

Article

Inference of Local Climate Zones from GIS Data, and Comparison to WUDAPT Classification and Custom-Fit Clusters

Fadel Muhammad ¹, Changkun Xie ^{1,2}, Julian Vogel ^{1,*} and Afshin Afshari ^{1,*}

¹ Fraunhofer Institute for Building Physics, Fraunhoferstraße 10, 83626 Valley, Germany; fadel.muhammad@ibp.fraunhofer.de (F.M.); xiechangkun@sjtu.edu.cn (C.X.)

² School of Design, Shanghai Jiao Tong University, Shanghai 200240, China

* Correspondence: julian.vogel@ibp.fraunhofer.de (J.V.); afshin.afshari@ibp.fraunhofer.de (A.A.)

Abstract: A GIS-based approach is used in this study to obtain a better LCZ map of Berlin in comparison to the remote-sensing-based WUDAPT L0 approach. The LCZ classification of land use/cover can be used, among other applications, to characterize the urban heat island. An improved fuzzy logic method is employed for the purpose of classification of the zone properties to yield the GIS-LCZ map over 100 m × 100 m grid tiles covering the Berlin region. The zone properties are calculated from raster and vector datasets with the aids of the urban multi-scale environmental predictor (UMEP), QGIS and Python scripts. The standard framework is modified by reducing the threshold for the zone property impervious fraction for LCZ E to better detect paved surfaces in urban areas. Another modification is the reduction in the window size in the majority filter during post-processing, compared to the WUDAPT L0 method, to retain more details in the GIS-LCZ map. Moreover, new training areas are generated considering building height information. The result of the GIS-LCZ approach is compared to the new training areas for accuracy assessment, which shows better overall accuracy compared to that of the WUDAPT L0 method. The new training areas are also submitted to the LCZ generator and the resulting LCZ-map gives a better overall accuracy value compared to the previous (WUDAPT) submission. This study shows one shortcoming of the WUDAPT L0 method: it does not explicitly use building height information and that leads to misclassification of LCZs in several cases. The GIS-LCZ method addresses this shortcoming effectively. Finally, an unsupervised machine learning method, k-means clustering, is applied to cluster the grid tiles according to their zone properties into custom classes. The custom clusters are compared to the GIS-LCZ classes and the results indicate that k-means clustering can identify more complex city-specific classes or LCZ transition types, while the GIS-LCZ method always divides regions into the standard LCZ classes.

Keywords: urban heat island (UHI); local climate zones (LCZ); WUDAPT; LCZ generator; k-means clustering



Citation: Muhammad, F.; Xie, C.; Vogel, J.; Afshari, A. Inference of Local Climate Zones from GIS Data, and Comparison to WUDAPT Classification and Custom-Fit Clusters. *Land* **2022**, *11*, 747. <https://doi.org/10.3390/land11050747>

Academic Editor: Guoyu Ren

Received: 14 April 2022

Accepted: 11 May 2022

Published: 18 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The urban heat island (UHI) effect has been defined as a phenomenon whereby the near-surface urban canopy air temperature is, on average, higher than that of its surrounding countryside [1]. Its intensity is likely to keep increasing in the future due to population and urbanization growth [2]. The UHI intensity characterizes urban climates and is related to a negative impact on the environment by increasing energy demand due to an increase in air conditioning, elevating emission of greenhouse gases and air pollutants, endangering human health and comfort and impairing water quality [3]; therefore, the prediction of UHI is becoming significantly important.

Mesoscale numerical climate simulation software such as the Weather Research and Forecasting (WRF) [4] as well as simplified urban canopy models [5] is used to estimate the UHI intensity; however, they require highly resolved and accurate land use/land cover

(LULC) maps. For this purpose, LULC data sources such as CORINE by the Copernicus Land Monitoring Service have been used, where urban classes are derived from spectral features of the surface, neglecting the importance of the urban geometry on the microclimate [6]. More recently, local climate zone (LCZ) classification maps have been proposed for LULC characterization considering the urban geometry property. The LCZ method was developed by Stewart and Oke [7] for urban climate study purposes by classifying urban and rural sites into 10 built-up classes and 7 land cover classes. According to the standard framework, these classes are defined by 10 zone properties such as sky view factor, mean building height, and surface fractions. The LCZ scheme has been used extensively for characterizing UHI.

As described by Lehnert et al. [8], there are four approaches that have been applied to generate LCZ classification maps: expert knowledge, remote sensing imagery (RS), geographic information system (GIS), and combined approaches. In this paper, the RS-LCZ and GIS-LCZ are applied to derive the LCZ of Berlin. The RS-LCZ is the approach adopted by the World Urban Database and Access Portal Tools (WUDAPT) project [9], which produces its Level 0 (L0) data comprising the LCZ map derived from machine learning supervised classification of Landsat or Sentinel images [10–12]. The classification is carried out by using an open source software, the System for Automated Geoscientific Analyses (SAGA) [13] by applying training areas that are digitized in Google Earth by volunteer local experts. The classification step can be performed several times to improve the classification result. The result is checked with Google imagery, and existing training areas (TAs) can be modified or additional TAs can be digitized to repeat the classification process. If the classification result is too heterogeneous, it can be filtered using a majority filter. One major downside of the WUDAPT L0 method in deriving LCZs is that its accuracy highly depends on the number and quality of TAs, which are generated from satellite imagery relying on a 2D perspective neglecting the building height information [14].

The WUDAPT L0 method has been further simplified with the introduction of a new platform named the LCZ Generator [15]. This platform is an online-based processing tool [16] that integrates the state-of-the-art of LCZ mapping, and simultaneously provides an automated accuracy assessment, training data derivatives, and a novel approach to identify suspicious training areas. The LCZ Generator web application consists of three major steps: In the first step, personal and training information needs to be submitted via the web application by the user. In the second step, upon successful submission, the LCZ classification and quality control is launched in the back-end to produce a quality-controlled LCZ map, metadata statistics, and labels for suspicious polygons. In a third and final step, compressed results are sent to the user via e-mail, and simultaneously added to the online submission table. These steps require the users to only submit their TAs to the website to obtain an automated LCZs result after less than half an hour.

Another method is the so called GIS-based method, in which vector or raster datasets of buildings, roads, vegetation, and LULC are used to estimate the zone properties to define LCZ classes. The GIS-LCZ method is claimed to primarily have a better accuracy for deriving LCZ compared to the WUDAPT L0 method since it is particularly considering the building height information [14]. As summarized by Quan and Bansal [17], the first GIS-based approach employed in deriving LCZ was described by Lelovics et al. [18] and Unger et al. [19]. Since then, several representative studies implementing the GIS-based LCZ mapping have been published.

Lelovics et al. [18] and Unger et al. [19] focused on the city of Szeged in Hungary and applied the lot area polygon as basic spatial units (BSU). They used seven zone properties: sky view factor, building height, surface roughness, surface albedo, and three surface fractions (building fraction, impervious fraction, and pervious fraction). These zone properties were classified using fuzzy logic. The classification result was further processed by merging the BSUs. The resulting map has six built-up classes of GIS-LCZ. The map was used to plan the location of an urban temperature measurement network. Other related studies used this map for the integration with WUDAPT L0 map [20] and for analyzing temperature conditions in an urban meteorological network that is designed

based on the LCZ scheme [21]. A GIS-LCZ map was also generated for Vienna, Austria by Hammeberg et al. [22]. This work applies $100\text{ m} \times 100\text{ m}$ grid tiles as the BSU and employs five zone properties: building height, aspect ratio, and the usual three surface fractions. The classification of the zone properties utilizes a naive Bayes classifier. There was no post-processing performed in this study. The evaluation for the classification result was carried out by comparing it to the WUDAPT L0 map.

Wang et al. [14] derived the GIS-LCZ map for Hong Kong applying a $100\text{ m} \times 100\text{ m}$ grid tile as the BSU. They used three zone properties and one additional land use data. The zone property's building height, building surface fraction, and sky view factor were employed to classify the built-up classes (LCZs 1–10). Additional land use data are used for the classification of the land cover classes (LCZs A–G). The classification was achieved by modifying the standard rules proposed by Stewart and Oke [7]. An accuracy assessment was carried out for the resulting GIS-LCZ map by comparing it to the established validation samples.

Another GIS-LCZ method was conducted by Estacio et al. [23] for Quezon City, Philippines. The study employed seven zone properties: sky view factor, building height, roughness length, surface albedo, and the three usual surface fractions. These zone properties were calculated over $100\text{ m} \times 100\text{ m}$ grid tiles, which were further classified by applying a fuzzy logic algorithm modified from Lelovics et al. [18]. The result of the classification was aggregated by using cellular automata to derive the LCZ map. The map was validated using expert knowledge. The land surface temperature profile for each LCZ type was also assessed in this research.

A clustering method was also applied in the work of Hidalgo et al. [24] to classify three cities in France (Toulouse, Paris and Nantes) based on the GIS datasets. They used eight parameters to categorize the built up classes (LCZ 1–10), and from those parameters, there was only one parameter that corresponds to the LCZ framework, which is mean building height. The other parameters are building density, building typology, majority building typology within the polygon, mean minimum distance, median minimum distance, main morphological group, and number of buildings. Morphological groups are used to identify LCZs 7 and 8. For the other built up classes, k-means clustering is applied for the classification. LCZs A–G were classified with another set of parameters: urban vegetation data from satellite images (LCZs A and B), road fraction indicator (LCZ E), and water fraction indicator (LCZ G).

