Geographical Information Management and Applications

Master Thesis
Assessment of heating energy use at postal code level 6: influence of the local climate

Author: Bram Pijnenburg
b.j.pijnenburg@students.uu.nl

Supervisor: Prof.dr. Laure Itard
L.C.M.Itard@tudelft.nl

Responsible professor: Prof.dr.ir. Peter van Oosterom
P.J.M.vanOosterom@tudelft.nl

June 2, 2021

Universiteit Utrecht

TU Delft

University of Twente

Wageningen UR

Acknowledgement

I wouldn't have completed my thesis without the support and feedback from my supervisor Prof. dr. Laure Itard. The field of energy consumption in the built environment was new for me and without the right input I wouldn't't have understand this topic completely. Besides the moments of feedback I also joined the weekly online meetings with other Master students, PhD students and professors from the TU Delft, initiated by Prof. dr. Laure Itard. These meeting helped me understand the field of energy transition and energy consumption in the built environment a lot better. Besides the scientific discussions and presentations of various attendees it was also a joyful gathering with people in this time of social distancing. I would also like to thank Prof. dr. ir. Peter van Oosterom for his feedback and his role as responsible professor. As an experienced professor of GIS Technology he explained clearly his feedback on my thesis, especially on the topics of the GIS field. He also introduced me the open access Netatmo dataset that contributed much to this thesis.

I would also like to thank Dr. Ir. Ken Arroyo Ohori to give me access to his own gathered database of Netatmo sensed data. Besides sharing his dataset he also put me in the right direction of adjusting the data that fits this thesis. I would also thank Andrea Joseph Thaddeus and Björn Bogers for explaining me the difficult parts of Python. Without their explanation it would be a lot harder to create the Python scripts for this thesis.

This thesis is the product of attending courses in my Bachelor *Sociale Geografie en Planologie* and my Master GIMA (Geographical Information Management and Applications). I would like to thank all my professors from the University of Utrecht, TU Delft, Wageningen University and the University of Twente. I have followed a lot of interesting courses in the field of GIS, spatial planning and social geography. Different courses in my Bachelor did me realize that I would continue in the field of geographical information science. And in my Master the in-depth courses of this field prepared me for a spot on the labour market.

At last I would thank all my fellow students from both my Bachelor and my Master for the fantastic time during and after the study hours. I have worked on several great projects such as *Atelier II* and the *Phetchabun erosion project*. I also had a great time during the trip to London in my first year of the Bachelor and the contact weeks during my Master were always a delight with different activities initiated by *NODE*. I'm happy to finish my years as a student with this thesis and ready for a new adventure.

Summery

Buildings in the Netherlands are responsible for 40% of the energy use. One of the main contributors of this energy use is heating the buildings in the colder periods of the year. There is an energy transition from fossil fuel to renewable energy but most of the heating of buildings is still due gas consumption. There are multiple studies carried out towards the reduction of the use the gas consumption and also multiple studies showed how to make buildings more sustainable. To a lesser extent the influence of the local climate has been investigate and what the influence is on the gas consumption. To compare the amount of gas consumption through different years the concept of degree days is created. With the degree day method the energy consumption of different years can be compared without the influence of the temperature. In this case the colder years can be compared with the warmer years.

This degree method is a broad concept that doesn't account for the differences of the local climate. The local climate has been a great topic of research in another spatial phenomenon, namely: urban heat islands. Climatic factors such as wind, irradiation and rain are account for to explain the differences in local temperature. Also non-climatic factors such as green areas and densely built areas with a lot of concrete are accounted for. Big differences between neighbourhoods or urban and rural areas are detected in these studies towards urban heat islands.

This thesis focus on the local differences of gas use in the city The Hague on a postal code 6 level. The degree days for these postal code 6 levels and the gas use per postal code 6 level will be mapped and explained. The method to investigate the local climatic differences in urban heat island will be applied in this thesis to explain the differences in degree days per postal code 6 level. Using open data from Netatmo the local temperature will be mapped and due these outcomes the degree days will be created. With data from the CBS the gas consumption per postal code 6 level will be visualized. Combining these two datasets with each other will explain to a certain extent the differences of gas consumption due the local climate. With other open data sources the climatic factors: wind, rain and irradiation will be investigated and explained to what extend the impact is on the local climate. Besides the climatic factors the non-climatic factors: green areas and the energy labels will explain the differences in gas use that are not influenced by the local weather.

The results of this thesis shows that there is a significant difference between the gas use and degree days in both the years 2018 and 2019. Also the gas use corrected with degree days shows that there is a difference in these years. Although this difference is very small. To put this in perspective: another statistical test is carried out with categorized gas use corrected with degree days that stated that there is no significant difference. With this information we can conclude that the climatic factors (besides temperature) plays an small role in the difference of gas use between the 2018 and 2019. It also seems that there are variations in the local climate. But without the local wind, rain and irradiation data it is hard to say how these factors influence the local climate.

Contents

Acronyms

ADD: Actual Degree Days
API: Application Programming Interface
BAG: Basisregistratie Adressen en Gebouwen
BGT: Basisregistratie Grootschalige Topografie
CBS: Centraal Bureau voor de Statistiek
CDD: Cooling Degree Days
CDH: Cooling Degree Hours
CSV: Comma Separated Values
DD: Degree Day
DDC: Degree Day Correction
ELD: Enthalpy Latent Days
FME: Feature Manipulation Engine
FOSS: Free and Open Source Software
HDD: Heating Degree Days
HEC: Household Energy Consumption
IDW: Inverse Distance Weighting
JSON: JavaScript Object Notation
KNMI: Koninklijk Nederlands Meteorologisch Instituut
LST: Land Surface Temperature
mBar: millibar
MCDH: Mean Cooling Degree Hours method
MDD: Mean Degree-Days
MDT: Mean Daily Temperature method
MDH: Mean Degree-Hours
mm: millimeters
PC4: Postal code 4
PC5: Postal code 5
PC6: Postal code 6
PET: Physiological Equivalent Temperature
PDOK: Publieke Dienstverlening Op de Kaart
RECS: Residential Energy Consumption Survey
RES: Regional Energy Strategy
SHI: Surface Heat Islands
SMY: Standard Meteorological Year
TMY: Typical Meteorological Year
TRY: Test Reference Year
RDD: Relative Degree-Days
UHI: Urban Heat Islands
WDD: Weighted Degree Days
WFS: Web Feature Service
WMS: Web Map Service

Chapter 1

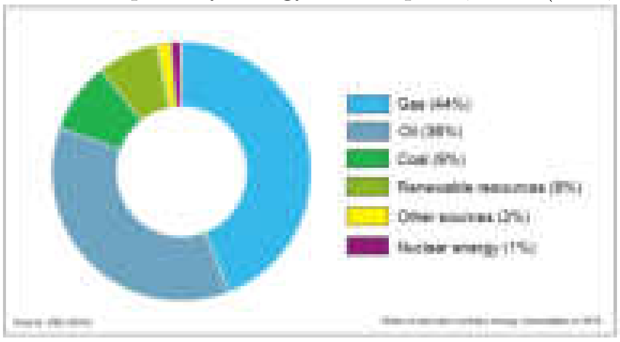
Introduction

In this following chapter the the problem statement will be explained. Buildings consume a significant amount of the energy, in the Netherlands and in the rest of the world. This section will explain what the problem is and why this thesis tries to give a explanation and solution. The section about research objectives will put this problem in a scientific perspective. It will also formulate the main question and the research gaps will be explained. The third section will explain the scientific and social relevance. It will give an answer why it is important to address the problem. The last section will give an overview of the thesis.

1.1 Problem statement

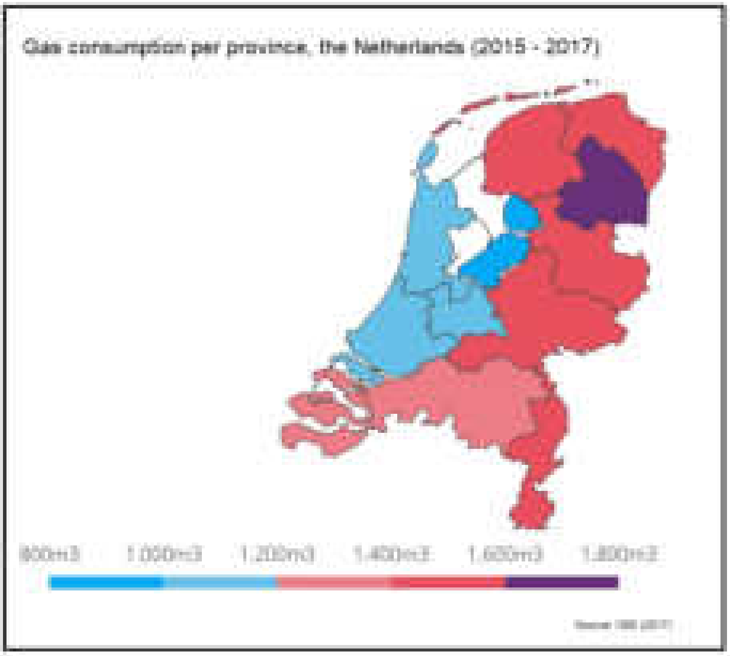
Buildings are responsible for about 40% of the **Dutch** national energy use (Energie in Nederland, 2019). They are therefore a main target of energy transition policies. In the Netherlands most of the resources for energy consumption are still fossil fuels. Figure 1.1 shows the share of resources in the Netherlands for 2019. Only 8% is renewable energy. The publicly available energy data from the Central Bureau of Statistics (CBS) in the Netherlands make it possible to map energy use geographically. The CBS provides open access data on the average annual gas and electricity consumption for postal code 6 areas (PC6) (CBS, 2019). Data for other areas, such as municipalities, districts, neighborhoods, and postal code 5 (PC5) and postal code 4 (PC4) are also available (Statline, 2020). Besides energy and gas use, the CBS also provides multiple maps, statistics and data about sustainability, energy labels and solar energy. A great deal of this information is available from databases such as *Publieke Dienstverlening Op de Kaart (2020)* and *energieatlas (2020)*. It has been demonstrated that the weather plays a part in the gas and energy consumption of individual buildings. The outside temperature is one of the main reasons for turning up the heating inside a house. There are significant differences between cities and rural neighbourhoods that can partly be explained by socioeconomic factors and construction types (Mashhoodi, 2019). But can the local climate also contribute to differences in energy consumption between postal code 6 levels? Energy models usually use climate data from the *Koninklijk Nederlands Meteorologisch Instituut* (Royal Netherlands Meteorological Institute or KNMI) meteorological stations. Generally, models use weather station **De Bilt** as meteorological reference, which makes it difficult to know how local weather differences influence the results. A first step understanding this is to investigate the variations of local weather and how these variations influence energy use for heating. To determine whether it is correct to use one central weather station or whether it would be better to use various local stations.

Figure 1.1: Share of resources in primary energy consumption, 2019 (CBS, 2019)



Today, buildings have become the main consumers of worldwide energy use. Research has indicated that 50% of energy is spent on heating and cooling buildings and industry in **Europa** (Popovski, Fleiter, Santos, Leal, & Fernandes, 2018). Figure 1.2 shows that there are even differences of gas consumption between the provinces in the Neterlands. It seems that the provinces in the west consume less gas than in the north and east. According to Popovski et al. (2018): *"The demand for heating is affected by several factors, such as the building shell, the type of heating system, the outdoor temperature, and occupant behaviour"*. Of these factors, outdoor temperature, irradiation and the factor wind are directly affected by climate change and are the subject of change. The research towards the cohesion between temperature and gas consumption are carried out broadly but variations in local climate differences are less common. Therefore, it is interesting and relevant to dive deeper in this particular topic. What is the relation between the difference in gas consumption and the climatic variations on a local level? This research will focus on this topic.

Figure 1.2: Gas consumption per province (2015 till 2017) (CBS, 2017)



1.2 Research objectives

Multiple studies on the effects of outdoor climatic variables have been conducted to date (Kohler, Blond, & Clappier, 2016; Britter & Hanna, 2003). This also applies to studies on energy consumption by households (Mashhoodi, 2019; Kavgic et al., 2010). The general objective of this research is to get to know how the local climate affects energy use, in order to improve existing models by using climate data at the right level. Another objective of this research is to visualize the climatic influences on cities' heating energy on a local scale. In general understanding, visualization is of great help to give an inside of what happens and geographical differences call for mapping the phenomena. To accomplish this goal, the following research question was formulated:

- **Are there outdoor climatic differences at postal code 6 level within large cities and what are the effects of these variations on heating energy consumption?**

To formulate a proper conclusion about the main question a literature study needs to be carried out first to find out where the research gaps about this topic are. The main question has a focus on the heating energy consumption. The most common type of heating energy in the Netherlands is still the consumption of gas *(Energie in Nederland, 2018)*. In the literature study a section needs to address the variations in heating energy use at a local level. The geographical scope of this research is the postal code 6 level. This is the smallest public local level available in the Netherlands that is also providing the average gas consumption per household for that specific area (CBS, 2019). A common way to measure the climatic influence on heating energy is to make use of the concept: degree day (Mindergas, 2020). Degree days are all the days in a year where the average outdoor temperature in 24 hours is below 18°C. It is assumed that for these days, heating is needed. Degree days are useful to increase comparability of energy use between different years and different geographical zones with different climates. By correcting heating energy usage with the corresponding degree days, different time periods can be compared without the temperature variations. The degree days are accounted for the difference in temperature but does it also account for the differences in other climatic factors such as wind, rain and irradiation? The literature study needs to investigate which kind of other climatic factors are influencing the cities' heating energy. It is also useful to see how other scientific literature tackles the question about climatic influence on heating energy on a local scale. After the literature study the sub-questions that will support the main question will be properly reformulated according to the exact research gaps. For now the intended research sub-questions are:

1. Are there variations in heating energy use on a local geographical level?
2. What are the effects of climatic variables on heating energy?
3. How can degree days helps with comparing gas consumption in different years and locations?

These sub-questions will be researched through literature study and refined then. It is important to clarify the scope of the research and its limitations. First of all, the focus of this research will be on the difference in heating energy level on the local level. The literature study could define the climatic variables. Later on in this research different time frames will be compared to investigate whether there is a difference between gas consumption and degree days. This research will not address solutions for energy transition, but seek to explain the effects of different climatic variables on cities' heating energy consumption.

The Royal Netherlands Meteorological Institute has 48 different weather stations; not all of these stations will be used, in order to make the research study manageable. A selection of these weather stations, in combination with other measurements from nongovernmental institutions such as the dataset from the website netatmo.com (Netatmo, 2020), will contribute additional climatic information for the case region of The Hague. The case region The Hague is chosen because of a collected and updated dataset from Netatmo. There are interesting climatic factors available in The Hague such as urban and green areas and it is placed directly at the coast. This will not explain the difference between other cities in the Netherlands but this data is (not yet) available for cities such as: Groningen, Maastricht or Enschede.

