

## Bachelor Thesis

---

# STRUCTURAL ANALYSIS OF THE INTRA URBAN MORPHOLOGY AT A GLOBAL SCALE USING UNSUPERVISED GENERATED DATA

---

written by  
**Lukas Müller**

submitted to  
*Prof. Dr.-Ing. Andreas Schmitt*

supervised by  
*Henri Debray*

Munich,  
May 13, 2024

---

This bachelor thesis was elaborated at the German Remote Sensing Data Center (DFD) within the Earth Observation Center (EOC) of the German Aerospace Center (DLR) Oberpfaffenhofen.

Department Geo-Risks and Civil Security

Heads of Department: Prof. Dr.-Ing. Günter Strunz and Prof. Dr. Hannes Taubenböck

# Kurzfassung

---

Das Jahrhundert der Urbanisierung — ein Name, der dem 21. Jahrhundert aufgrund des Ausmaßes der aktuell stattfindenden Urbanisierung gegeben wurde, mit einer voraussichtlichen städtischen Bevölkerung von fast 7 Milliarden Menschen in 2050. Dieser Prozess stellt Herausforderungen an das Wachstum und die Struktur von Städten. Doch um diese Probleme anzugehen ist Wissen über die Struktur von Städten notwendig. Diese Arbeit bietet ein Framework, um sowohl urbane Muster, die die innerstädtische Morphologie beschreiben, als auch Stadtmodelle, die verschiedene Strukturen von Städten auf globaler Ebene beschreiben zu erhalten, basierend auf der Analyse von 1520 unterschiedlichen Städten. Im ersten Teil werden Deep Neural Networks verwendet um Patches zu clustern, die eine Landcoverklassifizierung, die „local climate zones (LCZs)“ erhalten und von Städten extrahiert werden, die durch die „morphological urban areas (MUAs)“ beschrieben werden. Um die morphologische Realität in den Prozess einzubringen, wird eine parameterbasierte Kodierung verwendet, um die Dichte und Höhe der verschiedenen Elemente der Stadtform zu beschreiben. Anschließend wird im zweiten Teil auf der Grundlage der erhaltenen Muster eine Graphdarstellung jedes Stadtgebiets erstellt und mit einem k-Means-Algorithmus geclustert. Der Abstand zwischen den Graphen wird mithilfe der „Fused Gromov-Wasserstein (FGW)-Distanz“ berechnet, die sowohl die Knotenmerkmale als auch die Struktur der Graphen nutzt. Zusätzlich werden die Clusterzentren (Barycenter) berechnet, die als Stadtmodelle jedes Clusters dienen. Das Framework ergab 66 Stadtmuster, die jeweils ein morphologisches Merkmal beschreiben, und 16 Stadtmodelle, wobei die jedem Modell zugeordneten Städte entweder eine klare geografische Verteilung oder andere Unterscheidungsmerkmale aufweisen.

# Abstract

---

The century of urbanization — a name given to the 21st century based on the amount of urbanization currently happening, with a projected urban population of almost 7 billion people in 2050. This process provides challenges to the growth and structure of cities. But to tackle these challenges, knowledge about the structure of cities is necessary. This thesis provides a framework to obtain both urban patterns to describe the intra urban morphology, and city models to describe different structures of cities on a global scale, based on the analysis of 1520 different cities. During the first part deep neural networks are utilized to cluster patches containing a land cover classification, the local climate zones (LCZs), extracted from cities delineated by the morphological urban areas (MUAs). To introduce the morphological reality into the process, a parameter based encoding is used to describe the density and height of the different elements of the urban form. Afterwards, during the second part a graph representation of the obtained patterns is created for each urban area and clustered using a k-Means algorithm. The distance between the graphs is computed using the Fused Gromov-Wasserstein (FGW) distance, which utilizes both node features and structure of the graphs. Additionally, the cluster centers (barycenter) are computed, serving as city models of each cluster. The framework resulted in 66 urban patterns, each describing a morphological feature, and 16 city models, with the cities assigned to each model having either a clear geographic distribution or other distinguishing features.

# Contents

---

<b>Erklärung</b>	<b>ii</b>
<b>Kurzfassung</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Abbreviations</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Research Framework and Question . . . . .	2
<b>2 Related Work</b>	<b>3</b>
2.1 Urban Morphology . . . . .	3
2.1.1 Urban Remote Sensing . . . . .	5
2.1.2 Urban Remote Sensing for Urban Morphology . . . . .	7
2.2 Methodological Background . . . . .	9
2.2.1 Deep Learning . . . . .	9
2.2.2 Optimal Transport . . . . .	14
<b>3 Tools and Data</b>	<b>17</b>
3.1 Tools . . . . .	17
3.1.1 Hardware . . . . .	17
3.1.2 Software . . . . .	17
3.2 Data Basis . . . . .	19
3.2.1 Local Climate Zones . . . . .	19
3.2.2 Morphological Urban Areas . . . . .	19
<b>4 Methodology</b>	<b>21</b>
4.1 Unsupervised Deep Learning Approach to find Urban Patterns . . . . .	21
4.1.1 Data Preparation . . . . .	22
4.1.2 Semantic Clustering by Adopting Nearest Neighbors . . . . .	24
4.1.3 Robust Learning for Unsupervised Clustering . . . . .	29
4.1.4 Mosaicking . . . . .	33
4.2 Using Graph Representations to find Different Models of Cities . . . . .	34

4.2.1	Data Preparation . . . . .	35
4.2.2	k-Means Clustering using a Optimal Transport Based Distance Metric . .	35
<b>5</b>	<b>Results</b>	<b>41</b>
5.1	Unsupervised Deep Learning Approach to find Urban Patterns . . . . .	41
5.2	Using Graph Representations to find Different Models of Cities . . . . .	43
<b>6</b>	<b>Discussion</b>	<b>61</b>
6.1	Unsupervised Deep Learning Approach to find Urban Patterns . . . . .	61
6.2	Using Graph Representations to find Different Models of Cities . . . . .	64
<b>7</b>	<b>Conclusion</b>	<b>67</b>
	<b>Bibliography</b>	<b>68</b>
<b>A</b>	<b>Additional Information</b>	<b>75</b>
A.1	Local Climate Zone Properties . . . . .	75
A.2	Image Augmentations . . . . .	76
A.3	Overview of all Experiments . . . . .	79
<b>B</b>	<b>Overview Results</b>	<b>80</b>
B.1	Overview Patches . . . . .	80
B.2	Two Random Cities of each City Model . . . . .	112
B.3	Overview Urban Areas . . . . .	128

# List of Figures

---

1.1	Development of Urban and Rural Population . . . . .	1
2.1	Three City Models . . . . .	4
2.2	Seven City Types . . . . .	8
2.3	Urban Patterns of Morphological Configurations . . . . .	8
2.4	The Perceptron . . . . .	9
2.5	Multilayer Perceptron . . . . .	10
2.6	Three Activation Functions . . . . .	11
2.7	Convolutional Neural Network . . . . .	12
2.8	Convolution and Pooling Operations . . . . .	13
2.9	Residual Blocks . . . . .	13
2.10	Wasserstein and Gromov-Hausdorff Distance . . . . .	15
3.1	Study Sites . . . . .	19
3.2	Example of one Study Site . . . . .	20
4.1	Workflow to find Urban Patterns . . . . .	21
4.2	Overview of Data for Guangzhou . . . . .	22
4.3	Overview of Patches . . . . .	23
4.4	Example of Encoding . . . . .	24
4.5	ResNet-50 and Contrastive Head used for Pretext Task . . . . .	25
4.6	SimCLR Components . . . . .	25
4.7	ResNet-50 and Cluster Head used for Clustering . . . . .	26
4.8	SCAN Components . . . . .	26
4.9	Overview of Nearest Neighbors . . . . .	27
4.10	Sample Confidences after second Step of the SCAN Framework . . . . .	28
4.11	Self-label Components . . . . .	29
4.12	Sample Confidences after final Step of the SCAN Framework . . . . .	30
4.13	RUC Components . . . . .	30
4.14	Sample Confidences after RUC Framework . . . . .	33
4.15	Workflow to find Models of Cities . . . . .	34
4.16	Paris with Urban patterns and its Graph Representations . . . . .	36
4.17	Gap Statistic Results . . . . .	39
5.1	Overview of Patches in $c_{UB_3}$ . . . . .	41
5.2	Overview of Patches in $c_{UB_{24}}$ . . . . .	42
5.3	Overview of Patches in $c_{UB_{62}}$ . . . . .	42
5.4	Overview of all found City Clusters . . . . .	44
5.5	Overview of Cities in $c_{CM_1}$ . . . . .	45
5.6	Overview of Cities in $c_{CM_2}$ . . . . .	46

5.7	Overview of Cities in $c_{CM_3}$	47
5.8	Overview of Cities in $c_{CM_4}$	48
5.9	Overview of Cities in $c_{CM_5}$	49
5.10	Overview of Cities in $c_{CM_6}$	50
5.11	Overview of Cities in $c_{CM_7}$	51
5.12	Overview of Cities in $c_{CM_8}$	52
5.13	Overview of Cities in $c_{CM_9}$	53
5.14	Overview of Cities in $c_{CM_{10}}$	54
5.15	Overview of Cities in $c_{CM_{11}}$	55
5.16	Overview of Cities in $c_{CM_{12}}$	56
5.17	Overview of Cities in $c_{CM_{13}}$	57
5.18	Overview of Cities in $c_{CM_{14}}$	58
5.19	Overview of Cities in $c_{CM_{15}}$	59
5.20	Overview of Cities in $c_{CM_{16}}$	60
A.1	Overview of Augmentations	78
B.1	Overview of Patches in $c_{UB_1}$	80
B.2	Overview of Patches in $c_{UB_2}$	80
B.3	Overview of Patches in $c_{UB_4}$	81
B.4	Overview of Patches in $c_{UB_5}$	81
B.5	Overview of Patches in $c_{UB_6}$	82
B.6	Overview of Patches in $c_{UB_7}$	82
B.7	Overview of Patches in $c_{UB_8}$	83
B.8	Overview of Patches in $c_{UB_9}$	83
B.9	Overview of Patches in $c_{UB_{10}}$	84
B.10	Overview of Patches in $c_{UB_{11}}$	84
B.11	Overview of Patches in $c_{UB_{12}}$	85
B.12	Overview of Patches in $c_{UB_{13}}$	85
B.13	Overview of Patches in $c_{UB_{14}}$	86
B.14	Overview of Patches in $c_{UB_{15}}$	86
B.15	Overview of Patches in $c_{UB_{16}}$	87
B.16	Overview of Patches in $c_{UB_{17}}$	87
B.17	Overview of Patches in $c_{UB_{18}}$	88
B.18	Overview of Patches in $c_{UB_{19}}$	88
B.19	Overview of Patches in $c_{UB_{20}}$	89
B.20	Overview of Patches in $c_{UB_{21}}$	89
B.21	Overview of Patches in $c_{UB_{22}}$	90
B.22	Overview of Patches in $c_{UB_{23}}$	90
B.23	Overview of Patches in $c_{UB_{25}}$	91
B.24	Overview of Patches in $c_{UB_{26}}$	91
B.25	Overview of Patches in $c_{UB_{27}}$	92



---

B.26 Overview of Patches in $c_{UB_{28}}$ . . . . .	92
B.27 Overview of Patches in $c_{UB_{29}}$ . . . . .	93
B.28 Overview of Patches in $c_{UB_{30}}$ . . . . .	93
B.29 Overview of Patches in $c_{UB_{31}}$ . . . . .	94
B.30 Overview of Patches in $c_{UB_{32}}$ . . . . .	94
B.31 Overview of Patches in $c_{UB_{33}}$ . . . . .	95
B.32 Overview of Patches in $c_{UB_{34}}$ . . . . .	95
B.33 Overview of Patches in $c_{UB_{35}}$ . . . . .	96
B.34 Overview of Patches in $c_{UB_{36}}$ . . . . .	96
B.35 Overview of Patches in $c_{UB_{37}}$ . . . . .	97
B.36 Overview of Patches in $c_{UB_{38}}$ . . . . .	97
B.37 Overview of Patches in $c_{UB_{39}}$ . . . . .	98
B.38 Overview of Patches in $c_{UB_{40}}$ . . . . .	98
B.39 Overview of Patches in $c_{UB_{41}}$ . . . . .	99
B.40 Overview of Patches in $c_{UB_{42}}$ . . . . .	99
B.41 Overview of Patches in $c_{UB_{43}}$ . . . . .	100
B.42 Overview of Patches in $c_{UB_{44}}$ . . . . .	100
B.43 Overview of Patches in $c_{UB_{45}}$ . . . . .	101
B.44 Overview of Patches in $c_{UB_{46}}$ . . . . .	101
B.45 Overview of Patches in $c_{UB_{47}}$ . . . . .	102
B.46 Overview of Patches in $c_{UB_{48}}$ . . . . .	102
B.47 Overview of Patches in $c_{UB_{49}}$ . . . . .	103
B.48 Overview of Patches in $c_{UB_{50}}$ . . . . .	103
B.49 Overview of Patches in $c_{UB_{51}}$ . . . . .	104
B.50 Overview of Patches in $c_{UB_{52}}$ . . . . .	104
B.51 Overview of Patches in $c_{UB_{53}}$ . . . . .	105
B.52 Overview of Patches in $c_{UB_{54}}$ . . . . .	105
B.53 Overview of Patches in $c_{UB_{55}}$ . . . . .	106
B.54 Overview of Patches in $c_{UB_{56}}$ . . . . .	106
B.55 Overview of Patches in $c_{UB_{57}}$ . . . . .	107
B.56 Overview of Patches in $c_{UB_{58}}$ . . . . .	107
B.57 Overview of Patches in $c_{UB_{59}}$ . . . . .	108
B.58 Overview of Patches in $c_{UB_{60}}$ . . . . .	108
B.59 Overview of Patches in $c_{UB_{61}}$ . . . . .	109
B.60 Overview of Patches in $c_{UB_{63}}$ . . . . .	109
B.61 Overview of Patches in $c_{UB_{64}}$ . . . . .	110
B.62 Overview of Patches in $c_{UB_{65}}$ . . . . .	110
B.63 Overview of Patches in $c_{UB_{66}}$ . . . . .	111
B.64 Overview of two Cities in $c_{CM_1}$ . . . . .	112
B.65 Overview of two Cities in $c_{CM_2}$ . . . . .	113
B.66 Overview of two Cities in $c_{CM_3}$ . . . . .	114

---

B.67 Overview of two Cities in $c_{CM_4}$ . . . . .	115
B.68 Overview of two Cities in $c_{CM_5}$ . . . . .	116
B.69 Overview of two Cities in $c_{CM_6}$ . . . . .	117
B.70 Overview of two Cities in $c_{CM_7}$ . . . . .	118
B.71 Overview of two Cities in $c_{CM_8}$ . . . . .	119
B.72 Overview of two Cities in $c_{CM_9}$ . . . . .	120
B.73 Overview of two Cities in $c_{CM_{10}}$ . . . . .	121
B.74 Overview of two Cities in $c_{CM_{11}}$ . . . . .	122
B.75 Overview of two Cities in $c_{CM_{12}}$ . . . . .	123
B.76 Overview of two Cities in $c_{CM_{13}}$ . . . . .	124
B.77 Overview of two Cities in $c_{CM_{14}}$ . . . . .	125
B.78 Overview of two Cities in $c_{CM_{15}}$ . . . . .	126
B.79 Overview of two Cities in $c_{CM_{16}}$ . . . . .	127

# List of Tables

---

2.1	The Local Climate Zones . . . . .	6
2.2	The ResNet-Architectures . . . . .	14
3.1	Overview of used Python Packages . . . . .	18
A.1	Values of Properties for Local Climate Zones . . . . .	75
A.2	$T_{pre}$ Augmentations . . . . .	76
A.3	$T_{scan}/T_{self}$ Augmentations . . . . .	76
A.4	$T_{weak}$ Augmentations . . . . .	76
A.5	$T_{strong}$ Augmentations . . . . .	77
A.6	Randaugment Augmentations . . . . .	77
A.7	Overview of Experiments to find Urban Patterns . . . . .	79
A.8	Overview of Experiments to cluster Graphs . . . . .	79
B.1	Overview of all Morphological Urban Areas . . . . .	128

# List of Abbreviations

---

<b>LCZ</b>	local climate zone
<b>MUA</b>	morphological urban area
<b>ANN</b>	artificial neural network
<b>MLP</b>	multilayer perceptron
<b>ReLU</b>	rectified linear unit
<b>BN</b>	batch normalization
<b>CNN</b>	convolutional neural network
<b>ResNet</b>	residual neural network
<b>OT</b>	optimal transport
<b>SCAN</b>	semantic clustering by adopting nearest neighbors
<b>RUC</b>	robust learning for unsupervised clustering
<b>SimCLR</b>	simple framework for contrastive learning
<b>IQR</b>	interquartile range
<b>FGW</b>	Fused Gromov-Wasserstein

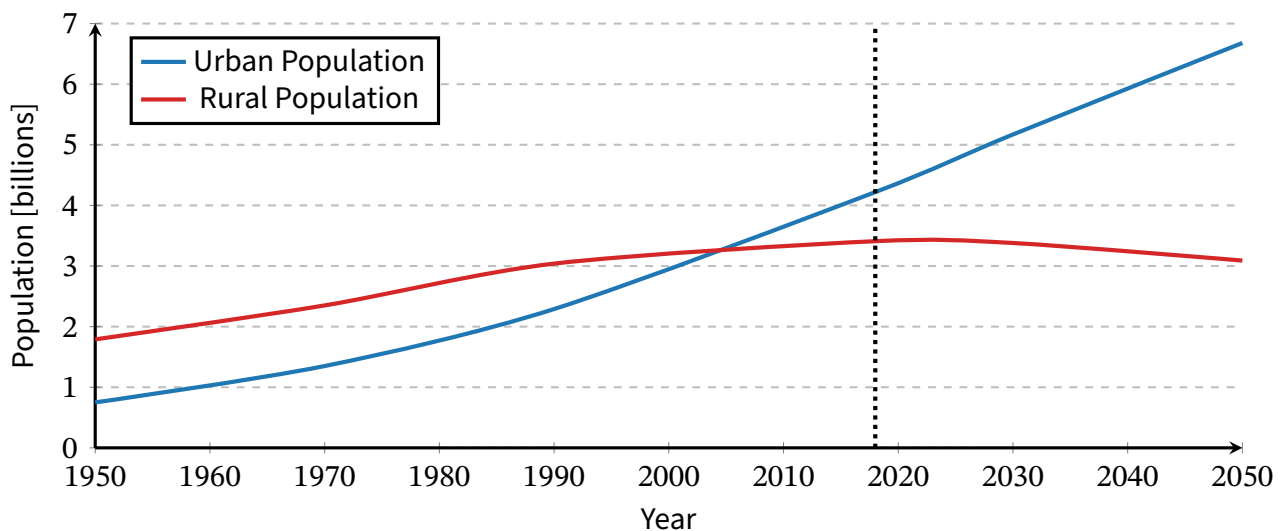
# 1 Introduction

---

While the title — Structural analysis of the intra urban morphology at a global scale using unsupervised generated data — seems quite long, it perfectly describes the scope of this thesis: To first discover intra urban patterns in over 1500 cities around the globe with the means of unsupervised deep learning. Afterwards, to create a graph representation of each city based on the found patterns and, once again through unsupervised learning, to find different models of cities. But before this can be discussed, an overview about the motivation of this work and why these topics are important will be presented.

## 1.1 Motivation

The century of urbanization — a fitting name given to the 21st century (OECD, 2015, p. 15). After all, by 2050 over 6 billion people, more than two-thirds of the human population will live in cities (UN DESA, 2019, p. 9), an increase of over 5 billion people in comparison to 1950 (Figure 1.1). But this enormous increase also brings challenges with it, as already “most cities are far from meeting the triple objective of being economically productive, socially inclusive and environmentally sustainable” (Sachs et al., 2019).



**Figure 1.1** Development of both urban and rural population from 1950 to 2018 and projected afterwards until 2050 (own illustration after UN DESA, 2019). Urban population surpassed rural population in the early 2000s. Future population growth will mainly occur in cities, while the rural population begins to decline.

According to the World Bank (2016, p. 240), this rapid urbanization “creates urgency to get our cities ‘right’ because the global response to our most pressing challenges—from climate change to rising inequality—will likely succeed or fail in cities”. But this transformation will be challenging, especially because it happens largely in emerging economies without government capacities to regulate and plan these processes, which will lead to results contrary to the objectives mentioned above (De Souza, 2018).

However, to counter these problems, knowledge on the urban structure is needed, which is provided by the field of Urban Morphology. Important areas inside this field are both urban forms and functions, with the analysis of the urban form conducted on a city scale as well as on a smaller scale (districts and neighborhoods). Similar levels are also used to analyze urban functions (Živković, 2019).

Based on that it is clear, that the analysis of both intra urban patterns and city models is very important. Although studies to both subjects already exist (which will be discussed later), this thesis offers a new approach to those topics.

## 1.2 Research Framework and Question

As discussed earlier this thesis consists of two parts: To find intra urban patterns in cities and afterwards, based on a graph representation of those patterns, to find different models of cities.

The first part is based on the master thesis of Gassilloud (2022), in which he developed an unsupervised framework to find urban patterns based on the Local climate zone (LCZ) classification developed by Stewart and Oke (2012). The boundaries of those analyzed cities were described by the Morphological urban areas (MUAs) developed by Taubenböck, Weigand, et al. (2019). This thesis adapts the framework by using a *parameter based encoding*, which introduces geographical and morphological information into the unsupervised process.

Afterwards a *graph representation* for each city based on the discovered patterns will be created. To group those cities into different types, a clustering using a *Optimal transport* based distance metric (Vayer et al., 2019) will be performed. Additionally, for each cluster, the *barycenter* (centroid of a cluster) will be computed, with the aim to produce a model of each city type.

Based on this framework, there are two major research questions to be answered:

1. *Is it possible to find urban patterns through the use of an unsupervised framework using a parameter based encoding?*
2. *Is it possible to discover different models of cities by using graph representations of the previous identified patterns?*

The structure of the remaining thesis is as following: Chapter 2 provides an overview of existing research as well as the methodological background. Chapter 3 presents the used hard- and software as well as the data basis. Chapter 4 describes the methodology, the results of which are presented in Chapter 5. Finally, those results are discussed in Chapter 6 and Chapter 7 gives a final conclusion as well as an outlook.

# 2 Related Work

---

This chapter presents an overview on the theoretical background of this thesis. It first provides an overview on urban morphology in general and its applications on urban patterns and city models in detail, as well as an overview of urban remote sensing. Additionally, it provides an outline of the possibilities of using urban remote sensing for urban morphology. Afterwards, it introduces the main concepts on which the methodology is based, namely deep learning and optimal transport.

## 2.1 Urban Morphology

Urban morphology is defined by Kropf (2017, p. 9) as follows:

... the study of human settlements, their structure and the process of their formation and transformation, ... a wide-ranging interdisciplinary field contributing to both academic research and professional practice in the built environment, ... concerned with the form and structure of cities, towns and villages, the way that they grow and change and their characteristics.

It is able to provide a clear description of the city and tools to understand and compare built environments and thus can provide valuable information to geographers, city planners and historians (Kropf, 2017, p. 10). The main elements of the studied area, the urban form, are described by Oliveira (2016, pp. 7—30) in increasing order of resolution:

- *The Urban Tissue* — The composition of a city at a general level, constituted by streets, blocks, plots, and buildings, which are combined in a specific way for each city.
- *The Natural Context* — The impact of the land relief, climate, and natural landscape on the establishment of settlements, from foundation to path, streets, and subdivision of the land.
- *The Street System* — The most stable element of urban form, defining street blocks and distinguishing public and private space.
- *The Plots System* — An essential element of the urbanization process with high stability over time, separating public and private spaces, but often overlooked by stakeholders.
- *The Buildings System* — The (possible) most visible element of urban form, consisting of two different types: ordinary buildings, which are making up most buildings of a city and having more similarities than differences, and exceptional buildings, which are clearly distinguishable in the urban landscape.

While those elements are described using a *top-down approach*, other definitions of the urban form are possible, with one example described by Wentz et al. (2018), in which the urban form is separated into three main components:

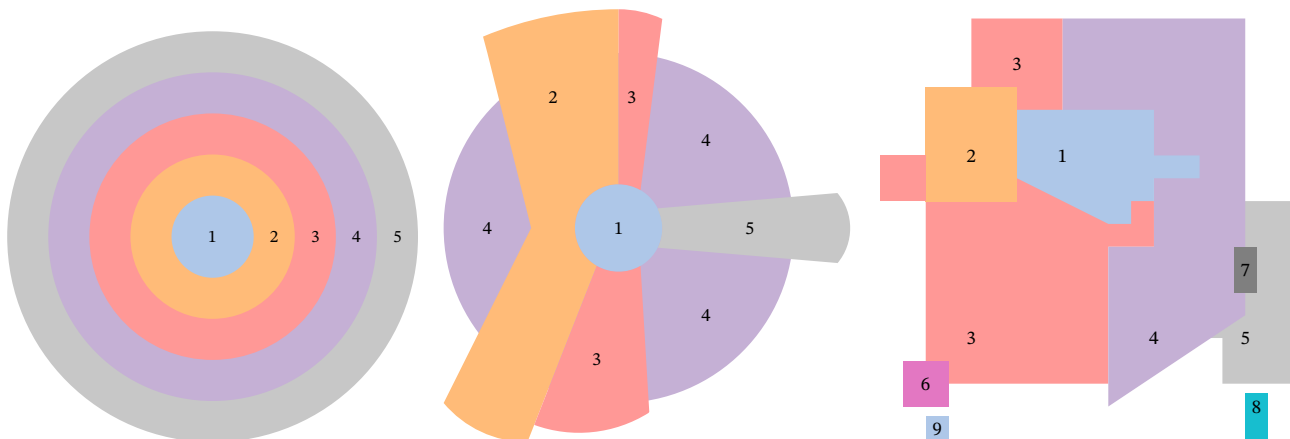
- Materials
  - Human constructions: buildings, roads, above ground utilities

- Soil-Plant continuum: organic matter (living and dead), soil processes
- Surface waters: streams, canals, lakes, ponds
- Configuration
  - Dimensionality: height information of the material elements
  - Spatial pattern: arrangement of the material elements
- Time

Those different entities are interacting with each other and thus are able to generate urban patterns (Huynh, 2019), which are then able to form a city (Kivell, 1992, p. 1).

### Urban Patterns and City Models

A pattern is defined as “a form or model proposed for imitation” (Merriam-Webster.com, 2024), meaning a kind of “blueprint” on which other entities are constructed. In the context of urban morphology, those patterns can take different forms such as land use patterns (Tsilimigkas, Stathakis, and Pafi, 2016) or urban growth patterns (Reis, Silva, and Pinho, 2016).



**(a)** The concentric city model developed by R. E. Park and Burgess ([1925] 2019, pp. 50—52). It consists of a *Central business district* (1), a *Zone of transition* (2), a *Zone of independent workers homes* (3), a *Zone of better residences* (4), and a *Commuter's zone* (5), all arranged in a concentric way.

**(b)** The sector model developed by Hoyt (1939). From the *Central business district* (1) originate *Transportation and industry* (2), *Low-class residential* (3), *Middle-class residential* (4), and *High-class residential* (5) districts as sectors.

**(c)** The multiple nuclei model developed by Harris and Ullman (1945). While some of those nuclei have existed from the city origins, others have developed over time. While the exact composition differs from city to city, most cities developed following districts: *Central business district* (1), *Light manufacturing* (2), *Low-class residential* (3), *Middle-class residential* (4), *High-class residential* (5), *Heavy manufacturing* (6), *Outlying business district* (7), *Residential suburb* (8), and *Industrial suburb* (9).

**Figure 2.1** Three different city models (own illustrations after Rubenstein (2020, pp. 468—469)).

As discussed earlier, urban patterns are able to form a city, which can be described using a city model, with historically three models standing out. One of the earliest models was the *concentric city model* developed by R. E. Park and Burgess ([1925] 2019, pp. 50—52), which consists of multiple homogeneous zones arranged in a concentric way (Figure 2.1a). This layout is the result of an expansion process, in which each inner zone extends its area in the next outer zone. As an example, applied to Chicago, all four outer zones were historically included in the area of the inner zone and only moved outwards over time. While this model was able to influence other models (like the sector model and the multiple nuclei model discussed below), it was also subjugated to



criticism, one example including “the fact that various cities do not actually conform to an ideal circular spatial pattern” (Quinn, 1940).

Another model is the *sector model* developed by Hoyt (1939) that has, while still retaining the CBD in a circular form, other homogeneous zones emerging from it as sectors (Figure 2.1b). Those sectors are again created during the expansion of the city with activities growing outwards from the center, instead of replacing the outer zones (Rubenstein, 2020, p. 468).

The third model is the *multiple nuclei model* developed by Harris and Ullman (1945). Unlike the other models, this one is not built around a single center but instead around different nuclei (Figure 2.1c). While those have existed sometimes from the very beginnings, they developed over time during the growth of the city in other cases. Reasons for the development of the nuclei is based on a combination of different factors, mostly the facts that some activities require specialized facilities as well as the grouping or antagonism of specific activities.

But there are also modern approaches to city models, as discussed by Heineberg et al. (2022, pp. 77—79), with cities divided by their locations, functions, history, and culture. One of those approaches was developed by Lemoine-Rodríguez, Inostroza, and Zepp (2020), in which four different types of urban form were found.

However, all the mentioned different approaches have one thing in common: before the urban form can be studied, the appropriate data needs to be collected, which happened traditionally through modes like physical surveys or censuses (Crooks et al., 2016). There are, however, different modern approaches to it, such as the use of “user-generated big data” (Crooks et al., 2016) or remote sensing (Bhatta, 2010).

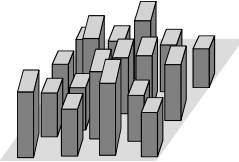
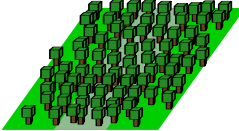
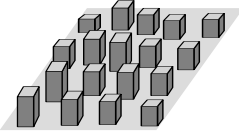
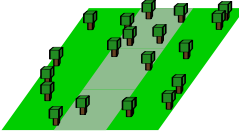
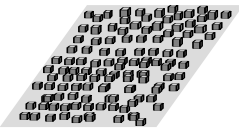
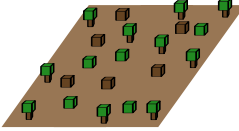
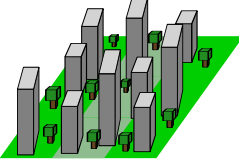

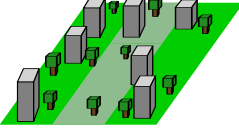

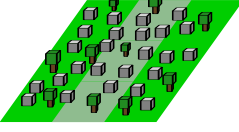

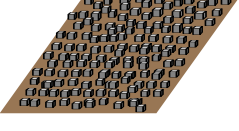


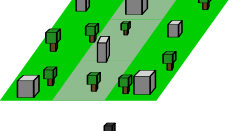
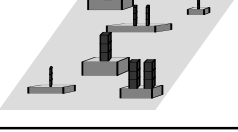
### 2.1.1 Urban Remote Sensing

As defined by NOAA (2024) “Remote sensing is the science of obtaining information about objects or areas from a distance, typically from aircraft or satellites”. Urban remote sensing is thus defined as the use of remote sensing in urban areas (Yang, 2021), a field that has made several advances over the last years. Those are based on the improvement of data availability, e.g., high resolution visual and radar imagery at no charge and inexpensive drone usage (Yang, 2021, p. 5), as well as advances in methodology on data fusion (A. Schmitt, Wendleder, et al., 2020; M. Schmitt and Zhu, 2016) and classification through the use of artificial intelligence and ensemble classifiers (Yang, 2021, p. 7).

Those advances allow to tackle a wide range of topics, with some examples such as:

- Delineating the urban settlements on a global scale (Marconcini et al., 2020) using ensemble classifiers on radar and optical imagery times series
- Spatialization of urban population with different classifiers on nighttime and optical imagery as well as elevation data (M. He, Y. Xu, and N. Li, 2020)
- Mapping of slum areas through various methods, such as using radar data (A. Schmitt, Sieg, et al., 2018) or optical imagery (Wurm and Taubenböck, 2018)

**Table 2.1** The 17 Local climate zones (LCZs) (own illustrations after Stewart and Oke, 2012): LCZ 1 — 10 represent built up types, while LCZ A — G represent non-built/land cover types.

Built types		Land cover types	
	LCZ 1 Compact high-rise		LCZ A Dense trees
	LCZ 2 Compact mid-rise		LCZ B Scattered trees
	LCZ 3 Compact low-rise		LCZ C Bush, scrub
	LCZ 4 Open high-rise		LCZ D Low plants
	LCZ 5 Open mid-rise		LCZ E Bare rock or paved
	LCZ 6 Open low-rise		LCZ F Bare soil or sand
	LCZ 7 Lightweight low-rise		LCZ G Water
	LCZ 8 Large low-rise		
	LCZ 9 Sparsely built		
	LCZ 10 Heavy industry		

- Damage assessment after floods (Jiménez-Jiménez et al., 2020) or earthquakes (Dell’Acqua and Gamba, 2012)
- Monitoring of pollution by combining optical imagery and dust measurement (Bakaeva and Le, 2022)
- Analysis of urban climate by classifying urban heat islands (Lemoine-Rodríguez, Inostroza, and Zepp, 2022b) and urban vegetation (Barbierato et al., 2020)

One other important aspect of not only urban remote sensing, but also remote sensing in general, is the development of land cover classifications (Phiri and Morgenroth, 2017). One corresponding product in the urban area are the Local climate zones (LCZs) developed by Stewart and Oke (2012). Although this classification was originally designed for urban temperature studies, it also offers a great value for the study of the intra urban morphology, as each of the 17 zones (Table 2.1) is defined as a “region of uniform surface cover, structure, material, and human activity that span hundreds of meters to several kilometers in horizontal scale” (Stewart and Oke, 2012). Additionally, for each class there are multiple geometrical and morphological parameters defined (Appendix A.1).

Multiple attempts to produce LCZ mapping for urban areas exist. Some of those will be, in addition with other applications of urban remote sensing for urban morphology, presented next.

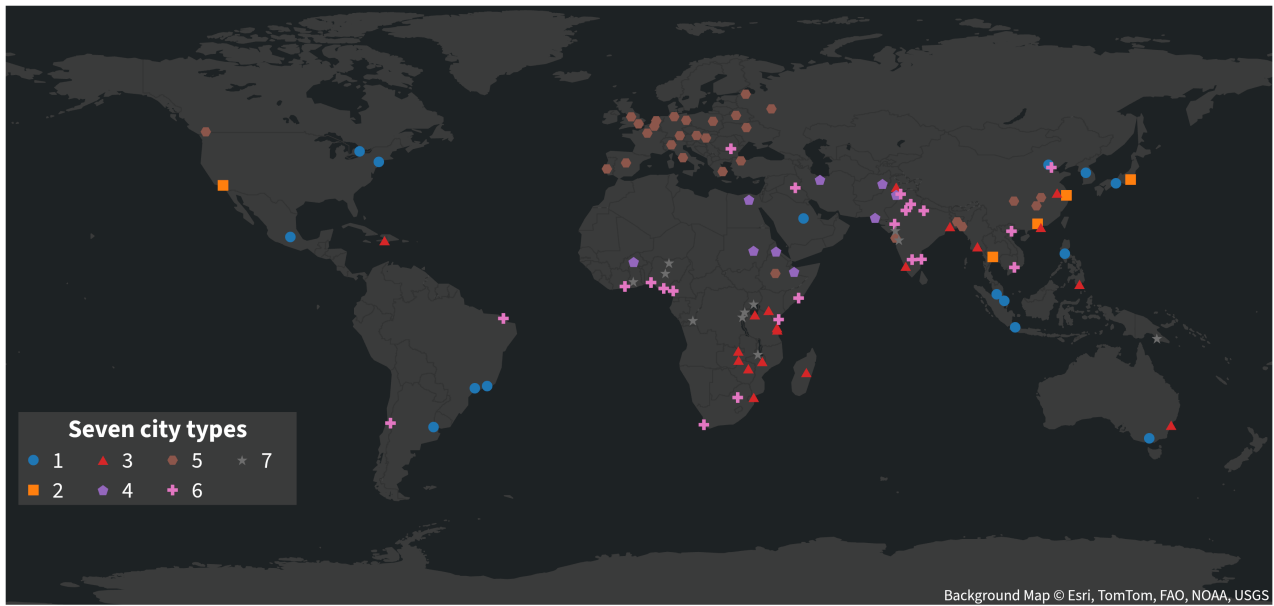
### 2.1.2 Urban Remote Sensing for Urban Morphology

There exist multiple remote sensing based LCZ mapping methods (Huang et al., 2023). One of the first attempts to map LCZs using a supervised, pixel based approach was developed by Bechtel et al. (2015). This attempt showed also difficulties for mapping LCZs, as the individual classes have dissimilar spectral properties across different regions. Another approach was developed by Qiu, M. Schmitt, and Zhu (2019) using a residual neural network to produce land cover maps of several cities with an accuracy of 86.7%. Additionally, Demuzere et al. (2022) created a global map of LCZs with an accuracy over 70% using a pixel based random forest approach.

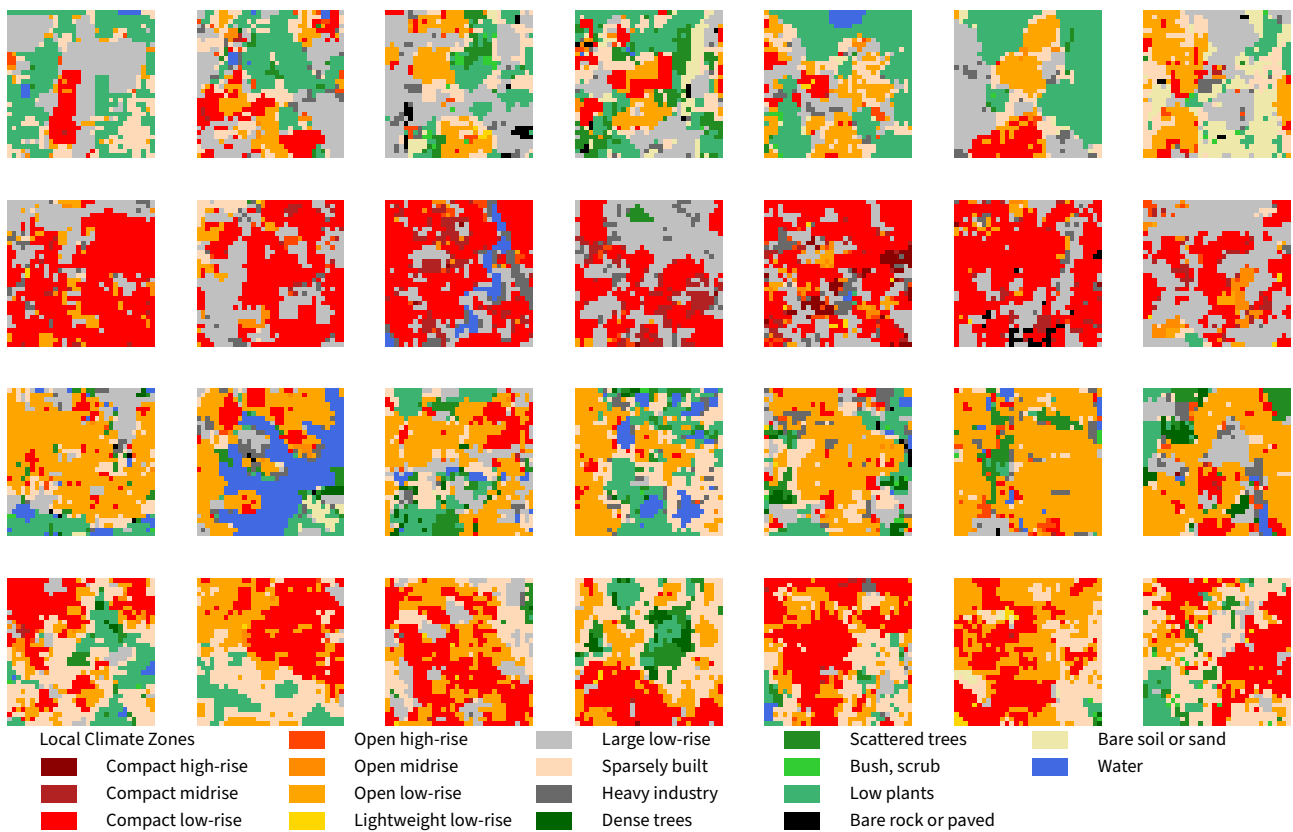
While those products help to understand the *composition* of cities, additional data is needed to *delineate* them. One approach to that was developed by Taubenböck, Weigand, et al. (2019), who aimed to show the discrepancy of administrative boundaries and real morphological configurations, resulting in the *Morphological urban areas (MUAs)*.

Using both the MUA and the LCZ mapping produced by Qiu, M. Schmitt, and Zhu (2019), Taubenböck, Debray, et al. (2020) developed a clustering process of 110 cities based on “the morphological-spatial configurations of urban landscapes”. During this process seven different city types were found (Figure 2.2):

1. Asian and American cities
2. Very large cities
3. Eastern African and Eastern Asian cities
4. Cities of the Islamic world
5. European cities



**Figure 2.2** An overview of 110 different urban areas, clustered in 7 different city types sharing morphological configurations (reprinted from Taubenböck, Debray, et al., 2020)



**Figure 2.3** An overview of four different urban patterns (one per row), each representing an urban morphological configuration, discovered through unsupervised learning (adapted from Gassilloud, 2022).

6. Asian and African cities
7. Cities of Central Africa

Another study, using the same dataset, but this time focusing on the intra urban morphology, was developed by Gassilloud (2022). In this case, patches of uniform size were extracted from the LCZ mapping inside the MUAs, with the aim to cluster them. In this case 138 different urban pattern types were found, with a subset of those shown in Figure 2.3.

As stated above, this thesis has similar aims (to find both intra urban patterns and models of cities). But to do this, two different approaches are needed, whose background will be discussed next.

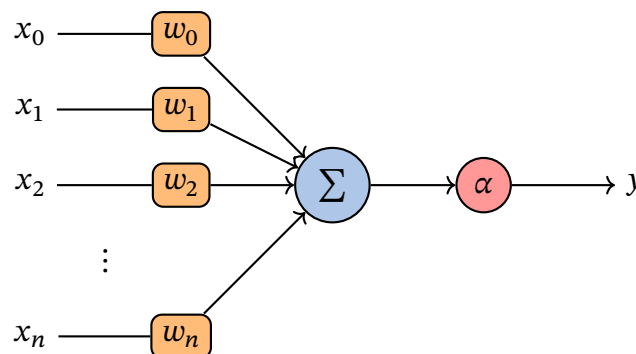
## 2.2 Methodological Background

As discussed earlier, the thesis consists of two different parts, with both having different methodological backgrounds, respectively *Deep Learning* and *Optimal Transport*.

### 2.2.1 Deep Learning

Deep learning is defined as a subset of machine learning utilizing *artificial neural networks (ANNs)* with multiple layers (LeCun, Bengio, and G. Hinton, 2015). Each layer consists of at least one artificial neuron inspired by biological nerve cells (Graupe, 2013, pp. 5—8). With the first principles of ANNs already described in the middle of the 20th century by McCulloch and Pitts (1943), deep learning today is a powerful tool with many applications.

#### The Perceptron



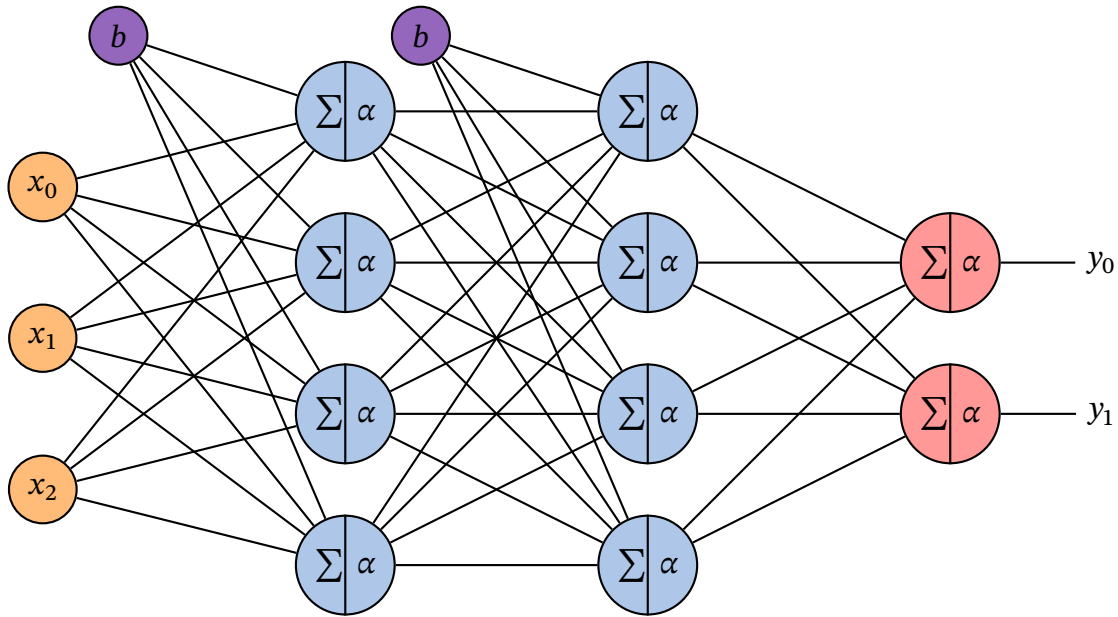
**Figure 2.4** The perceptron as developed by Rosenblatt (1958) (own illustration). The output ( $y$ ) is determined by applying the activation function ( $\alpha$ ) on weighted sum of the inputs ( $x_n$ ).

The first ANN implemented was the *perceptron* developed by Rosenblatt (1958), the design in itself quite simple (Figure 2.4). The weighted sum of the input vector  $X = [x_0, x_1, \dots, x_n]$  is, based on the weights  $W = [w_0, w_1, \dots, w_n]$ , computed. Afterwards the activation function  $\alpha$  is applied on the sum to compute the output  $y$  (Equation (2.1)). The decisive element to determine the output are thus the weights, which are discovered using a training process (Goodfellow, Bengio, and Courville, 2016, p. 13) as discussed below.

$$y = \alpha\left(\sum_{i=0}^n x_i \cdot w_i\right) \quad (2.1)$$

This perceptron, while in itself limited in its applications, remains the foundation of bigger ANN, such as in *multilayer perceptrons (MLPs)*.

### Multilayer Perceptrons



**Figure 2.5** A multilayer perceptron with two hidden layers and an output layer (own illustration). The output of each neuron is computed by applying the activation function ( $\alpha$ ) on the sum of the bias ( $b$ ) and the weighted sum of the input (either  $x_n$  or the output of the previous layer).

Multilayer perceptrons (MLPs), also called feedforward neural networks, are the central part of modern deep learning models (Goodfellow, Bengio, and Courville, 2016, p. 163). A MLP consists of multiple *hidden layers* and one *output layer* (Figure 2.5). Each layer consists of multiple neurons, who are applying the weighted sum (with the weights  $w = \begin{bmatrix} w_{00} & \dots & w_{0n} \\ w_{m1} & \dots & w_{mn} \end{bmatrix}$ ) of the inputs to the activation function  $\alpha$ , with the inputs  $X = [x_0, \dots, x_n]$  either the original input of the network or the result of the previous layer. An addition is the so-called *bias* neuron  $b$  for each layer, which value is added to each weighted sum. Thus, each layer forms the function  $f^{(\cdot)}(x)$  (Equation (2.2)) and each of those functions connected in a chain form the MLP (Equation (2.3)).

$$f^{(\cdot)}(x) = \alpha(X \cdot W + b) \quad (2.2)$$

$$f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x))) \quad (2.3)$$

As discussed, an important part of a MLP are activation functions, who are defined as *fixed nonlinear functions* (Goodfellow, Bengio, and Courville, 2016, p. 167). Historically, one of the most common forms used for hidden layers is the sigmoid, which is defined as “a monotonically increasing function that asymptotes at some finite value as  $\pm\infty$  is approached” (LeCun, Bottou, Orr,

et al., 1998), with the two most prevalent forms being the *standard logistic function* (Equation (2.4), Figure 2.6a) and the *hyperbolic tangent* (Equation (2.5), Figure 2.6b). In comparison between those two the hyperbolic tangent is preferred, as it is able to converge faster (LeCun, Bottou, Orr, et al., 1998). Today, however, the most commonly used activation function is the *rectified linear unit (ReLU)* (Equation (2.6), Figure 2.6c), which has shown to achieve better performance than other functions (Glorot, Bordes, and Bengio, 2011).

For the use in the output layer, however, the standard logistic function is still prevalent, as it is able to produce probabilities if applied on the output vector  $K = [k_0, k_1, \dots, k_n]$ . The values of this vector are defined as *logits*, “the vector of raw (non-normalized) predictions that a classification model generates” (Google Machine Learning Glossary, n.d.[b]). An alternative here is the *softmax activation function* (Equation (2.7)), which produces, instead of independent probabilities for each class, probabilities, whose sum for all classes equals 1.

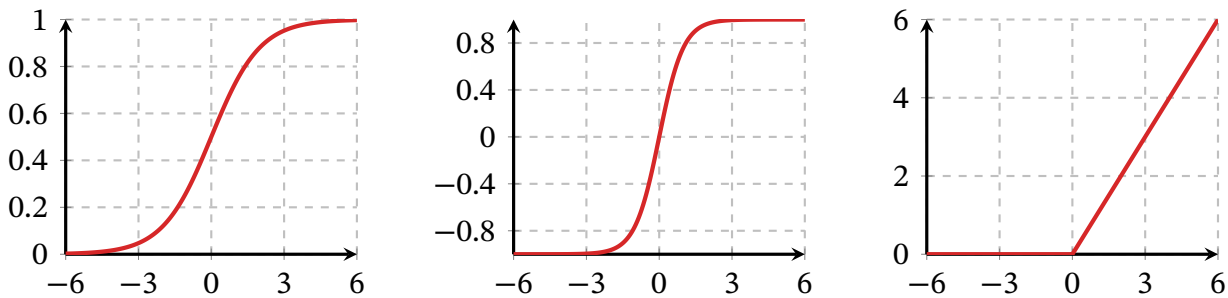
While not an activation function, an additional element used in tandem with them is the so-called *batch normalization (BN)* developed by Ioffe and Szegedy (2015), which helps during the updating of the weights during the training process (Goodfellow, Bengio, and Courville, 2016, p. 310).

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (2.4)$$

$$f(x) = \tanh(x) \quad (2.5)$$

$$f(x) = \max(0, x) \quad (2.6)$$

$$f(x)_i = \frac{\exp(x_i)}{\sum_j^K \exp(x_j)}, \quad \text{for } i = 1, \dots, |K| \quad (2.7)$$



(a) The standard logistic activation function. (b) The Hyperbolic tangent activation function. (c) The ReLU activation function.

**Figure 2.6** Three commonly used activation functions. A fourth widely used activation function is the softmax function.

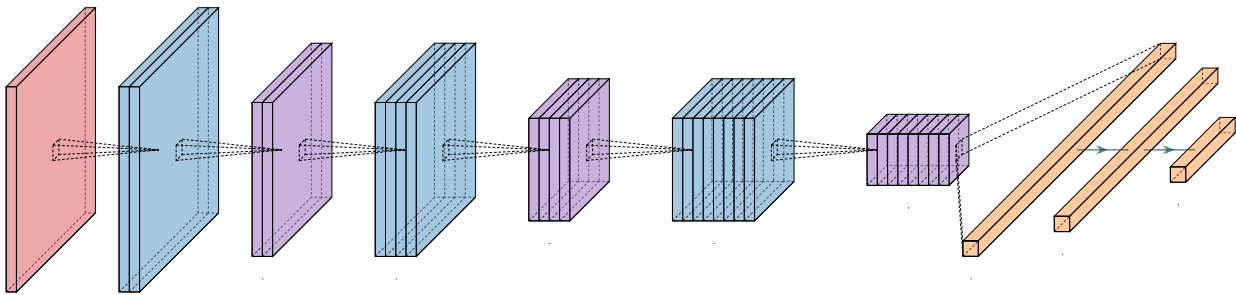
Training is defined as the process of updating the weights, who are the most decisive element to determine the output of the MLP. During this process, a loss function is minimized using *gradient descent*. The gradients are computed by the *backpropagation algorithm* discovered independently by both Werbos (1982) and Rumelhart, G. E. Hinton, and Williams (1986) (Goodfellow, Bengio, and Courville, 2016, p. 218). The training is normally done in batches, with different strategies for different sizes (Google Machine Learning Glossary, n.d.[a]):

- *Stochastic Gradient Descent*: batch size is 1

- *Full batch*: batch size equals size of training set (usually inefficient)
- *Mini-batch*: batch size is between 1 and number of training set, often between 10 and 1000 (usually most efficient)

MLPs today are normally not used on their own, even though they can achieve impressive results (Ciresan et al. (2010) were able to achieve an error rate of 0.35% on the MNIST handwritten digits benchmark using a very big MLP). One application nowadays is the use as head for further developed ANN such as *convolutional neural networks*.

## Convolutional Neural Networks



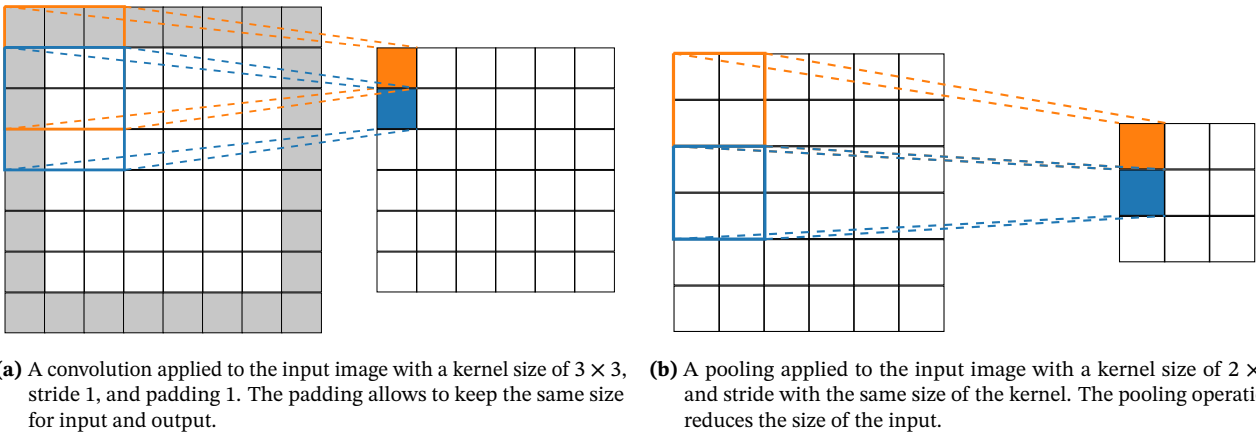
**Figure 2.7** A convolutional neural network with three convolutional layers (own illustration). A first convolution (blue) is applied on the input image (red), followed by a pooling operation (purple). It is possible, by using multiple filters, to increase the amount of layers after each convolution. Convolution and pooling operations are repeated until the final pooling layer. Here the result are flattened and passed to a MLP (orange) serving as a cluster head (LeCun, Bottou, Bengio, et al., 1998).

Convolutional neural networks (CNNs) are specialized ANNs utilizing *convolutions* instead of the general matrix multiplication (Figure 2.7), which are mostly used for processing image data (Goodfellow, Bengio, and Courville, 2016), with the first modern CNN developed by LeCun, Bottou, Bengio, et al. (1998). In general, a convolution is defined as “an operation on two functions of a real-valued argument” (Goodfellow, Bengio, and Courville, 2016, p. 322). In the context of a CNN it takes an input (either the input image or the result of the previous layer) and applies a *filter kernel* on it, who moves over the input to produce an output (Figure 2.8a) with the aim of feature extraction. The decisive elements here are the *kernel size* (size of the filter kernel), *stride* (movement of the filter kernel after each operation), and *padding* (values added to all four sides of the input). After the convolution the output is typically applied to an activation function.

Afterwards a pooling is performed with the aim to reduce dimensionality and make “the representation approximately invariant to small translations of the input” (Goodfellow, Bengio, and Courville, 2016, p. 332). Once again, a kernel is moved over the input (Figure 2.8b). Two different pooling operations are normally used: *max pooling* (which selects the maximum value inside the kernel) or *average pooling* (which take the average of the values inside the kernel). The result is either passed to a new convolutional layer (which repeats the process) or is flattened and passed to a MLP to perform a final task (e.g., clustering).

While large, deep CNNs can achieve impressive results (Krizhevsky, Sutskever, and G. E. Hinton, 2017), a problem with deep ANN is vanishing or exploding gradients (Glorot and Bengio, 2010).

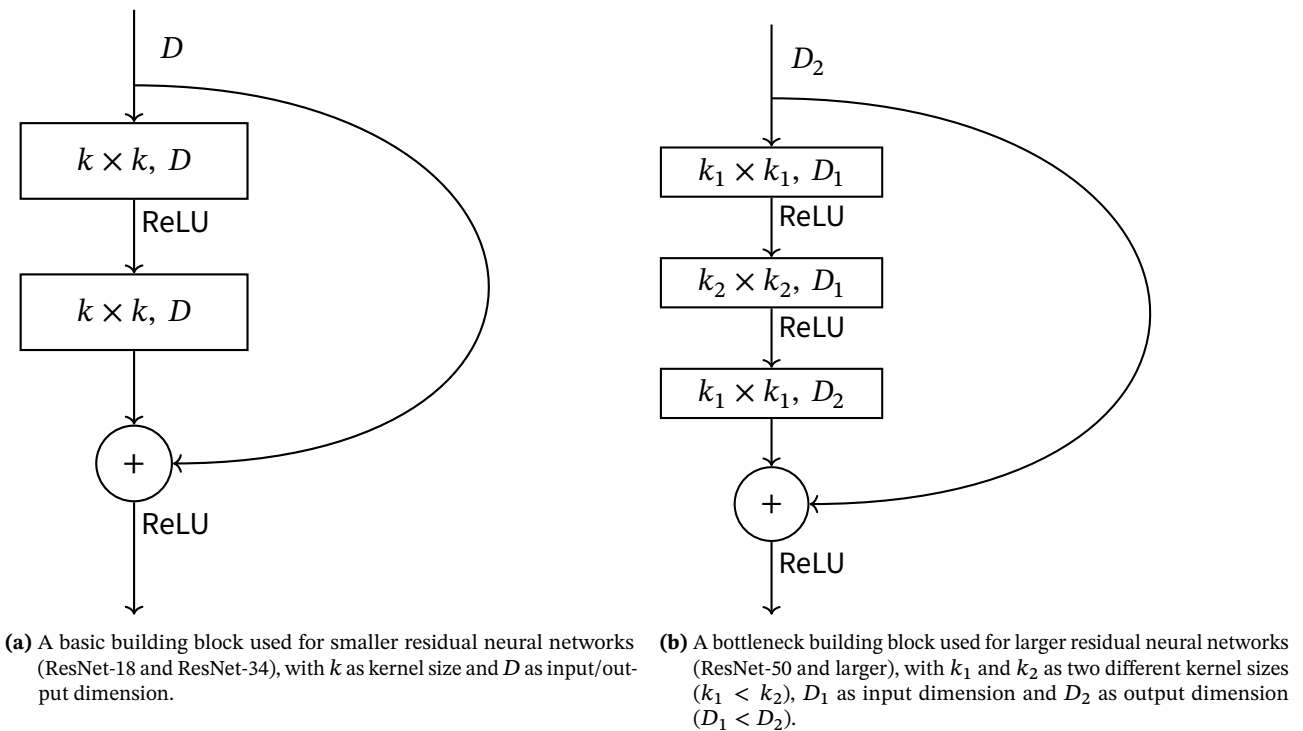




**Figure 2.8** Examples of convolution and pooling operations (own illustrations), both applied on an image with the size  $6 \times 6$ .

A solution for that problem are *residual neural networks (ResNets)*.

### Residual Neural Networks



**Figure 2.9** Two different building blocks for residual neural networks (own illustrations after K. He et al. (2016)), both consisting of multiple convolutional layers.

Residual neural networks (ResNets) are specialized CNNs developed by K. He et al. (2016) to counter vanishing or exploding gradients common with deep CNNs. A ResNet consists of multiple *residual blocks*, each consisting of multiple convolutional layers and a *residual connection* (Figure 2.9). The residual connection adds the identity of the input  $x$  to the output of the convolutional layers:  $g(x) = f(x) + x$ , with  $g(x)$  as the output of the residual block and  $f(x)$  as the output of the stacked convolutional layers. By adding multiple residual blocks together (Table 2.2) it is possible to achieve a very good performance even with very deep networks consisting of over a hundred layers.

**Table 2.2** The different ResNet-Architectures. Each inner layer (conv2\_x to conv5\_x) consists of multiple building blocks, who each consists of two or three convolution layers (Figure 2.9). For each layer the kernel size  $k$  and dimension  $D$  are given. Downsampling is performed at cov3\_1, cov4\_1, and cov5\_1 with a stride of 2. The kernel size  $k_1$ , stride  $s$ , and padding  $p$  of the first layer can differ and the max pooling after the first layer is not always present.

Layer name	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	$k_1 \times k_1, s, p, 64$ Max Pooling				
conv2_x	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
Average Pooling and MLP					

## 2.2.2 Optimal Transport

While deep learning is the foundation of the first part of this thesis, the second part is based on the concept of *optimal transport (OT)*, a mathematical theory based on the transportation problem originally introduced by the French mathematician Gaspard Monge in 1781 (Peyré and Cuturi, 2019, p. 9). The problem was originally formulated as to minimize the cost of moving a distribution of particles to another distribution (Santambrogio, 2015, p. xiv).

This so-called *monge problem* can generally be described as in Equation (2.8). In that case  $(\mu, \nu)$  are two arbitrary probability measures, supported on two probability spaces  $(\mathcal{X}, \mathcal{Y})$ , which can be linked through a map  $T : \mathcal{X} \rightarrow \mathcal{Y}$ . The constraint  $T_{\#}\mu = \nu$  ensures that  $\mu$  is completely transferred to  $\nu$  as defined in the *push-forward operator*  $T_{\#} : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{M}(\mathcal{Y})$ .  $c(x, y)$  is defined as the ground cost, with associated pairwise cost matrix  $C_{i,j}$  as defined in Equation (2.9) (Peyré and Cuturi, 2019, pp. 5, 11—12; Santambrogio, 2015, p. xiv).

$$\inf \left\{ \int_{\mathcal{X}} c(x, T(x)) d\mu(x) : T_{\#}\mu = \nu \right\} \quad (2.8)$$

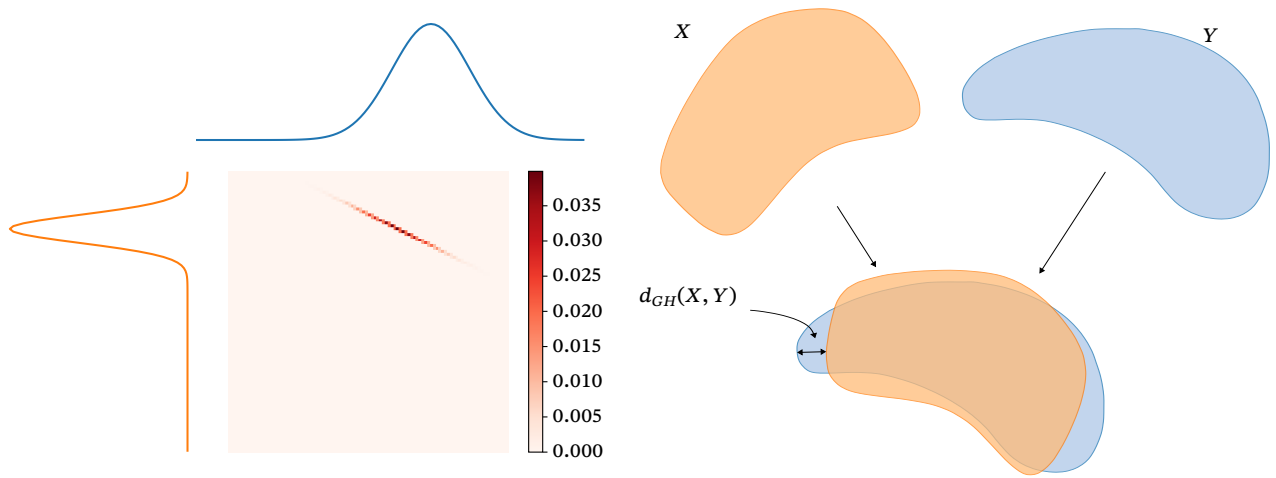
$$C_{i,j} := (c(x_i, y_j))_{i,j} \quad (2.9)$$

OT is able to define a distance between probability measures, the so-called *Wasserstein distance*  $W_p(\mu, \nu)$  (Figure 2.10a) as defined in Equation (2.10).  $\pi$  is defined as the *coupling measure* between  $\mu$  and  $\nu$ , which is defined as constructing two new random variables  $X$  and  $Y$  on a new probability space, such that  $\text{law}(X) = \mu$  and  $\text{law}(Y) = \nu$ , with an additional condition defined in Equation (2.11).  $\Pi(\mu, \nu)$  serves as a set of coupling measures (Peyré and Cuturi, 2019, pp. 5, 18—19; Villani, 2009, pp. 17, 105).

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{\mathcal{X} \times \mathcal{Y}} c(x, y)^p d\pi(x, y) \right\}^{\frac{1}{p}} \quad (2.10)$$

$$\begin{aligned} A \subset \mathcal{X}, \pi(A \times \mathcal{Y}) &= \mu(A) \\ B \subset \mathcal{Y}, \pi(\mathcal{X} \times B) &= \nu(B) \end{aligned} \quad (2.11)$$

It is possible to use the Wasserstein distance to compute a distance metric between structured data such as graphs. As an example, Kolouri et al. (2020) proposed a framework to measure dissimilarity between node embeddings using the Wasserstein distance, while Petric Maretic et al. (2019) proposed a framework to use the Wasserstein distance to compare the structural information of graphs.



(a) Visualization of the Wasserstein distance (reprinted from Flamary and Gramfort (2022)) between an exemplary source (orange) and target (blue) distribution, with  $W_1 = 0.17$ . The OT matrix (red) shows how much “mass” should be shifted from the source to the target distribution.

(b) Visualization of the Gromov-Hausdorff distance (reprinted from Chazal and Michel (2021)) between two metric spaces  $X$  and  $Y$ .

**Figure 2.10** Visualization of Wasserstein and Gromov-Hausdorff distances.

Another way to analyze datasets is to model them as metric spaces, with two arbitrary metric spaces defined as  $(X, d_X)$  and  $(Y, d_Y)$ . Then  $\phi$  is defined as the *isometric embedding* if  $\phi$  is surjective and Equation (2.12) is satisfied. Based on that, it is possible to define a distance between the two metric spaces. One example is the *Gromov-Hausdorff distance* (Figure 2.10b) as defined in Equation (2.13), with  $R$  ranging over the set of correspondences between  $X$  and  $Y$ . By utilizing the principles of OT it is possible to modify the Gromov-Hausdorff distance and to define the *Gromov-Wasserstein distance* as in Equation (2.14) between two *metric measure spaces*  $(X, d_X, \mu_X)$  and  $(Y, d_Y, \nu_Y)$ , each consisting of a metric space and an additional probability measure (Mémoli, 2011; Mémoli, 2014).

$$d_X(x, x') = d_Y(\phi(x), \phi(x')), \quad \text{for } x, x' \in X \quad (2.12)$$

$$d_{GH}(X, Y) := \frac{1}{2} \inf_R \sup_{(x,y), (x',y') \in R} |d_X(x, x') - d_Y(y, y')| \quad (2.13)$$

$$d_{GW,p}(X, Y) := \frac{1}{2} \inf_{\mu} \left( \int \int |d_X(x, x') - d_Y(y, y')|^p \mu(dx \times dy) \mu(dx' \times dy') \right)^{\frac{1}{p}} \quad (2.14)$$

Corresponding to the above, it is also possible to use the Gromov-Wasserstein distance to compute distances between graphs, with one example being a Gromov-Wasserstein learning framework proposed by H. Xu et al. (2019) to measure dissimilarity between graphs based on node embeddings. Another example is the combination of spectral clustering and Gromov-Wasserstein Learning developed by Chowdhury and Needham (2021).

# 3 Tools and Data

---

This chapter provides an overview of the used tools (hard- and software) as well as the data basis.

## 3.1 Tools

The tools can be divided in two groups: hardware and software. Hardware was provided by the German Aerospace Center, the used software is free and open-source.

### 3.1.1 Hardware

Two different hardware environments were used. A personal workstation (Lenovo ThinkStation P520), utilizing an eight-core Intel Xeon CPU, 64 GB RAM, and an Nvidia RTX A4000 GPU with 16 GB VRAM, used for development, testing, and small scale processing.

The other environment is the *terrabyte platform*, a new High Performance Data Analytics platform operated by the German Aerospace Center and the Leibniz Supercomputing Center<sup>1</sup>. The platform offers two different partition types: one for CPU based processing, utilizing two 80-core Intel Xeon CPUs and 1024 GB RAM per node, the other for GPU based processing, with two 48-core CPUs, 1024 GB RAM and four Nvidia HGX A100 GPUs, each with 80 GB VRAM, per node.

### 3.1.2 Software

The used Software products were *QGIS* in the version 3.34.0-Prizren, a free and open-source geographic information system and an official project of the Open Source Geospatial Foundation<sup>2</sup>, mainly used for visual analysis, and *PyCharm Community Edition* in the version 2023.2.1, a free and open-source integrated development environment developed by JetBrains<sup>3</sup>, used for code development.

All processing was done with the use of software scripts developed in *Python*, a free and open-source programming language with widespread use in data science and scientific computing<sup>4</sup>, used in the versions 3.8, 3.9 and 3.10. A powerful feature of Python is the large amount of packages provided and easy integration of those using package managers like *conda*<sup>5</sup>. An overview over the packages used is provided in Table 3.1.

---

<sup>1</sup><https://docs.terrabyte.lrz.de/> — last accessed: May 3, 2024

<sup>2</sup><https://qgis.org/> — last accessed: Apr. 18, 2024

<sup>3</sup><https://jetbrains.com/pycharm/> — last accessed: Apr. 18, 2024

<sup>4</sup><https://python.org/> — last accessed: Apr. 18, 2024

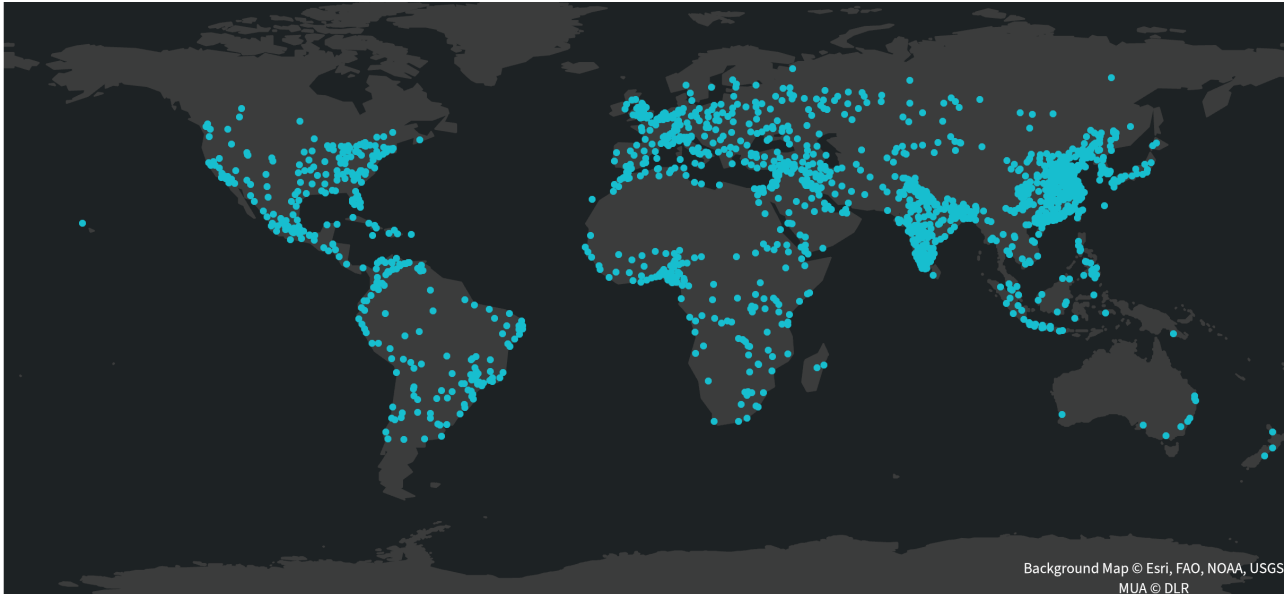
<sup>5</sup><https://docs.conda.io/> — last accessed: Apr. 18, 2024

**Table 3.1** Overview of used Python packages. As stated earlier, three different Python versions were used, for all the used package version is provided (not all packages were used for all versions).

Package	Description	Version 3.8	Version 3.9	Version 3.10
Faiss	Efficient similarity search and clustering of dense vectors <a href="https://faiss.ai/">https://faiss.ai/</a> — last accessed: Apr. 18, 2024	1.7.2		
GDAL	Geospatial data processing <a href="https://gdal.org/">https://gdal.org/</a> — last accessed: Apr. 18, 2024			3.7.2
GeoPandas	Extension of pandas for support of geospatial vector data <a href="https://geopandas.org/">https://geopandas.org/</a> — last accessed: Apr. 18, 2024			0.14.1
GeoUtils	Geospatial data processing <a href="https://geoutils.readthedocs.io/">https://geoutils.readthedocs.io/</a> — last accessed: Apr. 18, 2024			0.0.15
libpysal	Spatial analytical methods <a href="https://pysal.org/libpysal/">https://pysal.org/libpysal/</a> — last accessed: Apr. 18, 2024			4.9.2
Matplotlib	Data visualization <a href="https://matplotlib.org/">https://matplotlib.org/</a> — last accessed: Apr. 18, 2024	3.3.2	3.8.2	3.8.2
Momepy	Quantitative analysis of urban form <a href="https://docs.momepy.org/">https://docs.momepy.org/</a> — last accessed: Apr. 18, 2024			0.7.0
NetworkX	Creation, manipulation, and study of graphs <a href="https://networkx.org/">https://networkx.org/</a> — last accessed: Apr. 18, 2024		3.2.1	3.2.1
NumPy	Large and multidimensional arrays <a href="https://numpy.org/">https://numpy.org/</a> — last accessed: Apr. 18, 2024	1.24.4	1.26.4	1.26.0
pandas	Fast and efficient tables and lists <a href="https://pandas.pydata.org/">https://pandas.pydata.org/</a> — last accessed: Apr. 18, 2024		2.2.2	2.1.3
Pillow	Image processing <a href="https://python-pillow.org/">https://python-pillow.org/</a> — last accessed: Apr. 18, 2024	9.4.0	10.2.0	10.0.1
POT	Efficient solvers for optimization problems related to OT <a href="https://pythonot.github.io/">https://pythonot.github.io/</a> — last accessed: Apr. 18, 2024		0.9.3	
PyTorch	Optimized tensor operations for deep learning <a href="https://pytorch.org/">https://pytorch.org/</a> — last accessed: Apr. 18, 2024	1.12.0		
scikit-learn	Machine learning <a href="https://scikit-learn.org/">https://scikit-learn.org/</a> — last accessed: Apr. 18, 2024	1.3.2	1.4.0	1.3.2
SciPy	Fundamental scientific algorithms <a href="https://scipy.org/">https://scipy.org/</a> — last accessed: Apr. 18, 2024	1.10.1	1.12.0	1.11.3
Shapely	Manipulation and analysis of geometric objects <a href="https://shapely.readthedocs.io/">https://shapely.readthedocs.io/</a> — last accessed: Apr. 18, 2024			2.0.2

## 3.2 Data Basis

The data basis used consists of two datasets, both closed data provided by the German Aerospace Center. Overall there are 1520 study sites (Figure 3.1), with the urban morphology described by the LCZs and the urban boundaries delineated by the MUAs.



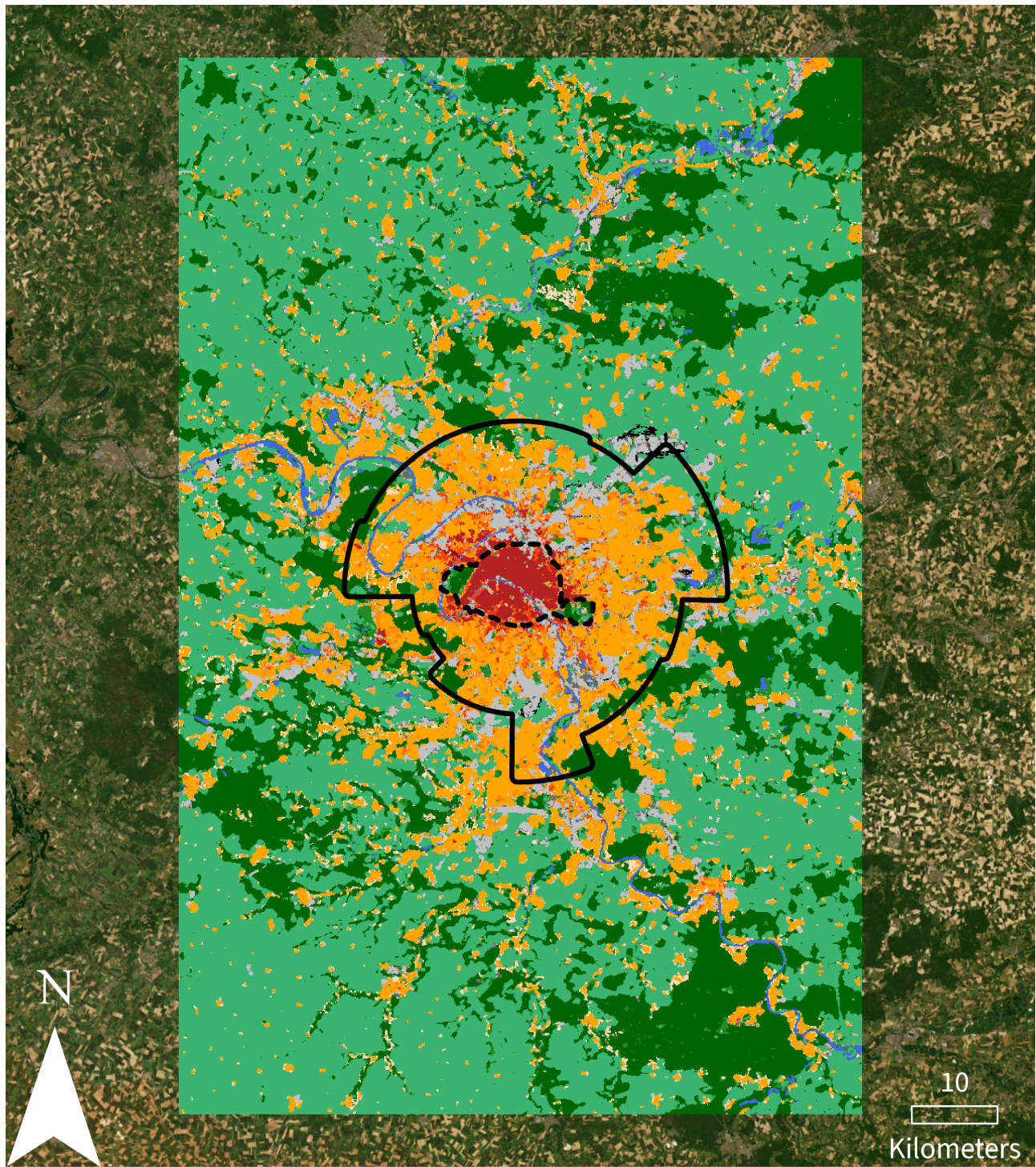
**Figure 3.1** An overview of all 1520 study sites.

### 3.2.1 Local Climate Zones

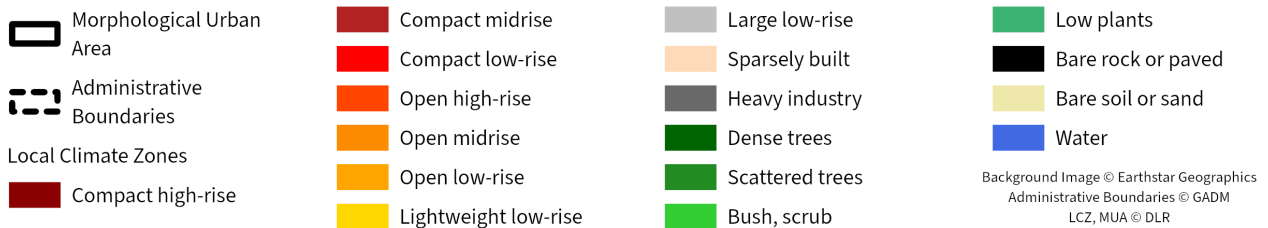
The LCZ data provided was created in 2019 with the method developed by Qiu, M. Schmitt, and Zhu (2019) and is provided in their local WGS84 UTM zone projection. It consists of 1692 raster tiles covering urban areas (with a pixel size of 100 m), who sometimes have overlaps between them. The outer 3% of the original raster tiles were already removed due to frequent missclassifications in this area (Gassilloud, 2022). The overall accuracy of the data is given with 86.7% (Qiu, M. Schmitt, and Zhu, 2019). An example for a raster tile is shown in Figure 3.2.

### 3.2.2 Morphological Urban Areas

The MUA data provided was created in 2019 with the method developed by Taubenböck, Weigand, et al. (2019) and is provided in WGS 1984 Web Mercator (EPSG:3857). It consists of 1520 shapefiles, with the original data already extended with a 1000 m buffer. The MUA delineate the urban boundaries, but do not abide by administrative boundaries as seen in Figure 3.2. A full overview is provided Table B.1.



Paris with Morphological Urban Area, Administrative Boundaries and Local Climate Zones



**Figure 3.2** Paris as an example for one investigated city with the corresponding raster tile containing the LCZ data and the extent of the MUA, with the administrative boundaries of Paris as comparison.



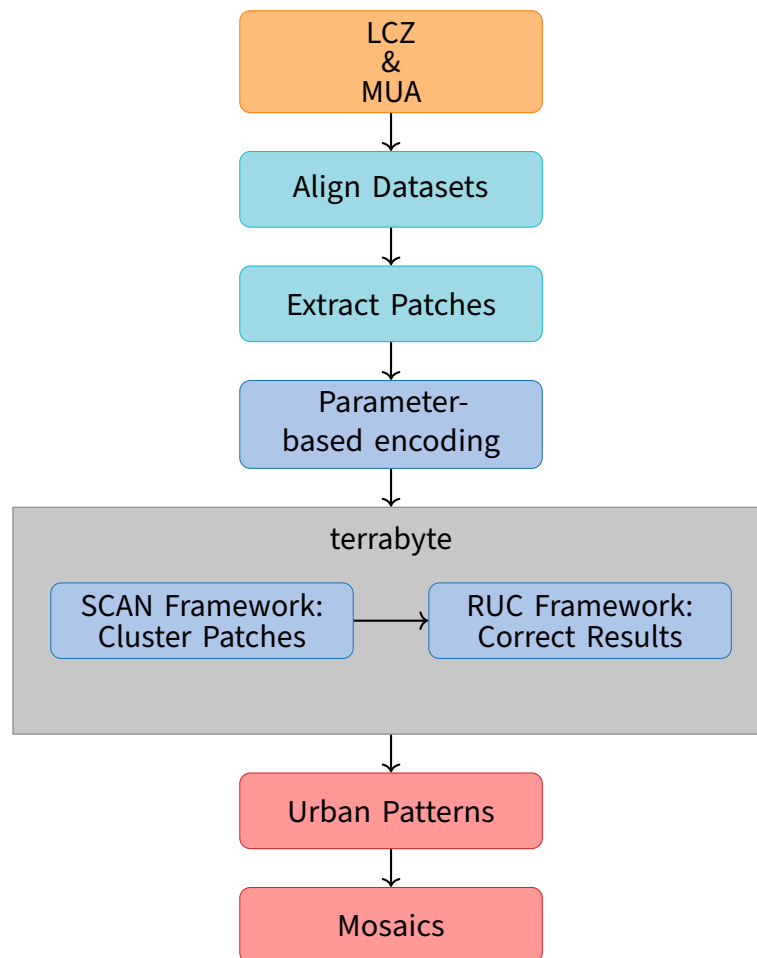
# 4 Methodology

---

This chapter provides an overview over the methodology. As stated earlier, the thesis consists of two parts with different methods. In Section 4.1 the approach to find urban patterns using a parameter based encoding will be presented, Section 4.2 covers the identification of different city models based on the previous discovered patterns.

