

## ***Comparing automated methods for identifying areas of critical heat demand in urban space***

*Hannes Seller*

Hannes Seller M.Sc., HafenCity Universität Hamburg, Überseeallee 16, 20457 Hamburg, [hannes.seller@hcu-hamburg.de](mailto:hannes.seller@hcu-hamburg.de)

### **1 ABSTRACT**

In recent years, urban heat supply has shifted to the center of attention of German energy policy. It is believed that heating grids are an important instrument for climate protection. For one, they open up a heat sink (i.e. a circle of heat customers) large enough to be able to take up heat from cogeneration, which needs a certain minimum scale of operation to be economically viable. Secondly, they allow the relatively easy tying-in of renewable energy sources.

However, heating grids are not the one-fits-all solution. As heat transport is associated with losses, a minimum heat density in urban space (that is: MWh per hectare urban space) is needed to make a district heating grid lucrative (and, possibly, ecologically worthwhile – depending on the source of the heat). At the same time, given the nature of the heat generator, a larger area served may offer economies of scale.

Opportunities to construct small and medium-sized grids often are overlooked, as information about critical parameters like heat density in a neighborhood are not obvious to potential initiators of such grids.

This paper offers a comparison of methods to systematically search an urban heat demand map for areas of critical heat density. Urban heat demand maps are now developed by many municipalities; they are usually constructed using electronic cadastre data, combined with an energetic building typology into which the buildings in the cadastre are mapped. Some potentially interesting opportunities for developing district heating grids may be visible to the experienced eye; algorithms that automatically search over the entire heat map may offer yet more insights. As algorithms I apply (1) a tessellation of the city into tiles of comparable size, and (2) a clustering method used to identify hot spots with two different approaches. I use selected neighborhoods in Hamburg to compare the results of both methods.

Keywords: District Heating, Urban Heat Demand, Energy GIS, Energy Planning

### **2 INTRODUCTION**

This paper's goal is to examine the effectiveness of different algorithms designed for the purpose of finding heat densities in urban areas. It reflects on the work done for the GEWISS project (*GEographisches WärmeInformations- und SimulationsSystem Hamburg*, verbatim: Geographical Heat Information and Simulation System Hamburg). GEWISS is a research project within the *EnEff:Stadt* (EnEff:City) programme funded by the German Federal ministry for Economy and Energy (BMWi). It strives to provide an interactive information tool on the spatial distribution of heat demand and supply in Hamburg, as it is today and could develop over time (Peters 2015).

Space heating is Germany's second largest CO<sub>2</sub> emitter (Peters 2015). New and energy-efficient constructions make up only a small percentage of the German building stock (around 1 % each year). A high-leverage option to reduce CO<sub>2</sub> emissions is therefore the retrofitting of the existing building stock. For an effective retrofitting strategy we need to acquire knowledge about the patterns of urban heat flows. CO<sub>2</sub> emission reduction measures should focus on urban areas with high heat density and promote (1) a reduction in primary energy demand and (2) an increase in the share of renewables in heat supply – this without additional environmental burden. The algorithms tested for this paper are primarily designed to identify urban areas where combined heat power plants seem particularly lucrative.

Automated algorithms become handy as nowadays we tend to have more data available than we can assess without the use of electronic devices. Although they do not make the urban planner superfluous, their application helps them to spend their time more efficiently. The algorithms presented in this paper are all meant to be visual and statistical aids to find areas of interest faster. They do not provide a result that could stand without human interpretation. The algorithms discussed in this paper are specifically programmed to

work with a minimum input of data to ensure that planners can apply them to their specific urban context. The required attributes--geographic coordinates, floor numbers and footprint sizes--surely exist in the electronical cadastre (ALKIS) of each city. The mapping of construction years is at least in Hamburg-- not mandatory and therefore incomplete, I therefore need to reduce the building stock to buildings with known construction years.

To test the algorithms applicability I chose two urban neighborhoods in Hamburg. After describing the neighborhood selection criteria I will explain the algorithms' working and assess their overall usefulness to determine urban heat densities.

### 3 NEIGHBORHOOD SELECTION

For the paper I selected two urban areas for which information about building stock--primarily the construction periods--is sufficient to estimate an annual heat demand using a building typology. The geographical data for this paper was provided by the Transparenzportal Hamburg, which is maintained by the city of Hamburg and can be accessed by everyone. As building typology I used the commonly used IWU-de (*Institut für Wohnen und Umwelt*, verbatim: Institute for Housing and Environment) to estimate the heat demand for individual buildings. Further data sources are not required.

