings (LCZ-6) with extents of low plants (LCZ-D) and pockets of large low-rise buildings (LCZ-8)</p> <p>Rivers of water (LCZ-G) with agglomerations of large low-rise buildings (LCZ-8) with compact low-rise buildings (LCZ-3) pockets and extents of low plants (LCZ-D) on the shores</p> <p>Agglomerations of open high-rise buildings (LCZ-4) with scattered open midrise buildings (LCZ-5) and open low-rise buildings (LCZ-6), bordered by agglomerations of large low-rise buildings (LCZ-8) and forests of dense trees (LCZ-A) with scattered trees (LCZ-B) edges</p> <p>Vast areas of large low-rise buildings (LCZ-8) with fragmented extents of low plants (LCZ-D) with dense trees (LCZ-A) and LCB cores and with agglomerations of open low-rise buildings (LCZ-6) with edges of sparsely built buildings (LCZ-9)</p> <p>Agglomerations of mixed compact low-rise buildings (LCZ-3) and lightweight low-rise buildings (LCZ-7) with edges of sparsely built buildings (LCZ-9) and pockets of large low-rise buildings (LCZ-8), bordered by extents of low plants (LCZ-D)</p> <p>Agglomerations of mixed open low-rise buildings (LCZ-6) and sparsely built buildings (LCZ-9) with pockets of large low-rise buildings (LCZ-8) and of bare soil or sand (LCZ-F), bordered by extents and pockets of low plants (LCZ-D)</p> <p>Thin rivers and bodies of water (LCZ-G) bordered by agglomerations of open low-rise buildings (LCZ-6) with inlays of sparsely built buildings (LCZ-9) and pockets of large low-rise buildings (LCZ-8)</p> <p>Agglomerations of sparsely built buildings (LCZ-9) with scattered compact low-rise buildings (LCZ-3) and open low-rise buildings (LCZ-6) bordered by low plants (LCZ-D) extents and pockets</p> <p>Agglomerations of large low-rise buildings (LCZ-8) with scattered pockets of low plants (LCZ-D)</p> <p>Agglomerations of open low-rise buildings (LCZ-6) with edges of sparsely built buildings (LCZ-9) bordered by stretches of low plants (LCZ-D)</p> <p>Vast areas of large low-rise buildings (LCZ-8) with agglomerations of mixed open high-rise buildings (LCZ-4) and open midrise buildings (LCZ-5) bordered by fragmented low plants (LCZ-D) extents</p> <p>Agglomerations of open low-rise buildings (LCZ-6) with pockets of large low-rise buildings (LCZ-8) and bordered by pockets low plants (LCZ-D)</p>
																																			<p>Tightly and lowly built, mostly residential areas at the edge of meadows or farmlands</p> <p>Lowly built tightly developed compounds in arid contexts</p> <p>Mixed residential areas with industrial areas or commercial and business parks at the edge of farmlands on the shores of rivers or canals</p> <p>Lowly built residential areas with industrial areas or commercial and business parks at the edge of farmlands on the shores of rivers or canals</p> <p>Mix of large residential compounds and industrial areas at the edge of forests</p> <p>Small residential settlements in direct with commercial and industrial parks at the edge of small farm lands</p> <p>Industrial and commercial areas on riversides with small residential areas</p> <p>commercial, business or industrial areas bordered by lowly built residential areas at the edge of meadows or agricultural lands</p> <p>Industrial and commercial areas on banks of canals or lakes with few small residential areas and some undeveloped banks</p> <p>Densely and very tightly developed river banks with mixed developments</p> <p>Industrial and commercial areas with some high built residential areas at the edge of woodlands</p> <p>Farmlands mixed with residential and industrial extensions at the edge of woodlands</p> <p>Farmland at edges of compact to open settlements with industrial areas</p> <p>River shores developed with high density mixed residential and industrial areas</p> <p>High density mixed residential and industrial areas at the edge of woodlands</p> <p>Residential areas around industrial areas at the edge of farmlands</p> <p>Densely mid-rise built town centers surrounded by commercial and industrial parks at the edge of agricultural lands</p> <p>Lowly built tightly developed compounds mixed with industrial and commercial areas in arid contexts</p> <p>Densely built compounds mixed with industrial and commercial areas surrounded by agricultural lands in arid contexts</p> <p>Agricultural lands with residential and industrial encroachments</p> <p>Farmlands villages with small industrial areas</p> <p>Tightly and lowly built residential areas at the edge of industrial areas or commercial and business parks in arid context</p> <p>Farmlands at the edges of industrial areas</p> <p>Low density homogeneous settlements with integrated commercial area surrounded by meadows</p> <p>Industrial or commercial areas and harbors bordered by residential area or farmlands on the shore of rivers</p> <p>Mixed midrise residential compounds with industrial areas or commercial and business parks at the edge of meadows or woods</p> <p>commercial, business or industrial areas bordered by lowly built low density residential areas at the edge of meadows or agricultural lands</p> <p>Compact low settlements with a part of makeshift habitations around industrial areas</p> <p>Mixed residential areas bordered by commercial and business parks and surrounded by agricultural lands</p> <p>Low density residential areas at the edge of water bodies with integrated commercial and business parks and leisure meadows</p> <p>Large very low density settlements at the edge of agricultural lands</p> <p>Large commercial, business or industrial parks at the edge of farmlands</p> <p>Low density residential developments at the edge of farmlands</p> <p>Large housing compounds mixed with business or commercial areas and with leisure parks</p> <p>Low density settlements with integrated commercial area surrounded by meadows and agricultural lands</p>

	Vast areas of large low-rise buildings (LCZ-8) bordering agglomerations of mixed open low-rise buildings (LCZ-6) and open midrise buildings (LCZ-5) and extents of low plants (LCZ-D)	Large commercial, business or industrial parks at the fringe of small towns, bordered by agricultural lands