In this paper, the GIS-LCZ method was applied with several novelties. It was used to define the LCZs in the city of Berlin, which, to the best of the authors' knowledge, has not been classified with the GIS-based method before. The resulting GIS-LCZ map is expected to improve the existing remote sensing-based LCZ map (WUDAPT L0). The classification of the zone properties employs an improved fuzzy logic algorithm and modification of the standard LCZ framework. The post-processing applied to the result of the GIS-LCZ classification is a majority filter as applied by the WUDAPT L0 approach, but in this paper, the window size in the majority filter is reduced to retain more detail from the GIS-LCZ data. New TAs of LCZ classes for Berlin are also generated considering the building height information, which is not considered by the TAs generated for WUDAPT L0. The resulting GIS-LCZ map and the WUDAPT L0 map are compared to the TAs for assessing the accuracy. Moreover, the TAs are also submitted to the LCZ generator website to derive an improved RS-LCZ. Furthermore, the k-means clustering method is used to cluster the grid tiles according to their zone properties into custom classes. Finally, the custom clusters are compared to the GIS-LCZ classes to investigate their similarities and differences. Section 2 in this paper explains the study area and methodology. Section 3 describes the classification results and discussions. Section 4 contains conclusions and research outlook.

2. Methodology

In this study, a GIS-based method was employed to derive the LCZs for Berlin. The general GIS-LCZ mapping process is illustrated in Figure 1. The process was initialized by collecting vector and raster data that are later used to calculate the zone properties. The

zone properties are calculated based on a basic spatial unit (BSU), which is defined as a grid tile. Each BSU is classified applying a fuzzy logic algorithm to derive the LCZs. During post-processing, a filtering is carried out to the classification result of LCZs. An evaluation was performed on the final GIS-LCZ by comparing it to the new training areas. The new training areas were further used to evaluate the existing WUDAPT L0 map and to generate an LCZ map from the LCZ generator. Furthermore, the zone properties of the grid tiles were clustered, and then compared to the GIS-LCZ result.

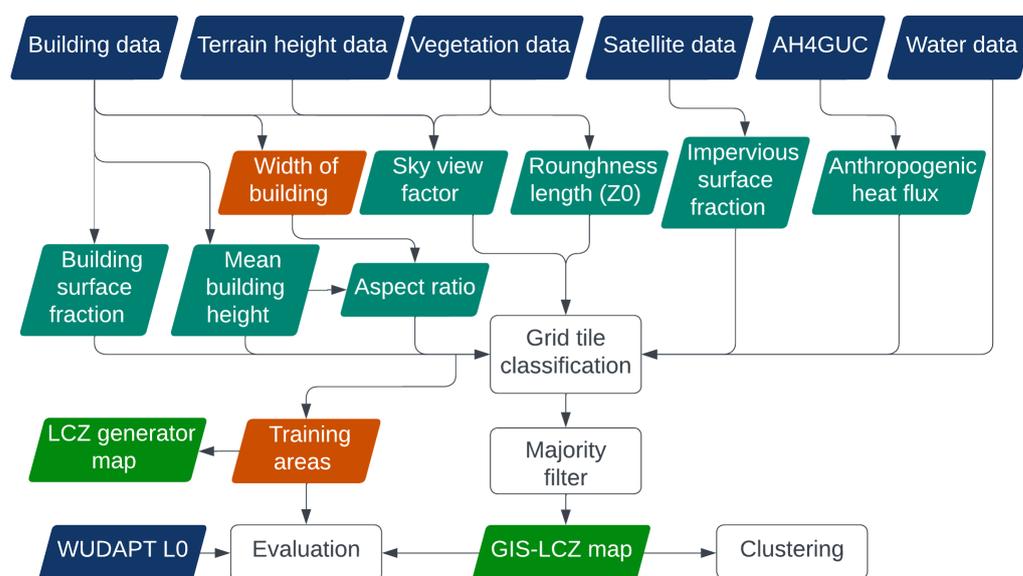


Figure 1. GIS-LCZ mapping method of Berlin.

2.1. Study Area and Datasets

The focus of this study is the city of Berlin. Berlin has an area of 892 km². As of 2019, its population was around 3.8 million, which makes it the most populous city in Germany. One of the reasons for choosing this city is the availability of the GIS data, which we require for deriving the GIS-LCZ map. A Google Earth image of Berlin and the newly created training areas (see Section 2.5) shown in Figure 2.

The datasets used in this study are from Deutsches Zentrum für Luft- und Raumfahrt (DLR) [25], OpenStreetMap (OSM) [26], WUDAPT L0, Copernicus [27], and Anthropogenic Heat for Global Urban Climatology (AH4GUC) [28].

The DLR dataset is obtained from the work of Heldens et al. [25]. They generated raster data of Berlin for a microclimate simulation. The raster dataset includes rasters of building height, terrain height, vegetation height, water, streets, and bridges. The dataset provides several resolution ranges from 1 m to 16 m. On the other hand, OSM is a vector dataset containing primarily building land use, road, and water features. OSM is an open source data generated by a community of mappers [29].

Satellite imagery from Copernicus is also used in this study. Copernicus is the European Earth monitoring system where data are acquired from different sources, such as in situ sensors and Earth observation satellites. Raster data of land cover and high resolution layers, such as imperviousness density (IMD) are provided by Copernicus [27].

As mentioned previously, WUDAPT data are used for evaluation purposes in this paper. WUDAPT is a project initiated by urban climate research to provide universally coherent and consistent information on form and function of urban morphology for climate studies [9]. WUDAPT information consists of three levels of detail (L) and is gathered using distinct methodologies. Level 0 (L0) data comprises a local climate zone map, which is based on the work of Stewart and Oke [7]. On the other hand, Level 1 (L1) data provide a better representation for each LCZ through sampling by providing information regarding urban form and function in a finer spatial resolution. L1 data have a representation in

three-dimensional form. Furthermore, Level 2 (L2) data provide more information by giving detailed descriptions of urban parameter values for boundary layer modeling.

The WUDAPT project provides level 0 (L0) maps for many cities around the world. The WUDAPT L0 map for Berlin was downloaded from the WUDAPT portal [30]. The map was produced in 2016 and was derived from Landsat 8 Images from March and April 2015. The resolution of the map is 100 m and it was resampled from 30 m Landsat 8 input image resolution. The training areas can also be downloaded through the WUDAPT website [31].

The thermal property used for the classification of GIS-LCZ of Berlin is anthropogenic heat flux that is obtained from AH4GUC. This dataset provides global present and future hourly data of anthropogenic heat flux (AHF) with a resolution of 1 km, which is derived from a global population density map, global gridded monthly air temperature, a global nighttime lights map, and a global combustion sources map [28].

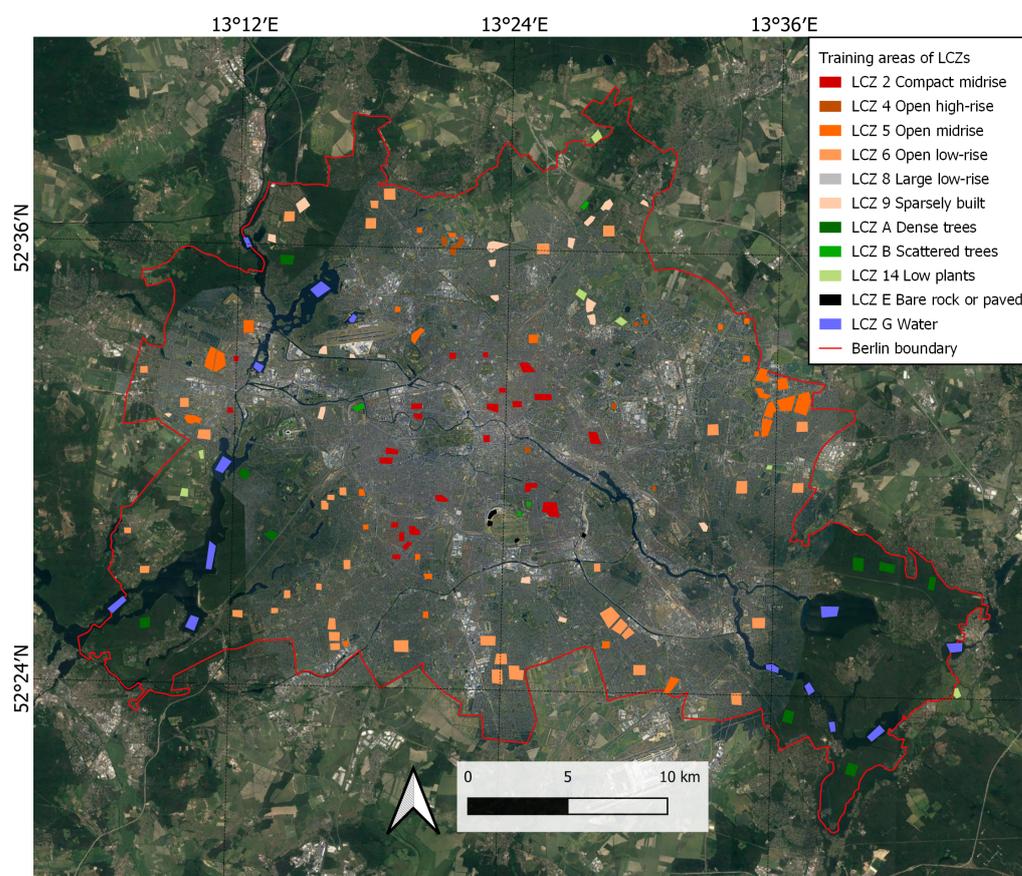


Figure 2. Local climate zones: Berlin’s training areas, see Section 2.5 (Google Earth image in QGIS).