Hamburg's building stock, as represented in the ALKIS, consists of roughly 372,000 buildings. I used the attribute *Gebäudefunktion* (GFK, building use), coded as four digits, to identify 220,000 residential buildings. This includes all buildings that have a GFK with the pattern "1XXX", excluding garden houses (1313) as they only have seasonal use and usually do not have a heat demand. Only for half of the building stock (108,000 buildings) construction years are known, which are paramount to determine building types and thus heat demands. The documentation for modelling the geodatabase reveals that the mapping of construction years is not mandatory for a complete data set (AdV 2015, p.245). I will therefore exclude all residential buildings without known construction years from the analysis. This of course implies that results will not capture urban heat flows accurately. For the purpose of testing and comparing algorithms, however, the reduced data set suffices.

The IWU typology primarily takes two attributes into consideration to determine a building's heat demand: (1) its size class and (2) its construction age class (*Größenklasse* and *Baualtersklasse*, IWU 2015, p.9). The age classes reflect historical and architectural epochs, like the Wilhelminian epoche (1860-1918) or the reconstruction phase after World War II (1949-1957). The size class depends on architectural characteristics, primarily the number of floors and floor area. For the purpose of this paper--its automated applicability to a set of geographical building stock data--I chose a simplistic approach that only takes the buildings' number of floors into consideration (see Table 1). Inarguably, one family houses with more than one floor exist, but since the average floor number in Hamburg is already rather small<sup>1</sup> in comparison to other cities, I strive for more diversity in size classes to test the algorithms.

Table 1: IWU size classes determined by number of building floors (own representation)

Größenklasse (Size class)	Number of floors
Einfamilienhaus (EFH, one-family house)	1
Reihenhaus (RH, row house)	2 or 3
Mehrfamilienhaus (MFH, multi-family house)	4 or 5
Großes Mehrfamilienhaus (GMH, big multi-family house)	6 and more

---

<sup>1</sup> Calculating the 1 % quantiles of the ALKIS floor numbers, I determined that 68% of Hamburg's building stock have less than two floors, while 99 % have less than six.

Conversations with the *Landesbetrieb Geoinformation und Vermessung* (LGV) revealed that the mapping of building characteristics are--to a certain degree--subjective and may vary from surveyor to surveyor. I therefore find this single objective criterion sufficient to determine the *Größenklasse* for the purpose of this study.

An R script matched the abbreviations of size and age classes for each individual building which then can be linked to a certain specific heat demand, denoted as kWh/(m<sup>2</sup>a), typical to the respective building type. I calculate the demand with the assumption that around 60% of each floor is heated. The formula, applied to each building, is:

$$(1) \text{ heat demand [kWh/a]} = \text{specific heat demand [kWh/(m}^2\text{a)]} * \text{floor area [m}^2\text{]} * \text{floor number} * 0.6$$

However, some manual adjustments had to be made since not all automated matches actually exist within the IWU typology. Adjustments were usually in favor of multi-family houses (see table A.1). For the sake of simplicity I regarded each building as not retrofitted (*Ist-Zustand*) and put buildings from the age class 2010-2015 (K) into the class of 2002-2009 (J), as the new constructions have slightly different characteristics and would demand more specifications (IWU 2015, Appendix C.3). The share of these buildings, however, is rather small. As all buildings are regarded as not retrofitted, the heat demand of my selection will be higher than the actual heat demand of the current building stock.

For the scale of the selected area I examined the 941 *Statistische Gebiete* (SG, Statistical Units) and the roughly 10,600 *Baublücke* (BB, City Blocks). The geographical boundaries of both levels are manually chosen by the *Statistikamt Nord*, the Statistical Office for Hamburg and Schleswig-Holstein. However while the SG aim for statistical comparability across the city of Hamburg, BB are primarily cut by traffic infrastructure. The scale of BB was impractical for further analysis; 90 % of all BB had only 29 or less residential buildings left after the automated type matching (with a maximum of 232). For the SG the remaining building stock was approximately tenfold. I determined the absolute building count, the total energy demand of the heated building area (kWh/a) and the heat density (MWh/(ha\*a)<sup>2</sup>) for each SG to find a neighborhood with a sufficiently large building stock and tangible density. Since the SG vary in size and construction type this was but a visual aid to select a fitting area. I eventually chose the following two SG:

- 76010: The neighborhood *Neuallermöhe* in the East of the district Bergedorf which mostly consists of one-family houses (total: 511 residential buildings) from different age classes; in the following denoted as **Bergedorf**; (Figure 1)
- 93001: A settlement, *Langenbek*, of one- and multi-family houses (total: 709 residential buildings) in the south of the district Harburg; in the following denoted as **Harburg**. (Figure 2)

<sup>2</sup> hectar of each SG's area urban space

While Bergedorf has many small structures with homogenous distances, Harburg has both denser and sparser areas of rowed houses mixed with larger scale buildings. The difference between both areas is paramount to test the algorithms applicability to diverse urban contexts.



Figure 1: Bergedorf, 511 buildings, 128 ha  
(own representation)



Figure 2: Harburg, 709 buildings, 79 ha (own representation)

## 4 COMPARISON OF AUTOMATED ALGORITHMS

I developed three algorithms to determine heat densities. In this chapter I shortly describe their unique approaches and compare their usefulness to find urban areas that fulfill certain criteria that would help to decide whether the installation of cogeneration plants seem feasible.

