Technical University of Munich (TUM)
Department of Aerospace and Geodesy
Chair of Geoinformatics
Univ.-Prof. Dr. rer. nat. Thomas H. Kolbe

Master Thesis

Title:

Inference of Urban 3D Morphology from Satellite Imagery and GIS Datasets - Application to Urban Climate Characterization

Fadel Muhammad

Course   :   Geodesy and Geoinformation (Master)

Supervisor   :   Dr. rer. nat. Tatjana Kutzner (Chair of Geoinformatics, TUM)

Dr.-Ing. Julian Vogel (Fraunhofer Institute for Building Physics)

Dr. Afshin Afshari (Fraunhofer Institute for Building Physics)

In cooperation with Fraunhofer Institute for Building Physics (IBP)

2021

Declaration of Authorship

With this statement I declare that I have independently completed this Master's thesis. The thoughts taken directly or indirectly from external sources are properly marked as such. This thesis was not previously submitted to another academic institution and has also not yet been published.

Munich, 01.10.2021   Fadel Muhammad

Acknowledgement

Abstract

The study of urban heat island (UHI) is becoming more significant since the growing of population and rapid urbanization magnify the effect of UHI in the urban area. In comparison to the rural area surrounding the city, the UHI effect leads to a lower air temperature during the day and a higher air temperature at night due, in large part, to solar radiation heat entrapment in the urban area. Mesoscale numerical modelling, such as the Weather Research and Forecasting (WRF) model, has been used for simulating the UHI intensity (UHII) in the urban area. This model requires, among other input data, the land use/land cover (LULC) map to run the simulation. However, the available LULC default datasets are too coarse for urban simulation. The so-called local climate zone (LCZ) is an approach that is introduced to classify the urban area into 17 classes, which are defined by the zone properties. The LCZ map can be used as LULC input for UHII simulation. It has a finer spatial resolution in comparison to the default LULC used by the WRF model, which is expected to give a better simulation result.

To generate the LCZ map, three methods can be used: in-situ measurements, machine learning classification on satellite imagery, and GIS-based method. For obvious reasons, urban in-situ measurements conducted over a large urban area do not constitute a practical approach. A more convenient approach has been proposed by a project, namely the World Urban Database and Access Portal Tools (WUDAPT) to generate the LCZ map. It is called WUDAPT level 0 (L0). The WUDAPT L0 LCZ map is obtained by classifying the satellite imagery, such as Landsat images, with the training areas digitized over Google Earth imagery. However, this method is dependent on the number and quality of the training areas. Another approach to derive the LCZ map is based on the Geographic Information System (GIS) method. This method has a better accuracy than WUDAPT L0 since it is able to quantify the zone properties of LCZ classes based on high-resolution vector or raster datasets.

The GIS approach is used in this thesis to obtain the LCZ map of Berlin. Fuzzy logic is employed for the purpose of classification of the zone properties to yield the GIS-LCZ map over 100 x 100 meter 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. After the classification, a post processing is carried out by applying a majority filter to the result of the classification. This result of GIS-LCZ is then compared to WUDAPT's training area for accuracy assessment, which shows good overall accuracy and kappa values. The GIS-LCZ is also compared to WUDAPT L0, which leads to a conclusion: lack of WUDAPT in detecting building height and inadequate data for the classification of GIS-LCZ lead to the misclassification of the LCZ. The GIS-LCZ map is further correlated with the surface urban heat island intensity (SUHII) and the near-surface air UHII. At the end, an analysis is conducted to look for the possibility to enrich the 3D city model of CityGML using the result of GIS-LCZ classification.

Table of Contents

1 Introduction

1.1 Motivation

The urban heat island (UHI) effect has been defined as a phenomenon whereby the near-surface urban canopy air temperature is, in average, higher than that of its surrounding countryside (Oke 1982). Its intensity is likely to keep increasing in the future due to population and urbanization growth (McCarthy 2010). The UHI intensity characterizes urban climates and is related to a negative impact on the sustainable 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 (EPA 2021). Therefore, forecasting of UHI is becoming significantly important.

Numerical modelling such as with the Weather Research and Forecasting (WRF) model (Skamarock 2019) can be applied to simulate the UHI effect (Vogel and Afshari 2020). The WRF model requires land use/land cover (LULC) information for the simulation. Often, a conventional LULC data source such as CORINE is used for this purpose. A local climate zone (LCZ) based map is another possible data source for LULC characterization, which can be used in a WRF simulation. The LCZ method was developed by Stewart and Oke (2012) for urban climate study purposes by classifying urban and rural sites into 10 urban classes and 7 natural classes. These classes are defined by 10 zone properties such as sky view factor, mean building height, and surface fractions. This LCZ scheme has been used extensively in studying UHI in different approaches.

The World Urban Database and Access Portal Tools (WUDAPT) (WUDAPT 2021) is a project deriving the LCZ map or the so-called WUDAPT level 0 (L0) data by running a machine learning supervised classification approach over satellite imagery such as Landsat 8 or Sentinel 2 (Bechtel et al. 2019). Another method to derive the LCZ is the GIS-based method where vector or raster datasets of building, roads, vegetation, land use or land cover are used for quantifying the zone properties to define LCZ classes. The GIS method is claimed to have primarily a better accuracy for deriving LCZ compared to WUDAPT L0. One major downside of WUDAPT is that its accuracy highly depends on the number and quality of training areas (Wang et al. 2018; Hammerberg et al. 2018).

The aim of this thesis is to derive the LCZ map based on the GIS-based method for the city of Berlin. The classification result will be compared with the available WUDAPT L0 of Berlin and will be used as an input of land cover/land use for the WRF model. The WUDAPT L0 of Berlin has a resolution of 100 m derived from Landsat Imagery of generally 30 m resolution. Applying the GIS-LCZ map, which will be derived from a much higher resolution dataset, will give a possibility to have a better simulation result with the WRF model.

1.2 Objectives

There are four objectives of this thesis:

a. Developing an algorithm to derive the LCZ classification from GIS data to obtain detailed and accurate land use and urban morphology.
b. Assessing the accuracy of the GIS-LCZ map and comparing the map to WUDAPT L0.
c. Finding a correlation between zone properties on the one hand and surface urban heat island intensity (SUHII) and near-surface air urban heat island intensity (UHII) on the other hand.
d. Analyzing possible enhancement to CityGML with the calculation result of zone properties and the classification of the GIS-LCZ map.

1.3 Research questions

a. How can the LCZ classes accurately and unambiguously be classified from the calculated zone properties and what kind of algorithm can be used?
b. How is the accuracy of the GIS-LCZ map compared to the WUDAPT training areas?
c. How is the agreement between the GIS-LCZ map and WUDAPT L0 map?
d. What correlation between the result of GIS-LCZ and the surface/near-surface urban heat island intensity can be found?
e. How can CityGML be enriched with the calculated zone properties and the result of the GIS-LCZ classification?
f. How will the new concept of CityGML 3.0 support the GIS-LCZ classification?

2 State of the Art

This chapter comprises the fundamentals related to the thesis, which are urban heat island, local climate zones (LCZ), the World Urban Database and Access Portal Tools (WUDAPT), and geographic information system-based local climate zones (GIS-LCZ). Theories regarding the evaluation and analysis of the classification result, which are confusion matrix, correlation analysis, and City Geography Markup Language (CityGML) are also disclosed in this chapter. Finally, in the last subsections, the dataset and software used in this thesis are also explained.

2.1 Urban heat island (UHI)

Urban heat island (UHI) refers to "the atmospheric warmth of the city compared to its countryside which is traditionally measured at standard screen height (1-2 m above ground), below the city's mean roof height in a thin section of the boundary layer atmosphere called the urban canopy layer" (Stewart and Oke 2012). The air temperature in the urban canopy layer (UCL) is usually warmer compared to the air in the countryside at the standard screen height as illustrated in Figure 1. The UHI phenomenon is caused mainly by the differences in land cover and structure between urban and rural areas. These differences increase with population growth and rapid urbanization happening in the urban areas.
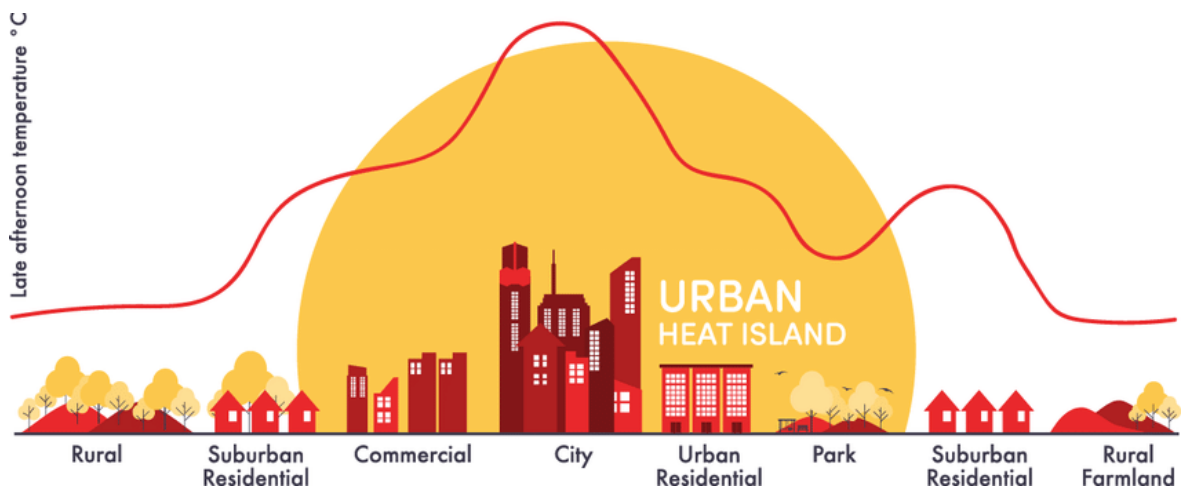
Figure 1 Urban heat island (Fuladlu et al. 2018)

There are factors, which cause the UHI in the urban area (EPA, 2008):

- Reduced vegetation
  Urban areas are characterized by reduced trees and vegetation compared to rural areas. Trees and vegetation reduce air temperatures through evapotranspiration. In this process, plants release water to the air while absorbing ambient heat, which leads to a reduction of air temperature. In urban areas, this process is less due to the predomination of dry and impervious surfaces. This will contribute to the increment of surface and air temperatures.

- Properties of urban materials
  Solar reflectance (albedo), thermal emissivity, and heat capacity are properties of urban materials that influence UHI intensity, since they are the parameters related to the reflection, emission, and absorption of solar energy. Albedo determines the percentage of solar radiation reflected by a surface. Low albedo materials will reflect only a small part of the radiation and will increase the UHI effect.

  Thermal emissivity is the ability of a surface to shed heat or emit long-wave radiation. Surfaces with high thermal emissivity will be cooler as they lose heat more rapidly. The other property is heat capacity, which describes the ability of materials to store heat. Building materials, such as stone and steel, are high heat capacitors compared to materials from rural areas, such as sand and dry soil.

- Urban geometry
  Urban geometrical parameters that are related to the UHI are the spacing and dimensions of buildings. They influence energy absorption, wind flow, and a given ability of a surface to release long-wave radiation back to space. In the city, a building is surrounded by neighboring buildings that obstruct the release of heat from the building. This will lead to a high thermal mass inside the city. During the night the UHI effect will be clearer when the temperature of the city is warmer compared to the rural area due to the hindrance of the heat dissipation between the buildings.

- Anthropogenic heat
  Anthropogenic heat refers to heat generated by human activities and contributes to the urban heat island. This heat is estimated by summing up the energy consumed for cooling and heating, transportation, running appliances, and industrial processes.

- Additional factors: weather and geographic location
  The primary properties of weather that influence the UHI are wind and cloud cover. Generally, UHI occurs during the periods of clear skies and calm winds where the sun radiation is at a maximum during these conditions and the amount of heat that is convected away is at a minimum. The geographic location determines the climate by the form and features of the topography. A lake might reduce the temperature of an area by cooling through convection or evaporation.

The increasing urban temperature could cause several impacts (EPA, 2021):

- Increased energy consumption
  In the city, the use of air conditioning to cool down the temperature of buildings increases during summer. This will lead to an increase of energy consumption.

- Elevated emissions of greenhouse gases and air pollutants
  The rising energy supply during summer is still mainly generated from fossil fuel. This will add more emissions of greenhouse gasses and air pollutants. The

emission of greenhouse gasses (e.g. carbon dioxide) contributes to global climate change. On the other hand, the air pollutants harm the human health and generate air quality problems such as the creation of smog (ground-level ozone), acid rain, and fine particulate matter.

- Endanger human health and comfort
  The urban area is getting warmer with the increment of UHI intensity. This could lead to heat stress, illness and even death for people living in the city. Older adults are the most vulnerable to extreme heat conditions because they are more sensitive to heat, more likely in poor health condition and more isolated. Young children are also susceptive to extreme heat since they have a small body size and a still developing respiratory system. They have an increased risk of getting aggravated asthma and other lung related diseases caused by smog and ozone air pollution.

- Impaired water quality
  The high temperatures of rooftop and pavement surfaces can heat up the rain water runoff which flows into sewers, streams, rivers, ponds, and lakes. The rising water temperatures can be stressful to the aquatic life since the water temperature affects the reproduction and metabolism of many aquatic species.

After realizing these impacts, it is becoming significantly important to predict UHI intensity in urban areas. UHI phenomenon is clearly a worldwide issue, since urban areas keep expanding all over the world. By the end of 2050, it is predicted that 68% of the world population will live in urban areas (UN 2018). In the future, if there is no effective mitigation from the stakeholders, the UHI phenomenon could worsen and lead to more severe impacts.

There are four types of UHI as explained by Oke et al. (2017):

- Subsurface urban heat island: underground temperature difference between the city and rural area.
- Surface urban heat island: surface temperature difference between the ground in the city and the ground in a rural area
- Canopy layer urban heat island: air temperature difference between the UCL of the urban area and the near-surface layer of the rural area.
- Boundary layer urban heat island: air temperature difference between a layer, which is between the top part of the UCL and the top part of the urban boundary layer (UBL), and the similar layer height of atmospheric boundary layer (ABL) in the rural area.

In this thesis, there are two types of UHI, which are used for the evaluation: the surface urban heat island intensity (SUHII) and canopy layer near-surface urban heat island intensity (UHII). The SUHII can be obtained from land surface temperature (LST) derived from thermal satellite imaging instruments such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Spinning Enhanced Visible and Infrared Imager (SEVIRI). On the other hand, the UHII is best investigated with mesoscale numerical weather prediction models, such as the WRF model (WRF 2021). These mesoscale models require land use/land cover (LULC) data for the simulation. However, the

available LULC default datasets are usually too coarse for urban simulation. The WRF model, for example, uses as default a LULC map derived from the (MODIS) dataset with 30 s resolution.

2.2 Local climate zones (LCZs)

The conventional approach to determine the magnitude of urban heat island is to simply measure temperatures at two or more sites and classify these sites as urban or rural. The urban heat island magnitude is then the temperature difference between the urban and rural sites. However, this method is claimed to suffer in giving quantitative metadata of the site exposure or the land cover that influences the screen-height temperature at the specific site (Stewart and Oke (2012)).

Stewart and Oke (2012) proposed LCZs as a climate-based classification of urban and rural regions by defining the regions with uniform surface structure, surface cover, surface fabric, and human activity which could span from 100 m to few kilometers in horizontal scale. This approach is intended to be universally and relatively easily implemented to local temperature studies applying screen-level observations. This approach conceptually aims to assist consistent documentation of the site metadata and to provide an objective guideline in measuring the urban heat island effect's magnitude in any city in the world.

The name "local climate zones" has a meaning of *local* in scale and *climate* in nature as well as *zonal* in representation. Every zone in LCZs has a representative screen height temperature regime, which is most obvious over dry surfaces, during calm and clear nights, as well as in areas with simple relief. These temperature regimes prevail year-round and are related with the homogeneous environments of cities, natural biomes, and agricultural lands.

The LCZ methodology comprises 17 zones, 10 of which are built environments and the other seven are land cover types as shown in Figure 2. These zones are characterized by its 10 properties, which are related to surface structure, surface cover, surface fabric and human activity. The surface structure influences local climate by modification of the airflow, the atmospheric heat transport, and the balances of shortwave and longwave radiations. The surface cover modifies the moisture availability, the albedo, and the cooling/heating potential of the ground.

The 10 properties that define LCZs are as follows:

- **Sky view factor (SVF)**: fraction of sky hemisphere which is visible from ground level. The value ranges from 0 to 1 and it depends on variation of spacing and height of building and trees. The value of 0 represents that the view from this area is completely blocked from the sky and the value of 1 tells that the area can see the whole sky.

- **Aspect ratio (H/W)**: aspect ratio of street canyons (for LCZ 1-7), building spacing (for LCZ 8-10), tree spacing (for LCZ A-G). The aspect ratio is calculated between the mean height and the mean width of the specified object (street canyons, building or tree).

- **Mean building/tree height (height of roughness element)**: average value of building heights (LCZ 1-10) and tree/plant heights (LCZ A-F).

- **Terrain roughness class (TRC)**: classification of terrain roughness length based on Davenport et al. (2000). The terrain roughness length is a wind profile parameter used to define the roughness of a surface area. The roughness is related to the efficacy of a surface area in transforming the average wind energy, which flows over it, into a turbulent motion above the surface (Davenport et al. 2000)
- **Building surface fraction (BSF)**: fraction of land which is covered by buildings.

- **Impervious surface fraction (ISF)**: fraction of land which is paved or covered by rock.

- **Pervious surface fraction (PSF)**: fraction of land which is covered by vegetation, water or bare soil.

- **Surface admittance**: measure of a surface's ability to absorb or release heat.

- **Surface albedo**: fraction of solar radiation reflected by a surface.

- **Anthropogenic heat flux (Q_F)**: mean annual heat flux density where the source of the heat could be from generally human activity and fuel combustion.

The values for some of these zone properties are shown in Figure 3.

To classify the LCZs, users should examine three steps:

- Collecting site metadata: to quantify the surface properties of the area of the temperature sensor which is optimally done by conducting field survey in person. The users acquire data regarding local horizon, building geometry, land cover, surface relief, surface wetness and population density.
- Defining the thermal source area, which is the surface area for which the sensor measures the temperature.
- Selecting the local climate zone. The site metadata collected in the first step will lead users to the suitable LCZ classes. The suitable LCZ class does not have to be a perfect match for the metadata. The process of assigning a best-fit LCZ class is rather an interpolation approach than a straight matching.
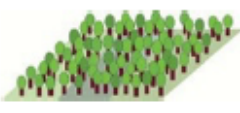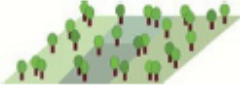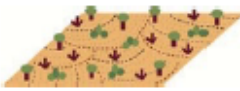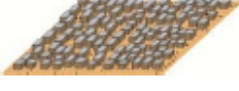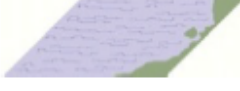
Built types	Definition	Land cover types	Definition
1. Compact high-rise	Dense mix of tall buildings to tens of stories. Few or no trees. Land cover mostly paved. Concrete, steel, stone, and glass construction materials.	A. Dense trees	Heavily wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
2. Compact midrise	Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	B. Scattered trees	Lightly wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
3. Compact low-rise	Dense mix of low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	C. Bush, scrub	Open arrangement of bushes, shrubs, and short, woody trees. Land cover mostly pervious (bare soil or sand). Zone function is natural scrubland or agriculture.
4. Open high-rise	Open arrangement of tall buildings to tens of stories. Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	D. Low plants	Featureless landscape of grass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park.
5. Open midrise	Open arrangement of midrise buildings (3–9 stories). Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	E. Bare rock or paved	Featureless landscape of rock or paved cover. Few or no trees or plants. Zone function is natural desert (rock) or urban transportation.
6. Open low-rise	Open arrangement of low-rise buildings (1–3 stories). Abundance of pervious land cover (low plants, scattered trees). Wood, brick, stone, tile, and concrete construction materials.	F. Bare soil or sand	Featureless landscape of soil or sand cover. Few or no trees or plants. Zone function is natural desert or agriculture.
7. Lightweight low-rise	Dense mix of single-story buildings. Few or no trees. Land cover mostly hard-packed. Lightweight construction materials (e.g., wood, thatch, corrugated metal).	G. Water	Large, open water bodies such as seas and lakes, or small bodies such as rivers, reservoirs, and lagoons.

VARIABLE LAND COVER PROPERTIES

Variable or ephemeral land cover properties that change significantly with synoptic weather patterns, agricultural practices, and/or seasonal cycles.

Built types	Definition	Land cover types	Definition
8. Large low-rise	Open arrangement of large low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Steel, concrete, metal, and stone construction materials.		
9. Sparsely built	Sparse arrangement of small or medium-sized buildings in a natural setting. Abundance of pervious land cover (low plants, scattered trees).	b. bare trees	Leafless deciduous trees (e.g., winter). Increased sky view factor. Reduced albedo.
		s. snow cover	Snow cover >10 cm in depth. Low admittance. High albedo.
10. Heavy industry	Low-rise and midrise industrial structures (towers, tanks, stacks). Few or no trees. Land cover mostly paved or hard-packed. Metal, steel, and concrete construction materials.	d. dry ground	Parched soil. Low admittance. Large Bowen ratio. Increased albedo.
		w. wet ground	Waterlogged soil. High admittance. Small Bowen ratio. Reduced albedo.

Figure 2. Local climate zones (Stewart and Oke 2012)

14

TABLE 3. Values of geometric and surface cover properties for local climate zones. All properties are unitless except height of roughness elements (m).							
Local climate zone (LCZ)	Sky view factor[a]	Aspect ratio[b]	Building surface fraction[c]	Impervious surface fraction[d]	Pervious surface fraction[e]	Height of roughness elements[f]	Terrain roughness class[g]
LCZ 1 Compact high-rise	0.2–0.4	> 2	40–60	40–60	< 10	> 25	8
LCZ 2 Compact midrise	0.3–0.6	0.75–2	40–70	30–50	< 20	10–25	6–7
LCZ 3 Compact low-rise	0.2–0.6	0.75–1.5	40–70	20–50	< 30	3–10	6
LCZ 4 Open high-rise	0.5–0.7	0.75–1.25	20–40	30–40	30–40	>25	7–8
LCZ 5 Open midrise	0.5–0.8	0.3–0.75	20–40	30–50	20–40	10–25	5–6
LCZ 6 Open low-rise	0.6–0.9	0.3–0.75	20–40	20–50	30–60	3–10	5–6
LCZ 7 Lightweight low-rise	0.2–0.5	1–2	60–90	< 20	<30	2–4	4–5
LCZ 8 Large low-rise	>0.7	0.1–0.3	30–50	40–50	<20	3–10	5
LCZ 9 Sparsely built	> 0.8	0.1–0.25	10–20	< 20	60–80	3–10	5–6
LCZ 10 Heavy industry	0.6–0.9	0.2–0.5	20–30	20–40	40–50	5–15	5–6
LCZ A Dense trees	<0.4	>1	<10	<10	>90	3–30	8
LCZ B Scattered trees	0.5–0.8	0.25–0.75	<10	<10	>90	3–15	5–6
LCZ C Bush, scrub	0.7–0.9	0.25–1.0	<10	<10	>90	<2	4–5
LCZ D Low plants	>0.9	<0.1	<10	<10	>90	<1	3–4
LCZ E Bare rock or paved	>0.9	<0.1	<10	>90	<10	<0.25	1–2
LCZ F Bare soil or sand	>0.9	<0.1	<10	<10	>90	< 0.25	1–2
LCZ G Water	>0.9	<0.1	<10	<10	>90	–	1

Figure 3. Zone properties of LCZ (Stewart and Oke 2012)

2.3 The World Urban Database and Access Portal Tools (WUDAPT)

The initial approach in deriving local climate zones is by doing in-situ measurements. On the other hand, there are two approaches applied in generating local climate zones namely the WUDAPT-based method and the GIS-based method. Those methods utilize data that is readily available without the need of observing the field site directly to generate the classification of LCZs.

WUDAPT is a project initiated by urban climate researches to provide universally coherent and consistent information on form and function of urban morphology for climate studies (Ching et al. 2018). The form comprises surface cover, urban geometry and construction materials. On the other hand, the function is activities that lead to consumption of energy, materials and water and lead to emission of waste heat, particulates, gases and water (Mills et al. 2015).

WUDAPT information consists of three levels of detail (L) and is gathered using distinct methodologies:

- Level 0 data comprises a local climate zone map which is based on the work of Stewart and Oke (2012)
- Level 1 data gives a better representation for each LCZ through sampling by providing information regarding urban form and function in a finer spatial resolution. Level 1 data has a representation in three-dimensional form.
- Level 2 data refines the data further by giving detailed description of urban parameter values for boundary layer modelling.
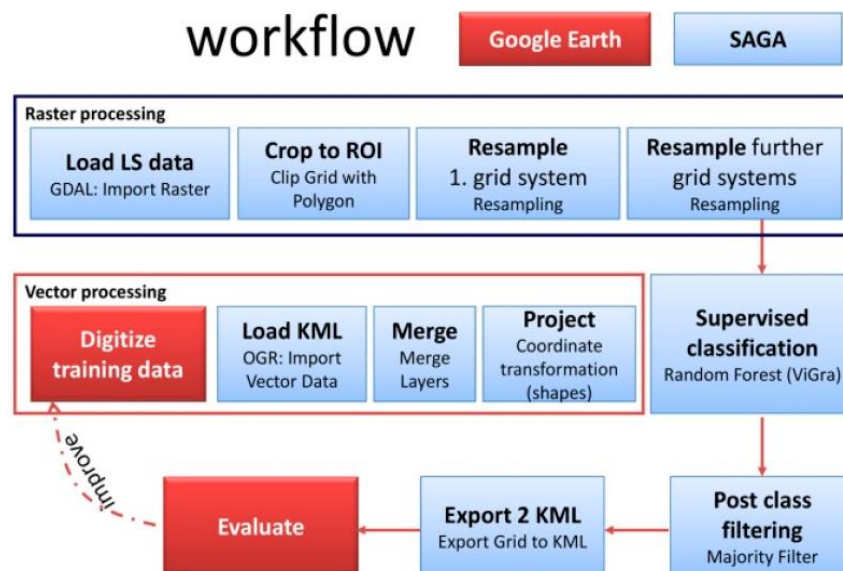
Figure 4. Workflow of generation of WUDAPT L0 (Bechtel et al. 2015)

WUDAPT Level 0 (L0) data is produced by applying machine learning supervised classification of Landsat or Sentinel images. The workflow of generating L0 data is as follows (Figure 4) (Bechtel et al. 2019):

- Digitizing training areas (TA) in Google Earth:
  Volunteer local experts who know the respective city to be mapped digitize polygon areas that represent local climate zones occurring in the city. A training area would have ideally about 1 km² area. The more training areas are digitized the better the classification result should be. Figure 5 shows the training areas for Berlin.

Figure 5. Training areas digitized over Berlin (Fenner, 2015)

- Raster processing in SAGA GIS:
  Satellite imagery from Landsat or Sentinel can be employed as an input raster for the classification. A Local Climate Zone Classification tool is provided by an open source software, the System for Automated Geoscientific Analyses (SAGA) (Conrad et al. 2015). Before the classification, preprocessing to the satellite imagery needs to be done by cropping the raster with respect to the region-of-interest (ROI) and resampling of the raster to a defined grid size.

  After that, the classification can be carried out with the Local Climate Zone Classification tool. The classification step can be performed several times to improve the classification result. The classification result is then checked with the Google imagery and existing TA can be modified or additional TA can be digitized to repeat the classification process. If the classification result is too granular, it can be filtered using a majority filter.

