Geographically weighted Urban Heat Island modelling using the Netatmo sensors. The case of the Hague.

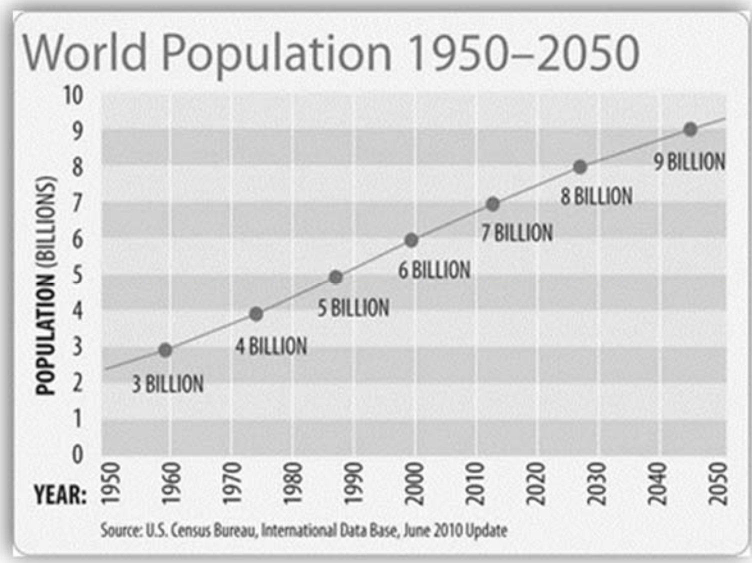
Lilia Angelova | 4620380

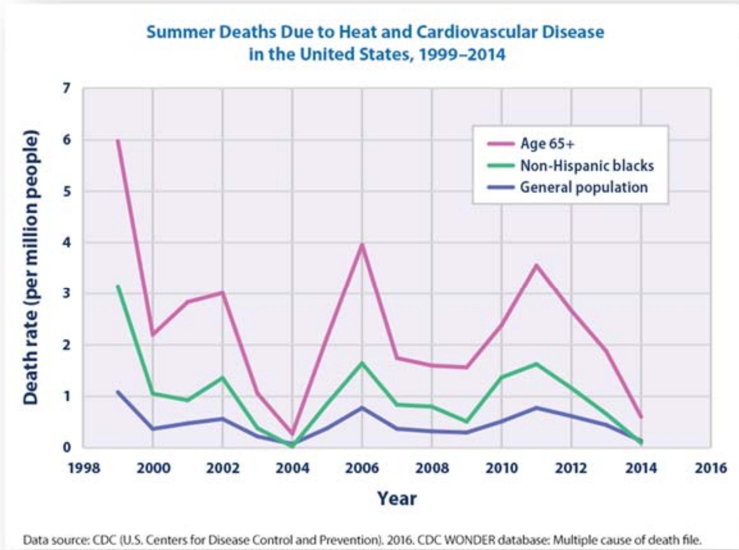
Mentor #1: Jorge Gil
Mentor #2: Alexander Wandl

Content

Motivation

- Increasing population
- More people in the cities
- Sustainable development
- Climate change
- Heat waves
- UHI leads to health-related issues, higher energy demands, economic losses





UHI types

Different characteristics

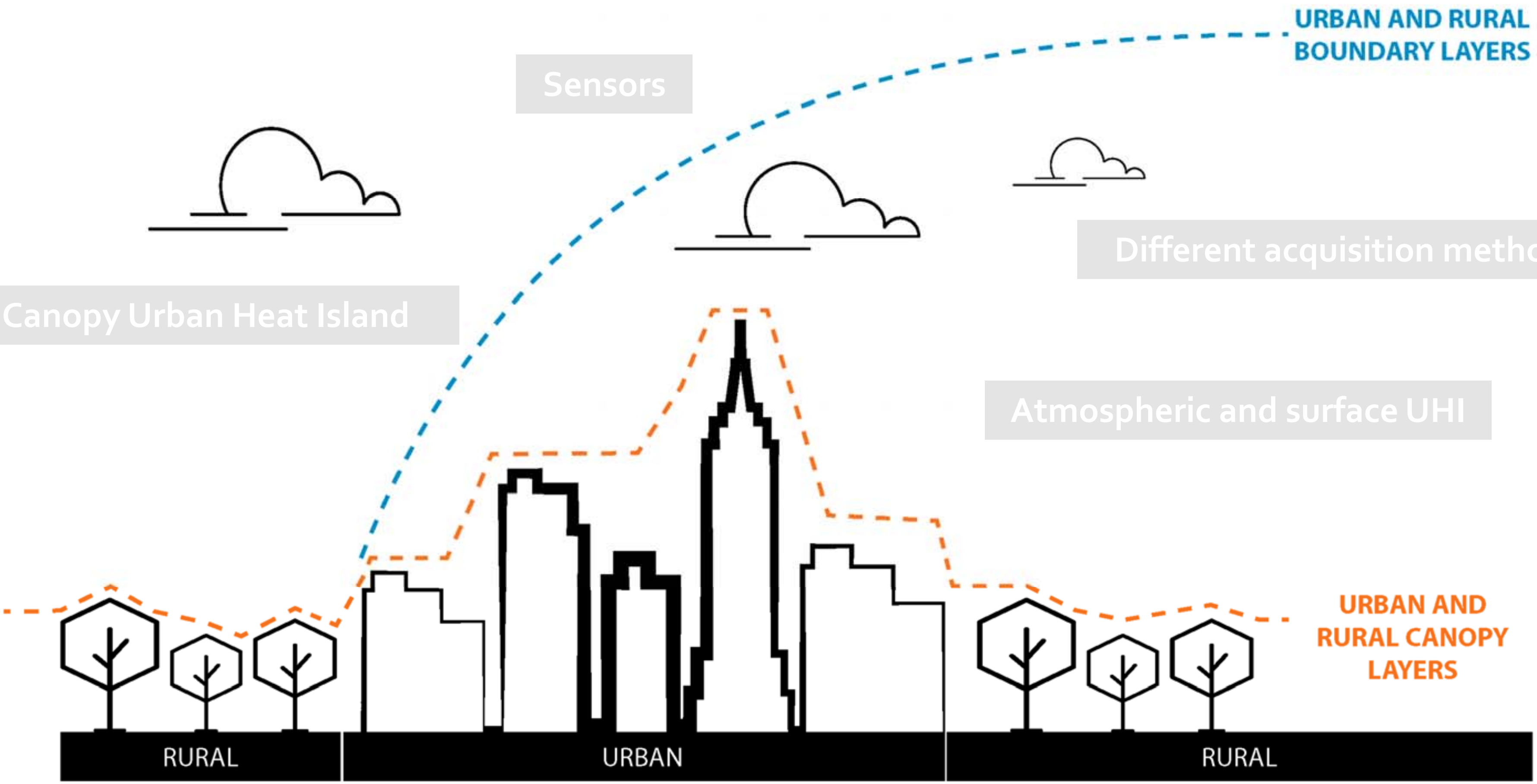
Satellites

Sensors

Different acquisition methods

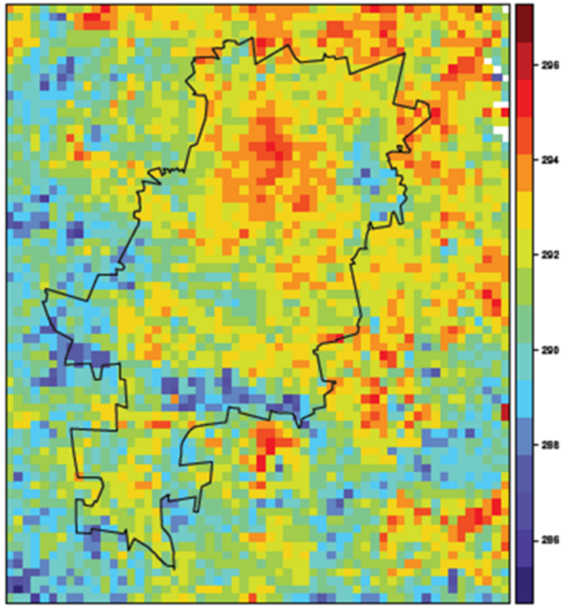
Canopy Urban Heat Island

Atmospheric and surface UHI

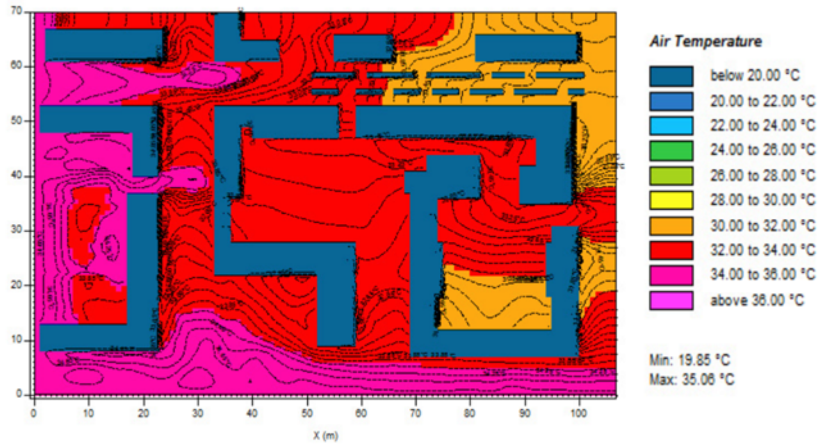


Problem

- Understanding the UHI pattern.
    - can be achieved by understanding its variability and the factors that define it.
- Therefore, we need highly detailed UHI models.
- Until now – such models exist but with limited spatial extent.
- City scale models – very coarse – don't provide info about intraurban variability.



Hardy, C. H., & Nel, A. L. (2015). Data and techniques for studying the urban heat island effect in Johannesburg.

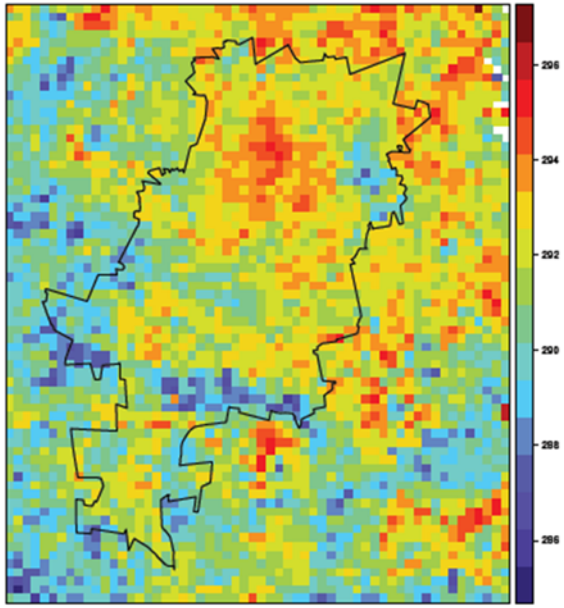


Gaspari, J., & Fabbri, K. (2017). A Study on the Use of Outdoor Microclimate Map to Address Design Solutions for Urban Regeneration.

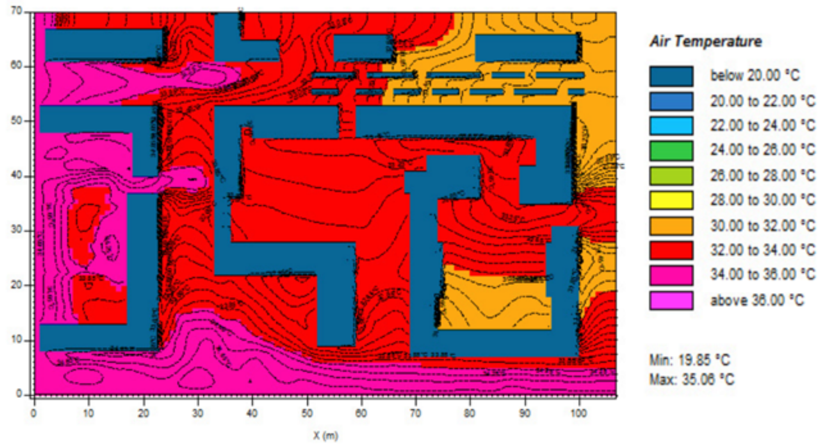
Problem

- Understanding the UHI pattern.
    - can be achieved by understanding its variability and the factors that define it.
- Therefore, we need highly detailed UHI models.
- Until now – such models exist but with limited spatial extent.
- City scale models – very coarse – don't provide info about intraurban variability.