## 4.1 Unsupervised Deep Learning Approach to find Urban Patterns

This part is mostly based on the thesis of Gassilloud (2022), who developed a framework to find urban patterns using methods developed for image clustering. The aim of this part is to adapt his work by using a *parameter based encoding* instead of a one-hot encoding with the objective to provide more information about the urban morphology to the framework.



**Figure 4.1** Proposed workflow to find urban patterns. The two datasets are aligned, patches are extracted, and each patch is encoded based on their LCZ properties. Afterwards the clustering, using the SCAN and RUC framework, is performed on the terrabyte platform. Based on the resulting urban patterns, a mosaic is computed for each urban area.

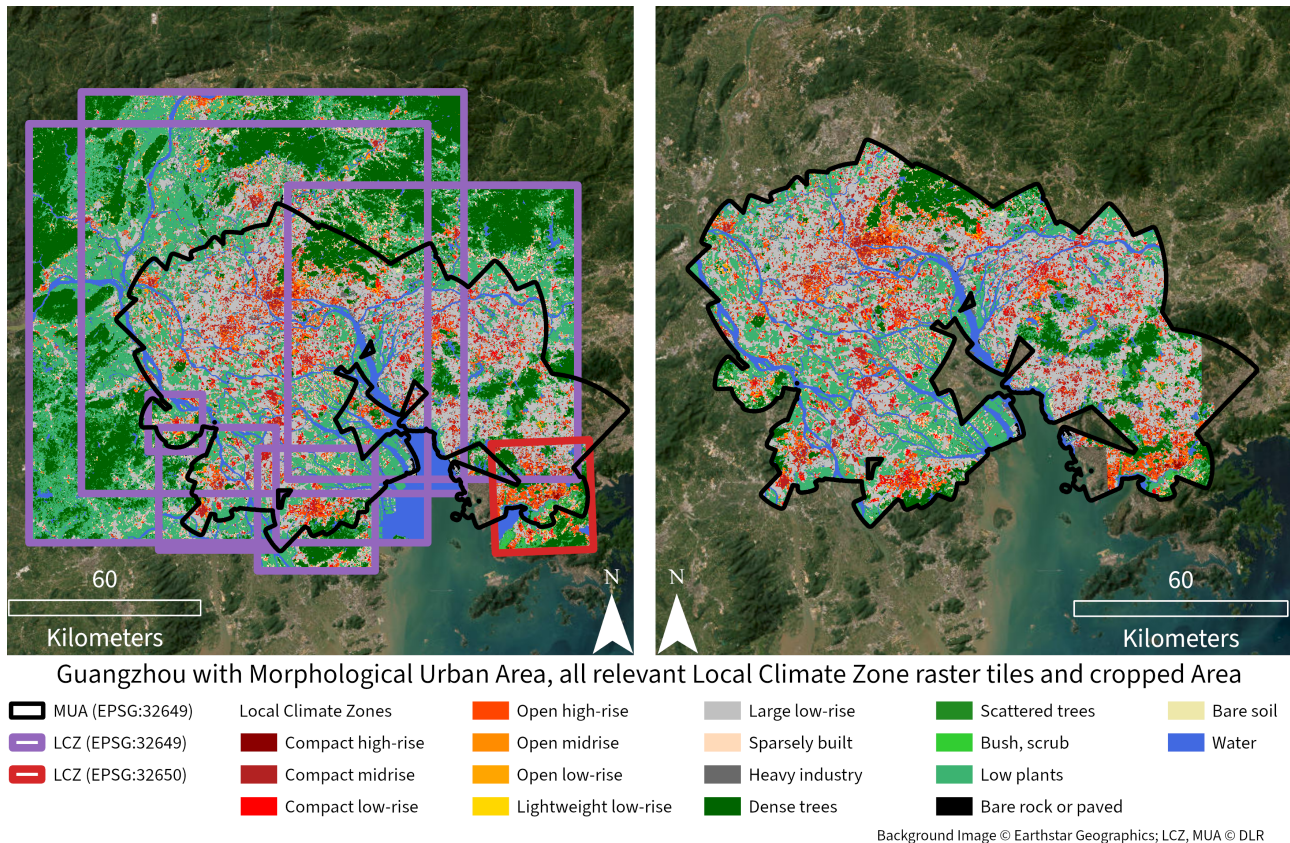
The workflow to do this is described in Figure 4.1: The data, consisting of LCZ mappings describing the urban morphology and MUAs to delineate the urban areas, needs to be prepared (Section 4.1.1).

Afterwards, the clustering is performed using two different frameworks: the *semantic clustering by adopting nearest neighbors (SCAN)* framework developed by Van Gansbeke et al. (2020) to cluster images, and the *robust learning for unsupervised clustering (RUC)* framework developed by S. Park et al. (2021) to improve the previous obtained results (Sections 4.1.2 and 4.1.3). After obtaining the urban patterns, a mosaic of each urban area consisting of the patterns is computed (Section 4.1.4).

### 4.1.1 Data Preparation

As stated earlier, the data basis consists of two different datasets, shapefiles and raster images, with the size of the images inconsistent to each other and not consistent with respect to the urban areas. To create a consistent dataset, that can be used to perform the image clustering, multiple steps are performed:

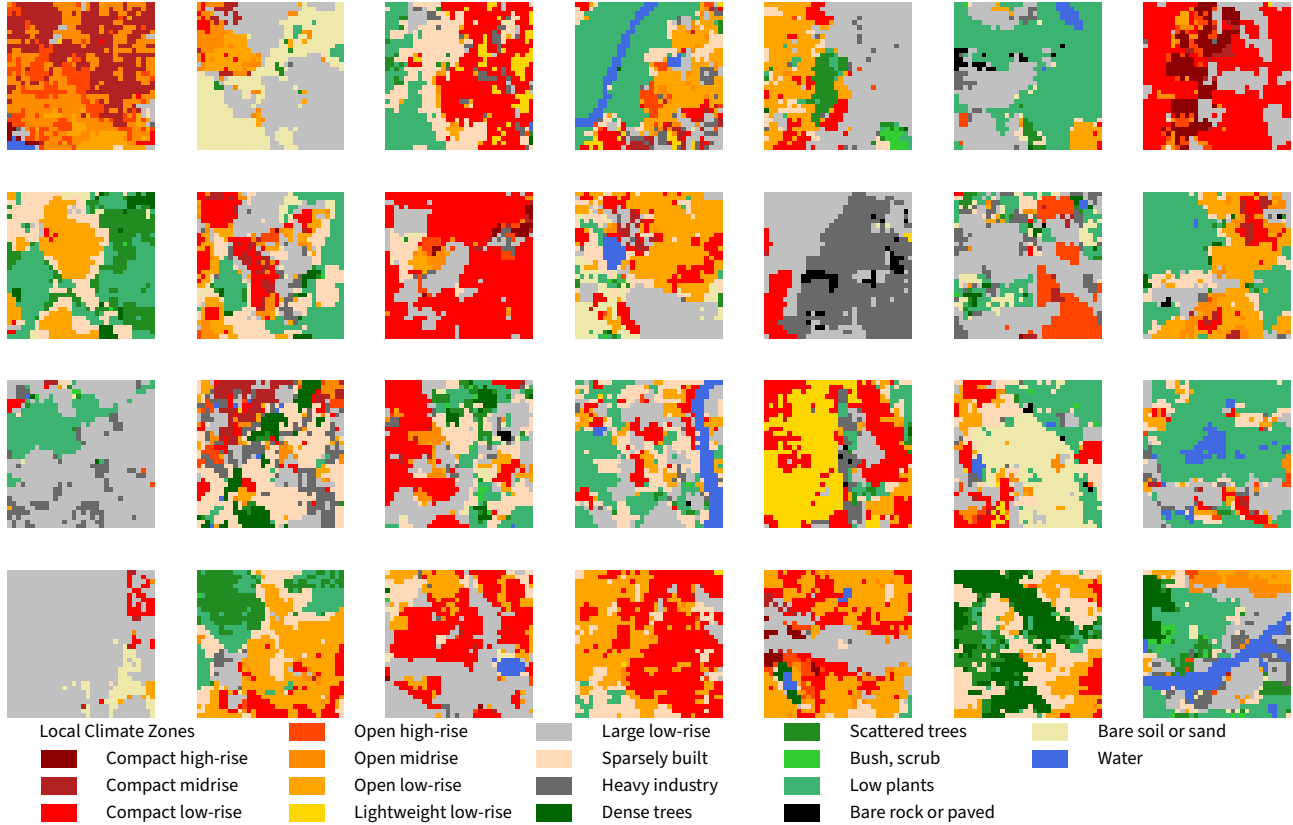
**Crop local climate zones to morphological urban areas** The first step is to crop the raster tile containing the LCZ mapping to the corresponding MUAs. Unfortunately, sometimes multiple raster tiles are needed to cover one MUA, who sometimes even have different reference systems. If that is the case, a mosaic containing all relevant raster tiles for this area is created. Afterwards the raster data is cropped to each MUA (Figure 4.2) (Gassilloud, 2022).



**Figure 4.2** Overview of Guangzhou (reprinted from Gassilloud (2022)) with MUA, all relevant LCZ raster tiles in their respective coordinate reference system, and the cropped raster data. As visible, the MUA is not always completely covered with the LCZ data.

**Patch extraction** After aligning both datasets, the data to be clustered is extracted. The size of those patches is set to  $32 \times 32$  pixel, thus each patch covers a size of  $10.24 \text{ km}^2$ . That can be

seen as a medium scale, in which the urban patterns are neither too generalized nor too specific. The patch extraction is done with a stride of 10 pixels. Thus, each patch overlaps the previous by around  $\frac{2}{3}$ , which allows “a more continuous representations of cities and search of patterns” (Gassilloud, 2022). Overall a total of  $n = 315\,992$  patches were extracted forming the dataset  $D = \{x_0, \dots, x_n\}$ . A random subset of patches is shown in Figure 4.3.

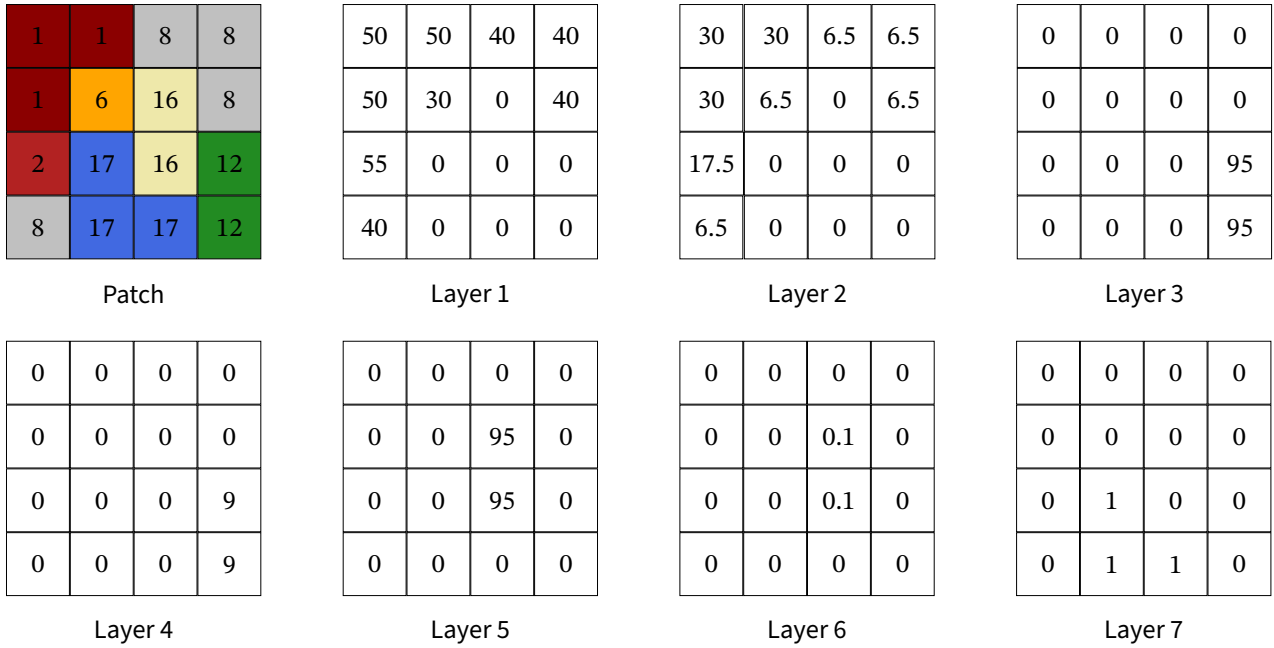


**Figure 4.3** An overview of a random sample of extracted patches.

**Encoding** As stated before, a parameter based encoding will be used utilizing the LCZ properties defined in Appendix A.1. For each data point, a stack of seven layers will be created, with Layers 1 and 2 defining LCZ 1—10, Layers 3 and 4 defining LCZ A—D, Layers 5 and 6 defining LCZ E and F, and Layer 7 defining LCZ G (Figure 4.4). Thus, Layers 1, 3, and 5 contain data of density for respectively built types, vegetation, and bare surface. Layers 2, 4, and 6 on the other hand contain height data of those elements:

- Layer 1: *building surface fraction* for LCZ 1—10
- Layer 2: *height of roughness elements* for LCZ 1—10
- Layer 3: *pervious surface fraction* for LCZ A—D
- Layer 4: *height of roughness elements* for LCZ A—D
- Layer 5: *pervious surface fraction* for LCZ E and F
- Layer 6: *height of roughness elements* for LCZ E and F
- Layer 7: *existence* of LCZ G

With the dataset used for the clustering defined, the step is to assign each patch  $x \in D$  to a cluster



**Figure 4.4** Encoding shown on an exemplary patch. Each Layer contains the data defined above for their corresponding patch for each pixel or zero if the pixel defines the wrong LCZ.

$c_i \in C_{UB}$  using the semantic clustering by adopting nearest neighbors (SCAN) and robust learning for unsupervised clustering (RUC) framework.

#### 4.1.2 Semantic Clustering by Adopting Nearest Neighbors

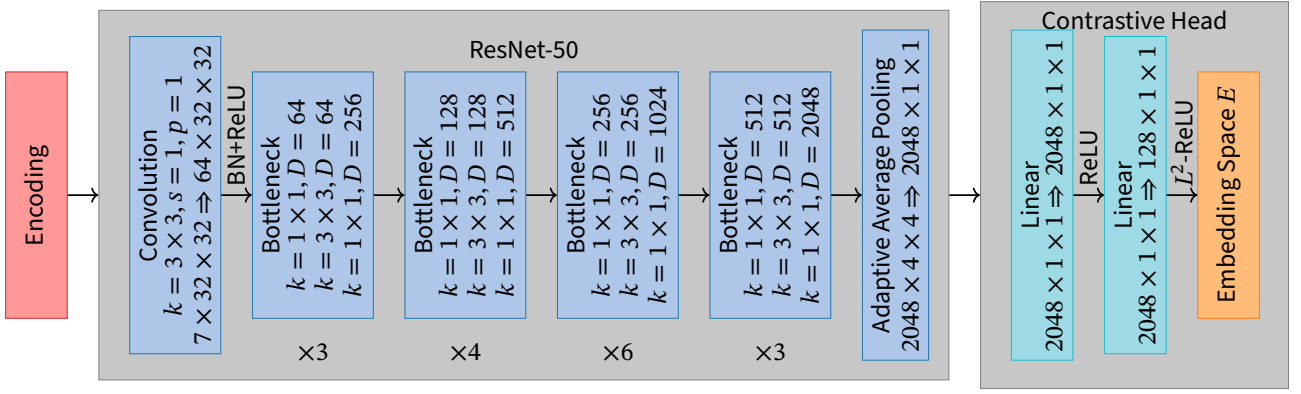
The SCAN framework was developed to “group images into semantically meaningful clusters when ground-truth annotations are absent” (Van Gansbeke et al., 2020). In relation to this thesis, it means to assign each patch  $x \in D$  to a cluster  $c_i$ , which each cluster representing one specific urban patterns, without a priori knowledge about expected urban patterns. The framework itself is separated into three different parts (Van Gansbeke et al., 2020):

1. Representation learning through solving a pretext task
2. Clustering patches through mining nearest neighbors
3. Improving of clustering results through self-labeling

##### Representation Learning through Solving a Pretext Task

Representation learning is the use of a pretraining stage to generate feature representations only from images. That means, a pretext task  $\tau$  trains a neural network  $\Phi_\theta$  (Figure 4.5) to map images into feature representation. In this case, the simple framework for contrastive learning (SimCLR) developed by Chen et al. (2020) is used to solve the pretext task. SimCLR works by maximizing agreement between an image  $x_i$  and their augmentations  $T[x_i]$  (Equation (4.1)) and consists of the components shown in Figure 4.6, with the detailed procedure as follows:

For each patch  $x \in B$  two data augmentations ( $t_i \sim T_{pre}$  and  $t_j \sim T_{pre}$ ) are applied to obtain the augmented patches  $\tilde{x}_i$  and  $\tilde{x}_j$ . The augmentations of the original SimCLR and SCAN framework are adapted to the reality of the LCZ data (Overview of all augmentations in Appendix A.2).

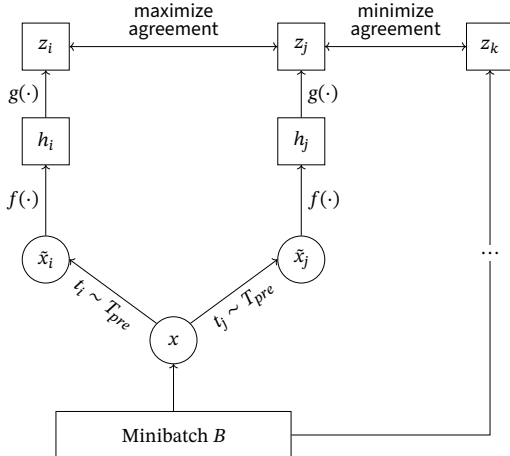


**Figure 4.5** A ResNet-50  $f(\cdot)$  and a Contrastive Head  $g(\cdot)$  form the network  $\Phi_\theta$  (own illustration). An image  $x$  is passed through the ResNet-50, with specifications of each layer shown. After the first convolution the output is applied to a ReLU activation function and a BN. Afterwards the data is passed to the contrastive head, a fully connected MLP.

After the encoding, both augmented patches are passed through the ResNet-50  $f(\cdot)$  to obtain the representations  $h_i = f(\tilde{x}_i)$  and  $h_j = f(\tilde{x}_j)$ . Both are afterwards passed through the contrastive head  $g(\cdot)$ , a MLP with one hidden layer, to obtain the representations  $z_i = g(h_i)$  and  $z_j = g(h_j)$  in the embedding space  $E$ . With the minibatch  $B$  this results in  $2B$  data points and for each positive pair, the other  $2(B - 1)$  augmented patches of the minibatch are treated as negative examples. The loss function  $\ell_{pre}(i, j)$  is defined in Equation (4.2), with  $\text{sim}(u, v) = u^T v / \|u\| \|v\|$  defined as the cosine similarity between  $u$  and  $v$  (Chen et al., 2020; Gassilloud, 2022).

$$\min_{\theta} d(\Phi_\theta(x_i), \Phi_\theta(T[x_i])) \quad (4.1)$$

$$\ell_{pre}(i, j) = -\log \frac{\exp(\text{sim}(z_i, z_j)/\kappa)}{\sum_{k=1}^{2B} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\kappa)} \quad (4.2)$$



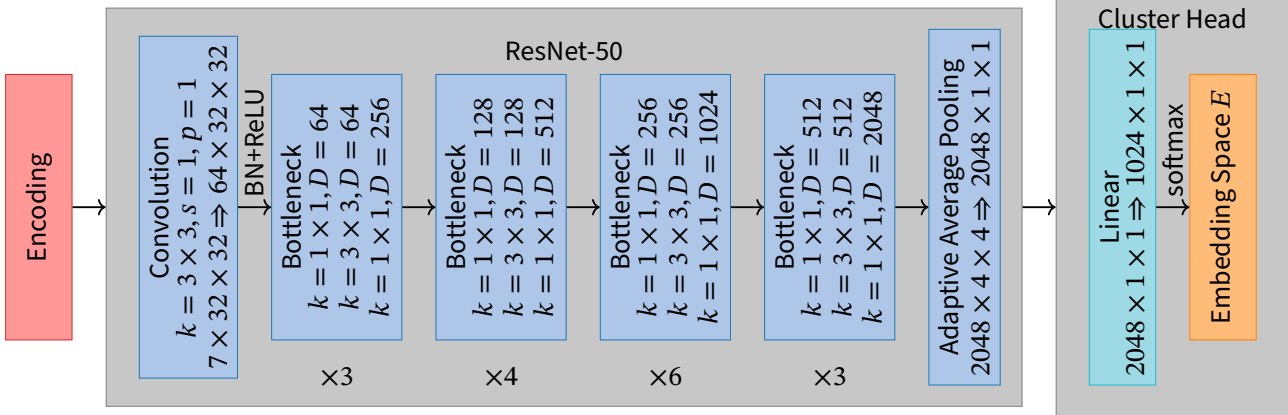
**Figure 4.6** Components of the SimCLR framework (own illustrations after Gassilloud (2022)). Two data augmentations  $\tilde{x}_i$  and  $\tilde{x}_j$  are created for each patch  $x \in B$  and passed through the ResNet-50  $f(\cdot)$  and contrastive head  $g(\cdot)$  to obtain the representations  $z_i$  and  $z_j$ . The loss function  $\ell_{pre}(i, j)$  is used to maximize agreement between those representations and to minimize agreement between all other representations  $z_k$  (with  $k \neq i, j$ ).

The indicator function  $\mathbb{1}_{[k \neq i]} \in \{0, 1\}$  evaluates to 1 if  $k \neq i$  and the temperature parameter  $\kappa$  regulates the training process. The final loss of a minibatch is computed across all positive pairs,  $(i, j)$  and  $(j, i)$ . Since SimCLR benefits from a large batch size (as it allows for more negative examples per minibatch), the batch size is set to 512 and the network  $\Phi_\theta$  is trained for 500 epochs (Chen et al., 2020; Gassilloud, 2022).

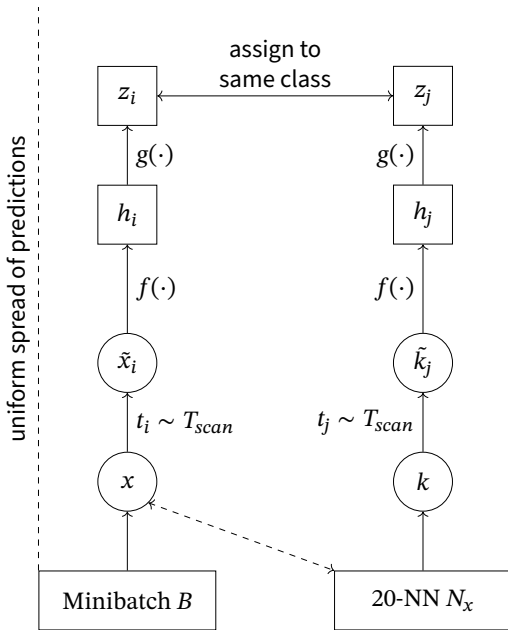
To summarize, the loss function minimizes the distance (*maximizing agreement*) between an augmented pair of patches in the embedding space  $E$  and maximizes the distance to all other augmented patches of the minibatch  $B$ . This allows the network  $f(\cdot)$  to learn high-level characteristics, which translates to morphological

features in this use case. Furthermore, it is possible to use the trained network to mine nearest neighbors of a patch, which, in addition to the clustering, is part of the next step (Van Gansbeke et al., 2020; Gassilloud, 2022).

### Clustering Patches through Mining Nearest Neighbors

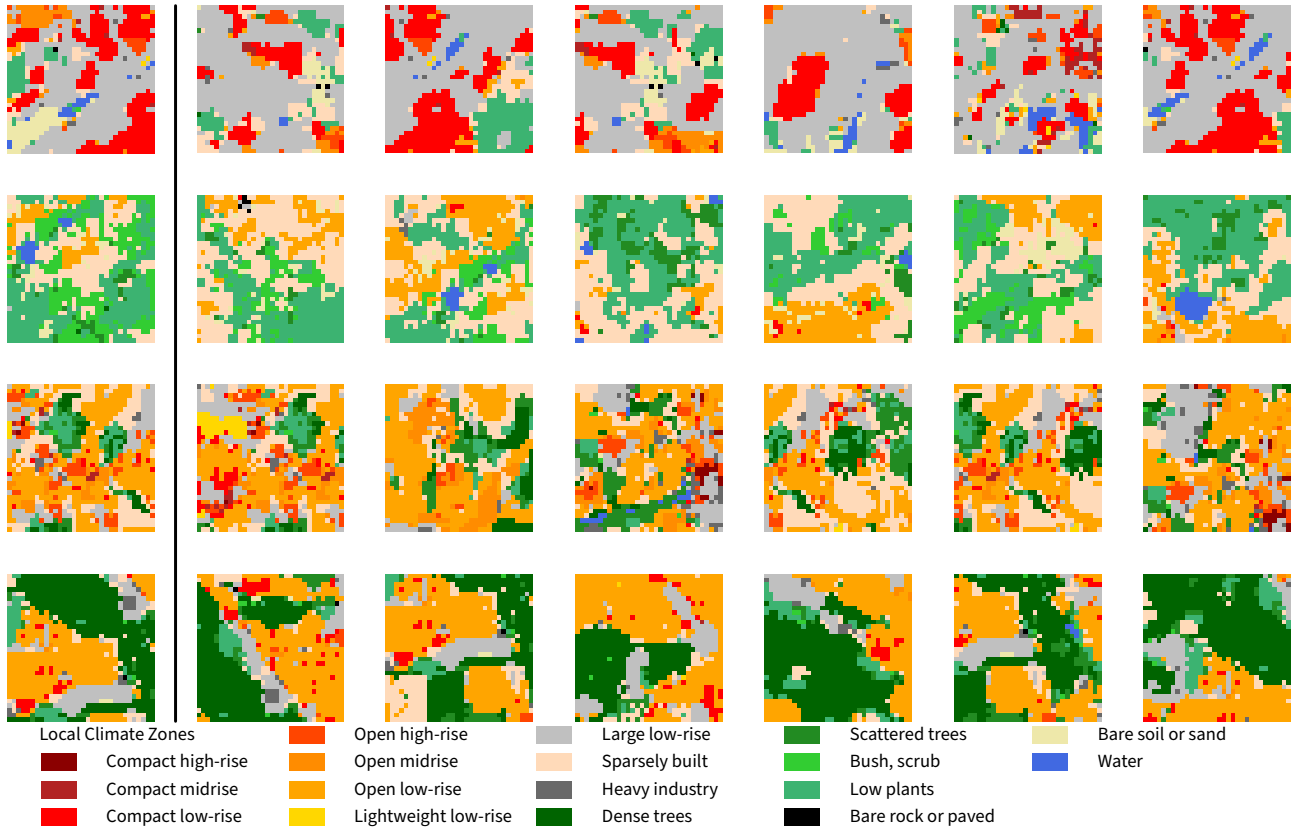


**Figure 4.7** A ResNet-50  $f(\cdot)$  and a Cluster Head  $g(\cdot)$  form the network  $\Phi_\eta$  (own illustration).  $f(\cdot)$  is initialized with the weights  $\theta$  of the previous trained network  $\Phi_\theta$ , which allows it to retain the learned features. The cluster head is a fully connected MLP with the output dimension  $1024 \times 1 \times 1$ .



**Figure 4.8** Components of the second step of the SCAN framework (own illustrations after Gassilloud (2022)). Two data augmentations  $\tilde{x}_i$  and  $\tilde{k}_j$  are created for each patch  $x \in B$  and  $k \in N_x$ . They are passed through the ResNet-50  $f(\cdot)$  and cluster head  $g(\cdot)$  to obtain the representations  $z_i$  and  $z_j$ . The loss function  $\ell_{scan}$  aims to assign both of those representations to the same cluster  $C_i$ , while spreading the predictions uniformly across the cluster.

The aim of the second step is to assign each  $x \in D$  to a cluster  $c_i$ . But in order to achieve that, the first step is to mine the  $k$  nearest neighbors of each patch  $x \in D$  in the embedding space  $E$  using the previous trained network  $\Phi_\theta$ . The set  $N_x$  is defined as the neighboring samples of each patch  $x \in D$ , with the assumption that the nearest neighbors are instances of the same semantic cluster (Van Gansbeke et al., 2020; Gassilloud, 2022). A random selection of patches and 5 of their nearest neighbors is shown in Figure 4.9. As this thesis does not have any ground truth about the expected urban patterns, it is impossible to measure the “semantic similarity” of those patches. But a visual analysis shows an existing similarity between the anchor images and their nearest neighbors. As the results are very insensitive to the value of  $k$  in the range of  $[5, 50]$  (Van Gansbeke et al., 2020), the 20 nearest neighbors for each patch are used in this thesis. With the nearest neighbors for each patch mined, the actual clustering can be performed.



**Figure 4.9** An overview of a random sample of patches (left, anchor images), with 6 of their 20 nearest neighbors each (right). As visible, some of the nearest neighbors show overlaps to their anchor images.

The aim here is to train a neural network  $\Phi_\eta$  (Figure 4.7) to classify a sample  $x$  and its nearest neighbors  $N_x$  together. The detailed components are shown in Figure 4.8. For each patch  $x \in B$  is one its nearest neighbors  $k \in N_x$  selected. A data augmentation is applied to each patch ( $t_i \sim T_{scan}$  and  $t_j \sim T_{scan}$ ) to obtain the augmented patches  $\tilde{x}_i$  and  $\tilde{k}_j$ . Once again the original augmentations of the SCAN framework are adapted, with an overview shown in Appendix A.2. The augmented patches are passed through the networks  $f(\cdot)$  and  $g(\cdot)$  to obtain the representations  $z_i = g(f(\tilde{x}_i))$   $z_j = g(f(\tilde{k}_j))$ . The network  $g(\cdot)$  is a MLP, which terminates in a softmax activation function. That allows the network  $\Phi_\eta$  to perform a soft assignment over the clusters  $C = \{0, \dots, c_{1023}\}$  by minimizing the loss function  $\ell_{scan}$ , as defined as in Equation (4.3), consisting of two terms. The first term aims to impose  $\Phi_\eta$  to make consistent predictions utilizing the dot product  $\langle \cdot, \cdot \rangle$ , which is maximal when the predictions are confident and consistent. The second term assures that not all samples are assigned to a single cluster by including a cross entropy term, which ensures a uniform spread of predictions.  $\Phi_\eta^c(x)$  is defined as the probability of patch  $x$  assigned to cluster  $c$  (Van Gansbeke et al., 2020; Gassilloud, 2022).

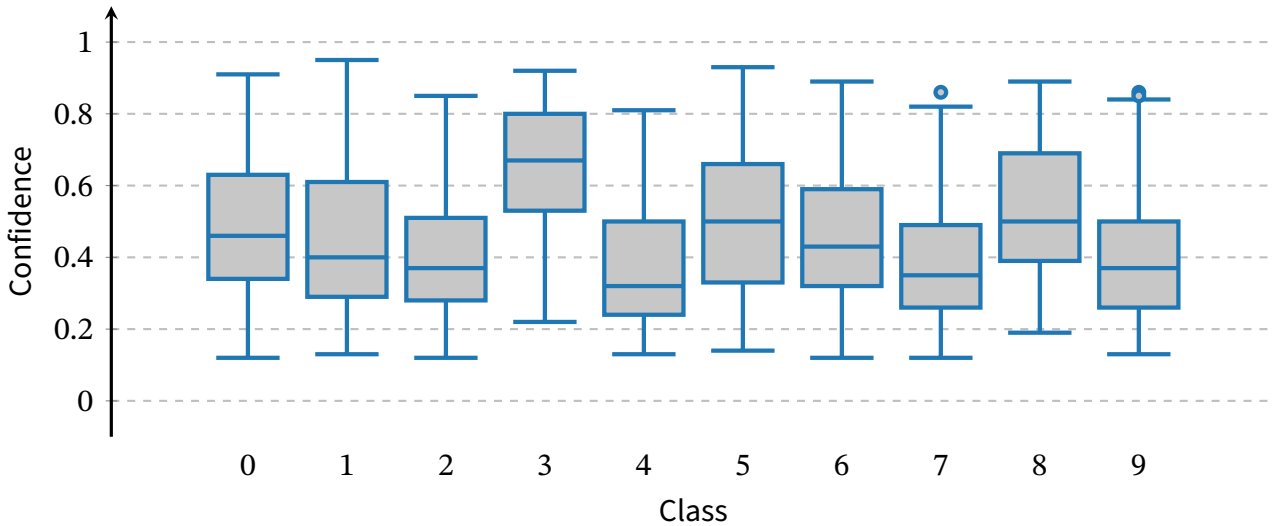
$$\ell_{scan} = -\frac{1}{|D|} \sum_{x \in D} \sum_{k \in N_x} \log \langle \Phi_\eta(x), \Phi_\eta(k) \rangle + \lambda \sum_{c \in C} \Phi_\eta^c \log \Phi_\eta^c \quad (4.3)$$

$$\text{with } \Phi_\eta^c = \frac{1}{|D|} \sum_{x \in D} \Phi_\eta^c(x)$$

The exact number of final classes in  $C$  is unknown. Under the assumption that a small number

of classes might generalize too much and a large number allows for a more detailed distinction, the number of classes is set to 1024 (Gassilloud, 2022). The confidence for an exemplary subset of classes is shown in Figure 4.10. The batch size is set to 1024 and the network is trained for 50 epochs.

Again to summarize, the loss function aims to assign an augmented pair of patches consisting of  $x$  and one of its nearest neighbors  $k$  to the same cluster, while maintaining a uniform spread of predictions across all clusters. The next step aims to improve the obtained results.



**Figure 4.10** Boxplot showing sample confidences after the second step of the SCAN framework for a subset of classes (10 out of 1024), with outliers 1.5×interquartile range (IQR) above/below the third/first quartile marked as dots. The median confidence on the full dataset lies at 44%.

### Improving of Clustering Results through Self-labeling

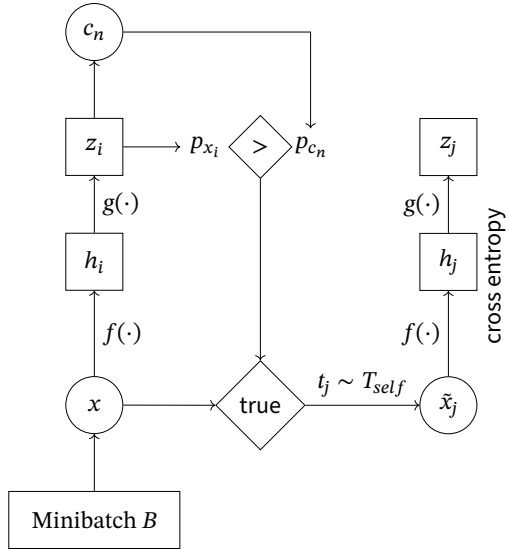
In the previous step the network  $\Phi_\eta$  was trained with the aim to assign patches, who share similar semantic features, to the same cluster. Semantic similar patches were defined for each patch  $x \in D$  by its 20 nearest neighbors  $N_x$ . But an assumption can be made, that some of the nearest neighbors do not share a high enough semantic similarity with the patch to be assigned to the same cluster. That leads to predictions with a lower confidence, which is defined for each patch by applying the softmax activation function to the logits. To counter this problem, a self-labeling step will be performed to retrain the network  $\Phi_\eta$  using patches with a high confidence. Those patches can be seen as *prototypes* of each class. The original implementation of the SCAN framework uses a threshold of  $p = 99.9\%$  to select the prototypes. As visible in Figure 4.10, a confidence that high is not achieved, and while the threshold could be lowered, a fixed one is not a good solution due to the variance of the confidence for each class (Van Gansbeke et al., 2020; Gassilloud, 2022).

The chosen solution is a flexible threshold  $p_c$  with a confidence based on each class  $c \in C$ , as defined in Equation (4.4), with  $cf_c$  as the confidence over samples of the class  $c$  and  $\sigma(cf_c)$  as the standard deviation of those samples. Additionally, the confidence  $p_x$  to select a patch  $x$  as a prototype is not based on the confidence of belonging to its assigned cluster alone ( $\max(z_i)$ ), but on its three highest confidence values as defined in Equation (4.5).



$$p_c = \max(cf_c) - \sigma(cf_c) \quad (4.4)$$

$$p_x = \sum_3 \max(z_i) \quad (4.5)$$



**Figure 4.11** Components of the third step of the SCAN framework (own illustrations after Gassilloud (2022)). Each patch  $x \in B$  is passed through  $f(\cdot)$  and  $g(\cdot)$  to obtain its class  $c_n = \arg \max(z_i)$ . The  $\max_3$  confidence  $p_{x_i}$  of the patch is compared with the confidence threshold  $p_{c_n}$  of the class. If greater, the patch is augmented and used to retrain the network using the loss function  $l_{CE}$ .

That means each patch  $x \in B$  is passed through the network  $\Phi_\eta$  to obtain its class  $c_n$  and  $\max_3$  confidence  $p_{x_i}$ . If the confidence is greater than the threshold of the class ( $p_{x_i} > p_{c_n}$ ), the patch is designated as  $k$  and a data augmentation  $t_j \sim T_{self}$  is applied to it (with details again provided in Appendix A.2). The obtained patch  $\tilde{k}_j$  is then used to retrain the network by minimizing the cross entropy loss function  $l_{CE}$  as defined in Equation (4.6) (Van Gansbeke et al., 2020; Gassilloud, 2022). The network is trained for 200 epochs with a batch size of 1000.

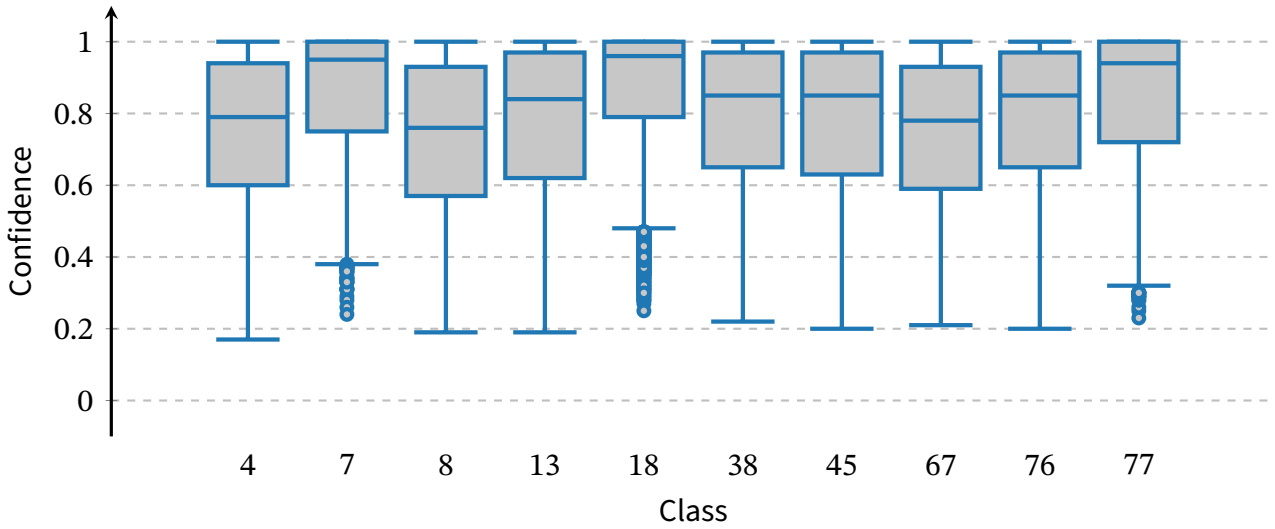
$$l_{CE} = - \sum_{n=1}^C c_n \log(z_n) \quad (4.6)$$

To summarize, the self-label step retrains the network  $\Phi_\eta$  with prototypes of each class to improve results. This results in the updated dataset  $D := \{(x, y = \Phi_\eta(x)) | x \in D\}$  which contains for each patch  $x_i$  the label  $y_i$ . While still maintaining 1024 classes in the cluster head  $g(\cdot)$ , patches are now only assigned to 66 different classes, with Figure 4.12 showing the confidence for an exemplary subset of classes. As visible the confidence for each class is quite high, with a big improvement in comparison to the results of the previous step. Those high confidence values can be interpreted as overconfident results, as discussed by S. Park et al. (2021). To counter this overconfidence, the RUC framework will be used as the next step to obtain the final assignment of each patch  $x \in D$  to a cluster  $c_i \in C_{UB}$ .

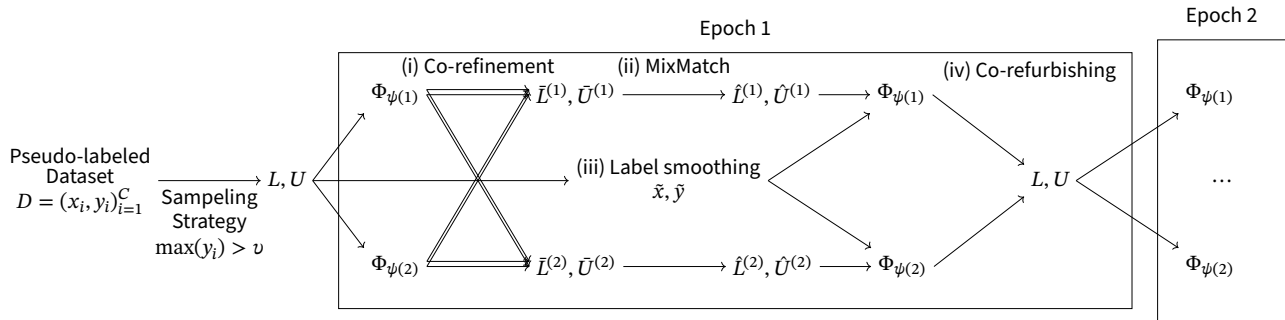
### 4.1.3 Robust Learning for Unsupervised Clustering

The RUC framework was developed by S. Park et al. (2021) with the aim to improve the results of existing methods for unsupervised clustering, as “these ... objectives are bound to producing overconfident results ... due to the dense grouping among clusters” (S. Park et al., 2021) as well as increasing robustness. To achieve that, the existing model  $\Phi_\eta$  is retrained using clean samples. The framework itself consists of two different components, with the second component further divided:

1. Extracting clean samples using a confidence based strategy



**Figure 4.12** Boxplot showing sample confidences after the final step of the SCAN framework for a subset of classes (10 out of 1024), with outliers  $1.5 \times \text{IQR}$  above/below the third/first quartile marked as dots. The median confidence on the full dataset lies at 88%.



**Figure 4.13** Components of the RUC framework (own illustration). The dataset  $D$  is divided in two sets  $L$ , denoting clean samples with a pseudo-label, and  $U$ , denoting unclean samples. Two networks  $\Phi_{\psi(1)}$  and  $\Phi_{\psi(2)}$  are trained for each epoch, in which the labels are co-refined (i) and the data is augmented using MixMatch (ii). Parallel to that a Label smoothing is performed (iii). After performing a training step on the networks, the data is co-refurbished by updating  $L$  (iv).

## 2. Retraining the model with the refined dataset

- (i) Co-refinement
- (ii) MixMatch
- (iii) Label smoothing
- (iv) Co-refurbishing

An overview of the components is shown in Figure 4.13.

### Extracting Clean Samples using a Confidence Based Strategy

As stated, the first step of the RUC framework is to apply a confidence based strategy to extract clean samples on the dataset  $D = \{(x_i, y_i)\}_{i=1}^C$ .  $D$  is the result of the SCAN framework, containing for each  $x_i$  the *pseudo-label*  $y_i = \Phi_{\gamma}(x_i)$ . This dataset is now divided into two disjoint sets  $D = L \cup U$ , with  $L$  considered as clean with a credible pseudo-label thus forming the labeled dataset  $(x, y) \in L$ .  $U$  on the other side is considered as unclean, which means the pseudo-labels are discarded, resulting in the unlabeled dataset  $u \in U$ . The distinction is based on the confidence value of the pseudo-label  $y$ , which means if  $\max(y) > \nu$  the pseudo-label is considered credible and

is, with its patch  $x$ , assigned to  $L$ . Otherwise, only the patch is assigned to  $U$ . The threshold  $v$  is set to 99%. While that value is very high, it should work well enough considered the high confidence values achieved with the network  $\Phi_\eta$  as seen in Figure 4.12 (S. Park et al., 2021; Gassilloud, 2022). Using the two datasets  $L$  and  $U$ , the next step is to train the refined network  $\Phi_\psi$ .

### Retraining the Model with the Refined Dataset

Since a single network has a vulnerability of overfitting, two networks  $\Phi_{\psi(1)}$  and  $\Phi_{\psi(2)}$  are trained each epoch iteratively (meaning each component described below is performed twice, with the networks  $\Phi_{\psi(1)}$  and  $\Phi_{\psi(2)}$  switched), with each minibatch  $B$  sampled from  $D = L \cup U$ . The networks are initialized using the weights of the previously trained network  $\Phi_\eta$ . The following (re-) training process consists of three training components: (i) Co-refinement, (ii) MixMatch, (iii) Label smoothing, and an additional component (iv) Co-refurbishing to update  $L$  (S. Park et al., 2021; Gassilloud, 2022):

**(i) Co-refinement** This first component is a label refinement process, using both networks predictions, to create reliable labels on both the labeled dataset  $L$  and the unlabeled dataset  $U$ . This process was originally developed by J. Li, Socher, and Hoi (2019), functioning as follows:

Starting with the clean samples  $L = \{(x_b, y_b) | b \in B\}$  each label  $y$  is co-refined to  $\bar{y}$  by combining  $y$  with the prediction of  $\Phi_{\psi(2)}$  (Equation (4.7)) and sharpening the result (Equation (4.8)). Here  $\omega^{(2)}$  denotes confidence of the other network and  $\kappa$  denotes the sharpening temperature parameter.

$$\bar{y} = (1 - \omega^{(2)}) \cdot y + \omega^{(2)} \cdot \Phi_{\psi(2)}(x) \quad (4.7)$$

$$\bar{y} = \text{Sharpen}(\bar{y}, \kappa) = \bar{y}^{c \frac{1}{\kappa}} / \sum_{c=1}^C \bar{y}^{c \frac{1}{\kappa}}, \quad \text{for } c = 1, 2, \dots, C \quad (4.8)$$

For the unclean samples  $U = \{(u_b) | b \in B\}$  a pseudo-label  $\bar{q}$  is guessed for each patch  $u$  using Equations (4.9) and (4.10). Each patch is augmented using weak augmentations ( $t_w \sim T_{weak}$ ) resulting in  $\tilde{u}_m$  (augmentations are described in Appendix A.2), with  $m = 1, \dots, M$  denoting the  $m$ -th augmentation. In this case  $M = 2$  is used.

$$\bar{q} = \frac{1}{2M} \sum_{m=1}^M (\Phi_{\psi(1)}(\tilde{u}_m) + \Phi_{\psi(2)}(\tilde{u}_m)) \quad (4.9)$$

$$\bar{q} = \text{Sharpen}(\bar{q}, \kappa) = \bar{q}^{c \frac{1}{\kappa}} / \sum_{c=1}^C \bar{q}^{c \frac{1}{\kappa}}, \quad \text{for } c = 1, 2, \dots, C \quad (4.10)$$

Co-refinement produces for  $L^{(1)}$  and  $U^{(2)}$  the refined datasets  $(x, \bar{y}) \in \bar{L}^{(1)}$  and  $(u, \bar{q}) \in \bar{U}^{(1)}$ , who are used in the next step (S. Park et al., 2021; Gassilloud, 2022).

**(ii) MixMatch** The next component is based on the MixMatch framework developed by Berthelot et al. (2019), which is able to increase noise resistance by mixing a patch and its label  $(x, y)$  with the label of a random patch. This is achieved by augmenting two data points  $x_1, y_1$  and  $x_2, y_2$ , sampled from  $\bar{L}^{(1)} \cup \bar{U}^{(1)}$  using Equations (4.11) to (4.14). Here  $\lambda$  is denoted as a random sample from a beta distribution with  $\alpha=4$  (S. Park et al., 2021; Gassilloud, 2022).

$$\lambda \sim \text{Beta}(\alpha, \alpha) \quad (4.11)$$

$$\lambda' = \max(\lambda, 1 - \lambda) \quad (4.12)$$

$$x' = \lambda' x_1 + (1 - \lambda') x_2 \quad (4.13)$$

$$y' = \lambda' y_1 + (1 - \lambda') y_2 \quad (4.14)$$

MixMatch results in the datasets  $(\hat{x}, \hat{y}) \in \hat{L}^{(1)}$  and  $(\hat{u}, \hat{q}) \in \hat{U}^{(1)}$  used to build two different loss functions  $\ell_{\hat{L}^{(1)}}$  and  $\ell_{\hat{U}^{(1)}}$  as defined in Equations (4.15) and (4.16). Both loss functions are used later to define the final loss  $\ell_{ruc^{(1)}}$  (S. Park et al., 2021; Gassilloud, 2022).

$$\ell_{\hat{L}^{(1)}} = \frac{1}{|\hat{L}^{(1)}|} \sum_{\hat{x}, \hat{y} \in \hat{L}^{(1)}} \sum_{n=1}^C \hat{y}_n \log(\Phi_{\psi^{(1)}}(\hat{x}_n)) \quad (4.15)$$

$$\ell_{\hat{U}^{(1)}} = \frac{1}{|\hat{U}^{(1)}|} \sum_{\hat{u}, \hat{q} \in \hat{U}^{(1)}} \|\hat{q} - \Phi_{\psi^{(1)}}(\hat{u})\|_2^2 \quad (4.16)$$

**(iii) Label smoothing** The third component aims to regulate the model to prevent it from making overconfident noisy predictions by adding uniform noise to each label  $y$ , corresponding to a patch  $x$  of the clean dataset  $L = \{(x_b, y_b) | b \in B\}$  resulting in  $\tilde{y}$  (Equation (4.17)). Here  $\varepsilon \sim \text{Uniform}(0, 1)$  is denoted as a smoothing hyper-operator (set to  $\varepsilon = 0.5$ ) and  $|C|$  as the number of classes. In addition, each patch  $x$  is augmented using strong augmentations ( $t_s \sim T_{strong}$ ) resulting in  $\tilde{x}$  (augmentations are described in Appendix A.2) (S. Park et al., 2021; Gassilloud, 2022).

$$\tilde{y} = (1 - \varepsilon) \cdot y + \frac{\varepsilon}{|C| - 1} \cdot (1 - y) \quad (4.17)$$

With  $(\tilde{x}, \tilde{y})$  defined the loss function  $\ell_{L^{(1)}}$  can be build as described in Equation (4.18). The final loss function  $\ell_{ruc^{(1)}}$  can be defined after processing each mini batches in an epoch as in Equation (4.19), with  $\lambda_U$  as a hyperparameter to control the effect of  $\ell_{\hat{U}^{(1)}}$  (the unsupervised loss of MixMatch) (S. Park et al., 2021; Gassilloud, 2022).

$$\ell_{L^{(1)}} = \frac{1}{|L^{(1)}|} \sum_{\tilde{x}, \tilde{y} \in L^{(1)}} \sum_{n=1}^C \tilde{y}_n \log(\Phi_{\psi^{(1)}}(\tilde{x}_n)) \quad (4.18)$$

$$\ell_{ruc^{(1)}} = \ell_{L^{(1)}} + \ell_{\hat{L}^{(1)}} + \lambda_U \cdot \ell_{\hat{U}^{(1)}} \quad (4.19)$$

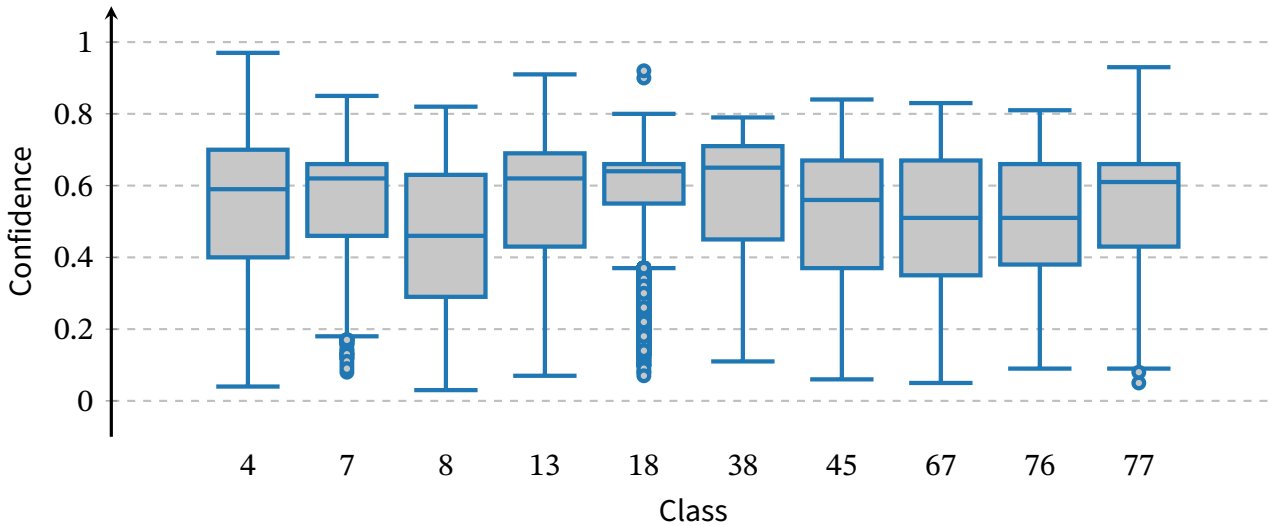
Those three components conclude the training process for each epoch of the RUC framework.

**(iv) Co-refurbishing** This final component aims to update the clean samples  $L$  following Equation (4.20): The prediction of each unclean sample  $u \in U$  of both networks  $\Phi_{\psi(1)}(u)$  and  $\Phi_{\psi(2)}(u)$  is compared. If both predict the same label and the confidence value is greater than 99% the sample and its label are assigned to  $L$  (S. Park et al., 2021; Gassilloud, 2022).

$$L := L \cup \{u \in U \mid \arg \max(\Phi_{\psi(1)}(u)) = \arg \max(\Phi_{\psi(2)}(u)) \wedge \max(\Phi_{\psi(2)}(u) \geq 0.99)\} \quad (4.20)$$

As stated above all four components are repeated in each epoch with  $\Phi_{\psi(1)}$  and  $\Phi_{\psi(2)}$  switched. The RUC framework is trained for 200 epochs with a batch size of 250. This results in the updated dataset  $D := \{(x, y = \Phi_{\psi}(x)) \mid x \in D\}$  containing the new results. Thus, each patch  $x \in D$  can be assigned to a cluster  $c_i \in C_{UB}$ , with  $|C_{UB}| = 66$ . Figure 4.14 shows the confidence for an exemplary subset of classes (10 out of 66). As visible, the confidence is higher than after the second step of the SCAN framework, but lower than after the third one.

This concludes the adaption of the master thesis of Gassilloud (2022). However, before the second part of this thesis, the identification of different city models based on the previous discovered patterns, an additional step needs to be performed: While each pattern is now assigned to one urban patterns, the urban areas themselves need to be reassembled.



**Figure 4.14** Boxplot showing sample confidences after the RUC framework for a subset of classes (10 out of 1024), with outliers  $1.5 \times \text{IQR}$  above/below the third/first quartile marked as dots. The median confidence on the full dataset lies at 61%.

#### 4.1.4 Mosaicking

After the clustering, a label  $y$  is known for each patch  $x \in D$ , as well as their confidence. Additionally, for each patch the position relative to the whole raster image is known. Based on this, as *mosaicking* is performed with the aim to produce a new raster image for each urban area.

As each patch has an overlap, a weighted majority voting is performed to decide the value of each pixel. The weights  $w$  are based on the *confidence* of the patch  $p_x$  as well as a *gaussian kernel* used

for smoothing, as defined in Equation (4.21), with  $v = [-15.5, -14.5, \dots, 14.5, 15.5]$  and  $\sigma = 8$ . Additionally, as the labels are in the range of  $[0, \dots, 1023]$  they are relabeled to the range  $[1, \dots, 66]$ , with 1 having the relative least amount of built-up LCZs and 66 the most.

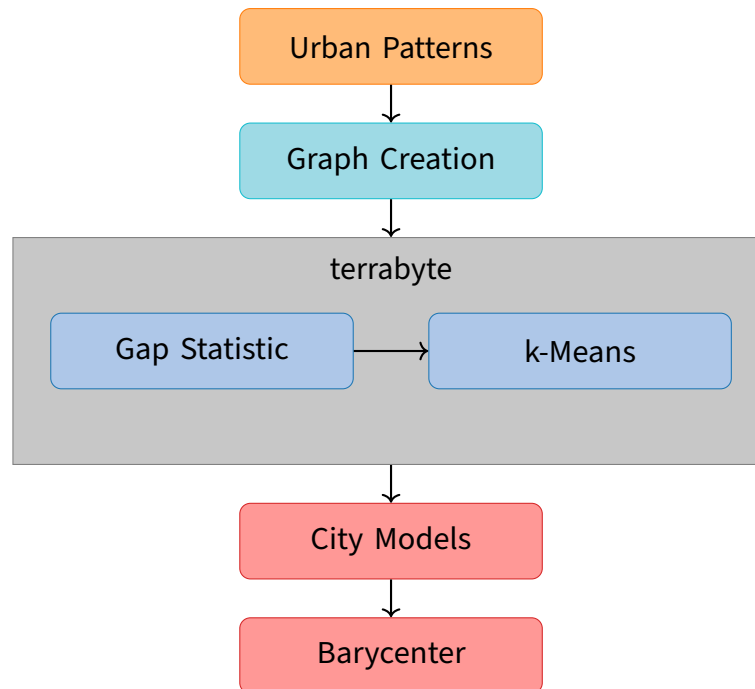
$$w = p_x \cdot (k \otimes k) \quad (4.21)$$

with  $k = \exp\left(-\frac{v}{2 \cdot \sigma^2}\right)$

This results in a raster image for each urban area, with each pixel assigned to one cluster  $c_i \in C_{UB}$ . Based on that, the second part of this thesis can be performed.

## 4.2 Using Graph Representations to find Different Models of Cities

This part aims to cluster graph representations based on the urban areas containing the previous discovered urban patterns. In addition, the barycenter of each cluster shall be created, with each center serving as one specific city model.



**Figure 4.15** Proposed workflow to find models of cities. Based on the urban pattern mosaic created in the last part, a graph representation is created for each urban area. Afterwards the clustering is performed, with the optimal amount of clusters determined by a gap statistic. Based on the resulting clusters, a barycenter for each city model is computed.

The workflow to do this is described in Figure 4.15: A graph is computed based on the previous mosaicked raster images (Section 4.2.1). Afterwards, the clustering is performed using the OT based FGW distance created by Vayer et al. (2019) (Section 4.2.2).

### 4.2.1 Data Preparation

As stated above, the data is available in the form of a raster images describing the patterns in its respective urban area (Figure 4.16a). To create a graph representation, the data is first converted into a vector format, with each region being represented by one polygon. Following that, an adjacency matrix is computed, with edges between polygons defined, if they share at least one vertex. The nodes themselves are initialized with the label of their corresponding region as their value (Figure 4.16b). Due to the reality of the original data it occasionally happens that some graphs have multiple components (one example is shown in Figure B.77a). If that is the case, the biggest connected component is chosen as the graph, and smaller components are discarded. An additional fully connected graph is computed based on the size of each region as node value and the distance between the regions as edge value (Figure 4.16c). This graph is used later to create an alternative representation of the barycenters.

With the data in graph form defining the dataset  $G$ , the clustering using the FGW distance can now be performed.

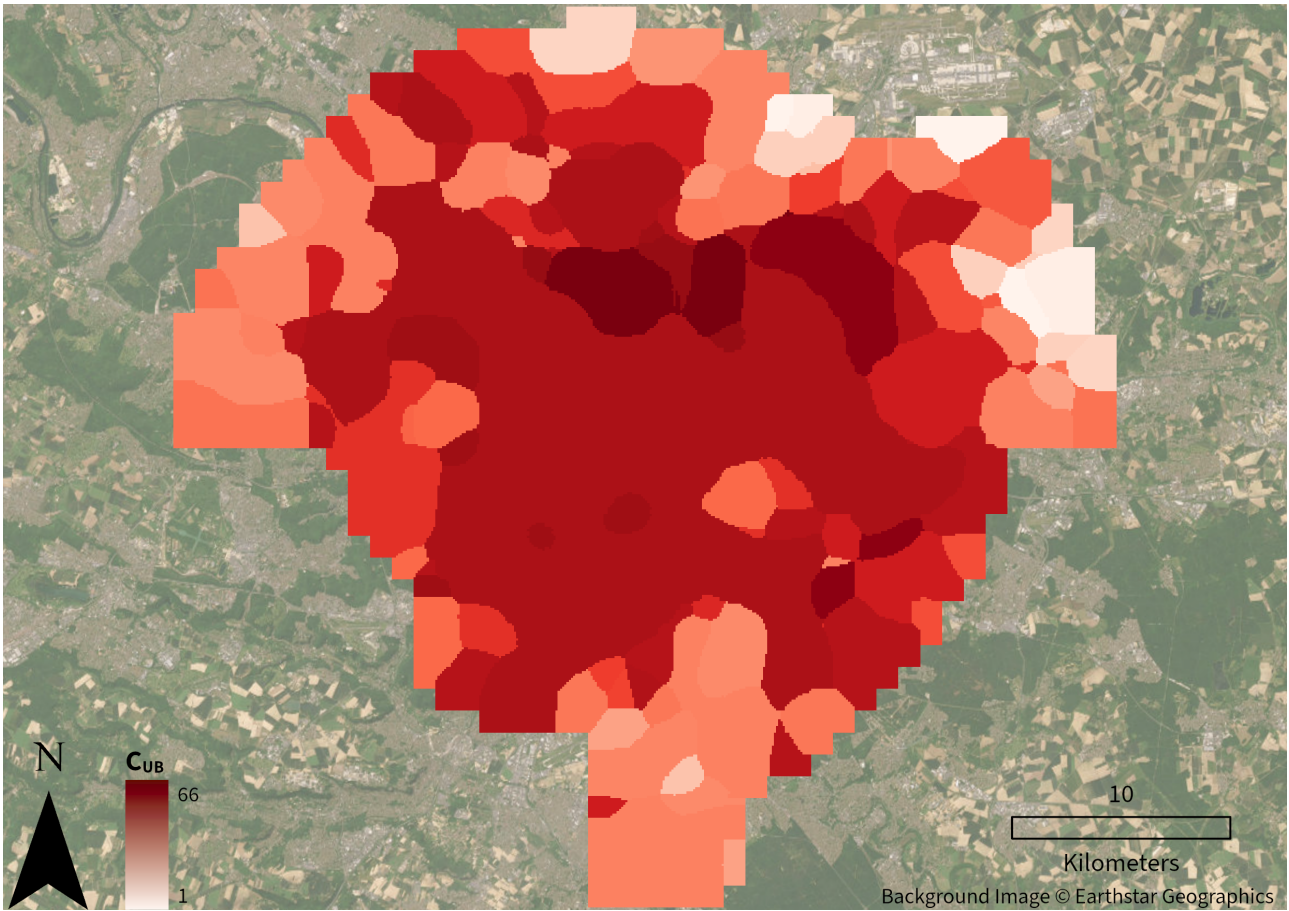
### 4.2.2 k-Means Clustering using a Optimal Transport Based Distance Metric

The k-Means algorithm is a clustering algorithm used to partition  $n$  data points into  $k$  clusters. To do that,  $k$  initial centroids are chosen randomly. Afterwards, each data point is assigned to the centroid with the shortest distance to it. Then new centroids are computed, this time the means of all data points assigned to them. Those two steps are repeated until a termination criterion is achieved or for a specific amount of iterations (Arthur and Vassilvitskii, 2006). In relation to this thesis, it means to assign each graph  $g \in G$  to a cluster  $c_i \in C_{CM}$ , with the centroid of a cluster  $c_i$  defined as the barycenter  $b_i$ . A specific distance metric, the *FGW distance* developed by Vayer et al. (2019) is used to compute the distances between graphs, as well as the barycenters.

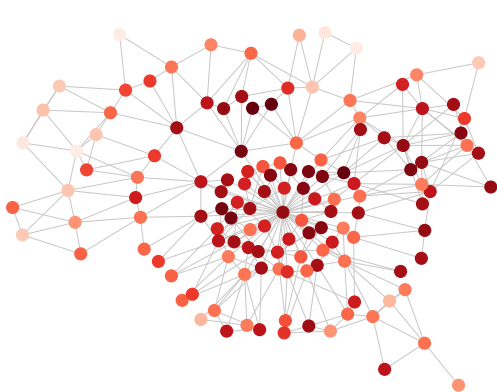
#### k-Means Clustering

As stated, the distance metric to compare graphs used in this thesis is the OT based *Fused Gromov-Wasserstein (FGW) distance*, which was developed by Vayer et al. (2019) to compare graphs based on both features and structure and is a generalization of the Wasserstein and Gromov-Wasserstein distance:

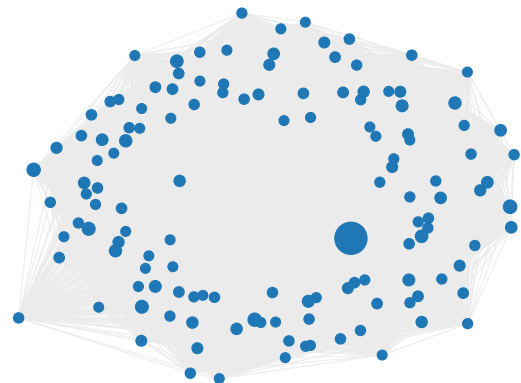
In this case each graph  $g \in G$  is characterized by the tuple  $g = (\mathcal{V}, \mathcal{E}, \mathcal{L}_f, \mathcal{L}_s)$ , with  $\mathcal{V}$  as the nodes and  $\mathcal{E}$  as the edges of the graph.  $\mathcal{L}_f : \mathcal{V} \rightarrow \Omega_f$  denotes the labelling function associating each node  $v_i \in V$  with a feature  $a_i := \mathcal{L}_f(v_i)$  in the feature metric space  $(\Omega_f, d)$  (Vayer et al., 2019). This results in the set of all features  $(a_i)_i$ , who are in this case the *logits*, defined as the output of the network  $\Phi_\psi$  (Section 4.1.3) before applied to the softmax function. Thus, each feature  $a_i$  is defined as a vector with the size of 1024, with each logit calculated as the mean of all logits of the corresponding urban pattern.



(a) Discovered urban patterns of Paris. As visible each pattern is able to form a region.



(b) Graph representation of Paris and its urban patterns. Each node represent one region defined above.



(c) Graph representation of Paris. Instead of its urban patterns and their structure the size and distance between nodes is represented

**Figure 4.16** Paris with urban patterns and its graph representation.



$\mathcal{L}_s : \mathcal{V} \rightarrow \Omega_s$  denotes the mapping of each node  $v_i \in V$  to its structure representation  $o_i := \mathcal{L}_s(v_i)$  in the structure space  $(\Omega_s, C)$ , with  $C : \Omega_s \times \Omega_s \rightarrow \mathbb{R}_+$  as a symmetric application measuring similarity between graph nodes (Vayer et al., 2019). In this case  $(C(i, k) = C(o_i, o_k))_{i,k}$  is the *adjacency matrix* of the graph  $g \in G$ .

With each graph  $g \in G$  defined, it is possible to extend the graphs by describing them as *structured data* defined as a tuple  $\mathcal{S} = (g, r_g)$ .  $r_g$  associates for each node its *weight* (relative importance), described by the histogram  $r \in \sum_n$ , with  $n$  as the number of nodes of the graph  $g$ . In this case, a uniform distribution ( $r_i = 1/n$ ) of the node weights is assumed. This allows each graph  $g \in G$  to be represented as a probability measure as defined in Equation (4.22) (Vayer et al., 2019). Then  $\Pi(r, s)$  is defined as the set of all couplings between  $r$  and  $s$  (Equation (4.23)), with  $\pi_{i,j}$  representing the amount of mass shifted from  $r_i$  to  $s_j$  (Vayer et al., 2019).

$$\mu = \sum_{i=1}^n r_i \delta_{(o_i, a_i)} \quad (4.22)$$

$$\Pi(r, s) = \{\pi \in \mathbb{R}_+^{n \times m}\} \text{ s.t. } \sum_{i=1}^n \pi_{i,j} = r_j, \sum_{j=1}^m \pi_{i,j} = s_i \quad (4.23)$$

Based on that it is possible to define the *FGW distance* between two graphs  $g_1$  and  $g_2$  as in Equation (4.24), with  $M_{AB} = (d(a_i, b_j))_{i,j}$  as  $n \times m$  matrix containing the distance between the features of the two graphs and the two structure matrices  $C_1$  and  $C_2$ , with the similarity between them defined as  $|C_1(i, k) - C_2(j, l)|$ . The trade-off parameter  $\alpha \in [0, 1]$  is set to 0.5 two ensure equal contribution of features and structure (Vayer et al., 2019).

$$\begin{aligned} FGW_{q,\alpha}(\mu, \nu) &= \min_{\pi \in \Pi(r,g)} E_q(M_{AB}, C_1, C_2, \pi) \\ &= \min_{\pi \in \Pi(r,g)} \langle (1 - \alpha)M_{AB}^q + \alpha L(C_1, C_2)^q \otimes \pi, \pi \rangle \\ &= \min_{\pi \in \Pi(r,g)} \sum_{i,j,k,l} (1 - \alpha)d(a_i, b_j)^q + \alpha |C_1(i, k) - C_2(j, l)|^q \pi_{i,j} \pi_{k,l} \end{aligned} \quad (4.24)$$

With this distance metric defined, it is possible to perform a k-Means clustering to compute for each graph  $g \in G$  a label  $h$ , which assigns the graph to a cluster  $c_i \in C_{CM}$ , as described in Algorithm 1. In this case, the initial centroids are chosen randomly from the existing dataset  $G$ . The maximum iteration count is set to 50, with a termination criterion comparing the average distance in a cluster (inertia  $z$ ) of the current iteration with the previous. The clustering is performed for a total of 5 times, with the labels  $h$  chosen from the attempt with the lowest inertia  $z$ .

But one final element of the clustering is still missing: the barycenter as centroids of each cluster. They are defined as the structured data  $\mu$  with minimal distances to a set of structured data  $(\mu_k)_k$  with given structure matrices  $(C_k)_k$ , features  $(B_k)_k$ , and histograms  $(r_k)_k$ . The node count of the barycenter  $N \in \mathbb{N}$  is set the median node count of all the graphs in the set  $(\mu_k)_k$ . Based on that

**Algorithm 1** Pseudocode for k-Means clustering of graphs

---

```

for  $i = 1, \dots, k$  do                                ▷ choose random cities as initial centroids
     $b_i \leftarrow$  choose random graph  $g \in G$ 
end for
for  $i = 0, \dots, 50$  do
    for  $g_j \in G$  do                                    ▷ assign each graph to closest barycenter
         $h_j \leftarrow \arg \min(\text{FGW}_{g_j}^B)$ 
    end for
    for  $j = 1, \dots, k$  do                                ▷ compute new barycenter
         $b_j \leftarrow \min(\sum \text{FGW}_{\{g|h=j\}}^{b_j})$ 
    end for
    if  $z_{i-1} - z < 0.0001$  then                            ▷ check if termination criterion is achieved
        terminate
    end if
end for

```

---

and  $(\lambda_k)_k$  (with  $\sum_k \lambda_k = 1$ ) the aim is to find the features  $A = (a_i)_i$  and structure matrix  $C$  by minimizing Equation (4.25).

While the clustering can now be performed, one problem remains: The optimal amount of clusters  $k$  is unknown. To find it, a *gap statistic* is performed.

$$\begin{aligned}
 & \min_{\mu} \sum_k \lambda_k \text{FGW}_{q,\alpha}(\mu, \mu_k) \\
 & = \min_{C \in \mathbb{R}^{N \times N}, A \in \mathbb{R}^{N \times b}, (\tau_k)_k} \sum_k \lambda_k E_q(M_{AB_k}, C, C_k, \tau_k)
 \end{aligned} \tag{4.25}$$

### Gap Statistic to find the Optimal Number of Clusters

The gap statistic was developed by Tibshirani, Walther, and Hastie (2001) with the aim to provide a clear statistical algorithm for estimating the ideal amount of clusters, instead of relying on visual, like the “elbow method”, interpretations. The procedure works by “comparing the change in within-cluster dispersion” (Tibshirani, Walther, and Hastie, 2001) and is adapted for the graph data as follows:

For two graphs  $g_i \in G$  and  $g_j \in G$  the distance between them is described as  $\text{FGW}(g_i, g_j)$ . Each graph is assigned to one cluster  $c_r \in C_{CM}$ , with  $n_r = |c_r|$  and  $k = |C_{CM}|$ . The  $D_r$  is defined as the sum of pairwise distances of all graphs in cluster  $c_r$  (Equation (4.26)) and  $W_k$  as a measure for the within-cluster variance. The gap value  $Gap(k)$  is then calculated as the comparison between  $\log(W_k)$  and its expectation under a reference distribution of the data (Equation (4.28)) (Tibshirani, Walther, and Hastie, 2001).

$$D_r = \sum_{g_i, g_r \in C_r} \text{FGW}(g_i, g_r) \quad (4.26)$$

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r \quad (4.27)$$

$$\text{Gap}(k) = E^*\{\log(W_k)\} - \log(W_k) \quad (4.28)$$

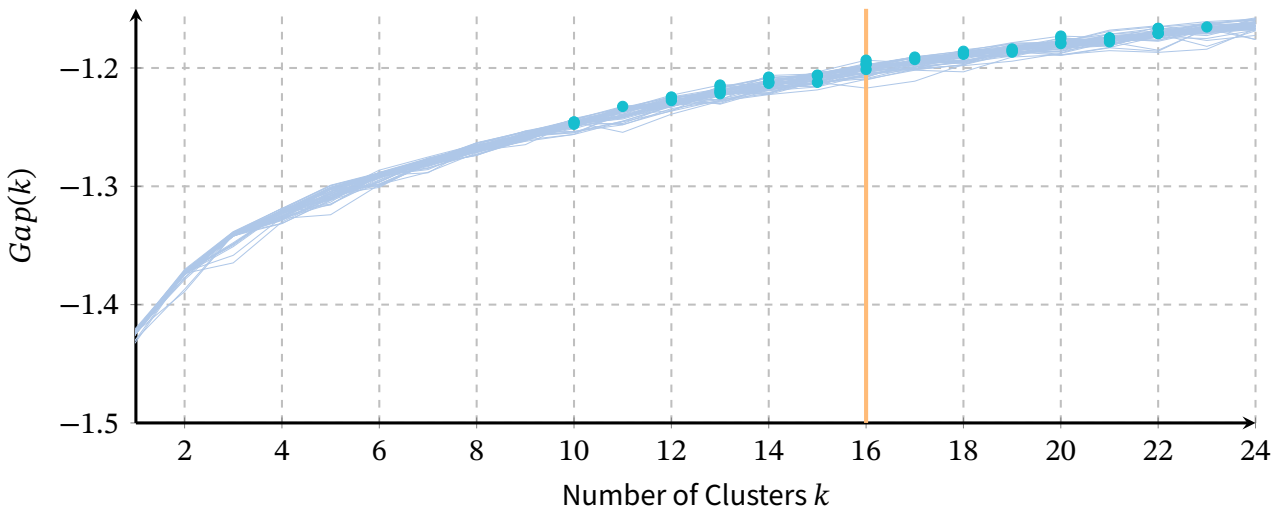
The reference distribution is based on random barycenters forming the dataset  $G^*$ , with each barycenter computed on 100 graphs selected on a uniform distribution. Thus,  $E^*\{\log(W_k)\}$  is defined as the average on  $Q = 10$  copies (Equation (4.29)). With  $sd_k$  defined as the standard deviation of the  $Q$  samples, it is possible to define the metric  $s_k$  as in Equation (4.30). This allows to compute the optimal number of clusters  $\hat{k}$  using Equation (4.31) (Tibshirani, Walther, and Hastie, 2001).

$$E^*\{\log(W_k)\} = \frac{1}{Q} \sum_q \log(w_k^*) \quad (4.29)$$

$$s_k = sd_k \cdot \sqrt{1 + \frac{1}{Q}} \quad (4.30)$$

$$\hat{k} = \min(k) \text{ s.t. } \text{Gap}(k) \geq \text{Gap}(k+1) - s_{k+1} \quad (4.31)$$

As  $\hat{k}$  can vary due to the randomness described above, the gap statistic is performed 60 times in the range 1 to 25. Out of those 60 attempts,  $\hat{k}$  was found 47 times, with the median for  $\hat{k}$  at 16 (Figure 4.17). This makes it possible to assign each graph  $g \in G$  to a cluster  $c_i \in C_{CM}$ , with  $k = |C_{CM}| = 16$ .



**Figure 4.17** Result for the gap statistic, performed 60 times, with  $\hat{k}$  found 47 times. The found optima are depicted with light blue dots, the median optimum is additionally marked.

### Barycenter as Models of Cities

After the clustering the final barycenters, which serve as city models, are computed twice:

One is computed on the standard dataset  $G$  with the logits as features  $(a_i)_i$  and the adjacency matrix as structure matrix  $C$ . It should be noted here that the computed structure matrix  $C$  of each barycenter contains values in the range  $[0, 1]$ . Thus, a threshold needs to be set to transform this matrix into a standard adjacency matrix, which is discovered iteratively, and is defined as the minimum value for which the graph forms one connected component. Additionally, each feature of the barycenter is applied to a softmax function with the aim to get class labels again as the final node feature.

The second barycenter is computed with the area of the regions corresponding to each node as features, and the distance between them as structure matrix, with the aim to compute an alternative city model showing geographic and not pure structural information.

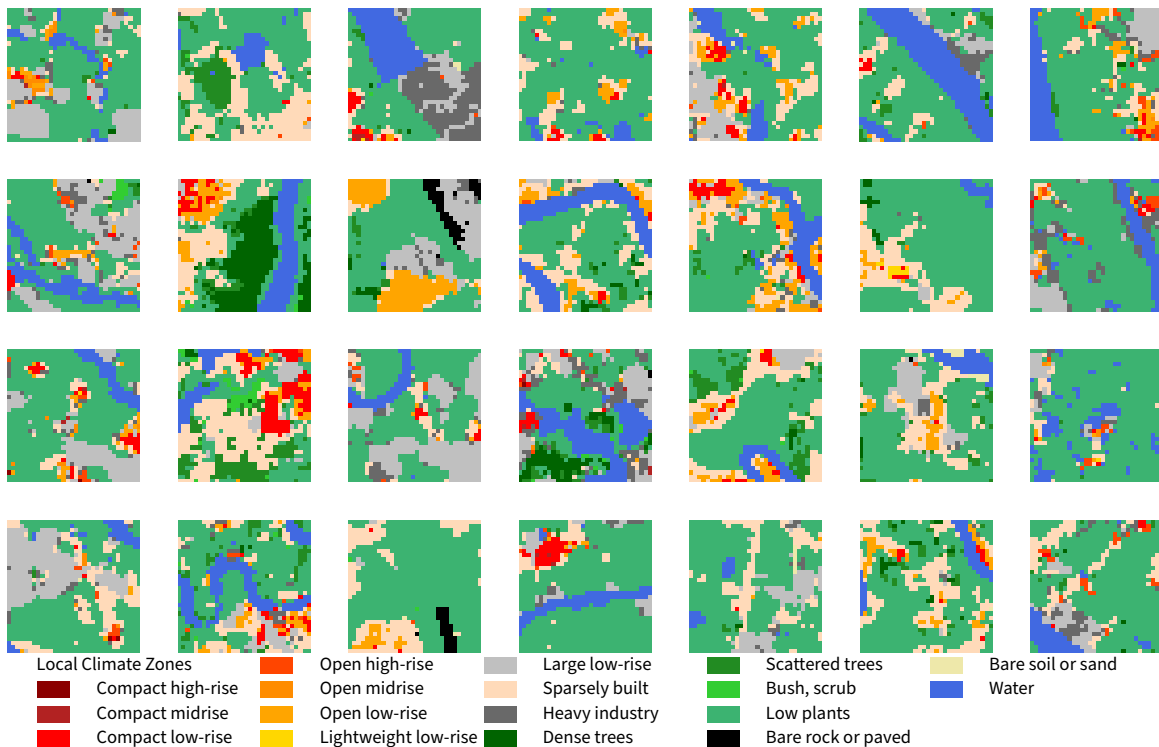
# 5 Results

This chapter presents the results, obtained from the methodology described in Chapter 4, applied on the data presented in Section 3.2. As with the rest of this thesis, this chapter is separated in two parts, with Section 5.1 covering the found urban patterns and Section 5.2 the city models.

## 5.1 Unsupervised Deep Learning Approach to find Urban Patterns

The aim of the part of this thesis was to find urban patterns through unsupervised learning, using the following summarized methodology: A clustering based on two deep learning frameworks was applied on patches of uniform size, extracted from a land cover classification inside urban areas using a parameter based encoding with the aim to introduce geographical and morphological parameters in the process. That resulted in the assignment of each patch  $x \in D$  to one class  $c_i \in C_{UB}$ , with  $|C_{UB}| = 66$  and each patch forming one urban pattern.

A random extract of patches within  $c_{UB_3}$  is shown in Figure 5.1. While this thesis does not have any ground truth data, and as such cannot compute any accuracy measures, a visual interpretation shows similarity across all patches. As visible, the most occurring element is *LCZ D – Low plants*, in combination with *LCZ G – Water* forming a line through it, as well as *LCZ 8 – Large low-rise* and *LCZ 9 – Sparsely built*.