2.2. Inference of Urban Morphology from GIS Data

According to the standard LCZ methodology, there are 10 zone properties defining the 17 local climate zones. These properties are: sky view factor, aspect ratio, building surface fraction, impervious surface fraction, pervious surface fraction, height of roughness elements (building heights), terrain roughness class, surface admittance, surface albedo, and anthropogenic heat flux. However, due to limited data sources, in practice, only a subset of those properties can be used for the classification of the LCZs.

For this study, seven zone properties are calculated to generate the LCZ map: sky view factor (SVF), mean building height (H) or mean vegetation height (H_v), aspect ratio (H/W), building surface fraction (BSF), impervious surface fraction (ISF), anthropogenic heat flux (AHF), and roughness length (z_0). The basic spatial unit for the classification of the zone properties is in the form of grid tiles with the size of $100\text{ m} \times 100\text{ m}$. Every zone property will be resampled to this resolution. The polygon of the grid tiles is created in QGIS in the

shapefile format with the extent of the area of Berlin with the coordinate reference system (CRS) of European Petroleum Survey Group Geodesy (EPSG) 25833 (ETRS89/UTM zone 33N). This CRS is used as the default CRS for all the calculation of the zone properties. The total number of grid tiles for the area of Berlin is 90517.

The sky view factor (SVF) is calculated by applying raster height data of building, terrain, and vegetation patch derived from the DLR dataset with a resolution of 5 m. The calculation is performed by employing the Urban Multi-scale Environmental Predictor (UMEP) plugin in QGIS by Lindberg et al. [32]. The DLR rasters of building and vegetation patch height with a resolution of 1 m are also used for the calculation of mean building height (H) for urban classes and mean vegetation height (H_v) for natural classes, respectively. The building surface fraction (BSF) is calculated to define the percentage of building area in a grid tile. Because not all building areas in Berlin are covered by the DLR data, additional building polygon data from OSM are used to define the BSF.

The impervious surface fraction (ISF) is the percentage of the area covered by impervious (paved or rock) materials in a grid tile. The Copernicus Impervious Density (IMD) raster is used for the calculation of the ISF by calculating the mean of the 20 m resolution raster data over the grid tile; however, the IMD cannot be directly used to represent the ISF needed by the LCZ framework since the IMD also includes building information. The ISF as a zone property in the LCZ classification excludes the information of buildings since that information is already covered by BSF. Thus, the BSF should be subtracted from the IMD in order to obtain the ISF, which can be formulated as: $ISF = IMD - BSF$; however, when this formula is implemented, it results in negative ISF values in several grid tiles. This can be due to the fact that the IMD raster is not fully harmonized with the BSF value since they are from different data sources and have different resolutions and acquisition methods. To avoid negative values in the ISF, the IMD is corrected by taking the maximum between the original IMD and the BSF which can be formulated as $IMD = \max(IMD, BSF)$.

The aspect ratio (H/W) is the ratio between mean building height (H) and building spacing (or street width). The width of building spacing (W) is estimated by the equation modified from Samsonov and Varentsov [33]:

$$W = \sqrt{S_0} \cdot (1 - \sqrt{BSF}) / N_{BLD} \quad (1)$$

S_0 is the grid tile area, which, in our case, is 10,000 m². N_{BLD} is the number of buildings that is obtained from building data of OSM and DLR. The resulting H/W calculation contains outliers where the grid tiles have a mean building height H but either they have no building width W or the value of it is small (less than 1 m). This results in incorrect values of H/W . These outliers occur, for example, in the grid tiles covered mostly or fully by buildings. To solve this issue, another H/W is calculated from the grid tiles of 250 m × 250 m to obtain a broader perspective of H/W . The new H/W is resampled to 100 m × 100 m grid tiles and assigned to the grid tiles that have incorrect H/W values or having a mean width W of less than 1 m.

The anthropogenic heat flux (AHF) is obtained from the global AHF dataset with the resolution of 1 km. The raster of AHF is resampled to 100 m × 100 m to be further used to calculate the mean AHF in a grid tile. The roughness length (z_0) is calculated from the formula suggested by Oke (cited from Estacio et al. [23]):

$$z_0 = f_0 \cdot \bar{z}_H \quad (2)$$

f_0 is 0.1, which is a constant value generally used for surfaces and \bar{z}_H is the average difference of DSM and DTM, which is calculated as the addition of two rasters: building and vegetation patch heights with their 1 m resolution. Water raster data from DLR with a resolution of 1 m are also used in the classification of GIS-LCZ. The raster is applied directly to classify grid tiles that contain mainly water into LCZ G. The calculation is performed in QGIS by applying the Zonal Statistic tool. Table 1 summarizes the data sources and the calculation method to derive the zone properties and LCZ G. The calculation result of the

zone properties, which are SVF, H , H_v , H/W , BSF, ISF, AHF, and z_0 , are shown in Figure 3. The zone properties are visualized in QGIS with the equal count (quantile) of 5 categories.

Table 1. Data sources and calculation method of zone properties and LCZ G.

Indicators	Data	Calculation	Outlier Treatment
SVF	DLR rasters of building and vegetation heights 5 m	UMEP in QGIS	
H	DLR raster of building height 1 m	QGIS	
H_v	DLR raster of vegetation height 1 m	QGIS	
BSF	DLR raster of building height 1 m and rasterized OSM building data 1 m	QGIS	
H/W	H , W (obtained from BSF and N_{BLD}), N_{BLD} is from DLR raster of building height 1 m and rasterized OSM building data 1 m	QGIS, Python script	H/W calculated from greater grid tile of 250 m
z_0	DLR rasters of building and vegetation heights 1 m	QGIS	
AHF	Raster of AH4GUC 1 km	QGIS	
LCZ G	DLR raster of water coverage 1 m	QGIS	

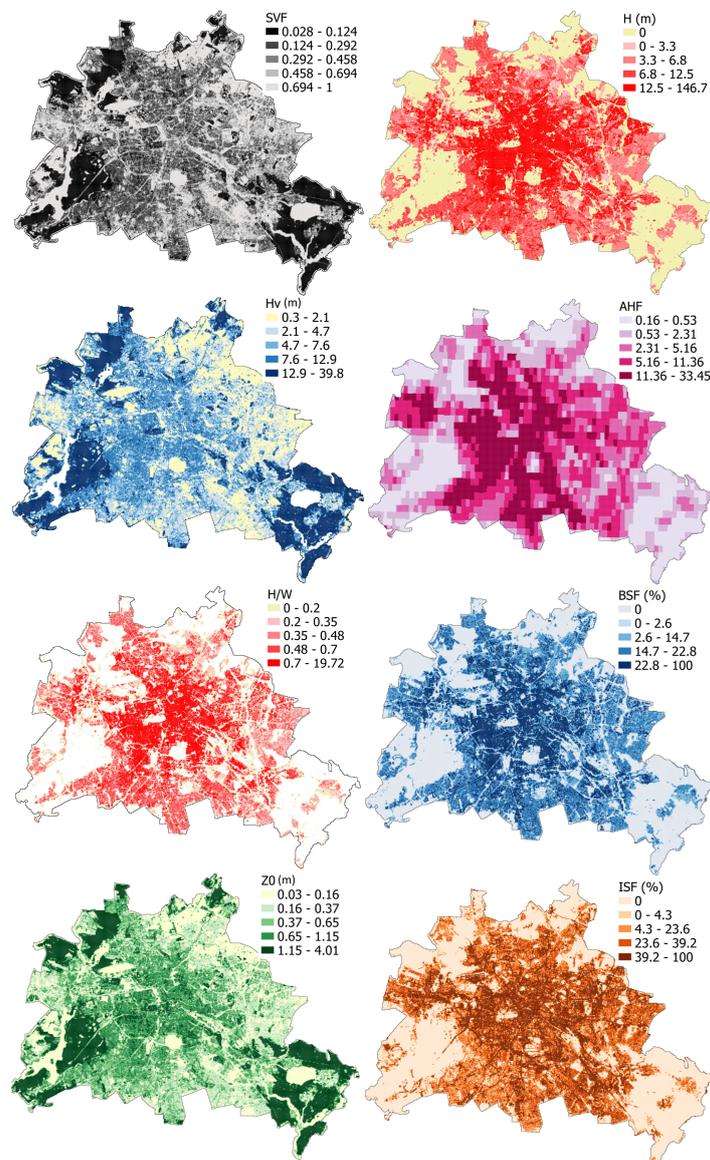


Figure 3. Zone properties of GIS-LCZ for Berlin.

