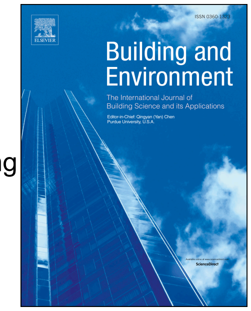
Journal Pre-proof



Identifying urban morphological archetypes for microclimate studies using a clustering approach

Mitali Yeshwant Joshi, Auline Rodler, Marjorie Musy, Sihem Guernouti, Mario Cools, Jacques Teller

PII: S0360-1323(22)00804-6

DOI: https://doi.org/10.1016/j.buildenv.2022.109574

Reference: BAE 109574

To appear in: *Building and Environment*

Received Date: 1 June 2022

Revised Date: 29 July 2022

Accepted Date: 31 August 2022

Please cite this article as: Joshi MY, Rodler A, Musy M, Guernouti S, Cools M, Teller J, Identifying urban morphological archetypes for microclimate studies using a clustering approach, *Building and Environment* (2022), doi: https://doi.org/10.1016/j.buildenv.2022.109574.

Identifying urban morphological archetypes for microclimate studies using a clustering approach

Authors: Mitali Yeshwant Joshi[1*], Auline Rodler[2], Marjorie Musy[2], Sihem Guernouti[2], Mario Cools[1], Jacques Teller[1]

[1]LEMA research group, Urban & Environmental Engineering Department, University of Liège, Liege, Belgium
(mjoshi@uliege.be, mario.cools@uliege.be, jacques.teller@uliege.be)

[2]Cerema, Equipe de Recherche BPE;CNRS UMR 5271 - Université Savoie Mont Blanc, Chambéry, France ; Institut de Recherche en Sciences et Techniques de la Ville (IRSTV), Nantes, France
(auline.rodler@cerema.fr, marjorie.musy@cerema.fr, sihem.guernouti@cerema.fr)

[*]Corresponding author (mjoshi@uliege.be)

Abstract:

Urban morphology relates to the form, structure, physical characteristics, and arrangement of buildings affecting the urban microclimate. As the morphological characteristics vary across the city, small units such as urban blocks are analysed for microclimate estimation. However, microclimatic analysis of all the blocks in a city is computationally challenging and time-consuming. Therefore, it is vital to identify representative blocks in a city to obtain a general overview of the microclimate. Urban morphological archetypes are the representative units of a homogenous group of blocks based on morphological parameters. Here, we propose a systematic approach for identifying urban morphological archetypes suited for microclimatic analysis. Specifically, we employ a well-defined, PCA-based *k-means* clustering approach supported by validation using external criterion analysis. We use urban morphological parameters based on form, shape, arrangement, and variations within a block in Liege, Belgium. We use the cubic clustering criterion and pseudo F statistic to identify nine distinct homogenous clusters. Then, we propose a validation approach in the absence of existing typologies using ANOVA analysis on the external criterion of land surface temperature, a proxy for measuring microclimate. The validation suggests that the clusters are significantly different, indicating successful clustering. We also compare our classification to the existing local climate zone (LCZ) classification. We identify relevant sub-classes within the broader LCZ classes essential for capturing microclimatic variation. Finally, the study provides realistic archetypes for performing microclimatic simulations at a city scale. The proposed approach can be effectively applied to other cities for urban microclimate studies.

Keywords: Microclimate, urban morphological archetypes, clustering, *k-means*, local climate zones

1 Introduction

Urban settlements alter the natural environment surrounding them, creating a unique microclimate [1]. Urban microclimates tend to produce and retain more heat, resulting in comparatively higher temperatures than their rural counterparts. This phenomenon is known as the urban heat island (UHI) effect [2,3]. The overall air and surface temperature in cities have risen gradually over the years, resulting in a significant air UHI and surface UHI effect, respectively [4,5]. The spatial variation of the UHI effect largely depends on urban morphological factors such as built-up intensity, presence of vegetation, building heights, albedo and sky-view factor (SVF) [6,7]. Therefore, microclimates encompassing buildings and urban blocks are studied to understand and mitigate the UHI effect.

The urban microclimate is the outcome of dynamic interactions between the macroclimate and urban morphology [8,9]. Thus, urban morphology is a vital part of UHI-related microclimatic studies owing to its importance also for assessing mitigating solutions for the UHI effect [10,11]. Urban morphology relates to the form and the structure of an urban area and the buildings' physical characteristics and arrangement. Urban morphological parameters influence the urban microclimate in several ways. For example, building density reduces the average wind velocity worsening the urban ventilation and intensifying the UHI effect [12,13]. Longwave radiation gets blocked in the streets due to low SVF on the streets, retaining more heat in the region and escalating the UHI effect [14,15].

Urban morphological parameters, such as the size of the building façades, also impact wind velocity, thus intercepting solar radiation and contributing to solar trapping, which primarily causes the UHI effect [16]. Sometimes, urban morphological factors can also help in regulating the UHI effect. For example, arrangements of buildings in a block, such as U-shaped blocks, blocks with courtyards, or multiple courtyards, have improved microclimates compared to detached, attached and linear blocks [17]. Additionally, a block's variation in the height of buildings is also observed to reduce outdoor temperatures better than the blocks with uniform building heights [12]. In winters, mid-rise blocks are more suitable than high-rise blocks with open spaces to block the cold winds (Xu et al., 2019). These studies demonstrate that urban morphological patterns largely influence microclimate.

The morphological patterns within the city vary based on building properties, street-related properties, and properties of urban blocks. Due to such variations in the city, many researchers suggest analysing individual urban units, such as urban blocks, to study the UHI effect at the microscale [18,19]. Evaluating microclimate in urban blocks also allows the urban planners and designers to play a pivotal role in providing balanced strategies and techniques related to its local context [20]. However, analysing the microclimate of all the blocks in the city can be computationally challenging and time-consuming. In this scenario, identifying typical urban morphological archetypes representing a homogenous group of blocks [21,22] in a city can simplify microclimatic studies and aid in generalising the results at a city scale. Therefore, we propose a systematic approach to identifying urban morphological archetypes suited for microclimatic studies.

Typically, the first step in identifying urban morphological archetypes involves classifying the entire city into different types. There have been previous attempts to classify the urban areas into homogenous units for UHI-related studies. Stewart and Oke [23] introduced the local climate zones (LCZs) as homogenous regions in terms of surface cover, structure, material and human activity that stretch over hundreds of meters to several kilometres horizontally. Based on LCZ, several researchers have generated LCZs specific to their areas, substituting or adding a few parameters in the process [24–26]. Although generating LCZs is the widely used approach, Stewart and Oke [23] highlight that the LCZ system is generic and cannot capture the peculiarities of every urban area and is adapted to catch the microclimate effect at the scale of a few hundred meters. Thus, they suggest that users can create new sub-classes in the city if needed. Apart from this, the LCZ parameters often remain insufficient while describing the urban canopy in detail, especially when the end goal is to analyse the UHI effect at the microscale [7]. Therefore, additional morphological parameters that describe the urban canopy in detail are necessary for classifying the urban area into morphological archetypes. Apart from this, as the basic unit of design for urban planners and designers is an urban block, rasterised output like LCZs does not provide precise typologies at the block level [27,28].

Other approaches have been employed previously, such as rule-based classification, machine learning based classification algorithms, and classification using spatial multi-criteria analysis to identify urban morphological archetypes [29–31]. While these methods are systematic and replicable, they are only applicable in the case of predefined urban morphological archetypes. In the absence of predefined archetypes, the clustering approach can be helpful as it is data-driven [19,32–35]. Furthermore, clustering allows simultaneous assessment of various variables by grouping the elements based on similarities [35]. Thus, the clustering approach is more logical and effective in identifying urban morphological typologies. However, prior studies have not employed the clustering approach with external validation to form urban archetypes to identify local climate zones for microclimate analysis. Therefore, in this study, we propose a clustering approach to identify the urban morphological archetypes based on important urban morphological parameters.

Once the clusters are generated, we need to check whether the clustering has delivered unique and distinct archetypes. One way to check if the clusters are significantly different is with the help of predefined or pre-existing rules for archetypes. But, in many cases, there is a lack of predefined rules wherein the validation requires some logical basis. An approach popular in the statistical community but not widely used in clustering urban blocks is the external validation criterion using ANOVA analysis. However, such a validation requires an external parameter not used in clustering [36], which, in this case, is a non-morphological parameter (i.e. a parameter not related to building geometry) to prove that significant differences exist between the clusters. Therefore, in this study, we propose a validation using an external parameter to confirm the adequacy of the clusters' diversity.

Summing up, in this paper, we propose a systematic approach to identifying the urban morphological archetypes particularly suited for UHI-related microclimatic analysis. The approach involves:

1. Clustering: classifying city blocks based on urban morphological parameters.
2. Validation of the clustering-based classification using an external criterion.
3. Determining the unique urban morphological archetypes that represent the different clusters in the city.

We also compare the clustering-based classification results with the LCZ data created for Europe by the world urban database and access portal tools (WUDAPT) [37,38].

The present approach provides realistic urban blocks for microclimate analysis, including CFD simulations instead of a simplistic representation of urban blocks. Furthermore, it reduces the computational time for analysing the microclimate at a higher resolution as it alleviates the need for simulating the microclimate of an entire city. Instead, the simulations can be carried out on the identified archetypes to arrive at a general overview of the microclimatic situation in the city. Moreover, the properties of these archetypes can be helpful to further generate modelled blocks in the city that will be a better representation of reality.

2 Methodology

2.1 Study area and dataset

In this study, we examine the city of Liege in the Wallonia region of Belgium. It is the third most populous city in the country, with 196,296 inhabitants [39] and an area of 69 km^2. The city is densely occupied by buildings in the centre, leaving few open green spaces. The surface temperatures in the city are observed to be high during the summer, indicating a significant surface UHI effect [40].

The dataset includes the building footprints from the PICC (Projet Informatique de Cartographie Continue) dataset, which has an accuracy of less than 25 cm. This data is retrieved from the geoportal of Wallonia (https://geoportail.wallonie.be). We also use the parcels from cadastre data (2018) for block generation. For the height of the buildings, we use the LiDAR point cloud data from 2014 with a point density of 0.8 points/m^2, retrieved from the geoportal of Wallonia.

2.2 Selection of blocks

Urban blocks are an area with one building or a group of buildings surrounded by streets [29,34]. Thus, firstly, the city is divided into several blocks using the parcels in the cadastre data. The border of the enclosing streets delineates a block, and cadastral information on streets or parcels can help in demarcating the blocks. Blocks can also be derived if the street width information is available in a city [34].

For Liege city, the cadastral information is readily available; therefore, we use parcels to define the blocks for this study. The parcels are the plot boundaries for each building in the city, as depicted in Figure 1(a). We merge all the parcels and transform them into blocks using ArcGIS Pro (Version 2.9.1) (Figure 1(b)).

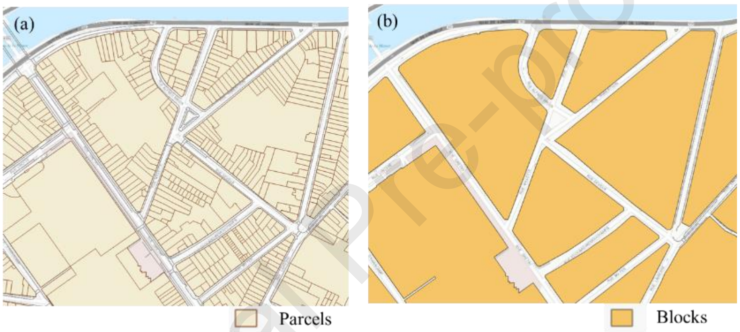


Figure 1 (a) Parcels based on cadastral data (b) Defined blocks with the help of parcels

Blocks with fewer buildings and larger open spaces are not relevant for this study. Thus, the first selection criteria for the blocks in this analysis is the ground space index (*GSI*). *GSI* refers to the area occupied by buildings (A_{bu}) in the block per area of the block (A_{bl}) as illustrated in the following equation [41]:

$$GSI = \frac{\sum_{i=1}^{n} A_{bu(i)}}{A_{bl}} \quad (1)$$

We only consider the blocks with *GSI* greater than 0.2 as we focus on urban climate zones [23]. In addition, we choose the blocks with at least three buildings within their perimeter to filter smaller blocks with just one or two buildings.