Bridging the gap between the highly detailed small-scale models and the coarse city scale ones.



Hardy, C. H., & Nel, A. L. (2015). Data and techniques for studying the urban heat island effect in Johannesburg.



Gaspari, J., & Fabbri, K. (2017). A Study on the Use of Outdoor Microclimate Map to Address Design Solutions for Urban Regeneration.

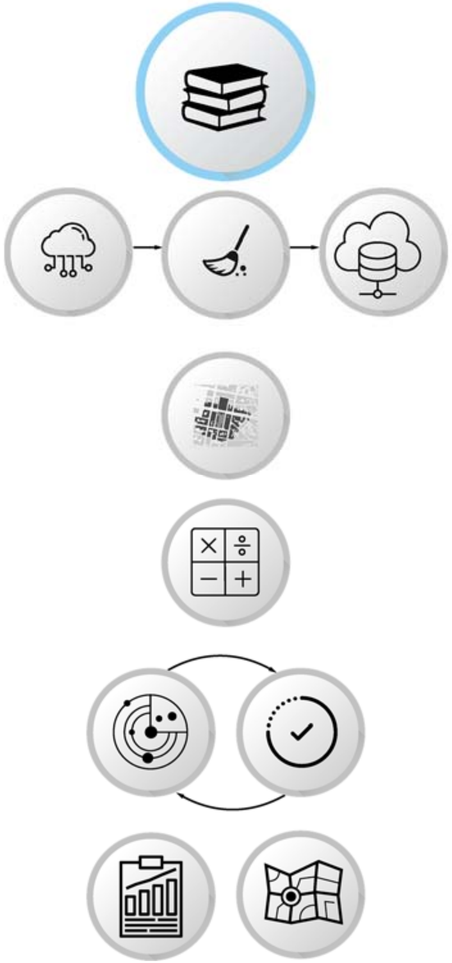
RESEARCH QUESTION

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

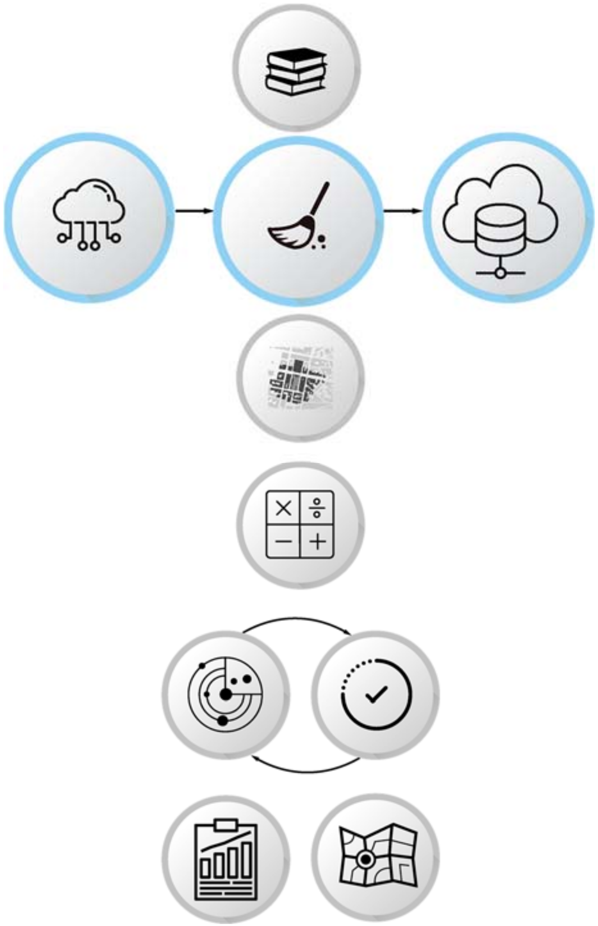
Sub questions

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?

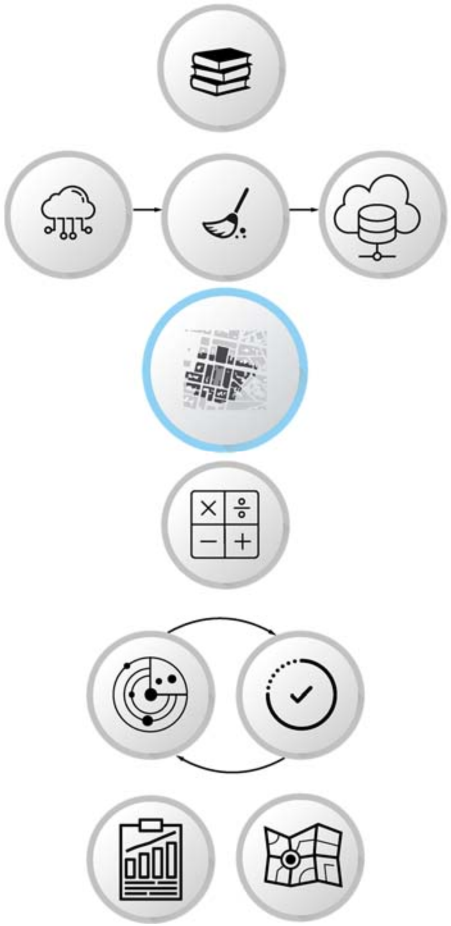
METHODOLOGY



Literature review of the current state of the art in Urban Heat Island studies and models.

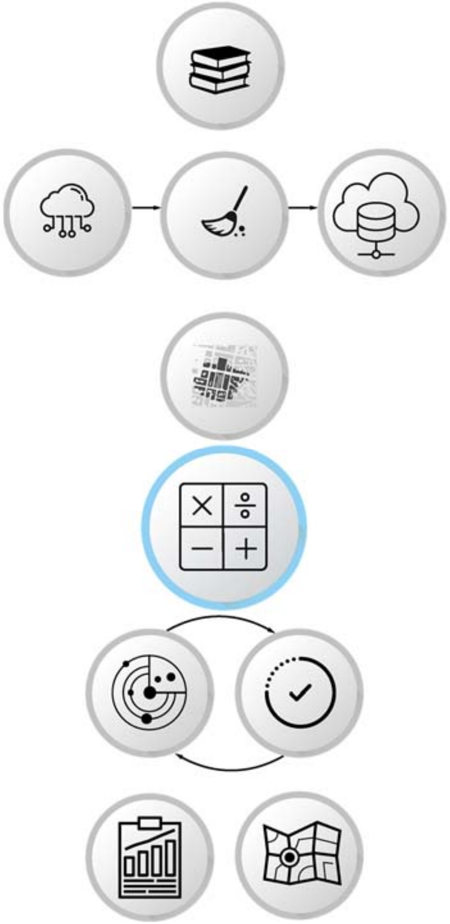


Sensor data retrieval, analysis, cleaning and storage in a spatial data base.

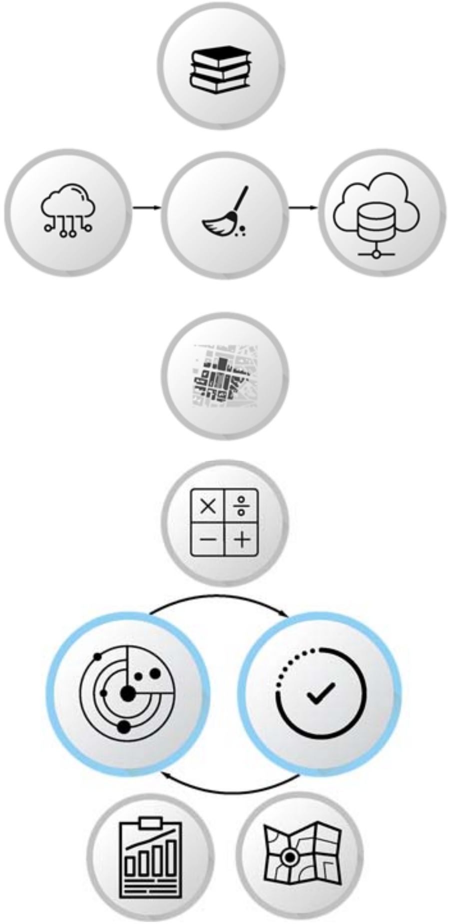


Collection, aggregation and analysis of additional relevant spatial data.

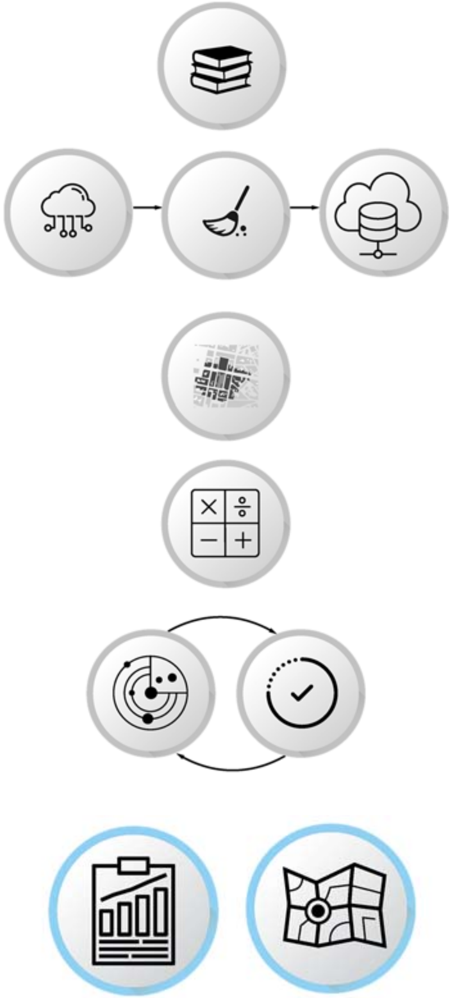
Methodology



Calculation of the spatial input parameters for the regression models.



Recursive building and validation of the alternative models.