1.3 Scientific & social relevance

As stated in Mashhoodi (2019) a great deal of research has been conducted in the field of energy transition, energy consumption and the sustainability of households. Extensive research has also been done on the various forms of heating energy, but there are still scientific research gaps in the literature on the influence of climatic factors on cities' heating energy use on a local level. Also the study towards the concept degree day has it's main focus on the differences in temperature. This research also tries to investigate a more in depth focus on the other climatic factors. Especially on a local scale the climatic variables can differ from each other. A common phenomenon in the scientific literature is the concept of urban heat islands (UHI). Less commonly as addressed are the influences of wind, irradiation and rainfall. Therefore this research will conduct the research in the direction of these climatic factors. Are these factors influencing the cities' heating energy on a local level, in the form of PC6 areas?

Besides the scientific relevance, this research also has more social relevance. There is no doubt that the gas consumption of households and businesses needs to be reduced. Most of the factors that contribute to this consumption are already known. More sustainable electric devices, better home isolation and generating one's own energy using solar panels are possible solutions. More knowledge about the concept degree day and the influence of climatic phenomena can improve the houses and development of urban areas of the future to contribute to a more sustainable environment. With this knowledge municipalities can prioritize which neighborhoods to address first when making plans for the energy transition. In the Netherlands there is a program called *Regionale Energie Strategie*, translated in Regional Energy Strategy (RES). According to Rijksoverheid (2019), the national program supports the region in making the RES by developing and sharing knowledge, offering process support and facilitating a learning community. With this research an example can be made for creating maps that shows which neighborhoods are consuming most energy.

1.4 Overview of the thesis

This thesis consists of the following chapters: literature study, methodology, results, discussion and the conclusion. The second chapter: literature study, presents the theoretical background on geographical studies of energy use, the influences of the weather on the energy use and the concept of degree day and what the cohesion is with cities' heating energy use. This will be ended by reformulating and refining the research questions. The third chapter addresses the methodology used to answer these research questions. Also the geographical scope of this thesis will be discussed and what kind of data is required to execute the research. This chapter will also explain how this data has been processed for further research. The fourth chapter of this thesis explains the results and how they were obtained. This chapter will address the variation of gas consumption on postal code 6 level and the variation of the climatic factors on postal code 6 level. Chapter five will pinpoint and discuss the strong and weak spots of this research. It will also propose ideas for further research on this topic and will draw conclusions from the found results. Also the main- and sub-questions will be answered.

Chapter 2

Literature review

This chapter introduces the scientific background to this thesis. As stated in the introduction, much research on energy consumption has already been conducted. In recent decades concern about environmental pollution, exhausting our nonrenewable energy sources and our role in the greenhouse effect has increased. It is no surprise that 195 countries, including the biggest polluters, have agreed to the Paris Agreement (UnitedNations, 2015). This literature review focuses on geographical studies on energy use in the building stock, while the second section focuses on studies that have used districts, departments and postal code areas. Section 2.3 will address the influence of the weather on the energy use. Climatic and non-climatic factors will be discussed. Section 2.4 will discuss the degree days. It will explain how the degree day method works and will elaborate on the different types of degree days. The final section of the literature review will investigate what we know about the research gaps and what still needs to be investigated.

2.1 Introduction

According to D'Amico, Ciulla, Panno, and Ferrari (2019), the weather is an important factor to account for in the process to design an building. They stated that the weather represents the most import boundary condition to affect the dynamic behaviour of the building. Various studies have examined individual buildings and calculated their consumption of cooling and heating energy (Amasyali & El-Gohary, 2018; Clarke & Clarke, 2001; Storey, 2012). Also many studies are carried out with the use of the degree day method to predict the energy demand of buildings (D'Amico et al., 2019). Other studies have examined different cities as a whole and compared their sustainability, and discussed improvements on a much larger scale, such as that of Robinson et al. (2017). According to D'Amico et al. (2019), correcting results obtained from a generic building simulation tool with degree days will not give a reliable energy evaluation, although this is the most applied method. In the study of D'Amico et al. (2019) they stated: *"to demonstrate that the assessment of building energy demand through the use of the degree day is correct only if the determination of the climate index is a function of the same weather data"*. This climate index can vary per region and per period of measuring. Hence the *function of the same weather data.* D'Amico et al. (2019) created a climate index for his case region in Italy.

In literature, the influence of local climate on energy use is approached essentially from the point of view of urban heat islands (UHIs). These studies have investigated densely populated and mostly concrete cities. Through the irradiation of the sun, the lack of wind flow, the heavy movement of traffic, the outside temperature and energy use interact with each other (Britter & Hanna, 2003). Most of these studies have focused on warm summers and cooling energy, for instance air conditioners. This literature review will address this issue but also focus on the cooler winters and heating energy in the form of gas. Stated in the study from D'Amico et al. (2019), the assessment of buildings energy demand through the use of the degree day method is only valid if the climatic index is a function of the up-to-date weather data from the investigated case area. The climatic index D'Amico et al. (2019) created is based on the heating degree days per hour, the Mean Degree-Hours (MDH) for a day (24 hours). This section will investigate if it is better to use weather data closer to the investigated case region and if the degree day method can be improved with also accounting the historical solar radiation, wind and rain data. Weather conditions such as wind, rain and irradiation are common factors in studies on UHIs and on the local climatic level. Therefore it is important that these factors are also investigated in terms of the effect of local climate on postal code 6 areas and whether they influence gas consumption.

As stated in the previous paragraphs, one topic that has emerged in both energy consumption studies and papers on local climate is that of degree days. The degree days method of both cooling and heating is linked to weather conditions (Holmes, Tett, & Butler, 2017). The degree days method seems to be indispensable if we want to combine gas use and weather data from the local climate. Although the degree days method applies on the variable temperature and the local climate also need other indicators. When the actual energy use is analysed these results are normalized through using the degree day method to compare different years and regions with each other. The fourth section will investigate degree days and the degree days method and explain why it is crucial for this research study.

To date, multiple studies on degree days have been carried out. This also applies to studies on gas consumption by households. The general research objective is to combine the important scientific findings about these topics and use the relevant variables from both types of studies. This literature review seeks to create an up-to-date overview that substantiates the choice of certain variables to create a visual map of the climatic influences on cities' heating energy use.

2.2 Geographical studies on energy use of the building stock

According to Kavgic et al. (2010), in the **EU**, the **residential building sector** is responsible for approximately 22% of total energy consumption. In recent decades policymakers have acknowledged the problem of buildings' high energy use and seeing some possibilities in reducing energy consumption and lowering down the CO2 emissions in the residential building sector. There are two approaches to model possible reduction of energy consumption. The first is the top-down approach and the second the bottom-up approach (Kavgic et al., 2010). According to Kavgic et al. (2010), top-down modelling is on a aggregated level. This type of modelling is aiming to create series of data of energy (national) consumption of energy or CO2 in a historical perspective. These type of models try to examine the correlations between the energy sector and the economy. According to Kavgic et al. (2010), they can broadly categorised as econometric and technological top-down models. The bottom-up approach brings hierarchy in the data on disaggregated components. Afterwards these are combined according to some estimate to measure their individual impact on energy usage (Kavgic et al., 2010). Both approaches are useful, but working with postal code 6 data means that this thesis employs a top-down approach.

A study by Amasyali and El-Gohary (2018) indicates that the **building sector** represents 39% and 40% of energy consumption and 38% and 36% of CO2 emissions in the **U.S.** and **Europe**, respectively. These are higher numbers than those in the study by Kavgic et al. (2010) but they include the whole building sector. A large part of this energy is generated by fossil fuels that, as we know, contribute to CO2 emissions and cause air pollution and global warming. Amasyali and El-Gohary (2018) talked in their paper about the prediction of building energy consumption. This can be a main factor when it comes to policies and decisions for decreasing the energy consumption and reducing CO2 emissions. It can be of assistance in thinking about different type of building design and building operation strategies (Amasyali & El-Gohary, 2018). They mentioned two approaches in their paper: physical modelling and data-driven modelling. The physical modelling approach calculates building energy consumption based on detailed building and environmental parameters. Data-driven modelling does not perform such energy analysis or require detailed data on the simulated building, but instead learns from available historical data to make predictions. Many studies show that there is a difference between the actual and the calculated energy in the models, this difference is called the energy-performance gap (van den Brom, Meijer, & Visscher, 2018). Multiple researchers showed that there is a discrepancies between the actual and theoretical energy consumption (Majcen, Itard, & Visscher, 2013; Guerra-Santin & Itard, 2010). This thesis will not discuss the prediction aspect of the modelling, but the data-driven approach provides a good example of how to analyse the current energy consumption situation and learn from it to make better models.

Howard et al. (2012) conducted a study on energy consumption in New York City according to zip code areas. This extensive study used the annual electricity and natural gas, steam, and fuel oil consumption for 191 zip codes. The paper mentioned that it is important to note that weather has a large impact on energy consumption from year to year, indicated by the high correlation between the consumption of fuel oil, natural gas, and to some extent steam, and heating degree days. In addition to annual energy consumption, the study collected information about the building stock. Howard et al. (2012) collected their data from a geo-database called PLUTO, used by the New York City Department for city planning. The Dutch municipalities and government bodies also collect geo-rectified data and store these in different open source databases, such as *Publieke Dienstverlening Op de Kaart (2020)* and *NGR (2020)*. As for the categorisation of the zip code levels, Howard et al. (2012) chose the following: two-family dwellings; primarily one-family dwellings with two stores or offices; primarily one-family dwellings with one store or office; primarily two-family dwellings with one store or office; primarily three-family dwellings with one store or office; primarily four-family dwellings with one store or office; and three-family walk-up apartment or four-family walk-up apartment, placing the residential areas into the category 'residential 1–4 family'. This thesis does not differentiate between types of houses, but distinguishes different built-up areas, such as industrial areas and office buildings that will not be a part in this thesis because this study has a focus on the residental areas. Howard et al. (2012) commented on their findings as follows:

"As one would expect a consumption normalized by block area would show particularly high values for parts of the city where the buildings are tightly packed and tall. Hence a block located in midtown Manhattan, has one of the largest annual energy consumption when normalized by block area consuming as much as 8000 kWh/m2. To provide a sense of scale in terms of power, the power consumption of the block averaged over all hours in the year is about 17.6 MW." (Howard et al., 2012)

The study from Howard et al. (2012) uses more than only the gas use for households and compares more parameters than only the residential areas but it is usefull to see if the case area for this thesis shows some of the same results.

Most studies have conducted research in just one city and the question is whether this is representative for a whole country (Mashhoodi, 2019). The results of the study by Mashhoodi (2019) demonstrated that the effect of land surface temperature (LST) on the heating of buildings (household energy consumption or HEC) is a spatial variant. In other words, its impact cannot be generalised for all urbanised neighbourhoods in the Netherlands. The impact of LST on HEC is significant in roughly one third of neighbourhoods, where it accounts on average for 6% of total HEC, and is often exceeded by the impact of other determinants of HEC (Mashhoodi, 2019). In this respect, while studies and policies regarding HEC acknowledge the impact of LST, it should be noted that this impact is only meaningful if it is studied alongside other socioeconomic, stated that low-income households consume more per m2 that the high-income households, this occurs in all types of houses. According to van den Brom et al. (2018), besides occupant behaviour the building characteristics are also affect energy use for heating.

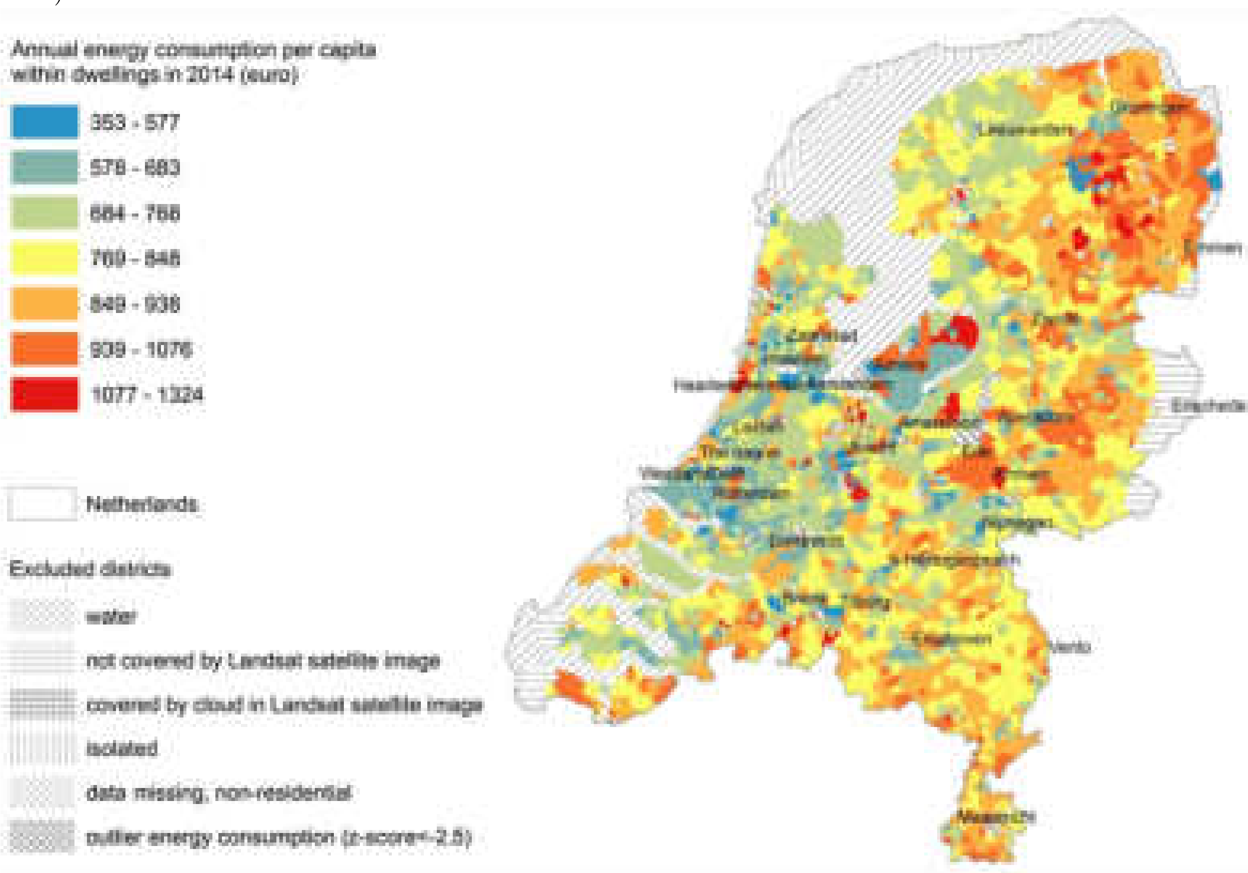
Zhang et al. (2018) created a model to determine the energy footprint of the residential sector in a metropolitan region at neighborhood scale. According to them this was challenging, mostly because information on household energy use is difficult to obtain from utility companies, due to legal and privacy concerns. Zhang et al. (2018) made use of household surveys. Besides these surveys, they also collected data such as the type of housing, occupant behaviours, types and number of appliances and devices, and built environment characteristics. Especially the last aspect can be useful for this thesis. This thesis will focus on the local climate factors, but these are not the only factors that influence gas consumption. It is necessary to keep an eye on other factors, such as the characteristics of buildings and neighbourhoods. *"Model coefficients can be interpreted to identify the sensitivity of residential energy consumption to property characteristics, user attributes, and urban form parameters such as density and land use diversity"* (Zhang et al., 2018).