Another criterion for selecting the blocks is their shape. Sometimes, the block shapes in a city can be irregular, as shown in figure 2. Such blocks usually have very few or no buildings or contain large landscapes. However, some blocks might have a significant number of buildings that are widely spaced. We use the shape factor (SF) as defined in Ma et al. [19]. To calculate SF, we first construct the minimum bounding circles around the blocks using ArcGIS's *minimum bounding geometry* tool. Thereafter, we compute SF using the following equation:

$$SF = \frac{A_{bl}}{\pi r_{circle}^2} \tag{2}$$

where A_{bl} is the area of the block and r_{circle} is the radius of minimum bounding circle around the block. Preliminary statistical analysis of the blocks in the city indicated that 95% of the blocks have a shape factor greater than 0.15. Further observation of blocks with *SF* less than 0.15 indicates that these blocks are irregular and non-repeating. Therefore, we consider blocks with *SF* greater than 0.15 in this study.

Figure 2 Blocks with irregular shapes

2.3 Parameters affecting microclimate

In this paper, we consider 17 morphological parameters that can potentially affect the microclimate (Table 1). Along with the parameters proposed for classifying LCZs by [23], we identify the parameters based on the categories that broadly define the morphology of a block, namely, form, shape, arrangement, and variations within the block. We also utilise the parameters that influence the wind flow in the urban area. These parameters are observed to influence urban microclimate in the literature, as mentioned in Table 1.

Table 1 Urban morphological parameters

Categories	Parameter		References
Parameters used for LCZ	*SVF*	Sky view factor	[23]
	AR	Aspect ratio	[23]
	GSI	Ground space index/Building surface factor	[16] [19] [42]
	ISF	Impervious surface fraction	[23]
	PSF	Pervious surface fraction	[23]
	HRE	Height of roughness elements	[17] [43] [44]
Arrangement (Density and arrangement of the buildings within the block)	*OSR*	Open Space Ratio	[45]
	MA	Mean building areas	[19] [46]
	NB	Number of buildings per unit area of the block	[47] [26]

Variation (Variations between the buildings within a block)	*SH*	Standard deviation of building heights	[12] [48]
	SA	Standard deviation of building areas	[19] [46]
Form (Compactness/sprawl of buildings within a block)	*DB*	Average distance between the nearby buildings	[17] [19]
	DCR	Ratio of distance from the centroid to buildings to the radius of minimum bounding circle to block ratio	[19]
Shape (Shape factor of the block)	*SF*	Block shape factor	[19] [49]
Wind flow	*FAI*	Frontal area index	[50]
	AH	Average height of the buildings	[51]
	Po	Porosity	[52]

2.3.1 LCZ parameters

The parameters used for classifying LCZs are the sky view factor (*SVF*), aspect ratio (*AR*), building surface fraction (*BSF*), impervious surface fraction (*ISF*), and pervious surface fraction (*PSF*), the height of roughness elements (*HRE*) and terrain roughness class. We do not consider the terrain roughness class (*TRC*) parameter of LCZ classification as the blocks fall into the *'very rough'* class of Davenport classification of effective terrain roughness, where the roughness length is 0.5 m [23].

<u>Sky view factor:</u>

SVF indicates the amount of sky visible from the ground at a given position, referring to the proportion of sky not obstructed by the surrounding built-up [6,53–55]. We calculate the *SVF* using the Relief Visualisation Toolbox of QGIS 3 [56,57].

We use the digital surface model (DSM) and the building footprint dataset to generate the raster with building height information. We consider the open spaces and roads along with the bottom of the buildings at 0 m. Moreover, for better accuracy in urban areas, we only consider the building heights and the obstructions like trees are non-existent in this analysis. We consider a search radius of 100 m and the number of directions as 16 based on [54] for *SVF* calculation.

Based on the output from the relief visualisation toolbox, we aggregate the *SVF* values in the output raster for every block. Therefore, it is crucial that we consider *SVF* values in streets and open spaces. However, as the blocks do not consist of streets surrounding them, we create a buffer zone to include the *SVF* values at street level on the streets surrounding the block. To create the buffer zone, firstly, we calculate the distance of each block from the nearest four blocks located within 33 meters of the block. Thirty three meters is the maximum street width of major streets in the Walloon region [58]. We consider the buffer as follows:

$$D_{\text{buffer}} = max(d_i) \tag{3}$$

where, d_i is the block's distance from the adjacent block, and n is the number of adjacent blocks within 33 meters of the block Figure 3(a).

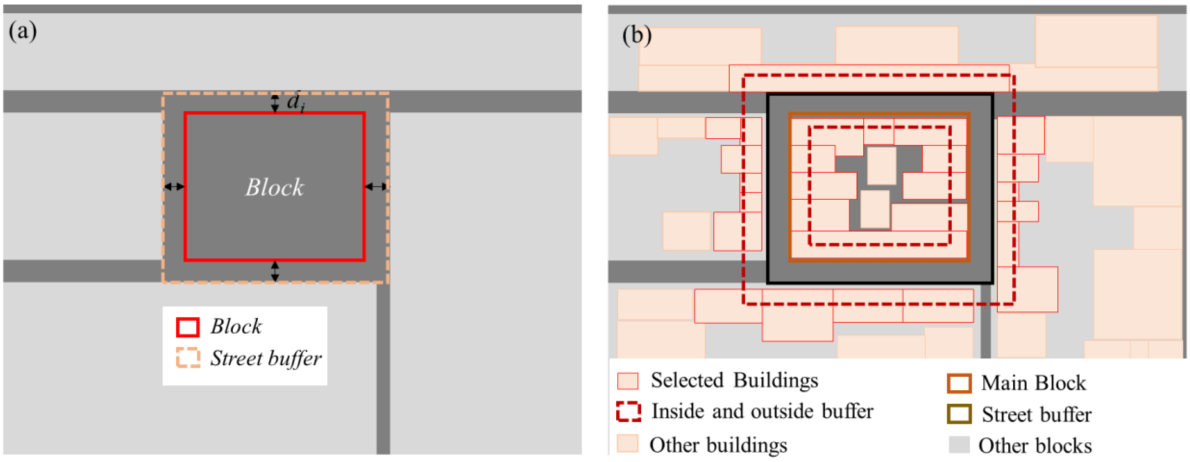


Figure 3 (a) Street width between the block (b) Buffers creates for selecting the buildings on both sides of roads

Next, we remove the buildings from the obtained *SVF* raster by setting the raster values corresponding to the building footprint as null. Thereafter, we consider the average *SVF* values on open spaces and streets in the buffer block as the *SVF* of that particular block. We aggregate the *SVF* values to the blocks using the *zonal statistics as table* tool in ArcGIS Pro 2.9.1.

<u>Aspect Ratio:</u>

Aspect ratio (*AR*) is the building height to street width ratio (H/W). As streets surround blocks on all sides, we consider the average value of *AR* of all streets surrounding the block. We compute the street width based on the distance of blocks from the adjacent blocks (d_i) as shown in figure 3(a). For AR, we consider the average street width calculated as follows:

$$w_{\text{avg}} = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{4}$$

where, *n* is the number of adjacent blocks within 33 meters of the block.

To estimate the building height, we first identify the buildings on both sides of the road. To do this, we create a buffer of 5 m inside the block and 3 m outside the street buffer, as shown in figure 3(b). Then, we select the buildings that are crossed by the outline of these buffers using ArcGIS Pro 2.9.1. After that, we calculate the average height of the selected buildings as follows:

$$h_{\text{avg}} = \frac{1}{n}\sum_{i=1}^{n} h_i \tag{5}$$

where, h_i is the height of the building on either side of the road, and *n* is the number of buildings that are on both sides of the road surrounding the block. Subsequently, we calculate the *AR* of a block as follows:

$$AR = \frac{h_{\text{avg}}}{w_{\text{avg}}} \tag{6}$$

<u>Impervious surface fraction (*ISF*):</u>

ISF indicates the area occupied by impervious surfaces such as pavements, rocks, and buildings. Zha et al. [59] defined the normalised difference built-up index (NDBI) to determine urban and built-up areas. It is used to express the intensity of urbanisation [60] and can be used as a substitute to indicate urban impervious surfaces [61]. Thus, in this paper, we use NDBI as a proxy for *ISF*. We calculate NDBI using the Sentinel-2A satellite imagery captured on 21st July 2021 from the United States geological survey (USGS). We choose the image on this date as July and August experience higher temperatures. Moreover, among the images available for this time frame, the selected image had the lowest and most acceptable cloud coverage of less than one per cent. The NDBI was calculated as follows:

$$\mathrm{NDBI} = \frac{SWIR1 - NIR}{SWIR1 + NIR} \tag{7}$$

where *SWIR1* is the shortwave infrared band (Band 11) with a resolution of 20 m, and *NIR* is the near-infrared band (Band 8) with a resolution of 10 m. For calculating the NDBI, we resample the *SWIR1* band to 10m and compute the NDBI at 10 m resolution. To estimate the NDBI of a block, we calculate the average NDBI of a block as *ISF* using the *zonal statistics as table tool* in ArcGIS Pro 2.9.1.

<u>Pervious surface fraction (*PSF*):</u>

PSF refers to the area occupied by pervious surfaces like bare soil, vegetation or water. Normalised difference vegetation index (NDVI) is used to detect bare soil and vegetation [62,63]. Thus, in this study, we consider NDVI to inform the perviousness in the block. We calculate the NDVI using the Sentinel-2A image used for calculating NDBI. It is computed as follows:

$$\mathrm{NDVI} = \frac{NIR - R}{NIR + R} \tag{8}$$

where *NIR* and *R* are the near-infrared (band 8) and red (band 4) bands with a resolution of 10 m. Similar to NDBI, we calculate the average NDVI of a block as the *PSF* using the *zonal statistics as table* tool in ArcGIS Pro 2.9.1.

<u>Height of roughness elements (*HRE*):</u>

As the roughness of the neighbourhood can influence the aerodynamic properties, we compute the *HRE* for the buffer around the block as used in the *SVF* calculation [50,64]. *HRE* is the average of building heights in the urban canopy. Here, we calculate it as follows :

$$HRE = \frac{\sum_{i=1}^{n} A_{bu(i)} \times h_{bu(i)}}{A_{buffer}} \tag{9}$$

where A_{bu} is the area of building and h_{bu} is the height of the buildings within the outer buffer of the block as shown in figure 3 (b), A_{buffer} is the area of the outer buffer and *n* is the number of buildings within the block. We calculate the height of buildings using digital surface model (DSM) data provided by the geoportal of Wallonia.

2.3.2 Parameters informing arrangement

The parameters in the arrangement category represent the parameters that inform the open spaces and area and the number of buildings. Open space ratio (*OSR*) is defined as the ratio of open areas to the built area, and it describes the intensity of use of non-built ground [45]. We compute it as follows:

$$OSR = \frac{1 - GSI}{GSI} \tag{10}$$

MA is the average area of buildings within the block. *NB* is the number of buildings per unit area in a block. We compute it as follows:

$$NB = \frac{n}{A_{bl}} \tag{11}$$

where *n* is the number of buildings within the block.

2.3.3 Parameters informing variation

The variation category consists mainly of two parameters. Studies have indicated that variation in heights can influence the microclimate [12,48]. In computational fluid dynamics (CFD) studies, the area of the object also affects the wind direction [46]. Thus, we also consider variation in building area in this study as there is a potential effect on the UHI. *SH* is the standard deviation in building heights, and *SA* is the standard deviation in building areas within the block. Since height informs the roughness of the block and its neighbourhood, we consider *SH* to be the standard deviation of building heights within the outer buffer of the block as considered for *SVF* calculations.

2.3.4 Parameters informing the form

The form represents the compactness or sprawl of buildings within the block. *GSI* can also demonstrate the form, but there can be variations in the patterns. Therefore, we consider parameters like *DB* and *DCR*. *DB* is the average distance between the adjacent buildings in a block. We calculate as follows:

$$D_{min} = \frac{1}{n}\sum_{i=1}^{n} \min_{1 \le j \le n-1} (D_{ij}) \tag{12}$$

where D_{ij} is the distance between one building to the rest of the buildings, n is the number of buildings in the block (Figure 4 (a)).

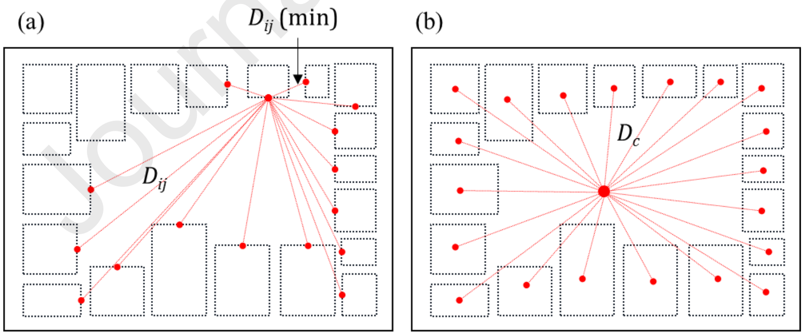


Figure 4 (a) Distance between the buildings (b) Distance from the centre of the block to the building

DC is the average distance between the block's centre and the building's centre (Eq. 13). As the value of *DC* depends upon the block's size, we normalise the parameter using the radius of the minimum bounding circle (Eq. 14). Thus, *DCR* is the parameter indicating the average distance of buildings from the block's centre.

$$DC = \frac{\sum_{i=1}^{n} D_{c(i)}}{n} \tag{13}$$

$$DCR = \frac{DC}{r_{circle}} \tag{14}$$

where, D_{c_i} is the distance from the block's centre to the building's centre, and *n* is the number of buildings in a block.