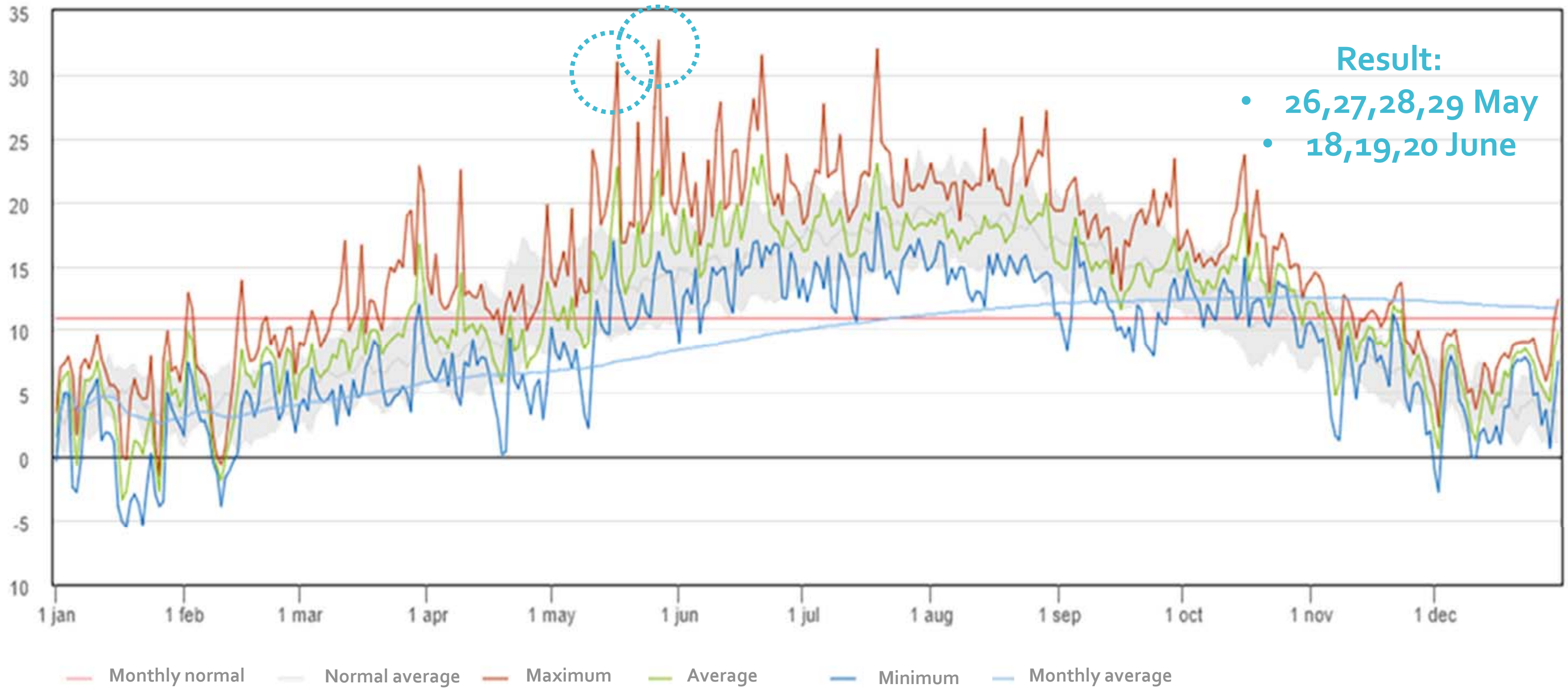
Analysis and comparison of the obtained results from the models.

SENSOR DATA

But what is a Heat Wave?

For the Netherlands **- Five consecutive** days in which the maximum temperatures exceed 25º (**at least 3 days** with temperatures higher than 30º).

Hottest days in 2017



Result:

- 26,27,28,29 May
- 18,19,20 June

Source: https://weerstatistieken.nl/hoek-van-holland/2017

Sensor data

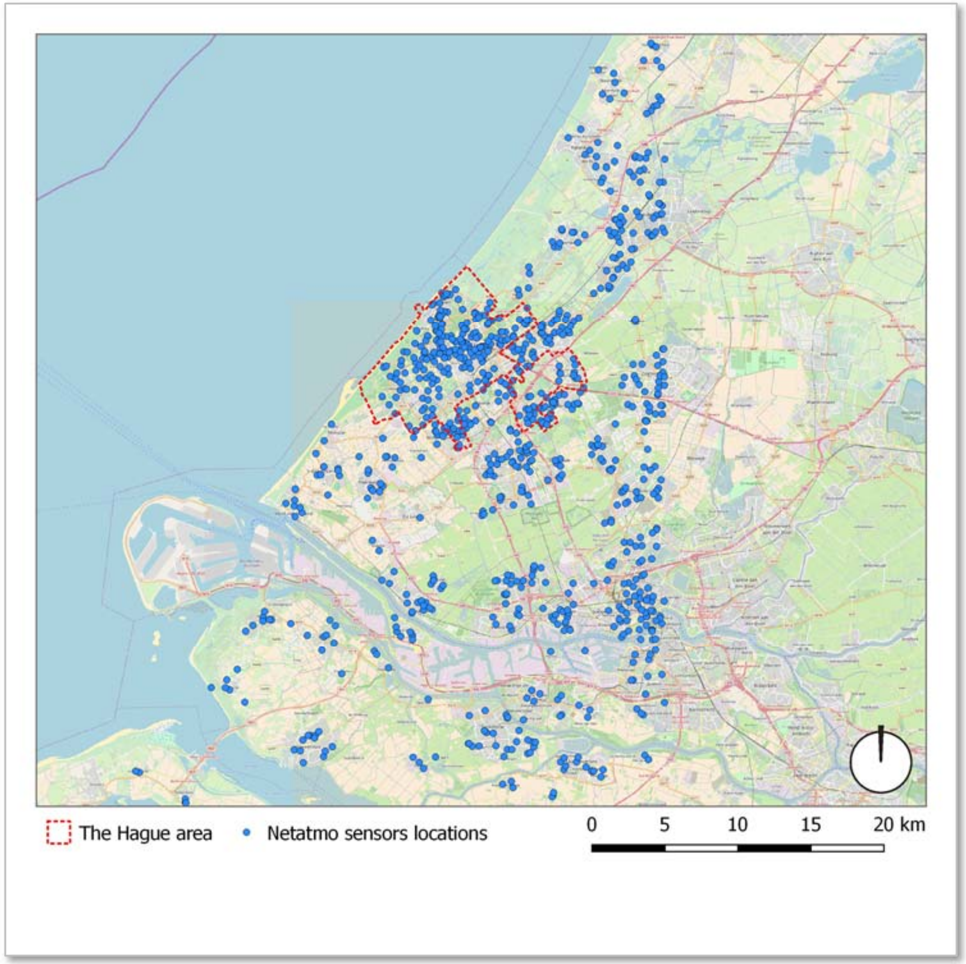
Over 800 Netatmo sensors measuring:

- Temperature
- Humidity
- Rain for the last Hour
- Rain for the last 24 Hours
- Pressure

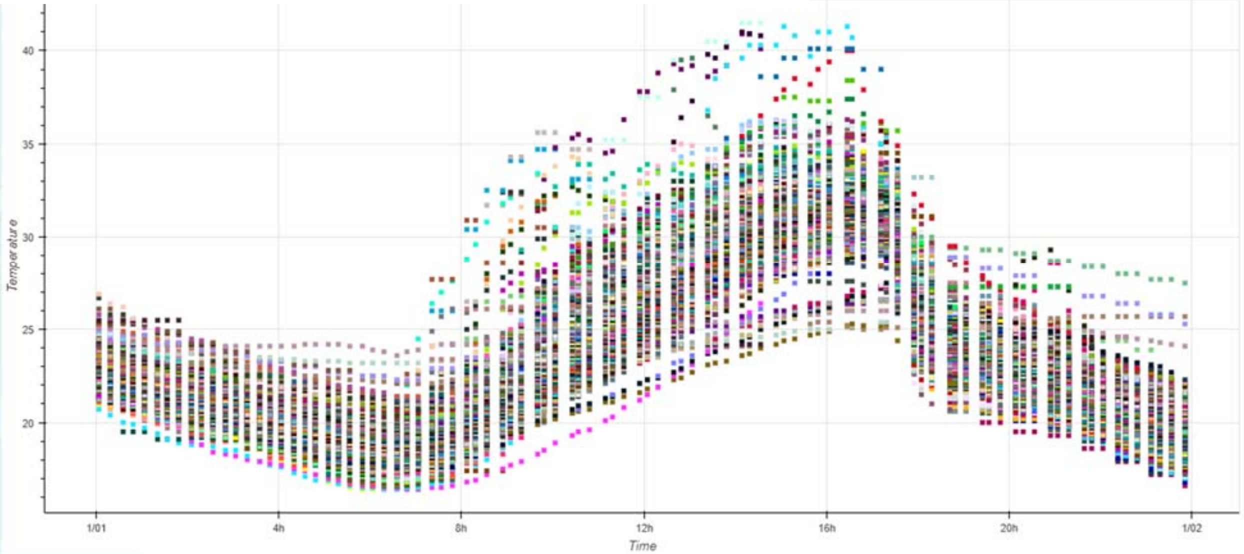
22 271 files for April – December 2017

Data Output | Explain | Messages | History

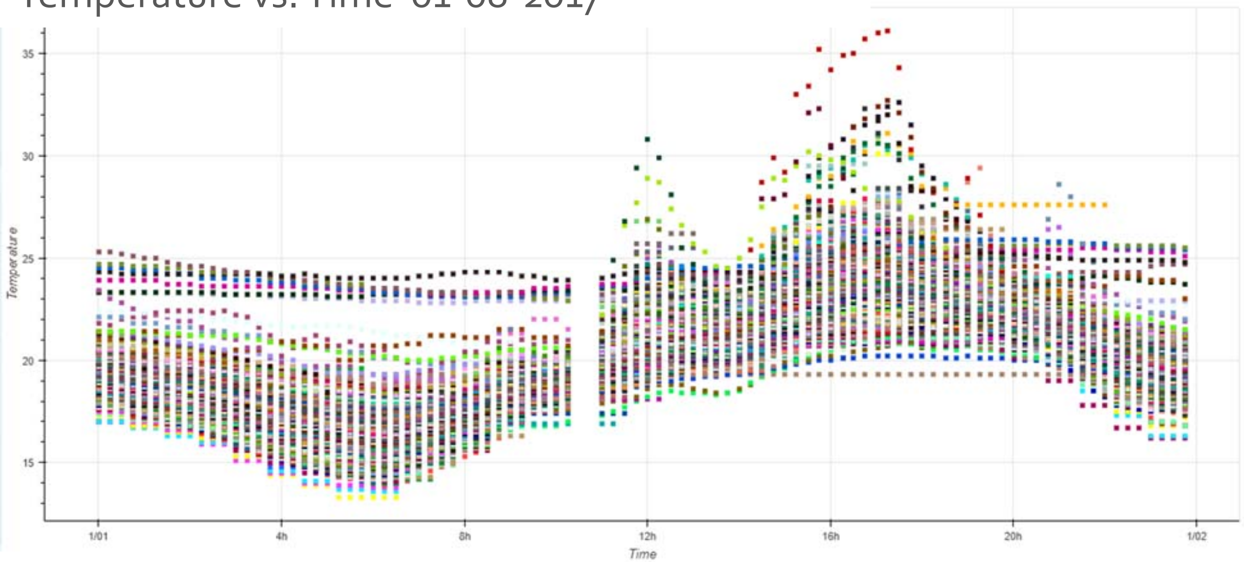
measure... [PK] dou...	sensor geometry	date date	time time wit...	mac text	altitude double p...	tempera... double p...	humidity double p...	rain_60... double p...	rain_24h double p...	pressure double p...
1	0101000...	2017-04-...	04:36:17	70:ee:50...	3	8.9	92			1020.2
2	0101000...	2017-04-...	04:36:17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