All three algorithms are defined as functions--programmed with R--and can be found in the following github repository: <https://github.com/hannes-seller/CORP2016>

### 4.1 Tessellation of urban space into raster tiles

The first algorithm uses a simple spatial approach that partitions the area under investigation into equally sized raster tiles. Besides a data frame with spatial data of the desired area, the function requires five numeric inputs:

- both eastern and western boundaries of the area as X coordinates
- both northern and southern boundaries of the area as Y coordinates
- the edge length of the square raster tile in meters

The function uses the coordinates to determine the area's spread as well as its geographical location and creates a grid of square tiles whose edges are equal to the length of the user's input. All tiles are consecutively numbered from north-west to south-east. In a second step, the function loops through all buildings of the data set and determines in which tile it is situated. Grouping the buildings by their tile numbers, it determines for each square:

- Total annual heat demand (MWh/a)
- Total heated floor area (m<sup>2</sup>)
- Heat density (MWh/(ha\*a))

I applied this algorithm to neighbourhoods with the raster lengths 100 m, 150 m and 200 m (1 ha, 2.25 ha, 4 ha), leading to six results which I display with histograms (see Figure A1). The outcomes for Harburg and Bergedorf are quite different (see Table 2). Bergedorf has a generally low density throughout all raster sizes. The area's density does not change much, probably due to its homogenous building structure. In Harburg,

however, the densities are less consistent. The 150 m raster has a much lower density which might be caused by the area's rather amorphous shape (see Figure 2).

Table 2: Results from tessellation, using 100, 150 and 200 m raster

Raster	Bergedorf		Harburg	
	Tiles	Med. density [MWh/(ha*a)]	Tiles	Med. density [MWh/(ha*a)]
100 m	106	53,574	78	111,454
150 m	55	45,781	43	78,88
200 m	36	47,846	23	103,038

This method's advantage is the comparability among tiles of urban space that have the same area. Thus it provides a good overview about heat densities in the area with the aim to identify those locations with tangible heat sinks. The algorithm's application only takes little time and processing power since it is basically a table join mechanism. Nonetheless, the algorithm has several disadvantages. Mostly, to ensure comparability among the raster tiles, the built-up urban structure requires a certain degree of homogeneity. As seen in Harburg (Figure 2), non-built up areas in the center and the amorphous shape (especially in the east) lead to tiles that only contain a few buildings and therefore small densities. Also, tiles closer to the edge will include urban space that is outside the area of interest which lowers the heat density as well. Since bigger buildings are assigned to the tile where their X and Y coordinates are located, they can increase the density of a certain tile while surrounding tiles can have significantly lower densities. This algorithm works better for Bergedorf--or in general--for areas with homogeneous structure, both in built-up density and construction sizes, and a more rectangular shape. Regarding raster size, lengths of roughly 100 m appear to be optimal, as bigger areas include too much space outside the area, leading to lower densities, and smaller sizes (< 1 ha) lead to higher calculated densities. The Figures A2 and A3 show the rasters with 100 m and 150 m for both Harburg and Bergedorf. Bigger raster tile have noticeably lower heat densities. Therefore, to assess the outcome, both the density and the total heat demand have to be considered and put into perspective of technical necessities of desired cogeneration power plants.

## 4.2 Clustering urban space by desired area size

The two clustering algorithms work similarly, but group buildings into clusters by assessing different criteria to determine heat densities. Since heat density is defined as MWh/a divided by hectares urban area (i.e. energy demand over area), the highest densities depends on (1) a high numerator and/or (2) a low denominator. The first algorithm aims to reach a predetermined cluster area (e.g. 1 ha) and calculates the heat density of each cluster found: The smallest area with the highest demand. The second one strives to reach a certain total heat demand with as few buildings as possible: The highest heat demand with the smallest area.

Both functions operate with two while-loops, leading to longer computation times in comparison to the raster algorithm. However, in the examples, the computation time was not longer than a few seconds with any of the algorithms. I describe the first algorithm with the following pseudo-code:

- Input: A data set with buildings that have X and Y coordinates and calculated heat demands
- while 1: repeat as long as the data set contains at least one building:
  - arrange data set, sort buildings by their demands in descending order
  - cut out the first entry (highest demand) and paste it to a temporary data set (temp)
  - while 2: add closest neighbours to temp until desired area is reached
    - calculate the distance of remaining buildings to the cut out one
    - cut out closest neighbour and add to temp
    - calculate the area described by buildings from temp (distance between smallest and biggest X coordinates times distances between smallest and highest Y coordinates)
    - repeat while 2 until temp reaches desired area (e.g. 1 ha)
  - assign a cluster number, its density and total heat demand to temp buildings
  - save temporary buildings into output data set
  - repeat while 1 until all buildings are clustered and brought to the output data set



The longer computation time comes from the necessity to re-calculate the building distances to each other every time the while 2 loop ends. The first clusters found come very close to the desired area and enable comparable densities. The later cases reach larger areas since the algorithm might need to add neighbours that are more distant, as the number of buildings in the original data set decreases after each loop.