**Figure 5.1** An overview of a random sample of extracted patches in  $c_{UB_3}$  with a total of 1998 patches.

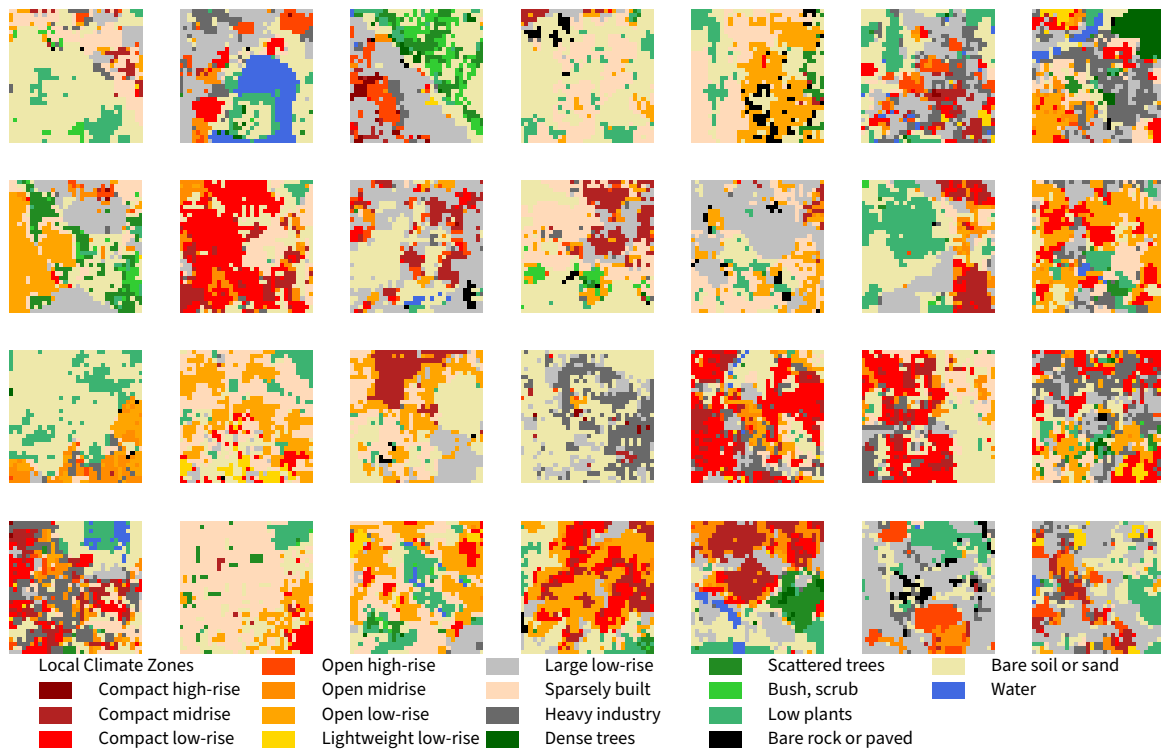


Figure 5.2 An overview of a random sample of extracted patches in  $c_{UB_{24}}$  with a total of 3189 patches.

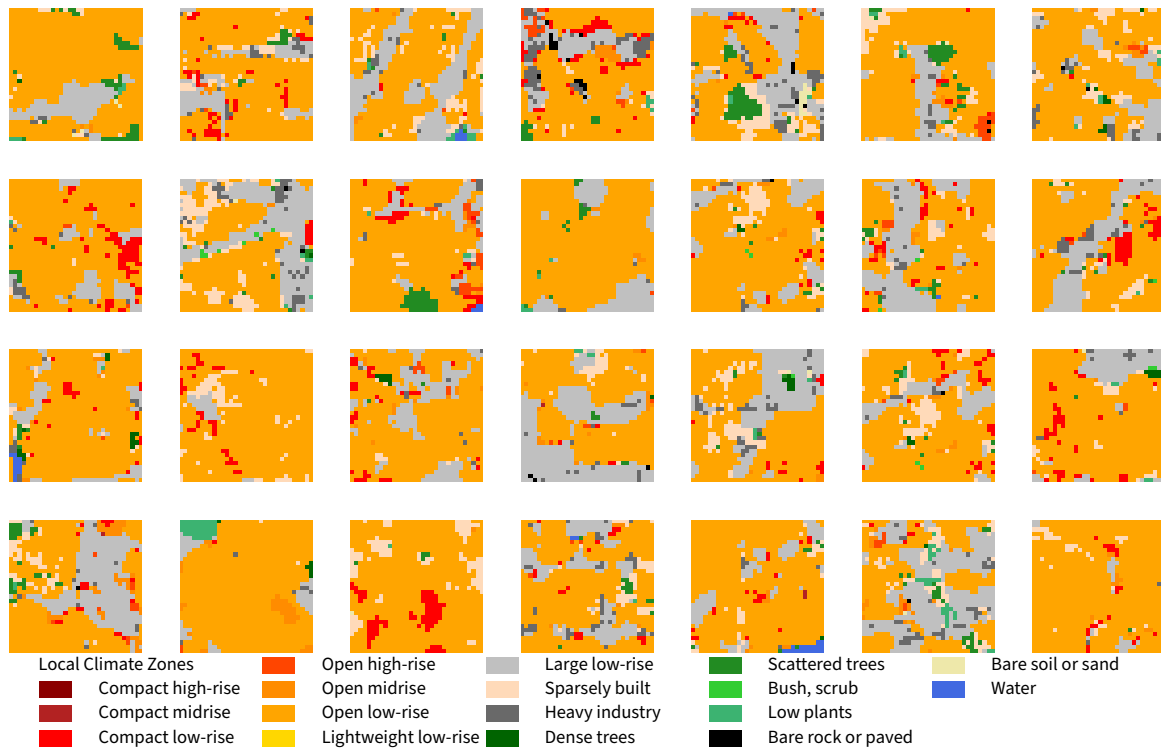


Figure 5.3 An overview of a random sample of extracted patches in  $c_{UB_{62}}$  with a total of 2070 patches.

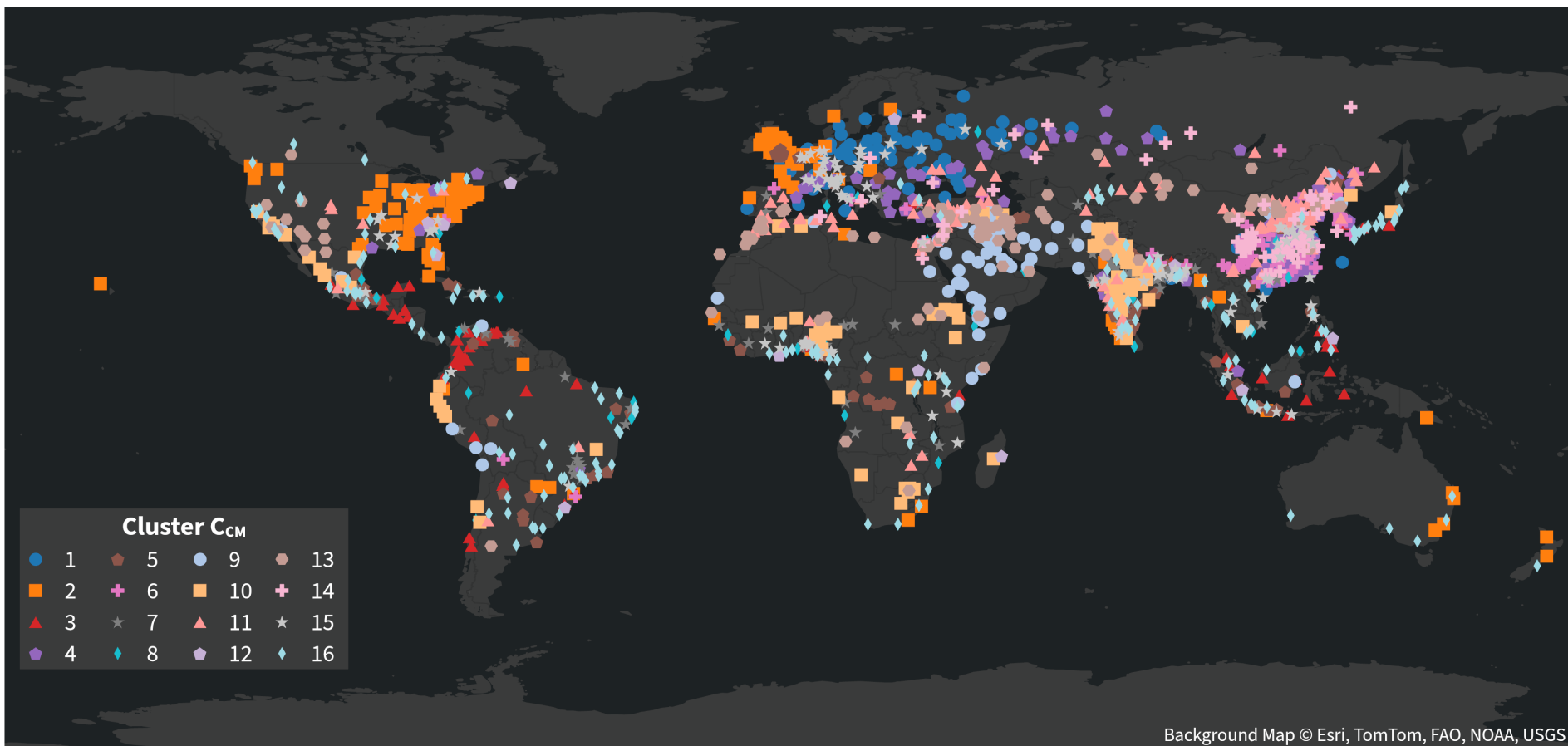
Another example is shown in Figure 5.2, with patches within  $c_{UB_{24}}$ . A visual similarity is also visible here, but it is less remarkable than in the previous example. In most patches *LCZ 9 – Sparsely built* as well as *LCZ F – Bare soil or sand* is very noticeable, often in combination with *LCZ 3 – Compact low-rise*, *LCZ 6 – Open low-rise*, and *LCZ 8 – Large low-rise*.

A third example with patches within  $c_{UB_{62}}$  is shown in Figure 5.3. Again, a high visual similarity is visible, with the patches mostly composed of *LCZ 6 – Open low-rise* and *LCZ 8 – Large low-rise*, with *LCZ 3 – Compact low-rise* and *LCZ 10 – Heavy industry* as additional elements.

Additional examples of all other classes with random patches within them are shown in Appendix B.1. While this thesis does not have a way to specify accuracy or correctness of the results (as stated before), an overview of them shows generally consistent patterns in one class, with different patterns across all patches. A discussion of those results will be presented in Section 6.1.

## 5.2 Using Graph Representations to find Different Models of Cities

The aim of the second part of this thesis was to cluster graph representations based on the previous obtain urban patterns, and afterwards to compute barycenter serving each as a model of one city cluster. The clustering was performed using a k-Means algorithm with a custom distance function, the FGW distance, resulting in the assignment of each graph  $g \in G$  to one cluster  $c_i \in C_{CM}$ , with  $|C_{CM}| = 16$ . Afterwards two different barycenters were computed for each cluster: one based on the structure and the urban patterns, the other one based on the size of the regions formed by the urban patterns and the distance between them. An overview of the final result is shown in Figure 5.4, the individual classes are shown in Figures 5.5 to 5.20, which will be discussed in Section 6.2.

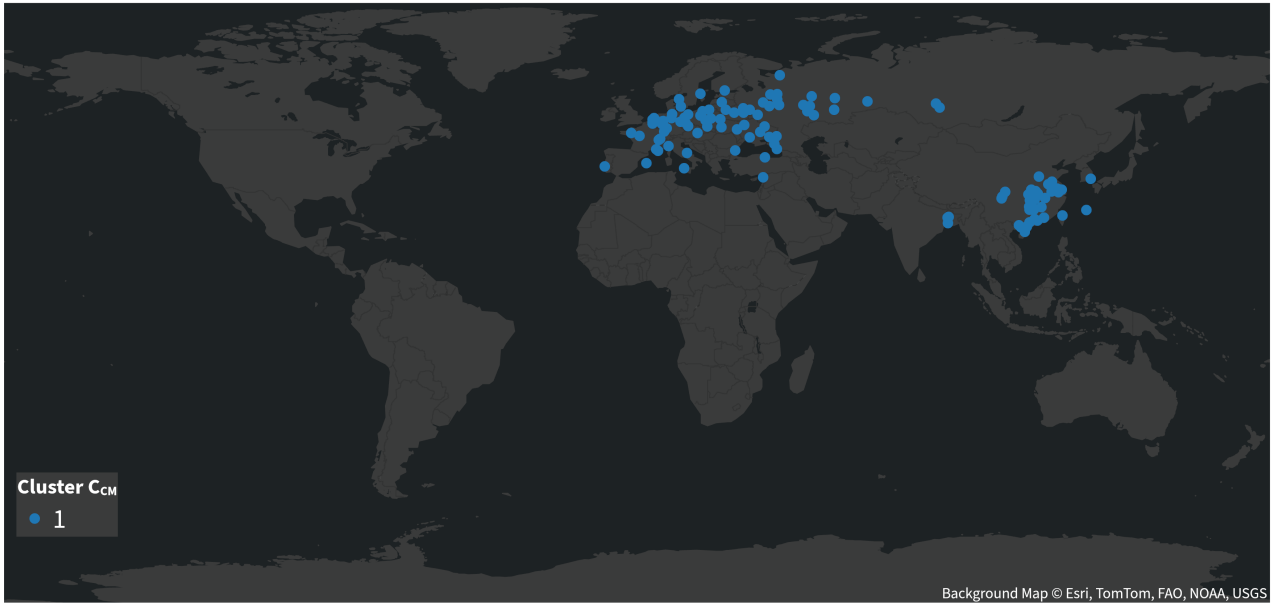


**Figure 5.4** World map showing an overview of all obtained clusters  $C_{CM}$ .

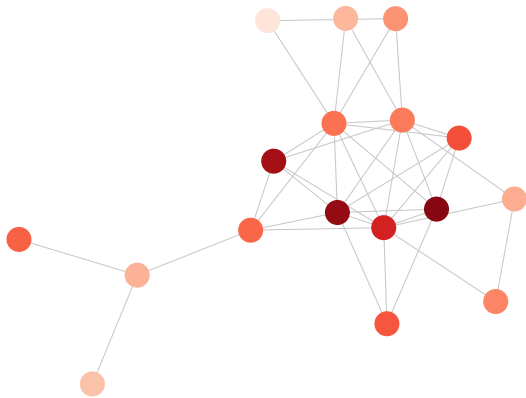


### Cluster $c_{CM_1}$

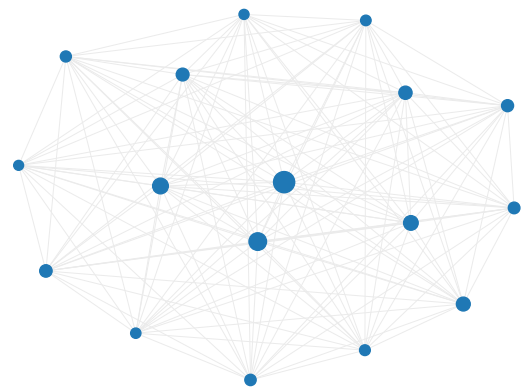
Cluster  $c_{CM_1}$  with cities mostly in Middle and Eastern Europe (including Russia), as well as Eastern China. Three additional cities are near the border of India and Bangladesh, one on a small Japanese island, and one in South Korea. A total of 122 cities are assigned to this cluster. The left barycenter is showing one main component, with multiple nodes and smaller component spreading out from it, with the inner nodes having more built-up area than the outer nodes. The right barycenter is showing two bigger nodes in the center, surrounded with smaller ones. Two cities and their graphs assigned to this cluster are shown in Figure B.64.



(a) World map showing all cities in  $c_{CM_1}$



(b) Barycenter of  $c_{CM_1}$  based on urban patterns and structure

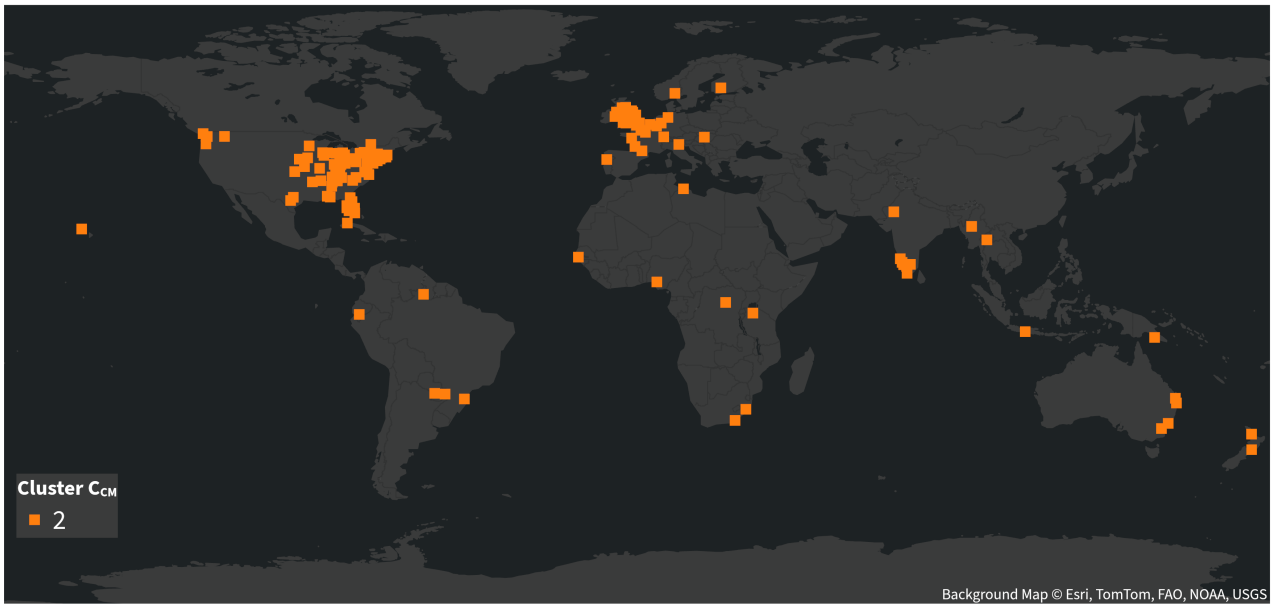


(c) Barycenter of  $c_{CM_1}$  based on area and distance

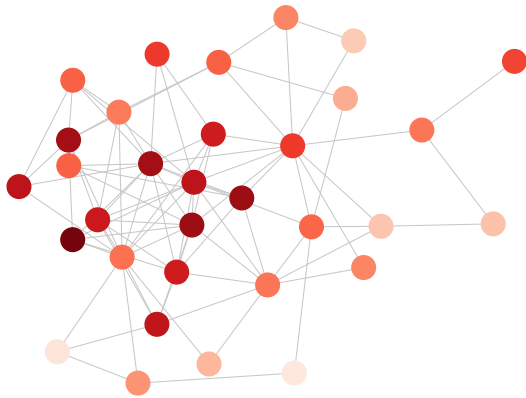
**Figure 5.5** Overview of all cities within  $c_{CM_1}$  and two different barycenters of that cluster.

### Cluster $c_{CM_2}$

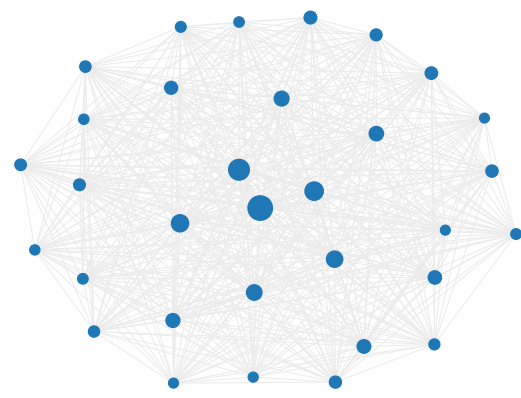
Cluster  $c_{CM_2}$  with cities mostly in Great Britain, Eastern USA, South India, and Eastern Australia. Additional cities are in Western Europe, South and Middle Africa, South America, Hawaii, and Oceania. A total of 136 cities are assigned to this cluster. The left barycenter is showing a lot of closely connected nodes, and additional nodes on one side with less built-up area on average. The right barycenter is showing the three biggest nodes in the center, surrounded by two rings of smaller nodes. Two cities and their graphs assigned to this cluster are shown in Figure B.65.



(a) World map showing all cities in  $c_{CM_2}$



(b) Barycenter of  $c_{CM_2}$  based on urban patterns and structure

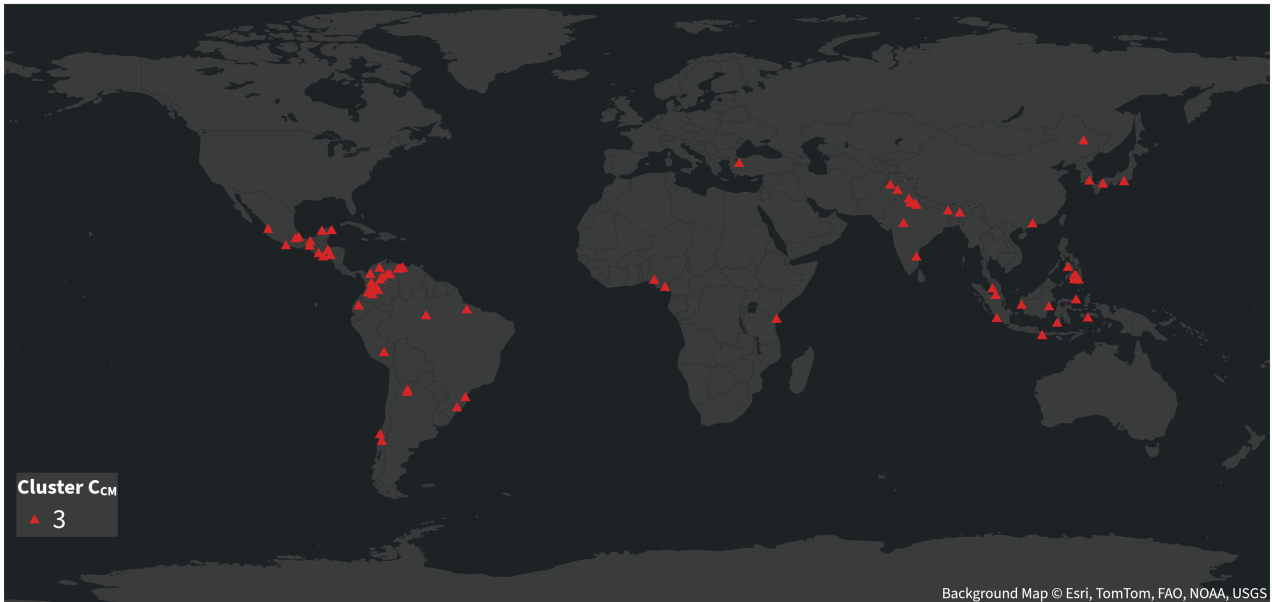
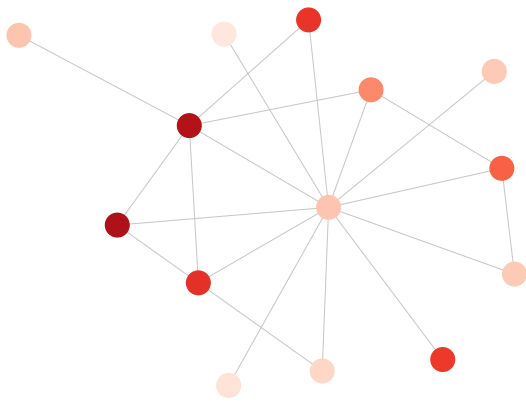
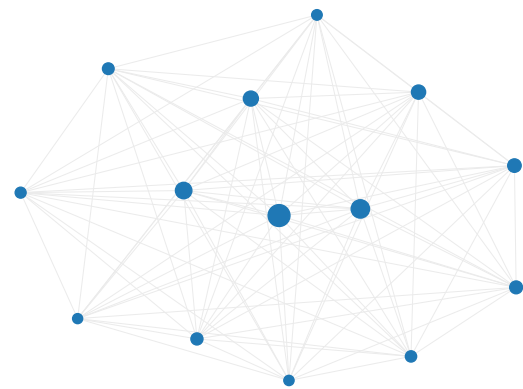


(c) Barycenter of  $c_{CM_2}$  based on area and distance

**Figure 5.6** Overview of all cities within  $c_{CM_2}$  and two different barycenters of that cluster.

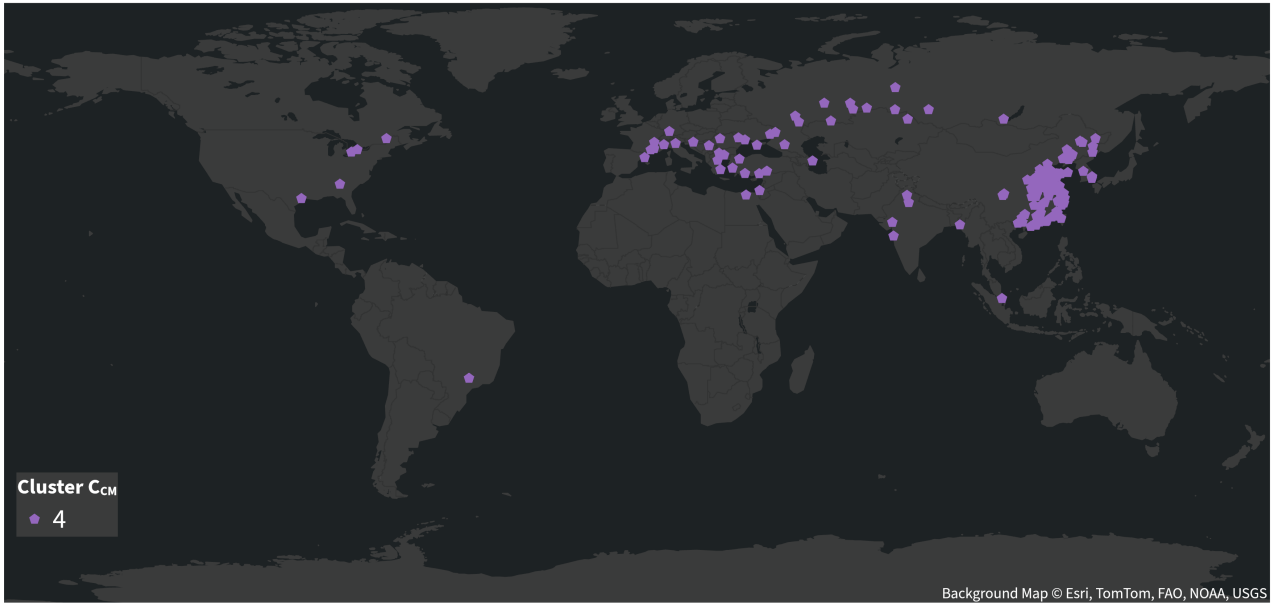
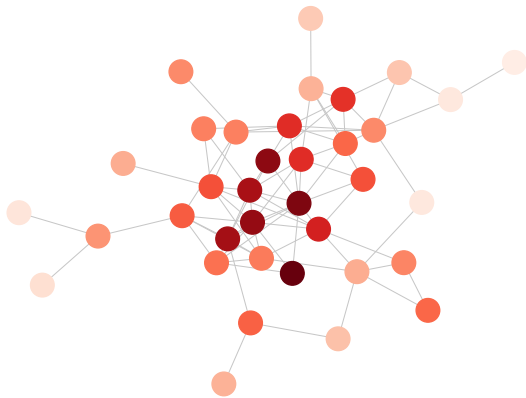
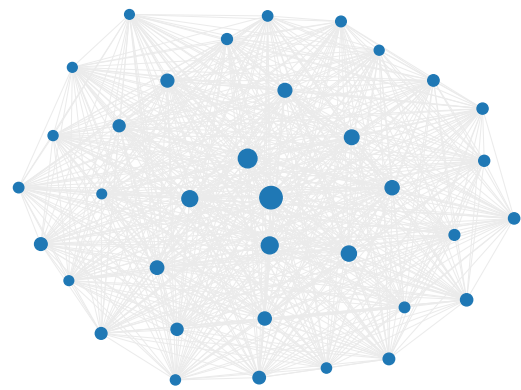
**Cluster  $c_{CM_3}$** 

Cluster  $c_{CM_3}$  with cities mostly in Middle America, Northern India and Pakistan, and South East Asia and Oceania. Additional cities are in South America, Middle Africa, Japan, China, and Korea. A total of 75 cities are assigned to this cluster. The left barycenter is showing one central node with less built-up area than the surrounding ring, the right barycenter one big central node. Two cities and their graphs assigned to this cluster are shown in Figure B.66.

(a) World map showing all cities in  $c_{CM_3}$ (b) Barycenter of  $c_{CM_3}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_3}$  based on area and distance**Figure 5.7** Overview of all cities within  $c_{CM_3}$  and two different barycenters of that cluster.

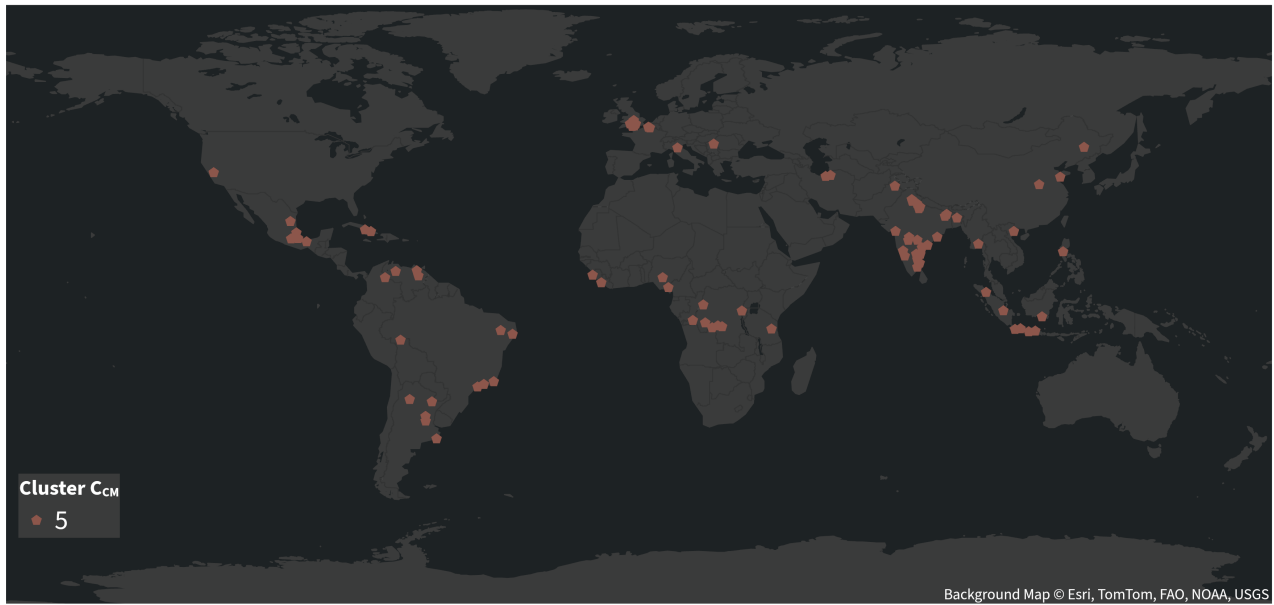
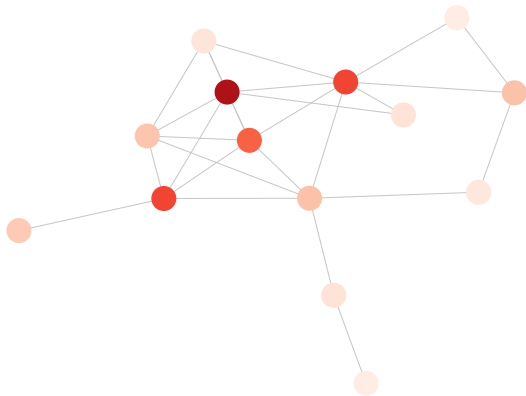
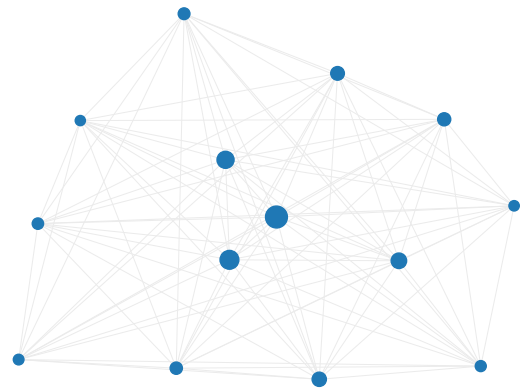
**Cluster  $c_{CM_4}$** 

Cluster  $c_{CM_4}$  with cities mostly in South and East Europa (including Russia) and East China. Other cities are in Eastern USA and Canada, as well as India. One city each is in Brazil and Singapore. A total of 158 cities are assigned to this cluster. The left barycenter is showing a closely connected component in the center with a high amount of built-up area, with additional nodes at the edge. The right barycenter is showing the four biggest nodes in the center, surrounded by two rings of smaller nodes. Two cities and their graphs assigned to this cluster are shown in Figure B.67.

(a) World map showing all cities in  $c_{CM_4}$ (b) Barycenter of  $c_{CM_4}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_4}$  based on area and distance**Figure 5.8** Overview of all cities within  $c_{CM_4}$  and two different barycenters of that cluster.

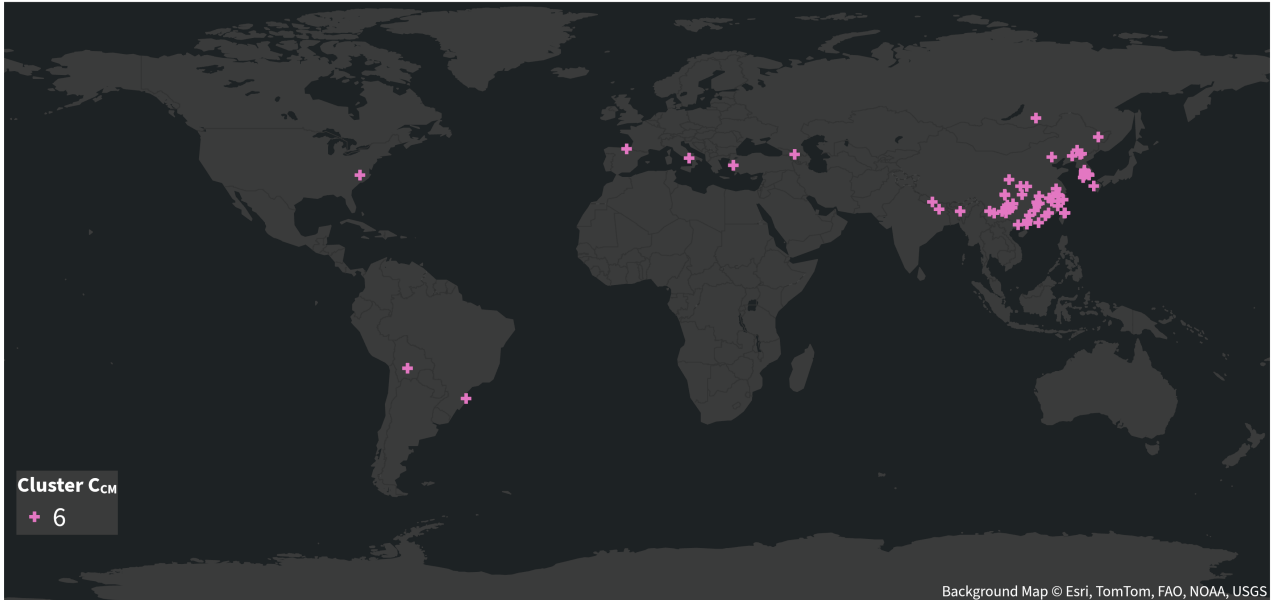
**Cluster  $c_{CM_5}$** 

Cluster  $c_{CM_5}$  with cities mostly in South America, Middle Africa, Mexico, and India. Additional cities are in Europe (mostly Great Britain), South East Asia, Northern China, and Cuba. A total of 82 cities are assigned to this cluster. The left barycenter is showing a few closely connected nodes and pathlike extensions, with the inner nodes having more built-up area. The right barycenter is showing the three biggest nodes in the center, with most of the others as a ring surrounding them. Two cities and their graphs assigned to this cluster are shown in Figure B.68.

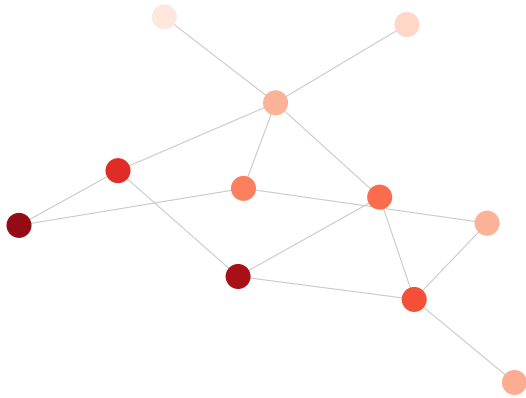
(a) World map showing all cities in  $c_{CM_5}$ (b) Barycenter of  $c_{CM_5}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_5}$  based on area and distance**Figure 5.9** Overview of all cities within  $c_{CM_5}$  and two different barycenters of that cluster.

**Cluster  $c_{CM_6}$** 

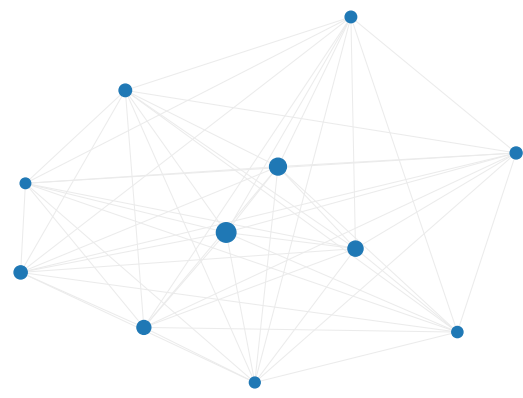
Cluster  $c_{CM_6}$  with cities mostly in Middle and South Eastern China as well as South Korea. A few are in America, Europe, and Central Asia. A total of 63 cities are assigned to this cluster. The left barycenter is showing most nodes with only two or three edges, the right barycenter is having the three biggest nodes in the center. Two cities and their graphs assigned to this cluster are shown in Figure B.69.



(a) World map showing all cities in  $c_{CM_6}$



(b) Barycenter of  $c_{CM_6}$  based on urban patterns and structure

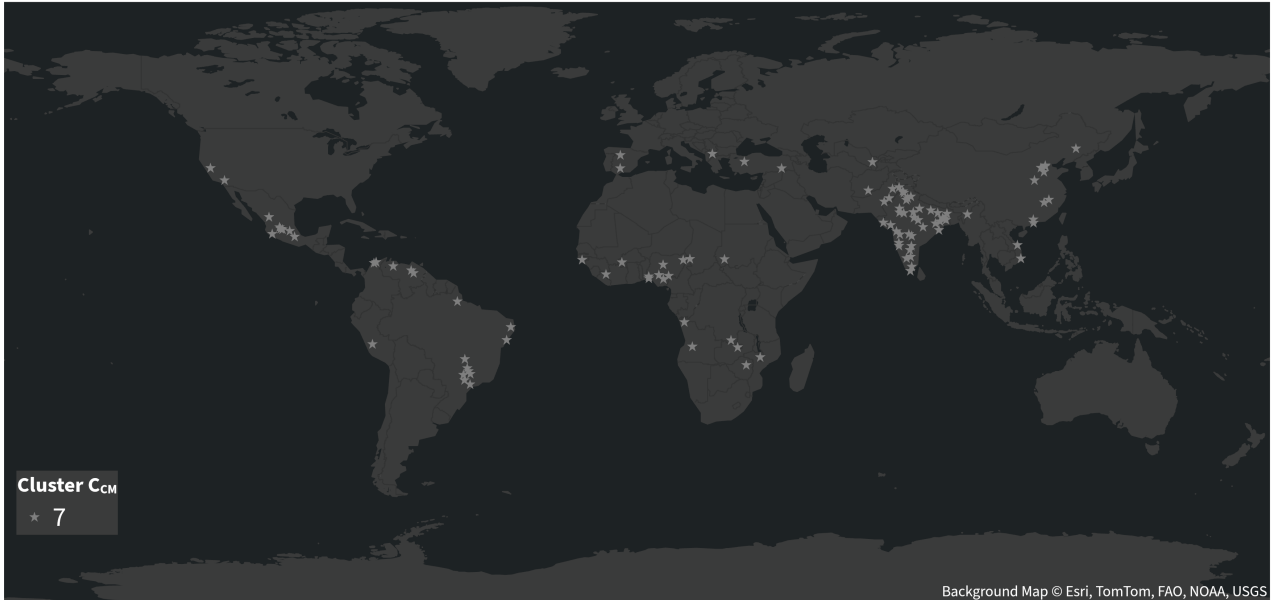
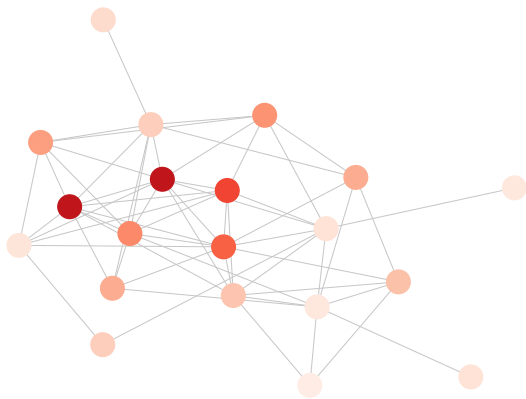
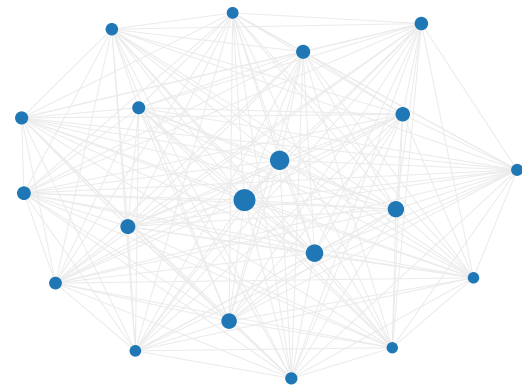


(c) Barycenter of  $c_{CM_6}$  based on area and distance

**Figure 5.10** Overview of all cities within  $c_{CM_6}$  and two different barycenters of that cluster.

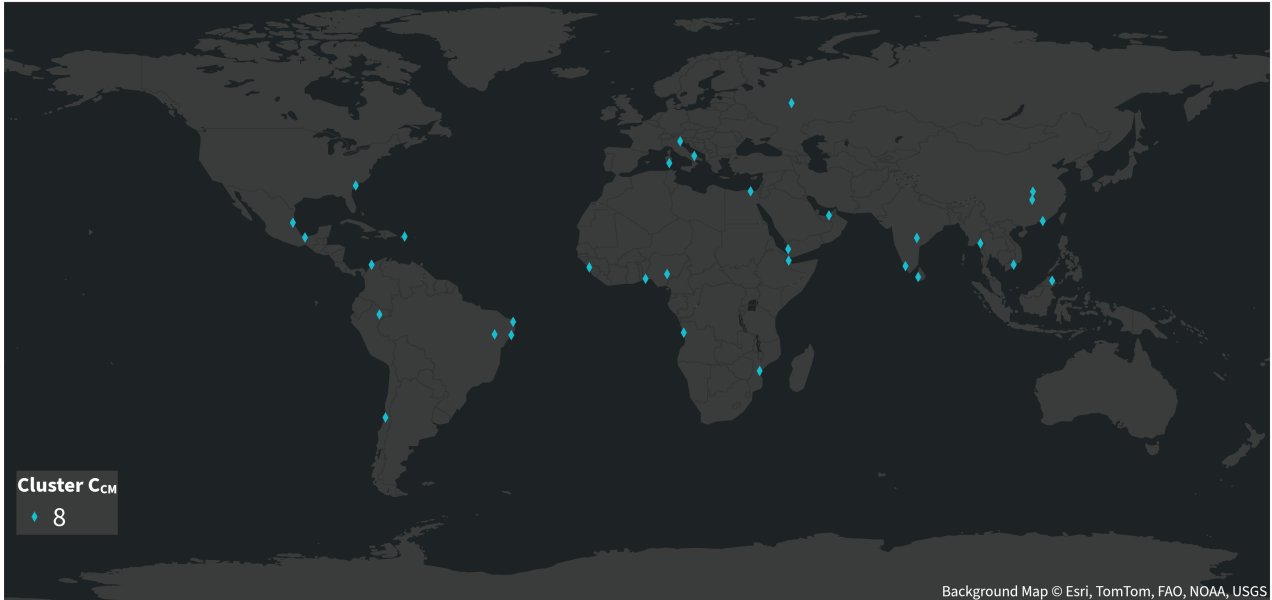
**Cluster  $c_{CM_7}$** 

Cluster  $c_{CM_7}$  with cities mostly in India, Western and Southern Africa, Middle and South America and East Asia. A total of 103 cities are assigned to this cluster. The left barycenter is showing almost all nodes (except three) closely connected with no clear center, but only a few nodes with a high amount of built-up area. The right barycenter has two bigger nodes surrounded by two rings. Two cities and their graphs assigned to this cluster are shown in Figure B.70.

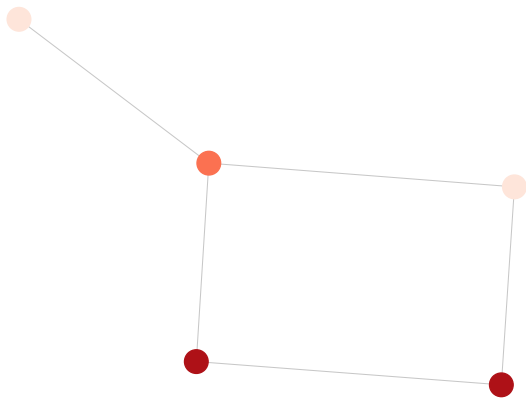
(a) World map showing all cities in  $c_{CM_7}$ (b) Barycenter of  $c_{CM_7}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_7}$  based on area and distance**Figure 5.11** Overview of all cities within  $c_{CM_7}$  and two different barycenters of that cluster.

**Cluster  $c_{CM_8}$** 

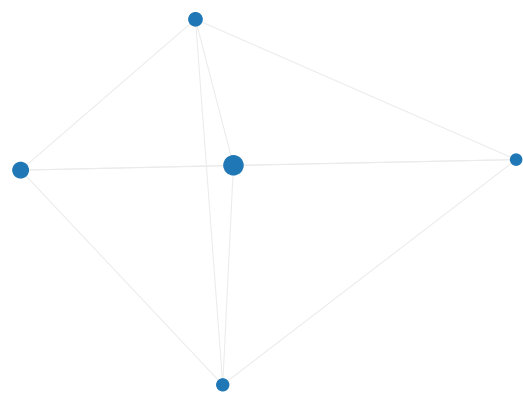
Cluster  $c_{CM_8}$  with cities without a clear geographical center, and the cities with very few nodes according to the barycenter. Two cities and their graphs assigned to this cluster are shown in Figure B.71.



(a) World map showing all cities in  $c_{CM_8}$



(b) Barycenter of  $c_{CM_8}$  based on urban patterns and structure



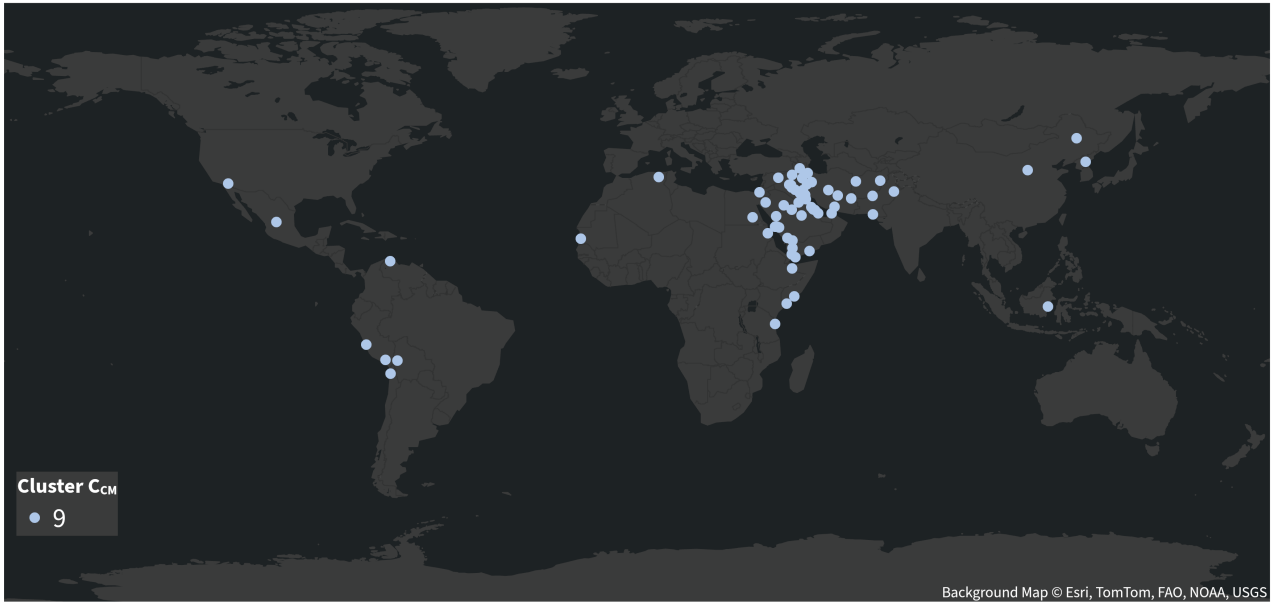
(c) Barycenter of  $c_{CM_8}$  based on area and distance

**Figure 5.12** Overview of all cities within  $c_{CM_8}$  and two different barycenters of that cluster.

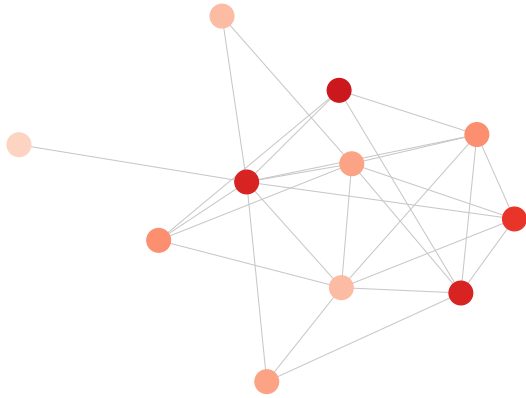


### Cluster $c_{CM_9}$

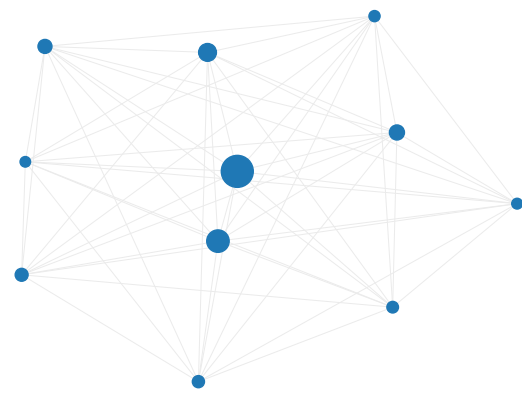
Cluster  $c_{CM_9}$  with cities mostly in the Middle East. Additional cities are in Eastern Africa and Western America. A total of 64 cities are assigned to this cluster. The left barycenter is showing most nodes very closely connected, the right barycenter two bigger, central nodes. Two cities and their graphs assigned to this cluster are shown in Figure B.72.



(a) World map showing all cities in  $c_{CM_9}$



(b) Barycenter of  $c_{CM_9}$  based on urban patterns and structure

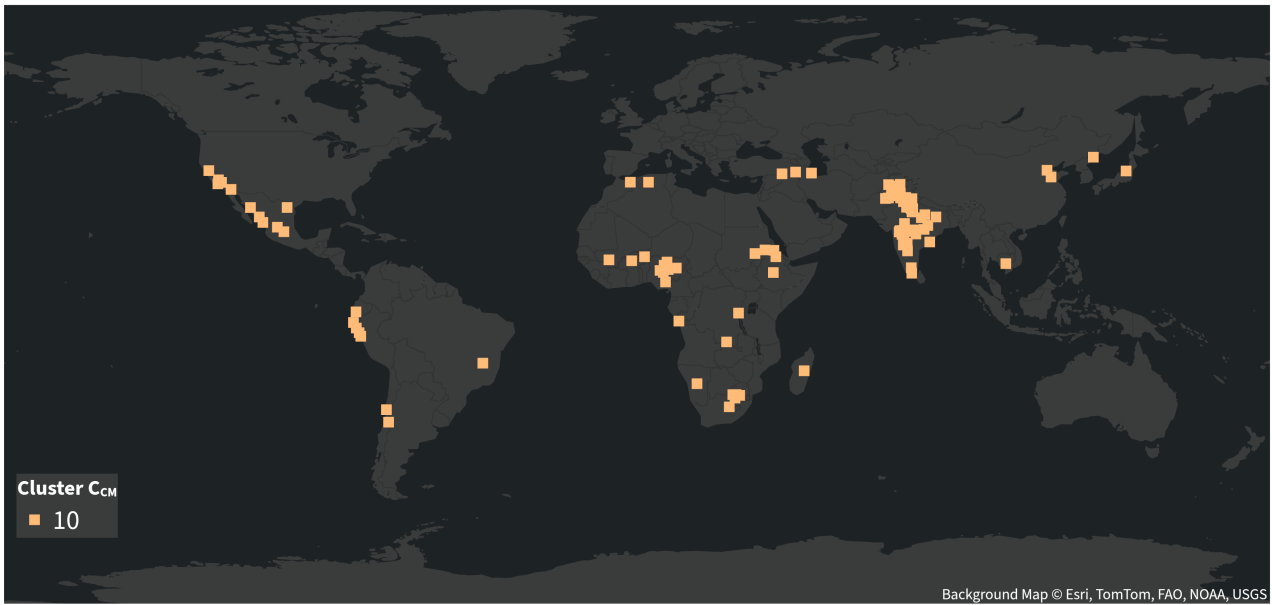
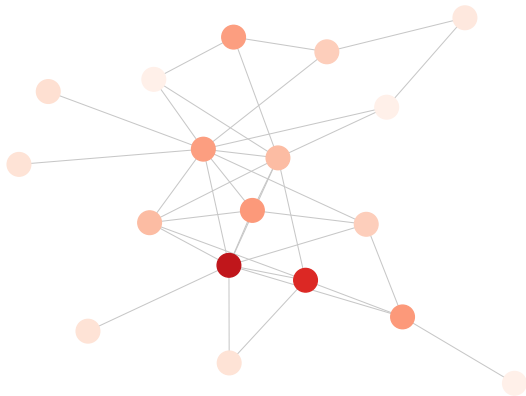
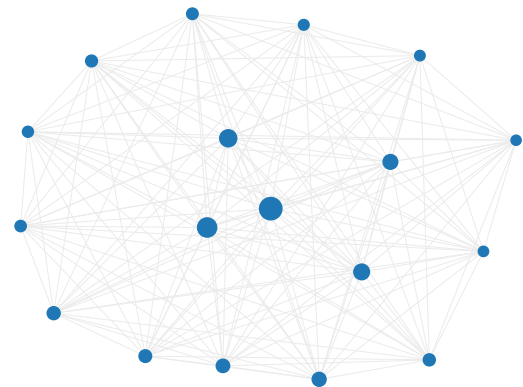


(c) Barycenter of  $c_{CM_9}$  based on area and distance

**Figure 5.13** Overview of all cities within  $c_{CM_9}$  and two different barycenters of that cluster.

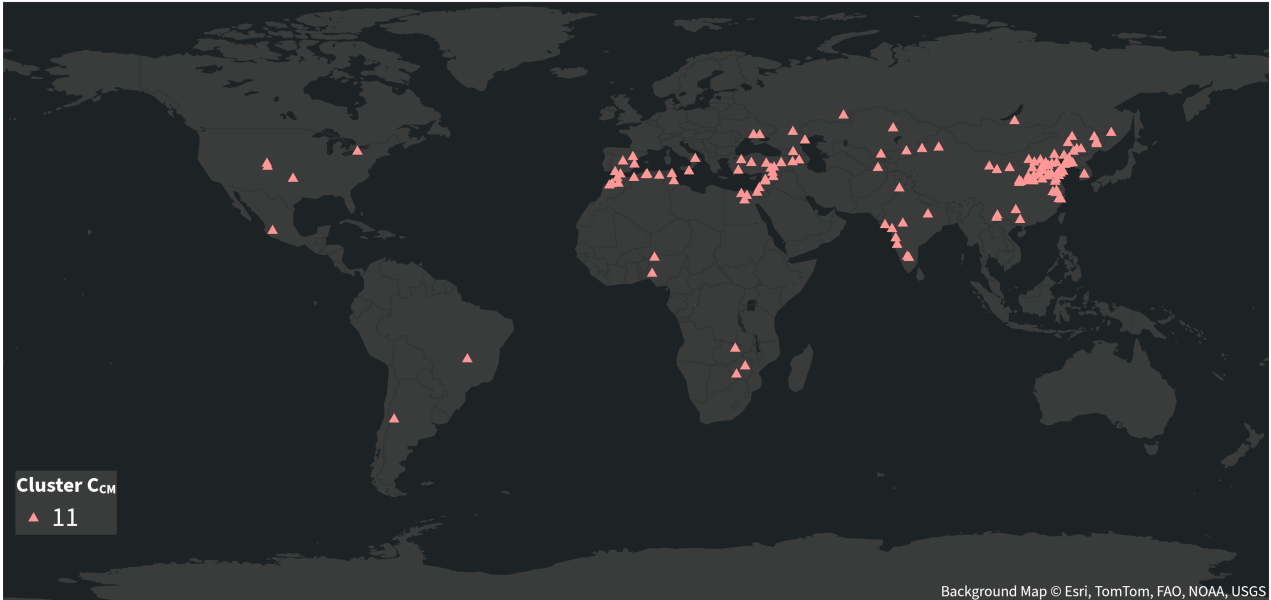
**Cluster  $c_{CM_{10}}$** 

Cluster  $c_{CM_{10}}$  with cities mostly in India, Middle and Southern Africa, and Eastern America. Additional cities are in North Africa, Middle East and East Asia. A total of 98 cities are assigned to this cluster. The left barycenter is showing a central structure with multiple extrusions and only two nodes with a lot of built-up area. The right barycenter shows three central nodes, with two additional nodes on the inside of a surrounding ring. Two cities and their graphs assigned to this cluster are shown in Figure B.73.

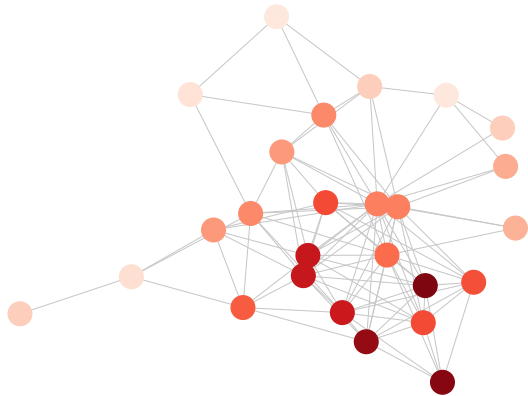
(a) World map showing all cities in  $c_{CM_{10}}$ (b) Barycenter of  $c_{CM_{10}}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_{10}}$  based on area and distance**Figure 5.14** Overview of all cities within  $c_{CM_{10}}$  and two different barycenters of that cluster.

**Cluster  $c_{CM_{11}}$**

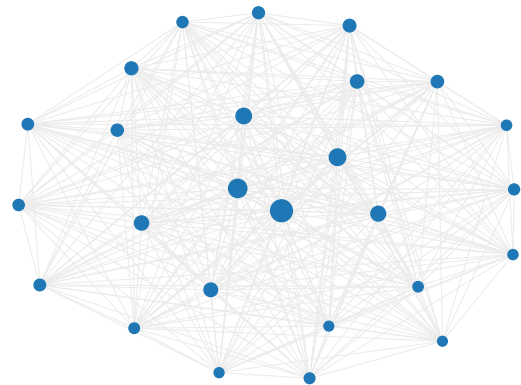
Cluster  $c_{CM_{11}}$  with cities mostly in Northern China and around the Mediterranean Sea. Additional cities are in India and Central Asia, Africa, and America. A total of 130 cities are assigned to this cluster. The left barycenter is showing a concentric structure, with some extrusions and built-up area only in one direction. The right barycenter hat multiple nodes on the inside, surrounded by a ring. Two cities and their graphs assigned to this cluster are shown in Figure B.74.



(a) World map showing all cities in  $c_{CM_{11}}$



(b) Barycenter of  $c_{CM_{11}}$  based on urban patterns and structure

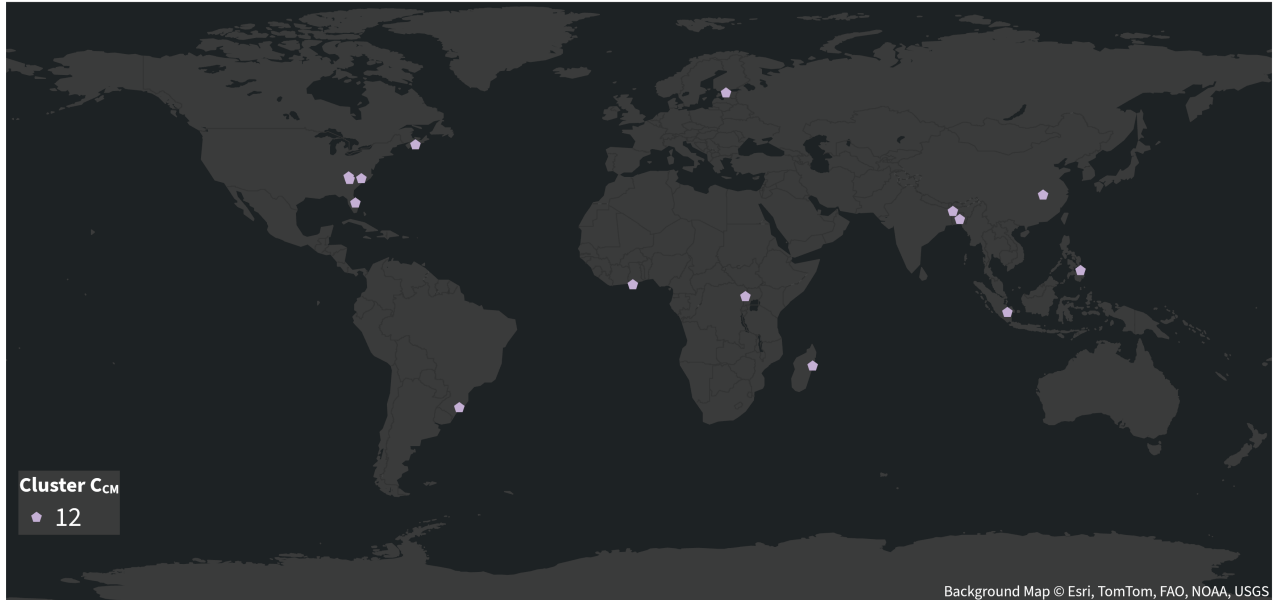


(c) Barycenter of  $c_{CM_{11}}$  based on area and distance

**Figure 5.15** Overview of all cities within  $c_{CM_{11}}$  and two different barycenters of that cluster.

**Cluster  $c_{CM_{12}}$** 

Cluster  $c_{CM_{12}}$  with cities not showing a geographical center, and the barycenters only showing two nodes. A total of 15 cities are assigned to this cluster. Two cities and their graphs assigned to this cluster are shown in Figure B.75.



(a) World map showing all cities in  $c_{CM_{12}}$



(b) Barycenter of  $c_{CM_{12}}$  based on urban patterns and structure

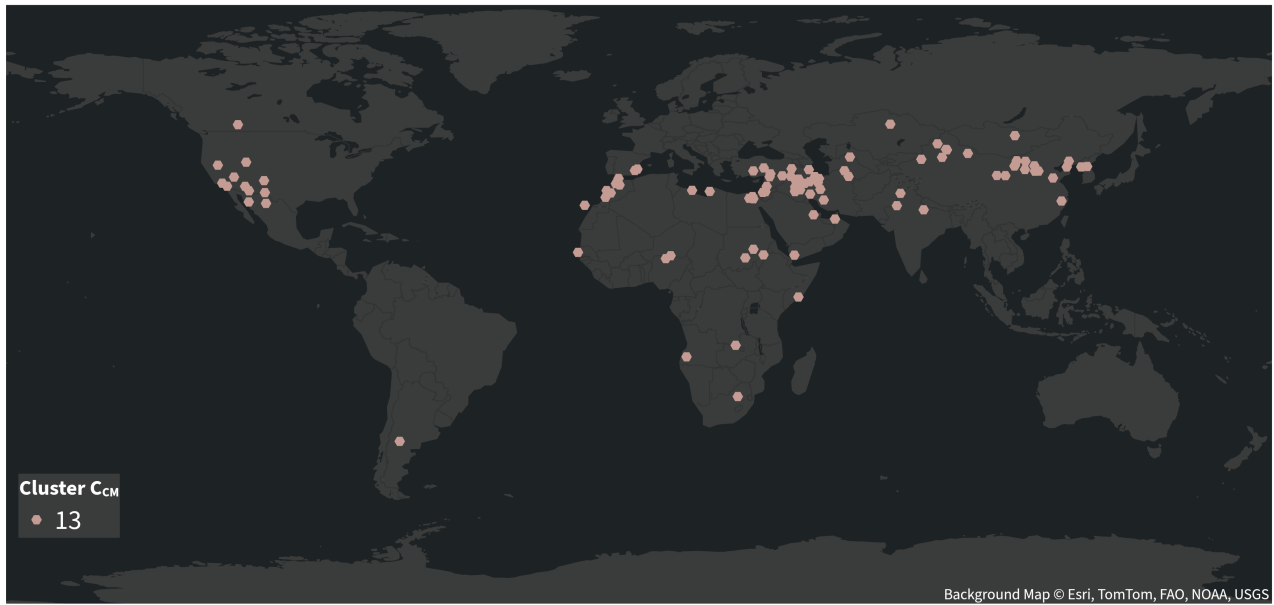


(c) Barycenter of  $c_{CM_{12}}$  based on area and distance

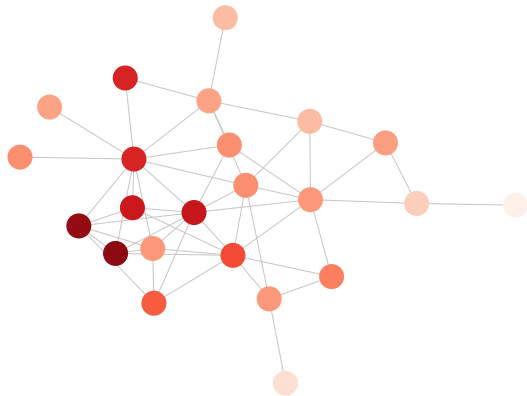
**Figure 5.16** Overview of all cities within  $c_{CM_{12}}$  and two different barycenters of that cluster.

**Cluster  $c_{CM_{13}}$** 

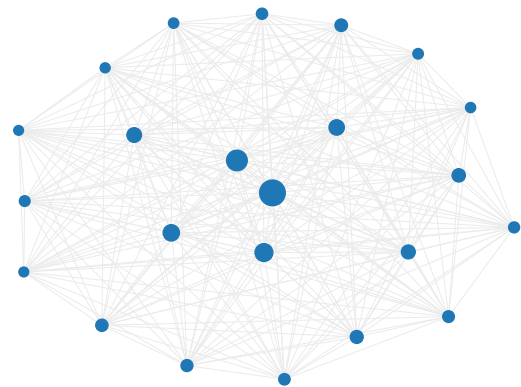
Cluster  $c_{CM_{13}}$  with cities mostly in the Middle East, Northern Africa, Western USA, and Northern China. Additional cities are in Canada and Argentina, Middle and Southern Africa, and Central Asia. A total of 96 cities are assigned to this cluster. The left barycenter is showing mostly closely connected nodes, with some branching off and having less built-up area. The right barycenter is showing two central nodes, with four additional nodes on the inside, surrounded by a ring. Two cities and their graphs assigned to this cluster are shown in Figure B.76.



(a) World map showing all cities in  $c_{CM_{13}}$



(b) Barycenter of  $c_{CM_{13}}$  based on urban patterns and structure

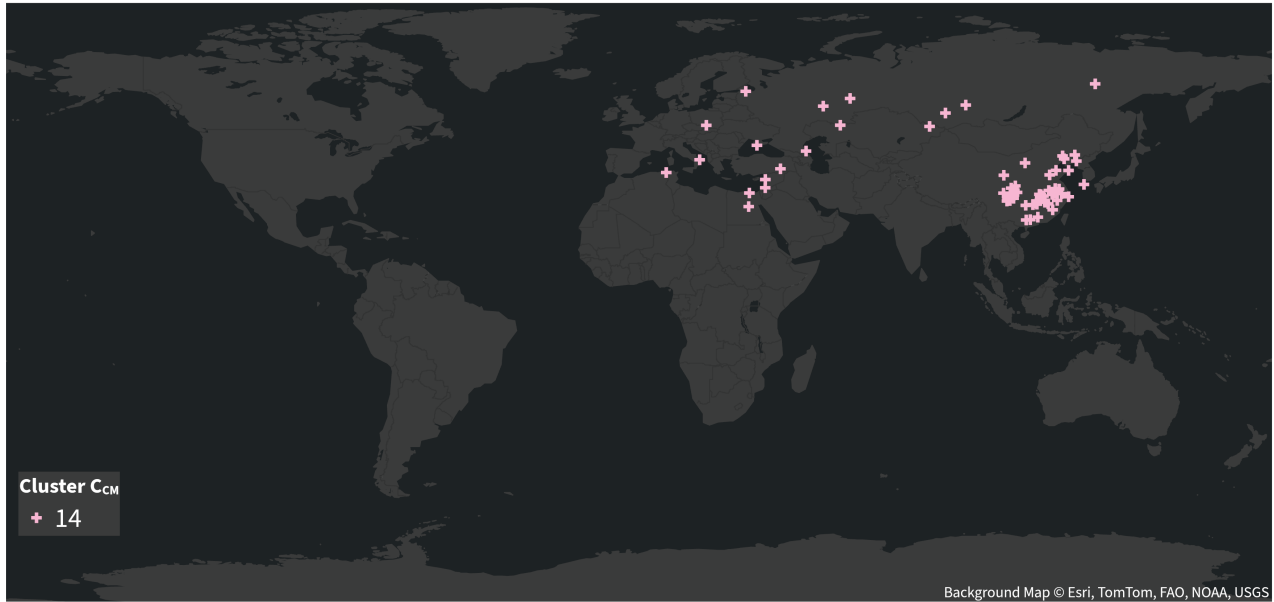
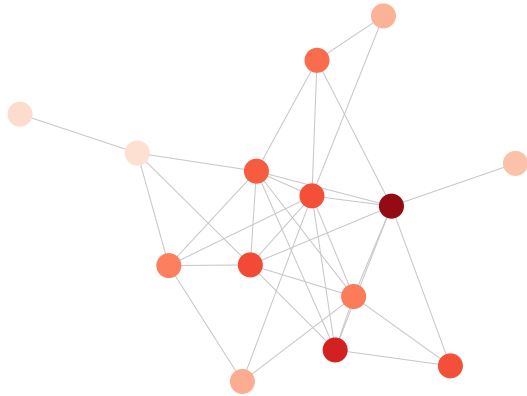
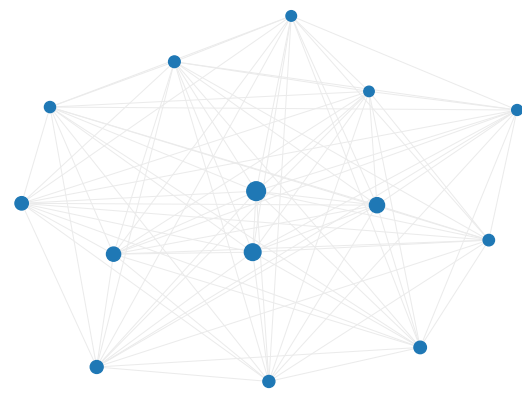


(c) Barycenter of  $c_{CM_{13}}$  based on area and distance

**Figure 5.17** Overview of all cities within  $c_{CM_{13}}$  and two different barycenters of that cluster.

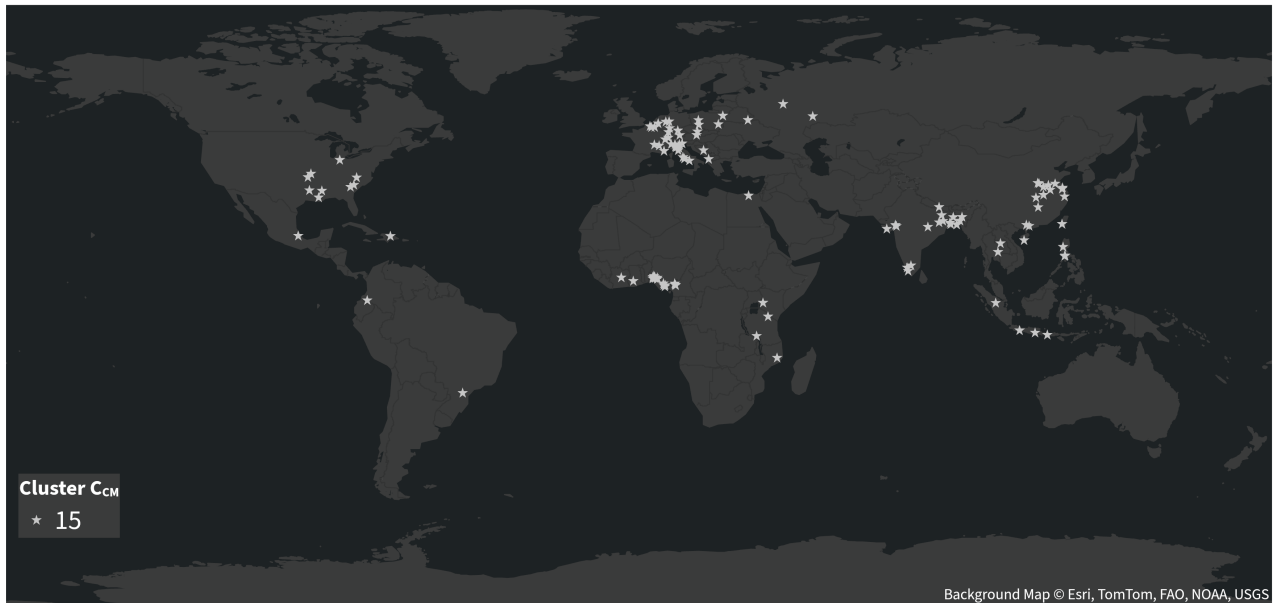
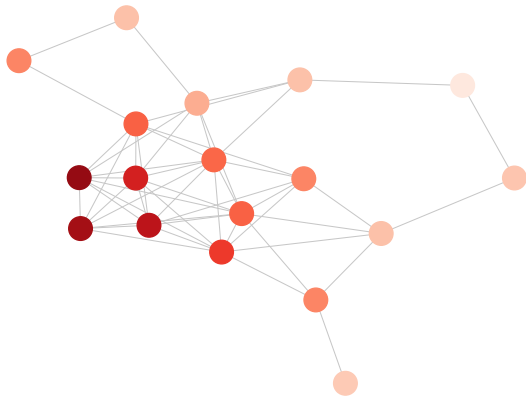
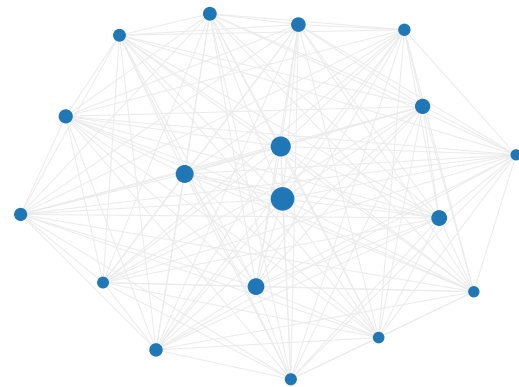
**Cluster  $c_{CM_{14}}$** 

Cluster  $c_{CM_{14}}$  with cities mostly in Eastern China, Russia, and Middle East. Additional cities are in Europe. A total of 65 cities are assigned to this cluster. The left barycenter is showing a very loose structure, the right one two central nodes. Two cities and their graphs assigned to this cluster are shown in Figure B.77.

(a) World map showing all cities in  $c_{CM_{14}}$ (b) Barycenter of  $c_{CM_{14}}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_{14}}$  based on area and distance**Figure 5.18** Overview of all cities within  $c_{CM_{14}}$  and two different barycenters of that cluster.

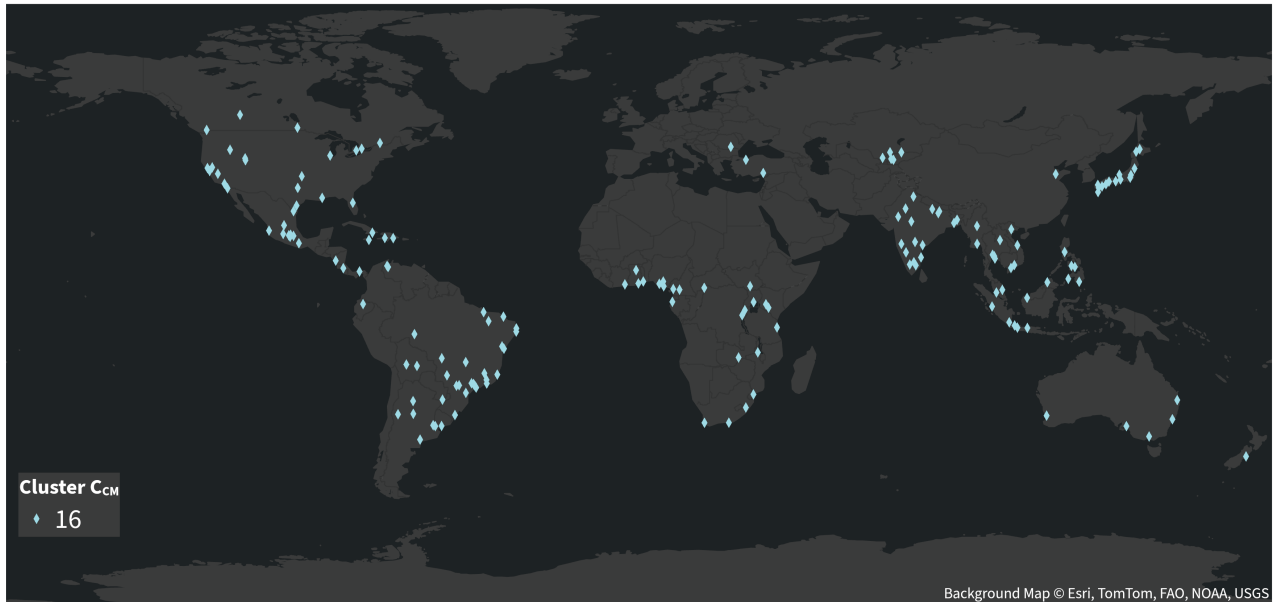
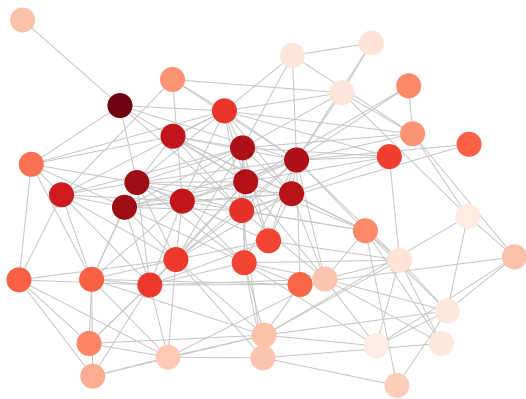
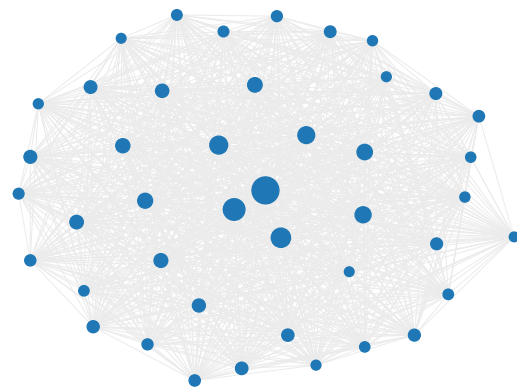
**Cluster  $c_{CM_{15}}$** 

Cluster  $c_{CM_{15}}$  with cities mostly in Middle and Eastern Europe, Eastern USA, Middle Africa, India, Eastern China, and South East Asia. A total of 113 cities are assigned to this cluster. The left barycenter is showing multiple closely connected nodes with a lot of built-up area and a pathlike structure. The right barycenter shows three central nodes, one additional node on the inside and a ring surrounding them. Two cities and their graphs assigned to this cluster are shown in Figure B.78.

(a) World map showing all cities in  $c_{CM_{15}}$ (b) Barycenter of  $c_{CM_{15}}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_{15}}$  based on area and distance**Figure 5.19** Overview of all cities within  $c_{CM_{15}}$  and two different barycenters of that cluster.

**Cluster  $c_{CM_{16}}$** 

Cluster  $c_{CM_{16}}$  with cities with no clear geographical center, but no city in Europe (including Russia) and only one in China. A total of 168 cities are assigned to this cluster. The left barycenter is showing a lot of closely connected nodes, with nodes having less built-up area on the outside. The right barycenter is showing multiple nodes on the inside with a surrounding ring. Two cities and their graphs assigned to this cluster are shown in Figure B.79.

(a) World map showing all cities in  $c_{CM_{16}}$ (b) Barycenter of  $c_{CM_{16}}$  based on urban patterns and structure(c) Barycenter of  $c_{CM_{16}}$  based on area and distance**Figure 5.20** Overview of all cities within  $c_{CM_{16}}$  and two different barycenters of that cluster.



# 6 Discussion

---

This chapter provides a discussion of both parts of this thesis, as well as an answer to the research questions formulated in Section 1.2, with Section 6.1 covering the first part and Section 6.2 the second part.

## 6.1 Unsupervised Deep Learning Approach to find Urban Patterns

The first part of this thesis, which was based on the work of Gassilloud (2022), aimed to find urban patterns through the means of an unsupervised framework using a parameter based encoding. To achieve this, patches containing the local climate zone (LCZ) land cover classification were extracted from cities delineated by the morphological urban areas (MUAs). Each patch was pixel wise encoded, using properties for density and height defined by the LCZs. Afterwards the patches were clustered using the SCAN and RUC deep learning frameworks, which resulted in the assignment of each patch to one of 66 different clusters, each representing one urban pattern. Additionally, a mosaic for each urban area, based on the urban patterns, was created. Those steps will now be discussed in detail, with the research question answered at the end.

### Data

The data basis was composed of two different datasets, raster tiles containing the LCZ mapping and MUA shapefiles delineating the urban areas. Those two datasets did not always align, which sometimes led to a low covering of the MUA by the LCZs, with the covering in some extreme cases as low as 20% (Table B.1). On the other side, the raster tiles often covered more area than the MUAs (with Figures 3.2 and 4.2 as examples). This leads to two different conclusions, who should be addressed in future studies:

1. The necessity of better aligned datasets, providing a better coverage of the MUAs by the LCZs.
2. The possibility of inclusion of LCZ data outside the MUAs.

The first conclusion is self-evident. A better coverage of the urban areas does not only improve results by having a more complete description of the data inside the analyzed areas, it also allows an improvement of the dataset used in the second part, which will be discussed below.

This aspect was also the reason behind the one change made to the data basis in comparison to the work of Gassilloud (2022): Every patch inside the MUAs was used, instead of excluding patches containing more than 85% water, thus improving the second dataset.

The second conclusion is based on the possible assumption that the morphology inside an urban area is also influenced by the area surrounding it. Although an experiment including the complete

LCZ dataset was performed (Table A.7), it did not yield expected results and led to a too high generalization of urban patterns. Thus, a compromise between using only LCZ data inside the MUAs and using the complete LCZ dataset should be made in the future and the results compared to the one obtained in this thesis.

## Encoding

The parameter based encoding used in this thesis was the biggest change to the work of Gassilloud (2022), with the aim to train the networks not only on the “existence” of LCZ classes, but on similarities and differences between them.

The encoding was based on two different properties defined by the LCZs (Appendix A.1), namely *surface fraction* and *height of roughness elements*, which resulted in a seven dimensional encoding (as described in Section 4.1.1). Those properties are serving as “proxies” for density and height.

While additional properties exist, and one could make an argument to use them as well, this encoding covers five of the six elements (except time) of the urban form as defined by Wentz et al. (2018) (Section 2.1). The materials are described by the separation into 7 different layers (layer 1 and 2 cover human constructions, layer 3—6 cover soil-plant continuum, and layer 7 covers water), dimensionality is covered by *height of roughness elements*, and spatial patterns by *surface fraction*.

Therefore, it can be assumed that the chosen encoding is a good representation of the urban form and that the inclusion of more parameters would not improve the results.