	Vast areas of mixed compact low-rise buildings (LCZ-3) and open low-rise buildings (LCZ-6) inlaid with pockets and stretches of large low-rise buildings (LCZ-8)	Commercial or business parks in the middle of lowly built residential areas at the edge of leisure parks
	Vast areas of mixed open low-rise buildings (LCZ-6) and sparsely built buildings (LCZ-9) cored with compact low-rise buildings (LCZ-3) with pockets of large low-rise buildings (LCZ-8) and pockets of mixed dense trees (LCZ-A) and scattered trees (LCZ-B)	Cluster of farmlands or small, low density agricultural settlements bordered by more compact residential areas and commercial or business areas
	Mix of Vast areas of large low-rise buildings (LCZ-8) with pockets of compact low-rise buildings (LCZ-3) and open high-rise buildings (LCZ-4) bordered by fragmented extents of low plants (LCZ-D)	Mixed dense residential and industrial areas bordered by agricultural lands
	Agglomerations of mixed open low-rise buildings (LCZ-6), compact low-rise buildings (LCZ-3) and lightweight low-rise buildings (LCZ-7) with edges of sparsely built buildings (LCZ-9), bordered by extents of LCZ with patches of scattered trees (LCZ-B)	Edge of lowly built predominantly residential settlements presenting a gradient of density from tightly packed to merging with surrounding agricultural lands
	Rivers or bays of water (LCZ-G) bordered by agglomerations of large low-rise buildings (LCZ-8) with heavy industry areas (LCZ-10) on the shores and mixed agglomerations of open low-rise buildings (LCZ-6), open midrise buildings (LCZ-5), compact low-rise buildings (LCZ-3) and compact midrise buildings (LCZ-2)	Industrial harbors bordered by commercial or business parks and residential areas
	Vast areas of sparsely built buildings (LCZ-9) with scattered pockets of open low-rise buildings (LCZ-6) and scattered pockets of low plants (LCZ-D)	Very low density residential developments into farmlands or meadows
	Thin rivers of water (LCZ-G) with agglomerations of large low-rise buildings (LCZ-8) and heavy industry areas (LCZ-10) on the shores and core pockets of open low-rise buildings (LCZ-6)	Fluvial industrial harbors with some low density residential settlements
	Vast areas of large low-rise buildings (LCZ-8) with scattered agglomerations of open high-rise buildings (LCZ-4) with cores of compact midrise buildings (LCZ-2) bordered by pockets of dense trees (LCZ-A) with edges of scattered trees (LCZ-B) and low plants (LCZ-D), with some rivers and bodies of water (LCZ-G)	High residential compounds surrounded by Industrial, business or commercial areas on the shore of rivers or ponds with leisure parks
	Agglomerations of open low-rise buildings (LCZ-6) with compact low-rise buildings (LCZ-3) cores and sparsely built buildings (LCZ-9) edges, bordered by fragmented extents of low plants (LCZ-D) with scattered trees (LCZ-B) patches	Lowly built mixed settlements of commercial and residential with a decreasing gradient of density from center to the edges shared with meadows or large agricultural parcels
	Thin rivers of water (LCZ-G) bordered by agglomerations of open high-rise buildings (LCZ-4), open low-rise buildings (LCZ-6) and compact midrise buildings (LCZ-2), and scattered pockets of large low-rise buildings (LCZ-8)	River shores developed with mixed of high and midrise residential compounds and commercial, business or industrial areas
	Vast areas of large low-rise buildings (LCZ-8) with scattered pockets of low plants (LCZ-D) and open high-rise buildings (LCZ-4)	Large industrial or logistical compounds and areas with some undeveloped land and some high residential compounds
	Thin rivers and bodies of water (LCZ-G) bordered by vast areas of large low-rise buildings (LCZ-8) with pockets of heavy industry areas (LCZ-10) and open high-rise buildings (LCZ-4)	River shores developed with mixed of high and midrise residential compounds and industrial areas or facilities
	Agglomerations of compact low-rise buildings (LCZ-3) bordered by agglomerations of sparsely built buildings (LCZ-9), agglomerations of large low-rise buildings (LCZ-8) and pockets and extents of low plants (LCZ-D)	Small settlements mixed of lowly built residential areas and commercial or industrial areas surrounded by farmlands
	Agglomerations of compact low-rise buildings (LCZ-3) with inner pockets and stretches of large low-rise buildings (LCZ-8) and scattered open low-rise buildings (LCZ-6) and lightweight low-rise buildings (LCZ-7), bordered by forests of dense trees (LCZ-A)	Tightly and lowly built residential areas surrounding industrial areas or commercial or business parks at the edge of forests and meadows or leisure parks
	Vast areas of sparsely built buildings (LCZ-9) with scattered pockets of open low-rise buildings (LCZ-6), stretches of large low-rise buildings (LCZ-8) and scattered pockets of dense trees (LCZ-A) and scattered trees (LCZ-B)	Low density and lowly built residential areas surrounding commercial or business parks at the edge of woodlands and meadows or leisure parks
	Vast areas of sparsely built buildings (LCZ-9) with scattered open low-rise buildings (LCZ-6), bordered by stretches of large low-rise buildings (LCZ-8) and with scattered pockets of low plants (LCZ-D) and scattered trees (LCZ-B)	Low density and lowly built residential areas surrounding commercial or business arteries with some meadows or leisure parks
	Agglomerations of interlocked compact low-rise buildings (LCZ-3) and open low-rise buildings (LCZ-6) with pockets of large low-rise buildings (LCZ-8) bordered by extents of low plants (LCZ-D)	Lowly built mixed residential areas with some small commercial or business areas bordered by meadows, leisure parks or agricultural lands
