



Patrícia Alexandra Fernandes Lopes

# LOCAL CLIMATE ZONES CLASSIFICATION WITH SATELLITE IMAGERY AND VOLUNTEERED GEOGRAPHIC INFORMATION

Master's degree in Systems and Technology of Geographic Information, area of specialization in Geographic Information Science and Technology, supervised by Professor Cidália Maria Parreira da Costa Fonte, co-supervised by Doctor Linda See and presented to the Department of Mathematics of the Faculty of Sciences and Technology of the University of Coimbra

2017



UNIVERSIDADE DE COIMBRA

Patrícia Alexandra Fernandes Lopes

**LOCAL CLIMATE ZONE  
CLASSIFICATIONS WITH SATELLITE  
IMAGERY AND VOLUNTEERED  
GEOGRAPHIC INFORMATION**

Master's degree in Systems and Technology of Geographic Information, area of specialization in Geographic Information Science and Technology, supervised by Professor Cidália Maria Parreira da Costa Fonte, co-supervised by Doctor Linda See and presented to Department of Mathematics of the Faculty of Sciences and Technology of the University of Coimbra

2017



UNIVERSIDADE DE COIMBRA

“Happiness can be found, even in the darkest of times, if one only remembers to turn on the light” —  
J.K. Rowling

## Agradecimentos

---

A conclusão de uma dissertação de mestrado é, sem dúvida, um importante marco num percurso de vida. É importante agradecer a todos que contribuíram para que isso fosse possível, de uma forma direta ou indireta.

Em primeiro lugar, expresso os meus mais sinceros agradecimentos à Professora Cidália Fonte, a minha orientadora, pela constante disponibilidade e apoio na realização desta etapa, pelas revisões, pela preocupação e pela confiança depositada durante todo este sinuoso percurso. Revelando-se decisiva para que este trabalho chegasse a bom porto.

Em segundo lugar, um especial agradecimento à Doutora Linda See, minha co-orientadora, por toda a disponibilidade quer na obtenção de recursos, nas soluções apresentadas e na revisão científica. Expresso igualmente, um especial agradecimento ao Doutor Benjamin Betchel, pela disponibilidade, ideias e sugestões apresentadas

Não seria possível esquecer de expressar um reconhecimento a todos os amigos, companheiros e professores de licenciatura em Geografia, que foram essenciais para que chegasse a esta etapa. Agradeço, igualmente, às funcionárias da Biblioteca do São Jerónimos, da Ciências da Terra e da Geral, pela preciosa ajuda com todos os materiais.

Uma palavra de gratidão a todos os meus colegas de mestrado, pela aprendizagem colaborativa e pela interajuda que foi essencial. A todos os meus professores do mestrado também expresso um sentido “obrigado” pelo fornecimento de ferramentas que tornaram este projeto bem-sucedido.

É indispensável agradecer à minha amiga Ana Sousa, pela sua paciência, apoio, incentivo e por mostrar diariamente o que significa uma amizade verdadeira. Um muito obrigado igualmente, à minha amiga Luciana Costa, por toda a compreensão e confiança.

Á minha mãe, por me incentivar a nunca desistir perante os desafios, a ser independente e a procurar e implantar sempre a solução para os problemas.

Á minha tia, pela confiança, acompanhamento, incentivo sempre a fazer mais e melhor e por toda a paciência.

Á minha irmã, por todo o apoio, incentivo e sinceridade sempre que esta se revelava necessária.

Obrigado a todos!



## Acknowledgements

---

The conclusion of a master's thesis is undoubtedly an important milestone in a life course. It is therefore important to thank all those who contributed to making this possible, directly or indirectly.

First of all, I would like to express my sincere thanks to Professor Cidália Fonte, my supervising tutor, for the constant availability and support of this stage, for the reviews, for the concern and for the trust placed throughout this winding course. Turning out to be decisive for this work to come to fruition.

Secondly, a special thanks to Dr. Linda See, my co-tutor supervisor, for all the availability in obtaining resources, solutions presented and scientific review. I also express a special thanks to Dr. Benjamin Betchel for all the availability, ideas and suggestions presented throughout this dissertation.

It would not be possible to forget to express an acknowledgment to all the friends, classmates and professors of degree in Geography, who were essential for it to arrive at this stage. Thank you also to the staff of the Library of São Jerónimos, Ciências da Terra and Geral, for the precious help with all materials.

A word of gratitude to all my master's colleagues, for the collaborative learning and for the inter-help that was essential. All my master's teachers also expressed a "thank you" for the provision of tools that made this project successful.

It is indispensable to thank my friend of one life, Ana Sousa, for her patience, support, encouragement and for showing daily what true friendship means. A big thank you also to my friend Luciana Costa for all the support, understanding and confidence shown.

To my mother, for encouraging me to never give in to the challenges, to be independent and to seek and always implant the solution to the problems.

To my aunt, for the trust, accompaniment, encouragement always to do more and better and for all the patience.

To my sister, for all the support, encouragement and sincerity whenever this was necessary.

Thanks to all!

Nº do aluno: [2011184019]

Nome: [Patrícia Alexandra Fernandes Lopes]

Título da dissertação:

[Local Climate Zones classifications with satellite imagery and Volunteered Geographic Information]

Palavras-Chave:

- Local Climate Zones
- Remote sensing
- World Urban Database and Access Portal Tools
- Volunteered Geographic Information
- OpenStreetMap

Student's number: [2011184019]

Name: [Patrícia Alexandra Fernandes Lopes]

Title of dissertation:

[Local Climate Zones classifications with satellite imagery and Volunteered Geographic Information]

Key-words:

- Local Climate Zones
- Remote sensing
- World Urban Database and Access Portal Tools
- Volunteered Geographic Information
- OpenStreetMap

## List of common acronyms

---

LCZ - Local Climate Zones

WUDAPT - World Urban Database and Access Portal Tools

OSM - OpenStreetMap

LULC - Land Use and Land Cover

UHI - Urban Heat Island

GIS - Geographic Information Systems

VGI - Volunteered Geographic Information

SAGA - System for Automated Geoscientific Analyses

ESRI - Environmental Systems Research Institute

OLI - Operational Land Imager

TIRS - Thermal Infrared Sensor

POIs - Points of interest

ROI - Region of Interest

KML - Keyhole Markup Language

OA - Overall Accuracy

UA - User Accuracy

PA - Producer Accuracy

## Resumo

---

O aumento da população mundial e das suas atividades e a crescente concentração de população nos centros urbanos resultam em alterações não só nas características físicas da superfície terrestre como igualmente, nas características climáticas. Como resultado desta ação antropogénica, as cidades registam valores mais elevados de temperatura em comparação com as áreas rurais circundantes. A constatação de que as temperaturas da cidade são mais elevadas, já foi estudada e documentada por um longo período temporal, referindo-se a este fenómeno como “ilhas de calor urbano”.

Ao longo dos últimos anos, esta temática tem registado grande destaque, em particular, nos meios de comunicação social, pela existência de um fenómeno denominado aquecimento global. Este fenómeno pode ser descrito como o aumento significativo das temperaturas a nível mundial e que com uso de modelos matemáticos de previsão do futuro climático foi perspectivado como sendo bastante significativo no futuro.

Contudo, para que possa ser efetivamente possível avaliar de uma forma mais precisa as futuras características climáticas do planeta, é importante perceber qual são as características do local onde estão implantadas as estações meteorológicas usadas nos modelos matemáticos. Isto porque nestas projeções são utilizados dados históricos das estações. Ora, se uma estação estava anteriormente localizada num ambiente rural e se, com o passar do tempo, sofreu ao seu redor um processo de urbanização, as temperaturas registadas passaram a verificar um aumento significativo. O uso de mais estações localizadas em ambiente urbano, em comparação com estações localizadas em ambiente rural nestes modelos tem suscitado o interesse científico, dado que pode ter ‘enviesado’ os resultados dos modelos das projeções climáticas.

Assim o principal objetivo deste trabalho é testar uma metodologia que ajude a melhorar a classificação de áreas urbanas a nível mundial feita no âmbito do projeto World Urban Database and Access Portal Tools (WUDAPT), para que se possa conhecer o ambiente envolvente das estações meteorológicas. Para isso, propõe-se uma metodologia que recorre a dados de Informação Geográfica Voluntária, especificamente, a dados do OpenStreetMap (OSM), para complementar os resultados da classificação de imagens de satélite nestas áreas.

A metodologia baseia-se na combinação, usando regras previamente estabelecidas, dos resultados obtidos através da classificação de imagens de satélite com a informação obtida a partir do OSM, depois de associada às classes usadas no projeto WUDAPT.

Os objetivos propostos inicialmente foram cumpridos, no entanto existem ainda limitações associadas à metodologia proposta, que deverão ser abordadas em trabalho futuro.

## Abstract

---

The increase in the world's population and its activities, and the increasing concentration of population in the urban centers, has resulted in alterations, not only in the physical characteristics of the terrestrial surface, but also in terms of the climate. As a result of this anthropogenic action, cities are experiencing higher temperature values compared to the surrounding rural areas. The evidence that the city temperatures are higher has been studied and documented for a long time, referred to as the "urban heat island."

Over the last few years, this issue has become particularly prominent in the media, due to the existence of a phenomenon called global warming. This phenomenon can be described as the significant increase of temperatures worldwide, which, with the use of mathematical models of the prediction of the future climate, has been foreseen as being quite significant in the future.

However, in order to be able to accurately assess the future climatic characteristics of the planet, it is important to understand the characteristics of the location of the meteorological stations used in mathematical models. This is important because, in these projections, historical data of the stations are used. However, if a station was previously located in a rural environment and if, over time, it had an urbanization process around it, the temperatures registered will have increased significantly. The use of more stations located in an urban environment compared to stations located in rural environments in these models has aroused scientific interest, since it may have 'biased' the results of the climate projection models.

Thus, the main objective of this work is to contribute and propose a methodology that will help to improve the precision of the classification of urban areas of the world made in the framework of the World Urban Database and Access Portal Tools (WUDAPT) project, so that the surrounding environment of the meteorological stations is better known. For this, a methodology is proposed that uses Volunteered Geographic Information, specifically OpenStreetMap data, to complement the results of the classification of satellite images in these areas.

The methodology was based on the combination, using previously established rules, of the results obtained in the classification of satellite images and the data available in OpenStreetMap once they are associated to the classes used in the WUDAPT project.

The objectives initially proposed were fulfilled; however, there are still some limitations associated with the proposed methodology, which should be addressed in future work.

# Summary

---

1. Introduction.....	1
1.1. Framework and motivation.....	1
1.2. Objectives and structure of the thesis .....	3
2. Theoretical background.....	5
2.1. GIS and Remote Sensing.....	5
2.1.1. Geographic Information.....	5
2.1.2. Remote Sensing.....	9
2.2. Volunteered Geographic Information .....	14
2.2.1. The OpenStreetMap project .....	15
2.2.2. VGI Quality .....	18
2.3. Local Climate Classification .....	20
2.4. World Urban Database and Access Portal Tools.....	28
3. Methodology.....	31
3.1. The World Urban Database and Access Portal Tools .....	31
3.2. Conversion of OSM data into LCZs.....	36
3.2.1. Pre-processing phase.....	36
3.2.2. Processing phase .....	37
3.3. Integration and combination of data.....	43
3.3.1. Integration of the outputs of the image classification and OSM data conversion .....	43
3.3.2. Combination of the data.....	44
3.4. Accuracy Assessment .....	46
4. Case studies.....	47
4.1. Study areas and data .....	47
4.1.1. Coimbra .....	47
4.1.2. Hamburg.....	50
4.2. Classification with the WUDAPT approach.....	52
4.2.1. Coimbra .....	52
4.2.2. Hamburg.....	59
4.3. Conversion of the OSM data into LCZ classes.....	67
4.3.1. Coimbra .....	67
4.3.2. Hamburg.....	70
4.4. Results of the data combination.....	78



4.4.1. Coimbra .....	78
4.4.2. Hamburg.....	84
5. Conclusions.....	91
6. References .....	93
Appendix A – Model I: Preparation of the variables.....	105
Appendix B – Model II: Preparation of the variables (Aggregation).....	106
Appendix C – Model III: OSM to LCZ A or B .....	107
Appendix D – Model IV: OSM to LCZ C.....	108
Appendix E – Model V: OSM to LCZ D.....	109
Appendix F – Model VI: OSM to LCZ G .....	110
Appendix G – Model VII: Obtain Impervious Surface Fraction (Roads and Railways).....	111
Appendix H – Model VIII: Obtain Building Surface Fraction .....	112
Appendix I – Model IX: OSM to urban LCZ class (LCZ 1 to LCZ 10) .....	113
Appendix J – Model X: Extract data from OSM maps.....	115
Appendix K – Model XI: Final procedure .....	116
Appendix L – Code Block to extract the percentage of times that a determinate class repeats on all classifications (summer, spring, autumn and winter) in each cell of the GRID .....	118
Appendix M – Code Block to extract the percentage of area occupied by each in each cell of the GRID .....	119
Appendix N – Confidence for each class regarding to satellite images for Coimbra .....	120
Appendix O – Confidence for each class regarding to satellite images for Hamburg.....	123

## Index of Figures

---

Figure 1: Raster model on the left and vector model on the right ( <i>in</i> Humboldt State University 2014; and Science 2017) .....	6
Figure 2: Raster and vector models schemes ( <i>in</i> Heywood, Cornelius, and Carver 2006, 78) .....	6
Figure 3: Google Earth vs Google Earth Pro tools ( <a href="https://www.google.com/intl/en-EN/earth/explore/products/desktop.html">https://www.google.com/intl/en-EN/earth/explore/products/desktop.html</a> ) .....	7
Figure 4: Landsat 8 satellite (U.S. Geological Survey 2016) .....	10
Figure 5: Random Forest scheme (Isied and Tamimi 2015) .....	13
Figure 6: Creating/Editing on OSM (December 13, 2016) .....	16
Figure 7: OpenStreetMap Primitives .....	16
Figure 8: Types of OSM features .....	17
Figure 9: Local Climate Zone types (Stewart and Oke, 2012) .....	23
Figure 10: WUDAPT's data hierarchy based on Mills et al. (2015); See et al. (2015) .....	29
Figure 11: Hierarchal methodology to collect data for levels 1, 2 and 3 (based on Mills et. al. (2015) and See et. al. (2015)).....	30
Figure 12: Workflow for the WUDAPT level 0 LCZ mapping procedure ( <i>in</i> Bechtel et al. 2015) .....	32
Figure 13: Step 1: "Digitize training areas" .....	33
Figure 14: Process and location of tools in SAGA software (Procedure 1) .....	34
Figure 15: Process and location of tools in SAGA software (Procedure 2) .....	34
Figure 16: Process and location of tools in SAGA software (Procedure 3) .....	35
Figure 17: Colouring of WUDAPT project.....	35
Figure 18: Scheme of the methodology used for the conversion of OSM data into LCZ classes A (Dense trees) or B (Scattered trees). .....	38
Figure 19: Scheme of the methodology used for the conversion of OSM data into LCZ class C (Bush, srub).....	38
Figure 20: Schematic of the methodology used for the conversion of OSM data into LCZ class D (Low plants) .....	39
Figure 21: Procedure for converting linear features into polygons.....	40
Figure 22: Schematic of the methodology used for classes LCZ G (Water) .....	40
Figure 23: Schematic of the methodology used for classes LCZ 1 to 10.....	41
Figure 24: Cells with different LCZ class .....	42

Figure 25: Representation of the created point feature and the vector GRID, overlaid with one map resulting from the classification of a Landsat 8 image into an LCZ map. ....	44
Figure 26: Example of the extraction of the percentage of the representative of this class for each pixel and for a particular class (x).....	45
Figure 27: Study area and the available OSM data .....	48
Figure 28: LCZ map created by the WUDAPT project for the region of Coimbra .....	48
Figure 29: Landsat 8 true color imagery of Coimbra in a) winter, b) autumn, c) summer and d) spring.....	49
Figure 30: False color composite RGB = 543 of the Landsat 8 imagery of Coimbra in a) winter, b) autumn, c) summer and d) spring.....	49
Figure 31: Study area and the available OSM data .....	50
Figure 32: LCZ map for Hamburg created by the WUDAPT project.....	51
Figure 33: Landsat 8 true colors imagery of Hamburg in a) winter, b) autumn, c) summer and d) spring.....	51
Figure 34: False color composite RGB = 543 of the Landsat 8 imagery of Hamburg in a) winter, b) autumn, c) summer and d) spring.....	52
Figure 35: Training areas over the Landsat 8 image (True-Color RGB) of Coimbra in a) winter, b) autumn, c) summer and d) spring .....	53
Figure 36: LCZs maps obtained for the city of Coimbra with a spatial resolution of 120 m with a filter of twenty-five neighboring cells for a) winter, b) autumn, c) summer and d) spring.....	53
Figure 37: LCZs maps obtained for the city of Coimbra with a spatial resolution of 30 m with a filter of sixty-four neighboring cells for a) winter, b) autumn, c) summer and d) spring.....	54
Figure 38: Training areas over the Landsat 8 image (True-Color RGB) of Hamburg in a) winter, b) autumn, c) summer and d) spring.....	60
Figure 39: LCZs maps for the city of Hamburg with a spatial resolution of 120 m with a filter of twenty-five neighboring cells for a) winter, b) autumn, c) summer and d) spring.....	60
Figure 40: LCZs maps for the city of Hamburg with a spatial resolution of 30 m with a filter of sixty-four neighboring cells for a) winter, b) autumn, c) summer and d) spring .....	61
Figure 41: a) Landsat 8 true color imagery (RGB 432). Results obtained from OSM for the percentage of area occupied in each cell by: b) LCZ A (Dense Trees) or LCZ B (Scattered Trees), c) LCZ class D (Low Plants), d) LCZ class G (Water), e) building surface fraction and f) impervious surface fraction.....	68
Figure 42: Confidence of presence of LCZ A in each pixel based on the classification of four satellite images, in percentage, for Coimbra.....	69

Figure 43: Confidence of presence of LCZ D in each pixel based on the classification of four satellite images, in percentage, for Coimbra.....	69
Figure 44: Results obtained from OSM for the percentage of area occupied in each cell by LCZ A (Dense Trees) or LCZ B (Scattered Trees).....	71
Figure 45: Results obtained from OSM for the percentage of area occupied in each cell by LCZ C (Bush, scrub) .....	71
Figure 46: Results obtained from OSM for the percentage of area occupied in each cell by LCZ D (Low plants).....	72
Figure 47: Results obtained from OSM for the percentage of area occupied in each cell by LCZ G (Water).....	72
Figure 48: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 3 (Compact low-rise).....	73
Figure 49: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 4 (Open high-rise) .....	74
Figure 50: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 5 (Open mid-rise).....	74
Figure 51: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 6 (Open low-rise).....	75
Figure 52: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 7 (Lightweight low-rise).....	75
Figure 53: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 9 (Sparsely built) .....	76
Figure 54: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 10 (Heavy industry).....	76
Figure 55: Confidence of the presence of LCZ A as a percentage for Hamburg.....	77
Figure 56: Confidence of presence of LCZ D as a percentage for Hamburg.....	77
Figure 57: Result of the procedure applied to Coimbra (the most likely class).....	78
Figure 58: Result of the procedure applied to Coimbra (the most likely class) for the Winter LCZ map .....	79
Figure 59: Result of the procedure applied to Coimbra (the most likely class) for the Autumn LCZ map .....	79
Figure 60: Results of the procedure applied to Coimbra (the most likely class) for the Summer LCZ map .....	80

Figure 61: Result of the procedure applied to Coimbra (the most likely class) for the Spring LCZ map .....	80
Figure 62: Result of the procedure applied to Hamburg (the most likely class) .....	84
Figure 63: Result of the procedure applied to Hamburg (the most likely class) for the Winter LCZ map .....	85
Figure 64: Results of the procedure applied to Hamburg (the most likely class) for the Autumn LCZ map .....	85
Figure 65: Result of the procedure applied to Hamburg (the most likely class) for the Summer LCZ map .....	86
Figure 66: Result of the procedure applied to Hamburg (the most likely class) for the Spring LCZ map .....	86
Figure 67: Confidence of presence of LCZ B (Scattered Trees) in values of percentage .....	120
Figure 68: Confidence of presence of LCZ G (Water) in values of percentage .....	120
Figure 69: Confidence of presence of LCZ 3 (Compact low-rise) in values of percentage.....	121
Figure 70: Confidence of presence of LCZ 6 (Open low-rise) in values of percentage.....	121
Figure 71: Confidence of presence of LCZ 8 (Large low-rise) in values of percentage .....	122
Figure 72: Confidence of presence of LCZ 9 (Scattered built) in values of percentage .....	122
Figure 73: Confidence of presence of LCZ B (Scattered Trees) in values of percentage .....	123
Figure 74: Confidence of presence of LCZ G (Water) in values of percentage .....	123
Figure 75: Confidence of presence of LCZ 1 (Compact high-rise) in values of percentage .....	124
Figure 76: Confidence of presence of LCZ 2 (Compact mid-rise) in values of percentage .....	124
Figure 77: Confidence of presence of LCZ 4 (Open high-rise) in values of percentage.....	125
Figure 78: Confidence of presence of LCZ 5 (Open midrise) in values of percentage .....	125
Figure 79: Confidence of presence of LCZ 6 (Open low-rise) in values of percentage.....	126
Figure 80: Confidence of presence of LCZ 8 (Large low-rise) in values of percentage .....	126
Figure 81: Confidence of presence of LCZ 10 (Heavy Industry) in values of percentage.....	127

## Index of Tables

---

Table 1: Characteristics of the Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) (from <a href="http://landsat.usgs.gov/best_spectral_bands_to_use.php">http://landsat.usgs.gov/best_spectral_bands_to_use.php</a> ).....	11
Table 2: Some multispectral classification methodologies (Jensen 2004; Alba 2014; Cortijo and Blanca 1996).....	12
Table 3: Description of surface cover and geometric properties for LCZs (Stewart and Oke 2012, 1886–87) .....	24
Table 4: Values of surface cover and geometric properties for LCZs (Stewart and Oke 2012, 1886–87) .....	24
Table 5: OSM keys and values .....	37
Table 6: Percentage of building and impervious surface fraction for classes LCZ 1 to 10 (in Stewart and Oke 2012).....	42
Table 7: Example of part of the attribute values associated with the GRID .....	43
Table 8: Dates of satellite images utilized for the LCZ classification for Coimbra.....	47
Table 9: Dates of satellite images utilized for LCZ classification for Hamburg.....	50
Table 10: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the winter satellite image for Coimbra. The user accuracy (UA), producer accuracy (PA) and overall accuracy (OA) are provided. Rows refer to classification output and columns to reference data. ....	55
Table 11: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the winter satellite image for Coimbra. ....	55
Table 12: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the autumn satellite image for Coimbra,.....	56
Table 13: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the autumn satellite image for Coimbra.....	56
Table 14: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the summer satellite image for Coimbra. ....	57
Table 15: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the summer satellite image for Coimbra. ....	58
Table 16: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the spring satellite image for Coimbra.....	58

Table 17: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the spring satellite image for Coimbra. ....	59
Table 18: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the winter satellite image for Hamburg.....	62
Table 19: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the winter satellite image for Hamburg.....	62
Table 20: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the autumn satellite image for Hamburg. ....	63
Table 21: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the autumn satellite image for Hamburg. ....	64
Table 22: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the summer satellite image for Hamburg.....	64
Table 23: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the summer satellite image for Hamburg. ....	65
Table 24: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the spring satellite image for Hamburg.....	66
Table 25: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the spring satellite image for Hamburg.....	66
Table 26: Overall accuracy before and after the data combination to Coimbra.....	81
Table 27: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the winter Landsat 8 image with a spatial resolution of 120 m. ....	81
Table 28: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the autumn Landsat 8 image with a spatial resolution of 120 m. ....	82
Table 29: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the summer Landsat 8 image with a spatial resolution of 120 m. ....	83
Table 30: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the spring Landsat 8 image with a spatial resolution of 120 m. ....	83
Table 31: Overall accuracy before and after the data combination was applied to Hamburg.....	87



Table 32: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the winter Landsat 8 image with a spatial resolution of 120 m. ....	87
Table 33: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the autumn Landsat 8 image with a spatial resolution of 120 m. ....	88
Table 34: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the summer Landsat 8 image with the spatial resolution of 120 m. ....	89
Table 35: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the spring Landsat 8 image with a spatial resolution of 120 m. ....	90

# I. Introduction

## I.1. Framework and motivation

Knowledge of the Land Use and Land Cover (LULC) of the Earth's surface is undoubtedly an important resource to assist in the development of local, regional and national-scale policies. Currently, it is recognized that the climate system is subject to induced environmental changes from natural or man-induced sources, so it is important that climate projections consider LULC change, since this can support the assessment of impacts in biodiversity as well as climate and ecosystem changes (Kondratyev and Cracknell 1998; Aquilué et al. 2017; Stewart and Oke 2012). In other words, the representation of the historical series of LULC patterns with the analyses of historical series of climatological variables (for example, temperature and precipitation) can be used to identify not only trends and their causes, but also to predict future climatological trends, based on updated data (Stewart and Oke 2012).

A large variety of climate models have been developed over several decades and provide results with considerable confidence, particularly at the global and continental scales. These models have generally shown a trend in the Earth's global warming and have raised concerns regarding greenhouses gases and the hole in the stratospheric ozone layer (IPCC 2007). However, these models, as with any mathematical representation of reality, still show significant errors that are greater overall when local scales are considered (IPCC 2007). The main source of most errors is the integration and representation of small-scale processes, since they cannot be represented explicitly in the models, and hence are integrated in a simplistic way and interact with larger-scale features (IPCC 2007).

The use of historical temperature series without consideration of the environment in which meteorological stations are located, i.e. whether in an urban or a rural landscape, can lead to misinterpretations and misclassifications of future climate trends. Hence, it becomes important to determine the influence of the values of temperature on the stations located in urban environments in the model's outputs. This could occur if, for example, a station, which was originally located in a rural environment, is subject to population growth and increasing urbanization (Stewart and Oke 2012), and hence is now located in an urban environment. In this case, the increase in

temperature values recorded at the station may not be due not to global temperature increases but rather to a change in the station environment. Moreover, the representativeness of rural and urban areas needs to be retained in the modelling process to eliminate any bias towards higher temperatures in the historical series that does not exist. Hence, knowledge of the actual land cover patterns of the Earth's surface at global scale is important to support the creation of future climate change trends and projections worldwide.

According to the United Nation World Population Prospects: the 2012 revision, the world's population of 7.2 billion in mid-2013 is expected to increase by one billion people within the next twelve years, reaching 9.6 billion in 2050 and 10.9 billion by 2100. The increase of world population and their activities influence the natural surface energy and radiation balance, so cities are relatively warm places compared to rural areas (Stewart and Oke 2012).

Evidence that air temperatures are often higher in cities than in the surrounding countryside was identified by Luke Howard in 1833. However, the scientific term "Urban Heat Island" (UHI), which is used to express this phenomenon, was only first recognized in the 40s, by Balchin and Pye (1947). This phenomenon is typical of the urban climate and occurs in almost all urban areas, large or small, in warm or cold climates (Stewart and Oke 2012). It also is an example of unintentional climate modification as a result of urbanization and anthropogenic activity. These factors induce changes in the physical characteristics of the surface (albedo, thermal capacity, heat conductivity, moisture), in radiative fluxes and in the near surface flow (Yang and Liu 2006).

However, for these types of studies, the classification of the environment into urban/rural has been shown to be too simplistic and therefore new classifications have been proposed to characterize the environment (further explained in section 2.3). Stewart and Oke (2012) proposed the Local Climate Zones (LCZ) classification based on a number of factors that characterize the physical characteristics of the region. A project has also been initiated to collect data on the form and function of cities around the world, namely the World Urban Database and Access Portal Tools (WUDAPT) project, which proposes a methodology to classify cities into LCZ classes and enable citizens around the world to participate. Due to the large quantities of information made available by citizens in other collaborative projects, additional data may be used to improve the LCZ classification. In this context, the main aim of this work is to identify if the data available in the collaborative project, OpenStreetMap (OSM), can be used to assist the process of creating more accurate LCZ classifications.

The importance of the topic described above and my personal interest in environmental issues were important motivations for the choice of this theme, in particular, the possibility to collaborate with the WUDAPT project and contribute to improving knowledge of the Earth's

surface that is used for the creation of more accurate models of climate prediction. Another key-factor for the choice of this theme was that it would give me the opportunity to work in several of the topics and disciplines studied during the Master course, namely Remote Sensing and Geographic Information Systems (GIS).

## **1.2. Objectives and structure of the thesis**

The main objective of this dissertation is to develop and test a methodology that explores the possibility of using Volunteered Geographic Information (VGI), namely data extracted from OSM, to assist in the creation of more accurate LCZ maps. The methodology proposed in the WUDAPT project is used to create the initial LCZ maps, and a methodology is developed to insert the data available in OSM into this process, in order to create more accurate LCZ maps.

The work was divided into several stages. In the first stage, the creation of LCZ maps using the methodology proposed by WUDAPT was carried out, including the accuracy assessment of the results. Subsequently, several processes for extracting, processing and transforming the data available in OSM were tested and developed. Then a methodology was developed to integrate the data created with the processes proposed by the WUDAPT project and the data extracted from OSM. Finally, the last step consisted of an analysis of the final results and the assessment of their accuracy.

This dissertation is subdivided into five chapters. In the first chapter an introduction is made to the dissertation, its objectives and the motivation that led to the choice of this subject. Subsequently, in the second chapter, some theoretical aspects are presented that can help in the contextualization of this work. This chapter describes some concepts of Remote Sensing, GIS, and VGI, as well as some of the tools and processes used within the dissertation. A more developed description of the local climate classification is presented, including the classification into LCZs. Finally, the WUDAPT project is presented.

Then, in the third chapter, the methodology used is described, including: 1) the methodology proposed by the WUDAPT project, which was also used in this thesis; 2) the processes used for converting OSM data into LCZ classes and 3) the integration of OSM data with the outputs of the LCZ classification performed with the WUDAPT project. The methodology used for the accuracy assessment of all the results is also described in this chapter.

In the fourth chapter, the methodology described in the previous chapter is applied to two areas of study, namely Coimbra and Hamburg. In this chapter, the location and characteristics of

the study areas are described, the results of all the processing is presented, as well as the confusion matrices and the user's, producer's and overall accuracy indices.

Finally, chapter five presents the conclusions, the main difficulties and limitations that have occurred throughout the study, as well as proposals for future developments and improvements.

## 2. Theoretical background

### 2.1. GIS and Remote Sensing

#### 2.1.1. Geographic Information

In this section, some basic notions of GIS that are used in this dissertation are presented, as well as the GIS software used.

Geographical information can be represented in geographical space using vector or raster model (Figures 1 and 2).

The raster model corresponds to a matrix of square cells. An attribute is associated with each cell of the matrix. In this model, the geographical space is considered discrete, since each cell (pixel) is considered as a spatial unit (Bonham-Carter 2014). The raster data structure can represent zero-dimensional objects by a cell or pixel, unidimensional elements by a sequence of cells or neighboring pixels and bi-dimensional elements, also by a sequence of cells or neighboring pixels (Bonham-Carter 2014).

The vector model, on the other hand, represents the phenomena or target-object as points, lines or polygon features (that can be filled – areas or features bound by lines). This model is appropriate to represent a large variety of different types of objects in comparison with the raster model, since in the raster model it is more difficult to capture the spatial position accurately, without making pixels very small, which would require higher storage costs and will make the processing of data much slower. In the vector model, in addition to the positional information and topology of the entities, it is possible to store additional information about each of the represented phenomena, called attributes.

It is possible to convert between these data structures but the use of one model over another depends on the aims of the project. In particular, the processing tools to use in each case are very different.

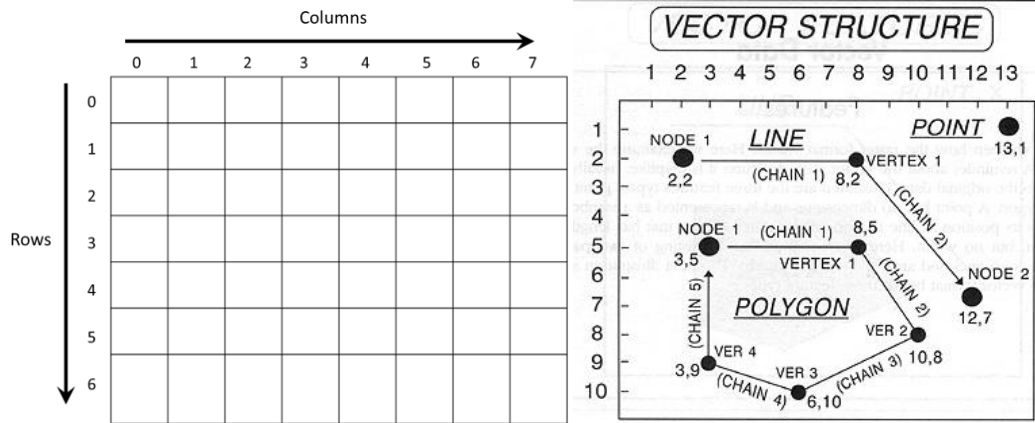


Figure 1: Raster model on the left and vector model on the right (in Humboldt State University 2014; and Science 2017)

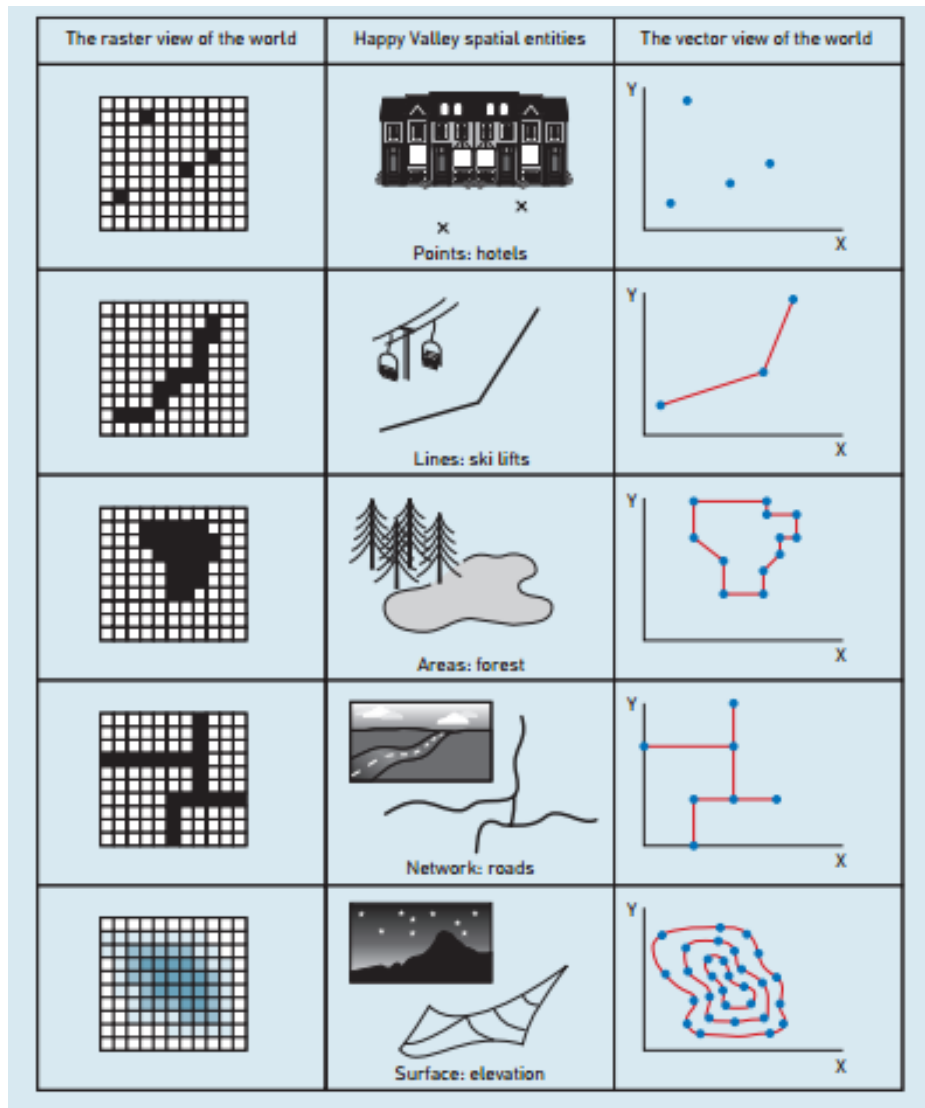


Figure 2: Raster and vector models schemes (in Heywood, Cornelius, and Carver 2006, 78)

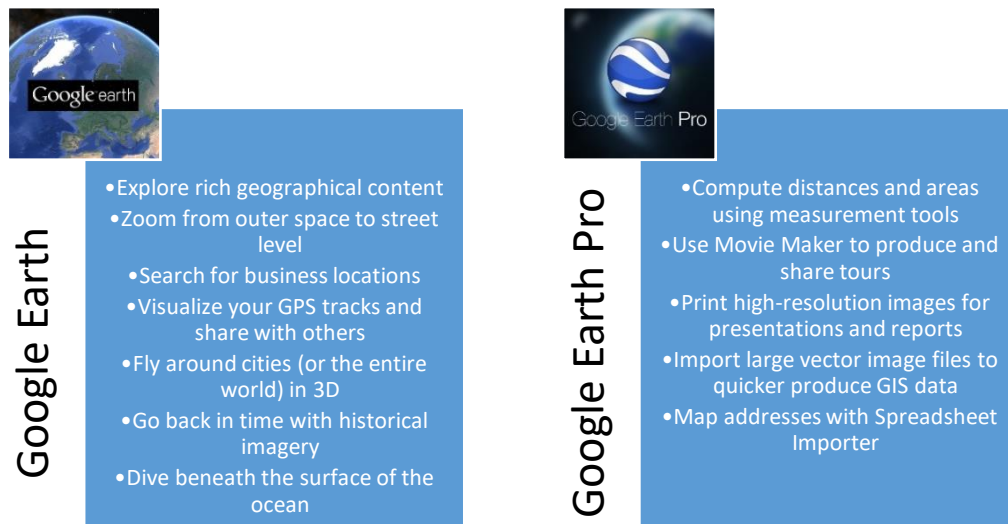


In this project, several software packages were used in different parts of the research, namely Google Earth, the System for Automated Geoscientific Analyses (SAGA) and ArcGIS software.

Google Earth software corresponds to ‘virtual globe’ software that allows 3D environmental data to be viewed, using a combination of digital elevation models, satellite imagery and 3D building envelopes (in some cities). This free software, which has been called “A 3D Interface to the Planet”, was presented to public in June 2005 and has attracted considerable public use and attention due to its ability to visualize landscapes in realistic three dimensions (Sheppard and Cizek 2009). The Google Earth software can be download as Google Earth or Google Earth Pro.

Google Earth allows the user to view any location on Earth, visualize space (i.e. the Sky, Moon and Mars - new features available only from Google Earth 5.0 or later versions), see geographical content, save toured places, and also share data with others.

Additionally, Google Earth Pro (Figure 3) provides extra capabilities such as Movie Maker and GIS data import capabilities (<https://www.google.com/earth/outreach/tools/index.html>). For over ten years, Google Earth Pro was not freely available, but since 30 January 2015, the software was made available for free download.



**Figure 3: Google Earth vs Google Earth Pro tools (<https://www.google.com/intl/en-EN/earth/explore/products/desktop.html>)**

The SAGA software can be classified as an open source system of geographic information, that was classified as free and available for public use since it was published in 2004. With the exception of the Application Programming Interface (API), the source code is licensed under the terms of the GNU General Public License.

SAGA has experienced a relatively fast evolution from a particular tool for digital terrain analysis to global GIS software, useful for geographical and scientific analysis and modeling

procedures. This software is coded in the C ++ programming language, it can be used in diverse operating systems, such as Windows and Linux, and is based on an object-oriented approach (Conrad et al. 2015).

SAGA was designed as a modular architecture that provides a user-friendly approach to users with little or no programming knowledge, with several visualization options. For users with programming knowledge, a command line interpreter is available that can interpret R and Python languages (Conrad et al. 2015).

ArcGIS is also a GIS software, but unlike SAGA, it is proprietary software. It is developed and commercialized by the Environmental Systems Research Institute, Inc (ESRI). This software has suffered a long evolution since it was originally developed for mainframe computers in a system based on command line into one with a graphical user interface (GUI), making the software more user-friendly and therefore available to a larger number of users. The old core of the ArcGIS system was called Arc/Info and included a basic set of programs: Arc, ArcPlot and ArcEdit and was command line based. As an intermediate product, ESRI developed ArcView, which was created primarily to view and analyze spatial data (Dixon and Uddameri 2016). ArcGIS was released in 2001 and emerged as a synthesis of both the powerful Arc/Info system with the easy-to-use interface ArcView.

ArcGIS contains several components, which include ArcMap (to create, display, analyze and edit spatial data and tables) and ArcCatalog (to view and manage spatial data files). It also includes ArcToolbox, a collection of tools and operations to which users can create and add their own tools for special or even often-used tasks (“ArcGIS Desktop” 2015). In ArcGIS 9.0, the Python programming language (a free, open-source, cross-platform and interpreted programming language) was introduced as a scripting language (“ArcGIS Desktop” 2015).

All the tasks that were performed in this research used the ModelBuilder, which is a visual programming language for building geoprocessing workflows, which allows spatial analysis and data management processes to be automated and documented. Several iterative models were created, which also included the creation of some block code in the Python language, mostly within the “Calculate field” tool.

### 2.1.2. Remote Sensing

Remote Sensing science is based on the fact that the objects on the terrestrial surface reflect electromagnetic radiation from the Sun. This is possible using sensors on board of satellites to infer the physical and chemical characteristics of objects on the Earth's surface, through the electromagnetic radiation reflected by those objects. Besides the reflected radiation, the objects emit their own radiation (with intensity and spectral composition dependent on the temperature), which is also used in remote sensing (Fonseca and Fernandes 2004).

There are many satellite platforms collecting data about the Earth's surface. However, the longest historical records of space-based Earth surface observations is provided by the Landsat series, with the first launch in 1972, followed by successive launches with chronological overlapping, namely, Landsat observatories (Landsat 2, 3, 4, 5, 7 and 8) that contributed to the knowledge of the Earth's surface by capturing images with higher spatial and spectral fidelity (Roy et al. 2014).

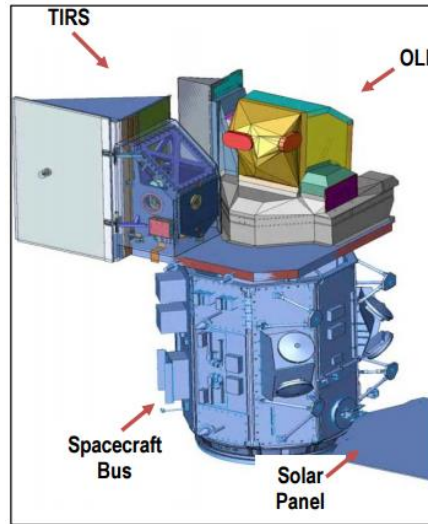
Landsat-8 satellite images are very important and useful for monitoring trends and evaluating land use changes because these satellites have collected data from the entire Earth's land surface with moderate resolution since they were launched on 11 February 2013. The applications are diverse, including land mapping and identification of changes in land cover (Knight and Kvaran 2014).

Landsat 8 satellites have a good temporal resolution, covering the entirety of the Earth's surface every 16 days, in an 8-day offset from Landsat 7. The data can be downloaded for free from GloVis, EarthExplorer or via the LandsatLook Viewer 24h after being received (<http://landsat.usgs.gov/landsat8.php>).

Landsat 8 transports two instruments (Figure 4): the Operational Land Imager (OLI) sensor, that measures nine spectral bands, where eight are multispectral, one is a pancromatic band and also a Quality Assessment band; and the Thermal Infrared Sensor (TIRS), that contributes to data acquisition with two thermal bands (Knight and Kvaran 2014).

The OLI band signal-to-noise ratios exceed those achieved by previous sensors, namely, by the Landsat "Enhanced Thematic Mapper Plus" (ETM+) of the Landsat 7 mission. These enhancements allow a 12-bit quantization (4096 levels) by the OLI and TIRS analog-to-digital converters, rather than the 8 bits (256 levels) used by Landsat ETM+. This upgraded signal-to-noise characteristics have improved measurements of land cover, reducing band saturation over highly reflective surfaces, such as snow or cloud (Roy et al. 2014). The 12-bit data are transformed into 16-bit integer data and distributed as Level 1 data products. All products are transformed into

55 000 levels of grey and this product can also be transformed by users into the Top of the Atmosphere (TOA) reflectance and/or radiance through the use of radiometric rescaling coefficients available on product metadata file (MTL file). Thus, Landsat 8 products are delivered as 16-bit images (U.S. Geological Survey 2016).



**Figure 4: Landsat 8 satellite (U.S. Geological Survey 2016)**

Landsat 8 images (Table 1) have eight spectral bands at 30 m resolution (OLI multispectral bands), from band 1 to 7 and 9, one panchromatic band with 15 m resolution (band 8) and two other thermal bands (TIRS) collected at 100 m resolution (band 10 and 11), which can be used to distinguish between surface covers.

The use of satellite data images involves performing image processing operations, namely, the pre-processing step. All Landsat standard products, namely, the data used in this study, are processed using the Level 1 Product, where a GeoTIFF output format is applied with a Cubic Convolution (CC) resampling technique. The map projection is Universal Transverse Mercator (UTM) with a World Geodetic System (WGS) 84 datum and a Map (North-up) image orientation ([http://landsat.usgs.gov/Landsat\\_Processing\\_Details.php](http://landsat.usgs.gov/Landsat_Processing_Details.php)).

The Cubic Convolution (CC) corresponds to an image resampling process, where each new pixel is interpolated from the existing pixel values. The image resampling process is necessary whenever the raster's structure (number of rows and columns) is modified during the projection, datum transformation or cell resizing operations (enlarged or reduced). CC resampling uses a weighted average of the 16 pixels nearest to the focal cell and produces the smoothest image compared to bilinear interpolation or nearest-neighbor resampling (Studley and Weber 2011).

**Table 1: Characteristics of the Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) (from [http://landsat.usgs.gov/best\\_spectral\\_bands\\_to\\_use.php](http://landsat.usgs.gov/best_spectral_bands_to_use.php))**

Band	Wavelength	Useful for mapping
Band 1: coastal aerosol	0,43 – 0,45 $\mu\text{m}$	Coastal and aerosol studies
Band 2: blue	0,45 – 0,51 $\mu\text{m}$	Bathymetric mapping, differentiate soil from vegetation and deciduous from coniferous vegetation
Band 3: green	0,53 – 0,59 $\mu\text{m}$	Highlights peak vegetation, which is useful for assessing plant vigor
Band 4: red	0,64 – 0,67 $\mu\text{m}$	Discriminates vegetation slopes
Band 5: Near Infrared (NIR)	0,85 – 0,88 $\mu\text{m}$	Emphasizes biomass content and shorelines
Band 6: Short-wave Infrared (SWIR) 1	1,57 – 1,65 $\mu\text{m}$	Distinguishes moisture content of soil and vegetation; penetrates thin clouds
Band 7: Short-wave Infrared (SWIR) 2	2,11 – 2,29 $\mu\text{m}$	Enhanced moisture content of soil and vegetation and thin cloud penetration
Band 8: Panchromatic	0,50 – 0,68 $\mu\text{m}$	15-meter resolution, sharper image definition
Band 9: – Cirrus	1,36 – 1,38 $\mu\text{m}$	Improved detection of cirrus cloud contamination
Band 10: TIRS 1	10,60 – 11,19 $\mu\text{m}$	100-meter resolution, thermal mapping and estimated soil moisture
Band 11: TIRS 2	11,5 – 12,51 $\mu\text{m}$	100-meter resolution, improved thermal mapping and estimated soil moisture

According to the U.S. Geological Survey (2016), the standard terrain correction (Level 1T -precision and terrain correction) is applied to Landsat 8 products and is available to users as radiometrically and geometrically corrected images. For that, inputs from both sensors and the spacecraft are used, as well as ground control points (GCPs) and Digital Elevation Models (DEMs). Hence, the result is “a geometrically rectified product free from distortions related to the sensor (e.g., view angle effects), satellite (e.g., attitude deviations from nominal), and Earth (e.g. rotation, curvature, relief). The image is also radiometrically corrected to remove relative detector differences, dark current bias, and some artifacts.” (U.S. Geological Survey 2016).

Remotely sensed data can be analyzed to extract thematic information, so *data* is transformed into *information*. The utilization of remote sensing as a source of information for characterizing LULC at local, regional and global scales has been increasing over time. One of the most often used techniques to extract information is image classification into LULC maps based on statistical pattern recognition procedures applied to multispectral remote sensing data (Jensen 2004). Image classification aims to automatically categorize the pixels in an image into a LULC class. It is therefore a process of reducing an image to informative classes. The categorization of

pixels is based on the values of grey from one or more spectral bands (Jensen 2004; Sousa and Marques 2011).

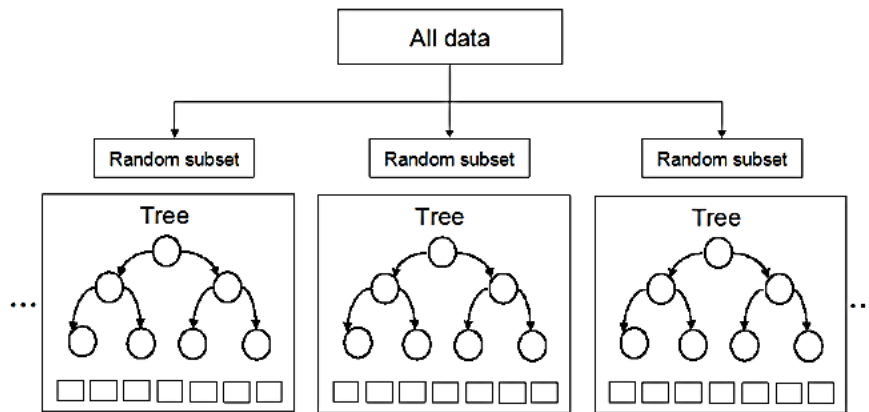
According to Jensen (2004), multispectral classification can be performed using a large variety of methodologies. These methods include parametric and nonparametric statistical algorithms (that use ratio and interval-scaled data) or non-metric procedures (that can similarly integrate nominal scale data). The classification methods can be classified into supervised or unsupervised, hard or soft (fuzzy), per-pixel or object-oriented or hybrid classification approaches. Table 2 summarizes these.