2.3.5 Parameters for block shape

As mentioned in section 2.2, *SF* informs the shape of the block. Moreover, there might be differences in the blocks based on the shape of the block. Thus, we use this parameter as well in our analysis.

We calculate the parameter values for blocks using the Geopandas package in python (https://geopandas.org/en/stable/). We convert each shapefile to a geodata frame to proceed with further analysis.

2.3.6 Parameters influencing the wind flow

Urban morphology influences the urban air ventilation environment [50,51]. Several indicators such as *GSI*, *SH*, *HRE*, average height (*AH*), frontal area index (*FAI*) and porosity (*Po*) are considered in analysing the urban wind environment as they indicate surface roughness [51,65,66]. In this paper, *GSI*, *SH* and *HRE* parameters are already considered in other categories of urban morphology. Therefore, in this section, we explain the remaining parameters such as *AH*, *FAI* and *Po*. As *AH*, *FAI* and *Po* indicate roughness in the urban block, we consider the buildings in the buffer area (Figure 3(a)) for calculating the indicators.

AH is the average height of buildings within the buffer of the block.

FAI measures building walls facing the wind flow in a particular direction [52]. We compute FAI using the methodology from [65] in this paper. The method involves rasterisation of the building height and area and computing the *FAI* at 100m resolution. The *FAI* is only calculated for northerly/easterly winds. We computed the *FAI* for the blocks using *zonal statistics as table tool* in ArcGIS Pro 2.9.1 and considered the mean of *FAI* pixels overlapping the buffer block as *FAI* of the block.

Po is the ratio of the empty volume in an urban canopy to the volume of the urban canopy. In this paper, we consider the volume of the urban canopy of a block (*UCLV*) as follows [67]:

$$UCLV = max(h_i) \times A_{buffer} \tag{15}$$

where h_i is the height of the buildings in the buffer block. The building volume (*BV*) in the buffer block is computed as follows:

$$BV = \sum_{i=1}^{n} A_i \times h_i \tag{16}$$

where A_i is the area of buildings located in the buffer block. Therefore, *Po* is defined as:

$$Po = \frac{UCLV - BV}{UCLV} \tag{17}$$

2.4 Clustering approach

We use *k-means* clustering in this paper for the reasons explained below:

- Firstly, Liege city does not have any pre-existing classification of urban built form.
- Secondly, the *k-means* clustering algorithm is a hard clustering method that provides distinct clusters.
- Lastly, it is very efficient for large and high-dimensional datasets.
- Moreover, several studies [32–34] have demonstrated that *k-means* have provided logical results for identifying urban typologies.

2.4.1 Pre-processing of data

It is crucial to scale and normalise the data before performing the *k-means* clustering [35,68], as the presence of outliers or skewed distributions can influence the optimal number of clusters generated by algorithms. Therefore, first, we scale the data to guarantee that no particular weight is given to any specific variable or feature.

Researchers often couple principal component analysis (PCA) with *k-means* in order to reduce dimensionality and ensure non-collinearity among the variables [33,69]. Moreover, PCA-based *k-means* are observed to generate a better clustering result [70]. A PCA is a linear transformation of variables into reduced dimensional space while retaining the maximum variance [33,35].

In this paper, we are not confronted with a large number of dimensions. However, there may be collinearity between the variables. As collinearity may influence the results, we do a PCA analysis to obtain principal components (PCs) that are non-collinear and explain maximum variance in the data. We transform the data into PCs using classical PCA and decide on the number of PCs based on Kaiser criteria [71]. Thus, we select the PCs based on the cumulative variance explained by each component that has an eigenvalue greater than one. We further discuss the loading of each parameter on the chosen PCs using varimax rotation to estimate their influence on the clustering outcome. We consider the rotated principal component (RC) scores as input for the k-means clustering. We use the R *stats* package version 4.0.3. for PCA analysis.

2.4.2 *K-means* clustering

K-means clustering algorithm is an unsupervised clustering technique that attempts to determine *k* non-overlapping clusters to maximise the distance between the clusters and minimise the distance within the cluster. Given the set of *n* data points and a predefined number of clusters (*k*), the algorithm randomly selects *k* cluster centres initially. It classifies the data points to the nearest cluster centre. Then, it calculates the within-cluster sum of squares and reassigns the cluster centres to result in a final partition that optimises the clustering quality by minimising the intracluster sum of squares distances of any data point to its nearest cluster centre as defined by the following equation [72]:

$$J(C) = \sum_{i=1}^{k} \sum_{j=1}^{n} \parallel x_j - c_i \parallel^2 \tag{18}$$

K-means algorithm generally picks up the centroids randomly. Thus, the result depends upon how the initial centroids were selected. Therefore, to avoid this problem, we use the *k-means* ++ initialisation which is a smart centroid initialisation technique. With this technique, the best possible initial centroid is selected and the replicability of results is ensured with significant iterations [73]. We use the *clusterR* package of R version 4.0.3 for clustering the data using the *k-means* algorithm.

2.4.3 Determining the number of clusters

There are numerous varieties of methods available to identify the number of clusters. However, identifying the optimal number of clusters is always challenging for clustering analyses. Thus, many studies use more than one method to determine an optimal number of clusters [74,75]. In this paper, we use the cubic clustering criterion (*CCC*) and pseudo *F* statistic to select the optimal number of clusters.

CCC is a test statistic developed by the SAS programming package [76] for identifying an optimal number of clusters. This index is the measure of within-cluster homogeneity compared to between-cluster heterogeneity. For identifying the optimal number of clusters, the values of *CCC* are plotted against the number of clusters and the peak value is chosen as appropriate. However, the peak value of *CCC* should be positive and preferably greater than two or three [77].

Caliñski and Harabasz [78] developed the pseudo *F* statistic and defined it as the ratio of between-cluster variance to within-cluster variance. Similar to *CCC*, the pseudo *F* values are plotted against the number of clusters and the peak value is chosen as the optimal number of clusters. The large peaks in the pseudo *F* statistic are indicators of greater cluster separation.

Milligan and Cooper [77] examined 30 indexes developed to identify the number of clusters. According to the study, *CCC* performed at a competitive rate; however, it may suggest too many clusters in some cases. Additionally, pseudo *F* statistic has performed well and is less prone to errors, according to Milligan and Cooper [77]. Thus, we compare these two indices and decide the optimal number of clusters. To compute these indices, we use the *NbClust* package of R.

2.4.4 Validating the clusters

For external criterion analysis, standard parametric analyses such as analysis of variance (ANOVA) or multivariate analysis of variance (MANOVA) are used to validate the clustering result using a variable excluded from clustering analysis [36].

Differences in surface temperature are related to different microclimates. So, the average surface temperature of the blocks can be one of the proxies for measuring microclimate conditions [79,80]. In this study, the external variable cannot be a parameter that influences or informs urban morphology. Thus, we choose the block's average land surface temperature (LST) as a dependent variable for external criterion analysis. We compute it using *zonal statistics as table tool* in ArcGIS Pro 2.9.1. We consider using a one-way ANOVA analysis to validate the clustering result as we have one parameter. Moreover, several studies have effectively used one-way ANOVA for validating the clustering analysis results [81–83]. Therefore, we validate the clustering result with the help of a one-way ANOVA to see whether the mean land surface temperature (LST) of the blocks varies across the clusters.

We calculate the land surface temperature (LST) using the LANDSAT-8 level 1 image captured on 18th July 2022. We choose the image on this date as July and August experience higher temperatures. Among the images available for July and August, the image on this date had the lowest and most acceptable cloud coverage of less than one per cent. In addition, we procured the image from the USGS at a resolution of 30 m. We use the thermal band 10 to compute the LST (in Kelvin (K)) using the following equations [84]:

$$L_\lambda = M_L Q_{cal} + A_L \quad (19)$$

where L_λ = TOA (Top of Atmosphere) spectral radiance (Watts/(m2 * srad * µm)), M_L = Band-Specific multiplicative rescaling factor from the metadata, A_L = Band-specific additive rescaling factor from the metadata, Q_{cal} = Quantized and calibrated standard product pixel values (DN)

$$T = \frac{K_2}{ln\left(\frac{K_1}{L_\lambda} + 1\right)} \quad (20)$$

where T = TOA brightness temperature (K), K_1 = Band-specific thermal conversion constant from the metadata, K_2 = Band-specific thermal conversion constant from the metadata

We further convert the LST values to degrees Celsius (^{o}C).

ANOVA analysis enables comparing variances of more than two populations to determine equality of means. The F-test is performed against the null hypothesis, where the means of LST of each cluster are equal. The alternate hypothesis would be that not all the means of LST are equal. If the p-value of the *F*-statistic (Pr (>F)) is less than 0.05, then the null hypothesis will be rejected, and the alternate hypothesis will be accepted.

2.5 Determining predefined LCZ for blocks

To compare the clustering results with LCZ, we first determine the LCZ of the blocks using the LCZ map by WUDAPT [38] for entire Europe. The LCZ map [37] is at a resolution of 100 m, with each pixel indicating the type of LCZ. We use *zonal statistics as a table* tool of ArcGIS Pro 2.9.1 and identify the value of the majority of pixels in the block as LCZ.

3 Results and discussion

3.1 Selected blocks

Figure 5(a) depicts the total blocks in the city of Liege. There are a total of 1,441 blocks in the city. Out of these blocks, we select 1007 blocks in total for the analysis based on the selection criteria explained in section 2.1. Figure5(b) illustrates the selected blocks in the city. From figure 5(b), we observe that the larger blocks with fewer buildings are effectively filtered along with irregularly shaped blocks.

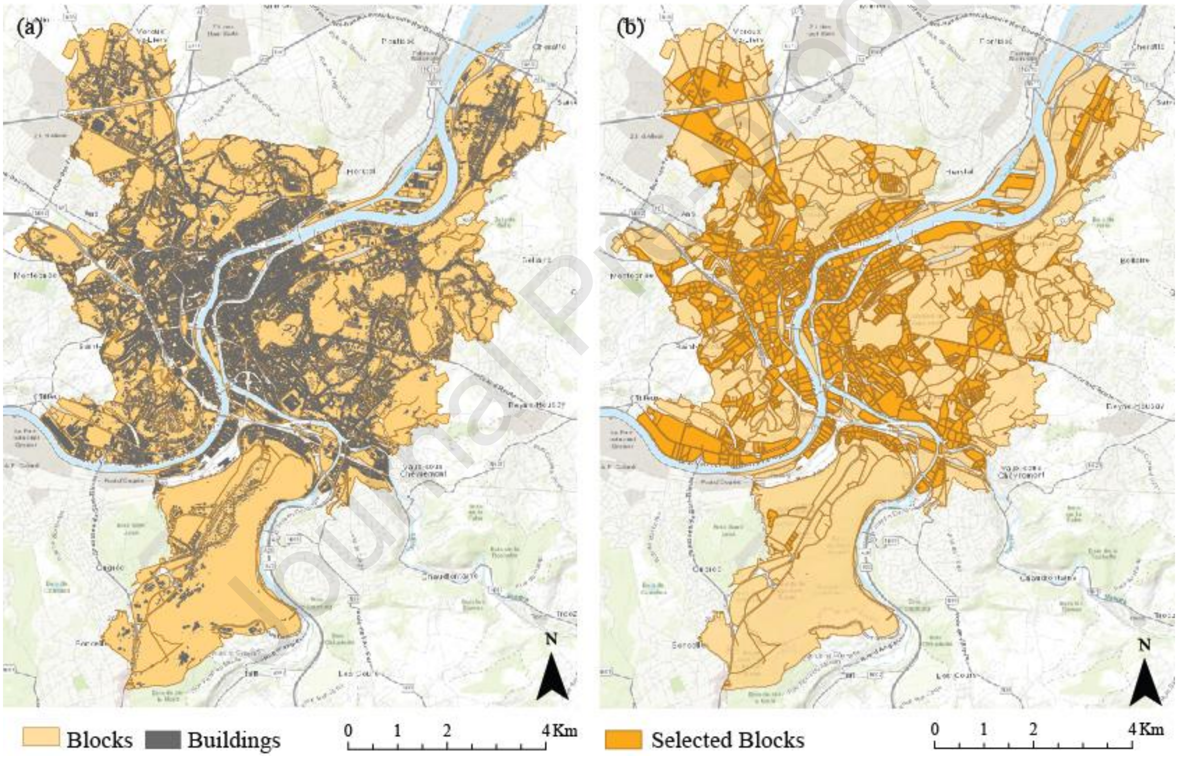


Figure 5 (a) Blocks in the city of Liege (b)Selected blocks for this study

3.2 Principal component analysis

Pairwise correlation of the 17 parameters indicates a high (>50%) and significant correlation between some variables, as shown in figure 6. The correlated variables may influence the clustering results, given the higher magnitude. Therefore, PCA analysis is relevant in this case.