	Vast areas of open low-rise buildings (LCZ-6) interlocked with stretches of sparsely built buildings (LCZ-9) and with pockets of large low-rise buildings (LCZ-8), pockets of dense trees (LCZ-A) and scattered compact low-rise buildings (LCZ-3)	Low density, lowly built mixed residential areas with some small commercial or business areas and some leisure parks or woodlands
	Extents of large low-rise buildings (LCZ-8) with scattered pockets of compact low-rise buildings (LCZ-3), pockets of low plants (LCZ-D), bodies of water (LCZ-G) and scattered heavy industry areas (LCZ-10)	Large industrial or logistical compounds and areas with some undeveloped land and some compact and lowly built small residential areas
	Vast areas of compact low-rise buildings (LCZ-3) with extents of low plants (LCZ-D) and of sparsely built buildings (LCZ-9) and open low-rise buildings (LCZ-6)	Compact residential areas with some scattered fringes at the edges of meadows or leisure parks
	Stretches of large low-rise buildings (LCZ-8) fragmented by agglomerations of mixed compact low-rise buildings (LCZ-3) and with pockets of large low-rise buildings (LCZ-8) and by extents of low plants (LCZ-D)	Industrial, business or commercial areas bordered by high density residential settlements with some agricultural lands or meadows
	Thin rivers and bodies of water (LCZ-G) bordered by agglomerations of open low-rise buildings (LCZ-6) and agglomerations of large low-rise buildings (LCZ-8) on the shores	Low density residential neighborhood on canals or riversides with commercial parks
	Some thin rivers and water bodies bordered by Vast areas of large low-rise buildings (LCZ-8) fragmented by scattered pockets and agglomerations of compact low-rise buildings (LCZ-3), open low-rise buildings (LCZ-6) and sparsely built buildings (LCZ-9) and scattered pockets of low plants (LCZ-D)	Predominantly industrial areas mixed with some dense lowly built residential areas with some agricultural lands
	Interlocked agglomerations of open midrise buildings (LCZ-5), open low-rise buildings (LCZ-6) and open high-rise buildings (LCZ-4) with thin stretches of large low-rise buildings (LCZ-8) and heavy industry areas (LCZ-10) and some pockets of dense trees (LCZ-A)	Residential compounds of mixed heights mixed with commercial, business or industrial areas and some leisure parks
	Vast areas of large low-rise buildings (LCZ-8) with agglomerations and pockets of compact low-rise buildings (LCZ-3), pockets of sparsely built buildings (LCZ-9) and of mixed low plants (LCZ-D), dense trees (LCZ-A) and LCB and scattered heavy industry areas (LCZ-10)	Large industrial or logistical compounds and areas with some compact and lowly built small residential areas and some leisure parks
	Agglomerations of large low-rise buildings (LCZ-8) bordered by mixed agglomerations of open high-rise buildings (LCZ-4), open midrise buildings (LCZ-5), open low-rise buildings (LCZ-6) and compact midrise buildings (LCZ-2) and agglomerations and pockets of sparsely built buildings (LCZ-9) bordered by low plants (LCZ-D)	Mixed areas of commercial, industrial or business areas and diverse residential areas
	Vast areas of large low-rise buildings (LCZ-8) with scattered heavy industry areas (LCZ-10) fragmented by agglomerations mixed of open high-rise buildings (LCZ-4), compact midrise buildings (LCZ-2), open midrise buildings (LCZ-5), compact low-rise buildings (LCZ-3) and open low-rise buildings (LCZ-6) and agglomerations and pockets of dense trees (LCZ-A)	Mixed areas of commercial, industrial or business areas and highly built with mixed densities residential areas with some woodlands
	Agglomerations with compact low-rise buildings (LCZ-3) cores, surrounded by open low-rise buildings (LCZ-6) and periphery of sparsely built buildings (LCZ-9) bordering pockets and extents of low plants (LCZ-D) and pockets of large low-rise buildings (LCZ-8)	Small industrial, commercial or business parks bordered by lowly built residential areas with decreasing density from core to outskirts mixing with meadows or farmlands
	Agglomerations of compact low-rise buildings (LCZ-3) with edges of open low-rise buildings (LCZ-6) and sparsely built buildings (LCZ-9), with pockets of low plants (LCZ-D) and of large low-rise buildings (LCZ-8)	Lowly built residential areas of mixed densities bordered by some industrial, commercial or business areas and some leisure parks or meadows
	Stretches of large low-rise buildings (LCZ-8) bordered by agglomerations of compact low-rise buildings (LCZ-3) interlocked with open low-rise buildings (LCZ-6) and with edges of sparsely built buildings (LCZ-9)	Lowly built residential compounds around scattered small industrial areas with some agricultural lands
	Vast areas of large low-rise buildings (LCZ-8) with scattered pockets of compact low-rise buildings (LCZ-3) and sparsely built buildings (LCZ-9) and scattered pockets of low plants (LCZ-D)	Large industrial or logistical compounds and areas with some undeveloped land and some lowly built small residential areas
	Vast areas of large low-rise buildings (LCZ-8) bordered by agglomerations of open low-rise buildings (LCZ-6) with edges of sparsely built buildings (LCZ-9) and with pockets of scattered trees (LCZ-B) and low plants (LCZ-D)	Large strips of commercial or business parks in the middle of low density, lowly built residential areas with small leisure parks
	Agglomerations of open low-rise buildings (LCZ-6) with interspersed sparsely built buildings (LCZ-9), fragmented by stretches of large low-rise buildings (LCZ-8) and with pockets of scattered trees (LCZ-B) and dense trees (LCZ-A)	Low density, lowly built residential areas with small leisure parks around strips of commercial or business areas
	Vast areas of compact midrise buildings (LCZ-2) with pockets of large low-rise buildings (LCZ-8) and scattered mixes of compact low-rise buildings (LCZ-3), open midrise buildings (LCZ-5), open low-rise buildings (LCZ-6) and compact high-rise buildings (LCZ-1)	Large compact midrise predominantly residential areas with some commercial or business areas



Data availability

The authors do not have permission to share data.