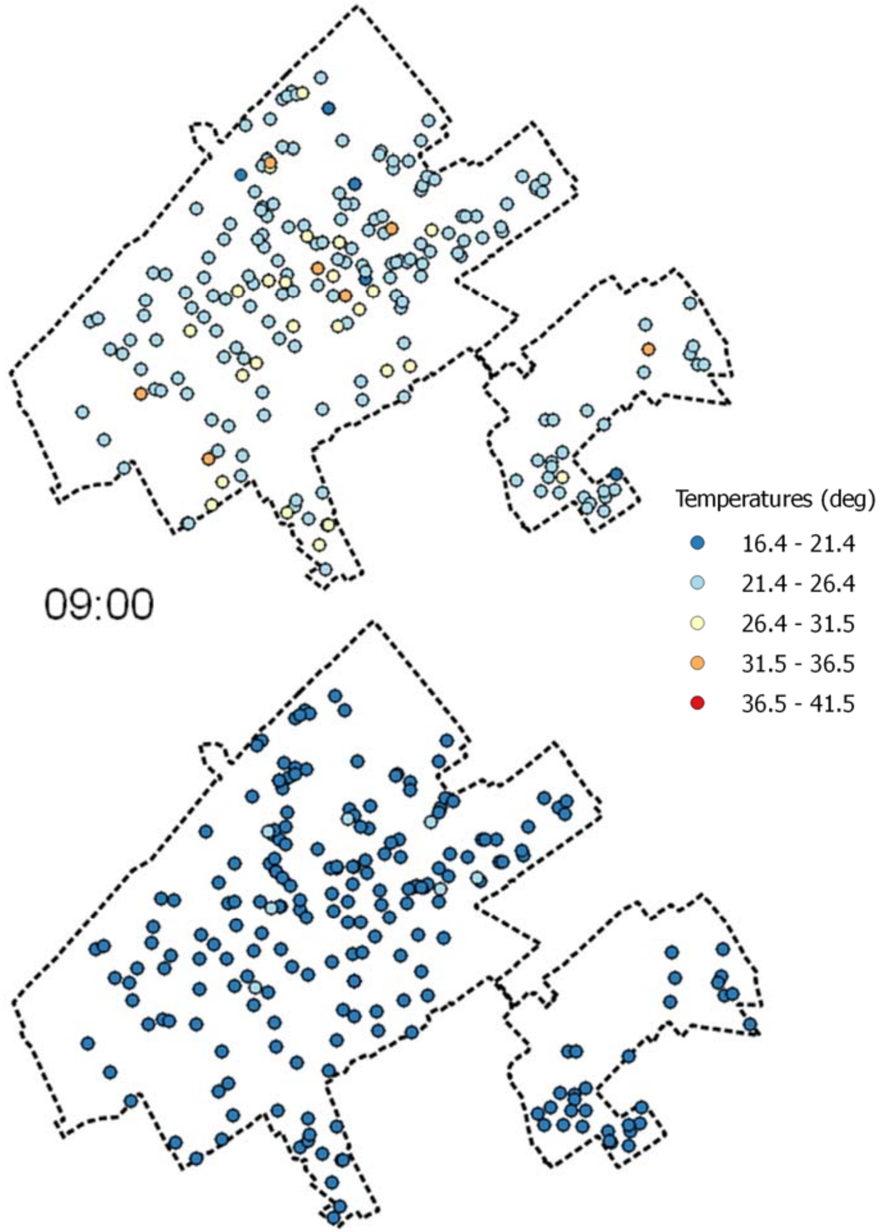


Temperature vs. Time 27-05-2017



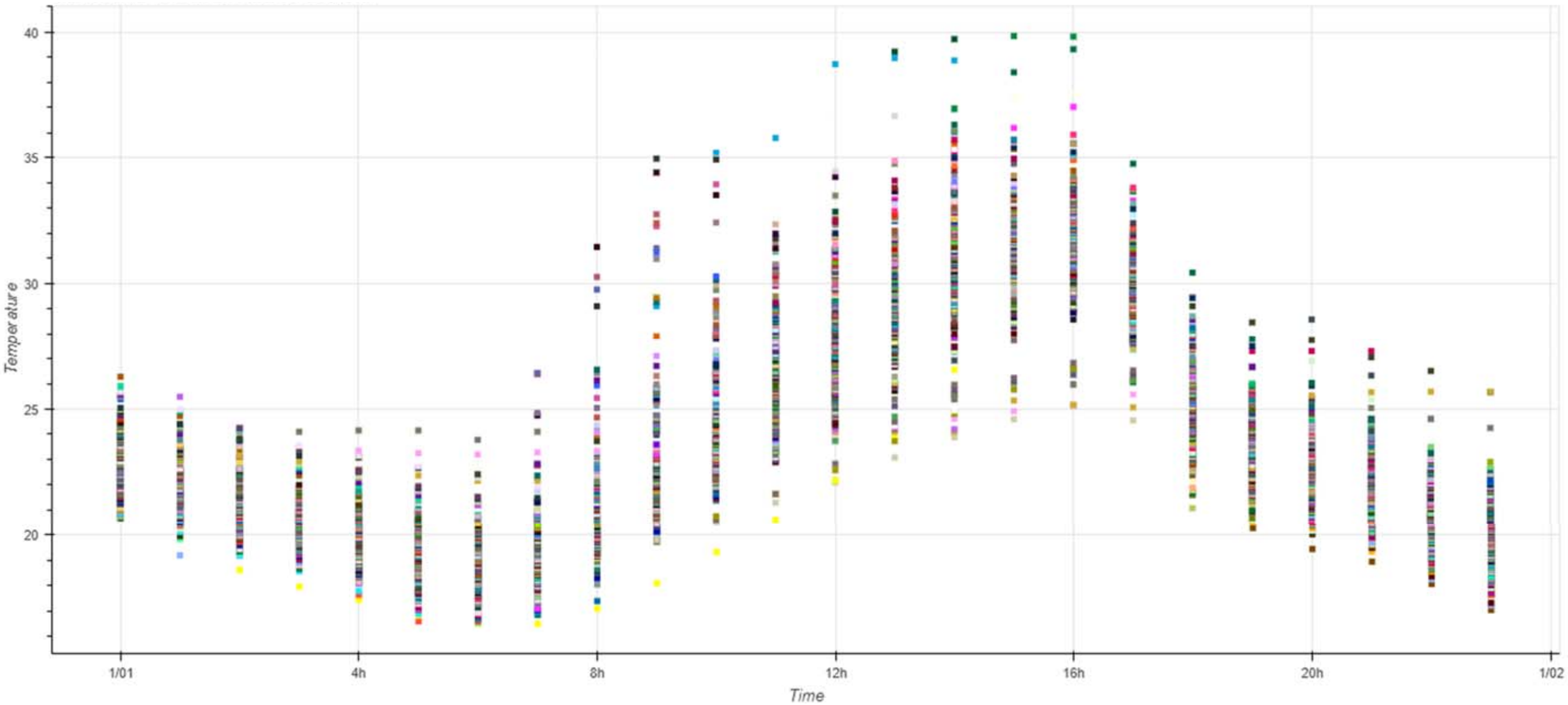
Temperature vs. Time 01-08-2017





The hottest day of 2017

Temperature vs. Time 27-05-2017 (averaged per hour)

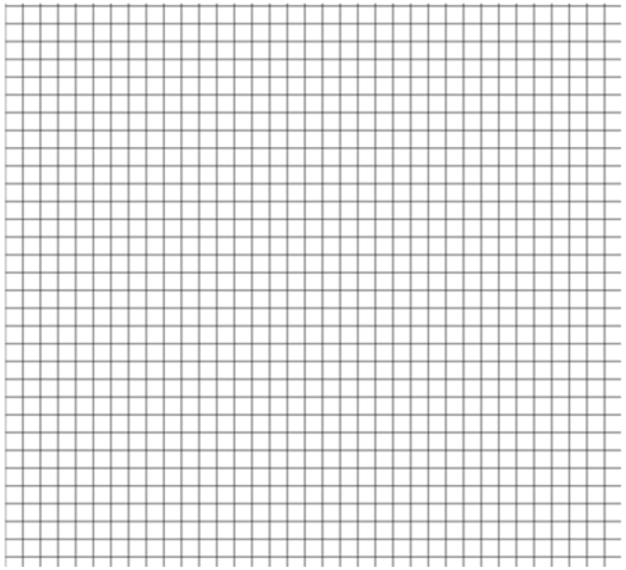


Dependent variable - the hour with the highest difference in the temperatures between the urban and rural areas – **1 am**

SPATIAL MODELS

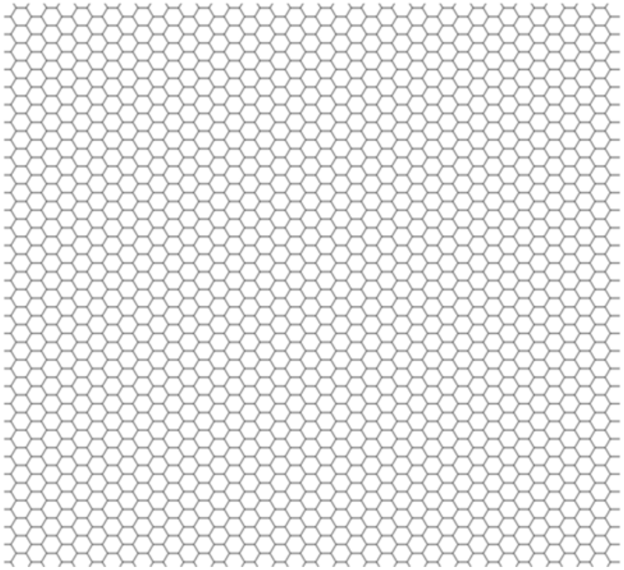
Spatial models

Rectangular grid – 9062 cells
100m side



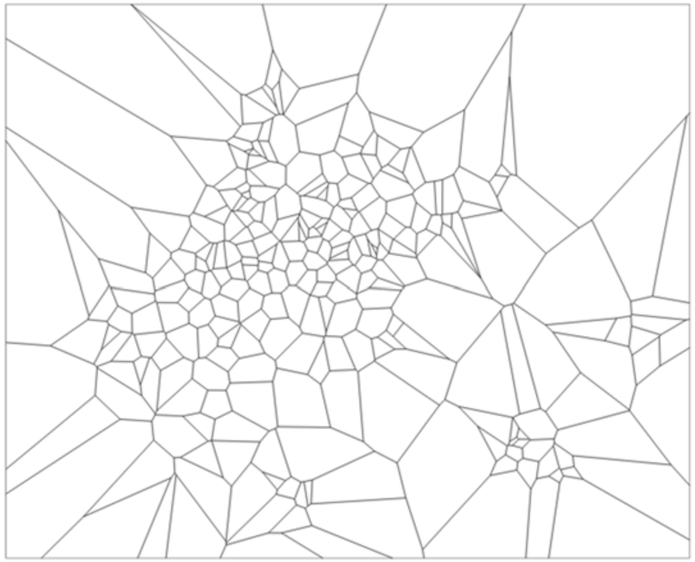
simpler definition, easily scalable,
easier raster operations

Hexagonal grid – 10359 cells
100m side distance



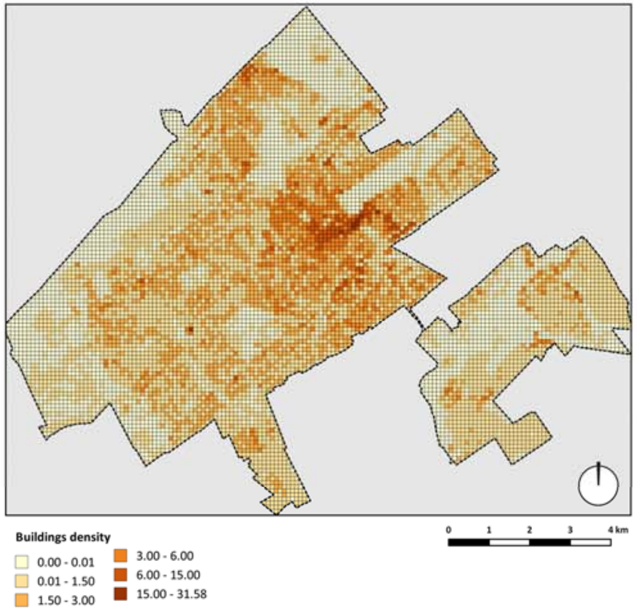
reduce edge effects, more compact,
curvature patterns of data better
visualized