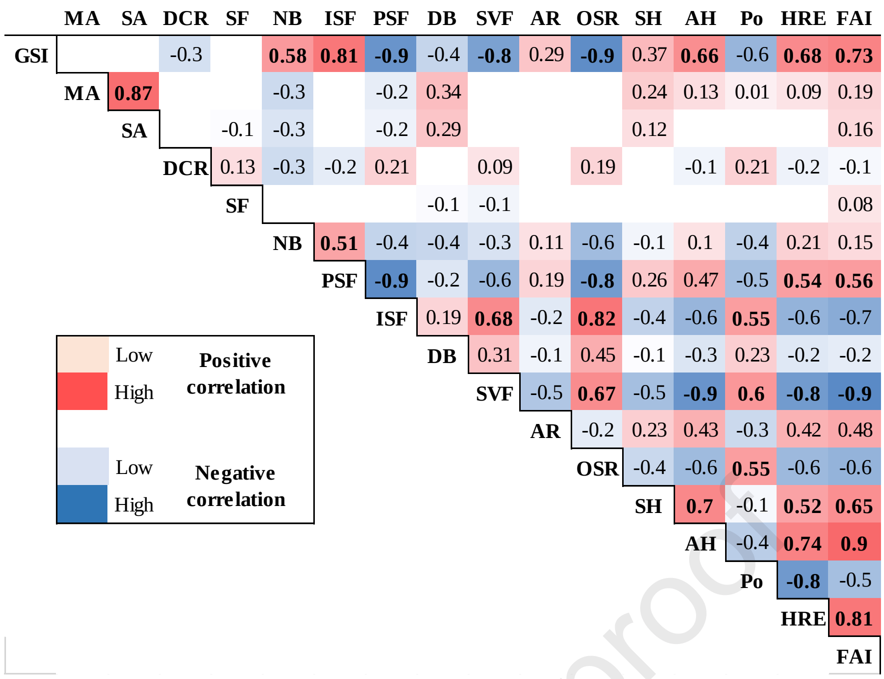


Figure 6 Correlation Matrix

We observe from figure 7 that first four principal components have eigenvalue greater than one. Moreover, these four PCs explain around 75% of the variance in the dataset. Therefore, we select these four PCs as variables for clustering.

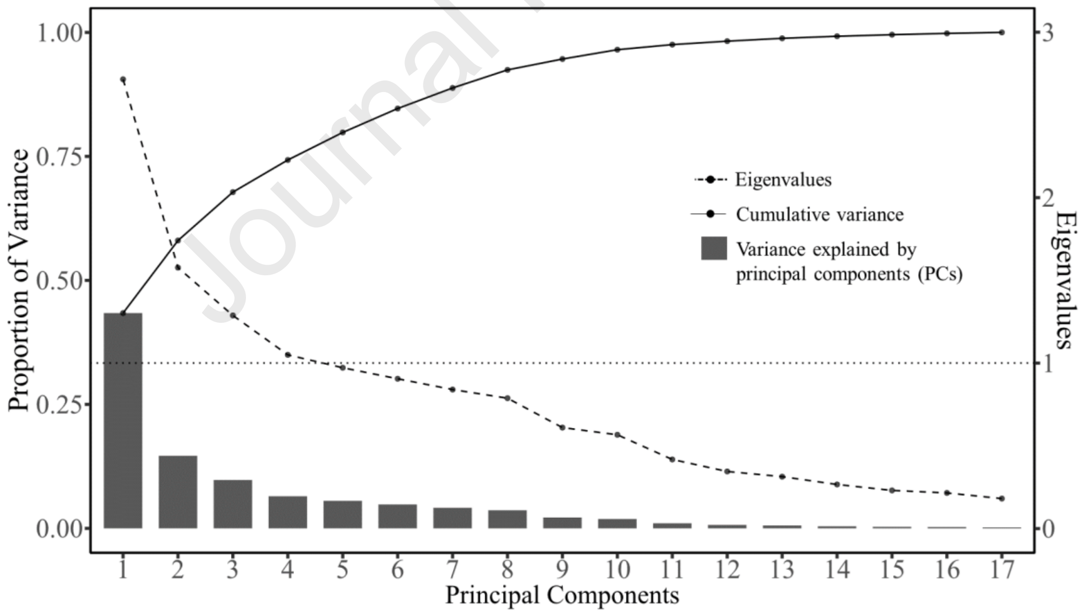


Figure 7 Principal Components Analysis

Table 2 Loadings of parameters on PCs and RCs

		PC loadings			
Parameters		PC1	PC2	PC3	PC4
LCZ	*GSI*	0.338	-0.099	0.162	-0.068
	PSF	-0.318	-0.022	-0.242	0.164
	ISF	0.293	-0.092	0.288	-0.195

	SVF	-0.332	-0.047	0.216	-0.019
	AR	0.164	0.051	-0.278	0.394
	HRE	0.314	0.080	-0.138	0.135
Arrangement	*OSR*	-0.315	0.119	-0.190	0.200
	MA	0.031	0.536	0.312	-0.058
	NB	0.163	-0.409	0.282	0.040
Variations	*SH*	0.200	0.277	-0.277	-0.009
	SA	0.017	0.498	0.344	-0.104
Form	*DCR*	-0.087	0.111	-0.309	-0.518
	DB	-0.135	0.318	0.162	0.169
Shape	*SF*	0.027	-0.031	-0.246	-0.615
Wind environment	*AH*	0.305	0.153	-0.252	0.081
	Po	-0.262	0.087	-0.060	-0.129
	FAI	0.328	0.176	-0.180	0.049
Eigen Values		2.717	1.577	1.288	1.050
Explained variance		43%	15%	10%	6%
Cumulative variance		43%	58%	68%	74%

RC loadings

Parameters		**RC1**	**RC2**	**RC3**	**RC4**
LCZ	*GSI*	**0.851**	0.000	-0.433	0.000
	PSF	**0.885**	0.000	-0.223	0.000
	ISF	**-0.838**	-0.176	0.374	0.000
	SVF	-0.481	0.000	**0.808**	0.000
	AR	0.000	-0.147	**-0.656**	0.227
	HRE	0.446	0.000	**-0.771**	0.000
Arrangement	*OSR*	**-0.875**	0.000	0.319	0.000
	MA	0.000	**0.935**	-0.107	0.000
	NB	**0.698**	-0.413	0.12	0.275
Variations	*SH*	0.000	0.211	**-0.725**	-0.19
	SA	0.000	**0.906**	0.000	0.000
Form	*DCR*	-0.269	0.000	0.000	**-0.683**
	DB	-0.362	**0.526**	0.144	0.18
Shape	*SF*	0.000	-0.123	0.000	**-0.707**
Wind environment	*AH*	0.332	0.000	**-0.86**	0.000
	Po	**-0.56**	0.113	0.432	-0.191
	FAI	0.428	0.114	**-0.854**	0.000
Eigen Values		4.723	2.297	4.386	1.222
Explained variance		28%	14%	26%	7%
Cumulative variance		28%	41%	67%	74%

Table 2 shows principal component loadings and rotated component loadings obtained with varimax rotation. We observe that it is difficult to interpret the loadings from the PCs as some parameters have significant loadings on more than one PC (for example, GSI, DCR and OSR). The Varimax rotated solution (RCs) results in large loadings on a single component and small cross-loadings on the other components, facilitating the interpretation.

The RC1 and RC3 account for the higher variance in the data, which is about 28% and 26%, respectively. Thus, the following parameters, such as *GSI*, *PSF*, *ISF*, *SVF*, *AR*, *HRE*, *OSR*, *NB*, *SH*, *AH*, and *FAI*, influence the clustering the most. These parameters mainly correspond to the LCZ parameters, the parameters informing the arrangement and the parameters related to the wind environment. Parameters related to variation in the block, form and shape account for 21% of the variance (combined variance of RC2 and RC4), indicating their significance as well.

3.3 Number of clusters (*k*)

Figure 8 demonstrates the *CCC* and pseudo *F* statistic values for the number of clusters (*k* = 1 to 30). The value of *CCC* peaks first at *k*= 11, then at *k*=13, followed by *k*=23 and lastly, *k*= 27. As per Sarle [76], the highest value of *CCC* corresponds to the optimal number of clusters. However, in the case of distinct non-hierarchical elliptical clusters, the graph often shows a sharp rise to the correct number of clusters, followed by a gradual increase and eventually a gradual decline. In the plot of CCC (figure 8), a sharp rise is observed at k=9.

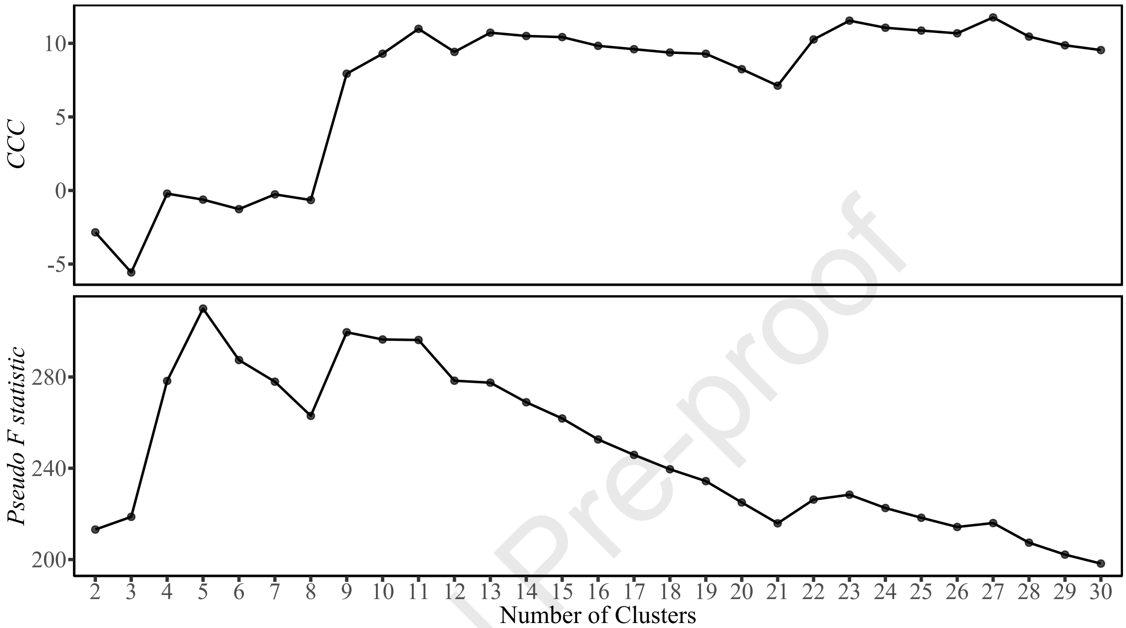


Figure 8 *CCC* and pseudo *F* statistic values for different numbers of clusters (k)

The pseudo *F* statistic value, on the other hand, peaks at *k*=5, then at *k*= 9, followed by *k*=9, k=11, k=13, k=23 and k=27. As the suitable value matches at k=9, we choose the number of optimal clusters as 9.

3.4 Validation – ANOVA analysis

Figure 9 demonstrates the mean LSTs of blocks within each cluster, whereas figure 10 depicts the spatial variation of the mean LST of blocks. We observe from figure 9 that there is a variation in the average mean LST across the clusters. Moreover, the variation can also be observed in figure 10 as the blocks in the city centre have higher LST as compared to the blocks in the outskirts of the city.

