

**THESIS IN GEOMATICS FOR THE BUILT ENVIRONMENT**

**JUNE 2018**

**GEOGRAPHICALLY WEIGHTED  
URBAN HEAT ISLAND MODELING  
USING THE NETATMO SENSORS.  
THE CASE OF THE HAGUE.**

**LILIA ANGELOVA | 4620380**

# Geographically weighted Urban Heat Island modeling using the Netatmo sensors.

## The case of The Hague.

A thesis submitted to the Delft University of Technology in partial fulfillment of the  
requirements for the degree of

Master of Science in Geomatics for the Built Environment

At Delft University of Technology

By

**Lilia Angelova**

Student number: 4620380

Supervisors: Jorge Lopes Gil

Alexander Wandl

Co-reader: Dr. Ir. F.D. van der Hoeven

June 2018

## Acknowledgments

Here I would like to thank all people who were around me in my journey of studying abroad. First of all, I want to thank my main supervisor Jorge Gil for the big amount of positive energy, always helpful conversations and constructive comments. Additionally, I want to thank Alexander Wandl for the positivism, the data that he provided me with and the helpful comments and advices.

For me, the challenge of following the master of Geomatics was long and sometimes tough but thanks to the nice atmosphere, interesting talks and the nicest coffee breaks the time went by faster than I thought. So, I want to thank my Geomatics peers for being so inspired and motivated. They were the most positive competition I have ever had.

Additionally, I want to thank Daniël Kersbergen, Laurens Oostwegel and Dimitris Xenakis for the time we spent bouldering and talking. This was most stress relieving time and the physical and mental training that I needed very much.

Further, I want to thank my best friend Iliyana Miteva, who was my family here for the last year. Thank you for the tremendous amount of support and love that you gave me, for sharing my best and worst moments. I want to thank Joan Sala Calero too for the support, understanding and nice time we shared together and to all my other close people and friends for helping me being myself but pushing my limits as well.

Finally, I want to thank my family for the love, the support, the trust and everything that they gave me. Without you, I wouldn't have been the person I am now and wouldn't have been able to achieve even part of this. Lastly, I want to dedicate this work to my two grandfathers who are not here with us anymore.

## Abstract

The continuously increasing population within cities imposes the future challenges related to planning and managing the sustainable environment, where people's health and wellbeing is prioritized. Currently more than half of the world's population live in cities which results in the rise of the human footprint, affecting the local climate. Simultaneously the planet's climate is changing towards less predictable weather conditions with high extremes, e.g., heavy rainfalls, followed by long dry periods or severe heatwaves. Frequently extreme weather conditions are associated with flooding and hurricanes, while heatwaves represent an equally important danger for the health of city's citizens.

UHI has been recognized as one of the leading environmental issues of the 21<sup>st</sup> century. The UHI is defined as the area within a city with the higher surface or air temperatures compared to its surroundings. The higher urban temperatures are resulting in health-related issues among the population, greater energy demands, and various economic losses.

Thus, the current work is focusing on researching the air temperatures in the urban canopy layer, which are mainly affected by the heat radiated from the urban surfaces during the night. Moreover, there are differences in the heat exchange phases between the neighborhoods, which are caused by the heterogeneity of the morphological characteristics of the city. Such information can be of great use for the development of UHI mitigation strategies, but it is currently very sparse or not detailed enough. Therefore this master thesis investigates the possibility for generation of more detailed models, taking into account the intraurban variability.

For the development of the statistical models explaining the UHI effect in the city of the Hague different spatial and sensor datasets have been used. The data about the temperatures in the city for 2017 have been collected with the means of the Netatmo weather stations. Additionally, different spatial representations and their effect on the statistical analysis have been investigated. In combination with the different spatial models, six distinct UHI contributing factors have been researched, namely the Buildings density, the Land cover index, the Vegetation index, the Sky View Factor, the Non-permeable surfaces in the city and the Vehicle traffic density. These variables have been calculated and utilized in the statistical analysis of their relationship with the air temperatures in the Hague. The results indicated a weak relationship between the air temperatures and the different spatial characteristics of the city where only the Sky View Factor and the Non-permeable surfaces proved to be statistically significant variables.

## Table of Contents

Acknowledgments .....	2
Abstract .....	3
List of Figures.....	6
List of Tables.....	7
Abbreviations .....	8
1. Introduction.....	9
1.1. The Urban Heat Island Effect.....	9
1.2. Problem Statement .....	11
1.3. Research aim and questions.....	11
1.4. Methodology .....	12
1.4.1. A literature review of the current state of the art in Urban Heat Island studies and models	12
1.4.2. Sensor data retrieval, analysis, cleaning and storage in a spatial database .....	12
1.4.3. Collection, aggregation and analysis of additional relevant spatial data.....	13
1.4.4. Calculation of the spatial input parameters for the regression models .....	13
1.4.5. Definition of the relationship between the spatial characteristics and the temperatures.....	13
1.4.6. Recursive building and validation of the alternative models.....	14
1.4.7. Analysis and comparison of the obtained results from the models .....	14
2. Related work .....	17
2.1. Challenges in modeling the UHI effect.....	17
2.2. Mapping of the UHI effect using satellite and sensor data.....	18
2.3. Research on the UHI contributing factors.....	20
2.4. Spatio-temporal variability of UHI.....	22
2.5. Statistical modeling .....	23
3. Sensor data.....	27
3.1. Sensors setup .....	27
3.2. Database building.....	28
3.3. Temperature data analysis.....	29
3.4. Aggregation .....	37
4. Spatial modeling.....	41
4.1. Space division .....	41
4.2. Calculation of the spatial indicators.....	43
4.2.1. Buildings density.....	43

4.2.2.	Land cover ratio.....	46
4.2.3.	Non-permeable surfaces index .....	47
4.2.4.	Vegetation coverage .....	49
4.2.5.	Sky View Factor .....	51
4.2.6.	Vehicle traffic density.....	53
4.3.	Results comparison and conclusions.....	55
4.4.	Spatial join .....	59
5.	Statistical modeling .....	62
5.1.	Analysis of the independent variables .....	63
5.2.	The first stage of modeling.....	66
5.3.	Improvement of the models .....	70
5.4.	Final model.....	73
6.	Conclusions and discussion .....	77
6.1.	Answers to the research questions .....	77
6.2.	Conclusions.....	79
6.3.	Discussion .....	81
	References:.....	84
	Appendix 1: Maps of calculated spatial indicators.....	88
	Appendix 2: Database creation .....	97
	Appendix 3: Python Bokeh Plots .....	99
	Appendix 4: Reports from the final statistical models.....	102

# List of Figures

FIGURE 1: METHODOLOGY OF THE RESEARCH .....	15
FIGURE 2: MAIN TESSELLATION OF THE URBAN ATMOSPHERE.....	18
FIGURE 3: (A) THE NETATMO WEATHER STATION; (B) SPATIAL EXTENT OF THE NETATMO SENSORS .....	27
FIGURE 4: EXCERPT FROM THE POSTGRES SPATIAL DATABASE .....	29
FIGURE 5: YEARLY WEATHER STATISTICS (HOEK VAN HOLLAND) .....	29
FIGURE 6: (A) TEMPERATURE VS. TIME (19-06-2017); (B) REFINED DATA .....	30
FIGURE 7: (A-G) PATTERNS OF TEMPERATURES IN THE HAGUE FOR THE HOTTEST DAYS (TEMPERATURES VS. TIME) .....	33
FIGURE 8: (A-B) TEMPERATURE PATTERNS FOR 2 REGULAR SUMMER DAYS (TEMPERATURE VS. TIME).....	34
FIGURE 9: (A-N) FRAMES FROM THE DYNAMIC VISUALIZATION OF THE SPATIAL PATTERN OF THE TEMPERATURE CHANGE (03-08-2017) .....	36
FIGURE 10: (A-N) FRAMES FROM THE DYNAMIC VISUALIZATION OF THE SPATIAL PATTERN OF THE TEMPERATURE CHANGE (03-08-2017) .....	37
FIGURE 11: (A-G) TEMPERATURES VS. TIME (AGGREGATED DATA FOR THE HOTTEST 7 DAYS).....	39
FIGURE 12: (A) RECTANGULAR TESSELLATION; (B) HEXAGONAL GRID; (C) VORONOI DIAGRAM .....	42
FIGURE 13: (A) AHN3 SPATIAL EXTENT; (B) AHN3 RESOLUTION .....	44
FIGURE 14: CALCULATION METHOD OF THE BUILDINGS DENSITY.....	45
FIGURE 15: CALCULATION METHOD OF THE LAND COVER INDEX .....	46
FIGURE 16: (A) PAVED OPEN AND TOTAL OPEN AREAS; (B) PAVED OPEN AND TOTAL OPEN AREAS (DETAIL) .....	48
FIGURE 17: CALCULATION METHOD OF THE NON-PERMEABLE SURFACES INDEX.....	48
FIGURE 18: NDVI MAP .....	50
FIGURE 19: RECLASSIFIED VEGETATION RASTER.....	51
FIGURE 20: VEGETATION INDEX CALCULATION METHOD .....	51
FIGURE 21: (A) SKY VIEW FACTOR;(B) SKY VIEW FACTOR (DETAIL).....	52
FIGURE 22: SVF CALCULATION METHOD .....	52
FIGURE 23:(A) ROAD NETWORK (LINES); (B) ROAD NETWORK (POLYGONS).....	54
FIGURE 24:VTD CALCULATION METHOD .....	54
FIGURE 25: (A) BUILDINGS DENSITY (HEXAGONAL GRID); (B) BUILDINGS DENSITY (RECTANGULAR GRID); (C) BUILDINGS DENSITY (VORONOI DIAGRAM).....	55
FIGURE 26:(A) LAND COVER INDEX (HEXAGONAL GRID); (B) LAND COVER INDEX (RECTANGULAR GRID); (C) LAND COVER INDEX (VORONOI DIAGRAM).....	56
FIGURE 27: (A) NON-PERMEABLE SURFACES (HEXAGONAL GRID); (B) NON-PERMEABLE SURFACES (RECTANGULAR GRID); (C) NON-PERMEABLE SURFACES (VORONOI DIAGRAM) .....	57
FIGURE 28: (A) SKY VIEW FACTOR (HEXAGONAL GRID); (B) SKY VIEW FACTOR (RECTANGULAR GRID); (C) SKY VIEW FACTOR (VORONOI DIAGRAM).....	57
FIGURE 29: (A) VEGETATION COVERAGE (HEXAGONAL GRID); (B) VEGETATION COVERAGE (RECTANGULAR GRID); (C) VEGETATION COVERAGE (VORONOI DIAGRAM).....	58
FIGURE 30: (A) VEHICLE TRAFFIC DENSITY (HEXAGONAL GRID); (B) VEHICLE TRAFFIC DENSITY (RECTANGULAR GRID); (C) VEHICLE TRAFFIC DENSITY (VORONOI DIAGRAM).....	58
FIGURE 31: METHOD OF DIRECT SPATIAL RELATION .....	60
FIGURE 32: AREA-WEIGHTED AVERAGED INDICATORS .....	60
FIGURE 33: (A) MORAN'S I SCATTERPLOT, (B) REFERENCE DISTRIBUTION .....	64
FIGURE 34: LOCAL INDICATORS OF SPATIAL ASSOCIATION (LISA) (A) CLUSTER MAP; (B) SIGNIFICANCE MAP .....	64
FIGURE 35: (B) SCATTERPLOT MATRIX (VORONOI DIAGRAM) .....	66
FIGURE 36: (A-F) T-SCORES AGAINST BUFFER RADII .....	72

# List of Tables

TABLE 1: UHI CONTRIBUTING FACTORS AND THEIR CALCULATION METHODS .....	21
TABLE 2: BUILDINGS DENSITY INDICATOR INFORMATION .....	43
TABLE 3: LAND COVER INDEX INFORMATION .....	46
TABLE 4: NON-PERMEABLE SURFACES INDEX INFORMATION .....	47
TABLE 5: VEGETATION INDEX INFORMATION.....	49
TABLE 6: SKY VIEW FACTOR INFORMATION .....	51
TABLE 7: INFORMATION ABOUT THE VEHICLE TRAFFIC DENSITY INDEX.....	53
TABLE 8: COMPARISON OF MODELS' PERFORMANCE .....	67
TABLE 9: RESULTS OF THE STATISTICAL TESTS .....	68
TABLE 10: INDEPENDENT VARIABLES COEFFICIENTS AND SIGNIFICANCE .....	69
TABLE 11: MODEL'S PERFORMANCE .....	73
TABLE 12: STATISTICAL TESTS .....	74
TABLE 13: COEFFICIENTS AND SIGNIFICANCE OF THE EXPLANATORY VARIABLES .....	75

## Abbreviations

<b>AHN</b>	Actueel Hoogtebestand Nederland
<b>DSM</b>	Digital Surface Model
<b>GIS</b>	Geographic Information System
<b>GWR</b>	Geographically Weighted Regression
<b>RA</b>	Regression Analysis
<b>UHI</b>	Urban Heat Island
<b>CUHI</b>	Canopy Urban Heat Island
<b>OLS</b>	Ordinary Least Squares
<b>BD</b>	Buildings density
<b>LC</b>	Land Cover
<b>NonPS</b>	Non-permeable Surfaces
<b>SVF</b>	Sky View Factor
<b>VTD</b>	Vehicle Traffic Density
<b>KNMI</b>	Koninklijk Nederlands Meteorologisch Instituut
<b>NDVI</b>	Normalized Difference Vegetation Index

# 1. Introduction

## 1.1. The Urban Heat Island Effect

The continuously increasing population within cities imposes the future challenges related to planning and managing the sustainable environment, where people's health and wellbeing is prioritized. Currently more than half of the world's population lives in cities (O'Malley, Piroozfarb, Farr, & Gates, 2014) which results in the rise of the human footprint, affecting the local climate. Simultaneously the planet's climate is changing towards less predictable weather conditions with high extremes, e.g., heavy rainfalls, followed by long dry periods, or severe heatwaves (Huber & Gullledge, 2011). Frequently extreme weather conditions are associated with flooding and hurricanes, while heatwaves represent an equally important danger for the health of the city's citizens.

In the last ten years, extreme heat conditions were observed more frequently, which provided more information about the related negative effects on the population and the economy. An example is a heatwave, which affected Paris in August 2003 (Hoeven & Wandl, n.d.). During this event the elderly population within the city was affected by the heat, peaking at eight times more deaths compared to an average number of deaths for the year. Studies mention that up to 14 800 people died during this period because of the heat. Furthermore, it was observed that public authorities and structures are not well prepared to respond adequately to such events, resulting in chaos and low-quality services.

In relation to the previously mentioned phenomena, a lot of studies are investigating the Urban Heat Island (UHI) effect within different major cities in the world. UHI has been recognized as one of the leading environmental issues of the 21<sup>st</sup> century cities (Memon, Leung, & Chunho, 2008). The UHI is defined as the area within a city with the higher surface or air temperatures compared to its surroundings (South, Working, & October, 2008). The higher urban temperatures are resulting in health-related issues among the population, higher energy demands, and various economic losses. Therefore UHI studies are important for the analysis and the mitigation of this phenomenon.

There are two main types of UHI effects recognized, namely surface and atmospheric. Surface UHI refers to the higher temperatures of the surfaces within the city than its surroundings and the atmospheric UHI is related to the air temperatures in the built environment. These two phenomena have different characteristics and are observed with the means of different techniques. The atmospheric UHI, for example, is mainly a nocturnal phenomenon, this means that it is more pronounced at night. The reason for that is the slow release of the stored during the day heat from the city's man-made surfaces like roads, buildings, parkings, etc. The atmospheric UHI is measured using in situ sensors – these could be for example weather stations, small distributed sensors

measuring temperatures outside buildings or car traverses. In contrast, the surface urban heat island has been measured using remote sensing techniques like satellite imagery or airborne sensors. In this case, the temperatures of the city's surfaces are of interest. The focus of the current work is on the Canopy Urban Heat Island, which refers to the air temperatures of Urban Canopy Layer – the layer between the ground and the mean height of the buildings within the city. This type of UHI has a direct influence on the human's health since this is where people are living.

In order for the UHI effect to be studied in detail, the contributing factors have to be identified. Broadly the UHI generating factors can be classified as controllable and uncontrollable (Memon et al., 2008). The uncontrollable factors are the ones, which are related to climate properties, such as airspeed, cloud coverage, solar radiation, etc. On the other hand, controllable are the factors describing the structure of the city – green areas, buildings volume, sealed surfaces, types of material that have been used, Sky View Factor, anthropogenic heat and others. All these factors have a different influence on the UHI effect. For example, the heat, which is generated by the anthropogenic activities (automobiles, air-conditioners, etc.), is being released to the environment directly and instantly, while only part of the solar radiation heats up the environment directly because man-made structures such as buildings, roads and roofs are absorbing and storing it. Thus, the whole process of surface energy balance has to be followed, which is based on the principle that energy cannot be lost (Nunez & Oke, 1980). Therefore, the total energy that the sun radiates to the city's surfaces is equal to the energy that is exchanged between the surfaces, the air, the water and the vegetation (sensible heat, evaporation and transpiration).

$$Q^* = QE + QH + QS$$

$Q^*$  - Net solar radiation, received by the earth surface

$QE$  – Energy exchanged through evaporation (by water and greenery)

$QH$  – Sensible heat (conversion of heat from the surface to air)

$QS$  – Energy absorbed by the ground, buildings and surface water

Thus, the current work is focusing on researching the air temperatures in the urban canopy layer, which is the result of the heat radiated from the urban surfaces during the night. There are differences in the heat exchange phases between the neighborhoods, which is caused by the heterogeneity of their morphological characteristics. Such information can be of great use for the development of UHI mitigation strategies, but it is currently not available. Therefore this master thesis will investigate the possibility for generation of more detailed models, taking into account the intraurban variability.

## 1.2. Problem Statement

The problem of Urban Heat Islands has been widely recognized, therefore the scientific interest of the current work is to gain a deep understanding of the pattern of this phenomenon. In order for the UHI effect to be effectively tackled, its variability within the city and the factors that define this pattern have to be identified. This can be achieved by using highly detailed UHI models, which is the main focus of the current work. Currently, there is lack of such models for the whole city. Some studies provide such information but for limited spatial extent – usually for a maximum of few buildings up to a neighborhood. On the other hand, a lot of research has been conducted on a city scale, but with very coarse models, which prove the existence of UHI effect in the study area but can't provide more information of its intraurban variability and the weight of the contributing factors. The satellite data that has been mostly incorporated in such studies provide an overview of the conditions of the area of interest, but it has poor spatial and temporal resolution. Hence, the problem that this work is focusing on is bridging the gap between the highly detailed small-scale models and the coarse city scale ones.

The challenge of the current work is to provide the urban planners with a tool, with which they can address the Urban Heat Island problem effectively. This can be achieved with the means of highly detailed data, which will provide inside in the relevant urban characteristics, causing the higher temperatures in the cities. In this way, urbanists will be able to customize their actions, depending on the city's location and its neighborhoods properties.

## 1.3. Research aim and questions

The aim of the current work is to model the atmospheric Urban Heat Island variability, based on the measurements of the Netatmo sensors in the city of The Hague for the year of 2017. In order to achieve this, different statistical modeling methods will be analyzed and compared. In addition, the main UHI contributing factors will be derived based on the morphological characteristics of the city and their correlation to the UHI temperatures and the statistical modeling significance will be analyzed. The scope of the work is limited to the temperatures within the urban canopy layer of the city, therefore surface temperatures will not be considered in this study. In order to achieve this, the following main research question and the corresponding sub-questions have been defined:

## **How to accurately model the spatial and time variability of the Canopy Urban Heat Island (CUHI) effect in the city of The Hague?**

- 1. How to collect, clean and validate the raw sensor data?**
- 2. Which factors or combination of factors influences the UHI effect at most?**
- 3. What level of detail is needed to model the UHI variability?**
- 4. How does the UHI effect variate in the different parts of the city during the day and night?**

The goal of the current research is highly connected with the Geomatics field of knowledge because it covers the whole process from spatial data acquisition, storage, processing, modeling and representation. For the sake of this scientific research, data from over 200 sensors, deployed in the area of The Hague has been used. These data have been continuously gathered and stored in a server property of the TU Delft. Broadly the methodology that is going to be followed will consist of retrieving the data, cleaning and storing it in a spatial database, then these sensor data will be used in combination with another relevant spatial information in the process of modeling the Urban Heat Island effect in The Hague.

### **1.4. Methodology**

#### **1.4.1. A literature review of the current state of the art in Urban Heat Island studies and models**

In this step, a comprehensive literature review has been performed, which leads to the definition of the main contributing factors to the UHI effect. These factors are further incorporated in the definition of the spatial model and the regression models as well. The literature review provides an overview of the spatial, statistical and analytical techniques that have been used in the process of the Urban Heat Island study. Further, the technological and terminological basis of the current work is built upon the preceding studies in the field of Urban Heat Island mitigation.

#### **1.4.2. Sensor data retrieval, analysis, cleaning and storage in a spatial database**

The second phase of this research will focus on the processing of the temperature data and namely the sensor measurements have been retrieved, processed and stored in a spatial database. This is particularly important due to the enormous amount of data, which is stored in separate files for every timestamp, resulting in almost 20 000 files for April - December 2017. Therefore, the storage of these files in a spatial database provides the opportunity to combine, process and query the data. In this way, the spatial extent of the temperature data can be refined and the hottest days to be extracted.

Further, these raw measurements have been analyzed regarding the errors and outliers that they contain, which can be a result of system failures or wrong positioning of the sensors (e.g., under direct sun radiation).

#### 1.4.3. Collection, aggregation and analysis of additional relevant spatial data

Further, the spatial data needed for the representation of the Urban Heat Island contributing factors have been collected. The data are analyzed and aggregated to the area of interest. Some of the parameters are retrieved after spatial processing operations like the Normalized Difference Vegetation Index for example, which has been extracted from satellite imagery. This index provides information about the green areas, water and ground surfaces within the city. Another important indicator is the Sky View Factor, which is calculated from Digital Elevation Models using GIS processing tools. Thus, in this step of the methodology the input data for the spatial model, representing the UHI explanatory characteristics, are prepared.

#### 1.4.4. Calculation of the spatial input parameters for the regression models

Following is the calculation of the explanatory variables, which are the input parameters for the regression analysis. The choice of these factors is based on the preceding literature review, where the main UHI contributing factors and their calculation methods are defined. In this phase, the collected in the previous step spatial data will be processed using Geographical Information Systems software, resulting in a set of values for every explanatory variable. The most important characteristics are the Buildings Density explained as the ratio between the buildings volumes and the total area, followed by the Land Cover which incorporates the areas of the planar surfaces of the buildings, the Sky View Factor, the Vegetation areas, the Non-permeable surfaces areas and the Vehicle traffic density. All of these indicators are calculated in the GIS software, based on different datasets. The calculation methods and the datasets used are further explained in the Spatial modeling section.

#### 1.4.5. Definition of the relationship between the spatial characteristics and the temperatures

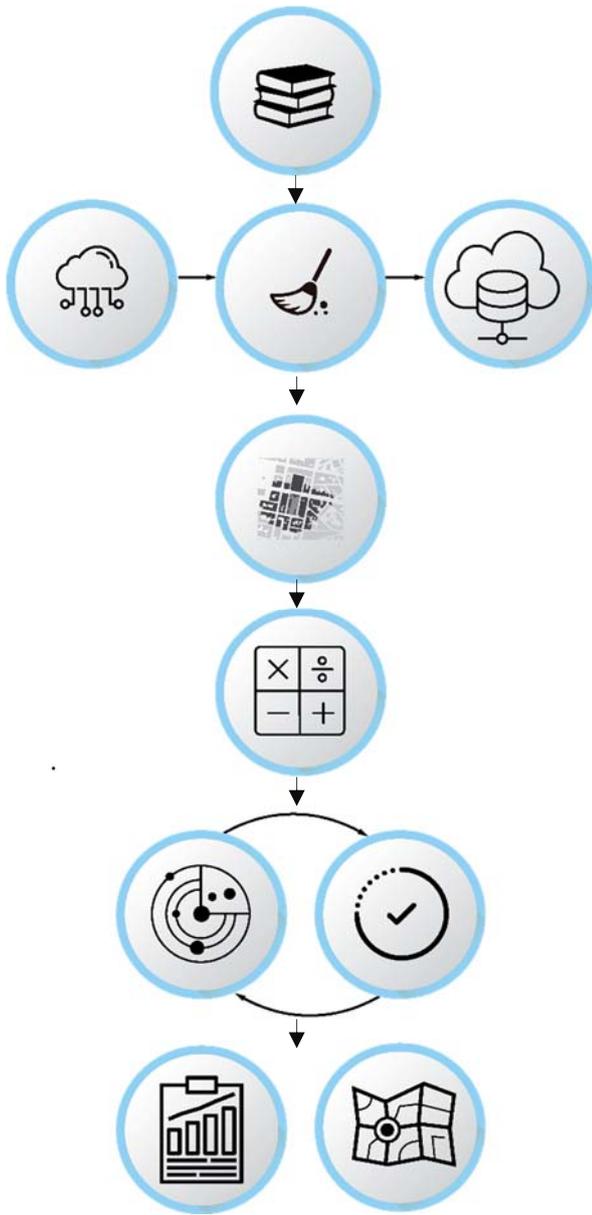
After obtaining all crucial spatial indicators, the relationship between them and the temperatures has to be defined in terms of the spatial model. Different ways of spatial association are explored in this step, resulting in different types of spatial joins between the aggregated temperatures and the cities characteristics. The output from this step is crucial and it essentially represents all dependent and independent variables, which are defining the input for the statistical analysis.

#### 1.4.6. Recursive building and validation of the alternative models

The following methodological phase consists of the creation and validation of the statistical models. During this phase, the difference between the ordinary and the spatial regression models is explored. The interest is focused on the performance of the spatial models and how well the Urban Heat Island phenomenon can be explained. The models use the Vegetation coverage, the Non-permeable surfaces, the Building density, the Land cover ratio, the Vehicle traffic density and the Sky View Factor as independent variables and the temperature difference between rural and urban areas as the dependent variable.

#### 1.4.7. Analysis and comparison of the obtained results from the models

Finally, the obtained models are statistically analyzed and compared. The results of the regression models are visualized and further analyzed, resulting in the definition of the most important for the UHI effect factors and explanation of the intraurban temperature variability. The comparison between the models provides information about the advantages and disadvantages of the different techniques and their appropriateness for the explanation of the Urban Heat Island phenomenon. Also, the statistical analysis indicates the correlation between the chosen variables and the Urban Heat Island effect.



- (1) Literature review of the current state of the art in Urban Heat Island studies and models.
- (2) Sensor data retrieval, analysis, cleaning and storage in a spatial data base.
- (3) Collection, aggregation and analysis of additional relevant spatial data.
- (4) Calculation of the spatial input parameters for the regression models.
- (5) Recursive building and validation of the alternative models.
- (6) Analysis and comparison of the obtained results from the models.

Figure 1: Methodology of the research



# LITERATURE REVIEW

## 2. Related work

This section provides an overview of the relevant literature, explaining the essence of the Urban Heat Island effect and its relationship with the structure of the city. Furthermore, review of the related acquisition and modeling methods has been provided.

### 2.1. Challenges in modeling the UHI effect

Considering the complexity of the UHI phenomenon different studies are approaching the problem in various ways, employing different design complexity and scale of application. The main challenge that has been recognized in the process of studying the UHI effect is to take into account the complexity of the phenomena, which occur in the city (Mirzaei, 2015). The intricacy of the urban physics is rooted in the fact that these processes interact at a different scale – from the human body to city scale, therefore the important first step is the simplification of the problem according to the goal of the study. As an example, some of the widely investigated UHI topics are an evaluation of mitigation strategies, improvement of citizen's comfort and health, building energy demand and others (Mirzaei, 2015). Considering this diversity of topics, one can understand the variety of UHI models, which can be observed.

Depending on the scale of the research area, UHI studies can be divided into three main classes – building-scale models, micro-scale models and city-scale models (Mirzaei, 2015). The first class is characterized by the largest detail of representation where a big amount of complex factors are incorporated, but usually, these models are highly limited in terms of spatial extent. In such models, multiple different alternatives for the development of the buildings are investigated, where the focus is on the influence of the architecture on the temperatures in the study area. Further, the micro-scale models are investigating the microclimate in a certain area and the interaction of a building with its environment. Usually, these models are incorporating complex data such as air flow patterns, which are modeled using computational fluid dynamics techniques (CFD). In general, these models can be utilized in the investigation of different parameters of the city's structure, such as buildings orientation, materials, sky view factors, vegetation and others. Despite the greater complexity of factors, which these models incorporate, they are still limited in their spatially. In the third class of models – city-scale models – one can observe the variation of the UHI effect. These models are mainly used for the development of mitigation strategies and their evaluation. Compared to the micro-scale models the city-scale ones incorporate fewer factors and model the temperatures in the cities with less detail. It is important to mention that the city-scale models often lack resolution quality because they have been applied on very coarse cells, which eventually affects the possibility of one to observe the interaction between the buildings and their environment.

## 2.2. Mapping of the UHI effect using satellite and sensor data

The definition of Urban Heat Island – increased temperatures of the surface or the air within the city, compared to the rural areas (Gomez, 2011) – implies that there are different types of heat islands that can be observed in the city. This differentiation is based on the theoretical division of the space to surface, canopy and boundary layers (see Figure 2).

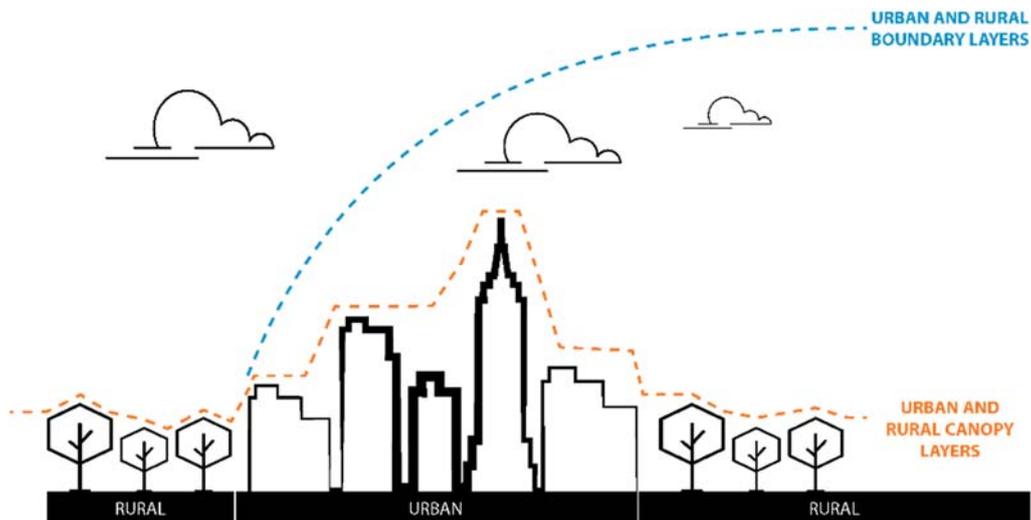


Figure 2: Main tessellation of the urban atmosphere.

This tessellation of the space refers to the different physical phenomena, which have been studied – the Canopy UHI (CUHI) and the Boundary Urban Heat Island (BUHI) are related to the higher air temperatures within the city. On the other hand, the Surface Urban Heat Island refers to the higher temperatures of the buildings, roads and other man-made surfaces within the city. The urban canopy layer is defined as the layer between the ground and the mean height of the buildings within the city or in the rural areas. The boundary layer is theoretically positioned above the canopy layer up to 1 km in the atmosphere (Gomez, 2011).

The Canopy Urban Heat Island is the most studied one because of its direct relevance to the people's health, therefore CUHI is the type that is mostly discussed concerning the UHI topics. There is an important difference in the behavior of the surface and air temperatures in the city – the atmospheric urban heat island is mainly a nocturnal phenomenon (it becomes more pronounced during the evenings, because of the slow release of heat from the buildings and surfaces). Therefore the CUHI has been mostly studied by measuring the temperatures of the air after sunset, when the maximum intensity of this effect can be measured (van Hove et al., 2015).

Depending on the different types of UHI two main acquisition methods have been used – remote sensing or ground-based temperature measuring stations. The main difference between the two methods lies in their spatiotemporal characteristics - ground-based measuring has the advantage of

high temporal resolution, but on the other hand, it has a poor spatial resolution, while on the contrary, remotely-sensed data has a higher temporal resolution, but the data describes one point in the time.

Regarding the remote sensing acquisition techniques, the study conducted by Hardy and Nel for the city of Johannesburg (Hardy & Nel, 2015) describes a methodology based on multisensorial multi-temporal remote sensed data collected for the purpose of building a predictive model, based on multiple influencing factors. In this study, the authors recognize the high importance of the type of urban landscape regarding the temperatures within the city. The different surfaces have different properties – e.g., thermal characteristics, placement, and others – and they are primarily described by the land use types in the urban area. Furthermore, this study relies on the cloud-free thermal data acquired from Landsat 5 TM, Landsat 7 ETM+, and ENVISAT AATSR. The Landsat thermal bands were converted to temperature maps and unsupervised k-means classification has been performed in order to obtain the underlying land-use classes. In addition, the mean surface temperature values were calculated from the data in order to minimize the cloud contamination. Then the UHI intensity was calculated by subtracting the mean temperatures from the rural areas for every type of urban land use class. The results showed that there is a considerable difference in the temperatures between the areas with high presence of vegetation compared to the densely populated built-up areas. This study describes a suitable methodology for studying the UHI effect for areas with a sparse availability of spatial data, where satellite data could provide the needed information.

The other acquisition method for studying the UHI is based on the data provided by ground-based sensors. The challenge which this technique presents lies in the fact that these measurements give information about the temperatures within the city for a limited number of specific locations, therefore the performance of the interpolation method to be used plays crucial role for the quality of the results. Szymanowski and Kryza (Szymanowski & Kryza, 2009) investigate different interpolation algorithms and evaluate them based on the statistical errors of the results and their visual representation. In addition to the temperature data derived from mobile traverses and ground-based weather stations, supplementary spatial data related to the UHI effect was incorporated. These data consist of information about the artificial and the non-built-up areas, the Normalized Difference Vegetation Index (NDVI), the roughness length, the thermal admittance (estimated on the ratio between vegetated and artificial surfaces) and the anthropogenic heat emission. These parameters were then approximated in relation to the land use class that they belong to, resulting in a set of values, which were used for the replacement of the land use categories. For this study, stochastic and deterministic interpolation methods were applied. The other classification of interpolation techniques divides them into global and local methods. The difference between them is that the global methods use all available data to predict the phenomenon, while the local ones use data from the direct

neighborhood of the point, which is going to be estimated. Another subdivision of these methods considers if they preserve the measured values or not. Finally, one can consider using the multidimensional interpolation techniques, which are incorporating additional explanatory data. The methods that were compared in this study are Inverse Distance Weighting (IDW), Regularized Spline with Tension (RST), Ordinary Kriging (OK), Multiple Linear Regression (MLR) and Residual Kriging (RK). As a result, the RK and OK methods were statistically more accurate, while with IDW and MLR the largest errors were observed. Further, it was observed that the IDW method overestimates the temperatures within the city. The preceding overview highlights the importance of the choice of interpolation algorithms for the results of the UHI modeling. Important to be considered is the diverse characteristics of these methods, which affect the methodology of the study, for example with some interpolation techniques additional data can be incorporated, while with others this is not possible.

### 2.3. Research on the UHI contributing factors

An important part of the UHI studies is the identification of the explanatory factors of this phenomenon. The knowledge about these factors provides the opportunity for adequate mitigation measures to be taken as well as more accurate prediction models to be built.

In their study, Stanganelli and Soravia are investigating the connections between the urban structure and the UHI effect. The indicators that have been used in the correlation analysis between the urban structure and the temperatures are the non-permeable surfaces index (ratio between paved and open areas), the percentage of green areas, the land cover ratio and the density and height of the buildings (Stanganelli, Marialuce, and Soravia, 2012). The main indexes have been calculated based on the relationships between the relevant areas and the total area. Consequently, each of these indicators was correlated to the measured temperatures within the city. It has been concluded that there is a direct relationship between the land cover ratio and the temperatures – with the increase of the land cover ratio the temperatures are increasing as well. On the other hand, green areas have a strong positive influence on the temperatures in the cities. Furthermore, the increase in the average height of the buildings leads to the increase in the temperatures as well. This can be explained by the decreased ability for longwave radiation during the night in densely populated areas.

Another important factor connected with the characteristics of the urban environment is the type of materials that have been used. Depending on their technical characteristics the materials absorb and reflect the radiated heat differently from the sun and thus affect the thermal comfort conditions of the buildings and the open spaces in a different way. Therefore, many studies have been conducted on the so-called cool materials, i.e., materials with high solar reflectance and infrared emittance properties. The reflectance is the ability of a material to reflect the solar radiation while the infrared

emittance measures the ability of a surface to release heat (Santamouris, Synnefa, & Karlessi, 2011). The beneficial applications of such materials for the mitigation of the UHI effect is for the coverage of roads, roofs and other surfaces in the city and this is another important indicator that can be incorporated in the analysis of the UHI effect generation.

From all previously mentioned factors affecting the UHI intensity within the cities, the presence of greenery has been defined as one of the best mitigations strategies that should be incorporated in the planning process. Considering the positive effect of the green areas, the extent of their influence has been studied. In their study, Luan et al. (Luan, Ye, Liu, & Li, 2014) use satellite imagery to retrieve the surface temperatures and the extent of the green areas in the city of Beijing with the goal of quantifying the influence area of the urban green lands on the surrounding buildings. The urban green lands were identified by calculating the Normalized Difference Vegetation Index (NDVI) and parameters like area, perimeter and shape were incorporated into the analysis. Further, these data were spatially overlaid with the available temperature data and statistically analyzed. It was observed that the vegetation significantly influences the cooling effect in the city, but in order for this effect to be quantified, analysis on the distance of influence was performed. For this purpose, buffer zones of different distances from the green area were created and the temperatures within the different areas were compared. As a result, it has been observed that the green areas are affecting the surrounding buildings only in the first 100m distance, which means that outside this area the greenery does not provide its temperature relieving effect. Furthermore, no significant evidence about the correlation between the green land parameters (area, perimeter, shape, etc.) and the temperature relief has been found.

In the following table (Table 1), a summary of the different UHI contributing factors and their calculation methods is provided.

Table 1. UHI contributing factors and their calculation methods

Factor	Calculation method	Study
Vegetation coverage	$f = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}$ <p> <math>NDVI_{min}</math> – min NDVI value for area  <math>NDVI</math> – NDVI value for each pixel  <math>NDVI_{max}</math> – max NDVI value for area </p>	(Luan et al., 2014)
Solar Reflectance Index	$SRI = \frac{T_{black} - T_{surface}}{T_{black} - T_{white}} * 100$	(Santamouris et al., 2011)

	$T_{black}$ – state temp. of black surface $T_{white}$ – state temp. of white surface $T_{surface}$ – state temp. of material surface	
Non-permeable Surfaces Index	$\frac{Paved\ open\ areas}{Total\ open\ areas}$	(Stanganelli, Marialuce, and Soravia, 2012)
Land cover ratio	$\frac{Built\ surface}{Land\ area}$ <i>Built surfaces</i> = sum of all building plots	(Stanganelli, Marialuce, and Soravia, 2012)
Buildings density	$\frac{Buildings\ volumes}{Entire\ area}$	(Stanganelli, Marialuce, and Soravia, 2012)
Anthropogenic heat	$VTD = SUM \left( \frac{1}{distance} * length \right)$ <i>VTD</i> – vehicle traffic density <i>Length</i> – road length <i>Distance</i> – distance to the road	(Yang & Chen, 2016)
Sky View Factor	the ratio between radiation received by a planar surface and that from the entire hemispheric radiating environment	(Svensson, 2004)

## 2.4. Spatio-temporal variability of UHI

The spatial variation of the urban microclimate and the UHI contributing factors have been studied by Yang and Chen (Yang & Chen, 2016). In their approach towards building a thermal atlas of spatial variations of the urban microclimate as a design decision-support tool, the authors incorporate urban morphological variables in the modeling process. They have recognized the role of the empirical models built upon carefully selected morphological variables, which can be used as an explanatory tool of the intraurban thermal variations. The set of variables that have been used consists of data about the land cover, the buildings form and density, the anthropogenic heat, the green areas and the proximity to heat sinks. Temperature data was retrieved from fixed weather stations for the period of 29 – 30<sup>th</sup>, July 2013. The methodology that has been followed consists of GIS-based spatial analysis

tools used for the calculation of the influence parameters and statistical methods like Pearson correlation and regression analyses. The results showed a considerable correlation between the measured influence factors and the physiological equivalent temperature (PET). For example, the Sky View Factor was positively correlated with the PET, accounting for more than 50% of the variations in the index. Contrarily the vegetation in the city showed opposite correlation than the expected, where green areas were hotter than the surrounding areas. Additionally, based on the Pearson correlation analysis the PET index was significantly related with the Sky View Factor, the Green plot ratio, the Frontal density area and the Proximity to heat sink at 0.01 significance level, while the Vehicle traffic density has been found not significant for the PET index. Further, a regression analysis of the variables and the PET index have been performed. For this, the researchers used the stepwise regression method which removes the insignificant variables progressively from the model. Thus, the final model included the Sky View Factor, the Green plot ratio and the measured temperatures as independent variables and the PET index as the dependent variable. The model explained 76% of the variability in the index.

The importance of studying the spatial and temporal variability of the UHI effect has been recognized by van Hoven et al. in their study (van Hove et al., 2015). The increased temperatures in the cities lead to many problems related to the quality of the life of the citizens and significant economic losses. Moreover, in comparison to other disasters such as floods, coastal erosion and others, the UHI effect has received the least attention from the research community. Also, it has been mentioned that there is lack of long-term observational data, which limits the quality of the performed studies. Further, it has been concluded that the intraurban variability of the UHI is strongly related to the buildings, the greenery, the sealed surfaces and the height of the buildings.

## 2.5. Statistical modeling

Regression analysis is a statistically based technique for exploration of the relationships between different phenomena. The simplest regression form models the relationship between a single pair of response and predictor variables. Since the aim is to explore the trend between two variables in mathematical terms this is done by defining the best line of fit, which can be expressed as follows:

$$y = \beta_0 + \beta_1x + u \quad (\text{Eq. 1})$$

This simplest form of linear regression is also called bivariate linear regression because it relates two variables – y (dependent variable) and x (independent or explanatory variable). The definition of the best fitting line is based on the exploration of the prediction errors, which have the smallest possible values. This is done by minimizing the squares of the prediction errors. (Wooldridge, 2013).

The Ordinary Least Squares (OLS) regression is the most common type of statistical model, where a linear relationship between one or multiple variables is developed. The OLS method is referred to as the Best Linear Unbiased Estimator (BLUE). Besides the simple OLS, there are different types of regression models depending on their main characteristics. OLS is a global model, where a single equation, explaining the overall relationship between the variables is built. The models can be divided into simple and multiple linear regressions, depending on the number of explanatory variables used. In contrast to the OLS, there are local statistical models which define single equations for all data points (Fotheringham, Brunson, & Charlton, 2002). Such models have been extensively used for modeling spatial data, where different phenomena and their explanatory processes have high variability in the space.