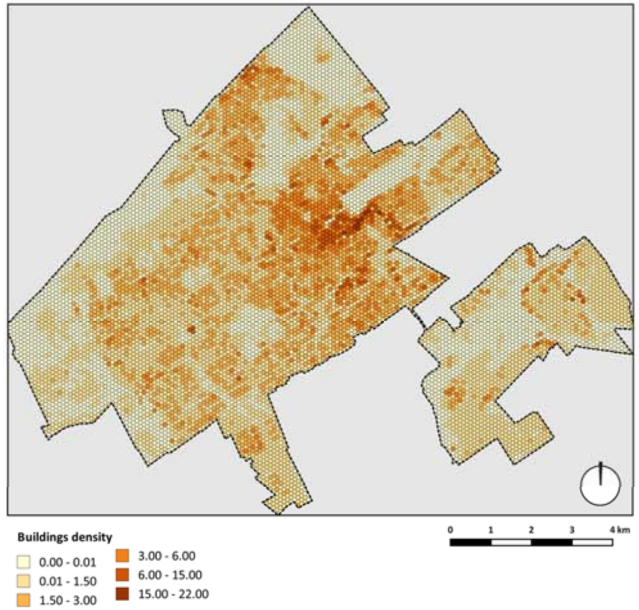
Voronoi tessellation - 273 cells
varying areas

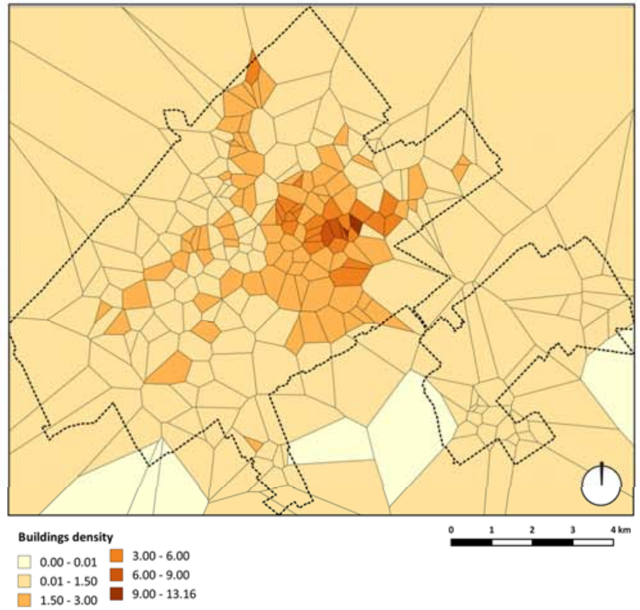


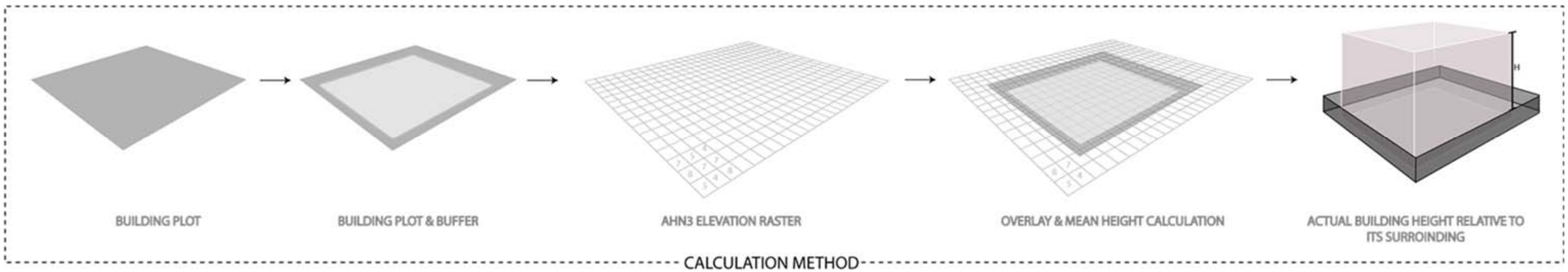
1 to 1 relation with sensor locations,
resembles data spread