2.3. Classification to Local Climate Zones

The zone properties used in the classification of the GIS-LCZ method in this study are simplified into 12 classes instead of 17 classes. LCZ 1 (compact high-rise), LCZ 7 (lightweight low-rise), LCZ C (bush and scrub), and LCZ F (soil/sand) are excluded due to the quasi-nonexistence of these LCZ classes in Berlin. For the classification, the grid tiles are divided into nine urban classes (LCZ 2, 3, 4, 5, 6, 8, 9, E) and three natural classes (LCZ A, B, D). The natural classes are categorized as grid tiles having $BSF \leq 10$ and $ISF \leq 10$ or containing water. The rest of the grid tiles, which are not natural classes, are classified as urban classes. Furthermore, the grid tiles are classified based on their zone properties into LCZ classes based on the range of values adapted from the standard LCZ framework of Stewart and Oke [7], which is shown in Table 2. In this study, LCZ E is considered as an urban class; therefore, its aspect ratio H/W is modified from the standard framework, where the W indicates building spacing instead of tree spacing. Its ISF range value is also modified from $ISF \geq 90$ to $ISF \geq 60$, so that it can better detect bare rock or paved surfaces.

Table 2. Zone properties of LCZ classes (adapted from Stewart and Oke [7]).

LCZ	SVF	H/W	BSF (%)	ISF (%)	H or H_v	AHF	z_0
2	-	0.75–2	40–70	30–50	10–25	≤ 75	-
3	-	0.75–1.5	40–70	20–50	3–10	≤ 75	-
4	-	0.75–1.25	20–40	30–40	≥ 25	≤ 50	-
5	-	0.3–0.75	20–40	30–50	10–25	≤ 25	-
6	-	0.3–0.75	20–40	20–50	3–10	≤ 25	-
8	-	0.1–0.3	30–50	40–50	3–10	≤ 50	-
9	-	0.1–0.25	10–20	≤ 20	3–10	≤ 10	-
10	-	0.2–0.5	20–30	20–40	5–15	≥ 300	-
A	≤ 0.4	-	≤ 10	≤ 10	3–30	-	≥ 2
B	0.5–0.8	-	≤ 10	≤ 10	3–15	-	0.25–0.5
D	≤ 0.9	-	≤ 10	≤ 10	≤ 1	-	0.03–0.1
E	-	≤ 0.1	≤ 10	≥ 60	≤ 2	-	-

The zone properties are classified into LCZ classes by applying fuzzy logic with a trapezoidal membership function modified from Estacio et al. [23]. The membership of every zone property for every LCZ class is determined as shown in Figure 4 (example case of LCZ 2 and its ISF property). The property value which is in the specified range will have a membership value of 1 and the membership value will linearly decrease from 1 to 0 when the property value is out of the range.

To understand how this membership function works, an example from Figure 4 is explained here. The zone property of the ISF of LCZ 2 has a range value from 30–50, which implies 30 as the left bound (LB), 50 as the right bound (RB) and length $L = RB - LB = 20$. When a grid tile has an ISF value, which is in this range (30–50), the membership value will be 1. On the other hand, when a grid tile has an ISF which is not in this range, the membership value will depend on how far it is away from the LB or RB. The membership value will become 0 for ISF values of less than 10 and more than 70. These values will be called as left zero bound (LZB) and right zero bound (RZB), respectively. The value of the RZB and LZB are defined as $LZB = LB - L = 10$ and $RZB = RB + L = 70$.

A problem arose for the zone properties that do not imply the RB, for example, the property value of H for LCZ 4, which is bigger than 25 m. The LB is 2 and the RB goes to infinity. It will not be a problem for defining the RZB because it can be set to infinity as well; however, it will be a problem for defining the LZB. The LZB cannot go to negative infinity, as it will overestimate the membership value of the zone property, which is less than LB. For this case, Estacio et al. [23] chose a value of 0 as the LZB; however, when choosing 0 as the LZB, it will still overestimate the membership value of H/W that is less than 25, because the H will obtain membership values from all the urban classes of LCZs which ranges from 0 (LZB) to 25 (LB). This is not desirable since the purpose of the classification is

to obtain a relatively distinct classification outcome. To tackle this issue, the LZB is chosen from the second highest RB value of the range values of H defined in Table 2, which is 15 m. The other zone properties, which do not define LB or RB, are SVE, BSF, and ISF. The common value is chosen as the LB or RB. For SVE, the LB would be 0 and the RB would be 1. For BSF and ISF, the LB would be 0 and the RB would be 100.

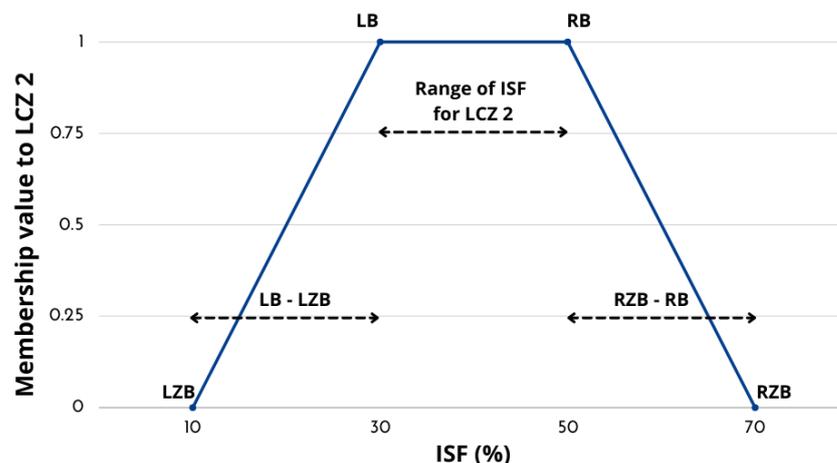


Figure 4. Trapezoidal membership function of fuzzy logic for ISF of LCZ 2 (adapted from Estacio et al. [23]).

The membership value for every property in every class of LCZ is calculated, and then the membership values of the zone properties for every class are summed up so that, at the end, for a grid tile of urban class, there are nine total membership values from the nine urban LCZ classes. On the other hand, for a grid tile of natural class, there are three total membership values from the three natural LCZ classes. These membership values are normalized by the number of zone properties available for the respective grid tile. From these normalized membership values, for a grid tile, maximum value is chosen and assigned as the LCZ class of the grid tile. The Python libraries used for the classification are NumPy, SciKit-Fuzzy, Pandas, and GeoPandas.

2.4. Filtering

The classification result is further processed by applying a post-processing step. As summarized by Quan and Bansal [17], previous GIS-LCZ studies carry out post-processing steps for two main reasons: diminishing unnecessary heterogeneity of the LCZ classes and maintaining their minimum area requirement.

In this study, during post-processing, a filtering was applied to obtain more homogeneous LCZ areas. The filter is a majority filter and it is applied using SAGA with a filter radius of one pixel and the search mode of square. These parameters yield a window filter of 3×3 pixels. The majority filter is also used by the WUDAPT L0 approach but with a greater size of the window filter, which is 5×5 pixels. We reduced the window size to retain more details of the GIS-LCZ classification.

2.5. New Training Areas

In generating WUDAPT L0 map, training areas (TAs) are digitized by the local experts who know the overview of the urban morphology of the city. Based on its guidelines [34], the WUDAPT L0 approach specifies TAs of LCZ classes in the respective city based on the view on Google Earth imagery. This implies that the approach relies only on the 2D perspective, which is inconsistent with the guidelines of the original framework of LCZ. Stewart and Oke [7] indicate that the geometric properties of the LCZ (mean building height, aspect ratio, and sky view factor) need building height information.

Therefore, in this study, new TAs are generated to improve the TAs of WUDAPT by considering the geometric properties calculated from the GIS-LCZ mapping method. The

new TAs are digitized in QGIS by the aid of satellite imagery and, at the same time, by considering the calculated zone properties. Figure 2 shows the new TAs, which are later used for the evaluation of the resulting GIS-LCZ map and the existing WUDAPT L0 map, as well as to derive a new LCZ map from the LCZ generator.

2.6. Custom Classification Using *k*-Means-Clustering

The classification of LCZ based on the standard framework is very general, as it was developed to fit a majority of cities around the world; however, it does not fit any specific city perfectly well and it may be that some classes of the standard framework do not exist in a city, or the city has specific classes not found in the standard framework. As summarized by Quan and Bansal [17], some studies modified the original LCZ classes by removing, mixing, and adding the standard LCZ classes. In the reviewed studies, not all the standard LCZ classes are found and the classified zones do not have zone properties that fit to the range values defined by the standard.

In this study, a custom classification to derive LCZs is introduced, where range values of the zone properties do not have to be defined. Instead, the grid tiles are clustered according to their zone properties into a number of custom classes, for which the average zone properties can be calculated to define the classes. This will give a result of distinct clusters, specific to the city, which contain grid tiles of similar zone properties. From this result, a custom land use/land cover map can be derived, and, together with the average zone properties, the urban morphology of a specific city may be described more accurately. This may also improve the accuracy of mesoscale climate simulation models that need highly resolved and accurate LULC maps.

For the clustering purpose, the *k*-means clustering method is applied. *k*-means clustering is one of the most popular methods in cluster analysis. It uses the vector quantization method to partition *N* observations into *k* clusters, in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid). *k*-means clustering minimizes within-cluster variances (squared Euclidean distances) and optimizes squared errors. This method has been widely used in the classification of land use/land cover [35].