This algorithm is better in finding the densest areas in regard of heat demand in comparison to the raster algorithm. While the boundaries of the latter ones are rather arbitrary, the first one aims directly for areas that promise a high density since it starts with the building with the highest total demand in the whole area (see Figure 3 and 4). The raster approach provides a better overview about the whole area; however, the cluster algorithm can spot islands of high densities within the area under investigation. A disadvantage is that it cannot aim for a desired heat density since the first loop iteration (highest demand building plus closest neighbour) usually creates very high densities (e.g. 5,000 MWh/(ha\*a)) as the denominator--the area of urban space between both buildings--can be very small. Therefore I recommend reaching a certain area (1 ha or more) of urban space before assessing the density.

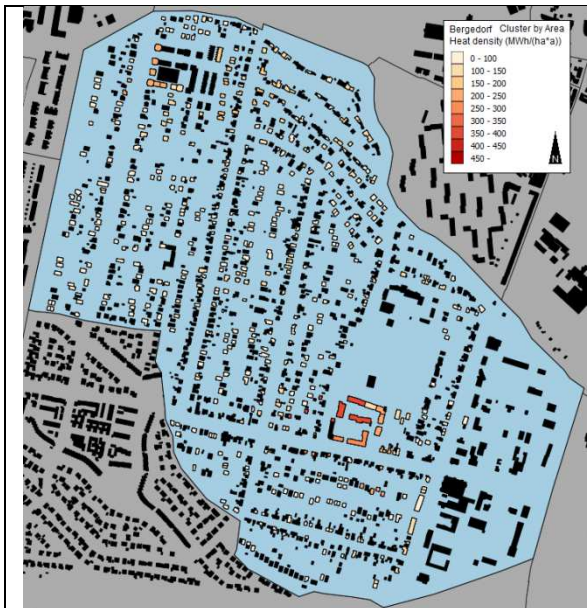


Figure 3: Bergedorf, Cluster by Area (own representation)

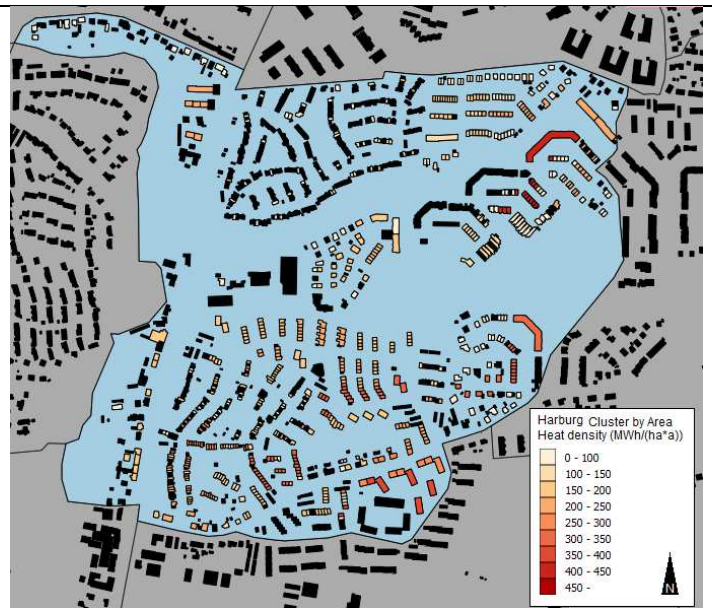


Figure 4: Harburg, Cluster by Area (own representation)

### 4.3 Clustering urban space by desired heat demand

The second cluster algorithm only differs in one aspect: The second while loop is repeated until a certain heat demand is reached, as opposed to a certain area of urban space. This method can be used to find building clusters that provide a desired total demand within the smallest possible area. However, the sizes of each cluster can vary strongly since the function cuts out buildings with highest demand first, forcing later loops to reach further to reach a certain total demand. For the investigated areas, the results look very similar to the results from the first clustering algorithm; I therefore do not include a visual representation here.

Both cluster algorithms can be used in addition to each other. While the first one finds spots with the highest density over an equal area of urban space, the second one can determine whether one of the found clusters has the highest demand of the whole area under investigation. This method helps as a visual and statistical aid to find a good location for a cogeneration power plant. However, I want to make clear that these algorithms are only meant to bring arguments to decision makings in regard of district heating planning; they do not deliver a result that makes further investigations obsolete. If a possible spot for a power plant is found, the buildings which actually should be connected to the plant need to be selected with more thorough investigation since the automatically formed clusters are based on simple algorithms that do not reflect on all relevant factors.