The study by Zhang et al. (2018) developed a model to address the previously mentioned objectives list using a novel bottom-up technique for generating estimates of neighbourhood-level residential energy consumption, by matching RECS (Residential Energy Consumption Survey) data with other publicly available datasets. They validated the model using data from the Atlanta metropolitan area. For this thesis, it can also be very useful to validate the obtained data with a reputable dataset, to avoid outliers in a less qualitative dataset. Zhang et al. (2018) concluded that cities continue to absorb growing populations and economic activity, invest in infrastructure, implement zoning requirements, building codes, road pricing and an array of other regulations and incentives, while urban forms will evolve, and so too will energy consumption patterns. Robinson et al. (2017) created a model for the same case area as Zhang et al. (2018), namely Atlanta. According to Robinson et al. (2017), the models are based on only five commonly accessible building and climate features, and can therefore be applied to diverse metropolitan areas in the United States and other countries by replicating their methodology. These five features are: square footage of the floors in the building, principal building activity, number of floors, and heating and cooling degree days.

Mashhoodi, Stead, and van Timmeren (2019) all ready carried out a case study towards energy consumption on dwellings. This case study has an spatial resolution on neighbourhood units shown in figure 2.1. The annual energy consumption per capita within dwellings is indicated with the colours from blue to red that is represented in euro's. Noticable is that there is also satalite data visable in figure 2.1, this data doesn't represent the energy consumption but the landuse types such as water or green areas. Mashhoodi et al. (2019) divided the neighbourhoods on annual energy consumption per capita per dwellings in euro's. At first sight it looks like the urban area's around the Randstad (Amsterdam, Rotterdam, The Hague and Utrecht) show lower values than for instance the area's around Groningen. According to Mashhoodi et al. (2019): *"The reason for excluding the latter is that the CBS database on households' gas and electricity consumption merely reports the consumption supplied from the distribution grid of gas and electricity in the neighbourhoods. The supply from district-heating systems or solar panels, however, is not reported by the CBS database. It is likely that a neighbourhood with an abnormally low level of consumption in the CBS database is provided with district-heating or a large number of solar panels".*

In this case study one of the conclusions is that there is significant impact from local variables on the differences between neighborhoods in the Netherlands. The variables *summer days* and *Household size* have the most impact on the household energy consumption, according to Mashhoodi et al. (2019). Local determinants all have a similar effect in all the investigated neighbourhoods but the magnitude can differ between these neighbourhoods. Mashhoodi et al. (2019) give an example:*" it is established that a higher Building age, as a proxy for buildings' energy efficiency, is associated with higher levels of HEC, such an effect is substantially smaller in highly-urbanised neighbourhoods".*

Figure 2.1: Annual energy consumption per capita per dwellings in 2014 (euro) (Mashhoodi et al., 2019)



2.3 How does the weather influences energy use?

In recent decades, concern about energy use and its impact on climate has become a recurring issue, according to Li, Yang, and Lam (2012). Reports by the Intergovernmental Panel on Climate Change (IPPC, 2021) have raised public awareness of energy use and climate change implications, and helped to generate interest in having a better understanding of energy consumption and its correlation with prevailing weather conditions. The phenomena of urban heat islands is a common one. Detailed descriptions of the UHIs are carried out. According to de Nijs et al. (2019), multiple factors are responsible for the UHIs, including climatic factors. This section will investigate what these climatic factors are and if these factors also apply for differences in gas consumption. Building energy consumption is weather dependent, according to Lazos, Sproul, and Kay (2014). As stated in the introduction of this chapter the common way to compare gas consumption through different years is by using the degree day method. Before this research will discuss the degree day method it is useful to evaluate the climatic factors that can have an influence on gas consumption. Besides the climatic factors there are also other aspects that influence the gas consumption, stated in the section: geographical studies on energy use of the building stock. The last part of this section will discuss these non-climatic factors and will explain how this further research will address these factors.

2.3.1 How weather influence energy use

In Chapter 5 from Visscher (2012) Laure Itard describes that the weather always influence the energy use. The energy balance closely relates to outdoor conditions like temperature, humidity, wind speed and solar radiation (Visscher, 2012). The weather variables are never constant and behave in a strongly dynamic way. The weather conditions can influence the energy use in the following ways:

1. temperature and wind influence heat losses through construction;
2. solar irradiation influences solar gains inside the building;
3. humidity influence the moisture balance and therefore also the heat gains and losses.

As said before, the weather conditions are never constant. The wind speed can varies from 0 to 100 m/s, the solar radiation can be influenced by cloud cover and the relative humidity may vary from almost zero to 100% (Visscher, 2012). All these differences in weather conditions are stored in historical databases collected by weather stations. The time resolution for these historical datasets can differ from yearly data to hourly data. According to Visscher (2012): *Some of these data, like in the so-called TRY (Test Reference Year), TMY (Typical Meteorological Year) or SMY (Standard Meteorological Year), are statistically worked out in order to be used for the specific aim of making accurate energy or heat/cooling load calculations."*

These statistical standards are constructed and expected to be valid for certain regions, but they don't account for local climatic distribution. According to Visscher (2012) the weather conditions that are of interest are influenced by other local variables such as the shadow of trees, other buildings, the protection of wind (also by trees and other buildings) or by the land use type that has a strong influence on the air temperature. The temperature in stony environment can be up to 5°C in comparison to a green environment (Visscher, 2012). These local variations of the climatic factors influence the local energy use. One of these phenomena is the urban heat island and will be discussed in the next paragraph.

2.3.2 Urban heat islands and the local climate

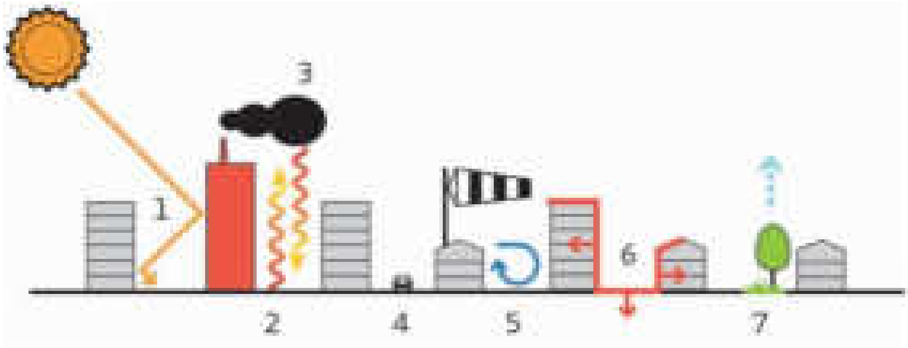
This thesis doesn't focus on the topic of UHI's but the urban heat islands can expose some interesting findings that also apply for the effects and extend of outdoor climatic factors of heating energy. The UHIs are local climatic phenomena that occurs with the right circumstances and differ within the city boundaries.

Figure 2.2 presents a schematic representation of the effects of UHIs. Urban areas usually experience higher external air temperatures and reduced natural ventilation compared to the rural countryside (Kohler et al., 2016). A combination of several factors contribute to this phenomenon, namely: waste heat that is discharged in the atmosphere due through human activity; a high urban density of buildings that reduce the skyview factors; a lack of cooler green areas that contribute to the evaporation; and higher heat capacity of built-up materials. Policy makers and urban planners are trying to create a solution for the problem of UHI mitigation, due to the warming of the global climate there need to be a solution for reducing the energy demand for building cooling. According to Kohler et al. (2016), urban heat islands have been shown to contribute to annual energy savings in buildings by greatly reducing energy demands for space heating (approximately 30%) in comparison to a much smaller increase in energy required for indoor cooling (more than 13%). Thus it is vital to assess more precisely the effects of UHI mitigation measures on energy requirements for both building space heating and cooling, based on a classification of the building types of the urban housing stock (Kohler et al., 2016).

Due to climate change, the temperature rises and causes more heat waves. This results in warmer days and nights. All governmental bodies in the Netherlands must identify the bottlenecks relating to flooding, heat and drought (de Nijs et al., 2019). According to de Nijs et al. (2019), a variety of heat maps is already available, but a standardized method to combine UHIs and household energy consumption is lacking. According to Mashhoodi (2019), it is widely accepted that UHIs affect household energy consumption. This could also be an explanation of the lower annual energy consumption in the Randstad in figure 2.1. Besides the effect of UHIs on the heating energy required by a city, it is also important to consider the effect when the city is cooling down. In other words: how are UHIs slowing down the cooling effect of a city? According to Pijpers-van Esch (2015), there are seven known effects that influence UHIs, namely:

1. Absorption of short-wave radiation from the sun;
2. Decreased long-wave radiation heat loss from street canyons;
3. Absorption and re-emission of long-wave radiation by air pollution in the urban atmosphere;
4. The release of anthropogenic heat by combustion processes, such as traffic, space heating and industries;
5. Decreased turbulent heat transport from within streets caused by a reduction of wind speed;
6. Increased heat storage by building materials with large thermal admittance; and
7. Decreased evaporation from urban areas because of waterproofed surfaces.

Figure 2.2: Urban Heat Islands(de Nijs et al., 2019)



In general, higher buildings mean less air circulation and more trapped heat energy in the urban canopy layer, and therefore higher air temperatures, according to Britter and Hanna (2003). Research has indicated that anthropogenic heat, such as waste heat from industry, cars, and air conditioning units, increases densely populated UHIs by 1–5°C. Lund et al. (2014) found that heavily trafficked streets in Tel Aviv accounted for up to 2°C warmer temperatures. The research into UHIs can be divided into broad-scale approaches that use remotely sensed satellite images to locate UHIs by surface temperature and fine-scale approaches that use small weather stations or mobile measurement devices to determine UHIs' air temperatures. Surface temperature variations are greatest during the day, but air temperature variations are greatest during the night. Jenerette et al. (2007) found that temperature differences within metropolitan areas varied systematically with neighbourhood-level social characteristics. Poorer neighbourhoods were significantly hotter, one the the reasons could be that there is a correlation with the type of buildings and the density of construction, for instance: more pavement and little green. Every 10,000 Dollars increase in a neighbourhood's annual median household income was associated with a 0.28°C decrease in the surface temperature at 10 a.m (Stone & Norman, 2006). All types of vegetation are not equally effective in mitigating UHIs. Grass is not as effective as trees that cast shade and contribute more moisture.

Besides the climatic factors the UHIs are also influenced by land use/ land cover. Combining measures of land use/land cover, building configuration, and adjacent heat sources and sinks that are easily collected, Stone Jr (2012) developed a practical set of metrics to understand the key drivers of elevated urban air temperatures. (Coseo & Larsen, 2014) conducted a study in 38 cities in the US, across eight bioclimatic regions. The percentage of impervious surface explained approximately 70% of the land surface temperature. Among the cities, the amount of variance explained by the percentage of impervious surface varied from 60% to 90%. Stone Jr (2012) believes that land use change and waste heat contribute more toward this warming than global climate change. The researchers identified six land use/land cover types:

- high-density urban use (largely commercial and industrial);
- low-density urban use (largely residential);
- cultivated or exposed land;
- cropland or grassland;
- forest and
- water.

Another study for land use cover by Lo and Quattrochi (2003) is executed. they stated:

"The temperature of UHIs was greater (6.5–9°C) in areas that had displaced forested environments relative to temperate grasslands and savannas (6.3°C) and tropical grasslands (5°C). Urban areas in temperate climates may therefore have greater UHI intensity relative to semi-arid and arid areas. Finally, we identify seven variables to represent adjacent heat sources, heat sinks, and contextual characteristics" (Lo & Quattrochi, 2003).

According to Santamouris, Cartalis, Synnefa, and Kolokotsa (2015), the increase of land surface temperature in hot climates can also be caused by the increase of the ambient temperature around buildings. The increase of ambient temperature around buildings is caused by the increase of energy consumption for space cooling. Various papers by Mashhoodi (2019) have also stated that a higher LST decreases the amount of energy consumed for space heating in colder periods. Although the ambient temperature increases in the vicinity of concrete buildings the ambient temperature will decrease around green areas. The vegetation absorbs solar radiation in order to grow, by which this radiation will not be absorbed by the ground. In other words, concrete buildings absorb solar radiation and radiate it to the surround environment, the green areas absorb the solar radiation and will not radiate it to the surround environment in order to cool down the ambient temperature.

UHI effects can somewhat reduce energy use for heating. However, most studies have verified that UHI leads to a dramatic increase of energy consumption for cooling in hot summer. This research has a focus in the opposite direction and is about the climatic factors influencing the gas consumption, mostly in the winter. Although UHIs focus mostly on summer days and the effects of this phenomenon on the surrounding environment, there are still similarities and overlapping topics worthwhile mentioning. The ambient temperature that is effected by radiation both from the sun or from concrete buildings and surrounding asphalt and pavement, the absorption from radiation by the green areas and changes in turbulent heat transport and speed of wind. Although this is less in wintertime because of a decrease of sun hours and a low sun. The urban form and wind direction also affect the UHI intensity (Sen & Roesler, 2020). Wind patterns in cities can be different in comparison with green areas. Wind can be blocked by buildings and create less wind on some spot while other places experience more through wind corridors. Sen and Roesler (2020) stated: *"The UHI intensity of an urban canyon averaged over the probability of wind blowing from different directions, was found to have a strong correlation with the probability that it is an interior canyon that experiences lower wind speed on account of turbulent dissipation"*. Also the decrease of evaporation from urban areas because of waterproofed surfaces plays a roll: water evaporated by plants contributes to cooler areas. Urban warming also affects water quality. Rainwater that comes off heated roofs and roads on hot summer days can become warmer from a few degrees to as much as 17 ° C (EPA, 2008). It is all ready known that the phenomena of UHIs affects the energy consumption on a local level. This research will further investigate in how far climatic factors influence heating' energy consumption on a local level. The climatic factors will be discussed in the upcoming sections. also the non-climatic factor: green areas will be discussed. The non-climatic factor: concrete buildings will be discussed in the chapter discussion. Besides the climatic effects another factor seems to play a part on the gas consumption per households, mentioned in the section: Geographical studies on energy use of the building stock. This factor is the improvement of the sustainability of the building stock. The last part of this section will focus on this factor.