**Table 2: Some multispectral classification methodologies (Jensen 2004; Alba 2014; Cortijo and Blanca 1996)**

Method	Description	Examples
Parametric	These techniques assume that the data are normally distributed, in other words, this classifier assumes that there is an underlying probability in the distribution of the data. The parameters (for example, mean vector and covariance matrix) are frequently created from training samples.	- Maximum likelihood - Unsupervised classification
Non-parametric	These techniques are based on the fact that no assumption about the data is needed, in other words, they do not assume anything about the probability distribution. This technique does not employ statistical parameters for the computation of the separation of class.	- Nearest neighbor classifiers - Fuzzy classifiers - Neural networks
Nonmetric	These techniques can be used in both remote sensing with real-value data (for example, reflectance values from 0 to 100%) or nominal scale data (for example, class 1 = agriculture; class 2 = water)	- Rule-based decision tree classifiers
Supervised classification	In this classification, samples of classes (e.g. types of soil cover) in the image are identified. In supervised classification the training samples need to have field information. The procedures of this classification are based on the extraction of the spectral characteristics (signatures) of the training samples. Then, the classifier uses these signatures to classify the input data into a thematic map.	- Parallelepiped - Minimum distance to mean - Maximum likelihood - Others (hyperspectral matched filtering, spectral angle mapper)
Unsupervised classification	In this classification, the identity of the soil cover types is specified in classes. However, unlike the supervised classification, it does not require any field information. In other words, no prior information is needed about the classes. Pixels with similar characteristics are grouped in different 'clusters' according to some statistical parameters. The classifier later reclassifies and combines these spectral clusters into information classes.	- Chain method - Multiple-pass ISODATA - Fuzzy $c$ -means)

As mentioned above, imagery classification is a process of reducing an image to informative classes. Nonparametric classification represents those methods that are not based on the assumptions that the population (from which the sample is drawn) has a specific distribution, such as normal distribution (Bhar 2014).

Random Forest (Figure 5) is a widely-used algorithm for remote sensing image classification because of its capability to manage high dimensional and non-normally distributed data, making it a powerful option for integrating different imagery sources. This classification method, used in this research, is an ensemble classifier that produces numerous Classification and Regression (CART)-like trees, where each tree varies according to the values of a bootstrapped sample of the training data. In each bootstrapped training set, approximately one-third of the training data are left out. The input variables are also randomly selected for building the trees. The trees grown are not pruned (Millard and Richardson 2015; Breiman 2001).



**Figure 5: Random Forest scheme (Isied and Tamimi 2015)**

The algorithm used by this classifier delivers an accuracy assessment called “out-of-bag” error (rfOOB error) that according to Breiman (2001), eliminates the necessity for a set-aside test set. These out-of-bag estimates are calculated using the withheld training data and the measures of variable importance based on the mean decrease in accuracy when a variable was not used in building a tree (Millard and Richardson 2015). Breiman (2001) considered this rfOOB error as an independent assessment of accuracy since sample points used for error calculation are not used in building the trees of the “forest” classification (Millard and Richardson 2015).



## 2.2. Volunteered Geographic Information

Geographic information has traditionally been acquired by specialists using, for example, remote sensing, land surveys or photogrammetric methods to capture data that characterize both social and environmental phenomena of the Earth (Castelein et al. 2010). The phenomenal development of many geographic information related technologies, such as GPS-enabled cell phones and sensor technology, have opened up the possibility for citizens to collect geographic data (Michael Goodchild 2007a; Sui, Goodchild, and Elwood 2013; Castelein et al. 2010). This user generated geographic information has been named Volunteered Geographic Information (VGI) by Goodchild (2007b).

The development of the new Web 2.0 has also been fundamental in the development of VGI. The early Web was one-directional, where a relatively small number of sites existed and a huge number of users, while the Web 2.0 is bi-directional, and users can interact and provide information to central sites, see the information collected and also make the data available to others. One example of this type of crowd-generated content is Wikipedia (<https://www.wikipedia.org>), where users can provide information to the project that is managed by a relatively small group of reviewers (Goodchild 2007b).

The combination of web developments along with portable positioning devices have triggered a large variety of VGI activities, such as the creation of maps by walking, cycling or driving or geocaching activities, where the participants have to find hidden destinations based on their coordinates (Goodchild 2007a).

Many examples of interactive platforms such as OSM (<https://www.openstreetmap.org/>), Wikimapia (<http://wikimapia.org>), Google Maps (<https://www.google.pt/maps>), Flickr (<https://www.flickr.com>) or the Geo-Wiki Project (<http://www.geo-wiki.org/>) make it possible for every citizen to disseminate their own georeferenced data, such as maps, photographs or other types of geographic information, contributing to sparking a large growth in VGI content (Castelein et al. 2010).

The use of VGI for charitable projects or the mapping of regions affected by natural disasters, such as the OSM mapping to support rescue response after Haiti's 7.0 magnitude earthquake in 2010, resulted in more attention being focused on VGI. Examples such as the Humanitarian OpenStreetMap Team (HOT) played a significant role through projects like the Sri Lanka Flooding in 2016 or the Nepal Earthquake Response in 2015, among others (Neis and Zielstra 2014). This has contributed to the acceptance of citizens as a legitimate source of observations that can be used by the scientific community, contributing to so called 'citizen science', which is now respected and recognized in some areas (Goodchild 2007b).

### 2.2.1. The OpenStreetMap project

The OSM project was started in 2004. In the beginning, all databases and web services were hosted on several servers at university domains of University College London, but over time, with the growth of the project and through numerous donations, it was possible to establish additional server infrastructures. The fundamental goal of the project was to build “a free database with geographic information” (Neis and Zielstra 2014, 79) and most of the all servers and interfaces of the project have been developed and administrated by volunteers (Neis and Zielstra 2014).

In this project, OSM volunteers contribute and maintain data on roads, railways, buildings, land use and many other types of information around the world. The OpenStreetMap Foundation (OSMF) has managed the basic design services and legal rights of this project and the data are available under an Open Database (ODbL) license. This allows commercial use of the data as long as a reference is made to the OSM project. Moreover, the data and their derivative products must be released under the same license or another compatible one (<https://www.openstreetmap.org/>).

The information available in OSM is in vector format and each element has an associated spatial dimension (geometry) and an informative dimension (attribute data). When an OSM contributor creates an object (Figure 6) that represents a real-world feature, the volunteer is able to use three types of primitives (Figure 7): nodes (points), ways (polylines, closed ways or areas/polygons) (<http://wiki.openstreetmap.org/wiki/Elements>) and relations (logical collections of two or more nodes, ways, areas or other relations) (Neis and Zielstra 2014).

In OSM, each element has a spatial and an informative dimension, so each of these primitives (namely, nodes, ways and relations) are associated with one or more attributes, also referred as tags. Tags are used to describe information such as the type of object (e.g. restaurant, street, etc.) and their most relevant details (e.g. address, if access is restricted, etc.). Each tag is formed by a key and a value; e.g. to identify a wood, the tag "natural=wood" must be used, where "natural" is the key and "wood" is the value. OSM contributors can use their own tags, although there is an official collection of tags that have been established and agreed by the OSM community, including explanations and examples on their use ([http://wiki.openstreetmap.org/wiki/Map\\_Features](http://wiki.openstreetmap.org/wiki/Map_Features)).

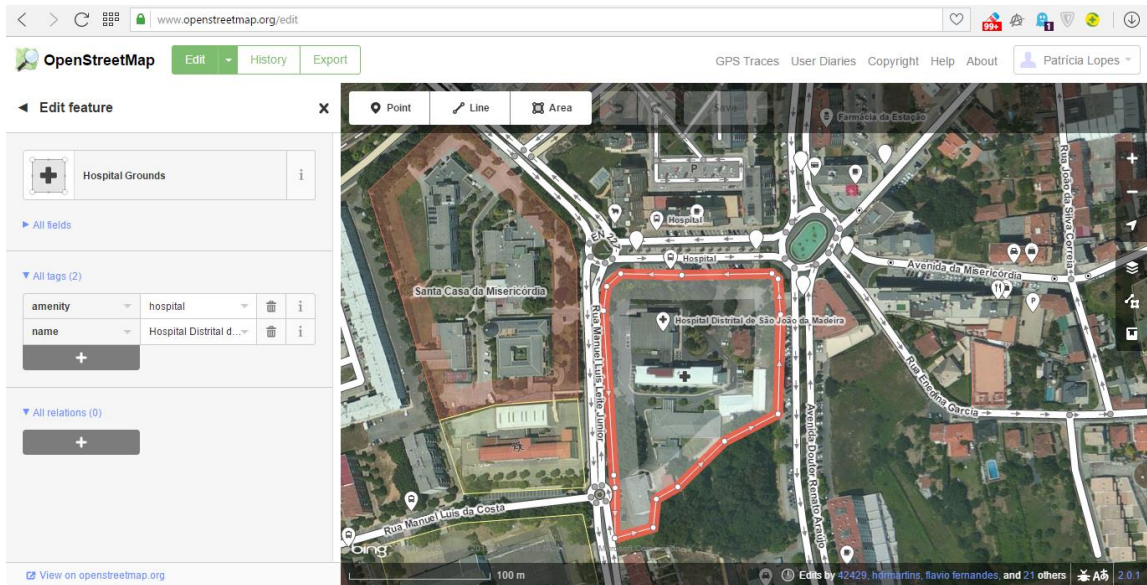


Figure 6: Creating/Editing on OSM (December 13, 2016)

Elements			
	<p><b>Node:</b></p> <ul style="list-style-type: none"> <li>- This is a unique point on the Earth and it is distinguished through its latitude, longitude coordinates and node id;</li> <li>-Altitude can also be included (optional).</li> </ul>	<p><b>Way:</b></p> <ul style="list-style-type: none"> <li>- Consists of a hierarchically organized list of successive nodes, namely, open ways or closed ways;</li> <li>- In open ways, the first point differs from the final point and describes linear features, such as many roads, railways and waterways;</li> <li>- In closed ways, the first point corresponds to the last point. Closed ways can be either an area or closed polyline or even both.</li> </ul>	<p><b>Relation:</b></p> <ul style="list-style-type: none"> <li>- Is a multi-purpose data structure that describes a relationship between at least two data components (nodes, ways, and/or other relations).</li> </ul>

Figure 7: OpenStreetMap Primitives

The OSM project organizes contributions according to numerous layers, including the ones represented in Figure 8, i.e. points of interest (POIs), roads, waterways, railways, land use, natural and buildings. In general, there are more contributions relating to roads, POIs and buildings than relate to other features (Arsanjani et al. 2013). It is important to note that the information available in OSM for each area varies widely, so in many regions, not all of the information represented in Figure 8 is available.

### Point Features



### Linear Features



### Polygon Features

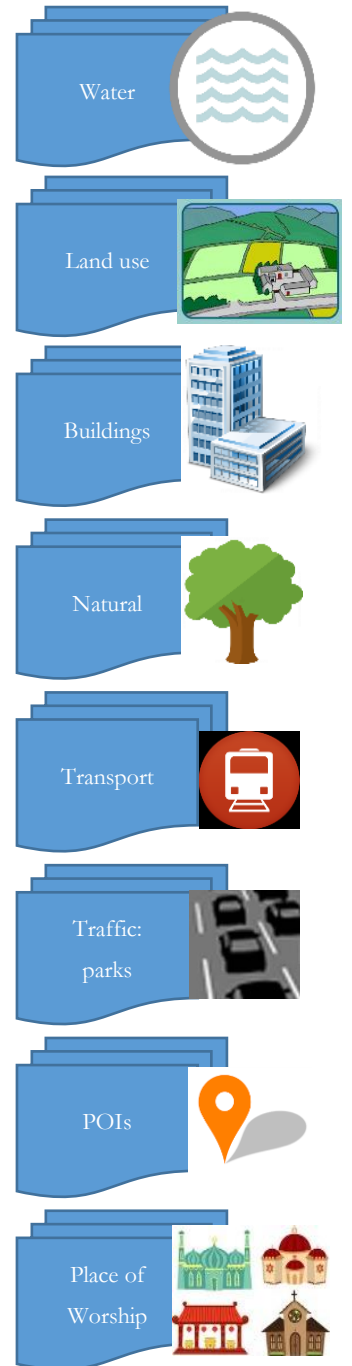


Figure 8: Types of OSM features

The use of OSM to generate land-use (LU) patterns was tested and applied by Arsanjani et al. (2013) in Vienna, Austria, where the OSM data are largely available. In this study, the Land-use map obtained using OSM data was compared with a propriety dataset, namely, the Global Monitoring for Environment and Security Urban Atlas (GMESUA). The outcomes showed acceptable accuracy. Thus, it was concluded that “VGI can be a potential source for mapping LU patterns” (Arsanjani et al. 2013, 13).

The use of OSM data as the only source of data for Land Use and Land Cover Maps (LULCM) has also been tested and confirmed in regions with high levels of available data in OSM by Martinho and Fonte (2015), namely for two regions located in London and in France. An automated methodology was developed to convert OSM features into LULCM, which through a hierarchical approach, a set of decision rules and spatial analysis, solves some types of inconsistencies (Fonte et al. 2016). The outcomes demonstrated satisfactory results but the problems inherent to this procedure are still present, namely, empty regions on the LULCM map or areas with no data in OSM, the fact that volunteers can classify each feature freely and also the predefined categories that can vary between study areas.

According to Arsanjani et al. (2013), the main advantage of VGI, and particularly of OSM, for the collection of LULC data, is that it is able to provide some data that have not been collected before and that are missing in authoritative databases.

### **2.2.2. VGI Quality**

One of the main problems associated with VGI data is its quality. The quality of geographic information includes a large variety of aspects and standards that should be respected. Namely, the lineage or chronological nature of information (history of dataset), the positional (exact accurate location) and attribute accuracy (the values on attribute information corresponding to the spatial data represented), logical consistency (rules to adapt and optimize data structures and storage), completeness (percentage of data available in relation to all data that should be represented), semantic accuracy (reliability of geographical entity types and their attributes ), usage, its purpose and constraints, and lastly, the temporal quality (Oort 2005).

There have been a number of research studies that have addressed this issue. The evaluation of all the quality parameters is not easy, so most authors only address a few of them. For example, Haklay (2010) only studied two elements of VGI quality, namely, positional accuracy and completeness of OSM data in relation to a United Kingdom (UK) government dataset provided by the Ordnance Survey, in the city of London – England (Haklay, 2010). The author concluded

that information from OSM can be fairly accurate, but also raised many questions about VGI quality that can and should be explored.

Ribeiro and Fonte (2015) developed an automated methodology to assess the degree of coverage of OSM data, computing an indicator of completeness.

Girres and Touya (2010) studied OSM data accuracy and selected France as the study area. Hence, they computed a comparison of OSM data with BD TOPO from the French National Institute of Geographic and Forest Information (IGN). The authors outlined the main advantages of OSM data such as its responsiveness and flexibility but also referred to the very high heterogeneity in several domains, such as attribute variety, limiting a large possibility of applications. These heterogeneities occur because of the lack of standardized and precise specifications.

Neis, Zielstra, and Zipf (2011) compared OSM roads or street networks with commercial spatial data (of TomTom Multinet 2011) in Germany. The completeness results of OSM were very satisfying, where the information of the total network exceeded that provided by the authoritative dataset by 27%, and for car routes and navigation, OSM only missed 9% of the commercial data.

Senaratne et al. (2017), in a literature review, used query techniques in Google Scholar with the aim of investigating all studies starting from 2007 up until the middle of 2015 that included some words or expressions both on title or abstract, namely, “quality assessment, methods and techniques, uncertainty, volunteered geographic information, map, microblog, photo” (Senaratne et al. 2016, 8). The results showed 425 academic studies. Thus, the authors minimized the requested parameters according to relevance ranking in Google Scholar, refining their assembly of papers by establishing some criteria. Namely, the papers must explore methodologies or tools for quality assessment, and when multiple studies used analogous procedures, the newest research was selected. This resulted in a total of 56 papers selected for further analysis.

This study, and others, demonstrate that despite the amount of literature produced over time on this subject, the advantages and disadvantages of VGI have been largely explored and analyzed. Nevertheless, the potential of this information has also been proven by a large variety of studies about urban management administration (Song, Chen, and Guo 2009), flood damage assessment (Poser and Dransch 2010), wild fire evacuation (Pultar et al. 2009) and natural disaster management or responses (Ostermann and Spinsanti 2011; Manfré et al. 2012; Horita and de Albuquerque 2013; M. Goodchild, Glennon, and Glennon 2010; Neis, Singler, and Zipf 2010; Bono and Gutiérrez 2011) (*in* Neis and Zielstra 2014).

### 2.3. Local Climate Classification

Over time, many studies of the UHI effect have been developed and the typical UHI studies were based on only two approaches; in one of them, the temperature is retrieved at meteorological stations located at urban or rural places, and the other is based on temperature values obtained along a transect from rural to urban areas using a mobile device (such as cars or bicycles, for example). In both approaches, sites are classified as either urban or rural and the UHI effect is estimated by comparing urban (U) values with rural (R) values or the background temperatures ( $\Delta T_{U-R}$ ) (Alexander and Mills 2014; Stewart and Oke 2012).

The background temperatures, or the UHI intensity ( $\Delta T_{U-R}$ ), is defined as the temperature difference between the maximum urban temperature and its surrounding rural temperatures (Oke 1987 in Santamouris 2013; Alexander and Mills 2014; János Unger, Lelovics, and Gál 2014; Stevan et al. 2013). Thus, the typical UHI measurement studies that have been developed for many decades were based on single comparisons of "urban" and "rural" air temperatures.

However, the terms rural and urban cannot be defined universally, given that they have a unique objective meaning, and thus also no climatological relevance. Furthermore, what is defined as urban or rural in a certain city can be different for another city (Stewart and Oke 2012).

In recent decades, urban theorists advocated that the spatial borderline between an urban region and the countryside is not natural, and that the relation should be more precisely represented as a continuum, as opposed to a dichotomy of urban/rural (Gugler 1996).

According to Chandler (1965), the climate of a city is a component of numerous factors, among which urban development is only one. He stated that there are three fundamental determining factors: the proprieties of the general climate of the region; the physical or geomorphic characteristics; and the built form of a town. According to Stewart and Oke (2012), Chandler (1965) was perhaps the first specialist to create a classification of a city based on climate characteristics.

Auer (1978) developed another urban-rural classification for St. Louis, located in Missouri, USA, where he partitioned the city into twelve types of land use, established according to the vegetation of the city and the utilization and the structures of the buildings.

Ellefsen (1991) grouped ten North American urban areas into a system of seventeen classes of neighborhood types referred to as "Urban Terrain Zones" (UTZ), which were based on building continuity, construction materials, and street configuration (Stewart and Oke, 2012). Combining the features of both Auer's and Ellefsen's classifications, Oke (2004, 2008) designed a classification called "urban climate zones" (UCZs), dividing the city terrain into seven homogenous regions, which ranged from semi-rural to intensely-developed cities.

Stewart and Oke (2009) applied a classification system that divides the landscape universe into nineteen local climate zone classes based on three landscape proprieties, namely, surface cover, surface structure and cultural activity. This was an initial classification system that was subject to upgrades. These nineteen classes are structured into four landscape series, namely city series, agricultural series, natural series and mixed series. For example, the natural series are divided into five local climate zones, namely, forest, wetland, grassland, tundra and hot desert. Their study area was the Nagano basin of central Honshu, Japan and they selected seven field sites from the observational urban heat island studies of Sakakibara (1999) and Sakakibara and Matsui (2005).

According to Stewart and Oke (2012), despite their many advantages, the previously mentioned classifications have limitations; namely, not all classifications use a full set of surface climate properties in the definitions of their classes. This full set must include “the urban structure (dimensions of buildings and the spaces between them, the street width and street spacing), the urban cover (built-up, paved, vegetated, bare soil, water), the urban fabric (construction and natural materials) and the urban metabolism (heat, water and pollutants due to human activity)”. Secondly, the systems that do not use any class for rural landscapes or systems that use elements (such as classes names or definitions), which are specific to a region, are not reliable for these studies. The authors state that all the previous classifications are based on developed cities, so if they were used in underdeveloped cities they would present several limitations.

To provide a universal answer for urban temperature studies, Stewart and Oke (2012) established another framework, a new upgraded classification system referred to as "Local Climate Zones" (LCZs) for urban temperature studies.

The LCZs classification system divides the landscape into seventeen classes (Figure 9). LCZs are areas of “uniform surface cover, structure, material, and human activity that span from hundreds of meters to several kilometers in horizontal scale” (Stewart and Oke 2012, 1884). This classification divides LCZs into ten built-up types (from LCZ 1 to 10), seven land cover types (from LCZ A to G), and additionally, variable seasonal types. These last types consist of variable seasonal or short period land cover properties, referenced, for example, by LCZ1s when the surface is covered by snow.

All classes emerge from the logical division of the landscape, according to different properties that influence screen-height temperatures, i.e. temperatures measured from one to two meters above the ground (Stewart, Oke, and Krayenhoff 2014), such as surface structure and surface cover. In this way, each "LCZ has a characteristic screen-height temperature regime that is most apparent on dry surfaces, on calm, clear night and in areas of simple relief" (Stewart and Oke 2012, 1884). LCZ types can be distinguished by ranges of typical values of measurable physical



properties, which characterize the geometry and surface cover, and the thermal, radiative and anthropogenic energy features of the surface. Tables 3 and 4 show the description of the properties used to characterize the LCZs and the values associated with each class, respectively.

The LCZ system can be classified as generic, as it is impossible to consider every peculiarity of each place, be it urban or rural. Moreover, in all classifications, the descriptive and explanatory capabilities are limited. However, the seventeen typologies should be familiar to the majority of any city's dwellers and should be adaptable to the local features (Stewart and Oke, 2012).

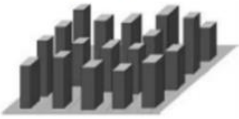
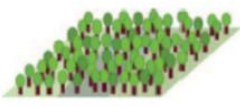

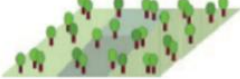

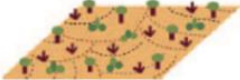
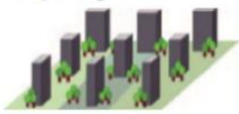



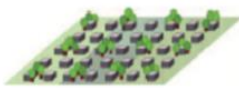

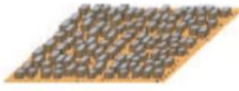
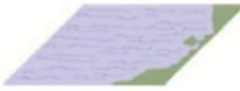



Built types	Definition	Land cover types	Definition
 <p>1. Compact high-rise</p>	Dense mix of tall buildings to tens of stories. Few or no trees. Land cover mostly paved. Concrete, steel, stone, and glass construction materials.	 <p>A. Dense trees</p>	Heavily wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
 <p>2. Compact midrise</p>	Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	 <p>B. Scattered trees</p>	Lightly wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
 <p>3. Compact low-rise</p>	Dense mix of low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	 <p>C. Bush, scrub</p>	Open arrangement of bushes, shrubs, and short, woody trees. Land cover mostly pervious (bare soil or sand). Zone function is natural scrubland or agriculture.
 <p>4. Open high-rise</p>	Open arrangement of tall buildings to tens of stories. Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	 <p>D. Low plants</p>	Featureless landscape of grass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park.
 <p>5. Open midrise</p>	Open arrangement of midrise buildings (3–9 stories). Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	 <p>E. Bare rock or paved</p>	Featureless landscape of rock or paved cover. Few or no trees or plants. Zone function is natural desert (rock) or urban transportation.
 <p>6. Open low-rise</p>	Open arrangement of low-rise buildings (1–3 stories). Abundance of pervious land cover (low plants, scattered trees). Wood, brick, stone, tile, and concrete construction materials.	 <p>F. Bare soil or sand</p>	Featureless landscape of soil or sand cover. Few or no trees or plants. Zone function is natural desert or agriculture.
 <p>7. Lightweight low-rise</p>	Dense mix of single-story buildings. Few or no trees. Land cover mostly hard-packed. Lightweight construction materials (e.g., wood, thatch, corrugated metal).	 <p>G. Water</p>	Large, open water bodies such as seas and lakes, or small bodies such as rivers, reservoirs, and lagoons.
 <p>8. Large low-rise</p>	Open arrangement of large low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Steel, concrete, metal, and stone construction materials.	<b>VARIABLE LAND COVER PROPERTIES</b>	
 <p>9. Sparsely built</p>	Sparse arrangement of small or medium-sized buildings in a natural setting. Abundance of pervious land cover (low plants, scattered trees).	<i>b. bare trees</i>	Leafless deciduous trees (e.g., winter). Increased sky view factor. Reduced albedo.
 <p>10. Heavy industry</p>	Low-rise and midrise industrial structures (towers, tanks, stacks). Few or no trees. Land cover mostly paved or hard-packed. Metal, steel, and concrete construction materials.	<i>s. snow cover</i>	Snow cover >10 cm in depth. Low admittance. High albedo.
		<i>d. dry ground</i>	Parched soil. Low admittance. Large Bowen ratio. Increased albedo.
		<i>w. wet ground</i>	Waterlogged soil. High admittance. Small Bowen ratio. Reduced albedo.

Figure 9: Local Climate Zone types (Stewart and Oke, 2012)

**Table 3: Description of surface cover and geometric properties for LCZs (Stewart and Oke 2012, 1886–87)**

	Properties	Properties description
Surface Structure	Sky view factor: $\psi_{sky}$	Proportion of the measure of half of earth sky seeable from terrain level to that of a clear hemisphere
	Aspect ratio: H/R	Mean height-to-width proportion of roads canyons (LCZs 1-7), building arrangement (LCZs 8-10), and tree arrangement (LCZs A-G)
	Building surface fraction: $\lambda_b$	Proportion of building cover area from total area (%)
	Impervious surface fraction: $\lambda_i$	Proportion of impervious area (namely, rock or paved) from total area (%)
	Pervious surface fraction: $\lambda_v$	Proportion of pervious area (bare soil, vegetation, water) from total area (%)
	Height of roughness elements (m): $Z_H$	Geometric mean of building statures (LCZs 1-10) and tree/plant statures (LCZs A-F)
	Terrain roughness class: $z_0$	Classification in Davenport et. al.'s (2000) of effective terrain harshness ( $z_0$ ) for city and nation scenes.
Surface Cover	Surface admittance ( $J m^{-2} s^{-1/2} K^{-1}$ ): $\mu$	The capacity of a surface to receive or discharge heat. This propriety differs with soil wetness and material thickness.
	Surface albedo: $\alpha$	The proportion of the quantity of solar radiation reflected by a surface to quantity received by it. This propriety differs with surface color, wetness, and roughness.
	Anthropogenic heat output ( $W m^{-2}$ ): $Q_F$	Mean annual warmth flux density from fuel burning and anthropological activity. Changes fundamentally with latitude, season and quantitate of the population.

**Table 4: Values of surface cover and geometric properties for LCZs (Stewart and Oke 2012, 1886–87)**

LCZ	$\psi_{sky}$	H/R	$\lambda_b$	$\lambda_i$	$\lambda_v$	$Z_H$	$z_0$	$\mu$	$\alpha$	$Q_F$
LCZ 1	0.2-0.4	>2	40-60	40-60	40-60	>25	8	1500-1800	0.10-0.20	50-300
LCZ 2	0.3-0.6	0.75-2	40-70	30-50	<20	10-25	6-7	1500-2200	0.10-0.20	<75
LCZ 3	0.2-0.6	0.75-1.5	40-70	20-50	<30	3-10	6	1200-1800	0.10-0.20	<75
LCZ 4	0.5-0.7	0.75-1.25	20-40	30-40	30-40	>25	7-8	1400-1800	0.12-0.25	<50
LCZ 5	0.5-0.8	0.3-0.75	20-40	30-50	20-40	10-25	5-6	1400-1200	0.12-0.25	<25
LCZ 6	0.6-0.9	0.3-0.75	20-40	20-50	30-60	3-10	5-6	1200-1800	0.12-0.25	<25
LCZ 7	0.2-0.5	1-2	60-90	<20	<30	2-4	4-5	800-1500	0.15-0.35	<35
LCZ 8	>0.7	0.1-0.3	30-50	40-50	<20	3-10	5	1200-1800	0.15-0.25	<50
LCZ 9	>0.8	0.1-0.25	10-20	<20	60-80	3-10	5-6	1000-1800	0.15-0.25	<10
LCZ 10	0.6-0.9	0.2-0.5	20-30	20-40	40-50	5-15	5-6	1000-2500	0.12-0.20	>300
LCZ A	<0.4	>1	<10	<10	>90	3-30	8	unknown	0.10-0.20	0
LCZ B	0.5-0.8	0.25-0.75	<10	<10	>90	3-15	5-6	1000-1800	0.15-0.25	0
LCZ C	0.7-0.9	0.25-1.0	<10	<10	>90	<2	4-5	700-1500	0.15-0.30	0
LCZ D	>0.9	<0.1	<10	<10	>90	<1	3-4	1200-1600	0.15-0.25	0
LCZ E	>0.9	<0.1	<10	<10	>90	<0.25	1-2	1200-2500	0.15-0.30	0
LCZ F	>0.9	<0.1	<10	<10	>90	<0.25	1-2	600-1400	0.20-0.35	0
LCZ G	>0.9	<0.1	<10	<10	>90	-	1	1500	0.02-0.10	0

The LCZ system can be classified as generic, as it is impossible to consider every peculiarity of each place, be it urban or rural. Moreover, in all classifications, the descriptive and explanatory capabilities are limited. However, the seventeen typologies should be familiar to the majority of any city's dwellers and should be adaptable to the local features (Stewart and Oke, 2012).

The classification methodology of LCZs has evolved throughout the years. Perera, Emmanuel, and Mahanama (2012) applied the classification to the city of Colombo, Sri Lanka, a warm and humid region. For the creation of the LCZ map, they used observational data, namely superficial temperature measures and photographs (fisheye lens photographs) and also published normative values (Oke 1987; Arnfield 1982). The authors also utilized the Surface Heat Island Model (SHIM), developed by Johnson et al. (1991), with the urban fabric LCZ classification, to simulate the local effects of warming.

A different methodology was used by Emmanuel and Krüger (2012) for the LCZ classification of the city of Glasgow, United Kingdom. The aim of the authors was to study the climatic changes of the area using three data sources, namely the UK Meteorological Office historical data for Glasgow (50-year historical series), the Weather Underground Network and the MIDAS Surface Weather Stations network. The LCZ map was created with the purpose of characterizing the Land Usage and Land Cover patterns of the regions surrounding the weather stations.

Also in 2012, and for the first time, Bechtel and Daneke used multiple observational data, namely thermal and multispectral satellite imagery, a normalized digital terrain model and different classifiers, including Support Vector Machines, Neural Networks, and Random Forest to map LCZs.

Meng and Liu (2013) used time series between 1981 and 2011, from Landsat Thematic Mapper (TM)/Enhanced Thematic Mapper Plus (ETM+) satellite imagery to extract LULC and Land Surface Temperature (LST) in the city of Jinan, China. The purpose was to analyze the Urban Heat Island patterns in a city that experienced a fast-urban growth during the aforementioned period. Although this study is not directly related to LCZ classification, it is mentioned here because it used satellite imagery to obtain LULC data. This procedure was later implemented by Alexander and Mills (2014), as will be explained in greater detail later.

The classification methodology without remote sensing images is again used for the LCZ classification by Stevan et al. (2013) and Thomas et al. (2014), for Novi Sad, Serbia, and Kochi, India, respectively. Stevan et al. (2013) used field work and data (using meteorological stations) aerial photographs, topographical maps, satellite imagery (Google Earth) and normative values. In the second study by Thomas et al. (2014), the study areas were classified based on site

measurements in all territories, with the creation of a grid of 100\*100 m within a 2 km radius around the meteorological stations, and a 500\*500 m grid for the rest of the area. Using an electronic distance meter and a GPS, measurements were made to obtain the following properties: building aspect ratio, building surface fraction and mean building height. The sky view factor was derived from the height and length of the buildings and the width of the roads. The remaining properties were obtained from field work and with the help of images from Google, so the study area was classified into LCZs.

Alexander and Mills (2014) applied the LCZ classification to the city of Dublin, Ireland, as grounds for a study of the UHI, as already undertaken by Emmanuel and Krüger (2012) in the past. The authors used LULC data from the CORINE program (CO-ordination of INformation on the Environment) for the transformation of the study field into LCZ classes. The process involved the creation of a grid of 1 km<sup>2</sup> resolution. In the first step, all the grid cells were transformed and coded into a LULC category using CORINE data and a random sample of cells were then selected for examination. The selected cells were scrutinized with data from Google Earth, Bing Maps and field work, and an LCZ class was then assigned. The data were subsequently used for the transformation of the all grid cells into corresponding LCZ classes.

Unger, Lelovics, and Gál (2014) developed a semiautomatic methodology based on geographic information systems and LCZ parameters. This procedure was also used to generate LCZ maps in Szeged, Hungary. The methodology was based on obtaining seven properties for the study area. Aspect ratio, surface admittance and anthropogenic heat output were properties that were omitted. The height of roughness elements, the sky view factor and the building surface fraction (BSF) properties were calculated using the 3D building database of Szeged. For the pervious surface fraction (PSF), RapidEye (2012) satellite imagery was used for calculating the Normalized Difference Vegetation Index (NDVI), while a 1:25000 topographic map, a road database and the CORINE Land Cover database were also employed. The impervious surface fraction (ISF) was calculated as the paved area outside the buildings, so  $ISF = 1 - (BSF + PSF)$ . Finally, the surface albedo was obtained using atmospherically corrected reflectance values of the five band RapidEye satellite imagery.

Middel et al. (2014) simulated the local thermal environment in the residential areas of Phoenix, Arizona, through the ENVI-met model. This model is a three-dimensional model that simulates surface-plant-air-interactions in urban environments. The model requires input data including air and soil temperature, relative humidity, soil moisture as well as wind speed and direction. It also requires information about vegetation, the surface characteristics and built structures that were collected in on-site measurements, digitization of data based on Bing Maps

and the North Desert Village (NDV) databases. This model was validated with meteorological data from the North Desert Village. The authors concluded that the “overall scale of the ENVI-met model and approach fits well with the concept of LCZs” and that “the LCZ classification scheme not only is a comprehensive framework for UHI research, but is also a useful concept for integrating local climate knowledge into urban planning and design practices” (Middel et al. 2014, 27).

Singapore has suffered fast landscape changes during the last 50 years. Ng (2015) studied the UHI map of this country, through LCZ classification. Three areas of study were used and monitored: one residential area, one area of parks/green spaces (CleanTech Park) and one commercial area (Central Business District), the Asia Square Tower 2. The author used photographs (using a Sony’s Nex3 camera with a 16mm f/2.8 wide-angle lens) for the extraction of the sky view factor property. The remaining properties, such as the aspect ratio, building surface fraction, impervious surface fraction, and pervious surface fraction were obtained through field surveys, the analysis of Google Earth, aerial photographs and LCLU maps. The terrain roughness class, surface admittance, surface albedo and anthropogenic heat, were obtained using the “LCZ data sheets as a guide” (Ng 2015, 122). At last, the LCZ map was created using guidelines developed by Stewart and Oke (2012).

Zheng et al. (2015) applied the LCZ classification to a city with a high population density, Hong Kong, and studied the UHI of that region. They used three datasets to classify the urban surface and quantify the relationships between urban morphology and climatic conditions at the local scale, namely, where one consisted of morphological data from the Planning Department, another was stationary observation data from automatic weather stations and the third included local meteorological data at a high spatial resolution captured by night-time traverse measurements across the study areas. The zonal average of the building coverage ratio and the building height were two properties calculated using a GIS methodology to classify the surface properties by density (compact/open) and height (high-rise/mid-rise/low-rise). The LCZ map was then generated.

In 2016, more studies appeared using the methodology proposed by the World Urban Database and Portal Tools (WUDAPT) project, which started in 2009, using remote sensing data and geographical information for the LCZ classification. This subject will be explored in detail in chapter 2.4.

Kaloustian and Bechtel (2016) applied the LCZ classification to the city of Beirut, Lebanon, located over the Mediterranean Sea. The classification was made using the WUDAPT methodology

as described in Bechtel, Foley, et al. (2015). A total of 251 training areas were collected using Google Earth and a template was provided by WUDAPT.

Danylo et al. (2016) studied two cities in Ukraine, namely Kyiv and Lviv, which differ in urban form and topography, and utilized three ways to validate and verify this classification. The classification was conducted according to the WUDAPT methodology, using Landsat 8 imagery for both cities. The classification of the study areas was done with images from different seasons; for Kyiv, four scenes representing spring (April and May, June) and autumn (October) were used while five scenes representing spring (May, June, March, April) and winter (March) were used for Lviv. However, the fifth scene that was downloaded for Kyiv resulted in linear artifacts (distortions on the image because of errors in the optical sensor of the satellite) in the LCZ map and was, therefore, omitted. A stratified random sample of 1125 pixels at the original resolution of 120 m was then selected from the city of Kyiv. This sample was used for independent validation of the LCZ map of Kyiv. The overall accuracy was 66%. Two different datasets were then used to make an independent comparison. The first comparison used the GlobeLand30 land cover dataset, at a resolution of 30 m, that has recently been (2010) developed by the National Geomatics Center of China. This dataset is divided into nine classes of landscape cover. The second dataset used was the POIs available in OSM, which were compared with the LCZ classification. The independent comparison with the GlobeLand30 land cover dataset in both cities showed an overall accuracy of 83% for Kyiv and 75% for Lviv. The comparison between LCZs and OSM data showed good correspondence between the POIs for the city, towns and villages and the LCZ classification.

Ren et al. (2016) studied two of the major cities in China, Wuhan and Hangzhou, using the methodology from the WUDAPT project. The authors also made an evaluation of the classification's accuracy, with a random sample corresponding to 0.5% of the number of pixels from each LCZ class. The assessment of the accuracy showed satisfactory results, namely an overall accuracy of 75.2% for Wuhan and 75.5% for Hangzhou.

## **2.4. World Urban Database and Access Portal Tools**

The World Urban Database and Access Portal Tools (shortly, WUDAPT) project was created to collect data about cities. It is an international collaborative project based on the National Urban Database and Portal Tool (NUDAPT), developed in 2009 (Ching et al. 2009). The NUDAPT project was designed as a resource of data for model development for the modeling community to facilitate addressing many of the evolving environmental problems of urban areas.

The NUDAPT characterized cities across the USA. However, in most regions of the world, this information is not available, since different methodologies for gathering the data must be developed (Feddemma, Mills, and Ching 2015).

According to Feddemma, Mills, and Ching (2015), over time, innumerable models of simulations of UHI and other urban climatology characteristics have been published, but a better correspondence of the parameters that represent the surface features is still lacking. In other words, the progress of urban climate science is limited as it is very dependent on the data that describes the form and functions of cities. The WUDAPT project was created to solve this issue. In 2012, this initiative started to collect and distribute urban information to provide parameters to be used in urban climate models using the same consistent methodology for all urban areas worldwide (Feddemma, Mills, and Ching 2015; Bechtel, Alexander, et al. 2015).

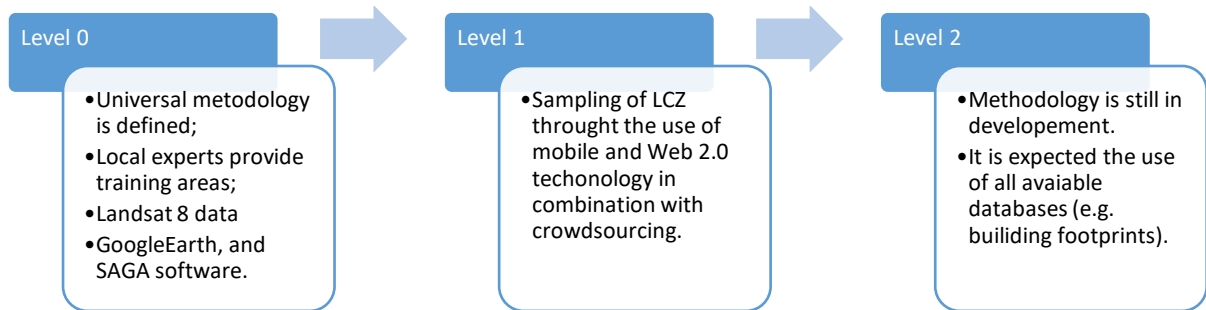
As mentioned previously, the data needed for climate research studies must report aspects related to urban form and function (UFF). Urban form depends on three aspects of cities: surface cover, proprieties of the construction materials and surface geometry. The urban form also describes the energy demands of the city, which may be represented by the anthropogenic heat flux (Mills et al. 2015). To capture these data, WUDAPT categorizes data acquisition into different levels, each of which represents a different level of detail (Figure 10).



**Figure 10: WUDAPT's data hierarchy based on Mills et al. (2015); See et al. (2015)**



The procedure for gathering Level 0 data has been established (and is further explained in section 3.1), but the methodology for collecting data at Levels 1 & 2 is currently being established. Nevertheless, it is planned to employ sampling schemas to Level 0 data for gathering data for levels 1 & 2; see Figure 11 (See et al. 2015).



**Figure 11: Hierarchical methodology to collect data for levels 1, 2 and 3 (based on Mills et. al. (2015) and See et. al. (2015))**

### **3. Methodology**

The methodology used in this thesis consists of: 1) using the procedure proposed by the WUDAPT project (level 0) to create an LCZ map for each study area using four different images, corresponding to four different times of the year. The detailed process used in this step is explained in section 3.1; 2) the data available in OSM is then converted to LCZ classes. This process is explained in section 3.2; 3) the outputs of both previous steps are integrated in order to solve inconsistencies obtained from the classification performed in the first step, where different LCZ classes are associated with the same pixel for the different satellite images used; and 4) the accuracy assessment of the results obtained in steps 1 and 3 is made, in order to determine if the data extracted from OSM improved the results obtained.

#### **3.1. The World Urban Database and Access Portal Tools**

The protocol to derive an LCZ map in level 0 of the WUDAPT project was established by Bechtel, Foley, et al. (2015), allowing local volunteers with varied levels of education to generate an LCZ map. The authors justified the procedure with several criteria that aim to achieve a method that is universal, quick and has no associated costs. In other words, it needed to be achievable, consist of a relatively fast procedure (a procedure that take less than 10 minutes on a standard computer) and without any economic costs (through the use of free software and data that needs to be available worldwide).

The resources needed to apply the level 0 procedure are: the local knowledge of the city, the study of the LCZ scheme available on Stewart and Oke (2012), the Google Earth and SAGA software and Landsat 8 scenes available on <http://earthexplorer.usgs.gov>. The protocol for gathering Level 0 data for each city can be divided into several steps (Figure 12), that can be summarized in the following list:

1. Digitize training data
2. Load satellite data
3. Import satellite data
4. Import vector data

5. Merge vector data
6. Project the vector data to the same coordinate system as the satellite images
7. Clip the satellite images with the ROI (Region of Interest)
8. Resample satellite images
9. Supervised Classification
10. Post classification filtering
11. Export the result to KML (Keyhole Markup Language)

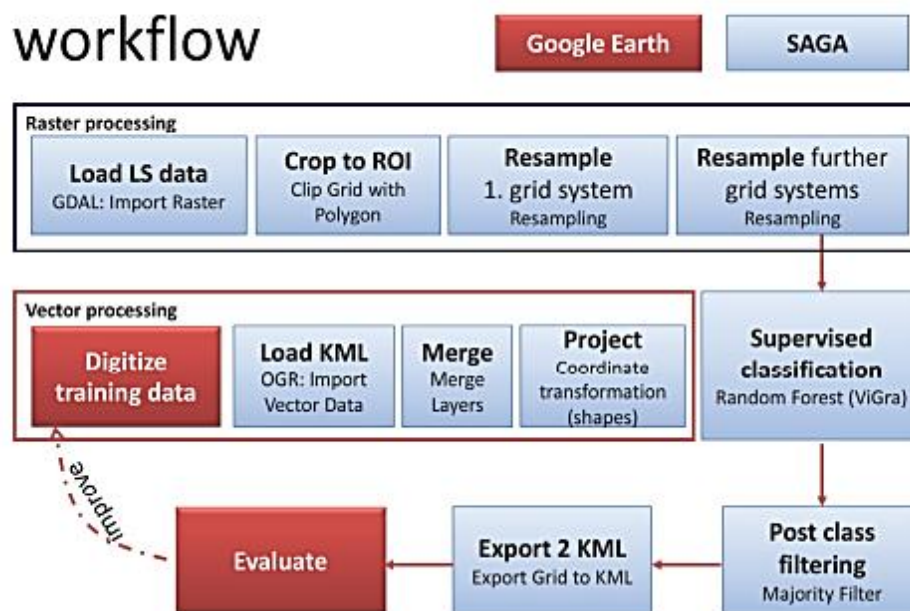
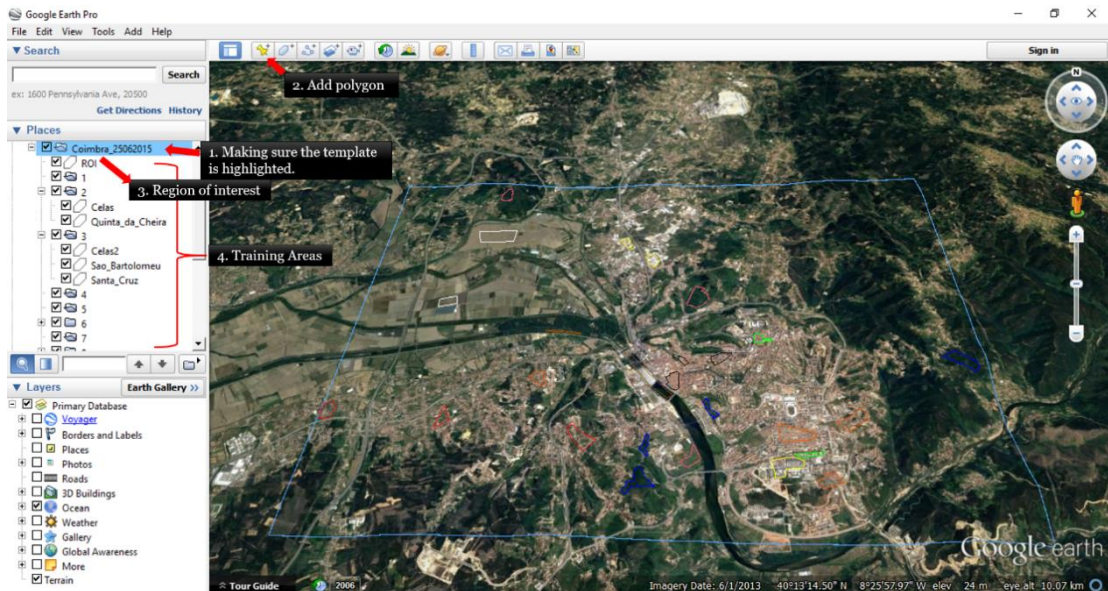


Figure 12: Workflow for the WUDAPT level 0 LCZ mapping procedure (*in* Bechtel et al. 2015)

The first step is a vector processing procedure, i.e. “Digitize training data”, where an individual with an interest and knowledge about a city (i.e. the expert) outlines an area that encloses the region of interest (ROI) and identifies parts of the natural and urban landscape that typify LCZ types. Thus, the expert creates training areas for each LCZ class present in the ROI. These training areas are created as polygons using Google Earth (Figure 13).

In the WUDAPT project website (<http://www.wudapt.org/create-lcz-training-areas/>), this procedure is explained in detail and a template for Google Earth is available, where each of the LCZ types is listed in folders. Classes 1 to 10 represent the urban landscape (LCZ 1 to LCZ 10) and classes 101 to 107 represent LCZ classes A to LCZ G (<https://drive.google.com/drive/folders/0B83nwq2eGktbUGxSV196a2tpYTA>).



**Figure 13: Step 1: “Digitize training areas”**

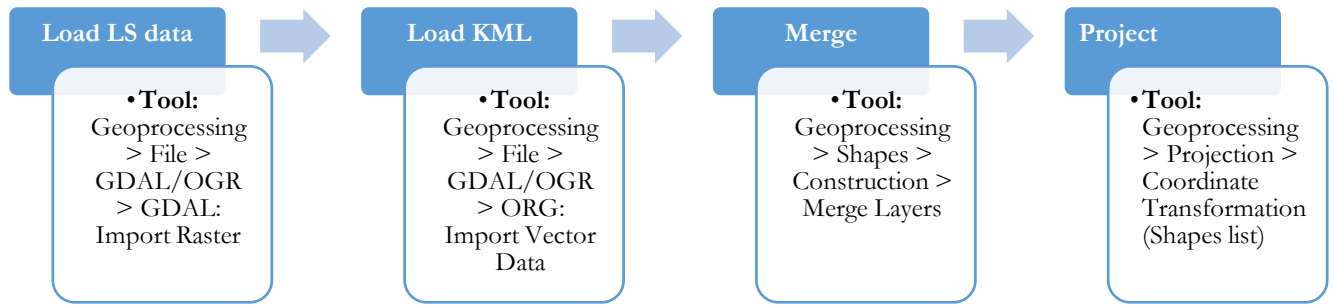
A polygon for the ROI is created around the urban area and also the training areas (that should be created within the template referred to previously).

The second step corresponds to downloading of Landsat 8 data (available on <http://earthexplorer.usgs.gov>).

The third step involves importing satellite data and is performed using the SAGA software (deselecting the “Transformation” option). The fourth step corresponds to importing the KML files created with Google Earth into SAGA, which contain both the ROI and the LCZ training areas. When importing the KML files, the option “Geometry Type” should be changed from “automatic” to “wkbPolygon”.

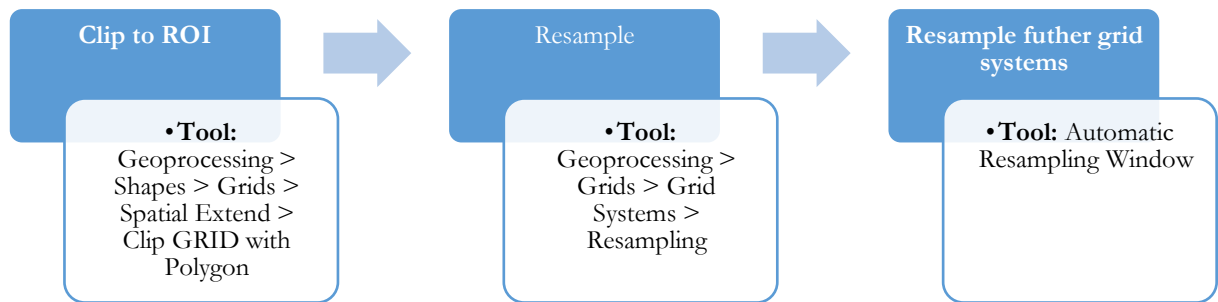
In the fifth step, the layers of the LCZ training areas are merged, so in the “Merge Layers” window, all of the LCZ training areas (not the ROI) were selected. It is still necessary to select the options, “Add Source information” and “Match Files by Name”, so that the attribute information of the layers is not lost.

In the sixth step, the coordinate transformation of the shape layers into the coordinate system of the satellite images is performed. In the “Coordinate Transformation (Shapes list)” window, the “Loaded Grid” was selected, and then the coordinate system information from the Landsat 8 image to geo-reference the KML vector file was imported. Then in the “Source” option, both the ROI and LCZ training area files were selected to geo-reference. At this stage (Figure 14), it is better to check the results, checking that the ROI and the training areas overlay correctly with the satellite images.