In the current work, four different types of models have been considered – OLS and three types of spatially varying statistical models (Spatial Lag, Spatial error and Geographically weighted regression). The spatial regression models are based on the standard linear regression with the difference that spatial dependence terms are added in the models' definition. There are two main approaches for the incorporation of the spatial dependence terms – in the form of the spatially lagged dependent variable ( $Wy$ , Eq.2) or in the error term ( $\varepsilon$ , Eq. 3) (Anselin, n.d.-a).

$$y = \rho Wy + x + \varepsilon, \text{ (Eq. 2 Spatial lag model),}$$

Where:

- $y$  is a vector of observations on the dependent variable;
- $Wy$  is a spatially lagged dependent variable for weights matrix  $W$ ;
- $x$  is a matrix of observations of the explanatory variables;
- $\varepsilon$  is the error term.

$$y = X\beta + \varepsilon, \text{ (Eq. 3 Spatial error model),}$$

Where:

- $y$  is a vector of observations on the dependent variable;
- $x$  is a matrix of observations of the explanatory variables;
- $\varepsilon$  is a vector of spatially autocorrelated error terms.

In addition to the two types of spatially varying statistical models, a local Geographically weighted regression model (GWR) is included. GWR (Eq.4) generates a separate regression equation for each observation. Each equation is calibrated using a different weighting of the observations contained in the data set (Brunson, Fotheringham, & Charlton, 2010):

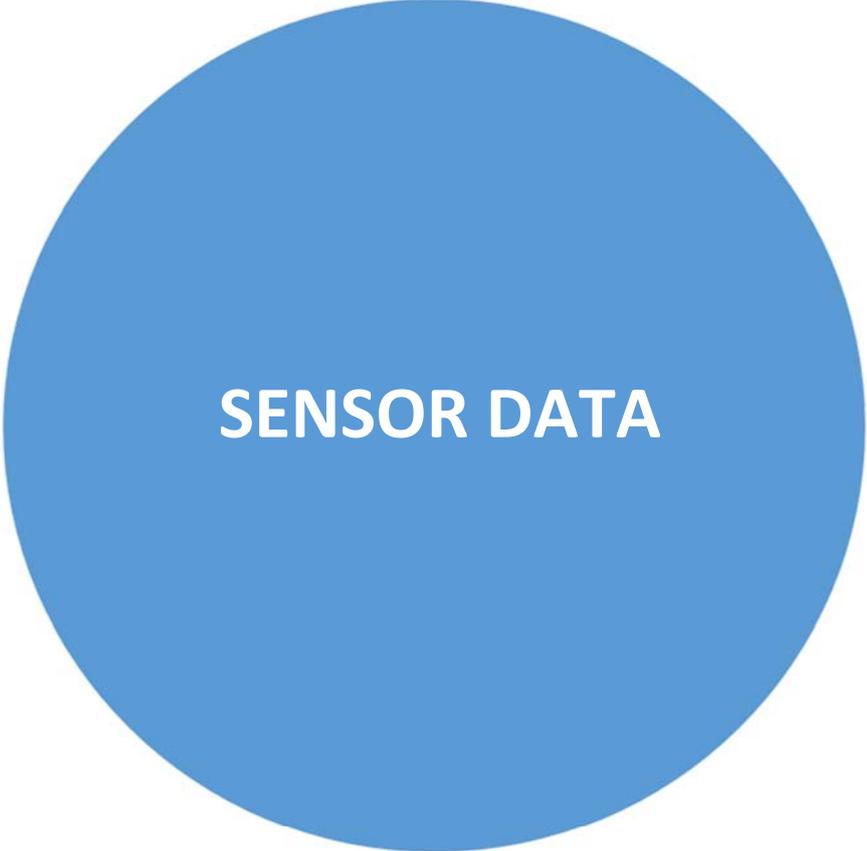
$$y_i = a_{i0} + \sum_{k=1,m} a_{ik}x_{ik} + \varepsilon_i, \text{ (Eq. 4 GWR),}$$

Where:

$y_i$  is the  $i$ -th value of the dependent variable;  
 $x_{ik}$  is the  $i$ -th observation of the  $k$ -th independent variable;  
 $\epsilon_i$  are the independent, normally distributed error terms;  
 $\alpha_{ik}$  is the value of the  $k$ th parameter at location  $i$ .

The main assumption behind the definition of the GWR is that observations are influenced by each other in the space, where closer ones have greater influence than the observations further apart. Therefore, a weight is assigned to each observation, which is based on a distance decay function centered at the current observation. The distance decay function, which may take a variety of forms, is modified by a bandwidth setting defining the extent of the function.

The development of the modeling methodology starts with the definition of the Ordinary Least Squares model, which is based on three main assumptions – the data sample is random and represents the general population, the sampling distribution is normal and the individual observations are independent. Based on these assumptions few main properties of the OLS models have to be investigated – the normal distribution of the random error (suggesting that there is no systematic misspecification or bias in the model), the constant variance of the errors (homoscedasticity) and the spatial independence of the variables and error terms. The reliability of the OLS regression model and the possibility to develop analysis and conclusions about the populations is strongly dependent on these prior assumptions. The violation of any of these assumptions can lead to bias, inefficiency in the regression estimates or unreliability of the confidence intervals, therefore any conclusions or statistical inference have to be carefully made. The notion of spatial dependency is another important impediment for the ordinary linear regression, which is carefully studied and incorporated in the spatial models. The presence of spatial dependency is confirmed when values at a location are affected by the neighboring locations. There are two main types of spatial dependence that can be observed in regression models - when the error terms across different observations are correlated (spatial error) or when the dependent variable is affected by independent variables from multiple locations (spatial lag). When a spatial error is observed in OLS regression, the assumption of uncorrelated error terms is violated, which results in inefficient estimates. On the other hand, if the OLS regression is affected by spatial lag, both assumptions of no correlation in the errors and the independence of the observations are violated, which results in bias and inefficiency of the model (Shoff, n.d.).



## 3. Sensor data

The following section provides an overview of the main dataset, which has been used for the development of the current research and namely the temperature measurements collected by the sensor network deployed in the area of The Hague. The data have been further processed and analyzed resulting in a temperature dataset, which represents the dependent variable in the developed statistical model.

### 3.1. Sensors setup

The temperature data has been collected from a network of small weather stations, deployed in the wider area of The Hague. The period that has been observed and analyzed covers the months from April until December 2017. The measurements are collected by the Netatmo weather stations (Figure 3a), which have been deployed for the purpose of analyzing the Urban Heat Island effect in The Hague (Hoeven & Wandl, 2018).

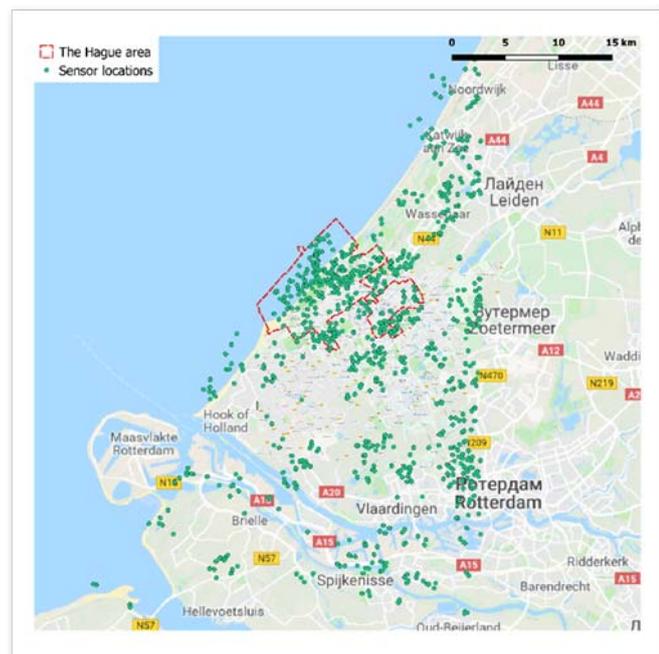


Figure 3: (a) The Netatmo weather station<sup>1</sup>; (b) Spatial extent of the Netatmo sensors

Over 850 small weather stations have been deployed with the help of citizens and volunteers. The 45x45x105 mm sensors have UV-resistant aluminum shell. The module is powered by batteries and it is equipped with WI-FI communication, which has been used for sending data to a remote server. The data is in comma separated format and consists of the id of the sensor, its position and altitude, the temperature, the humidity and the pressure and rain for the last day. The spatial extent of the raw

<sup>1</sup><https://www.netatmo.com/en-US/product/weather/>

dataset spreads out of The Hague area including cities like Leiden, Zoetermeer, Rotterdam and others (Figure 3b) and it is further refined to the area of interest and namely the area of The Hague.

Since the focus of the current work is on the UHI effect within the city, mainly the temperature measurements and the location of the sensors are included in the study, leaving the other attributes outside the scope of this research.

### 3.2. Database building

The whole dataset consists of roughly 22 thousand files for the period of April until December 2017, where every file represents a collection of all measurements for a time stamp (for around 870 different locations) and the frequency of measurements is roughly every 15 minutes. Considering the size of the data a spatial database was created, where the whole dataset was stored and therefore prepared for processing and analysis.

The database software, which was used is PostgreSQL. This is an open source object-relational database with full ACID compliance. In addition, it supports geographical data with its PostGIS extension, which provides the spatial data processing capabilities. This database is mainly used to maintain the enormous amount of sensor measurements, allowing their clustering in a single platform and therefore easier access to the needed information, querying and processing functionality. The version of the software that has been used is PostgreSQL 9.6.

In order to automatically process, parse and upload the great amount of data, a Python script has been developed. The script establishes an automatic connection to the local database and creates a table, where the data is stored. The fields, which were used for organizing the data are measurement id, sensor geometry, date, time, MAC, altitude, temperature, humidity, rain 60 min and rain 24h. Further, all files within a specified directory have been iterated and parsed. The location of every sensor has been defined as point geometry and stored in the database. Thus, the whole data for the year 2017 was organized in a single table (Figure 4) and could be further processed and analyzed using standard querying language for storing, manipulating and retrieving data in a database – Structured Query Language (SQL).

Data Output											
	measure... [PK] dou...	sensor geometry	date date	time time wit...	mac text	altitude double p...	tempera... double p...	humidity double p...	rain_60... double p...	rain_24h... double p...	pressure double p...
<input type="checkbox"/>	1	0101000...	2017-04-...	04:36:17	70:ee:50...	3	8.9	92			1020.2
<input type="checkbox"/>	2	0101000...	2017-04-...	04:36:17	70:ee:50...	2	8.9	89	0	0	1019
<input type="checkbox"/>	3	0101000...	2017-04-...	04:36:17	70:ee:50...	2	9.3	85			1025
<input type="checkbox"/>	4	0101000...	2017-04-...	04:36:17	70:ee:50...	4	9	87	0	0	1026.8
<input type="checkbox"/>	5	0101000...	2017-04-...	04:36:17	70:ee:50...	14	11.6	72			1024.4
<input type="checkbox"/>	6	0101000...	2017-04-...	04:36:17	70:ee:50...		11.6	74			1026.5
<input type="checkbox"/>	7	0101000...	2017-04-...	04:36:17	70:ee:50...	1	10.6	77	0	0	1024.1
<input type="checkbox"/>	8	0101000...	2017-04-...	04:36:17	70:ee:50...	6	9.9	82			1023.9
<input type="checkbox"/>	9	0101000...	2017-04-...	04:36:17	70:ee:50...	2	9.8	85			1028.5
<input type="checkbox"/>	10	0101000...	2017-04-...	04:36:17	70:ee:50...	2					1024.7
<input type="checkbox"/>	11	0101000...	2017-04-...	04:36:17	70:ee:50...	2	9.4	84	0	0	1023.6
<input type="checkbox"/>	12	0101000...	2017-04-...	04:36:17	70:ee:50...	1	9.2	82			1025.5
<input type="checkbox"/>	13	0101000...	2017-04-...	04:36:17	70:ee:50...	0	10.8	78			1027.2
<input type="checkbox"/>	14	0101000...	2017-04-...	04:36:17	70:ee:50...	0					1023.9

Figure 4: Excerpt from the Postgres spatial database

### 3.3. Temperature data analysis

Since the main aim of the current work is to study and model the Urban Heat Island effect in the city of the Hague, the analysis of the temperature data is targeted at the hottest days of 2017. According to the definition of the UHI effect, a period of 5 consecutive days, for which the maximum temperatures exceed 25<sup>o</sup> is qualified as a heat wave. In addition, for at least 3 of these 5 days, the temperatures are expected to be higher than 30<sup>o</sup>. According to this definition, an analysis of the temperatures for 2017 was performed based on the weather statistics for 2017. The Royal Dutch Meteorological Institute (Koninklijk Nederlands Meteorologisch Instituut, KNMI) provides historical weather data for the Netherlands. The overall temperature pattern for 2017 can be observed in Figure 5.

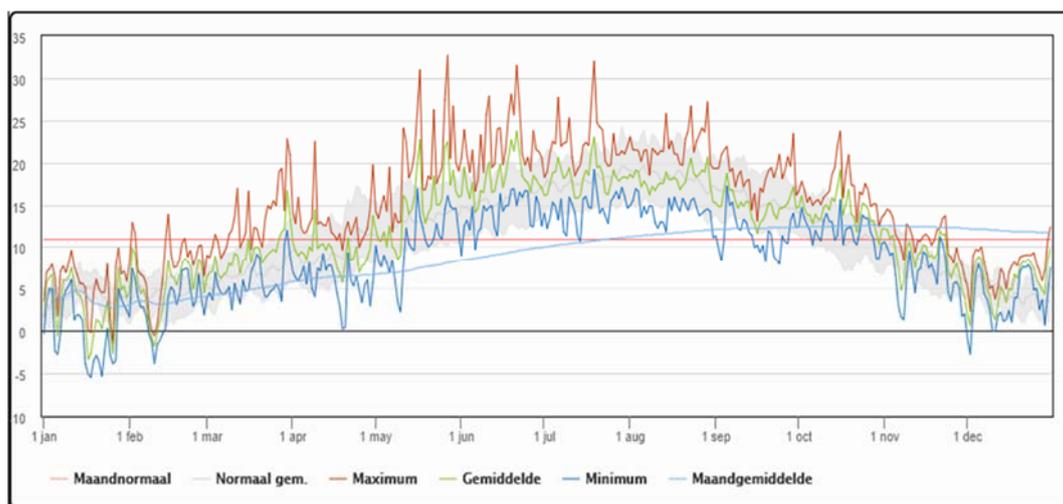


Figure 5: Yearly weather statistics (Hoek van Holland)<sup>2</sup>

<sup>2</sup> <https://weerstatistieken.nl/hoek-van-holland/2017>

The hottest days of the year were found to be in May and June, where temperatures above 30<sup>0</sup> were measured. Therefore, the spatial database containing all measurements was filtered resulting in separate tables for the hottest days. Due to a gap in the dataset, the final collection of the hottest days covers the period from the 26<sup>th</sup> to the 29<sup>th</sup> of May and from 18<sup>th</sup> to the 20<sup>th</sup> of June.

Further, in order to analyze the temporal and spatial pattern of the heating and cooling processes in the city, a set of analytical tools was developed. First, an open source Python library for interactive visualizations (Bokeh<sup>3</sup>) was used to create interactive plots, which can help in the analysis of the data. Based on the initial plots of the different datasets a data quality analysis was developed. As it can be seen in Figure 6a, the raw data contains systematic errors and outliers. In addition, the dataset contains a lot of empty measurements due to sensor failures. Therefore, all empty measurements, sensors yielding constant temperatures and extreme outliers were removed from the dataset based on the information provided from the plots (unique MAC addresses), resulting in refined visualization of the temperature patterns (Figure 6b).

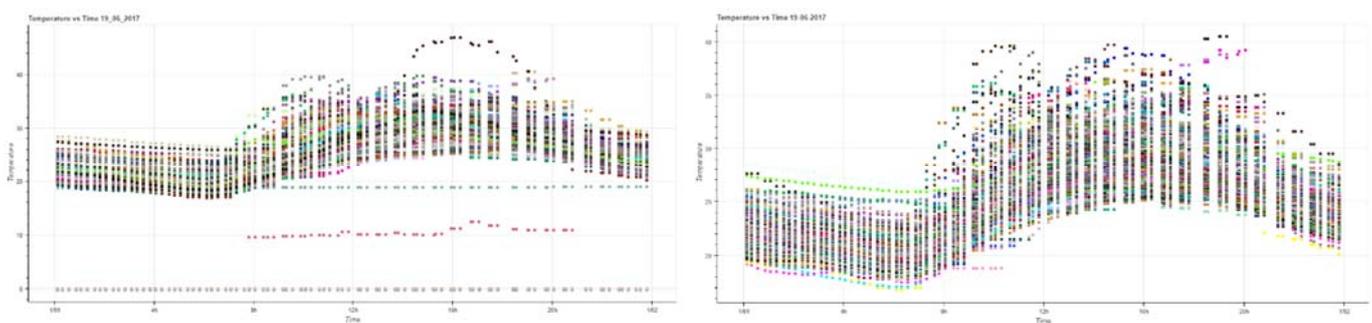
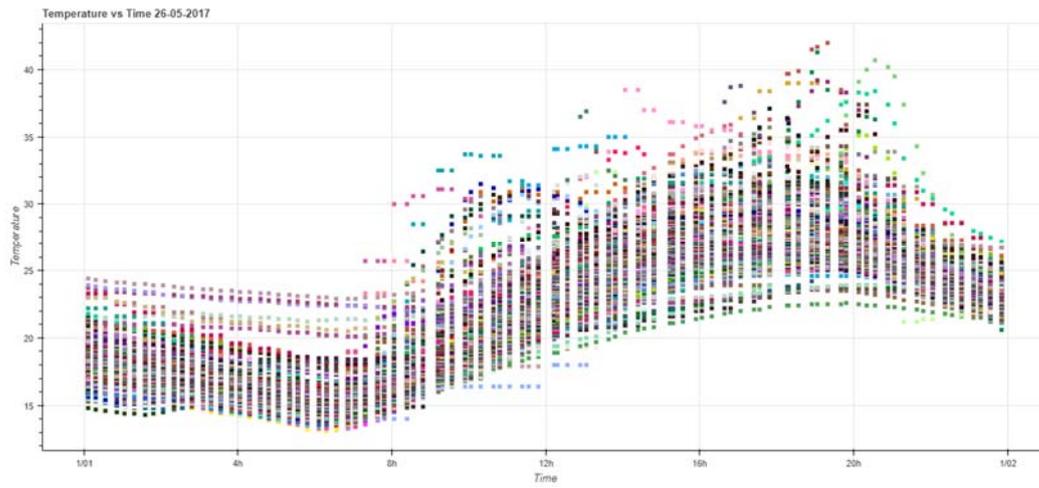


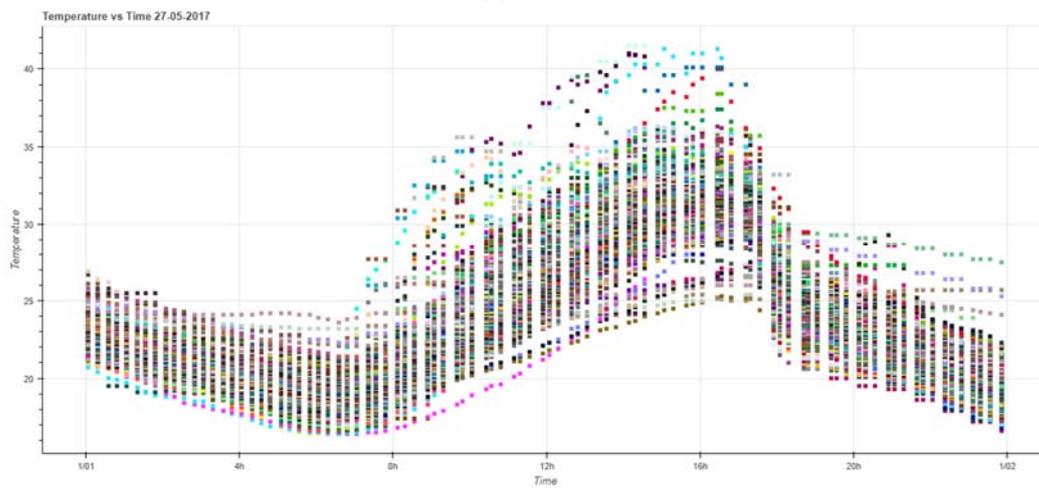
Figure 6: (a) Temperature vs. Time (19-06-2017); (b) refined data

Further, Figures 7 (a-g) provide an overview of the temporal patterns of the cooling and heating processes in the Hague for the hottest days of 2017. As it can be seen there is clearly expressed pattern of the heating process in the city between 10 am and 8 pm. In addition, the temperature range during the day is observably wider compared to the night-time measurements. This can be explained by the solar radiation during the day and the position of the sensors – some sensors had direct sunlight, which resulted in the measurement of extreme temperatures. However, the wider temperature range during the day can be attributed to the spatial differences between the areas, where the sensors are positioned as well. Therefore, the different spatial characteristics of the city influence the heating processes of the urban environment. These results give important insight into the temporal pattern of the development of the UHI effect.

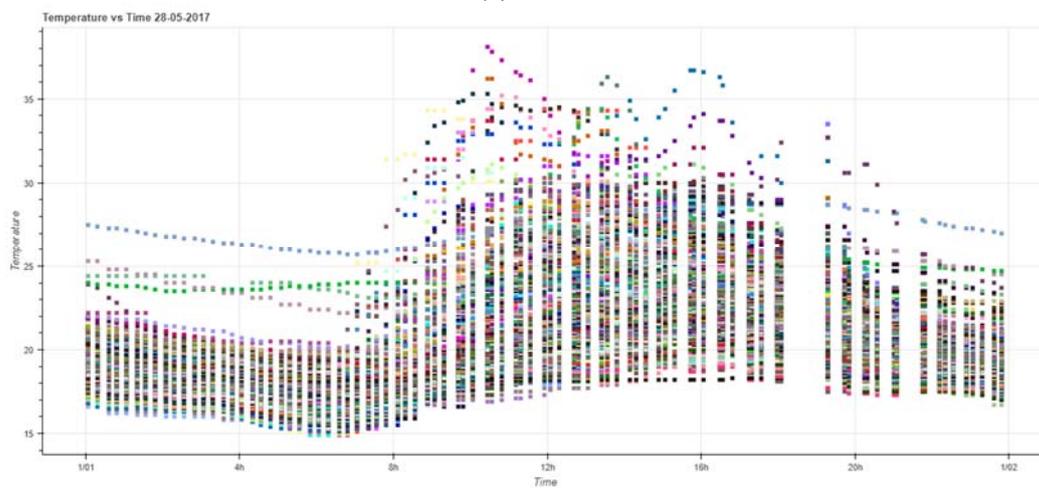
<sup>3</sup> <https://bokeh.pydata.org/en/latest/>



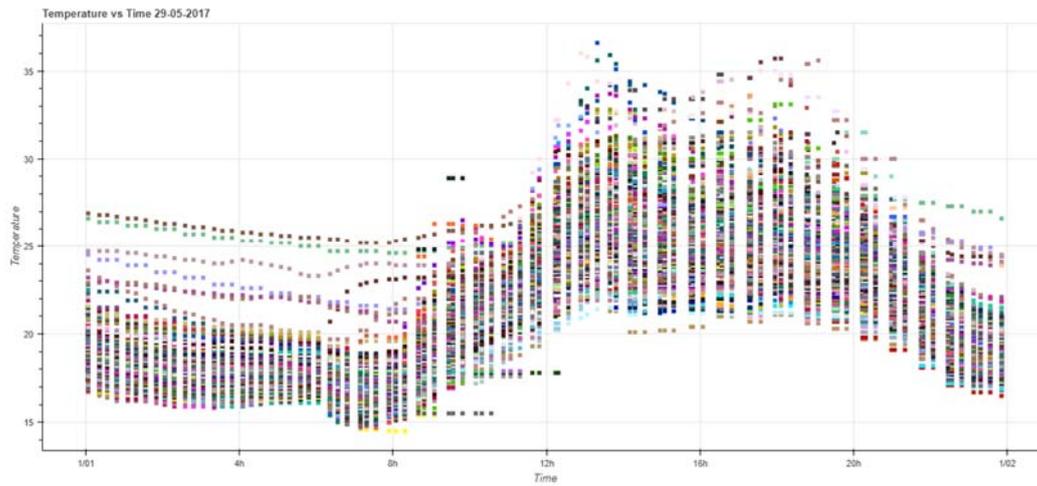
(a) 26-05-2017



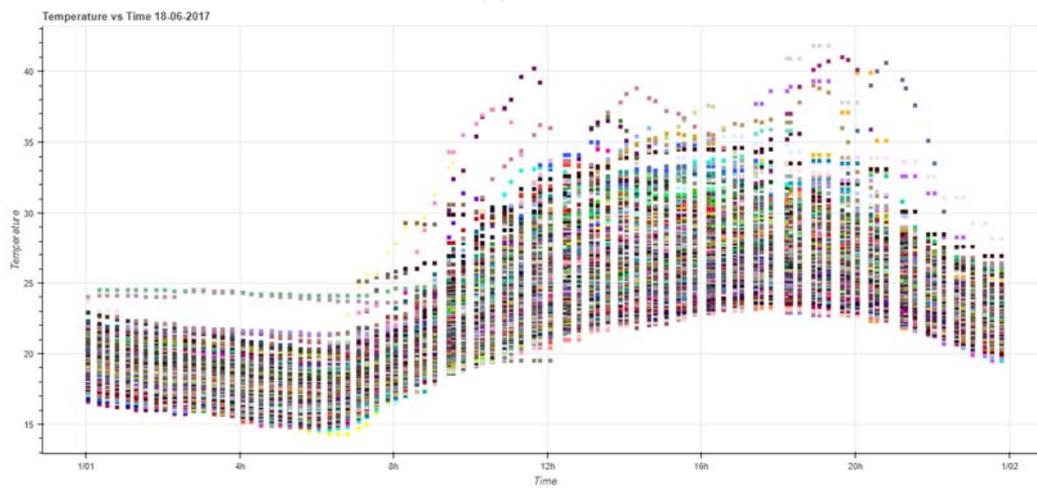
(b) 27-06-2017



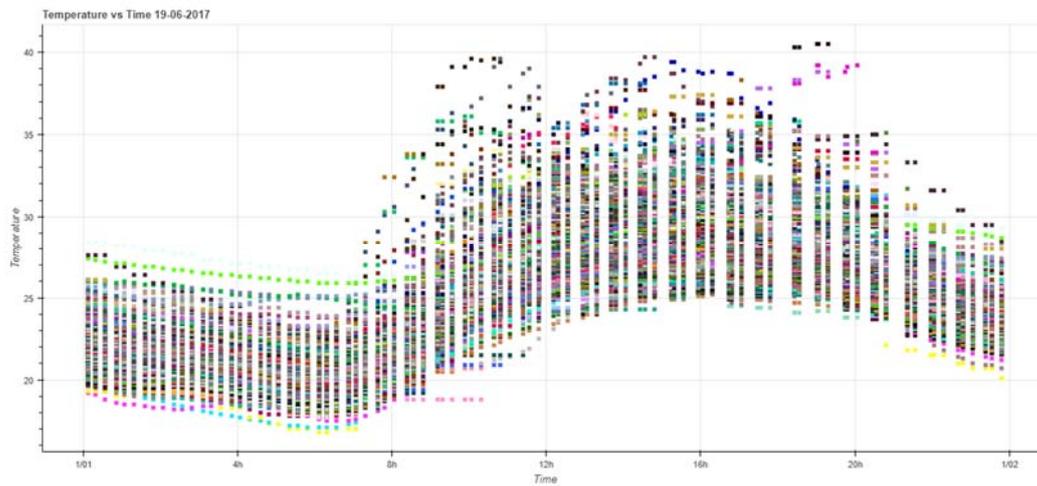
(c) 28-06-2017



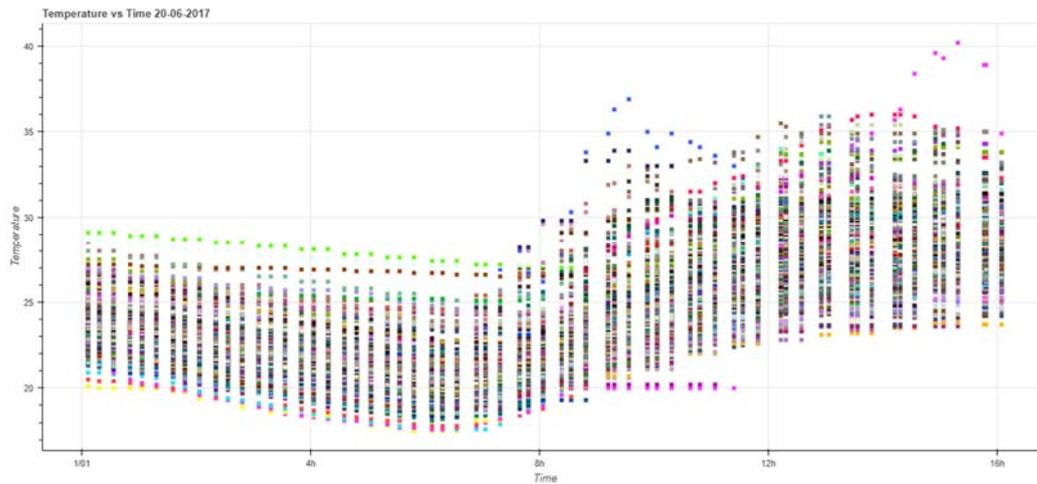
(d) 29-06-2017



(e) 18-06-2017



(f) 19-06-2017

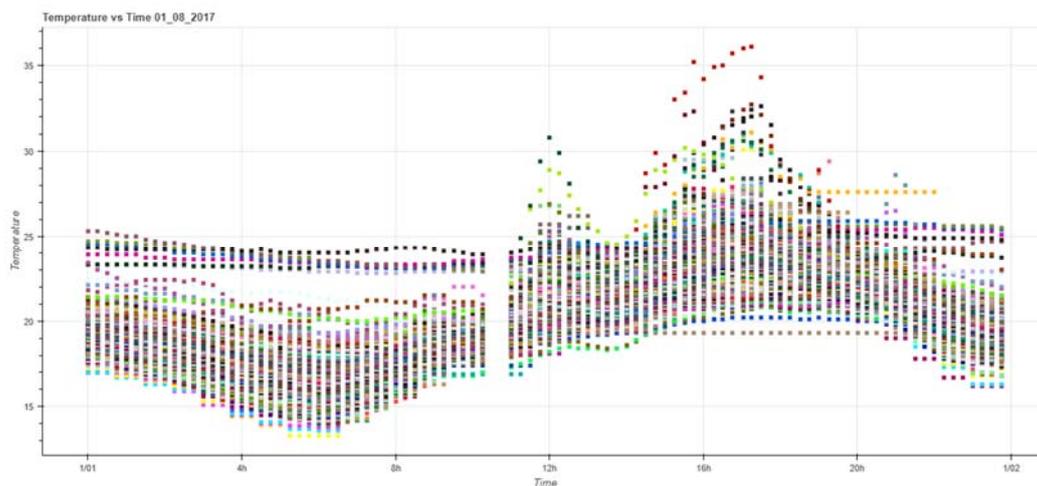


(g) 20-06-2017

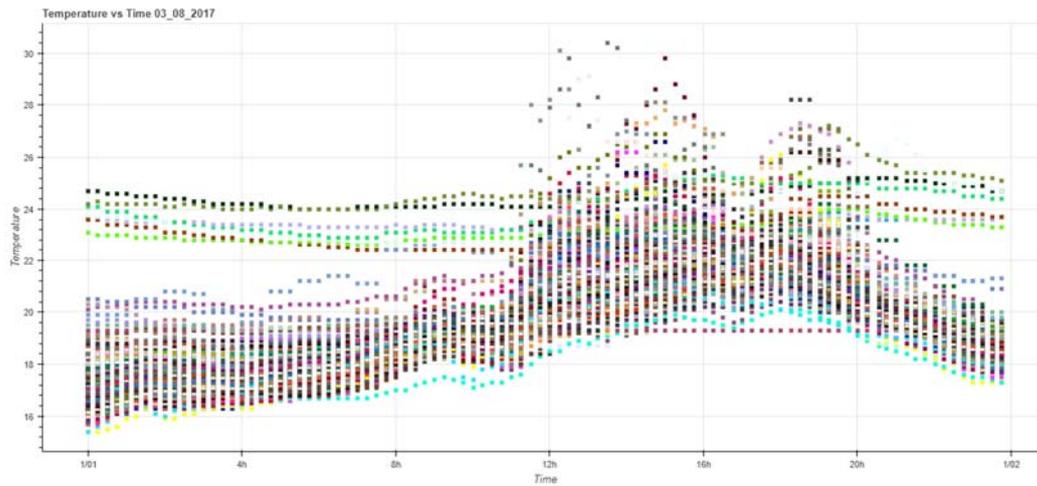
Figure 7: (a-g) Patterns of temperatures in the Hague for the hottest days (Temperatures vs. Time)

Moreover, the patterns of the temperatures of the hottest days of 2017 were compared to few regular summer days, where no heat wave was observed.

In Figures 8 a-b, two average summer days are presented for comparison purposes. Generally, the temperature ranges during these days differ noticeably from the days associated with the UHI effect in the city. The heating process of the urban environment follows less rapid development. In conclusion, the preceding observations provide evidence for the presence of the UHI effect in the city, the intraurban variability of the temperatures and the difference in the temperature patterns during extreme and normal summer days.



(a) 01-08-2017



(b) 03-08-2017

Figure 8: (a-b) Temperature patterns for 2 regular summer days (Temperature vs. time)

In addition, supplementary analysis of the spatial patterns of the temperatures within the city was developed using the QGIS software and the time manager plugin, which provides possibilities for animated visualization of spatial data. Thus, the change of the temperatures in the different locations was observed, where evidence for the intraurban variability of the Heat Island effect was noticed. Figures 9 (a-n) and Figure 10 (a-n) represent separate frames from the dynamic visualization of the change of the temperatures in the Hague for the 27<sup>th</sup> of May and the 3<sup>rd</sup> of August. It can be seen that there is a clear difference between the temperatures measured from the different sensors in the city.

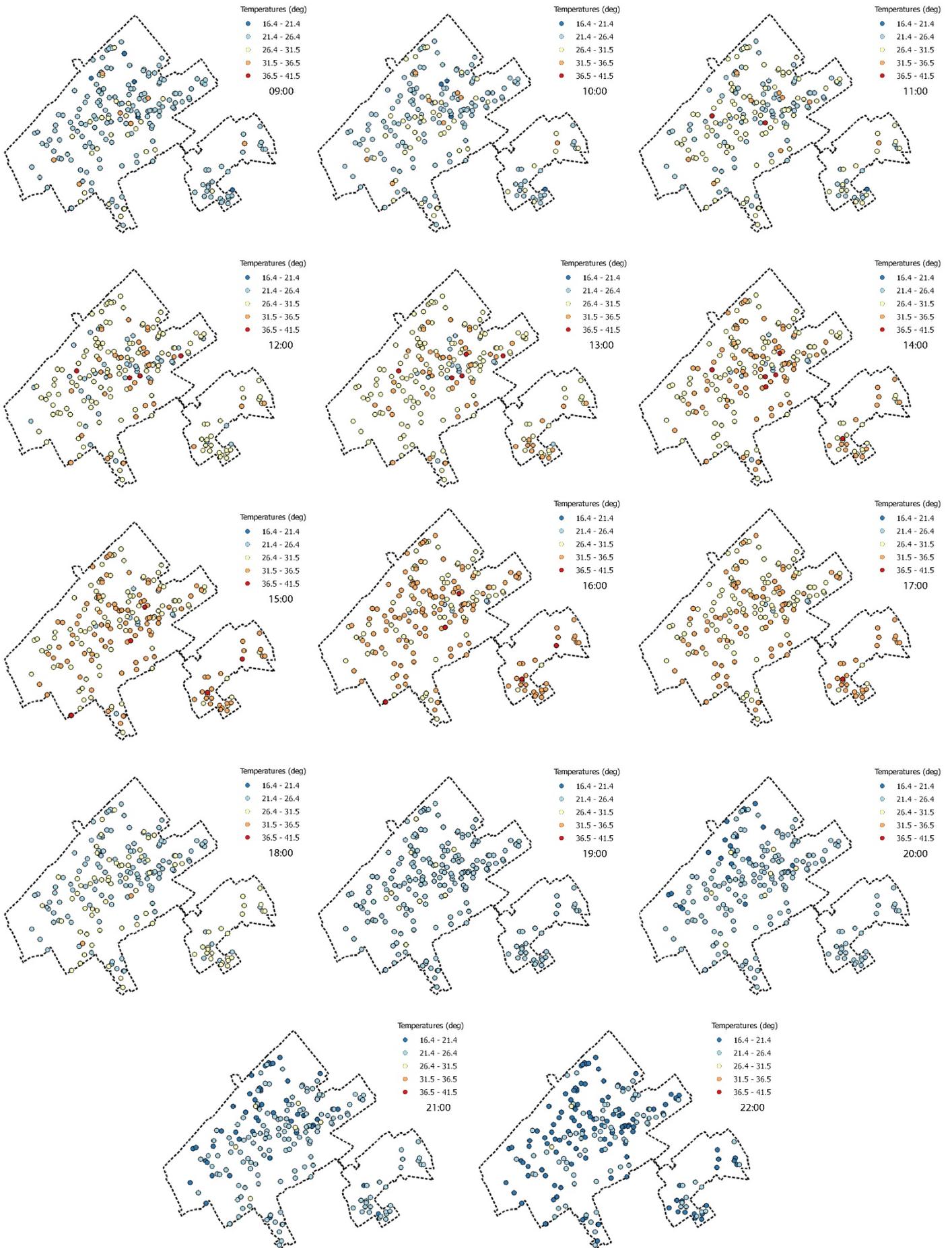
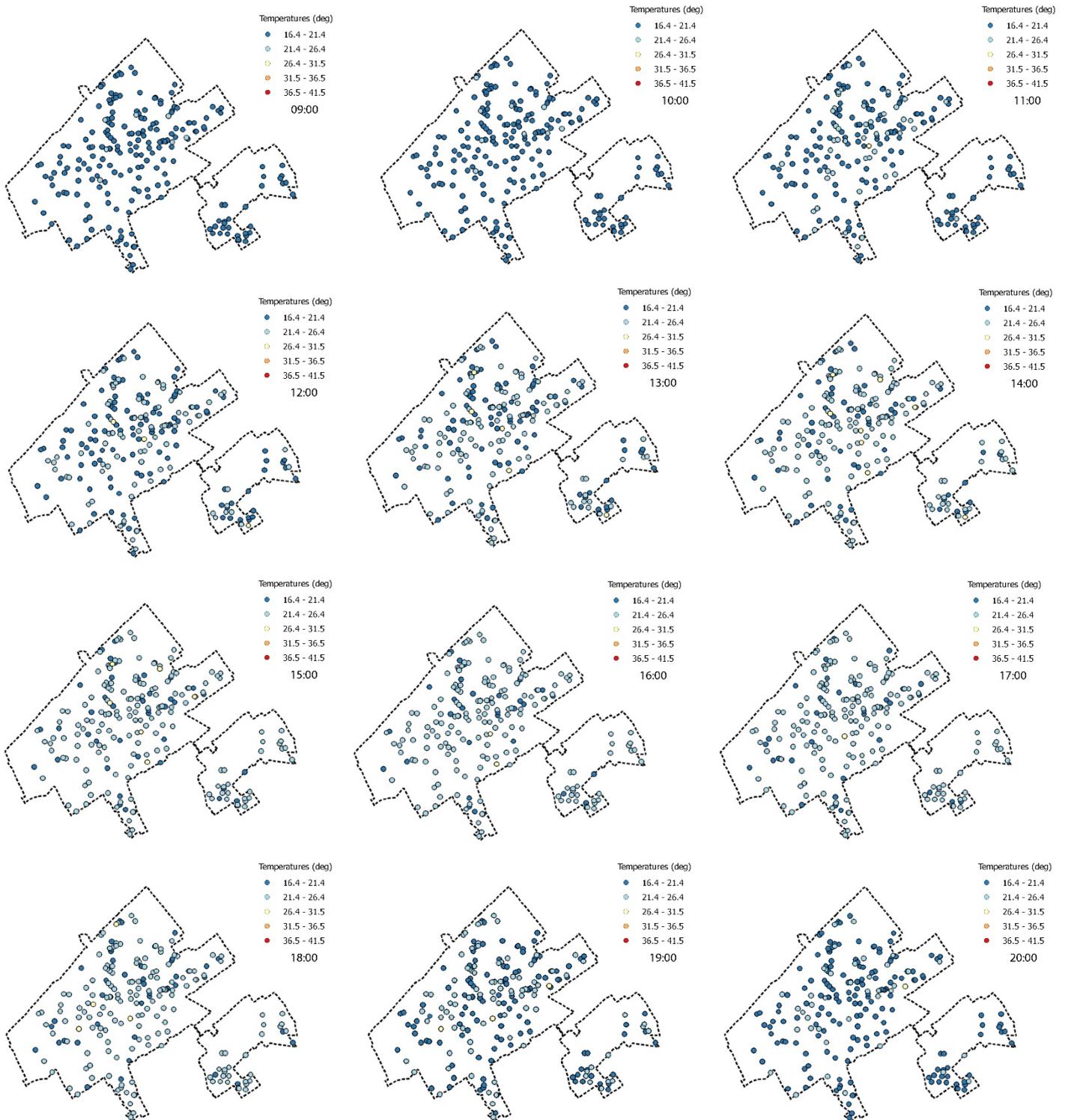


Figure 9: (a-n) Frames from the dynamic visualization of the spatial pattern of the temperature change (03-08-2017)



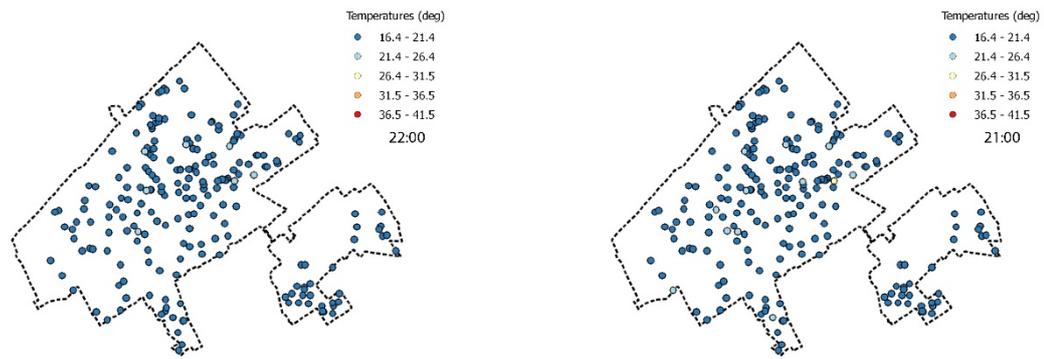


Figure 10: (a-n) Frames from the dynamic visualization of the spatial pattern of the temperature change (03-08-2017)

Also, it is worth mentioning that the development of statistically based methods for data cleaning was considered. For example, in many studies, the implementation of least squares fits for errors and noise detection is performed during the process of analysis of raw sensor data. However, since the focus of this research is targeted at understanding the temporal and spatial variability of the Urban Heat Island effect, the implementation of such methodologies can result in the loss of essential data. These methods would be suitable for models which are investigating averaged values or studies which are focused on similarities of certain phenomena rather than differences. In these cases, the extreme values and the noise are to be avoided for better performance of the models. On the contrary, the focus of the current work is on the extreme temperatures and their variability in the city. Therefore, the data cleaning process was limited to the removal of empty measurements and systematic errors. Thus, the variability of the data illustrating the complicated urban climate processes was preserved. However, the aggregation methods developed and explained in the following section are assisting in the removal of extreme outliers, which can affect the results of the modeling process.