Buildings density

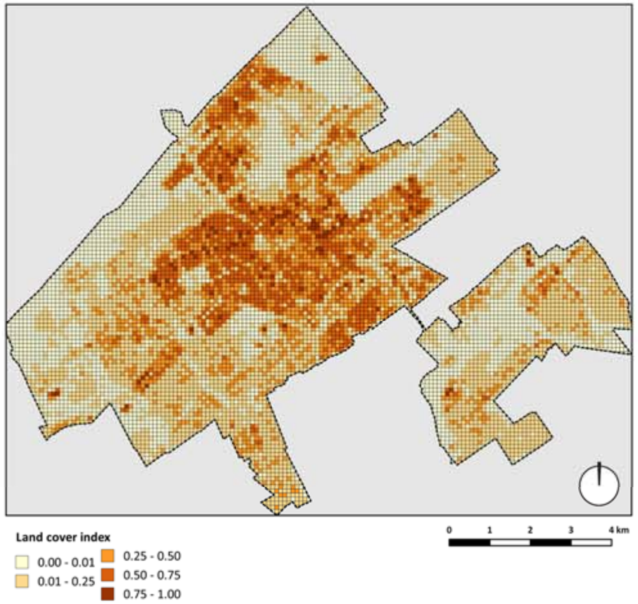


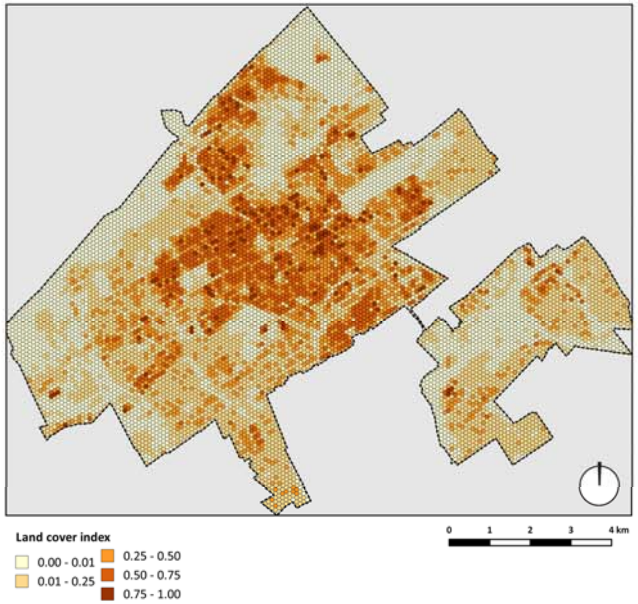


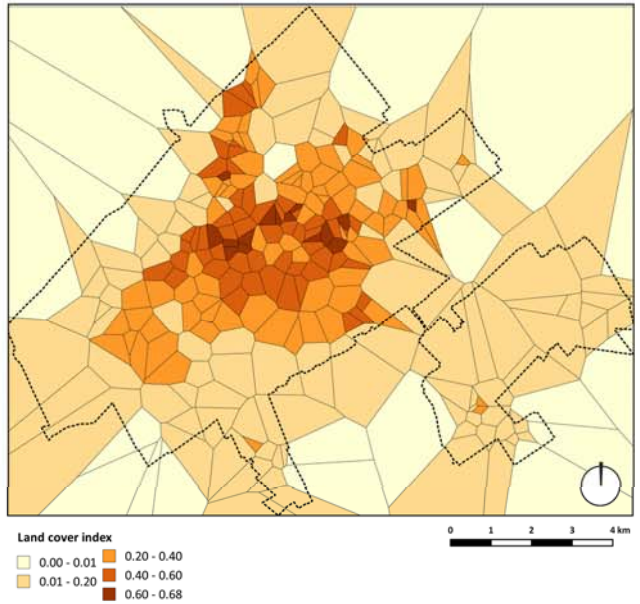


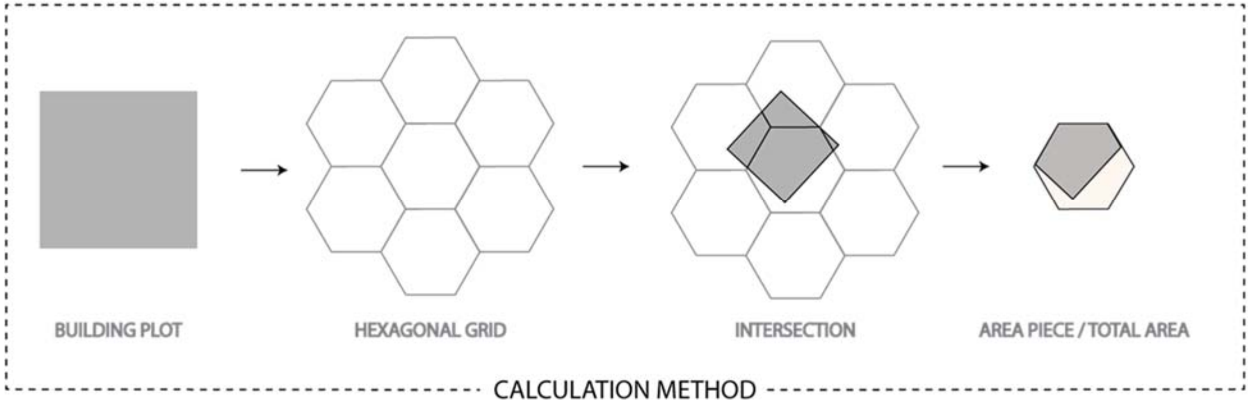


Land cover index

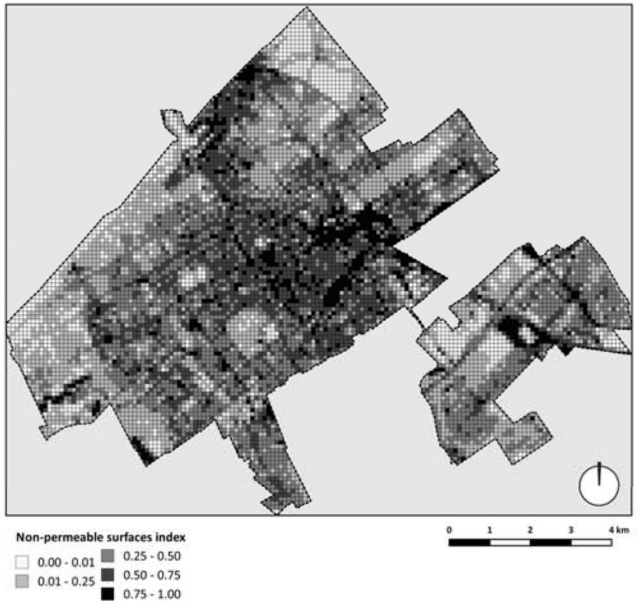


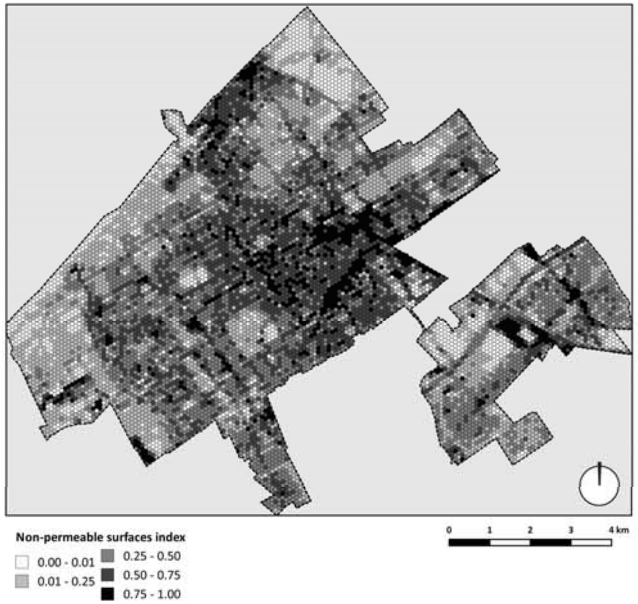


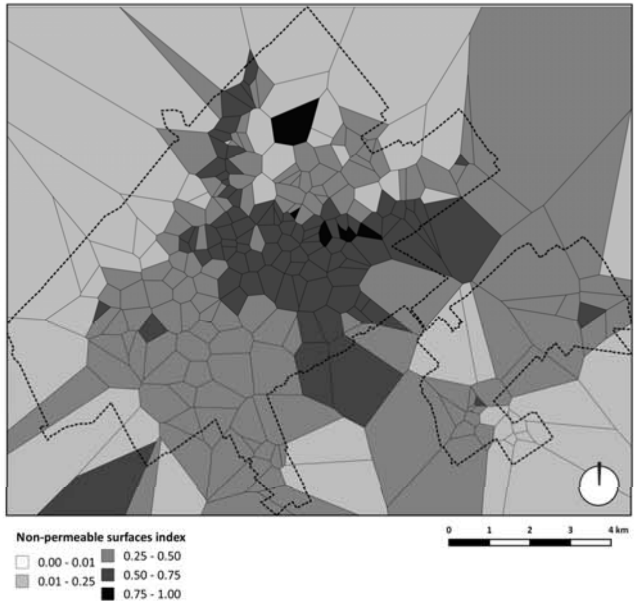


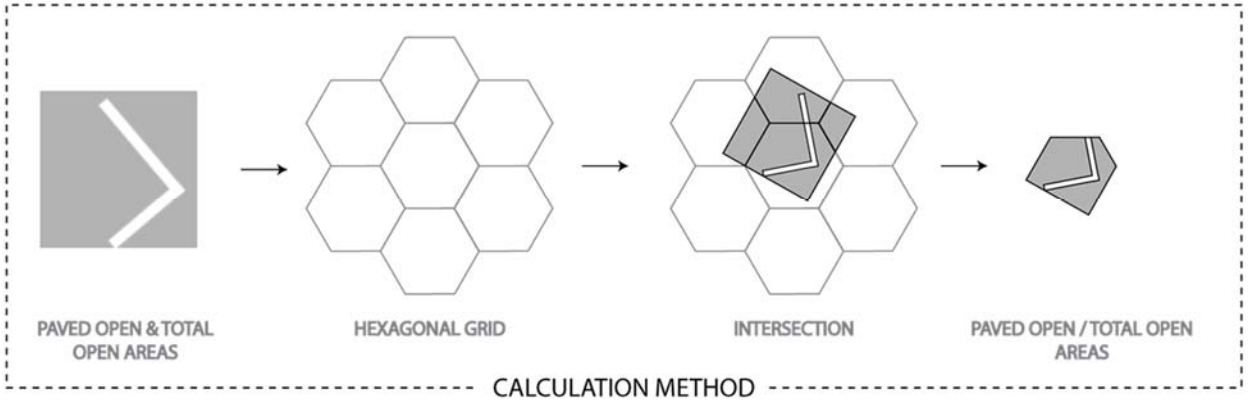


Non-permeable surfaces index

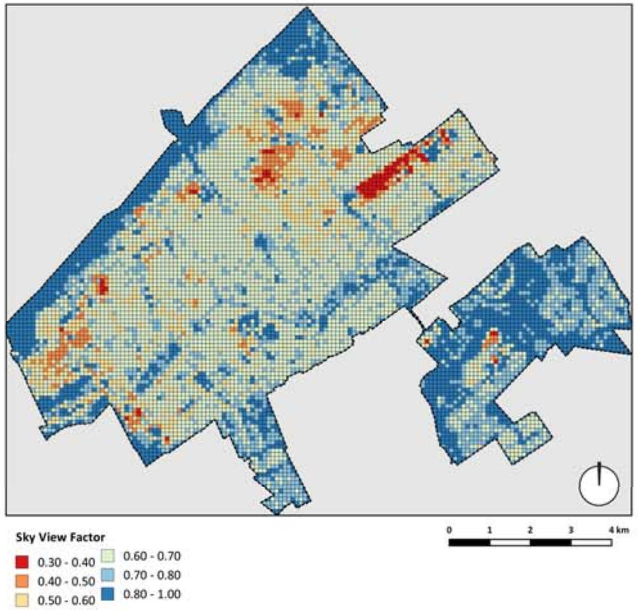


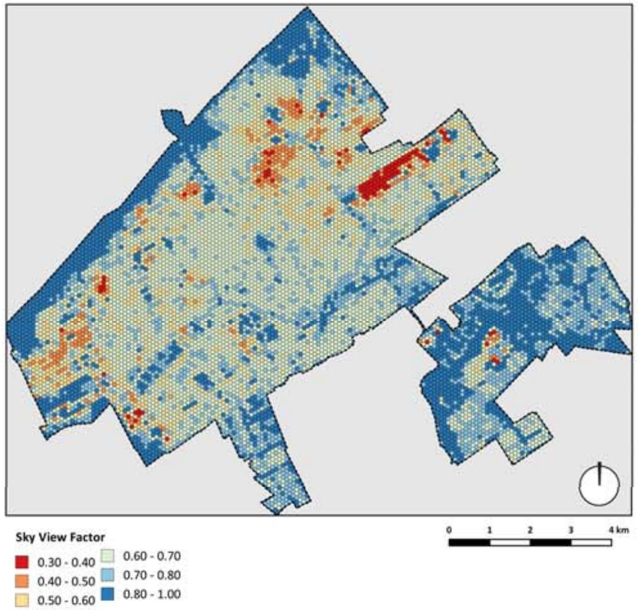


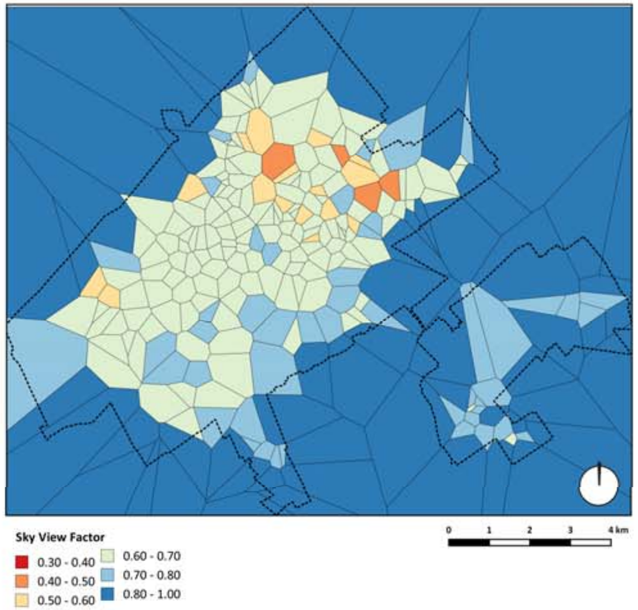


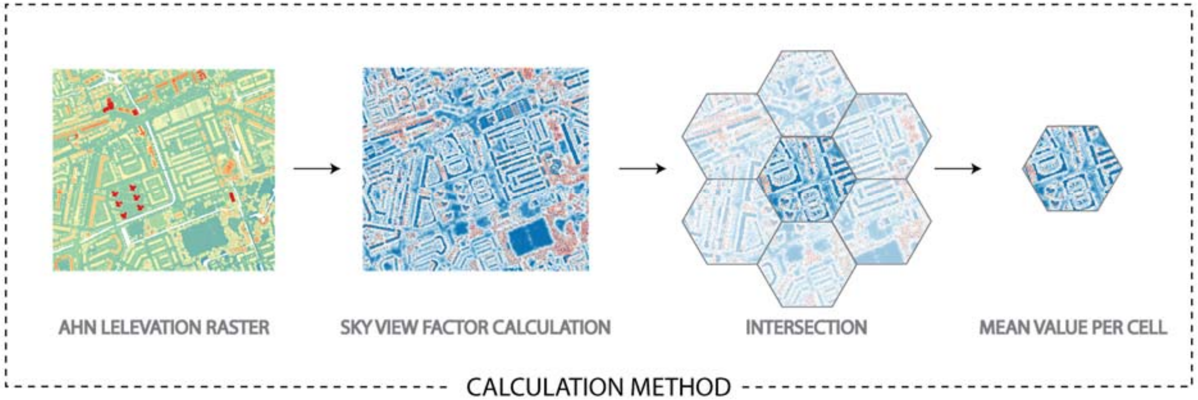


Sky View Factor

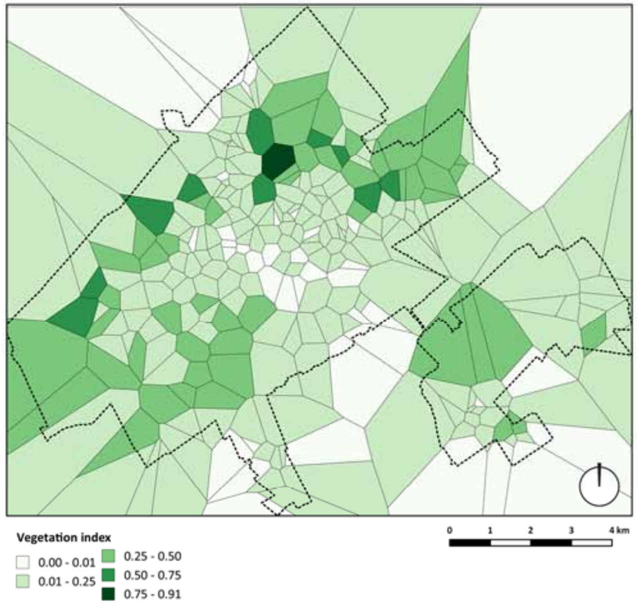


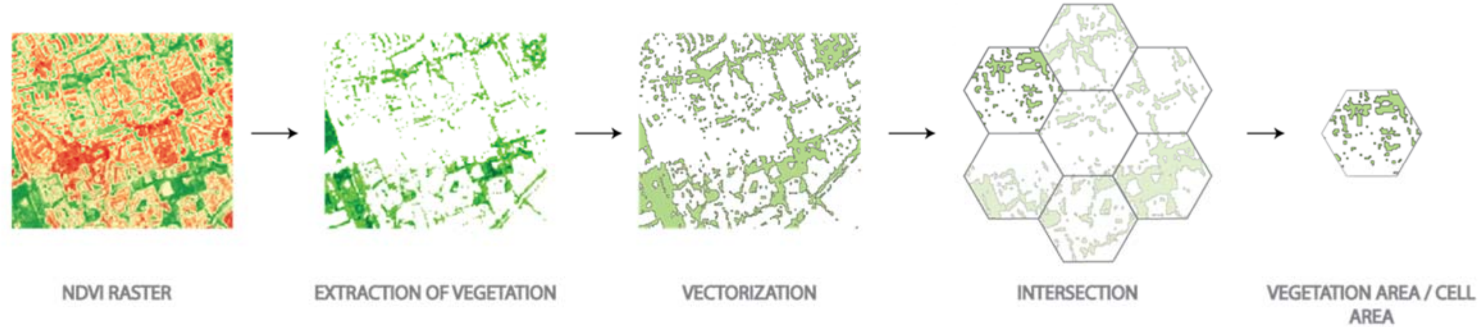






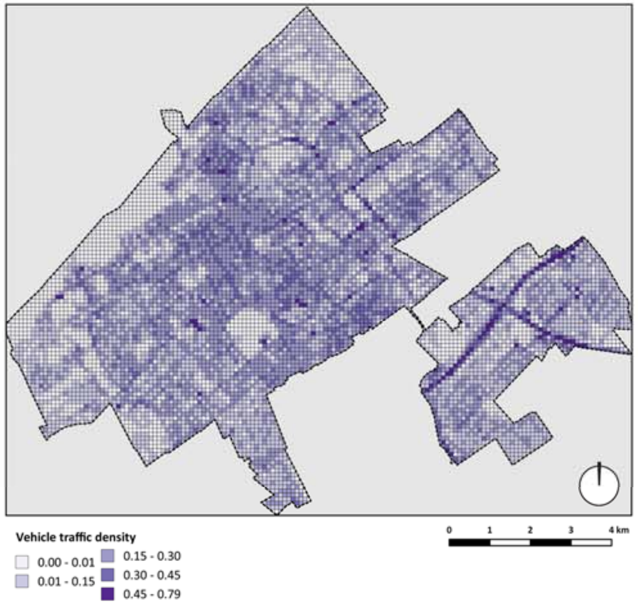
Vegetation index

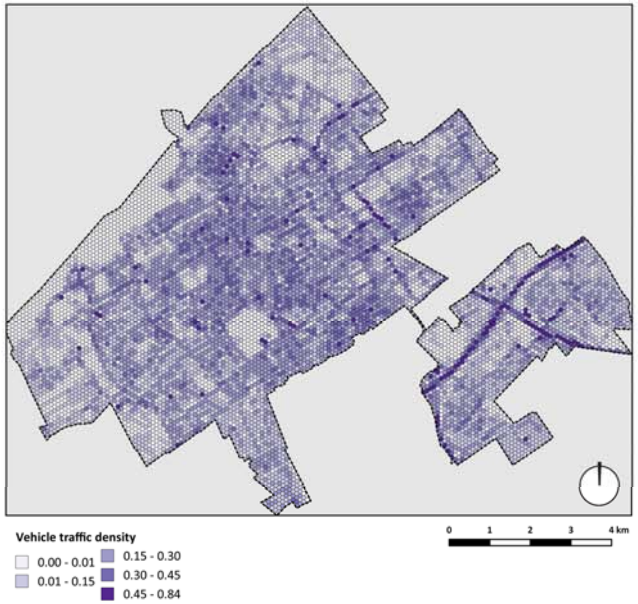


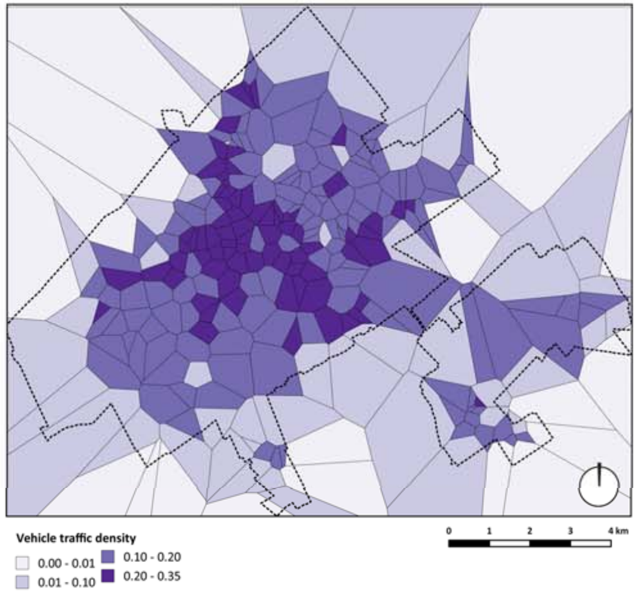


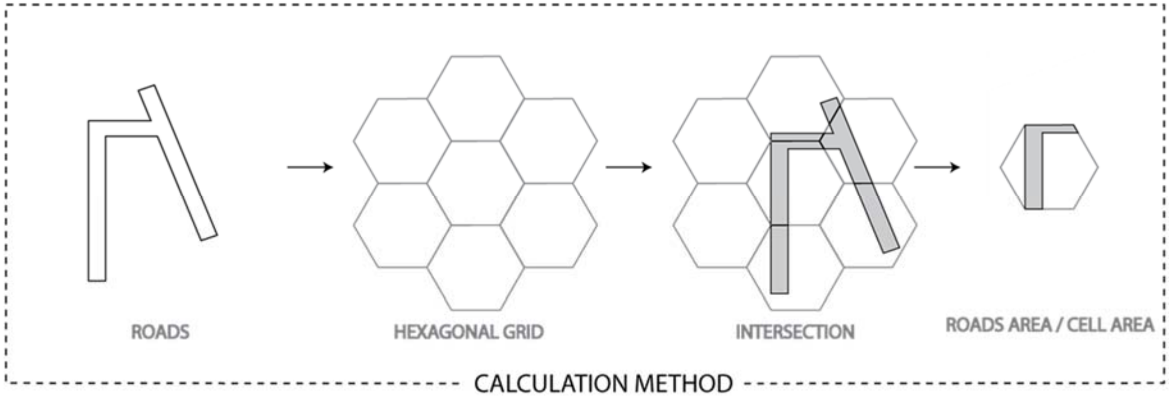
CALCULATION METHOD

Vehicle traffic density

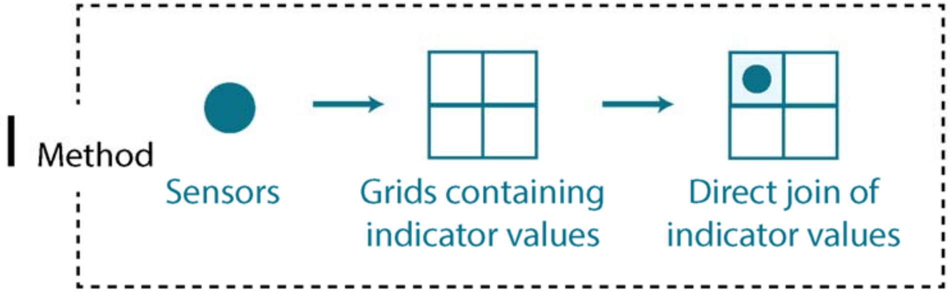




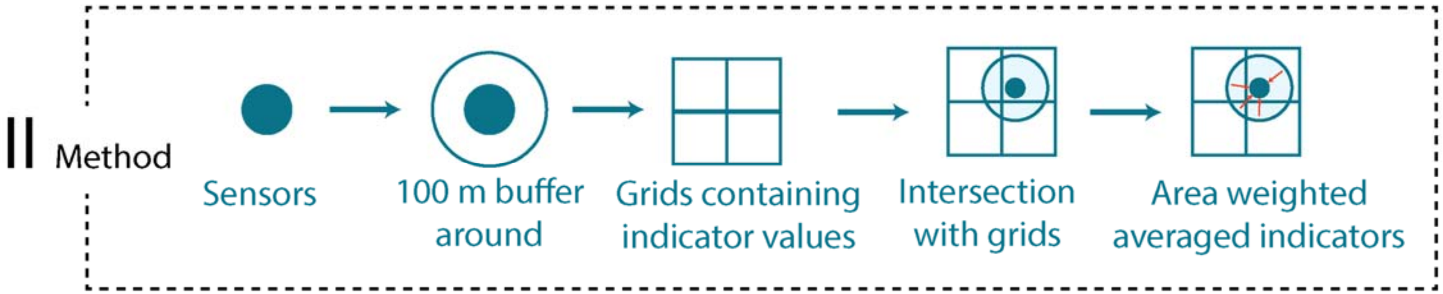




Joining



Direct join



Area weighted average

Final tables

X 3

	point_id	MAC	xcoord	ycoord	id	area	SVF	BD	LC	NonPS	Veg	VTD	Temp
1	1	70:ee:50:00:0a:3e	75452.85815	453101.00218	5640	8632.91...	0.84000	1.97000	0.13000	0.51000	0.01000	0.15000	3.70000
2	2	70:ee:50:00:38:bc	80512.73678	454281.14441	15245	8632.94...	0.73000	1.73000	0.34000	0.61000	0.04000	0.00000	2.58000
3	3	70:ee:50:00:3a:dc	81562.54547	457191.62886	17172	8632.85...	0.60000	1.22000	0.14000	0.37000	0.52000	0.20000	2.78000
4	4	70:ee:50:00:56:6c	81839.22567	454983.42010	17684	8632.94...	0.67000	8.30000	0.46000	0.99000	0.01000	0.27000	2.63000
5	5	70:ee:50:00:5d:42	79179.33785	457483.32328	12605	8632.82...	0.56000	1.64000	0.20000	0.47000	0.52000	0.15000	3.95000