- Evaluation:
  After a volunteer generated a sufficiently high quality LCZ classification map, the respective training areas and Landsat data will be submitted to the WUDAPT team and made available on the website. The submitted data will be

further processed in a local climate zone production and quality assessment workflow (as shown in Figure 6) before dissemination to the database and portal.
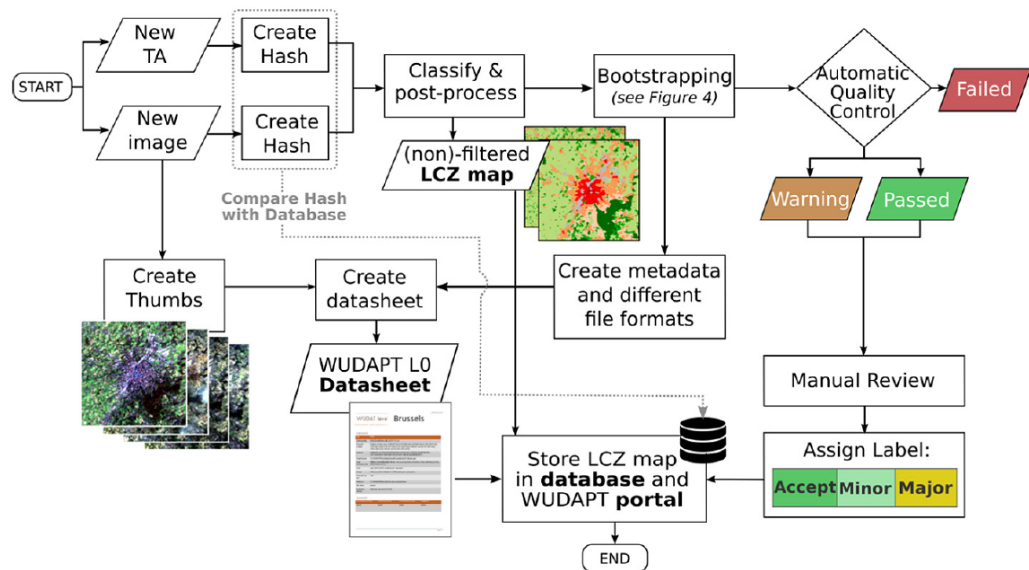
Figure 6. LCZ generation and quality assessment scheme in WUDAPT (Becthel et al. 2019)

Every data submitted to the WUDAPT team will be checked in the workflow whether it is a new data or not by comparing the hash value with the previously processed data. If the training data or the Landsat data is new, the classification, post-processing, and documentation step is activated. In this production chain, three approaches are applied for the quality assessment of the level 0 LCZ maps: cross validation, manual review and cross-comparison with other data. The cross validation and manual review by experts determine whether the LCZ map generated has an adequate quality and can be published on the WUDAPT portal.

The cross validation is applied automatically by employing a bootstrapping method as shown in Figure 7. This method is a random sampling scheme which estimate accuracy measures from the training areas. A number of sub-samples from the training areas is used for the classification and the accuracy of the classification result is evaluated with the samples that are not used.
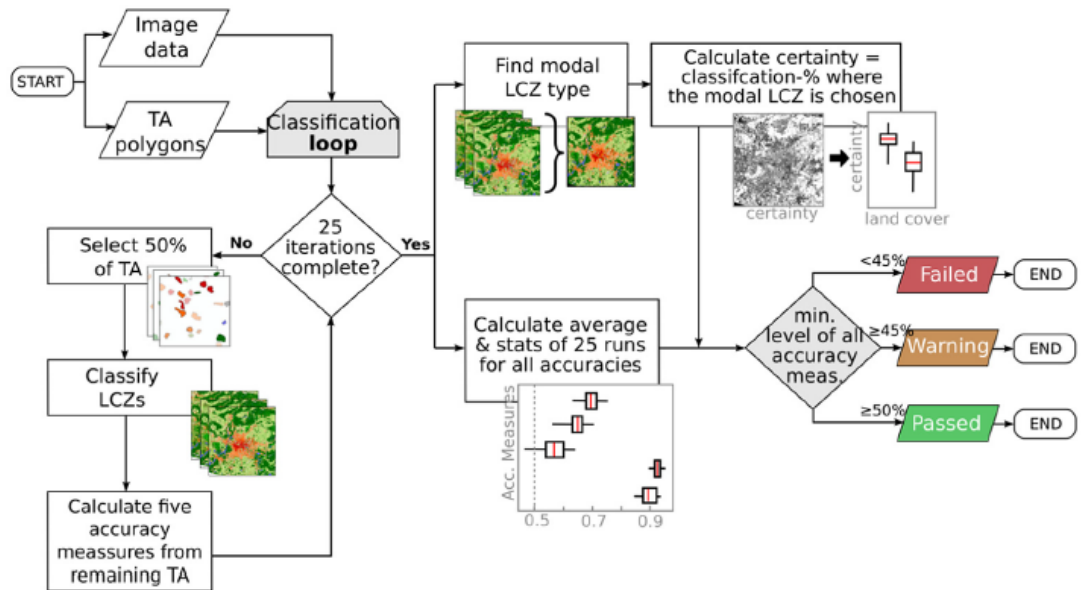
Figure 7. Automated quality assessment with bootstrapping (Bechtel et al. 2019)

WUDAPT provides level 0 LCZ maps for several cities around the world in the WUDAPT portal. However, this portal is no longer maintained and updated, and the user is asked to move to a new website, namely LCZ Generator released this year by Ruhr University Bochum (Demuzere et al. 2021). This website has published around 348 LCZ maps from around the world until now (LCZ Generator 2021). Users who want to generate their LCZ maps can easily submit the training areas to this website and the LCZ map will be generated automatically. The generated LCZ map is provided with the accuracy measures, which are openly published on the website together with training areas.

2.4 Geographic Information System (GIS) – based method for local climate zones

Another approach to derive local climate zones is by applying the so-called GIS-based method. As summarized by Quan and Bansal (2021), the first GIS-based approach employed in deriving LCZ was done by Lelovics et al. (2014) and Unger et al. (2014) and since then, 16 representative studies, which implemented the GIS-based LCZ mapping, have been published. From these studies, the general GIS-LCZ mapping process is illustrated in Figure 8.
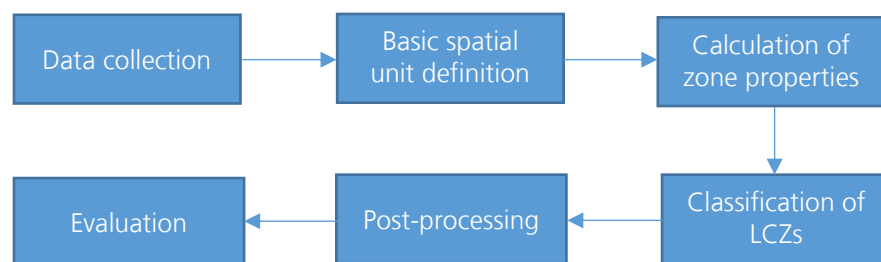
Figure 8. GIS-LCZ mapping method (adopted from Quan and Bansal (2021))

The data used to calculate the zone properties explained in section 2.2 can be in the form of vector or raster data. The vector data from OpenStreetMap, for example, is processed to derive building surface fraction. Raster data from satellite imagery such as Landsat and Copernicus, can also be employed to calculate the other properties, such as impervious surface fraction.

The basic spatial unit (BSU) is the minimum mapping unit of the classification. The classification is carried out for every BSU so that every BSU will have the values for the zone properties. The BSU has been defined in several representations. It can be in the form of grid, lot area polygon, neighborhood, urban block and sensor unit area. Geletič and Lehnert (2016) applied a 100 m x 100 m grid size for the classification. Unger et al. (2014) defined lot area polygon for the classification as illustrated in the left image in Figure 9. Quan et al. (2017) implemented urban block approach as the BSU as shown in the right image in Figure 9.
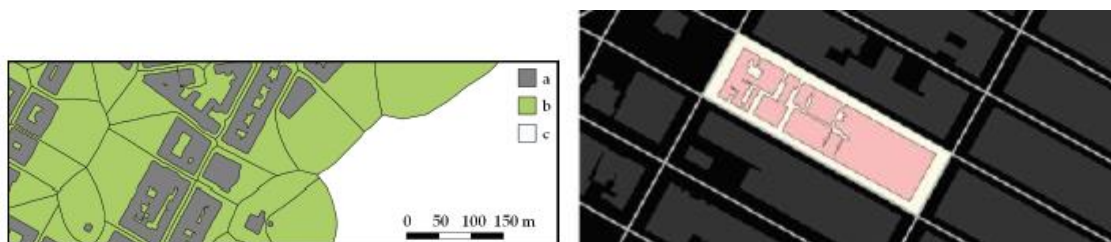
*Figure 9. Basic spatial unit. Left image: lot area polygon (Unger et al. 2014). Right image: urban block (Quan et al. 2017)*

The number of zone properties used in the 16 representative studies vary. One study only used 4 of the zone properties (Quan et al. 2017) and another study used only two of them but with 10 additional zone properties (Hidalgo et al. 2019). A study in Colombo, Sri Lanka used the default 10 zone properties (Perera and Emmanuel 2018).

After the zone properties are calculated, the classification is done by implementing a specific classifier. The representative studies used different kinds of classifiers: standard rule-based, modified standard rule-based, new rule-based, Naïve Bayes, K-means integrated rule-based, and Fuzzy rule-based. From the 16 studies, the most implemented classifiers are modified standard rule-based (six studies) and fuzzy rule based (six studies: Lelovics et al. 2014, Šećerov et al. 2015, Geletič and Lehnert 2016, Quan et al. 2017, Wu et al. 2018, and Estacio et al. 2019).

The classification result can be further processed by aggregating or merging the basic spatial unit. At the final step, evaluation is carried out on the classification result or post-processed result. The evaluation can be done by comparing the result with remote sensing-based method (WUDAPT L0 map) and by validating the result with expert knowledge and temperature observations.

2.5 Confusion matrix

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. An

example of a confusion matrix is shown in Table 1. This table denotes the result of classification of forest and water compared to the reference data.

Table 1. Confusion matrix

	Class	Reference		SumUser	AccUser
		Forest	Water		
Classified	Forest	20	10	30	66,67%
	Water	30	50	80	62,5%
SumProd		50	60		
AccProd		40%	83,33%		
Overall accuracy = 63,64%					

The first column shows the classified classes and the first row implies the reference classes. From this table, it indicates that there are 20 sites correctly classified as forest and 50 sites correctly classified as water.

The *SumProd* explains the total number of values in a class of reference data. The AccProd or producer accuracy for a class is defined as correctly classified class divided by *SumProd.* This value 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 classifed class is correctly classified in the reference class. This accuracy specifies map accuracy from the perspective of a map user and defines as the correctly classified class divided by *SumUser.* The *SumUser* is total number of values in a classified class.

Moreover, the overall accuracy value is the number of sites correctly classified divided by the total number of sites (total of *SumProd* or *SumUser*). Besides that, 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 (Accuracy Metrics, 2021). The kappa coefficient can be categorized as follows (Strunz, 2021):

- Excellent : >0.81
- Good : 0.80 – 0.61
- Moderate : 0.60 – 0.41
- Weak : 0.40 – 0.21
- Bad : ≤0.2

2.6 Correlation analysis

In this thesis, there are two correlation calculations. First is correlation of SUHII and urban fraction from GIS-LCZ over different GIS-LCZ classes. The second one is the correlation calculation of UHII and the urban fraction from WRF over GIS-LCZ. Correlation is a linear regression calculation, which yields a function that has minimum distance (in the least squares sense) between the fitted regression line and all data

points (Regression 2021). Beside the regression coefficients, there are three other main measures obtained from this regression analysis:

- Correlation value: defines the strength of a linear relation between two variables. This value is between -1 and 1, which both of them show a perfect correlation result where all data points are distributed along a straight line. A positive correlation implies a positive relation between the variables, while a negative correlation tells a negative association between them (Correlation 2021).

- R^2 value: denotes how close the data points to the fitted line. It is calculated as the division of the explained variable variation divided by the total of the variation. This will give a percentage value. 0 implies that none of variability is around the mean value and 1 implies that the regression line explains all the variability of the data around its mean (Regression 2021).

- Standard error: is the standard deviation of the coefficient estimates (BSE 2021).

2.7 City Geography Markup Language (CityGML)

CityGML is an international standard issued by the Open Geospatial Consortium (OGC) for representing and exchanging semantic 3D city models. This standard defines the geometrical, topological, semantic, and visual aspects of the 3D city models in different levels of detail (LODs) (Gröger et al. 2012) as shown in Figure 10 for the case of the building model. CityGML is designed as an Extensible Markup Language (XML)-based format and as an implementation of the application schema of Geography Markup Language 3 (GML3).

Figure 10. The five LODs of CityGML version 2.0 (Biljecki et al. 2016)

CityGML serves as a central information hub where the semantic 3D city model has been used by many disciplines to attach their specific information for different purposes as shown in Figure 11 (Kolbe 2009). Solar potential analysis and energy consumption analysis used the semantic 3D city model with CityGML format as part of the analysis and also at the end enrich the model by the result of the analysis (Chaturvedi and Kolbe 2017; Carrión et al. 2010).
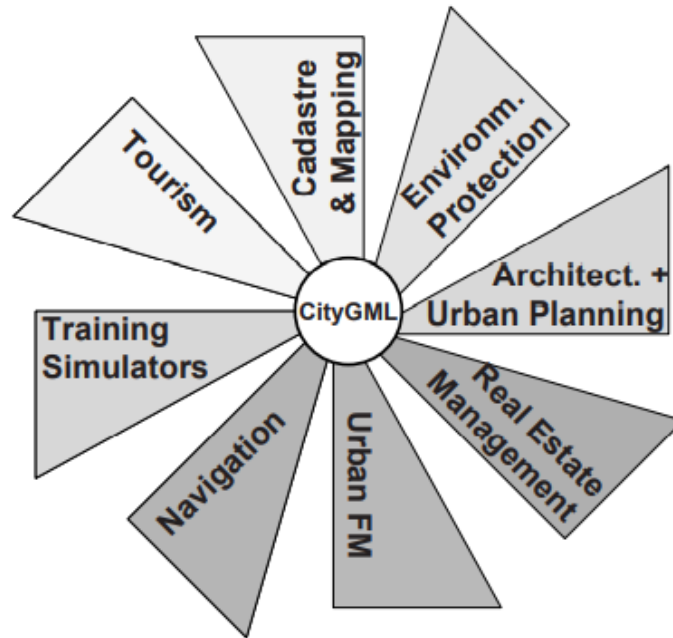
Figure 11. CityGML as information hub (Kolbe 2009)

CityGML comprises one core module and 12 thematic modules. The core module defines the basic elements of the CityGML data model. This consists of abstract base classes, from which the thematic classes in the thematic modules inherit. The 12 thematic modules are as follows (Gröger et al. 2012):

- **Appearance**: representation of the appearances of CityGML features.
- **Bridge**: representation of bridges in four levels of detail (LOD 1-4).
- **Building**: representation of buildings in five levels of detail (LOD 0-4).
- **CityFurniture**: representation of city furniture objects (immovable) like lanterns and traffic signs.
- **CityObjectGroup**: providing a grouping concept of CityGML objects.
- **Generics**: providing generic extensions for additional attributes and features which are not defined by the thematic classes of CityGML by introducing generic objects or generic attributes.
- **LandUse**: representation of land use and land cover.
- **Relief**: representation of the terrain.
- **Transportation**: representation of the transportation features.
- **Tunnel**: representation of tunnels in four level of detail (LOD 1-4).
- **Vegetation**: representation of vegetation objects.
- **WaterBody**: representation of rivers, canals, lakes, and basins.

- **TexturedSurface [deprecated]**: representation of visual appearance properties. However, this module is expected to be removed in future CityGML versions and can be converted into CityGML's *Appearance* module.

All modules including the core are defined by their own XML schema definition file (XSD) with a globally unique XML namespace for every module. Figure 12 shows the dependency relations of the CityGML core and extension modules.
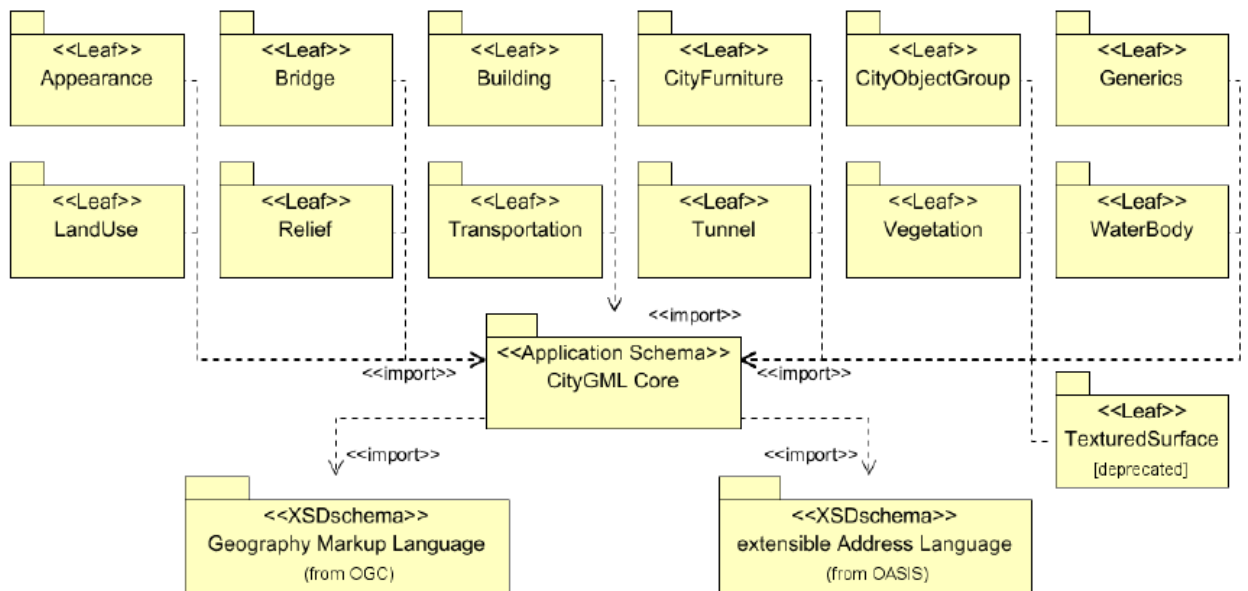
Figure 12. CityGML modules and their dependencies (Gröger et al. 2012)

The *Landuse* module is used to define areas of the earth's surface related to land use and land cover. Land use describes human activities on that specific land, while land cover describes its biological and physical cover. Every land use object can have the attributes of *class*, *function*, and *usage*. The *class* specifies the classification of the land use object e.g. settlement area, vegetation, etc. The possible values are defined in a code list. The *function* describes the nature or the purpose of the object e.g. residential, forest, etc. On the other hand, the *usage* is used when the way the object used is dissimilar from the function.

The *Generics* module allows extensions to the CityGML model that can be used to model and exchange 3D objects that are not covered by the predefined classes of CityGML. The generic extensions introduce two main concepts, i.e. generic city objects and generic attributes as shown in Figure 13. A *GenericCityObject* can have the attributes which are *class*, *function*, and *usage*. The *class* represents the object classification in the thematic area such as power line, the *function* describes to which thematic area the object belongs such as power supply, and the *usage* is used when the way the object used is dissimilar from the function. A generic attribute has *name* as an attribute which is mandatory to be filled and used as an identifier. The attribute *value* is associated with the generic attribute, which can be filled with different kind of data types.

Besides applying the *Generics* module, an extension to the CityGML data model can be realized by using Application Domain Extensions (ADEs). The extension can be in the form of addition of new properties to the existing CityGML modules or introduction of new object types. The main difference between generic extensions and the ADE approach is that a new XSD file and its own namespace have to be defined in the ADE. Besides defining on XSD, an ADE can also be specified using UML (Unified Modeling Language). Thus, the advantage of using ADEs is that the extension is formally described, which allows instance documents to be validated with the extended CityGML and the ADE schema.
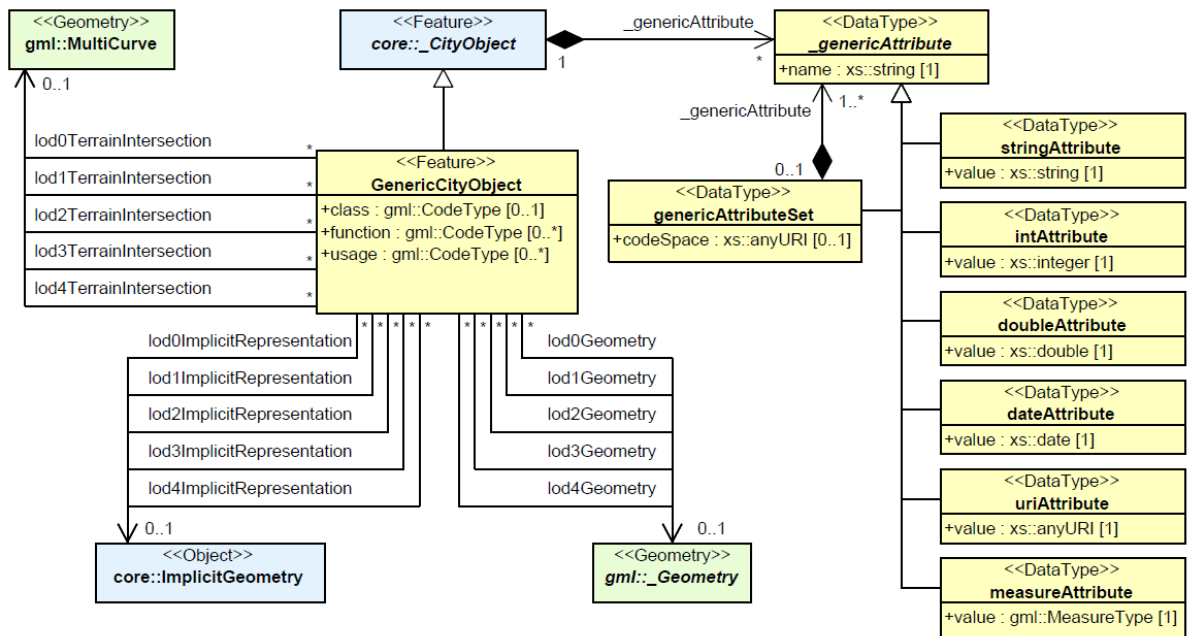
Figure 13. UML model of generic object and attribute (Gröger et al. 2012)

A new version 3.0 of CityGML was recently published (OGC 2021) introducing some new modules as shown in Figure 14. The new concept of CityGML 3.0 defines several new concepts including a space concept and the new modules *Dynamizer* and *Versioning* (Kutzner et al. 2020).

The space concept delineates spatial features into *spaces* and *space boundaries*. A space is a feature, which has a volumetric measure in the real world such as buildings, trees, and water bodies. On the other hand, a space boundary is a feature, which has an areal measure in the real world and connects as well as delimits the spaces. Examples for space boundaries are wall surfaces that bound buildings.

Furthermore, spaces can also be categorized into two forms, which are physical spaces and logical spaces. Physical spaces define spaces, which are fully or partially delimited by physical objects such as buildings and rooms. On the other hand, logical spaces are spaces bounded by thematic considerations such as a building unit, which is defined as an aggregation of rooms and delimited by a virtual boundary (Kutzner et al. 2020).
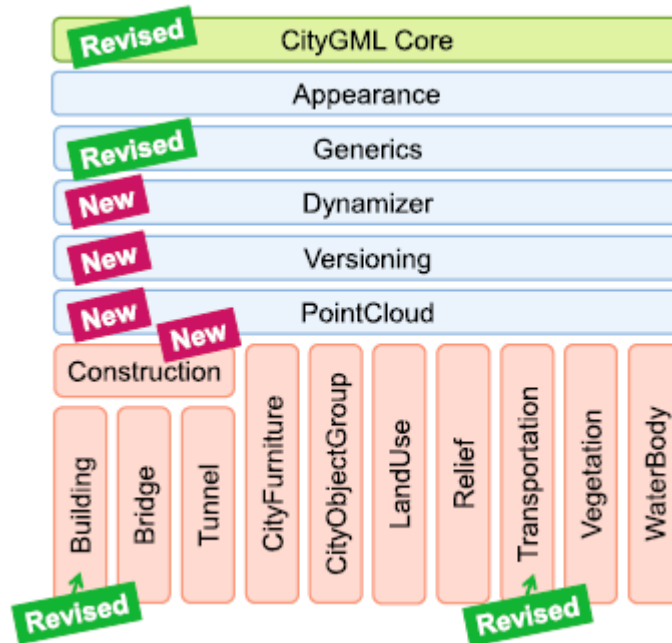
Figure 14. CityGML 3.0 module overview (Kutzner et al. 2020)

The *Versioning* and *Dynamizer* modules are introduced to represent the dynamics of change in the cities. A city object may change with respect to time. The change can be related to the geometry, semantics, appearance, and topology. A building that was renovated can have a modification in its geometry and appearance. Some sensors can also be attached to the building, which measure, for example temperature and humidity over time. The changing of the renovated building and the sensor values can be expressed with *Versioning* and *Dynamizer*.

The Versioning module handles qualitative changes that are considered slower in nature such as transformation of cities as well as their versions of city models. This module identifies the changes in a city object by specifying two identifiers which are *identifier* property and *feature:id* attribute. The *identifier* property is stable for the lifetime of the object. On the other hand, the *feature:id* attribute specifies the version of the object. The *Versioning* module also defines two feature types: *Version* and *VersionTransition*. *Version* is used to specify versions of the city objects and *VersionTransition* is applied to linking different versions of the city object by expressing the reason of modification and the changes applied to the model.

On the other hand, the *Dynamizer* module specifies the quantitative changes that represent high dynamic variations of city object properties such as thematic attributes (e.g. energy demands and solar irradiation levels), spatial properties (e.g. geometry change of a feature), and real-time sensor observations. The dynamic values in *Dynamizer* can be represented in three different classes:

- *AtomicTimeseries* class which tabulates the time/value pairs.
- *CompositeTimeseries* class which gives patterns of time/value pairs according to statistical rules.
- *SensorConnection* class which facilitates the retrieval of data from external sensor/IoT.

2.8 Dataset

The following datasets are used in this thesis:

- ***Deutsches Zentrum für Luft- und Raumfahrt* (DLR)**
  The DLR dataset is obtained from the work of Heldens et al. (2020). They generated raster data of Berlin for a microclimate simulation. The raster dataset includes rasters of building height (Figure 15), terrain height, vegetation, streets and bridges. The dataset provides several resolution ranges from 1 meter to 16 meter. For this thesis, rasters of building height with the resolution of 1 meter and 5 meter as well as raster of terrain height with the resolution of 1 meter are used to calculate the zone properties. The building raster data is derived from the 3D building model of CityGML LOD2, and the terrain height is derived from the LOD0 3D city model generated from airborne LiDAR (for the municipalities) and 30 m digital elevation model (DEM) of SRTM (for the surrounding areas).

Height (m)
6.1
10.3
17.9
140.7

Figure 15. Building height's raster data of DLR

- **OpenStreetMap (OSM)**
  OSM is a vector dataset containing primarily building (Figure 16), land use, road, and water features. OSM is an open source data generated by a community of mappers (OSM, 2021). For this thesis, building surface polygon of OSM is used to calculate the aspect ratio property of a local climate zone.