We use the *k*-means clustering method to find clusters of grid tiles according to their zone properties. The clustering is only performed for the urban classes. In cases where *H/W* is null because of no building height information, the zone property is set to 0. The five zone properties of urban classes are normalized and clustered applying the *k*-means clustering package of scikit-learn in Python. The number of clusters are set to be the same as the number of GIS-LCZ classes. The clustering result are compared to the GIS-LCZ classification result to investigate their relationship.

3. Results and Discussion

3.1. Local Climate Zones Classification from GIS Data

The GIS-LCZ of Berlin is shown in Figure 5. In general, the LCZ classes in Berlin comprise 61.13% urban classes and the rest are distributed over natural classes. Out of the urban classes, LCZ 6 (open low-rise) dominates the distribution with 25.70% followed by LCZ A (dense trees) with 21.68%. On the other hand, LCZ 5 (open midrise), LCZ 9 (sparsely built), LCZ 2 (compact midrise), and LCZ E (bare rock or paved) occupy 15.1%, 7.25%, 7.2%, and 4.81%, respectively. Moreover, the proportion of LCZ 3 (compact low-rise), LCZ 4 (open high-rise), and LCZ 8 (large low-rise) are only 0.37%, 0.64%, and 0.1%, respectively. These statistics imply that the urban classes in Berlin are dominated by LCZs with open areas. The other natural classes, which are LCZ B (scattered trees), LCZ D (low plants), and LCZ G (water), have proportions of 5.1%, 6.05%, and 6.04%, respectively.

The result of the GIS-LCZ method is compared to the new training areas generated for Berlin, which are shown in Figure 2. The comparison is performed using the confusion matrix calculation in SAGA. Using a confusion matrix is a common method to assess the accuracy of a classification, where the classification result is compared with the reference or ground truth data. AccProd or producer accuracy implies the map accuracy from the perspective of the producer (map maker) or the probability that the reference class is

correctly classified in the classification result. On the other hand, the AccUser or user accuracy is the probability that the classified class is correctly classified in the reference class. This accuracy specifies map accuracy from the perspective of a map user. Moreover, the overall accuracy value is the number of sites correctly classified divided by the total number of sites. The kappa coefficient can also be calculated. This value describes how well the classification was executed in comparison to just a random classification. The value ranges from 0 to 1, where 1 represents a perfect match between the classification result and the reference data, and 0 is the other way around where the classification result is considered completely random [36].

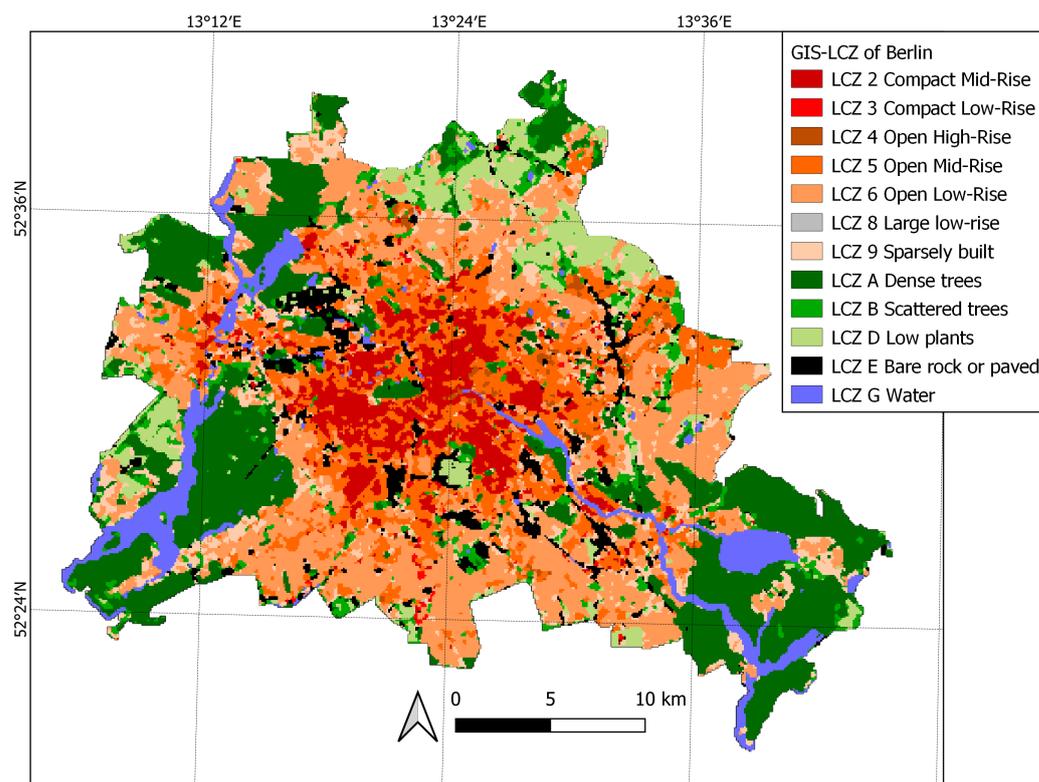


Figure 5. GIS-LCZ classification of Berlin.

The confusion matrix of GIS-LCZ compared to the TAs is tabulated in Table 3. The overall accuracy and kappa values (resp. 92.47% and 0.91) are excellent; however, the LCZ 8 does not have a good producer accuracy value since the training areas of LCZ 8 do not detect any LCZ 8 in the classification result. Nevertheless, LCZ 8 only corresponds to only 0.1% of GIS-LCZ, which is also the lowest proportion of the LCZ classes in the GIS-LCZ result. Moreover, the TA for LCZ 3 could not be created, because insufficient training areas were identified.

The WUDAPT L0 is also compared with the TAs in a confusion matrix shown in Table 4. The overall accuracy and kappa values are 74.95% and 0.69, respectively, which are lower than that of GIS-LCZ. Generally the producer accuracy of each LCZ classes is lower than that of GIS-LCZ. The producer accuracies of WUDAPT L0 of LCZ 5 and 9 have significant discrepancies compared to that of GIS-LCZ; however, the WUDAPT L0 detects the LCZ 8 far better compared to GIS-LCZ. Moreover, LCZ 3 and E are not classified in WUDAPT L0.

Table 3. Confusion Matrix of GIS-LCZ vs. TA.

LCZ	2	3	4	5	6	8	9	A	B	D	E	G	AccUser (%)
2	426	0	0	0	0	2	0	0	0	0	0	0	99.5
3	0	0	0	0	0	3	0	0	0	0	0	0	0
4	0	0	60	4	0	0	0	0	0	0	0	0	93.8
5	0	0	7	624	0	6	0	0	0	0	4	0	97.4
6	0	0	1	16	969	12	27	0	0	0	0	0	94.5
8	0	0	0	0	0	0	2	0	0	0	0	0	0
9	0	0	0	2	32	13	184	0	0	0	1	0	79.3
A	0	0	0	0	0	0	3	252	16	0	0	0	93.0
B	0	0	2	3	0	0	11	0	18	5	0	0	46.2
D	0	0	0	0	0	0	2	0	0	60	4	0	90.9
E	0	0	0	0	2	5	56	0	0	0	19	0	23.2
G	0	0	0	0	0	0	0	0	1	0	0	361	99.7
AccProd (%)	100	-	85.7	96.2	96.6	0.0	64.6	100	51.4	92.3	67.9	100	
Overall acc.	92.5%												
Kappa	0.91												

Table 4. Confusion Matrix of WUDAPT L0 vs. TA.

LCZ	2	4	5	6	8	9	A	B	D	E	G	AccUser (%)
2	406	0	1	0	2	0	0	0	0	0	0	99.8
4	0	56	340	0	0	0	0	0	0	0	0	14.1
5	20	13	217	20	4	1	0	0	0	0	0	78.9
6	0	0	81	980	2	234	0	6	0	0	0	75.2
8	0	1	9	3	35	2	0	0	0	0	20	50.0
9	0	0	0	0	0	33	0	0	0	0	0	100
A	0	0	0	0	0	0	252	3	0	0	1	96.6
B	0	0	1	0	0	10	0	21	7	0	0	53.9
D	0	0	0	0	0	0	0	5	67	8	0	83.8
E	0	0	0	0	0	0	0	0	0	0	0	-
G	0	0	0	0	0	0	0	0	0	0	386	100
AccProd (%)	95.3	80.0	33.4	97.7	85.4	11.6	100	60.0	69.1	0.0	99.7	
Overall acc.	74.6%											
Kappa	0.69											

It is observed that WUDAPT L0 is not effective at detecting the zone properties related to building geometry (H , H/W , and BSF), which leads to misclassification of some LCZs. This situation is observed in WUDAPT L0 for LCZ 2, LCZ 4, and particularly LCZ 5 as implied in the confusion matrix of Table 4. LCZ 5 has a very low producer accuracy since it classifies most of the TAs of LCZ 5 into LCZ 4. WUDAPT L0 also gives a completely random classification result on LCZ 6 and 9. It is found that the tiles classified in these classes are in reality natural classes instead of urban classes, based on the zone properties and the view from the satellite imagery as shown in Figure 6.