**Figure 14: Process and location of tools in SAGA software (Procedure 1)**

The seventh step (Figure 15) corresponds to the raster clip procedure, “Crop to ROI”, where the satellite images are clipped to the size of the ROI through the tool “Clip Grid with Polygon”. The grids (i.e. satellite images) should be selected to ‘clip’ and the ROI polygon imported within the same coordinate system.



**Figure 15: Process and location of tools in SAGA software (Procedure 2)**

The eighth step corresponds to the resampling of the satellite images, and consists of the integration of information contained in the Landsat 8 scenes, since as referred to previously, there are different spatial resolutions. Hence, it is necessary to resample these bands to the same grid. At the beginning of this study, the resolution advised to resample all Landsat scenes to a grid was 120 meters’ resolution, but in November 2016, the spatial resolution advised was altered to 100 meters ([http://www.wudapt.org/prepfeat\\_overview/path2step3c/](http://www.wudapt.org/prepfeat_overview/path2step3c/)). Thus, in this step, all Landsat scenes (B1 to B11) are resampled to a grid of 120 m resolution. The interpolation method that should be selected is “Mean Value (cell area weighted)”. All Landsat scenes were resampled to a grid of 30 m resolution to test if a greater spatial resolution produced a more accurate result.

The ninth step uses the LCZ training areas to classify the images in the ROI (Figure 16) into neighborhood types using the random forest classification scheme (Mills et. al., 2015). The resulting raster image, based on a per-pixel automatic classification, can create an outcome with pixels with isolated LCZs. Thus, the tenth step consists in the removal of this single pixels of one LCZ using a post classification filter, namely, a majority filter. At the beginning of this study, the WUADPT project advised a post classification majority filter with a 200 to 300 m radius. Thus, for

a grid of 120 m resolution, a majority filter considers 25 neighboring cells and for a grid of 30 m resolution, a majority filter considers 64 neighboring cells

In the eleventh step, the classification process was exported as KML files for integration into the Google Earth software. The lookup table with the same legend colors as other cities on the WUDAPT project website (Figure 17) was used; the coloring was set to “same as in graphical user interface” and the interpolation option was deselected.

At the beginning of this study, the color scheme for each LCZ class was not available but since November 2016, it can be download from the WUDAPT project website.

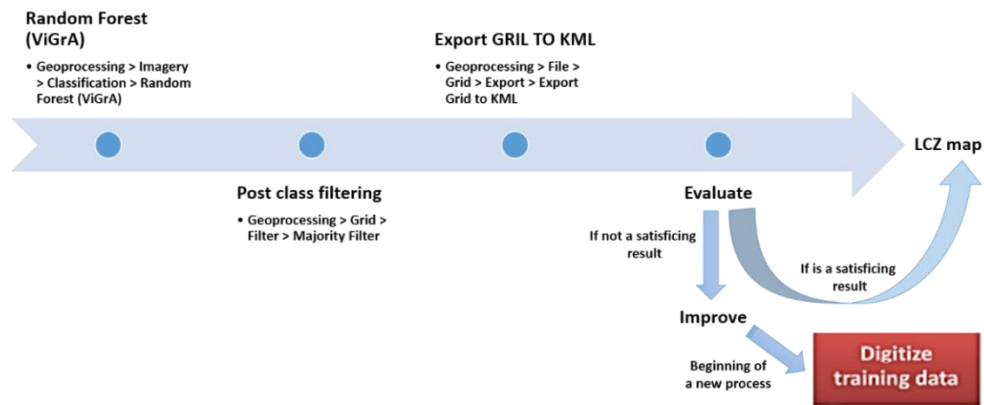


Figure 16: Process and location of tools in SAGA software (Procedure 3)

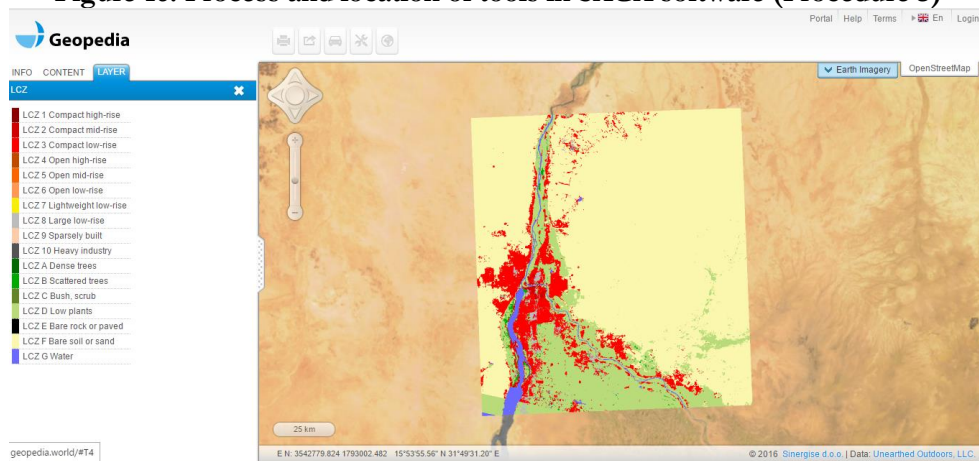


Figure 17: Colouring of WUDAPT project

Finally, in the last step, the expert scrutinizes the map and adds to (or adjusts) the LCZ training areas to account for misclassifications and repeats the process until satisfied (Figure 16).

In this stage, the expert with the knowledge of the city sets the transparency of the classification product image overlay to approximately 50% in Google Earth and examines the results to evaluate whether the process has generated an acceptable LCZ classification for the ROI or not. If not, the expert can readjust or improve it by adding other training areas until reaching a satisfactory result.

## 3.2. Conversion of OSM data into LCZs

In this project, the methodology of converting OSM features into an LCZ map was done via a sequence of steps that can be divided in two phases: data pre-processing and data processing.

### 3.2.1. Pre-processing phase

The data pre-processing phase includes: the download of OSM data, identification of the key/value combinations that can be used to identify LCZ classes, transformation of reference systems, and creating a grid corresponding to the limits of the pixels obtained with the application of the WUDAPT methodology, so that the outputs of the conversion of OSM data into the LCZ classes can be compared with the results of the satellite image classification. The procedures used to create the features in the pre-processing phase are shown in Appendix A and B.

The OSM data were downloaded from the Geofabrik portal (<http://www.geofabrik.de>) in shapefile format. To associate the OSM data with the LCZ classes, a previous analysis was done to associate the key/value combinations and their descriptions established by the OSM community and available online ([http://wiki.openstreetmap.org/wiki/Map\\_Features](http://wiki.openstreetmap.org/wiki/Map_Features)) with the most relevant LCZ classes. Table 5 shows the associations considered in this research. It should be noted that, in some cases, the data available do not allow differentiation between some classes. For example, in the OSM features that correspond to regions with trees, it is not possible from the data available in OSM to differentiate between dense or scattered trees. Therefore, in this case the conversion process just identifies that those regions that may be associated with class A or B.

As volunteers can create new values for the available keys, the analysis made previously is insufficient. Thus, an additional analysis must always be done for each study area because different key/value combinations may be available. Moreover, even for the same region, the OSM data may differ if obtained at different dates, since the information is continuously added and edited by the citizens. Therefore, for each study area, additional key/value combinations available in OSM data were identified and their correspondence to the LCZ classes of interest (LCZ 1 to LCZ 10 and LCZ A to LCZ G) was made.

The next step in the pre-processing phase is the transformation of the coordinate system of the LCZ maps obtained with the application of the WUDAPT methodology to the same coordinate system used for the processing of the OSM data, so that the data obtained from both approaches can then be compared and integrated. This requires the creation of a GRID in vector format, in such a way that each cell corresponds to the LCZ map resulting from the random forest classification.



**Table 5: OSM keys and values**

LCZ	Keys	Keys values
LCZ A or B	natural	Wood, trees, grass, tree_row, forest
	landuse	Forest, nature_reserve
LCZ C	natural	Scrub
	landuse	Heath, orchard, scrubs, vineyard, scrub, scrubs, plant_nursery
LCZ D	natural	Grass
	landuse	Farm, farmland, farmyard, meadow, greenfield, grass
LCZ G	natural	water
	waterways	river, stream, canal, drain, brook, ditch, riverbank
LCZ 1 to 10	roads	Bus_guideway, living_street, primary, primary_link, residential, raceway, road, secondary link, tertiary, tertiary link, trunk, trunk, trunk link
	railways	funicular, miniature, monorail, light_rail, narrow gauge, rail, tram, transfer table, mainline
	building	apartments, hotel, house, detached, residential, dormitory, terrace, houseboat, static_caravan, commercial, industrial, retail, warehouse, bakehouse, cathedral, chapel, church, mosque, temple, synagogue, shrine, civic, hospital, school, stadium, train_station, transportation, university, barn, public, bridge, bunker, cabin, ruins, construction, farm_auxiliary, garage, garages, carport, hangar, roof, shed, stable, transformer tower, kiosk

### 3.2.2. Processing phase

In the processing phase, the conversion of the OSM data into the LCZ classes is made. This phase consists of several steps, as different features available in OSM require different types of processing. Some OSM features have a direct association to LCZ classes, namely to the land cover classes such as LCZ A or B (dense trees and scattered trees), LCZ C (Bush or scrub) or LCZ D (low plants). The processing of these features requires the selection of the OSM polygons with a particular combination of key/value that satisfies the defined correspondences to the LCZ classes. Once all polygons corresponding to the LCZ classes are identified, they are merged into a single feature. Then an intersection of the resulting features with the GRID corresponding to the satellite images classification is done, and the area occupied by the specified LCZ in each cell of the GRID is computed.

Figures 18, 19 and 20 show the procedures applied to identify LCZ A or B, LCZ C and LCZ D, respectively. The procedures used to create the features corresponding to LCZ classes, namely, LCZ A or B, LCZ C and LCZ D, are shown in Appendix C, D and E, respectively.



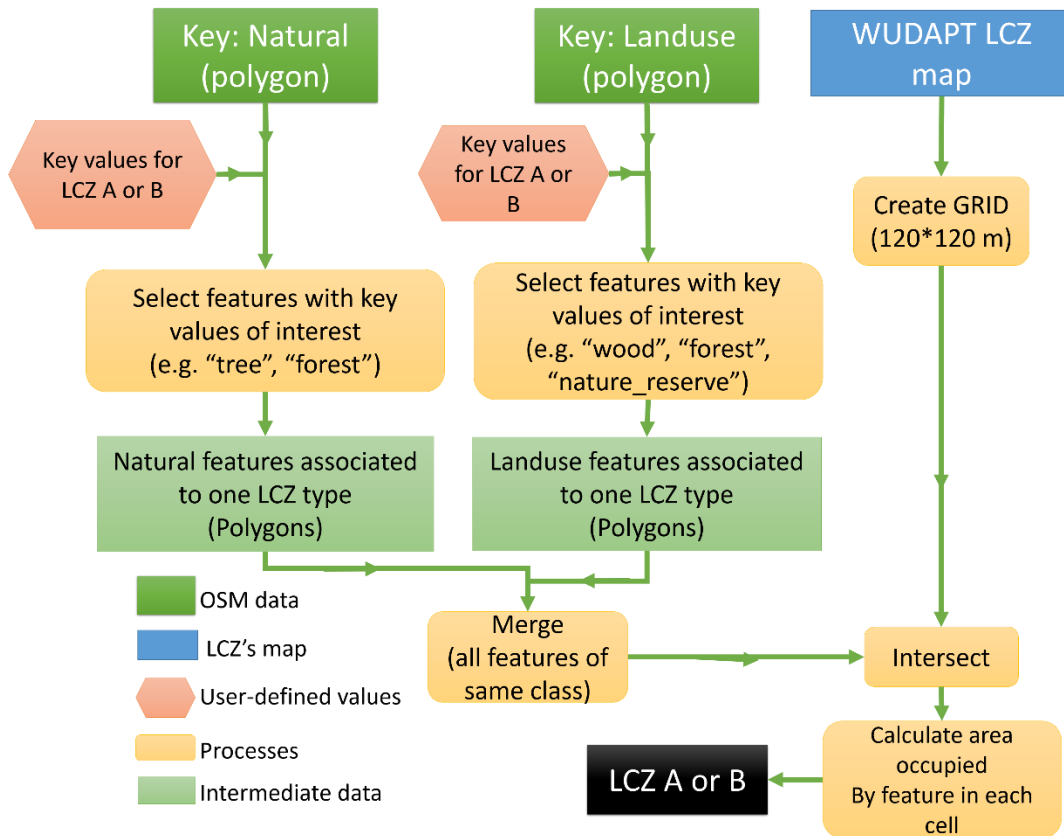


Figure 18: Scheme of the methodology used for the conversion of OSM data into LCZ classes A (Dense trees) or B (Scattered trees).

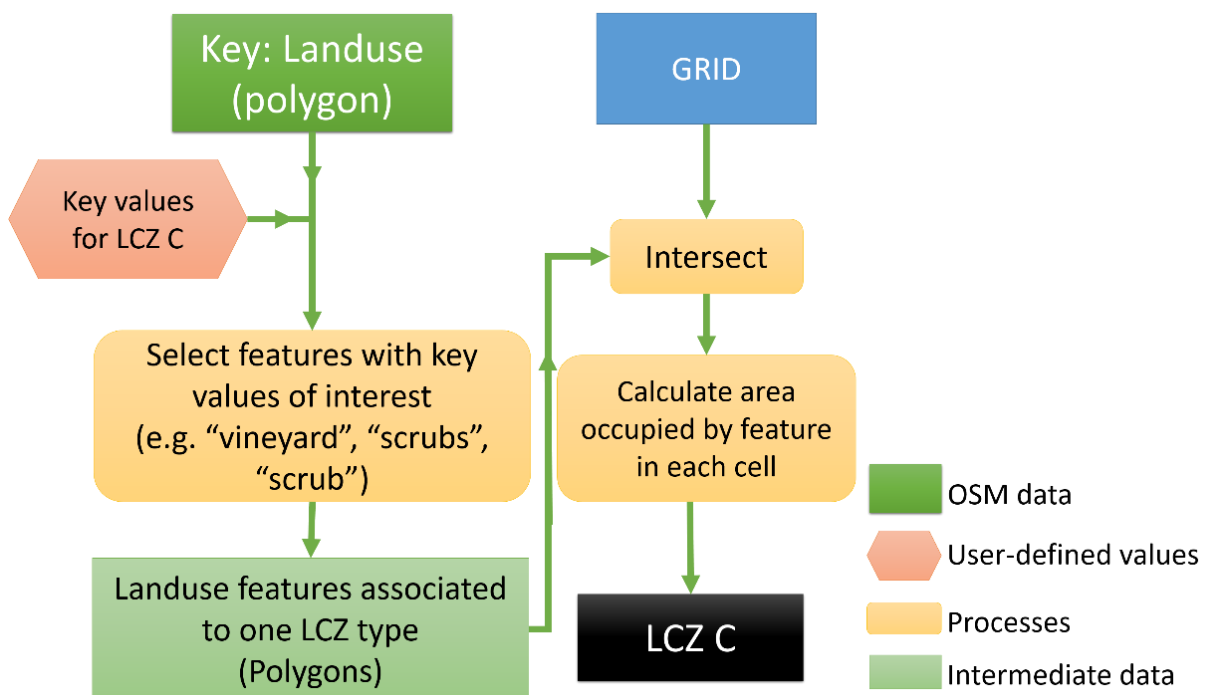
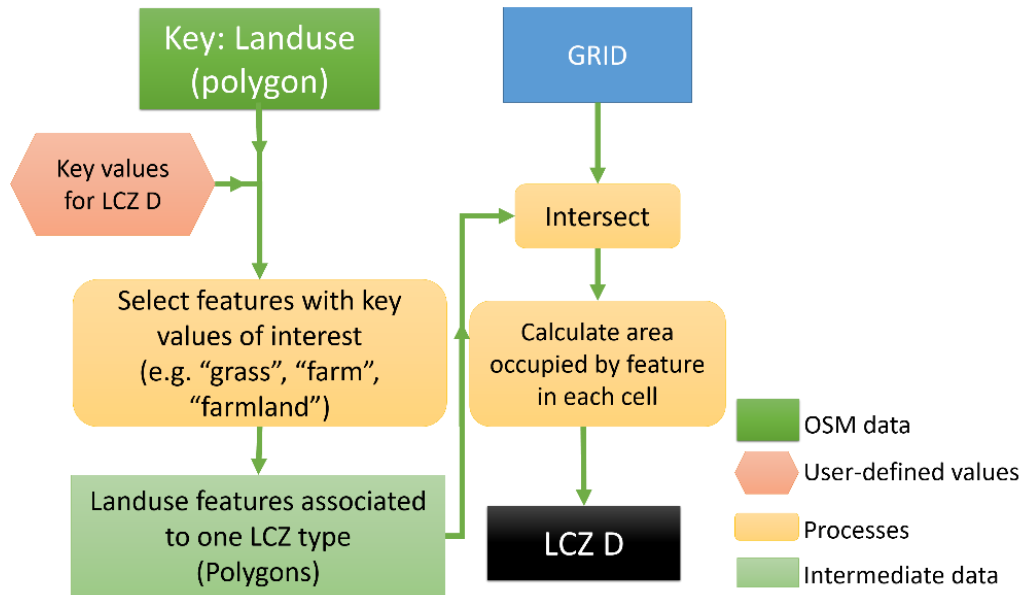


Figure 19: Scheme of the methodology used for the conversion of OSM data into LCZ class C (Bush, scrub)



**Figure 20: Schematic of the methodology used for the conversion of OSM data into LCZ class D (Low plants)**

Some of the features in OSM are represented by linear elements, such as roads, railways and waterways. To assign these features to classes in an LCZ map, they need to be converted into areas. This process was done clipping these features with the study area and creating a buffer around them. To define the width of the buffers to use, so that the resulting areas do not overlap other types of features existing in the vicinity, in particular the building along the streets, the lines are separated into segments, the distance of each segment to the OSM buildings is computed, and the obtained result is assigned to each segment as an additional attribute (distance to buildings). The process used to create the buffers is illustrated in Figure 21, where “d” represents the obtained distance to the buildings and “t” represents a value predefined for each type of feature, which is used when distance “d” is too large to be considered. That is, if the distance of the feature to the buildings (d) is greater than the user-defined threshold (t), this means that there are no buildings near the feature. Thus, the user-defined threshold is used to create the buffer. If the distance to buildings is inferior to this threshold, then that value is used to define the buffer. The threshold values need to be chosen according to the characteristics of the regions under analysis, as the different types of features may have very different typical widths in different cities and different parts of the world. The procedures used to create the various features are shown in Appendix G.

For LCZ G (Water), there are usually both linear features representing waterways but also polygonal features corresponding to the regions with water (mainly in the features with the key “natural”) in OSM. Hence, it is necessary to convert these linear features to areas, and then merge

them into polygonal features that also represent water. Figure 22 illustrates the procedure used to extract the water layer from the OSM data, which is shown in Appendix F.

As with the other classes, the merged results are then combined with the outputs obtained with the GRID previously created at a spatial resolution of 120 meters. An intersection of the GRID with the data obtained is then made and the area occupied by each feature per cell was computed.

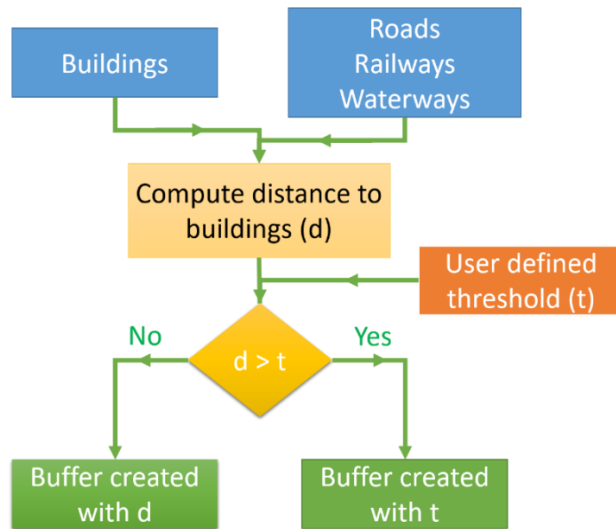


Figure 21: Procedure for converting linear features into polygons

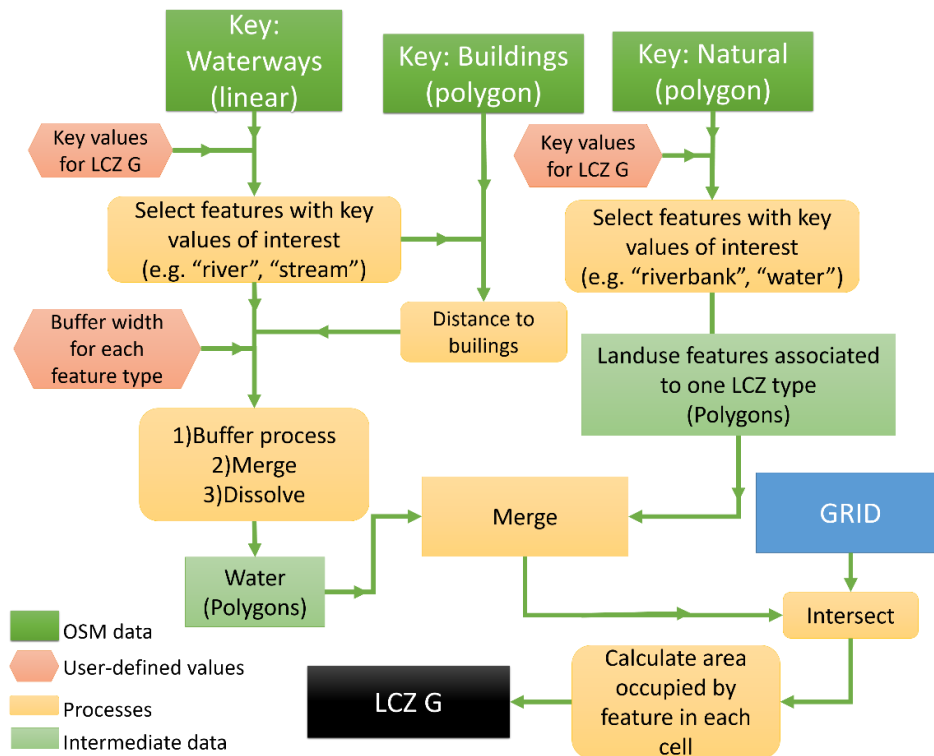
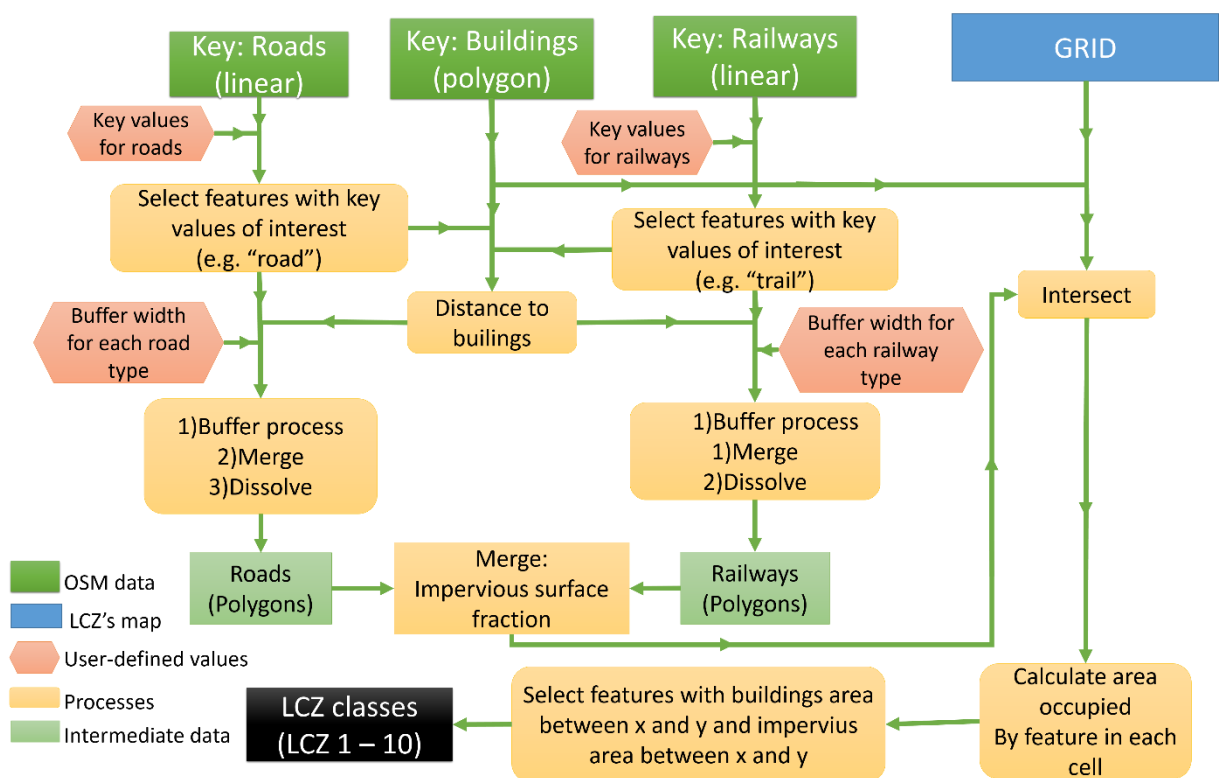


Figure 22: Schematic of the methodology used for classes LCZ G (Water)

For the LCZ urban classes (LCZ 1 to LCZ 10), the data available in OSM can provide information on the presence of buildings and impervious surfaces, such as roads or railways. To associate these data to the LCZ classes, the values established by Stewart and Oke (2012) regarding the building surface fraction and the impervious surface were considered, as well as the outputs of the described procedure, which was enable to the conversion of the polyline features (railways and roads) into polygons. Figure 23 illustrates the procedure used to extract this urban classes layers from the OSM data. The procedures used to create the various features are provided in Appendix H and I.



**Figure 23: Schematic of the methodology used for classes LCZ 1 to 10**

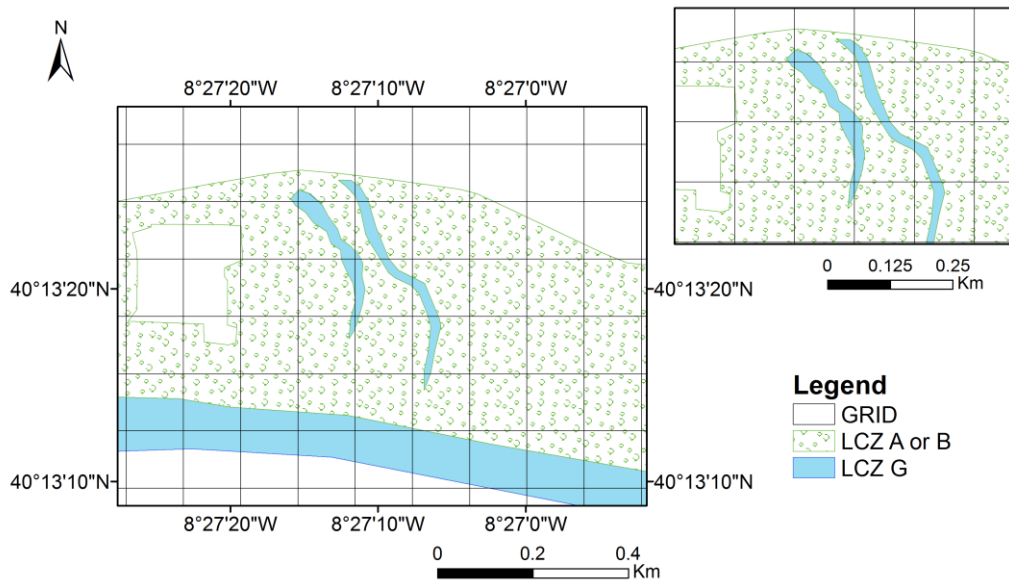
The data resulting from the conversion of the roads and railway linear features to areas was aggregated, where these resulting features correspond to the impervious regions. The buildings existing in OSM were extracted and the layer corresponding to the impervious regions and the buildings were intersected with the GRID corresponding to the cells in the raster file resulting from the classification of the satellite images. Then, the percentage of each cell occupied by the impervious layer and building was computed, generating a layer with the building surface fraction and another with the impervious surface fraction.

Once the information about the building surface fraction and impervious surface area in each cell of the GRID is available, the cells were assigned to LCZ urban classes according to the percentage of area covered by building and impervious surfaces. Table 6 shows the values defined by Stewart and Oke (2012) that were used in this assignment.

With this procedure, the area occupied by each LCZ class in each cell was obtained. This information is important since a cell may be occupied by more than one class with a different percentage of occupation (see Figure 24). Table 7 shows an example of the attribute table that is associated with the GRID features (in vector format), where the existence (or not) of each LCZ class is identified and the percentage of the cell covered by that class is also stored.

**Table 6: Percentage of building and impervious surface fraction for classes LCZ 1 to 10 (in Stewart and Oke 2012)**

LCZ	Building surface fraction (%)	Impervious surface fraction (%)
LCZ 1	40 – 60	40 – 60
LCZ 2	40 – 70	30 – 50
LCZ 3	40 – 70	20 – 50
LCZ 4	20 – 40	30 – 40
LCZ 5	20 – 40	30 – 50
LCZ 6	20 – 40	20 – 50
LCZ 7	60 – 90	< 20
LCZ 8	30 – 50	40 – 50
LCZ 9	10 – 20	> 20
LCZ 10	20 – 30	20 – 40



**Figure 24: Cells with different LCZ class**

**Table 7: Example of part of the attribute values associated with the GRID**

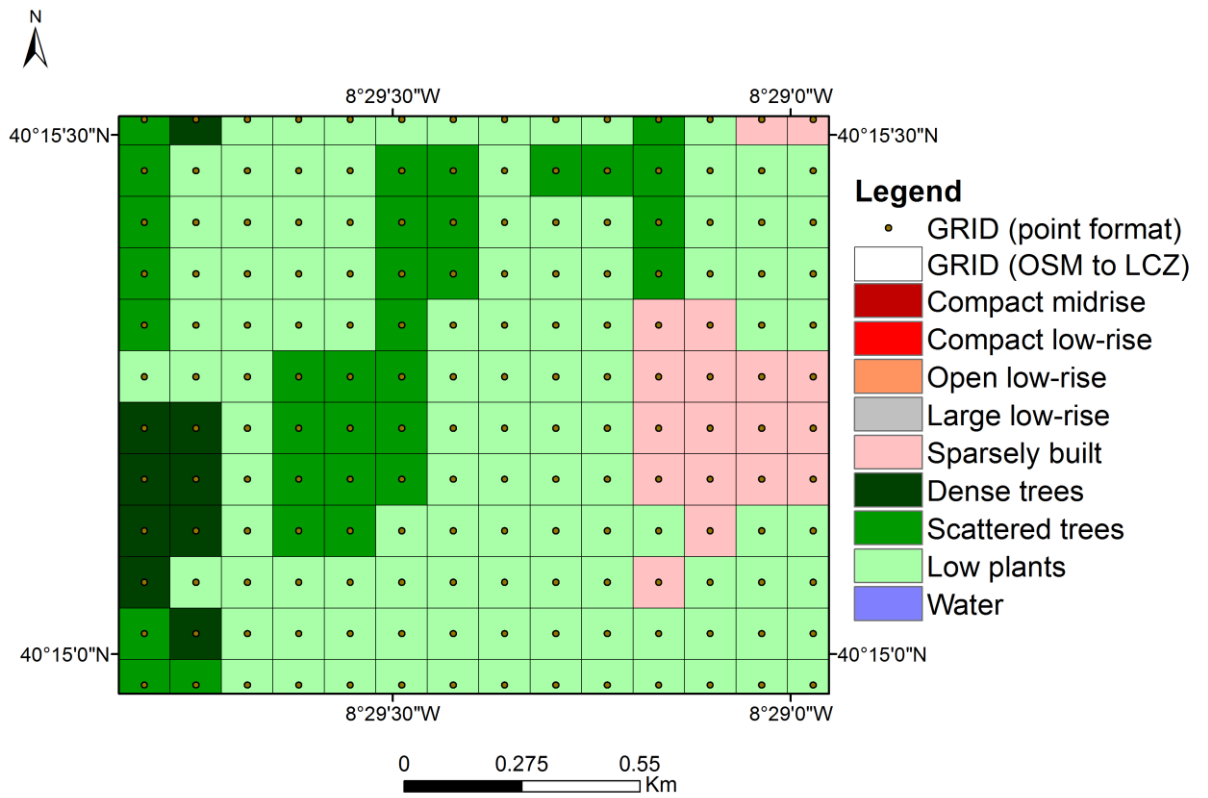
FID GRID	LCZ A or B	Area LCZ A or B (%)	LCZ G	Area LCZ G (%)	Buildings	Area Buildings (%)	LCZ 1	Area LCZ 1 (%)
1	1	79	1	21	<Null>	<Null>	<Null >	<Null>
2	1	20	<Null >	<Null>	1	50	1	60
3	1	100	<Null >	<Null>	<Null>	<Null>	<Null >	<Null>
4	<Null >	<Null>	1	100	<Null>	<Null>	<Null >	<Null>
5	<Null >	<Null>	1	50	<Null>	40	1	50
6	1	100	<Null >	<Null>	<Null>	<Null>	<Null >	<Null>

### 3.3. Integration and combination of data

#### 3.3.1. Integration of the outputs of the image classification and OSM data conversion

In the first step of the methodology, the classification of four Landsat 8 satellite images corresponding to different seasons (spring, summer, autumn and winter) is made using the methodology proposed by the WUDAPT project. As some physical characteristics of the territory, such as vegetation, change with the seasons, which translates into different spectral responses for the same area at different times of the year, the classification of the four images produces different results for some locations, resulting in inconsistent data. All the data are now combined at the pixel level. For this combination, a feature of points was created (using ArcGIS tool 'Feature to Point'), extracting the centroid of each pixel (see Figure 25). The value of the classes in the winter, summer, autumn and spring classified LCZ raster maps were then extracted and added as additional attributes to the table associated to the points (using ArcGIS tool 'Extract multivalues to points').

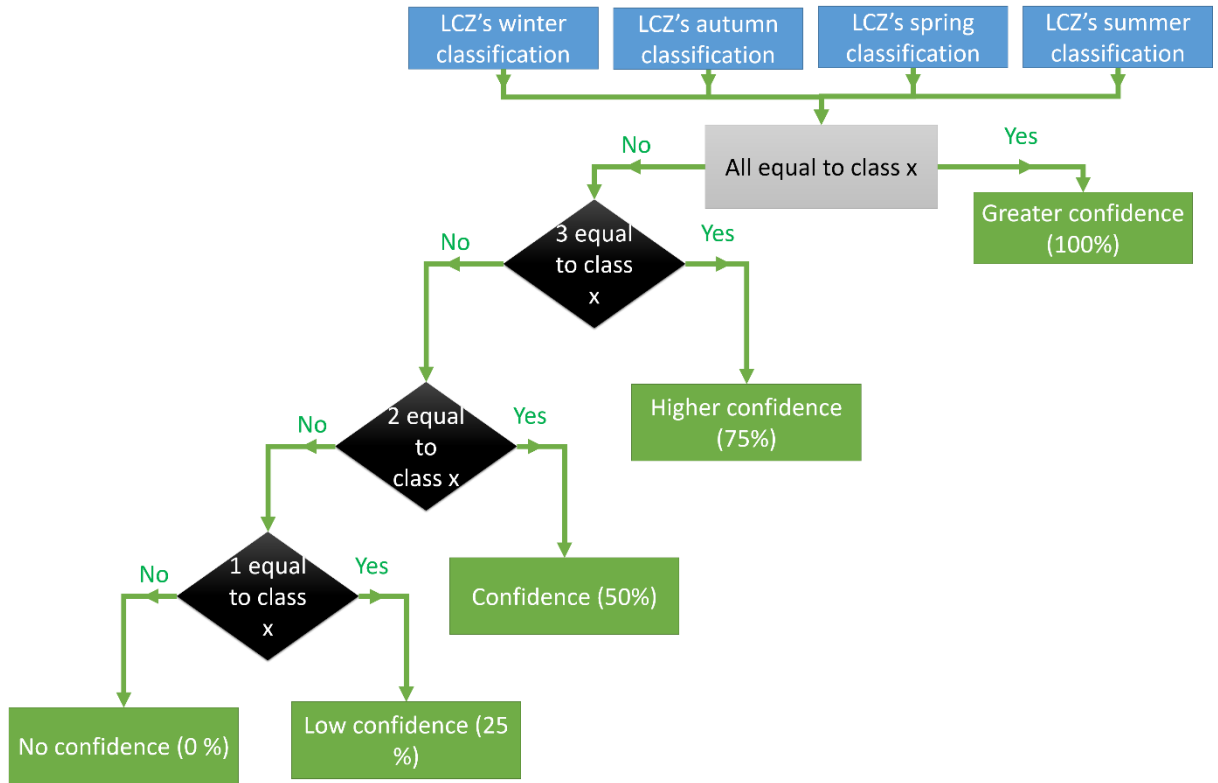
The created point features were associated with the GRID used in the conversion of the OSM data into LCZ classes through a spatial join. This aggregates all rows from the point feature with the polygons of the GRID (target features) considering their relative spatial positions (in this case inclusion). The procedures used to create the various features are shown in Appendix J.



**Figure 25: Representation of the created point feature and the vector GRID, overlaid with one map resulting from the classification of a Landsat 8 image into an LCZ map.**

### 3.3.2. Combination of the data

To combine the results obtained during the satellite image classification, and keeping the information about the reliability of the result, a decision tree was used, where a degree of confidence is associated to each cell, depending on the outputs of the classification of the four images used. Figure 26 illustrates the procedure, which is used for each pixel, generating for each class a degree of confidence. The procedures used to create the various features in this combination-of-data phase are shown in Appendix J as well as the code used to generate a degree of confidence for each class, which is provided in Appendix L.



**Figure 26: Example of the extraction of the percentage of the representative of this class for each pixel and for a particular class (x)**

A similar procedure was run for the data extracted from OSM, extracting the dominant LCZ class obtained for each cell of the GRID, which is the class with the greatest percentage of area occupied in the cell and also the second most representative LCZ class.

The last step in the combination of the results consists of assigning a class to each cell based on the data provided by the satellite image classification and the data extracted from OSM. Thus, once the information regarding the percentage of the occupation identified in OSM (see code in Appendix M) and the confidence in the assignment of a class based on the four LCZ maps resulting from the WUDAPT procedure is available for each cell, it is then necessary to aggregate all this information and evaluate the possibility associated to the assignment of each class to that cell. This was done using Eq. 1, where  $a(x)$  is the confidence rating associated with class  $x$ , and  $b(x)$  is the OSM area occupied by class  $x$  in the cell. The most likely class to assign to each cell is the class corresponding to the higher value of possibility.

$$\frac{a(x) + b(x)}{200} * 100 \quad (1)$$



### **3.4. Accuracy Assessment**

To assess the accuracy of all the maps, a reference database was created for all study areas considered. A stratified sample was used, selecting 200 points per class, and the strata were the classes obtained with the classification of the winter images. The reference data were created through photo-interpretation of the images available in Google Earth as a base map. Confusion matrices were created and accuracy indices were computed, namely the user's, producer's and overall accuracy.

## 4. Case studies

### 4.1. Study areas and data

Two study areas were used to test the proposed methodology. One includes the city of Coimbra, Portugal, while the other is the city of Hamburg in Germany. These two regions were chosen because their LCZ maps were already created by the WUDAPT project and they have very different data coverage in OSM.

#### 4.1.1. Coimbra

The first study area was the city of Coimbra, located in the center of continental Portugal. It is located on the Atlantic coast of the Iberian Peninsula, approximately 40 km from the western coast and has an area of 319.4 km<sup>2</sup> (Direção-Geral do Território 2016) and a population of 105 842 inhabitants (Instituto Nacional de Estatística 2013). This study area was chosen due to the possibility of using local knowledge and because of its relatively low-average cover of OSM data (12-13%) (Fonte et al. 2016), which allows for a better understanding of the usefulness of this procedure, even in cities where the OSM data are not abundant. Figure 27 shows the study area and the available OSM data while Figure 28 shows the LCZ map created by the WUADPT project for this area.

For the creation of LCZ maps using the methodology proposed in WUDAPT, four Landsat 8 images were chosen corresponding to spring, summer, winter and autumn. Table 8 shows the dates of the images used.

**Table 8: Dates of satellite images utilized for the LCZ classification for Coimbra**

	Spring	Summer	Autumn	Winter
Coimbra	23/04/2015	09/07/2015	29/10/2014	01/01/2015

Figure 29 and 30 show, respectively, the true color Landsat 8 images for Coimbra and a composite RGB = 543, to enhance the presence of vegetation in the study areas (highlighted in red).

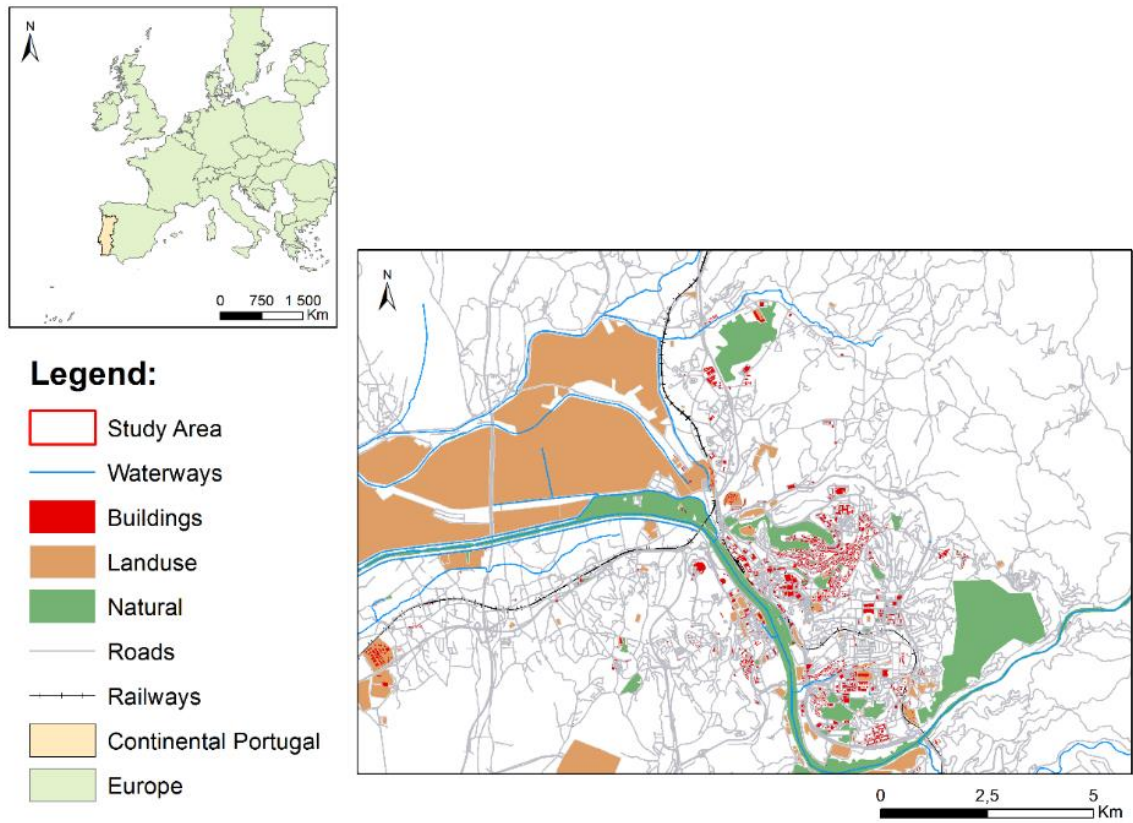


Figure 27: Study area and the available OSM data

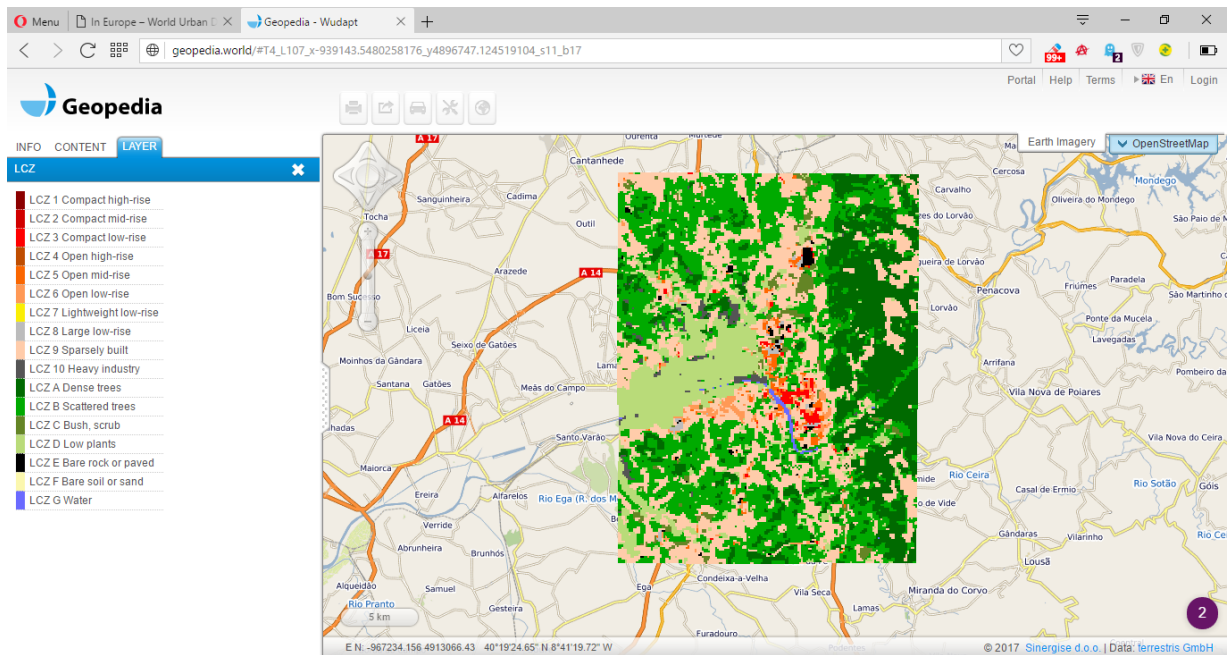


Figure 28: LCZ map created by the WUDAPT project for the region of Coimbra



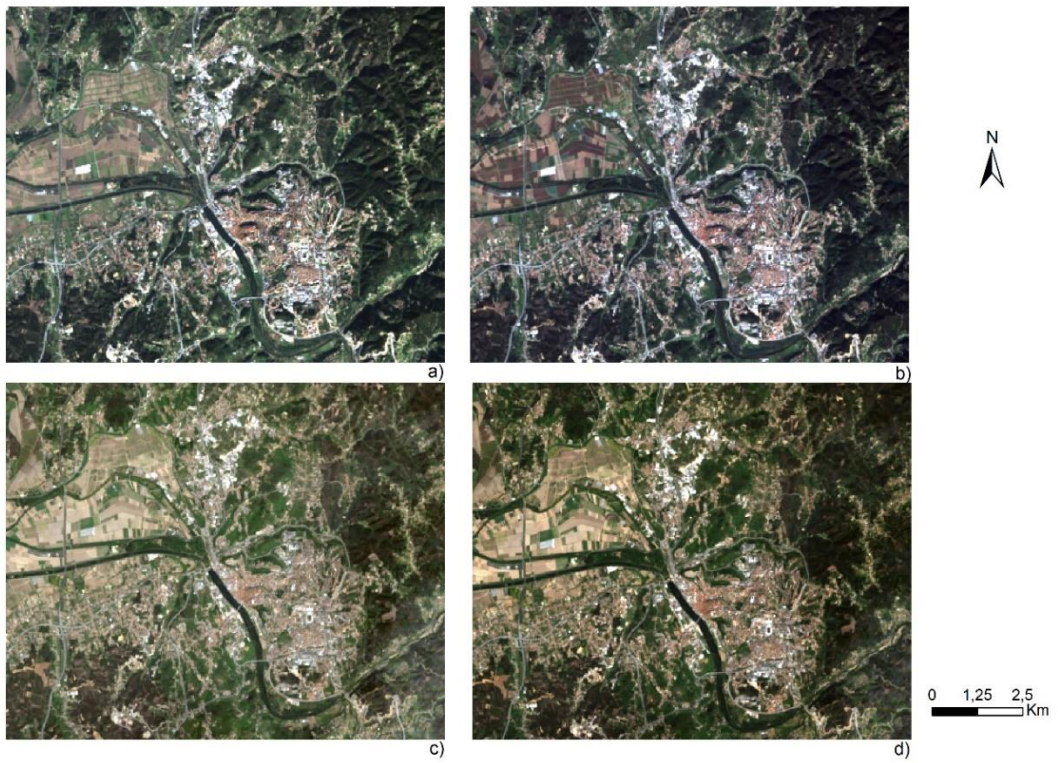


Figure 29: Landsat 8 true color imagery of Coimbra in a) winter, b) autumn, c) summer and d) spring

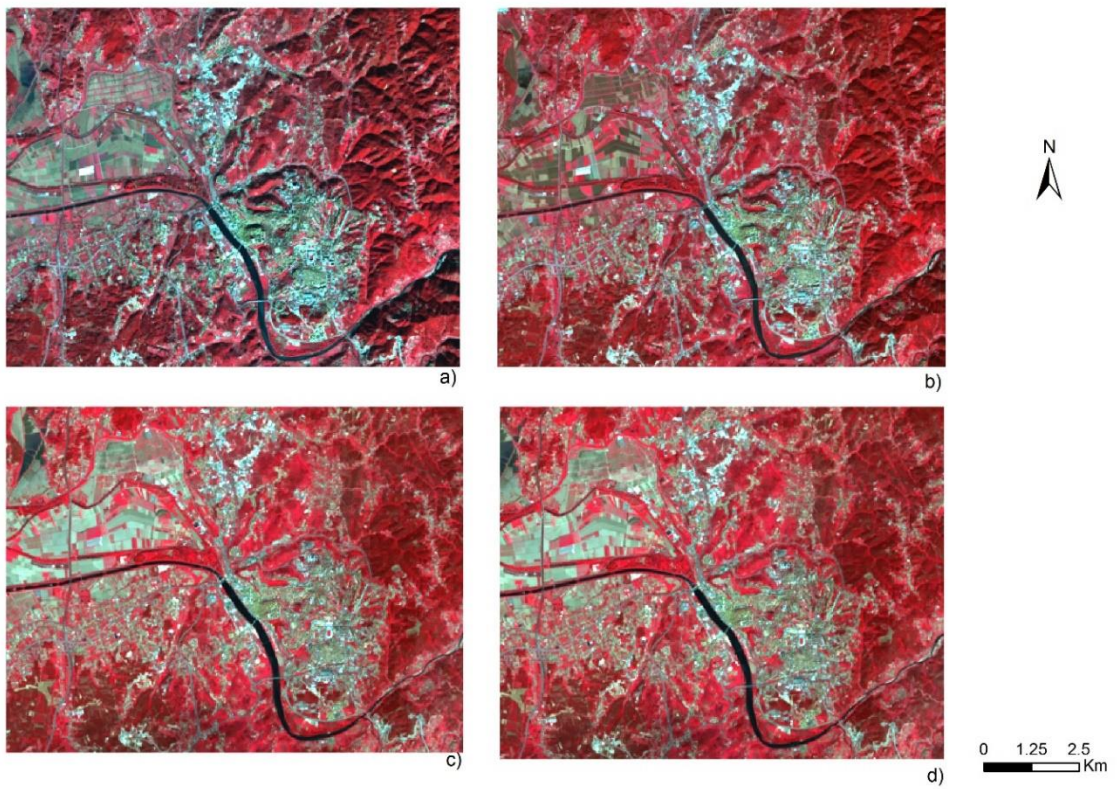


Figure 30: False color composite RGB = 543 of the Landsat 8 imagery of Coimbra in a) winter, b) autumn, c) summer and d) spring

#### 4.1.2. Hamburg

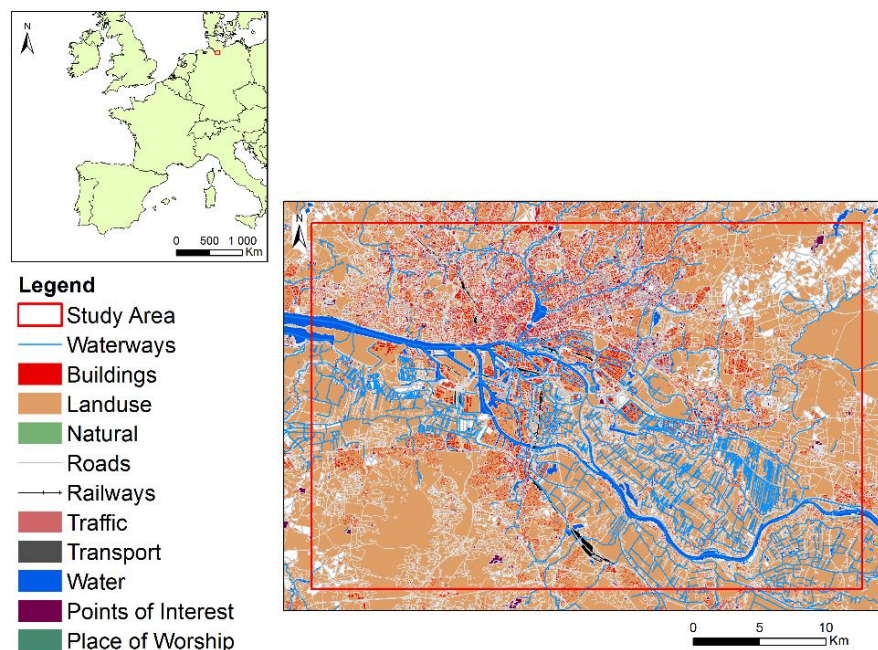
The second study area was the city of Hamburg located on the North of the German Plain in the lower slopes of Elbe River and 100 km upstream of the Elbe estuary. This is the second largest city in Germany, after Berlin, with 755.3 km<sup>2</sup> and a population of 1 814 597 inhabitants (Statistical Office of Hamburg and Schleswig-Holstein 2014; Rose and Wilke 2015). This city contains a major seaport in Germany, which makes the city a center of exchange, transport and services and it also has one of the most important industrial areas in all Germany (Rose and Wilke 2015). This study area presents a high-average cover of OSM data (see Figure 31), in contrast to Coimbra. Figure 32 shows the LCZ map created by the WUDAPT project.