References

- Adams, J.S., 2005. Hoyt, H. 1939: The structure and growth of residential neighborhoods in American cities. Washington, DC: Federal Housing Administration. Progress in Human Geography 29 (3), 321–325. <https://doi.org/10.1191/0309132505ph552xx>.
- Afchar, D., Hennequin, R., Guigue, V., 2022. Learning Unsupervised Hierarchies of Audio Concepts (arXiv:2207.11231). arXiv. <https://doi.org/10.48550/arXiv.2207.11231>.
- Alexander, C., Ishikawa, S., Silverstein, M., 1977. A pattern language: Towns, buildings, construction. Oxford University Press.
- Andrews, H.F., 1993. Durkheim and Social Morphology. Routledge, In Emile Durkheim.
- Angel, S., Blei, A., Parent, J., Lamson-Hall, P., Galarza Sánchez, N., 2016. Atlas of Urban Expansion—2016 Edition: Vol. Volume 1: Areas and Densities. Copublished by the NYU Urban Expansion Program at New York University, UN-Habitat, and the Lincoln Institute of Land Policy. <https://www.lincolnst.edu/publications/other/atlas-urban-expansion-2016-edition/>.
- Arribas-Bel, D., Fleischmann, M., 2022. Spatial signatures—Understanding (urban) spaces through form and function. Habitat International 128, 102641. <https://doi.org/10.1016/j.habitatint.2022.102641>.
- Arsanjani, J. J., Zipf, A., Mooney, P., & Helbich, M. (Eds.), 2015. OpenStreetMap in GIScience: Experiences, Research, and Applications. Springer International Publishing. <https://doi.org/10.1007/978-3-319-14280-7>.
- Batty, M., 2009. In: Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies. Springer, New York, pp. 1041–1071. https://doi.org/10.1007/978-0-387-30440-3_69.
- Batty, M., Longley, P., 1994. Fractal cities: A geometry of form and function. Acad Press.
- Bechtel, B., Foley, M., Mills, G., Ching, J., See, L., Alexander, P., O'Connor, M., Albuquerque, T., Andrade, M., Brovelli, M., Das, D., Fonte, C., Petit, G., Hanif, U., Jiménez, J., Lackner, S., Liu, W., Perera, N., Rosni, N. A., Gál, T., 2015. CENSUS of Cities: LCZ Classification of Cities (Level 0) – Workflow and Initial Results from Various Cities.
- Bertyák, Á., 2021. Urban morphology: The classical and modern research methodologies. Periodica Polytechnica Architecture 52 (2). <https://doi.org/10.3311/PPar.17988>. Article 2.
- Biljecki, F., Chow, Y.S., 2022. Global building morphology indicators. Computers, Environment and Urban Systems 95, 101809. <https://doi.org/10.1016/j.compenvurbsys.2022.101809>.
- Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems 65, 126–139. <https://doi.org/10.1016/j.compenvurbsys.2017.05.004>.
- Boeing, G., 2019. Urban spatial order: Street network orientation, configuration, and entropy. Applied Network Science 4 (1), 1–19. <https://doi.org/10.1007/s41109-019-0189-1>.
- Boeing, G., 2020. A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood. Environment and Planning B 47 (4), 590–608. <https://doi.org/10.1177/2399808318784595>.
- Bosch, M., 2019. PyLandStats: An open-source Pythonic library to compute landscape metrics. PLOS ONE 14 (12), e0225734. <https://doi.org/10.1371/journal.pone.0225734>.
- Braunfels, W., 1976. Abendländische Stadtbaukunst: Herrschaftsform u. Baugestalt, DuMont Schauberg.
- Carroll, R.M., Nordholm, L.A., 1975. Sampling characteristics of Kelley's ϵ and Hays' ω . Educational and Psychological Measurement 35 (3), 541–554. <https://doi.org/10.1177/001316447503500304>.
- Castex, J., Depaule, J.-C., Panerai, P., 1997. Formes urbaines: De l'îlot à la barre. Éd. Parenthèses.
- Cataldi, G., Maffei, G.L., Vaccaro, P., 2002. Saverio Muratori and the Italian school of planning typology. Urban Morphology 6 (1), 3–14. <https://doi.org/10.51347/jum.v6i1.3899>.
- Chakraborty, S., Novotný, J., Maity, I., Lemoine-Rodríguez, R., Follmann, A., 2024. Same planet but different worlds! Diverging convergence pattern of urban form typologies across 413 cities with million+ inhabitants and their sustainability trade-offs. Habitat International 145, 103024. <https://doi.org/10.1016/j.habitatint.2024.103024>.