### 3.4. Aggregation

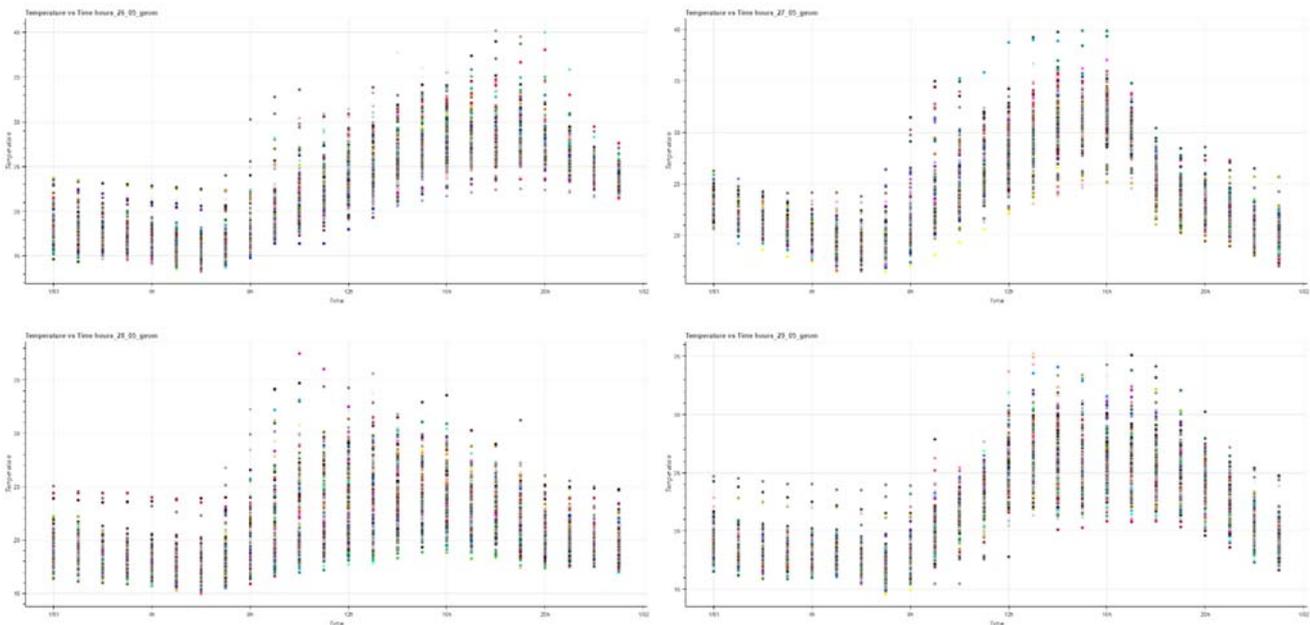
Following the analysis of the obtained data from the Netatmo weather network is the process of data aggregation. This step is required due to the great amount of data resulting from the high measuring frequency of the sensors (on average every 15 minutes). In addition, the temperatures represent the dependent variables of the statistical models and therefore they have to be reduced and processed in a way that will preserve the essential information, which the dataset provides. The whole process of temperature aggregation was based on trials of multiple possibilities and combinations, resulting in one final dataset, where one temperature value is used for every sensor.

Initially, the temperatures were averaged per hour for every sensor. The results are averaged temperatures for every hour of the 7 hottest days (Figures 11 a-g). In this way, the data was transformed into a better understandable form. In addition, this aggregation resulted in the finer

dataset, where part of the noise and errors were removed. Subsequently, based on the characteristics of those days, the hottest day of the 27<sup>th</sup> of May was chosen for the modeling phase.

Another approach of aggregating the values for the hottest period was considered but rejected due to the highest level of smoothing of the data signal. Such approach can result in undesired results and bad performance of the models. Also, the choice of aggregation method and final values directly depend on the purpose for which the statistical models are built. In the current case, the goal is to construct a model, which can predict the highest temperatures in the city during a heatwave and for that reason the hottest of the seven days was chosen.

Further, taking into account the definition of UHI effect and namely the deviation between the temperatures within the cities and their rural areas, additional analysis was performed. The difference between the temperatures, measured by every sensor in the Hague for the 24 hours of the 27<sup>th</sup> of May, and a rural location was calculated. The rural reference point was chosen from the Netatmo dataset based on a satellite image. The location was chosen to be as close to a rural area as possible. As a result, the highest disparity between the urban and rural temperatures was observed at 1 am with almost 3 degrees of difference. Therefore, the final dataset consists of the aggregated temperatures of the hour representing the highest difference between rural and urban.



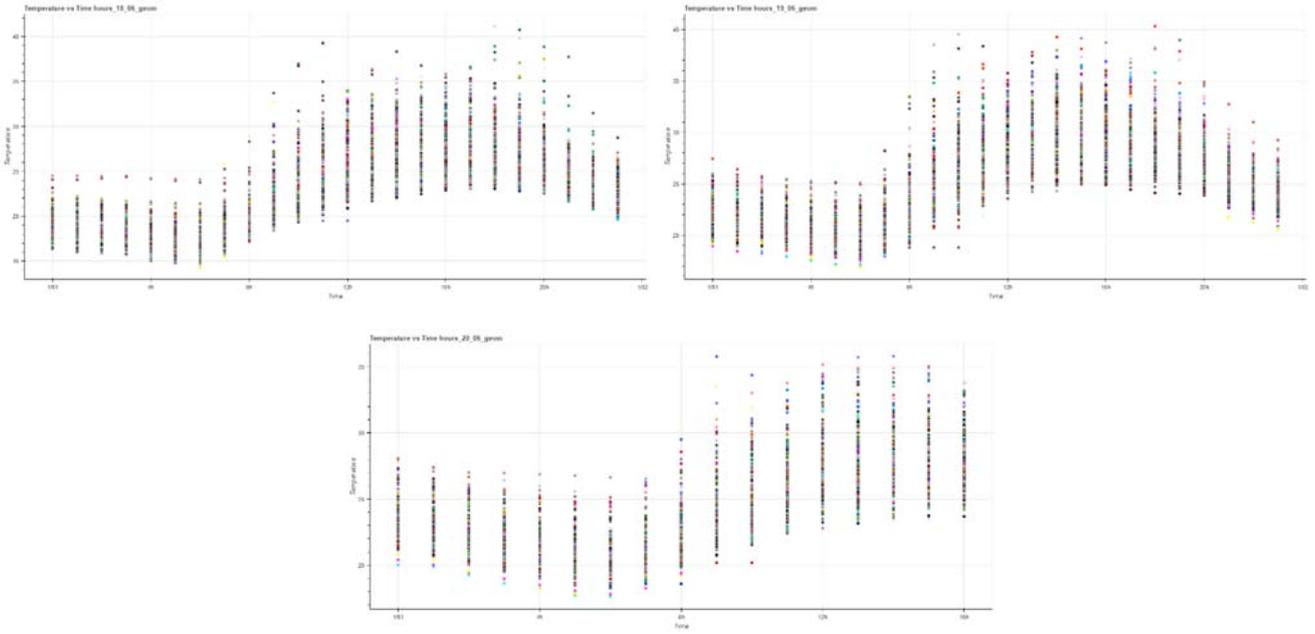
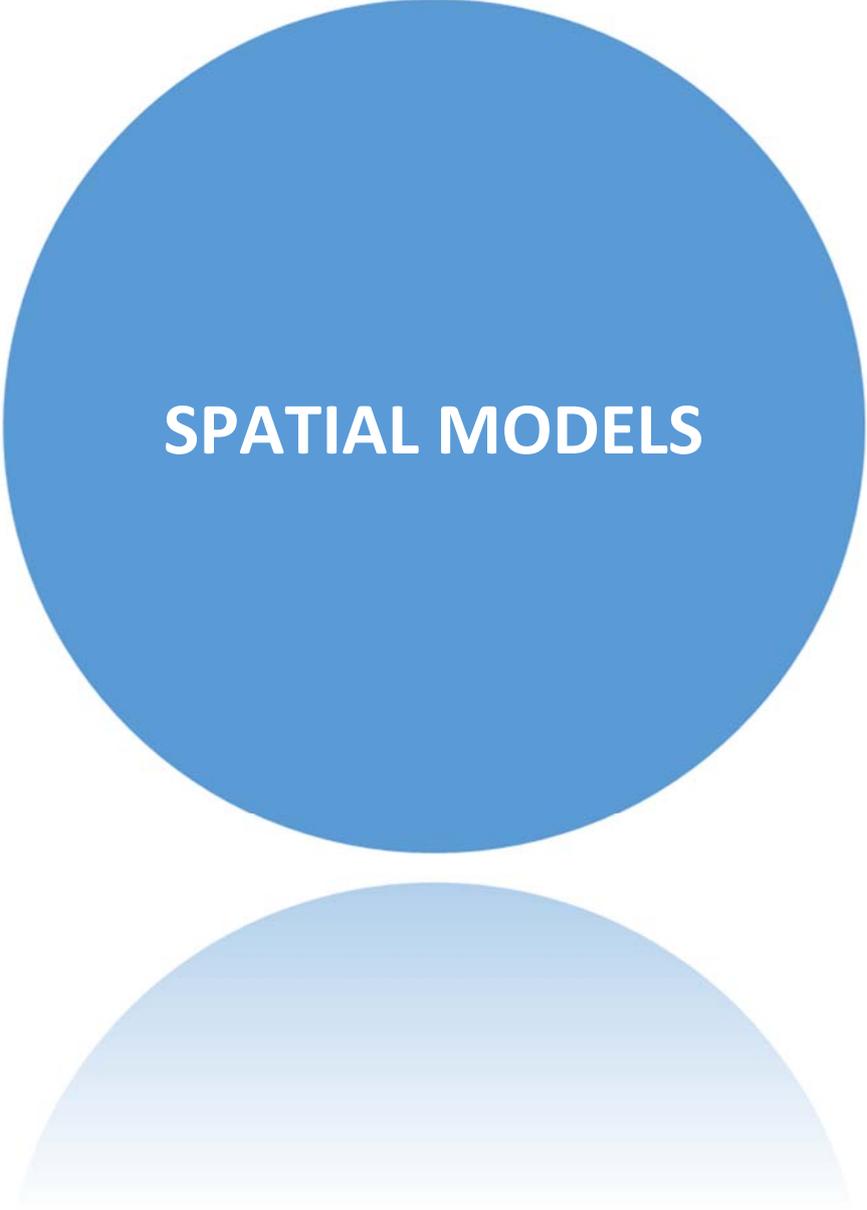


Figure 11: (a-g) Temperatures vs. time (aggregated data for the hottest 7 days)



# SPATIAL MODELS

## 4. Spatial modeling

The spatial modeling phase of the current research is one of the main aspects of the work, where the definitions of the space and the city's characteristics are created. The spatial parameters and their calculation methods represent important choices along the way of the creation of the statistical models.

The following section provides step by step overview of the process of city's structure parametrization. The chapter starts with the explanation of the design choices for the space tessellation. The different methods for grids creation are represented along with the analysis of their main characteristics regarding the spatial modeling performance. Further, the calculation of the spatial indicators takes place, where the GIS processing choices are explained and the obtained results are analyzed.

### 4.1. Space division

The main goal of this research is to model the UHI phenomenon in the city of The Hague. In order to achieve this, a substantial part of the work is to define the space and its characteristics. This abstraction will allow the creation of a mathematical representation of the relationships between the structure of the city and the observed temperatures. With the abstraction process, one aims at representing real-life features and their characteristics as measurable and quantifiable notions. The results from this process are the so-called "buckets", which store the processed data. There are two main types of space tessellations that one may consider – one that is primarily based on the features, which are going to be modeled (feature-primary tessellation) and one that decomposes space (space-primary tessellation)(Lee, Y., Li, Z., Li, 2000). Depending on the type of the tessellation process the results are either feature cells (FPT) or space cells, obtained from the space-primary tessellation (SPT).

Using either approach one obtains a representation of both the attributes and the space with the difference of where the focus is on. With the feature-primary tessellation (FTP) there is a direct relationship between the feature cell and the feature, in contrast to SPT where one cell can be associated with multiple features.

For the purposes of modeling the UHI phenomenon, the one to many relationship type has been chosen. In this way, the different characteristics of the city can be stored in the defined space cells. The space tessellation notion can be explored in a different manner, considering the expected results. In the current approach, two main types have been analyzed and namely the regular and unregular space divisions. For these purposes, three main grid types have been defined – equal sized rectangular

grid, equal sized hexagonal grid and a Voronoi tessellation. All tessellations have been produced as polygonal features with the means of GIS software and can be seen in Figures 12 a-c.

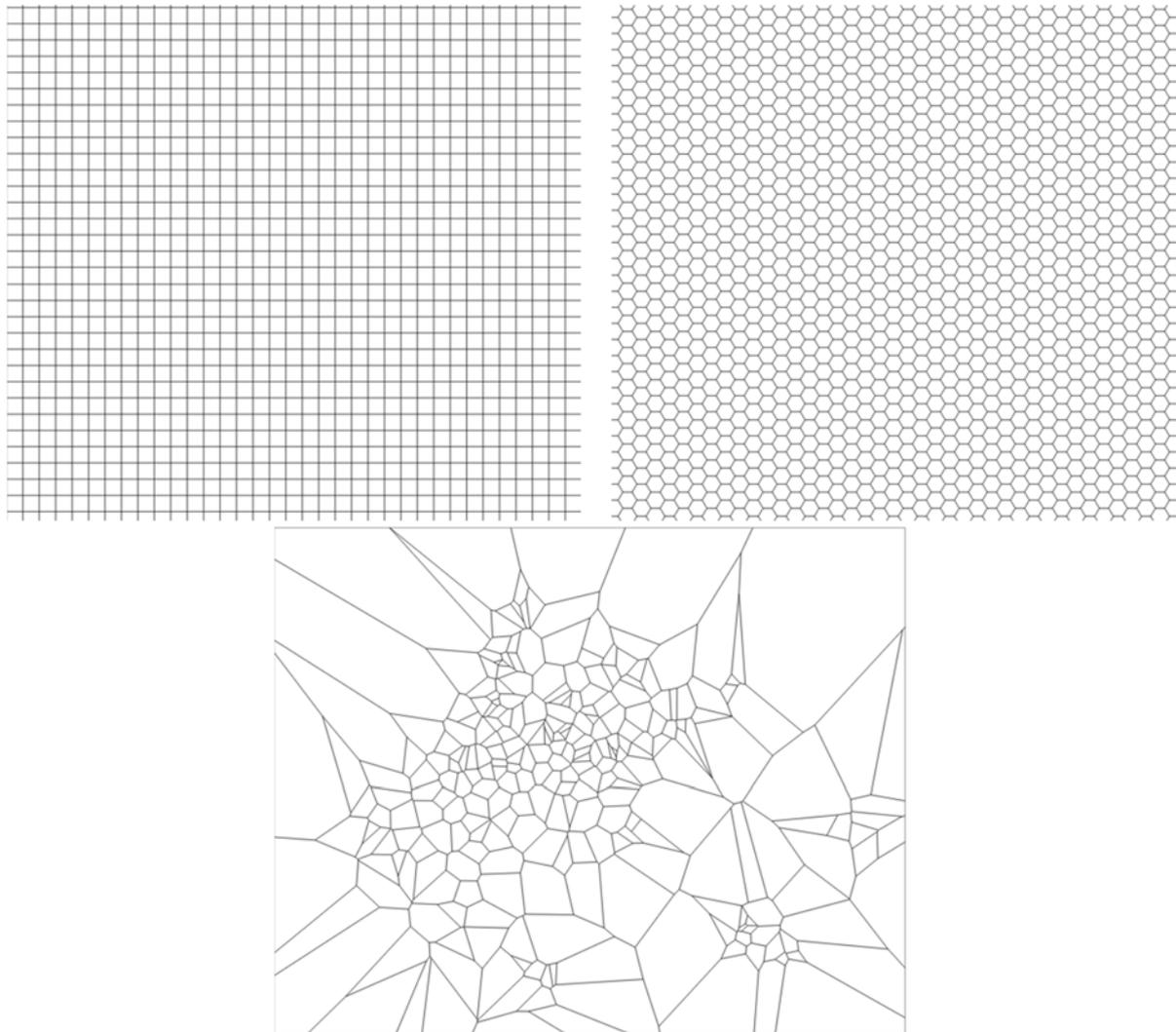


Figure 12: (a) Rectangular tessellation; (b) Hexagonal grid; (c) Voronoi diagram

The created rectangular grid covers the area of The Hague and it is based on the Dutch CBS (Centraal Bureau voor de Statistiek) grid, which is widely used for statistical purposes. In order to maintain the compatibility of the different tessellations, the origin, the cell size and the extent of the CBS grid have been used for the creation of the rectangular tessellation. The purpose of this space abstraction is to capture the detailed characteristics of the city and therefore a 100m cell size has been chosen as an appropriate detailed representation. The grid contains 9062 cells and represents a full space partitioning of The Hague area, where the comparability of the different units is secured.

The second chosen space partitioning method is based on the hexagonal shape. For comparison purposes, the distance between the two sides of the hexagons is 100m and the resulting grid contains 10 359 cells.

These two types of tessellations have different characteristics, which affect the way the spatial data is associated with the cells. The hexagons represent more compact shapes and one of their main advantages is that the edge effects of the data patterns are reduced. Also, the hexagons are preferable for visualization purposes due to their ability to map better the curvature patterns of the data. On the other hand, the rectangles are simpler shapes (easier to define and store) and are scalable, which means that the chosen approach can be easily transferred to different areas and cities. Furthermore, from a processing point of view rectangles can be more easily used in different raster operations.

Finally, a third space representation approach based on the Voronoi tessellation has been incorporated in this study. The Voronoi tessellation has been substantially used in the field of computer science and can be roughly described as a space tessellation based on a discrete number of points, where each side of the polygons is closer to the point that lies within it, rather than to any other (Aurenhammer, n.d.). For the creation of these polygons, the locations of the unique sensors in these areas have been used. First, the working sensors for the period of interest of this study have been retrieved from the spatial database. Then the unique sensors were filtered from the data files and their point geometry was used as seed points for the Voronoi algorithm. The generation of the vector data set is performed using GIS software and its embedded Voronoi algorithms. The resulting grid contains 273 cells with varying areas.

The three resulting types of data “buckets” are used in the following step of the calculation of the UHI contributing spatial indicators. This means that the grid cells are the aggregation unit for every indicator and essentially store all parametrized values.

## 4.2. Calculation of the spatial indicators

The essential work of the spatial modeling phase is the representation of the morphological characteristics of the city as measurable indicators, which will explain the difference in the temperatures within The Hague. These characteristics are defined as the main UHI contributing indicators. The methodology of deriving these indicators from raw spatial data, the processing steps and final calculation procedure are presented in the following section.

### 4.2.1. Buildings density

*Table 2: Buildings density indicator information*

Indicator	Values	Datasets
Buildings density	0 (no buildings) – max volume	AHN3, TOP10NL

The urban heat island phenomenon is characterized by the difference between the temperatures within the city and its surrounding areas (Stanganelli, Marialuce, and Soravia, 2012). The reason for

this difference is the subject of interest of this study and namely how to measure and quantify these factors. One of the most important characteristics is the city's structure and more precisely the built environment. The buildings within the city play an important role in the definition of the urban microclimate. The geometry of the built-up areas can be easily measured and it is crucial for the thermal energy exchange in the city. The heat storage and exchange properties of the different materials have been thoroughly studied and their contribution to the higher temperatures within the city has been established (Santamouris et al., 2011). Therefore, it is important to include the structure of the buildings in the development of the UHI studies.

In the current study, the geometry of the city is defined and measured as the Buildings density index. This index is expressed by the ratio between the sum of all buildings volumes falling within a grid cell to its area (Stanganelli, Marialuce, and Soravia, 2012). The datasets that have been incorporated in the calculation of the density of the buildings are the Dutch digital surface model (AHN3) and the topographic dataset containing the buildings footprints – the TOP10NL.

On Figures 13a and 13b, the spatial extent and the resolution of the AHN3 raster are represented. The surface elevation raster is derived from a lidar point cloud and has a resolution of 0.5 x 0.5m. Important to mention is that the values are stored as absolute numbers (not relative to the sea level), which essentially affects the calculation method of the heights of the buildings. First, the spatial extent of the dataset has been refined to the area of interested by combining all relevant tiles and clipping them to the area of The Hague.

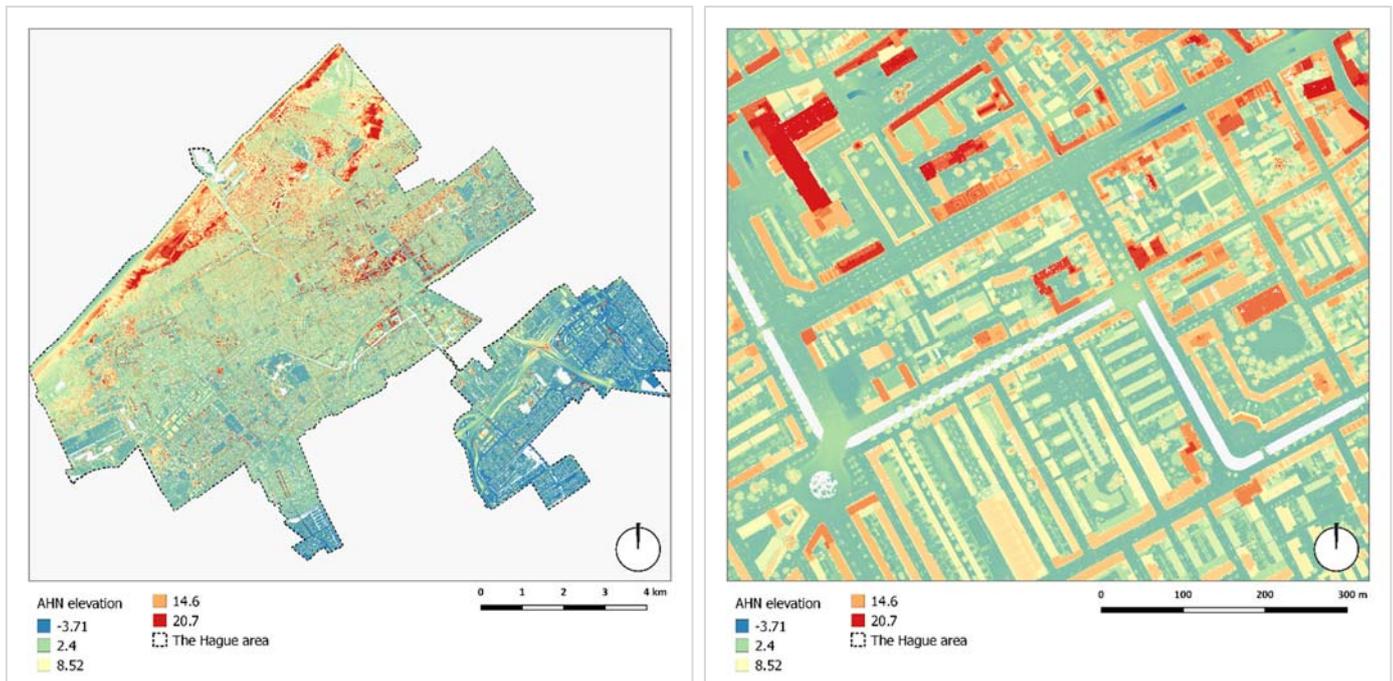


Figure 13: (a) AHN3 spatial extent; (b) AHN3 resolution

The calculation of the buildings density index is performed in multiple stages and is represented in Figure 14. The elevation raster has been used in the initial steps of the method where the volumes of the buildings are calculated. This has been done by multiplying the area of the buildings footprints by the extracted average height. The first important decision to be made at this stage is how to derive the heights of the buildings from the elevation raster accurately. The followed approach is to overlay the buildings footprints and the AHN3 datasets spatially. Further, the height values are derived by extracting summary statistics using the embedded in QGIS zonal statistics plugin. Essentially, this algorithm performs spatial overlay between the vector and raster datasets, where a statistic per every polygon, based on the underlying raster values is calculated. Important here is the choice of aggregation value from the different options – minimum, maximum, mean, median, etc. In the current method, the mean has been defined as the most appropriate value based on literature references and comparison between the different options. After performing this operation, a substantial amount of the resulted heights values was negative. This indicated that the elevation values stored in the raster file were absolute numbers. This problem has been overcome by creating a buffer around the buildings, which is then clipped with the buildings footprints in order to retrieve the surrounding areas only. Further, the mean heights of these areas are calculated from the AHN elevation raster. Therefore, the final heights of the buildings are defined as the difference between the average height of the building plot and the surrounding area. The height attribute is then joined to the buildings layer and used for further analysis.

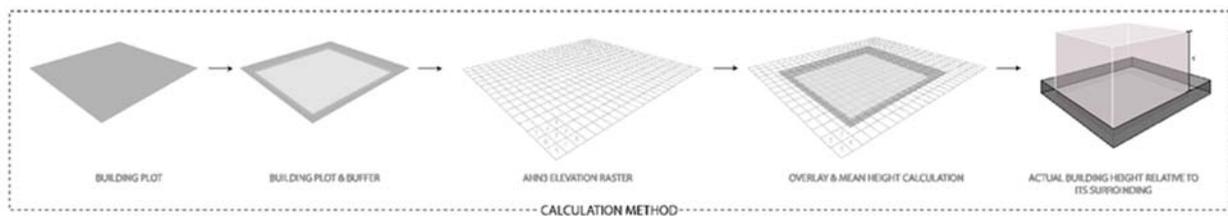


Figure 14: Calculation method of the Buildings density

Further, the buildings plots have been intersected with the three different grid types, resulting in partitioning of the areas of the buildings according to the cell that they belong to. In the next steps, the areas of the parts of the buildings, falling within every cell have been calculated and multiplied by their height, resulting in the calculation of the volumes of the buildings in every cell. Next, the unique identification numbers of the cells of the grids are joined to the buildings layer. An important step before the final calculation of the Buildings density index is the removal of the small pieces and silver polygons resulted from the intersection operation. Thus, all buildings pieces with areas smaller than 1 m<sup>2</sup> and all buildings heights smaller than 0.1 m are removed from the dataset and considered as irrelevant for the calculation of the current indicator. Finally, the Buildings density indicator is

calculated by summing all volumes of the buildings per cell id and dividing them by the area of the cell that they belong to.

4.2.2. Land cover ratio

Table 3: Land cover index information

Indicator	Values	Datasets
Land cover index	0 (no buildings) – 1 (entirely covered cell)	TOP10NL

The second indicator describing the geometry of the city which was included in this research is the Land cover ratio. This index is expressed as the ratio between the sum of all buildings plots and the areas of the cells. As the previous indicator, the relevance of the Land cover ratio is related to the heat storing properties of the materials and the decreased amount of vegetation within the city. It has been observed that vegetation has a positive influence on the urban microclimate and assists in the decreasing of the temperatures within the city (Akbari et al., n.d.). Thus, measuring the density of the built-up areas is highly relevant to the UHI studies. The difference between the Buildings density and the Land cover indices is that the first one represents the volumes of the building. The 3-dimensional indicator incorporates another important phenomenon for the heat exchange in the city and namely the so-called canyon effect, where temperatures are higher due to the limited long-wave radiation.

The calculation method of the Land cover index is very similar to the one followed in the previous section and is illustrated in Figure 15. The TOP10NL vector data set has been used again, and more precisely the buildings footprints, which are extracted from the original GML files. The spatial extent of the data is refined to the area of The Hague. Further, the resulting layer has been intersected with the three different grids.

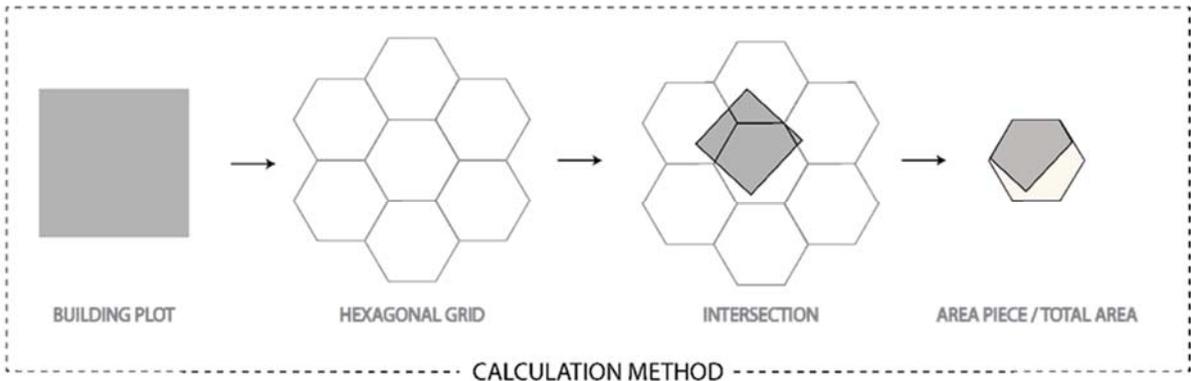


Figure 15: calculation method of the Land cover index

Thus, the buildings are partitioned according to the geometry of the different spatial tessellations and the areas of the buildings falling within the different cells can be obtained separately. Further, the values are summed per grid cell and divided by its area. The result is a normalized spatial indicator, varying from 0 (no buildings in the cell) to 1 (entirely covered by buildings cell). Finally, all small pieces

and sliver polygons, which are considered irrelevant for the research have been removed from the dataset.

Depending the tessellation type, it can be observed that the regular grids have a higher probability of being entirely covered by buildings plots due to their smaller cell sizes. This means that overall the regular grids yield more extreme values, while the Voronoi tessellation gives a smoother representation of the index.

#### 4.2.3. Non-permeable surfaces index

*Table 4: Non-permeable surfaces index information*

<b>Indicator</b>	<b>Values</b>	<b>Datasets</b>
Non-permeable surfaces	0 (no buildings) – 1 (entirely covered cell)	TOP10NL

The impermeable surfaces in the cities are identified as one of the major contributing factors to the higher ambient temperatures in the urban areas (Asaeda & Ca, 2000). The heating and cooling processes in the cities are highly related to the balance of the surface energy fluxes, where the transportation and transformation of water play a crucial role. The impermeable surfaces have high heat absorbing and storage capacities, which affect the temperatures in the cities directly. Due to the limited evaporation processes, the heat stored during the day cannot be released into the atmosphere or in the ground, which causes the perceptibly higher air temperatures. In contrast, natural surfaces covered by soil or vegetation are characterized by higher ability to cool down during the night. This means that ambient temperatures in natural areas are high during the day, but after sunset, a rapid decline can be observed (Asaeda & Ca, 2000). Therefore, the percentage of sealed surfaces within a certain area provides important information for the analysis of the higher temperatures within the city.

In this study, the Non-permeable surfaces indicator is expressed as the ratio between the paved open and the total open areas per grid cell. For the spatial processing, land use classification data provided by the municipality of The Hague has been used. The dataset contains a highly detailed classification of the surfaces within the city, divided by different categories such as types, used materials, permeability and others. For the calculation of the current indicator, the “type” attribute has been chosen. It includes categories like roads, roofs, gardens, buildings, parking lots, bike paths, etc. Based on this classification the dataset has been divided into two different layers – one representing the total amount of open areas in the city (the urban area with excluded buildings) and the second one describing the impermeable surfaces in the city (roads, parkings, bike lanes, construction sites, etc.). In Figure 16a and Figure 16b, the spatial extent and more detailed overview of the two layers are provided.



Figure 16: (a) Paved open and total open areas; (b) Paved open and total open areas (detail)

Both layers are intersected with the three different grids and dissolved by the cell id that they belong to. Further the areas of the resulted polygons are calculated and finally, the two tables of the paved open areas and total open areas have been joined for the calculation of the ratio between paved and open areas per grid cell (Figure 17).

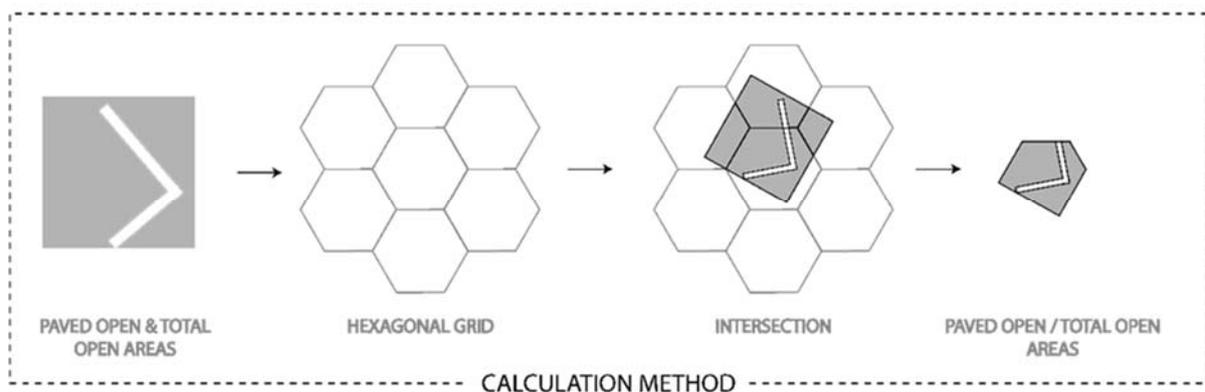


Figure 17: Calculation method of the Non-permeable surfaces index

The result is an indicator with values varying between 0 (cell covered with open, permeable surfaces) and 1 (entirely impermeable area). Important to be mentioned here is that the exclusion of the areas of the buildings from this indicator is beneficial for the developed statistical models. The multicollinearity between different independent variables introduces bias in the statistical models and it has to be avoided while creating the explanatory variables. Therefore, the surfaces of the buildings are purposefully excluded from the calculation of the impervious surfaces in the city, although essentially they belong to this category.

#### 4.2.4. Vegetation coverage

Table 5: Vegetation index information

Indicator	Values	Datasets
Vegetation index	0 (no vegetation) – 1 (entirely covered cell)	Sentinel-2

The influence of the vegetation on the temperatures within the cities have been studied extensively (Luan et al., 2014). Overall the positive effect of the green areas on the urban microclimate has been recognized and vegetation is described as one of the main UHI mitigation strategies (Gago, Roldan, Pacheco-Torres, & Ordonez, 2013; Memon et al., 2008; O'Malley et al., 2014). The different characteristics of the urban green areas are also important to consider, such as area, vegetation type, density, etc. Greenery can positively affect the Urban Heat Islands not only in the form of big park areas but as green roofs and facades as well (Wijerathne & Halwatura, n.d.). Thus, the inclusion of indices related to the vegetation coverage within the cities is crucial for the development of the UHI models.

The calculation of the number of green areas can be performed in multiple ways using diverse datasets. In the current work, the Sentinel-2 satellite collection has been utilized. Satellite images provide detailed information with high temporal frequency, which allows a detailed analysis to be performed on the urban greenery. The Sentinel-2 is an Earth observation mission, developed by the European Space Agency with the mission of providing data for environmental, disaster management, change detection purposes and others ("Sentinel-2: The operational Copernicus optical high resolution land mission," n.d.). The Sentinel-2 satellites provide multispectral data with 13 bands, covering the visible, near-infrared and short-wave radiation. The revisiting time is approximately every 5 days and the data have a spatial resolution of 10 m.

The processing of the satellite images has been performed using the Google Earth Engine platform (Gorelick et al., 2017). Using the web-based API one can retrieve and analyze satellite images entirely within the web browser environment.

In order to define the green areas from the Satellite images, the Normalized Difference Vegetation Index (NDVI) has been calculated. This index can determine the amount, the density and the condition of the vegetation, based on the analysis of the reflected visible and near-infrared light from the plants (Herring, 2000). It has been observed that healthy vegetation absorbs most of the visible light and reflects bigger part of the near-infrared wavelengths, while in sparse or unhealthy plants the opposite relation is detected. Thus, the NDVI index values vary between -1 and 1, where values close to 1 indicate high-density green areas. The index has been computed according to the following formula:  $RED - NIR / RED + NIR$  and the resulting NDVI map is illustrated in Figure 18.

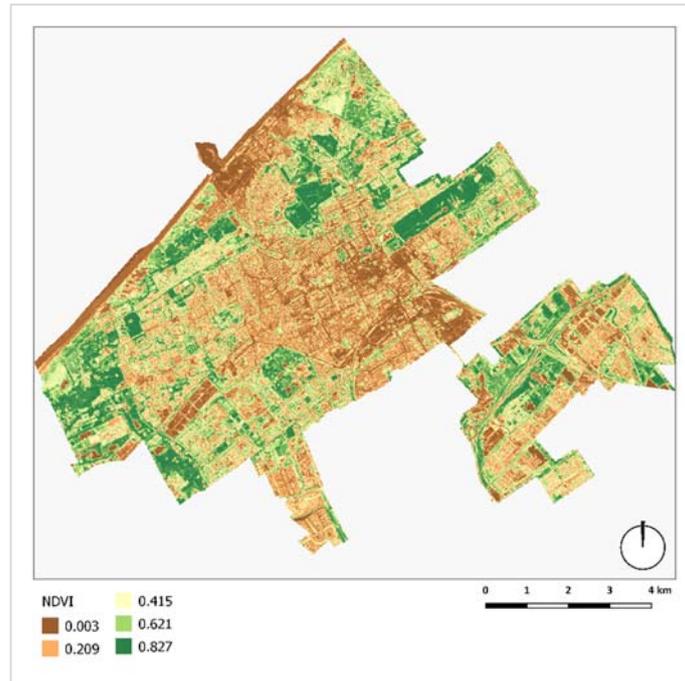


Figure 18: NDVI map

The obtained NDVI index is in raster format and has been further processed in QGIS for the definition of the ratio between the green areas and the area of the grid cell.

Further, the calculation method is illustrated in Figure 20 and it starts with the extraction of these values from the NDVI raster, which represent the densely vegetated areas. This is done by defining a vegetation class with values varying from 0.6 to 1. These cells are then extracted from the raster using a raster mask. Thus, the green areas within the city have a value of 1 in the reclassified raster and everything else is 0 (Figure 19).

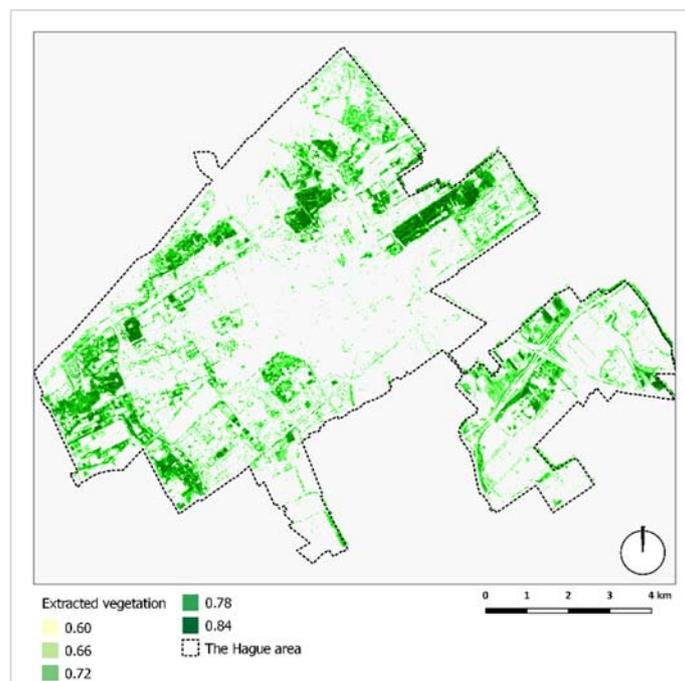


Figure 19: Reclassified vegetation raster

In order to calculate the vegetation areas, the raster file was vectorized. The resulting vector layer was intersected with the three grids, and thus the vegetated areas in every grid cell were obtained. Finally, the areas of the resulting polygons were calculated and divided by the cell areas, resulting in the ratio between the vegetation coverage and the area of the cell.

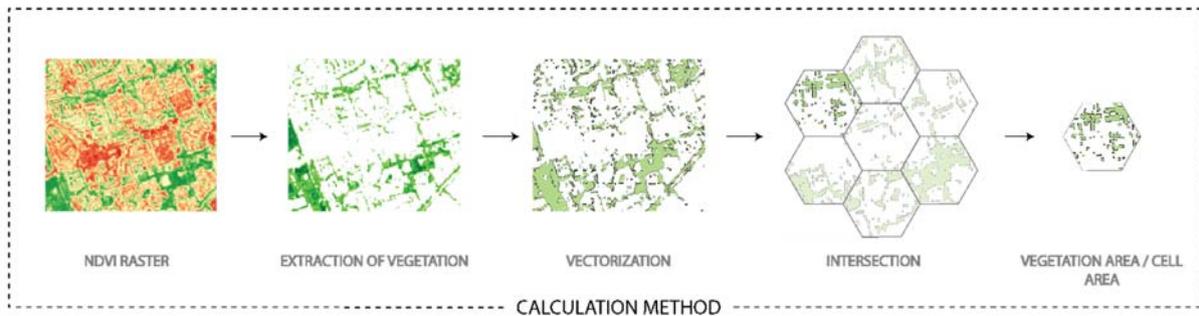


Figure 20: Vegetation index calculation method

#### 4.2.5. Sky View Factor

Table 6: Sky View Factor information

Indicator	Values	Datasets
Sky View Factor	0 (obstructed areas) – 1 (open areas)	AHN3

The Urban Heat Island effect illustrates the processes of higher temperatures in the cities due to the diminished cooling properties of the urban environment. During the day all man-made surfaces within the city accumulate heat, which is slowly released after sunset and during the night. Thus, the cooling processes in the city are very important and strongly related to the UHI effect. The Sky View Factor is an important indicator, which has been investigated excessively in the context of the UHI studies (Hämmerle, Gál, Unger, & Matzarakis, 2014; Svensson, 2004; Unger, 2009). The Sky View Factor expresses the ratio between the received radiation from the surface and the entire hemispheric radiation. In other words, the SVF denotes the fraction of visible sky from a certain point on the ground. It varies between 0 and 1, where 0 is a full obstruction in contrast to 1, which denotes open areas. The proportion of open sky is extremely important for the release of the heat and the cooling processes in the cities. The urban canyons are characterized by very high buildings, which essentially have higher heat storage properties due to the obstructed long-wave radiation. Thus, the low Sky View Factor values within a certain area can be connected to the presence of the UHI effect.

The calculation of the SVF can be performed in two ways – within QGIS using specifically developed algorithm by SAGA or using the Relief Visualization Toolbox (RVT) – a tool developed by the Institute of Anthropological and Spatial Studies in Slovenia (Kokalij, Zaksek, & Ostir, 2011). The computation of the SVF is based on an elevation model, which was derived from the PDOK portal. The AHN3 raster

files of the area of interest were derived and resampled to 1 m resolution due to the heavy load of the SVF computation. The algorithm implemented in the RVT computes the vertical elevation angle of the horizon from a certain point in a number of directions to the radius, specified by the user. The parameters that have been used are search radius of 10 pixels in 8 directions and the results can be seen in Figures 21 a-b.