Figure 16. Building data of OSM

- **Copernicus**
  Copernicus is the European Earth monitoring system where data is 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) is provided by Copernicus (Copernicus, 2021). For this thesis IMD raster from 2015 with the resolution of 20 meter is used to calculate the impervious surface fraction of the LCZ's zone property (Figure 17).
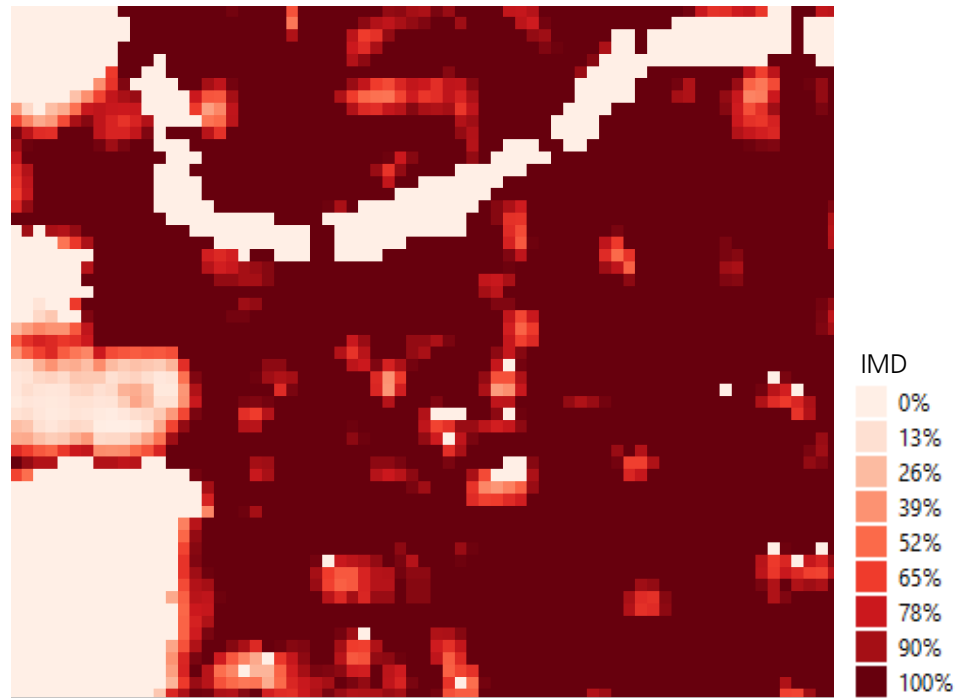
Figure 17. Imperviousness density 2015 of 20 meter resolution of Copernicus

- **WUDAPT L0**

  WUDAPT provides level 0 (L0) maps for many cities around the world. The Berlin L0 map is downloaded from the WUDAPT portal (WUDAPT Portal 2021) as it is shown in Figure 18. 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 meter and has been resampled from generally 30 meter Landsat 8 input image resolution. The training areas can also be downloaded through the WUDAPT website (WUDAPT TA 2021).
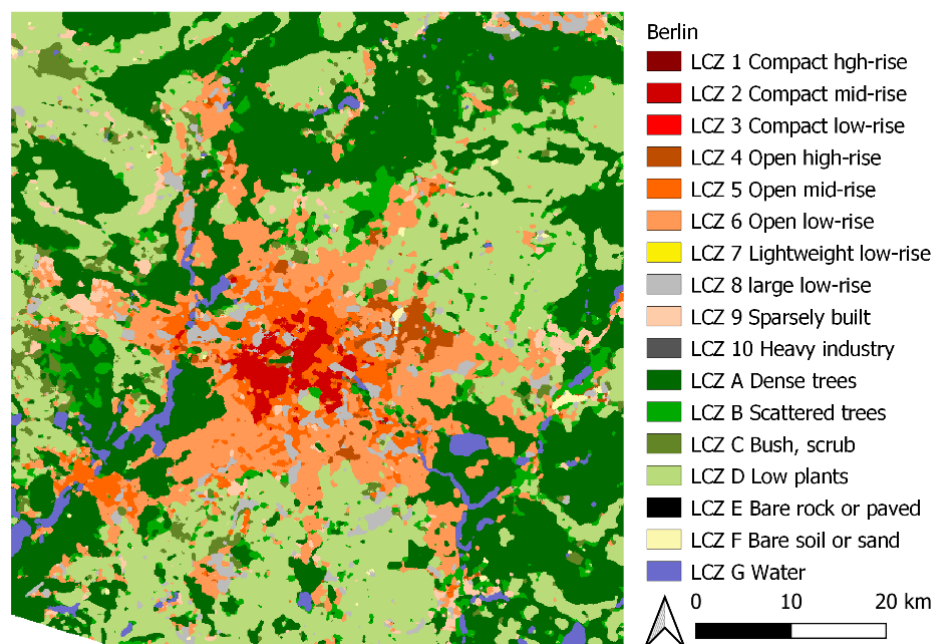
Figure 18. WUDAPT L0 of Berlin (Bechtel and Daneke 2012, Bechtel et al. 2015, Ching et al. 2017, Stewart and Oke 2012)

- **Measured land surface temperature (LST)**
  Measured land surface temperature is obtained from satellite imagery of MODIS Aqua with the resolution of around 700 meter. The data products used are MYD21A1D v006 and MYD21A1N v006 which correspond to day and night LST respectively (Hulley 2017). The Aqua satellite passes over every place on the earth at around 1:30 PM (day LST) and 1:30 AM (night LST).

  This dataset can be accessed or downloaded in various ways. For this thesis, a website application namely Application for Extracting and Exploring Analysis Ready Samples (AρρEEARS) is used to retrieve the LST raster data. This website is very convenient to use since you can define the area of interest, choose the period, select the LST data sources, and finally specify the format and the projection of the LST raster data (AρρEEARS 2021).

  The LST data retrieved is from 21st of June until 4th of July 2010. This period is chosen to fit the period of the available WRF simulation results, which will be used later for the evaluation/correlation of GIS-LCZ.

- **WRF simulation results**
  The WRF simulation was carried out over Berlin and surrounding area from 21st of June until 4th of July 2010 (Vogel and Afshari 2020). The result of the simulation is saved in netCDF format and contains data including urban fraction (UF) and urban heat island intensity (UHII) which are used in this thesis later for the correlation analysis.

2.9 Software

To support this thesis, QGIS and its plugins are used to do the calculation of the zone properties. QGIS is an open source software for Geographic Information System (GIS) which supports vector, raster, and database functionalities (QGIS 2021). The QGIS plugins relevant for this study, which are primarily used to support the data processing, are the Geospatial Data Abstraction Library (GDAL), the Geographic Resources Analysis Support System (GRASS), and the Urban Multi-scale Environmental Predictor (UMEP). GDAL is a geospatial data translator and processor library for raster and vector data formats. GRASS plugin allows access to databases and functionalities of GRASS GIS (GRASS QGIS 2021). UMEP is an open source climate service tool which combines models and tools for climate simulation (UMEP 2021). The UMEP pre-processor component (a tool to prepare input data for meteorological and surface information) is applied in this thesis (Lindberg et al. 2018).

Python libraries are also used to facilitate the calculation process of zone properties as well as the classification of the zone properties into local climate zones. After the GIS-LCZ is generated, an evaluation of the result is carried out. The evaluation comprises accuracy assessment and comparison between the GIS-LCZ map and WUDAPT dataset, which is done using SAGA. Another evaluation is correlation analysis between zone properties/LCZ classes and temperature values, which is calculated in a custom code written in Python. ArcGIS Pro is also used to retrieve the WRF simulation results for the correlation analysis. ArcGIS Pro is similar to QGIS but it is a commercial software (ArcGIS Pro 2021). Finally, Enterprise Architect is used to generate the UML model for the

analysis of enrichment of CityGML. This software is a modelling platform, which can be used for the analysis, design, implementation, assessment, and maintenance of models using UML, SysML and other open standards (EA 2021).

2.10 Related works

Lelovics et al. (2014) with its related work (Unger et al. 2014) is considered as the first research deriving local climate zones with the GIS-based method (Quan and Bansal (2021). The study is carried out in Szeged, Hungary applying lot area polygon as the BSU. They apply seven zone properties: sky view factor, building height, surface roughness, surface albedo, and the three surface fractions (building, impervious, and pervious surface fractions). These zone properties are classified using fuzzy logic. The classification result is further processed by merging the BSUs. The resulting map has six urban classes of GIS-LCZ. The map is used to plan the location of urban temperature measurement network. Other related studies use this map for the integration with WUDAPT L0 map (Gál et al. 2015) and for monitoring air temperature (Skarbit et al. 2017).

A GIS-LCZ map is also generated for Vienna, Austria by Hammeberg et al. (2018). This work applies 100 m x 100 m grid tile as the BSU and employs five zone properties: building height, aspect ratio and the three surface fractions. The classification of the zone properties utilizes Naive Bayes classifier. There is no post-processing done in this study. The evaluation for the classification result is carried out by comparing it to the WUDAPT L0 map.

Wang et al. (2018) derives the GIS-LCZ map for Hong Kong applying 100 m x 100 m grid tile as the BSU. They use three zone properties and 1 additional land use data. The zone properties, which are building height, building surface fraction, and sky view factor, are employed to classify the urban classes (LCZ 1 – 10). The additional land use data is used for the classification of the natural classes (LCZ A – G). The classification is done by the modification of the standard rule proposed by Stewart and Oke (2012). An accuracy assessment is carried out to the resulting GIS-LCZ map by comparing it to the established validation samples.

Another GIS-LCZ study is conducted by Estacio et al. 2019 in Quezon City, Philippines. The study employs seven zone properties: sky view factor, building height, roughness length, surface albedo and the three surface fractions. The classification is done by employing fuzzy logic by modifying trapezoidal membership function from Lelovics et al. (2014). The result of the classification is aggregated by using cellular automata to derive the LCZ map. The map is validated with the expert knowledge. The land surface temperature profile for each LCZ type is also assessed in this study.

The study that is considered as the latest GIS-LCZ study is the work of Chen et al. (2020) in Chenzou, China. The classification is done over 200 m x 200 m grid tiles with seven zone properties and one additional property: building height, the three surface fractions, sky view factor, aspect ratio, terrain roughness class, and the additional vegetation coverage ratio. The classification is carried out by modifying the standard

rule proposed by the LCZ framework. The result of the classification is compared to the test samples. Moreover, the result is also analyzed with the land surface temperature.

In this thesis, GIS-LCZ method is applied to the city of Berlin, Germany by dividing the city area into 100 m x 100 m grid tiles. Five zone properties are used: sky view factor, building height, aspect ratio, building surface fraction and impervious surface fraction. The classification of the zone properties employs fuzzy logic by modifying the rule applied in Estacio et al. (2019). The post-processing applied to the result of the classification is majority filter as applied by the WUDAPT L0 approach. The resulting GIS-LCZ map is compared to the training areas from WUDAPT for the accuracy assessment. The LCZ map from WUDAPT, which is WUDAPT L0, is also compared to the GIS-LCZ map to see the agreement between both of the maps.

The result of GIS-LCZ map can be further used to enhance the 3D city model of CityGML. The calculated zone properties and the LCZ class of the grid tiles can be represented with the statistical grid as done by the government of Japan. They introduce statistical grid module in the ADE of i-Urban Revitalization. This module specifies the statistical grid of thematic values of the city objects, which is intended for time series analysis and regional comparison (Ishimaru et al. 2020). Figure 19 illustrated statistical grids that represent the number and age of building in Munakata City, Japan.

Figure 19 Aggregated statistical grids (Height: Number of buildings, Color: Average of year of construction) (Ishimaru et al. 2020)

3   Methodology

The GIS-LCZ method employed in this thesis comprises the processing steps shown in Figure 20. Vector and raster datasets are used to calculate the zone properties needed to derive the LCZ map. Building, terrain, and land cover or land use datasets are obtained from different sources and formats. These datasets are georeferenced to the same coordinate system to calculate the zone properties.

The calculation of the zone properties is done with the aid of Python and QGIS applying plugins such as GDAL, GRASS, and UMEP. The calculated zone properties are classified applying Fuzzy logic to generate the LCZ map. A post classification is carried out by applying a majority filter to the classification result. The accuracy of the resulting map is calculated with the training areas provided from WUDAPT. Moreover, the comparison of the GIS-LCZ map with the WUDAPT L0 map is also evaluated.

The LCZ map derived from the GIS dataset is correlated with the surface UHI intensity (SUHII) which is calculated from the land surface temperature acquired from satellite imagery. The correlation is also carried out with the simulated near-surface air UHI intensity (UHII) resulting from the WRF simulation. Finally, the possibility of integrating the result of GIS-LCZ map to enrich the CityGML is also analyzed.
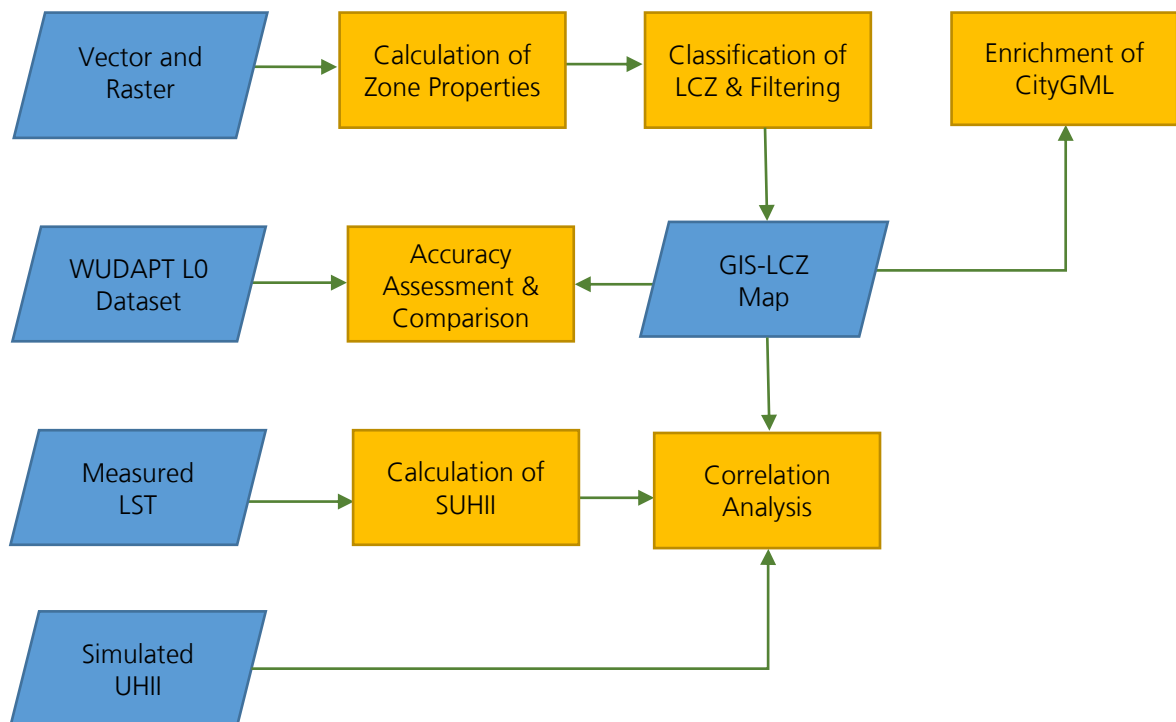

Figure 20.   Methodology

3.1 Study area

The study area of this thesis is located in Berlin with the border of the city shown in Figure 21. Berlin is the capital city of Germany with an area of 892 km². As of 2019, its population is around 3.8 million, which is the highest number among other cities in Germany. Berlin is chosen for this study because, contrary to most other German cities, the GIS data is abundantly available for deriving the GIS-LCZ map. DLR dataset, OSM, WUDAPT L0 map, and Copernicus dataset are also available for Berlin. Previous work related to WRF simulation from the Urban Physics Group of Fraunhofer Institute for Building Physics have also been conducted in Berlin. The results of this work will be used in this thesis for the validation/correlation purposes. To the best of the author's knowledge, Berlin is the first city in Germany to be classified in local climate zones with a GIS-based approach.
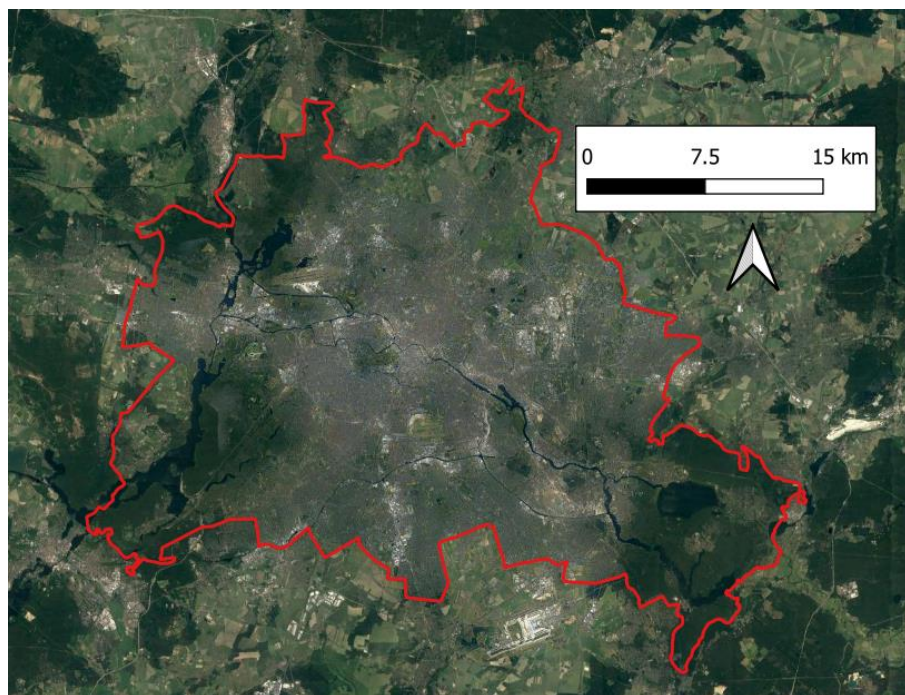
Figure 21. Berlin area

3.2 Calculation of zone properties

As explained in section 2.2, there are ideally 10 zone properties defining the 17 local climate zones which are sky view factor, aspect ratio, building surface fraction, impervious surface fraction, pervious surface fraction, height of roughness element, terrain roughness class, surface admittance, surface albedo, and anthropogenic heat output. However, due to limited data sources, in practice only a subset of those properties can be used for the classification of the LCZs. Chen et al. (2020) used seven of them excluding surface admittance, surface albedo, and anthropogenic heat flux for the city of Chenzou in China. Hammerberg et al. (2018) used five of them, which are average building height weighted by area, building surface fraction, impervious surface fraction, pervious surface fraction, and height to width ratio. Wang et al. (2018) only used three of them, which are building height, building surface fraction, and sky view factor for the city of Hong Kong.

For this thesis, 5 zone properties will be calculated to generate the LCZ map. Pervious surface fraction, terrain roughness class, surface admittance, surface albedo, and anthropogenic heat flux are excluded from this calculation, because the data are not easily obtained.

The basic spatial unit for the classification is in the form of grid tiles with the size of 100 m x 100 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 number of grid tiles in total for the area of Berlin is 90517 as shown in Figure 22. The extent of the grids is based on WUDAPT's Berlin L0 grid, which also has 100 m spatial resolution. The reason of choosing WUDAPT's Berlin L0 map as the reference is that after the classification of GIS-LCZ, the resulted map is compared with WUDAPT's Berlin L0 map. The comparison is done tile by tile. Thus, for later convenience, the WUDAPT's Berlin L0 map is chosen.
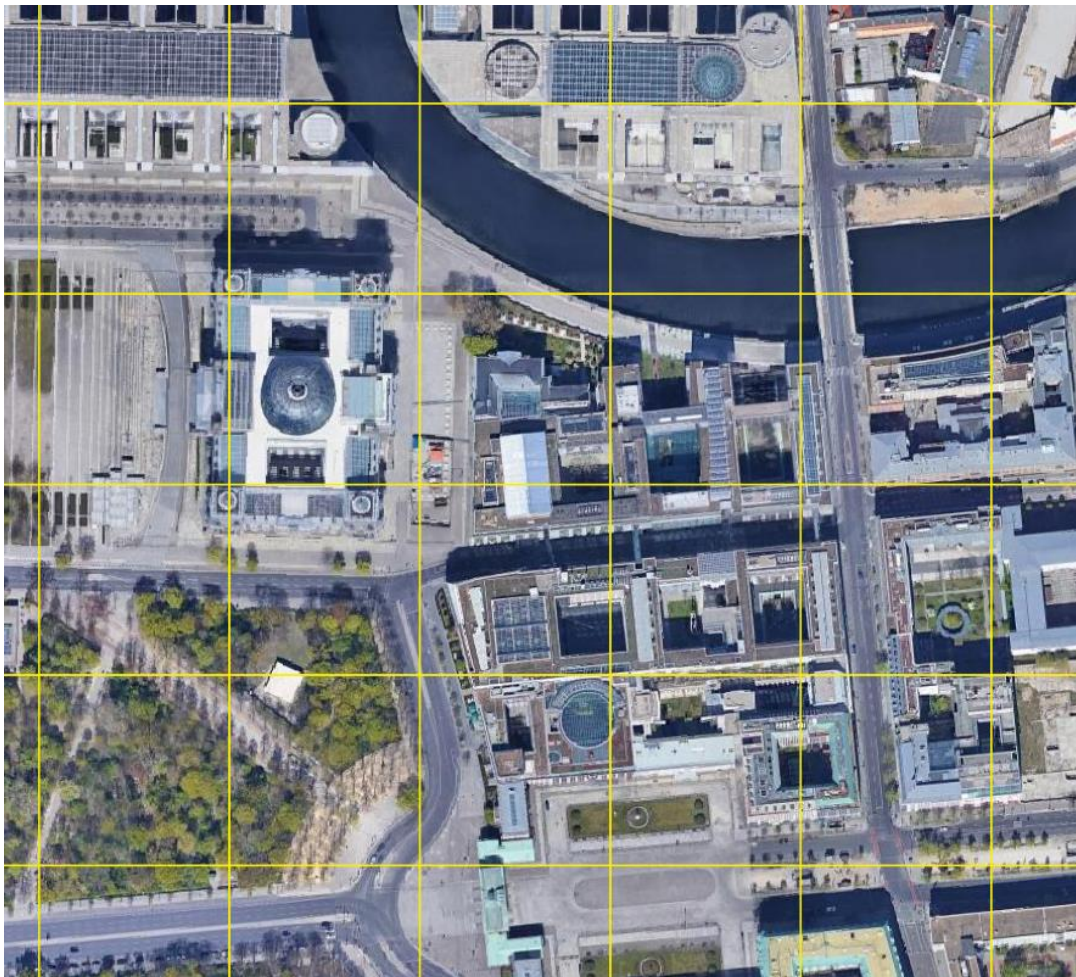

Figure 22. Grid tiles of GIS-LCZ for berlin (example)

The calculation of the 6 zone properties is explained as follows:

- **Sky View Factor (SVF)**
  The SVF is calculated with the Urban Geometry-Sky View Factor Calculator of the UMEP plugin in QGIS. The algorithm used to calculate the SVF is described in Lindberg and Grimmond (2010). For the calculation of the SVF, rasters of building and terrain heights are needed. Before starting the calculation, a conversion of null values (no building height information) to zero is done for DLR raster of building height with the resolution of 5 meter which has netCDF file format. This conversion is done using GRASS library namely *r.null* in QGIS and the result is saved in GeoTIFF format. The GeoTIFF file is then reprojected to EPSG 25833. The conversion to zero is done for convenience later in the process of addition with the raster of terrain height.

  The terrain height's raster is also processed by applying resampling to the raster of terrain height from 1 m to 5 m resolution. The resampling is done due to consideration of computation time that might take a longer time for a resolution of 1 meter. The result of SVF calculation with the 5 meter resolution is considered enough for the classification as it is later also averaged over 100 x 100 m grid tile. The resampling applies GRASS library namely *r.resample* in QGIS. The resampling method applied in this library is the method of nearest neighbors. The result is saved in GeoTIFF format. The conversion of null values to zero is also done. However, the result shows no null values in the raster of terrain height. The reprojection is not necessary since the original netCDF file of the terrain height's raster is already in EPSG 25833.

  A raster of Berlin city boundary is also applied to both rasters of the building and terrain heights. The boundary raster has zero pixel values inside the city boundary and no data/null values outside the city boundary. Thus, when both rasters of building and terrain heights are added with the boundary raster, the boundary sets all the pixel values outside the city to null values. The addition of these three rasters (boundary, building height, and terrain height) is done in *Raster Calculator* in QGIS. The result is shown in Figure 23. This raster is then used as input in the UMEP's SVF calculator. The resulting SVF of the calculation is in the form of raster with the same resolution as the source data which is 5 m.
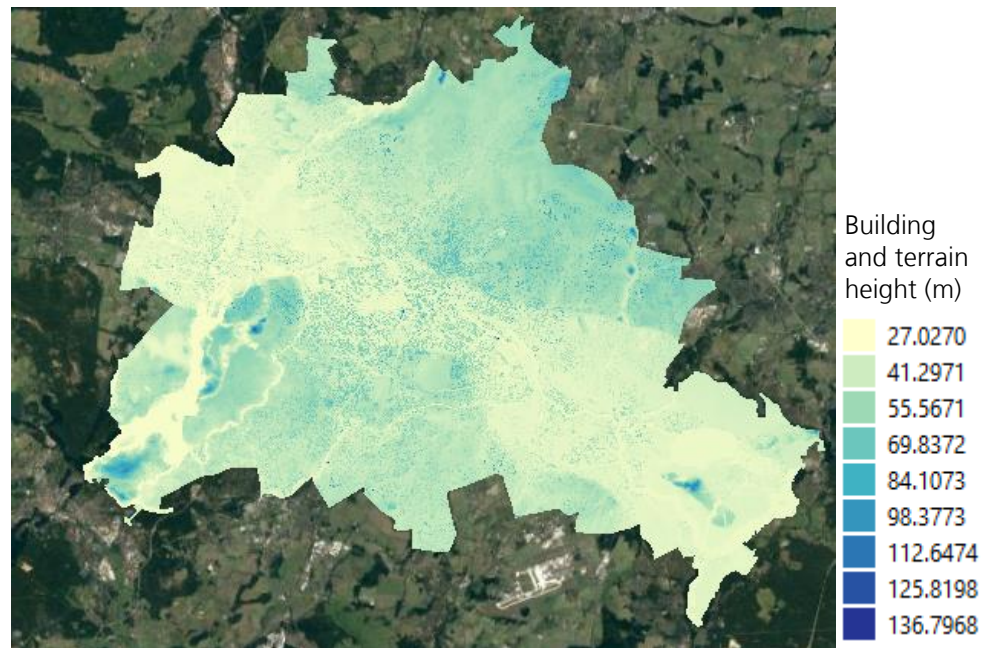
Figure 23. Raster of building and terrain height

For the LCZ classification purpose, only the SVF value from the ground-level is needed therefore the generated SVF raster is extracted with the value only from the terrain and the value from the building is set to no data. This is done by preparing a 0-1-building raster by converting the pixel values in building height's raster to the value of zero for the pixels with building height information and value of one for pixels with no building. This process is carried out in the *Raster Calculator*. The resulting raster is then multiplied with the SVF raster and yields an SVF raster with zero pixel values for building footprints. These zero pixel values are then translated to null values or no data applying *Translate* (Convert Format) tool from GDAL in QGIS. The resulting raster (Figure 24) is then resampled to the grid tile of 100 m x 100 m for calculation of zone property of SVF. The resampling is carried out by calculating the *mean* statistics using *Zonal Statistics* tool in QGIS. The mean value from this calculation result is assigned as the SVF value for the respective grid tile.