2.3.3 Rain and humidity

Most of the studies on UHIs are focusing on the evaporation of water and rain. As said in the previous section: urban warming also affects water quality. Rainwater that comes off heated roofs and roads on hot summer days can become warmer from a few degrees to as much as 17 ° C (EPA, 2008). The surface water where the rain eventually end up is also getting warmer. This can cause unwanted algae population that increases due to global warming and can lead to fish death and botulism. Certain substances can also pose a danger to humans (EPA, 2008). According to the KNMI (2014) the average rainfall over land has increased in the mid-latitudes of the Northern

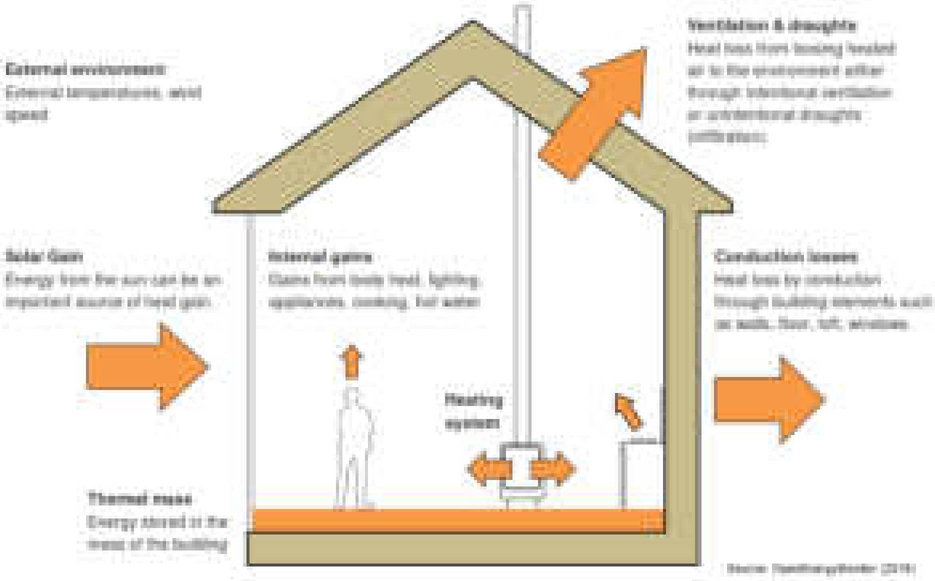
Hemisphere since 1901. They are also certain that the behavior of humans have an effect on this increased rainfall. The amount of water vapor in the air, the so-called relative humidity has increased since the 1970s. This is partly due to the warming, because warmer air can contain more moisture (KNMI, 2014). Additionally between 1910 and 2013, annual rainfall in the Netherlands increased by 26%. Between 1951 and 2013, the increase was 14%. All seasons except the summer are getting wetter. From 1951 the number of days per year increased in the Netherlands with at least 10 millimeters of precipitation in the winter or at least 20 millimeters of precipitation in the summer. The effect on heavy rain is even greater. Observations show that during the most extreme rainfall, the amount of rainfall per hour increases by about 12% per degree that it is getting warmer. The temperature had obviously an effect on the amount of rainfall. But does the rainfall also has an effect on the local temperature. The chapter: results tries to find out if this is the case for investigated region.

Besides the rain another factor has an influences on the temperature. In this case the windchill. The humidity plays a big part in windchill. In winter when the air is humid, people don't like it and are heating more because it then feels less humid. And in summer the humid air fells much warmer than dry air. The human body cannot evaporate sweat in a more humid environment so there is a wish for more cooling, that also influences the energy consumption as mentioned in earlier sections. Ihara, Genchi, Sato, Yamaguchi, and Endo (2008) all ready did a study towards city-block-scale sensitivity of electricity consumption to air temperature and air humidity in business districts of Tokyo, Japan. They found that the air humidity during summer will increase and the energy consumption for dehumidifying will increase.

2.3.4 Wind

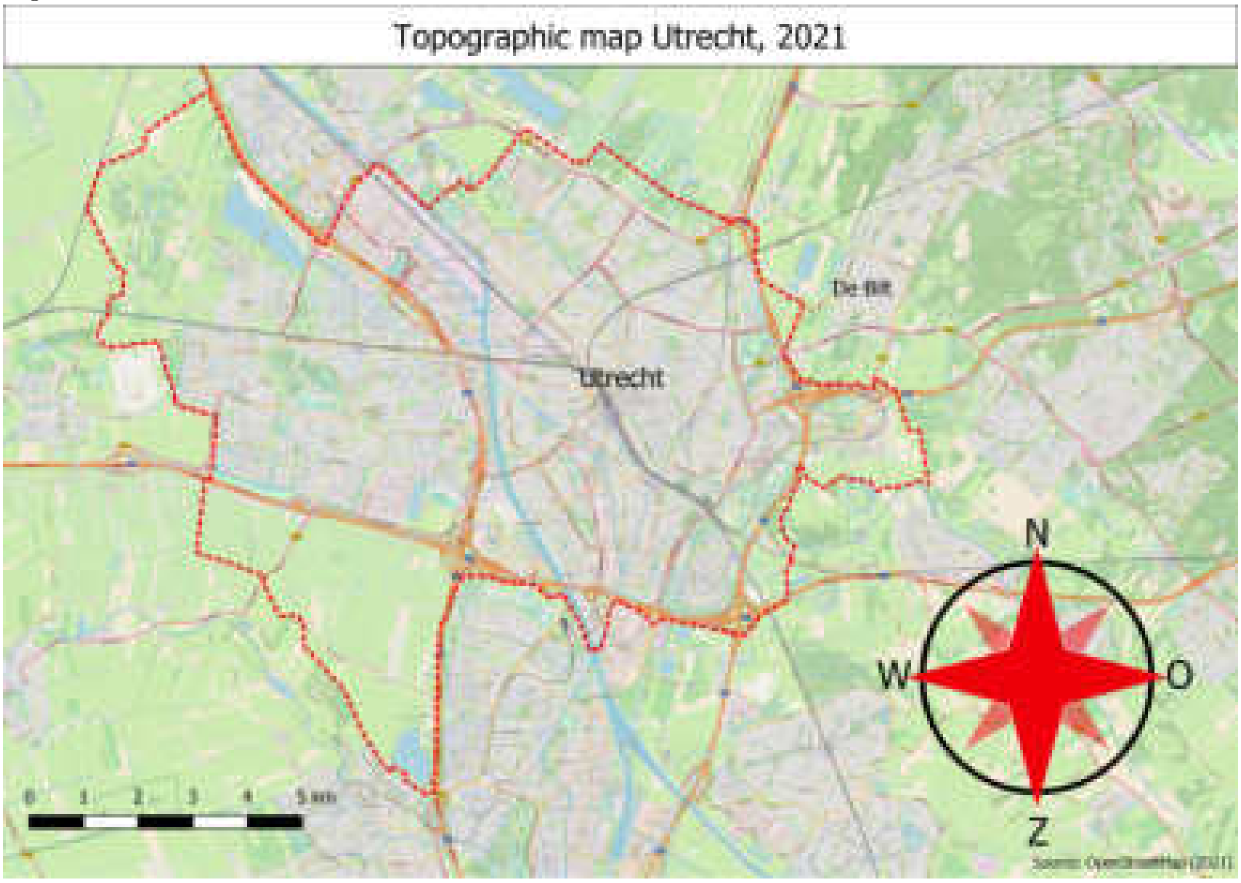
In Chapter 5 from Visscher (2012) Laure Itard describes that there are multiple possibilities for heat loss, the two most common through the climatic factor wind are infiltration and transmission. Figure 2.3 shows the infiltration through ventilation and draughts and transmission via building elements such as walls, floors and windows, also know as: conduction losses. The infiltration losses are very much dependent on wind speed, direction and the orientation of façade. The transmission losses are very much depended on the heat convection coefficient, which depends a lot on the wind.

Figure 2.3: Heat loss through infiltration and transmission (OpenEnergyMonitor, 2019)



In the section about UHIs it has been noticed that wind and wind direction have an effect on the ambient temperature. Horizontal transport (advection) from the heat loss from city centres depends on the wind direction (KNMI, 2010). The KNMI has investigated this effect in the city of Utrecht, and split the profiles into four wind direction classes: north, east, south and west. Figure 2.4 shows the municipality of Utrecht and the wind directions. It turns out that profiles for the wind directions west and south have a similar temperature shape and also the profiles for east and north. It is striking that for the directions west and south the temperature outside the city of Utrecht continues to drop, while for the directions north and east the temperature decrease stops around 12 kilometer. It is also noticeable that the temperature differences between the closest part and the farthest part of De Bilt (a small city close to Utrecht) are approximately 0.7 ° C smaller for west and south directions than for north east directions (KNMI, 2010). In the case of westerly wind, these aspects can be an indication of the advection of heat from Utrecht to De Bilt.

Figure 2.4: Utrecht



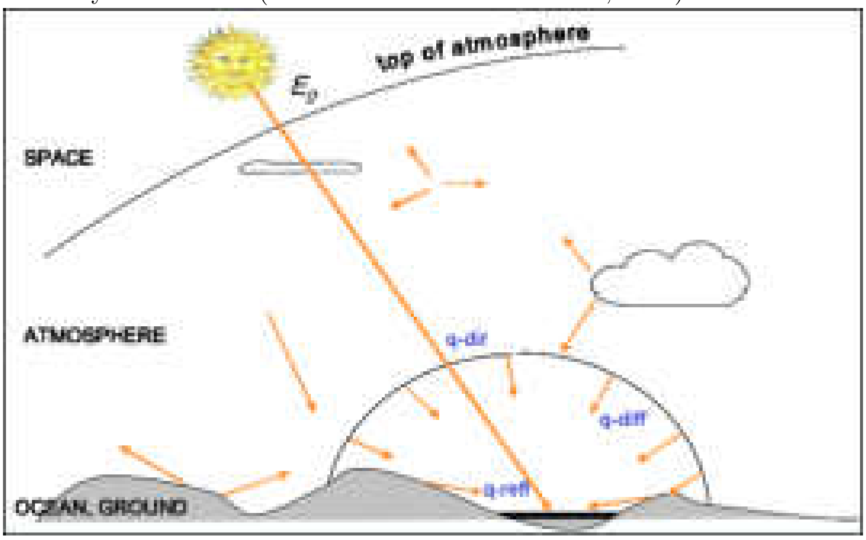
Wever (2008) has also used KNMI data from the measuring station in The Bilt. The wind experiences friction from the earth's surface. That friction ensures that the airflow, especially at the surface, is slowed down. This inhibition is stronger as higher and/or larger objects (such as trees, shrubs and buildings) on the surface of the earth are rougher. The roughness of the earth's surface is measured by the so-called roughness length. The roughness length increases as the objects on the earth's surface are higher. The roughness length can range from about 0.03 m for grassland up to 1.0 m for forests. Winds at night tend to be lighter than during the day, due to decreased solar heating and less atmospheric mixing (Britter & Hanna, 2003). According to de Nijs et al. (2019), the wind speed has an influence on the turbulent heat movement within the city. Therefore it is important to investigate the wind speed and direction in the chapter: results. This climatic factor probably has effects on the heating energy the winter.

2.3.5 Irradiation

According to van het Schip (2010), the radiation emitted by the sun is experienced by humans as heat. The amount of energy carried by this radiation is also an important aspect of the solar phenomenon.

The radiation entering the atmosphere is partly absorbed by the ozone, water vapor and dust particles. It is partly scattered by the air molecules and the dust particles, while the remainder will be transmitted. The transmitted radiation that reaches the earth's surface is called direct radiation. Diffuse radiation refers to the part of the scattered radiation that reaches the earth's surface. Some part of the total radiation incidence on the earth's surface is reflected, and is known as reflected radiation. The intensity of solar radiation is understood to be the amount of energy falling on a surface, measured per unit of time and per unit of surface. Outside the earth's atmosphere, the intensity is inversely proportional to the square of the distance between the sun and the earth. The intensity of direct, diffuse and reflected radiation is indicated by q-dir, q-diff and q-refl, respectively. All these factors represent radiation as a function of the place on the earth, the time, the orientation of the irradiated surface and the clarity of the atmosphere. Figure 2.5 illustrate these factors for the intensity of radiation.

Figure 2.5: Intensity of radiation (Schroedter-Homscheidt et al., 2016)



The so-called atmospheric turbidity factor (T) is used to measure the clarity of the atmosphere. The value of T varies from 2 to 6. The following guideline values can be used

- T = 2.7 for mountain areas
- T = 3.5 for rural areas
- T= 4.3 for large cities
- T= 6.0 for heavy industrial areas

The atmospheric turbidity factor differs for different types of land use cover. With an increase in T, direct radiation decreases and diffuse radiation increases. The q-dir is important as it indicates the direct radiation of the sun and the q-refl is important as it indicates the long-wave radiation heat from street canyons (Pijpers-van Esch, 2015). Also q-dir has a direct influence on buildings and the outside temperature.

The previous paragraphs described how the irradiation works but how does this solar radiation influences the heat demand of buildings? According to Laure Itard in the book *Sustainable Urban Environments: An Ecosystem Approach* from Visscher (2012), solar radiation penetrates into buildings through windows and is absorbed by the outside surface of the walls and roof. According to Visscher (2012), the solar radiation will be absorbed by the surrounding walls and floors through the multiple reflection on different surfaces instead of absorption by the indoor air, as what is the common expectation. *"In addition to the thermal mass, the heat gains through solar radiation depend essentially on the orientation, size and properties of the windows"* (Visscher, 2012). The solar radiation could be an important factor for local variations at postal code level. As said in the previous section about energy use in the building stock, cloud cover and the shadow of trees can create a difference in solar radiation on a local level and influences the energy use. Also the ground coverage can play its part due the absorption by plants in more green environments.

2.3.6 Sustainability

According to Marjaba, Chidiac, and Kubursi (2020), sustainability depends on three main aspects: the environmental impact, the social impact and the economic impact. For this thesis, the environmental impact is the most important. Marjaba et al. (2020) listed various sub-factors that contribute to the environmental impact and various factors that determine how sustainable buildings are. For example, environmental impacts are balanced with social and economic impacts, pollutants that impact land and habitat quality have to be minimized and the amount of resources used can be measured in terms of material volume or mass, energy consumption, water consumption, etc. Being more efficient with using energy is gaining more importance nowadays. Both the exhausting our energy resources and the negative effect it has on the global environment that result from increasing carbon dioxide emissions. According to Aydin and Brounen (2019), minimizing the thermal waste heat by buildings is one of the most important energy conservation measures. Therefore minimum thermal efficiency standards for new buildings are mandatory. The most common way to reduce the primary energy use, according to Laure Itard on chapter 5 in the book from Visscher (2012), is by following these steps:

- using renewable energy;
- using cogeneration; and
- using waste heat.