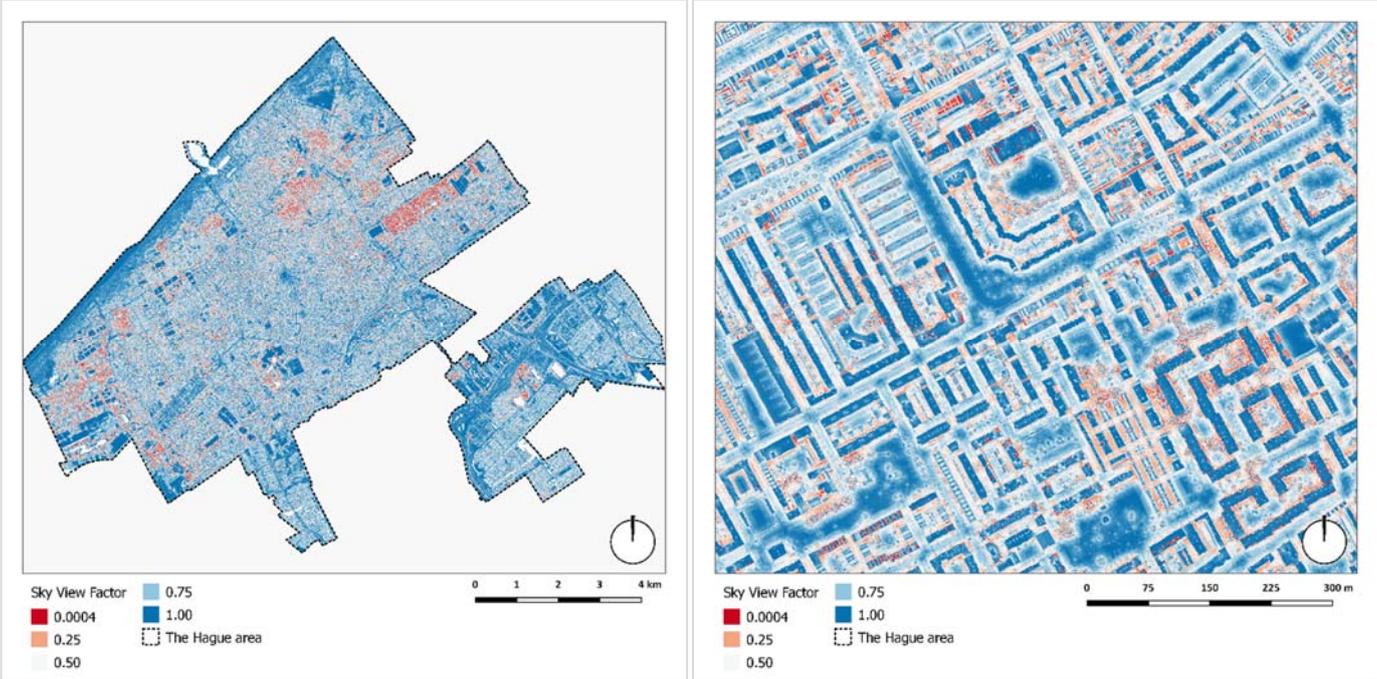


Figure 21: (a) Sky View Factor;(b) Sky View Factor (detail)

It should be mention that the AHN 3 digital surface model contains trees, which are included in the calculation of the Sky View Factor. Therefore, areas with high trees have low SVF values and are considered as obstructed.

Further, the obtained SVF raster file has a spatial resolution of 1 m, meaning that the obtained values have to be aggregated per every grid cell. For this purpose, the Zonal statistics tool has been used, where the average SVF value per cell was calculated and used as the final value for the current indicator (Figure 22).

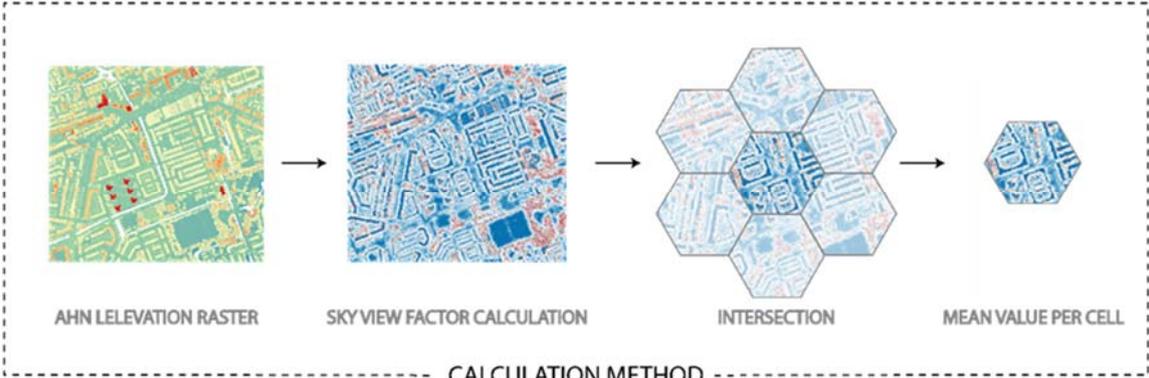


Figure 22: SVF calculation method

#### 4.2.6. Vehicle traffic density

Table 7: Information about the Vehicle traffic density index

Indicator	Values	Datasets
Vehicle traffic density	0 (no roads) – 1 (cell entirely covered by roads)	TOP10NL

The heat generated from the different anthropogenic activities is largely contributing to the difference in the temperatures between the city and its surrounding areas. Therefore, the inclusion of the indicators for the anthropogenic activities in the city is crucial for the reliability of the UHI studies. In the current work, the heat generated from the vehicles has been considered in the analysis of the UHI contributing factors. However, the calculation of the vehicle traffic density and the heat generated from the cars is very complicated and includes multiple factors, which are difficult to estimate and require empirical data. Therefore, this index has been simplified to the computation of the density of the road network within a grid cell. In this way the Vehicle Traffic Density indicator can be included in the study, considering the current availability of data.

Another approach would be the inclusion of data, representing the number of vehicles passing through certain roads. Based on this numbers an estimation of the generated heat could be produced and included in the set of UHI explanatory variables. However, it should be considered that this indicator has high spatial and temporal variability, meaning that the number of vehicles passing through some area depends on the time of the day and the main function of the area (e.g., city center, densely populated neighborhoods, commercial area, etc.). Therefore, the generated anthropogenic heat has high variability during the different times of the day and the different parts of the cities.

However, in the current work, a simplified approach based on the density of the roads has been adopted. For this purpose, the TOP10NL vector data, derived from the PDOK online data catalog, have been used. This dataset contains the road network of The Hague in two formats – either as polygons or lines. Thus the vehicle traffic density indicator has been calculated in two ways – including the areas of the roads or the length of the network within the grid cell (Figures 23 a-b).

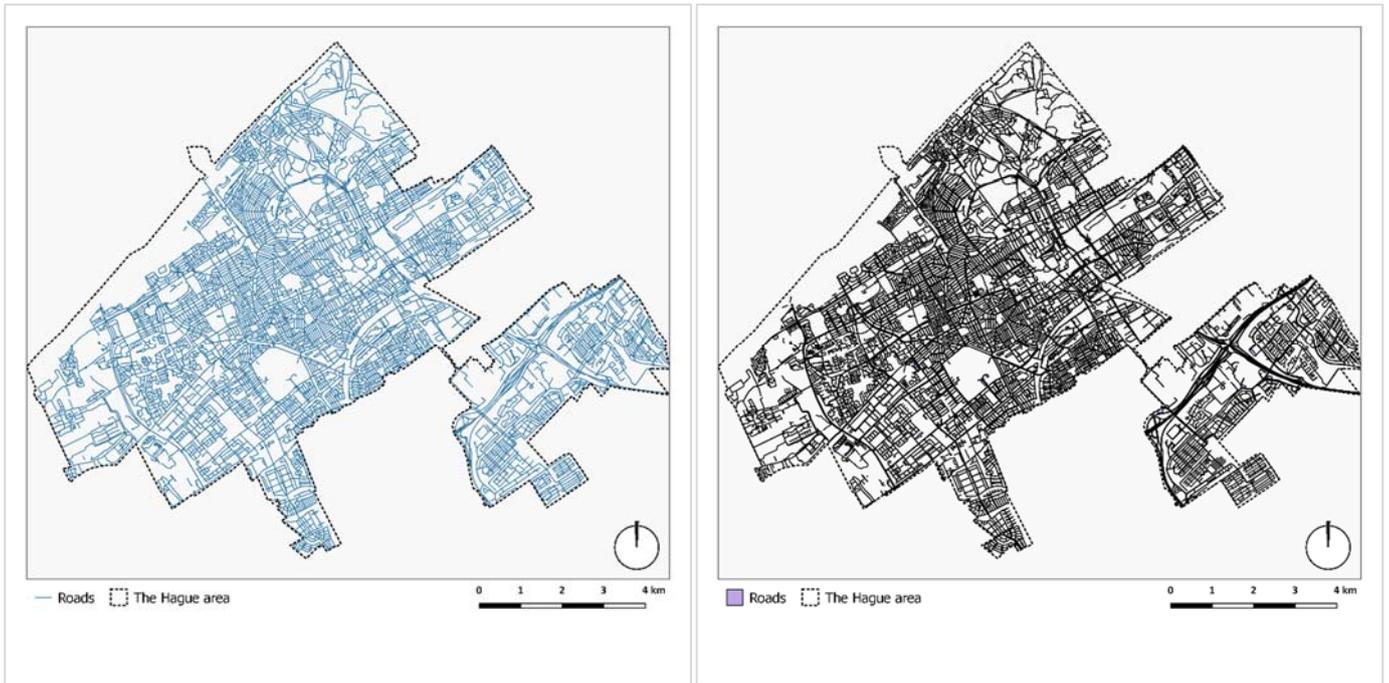


Figure 23:(a) Road network (lines); (b) Road network (polygons)

The data is provided as a tiled service and therefore had to be clipped to the area of interest. Further the intersection between the road polygons and lines with the different grid types was computed.

In addition, filtering based on the type of the main traffic has been performed. The dataset contains attribute classification of the roads, based on their main functions – bike paths, pedestrian lanes, roads, etc. Therefore, the roads, which don't contain automobile traffic were excluded from the dataset. Thus, the relation of this indicator to the possible heat generated from the vehicles has been preserved despite the simplified calculation process.

Further, the calculation of the final values was performed in a different manner depending on the different feature types. After obtaining the intersection of the roads with the grid cells, the road areas were dissolved by cell id, which resulted in one polygon per cell. Finally, the areas of all polygons were calculated and divided by the cell area. In contrast, the road lines were not dissolved, but the length of all roads was calculated and summed for every cell (Figure 24).

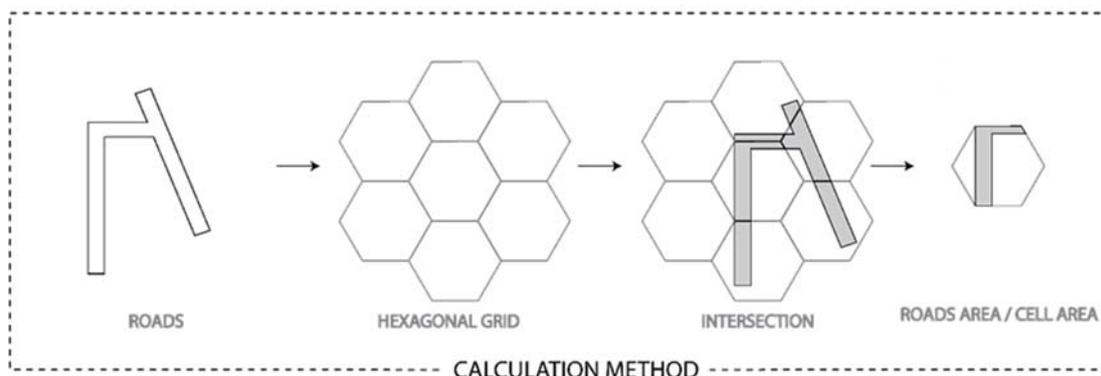


Figure 24:VTD calculation method

The outcomes indicated that using only the length of the roads results in very small values, which can be considered insignificant. In contrast, the road areas were more comparable to the cells and therefore resulted in preferable values for the modeling purposes.

### 4.3. Results comparison and conclusions

The results obtained from the spatial modeling phase represent six different indicators, which have been calculated for every grid cell from the hexagonal, rectangular and Voronoi grids. All indicators except the Buildings density are normalized and therefore vary from 0 to 1, where 1 represents full coverage of the cell. In addition, it is important to mention that all variables have purely spatial character and have been derived from raw spatial datasets with the means of the Geographical information system. In contrast, many different studies, which aim at statistical modeling use as independent variables indicators originating from diverse fields like economics, social sciences, health or others. The aim of the current work is to explore how the spatial characteristics of the city influence the Urban Heat Island effect. These results can provide input for further research and exploration of different planning alternatives and how they influence the temperatures in the cities. Therefore, the spatial indicators calculated in the previous step are the main input variables for the statistical models. Further, for the analysis of the obtained results, every indicator has been calculated and mapped separately for the different spatial tessellations. In Figures 25 a-c, the resulting maps of the Buildings density index are represented.

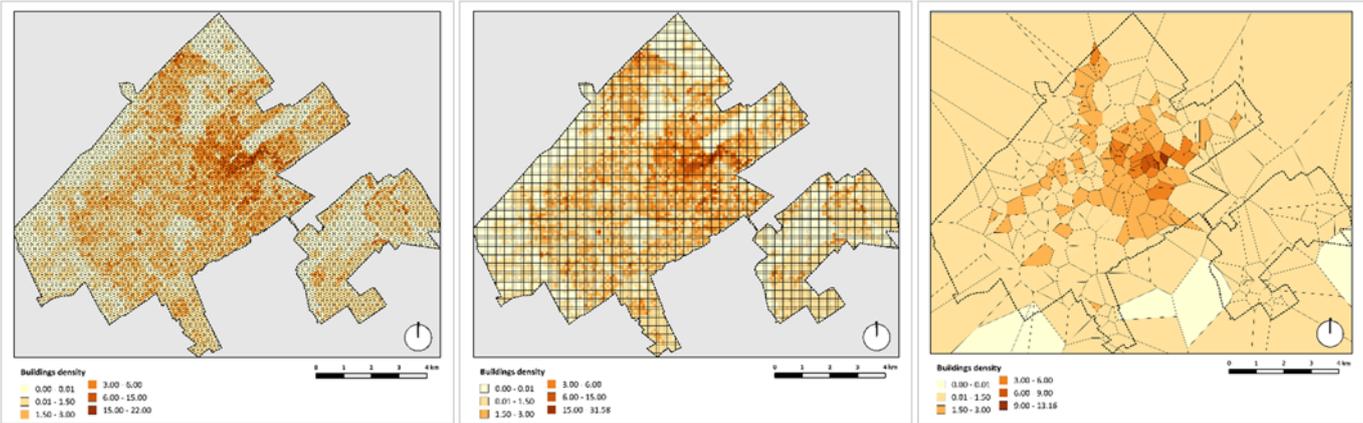


Figure 25: (a) Buildings density (hexagonal grid); (b) Buildings density (rectangular grid); (c) Buildings density (Voronoi diagram)

This is the only indicator that exceeds values of one due to the fact that the three-dimensional geometry of the buildings has been compared to the areas of the grid cells and therefore these values are not normalized. The resulting maps clearly show the spatial distribution of the densely populated areas in the city of The Hague, which correspond to the areas around the main train station. This is the main commercial and office area of the city, where the tallest buildings are concentrated. According

to the different grid types and the difference in the total areas of the cells, the maximum values of this indicator vary. The highest values of the Buildings density are observed for the rectangular spatial aggregation, while the smallest - for the Voronoi tessellation. This variation of the values is expected and it can be explained by the difference of the total areas of the cells.

Further, the second spatial indicator which was calculated is the Land cover index. The results can be seen in Figures 26 a-c.

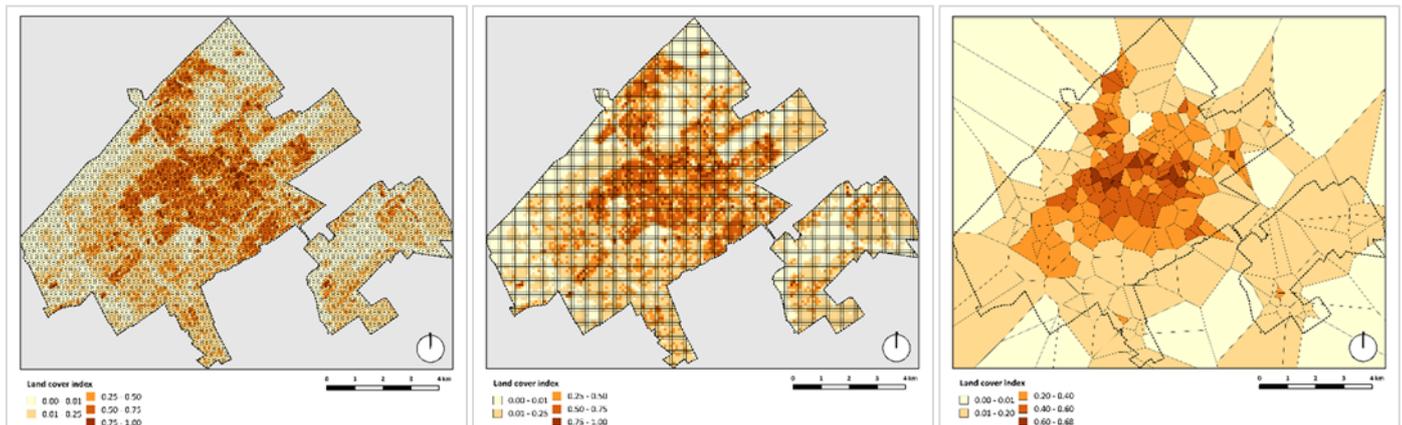


Figure 26:(a) Land cover index (hexagonal grid); (b) Land cover index (rectangular grid); (c) Land cover index (Voronoi diagram)

Although the Land cover index is directly connected to the built environment in the city (similar to the Buildings density), the spatial distribution of the highest values is different. The city center of The Hague is characterized by the highest density of buildings, where a lot of the cells of the grids have more than 50% of their areas occupied by buildings plots. Generally, the Land cover index represents well the structure of The Hague and provides a different perspective on the geometry of the city.

Further on Figures 27 a-c, the spatial distribution of the Non-permeable surfaces in the city can be observed. It can be noted that this indicator is calculated based on the open areas of the city, thus the buildings plots haven't been included in the index. Again, the highest concentration of sealed surfaces can be observed in the central parts of the city. In addition, it should be mentioned that compared to the Voronoi diagram, the regular grids provide a great amount of detail, where the streets structure of the city is visible. Generally, the Voronoi diagram provides a smoother representation of the indicator, where only a few cells have areas of 75% to 100% occupation with non-permeable surfaces. In contrast, visibly greater number of cells from the same class can be observed in the other two tessellations.

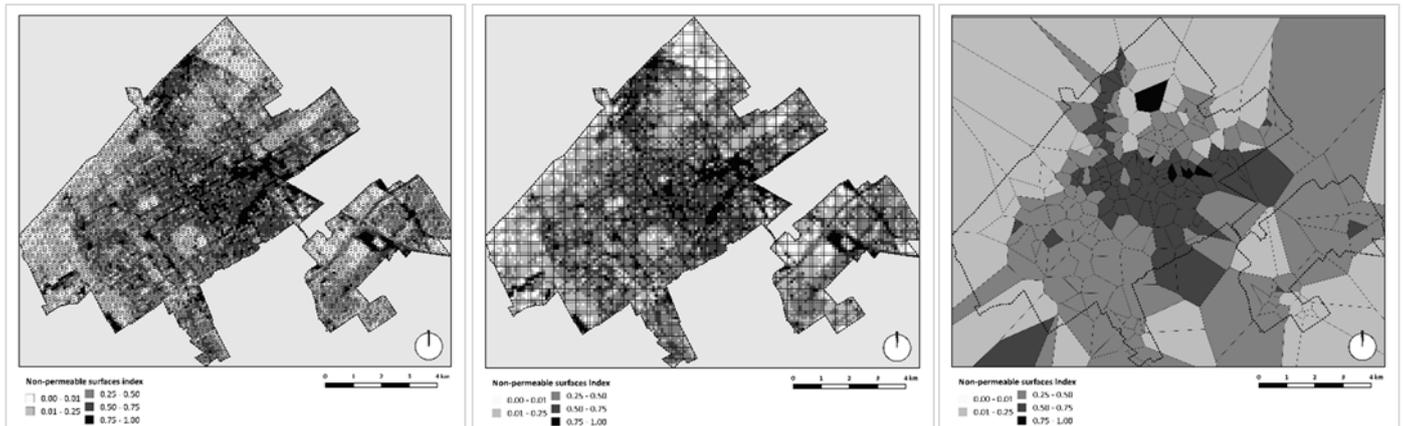


Figure 27: (a) Non-permeable surfaces (hexagonal grid); (b) Non-permeable surfaces (rectangular grid); (c) Non-permeable surfaces (Voronoi diagram)

The next indicator is the Sky View Factor and the obtained results are visualized in Figures 28 a-c. Similar to the previous variables, the SVF ranges between 0 and 1, where the lower values denote the bigger percentage of obstructed sky and higher values correspond to open areas. On the maps, it can be observed that mostly the peripheral areas of the city have higher SVF (the coastal area, industrial zones and big transport junctions). In contrast, the central parts of the city have between 60% to 70% of their areas open. Important to be mentioned here is that the trees in the city are included in the calculation of the indicator. Therefore the areas with the lowest SVF values correspond to the parks or other areas with high vegetation.

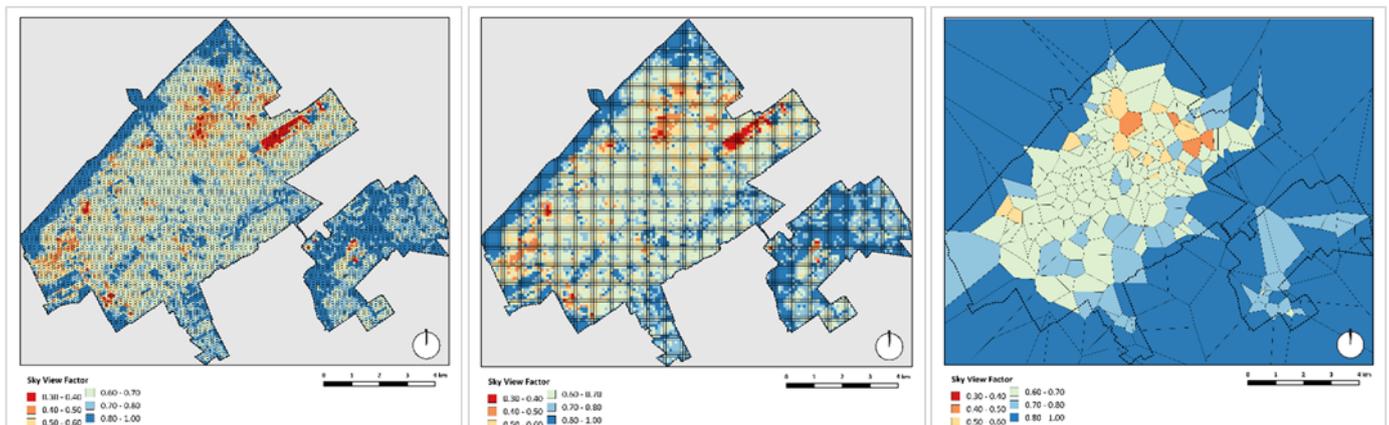


Figure 28: (a) Sky View Factor (hexagonal grid); (b) Sky View Factor (rectangular grid); (c) Sky View Factor (Voronoi diagram)

In the following figures (Figures 29 a-c) is the vegetation coverage of the city illustrated. The similarity of the spatial distributions of the SVF and the Vegetation index is observable. Mainly the boundary areas of the city have higher percentages of vegetation, which is in relation to the higher percentages of buildings and man-made surfaces in the central areas. In central parts of the city are concentrated the cells with less than 25% of their areas occupied with vegetation. In addition, a big part of these cells has less than 1% vegetation, while in the peripheral areas most of the cells contain more than

75% of vegetated areas. These results are highly affected by the chosen method of calculation of the Vegetation indicator, where the dense green areas have been extracted from the NDVI raster, leaving out the sparsely vegetated areas of the city.

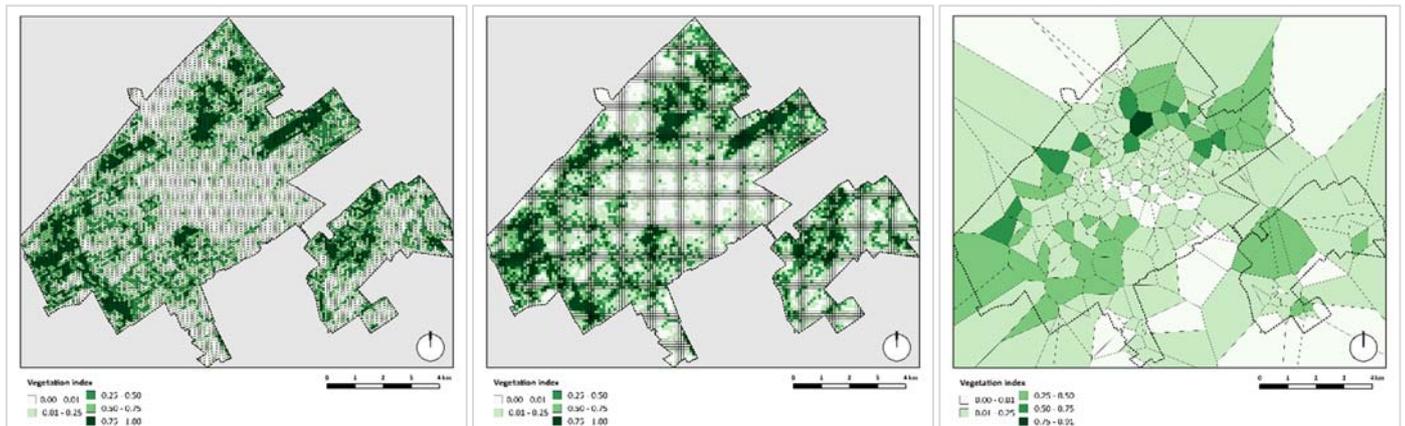


Figure 29: (a) Vegetation coverage (hexagonal grid); (b) Vegetation coverage (rectangular grid); (c) Vegetation coverage (Voronoi diagram)

The final calculated indicator is illustrated in Figures 30 a-c. The Vehicle traffic density is representing the density of the road network in the city of the Hague. Most of the grid cells have from 15% to 45% of roads coverage and only a few cells reach the highest values of roughly 80% road coverage. It should be mentioned that considering the detail of roads in the city (they don't represent big clustered areas but are linear structures) the Voronoi diagram doesn't provide detailed enough representation of the road network in the city. On the other hand in the regular grids, the structure of the road network in the city can be seen. For example, in the south-west part of The Hague, the main highway of the city is clearly visible. In the central parts these patterns are less clear due to the higher diversity and density of the urban fabric. In addition, it should be mentioned that considerable amount of these streets contain mixed traffic including bike paths and pedestrians, where cars are not that predominant.

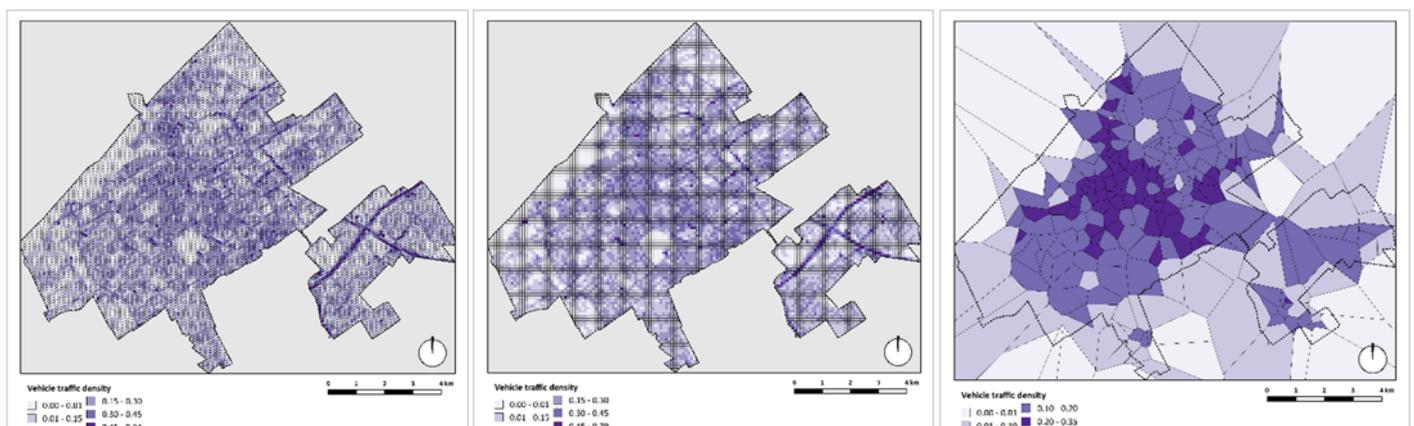


Figure 30: (a) Vehicle traffic density (hexagonal grid); (b) Vehicle traffic density (rectangular grid); (c) Vehicle traffic density (Voronoi diagram)

Generally, the difference between the two regular grids is small and it can be concluded that they provide a good amount of detail in the spatial representation of the morphological characteristics of the city of The Hague. In contrast, the varying sizes of the Voronoi tessellation result in bigger areas and overall averaging of the indicators. Further, compared to the rectangular grid, the hexagons provide a better visual representation of the curvature of the data, in contrast to the rectangular grids where the attention is focused on the straight parallel lines of the squares.

Some advantages of the rectangular tessellations are that they are simpler to define, which can be relevant for the storage of big datasets. In addition, this tessellation is easily scalable and it is suitable for raster operations. However, for the current research, the hexagonal grid is generally performing better than the other two tessellations, because of its better properties for visualization and its edge reducing effect.

#### 4.4. Spatial join

The final step preceding the statistical modeling of the temperatures in the city of The Hague is the definition of the relation between the calculated spatial indicators and the location of the sensors in the city. This is a crucial component of the final spatial model and has a direct implication on the statistics.

The definition of the morphology of the city has the purpose of constructing the main Urban Heat Island contributing factors, while the relationship between these variables and the sensors locations, implies the area of influence of these spatial indicators.

The inputs for this step are the three different grid types, which contain the values of every indicator for every grid cell and the locations of the sensors, which provide information about the temperatures in the city. In order to construct the input parameters for the statistical models (dependent and independent variables), the two different types of datasets have to be spatially associated. This can be achieved using GIS software and different spatial relation operations.

The first method that has been adopted (Figure 31) is to directly associate the spatial indicators with the temperatures, which are measured by the sensors. Therefore, the joining has been performed directly. Thus, the spatial indicators of the cell that the sensors belong to have been directly added to the sensor dataset. Theoretically, this association means that the temperatures have been directly influenced by the spatial characteristics of the city related only to the current cell.

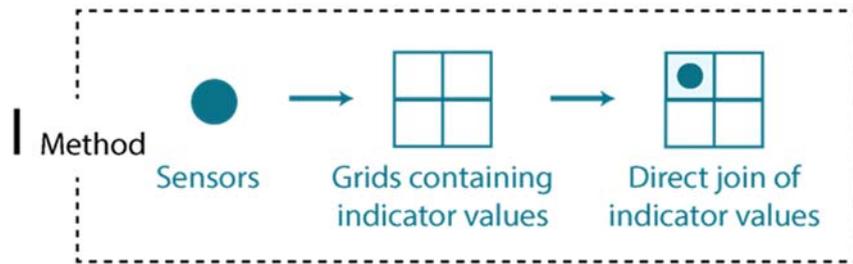


Figure 31: Method of direct spatial relation

Considering the complexity of the urban environment and the interdependency of all processes in the city, the pure direct association between the temperatures and the space characteristics can be considered as unrealistic and theoretical. Therefore, another approach of association has been adopted. The second method (Figure 32) takes into account the interdependency of the different grid cells, which effectively represent only a theoretical division of the space. Therefore, the association between the spatial indicators and the temperatures is performed considering an area of influence. For this purpose, a buffer of 100m around every sensor has been calculated and intersected with the three different grid types. The resulting polygons represent whole cells or parts of a grid cell, which fall within this area of influence. Hence, the resulting indicators represent area weighted averaged values.

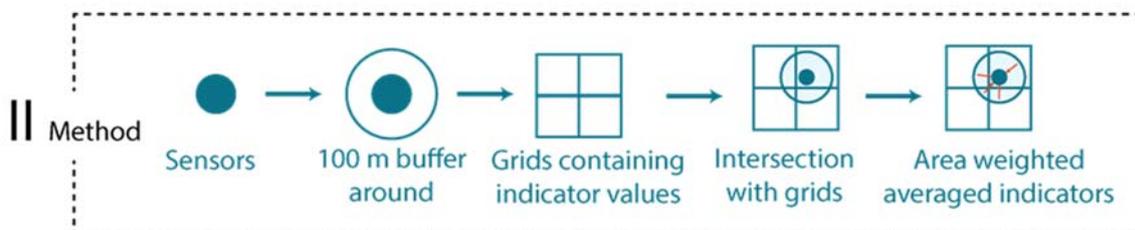


Figure 32: Area-weighted averaged indicators



# STATISTICAL ANALYSIS

## 5. Statistical modeling

The following section of the research examines the principles of the statistical modeling and the methodology and results obtained from this work. Different statistical models aiming at the prediction of extreme temperatures have been implemented and compared. The goal of this work is to create a predictive tool, which can be used for the analysis of different planning alternatives and their effect on the urban microclimate. Therefore, the relationship between the urban morphology and the air temperatures in the city has to be established. This has been done with the means of regression analysis, where different alternatives and assumptions have been evaluated. Further, the role of geographic weights and their modeling enhancement characteristics have been analyzed.

Four different types of models have been implemented and the performance and the quality of the results have been compared in this section. Different characteristics of the models have been considered in the analytical process. First, the coefficients of the independent variables, between the models have been analyzed. These coefficients define the direction and magnitude of the relationship between the explanatory and dependent variables. In addition, the significance of every variable was examined by the probability values (p-values) of every model. Finally, the performance and goodness of fit of the models have been compared considering the R-squared ( $R^2$ ) values and the Akaike information criterion (AIC). The so-called R-squared values represent a statistical measure of the closeness of the observed data to the estimated regression line. Its values vary between 0 and 1, where zero indicates no correlation between the modeled variables. The R-squared measure is the most common performance criteria for the simple statistical models, which is not always reliable measure for the spatial models though. Therefore, a more robust measure of the models' performance has been incorporated as well – the Akaike information criterion. This indicator provides information about the quality of the statistical model, based on its goodness of fit and complexity. AIC can be used only for comparison purposes, where lower values indicate better performance of the models (Anselin, n.d.-a).

In the modeling implementations, tests against all these criteria have been incorporated. The main software that has been used for the statistical analysis is Geoda<sup>4</sup>. This is an open source statistical software, developed by Luc Anselin and his team. The program incorporates a great amount of functionality, including implementation of the spatial lag and the spatial regression models, the correlation analysis and the basic assumptions tests. In addition, for the Geographically weighted

---

<sup>4</sup> <https://geodacenter.github.io/>

regression models a separate software – GWR4<sup>5</sup> - developed by Professor Tomoki Nakaya, from the Ritsumeikan University in Japan, was used.

## 5.1. Analysis of the independent variables

Based on the prior assumptions of the Ordinary Least Squares regression model (explained in Section 2.5) a statistical analysis of the dependent and independent variables was performed. The spatial autocorrelation is characterized by the clustering of similar values in nearby locations. When high values are found close to each other, e.g., big clusters of similar values, it is referred to as positive correlation. On the other hand, negative autocorrelation can be found in locations, where high values are surrounded by a cluster of low values or vice versa (Anselin, n.d.-b). The presence of spatial autocorrelation between the variables violates the general assumptions of independence and stationarity of the OLS regression models, which can affect their performance and reliability. In addition, the presence of spatial autocorrelation indicates that the relationships between the variables differ in space and therefore a global model could be unsuitable for the development of the statistical analysis. On the other hand, local models, which are estimated uniquely for every location based on its neighbors as well, can provide a better understanding of the phenomenon (Fotheringham, 2009).

For the analysis of the spatial autocorrelation of the measured temperature differences by the Netatmo sensors, global and local autocorrelation statistics have been utilized. The global Moran's I scatterplot is presented in Figure 33a. The Moran's I is a global inferential statistic measuring spatial autocorrelation. The interpretation of the results can be done by computing statistical significance and its p-values ("Moran's I: Definition, Examples," n.d.). The statistic analyzes the data under the null hypothesis of random distribution, therefore low significance values refer to the rejection of the null hypothesis and the presence of spatial autocorrelation. The scatterplot consists of the original (x-axis) and spatially lagged (y-axis) variables. In addition, each quadrant of the graph corresponds to a different type of spatial autocorrelation – positive (high-high and low-low clusters) and negative (high-low and low-high) ("Global Spatial Autocorrelation (1)," n.d.). On the scatterplot below (Figure 33a) the presence of distinct outliers can be seen in three out of the four quadrants in the plot. Additionally, the computed Moran's value is 0.15 with its pseudo p-value of 0.002 indicating strong rejection of the null hypothesis of independence. The combination of the visual information from the scatter plot and the computed significance value show clear evidence for the presence of spatial autocorrelation in the current model. Since the Moran's I is a general measure of the presence of spatial autocorrelation and it does not provide any further information about the specific locations and patterns, the Local Indicators of Spatial Autocorrelation (LISA) have been investigated as well. These indicators have been

---

<sup>5</sup> <http://gwr.maynoothuniversity.ie/gwr4-software/>

used for the definition of the location and the magnitude of the spatially autocorrelated values. In Figures 34 a-b are the Cluster and Significance maps of the correlated variables presented. It can be observed that 51 of the selected variables are autocorrelated and depending on the type of correlation they have been classified into four different groups, representing the positive and negative autocorrelations. According to the p-values, 30 of these locations are statistically significant, considering the threshold value of 0.05. In conclusion, based on the spatial autocorrelation analysis of the temperature measurements all autocorrelated dependent variables have been removed from the modeling group, assuring the preservation of the independence and stationarity of the variables.

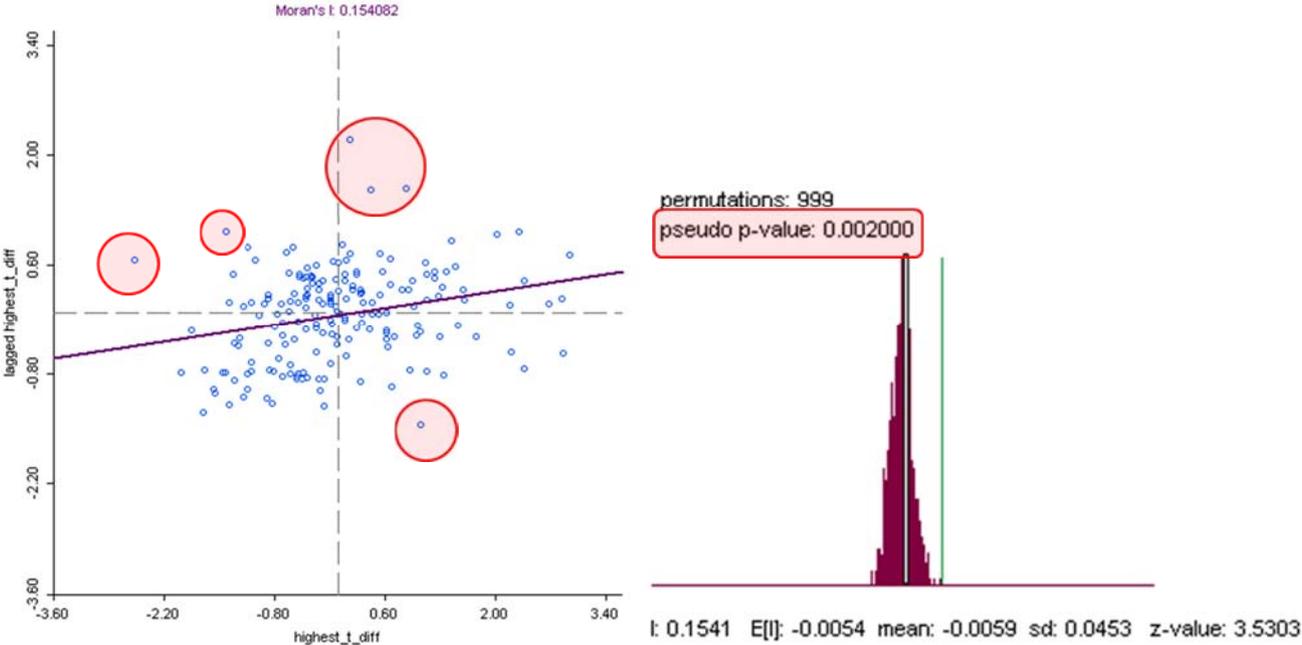


Figure 33: (a) Moran's I scatterplot, (b) Reference distribution

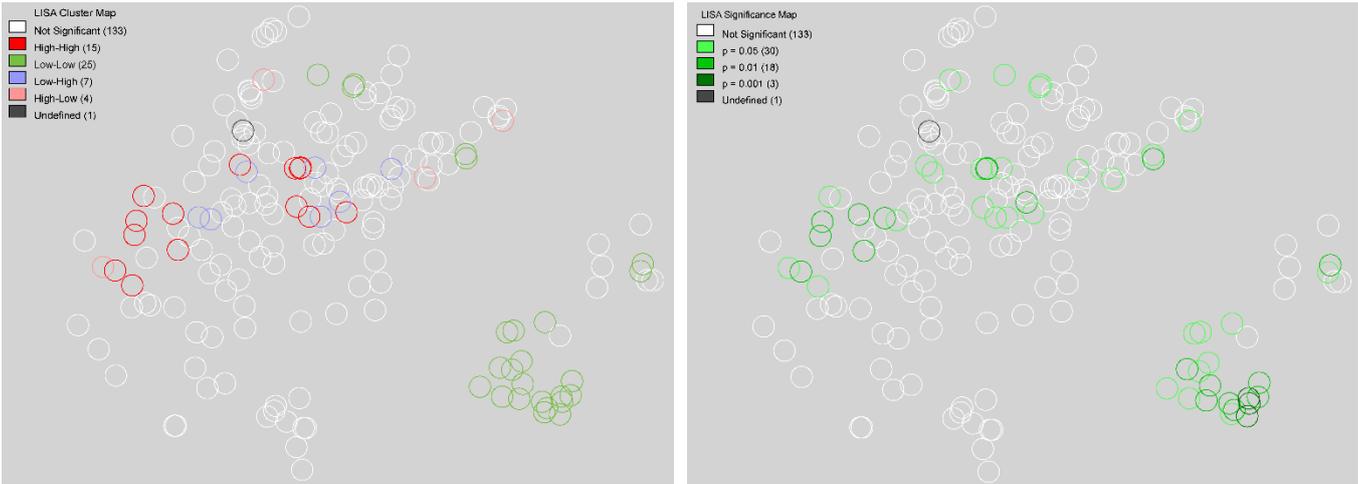


Figure 34: Local indicators of spatial association (LISA) (a) Cluster map; (b) Significance map

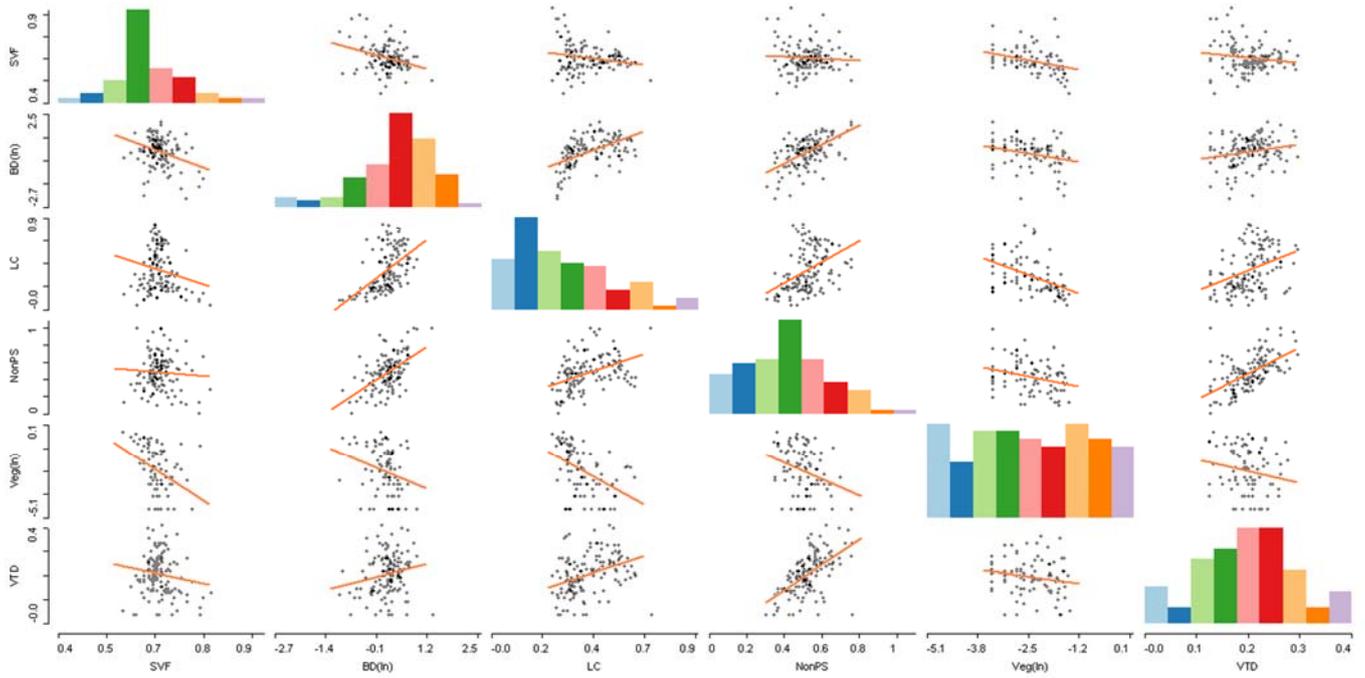
In addition to the autocorrelation analysis, multicollinearity tests of all independent variables have been performed. Multicollinearity is observed when independent variables are correlated with each

other. The presence of multicollinearity can affect the results and the accuracy of the regression model because of the higher standard errors of the variables in the equation. This can lead to false computations of the significance of some variables (“The problem of multicollinearity,” 1997).