When the original SVF raster is evaluated, there is an anomaly in the SVF pixel values at the border of Berlin. The outside pixel values that are near to the border should have null values for the SVF since the input raster data has null pixel values for the area outside the city boundary. However, these tiles have values so that the calculated mean SVF values for the tiles in the border are not correct. Therefore, the SVF values for these tiles are converted into no data. The SVF values in the forest area are also not correct, since the vegetation height is not available for the calculation of the SVF. Thus, the forest areas have SVF values of around 1. This is one of the limitations of this thesis: data about plants or tree heights is not available.
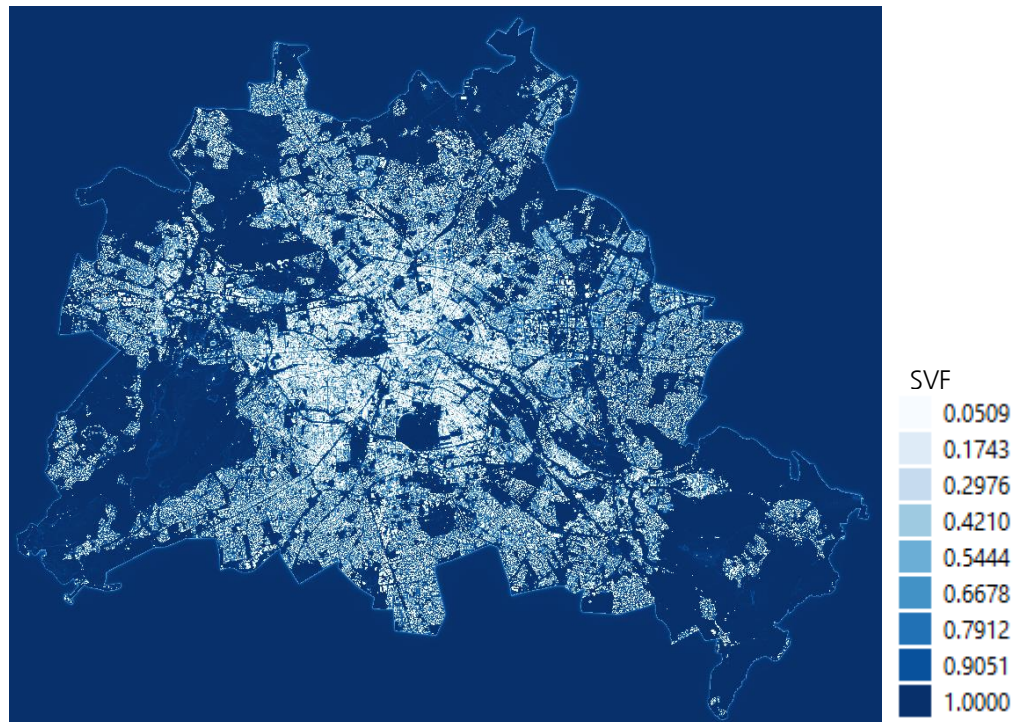
Figure 24. SVF

- **Building and impervious surface fractions (BSF and ISF)**
  BSF is defined as the percentage of the area covered by buildings in a 100 m x 100 m grid tile. It is calculated from the building footprint from the DLR building height's raster with the resolution of 1 meter. The raster is reprojected to EPSG 25833 and further processed with *Zonal Statistic* tool in QGIS to get the number of pixels that contain buildings in a specific grid tile. The number of pixels (NOP) can be derived from *count* calculation in the tool. The *count* value is divided by 100 to get the BSF percentage in a grid tile. The resulting BSF is shown in Figure 25.
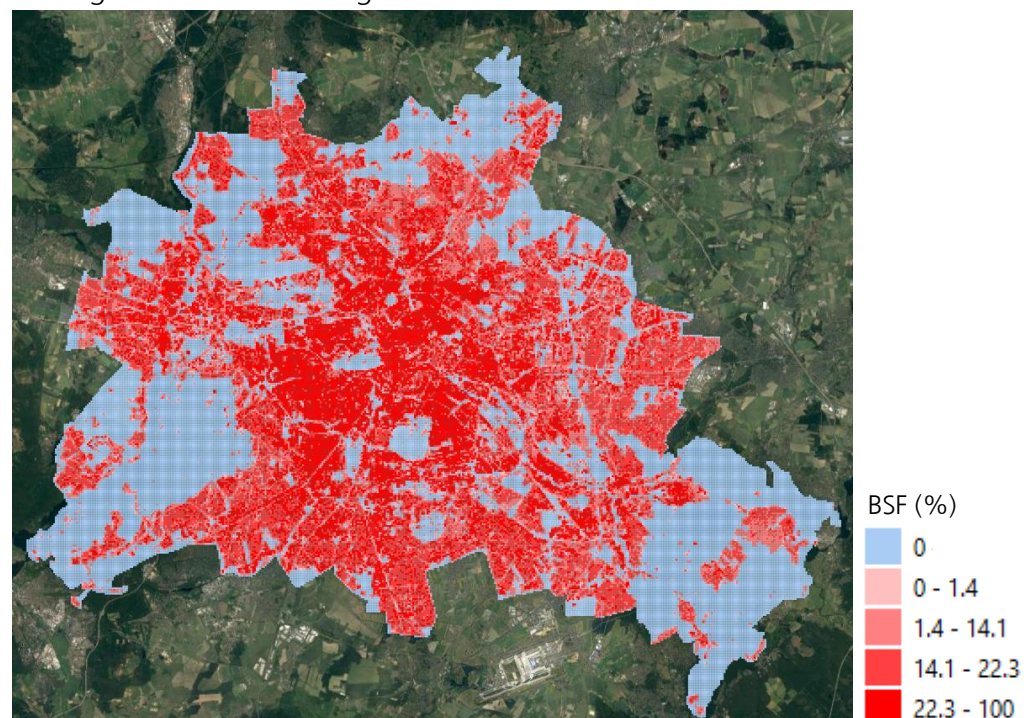
Figure 25. BSF

The ISF is the percentage of the area covered by impervious (paved or rock) materials in a 100 m x 100 m grid tile. Copernicus Impervious Density (IMD) raster is used for the calculation of ISF by calculating the *mean* of the 20 meter resolution raster data over 100 m x 100 m grid tile applying *Zonal Statistic* tool. However, the IMD cannot be directly used to represent the ISF needed by LCZs framework, since the IMD also includes building information. ISF as a zone property in the LCZs excludes the information of buildings since it is already covered by BSF. Thus, BSF should be subtracted from 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. It can be due to the fact that 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 the negative values of 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 corrected IMD will have more information about imperviousness in the study area since it gets additional information from the BSF. Finally, the ISF is defined as corrected IMD with the subtraction of BSF. Figure 26 shows the final ISF map.

Figure 26. ISF

- **Aspect ratio (H/W)**

The width of the street canyons or building spacing (w) is estimated by the equation modified from Samsonov and Varentsov (2020). The original equation is as follows:

$$w = [\sqrt{S_0}(\sqrt{\lambda_{urb\,0}} - \sqrt{\lambda_{OSM\,BLD}} \,] / \sqrt{N_{OSM\,BLD}} \quad (1)$$
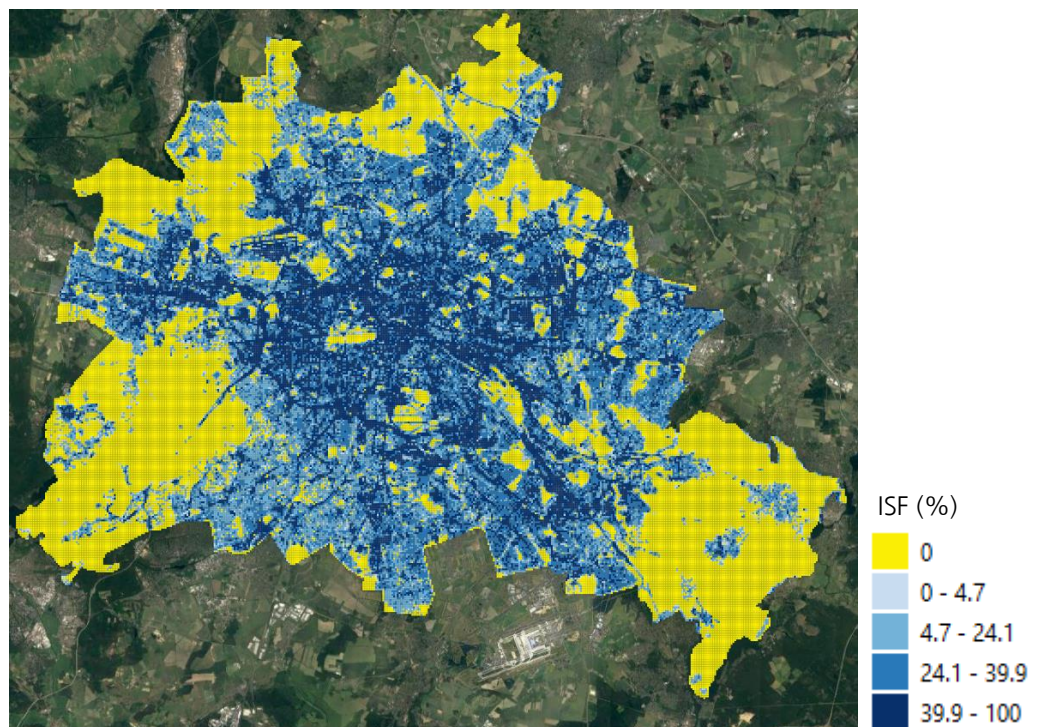
where S_0 is the grid tile area which, in our case, is 10.000 m², $\lambda_{urb\,0}$ is the urban surface fraction, $\lambda_{OSM\,BLD}$ is the building surface fraction obtained from OSM, and N_{BLD} is the number of buildings. This formula assumes that buildings in a grid tile are located regularly within its urban area and have the same size and also are square.

In this thesis, $\lambda_{urb\,0}$, modified to include all the surface fractions (urban/built-up and natural) in a grid tile, is set to 1 as shown in equation 2. The inclusion of all surface fractions is aimed to obtain the overall width of building spacing in LCZ 8-10. $\lambda_{OSM\,BLD}$ is also modified into λ_{BLD} to include the building surface fraction from DLR that was calculated before and defined as BSF.

$$w = \left[\sqrt{S_0}(1 - \sqrt{\lambda_{BLD}})\right] / \sqrt{N_{BLD}} \quad (2)$$

λ_{BLD} is then formulated as max($\lambda_{OSM\,BLD}$, BSF). N_{BLD} is obtained from building data of OSM and DLR. Building data from OSM is in the vector format and it is easier to be processed to get the number of buildings by applying methods in QGIS and Python. However, not all building data are detected by OSM, there are around 1500 grid tiles that have no building data in OSM but include buildings in the DLR raster data. Thus for these tiles, building data from DLR is used for the calculation of N_{BLD}. For the calculation, the DLR data is then clipped by applying a mask layer of grid tiles that do not have building information from OSM. This process is done using GDAL library of *Clip Raster by Mask Layer* in QGIS. The resulting masked raster is then converted into vector data in shapefile format.

For calculating the N_{BLD}, both OSM and DLR vector data are dissolved in QGIS. It is done by applying GDAL library of *Dissolve* in QGIS. For the building vector data from OSM, it is needed to be dissolved since buildings that are attached to each other have different identification numbers and they are counted as different features in the shapefile (left image in Figure 27). This condition is not desired in the calculation of N_{BLD}. Even though the building belongs to different persons, but if they are attached to each other (sharing the same walls), the buildings are considered as one feature for the calculation of N_{BLD}. For the vector DLR data, the building is also necessary to be dissolved since the data is in the form of grid tiles of 1 m x 1 m (left image in Figure 28) which is the original resolution of the building raster data from DLR.
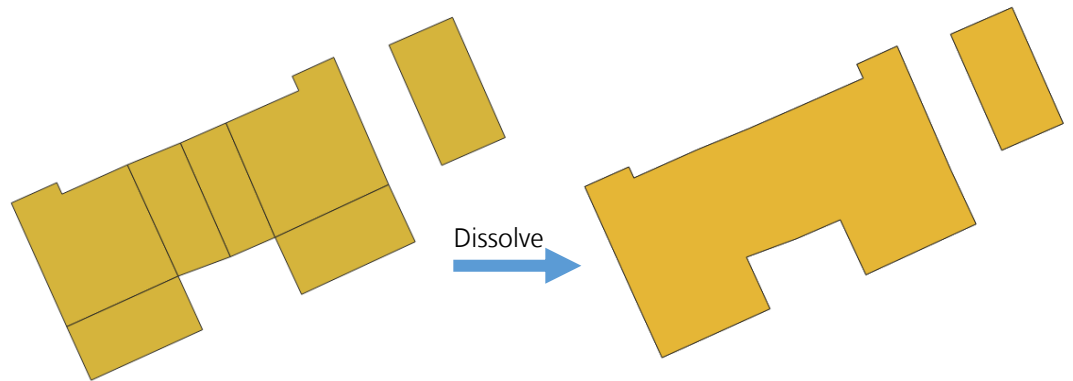
Figure 27. Dissolved OSM building data
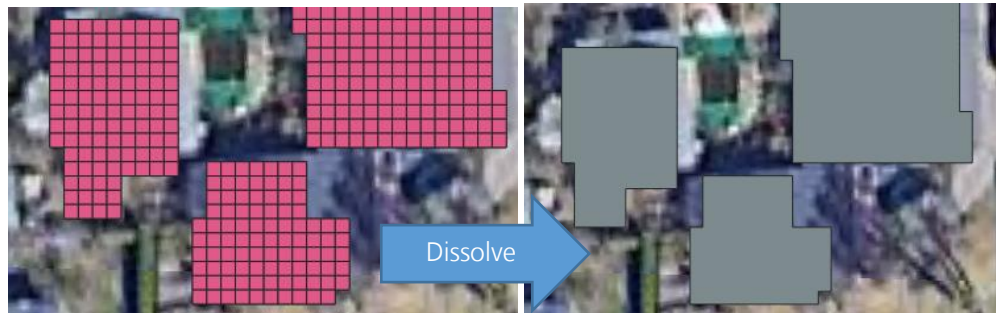
Figure 28. Dissolved DLR building data

After the dissolution is completed, the vector building data, either from OSM or DLR, is separated into single features since the dissolved result is in the form of merged buildings in one feature. The separation is done using *Multipart to singleparts* tool in QGIS. After the separation into single features, the building features are then intersected with the grid tiles of GIS-LCZ. The intersection is carried out by applying *Intersection* tool in QGIS. The intersected building features are further processed in Python with Geopandas library for calculating the number of buildings in each grid tile.

The resulting N_{BLD} together with other variables in Equation 2 is then used to calculate the mean width/space between buildings (*w*). After that, the aspect ratio can be calculated which is the ratio between mean building height (H) and mean width/space between building (*w*). The result is shown in Figure 29.
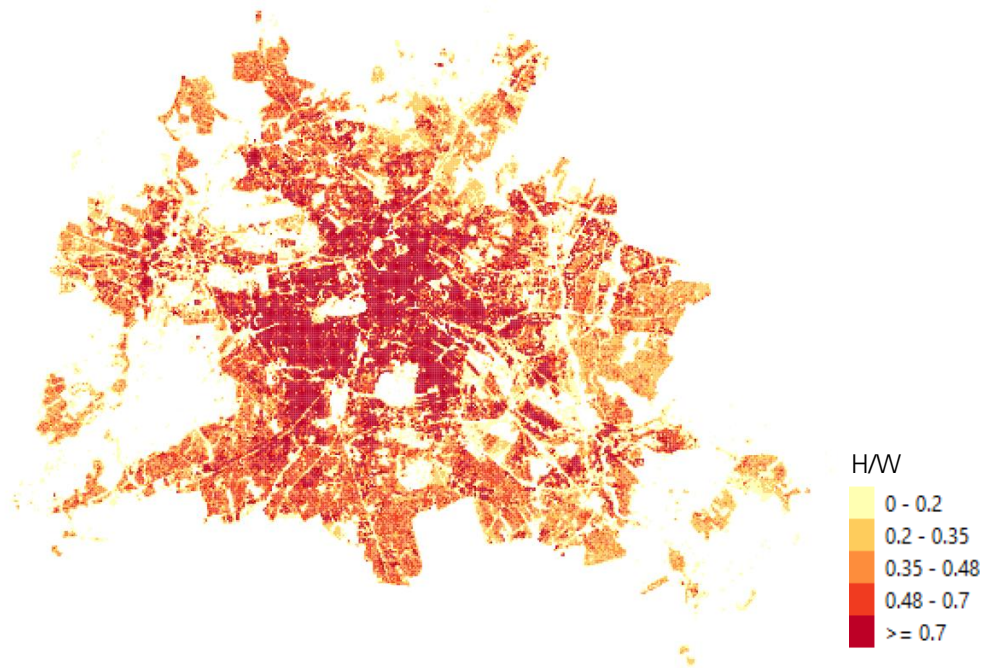
Figure 29. Aspect ratio (H/W)

- **Mean building height (H)**
  The DLR raster of building height with the resolution of 1 m as used for the calculation of BSF is also used to calculate the mean building height (H) in the grid tile. The calculation is done per grid tile with the *Zonal Statistic* tool by taking the *mean* value. The resulting H is shown in Figure 30.

Figure 30. Mean building height

After all the properties are calculated, all the grid tiles over Berlin have zone property values as exemplified in Figure 31. Every grid tile has an identification number (ID) that distinguishes the grid tile from the other grid tiles. These zone properties are further processed to classify the tiles into local climate zones.

| ID	81531
SVF	0.47
H	20.99 m
H/W	1.5
BSF	43.22 %
ISF	52.54 %

Figure 31. Zone properties of a grid tile

3.3 Classification of GIS-LCZ

The zone properties used in the classification of GIS-LCZ in this thesis are simplified into 12 classes instead of 17 classes. We retain the original 10 urban classes (LCZ 1 – LCZ 10); however, the seven natural classes (LCZ A – LCZ G) are simplified into two new classes: LCZ 11 and LCZ 15. LCZ A – LCZ D and LCZ F – LCZ G are combined into LCZ11 since they all clearly belong to natural classes. LCZ E is separated from the other natural classes since it could have a paved surface property which is manmade and cannot be systematically categorized as natural class. Moreover, WRF framework also recognizes LCZ E as an urban class. For the convenience in the classification process, LCZ E is named as LCZ 15. The simplification of these 17 classes is done because the classification of natural classes is not necessarily needed for the WRF simulation. On the other hand, we do not have sufficient data to calculate the zone properties in natural classes. The zone properties such as SVF and H/W require information about tree/plant height, which is not easily obtained.

The calculated zone properties are classified into LCZ classes by applying fuzzy logic with a trapezoidal membership function modified from Estacio et al. (2019). The membership of every zone property for every LCZ class is determined as shown in Figure 32 (example case of LCZ 1 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. The range

values for the zone properties (based on the framework of Stewart and Oke (2012)) used in the classification of the GIS-LCZ map is shown in Table 2.
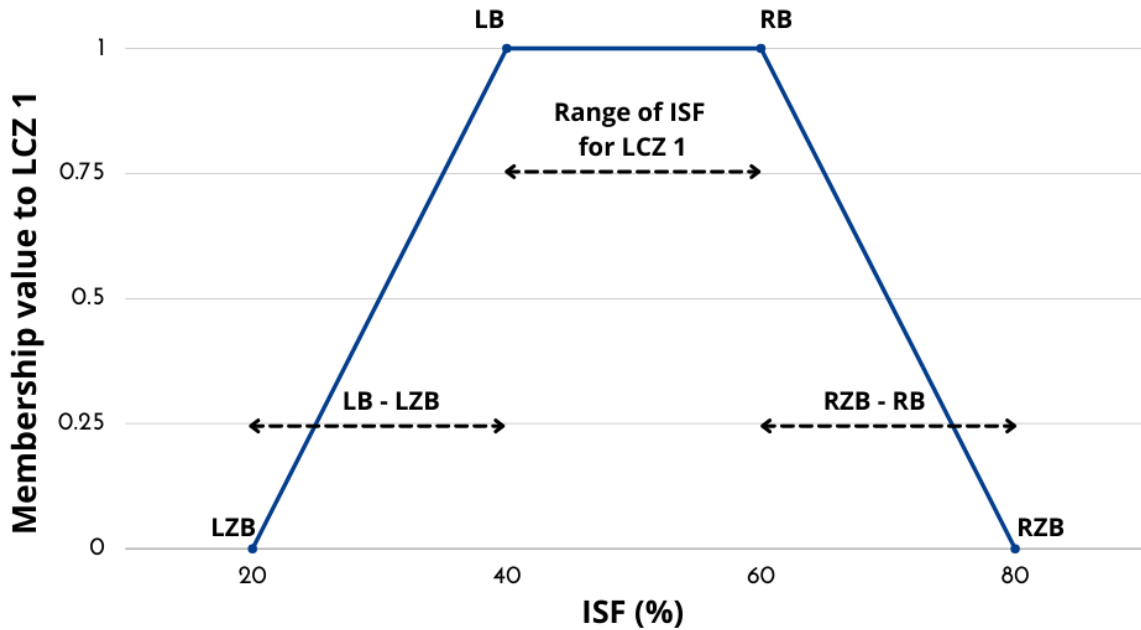

Figure 32. Trapezoidal membership function of fuzzy logic for ISF of LCZ 1 (adopted from Estacio et al. (2019))

To understand how this membership function works, an example from Figure 31 is explained here. The zone property of the ISF of LCZ 1 has a range value from 40 – 60, which implies 40 as the left bound (LB), 60 as the right bound (RB) and length (L) = RB-LB = 20. When a grid tile has a value of the ISF, which is in this range (40-60), the membership value will be 1. On the other hand, when a grid tile has 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 will become 0 for the ISF value of 20 and 80. 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 = 20 and RZB = RB – L = 80.

A problem arose for the zone properties that do not imply the RB, for example, the property value of H/W for LCZ 1, which is bigger than 2. 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 infinity, as it will overestimate the membership value of the zone property, which is less than LB. For this case, Estacio et al. (2019) chose 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 2, because the H/W will get membership values from all the urban classes of LCZs which ranges, from 0 (LZB) to 2 (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 HW's range values defined in Table 2, which is 1.5. This value will constrain the H/W membership to two LCZ classes: LCZ 1 and 2.

Table 2. Range values of zone properties for the classification of the GIS-LCZ map

LCZ	SVF	H/W	BSF (%)	ISF (%)	H (m)
1	0.2 – 0.4	>2	40 – 60	40 – 60	>25
2	0.3 – 0.6	0.75 – 2	40 – 70	30 – 50	10 – 25
3	0.2 – 0.6	0.75 – 1.5	40 – 70	20 – 50	3 – 10
4	0.5 – 0.7	0.75 – 1.25	20 – 40	30 – 40	>25
5	0.5 – 0.8	0.3 – 0.75	20 – 40	30 – 50	10 – 25
6	0.6 – 0.9	0.3 – 0.75	20 – 40	20 – 50	3 – 10
7	0.2 – 0.5	1 – 2	60 – 90	< 20	2 – 4
8	>0.7	0.1 – 0.3	30 – 50	40 – 50	3 – 10
9	>0.8	0.1 – 0.25	10 – 20	< 20	3 – 10
10	0.6-0.9	0.2 – 0.5	20 – 30	20 – 40	5 – 15
11			< 10	< 10	
15			< 10	>90	

The same approach is also used for the zone property H of LCZ1 and LCZ4, which has the range value of >25. The LB is then defined as the second highest RB value of the range values of H from Table 2, which is 15 m. The other zone properties, which do not define LB or RB, are SVF, BSF, and ISF. The common value is chosen as the LB or RB. For SVF, 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.

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, there are 12 total membership values from the 12 GIS-LCZ classes. These membership values are normalized by the number of zone properties available for the respective grid tile. From these 12 total membership values, the maximum value is chosen and assigned as the LCZ class of the grid tile. For an example, Table 3 shows the membership values calculated for every zone property for the grid tile shown in Figure 31. From Table 3, it implies that the final LCZ class assigned to the grid tile with the ID of 81531 is LCZ 2 as it has the highest normalized membership value.

Table 3. Membership values (example)

Zone Properties	Membership value for each LCZ											
ID: 81531	1	2	3	4	5	6	7	8	9	10	11	15
SVF: 0.47	0.65	1	1	0.85	0.9	0.56	1	0.23	0	0.56	-	-
H/W: 1.5	0.1	1	0.94	0.4	0	0	1	0	0	0	-	-
H: 20.99	0.6	1	0	0.6	1	0	0	0	0	0.4	-	-
BSF: 43.22%	1	1	1	0.84	0.84	0.84	0.44	1	0	0	0	0
ISF: 52.54%	1	0,873	0.92	0	0.87	0.92	0	0.75	0	0.3	0	0
Total	3.35	4.87	3.86	2.69	3.61	2.32	2.44	1.98	0	1.33	0	0
Normalized	0.67	0.97	0.77	0.54	0.72	0.46	0.49	0.4	0	0.27	0	0

The Python libraries used for the classification are *numpy*, *skfuzzy.membership*, *copy, pandas*, and *geopandas*. The classification result is further processed by applying a post-processing step namely filtering. The filter chosen is a majority filter and is applied using SAGA with the filter radius of two pixels and the search mode of square. These parameters yield a window filter of 5 x 5 pixels. The final result after filtering is shown in Figure 32.
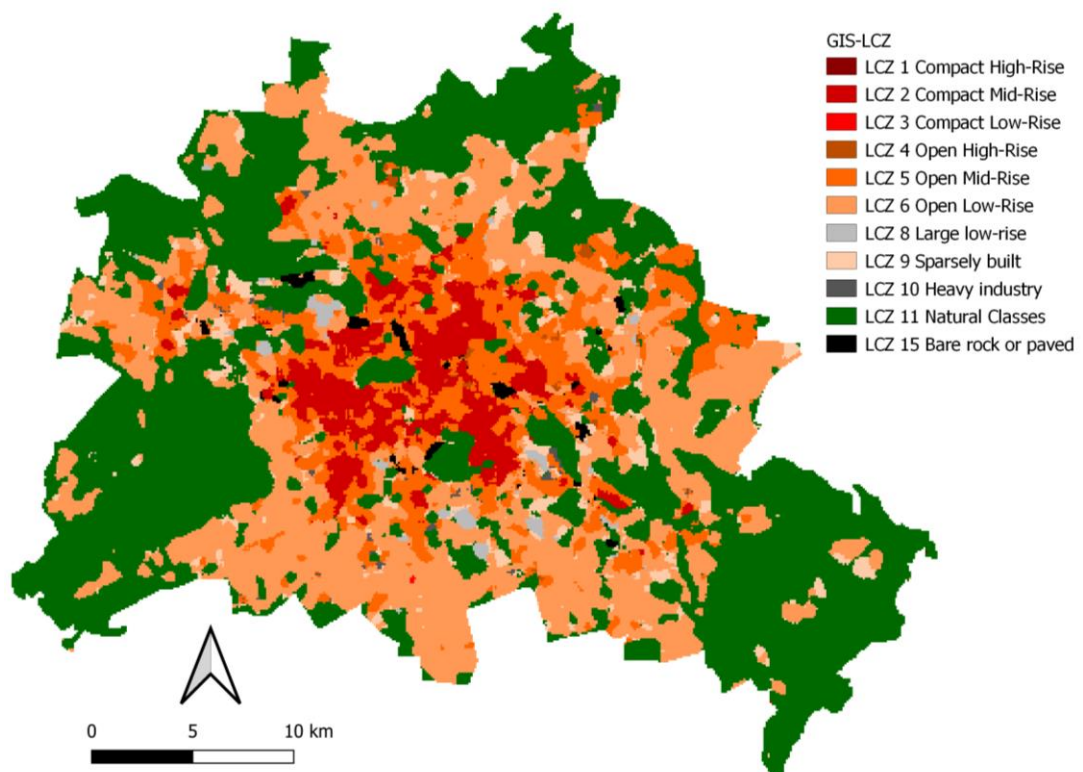

Figure 33. GIS-LCZ

3.4 Calculation of surface urban heat island intensity (SUHII)

In the evaluation, the result of GIS-LCZ map will be correlated with the SUHII. This section will explain the steps to retrieve the SUHII from land surface temperature (LST) obtained from the satellite imagery of MODIS Aqua instrument. The period of the LST is from 21st of June until 4th of July 2010.