## 5 CONCLUSION AND OUTLOOK

This paper's aim was to determine the usefulness of three algorithms with the purpose of finding urban areas with properties that might justify the installation of cogeneration power plants. Automated processes are helpful when data sets are large and not all attributes are known for each observation (i.e. each building). Especially in urban planning, knowledge of a place is needed to make decisions. If areas under investigation are too big to ensure a decent knowledge about all local characteristics, automated approaches can help planners to set their focus to areas that are worth a deeper and more time consuming investigation.

The algorithms I presented are all meant as tools to subset larger urban areas to spots of interest. The raster tessellation provides a comparability of urban heat densities as it creates tiles with equal area sizes. Since the cities usually develop naturally and in dependence of topographical necessities, the overlay of square shapes does not do justice to all urban contexts. Especially raster tiles close to borders, natural barriers like rivers or open spaces loose comparability. The algorithm is most useful for areas with homogeneous building structures and built-up densities. Both cluster algorithms have similar strengths and weaknesses. They are apt to find spots within the urban area that fulfil certain conditions: a desired area size and density or a desired minimum total heat demand. However, they do not provide a complete overview about the whole area under investigation like the raster algorithm does. Applying the cluster mechanisms to larger areas will lead to spots of interest while areas among these spots may be underrepresented as they draw data from less desired buildings. The advantage of all algorithms is the low demand of known building characteristics: They can be applied to each set of buildings that provides geographical coordinates, number of floors and construction years.

However, I'd also like to admit that all algorithms are rather crude and could be improved by increasing the complexity of their match making processes. Furthermore, it would be beneficial to investigate more diverse urban areas as the results of all algorithms in both Harburg and Bergedorf do not differ much from each other. For the raster algorithm it is worth investigating a possibility to shift raster tiles a few meters to all directions to reduce the arbitrariness of the raster's starting points. This would allow getting multiple density values for each construction which could be averaged to eliminate outliers caused by heterogeneous urban contexts. The cluster algorithms could be taught to choose a fitting neighbour by more than just the spatial vicinity. By applying weights to the data set's attributes, these mechanisms can favour neighbours with higher heat demands or buildings which are in need of retrofitting. These options would increase the effectiveness of strategies aiming for CO<sub>2</sub> emission reduction.

## 6 REFERENCES

- AdV - Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland: Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens. Munich, 2015.  
 <<http://www.adv-online.de/AAA-Modell/binarywriterservlet?imgUid=3c860f61-34ab-4a41-52cf-b581072e13d6&uBasVariant=11111111-1111-1111-1111-111111111111>> [retrieved: 28.3.2016]
- IWU - Institut für Wohnen und Umwelt GmbH: Deutsche Wohngebäudetypologie - Beispielhafte Maßnahmen zur Verbesserung der Energieeffizienz von typischen Wohngebäuden. Darmstadt, 2015.  
 <[http://www.building-typology.eu/downloads/public/docs/brochure/DE\\_TABULA\\_TypologyBrochure\\_IWU.pdf](http://www.building-typology.eu/downloads/public/docs/brochure/DE_TABULA_TypologyBrochure_IWU.pdf)> [retrieved: 28.3.2016]
- Peters, Irene: Short portrait of the GEWISS-Project in Hamburg. Hamburg, 2015.  
 <[https://www.hcu-hamburg.de/fileadmin/documents/Professoren\\_und\\_Mitarbeiter/Irene\\_Peters/GEWISS-PortraitShort\\_2015June29.pdf](https://www.hcu-hamburg.de/fileadmin/documents/Professoren_und_Mitarbeiter/Irene_Peters/GEWISS-PortraitShort_2015June29.pdf)> [retrieved: 28.3.2016]

## 7 APPENDIX

Table A1: Overview about matching IWU types (own representation)

Automated type matching				
Nr	IWU type	Count	Available	Transform to
1	no match	111,608	no	
2	EFH_A	1,318	yes	
3	EFH_B	1,519	yes	
4	EFH_C	6,329	yes	
5	EFH_D	5,876	yes	
6	EFH_E	21,065	yes	
7	EFH_F	6,052	yes	
8	EFH_G	2,743	yes	
9	EFH_H	10,715	yes	
10	EFH_J	9,559	yes	
11	EFH_K	4,401	no	EFH_J
12	GMH_A	1	no	MFH_A
13	GMH_B	4	yes	
14	GMH_C	10	yes	
15	GMH_D	53	yes	
16	GMH_E	619	yes	
17	GMH_F	511	yes	
18	GMH_G	76	no	MFH_G
19	GMH_H	49	no	MFH_H
20	GMH_J	85	no	MFH_J
21	GMH_K	72	no	MFH_J
22	MFH_A	17	yes	
23	MFH_B	828	yes	
24	MFH_C	477	yes	
25	MFH_D	1,130	yes	
26	MFH_E	1,890	yes	
27	MFH_F	2,220	yes	
28	MFH_G	573	yes	
29	MFH_H	636	yes	
30	MFH_J	304	yes	
31	MFH_K	274	no	MFH_J
32	RH_A	76	no	MFH_A
33	RH_B	1,015	yes	
34	RH_C	1,179	yes	
35	RH_D	4,036	yes	
36	RH_E	9,166	yes	
37	RH_F	5,574	yes	
38	RH_G	1,663	yes	
39	RH_H	2,324	yes	
Total		219,307		