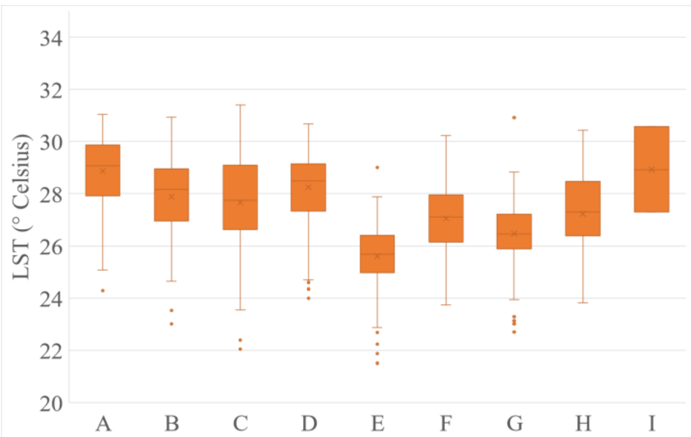


Figure 9 Cluster-wise mean LSTs of blocks

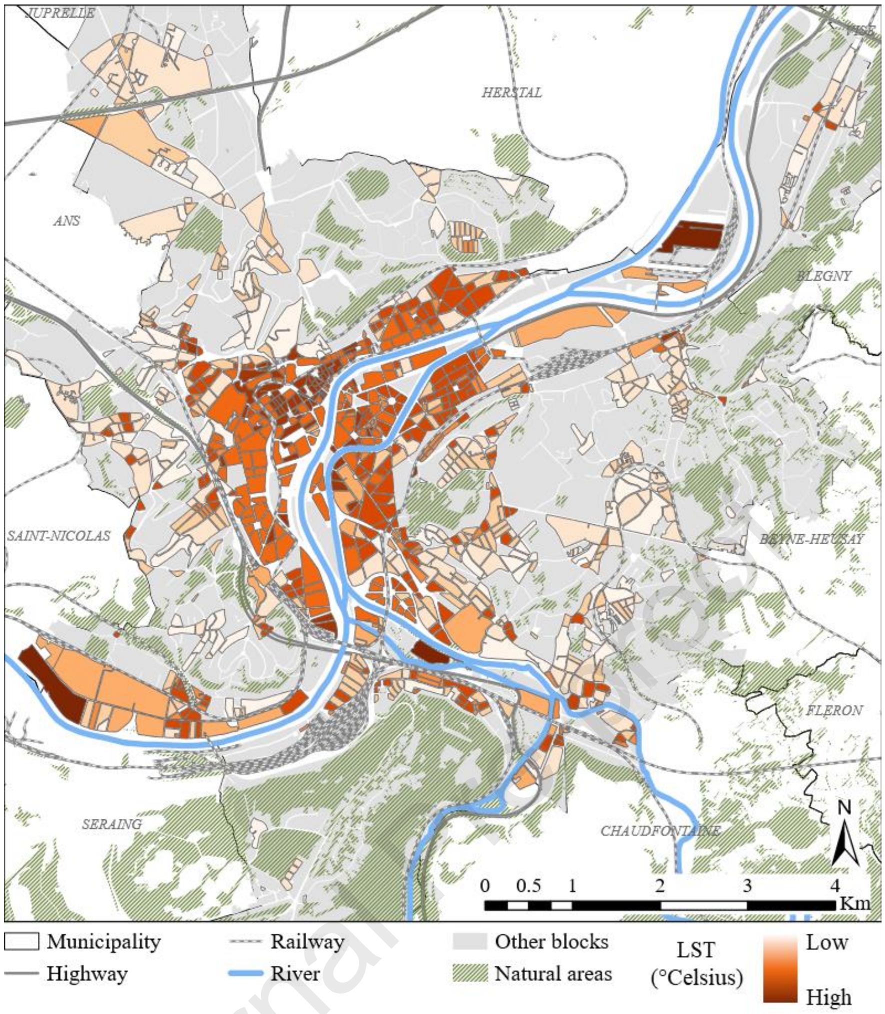


Figure 10 Spatial variation of mean LST of clusters

Table 3 provides the details of the ANOVA test. The *p-value* of the *F-statistic* is less than 0.01, implying that the means of LST in clusters A to I are not equal, indicating that the clusters are different from each other. Therefore, the ANOVA analysis validates the result of clustering. Altogether, the clustering result is acceptable, and the clusters are different from each other.

Table 3 Model summary of ANOVA

	Df (degrees of freedom)	*Sum of squares*	*Mean sum of squares*	*F-statistic*	*Pr(>F)*
Clusters	8	1117	139.59	79.36.95	0.000

3.5 Features of morphological clusters

After applying k-means to the four RCs, we obtain 9 clusters in the city of Liege. Clusters D, E, F and G have the largest number of blocks in Liege city, followed by the clusters C, B and A. The clusters H and I have the lowest number of blocks (Figure 11). Therefore, we consider blocks closest to the cluster centres and identify them as morphological archetypes. Given the intra-cluster homogeneity, these blocks can effectively represent the clusters. Figure 12 provides a two-dimensional view of the morphological archetypes in Liege obtained based on the clustering. Table A.1 (Appendix) provides the values of parameters of the archetypes. Figure 13 (a) demonstrates the spatial distribution of clusters in the city.

The clusters in the city centre consist of compact blocks with low-rise to high-rise buildings (Clusters A, B, C and D). The clusters on the fringe of the city are open and sparsely built compared to the clusters in the inner city (Clusters E, F, and G). Other clusters (clusters H and I) are spread across the city, and the blocks in these clusters have large-sized and a few buildings, with mostly homogenous mid-rise to low-rise buildings. Significant variation between the clusters in terms of morphological parameters can be observed in figure 14. Based on these characteristics and the LCZ nomenclature provided by Stewart and Oke [23], we name the clusters as given in figure 12.