## Methodology

The clustering process was based on the SCAN and RUC frameworks. While those frameworks were originally developed for image clustering, Gassilloud (2022) adapted them for clustering of categorical data, like the LCZs. Additional changes to his work beyond the encoding were made by using a bigger batch size as well as a deeper backbone. This was made possible due to the use of the terrabyte platform, as stated in Section 3.1.

The original SimCLR paper shows a clear correlation between bigger batch sizes and achieved accuracy. The deeper backbone was chosen due to the high dimensionality of the data, and the large number of samples. Based on that the assumption is made that those changes will yield a better final result, even though there is no actual way to compare the accuracy of the different results as discussed below.

No change was made to the number of output dimensions. It was kept at 1024 after trying different values, powers of two in the range 128 to 2048 (Table A.7). Under the assumption that “a small number of classes might generalize [the patterns] too much” (Gassilloud, 2022) and with the knowledge that the patches are assigned only to a small subset of clusters after the SCAN framework, the highest value still yielding a result was chosen, which was 1024, as the use of 2048 output dimensions led to multiple empty clusters. Values between 1024 and 2048 were not tested due to time constraints and should be analyzed in future studies.

The selection process for the prototypes during the self-labeling step of the SCAN framework was not changed either. This part was one of the biggest modifications made by Gassilloud (2022) to the original framework by using a flexible threshold and the  $\max_3$  confidence as described in Section 4.1.2. However, this confidence metric could result in the selection of too many prototypes, which would skew the results, and thus should be analyzed in future studies.

The final part of the methodology, the mosaicking, was newly developed in this thesis, with the aim to provide an overview of the distribution of urban patterns across the urban area, as well as to provide a data basis for the second part. The relabeling during this part was based only on the relative amount of built-up area of each pattern, and thus the labels are not usable to describe a distance between the patterns and should be improved in future studies.

## Results

*Is it possible to find urban patterns through the use of an unsupervised framework using a parameter based encoding?* That was the research question defined for this part, which resulted in 66 different urban patterns. As already stated, this thesis has no way to measure any accuracy of those found patterns, but visual analysis shows similarity between patches in one cluster and dissimilarity between patches across different clusters, even though the median confidence is quite low with 61%.

Based on that, and the knowledge that the framework is usable to discover urban patterns in general, it is a fair assumption to say yes, it is possible to find urban patterns through the use of an unsupervised framework using a parameter based encoding.

But what do these results exactly mean, especially in comparison to the 138 patterns obtained by Gassilloud (2022), and how do they help to tackle the challenges of urbanization? Unfortunately, those questions are not easy to answer. One clear difference to the work of Gassilloud (2022) is the significantly lower amount of patterns. Additionally, it can be assumed, due to the parameter based encoding, that the patterns discovered in this work are a better representation of the actual morphological configurations inside the urban areas. But beyond that, it is difficult to draw a clear conclusion, especially as the patterns are only defined by their morphological structure, which makes it difficult to assign a clear semantic label to them.

But that, and the large size of each pattern (10.24 km<sup>2</sup>), also makes it possible to use the patterns in a wide variety in combination with applications covering large areas, like the study of urban climate, which is often analyzed on a similar scale (Lemoine-Rodríguez, Inostroza, and Zepp, 2022a).

Beyond that, the relation between the patterns can be analyzed, with an approach to this discussed in the next section.

## 6.2 Using Graph Representations to find Different Models of Cities

The second part of this thesis aimed to find different city models using a graph representation of each city describing the relationship between the previous discovered urban patterns. To achieve this, a k-Means clustering was applied on the graphs, created from the mosaics, using the optimal transport based Fused Gromov-Wasserstein (FGW) distance. Afterwards, for each cluster, two different barycenters were computed serving as city models. Those steps, as well as the research question, will now be discussed in detail.

### Data

The data basis consisted of multiple graphs, each corresponding to one urban area, and with each node of a graph representing one *region*, an aggregation of one urban pattern as described by the mosaics computed during the previous step. Edges were drawn between regions who shared at least one pixel.

The graphs had occasionally multiple connected components, due to the unaligned datasets used in the first part. If that happened, the biggest component was chosen to represent the urban area and the other components were discarded. As discussed above, this problem could be improved by using a data basis with a higher coverage of the MUAs by the LCZs. In other cases this happens due to the shape of the MUAs, which means there is no other simple solution except the one chosen.

An additional fully connected graph with multiple edge and node values was created. The node values described in this case geographic and morphological parameters, such as the area or the fractal dimension of the corresponding region. The edge values provided more information about the layout of the regions, such as the distance or angle between them. This graph was not used during the clustering, but for the creation of an additional barycenter.

### Methodology

The clustering was performed using a k-Means algorithm based on the FGW metric, which is able to compute a distance between graphs based on both the node features (the class label) and the graph structure (adjacency matrix).

As stated above, it is not possible to describe a distance between the class labels, as they only describe the relative amount of built-up area of an urban pattern. Thus, the assumption is made that the mean logits of each label are a better node value in this case. The logits allow then to compute the distance between the labels as the euclidean distance between the logits.

While the FGW distance includes both node features and structure, there are two limitations to it, which should be addressed in future studies.

1. The node features are constricted to a single dimensional vector.

## 2. The structure is constricted to a single value.

The first limitation makes it impractical to use other nodes values in addition to the logits, as they are already defined as a 1024 dimensional vector. Thus, the influence of additional values would be negligible.

An alternative approach was tried (Table A.8), with a multidimensional node value consisting of the class label and principal components of the other parameters. But this approach had multiple disadvantages: the class label was used, which should be avoided as stated above, and the results differed from the expectations if compared to previous studies like the seven city types discovered by Taubenböck, Debray, et al. (2020). Thus, this approach is not usable in this thesis.

However, a third approach could exist, but was not tried in this thesis and should be addressed in the future: As stated in Section 4.2.2, in addition to the node features, the “relative importance” of each node is part of the FGW distance. This thesis assumed the same importance for each node, but it might be possible to describe the importance by using other node parameters.

The second limitation prevents the combination of different matrices, like the adjacency matrix and the distance. With this constriction, the adjacency matrix is chosen as structure matrix, as the usage of other matrices did again not seem to achieve good results (Table A.8).

The optimal amount of clusters was determined by a gap statistic, which was performed 60 times in the interval of 1—25. An optimum in this interval was only found during 47 attempts, with the median at 16. Thus, an argument could be made to increase the range of the interval in order to always find an optimum. But due to the position of the median in the interval, the assumption is made that the median is a good representation of the optimal value of clusters.

The final step of the methodology was the computation of two different barycenters for each cluster. One was based on the logits and the adjacency matrix with the aim to receive a pure structural representation of the urban morphology. The other one was based on the area of and distance between the regions, with the aim to obtain a model with geographic information. While a single barycenter containing all these information would be preferable, this approach was chosen due to the same restrictions as above.

## Results

*Is it possible to discover different models of cities by using graph representations of the previous identified patterns?* That was the research question defined for the second part, which resulted in 16 different city clusters, with two barycenters for each cluster serving as city models. Most of the clusters have a clear geographic distribution and unique barycenters.

This process was again unsupervised and without a way to measure the accuracy of the results. But based on a visual analysis, with comparison of the results to previous studies, the research question can be answered with yes, it is possible to find different models of cities by using graph representations of urban patterns.

But that again brings up the question, what is the exact meaning of the results and how do they compare to similar studies.

Most of the clusters (except  $c_{CM_8}$ ,  $c_{CM_{12}}$ , and  $c_{CM_{16}}$ ) have clear geographic centers (like China and Europe or the Middle East). Out of the other three, two clusters are composed of very small cities, with an average of two respective five regions. But even the last cluster has some specific features, like the almost complete absence of cities in Europe or China.

Other studies dealt with the unsupervised clustering of cities before, like Taubenböck, Debray, et al. (2020), who used morphological variances described by the LCZs, or Lemoine-Rodríguez, Inostroza, and Zepp (2020), who clustered cities based on principal components of landscape metrics. Both of those achieved results with the elements inside the different clusters sharing specific configurations and, in the case of Taubenböck, Debray, et al. (2020), clusters with clear geographic distributions. Thus, the assumption can be made, that the cities within each cluster obtained in this thesis share morphological characteristics, even within a cluster without a clear geographic center, like  $c_{CM_{16}}$ .

An advantage of the method used in this thesis are the barycenters, who represent the average features of each class and are able to serve as city models, similar to the models discussed in Section 2.1, although more complex. As stated above, while it would be preferable to have a single barycenter for each cluster, the two barycenter are still able to describe both morphological structure and geographic composition of each cluster. Thus, they are able to describe the general composition of cities, which is a crucial step for understanding cities and tackling the challenges of urbanization.

# 7 Conclusion

---

66 different urban patterns, 16 city types, and 32 city models — the results achieved in this thesis by extending the work of Gassilloud (2022) with a *parameter based encoding* to find patterns within urban areas. Afterwards a cluster analysis of graphs, *representing the relation of the urban patterns to each other* inside the urban areas, was performed with the aim to find different types of cities. Finally, for each found type, *two different city models* describing the average morphological and geographical structure of that type, were created.

The first part utilized the unsupervised SCAN and RUC deep learning frameworks, originally developed for image clustering, adapted by Gassilloud (2022) for the clustering of categorical data, like the LCZ land cover classification, and further extended in this work with a parameter based encoding. This encoding allowed the frameworks not only to learn the composition of the urban area by the individual land cover classes, but to recognize *similarities and differences of the urban form* between them. This process resulted in *66 different urban patterns*, with each pattern *describing one specific morphological configuration* of the urban area. While there are still some improvements to be made (mainly for the data basis), the discovered patterns are able to describe the intra urban morphology of cities around the globe at a medium scale. Furthermore, a *mosaic composed of the urban patterns* was created for each city, thus allowing to analyze the urban morphology, as described by the patterns, on a city scale.

With a clear description of the intra urban morphology of each city, the aim of the second part was to find morphological similarities between the urban areas. To achieve that, a graph representation of each city based on the mosaics was created, with the *nodes representing aggregations of urban patterns* in each urban area and the *edges showing the relation of the patterns to each other*. Utilizing the FGW metric, which is able to compute a distance between graphs based on both their nodes and edges, a k-Means clustering was performed. This process resulted in *16 different city types*, with the cities of each type having a *similar structure of urban patterns* and, in most cases, a *clear geographic distribution*. Afterwards, *two different city models*, once again represented using graph structures, were created for each city type. The first one describes the intra urban morphology represented by the *aggregation of and relationship between urban patterns*, the second one showcases the *relative geographic location of each aggregation*. While there are currently some limitations to this process (mainly the combination of morphological and geographical information), this part nonetheless provides a clear description of multiple city types around the globe, each sharing morphological configurations, as well as for the general composition of each city type.

In conclusion, although both parts of this thesis are highly explorative and currently without a way to verify or label the results, this thesis provides a *clear framework to support the understanding of the urban morphology* on multiple scales, and thus should be able to help tackle the challenges of urbanization.

# Bibliography

---

- Arthur, David and Sergei Vassilvitskii (2006). “How Slow Is the K-Means Method?” In: *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*. 22nd Annual Symposium on Computational Geometry. SCG '06. New York, NY, USA: Association for Computing Machinery, pp. 144–153. ISBN: 978-1-59593-340-9. DOI: 10.1145/1137856.1137880.
- Bakaeva, Natalia and Minh Tuan Le (2022). “Determination of Urban Pollution Islands by Using Remote Sensing Technology in Moscow, Russia.” In: *Ecological Informatics* 67, p. 101493. ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2021.101493.
- Barbierato, Elena et al. (2020). “Integrating Remote Sensing and Street View Images to Quantify Urban Forest Ecosystem Services.” In: *Remote Sensing* 12.2 (2), p. 329. ISSN: 2072-4292. DOI: 10.3390/rs12020329.
- Bechtel, Benjamin et al. (2015). “Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities.” In: *ISPRS International Journal of Geo-Information* 4.1 (1), pp. 199–219. ISSN: 2220-9964. DOI: 10.3390/ijgi4010199.
- Berthelot, David et al. (2019). “MixMatch: A Holistic Approach to Semi-Supervised Learning.” In: *Advances in Neural Information Processing Systems* 32. 33rd Conference on Neural Information Processing Systems. Vol. 32. Curran Associates, Inc. URL: <https://proceedings.neurips.cc/paper/2019/hash/1cd138d0499a68f4bb72bee04bbec2d7-Abstract.html> (visited on Apr. 22, 2024).
- Bhatta, Basudeb (2010). *Analysis of Urban Growth and Sprawl from Remote Sensing Data*. Advances in Geographic Information Science. Berlin, Heidelberg: Springer. ISBN: 978-3-642-05298-9. DOI: 10.1007/978-3-642-05299-6.
- Chazal, Frédéric and Bertrand Michel (2021). “An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists.” In: *Frontiers in Artificial Intelligence* 4. ISSN: 2624-8212. DOI: 10.3389/frai.2021.667963.
- Chen, Ting et al. (2020). “A Simple Framework for Contrastive Learning of Visual Representations.” In: *Proceedings of the 37th International Conference on Machine Learning*. The 37th International Conference on Machine Learning. Vol. 119. ICML'20. JMLR.org, pp. 1597–1607. arXiv: 2002.05709 [cs, stat].
- Chowdhury, Samir and Tom Needham (2021). “Generalized Spectral Clustering via Gromov-Wasserstein Learning.” In: *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*. International Conference on Artificial Intelligence and Statistics. PMLR, pp. 712–720. URL: <https://proceedings.mlr.press/v130/chowdhury21a.html> (visited on Apr. 18, 2024).
- Ciresan, Dan Claudiu et al. (2010). “Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition.” In: *Neural Computation* 22.12, pp. 3207–3220. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/NECO\_a\_00052. arXiv: 1003.0358 [cs].



- Crooks, Andrew T. et al. (2016). “User-Generated Big Data and Urban Morphology.” In: *Built Environment* 42.3, pp. 396–414. DOI: 10.2148/benv.42.3.396.
- Cubuk, Ekin D. et al. (2020). “Randaugment: Practical Automated Data Augmentation with a Reduced Search Space.” In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 3008–3017. DOI: 10.1109/CVPRW50498.2020.00359.
- Davenport, Alan G. et al. (2000). “Estimating the Roughness of Cities and Sheltered Country.” In: *12th Conference on Applied Climatology*, pp. 96–99.
- De Souza, Felipe Francisco (2018). “Concepts on Land Readjustment.” In: *Land Readjustment: Solving Urban Problems Through Innovative Approach*. Ed. by Felipe Francisco De Souza, Takeo Ochi, and Akio Hosono. Tokyo: JICA Research Institute, pp. 15–32. ISBN: 978-4-86357-078-8. URL: [https://www.jica.go.jp/english/jica\\_ri/publication/booksandreports/20180228\\_01.html](https://www.jica.go.jp/english/jica_ri/publication/booksandreports/20180228_01.html) (visited on Apr. 2, 2024).
- Dell’Acqua, Fabio and Paolo Gamba (2012). “Remote Sensing and Earthquake Damage Assessment: Experiences, Limits, and Perspectives.” In: *Proceedings of the IEEE* 100.10, pp. 2876–2890. ISSN: 1558-2256. DOI: 10.1109/JPROC.2012.2196404.
- Demuzere, Matthias et al. (2022). “A Global Map of Local Climate Zones to Support Earth System Modelling and Urban-Scale Environmental Science.” In: *Earth System Science Data* 14.8, pp. 3835–3873. ISSN: 1866-3508. DOI: 10.5194/essd-14-3835-2022.
- Flamary, Rémi and Alexandre Gramfort (2022). *Optimal Transport for 1D Distributions*. URL: [https://pythonot.github.io/auto\\_examples/plot\\_OT\\_1D.html](https://pythonot.github.io/auto_examples/plot_OT_1D.html) (visited on May 4, 2024).
- Gassilloud, Matthias (2022). “Exploring City Patterns Globally: The Intra-Urban Morphology through the Scope of Unsupervised Learning.” MA thesis. Heidelberg University. URL: <https://elib.dlr.de/189806/>.
- Glorot, Xavier and Yoshua Bengio (2010). “Understanding the Difficulty of Training Deep Feed-forward Neural Networks.” In: *Journal of Machine Learning Research - Proceedings Track 9*, pp. 249–256.
- Glorot, Xavier, Antoine Bordes, and Yoshua Bengio (2011). “Deep Sparse Rectifier Neural Networks.” In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, pp. 315–323. URL: <https://proceedings.mlr.press/v15/glorot11a.html> (visited on Apr. 14, 2024).
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press. ISBN: 978-0-262-03561-3.
- Google Machine Learning Glossary (n.d.[a]). *Batch Size*. URL: <https://developers.google.com/machine-learning/glossary#batch-size> (visited on Apr. 19, 2024).
- (n.d.[b]). *Logits*. URL: <https://developers.google.com/machine-learning/glossary/#logits> (visited on Apr. 14, 2024).
- Graupe, Daniel (2013). *Principles of Artificial Neural Networks*. 3rd ed. Vol. 7. Advanced Series in Circuits and Systems. WORLD SCIENTIFIC. ISBN: 978-981-4522-73-1. DOI: 10.1142/8868.

- Harris, Chauncy D. and Edward L. Ullman (1945). "The Nature of Cities." In: *The ANNALS of the American Academy of Political and Social Science* 242.1, pp. 7–17. ISSN: 0002-7162. DOI: 10.1177/000271624524200103.
- He, Kaiming et al. (2016). "Deep Residual Learning for Image Recognition." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- He, Miao, Yongming Xu, and Ning Li (2020). "Population Spatialization in Beijing City Based on Machine Learning and Multisource Remote Sensing Data." In: *Remote Sensing* 12.12 (12), p. 1910. ISSN: 2072-4292. DOI: 10.3390/rs12121910.
- Heineberg, Heinz et al. (2022). *Stadtgeographie*. 6th ed. Stuttgart, Deutschland: utb GmbH. ISBN: 978-3-8385-5856-1. DOI: 10.36198/9783838558561.
- Hoyt, Homer (1939). *The Structure and Growth of Residential Neighborhoods in American Cities*. U.S. Government Printing Office. Google Books: GXxAq1bof4kC.
- Huang, Fan et al. (2023). "Mapping Local Climate Zones for Cities: A Large Review." In: *Remote Sensing of Environment* 292, p. 113573. ISSN: 0034-4257. DOI: 10.1016/j.rse.2023.113573.
- Huynh, Hoai Nguyen (2019). "Spatial Point Pattern and Urban Morphology: Perspectives from Entropy, Complexity and Networks." In: *Physical Review E* 100.2, p. 022320. ISSN: 2470-0045, 2470-0053. DOI: 10.1103/PhysRevE.100.022320. arXiv: 1904.09787 [cond-mat, physics:physics].
- Ioffe, Sergey and Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." In: *Proceedings of the 32nd International Conference on Machine Learning*. The 32nd International Conference on Machine Learning. ICML'15. Lille, France: JMLR.org, pp. 448–456. arXiv: 1502.03167 [cs].
- Jiménez-Jiménez, Sergio Iván et al. (2020). "Rapid Urban Flood Damage Assessment Using High Resolution Remote Sensing Data and an Object-Based Approach." In: *Geomatics, Natural Hazards and Risk* 11.1, pp. 906–927. ISSN: 1947-5705. DOI: 10.1080/19475705.2020.1760360.
- Kivell, Philip (1992). *Land and the City: Patterns and Processes of Urban Change*. London: Routledge. ISBN: 978-0-203-42023-2. DOI: 10.4324/9780203420232.
- Kolouri, Soheil et al. (2020). "Wasserstein Embedding for Graph Learning." In: *International Conference on Learning Representations 2021*. arXiv: 2006.09430 [cs, stat]. URL: [https://openreview.net/forum?id=AAes\\_3W-2z](https://openreview.net/forum?id=AAes_3W-2z) (visited on Apr. 17, 2024).
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2017). "ImageNet Classification with Deep Convolutional Neural Networks." In: *Communications of the ACM* 60.6, pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386.
- Kropf, Karl (2017). *The Handbook Of Urban Morphology*. 1st ed. Wiley. ISBN: 978-1-118-74769-8. DOI: 10.1002/9781118747711.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep Learning." In: *Nature* 521.7553, pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539.

- LeCun, Yann, Leon Bottou, Yoshua Bengio, et al. (1998). “Gradient-Based Learning Applied to Document Recognition.” In: *Proceedings of the IEEE* 86.11, pp. 2278–2324. ISSN: 1558-2256. DOI: 10.1109/5.726791.
- LeCun, Yann, Leon Bottou, Genevieve B. Orr, et al. (1998). “Efficient BackProp.” In: *Neural Networks: Tricks of the Trade*. Ed. by Genevieve B. Orr and Klaus-Robert Müller. Red. by Gerhard Goos, Juris Hartmanis, and Jan Van Leeuwen. Vol. 1524. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 9–50. ISBN: 978-3-540-65311-0. DOI: 10.1007/3-540-49430-8\_2.
- Lemoine-Rodríguez, Richard, Luis Inostroza, and Harald Zepp (2020). “The Global Homogenization of Urban Form. An Assessment of 194 Cities across Time.” In: *Landscape and Urban Planning* 204, p. 103949. ISSN: 0169-2046. DOI: 10.1016/j.landurbplan.2020.103949.
- (2022a). “Does Urban Climate Follow Urban Form? Analysing Intraurban LST Trajectories versus Urban Form Trends in 3 Cities with Different Background Climates.” In: *Science of The Total Environment* 830, p. 154570. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2022.154570.
- (2022b). “Intraurban Heterogeneity of Space-Time Land Surface Temperature Trends in Six Climate-Diverse Cities.” In: *Science of The Total Environment* 804, p. 150037. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2021.150037.
- Li, Junnan, Richard Socher, and Steven C. H. Hoi (2019). “DivideMix: Learning with Noisy Labels as Semi-supervised Learning.” In: International Conference on Learning Representations 2020. URL: <https://openreview.net/forum?id=HJgExaVtwr> (visited on Apr. 22, 2024).
- Marconcini, Mattia et al. (2020). “Outlining Where Humans Live, the World Settlement Footprint 2015.” In: *Scientific Data* 7.1, p. 242. ISSN: 2052-4463. DOI: 10.1038/s41597-020-00580-5.
- McCulloch, Warren S. and Walter Pitts (1943). “A Logical Calculus of the Ideas Immanent in Nervous Activity.” In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259.
- Mémoli, Facundo (2011). “Gromov–Wasserstein Distances and the Metric Approach to Object Matching.” In: *Foundations of Computational Mathematics* 11.4, pp. 417–487. ISSN: 1615-3383. DOI: 10.1007/s10208-011-9093-5.
- (2014). “The Gromov–Wasserstein Distance: A Brief Overview.” In: *Axioms* 3.3 (3), pp. 335–341. ISSN: 2075-1680. DOI: 10.3390/axioms3030335.
- Merriam-Webster.com (2024). *Pattern*. URL: <https://www.merriam-webster.com/dictionary/pattern> (visited on Apr. 10, 2024).
- NOAA (2024). *What Is Remote Sensing?* URL: <https://oceanservice.noaa.gov/facts/remotesensing.html> (visited on Apr. 7, 2024).
- OECD (2015). *The Metropolitan Century: Understanding Urbanisation and Its Consequences*. Paris: Organisation for Economic Co-operation and Development. DOI: 10.1787/9789264228733-en.
- Oliveira, Vitor (2016). *Urban Morphology*. The Urban Book Series. Cham: Springer International Publishing. ISBN: 978-3-319-32081-6. DOI: 10.1007/978-3-319-32083-0.
- Park, Robert E. and Ernest W. Burgess ([1925] 2019). *The City*. Chicago: University of Chicago Press. ISBN: 978-0-226-63650-4. DOI: 10.7208/chicago/9780226636641.001.0001.

- Park, Sungwon et al. (2021). “Improving Unsupervised Image Clustering With Robust Learning.” In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12273–12282. DOI: 10.1109/CVPR46437.2021.01210.
- Petric Maretic, Hermina et al. (2019). “GOT: An Optimal Transport Framework for Graph Comparison.” In: *Advances in Neural Information Processing Systems 32*. 33rd Conference on Neural Information Processing Systems. Vol. 32. Curran Associates, Inc. arXiv: 1906.02085 [cs, stat]. URL: <https://papers.nips.cc/paper/2019/hash/fdd5b16fc8134339089ef25b3cf0e588-Abstract.html> (visited on Apr. 17, 2024).
- Peyré, Gabriel and Marco Cuturi (2019). “Computational Optimal Transport: With Applications to Data Science.” In: *Foundations and Trends® in Machine Learning* 11.5-6, pp. 355–607. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/22000000073. arXiv: 1803.00567 [stat].
- Phiri, Darius and Justin Morgenroth (2017). “Developments in Landsat Land Cover Classification Methods: A Review.” In: *Remote Sensing* 9.9 (9), p. 967. ISSN: 2072-4292. DOI: 10.3390/rs9090967.
- Qiu, Chunping, Michael Schmitt, and Xiao Xiang Zhu (2019). “Fusing Multi-Seasonal Sentinel-2 Images with Residual Convolutional Neural Networks for Local Climate Zone-Derived Urban Land Cover Classification.” In: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 5037–5040. DOI: 10.1109/IGARSS.2019.8898223.
- Quinn, James A. (1940). “The Burgess Zonal Hypothesis and Its Critics.” In: *American Sociological Review* 5.2, pp. 210–218. ISSN: 0003-1224. DOI: 10.2307/2083636. JSTOR: 2083636.
- Reis, José P., Elisabete A. Silva, and Paulo Pinho (2016). “Spatial Metrics to Study Urban Patterns in Growing and Shrinking Cities.” In: *Urban Geography* 37.2, pp. 246–271. ISSN: 0272-3638. DOI: 10.1080/02723638.2015.1096118.
- Rosenblatt, Frank (1958). “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” In: *Psychological Review* 65.6, pp. 386–408. ISSN: 1939-1471. DOI: 10.1037/h0042519.
- Rubenstein, James M. (2020). *The Cultural Landscape: An Introduction to Human Geography*. Thirteenth edition. Hoboken, NJ: Pearson. ISBN: 978-0-13-511615-9.
- Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (1986). “Learning Representations by Back-Propagating Errors.” In: *Nature* 323.6088, pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0.
- Sachs, Jeffrey D. et al. (2019). “Six Transformations to Achieve the Sustainable Development Goals.” In: *Nature Sustainability* 2.9, pp. 805–814. ISSN: 2398-9629. DOI: 10.1038/s41893-019-0352-9.
- Santambrogio, Filippo (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Vol. 87. Progress in Nonlinear Differential Equations and Their Applications. Cham: Springer International Publishing. ISBN: 978-3-319-20827-5. DOI: 10.1007/978-3-319-20828-2.

- Schmitt, Andreas, Tobias Sieg, et al. (2018). "Investigation on the Separability of Slums by Multi-Aspect TerraSAR-X Dual-Co-Polarized High Resolution Spotlight Images Based on the Multi-Scale Evaluation of Local Distributions." In: *International Journal of Applied Earth Observation and Geoinformation* 64, pp. 181–198. ISSN: 1569-8432. DOI: 10.1016/j.jag.2017.09.006.
- Schmitt, Andreas, Anna Wendleder, et al. (2020). "Multi-Source and Multi-Temporal Image Fusion on Hypercomplex Bases." In: *Remote Sensing* 12.6 (6), p. 943. ISSN: 2072-4292. DOI: 10.3390/rs12060943.
- Schmitt, Michael and Xiao Xiang Zhu (2016). "Data Fusion and Remote Sensing: An Ever-Growing Relationship." In: *IEEE Geoscience and Remote Sensing Magazine* 4.4, pp. 6–23. ISSN: 2168-6831. DOI: 10.1109/MGRS.2016.2561021.
- Stewart, Iain D. and Tim R. Oke (2012). "Local Climate Zones for Urban Temperature Studies." In: *Bulletin of the American Meteorological Society* 93.12, pp. 1879–1900. DOI: 10.1175/BAMS-D-11-00019.1.
- Taubenböck, Hannes, Henri Debray, et al. (2020). "Seven City Types Representing Morphologic Configurations of Cities across the Globe." In: *Cities* 105, p. 102814. ISSN: 0264-2751. DOI: 10.1016/j.cities.2020.102814.
- Taubenböck, Hannes, Matthias Weigand, et al. (2019). "A New Ranking of the World's Largest Cities—Do Administrative Units Obscure Morphological Realities?" In: *Remote Sensing of Environment* 232, p. 111353. ISSN: 0034-4257. DOI: 10.1016/j.rse.2019.111353.
- Tibshirani, Robert, Guenther Walther, and Trevor Hastie (2001). "Estimating the Number of Clusters in a Data Set Via the Gap Statistic." In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 63.2, pp. 411–423. ISSN: 1369-7412. DOI: 10.1111/1467-9868.00293.
- Tsilimigkas, Georgios, Demetris Stathakis, and Maria Pafi (2016). "Evaluating the Land Use Patterns of Medium-Sized Hellenic Cities." In: *Urban Research & Practice* 9.2, pp. 181–203. ISSN: 1753-5069. DOI: 10.1080/17535069.2015.1125940.
- UN DESA (2019). *World Urbanization Prospects: The 2018 Revision*. United Nations. ISBN: 978-92-1-004314-4. DOI: 10.18356/b9e995fe-en.
- Van Gansbeke, Wouter et al. (2020). "SCAN: Learning to Classify Images Without Labels." In: *Computer Vision – ECCV 2020*. 16th European Conference on Computer Vision. Ed. by Andrea Vedaldi et al. Cham: Springer International Publishing, pp. 268–285. ISBN: 978-3-030-58607-2. DOI: 10.1007/978-3-030-58607-2\_16. arXiv: 2005.12320 [cs].
- Vayer, Titouan et al. (2019). "Optimal Transport for Structured Data with Application on Graphs." In: *Proceedings of the 36th International Conference on Machine Learning*. The 36th International Conference on Machine Learning. PMLR, pp. 6275–6284. DOI: 10.48550/arXiv.1805.09114.
- Villani, Cédric (2009). *Optimal Transport*. Red. by Marcel Berger et al. Vol. 338. Grundlehren Der Mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-540-71049-3. DOI: 10.1007/978-3-540-71050-9.

- Wentz, Elizabeth A. et al. (2018). “Six Fundamental Aspects for Conceptualizing Multidimensional Urban Form: A Spatial Mapping Perspective.” In: *Landscape and Urban Planning* 179, pp. 55–62. ISSN: 0169-2046. DOI: 10.1016/j.landurbplan.2018.07.007.
- Werbos, Paul J. (1982). “Applications of Advances in Nonlinear Sensitivity Analysis.” In: *System Modeling and Optimization*. 10th IFIP Conference on System Modeling and Optimization. Ed. by Rudolf F. Drenick and Frank Kozin. Berlin, Heidelberg: Springer, pp. 762–770. ISBN: 978-3-540-39459-4. DOI: 10.1007/BFb0006203.
- World Bank (2016). *World Development Report 2016: Digital Dividends*. World Development Report. The World Bank. ISBN: 978-1-4648-0671-1. DOI: 10.1596/978-1-4648-0671-1.
- Wurm, Michael and Hannes Taubenböck (2018). “Detecting Social Groups from Space – Assessment of Remote Sensing-Based Mapped Morphological Slums Using Income Data.” In: *Remote Sensing Letters*. ISSN: 2150-704X. URL: <https://www.tandfonline.com/doi/abs/10.1080/2150704X.2017.1384586> (visited on Apr. 25, 2024).
- Xu, Hongteng et al. (2019). “Gromov-Wasserstein Learning for Graph Matching and Node Embedding.” In: *Proceedings of the 36th International Conference on Machine Learning*. The 36th International Conference on Machine Learning. PMLR, pp. 6932–6941. URL: <https://proceedings.mlr.press/v97/xu19b.html> (visited on Apr. 18, 2024).
- Yang, Xiaojun (2021). “Progress in Urban Remote Sensing.” In: *Urban Remote Sensing*. Ed. by Xiaojun Yang. John Wiley & Sons, Ltd, pp. 3–13. ISBN: 978-1-119-62586-5. DOI: 10.1002/9781119625865.ch1.
- Živković, Jelena (2019). “Urban Form and Function.” In: *Climate Action*. Ed. by Walter Leal Filho et al. Cham: Springer International Publishing, pp. 1–10. ISBN: 978-3-319-71063-1. DOI: 10.1007/978-3-319-71063-1\_78-1.

# A Additional Information

## A.1 Local Climate Zone Properties

**Table A.1** Local climate zone (LCZ) properties defined for each class by Stewart and Oke (2012). All values are unitless except height of roughness elements [m].

LCZ	Sky view factor	Aspect ratio	Building surface fraction	Impervious surface fraction	Pervious surface fraction	Height of roughness elements	Terrain roughness class
LCZ 1	0.2—0.4	>2	40—60	40—60	<10	>25	8
LCZ 2	0.3—0.6	0.75—2	40—70	30—50	<20	10—25	6—7
LCZ 3	0.2—0.6	0.75—1.5	40—70	20—50	<30	3—10	6
LCZ 4	0.5—0.7	0.75—1.25	20—40	30—40	30—40	>25	7—8
LCZ 5	0.5—0.8	0.3—0.75	20—40	30—50	20—40	10—25	5—6
LCZ 6	0.6—0.9	0.3—0.75	20—40	20—50	30—60	3—10	5—6
LCZ 7	0.2—0.5	1—2	60—90	<20	<30	2—4	4—5
LCZ 8	>0.7	0.1—0.3	30—50	40—50	<20	3—10	5
LCZ 9	>0.8	0.1—0.25	10—20	<20	60—80	3—10	5—6
LCZ 10	0.6—0.9	0.2—0.5	20—30	20—40	40—50	5—15	5—6
LCZ A	<0.4	>1	<10	<10	>90	3—30	8
LCZ B	0.5—0.8	0.25—0.75	<10	<10	>90	3—15	5—6
LCZ C	0.7—0.9	0.25—1.0	<10	<10	>90	<2	4—5
LCZ D	>0.9	<0.1	<10	<10	>90	<1	3—4
LCZ E	>0.9	<0.1	<10	>90	<10	<0.25	1—2
LCZ F	>0.9	<0.1	<10	<10	>90	<0.25	1—2
LCZ G	>0.9	<0.1	<10	<10	>90	—	1

Those properties are calculated as follows (Stewart and Oke, 2012):

- *Sky view factor*: Ratio of the amount of sky hemisphere visible from ground level to that of an unobstructed hemisphere
- *Aspect ratio*: Mean height-to-width ratio of street canyons (LCZ 1 — 7), building spacing (LCZ 8 — 10), and tree spacing (LCZ A — G)
- *Building surface fraction*: Ratio of building plan area to total plan area (%)
- *Impervious surface fraction*: Ratio of impervious plan area (paved, rock) to total plan area (%)
- *Pervious surface fraction*: Ratio of pervious plan area (bare soil, vegetation, water) to total plan area (%)
- *Height of roughness elements*: Geometric average of building heights (LCZ 1 — 10) and tree/plant heights (LCZ A — F) (m)

- *Terrain roughness class*: Davenport et al. (2000) classification of effective terrain roughness ( $z_0$ ) for city and country landscapes

## A.2 Image Augmentations

**Table A.2**  $T_{pre}$  augmentations used for the SimCLR pretext task, with both the original augmentations used by Van Gansbeke et al. (2020) and the adapted augmentations used in this thesis. Normalization and color augmentations are not used due to the nature of the LCZ data. Scale of *Random Resized Crop* is changed and a *Vertical Flip* and a *Rotation* is added.

Transformation	$T_{pre}$ Original	$T_{pre}$ Adaption
Random Resized Crop	Size: 32, Scale [0.2—1.0]	Size: 32, Scale [0.75—1.0]
Color Jitter	Probability: 80%	—
Random Grayscale	Probability: 20%	—
Normalization	<i>data dependent</i>	—
Horizontal Flip	Probability: 50%	Probability: 50%
Vertical Flip	—	Probability: 50%
Rotation	—	Range: [0°,360°]

**Table A.3**  $T_{scan}/T_{self}$  augmentations used for clustering and selflabeling, with both the original augmentations used by Van Gansbeke et al. (2020) and the adapted augmentations used in this thesis. Normalization is removed due to the nature of the LCZ data. Four *Randaugment* augmentations are used (Table A.6) and the cutout is changed to more, but smaller ones.

Transformation	$T_{scan}/T_{self}$ Original	$T_{scan}/T_{self}$ Adaption
Horizontal Flip	Probability: 50%	Probability: 50%
R <sub>SCAN</sub> Augmentations	4	6
Normalization	<i>data dependent</i>	—
Cutout	no. holes: 1, size: 16 × 16	no. holes: 32, size: 2 × 2

**Table A.4**  $T_{weak}$  augmentations used for co-refinement, with both the original augmentations used by S. Park et al. (2021) and the adapted augmentations used in this thesis. Normalization is removed due to the nature of the LCZ data and a *Rotation* is added.

Transformation	$T_{weak}$ Original	$T_{weak}$ Adaption
Random Resized Crop	Size: 32, Scale [0.2—1.0]	Size: 32, Scale [0.75—1.0]
Normalization	<i>data dependent</i>	—
Horizontal Flip	Probability: 50%	Probability: 50%
Vertical Flip	—	Probability: 50%
Rotation	—	Range: [0°,360°]

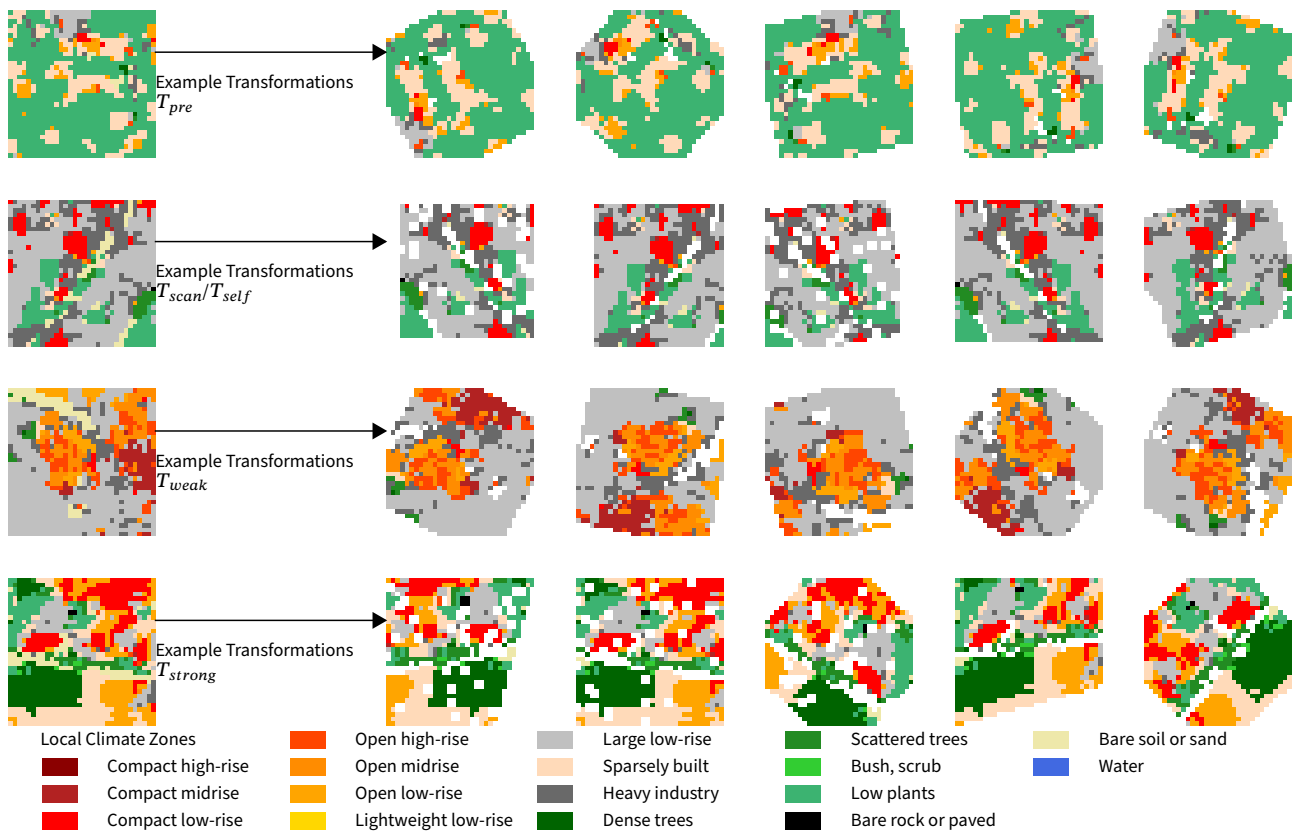


**Table A.5**  $T_{strong}$  augmentations used for label smoothing, with both the original augmentations used by S. Park et al. (2021) and the adapted augmentations used in this thesis. Normalization is removed due to the nature of the LCZ data. Two *Randaugment* augmentations are used (Table A.6) and the cutout is changed to more, but smaller ones.

Transformation	$T_{strong}$ Original	$T_{strong}$ Adaption
Random Resized Crop	Size: 32, Scale [0.2—1.0]	Size: 32, Scale [0.75—1.0]
Horizontal Flip	Probability: 50%	Probability: 50%
$R_{RUC}$ Augmentations	2	2
Normalization	<i>data dependent</i>	—
Cutout	no. holes: 1, size: $16 \times 16$	no. holes: 32, size: $2 \times 2$

**Table A.6** Randaugment augmentations developed by Cubuk et al. (2020), with both the original augmentations used by Van Gansbeke et al. (2020) ( $R_{SCAN}$ ) and S. Park et al. (2021) ( $R_{RUC}$ ) as well as the adapted augmentations used in this thesis. For both, a number of augmentations from this list is randomly chosen. Color augmentations are not used due to the nature of the LCZ data.

Transformation	$R_{SCAN}$ Original	$R_{SCAN}$ Adaption	$R_{RUC}$ Original	$R_{RUC}$ Adaption
Identity	Yes	Yes	Yes	Yes
Autocontrast	Yes	—	Yes	—
Equalize	Yes	—	Yes	—
Rotate	$[-30^\circ, 30^\circ]$	$[-30^\circ, 30^\circ]$	$[-45^\circ, 45^\circ]$	$[-45^\circ, 45^\circ]$
Solarize	[0, 256]	—	[0, 256]	—
Color	[0.05, 0.95]	—	[0.01, 0.99]	—
Contrast	[0.05, 0.95]	—	[0.01, 0.99]	—
Brightness	[0.05, 0.95]	—	[0.01, 0.99]	—
Sharpness	[0.05, 0.95]	—	[0.01, 0.99]	—
Shear X	[-0.1, 0.1]	[-0.1, 0.1]	[-0.3, 0.3]	[-0.3, 0.3]
Translation X	[-0.1, 0.1]	[-0.1, 0.1]	[-0.3, 0.3]	[-0.3, 0.3]
Shear Y	[-0.1, 0.1]	[-0.1, 0.1]	[-0.1, 0.1]	[-0.1, 0.1]
Translation Y	[-0.1, 0.1]	[-0.1, 0.1]	[-0.3, 0.3]	[-0.3, 0.3]
Posterize	[4, 8]	—	[1, 8]	—



**Figure A.1** An overview of a random sample of augmented patches.

## A.3 Overview of all Experiments

**Table A.7** Overview of all Experiments performed to find urban patterns during the first part of this thesis. Three different datasets were used: *MUA* containing all patches inside urban areas, *MUA\** excludes patches with more than 85% water, and *Tiles* includes patches from the whole raster tiles.

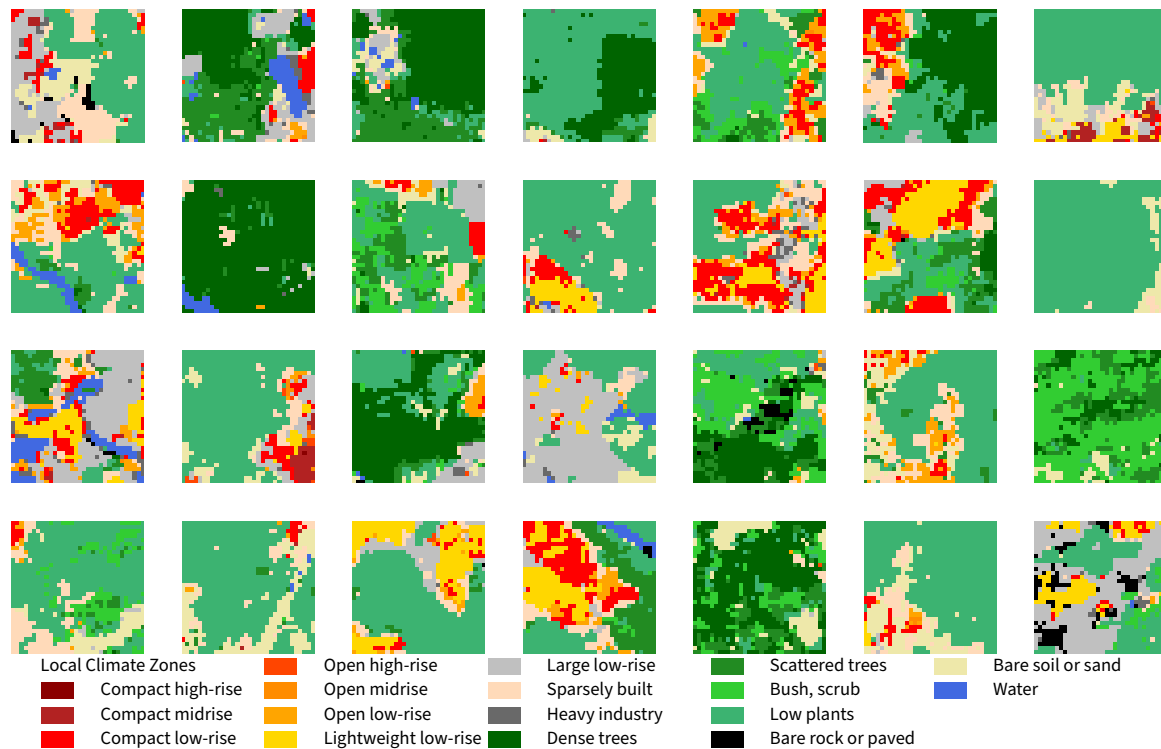
Backbone	Dataset	No. Classes	No. Output Classes
ResNet-18	MUA*	128	25
ResNet-18	MUA*	256	24
ResNet-18	MUA*	512	39
ResNet-18	MUA*	1024	40
ResNet-18	MUA	1024	64
ResNet-18	MUA*	2048	—
ResNet-18	Tiles	128	23
ResNet-18	Tiles	256	20
ResNet-18	Tiles	512	24
ResNet-18	Tiles	1024	25
ResNet-18	Tiles	2048	—
<b>ResNet-50</b>	<b>MUA</b>	<b>1024</b>	<b>66</b>

**Table A.8** Overview of all Experiments performed to cluster graphs during the second part of this thesis. *Node Feature* was either the class label (value between 1 and 66) or the mean logits of the class label (1024 dimensional vector). *Additional Node Features* contained principal components of multiple geometric and morphological parameters. *Structure Matrix* was either the adjacency matrix of the graph or the (Hausdorff) distance between the regions corresponding to the graph nodes.

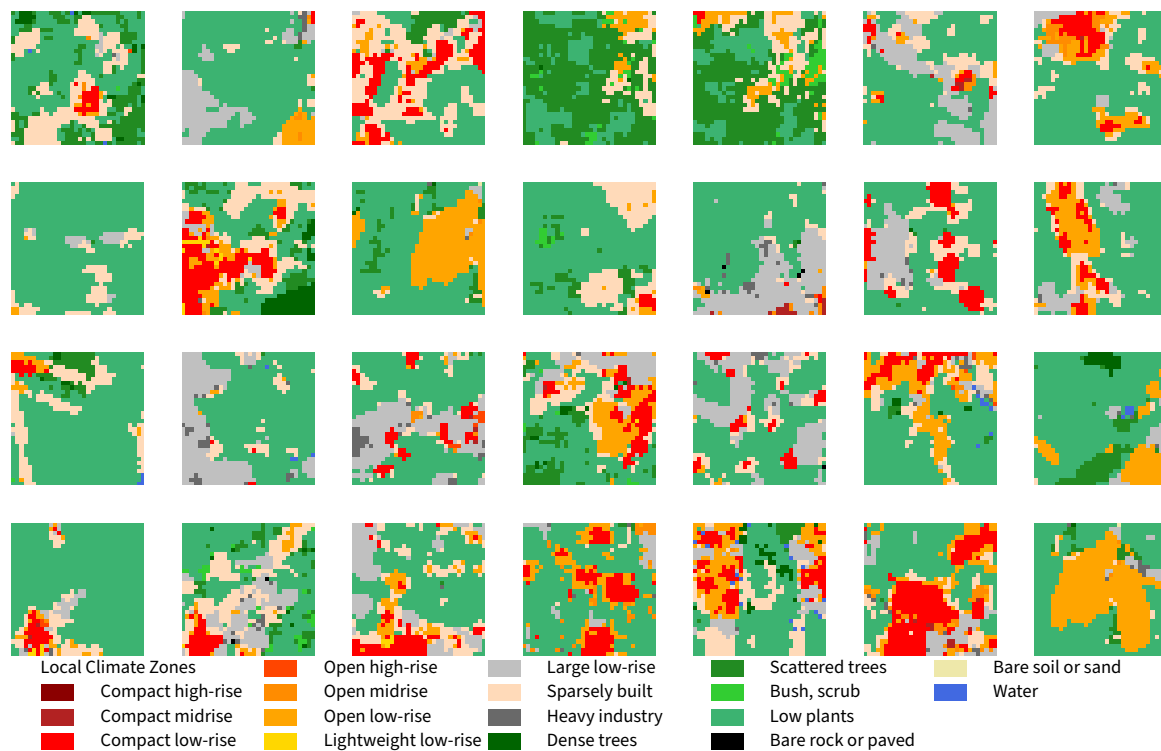
Node Feature	Additional Node Feature	Structure Matrix	No. Classes
Class Label	PCA	Adjacency matrix	10
Class Label	PCA	Distance	10
Class Label	PCA	Hausdorff distance	10
Logits	—	Adjacency matrix	10
Logits	—	Distance	10
Logits	—	Hausdorff distance	10
<b>Logits</b>	—	<b>Adjacency matrix</b>	<b>16</b>

# B Overview Results

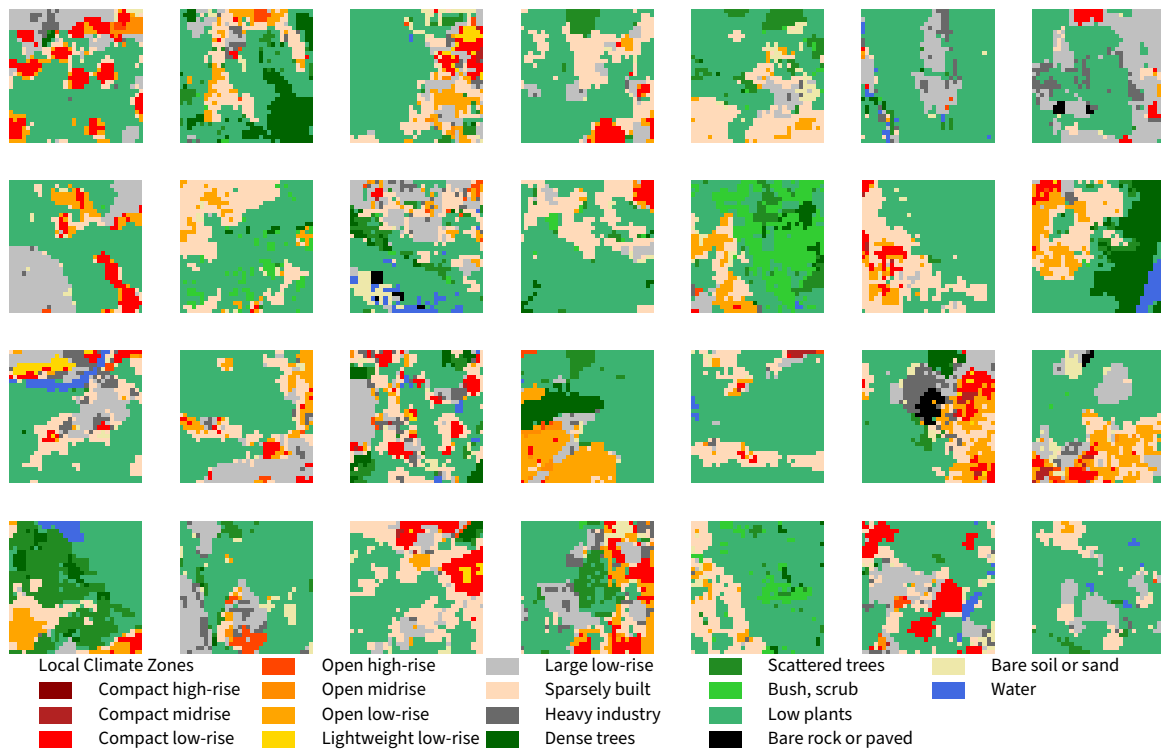
## B.1 Overview Patches



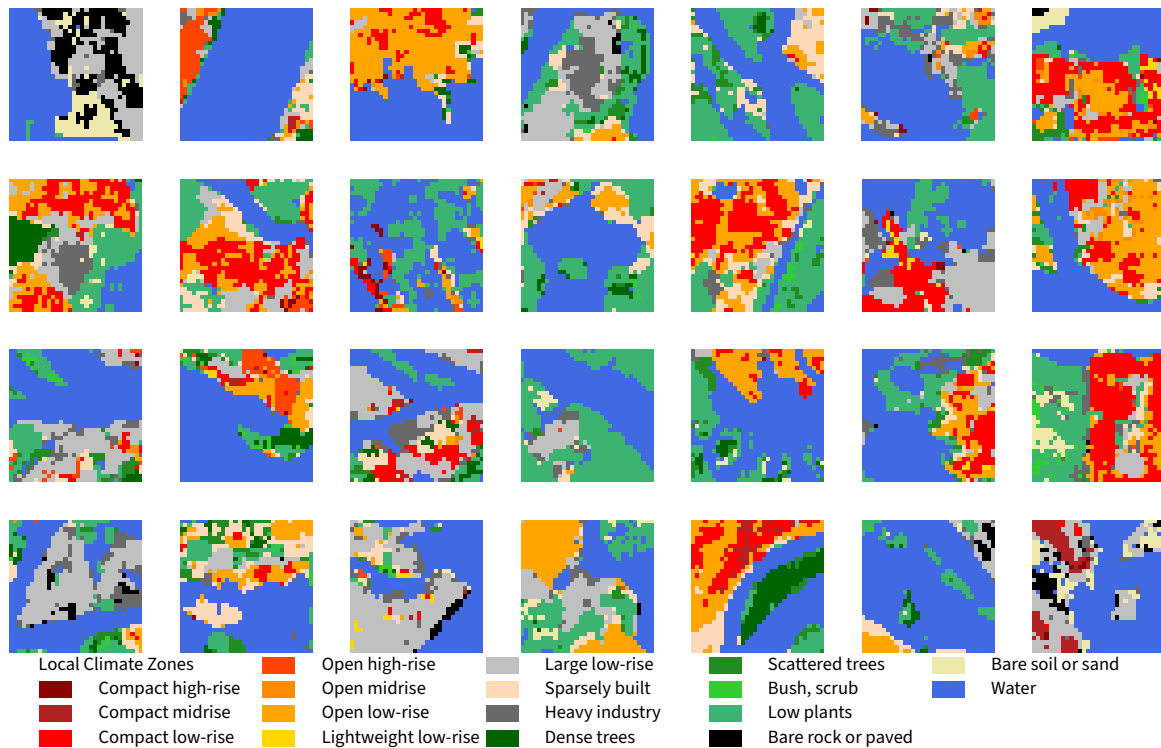
**Figure B.1** An overview of a random sample of extracted patches in  $c_{UB_1}$  with a total of 3029 patches.



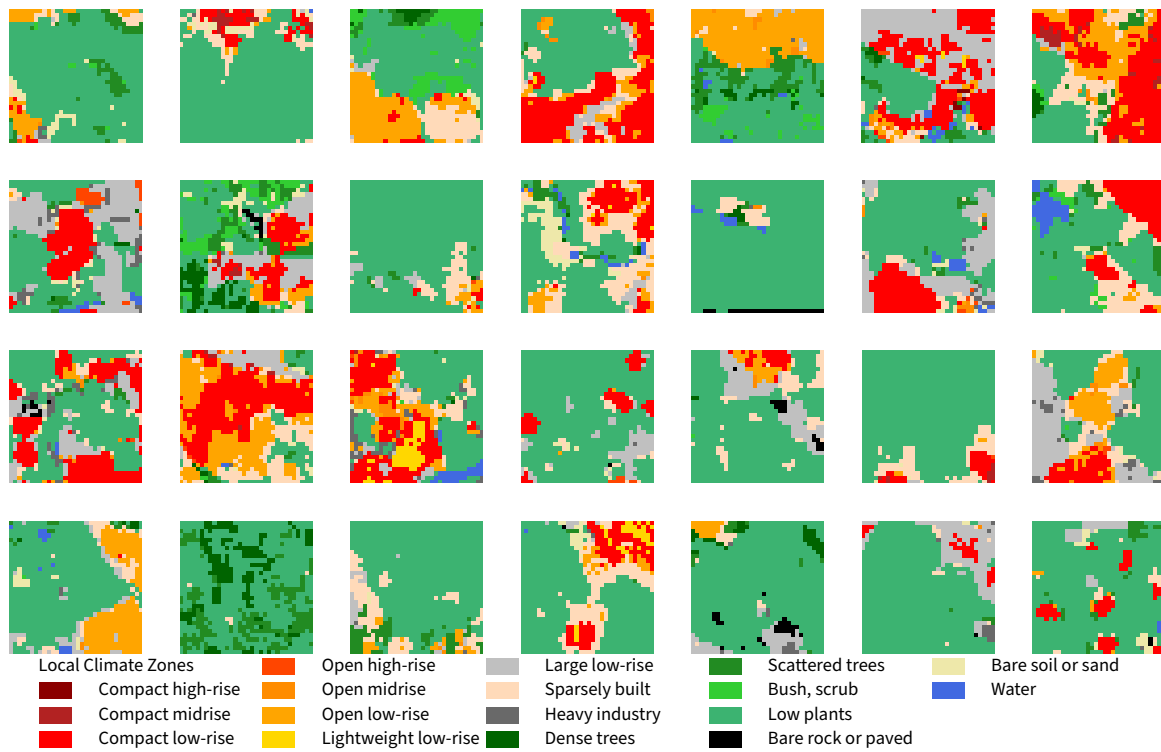
**Figure B.2** An overview of a random sample of extracted patches in  $c_{UB_2}$  with a total of 3135 patches.



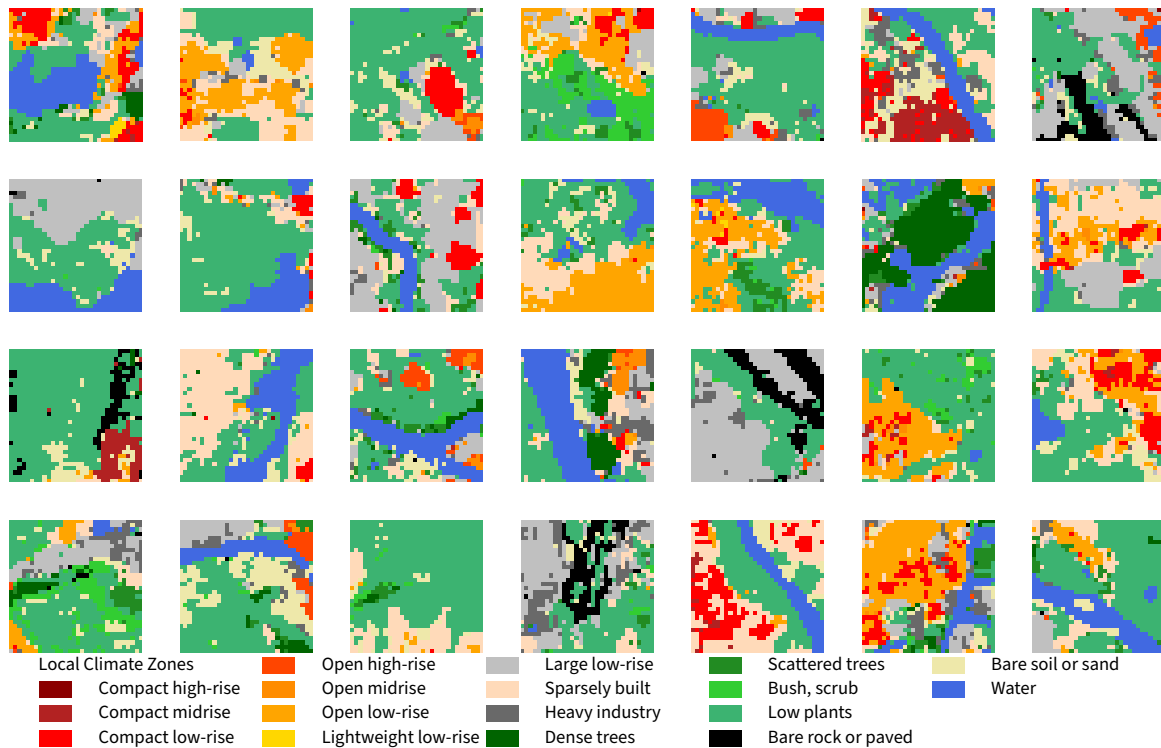
**Figure B.3** An overview of a random sample of extracted patches in  $c_{UB_4}$  with a total of 6753 patches.



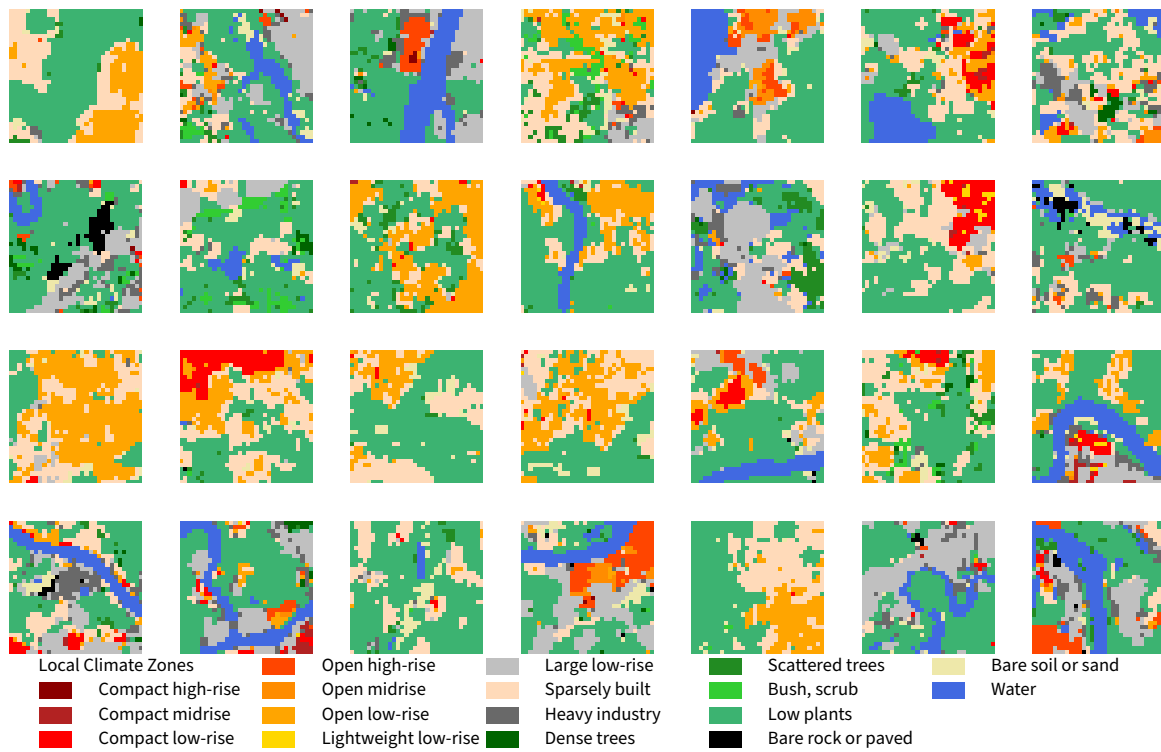
**Figure B.4** An overview of a random sample of extracted patches in  $c_{UB_5}$  with a total of 6752 patches.



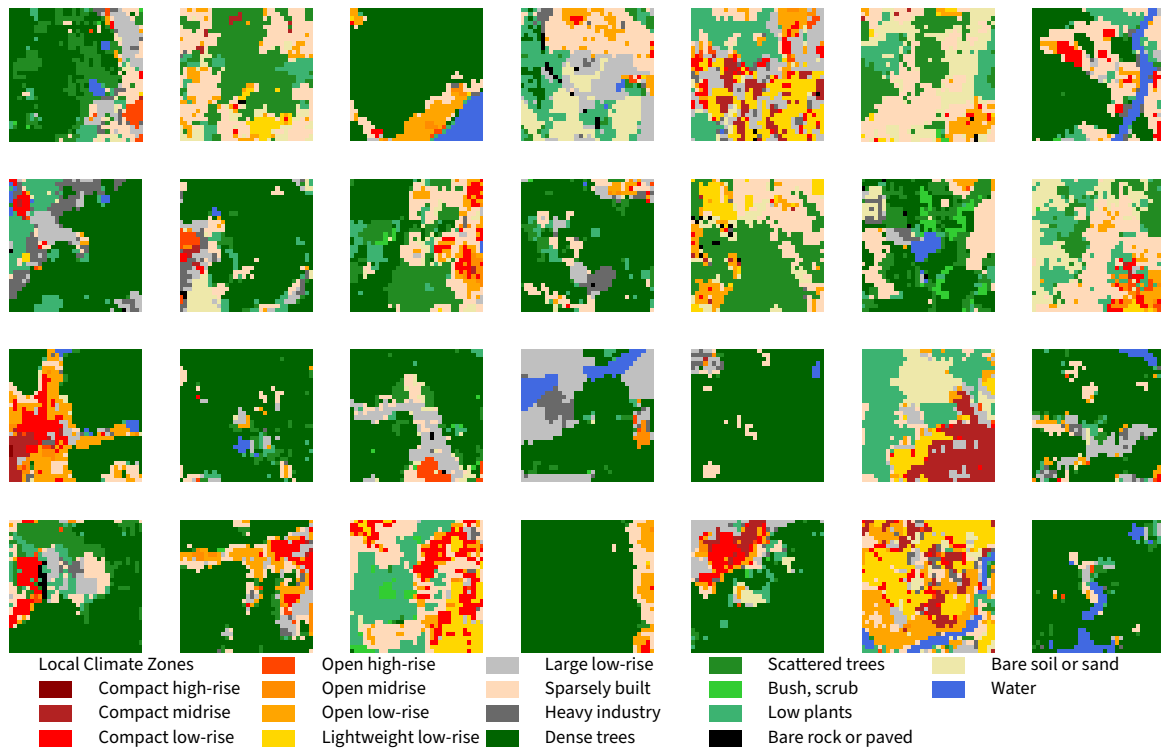
**Figure B.5** An overview of a random sample of extracted patches in  $c_{UB_6}$  with a total of 9257 patches.



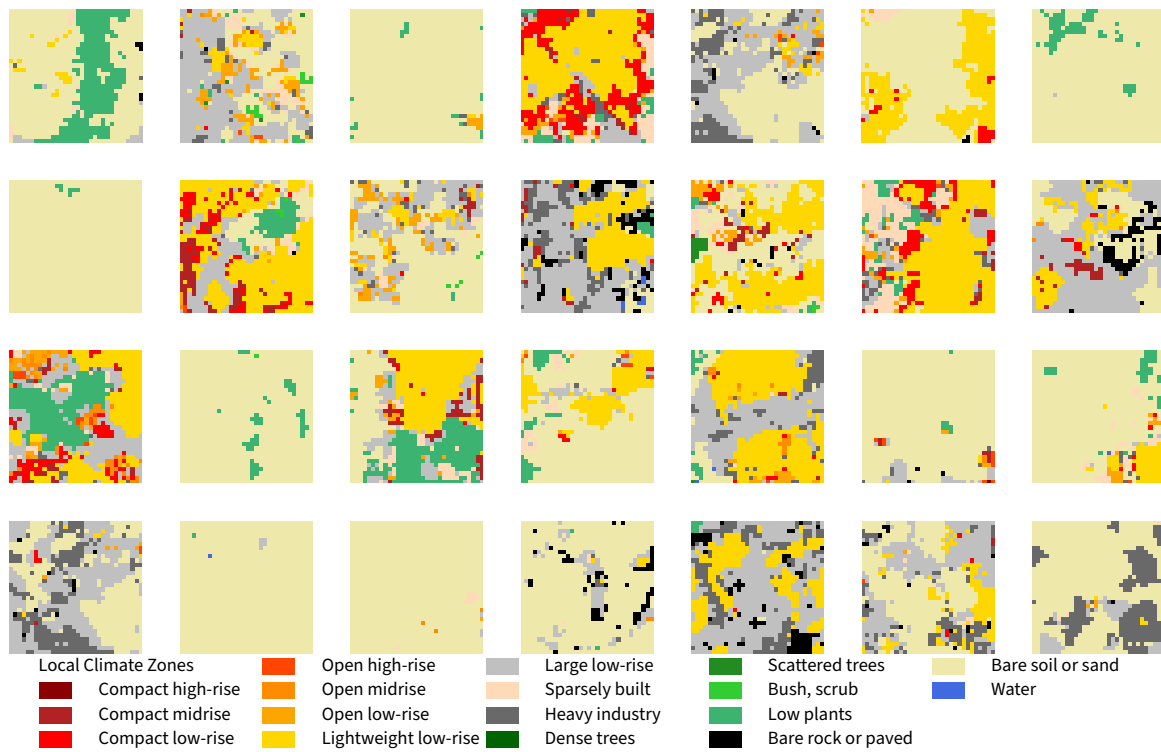
**Figure B.6** An overview of a random sample of extracted patches in  $c_{UB_7}$  with a total of 4530 patches.



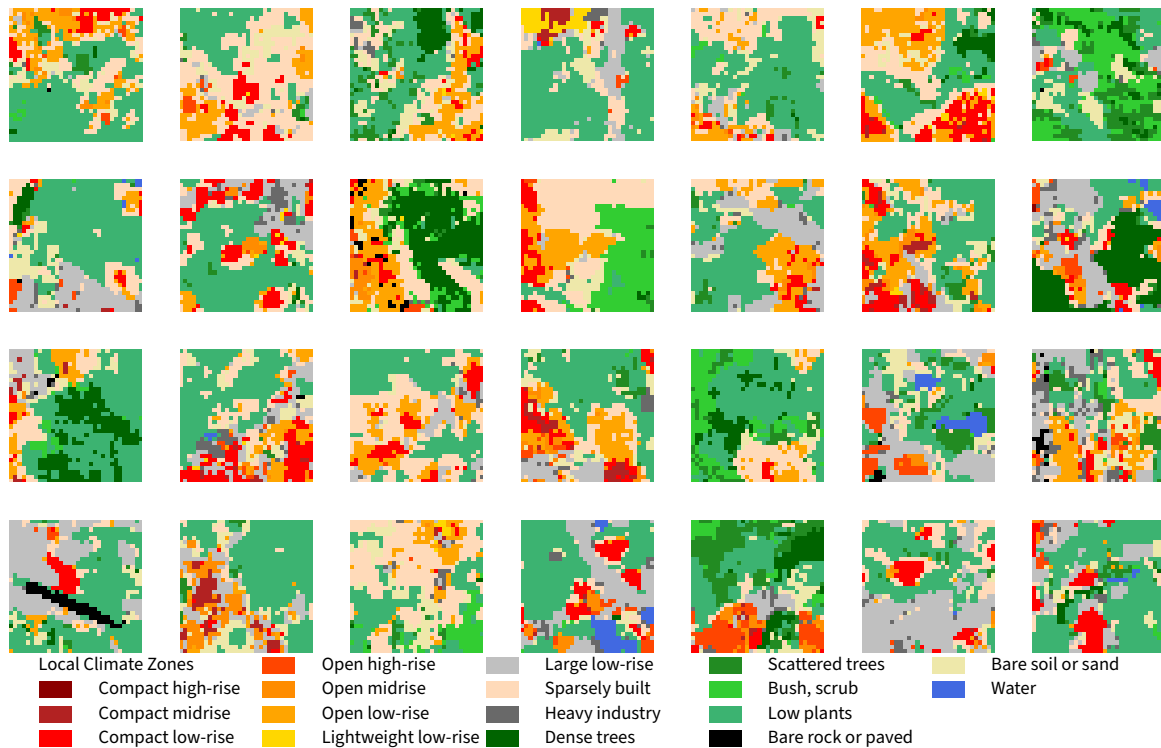
**Figure B.7** An overview of a random sample of extracted patches in  $c_{UB_8}$  with a total of 1772 patches.



**Figure B.8** An overview of a random sample of extracted patches in  $c_{UB_9}$  with a total of 2960 patches.

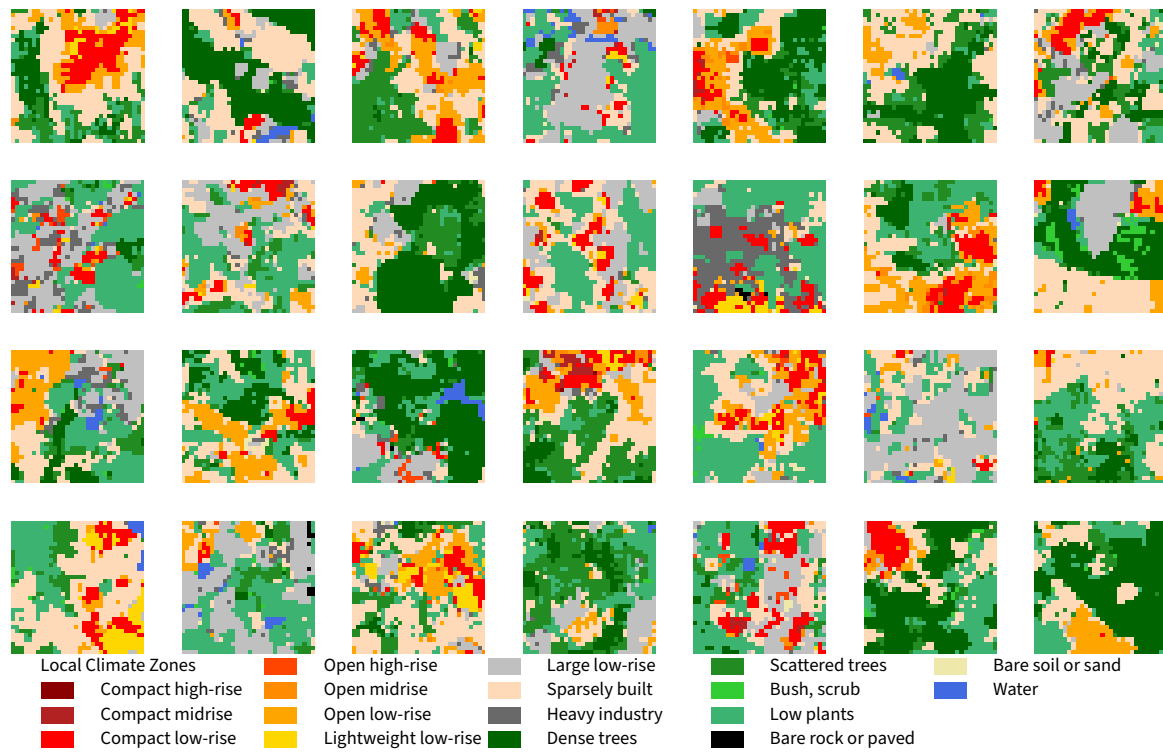


**Figure B.9** An overview of a random sample of extracted patches in  $c_{UB_{10}}$  with a total of 1539 patches.

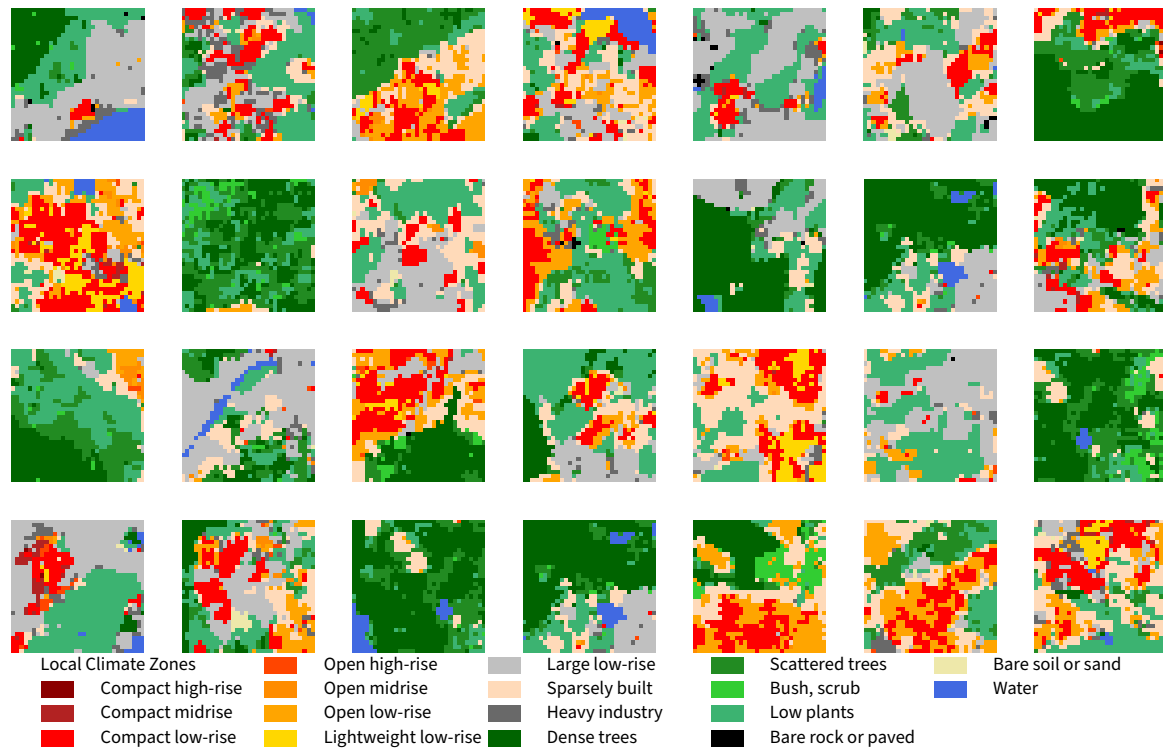


**Figure B.10** An overview of a random sample of extracted patches in  $c_{UB_{11}}$  with a total of 5482 patches.

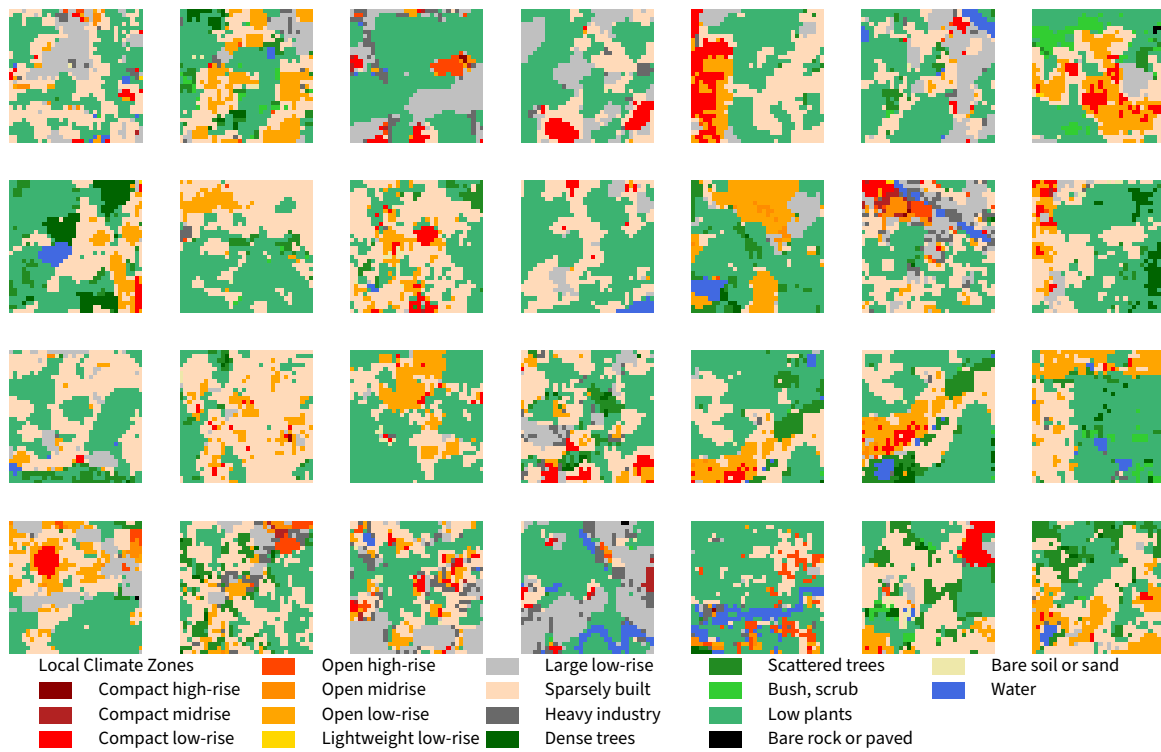




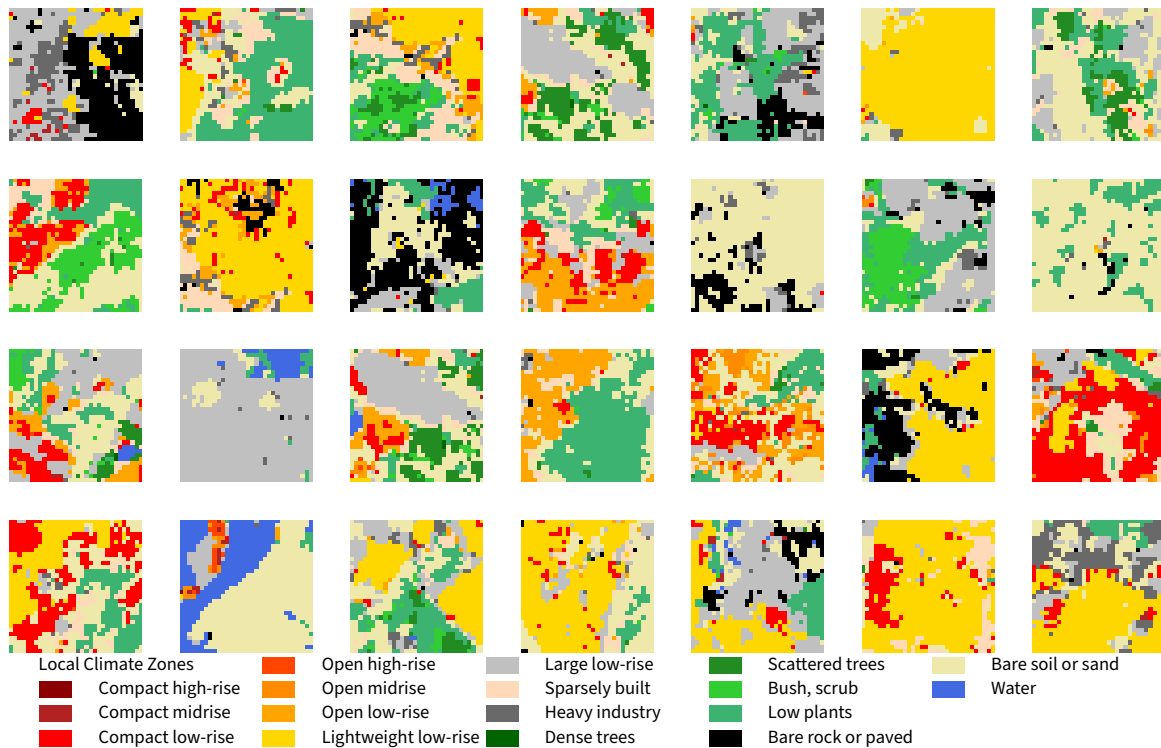
**Figure B.11** An overview of a random sample of extracted patches in  $c_{UB_{12}}$  with a total of 4018 patches.



