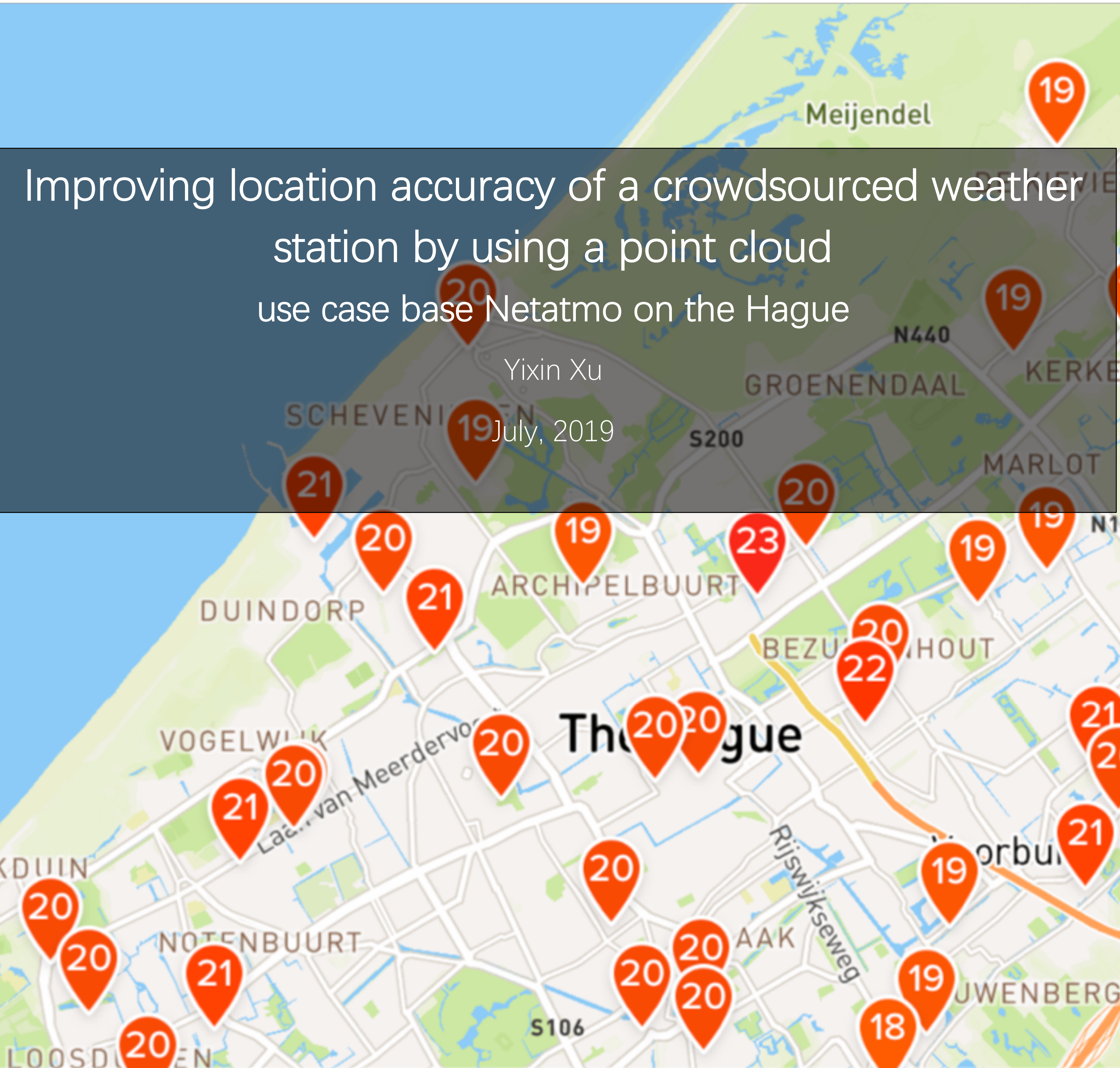
Delft University of Technology

Improving location accuracy of a crowdsourced weather station by using a point cloud

use case base Netatmo on the Hague

Yixin Xu

July, 2019



TU Delft logo by https://freebiesupply.com/logos/tu-delft-logo/

Improving location accuracy of a crowdsourced weather station by using a point cloud

----use case based Netatmo on the Hague

Master of Science in Geomatics for the Built Environment

at the Delft University of Technology,

by

Yixin Xu

July 2019

Mentors:
1st: Ken Arroyo Ohori
2nd: Alexander Wandl
Co-reader:
Agung Indrajit

An electronic version of this thesis is available at: http://repository.tudelft.nl/
Code is available at: https://github.com/yancyee/TUDMaster_code

Acknowledgements

My first gratitude is expressed to three mentors: Ken, Alexander and Agung who help me a lot in the thesis. The graduation cannot be finished without their support. Ken provides me many technical support and direction when I got lost in dealing with troubles. Due to abroad working, Ken sometime are not able to help me or have a meeting face to face. However, we still make it by online communication regardless the time lag. I could still remember he sometimes got up very early in order to attend my presentation in time. His conscientiousness as a mentor makes me warm and moved. Besides, Alex and Agung also offer suggestions which enable that I can enrich the project not only technically but also practically. I am honored to work with them in my graduation thesis.

This project is the result of 2-year abroad study in the Netherlands and the living cost and tuition fee here are not cheap especially for a non-EU student. I would like to thank my parents support and encourage me to finish M.Sc. program in the Netherlands, a country thousands of miles away from my home. Without their financial support, I do not have chance to do it by myself. Also, I would like to thank my girlfriend. Her accompany enables me to face challenges with courage but not alone. There is an old saying in China: Love is companionship. Now, I get the profound meaning behind this short sentence.

Thanks to all my friends and classmates who supported me in these two years. Special thanks to Qu Wang who always invite me to enjoy delicious food when I feel down and YuanPeng Zhang who proved that friendship has no geographical limitations. I am a big fan of Overwatch and I really enjoy time playing it with online friends after daily routine work, so I also want to express my gratitude to every Overwatch agent I fight together in the EU-sever, especially Andrenzo, Jenseits, LGDElieen, XCVBastions, SIEUJordi and Wisdom.

The gratitude towards people who help me in study life actually is beyond words. By quoting lines from Ironman, what I want to say in the end is: love you 3000 times.

Abstract

The world's continuously increasing population leads to environmental challenges, among which, the urban heat island effect has been recognized as one of the leading environmental issues recently. Using traditional weather station (usually one or two within one city and placed in rural area) to monitor and model the canopy layer urban heat phenomenon does not provide enough spatial resolution. Alternatively, the Netatmo weather station, a low cost and citizen science weather sensor, is able to collect crowdsourced temperature records and has significant strength in spatial and temporal resolution in temperature measurement. Thanks to the variety of uses of the Netatmo weather station and its open API, more temperature data could be used for UHI research. However, for scientific use, the main challenge is the data quality. For one thing, the stations' locations are set by users and are thus not accurate enough for temperature modeling in a complex city environment. For another, sensors some time generate unreliable records when exposed to solar radiance directly. These two things are actually highly interactional. Knowing the accurate location of stations could be helpful to calculate when the stations are exposed sun then filter outliers, and vice versa. However, the location information could be used to improve its accuracy is quite limited. Thus, the current work is focusing on develop an approach to determine the likely correct location of the stations.

For the development of the relocation method, different spatial and sensor datasets have been used. The temperature data in the Hague in May, 2018 have been collected from Netatmo weather stations. Additionally, the AHN3 points cloud for solar simulation and BGT shapefile for creating new location have been investigated. The methodology of relocation process is divided into 6 steps: Sensor data pre-processing, Detecting higher temperature time, Generating potential location of stations, Computing sky view (dome) and solar parameter, Finding the most likely horizontal location of the station, and Assigning height value to points. These steps also have been used with another period time in the Hague for validation and one sample Netatmo sensor experiment in Delft will be conducted.

The results proved the feasibility and rationality of the adopted methodology. Around 67% stations (new location) is shown more than 0.5 similarity when comparing with their solar simulation. Validation result detained by two period comparison indicates that over 70% Netatmo stations' new location show high quality on both the horizontal and vertical dimensions after applying the process. Validation experiment is shown a real example of fluctuated air temperature and how it will be influenced by solar radiance. In the experiment, the location error is reduced from 16 meters to 4 meters, which proves that the methodology adopted by the project is helpful to improve the station's location accuracy.

Contents

List of Figure

List of Table

1 Introduction

The following chapter will introduce scientific background related to this project. UHI has been recognized as one of the leading environmental issues and so far, and there have been several solutions to quantify the UHI. The canopy UHI is usually measured using traditional weather stations, which is not ideal as usually weather station could not offer enough spatial resolution. A new opportunity, Netatmo weather station has recently emerged in the form of cheap, citizen-used weather stations and their connection to smartphones and online map. As a new crowdsourced weather data, Netatmo has significant strength in spatial and time resolution in temperature measurement, however, challenges still exist in Netatmo station, which cannot be ignored in UHI quantifying research.

1.1 Urban Heat Island (UHI)

An urban heat island (UHI) is an urban area that is significantly warmer than the surrounding rural areas due to human activities. Recently, extreme weather records have been collected more frequently, which provided more proof about the related harmful effects on the people living in cities and the global economy. One example is the summer mortality rates in and around Shanghai, China has increasing heat-related mortality in urban regions and UHI has been proved to be directly responsible for it [1]. Akbari and Hashem also found that increase in air temperature is responsible for 5-10% of urban peak electric consumption for air conditioner use in America [2]. Besides, one study is shown that UHI in has an important impact on the primary and secondary area pollution, especially the ozone and the nitrogen oxide (NOx) [3].

Urban heat islands can be divided into three types based on different components (Figure 1):

1. canopy layer heat island (CLHI)
2. boundary layer heat island (BLHI)
3. surface heat island (SHI)

The CLHI and BLHI are warming of the urban atmosphere; the SHI is the relative warmth of urban surfaces. The urban canopy layer is the air layer around the surface in cities, extending upwards to approximately the mean building height. Above the urban canopy layer lies the urban boundary layer, which may be around 1 kilometer in thickness by day, shrinking to hundreds of meters after sunset [4]. The CLHI is the most directly connected to human's life among three UHI, therefore it is the most studied one. Also, CLHI is the type that is mostly discussed in this thesis.

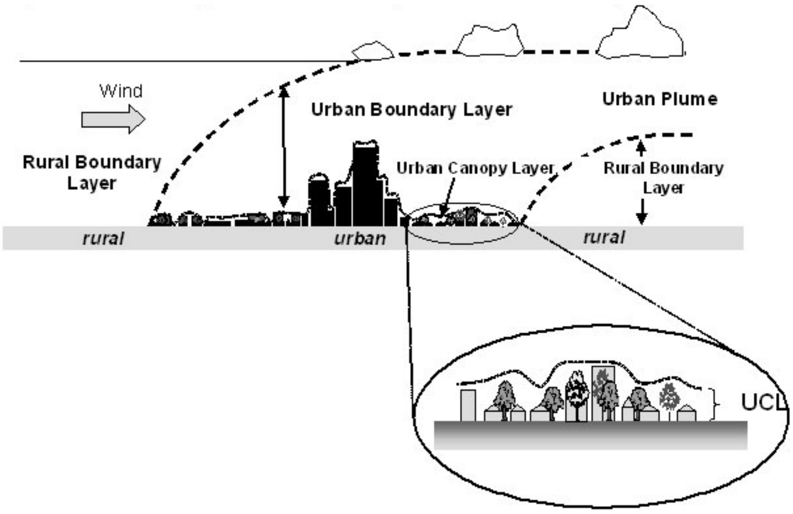


Figure 1. Depiction of the different components of the UHI. [5]

Based on the components of the UHI as discussed above, two main acquisition approaches are used to observe the UHI: remote sensing and ground-based weather stations. Researchers measure air temperatures for CLHI or BLHI directly using thermometers, whereas the SHI is measured by remote sensors mounted on satellites or aircraft [5]. The main difference between the two methods are spatial and temporal characteristics. Thermometer measurements have high temporal resolution, but only one location per sensor. On the contrary, remote sensing images have a better spatial resolution, but the data only describes the temperature once along period of time.

1.2 Quantifying the UHI

The CLHI is usually observed by using ground-based station which is not ideal as they have limited spatial resolution. Alternatively, many studies have attempted to quantify the UHI using remote sensing. This provides spatial data at a daily resolution, but it observes land surface temperatures and different to air temperatures. Given these restrictions, numerical models are frequently used instead to quantify the UHI [6]. However, due to lack of observation data, the validation of accurate UHI simulations is hard to guarantee [7].

A recent trend in urban climatology has seen an increasing use of high resolution urban meteorological networks as well as the decrease in the costs of instrumentation. A new opportunity, Netatmo weather station has recently emerged in the form of cheap, civil utilization weather stations that connect to local Wi-Fi networks and the crowdsourced data is available through their public API. But a considerable scientific challenge remains: can they provide sufficient quality to be accepted by the atmospheric science [8]?

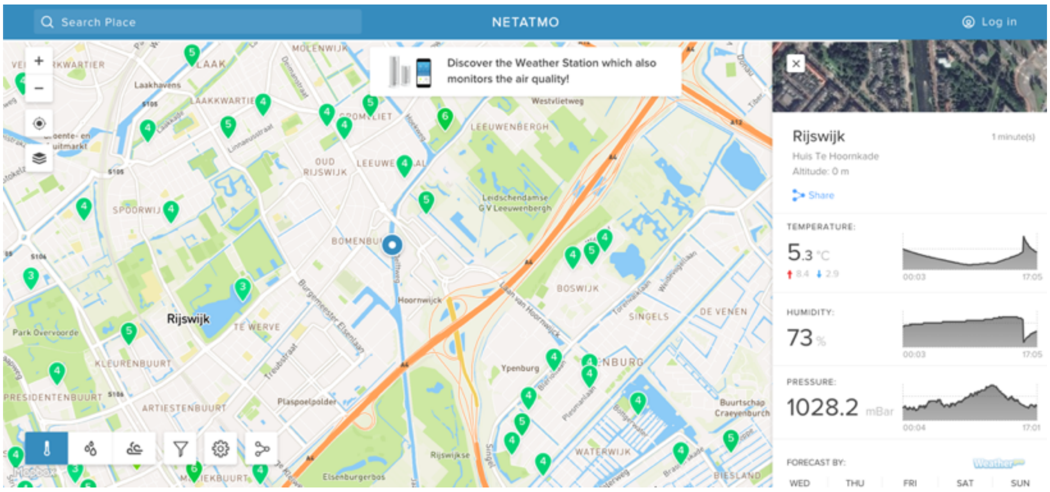
1.3 Crowdsourced data

Crowdsourced data is a sourcing model in which organizations or company can derive data from many users or publics. Advantages of using crowdsourced may include improved costs, speed, quality, flexibility, scalability, or diversity [9].

Crowdsourced was first termed by Howe referring to the idea of outsourcing to the crowd [10]. Linked with civil engagement activities, crowdsourced is now increasingly finding itself as a technique for gathering massive data in scientific subjects [11]. However, the use of crowdsourced data in the atmospheric sciences is very limited when compared with other scientific studies and the main reason is the difficulty in obtain an accurate typical observation [12]. Still, the results from the validation exercise of the Netatmo weather station with standard measurements have proved promising [13].

The crowdsourced weather station in this study is from the Netatmo company. Netatmo weather station is easily configured and controlled. Using a smartphone, users are able to monitor and record the meteorological data and the station's location. Besides, the spatial density of Netatmo station ensures that these stations could work as a network. The dense data means more choice when dealing with data, e.g. when one or two station is faulty, it's possible to replace by other stations.

Records from the station are transmitted wirelessly, using Wi-Fi and configuring by Bluetooth, to the Netatmo sever and available via a 'weathermap' (Figure 2) on the Netatmo website. All weather observations are updated every 15 min. It's also noticeable that the data is shown in the "weathermap" is already filtered by the Netatmo, so the data looks much smoother than the raw data from the API. The Netatmo API ensures retrieve publicly shared weather raw data from outdoor modules within a predefined area and that's also the data source in this project.



Discover the Weather Station which also monitors the air quality!

Rijswijk 1 minute(s)
Huis Te Hoornkade
Altitude: 0 m

TEMPERATURE: 5.3 °C ↑ 8.4 ↓ 2.9

HUMIDITY: 73 %

PRESSURE: 1028.2 mBar

Figure 2. Netatmo Weather Map

The Royal Netherlands Meteorological Institute (KNMI), provides hourly weather data including temperature for fixed positions. The KNMI sensors are usually placed in the rural areas: KNMI 330 is located at Hoek van Holland (next to the sea) and the KNMI 344 (The Hague airport) [14]. Consequently, KNMI for now is not suitable for researching temperature difference between urban and rural area.

Alternatively, although Netatmo weather station most of time can provide reliable temperature data, the raw data (very noisy) may not suitable enough for AUHI research before further processing. The first reason is that there is not case which can block radiance outside the weather station, thus, air temperature measurement influenced by sun radiance could be higher than true air temperature. Netatmo also mentions that temperature records when exposed to the sun could be 1-2 ℃ higher than that in the shadow and the accuracy of that would be 0.1 ℃ normally [15]. The second reason is that the most of Netatmo users are not experts and this means they might put the sensor anywhere or for different purposes (Figure 3) and therefore sensors could generate abnormal temperature pattern or extreme values. What’s more, the location of Netatmo stations are given by user’s smartphone or simply using the address or by clicking in a web map. Usually the accuracy of smartphone GNSS is about 15-30m and depends on smartphone itself and GNSS application [16]. Considering that most of the stations are put near buildings, the accuracy of the location may also suffer from multipath influence. Besides, some system errors are also found in the raw data, e.g. some sensors record temperature only 10 times per day, which is not enough for further research, and some sensors keep recording same value which may result from hardware issues. An example of system errors is shown in Figure 6. It’s obvious that the top two lines don’t change too much with time, which does not fit with common sense.

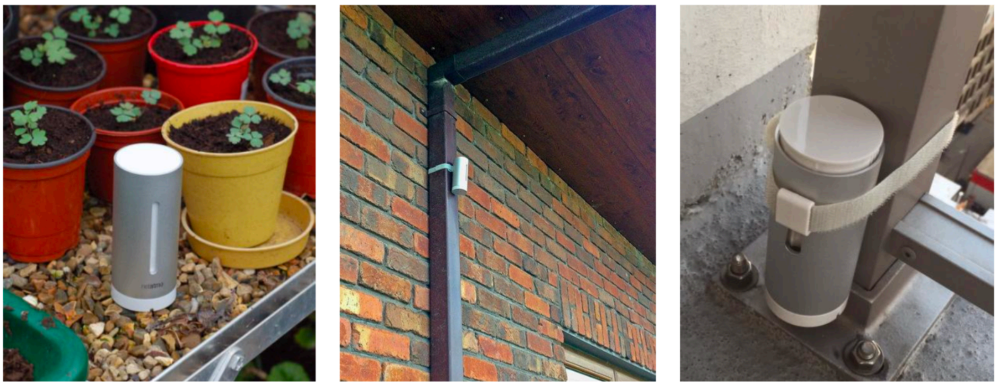


Figure 3. Different settings for the Netatmo weather station [17]

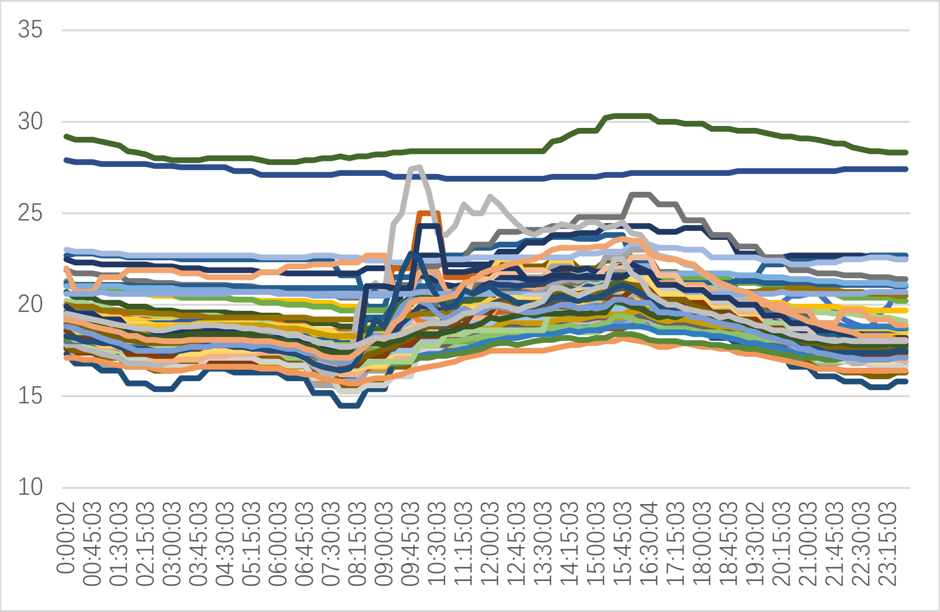


Figure 4. Example of system error (data source from Netatmo API on September 10th, 2018)

1.4 Problem statement

Quantifying the UHI using remote sensing now is not suitable for canopy layer heat island research, which focuses on air temperature above the surface. Air temperature is usually detected by traditional ground-based thermometers (Figure 5) which are located at certain height above the ground and placed in the shade. However, due to the lack of spatial resolution, many traditional thermometers are hard to cover rural areas or large city nevertheless further spatial UHI research.

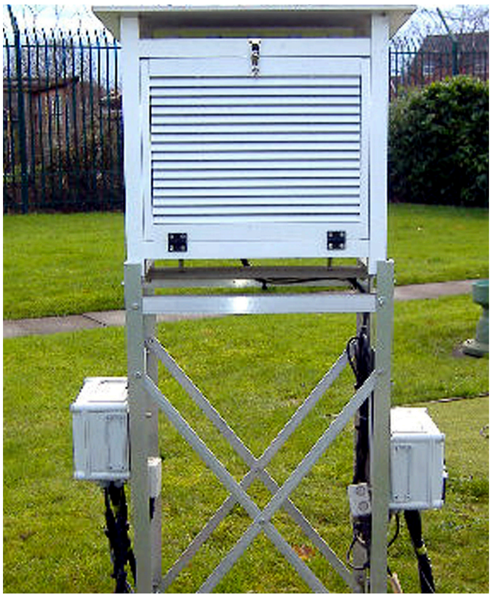


Figure 5. traditional ground - based thermometer [18]

Netatmo, as a new crowdsourced weather data, has significant strengths in spatial and time resolution in temperature measurement. However, two main challenges existing in Netatmo station cannot be ignored in UHI-quantifying research. The first challenge is outliers in the raw data. Outliers could come from system errors or solar radiance, as discussed in Section 1.4, but so far previous work mainly uses mathematical methods to remove system errors and there is lack of a method to remove outliers caused by solar radiance. Another challenge is the accuracy of the stations' location. The reason why accurate locations are important in UHI research is that the temperature divergency could be remarkable. Figure 6 is shown an example of UHI modeling result in the Hague. UHI could relate to many factors e.g. NDVI, building density, land surface character and, so on so the air temperature in different regions within one city could show significant variation (also shown in Figure 6). In other words, any specific place in the city has its corresponding UHI impact factor. Accordingly, obtaining an accurate location of each sensor should not be ignored in UHI quantifying or observation.

These two challenges are actually highly related and interactional. Knowing the accurate location of stations could be helpful to calculate when the stations are exposed sun then flite outliers, and vice versa. Unfortunately, both two data (coordinates and temperature records) more or less has their defect and the project need to determine which of them is easier to start with. Because lack of temperature contrast experiment, finding the time when a station records higher temperature data due to solar radiance is ambiguous. However, the location information could be used to improve its accuracy is even more limited. Thus, the current work is focusing on develop an approach to relocate the stations and based on new location, solar influence time in the research period will be calculated.

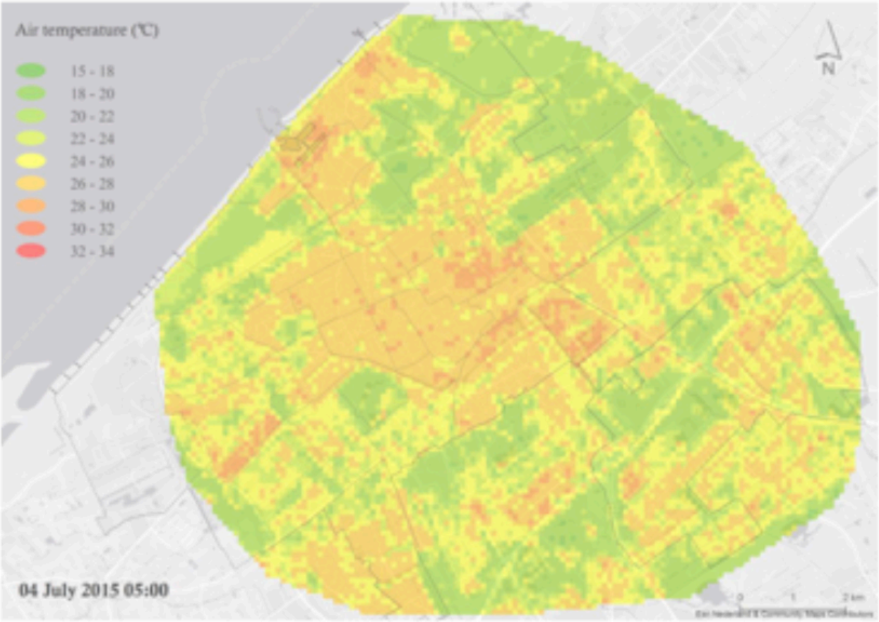


Figure 6. UHI modeling result in the Hague [20]

2 Research questions and scope

The aim of the current work is relocating the stations and finding solar influence time based on the temperature records of the Netatmo sensors in the Hague for the May, 2018. In order to achieve this, outliers caused by system errors will be removed and then remaining data is saved in the dataset. The project will also check for each sensor, when it is shown solar influenced symptoms and then compare it with solar simulations from candidate locations of the sensor.

This thesis will focus on the temperature data from the Netatmo outdoor modules in the Hague. These records are continuously collected and stored in a server in the TU Delft. Other data, e.g. indoor station modules or data from other cities will not be considered. Besides, the project will mainly analysis temperature pattern from direct solar radiance but other types of abnormal behaviors, e.g. system error will only be briefly discussed. Also, the project will not be concerned with details about UHI modeling or other factors that contribute to the UHI. Although the time when sensor is exposed to sun will be calculated, the further research about how to deal with these outliers will not be a part of this project.

The primary research question and the matching sub-questions are defined below:

- How to locate a more accurate position of NETATMO sensors?

    1. How to find potential locations for each sensor?
    2. How to know if a sensor’s record is higher than it should be?
    3. For a certain area, how to know when it receives direct solar radiance?
    4. How to compare station records with solar simulation?

3 Related work

This chapter states an overview of the concerned research, showing how the Urban Heat Island topics relate to sensed air temperature data. Furthermore, review of the data source and solar simulation method has been provided. Researches about using points cloud to correct location of sensors is a quite new topic, hence relevant papers cannot be found so far.

3.1 Related UHI research

So far, there have been three TU Delft MSc thesis projects related to the weather station and all of them stress the UHI modeling. All of them involve sensed air temperature data as a validation tool for modeling result.

Lilia Angelova has developed several statistical models showing the UHI in the Hague. Models are based on different geo-information and Netatmo records have been used for validation. Every spatial model and its impression on the statistical analysis was studied. Six distinct UHI contributing 2D factors have been researched: Building density, Land cover index, Vegetation index, Sky View Factor, Non-permeable surfaces and Vehicle traffic density. Her result is shown Sky View Factor and Non-permeable place are two main factors affecting the UHI [19].

Likewise, Anna-Maria Ntarladima studied a solution to analyze UHI and visualize dynamical change of it. Spatial factors influencing the UHI have been computed and a dataset of 140 Netatmo stations distributed in The Hague constitutes the most important data source. Combined with the KNMI temperatures, UHI is modelled in time and visualized. An UHI model is implemented for all cells in the grid by using the relationship between the temperature records with the spatial information. The Netatmo records were further used to validate the UHI models [20].

Iris A.H. Theunisse created a 3D temperature model by combining weather station data and CityGML. Temperature records (Netatmo included) from several station sources are collected and used to create the model. Data was coming from about 1300 weather stations that were distributed in Rotterdam in 2014. The sensed records are connected to the 3D model showing all locations within Rotterdam. The model could be useful for simulating indoor temperatures for all areas in the city and this is done by analyzing the relation between the temperature records and the environmental variables of each station’s address. The raw temperature data is also used to verify and to check the accuracy of the simulation results. But other factors which might bring influence on temperature in the city are not studied [21].

3.2 Netatmo weather station

Papers introducing scientific research based on the Netatmo sensor are not many and usages of this sensor are in the quite preliminary stage.

Lee Chapman applies air temperature from the Netatmo sensors to quantify the UHI effect of London in the summer of 2015. The results are highly similar to previous research. The normal observations showed a range of magnitudes of between 1 - 6 ◦C in the London subjected to atmospheric stability. However, the paper also indicates that some results cannot be clearly explained by weather conditions and therefore the data quality of crowdsourced data should be noticed [11].

Fred Meier believes Netatmo sensors play a role of a medium between citizen utilization and crowdsourced data. His research is shown that crowdsourced temperature records could be useful to UHI research and the result finds different UHI rules in Berlin during day and night. The density of Netatmo stations in Berlin exceeds that traditional weather stations networks by far, but the paper also proposes that observations of standardized, calibrated and their data quality would be important in order to validate such weather crowdsourced data [12].

3.3 Solar simulation

As Section 1.4 mentioned above, temperature records exposed to the Sun could be 1-2 ℃ higher than normal air temperature. It’s obvious that if the project wants to improve Netatmo data accuracy, solar simulation must be done, e.g. how that landing site is lit by the sun within one city.

High accuracy solar radiation in a large area can be simulated by using points cloud. Andereas Jochem use a points cloud based solar radiance model, which is embedded into Open Source SAGA GIS. It applies the 3-dimentional coordinates of each point cloud for modeling of the solar radiance. In order to handle the huge amount of spatial data with insufficient computer RAM, all the points cloud data is placed in the Laser data Information System [22].

However, putting the whole of the Hague into solar radiation model is complicated and if dynamic solar change is considered, the calculation would be more time consuming. Alternatively, this project uses sky view with sun’s parameters to simplify the simulation (Section 4.4).

3.3.1 Sky view creation

The SVF provides the relationship between the visible location of the sky and covered environments, e.g. man-made objects and vegetations, by the ratio of the total amount of radiation received from the earth surface to that available from a radiant environment [24].

Usually, sky view modeling research focuses only on fish-eye photography, vector, raster or 3D building models (Figure 7) which blocks the sky in urban. But these input data sources bring challenges in the study of the effect of plants because the complex geometry of plants is usually hard to render precisely in a 3D model, while they also play an important role in real dome reconstruction. Using point clouds with classification allows reconstruction to take plants into consideration. An, S. M proved that 3D point cloud source would be helpful for quantitative analysis of urban components by lowering the structured dimensional complexes, not only by shape itself but also with many meaningful indices such as SVF. He proposed a new SVF simulation solution (Figure 8) based on point clouds. Digital dome was applied to place points cloud on the digital sky view [27].

Similarly, the urban horizon group (Geomatics, TU Delft) has developed a more efficient method to reconstruct dome and sky view factor in using AHN3 points cloud. The main purpose of this work is the estimation of SVF, a necessary element for modern urban planning, for the Hague. To calculate SVF, the methodology used is based on 3D point clouds in order to incorporate the urban environment in its entirety (including vegetation) [28]. The algorithm is embedded in the webpage, and users are able to select different places in the Hague and check their digital dome and SVF. The result is shown that their solution provides a efficient and accurate simulation process for dome reconstruction.

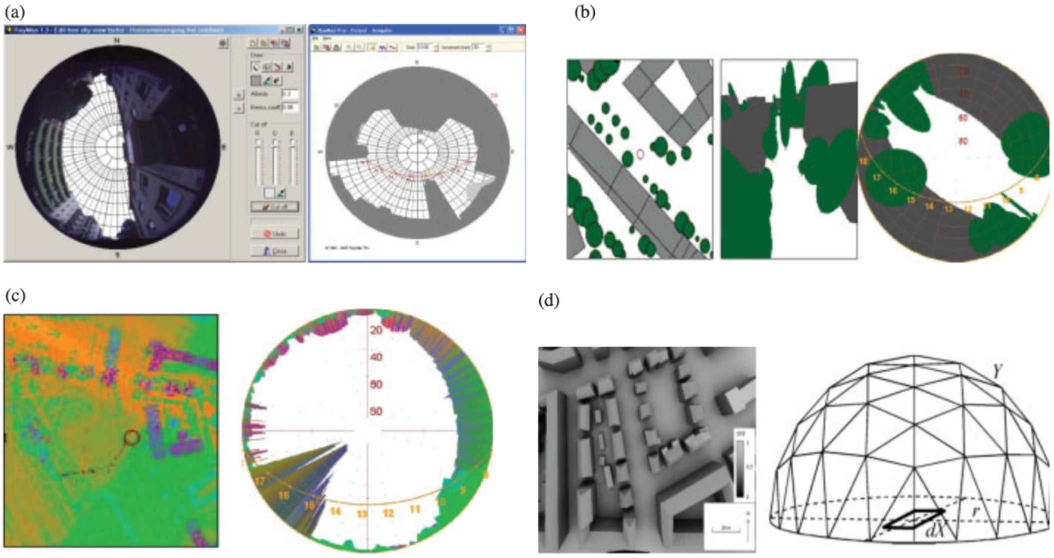


Figure 7. Dome reconstructions of data source types for SVF [25,26].

(a) Fish-eye photograph SVF, (b) vector SVF, (c) raster SVF, (d) 3D model SVF

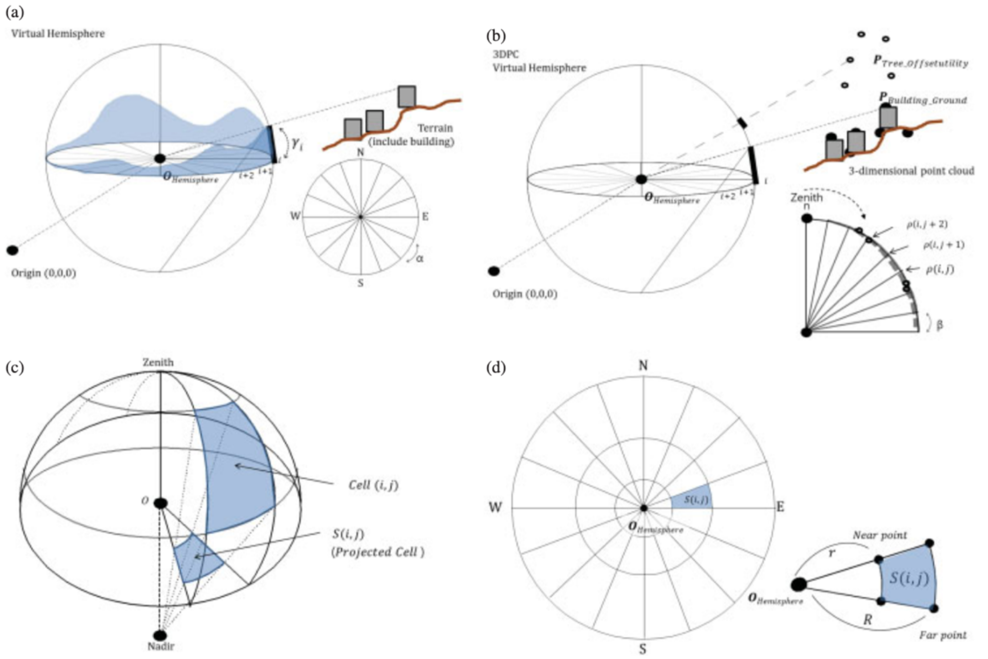


Figure 8. SVF analysis method using a point cloud [27] (a) Digital dome for raster SVF such as DEM (b) vertical unit angle of points clouds virtual hemisphere (c) virtual hemisphere array (d) the projected area calculation of the cell

3.3.2 Solar parameter

In order to compute solar simulation, solar parameters that describe sun’s track are also needed to know. Reach about solar parameters are common and the project mainly selects materials from University of Oregon Solar Radiation Monitoring Laboratory (UO SRML) as the reference. The UO SRML is a regional solar radiation data center, whose goal is to provide sound solar resource data for planning, design, deployment. During 2002, the UO SRML began developing educational material on the use of solar radiation data and provide a basic understanding of the solar resource and the uses of solar radiation data [29]. Detailed methodology about describing sun could be found at Section 4.4.1.

4 Methodology

In order in answer the research questions, the methodology of this project is divided into 6 steps: Sensor data pre-process, Detecting higher temperature time, Generating potential location of stations, Computing sky view (dome) and solar parameter, Finding the most likely horizontal location of the station, and Assigning height value to points.

Sensor data pre-processing is the first step-up of the methodology which aims to remove system failure outliers from raw temperature records and merge 7–day data as the research period into the dataset.

The second stage is detecting higher temperature data. The methodology here is trying to find when or whether the weather stations receive solar radiance and detecting higher temperature records. However, due to lack of comparison equipment, it is almost impossible to know an accurate real air temperature at each sensor's location. The solution in the project will use an "average temperature" to represent real temperature changing patterns and then follow an "increase check" algorithm.

Although the coordinates of stations are given online, they are actually uploaded by users (phone's GPS or click at online map) and not accurate enough for urban heat modeling. Therefore, the next step is that the project will generate scatter points which represent the candidates of a stations' real location. The scatter points with uniform density will be inside a "constrained buffer" which is made by a circle but avoid building and transportation areas.

In order to know, for each potential location, when it receives direct solar radiance, dome and sun position (elevation angle and azimuth angle) will be computed. For a location, the time it receives direct solar radiance is the time the sun is shown up in the sky part if its dome. Dome reconstruction is made using the AHN3 points cloud in the Hague.

The following step is finding the most likely horizontal location of the station. This is comparing the result from step 2 and step 4 for each sensor. For example, if the result derived from data is that from "7:50 am to 2:50 pm the sensor will detect higher temperature than it should be", then the potential locations whose output from solar simulation are the closest to "7:50 am to 2:50 pm" will be set as the location of the Netatmo station. This step will generate the 10 highest similarity 2D points and pass them to the next step.

The final one will be assigning a height value to the points. It is possible that users don't place their sensors directly on the ground. A higher view point has a bigger sky view factor and thus bring higher solar susceptibility to the weather stations, therefore, the project also take height into consideration not only horizontal coordinates. The approach here is similar to previous steps: vertically generating potential points for the previous 10 points and then comparing their solar simulation result with temperature records and finding which of these 3D points is most suitable.

4.1 Sensor data pre-process

Sensor data in this project is chosen from 22th, May, 2018 to 28th, May, 2018, a consecutive 7-day period without rain and mostly uncloudy in the Hague [22]. This is done to try to avoid influence from weather element. Also, this range of days ensures that the solar elevation angle is large enough for sky view factor research in the next section. Another important reason is AHN3 points cloud are collected in summer so the research period should be as close to this as possible in order to avoid seasonal influence e.g. leaf. The data pre-process part is divided into one-day data process and seven-day data process.

4.1.1 One-day sensor data

As Section 1.2 mentioned above, the raw data contains many outliers which not only result from solar radiance but also because of system errors. Obviously, system errors (e.g. hardware failure) are not concerned in this project so they should be removed before further processing. It’s noticeable that some extreme high temperature records do not necessarily mean system errors caused by direct solar radiance, but because it might result from users putting a station somewhere warmer than the environment e.g. next to the building’s wall in winter. The project will only confirm a sensor is “problematic” when it is shown a very abnormal pattern e.g. a temperature difference between sunset and sunrise is less than 3 ℃ (this is just a threshold ensures sensors will not keep recording same temperatures, so the phenomenon in Figure 4 will not appear) or wrong a too low frequency (e.g. only 50 records in one day which is not sufficient for research). Once a sensor is problematic, it will be deleted from the dataset because it is considered as unreliable. Also, the remaining points will be clipped, which ensures every sensor is located in the Hague (Figure 9).

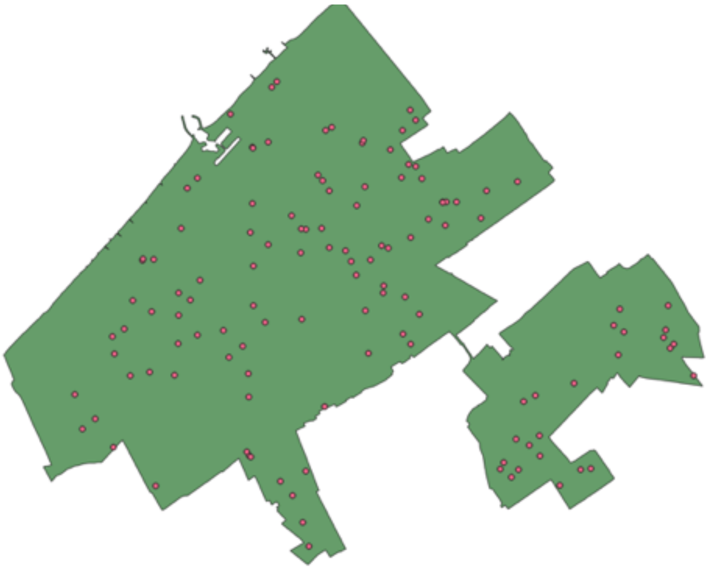
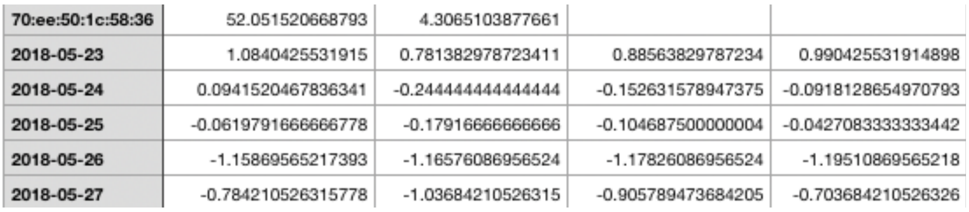


Figure 9. Sensors’ distribution after one-day pre-processing

4.1.2 Seven-day sensor data

However, it is possible that one sensor could work normally on Monday but abnormally on Tuesday or Wednesday. Hence, only use one-day filter is not “fair” enough for all sensors. Also, sensors that only work correctly one day in this week are not suitable for further analysis. Here, I have developed another method to deal with the issue mentioned above: a sensor can pass the seven-day filter and go to the dataset, if at least 5 days of data can pass through one day filter.

An example of a sensor in the dataset is shown below (Figure 10). The first row includes sensor’s MAC address and its coordinates (CRS = WGS84) while the first column is the dates pass one-day filter. This sensor cannot work properly at 22th, May and 28th, May, but can still save to dataset because it still has 5 working days (minimal criterion the in seven-day filter). The values in the table are the temperature differences between sensor records and average temperature, which will be elaborated in Section 4.2. In total, there are 186 sensors left in the datasets in the Hague. Notice that all sensors of the Netatmo weather station the paper mentions below are after pre-processing.



70:ee:50:1c:58:36	52.051520668793	4.3065103877661		
2018-05-23	1.0840425531915	0.781382978723411	0.88563829787234	0.990425531914898
2018-05-24	0.0941520467836341	-0.244444444444444	-0.152631578947375	-0.0918128654970793
2018-05-25	-0.0619791666666778	-0.179166666666666	-0.104687500000004	-0.0427083333333442
2018-05-26	-1.15869565217393	-1.16576086956524	-1.17826086956524	-1.19510869565218
2018-05-27	-0.784210526315778	-1.03684210526315	-0.905789473684205	-0.703684210526326

Figure 10. Example of a sensor in dataset after seven-day pre-process

4.2 Detecting higher temperature time

It’s important to know when the Netatmo station will record higher temperature than it should be, and this also means the time the station is exposed to solar radiance. This “time” can be used to locate sensor more precisely (shown in section below). However, it is almost impossible to know an accurate real air temperature at each sensor’s location, so the data (time and precise difference) will not be derived directly. An interpolation method is helpful to predict the value but will largely be influenced by the nearest points, and the reliability of the nearest points are unknown. According to this, the project plans to use “average temperature” to represent real temperature. Although the absolute value here is not accurate, the temperature changing pattern is more or less reliable because average relieves the outliers influence from solar radiance. The reason is that the time sensors are exposed to solar radiance is not identical and all sensors are taken into consideration with same weight therefore, outliers could be “diffuse”.

After that, it’s possible to know when the temperature is higher than it should be by using record data minus average data, following an algorithm to classify these points are increasing, decreasing or neither.

The algorithm is based on the current point, the previous point and the next point and then the gradient calculation. The reason consider 3 points is trying to avoid some influence from outliers and make the classification smoother. A pseudo-code for the increase check is shown below.

```
Increase_lis= [ ]

for index in Pts [sunrise: sunset]:

  if Pt[index] - Pt[index-1] > 0.2 or Pt[index+1] - Pt[index] > 0.2 or Pt[index+1] - Pt[index-1] > 0.4 :

  Increase_lis. append (Pt[index])
```

An example result (only one sensor in one day) is shown in picture below. Red dots mean that temperature difference is increasing while blue is decreasing. Here we obtain 95 data points (x-axis) because the raw data is updated 15 min per time and there will be 95 records in one day. The red dots are concentrated from x=40 to x=50 and the corresponding time is about 10:00am and 12:30pm respectively (Figure 11). To simplify the research, the project will assume that this sensor is likely to be exposed to sun radiance during this period of time.

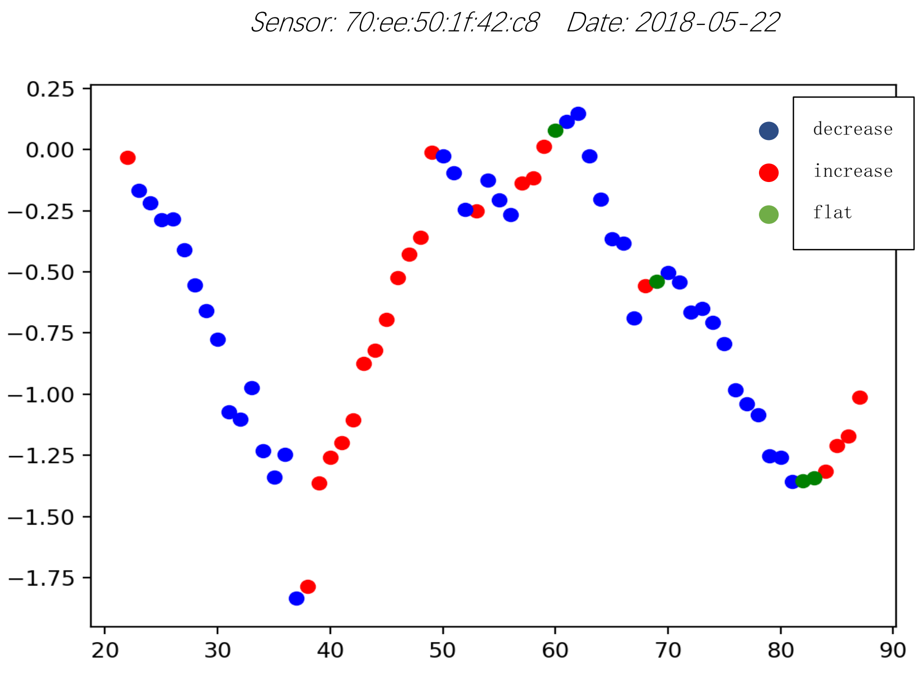


Figure 11. One example to classify temperature difference

The picture above only is shown one sensor in one day, however, for each sensor, the dataset store at least 5 days data. In order to consider data collected from different date, I have introduced a concept called “increase possibility”. For instance, for sensor A at 1:00 am, the increase/decrease checks in the 7 days are:

[05-22: increase ;05-23: decrease;05-24: increase;05-25: decrease;05-26: increase;

05-27: decrease;05-28: increase]

The increase possibility at 1:00 am for sensor A will thus be

increase days / all days = 4 / 7 = 0.571

Only the time where the increase possibility is bigger than 0.5 will be consider as "the time influenced by solar radiance". The result here will be used in Section 4.5.

4.3 Generating potential locations of Netatmo stations

Although the coordinate of each station is given in the dataset, it's actually a rough location info and a real location of a station could be inside a buffer of the given location. The radius of the buffer is the accuracy of the given location. The project selects 15m as radius, which was chosen according to the accuracy of mobile phone GPS.

However, resection of the buffer is buildings and transportation areas. All temperature data used in this project are collected from the Netatmo station outdoor module (it is possible that some users use outdoor module inside, but they are removed in pre-process because temperature inside room will not change too much with time, thus they cannot pass through the filter) so the part(s) where the buffer covers buildings will not be considered when generating potential locations. Likewise, no users will put sensors in a transportation area. The buildings' and roads' footprints will be extracted from the BGT dataset (The Basic Registration of topography map of the Netherlands. In the BGT, objects such as buildings, roads, water, railway lines and greenery are defined and classified). Then points are generated with same distance between them (Figure 12).

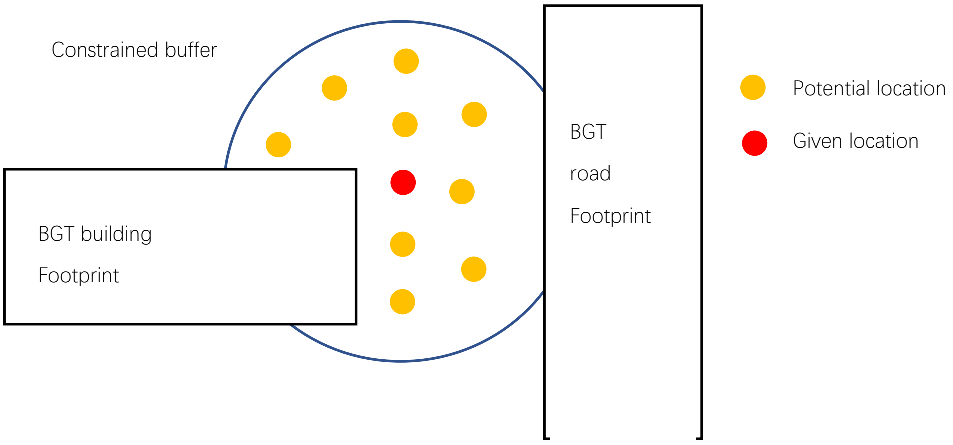


Figure 12. Principle to generate potential location

4.3.1 Generating scatter points around a given location

Usually, generating scatter points using algorithm is done using two nesting loops. An outside loop is generating points from initial point with a given distance while the inside loop is rotating the central point with same angle from 0 to 2pi. The disadvantage of this method is the density of points will be lager when close to the center of a circle. In the project, this means the real location has higher possibility when close to given location which is logical. However, the center circle would be extremely dense and if the density close to the edge of the circle will be very low and cause may data redundancy. Alternatively, the project creates points with evenly density inside a bounding box (length of side is equal to the double radius), then remove points whose distance to the initial point is larger than the radius. Both methods are shown in Figure 13.

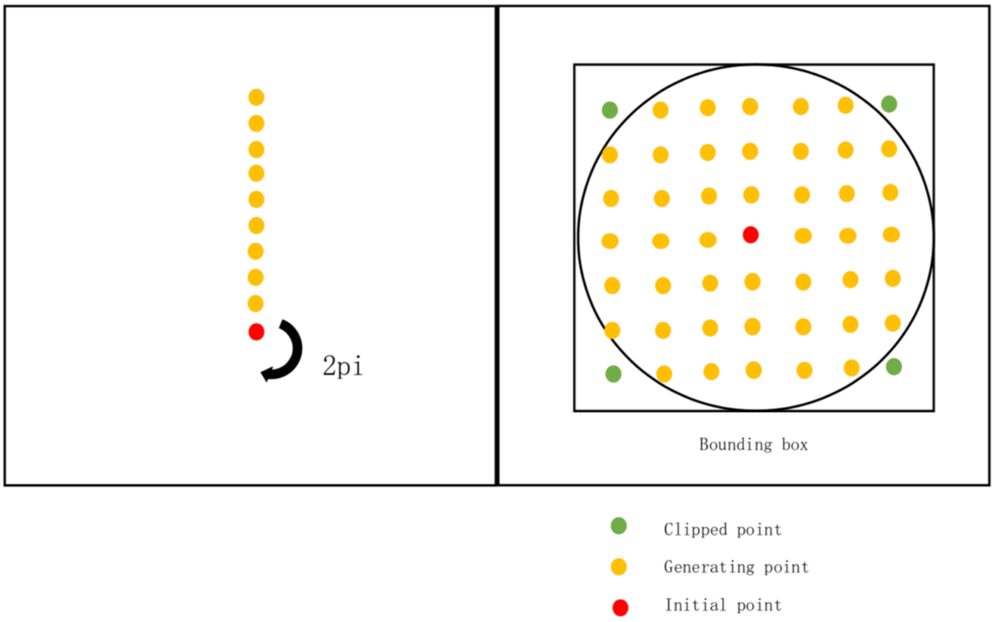


Figure 13. Principle of creating scatter points within a circle

(left: common method; right: project adoption)

4.3.2 Removing points inside polygons

Due to the thousands of building and road polygons in the Hague, the calculation of whether each point inside each polygon will consume very long time. Two methods are developed here in order to lower the calculation complexity.

The first method is called Tiling. All polygons in the Hague are divided into several square tile sand each of them has its ID. Depending on the given sensor’s location and bounding box of tiles, only one tile will be used for further calculation. One example is shown below.

In this case, only Tile #6 will be input. In this project, the Hague is divided into 12 tiles and the specific tile numbers can be found in the Appendix D.

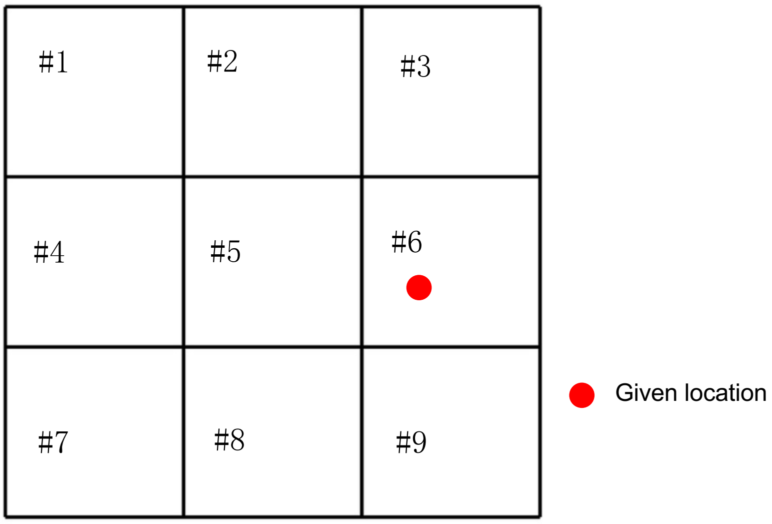


Figure 14. Principle to use tiling

However, polygons inside one tile still too many to efficient calculation, accordingly the project develop the second method. The second method called Find Nearest Polygons.

Polygons in a .shp file is stored with the coordinates of each vertex and can be extract by Python library: Fiona. The idea of the second method is for each potential location, calculate the average distance between it and all polygons in the tile is calculated. The average distance here is sum of all vertex (one polygon) distances to a potential location, which is divide by number of vertices. This method will return only nearest 20 polygons for inside/outside calculation. The inside/outside check could be implemented by the Python library shapely with the “within” function. The points within the 20 nearest polygons will be removed and the others will still remain.

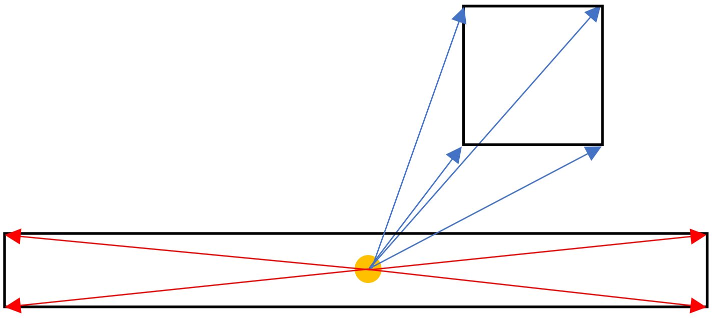


Figure 15. Example of a failure when using method 2

Ideally, only the nearest polygon needs to be checked inside or outside. However, since the definition of distance between a polygon and a point is vague and the average distance is not true distance, the method could cause failure if return only a few nearest polygons are returned. An example is shown above. Red arrows and blue arrows represent the distances to each polygon's vertices respectively and all red arrows are longer than all blue arrows which means that the average distance of the rectangle is larger than the square. Actually, it's not going to happen no matter what the distance means because the point is already inside the rectangle. Besides, the polygon's shape in the Hague could be much more complex than the example mentioned above. According to this, this method needs to return 20 even more polygons for inside/outside test. One result (radius = 25m and density = 1m) is shown in Figure 16. It's noticeable that the given sensor location is inside the building, hence being removed.

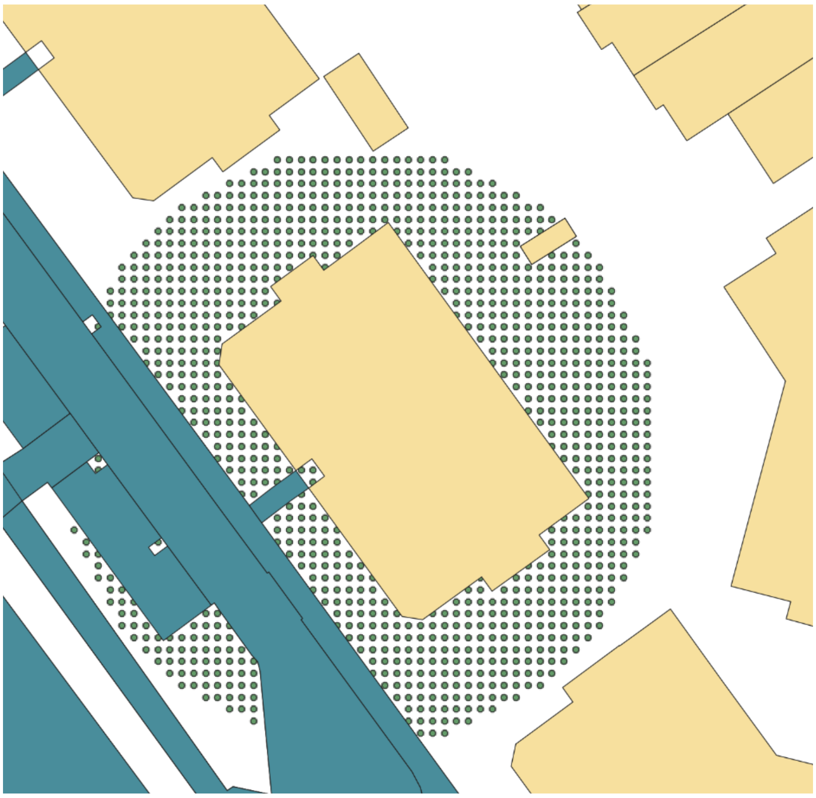


Figure 16. Result of generating potential sensor locations

Yellow parts are buildings and cyan parts are transportation areas.

4.4 Computing sky view (dome) and solar parameters

In order to know, for each potential location, when it receives solar radiance, sky view and sun position will be computed. Dome reconstruction is done by Urban Horizon project [24] and this project will add functions to their work to calculate the sun's position and the time a position receives direct solar radiance. The principle is to know when then sun will not be blocked by grey (buildings) and green (vegetation) part in Figure below.

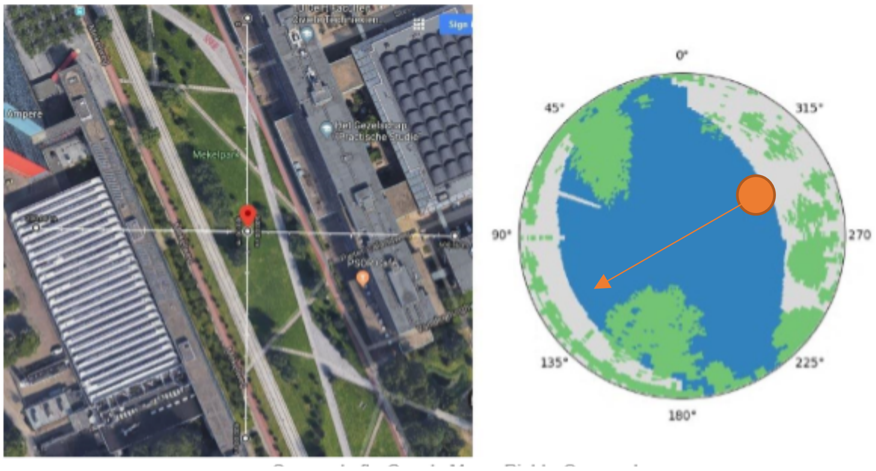
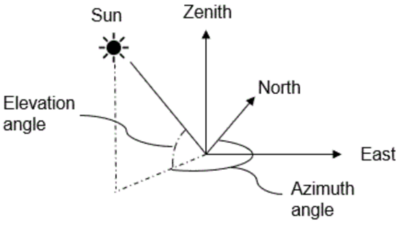


Figure 17. left: one selection point;

right: dome output from Urban Horizon project and orange point means sun' s position

4.4.1 Solar parameters

In order to know the Sun's position in the dome, the elevation angle as well as the azimuth angle must be known. The solar zenith angle is the angle between the zenith and the center of the Sun's disc. The solar elevation angle is the altitude of the Sun, the angle between the horizon and the center of the Sun's disc [30].



azimuth angle: north=0, east=90, south=180, west=270 degree

Figure 18. Elevation angle and azimuth angle of sun [31]

The elevation angle and azimuth angle cannot be computed directly from the local time and coordinates but are available from other intermediate parameters: hour angle, declination, solar time and equation of time.

The declination of the Sun is the angle between a plane perpendicular to a line between the Earth and the Sun and the Earth's axis. An approximate formula for the declination (D) of the Sun is [32]

$$D = 23.45 / 180 * \sin (2 * (284 + n) / 365)$$

n = number of days starting from 1^{th}, January ($n = 1$ now)

As the Earth moves around the Sun, the solar time changes slightly with respect to the local standard time. (This is mainly related to the conservation of angular momentum as the Earth moves around the Sun.) This time difference is called the equation of time and can be important when determining the position of the Sun for solar energy calculations. An approximate formula for the equation of time (E_{qt}) in minutes is [33].

$$E_{qt} = 4.0 \sin((n - 106) / 59)$$

for year day n between 107 and 166

$$E_{qt} = -6.5 \sin((n - 166) / 80)$$

for year day n between 167 and 246

$$E_{qt} = 16.4 \sin((n - 247) / 113)$$

for year day n between 247 and 365

To describe the position of the Sun in local standard time, one needs to know the relationship between solar time and local standard time. Local time is the same in the entire time zone whereas solar time relates to the position of the Sun with respect to the observer, and that is different depending on the exact longitude where solar time is calculated [34].

$$T_{solar} = T_{local} + E_{qt} / 60 + (Long_{local} - Long_{sm}) / 15$$

T_{local} = local time; $Long_{local}$ = local longitude; $Long_{sm}$ = central longitude of time zone

Hour angle (ω) is one of the coordinates used in the equatorial coordinate system to give the direction of a point on the celestial sphere. The hour angle of a point is the angle between two planes: one containing Earth's axis and the zenith, and the other containing Earth's axis and the given point [35].

$$\omega = \pi * (12 - T_{solar}) / 12$$

With the above information, one can now calculate the cosine of the zenith angle and elevation angle [36]:

$$\sin (E) = \cos(Z) = \sin (I) \sin (D) + \cos (I) \cos (D) \cos (\omega)$$

I = the latitude of the location; E = elevation angle; Z = zenith angle

4.4.2 Merging solar parameters into sky view

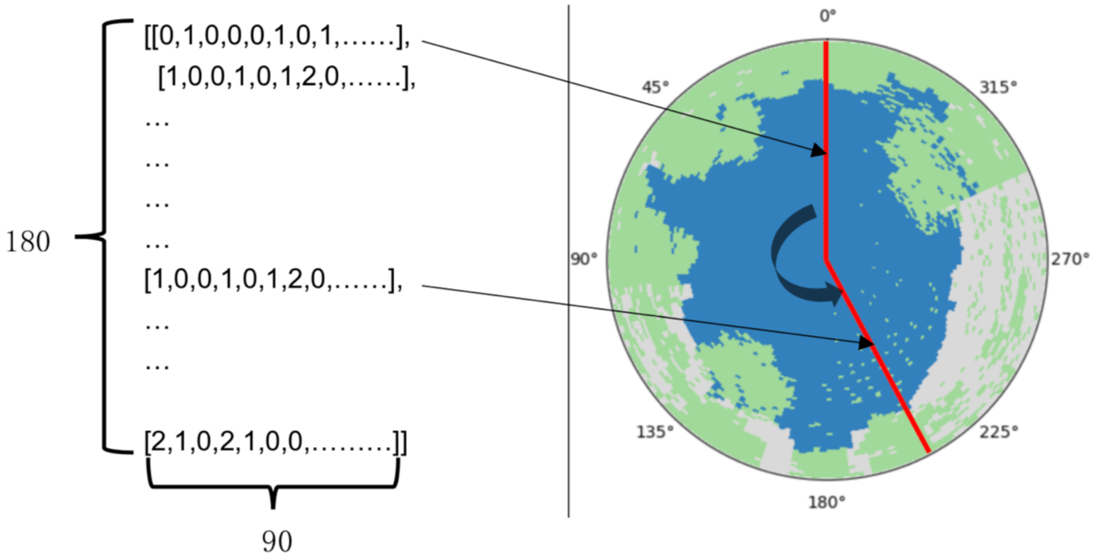


Figure 19. Data structure of sky view

The data structure used in the sky view is shown in Figure 19. The right dome circle is represented by a nested list as is shown above. Each inner list is one radius of the circle and has 90 items and these items are pixels (value is type of pixel: sky, building or vegetation). The outer list consists of 180 inner lists which means the whole circle is made by 180 rad (counter-clockwise).

Accordingly, for showing the Sun's position in the sky view, the position of the item in the nested list, which represents sun, must be computed based on solar parameters. The formula is shown below:

$$s\_z = Int\,(180 - Z\,/\,2)$$

$$s\_e = Int\,(90 * \cos\,(E\,/\,180 * \pi))$$

s_z = index in outer list; s_e = index in inner list; E = elevation angle; Z = zenith angle

An example of the merging result is shown in Figure 20. The white dashed line means Sun’s position every 15 minutes (same as sensor recoding frequency) between sunrise and sunset. The corresponding solar influence result is:

{5.5: 'not influenced', 5.75: 'not influenced', 6.0: 'not influenced', 6.25: 'not influenced', 6.5: 'not influenced', 6.75: 'not influenced', 7.0: 'not influenced', 7.25: 'not influenced', 7.5: 'not influenced', 7.75: 'not influenced', 8.0: 'not influenced', 8.25: 'not influenced', 8.5: 'not influenced', 8.75: 'not influenced', 9.0: 'not influenced', 9.25: 'not influenced', 9.5: 'not influenced', 9.75: 'not influenced', 10.0: 'not influenced', 10.25: 'not influenced', 10.5: 'influenced', 10.75: 'influenced', 11.0: 'influenced', 11.25: 'influenced', 11.5: 'not influenced', 11.75: 'influenced', 12.0: 'influenced', 12.25: 'not influenced', 12.5: 'influenced', 12.75: 'influenced', 13.0: 'influenced', 13.25: 'influenced', 13.5: 'influenced', 13.75: 'influenced', 14.0: 'influenced', 14.25: 'not influenced', 14.5: 'influenced', 14.75: 'influenced', 15.0: 'influenced', 15.25: 'influenced', 15.5: 'influenced', 15.75: 'influenced', 16.0: 'not influenced', 16.25: 'influenced', 16.5: 'influenced', 16.75: 'influenced', 17.0: 'influenced', 17.25: 'influenced', 17.5: 'influenced', 17.75: 'influenced', 18.0: 'influenced', 18.25: 'not influenced', 18.5: 'not influenced', 18.75: 'not influenced', 19.0: 'not influenced', 19.25: 'not influenced', 19.5: 'not influenced', 19.75: 'not influenced', 20.0: 'not influenced', 20.25: 'not influenced', 20.5: 'not influenced', 20.75: 'not influenced', 21.0: 'not influenced', 21.25: 'not influenced', 21.5: 'not influenced', 21.75: 'not influenced'}

(xx.25 = xx:15; xx.5 = xx:30; xx.75 = xx:45; xx is hour)

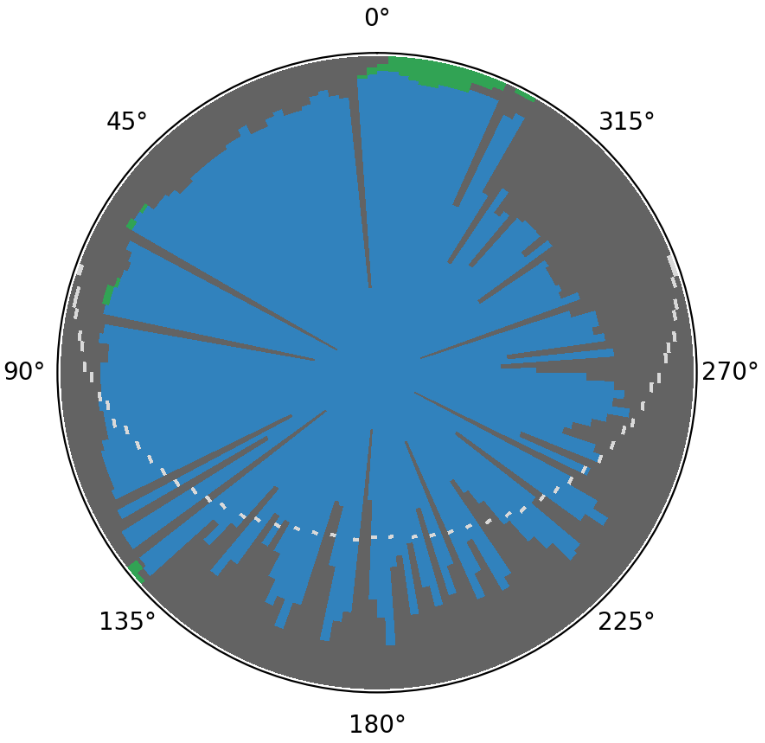


Figure 20. Merging the sky view and the sun’ s position

4.5 Finding the most likely horizontal location of the station

The idea of this section is by comparing the result from Section 4.2 to 4.4 for each sensor. More specifically, for a given sensor and its several days' data, the project will detect when it may suffer from solar radiance and its possibility (Section 4.2). Then, for each potential location of the sensor (Section 4.3), the project uses sky view with solar parameters to simulate time when each potential location suffers from solar radiance (Section 4.4). This section aims to compare the sensor data with the simulation results and return 10 highest similarity points. These 10 points will be used in the next section for generating the height value.

For example, if the result derived from Section 4.2 is "7:50 am to 2:50 pm the sensor will detect higher temperature than it should be", then the 10 potential locations whose output from Section 4.4 are the closest to "7:50 am to 2:50 pm" will be set as the candidate horizontal locations of the Netatmo station. A brief flow chart is shown below.

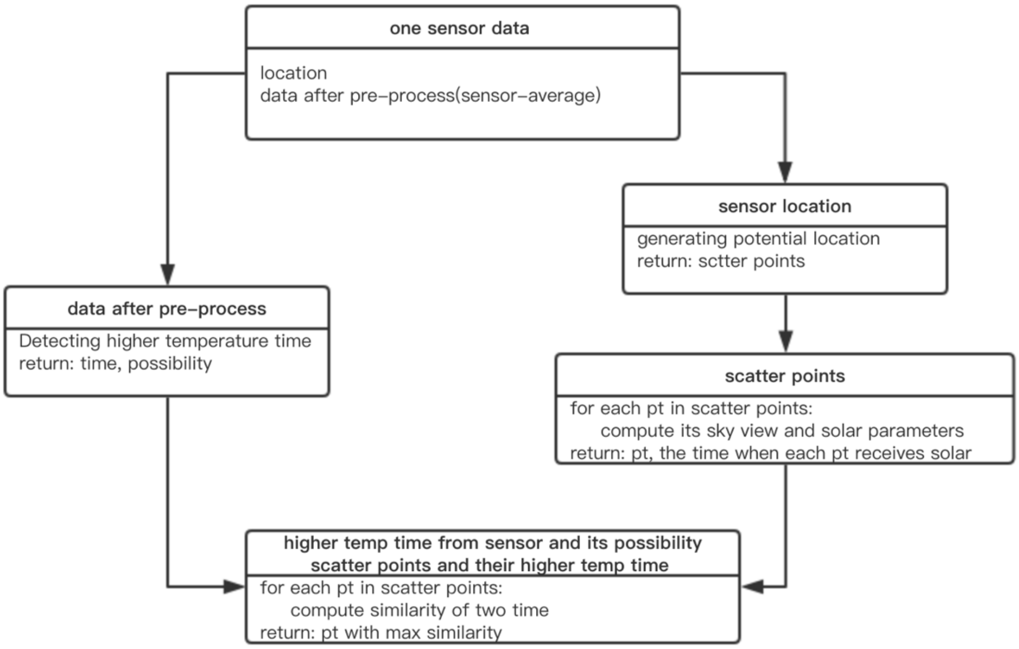


Figure 21. Idea of finding the most likely location of the station

Technically, only the stations where an "increase pattern" is found in Section 4.2 will be take into consideration from Section 4.3 to 4.5. Those without an "increase pattern" could means the station is in the shadow all the time. However, due to the complexity of temperature records, the average temperature could still not represent real temperature, so it is hard to find a threshold to determine which sensor are actually inside the shadow. So, all sensors are regard as "influenced by solar radiance" and this will be discussed in the discussion chapter.

4.5.1 Similarity calculation

Because temperature differences (average – records) from every day do not show an identical pattern, the project introduce “increase possibility” (see Section 4.2) to avoid occasional temperature behaviors and takes it into consideration when computing two time’s (solar influenced time from data and solar influenced time from sky view simulation) similarity.

The time similarity calculation pseudo code developed in the project is given by:

```
For time in sky view result (see section 4.4.2):
    If time == “influenced” :
        $S_{summary} = \sum P_{time}$;
    If time == “not influenced” :
        Continue;
$S_{similarity} = S_{summary} / \sum (1)$;
Return: $S_{similarity}$
```

Where:

$S_{summary}$ = *summary of increase possibility;* P_{time} = *Increase possibility when “time”*

P = *all increase possibility;* $S_{similarity}$ = *normalized* $S_{summary}$ *(ensure this value won’t > 1))*

Here are two examples explaining the calculation steps and why the project introduces possibility here.

Example one: let’s say a sky view result from a potential location of sky view simulation is:

13.25: 'influenced', 13.5: 'influenced', 13.75: 'not influenced'

While the result from increase/decrease (Section 4.2.2) check from the same sensor is:

13.25: '0.7', 13.5: '0.9', 13.75: '0.6' (0.7,0.9 and 0.6 are increase possibilities)

Then the $S_{summary} = 0.7 + 0.9 = 1.6$; $S_{similarity} = 1.6 / (1 + 1 + 1) = 0.53$. Alternatively, if the “increase possibility” is not introduced, then the increase/decrease result would be:

13.25: 'influenced', 13.5: 'influenced', 13.75: 'influenced'

Therefore, $S_{similarity}$ = 2("influenced" in simulation) / 3 ("influenced" in data") = 0.67, which is higher than previous similarity result. However, the data itself cannot guarantee that each of these 3 times is 100% influenced by solar radiance (none gets an 1.0 increase possibility here), so the real situation of similarity should be lower than the ideal (every time "increase possibility" = 1) one.

Example two: sky view result from a potential location α of sky view simulation is:

13.25: 'influenced', 13.5: 'influenced', 13.75: 'not influenced'

Potential location β (α and β are generated from same sensor) of sky view simulation is:

13.25: 'influenced', 13.5: ' not influenced ', 13.75: ' influenced '

result from increase/decrease check:

13.25: '0.7', 13.5: '1', 13.75: '0.6'

The result will be $S(\alpha)_{similarity}$ = 1.7 / 3 = 0.56 while $S(\beta)_{similarity}$ = 1.3 / 3 = 0.43. It's reasonable that $S(\alpha)_{similarity} > S(\beta)_{similarity}$ since increase possibility at 13.5 > 13.75, which means temperature difference at 13.5 are more likely is shown increase pattern than 13.75. Therefore, α which is influenced at 13.5, deserves higher similarity than β, which is influenced at 13.75. Also, $S(\alpha)_{similarity} = S(\beta)_{similarity}$ if this possibility is not considered, hence α and β may not be distinguished properly.

Based on the two examples, it's not hard to see that "increase possibility" plays the role of weight in similarity calculation. Higher "increase possibility" means the sensors at certain time are more likely to receive radiance thus bringing greater weight.

4.6 Assign height value to point

The result before this section is 2D points with only horizontal coordinates. However, it is possible that users place their sensors on the window, or shelf (or some not high place in the garden) instead of directly on the ground. Usually, a higher view point has a bigger sky view factor and thus brings higher solar susceptibility to the weather stations. Therefore, this step is trying to assign height value to the 10 points from Section 4.5.

The approach here is similar to previous steps: vertically (at 0.5, 1.0, 1.5 and 2.0 meter) generating potential points for these 10 points and then compare their solar simulation result with data as is shown in Section 4.5.1. Technically, creating a 3D points within 3D bounding box then calculate their similarity maybe more accurate however it consumes a lot more time. For instance, if there are 100 potential points for one sensor on the ground and create 3D points at first, then total 3D points will be 500 which means 500 times dome reconstruction, alternatively, assigning heights here only needs 140 times in total. Because horizontal range (0-30 meter) of potential location is much higher than vertical range (0-2 meter), the dome variance in the vertical direction usually is smaller. Therefore, it is almost impossible that a point on the ground with low similarity could be found high similarity at the height of 2 meters. It also is the reason why this section only considers 10 points with highest similarity on the ground from last step but not all points. The picture below is shown the principle of this section.

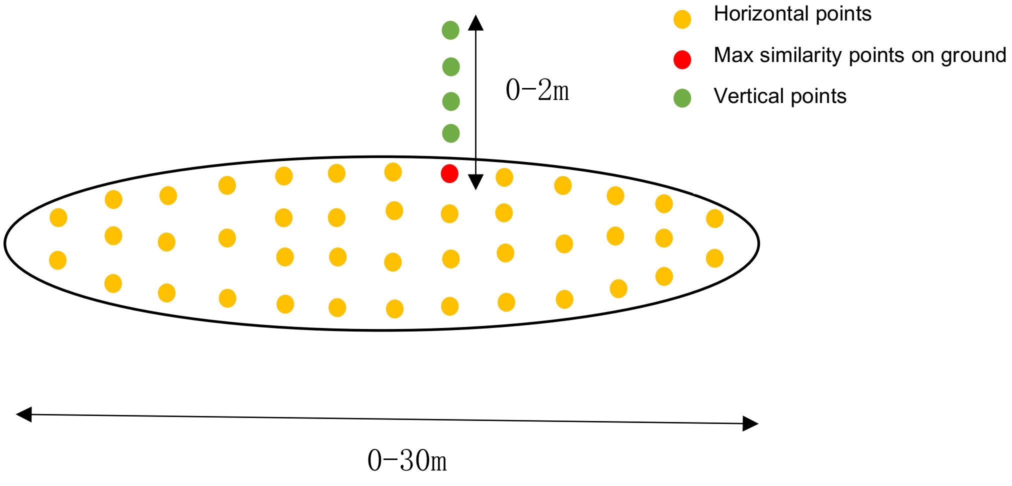


Figure 22. Assign height to the ground points

5 Result and validation

The following chapter of the research examines the principles of the similarity calculation and the relative methodology and results obtained from this work. The cleaned dataset and relocation results have been implemented and compared. Because the topic introduces a quite new approach to solve the problem, therefore, the goal of this chapter is to visualize the result and statistically analyse its rationality. Further, the two validation methods including comparing the results from different period of date and field work experiment done by an example Netatmo weather station will be shown.

5.1 Relocation of the stations

5.1.1 Overall of the dataset

After the data pre-processing step, there are 185 sensors saved in the dataset. And for each day in research period, the number of sensors that are "working normally" is shown below.

day (May, th)	22	23	24	25	26	27	28
number	177	180	162	183	175	178	170

Table 1. number sensors in the dataset per day

The main factor that leads to the variation of the sensors' number in research days is the data recording frequency. Only sensors that record the complete of 95 times temperature data in one day are qualified for further research (Section 4.1.1), however, some sensors show unstable recording behaviors, e.g. records 95 times in the 22th May but 60 times in the 23th May. The reason of this may be network connection issues or defects of Netatmo station itself but it is out of the research scope and will not be discussed further here. Although several approaches (e.g. spatial or temporal interpolation) could be helpful to fill the missing data and enable more stations enter the dataset for further analysis, it's hard to guarantee the simulation data quality, which brings more challenges to validate results. Considering this, the project tries to maintain the integrity of raw data, therefore just kick out defects in the raw data but not amend them.

5.1.2 Similarity interpolation

Interpolation is one of good ways to show similarity visualization results. In order to lower the complexity of this research, the result here is created by IDW (inverse distance weight) interpolation rather than Kriging interpolation.

According to Tobler's First Law of Geography, everything on a geographic surface is related to everything else, but near things are more related than distant things [37]. Although similarity result is not a nature phenomenon on the earth, it's actually computed based on sky view factor and temperature records, both of which are highly related to geographic surface. So, a validated similarity interpolation result should more or less obey this law and is shown smooth changing trend or some specific patterns in the interpolation map.

Due to context limitation, this section only is shown four similarity IDW interpolation maps based on different weather stations for analysis. The results are implemented with IDW tool embed in Qgis3 and with default interpolation parameters.

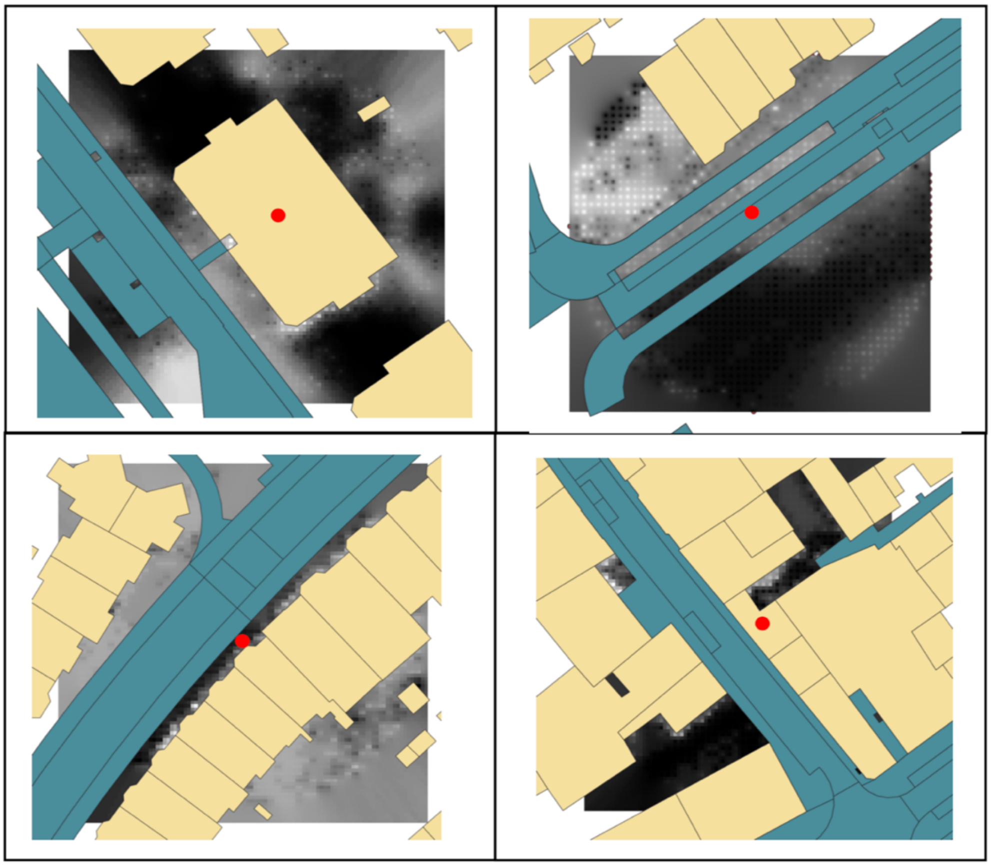


Figure 23. Interpolation result of four weather stations (red dot is given location)

70:ee:50:29:39:12(top left); 70:ee:50:03:db:a4(top right); 70:ee:50:1f:42:c8(bottom left); 70:ee:50:1a:fb:3a(bottom right)

White: higher possibility; Black: lower possibility

As shown in the picture above, although all 4 interpolation maps do not show very smooth changing trends (this may result from uneven distribution but high density of scatter points), the high possibility blocks and low possibility blocks are also not totally random. Besides, gradual changes between two blocks are obvious. In addition, high similarity pixels appear next to the edge of building polygons, which matches with reality. Usually, users won't put their own weather stations too far away from their home.

5.1.3 Relocation result

The whole relocation process is done by python3, among which, packages including shapely, fiona have been using to tackle with shapefile data; osgeo has been using to coordinate transformation; numpy and laspy has been using to speed up matrix calculations. The processing time for each sensor is about 3-5 minutes and whole running time to get full result for all sensors is about 12 hours (based on MacOS, i5-8259U, 8G RAM).

The picture below is shown the new location of all sensors after processing (obviously not much difference from Figure 6 because several meters differences are hard to see in the city scale map) and details about their coordinates and height value and similarity could be found in the appendix.

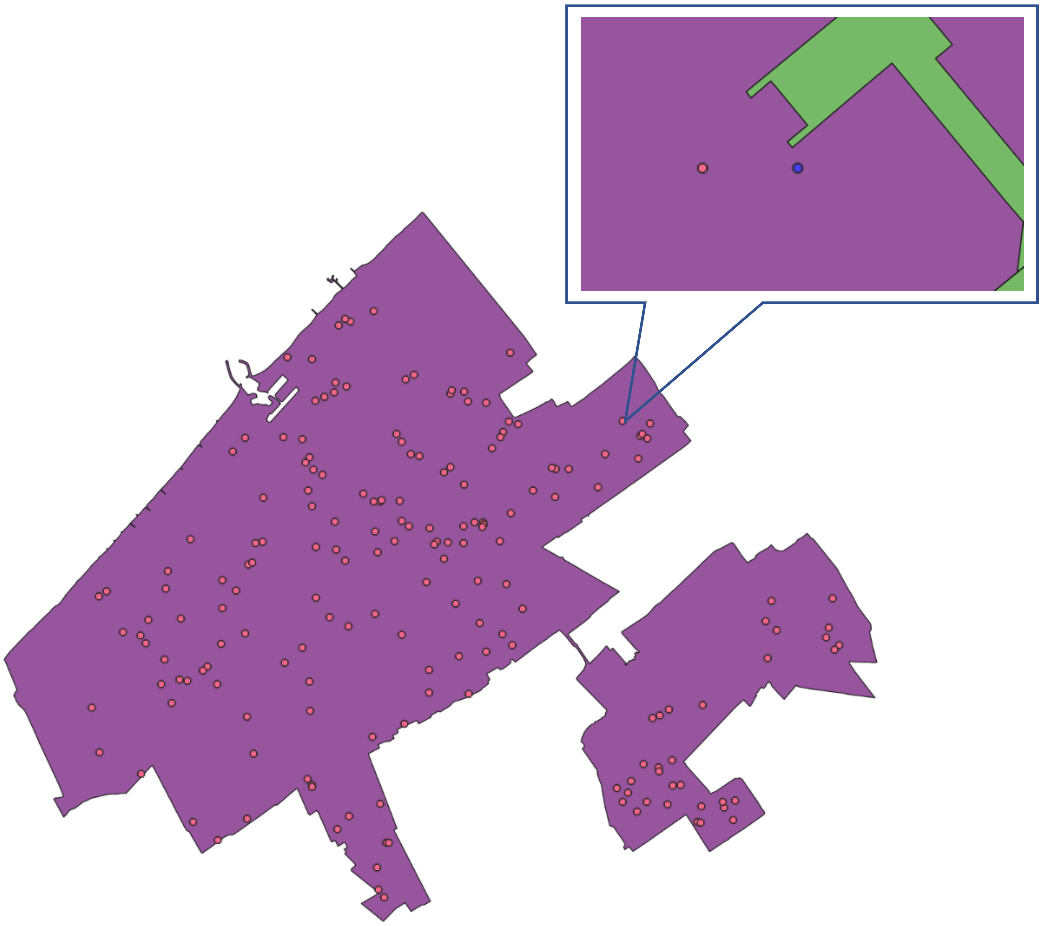


Figure 24. New location of the stations and one zoom in example

Orange dot = new location, blue dot = old location

The statistics distribution of the similarity result of all 185 sensors can be seen in Figure 25. The bar is shown the number of sensors at each similarity interval which ranges from 0.061 to 0.961 and the width of each interval is 0.1. The number of sensors goes up with the similarity and reaches the maximum at [0.661, 0.761], where 59 sensors are, then decrease sharply to 33 at the next interval and only 3 sensors are in the highest similarity interval.

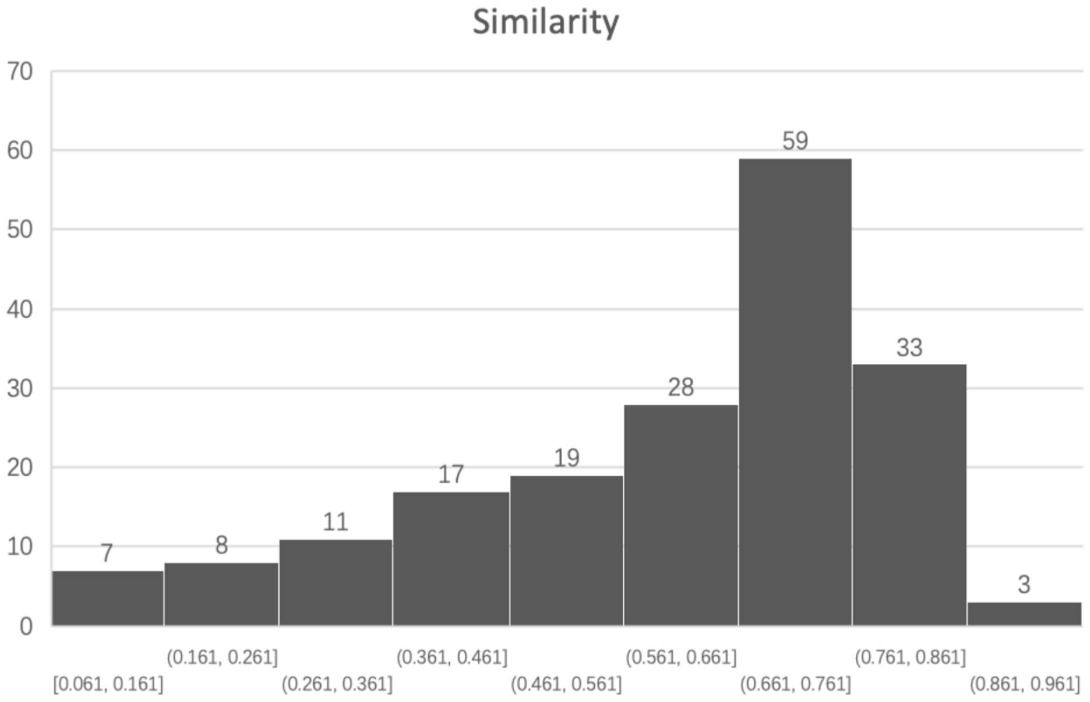


Figure 25. Distribution of similarity result

There are several explanations for the low similarity result. For example, it is quite possible that not all stations are expose to solar radiance and records from those who stay in shadow do not match the average temperature pattern, so the fake increase temperature pattern will appear in the calculation process. If this kind of fake pattern cannot fit the solar simulation result, then the similarity will accordingly be very low.

Another explanation is that the potential location failed to cover the real location of the stations. Usually, stations are placed in the garden behind their house, a place outside buildings and transportations areas and this is the principle how the project create potential location. However, this could be wrong if users live in an apartment not a detached house: apartment usually will not have a garden and even if, it would be more reasonable for users putting stations in their balcony. Also, in the BGT shapefiles, apartment polygons are very close to transportation area polygons and this means that there is not much space for creating potential locations and even more hard to guarantee their reliability.

A corresponding example is shown below (Figure 26): the MAC address of the sensor is 70:ee: 50:04:76:90 and its similarity result is 0.115.

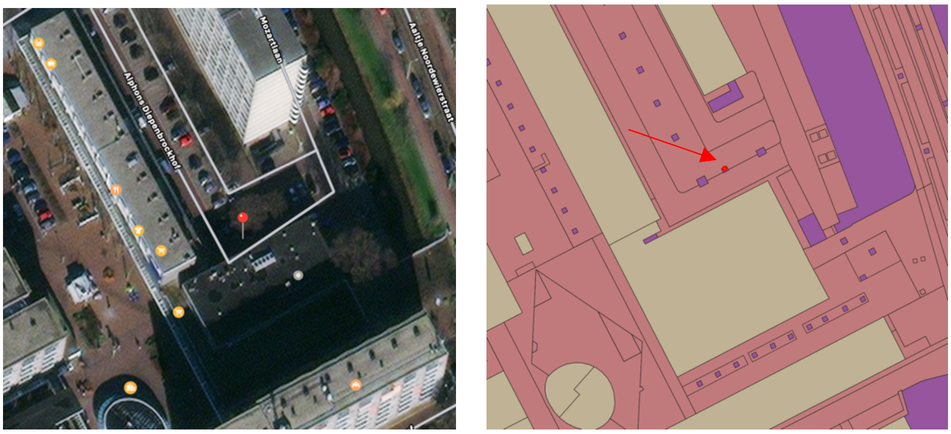


Figure 26. Station 70:ee: 50:04:76:90’ s given location in satellite map and shapefile

As in picture is shown above, there is not much space between the buildings (yellow polygons) and the transportation areas (pink polygons). Although an open space (purple polygon) on the right could be used for generating potential location, actually it is a public green land (left Figure) so there is a slim chance the user has put a station there. In this complex environment, determining an area for potential location would be extreme tough.

5. 2 Validation

Validation is an essential procedure through the re-location processes. Validation verifies that the new location represents a situation that is close to the reality. Unfortunately, because of privacy, making a Netatmo user survey is impractical. In order to solve this, the project will use two separate methods to validate the result, which are described in the sections below.

5.2.1 Comparison of result from two period of 7-days

The idea here is checking the results from different periods of date. More precisely, the methods mentioned above are all based on seven-day data, so every consecutive seven-day with good weather condition could be a control group. The sensors are fixed so the result of sensors’ location from different periods should be more or less same. If the location difference of each sensor is less than some threshold then it’s reasonable to say the method is validated.

The period for validation is from 17th, April 2018 to 23th, April 2018. The validation period of time also has a good weather condition in general [24] and is not far from research period, which could ensure the city environment won’t change too much and the all working sensors are very similar. The general weather data and working station numbers are integrated into Table 2. Besides, the Sun’s positions of two periods still have noticeable difference (Figure 27), therefore this period would be ideal for validation.

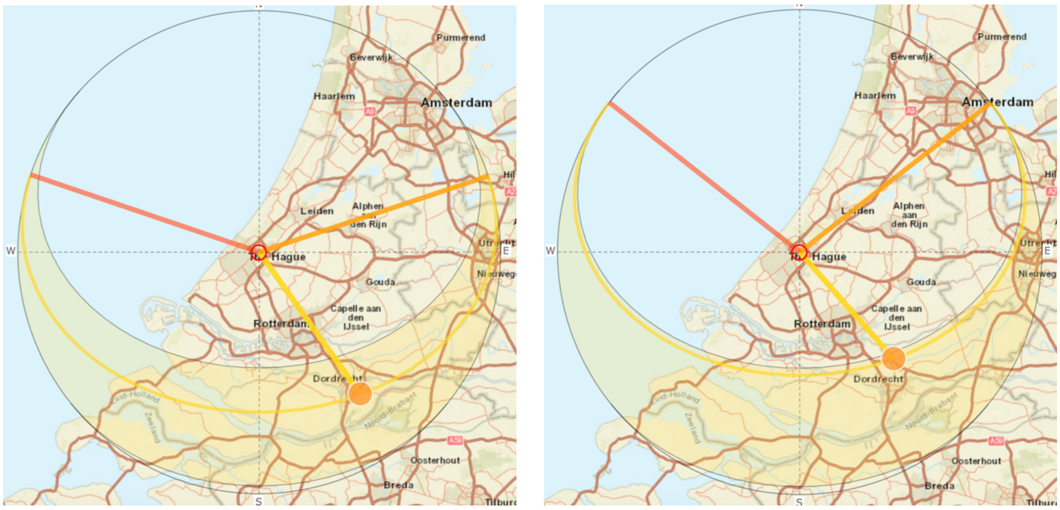


Figure 27. Sun' s position at validation period (left) and research period (right) [38]

Yellow arc is sun' s track and orange dot is sun' s position at 12:00 pm

research period					
Date (May)	working sensors	weather/temp(C°)			
		mid-night	morning	noon	afternoon
22th	177	clear,15-19	clear,15-20	clear,20-23	clear,18-22
23th	180	clear,14-18	mild,15-19	clear,18-24	clear,18-23
24th	162	clear,14-18	clear,14-19	clear,17-20	clear,18-21
25th	183	fog,16-17	fog,15-22	clear,22-23	clear,15-20
26th	175	fog,16-17	clear,16-26	clear,26-29	clear,19-28
27th	178	clear,17-20	clear,18-25	clear,23-25	clear,20-23
28th	170	clear,17-20	clear,17-27	clear,27-29	clear,22-29
validation period					
Date (April)	working sensors	weather/temp(C°)			
		mid-night	morning	noon	afternoon
17th	180	clear,8-9	clear,9-17	clear,17-19	clear,13-19
18th	176	clear,11-12	clear,12-21	clear,21-23	clear,13-22
19th	187	clear,13-17	clear,13-24	clear,24-27	clear,15-23
20th	173	clear,12-15	mild,14-23	fog,21-24	mild,10-18
21th	174	fog,8-10	mild,8-17	clear,18-22	clear,12-20
22th	169	clear,11-14	clear,11-29	clear,21-25	clear,16-20
23th	159	clear,11-15	cloudy,9-14	clear,13-14	clear,10-14

Table 2. Weather and stations condition of two period of result

horizontal distance (m)	Amount	height distance (m)	Amount
<5 (high quality)	130	0	95
		0.5	21
		1	8
		1.5	4
		2	2
average (m)	1.451	average (m)	0.219
5-10 (mid quality)	25	0	14
		0.5	5
		1	1
		1.5	4
		2	1
average (m)	7.427	average (m)	0.460
10-15 (low quality)	8	0	3
		0.5	3
		1	0
		1.5	1
		2	1
average (m)	12.020	average (m)	0.563
>15 (no quality)	24	0	7
		0.5	10
		1	2
		1.5	2
		2	1
average (m)	21.799	average (m)	0.545

Table 3. Comparison of two period of result

The comparison new location of stations from two 7-days data are shown in Table 3. The horizontal and vertical distances of each sensor’s location in research period and validation period have been calculated. The project divides the results’ quality into 4 levels based on horizontal distance: <5m is high quality; 5-10m is mid quality; 10-15m is low quality and >15m is no quality. The sensor amount and their average distance of each quality level are shown in the table. Also, each quality level will be sub-divided into 5 sections based on the vertical distance. However, since height range is significantly shorter than the width range as discussed in Section 4.6, the vertical distance will only be a secondary reference but not a quality standard.

Among 185 Netatmo stations, over 70% stations show high quality and the average horizontal distance of them is less than 1.5m. Also, height distance of high-quality stations is shown a good result, most of which are at identical height. The following quality levels are mid quality and no quality, the amounts of which are 25 and 24 respectively. The sensors classified into low quality are not many.

It is noticeable that similarity and quality are two different notions. Similarity is computed based on temperature records and solar simulation while quality is computed by horizontal distance for each sensor over two time period. Neither high similarity can guarantee high quality nor low similarity means low quality. Table 4 is shown the correlation between similarity result (research period) and its quality at 4 quality levels. All of them indicates a weak relationship between each other. (The similarity and quality interpolation maps in city scale can be found in the Appendix E. The paper will not analyse these interpolation results further because massive points there and irregular visualization in the maps.)

quality level	<5	5,10	10,15	>15
correlation	-0.0072	0.0301	0.0304	-0.0228

Table 4. Correlation of similarity (research period) and quality

From a macro level, similarity and quality have different interpretations and almost zero dependency, hence they could be regarded as attributes of the result or also two dimensions in classification. For example, high similarity with low quality means that more than one place has good chance to become real location of a station thus the result is not trustworthy; low similarity with high quality means that no place is suitable for the station but it is the real situation. A metaphor here to explain this would be “Get high marks by cheating & Failing but not cheating”. The idea is briefly shown in Table 4 and in order to lower the context complexity, different from the above, both similarity and quality are only divided into 2 level here. In this classification case, stations will be located in 4 quadrants and the number in each quadrant are also shown in Table 5.

similarity\quality	Low (>10m)	High (<10m)
low (<0.5)	12	38
high (>0.5)	18	117

Table 5. Classification by similarity (research period) and quality

In order to look deeper into stations with poor result quality, several of them has been selected for analysis. The pictures below show these stations’ locations (original location, new location from two periods data) in a shapefile map and a corresponding satellite image.

The first station is 70:ee:50:02:88:c0 (Figure 24) and the horizontal distance from two period result is about 20 meters. The given location of this station is inside a street thus being removed. Based on the research period, the station has been relocated to the left side of the street where there is a public greenland is shown in the satellite image. On the contrary, validation period data shows that the station should be on the right side of the street and just close to the house, which is a much more reasonable location.

The reasons causing the result from the research period not being as reliable as validation period could be various. One explanation is that the public greenland is not suitable for placing potential locations but it is not being removed in the shapefile. Another explanation could be, for some reason, the temperature difference in the research period is not suitable for representing the real temperature changing pattern.

In this case, the problem can be solved by introducing more than two periods of data for researching new a location and more restrictions for potential location need to be considered.

Figure 28. Result from two period, sensor: 70:ee:50:02:88:c0

3 dots on right picture are 3 locations, blue=given location, red=research period, yellow=validation period

Another station, 70:ee:50:02:9d:ee, with 42 meters (biggest in the whole dataset) divergence is shown in Figure 25. The given location is next to the corner of a theater; location computed by research period is inside a narrow gap of an open square ahead of the theater; location of validation period is next to a corner of another building, and it is hard to tell which one is shown a better prediction.

When man-made objects make the environment very complex, the real situation is not fully compatible with the shapefile map. Open spaces around given location in the shapefile are gaps, holes or corners in the square and it’s unlikely to create potential locations there. However, the suitable places, e.g. sides of the theater, are all covered by transportation areas thus no potential location will be here. Actually, even for human eyes, a rule-making process to determine which place the sensor might be is difficult in this kind of commercial, block nevertheless for a program.

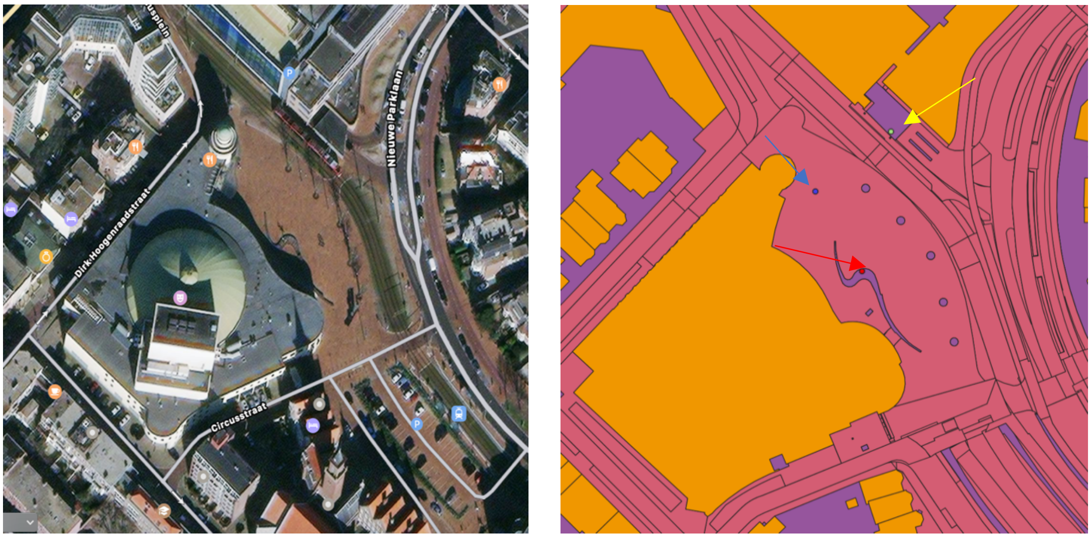


Figure 29. Result from two period, sensor: 70:ee:50:02:9d:ee

blue=given location, red=research period, yellow=validation period

The final example is station 70:ee:50:14:5f:f0 with around 10 meters divergence is shown in Figure 30. Although 10 meters is not so low in the quality validation, one issue could not be ignored. The buildings part in the shapefile map cannot match the satellite image at all. This could result from the unstable data quality of BGT itself or these blocks were built before the BGT data update.

Figure 30. Result from two periods, sensor: 70:ee:50:14:5f:f0

blue=given location, red=research period, yellow=validation period

5.2.2 Experiment by sample Netatmo weather station

This section aims to introduce a sample Netatmo weather station for experiment. The experiment will place the station at a known place and validate the relocation process by checking whether the adopted algorithm is able to find the sensor's location. The new location is in Hooikade 26, 2627 AB, Delft on a 0.5 meter-height table in the garden behind this building. Figure 31 is shown the details of the sensor's location.

Figure 31. location of experimental sensor (satellite images from google map)

The experiment started on 13th, May until 19th, May, 2019. The weather condition of this period is shown in the table below. The temperature records collected from the experimental station shows very abnormal record frequency in the sever and the reason for this is still unclear. Alternatively, the records are download from the private station management portal. Other data in the Delft still comes from its public API. Private portal provides weather data every 5 minutes while 15 minutes in the sever, thus the experimental station record frequency is switched into 15 minutes manually.

Date (May)	weather/temp(C°)			
	mid-night	morning	noon	afternoon
13th	clear,2-6	mild,2-13	clear,13-15	clear,8-14
14th	mild,4-8	mild,4-13	clear,13-16	clear,10-16
15th	clear,6-10	clear,6-15	clear,17-18	clear,11-17
16th	clear,5-12	clear,6-14	cloudy,13-18	clear,11-16
17th	cloudy,10-12	cloudy,10-13	mild,13-15	cloudy,12-14
18th	mild,6-12	clear,7-16	clear,17-20	clear,12-17
19th	mild,12-14	clear,12-16	cloudy,12-16	mild,11-12

Table 6. Weather condition in the experiment period

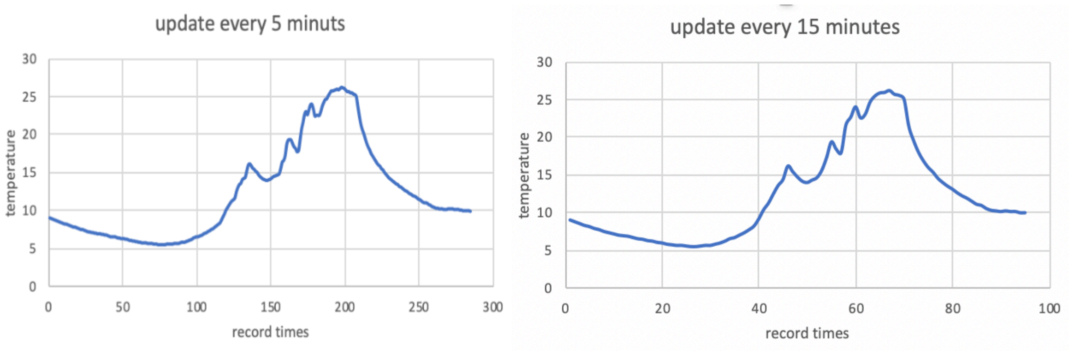


Figure 32. Temperature records on 16th, May in two recording frequency (left: 5 min; right: 15 min)

Figure 32 shows temperature changing pattern on 16th, May in different recording frequency (5 minutes and 15 minutes). Generally, the lower frequency is shown a bit more smoothness while doesn't lose much data changing characteristic, thus, 15 minutes, the recording frequency in the sever, is sufficient on this research.

Also, Figure 32 also shows the impact from the cloud space. When the station is placed without any shield, the temperature records will be influenced not only by the solar radiance but also the cloud distribution above its dome. The weather condition in this noon is cloudy (Table 6) and this makes the middle of two lines (Figure 32) up and down obviously. The cloud would protect the station free from direct radiance and records could closer to the real air temperature, but cloud appearing from time to time makes the detected temperature change unevenly and bring more challenges in analyzing the fluctuate pattern.

Another finding in this picture is, the temperature records error could be higher than 1-2 ℃ as the paper mentioned above. The weather report indicates that the highest air temperature on 16th, May will not exceed 20 ℃ but the detected temperature actually could reach around 26 ℃ (Table 6 and Figure 32). The short-time increasing temperature could be 1-2 ℃ but when the station is exposed to the Sun in long time, heat will accumulate at the metal shell of the station and cause the temperature keep rising.

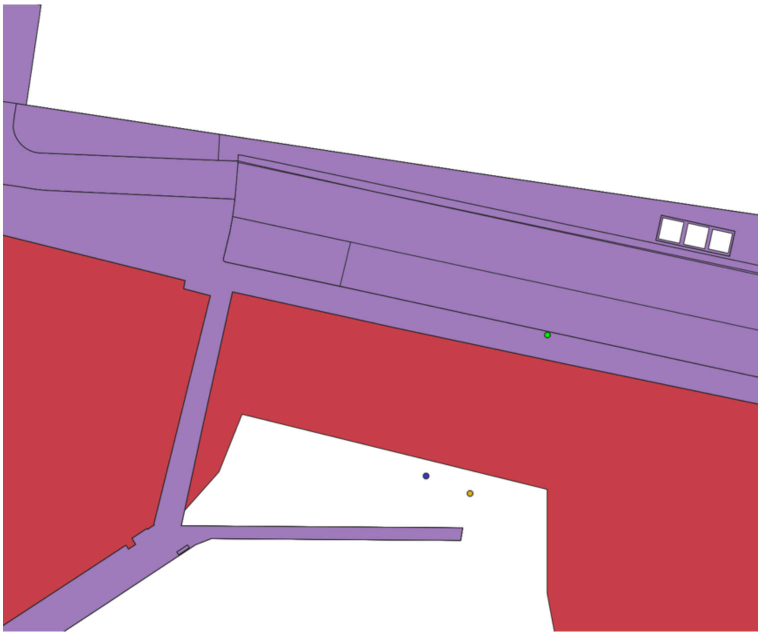


Figure 33. Location of the experimental weather station

green: default location; blue: calculation location; yellow: real location

red polygons are buildings and purple polygons are transportation areas

The final result of this experiment is shown in Figure 33. The 3 points represent 3 locations of the weather station (default location; calculation location; real location). The sensor has been located into the road in front of the building (the sensor's location is collected inside the building when the station is setting up as most of users did). The calculation location is in the backyard, which is showing a more reliable result and also closer to the real location. However, there still is an around 4 meters gap between the result and real location. The gap may result from the imperfect weather condition in the experiment period: not all days are fully sunny and clear.

Overall, this validation experiment is shown a real example of fluctuated air temperature and how it will be influenced from solar radiance. When the station is exposed to sun in a long time, heat will accumulate at the station and cause air temperature difference more than 1-2 ℃ as Netatmo mentioned. Although the real situation does not coincide with experimental result, the location difference is reduced from 16 meters to 4 meters roughly. Therefore, in terms of this experiment, the methodology adopted by the project is helpful to improve the station's location accuracy.

6 Conclusions and discussion

This final chapter of the project gives the summary of the obtained results. Also, the answers to the research questions are given, as well as the conclusions and discussion. Discussion section will analyse relative demerits of the methodology and future work will be implemented.

6.1 Answers to the research questions

How to find potential locations for each sensor?

Although the coordinate of each station is given, it’s actually a rough location info and the real horizontal location could be any point inside a buffer of the given location. The radius of the buffer is the accuracy of the given location. Another section of the buffer is buildings and transportation areas. All temperature data used in this project are collected from the Netatmo station outdoor module. Likewise, no users will put sensors in transportation area. So, the part of the buffer that covers buildings and transportation areas will not be considered when generating potential locations.

In order to achieve this, the project creates points with even density inside a bounding box, then removes points whose distance to the initial point is longer than the radius. Any remaining point that is outside road and building polygons will be considered as horizontal (on the ground and height = 0) potential point. Due to the thousands of building and road polygons in the Hague, the calculation whether each point inside each polygon consume very long time. Thus, two methods (tiling and find nearest polygon) were develop here in order to lower the calculation complexity.

After this, each ground point similarity value will be computed based on sensor data and solar simulation. The 10 highest similarity ground points will be returned for creating height value and then merging the height and horizontal location to 3D coordinates.

How to know if a sensor’s record is higher than it should be?

Because of the lack of a control experiment for each location, it is almost impossible to get the real air temperature around their environment. Instead of real air temperature value, the project uses temperature changing pattern as a substitute and this is done by daily average temperature collected by Netatmo weather stations in the Hague. Although the absolute value (average temperature) here is not accurate, the temperature changing pattern is more or less reliable because the average relieves the outliers influence from the solar radiance. The time sensors are exposed to solar radiance is not identical and all sensors are taken into consideration with same weight therefore, outliers could be “diffuse”.

After that, it's possible to know when the temperature is higher than it should be by using record data minus average data, following an algorithm to classify these result points as increase, decrease or neither. Increasing result points means they are higher than air temperature and likely to be exposed to solar radiance. The algorithm is based on conditional classification of the 3 gradients from the consecutive 3 and is trying to avoid some influence from outliers to make the classification smoother.

Besides, in order to consider data collected from different dates, the project introduced a concept called "increase possibility". Only time whose increase possibility bigger than 0.5 will be considered as "the time influenced by solar radiance".

For a certain area, how to know when it receives direct solar radiance?

In order to know, for each potential location, when it receives solar radiance, the sky view and sun position are computed. Dome reconstruction is based on the AHN3 point cloud and sun's track between sunrise and sunset will be added into digital dome. The dome will be classified into 3 parts: building, vegetation and sky. The time this area is expose to solar radiance is the time sun is on the sky part of its simulation dome.

In order to know the sun's position in the dome, elevation angle as well as azimuth angle must be known. The elevation angle and azimuth angle can be computed indirectly by geographic parameters: hour angle, declination, solar time and equation of time. Difference sun's track for all sensors in the Hague at the same day is negligible. Therefore, all research the sensors share same daily Sun's track for sake of lower complexity.

The data structure of the digital dome is a nested list could also be regarded as a matrix. Each inner list (or row of the matrix) is one radius of the dome and has 90 items. These items are pixels (value is type of pixel: sky, building or vegetation). The outer list (or column of the matrix) consists of 180 inner lists which means the whole dome is made by 180 radii. Accordingly, for showing the Sun's position in the dome, the corresponding index in the nested list is computed based on solar parameters.

How to compare a station records with solar simulation?

For a given sensor and its several days' data, the project will detect when it may suffer from solar radiance and its possibility. Then, for each potential location of the sensor, the project uses the sky view with solar parameters to simulate the time when each potential location suffers from solar radiance.

The similarity of the sensor data and the simulation result will be computed for each potential location. Because the temperature differences (average – records) from every day do not show an identical pattern, the project introduces the “increase possibility” to avoid occasional temperature behavior and take it into consideration when computing the similarity.

6.2 Conclusion

The Netatmo weather station relocation research aims to improve the location accuracy of sensors and provide a reference time when sensors are reliable or free from solar radiance. Improved crowding source temperature data would be helpful for deeper UHI research. For example, records after correction are more suitable for UHI observation or become a more robust temperature reference for UHI modelling.

The whole project was based on the exploration of a new approach to update the coordinates for every valid Netatmo stations in the Hague. In order to do this, different geo-information and Netatmo sensor temperature records have been used. The temperature data in the Hague in May, 2018 has been collected from Netatmo weather stations through its public API. Additionally, the AHN3 point cloud for solar simulation and the BGT shapefile for creating new locations have been investigated.

The methodology of this project is divided into 6 steps: Sensor data pre-processing to remove system failures, outliers and merging 7–days data into the dataset; Detecting higher temperature time to confirm whether the stations are exposed to direct solar radiance or not; Generating potential location of stations to create candidates of the real location; Computing sky view (dome) and solar parameter are used for calculating when potential locations are receiving direct solar radiance; Finding the most likely horizontal location of the station, comparing the result from the temperature records and solar simulation then returning the most likely horizontal locations; Assigning height value to points, making previous 2D locations 3D.

The results from this research proved the feasibility and rationality of the adopted methodology. The similarity interpolation maps show that the stations are prone to be placed next to the wall of buildings, which matches with reality because usually users won’t put their own weather stations too far away from their home. In the similarity result analysis, cases that sensor’s records are totally mismatched with solar simulation are not many, and around 67% stations (new location) could is shown more than 0.5 similarity when comparing with their solar simulation. Besides, the validation result performed by the two-period comparison indicates that over 70% Netatmo stations’ new location show high quality on both horizontal and vertical after the process. The project also demonstrated that similarity and quality have different interpretations and almost zero dependency, hence they are suitable to be two dimensional attributes of classification. The experiment shows a real example of fluctuated air temperature and how it will be influenced from solar radiance and the proved that the methodology is helpful to improve the station’s location accuracy.

The unstable station performance and intricate spatial information (especially shapefile) are two main limitations of result quality. Irregular temperature records make it difficult to distinguish whether a station is exposed to direct solar radiance during a certain period of time but the deeper reason causing irregular records, artificial issues or the complicated environment itself or something else, has not been studied. The study area, the Hague, is a metropolis in the Netherlands and is filled with man-made objects. Although it is covered by the BGT shapefile and most of areas have been classified into different properties of polygons, the project still finds incompatible cases. Therefore, some of the new locations can appear where they should not be.

Finally, the research time restricts the scope and only one set of methodology has been implemented, thus no alternative solutions to address the research question could be compared. On the other hand, because the research questions embed several practical and complex problems, introducing more than one solution is likely to cause chaos in the research, but guesses about perfection of this research will be discussed in the next section.

6.3 Discussion

The obtained result of this project is shown various limitations of the applied methods, which are already recognized and incorporated in the analysis. This section is only talking about the methodology itself while defects caused by imperfect spatial information or its quality are not considered.

As the paper mentioned above, two main issues could influence data quality when using Netatmo weather station for UHI research. For one thing, the stations’ locations are not accurate enough for temperature modeling in a complex city environment. For another, sensors could generate outliers when exposed to solar radiance directly. These two things are actually highly interrelated. Knowing the accurate location of the stations could be helpful to calculate when the stations are exposed to the Sun then flite outliers, and vice versa.

The issues are more or less like “two unknown parameters but only one binary equation is given” and with the deepening of the research, uncertainty and complexity of the temperature records bring further challenges into solving the problems. Fortunately, the high number of temperature records allow the project finds some temperature changing patterns while the location information could be used to improve its accuracy is quite limited. Thus, the project’s work starts with records that then combine with solar simulation to relocate the stations.

6.3.1 Demerits of the methodology

The leading limitation is that knowing an accurate real air temperature at each sensor's location is almost impossible without massive field work experiments. The project uses "average temperature" to represent real temperature, replacing absolute temperature values by "average temperature" changing pattern. Although the pattern is relatively reliable, but still has obvious disadvantages when becoming a substitute of real air temperature. Air temperature inside a city is varied due to several surface attributes and its changing behaviors as well. Therefore, it's still hard to confirm "average temperature" patterns are suitable enough to replace absolute air temperature values in this research. For example, assuming one place has better heat retaining property than other places, then in the afternoon when the air temperature starts to drop, this place will show slower temperature decrease than average temperature and its temperature difference will start increasing. In the methodology, this case will be taken as "direct solar influenced" since the spoofing increasing pattern is shown up, which may not be the real situation.

Another limitation results from the unknown air temperature for stations who stay in shadow all working time and completely free from solar radiance influence. Strictly speaking, the whole project is developed for "solar influenced" stations. Those "unpolluted" stations are impossible to locate when the project's approach is applied, on the other hand, their records are reliable therefore, further correction for temperature data is not needed. Ideally, for "unpolluted" records, they should show similar changing pattern and "temperature difference" should be a flat line. Since the temperature records vary with not only place but also time and date, almost no consistent flat line is found in the "temperature difference" dataset. Setting a threshold for the range of float up or down could be a solution and "temperature difference" curve within the threshold will be regarded as flat line. However, the lack of a control temperature experiment makes setting this kind of threshold subjective and arbitrary. So, even knowing it is a long shot, the project still assumes all sensors are "polluted".

Further, the utilization of spatial information used to compute the solar simulation and potential location could be improved. For the solar simulation, the project only considers if the weather is sunny or clear in choosing a research period while the cloudscape is not in the research. In reality, the dynamic cloud distribution in the sky view also should be taken into account to get more accurate simulation results. For a potential location, the points inside building polygons will be removed and its prerequisite is that the user lives in a detached house and places the station in the garden. However, the result is shown the methodology is not working when user lives in a high-rise apartment and places their station at a balcony. Also, more restrictions for potential location has not been taken into account e.g. public green land, water areas.

6.3.2 Future work

The uncertain data quality of crowdsourced temperature data limits its convenience and requires more data processing when applied to scientific uses. However, this does not mean that the crowdsourced data destined to serve civilian uses only. The result chapter already is shown the potential of increasing its quality and for scientific UHI research. Due to limited research time and paper length, only one prototype methodology to improve location accuracy has been implemented. Several improvement schemes could be helpful to improve the methodology.

The top priority, as the thesis has mentioned, would be more field-work experiment research. An experiment that looks deep into Netatmo weather station mechanism and working principle is helpful to realize the reason why it generates outliers and how to filter them out. Also, the experiment would be a baseline test for influence from direct solar radiance and a reference for developing a “solar influence removal” algorithm. Further, the project only finished solar simulation qualitative test but not quantify. For example, simulation result in a period of time is: {13:15: 'influenced', 13:30: 'influenced', 13:45: 'not influenced'}, and all 3 time are homogeneous after the calculation. However, different solar angles could bring different amount of radiance that accumulates at a station’s case and leads to different degrees of heating. Thus, combining radiance quantify with station experiments would be a good direction to perfect the research. Besides, the Netatmo station experiment found that the sensor could record temperature data and upload it to the user normally by private portal while the data collected from same sensor in its public API is fragmentary. The available data could be more if this problem is fixed.

Figure 34. Creating potential locations in a 3D apartment model

Yellow points: potential locations; arrows: two sides of a corner

(the building model comes from https://www.cgtrader.com/3d-models/exterior/house/apartment-building-9)

The thesis creates the 3D points (2.5D more precisely, because of assigning a certain height to horizontal points) to represent potential locations of the Netatmo stations and the previous sections have proved that this method is not ideal when applied to high rise apartment. In order to create more accurate potential location, 3D building models could be used in generating 3D scatter points (Figure 34). Using this model, the scatter points are able to be placed in the balcony but not only on the ground with certain height value. But this also depends on how details (LOD) the building model could be. For example, two sides (red and green arrow in the figure above) of the balcony corner could generate totally different solar simulation result even they are close, because they face different direction.

The project aims to provide a solution to improve the data quality collected from crowdsourced weather stations and this will also help other researchers realize the potential achievements by using this data, especially in the field of air temperature monitor as well as UHI dynamic modeling and promising in subtle spatial & time resolution. 3 previous TU Delft M.Sc. thesis [19-21] have already studied the Netatmo station for temperature modeling reference but the temperature and location records were used before further processing. After the proposal and methodology of this project are refined, the result could be a good tool for checking their modeling result. In the bigger picture, open data collected by all crowdsourced weather station, not only Netatmo, is useful to update its quality especially the location info and bringing more use possibilities in the atmospheric sciences.

Reference

[1] Tan, J., Zheng, Y., Tang, X. et al. Int J Biometeorol (2010) 54: 75. https://doi.org/10.1007/s00484-009-0256-x

[2] Akbari, Hashem. (2005). Energy Saving Potentials and Air Quality Benefits of Urban Heat Island Mitigation. Lawrence Berkeley National Laboratory: Lawrence Berkeley National Laboratory.

[3] C. Sarrat, A. Lemonsu, V. Masson, D. Guedalia. Impact of urban heat island on regional atmospheric pollution Atoms Environ, 40 (10) (2006), pp. 1743-1758

[4] Sun, Hao & Chen, Yun & Zhan, Wenfeng. (2015). Comparing surface- and canopy-layer urban heat islands over Beijing using MODIS data. International Journal of Remote Sensing. 36. 5448-5465. 10.1080/01431161.2015.1101504.

[5] http://www.actionbioscience.org/environment/voogt.html

[6] Grimmond CSB, Blackett M, Best MJ, Barlow J, Baik JJ,, Young D, Zhang N. 2010. The international urban energy balance models comparison project: first results from phase 1. J. Appl. Meteorol. Climatol. 49: 1268 – 1292.

[7] Bohnenstengel SI, Evans S, Clark PA, Belcher SE. 2011. Simulations of the London urban heat island. Q. J. R. Meteorol. Soc. 137: 1625 – 1640.

[8] Muller CL, Chapman L, Young DT, Grimmond CSB, Cai X. 2013.Sensors & The City: A Review of Urban Meteorological Sensor Networks. Int. J. Climatol. 33: 1585–1600.

[9] Buettner, Ricardo (2015). A Systematic Literature Review of Crowdsourced Research from a Human Resource Management Perspective. 48th Annual Hawaii International Conference on System Sciences. Kauai, Hawaii: IEEE. pp. 4609–4618. doi:10.13140/2.1.2061.1845. ISBN 978-1-4799-7367-5.

[10] Howe J. 2006. The rise of crowdsourced. *Wired Magazine*. http://www.wired.com/2006/06/crowds/

[11] Chapman, L. , Bell, C. and Bell, S. (2017), Can the crowdsourced data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. Int. J. Climatol., 37: 3597-3605. doi:10.1002/joc.4940

[12] Chapman L. 2015. Urban meteorological networks: an urban climatologists' panacea? Urban Clim. News 58: 7–12.

[13] Meier F, Fenner D, Grassmann T, Jänicke B, Otto M, Scherer D. 2015. Challenges and benefits from crowdsourced atmospheric data for urban climate research using Berlin, Germany, as testbed. In Proceedings of the 9th International Conference on Urban Climate,
Toulouse, France, 20–24 July 2015.

[14] KNMI, Koninklijk Nederlands Meteorologisch Instituut,
Available on: https://www.knmi.nl/nederland-nu/klimatologie/grafieken/jaar

[15] Netatmo. 2018. Netatmo personal weather station.
https://www.netatmo.com/en-US/product/weather-station (accessed December 2018).

[16] Bauer, Christine. (2013). On the (In-)Accuracy of GPS Measures of Smartphones: A Study of Running Tracking Applications. ACM International Conference Proceeding Series. 10.1145/2536853.2536893.

[17] Netatmo Weather Station review
https://www.techadvisor.co.uk/review/ipad-accessories/netatmo-personal-weather-station-review-3442508/ (accessed December 2018).

[18] https://en.wikipedia.org/wiki/Thermometer (accessed December 2018).

[19] Lilia Angelova. 2018. Geographically weighted Urban Heat Island modeling using the Netatmo sensors. The case of The Hague.

[20] Anna-Maria Ntarladima. 2016. Modelling the Atmospheric Urban Heat Island and its Contributing Spatial Characteristics

[21] The Visualization of Urban Heat Island Indoor Temperatures. 2015 . Iris A.H. Theunisse

[22] Jochem, Andreas & Wichmann, Volker & Höfle, Bernhard. (2019). LARGE AREA POINT CLOUD BASED SOLAR RADIATION MODELING.

[23] https://www.timeanddate.com/weather/netherlands/the-hague/historic?month=5&year=2018 (accessed April 2019).

[24] Souza LCL, Rodrigues DS, Mendes JFG. 2003. Sky view factors estimation using a 3D GIS extension. In Proceedings of the 8thInternational Conference on IBPSA, Eindhoven, The Netherlands,1227 – 1234.

[25] Matuschek O, Matzarakis A. 2010. Estimation of sky view factor in complex environment as a tool for applied climatological studies. Berichte des Meteorologischen Instituts der Albert-Ludwigs-Universit ̈at Freiburg20: 534 – 539.

[26] Kokalj Z, Zaksek K, Ostir K. 2011. Application of sky-view factor for the visualization of historic landscape features in LiDAR-derived relief models. Antiquity85: 263 – 273

[27] An, S. M., Kim, B. S., Lee, H. Y., Kim, C. H., Yi, C. Y., Eum, J. H. and Woo, J. H. (2014), Three-dimensional point cloud-based sky view factor analysis in complex urban settings. Int. J. Climatol., 34: 2685-2701. doi:10.1002/joc.3868

[28] Urban Horizon Webpage.2018 http://skyview.bk.tudelft.nl/SynthesisProject/Home.html (accessed December 2018).

[29] http://solardat.uoregon.edu (accessed May 2019).

[30] https://en.wikipedia.org/wiki/Solar_zenith_angle (accessed April 2019).

[31] https://keisan.casio.com/exec/system/1224682331 (accessed April 2019).

[32] Joe Michalsky, J. J. 1988. The astronomical almanac's algorithm for approximate solar position (1950-2050), Solar Energy, 40, 227-235.

[33] Watt Engineering Ltd. On the Nature and Distribution of Solar Radiation US Government Printing Office Stock No. 016-000-00044-5, March 1978

[34] https://en.wikipedia.org/wiki/Solar_time (accessed April 2019).

[35] https://en.wikipedia.org/wiki/Hour_angle (accessed April 2019).

[36] http://solardat.uoregon.edu/SolarRadiationBasics.html#Note1 (accessed April 2019).

[37] Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". Economic Geography, 46(Supplement): 234-240.

[38] https://www.suncalc.org/#/52.0798,4.3549,9/ (accessed April 2019).

Appendix

Appendix A: Similarity results from the research period

Id = MAC address; 293c3c = 70:ee:50:29:3c:3c

id	x	y	z	similarity
293c3c	81344.9039	453596.878	0	0.76190476
0580ca	78764.3686	451693.108	0	0.30952381
294a16	77183.3717	452882.127	1.5	0.57936508
038b78	80884.5103	454944.295	0	0.72222222
03c966	85196.4949	450814.242	0	0.7
1c861e	76684.0784	449713.823	2	0.66666667
01c876	85733.5487	451796.136	0	0.71794872
1afb3a	78791.4946	457938.496	1.5	0.25490196
1f5a80	88165.8749	452863.654	0	0.77083333
280000	79102.0486	453358.11	0	0.69230769
1c8940	77599.3997	456537.656	0	0.35833333
19d8b0	83121.3884	455987.223	0	0.74242424
1c7cc8	77108.6534	452163.783	0.5	0.45555556
145b4a	85113.544	450028.897	0	0.70238095
1c82f6	80260.9906	454711.053	0	0.7948718
2ce40a	85126.8798	451716.791	0	0.57894737
03bd3e	80003.45	455411.554	0.5	0.78888889
028f52	80827.544	453974.615	2	0.74074074
1c8836	76568.5065	452227.601	1.5	0.61594203
1c683e	84310.4949	456840.996	0	0.68181818
179498	82334.2066	455212.51	1.5	0.76923077
1c83ae	76157.0841	452604.773	1	0.46666667
2ccc46	82328.4268	452840.535	1	0.74358974
13416e	81258.614	457331.86	1	0.51388889
2af39e	79975.494	448510.496	2	0.5625
1c7f02	79954.2413	454506.674	0	0.6
2898c4	78860.9956	454600	0.5	0.53333333
1c65b8	76110.5907	452162.196	1	0.27083333
19368	79894.7278	458797.277	2	0.40350877
19261a	82185.8718	456645.26	1	0.83
126c3a	79399.2355	457454.774	0	0.375
193534	80069.0592	448376.766	1	0.63888889
1c836c	76934.7777	452483.617	1.5	0.67592593
05f64a	81230.0293	456014.216	0.5	0.725
124d42	80024.0234	455431.508	0	0.7

1c7478	75882.2665	453314.377	1	0.79545455
1c82d8	78743.0625	452214.695	2	0.36363636
000a3e	75427.9971	453095.997	2	0.28888889
047b9e	84963.4169	450712.947	0.5	0.49074074
1c7524	82170.7632	453079.64	1	0.94230769
12bdd2	76434.8428	452241.755	0	0.20588235
2ce616	76837.7529	452410.434	1.5	0.60185185
276a54	81885.9996	457173.996	1.5	0.67
1c7f00	77445.7385	453827.046	2	0.61818182
1c8088	78300.5075	452542.869	2	0.77272727
02be22	88044.1482	453698.752	0.5	0.76851852
02b7b8	84631.8057	456575.711	0	0.54166667
29367c	78701.8087	450476.799	0	0.6875
1.45E+05	80007.3225	450039.142	0	0.74358974
1e1d8e	79015.0187	457279.261	2	0.6875
007a7c	84974.9852	451610.81	2	0.66666667
0607ca	76190.1463	453864.565	0	0.55555556
1276f8	77642.9972	451588.999	0	0.125
0371ca	79410.9805	453176.171	2	0.70512821
02979c	75833.5038	452898.866	1.5	0.45454546
15e880	85339.9958	450371.283	2	0.68421053
1c8acc	76452.9945	453329.496	2	0.73333333
2794a8	80706.3061	456227.366	0	0.31578947
211728	75004.2351	453728.489	0	0.25
1c8bde	79915.4638	454878.893	2	0.54444444
06a884	82451.9938	456790.499	1	0.6631579
00e12a	80868.1313	452421.637	0	0.48888889
2af290	86262.2897	449737.283	1	0.76666667
02542c	81680.7814	455037.87	2	0.7
0076c4	86125.1169	449966.118	0	0.75
12d7fa	88074.6569	452776.798	1	0.73333333
1c888a	80386.9328	455066.393	0.5	0.53043478
293cb2	79850.9972	451228.999	2	0.39047619
1c80aa	77193.4032	453515.244	0	0.34444444
274e2e	74863.9634	451755.656	1	0.54166667
133f0a	79897.4459	455404.587	0	0.74444444
293f2e	81835.781	455043.122	0	0.58333333
293738	78974.4511	455887.62	0	0.65555556
0288c0	79261.9261	449589.255	2	0.45833333
00e7d0	85709.9997	449997.493	0	0.71428571
1c8a9a	76631.9267	454746.124	1.5	0.52631579
28a662	80107.8148	449347.433	2	0.6875

007aaa	81205.3892	454681.895	1.5	0.66666667
0607c2	81994.8127	456356.217	1	0.69090909
0177f2	81510.4289	457355.977	2	0.75641026
28ae76	87946.1739	452996.333	1.5	0.74242424
12d8ae	77392.8282	456293.635	2	0.45454546
1c673c	81768.8733	453256.991	1.5	0.75454546
12c3fc	86960.5338	453643.46	0	0.77380952
00566c	81836.2256	455011.92	0	0.06666667
1c6872	82140.8983	454709.705	0	0.75925926
294bdc	80386.6682	453047.804	0	0.64583333
294b22	79916.0486	453413.283	0	0.76190476
00f5b2	84399.7989	450228.664	0.5	0.76388889
03cbe2	77193.1443	454009.441	2	0.42222222
1c7b80	84949.4959	450604.998	0	0.73913044
277100	80546.441	456259.846	0	0.125
293912	79210.4962	457522.497	1.5	0.68
2719ba	77661.7841	454290.206	2	0.64814815
00f78c	78723.6562	455610.736	1.5	0.46666667
1c5836	80863.3656	452011.788	1.5	0.66071429
161d28	86891.2483	452629.85	2	0.79347826
04cd7e	80449.473	457592.861	0	0.4
4.73E+05	84662.1789	456620.12	0	0.12820513
1f5636	81494.0887	455717.853	2	0.51851852
00d02a	78746.5393	456194.239	1	0.45833333
29559a	78795.1525	450373.061	0.5	0.775
0292f0	78616.4981	452804.494	1	0.75
22d5a2	85200.5796	450366.729	0.5	0.77272727
1c832e	80170.9665	449354.942	1.5	0.57647059
145ab8	78251.9959	456547.491	2	0.58888889
1c7ed0	79383.3726	454363.086	0	0.56666667
294b08	78794.9082	450344.896	0	0.70344828
1c8478	81137.4188	454392.208	0.5	0.74074074
02f330	77921.3665	454720.529	2	0.30555556
1c7fca	79234.6173	454564.336	1.5	0.27083333
04d6fa	79260.5623	458551.202	2	0.44615385
176c92	84753.6147	456531.351	0	0.73333333
2cccfa	76293.0586	451828.114	1.5	0.51041667
0086d4	85643.0254	449711.371	1	0.78125
2772a6	80288.8135	456608.532	0	0.62121212
1f42c8	78850.8119	457209.409	2	0.52777778
1c6c64	79945.8959	448904.638	0.5	0.20588235
1f43f4	78679.3806	456104.093	1	0.75

03faae	78343.4455	457966.383	0	0.51754386
2cbfd2	81812.2854	454959.965	0	0.640625
003adc	81573.0454	457197.629	0	0.26388889
1c845c	84208.9656	450313.27	0	0.578125
027ee4	77759.0868	450916.92	1.5	0.51388889
1c8482	82722.9981	455616.495	2	0.68333333
49454	77790.8852	454667.202	1.5	0.38666667
1c576e	82313.5897	458062.309	0.5	0.3125
28f7f6	81008.1816	454698.47	0	0.78205128
276da6	83863.6376	455659.292	0.5	0.70434783
47690	75736.7428	453028.072	1.5	0.11538462
275f20	80533.016	454972.457	2	0.45454546
1c6e7c	79453.5607	449818.147	0	0.3
1c845a	81576.8308	451993.467	2	0.88157895
145ff0	82137.5663	456556.258	2	0.625
58188	80347.0608	455426.105	1.5	0.76984127
1c8872	78790.439	455331.132	0	0.45833333
1c88ac	81414.6507	452638.821	0	0.6969697
1c7b84	84469.4412	450446.287	0	0.73076923
013dba	81480.5114	454972.629	0	0.74242424
19adca	83109.403	455498.32	0	0.75
293718	77725.3984	454329.029	1.5	0.67647059
03f96e	77116.1989	449389.024	0.5	0.76923077
1c8380	87977.1302	453176.46	2	0.71111111
55434	81138.8794	455931.315	2	0.78070175
12770a	79188.3448	455052.758	0	0.4125
0551be	80428.6854	451465.366	2	0.60606061
2948de	82247.9984	453944.994	0	0.82142857
1e01a2	80389.2542	456472.132	1	0.6952381
2ae3e4	80600.6897	457663.828	1.5	0.14814815
28cc1a	75750.6626	450563.607	0	0.77777778
28a5b0	79469.2897	458616.465	0	0.22857143
03d80c	86094.993	450070.491	0.5	0.75396825
1c82da	77646.671	449768.269	0	0.42708333
1c8368	80966.1628	454644.986	0	0.95
17604	84847.6738	451566.931	1.5	0.7254902
0283b2	82528.9399	453508.061	0	0.56410256
1c835c	79706.6874	455554.266	2	0.4952381
015a20	85705.9813	449708.735	2	0.75
275a26	84815.9951	456801.681	0	0.06060606
2931c0	84319.4944	450071	0	0.76470588
0241b4	81278.479	457384.261	0	0.40909091

01f492	81479.9961	454678.493	0	0.71875
0195a4	77923.5362	455486.223	1.5	0.57
029dee	79392.6517	458697.197	2	0.39583333
1c85b4	81911.7244	452742.942	0	0.75
13213e	78616.6888	456517.212	0	0.47916667
1270ac	84569.4966	449893.498	0	0.78571429
1ded62	83046.3731	456014.66	1	0.62666667
030b06	82297.8737	456838.721	0	0.2962963
191a90	77605.2288	453072.544	1.5	0.76851852
2cc7d6	86857.3965	453296.934	0.5	0.76923077
1c8218	83355.3023	455995.82	1.5	0.74242424
271338	86322.4723	450104.583	1	0.75362319
1e0c34	84678.7843	450746.437	1	0.704
28a02a	79173.9935	457357.696	0	0.22222222
134c28	76223.3764	454181.621	0	0.54385965
1c8434	78857.5045	453707.384	0	0.76666667
1c8352	84587.4597	456180.241	0	0.78333333
1c8636	84000.2376	456255.09	0	0.80208333
292a4a	75013.509	450961.274	2	0.50925926
03dba4	78819.691	455983.379	0.5	0.7037037
13233e	87055.9978	453125.499	1.5	0.75641026
1c822a	75146.4977	453808.994	0	0.81818182
1333b8	84732.177	450059.53	0	0.75490196
1c7458	81738.5664	454005.535	0	0.83333333

Appendix B: Solar influenced time based on the research period

Only sensors who similarity > 0.5 are considered; “-” = from… to

id	not influenced time														
2af290	6	-	6.75	8.5	20	-	20.5								
1c7478	19.5	19.75	20	20.25	20.5	20.75	21	21.25	21.5	21.75					
1c8218	7	8.25	9.75	-		10.5	11.25	20.25	20.5	21.25	21.75				
15e880	5.5	-	6.5	7	7.5	16.75	20	-	20.5	21.5	21.75				
1c7524	5.5	-	6.75	18.75	19.25	-	19.75	20.5	20.75	21.25	21.5				
271338	5.75	-	6.25	6.75	-	7.25	11.5	18.75	20.5	-	21.75				
0371ca	6.5	7	7.5	18.5	19.25	-	21.75								
0177f2	5.5	-	7.5	20.5	-	21.75									
58188	5.5	-	6.75	13.75	17	19.25	19.75	20	-	21.75					
191a90	5.5	-	7.25	8.75	20.5	-	21.75								
179498	6.75	-	7.5	14	-	17.75	20								
1c8acc	5.5	-	6.75	19.25	-	21.75									
12d7fa	5.5	-	6.75	-	7.5	19.5	20.25	-	21.75						
28ae76	5.5	-	6.75	19.25	-	21.75									
0292f0	5.5	-	7.25	19.5	20.25	-	21.75								
0086d4	5.5	-	7.5	15.5	16.25	20.5	-	21.75							
1c7f00	5.5	-	6.25	7	7.25	18.25	19.5	-	21.75						
1f43f4	5.5	-	8	12	17.5	21	-	21.75							
1e0c34	5.5	-	9	9.5	19.75										
028f52	5.5	-	6.75	8.25	9	19.5	-	21.75							
19368	5.5	-	7.5	19.5	-	21.75									

1c8a9a	5.5	-	7	13.5	13.75	19.5	-	21.75							
1c673c	5.5	-	8.75	10.5	11.5	12.25	13.5	19.5							
015a20	5.5	-	8	17.5	17.75	18.75	19.25	20.5	21	21.25	21.75				
1c861e	5.5	-	8.5	10.25	19.5	20.5	-	21.75							
19261a	5.5	-	7.75	17	17.25	19.25	-	19.75	20	-	21.75				
1c8088	5.5	-	9	12.25	20.5	-	21.75								
1f5636	5.5	-	8.25	19.5	20.25	-	21.75								
17604	5.5	-	7	14	14.75	16.5	17	18.75	20	-	21.75				
1c8482	5.5	-	8.5	19.75	20.25	-	21.75								
49454	5.5	-	8.25	19.75	20	-	21.75								
007aaa	5.5	-	7.5	18.75	-	21.75									
275f20	5.5	-	7.75	19	-	21.75									
007a7c	5.5	-	7.5	12.75	18.25	-	21.75								
1c832e	5.5	-	8	19	-	21.75									
1ded62	5.5	-	8.5	18.75	-	21.75									
292a4a	5.5	-	8	15.25	18.75	19	19.5	-	21.75						
1c8836	5.5	-	7.25	8	-	8.5	18.25	-	21.75						
193534	5.75	-	8.5	9	-	9.75	10.25	-	10.75	19.5	-	21.75			
12d8ae	5.5	-	7.75	18.5	-	21.75									
161d28	5.5	-	7.5	8.25	13.5	16	16.75	-		18.75	19.5	20.25	-	21.75	
1c845a	5.5	-	9.5	16.25	19.75	20.25	-	21.75							
1c8380	6	-	8.5	14.25	15.5	17	19	-	21.5						
1c83ae	5.5	-	8	8.5	-	9	19.25	-	21.75						
06a884	5.5	-	7.75	10.5	12.25	12.5	14	15.5	17.5	18	18.5	19.75	20.25	-	21.75

13416e	5.5	7	-	7.5	9.25	10	10.5	10.75	-	12.75	17.5	18.5	19	-	21.75
28a662	5.5	5.75	-	8	8.75	9.25	17.75	-	21.75						
2af39e	5.5	-	20.75												
02542c	5.5	-	10	-	10.5	10.75	19.25	-	21.75						
55434	5.5	-	13	14.5	19	-	21.75								
1e01a2	6.25	-	12	13.25	13.5	14	19.75	-	20.5	21.75					
293cb2	5.5	-	8.75	17.25	-	17.75	18	18.25	19.75	-	21.75				
2719ba	5.5	-	17.25	-	21.75										
1c836c	5.5	-	11.5	20.25	-	21.75									
145ff0	5.5	-	9	10	-	10.5	11.25	16.25	17.5	18.5	19.5	-	21.75		
276a54	5.5	-	9.25	15.5	15.75	16.5	-	18	19.25	-	21.75				
293718	5.5	-	9.25	17.75	18	18.25	-	21.75							
2ce616	5.5	-	9.25	9.75	10.25	-	21.75								
027ee4	5.5	-	10	17.5	17.75	-	21.75								
1afb3a	5.5	-	8.75	16.75	-	21.75									
0288c0	5.5	-	12.5	19.75	-	21.75									
1c5836	5.5	-	10.5	15.75	-	21.75									
0195a4	5.5	-	10.75	18	-	21.75									
293912	5.5	-	11.75	14.75	15	16.5	-	21.75							
1c835c	5.5	-	8	9.25	15.5	-	21.75								
029dee	5.5	-	11.75	15	-	15.5	17.75	18.25	-	21.75					
294a16	5.5	-	9.25	9.75	11.5	-	12	17.25	-	21.5	21.75				
274e2e	5.5	-	12.5	19	19.5	-	21.75								
145ab8	5.5	-	9.5	16.25	-	21.75									

2cccfa	5.5	-	12	18	-	21.75									
1c65b8	5.5	-	8.5	11.75	15	-	21.75								
02979c	5.5	-	8.75	-	11	16	17.25	-	21.75						
1c8bde	5.5	-	8	14.25	-	21.75									
0551be	5.5	-	10.25	-	21.75										
00d02a	5.5	-	10.5	15	-	21.75									
1c82d8	5.5	-	10.75	16.5	17	-	21.75								
1e1d8e	5.5	-	13.25	14	-	21.75									
0607c2	5.5	-	13.25	19	-	21.75									
03cbe2	5.5	-	10.25	15.5	-	21.75									
000a3e	5.5	-	-	14.25	19.5	-	21.75								
00f78c	5.5	-	9.5	13.5	14	-	21.75								
47690	5.5	-	8.75	10.25	-	13.5	16	-	21.75						
1c7fca	5.5	-	10	11.5	13.5	-	21.75								
1f42c8	5.5	-	10	-	15	18.5	-	21.75							
2ae3e4	5.5	-	10	-	18	18.75	21.75								
04d6fa	5.5	-	10	-	15.5	18.75	-	21.75							
02f330	5.5	-	10	-	21.75										
13233e	5.5	-	8.25	9.5	19.25	-	21.75								

Appendix C: Validation result (comparison of two periods)

id	Vertical D	Horizontal D	id	Vertical D	Horizontal D
293c3c	0	2.23605607	1c7524	0.5	21.7829834
0580ca	0	7.56636959	12bdd2	0	0.99999691
294a16	1.5	8.54400716	2ce616	0.5	17.6139153
038b78	0	3.8591E-05	276a54	0	2.3302E-05
03c966	0	4.4033E-05	1c7f00	0	1.41419464
1c861e	0	2.49997308	1c8088	0	0.49996642
01c876	0	4.94973189	02be22	1.5	9.30053261
1afb3a	1.5	4.30115345	02b7b8	0	7.7607E-06
1f5a80	0	0.70708456	29367c	0	5.09901527
280000	2	2.49999838	145000	0	4.03117655
1c8940	0.5	3.53551818	1e1d8e	1	0.70712865
19d8b0	0	4.9243E-05	007a7c	0	0.70707492
1c7cc8	0	5.0103E-05	0607ca	0	1.58115326
145b4a	0	1.41420359	1276f8	0	2.82845666
1c82f6	0	1.7404E-05	0371ca	0.5	29.2617579
2ce40a	0	2.54951134	02979c	1.5	5.85239618
03bd3e	0	0.70708457	15e880	0.5	1.80276469
028f52	0	3.20155099	1c8acc	0	0.49995419
1c8836	0.5	1.11799081	2794a8	0	15.3052356
1c683e	0	0.50000161	211728	0	2.4733E-05
179498	0.5	9.3407686	1c8bde	0	0.50000066
1c83ae	1	5.85236921	06a884	0	2.06156535
2ccc46	1	26.0048358	00e12a	2	1.11800103
13416e	0.5	3.90512629	2af290	0.5	20.5912339
2af39e	0	4.30116323	02542c	1.5	7.28014429
1c7f02	0.5	9.01387953	0076c4	0	2.54951458
2898c4	0	4.9548E-05	12d7fa	0	1.58112649
1c65b8	0	4.71697631	1c888a	0.5	0.70708515
19368	0	4.03111224	293cb2	0.5	12.1655295
19261a	0	3.598E-05	1c80aa	1	2.2360542
126c3a	0	2.23605087	274e2e	1	10.1242211
193534	0	3.9396E-05	133f0a	0	6.40311867
1c836c	0.5	2.54949994	293f2e	0	0.70710349
05f64a	0	21.5058087	293738	0	0.70708121
124d42	0	2.50003677	0288c0	2	19.7294251
1c7478	1	0.70708674	00e7d0	0	0.50004257
1c82d8	0	1.58112742	1c8a9a	0.5	1.80277491
000a3e	0	7.5605E-06	28a662	0	0
047b9e	0.5	27.6812103	007aaa	0.5	1.11803227

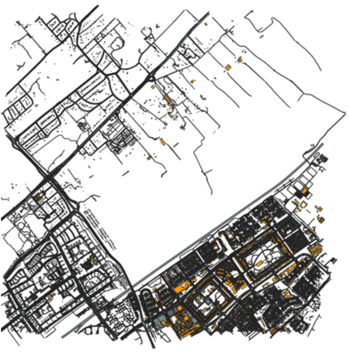
0607c2	1	1.58115701	1c6c64	0	1.11804748
0177f2	0.5	16.0078213	1f43f4	0	0.49999544
28ae76	0.5	14.0890026	03faae	0	9.21951496
12d8ae	2	14.1597999	2cbfd2	0	0.5000016
1c673c	1.5	1.80276753	003adc	0	10.7354531
12c3fc	0	0.99999609	1c845c	0	0.70713362
00566c	0	0.50004099	027ee4	0.5	17.3565432
1c6872	0	5.59019197	1c8482	0.5	4.71696978
294bdc	0	4.03113273	49454	0	1.41423923
294b22	0	1.11805713	1c576e	0.5	0.70710347
00f5b2	0.5	4.99997958	28f7f6	0	1.58113296
03cbe2	0.5	0.70708833	276da6	0.5	17.3349214
1c7b80	0.5	10.9658343	47690	0	0.50000281
277100	0	0.50000307	275f20	1.5	22.5610276
293912	0	0.7070987	1c6e7c	0	0.70713898
2719ba	0.5	3.80787681	1c845a	0	2.54950346
00f78c	0	0.50000203	145ff0	0	9.01389529
1c5836	0	5.70090993	58188	0.5	3.04135727
161d28	0	1.11804465	1c8872	0	3.90509556
04cd7e	0	3.64009888	1c88ac	0	21.2720168
4.73E+06	0	0.49997679	1c7b84	0	0.50001435
1f5636	0.5	0.70707773	013dba	0	0.49999833
00d02a	0.5	0.49999547	19adca	0	1.41418283
29559a	0	2.0615116	293718	0	4.9355E-05
0292f0	0	11.0679618	03f96e	1	2.23605789
22d5a2	0.5	0.70710647	1c8380	2	8.60232748
1c832e	0.5	22.0056675	55434	0.5	0.70713574
145ab8	1	27.6811999	12770a	0	8.90225443
1c7ed0	0	3.6227E-05	0551be	0.5	1.58113336
294b08	0	1.9852E-05	2948de	0	7.6322E-06
1c8478	0	0.70711189	1e01a2	1	0.49999618
02f330	0.5	19.906038	2ae3e4	1.5	2.23608827
1c7fca	0	15.3378695	28cc1a	0	3.53553444
04d6fa	0	9.19242548	28a5b0	0	0.70708382
176c92	0	0.5000135	03d80c	0	0.50003058
2cccfa	0	8.4604E-06	1c82da	0	12.8549554
0086d4	0	1.11805025	1c8368	0	2.91549167
2772a6	0	1.41424447	17604	0.5	1.00000387
1f42c8	0	2.50003539	0283b2	0	1.11805428
015a20	0	8.2006221	1c835c	0	2.91548043

275a26	0	20.7183402
2931c0	0	4.03116733
0241b4	0	1.58113239
01f492	0	1.11803113
0195a4	0.5	0.50001443
029dee	1.5	42.851441
1c85b4	0	20.7183514
13213e	0	3.3473E-05
1270ac	0	5.522635
1ded62	0.5	6.10329023
030b06	0	6.2482E-06
191a90	1	3.16225168
2cc7d6	0.5	9.70825633
1c8218	1.5	0.49999846
271338	0.5	9.30054529
1e0c34	0	2.3379E-05
28a02a	0	5.09903106
134c28	0	3.3623E-05
1c8434	0	1.80278589
1c8352	0	1.7816E-05
1c8636	0	0.99999887
292a4a	1	4.00002473
03dba4	0	1.9469E-05
13233e	0	2.4999792
1c822a	0	5.22013174
1333b8	0	16.347785
1c7458	0	6.04152536

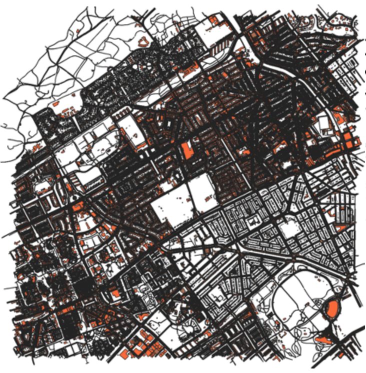
Appendix D: Tile number


Tile 1


Tile 2


Tile 3


Tile 4


Tile 5


Tile 6


Tile 7


Tile 8

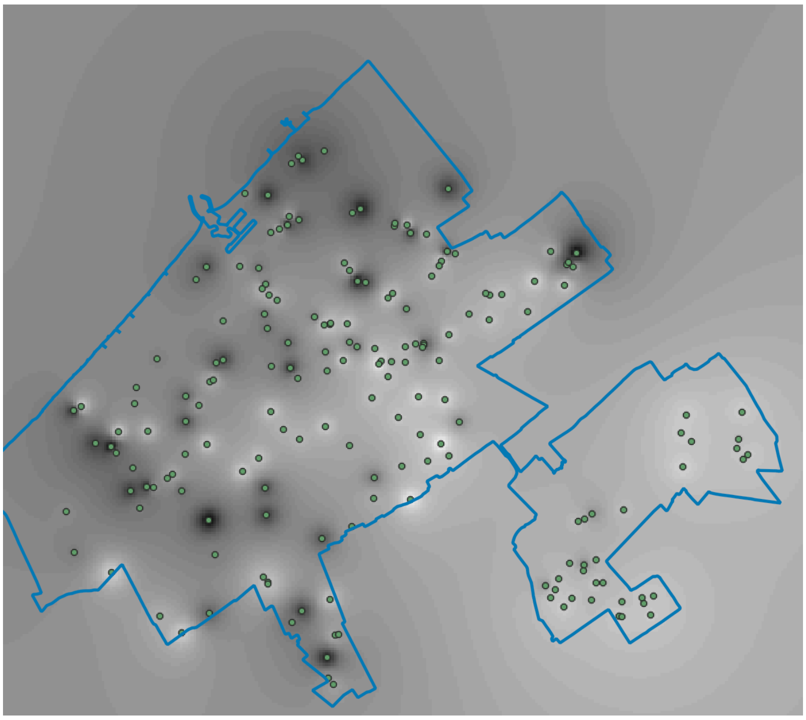

Tile 9


Tile 10

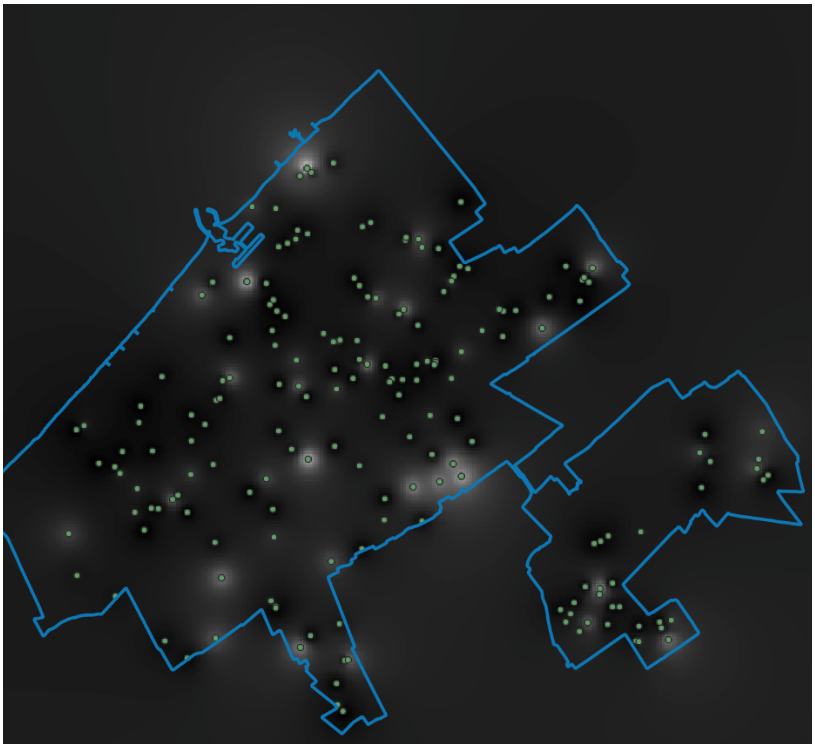

Tile 11


Tile 12

Appendix E: Similarity (research period) and location difference interpolation (IDW) maps



Similarity Interpolation (while pixel = 1; black = 0)



Location difference interpolation (while pixel = 42.8; black = 0)