Depending on the definition of the spatial representation of the UHI contributing factors and the different association methods which were tested, five different statistical models have been explored. Since the models are based on different spatial tessellations and aggregation areas, the relationships between the different variables can differ. Therefore, the multicollinearity of the independent variables has been explored and compared between the five models using scatterplot matrices, which define the correlation between the variables. It has been observed that the correlation between the explanatory variables is very similar between the rectangular and hexagonal models, including their area-weighted averaged (AWA) variations. Therefore, the scatterplot matrices of the hexagonal and Voronoi models have been presented in Figures 35 a-b. Generally, substantial differences between the multicollinearity of the variables can be observed between these two models only, which can be attributed to the significant difference between the sizes of the cells of the two models.

Regarding the Figures 35 a-b, the analysis of the correlation between the variables is based on the visual inspection of the plots where the clustering and the closeness of the data points to the regression line indicate the interdependence of the variables. In the graphs below it can be observed that the Buildings density index is related to the Sky View Factor, the Land cover index, the Non-permeable surfaces, the Vegetation index and the Vehicle traffic density. Therefore this variable has to be used with caution in the statistical phase and if not significant, it has to be removed from the model. In addition, the Non-permeable surfaces index is related to the Vehicle traffic density, which is an expected result, because the Vehicle traffic density represents the roads density, which is part of the impermeable surfaces in the city. Generally, the correlations of the Buildings density index with the other indicators are weaker in the Hexagonal model, but still present. In addition to that, the aggregation in the Voronoi representation leads to a more pronounced correlation between the Non-permeable surfaces and the Land cover index and the Vehicle traffic density; and between the Land cover index and the Vehicle traffic density. These differences can be attributed to the generally bigger areas of the Voronoi cells, where the spatial characteristics tend to be more correlated.

These observations suggest, that these variables have to be used with caution in the regression models and if multicollinearity problems occur, the highly correlated variables in the model can be removed.



(a) Scatter plot matrix (hexagonal grid)

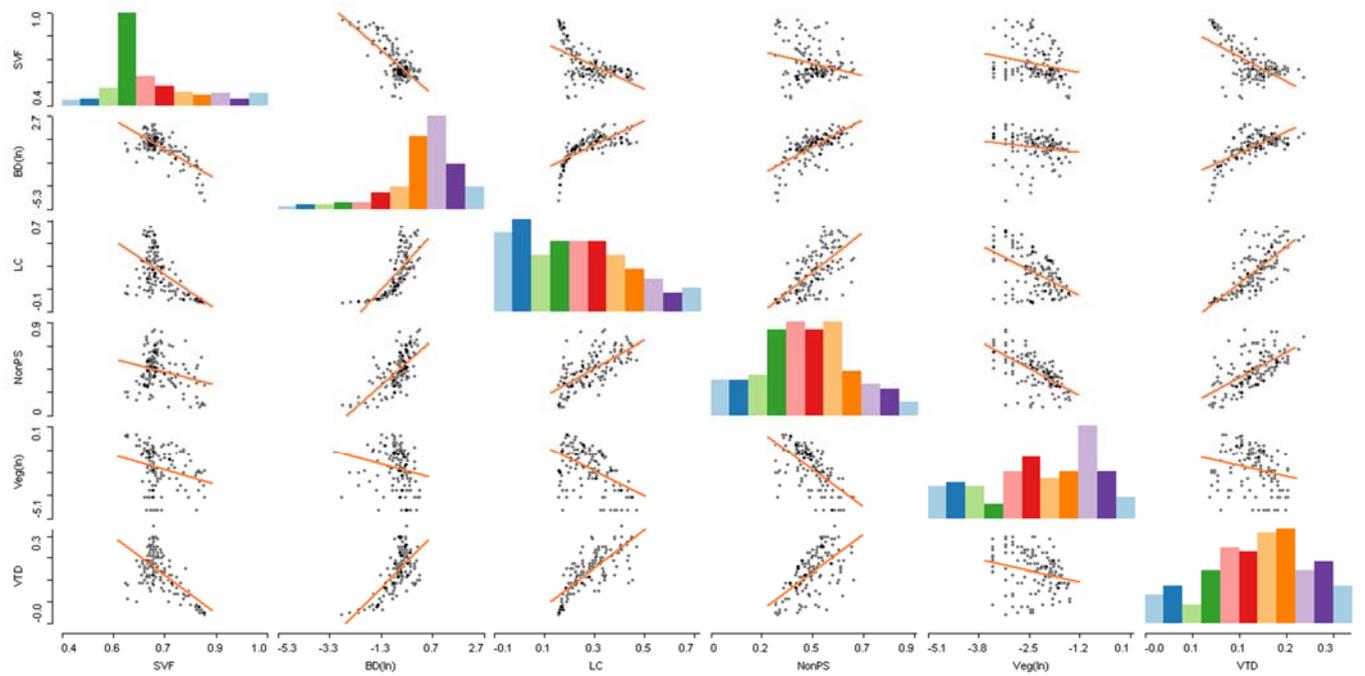


Figure 35: (b) Scatterplot matrix (Voronoi diagram)

## 5.2. The first stage of modeling

Following the analysis of the independent variables, the first stage of statistical modeling was performed. In this phase, the Ordinary Least Squares regression model was developed for all spatial models and the results have been compared.

As explained in Section 4.4, five different spatial models representing the morphological characteristics of the Hague have been developed. They can be differentiated by two main

characteristics – the type of spatial tessellation (hexagonal grid, rectangular grid and Voronoi diagram) and the association methods between the observed temperatures and the spatial indicators (direct joins of the indicators to the sensors, based on the grid cell they belong to and the area-weighted averaged indicators). Considering these differences, the initial statistical analysis has been performed on all five models, with the goal of comparing their performance.

The first step of the development of the spatial or non-spatial regression models starts with the definition of the OLS model, which is tested against different assumptions and the presence of spatial dependency between the variables. For this purpose, the Geoda software has been used, where all spatial models have been imported separately. In order to define and test the spatial dependency of the data, a spatial weights matrix has to be computed. Since the dataset is representing points (the sensors locations with their attached spatial parameters) the most suitable weights representation is defined by the distance between the points. Thus, for every point, its neighbors and the distances between them have been defined based on a threshold value. The default threshold distance value is computed in such way that every point in the dataset has at least one neighbor.

After obtaining the spatial weights, the Ordinary Least Squares regression model can be created, including the diagnostics for spatial dependence. These tests essentially explore the two other alternative models, which are included in the statistical software – the spatial lag and spatial error models. The results from the OLS regression models are presented in Tables 8-10, where the general model performance in terms of goodness of fit, including all tests, coefficients and the significance of the independent variables are presented.

The initial R<sup>2</sup> results from the regression models are in the range of 0.03 and 0.04 (Table 8), which indicates an extremely small correlation between the independent and dependent variables. Also, all models experience multicollinearity issues (values above 30 are considered as problematic), which is expected result considering that all independent variables have been included at this stage. The other indices, which represent the goodness of fit of the models (AIC and Log-likelihood) indicate the presence of a very small difference between the five models.

*Table 8: Comparison of models' performance*

<b>MODEL</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup> - ADJUSTED</b>	<b>AIC</b>	<b>LOG - LIKELIHOOD</b>	<b>MULTICOLLINEARITY CONDITION NUMBER</b>
<b>HEX (DIRECT)</b>	0.03	-0.012	415.682	-200.841	45.70
<b>RECT (DIRECT)</b>	0.03	-0.003	414.425	-200.213	47.92
<b>VORONOI</b>	0.04	0.002	413.592	-199.796	71.12

<b>HEX (AWA)</b>	0.04	0.004	413.283	-199.642	62.81
<b>RECT (AWA)</b>	0.04	0.005	413.051	-199.526	62.50

Further, Table 9 represents the results of the tests against the main OLS assumptions – the normality of the errors, the presence of heteroscedasticity and the two alternative spatial models. Generally, for all models, it can be observed that there is evidence for the non-normal distribution of the errors – the low probability values of the Jarque-Bera tests indicate rejection of the null hypothesis of normality. Except this, all other tests have probability values exceeding the most commonly used significance threshold value of 0.05. Therefore, it can be concluded that there is no indication for the presence of heteroscedasticity in the models and that the spatial lag or regression models are not indicated as better alternatives at this stage.

Table 9: Results of the statistical tests

	JARQUE-BERA TEST (NORMALITY OF ERRORS)	BREUSCH-PAGAN TEST (HS)	KOENKER-BASSETT TEST (HS)	WHITE (HS)	MORAN'S I	LM (LAG)	ROBUST LM (LAG)	LM (ERROR)	ROBUST LM (ERROR)
<b>HEXAGONAL GRID (DIRECT)</b>									
<b>VALUE</b>	9.846	7.144	5.769	16.478	-1.230	1.432	2.547	2.060	3.175
<b>PROBABILITY</b>	<b>0.0072</b>	0.3077	0.4496	0.9432	0.2187	0.2315	0.1105	0.1512	0.0748
<b>RECTANGULAR GRID (DIRECT)</b>									
<b>VALUE</b>	10.629	5.665	4.407	24.108	-1.245	1.880	0.371	2.148	0.637
<b>PROBABILITY</b>	<b>0.0049</b>	0.4618	0.6218	0.6243	0.2132	0.1702	0.5426	0.1428	0.4247
<b>VORONOI DIAGRAM (DIRECT)</b>									
<b>VALUE</b>	13.301	4.797	3.751	23.611	-1.358	2.014	0.726	2.511	1.223
<b>PROBABILITY</b>	<b>0.0019</b>	0.5701	0.7104	0.6518	0.1745	0.1558	0.3941	0.1131	0.2688
<b>HEXAGONAL GRID (AWA)</b>									
<b>VALUE</b>	9.7408	7.9077	6.6003	27.3523	-1.2625	1.7813	0.9030	2.3364	1.4580
<b>PROBABILITY</b>	<b>0.0077</b>	0.2449	0.3594	0.4449	0.2068	0.1820	0.3420	0.1264	0.2272
<b>RECTANGULAR GRID (AWA)</b>									
<b>VALUE</b>	11.131	5.429	4.452	23.515	-1.252	1.818	0.608	2.294	1.085
<b>PROBABILITY</b>	<b>0.0038</b>	0.4900	0.6158	0.6570	0.2105	0.1776	0.4354	0.1298	0.2976

Finally, the coefficients of the estimated regression variables and their significance levels are presented in Table 10. As a whole, the significance level of all variables is lower than the presumed thresholds ( $t\text{-value} > |2|$  or  $p\text{-value} < 0.05$ ). It can be observed that the models representing the direct join between the temperatures and the spatial indicators have the lowest performance in

terms of the significance of the coefficients, where only the Non-preamble surfaces index for the hexagonal grid can be considered as significant. Overall the differences between the hexagonal, rectangular and Voronoi grids are very small. In contrast, both representations of the area weighted averaged indicators have slightly higher significance values, where Sky View Factor, Non-permeable surfaces and Vehicle traffic density are approaching the significance threshold. It is important to consider that these values are affected by all variables included in the model, which means that the removal of all insignificant indicators can improve the overall performance of the others.

However, generally the results from the initial OLS models give very low indications of the correlation between the measured extreme temperatures and the morphological characteristics of The Hague. In addition, the difference between the performance of the direct models is very small. Compared to the direct relationships, the models incorporating the broader area of influence of the variables yield better results. Therefore, for the second stage, where different approaches for improving the results are explored, the area-weighted averaged models have been considered only.

Table 10: Independent variables coefficients and significance

	SVF	BD	LC	NONPS	VEG	VTD
HEXAGONAL GRID (DIRECT)						
COEFFICIENT	-1.2818	-0.0586	-0.2475	1.3438	-0.2010	-1.0807
T-STATISTIC	-0.7739	-0.9129	-0.4127	<b>1.7952</b>	-0.2506	-0.7489
P-VALUE	0.44039	0.3629	0.6804	0.0748	0.8024	0.4552
RECTANGULAR GRID (DIRECT)						
COEFFICIENT	-0.9518	0.0325	-0.3079	0.6709	-0.1239	0.2965
T-STATISTIC	-0.5621	0.5851	-0.5267	0.8781	-0.1441	0.1898
P-VALUE	0.5749	0.5594	0.5992	0.3814	0.8856	0.8497
VORONOI DIAGRAM						
COEFFICIENT	-2.1701	-0.0334	0.3877	1.0948	-0.6776	-3.4852
T-STATISTIC	-1.0278	-0.2994	0.3308	1.1077	-0.5313	<b>-1.4100</b>
P-VALUE	0.3059	0.7651	0.7412	0.2699	0.5960	0.1608
HEXAGONAL GRID (AWA)						
COEFFICIENT	-2.4328	-0.0751	0.3085	1.9227	0.1071	-3.6976
T-STATISTIC	<b>-1.1616</b>	-0.8635	0.3567	<b>1.7670</b>	0.0898	<b>-1.6608</b>
P-VALUE	0.24747	0.3893	0.7218	0.0795	0.9285	0.0991
RECTANGULAR GRID (AWA)						
COEFFICIENT	-2.8955	-0.0784	0.1512	2.0012	-0.0317	-3.4379

<b>T-STATISTIC</b>	<b>-1.3974</b>	-0.9200	0.1687	<b>1.7872</b>	-0.0261	<b>-1.4907</b>
<b>P-VALUE</b>	0.1646	0.3592	0.8662	0.0761	0.9792	0.1384

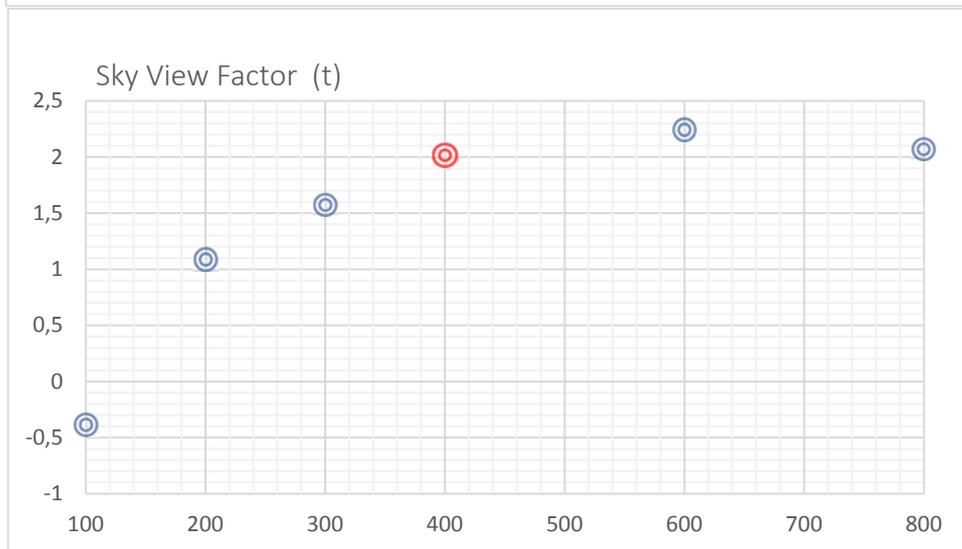
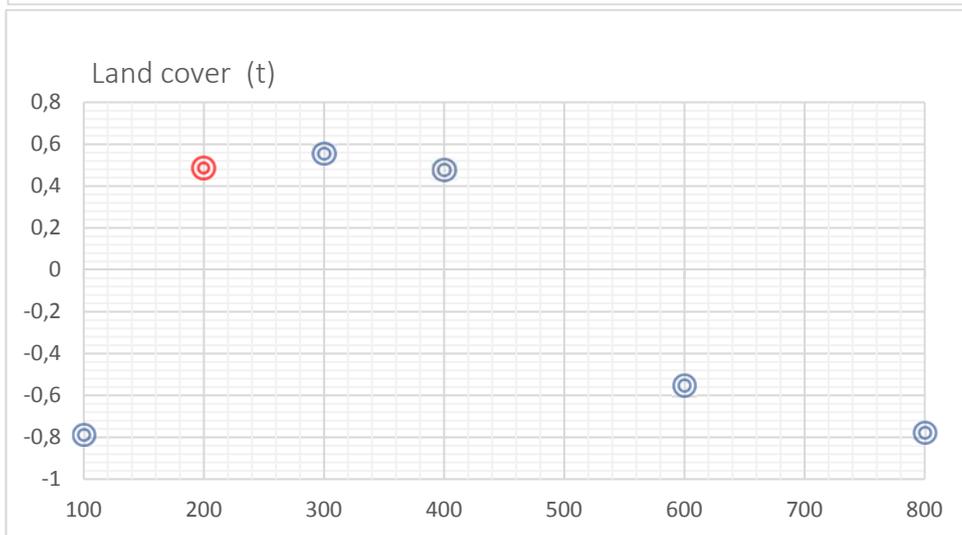
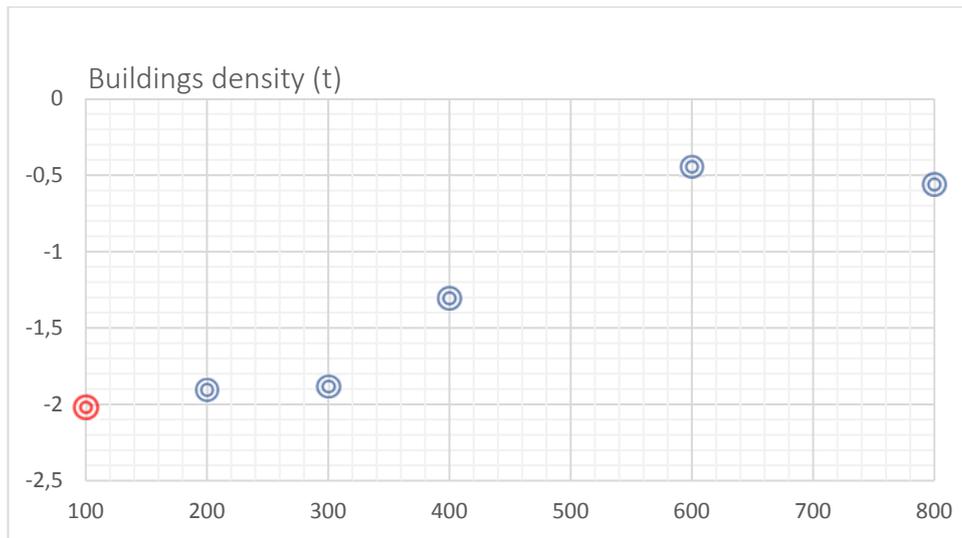
### 5.3. Improvement of the models

Following the results of the initial statistical analysis, multiple options for the improvement of the models were explored. The initial stage of the modeling showed a very limited correlation between the morphological characteristics of the city and the measured extreme temperatures. However, the comparison between the different models indicated that broader areas of influence of the computed indicators could be considered. Therefore, different areas of influence for the independent variables were explored by computing their area-weighted averaged values for different buffer sizes, generated from the sensor dataset. Further, the statistical significance of all variables considering the different radiuses of influence was compared. For this purpose, all indicators have been computed using six different buffer sizes – 100m, 200m, 300m, 400m, 600m and 800m of their areas of influence. At this stage, the area-weighted averaged hexagonal spatial model was used only. Moreover, six different OLS models were developed, considering the diverse buffer sizes. The significance of the explanatory variables has been compared in terms of the computed t-values and the results are illustrated in Figure 36 a-f. Generally, it can be observed that the all variables have different areas of influence, which are relevant for the OLS models. Some variables improve their importance with the bigger aggregation areas – Sky View Factor, Vegetation index and Vehicle traffic density. On the other hand, Buildings density, Land cover index and Non-permeable surfaces have higher absolute t-values for the smaller areas.

On the graphs below (Figure 36 a-f) are the best-suited radii for every variable shown in red. The choice of the best averaging radius was driven by the trade-off between the significance of the factors and the size of the buffer. The goal was to achieve as high importance as possible for the smallest possible radii. The reason for this choice lies in the fact that the usage of big buffer sizes leads to a smoothing of the results. Also, when too big aggregation areas are used in spatial studies, the data can experience high levels of bias.

Further, based on the combination of optimal areas of influence an improved version of the initial hexagonal area-averaged model was developed. Again, the significance levels of all variables have been analyzed. The insignificant variables were excluded from the final model in order to increase its performance and to remove the correlated indicators. Moreover, the spatial lag, spatial error and Geographically weighted regression models were developed and compared to the OLS regression.

The results of the optimized models are presented and further analyzed in the following section.



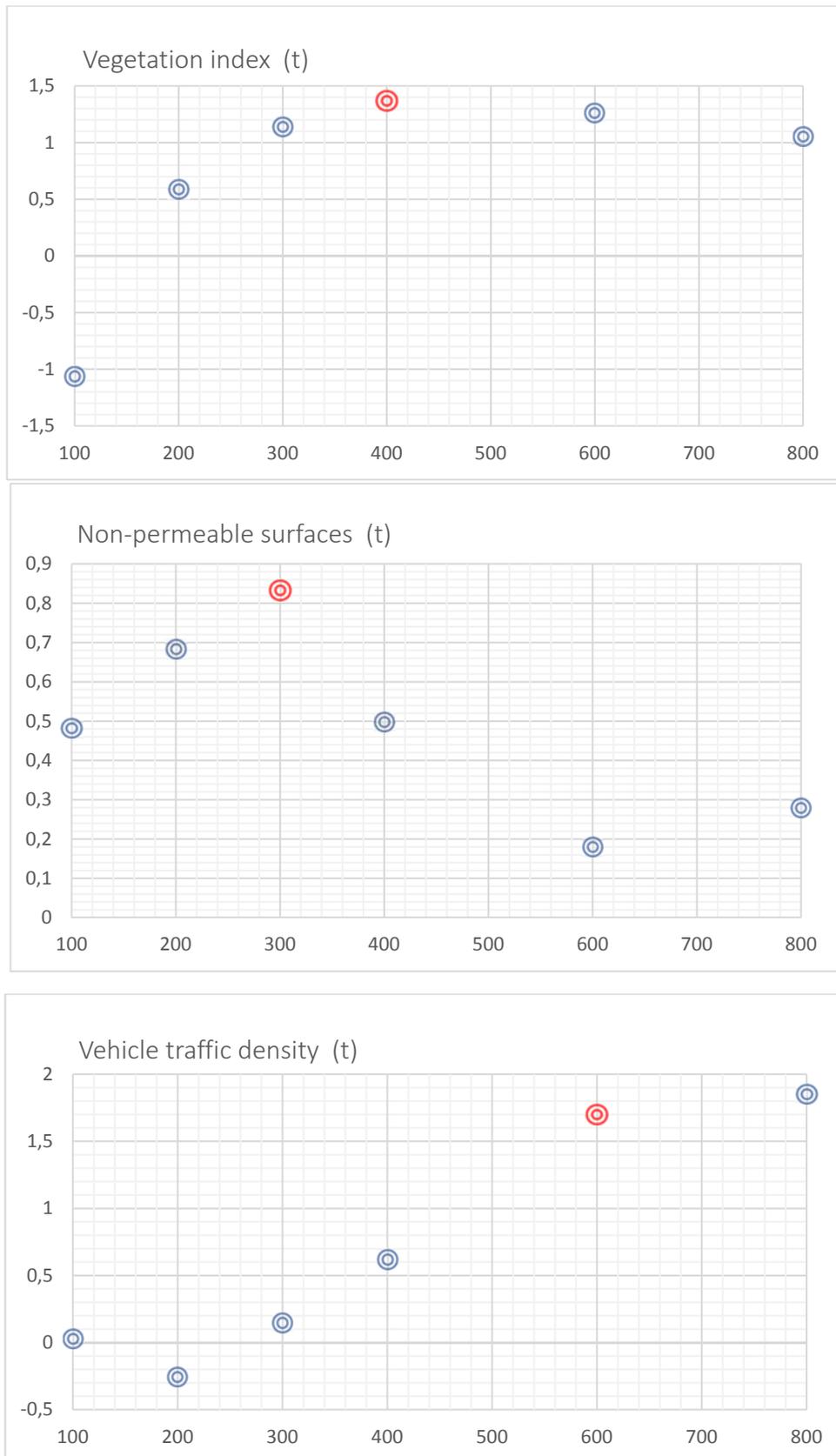


Figure 36: (a-f) T-scores against buffer radii

## 5.4. Final model

Based on the preceding steps of the initial statistical analysis and the exploration of the methods for improvement of these results, the final models were developed. These models were built using the optimal combination of explanatory variables, which was selected in the previous step. The results are represented in Tables 11-13 and for comparison purposes, the initial OLS hexagonal model is included.

As indicated in the previous chapters, the statistical analysis starts with the Ordinary Least Squares model, against which all other variations are compared. At this stage, the OLS model was developed using all indicators with their varying areas of influence. This model was then reduced to the significant variables only, which are Sky View Factor and Non-permeable surfaces. The exclusion of the insignificant indices is an important step in the regression analysis, which results in overall better performance and stability of the model and thus reduction of the multicollinearity problems. This can be observed in the adjusted  $R^2$  value, which is higher for the refined OLS model (Table 11). Also, compared to the first models which were developed, the refined simple linear regression has considerably improved.

*Table 11: Model's performance*

<b>MODEL</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup>- ADJUSTED</b>	<b>AIC</b>	<b>LOG - LIKELIHOOD</b>	<b>MULTICOLLINEARITY CONDITION NUMBER</b>
<b>OLS (INITIAL STAGE)</b>	0.04	0.004	413.283	-199.642	62.810
<b>OLS</b>	0.11	0.07	431.275	-207.638	163.427
<b>OLS (REFINED)</b>	0.10	0.09	422.786	-208.393	31.8182
<b>SPATIAL LAG</b>	0.11	-	424	-208	-
<b>SPATIAL ERROR</b>	0.11	-	422.221	-208.110	-
<b>GWR</b>	0.20	0.15	424.767	-	-

Further, all statistical tests except one suggest that there is no rejection of the null hypothesis, therefore no heteroscedasticity or autocorrelation is present in the model (Table 12). The single positive test indicates that the residuals of the OLS regression are not normally distributed. This means that there is a considerable difference between the predicted and observed values in the model. This disparity can be attributed to the spatial variance of the predictors in the model.

However, the performance of the two global spatial models in terms of suitability – spatial lag and spatial error - has no substantial difference. These two models include diagnostics for spatial

dependence which are not statistically significant for both of them, meaning that there are no other spatial effects present in the model. Moreover, the additional variables, which the two models introduce – the spatial lag and error terms are not statistically significant as well. Finally, compared to the refined OLS model the two explanatory variables – the Sky View Factor and the Non-permeable surfaces – have higher significance in the spatial regressions (Table 13). In conclusion, it can be observed that there is no significant difference between the global OLS, Spatial lag and Spatial error models.

Table 12: Statistical tests

	JARQUE-BERA TEST (NORMALITY OF ERRORS)	BREUSCH-PAGAN TEST (HS)	KOENKER-BASSETT TEST (HS)	WHITE (HS)	MORAN'S I	LM (LAG)	ROBUST LM (LAG)	LM (ERROR)	ROBUST LM (ERROR)
<b>OLS (INITIAL STAGE)</b>									
<b>VALUE</b>	9.741	7.908	6.600	27.352	-1.262	1.781	0.903	2.336	1.458
<b>PROBABILITY</b>	0.0077	0.2449	0.359	0.4449	0.2068	0.1820	0.3420	0.1264	0.2272
<b>OLS</b>									
<b>VALUE</b>	16.484	7.055	4.978	21.307	1.474	1.092	0.591	0.739	0.239
<b>PROBABILITY</b>	0.0003	0.4232	0.6626	0.9667	0.1404	0.2961	0.4418	0.3898	0.6248
<b>OLS (REFINED)</b>									
<b>VALUE</b>	17.766	2.323	1.641	3.328	1.191	0.942	0.513	0.636	0.121
<b>PROBABILITY</b>	0.0001	0.3129	0.4401	0.6496	0.2335	0.3316	0.4739	0.4250	0.6493
<b>SPATIAL LAG</b>									
<b>VALUE</b>	-	1.6069	-	-	-	-	-	-	-
<b>PROBABILITY</b>	-	0.44778	-	-	-	-	-	-	-
<b>SPATIAL ERROR</b>									
<b>VALUE</b>		1.7266							
<b>PROBABILITY</b>		0.42178							

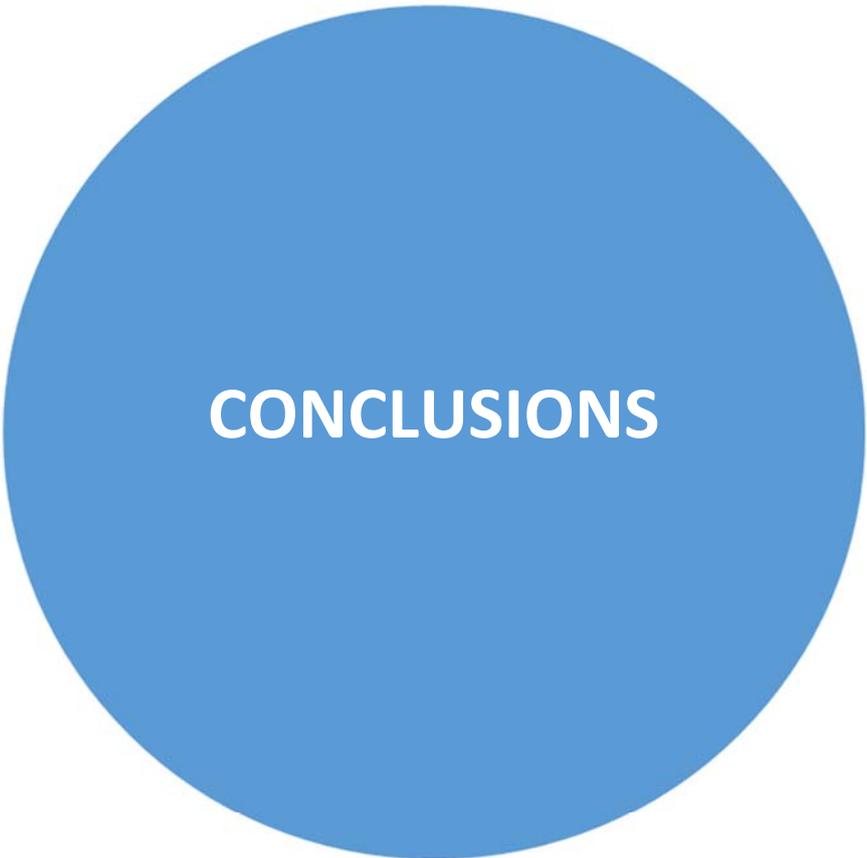
Lastly, the GWR model was developed as an alternative to the three global models. The GWR is a local model, which creates a separate equation for every location, which is calibrated based on the neighboring data points. From the results, it can be observed that the GWR model improves the regression outcomes considerably with adjusted R<sup>2</sup> values almost twice higher than the OLS model (Table 11). Therefore, it can be concluded that the global models, which imply that the relationships between the variables are constant in space, are less suitable for the analysis of the UHI effect. On the other hand, the use of locally varying coefficients improves the overall correlation between the dependent and the independent variables.

In conclusion, the different steps, which aimed at the improvement of the statistical analysis resulted in the considerable improvement of the refined models. Therefore, the regression analysis of the UHI effect requires a custom approach to the problem, where the influence of the different variables has to be explored prior to the modeling phase. In addition, the approach, which was adopted in the current work focuses on the explanation of the UHI phenomenon based on the space and the morphology of the city. This essentially affects the performance of the simple linear regression models, where spatial dependence between the explanatory and the independent variables can be observed. In addition, the exclusion of the autocorrelated sensors locations and multicollinear variables yielded better results in the regression analysis.

Finally, the locally varying GWR model proved to be a better alternative for the analysis of the UHI effect with significantly higher suitability values.

Table 13: Coefficients and significance of the explanatory variables

	<b>SVF (400M)</b>	<b>BD (100M)</b>	<b>LC (200M)</b>	<b>NONPS (300M)</b>	<b>VEG (400M)</b>	<b>VTD (600M)</b>
<b>OLS</b>						
<b>COEFFICIENT</b>	-4.816	-0.012	0.035	2.932	-0.156	-3.121
<b>T-STATISTIC</b>	-1.926	-0.085	0.036	2.126	-0.133	-0.948
<b>P-VALUE</b>	0.0562	0.9326	0.9716	0.0353	0.8947	0.3446
<b>OLS (REFINED)</b>						
<b>COEFFICIENT</b>	-3.934	-	-	2.032	-	-
<b>T-STATISTIC</b>	-2.482	-	-	3.147	-	-
<b>P-VALUE</b>	0.0142	-	-	0.0020	-	-
<b>SPATIAL LAG</b>						
<b>COEFFICIENT</b>	-3.735	-	-	1.889	-	-
<b>T-STATISTIC</b>	-2.325	-	-	2.862	-	-
<b>P-VALUE</b>	0.0200	-	-	0.0042	-	-
<b>SPATIAL ERROR</b>						
<b>COEFFICIENT</b>	-3.974	-	-	2.011	-	-
<b>T-STATISTIC</b>	-2.410	-	-	2.994	-	-
<b>P-VALUE</b>	0.0159	-	-	0.0027	-	-



# CONCLUSIONS



## 6. Conclusions and discussion

This final section of the research gives an overview of the obtained results. Also, the answers to the main research questions are provided along with the conclusions and discussion, which resulted from the exploration process of the work.

### 6.1. Answers to the research questions

#### **How to collect, clean and validate the raw sensor data?**

The data representing the temperatures in the Hague for 2017 was collected by the Netatmo sensor network. The network is built from small weather stations which are deployed with the help of the citizens. The data was collected using the private WI-FI networks of the house owners and it was directly sent to a remote server. The files were transferred using File Transfer Protocol (FTP), which is a standard protocol for establishing client-server communication. The collected data was further organized in a spatial database. This step is essential in the process of analyzing the data due to the great number of files (22 271 separate files for the period April – December). The spatial information is imported as points in the database, defined by the coordinates of the sensors locations. With the use of the spatial database, the sensor data could be further processed and analyzed.

Since the focus of this study is on the extreme temperatures, representing the UHI effect in the city, the extreme measurements were not removed from the dataset. In addition, using the Bokeh python library, the data was visualized. Thus, all systematic errors and empty measurements were identified and removed from the dataset.

Finally, by aggregating all measurements per hour for the different sensors, the noise was removed from the dataset. Thus, the essential information, which the data provides was preserved, while the noise in the dataset was notably decreased.

#### **Which factors or combination of factors influences the UHI effect at most?**

The spatial indicators, which were explored in the current work are the Buildings density, the Land cover index, the Non-permeable surfaces, the Vegetation index, the Sky View Factor and the Vehicle traffic density. Based on the statistical analysis, it was observed that the different indicators have a different area of influence, which is relevant to the Urban Heat Island phenomenon. Therefore, multiple different variations of the aggregations were explored - 100, 200, 300, 400, 600 and 800 m wide buffers, containing the area-averaged values of the indicators.

However, the analysis of the Ordinary least squares regression model indicated that only the Sky View Factor and the Non-permeable surfaces are statistically significant variables. Also, it was observed that

when the statistically insignificant variables are removed, the results from the model improve, which can be explained by the presence of multicollinearity issues between the variables.

Finally, important to mention is that the Sky View Factor and the Non-permeable surfaces are the indicators that influence the night time temperatures at most. The independent variables in the model are defined by the highest difference between the temperatures in the city and the rural areas. The highest variation was observed at 1 am (the 27<sup>th</sup> of May 2017), when the UHI effect was pronounced at most. However, if different temperature patterns were used in the statistical models, the resulting significance of the variables would have been different.

#### **What level of detail is needed to model the UHI variability?**

For the definition of the spatial model of the city, different techniques for space tessellation were explored. The performance of two regular grids (hexagonal and rectangular) was compared to the Voronoi tessellation, which is built based on the locations of the Netatmo sensors. The regular grids were defined with cell sizes of 100 m. As a result, it was observed that generally, the regular grids are more suitable for the representation of the morphological characteristics of the city, because of the smaller cell sizes which provide enough detail. On the other hand, the Voronoi diagram with its varying size provides less detail, which leads to an overall smoothing of the data, where the intraurban spatial variations are not distinguishable.

In addition, in between the two types of regular grids, the hexagonal representation was selected as more suitable, due to its edge reducing effects. Thus, the hexagonal grid provides a detailed representation of the morphological characteristics of the city, where the patterns of the data can be easily observed.

#### **How does the UHI effect variate in the different parts of the city during the day and night?**

The analysis of the temperature data was performed using interactive scatterplots and time-varying animations, which provided a good overview of the heating and cooling processes in the city. Generally, it was observed that the temperatures have higher ranges of variability during the day, compared to the night. Additionally, the same pattern was present for the days where a heat wave is present compared to the regular summer days.

Further, the animations, which were generated with the QGIS Time manager, provided a suitable illustration of the temperature change for the different locations of the sensors. Usually, few different locations close to the city center were characterized by higher temperatures and slower cooling processes.

## **How to accurately model the spatial and time variability of the Canopy Urban Heat Island (CUHI) effect in the city of The Hague?**

In the process of the statistical analysis of all collected data, multiple different variations were explored. The results indicated that the morphological characteristics of the city have different areas of influence, which can explain the UHI effect in the city. Further, in order to create valid models, the data has to be analyzed beforehand. The presence of autocorrelated measurements proved to have a high negative influence on the performance of the models. Additionally, the intra-correlated explanatory variables introduce bias in the model, resulting in low reliability of the regression results.

Further, for the statistical analysis global and local regression models were explored. The Ordinary Least Squares regression is always used in the first stage of the analysis. At this stage, the results from the OLS regression provide information about the data and the most suitable statistical approach. The outcomes from all tests give an indication of the stability of the model regarding the presence of spatial dependence in the data, heteroscedasticity, multicollinearity and autocorrelation.

Finally, in comparison to the OLS regression, the Spatial lag, and Spatial error models, the GWR proved to provide better statistical results. This outcome can be attributed to the local nature of the GWR model, which creates and calibrates a regression equation individually for every data point, which proved to be the more suitable approach for the analysis of the UHI effect in The Hague.

## **6.2. Conclusions**

The Geographically weighted Urban Heat Island modeling research focuses on the exploration of a modeling technique, which is primarily based on spatial data. Multiple reasons guide the choice of such approach. The use of spatial data exclusively will provide relative independence of the model from the availability of current and accurate weather data. Thus, the model could be used as a tool, assisting different planning strategies and the development of different UHI mitigation measurements. The model can provide valuable information about the patterns and the influencing factors on the UHI effect. In this way, different experts can address the UHI problem by evaluating the significance of the different spatial indicators for the temperatures in the city. For example, urban planners can develop their designs using such models as a tool. The parametrization of the urban structure provides the possibilities to analyze its quality. In this way, for example, the best suitable heat relief measurements can be designed for the hottest neighborhoods in the city by comparing different alternatives. Additionally, by using the model and predicting the temperatures in the cities the preparedness for responding to heat waves can be improved. Therefore, the development and improvement of such models are of great importance for minimizing the negative consequences of the climate extremes.

However, there are different types of UHI effects, which are recognized in the literature – the surface and the atmospheric UHI. This research is primarily focused on the Canopy Urban Heat Island, which is defined as the air temperature differences between the urban and rural areas. Therefore, the data that was extensively used in this work is collected from small temperature sensors, deployed in the broader area of the Hague.

The whole research process was based on the exploration of different techniques and solutions of the adopted methodology. Therefore, every step and decision in the process have a direct effect on the final results of the developed models.

Nevertheless, the results from this research proved the complexity and interdependency of the Canopy Urban Heat island phenomenon on different factors. The urban morphology is relevant and it contributes to the development of the higher temperatures in the cities, but additional factors have to be included in the analysis of the extreme ambient temperatures in the urban areas. In addition, since the air has lower heat capacity, the heating and cooling processes have higher frequencies and are less stable. Therefore, meteorological processes like wind and rain have a direct and immediate effect on the air temperatures in the cities. On the other hand, the temperatures of the surfaces are more constant and therefore easily predictable.

Further, the development of such models is highly dependent on the availability of current spatial and meteorological data. In this study, information from more than 200 separate sensors was used. Since the whole research is based on the weather data, its quality and reliability are of great importance. Therefore, more research has to be performed regarding the best suitable locations of the sensors and the quality of the information that they provide.

The spatial information, which was used for the development of this research was openly available vector and satellite data. Thus, the advancement of the open data movement provides opportunities for the development of open and freely accessible spatial and statistical models, which can be used from broader target groups for different purposes. Thus, the overall quality of such studies can be increased, leading to broader positive effects and greater added value.

Finally, the complexity of all spatial processes requires a lot of investigation and exploration of different alternatives. On the other hand, this provides a certain degree of freedom, regarding the data and methodologies, which one can use.