For the calculation of SUHII, the LST values per pixel for each day in the period and for day and night are averaged. As a result, there are two different mean raster, namely average daytime LST and average nighttime LST. The average calculation is done using Python libraries (*numpy* and *rasterio*). The resulting raster of the average LST needs to be scaled by the factor of 0.02 to obtain the LST value in Kelvin (MYD 2021). The pixel values are then subtracted with 273.15 to get the LST values in degree Celsius. The scaling and subtraction are carried out by using *Raster Calculator* in QGIS. The LST raster is reprojected to EPSG 25833 from the CRS of the original data source WGS 84 applying *Warp (Reproject)* tool in QGIS. Figure 34 shows the average daytime LST of the 2010 period.
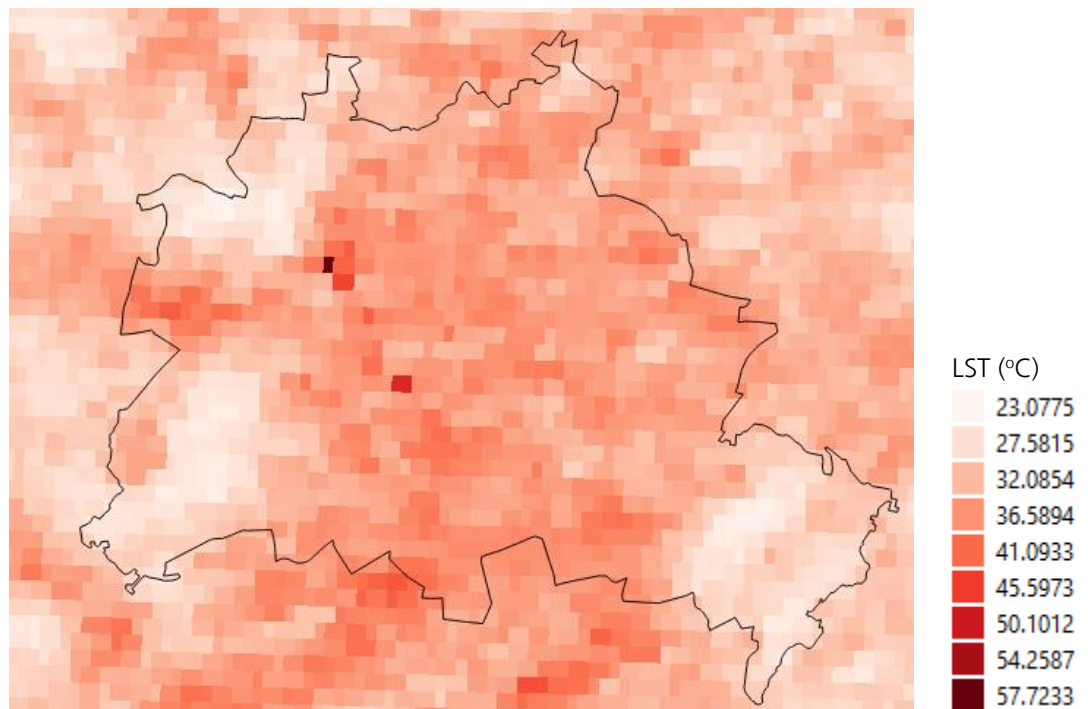

Figure 34. Average day LST of 2010 period
(from 21st of June until 4th of July 2010)

After the average LST values are obtained, a ring rural reference (RRR) is defined. The RRR (shown as green colored ring in Figure 35) is specified with the inner radius of 26 km from the city center and with the thickness of the ring of 3.765 km. It has the same area as the city. This ring is determined based on the work of Vogel and Afshari (2020). The average day and night LST are calculated for the area of RRR with the resulting average LST values of 31.24 °C for the day and 10.84 °C for the night.
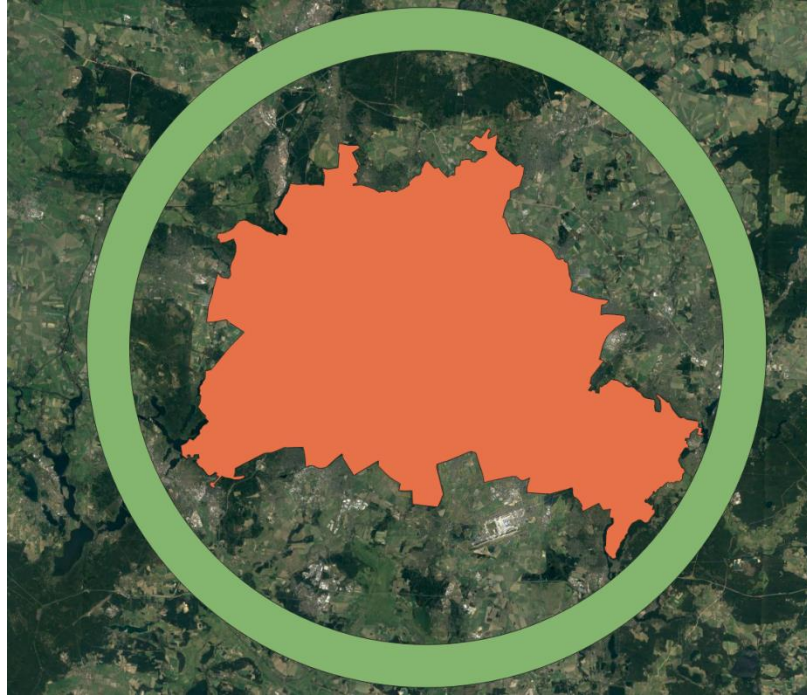

Figure 35. Ring rural reference outside Berlin

Before calculating the SUHII, the GIS-LCZ's grid tiles are resampled to the size of LST's pixel resolution (683 meter) and with the extent of the LST raster data. Resampling is done by applying GRASS library *r.resamp.stats* in QGIS. This calculation is saved in GeoTIFF and then converted to vector with shapefile format applying the *Raster Pixels to Polygon tool*.

For the calculation of SUHII, new attributes of the average day and night LST in the area of berlin are added to the resampled grid tile. The addition of the attributes is done with the *Zonal Statistics* tool by taking the *mean statistics* of the rasters of average day and night LST over the resampled vector grid tile. SUHII values are calculated for day SUHIII, night SUHII, and average SUHII. Day or night SUHII are obtained from the subtraction of the average day or night LST of every grid tile of Berlin area, with the average day or night LST of the RRR. The average SUHII, shown in Figure 36, is the mean of day SUHII and night SUHII. The calculation of these three SUHII values is done in *Field Calculator* in QGIS in the attribute table of the resampled vector shapefile.

The average SUHII is used for the correlation analysis with the urban fraction calculated from GIS-LCZ. The urban fraction (UF) is the addition between building surface fraction (BSF) and impervious surface fraction (ISF) of the zone property. For the correlation calculation, the BSF and ISF are converted into raster applying GDAL library *Rasterize (vector to raster)* tool. In the tool, no data value needs to be specified by assigning arbitrary number other than 0 until 100, because the values in this range are the BSF or ISF value. If the no data value is not specified here, the value of zero of the BSF or

ISF will be assigned as no data since the default value of no data in this tool is zero. If the default value is already used, the assignment of no data can also be later done using *Translate* tool from GDAL by assigning a specified no data value to output bands. Finally, the rasters of BSF and ISF are added to the resampled grid tile applying *Zonal Statistics* tool by taking the *mean* statistics of the rasters. The UF is simply the addition between the mean values of ISF and BSF in the resampled grid tile, and this addition is carried out applying the *Field Calculator*. The GIS-LCZ classes are also added to the resampled grid tile by calculating the *majority* statistics of the GIS-LCZ raster using the *Zonal Statistics* tool.
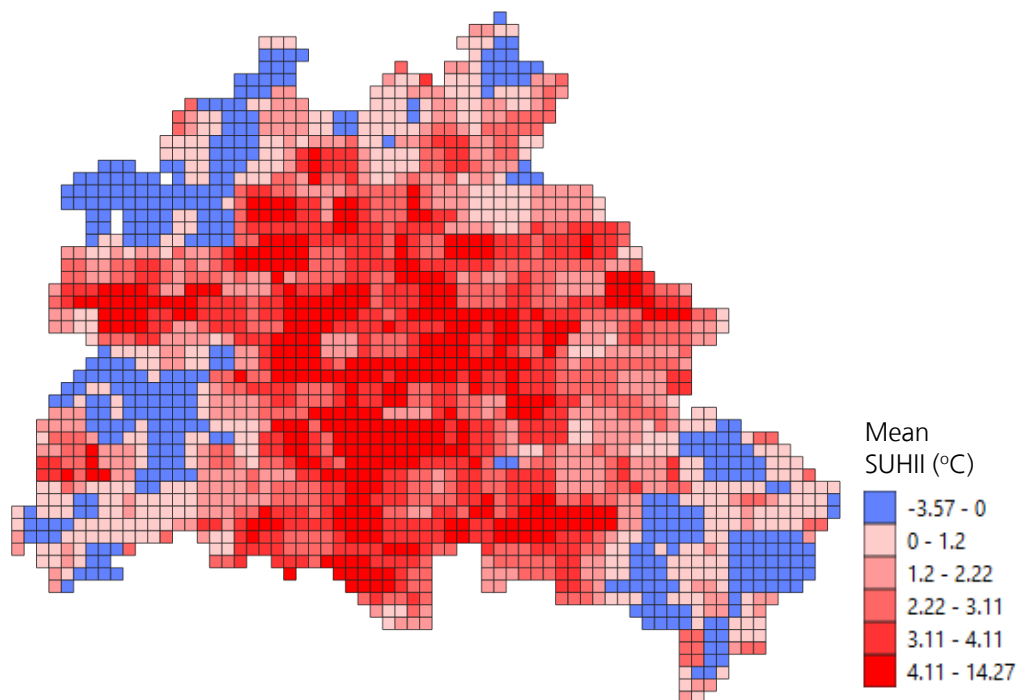
Figure 36. Average SUHII 2010

3.5 Retrieval of urban heat island intensity (UHII)

The result of GIS-LCZ will also be correlated with the UHII. The UHII is obtained from the WRF simulation result that is stored in netCDF file format. The file contains several variables including:

- XLON   : longitude
- XLAT   : latitude
- UF   : urban fraction
- UHII_vrr_tavg   : UHII calculated with the virtual rural reference (VRR). VRR is a fictitious rural area created from urban area that is replaced by natural classes (Vogel and Afshari (2020)).

These variables are used to calculate the correlation of UHII and UF with the GIS-LCZ. To retrieve these variables from the netCDF file, a tool called *Make netCDF feature layer* in ArcGIS Pro is used. The UHII_vrr_tavg and UF are read separately.

To read the UHII_vrr_tavg from the netCDF file, six parameters need to be defined in the tool. They are variables, X variable, Y variable, output feature layer, row dimensions

and value selection method. Variables are the data from netCDF file that will be read which is the UHII_vrr_tavg. The X and Y variables can be filled with the XLON and XLAT respectively. The output feature layer is the name of the output layer. The row dimensions are used to separate the value of UHII for every grid point of WRF. The dimensions used are *west_east* and *south_north*. The value selection method used is *by value* which is the default method. After all of these parameters are set, the tool can be run. The output layer is in the form of grid points, and then it is saved as a shapefile and further processed in QGIS. The grid vector is reprojected in QGIS to EPSG 25833 and buffered with the distance of 500 meter to get a square-grid form as shown in Figure 37. The buffering is done with *Buffer* tool with the segment of one and end cap style of square.

The same steps applying ArcGIS Pro are also done to retrieve the UF data from the netCDF file. After reprojection to EPSG 25833 in QGIS, the UF is added to the buffered grid of UHII by applying *Join Attributes by Location* tool. The GIS-LCZ classes are also added to the grid by calculating *majority* statistics of *Zonal Statistics* tool over the GIS-LCZ raster.
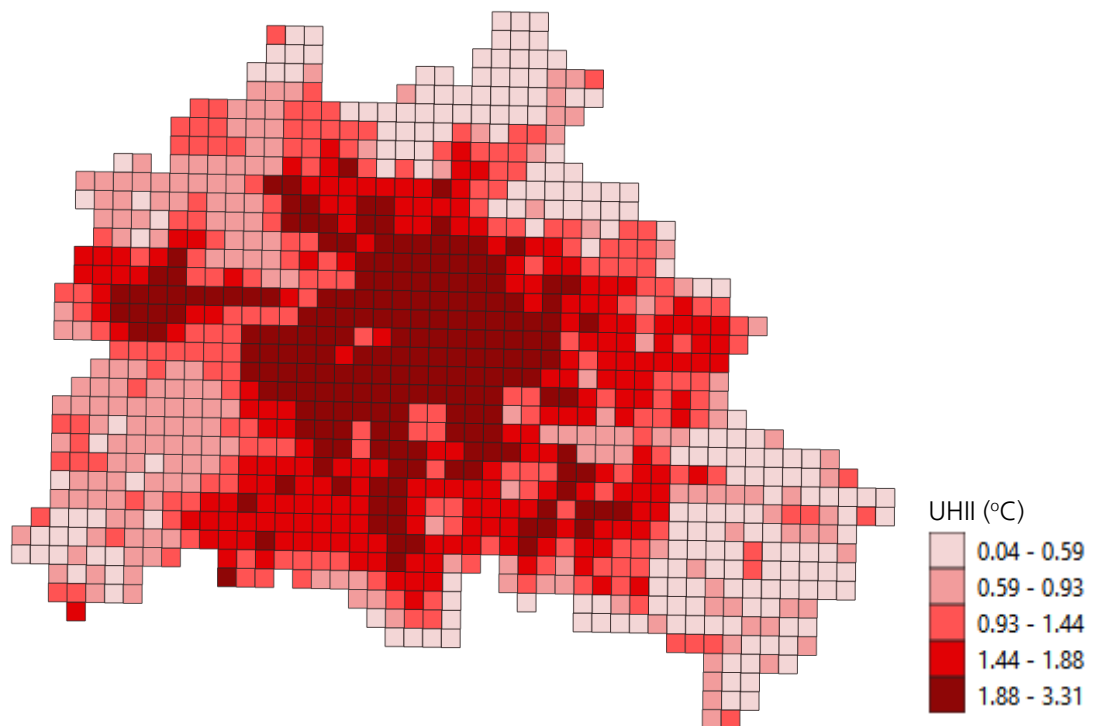

Figure 37. Simulated UHII

4   Evaluation and Analysis

In this chapter, the result of GIS-LCZ for Berlin is evaluated and analyzed. First by comparing the GIS-LCZ with the training areas from WUDAPT. Second by comparing the map to WUDAPT L0. Third by correlating the urban fraction from GIS-LCZ with the SUHII for each LCZ class. Fourth by correlating the urban fraction and UHII from WRF simulation for each LCZ class. Fifth by analyzing the enhancement of CityGML model with the result of GIS-LCZ.

4.1   Comparison to WUDAPT's training areas

The result of GIS-LCZ is compared to training areas from WUDAPT using the confusion matrix calculated in SAGA applying *Confusion Matrix (Polygons / Grid)* tool. Before the calculation, the GIS-LCZ map is rasterized and loaded as a grid in SAGA. Training areas from WUDAPT, which are in the vector shapefile format, are loaded in SAGA as *Shapes* in the form of polygons. The confusion matrix tool evaluates whether the grid or pixel center falls into the polygon or not. If it falls, it will be counted as one value of agreement or correctly classified pixel in the confusion matrix.

The result of GIS-LCZ is compared to the training areas (TA) from WUDAPT using confusion matrix, which is tabulated in Table 4.

Table 4. Confusion matrix of GIS-LCZ and WUDAPT's TA

	CLASS	WUDAPT's Training Area							SumUser	AccUser
		2	4	5	6	8	9	11		
GIS-LCZ	1	0	0	0	0	0	0	0	0	0
	2	1224	2	101	1	3	0	0	1331	91.96
	3	0	0	0	0	3	0	0	3	0
	4	0	37	27	0	0	0	0	64	57.81
	5	83	422	863	7	49	0	0	1424	60.6
	6	0	45	182	2007	387	43	0	2664	75.34
	8	0	0	0	0	12	0	0	12	100
	9	0	4	12	56	26	0	3	101	0
	10	0	1	36	0	2	0	0	39	0
	11	0	87	40	3	8	2	2158	2298	93.91
	15	0	0	2	0	33	0	0	35	0
	SumProd	1307	598	1263	2074	523	45	2161		
	AccProd	93.65	6.19	68.33	96.77	2.29	0	99.86		
	Overall Accuracy = 79,05%									
	Kappa = 73,09%									

The training area data from WUDAPT comprises LCZ 2, 4, 5, 6, 8, and 11. It does not have LCZ 1, 3 and 10 which GIS-LCZ has. The accuracy or the agreement calculation between the WUDAPT's training areas and GIS-LCZ can only be conducted between the classes which exist in WUDAPT. From the table, the green color shows the value of the correctly classified GIS-LCZ pixels compared to WUDAPT's training area polygons.

The producer accuracy for LCZ 4, 8, and 9 does not show good values. On the other hand, the user accuracy does not have good values for LCZ 1, 3, 9, 10, and 15. LCZ 1,

3, 10, and 15 do not exist in WUDAPT, and thus the user accuracy is zero. For the LCZ 9, it has producer and user accuracies of zero since there is no LCZ 9 of WUDAPT correctly classified in GIS-LCZ. The overall accuracy calculated for the classification result of GIS-LCZ is 79.05 % with the kappa coefficient of 0.73 which is considered as a good kappa value.

4.1.1   LCZ 4

For LCZ 4 in WUDAPT's training area, many of the tiles are classified as LCZ 5 in GIS-LCZ. LCZ 4 and 5 are indeed in a similar group of zones, which is open zone. LCZ 4 is an open high rise and LCZ 5 is open mid-rise. The significant differences between these two LCZ classes are in the range values of aspect ratio (LCZ 4: 0.75-1.25, LCZ 5: 0.3 - 0.75) and building height (LCZ 4: >25 meter, LCZ 5: 3-10 meter). Figure 38 shows one of the example of LCZ 4 in WUDAPT's training areas (yellow polygon) but it is LCZ 5 in GIS-LCZ. When the aspect ratio (H/W) and mean building height (H) for the WUDAPT's training area are calculated, it shows the value of 0.57 and 14 meter respectively. These values are suitable for LCZ 5, not LCZ 4. Moreover, the H/W and H of the LCZ 5 of the GIS-LCZ shown in this figure is 0.6 and 16 meter respectively. These values are also fit for the zone properties of LCZ 5 but not LCZ 4. Thus, this case shows the lack of WUDAPT method in deriving LCZs where it cannot quantify the H/W and H since the decision of the class of LCZ in WUDAPT is based on the visual perspective obtained from satellite imagery.
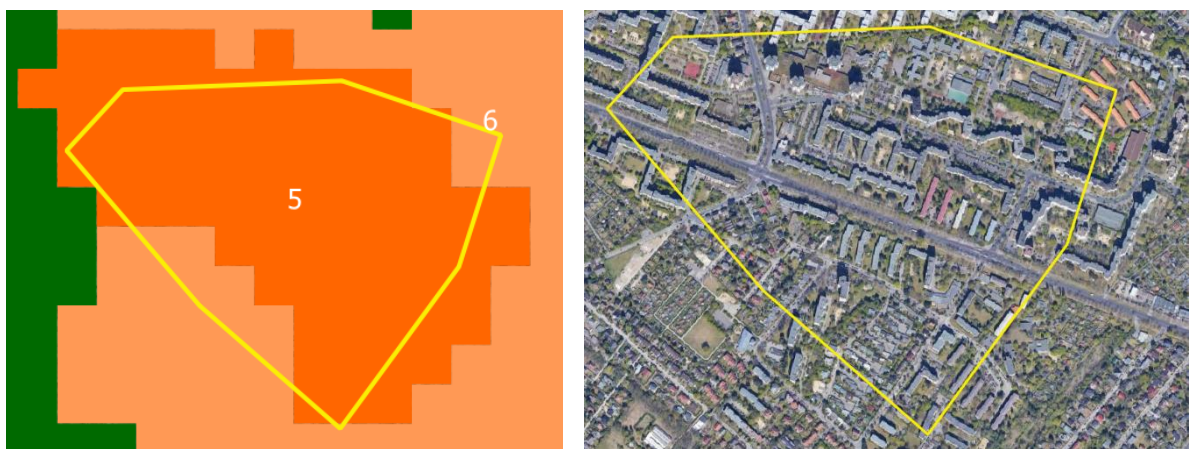

Figure 38. LCZ 4 in WUDAPT'S TA and LCZ 5 in GIS-LCZ. Left image: WUDAPT'S TA (yellow polygon) and GIS-LCZ (colored polygon) with its label. Right image: satellite view from the left image. (id: 24545 in shapefile)

4.1.2   LCZ 8

For LCZ 8 in WUDAPT's training areas, many of the tiles are classified as LCZ 6 in GIS-LCZ. As shown in Figure 37, the yellow polygon represents WUDAPT's TA which is LCZ 8 but classified as LCZ 6 in GIS-LCZ. Visually the yellow polygon is LCZ 8, but in GIS-LCZ it is considered as LCZ 6 after applying majority filter. Before filtering, there exists other classes such as LCZ 2, 3, 5, and 15 (represented as red grid tiles with labels in Figure 39). However, LCZ 8 is not there but in the periphery. These periphery tiles of LCZ 8 do not give significant impact for the majority filter to categorize the yellow polygon as LCZ 8.

Figure 39. LCZ 8 in WUDAPT's TA and LCZ 6 in GIS-LCZ. The labels inside the grid tiles are the unfiltered GIS-LCZ. (id: 80460 in shapefile)

LCZ 6 and LCZ 8 are respectively open low rise and large low rise. From the names of the LCZs, it can be inferred that they have the same building height's range values (3-10 m) but differ in the width of space between the buildings. This will eventually differ them in H/W value. Because the space is wider in LCZ 8, the H/W would be smaller (0.1-0.3) than LCZ 6 (0.3-0.75). When this yellow polygon is evaluated, the H value is 11.18 m which is still in the RB-RZB range value of LCZ 6 and 8. The average H/W value is 0.72 which is suitable to LCZ 6. As the area of the yellow polygon viewed from the satellite imagery, it obviously does not have that much space to be categorized as LCZ 8. On the other hand, LCZ 6 is also not really suitable for this area as the satellite image shows only small portions of vegetation. LCZ 6 should have pervious surface with the LB of 30%. This is the reason, in the unfiltered version, the tiles are classified diversely to classes that do not have that much of pervious surface fraction: LCZ 2, 3, 7, and 15. From this case, it implies that there is no perfect match in a classification result as also mentioned by Stewart and Oke (2012). There is always a case where the zone properties could lie between two different LCZ classes. This is the reason that in some studies, they add additional LCZ such as LCZ 34 which represent the mix of compact low-rise and high-rise (Chen et al. 2020).

Another example for LCZ 8 in WUDAPT's training areas classified as another LCZs is shown in Figure 40. In this Figure, there is one dominant class of the GIS-LCZ which is LCZ 15 (bare rock or paved surface). In the satellite image, the LCZ 15 is also dominant as shown in the right image. WUDAPT'S training areas generalizes the area as one LCZ class which is LCZ 8 which represents an airport area based on the framework. On the other hand, GIS-LCZ gives detail since the tiles are in 100 m x 100 m. For this case, classification of GIS-LCZ are successful at detecting the LCZ 15, but the classification cannot recognize LCZ 8 which represents the buildings and the paved surfaces of the airport. In the unfiltered version, there exists few tiles which are LCZ 8 and they are in the periphery. After filtering, these tiles are removed. LCZ 8 cannot be represented in

these building area because of the building height in this area is around 20 meter which is not the property of LCZ 8 (3-10 m). The H in this area are for LCZ 2 and 5. This is the reason that in the unfiltered version, the buildings area of the airport are classified as LCZ 2 or 5. Finally, after filtering, some parts of the building are generalized in to LCZ 5. From this case, it is also found that the framework of LCZ is not ideal for every city. The visual inspection done by WUDAPT categorizes the area as LCZ 8, which is supposed to represent airport area. However, based on the zone property of H, the area cannot be in the LCZ 8.
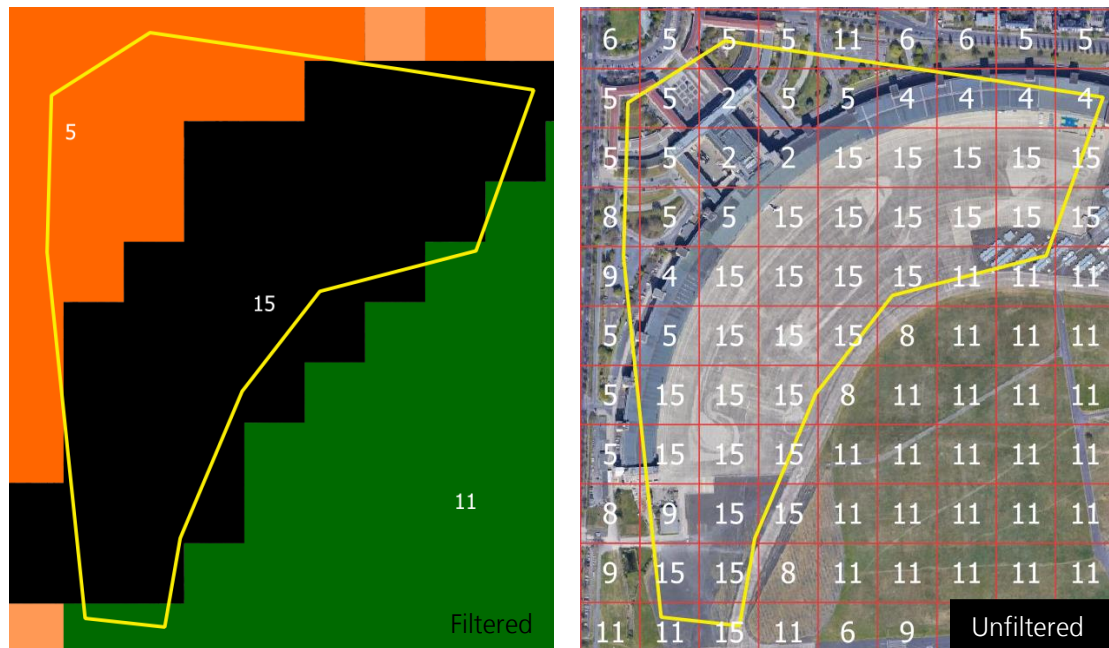

Figure 40. LCZ 8 in WUDAPT'S TA and LCZ 15 in GIS-LCZ. Left image: yellow polygon represents LCZ 8 in WUDAPT'S TA and filled-colored polygons represents GIS-LCZ with its LCZ class label. Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label. (id: 85119 in shapefile)

4.1.3   LCZ 9

For LCZ 9 in WUDAPT's TA, there is only one polygon of training area representing this class of LCZ. However, this class is not found as LCZ 9 in GIS-LCZ but it is found primarily as LCZ 6 in GIS-LCZ as shown in the right image of Figure 41. The left image shows that the area is an allotment where the small gardens exist with small buildings. However, most of these buildings are not detected either by the DLR raster or by OSM. Thus, calculation of zone properties for defining the LCZs are not optimal in this area. Building height, sky view factor, building surface fraction of SVF cannot be calculated correctly. Only impervious surface fraction can be obtained from the IMD of Copernicus. The mean ISF for the yellow polygon is 32.63%. This value is in the range value of LCZ 6, not LCZ 9. Visually from the satellite image, the area is also more to LCZ 6 rather than LCZ 9 because LCZ 9 represents a sparsely built area which is spaced widely in comparison to LCZ 6. Thus, the result of the GIS-LCZ in this area is favored.
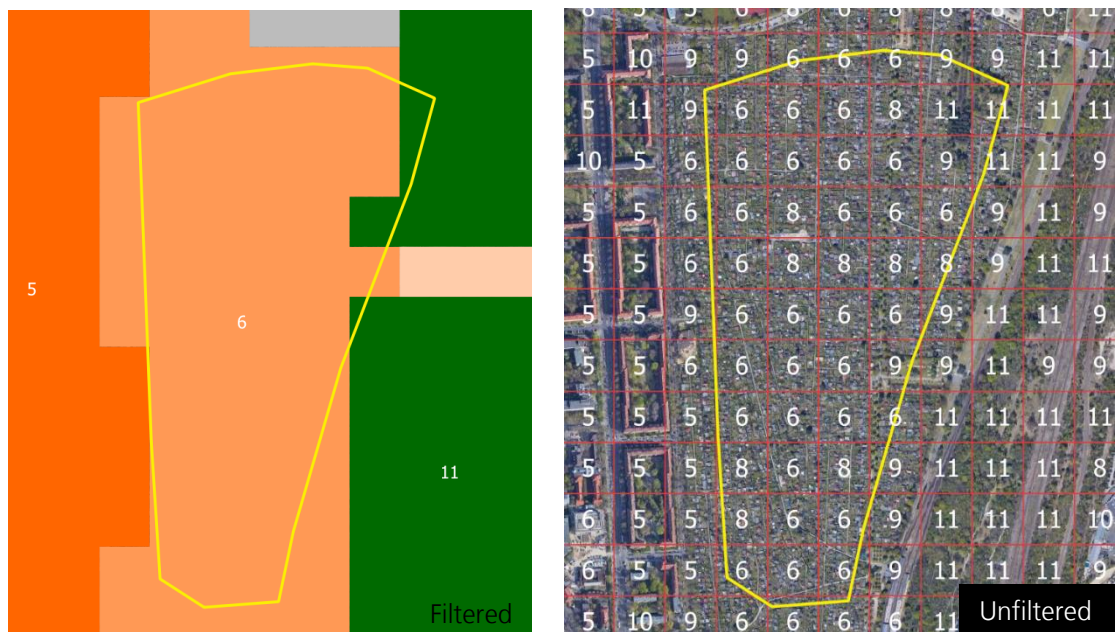
Figure 41. LCZ 9 in WUDAPT'S TA and LCZ 6 in GIS-LCZ. Left image: yellow polygon represents LCZ 9 in WUDAPT's TA and filled-colored polygons represents GIS-LCZ with its LCZ class label. Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label (id: 74127 in shapefile)