On the other hand, GIS-LCZ considers the building geometry, which enables it to correctly detect LCZs that are misclassified by WUDAPT L0: LCZ 2 instead of LCZ 5, LCZ 5 instead of LCZ 2 and LCZ 4, and LCZ 6 instead of LCZ 5 as indicated in the confusion matrix of Table 3. Figure 7 shows misclassification in WUDAPT L0, where the method classified the tiles as LCZ 4, but the mean building height values of these tiles are actually in the range of LCZ 5. The GIS-LCZ method addresses this shortcoming effectively.

One shortcoming of the GIS-LCZ method is its strong dependency on the availability of input data. It is observed that some tiles cannot be classified correctly due to the inadequacy of the data to calculate the zone properties. GIS-LCZ also does not manage

to detect LCZ 8, which is probably due to the unavailability of pervious surface fraction property or the small extent of the grid tile, which is insufficient for defining LCZ 8. The majority filter applied can remove the granular view and produce more homogeneous LCZ classes; however, this filter can also diminish the correctly classified GIS-LCZ classes. Some tiles of GIS-LCZ are found to belong partially to LCZ 6 or 8, which implies that the combination of two LCZs is possible to represent a local climate zone as it was also noted by other researchers [37]. The airport areas are classified as LCZ 5 in GIS-LCZ (rather than LCZ 8) because the building height property suits this class. This highlights a limitation of the LCZ framework, showing that its implementation cannot always be ideal for every city.



Figure 6. Misclassification in WUDAPT L0: LCZ 6 and 9 instead of LCZ A (Google Earth image in QGIS).



Figure 7. Misclassification in WUDAPT L0: LCZ 4 instead of LCZ 5 (Google Earth image in QGIS).

3.2. LCZ Generator

The new TAs based on GIS-LCZ data, see Figure 2, were also submitted to the LCZ generator and the result of the new LCZ map gives an accuracy value of 84% (see Figure 8). The previous (WUDAPT) submission for Berlin to the LCZ generator obtained a lower accuracy value of 61% (see Figure 9). The accuracy evaluation of the LCZ Generator refers to five indexes [38]: overall accuracy (OA), overall accuracy for the urban LCZ classes only (OAu), overall accuracy of the built vs. natural LCZ classes only (OAbu), a weighted accuracy (OAw), and the class-wise metric F1. The accuracy indexes OA, OAu, OAbu, and

OAw of the new LCZ map are 84%, 82%, 98%, and 97%, respectively. The relevant indexes of WUDAPT are 61%, 59%, 94%, and 92%, respectively. The class-wise metric F1 evaluation also shows that in the new LCZ map, except for LCZ 8 and LCZ B, the accuracy of the other eight LCZ classes are better than that of the previous submission.

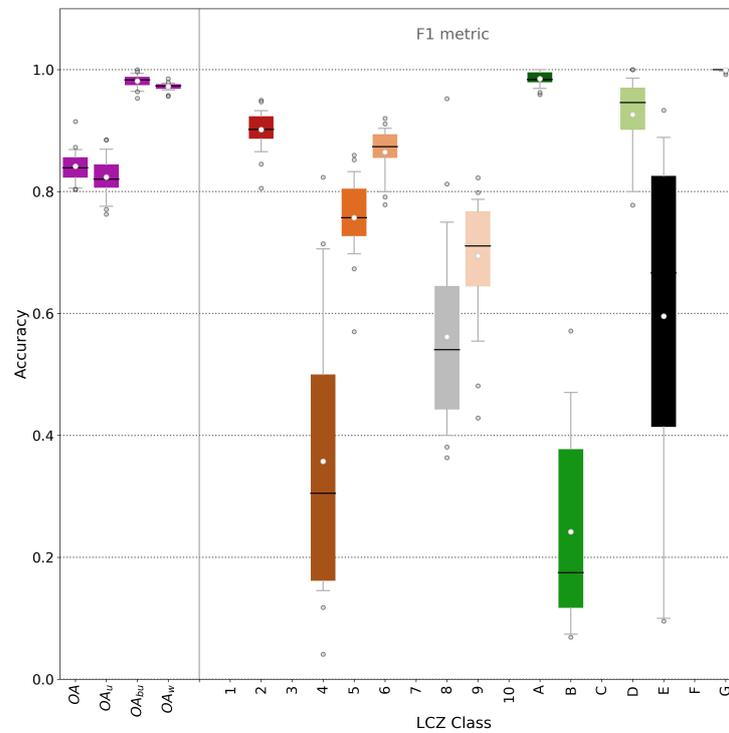


Figure 8. Boxplot accuracy of the LCZ generator map for the new training areas of Berlin; data from [39] (see also [15]).

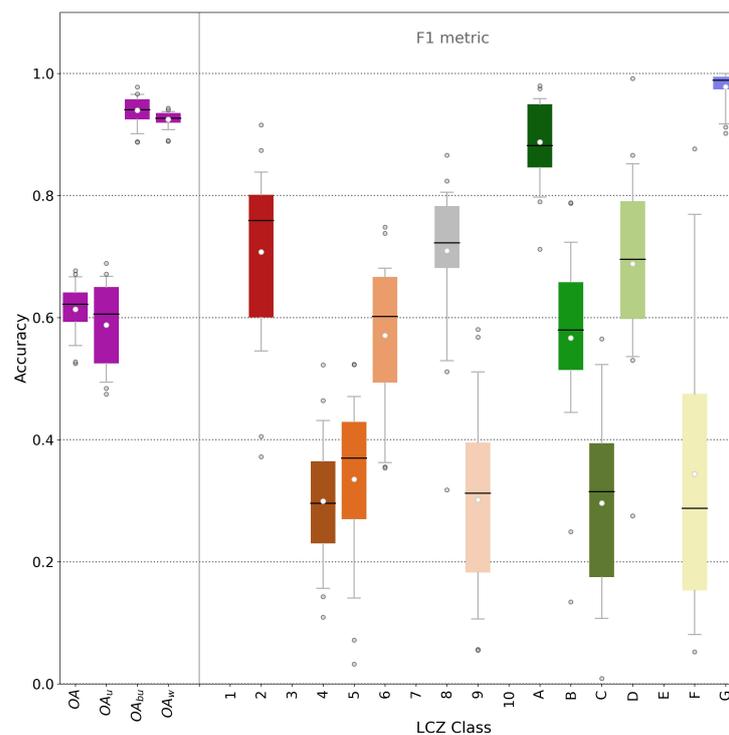


Figure 9. Boxplot accuracy of the LCZ generator map for the previous WUDAPT training areas of Berlin; data from [40] (see also [15]).

3.3. k-Means Clustering and Comparison to GIS-LCZ Classes

First, we compare the result of the k-means clustering to the previously determined GIS-LCZ classes using average cluster properties. The average of the zone properties for each cluster is shown in Table 5.

Table 5. Mean and standard deviation (SD) of properties for different clusters.

Clusters	H (m)		H/W		BSF (%)		ISF (%)		AHF (W/m ²)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	10.2	5.6	0.45	0.24	17.1	8.6	20.1	10.3	15.6	3.4
1	19.8	7.9	0.68	0.26	20.3	7.7	28.5	10.5	7.4	2.9
2	13.5	7.6	0.52	0.33	19.1	9.7	56.0	12.0	16.1	3.8
3	8.5	6.4	0.28	0.24	13.1	10.1	66.7	12.4	5.4	3.1
4	18.5	5.1	1.13	0.57	42.7	11.0	39.0	10.5	17.5	4.0
5	6.7	3.9	0.32	0.19	13.7	8.2	9.7	7.3	3.8	2.4
6	5.8	3.2	0.32	0.18	14.6	7.6	33.8	7.4	4.6	2.6
7	13.6	5.5	0.94	0.82	45.6	13.4	39.5	12.3	6.4	3.0

If we analyze each average cluster property separately, we can make the following observations. The clusters 0, 1, 2, 4 and 7 with building height between 10 m and 20 m may be middle-rise areas, the clusters 3, 5, 6 having building heights between 5 m and 10 m may be low-rise or sparsely built areas. The clusters 4 and 7 with H/W greater than 0.75 may be compact, open high-rise or lightweight areas. The cluster 3 with an H/W value of 0.28, may indicate large low-rise or heavy industry areas. The other clusters with H/W between 0.3 and 0.75 may indicate open middle-rise or open low-rise areas. The BSF values of clusters 4 and 7 are greater than 40%. These areas may be compact, lightweight low-rise or large low-rise areas. The BSF values of the other clusters are between 10% and 20%, buildings in these areas are sparsely built. This shows that the BSF of most clusters in Berlin is small. The ISF value of cluster 3 is 66.7%, which may correspond to bare rock or paved areas. The ISF value of cluster 2 is 56.0%, which may correspond to compact areas. The ISF values of clusters 0, 1, 4, and 7 are between 20% and 40% indicating open or sparsely built areas. The cluster 5 with an ISF value of 9.7% may belong to sparsely built areas. The clusters 2, 4, and 7 with an AHF more than 10 W m^{-2} may indicate compact, open or lightweight areas. On the other hand, the other clusters with an AHF less than 10 W m^{-2} may correspond to sparsely built or paved areas.

Overall, it seems difficult to establish a clear relationship between the clusters and the LCZ classes using only average cluster properties, because, depending on the property analyzed, the most likely LCZ class corresponding to each cluster differs.

Next, we performed a cross-analysis of the average zone-properties between the custom clusters and the GIS-LCZ classification. The resulting agreement values in percent are shown in Table 6. Percentages greater than 20 are marked to identify GIS-LCZ classes that are similar to the k-means clusters.