Manual adjustment		
Nr	IWU type	Count
1	no match	111,608
2	EFH_A	1,318
3	EFH_B	1,519
4	EFH_C	6,329
5	EFH_D	5,876
6	EFH_E	21,065
7	EFH_F	6,052
8	EFH_G	2,743
9	EFH_H	10,715
10	EFH_J	13,960
11	GMH_B	4
12	GMH_C	10
13	GMH_D	53
14	GMH_E	619
15	GMH_F	511
16	MFH_A	94
17	MFH_B	828
18	MFH_C	477
19	MFH_D	1,130
20	MFH_E	1,890
21	MFH_F	2,220
22	MFH_G	649
23	MFH_H	685
24	MFH_J	735
25	RH_B	1,015
26	RH_C	1,179
27	RH_D	4,036
28	RH_E	9,166
29	RH_F	5,574
30	RH_G	1,663
31	RH_H	2,324
32	RH_J	3,260
Total		219,307



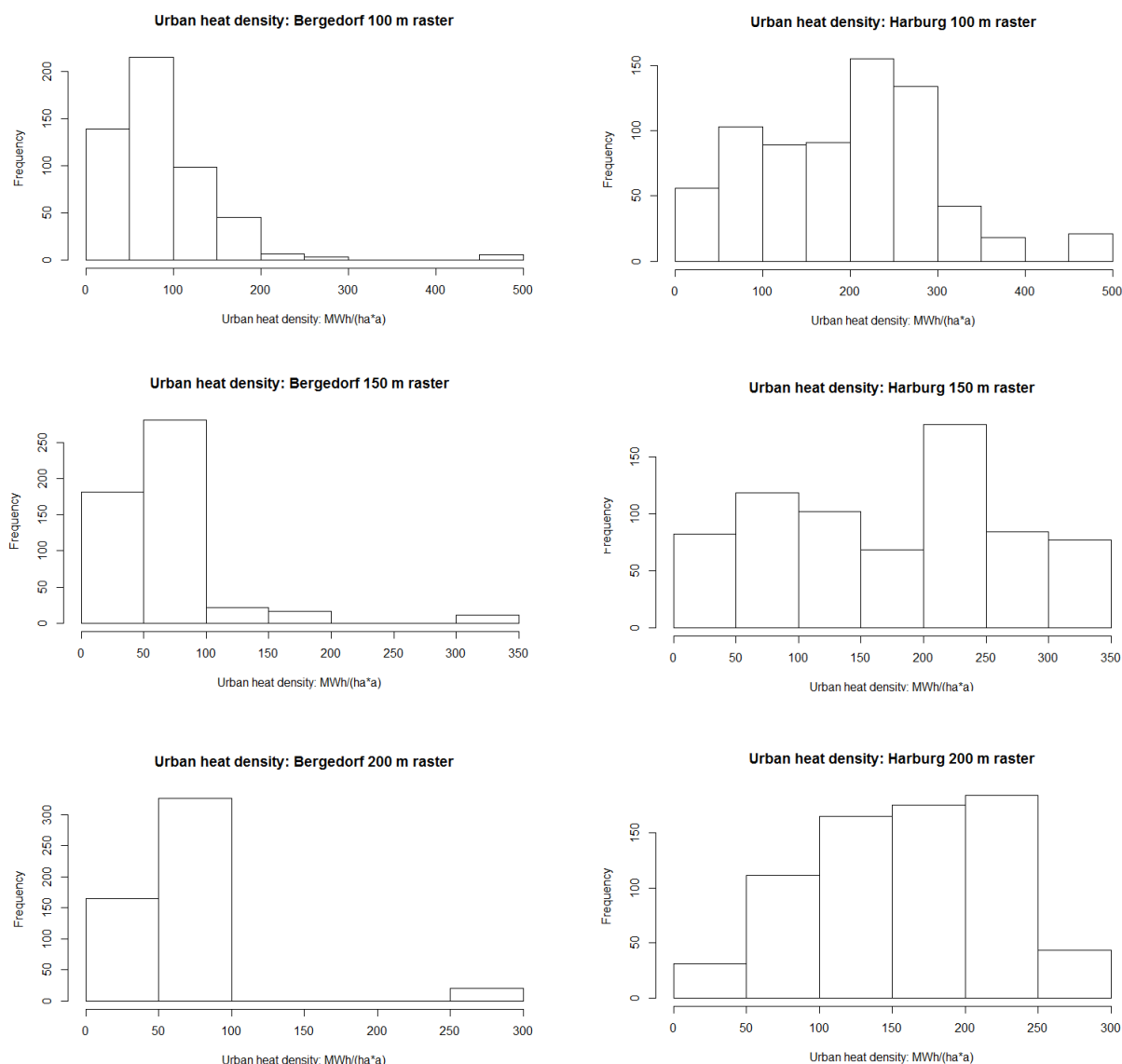


Figure A1: Histogrammes of heat density in MWh/(ha\*a) for Harburg and Bergedorf with 100, 150 and 200 m raster tiles (own representation)