- Chang, J., Qu, Z., Xu, R., Pan, K., Xu, B., Min, Y., Ren, Y., Yang, G., Ge, Y., 2017. Assessing the ecosystem services provided by urban green spaces along urban center-edge gradients. Scientific Reports 7 (1), 11226. <https://doi.org/10.1038/s41598-017-11559-5>.
- Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A Simple Framework for Contrastive Learning of Visual Representations (arXiv:2002.05709). arXiv. <https://doi.org/10.48550/arXiv.2002.05709>.
- Chen, W., Huang, H., Liao, S., Gao, F., Biljecki, F., 2024. Global urban road network patterns: Unveiling multiscale planning paradigms of 144 cities with a novel deep learning approach. Landscape and Urban Planning 241, 104901. <https://doi.org/10.1016/j.landurbplan.2023.104901>.
- Coates, A., Ng, A., Lee, H., 2011. An analysis of single-layer networks in unsupervised feature learning. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 215–223.
- Cohen, J., 1973. Eta-squared and partial eta-squared in fixed factor anova designs. Educational and Psychological Measurement 33 (1), 107–112. <https://doi.org/10.1177/001316447303300111>.
- Cohen, J., 1988. Statistical power analysis for the behavioral sciences. Erlbaum.
- Corbane, C., Pesaresi, M., Kemper, T., Politis, P., Florczyk, A.J., Syrris, V., Melchiorri, M., Sabo, F., Soille, P., 2019. Automated global delineation of human settlements from 40 years of Landsat satellite data archives. Big Earth Data 3 (2), 140–169. <https://doi.org/10.1080/20964471.2019.1625528>.
- Cozzolino, S., 2020. The (anti) adaptive neighbourhoods. Embracing complexity and distribution of design control in the ordinary built environment. Environment and Planning b: Urban Analytics and City Science 47 (2), 203–219. <https://doi.org/10.1177/2399808319857451>.
- Cramér, H., 1991. Mathematical Methods of Statistics. Princeton University Press.
- de Amorim, R.C., 2015. Feature relevance in ward's hierarchical clustering using the LpNorm. Journal of Classification 32 (1), 46–62. <https://doi.org/10.1007/s00357-015-9167-1>.
- Debray, H., Kraff, N.J., Zhu, X., Taubenböck, H., 2023. Planned, unplanned, or in-between? A concept of the intensity of plannedness and its empirical relation to the built urban landscape across the globe. Landscape and Urban Planning 233, 104711. <https://doi.org/10.1016/j.landurbplan.2023.104711>.
- Debray, H., Qiu, C., Schmitt, M., Wang, Y., Zhu, X., Taubenböck, H., 2021. Types of Morphological Configurations of the City across the Globe – a Remote Sensing based Comparative Approach [Application/pdf]. CITIES 20.50 – Creating Habitats for the 3rd Millennium: Smart – Sustainable – Climate Neutral. Proceedings of REAL CORP 2021, 26th International Conference on Urban Development and Regional Planning in the Information Society.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (methodological) 39 (1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Demuzere, M., Kittner, J., Bechtel, B., 2021. LCZ generator: A web application to create local climate zone maps. Frontiers in Environmental Science 9. <https://www.frontiersin.org/articles/10.3389/fenvs.2021.637455>.
- Demuzere, M., Kittner, J., Martilli, A., Mills, G., Moede, C., Stewart, I.D., van Vliet, J., Bechtel, B., 2022. A global map of local climate zones to support earth system modelling and urban-scale environmental science. Earth System Science Data 14 (8), 3835–3873. <https://doi.org/10.5194/essd-14-3835-2022>.
- Dibble, J., Prelorandjos, A., Romice, O., Zanella, M., Strano, E., Pagel, M., Porta, S., 2017. On the origin of spaces: Morphometric foundations of urban form evolution. Environment and Planning b: Urban Analytics and City Science 46 (4), 707–730. <https://doi.org/10.1177/2399808317725075>.
- Dovey, K., 2020. Towards a Morphogenesis of Informal Settlements. Habitat International 14.
- Droin, A., Wurm, M., Weigand, M., Gawlas, C., Köberl, M., Taubenböck, H., 2024. How does pedestrian permeability vary in and across cities? A fine-grained assessment for all large cities in Germany. Computers, Environment and Urban Systems 110, 102115. <https://doi.org/10.1016/j.compenvurbsys.2024.102115>.
- Durand, J.-N.-L., 1805. Précis des leçons d'architecture données à l'École polytechnique. À Paris, chez L'auteur. <http://archive.org/details/prcisdesleons02dura>.
- Durand, J.-N.-L., 1825. Précis des leçons d'architecture données à l'École polytechnique, par J.-N.-L. Durand,... <https://gallica.bnf.fr/ark:/12148/bpt6k5762681g>.
- Eiben, E., Galian, R., Kanj, I., Ordyniak, S., Szeider, S., 2021. The Parameterized Complexity of Clustering Incomplete Data (arXiv:1911.01465). arXiv. <http://arxiv.org/abs/1911.01465>.
- Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., Strano, E., 2017. Breaking new ground in mapping human settlements from space – The Global Urban Footprint. ISPRS Journal of Photogrammetry and Remote Sensing 134, 30–42. <https://doi.org/10.1016/j.isprsjprs.2017.10.012>.
- Fay, M.P., Proschian, M.A., 2010. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Statistics Surveys 4, 1–39. <https://doi.org/10.1214/09-SS051>.