With the cogeneration is meant combined heat and power. According to the book from Visscher (2012): *"Using cogeneration as much as possible (heat and power, heat and cold, power heat and cold)"*. Individual sustainability factors are hard to measure on their own. According to *Milieudefensie (2019)*, isolation a house is a way of saving energy. How do we measure isolation per PC6 level? How do we know if buildings are designed to save energy? All these individual factors are brought together under energy labels for residential buildings. *"In the Netherlands, the energy label is calculated based on both the building characteristics and modelled heating behaviour of occupants"* (van den Brom et al., 2018). The energy label indicates how energy efficient a building is and makes it clear which energy-saving measures are possible (Rijksoverheid, 2020). According to Majcen et al. (2013), the energy labels are theoretical gas and electricity consumption, obtained by a model based on the characteristics of the building like insulation, floor area and its heating, ventilation and cooling systems. The label classes for buildings range from A to G, with buildings with an A label being the most energy efficient and buildings with a G label being the least energy efficient. The label also provides an overview of the housing characteristics, such as the housing type, insulation, double glazing and heating.

Eventually the main point of the theses of Majcen et al. (2013) and van den Brom et al. (2018) is that the energy models used to calculate the heating energy consumption are very often far from reality. In this case the theoretical versus the actual situation. So, because the models have been shown to not correspond to reality, they try to make them better. One of the problems may be that they do not account for local variation in climate.

2.4 Comparability of energy use under differing climates: degree day method

In the earlier sections the term degree day came across. The degree day method is the most common way to calculate and compare the gas consumption between different years and regions. The following sections will explain the degree day method. First the common degree day will be explained and the term weighed degree days will be introduced. Second the degree day correction will be discussed. With this correction the gas use of different years can be compared with taking the climatic factors into account. The last part will discuss the difference between heating degree days and cooling degree days.

2.4.1 Degree days & weighted degree days

If one compares gas consumption for a certain period with the same period in another year there can be significant differences in the amount of consumption. However, one cannot draw any conclusions about whether there has been a saving or waste in gas consumption. Perhaps it was colder in one year and milder in the other. Weather influences can give a distorted picture of the figures. By calculating degree days when comparing natural gas consumption, one can minimize the influence of the varying outside temperature.

To determine degree days, the starting point is that there is no gas consumption on a day when the average outside temperature is higher than the desired average inside temperature. However, if the outside temperature is lower, buildings need to be heated and degree days are counted. For degree days calculation, 18°C is taken as the value for the average indoor temperature in the Netherlands, according to the Dutch KNMI. Not in all parts of the world applies the same boundary. For instance in the United States the boundary for degree days is 15°C (Visscher, 2012). This thesis will focus on the Dutch degree days and in further discussion of the degree days the boundary of 18°C will be used. The daily (24 hours) average outside temperature of colder days is subtracted from the daily average inside temperature of 18°C. If on one day (24 hours) the outside temperature was an average of 10°C, the calculation is as follows: 18 - 10 = 8 degree days. If the average outside temperature over 24 hours was higher than 18 degrees, one always arrives at 0 degree days. This is done for each day of the period under consideration and the daily degree days are summed up for all days of this period.

In addition to the outside temperature, other weather conditions affect the heating energy consumption for example, the irradiation of the sun on the house, a rainy day and the wind factor. According to the KNMI, to minimize the influence of these factors on the calculations, degree days are multiplied by a seasonally dependent weighting factor, called the weighted degree days (WDD). The weighting factor is as follows throughout the year:

- From April to September: 0,8
- March and October: 1,0
- From November to February: 1,1

These correction factors do not always give credit to the weather circumstances in the Netherlands, and certainly not for local variations. In the calculations from the website: *www.mindergas.nl* (Mindergas, 2020) these weighting factors are not being account for. In figure 2.6 the calculation for heating degree days (HDD) is given. Where *H(x)* is the Heaviside step function such that *H(x)* = 0 if x is smaller than 0 and *H(x)* = 1 otherwise. Heating degree days only occurs on days below the threshold (TH). Where N is the total days (in the case of one year = 365) and the Xi stands for the daily temperatures. Previous studies in Europe and the United States have used different thresholds, often, 18.3°C (Holmes et al., 2017).

Figure 2.6: HDD (Holmes et al., 2017)

$$\mathrm{HDD} = \sum_{i=1}^{N} (T_H - x_i) H(T_H - x_i).$$

2.4.2 What is the degree days correction?

One can take an example from the website *mindergas.nl* (Mindergas, 2020), which calculated the percentage of gas used over two different years for one building. This includes the gas that is used for heating the building but also for cooking or heating water. It is not possible to extract these different types of gas use because they use the same supply tube. Although these other types of gas use represent a small percentage of the total gas use. In 2006, 1800 cubic meters of gas were used and there were 2718 weighted degree days. After taking energy-saving measures, one would use 1100 cubic meters of gas in 2007 and there were 2565 weighted degree days. As can be seen from the number of degree days, 2006 with 2718 weighted degree days was colder than 2007 with only 2565 weighted degree days. In 2006, the gas consumption per weighted degree day was as follows: 1800 ÷ 2718 = 0.66 m3 / degree day. In 2007 this was: 1100 ÷ 2565 = 0.43 m3 / degree day. Gas consumption per weighted degree day in 2007 was considerably lower than in 2006. After correction for weather influences, energy-saving measures resulted in 35% less energy being used: 0.43 m3 / degree day corresponds to 65% of 0.66 m3 / degree day, resulting in 35% less consumption. If one did not take the weather influences into account, one would end up with a saving of 39%, as 1100 m3 corresponds to 61% of 1800 m3, meaning 39% less. The energy savings would then have been overestimated due to the milder weather in 2007 compared to 2006.

2.4.3 What is the difference between heating degree days and cooling degree days?

Degree days refer to a temperature index used to understand the impact of weather. In general, degree days are being calculated as the sum of all the measurements that are above and below a certain threshold. The heating degree days (HDD) are the average temperatures below this certain threshold and the cooling degree days (CDD) are the average temperatures above this threshold. Just like degree days for heating were defined in 2.4.1, one can also define degree days for cooling as showed in figure 2.7, the only difference is the use of the threshold (TC) in the case of Holmes et al. (2017) they use the threshold of 22.0°C.

Figure 2.7: CDD(Holmes et al., 2017)

$$\mathrm{CDD} = \sum_{i=1}^{N} (x_i - T_C) H(x_i - T_C).$$

According to Krese, Prek, and Butala (2012): *"Cooling degree days are defined as the sum of positive differences between outdoor air temperature and a reference temperature over a certain time period. The reference temperature is the building's balance point with the maximum outside temperature, at which no cooling is required to hold the same thermal comfort inside the building"*. The reference temperature is also called the base temperature (Day, Knight, Dunn, & Gaddas, 2003). Heating consumption depends for the largest part on the outdoor temperature, which makes it easy to define a heating reference temperature. However, cooling depends much less on the outdoor temperature than on other factors, by which the reference temperature may differs from building to building. Some of these factors are the building's characteristics, such as the materials used for the building, and much more important the solar radiation coming into the building through windows. Also internal factors (light and a crowded building) or external factors (infiltration and fenestration) can play a role. These factors are specific for each building, so the base temperature should be determined for each building separately, according to Day et al. (2003).

Besides the previously mentioned problem, the cooling degree days method has another flaw: it assumes that a building's total cooling load consists only of sensible load components (Huang, Ritschard, Bull, & Chang, 1986). According to Day et al. (2003): *"the most basic way to describe the CDD is the time integral of temperature differences between a defined base temperature and the outside air temperatures"*. This means that, if we see it mathematically, the positive area between the outdoor temperature curve and the threshold temperature line is taken into account (Day et al., 2003). In this case there can occur a difference in CDD due the quality of weather data used to calculate it. Hourly data is more accurate than daily data. The way of calculating is the same but instead of subtracting the outside daily temperature by the threshold temperature it is done per hour. by averaging the sum of the positive hourly differences, which are called cooling degree hours, over the day can create the daily CDD (Day et al., 2003). In the end the most simple technique is just to calculate the CDD over the mean daily temperature. In this way, it is less precise than the mean cooling degree hours method, because it only takes into account average daily temperatures which are higher than the reference temperature (Day et al., 2003).

2.4.4 Heating degree days

According to Meng and Mourshed (2017): *"Heating degree days is a versatile climatic indicator that encapsulates the severity and duration of cold weather in one index, enabling the weather-related analysis of the consumption of fuels such as natural gas, coal and electricity. The versatility of HDD is due to its simplicity in reducing the dimensions required to characterise a given weather"*. This is in contrast with the cooling degree days. As stated before, the threshold or in other words the balance point is the ambient air temperature that doesn't require indoor heating. When the ambient air temperature is below this threshold a building needs to heat to maintain a desirable indoor environmental condition (Meng & Mourshed, 2017). As stated before, the daily energy calculations based on the heating degree days are simpler than the hourly methods and known as an effective way to manage the energy demand. Methods based on degree days are used to estimate building energy consumption and to determine energy performance rating (Meng & Mourshed, 2017).

Building energy consumption is weather dependent, according to Lazos et al. (2014). According to Shin and Do (2016), the cooling load is also very much dependent on the humidity of the air. The heating degree days are well known as the indicator for cold weather and the duration of it. According to Lazos et al. (2014), the methods used for the heating degree days only account for the effects of the temperature and underestimate the other climatic factors such as rain and humidity, wind speed and solar radiation. Lazos et al. (2014) suggest that besides the ambient temperature also the other climatic factors need to take into account because buildings are getting also exposed to these factors. Although much research has been conducted on degree days and the determination of base temperature, a study such as that by Lindelöf (2017) focused on one parameter (temperature). Few studies have incorporated more than one parameter into the approach. A fortunate event is that the KNMI proposes correction coefficients so additional weather variables can be considered along with ambient temperature.

2.5 Conclusions

As stated in the chapter: introduction the main focus of the literature study was threefold. In all the sections of literature study three theme's came across, namely: variation in heating energy on a local level, the climatic variables on heating energy and the concept of degree days in interaction with gas consumption. The introduction stated these theme's into sub-questions namely:

1. Are there variations in heating energy use on a local geographical level?
2. What are the effect of climatic variables on heating energy?
3. How can degree days helps with comparing gas consumption in different years and locations?

To start with the first topic: the variations in heating energy use on a local geographical level. The second section of this chapter showed that the energy use of the building stock plays a big part in the overall energy use. Also multiple studies showed that the differences in energy consumption could occur due to the characteristics of the building, the socio-economic situation of the residents of the building and that the surrounding area of the building could play a part. As shown in figure 2.1 there has been done some study towards energy consumption within dwellings but the specific research for heating energy on a postal code 6 level is not yet carried out.

The second section showed the importance of temperature, solar radiation, humidity and wind on on the heating energy use. The study of UHIs showed that local variations in temperature. The specific topic of climatic variables on heating energy are less often carried out. Especially the studies on the local climatic influences interacting with heating energy are scarce. Therefore studies toward the heating energy use in different postal code 6 levels are needed. The further research will focus on the climatic influences and the same factors that play a part as in UHIs, namely: rain, wind, irradiation and sustainability. This research will be conducted on postal code 6 areas.

The last part of this chapter focused on the concept of degree days. A lot of research has been carried out towards the cooling and heating degree days. The section about the correction of gas consumption with degree days showed that this is the most suitable method to compare a region in different years with each other. Also the weighted degree days where mentioned. However the weights seems to be very rough and have no geographical dependence, while they are expected to vary a lot depending on location. The further research will investigate if the weighted degree days can explain differences in on a local level in gas use.

The research questions above are partly explained by the theory. Nevertheless there are still some research gaps on this topic to fully answer the main question. Are there local differences of degree days within cities? How does the heating degree days help to understand the local heating energy use? Are the local climatic factors accounted for in the weighted degree days? The following section will redefine the sub-questions that will be investigated in the remaining research.

2.6 Research questions

The literature review discussed which aspects have already been covered in the scientific literature and where there are still research gaps. The research question gave rise to various sub-questions, which need to be answered in order to answer the main question. The main research question is the following:

- **Are there outdoor climatic differences at postal code 6 level within large cities and what are the effects of these variations on heating energy consumption?**

The research question gave rise to the following five sub-questions:

1. Are there local differences within a city between the degree days at postal code level 6?
2. Does a correction with local HDD helps to understand the local heating energy use?
3. Is the local influence of wind, solar irradiation and rainfall well accounted for in the weighted HDD?
4. Are there other factors like the ones captured by the energy label explaining differences?
5. How will the local weather- and gas use data be obtained?

The first sub-question will explain the difference between the degree days on the local scale of postal code 6. Stated in the literature study the local climate can vary on a small scale through all kinds of different factors, one of the examples is the phenomena UHIs. This research will further investigate if and why there are local differences in degree days. Besides the degree days per postal code 6 level also the gas consumption on this scale will be investigated. The second sub-question will try to answer the question who the local (heating) degree days can explain the differences in heating energy use. Also stated in the literature study are the climatic factors. The weighted degree days are created to understand the effects of these climatic factors and been accounted for in the degree day method. Is the concept of WDD also applying differences in the local climate or is this method not suitable for such a small scale? The following chapters tries to explain this. The fourth sub-question will try to explain if other non-climatic factors also play a part in explaining the differences in gas consumption. Both with the climatic and non-climatic factors the effects of variations on heating energy consumption can be explained. The last sub-question will be carried out in the methodology chapter. This chapter will explain how the data needed for the other sub-questions is obtained and how this data will be prepared.

Chapter 3

Methodology

The methodology chapter is divided into four parts:

- The case study region: The Hague
- The data types and sources
- Detailed flowchart of the research steps
- Processing of the data

The first part of this chapter discusses the case study region of The Hague. Why was this municipality chosen? Second, the useful data are explained. Also the validation of the source is key for the thesis. The use of reputable institutes was necessary to validate the more unreliable data sources. For example, amateur weather stations need to be validated by professional weather measurements. The combination of both datasets is useful, because the amateur data providing local data that has the quantity but the professional data provide quality. In the second part the different software packages are discussed: this part will explain the methods and techniques that will be used to obtain these results. The third part of this chapter visualizes how the results will be obtained using a flowchart. Subsequently, the preparation of the data is discussed, in other words how the data will be pre-processed to fit this specific research study. Validation, extraction and the combination of different datasets will be discussed and the use of the software will be explained.

3.1 Methodological approach and data

To answer the research question and the sub-questions, it is key to work with accurate data. For this research study, open data sources were used. This has the benefit that it is free and that the thesis can be reproduced by anyone else. The main question addresses the differences in postal code 6 areas within large cities. This research will focus on one city to make it comprehensible. It needs to be a city with some characteristics, namely: a combination of green areas and densely built areas to illustrate the difference in the local climate. Also the different climatic factors such as wind, irradiation and rainfall need to be shown properly in this case city. In this research the city of The Hague will be used. It has both dense areas (city centre) and green areas (*Haagse Bos*). It is also close to the sea where the wind plays an important role but also has some areas more inland such as *Ypenburg*. Besides the geographical and climatic factors it has a well constructed open database facilitated by the municipality. The case region will be further explained in the following chapter.