4.2 Comparison to WUDAPT L0

A comparison is carried out between the WUDAPT L0 map and the GIS-LCZ map of Berlin. Both maps are represented in Figure 42. Each grid tile from the WUDAPT L0 map and the GIS-LCZ map is compared to find the agreement between these two mapping methods, i.e. whether the LCZ classes match. The comparison is done applying the confusion matrix in SAGA. For the calculation of the confusion matrix, the WUDAPT L0 is converted into vector shapefile in QGIS and loaded as *Shapes* in SAGA. The resulting confusion matrix is tabulated in Table 5. The overall accuracy calculated from this matrix is 67.33%. This value denotes the relatively good agreement between WUDAPT L0 and the GIS-LCZ.
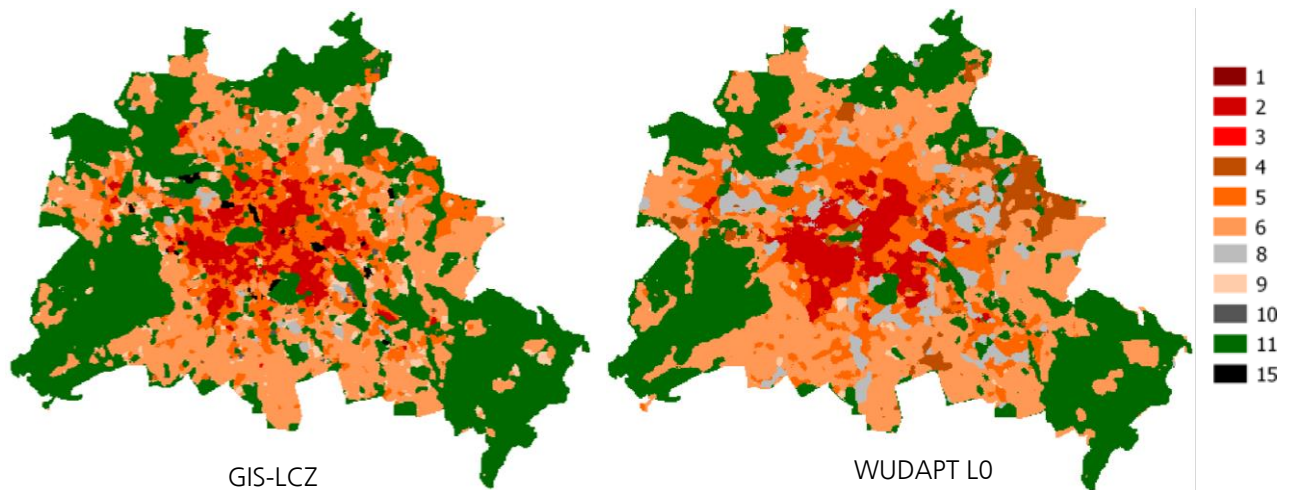
Figure 42. GIS-LCZ vs WUDAPT L0

Table 5. Confusion matrix of GIS-LCZ and WUDAPT L0

	CLASS	WUDAPT L0							SumUser	AccUser
		2	4	5	6	8	9	11		
GIS-LCZ	1	1	2	0	2	2	0	1	8	0
	2	4545	31	870	30	223	0	1	5700	79.74
	3	0	0	0	0	15	0	0	15	0
	4	9	114	52	3	3	0	0	181	62.98
	5	1453	1985	6363	1204	815	2	50	11872	53.6
	6	52	692	3375	19566	3168	160	184	27197	71.94
	8	0	36	248	312	72	41	11	720	10
	9	9	131	411	1874	481	68	105	3079	2.21
	10	0	21	164	76	14	0	3	278	0
	11	52	1067	892	7762	667	248	30221	40909	73.87
	15	73	6	57	0	410	0	12	558	0
	SumProd	6194	4085	12432	30829	5870	519	30588		
	AccProd	73.38	2.8	51.18	63.47	1.23	13.10	98.80		
	Agreement ("overall accuracy") : 67.33%									

4.2.1 CZ 2

In the confusion matrix, the producer accuracy of LCZ 2 is considered good. However many of LCZ 2 from WUDAPT L0 are classified as LCZ 5 in GIS-LCZ. These classes are found in the GIS-LCZ as neighboring classes. Based on the definition, these classes belongs to *midrise* group where LCZ 2 is compact midrise and LCZ 5 is open midrise. From the name of the classes, it can be inferred that LCZ 5 has more spaces compared to the LCZ 2. The zone properties of these classes are different significantly in H/W and BSF. They have the same range value for H and ISF as well as a bit of overlap in the SVF's range value. Figure 43 shows two images. Left image is the grid tiles which are colored and labelled. The color represents WUDAPT L0' LCZ classes (red: LCZ 2, orange: LCZ 5). The label represent the LCZ classes from GIS-LCZ. Right image is the representation of the satellite imagery of google maps in QGIS.

Figure 43. LCZ 2 in WUDAPT L0 and LCZ 5 in GIS-LCZ. Left image: color *represents WUDAPT L0 and* label *represent GIS-LCZ. Right image: satellite imagery from tiles of the left image. (id: 76418 in shapefile)*

Based on the satellite image, the area of the tiles has quite large spaces between building that it is supposed to be LCZ 5 rather than LCZ 2. When the zone properties of the four tiles (yellow box) are evaluated, the H values are found to be in the range of LCZ 2 and 5 (10-25 m). However, the ISF value for these four tiles is not in the range of both LCZs range values (30-50%). It has lower value than that which is around 20%. Nevertheless, this value is not that far from the range, which is still in the LZB-LB range. For the H/W property, the average value of H/W for these tiles is 0.75, which is the upper range (RB) for LCZ 5 but lower limit (LB) for LCZ 2. Moreover, the BSF and SVF values in these four tiles are around 29% and 0.6, respectively, which are in the range of LCZ 5 but not LCZ 2. The zone properties calculated for these tiles are more for LCZ 5 rather than LCZ 2. Thus, the classification of WUDAPT in this case is not correct.

The other way around, in Figure 44, it is shown that the tiles of LCZ 2 in GIS-LCZ are classified as LCZ 5 in WUDAPT L0. This also happens between the border of LCZ 2 and 5. The yellow box seems to be LCZ 2 when it is inspected with the satellite imagery, which is the same with the class of GIS-LCZ. The H, SVF, and ISF values for the yellow box are in the range of LCZ 2 and 5. However, the BSF and H/W values for these two tiles in the yellow box are in the range of LCZ 2, not LCZ 5. Thus, the classification of these tiles are again correct based on GIS-CLZ, but not WUDAPT L0.

However, the neighboring white box does not look like LCZ 2 in the satellite image, but LCZ 5. Although the H & ISF values form this tile is in the range of LCZ 2 and 5, the other properties, which are SVF, BSF and H/W, are in the range values of LCZ 5, not LCZ 2. Therefore, the classification of GIS-LCZ here is not correct. When it is traced back, this tile was originally LCZ 5 in GIS-LCZ before majority filtering applied. This case denotes the drawback of the majority filter in removing detail of the LCZ classes in GIS-LCZ.

Figure 44. LCZ 5 in WUDAPT L0 and LCZ 2 in GIS-LCZ. Left image: color *represents WUDAPT L0 and* label *represent GIS-LCZ. Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label. (id: 74444 in shapefile)*

4.2.2 LCZ 4

LCZ 4 from WUDAPT L0 are classified mostly to LCZ 5 in GIS-LCZ, which is the case for 1985 tiles compared to its agreement for only 114 tiles. A similar case is also found in the accuracy assessment in section 4.1 when the confusion matrix of GIS-LCZ and WUDAPT's TA is evaluated. From the previous case, it is concluded that the shortcoming of WUDAPT in detecting H/W and H leads to the misclassification of LCZ 4 to LCZ 5.

4.2.3 LCZ 5

For LCZ 5 in WUDAPT L0, there are many of them that are classified as LCZ 6 in GIS-LCZ. The orange tiles in Figure 45 are LCZ 5 in WUDAPT L0 but LCZ 6 in GIS-LCZ. LCZ 5 and 6 are in the same group of *open* LCZ. LCZ 5 is open midrise and LCZ 6 is open low-rise. From the name of each class, it can be inferred that the significant difference between them is in the building height. When the zone properties are evaluated, the H values for these tiles are actually in the range of LCZ 6, not LCZ 5. Thus, the classification of WUDAPT L0 is not correct in this case.
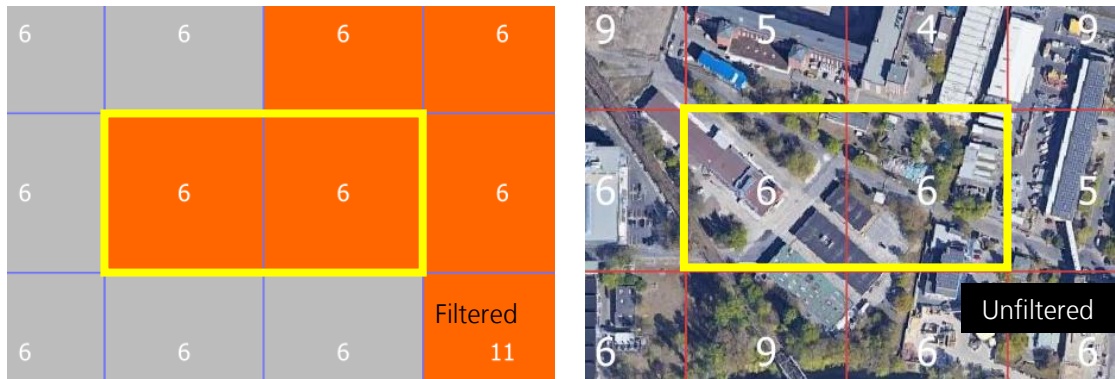
Figure 45. LCZ 5 in WUDAPT L0 and LCZ 6 in GIS-LCZ. Left image: color *represents WUDAPT L0 and* label *represent GIS-LCZ. Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label. (id: 55188 in shapefile)*

Another case is shown in Figure 46. The yellow box is LCZ 5 in WUDAPT L0 but LCZ 6 in GIS-LCZ. The satellite image shows that these two tiles are more for LCZ 15 or probably a bit part of LCZ 8 (as the unfiltered GIS-LCZ). In these two tiles, there is no building that the H and H/W are null and BSF will be 0. The ISF value for these tiles is in the range of several LCZ classes: LCZ 1, 2, 3, 5, 6, and 8. From these classes, the SVF values for these tiles fit only to LCZ 8. The unfiltered of GIS-LCZ has the same LCZ. After the filtering, the class becomes LCZ 6. Thus, none of WUDAPT L0 or GIS-LCZ result is correct here.
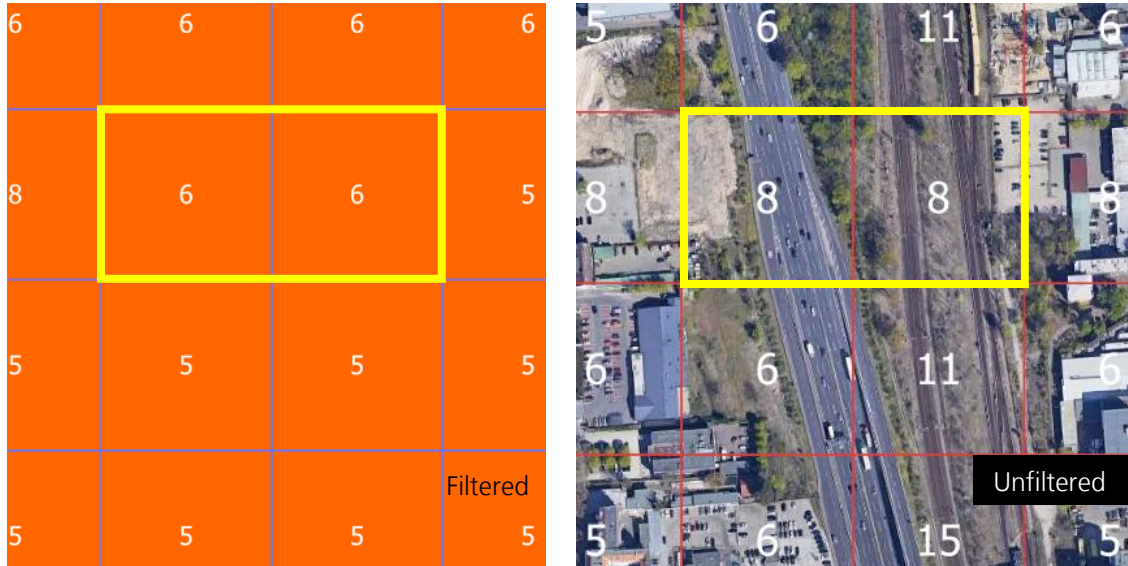

Figure 46. LCZ 5 in WUDAPT L0 and LCZ 6 in GIS-LCZ. Left image: color *represents WUDAPT L0 and* label *represent GIS-LCZ. Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label. (id: 55986 in shapefile)*

There are also tiles of LCZ 5 in WUDAPT L0 but LCZ 8 in GIS-LCZ as shown in this Figure 47. Based on the satellite image, it can be perceived as LCZ 6. The classification of GIS-LCZ in these tiles cannot be conducted correctly since the building data to calculate related zone properties (SVF, H/W, H, BSF) are mostly not available from either DLR or OSM. Thus, neither WUDAPT nor GIS-LCZ has a correct classification in this case.

Figure 47. LCZ 5 in WUDAPT L0 and LCZ 8 in GIS-LCZ. Label in satellite image represents GIS-LCZ classes before and after filtering. (id: 63038 in shapefile)

4.2.4 LCZ6

For LCZ 6 in WUDAPT L0, there are tiles, which are classified as LCZ 11 as shown in Figure 48. From the zone properties, the tiles have no building information as well as impervious surface that it leads to the classification of GIS-LCZ to LCZ 11. From the satellite image, the tiles are also easily recognized as LCZ 11. Thus, WUDAPT L0 classification is not correct here.

Figure 48. LCZ 6 in WUDAPT L0 and LCZ 11 in GIS-LCZ. Label in satellite image represents GIS-LCZ classes before and after filtering. (id: 68923 in shapefile)

The tiles shown in Figure 49 below are LCZ 6 in WUDAPT L0 but LCZ 9 in the filtered GIS-LCZ. Before filtering, they are LCZ 6 for the upper tiles and LCZ 9 for the lower tiles. Based on the satellite image shown, these tiles are between LCZ 6 and 9 which are open low-rise zone and sparsely built zone respectively. From the name of the zones, LCZ 9 should have more spaces than LCZ 6. Based on the properties, the range value for H of both LCZs are the same (3 – 10 m) and their SVF's range values are a bit overlapped. The significant difference of range values are for H/W, BSF, and ISF. The H/W, BSF, and SVF for the four tiles are LCZ 9. However, the ISF value for the tiles are

for LCZ 6. Since most of the zone properties are classified as LCZ 9, the GIS-LCZ approach here is considered correct.

Figure 49. LCZ 6 in WUDAPT L0 and LCZ 9 in GIS-LCZ. Label in satellite image represents GIS-LCZ classes before filtering. (id: 99145 in shapefile)

The tiles in Figure 50 are LCZ 6 in WUDAPT L0 but LCZ 5 in GIS-LCZ. Before filtering, the tiles were from two different zones: LCZ 5 and 11. As explained before, LCZ 5 and 6 are in the same group of *open* LCZ. LCZ 5 is open midrise and LCZ 6 is open low-rise. The significant difference between them is in the range values of the building height property. When the zone properties are evaluated, the H values for these tiles are actually in the range of LCZ 5, not LCZ 6. Thus, the classification of WUDAPT L0 is not correct in this case.

Figure 50. LCZ 6 in WUDAPT L0 and LCZ 5 in GIS-LCZ. Label in satellite image represents GIS-LCZ classes before filtering. (id: 94434 in shapefile)

There are also tiles which are classified as LCZ 6 in WUDAPT L0 but LCZ 8 in GIS-LCZ as shown in Figure 51. These tiles do not have building data that the related zone properties cannot be calculated correctly with the GIS-LCZ approach. From the image, these area looks like more to LCZ 6 as in WUDAPT L0.

Figure 51. LCZ 6 in WUDAPT L0 and LCZ 8 in GIS-LCZ. Label in satellite image represents GIS-LCZ classes before filtering. (id: 52050 in shapefile)

4.2.5 LCZ 8

The agreement of LCZ 8 of WUDAPT L0 to GIS-LCZ is very low that the producer accuracy is around 1.23% only. Most of the tiles are classified as LCZ 6 in GIS-LCZ for around 3168 tiles. Figure 52 shows the tiles where they are LCZ 6 in GIS-LCZ after the filtering. The labels shows the classes before filtering. The unfiltered version has more variety of classes: LCZ 3 and 9. From the satellite imagery, the area looks like LCZ 8 where the offices or workshops take place, which agrees with the classification from WUDAPT L0. When the tile in the middle of the image is evaluated, it has H of 6.3 m which is in the range of LCZ 3, 6, 8, 9, and 10. The BSF value is 24.4% which is in the range of LCZ 4, 5, 6, and 10. The H/W is 0.21 which is in the range of LCZ 8, 9, and 10. The ISF is 71.6% is in the range RB-RZB of LCZ 1 and 3. The SVF value is 0.9 which is in the range of LCZ 6, 8, 9, and 10. From the classification of zone properties, the tile is definitely assigned as LCZ 6 and 8. However, from the satellite image, the area has a small amount of pervious surface fraction that it cannot be classified as LCZ 6. Inclusion of pervious surface fraction (PSF) in the classification could lead this area to LCZ 8 since LCZ 8 is identified with PSF of less than 20%.

Figure 52. LCZ 8 in WUDAPT L0 and LCZ 6 in GIS-LCZ. Label in satellite image represents GIS-LCZ classes before filtering. (id: 92069 in shapefile)

4.2.6 LCZ 9

There are not many tiles classified as LCZ 9 in WUDAPT L0 (519 tiles) compared to GIS-LCZ (3427 tiles). From the 519 tiles, only 69 tiles agree with GIS-LCZ. The rest are distributed to LCZ 5, 6, 8, 9, 10 and 11. The highest number of distributed tiles are 247 tiles, which are classified as LCZ 11 in GIS-LCZ. Figure 53 shows LCZ 9 in WUDAPT L0 which is classified as LCZ 11 in GIS-LCZ. From the visual inspection of the satellite imagery, the tiles are definitely LCZ 11. The zone properties of GIS-LCZ has no information of H/W and H because there is no building. Thus, the BSF is zero everywhere. The ISF is also almost zero in all the tiles. Therefore, the suitable LCZ class is LCZ 11 which agrees with GIS-LCZ.
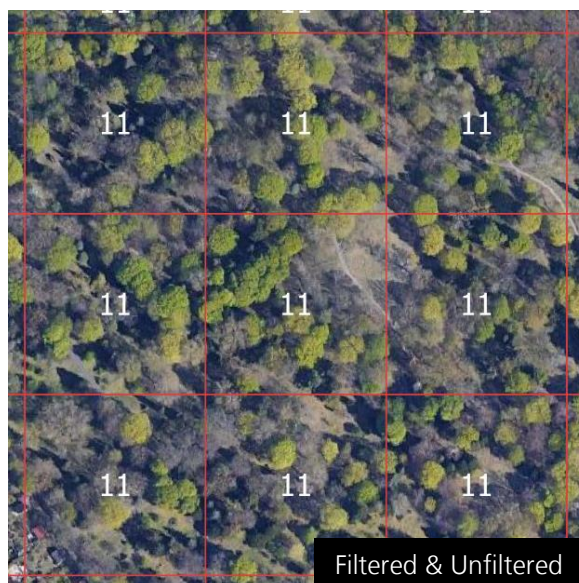

Figure 53. LCZ 9 in WUDAPT L0 and LCZ 11 in GIS-LCZ. Label in satellite image represents GIS-LCZ classes before and after filtering. (id: 66961 in shapefile)

The other way around, the GIS-LCZ 9 has the highest number of agreement to LCZ 6 of WUDAPT L0 for 1968 tiles. This agreement case is explained in section 4.2.4. which denotes the favored GIS-LCZ approach.

4.2.7 LCZ 11

LCZ 11 of GIS-LCZ has a good agreement with the WUDAPT L0. The producer accuracy is almost 99 % and the user accuracy is 74%. The user accuracy is lower because many of the tiles (around 19%) from GIS-LCZ 11 are classified as LCZ 6 in WUDAPT L0. This case is already discussed in section 4.2.4.

4.2.8 Disagreement: LCZ 1, 3, 10 and 15

There is no agreement found for GIS-LCZ 1, 3, and 10 to WUDAPT L0 since these classes are not available in WUDAPT L0 of Berlin. The number of tiles that belongs to these LCZs are also not significant compared to 90517 tiles in total. GIS-LCZ 1 has eight tiles which are classified to all LCZs in WUDAPT L0 except for LCZ 5 and 9.

Figure 54 shows the tiles of LCZ 1 of GIS-LCZ (bigger label in the left image) which are classified to several classes in WUDAPT L0 (smaller label in left image). As it is visualized in the satellite image, the area might be an LCZ 2. However when the zone properties are evaluated in these tiles, the building height values are more than 25 meter (one of the tiles has H of 55.41 meter). If the two lower tiles are evaluated, the H and H/W values belong to LCZ 1, the SVF and BSF are both for LCZ 1 and 2. Moreover, the ISF are for LCZ 1 for the left tile, but for the right tile is for LCZ 2 and still in the LZB-LB range of LCZ 1. Thus, the classification of GIS-LCZ here is correct.

Figure 54. LCZ 1 in GIS-LCZ and WUDAPT L0 LCZs. Left image: color *represents WUDAPT L0 and* label *represents GIS-LCZ (bigger label) and WUDAPT L0 (smaller label). Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label. (id: 79973 in shapefile)*

For LCZ 3 in GIS-LCZ, there exist only 15 tiles and they are all classified to LCZ 8 of WUDAPT L0. The visual imagery of some of the tiles of the LCZ 3 is shown in Figure 55. The rest of the tiles are mostly found in the similar places as viewed in the satellite

image. Based on the zone properties, LCZ 3 and 8 have the same range value for H and overlaps in BSF and ISF. Significant differences are found for the properties of SVF and H/W. When the zone properties of these tiles are evaluated, the mean values of H (9.3 m) and ISF (40.6%) are for LCZ 3 and 8. However, the BSF (55.6%), SVF (0.8), and H/W (0.8) are a fit for LCZ 3. Thus, WUDAPT L0 classification here is not correct.
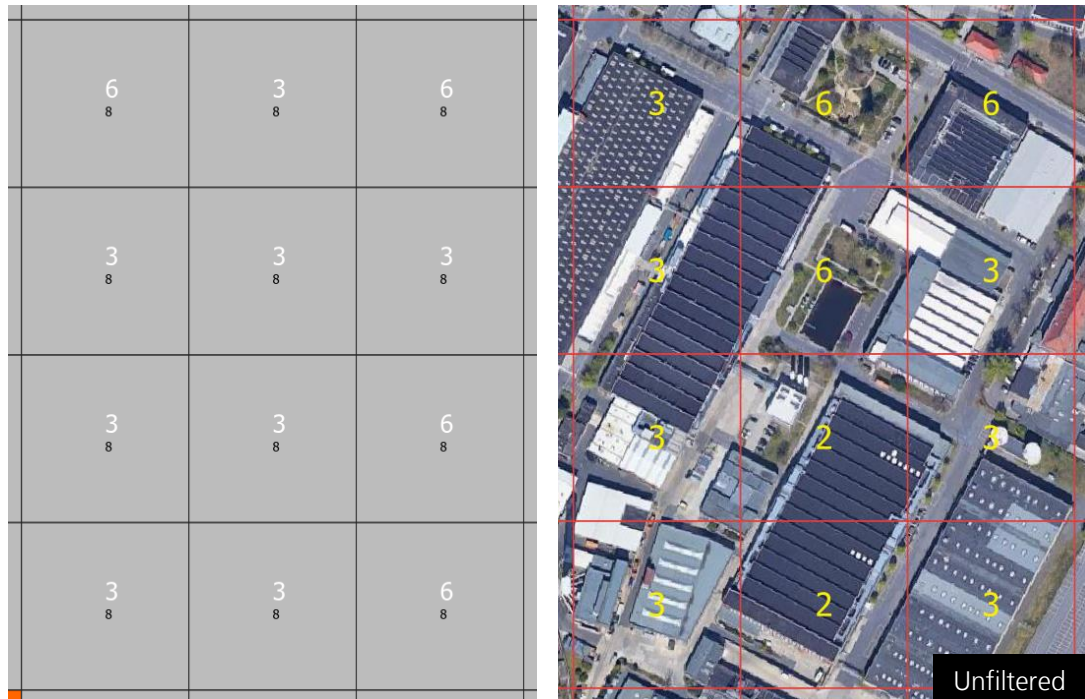

Figure 55. LCZ 3 in GIS-LCZ and LCZ 8 in WUDAPT L0. Left image: color represents WUDAPT L0 and label represents GIS-LCZ (bigger label) and WUDAPT L0 (smaller label). Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label. (id: 80856 in shapefile)

For LCZ 10 in GIS-LCZ, there are in total 278 tiles, which are mostly classified to LCZ 5 of WUDAPT L0 for 164 tiles. Some of these tiles are shown in Figure 56. Based on the satellite imagery on the right image, the area seems to be LCZ 5. When the tiles of LCZ 10 of the GIS-LCZ are evaluated, their mean values of H (12.6 m), H/W (0.47), BSF (23%), and SVF (0.8) are both for LCZ 5 and 10. But the mean value of ISF (20.4) is for LCZ 10, but still in LZB-LB of LCZ 5. Based on the zone properties, these tiles are suitable for LCZ 10. LCZ 10 is actually representing an area of heavy industry where few or no tress in the area. In the satellite image, the area does not look like the heavy industry area at all. The inclusion of other zone properties in the framework that are not included in the classification of GIS-LCZ such as pervious surface fraction and anthropogenic heat could further give a distinct decision of the final LCZ class.

Figure 56. LCZ 10 in GIS-LCZ and LCZ 5 in WUDAPT L0. Left image: color represents WUDAPT L0 and label represents GIS-LCZ (bigger label) and WUDAPT L0 (smaller label). Right image: satellite imagery from tiles of the left image with the unfiltered GIS-LCZ as label. (id: 84367 in shapefile)

For LCZ 15 in GIS-LCZ, the tiles are mostly classified as LCZ 8 in WUDAPT L0 for 410 tiles. This case has been explained in section 4.1.2.

4.3   Correlation analysis of GIS-LCZ with the calculated SUHII

The correlation between the GIS-LCZ classification result and the calculated SUHII from satellite imagery is evaluated by employing urban fraction from the GIS-LCZ. The correlation results for all the LCZs and each LCZ are shown in Figure 57. The LCZ class existing in the resampled GIS-LCZ are LCZ 2, 4, 5, 6, 8, 9, 11 and 15, but LCZ 4, 8, and 15 correlation are not shown here since mot enough data is available for the correlations.