This work was developed as part of the Geomatics programme and follows the principles and the knowledge obtained in this study. The whole process from data collection, storage, analysis, processing and modeling has been covered. Further, different types of spatial and sensor data have been used in this thesis and diverse spatial, geographical and statistical methods have been utilized in

the process of the UHI effect study. Thus, this work is in full compliance with the methodological line of approach of the Master in Geomatics. Additionally, this project has considerable influence on the wider social context due to the importance of the UHI studies for the development of healthy and sustainable environments for the increasing urban populations.

### 6.3. Discussion

The development process of this research revealed different shortcomings of the applied methods, which have to be recognized and incorporated in the analysis of the obtained results.

Generally, the statistical analysis showed a very limited correlation between the chosen spatial indicators and the temperature measurements. These results can be attributed to two main reasons – the more inconstant nature of the air temperature fluctuations and the errors in the sensor measurements. Since the focus of the research is targeted at the Canopy Urban Heat Island, the air temperatures were investigated only. The analysis of the heat variations during the day and the night revealed the unstable variations of the air temperatures. This means that different meteorological processes can cause an immediate change in the observed measurements – for example, when wind appears the air temperatures can drop instantly with few degrees. These characteristics of the air fluctuations have a direct effect on the performance of the models, which are less stable and therefore have limited ability for explaining the variations in the data.

Additionally, the measurements, which are obtained from the Netatmo weather stations contain a significant amount of error. These errors can be caused by the position of the sensors or different technical failures. The location of the sensors has a direct effect on the reliability of the obtained results. Direct sunlight can cause unrealistically high measured temperatures for example. Therefore, the positioning of the weather stations has to be controlled and analyzed during their deployment. For example, a series of different recommendations about the dispose and the maintenance of the sensors can be prepared for the citizens. In this way, the reliability of the data can be drastically improved. Also, the quality of the measurements can be continuously observed, resulting in the exclusion of failing sensors – e.g., sensors that are measuring continuously the same temperatures.

Further, the performance of the statistical analysis can be further improved by exploring different types of models. In the current work, global and local linear models have been used. These models assume that the relationship between the dependent and independent variables has linear character. However, different types of non-linear models can be further incorporated in the analysis of the Urban Heat Island effect in the cities.

Besides the quality of the temperature data, the spatial datasets influence the final results as well. In order to obtain the final spatial indices, every spatial dataset has undergone multiple processing steps, which influence the outcomes. Therefore, all conversion and analysis algorithms have to be used with caution and awareness of the possible shortcomings. For example, the conversion between different data formats usually leads to a loss in the quality or resolution of the data. In addition, all aggregation techniques change essentially the information which the data provides. For example, when the data is aggregated to very big areas, this leads to an overall smoothing of the information, where all extremes are removed from the data. This could be the desired outcome for some methodologies or unnecessary loss of vital information for others. Therefore, every step of the spatial analysis has to be done with the consideration of the desired results and the main aim of the research. Part of the definition of the spatial model is the choice of the space tessellation technique. For this work, three different representations have been compared. Based on the different outcomes, it was observed that both the area and the shape of the grids influence the statistical analysis of the Urban Heat Island effect.

Additionally, the way some of the indicators have been calculated or the data that has been used can be further improved. For example, the Vehicle traffic density index has been simplified in this work to the calculation of the roads density. The indicator could be considerably improved if additional traffic data is included in the calculation. Data about the number of vehicles passing by these roads can be combined with the expected amount of heat released from the cars resulting in more sophisticated indicator measuring the released anthropogenic heat from the cars. Except for the Vehicle traffic density, the calculation method of the vegetation index could be further improved as well. Currently, the indicator is defined by the values from the NDVI index, which correspond to the densely vegetated areas. In this way, small green spots along roads or around buildings are disregarded from the research. Thus, the inclusion of all values from the NDVI index, which correspond to the different land type classes (water, soil and vegetation) could improve the performance of the Vegetation index.

Finally, the complexity of the Urban Heat Island phenomenon revealed the necessity for the inclusion of different indicators in the modeling process. Therefore, it is advised that in the future works broader set of indicators is included in the statistical analysis. In addition, the choice of independent variables has to be driven by the knowledge of the primary requirements of the statistical models. Thus, problems like multicollinearity of the variables can be resolved in the initial stages of the research and such approach can improve the stability and the reliability of the models as well.

The reusability of the methodology of this research is highly dependent on the availability of spatial and temperature data. The datasets that have been used are obtained from the open geospatial portal, developed for the Netherlands. However, such availability and quality of open spatial data are

still limited for many countries and regions. Thus, the results of the implementation of this methodology can be different. On the other hand, the spatial indicators that have been used in this work are widely recognized and have been used in multiple different studies. Additionally, these indicators can be derived from different datasets using diverse methods, which provides certain freedom in the development of the spatial analysis of the territory.

In conclusion, the process of modeling and analyzing the Urban Heats Island effect contains a great amount of exploration of different alternatives. Therefore, there is no single approach towards this problem, but the complexity of factors drives the whole process. Thus, the characteristics of the research methodology are defined by the main goal of the work and the current state of the art in the spatial and statistical domains.

## References:

- Akbari, H., Bell, R., Brazel, T., Cole, D., Estes, M., Heisler, G., ... Zalph, B. (n.d.). Reducing Urban Heat Islands: Compendium of Strategies - Urban Heat Island Basics. Retrieved from <https://www.epa.gov/sites/production/files/2014-06/documents/basicscompendium.pdf>
- Anselin, L. (n.d.-a). Exploring Spatial Data with GeoDa TM : A Workbook. Retrieved from <http://sal.agecon.uiuc.edu/>
- Anselin, L. (n.d.-b). Local Indicators of Spatial Association-LISA. Retrieved from [http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4\\_Anselin1995.pdf](http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4_Anselin1995.pdf)
- Asaeda, T., & Ca, V. (2000). Characteristics of permeable pavement during hot summer weather and impact on the thermal environment. *Building and Environment*, (35), 363–375. Retrieved from [https://ac.els-cdn.com/S0360132399000207/1-s2.0-S0360132399000207-main.pdf?\\_tid=d5228077-6608-4b2b-9b4d-ae88b9e15a96&acdnat=1525770060\\_89e8aaafb3b71658bf5767143d3967e3](https://ac.els-cdn.com/S0360132399000207/1-s2.0-S0360132399000207-main.pdf?_tid=d5228077-6608-4b2b-9b4d-ae88b9e15a96&acdnat=1525770060_89e8aaafb3b71658bf5767143d3967e3)
- Aurenhammer, F. (n.d.). Voronoi Diagrams — A Survey of a Fundamental Geometric Data Structure. Retrieved from [http://delivery.acm.org/10.1145/120000/116880/p345-aurenhammer.pdf?ip=145.94.172.99&id=116880&acc=ACTIVE-SERVICE&key=0C390721DC3021FF.512956D6C5F075DE.4D4702B0C3E38B35.4D4702B0C3E38B35&\\_\\_acm\\_\\_=1525271111\\_d5899931ea589ccc0c97ee893c7a99a1](http://delivery.acm.org/10.1145/120000/116880/p345-aurenhammer.pdf?ip=145.94.172.99&id=116880&acc=ACTIVE-SERVICE&key=0C390721DC3021FF.512956D6C5F075DE.4D4702B0C3E38B35.4D4702B0C3E38B35&__acm__=1525271111_d5899931ea589ccc0c97ee893c7a99a1)
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (2010). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- Fotheringham, A. S. (2009). “The Problem of Spatial Autocorrelation” and Local Spatial Statistics. *Geographical Analysis*, 41(4), 398–403. <https://doi.org/10.1111/j.1538-4632.2009.00767.x>
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression : the analysis of spatially varying relationships*. Wiley. Retrieved from <https://www.wiley.com/en-us/Geographically+Weighted+Regression%3A+The+Analysis+of+Spatially+Varying+Relationships+-p-9780471496168>
- Gaetani, I. (2013). The main components of the urban atmosphere. Retrieved December 10, 2017, from [https://www.researchgate.net/figure/262068916\\_fig3\\_Figure-23-The-main-components-of-the-urban-atmosphere](https://www.researchgate.net/figure/262068916_fig3_Figure-23-The-main-components-of-the-urban-atmosphere)
- Gago, E. J., Roldan, J., Pacheco-Torres, R., & Ordóñez, J. (2013). The city and urban heat islands: A

- review of strategies to mitigate adverse effects. *Renewable and Sustainable Energy Reviews*, 25, 749–758. <https://doi.org/10.1016/j.rser.2013.05.057>
- Global Spatial Autocorrelation (1). (n.d.). Retrieved May 15, 2018, from [https://geodacenter.github.io/workbook/5a\\_global\\_auto/lab5a.html#morans-i](https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html#morans-i)
- Gomez, R. F. A. de S. A. (2011). Satellite and ground-based sensors for Urban Heat Island analysis in the city of Madrid. <https://doi.org/10.1109/JURSE.2011.5764791>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.
- Hämmerle, M., Gál, T., Unger, J., & Matzarakis, A. (2014). DIFFERENT ASPECTS IN THE QUANTIFICATION OF THE SKY VIEW FACTOR IN COMPLEX ENVIRONMENTS, 47–48. Retrieved from <http://www2.sci.u-szeged.hu/eghajlattan/akta14/2014-Acta-Clim-chor-47-48-Hammerle-et-al.pdf>
- Hardy, C. H., & Nel, A. L. (2015). Data and techniques for studying the urban heat island effect in Johannesburg. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W3(May), 203–206. <https://doi.org/10.5194/isprsarchives-XL-7-W3-203-2015>
- Herring, J. W. and D. (2000, August 30). Measuring Vegetation (NDVI & EVI). Retrieved May 9, 2018, from [https://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring\\_vegetation\\_2.php](https://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_2.php)
- Hoeven, F. van der, & Wandl, A. (2018). Haagse Hitte. *BK BOOKS*. Retrieved from <https://books.bk.tudelft.nl/index.php/press/catalog/view/haagsehitte/705/546-1>
- Hoeven, F. Van Der, & Wandl, A. (n.d.). *Hotterdam*.
- Huber, D., & Gullede, J. (2011). EXTREME WEATHER & CLIMATE CHANGE: UNDERSTANDING THE LINK AND MANAGING THE RISK. Retrieved from <https://www.c2es.org/site/assets/uploads/2011/12/white-paper-extreme-weather-climate-change-understanding-link-managing-risk.pdf>
- Kokalij, Z., Zaksek, K., & Ostir, K. (2011). Application of Sky-View Factor for the Visualization of Historic Landscape Features in Lidar-Derived Relief Models. *Antiquity*, 85(327), 263–273.
- Lee, Y., Li, Z., Li, Y. (2000). Taxonomy of space. *Journal of Photogrammetry & Remote Sensing*, (55), 139–149. Retrieved from <https://ac.els-cdn.com/S0924271600000150/1-s2.0->

S0924271600000150-main.pdf?\_tid=52f0d9bf-a3f5-452d-ab31-c37824f0a43c&acdnat=1525260823\_f337af3dc61d1d74c1bb0706ce2b8b81

Luan, Q., Ye, C., Liu, Y., & Li, S. (2014). Study of Effect of Urban Green Land on Thermal Environment of Surrounding Buildings : A Case Study in Beijing , China, 299–314.

Memon, R. A., Leung, D. Y. C., & Chunho, L. I. U. (2008). A review on the generation, determination and mitigation of UHI.pdf, 20, 120–128.

Mirzaei, P. A. (2015). Recent challenges in modeling of urban heat island. *Sustainable Cities and Society*, 19, 200–206. <https://doi.org/10.1016/j.scs.2015.04.001>

Moran's I: Definition, Examples. (n.d.). Retrieved May 15, 2018, from <http://www.statisticshowto.com/morans-i/>

Nunez, M., & Oke, T. R. (1980). Modeling the Daytime Urban Surface Energy Balance. Retrieved from <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1538-4632.1980.tb00043.x>

O'Malley, C., Piroozfarb, P. A. E., Farr, E. R. P., & Gates, J. (2014). An investigation into minimizing urban heat island (UHI) effects: A UK perspective. *Energy Procedia*, 62(0), 72–80. <https://doi.org/10.1016/j.egypro.2014.12.368>

Santamouris, M., Synnefa, A., & Karlessi, T. (2011). Using advanced cool materials in the urban built environment to mitigate heat islands and improve thermal comfort conditions. *Solar Energy*, 85(12), 3085–3102. <https://doi.org/10.1016/j.solener.2010.12.023>

Sentinel-2: The operational Copernicus optical high resolution land mission. (n.d.). Retrieved from [http://esamultimedia.esa.int/docs/S2-Data\\_Sheet.pdf](http://esamultimedia.esa.int/docs/S2-Data_Sheet.pdf)

Shoff, C. (n.d.). Exploratory Spatial Data Analysis. Retrieved May 15, 2018, from <http://gispopsci.org/exploratory-spatial-data-analysis/>

South, S., Working, B., & October, P. (2008). Urban Heat Island Mitigation Can Improve New York City's Environment: Research on the Impacts of Mitigation Strategies, (October).

Stanganelli, Marialuce, and Soravia, M. (2012). Connections Between Urban Structure and Urban Heat Island Generation: An Analysis Through Remote Sensing and GIS. *Computational Science and Its Applications - ICCSA 2012*, 7334, 599–608. <https://doi.org/10.1007/978-3-642-31075-1>

Svensson, M. K. (2004). Sky view factor analysis – implications for urban air temperature differences. *Meteorological Applications*, 11(3), 201–211. <https://doi.org/10.1017/S1350482704001288>

Szymanowski, M., & Kryza, M. (2009). GIS-based techniques for urban heat island spatialization.

*Climate Research*, 38(2), 171–187. <https://doi.org/10.3354/cr00780>

The problem of multicollinearity. (1997). In *Understanding Regression Analysis* (pp. 176–180). Boston, MA: Springer US. [https://doi.org/10.1007/978-0-585-25657-3\\_37](https://doi.org/10.1007/978-0-585-25657-3_37)

Unger, J. (2009). Connection between urban heat island and sky view factor approximated by a software tool on a 3D urban database. *Int. J. Environment and Pollution*, 363(12). Retrieved from <http://www2.sci.u-szeged.hu/eghajlattan/ungercikk/2009-IJEP-36-Unger.pdf>

Urban heat island.svg - Wikimedia Commons. (n.d.). Retrieved December 8, 2017, from [https://commons.wikimedia.org/wiki/File:Urban\\_heat\\_island.svg](https://commons.wikimedia.org/wiki/File:Urban_heat_island.svg)

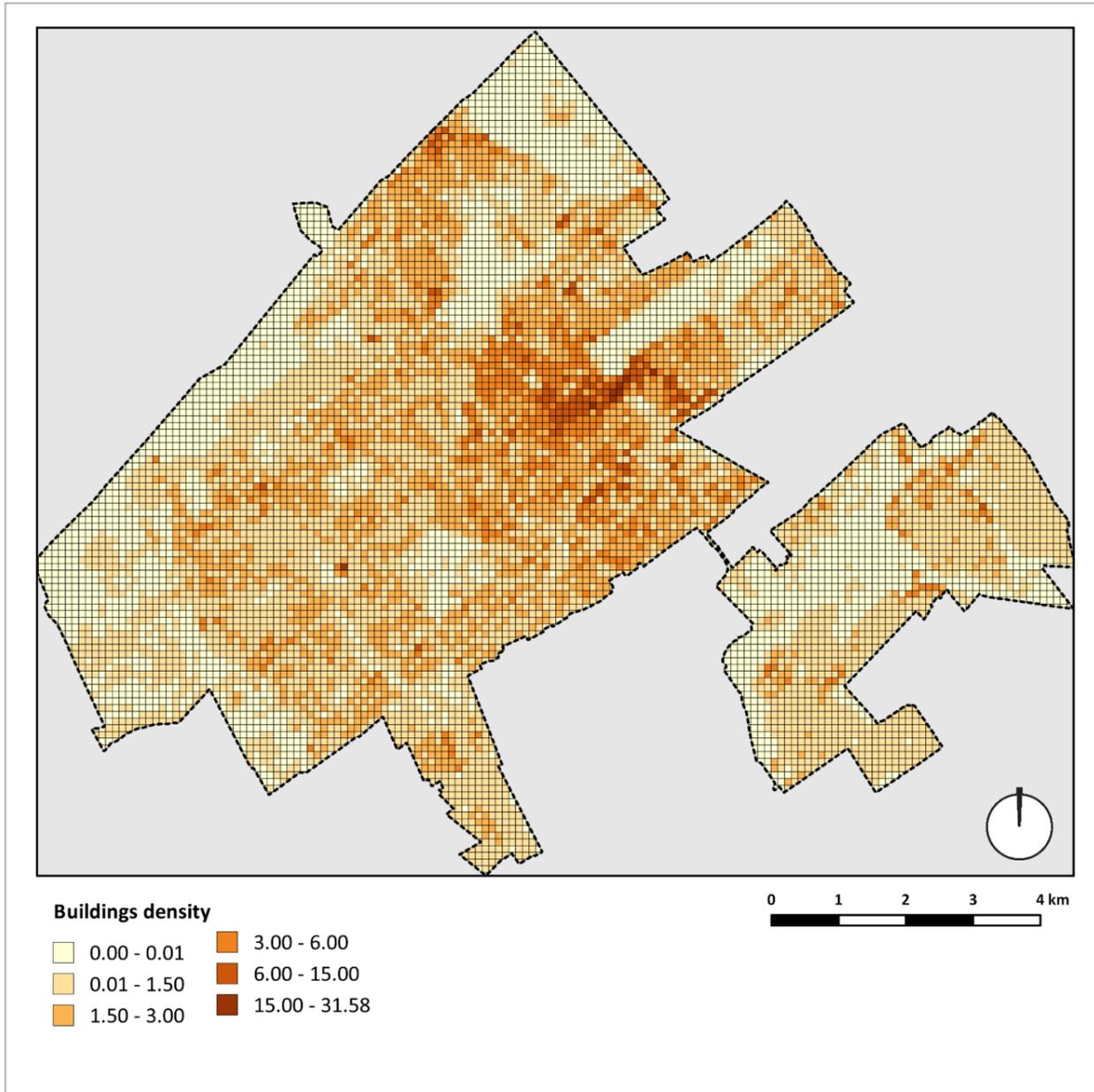
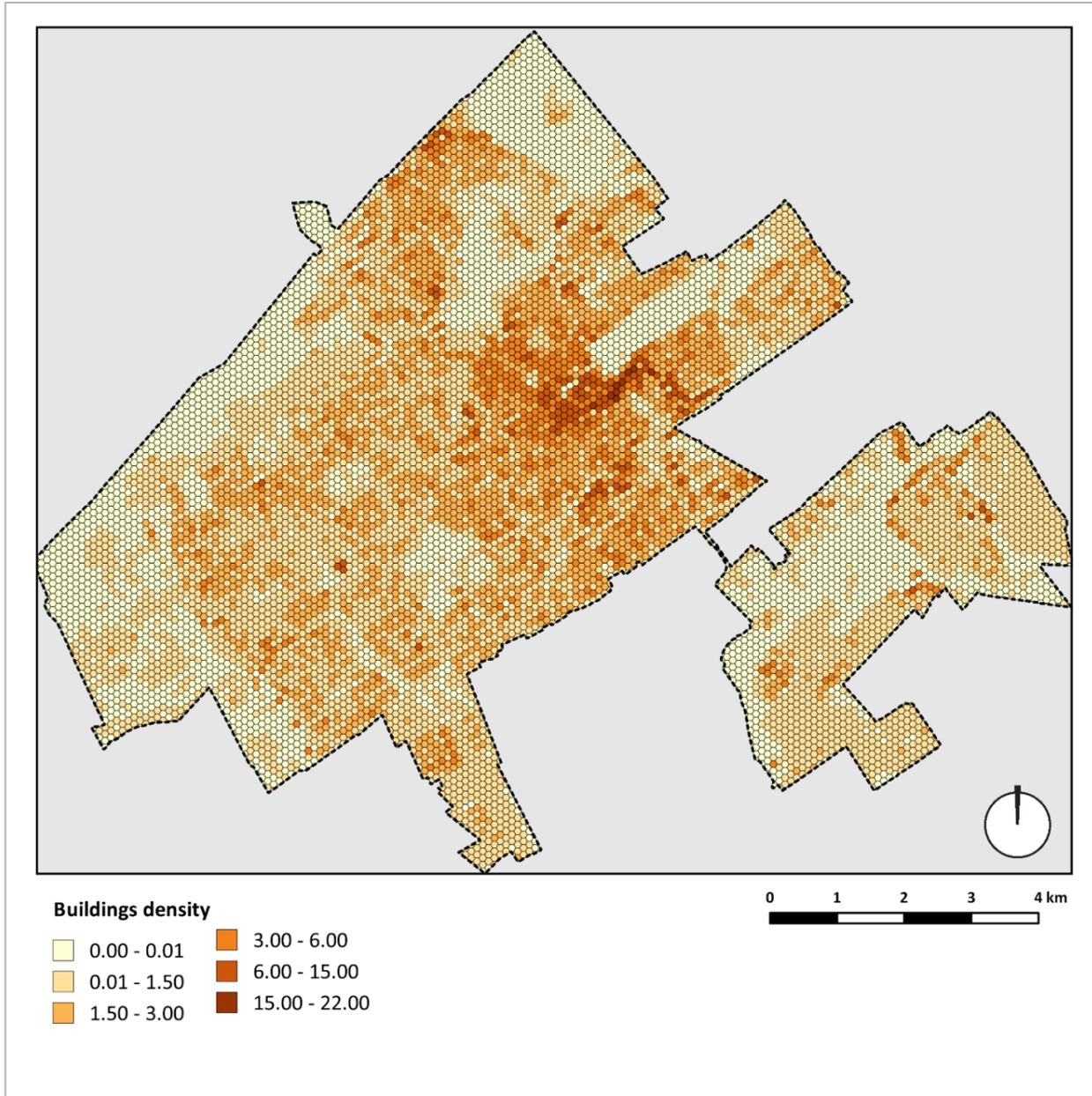
van Hove, L. W. A., Jacobs, C. M. J., Heusinkveld, B. G., Elbers, J. A., Van Driel, B. L., & Holtslag, A. A. M. (2015). Temporal and spatial variability of urban heat island and thermal comfort within the Rotterdam agglomeration. *Building and Environment*, 83, 91–103. <https://doi.org/10.1016/j.buildenv.2014.08.029>

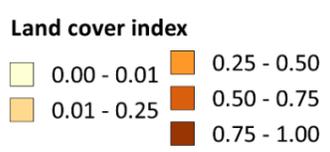
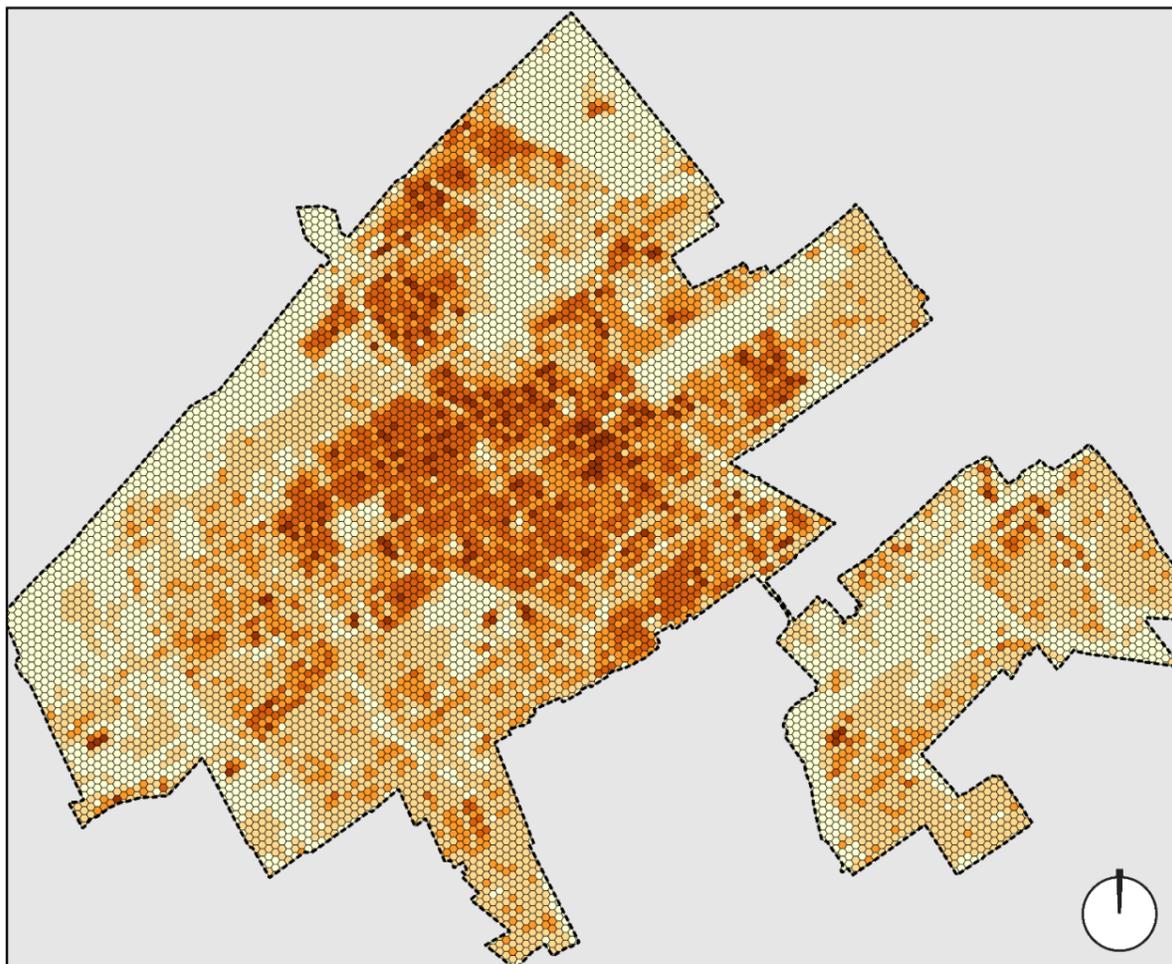
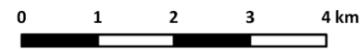
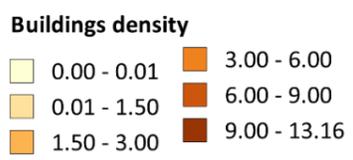
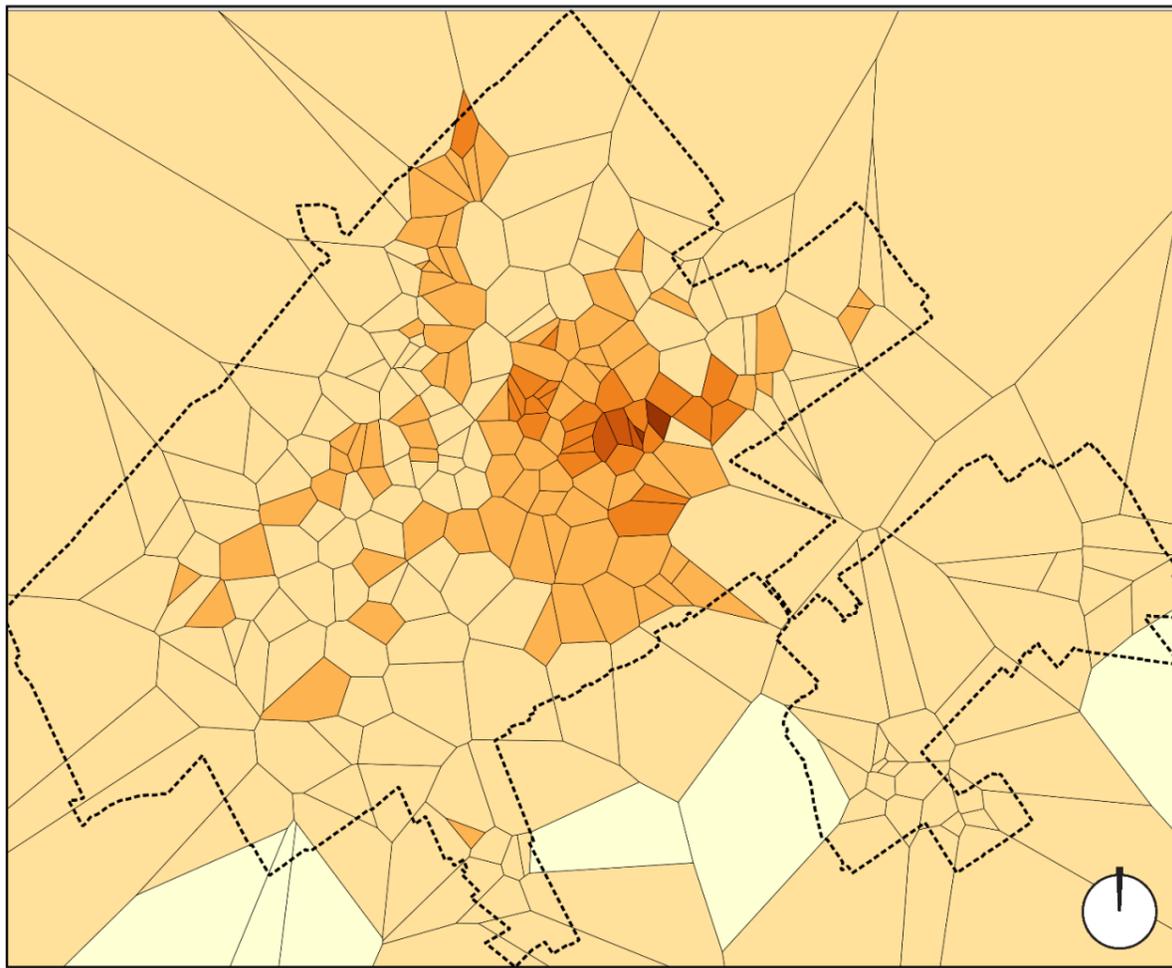
Wijerathne, S. N., & Halwatura, R. U. (n.d.). The Impact of Green Roofs on Urban Heat Island Effect. Retrieved from <https://pdfs.semanticscholar.org/6b73/06595cc44a158a8e0951e82d8cc8d8f44d2b.pdf>

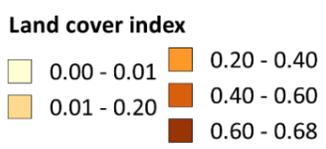
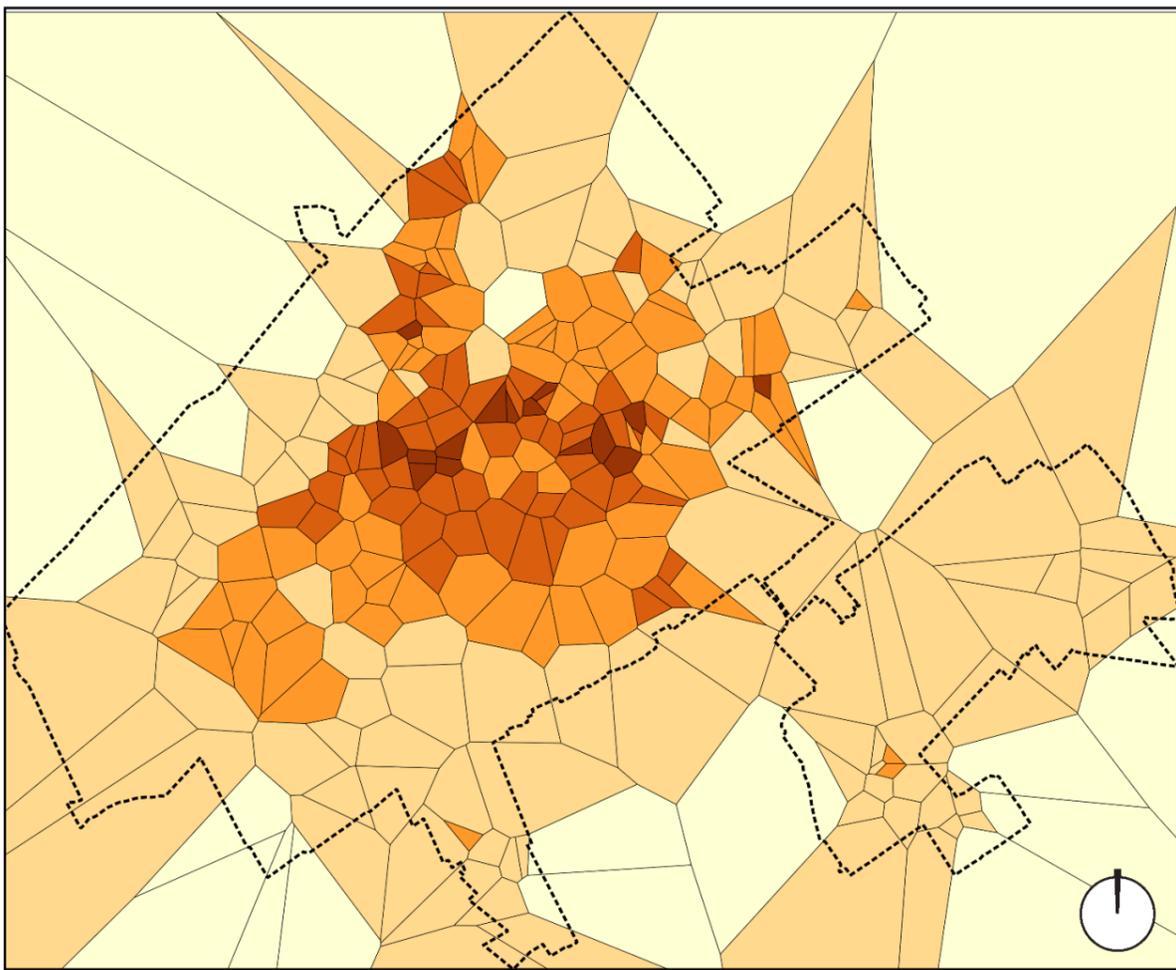
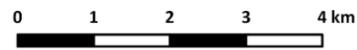
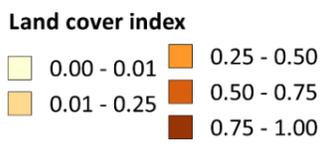
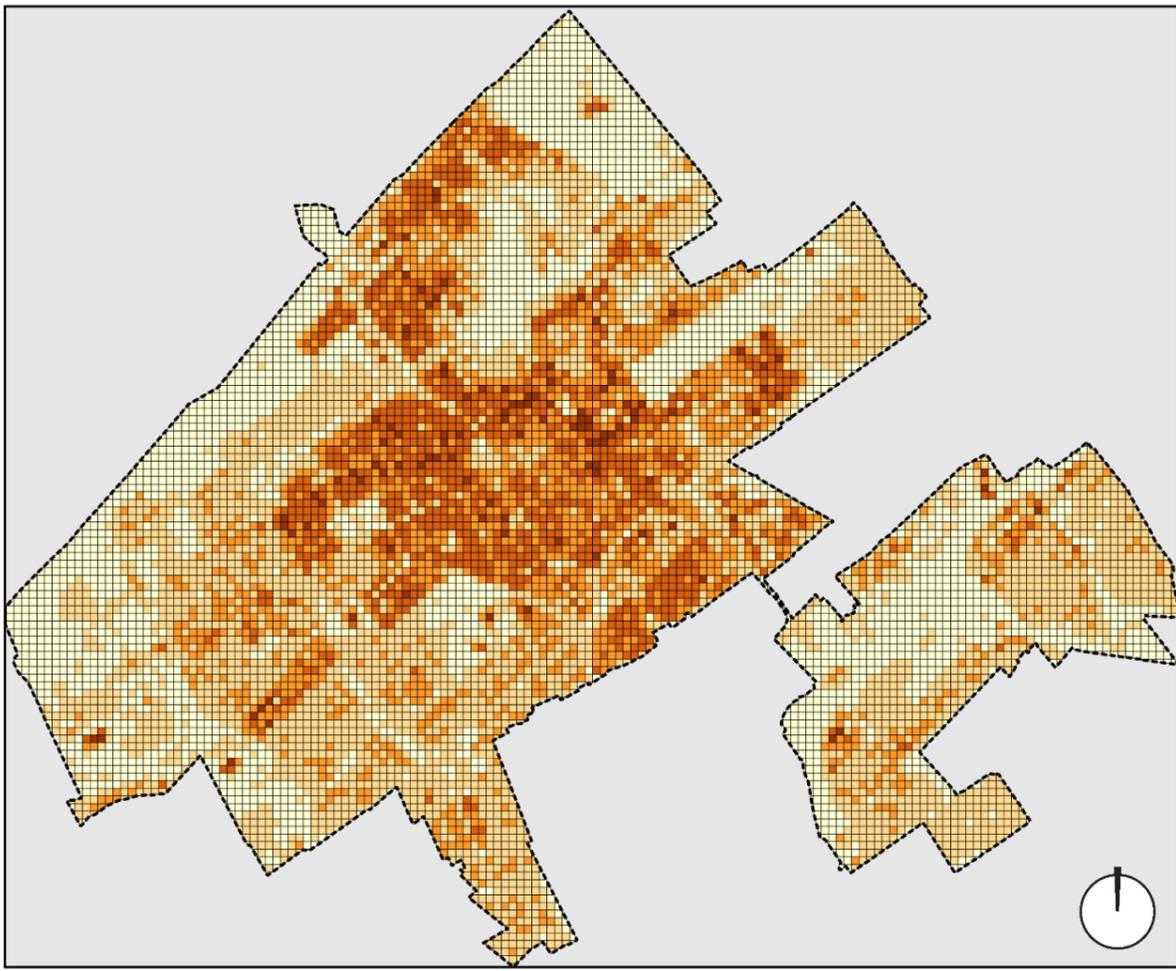
Wooldridge, J. (2013). *Full-Text* (5th ed.). Retrieved from [http://economics.ut.ac.ir/documents/3030266/14100645/Jeffrey\\_M.\\_Wooldridge\\_Introductory\\_Econometrics\\_A\\_Modern\\_Approach\\_\\_2012.pdf](http://economics.ut.ac.ir/documents/3030266/14100645/Jeffrey_M._Wooldridge_Introductory_Econometrics_A_Modern_Approach__2012.pdf)

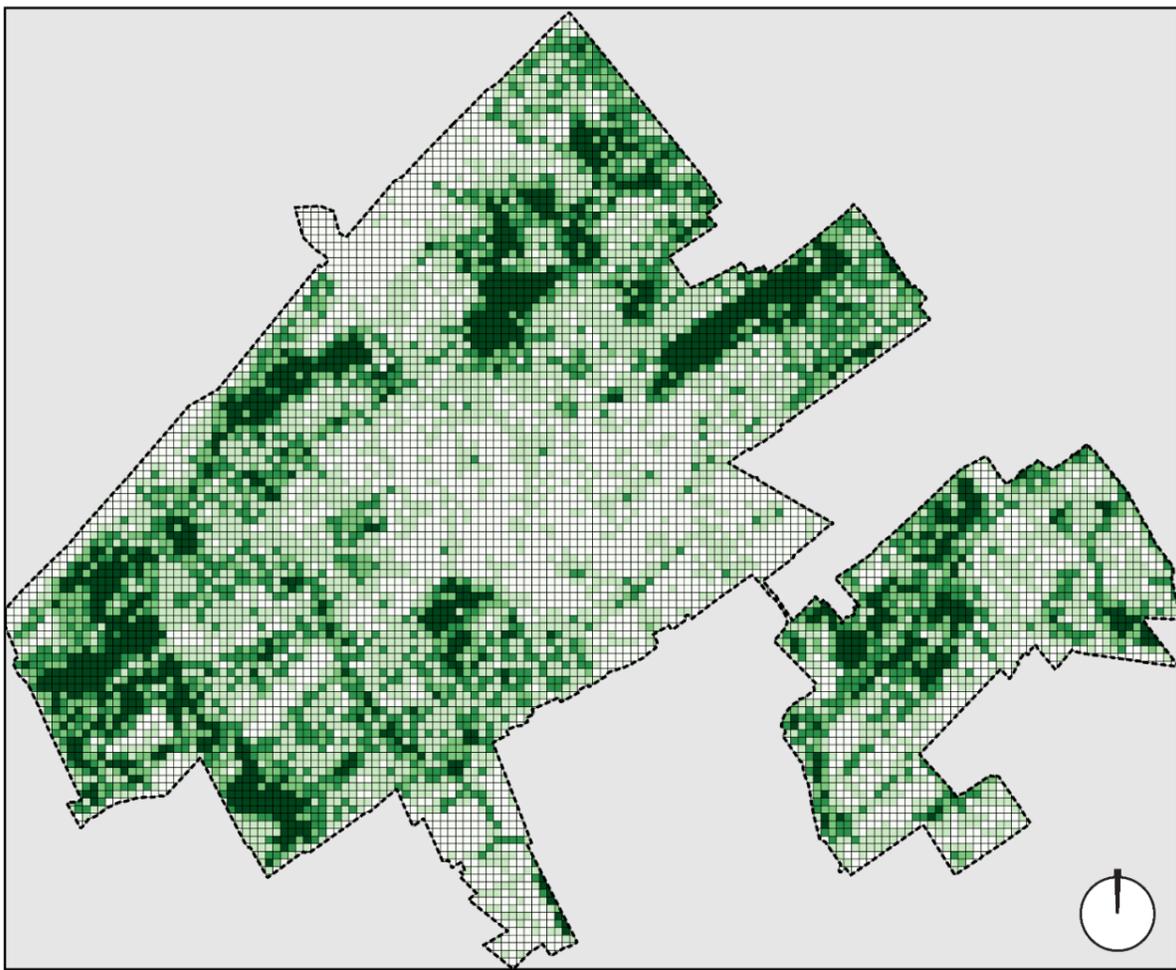
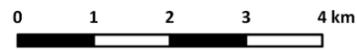
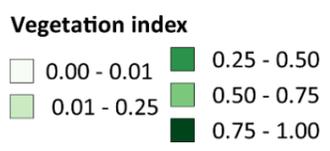
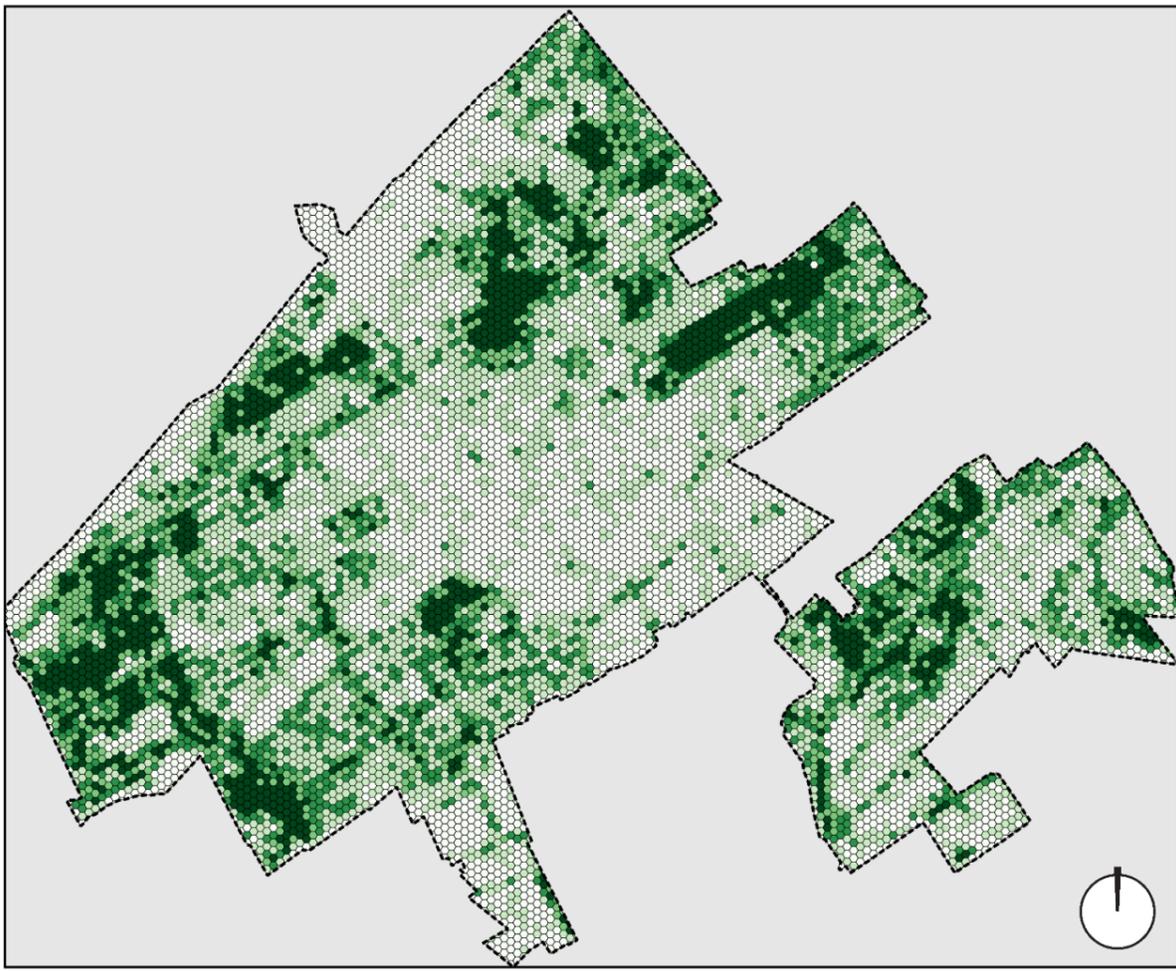
Yang, F., & Chen, L. (2016). Developing a thermal atlas for climate-responsive urban design based on empirical modeling and urban morphological analysis. *Energy and Buildings*, 111, 120–130. <https://doi.org/10.1016/j.enbuild.2015.11.047>