It can be seen from Table 6 that 71.8% of cluster 1 belong to GIS-LCZ 5, 67.9% of cluster 6 belong to GIS-LCZ 6, and 69.3% of cluster 4 belong to GIS-LCZ 2. Most of the members of cluster 0, 2, and 3 belong to GIS-LCZ 5 and 6. Relatively speaking, the mean building height of cluster 2 is closer to that of a midrise area, while the mean building height of cluster 0 and 3 are closer to that of a low-rise area. In addition, there may be a large number of paved areas in cluster 3. Moreover, cluster 5 is mainly composed of GIS-LCZs 6 and 9, and cluster 7 is distributed among GIS-LCZs 2, 5 and 6.

Table 6. Crosstab between GIS-LCZ classes and k-mean clusters (numbers in bold font are above 20%).

	k-Means Clusters								Total	
	0	1	2	3	4	5	6	7		
GIS-LCZ	2	4.7%	7.8%	17.4%	3.6%	69.3%	0.4%	0.2%	37.6%	11.7%
	3	0.0%	0.1%	0.4%	0.7%	1.2%	0.1%	0.1%	5.5%	0.6%
	4	0.5%	7.6%	1.5%	0.8%	0.5%	0.2%	0.2%	0.2%	1.0%
	5	33.8%	71.8%	44.7%	26.4%	22.7%	9.7%	8.9%	25.7%	24.7%
	6	52.5%	11.0%	28.3%	34.8%	6.0%	46.2%	67.9%	25.7%	42.0%
	8	0.1%	0.0%	0.1%	0.4%	0.0%	0.0%	0.3%	0.5%	0.2%
	9	3.5%	1.2%	0.5%	9.2%	0.1%	35.7%	10.8%	2.9%	11.9%
	15	4.9%	0.6%	7.0%	24.1%	0.1%	7.6%	11.6%	2.0%	7.9%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	

After the cross analysis of the GIS-LCZ classification and the k-means clustering, it can be considered that clusters 1, 6, and 4 represent open midrise areas, open low-rise areas and compact midrise areas, respectively. On the other hand, clusters 0, 2, and 3 could be transition areas from open midrise to open low-rise. Moreover, cluster 5 can be assumed to be a transition class between open low-rise and sparsely built areas, and cluster 7 could be a transition class between compact midrise and open midrise areas.

The classification of the GIS-LCZ method is based on the standard LCZ framework; therefore, the GIS-LCZ method is not able to identify zone types outside of the standard LCZ framework, in which the zone types have different range values than that defined in the standard. On the other hand, k-means clustering is based entirely on quantitative properties, which can effectively avoid the constraints of the standard LCZ framework and may distinguish zone types other than that defined by the standard LCZ framework; however, the resulting clusters do not correspond to clearly defined urban topologies, and so the researchers need to name them according to the characteristics of each cluster on a case by case basis.

4. Conclusions

In this study, it is shown that the GIS-LCZ method can improve the accuracy over the remote sensing-based WUDAPT L0 approach in deriving an LCZ map of Berlin. When compared to the new training areas, a high overall accuracy of 92.47% and a high kappa value of 0.91 were found for the GIS-LCZ map. This indicates a highly accurate classification result and a strong improvement over the previous WUDAPT L0 result with an overall accuracy of 74.95% and a kappa value of 0.69.

It is observed that the WUDAPT L0 method misclassified some LCZs due to its shortcoming in detecting the zone properties related to building geometry. On the other hand, the GIS-LCZ approach calculated the zone properties from vector and raster datasets, which correctly detects LCZs that were misclassified by WUDAPT L0. Nevertheless, the GIS-LCZ approach strongly depends on the availability of the data to calculate the zone properties. It can lead to a misclassification of LCZs when the zone properties are not available. This study also shows the limitation of the LCZ framework in addressing the finer variety of zone property values of the grid tiles in Berlin, where the LCZ classes from the standard do not fit perfectly to the calculated zone properties.

The GIS-LCZ approach classified the zone properties using a fuzzy logic algorithm that was adapted from Estacio et al. [23] and modified in order to solve the membership value problem of the left zero bound of the trapezoidal function. The standard framework for LCZ E is also modified by defining the impervious surface fraction bigger than or equal to 60%, so that it can better detect paved surfaces in urban areas. The majority filter used in the post-processing employs a window filter of 3×3 pixels instead of 5×5 pixels used in the standard WUDAPT L0, in order to keep the details of the GIS-LCZ classification.

This study also shows that the new training areas generated with the consideration of building height information can increase the accuracy of the LCZ map produced by the LCZ generator. Moreover, the k-means clustering result shows that the GIS-LCZ method tends to divide regions into clearly differentiated LCZ types according to the standard framework, while k-means clustering can identify regions with city-specific characteristics or inter-LCZ transition types. From the cross analysis between the clusters and the GIS-LCZ classes, it is found that some of the GIS-LCZ classes seem to be naturally distinguishable but some are difficult to separate.

Future improvement of the methodology presented here could proceed in several directions. Only seven of the zone properties were used here for the classification of GIS-LCZ. It would be interesting to add other zone properties (pervious surface fraction, surface albedo, and surface admittance) and to see whether the result improves. The WUDAPT L0 map has been inefficient in detecting certain LCZs, where geometric zone properties, such as H and H/W , are critical. The classification in WUDAPT L0 could be improved by introducing building height information to the machine learning algorithm. Integration of the GIS-LCZ and the WUDAPT L0 method could also be possible to obtain advantages from both of these methods by replacing the post-processing step (majority filter) of WUDAPT L0 with an aggregation step applied in GIS-LCZ method by Gál et al. [20].

Existing investigations [21,41–44] can be enhanced or reapplied by using the result of the GIS-LCZ map of Berlin. This approach is particularly suitable for practitioners attempting to characterize UHI via climate simulation as it can improve the accuracy of the simulation results by reflecting terrain features more precisely and consistently. This is especially valuable in the case of high horizontal grid resolution. The GIS-LCZ classes can also be used in the evaluation of simulation results from microscale numerical models, such as PALM-4U. Furthermore, studies that correlate UHI intensity with land cover characteristics can be extended to establish a correlation with the more accurate urban classification of the GIS-LCZ method. Another type of correlation can be carried out directly between the GIS-LCZ classes and remotely sensed land surface temperature (LST) or in situ measurements of near-surface air temperature. The GIS-LCZ classes can also be applied to design urban temperature measurement networks for the purpose of understanding spatial and temporal characteristics of urban climate. This network can be further used to analyze the air temperature conditions within the GIS-LCZ classes.

Author Contributions: Conceptualization, A.A.; Formal analysis, F.M., C.X. and J.V.; Funding acquisition, A.A.; Investigation, F.M. and C.X.; Methodology, J.V. and A.A.; Project administration, A.A.; Software, F.M. and J.V.; Supervision, J.V. and A.A.; Validation, F.M., J.V. and A.A.; Visualization, F.M. and C.X.; Writing—original draft, F.M., C.X., J.V. and A.A.; Writing—review and editing, F.M., C.X., J.V. and A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Fraunhofer Internal Programs under Grant No. Attract 003-695033.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge all WUDAPT contributors for providing the training areas for our city of interest. Map data copyrighted OpenStreetMap contributors and available from <https://www.openstreetmap.org> (accessed on 1 April 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BSF	Building Surface Fraction
CORINE	Coordination of Information on Environment
DLR	Deutsches Zentrum für Luft- und Raumfahrt
GIS	Geographic Information System
IMD	Imperviousness Density
ISF	Impervious Surface Fraction
LCZ	Local Climate Zone
OSM	OpenStreetMap
UHI	Urban Heat Island
UHII	Urban Heat Island Intensity
SUHII	Surface Urban Heat Island Intensity
SVF	Sky View Factor
WRF	Weather Research and Forecasting
WUDAPT	World Urban Database and Access Portal Tools