The correlation results indicate a high correlation value of 0.708 for the correlation between urban fraction of GIS-LCZ and the average SUHII for all the LCZ classes. However, when the correlation calculation is conducted separately for each LCZ, the correlation values decrease. LCZ 11 has the highest correlation of 0.586 and LCZ 2 corresponds to the lowest correlation of 0.163. The correlation values are LCZ 11 > 9 > 6 > 5 > 2. The R^2 value for all LCZs is 0.501. However, when it is evaluated for each class of LCZ, the R^2 decreases significantly. The highest R^2 value for an LCZ class is LCZ 11 which is 0.343 and the lowest R^2 is 0.027 which is LCZ 2.

The correlation result of LCZ 11 leads to a question on why such a class has a high urban fraction which is until around 70 %. This is due to the result of the filtering of GIS-LCZ classes and the resampling to LST pixel resolution which takes the mode of LCZ classes (see section 3.4) existing in the extent of the LST pixel. The final result could have LCZ 11 as the mode class, but the urban fraction from the other urban classes existing in the unfiltered and non-resampled grid tiles are still considered in the final resampled tile.

Nevertheless, from all the correlation and R^2 values of the classes of LCZ, it shows that SUHII derived from LST cannot segregate the LCZ classes based on its urban fraction.
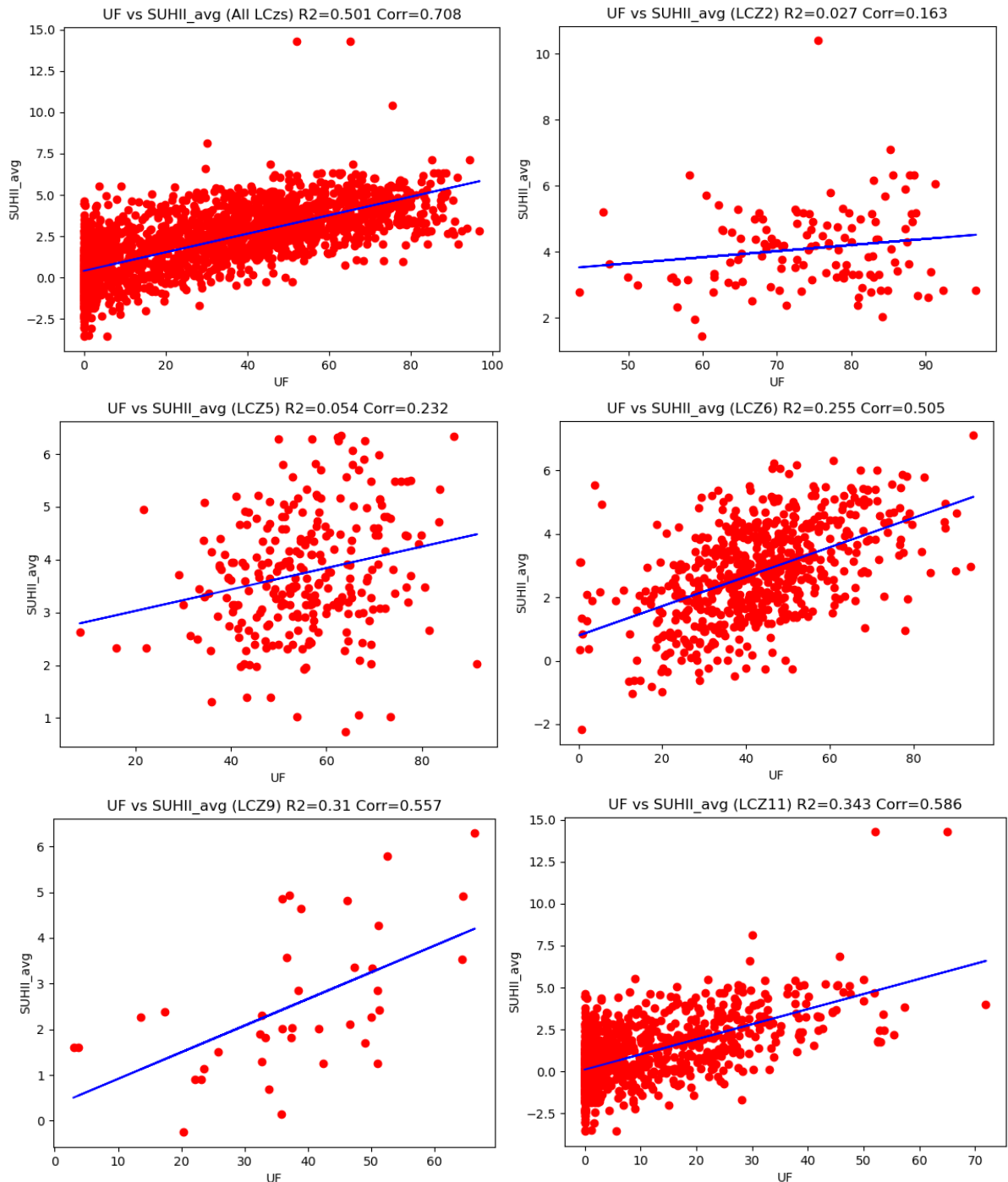
Figure 57. Correlation of SUHII and urban fraction of GIS_LCZ over the LCZ classes

4.4 Correlation analysis of GIS-LCZ with the simulated UHII

The correlation calculation is also conducted between the simulated UHII and the urban fraction from WRF considering the GIS-LCZ. The correlation is conducted for every GIS-LCZ class. After calculating the *majority* statistics, there are seven GIS-LCZ classes available over the grid of UHII. They are LCZ 2, 5, 6, 8, 9, 11, and 15. LCZ 8 and 15 are omitted for the correlation analysis due to less representation of these classes over the grid of UHII. The correlation result for the rest of the GIS-LCZ classes are shown in Figure 58. The upper left image shows the correlation of the urban fraction from WRF

and its UHII for all the LCZ classes, and the rest of the images shows a separate correlation analysis for each LCZ class.
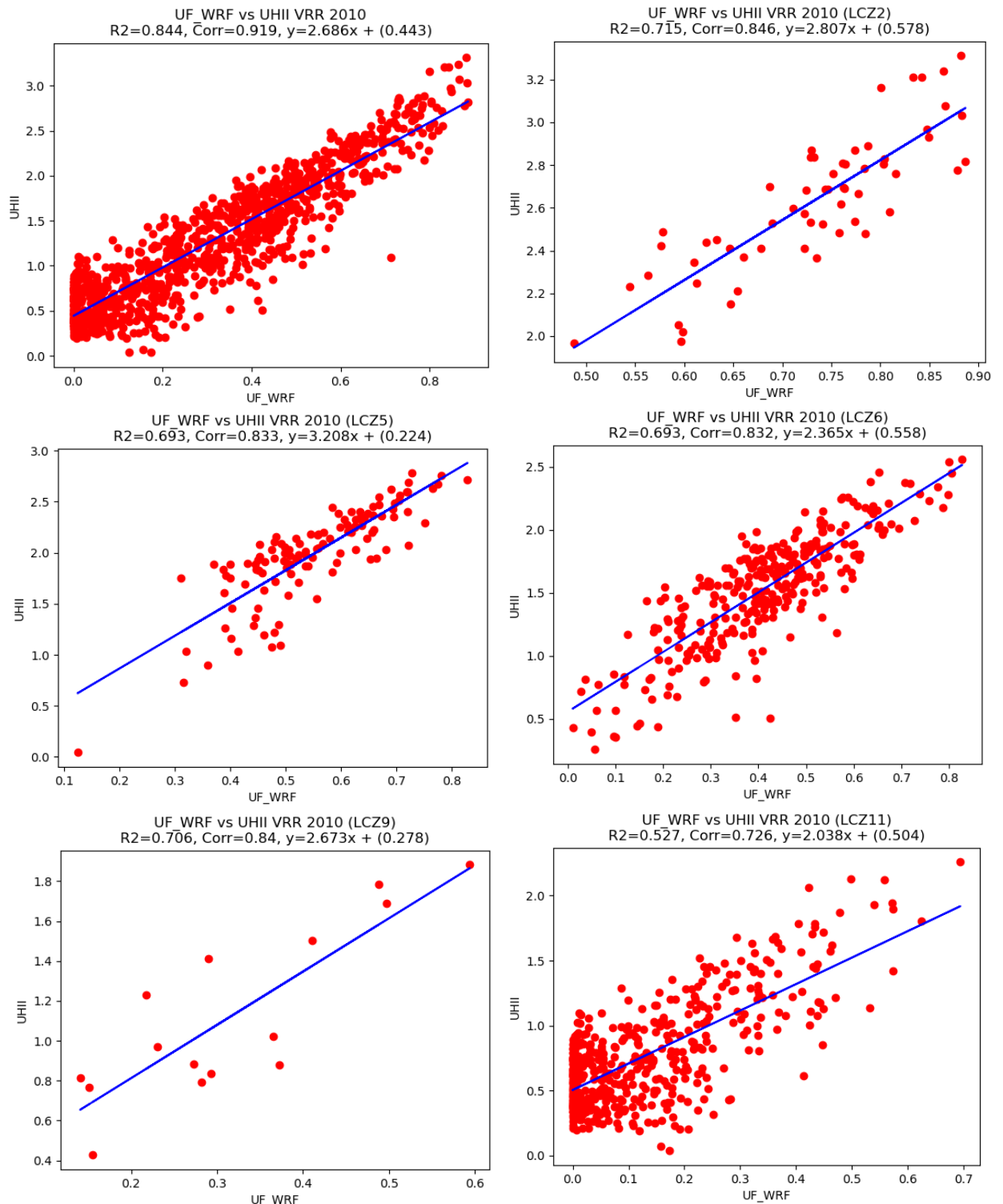

Figure 58. Correlation between UHII and urban fraction of WRF over GIS-LCZ classes

For the LCZ classes, the correlation result denotes LCZ 2 has the highest correlation value of 0.846 and LCZ 11 is the lowest one with the value of 0.727. The correlation values are LCZ 2 > 9 > 5 > 6 > 11. The highest R^2 value is also found in LCZ 2, which is 0.715, and the lowest one is also LCZ 11 with 0.529. The R^2 values are LCZ 2 > 9 > 5 > 6 > 11. From these measure values, it is found that the distribution of GIS-LCZ

classes can be well represented with the correlation of UHII and urban fraction from the WRF simulation result.

The standard error of the slopes and intercepts from the coefficients of the regression line are listed in Table 6 and illustrated in Figure 59. The standard error for slope are LCZ 11 < 6 < 5 < 2 < 9, and for intercept are LCZ 11 < 6 < 5 < 9 < 2.

Table 6. Standard error of slopes and intercepts

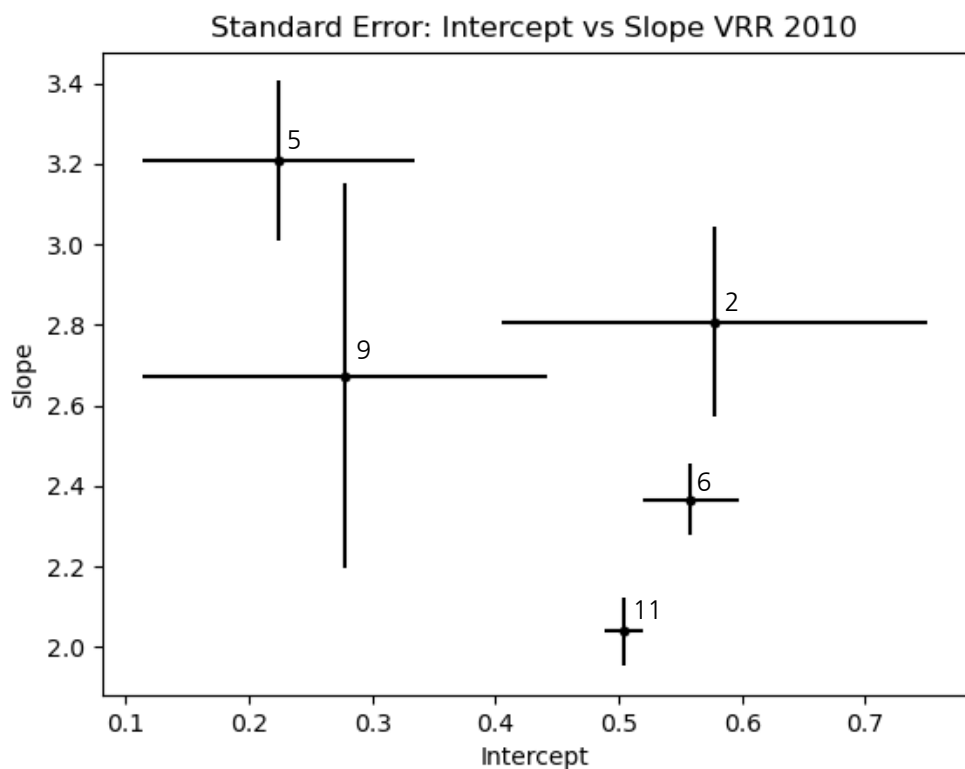| GIS-LCZ	Slope		Intercept	
	Coefficient	Standard error	Coefficient	Standard error
LCZ 2	2.807	0.235	0.578	0.173
LCZ 5	3.208	0.197	0.224	0.111
LCZ 6	2.365	0.089	0.558	0.039
LCZ 9	2.673	0.478	0.278	0.164
LCZ 11	2.038	0.086	0.504	0.016

Figure 59. Standard error of coefficients of linear regression between urban fraction and UHII

4.5 Enrichment of CityGML with the GIS-LCZ

CityGML as the information hub can be enriched with the result of the GIS-LCZ map. From the GIS-LCZ classification, every grid over Berlin comprises two main attributes: the LCZ class and the zone properties which are sky view factor (SVF), mean building height (H), aspect ratio (H/W), building surface fraction (BSF), and impervious surface fraction (ISF).

There are two main approaches, which are provided by CityGML to augment the model beyond its fundamental scope. The first approach is by defining generic objects and/or attributes, and the second one is by defining Application Domain Extensions (ADEs). In this section, these two approaches will be analyzed on how they could be applied to enrich the CityGML model with the result of GIS-LCZ.

At the end of this section, an analysis on the possibility of the new concept of CityGML 3.0 in assisting or supporting the GIS-LCZ method will also be conducted.

4.5.1 Defining generic attributes to the *LandUse* module

In this section, the generic attributes will be analyzed towards its capability on specifying the result of GIS-LCZ map for an existing CityGML class. The existing module, which is closely related to the result of GIS-LCZ, is the *LandUse* module. Figure 60 describes the UML diagram of the *LandUse* class and the generic attributes. The GIS-LCZ map can be represented with the *class* and *function* attributes. The *class* can be filled with '*Local Climate Zone*' and the *function* can be filled with the local climate zone class it belongs to.
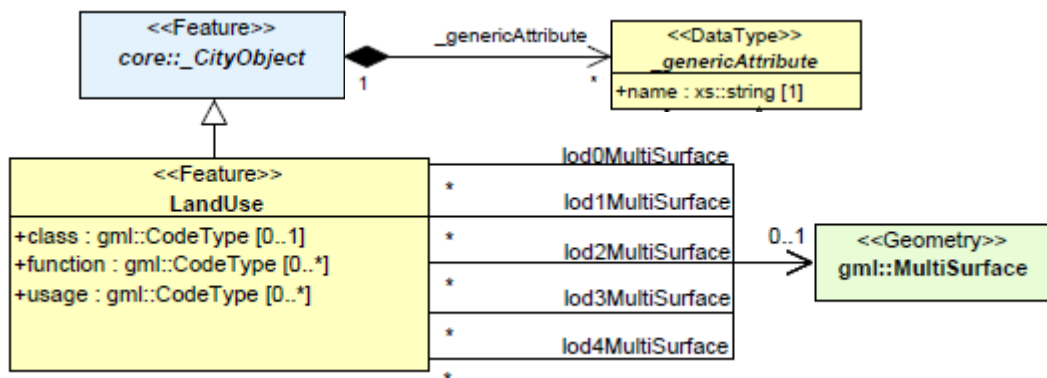

Figure 60. LandUse module with generic attribute (adopted from Gröger et al. 2012)

For the zone properties of the LCZ classes, generic attributes can be introduced. The geometry properties (street sky view factor, aspect ratio, mean building height) cannot be really represented by the definition of the *LandUse* object because they do not explicitly represent the human activities and the physical or biological cover of the earth's surface. On the other hand, the surface cover properties (building surface fraction and impervious surface fraction) are obviously related to the land cover definition of the *LandUse* object. Thus, the surface cover properties can be added to the *LandUse* class by defining the generic attributes.

There will be 2 instances of generic attribute which will be defined for GIS-LCZ. They are BSF and ISF. However, these instances cannot be modelled in the UML diagram.

The data type for these instances can also be defined as represented in Figure 13. The data type for the LCZ class would be *intAttribute* and for the zone properties would be *doubleAttribute*. The geometry of the GIS-LCZ map is in the form of polygon surface which can also be represented by the geometry of *LandUse* object. As illustrated in the UML model, *LandUse* object has the geometry of *MultiSurface, which* is derived from GML.

As shown in Figure 60, the generic attribute is related to the *CityObject* (core module) with the composition relation as shown in the UML model. The *LandUse* module inherits the definitions from the core. Thus, the land use object inherits generic attributes which allows to represent the zone properties of GIS-LCZ. However, other thematic modules will also inherit these generic attributes since they are all derived from the core module.

4.5.2 Defining a generic city object and its generic attributes

Another approach to enrich CityGML is by introducing a generic city object as well as its generic attributes. A grid tile of GIS-LCZ map can be defined as a generic city object. The attribute *class* of this object can be filled with '*Local Climate Zone*' and the *function* can be filled with the local climate zone class it belongs to. The geometry can be specified as *Surface*. There are in total 90.517 grid tiles inside the GIS-LCZ map. So there will be the same number of instances which are defined in the instance documents. However, these instances cannot be modelled in UML. The zone properties of the GIS-LCZ can be defined with the generic attributes as explained in the previous section. However, in this approach, not only surface cover fractions can be added as generic attributes, but also the rest of the zone properties: thermal and geometry properties. These properties would also be defined with *doubleAttribute* as the surface properties.

4.5.3 Defining a new Application Domain Extension (ADE)

Another possibility on defining new attributes or objects is to specify an Application Domain Extension (ADE). A UML model designed for the ADE of GIS-LCZ is depicted in Figure 61. The name of this ADE is Local Climate Zone and *lcz* is defined as a namespace. Every grid tile of the GIS-LCZ map is represented as *LocalClimateZone* feature. This feature has six attributes, which are the *zone* and its properties. The *zone* indicates the LCZ class and has value of *CodeType.* The codes and its values are listed in Table 7. After the *zone*, the five zone properties are also specified for every grid tile of the GIS-LCZ map and have a value of *double*. The six attributes of *LocalClimateZone* can only occur maximum once. Finally, the geometry of the *LocalClimateZone* feature is defined as a *Surface*. The advantage of defining ADE compared to *Generics* is that the zone properties can be modeled directly in the UML model. This will better illustrate the information comprised by the GIS-LCZ map.
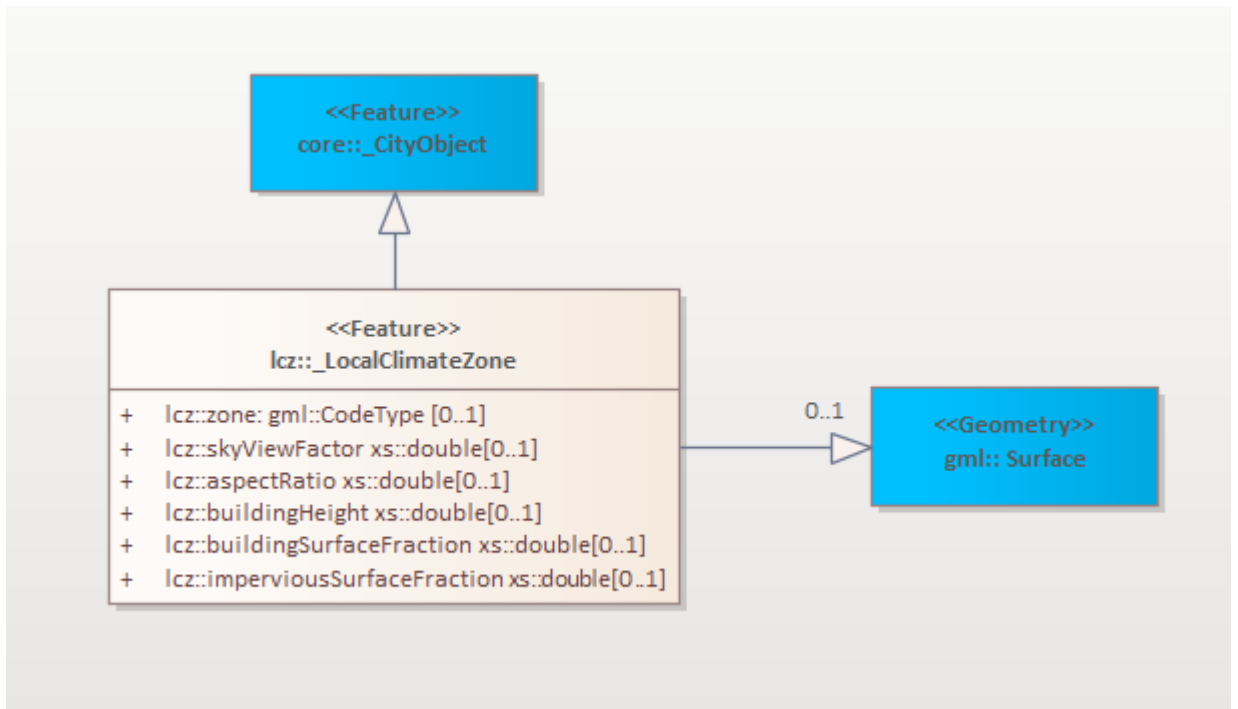
Figure 61. UML model of Local Climate Zone ADE

Table 7. Code list of zone attribute of Local Climate Zone ADE

Code list of the attribute *zone* of *LocalClimateZone*			
1	LCZ 1	7	LCZ 7
2	LCZ 2	8	LCZ 8
3	LCZ 3	9	LCZ 9
4	LCZ 4	10	LCZ 10
5	LCZ 5	11	LCZ 11
6	LCZ 6	15	LCZ 15

4.5.4 Analysis on whether the new concept of CityGML 3.0 could support the GIS-LCZ classification

The new space concept can be applied in defining local climate zones as logical spaces. The 12 zones of local climate defined by GIS-LCZ delineate Berlin area in the form of grid tiles. However, applying such definition to the zones will probably not give an impact for the classification of GIS-LCZ and its application.

The *Versioning* and *Dynamizer* modules could be used to support the GIS-LCZ classification. The changing in the city can definitely happen. New buildings are constructed and green spaces are reduced. This will lead to the change of the geometry, surface and thermal properties of a city. These are indeed the properties from local climate zones. Therefore, the changing city could also change the existing local climate zones inside the city. When a new GIS-LCZ map is derived based on these changes, the new GIS-LCZ map can be specified with the new Versioning concept.

The changes in the zone properties of local climate zones could be defined with the *Dynamizer* concept. This concept is essentially developed to enhance the application of CityGML for different types of simulations and to assist in facilitating the assimilation of sensors with 3D city models. Even though GIS-LCZ is not a result of simulation but more as input for the WRF simulation, the changes happening in the zone properties and the LCZ class can still be quantified using the dynamic concept of *Dynamizer*. The value of these properties can be specified with the *AtomicTimeseries*. By applying *Dynamizer* to the result of the GIS-LCZ map, the changes in the zone properties as well as the LCZ class can also be tracked and analyzed for further application, such as trend analysis and mitigation related to the changes of local climate.

5   Discussion

The research questions asked in the first chapter are answered in this chapter. The first and second questions are related to each other so that they will be answered in the same section. For the rest of the questions, they will be answered separately.

- **How can the LCZ classes be accurately and unambiguously classified from the calculated zone properties and what kind of algorithm can be used? Moreover, how is the accuracy of the GIS-LCZ map compared to the WUDAPT training areas?**

To know whether the classification result is accurate and unambiguous, an accuracy assessment needs to be done. The accuracy calculation is carried out with the reference to WUDAPT's training area and yields an overall accuracy of 79.05 % with the kappa coefficient of 0.73 which is considered as a good value. This kappa value shows that the classification result is not completely random or ambiguous. The highest producer accuracy is found for LCZ 11, which is 99.86%, and followed by LCZ 6 and 2 which are 96.77% and 93.65% respectively.

The producer accuracy does not show good values for LCZ 4, 8, and 9, which are 6.19%, 2.29%, and 0%, respectively. Many of LCZ 4 from WUDAPT's TA are classified as LCZ 5 in GIS-LCZ, and it is found for some tiles that WUDAPT's TA fails to detect the H and H/W properties of the selected zone, which leads to the misclassification of LCZ 5 into LCZ 4.

For LCZ 8 of WUDAPT's TA, many of the tiles are classified as LCZ 6 in GIS-LCZ and the reason identified at some tiles is due to the H/W property is more to LCZ 6 which WUDAPT fails to recognize. However, LCZ 6 does not fit for these tiles due to few of pervious surfaces shown in the satellite image. Thus, the LCZ class for these tiles lies between LCZ 6 and 8, which implies that there is not a perfect classification. The LCZ 8 is also classified as LCZ 5 and 15 in GIS-LCZ. This happens at some tiles due to the H property, which is not detected by WUDAPT to classify the area into LCZ 5. Even though the tiles are airport area (LCZ 8) as viewed in satellite imagery, but the H property does not fit LCZ 8. This implies the shortcoming of LCZ framework that cannot be ideally implemented for every city. It is also found that the size of the grid tiles of GIS-LCZ are able to classify the paved surfaces in the airport areas as paved class which is LCZ 15.

For LCZ 9 in WUDAPT's TA, there is only one polygon representing this TA and it is classified as LCZ 6 in GIS-CLZ due to the ISF value of the area is for LCZ 6. In the WUDAPT's TA, there is no LCZ 1, 3, 10 and 15 which leads to zero value of user accuracy

- **How is the agreement between the GIS-LCZ map and WUDAPT L0 map?**

The agreement found between GIS-LCZ and WUDAPT L0 for Berlin is 67.33%. The same agreement value is also found for the GIS-LCZ of Vienna when it is compared to the WUDAPT L0 (Hammerberg et al. 2018). A full disagreement between GIS-LCZ and WUDAPT L0 of Berlin is found in LCZ 1, 3, 10, and 15 since these LCZs do not exist in WUDAPT classification. The number of grid tiles belongings to these LCZs are not significant compared to the total of 90.517 tiles. LCZ 1 is 8 tiles, LCZ 3 is 15 tiles, LCZ 10 is 278 tiles, and LCZ 15 is 558 tiles. For the case of LCZ 1 which is classified into LCZ 2 and 8 in WUDAPT L0, it is found that based on the zone properties it is actually LCZ 1.

For the tiles of LCZ 3, they are all classified as LCZ 8 in WUDAPT L0. When some of these tiles are evaluated based on the zone properties, the tiles are fit to LCZ 3 rather than to LCZ 8 as classified by WUDAPT L0. For LCZ 10, they are mostly classified into LCZ 5 in WUDAPT L0. Based on the zone properties, the tiles are suitable for LCZ 10. However, based on the image, they seem to be LCZ 5. Inclusion of other zone properties such as pervious surface fraction and anthropogenic heat could give a better decision of the final LCZ class. LCZ 15 in GIS-LCZ is mostly classified as LCZ 8 in WUDAPT L0. This case has been discussed in the previous discussion of training areas.