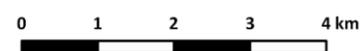
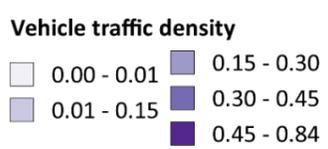
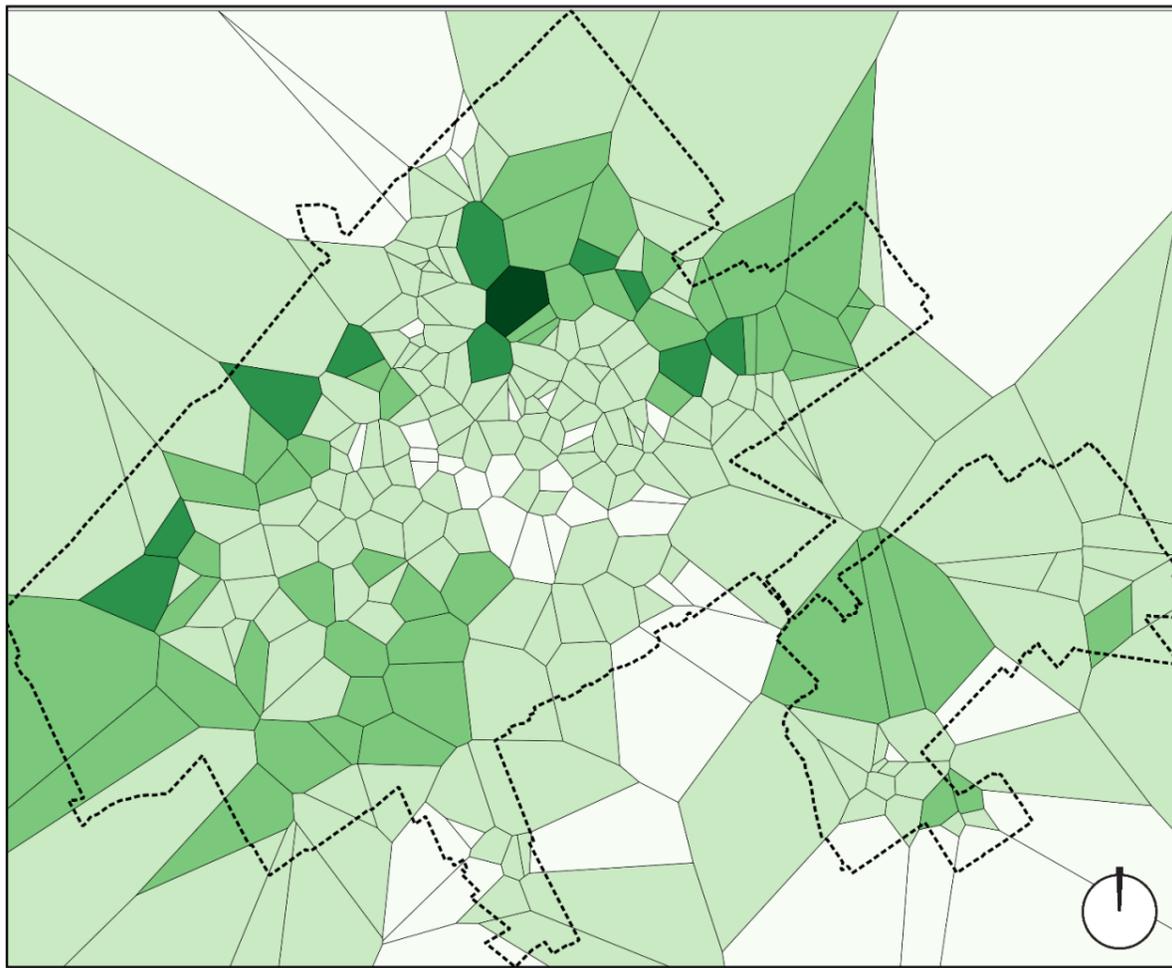
Appendix 1: Maps of calculated spatial indicators

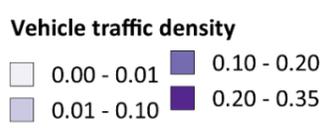
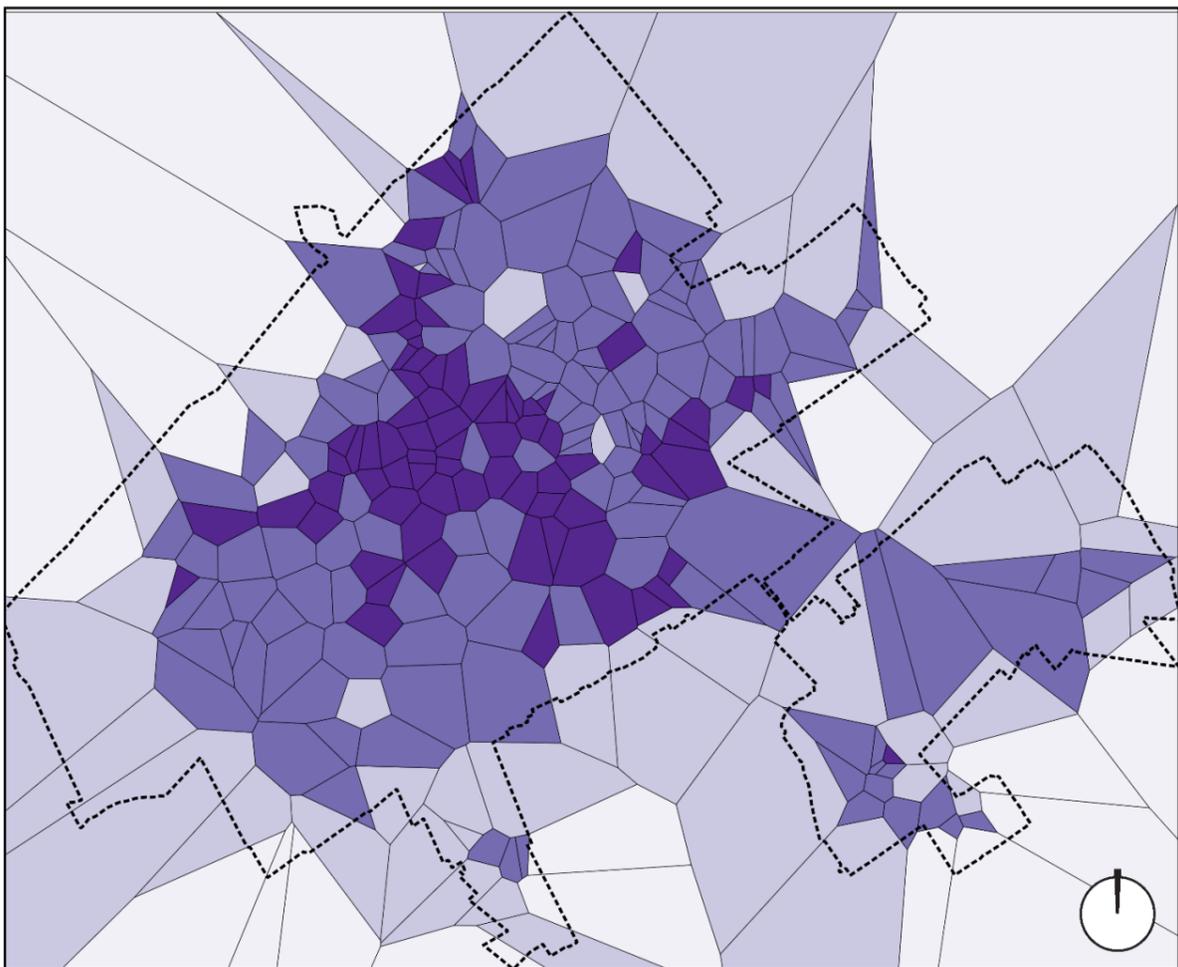
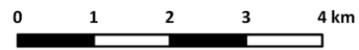
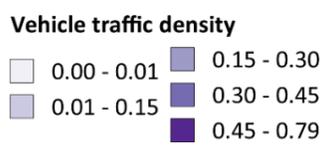
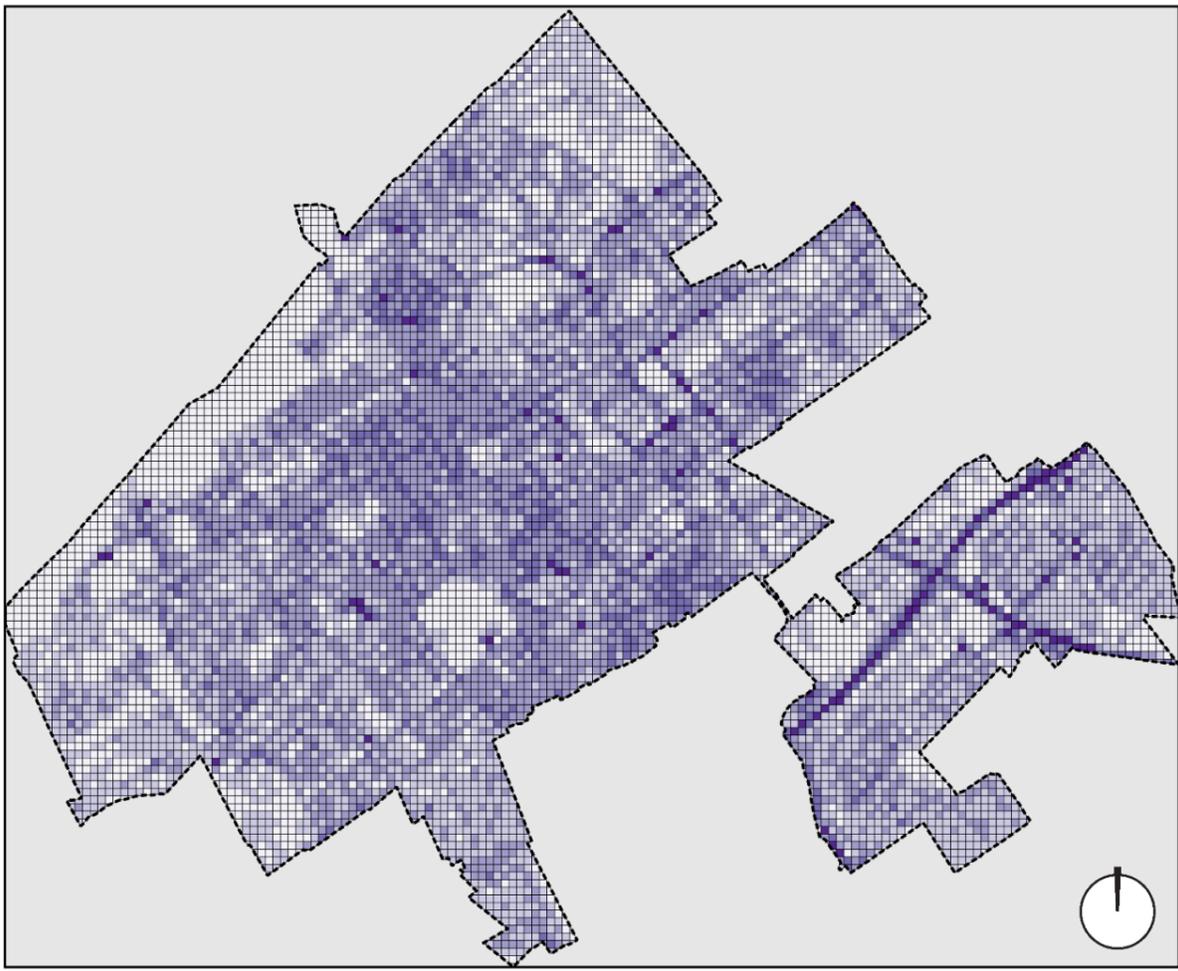


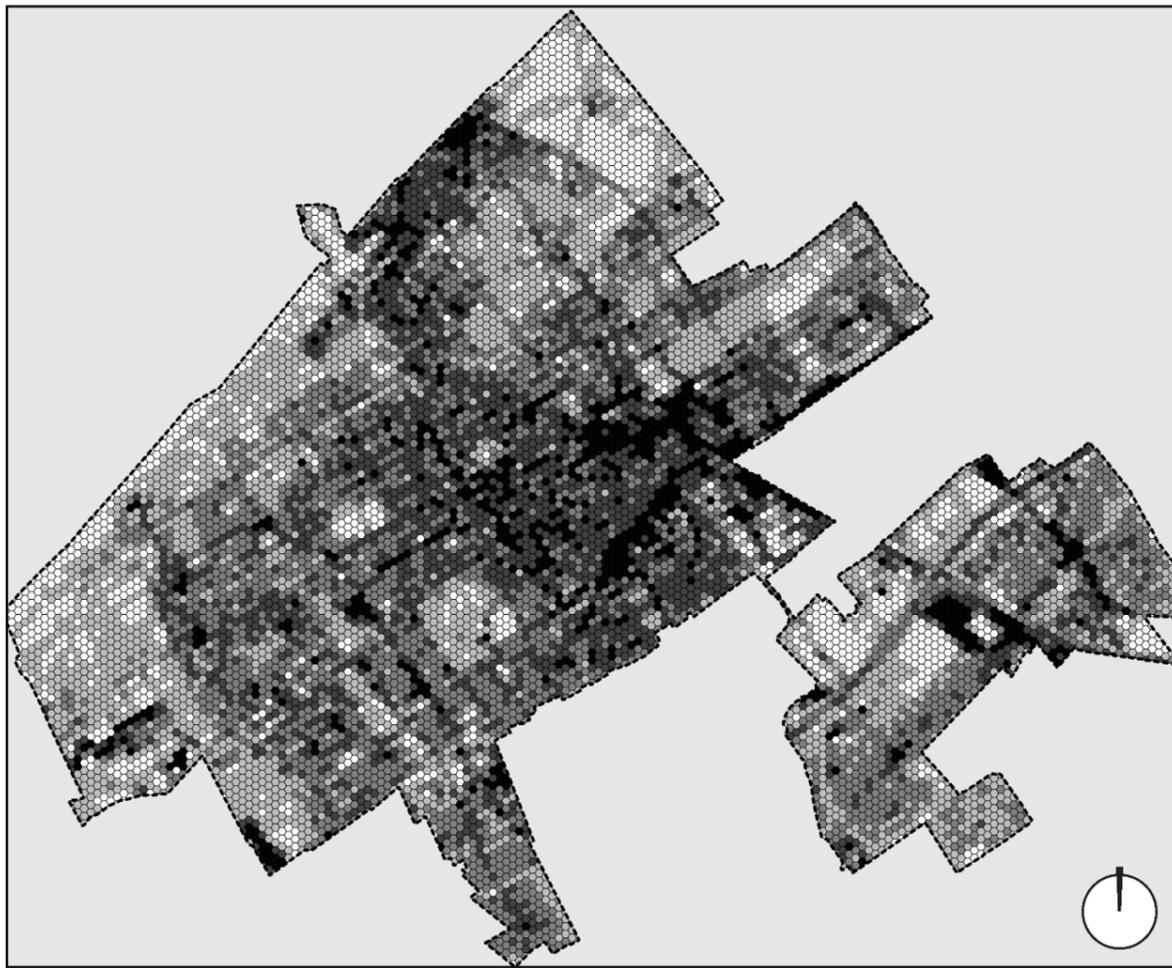




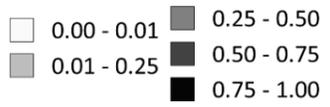




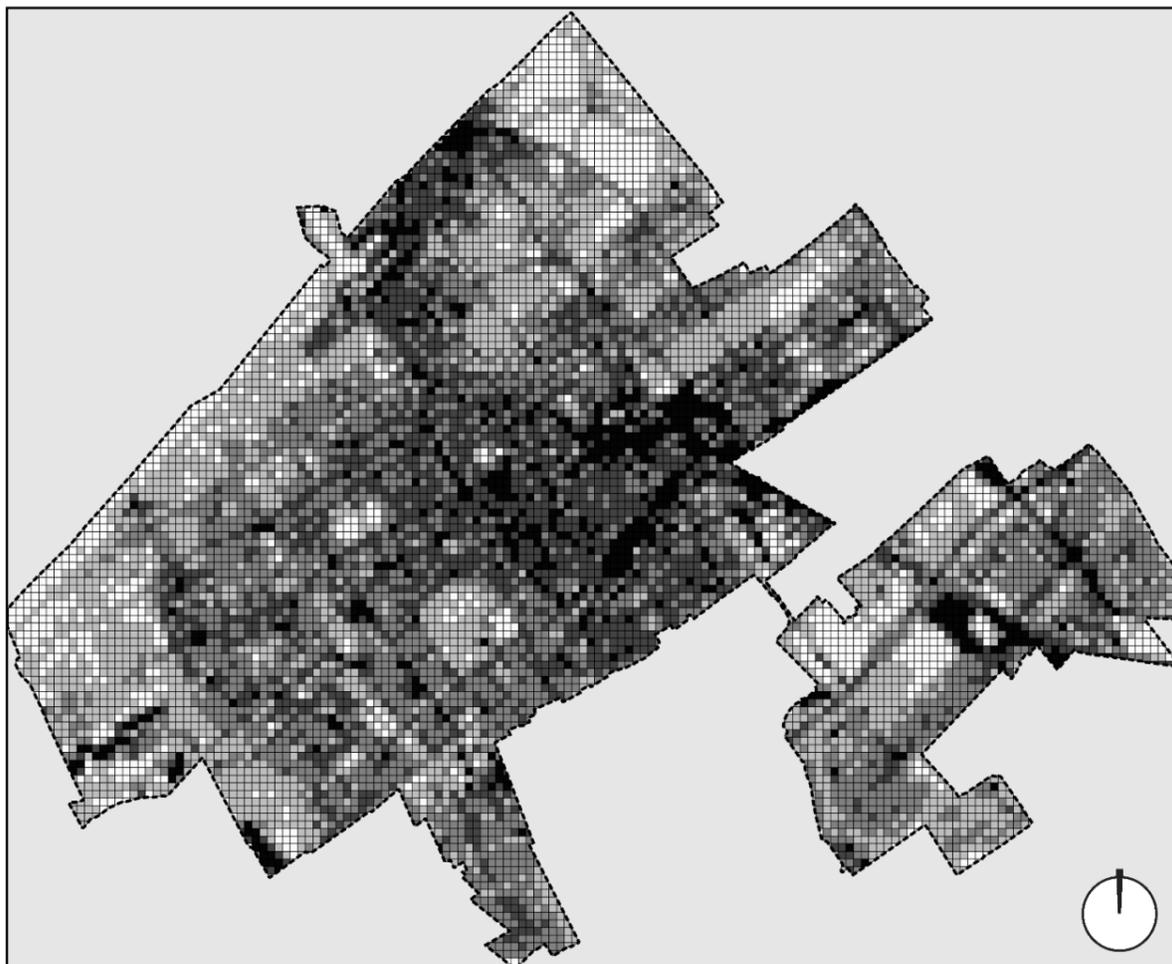




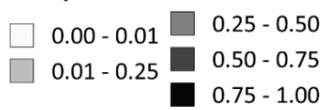
**Non-permeable surfaces index**



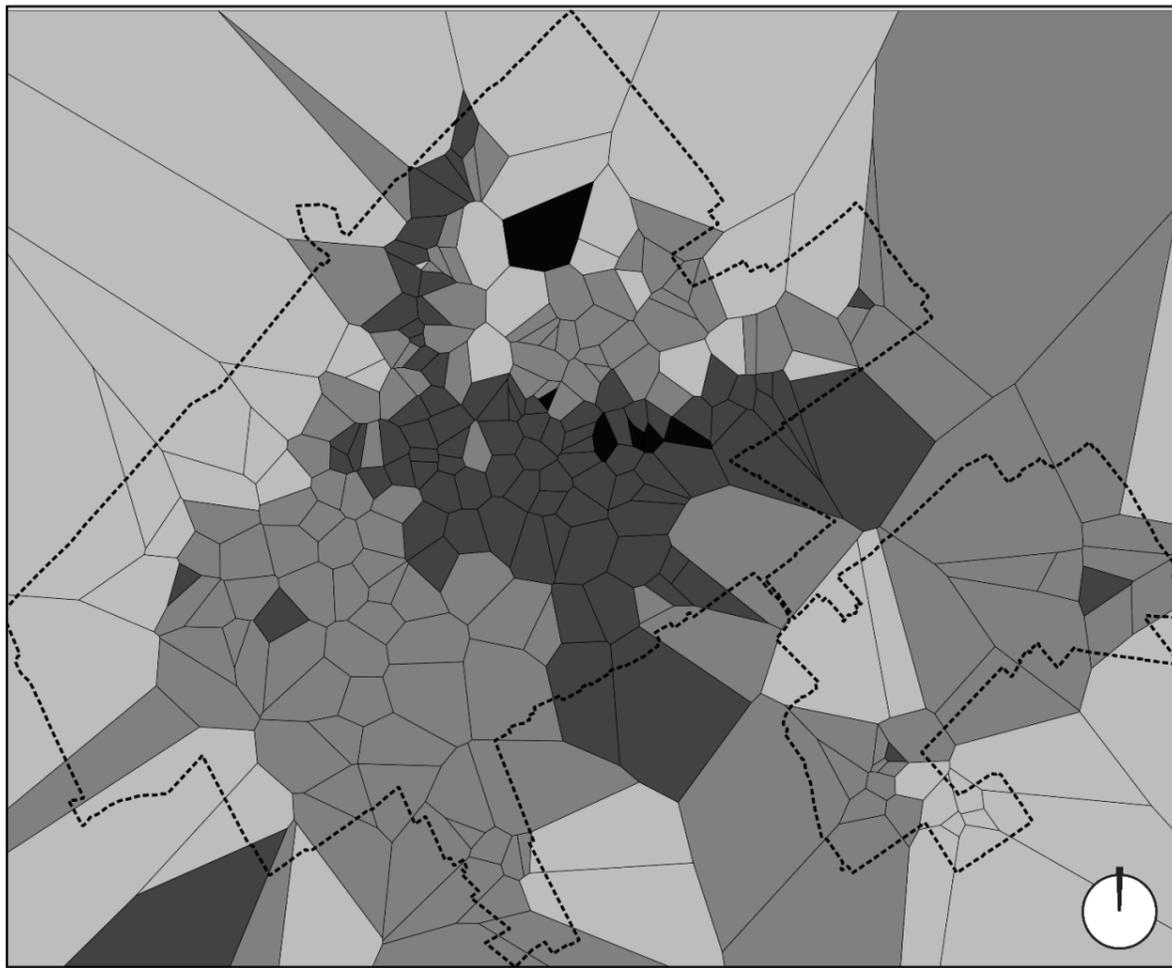
0 1 2 3 4 km



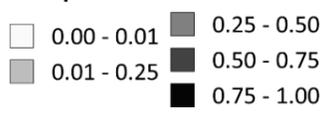
**Non-permeable surfaces index**



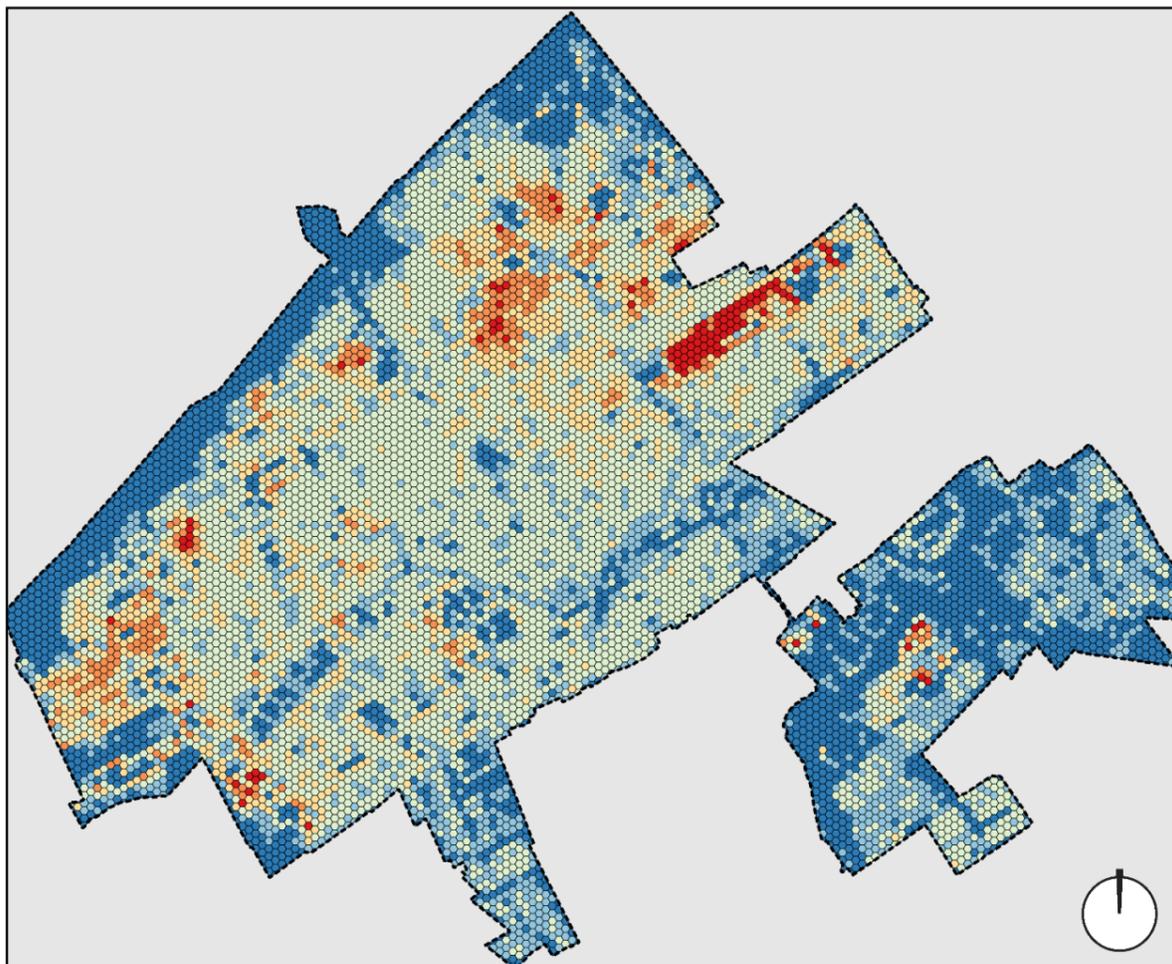
0 1 2 3 4 km



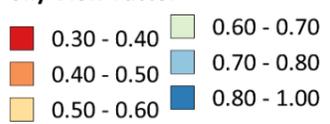
**Non-permeable surfaces index**



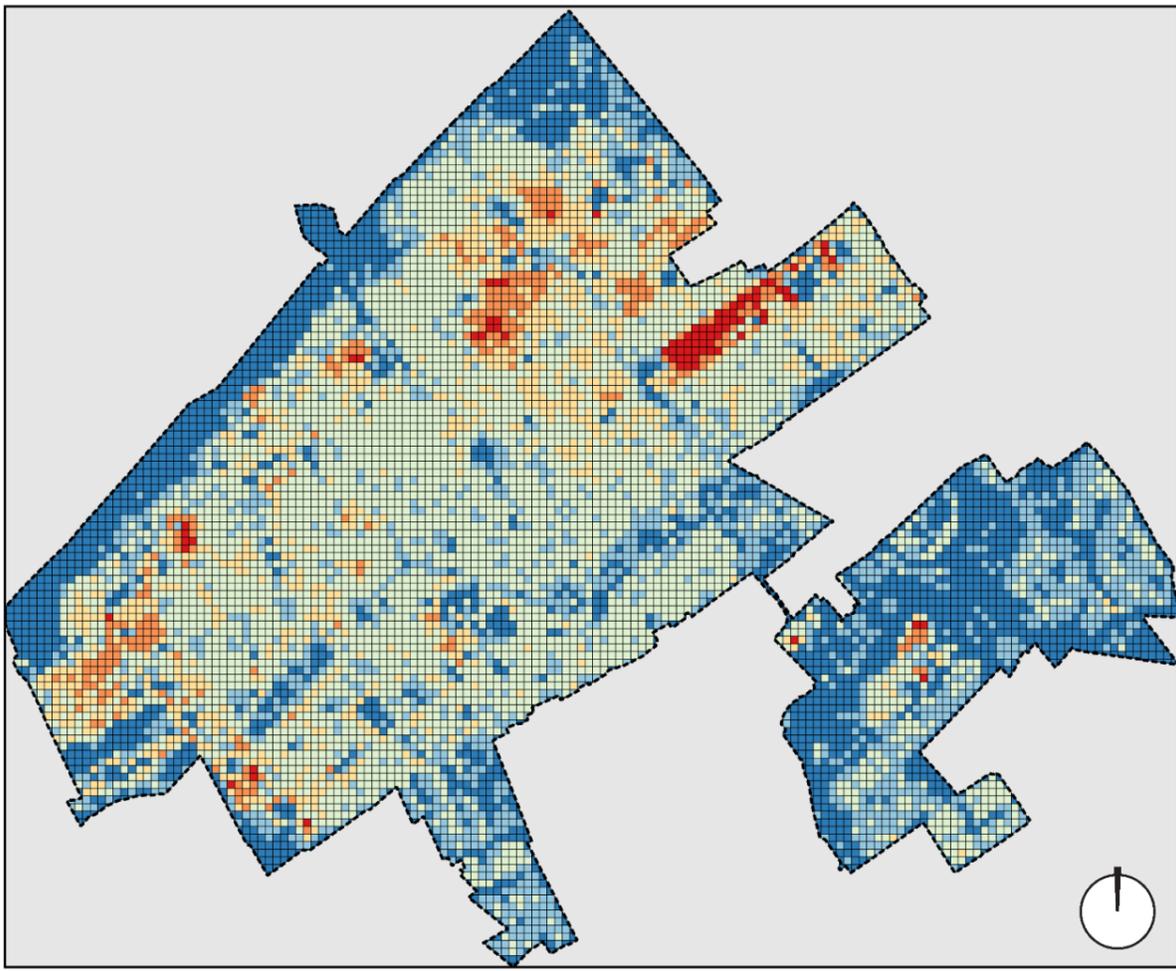
0 1 2 3 4 km



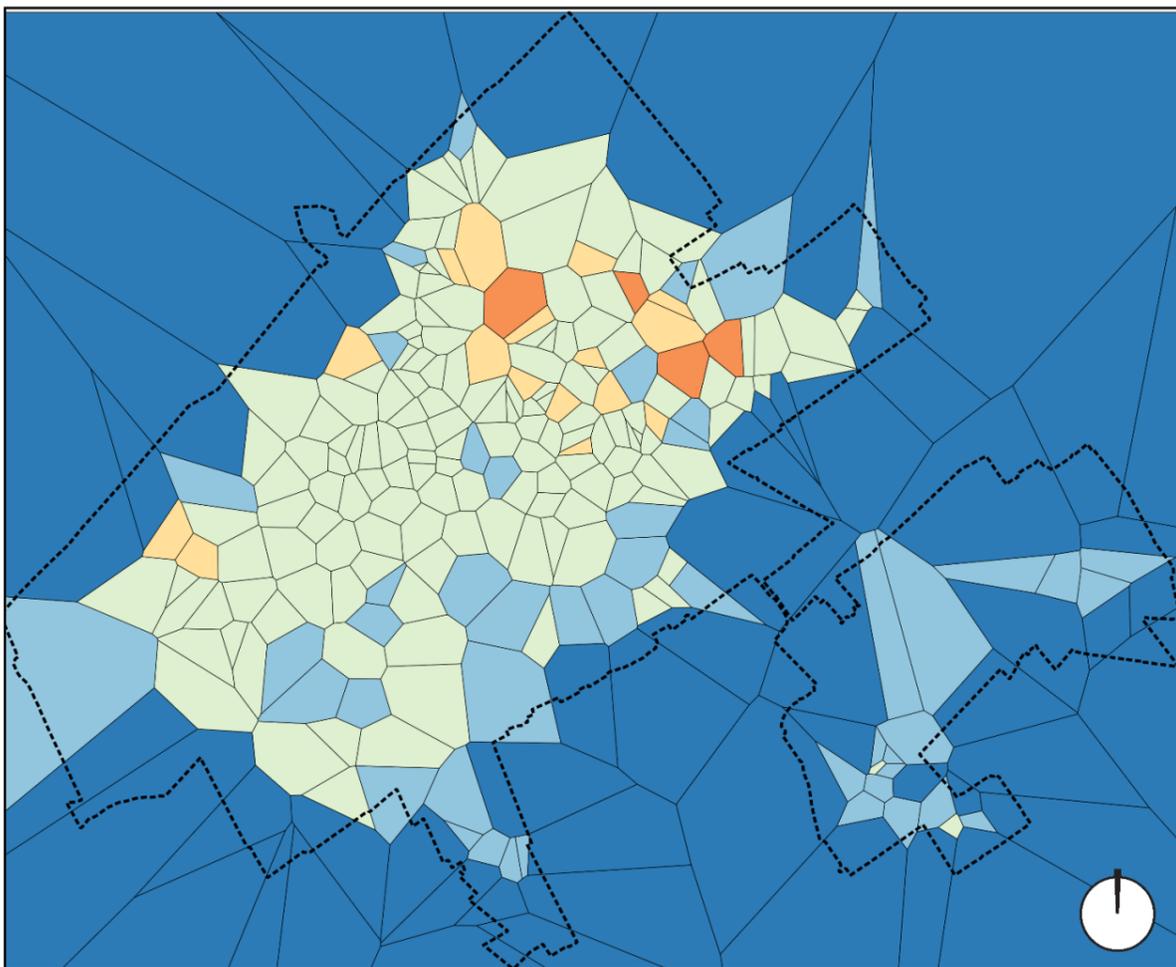
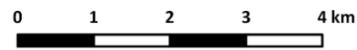
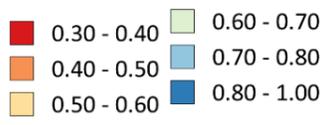
**Sky View Factor**



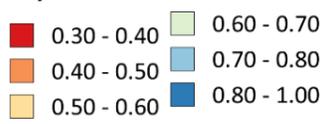
0 1 2 3 4 km



**Sky View Factor**



**Sky View Factor**



## Appendix 2: Database creation

```
import os
import psycopg2
import datetime

directory = "D:\MASTER THESIS\Data\sensor measurements"

def connect_db():
    try:
        global conn
        # creating db connection object
        conn = psycopg2.connect("host=localhost password=heat
dbname=Temp_data user=postgres")
        global cur
        cur = conn.cursor() #used to execute commands
        print "Connected."
    except (Exception, psycopg2.DatabaseError) as error:
        print (error)

def create_table():
    cur.execute(""" Drop table IF EXISTS Temp_data3""")
    cur.execute("""
        CREATE TABLE Temp_data3(
            measure_id float NOT NULL PRIMARY KEY,
            Sensor geometry,
            Date date,
            Time time,
            MAC text,
            Altitude float,
            Temperature float,
            Humidity float,
            Rain_60min float,
            Rain_24h float,
            Pressure float)
        """)
    print "Table created."
    conn.commit()

def insert_to_db():
    measure_id = 0
    inserted = 0
    failed = 0
    for filename in os.listdir(directory):
        if filename.endswith(".txt"):

            with open(os.path.join(directory, filename)) as f:
                content = f.readlines()[1:]
                id = 0
                for line in content:
                    row = line.strip().split(",")
                    coordinates = 'POINT(%s %s)' % (row[1], row[2])

                    # Specify time
                    time = filename.split(" ")
                    day = datetime.datetime.strptime(time[0], '%Y-%m-
%d').date()

                    time = datetime.datetime.strptime(time[1],
'%H_%M_%S').time()

                    row = ['NULL' if v is "" else v for v in row]
                    id+=1
```

```

        measure_id+=1
        strsql = "INSERT INTO Temp_data3(measure_id,
Sensor,Date, Time, MAC, Altitude, Temperature,Humidity,Rain_60min,Rain_24h,
Pressure) VALUES ({} , ST_GeomFromText('{}', {}), '{}',
'{}','{}',{},{},{},{},{},{},{},{},{})".format(measure_id, coordinates, 4326,
day,time, row[0], row[3], row[4], row[5], row[6], row[7],row[9])
        try:
            cur.execute(strsql)
            conn.commit()
            inserted+=1
        except psycopg2.Error as e:
            print('Insert failed: '+ str(e))
            failed+=1
            print strsql

    print filename
    print 'Number of rows inserted {}'.format(inserted)
    print 'Number of rows failed {}'.format(failed)

if __name__ == '__main__':
    connect_db()
    create_table()
    insert_to_db()

```

## Appendix 3: Python Bokeh Plots

```
from bokeh.plotting import figure, show, save, ColumnDataSource
from bokeh.models import HoverTool
from bokeh.io import output_file
import datetime as dt
import os
import numpy as np

# Color schema [269 different colors]
# source = https://graphicdesign.stackexchange.com/revisions/3815/8
colors_269 = [
"#000000", "#FFFFFF", "#1CE6FF", "#FF34FF", "#FF4A46", "#008941",
"#006FA6", "#A30059", "#FFDBE5",
"#7A4900", "#0000A6", "#63FFAC", "#B79762", "#004D43", "#8FB0FF",
"#997D87", "#5A0007", "#809693",
"#FEFFE6", "#1B4400", "#4FC601", "#3B5DFF", "#4A3B53", "#FF2F80",
"#61615A", "#BA0900", "#6B7900",
"#00C2A0", "#FFAA92", "#FF90C9", "#B903AA", "#D16100", "#DDEFFF",
"#000035", "#7B4F4B", "#A1C299",
"#300018", "#0AA6D8", "#013349", "#00846F", "#372101", "#FFB500",
"#C2FFED", "#A079BF", "#CC0744",
"#C0B9B2", "#C2FF99", "#001E09", "#00489C", "#6F0062", "#0CBD66",
"#EEC3FF", "#456D75", "#B77B68",
"#7A87A1", "#788D66", "#885578", "#FAD09F", "#FF8A9A", "#D157A0",
"#BEC459", "#456648", "#0086ED",
"#886F4C", "#34362D", "#B4A8BD", "#00A6AA", "#452C2C", "#636375",
"#A3C8C9", "#FF913F", "#938A81",
"#575329", "#00FECF", "#B05B6F", "#8CD0FF", "#3B9700", "#04F757",
"#C8A1A1", "#1E6E00", "#7900D7",
"#A77500", "#6367A9", "#A05837", "#6B002C", "#772600", "#D790FF",
"#9B9700", "#549E79", "#FFF69F",
"#201625", "#72418F", "#BC23FF", "#99ADC0", "#3A2465", "#922329",
"#5B4534", "#FDE8DC", "#404E55",
"#0089A3", "#CB7E98", "#A4E804", "#324E72", "#6A3A4C", "#83AB58",
"#001C1E", "#D1F7CE", "#004B28",
"#C8D0F6", "#A3A489", "#806C66", "#222800", "#BF5650", "#E83000",
"#66796D", "#DA007C", "#FF1A59",
"#8ADB84", "#1E0200", "#5B4E51", "#C895C5", "#320033", "#FF6832",
"#66E1D3", "#CFCDAC", "#D0AC94",
"#7ED379", "#012C58", "#7A7BFF", "#D68E01", "#353339", "#78AFA1",
"#FEB2C6", "#75797C", "#837393",
"#943A4D", "#B5F4FF", "#D2DCD5", "#9556BD", "#6A714A", "#001325",
"#02525F", "#0AA3F7", "#E98176",
"#DBD5DD", "#5EBCD1", "#3D4F44", "#7E6405", "#02684E", "#962B75",
"#8D8546", "#9695C5", "#E773CE",
"#D86A78", "#3E89BE", "#CA834E", "#518A87", "#5B113C", "#55813B",
"#E704C4", "#00005F", "#A97399",
"#4B8160", "#59738A", "#FF5DA7", "#F7C9BF", "#643127", "#513A01",
"#6B94AA", "#51A058", "#A45B02",
"#1D1702", "#E20027", "#E7AB63", "#4C6001", "#9C6966", "#64547B",
"#97979E", "#006A66", "#391406",
"#F4D749", "#0045D2", "#006C31", "#DDB6D0", "#7C6571", "#9FB2A4",
"#00D891", "#15A08A", "#BC65E9",
"#FFFFFFE", "#C6DC99", "#203B3C", "#671190", "#6B3A64", "#F5E1FF",
"#FFA0F2", "#CCAA35", "#374527",
"#8BB400", "#797868", "#C6005A", "#3B000A", "#C86240", "#29607C",
"#402334", "#7D5A44", "#CCB87C",
"#B88183", "#AA5199", "#B5D6C3", "#A38469", "#9F94F0", "#A74571",
"#B894A6", "#71BB8C", "#00B433",
```

```

"#789EC9", "#6D80BA", "#953F00", "#5EFF03", "#E4FFFC", "#1BE177",
"#BCB1E5", "#76912F", "#003109",
"#0060CD", "#D20096", "#895563", "#29201D", "#5B3213", "#A76F42",
"#89412E", "#1A3A2A", "#494B5A",
"#A88C85", "#F4ABAA", "#A3F3AB", "#00C6C8", "#EA8B66", "#958A9F",
"#BDC9D2", "#9FA064", "#BE4700",
"#658188", "#83A485", "#453C23", "#47675D", "#3A3F00", "#061203",
"#DFFB71", "#868E7E", "#98D058",
"#6C8F7D", "#D7BFC2", "#3C3E6E", "#D83D66", "#2F5D9B", "#6C5E46",
"#D25B88", "#5B656C", "#00B57F",
"#545C46", "#866097", "#365D25", "#252F99", "#00CCFF", "#674E60",
"#FC009C", "#92896B"
]
# Read a file
def reading_file(fpath):
    listx = []
    tseconds = []
    listy = []
    listmac = []
    color_counter = 0
    colors = []
    names = []
    macdict = {}

    # Read file
    with open(fpath, 'r') as f:
        # Skip header
        next(f)
        for line in f:
            t = line.strip().split(";")
            # Assign color [check if it is a 'new' mac]
            if t[2] in macdict:
                colors.append(macdict[t[2]])
            else: # we need a new color + we store it
                color_counter += 1
                newcolor = colors_269[color_counter]
                macdict[t[2]] = newcolor
                colors.append(newcolor)

            # List with time strings
            listx.append(t[1])
            #(h, m, s) = t[1].split(':')
            # List with mac addresses
            listmac.append(t[2])
            names.append(t[2])
            # List with temperatures
            listy.append(t[3])

    return listx, listy, listmac, names, colors

# Plot time series
def plot_tseries(fpath):
    # Get filename [without extension]
    f_name = os.path.basename(fpath).replace('.csv', '')
    listx, listy, listmac, names, colors = reading_file(fpath)

    # Convert datetimes to X axis
    x = [(dt.datetime.strptime(d, '%H:%M:%S').time()) for d in listx]
    temp = [xi + '0' for xi in listy] #add 0 to no data values
    y = [float(xi) for xi in temp] #convert temperatures to floats