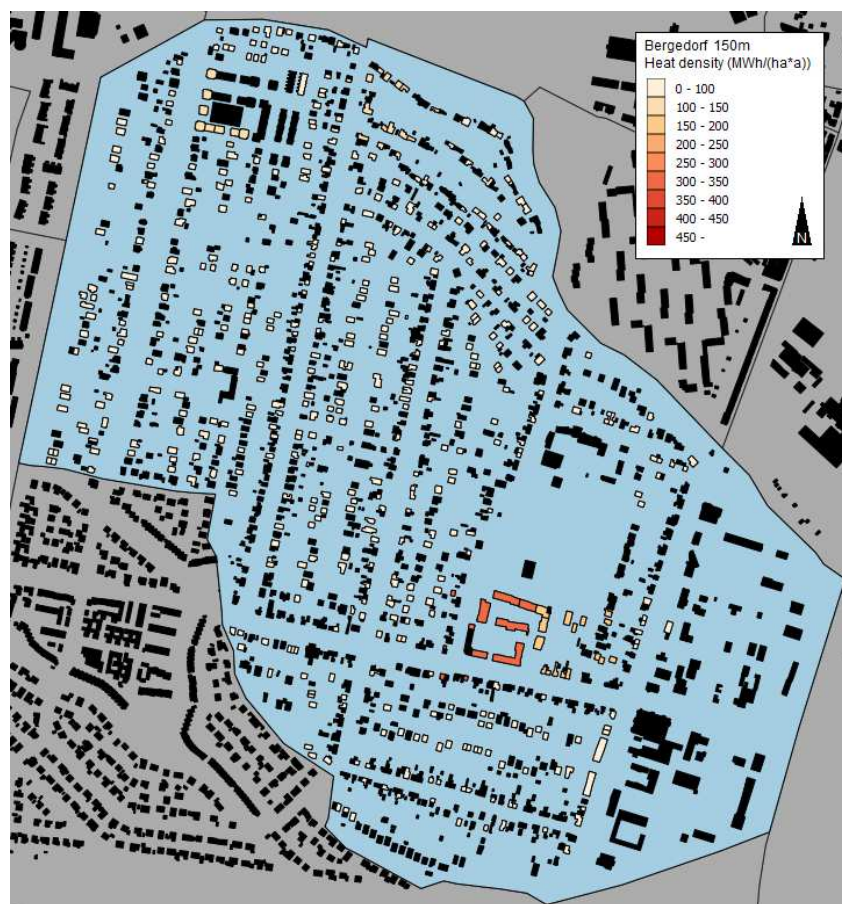
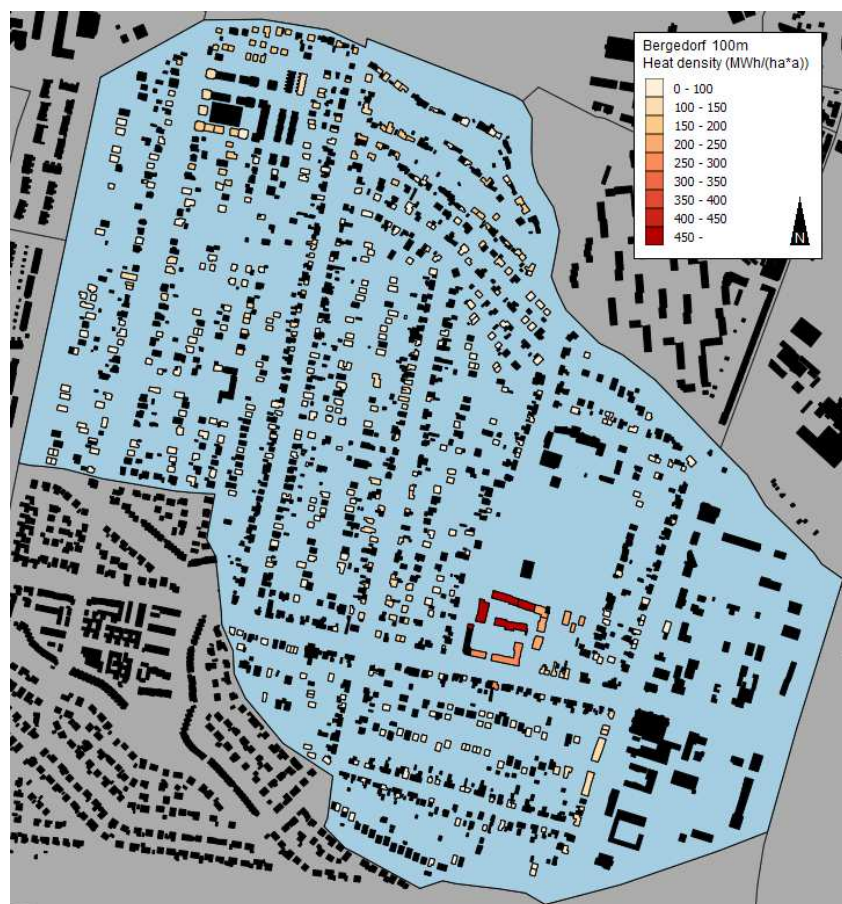