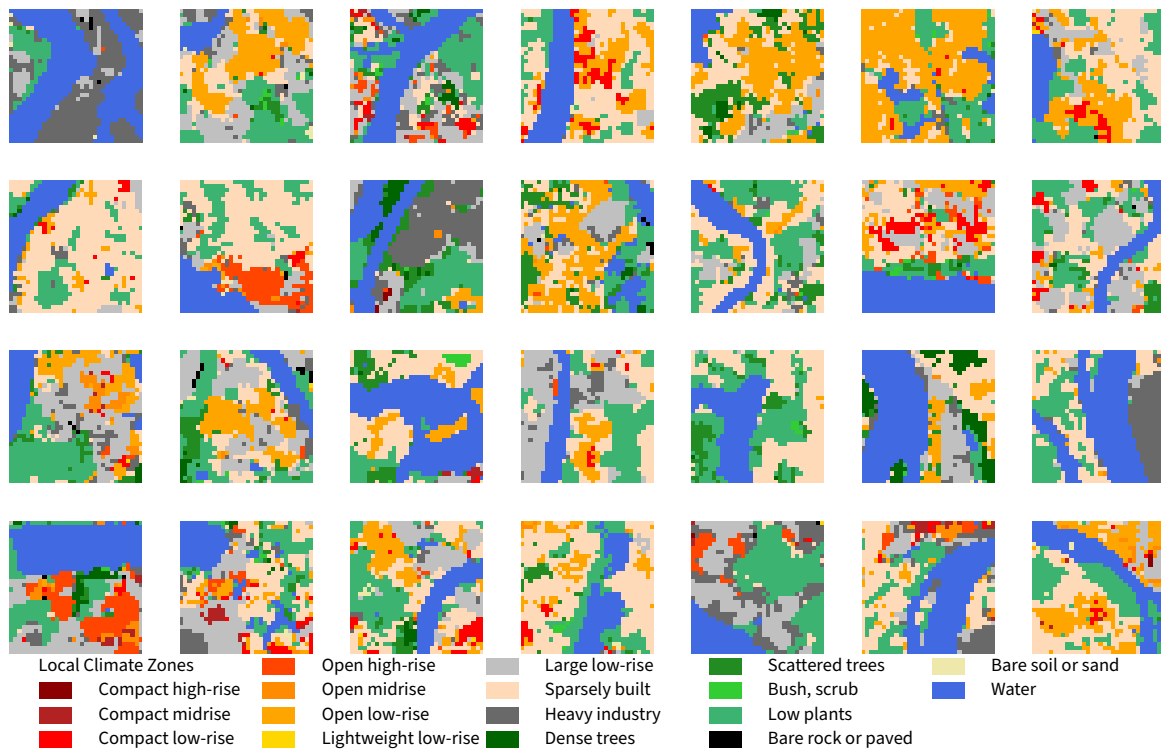
**Figure B.12** An overview of a random sample of extracted patches in  $c_{UB_{13}}$  with a total of 6757 patches.



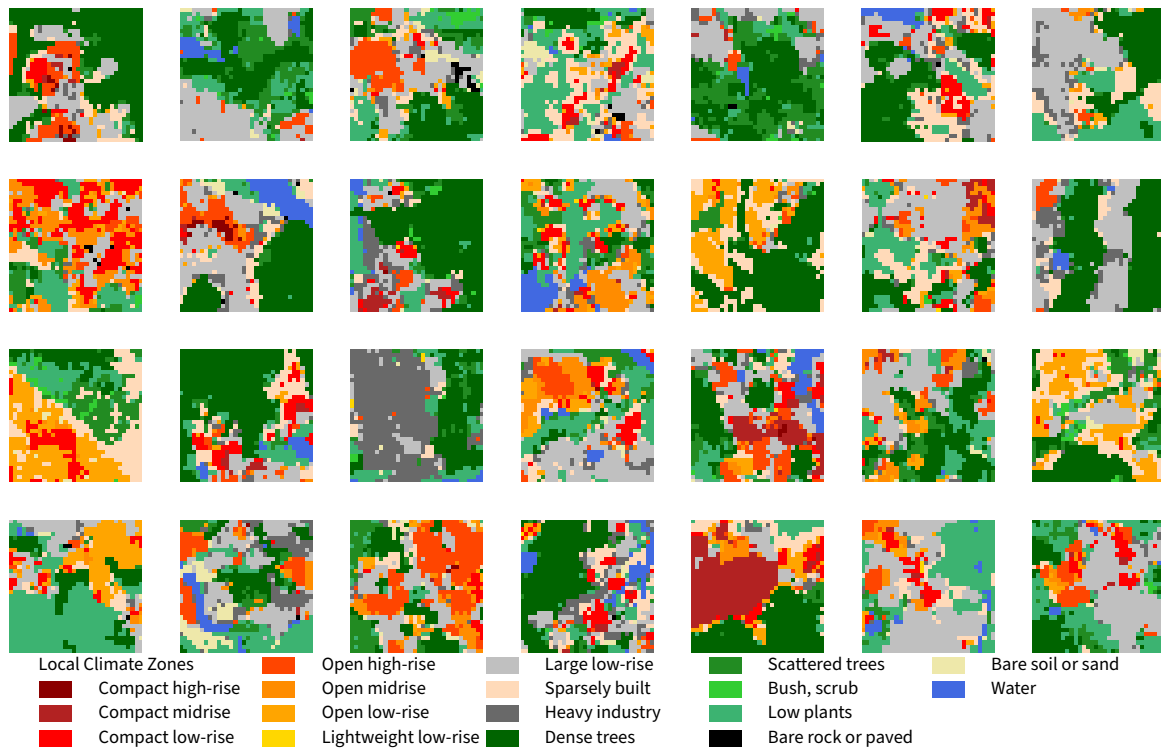
**Figure B.13** An overview of a random sample of extracted patches in  $c_{UB_{14}}$  with a total of 10145 patches.



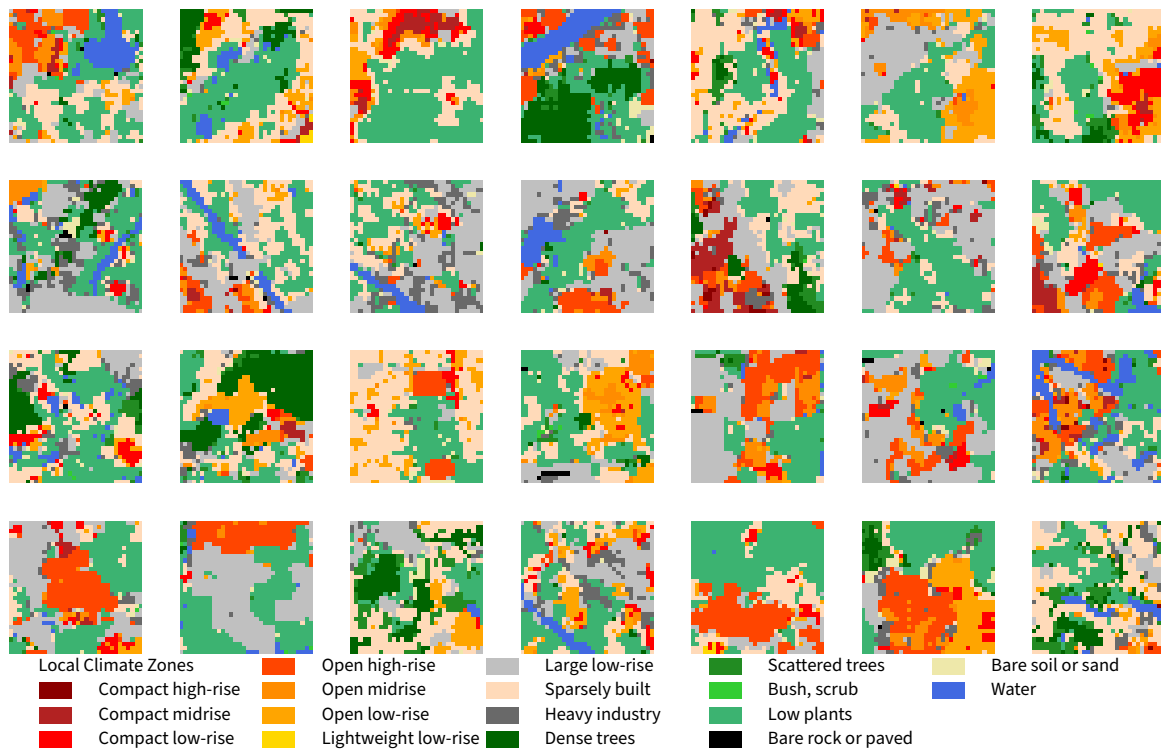
**Figure B.14** An overview of a random sample of extracted patches in  $c_{UB_{15}}$  with a total of 8784 patches.



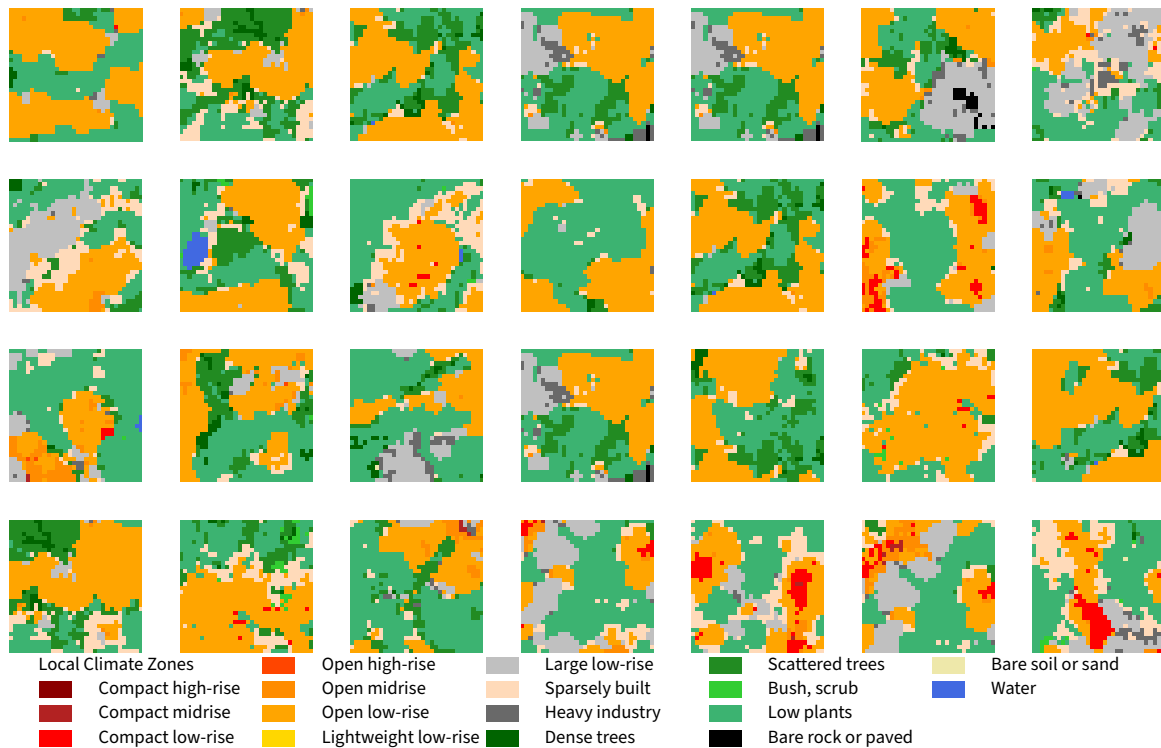
**Figure B.15** An overview of a random sample of extracted patches in  $c_{UB_{16}}$  with a total of 2088 patches.



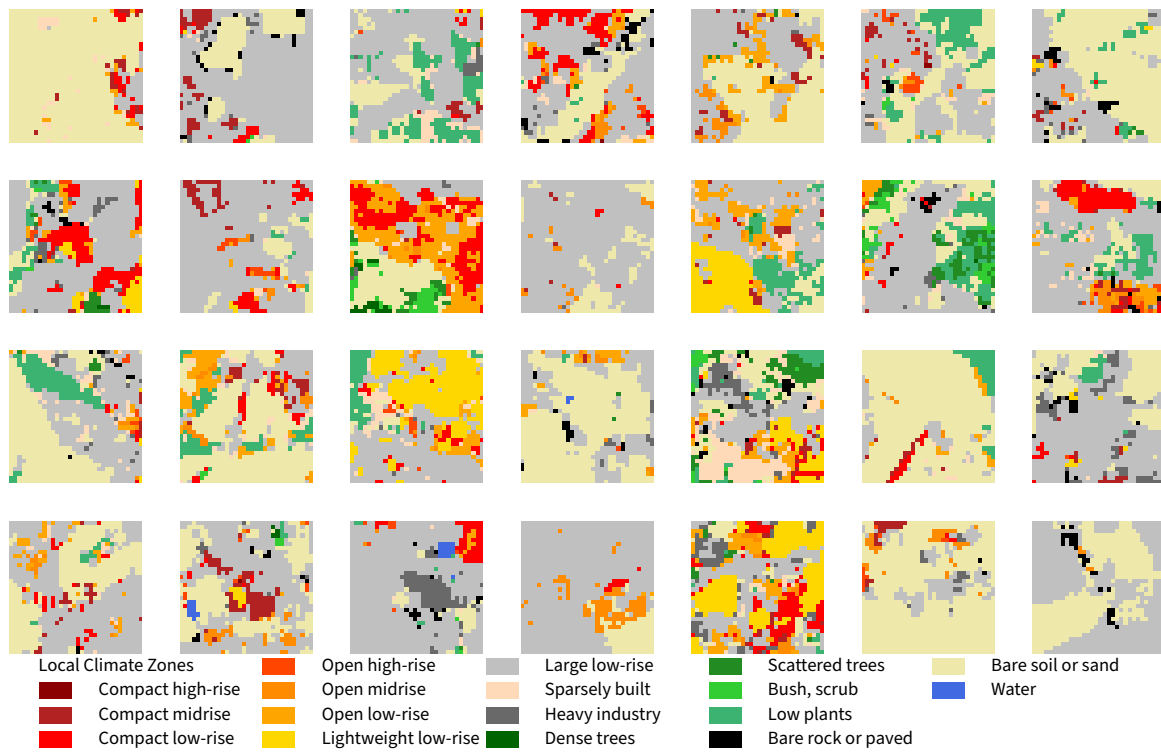
**Figure B.16** An overview of a random sample of extracted patches in  $c_{UB_{17}}$  with a total of 6414 patches.



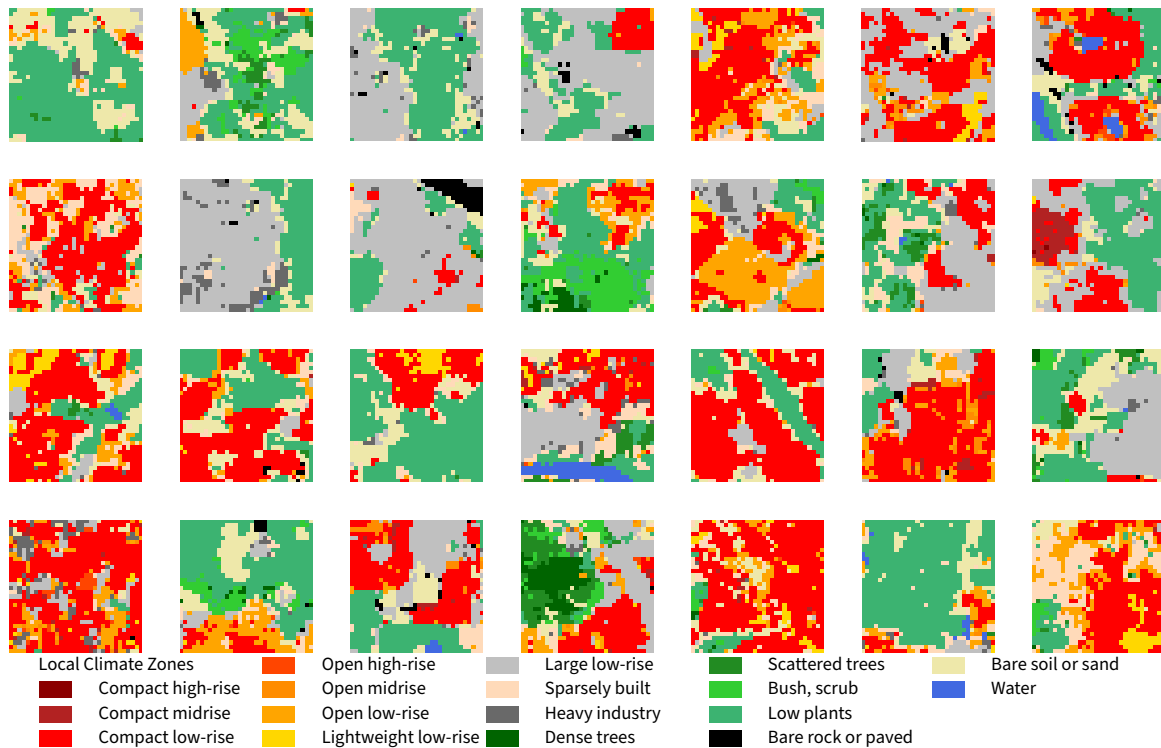
**Figure B.17** An overview of a random sample of extracted patches in  $c_{UB_{18}}$  with a total of 7667 patches.



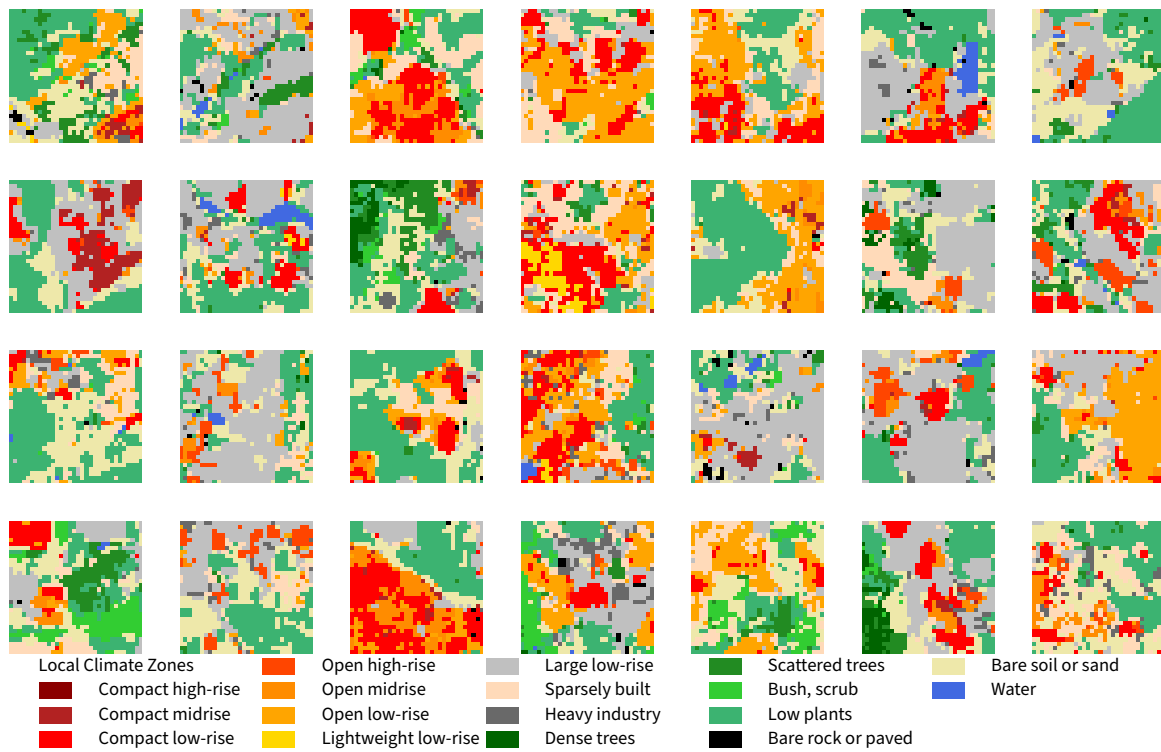
**Figure B.18** An overview of a random sample of extracted patches in  $c_{UB_{19}}$  with a total of 278 patches.



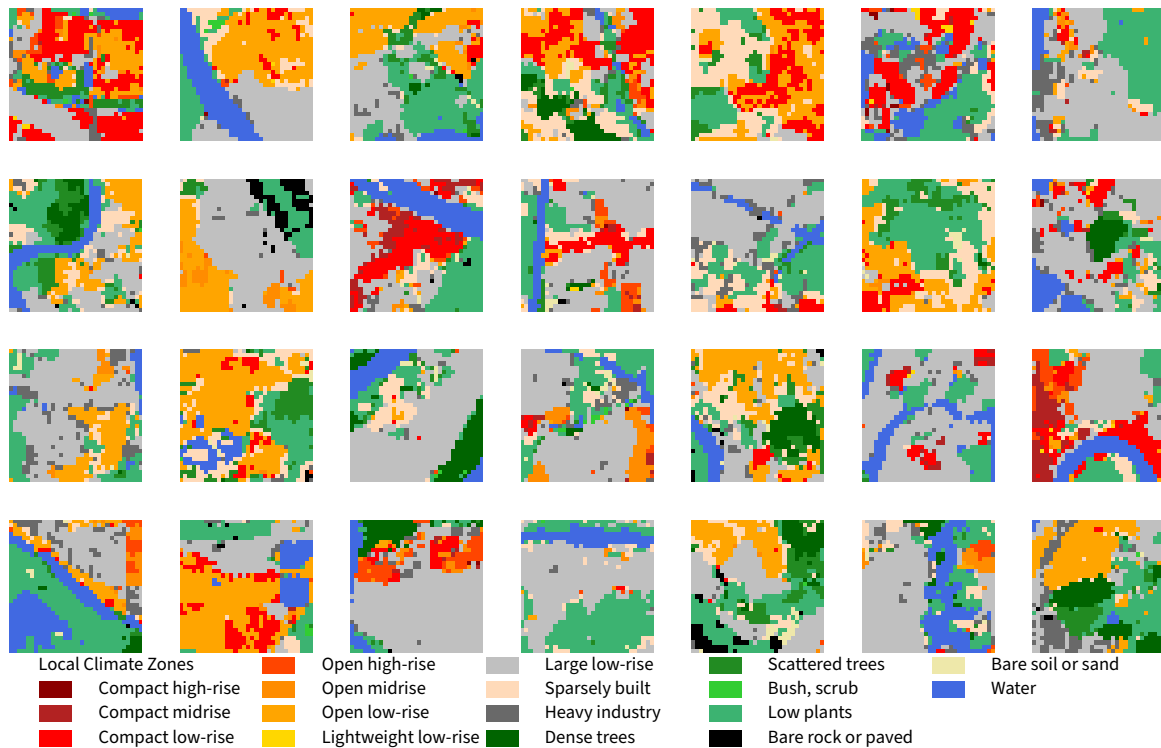
**Figure B.19** An overview of a random sample of extracted patches in  $c_{UB_{20}}$  with a total of 5599 patches.



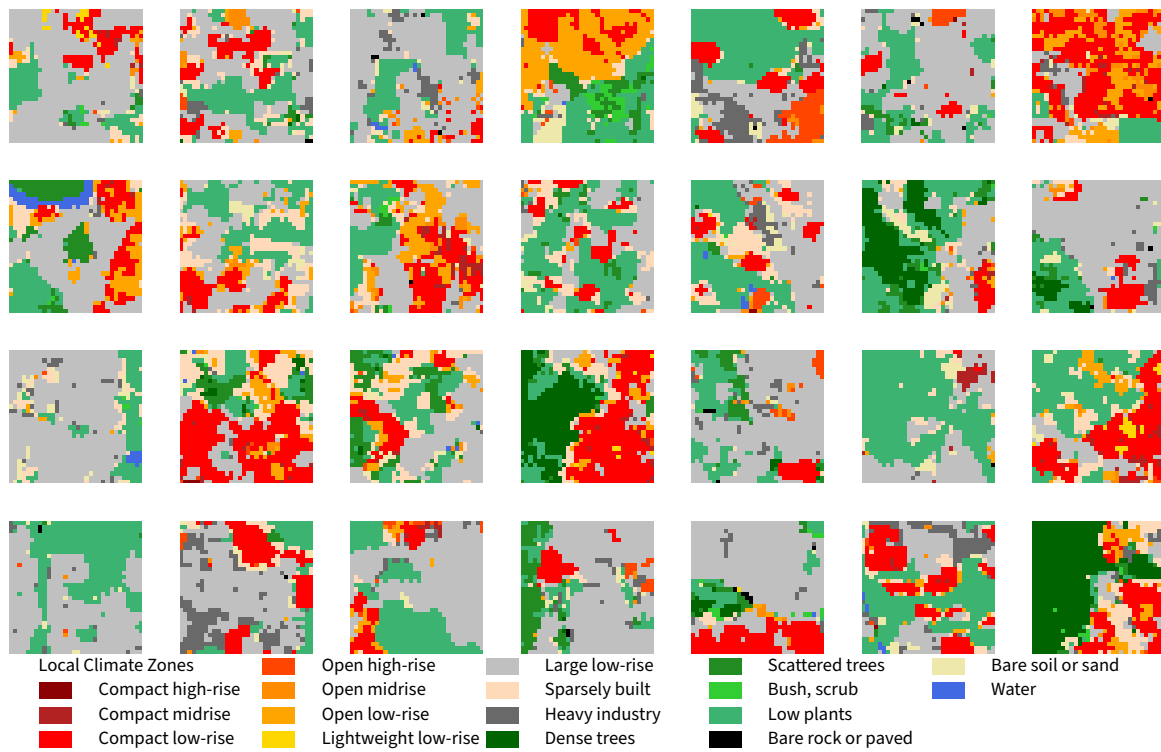
**Figure B.20** An overview of a random sample of extracted patches in  $c_{UB_{21}}$  with a total of 5311 patches.



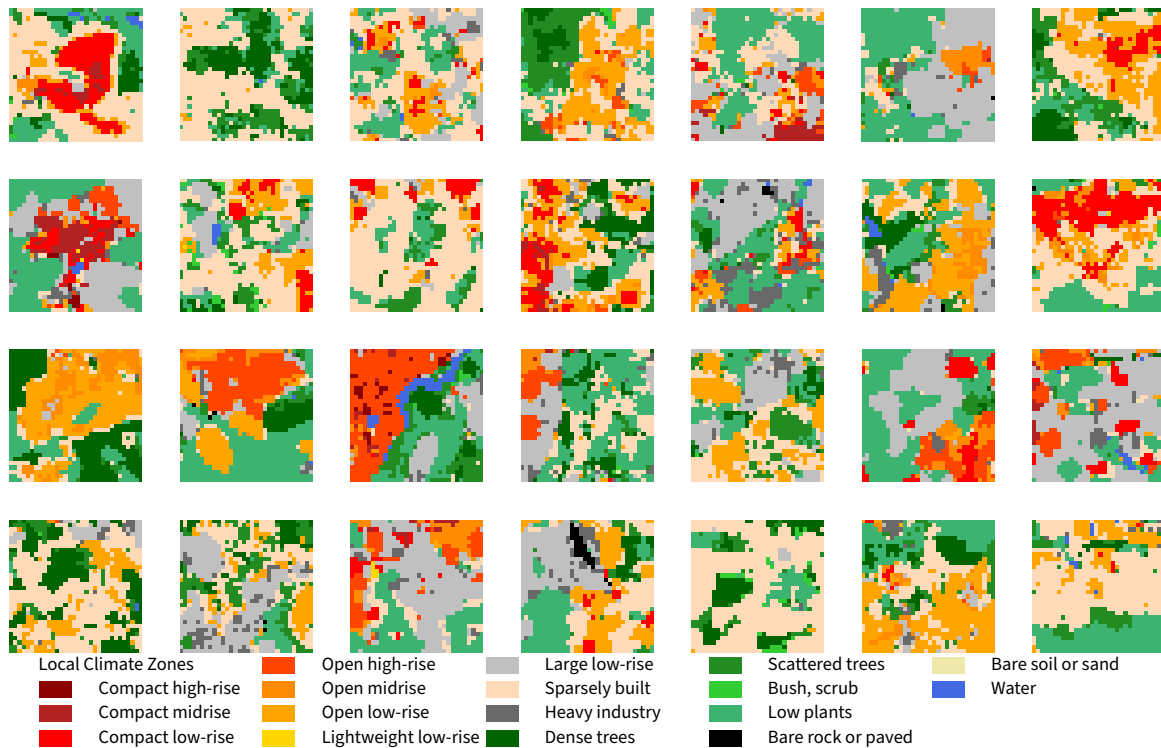
**Figure B.21** An overview of a random sample of extracted patches in  $c_{UB_{22}}$  with a total of 4654 patches.



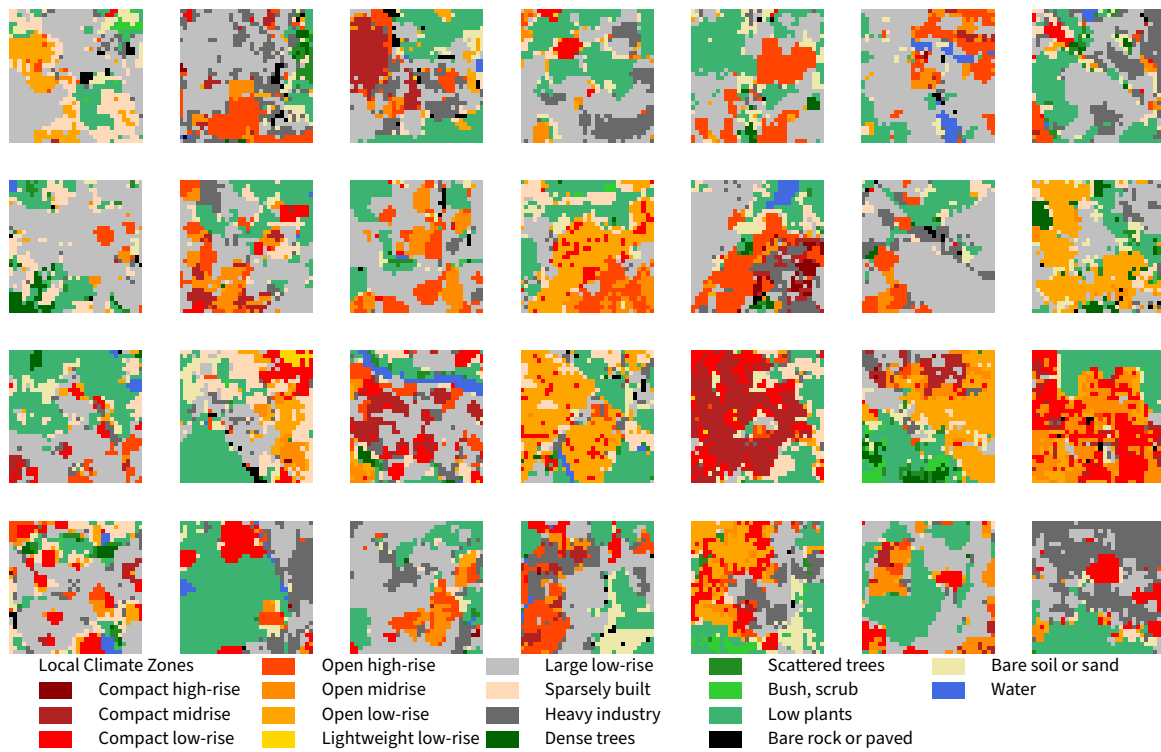
**Figure B.22** An overview of a random sample of extracted patches in  $c_{UB_{23}}$  with a total of 4298 patches.



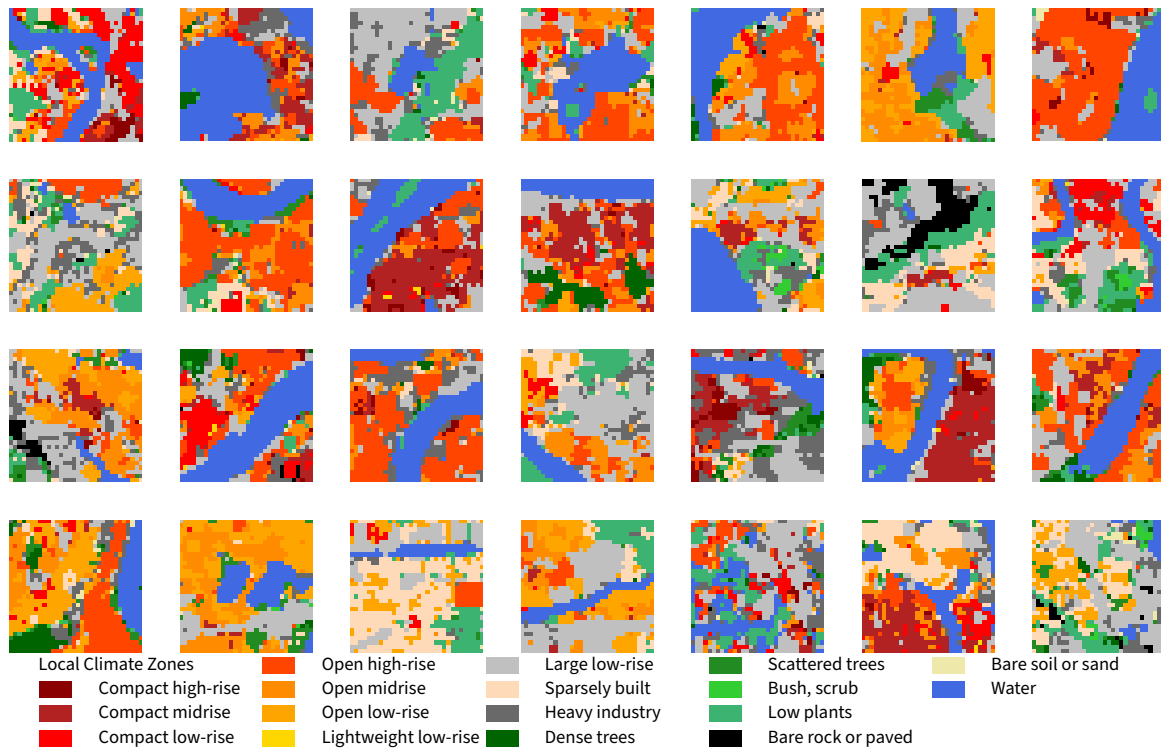
**Figure B.23** An overview of a random sample of extracted patches in  $c_{UB_{25}}$  with a total of 6548 patches.



**Figure B.24** An overview of a random sample of extracted patches in  $c_{UB_{26}}$  with a total of 5331 patches.

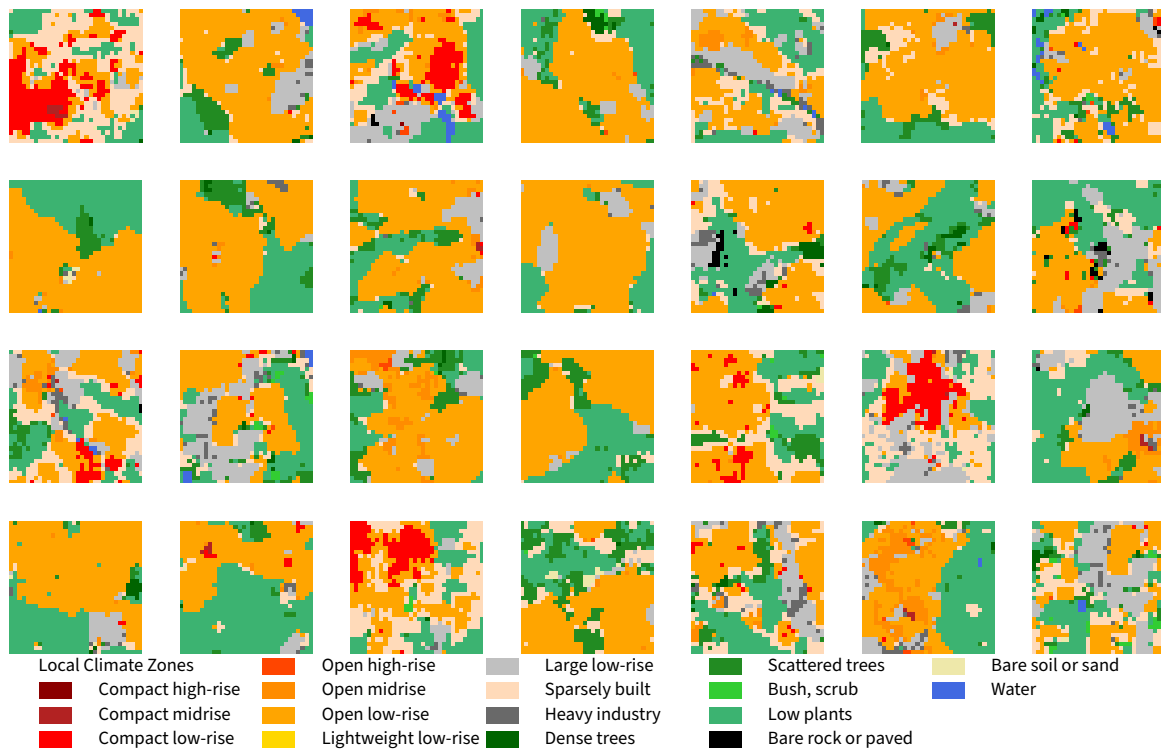


**Figure B.25** An overview of a random sample of extracted patches in  $c_{UB_{27}}$  with a total of 5940 patches.

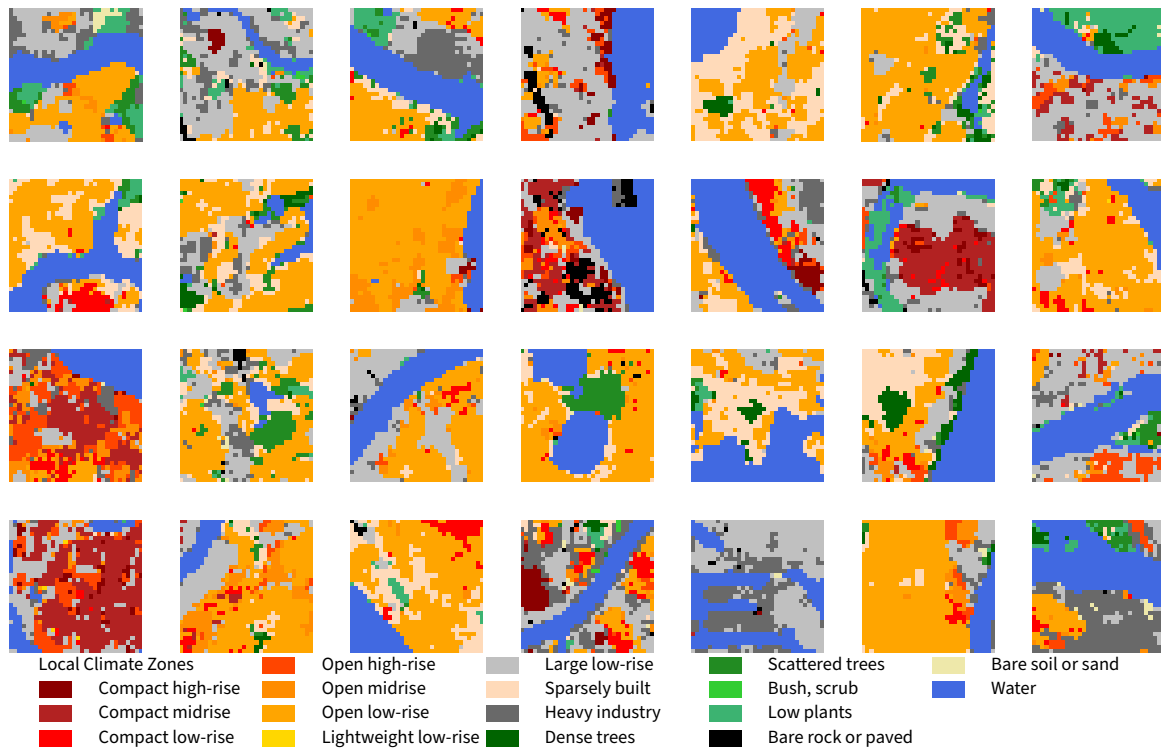


**Figure B.26** An overview of a random sample of extracted patches in  $c_{UB_{28}}$  with a total of 4678 patches.

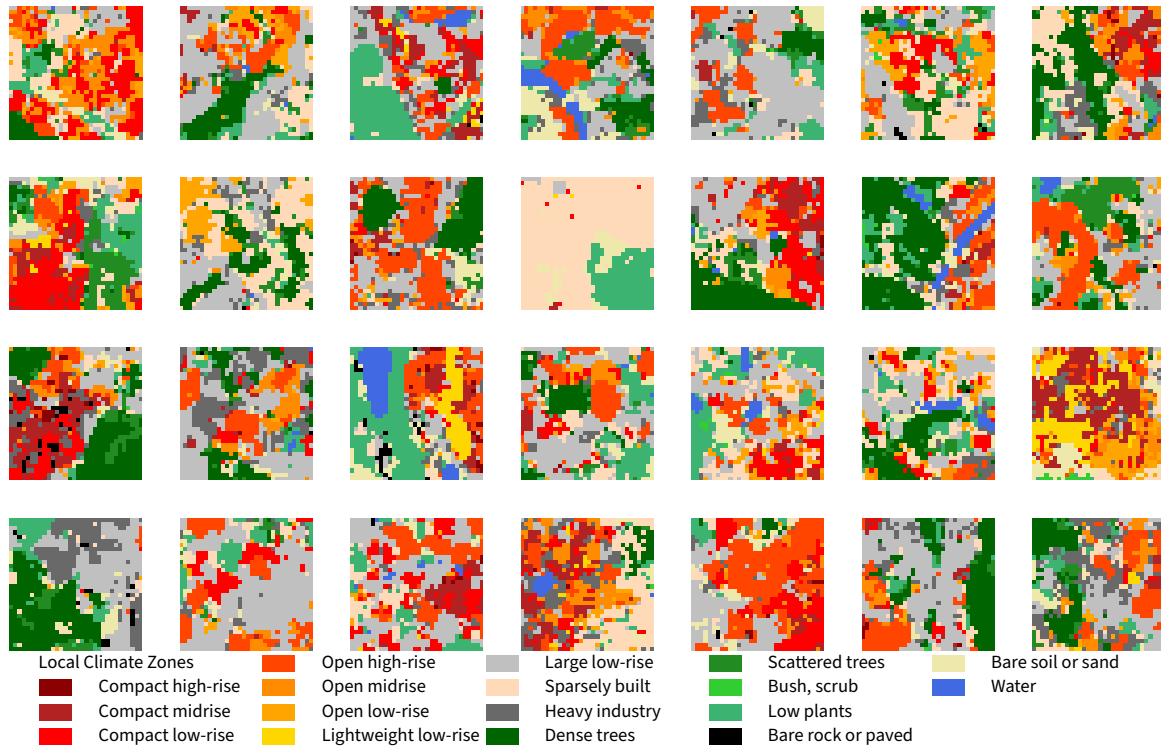




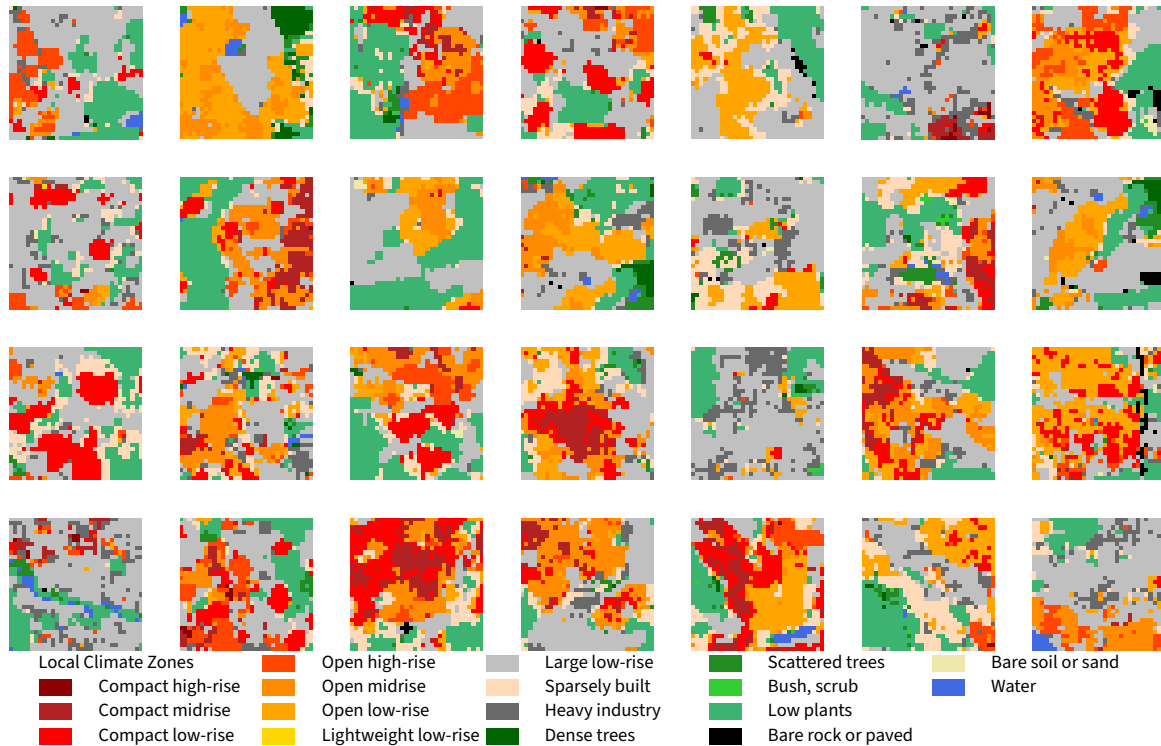
**Figure B.27** An overview of a random sample of extracted patches in  $c_{UB_{29}}$  with a total of 1740 patches.



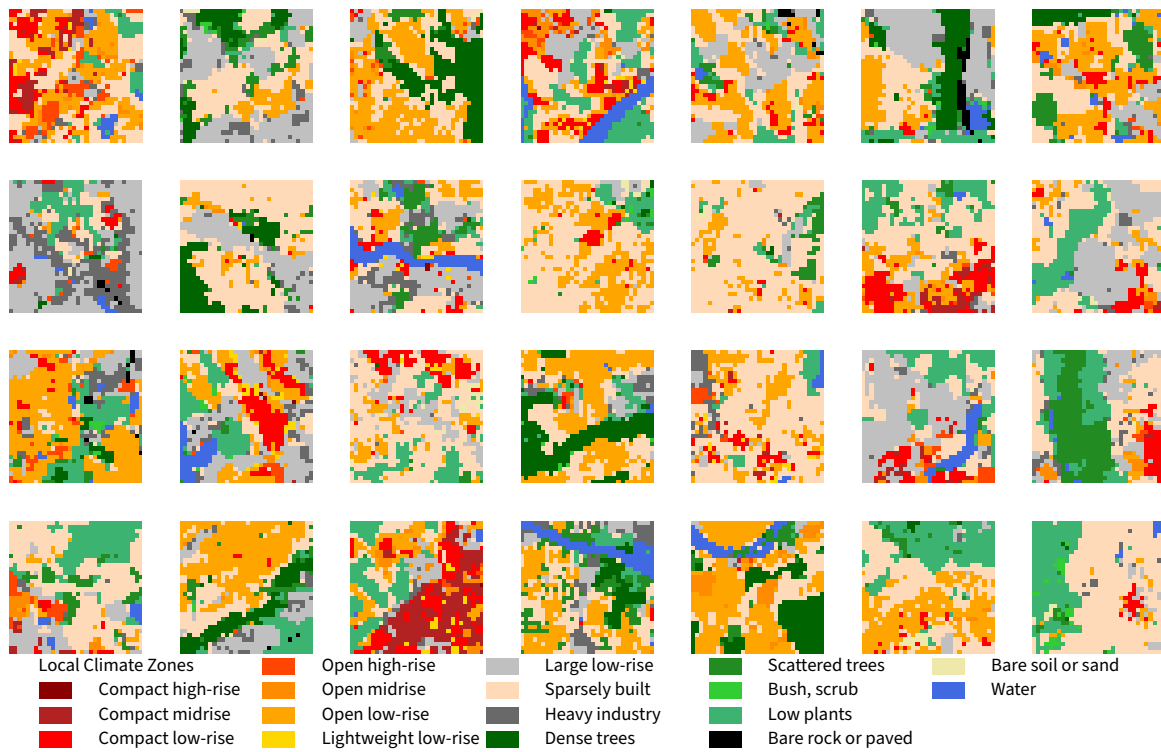
**Figure B.28** An overview of a random sample of extracted patches in  $c_{UB_{30}}$  with a total of 5214 patches.



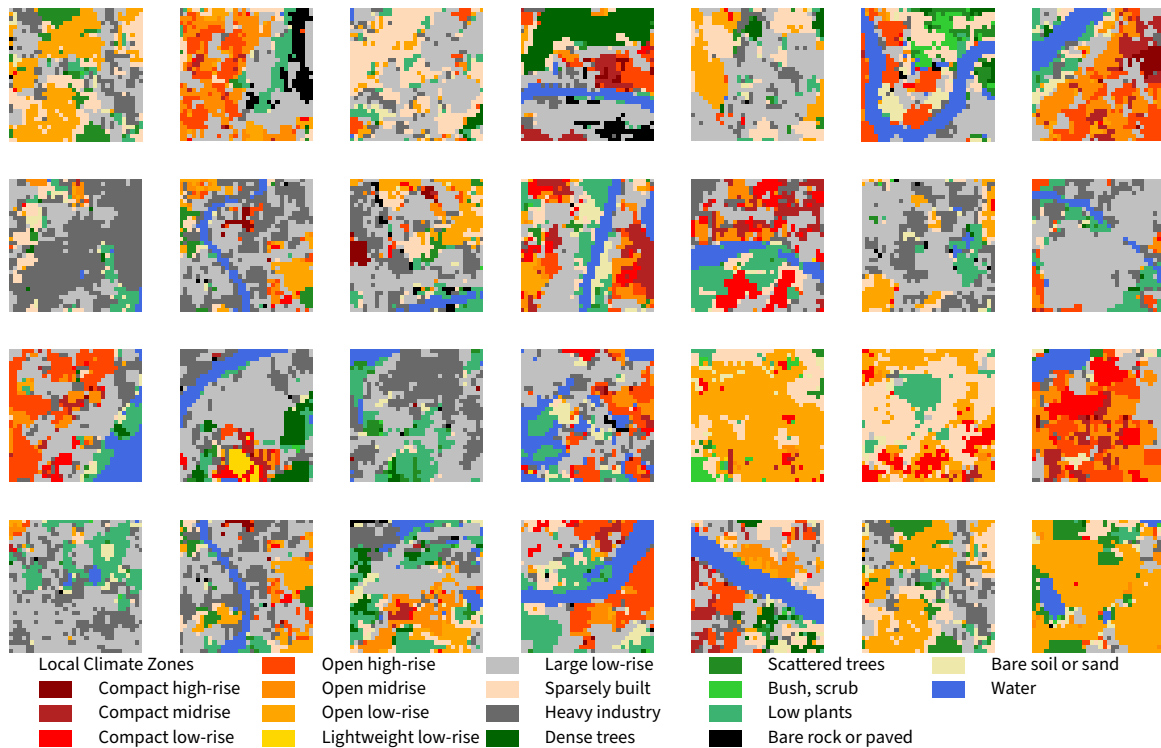
**Figure B.29** An overview of a random sample of extracted patches in  $c_{UB_{31}}$  with a total of 2301 patches.



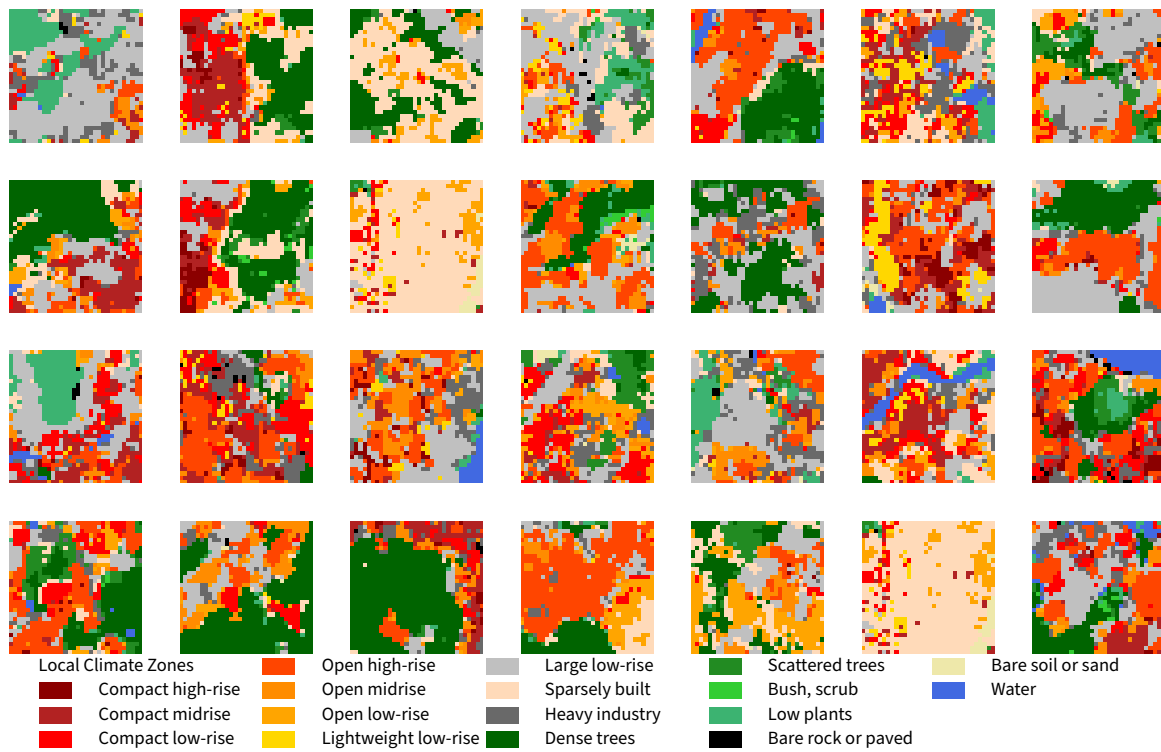
**Figure B.30** An overview of a random sample of extracted patches in  $c_{UB_{32}}$  with a total of 3634 patches.



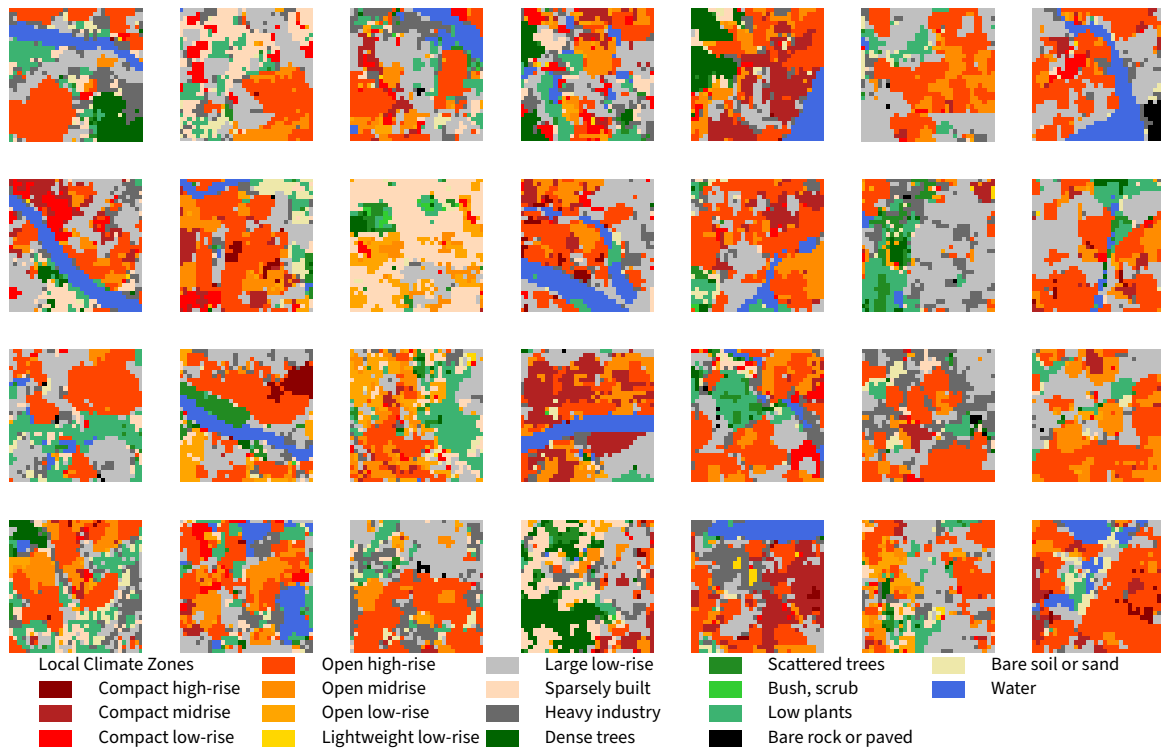
**Figure B.31** An overview of a random sample of extracted patches in  $c_{UB_{33}}$  with a total of 9335 patches.



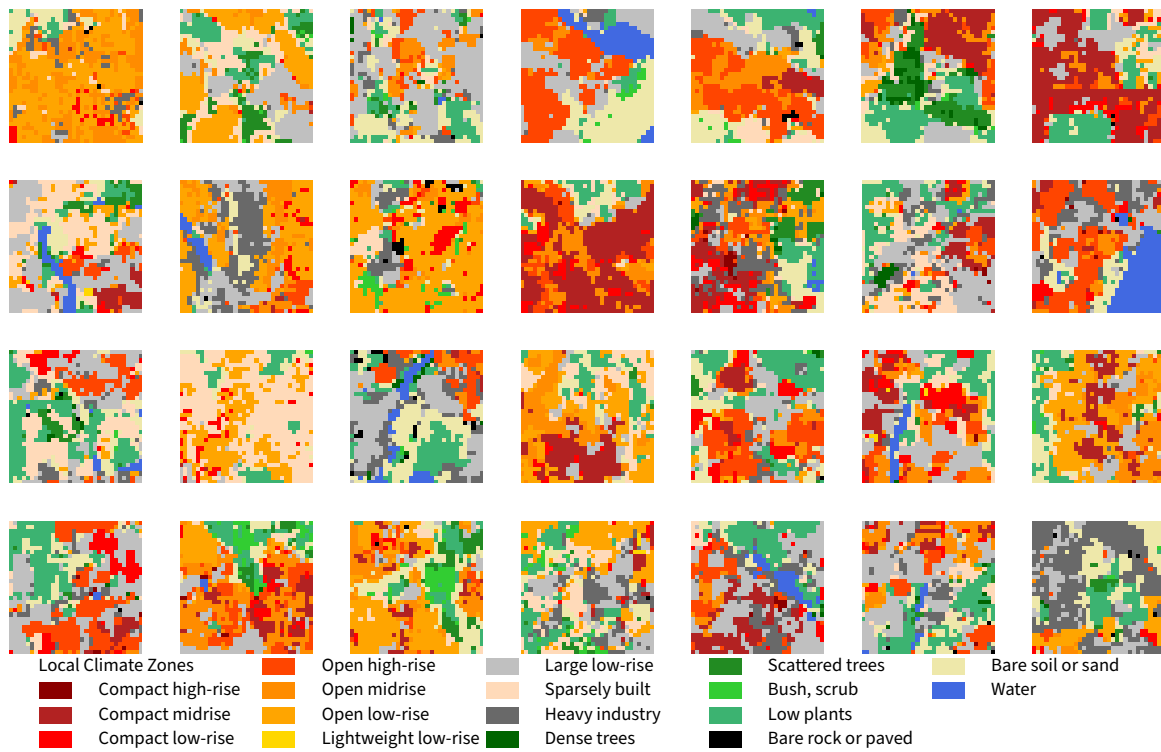
**Figure B.32** An overview of a random sample of extracted patches in  $c_{UB_{34}}$  with a total of 3647 patches.



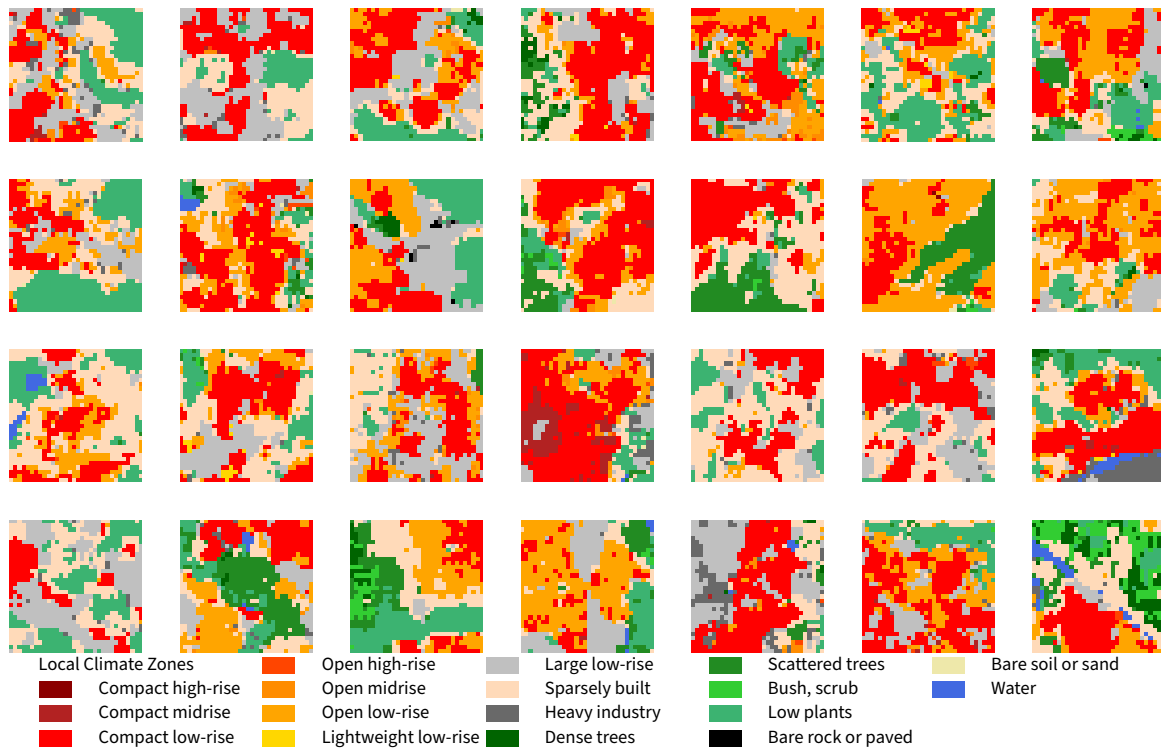
**Figure B.33** An overview of a random sample of extracted patches in  $c_{UB_{35}}$  with a total of 3109 patches.



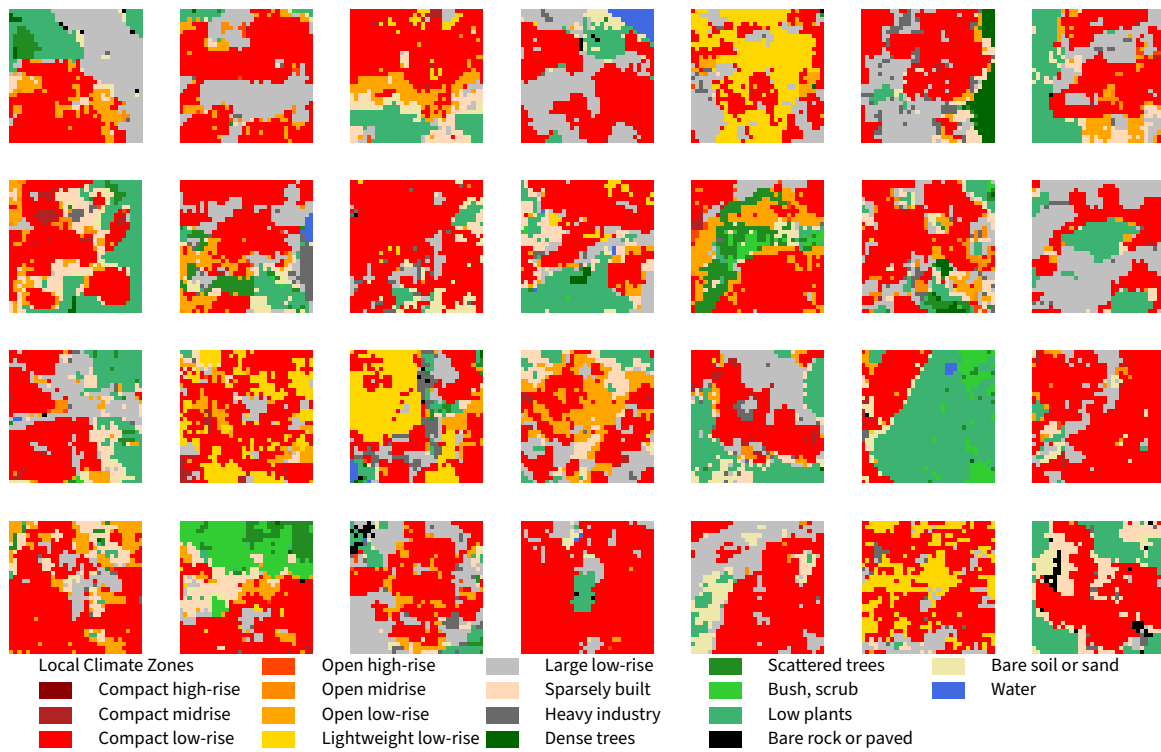
**Figure B.34** An overview of a random sample of extracted patches in  $c_{UB_{36}}$  with a total of 4493 patches.



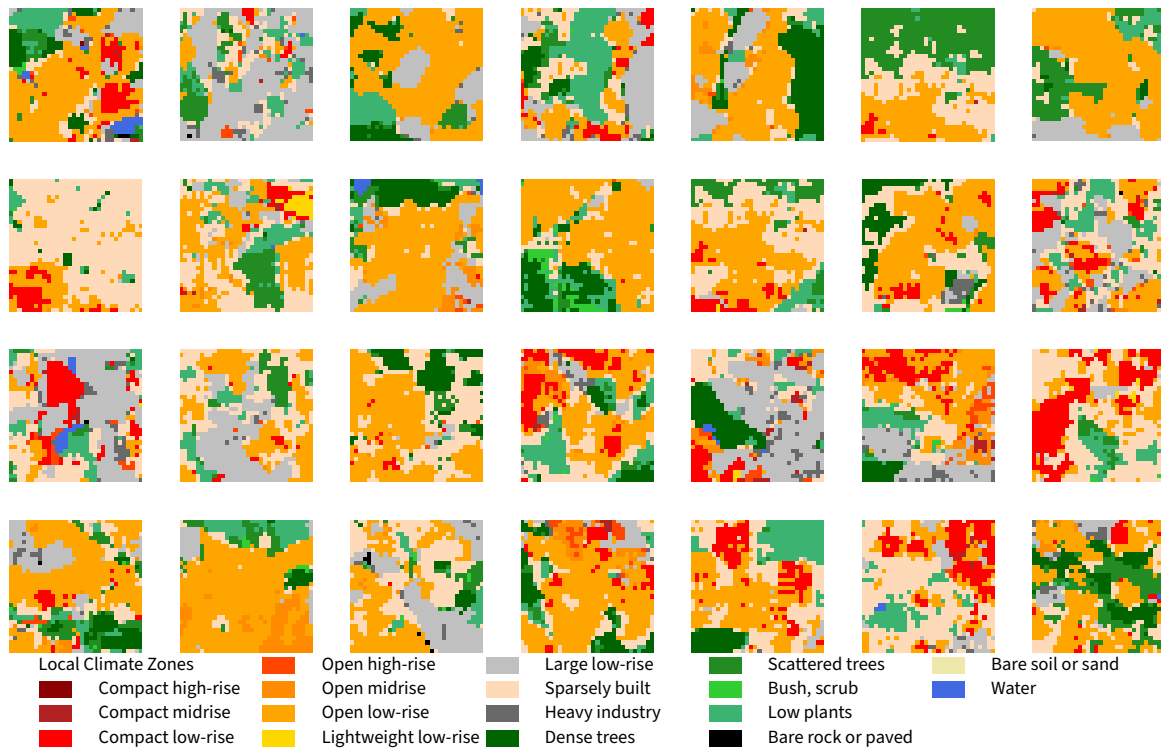
**Figure B.35** An overview of a random sample of extracted patches in  $c_{UB_{37}}$  with a total of 3979 patches.



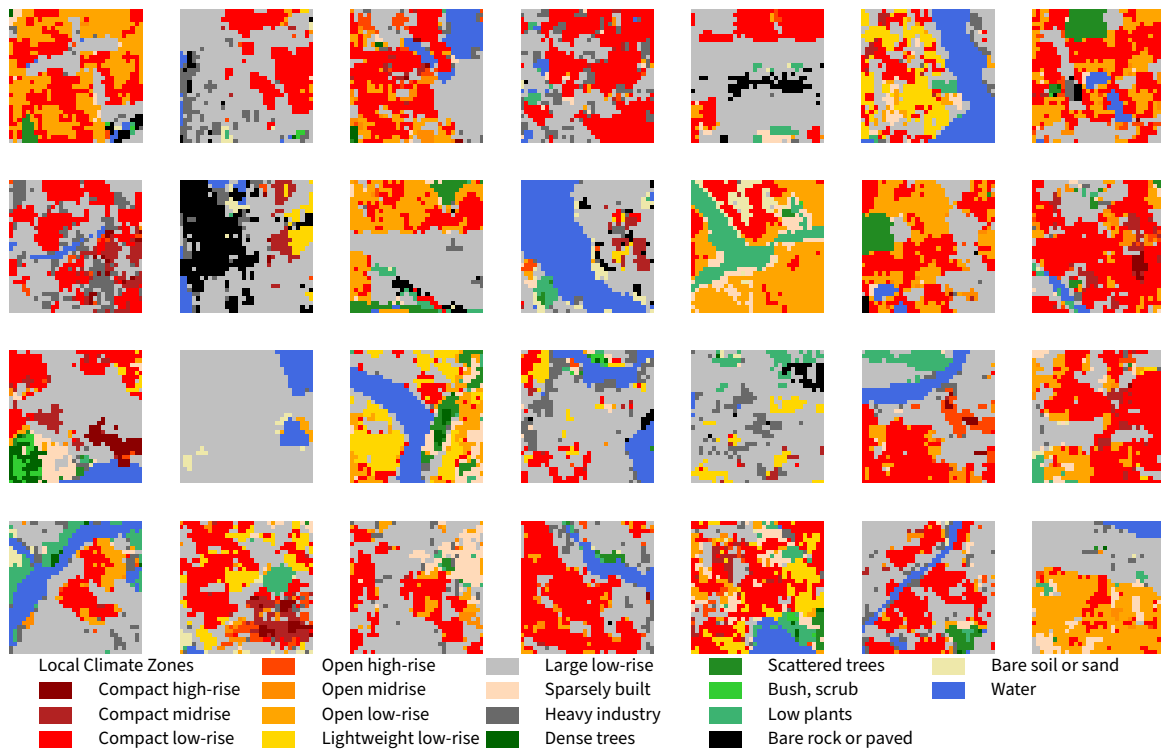
**Figure B.36** An overview of a random sample of extracted patches in  $c_{UB_{38}}$  with a total of 5074 patches.



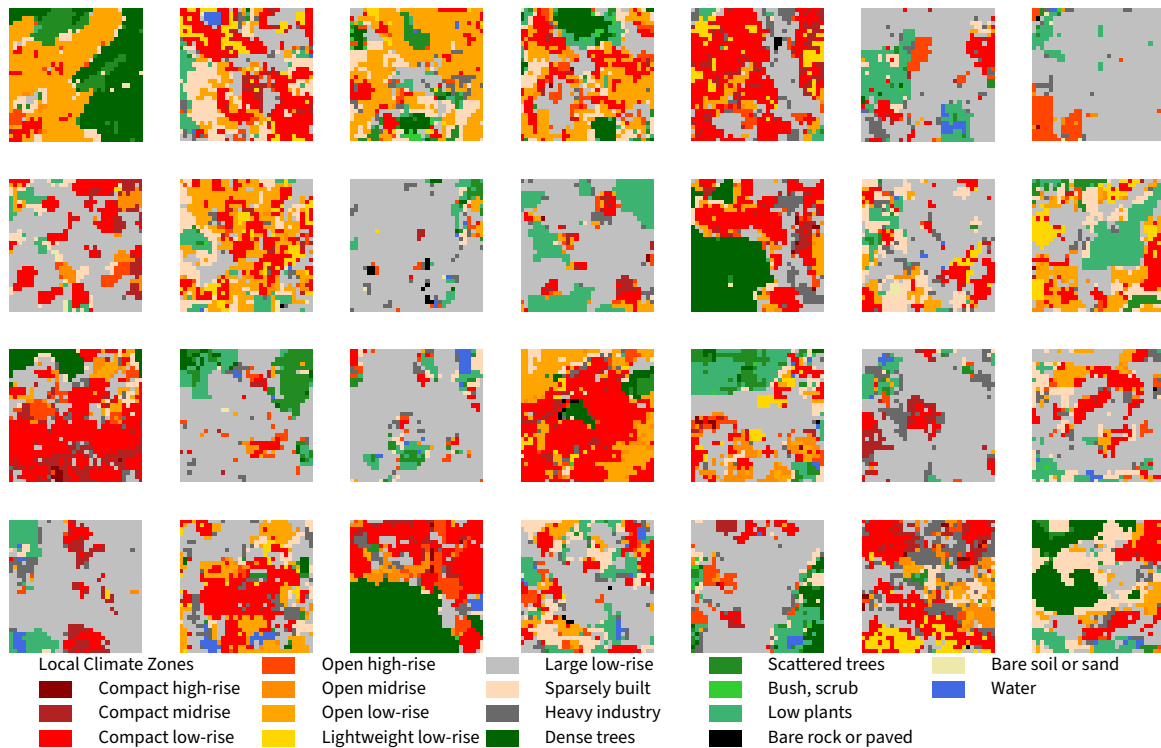
**Figure B.37** An overview of a random sample of extracted patches in  $c_{UB_{39}}$  with a total of 3149 patches.



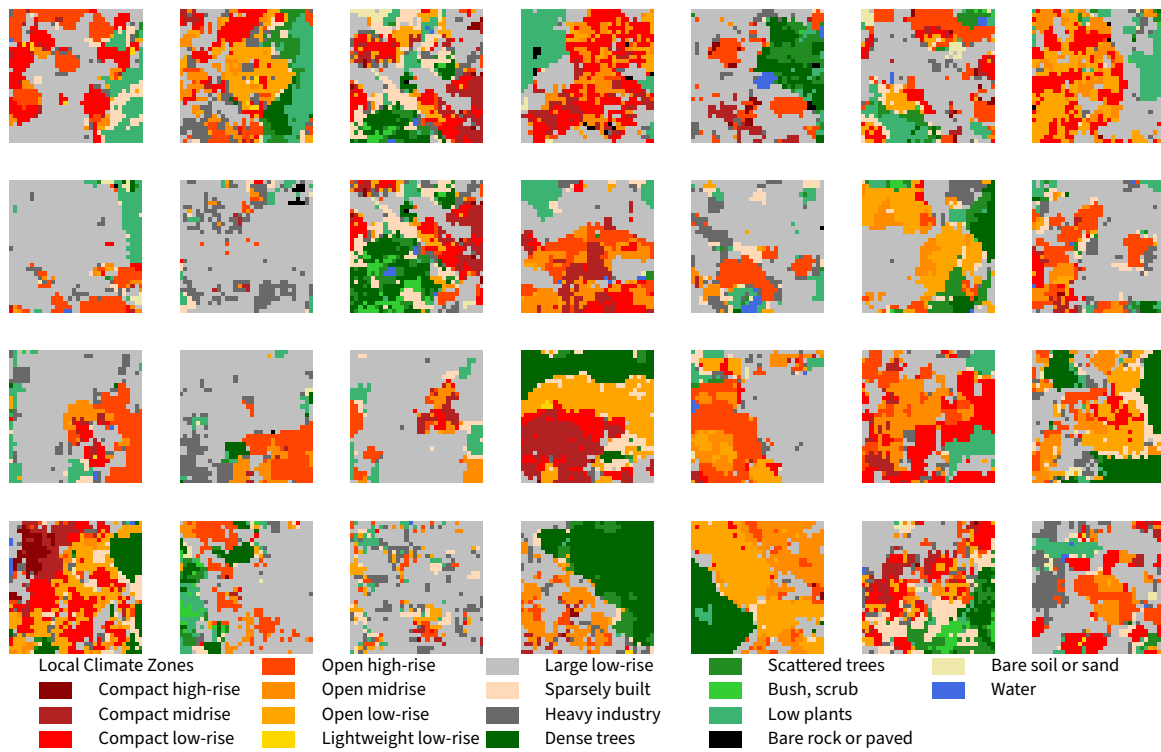
**Figure B.38** An overview of a random sample of extracted patches in  $c_{UB_{40}}$  with a total of 5247 patches.



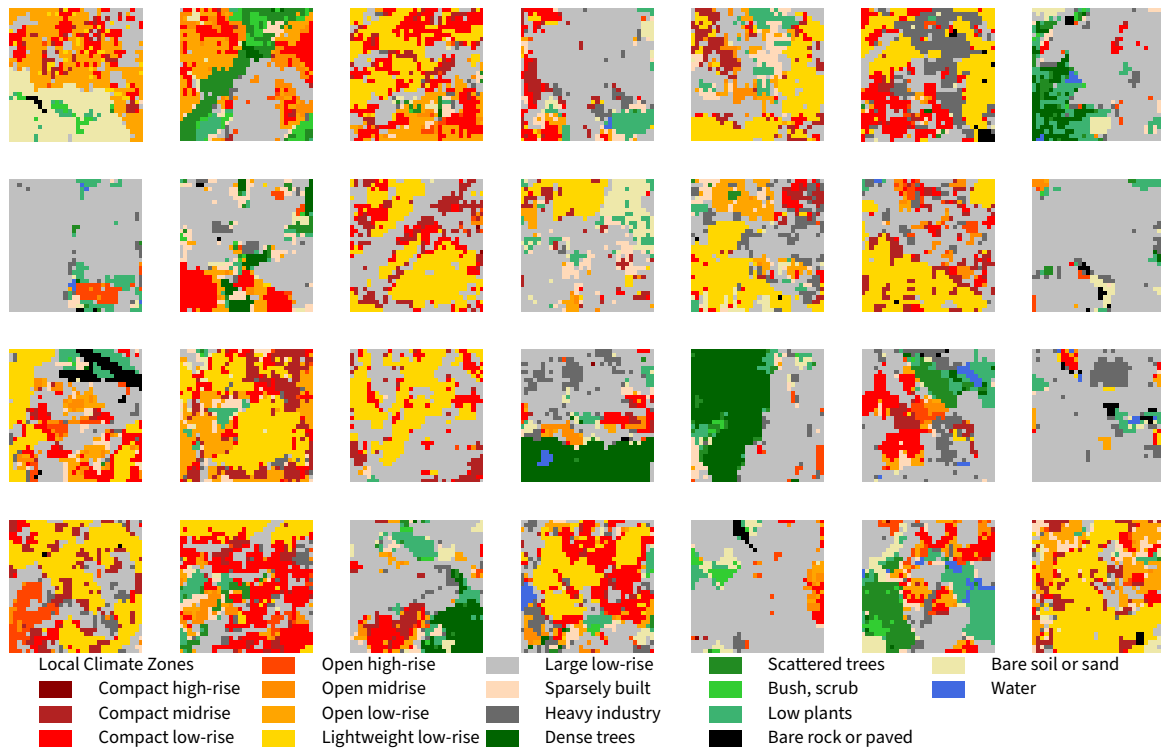
**Figure B.39** An overview of a random sample of extracted patches in  $c_{UB_{41}}$  with a total of 5181 patches.