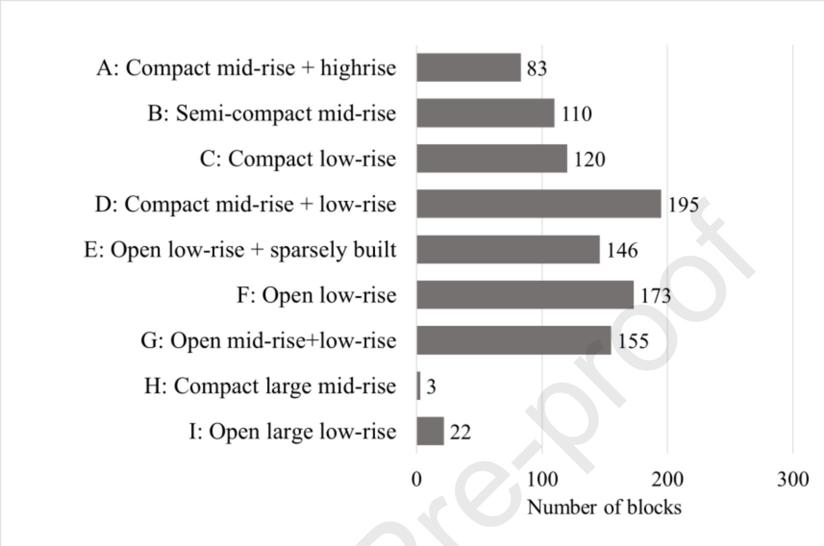


Figure 11 Number of blocks per cluster

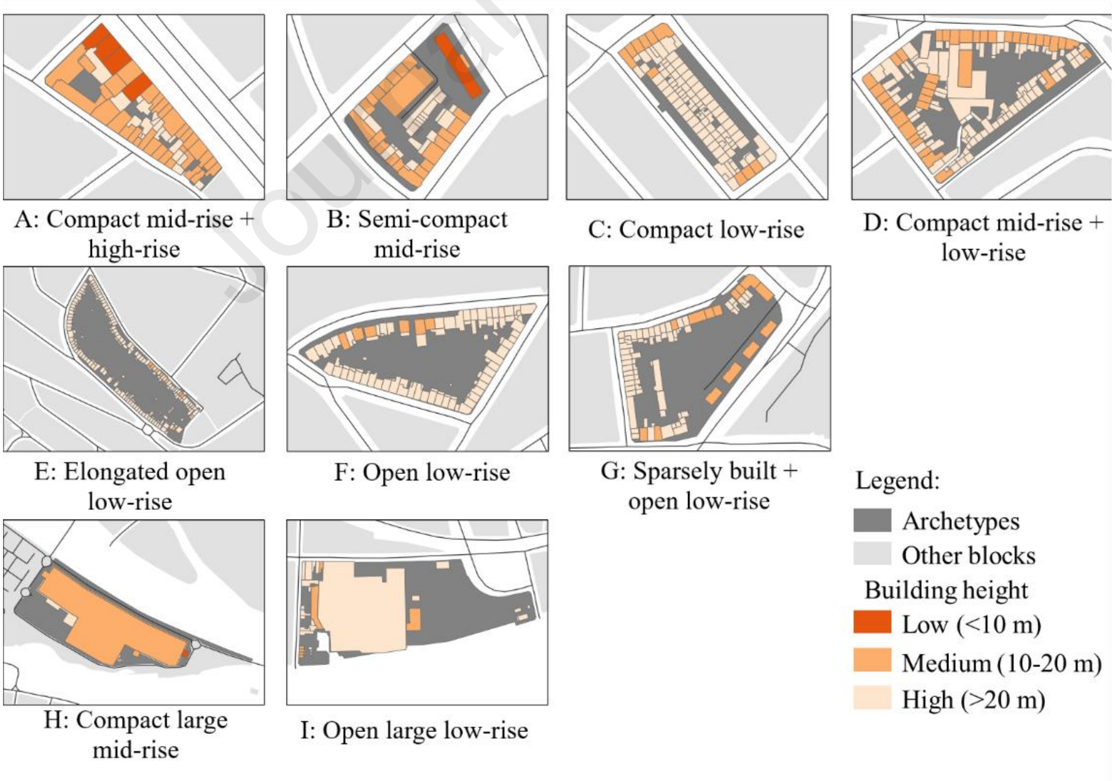


Figure 12 Two-dimensional view of morphological archetypes based on clustering

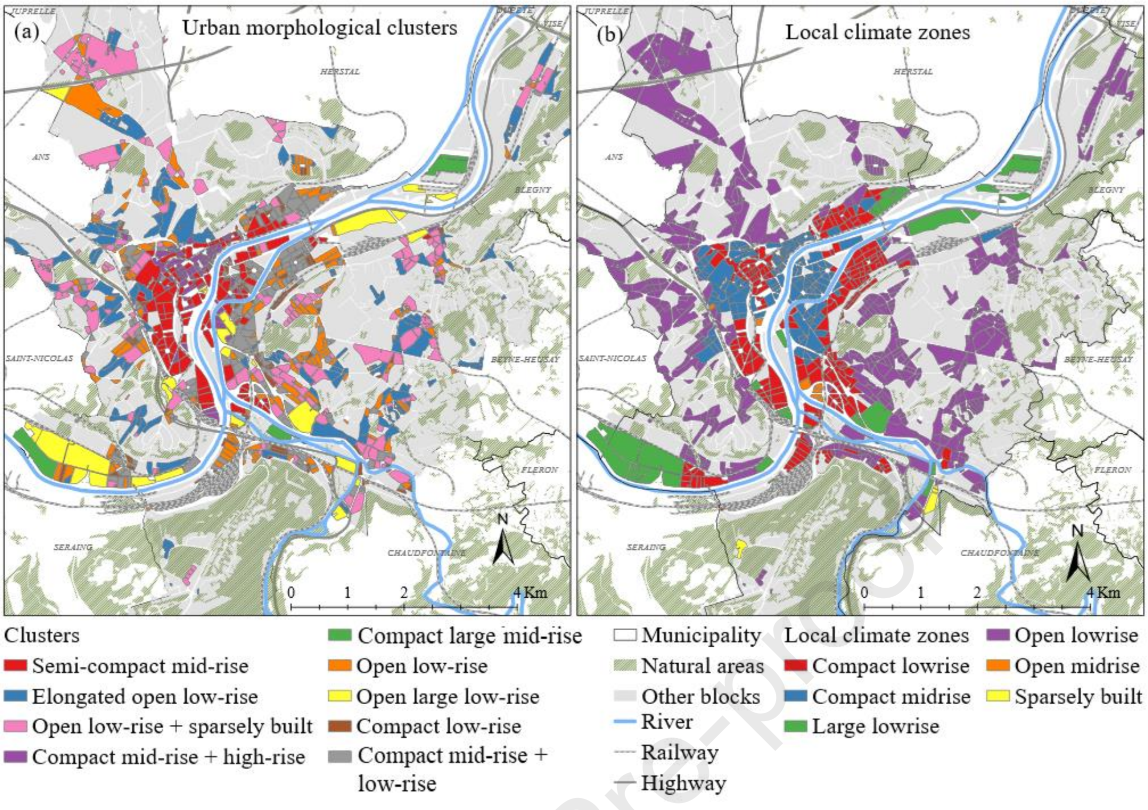


Figure 13 (a) Spatial distribution of clusters (b)Spatial distribution of LCZs

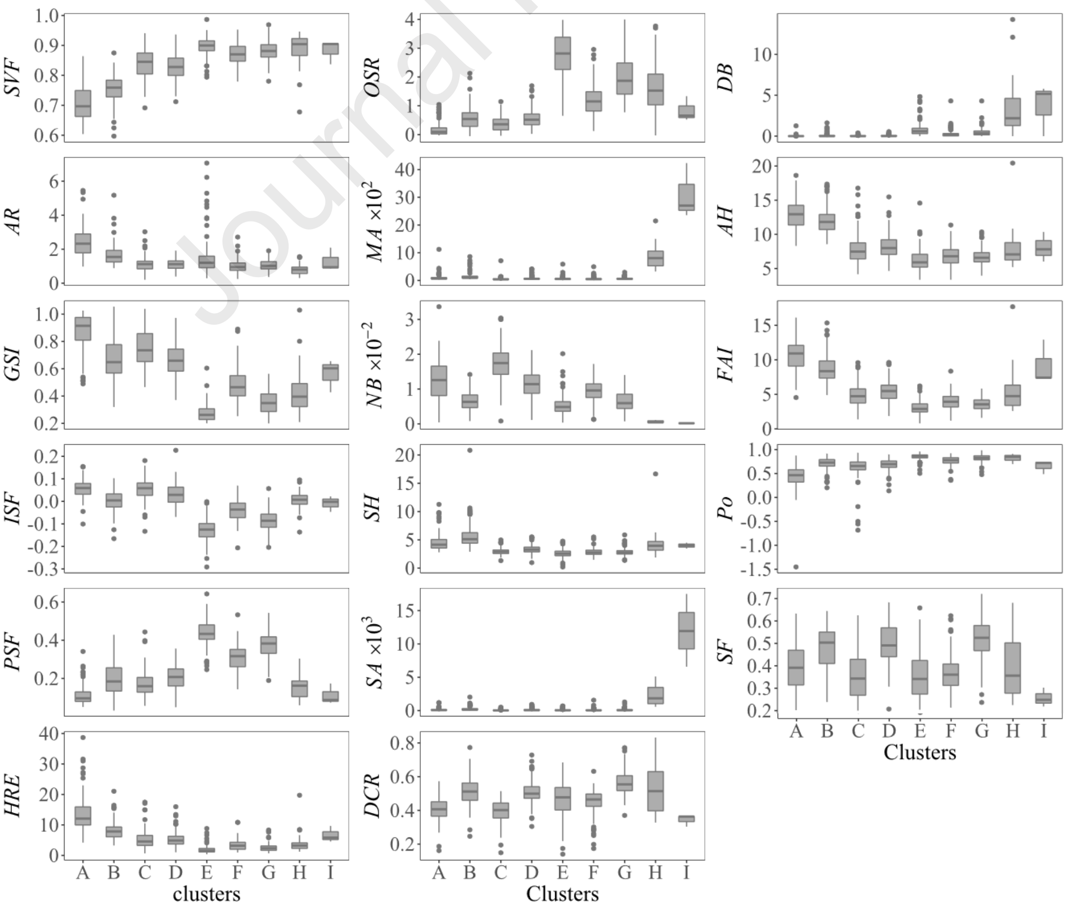


Figure 14 Variations in the values of parameters for different clusters

3.6 Comparing clusters with LCZ

As per figure 13 (b), Liege city is classified mainly into four LCZs: Open low-rise, compact low-rise, compact mid-rise, and large low-rise. A few blocks in the city are classified into open, mid-rise and sparsely built. According to LCZ classification, clusters A and B blocks are predominantly compact mid-rise (Table 4). As cluster A has the largest *GSI* and a significantly higher *HRE*, it fits the description of the corresponding LCZ. However, cluster B has a moderate *GSI* value but a comparable *HRE*. Therefore, an almost equal share of blocks in this cluster corresponds to the compact low-rise classification of LCZ. Moreover, around 39% of the blocks in cluster A are classified as compact low-rise.

Clusters C and D are the clusters with blocks mostly on the city's outskirts. The *GSI* range of these clusters largely falls within the specified range of LCZ class of compact low-rise (table 5). However, there are variations between the clusters in terms of the arrangement (*NB*) and the shape of the blocks (*SF*). There are also slight differences in terms of building heights. For example, blocks in cluster D have taller buildings than the blocks in cluster C. Consequently, the *AR* values are higher for blocks in cluster D as compared to blocks in cluster C. Additionally, an almost equal share of blocks of cluster C corresponds to open low-rise as per LCZ classification (table 4). Thus, clusters C and D represent distinct sub-classes within the LCZ of open low-rise.

Blocks in clusters E, F, and G are predominantly open low-rise as per LCZ classification. Around 80% of these blocks in the cluster fall into the open low-rise LCZ category (Table 4). However, they are different in terms of the shape of the block (*SF*) and open space in the block (*OSR*). Therefore, these clusters form essential sub-classes within the LCZ type of open low-rise.

The blocks in clusters H and I mainly belong to the large low-rise type of LCZ. These clusters have blocks with a large building area. However, cluster I has a *GSI* of 0.65, whereas the blocks in cluster H have a lower *GSI*. Thus, cluster I is more compact as compared to cluster H. The main characteristic of the blocks in these clusters are the large-sized buildings, so they are the essential sub-classes of the large low-rise LCZ classification. Figure 15 shows the proportion of clusters (sub-classes) in the existing LCZ classification. The distribution of sub-classes in compact mid-rise, compact open low-rise and large low-rise types of LCZ is almost equal. However, the dominant sub-class in the compact low-rise type of LCZ is the semi-compact low-rise cluster.

Table 4 Percentage of blocks in a cluster classified as an LCZ

Morphological clusters	Compact mid-rise	Compact low-rise	Open mid-rise	Open low-rise	Large low-rise	Sparsely built
A: Semi-compact mid-rise	44%	41%	5%	6%	1%	
B: Compact mid-rise + high-rise	57%	39%		2%	2%	1%
C: Compact low-rise	18%	39%		38%	3%	
D: Compact mid-rise + low-rise	23%	40%	1%	29%	5%	
E: Elongated open low-rise	2%	1%		96%	1%	
F: Open low-rise	4%	20%		73%	3%	
G: Open low-rise + Sparsely built	2%	8%	1%	86%	1%	
H: Compact large mid-rise				33%	67%	1%
I: Open large low-rise	18%	14%		23%	45%	1%

Table 5 Comparing parameter values of LCZs and archetypes

LCZs	Clusters	*GSI*	*BSF* (LCZ)	*HRE*	*HRE* (LCZ)	*AR*	*AR* (LCZ)	*SVF*	*SVF* (LCZ)

Compact mid-rise	Compact mid-rise + high-rise	0.89	0.4 - 0.7	13.29	10 -- 25	2.85	0.75 - 2	0.69	0.3 - 0.6
	Semi-compact mid-rise	0.57		9.69		0.88		0.78	
Compact low-rise	Compact low-rise	0.72	0.4 - 0.7	6.52	3 -- 10	1.08	0.3 - 0.75	0.808	0.2 - 0.6
	Compact mid-rise + low-rise	0.63		4.29		1.54		0.812	
Open low-rise	Elongated open low-rise	0.29	0.2 - 0.4	1.94	3 -- 10	1.47	0.3 - 0.75	0.876	0.6 - 0.9
	Open low-rise	0.43		2.9		0.72		0.864	
	Open low-rise + sparsely built	0.32		2.45		0.97		0.869	
Large low-rise	Compact large mid-rise	0.65	0.3 - 0.5	5.81	3 -- 10	2.09	0.1 - 0.3	0.9	> 0.7
	Open large low-rise	0.49		3.07		0.74		0.92	

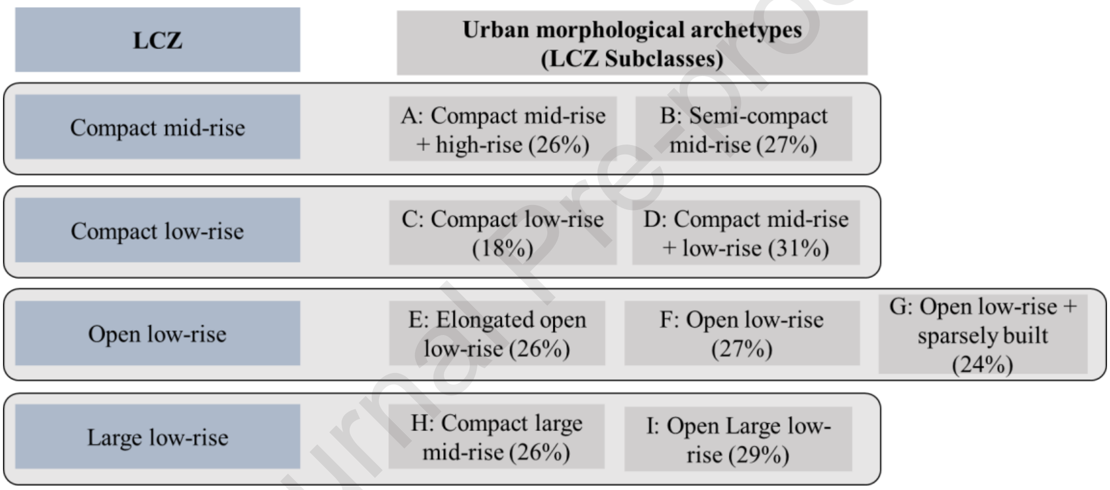


Figure 15 Proportion of clusters (sub-classes) in the existing LCZ classification

Altogether, we observe that the classification in this paper broadly corresponds to the LCZ classification. However, our approach provides more detailed sub-classes, which are necessary to capture the microclimatic variations in the city (figure 14). We also observe that there are a few of the blocks misclassified as open low-rise when they are supposed to be in the compact mid-rise or low-rise category and vice versa. Furthermore, the differences in terms of morphological parameters within the sub-classes of each LCZ are noteworthy and can affect the microclimate and UHI. Thus, the clustering approach has produced more meaningful homogenous clusters which deliver logical LCZs and sub-classes of LCZs suited to the region.

4 Discussion and Conclusions

Urban morphological archetypes are the urban blocks that represent a homogenous group of blocks in terms of urban morphology, identification of which is vital for microclimatic analyses. In this paper, we propose a well-defined PCA-based *k-means* clustering approach supported by an external criterion validation using ANOVA analysis to identify urban archetypes. We choose the *k-means* clustering approach as it is an unsupervised data-driven algorithm that is robust in the absence of an existing classification of built-form in the city. We use seventeen urban morphological parameters based on LCZ, the categories such as form, shape, arrangement, and variations within the block, along with the parameters influencing the wind flow, defining the morphology of the block. Moreover, we support the choice of morphological parameters by identifying their influence on the microclimate, as mentioned in the literature.

We propose the following steps to identify urban morphological archetypes for microclimate studies.

1. Filter the urban blocks based on building density, regularity of block shape and the number of buildings per block to identify blocks fit for analysis.
2. To reduce the dimensionality and verify non-collinearity among the variables, use varimax rotated PCA to transform the urban morphological parameters into principal components that explain the most variance present in the data.
3. Cluster the RCs using the *k-means* clustering algorithm with *k-means* ++ initialisation; identify the best number of clusters using the *CCC* and pseudo *F* statistic.
4. Validate the clusters using ANOVA analysis using the mean LST of blocks as a dependent variable, which is one of the proxies for measuring microclimate quality.
5. Identify the blocks that are nearest to the cluster centre as an urban morphological archetype.

We apply the aforementioned steps to the city of Liege in Belgium. Firstly, by filtering the city blocks using *GSI* and *SF*, we obtain 1007 blocks qualifying for the analysis. Subsequently, we obtain four principal components based on the PCA, which are rotated using varimax rotation before clustering them using the *k-means* algorithm. We determine the number of clusters based on the best values of indices *CCC* and pseudo *F* statistic, resulting in nine homogenous clusters different from each other. Lastly, the validation indicates that the clustering is also successful in terms of LST, making the clusters logical for analysing the microclimate. Finally, we put forward a representative block from each cluster resulting in nine urban morphological archetypes that can be used as input for microclimatic analyses. We also compare the clusters with the existing LCZ map for Europe. The clustering-based classification in this study broadly corresponds to the LCZ classification. However, LCZ fails to capture morphological variety that can influence the microclimate. The approach used in this study identifies the relevant sub-classes that fall within the broad LCZ classes.

Although LCZ maps are available for Europe, the method proposed in this study is advantageous for the following reasons: the classification accuracy of LCZ maps is 70% [38] and using them directly might include misclassifications, as explained in section 3.7. Researchers also compute LCZs for specific regions for better accuracy [26,85]; however, implementing a methodology to estimate LCZs in a city is not straightforward. Due to the wide-ranged parameter values of LCZ, there are often overlaps in the LCZ categories. Thus, sub-classes or combinations of various LCZ classes are often proposed as it gets extremely challenging to assign LCZ class to a block [6]. Moreover, the LCZ dataset is insufficient as the urban canopy must be described explicitly for analysing microclimate [7]. Furthermore, integrating raster-based LCZs into urban planning can be challenging as urban blocks are the urban design unit for urban planners and architects [28]. The urban block scale is also considered appropriate to analyse the heterogeneity of microclimate within the urban fabric [86]. Therefore, urban blocks are the appropriate units for identifying homogenous climate zones in a region.