6	6	70:ee:50:00:69:62	75345.43971	451174.73352	5496	8632.98...	0.78000	0.48000	0.08000	0.28000	0.09000	0.17000	3.10000

X 2

	point_id	MAC	xcoord	ycoord	SVF_AWA	BD_AWA	LC_AWA	NonPS_AWA	Veg_AWA	VTD_AWA	Temp
1	1	70:ee:50:00:0a:3e	75452.85815	453101.00218	0.71000	1.58000	0.15900	0.48000	0.15000	0.15000	3.70000
2	2	70:ee:50:00:38:bc	80512.73678	454281.14441	0.72000	1.29000	0.31800	0.68000	0.01000	0.23000	2.58000
3	3	70:ee:50:00:3a:dc	81562.54547	457191.62886	0.63000	1.46000	0.29900	0.39000	0.27000	0.24000	2.78000
4	4	70:ee:50:00:56:6c	81839.22567	454983.42010	0.66000	9.78000	0.43800	0.92000	0.01000	0.15000	2.63000
5	5	70:ee:50:00:5d:42	79179.33785	457483.32328	0.56000	1.56000	0.20900	0.36000	0.39000	0.15000	3.95000
6	6	70:ee:50:00:69:62	75345.43971	451174.73352	0.79000	0.51000	0.10400	0.40000	0.05000	0.12000	3.10000

STATISTICAL MODELING

Ordinary least squares regression – simplest form of linear model
Spatial lag - added spatially lagged dependent variable
Spatial error – added spatially autocorrelated error term
Geographically weighted regression – locally varying spatial model

$$y = \beta_0 + \beta_1 x + u \quad (OLS)$$

Dependent variable - t°

Independent variables – Spatial indicators

$$y = \rho W y + x + \varepsilon, \ \textit{(Spatial lag model)}$$

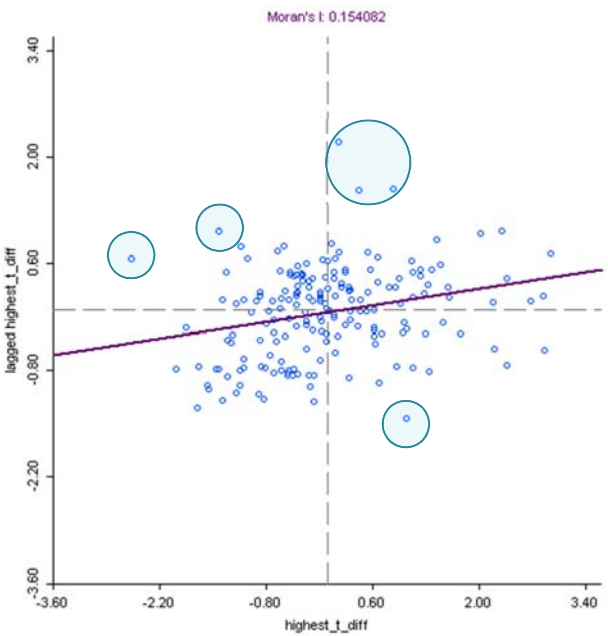
$$y = X\beta + \varepsilon, \ \textit{(Spatial error model)}$$

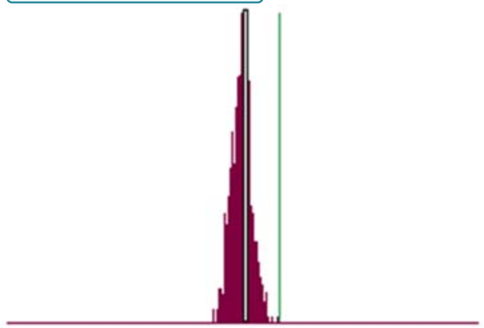
$$y_i = a_{i0} + \sum_{k=1,m} a_{ik} x_{ik} + \epsilon_i, \ (GWR)$$

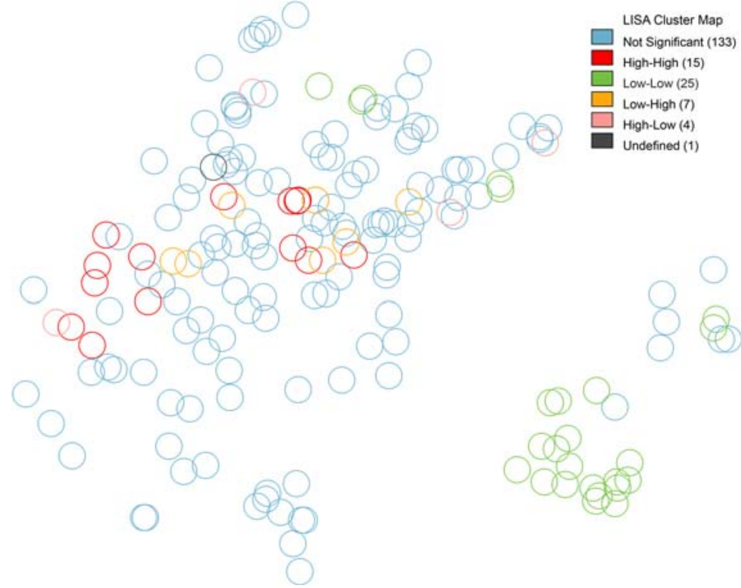
1) Random data sample representing the general population.
2) The random error term has normal distribution (no systematic misspecification or bias in the model).
3) The errors have constant variance (homoscedasticity).
4) The individual observations are independent.