- Fleischmann, M., 2019. momepy: Urban morphology measuring toolkit. Journal of Open Source Software 4 (43), 1807. <https://doi.org/10.21105/joss.01807>.
- Fleischmann, M., Arribas-Bel, D., 2022. Geographical characterisation of British urban form and function using the spatial signatures framework. Scientific Data 9 (1). <https://doi.org/10.1038/s41597-022-01640-8>. Article 1.
- Fleischmann, M., Feliciotti, A., Romice, O., Porta, S., 2021a. Methodological foundation of a numerical taxonomy of urban form. Environment and Planning b: Urban Analytics and City Science 239980832110598. <https://doi.org/10.1177/23998083211059835>.

- Fleischmann, M., Romice, O., Porta, S., 2021b. Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environment and Planning B: Urban Analytics and City Science* 48 (8), 2133–2150. <https://doi.org/10.1177/2399808320910444>.
- Gadetsky, A., Jiang, Y., Brbic, M., 2024. June 6). Let Go of Your Labels with Unsupervised Transfer. Forty-First International Conference on Machine Learning.
- Gil, J., Beirão, J.N., Montenegro, N., Duarte, J.P., 2012. On the discovery of urban typologies: Data mining the many dimensions of urban form. *Urban Morphology* 16 (1), 27–40.
- Han, S., Park, S., Park, S., Kim, S., Cha, M., 2020. Mitigating Embedding and Class Assignment Mismatch in Unsupervised Image Classification. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (Vol. 12369, pp. 768–784). Springer International Publishing. https://doi.org/10.1007/978-3-030-58586-0_45.
- Hays, W.L., 1973. *Statistics for the Social Sciences*, 2. ed. Holt, Rinehart and Winston.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition (arXiv:1512.03385). arXiv. <https://doi.org/10.48550/arXiv.1512.03385>.
- Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., Zipf, A., 2021. The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports* 11 (1). <https://doi.org/10.1038/s41598-021-82404-z>. Article 1.
- Hu, J., Wang, Y., Taubenböck, H., Zhu, X.X., 2021. Land consumption in cities: A comparative study across the globe. *Cities* 113, 103163. <https://doi.org/10.1016/j.cities.2021.103163>.
- Huang, J., Lu, X.X., Sellers, J.M., 2007. A global comparative analysis of urban form: Applying spatial metrics and remote sensing. *Landscape and Urban Planning* 82 (4), 184–197. <https://doi.org/10.1016/j.landurbplan.2007.02.010>.
- Inostroza, L., Baur, R., Csaplovics, E., 2013. Urban sprawl and fragmentation in Latin America: A dynamic quantification and characterization of spatial patterns. *Journal of Environmental Management* 115, 87–97. <https://doi.org/10.1016/j.jenvman.2012.11.007>.
- Ji, X., Vedaldi, A., Henriques, J., 2019. Invariant information clustering for unsupervised image classification and segmentation. *IEEE/CVF International Conference on Computer Vision (ICCV)* 2019, 9864–9873. <https://doi.org/10.1109/ICCV.2019.00996>.
- Jin, X., Han, J., 2010. K-means clustering. In: Sammut, C., Webb, G.I. (Eds.), *Encyclopedia of Machine Learning*. Springer, US, pp. 563–564. https://doi.org/10.1007/978-0-387-30164-8_425.
- Khalilzadeh, J., Tasci, A.D.A., 2017. Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management* 62, 89–96. <https://doi.org/10.1016/j.tourman.2017.03.026>.
- Kostof, S., 1991. *The city shaped: Urban patterns and meanings through history* (1. paperback ed). Thames & Hudson.
- Kostof, S., 1992. *The city assembled: The elements of urban form through history*. Thames and Hudson.
- Krier, L., 1998. *Architecture: Choice or Fate*. Andreas Papadakis.
- Krier, R., Rowe, C., 1991. *Urban Space* (5. impr). Academy Editions.
- Krizhevsky, A., 2009. Learning Multiple Layers of Features from Tiny Images.
- Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4. <https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00863>.
- Lefebvre, H., 1974. *La production de l'espace* (4. éd). Éd. Anthropos.
- Lemoine-Rodríguez, R., Inostroza, L., Zepp, H., 2020. The global homogenization of urban form. An assessment of 194 cities across time. *Landscape and Urban Planning* 204, 103949. <https://doi.org/10.1016/j.landurbplan.2020.103949>.
- Lemoine-Rodríguez, R., Inostroza, L., Zepp, H., 2022. Does urban climate follow urban form? Analysing intraurban LST trajectories versus urban form trends in 3 cities with different background climates. *Science of the Total Environment* 830, 154570. <https://doi.org/10.1016/j.scitotenv.2022.154570>.
- Lim, T.K., Ignatius, M., Miguel, M., Wong, N.H., Juang, H.-M.-H., 2017. Multi-scale urban system modeling for sustainable planning and design. *Energy and Buildings* 157, 78–91. <https://doi.org/10.1016/j.enbuild.2017.02.024>.
- Liu, J., Ma, S., Xu, W., Zhu, L., 2022. A generalized Wilcoxon–Mann–Whitney type test for multivariate data through pairwise distance. *Journal of Multivariate Analysis* 190, 104946. <https://doi.org/10.1016/j.jmva.2022.104946>.
- Liu, P., Biljecki, F., 2022. A review of spatially-explicit GeoAI applications in Urban Geography. *International Journal of Applied Earth Observation and Geoinformation* 112, 102936. <https://doi.org/10.1016/j.jag.2022.102936>.
- Lynch, K., 1992. *The image of the city*, (21st ed). Mit press.