The archetypes identified using this method can be used as a database to inform urban planners in optimising the urban forms to regulate the microclimate [87]. The approach used in this study is also helpful in the absence of predefined typologies or classifications. It can effectively be applied to other cities worldwide for analysing the microclimate based on urban morphology. For instance, we apply it to Liege city, but the approach can be applied to the entire Wallonia region or Belgium to identify local archetypes for microclimate analysis. Although the approach is straightforward, it is essentially dependent on the data availability. Thus, preparing the data can be challenging when the datasets are not readily available. Approaches such as the Geoclimate tool by [66] can help derive LCZ parameters. Further studies can include developing data using open-access datasets to identify urban morphological parameters.

The present approach aids the microclimatic analysis by providing realistic urban blocks for microclimate analysis, including CFD simulations instead of a simplistic representation of urban blocks. Furthermore, it reduces the computational expense of analysing the microclimate at a higher resolution as it alleviates the need for simulating the microclimate of an entire city. Instead, the simulations can be carried out on the identified archetypes to arrive at a general overview of the microclimatic situation in the city. Moreover, the properties of these archetypes can be helpful to further generate modelled blocks in the city that will be a better representation of reality.

Appendix

Table A.1 provides the values of parameters of the archetypes.

Table A.1 Characteristics of morphological archetypes belonging to each cluster

Cluster	*SVF*	*AR*	*GSI*	*ISF*	*PSF*	*HRE*	*MA*	*SA*	*DCR*	*DB*	*SF*	*NB*	*OSR*	*SH*	*AH*	*Po*	*FAI*
A	0.69	2.85	0.90	0.05	0.09	13.30	80.63	71.37	0.43	0.00	0.35	0.01	0.11	4.55	12.57	0.51	10.24
B	0.78	0.88	0.57	-0.01	0.20	9.70	109.77	149.22	0.41	0.00	0.55	0.01	0.75	6.76	10.69	0.76	7.68
C	0.81	1.08	0.72	0.04	0.19	6.52	33.61	18.21	0.42	0.00	0.35	0.02	0.39	3.08	7.54	0.63	5.84
D	0.81	1.54	0.63	0.03	0.21	4.29	46.65	61.27	0.58	0.01	0.44	0.01	0.58	2.99	7.78	0.70	5.31
E	0.88	1.48	0.29	-0.15	0.45	1.94	41.44	24.87	0.54	0.48	0.28	0.01	2.43	2.42	6.09	0.83	3.58
F	0.86	0.72	0.44	-0.07	0.33	2.91	45.49	27.60	0.49	0.20	0.33	0.01	1.29	2.37	6.72	0.80	4.45
G	0.87	0.98	0.33	-0.08	0.39	2.46	54.61	45.16	0.56	0.48	0.48	0.01	2.03	3.00	7.18	0.82	3.39
H	0.90	2.09	0.65	-0.05	0.17	5.82	4234.20	11929.14	0.30	0.00	0.22	0.00	0.53	4.50	10.35	0.74	7.35
I	0.92	0.74	0.49	-0.02	0.17	3.08	807.40	3789.92	0.70	2.60	0.35	0.00	1.03	3.17	8.40	0.79	4.61

Acknowledgement

The research was funded through the ARC grant for Concerted Research Actions for project number 19/23-28 “CityRoof” financed by the French Community of Belgium (Wallonia- Brussels Federation). This work was also supported by FRS-FNRS mobility fund for carrying out the research at Cerema, in Nantes, France.

References

[1] Y. Toparlar, B. Blocken, B. Maiheu, G.J.F. van Heijst, A review on the CFD analysis of urban microclimate, Renew. Sustain. Energy Rev. 80 (2017) 1613–1640. https://doi.org/10.1016/j.rser.2017.05.248.

[2] M. Santamouris, Energy and Climate in the Urban Built Environment, Routledge, London, 2013. https://doi.org/10.4324/9781315073774.

[3] T.R. Oke, Boundary layer Climates, Routledge, New York, 1987.

[4] L. Kleerekoper, M. van Esch, T.B. Salcedo, How to make a city climate-proof, addressing the urban heat island effect, Resour. Conserv. Recycl. 64 (2012) 30–38. https://doi.org/10.1016/j.resconrec.2011.06.004.

[5] S.W. Kim, R.D. Brown, Urban heat island (UHI) variations within a city boundary: A systematic literature review, Renew. Sustain. Energy Rev. 148 (2021) 111256. https://doi.org/10.1016/j.rser.2021.111256.

[6] A. Rodler, T. Leduc, Local climate zone approach on local and micro scales: Dividing the urban open space, Urban Clim. 28 (2019) 100457. https://doi.org/10.1016/j.uclim.2019.100457.

[7] V. Masson, W. Heldens, E. Bocher, M. Bonhomme, B. Bucher, C. Burmeister, C. de Munck, T. Esch, J. Hidalgo, F. Kanani-Sühring, Y.-T. Kwok, A. Lemonsu, J.-P. Lévy, B. Maronga, D.

Pavlik, G. Petit, L. See, R. Schoetter, N. Tornay, A. Votsis, J. Zeidler, City-descriptive input data for urban climate models: Model requirements, data sources and challenges, Urban Clim. 31 (2020) 100536. https://doi.org/10.1016/j.uclim.2019.100536.

[8] A. Boccalatte, M. Fossa, L. Gaillard, C. Menezo, Microclimate and urban morphology effects on building energy demand in different European cities, Energy Build. 224 (2020) 110129. https://doi.org/10.1016/j.enbuild.2020.110129.

[9] D. Xu, D. Zhou, Y. Wang, W. Xu, Y. Yang, Field measurement study on the impacts of urban spatial indicators on urban climate in a Chinese basin and static-wind city, Build. Environ. 147 (2019) 482–494. https://doi.org/10.1016/J.BUILDENV.2018.10.042.

[10] A. Kamal, S.M.H. Abidi, A. Mahfouz, S. Kadam, A. Rahman, I.G. Hassan, L.L. Wang, Impact of urban morphology on urban microclimate and building energy loads, Energy Build. 253 (2021) 111499. https://doi.org/10.1016/j.enbuild.2021.111499.

[11] K.K.L. Lau, Z. Tan, T.E. Morakinyo, C. Ren, Urban Greening Strategies for Enhancing Outdoor Thermal Comfort, SpringerBriefs Archit. Des. Technol. (2022) 85–100. https://doi.org/10.1007/978-981-16-5245-5_6.

[12] S. Shareef, B. Abu-Hijleh, The effect of building height diversity on outdoor microclimate conditions in hot climate. A case study of Dubai-UAE, Urban Clim. 32 (2020) 100611. https://doi.org/10.1016/j.uclim.2020.100611.

[13] Y. Gao, J. Zhao, L. Han, Exploring the spatial heterogeneity of urban heat island effect and its relationship to block morphology with the geographically weighted regression model, Sustain. Cities Soc. 76 (2022) 103431. https://doi.org/10.1016/j.scs.2021.103431.

[14] A. Bernabé, J. Bernard, M. Musy, H. Andrieu, E. Bocher, I. Calmet, P. Kéravec, J.M. Rosant, Radiative and heat storage properties of the urban fabric derived from analysis of surface forms, Urban Clim. 12 (2015) 205–218. https://doi.org/10.1016/J.UCLIM.2015.04.001.

[15] C. Yuan, L. Chen, Mitigating urban heat island effects in high-density cities based on sky view factor and urban morphological understanding: a study of Hong Kong, Http://Dx.Doi.Org/10.1080/00038628.2011.613644. 54 (2011) 305–315. https://doi.org/10.1080/00038628.2011.613644.

[16] O. Palusci, P. Monti, C. Cecere, H. Montazeri, B. Blocken, Impact of morphological parameters on urban ventilation in compact cities: The case of the Tuscolano-Don Bosco district in Rome, Sci. Total Environ. 807 (2022) 150490. https://doi.org/10.1016/J.SCITOTENV.2021.150490.

[17] C. Apreda, A. Reder, P. Mercogliano, Urban morphology parameterization for assessing the effects of housing blocks layouts on air temperature in the Euro-Mediterranean context, Energy Build. 223 (2020) 110171. https://doi.org/10.1016/j.enbuild.2020.110171.

[18] A. Meek, N. Jayasuriya, ; Edmund Horan, R. Adams, Environmental Benefits of Retrofitting Green Roofs to a City Block, J. Hydrol. Eng. 20 (2014) 05014020. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001048.

[19] R. Ma, X. Li, J. Chen, An elastic urban morpho-blocks (EUM) modeling method for urban building morphological analysis and feature clustering, Build. Environ. 192 (2021) 107646. https://doi.org/10.1016/j.buildenv.2021.107646.

[20] H.A.S. Othman, A.A. Alshboul, The role of urban morphology on outdoor thermal comfort: The case of Al-Sharq City – Az Zarqa, Urban Clim. 34 (2020) 100706. https://doi.org/10.1016/J.UCLIM.2020.100706.

[21] M.H. Kristensen, R.E. Hedegaard, S. Petersen, Hierarchical calibration of archetypes for urban building energy modeling, Energy Build. 175 (2018) 219–234. https://doi.org/10.1016/J.ENBUILD.2018.07.030.

[22] J. Sokol, C. Cerezo Davila, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, Energy Build. 134 (2017) 11–24. https://doi.org/10.1016/J.ENBUILD.2016.10.050.

[23] I.D. Stewart, T.R. Oke, Local Climate Zones for Urban Temperature Studies, Bull. Am. Meteorol. Soc. 93 (2012) 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1.

[24] R. Kotharkar, A. Bagade, Local Climate Zone classification for Indian cities: A case study of Nagpur, Urban Clim. 24 (2018) 369–392. https://doi.org/10.1016/J.UCLIM.2017.03.003.

[25] C. Zhao, J. Jensen, Q. Weng, N. Currit, R. Weaver, Application of airborne remote sensing data on mapping local climate zones: Cases of three metropolitan areas of Texas, U.S., Comput. Environ. Urban Syst. 74 (2019) 175–193. https://doi.org/10.1016/j.compenvurbsys.2018.11.002.

[26] J. Geletič, M. Lehnert, GIS-based delineation of local climate zones: The case of medium-sized Central European cities, Morav. Geogr. Reports. 24 (2016) 2–12. https://doi.org/10.1515/MGR-2016-0012.

[27] M. Taleghani, M. Tenpierik, A. Van Den Dobbelsteen, R. De Dear, Energy use impact of and thermal comfort in different urban block types in the Netherlands, Energy Build. 67 (2013) 166–175. https://doi.org/10.1016/J.ENBUILD.2013.08.024.

[28] C.M. Hsieh, H.C. Huang, Mitigating urban heat islands: A method to identify potential wind corridor for cooling and ventilation, Comput. Environ. Urban Syst. 57 (2016) 130–143. https://doi.org/10.1016/J.COMPENVURBSYS.2016.02.005.

[29] G. Meinel, R. Hecht, H. Herold, Analyzing building stock using topographic maps and GIS, Build. Res. Inf. 37 (2009) 468–482. https://doi.org/10.1080/09613210903159833.

[30] S. Vanderhaegen, F. Canters, Mapping urban form and function at city block level using spatial metrics, Landsc. Urban Plan. 167 (2017) 399–409. https://doi.org/10.1016/J.LANDURBPLAN.2017.05.023.

[31] H. López-Moreno, M. Núñez-Peiró, C. Sánchez-Guevara, J. Neila, On the identification of Homogeneous Urban Zones for the residential buildings' energy evaluation, Build. Environ. 207 (2022) 108451. https://doi.org/10.1016/j.buildenv.2021.108451.

[32] J. Gil, J. Beirão, N. Montenegro, J. Duarte, On the discovery of urban typologies: Data mining the many dimensions of urban form, Urban Morphol. 16 (2012) 27–40.

[33] P.M. Schirmer, K.W. Axhausen, A Multiscale Clustering of the Urban Morphology for Use in Quantitative Models, Model. Simul. Sci. Eng. Technol. (2019) 355–382. https://doi.org/10.1007/978-3-030-12381-9_16.

[34] P.M. Schirmer, K.W. Axhausen, A multiscale classification of urban morphology, J. Transp. Land Use. 9 (2016) 101–130. http://www.jstor.org/stable/26203210.

[35] G. Tardioli, R. Kerrigan, M. Oates, J. O'Donnell, D.P. Finn, Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach, Build. Environ. 140 (2018) 90–106. https://doi.org/10.1016/J.BUILDENV.2018.05.035.

[36] G.W. Milligan, M.C. Cooper, Methodology Review: Clustering Methods, Appl. Psychol. Meas. 11 (1987) 329–354. https://doi.org/10.1177/014662168701100401.