**Figure B.40** An overview of a random sample of extracted patches in  $c_{UB_{42}}$  with a total of 4469 patches.

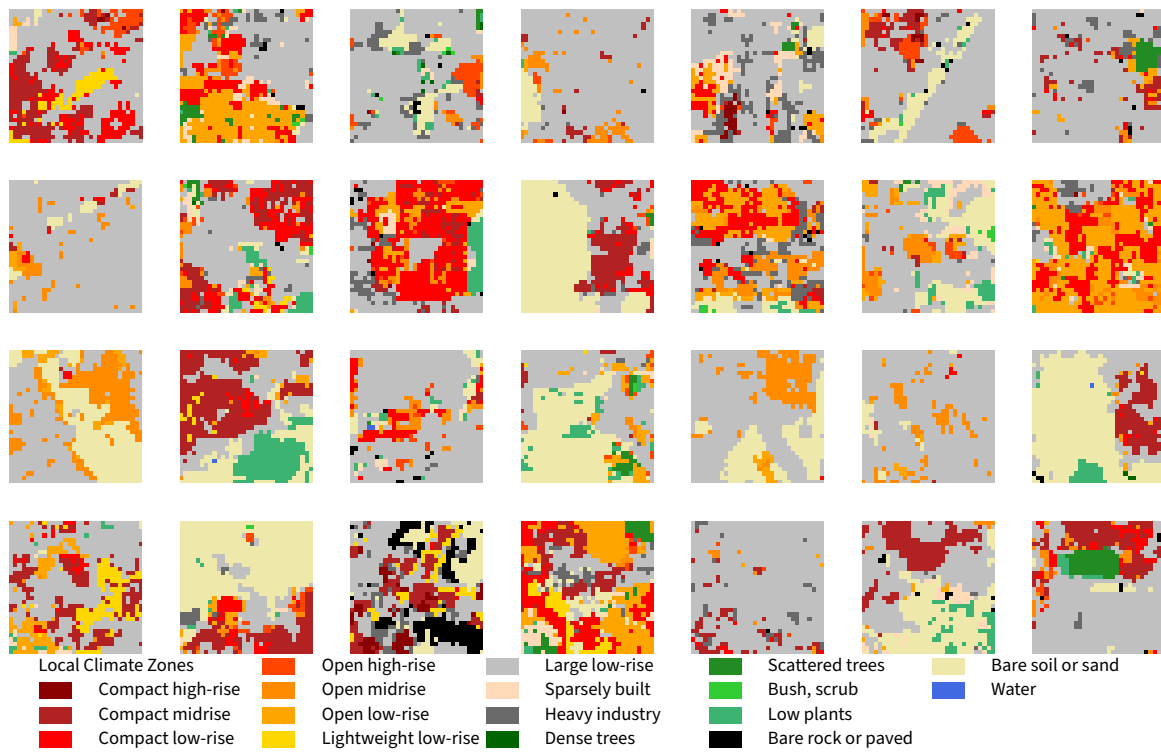


**Figure B.41** An overview of a random sample of extracted patches in  $c_{UB_{43}}$  with a total of 2419 patches.

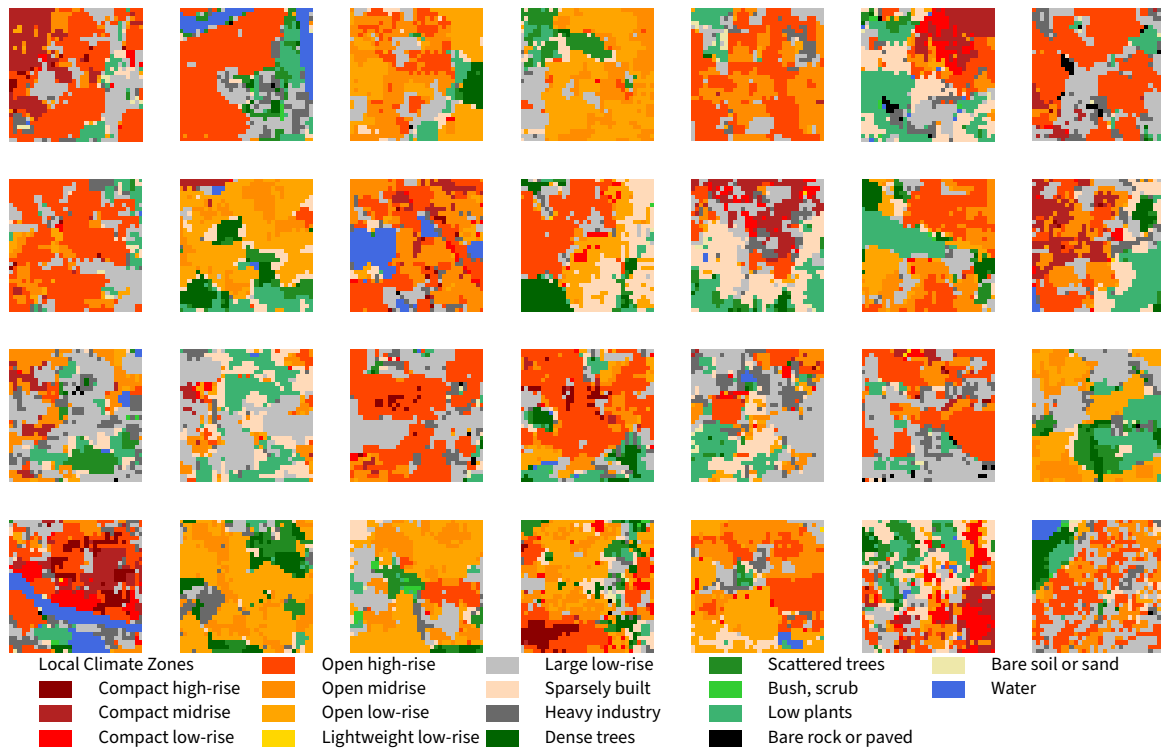


**Figure B.42** An overview of a random sample of extracted patches in  $c_{UB_{44}}$  with a total of 2260 patches.

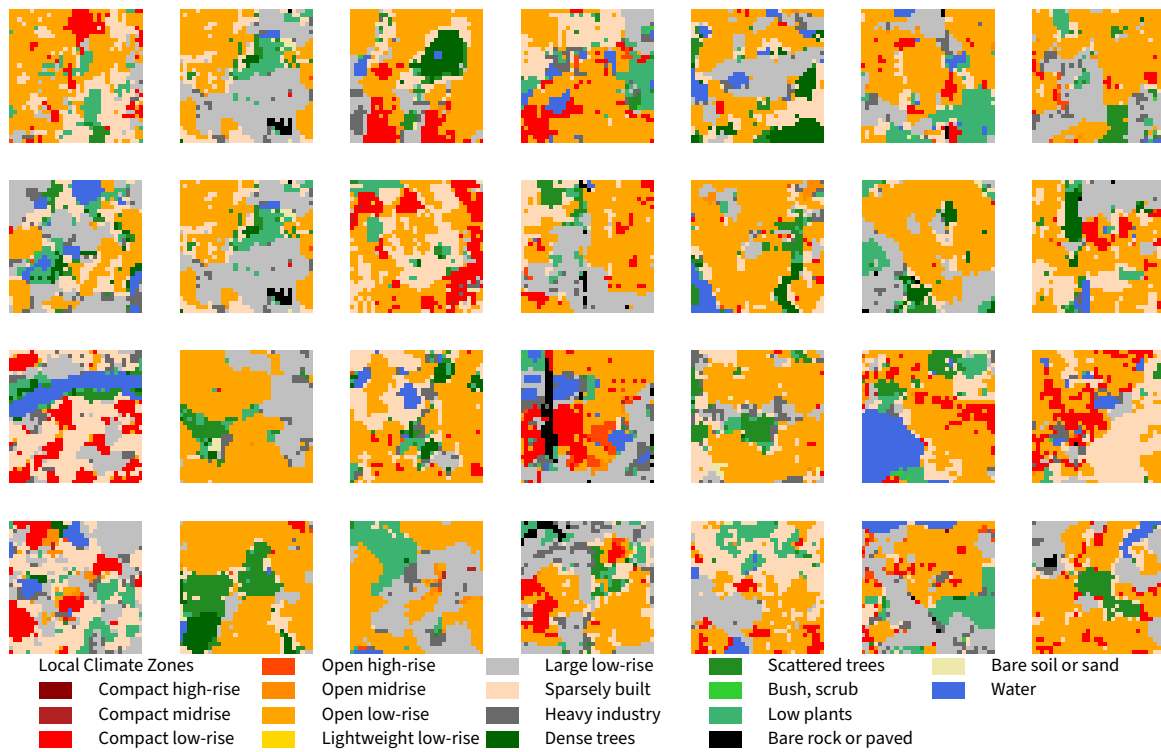




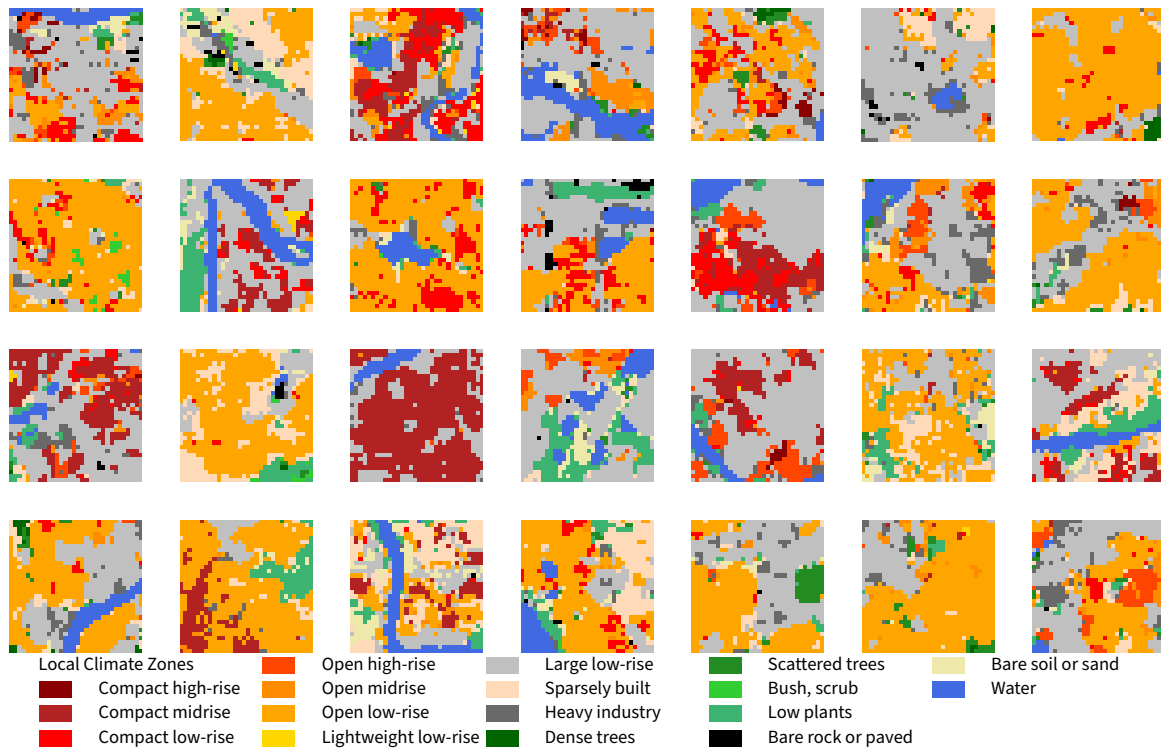
**Figure B.43** An overview of a random sample of extracted patches in  $c_{UB_{45}}$  with a total of 5064 patches.



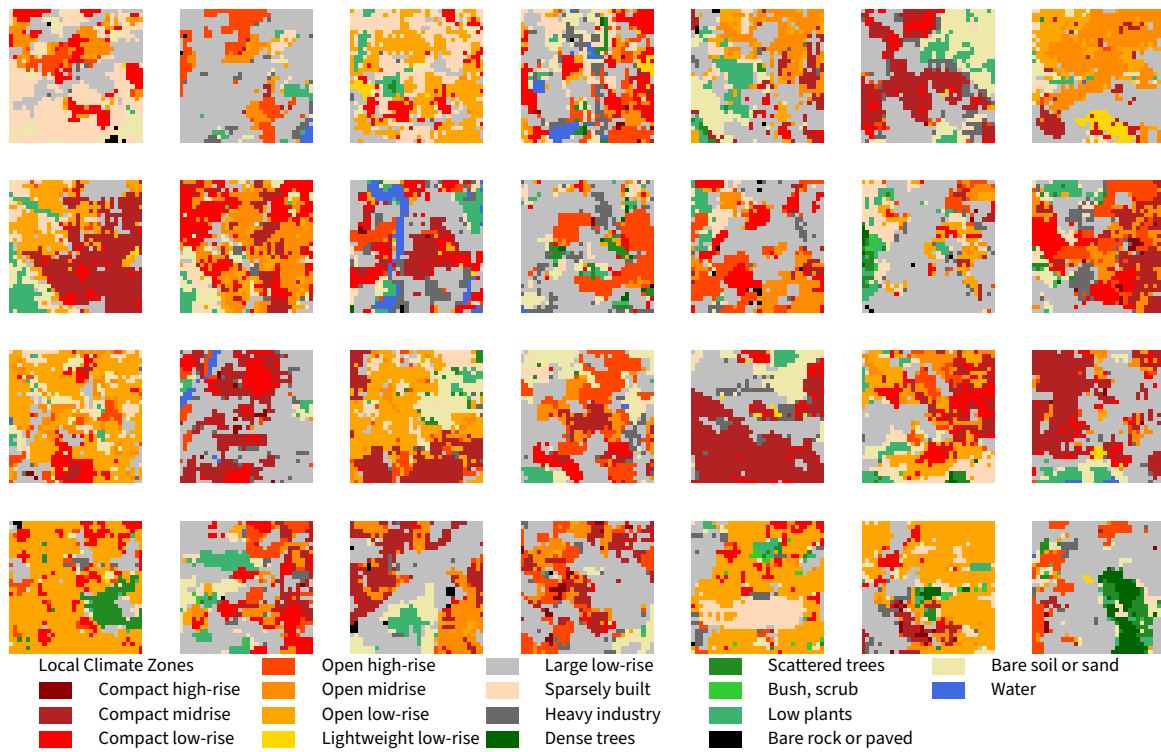
**Figure B.44** An overview of a random sample of extracted patches in  $c_{UB_{46}}$  with a total of 6580 patches.



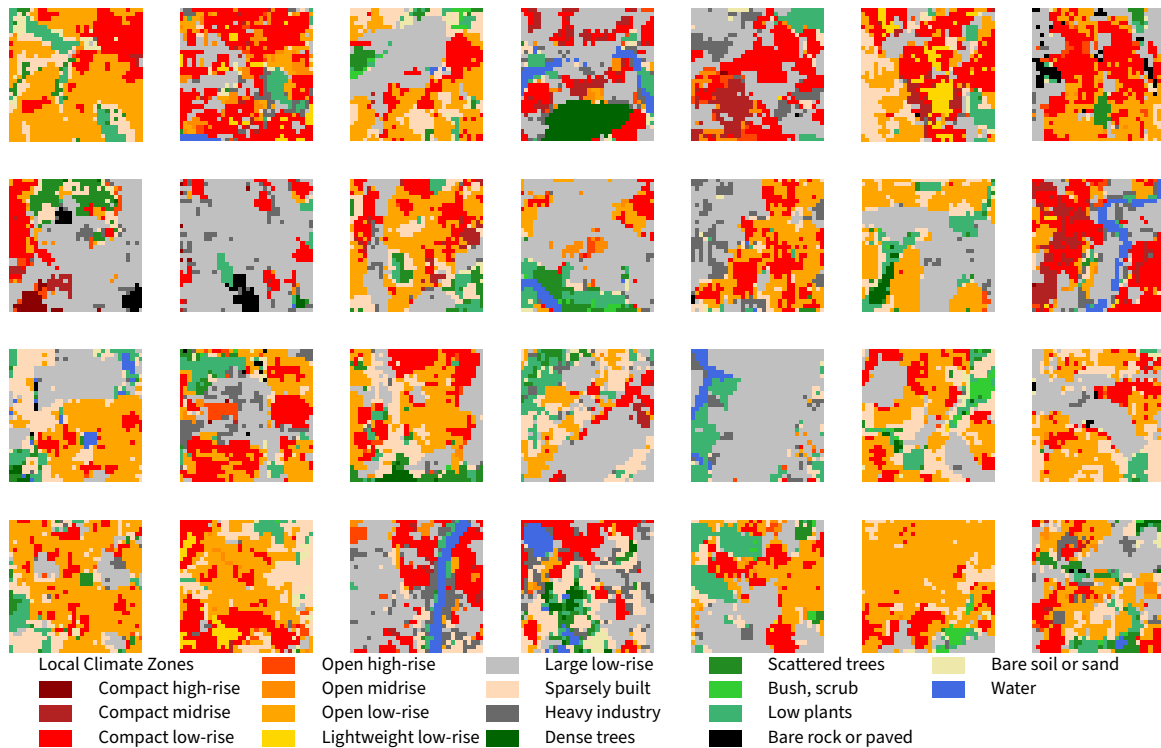
**Figure B.45** An overview of a random sample of extracted patches in  $c_{UB_{47}}$  with a total of 2500 patches.



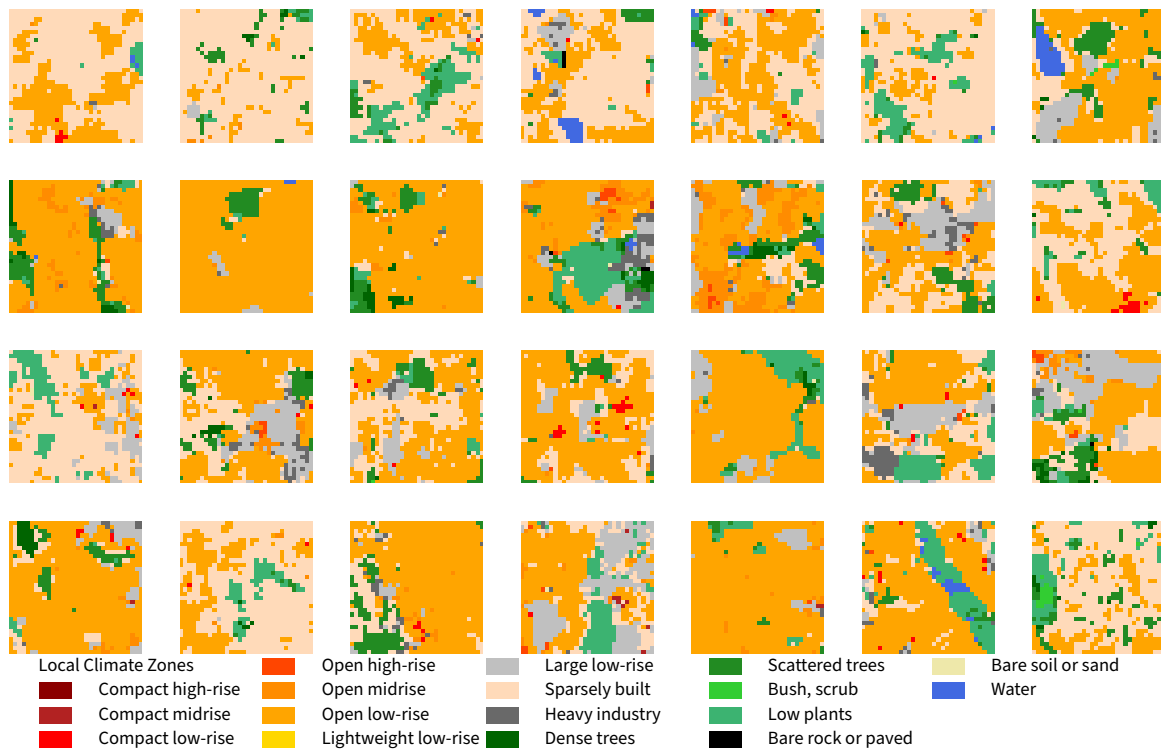
**Figure B.46** An overview of a random sample of extracted patches in  $c_{UB_{48}}$  with a total of 3928 patches.



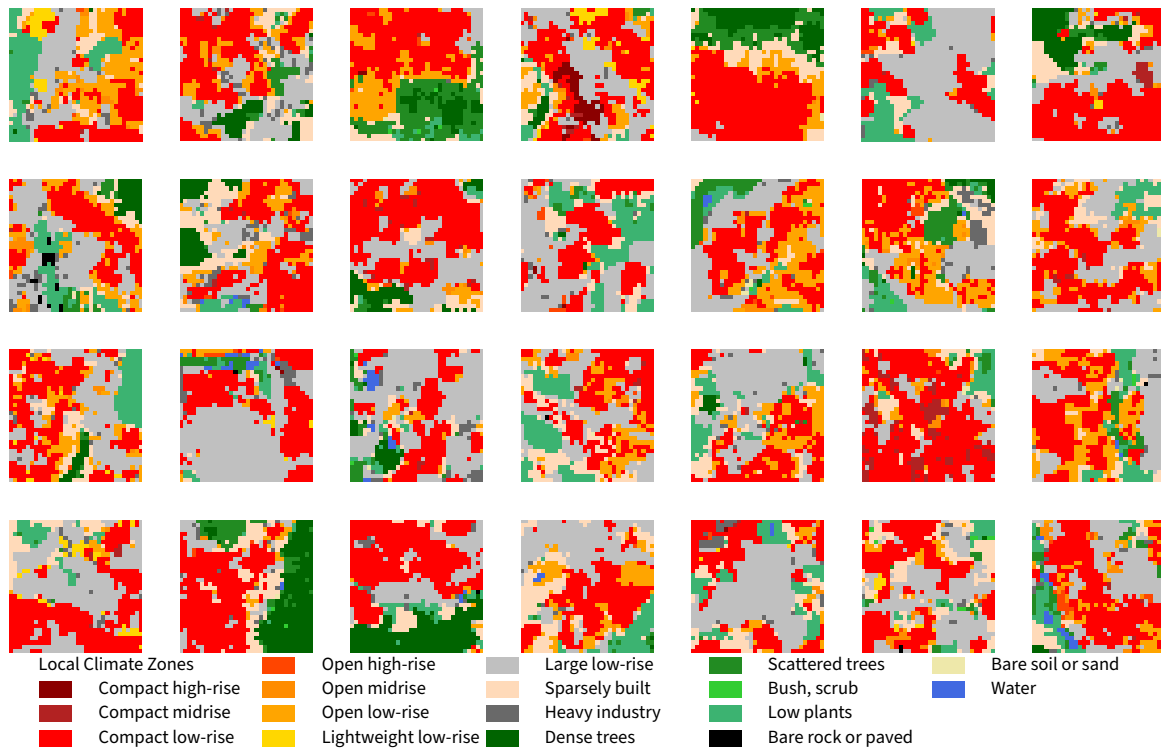
**Figure B.47** An overview of a random sample of extracted patches in  $c_{UB_{49}}$  with a total of 5299 patches.



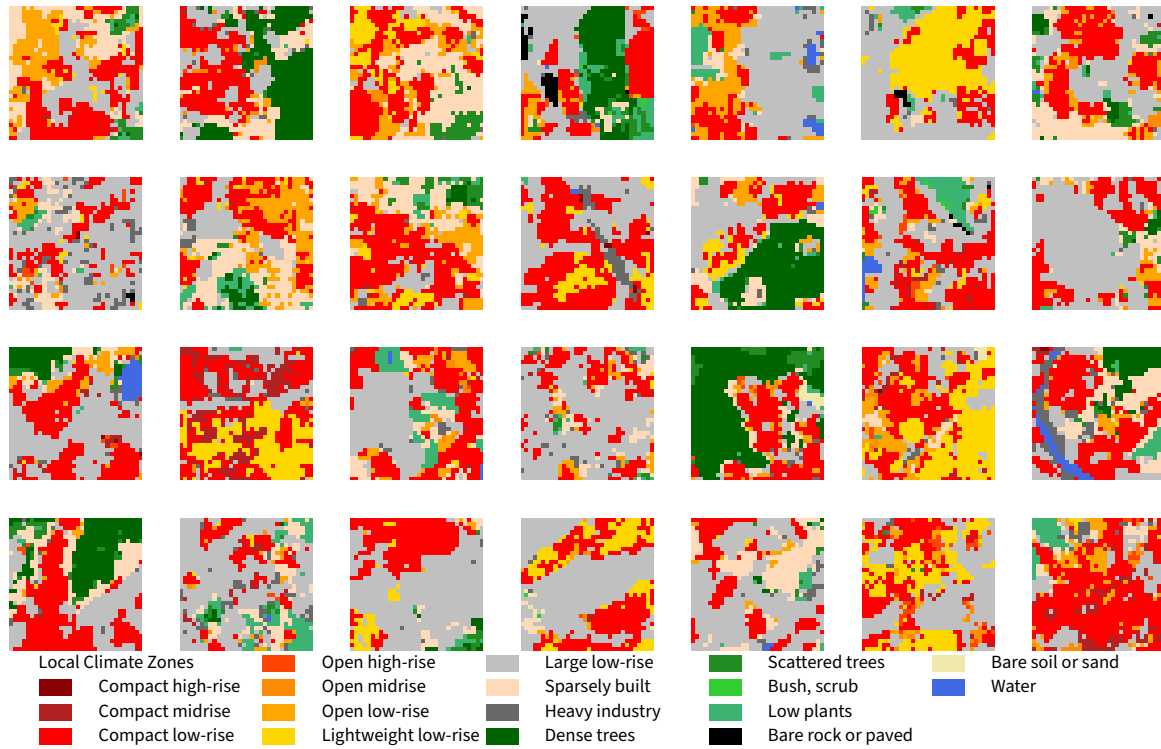
**Figure B.48** An overview of a random sample of extracted patches in  $c_{UB_{50}}$  with a total of 6340 patches.



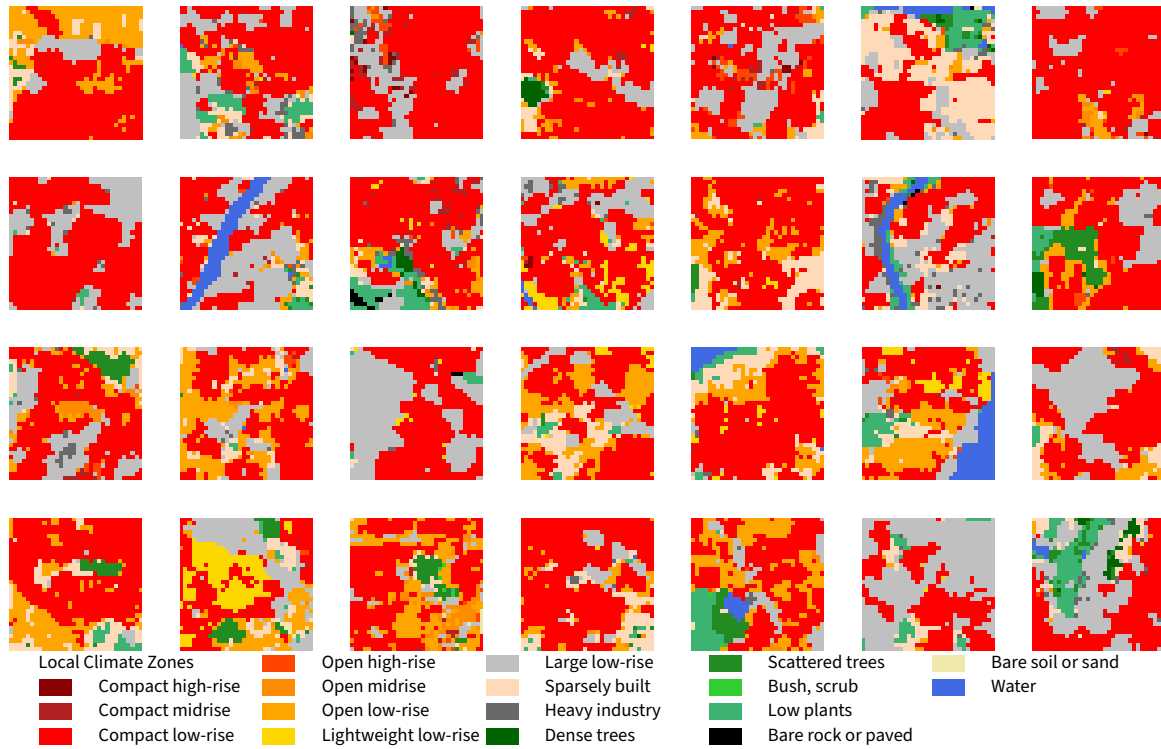
**Figure B.49** An overview of a random sample of extracted patches in  $c_{UB_{51}}$  with a total of 5385 patches.



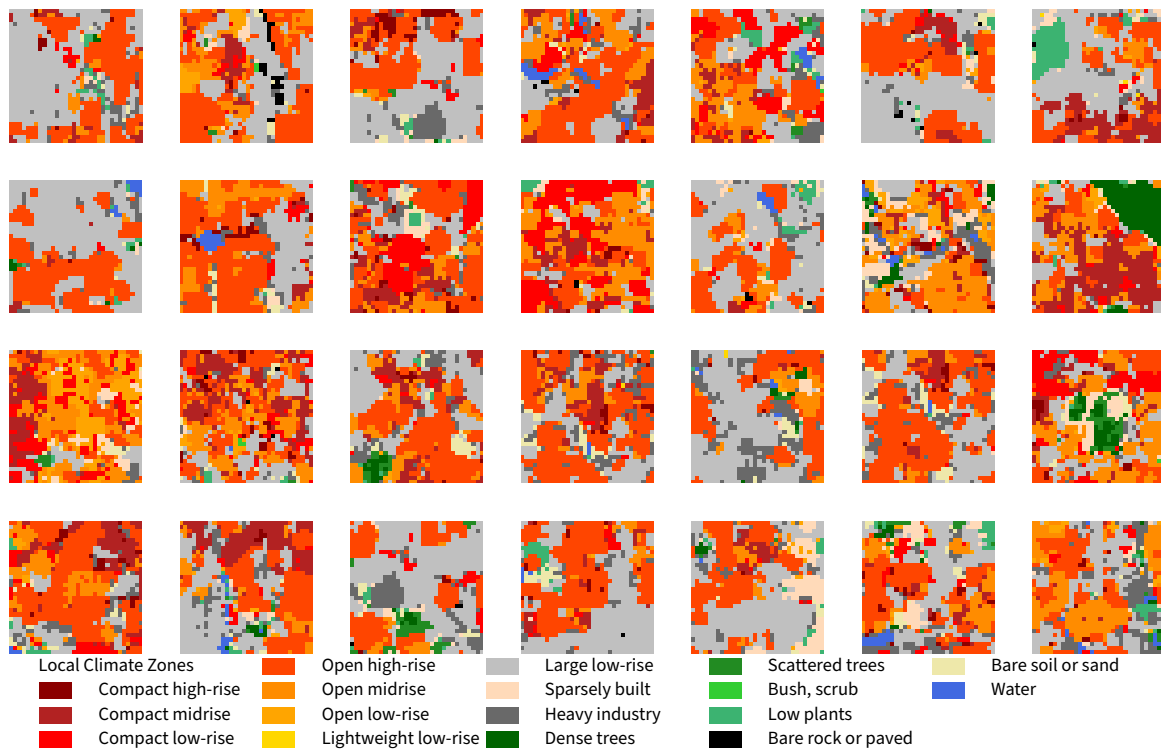
**Figure B.50** An overview of a random sample of extracted patches in  $c_{UB_{52}}$  with a total of 1098 patches.



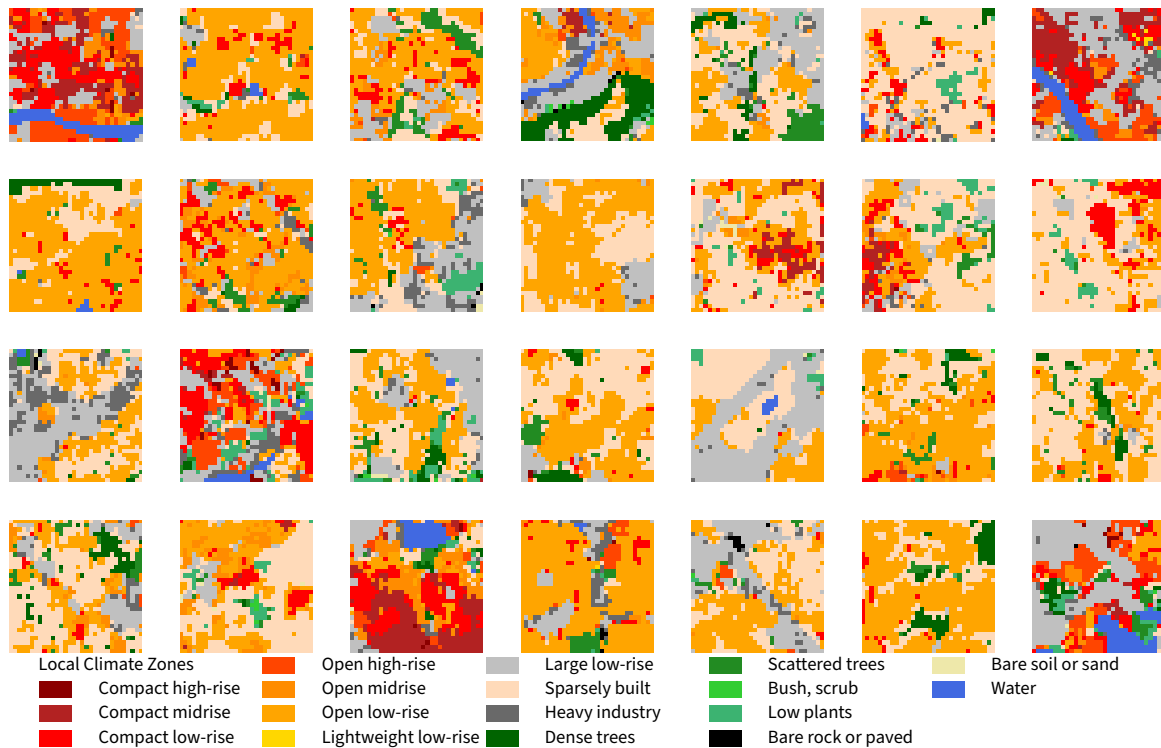
**Figure B.51** An overview of a random sample of extracted patches in  $c_{UB_{53}}$  with a total of 5388 patches.



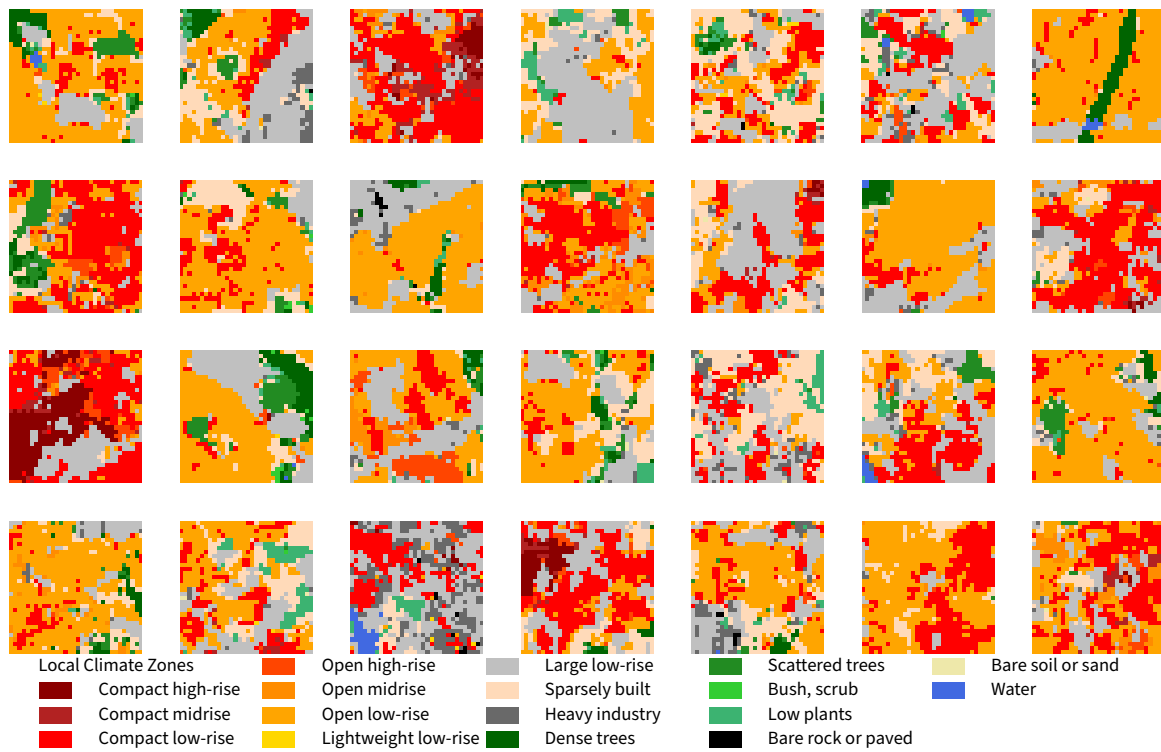
**Figure B.52** An overview of a random sample of extracted patches in  $c_{UB_{54}}$  with a total of 13906 patches.



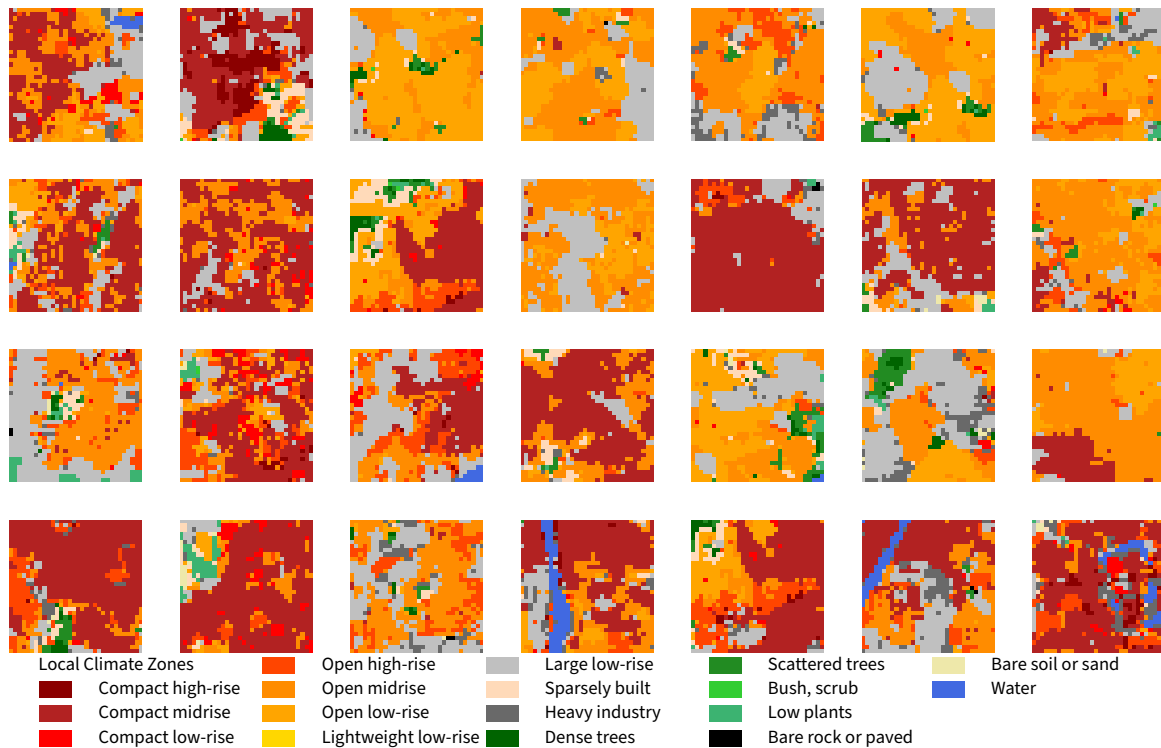
**Figure B.53** An overview of a random sample of extracted patches in  $c_{UB_{55}}$  with a total of 2004 patches.



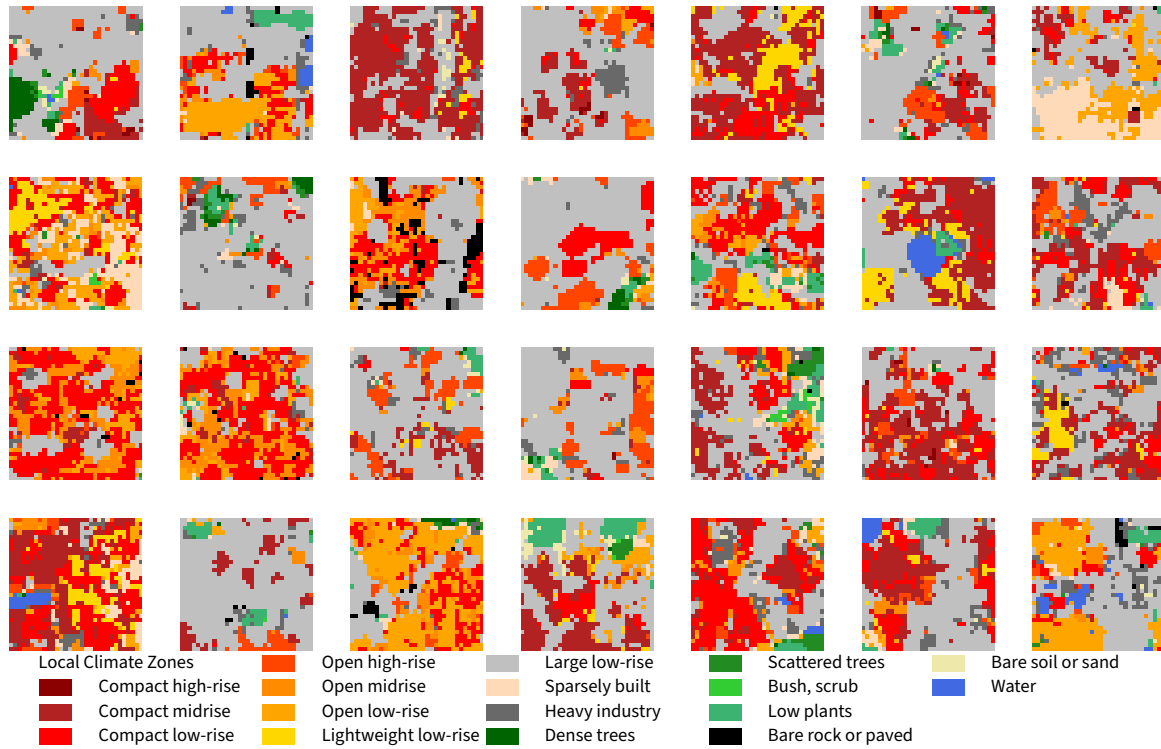
**Figure B.54** An overview of a random sample of extracted patches in  $c_{UB_{56}}$  with a total of 8518 patches.



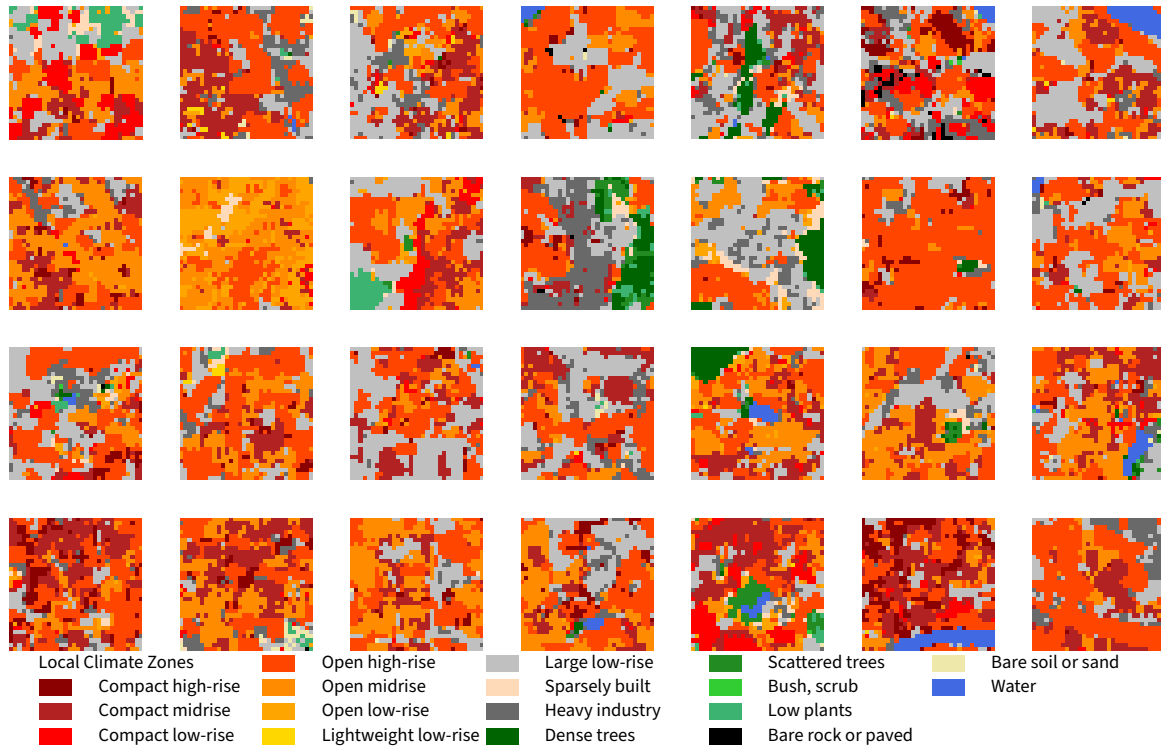
**Figure B.55** An overview of a random sample of extracted patches in  $c_{UB_{57}}$  with a total of 8158 patches.



**Figure B.56** An overview of a random sample of extracted patches in  $c_{UB_{58}}$  with a total of 7023 patches.

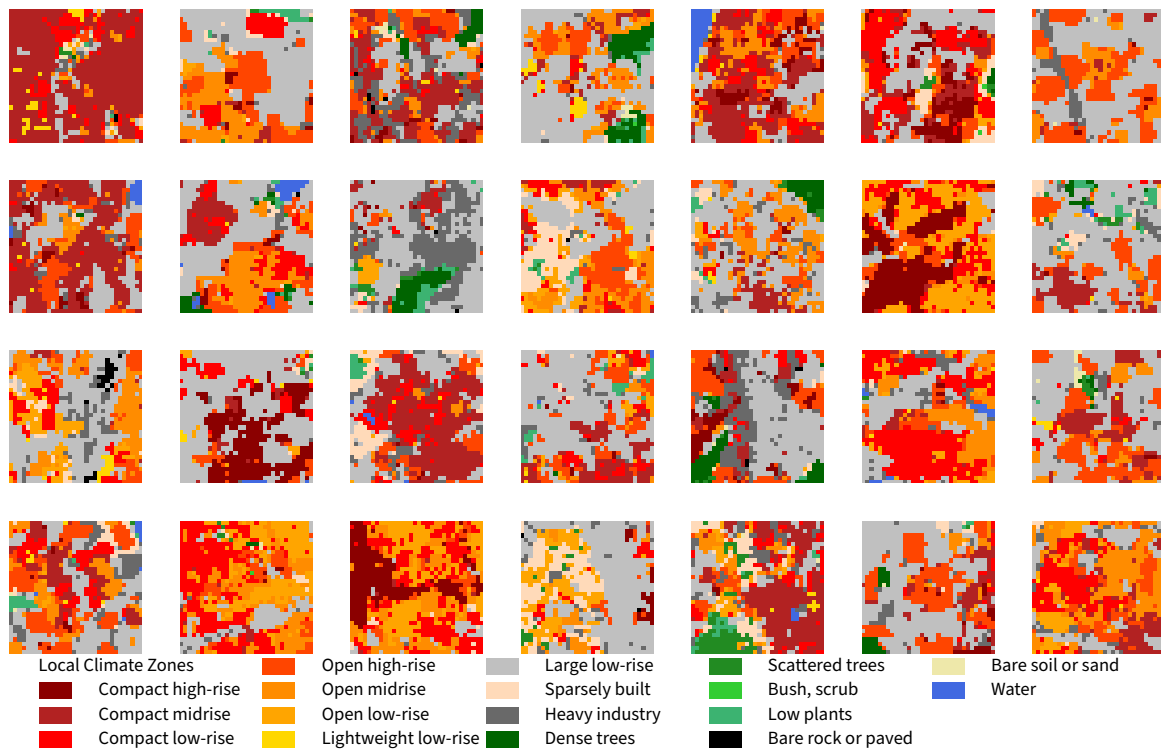


**Figure B.57** An overview of a random sample of extracted patches in  $c_{UB_{59}}$  with a total of 4709 patches.

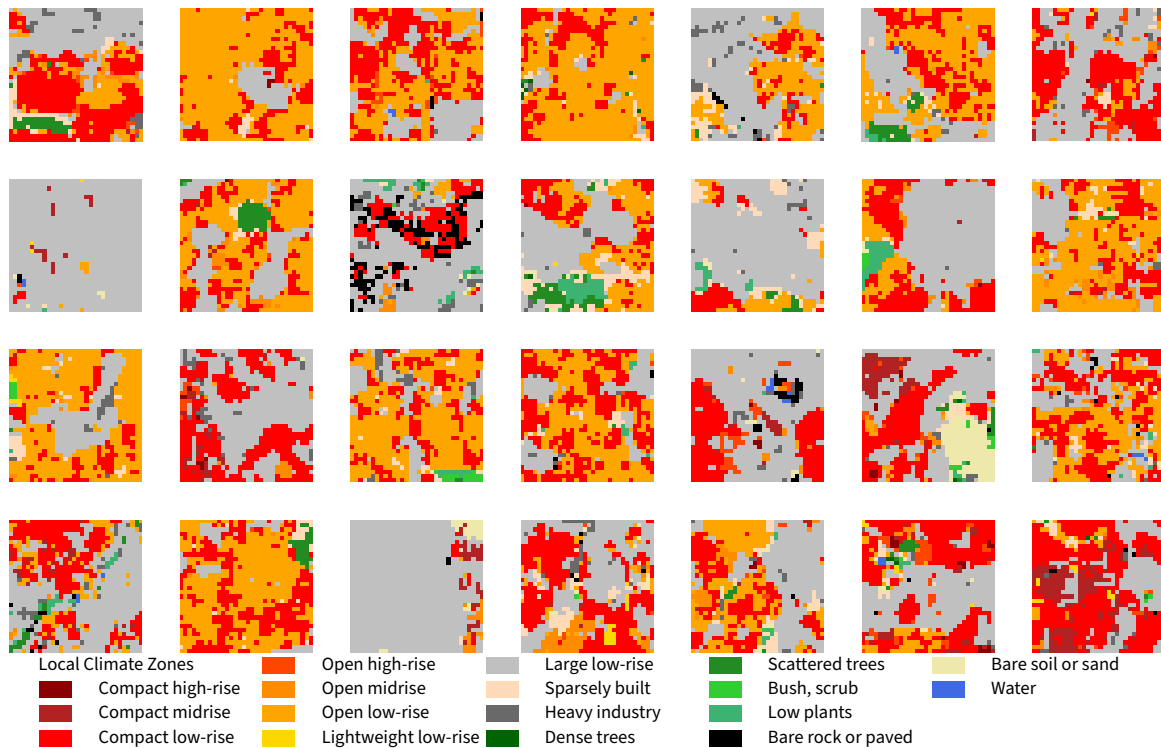


**Figure B.58** An overview of a random sample of extracted patches in  $c_{UB_{60}}$  with a total of 4630 patches.

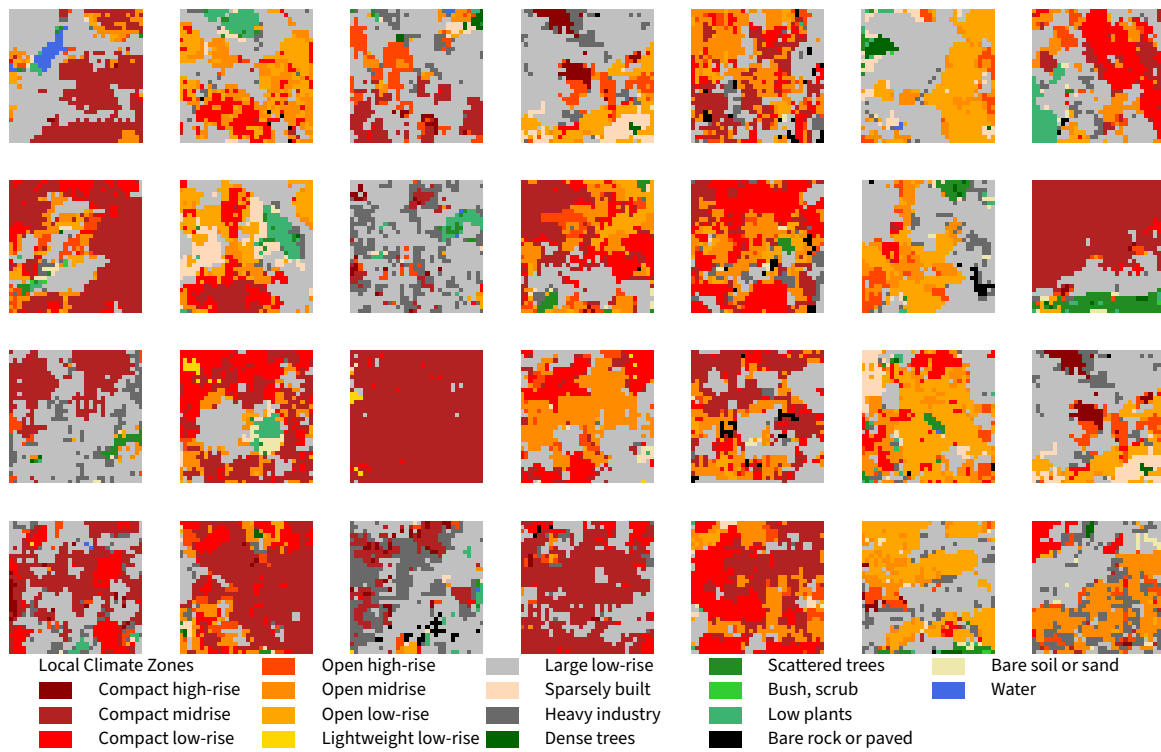




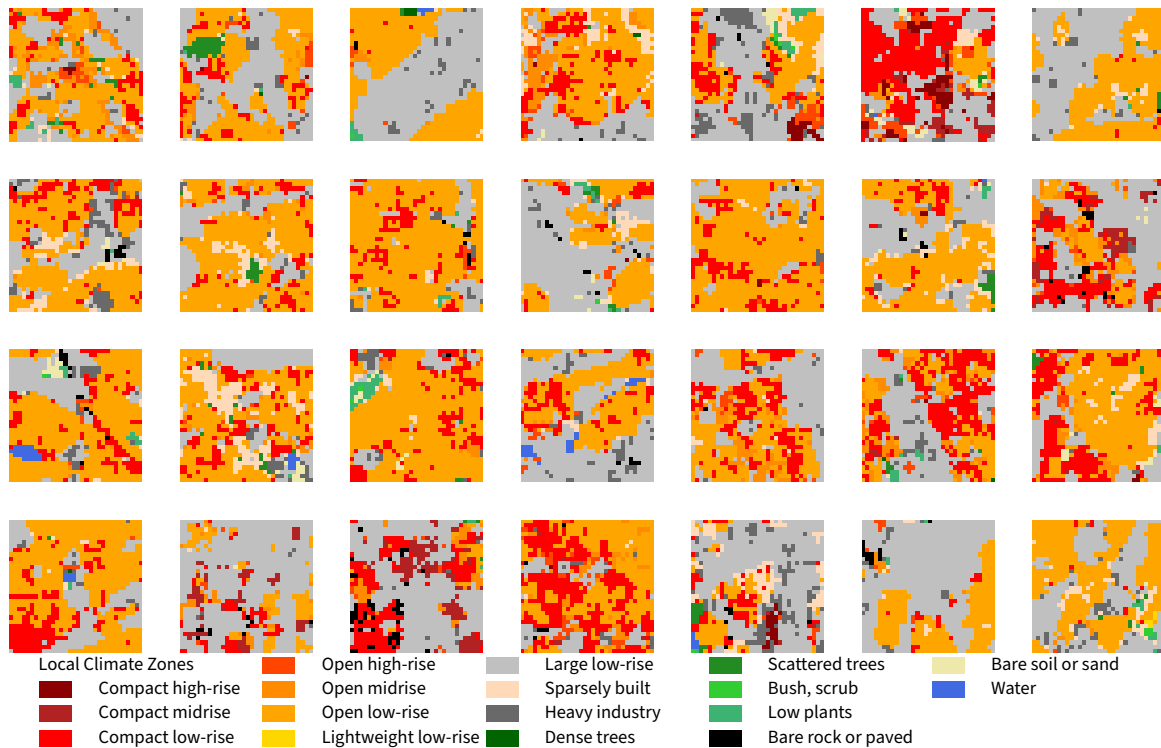
**Figure B.59** An overview of a random sample of extracted patches in  $c_{UB_{61}}$  with a total of 3193 patches.



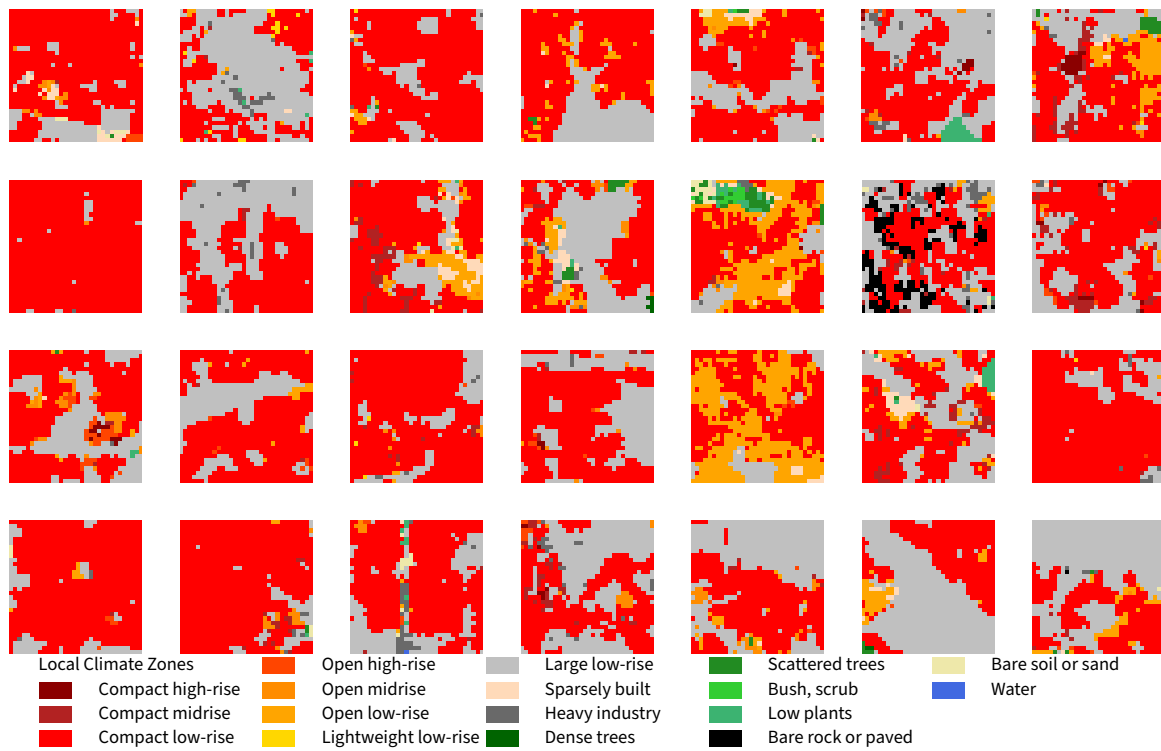
**Figure B.60** An overview of a random sample of extracted patches in  $c_{UB_{63}}$  with a total of 6054 patches.



**Figure B.61** An overview of a random sample of extracted patches in  $c_{UB64}$  with a total of 3066 patches.

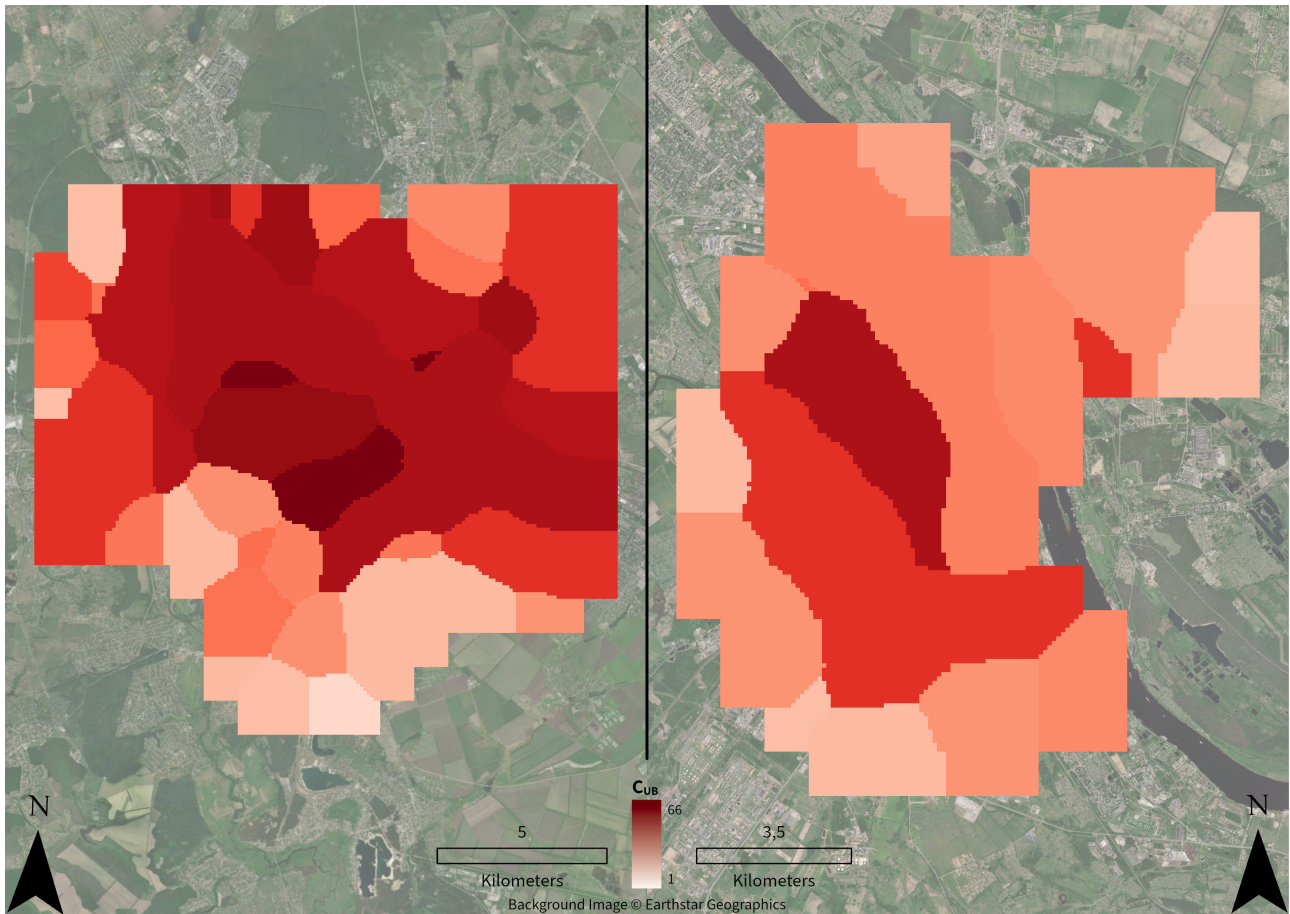


**Figure B.62** An overview of a random sample of extracted patches in  $c_{UB65}$  with a total of 2048 patches.

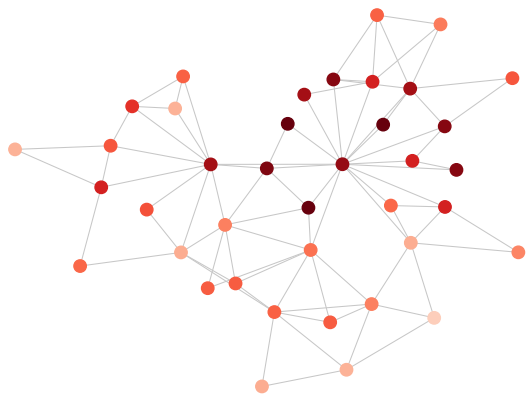


**Figure B.63** An overview of a random sample of extracted patches in  $c_{UB_6}$  with a total of 4874 patches.

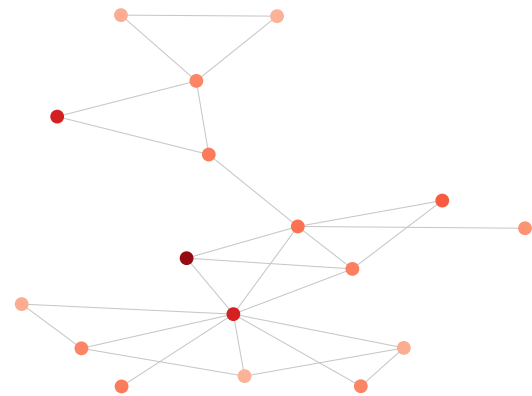
## B.2 Two Random Cities of each City Model



(a) Overview of the urban patterns of two cities within  $c_{CM_1}$

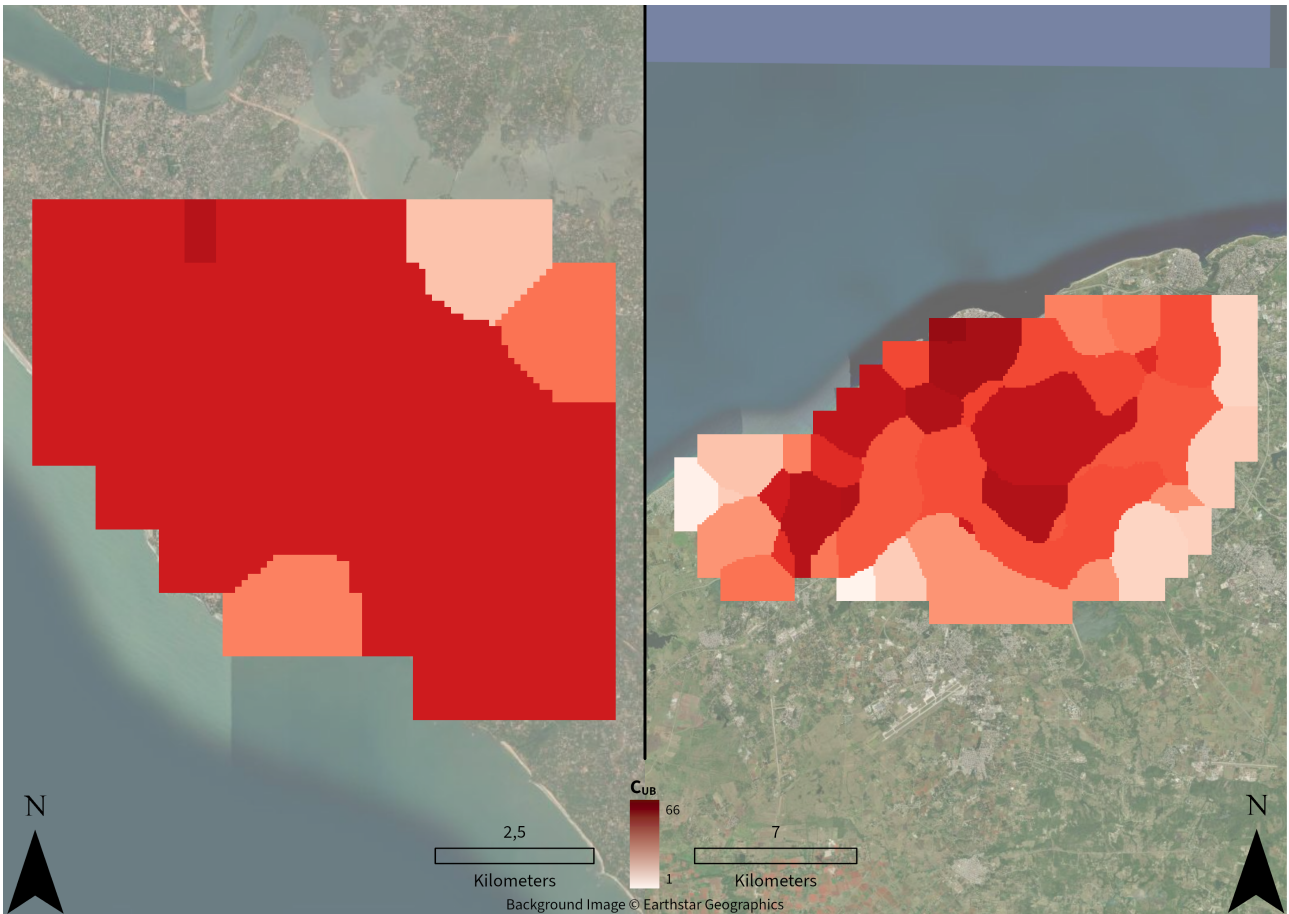


(b) Graph representation of Kharkiv

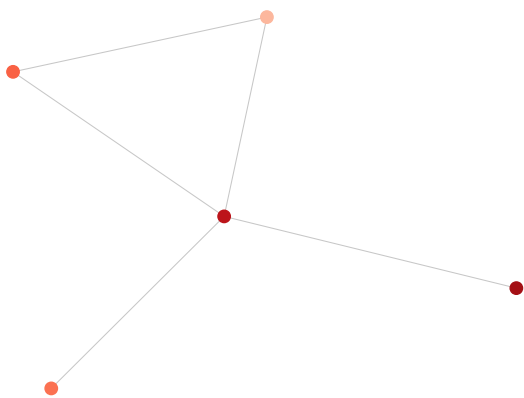


(c) Graph representation of Yaroslavl

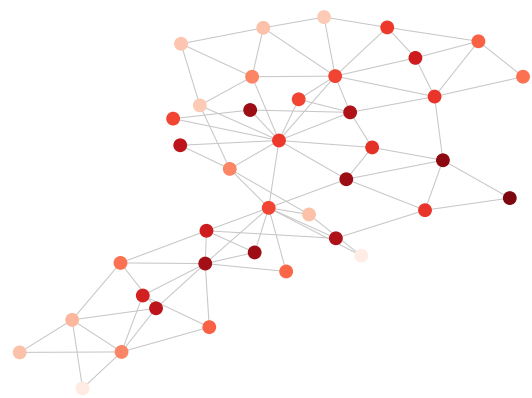
**Figure B.64** Overview of two cities within  $c_{CM_1}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_2}$

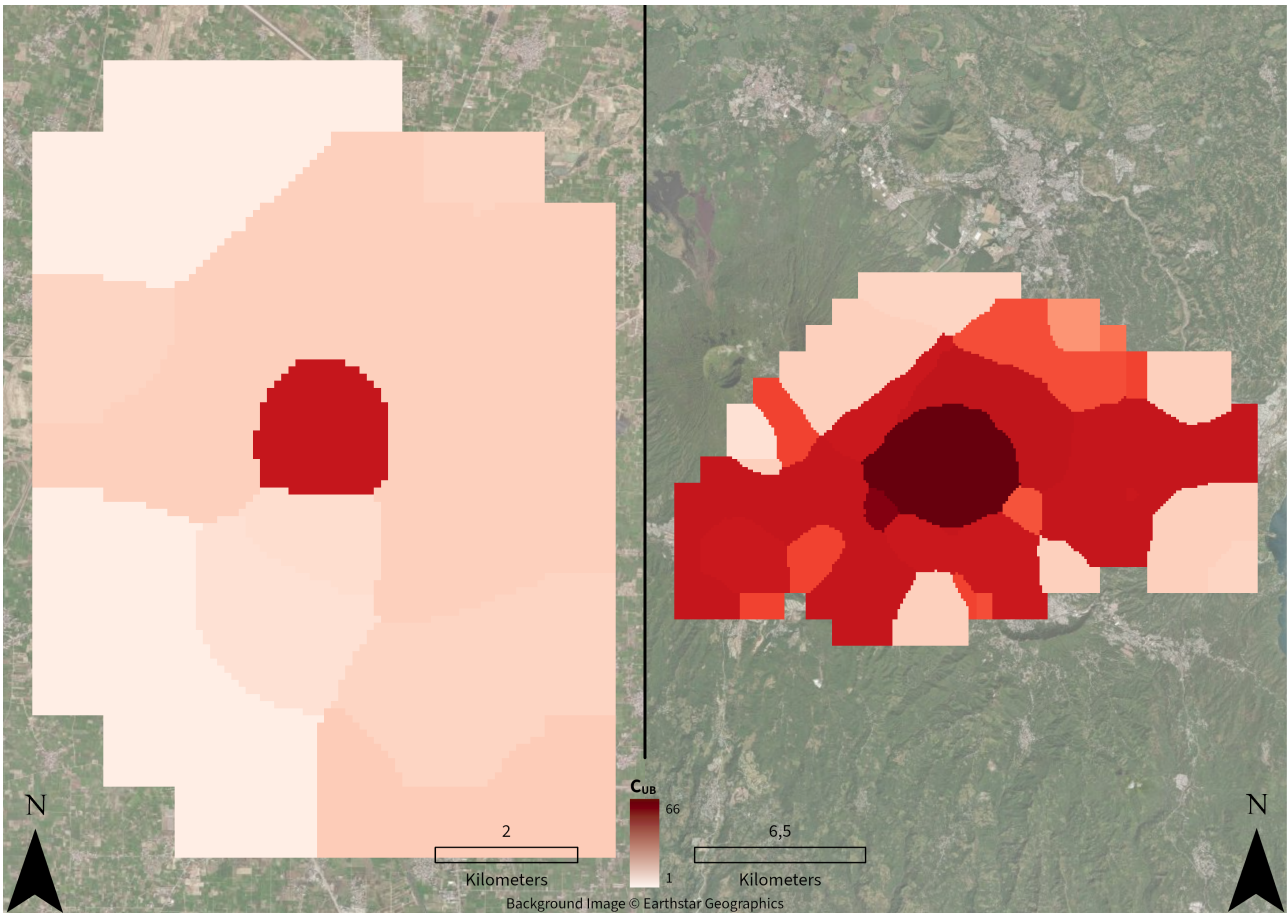


(b) Graph representation of Kannur

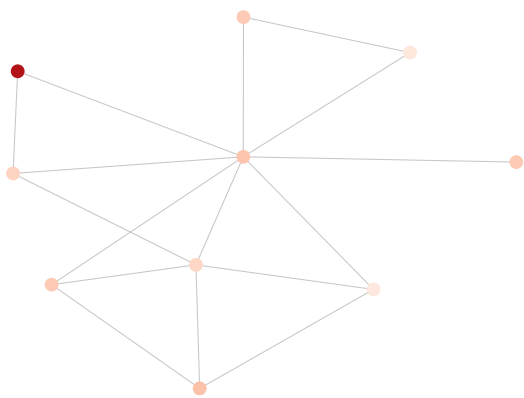


(c) Graph representation of La Habana (Havana)

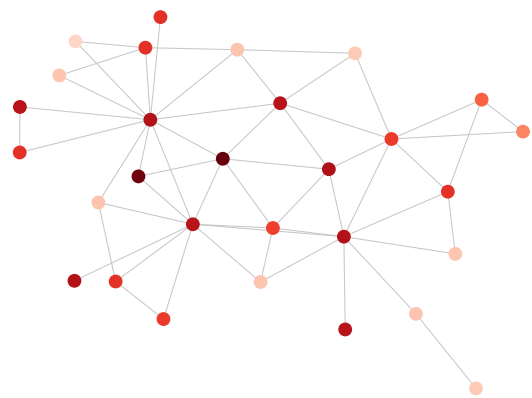
**Figure B.65** Overview of two cities within  $c_{CM_2}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_3}$

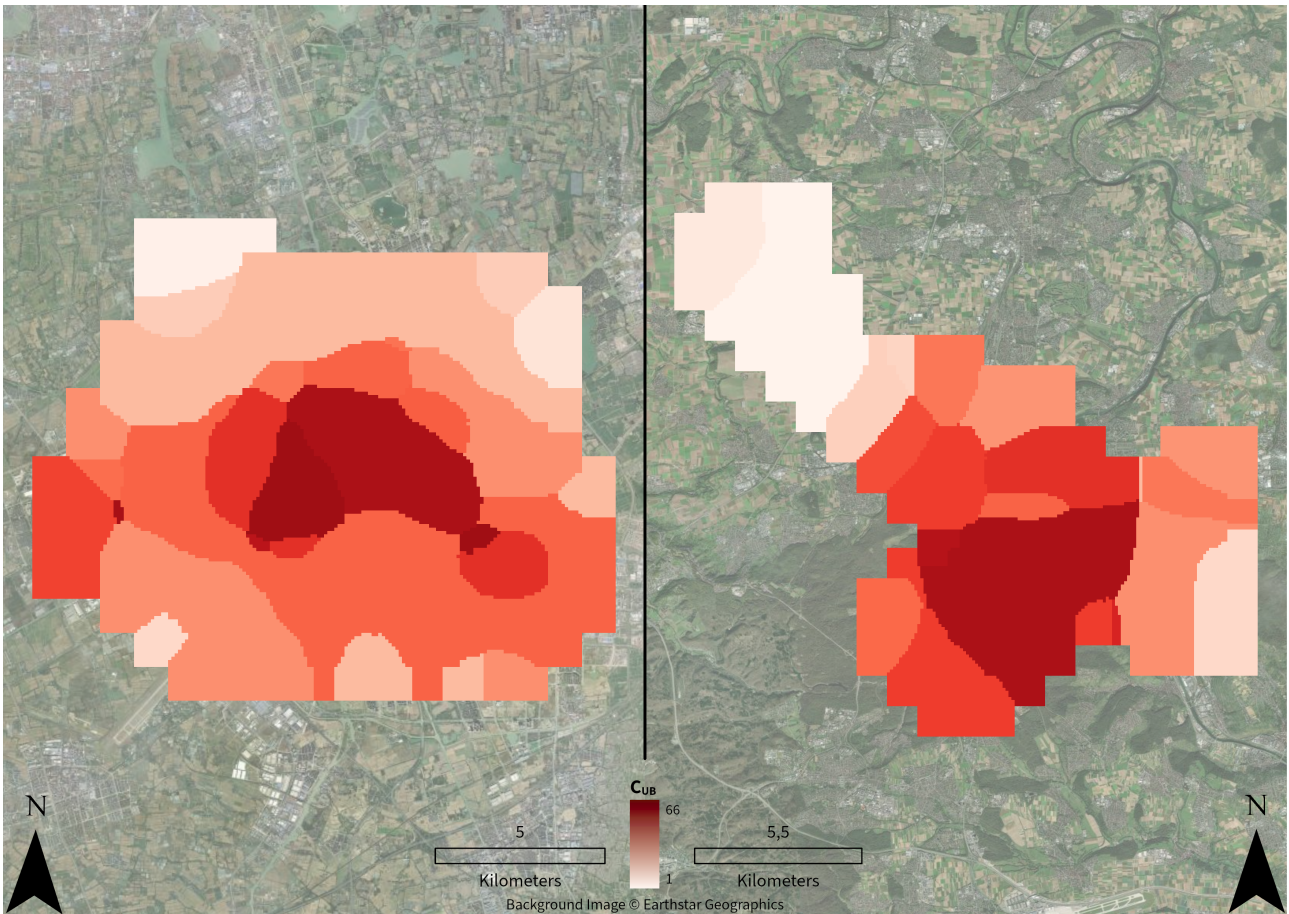


(b) Graph representation of Gujrat

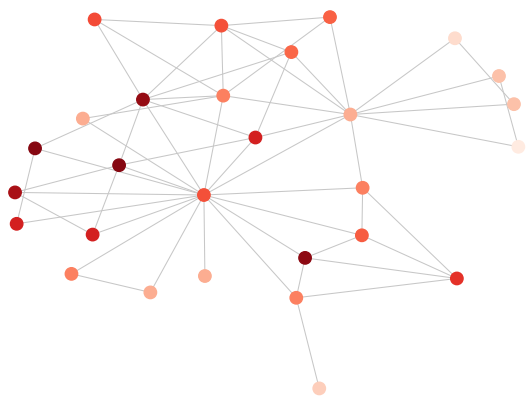


(c) Graph representation of San Salvador

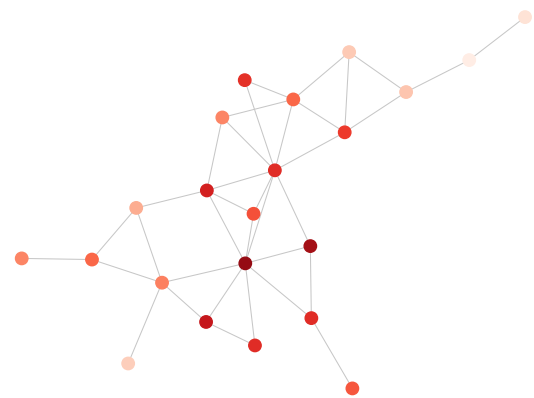
**Figure B.66** Overview of two cities within  $c_{CM_3}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_4}$

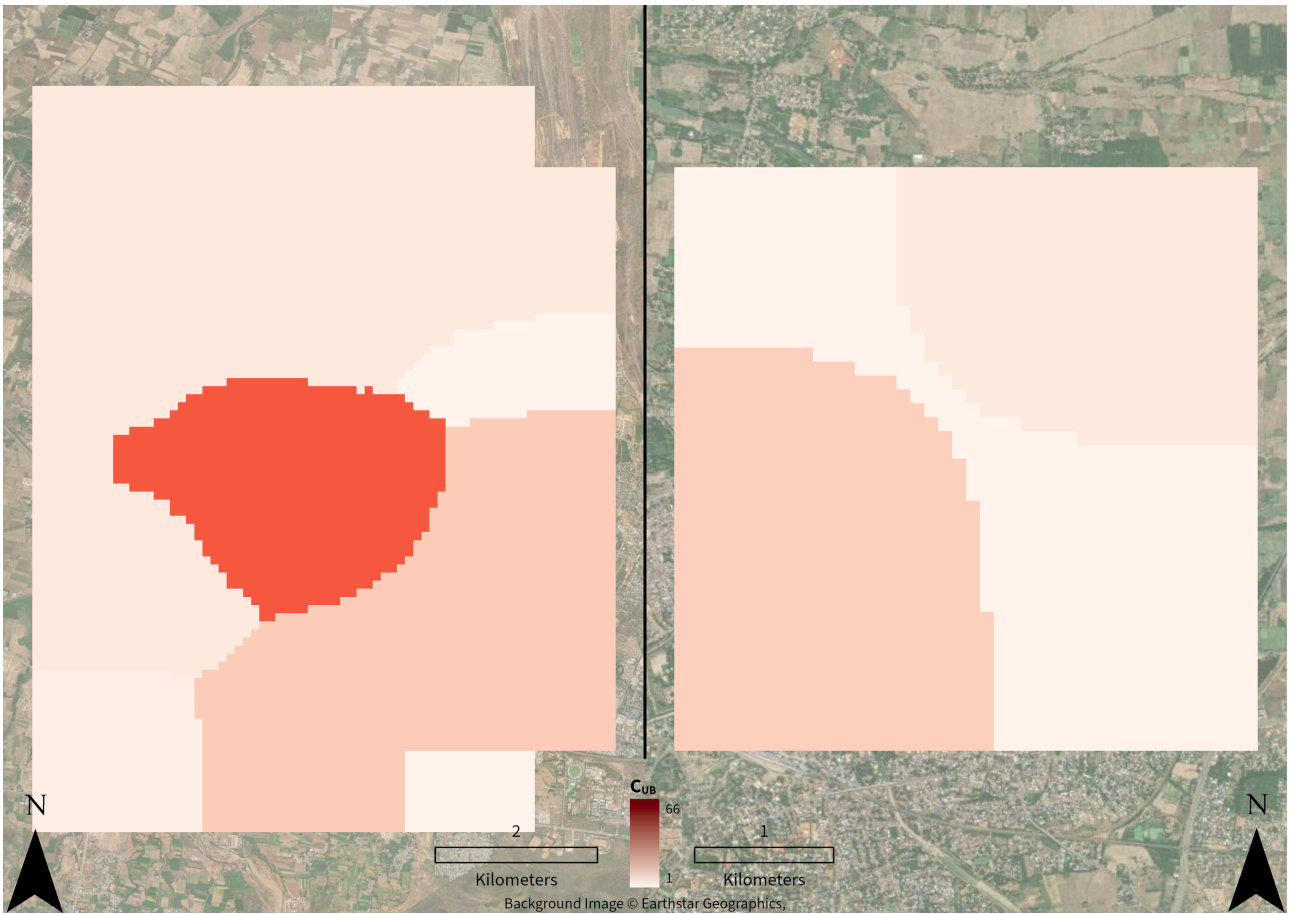


(b) Graph representation of Jiaying

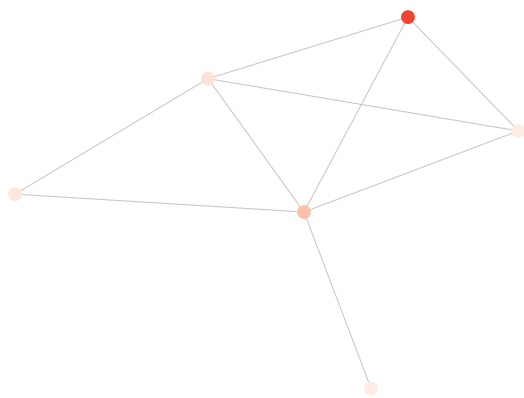


(c) Graph representation of Stuttgart

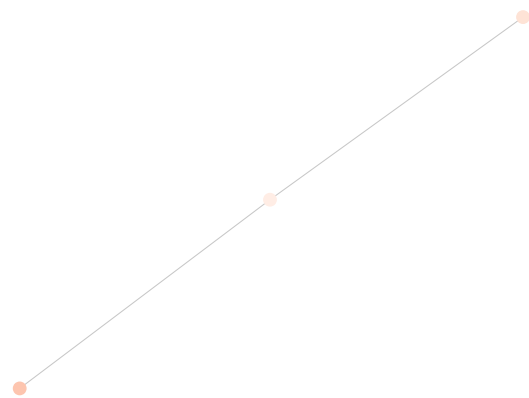
**Figure B.67** Overview of two cities within  $c_{CM_4}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_5}$