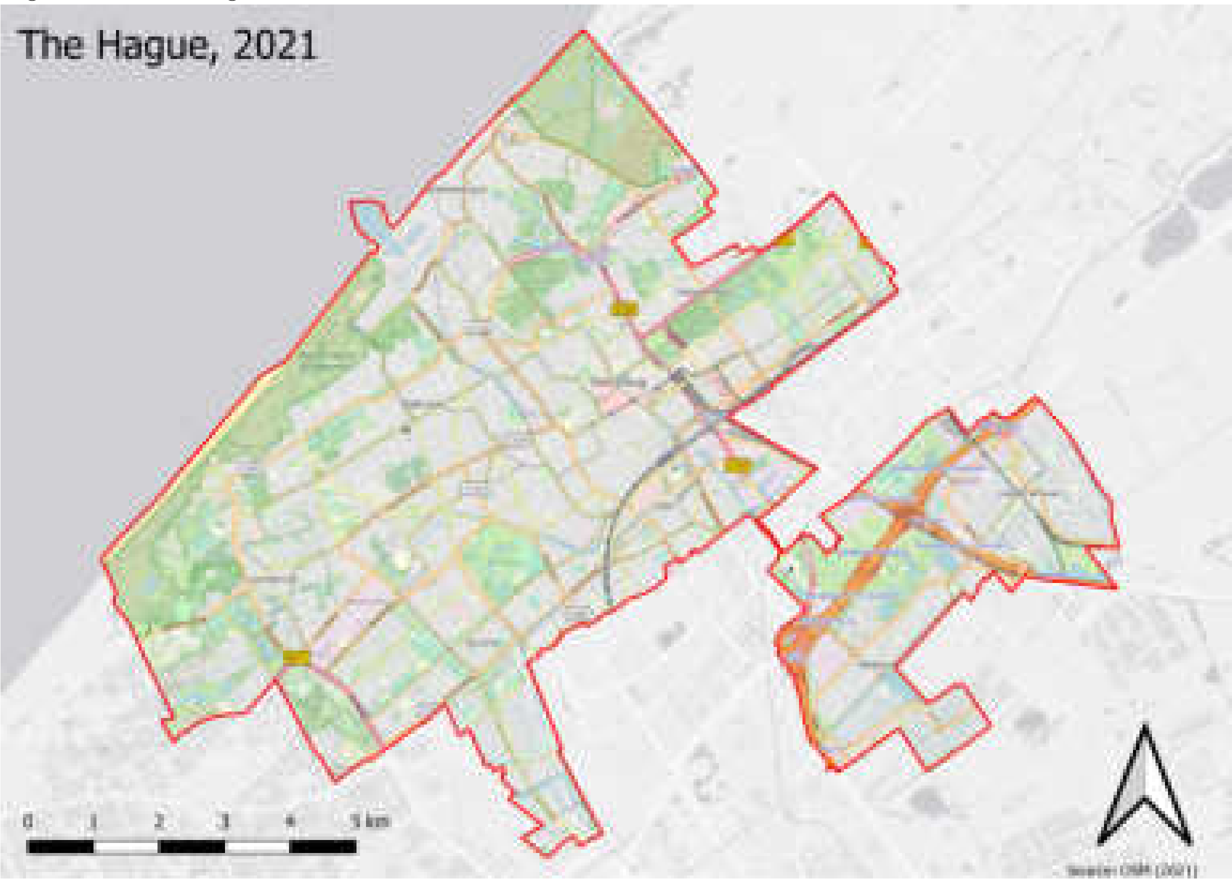
In the previous chapter the sub-questions are discussed. The following enumeration will explain how these sub-questions will be researched in the remaining chapters of this thesis and what kind of data is needed to visualize and analyse these topics.

1. Sub-question 1: For this sub-question the use of existing local historical weather data in the city of The Hague will be used to map the local differences in degree days. Also these differences will be statically tested.
2. Sub-question 2: the data for this sub-question will make use of postal code level 6 heating energy data (gas). This is publicly available data and will later on be combination with the degree day data from sub-question 1 to analyze and compare raw local heating energy on a local level. This research will be conducted over 2 different years and to also compare an statistically analyse with HDD corrected energy data.
3. Sub-question 3: With the available climatic data the averages for wind, irradiation and rain will be researched if the weighted degree days method is suitable to explain the local differences. This will be statistically analysed and tries to explain what their contribution to local heating energy use is.
4. Sub-question 4: Non-climatic data such as energy labels will be investigated and mapped to explore if these non-climatic factors also contribute to the difference in gas use on a local level.
5. Sub-question 5: will explain what kind of data is needed to complete the chapter: results. The methodology chapter will explain how this data is obtained and prepared. The data that is not been found and the limitations of the data will be discussed in the last chapter: discussion and conclusion.

3.2 The case study region: The Hague

The Hague is an interesting city to use as a research region. There are a few reasons this municipality was chosen to execute the research. Figure 3.1 shows the case study region of The Hague. First of all, this municipality has its spatial orientation at the coast. The literature study showed that the wind factor is a climatic influence that may contribute to differences in gas use. A coastal municipality experiences in general more influence from the wind than a more inland region. With a higher wind speed from time to time, it may be possible that this factor plays a more significant part and it is therefore important to include. Coastal placed regions are also known for a more moderate climate than the inland regions. The second reason is that The Hague is the third biggest city in the Netherlands, with a high diversity in building types. The Hague has a centre with high-rise buildings, but also features green areas, both on the coast and in the city. These contradictions can lead to interesting results. The third reason for choosing The Hague is the extended open source database provided by the CBS in cooperation with the municipality of The Hague. Without data, this research could not lead to good results.The extensive database of The Hague provided sufficient data on buidlings and energy use. For this research local meteorological data is also needed at a much higher granularity than KNMI data. This research study made use of an amateur weather database that only has data about Delft and The Hague. Local weather data are crucial for determining the local climatic differences.

Figure 3.1: The Hague



3.3 The data

The data of this research study can be divided into four groups:

- Geographical PC6 levels descriptions;
- Weather data;
- Degree days; and
- Gas consumption.

The table 3.1 gives an overview of the most important data that will be discussed in the following sections.

Table 3.1: The data composition

Data source	**Data**	**Format**	**Temporal resolution**	**Spatial resolution**
PDOK	Postal code 6	Shapefile	Yearly	Postal code 6
KNMI	Weather data	CSV	Hourly/Daily/Yearly	48 verified weather stations in NL
Netatmo	Weather data	JSON	Daily	Various weather stations in DH
Mindergas	Degree days	CSV	Daily	48 verified weather stations in NL
CBS	Gas	CSV	Yearly	Postal code 6
Stedin	Gas	CSV	Yearly	Postal code 6
Open data The Hague	Energy labels	CSV	Yearly	Values per neighborhood

The first data type that will be discussed is the postal code 6 levels obtained from the website *PDOK.nl* (Publieke Dienstverlening Op de Kaart, 2020). The research focuses on the local climate. Besides the PC6 there are also other boundary types such as neighbourhood level, district level, PC5 and PC4. The building type, the green areas and the density of buidlings in the street can be diverse on a small area. To measure the changes in the local climate that these diversion of streets could have the PC6 levels are the most suitable way to map these changes.

The second kind of data required for investigating the local climate is weather data. To do research on such a small area local weather data is required to map the differences on PC6 level. This weather data will be obtained from the dataset of the website Netatmo.com (Netatmo, 2020). This dataset contains different types of weather stations. Some are from a high quality and some are not. To verify and select the high quality stations the data from Netatmo will be compared with data from the KNMI (KNMI, 2020). The main goal of processing weather data is to obtain the degree day information at high spatial resolution. As said in the chapter: literature review, the degree days can be calculated from the daily temperature. To verify these calculations the data from the website *Mindergas.nl* (Mindergas, 2020), will be compared with the calculated temperature from the KNMI and the data from Netatmo.

The third datatype that will be discussed is the gas consumption per households. This data can be retrieved from both the CBS (CBS, 2020) and the network operator *Stedin* (Stedin, 2020). Both these datasets contain yearly average gas consumption per PC6 level. Because this data is only available in the temporal resolution of a year all the other data will be transformed to yearly data so they can be compared and be merged.

Finally, as said in the chapter: literature review, also the energy labels will be used in the further research to measure geographical variations in sustainability in The Hague. This data will come from the database from the municipality The Hague (Gemeente Den Haag, 2020) and the *energiematrix* from the CBS (CBS, 2018).

3.4 Geographical data: PC6 levels and application to The Hague

In the collection of open data, multiple spatial formats are available, such as municipalities, districts, neighbourhoods, and PC4, PC5 and PC6 areas. The municipality of The Hague delivers all these formats. PC6 levels will be used as much as possible. Almost all of the data are categorized in this format. The most important reason for this is that the gas use data, reflecting the heating energy use, are provided in terms of the PC6 level. The PC6 level is the division of The Hague with the smallest areas and therefore contains the most information. The PC6 levels are the smallest postal code areas available in the Netherlands. Because of the large scale of these areas other information that is assigned to a certain PC6 level can be privacy-sensitive. In the Netherlands the CBS will not address information of any type (this also includes gas consumption) to PC6 areas with less than 5 residences (CBS, 2020). Nevertheless, PC6 will be used instead of, for example, PC4. The reason is simply that this is the most accurate dataset spatially. The division of the PC areas is shown in table 3.2.

Table 3.2: The postal code distribution

Postal code	**Example numeration**	**Total areas in The Hague**
PC6	1234 AB	13.458
PC5	1234 A	844
PC4	1234	88

The benefit of using PC4 areas instead of the PC6 areas is that the data doesn't have any missing values. If we use the PC4 level, blank spaces with 0 values will disappear from the data because we the PC6 areas, in which areas with less than 5 buildings are not taken into account for privacy reasons, are merged into the bigger PC4 areas. The disadvantage of this method is that the calculation will be for the average of the larger PC4 areas. Because the PC6 gas consumption values are already averages and the 0 values will be included in the average for the PC4 level, this would distort the representation of the reality.

The different PC levels can be found on the website of *Publieke Dienstverlening Op de Kaart (2020)*. This is the open source database of geographical information from and for Dutch government institutes. Figure 3.2 illustrates a section of The Hague Centre region. The most important part of this figure is that it shows the proportions the PC areas have to each other. The PC4 area, displayed with the red line, covers a larger part of The Hague. Within the PC4 are the area is divided in the smaller areas PC5 (orange-lines) and the PC6 regions (green-lines). In order to keep figure 3.2 clear, not all the PC4, PC5 and PC6 levels of The Hague are visualized. The map is only to illustrate the different types of PC6 areas and the distribution of it. Figure 3.3 shows all the PC6 areas in the Hague.

Figure 3.2: PC areas distribution in The Hague centre (Publieke Dienstverlening Op de Kaart, 2020)

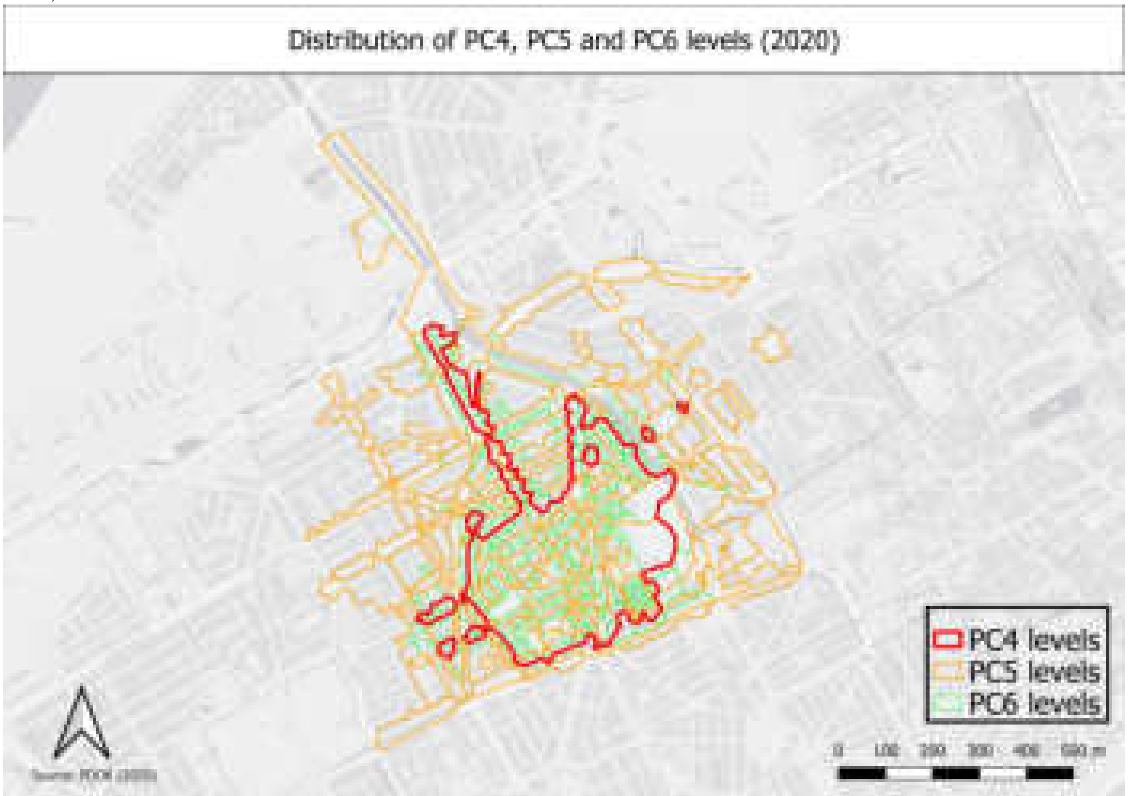
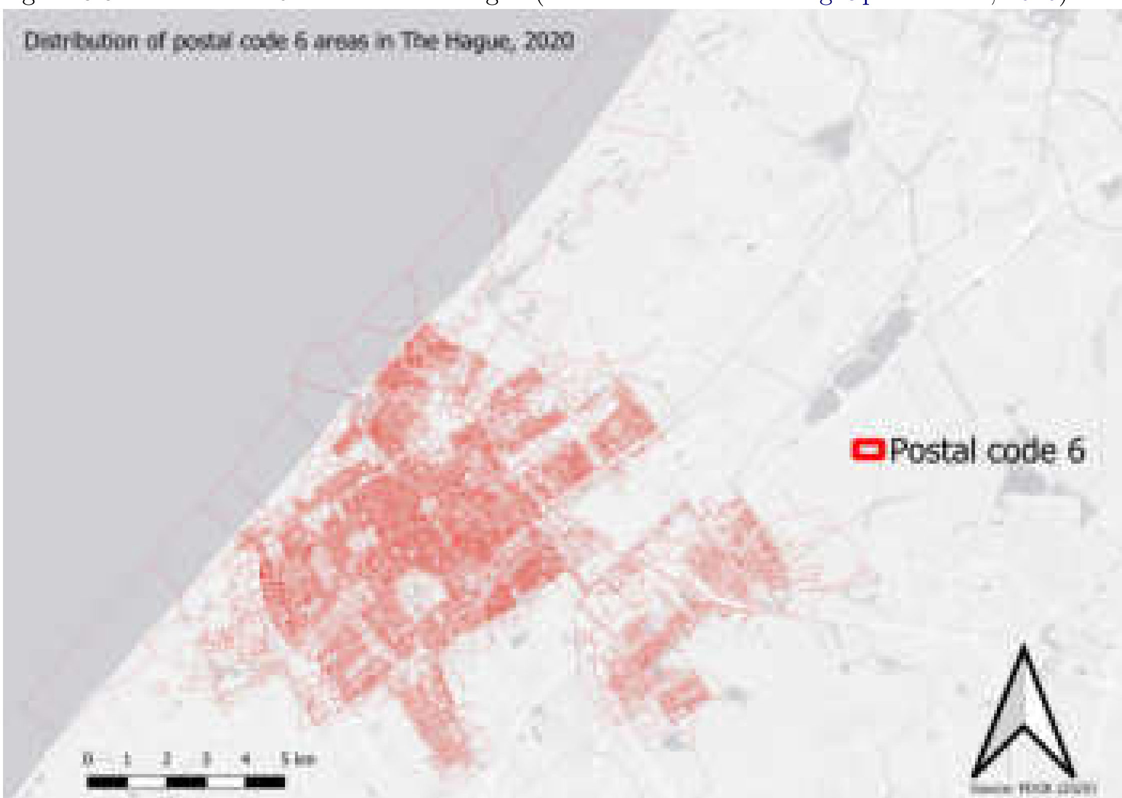