References

- Oke, T.R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **1982**, *108*, 1–24. [CrossRef]
- McCarthy, M.P.; Best, M.J.; Betts, R.A. Climate change in cities due to global warming and urban effects. *Geophys. Res. Lett.* **2010**, *37*. [CrossRef]
- EPA; OAR. Heat Island Impacts | US EPA. 2014. Available online: <https://www.epa.gov/heatislands/heat-island-impacts> (accessed on 11 March 2022).
- Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Liu, Z.; Berner, J.; Wang, W.; Powers, J.G.; Duda, M.G.; Barker, D.M.; et al. A Description of the Advanced Research WRF Model Version 4. NCAR Technical Notes NCAR/TN-556+STR; National Center for Atmospheric Research: Boulder, CO, USA, 2019. [CrossRef]
- Afshari, A.; Ramirez, N. Improving the accuracy of simplified urban canopy models for arid regions using site-specific prior information. *Urban Clim.* **2021**, *35*, 100722. [CrossRef]
- Molnár, G.; Gyöngyösi, A.Z.; Gál, T. Integration of an LCZ-based classification into WRF to assess the intra-urban temperature pattern under a heatwave period in Szeged, Hungary. *Theor. Appl. Climatol.* **2019**, *138*, 1139–1158. [CrossRef]
- Stewart, I.D.; Oke, T.R. Local Climate Zones for Urban Temperature Studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900. [CrossRef]
- Lehnert, M.; Savić, S.; Milošević, D.; Dunjić, J.; Geletič, J. Mapping Local Climate Zones and Their Applications in European Urban Environments: A Systematic Literature Review and Future Development Trends. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 260. [CrossRef]
- Ching, J.; Mills, G.; Bechtel, B.; See, L.; Feddema, J.; Wang, X.; Ren, C.; Brousse, O.; Martilli, A.; Neophytou, M.; et al. WUDAPT: An Urban Weather, Climate, and Environmental Modeling Infrastructure for the Anthropocene. *Bull. Am. Meteorol. Soc.* **2018**, *99*, 1907–1924. [CrossRef]
- Bechtel, B.; Daneke, C. Classification of Local Climate Zones Based on Multiple Earth Observation Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1191–1202. [CrossRef]
- Bechtel, B.; Alexander, P.; Böhner, J.; Ching, J.; Conrad, O.; Feddema, J.; Mills, G.; See, L.; Stewart, I. Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 199–219. [CrossRef]
- Bechtel, B.; Alexander, P.J.; Beck, C.; Böhner, J.; Brousse, O.; Ching, J.; Demuzere, M.; Fonte, C.; Gál, T.; Hidalgo, J.; et al. Generating WUDAPT Level 0 data – Current status of production and evaluation. *Urban Clim.* **2019**, *27*, 24–45. [CrossRef]
- Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [CrossRef]
- Wang, R.; Ren, C.; Xu, Y.; Lau, K.K.L.; Shi, Y. Mapping the local climate zones of urban areas by GIS-based and WUDAPT methods: A case study of Hong Kong. *Urban Clim.* **2018**, *24*, 567–576. [CrossRef]
- Demuzere, M.; Kittner, J.; Bechtel, B. LCZ Generator: A Web Application to Create Local Climate Zone Maps. *Front. Environ. Sci.* **2021**, *9*. [CrossRef]
- LCZ Generator. LCZ Generator Website. Available online: <https://lcz-generator.rub.de/> (accessed on 18 March 2022).
- Quan, S.J.; Bansal, P. A systematic review of GIS-based local climate zone mapping studies. *Build. Environ.* **2021**, *196*, 107791. [CrossRef]
- Lelovics, E.; Unger, J.; Gál, T.; Gál, C.V. Design of an urban monitoring network based on Local Climate Zone mapping and temperature pattern modelling. *Clim. Res.* **2014**, *60*, 51–62. [CrossRef]
- Unger, J.; Lelovics, E.; Gál, T. Local Climate Zone mapping using GIS methods in Szeged. *Hung. Geogr. Bull.* **2014**, *63*, 29–41. [CrossRef]

20. Gál, T.; Bechtel, B.; Unger, J. Comparison of two different Local Climate Zone mapping methods. In Proceedings of the ICUC9—9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment, Toulouse, France, 20–24 July 2015.
21. Skarbit, N.; Stewart, I.D.; Unger, J.; Gál, T. Employing an urban meteorological network to monitor air temperature conditions in the ‘local climate zones’ of Szeged, Hungary. *Int. J. Climatol.* **2017**, *37*, 582–596. [[CrossRef](#)]
22. Hammerberg, K.; Brousse, O.; Martilli, A.; Mahdavi, A. Implications of employing detailed urban canopy parameters for mesoscale climate modelling: A comparison between WUDAPT and GIS databases over Vienna, Austria. *Int. J. Climatol.* **2018**, *38*, e1241–e1257. [[CrossRef](#)]
23. Estacio, I.; Babaan, J.; Pecson, N.J.; Blanco, A.C.; Escoto, J.E.; Alcantara, C.K. Gis-based mapping of local climate zones using fuzzy logic and cellular automata. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2019**, *XLII-4/W19*, 199–206. [[CrossRef](#)]
24. Hidalgo, J.; Dumas, G.; Masson, V.; Petit, G.; Bechtel, B.; Bocher, E.; Foley, M.; Schoetter, R.; Mills, G. Comparison between local climate zones maps derived from administrative datasets and satellite observations. *Urban Clim.* **2019**, *27*, 64–89. [[CrossRef](#)]
25. Heldens, W.; Burmeister, C.; Kanani-Sühring, F.; Maronga, B.; Pavlik, D.; Sühring, M.; Zeidler, J.; Esch, T. Geospatial input data for the PALM model system 6.0: Model requirements, data sources, and processing. *Geosci. Model Dev.* **2020**, *13*, 5833–5873. [[CrossRef](#)]
26. OpenStreetMap Contributors. Planet Dump. 2017. Available online: <https://planet.osm.org> (accessed on 1 April 2022).
27. Copernicus. Copernicus Land Monitoring Service. Available online: <https://land.copernicus.eu/> (accessed on 11 March 2022).
28. Varquez, A.C.G.; Kiyomoto, S.; Khanh, D.N.; Kanda, M. Global 1-km present and future hourly anthropogenic heat flux. *Sci. Data* **2021**, *8*, 64. [[CrossRef](#)] [[PubMed](#)]
29. OSM. About OpenStreetMap. Available online: <https://www.openstreetmap.org/about> (accessed on 11 March 2022).
30. WUDAPT Portal. LCZ Maps Generated Based on WUDAPT Framework. Available online: <https://wudapt.cs.purdue.edu/wudaptTools/default/getlcz> (accessed on 16 March 2022).
31. WUDAPT TA. Training Area Dataset to Generate LCZ Map Based on WUDAPT Framework. Available online: <https://www.wudapt.org/the-wudapt-portal/> (accessed on 16 March 2022).
32. Lindberg, F.; Grimmond, C.; Gabey, A.; Huang, B.; Kent, C.W.; Sun, T.; Theeuwes, N.E.; Järvi, L.; Ward, H.C.; Capel-Timms, I.; et al. Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services. *Environ. Model. Softw.* **2018**, *99*, 70–87. [[CrossRef](#)]
33. Samsonov, T.E.; Varentsov, M.I. Computation of City-descriptive Parameters for High-resolution Numerical Weather Prediction in Moscow Megacity in the Framework of the COSMO Model. *Russ. Meteorol. Hydrol.* **2020**, *45*, 515–521. [[CrossRef](#)]
34. Creating WUDAPT TA. Create LCZ Training Areas—World Urban Database. Available online: <https://www.wudapt.org/digitize-training-areas/> (accessed on 7 May 2022).
35. Navin, M.S.; Agilandeswari, L. Land use Land Cover Change Detection using K-means Clustering and Maximum Likelihood Classification Method in the Javadi Hills, Tamil Nadu, India. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 51–56. [[CrossRef](#)]
36. Accuracy Metrics. GSP 216 Introduction to Remote Sensing. Available online: http://gsp.humboldt.edu/olm/_2019/courses/GSP/_216/_Online/lesson6-2/metrics.html (accessed on 17 March 2022).
37. Chen, Y.; Zheng, B.; Hu, Y. Mapping Local Climate Zones Using ArcGIS-Based Method and Exploring Land Surface Temperature Characteristics in Chenzhou, China. *Sustainability* **2020**, *12*, 2974. [[CrossRef](#)]
38. Demuzere, M.; Hankey, S.; Mills, G.; Zhang, W.; Lu, T.; Bechtel, B. Combining expert and crowd-sourced training data to map urban form and functions for the continental US. *Sci. Data* **2020**, *7*, 264. [[CrossRef](#)]
39. Muhammad, F.; Xie, C. WUDAPT Level 0 Training Data for Berlin (Germany, Federal Republic of) 2022, Submitted to the LCZ Generator. This Dataset Is Licensed under CC BY-SA. Available online: https://lcz-generator.rub.de/factsheets/d696992a4762392aff2fd540d269705fb247ea6f_factsheet.html (accessed on 17 March 2022).
40. Fenner, D. WUDAPT Level 0 Training Data for Berlin (Germany, Federal Republic of) 2018, Submitted to the LCZ Generator. This Dataset Is Licensed under CC BY-SA. Available online: https://lcz-generator.rub.de/factsheets/6d7f501c212dc888e32a3a9a5740ce930add3bc/6d7f501c212dc888e32a3a9a5740ce930add3bc_factsheet.html (accessed on 18 March 2022).
41. Ianger, I.; Pasternack, A.; Ulbrich, U.; Rust, H. Comparison of urban climate measurements in Berlin and LES model output for a special observation period. In Proceedings of the EGU General Assembly 2021, Virtual Event, 19–30 April 2021. [[CrossRef](#)]
42. Vogel, J.; Afshari, A. Comparison of Urban Heat Island Intensity Estimation Methods Using Urbanized WRF in Berlin, Germany. *Atmosphere* **2020**, *11*, 1338. [[CrossRef](#)]
43. Yang, J.; Ren, J.; Sun, D.; Xiao, X.; Xia, J.; Jin, C.; Li, X. Understanding land surface temperature impact factors based on local climate zones. *Sustain. Cities Soc.* **2021**, *69*, 102818. [[CrossRef](#)]
44. Unger, J.; Savić, S.; Gál, T.; Milošević, D.; Marković, V.; Gulyás, Á.; Arsenović, D. Urban climate monitoring networks based on LCZ concept. In Proceedings of the ICUC9—9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment, Toulouse, France, 20–24 July 2015.