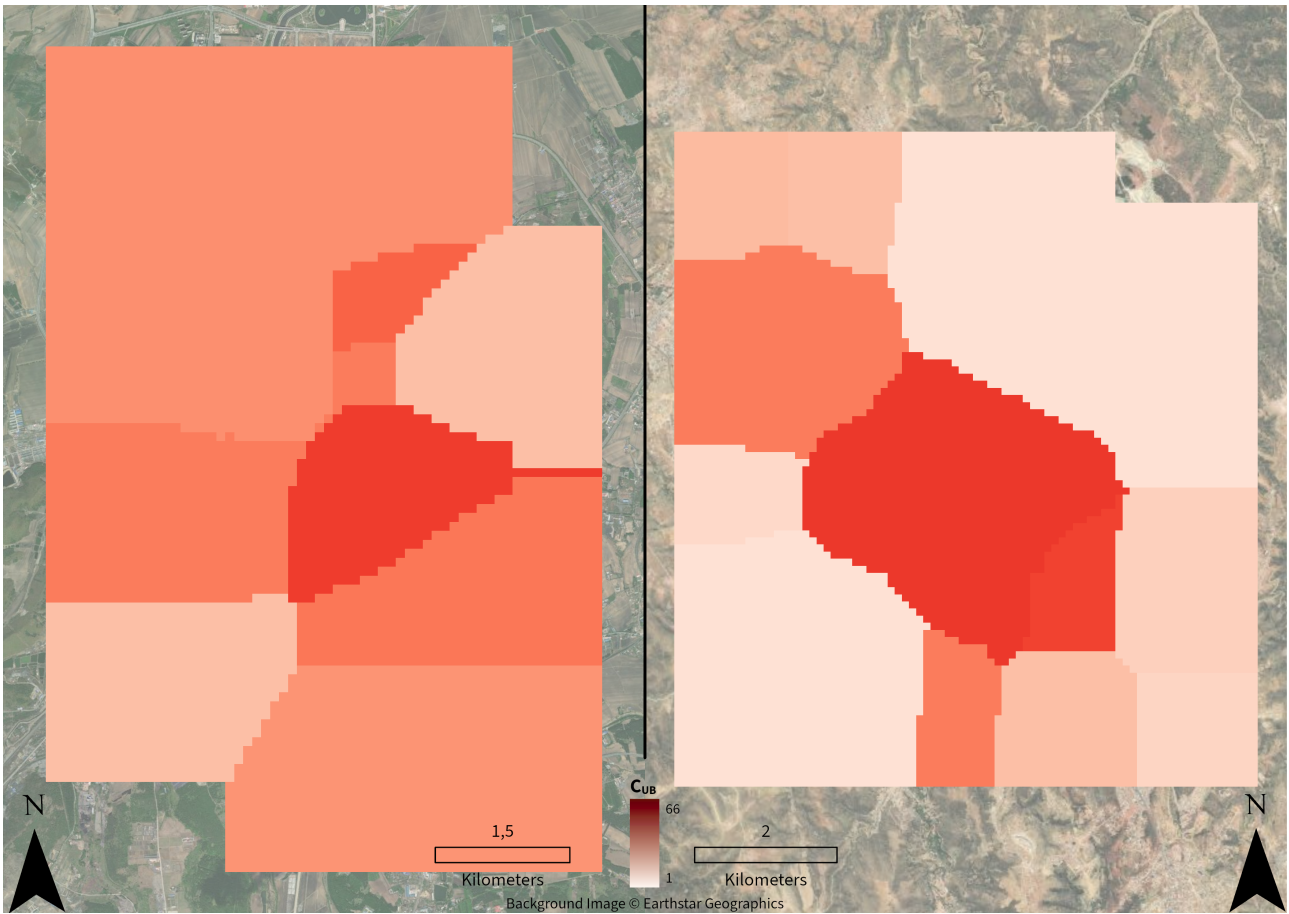
(b) Graph representation of Kadapa



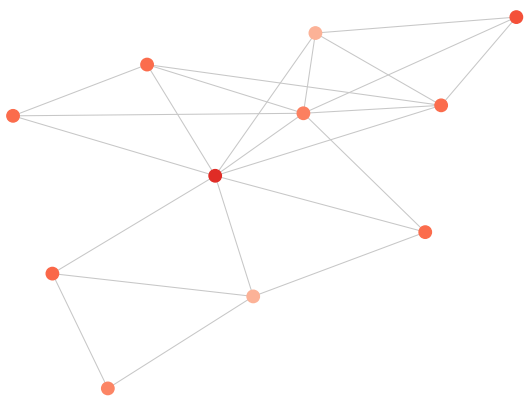
(c) Graph representation of Thanjavur

**Figure B.68** Overview of two cities within  $c_{CM_5}$  and their graph representations.

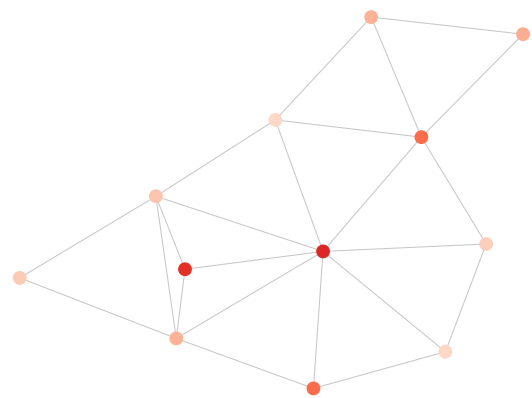




(a) Overview of the urban patterns of two cities within  $c_{CM_6}$

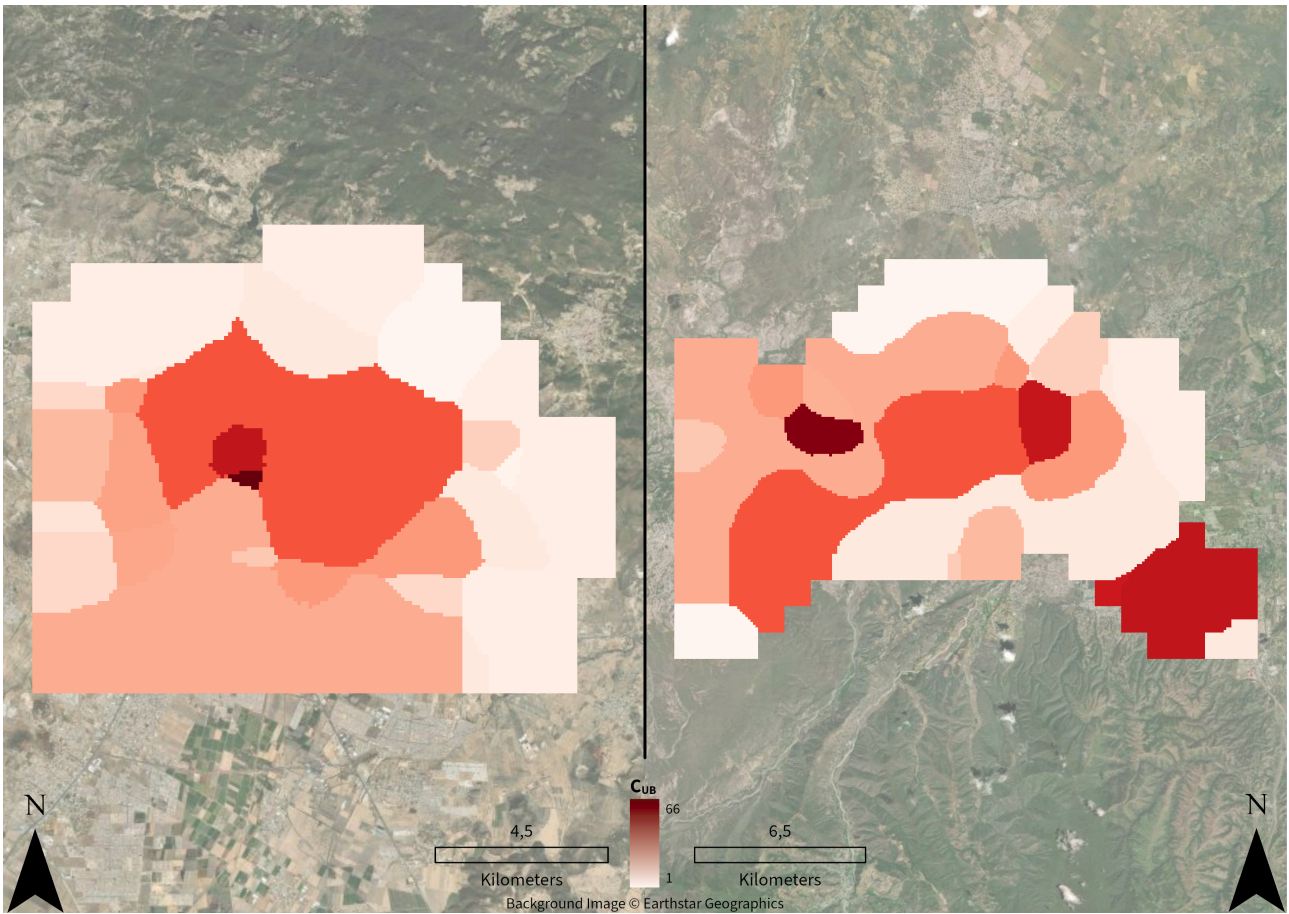


(b) Graph representation of Shuangyashan

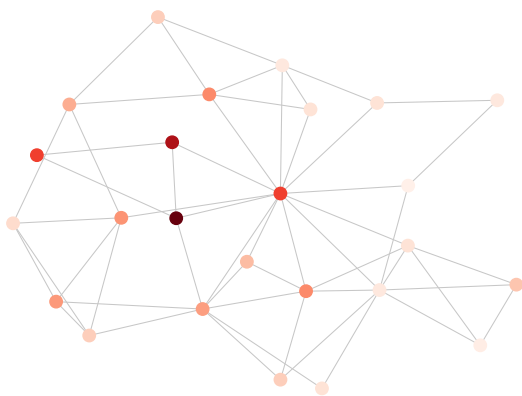


(c) Graph representation of Sucre

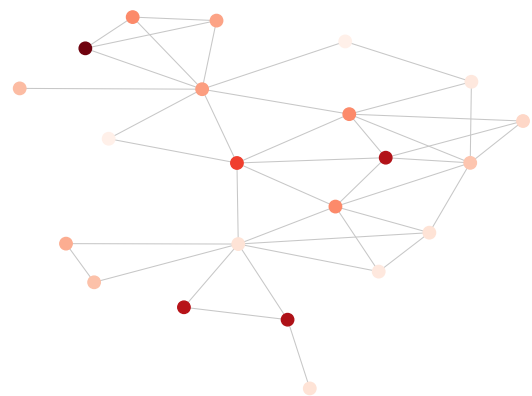
**Figure B.69** Overview of two cities within  $c_{CM_6}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_7}$

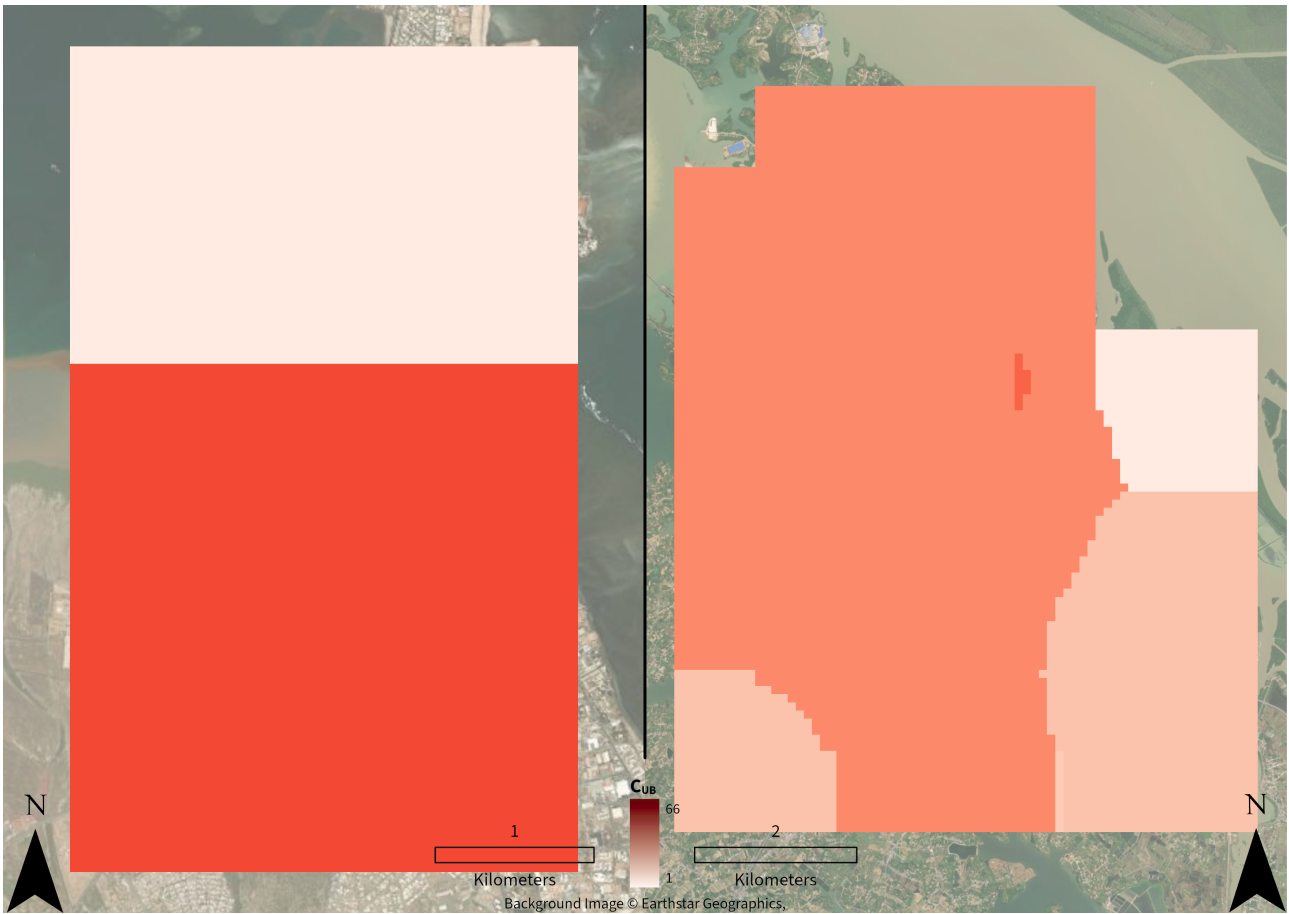


(b) Graph representation of Pachuca de Soto



(c) Graph representation of Barquisimeto

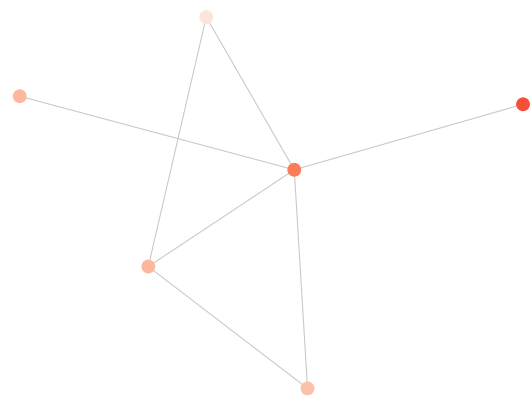
**Figure B.70** Overview of two cities within  $c_{CM_7}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_8}$

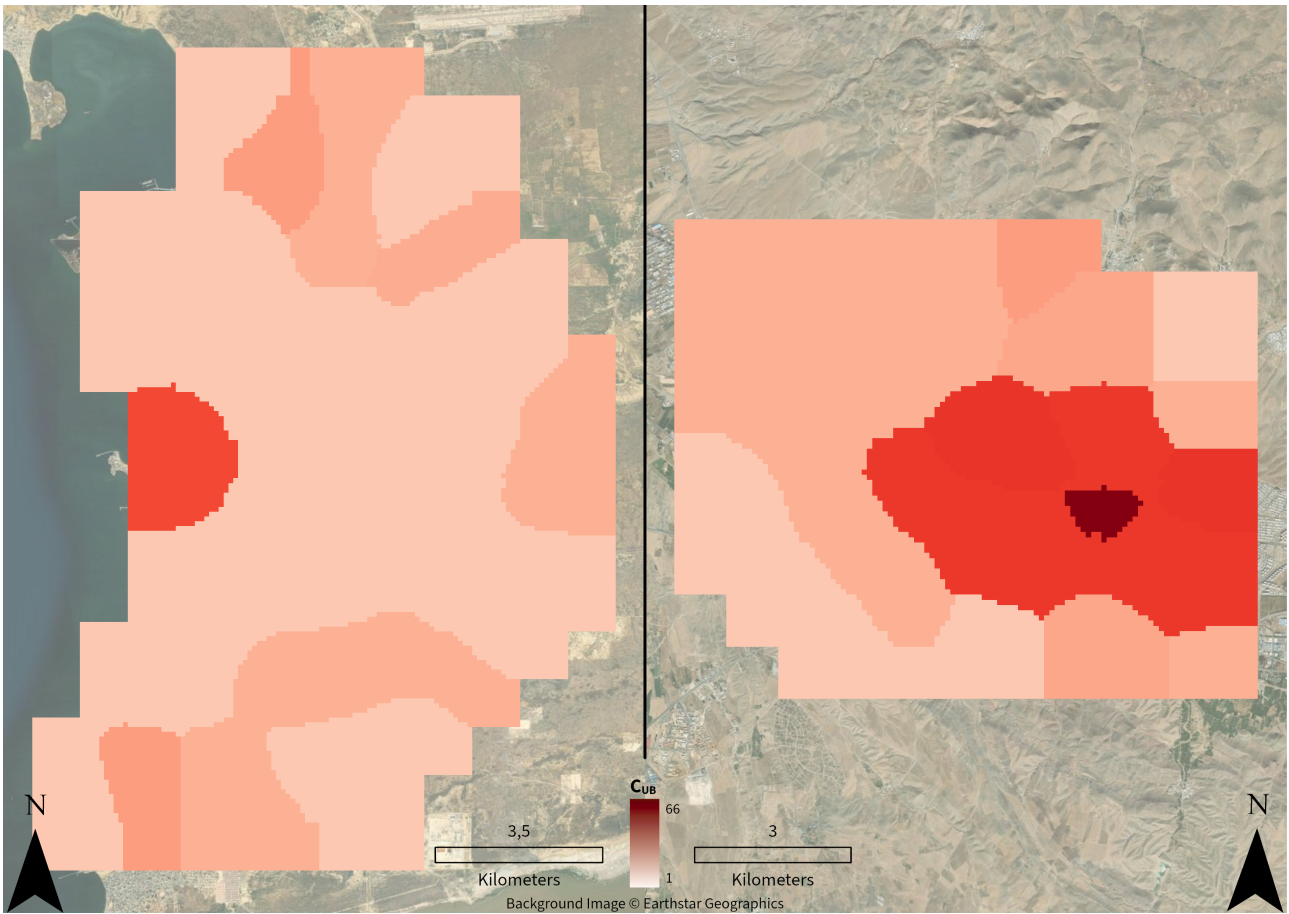


(b) Graph representation of Djibouti

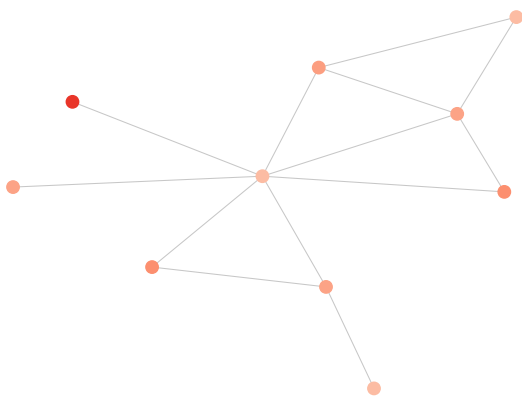


(c) Graph representation of Yuanjiang

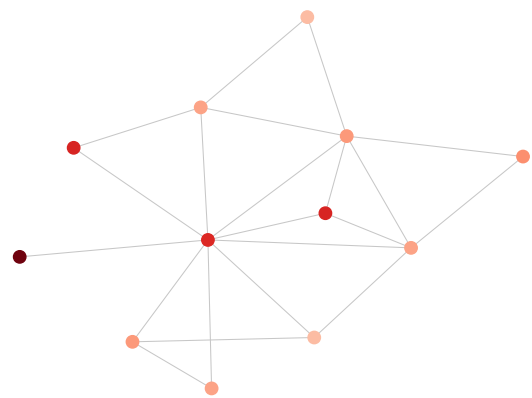
**Figure B.71** Overview of two cities within  $c_{CM_8}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_9}$

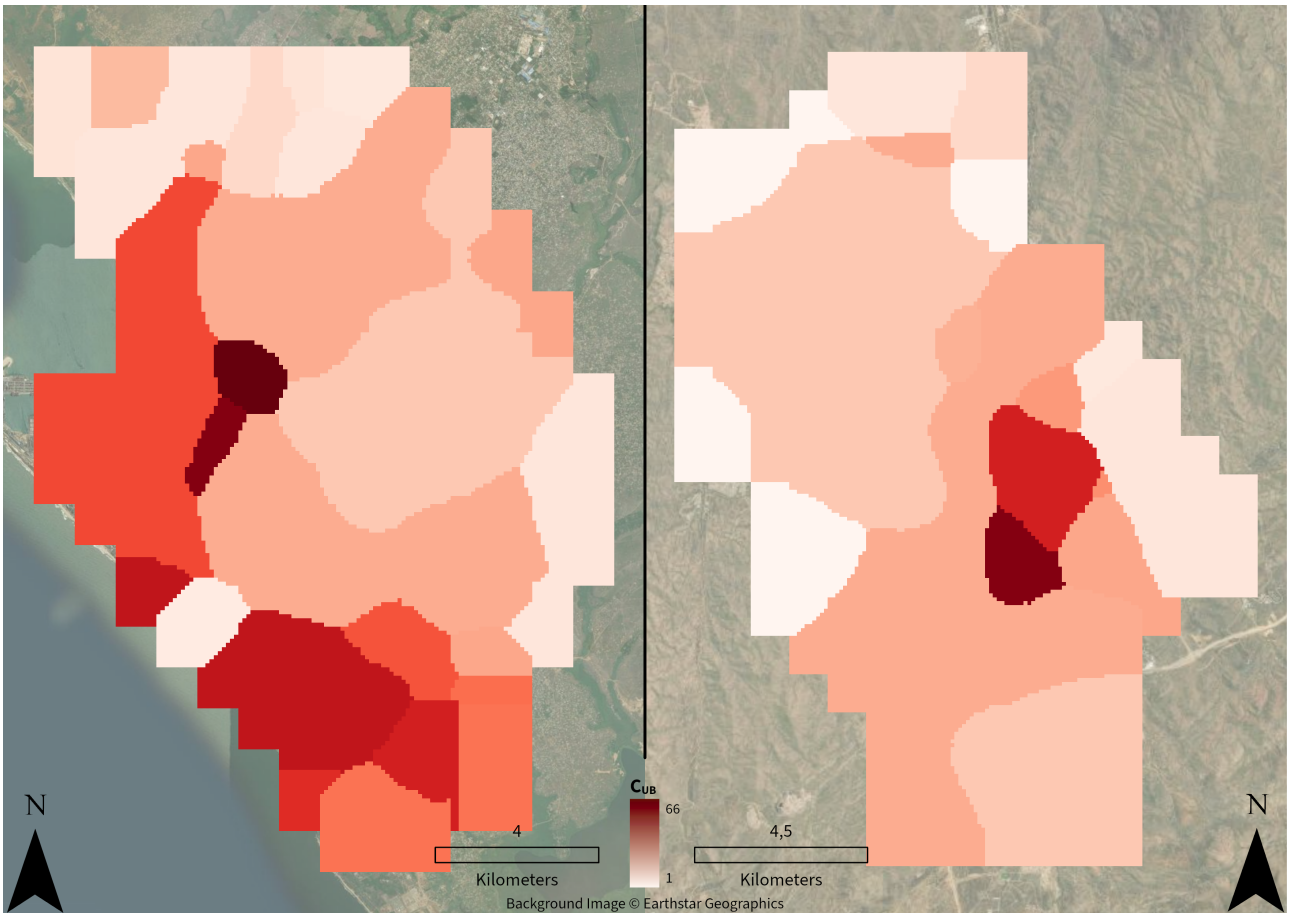


(b) Graph representation of Punto Fijo

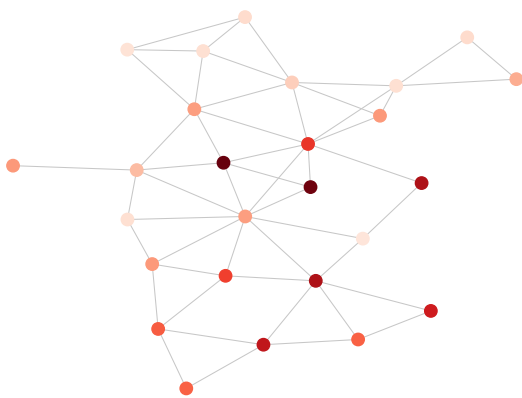


(c) Graph representation of Zanjan

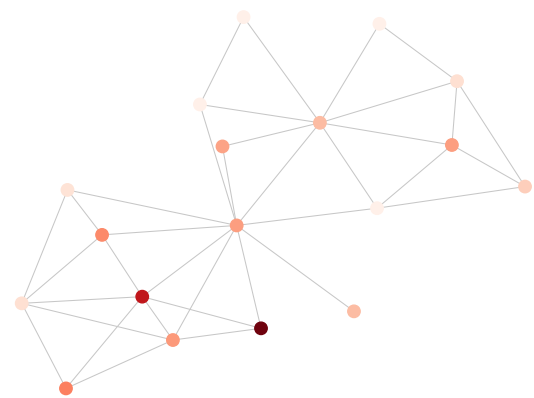
**Figure B.72** Overview of two cities within  $c_{CM_9}$  and their graph representations<sup>9</sup>.



(a) Overview of the urban patterns of two cities within  $c_{CM_{10}}$

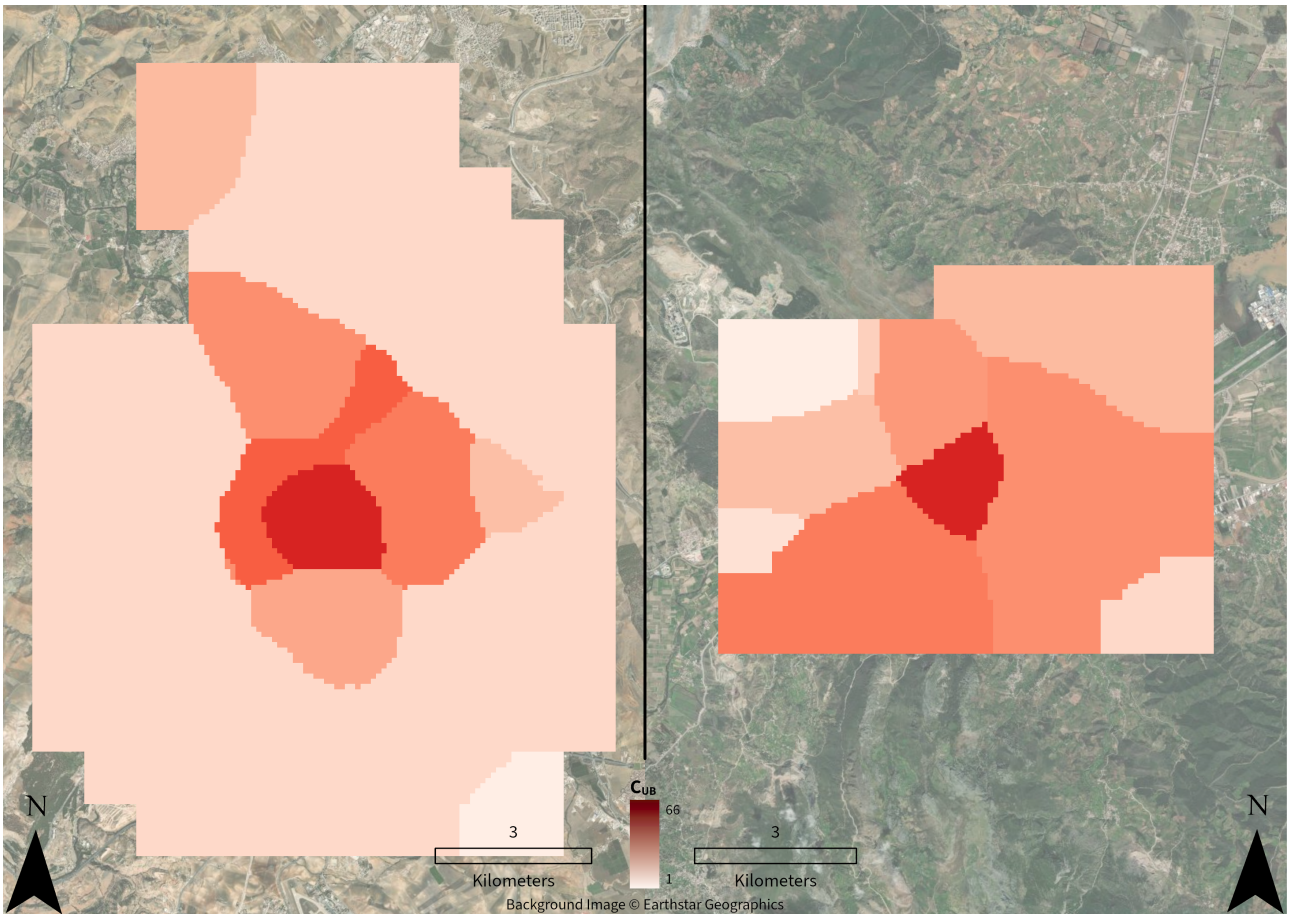


(b) Graph representation of Pointe-Noire

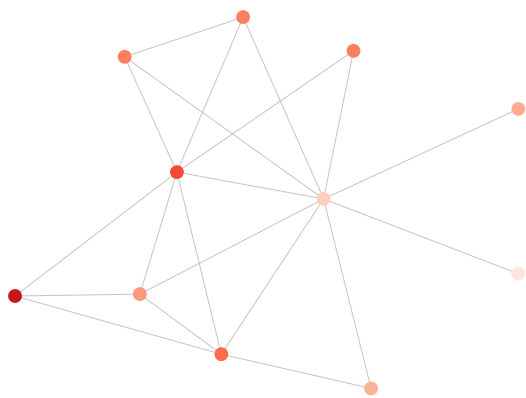


(c) Graph representation of Windhoek

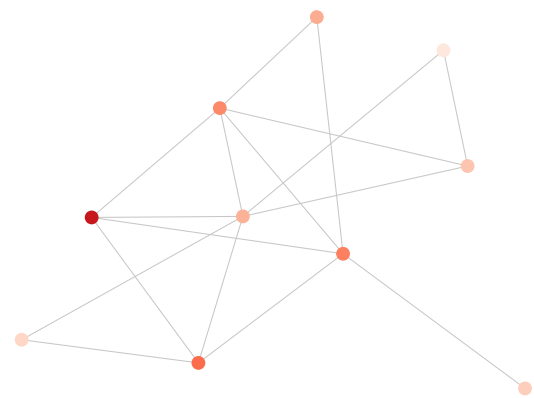
**Figure B.73** Overview of two cities within  $c_{CM_{10}}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_{11}}$

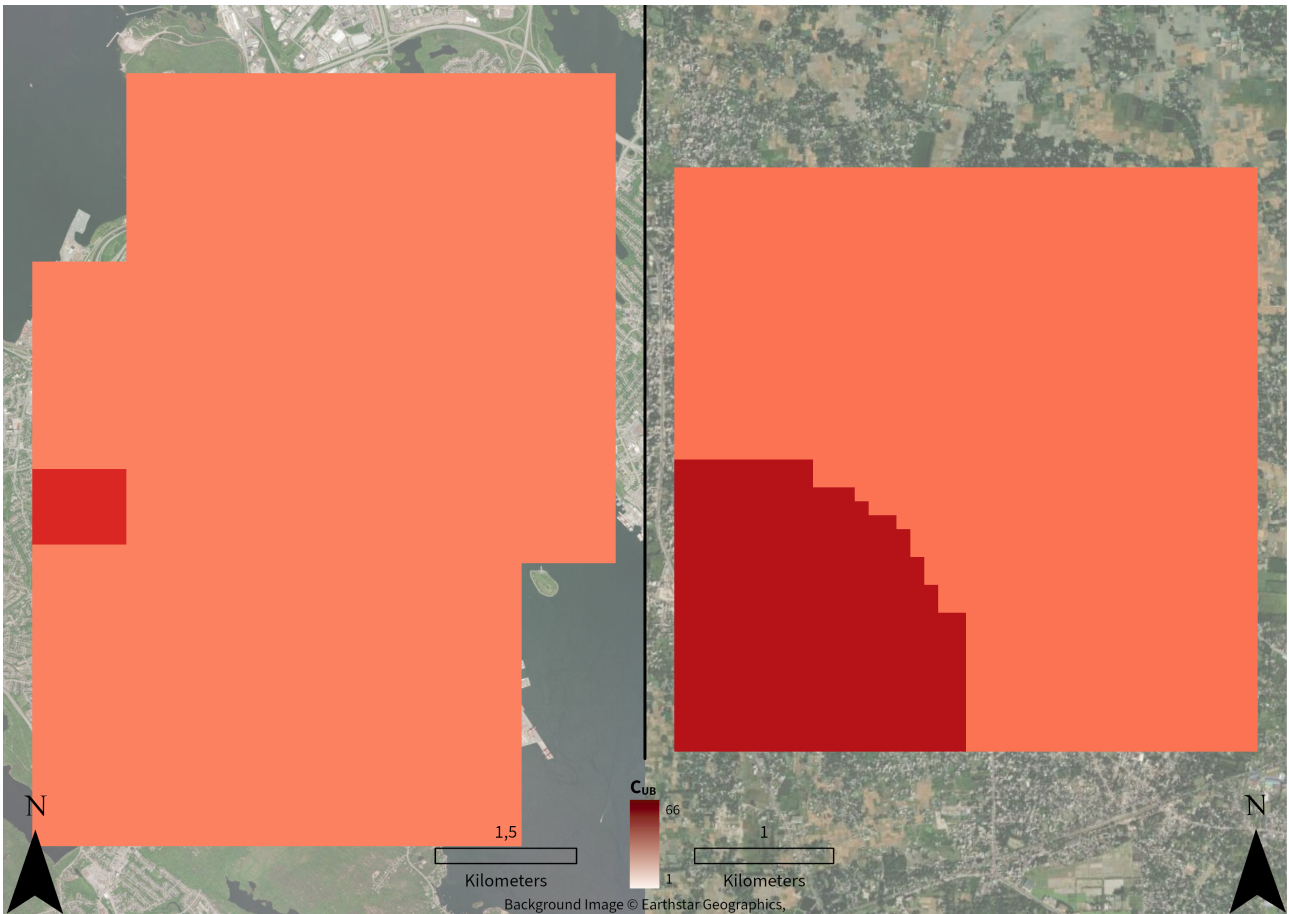


(b) Graph representation of Qacentina



(c) Graph representation of Tétouan

**Figure B.74** Overview of two cities within  $c_{CM_{11}}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_{12}}$

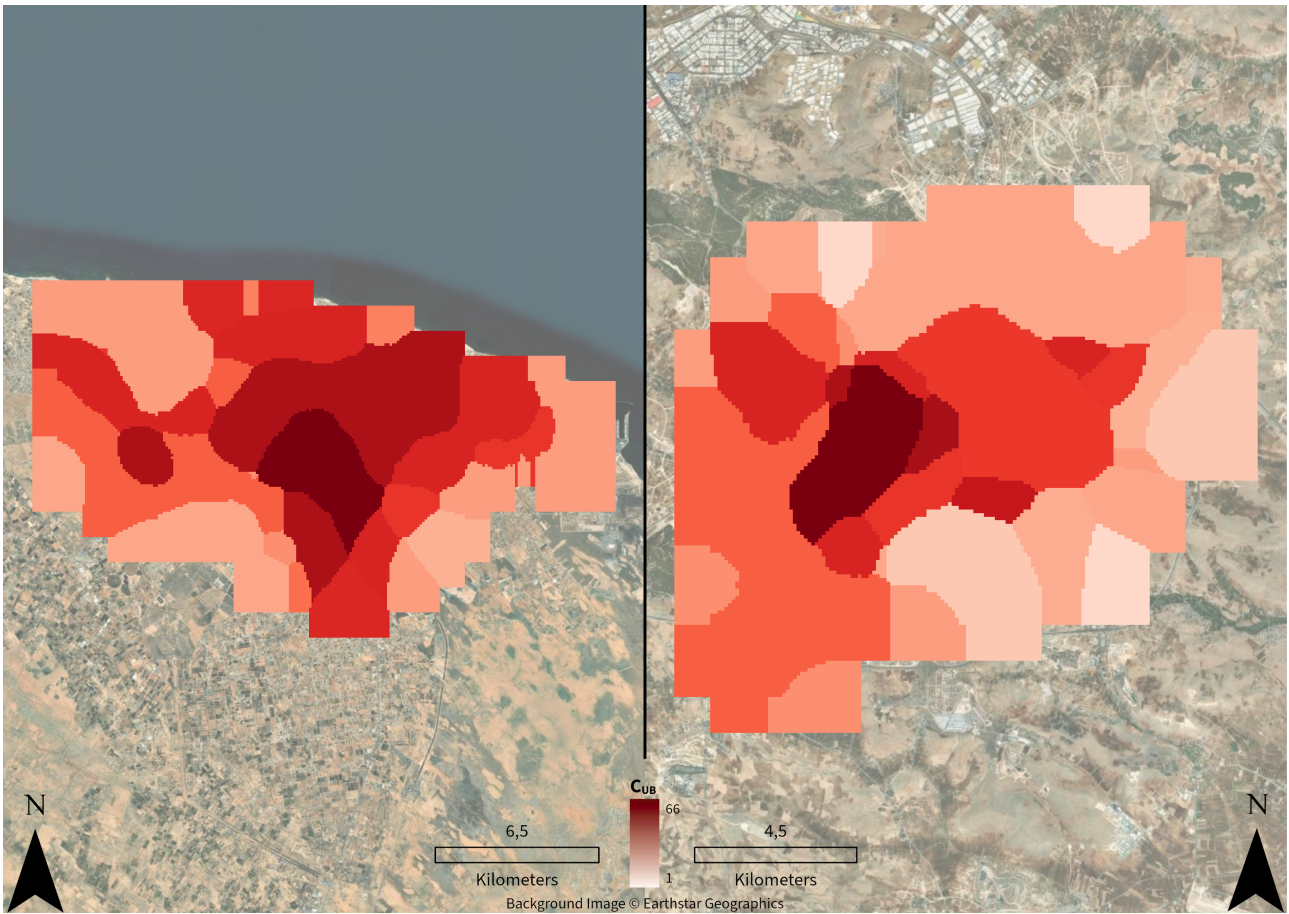


(b) Graph representation of Halifax

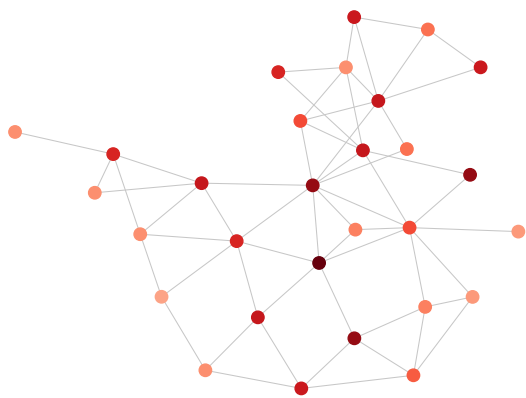


(c) Graph representation of Rangpur

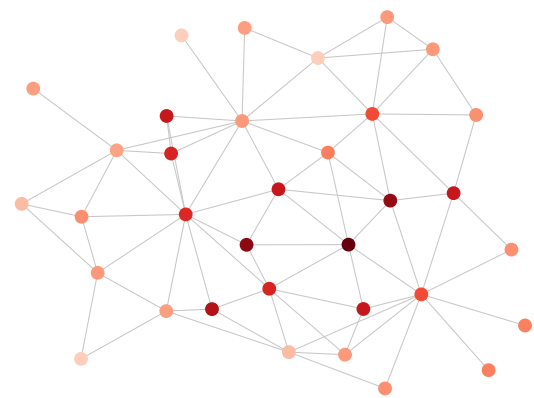
**Figure B.75** Overview of two cities within  $c_{CM_{12}}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_{13}}$



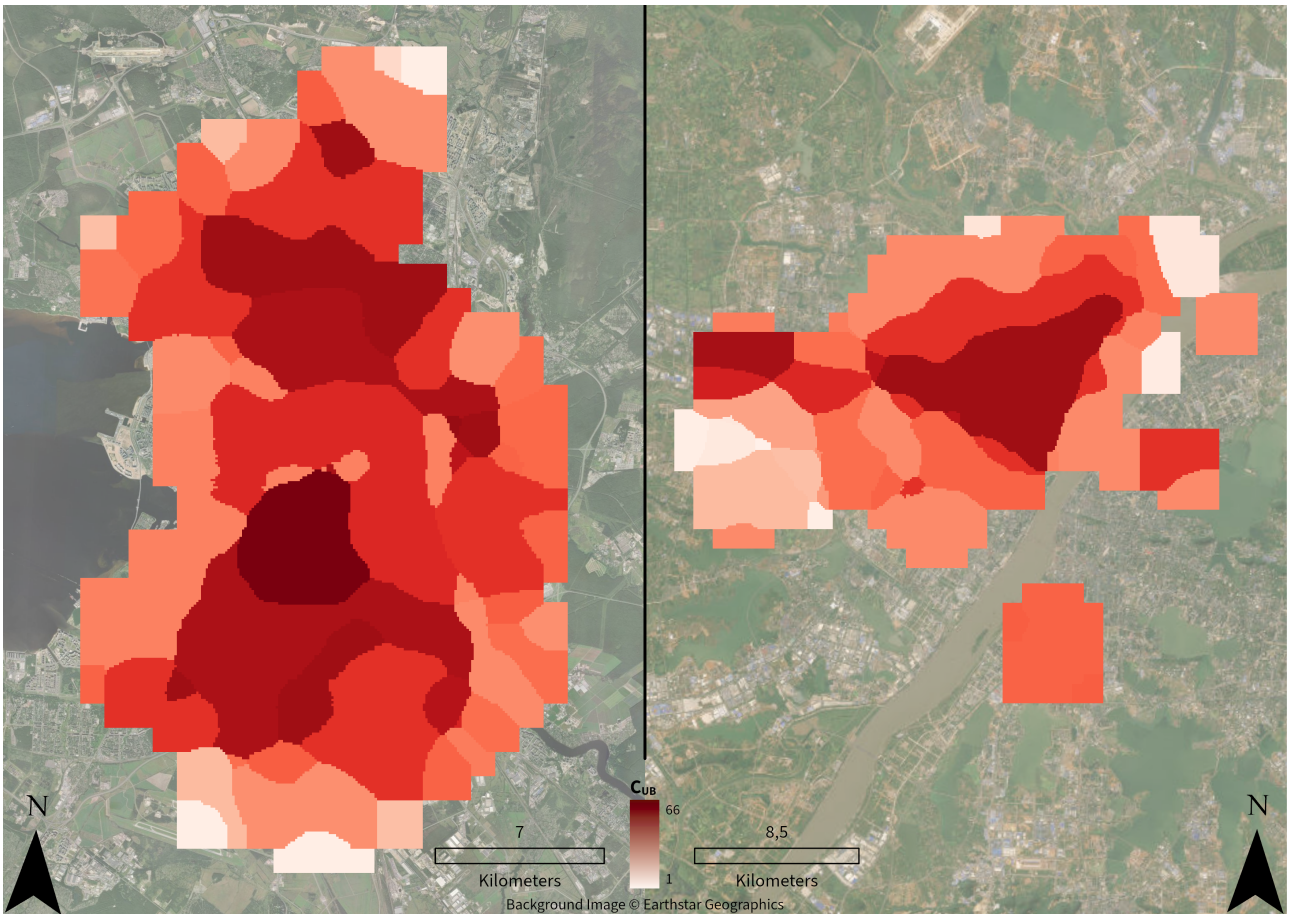
(b) Graph representation of Misratah



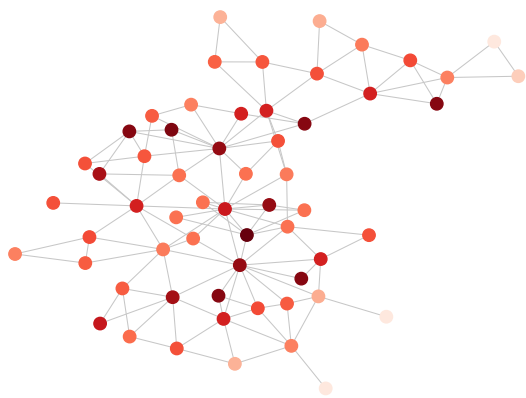
(c) Graph representation of Gaziantep

**Figure B.76** Overview of two cities within  $c_{CM_{13}}$  and their graph representations.

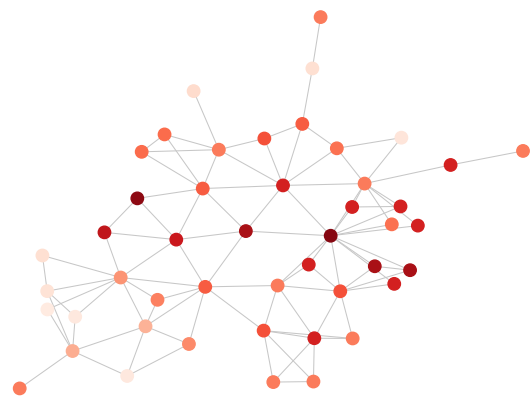




(a) Overview of the urban patterns of two cities within  $c_{CM_{14}}$

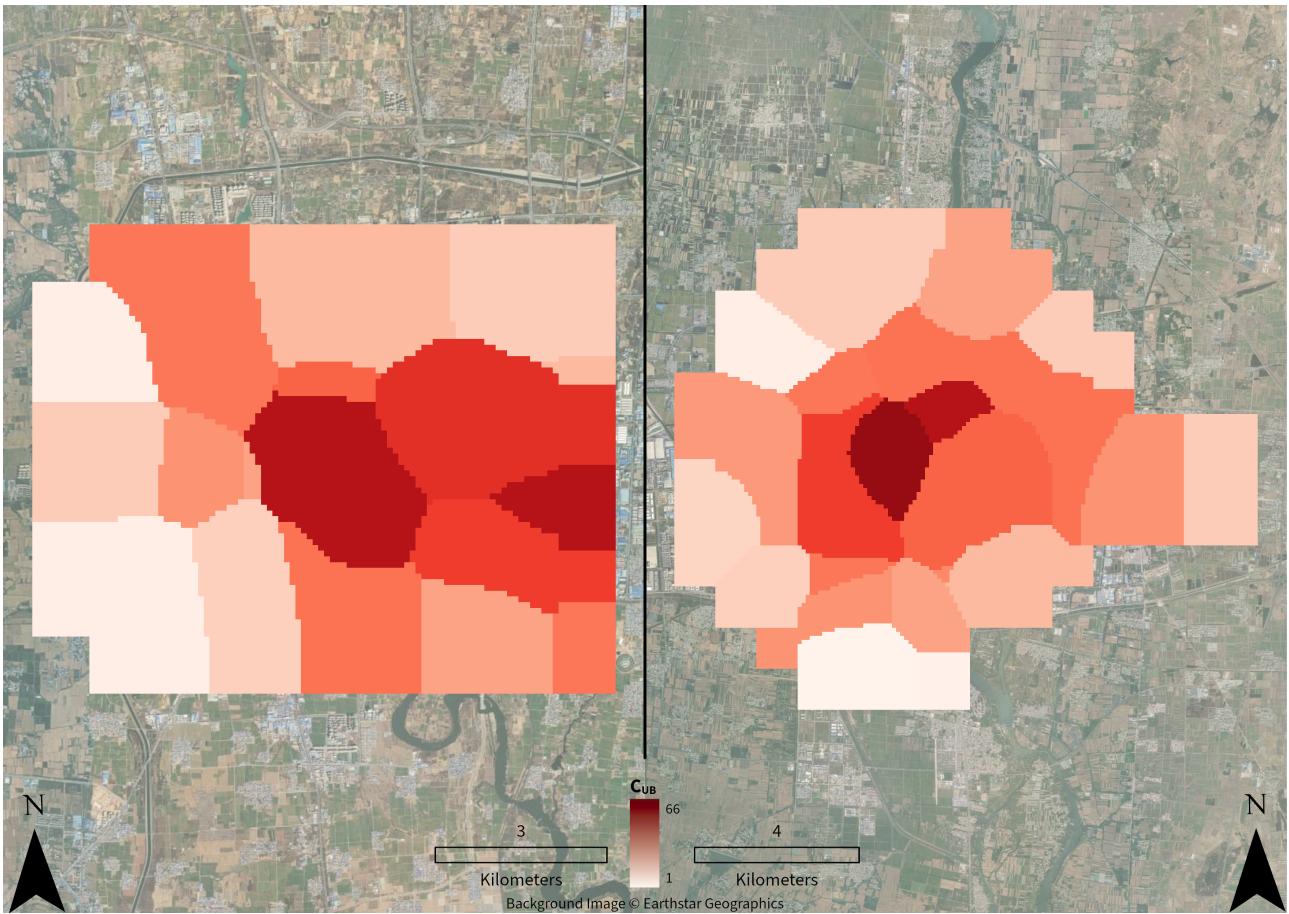


(b) Graph representation of Sankt Peterburg (Saint Petersburg)

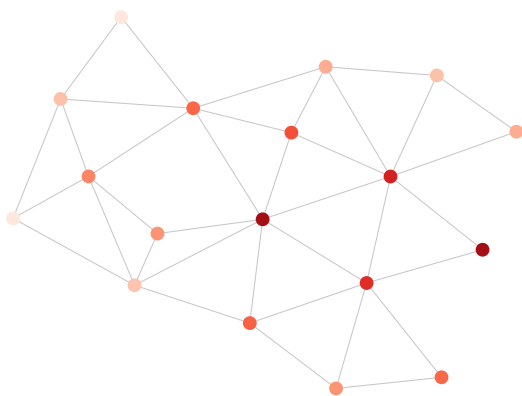


(c) Graph representation of Wuhan

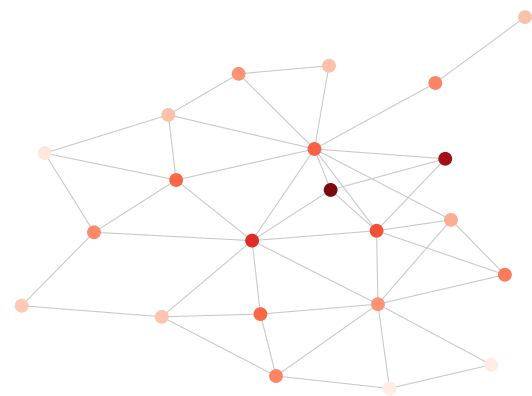
**Figure B.77** Overview of two cities within  $c_{CM_{14}}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_{15}}$

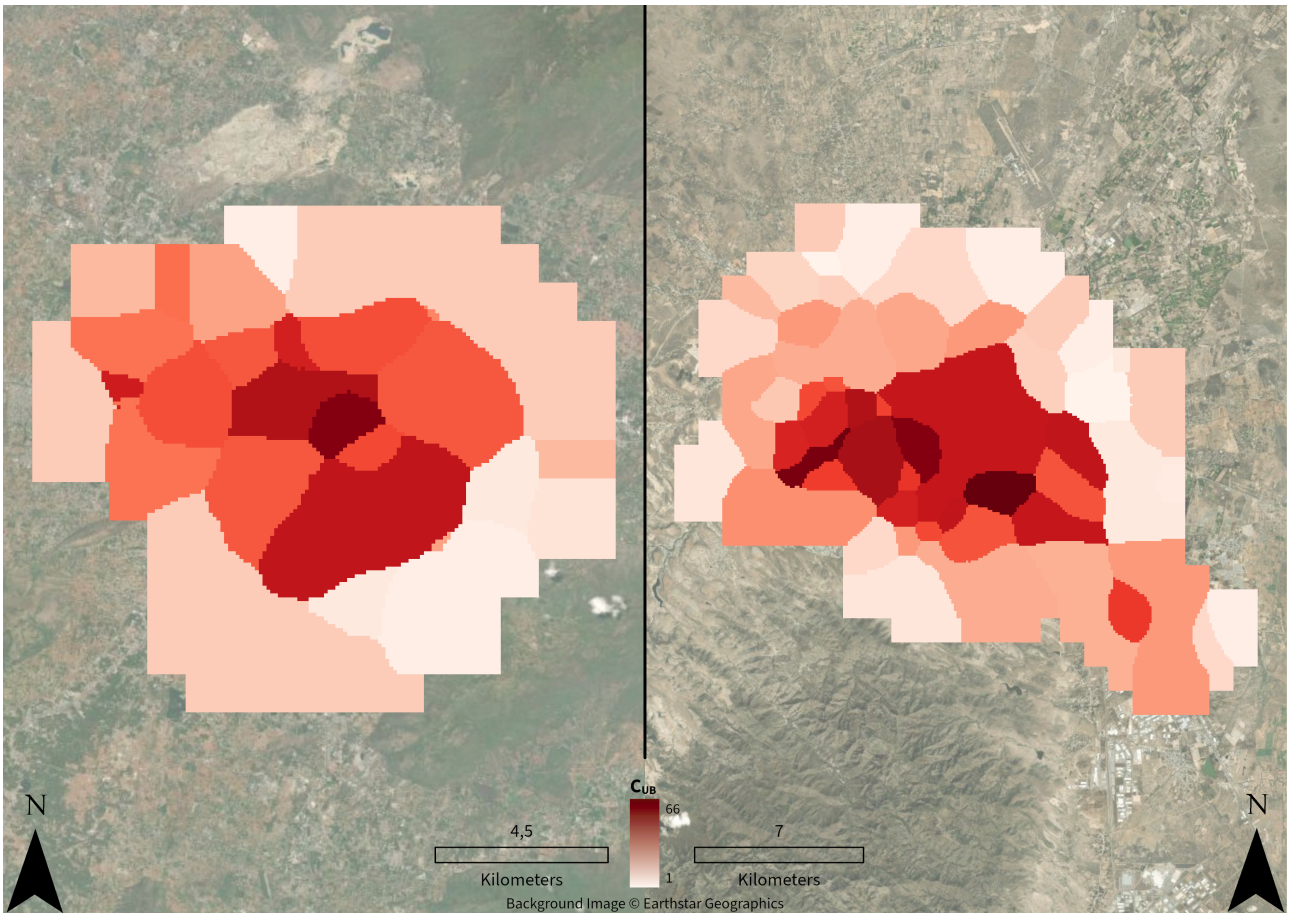


(b) Graph representation of Xinzheng

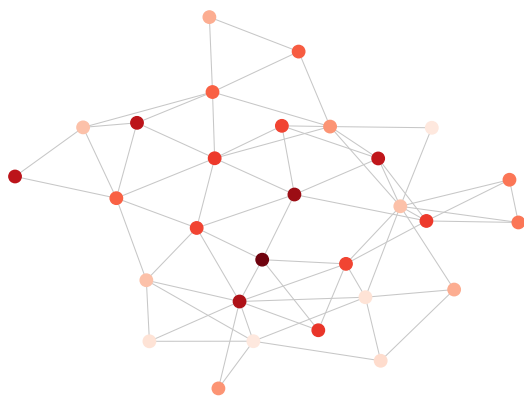


(c) Graph representation of Xinyi

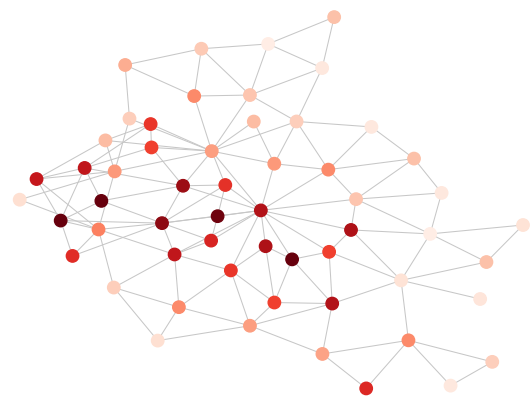
**Figure B.78** Overview of two cities within  $c_{CM_{15}}$  and their graph representations.



(a) Overview of the urban patterns of two cities within  $c_{CM_{16}}$



(b) Graph representation of Salem



(c) Graph representation of San Luis Potosí

**Figure B.79** Overview of two cities within  $c_{CM_{16}}$  and their graph representations.

## B.3 Overview Urban Areas

**Table B.1** Overview of all MUAs with corresponding country, size of MUA and LCZ as well as assigned class  $C_{CM}$ .

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
0	Winnipeg	Canada	364	553.54	543.28	98.15	16
1	Calabar	Nigeria	83	207.90	158.56	76.27	5
2	Anqiu	China	93	195.23	194.99	99.88	4
3	Maracay	Venezuela (Bolivarian Republic of)	139	449.22	278.15	61.92	3
4	Tel Aviv-Yafo (Tel Aviv-Jaffa)	Israel	157	920.49	269.43	29.27	4
5	Sanmenxia	China	49	162.36	107.50	66.21	11
6	Abidjan	Côte d'Ivoire	371	598.77	598.86	100.01	16
7	Abomey-Calavi Cotonou	Benin	136	640.22	296.25	46.27	8
8	Abuja	Nigeria	208	645.31	361.80	56.07	10
9	Abu Zaby (Abu Dhabi)	United Arab Emirates	122	930.53	244.01	26.22	8
10	Acapulco de Juárez	Mexico	50	262.15	151.66	57.85	3
11	Accra	Ghana	511	908.42	699.78	77.03	16
12	Adana	Turkey	173	337.62	337.60	99.99	16
13	Ad-Dammam	Saudi Arabia	386	775.73	663.12	85.48	9
14	Ad-Dawhah (Doha) Ad-Dawhah (Doha) Ar-Rayyan	Qatar	606	923.98	923.99	100.00	9
15	Addis Ababa	Ethiopia	376	566.62	566.33	99.95	10
16	Adelaide	Australia	713	1008.06	1007.73	99.97	16
17	Ado-Ekiti	Nigeria	85	180.47	177.90	98.58	15
18	Agadir	Morocco	27	259.45	79.63	30.69	13
19	Agartala	India	58	133.56	133.38	99.86	15
20	Aguascalientes	Mexico	179	314.29	314.07	99.93	10
21	Ahmadnagar	India	37	163.85	102.14	62.34	7
22	Ahvaz	Iran (Islamic Republic of)	293	445.30	444.35	99.79	13
23	Dubayy (Dubai) Sharjah Ajman	United Arab Emirates	556	1291.06	782.13	60.58	9
24	Ajmer	India	89	186.15	183.16	98.39	7
25	Akesu	China	87	241.88	205.89	85.12	13
26	Cleveland Akron	United States of America	828	1583.69	1241.18	78.37	2
27	Aktyubinsk	Kazakhstan	22	226.04	75.34	33.33	14
28	Al-Ain	United Arab Emirates	303	528.47	511.70	96.83	13
29	Albany	United States of America	99	352.86	194.77	55.20	2
30	Al-Basrah (Basra)	Iraq	149	279.56	276.55	98.92	9
31	Albuquerque	United States of America	604	856.44	856.44	100.00	13

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
32	Al-Fayyum	Egypt	16	84.03	61.24	72.88	11
33	Al Gadarif	Sudan	40	126.85	88.05	69.41	13
34	Al-Hasakah	Syrian Arab Republic	43	119.38	115.55	96.79	13
35	Al-Hudaydah	Yemen	52	138.81	120.18	86.58	8
36	Aligarh	India	76	158.64	158.47	99.89	10
37	Al-Iskandariyah (Alexandria) Kafr-ad-Dawwar	Egypt	63	772.97	204.57	26.47	11
38	Al-Ismailiyah	Egypt	55	158.29	122.05	77.10	13
39	Al-Kamishli	Syrian Arab Republic	54	128.09	123.02	96.04	10
40	Al-Khartum (Khartoum)	Sudan	849	1213.37	1213.39	100.00	13
41	Allahabad	India	122	286.89	256.66	89.46	10
42	Allentown-Bethlehem	United States of America	36	117.47	117.47	100.00	2
43	Al-Madinah (Medina)	Saudi Arabia	337	533.45	531.65	99.66	9
44	Al-Manamah (Manama)	Bahrain	282	637.59	483.96	75.91	9
45	Almaty	Kazakhstan	394	597.17	597.00	99.97	11
46	Al-Mawsil (Mosul)	Iraq	223	373.10	353.28	94.69	13
47	Al-Mukalla	Yemen	8	161.96	33.97	20.97	9
48	Al Obeid (Al Ubayyid)	Sudan	134	237.06	236.12	99.60	13
49	Al-Qahirah (Cairo)	Egypt	1043	1388.58	1373.16	98.89	13
50	Al-Raqqa	Syrian Arab Republic	44	108.93	107.16	98.37	11
51	Alwar	India	8	198.62	35.19	17.72	10
52	Amara	Iraq	80	170.50	155.31	91.09	9
53	Amman Zarqa Ar-Rusayfah	Jordan	684	978.26	976.26	99.80	13
54	Amritsar	India	132	238.95	238.16	99.67	10
55	Amsterdam	Netherlands	318	517.38	506.20	97.84	1
56	Anantapur	India	47	125.90	99.56	79.08	7
57	Anápolis	Brazil	131	254.69	254.29	99.84	7
58	Andizhan	Uzbekistan	130	250.03	249.80	99.91	16
59	Ankang	China	31	94.11	88.46	94.00	6
60	Ankara	Turkey	376	707.38	560.44	79.23	11
61	Ann Arbor	United States of America	42	143.00	142.89	99.92	2
62	Binzhou	China	157	381.16	275.69	72.33	4
63	Seoul Incheon Suweon Seongnam Goyang Bucheon Ansan Anyang Uijeongbu Siheung Gwangmyeong	Republic of Korea	2525	3187.37	3187.37	100.00	4
64	Anshan Liaoyang	China	483	723.80	723.86	100.01	11
65	Anshun	China	34	103.33	83.82	81.12	6

continued on next page

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
66	Antalya	Turkey	133	304.53	288.13	94.61	4
67	Antananarivo	Madagascar	220	397.49	352.64	88.72	10
68	Antioch	United States of America	107	353.46	213.96	60.53	10
69	Anyang	China	172	294.36	293.68	99.77	4
70	Aracaju	Brazil	114	276.88	208.30	75.23	7
71	Ardabil	Iran (Islamic Republic of)	72	149.54	148.81	99.51	13
72	Arequipa	Peru	163	314.95	302.14	95.93	9
73	Arkhangelsk	Russian Federation	25	99.89	81.38	81.47	1
74	Armenia	Colombia	51	139.01	113.56	81.69	3
75	Ar-Riyadh (Riyadh)	Saudi Arabia	1439	1786.88	1786.83	100.00	9
76	Arusha	United Republic of Tanzania	107	207.52	205.82	99.18	15
77	Asahikawa	Japan	123	226.85	226.75	99.96	16
78	Asheville	United States of America	2	207.16	21.73	10.49	12
79	Ashgabat	Turkmenistan	152	299.13	284.89	95.24	13
80	Asmara	Eritrea	69	144.78	137.73	95.13	10
81	As-Suways	Egypt	27	196.89	94.64	48.07	13
82	Astana	Kazakhstan	138	264.59	246.21	93.05	13
83	Astrakhan	Russian Federation	212	360.57	355.77	98.67	11
84	Aswan	Egypt	44	130.93	105.08	80.26	9
85	Asyut	Egypt	36	108.35	89.77	82.85	14
86	Atlanta	United States of America	142	487.49	284.49	58.36	4
87	Auckland	New Zealand	515	910.09	909.96	99.99	2
88	Augusta-Richmond County	United States of America	47	522.20	110.19	21.10	15
89	Aurangabad	India	113	269.96	220.52	81.69	10
90	Austin	United States of America	299	954.70	533.75	55.91	2
91	Az-Zaqazig	Egypt	35	104.40	86.37	82.73	11
92	Baaqoobah	Iraq	52	171.25	107.05	62.51	13
93	Baghdad	Iraq	722	974.39	974.19	99.98	13
94	Bago	Myanmar	46	119.82	106.93	89.24	5
95	Baguio City	Philippines	60	137.37	137.18	99.86	15
96	Baharampur	India	5	109.31	36.96	33.81	1
97	Bahawalpur	Pakistan	73	154.45	153.83	99.60	7
98	Baicheng	China	36	99.28	99.20	99.92	11
99	Baishan	China	34	95.51	91.59	95.90	6
100	Baixada Santista	Brazil	137	296.47	296.28	99.93	16
101	Baiyin	China	78	166.88	166.41	99.72	13
102	Bakersfield	United States of America	458	654.73	654.36	99.94	10

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
103	Baku Sumquayit	Azerbaijan	676	1332.50	1024.54	76.89	4
104	Washington, D,C, Baltimore	United States of America	1623	2289.15	2288.70	99.98	2
105	Bamako	Mali	325	497.27	492.39	99.02	10
106	Bamenda	Cameroon	77	159.92	159.88	99.97	15
107	Bandar Abbas	Iran (Islamic Republic of)	45	404.00	116.51	28.84	9
108	Bandung	Indonesia	353	533.92	533.61	99.94	16
109	Bangalore	India	659	863.52	863.49	100.00	11
110	Banghazi	Libya	73	361.74	161.93	44.76	13
111	Bangui	Central African Republic	103	216.99	209.66	96.62	16
112	Banjarmasin	Indonesia	105	200.83	200.75	99.96	5
113	Banjul	Gambia	89	248.74	189.50	76.18	2
114	Baoding	China	399	704.63	699.86	99.32	11
115	Baotou	China	373	635.29	550.14	86.60	13
116	Barcelona	Spain	192	379.06	379.36	100.08	4
117	Bareilly	India	79	185.94	180.36	97.00	3
118	Bari	Italy	59	223.08	128.15	57.44	11
119	Barinas	Venezuela (Bolivarian Republic of)	96	186.55	177.94	95.38	3
120	Barisal	Bangladesh	46	111.40	93.21	83.67	15
121	Barletta	Italy	1	70.42	18.95	26.91	8
122	Basel	Switzerland	106	217.40	214.16	98.51	2
123	Bathinda	India	60	142.13	134.46	94.61	10
124	Batman	Turkey	43	107.95	96.18	89.09	7
125	Batna	Algeria	50	117.64	115.65	98.30	9
126	Baton Rouge	United States of America	307	528.73	474.53	89.75	15
127	Bauchi	Nigeria	73	176.18	169.81	96.38	10
128	Bauru	Brazil	126	228.64	228.45	99.92	7
129	Be'er Sheva	Israel	56	171.60	131.19	76.45	9
130	Begusarai	India	25	199.48	68.29	34.23	15
131	Beijing Sanhe	China	2075	2667.72	2667.28	99.98	11
132	Belém	Brazil	71	477.78	153.15	32.05	3
133	Belfast	United Kingdom	89	327.31	211.01	64.47	2
134	Bellary	India	10	136.07	38.76	28.49	10
135	Bengbu	China	107	209.03	208.85	99.91	4
136	Bengkulu	Indonesia	22	149.20	81.80	54.83	3
137	Benin City	Nigeria	337	510.73	510.46	99.95	2
138	Benxi	China	74	153.24	153.24	100.00	6
139	Beograd (Belgrade)	Serbia	128	270.23	267.17	98.87	4

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
140	Milano (Milan) Bergamo Busto Arsizio Seregno Como	Italy	1572	2156.30	2156.11	99.99	15
141	Berlin	Germany	491	699.52	699.42	99.99	1
142	Bern	Switzerland	53	146.07	146.09	100.01	15
143	Bhagalpur	India	34	124.79	84.89	68.03	5
144	Bhilwara	India	53	141.06	121.53	86.15	7
145	Mumbai (Bombay) Bhiwandi	India	451	1309.42	845.99	64.61	4
146	Bhopal	India	190	359.29	338.02	94.08	16
147	Bhubaneswar	India	119	456.51	279.02	61.12	7
148	Bielefeld	Germany	36	112.73	112.71	99.98	15
149	Bijapur	India	41	146.11	93.81	64.20	10
150	Bijie	China	38	105.78	87.97	83.17	6
151	Bikaner	India	123	222.44	211.85	95.24	13
152	Bilaspur	India	55	137.36	118.12	85.99	15
153	Bilbao	Spain	17	261.53	47.32	18.09	6
154	Liverpool Birkenhead	United Kingdom	275	587.96	552.20	93.92	2
155	Birmingham (West Midlands)	United Kingdom	499	751.67	751.55	99.98	2
156	Bishkek	Kyrgyzstan	237	374.84	374.51	99.91	16
157	Bissau	Guinea-Bissau	54	142.58	130.17	91.30	7
158	Blantyre-Limbe	Malawi	205	380.16	335.66	88.29	7
159	Bloemfontein	South Africa	250	389.06	388.97	99.98	10
160	Boa Vista	Brazil	43	249.93	102.33	40.94	2
161	Bobo-Dioulasso	Burkina Faso	147	264.08	263.83	99.91	7
162	Jakarta Bogor	Indonesia	2642	3206.04	3206.20	100.00	16
163	Bogra	Bangladesh	41	118.01	105.19	89.14	15
164	Boise City	United States of America	173	521.11	313.24	60.11	16
165	Cape Coral Bonita Springs-Naples	United States of America	393	2549.23	600.40	23.55	2
166	Bonn	Germany	29	122.05	122.02	99.97	1
167	Bordeaux	France	214	364.45	364.28	99.95	2
168	Boston	United States of America	368	682.10	682.08	100.00	2
169	Botou	China	47	112.99	111.70	98.86	10
170	Bouake	Côte d'Ivoire	110	205.94	202.33	98.25	15
171	Bozhou	China	78	162.97	161.61	99.16	4
172	Brahmapur	India	39	118.41	92.67	78.26	5
173	Brasília	Brazil	57	892.71	126.68	14.19	11
174	Bratislava	Slovakia	64	196.98	176.34	89.52	1
175	Kinshasa Brazzaville	Democratic Republic of the Congo	630	927.13	927.09	100.00	5

*continued on next page*



ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
176	Bremen	Germany	103	309.86	223.70	72.19	2
177	Brescia	Italy	101	328.82	238.29	72.47	15
178	Brest	Belarus	74	181.09	169.80	93.76	1
179	Brighton-Worthing-Littlehampton	United Kingdom	28	109.40	80.06	73.18	2
180	Brisbane	Australia	953	1325.16	1325.14	100.00	16
181	Bristol	United Kingdom	178	310.46	310.20	99.92	5
182	Brno	Czech Republic	86	198.90	198.97	100.04	15
183	Bryansk	Russian Federation	139	294.48	270.59	91.89	1
184	Bucaramanga	Colombia	46	167.41	103.46	61.80	3
185	Bucuresti (Bucharest)	Romania	324	487.66	487.22	99.91	16
186	Budapest	Hungary	390	590.39	590.37	100.00	2
187	Buffalo	United States of America	519	738.41	736.08	99.68	2
188	Bujumbura Uvira	Burundi	130	443.55	269.71	60.81	16
189	Bukavu	Democratic Republic of the Congo	33	127.02	126.18	99.34	10
190	Bulawayo	Zimbabwe	204	366.61	351.79	95.96	11
191	Bunia	Democratic Republic of the Congo	16	122.15	59.85	49.00	12
192	Buraydah	Saudi Arabia	130	357.08	235.11	65.84	9
193	Bur Sa'id	Egypt	27	117.48	99.29	84.51	8
194	Bydgoszcz	Poland	44	141.07	120.41	85.35	1
195	Cagliari	Italy	15	198.59	67.50	33.99	8
196	Calgary	Canada	637	875.05	875.02	100.00	13
197	Camaguey	Cuba	59	135.31	131.96	97.52	5
198	Campina Grande	Brazil	94	193.87	174.39	89.95	7
199	Campinas	Brazil	278	507.52	507.54	100.00	16
200	Campo Grande	Brazil	318	486.82	486.55	99.94	16
201	Campos dos Goytacazes	Brazil	85	198.50	190.84	96.14	5
202	Canberra	Australia	50	698.33	107.42	15.38	2
203	Cancún	Mexico	109	254.17	241.55	95.04	3
204	Can Tho	Viet Nam	58	139.13	139.11	99.98	16
205	Caracas	Venezuela (Bolivarian Republic of)	283	462.09	438.17	94.82	3
206	Caruaru	Brazil	66	169.25	139.49	82.42	5
207	Napoli (Naples) Caserta Nola	Italy	784	1446.75	1181.71	81.68	15
208	Catania	Italy	260	465.63	432.91	92.97	11
209	Caxias Do Sul	Brazil	99	201.59	201.50	99.96	3
210	Cenxi	China	31	96.98	80.98	83.51	6
211	Chandrapur	India	44	154.02	96.27	62.51	10
212	Changchun	China	380	573.82	573.75	99.99	11

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
213	Changde	China	102	192.09	191.88	99.89	1
214	Changji	China	53	128.02	127.67	99.72	13
215	Changning	China	27	93.77	67.40	71.88	1
216	Changshu	China	240	433.01	433.05	100.01	4
217	Changyi	China	44	106.87	106.69	99.83	4
218	Changzhi	China	146	487.41	270.67	55.53	11
219	Chaohu	China	50	127.87	127.85	99.99	1
220	Chaoyang	China	42	139.53	120.79	86.57	14
221	Charleroi	Belgium	83	189.42	189.23	99.90	15
222	Charleston-North Charleston	United States of America	8	782.12	33.15	4.24	8
223	Charlotte	United States of America	83	366.15	192.80	52.66	2
224	Chattanooga	United States of America	104	498.37	188.40	37.80	2
225	Cheboksary	Russian Federation	39	155.26	107.50	69.24	1
226	Chelyabinsk	Russian Federation	180	377.10	377.03	99.98	4
227	Chennai (Madras)	India	665	910.63	910.39	99.97	16
228	Chenzhou	China	72	171.89	159.01	92.51	6
229	Cheonan	Republic of Korea	63	144.43	144.28	99.89	6
230	Cherepovets	Russian Federation	36	154.08	100.81	65.43	1
231	Chicago Round Lake Beach-McHenry-Grayslake	United States of America	2687	3590.70	3332.20	92.80	16
232	Chifeng	China	116	251.65	241.14	95.82	11
233	Chihuahua	Mexico	230	415.11	397.38	95.73	13
234	Chimbote	Peru	21	169.03	87.01	51.48	10
235	Chita	Russian Federation	49	197.37	144.61	73.27	6
236	Chittagong	Bangladesh	135	248.81	248.73	99.97	4
237	Chitungwiza	Zimbabwe	65	162.89	139.76	85.80	7
238	Chongqing	China	648	912.99	902.33	98.83	14
239	Christchurch	New Zealand	204	336.50	336.36	99.96	16
240	Chukyo M.M.A. (Nagoya)	Japan	2316	2926.26	2926.25	100.00	16
241	Chuxiong	China	52	143.46	142.45	99.29	6
242	Chuzhou	China	96	205.79	203.77	99.02	14
243	Cincinnati	United States of America	151	655.63	284.68	43.42	2
244	Cirebon	Indonesia	63	337.80	156.88	46.44	5
245	Ciudad Bolivar	Venezuela (Bolivarian Republic of)	112	295.94	244.97	82.78	7
246	Ciudad de Guatemala (Guatemala City)	Guatemala	324	528.86	528.69	99.97	3
247	Ciudad del Este	Paraguay	248	419.42	418.74	99.84	2
248	Ciudad de México (Mexico City)	Mexico	2057	2529.97	2529.60	99.99	16

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
249	Ciudad Guayana	Venezuela (Bolivarian Republic of)	101	534.61	238.19	44.55	5
250	Ciudad Obregón	Mexico	88	175.28	175.20	99.95	10
251	Ciudad Victoria	Mexico	101	191.66	191.44	99.89	5
252	Cluj-Napoca	Romania	51	121.34	121.18	99.86	4
253	Coatzacoalcos	Mexico	30	241.82	78.73	32.56	8
254	Cochabamba	Bolivia (Plurinational State of)	276	501.60	470.22	93.74	16
255	Coimbatore	India	275	451.65	434.02	96.10	16
256	Colima	Mexico	88	180.49	164.67	91.24	7
257	Colombo	Sri Lanka	22	725.69	77.16	10.63	8
258	Colorado Springs	United States of America	444	639.25	639.13	99.98	11
259	Columbia, South Carolina	United States of America	395	855.20	684.44	80.03	15
260	Columbus, Ohio	United States of America	532	876.97	876.54	99.95	2
261	Comilla	Bangladesh	2	147.17	24.44	16.61	12
262	Concepción	Chile	38	286.07	96.87	33.86	3
263	San Francisco-Oakland San Jose Concord	United States of America	1482	3066.08	2266.79	73.93	16
264	Córdoba	Argentina	297	458.58	458.15	99.91	16
265	Corpus Christi	United States of America	201	522.90	383.76	73.39	16
266	Resistencia Corrientes	Argentina	148	350.29	319.10	91.10	5
267	Coventry-Bedworth	United Kingdom	86	218.86	218.67	99.91	5
268	Cuatla Morelos	Mexico	42	153.38	94.87	61.85	5
269	Cuenca	Ecuador	98	230.55	189.90	82.37	2
270	Cuiabá	Brazil	302	467.76	467.91	100.03	16
271	Culiacán	Mexico	179	303.49	303.34	99.95	10
272	Curitiba	Brazil	600	814.75	814.51	99.97	16
273	Cusco	Peru	29	147.38	75.48	51.22	3
274	Dafeng	China	70	157.26	148.44	94.39	15
275	Dakar	Senegal	51	364.95	148.05	40.57	13
276	Dali	China	39	166.86	118.54	71.04	6
277	Dalian	China	308	724.20	532.07	73.47	13
278	Dallas-Fort WorthDallas-Fort Worth Denton-Lewisville	United States of America	818	2680.61	1518.87	56.66	16
279	Da Nang	Viet Nam	101	298.44	262.84	88.07	7
280	Dandong Sinuiju	China	81	249.73	168.12	67.32	14
281	Danyang	China	83	172.18	172.12	99.97	11
282	Daqing	China	38	548.05	92.26	16.83	9
283	Darbhanga	India	2	132.91	23.92	18.00	6
284	Dar-el-Beida (Casablanca)	Morocco	192	362.35	336.32	92.82	11

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
285	Dar es Salaam	United Republic of Tanzania	429	661.12	661.15	100.00	16
286	Datong	China	155	335.12	314.51	93.85	11
287	Davangere	India	34	130.39	77.79	59.66	16
288	Huangshi Daye	China	59	477.48	142.17	29.78	14
289	Dayton	United States of America	74	195.99	195.87	99.94	2
290	Daytona Beach-Port Orange	United States of America	153	434.73	266.27	61.25	2
291	Dazhou	China	4	120.57	22.95	19.03	14
292	Deir El-Zor (Deir ez-Zor)	Syrian Arab Republic	5	103.36	33.11	32.03	9
293	Delhi	India	1000	1326.30	1325.99	99.98	4
294	Dengfeng	China	36	97.21	92.55	95.21	7
295	Dengzhou	China	48	122.70	103.59	84.43	4
296	Denizli	Turkey	71	165.48	164.55	99.44	11
297	Denver-Aurora	United States of America	1520	1897.50	1896.91	99.97	11
298	Dera Ghazikhan	Pakistan	49	120.38	104.94	87.18	10
299	Des Moines	United States of America	174	480.46	333.43	69.40	2
300	Detroit Windsor	United States of America	2837	3356.05	3355.54	99.98	2
301	Dezhou	China	122	252.49	252.39	99.96	4
302	Dhaka	Bangladesh	402	640.26	613.91	95.88	16
303	Dimashq (Damascus)	Syrian Arab Republic	315	499.06	498.85	99.96	13
304	Dindigul	India	19	118.27	58.84	49.75	10
305	Diwanayah	Iraq	67	150.72	150.74	100.01	9
306	Diyarbakir	Turkey	36	127.83	102.55	80.22	14
307	Dnipropetrovsk	Ukraine	218	374.48	374.42	99.98	1
308	Donetsk Makeyevka	Ukraine	405	613.84	605.17	98.59	4
309	Dongtai	China	49	129.47	129.11	99.73	15
310	Yiwu Dongyang	China	353	546.37	540.39	98.90	13
311	Dortmund	Germany	28	110.45	110.31	99.88	1
312	Douala	Cameroon	191	372.16	372.06	99.97	16
313	Dresden	Germany	103	224.64	224.59	99.98	1
314	Dublin	Ireland	288	492.01	487.19	99.02	2
315	Durango	Mexico	132	253.87	235.20	92.65	7
316	Durgapur	India	16	314.24	54.67	17.40	7
317	Raleigh Durham	United States of America	144	1127.92	293.18	25.99	6
318	Dushanbe	Tajikistan	227	389.90	389.50	99.90	11
319	East London	South Africa	30	204.34	86.18	42.18	2
320	Edmonton	Canada	420	678.94	671.96	98.97	16
321	Eindhoven	Netherlands	86	204.42	204.03	99.81	15

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
322	Elazig	Turkey	14	128.62	52.52	40.83	11
323	El Djelfa	Algeria	42	104.86	100.36	95.71	10
324	Eldoret	Kenya	43	168.01	100.90	60.06	15
325	El Tigre-San José de Guanipa	Venezuela (Bolivarian Republic of)	56	244.39	116.92	47.84	7
326	English Bazar	India	24	105.99	69.19	65.28	7
327	Enshi	China	25	115.27	77.80	67.49	6
328	Erduosi (Ordoss)	China	6	142.12	31.80	22.38	14
329	Erzurum	Turkey	23	111.34	72.72	65.32	11
330	Esfahan	Iran (Islamic Republic of)	382	823.50	571.31	69.38	13
331	Eskisehir	Turkey	86	211.17	184.69	87.46	7
332	TehranTehran Karaj Eslamshahr Malard Qods	Iran (Islamic Republic of)	1980	3454.56	2476.92	71.70	13
333	Ezhou	China	8	209.32	34.45	16.46	12
334	Faisalabad	Pakistan	210	338.19	338.12	99.98	9
335	Faloojah	Iraq	48	110.58	110.36	99.80	9
336	Farrukhabad	India	8	71.23	36.00	50.54	5
337	Fayetteville	United States of America	2	737.82	21.93	2.97	12
338	Fayetteville-Springdale	United States of America	80	453.49	167.28	36.89	15
339	Feicheng	China	41	107.44	107.45	100.01	14
340	Feira De Santana	Brazil	115	218.44	216.76	99.23	16
341	Firozabad	India	26	111.08	67.23	60.52	10
342	Flint	United States of America	158	302.90	285.93	94.40	2
343	Florence	Italy	81	466.83	199.54	42.74	15
344	Fortaleza	Brazil	414	608.73	608.49	99.96	16
345	Fort Wayne	United States of America	256	425.65	425.39	99.94	15
346	Franca	Brazil	127	236.96	236.87	99.96	7
347	Frankfurt am Main	Germany	175	468.28	367.70	78.52	15
348	Freetown	Sierra Leone	48	350.03	141.29	40.37	5
349	Fresno	United States of America	488	685.92	685.76	99.98	16
350	Fuan	China	16	101.41	53.96	53.21	6
351	Fuqing	China	105	1418.69	239.47	16.88	4
352	Fushun, Liaoning	China	117	346.98	298.05	85.90	14
353	Fuxin	China	111	268.86	234.91	87.37	11
354	Fuyang	China	115	225.98	225.72	99.89	4
355	Fuyang	China	95	210.33	210.12	99.90	6
356	Fuzhou, Fujian	China	384	574.79	574.35	99.92	4
357	Fuzhou, Jiangxi	China	63	178.00	158.56	89.08	14
358	Ganja	Azerbaijan	80	187.21	180.58	96.46	11

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
359	Ganzhou Nankang	China	329	910.09	539.63	59.29	4
360	Gaocheng	China	31	94.30	94.25	99.95	7
361	Gaoyou	China	58	134.33	134.15	99.87	4
362	Gaozhou	China	38	108.44	90.94	83.86	6
363	Gaya	India	41	117.46	99.37	84.60	10
364	Gaza (incl. Ash Shati Camp)	State of Palestine	67	221.55	155.20	70.05	11
365	Gboko	Nigeria	65	139.25	138.26	99.29	7
366	Gdańsk	Poland	85	271.37	209.16	77.08	1
367	Istanbul Gebze	Turkey	554	1441.48	992.85	68.88	4
368	General Santos City	Philippines	41	173.19	121.93	70.40	16
369	Geneva	Switzerland	23	127.33	120.13	94.34	1
370	Genova (Genoa)	Italy	14	111.43	54.68	49.07	1
371	Glasgow	United Kingdom	301	531.77	477.74	89.84	2
372	Goiânia	Brazil	630	861.95	861.71	99.97	16
373	Gold Coast	Australia	66	431.06	141.74	32.88	2
374	Gombe	Nigeria	85	168.40	167.41	99.41	10
375	Gomel	Belarus	84	192.84	192.77	99.96	15
376	Gongyi	China	59	140.61	132.96	94.56	11
377	Gongzhuling	China	32	93.57	93.40	99.82	7
378	Gorakhpur	India	91	203.69	200.79	98.57	16
379	Gorgan	Iran (Islamic Republic of)	46	116.54	116.22	99.73	5
380	Göteborg	Sweden	53	196.83	196.64	99.91	1
381	Grand Rapids	United States of America	313	501.66	501.40	99.95	2
382	Greensboro	United States of America	54	321.85	180.28	56.01	2
383	Greenville	United States of America	50	553.08	118.84	21.49	12
384	Grenoble	France	63	139.97	139.95	99.98	15
385	Grodno	Belarus	53	151.64	142.20	93.78	15
386	Guadalajara	Mexico	678	904.13	904.04	99.99	11
387	Guang'an	China	25	128.19	67.01	52.27	14
388	Guangyuan	China	45	165.31	104.49	63.21	14
389	Guayaquil	Ecuador	358	581.72	581.20	99.91	10
390	Guigang	China	78	169.26	169.24	99.99	4
391	Guilin	China	169	337.28	296.16	87.81	4
392	Guiping	China	4	117.22	36.66	31.28	14
393	Guiyang	China	167	403.98	346.40	85.75	6
394	Gujranwala	Pakistan	113	215.05	215.02	99.99	10
395	Gujrat	Pakistan	46	119.29	113.84	95.43	3