- McGarigal, K., Cushman, S., Ene, E., 2023. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical Maps (Version v4) [Computer software]. <https://fragstats.org/index.php>.
- Microsoft. (2020, February 12). Building Footprints—Bing Maps. <https://www.microsoft.com/en-us/maps/building-footprints>.
- Mu, Q., Miao, S., Wang, Y., Li, Y., He, X., Yan, C., 2020. Evaluation of employing local climate zone classification for mesoscale modelling over Beijing metropolitan area. *Meteorology and Atmospheric Physics* 132 (3), 315–326. <https://doi.org/10.1007/s00703-019-00692-7>.
- Muratori, S., 1960. *Studi per una operante storia urbana di Venezia*. Istituto Poligrafico Dello Stato.
- Nakagawa, S., Cuthill, I.C., 2007. Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews* 82 (4), 591–605. <https://doi.org/10.1111/j.1469-185X.2007.00027.x>.
- Openshaw, S., 1984. *The modifiable areal unit problem*. Geo.
- Panerai, P., Demorgon, M., Depaule, J.-C., 1999. *Analyse urbaine*. Parenthèses.
- Park, R.E., Burgess, E.W., McKenzie, R.D., 1925. *The City*. University of Chicago Press.
- Park, S., Han, S., Kim, S., Kim, D., Park, S., Hong, S., Cha, M., 2021. Improving Unsupervised Image Clustering With Robust Learning (arXiv:2012.11150; Version 2). arXiv. <http://arxiv.org/abs/2012.11150>.
- Pont, M. B., Olsson, J., 2018, May 15. Typology based on three density variables central to Spacematrix using cluster analysis. 24th ISUF 2017 - City and Territory in the Globalization Age. 24th ISUF 2017 - City and Territory in the Globalization Age. <http://ocs.editorial.upv.es/index.php/ISUF/ISUF2017/paper/view/5319>.
- Qian, Q., Xu, Y., Hu, J., Li, H., Jin, R., 2022. Unsupervised visual representation learning by online constrained K-means. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2022, 16619–16628. <https://doi.org/10.1109/CVPR52688.2022.01614>.
- Ramaramanana, F.N., Teller, J., Sliuzas, R., Kuffer, M., 2025. Using comparative approaches to model deprivation in Antananarivo, Madagascar: A multidimensional analysis using principal components analysis and weighting system across meso and macro scales. *Habitat International* 159, 103359. <https://doi.org/10.1016/j.habitatint.2025.103359>.
- Rode, P., Floater, G., Thomopoulos, N., Docherty, J., Schwinger, P., Mahendra, A., Fang, W., 2017. Accessibility in Cities: Transport and Urban Form. In: Meyer, G., Shaheen, S. (Eds.), *Disrupting Mobility: Impacts of Sharing Economy and Innovative Transportation on Cities*. Springer International Publishing, pp. 239–273. https://doi.org/10.1007/978-3-319-51602-8_15.
- Rubin, D.B., 1976. Inference and missing data. *Biometrika* 63 (3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>.
- Sawilowsky, S.S., 2009. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods* 8, 597–599. <https://doi.org/10.56801/10.56801/v8.i.452>.
- Schirmer, P.M., Axhausen, K.W., 2016. A multiscale classification of urban morphology. *Journal of Transport and Land Use* 9 (1). <https://doi.org/10.5198/jtlu.2015.667>. Article 1.
- Sharifi, A., 2019. Urban form resilience: A meso-scale analysis. *Cities* 93, 238–252. <https://doi.org/10.1016/j.cities.2019.05.010>.
- Staab, J., Stark, T., Wurm, M., Wolf, K., Dallavalle, M., Schady, A., Lakes, T., Taubenböck, H., 2023. Using CNNs on Sentinel-2 data for road traffic noise modelling. *Joint Urban Remote Sensing Event (JURSE)* 2023, 1–4. <https://doi.org/10.1109/JURSE57346.2023.10144160>.
- Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society* 93 (12), 1879–1900. <https://doi.org/10.1175/BAMS-D-11-00019.1>.
- Tammaru, T., Sinitsyna, A., Akhaviadegan, A., van Ham, M., Marciniak, S., Musterd, S., 2021. Income inequality and residential segregation in European cities. In: Pryce, G., Wang, Y.P., Chen, Y.P., Shan, J., Wei, H. (Eds.), *Urban Inequality and Segregation in Europe and China: towards a New Dialogue*. Springer International Publishing, pp. 39–54. https://doi.org/10.1007/978-3-030-74544-8_3.
- Taubenböck, H., 2021. How we live and what that means—A character study with data from space. *IEEE International Geoscience and Remote Sensing Symposium IGARSS* 2021, 1190–1193. <https://doi.org/10.1109/IGARSS47720.2021.9553084>.
- Taubenböck, H., Debray, H., Qiu, C., Schmitt, M., Wang, Y., Zhu, X., 2020. Seven city types representing morphologic configurations of cities across the globe. *Cities* 105, 102814. <https://doi.org/10.1016/j.cities.2020.102814>.
- Taubenböck, H., Kraff, N.J., Wurm, M., 2018. The morphology of the Arrival City—A global categorization based on literature surveys and remotely sensed data. *Applied Geography* 92, 150–167. <https://doi.org/10.1016/j.apgeog.2018.02.002>.
- Taubenböck, H., Mast, J., Lemoine Rodríguez, R., Debray, H., Wurm, M., Geiß, C., 2025. Was global urbanization from 1985 to 2015 efficient in terms of land consumption? *Habitat International* 160, 103397. <https://doi.org/10.1016/j.habitatint.2025.103397>.