As for the first study area, for the creation of LCZ maps for the region of Hamburg using the methodology proposed in WUDAPT, four Landsat 8 images were also used, corresponding to the spring, summer, winter and autumn. Table 9 shows the dates of the images used.

**Table 9: Dates of satellite images utilized for LCZ classification for Hamburg**

	Spring	Summer	Autumn	Winter
Hamburg	30/04/2016	21/08/2015	06/12/2016	03/01/2017

Figures 33 and 34 show, respectively, the true color Landsat 8 images for Hamburg and a composite RGB = 543.



**Figure 31: Study area and the available OSM data**



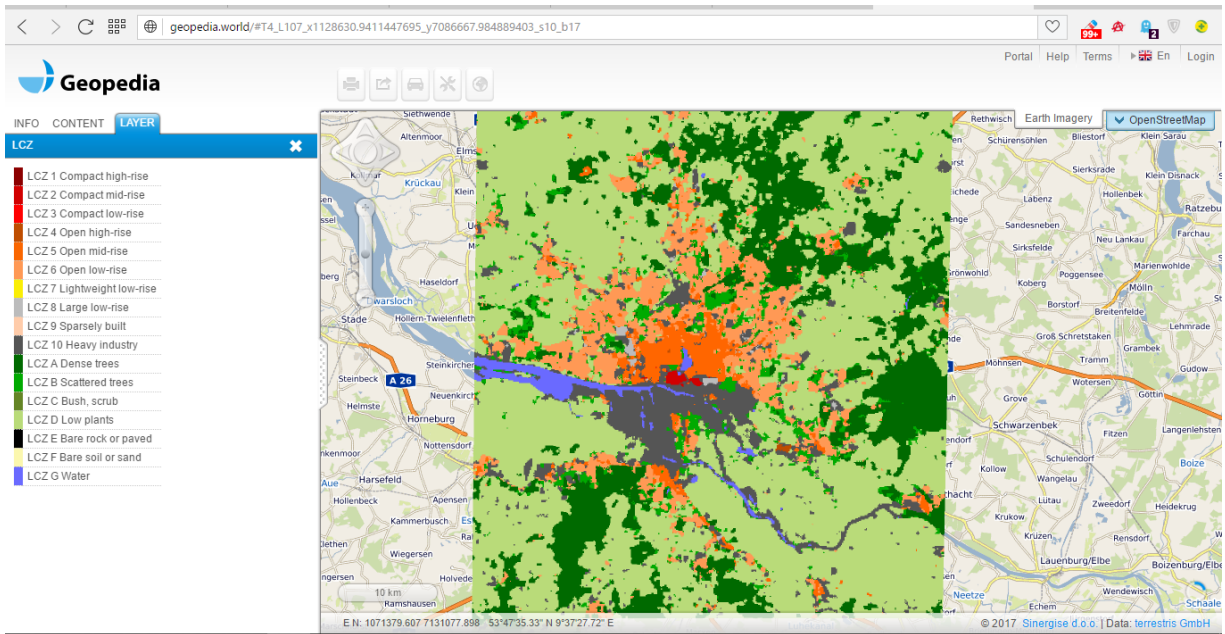


Figure 32: LCZ map for Hamburg created by the WUDAPT project

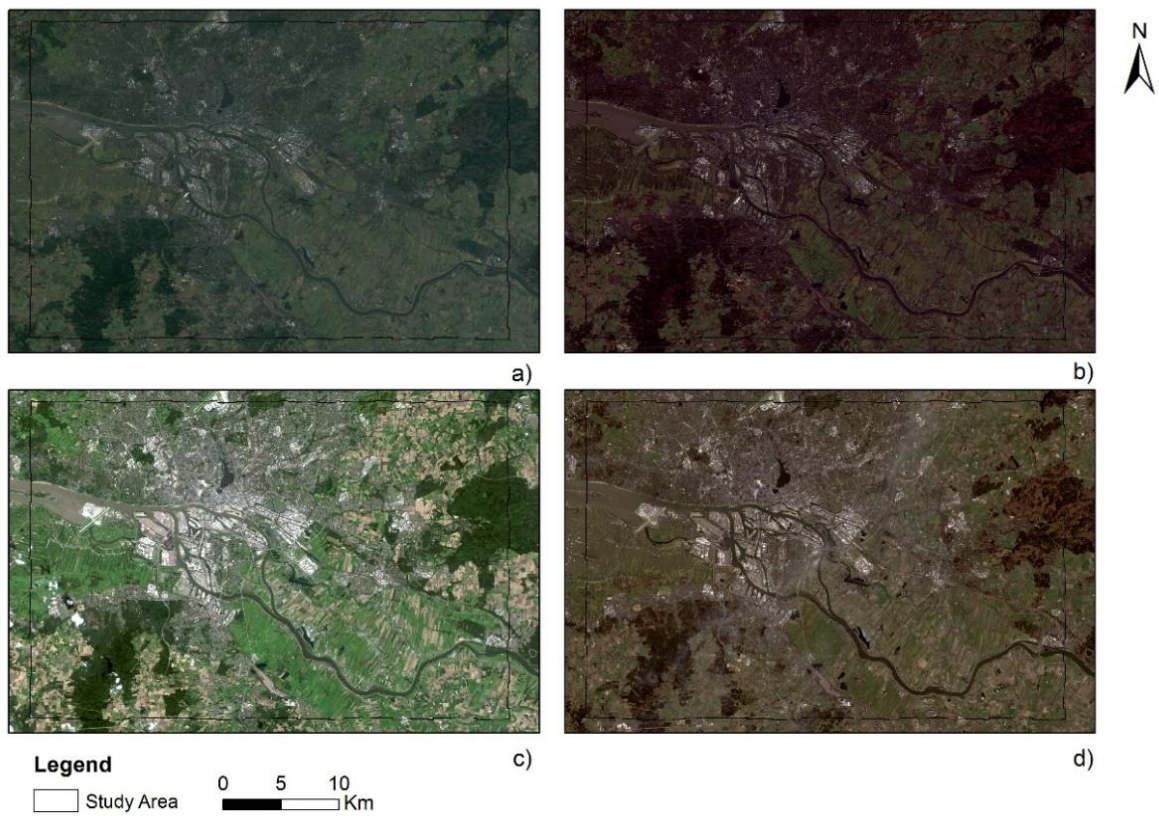


Figure 33: Landsat 8 true colors imagery of Hamburg in a) winter, b) autumn, c) summer and d) spring

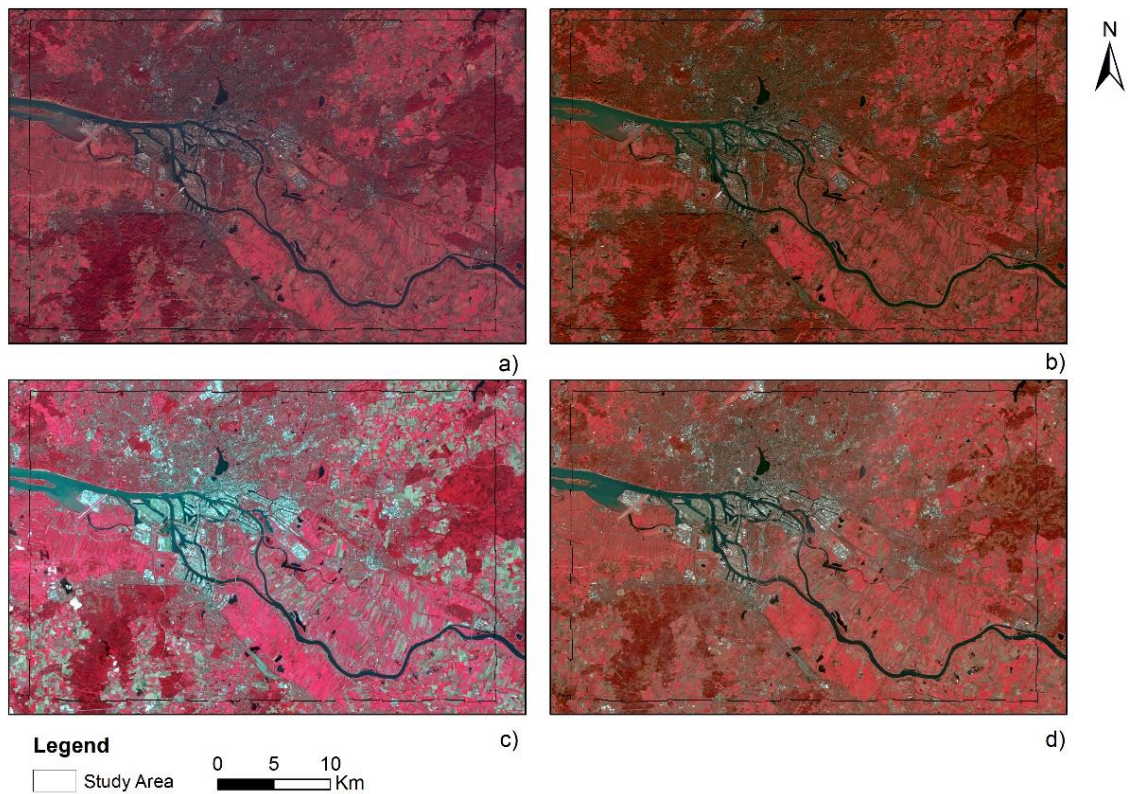


Figure 34: False color composite RGB = 543 of the Landsat 8 imagery of Hamburg in a) winter, b) autumn, c) summer and d) spring

## 4.2. Classification with the WUDAPT approach

### 4.2.1. Coimbra

The classification into LCZ classes proposed in the WUDAPT project was applied to the four images of each study area, providing four LCZ maps per area. After the definition of the ROI, it is necessary to identify training areas for each one of the LCZ classes. As some physical characteristics of the territory change with the seasons, which translate into different spectral responses for the same area at different times of the year, different training areas were selected for the images in each season, mainly for the vegetation classes. The training regions are shown in Figure 35 for Coimbra.





Figure 35: Training areas over the Landsat 8 image (True-Color RGB) of Coimbra in a) winter, b) autumn, c) summer and d) spring

The classification of the images for each season was made using the methodology presented in section 3.1. Figures 36 and 37 show, respectively, the classification results of the four images with a spatial resolution of 120 m using a filter of twenty-five neighboring cells and a spatial resolution of 30 m using a filter of sixty-four neighboring cells.

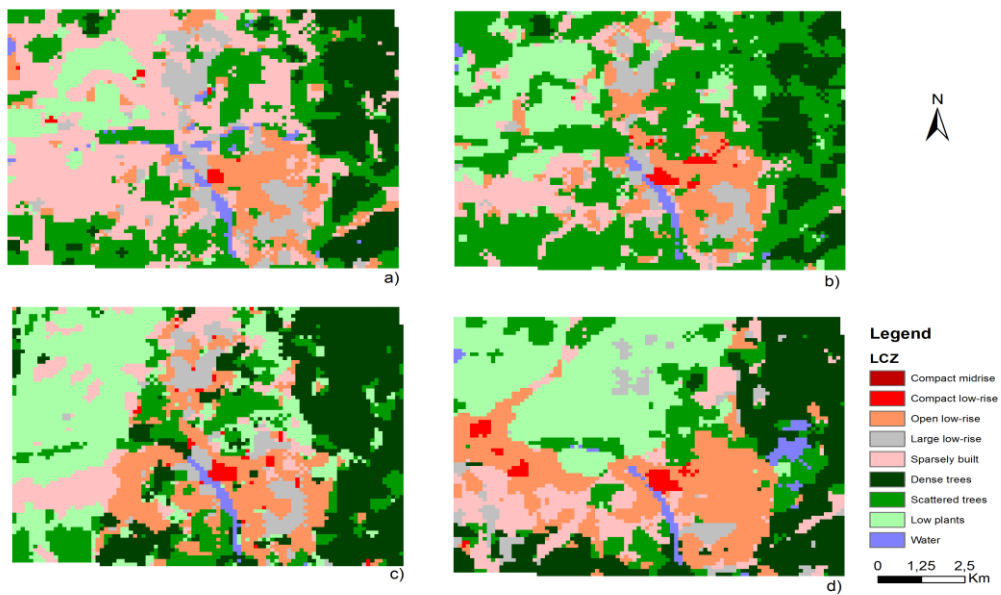
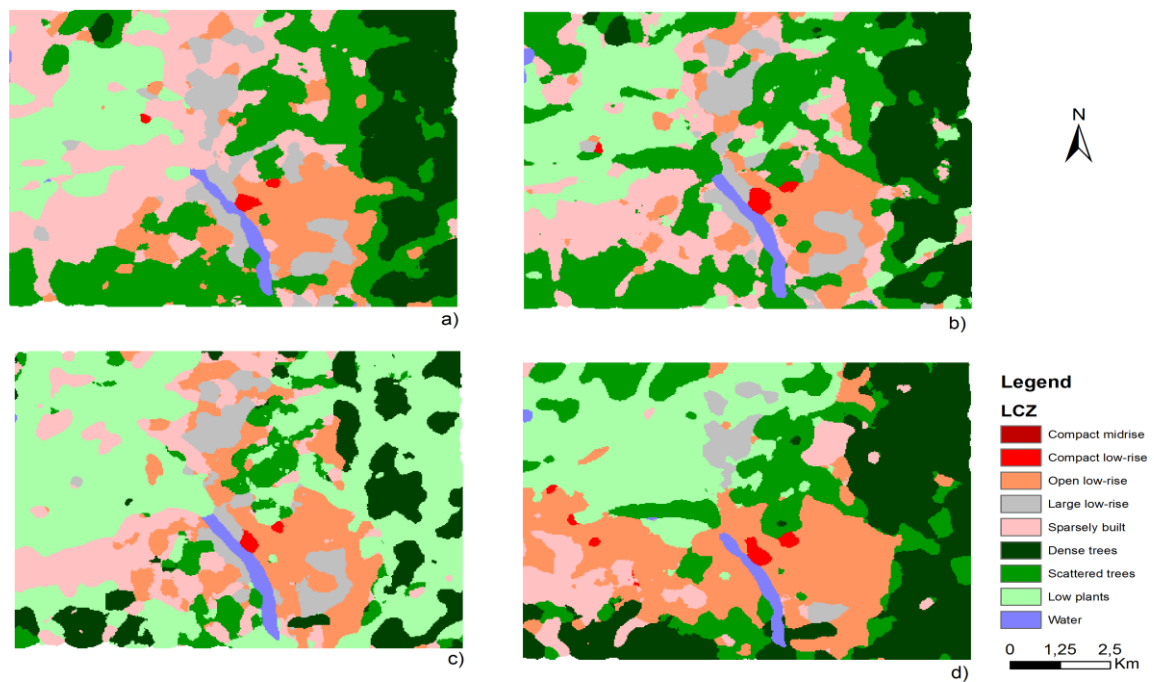


Figure 36: LCZs maps obtained for the city of Coimbra with a spatial resolution of 120 m with a filter of twenty-five neighboring cells for a) winter, b) autumn, c) summer and d) spring





**Figure 37: LCZs maps obtained for the city of Coimbra with a spatial resolution of 30 m with a filter of sixty-four neighboring cells for a) winter, b) autumn, c) summer and d) spring**

Based on a visual comparison of Figures 36 and 37, it can be stated that the results show a relevant correspondence with the land cover. With the first figure, the misclassification in rice Mondego fields is evident, particularly in winter and spring time. The best results for this area were obtained from the summer imagery. The second image shows satisfying results overall, where the major misclassification occurs in the summer image on the right side of the image with the attribution of LCZ D class (Low plants) to forest areas.

To validate the results, the methodology explained in section 3.4 was used.

Tables 10, 12, 14 and 16 show the confusion matrices for the classification of Coimbra using Landsat 8 data with a 120 m spatial resolution for winter, autumn, summer and spring, respectively, and Tables 11, 13, 15 and 17 show the confusion matrices for the classification of Coimbra using Landsat 8 data with a 30 m spatial resolution for winter, autumn, summer and spring, respectively.

**Table 10: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the winter satellite image for Coimbra. The user accuracy (UA), producer accuracy (PA) and overall accuracy (OA) are provided. Rows refer to classification output and columns to reference data.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	17	3	107	11	37	2	15	4	4	200	9%
3	0	119	67	0	0	0	10	4	0	200	60%
6	4	10	80	4	60	3	10	29	0	200	40%
8	7	5	17	85	52	1	17	13	3	200	43%
9	0	0	6	3	93	10	48	38	2	200	47%
A	0	0	0	0	22	47	125	6	0	200	24%
B	0	0	0	0	23	31	137	6	3	200	69%
D	0	0	0	0	24	10	23	139	4	200	70%
G	0	2	3	4	19	46	77	19	30	200	15%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	61%	86%	29%	79%	28%	31%	30%	54%	65%	OA	42%

The LCZ map (120 m spatial resolution) obtained with the classification of the winter satellite image shows a higher user's accuracy for the classes LCZ D (Low Plants) and LCZ B (Scattered Trees) with 70% and 69%, respectively and a lower user's accuracy for classes LCZ 2 (Compact midrise) – 9% and LCZ G (Water) – 15%. The producer's accuracy is higher for classes LCZ 3 (Compact low-rise) (86%) and LCZ 8 (Large Low-rise) (79%) and lower for classes LCZ 9 (Sparsely built) – 28%. There appears to be a considerable amount of confusion between classes LCZ A and B, and also between classes LCZ 2 and LCZ 6.

**Table 11: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the winter satellite image for Coimbra.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	7	0	0	0	0	0	0	0	0	7	100%
3	2	116	61	0	0	0	0	9	0	188	62%
6	10	18	196	10	117	10	38	30	2	431	45%
8	8	4	12	72	25	2	3	3	1	130	55%
9	1	0	8	22	106	19	77	75	13	321	33%
A	0	0	1	1	26	60	148	5	3	244	25%
B	0	0	0	1	44	55	175	14	1	290	60%
D	0	0	0	0	9	3	13	118	1	144	82%
G	0	1	2	1	3	1	8	4	25	45	56%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	25%	83%	70%	67%	32%	40%	38%	46%	54%	OA	49%

Nevertheless, for the winter 30 m resolution LCZ map the higher user's accuracy was obtained for classes LCZ 2 - 100% and LCZ D – 82%, and lower user's accuracy for classes LCZ A (25%) and LCZ 9 (33%). The producer's accuracy is higher for classes LCZ 3 – 83% and LCZ

6 – 70% and lower for classes LCZ 2 – 25% and LCZ 9, with 32%. There is a certain amount of confusion between classes LCZ A and LCZ B and between classes LCZ 6 and LCZ 9.

The biggest differences between the two results was in LCZ 2, with a gain in user accuracy of 91% in the 30 m resolution LCZ map and in the LCZ G with a gain of 41% also in 30 m resolution map. The producer accuracy had a major loss in LCZ 2, namely, 36%

**Table 12: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the autumn satellite image for Coimbra,**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	12	1	35	1	0	0	0	0	1	50	24%
3	2	115	78	0	0	1	1	1	0	198	58%
6	11	17	112	19	102	2	16	18	0	297	38%
8	2	3	34	80	26	1	8	5	0	159	50%
9	0	0	12	3	115	12	55	61	0	258	45%
A	0	1	4	1	51	45	131	28	6	267	17%
B	0	0	0	0	3	44	150	9	1	207	72%
D	0	0	1	1	23	9	34	120	3	191	63%
G	1	2	4	2	10	36	67	16	35	173	20%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	43%	83%	40%	75%	35%	30%	32%	47%	76%	OA	44%

For the image classification using the 120 m resolution autumn satellite image, the results show that there is also higher user's accuracy for the classes LCZ B - 72% and LCZ D – 63%, and a lower user's accuracy for classes LCZ A (17%) and LCZ G (20%). The producer's accuracy is higher for classes LCZ 3 – 83% and LCZ G – 76% and lower for classes LCZ A – 30% and LCZ B, with 32%. As for the image classification using the winter satellite image, in this image a certain amount of confusion between classes LCZ A and LCZ B also occurs as well as between classes LCZ 6 and LCZ 9.

**Table 13: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the autumn satellite image for Coimbra.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	0	0	0	0	0	0	0	0	0	0	0%
3	0	111	51	0	1	0	1	6	0	170	65%
6	21	20	201	13	143	14	29	35	1	477	42%
8	6	7	17	86	16	0	6	3	1	142	61%
9	0	0	9	6	120	12	90	116	4	357	34%
A	0	0	0	0	30	63	157	14	1	265	24%
B	0	0	0	0	15	42	147	10	2	216	68%
D	0	0	0	0	0	1	3	67	2	73	92%
G	1	1	2	2	5	18	29	7	35	100	35%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	0%	80%	72%	80%	36%	42%	32%	26%	76%	OA	46%

The overall accuracy of the classification using the 30 m autumn satellite image was 46 %, with a higher user’s accuracy for the classes LCZ D (Low plants) – 92 % and LCZ B (Scattered trees) – 68%. The producer’s accuracy is higher for classes LCZ 3 and LCZ 8, both with 80%. A lower user’s and producer’s accuracy was obtained for class LCZ 2 (0%), since no pixel was classified as belonging to that class. In this image, it appears that also occurred a certain amount of misclassification between classes LCZ 6 and LCZ 9 and between classes LCZ A and LCZ B. The biggest different in user accuracy was the gain of 29% in LCZ D in the classification using the 30 m resolution satellite image and for the producer accuracy was the gain of 32% in LCZ 6, also in the classification using the 30 m satellite image.

**Table 14: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the summer satellite image for Coimbra.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	13	4	27	1	5	0	1	1	0	52	25%
3	2	121	45	2	1	0	0	1	0	172	70%
6	6	11	168	32	109	5	27	14	0	372	45%
8	5	1	11	63	6	0	6	0	1	93	68%
9	0	0	15	0	68	0	26	18	0	127	54%
A	0	0	1	2	37	95	219	15	10	379	25%
B	0	0	0	0	18	34	107	84	6	249	43%
D	1	2	9	6	85	16	73	117	6	315	37%
G	1	0	4	1	1	0	3	8	23	41	56%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	46%	87%	60%	59%	21%	63%	23%	45%	50%	OA	43%

For the image classification using the 120 m resolution summer satellite image, the overall accuracy was 43% with a higher user’s accuracy for the classes LCZ 3 and LCZ 8, with 70% and 68% respectively, and a lower user’s accuracy for classes LCZ 2 and LCZ A, both with 25%. The producer’s accuracy is higher for classes LCZ 3 (87%) and LCZ A (63%) and lower for LCZ 9 (21%) and LCZ B (23%). The major misclassification occurs between the LCZ A and LCZ B.

**Table 15: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the summer satellite image for Coimbra.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	8	2	4	1	3	1	0	0	0	19	42%
3	0	103	32	0	0	0	1	0	0	136	76%
6	13	24	215	10	87	7	31	17	0	404	53%
8	6	9	14	88	9	0	13	0	1	140	63%
9	0	0	10	5	129	2	51	36	0	233	55%
A	0	0	1	2	43	75	177	35	3	336	22%
B	0	0	0	0	35	59	164	123	17	398	41%
D	0	0	1	1	20	5	18	43	0	88	49%
G	1	1	3	0	4	1	7	4	25	46	54%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	29%	74%	77%	82%	39%	50%	35%	17%	54%	OA	47%

The LCZ classification using the 30 m resolution summer satellite image had an overall accuracy of 47%. The higher user's accuracy appears for classes LCZ 3 (76 %) and LCZ 8 (63%) while the classes LCZ A and LCZ B had lower user's accuracies, 22% and 41%, respectively. The producer's accuracy is higher for classes LCZ 8 (82%) and LCZ 6 (77%) and lower for classes LCZ D (17%) and LCZ 2 (29%). There was a certain amount of confusion between the classes LCZ A and LCZ B, and also between LCZ 6 and LCZ 9.

The biggest differences between the two results was the loss in producer accuracy of 28% in LCZ D and the gain in producer accuracy of 23% in LCZ 8.

**Table 16: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the spring satellite image for Coimbra.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	9	2	4	3	3	0	0	0	0	21	43%
3	0	122	62	0	0	1	4	14	0	203	60%
6	10	9	142	7	71	2	16	26	3	286	50%
8	6	2	8	59	6	0	12	26	1	120	49%
9	1	1	12	12	66	1	13	38	0	144	46%
A	0	2	36	5	110	123	313	23	16	628	20%
B	0	0	0	0	26	18	67	21	4	136	49%
D	1	1	12	20	46	5	33	102	3	223	46%
G	1	0	4	1	2	0	4	8	19	39	49%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	32%	88%	51%	55%	20%	82%	15%	40%	41%	OA	39%

The LCZ classification using the 120 m resolution spring satellite image showed an overall accuracy of 39%. The higher user's accuracy was obtained for classes LCZ 3 (60%) and LCZ 6 (50%) while the classes LCZ A and LCZ 2 have a lower user's accuracy, 20% and 43%, respectively.

The producer's accuracy is higher for classes LCZ 3 (88%) and LCZ A (82%) and lower for classes LCZ B (15%) and LCZ 9 (20%). It also appears that there is a certain amount of confusion between the classes LCZ A and LCZ B, and also between LCZ A and LCZ 9.

The overall accuracy of the classification using the 30 meters' spring satellite image also was 39%, with a higher user's accuracy for classes LCZ 3 (Compact low-rise) and LCZ D (Low plants), both with 70%, and a low user's accuracy for LCZ A (Dense Trees) with 21%. The producer's accuracy is higher for classes LCZ 3 with 85% and a lower user's accuracy was obtained for class LCZ D (16%). Similar to the previous image classifications, there was a certain amount of confusion between classes LCZ A and LCZ B.

Comparing both results, the class LCZ D presents the major gain in user accuracy and the major loss in producer accuracy, both with 24%.

**Table 17: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the spring satellite image for Coimbra.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
2	11	1	12	0	1	1	0	0	0	26	42%
3	0	118	44	0	1	0	1	5	0	169	70%
6	9	14	191	31	121	11	50	76	9	512	37%
8	5	4	11	71	10	0	17	50	0	168	42%
9	0	0	4	1	59	2	14	27	0	107	55%
A	2	0	11	3	75	106	267	24	8	496	21%
B	0	1	0	0	50	23	84	24	4	186	45%
D	0	0	3	0	9	2	3	40	0	57	70%
G	1	1	4	1	4	5	26	12	25	79	32%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	39%	85%	68%	66%	18%	71%	18%	16%	54%	OA	39%

#### 4.2.2. Hamburg

The creation of the training sites for Hamburg was done using the same approach (training sets were used in each of the images). Figure 38 shows the training areas used for the Hamburg study area.

For Hamburg, the results of the classification process of the ROI into LCZ maps are illustrated in Figures 39 and 40, for the classification of the Landsat 8' images with a 120 m and 30 m spatial resolution, respectively.

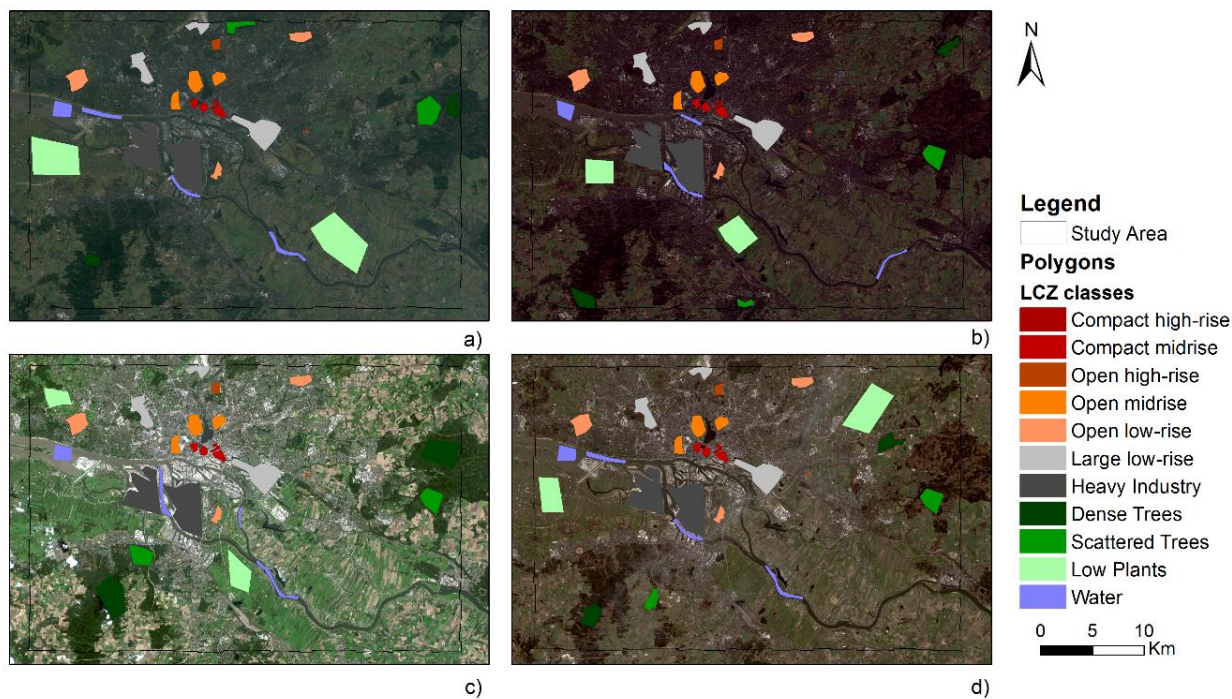


Figure 38: Training areas over the Landsat 8 image (True-Color RGB) of Hamburg in a) winter, b) autumn, c) summer and d) spring

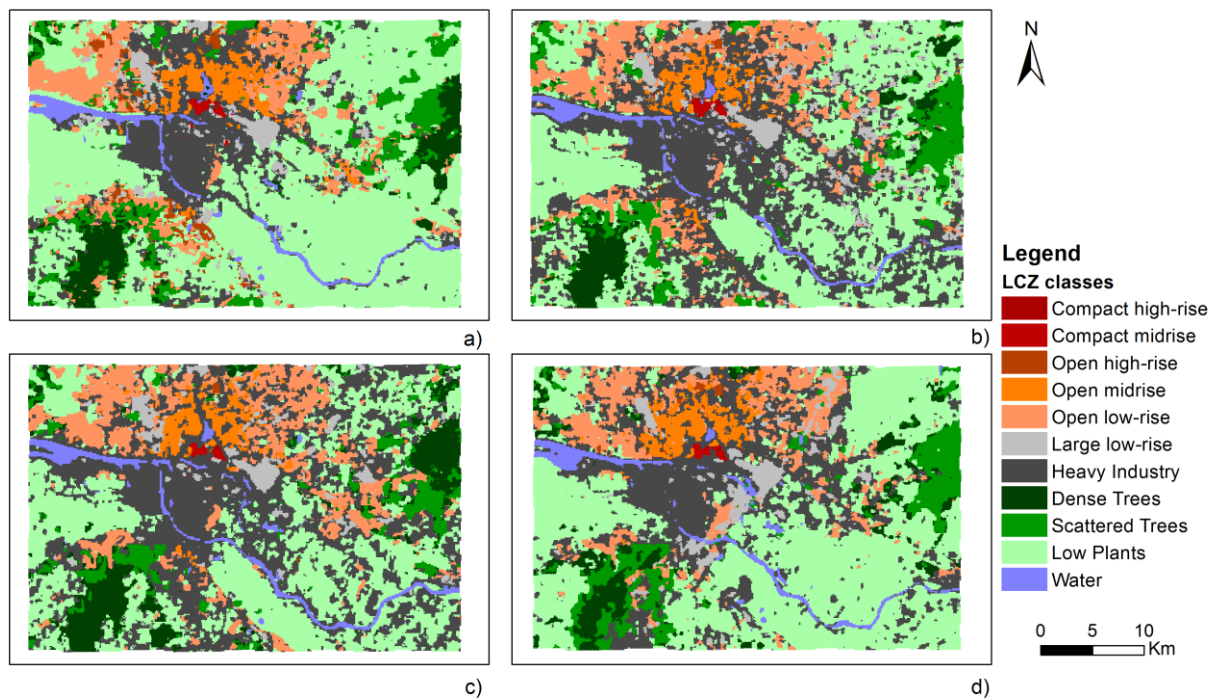
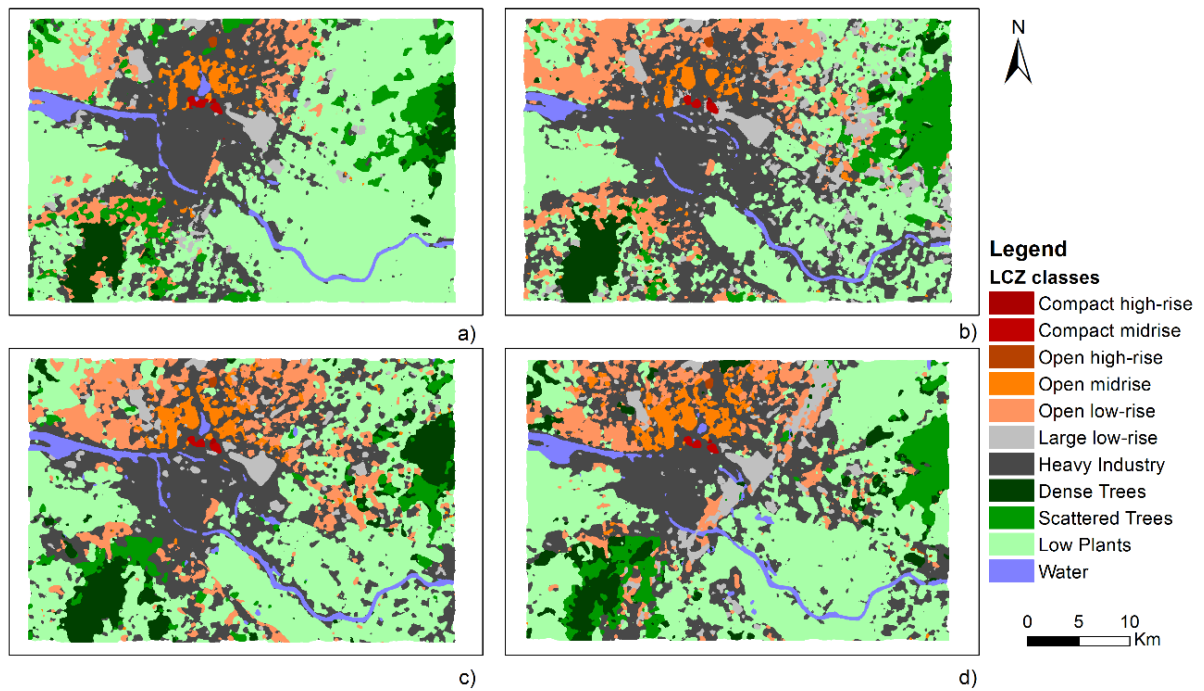


Figure 39: LCZs maps for the city of Hamburg with a spatial resolution of 120 m with a filter of twenty-five neighboring cells for a) winter, b) autumn, c) summer and d) spring





**Figure 40: LCZs maps for the city of Hamburg with a spatial resolution of 30 m with a filter of sixty-four neighboring cells for a) winter, b) autumn, c) summer and d) spring**

The results obtained for Hamburg were also validated using a stratified random sample of points as for Coimbra, where the classes were obtained using the classification of the winter image as strata (also considering 200 per class). As in the previous study area, the reference dataset was also created through photo-interpretation of the images available in Google Earth.

Tables 18, 20, 22 and 24 show the confusion matrices for the classification of Hamburg using Landsat 8 data with a 120 m spatial resolution for winter, autumn, summer and spring, respectively, and Tables 19, 21, 23 and 25 show the confusion matrices for the classification of Hamburg using Landsat 8 data with a 30 m spatial resolution for winter, autumn, summer and spring, respectively.



**Table 18: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the winter satellite image for Hamburg**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	Σ	UA
1	182	18	0	0	0	0	0	0	0	0	0	0	200	91%
2	1	156	0	2	0	0	0	23	0	0	0	18	200	78%
4	0	0	112	4	42	5	2	0	9	20	5	1	200	56%
5	1	0	1	158	21	8	0	1	0	3	1	6	200	79%
6	0	0	2	1	152	5	0	2	11	21	6	0	200	76%
8	0	0	3	5	28	141	0	13	0	3	3	4	200	71%
10	0	1	5	22	43	4	1	67	1	9	23	24	200	34%
A	0	0	0	0	4	0	0	0	162	26	7	1	200	81%
B	0	0	0	0	21	0	2	0	51	114	12	0	200	57%
D	0	0	1	0	18	0	2	0	6	12	159	2	200	80%
G	0	1	0	0	1	0	1	2	0	3	3	189	200	95%
Σ	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	99%	89%	90%	82%	46%	87%	0%	62%	68%	54%	73%	77%	OA	72%

The classification using the winter image for Landsat 8 data with 120 meters of resolution has the best overall accuracy of the four classifications, namely, 72%.

For this classification, the higher user's accuracy is for classes G (Water) and LCZ 1 (Compact high-rise) with 95% and 91%, respectively and lower user's accuracy for classes LCZ 10 (Heavy industry) – 34% and LCZ 4 (Open high-rise) – 56%. On the other hand, the producer's accuracy is higher for classes LCZ 1 – 99% and LCZ 4 – 90% and lower for classes LCZ 9 (Sparsely built) (0%) and LCZ 6 (Open low-rise) – 46%. There also appears to be considerable confusion between classes LCZ 10 and 2 (Compact mid-rise), and also between classes LCZ 6 and LCZ 10.

**Table 19: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the winter satellite image for Hamburg.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	Σ	UA
1	161	8	0	0	0	0	0	0	0	0	0	0	169	95%
2	1	123	0	0	0	0	0	1	0	0	0	7	132	93%
4	0	0	8	0	0	0	0	0	0	0	0	0	8	100%
5	6	4	0	74	5	0	0	0	0	2	1	2	94	79%
6	0	0	6	6	90	2	1	0	25	28	6	0	164	55%
8	0	8	7	4	15	100	0	7	0	2	1	4	148	68%
10	16	33	80	100	127	50	2	95	1	30	34	76	644	15%
A	0	0	0	0	1	0	0	0	148	24	2	0	175	85%
B	0	0	1	0	22	0	2	1	54	100	8	0	188	53%
D	0	0	22	8	70	11	3	3	12	25	166	1	321	52%
G	0	0	0	0	0	0	0	1	0	0	1	155	157	99%
Σ	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	88%	70%	6%	39%	27%	61%	0%	88%	62%	47%	76%	63%	AO	55%

For the classification using the winter image for Landsat 8 data with 30 meters of resolution, a higher user's accuracy is obtained for classes LCZ 4 (Open high-rise), LCZ G (Water) and LCZ 1 (Compact high-rise) with 100%, 99% and 95%, respectively and a lower user's accuracy for classes LCZ 10 (Heavy industry) – 15% and LCZ D (Low plants) – 52%. However, the producer's accuracy is higher for classes LCZ 1 and LCZ 10, both with 88%, but is lower for classes LCZ 9 (Sparsely built) (0%) and LCZ 4 (Open high-rise) – 6%. There also appears to be a considerable amount of confusion between classes LCZ A (Dense trees) and LCZ B (Scattered trees) and also between LCZ B and LCZ D (Low plants).

**Table 20: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the autumn satellite image for Hamburg.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	173	18	0	0	1	0	0	0	0	0	0	0	200	87%
2	11	139	0	1	0	1	0	0	0	0	0	10	200	70%
4	0	1	5	3	4	1	0	0	0	0	1	0	200	3%
5	0	2	5	104	7	2	0	3	0	3	0	1	200	52%
6	0	0	50	11	150	4	1	0	5	29	4	6	200	75%
8	0	5	7	7	21	101	0	22	0	3	6	5	200	51%
10	0	11	57	66	120	54	3	82	8	23	64	74	200	41%
A	0	0	0	0	3	0	1	0	111	32	4	0	200	56%
B	0	0	0	0	7	0	1	0	99	96	7	16	200	48%
D	0	0	0	0	17	0	2	1	17	25	132	1	200	66%
G	0	0	0	0	0	0	0	0	0	0	1	132	200	66%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	94%	79%	4%	54%	45%	62%	0%	76%	46%	45%	60%	54%	AO	56%

The classification using the 120 m resolution autumn image has the second-best value of overall accuracy of all images (56%) and it has a higher user's accuracy for classes LCZ 1 (Compact high-rise) (87%) and LCZ 6 (75%), and lower user's accuracy for classes LCZ 4 (3%) and LCZ 10 (41%). The producer's accuracy is higher for classes LCZ 1 – 94% and LCZ 2 – 79% and lower for classes LCZ 9 (0%) and LCZ 4 (4%). In this image, it appears that a certain amount of confusion also occurs between classes LCZ 6 and LCZ 10 and between classes LCZ A and LCZ B (Scattered Trees).

**Table 21: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the autumn satellite image for Hamburg.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	153	15	0	0	0	0	0	0	0	0	0	0	168	91%
2	1	112	0	0	0	0	0	0	0	0	0	6	119	94%
4	0	0	7	0	0	0	0	0	0	0	0	0	7	100%
5	0	0	1	47	2	0	0	0	0	1	0	0	51	92%
6	0	0	37	15	157	2	0	1	26	62	9	4	313	50%
8	17	4	4	8	21	89	0	21	0	3	13	5	185	48%
10	13	45	75	122	122	72	4	85	8	29	73	130	778	11%
A	0	0	0	0	6	0	2	0	120	15	3	0	146	82%
B	0	0	0	0	14	0	1	0	76	77	8	0	176	44%
D	0	0	0	0	7	0	1	0	10	24	113	1	156	72%
G	0	0	0	0	1	0	0	1	0	0	0	99	101	98%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	83%	64%	6%	24%	48%	55%	0%	79%	50%	36%	52%	40%	AO	48%

The classification using the autumn image has an overall accuracy of 48%. The higher user's accuracy is obtained for classes LCZ 4 and LCZ G with 100% and 98%, respectively and lower user's accuracy for classes LCZ 10 – 11% and LCZ B – 44%. On other hand, the producer's accuracy is higher for classes LCZ 1 – 83% and LCZ 10 – 79% and lower for classes LCZ 9 - 0% and LCZ 4 – 6%.

It also appears to be a lot of confusion between classes LCZ 2 and 10 and also between classes LCZ 6 and LCZ 10.

**Table 22: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the summer satellite image for Hamburg.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	173	18	0	0	0	0	0	0	0	0	0	0	200	87%
2	1	117	0	0	0	1	0	4	0	0	0	5	200	59%
4	0	0	6	0	0	0	0	0	0	0	0	0	200	3%
5	0	1	6	102	7	5	0	2	0	2	1	0	200	51%
6	0	0	69	25	160	2	1	1	3	22	3	0	200	80%
8	0	5	2	15	12	89	0	21	0	0	3	4	200	45%
10	10	35	41	50	128	66	3	79	10	28	70	66	200	40%
A	0	0	0	0	6	0	2	1	137	78	9	0	200	69%
B	0	0	0	0	6	0	0	0	82	62	6	2	200	31%
D	0	0	0	0	11	0	2	0	8	18	125	1	200	63%
G	0	0	0	0	0	0	0	0	0	1	2	167	200	84%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	94%	66%	5%	53%	48%	55%	0%	73%	57%	29%	57%	68%	AO	55%

The LCZ classification using the 120 meters' summer satellite image shows an overall accuracy of 55% with the higher user's accuracy at the classes LCZ 1 (Compact high-rise) and LCZ G (Water), with 87% and 84% respectively, and lower user's accuracy for class LCZ 4 (Open high-rise) with 3%.

The producer's accuracy is higher for classes LCZ 1 – (94%) and LCZ 10 (73%) and lower at LCZ 9 (0%) and LCZ 4 (5%). The major misclassification occurs between the LCZ 6 and LCZ 10.

**Table 23: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the summer satellite image for Hamburg.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	154	15	0	0	0	0	0	0	0	0	0	0	169	91%
2	1	108	0	0	0	0	0	0	0	0	0	4	113	96%
4	0	0	7	0	0	0	0	0	0	0	0	0	7	100%
5	0	0	3	100	5	1	0	1	0	2	0	0	112	89%
6	0	0	67	20	149	2	0	0	3	12	2	0	255	58%
8	0	7	0	6	7	91	0	13	0	0	2	3	129	71%
10	29	46	47	66	136	68	3	93	9	30	47	67	641	15%
A	0	0	0	0	3	0	1	0	142	77	7	0	230	62%
B	0	0	0	0	4	0	0	0	67	45	4	2	122	37%
D	0	0	0	0	26	1	4	1	19	44	156	2	253	62%
G	0	0	0	0	0	0	0	0	0	1	1	167	169	99%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	84%	61%	6%	52%	45%	56%	0%	86%	59%	21%	71%	68%	AO	55%

The LCZ classification using the 30 meters' summer satellite image has an overall accuracy of 55%. The higher user's accuracy it appears for classes LCZ 4 (100%) and LCZ G (99%) while the classes LCZ 10 and LCZ B has the lower user's accuracy, 15% and 37%, respectively.

The producer's accuracy is higher for classes LCZ 1 (84%) and LCZ D (71%) and lower for classes LCZ 9 (0%) and LCZ 4 (6%). It also appears a certain amount of confusion between the classes LCZ 6 and LCZ 10, and also between LCZ B and LCZ A.

**Table 24: Confusion matrix of the LCZs map with a spatial resolution of 120 m obtained using the spring satellite image for Hamburg.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	160	18	0	0	0	0	0	0	0	0	0	0	200	80%
2	24	124	1	0	0	0	0	1	0	0	0	5	200	62%
4	0	0	7	0	6	0	0	0	1	0	0	0	200	4%
5	0	1	6	103	18	6	0	0	0	3	0	0	200	52%
6	0	0	55	11	128	7	0	0	0	15	9	0	200	64%
8	0	4	3	7	13	90	0	33	1	1	2	5	200	45%
10	0	29	47	70	93	58	3	73	4	24	52	80	200	37%
A	0	0	0	0	10	0	0	0	111	53	3	0	200	56%
B	0	0	0	0	6	0	3	0	107	82	11	2	200	41%
D	0	0	5	1	56	2	2	1	16	32	142	5	200	71%
G	0	0	0	0	0	0	0	0	0	1	0	148	200	74%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	87%	70%	6%	54%	39%	55%	0%	68%	46%	39%	65%	60%	AO	53%

In the confusion matrix for the 120 meters' spring image, the classes with higher user accuracy are LCZ 1 (Compact high-rise): (80%) and LCZ G (Water) – 74% and a lower user accuracy was found for class LCZ 4 – Open mid-rise (4%).

The producer's accuracy is higher for classes LCZ 1 (87%) and LCZ 2 (Compact mid-rise) - (70%) and lower for classes LCZ 9 – Sparsely Built (0%), and LCZ 4 – Open mid-rise (6%). A certain amount of confusion also appears between the classes LCZ 4 and LCZ 6, and also between LCZ 6 and LCZ 10.