The violation of any of these assumptions can lead to bias, inefficiency in the regression estimates or unreliability of the confidence intervals.

Spatial autocorrelation



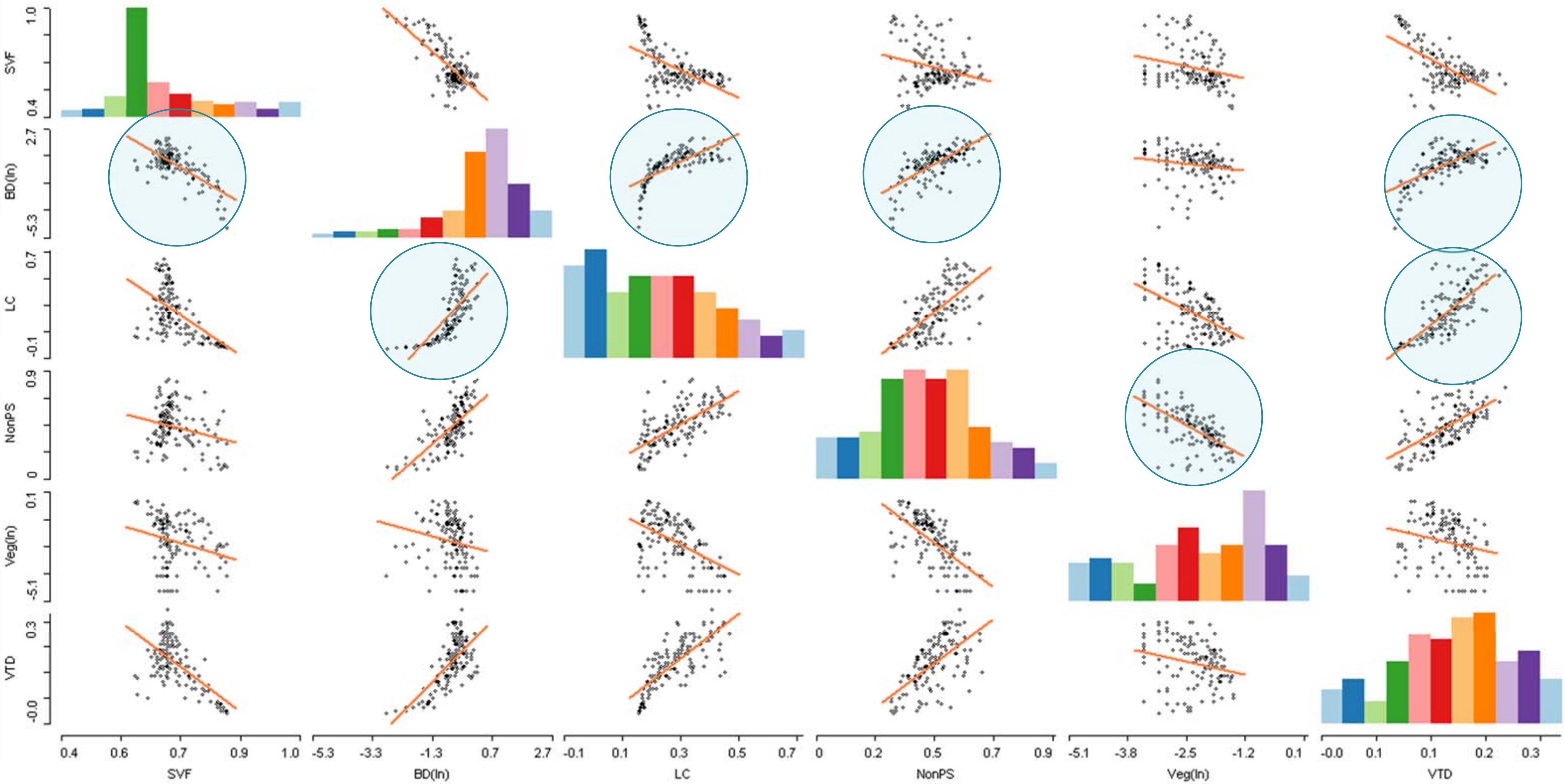




Local indicators of spatial association (LISA) - Cluster map

LISA – significance map

Multicollinearity



Correlated indicators:
BD with SVF, LC, NonPS, VTD; Land cover –BD, VTD; NonPS - Vegetation

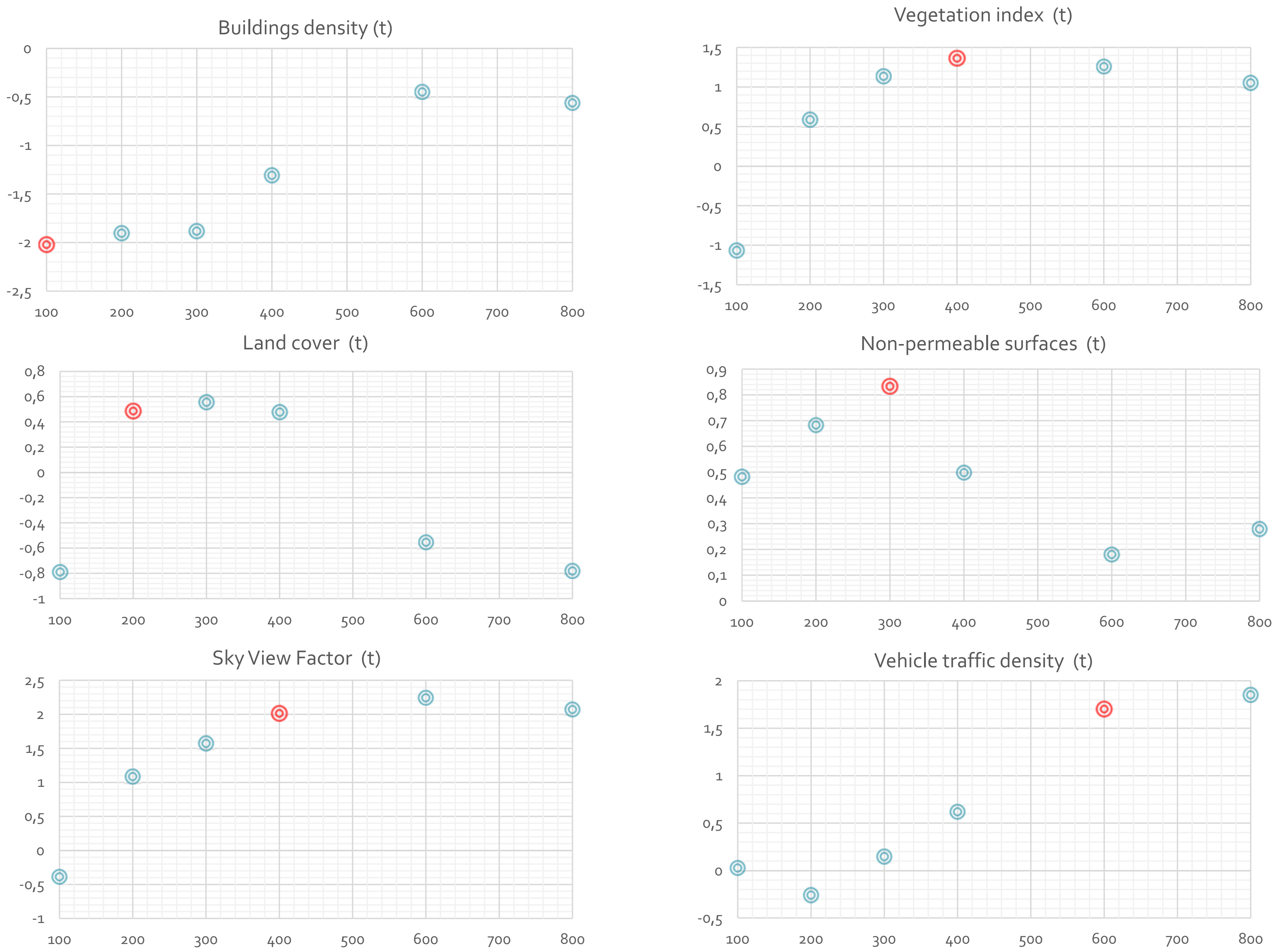
Initial models

Goodness of fit indicators

MODEL	R^2	R^2 - ADJUSTED	AIC	LOG -LIKELIHOOD	MULTICOLL INEARITY CONDITION NUMBER
HEX (DIRECT)	0.03	-0.012	415.682	-200.841	45.70
RECT (DIRECT)	0.03	-0.003	414.425	-200.213	47.92
VORONOI	0.04	0.002	413.592	-199.796	71.12
HEX (AWA)	0.04	0.004	413.283	-199.642	62.81
RECT (AWA)	0.04	0.005	413.051	-199.526	62.50

Variables 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	1.7952	-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	-1.4100
P-VALUE	0.3059	0.7651	0.7412	0.2699	0.5960	0.1608
RECTANGULAR GRID (AWA)						
COEFFICIENT	-2.8955	-0.0784	0.1512	2.0012	-0.0317	-3.4379
T-STATISTIC	-1.3974	-0.9200	0.1687	1.7872	-0.0261	-1.4907
P-VALUE	0.1646	0.3592	0.8662	0.0761	0.9792	0.1384
HEXAGONAL GRID (AWA)						
COEFFICIENT	-2.4328	-0.0751	0.3085	1.9227	0.1071	-3.6976
T-STATISTIC	-1.1616	-0.8635	0.3567	1.7670	0.0898	-1.6608
P-VALUE	0.24747	0.3893	0.7218	0.0795	0.9285	0.0991



Statistical significance against buffer sizes

Improved models by:

- Different areas of influence included
- Removed insignificant variables
- Excluded correlated variables
- Applied spatial models

Geographically weighted regression has the highest performance.

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

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

What?

Model predicting **extreme temperatures in the city**; primarily based on **spatial data**, thus not dependent on weather data.

How?

Based on **spatial and statistical techniques for data modelling**; using **spatial indicators** as independent variables; using Geographically weighted regression model.

Why?

Tool, assisting **different planning strategies** and the development of different **UHI mitigation measurements**; different experts can address the UHI problem by **evaluating the significance of the different spatial indicators** for the temperatures in the city.

Future work

Currently the model has limited explanatory power, because of:

1) The inconsistent nature of the air – e.g. wind.
2) The errors in the sensor data – position, technical characteristics.

Improvement of the model:

1) Inclusion of more diverse indicators.
2) Improvement of the calculation methods for these indicators – e.g. Vehicle traffic density, Vegetation.
3) Statistical analysis of the variables in the selection process.
4) Exploration of different non-linear models.

THANK YOU