- Taubenböck, H., Wegmann, M., Roth, A., Mehl, H., Dech, S., 2009. Urbanization in India – Spatiotemporal analysis using remote sensing data. *Computers, Environment and Urban Systems* 33 (3), 179–188. <https://doi.org/10.1016/j.compenvurbysys.2008.09.003>.
- Taubenböck, H., Weigand, M., Esch, T., Staab, J., Wurm, M., Mast, J., Dech, S., 2019. A new ranking of the world's largest cities—Do administrative units obscure morphological realities? *Remote Sensing of Environment* 232, 111353. <https://doi.org/10.1016/j.rse.2019.111353>.
- Ulvi, H., Yerlikaya, M.A., Yildiz, K., 2024. Urban traffic mobility optimization model: A novel mathematical approach for predictive urban traffic analysis. *Applied Sciences* 14 (13). <https://doi.org/10.3390/app14135873>. Article 13.
- UN. (2018, May 16). 2018 Revision of World Urbanization Prospects | Multimedia Library - United Nations Department of Economic and Social Affairs. <https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html>.
- United Nations., 2015. Sustainable Development Goals (SDG 11) | UN Western Europe. United Nations Western Europe. <https://unric.org/en/sdg-11/>.
- Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., Van Gool, L., 2020. SCAN: Learning to Classify Images without Labels (arXiv:2005.12320; Version 2). arXiv. <http://arxiv.org/abs/2005.12320>.

- von Szombathely, M., Albrecht, M., Antanaskovic, D., Augustin, J., Augustin, M., Bechtel, B., Bürk, T., Fischereit, J., Grawe, D., Hoffmann, P., Kaveckis, G., Krefis, A., Oßenbrügge, J., Scheffran, J., Schlünzen, K., 2017. A conceptual modeling approach to health-related urban well-being. *Urban Science* 1 (2), 17. <https://doi.org/10.3390/urbansci1020017>.
- Wandl, A., Hausleitner, B., 2021. Investigating functional mix in Europe's dispersed urban areas. *Environment and Planning b: Urban Analytics and City Science* 2399808320987849. <https://doi.org/10.1177/2399808320987849>.
- Wang, J., Biljecki, F., 2022. Unsupervised machine learning in urban studies: A systematic review of applications. *Cities* 129, 103925. <https://doi.org/10.1016/j.cities.2022.103925>.
- Wang, J., Fleischmann, M., Venerandi, A., Romice, O., Kuffer, M., Porta, S., 2023. EO + Morphometrics: Understanding cities through urban morphology at large scale. *Landscape and Urban Planning* 233, 104691. <https://doi.org/10.1016/j.landurbplan.2023.104691>.
- Wang, J., Huang, W., Biljecki, F., 2024. Learning visual features from figure-ground maps for urban morphology discovery. *Computers, Environment and Urban Systems* 109, 102076. <https://doi.org/10.1016/j.compenvurbsys.2024.102076>.
- Weigand, M., Worbis, S., Sapena, M., Taubenböck, H., 2023. A structural catalogue of the settlement morphology in refugee and IDP camps. *International Journal of Geographical Information Science*. <https://www.tandfonline.com/doi/abs/10.1080/13658816.2023.2189724>.
- Wentz, E.A., York, A.M., Alberti, M., Conrow, L., Fischer, H., Inostroza, L., Jantz, C., Pickett, S.T.A., Seto, K.C., Taubenböck, H., 2018. Six fundamental aspects for conceptualizing multidimensional urban form: A spatial mapping perspective. *Landscape and Urban Planning* 179, 55–62. <https://doi.org/10.1016/j.landurbplan.2018.07.007>.
- Whitehand, J. (2001). British urban morphology: The Conzenian tradition. Undefined. <https://www.semanticscholar.org/paper/British-urban-morphology%3A-the-Conzenian-tradition-Whitehand/b9fd4ff8d85095a871183c75263de48be98e7049>.
- Wilcoxon, F., 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1 (6), 80–83. <https://doi.org/10.2307/3001968>.
- Wu, C., Wang, J., Wang, M., Biljecki, F., Kraak, M.-J., 2025. Formalising the urban pattern language: A morphological paradigm towards understanding the multi-scalar spatial structure of cities. *Cities* 161, 105854. <https://doi.org/10.1016/j.cities.2025.105854>.
- Zeng, L., Zhang, X., Lu, J., Li, Y., Hang, J., Hua, J., Zhao, B., Ling, H., 2024. Influence of various urban morphological parameters on urban canopy ventilation: A parametric numerical study. *Atmosphere* 15 (3). <https://doi.org/10.3390/atmos15030352>. Article 3.
- Zhou, Q., Bao, J., Liu, H., 2022. Mapping urban forms worldwide: An analysis of 8910 street networks and 25 indicators. *ISPRS International Journal of Geo-Information* 11 (7). <https://doi.org/10.3390/ijgi11070370>. Article 7.
- Zhu, X., Qiu, C., Hu, J., Shi, Y., Wang, Y., Schmitt, M., 2021. So2Sat GUL - So2Sat Global Urban LCZs [Dataset]. Technical University of Munich. <https://doi.org/10.14459/2021MP1633461>.
- Zhu, X., Qiu, C., Hu, J., Shi, Y., Wang, Y., Schmitt, M., Taubenböck, H., 2022. The urban morphology on our planet – Global perspectives from space. *Remote Sensing of Environment* 269, 112794. <https://doi.org/10.1016/j.rse.2021.112794>.