**Table 25: Confusion matrix of the LCZs map with a spatial resolution of 30 m obtained using the spring satellite image for Hamburg.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	122	0	0	0	0	0	0	0	0	0	0	0	122	100%
2	1	107	0	0	0	0	0	0	0	0	0	4	112	96%
4	0	0	6	0	0	0	0	0	0	0	1	0	7	86%
5	6	0	16	97	15	6	0	1	0	2	0	0	143	68%
6	0	0	32	7	96	4	0	0	1	9	8	0	157	61%
8	0	4	5	9	15	91	0	27	1	2	6	6	166	55%
10	55	65	58	77	138	59	4	79	4	31	43	71	684	12%
A	0	0	0	0	12	0	0	0	114	56	6	0	188	61%
B	0	0	0	0	5	0	2	0	105	74	4	0	190	39%
D	0	0	7	2	49	3	2	1	15	37	150	9	275	55%
G	0	0	0	0	0	0	0	0	0	0	1	155	156	99%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	66%	61%	5%	51%	29%	56%	0%	73%	48%	35%	68%	63%	AO	50%

The classification using 30 meters' spring image has the third larger value of overall accuracy of all images (50%) and the classes with higher user's accuracy are LCZ 1 (Compact high-rise) - (100%) and LCZ G (Water) – 99% and lower user's accuracy was found for classes LCZ 10 (12%) and LCZ B (39%). The producer's accuracy is higher for classes LCZ 10 – 73% and LCZ D – 68% and lower for classes LCZ 9 (0%) and LCZ 4 with 5%. In this image, it appears that occurs a certain amount of confusion also between classes LCZ 4 and LCZ 6.

### 4.3. Conversion of the OSM data into LCZ classes

#### 4.3.1. Coimbra

Figure 41 a) shows Landsat 8 true color imagery of the study area. The images shown in Figure 41 b) to f) represent the percentage of area occupied in each cell by: b) LCZ class A (Dense Trees) or B (Scattered trees), c) LCZ class D (Low plants), d) LCZ class G (Water), e) the building surface fraction, and f) the impervious surface fraction.

A visual analysis of the results shows that for classes LCZ A (Dense Trees) or LCZ B (Scattered Trees) (Figure 41 b), as well as for class LCZ D (Low Plants) (Figure 41 c), compact regions of cells with more than 80% occupation were obtained and only a few cells with low occupation can be seen, located mainly in the borders of the larger regions.

For class LCZ G (Water) (Figure 41 d), a large number of cells have a low percentage of occupation, corresponding to existing creeks, and only the main watercourse of the study area (Mondego River) has high values of occupation.

Relative to the building surface fraction (Figure 41 e), it can be seen that there is little information for the study area and that most of the available information indicates a low percentage of area occupied by buildings due to missing data in OSM. On the other hand, the impervious surface fraction (Figure 41 f) shows the railway and road network mapped in great detail, even though these structures only occupy a small percentage of the cell's area, i.e. usually less than 20%.

Even though there is a considerable amount of missing data in OSM for the study area (Figure 41), the visual comparison of the results obtained for the existing data with the Landsat true color image shows a good level of correspondence with the ground cover.

In general, a visual analysis of the results obtained for the natural LCZ classes (LCZ A or B, LCZ D and LCZ G), in comparison with the satellite image, showed relevant correspondence with the land cover.

Another procedure that was undertaken was the extraction of the confidence of each class ( $x$ ) in each cell depending on the number of times that this class was assigned to the cell in the four LCZ maps (corresponding to different seasons). This procedure was implemented on all LCZ classes. Figures 42 and 43 show two examples of the results obtained for LCZ A and LCZ D, respectively. The results for the other classes are in Appendix N, from Figures 67 to 72.

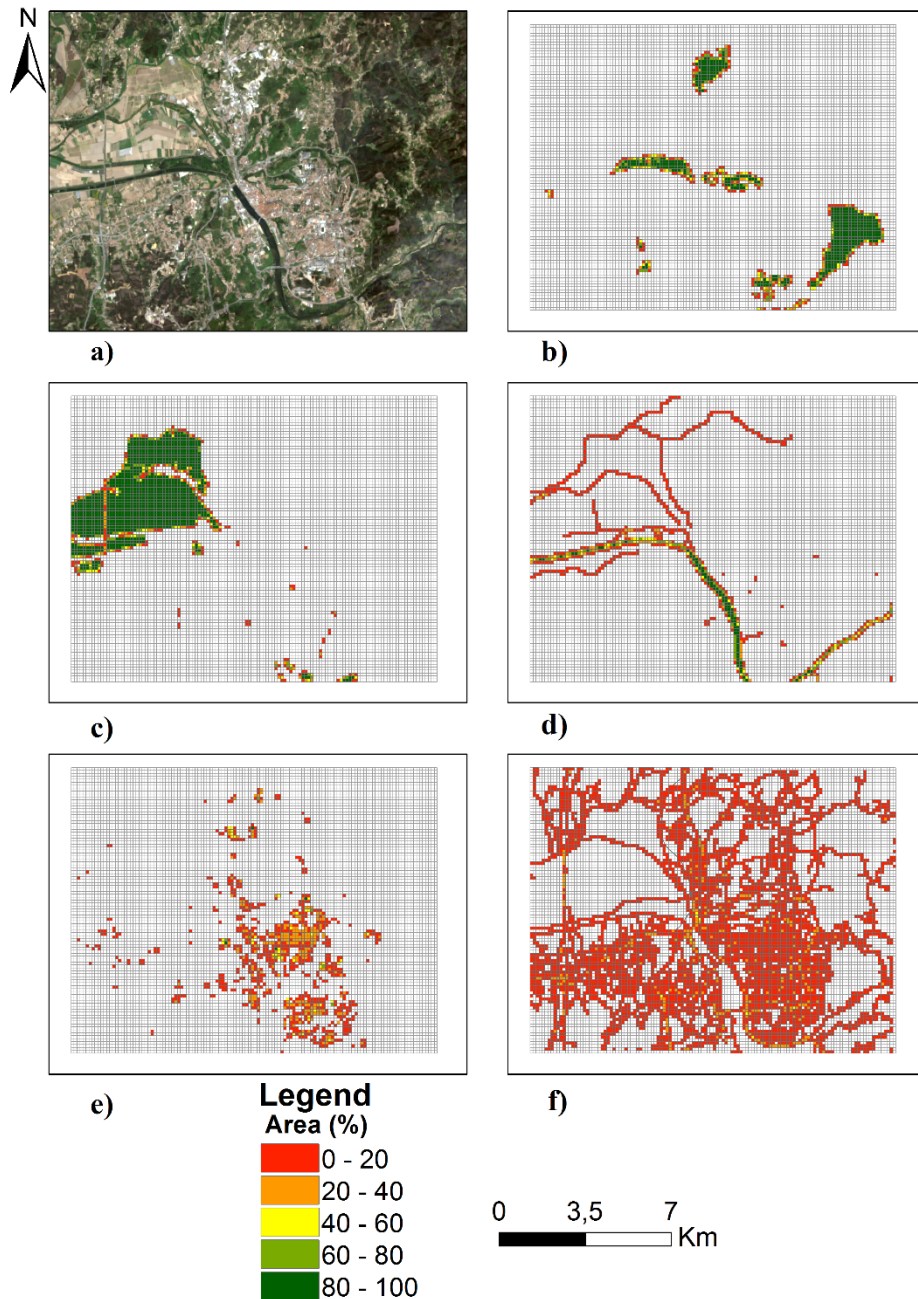


Figure 41: a) Landsat 8 true color imagery (RGB 432). Results obtained from OSM for the percentage of area occupied in each cell by: b) LCZ A (Dense Trees) or LCZ B (Scattered Trees), c) LCZ class D (Low Plants), d) LCZ class G (Water), e) building surface fraction and f) impervious surface fraction

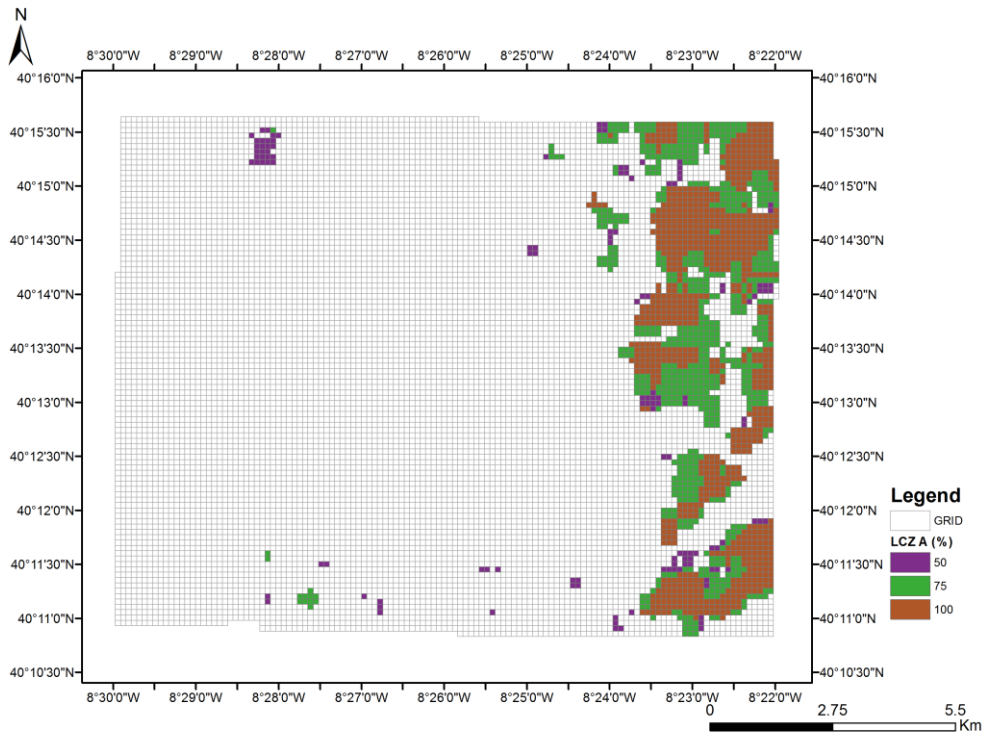


Figure 42: Confidence of presence of LCZ A in each pixel based on the classification of four satellite images, in percentage, for Coimbra

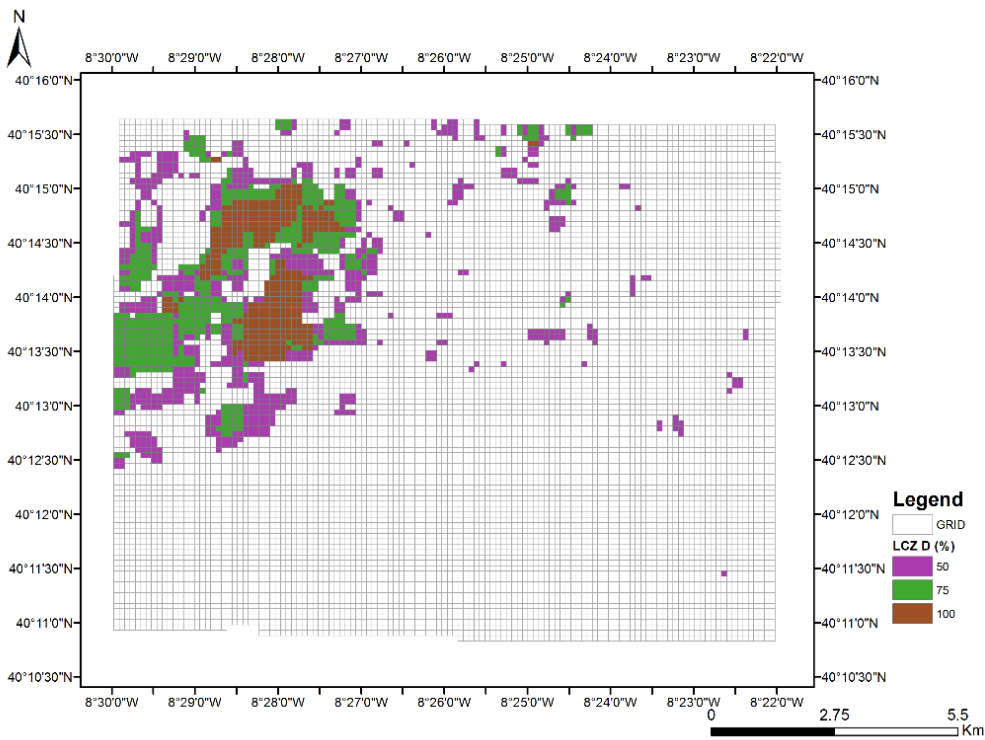


Figure 43: Confidence of presence of LCZ D in each pixel based on the classification of four satellite images, in percentage, for Coimbra



### 4.3.2. Hamburg

Figures 44 to 47 show the results of the conversion of OSM data into LCZ classes by class and percentage of area occupied in each cell by the class, for classes LCZ A (Dense Trees) or B (Scattered trees) – Figure 44, LCZ C (Bush, scrub) – Figure 45, LCZ D (Low plants) – Figure 46 and LCZ G (Water) – Figure 47. The urban classes are presented in Figures 52 to 58 and represent the classes LCZ 3 (Compact low-rise), LCZ 4 (Open high-rise), LCZ 5 (Open mid-rise), LCZ 6 (Open low-rise), LCZ 7 (Lightweight low-rise), LCZ 9 (Sparsely built) and LCZ 10 (Heavy industry).

As for Coimbra, a visual analysis of the results shows that for classes LCZ A (Dense Trees) or LCZ B (Scattered Trees) (Figure 44), as well as for class LCZ D (Low Plants) (Figure 46), compact regions of cells with more than 80% occupation were obtained and only a few cells with low occupation can be seen, located mainly in the borders of the larger regions. The LCZ C (Figure 45) has low representation in the study area, where numerous cells with a percentage of occupation of 0 to 20% exist but only a relatively small compact region with large values of occupation can be seen.

Class LCZ G (Water) (Figure 47) has a small number of cells with high values of occupation corresponding to the main watercourse of the study area (Elba River) and has a large number of cells with a low percentage of occupation, corresponding to existing streams.

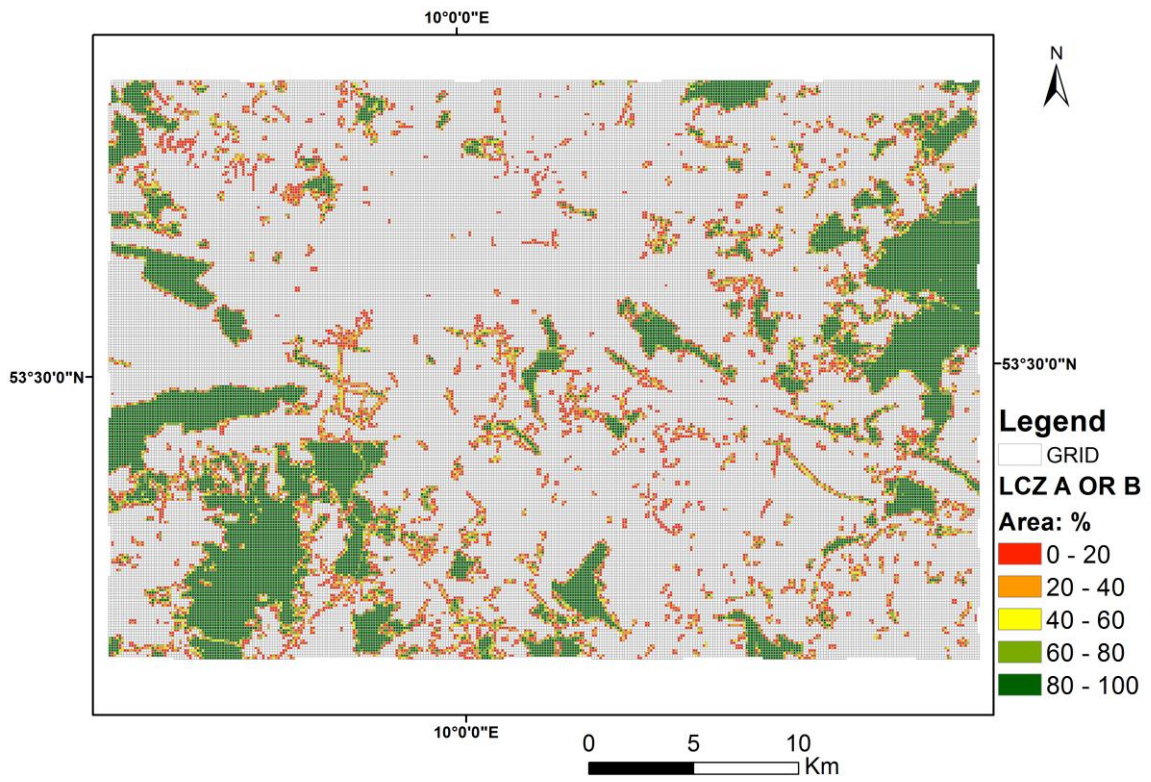


Figure 44: Results obtained from OSM for the percentage of area occupied in each cell by LCZ A (Dense Trees) or LCZ B (Scattered Trees)

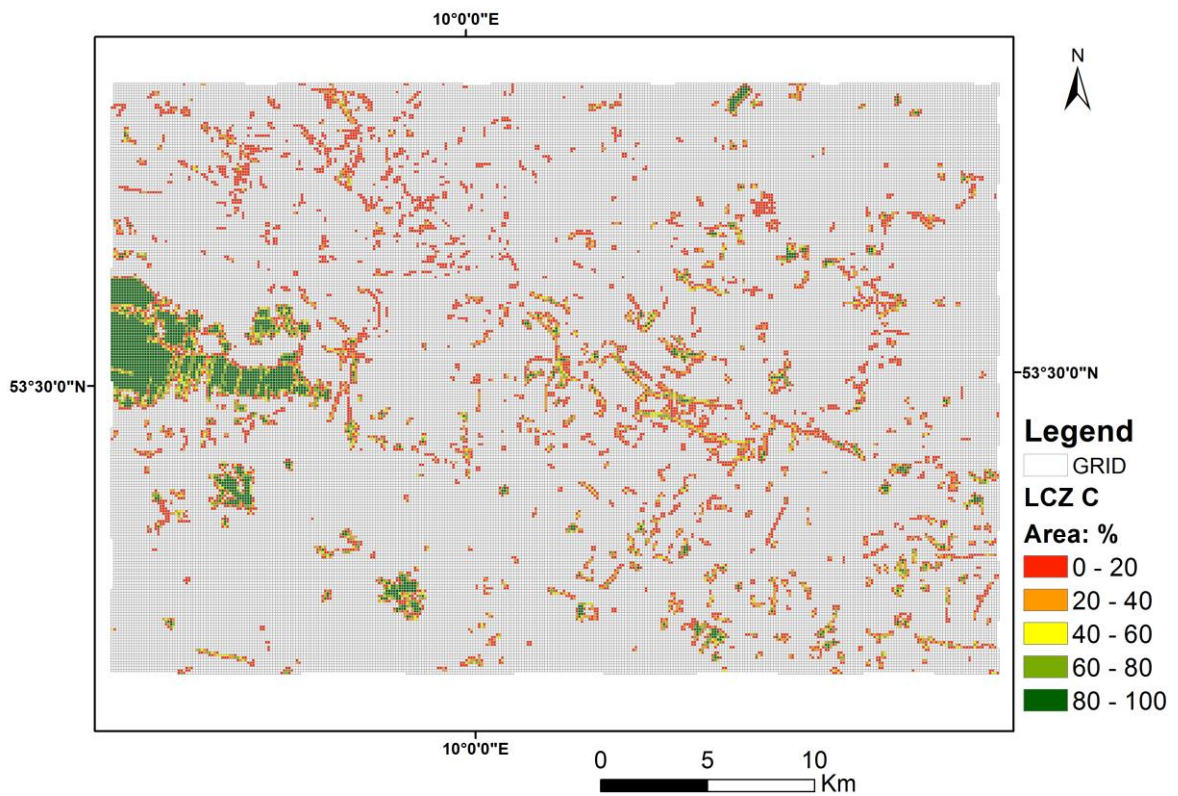


Figure 45: Results obtained from OSM for the percentage of area occupied in each cell by LCZ C (Bush, scrub)



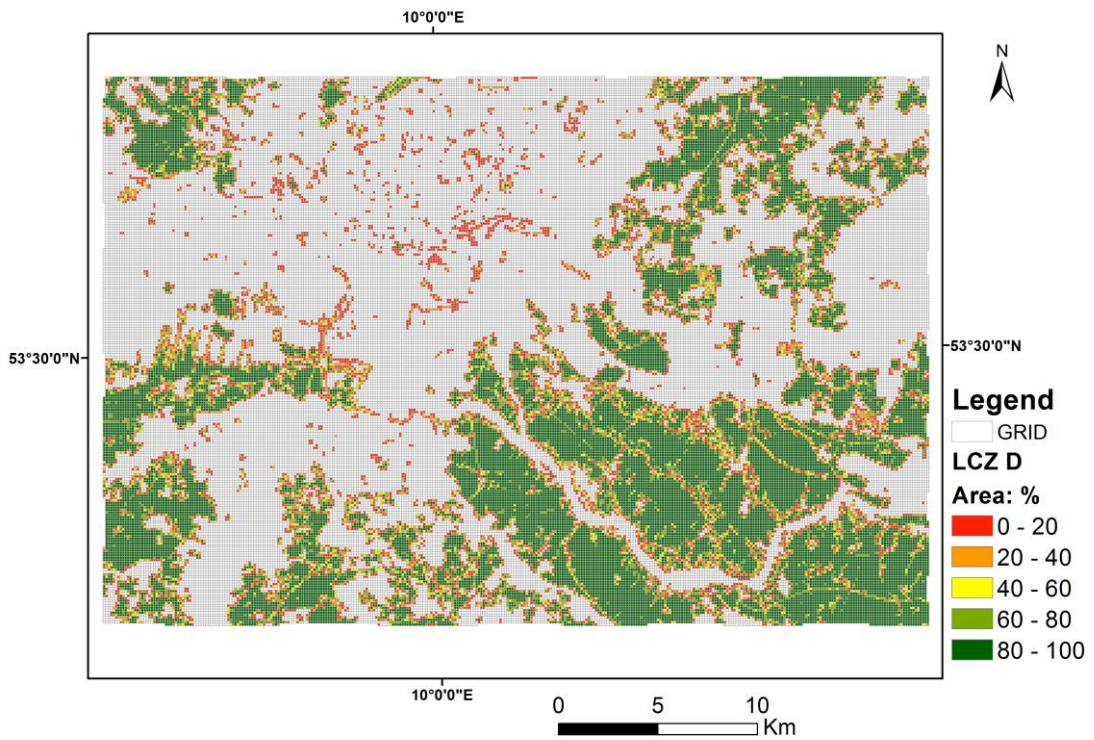


Figure 46: Results obtained from OSM for the percentage of area occupied in each cell by LCZ D (Low plants)

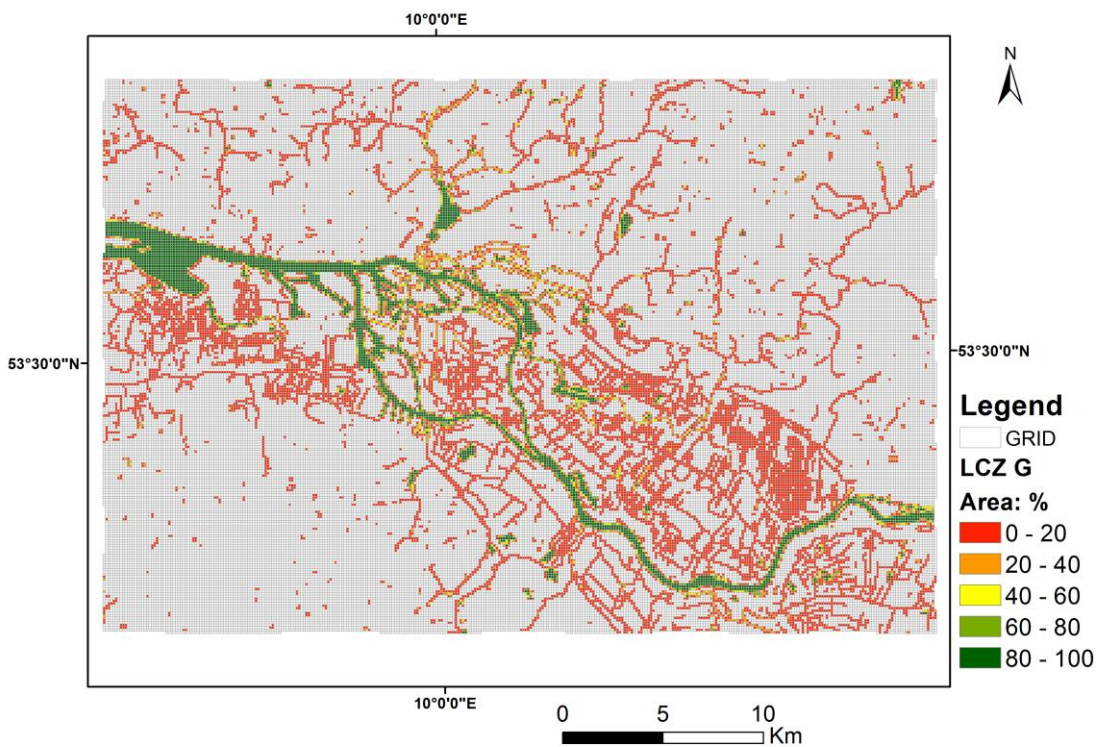
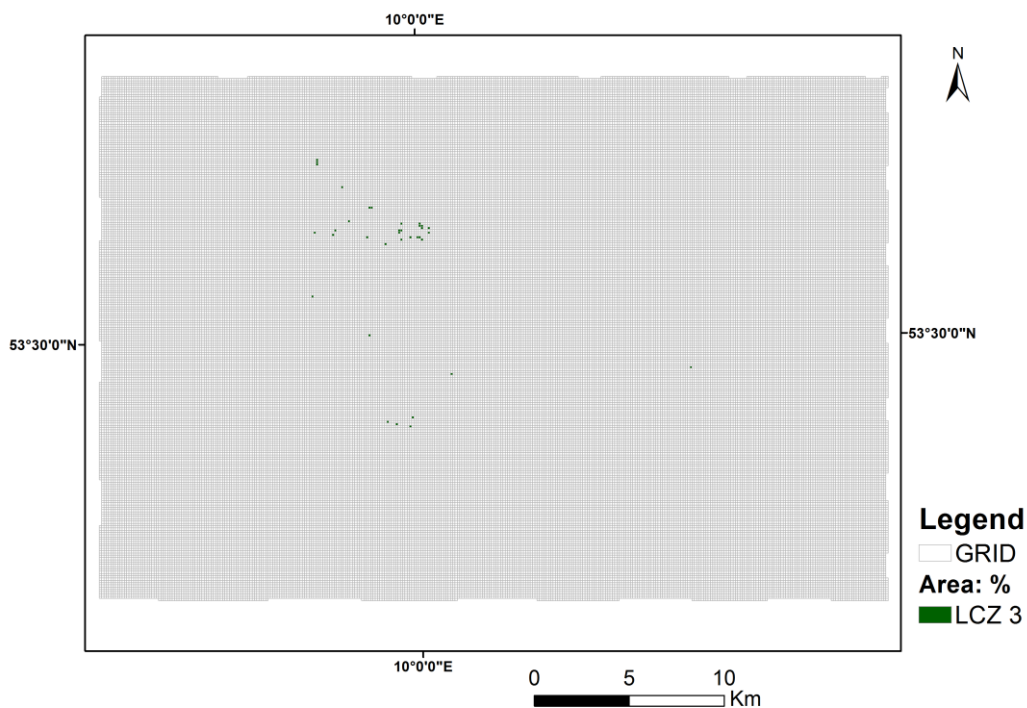


Figure 47: Results obtained from OSM for the percentage of area occupied in each cell by LCZ G (Water)

Relative to urban classes, the LCZ 1 (Compact high-rise) and LCZ 2 (Compact midrise) were not obtained for this study area. Classes LCZ 3 (Figure 48), LCZ 6 (Figure 51), LCZ 7 (Figure 52) and LCZ 9 (Figure 53) have a low percentage of occupation across the study area. In contrast, the class LCZ 4 (Figure 49), LCZ 5 (Figure 50) and LCZ 10 (Figure 54) have a large number of cells with high values of occupation. The LCZ 4 and LCZ 5 have many corresponding cells since the similarity between these classes is very similar, so they have similar selection criteria.

Overall, the visual comparison of the results obtained for the existing data with the Landsat true color image shows a great level of correspondence with the ground cover. In particular, for the natural LCZ classes (LCZ A or B, LCZ D and LCZ G), in comparison with the satellite image, a relevant correspondence with the land cover can be seen.

The confidence of a particular class ( $x$ ) was also extracted for each cell, depending on the number of times that the class appears in all four LCZs maps for the different seasons. This procedure was applied to all LCZ classes, but here only two examples are presented, in Figures 55 (LCZ A) and 56 (LCZ D). The results for the other classes are in Appendix O, in Figures 73 to 81.



**Figure 48: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 3 (Compact low-rise)**



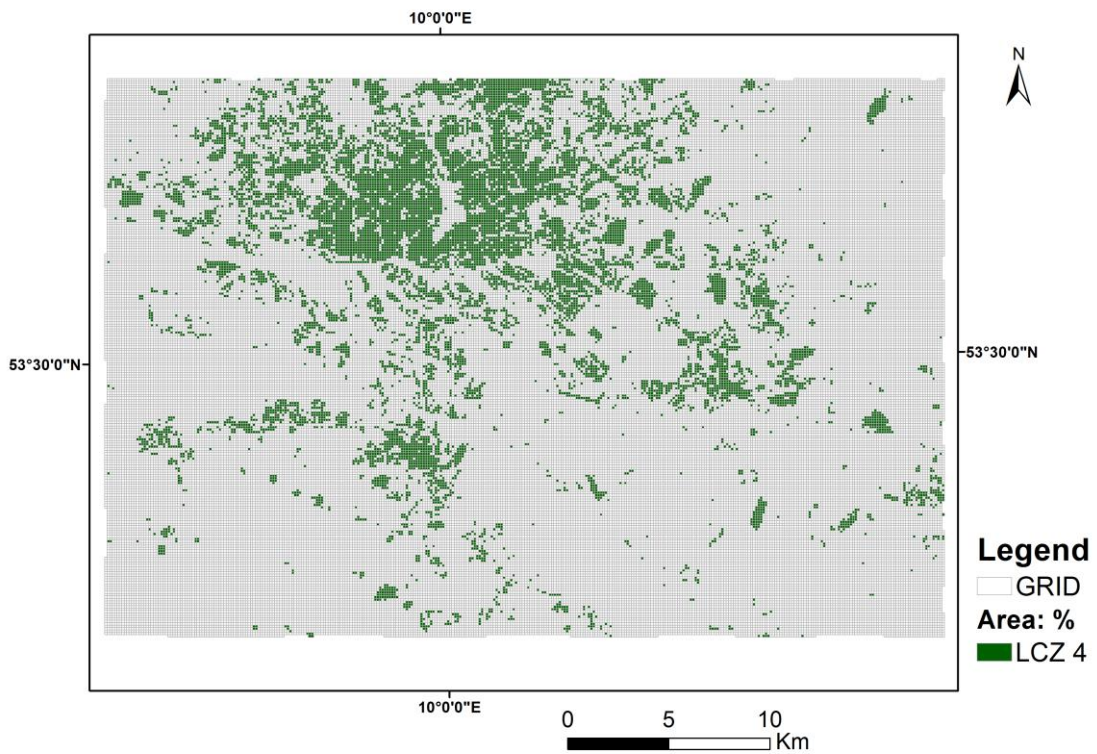


Figure 49: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 4 (Open high-rise)

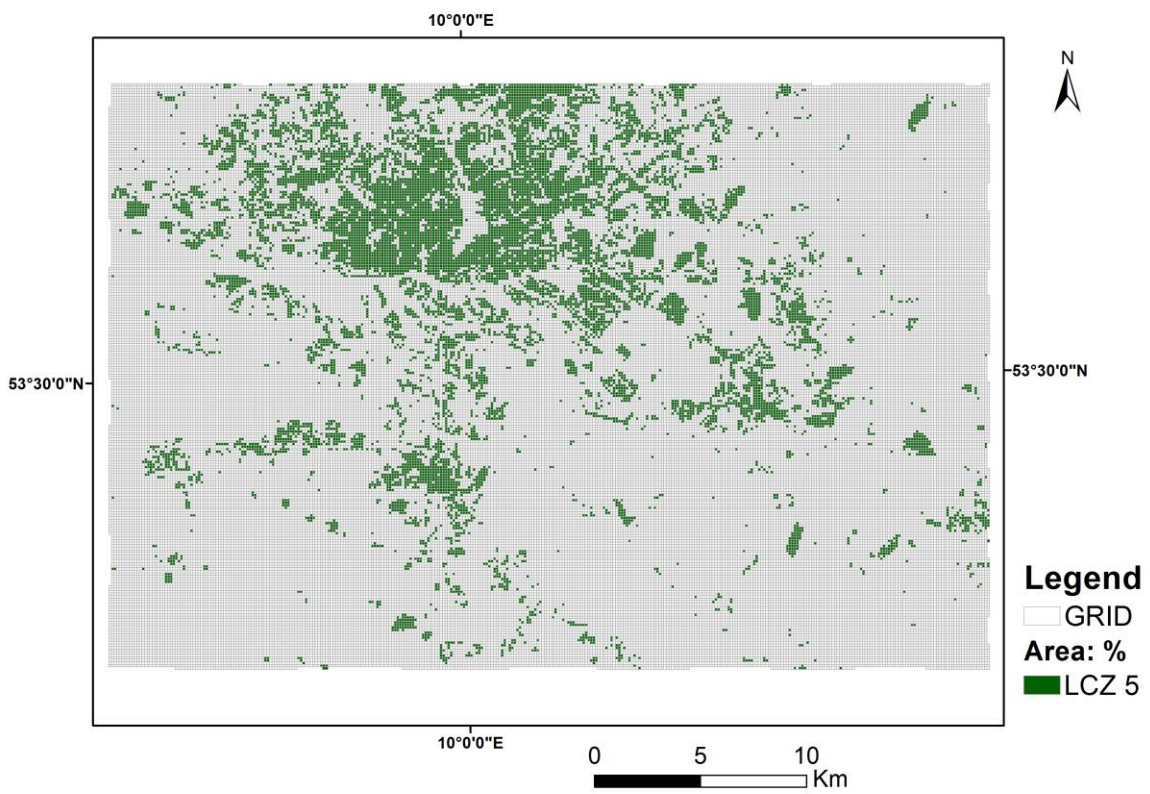


Figure 50: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 5 (Open mid-rise)

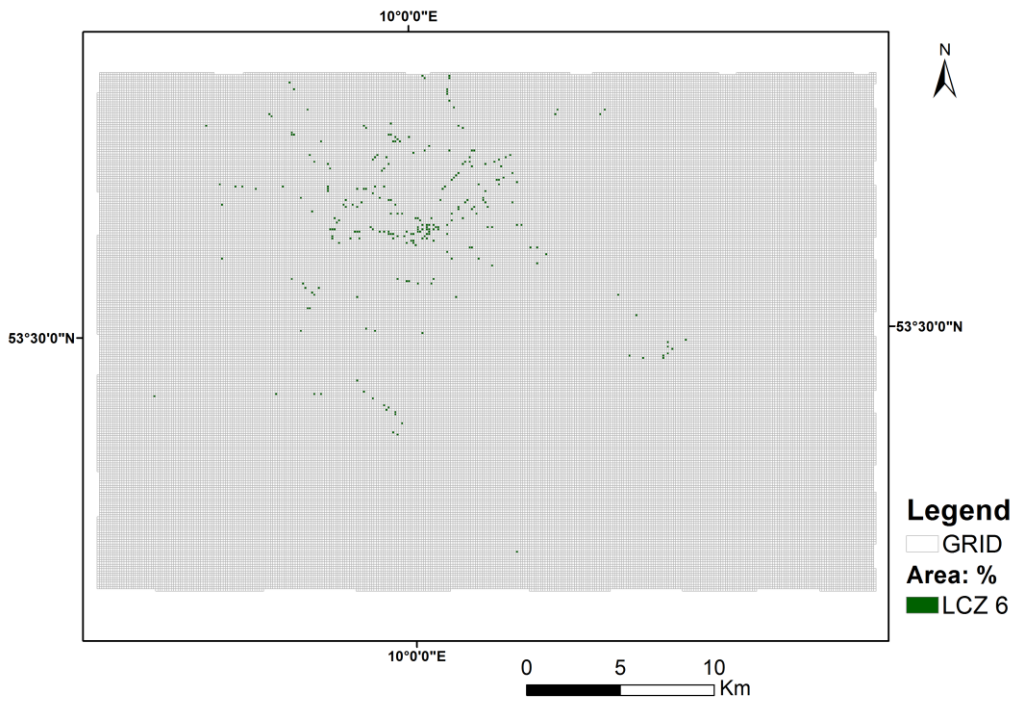


Figure 51: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 6 (Open low-rise)

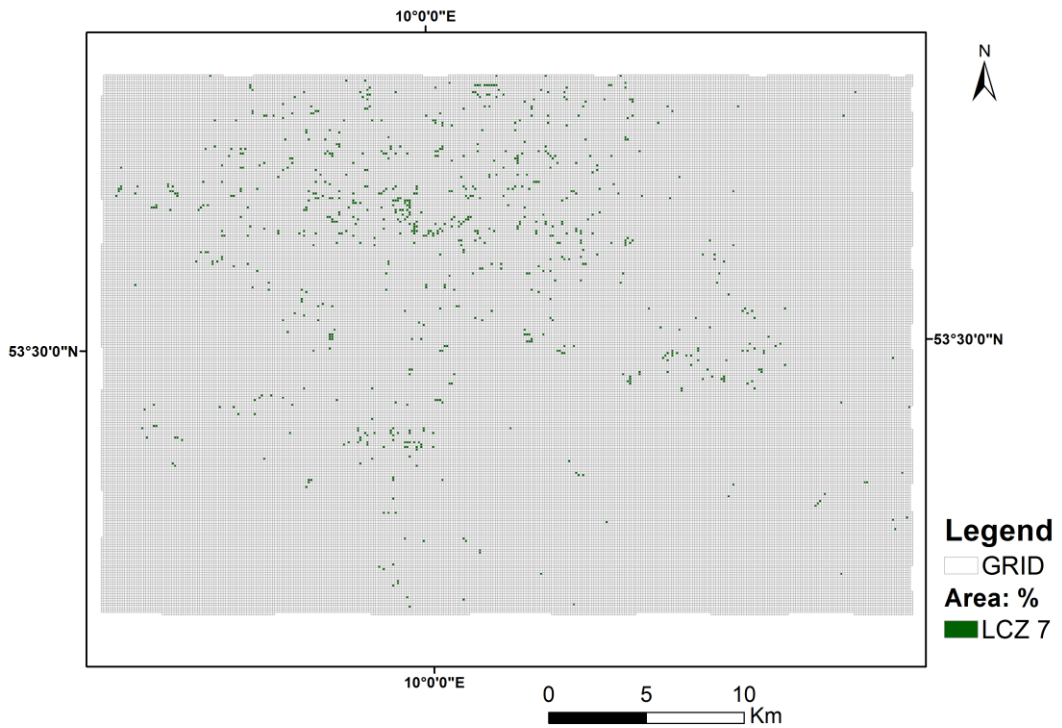


Figure 52: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 7 (Lightweight low-rise)



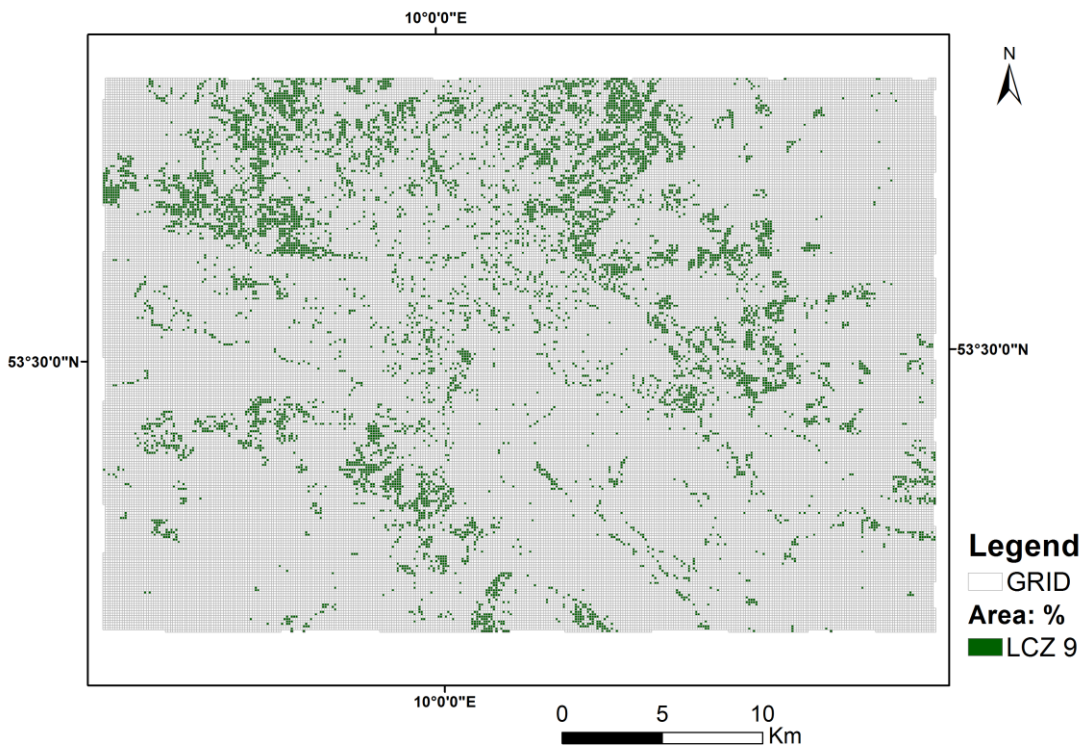


Figure 53: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 9 (Sparsely built)

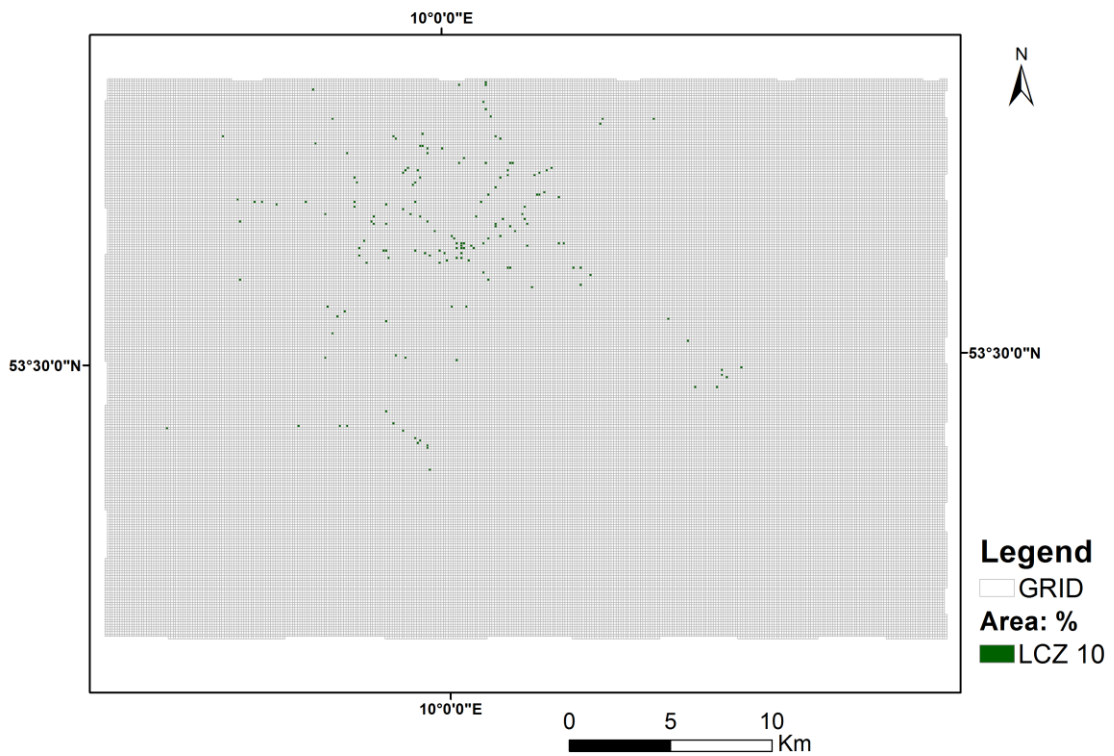


Figure 54: Results obtained from OSM for the percentage of area occupied in each cell by LCZ 10 (Heavy industry)

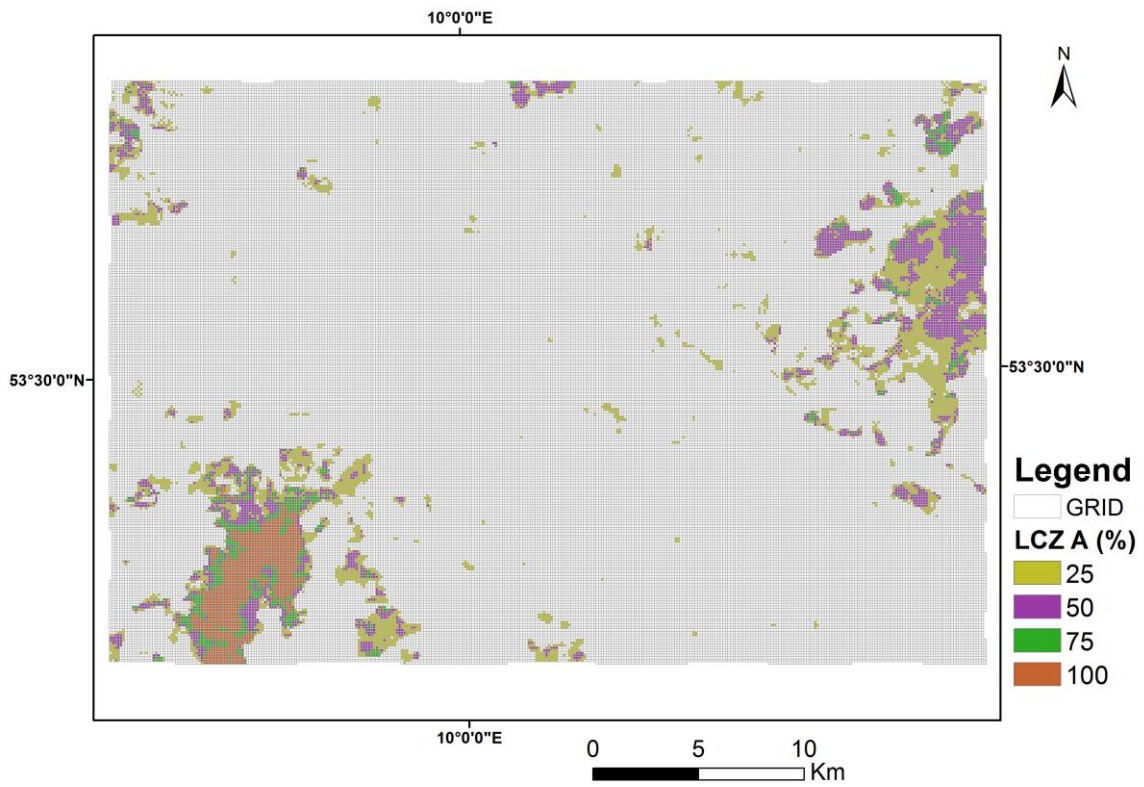


Figure 55: Confidence of the presence of LCZ A as a percentage for Hamburg

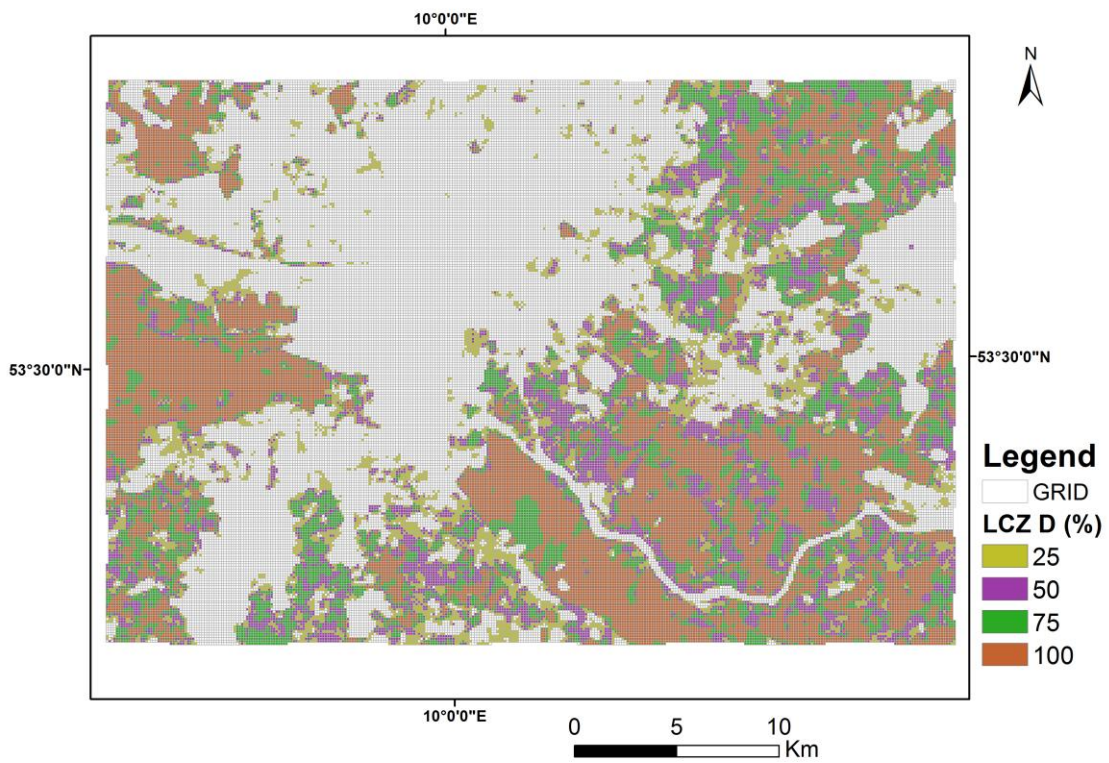


Figure 56: Confidence of presence of LCZ D as a percentage for Hamburg



## 4.4. Results of the data combination

### 4.4.1. Coimbra

The final result of the proposed methodology for Coimbra is shown in Figure 57, where the cells are filled with the most likely class. In the case where a predominant class does not exist, the cells are not filled. Thus, to evaluate the accuracy of these final maps and compare them with the validation dataset created previously, the “no data” class was filled with the data from the LCZ map using the 120 m Landsat 8 images for each season. Figures 57 to 61 represent the combination of the most likely class and the values for winter (Figure 57), autumn (Figure 58), summer (Figure 60) and spring images (Figure 61).