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
396	Gulbarga	India	77	164.62	153.94	93.51	10
397	Guntur	India	50	138.26	97.61	70.60	5
398	Guwahati (Gauhati)	India	179	318.58	292.58	91.84	3
399	Gwalior	India	121	228.85	228.81	99.98	10
400	Gwangju	Republic of Korea	117	360.87	214.07	59.32	6
401	Habra	India	42	117.97	108.87	92.28	15
402	Haerbin	China	246	395.69	395.53	99.96	4
403	Hafar al-Batin	Saudi Arabia	137	276.47	238.33	86.20	9
404	Haikou	China	165	526.61	309.46	58.76	1
405	Ha'il	Saudi Arabia	180	332.11	328.82	99.01	9
406	Nantong Haimen	China	707	1729.34	1101.53	63.70	1
407	Hải Phòng	Viet Nam	206	609.35	359.43	58.99	5
408	Halab (Aleppo)	Syrian Arab Republic	169	297.20	297.02	99.94	13
409	Halifax	Canada	19	260.90	69.55	26.66	12
410	Hamadan	Iran (Islamic Republic of)	98	185.35	184.81	99.71	13
411	Hamah	Syrian Arab Republic	59	133.36	131.93	98.93	11
412	Hamburg	Germany	154	290.92	290.80	99.96	1
413	Hamhung	Dem. People's Republic of Korea	34	147.71	92.81	62.83	9
414	Hami	China	64	169.94	147.21	86.62	13
415	Toronto Hamilton	Canada	1897	2639.85	2565.43	97.18	4
416	Handan	China	195	415.00	358.00	86.26	11
417	Hannover	Germany	75	207.62	207.43	99.91	1
418	Hanoi	Viet Nam	302	537.16	536.83	99.94	16
419	Hanzhong	China	78	208.43	169.89	81.51	14
420	Harare	Zimbabwe	453	680.25	678.61	99.76	11
421	Hargeysa	Somalia	79	156.70	156.55	99.90	9
422	Harrisburg	United States of America	204	404.16	384.93	95.24	2
423	Hartford	United States of America	242	559.92	452.39	80.80	2
424	Hebi	China	60	140.92	136.57	96.91	1
425	Hefa (Haifa)	Israel	58	216.96	120.31	55.45	11
426	Hefei	China	562	818.12	818.00	99.99	4
427	Hegang	China	83	240.79	175.63	72.94	11
428	Hejian	China	35	92.28	92.13	99.84	7
429	Helsinki	Finland	47	353.68	132.18	37.37	1
430	Hengshui	China	74	165.90	165.77	99.92	11
431	Hengyang	China	184	333.19	332.90	99.91	1
432	Herat	Afghanistan	73	154.20	154.10	99.94	9

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
433	Hermosillo	Mexico	211	356.72	356.74	100.01	13
434	Heyuan	China	58	138.23	137.07	99.16	4
435	Heze	China	119	257.28	230.93	89.76	11
436	Hezhou	China	8	165.54	36.92	22.30	6
437	Hims (Homs)	Syrian Arab Republic	79	158.54	158.46	99.95	11
438	Hiroshima	Japan	212	437.61	432.70	98.88	3
439	Hisar	India	38	137.93	89.37	64.79	10
440	Hohhot	China	307	487.47	487.23	99.95	11
441	Holguin	Cuba	28	116.72	77.15	66.10	5
442	Hong Kong	China, Hong Kong SAR	19	378.56	87.53	23.12	6
443	Honolulu	United States of America	143	737.13	288.12	39.09	2
444	Hosur	India	58	182.66	131.21	71.83	7
445	Houston The Woodlands	United States of America	2865	3543.00	3542.69	99.99	4
446	Hsinchu	China	104	219.59	218.03	99.29	1
447	Huai'an	China	207	460.36	363.28	78.91	4
448	Huaibei	China	177	325.36	320.10	98.38	4
449	Huaihua	China	59	139.74	129.30	92.53	14
450	Huainan	China	118	473.64	256.58	54.17	15
451	Huambo	Angola	88	186.41	186.15	99.86	7
452	Huancayo	Peru	118	284.85	228.31	80.15	7
453	Huangshan	China	47	223.77	106.54	47.61	6
454	Hue	Viet Nam	60	148.68	124.93	84.03	16
455	Hufuf-Mubarraz	Saudi Arabia	346	665.38	515.94	77.54	13
456	Huizhou	China	132	297.81	254.40	85.42	4
457	Huntsville	United States of America	77	429.47	169.56	39.48	2
458	Huzhou	China	100	332.76	206.36	62.02	14
459	Hyderabad	India	763	1002.35	1002.57	100.02	16
460	Ibagué	Colombia	14	130.08	53.56	41.18	3
461	Ibb	Yemen	25	108.83	73.45	67.49	13
462	Ife	Nigeria	118	217.41	217.25	99.92	15
463	Imphal	India	85	177.15	177.01	99.92	7
464	Indianapolis	United States of America	714	999.29	997.59	99.83	2
465	Indio-Cathedral City-Palm Springs	United States of America	194	494.20	343.43	69.49	9
466	Ipoh	Malaysia	338	541.00	533.64	98.64	3
467	Iquique	Chile	9	132.40	52.89	39.95	9
468	Iquitos	Peru	32	131.87	77.83	59.02	8
469	Irbid	Jordan	113	216.07	214.77	99.40	14

continued on next page



ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
470	Irbil (Erbil)	Iraq	174	349.43	287.04	82.15	9
471	Irkutsk	Russian Federation	153	316.27	292.14	92.37	4
472	Rawalpindi Islamabad	Pakistan	488	738.86	738.79	99.99	10
473	Ivanovo	Russian Federation	89	170.91	170.75	99.91	15
474	Izhevsk	Russian Federation	76	211.38	198.55	93.93	4
475	Izmir	Turkey	226	463.81	462.71	99.76	4
476	Izmit	Turkey	40	298.87	151.65	50.74	11
477	Jabalpur	India	130	248.61	248.72	100.04	7
478	Jackson, Mississippi	United States of America	30	326.57	90.32	27.66	15
479	Jacksonville, Florida	United States of America	598	908.16	894.86	98.54	2
480	Jaipur	India	328	540.83	532.60	98.48	16
481	Jalandhar	India	127	239.40	239.28	99.95	7
482	Jalgaon	India	61	138.27	138.34	100.05	10
483	Jalna	India	33	124.47	88.50	71.10	10
484	Jambi	Indonesia	142	269.56	247.61	91.86	5
485	Jammu	India	134	272.25	270.33	99.29	11
486	Jamshedpur	India	186	338.17	331.27	97.96	7
487	Jhang	Pakistan	40	131.08	92.79	70.79	10
488	Jiamusi	China	50	172.64	117.76	68.21	4
489	Ji'an, Jiangxi	China	60	149.25	127.36	85.33	1
490	Jianyang	China	5	83.06	34.19	41.16	6
491	Jiaxing	China	136	257.47	251.30	97.60	4
492	Jiddah	Saudi Arabia	582	825.63	825.05	99.93	9
493	Jilin	China	146	287.10	286.86	99.92	11
494	Ji'nan, Shandong Zhangqiu	China	717	1253.52	1090.62	87.00	11
495	Jincheng	China	73	188.56	155.60	82.52	11
496	Jingdezhen	China	98	228.01	202.90	88.99	6
497	Jingmen	China	66	222.55	147.03	66.07	1
498	Quanzhou Shishi Jinjiang	China	1268	1750.28	1741.02	99.47	4
499	Taiyuan, Shanxi Jinzhong	China	625	890.31	890.28	100.00	13
500	Jiujiang	China	152	307.26	295.46	96.16	1
501	Jixi, Heilongjiang	China	48	157.43	143.14	90.92	11
502	João Pessoa	Brazil	193	358.37	336.53	93.91	16
503	Jodhpur	India	180	338.87	302.51	89.27	2
504	Johannesburg	South Africa	1361	1797.14	1797.38	100.01	13
505	Jos	Nigeria	153	354.37	273.78	77.26	10
506	Juazeiro Do Norte	Brazil	62	210.82	140.73	66.75	5

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
507	Juba	South Sudan	90	186.18	171.51	92.12	16
508	Jubayl	Saudi Arabia	4	484.13	34.59	7.14	9
509	Juiz De Fora	Brazil	104	237.03	205.62	86.75	16
510	Junagadh	India	35	132.61	89.63	67.59	15
511	Jurong	China	43	109.47	107.63	98.32	6
512	Kabul	Afghanistan	373	602.15	562.77	93.46	9
513	Kadapa	India	33	115.66	74.19	64.14	5
514	Kaduna	Nigeria	301	452.90	453.00	100.02	7
515	Kagoshima	Japan	121	280.17	239.65	85.54	16
516	Kaifeng	China	121	254.82	242.94	95.34	4
517	Kaili	China	41	115.63	90.85	78.57	11
518	Taishan Kaiping	China	174	456.66	362.38	79.35	7
519	RajahmundryRajahmundry Kakinada	India	94	1076.10	221.83	20.61	5
520	Kaliningrad, Kaliningrad Oblast	Russian Federation	114	208.89	208.61	99.87	1
521	Kampala	Uganda	424	610.04	609.94	99.98	16
522	Kananga	Democratic Republic of the Congo	79	163.46	160.02	97.89	5
523	Kanazawa	Japan	166	377.67	335.99	88.96	16
524	Kandahar	Afghanistan	90	302.48	167.95	55.52	7
525	Kannur	India	29	140.70	83.88	59.62	2
526	Kano	Nigeria	284	447.08	446.94	99.97	10
527	Kanpur	India	217	390.48	384.07	98.36	7
528	Kansas City	United States of America	140	383.54	256.39	66.85	2
529	KaohsiungKaohsiung Tainan	China	653	2437.78	904.13	37.09	15
530	Karachi	Pakistan	738	1022.48	1022.22	99.97	9
531	Karaganda	Kazakhstan	69	164.96	158.20	95.90	11
532	Karbala	Iraq	88	186.37	186.13	99.87	9
533	Kassala	Sudan	99	249.63	179.27	71.81	10
534	Kathmandu	Nepal	162	270.69	270.50	99.93	15
535	Katowice	Poland	41	243.88	133.24	54.63	14
536	Katsina	Nigeria	88	174.12	172.69	99.18	13
537	Thiruvananthapuram Kollam Kayamkulam	India	353	2152.26	580.56	26.97	2
538	Kayseri	Turkey	134	286.72	252.94	88.22	13
539	Kazan	Russian Federation	182	372.27	365.77	98.25	1
540	Keelung	China	47	150.22	143.85	95.76	6
541	Kelamayi	China	38	100.10	98.51	98.41	13
542	Kemerovo	Russian Federation	89	256.28	184.57	72.02	1
543	Kenitra	Morocco	40	165.48	99.83	60.33	13

continued on next page

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
544	Kerman	Iran (Islamic Republic of)	146	268.33	254.20	94.73	9
545	Kermanshah	Iran (Islamic Republic of)	107	241.25	211.00	87.46	13
546	Khamis Mushayt	Saudi Arabia	183	624.89	327.91	52.47	9
547	Kharkiv	Ukraine	161	409.97	270.49	65.98	1
548	Khorramabad	Iran (Islamic Republic of)	22	161.46	71.81	44.48	9
549	Khulna	Bangladesh	39	151.05	94.86	62.80	16
550	Kigali	Rwanda	138	270.65	236.95	87.55	16
551	Kikwit	Democratic Republic of the Congo	64	149.73	146.63	97.93	5
552	Kingston	Jamaica	127	334.72	277.03	82.76	16
553	Kinki M.M.A. (Osaka)	Japan	2046	2723.23	2723.29	100.00	16
554	Kirkuk	Iraq	138	249.94	249.73	99.91	13
555	Kirov	Russian Federation	68	196.47	158.56	80.70	1
556	Kisangani	Democratic Republic of the Congo	85	186.03	172.81	92.90	2
557	Kismaayo	Somalia	2	77.02	26.29	34.13	9
558	Orlando Kissimmee	United States of America	1648	2158.66	2148.82	99.54	16
559	Kitchener	Canada	258	561.85	407.46	72.52	16
560	Knoxville	United States of America	106	343.28	219.50	63.94	2
561	Kolwezi	Democratic Republic of the Congo	49	305.83	104.69	34.23	10
562	Konya	Turkey	173	293.18	293.03	99.95	13
563	Kota	India	120	240.14	239.98	99.94	7
564	Kota Kinabalu	Malaysia	89	251.75	208.21	82.70	16
565	Kottayam	India	96	224.55	200.18	89.15	15
566	Kraków	Poland	47	139.48	139.24	99.83	1
567	Krasnodar	Russian Federation	186	400.93	355.27	88.61	1
568	Krasnoyarsk	Russian Federation	59	354.14	148.46	41.92	14
569	Krivoi Rog	Ukraine	222	399.87	378.08	94.55	11
570	Kuala Lumpur	Malaysia	1620	1996.46	1996.62	100.01	16
571	Kuantan	Malaysia	89	198.92	181.49	91.24	16
572	Kuerle	China	148	304.17	288.05	94.70	13
573	Kumamoto	Japan	271	450.95	450.91	99.99	16
574	Kumasi	Ghana	517	704.47	704.08	99.94	15
575	Kunming	China	455	691.29	691.18	99.98	11
576	Shanghai Kunshan Taicang	China	2837	3481.52	3480.96	99.98	4
577	Kurgan	Russian Federation	68	152.84	148.96	97.46	4
578	Kurnool	India	57	145.34	119.49	82.21	7
579	Kut	Iraq	66	145.74	134.72	92.44	13
580	Kyiv (Kiev)	Ukraine	387	604.42	604.03	99.94	1

continued on next page

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
581	La Habana (Havana)	Cuba	178	321.84	316.89	98.46	2
582	Laibin	China	55	128.79	126.91	98.54	11
583	Laixi	China	81	180.98	180.74	99.87	4
584	Laiyang	China	93	203.06	174.65	86.01	11
585	Lampang	Thailand	68	145.06	144.97	99.94	2
586	Lancaster-Palmdale	United States of America	238	407.32	406.81	99.88	10
587	Lansing	United States of America	38	183.69	178.72	97.29	2
588	Lanzhou	China	38	479.37	99.13	20.68	14
589	La Paz	Bolivia (Plurinational State of)	263	441.39	430.10	97.44	9
590	La Plata	Argentina	143	263.09	263.04	99.98	16
591	La Serena-Coquimbo	Chile	61	243.12	149.31	61.42	10
592	Las Palmas Gran Canaria	Spain	27	104.43	78.28	74.96	13
593	Las Vegas	United States of America	1160	1461.87	1461.60	99.98	13
594	Lattakia	Syrian Arab Republic	20	102.13	84.12	82.36	1
595	Latur	India	44	138.07	96.22	69.69	5
596	Lausanne	Switzerland	23	192.91	99.23	51.44	1
597	Leicester	United Kingdom	105	213.93	213.60	99.85	2
598	Leipzig	Germany	47	149.65	149.50	99.90	1
599	Leiyang	China	48	124.16	124.08	99.93	1
600	Leping	China	55	145.47	136.04	93.52	7
601	Lexington-Fayette	United States of America	231	360.48	360.25	99.94	2
602	Liaocheng	China	103	222.56	222.32	99.89	4
603	Liaoyuan	China	44	126.15	125.96	99.85	6
604	Liège	Belgium	118	243.95	243.79	99.93	2
605	Likasi	Democratic Republic of the Congo	58	153.64	128.39	83.57	7
606	Liling	China	65	140.92	135.81	96.37	15
607	Lilongwe	Malawi	279	453.24	420.68	92.82	16
608	Lin'an	China	27	120.52	79.49	65.96	6
609	Linfen	China	74	184.02	178.64	97.07	11
610	Linhai	China	71	220.08	186.11	84.57	6
611	Linqing	China	46	111.74	111.54	99.82	7
612	Linyi, Shandong	China	479	712.83	712.67	99.98	4
613	Linzhou	China	57	151.74	151.64	99.93	4
614	Lipa City	Philippines	10	95.26	39.00	40.94	15
615	Lisboa (Lisbon)	Portugal	407	873.15	781.19	89.47	1
616	Lishui, Zhejiang	China	26	108.09	77.16	71.39	6
617	Little Rock	United States of America	105	378.04	266.46	70.48	2

continued on next page

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
618	Liuan	China	82	226.16	180.78	79.94	14
619	Liupanshui	China	16	141.21	60.16	42.60	6
620	Liuzhou	China	267	425.84	425.58	99.94	4
621	Liyang	China	69	158.69	153.37	96.65	11
622	London	United Kingdom	834	1205.49	1205.03	99.96	2
623	London	Canada	239	377.95	377.60	99.91	4
624	Londrina	Brazil	165	291.57	291.39	99.94	16
625	Xiamen Zhangzhou Longhai	China	395	1771.09	818.58	46.22	4
626	Longyan	China	99	241.62	221.97	91.87	6
627	Los Angeles-Long Beach-Santa Ana Riverside-San Bernardino Mission Viejo	United States of America	5534	6680.38	6680.55	100.00	16
628	Loudi	China	77	165.27	165.32	100.03	14
629	Louisville	United States of America	293	496.76	491.85	99.01	2
630	Luanda	Angola	347	844.71	565.03	66.89	8
631	Lubango	Angola	68	188.80	140.51	74.42	13
632	Lublin	Poland	71	157.63	155.66	98.75	15
633	Lubumbashi	Democratic Republic of the Congo	243	378.86	378.68	99.95	13
634	Lucknow	India	284	459.14	458.64	99.89	13
635	Ludhiana	India	209	336.20	336.13	99.98	7
636	Lugansk	Ukraine	111	202.93	202.83	99.95	4
637	Luoding	China	26	111.68	71.94	64.42	1
638	Luohe	China	97	194.84	194.73	99.94	4
639	Luoyang Yanshi	China	650	1090.21	964.86	88.50	4
640	Lusaka	Zambia	350	514.07	513.61	99.91	16
641	Luzhou	China	40	216.32	94.13	43.51	14
642	Lvliang	China	30	109.90	91.26	83.04	9
643	Lvov	Ukraine	72	170.56	170.39	99.90	1
644	Lyon	France	168	307.58	307.47	99.97	4
645	Zhuhai Macao	China	206	462.62	360.06	77.83	4
646	Macaé	Brazil	50	189.71	119.35	62.91	7
647	Macheng	China	32	92.52	78.04	84.35	15
648	Madison	United States of America	20	267.23	119.39	44.68	2
649	Madrid	Spain	541	801.13	801.09	99.99	11
650	Madurai	India	133	234.14	234.15	100.00	7
651	Magnitogorsk	Russian Federation	135	300.57	281.00	93.49	11
652	Maiduguri	Nigeria	186	314.11	313.42	99.78	7
653	Makhachkala	Russian Federation	33	252.04	109.28	43.36	14

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
654	Makkah (Mecca)	Saudi Arabia	274	469.50	468.78	99.85	9
655	Makurdi	Nigeria	28	214.95	79.31	36.90	8
656	Malaga	Spain	47	220.28	130.27	59.14	11
657	Malatya	Turkey	45	118.55	113.68	95.89	11
658	Managua	Nicaragua	130	242.73	242.67	99.98	16
659	Manaus	Brazil	339	520.19	501.87	96.48	3
660	Manchester	United Kingdom	372	611.71	611.45	99.96	2
661	Mandalay	Myanmar	145	281.63	256.49	91.07	16
662	Mangalore	India	121	310.43	216.07	69.60	2
663	Manisa	Turkey	6	143.76	32.24	22.43	6
664	Manizales	Colombia	9	127.61	40.50	31.74	3
665	Mardan	Pakistan	31	288.26	81.61	28.31	3
666	Mar Del Plata	Argentina	95	198.79	198.77	99.99	5
667	Maringá	Brazil	170	280.47	280.21	99.91	16
668	Mariupol	Ukraine	28	312.91	94.96	30.35	1
669	Marrakech	Morocco	146	253.31	253.14	99.93	13
670	Mashhad	Iran (Islamic Republic of)	326	515.31	512.20	99.40	13
671	Matadi	Democratic Republic of the Congo	38	126.98	87.03	68.54	7
672	Mataram	Indonesia	88	179.00	177.36	99.08	15
673	Maturin	Venezuela (Bolivarian Republic of)	139	309.43	292.62	94.57	5
674	Maunath Bhanjan	India	12	95.49	47.97	50.23	7
675	Mawlamyine	Myanmar	4	142.99	25.48	17.82	8
676	Mazatlán	Mexico	88	200.52	200.08	99.78	10
677	Mbandaka	Democratic Republic of the Congo	36	126.79	78.00	61.52	5
678	Mbeya	United Republic of Tanzania	79	263.81	159.86	60.60	15
679	Mbuji-Mayi	Democratic Republic of the Congo	4	385.47	27.04	7.01	5
680	McAllen Reynosa	United States of America	615	1063.88	911.96	85.72	16
681	Medan	Indonesia	440	715.68	711.03	99.35	5
682	Medellín	Colombia	168	378.79	303.09	80.02	3
683	Meerut	India	93	205.70	205.43	99.87	3
684	Meishan	China	75	164.25	155.16	94.46	1
685	Meizhou	China	101	196.97	196.96	100.00	6
686	Mekele	Ethiopia	38	124.44	109.47	87.97	10
687	Meknès	Morocco	81	195.97	173.67	88.62	13
688	Melbourne	Australia	1698	2152.47	2152.61	100.01	16
689	Memphis	United States of America	143	947.08	264.55	27.93	2
690	Mendoza	Argentina	249	391.07	390.91	99.96	11

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
691	Merca	Somalia	1	60.94	22.30	36.59	9
692	Mersin	Turkey	70	404.48	176.27	43.58	4
693	Mexicali	Mexico	262	428.37	428.27	99.98	10
694	Miami	United States of America	1355	2493.56	1764.20	70.75	2
695	Mianyang, Sichuan	China	148	360.66	263.74	73.13	1
696	Miluo	China	16	84.77	51.86	61.18	14
697	Milwaukee	United States of America	576	841.30	787.57	93.61	2
698	Minatitlán	Mexico	78	154.78	154.78	100.00	5
699	Minna	Nigeria	106	213.91	209.68	98.02	10
700	Minsk	Belarus	254	419.31	419.24	99.98	1
701	Misratah	Libya	157	328.14	280.71	85.55	13
702	Mobile	United States of America	39	325.20	130.64	40.17	2
703	Mogilev	Belarus	95	199.53	192.36	96.41	1
704	Mombasa	Kenya	113	269.87	264.24	97.91	3
705	Monrovia	Liberia	32	399.71	85.60	21.42	5
706	Monteria	Colombia	53	144.30	103.07	71.43	3
707	Montes Claros	Brazil	88	179.74	178.81	99.48	10
708	Montevideo	Uruguay	215	396.04	378.47	95.56	16
709	Montpellier	France	135	322.20	320.47	99.46	4
710	Monywa	Myanmar	27	122.24	72.47	59.28	2
711	Moradabad	India	56	141.55	130.43	92.15	3
712	Morelia	Mexico	129	233.23	232.87	99.85	16
713	Morogoro	United Republic of Tanzania	74	191.78	158.61	82.70	5
714	Moskva (Moscow)	Russian Federation	648	947.72	947.19	99.94	1
715	Mudanjiang	China	81	181.49	180.83	99.64	4
716	Muenster (Westfalen)	Germany	48	137.12	136.89	99.83	15
717	Multan	Pakistan	170	284.18	283.92	99.91	10
718	Muqdisho (Mogadishu)	Somalia	107	228.78	215.97	94.40	13
719	Muzaffarpur	India	24	119.40	77.56	64.96	16
720	Mwanza	United Republic of Tanzania	42	230.25	95.24	41.36	2
721	Mymensingh	Bangladesh	45	106.08	101.74	95.91	5
722	Mysore	India	159	291.59	282.91	97.02	7
723	Nagasaki	Japan	42	275.46	108.69	39.46	6
724	Nagpur	India	224	378.15	358.49	94.80	10
725	Naha	Japan	161	535.00	317.49	59.34	1
726	Nairobi	Kenya	455	694.67	694.34	99.95	16
727	Najran	Saudi Arabia	67	441.32	140.75	31.89	9

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
728	Nakhon Ratchasima	Thailand	115	242.04	238.38	98.49	15
729	Nakuru	Kenya	41	184.63	95.03	51.47	16
730	Namangan	Uzbekistan	167	340.41	298.72	87.75	16
731	Nampula	Mozambique	87	172.70	171.85	99.51	15
732	Nanchong	China	85	229.98	188.67	82.04	14
733	Nanded Waghala	India	53	155.58	112.41	72.25	7
734	Nanjing, Jiangsu	China	848	1179.43	1179.32	99.99	4
735	Nanning	China	390	617.50	597.84	96.82	4
736	Nantes	France	223	392.31	392.22	99.98	2
737	Nanyang, Henan	China	136	241.46	241.35	99.95	4
738	Nashik	India	182	329.74	319.22	96.81	7
739	Nashville-Davidson	United States of America	355	731.84	578.91	79.10	2
740	Nasiriyah	Iraq	65	168.70	132.74	78.68	9
741	Natal	Brazil	45	448.53	114.30	25.48	8
742	Navsari	India	44	119.23	106.07	88.96	5
743	N'Djamena	Chad	268	451.37	411.56	91.18	7
744	Ndola	Zambia	118	253.28	227.41	89.79	7
745	Neijiang	China	4	120.64	35.07	29.07	14
746	Neiva	Colombia	60	139.32	138.69	99.55	3
747	Nellore	India	50	155.13	128.14	82.60	5
748	Neuquén -Plottier-Cipolletti	Argentina	64	212.88	141.50	66.47	13
749	Newcastle and Lake Macquarie	Australia	116	435.00	233.95	53.78	16
750	Newcastle upon Tyne Sunderland	United Kingdom	284	457.21	452.96	99.07	2
751	New Orleans	United States of America	475	716.72	716.49	99.97	16
752	New York-Newark	United States of America	1549	2320.03	2319.46	99.98	2
753	Niamey	Niger	203	343.79	342.85	99.73	10
754	Nice-Cannes	France	101	517.95	184.71	35.66	15
755	Niigata	Japan	111	348.96	216.37	62.01	10
756	Nikolaev	Ukraine	67	191.08	181.34	94.90	1
757	Ningbo	China	884	1320.57	1174.74	88.96	4
758	Nizamabad	India	42	105.88	101.36	95.73	7
759	Nizhniy Novgorod	Russian Federation	4	442.84	22.26	5.03	8
760	Nizhny Tagil	Russian Federation	61	196.42	146.04	74.35	14
761	Nottingham	United Kingdom	131	255.71	251.21	98.24	2
762	Nouakchott	Mauritania	166	291.78	286.52	98.20	9
763	Novokuznetsk	Russian Federation	37	130.87	101.72	77.73	14
764	Novosibirsk	Russian Federation	173	375.98	297.68	79.18	4

*continued on next page*



ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
765	Nuevo Laredo	Mexico	241	420.94	399.56	94.92	10
766	Nurenbeg	Germany	167	302.87	291.28	96.17	15
767	Nyala	Sudan	142	248.91	242.49	97.42	7
768	Nzérékoré	Guinea	47	123.93	109.69	88.51	7
769	Oaxaca de Juárez	Mexico	144	273.40	271.28	99.22	16
770	Łódź	Poland	95	213.08	212.59	99.77	1
771	Odesa	Ukraine	118	440.15	289.15	65.69	4
772	Ogden-Layton	United States of America	195	499.28	348.33	69.77	16
773	Okara	Pakistan	20	99.77	69.58	69.74	10
774	Okayama	Japan	356	1676.95	559.78	33.38	16
775	Okene	Nigeria	45	191.35	111.76	58.40	7
776	Oklahoma City	United States of America	725	1038.81	1038.78	100.00	11
777	Omaha	United States of America	280	664.10	490.68	73.89	2
778	Omsk	Russian Federation	161	384.49	301.66	78.46	4
779	Ondo	Nigeria	79	158.83	149.49	94.12	3
780	Orenburg	Russian Federation	97	219.80	211.95	96.43	4
781	Orumiyeh	Iran (Islamic Republic of)	122	287.81	242.44	84.24	10
782	Oshawa	Canada	117	247.76	218.39	88.14	16
783	Oslo	Norway	80	233.15	201.54	86.44	2
784	Ottawa-Gatineau	Canada	308	487.69	487.77	100.02	2
785	Ouagadougou	Burkina Faso	430	604.92	604.67	99.96	10
786	Oujda	Morocco	84	165.37	165.39	100.01	10
787	Oxnard	United States of America	136	268.09	266.09	99.25	10
788	Pachuca de Soto	Mexico	104	294.49	186.39	63.29	7
789	Padang	Indonesia	75	229.98	168.40	73.22	16
790	Palakkad	India	62	162.62	125.64	77.26	2
791	Palembang	Indonesia	10	525.90	40.56	7.71	12
792	Palermo	Italy	91	256.53	217.64	84.84	1
793	Palma	Spain	58	200.97	150.89	75.08	1
794	Palm Bay-Melbourne	United States of America	6	771.51	33.80	4.38	12
795	Panjin	China	166	367.25	324.08	88.24	4
796	Parbhani	India	30	115.33	73.43	63.67	5
797	Paris	France	886	1161.10	1160.78	99.97	2
798	Parma	Italy	46	118.04	117.84	99.83	15
799	Pasto	Colombia	41	121.73	99.13	81.43	15
800	Pavlodar	Kazakhstan	67	211.08	145.80	69.07	4
801	Pekan Baru	Indonesia	257	442.27	406.20	91.84	15

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
802	Pelotas	Brazil	70	161.70	148.36	91.75	16
803	Pensacola	United States of America	164	542.46	329.03	60.65	2
804	Penza	Russian Federation	112	233.03	227.97	97.83	4
805	Pereira	Colombia	54	207.83	118.57	57.05	3
806	Perm	Russian Federation	65	288.35	158.53	54.98	1
807	Pescara	Italy	21	114.10	106.10	92.99	1
808	Peshawar	Pakistan	154	341.06	288.64	84.63	10
809	Petrolina	Brazil	4	214.01	27.04	12.64	8
810	Phnom Penh (Phnom Penh)	Cambodia	194	349.96	338.19	96.64	10
811	Phoenix-Mesa	United States of America	2875	3405.15	3405.04	100.00	13
812	Pingdu	China	102	195.87	195.72	99.92	4
813	Pinghu	China	63	137.28	137.12	99.89	15
814	Pingxiang, Jiangxi	China	77	173.33	160.95	92.86	6
815	Piracicaba	Brazil	133	257.84	257.72	99.96	16
816	Pittsburgh	United States of America	108	232.63	232.71	100.03	2
817	Piura	Peru	76	173.95	167.17	96.10	10
818	Plovdiv	Bulgaria	66	138.53	138.47	99.96	4
819	Pohang	Republic of Korea	81	210.65	192.26	91.27	4
820	Pointe-Noire	Congo	133	255.06	252.81	99.12	10
821	Pokhara	Nepal	2	128.13	23.97	18.71	6
822	Ponta Grossa	Brazil	146	266.09	265.91	99.93	15
823	Pontianak	Indonesia	113	224.16	223.89	99.88	3
824	Port-au-Prince	Haiti	215	399.79	376.02	94.06	16
825	Port Elizabeth	South Africa	235	407.02	406.91	99.97	16
826	Portland	United States of America	718	961.59	961.60	100.00	2
827	Porto	Portugal	341	569.43	569.18	99.96	2
828	Porto Alegre	Brazil	2	763.63	23.06	3.02	12
829	Port St. Lucie	United States of America	410	891.69	625.19	70.11	2
830	Port Sudan (Bur Sudan)	Sudan	145	277.22	270.48	97.57	9
831	Posadas	Argentina	70	273.92	210.34	76.79	16
832	Poza Rica de Hidalgo	Mexico	90	179.63	175.35	97.62	5
833	Pózná	Poland	104	270.07	267.52	99.06	15
834	Praha (Prague)	Czech Republic	157	286.35	286.28	99.98	1
835	Preston	United Kingdom	55	136.51	136.25	99.81	2
836	Pretoria Soshanguve	South Africa	967	1461.14	1440.20	98.57	10
837	Providence	United States of America	173	331.38	331.16	99.93	2
838	Provo-Orem	United States of America	100	529.45	211.80	40.00	13

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
839	Puducherry	India	50	133.69	123.41	92.31	5
840	Puebla Tlaxcala	Mexico	713	1235.29	1017.34	82.36	16
841	Pulandian	China	33	98.76	75.83	76.78	11
842	Pune (Poona)	India	590	823.50	823.13	99.95	11
843	Punto Fijo	Venezuela (Bolivarian Republic of)	97	239.52	205.08	85.62	9
844	Purnia	India	20	105.09	70.30	66.89	5
845	Putian	China	209	904.83	364.15	40.25	4
846	Puyang	China	105	217.37	217.26	99.95	4
847	Pyongyang	Dem. People's Republic of Korea	228	379.50	379.38	99.97	13
848	Qacentina	Algeria	88	325.63	172.33	52.92	11
849	Qazvin	Iran (Islamic Republic of)	42	117.60	112.56	95.72	13
850	Qinhuangdao	China	147	273.63	273.49	99.95	11
851	Qinzhou	China	68	146.15	140.65	96.24	1
852	Qiqihaer	China	113	218.69	218.24	99.79	11
853	Qom	Iran (Islamic Republic of)	123	250.08	230.38	92.12	13
854	Québec	Canada	235	495.85	484.25	97.66	4
855	Quetta	Pakistan	140	266.51	265.23	99.52	9
856	Quito	Ecuador	371	836.82	549.62	65.68	16
857	Qujing	China	96	238.79	200.44	83.94	6
858	Quzhou	China	10	265.94	36.19	13.61	14
859	Rabat	Morocco	132	304.67	272.64	89.49	11
860	Rahim Yar Khan	Pakistan	55	126.94	126.94	100.00	7
861	Rajkot	India	120	218.67	217.43	99.43	11
862	Rajshahi	Bangladesh	35	118.64	82.66	69.67	1
863	Ramadi	Iraq	56	166.96	144.96	86.82	9
864	Rampur	India	33	104.62	86.19	82.38	5
865	Rangpur	Bangladesh	4	107.17	24.48	22.84	12
866	Rasht	Iran (Islamic Republic of)	73	157.26	157.11	99.90	10
867	Raurkela	India	51	203.41	109.53	53.85	7
868	Rayong	Thailand	91	350.88	187.57	53.46	16
869	Reading-Wokingham	United Kingdom	11	114.93	43.62	37.95	5
870	Recife	Brazil	293	544.92	463.75	85.10	16
871	Rennes	France	47	130.16	130.14	99.98	1
872	Reno	United States of America	340	524.66	524.47	99.96	13
873	Renqiu	China	74	184.70	184.47	99.88	7
874	Richmond	United States of America	94	282.61	266.30	94.23	2
875	Riga	Latvia	75	204.47	186.71	91.31	1

continued on next page

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
876	Rio Branco	Brazil	108	236.91	198.04	83.59	5
877	Rio de Janeiro	Brazil	110	1509.26	276.53	18.32	16
878	Rizhao	China	183	642.52	377.15	58.70	11
879	Rochester	United States of America	265	439.17	434.61	98.96	2
880	Rockford	United States of America	162	348.46	282.75	81.14	2
881	Rohtak	India	54	139.86	114.70	82.01	7
882	Roma (Rome)	Italy	558	840.02	834.58	99.35	15
883	Rongcheng	China	52	219.47	123.92	56.46	4
884	Rosario	Argentina	179	353.20	319.98	90.59	5
885	Rostov-na-Donu (Rostov-on-Don)	Russian Federation	222	369.51	369.15	99.90	1
886	Rotterdam s-Gravenhage (The Hague)	Netherlands	733	1250.58	1067.34	85.35	1
887	Rouen	France	94	197.06	197.01	99.97	2
888	WenzhouWenzhou Ruian Yueqing	China	612	1723.88	966.67	56.08	4
889	Rustenburg	South Africa	98	192.44	191.68	99.61	10
890	Sacramento	United States of America	1046	1487.28	1353.88	91.03	16
891	Safaqis	Tunisia	170	325.97	311.59	95.59	11
892	Safi	Morocco	20	97.90	73.65	75.23	13
893	Sagar	India	34	97.90	83.92	85.72	7
894	Saint-Étienne	France	64	194.10	168.69	86.91	15
895	Sakarya	Turkey	71	158.39	158.27	99.92	16
896	Salem	India	99	199.05	198.97	99.96	16
897	Salerno	Italy	3	83.20	33.02	39.69	6
898	Salta	Argentina	89	257.14	185.71	72.22	3
899	Salt Lake City	United States of America	651	949.78	889.32	93.63	16
900	Samara	Russian Federation	36	406.21	132.32	32.57	1
901	Samarinda	Indonesia	117	221.74	221.57	99.92	3
902	Samarkand	Uzbekistan	179	297.51	297.32	99.94	7
903	Samsun	Turkey	24	95.83	93.39	97.45	1
904	Sanandaj	Iran (Islamic Republic of)	66	166.72	135.46	81.25	9
905	San Antonio	United States of America	1095	1411.29	1410.14	99.92	2
906	San Diego Tijuana	United States of America	1271	1840.93	1763.61	95.80	16
907	Sangali	India	39	172.67	90.85	52.61	16
908	San Juan	Puerto Rico	3	619.00	31.93	5.16	8
909	San Juan	Argentina	148	264.06	263.64	99.84	16
910	Sankt Peterburg (Saint Petersburg)	Russian Federation	365	580.03	579.98	99.99	14
911	Sanliurfa	Turkey	51	157.88	144.66	91.63	11
912	San Miguel de Tucumán	Argentina	193	326.56	326.35	99.94	5

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
913	Sanming	China	4	125.20	24.44	19.52	14
914	San Salvador	El Salvador	156	300.40	281.07	93.57	3
915	San Salvador de Jujuy	Argentina	70	191.50	152.67	79.72	3
916	Santa Clarita	United States of America	143	258.58	258.42	99.94	13
917	Santa Cruz	Bolivia (Plurinational State of)	307	457.03	456.93	99.98	16
918	Santa Fe	Argentina	71	210.19	196.51	93.49	5
919	Santa Rosa	United States of America	196	362.42	336.20	92.76	16
920	Santiago	Chile	824	1091.02	1090.75	99.97	10
921	Santiago	Dominican Republic	123	228.32	228.27	99.98	15
922	Santiago de Cuba	Cuba	58	134.09	125.50	93.59	16
923	Santiago Del Estero	Argentina	108	233.69	228.20	97.65	16
924	Santipur	India	8	93.48	36.71	39.27	15
925	Santo Domingo	Dominican Republic	274	464.03	459.12	98.94	16
926	Santo Domingo	Ecuador	75	157.41	144.95	92.08	3
927	Sanya	China	79	236.26	211.87	89.68	15
928	São José do Rio Preto	Brazil	142	245.22	245.11	99.96	7
929	São José dos Campos	Brazil	86	395.00	179.84	45.53	5
930	São Paulo	Brazil	2061	2520.81	2520.53	99.99	16
931	Sapporo	Japan	326	515.78	515.75	99.99	16
932	Sarajevo	Bosnia and Herzegovina	66	187.95	143.01	76.09	15
933	Sarasota-Bradenton	United States of America	503	878.37	830.13	94.51	2
934	Saratov	Russian Federation	85	353.82	279.16	78.90	4
935	Sargodha	Pakistan	62	144.76	141.33	97.63	7
936	Sari	Iran (Islamic Republic of)	53	121.35	120.84	99.58	5
937	Satna	India	58	152.27	114.39	75.12	10
938	Scranton	United States of America	72	226.37	157.26	69.47	2
939	Seattle	United States of America	552	1391.41	949.41	68.23	2
940	Sekondi Takoradi	Ghana	4	199.93	33.82	16.92	12
941	Semarang	Indonesia	251	414.71	414.63	99.98	16
942	Sendai	Japan	396	612.96	612.73	99.96	16
943	Seremban	Malaysia	191	330.11	328.78	99.60	3
944	Sevastopol	Ukraine	50	166.24	157.16	94.54	14
945	Sevilla	Spain	252	442.38	410.16	92.72	11
946	Shahjahanpur	India	9	138.03	55.86	40.47	5
947	Shangqiu	China	154	369.61	285.77	77.32	4
948	Shangrao	China	89	210.27	174.67	83.07	6
949	Shaoguan	China	76	314.06	195.49	62.25	4

continued on next page

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
950	Shaoyang	China	83	206.66	179.93	87.07	1
951	Sheffield	United Kingdom	173	354.81	299.35	84.37	2
952	Shengzhou	China	81	176.29	175.92	99.79	11
953	Shenyang	China	672	917.46	917.21	99.97	4
954	Shihezi	China	96	200.39	181.15	90.40	11
955	Shijiazhuang	China	591	882.91	882.77	99.98	4
956	Shillong	India	48	123.36	112.44	91.15	6
957	Shimkent	Kazakhstan	241	399.60	382.55	95.73	11
958	Shimoga	India	34	124.70	87.51	70.17	5
959	Shiraz	Iran (Islamic Republic of)	234	391.84	391.61	99.94	13
960	Shiyan	China	83	219.01	170.49	77.84	6
961	Shizuishan	China	24	188.22	77.38	41.11	13
962	Shizuoka-Hamamatsu M.M.A.	Japan	140	531.66	277.15	52.13	3
963	Shreveport	United States of America	274	456.12	428.22	93.88	15
964	Shuangyashan	China	24	174.18	70.87	40.69	6
965	Shuozhou	China	66	151.30	151.10	99.87	13
966	Sialkot	Pakistan	93	207.23	188.89	91.15	7
967	Sihui	China	46	122.29	122.21	99.94	7
968	Siliguri	India	94	197.00	196.62	99.81	3
969	Simferopol	Ukraine	81	188.09	187.95	99.92	4
970	Sivas	Turkey	19	104.62	71.47	68.31	11
971	Skopje	TFYR Macedonia	103	221.97	199.60	89.92	7
972	Smolensk	Russian Federation	56	127.36	125.14	98.26	1
973	Sochi	Russian Federation	19	91.31	89.15	97.64	1
974	Sofia	Bulgaria	205	343.68	343.61	99.98	4
975	Sokoto	Nigeria	124	229.41	228.79	99.73	11
976	Solapur	India	79	185.09	158.81	85.80	10
977	Southampton/Portsmouth (South Hampshire)	United Kingdom	126	443.29	287.27	64.80	5
978	Spokane	United States of America	188	448.40	330.79	73.77	2
979	Springfield, Massachusetts, Connecticut	United States of America	158	297.80	294.83	99.00	2
980	Springfield, Missouri	United States of America	213	356.43	356.36	99.98	15
981	Srinagar	India	112	240.07	239.85	99.91	10
982	Stavropol	Russian Federation	148	282.87	282.71	99.94	4
983	St. Catharines-Niagara	Canada	84	181.19	181.17	99.99	11
984	St. Louis	United States of America	208	384.31	341.50	88.86	2
985	Stockholm	Sweden	48	278.75	212.24	76.14	1
986	Stoke-on-Trent (The Potteries)	United Kingdom	91	198.55	198.29	99.87	2

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
987	Strasbourg	France	132	242.72	242.32	99.84	1
988	Stuttgart	Germany	79	220.38	220.30	99.96	4
989	Sucre	Bolivia (Plurinational State of)	40	128.74	94.53	73.43	6
990	Suihua	China	30	85.14	84.93	99.75	3
991	Suining, Sichuan	China	51	151.68	117.65	77.56	14
992	Suizhou	China	51	127.13	127.09	99.97	1
993	Sukabumi	Indonesia	75	186.26	162.92	87.47	5
994	Sulaimaniya	Iraq	115	240.08	220.48	91.84	13
995	Sunshine Coast	Australia	11	289.02	45.31	15.68	2
996	Suqian	China	234	611.83	414.85	67.80	4
997	Surakarta	Indonesia	156	285.95	285.78	99.94	5
998	Surat	India	304	628.20	491.18	78.19	11
999	Surgut	Russian Federation	37	110.80	96.21	86.83	4
1000	Suzhou, Anhui	China	75	154.41	153.02	99.10	1
1001	Sydney	Australia	1019	1637.64	1390.13	84.89	2
1002	Sylhet	Bangladesh	56	129.09	127.29	98.60	15
1003	Syracuse	United States of America	109	266.62	244.05	91.53	2
1004	Szczecin	Poland	49	141.28	138.91	98.32	1
1005	Tabriz	Iran (Islamic Republic of)	130	257.44	252.09	97.92	9
1006	Tabuk	Saudi Arabia	135	259.33	258.79	99.79	9
1007	Taichung	China	792	1402.61	1058.51	75.47	4
1008	Taif	Saudi Arabia	138	275.29	258.54	93.91	9
1009	Taipei	China	364	575.12	575.08	99.99	6
1010	Taizhou, Jiangsu	China	177	506.77	335.56	66.22	4
1011	Taizhou, Zhejiang Wenling	China	832	1327.29	1147.33	86.44	4
1012	Ta'izz	Yemen	61	159.68	135.04	84.57	9
1013	Tallinn	Estonia	8	155.79	84.28	54.10	12
1014	Tamale	Ghana	93	182.46	176.96	96.98	16
1015	Tampa-St. Petersburg	United States of America	721	1448.35	1133.71	78.28	2
1016	Tampere	Finland	12	121.20	78.10	64.44	2
1017	Tampico	Mexico	51	338.45	133.73	39.51	8
1018	Tangshan, Hebei	China	603	1019.68	990.34	97.12	11
1019	Tarabulus (Tripoli)	Libya	482	868.17	693.50	79.88	2
1020	Taraz	Kazakhstan	110	204.27	204.02	99.88	16
1021	Tartus	Syrian Arab Republic	13	89.10	64.89	72.83	14
1022	Tashauz	Turkmenistan	55	131.30	125.78	95.79	13
1023	Tashkent	Uzbekistan	573	799.04	798.81	99.97	16

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1024	Tasikmalaya	Indonesia	77	165.21	165.11	99.94	15
1025	Tbilisi	Georgia	41	380.39	95.67	25.15	6
1026	Teeside (Middlesbrough)	United Kingdom	60	258.99	140.91	54.41	2
1027	Tegucigalpa	Honduras	143	249.73	249.60	99.95	3
1028	Tehuacán	Mexico	84	255.02	151.33	59.34	7
1029	Temecula-Murrieta	United States of America	152	326.24	286.59	87.85	13
1030	Temuco	Chile	38	145.22	94.72	65.23	3
1031	Tengzhou	China	102	205.76	205.50	99.88	4
1032	Tepic	Mexico	82	189.01	159.16	84.21	3
1033	Teresina	Brazil	238	375.01	374.71	99.92	16
1034	Tétouan	Morocco	31	177.67	79.36	44.67	11
1035	Thessaloniki	Greece	90	273.66	242.71	88.69	4
1036	Thoothukkudi (Tuticorin)	India	48	129.42	120.23	92.90	10
1037	Tianjin	China	1018	1351.43	1351.30	99.99	11
1038	Tianmen	China	43	107.57	99.84	92.81	1
1039	Tianshui	China	30	224.44	78.15	34.82	6
1040	Tieling	China	46	139.05	119.42	85.88	4
1041	Timisoara	Romania	68	153.19	153.13	99.96	5
1042	Tirana (Tirana)	Albania	89	178.24	173.22	97.18	15
1043	Tirunelveli	India	10	268.72	39.27	14.61	7
1044	Toamasina	Madagascar	4	140.19	24.89	17.75	12
1045	Tokyo	Japan	4290	5087.06	5086.40	99.99	16
1046	Toledo	United States of America	334	594.67	519.42	87.35	2
1047	Tomsk	Russian Federation	64	156.90	144.15	91.88	1
1048	Tongchuan	China	49	172.69	104.57	60.55	11
1049	Tonghua	China	30	167.97	81.01	48.23	6
1050	Tongliao	China	105	219.02	218.78	99.89	4
1051	Tongling	China	124	244.16	243.95	99.91	14
1052	Tongxiang	China	88	191.85	191.75	99.95	4
1053	Torino (Turin)	Italy	164	336.51	330.18	98.12	4
1054	Toulon	France	46	262.25	145.17	55.36	1
1055	Toulouse	France	208	382.59	382.12	99.88	2
1056	Tours	France	55	170.74	170.51	99.86	1
1057	Trujillo	Peru	118	228.40	226.88	99.34	10
1058	Tshikapa	Democratic Republic of the Congo	49	146.00	114.37	78.34	5
1059	Tucson	United States of America	706	1026.02	1025.66	99.97	13
1060	Tulsa	United States of America	504	723.23	721.86	99.81	16

*continued on next page*



ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1061	Tumkur	India	34	144.24	92.63	64.22	11
1062	Tunis	Tunisia	292	566.10	560.89	99.08	11
1063	Tuxtla Gutierrez	Mexico	136	261.12	253.96	97.26	3
1064	Tver	Russian Federation	71	156.76	156.71	99.97	1
1065	Tyumen	Russian Federation	118	239.44	227.99	95.22	1
1066	Uberaba	Brazil	108	206.90	206.83	99.97	7
1067	Uberlândia	Brazil	196	328.94	328.66	99.92	7
1068	Udaipur	India	117	232.36	206.80	89.00	16
1069	Udon Thani	Thailand	71	163.70	158.17	96.62	15
1070	Ufa	Russian Federation	52	260.35	130.86	50.26	1
1071	Ulaanbaatar	Mongolia	186	343.17	319.77	93.18	13
1072	Ulan-Ude	Russian Federation	71	196.54	154.98	78.85	11
1073	Ulsan	Republic of Korea	161	324.52	305.56	94.16	4
1074	Ulyanovsk	Russian Federation	24	244.57	81.79	33.44	1
1075	Ürümqi (Wulumqi)	China	364	629.95	576.22	91.47	13
1076	Ust-Kamenogorsk	Kazakhstan	48	178.79	130.95	73.24	14
1077	Utsunomiya	Japan	386	666.70	666.53	99.97	16
1078	Valencia	Venezuela (Bolivarian Republic of)	307	585.25	499.80	85.40	3
1079	Valencia	Spain	298	499.70	468.98	93.85	11
1080	Valenciennes	France	62	153.04	141.40	92.39	15
1081	Valladolid	Spain	81	173.39	172.30	99.37	7
1082	Valledupar	Colombia	65	149.36	143.50	96.08	3
1083	Valparaíso	Chile	2	223.20	21.74	9.74	8
1084	Vancouver	Canada	364	953.66	644.46	67.58	16
1085	Varanasi (Benares)	India	141	271.60	271.35	99.91	11
1086	Varna	Bulgaria	39	146.47	125.73	85.84	1
1087	Veracruz	Mexico	60	242.95	137.96	56.79	3
1088	Vereeniging	South Africa	155	397.39	299.80	75.44	10
1089	Verona	Italy	61	204.59	157.18	76.83	4
1090	Victorville-Hesperia-Apple Valley	United States of America	22	646.18	122.83	19.01	7
1091	Vientiane	Lao People's Democratic Republic	160	415.73	325.15	78.21	16
1092	Vijayawada	India	69	173.51	157.48	90.76	16
1093	Villahermosa	Mexico	86	232.24	155.81	67.09	3
1094	Villavicencio	Colombia	48	135.28	108.56	80.25	3
1095	Vilnius	Lithuania	77	168.12	168.03	99.94	1
1096	Vinnitsa	Ukraine	80	167.37	155.17	92.71	1
1097	Virginia Beach	United States of America	285	1075.40	486.54	45.24	2

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1098	Visakhapatnam	India	49	345.03	113.70	32.95	10
1099	Vitebsk	Belarus	60	157.25	139.42	88.66	1
1100	Vladimir	Russian Federation	53	141.35	136.87	96.83	1
1101	Volgograd Volzhsky	Russian Federation	307	644.80	592.74	91.93	11
1102	Vologda	Russian Federation	33	120.94	109.34	90.41	1
1103	Volta Redonda	Brazil	88	227.47	181.58	79.83	5
1104	Wad Medani	Sudan	125	278.98	226.47	81.18	10
1105	Wafangdian	China	51	168.25	132.53	78.77	11
1106	Wah	Pakistan	50	230.95	103.16	44.67	5
1107	Wahran (Oran)	Algeria	60	274.85	120.53	43.85	11
1108	Warri	Nigeria	150	293.05	285.50	97.42	16
1109	Warszawa (Warsaw)	Poland	262	421.56	421.19	99.91	1
1110	Weihai	China	31	330.46	120.48	36.46	14
1111	Weinan	China	51	166.54	119.04	71.48	11
1112	West Yorkshire	United Kingdom	94	340.06	204.96	60.27	2
1113	Wichita	United States of America	369	563.02	551.50	97.95	2
1114	Wien (Vienna)	Austria	216	421.75	421.51	99.94	15
1115	Windhoek	Namibia	139	273.59	260.10	95.07	10
1116	Winston-Salem	United States of America	212	417.33	321.77	77.10	15
1117	Witbank	South Africa	55	300.79	116.97	38.89	10
1118	Worcester	United States of America	24	85.27	85.08	99.78	2
1119	Wroclaw	Poland	77	170.57	170.55	99.99	15
1120	Wuhai	China	57	518.82	145.69	28.08	13
1121	Wuwei	China	45	111.72	111.36	99.68	11
1122	Wuzhou	China	37	154.02	87.45	56.78	14
1123	Xalapa	Mexico	89	188.47	188.27	99.89	15
1124	Xiangcheng	China	39	210.35	101.11	48.07	15
1125	Xiangyang	China	53	359.23	108.02	30.07	1
1126	Xianning	China	44	185.99	103.38	55.58	6
1127	Xi'an, Shaanxi Xianyang, Shaanxi	China	1090	1411.23	1405.82	99.62	11
1128	Xiantao	China	65	147.68	147.50	99.88	4
1129	Xinghua	China	77	162.00	159.97	98.75	4
1130	Xingning	China	46	121.15	119.40	98.55	1
1131	Xingtai	China	195	397.84	340.67	85.63	4
1132	Zhengzhou Xinyang	China	590	959.68	887.48	92.48	4
1133	Xingyi, Guizhou	China	34	139.77	85.56	61.21	6
1134	Xining	China	107	321.98	258.15	80.18	13

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1135	Xinmi	China	51	117.30	117.11	99.84	15
1136	Xintai	China	59	183.73	127.39	69.34	13
1137	Xinyang	China	97	184.39	184.01	99.80	4
1138	Xinyi	China	33	93.46	83.98	89.86	15
1139	Xinyu	China	120	233.53	229.11	98.11	4
1140	Xinzheng	China	46	156.00	107.79	69.10	15
1141	Xinzhou	China	43	115.58	115.55	99.98	11
1142	Xuancheng	China	40	124.47	98.07	78.79	14
1143	Xuzhou	China	224	398.33	398.16	99.96	4
1144	Yakutsk	Russian Federation	57	152.30	148.41	97.44	14
1145	Yancheng, Jiangsu	China	216	364.23	357.15	98.06	4
1146	Yangon	Myanmar	402	629.23	589.37	93.66	16
1147	Yangquan	China	87	216.14	200.25	92.65	13
1148	Yangzhou	China	257	482.08	451.37	93.63	1
1149	Yanji	China	62	142.07	141.27	99.44	4
1150	Yantai	China	122	698.54	268.59	38.45	11
1151	Yaoundé	Cameroon	246	388.91	388.65	99.93	16
1152	Yaroslavl	Russian Federation	71	179.75	176.49	98.18	1
1153	Yazd	Iran (Islamic Republic of)	167	296.16	296.16	100.00	9
1154	Yekaterinburg	Russian Federation	164	327.34	327.01	99.90	4
1155	Yerevan	Armenia	201	335.98	335.93	99.99	11
1156	Yibin	China	42	182.50	117.69	64.49	14
1157	Yichang	China	104	259.10	242.69	93.67	1
1158	Yichun, Jiangxi	China	65	187.02	143.80	76.89	6
1159	Yinchuan	China	181	451.46	367.59	81.42	11
1160	Yingde	China	4	118.77	24.48	20.61	14
1161	Yining	China	96	199.61	199.58	99.99	11
1162	Yogyakarta	Indonesia	195	323.91	323.89	100.00	2
1163	Yongcheng	China	44	158.49	134.86	85.09	15
1164	Yongin	Republic of Korea	20	89.36	89.16	99.78	6
1165	Yongzhou	China	42	161.71	95.65	59.15	1
1166	Youngstown	United States of America	155	456.24	310.70	68.10	2
1167	Yulin, Guangxi	China	126	238.86	238.65	99.91	15
1168	Yulin, Shaanxi	China	86	180.09	177.33	98.46	13
1169	Yushu	China	18	75.70	61.92	81.80	5
1170	Yuxi	China	85	207.72	192.91	92.87	11
1171	Zacatecas	Mexico	31	206.58	85.63	41.45	9

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1172	Zagreb	Croatia	122	245.66	237.78	96.79	4
1173	Zahedan	Iran (Islamic Republic of)	106	194.30	194.11	99.90	9
1174	Zanjan	Iran (Islamic Republic of)	57	162.52	128.58	79.12	9
1175	Zaoyang	China	37	124.68	88.44	70.93	4
1176	Zaozhuang	China	103	224.81	210.11	93.46	4
1177	Zaporizhzhya	Ukraine	146	292.14	270.09	92.45	11
1178	Zaragoza	Spain	77	209.79	190.45	90.78	11
1179	Zaria	Nigeria	113	259.21	211.99	81.78	10
1180	Zengcheng	China	62	135.43	135.29	99.89	1
1181	Zhangjiakou	China	79	416.73	196.18	47.08	11
1182	Zhangye	China	38	96.78	96.53	99.75	11
1183	Zhaodong	China	15	87.44	55.69	63.69	4
1184	Zhenjiang, Jiangsu	China	126	274.82	263.68	95.95	14
1185	Zhongxiang	China	47	117.50	102.25	87.02	8
1186	Zhoukou	China	87	201.81	201.66	99.92	4
1187	Zhoushan	China	34	288.89	99.49	34.44	14
1188	Zhuji	China	145	305.36	272.12	89.11	11
1189	Zhumadian	China	96	292.75	204.67	69.91	4
1190	Zibo	China	451	762.77	732.01	95.97	11
1191	Zigong	China	96	223.31	183.51	82.18	14
1192	Zinder	Niger	48	128.72	98.37	76.42	13
1193	Ziyang	China	42	125.06	103.12	82.46	6
1194	Zunyi	China	79	187.98	162.60	86.50	6
1195	Minneapolis-St. Paul	United States of America	332	645.57	645.13	99.93	2
1196	Xiaogan	China	14	57.89	57.73	99.73	1
1197	Merida	Venezuela (Bolivarian Republic of)	8	145.24	39.78	27.39	3
1198	San José	Costa Rica	378	605.77	603.78	99.67	16
1199	Cordoba	Spain	50	145.61	125.18	85.97	7
1200	Jining, Shandong	China	109	214.50	204.59	95.38	4
1201	Douai-Lens	France	39	115.80	115.85	100.04	5
1202	Zhaoqing	China	16	237.40	99.18	41.78	3
1203	Qingyuan	China	57	213.84	147.93	69.18	1
1204	Guangzhou, Guangdong Shenzhen Dongguan Foshan Zhongshan Jiangmen Heshan	China	8214	10121.98	9791.26	96.73	4
1205	Donggang	China	41	133.08	128.46	96.53	11
1206	Djibouti	Djibouti	3	135.10	30.73	22.75	8
1207	Dhule	India	71	152.51	149.35	97.93	10

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1208	Dewas	India	4	142.95	24.48	17.12	10
1209	Ujjain	India	31	125.93	81.76	64.92	3
1210	Cuernavaca	Mexico	257	481.75	391.96	81.36	16
1211	San Cristóbal	Venezuela (Bolivarian Republic of)	66	191.92	142.90	74.46	5
1212	Cotabato	Philippines	4	108.13	27.04	25.01	3
1213	Córdoba	Mexico	52	139.60	112.07	80.28	5
1214	Conakry	Guinea	48	340.48	156.83	46.06	8
1215	Haining	China	75	172.46	155.98	90.44	4
1216	Chongjin	Dem. People's Republic of Korea	24	118.69	69.71	58.73	10
1217	Choloma	Honduras	6	126.55	53.77	42.49	3
1218	Chişinău	Republic of Moldova	127	230.54	225.57	97.85	4
1219	Dujiangyan	China	67	165.60	154.74	93.44	14
1220	Chiclayo	Peru	91	178.50	176.32	98.78	10
1221	Cheongju	Republic of Korea	88	210.73	187.34	88.90	11
1222	Deyang	China	120	224.91	224.83	99.96	4
1223	Puning	China	132	712.75	252.90	35.48	4
1224	Daejeon	Republic of Korea	165	323.72	266.89	82.45	6
1225	Chengde	China	19	74.18	68.88	92.86	6
1226	Shantou Chaozhou Jieyang	China	1075	1930.30	1596.89	82.73	4
1227	Maracaibo	Venezuela (Bolivarian Republic of)	298	589.31	506.96	86.03	16
1228	Suzhou, Jiangsu Wuxi, Jiangsu Changzhou, Jiangsu Jiangyin Zhangjiagang Jingjiang	China	4835	6141.05	6104.91	99.41	4
1229	Yixing	China	51	287.14	108.79	37.89	1
1230	Cabimas Lagunillas	Venezuela (Bolivarian Republic of)	88	441.82	196.78	44.54	16
1231	Denpasar	Indonesia	140	338.84	323.01	95.33	3
1232	Hardwar	India	3	385.68	22.79	5.91	10
1233	Davao City	Philippines	63	262.25	159.83	60.95	3
1234	Yingkou	China	130	253.32	253.26	99.97	13
1235	Dashiqiao	China	67	198.61	192.48	96.91	11
1236	Haicheng	China	52	144.21	139.84	96.97	4
1237	Pingdingshan, Henan	China	110	238.07	231.98	97.44	4
1238	Xinyi	China	63	143.04	142.90	99.90	15
1239	Puerto Vallarta	Mexico	18	118.37	65.28	55.15	16
1240	Qitaihe	China	6	230.32	46.50	20.19	11
1241	Taixing	China	88	203.07	195.54	96.29	1
1242	Changwon	Republic of Korea	84	265.44	184.86	69.64	3
1243	Liuyang	China	47	135.58	103.39	76.26	6

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1244	Xiangtan, Hunan	China	245	419.20	404.46	96.48	4
1245	Xuchang	China	103	197.27	197.17	99.95	4
1246	Patiala	India	102	196.75	196.03	99.63	10
1247	Celaya	Mexico	97	184.32	184.12	99.89	7
1248	Querétaro	Mexico	238	410.88	410.66	99.95	10
1249	Cardiff	United Kingdom	91	219.69	217.46	98.99	2
1250	Cape Town	South Africa	20	855.72	60.83	7.11	16
1251	Cangzhou	China	93	190.20	190.09	99.94	4
1252	Cagayan de Oro City	Philippines	57	145.93	141.08	96.68	3
1253	Butuan	Philippines	4	102.77	26.52	25.81	12
1254	Gimhae	Republic of Korea	232	652.58	455.39	69.78	1
1255	Gumi	Republic of Korea	38	202.00	103.17	51.08	6
1256	Bursa	Turkey	59	299.64	142.22	47.46	3
1257	Bridgeport-Stamford New Haven	United States of America	279	667.37	590.86	88.54	2
1258	Modena	Italy	56	130.21	127.13	97.63	5
1259	Blumenau	Brazil	49	267.73	105.67	39.47	2
1260	Bologna	Italy	86	198.26	185.76	93.69	15
1261	Blida	Algeria	70	158.18	157.82	99.77	11
1262	Bihar Sharif	India	6	115.33	36.57	31.71	7
1263	Vungtau	Viet Nam	11	120.64	53.53	44.37	8
1264	Béthune	France	22	70.76	69.13	97.70	5
1265	Hubli-Dharwad	India	46	375.64	99.73	26.55	5
1266	Beira	Mozambique	14	148.90	51.24	34.41	8
1267	Beihai	China	73	194.60	189.83	97.55	1
1268	Bazhong	China	4	125.98	22.95	18.22	14
1269	Bayannaer	China	63	143.36	143.30	99.96	13
1270	Batangas City	Philippines	24	164.52	80.38	48.86	15
1271	Zamboanga City	Philippines	25	163.85	88.64	54.10	16
1272	Santa Marta	Colombia	48	137.15	135.17	98.56	7
1273	Barddhaman	India	60	134.56	132.23	98.27	7
1274	Dhanbad	India	50	252.28	104.32	41.35	15
1275	Bandar Lampung	Indonesia	172	321.78	307.49	95.56	16
1276	Bahia Blanca	Argentina	77	175.34	173.44	98.92	16
1277	Mbouda	Cameroon	14	106.44	51.56	48.44	15
1278	Bacolod	Philippines	80	192.91	160.72	83.31	16
1279	Marseille-Aix-en-Provence	France	27	82.08	81.96	99.85	1
1280	Asansol	India	40	280.33	94.95	33.87	7

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1281	Arak	Iran (Islamic Republic of)	76	167.96	157.88	94.00	9
1282	Gent	Belgium	86	186.81	186.59	99.88	2
1283	Bruxelles-Brussel	Belgium	137	260.37	260.38	100.00	1
1284	Qingzhou	China	117	227.33	227.17	99.93	16
1285	Shouguang	China	680	1169.52	1089.19	93.13	4
1286	Dongying	China	66	600.04	165.34	27.55	14
1287	Tongcheng	China	42	131.64	102.61	77.95	4
1288	Annaba	Algeria	11	170.45	62.95	36.93	14
1289	Angeles City	Philippines	84	166.42	166.39	99.98	5
1290	Ambon	Indonesia	25	113.80	86.34	75.87	3
1291	Taian, Shandong	China	198	383.70	352.42	91.85	11
1292	Qufu	China	46	115.02	111.94	97.32	4
1293	Yongkang	China	110	395.16	241.20	61.04	4
1294	Jingzhou, Hubei	China	140	321.62	292.20	90.85	4
1295	Jhansi	India	96	190.12	186.08	97.88	7
1296	Jerusalem	Israel	194	374.84	339.93	90.69	13
1297	Jeonju	Republic of Korea	87	190.78	177.55	93.07	6
1298	Jeju	Republic of Korea	46	149.03	120.68	80.98	14
1299	Jamnagar	India	79	174.45	168.71	96.71	7
1300	Iligan	Philippines	20	95.33	67.28	70.58	3
1301	Irapuato	Mexico	82	166.16	166.04	99.93	7
1302	Kolhapur	India	79	191.01	174.81	91.52	11
1303	Oyo	Nigeria	93	186.27	185.91	99.81	15
1304	Hillah	Iraq	97	197.30	196.65	99.67	13
1305	Jinzhou	China	102	201.06	201.03	99.98	4
1306	Huludao	China	77	237.95	174.72	73.43	14
1307	Hanchuan	China	22	103.19	66.57	64.51	14
1308	Guarenas-Guatire	Venezuela (Bolivarian Republic of)	40	193.68	96.22	49.68	3
1309	Goma	Democratic Republic of the Congo	65	163.92	154.46	94.23	5
1310	Kahramanmaras	Turkey	42	117.79	95.00	80.65	4
1311	Zhucheng	China	138	272.59	272.50	99.97	4
1312	Gaomi	China	106	201.32	201.32	100.00	5
1313	Fengcheng	China	51	178.80	125.21	70.03	7
1314	Fangchenggang	China	24	93.49	67.55	72.26	6
1315	Erode	India	78	166.66	166.44	99.87	2
1316	Edinburgh	United Kingdom	86	188.59	172.02	91.21	2
1317	Durg-Bhilainagar	India	63	292.12	137.06	46.92	7

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1318	Durban	South Africa	225	933.72	343.40	36.78	2
1319	Lima	Peru	740	1101.19	1080.86	98.15	9
1320	Lianyungang	China	123	253.84	241.29	95.06	11
1321	Libreville	Gabon	146	300.46	291.92	97.16	16
1322	Leshan	China	53	242.37	110.04	45.40	1
1323	Latina	Italy	29	85.04	84.93	99.87	15
1324	León de Los Aldama	Mexico	249	404.70	404.59	99.97	7
1325	Langfang	China	83	198.53	198.52	100.00	4
1326	Lancaster	United States of America	45	142.44	122.87	86.26	2
1327	Longkou	China	124	443.24	241.06	54.39	11
1328	Zhaoyuan	China	60	136.38	130.20	95.47	11
1329	Laizhou	China	80	300.20	197.26	65.71	11
1330	Kuching	Malaysia	172	305.14	301.23	98.72	16
1331	Korba	India	2	337.38	24.99	7.41	10
1332	Thrissur	India	96	196.73	196.73	100.00	15
1333	Kitwe	Zambia	115	365.56	209.70	57.36	11
1334	Kitakyushu-Fukuoka M.M.A.	Japan	460	802.50	778.12	96.96	16
1335	Kingston upon Hull	United Kingdom	73	183.90	169.17	91.99	2
1336	Khabarovsk	Russian Federation	123	270.33	253.00	93.59	11
1337	Mannheim	Germany	83	218.39	218.41	100.01	1
1338	Karimnagar	India	5	124.63	37.44	30.04	5
1339	Sorocaba	Brazil	181	311.69	311.31	99.88	7
1340	Joinville	Brazil	225	375.71	375.47	99.94	3
1341	Thanjavur	India	4	179.13	25.95	14.49	5
1342	Tirupati	India	2	125.63	26.00	20.70	3
1343	Tolyatti	Russian Federation	15	450.31	48.19	10.70	15
1344	Van	Turkey	75	159.72	159.65	99.96	13
1345	Victoria	Canada	75	269.63	174.70	64.79	2
1346	Vladikavkaz	Russian Federation	78	173.15	172.07	99.38	11
1347	Newport	United Kingdom	40	143.35	127.83	89.17	2
1348	Nha Trang	Viet Nam	51	243.12	118.94	48.92	7
1349	Nnewi	Nigeria	121	351.48	264.68	75.31	15
1350	Owerri	Nigeria	123	227.01	226.93	99.96	15
1351	Umuahia	Nigeria	49	262.02	102.82	39.24	15
1352	Oita	Japan	151	306.63	259.68	84.69	16
1353	Perth	Australia	889	1211.08	1211.11	100.00	16
1354	Ribeirão Preto	Brazil	215	349.16	349.08	99.98	4

*continued on next page*



ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1355	Saharanpur	India	51	128.79	111.83	86.83	5
1356	Yamunanagar	India	23	198.62	79.32	39.94	3
1357	Salvador	Brazil	94	538.24	186.41	34.63	16
1358	Sandakan	Malaysia	4	274.23	29.75	10.85	8
1359	Wuhu, Anhui	China	158	322.60	293.53	90.99	14
1360	Chizhou	China	38	117.42	95.78	81.57	14
1361	Jiaozuo	China	149	298.32	290.78	97.47	4
1362	Muzaffarnagar	India	5	109.07	35.70	32.73	5
1363	Makassar (Ujung Pandang)	Indonesia	129	356.94	238.83	66.91	3
1364	Shanwei	China	17	102.69	65.49	63.77	8
1365	Siping	China	69	159.99	159.96	99.98	11
1366	Southend-On-Sea	United Kingdom	9	266.23	50.08	18.81	2
1367	Swansea	United Kingdom	16	118.43	54.62	46.12	2
1368	Tanger	Morocco	59	206.24	132.20	64.10	13
1369	Treviso	Italy	52	328.61	162.13	49.34	15
1370	Utrecht	Netherlands	90	247.78	193.61	78.14	1
1371	Bafoussam	Cameroon	86	176.74	152.02	86.01	15
1372	Barranquilla	Colombia	178	316.04	311.34	98.51	7
1373	Akure	Nigeria	158	287.40	287.23	99.94	15
1374	Port Harcourt	Nigeria	270	470.88	421.78	89.57	16
1375	Lagos Ikorodu	Nigeria	336	2089.16	599.88	28.71	7
1376	Barquisimeto	Venezuela (Bolivarian Republic of)	144	314.06	290.53	92.51	7
1377	Abakaliki	Nigeria	52	154.13	119.59	77.59	10
1378	Ilorin	Nigeria	200	330.35	326.55	98.85	11
1379	Iloilo City	Philippines	25	187.99	75.40	40.11	3
1380	Kolkata (Calcutta)	India	882	1200.95	1200.60	99.97	1
1381	Singapore Johor Bahru	Singapore	1271	1904.64	1680.62	88.24	4
1382	Belgaum	India	83	185.31	163.82	88.40	7
1383	Thành phố Hồ Chí Minh (Ho Chi Minh City) Thành phố Hồ Chí Minh (Ho Chi Minh City) Biên Hòa	Viet Nam	1181	1486.93	1487.43	100.03	16
1384	Patna	India	140	290.18	288.04	99.26	16
1385	Manila	Philippines	901	1537.42	1263.08	82.16	16
1386	Ahmadabad	India	360	524.45	524.22	99.96	4
1387	Amravati	India	66	146.21	142.78	97.65	10
1388	Ranchi	India	108	214.33	214.26	99.97	15
1389	Avignon	France	43	199.59	126.04	63.15	4

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1390	Antwerpen	Belgium	143	313.69	313.73	100.01	1
1391	Alicante	Spain	54	214.33	147.97	69.04	13
1392	Anqing	China	108	207.06	201.68	97.40	4
1393	Tanta	Egypt	54	122.52	121.37	99.06	4
1394	Agra	India	196	318.77	317.75	99.68	10
1395	Toluca de Lerdo	Mexico	243	397.33	396.85	99.88	16
1396	Orizaba	Mexico	63	147.39	145.34	98.61	3
1397	Hangzhou Shaoxing Cixi Yuyao Shangyu	China	3418	4798.51	4414.52	92.00	4
1398	San Pedro Sula	Honduras	92	193.75	191.90	99.04	3
1399	Changge	China	48	135.52	134.03	98.90	5
1400	Changsha	China	641	908.98	905.16	99.58	14
1401	Zhuzhou	China	160	291.78	291.73	99.98	14
1402	Chengdu	China	1062	1368.22	1368.17	100.00	4
1403	Reggio Emilia	Italy	38	100.53	100.25	99.72	15
1404	Chandigarh	India	253	439.93	407.86	92.71	4
1405	Cali	Colombia	193	319.92	319.73	99.94	3
1406	Daegu	Republic of Korea	173	300.04	300.10	100.02	6
1407	Cúcuta	Colombia	166	322.72	279.69	86.67	3
1408	Pietermaritzburg	South Africa	165	354.82	294.01	82.86	16
1409	Enugu	Nigeria	145	259.90	259.68	99.92	7
1410	Ibadan	Nigeria	512	735.13	734.71	99.94	15
1411	Najaf	Iraq	119	250.28	216.93	86.67	13
1412	Gaziantep	Turkey	125	234.58	234.55	99.99	13
1413	Lille	France	222	397.57	381.47	95.95	15
1414	Ichalakaranji	India	47	142.59	99.54	69.81	7
1415	Wuhan	China	243	502.74	502.53	99.96	14
1416	Qingdao Jimo Jiaozhou	China	1239	2319.82	1838.97	79.27	11
1417	Dehradun	India	120	223.05	223.13	100.04	16
1418	Roorkee	India	32	109.03	68.38	62.72	7
1419	Malegaon	India	42	123.82	91.97	74.28	10
1420	Indore	India	180	307.40	300.56	97.78	11
1421	Raipur	India	136	323.92	242.90	74.99	10
1422	Tiruppur	India	94	254.72	184.16	72.30	15
1423	Nanchang	China	549	785.70	785.59	99.99	4
1424	Gaoan	China	25	123.53	67.11	54.33	14
1425	Vellore	India	102	212.89	200.64	94.25	5
1426	Onitsha	Nigeria	141	644.36	310.38	48.17	16

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1427	Panipat	India	77	157.91	157.71	99.87	7
1428	Stockton	United States of America	302	517.99	481.67	92.99	7
1429	Zhanjiang	China	149	369.17	317.36	85.97	1
1430	Surabaya	Indonesia	635	1027.68	889.41	86.55	5
1431	Ma'anshan	China	199	441.95	339.88	76.91	11
1432	Kochi (Cochin)	India	38	569.80	104.47	18.33	8
1433	Kozhikode (Calicut)	India	100	239.79	224.27	93.53	2
1434	Pizhou	China	25	158.57	66.61	42.01	1
1435	Jiyuan	China	97	215.70	215.75	100.03	4
1436	Yiyang, Hunan	China	92	198.76	198.76	100.00	1
1437	Xinxiang	China	171	394.23	327.75	83.14	4
1438	Qianjiang	China	34	96.11	93.91	97.71	15
1439	Jinhua	China	165	326.75	309.95	94.86	4
1440	Tiruchirappalli	India	86	200.04	149.90	74.94	16
1441	Laiwu	China	240	520.07	401.70	77.24	10
1442	Jundiaí	Brazil	137	271.21	270.86	99.87	16
1443	Warangal	India	8	226.78	37.96	16.74	8
1444	Karlsruhe	Germany	68	196.82	159.41	80.99	15
1445	Lahore	Pakistan	431	638.90	638.70	99.97	13
1446	Port Moresby	Papua New Guinea	61	237.79	141.73	59.60	2
1447	Bournemouth/Poole	United Kingdom	68	301.76	161.70	53.59	5
1448	Yuncheng	China	82	179.73	173.76	96.68	4
1449	Cebu City Lapu-Lapu City Mandaue City	Philippines	178	512.25	443.64	86.61	16
1450	Buenos Aires	Argentina	1771	2234.79	2171.57	97.17	16
1451	Bhavnagar	India	75	163.91	141.38	86.25	7
1452	Birmingham	United States of America	80	228.25	228.40	100.07	2
1453	Cartagena	Colombia	61	251.87	190.39	75.59	8
1454	Taranto	Italy	11	200.66	83.72	41.72	14
1455	Venezia	Italy	3	182.12	27.40	15.04	8
1456	Naberezhnye Tchelny	Russian Federation	66	282.19	139.64	49.48	14
1457	Lokoja	Nigeria	35	135.46	77.52	57.23	5
1458	Porto Velho	Brazil	86	225.47	165.14	73.24	16
1459	Duesseldorf	Germany	119	284.36	284.22	99.95	1
1460	Wuppertal	Germany	33	102.54	102.45	99.91	1
1461	Mérida	Mexico	280	422.92	422.77	99.97	3
1462	Zhuanghe	China	42	142.77	114.11	79.93	11
1463	Sheikhupura	Pakistan	36	119.20	83.43	69.99	10

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1464	Kasur	Pakistan	26	103.65	65.65	63.34	10
1465	Padova	Italy	88	236.31	236.34	100.01	2
1466	Vicenza	Italy	33	194.39	107.75	55.43	15
1467	Philadelphia Trenton	United States of America	796	1224.88	1224.76	99.99	2
1468	Duisburg Essen Bochum	Germany	778	1107.19	1088.25	98.29	2
1469	Aba	Nigeria	155	263.99	263.90	99.96	15
1470	Abeokuta	Nigeria	160	278.59	278.22	99.87	7
1471	Acarigua-Aruare	Venezuela (Bolivarian Republic of)	94	193.79	181.19	93.50	5
1472	Adan (Aden)	Yemen	21	225.31	86.35	38.32	9
1473	Uyo	Nigeria	4	321.90	27.04	8.40	3
1474	Mathura	India	69	173.31	151.24	87.26	10
1475	Al Kuwait (Kuwait City)	Kuwait	321	620.42	521.12	84.00	9
1476	Matsuyama	Japan	131	295.06	294.95	99.96	16
1477	Modesto	United States of America	244	487.23	443.96	91.12	5
1478	Maputo Matola	Mozambique	201	729.09	396.44	54.37	16
1479	Wuchuan	China	53	123.20	122.31	99.28	1
1480	Maoming	China	91	201.28	193.20	95.98	1
1481	Manado	Indonesia	66	180.15	168.33	93.44	3
1482	Malang	Indonesia	128	227.35	227.41	100.03	15
1483	Balikpapan	Indonesia	16	219.44	66.70	30.40	9
1484	Vadodara	India	180	304.42	304.31	99.96	15
1485	Zanzibar	United Republic of Tanzania	1	216.09	30.43	14.08	9
1486	Yueyang	China	107	230.54	230.54	100.00	1
1487	Yuanjiang	China	28	125.19	103.02	82.29	8
1488	Yangjiang	China	107	222.51	214.90	96.58	4
1489	Wonsan	Dem. People's Republic of Korea	13	81.94	65.36	79.77	13
1490	Wellington	New Zealand	18	117.05	75.58	64.57	2
1491	Anand	India	41	312.29	91.37	29.26	15
1492	Akola	India	8	167.66	37.20	22.19	10
1493	Oshogbo	Nigeria	125	407.34	215.92	53.01	15
1494	Ogbomosho	Nigeria	70	177.63	142.04	79.96	15
1495	Al-Mansurah	Egypt	47	117.42	116.84	99.50	14
1496	Al-Mahallah al-Kubra	Egypt	44	109.96	106.11	96.50	15
1497	Murcia	Spain	238	411.24	385.40	93.72	13
1498	Chon Buri	Thailand	146	395.55	301.81	76.30	16
1499	Krung Thep (Bangkok) Krung Thep (Bangkok) Samut Prakan Nonthaburi	Thailand	3536	4106.39	4100.75	99.86	16

*continued on next page*

ID	Aggregated Name	Country	No. Patches	Area MUA [km <sup>2</sup> ]	Area LCZ [km <sup>2</sup> ]	Coverage LCZ [%]	$C_{CM}$
1500	Bogotá	Colombia	485	696.83	695.74	99.84	3
1501	Belo Horizonte Vale do Aco	Brazil	604	834.35	834.17	99.98	16
1502	Montreal	Canada	912	1329.82	1309.31	98.46	16
1503	Sanaa	Yemen	247	447.05	406.72	90.98	9
1504	Athína (Athens)	Greece	486	773.30	771.20	99.73	4
1505	El Djazaïr (Algiers)	Algeria	279	749.50	461.92	61.63	11
1506	Asunción	Paraguay	289	567.43	472.10	83.20	2
1507	Ciudad Juarez El Paso	Mexico	964	1324.57	1324.45	99.99	13
1508	Ciudad de Panamá (Panama City)	Panama	145	363.63	279.76	76.93	16
1509	Grande Vitória	Brazil	105	602.19	232.17	38.55	16
1510	Grande São Luís	Brazil	199	516.28	358.01	69.34	16
1511	München (Munich)	Germany	253	442.31	442.33	100.00	15
1512	Maceió	Brazil	29	244.43	100.53	41.13	8
1513	Zürich (Zurich)	Switzerland	133	785.00	317.98	40.51	15
1514	København (Copenhagen)	Denmark	206	418.66	354.48	84.67	1
1515	Florianópolis	Brazil	12	405.75	61.61	15.18	6
1516	Fès	Morocco	102	211.40	202.33	95.71	11
1517	San Luis Potosí	Mexico	222	384.77	382.73	99.47	16
1518	Köln (Cologne)	Germany	55	139.80	139.82	100.02	1
1519	Lomé	Togo	155	440.69	297.02	67.40	16