Figure 3.3: Postal code 6 areas in The Hague (Publieke Dienstverlening Op de Kaart, 2020)



3.5 Weather data, including degree days

3.5.1 The KNMI weather stations

The KNMI has 48 verified weather stations located throughout the Netherlands, illustrated in figure 3.4. The figure indicates that most of the urban and rural areas on the west side of the Netherlands are covered. This thesis will focus on the municipality of The Hague and in Figure 3.4 there is a zoomed image that illustrates that around The Hague the coverage of KNMI stations is above average. Unfortunately, there is no KNMI station within the borders of The Hague. As we look at the zoomed image it seems that four stations have the potential to cover the municipality of The Hague, namely:

- Hoek van Holland
- Rotterdam
- Voorschoten
- Valkenburg

Unfortunately only station Rotterdam and Hoek van Holland provide the necessary information: stations Voorschoten and Valkenburg are not accessible as open data sources for the public in the KNMI historical database.

The benefit of KNMI data are the hourly measurements of these stations. Also these stations measure extensive data, including factors such as wind direction, wind speed, temperature, average radiation, rainfall and humidity. This information has been collected and saved in a historical database. For this thesis, the yearly averages will be used. As stated in the literature study, heating degree days can be calculated per year .

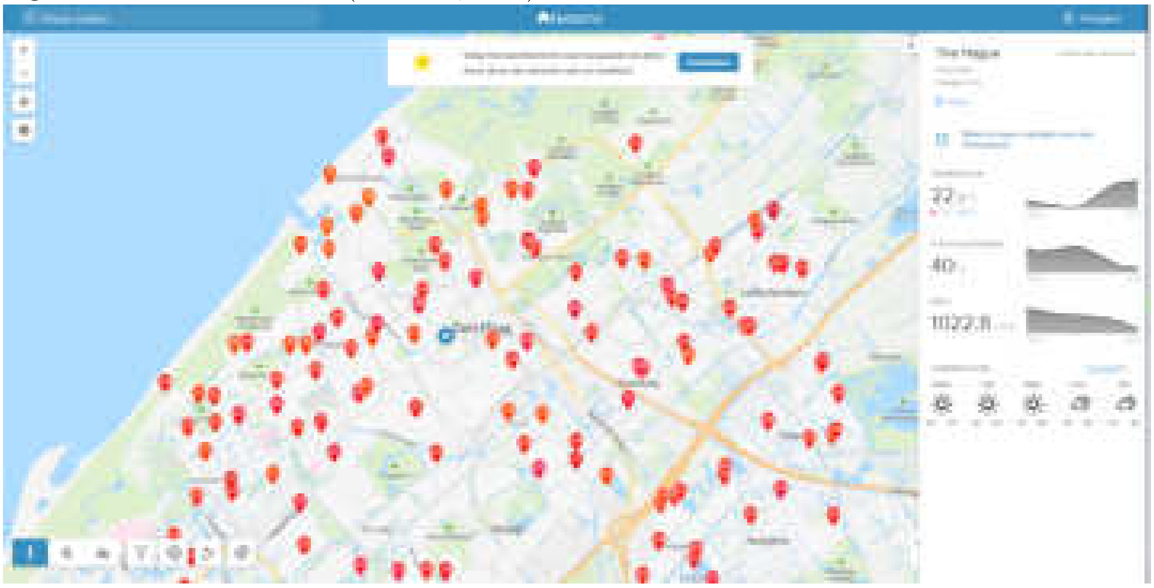
As said before, since the official meteorological stations operated by the KNMI are not located in The Hague itself, measurements from weather hobbyists will be included. The scope of the thesis is at the PC6 level and multiple measurement stations need to be used to conclude if there are differences within the municipality. These will be obtained from the Netatmo website, where hobbyists can upload their measurements in an automated way.

Figure 3.4: Weather stations KNMI (KNMI, 2020)

3.5.2 Netatmo amateur weather stations

Netatmo is a company that sells multiple types of sensors, such as smart doorbells, indoor climate control modules and of course outdoor weather stations. Although Netatmo is a commercial company, where the first priority is to sell these sensors, they also provide a map (figure 3.5) where weather stations are displayed. It is possible to check the local weather via this interactive map, but it is also possible to gather this information.

Figure 3.5: Weather stations (Netatmo, 2020)



There are two ways to collect and store data from the Netatmo database. The first is to ask permission to collect data from an individual sensor through the owner; all these sensors are displayed on the Netatmo map. The second is to collect all the measurements for a spatial area through an application programming interface (API). For this research study, the second option is used, as it is the easiest. The advantage of this method is that all the data can be collected at once. The disadvantage is that one can only collect the latest measurements. Every 15 minutes, the real-time weather data are refreshed and the older measurements disappear. It is possible that the historical weather data have been stored by the owner of a certain sensor, but it is nearly impossible to ask permission to all of the owners to collect their data, station by station.

Fortunately, Dr. Ir. Ken Arroyo Ohori (MSc), a postdoctoral researcher at the 3D geoinformation group of the Delft University, discovered the same limitations of the API years ago. Since 2015, he has been been authorized to collect the 15-minute timeframe data for the spatial regions of Delft and The Hague. He has created his own dataset. From 2017 onward, most days are covered, and from 2018 all 15-minute timeframe data have been collected. This thesis makes use of Ohori's dataset.

The dataset includes the ID for every individual sensor, including the longitude, latitude and altitude, so that the sensor can be geographically placed. Almost all sensors collect temperature and pressure data. Some sensors also collect rain and humidity data. For the further use of this dataset, the most important aspects are the X and Y coordinates and the temperature to calculate the degree days per weather station. Figure 3.6 shows an example of the raw Netatmo data (after converting from txt to CSV). The figure illustrate the first ten stations in the dataset from 00:07 a.m. January 1, 2018. Notice that the day and time are not present in the attribute table, these are mentioned in every individual filename.

Figure 3.6: Netatmo raw data (Netatmo, 2020)

id	longitude	latitude	altitude	temperature	humidity	rain_60min	rain_24h	rain_live	pressure
70:ee:50:04:6f:1c	52.143	4.388164	2	8.5	85	0	8.888	0	988.2
70:ee:50:2b:58:b0	52.150682	4.4033974	3	8.3	92	0	19.594	0	1000.8
70:ee:50:00:5a:16	52.150683	4.397504	1	8.4	83	0	10.908	0	996.4
70:ee:50:01:01:b8	52.205914	4.404077	4	7.7	96	0.404	8.282	0	995.8
70:ee:50:03:a4:30	52.147131	4.4204433	1	8.3	88	0	0	0	992.7
70:ee:50:05:58:94	52.150392	4.4066107	1	8.4	87	0.101	5.656	0	992.9
70:ee:50:02:3f:66	52.14868814	4.407957928	3	8.8	86	0	6.06	0	996.7
70:ee:50:02:f5:8c	52.1957179	4.4206633	4			0	9.999	0	996.7
70:ee:50:03:f9:a0	52.211199	4.416348	2	8.6	88				993.6

[1]

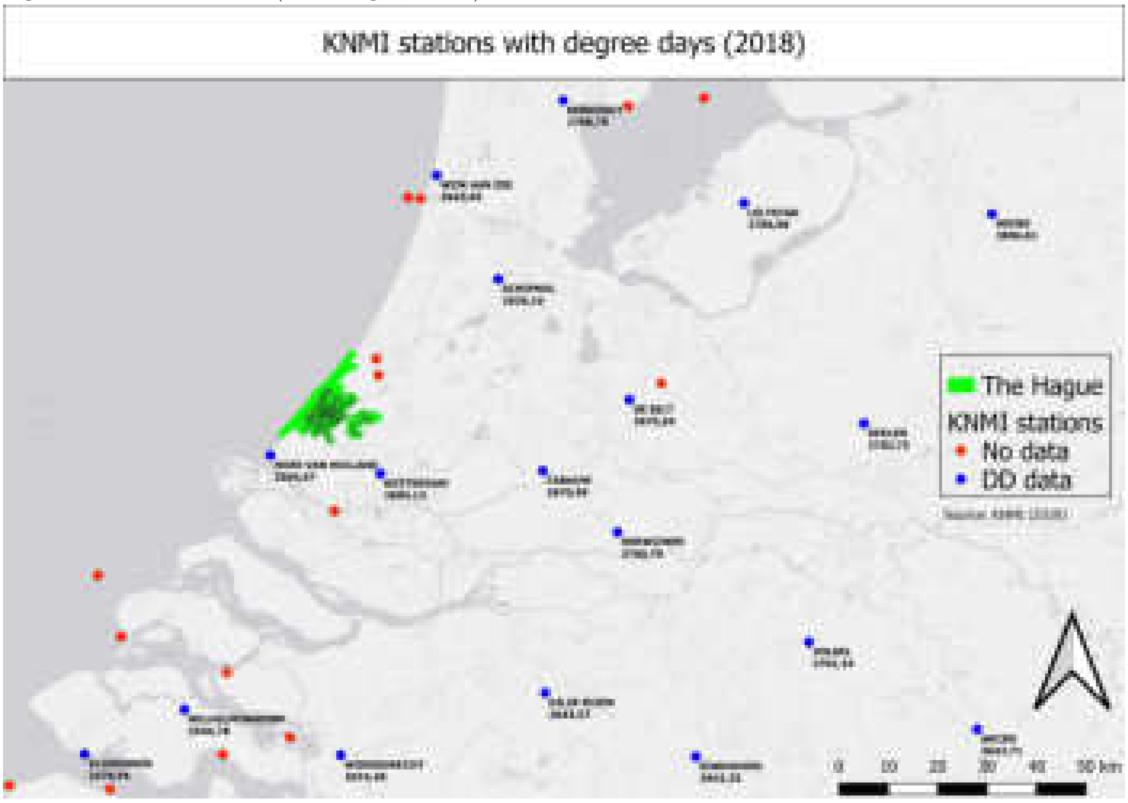
[1] The Netatmo data is collected by Dr. Ir. Ken Arroyo Ohori (MSc) and is used in this thesis with his permission.

3.5.3 Degree days

In this research study, the degree days correction of the gas use will be calculated using the Netatmo temperature data for PC6 levels. This degree day correction will be excecuted as explained in section 2.4.2. namely, gas consumption year (x) / total degree days year (x) compared to gas consumption year (y) / total degree days year (y). Because of the possible presence of low quality data and to prevent miscalculations, a reference dataset was necessary. Besides weather measurements, the KNMI also calculates and publish the degree days per station per year.

Figure 3.7 illustrate the KNMI stations in the west of the Netherlands. Said in the section about the KNMI stations that the KNMI features 48 verified weather stations, only 34 are accessible when it comes to open data. The blue stations are provided with the degree day data from 2018 from the website *mindergas.nl* (Mindergas, 2020) and the red stations are not accessible. Mindergas (2020) has a degree day calculator that calculates the degree days for a given time and date of your choice. In this thesis the calculator is used to calculate the degree days per year. This calculator can compute back till 2006.

Figure 3.7: WDD 2018 (Mindergas, 2020)



3.6 Energy data

The CBS collects gas and electricity data per PC6. These data represent the averages for a whole year for the PC6 areas, to protect the privacy of individual users. Both electricity and gas supply are divided into the average for houses and businesses. This thesis focuses on household gas consumption, so the gas consumption for business parks and electricity data were filtered.

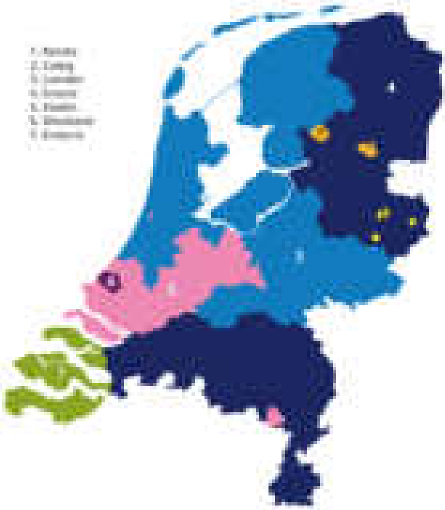
Besides the CBS, individual network operators also provide open source gas use data. There are two reasons to use CBS data instead of the data from network operators. The first reason is that there are multiple network operators. Figure 3.8 presents the distribution of these operators throughout the Netherlands. This thesis focuses on the municipality of The Hague, where Stedin is the network operator. If in future the someone want to expand this research to other municipalities or regions, it is possible to extract data from the operators. The CBS has already collected these different datasets and combined them. The other reason there is no use of the Stedin dataset is because of the format of the data. This thesis makes use of PC6 levels. The Stedin dataset uses the format from PC6-1 to PC6-4. This means that PC6-2 and PC6-3 are included in this range. The CBS provides data for all the individual PC6 levels, so it will be easy to use in further research. Table 3.3 illustrates the differences in notation for both CBS and Stedin.

Table 3.3: Difference in data notation CBS and Stedin

Data source	**Data notation (PC6)**
CBS	1234 AA, 1234 AB, 1234 AC, 1234 AD
Stedin	1234 AA t/m 1234 AD

The disadvantage of the CBS method is that missing values are included in the dataset. As stated previously, the CBS has a privacy regulation, stipulating that no PC6 level with fewer than five households will be included in the dataset, creating missing data in the data for this research. The CBS makes use of the datasets from individual operators such as Stedin. One of the reasons that Stedin groups the data together could be to avoid these missing data fields.

Figure 3.8: Network operators in the Netherlands (*energieleveranciers.nl*, n.d.)