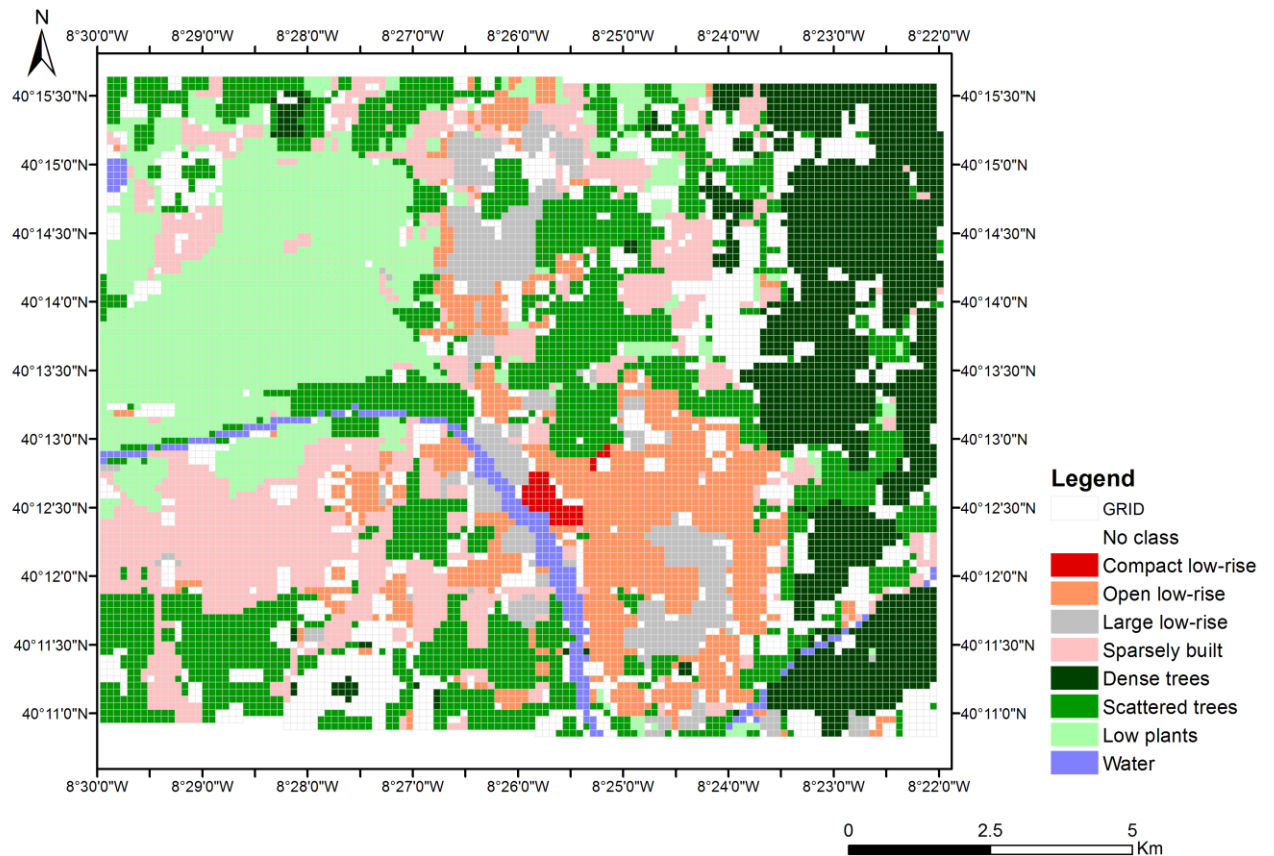


Figure 57: Result of the procedure applied to Coimbra (the most likely class)

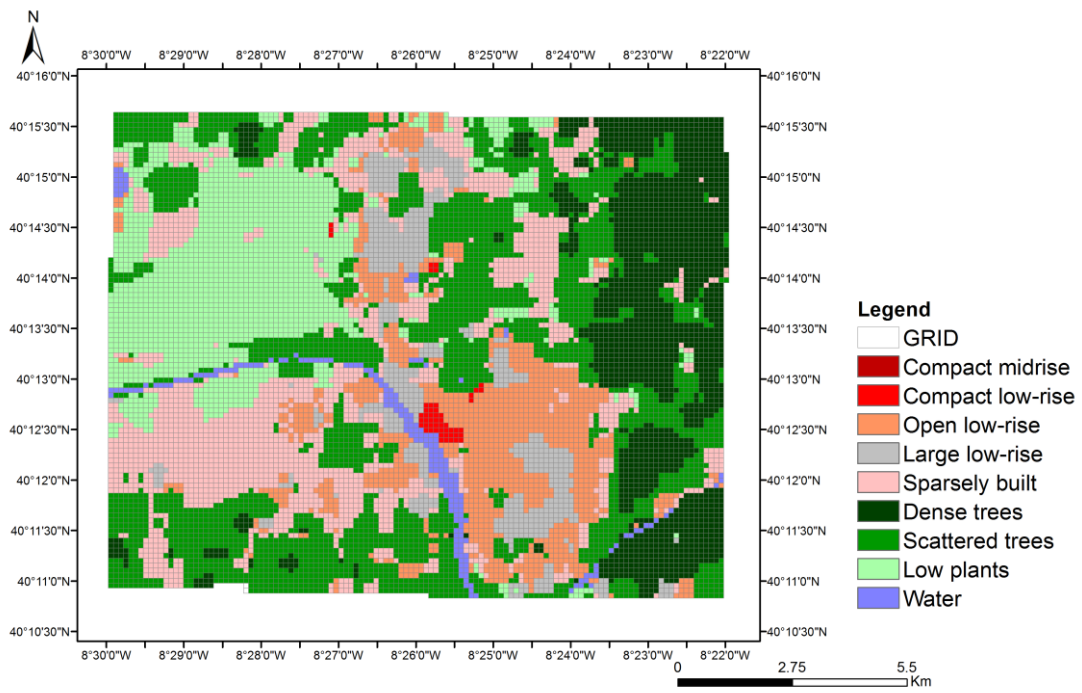


Figure 58: Result of the procedure applied to Coimbra (the most likely class) for the Winter LCZ map

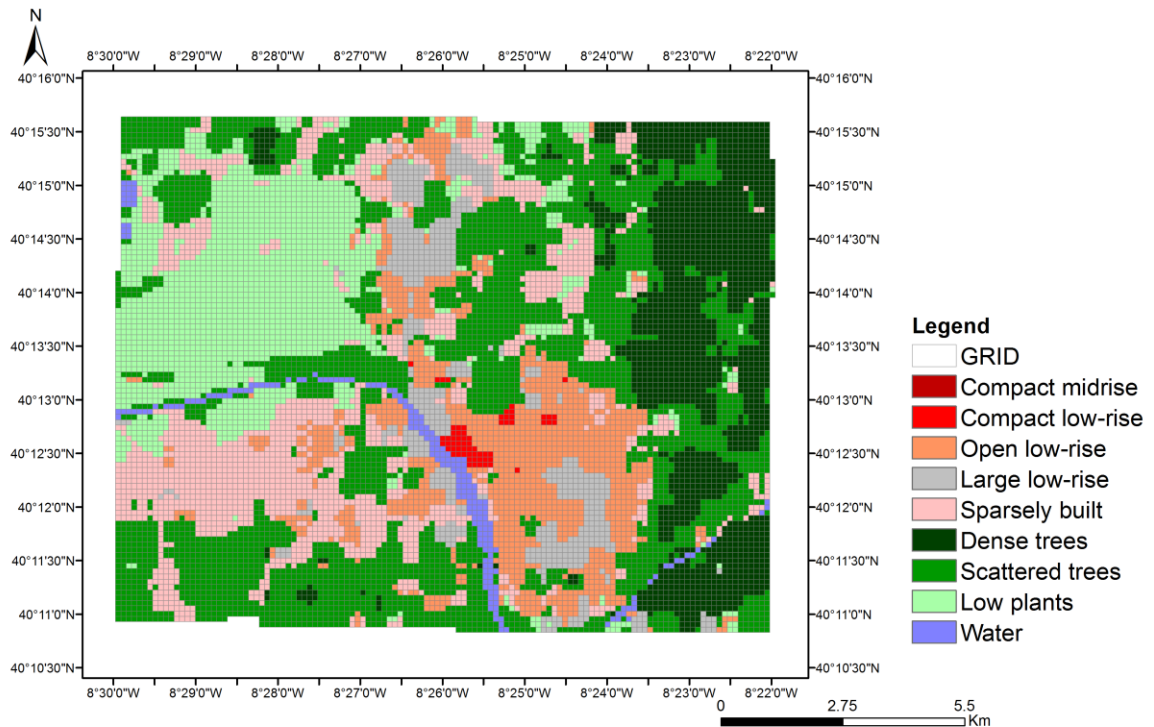


Figure 59: Result of the procedure applied to Coimbra (the most likely class) for the Autumn LCZ map

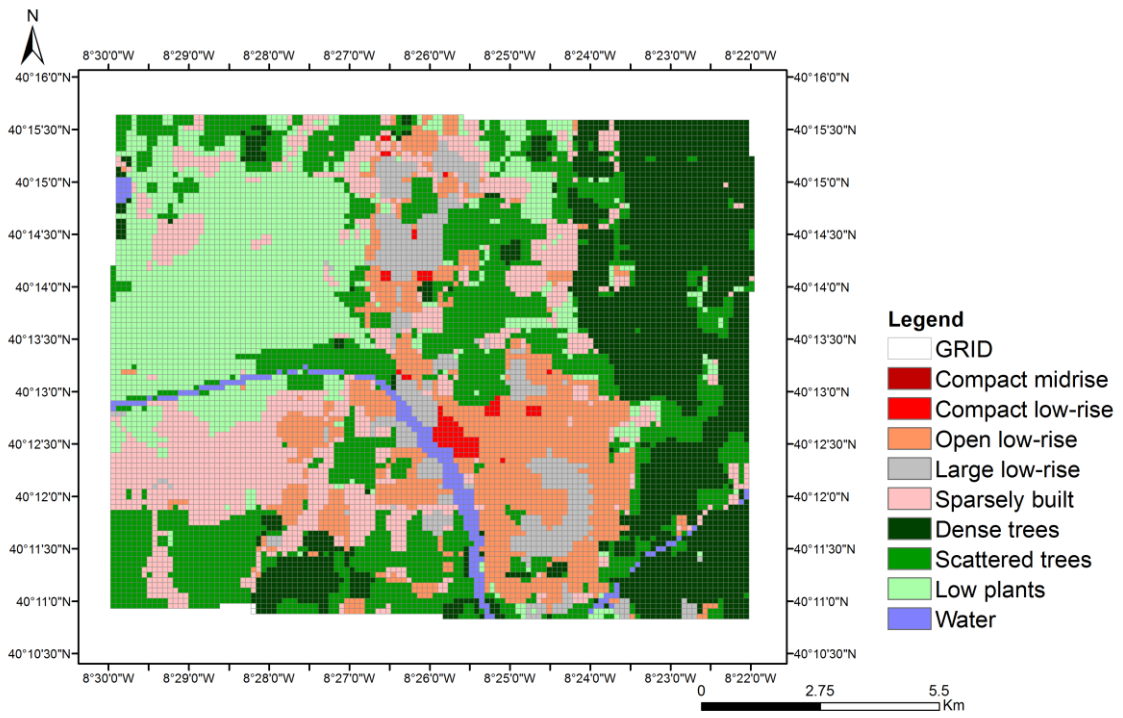


Figure 60: Results of the procedure applied to Coimbra (the most likely class) for the Summer LCZ map

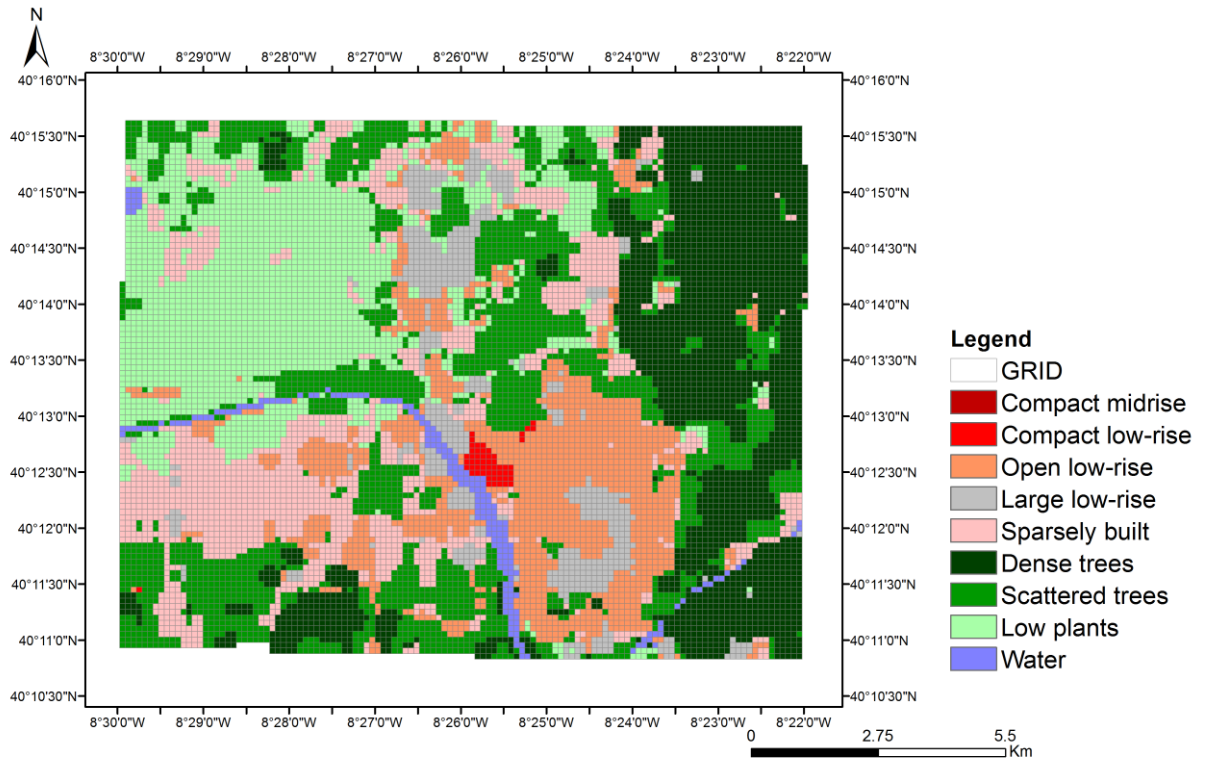


Figure 61: Result of the procedure applied to Coimbra (the most likely class) for the Spring LCZ map

The results show that the overall accuracy of all images for Coimbra improved by at least 7% (Table 26), where the spring image showed the largest increase (12%).

Tables 27 to 30 were generated for the classification of Coimbra using the combination of OSM data and the LCZ map from the classification using 120 m Landsat 8 data for winter, autumn, summer and spring, respectively.

**Table 26: Overall accuracy before and after the data combination to Coimbra**

	OA (Before)	OA (After)	Difference
<b>Winter</b>	42%	53%	+11%
<b>Autumn</b>	44%	53%	+9%
<b>Summer</b>	43%	50%	+7%
<b>Spring</b>	39%	51%	+12%

**Table 27: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the winter Landsat 8 image with a spatial resolution of 120 m.**

LCZ	2	3	6	8	9	A	B	D	G	∑	UA
3	0	102	12	0	0	1	0	0	0	115	89%
6	10	23	206	10	107	8	24	10	0	398	52%
8	18	3	53	89	44	3	17	2	2	231	39%
9	0	1	5	3	106	8	47	33	0	203	52%
A	0	0	0	0	9	73	166	8	3	259	28%
B	0	1	2	2	42	47	170	20	5	289	59%
D	0	0	2	1	19	8	30	179	5	244	73%
G	0	9	0	2	3	2	8	6	31	61	51%
∑	28	139	280	107	330	150	462	258	46	1800	
PA	0%	73%	74%	83%	32%	49%	37%	69%	67%	OA	53%

The LCZ map combining the data extracted from OSM and the winter satellite derived LCZ map has an overall accuracy of 53%. It can be concluded that this represents a significant improvement (11%) in the overall accuracy. A higher user's accuracy appears for classes LCZ 3 (Compact mid-rise) – (89%) and LCZ D (Low plants) - (73 %), while the classes LCZ A (Dense Trees) – 28% and LCZ 8 (Large low-rise) – 39% have a lower user's accuracy. The producer's accuracy is higher for classes LCZ 8 (83%) and LCZ 6 (Open low-rise) (74%) and lower for classes LCZ 2 (Compact mid-rise) (0%), LCZ 9 (Sparsely Built) (32%) and LCZ B (Scattered trees) (37%). There is a certain amount of confusion between the classes LCZ A and LCZ B. Comparing these results with the confusion matrix obtained before the procedure for the winter image, there is a

major improvement in the producer’s accuracy for LCZ 6 (from 29% to 74%). There was also a major improvement in the user’s accuracy for class LCZ G (36%).

**Table 28: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the autumn Landsat 8 image with a spatial resolution of 120 m.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
3	0	102	32	1	0	1	1	0	0	137	74%
6	19	23	197	14	110	8	25	6	0	402	49%
8	9	3	42	85	38	3	17	2	2	201	42%
9	0	1	5	2	99	6	35	24	0	172	58%
A	0	0	0	0	7	73	155	7	3	245	30%
B	0	1	2	2	52	49	187	30	5	328	57%
D	0	0	2	2	21	8	35	180	5	253	71%
G	0	9	0	1	3	2	7	9	31	62	50%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	0%	73%	70%	79%	30%	49%	40%	70%	67%	OA	53%

In the confusion matrix for the combined results of autumn, the classes with a higher user accuracy are LCZ 3 (Compact low-rise) – 74% and LCZ D (Low plants) - (71%) and it shows a lower user’s accuracy for classes LCZ A – Dense Trees (30%) and LCZ 8 – Large low-rise (42%). The producer’s accuracy is higher for classes LCZ 6 – Open low-rise (79%) and LCZ 3 (Compact low-rise) – (73%). There is also a certain amount of confusion between the classes LCZ 9 and LCZ 6. Comparing the results with the confusion matrix before the procedure for the autumn image, the major gain was in the producer’s accuracy for LCZ 6 (from 40% to 70%) and for LCZ D (from 47% to 70%). As in the previous image, the class that experienced a major improvement was the user’s accuracy for LCZ G (from 20% to 50%).

**Table 29: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the summer Landsat 8 image with a spatial resolution of 120 m.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
3	0	109	42	2	0	1	1	0	0	155	70%
6	11	16	179	12	123	7	26	8	0	382	47%
8	17	3	47	86	36	3	19	2	2	215	40%
9	0	1	5	3	94	7	34	22	0	166	57%
A	0	0	0	0	10	82	204	9	4	309	27%
B	0	1	2	2	41	37	132	23	4	242	55%
D	0	0	2	1	23	11	39	188	5	269	70%
G	0	9	3	1	3	2	7	6	31	62	50%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	0%	78%	64%	80%	28%	55%	29%	73%	67%	OA	50%

Table 29 shows an overall accuracy of 50%. A higher user's accuracy is obtained for classes LCZ D (Low plants) and LCZ 3 (Compact low-rise), both with 70%, while the classes LCZ A (Dense Trees) and LCZ 8 (Large low-rise) have a lower user's accuracy, with 27% and 40%, respectively.

The producer's accuracy is higher for classes LCZ 3 (78%), LCZ D (73%) and lower for classes LCZ 2 (0%), LCZ 8 (28%) and LCZ B (29%). There is also a certain amount of confusion between the classes LCZ A (Dense trees) and LCZ B (Scattered Trees). Comparing these results with the confusion matrix before the procedure for the summer image, a major gain in the producer's accuracy can be seen in LCZ 8 (from 59% to 80%) and in the user's accuracy for LCZ D (+33%).

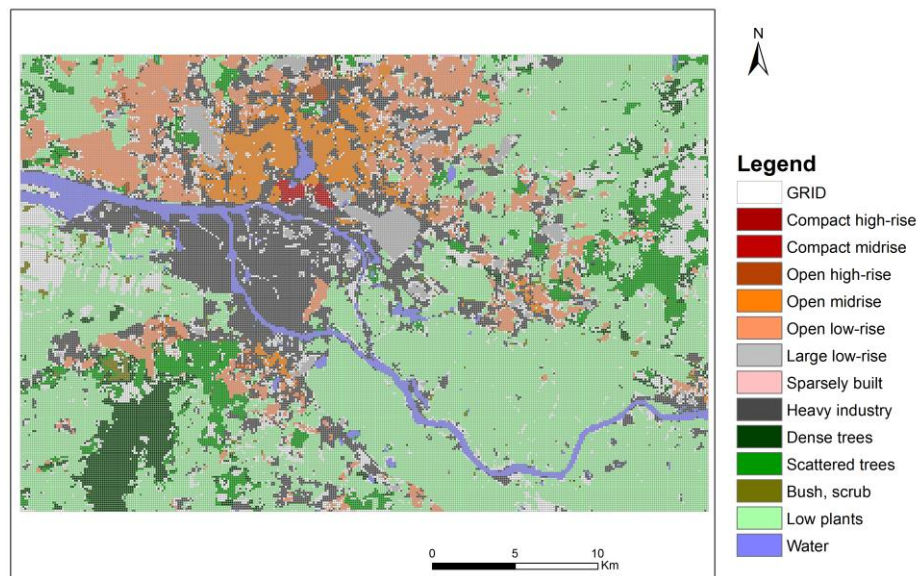
**Table 30: Confusion matrix for the final results after the data combination for Coimbra, using the OSM and LCZ map derived from the spring Landsat 8 image with a spatial resolution of 120 m.**

LCZ	2	3	6	8	9	A	B	D	G	$\Sigma$	UA
3	0	109	25	0	0	1	1	0	0	136	80%
6	20	16	206	13	121	9	26	11	0	422	49%
8	8	3	40	84	36	3	18	3	2	197	43%
9	0	1	4	2	92	7	36	23	0	165	56%
A	0	0	0	0	11	81	199	6	4	301	27%
B	0	1	2	2	39	37	130	21	4	236	55%
D	0	0	3	5	28	10	45	188	5	284	66%
G	0	9	0	1	3	2	7	6	31	59	53%
$\Sigma$	28	139	280	107	330	150	462	258	46	1800	
PA	0%	78%	74%	79%	28%	54%	28%	73%	67%	OA	51%

In the confusion matrix for the combined results of spring, the classes with a higher user's accuracy are LCZ 3 (Compact low-rise) – 80% and LCZ D (Low plants) - 66%, and it shows a lower user's accuracy for classes LCZ A – Dense Trees (27%) and LCZ 8 – Large low-rise (43%). The producer's accuracy is higher for classes LCZ 8 (79%) and LCZ 3 (Compact low-rise) – 78% and lower for classes LCZ 2 – Compact mid-rise (0%), and also LCZ B (Scattered Trees) and LCZ 9 – Scattered Trees, both with (28%). There is also a certain amount of confusion between the classes LCZ A and LCZ B. Comparing this results with the confusion matrix before the procedure for the spring image it appears that a major gain was achieved in the producer's accuracy for LCZ G (from 41% to 67%), LCZ 8 (from 55% to 79%) and in LCZ 6 (from 51% to 74%). For the user's accuracy, major gains were made in LCZ 3 (from 60 to 80%) and LCZ D (from 46% to 66%).

#### 4.4.2. Hamburg

The final result of the proposed methodology is shown in Figure 62, where the cells are filled with the most likely class. In the case where a predominant class does not exist, the cells were not filled. Figures 63 to 66 represent the combination of the most likely class and the values for winter (Figure 63), autumn (Figure 64), summer (Figure 65) and spring images (Figure 66).



**Figure 62: Result of the procedure applied to Hamburg (the most likely class)**



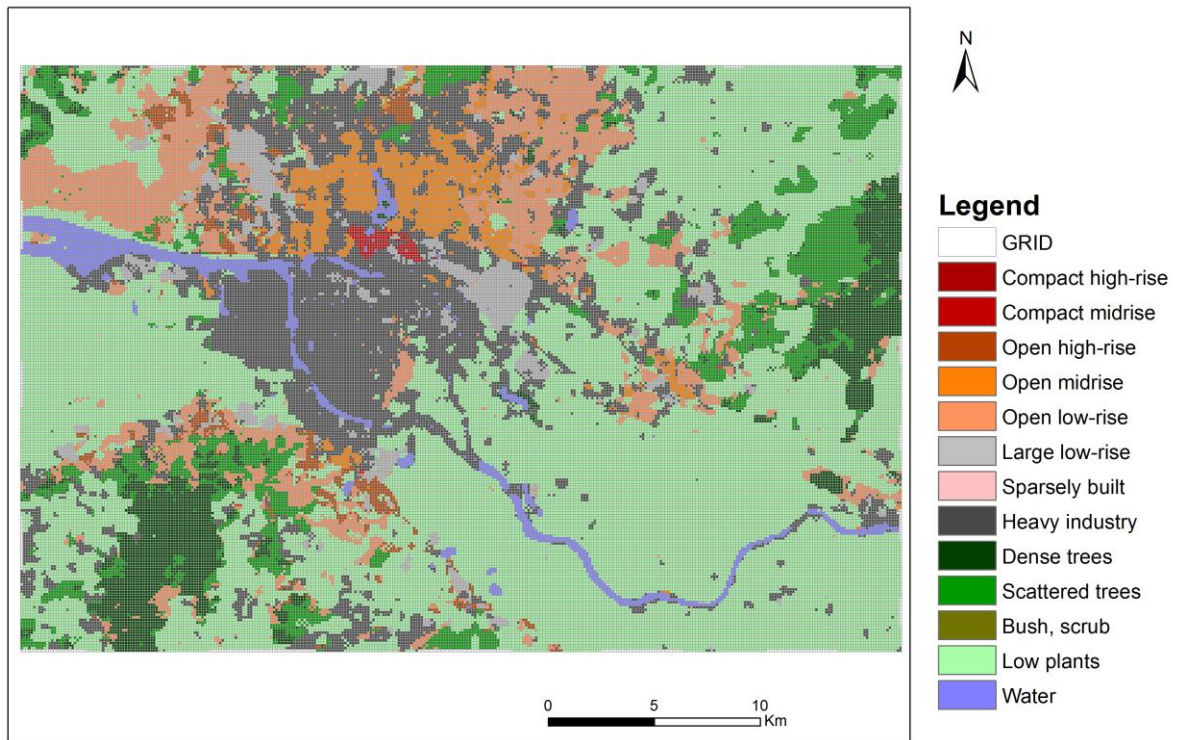


Figure 63: Result of the procedure applied to Hamburg (the most likely class) for the Winter LCZ map

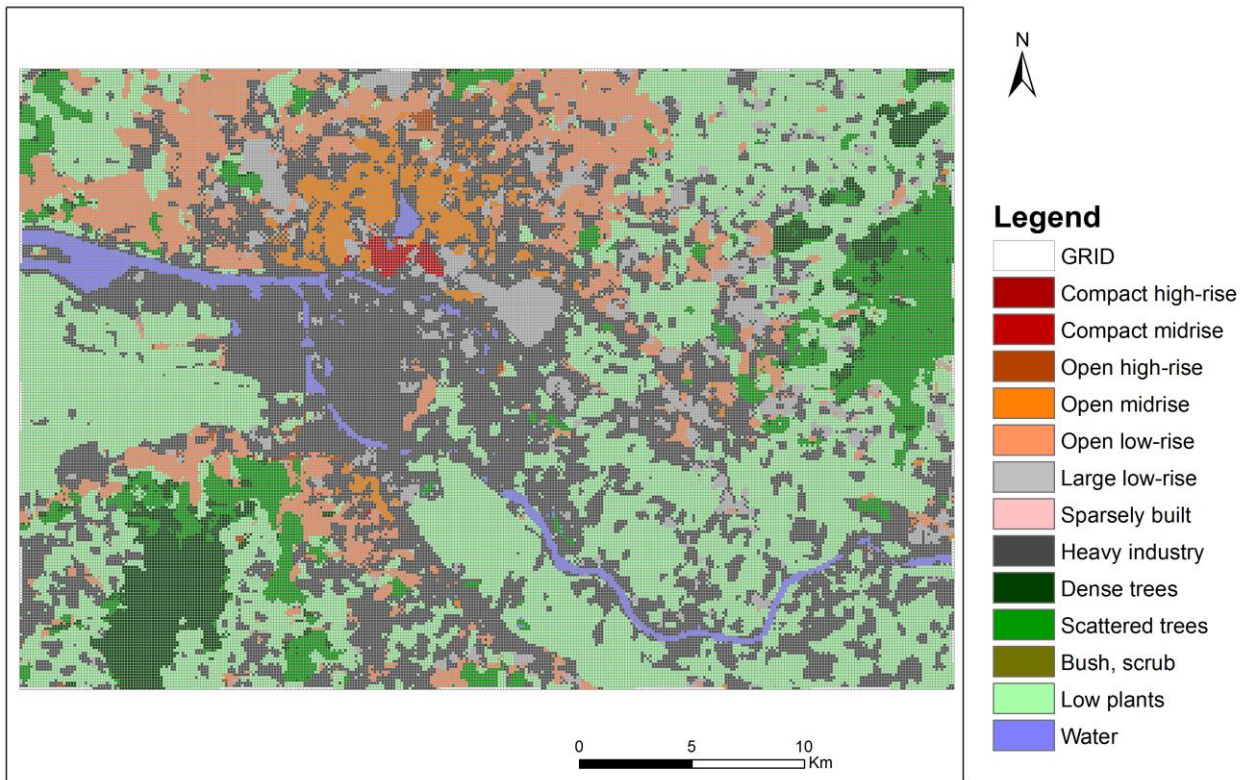


Figure 64: Results of the procedure applied to Hamburg (the most likely class) for the Autumn LCZ map



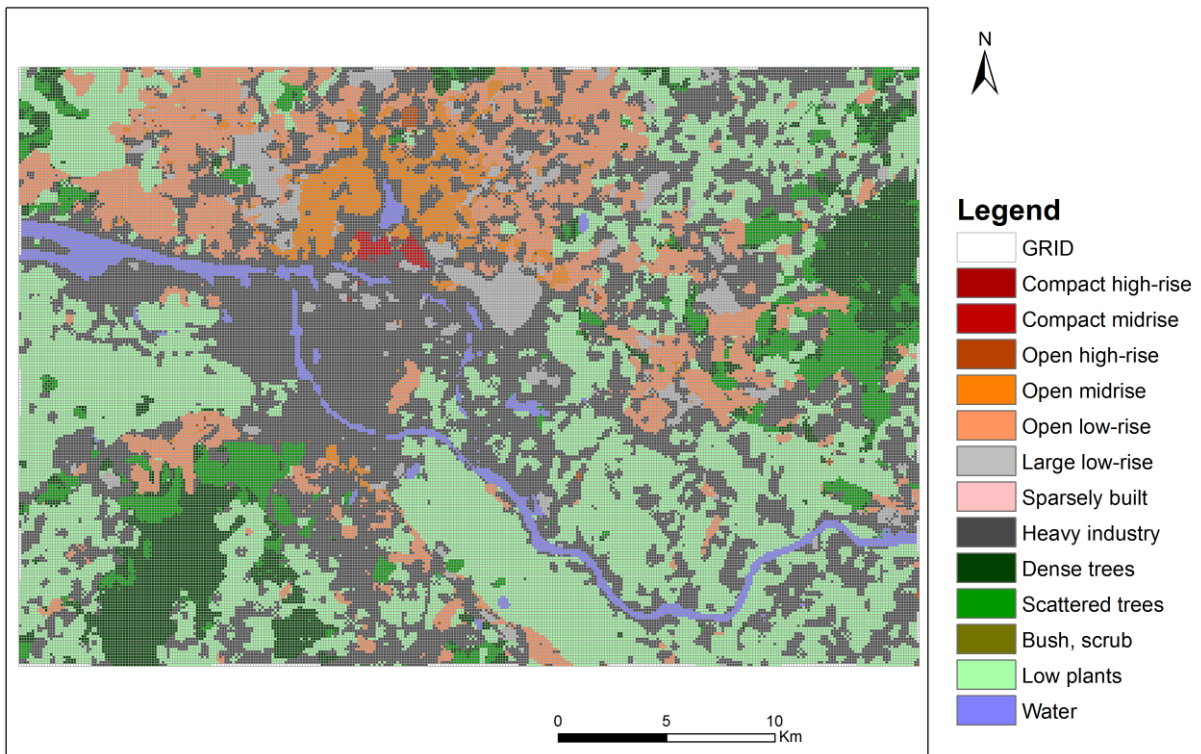


Figure 65: Result of the procedure applied to Hamburg (the most likely class) for the Summer LCZ map

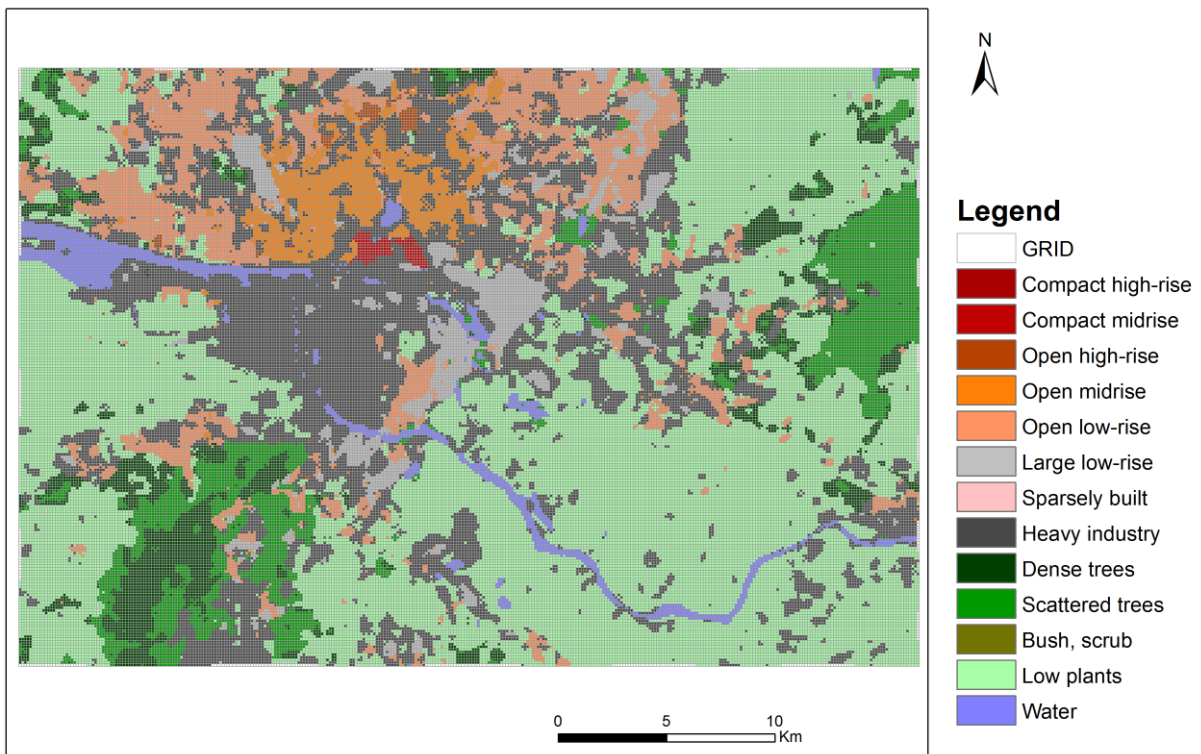


Figure 66: Result of the procedure applied to Hamburg (the most likely class) for the Spring LCZ map

For Hamburg, one of the final maps showed a decrease in the overall accuracy, namely, the winter image (-5%). Nevertheless, all other maps showed an increase in overall accuracy of at least of 9% (Table 31).

Tables 32 to 35 show the user accuracy (UA) and producer accuracy (PA) for each class obtained in both confusion matrices of 120 meters' LCZs maps (before and after the OSM procedure) and their differences in Hamburg.

**Table 31: Overall accuracy before and after the data combination was applied to Hamburg**

	(OA) Before	(OA) After	Difference
<b>Winter</b>	72%	67%	-5%
<b>Autumn</b>	56%	65%	+9%
<b>Summer</b>	55%	64%	+9%
<b>Spring</b>	53%	65%	+12%

**Table 32: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the winter Landsat 8 image with a spatial resolution of 120 m.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	∑	UA
1	173	18	0	0	0	0	0	0	0	0	0	0	191	91%
2	10	141	0	0	0	0	0	7	0	0	0	8	166	85%
4	0	1	33	3	15	1	2	0	2	5	1	0	63	52%
5	1	4	4	149	15	8	0	4	0	3	0	1	189	79%
6	0	0	49	8	163	3	0	0	5	18	2	0	248	66%
8	0	1	1	3	15	118	0	18	0	2	0	3	161	73%
9	0	0	1	0	4	0	0	0	0	0	0	0	5	0%
10	0	10	35	29	76	33	1	75	3	12	18	33	325	23%
A	0	0	0	0	4	0	0	0	144	40	7	0	195	74%
B	0	0	0	0	16	0	2	0	78	108	10	1	215	50%
C	0	0	0	0	0	0	0	0	1	1	2	0	4	0%
D	0	0	1	0	22	0	2	1	7	21	177	1	232	76%
G	0	1	0	0	0	0	1	3	0	1	2	198	206	96%
∑	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	94%	80%	27%	78%	49%	72%	0%	69%	60%	51%	81%	81%	AO	67%

The LCZ classification using the integrated values of the winter satellite LCZ map resulted in an overall accuracy of 67%, which represents a reduction of 5%. A higher user's accuracy is obtained for classes LCZ G (Water) – (96%) and LCZ 1 (Compact high-rise) - (91 %), while the classes LCZ 9 (Scattered trees), LCZ C (Bush, scrub) and LCZ 10 (Heavy industry) have lower user's accuracies, the first two both 0% and 23%, respectively. The producer's accuracy is higher for classes LCZ 1 (94%), LCZ D and LCZ G, both with 81% and lower for classes LCZ 9 (0%),

LCZ 4 (27%) and LCZ 6 (49%). There is also a certain amount of confusion between the classes LCZ 4 and LCZ 6, and also between LCZ A (Dense trees) and LCZ B (Scattered Trees). Comparing these results with the confusion matrix before the procedure for the winter image, it appears that the major loss in producer's accuracy was in LCZ 4 (from 90% to 27%).

**Table 33: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the autumn Landsat 8 image with a spatial resolution of 120 m.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	∑	UA
1	173	18	0	0	0	0	0	0	0	0	0	0	191	91%
2	10	140	0	0	0	0	0	0	0	0	0	8	158	89%
4	0	1	27	3	8	2	0	0	0	0	1	0	42	64%
5	1	4	5	148	14	8	0	4	0	3	0	1	188	79%
6	0		49	9	164	2	1	0	3	18	3	0	249	66%
8	0	1	2	3	12	102	0	21	0	2	0	4	147	69%
9	0	0	1	0	4	0	0	0	0	0	0	0	5	0%
10	0	11	40	29	92	49	1	79	6	15	19	37	378	21%
A	0	0	0	0	4	0	1	0	120	38	7	0	170	71%
B	0	0	0	0	11	0	1	0	101	112	9	1	235	48%
C	0	0	0	0	0	0	0	0	1	1	2	0	4	0%
D	0	0	0	0	21	0	3	1	9	21	176	1	232	76%
G	0	1	0	0	0	0	1	3	0	1	2	193	201	96%
∑	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	94%	80%	22%	77%	50%	63%	0%	73%	50%	53%	80%	79%	AO	65%

In the confusion matrix for the combined results of autumn, the higher user accuracy classes are LCZ G (Water) – 96% and LCZ 1 (Compact high-rise): (91%) and it shows a lower user accuracy for classes LCZ 9 – Sparsely Built and LCZ C - Bush, scrub (both with 0%) and LCZ 10 – Heavy Industry (21%). The producer's accuracy is higher for classes LCZ 1 (94%) and LCZ 2 (Compact mid-rise) and LCZ D - (both with 80%) and lower for classes LCZ 9 – Sparsely Built (0%), and LCZ 4 – Open mid-rise (22%). It also appears a certain amount of confusion between the classes LCZ 6 and LCZ 10. Comparing this results with the confusion matrix before the procedure for the autumn image it appears that the major gain in producer accuracy was in LCZ G (from 54% to 79%) and in LCZ 5 (from 54% to 77%) and in user accuracy was in LCZ 4 (from 3% to 64%).

**Table 34: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the summer Landsat 8 image with the spatial resolution of 120 m.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	173	18	0	0	0	0	0	0	0	0	0	0	191	91%
2	10	127	0	0	0	1	0	4	0	0	0	5	147	86%
4	0	1	27	2	5	1	0	0	0	0	0	0	36	75%
5	1	4	6	146	13	9	0	4	0	3	0	0	186	78%
6	0	0	52	10	160	3	1	0	2	19	2	0	249	64%
8	0	1	2	3	10	92	0	19	0	0	1	4	132	70%
9	0	0	1	0	4	0	0	0	0	0	0	0	5	0%
10	0	24	36	31	106	57	1	77	5	16	17	39	409	19%
A	0	0	0	0	4	0	1	0	135	59	9	0	208	65%
B	0	0	0	0	10	0	1	0	90	90	9	1	201	45%
C	0	0	0	0	0	0	0	0	1	1	2	0	4	0%
D	0	0	0	0	18	0	3	1	7	21	177	1	228	78%
G	0	1	0	0	0	0	1	3	0	2	2	195	204	96%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	94%	72%	22%	76%	48%	56%	0%	71%	56%	43%	81%	80%	AO	64%

The final results after the data combination for Hamburg, using OSM and LCZ's map summer Landsat 8 image with the spatial resolution of 120 m combination shows an overall accuracy of 64%, that represent an improvement of 9%. Similarly, to the previous results, the higher user's accuracy it appears for classes LCZ G (Water) – (96%) and LCZ 1 (Compact high-rise) - (91%), while the classes LCZ 9 (Scattered trees), LCZ C (Bush, scrub) and LCZ 10 (Heavy industry) has the lower user's accuracy, first two with 0% and the third with 19%.

The producer's accuracy is higher for classes LCZ 1 (94%), LCZ D and LCZ G, with 81% and 80%, respectively and lower for classes LCZ 9 (0%), LCZ 4 (22%) and LCZ B (43%). It also appears a lot of confusion between the classes LCZ 6 and LCZ 10, and also between LCZ A (Dense trees) and LCZ B (Scattered Trees). Comparing this results with the confusion matrix before the procedure for the summer image it appears that the major's gains in producer accuracy was in LCZ D (from 57% to 81%) and in LCZ 4 (from 53% to 76%) and in user accuracy was LCZ 4 (from 3% to 75%).

**Table 35: Confusion matrix for the final results after the data combination for Hamburg, using the OSM and LCZ map derived from the spring Landsat 8 image with a spatial resolution of 120 m.**

LCZ	1	2	4	5	6	8	9	10	A	B	D	G	$\Sigma$	UA
1	173	18	0	0	0	0	0	0	0	0	0	0	191	91%
2	10	126	0	0	0	0	0	0	0	0	0	5	141	89%
4	0	1	27	2	6	1	0	0	1	0	0	0	38	71%
5	1	4	5	149	14	9	0	4	0	3	0	0	189	79%
6	0	0	50	8	152	3	0	0	1	16	2	0	232	66%
8	0	1	2	4	10	93	0	23	0	0	1	3	137	68%
9	0	0	1	0	4	0	0	0	0	0	0	0	5	0%
10	0	25	36	29	96	57	2	77	2	17	17	36	394	20%
A	0	0	0	0	7	0	0	0	130	34	7	0	178	73%
B	0	0	0	0	10	0	3	0	96	115	10	1	235	49%
C	0	0	0	0	0	0	0	0	1	1	2	0	4	0%
D	0	0	3	0	31	0	2	1	9	23	178	1	248	72%
G	0	1	0	0	0	0	1	3	0	2	2	199	208	96%
$\Sigma$	184	176	124	192	330	163	8	108	240	211	219	245	2200	
PA	94%	72%	22%	78%	46%	57%	0%	71%	54%	55%	81%	81%	AO	65%

In the confusion matrix for the combined results for spring, the classes with a higher user's accuracy are LCZ G (Water) – 96% and LCZ 1 (Compact high-rise) - (91%) while a lower user's accuracy is obtained for classes LCZ 9 – Sparsely Built and LCZ C - Bush, scrub (both with 0%) and LCZ 10 – Heavy Industry (20%). The producer's accuracy is higher for classes LCZ 1 (94%), LCZ D (Low plants) and LCZ G - (both with 81%) and lower for classes LCZ 9 – Sparsely Built (0%), and LCZ 4 – Open mid-rise (22%). There is also a certain amount of confusion between the classes LCZ 4 and LCZ 10. Comparing this results with the confusion matrix before the procedure for the autumn image, it appears that the major gain in the producer's accuracy was in LCZ 5 (from 54% to 78%) and in LCZ G (from 60% to 81%). Similarly, with the previous results from the other images, the major gain in user's accuracy was in LCZ 4 (from 4% to 71%).

## 5. Conclusions

It can be concluded that the objectives of this research have been achieved, i.e. the procedure developed here has been shown to improve the accuracy of the LCZ maps in a consistent manner. Moreover, it was shown that VGI can be an important source of spatial data although it is based on data collected by volunteers rather than scientists so there are limitations in the use of the data for scientific studies. Nevertheless, the use of the same validation points (200 per class) for the results, before and after the procedure was applied, shows a clear improvement in the accuracy of the results, validating the importance of VGI.

Although the methodology proposed here obtained good results, as in any application in the context of this work, there were some difficulties encountered, especially in the urban LCZ classes. Some of these difficulties are related to the lack of OSM data related to buildings and to the intervals of the values of the properties that distinguish the LCZ classes, such as building surface fraction and impervious surface fraction. Some of this information is not present in OSM, e.g. information on the number of floors, which is one of the main characteristics that distinguishes between the diverse types of urban LCZ classes.

The LCZ classes that represent natural land cover types experienced the greatest improvements in producer accuracy, namely, LCZ D (Low plants), which showed improvements of 8% in all images, for the both study areas. The Water (LCZ G) class also showed improvements in the majority of images, where the improvement was greater than 10% in both cases. In general, the largest improvement in producer accuracy was in class LCZ 6 (Open low-rise) in Coimbra and in class LCZ 5 (Open mid-rise) in Hamburg.

To evaluate the final results after the procedure was implemented, a comparison could be made with the results currently present in WUDAPT (<http://www.wudapt.org/>), which have been developed using the workflow presented in Bechtel, Alexander, et al. (2015) and See et al. (2015), which could be done as part of future research. However, in line with the philosophy of WUDAPT, it should be easy to implement and freely available for WUDAPT participants to try out.

As a proposal for future work, this procedure can be developed in open source software and be made freely available to all WUDAPT contributors or a webpage could be developed that allows the citizens to submit their LCZ maps and download the OSM data for their region of interest.

The application could also allow users to make choices on some of their parameters regarding the OSM data available, or even use OSM data as training data input for the LCZs classifications.

It will also be interesting to test this methodology in more areas of study and to evaluate the results. It may also be interesting to study the differences obtained if a historical series of satellite images (for each season of the year) and the historical records corresponding to OSM data were compared in order to understand the evolution of the LCZ classes and hence changes in the land use and cover within urban areas.

## 6. References

- Alba, German. 2014. "Remote Sensing Classification. Algorithms Analysis Applied to Land Cover Change." Argentina: Mario Gulich Institute, CONAE.
- Alexander, Paul, and Gerald Mills. 2014. "Local Climate Classification and Dublin's Urban Heat Island." *Atmosphere* 5 (4): 755–74. doi:10.3390/atmos5040755.
- Aquilué, Núria, Miquel De Cáceres, Marie-Josée Fortin, Andrew Fall, and Lluís Brotons. 2017. "A Spatial Allocation Procedure to Model Land-Use/land-Cover Changes: Accounting for Occurrence and Spread Processes." *Ecological Modelling* 344: 73–86. doi:10.1016/j.ecolmodel.2016.11.005.
- "ArcGIS Desktop." 2015. Accessed June 8.  
<http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/009z00000054000000.htm>.
- Arnfield, A John. 1982. "AN APPROACH TO THE ESTIMATION OF THE SURFACE RADIATIVE PROPERTIES AND RADIATION BUDGETS OF CITIES." *Physical Geography* 3 (2). Taylor & Francis: 97–122. doi:10.1080/02723646.1982.10642221.
- Arsanjani, Jamal, Marco Helbich, Mohamed Bakillah, Julian Hagenauer, and Alexander Zipf. 2013. "Toward Mapping Land-Use Patterns from Volunteered Geographic Information." *International Journal of Geographical Information Science* 27 (12). Taylor & Francis: 2264–78. doi:10.1080/13658816.2013.800871.
- Auer, August H. 1978. "Correlation of Land Use and Cover with Meteorological Anomalies." *Journal of Applied Meteorology* 17 (5): 636–43.
- Bechtel, B., P. Alexander, J. Böhner, J. Ching, O. Conrad, J. Feddema, G. Mills, L. See, and I. Stewart. 2015. "Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities." *ISPRS International Journal of Geo-Information* 4 (1): 199–219.
- Bechtel, B., and C. Danek. 2012. "Classification of Local Climate Zones Based on Multiple Earth Observation Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (4): 1191–1202. doi:10.1109/JSTARS.2012.2189873.
- Bechtel, B., M. Foley, G. Mills, J. Ching, L. See, P. Alexander, M. O'Connor, et al. 2015. "CENSUS of Cities: LCZ Classification of Cities (Level 0) – Workflow and Initial Results from Various Cities." In *ICUC9 - 9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment*. Toulouse, France. doi:10.13140/RG.2.1.4028.5206.
- Bhar, Lalmohan. 2014. *Non-Parametric Tests*. Edited by Lal Mohan Bhar and VK Bhatia Rajender Parsad, VK Gupta. E-book on Advances in Data Analytical Techniques.  
[http://link.springer.com/10.1007/978-94-007-1211-9\\_4](http://link.springer.com/10.1007/978-94-007-1211-9_4).
- Bonham-Carter, Graeme. 2014. "Geographic Information Systems for Geoscientists : Modelling with GIS." In *Computer Methods in the Geosciences*, 416. Elsevier.



- Bono, Flavio, and Eugenio Gutiérrez. 2011. "A Network-Based Analysis of the Impact of Structural Damage on Urban Accessibility Following a Disaster: The Case of the Seismically Damaged Port Au Prince and Carrefour Urban Road Networks." *Journal of Transport Geography* 19 (6): 1443–55. doi:10.1016/j.jtrangeo.2011.08.002.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1). Kluwer Academic Publishers: 5–32. doi:10.1023/A:1010933404324.
- Castelein, Watse, Lukasz Grus, Joep Cromptvoets, and Arnold Bregt. 2010. "A Characterization of Volunteered Geographic Information." In *Proceedings of the 13th AGILE International Conference on Geographic Information Science 2010 | 13th AGILE International Conference on Geographic Information Science 2010 | 10/05/2010 - 14/05/2010 | Guimarães, Portugal*, 1–10. Guimarães: E.T.S.I. en Topografía, Geodesia y Cartografía (UPM).
- Chandler, T. J. 1965. "The Climate of London." *Quarterly Journal of the Royal Meteorological Society* 92 (392). John Wiley & Sons, Ltd.
- Ching, J., M. Brown, T. McPherson, S. Burian, F. Chen, R. Cionco, A. Hanna, et al. 2009. "National Urban Database and Access Portal Tool." *Bulletin of the American Meteorological Society* 90 (8): 1157–68. doi:10.1175/2009BAMS2675.1.
- Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, and J. Böhner. 2015. "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4." *Geoscientific Model Development* 8 (7). Copernicus GmbH: 1991–2007. doi:10.5194/gmd-8-1991-2015.
- Cortijo, F., and N. Blanca. 1996. "Image Classification Using Non-Parametric Classifiers and Contextual Information." *International Archives of Photogrammetry and Remote Sensing XXXI*: 120–24.
- Danylo, O., L. See, B. Bechtel, D. Schepaschenko, and S. Fritz. 2016. "Contributing to WUDAPT: A Local Climate Zone Classification of Two Cities in Ukraine." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (5): 1841–53. doi:10.1109/JSTARS.2016.2539977.
- Direção-Geral do Território. 2016. "Carta Administrativa Oficial de Portugal - Versão 2016." [http://www.dgterritorio.pt/cartografia\\_e\\_geodesia/cartografia/carta\\_administrativa\\_oficial\\_de\\_portugal\\_\\_caop\\_/caop\\_\\_download\\_/carta\\_administrativa\\_oficial\\_de\\_portugal\\_\\_versao\\_2016/](http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal__caop_/caop__download_/carta_administrativa_oficial_de_portugal__versao_2016/).
- Dixon, Barnali, and Venkatesh Uddameri. 2016. *GIS and Geocomputation for Water Resource Science and Engineering*.
- Ellefsen, Richard. 1991. "Mapping and Measuring Buildings in the Canopy Boundary Layer in Ten U.S. Cities." *Energy and Buildings* 16 (3–4): 1025–49.
- Emmanuel, R., and E. Krüger. 2012. "Urban Heat Island and Its Impact on Climate Change Resilience in a Shrinking City: The Case of Glasgow, UK." *Building and Environment* 53: 137–49. doi:10.1016/j.buildenv.2012.01.020.
- Feddema, J., G. Mills, and J. Ching. 2015. "Demonstrating the Added Value of WUDAPT for Urban Climate Modelling." In *ICUC9 - 9 Th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment*. Toulouse, France.
- Fonseca, A., and C. Fernandes. 2004. *Detecção Remota*. Lisboa: Lidel - Edições Técnicas, Lda.

- Fonte, Cidália, Marco Minghini, Vyron Antoniou, Linda See, Joaquim Patriarca, Maria Brovelli, and Grega Milcinski. 2016. "An Automated Methodology for Converting OSM Data into a Land Use/cover Map." *Proceedings of the 6th International Conference on Cartography & GIS* 1: 462–73.  
[http://vgibox.eu/repository/index.php/An\\_automated\\_methodology\\_for\\_converting\\_OSM\\_data\\_into\\_a\\_land\\_use/cover\\_map](http://vgibox.eu/repository/index.php/An_automated_methodology_for_converting_OSM_data_into_a_land_use/cover_map).
- Girres, Jean-François, and Guillaume Touya. 2010. "Quality Assessment of the French OpenStreetMap Dataset." *Transactions in GIS* 14 (4). Blackwell Publishing Ltd: 435–59. doi:10.1111/j.1467-9671.2010.01203.x.
- Goodchild, M., J. Glennon, and J Alan Glennon. 2010. "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier." *International Journal of Digital Earth* 3 (3): 231–41. doi:10.1080/17538941003759255.
- Goodchild, Michael. 2007a. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 29 (4): 211–21. doi:10.1007/s10708-007-9111-y.
- . 2007b. "Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0." *International Journal of Spatial Data Infrastructures Research* 2: 24–32.
- Gugler, J. 1996. "The Urban Transformation of the Developing World." New York : Oxford University Press.
- Haklay, Mordechai. 2010. "How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets." *Environment and Planning B: Planning and Design* 37 (4). SAGE Publications: 682–703. doi:10.1068/b35097.
- Heywood, Ian, Sarah Cornelius, and Steve Carver. 2006. *An Introduction to Geographical Information Systems*. 3rd editio. Prentice Hall, Inc.
- Horita, Flávio, and João de Albuquerque. 2013. "An Approach to Support Decision-Making in Disaster Management Based on Volunteer Geographic Information (VGI) and Spatial Decision Support Systems (SDSS)." In *Proceedings of the 10th International ISCRAM Conference*, edited by J. Geldermann and T.Müller T. Comes, F. Fiedrich, S. Fortier, 301–6. Baden-Baden, Germany: Supporting Disaster Management using VGI and SDSS.
- Humboldt State University. 2014. "HSU OLM - About Online Learning."  
[http://gsp.humboldt.edu/OLM/Lessons/GIS/08\\_Rasters/Images/RasterGrid.png](http://gsp.humboldt.edu/OLM/Lessons/GIS/08_Rasters/Images/RasterGrid.png).
- Instituto Nacional de Estatística. 2013. "População Residente." Lisboa.  
[https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_indicadores&indOcorrCod=0008306&contexto=bd&selTab=tab2](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008306&contexto=bd&selTab=tab2).
- IPCC. 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Miller, and H. Miler. United Kingdom and New York, NY, USA: Cambridge University Press.
- Isied, Anwar, and Hashem Tamimi. 2015. "Using Random Forest (RF) as a Transfer Learning Classifier for Detecting Error-Related Potential (ErrP) within the Context of P300-Speller." In *Bernstein Conference*. Berlin. doi:10.12751/NNCN.BC2015.0143.
- Jensen, John. 2004. *Introductory Digital Image Processing: A Remote Sensing Perspective*. 3rd Editio. Upper Saddle River, NJ: Prentice Hall, Inc.

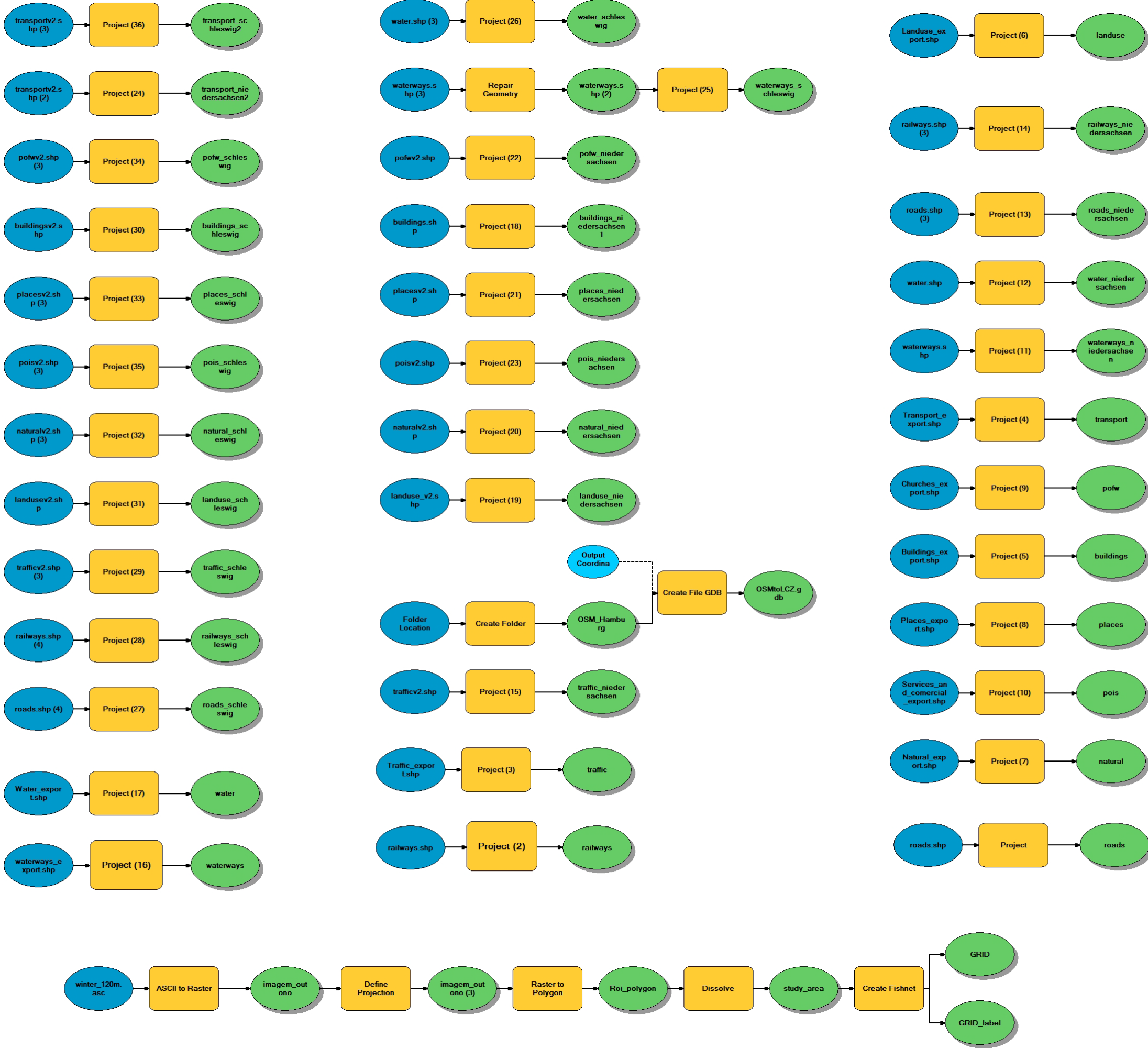
- Johnson, G., T. Oke, T. Lyons, D. Steyn, I. Watson, and J. Voogt. 1991. "Simulation of Surface Urban Heat Islands under IDEAL Conditions at Night Part 1: Theory and Tests against Field Data." *Boundary-Layer Meteorology* 56 (3). Kluwer Academic Publishers: 275–94. doi:10.1007/BF00120424.
- Kaloustian, N., and B. Bechtel. 2016. "Local Climatic Zoning and Urban Heat Island in Beirut." In *4 Th International Conference on Countermeasures to Urban Heat Island*, 1–10. Singapore.
- Knight, Edward, and Geir Kvaran. 2014. "Landsat-8 Operational Land Imager Design, Characterization and Performance." *Remote Sensing* 6 (11). Multidisciplinary Digital Publishing Institute: 10286–305. doi:10.3390/rs61110286.
- Kondratyev, K., and A. Cracknell. 1998. *Observing Global Climate Change*. Edited by Taylor and Francis Ltd. London.
- Manfré, Luiz, Eliane Hirata, Janaína Silva, Eduardo Shinohara, Mariana Giannotti, Ana Paula Larocca, and José Quintanilha. 2012. "An Analysis of Geospatial Technologies for Risk and Natural Disaster Management." *ISPRS International Journal of Geo-Information* 1 (3). Molecular Diversity Preservation International: 166–85. doi:10.3390/ijgi1020166.
- Martinho, Nuno, and Cidália Fonte. 2015. "Assessing the Applicability of OpenStreetMap Data to Validate Land Use/Land Cover Maps." *Instituto de Engenharia de Sistemas E Computadores de Coimbra* 5.
- Meng, F., and M. Liu. 2013. "Remote-Sensing Image-Based Analysis of the Patterns of Urban Heat Islands in Rapidly Urbanizing Jinan, China." *International Journal of Remote Sensing* 34 (24): 8838–53. doi:10.1080/01431161.2013.853895.
- Middel, A., K. Häb, A. Brazel, C. Martin, and S. Guhathakurta. 2014. "Impact of Urban Form and Design on Mid-Afternoon Microclimate in Phoenix Local Climate Zones." *Landscape and Urban Planning* 122: 16–28. doi:10.1016/j.landurbplan.2013.11.004.
- Millard, Koreen, and Murray Richardson. 2015. "On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping." *Remote Sensing* 7 (7). Multidisciplinary Digital Publishing Institute: 8489–8515. doi:10.3390/rs70708489.
- Mills, G., B. Bechtel, J. Ching, L. See, J. Feddema, M. Foley, and P. Alexander. 2015. "An Introduction to the WUDAPT Project." In *ICUC9 - 9 Th International Conference on Urban Climate Jointly with 12 Th Symposium on the Urban Environment*. Toulouse, France.
- Neis, Pascal, Peter Singler, and Alexander Zipf. 2010. "Collaborative Mapping and Emergency Routing for Disaster Logistics - Case Studies from the Haiti Earthquake and the UN Portal for Afrika." In *Proceedings of the Geoinformatics Forum*. Salzburg, Austria.
- Neis, Pascal, and Dennis Zielstra. 2014. "Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap." *Future Internet* 6 (1). Multidisciplinary Digital Publishing Institute: 76–106. doi:10.3390/fi6010076.
- Neis, Pascal, Dennis Zielstra, and Alexander Zipf. 2011. "The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011." *Future Internet* 4 (4). Molecular Diversity Preservation International: 1–21. doi:10.3390/fi4010001.
- Ng, Y. 2015. "A Study of Urban Heat Island Using 'Local Climate Zones' – The Case of Singapore." *British Journal of Environment and Climate Change* 5 (2): 116–33. doi:10.9734/BJECC/2015/13051.

- Oke, T. 1987. *Boundary Layer Climates*. Edited by Psychology Press. 2nd ed. London: Routledge.
- . 2004. “Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites.” <http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-81/IOM-81-UrbanMetObs.pdf>.
- . 2008. “Guide to Meteorological Instruments and Methods of Observation,” no. 8. Geneva: 681.
- Oort, Pepijn van. 2005. *Spatial Data Quality: From Description to Application*. PhD thesis, NCG, Nederlandse Commissie voor Geodesie.
- Ostermann, Frank, and Laura Spinsanti. 2011. “A Conceptual Workflow for Automatically Assessing the Quality of Volunteered Geographic Information for Crisis Management.” In *Proceedings of the 14th AGILE International Conference on Geographic Information Science*. University of Utrecht.
- Perera, N., M. Emmanuel, and P. Mahanama. 2012. “Mapping ‘Local Climate Zones’ and Relative Warming Effects in Colombo, Sri Lanka.” In *CUC8 – 8th International Conference on Urban Climates*. Dublin, Ireland.
- Poser, K., and D. Dransch. 2010. “Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation.” *Geomatica* 64 (1): 89–98.
- Pultar, Edward, Martin Raubal, Thomas J Cova, and Michael F Goodchild. 2009. “Dynamic GIS Case Studies: Wildfire Evacuation and Volunteered Geographic Information.” *Transactions in GIS* 13 (s1): 85–104. doi:10.1111/j.1467-9671.2009.01157.x.
- Ren, C., M. Cai, M. Wang, Y. Xu, and E. Ng. 2016. “Local Climate Zone (LCZ) Classification Using the World Urban Database and Access Portal Tools (WUDAPT) Method: A Case Study in Wuhan and Hangzhou.” In *The Fourth International Conference on Countermeasure to Urban Heat Islands (4th IC2UHI)*, 1–12. Nus, Singapore.
- Ribeiro, A., and C. Fonte. 2015. “A Methodology for Assessing Openstreetmap Degree of Coverage for Purposes of Land Cover Mapping.” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W5 (August): 297–303. doi:10.5194/isprsannals-II-3-W5-297-2015.
- Rose, Julia, and Christina Benita Wilke. 2015. *Climate Change Vulnerability in Cities – the Case of Hamburg*. Hamburg Institute of International Economics (HWWI). Hamburg. [http://www.hwwi.org/fileadmin/hwwi/Publikationen/Research/HWWI\\_ResearchPaper\\_167-kl.pdf](http://www.hwwi.org/fileadmin/hwwi/Publikationen/Research/HWWI_ResearchPaper_167-kl.pdf).
- Roy, D.P., M.A. Wulder, T.R. Loveland, Woodcock C.E., R.G. Allen, M.C. Anderson, D. Helder, et al. 2014. “Landsat-8: Science and Product Vision for Terrestrial Global Change Research.” *Remote Sensing of Environment* 145: 154–72. doi:10.1016/j.rse.2014.02.001.
- Sakakibara, Yasushi. 1999. “The Relationship between Heat Island Intensity and Rural Land Coverage in Obuse, Nagano.” *Tenki* 46: 567–75.
- Sakakibara, Yasushi, and Emi Matsui. 2005. “Relation between Heat Island Intensity and City Size Indices/urban Canopy Characteristics in Settlements of Nagano Basin, Japan.” *Geographical Review of Japan* 78 (12): 812–24.
- Santamouris, M. 2013. *Environmental Design of Urban Buildings: An Integrated Approach*. London:

- Earthscan. <https://books.google.com/books?id=c-v7OSaxGSsC&pgis=1>.
- Science, Indian Institute of. 2017. “ENVironmental Information System.” [http://ces.iisc.ernet.in/hpg/envis/Remote/section156\\_files/Fig15\\_21.jpeg](http://ces.iisc.ernet.in/hpg/envis/Remote/section156_files/Fig15_21.jpeg).
- See, L., J. Ching, V. Masson, J. Feddema, G. Mills, M. Neophytou, M. Foley, et al. 2015. “Generating WUDAPT’s Specific Scale-Dependent Urban Modeling and Activity Parameters: Collection of Level 1 and Level 2 Data.” In *ICUC9 - 9 Th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment*. Toulouse, France.
- Senaratne, H., A. Mobasher, A. Ali, C. Capineri, and M. Haklay. 2016. “A Review of Volunteered Geographic Information Quality Assessment Methods.” *International Journal of Geographical Information Science* 31 (1). Taylor & Francis: 139–67. doi:10.1080/13658816.2016.1189556.
- Sheppard, S., and P. Cizek. 2009. “The Ethics of Google Earth: Crossing Thresholds from Spatial Data to Landscape Visualisation.” *Journal of Environmental Management* 90 (6): 2102–17. doi:10.1016/j.jenvman.2007.09.012.
- Song, M, X Chen, and P Guo. 2009. “A Fusion Method for Multispectral and Panchromatic Images Based on HSI and Contourlet Transformation.”
- Sousa, A., and J. Marques. 2011. *Fundamentos Teóricos de Detecção Remota*. Évora: Departamento de Engenharia Rural.
- Statistical Office of Hamburg and Schleswig-Holstein. 2014. *Facts and Figures*. Kiel: Grafik & Druck GmbH & Co KG. [https://www.statistik-nord.de/fileadmin/Dokumente/Faltblätter/Faltblatt\\_Stadtportrait\\_2013\\_E\\_Internet.pdf](https://www.statistik-nord.de/fileadmin/Dokumente/Faltblätter/Faltblatt_Stadtportrait_2013_E_Internet.pdf).
- Stevan, Savić, Milošević Dragan, Lazić Lazar, Marković Vladimir, Arsenović Daniela, and Pavić Dragoslav. 2013. “Classifying Urban Meteorological Stations Sites By ‘local Climate Zones’: Preliminary Results for the City of Novi Sad (Serbia).” *Geographica Pannonica* 17 (3): 60–68.
- Stewart, I., and T. Oke. 2009. “Classifying Urban Climate Field Sites By ‘Local Climate Zones’: The Case of Nagano, Japan.” *The Seventh International Conference on Urban Climate*, no. July: 10–13.
- . 2012. “Local Climate Zones for Urban Temperature Studies.” *Bulletin of the American Meteorological Society* 93 (12). American Meteorological Society: 1879–1900.
- Stewart, I., T. Oke, and E. Krayenhoff. 2014. “Evaluation of the ‘local Climate Zone’ Scheme Using Temperature Observations and Model Simulations.” *International Journal of Climatology* 34 (4). John Wiley & Sons, Ltd: 1062–80.
- Studley, Heather, and Keith Weber. 2011. “Comparison of Image Resampling Techniques for Satellite Imagery.” In *Final Report: Assessing Post-Fire Recovery of Sagebrush-Steppe Rangelands in Southeastern Idaho*, edited by K. T. Webber and K. Davis, 252.
- Sui, D., M. Goodchild, and S. Elwood. 2013. “Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide.” In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, 1–12. Dordrecht: Springer Netherlands.
- Thomas, George, A.P. Sherin, Shareekul Ansar, and E.J. Zachariah. 2014. “Analysis of Urban Heat Island in Kochi, India, Using a Modified Local Climate Zone Classification.” *Procedia Environmental Sciences* 21: 3–13.

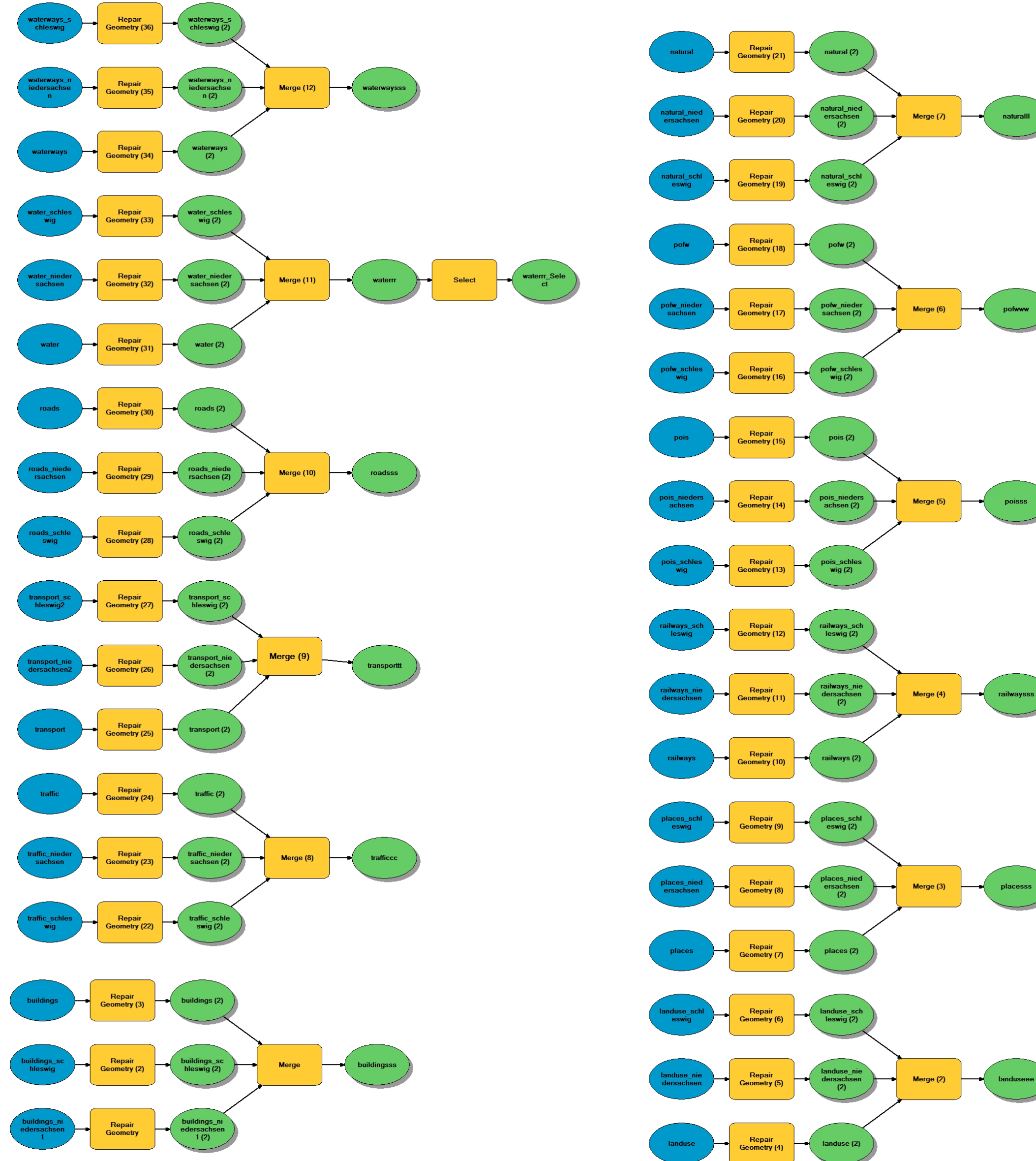
- U.S. Geological Survey, Department of the Interior. 2016. *Landsat 8 (L8) Data Users Handbook*. Sioux Falls: USGS Landsat User Services.
- Unger, J., E. Lelovics, and T. Gál. 2014. "Local Climate Zone Mapping Using GIS Methods in Szeged." *Hungarian Geographical Bulletin* 63 (1): 29–41. doi:10.15201/hungeobull.63.1.3.
- Unger, János, Enikő Lelovics, and Tamás Gál. 2014. "Local Climate Zone Mapping Using GIS Methods in Szeged." *Hungarian Geographical Bulletin* 63 (1): 29–41. doi:10.15201/hungeobull.63.1.3.
- Zheng, Y., C. Ren, Y. Shi, K. Ka-Lun, S. Yim, and D. Lau. 2015. "Applying 'Local Climate Zone (LCZ)' into a High-Density High-Rise Cities - A Pilot Study in Hong Kong." In *ICUC9 - 9 Th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment*. Toulouse, France.

# Appendix A – Model I: Preparation of the variables



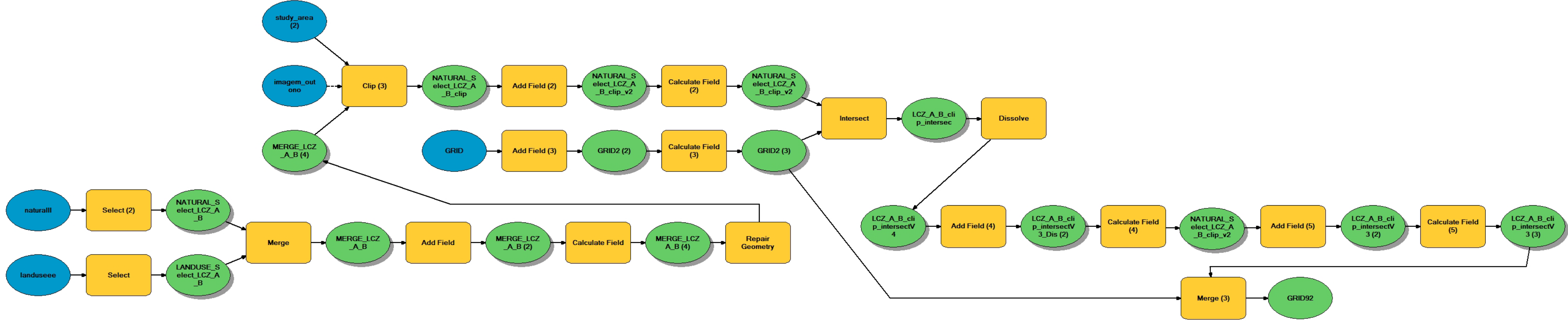


# Appendix B – Model II: Preparation of the variables (Aggregation)

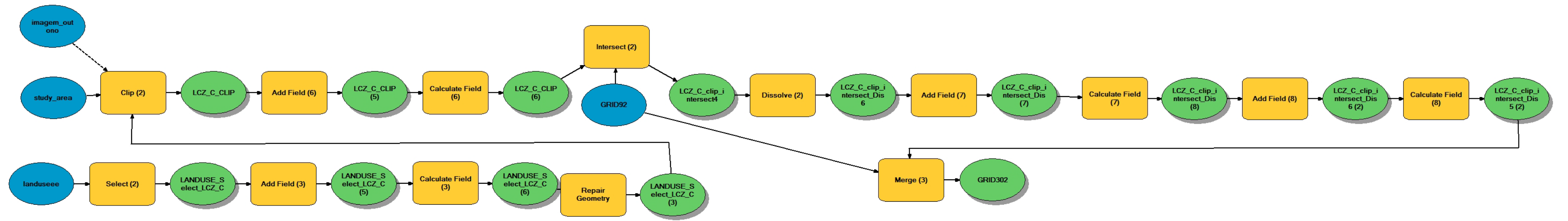




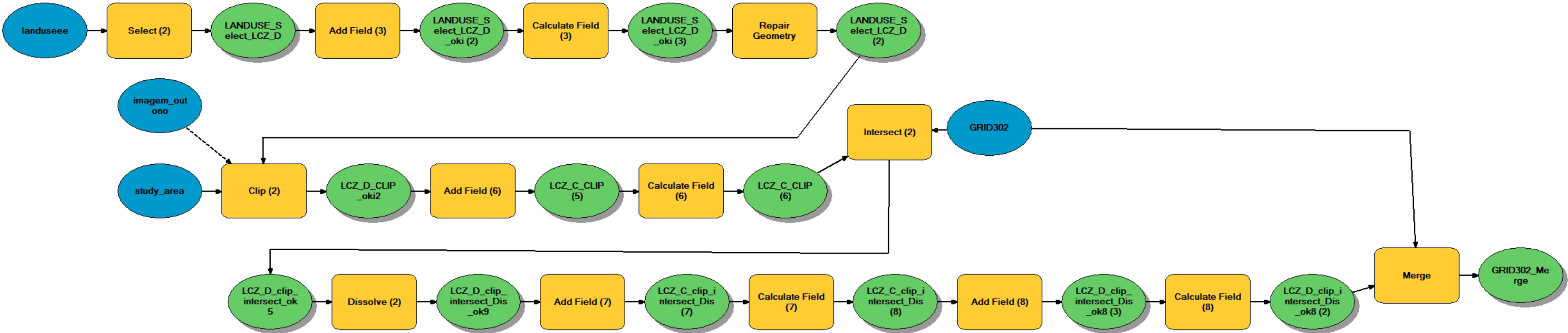
# Appendix C — Model III: OSM to LCZ A or B



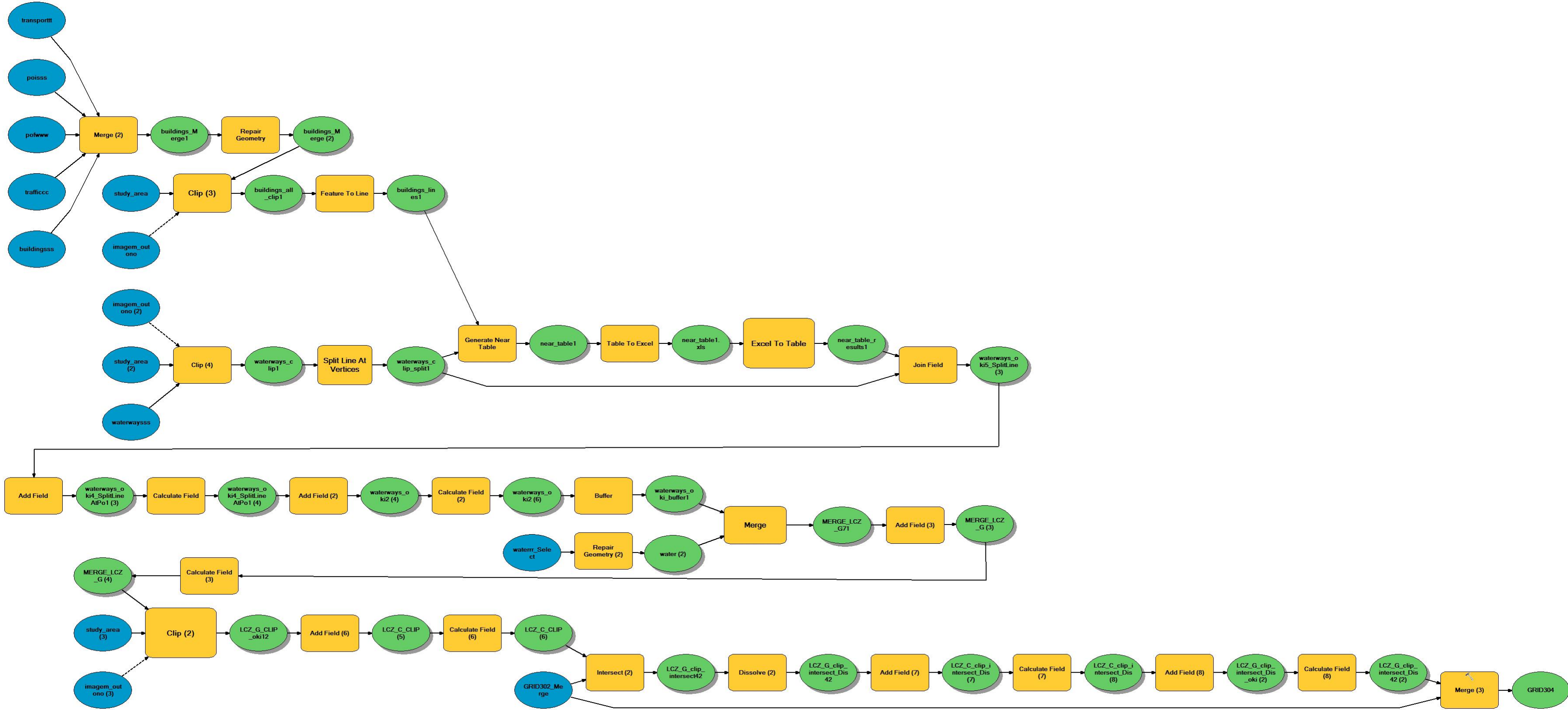
# Appendix D — Model IV: OSM to LCZ C



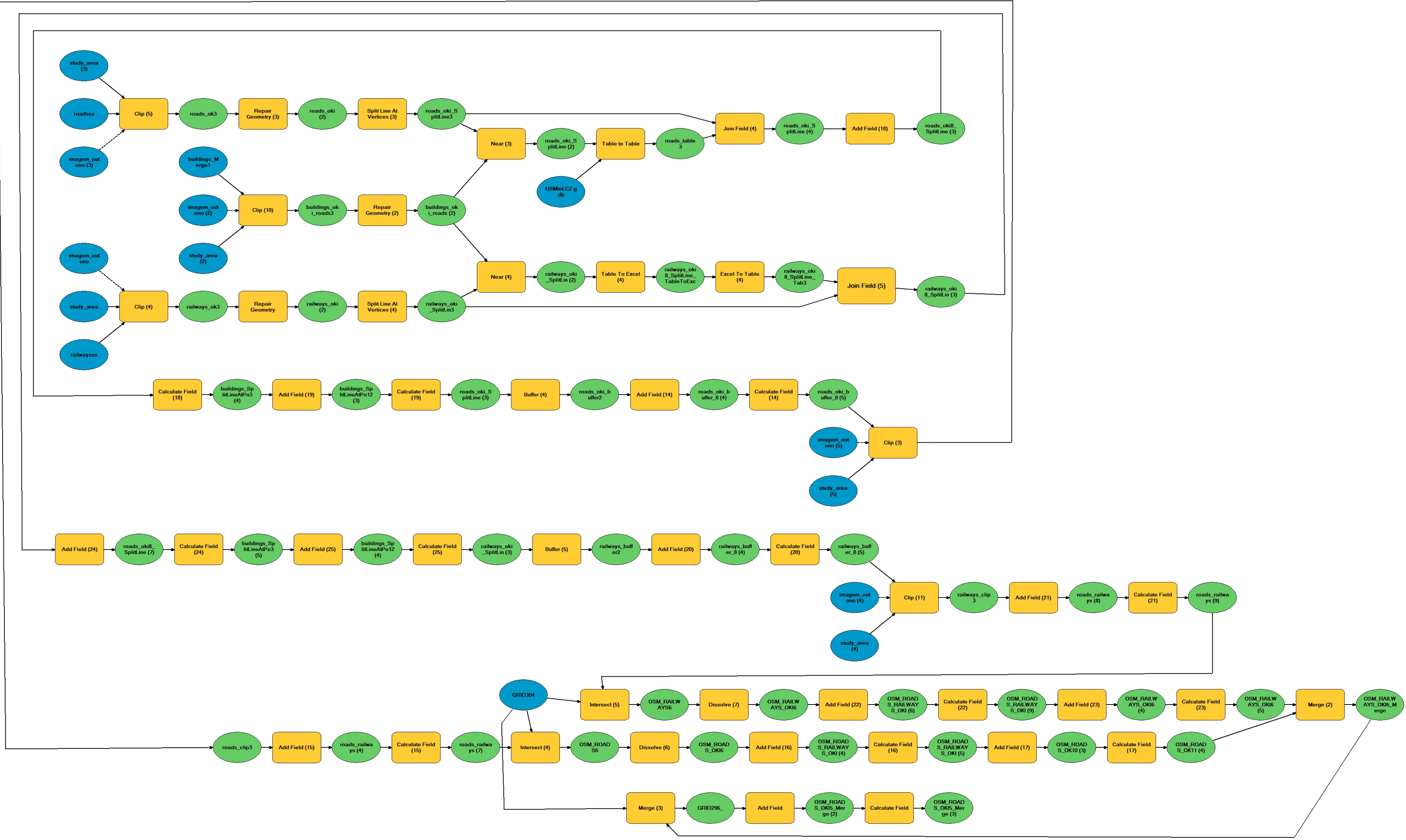
# Appendix E — Model V: OSM to LCZ D



# Appendix F — Model VI: OSM to LCZ G

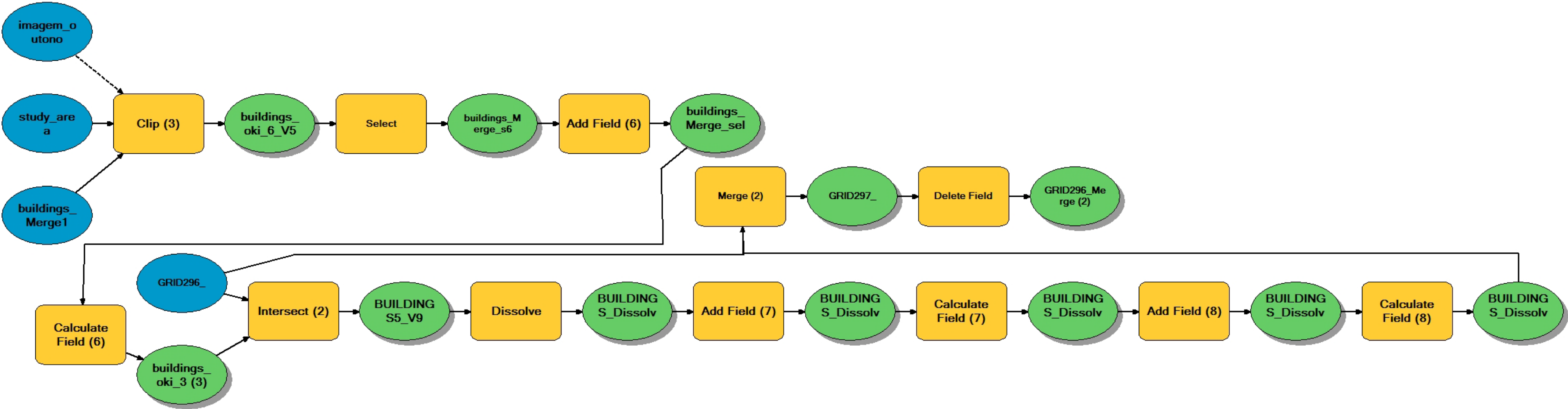


# Appendix G – Model VII: Obtain Impervious Surface Fraction (Roads and Railways)

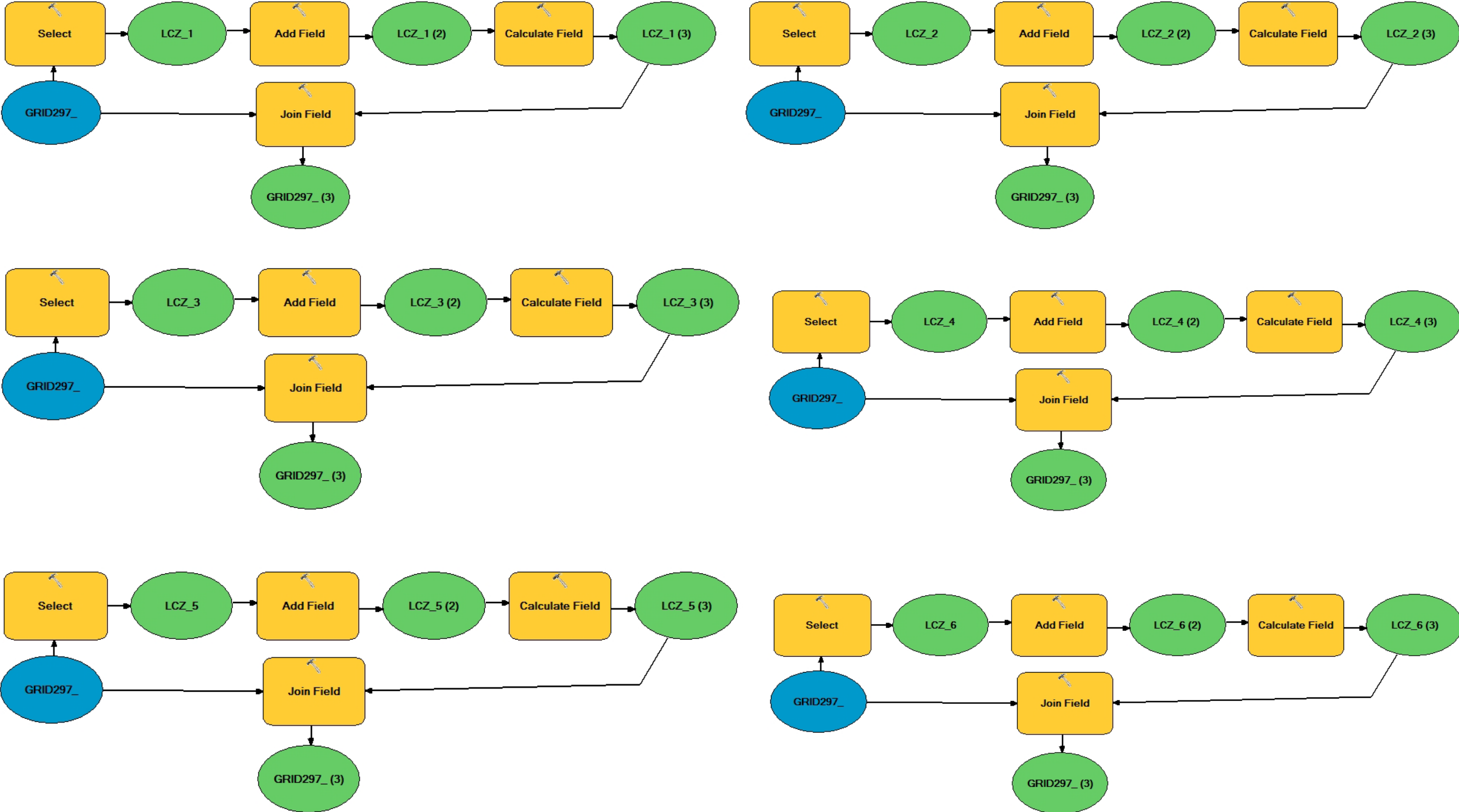


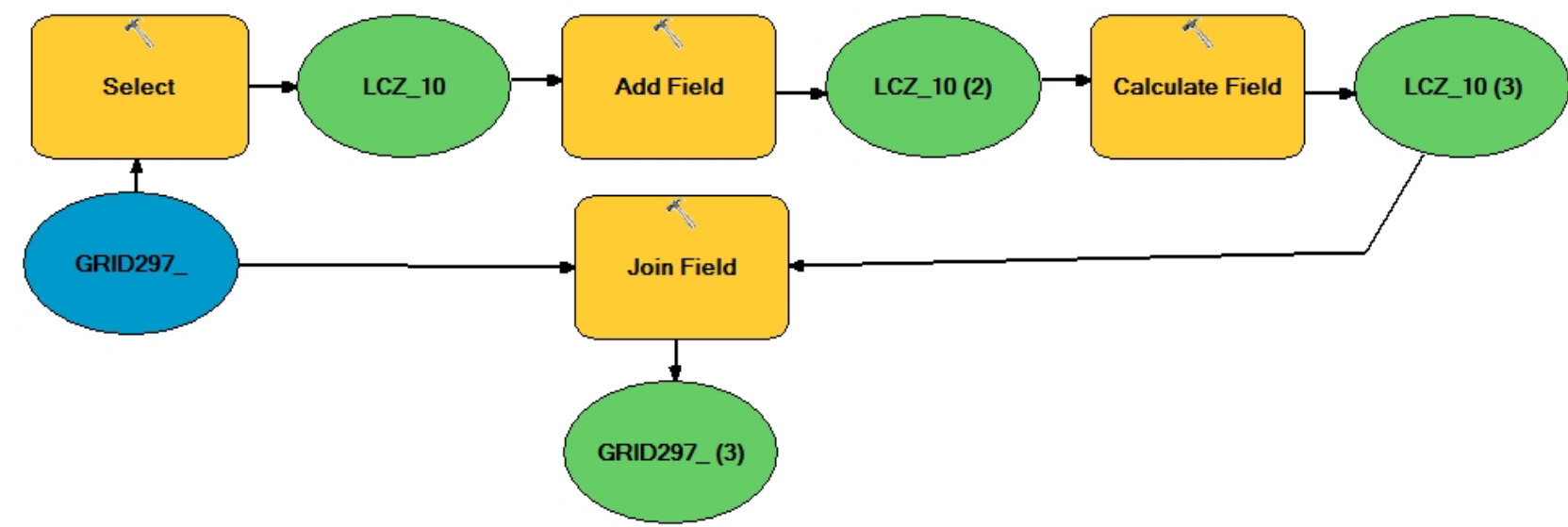
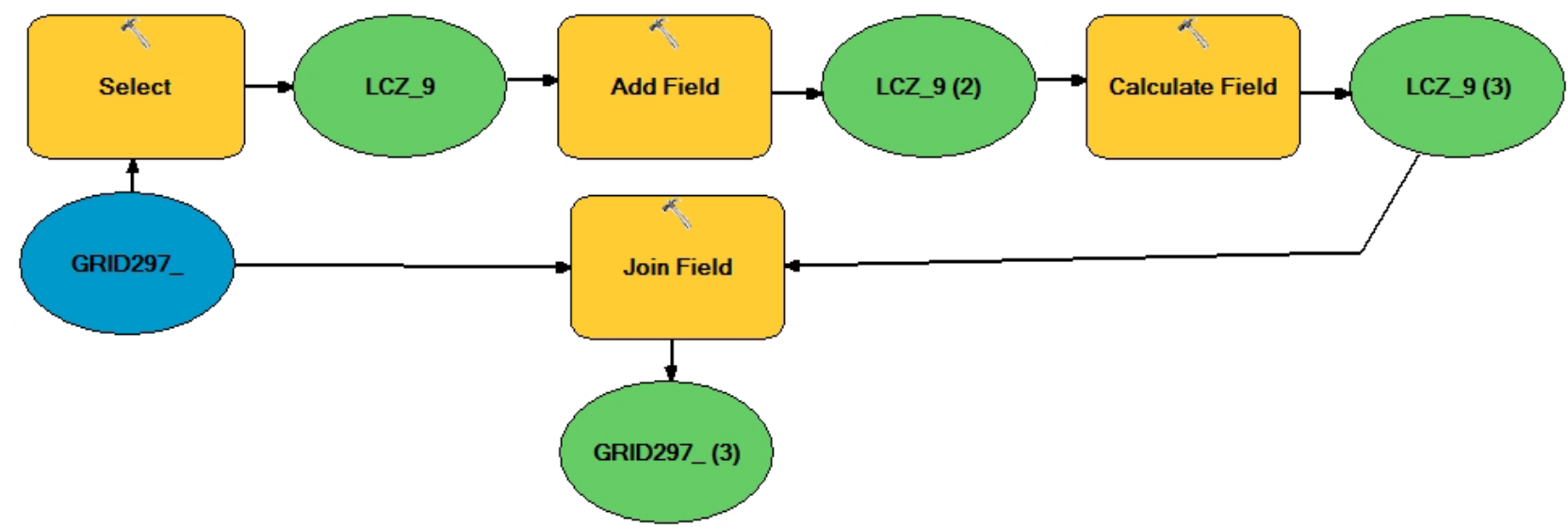
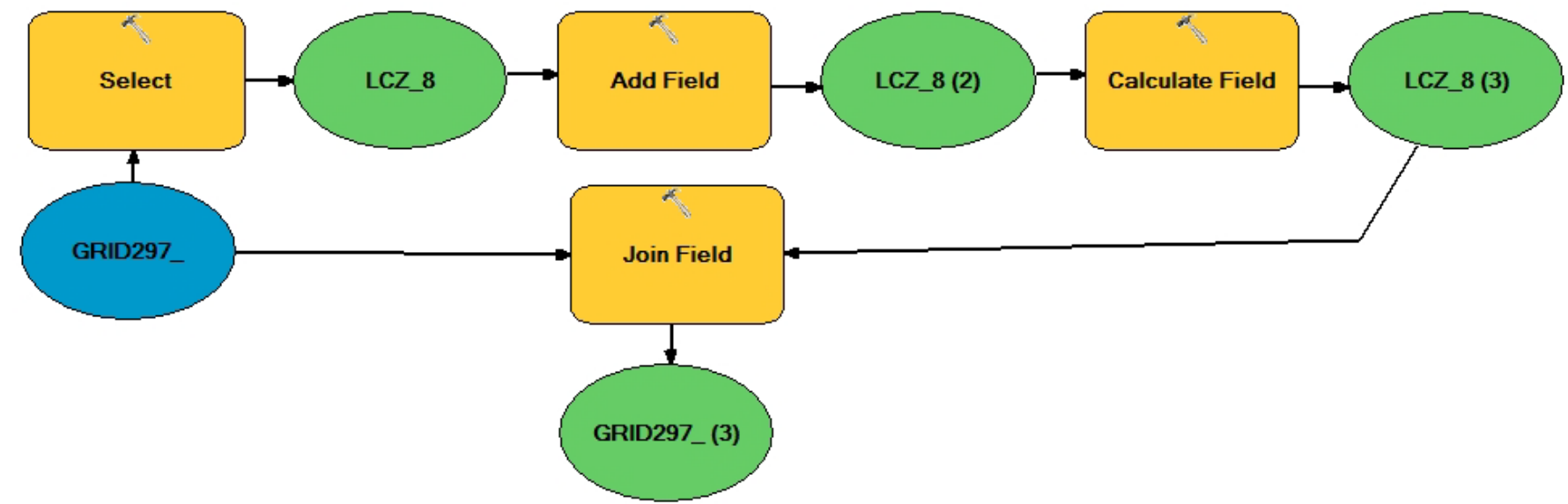
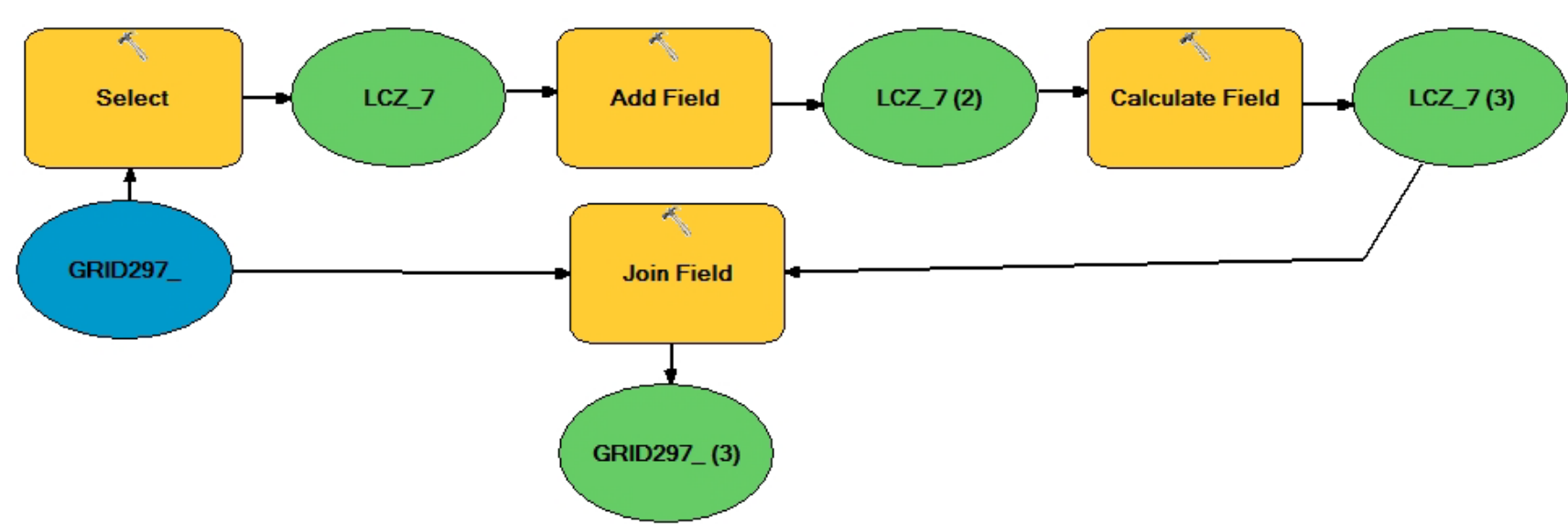


# Appendix H – Model VIII: Obtain Building Surface Fraction



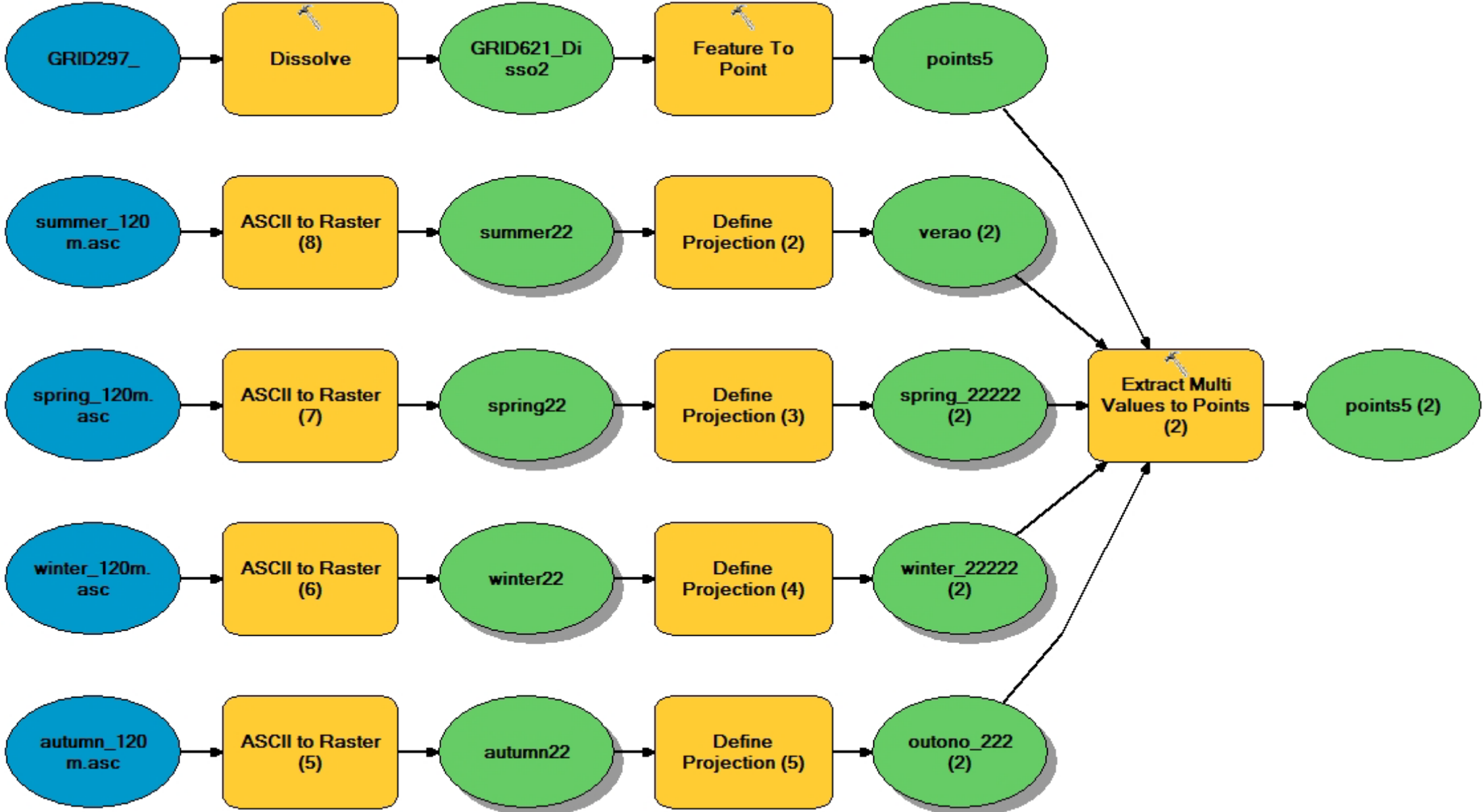
# Appendix I – Model IX: OSM to urban LCZ class (LCZ 1 to LCZ 10)



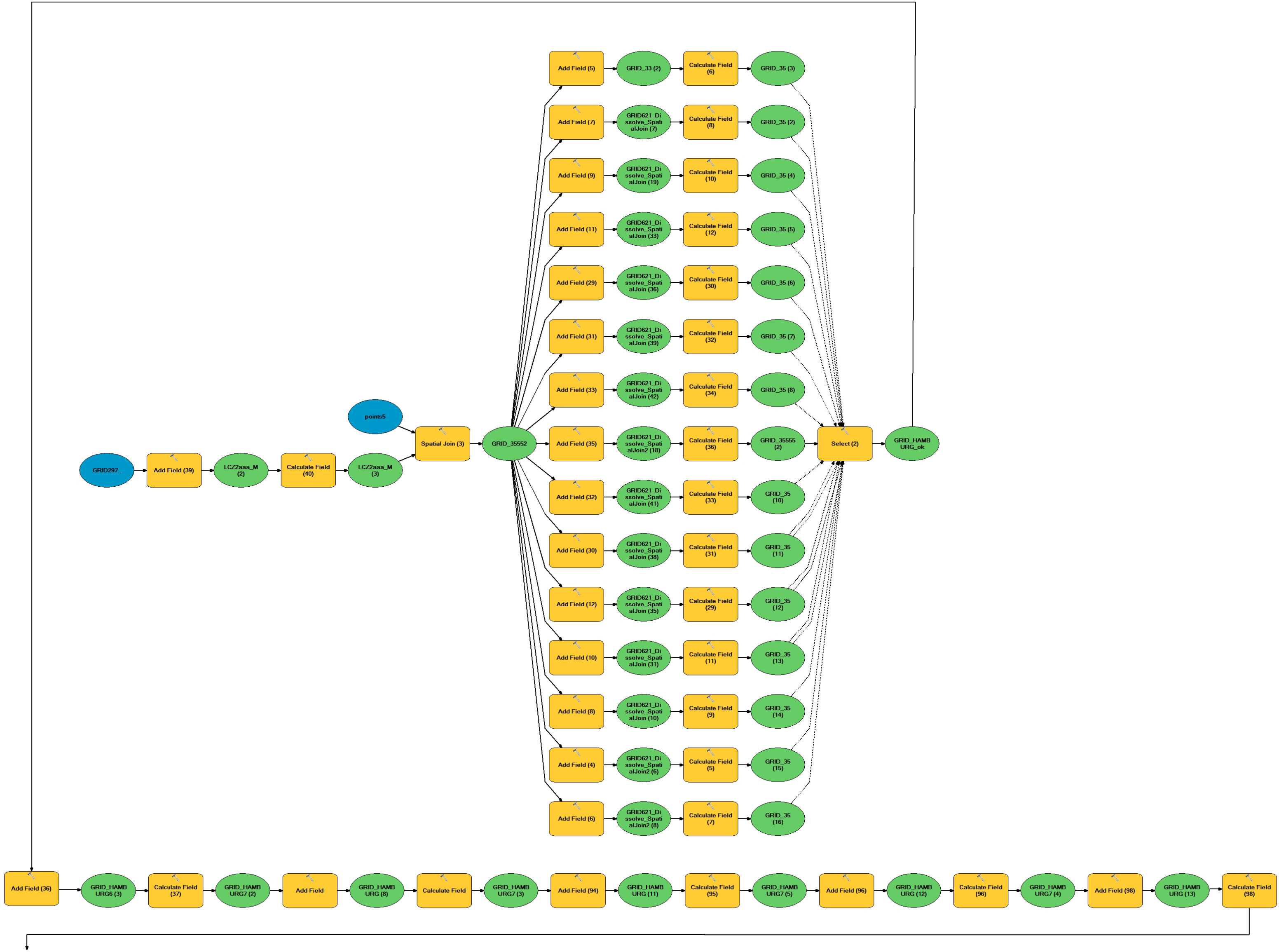




# Appendix J – Model X: Extract data from OSM maps



# Appendix K — Model XI: Final procedure





## Appendix L – Code Block to extract the percentage of times that a determinate class repeats on all classifications (summer, spring, autumn and winter) in each cell of the GRID

---

```
def Reclass (one,two,three,four):  
    valor = 0  
    while valor == 0:  
        if one == 104:  
            valor = 0 + 25  
        if two == 104:  
            valor = valor + 25  
        if three == 104:  
            valor = valor + 25  
        if four == 104:  
            valor = valor + 25  
    return valor
```

## Appendix M – Code Block to extract the percentage of area occupied by each in each cell of the GRID

---

```
def Reclass (dom, dom_area, sec, sec_area):  
    valor = 0  
    while valor == 0:  
        if dom == "LCZ D":  
            valor = 0 + dom_area  
        if sec == "LCZ D":  
            valor = valor + sec_area  
    return valor
```



## Appendix N – Confidence for each class regarding to satellite images for Coimbra

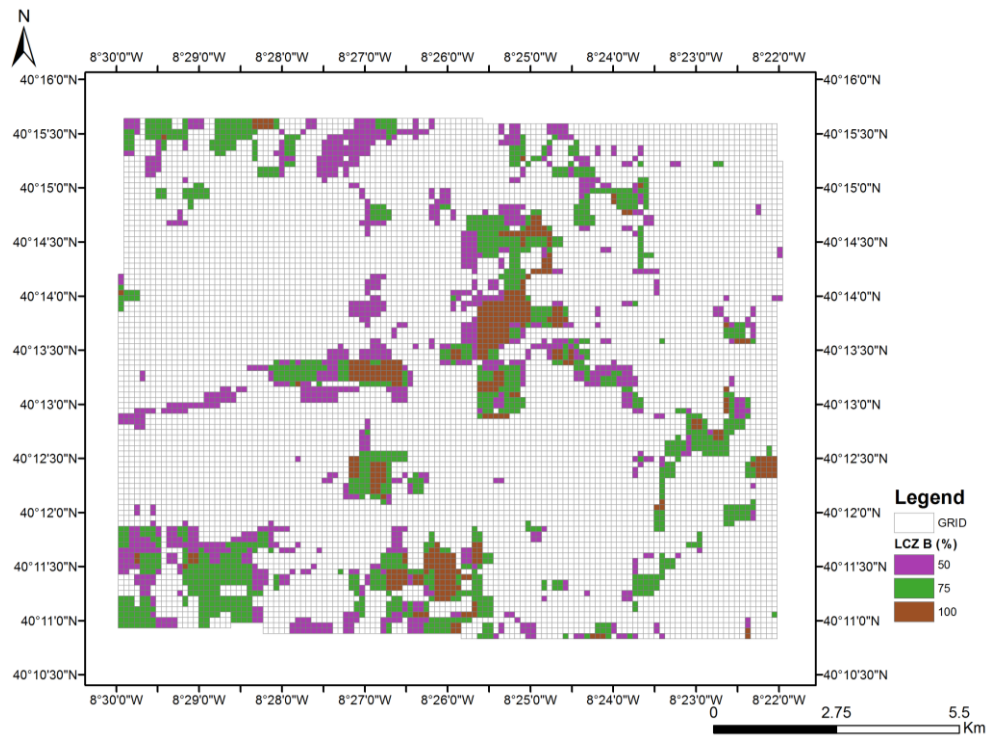


Figure 67: Confidence of presence of LCZ B (Scattered Trees) in values of percentage

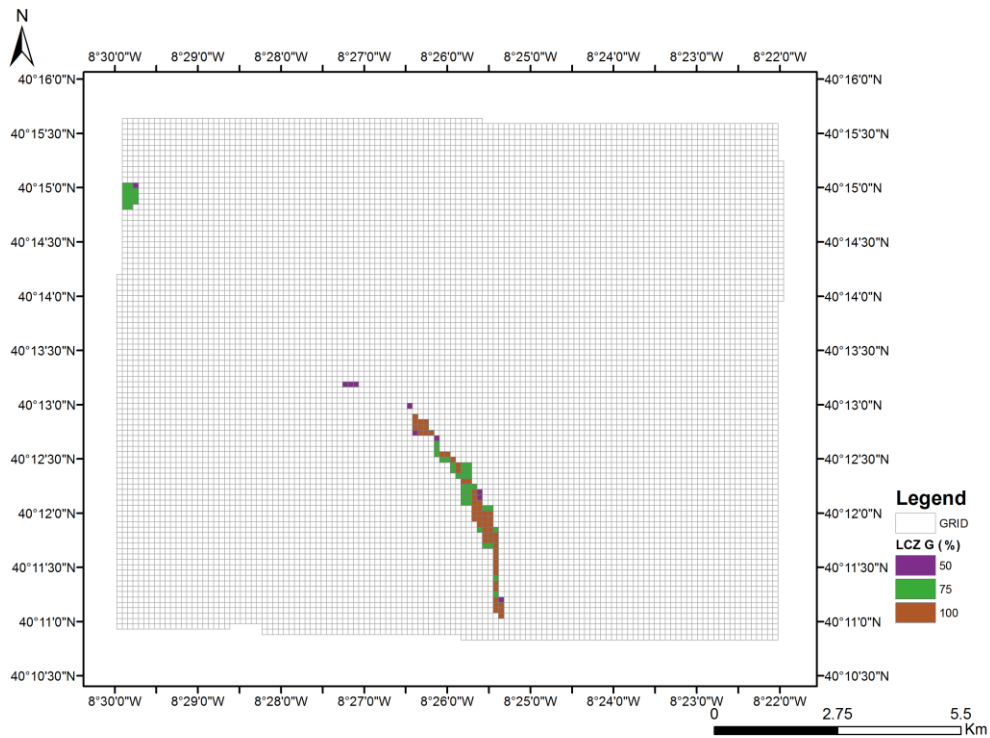


Figure 68: Confidence of presence of LCZ G (Water) in values of percentage

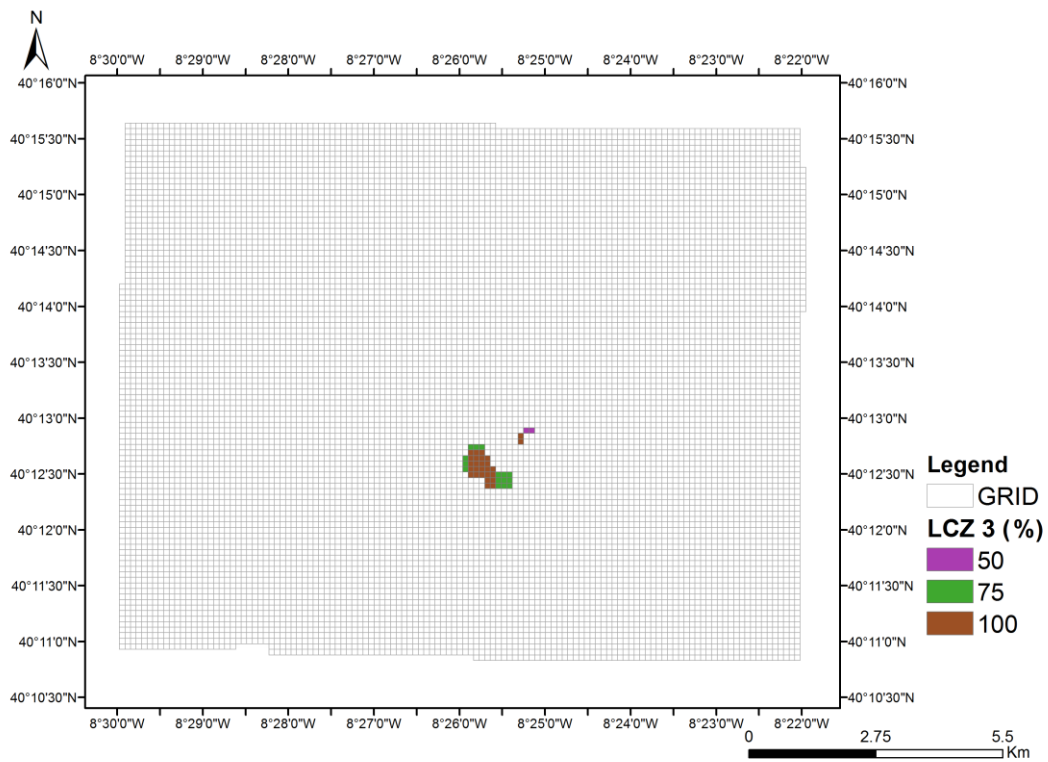


Figure 69: Confidence of presence of LCZ 3 (Compact low-rise) in values of percentage

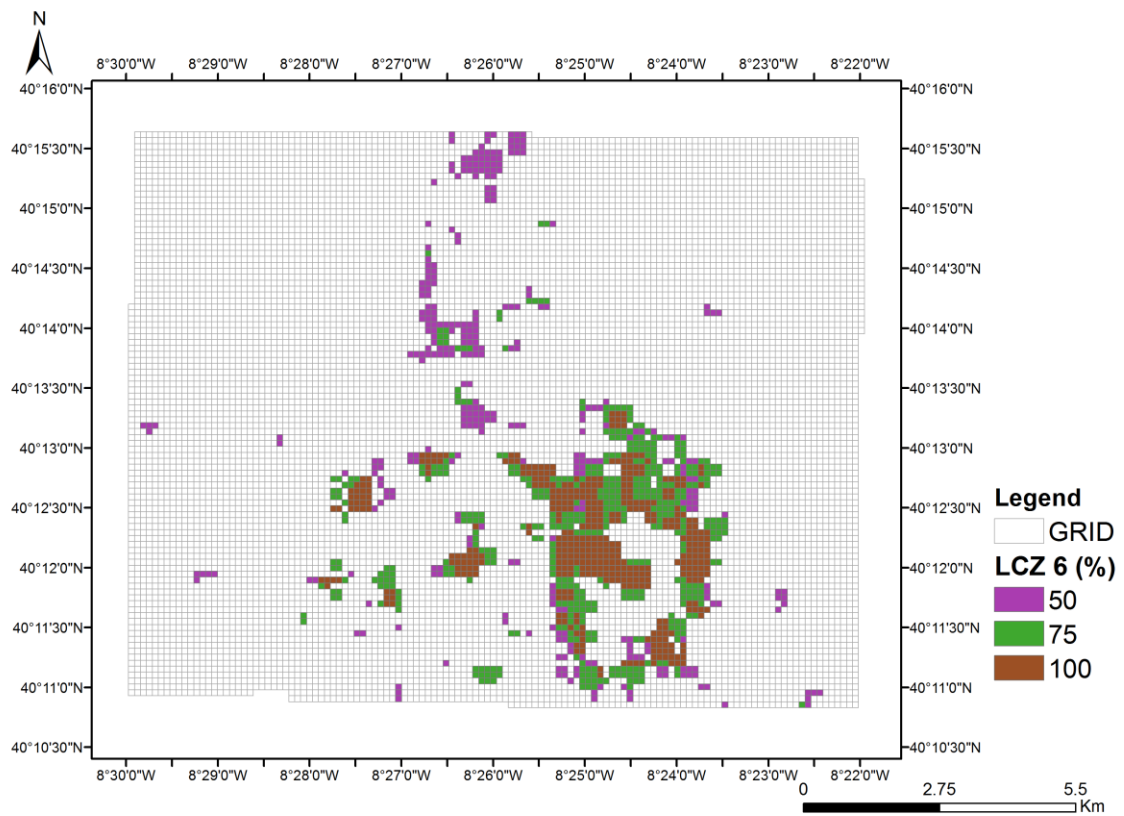


Figure 70: Confidence of presence of LCZ 6 (Open low-rise) in values of percentage



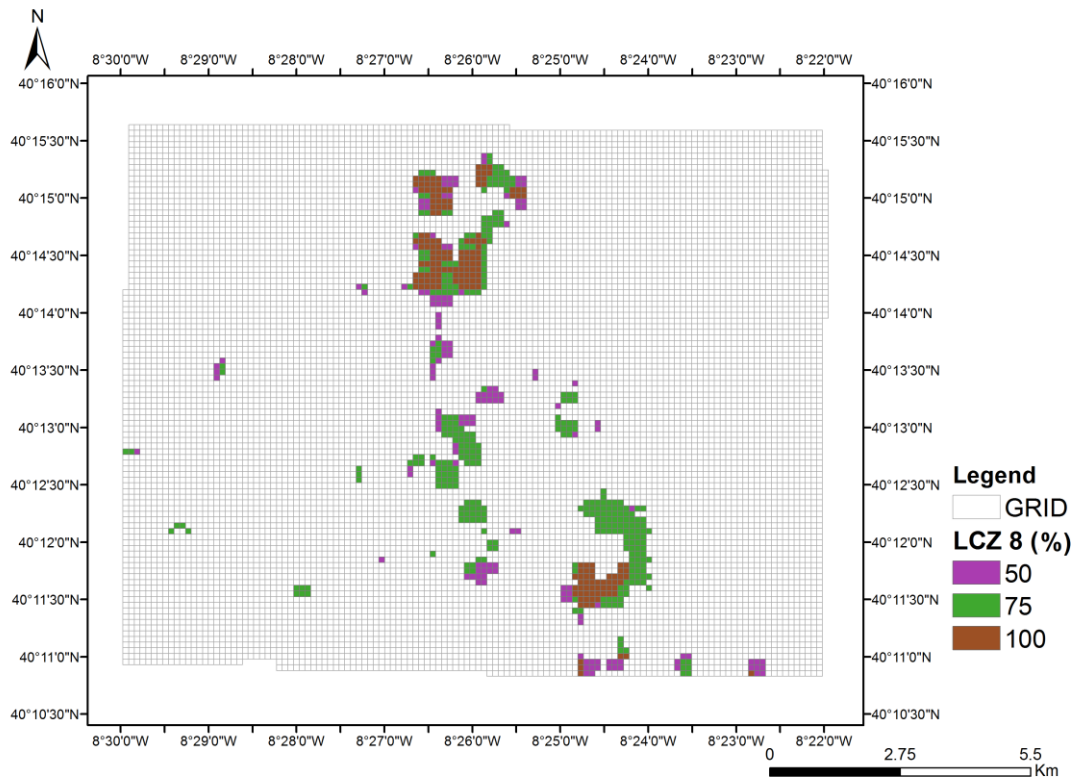


Figure 71: Confidence of presence of LCZ 8 (Large low-rise) in values of percentage

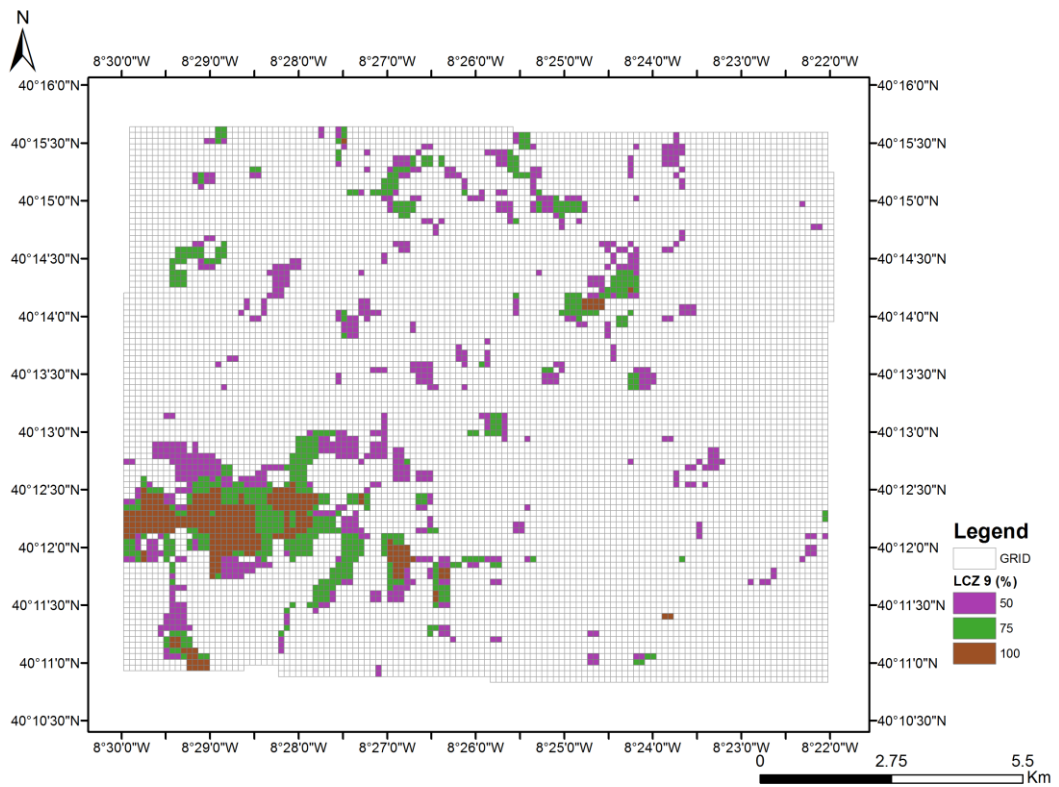


Figure 72: Confidence of presence of LCZ 9 (Scattered built) in values of percentage

## Appendix 0 – Confidence for each class regarding to satellite images for Hamburg

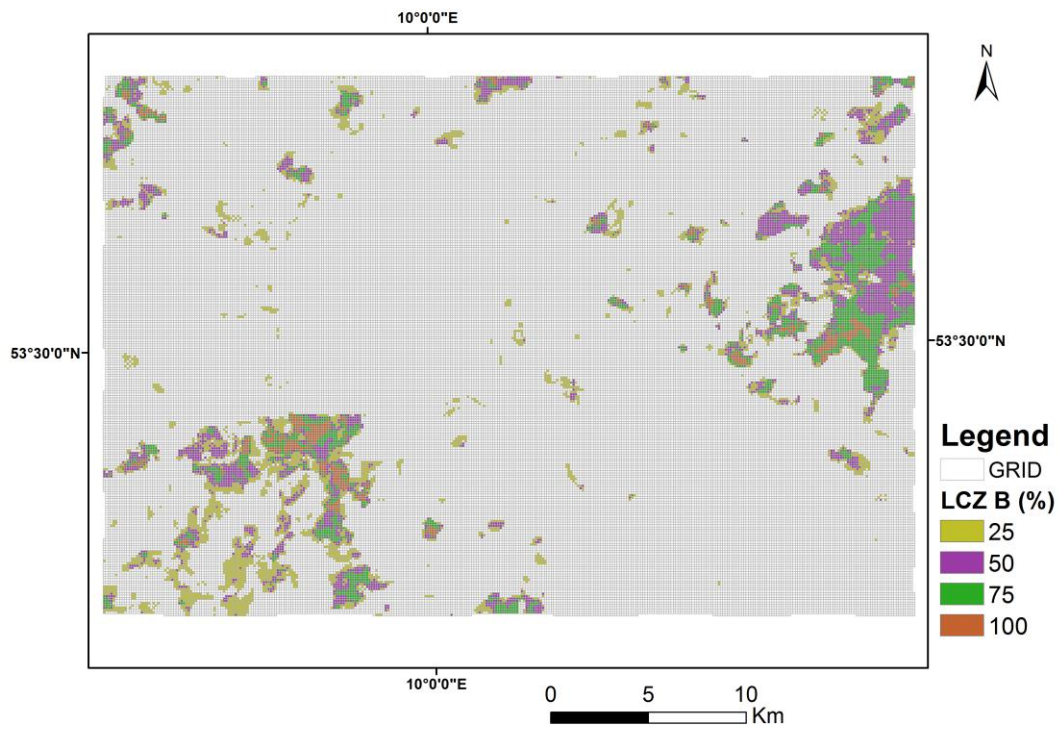


Figure 73: Confidence of presence of LCZ B (Scattered Trees) in values of percentage

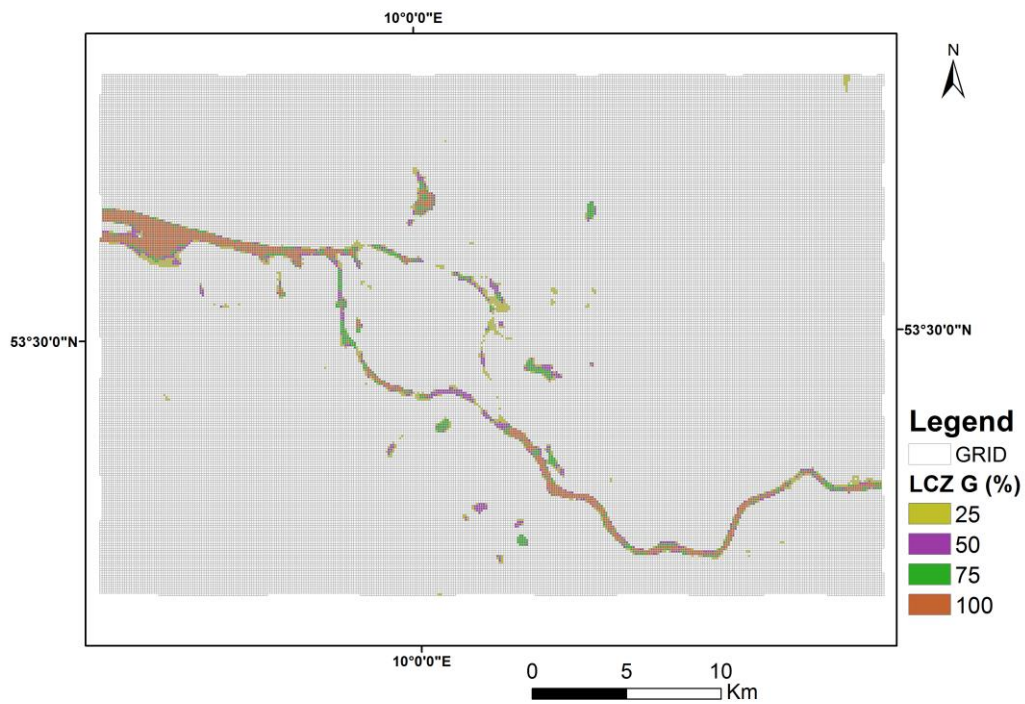


Figure 74: Confidence of presence of LCZ G (Water) in values of percentage

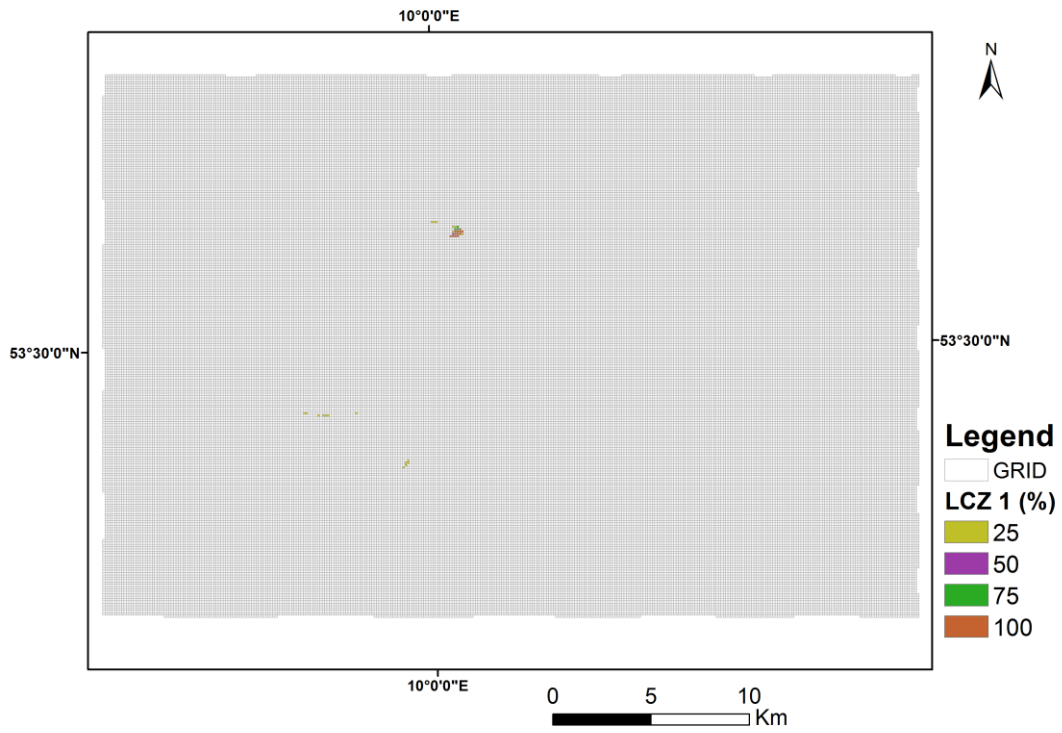


Figure 75: Confidence of presence of LCZ 1 (Compact high-rise) in values of percentage

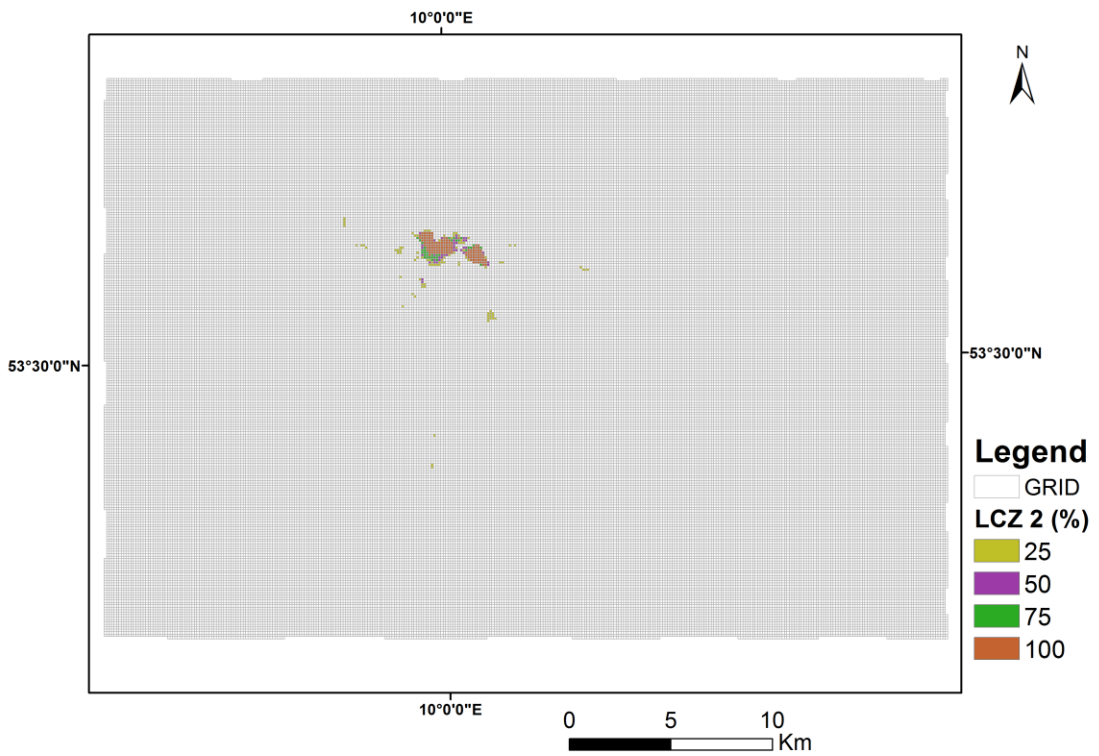


Figure 76: Confidence of presence of LCZ 2 (Compact mid-rise) in values of percentage



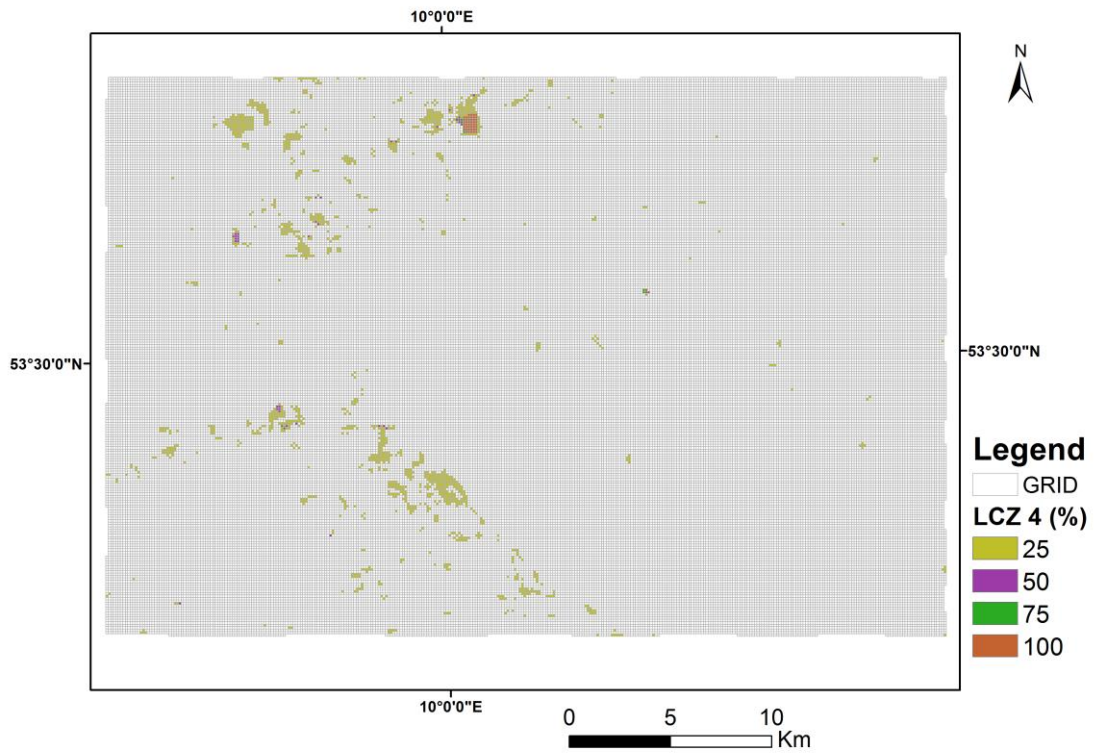


Figure 77: Confidence of presence of LCZ 4 (Open high-rise) in values of percentage

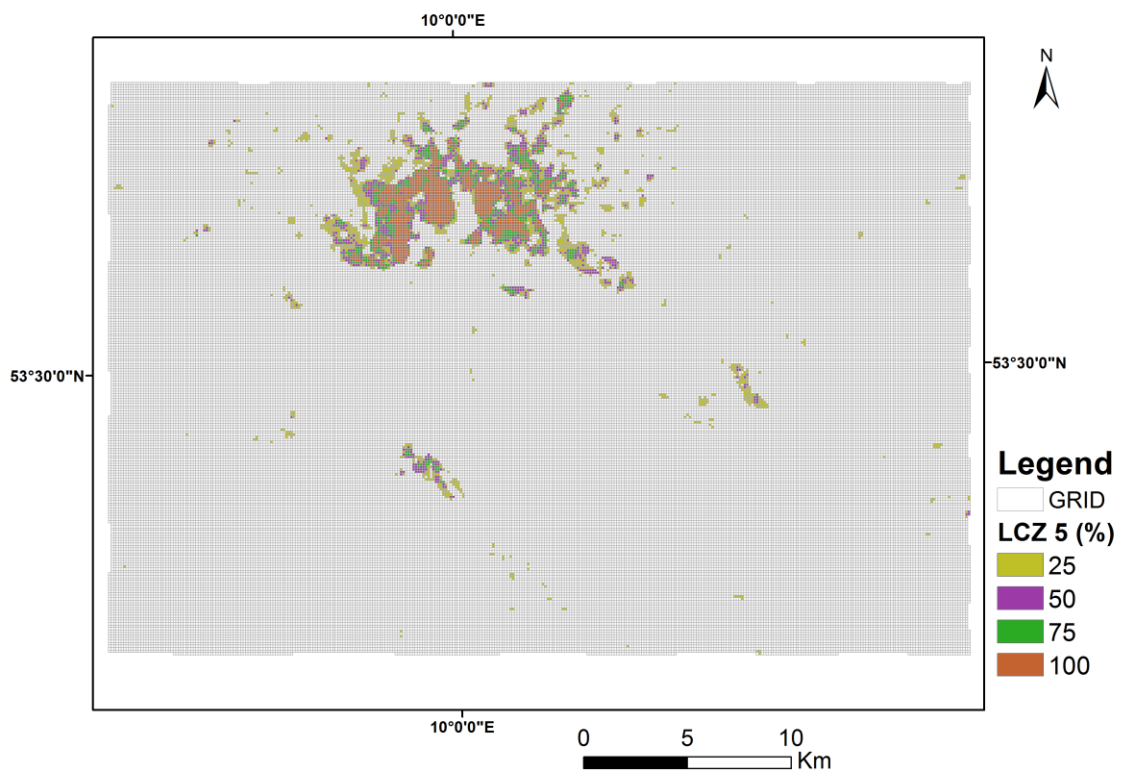


Figure 78: Confidence of presence of LCZ 5 (Open midrise) in values of percentage

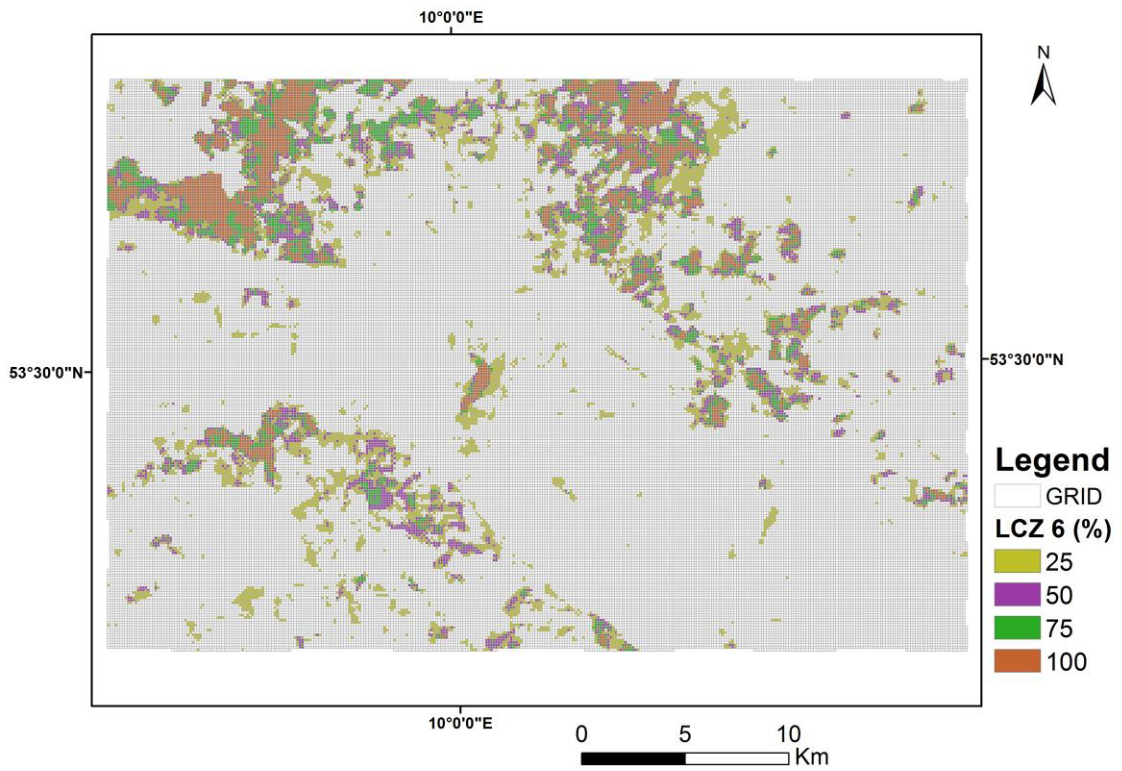


Figure 79: Confidence of presence of LCZ 6 (Open low-rise) in values of percentage

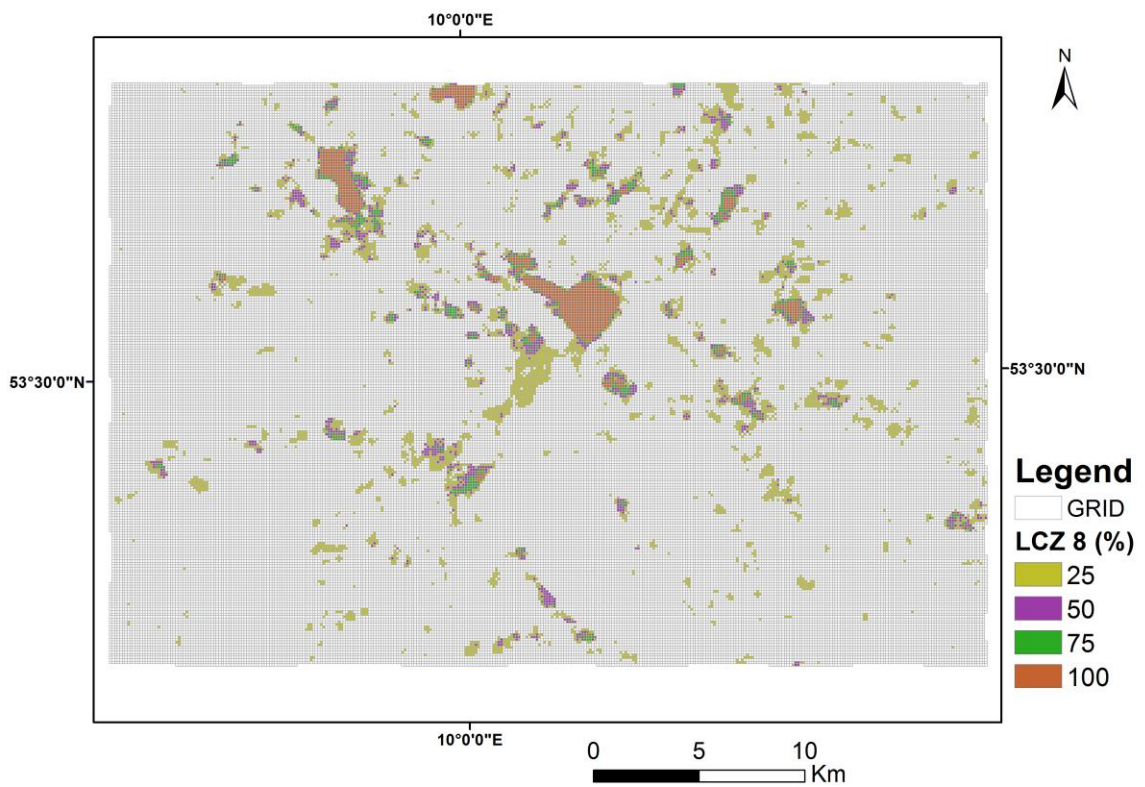


Figure 80: Confidence of presence of LCZ 8 (Large low-rise) in values of percentage



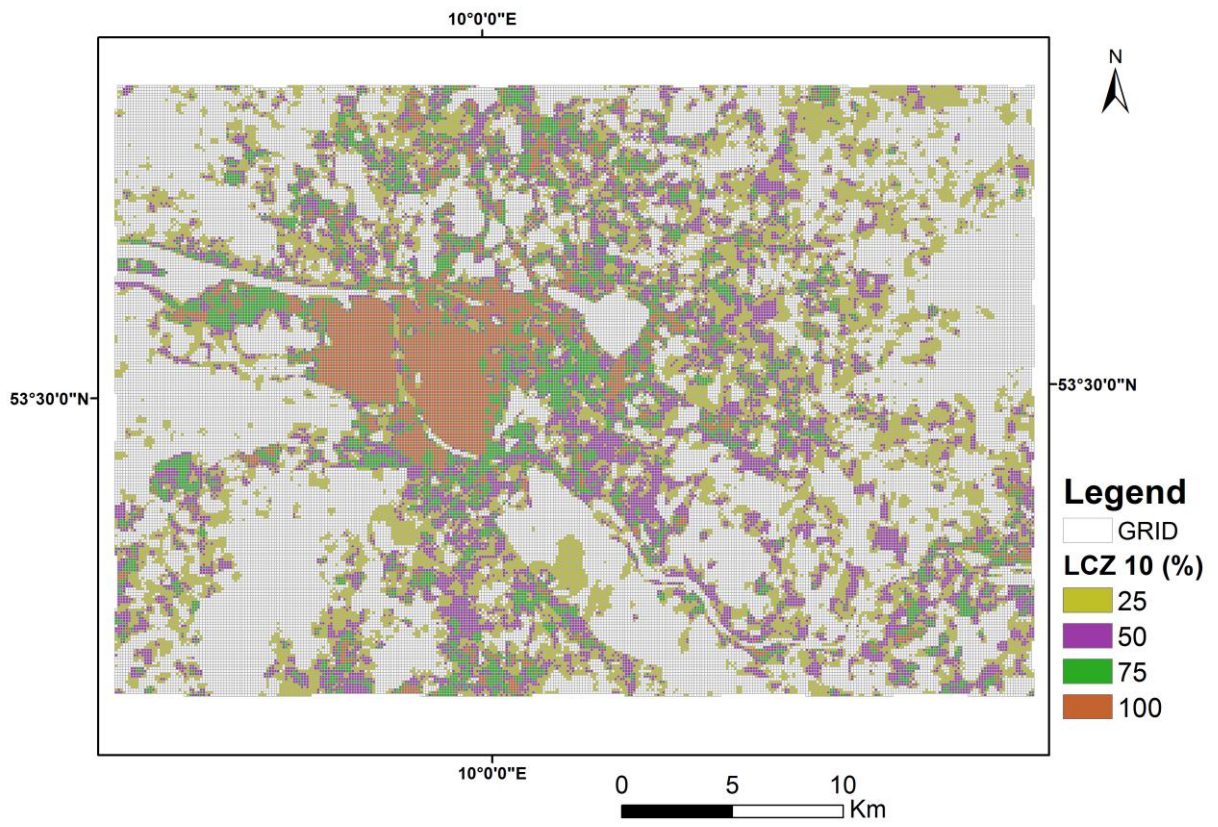


Figure 81: Confidence of presence of LCZ 10 (Heavy Industry) in values of percentage