```

```

# Prepare tooltips
# - Source data dict
source = ColumnDataSource(data=dict(
    x=x,
    y=y,
    color=colors,
    names=names,
    dates=listx
))

# Plot with bokeh
TOOLS = 'pan,wheel_zoom,box_zoom,reset,hover,save'
p = figure(width=1300, height=600, title='Temperature vs Time
{}'.format(f_name), x_axis_type='datetime', tools=TOOLS)
p.xaxis.axis_label = 'Time'
p.yaxis.axis_label = 'Temperature'
p.square(x='x', y='y', color='color', source=source)

# - Mouse hover
hover = p.select_one(HoverTool)
hover.point_policy = "follow_mouse"
hover.tooltips = """
<div>
  <div>
    <span style="font-size: 12px;font-weight: bold;">Sensor:</span>
    <span style="font-size: 12px; color: #777777;">@names</span>
  </div>
  <div>
    <span style="font-size: 12px;font-weight:
bold;">Temperature:</span>
    <span style="font-size: 12px; color: #777777;">@y</span>
  </div>
  <div>
    <span style="font-size: 12px;font-weight: bold;">Time:</span>
    <span style="font-size: 12px; color: #777777;">@dates</span>
  </div>
</div>
"""

# Save to file
output_file(fpath.replace('.csv', '.html'))
save(p)

## MAIN
if __name__ == "__main__":
    # For every csv in a directory
    dir = r'D:\MASTER THESIS\Data\sensor measurements\hot
days\raw_plots\raw_plots'
    for f in os.listdir(dir):
        if f.endswith('.csv'):
            print 'Processing: {}'.format(f)
            plot_tseries(os.path.join(dir, f)

```

## Appendix 4: Reports from the final statistical models

>>05/18/18 12:10:48

### REGRESSION

#### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```

Data set      : combined_best_ind_fin_clear
Dependent Variable : highest_t_diff
Number of Observations: 146
Mean dependent var : 3.02151   Number of Variables : 8
S.D. dependent var : 1.0686    Degrees of Freedom  : 138

R-squared      : 0.118578   F-statistic      : 2.65217
Adjusted R-squared : 0.073868   Prob(F-statistic) : 0.0132556
Sum squared residual: 146.948   Log likelihood   : -207.638
Sigma-square    : 1.06484   Akaike info criterion : 431.275
S.E. of regression : 1.03191   Schwarz criterion : 455.144
Sigma-square ML  : 1.0065
S.E of regression ML: 1.00324
  
```

```

-----
--
Variable      Coefficient      Std.Error      t-Statistic
Probability
-----
--
CONSTANT      3.77755          4.28657        0.881251      0.37972
net_rad       0.00304949      0.00536894     0.567988      0.57096
BD_AWA_100m  -0.0125145      0.147797       -0.0846737    0.93263
LC_AWA_200m  0.0353911       0.991983       0.0356771     0.97157
SVF_AWA_400m -4.81558        2.50069        -1.9257       0.05620
Veg_AWA_400m -0.156079       1.17649        -0.132666     0.89466
NonPS_AWA_300m 2.93255        1.37961        2.12564       0.03531
VTD_AWA_600m -3.12084        3.29058        -0.948416     0.34458
-----
--
  
```

#### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 163.427989

#### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	16.4840	0.00026

#### DIAGNOSTICS FOR HETEROSKEDASTICITY

##### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	7	7.0551	0.42317
Koenker-Bassett test	7	4.9782	0.66263

##### SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	35	21.3067	0.96675

#### DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : combined\_best\_ind\_fin\_clear

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.0455	1.4742	0.14043
Lagrange Multiplier (lag)	1	1.0917	0.29610
Robust LM (lag)	1	0.5914	0.44187
Lagrange Multiplier (error)	1	0.7395	0.38984
Robust LM (error)	1	0.2392	0.62480

Lagrange Multiplier (SARMA)            2                    1.3309                    0.51405

COEFFICIENTS VARIANCE MATRIX

CONSTANT	net_rad	BD_AWA_100m	LC_AWA_200m	SVF_AWA_400m
18.374720	-0.019763	-0.070667	-1.338658	-6.651736
-0.019763	0.000029	-0.000073	0.001146	0.002059
-0.070667	-0.000073	0.021844	-0.026779	0.138384
-1.338658	0.001146	-0.026779	0.984031	0.788677
-6.651736	0.002059	0.138384	0.788677	6.253434
-2.702396	0.001143	0.027532	0.460799	1.863773
-1.444711	0.002579	-0.104110	-0.137037	-0.631743
-2.981668	-0.001377	0.195285	-0.956838	3.758175

Veg_AWA_400m	NonPS_AWA_300m	VTD_AWA_600m
-2.702396	-1.444711	-2.981668
0.001143	0.002579	-0.001377
0.027532	-0.104110	0.195285
0.460799	-0.137037	-0.956838
1.863773	-0.631743	3.758175
1.384118	0.300129	1.100732
0.300129	1.903321	-1.911208
1.100732	-1.911208	10.827886

OBS	highest_t_diff	PREDICTED	RESIDUAL
1	3.70000	3.37422	0.32578
2	2.78000	3.15465	-0.37465
3	3.95000	3.02146	0.92854
4	3.10000	2.78045	0.31955
5	2.10000	2.35939	-0.25939
6	2.25000	2.41275	-0.16275
7	1.53000	3.15822	-1.62822
8	3.08000	3.18799	-0.10799
9	1.60000	2.22073	-0.62073
10	3.00000	2.83915	0.16085
11	1.83000	2.48074	-0.65074
12	4.00000	3.10026	0.89974
13	2.75000	2.51499	0.23501
14	3.28000	3.89142	-0.61142
15	1.48000	2.03426	-0.55426
16	0.90000	2.83703	-1.93703
17	2.70000	2.75707	-0.05707
18	3.18000	3.04490	0.13510
19	2.88000	3.37347	-0.49347
20	2.65000	2.87773	-0.22773
21	2.23000	3.42805	-1.19805
22	2.70000	3.31159	-0.61159
23	2.15000	3.21839	-1.06839
24	3.30000	3.06523	0.23477
25	2.88000	2.97193	-0.09193
26	3.55000	3.47826	0.07174
27	2.53000	2.87223	-0.34223
28	2.55000	3.28186	-0.73186
29	2.55000	2.76702	-0.21702
30	2.80000	2.34969	0.45031
31	2.95000	3.32565	-0.37565
32	3.63000	2.93126	0.69874
33	2.58000	3.48362	-0.90362
34	1.65000	3.40879	-1.75879
35	3.43000	3.41692	0.01308
36	1.45000	2.85270	-1.40270

37	1.93000	3.14618	-1.21618
38	1.18000	3.13696	-1.95696
39	4.05000	3.51650	0.53350
40	3.98000	2.97439	1.00561
41	2.40000	2.67020	-0.27020
42	4.18000	3.21577	0.96423
43	3.23000	3.41110	-0.18110
44	2.88000	3.09873	-0.21873
45	2.63000	3.37288	-0.74288
46	2.70000	3.03441	-0.33441
47	3.15000	2.99098	0.15902
48	3.13000	3.44615	-0.31615
49	1.60000	3.14752	-1.54752
50	4.58000	3.33259	1.24741
51	5.80000	2.98270	2.81730
52	3.15000	3.49631	-0.34631
53	3.00000	2.85340	0.14660
54	2.88000	3.11086	-0.23086
55	4.40000	2.99711	1.40289
56	3.63000	3.17239	0.45761
57	4.18000	2.62259	1.55741
58	1.48000	2.60224	-1.12224
59	1.35000	2.37108	-1.02108
60	2.25000	3.37929	-1.12929
61	2.18000	2.59294	-0.41294
62	4.08000	3.17971	0.90029
63	4.15000	3.45864	0.69136
64	2.35000	2.77952	-0.42952
65	3.90000	2.83765	1.06235
66	2.65000	3.16329	-0.51329
67	3.63000	3.14388	0.48612
68	2.48000	3.05831	-0.57831
69	3.23000	2.57287	0.65713
70	2.73000	3.33380	-0.60380
71	2.13000	2.16809	-0.03809
72	5.48000	3.20604	2.27396
73	1.53000	2.25244	-0.72244
74	3.65000	2.87644	0.77356
75	3.10000	3.29118	-0.19118
76	2.35000	2.93074	-0.58074
77	2.45000	2.86354	-0.41354
78	2.05000	3.15866	-1.10866
79	2.58000	3.19738	-0.61738
80	3.73000	2.71456	1.01544
81	1.78000	3.06819	-1.28819
82	3.68000	3.29723	0.38277
83	2.90000	2.86959	0.03041
84	2.73000	3.51032	-0.78032
85	1.83000	3.18074	-1.35074
86	2.98000	2.83975	0.14025
87	2.88000	3.36306	-0.48306
88	3.33000	2.96165	0.36835
89	2.95000	3.65044	-0.70044
90	6.00000	3.02024	2.97976
91	1.33000	2.53993	-1.20993
92	3.03000	3.07764	-0.04764
93	2.38000	2.94322	-0.56322
94	4.28000	3.20911	1.07089
95	3.58000	3.49063	0.08937
96	5.48000	3.07389	2.40611
97	4.55000	3.15079	1.39921

98	3.68000	3.30473	0.37527
99	3.08000	2.93641	0.14359
100	2.93000	2.86634	0.06366
101	5.10000	3.14636	1.95364
102	3.30000	3.08451	0.21549
103	3.45000	3.10848	0.34152
104	4.33000	3.27225	1.05775
105	2.05000	2.48893	-0.43893
106	5.98000	3.49820	2.48180
107	1.83000	2.54052	-0.71052
108	1.88000	3.98009	-2.10009
109	3.23000	2.53345	0.69655
110	2.53000	3.15956	-0.62956
111	2.20000	3.13917	-0.93917
112	4.28000	3.59355	0.68645
113	2.90000	2.89859	0.00141
114	3.53000	3.06489	0.46511
115	4.43000	3.18606	1.24394
116	3.93000	3.45117	0.47883
117	1.68000	2.95615	-1.27615
118	3.40000	2.97833	0.42167
119	3.05000	3.17362	-0.12362
120	3.55000	4.14096	-0.59096
121	2.25000	3.00339	-0.75339
122	3.28000	2.83566	0.44434
123	3.10000	2.72899	0.37101
124	6.08000	3.17052	2.90948
125	2.68000	3.34175	-0.66175
126	4.50000	2.81782	1.68218
127	2.08000	2.33456	-0.25456
128	3.30000	3.16472	0.13528
129	5.28000	3.11268	2.16732
130	2.43000	2.46910	-0.03910
131	0.27000	2.34775	-2.07775
132	2.58000	2.49599	0.08401
133	2.03000	2.38934	-0.35934
134	2.30000	3.14337	-0.84337
135	2.78000	2.81902	-0.03902
136	2.55000	3.27848	-0.72848
137	3.63000	2.99478	0.63522
138	1.78000	2.78486	-1.00486
139	4.03000	3.35356	0.67644
140	4.35000	3.24902	1.10098
141	3.45000	3.36319	0.08681
142	5.40000	2.67288	2.72712
143	2.45000	3.46287	-1.01287
144	1.73000	2.50657	-0.77657
145	2.98000	2.49968	0.48032
146	2.00000	2.60470	-0.60470

===== END OF REPORT  
=====

>>05/18/18 12:11:46

REGRESSION

-----

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : combined\_best\_ind\_fin\_clear  
Dependent Variable : highest\_t\_diff  
Number of Observations: 146  
Mean dependent var : 3.02151 Number of Variables : 3  
S.D. dependent var : 1.0686 Degrees of Freedom : 143  
  
R-squared : 0.109414 F-statistic : 8.78422  
Adjusted R-squared : 0.096958 Prob(F-statistic) : 0.000252245  
Sum squared residual: 148.476 Log likelihood : -208.393  
Sigma-square : 1.0383 Akaike info criterion : 422.786  
S.E. of regression : 1.01897 Schwarz criterion : 431.736  
Sigma-square ML : 1.01696  
S.E of regression ML: 1.00844

-----

--

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	4.79853	1.14541	4.18934	0.00005
SVF_AWA_400m	-3.93405	1.58479	-2.48238	0.01421
NonPS_AWA_300m	2.0318	0.64562	3.14704	0.00201

-----

--

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 31.818256

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	17.7657	0.00014

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	2.3235	0.31293
Koenker-Bassett test	2	1.6415	0.44010

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	5	3.3276	0.64962

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : combined\_best\_ind\_fin\_clear

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.0422	1.1913	0.23354
Lagrange Multiplier (lag)	1	0.9425	0.33162
Robust LM (lag)	1	0.5130	0.47386
Lagrange Multiplier (error)	1	0.6364	0.42503
Robust LM (error)	1	0.2068	0.64930
Lagrange Multiplier (SARMA)	2	1.1493	0.56289

COEFFICIENTS VARIANCE MATRIX

CONSTANT	SVF_AWA_400m	NonPS_AWA_300m
1.311975	-1.752409	-0.246814
-1.752409	2.511549	0.089565
-0.246814	0.089565	0.416825

OBS	highest_t_diff	PREDICTED	RESIDUAL
1	3.70000	3.23569	0.46431
2	2.78000	3.14794	-0.36794
3	3.95000	3.04894	0.90106
4	3.10000	2.73940	0.36060
5	2.10000	2.35465	-0.25465
6	2.25000	2.41431	-0.16431
7	1.53000	3.27632	-1.74632
8	3.08000	3.30053	-0.22053
9	1.60000	2.27467	-0.67467
10	3.00000	3.01737	-0.01737
11	1.83000	2.47397	-0.64397
12	4.00000	3.11119	0.88881
13	2.75000	2.45235	0.29765
14	3.28000	3.82491	-0.54491
15	1.48000	2.04511	-0.56511
16	0.90000	2.92703	-2.02703
17	2.70000	2.58074	0.11926
18	3.18000	3.18121	-0.00121
19	2.88000	3.38439	-0.50439
20	2.65000	3.03769	-0.38769
21	2.23000	3.38698	-1.15698
22	2.70000	3.13539	-0.43539
23	2.15000	3.16437	-1.01437
24	3.30000	3.03510	0.26490
25	2.88000	2.97933	-0.09933
26	3.55000	3.49852	0.05148
27	2.53000	2.76231	-0.23231
28	2.55000	3.25600	-0.70600
29	2.55000	2.72167	-0.17167
30	2.80000	2.37497	0.42503
31	2.95000	3.32084	-0.37084
32	3.63000	2.94905	0.68095
33	2.58000	3.37532	-0.79532
34	1.65000	3.15571	-1.50571
35	3.43000	3.29794	0.13206
36	1.45000	2.84358	-1.39358
37	1.93000	3.17862	-1.24862
38	1.18000	3.04764	-1.86764
39	4.05000	3.36666	0.68334
40	3.98000	2.99317	0.98683
41	2.40000	2.63522	-0.23522
42	4.18000	3.18121	0.99879
43	3.23000	3.49982	-0.26982
44	2.88000	3.10512	-0.22512
45	2.63000	3.40082	-0.77082
46	2.70000	3.09476	-0.39476
47	3.15000	2.97414	0.17586
48	3.13000	3.30571	-0.17571
49	1.60000	3.15053	-1.55053
50	4.58000	3.27632	1.30368
51	5.80000	3.09994	2.70006
52	3.15000	3.49852	-0.34852
53	3.00000	2.88510	0.11490
54	2.88000	3.04505	-0.16505
55	4.40000	2.93739	1.46261
56	3.63000	3.05671	0.57329
57	4.18000	2.62526	1.55474
58	1.48000	2.56042	-1.08042

59	1.35000	2.37367	-1.02367
60	2.25000	3.44146	-1.19146
61	2.18000	2.64040	-0.46040
62	4.08000	3.27891	0.80109
63	4.15000	3.49852	0.65148
64	2.35000	2.90671	-0.55671
65	3.90000	2.81419	1.08581
66	2.65000	3.17085	-0.52085
67	3.63000	3.33857	0.29143
68	2.48000	3.25989	-0.77989
69	3.23000	2.48692	0.74308
70	2.73000	3.25082	-0.52082
71	2.13000	2.50206	-0.37206
72	5.48000	3.11637	2.36363
73	1.53000	2.23533	-0.70533
74	3.65000	2.72296	0.92704
75	3.10000	3.21278	-0.11278
76	2.35000	2.80424	-0.45424
77	2.45000	2.87255	-0.42255
78	2.05000	3.20023	-1.15023
79	2.58000	3.13798	-0.55798
80	3.73000	2.59976	1.13024
81	1.78000	3.07573	-1.29573
82	3.68000	3.14705	0.53295
83	2.90000	2.95124	-0.05124
84	2.73000	3.46437	-0.73437
85	1.83000	3.25600	-1.42600
86	2.98000	2.72296	0.25704
87	2.88000	3.46048	-0.58048
88	3.33000	3.09735	0.23265
89	2.95000	3.48727	-0.53727
90	6.00000	3.01608	2.98392
91	1.33000	2.51201	-1.18201
92	3.03000	3.17603	-0.14603
93	2.38000	2.91578	-0.53578
94	4.28000	3.23569	1.04431
95	3.58000	3.44275	0.13725
96	5.48000	3.09605	2.38395
97	4.55000	3.22573	1.32427
98	3.68000	3.38050	0.29950
99	3.08000	3.01348	0.06652
100	2.93000	2.97933	-0.04933
101	5.10000	3.13539	1.96461
102	3.30000	3.01737	0.28263
103	3.45000	3.31955	0.13045
104	4.33000	3.21537	1.11463
105	2.05000	2.51849	-0.46849
106	5.98000	3.42373	2.55627
107	1.83000	2.69617	-0.86617
108	1.88000	3.70170	-1.82170
109	3.23000	2.37756	0.85244
110	2.53000	3.28150	-0.75150
111	2.20000	3.30182	-1.10182
112	4.28000	3.48080	0.79920
113	2.90000	3.13928	-0.23928
114	3.53000	3.23828	0.29172
115	4.43000	3.21666	1.21334
116	3.93000	3.54564	0.38436
117	1.68000	3.05282	-1.37282
118	3.40000	2.93480	0.46520
119	3.05000	3.09346	-0.04346



>>05/18/18 12:12:52

REGRESSION

-----  
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : combined\_best\_ind\_fin\_clear  
Spatial Weight : combined\_best\_ind\_fin\_clear  
Dependent Variable : highest\_t\_diff  
Number of Observations: 146  
Mean dependent var : 3.021507 Number of Variables : 3  
S.D. dependent var : 1.068597 Degrees of Freedom : 143  
Lag coeff. (Lambda) : 0.092735

R-squared : 0.114178 R-squared (BUSE) : -  
Sq. Correlation : - Log likelihood : -208.110430  
Sigma-square : 1.01152 Akaike info criterion : 422.221  
S.E of regression : 1.00574 Schwarz criterion : 431.172

-----  
--

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	4.83313	1.18686	4.07221	0.00005
SVF_AWA_400m	-3.97429	1.6491	-2.40998	0.01595
NonPS_AWA_300m	2.01155	0.67174	2.99454	0.00275
LAMBDA	0.0927353	0.1308	0.708984	0.47833

-----  
--

REGRESSION DIAGNOSTICS  
DIAGNOSTICS FOR HETEROSKEDASTICITY  
RANDOM COEFFICIENTS

TEST DF VALUE PROB  
Breusch-Pagan test 2 1.7266 0.42178

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : combined\_best\_ind\_fin\_clear  
TEST DF VALUE PROB  
Likelihood Ratio Test 1 0.5647 0.45239

COEFFICIENTS VARIANCE MATRIX

CONSTANT	SVF_AWA_400m	NonPS_AWA_300m	LAMBDA
1.408629	-1.888330	-0.251874	0.000000
-1.888330	2.719523	0.074742	0.000000
-0.251874	0.074742	0.451234	0.000000
0.000000	0.000000	0.000000	0.017109

OBS	highest_t_diff	PREDICTED	RESIDUAL	PRED ERROR
1	3.7	3.23502	0.44340	0.46498
2	2.78	3.15211	-0.33464	-0.37211
3	3.95	3.05251	0.94105	0.89749
4	3.1	2.73652	0.38702	0.36348
5	2.1	2.35481	-0.21414	-0.25481
6	2.25	2.41467	-0.12296	-0.16467
7	1.53	3.27525	-1.74298	-1.74525
8	3.08	3.29683	-0.20129	-0.21683
9	1.6	2.27484	-0.63121	-0.67484
10	3	3.01570	-0.02567	-0.01570
11	1.83	2.47453	-0.60838	-0.64453

12	4	3.11335	0.85835	0.88665
13	2.75	2.45393	0.31439	0.29607
14	3.28	3.81837	-0.52341	-0.53837
15	1.48	2.03882	-0.83562	-0.55882
16	0.9	2.93182	-1.97837	-2.03182
17	2.7	2.57706	0.09503	0.12294
18	3.18	3.17711	-0.01634	0.00289
19	2.88	3.37827	-0.50956	-0.49827
20	2.65	3.03581	-0.43282	-0.38581
21	2.23	3.37924	-1.15289	-1.14924
22	2.7	3.13493	-0.40374	-0.43493
23	2.15	3.17076	-1.00182	-1.02076
24	3.3	3.03484	0.29647	0.26516
25	2.88	2.97644	-0.11043	-0.09644
26	3.55	3.49603	0.03782	0.05397
27	2.53	2.75761	-0.19380	-0.22761
28	2.55	3.25513	-0.80302	-0.70513
29	2.55	2.71738	-0.13291	-0.16738
30	2.8	2.37493	0.47362	0.42507
31	2.95	3.31694	-0.36664	-0.36694
32	3.63	2.94012	0.72618	0.68988
33	2.58	3.37485	-0.80497	-0.79485
34	1.65	3.15504	-1.50351	-1.50504
35	3.43	3.29585	0.11121	0.13415
36	1.45	2.83807	-1.36127	-1.38807
37	1.93	3.17613	-1.24255	-1.24613
38	1.18	3.05202	-1.77192	-1.87202
39	4.05	3.35913	0.69441	0.69087
40	3.98	2.99412	0.95723	0.98588
41	2.4	2.63497	-0.26133	-0.23497
42	4.18	3.17711	0.96743	1.00289
43	3.23	3.49652	-0.27222	-0.26652
44	2.88	3.09860	-0.20670	-0.21860
45	2.63	3.39692	-0.77324	-0.76692
46	2.7	3.09470	-0.42553	-0.39470
47	3.15	2.97449	0.06350	0.17551
48	3.13	3.29878	-0.17664	-0.16878
49	1.6	3.15309	-1.50723	-1.55309
50	4.58	3.27525	1.32248	1.30475
51	5.8	3.09665	2.70856	2.70335
52	3.15	3.49603	-0.36555	-0.34603
53	3	2.89110	0.08520	0.10890
54	2.88	3.05105	-0.11558	-0.17105
55	4.4	2.93572	1.48890	1.46428
56	3.63	3.05544	0.53608	0.57456
57	4.18	2.61876	1.60853	1.56124
58	1.48	2.55695	-1.05451	-1.07695
59	1.35	2.37444	-0.99269	-1.02444
60	2.25	3.43715	-1.18549	-1.18715
61	2.18	2.63692	-0.46053	-0.45692
62	4.08	3.27622	0.83306	0.80378
63	4.15	3.49603	0.64524	0.65397
64	2.35	2.91170	-0.49121	-0.56170
65	3.9	2.81454	0.97114	1.08546
66	2.65	3.17320	-0.54138	-0.52320
67	3.63	3.33608	0.30207	0.29392
68	2.48	3.25660	-0.79606	-0.77660
69	3.23	2.47941	0.73238	0.75059
70	2.73	3.25318	-0.53370	-0.52318
71	2.13	2.49758	-0.41878	-0.36758
72	5.48	3.11530	2.43800	2.36470

73	1.53	2.23510	-0.84988	-0.70510
74	3.65	2.71787	0.91673	0.93213
75	3.1	3.21392	-0.08701	-0.11392
76	2.35	2.79833	-0.38933	-0.44833
77	2.45	2.87391	-0.39995	-0.42391
78	2.05	3.19674	-1.09477	-1.14674
79	2.58	3.13590	-0.56693	-0.55590
80	3.73	2.59669	1.12102	1.13331
81	1.78	3.07507	-1.39936	-1.29507
82	3.68	3.13932	0.51916	0.54068
83	2.9	2.95340	-0.02494	-0.05340
84	2.73	3.45824	-0.78150	-0.72824
85	1.83	3.25513	-1.42305	-1.42513
86	2.98	2.71787	0.23637	0.26213
87	2.88	3.45677	-0.59199	-0.57677
88	3.33	3.09567	0.22023	0.23433
89	2.95	3.47933	-0.52975	-0.52933
90	6	3.01521	3.03331	2.98479
91	1.33	2.51379	-1.15052	-1.18379
92	3.03	3.17516	-0.19075	-0.14516
93	2.38	2.91512	-0.57332	-0.53512
94	4.28	3.23502	1.05134	1.04498
95	3.58	3.43764	0.12908	0.14236
96	5.48	3.09518	2.33115	2.38482
97	4.55	3.21881	1.34239	1.33119
98	3.68	3.37680	0.28694	0.30320
99	3.08	3.01423	0.03381	0.06577
100	2.93	2.97644	-0.07699	-0.04644
101	5.1	3.13493	1.90542	1.96507
102	3.3	3.01570	0.23933	0.28430
103	3.45	3.31645	0.14077	0.13355
104	4.33	3.21490	1.09338	1.11510
105	2.05	2.51623	-0.49201	-0.46623
106	5.98	3.41801	2.59741	2.56199
107	1.83	2.69531	-0.77736	-0.86531
108	1.88	3.69718	-1.84232	-1.81718
109	3.23	2.37591	0.79284	0.85409
110	2.53	3.27720	-0.75764	-0.74720
111	2.2	3.29732	-1.02772	-1.09732
112	4.28	3.47689	0.83148	0.80311
113	2.9	3.13639	-0.19235	-0.23639
114	3.53	3.23599	0.28443	0.29401
115	4.43	3.21539	1.16340	1.21461
116	3.93	3.53870	0.40985	0.39130
117	1.68	3.05398	-1.34595	-1.37398
118	3.4	2.93475	0.42508	0.46525
119	3.05	3.09421	-0.09155	-0.04421
120	3.55	3.83751	-0.28960	-0.28751
121	2.25	3.11823	-0.82326	-0.86823
122	3.28	2.99558	0.39076	0.28442
123	3.1	2.67276	0.35862	0.42724
124	6.08	3.05446	2.96525	3.02554
125	2.68	3.49896	-0.89058	-0.81896
126	4.5	2.85331	1.58274	1.64669
127	2.08	2.45686	-0.33072	-0.37686
128	3.3	3.21588	0.09893	0.08412
129	5.28	3.21685	2.09278	2.06315
130	2.43	2.45686	0.07301	-0.02686
131	0.27	2.39700	-2.15478	-2.12700
132	2.58	2.45686	0.16777	0.12314
133	2.03	2.39456	-0.31790	-0.36456



>>05/18/18 12:12:16

REGRESSION

-----  
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION  
Data set : combined\_best\_ind\_fin\_clear  
Spatial Weight : combined\_best\_ind\_fin\_clear  
Dependent Variable : highest\_t\_diff  
Number of Observations: 146  
Mean dependent var : 3.02151 Number of Variables : 4  
S.D. dependent var : 1.0686 Degrees of Freedom : 142  
Lag coeff. (Rho) : 0.101731  
  
R-squared : 0.115792 Log likelihood : -208  
Sq. Correlation : - Akaike info criterion : 424  
Sigma-square : 1.00968 Schwarz criterion : 435.934  
S.E of regression : 1.00483

-----  
--

Variable	Coefficient	Std.Error	z-value	Probability
W_highest_t_	0.101731	0.125289	0.811975	0.41681
CONSTANT	4.4156	1.25908	3.50699	0.00045
SVF_AWA_400m	-3.73532	1.6063	-2.32542	0.02005
NonPS_AWA_300m	1.88879	0.659903	2.86222	0.00421

-----  
--

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	1.6069	0.44778

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : combined\_best\_ind\_fin\_clear

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	0.7857	0.37540

COEFFICIENTS VARIANCE MATRIX

CONSTANT	SVF_AWA_400m	NonPS_AWA_300m	W_highest_t_
1.585294	-1.910665	-0.143439	-0.069699
-1.910665	2.580187	0.022641	0.046519
-0.143439	0.022641	0.435471	-0.021749
-0.069699	0.046519	-0.021749	0.015697

OBS	highest_t_diff	PREDICTED	RESIDUAL	PRED ERROR
1	3.7	3.21492	0.46214	0.48508
2	2.78	3.14901	-0.32760	-0.36901
3	3.95	3.06834	0.92968	0.88166
4	3.1	2.71223	0.41253	0.38777
5	2.1	2.32816	-0.18643	-0.22816
6	2.25	2.38369	-0.09076	-0.13369
7	1.53	3.27611	-1.74299	-1.74611
8	3.08	3.29668	-0.19884	-0.21668
9	1.6	2.26131	-0.61676	-0.66131
10	3	3.04084	-0.05129	-0.04084
11	1.83	2.43922	-0.57234	-0.60922

12	4	3.11635	0.85362	0.88365
13	2.75	2.45401	0.31254	0.29599
14	3.28	3.82016	-0.52166	-0.54016
15	1.48	2.09218	-0.91914	-0.61218
16	0.9	2.94232	-1.98361	-2.04232
17	2.7	2.55623	0.11099	0.14377
18	3.18	3.17802	-0.01809	0.00198
19	2.88	3.41001	-0.54098	-0.53001
20	2.65	3.04420	-0.44554	-0.39420
21	2.23	3.36525	-1.13861	-1.13525
22	2.7	3.12595	-0.39152	-0.42595
23	2.15	3.16570	-0.99454	-1.01570
24	3.3	3.04167	0.29275	0.25833
25	2.88	2.99764	-0.13235	-0.11764
26	3.55	3.50356	0.03008	0.04644
27	2.53	2.71964	-0.15465	-0.18964
28	2.55	3.23433	-0.79125	-0.68433
29	2.55	2.68579	-0.10044	-0.13579
30	2.8	2.34681	0.50359	0.45319
31	2.95	3.31734	-0.36618	-0.36734
32	3.63	2.96269	0.70728	0.66731
33	2.58	3.39886	-0.82867	-0.81886
34	1.65	3.15063	-1.49910	-1.50063
35	3.43	3.27405	0.13134	0.15595
36	1.45	2.79046	-1.31317	-1.34046
37	1.93	3.18272	-1.24816	-1.25272
38	1.18	3.04838	-1.75764	-1.86838
39	4.05	3.34290	0.71216	0.70710
40	3.98	2.99321	0.95650	0.98679
41	2.4	2.65792	-0.28767	-0.25792
42	4.18	3.15970	0.98184	1.02030
43	3.23	3.49729	-0.27257	-0.26729
44	2.88	3.10999	-0.21638	-0.22999
45	2.63	3.40402	-0.78000	-0.77402
46	2.7	3.10552	-0.43917	-0.40552
47	3.15	2.97160	0.05515	0.17840
48	3.13	3.28705	-0.16455	-0.15705
49	1.6	3.15144	-1.50024	-1.55144
50	4.58	3.26981	1.32956	1.31019
51	5.8	3.11646	2.69008	2.68354
52	3.15	3.50356	-0.37362	-0.35356
53	3	2.90884	0.06593	0.09116
54	2.88	3.06678	-0.12597	-0.18678
55	4.4	2.93143	1.49533	1.46857
56	3.63	3.05629	0.53129	0.57371
57	4.18	2.58205	1.64842	1.59795
58	1.48	2.53236	-1.03010	-1.05236
59	1.35	2.34639	-0.96443	-0.99639
60	2.25	3.42666	-1.17400	-1.17666
61	2.18	2.60840	-0.43490	-0.42840
62	4.08	3.27814	0.83479	0.80186
63	4.15	3.49867	0.64321	0.65133
64	2.35	2.92366	-0.49629	-0.57366
65	3.9	2.81269	0.96208	1.08731
66	2.65	3.19409	-0.56303	-0.54409
67	3.63	3.33982	0.29973	0.29018
68	2.48	3.26840	-0.80910	-0.78840
69	3.23	2.46586	0.74188	0.76414
70	2.73	3.24957	-0.53123	-0.51957
71	2.13	2.48381	-0.41330	-0.35381
72	5.48	3.11402	2.44618	2.36598

73	1.53	2.23608	-0.86864	-0.70608
74	3.65	2.68624	0.94488	0.96376
75	3.1	3.20786	-0.07810	-0.10786
76	2.35	2.75781	-0.34454	-0.40781
77	2.45	2.89008	-0.41386	-0.44008
78	2.05	3.19653	-1.08877	-1.14653
79	2.58	3.13060	-0.56286	-0.55060
80	3.73	2.57439	1.13999	1.15561
81	1.78	3.05682	-1.39194	-1.27682
82	3.68	3.15877	0.49939	0.52123
83	2.9	2.96511	-0.03437	-0.06511
84	2.73	3.48514	-0.81157	-0.75514
85	1.83	3.25632	-1.42336	-1.42632
86	2.98	2.68624	0.26352	0.29376
87	2.88	3.45501	-0.59067	-0.57501
88	3.33	3.11480	0.20018	0.21520
89	2.95	3.50259	-0.55128	-0.55259
90	6	2.98262	3.07132	3.01738
91	1.33	2.47889	-1.11525	-1.14889
92	3.03	3.17436	-0.19401	-0.14436
93	2.38	2.91773	-0.57941	-0.53773
94	4.28	3.25276	1.03485	1.02724
95	3.58	3.43989	0.12639	0.14011
96	5.48	3.09843	2.32288	2.38157
97	4.55	3.25108	1.31198	1.29892
98	3.68	3.38304	0.28005	0.29696
99	3.08	3.02319	0.02200	0.05681
100	2.93	2.98314	-0.08663	-0.05314
101	5.1	3.13270	1.90183	1.96730
102	3.3	3.00158	0.24814	0.29842
103	3.45	3.32147	0.13705	0.12853
104	4.33	3.20287	1.10350	1.12713
105	2.05	2.50261	-0.48345	-0.45261
106	5.98	3.44580	2.57476	2.53420
107	1.83	2.71045	-0.78460	-0.88045
108	1.88	3.70524	-1.85082	-1.82524
109	3.23	2.41144	0.75204	0.81856
110	2.53	3.28151	-0.76223	-0.75151
111	2.2	3.29220	-1.01524	-1.09220
112	4.28	3.50248	0.81018	0.77752
113	2.9	3.14481	-0.19567	-0.24481
114	3.53	3.24186	0.27800	0.28814
115	4.43	3.20679	1.16722	1.22321
116	3.93	3.55106	0.40123	0.37894
117	1.68	3.04809	-1.33727	-1.36809
118	3.4	2.94486	0.41072	0.45514
119	3.05	3.09436	-0.09675	-0.04436
120	3.55	3.83502	-0.28539	-0.28502
121	2.25	3.12813	-0.82822	-0.87813
122	3.28	3.01295	0.38369	0.26705
123	3.1	2.68399	0.34063	0.41601
124	6.08	3.05684	2.95647	3.02316
125	2.68	3.49658	-0.89273	-0.81658
126	4.5	2.85964	1.57042	1.64036
127	2.08	2.45088	-0.32405	-0.37088
128	3.3	3.21973	0.09716	0.08027
129	5.28	3.22205	2.09121	2.05795
130	2.43	2.43921	0.09784	-0.00921
131	0.27	2.45542	-2.21803	-2.18542
132	2.58	2.43102	0.19506	0.14898
133	2.03	2.36426	-0.28595	-0.33426



\*\*\*\*\*  
\* Semiparametric Geographically Weighted Regression  
\* Release 1.0.90 (GWR 4.0.90)  
\* 12 May 2015  
\* (Originally coded by T. Nakaya: 1 Nov 2009)  
\* Tomoki Nakaya(1), Martin Charlton(2), Chris Brunson (2)  
\* Paul Lewis (2), Jing Yao (3), A Stewart Fotheringham (4)  
\* (c) GWR4 development team  
\* (1) Ritsumeikan University, (2) National University of Ireland, Maynooth,  
\* (3) University of Glasgow, (4) Arizona State University  
\*\*\*\*\*

Program began at 18.5.2018 r. 14:06:26  
\*\*\*\*\*  
\*\*

Session:  
Session control file: D:\MASTER  
THESIS\Data\Modelling\final\_joined\CSVs\AWA\_differrent\_buffers\combined\_best\_indicators\_FINAL\new\_t\_diff\gwr\GWR\_2\_ind.ctl  
\*\*\*\*\*  
\*\*

Data filename: D:\MASTER  
THESIS\Data\Modelling\final\_joined\CSVs\AWA\_differrent\_buffers\combined\_best\_indicators\_FINAL\new\_t\_diff\combined\_best\_ind\_fin\_clear\_GWR.csv  
Number of areas/points: 147

Model settings-----  
Model type: Gaussian  
Geographic kernel: adaptive Gaussian  
Method for optimal bandwidth search: Golden section search  
Criterion for optimal bandwidth: AIC  
Number of varying coefficients: 3  
Number of fixed coefficients: 0

Modelling options-----  
Standardisation of independent variables: OFF  
Testing geographical variability of local coefficients: OFF  
Local to Global Variable selection: OFF  
Global to Local Variable selection: OFF  
Prediction at non-regression points: OFF

Variable settings-----  
Area key: field2: MAC  
Easting (x-coord): field3 : xcoord  
Northing (y-coord): field4: ycoord  
Cartesian coordinates: Euclidean distance  
Dependent variable: field16: highest\_t\_diff  
Offset variable is not specified  
Intercept: varying (Local) intercept  
Independent variable with varying (Local) coefficient: field11:  
SVF\_AWA\_400m  
Independent variable with varying (Local) coefficient: field13:  
NonPS\_AWA\_300m  
\*\*\*\*\*  
\*\*

\*\*\*\*\*  
\*\*  
Global regression result  
\*\*\*\*\*  
\*\*

< Diagnostic information >

```

Residual sum of squares:          159,483691
Number of parameters:             3
(Note: this num does not include an error variance term for a Gaussian
model)
ML based global sigma estimate:   1,041596
Unbiased global sigma estimate:   1,052390
-2 log-likelihood:               429,149764
Classic AIC:                     437,149764
AICc:                            437,431454
BIC/MDL:                         449,111494
CV:                              1,124697
R square:                        0,092734
Adjusted R square:               0,073700

```

Variable	Estimate	Standard Error	t (Est/SE)
Intercept	4,815323	1,182973	4,070527
SVF_AWA_400m	-3,877811	1,636671	-2,369328
NonPS_AWA_300m	1,857430	0,664499	2,795233

```

*****
**

```

GWR (Geographically weighted regression) bandwidth selection

```

*****
**

```

Bandwidth search <golden section search>

Limits: 46, 147

Golden section search begins...

Initial values

```

pL      Bandwidth: 46,000 Criterion: 423,879
p1      Bandwidth: 48,150 Criterion: 424,002
p2      Bandwidth: 49,479 Criterion: 424,144
pU      Bandwidth: 51,629 Criterion: 424,391
iter  1 (p1) Bandwidth: 48,150 Criterion: 424,002 Diff: 1,329
iter  2 (p1) Bandwidth: 47,329 Criterion: 423,907 Diff: 0,821
iter  3 (p1) Bandwidth: 46,821 Criterion: 423,879 Diff: 0,508

```

Best bandwidth size 46,000

Minimum AIC 423,879

```

*****
**

```

GWR (Geographically weighted regression) result

```

*****
**

```

Bandwidth and geographic ranges

```

Bandwidth size: 46,821192
Coordinate      Min          Max          Range
-----
X-coord        74876,696480  88218,453760  13341,757280
Y-coord        448367,765400  458968,992700  10601,227300

```

Diagnostic information

```

Residual sum of squares:          139,194055
Effective number of parameters (model: trace(S)):
6,365861
Effective number of parameters (variance: trace(S'S)):
4,353976
Degree of freedom (model: n - trace(S)):
140,634139
Degree of freedom (residual: n - 2trace(S) + trace(S'S)):
138,622255

```

ML based sigma estimate: 0,973087  
 Unbiased sigma estimate: 1,002060  
 -2 log-likelihood: 409,147087  
 Classic AIC: 423,878808  
 AICc: 424,767792  
 BIC/MDL: 445,905917  
 CV: 1,025838  
 R square: 0,208157  
 Adjusted R square: 0,159953

\*\*\*\*\*  
 << Geographically varying (Local) coefficients >>  
 \*\*\*\*\*

Estimates of varying coefficients have been saved in the following file.

Listwise output file: D:\MASTER  
 THESIS\Data\Modelling\final\_joined\CSVs\AWA\_different\_buffers\combined\_bes  
 t\_indicators\_FINAL\new\_t\_diff\gwr\GWR\_2\_ind\_listwise.csv

Summary statistics for varying (Local) coefficients

Variable	Mean	STD
Intercept	3,966493	1,855875
SVF_AWA_400m	-1,850484	2,743443
NonPS_AWA_300m	0,956365	0,980413

Variable	Min	Max	Range
Intercept	1,129191	7,266308	6,137117
SVF_AWA_400m	-6,020359	2,719052	8,739411
NonPS_AWA_300m	-0,613972	2,515170	3,129143

Variable	Lwr Quartile	Median	Upr Quartile
Intercept	2,184371	4,028858	5,759363
SVF_AWA_400m	-4,111420	-2,265463	0,908025
NonPS_AWA_300m	0,038746	0,846243	1,859715

Variable	Interquartile R	Robust STD
Intercept	3,574993	2,650106
SVF_AWA_400m	5,019445	3,720863
NonPS_AWA_300m	1,820970	1,349866

(Note: Robust STD is given by (interquartile range / 1.349) )

\*\*\*\*\*  
 \*\*

GWR ANOVA Table

\*\*\*\*\*  
 \*\*

Source	SS	DF	MS	F
Global Residuals	159,484	144,000		
GWR Improvement	20,290	5,378	3,773	
GWR Residuals	139,194	138,622	1,004	
	3,757390			

\*\*\*\*\*  
 \*\*

Program terminated at 18.5.2018 r. 14:06:28