[37] M. Demuzere, B. Bechtel, A. Middel, G. Mills, European LCZ Map (Version2).figshare., 2019. https://doi.org/https://doi.org/10.6084/m9.figshare.13322450.v2.

[38] M. Demuzere, B. Bechtel, A. Middel, G. Mills, Mapping Europe into local climate zones, PLoS One. 14 (2019) e0214474. https://doi.org/10.1371/journal.pone.0214474.

[39] Statbel, Population by statistical sector , (2021). https://statbel.fgov.be/nl/open-data/bevolking-statistische-sector-9 (accessed April 27, 2022).

[40] M.Y. Joshi, W. Selmi, M. Binard, G.-A. Nys, J. Teller, Potential for urban greening with green roofs: A way towards smart cities, in: 2020. https://doi.org/10.5194/isprs-annals-VI-4-W2-2020-87-2020.

[41] D. Godoy-Shimizu, P. Steadman, S. Evans, Density and morphology: from the building scale to the city scale, Build. Cities. 2 (2021) 92–113. https://doi.org/10.5334/BC.83.

[42] I.D. Dwiputra, I.W.K. Mt, H. Winarso, Impact of Urban Block Typology on Microclimate Performance in a Hot-Humid High-Density City, IOP Conf. Ser. Earth Environ. Sci. 738 (2021) 012067. https://doi.org/10.1088/1755-1315/738/1/012067.

[43] X. Zhou, H. Chen, Experimental Analysis of the Influence of Urban Morphological Indices on the Urban Thermal Environment of Zhengzhou, China, Atmos. 2021, Vol. 12, Page 1058. 12 (2021) 1058. https://doi.org/10.3390/ATMOS12081058.

[44] J. Yang, Y. Yang, D. Sun, C. Jin, X. Xiao, Influence of urban morphological characteristics on thermal environment, Sustain. Cities Soc. 72 (2021) 103045. https://doi.org/10.1016/J.SCS.2021.103045.

[45] M. Berghauser Pont, P. Haupt, The Spacemate: Density and the typomorphology of the urban fabric, Nord. Arkit. 18 (2005) 55–68.

[46] J. Allegrini, J. Carmeliet, Simulations of local heat islands in Zürich with coupled CFD and building energy models, Urban Clim. 24 (2018) 340–359. https://doi.org/10.1016/J.UCLIM.2017.02.003.

[47] A. Abd Razak, A. Hagishima, N. Ikegaya, J. Tanimoto, Analysis of airflow over building arrays for assessment of urban wind environment, Build. Environ. 59 (2013) 56–65. https://doi.org/10.1016/j.buildenv.2012.08.007.

[48] M.P. Heris, A. Middel, B. Muller, Impacts of form and design policies on urban microclimate: Assessment of zoning and design guideline choices in urban redevelopment projects, Landsc. Urban Plan. 202 (2020) 103870. https://doi.org/10.1016/J.LANDURBPLAN.2020.103870.

[49] R. Louf, M. Barthelemy, A typology of street patterns, J. R. Soc. Interface. 11 (2014). https://doi.org/10.1098/RSIF.2014.0924.

[50] E. Ng, C. Yuan, L. Chen, C. Ren, J.C.H. Fung, Improving the wind environment in high-density cities by understanding urban morphology and surface roughness: A study in Hong Kong, Landsc. Urban Plan. 101 (2011) 59–74. https://doi.org/10.1016/J.LANDURBPLAN.2011.01.004.

[51] B. Wang, S. Geoffroy, M. Bonhomme, Urban form study for wind potential development:, Https://Doi.Org/10.1177/2399808321994449. 49 (2021) 76–91. https://doi.org/10.1177/2399808321994449.

[52] M.S. Wong, J.E. Nichol, P.H. To, J. Wang, A simple method for designation of urban ventilation corridors and its application to urban heat island analysis, Build. Environ. 45 (2010) 1880–1889. https://doi.org/10.1016/j.buildenv.2010.02.019.

[53] J. Bernard, E. Bocher, G. Petit, S. Palominos, Sky View Factor Calculation in Urban Context: Computational Performance and Accuracy Analysis of Two Open and Free GIS Tools, Clim. 2018, Vol. 6, Page 60. 6 (2018) 60. https://doi.org/10.3390/CLI6030060.

[54] M. Dirksen, R.J. Ronda, N.E. Theeuwes, G.A. Pagani, Sky view factor calculations and its application in urban heat island studies, Urban Clim. 30 (2019) 100498. https://doi.org/10.1016/J.UCLIM.2019.100498.

[55] J. Teller, S. Azar, Townscope II - A computer systems to support solar access decision-making, Sol. Energy. 70 (2001) 187–200. https://doi.org/10.1016/S0038-092X(00)00097-9.

[56] K. Zakšek, K. Oštir, Ž. Kokalj, Sky-View Factor as a Relief Visualization Technique, Remote Sens. 2011, Vol. 3, Pages 398-415. 3 (2011) 398–415. https://doi.org/10.3390/RS3020398.

[57] Ž. Kokalj, M. Somrak, Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping, Remote Sens. 2019, Vol. 11, Page 747. 11 (2019) 747. https://doi.org/10.3390/RS11070747.

[58] M. Bruwier, A. Mustafa, D.G. Aliaga, P. Archambeau, S. Erpicum, G. Nishida, X. Zhang, M. Pirotton, J. Teller, B. Dewals, Influence of urban pattern on inundation flow in floodplains of lowland rivers, Sci. Total Environ. 622–623 (2018) 446–458. https://doi.org/10.1016/j.scitotenv.2017.11.325.

[59] Y. Zha, J. Gao, S. Ni, Use of normalized difference built-up index in automatically mapping urban areas from TM imagery, Https://Doi.Org/10.1080/01431160304987. 24 (2010) 583–594. https://doi.org/10.1080/01431160304987.

[60] X.-L. Chen, H.-M. Zhao, P.-X. Li, Z.-Y. Yin, Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes, Remote Sens. Environ. 104 (2006) 133–146. https://doi.org/10.1016/j.rse.2005.11.016.

[61] P. Macarof, F. Statescu, Comparasion of NDBI and NDVI as indicators of surface urban heat island effect in landsat 8 imagery: a case study of Iasi, Present Environ. Sustain. Dev. 2 (2017) 141–150. https://doi.org/10.1515/pesd-2017-0032.

[62] L.S. Ferreira, D.H.S. Duarte, Exploring the relationship between urban form, land surface temperature and vegetation indices in a subtropical megacity, Urban Clim. 27 (2019) 105–123. https://doi.org/10.1016/j.uclim.2018.11.002.

[63] L. MONTANDON, E. SMALL, The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI, Remote Sens. Environ. 112 (2008) 1835–1845. https://doi.org/10.1016/j.rse.2007.09.007.

[64] A. Martilli, E.S. Krayenhoff, N. Nazarian, Is the Urban Heat Island intensity relevant for heat mitigation studies?, Urban Clim. 31 (2020) 100541. https://doi.org/10.1016/J.UCLIM.2019.100541.

[65] H. Li, Y. Liu, H. Zhang, B. Xue, W. Li, Urban morphology in China: Dataset development and spatial pattern characterization, Sustain. Cities Soc. 71 (2021) 102981. https://doi.org/10.1016/j.scs.2021.102981.

[66] E. Bocher, J. Bernard, E. Wiederhold, F. Leconte, G. Petit, S. Palominos, C. Noûs, GeoClimate: a Geospatial processing toolbox for environmental and climate studies, J. Open Source Softw. 6 (2021) 3541. https://doi.org/10.21105/joss.03541.

[67] R. Burghardt, Development of an ArcGIS extension to model urban climate factors, (2014). www.lp-kassel.de (accessed July 18, 2022).

[68] P. Berkhin, A Survey of Clustering Data Mining Techniques, Group. Multidimens. Data Recent Adv. Clust. (2006) 25–71. https://doi.org/10.1007/3-540-28349-8_2.

[69] S. Chen, N.H. Wong, M. Ignatius, W. Zhang, Y. He, Z. Yu, D.J.C. Hii, ATLAS: Software for analysing the relationship between urban microclimate and urban morphology in a tropical city, Build. Environ. 208 (2022) 108591. https://doi.org/10.1016/J.BUILDENV.2021.108591.

[70] C. Ding, X. He, K-means Clustering via Principal Component Analysis, in: Proc. Twenty-First Int. Conf. Mach. Learn., 2004: pp. 29–41.

[71] H.F. Kaiser, The varimax criterion for analytic rotation in factor analysis, Psychom. 1958 233.

23 (1958) 187–200. https://doi.org/10.1007/BF02289233.

[72] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. 31 (2010) 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.

[73] D. Arthur, S. Vassilvitskii, k-means++: The Advantages of Careful Seeding, in: Proc. Eighteenth Annu. ACM-SIAM Symp. Discret. Algorithms. Soc. Ind. Appl. Math., Philadelphia PA, USA, 2007: pp. 1027–1035.

[74] X. Wang, Z. Zou, H. Zou, Using discriminant analysis to assess polycyclic aromatic hydrocarbons contamination in Yongding New River, Environ. Monit. Assess. 185 (2013) 8547–8555. https://doi.org/10.1007/S10661-013-3194-3/FIGURES/2.

[75] J. Rhee, J. Im, G.J. Carbone, J.R. Jensen, Delineation of climate regions using in-situ and remotely-sensed data for the Carolinas, Remote Sens. Environ. 112 (2008) 3099–3111. https://doi.org/10.1016/j.rse.2008.03.001.

[76] W.S. Sarle, Cubic clustering criterion, SAS Institute, 1983.

[77] G.W. Milligan, M.C. Cooper, An examination of procedures for determining the number of clusters in a data set, Psychom. 1985 502. 50 (1985) 159–179. https://doi.org/10.1007/BF02294245.

[78] T. Caliñski, J. Harabasz, A dendrite method for Cluster analysis, Commun. Stat. 3 (1974) 1–27. https://doi.org/10.1080/03610927408827101.

[79] L.S. Ferreira, D.H.S. Duarte, Exploring the relationship between urban form, land surface temperature and vegetation indices in a subtropical megacity, Urban Clim. 27 (2019) 105–123. https://doi.org/10.1016/j.uclim.2018.11.002.

[80] D. Maroni, G.T. Cardoso, A. Neckel, L.S. Maculan, M.L.S. Oliveira, E.T. Bodah, B.W. Bodah, M. Santosh, Land surface temperature and vegetation index as a proxy to microclimate, J. Environ. Chem. Eng. 9 (2021) 105796. https://doi.org/10.1016/J.JECE.2021.105796.

[81] K. jae Kim, H. Ahn, A recommender system using GA K-means clustering in an online shopping market, Expert Syst. Appl. 34 (2008) 1200–1209. https://doi.org/10.1016/J.ESWA.2006.12.025.

[82] R. Borge, J. Lumbreras, S. Vardoulakis, P. Kassomenos, E. Rodríguez, Analysis of long-range transport influences on urban PM10 using two-stage atmospheric trajectory clusters, Atmos. Environ. 41 (2007) 4434–4450. https://doi.org/10.1016/J.ATMOSENV.2007.01.053.

[83] L.A. Kort-Butler, K.A. Tyler, A cluster analysis of service utilization and incarceration among homeless youth, Soc. Sci. Res. 41 (2012) 612–623. https://doi.org/10.1016/J.SSRESEARCH.2011.12.011.

[84] USGS, Landsat 8 (L8) Data Users Handbook, USGS, Sioux Falls, South Dakota, 2019. https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1574_L8_Data_Users_Handbook-v5.0.pdf (accessed April 28, 2020).

[85] M. Lehnert, S. Savić, D. Milošević, J. Dunjić, J. Geletič, Mapping Local Climate Zones and Their Applications in European Urban Environments: A Systematic Literature Review and Future Development Trends, ISPRS Int. J. Geo-Information. 10 (2021) 260. https://doi.org/10.3390/ijgi10040260.

[86] E. Bocher, G. Petit, J. Bernard, S. Palominos, A geoprocessing framework to compute urban indicators: The MApUCE tools chain, Urban Clim. 24 (2018) 153–174. https://doi.org/10.1016/j.uclim.2018.01.008.

[87] N. Mohajeri, G. Upadhyay, A. Gudmundsson, D. Assouline, J. Kämpf, J.L. Scartezzini, Effects of urban compactness on solar energy potential, Renew. Energy. 93 (2016) 469–482. https://doi.org/10.1016/j.renene.2016.02.053.

Highlights

- We propose systematic PCA-based k-means clustering approach to find urban archetypes
- Validation - ANOVA with land surface temperature in absence of existing typologies
- Our clusters are compared with WUDAPT's local climate zones (LCZs)
- Our approach provides essential sub-classes to the existing LCZs
- We identify 9 urban morphological archetypes defining the morphology of Liege city

Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: