Personal information

Name	Yixin Xu
Student number	4684710
Address	Barabarasteeg 2A17
Postal code	2611BM
Place of residence	Netherlands, Delft
Telephone number	+31 649248206
E-mail address	Xuyxin12@gmail.com

Graduation committee

Main mentor	Ken Arroyo Ohori
Second Mentor	Alexander Wandl

Title

Title of the graduation research	Filtering an amateur weather station’s temperature data by using 3D model and points cloud ----Example based on the Hague

Research

see full proposal attached

Filtering an amateur weather station's temperature data by using 3D model and points cloud

----Example based on the Hague

MSc Geomatics Thesis Proposal

Yixin Xu

Mentors:
Ken Arroyo Ohori
Alexander Wandl

January 10, 2018

1 Introduction

1.1 Abstract

The continuously increasing population within cities imposes the future challenges related to planning and managing the sustainable environment and urban heat island has been recognized as one of the leading environmental issues of the 21st century. Thanks to the variety uses of Netatmo weather station, more and more temperature data could be used for UHI research. So far, there already have been two MSc projects (TU Delft) that relate Netatmo data to UHI modeling. However, due to character of this station, it sometimes could generate unreliable records when expose to solar radiance. Thus, the current work is focusing on develop a way, combining with points cloud to filter outliers from Netatmo records.

The following part in section 1 will introduce scientific background related to this project. Section 2 will propose the main question as well as sub-questions the project plans to solve and section 3 is relevant work done by others. The specific method to answer the proposal questions will be elaborated in section 4 and the schedule of this project will be shown in the final section.

1.2 Urban Heat Island (UHI)

An urban heat island (UHI) is an urban area or metropolitan area that is significantly warmer than its surrounding rural areas due to human activities. In the last ten years, extreme heat conditions were observed more frequently, which provided more information about the related negative effects on the population and the economy. One example is the summer mortality rates in and around Shanghai yields heightened heat-related mortality in urban regions and UHI is directly responsible [1]. One paper also found that this increase in air temperature is responsible for 5-10% of urban peak electric demand for a/c use in USA [2]. Besides, one study shows that UHI in has an important impact on the primary and secondary regional pollutants, more specifically the ozone and the nitrogen oxide (NOx) [3].

There are three types of heat islands based on different components (Figure 1):

- canopy layer heat island (CLHI)
- boundary layer heat island (BLHI)
- surface heat island (SHI)

The first two refer to a warming of the urban atmosphere; the last refers to the relative warmth of urban surfaces. The urban canopy layer is the layer of air closest to 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 1 kilometer or more in thickness by day, shrinking to hundreds of meters or less at night [4]. The Canopy Urban Heat Island is the most studied one because of its direct relevance to the people's health, therefore CUHI is the type that is mostly discussed concerning the UHI topics.

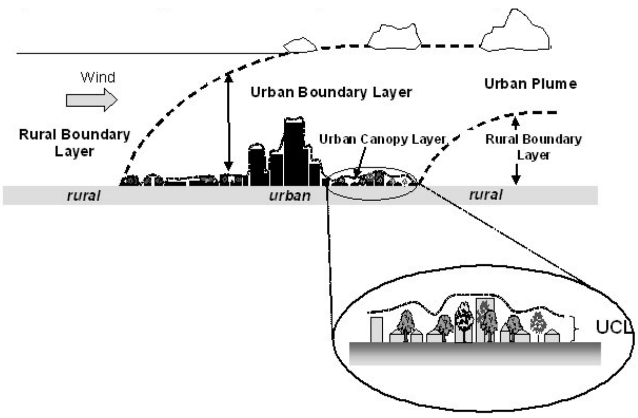

Figure 1. Schematic depiction of the main components of the urban atmosphere. [5]

Depending on the different types of UHI two main acquisition methods have been used – remote sensing or ground-based temperature measuring 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. The main difference between the two methods lies in their spatiotemporal characteristics - ground-based measuring has the advantage of high temporal resolution, but only one data per sensor. On the contrary, remotely-sensed data has a higher temporal resolution, but the data describes one point in the period of time.

1.3 Quantifying the UHI

There are numerous approaches available to quantify the UHI effect. The canopy UHI is traditionally measured using station which is not ideal as station pairs offer limited spatial information. As an alternative, many studies have attempted to quantify the UHI using remote sensing. This provides spatial data at a daily resolution but is limited as it observes land surface temperatures as opposed to canopy 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 a growing number of high resolution urban meteorological networks brought about by the decreasing costs of instrumentation. But there remains a considerable scientific challenge in sufficient quality to be accepted by the atmospheric science community [8]. A new opportunity, Netatmo weather station has recently emerged in this area in the form of low cost, citizen science weather stations that connect to smartphones and local Wi-Fi networks to relay crowdsourcing data in real time to the sever.

1.4 Crowdsourcing data

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

Crowdsourcing was first termed by Howe referring to the idea of outsourcing to the crowd [10]. Linked with public engagement activities via citizen science, crowdsourcing is now increasingly finding itself as an established technique for collecting mass data in many scientific disciplines [11]. However, with these few notable exceptions, the use of crowdsourced data in the atmospheric sciences is actually very limited when compared with other areas of scientific study and the main reason is how to obtain a precise and representative observation [12]. Despite these concerns, the results from the validation exercise of the Netatmo with standard measurements have proved promising [13].

The amateur weather station chosen to be used in this study is the Netatmo weather station, which is easily configured and controlled by a smartphone to monitor and record the local environment and coordinates. Besides, the spatial density of Netatmo station ensures that these stations could work as a network not singly. The dense data means more choice when dealing with data e.g. when one or two station is faulted, it’s possible to replace by other stations.

Records from the station is transmitted wirelessly, using Wi-Fi and configuring by bluetooth, to the cloud where it can be accessed via a smart device, as well as being made available online via a ‘weathermap’ (Figure 2) on the Netatmo website with observations updated every 15 min. It’s also noticeable that the data shows in the” weathermap” is already filtered by the “Netatmo” so the data looks much more harmony than the raw data from API. 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.

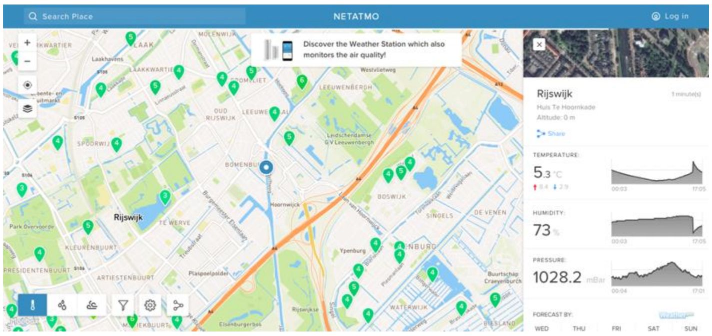

Figure 2. Netatmo Weather Map

The Royal Netherlands Meteorological Institute (KNMI), provides hourly temperatures for certain locations. The KNMI sensors are located in the rural environment: KNMI 330 is located at the 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 this 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, however, air temperature measurement influenced by sun radiance could be higher than true air temperature. Netatmo also mentions that temperature records exposed sun could be 1-2 ℃ higher than that in shadow [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]. Regarding that most of the stations are put near buildings, 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 shows in the Figure 6. It’s obvious find top two lines don’t change too much with time, which fail to fit with the common sense.

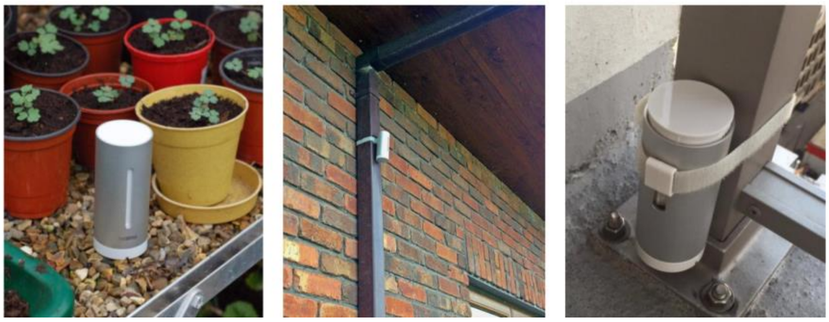

Figure 3. Different uses for Netatmo weather station[17]

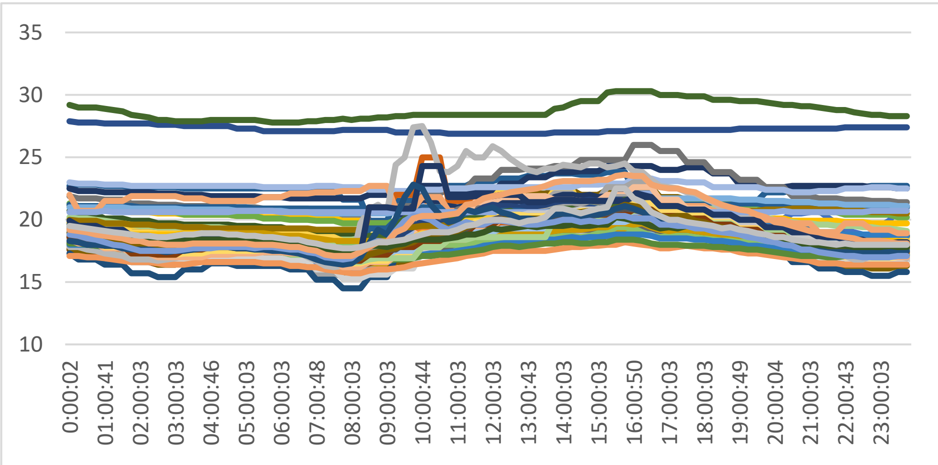

Figure 4. Example of system error (data source from Netatmo API on September 10th, 2018)

1.5 Problem statement

Urban heat island has been recognized as one of the leading environmental issues of the 21st century. In the last ten years, extreme heat conditions were observed more frequently, which provided more information about the related negative effects on the population and the economy.

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

Figure 5. traditional ground - based thermometer[18]

Netatmo, as a new crowdsourcing weather data, has significant strength 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 error or solar radiance as discussed in section 1.4, but so far previous work mainly use mathematical way to remove system error and there is lack of method to remove outliers caused by solar radiance. Another challenge is accuracy of stations' location. The reason why accurate locations are important in UHI research is that the temperature divergency could be remarkable. Figure 6 shows an example of UHI modeling result in the Haag. 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 shows significant variation (also shows in the figure 6). In another word, 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.

Hence, the problem that this work is focusing on is developing a method which is able to find accurate location of the Netatmo stations and filter outliers especially caused by solar radiance.

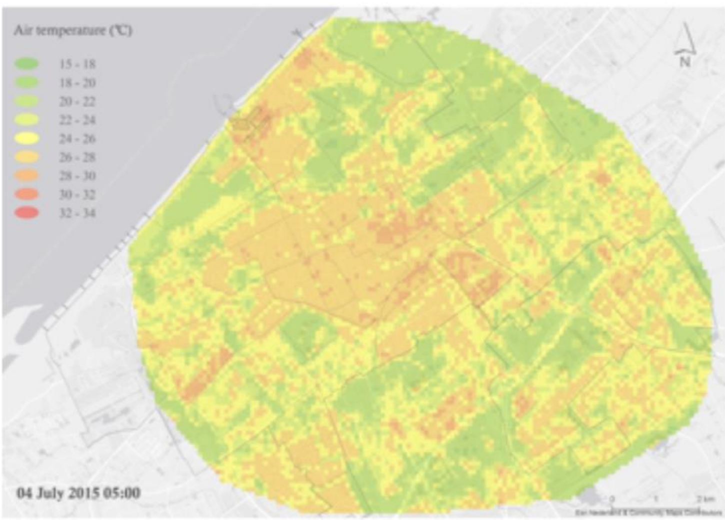

Figure 6. UHI modeling result in the Haag [20]

2 Research questions and scope

2.1 Research questions

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

    1. How to find potential location of real sensors are?
    2. How to know sensor's record is higher than it should be and filter outliers caused by direct solar radiance?
    3. For a certain area, how to know when it receives direct solar radiance?

2.2 Research scope

This thesis will focus on the temperature data from Netatmo outdoor modules in the Haag so other data e.g. indoor modules or data in other cities will not be considered. Besides, the project will mainly analysis outliers from direct solar radiance but other types of outliers e.g. system error will only be briefly discussed. Also, the project will not concern UHI modeling or other factors that contribute to UHI (except solar radiance).

3 related work

So far, there have been three Msc thesis projects related to the NETATMO weather station and all of them stress the UHI modeling. Lilia Angelova has developed a mathematical method to filter outliers in NETATMO temperature data as well as several UHI model indicators, e.g. sky view factor and NDVI and these method and indicators are based on 2D data [19]. Likewise, Anna-Maria Ntarladima focused on the dynamic UHI research as well as dynamic modeling [20]. Iris A.H. Theunisse created a 3D temperature model by combining weather station data and CityGML, but she didn't concern other factors which might bring influence on temperature in the city [21].

Other papers relate to UHI base on Netatmo sensor are not much. One finds not only promising benefits (high spatial resolution) from Netatmo data for urban climate research in Berlin but also challenges: user-specific and sensor-specific measurement errors[10].

Unfortunately, researches about using 3D model or points cloud to filter the data or find the location of sensors cannot be found so far.

4 Methodology

4.1 Data pre-process

As section 1.2 mentioned above, 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 process. It's noticeable that some extreme high temperature records not necessarily mean system errors but cause by direct solar radiance because it might result from that users put 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 shows very abnormal pattern e.g. temperature difference between sunset and sunrise less than 5 ℃ (this just a threshold on the initial stage; the temperature difference between sun rise and set is complex, so this threshold just needs to ensure that phenomenon in Figure 4 will not appear) or wrong recording frequency. Once a sensor is problematic, it will be deleted in the dataset because it is considered as unreliable. Also, the remaining points will be clipped in Qgis which ensures every sensor is located in the Haag (Figure 6). Notice that all sensors or Netatmo weather station the paper mentions below are after pre-process.

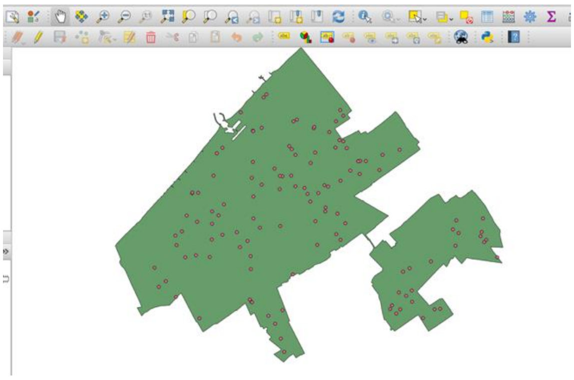

Figure 6. Sensors' distribution after pre-process

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 (shows 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. Interpolation method is helpful to predict the value but will largely 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 expose 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 temperature is higher than it should be by using record data minus average data, following non-linear LSA (least square adjustment) to get a smoother line. An example (only one sensor) shows in picture below. Red dots mean temperature differences between records and averages while blue line is a non-linear function with red dots. Here we obtain 95 data (x-axis) because raw data updates 15 min per time and there will be 95 records in one day. The blue line increases significantly from x=30 to x=60 and the corresponding time is about 7:50am and 2:50pm respectively (Figure 7). Accordingly, this sensor is likely to expose to sun radiance during this period of time.

It's noticeable that not all sensors will show the pattern like Figure 6. For example, it might be like a flat line with neither decrease nor increase, or the increase is not significant. Hence, defining a threshold to tell whether and when a sensor is influenced by solar radiance is required. Sensors who will be influenced will go to section 4.3.

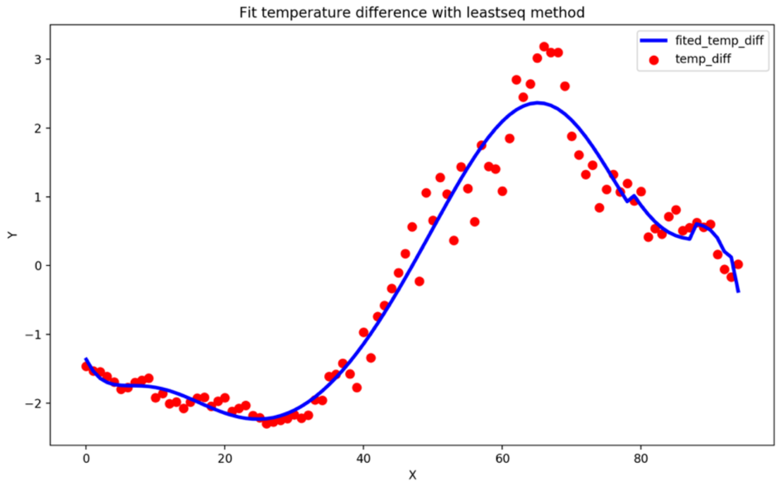

Figure 7. One example of using LSA to simulate temperature difference

4.3 Generating potential location of Netatmo stations

Although the coordinate of each station is given, 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.

Another resection of the buffer is buildings. All temperature data used in this project are collected from Netatmo station outdoor module (it is possible that some users use outdoor module inside, but they are removed in pre-process) so the part where the buffer covers buildings will not be considered when generating potential location. The buildings' footprint will be extracted from BAG. Then generate points with same distance (Figure 8).

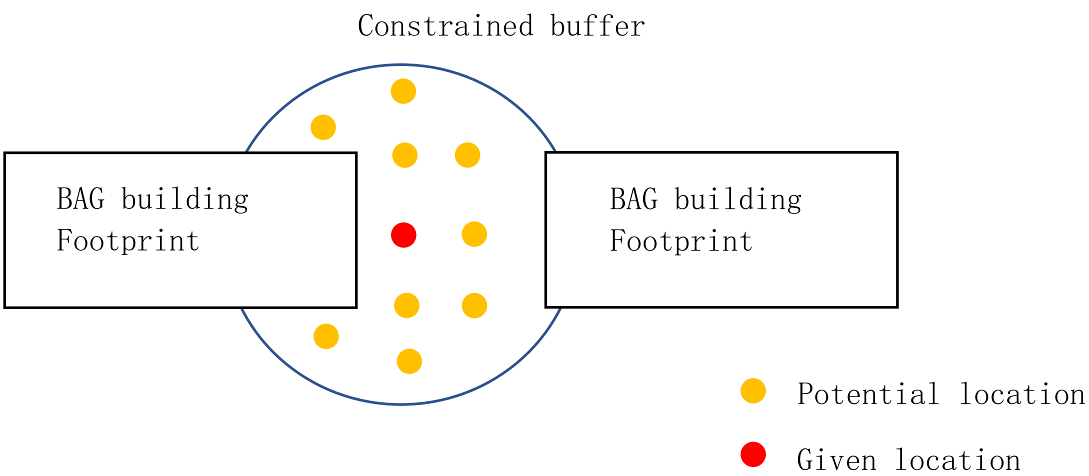

Figure 8. Principle to generate potential location

4.4 Computing sky view (dome) and sun parameter

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 [22] and this project will add functions on basis of their work to calculate 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 9. left: one selection point;

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

4.5 Finding the most likely location of the station

The idea of this section is by comparing the result from 4.2 to 4.4 for each sensor. For example, if the result derived from section 4.2 is "7:50am to 2:50pm the sensor will detect higher temperature than is should be", then the potential location whose output from section 4.4 is the closest to "7:50am to 2:50pm" will be set as the real location of the Netatmo station. It's noticeable that only station who can find "increase pattern" in section 4.2 will be take into consideration in 4.3 to 4.5. Those cannot find "increase pattern" could means the station is in the shadow all the time. Although locate them is impossible in this project, actually in another way, they almost don't have outliers cause by solar radiance.

4.6 Filtering outliers that influenced by solar radiance

Once the location of each sensor is known (exclude "shadow station" mentioned above), it could be more than one way to filter outliers. For instance, for each recording time it is possible to replace the noisy sensor's record by interpolation with sensor that not influenced by solar radiance. Alternatively, it is also possible to develop an algorithm to process data when it influenced by sun.

4.7 Validation

The idea is that checking the results from different date. More precisely, the methods mentioned above are all based on one-day data, so this means everyday data could be a control group. Then because the sensors are fixed so the result of sensors' location from different date should be same. If the location difference of each sensor is less than some threshold then it's reasonable to say the method is validated.

5 Schedule

Start	End	Activity
Sep.15	Nov.14	Exploring the topics
Nov.15		P1
Nov.16	Dec.1	Literature review
Dec.2	Dec.23	Pre-process of Netatmo data
Dec.24	Jan.9	Studying factors related to solar radiance
Jan.10		P2
Jan.11	Jan.31	Combining sky view factors related to solar radiance
Feb.1	Feb.25	Creating potential location of sensors
Feb.25	Mar.4	Finding the most likely location of the station
Mar.5		P3
Mar.6	Apr.1	Optimizing final implementation
Apr.2	Apr.30	Writing Thesis
May.1		P4
May.2	May.15	Finalizing thesis
May.16	May.31	Preparing presentation
Jun.1		P5

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 Crowdsourcing 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 crowdsourcing. *Wired Magazine*. http://www.wired.com/2006/06/crowds/ (accessed 11 July 2016).

[11] Chapman, L. , Bell, C. and Bell, S. (2017), Can the crowdsourcing 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 Visualisation of Urban Heat Island Indoor Temperatures. 2015 . Iris A.H. Theunisse
[22] Urban Horizon Webpage.2018
http://skyview.bk.tudelft.nl/SynthesisProject/Home.html (accessed December 2018).