Figure A2: Bergedorf, heat density in MWh/(ha\*a) (Raster 100 m and 150 m)



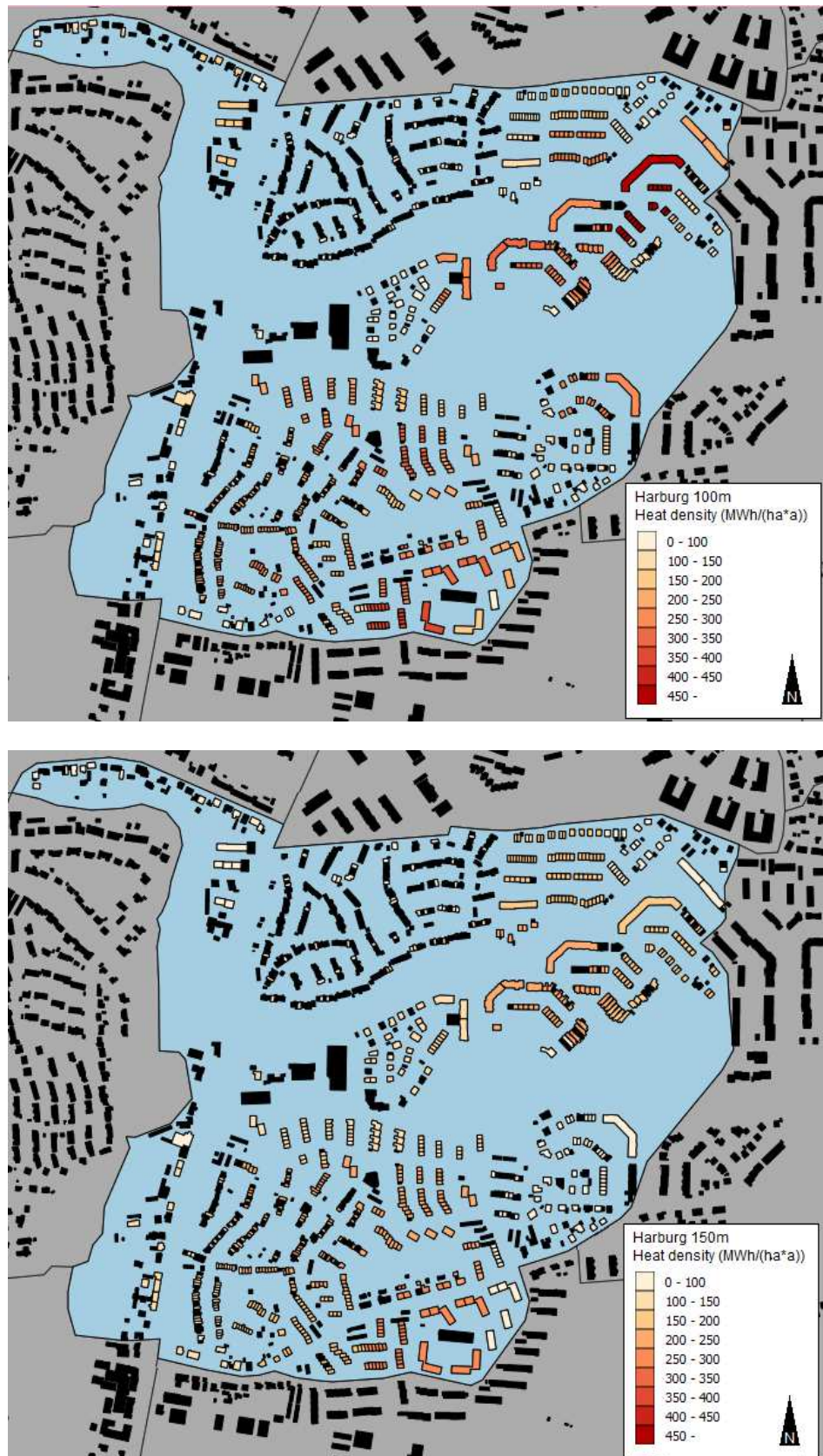


Figure A3: Harburg, heat density in MWh/(ha\*a) (Raster 100 m and 150 m)