The agreement between WUDAPT L0 and GIS-LCZ is found for the other LCZs: LCZ 2, 4, 5, 6, 8, 9, and 11. Many of LCZ 2 of WUDAPT L0 are classified as LCZ 5 in GIS-LCZ. The reason found in this thesis is that the lack of WUDAPT L0 in detecting BSF and SVF which are in the range of LCZ 5 as classified in GIS-LCZ. The effect of filtering also influences the final GIS-LCZ, in which the filter removes the correctly classified LCZ 5 in GIS-LCZ. It is also found that LCZ 2 in GIS-LCZ are classified to LCZ 5 in WUDAPT L0. After the evaluation, it is shown the lack of WUDAPT in detecting BSF and H/W to correctly classify the tiles into LCZ 2.

As in the previous discussion of training areas, many of LCZ 4 in WUDAPT L0 are also classified as LCZ 5 in GIS-LCZ. On the other hand, there are many of LCZ 5 in WUDAPT L0 which are classified as LCZ 6 in GIS-LCZ. The reason found in some of the tiles is the lack of WUDAPT in detecting H value to correctly classify the area into LCZ 6. It is also identified in some tiles, after the evaluation of the zone properties, that the tiles are actually LCZ 8 (same as the unfiltered GIS-LCZ). However, after filtering, it becomes LCZ 6 in GIS-LCZ, and it is found as LCZ 5 in WUDAPT L0. Thus, in this case none of WUDAPT L0 and GIS-LCZ are correct. There are also tiles of LCZ 5 in WUDAPT L0 but they are LCZ 8 in GIS-LCZ, which are not correctly classified by GIS-LCZ due to the unavailability of the building data to calculate the zone properties.

For LCZ 6 in WUDAPT L0, there are many of them are classified as LCZ 11 in GIS-LCZ. One reason found is misclassification done by WUDAPT L0 to classify the area without building and impervious surfaces as LCZ 6. There are also the tiles of LCZ 6 which are classified as LCZ 9 in GIS-LCZ. Based on the properties the area is more to LCZ 9 rather than LCZ 6. There are also tiles of LCZ 6 which are classified as LCZ 5 in GIS-CLZ. The significant different between the zone properties of LCZ 5 and 6 is in the H value, and in these tiles are found that the H values are for LCZ 5. Moreover, there are also the tiles of LCZ 6 which are classified as LCZ 8 in GIS-LCZ. However, the classification of GIS-LCZ here is not correct since the building data is not available for these tiles.

For LCZ 8 in WUDAPT L0, most of the tiles are classified to LCZ 6 in GIS-LCZ. One tile is identified to be LCZ 6 based on the zone properties, however the satellite imagery shows that the area has a very few of pervious surfaces that it cannot be classified as LCZ 6, but rather to LCZ 8. Inclusion of pervious surface fraction as a zone property might help to identify the class to LCZ 8.

For LCZ 9 in WUDAPT L0, the total of tiles are only 519 tiles compared to the GIS-LCZ which has 3427 tiles of LCZ 9. The agreement between them is only for 69 tiles. One of the disagreement found, where the tiles of LCZ 9 in WUDAPT L0 are classified as LCZ 11 in GIS-

LCZ. The classification from WUDAPT L0 is not correct here, since there is no building in the area and almost no ISF value but full of pervious surfaces as shown in the satellite imagery.

- **What correlation between the result of GIS-LCZ and the surface/near-surface urban heat island intensity can be found?**

The correlation analysis is done for two types of UHI, which are surface urban heat island intensity (SUHII) and near-surface air urban heat island intensity (UHII). For the correlation with SUHII, it is found that the overall correlation value for all urban fraction of GIS-LCZ has a good correlation value of 0.708 and R^2 of 0.501. However, these values are decreasing when the correlation is done separately for every LCZ class. It is concluded that the SUHII derived from LST cannot be used to distinguish between LCZs distribution. The LCZ framework is intended to study air UHI magnitude that is based on temperature measurements at a height of approximately 2 meter above ground. Land surface temperature measurement, even if it were accurate (which it is not, given that it is inferred from remote radiation readings) would not be expected to correlate with LCZ.

Therefore, when the correlation is carried out with the WRF simulation result, which is based on the urban canopy-layer air simulation, the result shows a better correlation and R^2 values, even for single LCZs. Based on the classes, LCZ 2 has the highest correlation and R^2 values which are 0.846 and 0.715 respectively. The correlation and R^2 values are LCZ 2 > 9 > 5 > 6 > 11.

- **How can the CityGML be enriched with the calculated zone properties and the result of the GIS-LCZ classification?**

CityGML can be enriched with the result of GIS-LCZ by applying three possible approaches:

- o Defining generic attributes to the *LandUse* module
  The *LandUse* module can be used to store the local climate zone classes of the grid tiles by defining the attribute *class* with 'Local Climate Zone' and the *function* with the class of local climate zone of the grid tile. However, based on the definition of the *LandUse* module, only the surface cover properties of GIS-LCZ can be integrated with this module. These properties of the GIS-LCZ can be defined with the generic attribute.

- o Defining a generic city object and its generic attributes
  A grid tile of GIS-LCZ map can be specified as a generic city object. The attribute *class* of this object can be filled with 'Local Climate Zone' and the *function* can be filled with the local climate zone class it belongs to. The geometry can be specified as *Surface* and the five zone properties can be defined with the generic attribute mechanism.

- o Defining a new Application Domain Extension (ADE)
  A new ADE can be defined to facilitate the local climate zone features. A *LocalClimateZone* class can be specified with its six attributes: zone, SVF value, H/W, H, BSF, and ISF. Attribute *zone* is defined as *CodeType* which lists the name of LCZ classes. The geometry of this feature is defined as *Surface*. The advantage of ADE is that the extension is formally described, which allows instance documents to be validated with the extended CityGML and the ADE schema.

- **How will the new concept of CityGML 3.0 support the GIS-LCZ classification?**

The two new modules, *Versioning* and *Dynamizer* of the new version of CityGML 3.0 could be used to facilitate the result of GIS-LCZ. *Versioning* can manage multiple versions of GIS-LCZ when in the future, a new GIS-LCZ classification is generated for Berlin. On the other hand, *Dynamizer* could help in quantifying the changes of LCZ class and zone properties inside the LCZ map. The values of the LCZ class and zone properties can be saved with the *AtomicTimeSeries.* Applying Dynamizer, it would be possible to track the changes of the LCZ class and its zone properties for further analysis.

6 Conclusion and outlook

6.1 Conclusion

The GIS-based method is applied to derive local climate zones over Berlin using Fuzzy logic algorithm. When compared to the training areas from WUDAPT, the overall accuracy found for the result of the classification is 79.05% with the kappa value of 0.73. Both show good values for a result of a classification. The agreement between the GIS-LCZ and WUDAPT L0 is 67.33%, and full disagreements are in LCZ 1, 3, 10, and 15 due to the nil of these classes in WUDAPT L0. However, the number of grid tiles that belong to these classes are not significant.

It is observed that WUDAPT L0 is lacking at detecting the zone properties related to building geometry (H, H/W, SVF, and BSF) which leads to the misclassification of LCZs. This case is found in WUDAPT L0 for LCZ 2, 4, and 5. WUDAPT L0 also gives a completely random classification result on LCZ 6 and 9. It is found in these classes that the tiles are actually natural class (LCZ 11) instead of urban classes, based on the zone properties and the view from the satellite imagery.

On the other hand, GIS-CLZ specifies the building geometry, which enables 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. However, GIS-LCZ is dependent to the availability of data. It is identified that some tiles cannot be classified correctly due to the inadequacy of the data to calculate the zone properties. GIS-LCZ also lacks on detecting LCZ 8, which is probably due to the unavailability of pervious surface fraction property. The majority filter applied can remove the granular view and give a more homogeneous LCZ classes. However, this filter can also diminish the correctly classified GIS-LCZ classes. Some tiles of GIS-LCZ are found between LCZ 6 and 8 which implies that the combination of two LCZs is possible to represent a local climate zone as it was also done by the other research (Chen et al. 2020). The airport areas are classified as LCZ 5 in GIS-LCZ because of building height property suits this class rather than LCZ 8. This highlights the limitation of LCZ framework, where its implementation cannot always be ideal for every city.

A correlation analysis is conducted between estimated GIS-LCZ and SUHII. It is found that the correlations of urban fraction from GIS-LCZ and SUHII over each LCZ class do not show good correlation values. Thus, it is implied that SUHII derived from land surface temperature cannot explain the distribution of LCZs. The LCZ framework is indeed not related to the measurement of surface temperature but near-surface air temperature, which is used to calculate the UHII. Therefore, the correlations carried out between urban fraction and UHII from the WRF simulation yield satisfying results for each LCZ class where the highest correlation and R^2 values are detected in LCZ 2 for 0.846 and 0.715 respectively.

The GIS-LCZ map can be further used in enhancing the CityGML model by applying the *LandUse* module, the *Generics* module or creating an ADE. The LCZ classes can be stored with the existing *LandUse* module or with a generic city object. On the other hand, the zone properties of each class can be specified with generic attributes. The

new version of CityGML 3.0 can be used to facilitate the result of GIS-LCZ by employing the *Versioning* and *Dynamizer* modules. *Versioning* is able to qualitatively describe the change of the GIS-LCZ map's version. On the other hand, *Dynamizer* can quantify the changes of LCZ classes and their zone properties within GIS-LCZ map over time.

6.2 Outlook

In this thesis, only five of the zone properties are used for the classification of GIS-LCZ. It would be interesting to add the other zone properties for the classification and to see whether the result is better. Data about plants and tree heights could be added for the future work so that the zone properties, such as H, SVF, and H/W, can be quantified to classify the natural classes (LCZ A-G).

WUDAPT method has been identified on its deficiency in detecting building height that the zone properties, such as H, H/W, and SVF are not considered. The classification in WUDAPT could be probably 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 the aggregation step from the GIS-LCZ method (Gál et al. 2015).

Using the existing 3D city model of CityGML of Berlin could be an option in the future to generate the LCZ map. The sky view factor calculator, which is developed for the purpose of solar potential analysis by the Chair of Geoinformatics at TUM, can be implemented to calculate the zone property SVF of the grid tile (Chaturvedi et al. 2017). The building height and the aspect ratio (H/W) can also be calculated with the 3D city model since the building spacing and the height of the building are given in the model. However, for a city scale, it could be computationally expensive to calculate the zone properties based on the 3D city model.

The result of GIS-LCZ can be further used for the WRF simulation to improve the spatial resolution of LULC and improve the accuracy of the simulation result. The simulation result, such as UHII, can be correlated with the LCZ classes for further analysis. Another correlation can also be carried out between the LCZ classes directly with LST without deriving it into the SUHII (Yang et al. 2021).

The simulation result of the microscale numerical model, such as from PALM4U model, can also be evaluated with the GIS-LCZ classes to assess the local climate at a very high spatial resolution of 10 m o 1 m (langer et al. 2021). The GIS-LCZ classes can also be applied to design urban temperature network for the purpose of understanding of spatial and temporal characteristics of urban climate (Unger et al. 2015). This network can be further used to analyze the air temperature conditions within the GIS-LCZ classes (Skarbit et al. 2017).

List of Figures

List of Tables

Acronyms

ADE	Application Domain Extension
BSF	Building Surface Fraction
CityGML	City Geography Markup Language
DLR	*Deutsches Zentrum für Luft- und Raumfahrt*
GIS	Geographic Information System
H	Mean Building Height
H/W	Aspect Ratio
IMD	Imperviousness Density
ISF	Impervious Surface Fraction
LCZ	Local Climate Zone
LB	Left Bound
LZB	Left Zero Bound
LST	Land Surface Temperature
OSM	OpenStreeMap
UCL	Urban Canopy Layer
UHI	Urban Heat Island
UHII	Urban Heat Island Intensity
RB	Right Bound
RZB	Right Zero Bound
SUHII	Surface Urban Heat Island Intensity
SVF	Sky View Factor
WRF	Weather Research and Forecasting
WUDAPT	World Urban Database and Access Portal Tools Level 0
WUDAPT L0	WUDAPT Level 0 (LCZ map)
WUDAPT's TA	WUDAPT's training area
UF	Urban Fraction
UML	Unified Modeling Language

References

Accuracy Metrics. [online] Accessed on 07.09.2021 from: http://gsp.humboldt.edu/olm_2019/courses/GSP_216_Online/lesson6-2/metrics.html.

Afshari, A., and Ramirez, N. Improving the accuracy of simplified urban canopy models for arid regions using site-specific prior information. Urban Clim. 2021, 35, 100722. https://doi.org/10.1016/j.uclim.2020.100722.

AρρEEARS. Application for Extracting and Exploring Analysis Ready Samples. [online] Accessed on 31.08.2021 from: https://lpdaacsvc.cr.usgs.gov/appeears/.

ArcGIS Pro. [online] Accessed on 14.09.2021 from: https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Bechtel, B. and Daneke, C. Classification of Local Climate Zones based on multiple Earth Observation data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2012, 5:1191–1202

Bechtel, B., Alexander, P.J., Böhner, J., Ching, J., Conrad, O., Feddema, J., Mills, G., See, L., and 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. https://doi.org/10.3390/ijgi4010199.

Bechtel, B., Alexander, P.J., Beck, C., Böhner, J., Brousse, O., Ching, J., Demuzere, M., Fonte, C., Gál, T., Hidalgo, J., Hoffmann, P., Middel, A., Mills, G., Ren, C., See, L., Sismanidis, P., Verdonck, M.-l., Xu, G., and Xu, Y. Generating WUDAPT Level 0 data – Current status of production and evaluation, Urban Climate, Volume 27, 2019, Pages 24-45, ISSN 2212-0955, https://doi.org/10.1016/j.uclim.2018.10.001.

Biljecki, F., Ledoux, H., and Stoter, J. An improved LOD specification for 3D building models. Computers, Environment and Urban Systems 2016, 59, 25-37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005.

BSE. Standard Error. [online] Accessed on 18.09.2021 from: https://www.statsmodels.org/stable/generated/statsmodels.regression.linear_model.OLSResults.bse.html?highlight=bse#statsmodels.regression.linear_model.OLSResults.bse.

Carrión, D., Lorenz, A., and Kolbe, T.H. Estimation of the energetic rehabilitation state of buildings for the city of Berlin using a 3D city model represented in CityGML. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2010, XXXVIII-4/W15, 31–35.

Chaturvedi, K., Willenborg, B., Sindram, M., and Kolbe, T. Solar Potential Analysis and Integration of the Time-Dependent Simulation Results for Semantic 3D City Models Using Dynamizers. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences 2017. IV-4/W5. 25-32. https://doi.org/10.5194/isprs-annals-IV-4-W5-25-2017.

Chaturvedi, K. and Kolbe, T.H. Future City Pilot 1 Engineering Report, OGC Doc. No. 19-098, 2017. [online] Accessed on 16.04.2021 from http://docs.opengeospatial.org/per/16-098.html.

Chen, Y., Zheng, B., and Hu, Y. Mapping Local Climate Zones Using ArcGIS-Based Method and Exploring Land Surface Temperature Characteristics in Chenzhou, China. Sustainability 2020, 12, 2974. https://doi.org/10.3390/su12072974.

Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X., Ren, C., Brousse, O., Martilli, A., Neophytou, M., Mouzourides, P., Stewart, I., Hanna, A., Ng, E., Foley, M., Alexander, P., Aliaga, D., Niyogi, D., Shreevastava, A., Bhalachandram, S., Masson, V., Hidalgo, J., Fung, J., Fatima-Andrade, M., Baklanov, A., Wei Dai, D., Milcinski, G., Demuzere, M., Brunsell, N., Pesaresi, M., Miao, S., Mu, Q., Chen, F., and Theeuwes, N. World Urban Database and Access Portal Tools (WUDAPT), an urban weather, climate and environmental modeling infrastructure for the Anthropocene. Bull. Am. Meteorol. Soc. 2018 (In press), https://doi.org/10.1175/BAMS-D-16-0236.1.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8, 1991–2007, 2015. https://doi.org/10.5194/gmd-8-1991-2015.

Copernicus. High Resolution Layers. [online] Accessed on 25.03.2021 from: https://land.copernicus.eu/.

Correlation. [online] Accessed on 18.09.2021 from: http://www.stat.yale.edu/Courses/1997-98/101/correl.htm.

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

Demuzere, M., Kittner, J., and Bechtel, B. LCZ Generator: a web application to create Local Climate Zone maps. Frontiers in Environmental Science 9:637455. 2021. https://doi.org/10.3389/fenvs.2021.637455.

EA. Enterprise Architect. [online] Accessed on 14.09.2021 from: https://sparxsystems.com/products/ea/.

EPA. United States Environmental Protection Agency: Heat Island impacts. [online] Accessed on 25.03.2021 from: https://www.epa.gov/heatislands/heat-island-impacts.

EPA. U.S. Environmental Protection Agency. Reducing urban heat islands: Compendium of strategies. Draft. 2008. https://www.epa.gov/heat-islands/heat-island-compendium.

Estacio, I., Babaan, J., Pecson, N. J., Blanco, A. C., Escoto, J. E., and Alcantara, C. K. GIS-based Mapping of Local Climate Zones Using Fuzzy Logic and Cellular Automata. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 2019, XLII-4/W19, 199–206, https://doi.org/10.5194/isprs-archives-XLII-4-W19-199-2019.

FIS-Broker. Geoportal Berlin. [online] Accessed on 25.03.2021 from: https://fbinter.stadt-berlin.de/fb/index.jsp.

Gál, T., Bechtel, B., and & Lelovics, E. Comparison of two different Local Climate Zone mapping methods. International Conference on Urban Climate 9 jointly with Symposium on the Urban Environment 12, 2015. http://real.mtak.hu/28577/1/GD2-6-1551002_a.pdf.

Geletič, J. and Lehnert, M. GIS-based delineation of local climate zones: The case of medium-sized Central European cities. Moravian Geographical Reports 2016. 24. 2-12. 10.1515/mgr-2016-0012.

GDAL. GDAL Documentation. [online] Accessed on 25.03.2021 from: https://gdal.org/.

GRASS QGIS. GRASS GIS Integration. [online] Accessed on 25.03.2021 from: https://docs.qgis.org/2.8/de/docs/user_manual/grass_integration/grass_integration.html.

Gröger et al. OGC City Geography Markup Language (CityGML) Encoding Standard, Version 2.0.0, 2012.

Hammerberg, K., Brousse, O., Martilli, A. and 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. https://doi.org/10.1002/joc.5447.

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. Discuss. 2020, 1–62. https://gmd.copernicus.org/preprints/gmd-2019-355/gmd-2019-355.pdf.

Hidalgo, J., Dumas, G., Masson, V., Petit, G., Bechtel, B., Bocher, E., Foley, M., Schoetter, R., and Mills, G. Comparison between local climate zones maps derived from administrative datasets and satellite observations. Urban Climate 2019. 27. 64-89. https://doi.org/10.1016/j.uclim.2018.10.004.

Hulley, G. MYD21A1D MODIS/Aqua Land Surface Temperature/3-Band Emissivity Daily L3 Global 1km SIN Grid Day V006 [Data set]. NASA EOSDIS Land Processes DAAC 2017. Accessed on 2021-09-21 from https://doi.org/10.5067/MODIS/MYD21A1D.006.

Hulley, G. MYD21A1N MODIS/Aqua Land Surface Temperature/3-Band Emissivity Daily L3 Global 1km SIN Grid Night V006 [Data set]. NASA EOSDIS Land Processes DAAC 2017. Accessed on 2021-09-21 from: https://doi.org/10.5067/MODIS/MYD21A1N.006.

langer, I., Pasternack, A., Ulbrich, U., and Rust, H. Comparison of urban climate measurements in Berlin and LES model output for a special observation period, EGU

General Assembly 2021, online, 19–30 Apr 2021, EGU21-2194, https://doi.org/10.5194/egusphere-egu21-2194.

Ishimaru, Nobuhiro; Kurokawa, Chikako; Tanaka, Yuichi; Oishi, Tomohisa; Akahoshi, Kentaro; Kutzner, Tatjana; Kolbe, Thomas H. CityGML Urban Planning ADE for i-Urban Revitalization. Open Geospatial Consortium 2020. https://portal.ogc.org/files/?artifact_id=92113.

Kanda, M., Kawai, T., Kanega, M. et al. A Simple Energy Balance Model for Regular Building Arrays. Boundary-Layer Meteorol 2005, 116, 423–443. https://doi.org/10.1007/s10546-004-7956-x.
Kolbe, T.H. Representing and exchanging 3D city models with CityGML. In 3D Geo-Information Sciences; Zlatanova, S., Lee, J., Eds.; Springer: Berlin, Germany, 2009; pp. 15–31.

Kutzner, T., Chaturvedi, K. and Kolbe, T.H. CityGML 3.0: New Functions Open Up New Applications. PFG 2020, 88, 43–61. https://doi.org/10.1007/s41064-020-00095-z.

Landsat surface temperature. [online] Accessed on 25.03.2021 from: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-surface-temperature.

Lelovics, E., Unger, J., Gál, T., and 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. https://doi.org/10.3354/cr01220.

LCZ Generator. [online] Accessed on 15.09.2021 from: https://lcz-generator.rub.de/submissions.

Lindberg, F., and Grimmond, CSB. Continuous sky view factor maps from high resolution urban digital elevation models. Clim Res 2010, 42:177-183m https://doi.org/10.3354/cr00882.

Lindberg F, Grimmond CSB, Gabey A, Huang B, Kent CW, Sun T, Theeuwes N, Järvi L, Ward H, Capel- Timms I, Chang YY, Jonsson P, Krave N, Liu D, Meyer D, Olofson F, Tan JG, Wästberg D, Xue L, Zhang Z. Urban Multi-scale Environmental Predictor (UMEP) - An integrated tool for city-based climate services. Environmental Modelling and Software.99, 70-87 (2018) https://doi.org/10.1016/j.envsoft.2017.09.020.

McCarthy, M. P., Best, M. J., and Betts, R. A. Climate change in cities due to global warming and urban effects, Geophys. Res. Lett., 2010, 37, L09705, doi:10.1029/2010GL042845.

Mills, G., Bechtel, B., Ching, J., See, L., Feddema, J., Foley, M., Alexander, P., and O'Connor, M. Introduction to the WUDAPT Project. Extended Abstracts, 9th Int. Conf. on Urban Climate/12th Symp. on the Urban Environment 2015, Toulouse, France, Météo-France, 6 pp., www.meteo.fr/icuc9/LongAbstracts/gd2-1-6521222_a.pdf.

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. https://doi.org/10.1007/s00704-019-02881-1.

MYD. MYD21A1D v006: MODIS/Aqua Land Surface Temperature/3-Band Emissivity Daily L3 Global 1 km SIN Grid Day. [online] Accessed on 30.09.2021 from: https://lpdaac.usgs.gov/products/myd21a1dv006/.

Nuruzzaman, Md. Urban Heat Island: Causes, Effects and Mitigation Measures -A Review. International Journal of Environmental Monitoring and Analysis 2015. 3. 67-73. 10.11648/j.ijema.20150302.15.

OGC. OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard. [online] Accessed on 28.09.2021 from: http://docs.ogc.org/is/20-010/20-010.html.

Oke, T.R., Mills, G., Christen, A., and Voogt, J.A. Urban Climates. Cambridge University Press: Cambridge, UK, 2017.

Oke, T.R. The energetic basis of the urban heat island. Q.J.R. Meteorol. Soc 1982, 108: 1-24. https://doi.org/10.1002/qj.49710845502.

OSM. OpenStreetMap. [online] Accessed on 25.03.2021 from: https://www.openstreetmap.org/about.

Perera, N.G.R. and Emmanuel, R. A "Local Climate Zone" based approach to urban planning in Colombo, Sri Lanka. Urban Climate 2018. 23. 188-203. https://doi.org/10.1016/j.uclim.2016.11.006.

Quan, S.J., and Bansal, P. A Systematic Review of GIS-based local climate zone mapping studies. Building and Environement 2021, 196, 107791, https://doi.org/10.1016/j.buildenv.2021.107791.

Quan, S.J., Dutt, F., Woodworth, E., Yamagata, Y., and Yang, P.P.-J. Local Climate Zone Mapping for Energy Resilience: A Fine-grained and 3D Approach. Energy Procedia 2017, 105, 3777-3783, https://doi.org/10.1016/j.egypro.2017.03.883.

QGIS. QGIS - The Leading Open Source Desktop GIS. [online] Accessed on 25.03.2021 from: https://www.qgis.org/en/site/about/index.html.

Regression. [online] Accessed on 18.09.2021 from: https://blog.minitab.com/en/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit.

Samsonov, T. and Varentsov, M. Computation of City-descriptive Parameters for High-resolution Numerical Weather Prediction in Moscow Megacity in the Framework of the COSMO Model. Russian Meteorology and Hydrology 2020. 45. 515-521. https://doi.org/10.3103/S1068373920070079.

Šećerov, I., Savić, S., Milošević, D., Marković, V., and Bajšanski, I. Development of an automated urban climate monitoring system in Novi Sad (Serbia). Geographica Pannonica 2015. 19. 174-183. 10.5937/GeoPan1504174S.

Sigma. [online] Accessed on 18.09.2021 from: https://itrevolution.com/whats-your-sigma/.

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., et al. A Description of the Advanced Research WRF Model Version 4; Technical Report NCAR/TN-556+STR; UCAR/NCAR: Boulder, CO, USA, 2019. http://dx.doi.org/10.5065/1dfh-6p97.

Skarbit, N., Stewart, I., Unger, J., and Gál, T. Employing an urban meteorological network to monitor air temperature conditions in the 'local climate zones' of Szeged, Hungary. International Journal of Climatology 37, 2017. https://doi.org/10.1002/joc.5023.

Stewart, I. D. and T. R. Oke.Local Climate Zones for Urban Temperature Studies. Bulletin of the American Meteorological Society 2012, 93, 12: 1879-1900, accessed Mar 24, 2021, https://doi.org/10.1175/BAMS-D-11-00019.1.

Strunz, G. Image Analysis and Classification [PowerPoint slides]. Earth Observation center, German Aerospace Center, DLR. Department of Aerospace and Geodesy, Technical University of Munich. 2021.

UMEP. Introduction. [online] Accessed on 25.03.2021 from: https://umep-docs.readthedocs.io/en/latest/Introduction.html.

UN. News on Urbanization. [online] Accessed on 15.09.2021 from: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

Unger, J., Lelovics, E., and Gál, T. Local Climate Zone mapping using GIS methods in Szeged. Hungarian Geographical Bulletin 2014, 63(1), 29-41. https://doi.org/10.15201/hungeobull.63.1.3.

Unger, J., Savic, S., Gál, T., Milosevic, D., Marković, V., Agnes, G., and Arsenovic, D. Urban climate monitoring networks based on LCZ concept. International Conference on Urban Climate 9 jointly with Symposium on the Urban Environment 12, 2015. https://core.ac.uk/download/pdf/35347224.pdf.

Vogel, J. and Afshari, A. Comparison of Urban Heat Island Intensity Estimation Methods Using Urbanized WRF in Berlin, Germany. Atmosphere 2020, 11, 1338. https://doi.org/10.3390/atmos11121338.

Wang, R., Ren, C., Xu, Y., Lau, K.K.-L., and 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. https://doi.org/10.1016/j.uclim.2017.10.001.

WRF. Website. [online] Accessed on 17.04.2021 from: https://www.mmm.ucar.edu/weather-research-and-forecasting-model.

Wu, Y., Sharifi, A., Yang, P., Borjigin, H., Murakami, D., and Yamagata, Y. Mapping building carbon emissions within local climate zones in Shanghai. Energy Procedia 2018. 152. 815-822. https://doi.org/10.1016/j.egypro.2018.09.195.

WUDAPT. WUDAPT website. [online] Accessed on 13.04.2021 from: https://www.wudapt.org/.

WUDAPT TA. [online] Accessed on 31.08.2021 from: https://www.wudapt.org/the-wudapt-portal/.

WUDAPT Portal. Get LCZ. [online] Accessed on 25.03.2021 from: https://wudapt.cs.purdue.edu/wudaptTools/default/getlcz.#

Yang, J., Ren, J., Sun, D., Xiao, X., Xia, C., Jin, C., and Li, X. Understanding land surface temperature impact factors based on local climate zones. Sustainable Cities and Society 2021. https://doi.org/10.1016/j.scs.2021.102818.