3.7 Flowchart

Figure 3.9 illustrates the flowchart of the process to the data modification to create the results. The blue boxes represent the data that was earlier mentioned in the section: the data. The diamonds in the upper-right corner gives the legend of the different type of software used to process the data. All three diamonds are provided with a colour, this colour can be found in the boxes and ovals in the flowchart. These colours are illustrate which software is required to complete the task. The boxes in the flowchart represent a product or a task that needs to be carried out. The ovals represent a process, mostly to complete a task or a product. The next paragraphs will describe per data group what the tasks and processes are.

The first data type is the degree days from *mindergas*. As said in the section: degree days, this data will be used to provide yearly degree days for 2018 and 2019 and validate the data from KNMI stations. Together with the weather data it will deliver the degree days per postal code 6 area for The Hague.

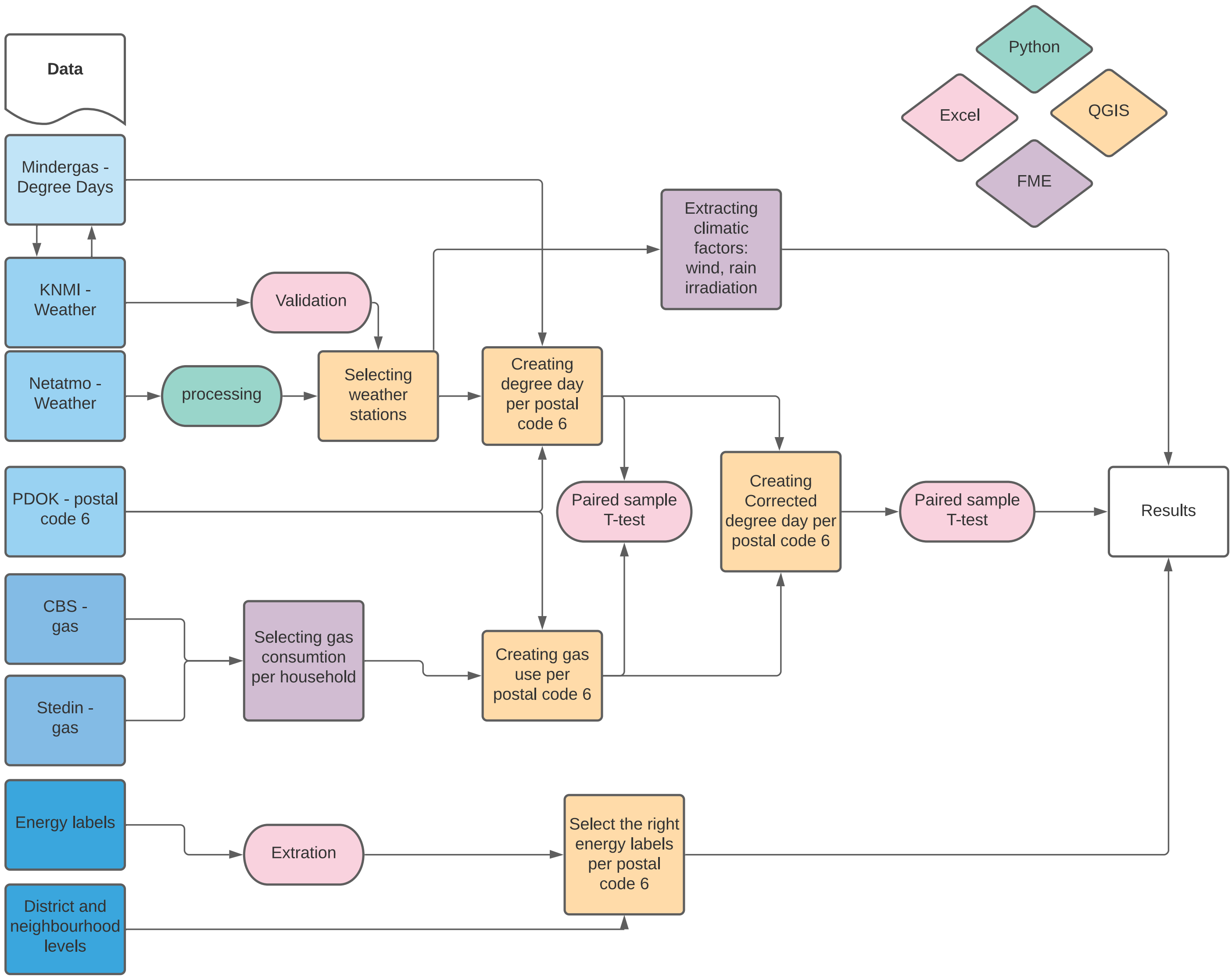
Before the weather data can be used to calculate the degree days per postal code 6 they need to be processed. The Netatmo data need to be extracted and transformed to yearly data before it is useful, this will be done with Python. After this process the KNMI stations will be used to validate the Netatmo data and filter the outliers. After validation the useful stations can be selected and together with the postal code 6 areas from (Publieke Dienstverlening Op de Kaart, 2020) the degree days per PC6 area can be calculated, this will be done with QGIS.

CBS and Stedin both provide gas consumption per household. With FME both datasets can be inspected and the gas consumption of The Hague will be extracted from the dataset of choice, in this case the CBS dataset. Besides the creation of the degree days per area the PC6 data will also be used for the creation for the gas consumption per PC6 area. This will be done the same way, via QGIS. The energy labels are stored in two different datasets and will be extracted with Excel and combined with district and neighbourhood levels. In the dataset for gas use the unfortunate phenomena occurs that the spatial data is not stored in the earlier mentioned PC6, PC5 of PC4 levels but in district and neighbourhood levels.

After creating both the degree days and gas use per PC6 level this data will be combined to calculate the gas use corrected with degree days. First both created databases need to be statistical tested if the data from 2018 and 2019 significantly differ from each other. When both datasets are merged together adjusted to calculate the gas use corrected with degree days the created data will be statistically tested if also this data significantly differs from each other. All these statistical tests will be done via Excel.

The energy labels and the gas use corrected with degree days per PC6 will be used together with the extracted climatic factors to come test their influence on the gas use at PC6 level. These climatic factors will be extracted via FME. Not all climatic factors are available in the weather databases, With FME the data can be inspected and the useful data can be used.

Figure 3.9: Flowchart



3.8 Processing of the data

3.8.1 The Netatmo data

The data from Netatmo (2020) collected by Dr. Ir. Ohori are stored in the format of txt-files with a timeframe of 15 minutes. This means that there are 96 files per day, resulting in 35,040 files per year. Each file includes approximately 875 individual stations spread over The Hague and Delft, resulting in over 30 million observations per year. Because this thesis focuses on daily and yearly observations of the individual stations, the Netatmo data needed to be processed. The first step was to convert the txt-files into CSV files and to merge the 96 files per day into one file per day. To accomplish this, a Python script was created, which can be found in Appendix 1. The main script was created as follows:

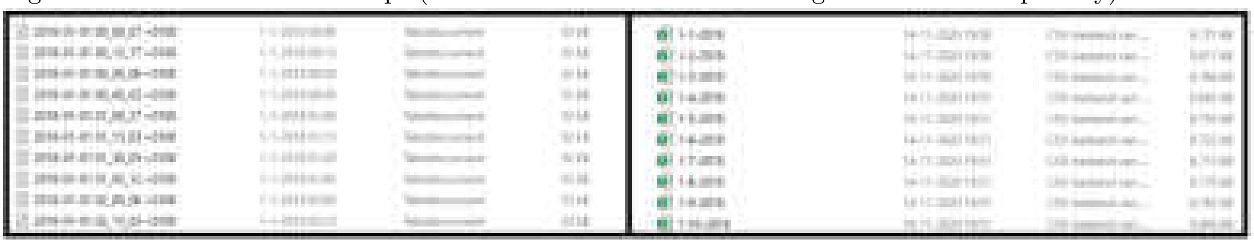
The first step was to generate a CSV day-based output by importing the following modules:

- os: the os module in Python provides functions for interacting with the operating system
- CSV: a comma-separated values (CSV) file is a type of plain text file that uses specific structuring to arrange tabular data
- pandas: pandas is a software library written for the Python programming language for data manipulation and analysis
- datetime: the datetime module supplies classes for manipulating dates and times
- calendar: an inbuilt module calendar which handles operations related to calendar
- sys: the Python sys module provides functions and variables which are used to manipulate different parts of the Python runtime environment

After importing the necessary modules, the class was created to store the weather station data. The same attributes from the txt-files are used in the new class. After creating the class, the csvWriter was used to fill the data frame. The years, months and days were looped and the data were written to CSV. Since the data were collected until 23-10-2020, the loop captured all these data.

When all the data were converted to CSV files, input and output files were selected. Also, there is an initiation of the array of objects for each year to save memory. The last step was to create the CSV files for one day instead of for every 15 minutes. These CSV files can be found in the last part of Appendix 1 . Files with errors were printed to control and keep track of failing files. The results of this script are illustrated in Figure 4.1. The first box indicates the first ten records of 2018 and the second box the converted CSV files, neatly aligned in days. The aggregation of the data is illustrated in figure 3.11. The temperature, humidity and rain-60min where averaged for day (x) for station (y). The rain-24h is the total rain for day (x) station (y).

Figure 3.10: Results of main script (left the 15 minute txt-files and right the CSV files per day)



The second script was created calculate the degree days for every day and for every station. The analysed script can be found in Appendix 2. The step-by-step creation of this script for degree days per station per day was accomplished as follows. First the following modules were imported: os, CSV, pandas, datetime and calendar. Also the input and output directories were created. As in the main script, a class was created to store the weather station data. The csvWriter again created new CSVs in the output file. Again, the definition was created for the temperature tipping point. In the literature study, the tipping point for the degree days was discussed, namely 18 degrees Celsius. According to this definition, the script can make a distinction between the values under and above this boundary.

As in the main script, the loop to the selected years was created, as well as the initiation of the arrays for objects for each year, to save memory. Again, the new CSV files are like the CSV files from the main script, but the degree day is added as the last column. With the lambda function in the last part of the script, the degree day column will be populated by the difference in temperature from the boundary of 18°C, if the temperature is below this value. So, for example, if station x on 1-1-2018 had a temperature value of 6°C, the lambda function will calculate 18-6 = 12 degree days, just like the examples in the theory chapter. If on 1-6-2018 the temperature was 21°C, the function will calculate a value of 0. The first box in Figure 3.11 presents the input data from Netatmo (2020), while the second box presents the created output from the analysed script.

Using the CSV files for every day of the year, another CSV file was created. Using the pivot table function and the sum function in Excel, the total degree days per station for a whole year were created. Together with the longitude and the latitude, a shapefile for the stations was created, ready for use for the geographical analysis. According to the website: *weathermap.netatmo.com*, where the individual stations are visualized, the temperature is in degrees Celsius, the humidity is a percentage from 0 to 100, the rain is in millimeters (mm) and the pressure is in millibar (mBar).

Figure 3.11: Results of analyse script

```
id,longitude,latitude,altitude,temperature,humidity,rain_60min,rain_24h,rain_live,pressure
70:ee:50:04:6f:1c,52.143,4.388164,2,8.5,85,0,8.888,0,988.2
70:ee:50:2b:58:b0,52.150882,4.4033974,3,8.3,92,0,19.594,0,1000.8
70:ee:50:00:5a:16,52.150685,4.397504,1,8.4,83,0,10.908,0,996.4
70:ee:50:01:01:b8,52.205914,4.404077,4,7.7,96,0.404,8.282,0,995.8
70:ee:50:03:ad:30,52.147131,4.4204433,1,8.3,88,0,0,0,992.7
```

id	longitude	latitude	altitude	temperat	humidity	rain_60mi	rain_24h	rain_live	pressure	datetime	degree_days
70:ee:50:04:6f:1c	52.143	4.388164	2	8.5	85	0	8.888	0	988.2	1-1-2018 00:00	9.5
70:ee:50:2b:58:b0	52.15088	4.403397	3	8.3	92	0	19.594	0	1000.8	1-1-2018 00:00	9.7
70:ee:50:00:5a:16	52.15069	4.397504	1	8.4	83	0	10.908	0	996.4	1-1-2018 00:00	9.6
70:ee:50:01:01:b8	52.20591	4.404077	4	7.7	96	0.404	8.282	0	995.8	1-1-2018 00:00	10.3
70:ee:50:03:ad:30	52.14713	4.420443	1	8.3	88	0	0	0	992.7	1-1-2018 00:00	9.7

3.8.2 Paired samples *T-test*

According to StatisticsSolution (2021): *"The paired samples t-test, sometimes called the dependent samples t-test, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired samples t-test, each subject or entity is measured twice, resulting in pairs of observations"*. One of the common ways to using the paired samples t-test is for case-control studies and for repeated-measures designs (StatisticsSolution, 2021). The following example is given: if a company wants to indicate if there training programme is effective two measurements would be compared. The first one is the measurement of the performance of employees before the training and the second measurement is after completing the programme. To analyse the differences the paired samples t-test is used (StatisticsSolution, 2021).

For most of the statistical procedures there are two competing hypotheses, the first one is the null hypothesis and the second one is the alternative hypothesis, this is also the case for the paired samples t-test. When it occurs that the null hypothesis is right the true mean difference between the paired samples is zero (StatisticsSolution, 2021). In this model all the observable differences can be explained as random variation. In contrary to the alternative hypothesis. When it occurs that the alternative hypothesis is right the true mean difference between the paired samples is not equal to zero. According to StatisticsSolution (2021): *"The alternative hypothesis can take one of several forms, depending on the expected outcome. If the direction of the difference does not matter, a two-tailed hypothesis is used. Otherwise, an upper-tailed or lower-tailed hypothesis can be used to increase the power of the test. The null hypothesis remains the same for each type of alternative hypothesis"*. In this research the two-tailed paired sample t-test is used because the direction of the difference does not matter.

In this thesis four paired samples t-test will be performed. A test for the degree days for the years 2018 and 2019, a test for the difference between gas use per PC6 level in the years 2018 and 2019, the paired samples t-test for the gas use corrected with degree days for both the years 2018 and 2019 and the same test but with categorized gas use corrected with degree days. The aim for this tests is to verify if there are differences (or similarities) between this concepts in the two different years. This will help to conclude if if and how different climatic factors have influence on the local weather and tries to put these climatic factors in perspective.

3.8.3 The software

To obtain the results, three different software packages were used: FME, QGIS and Excel. FME was used for the data preparation, QGIS for the visualisation of the data and Excel for executing the statistical tests.

"FME (aka. Feature Manipulation Engine) has built-in support for hundreds of formats and applications as well as transformation tools, allowing users to build and automate custom integration workflows without having to code" (SafeSoftware, 2021). FME is a data integration platform that can easily transform different formats to others and where adjustments can be made to databases. Using this software, different types of data were converted to shapefiles and unnecessary data were removed. As all ready illustrated in figure 3.9, FME is used to selected certain attribute tables from a CSV or text file with X and Y coordinates. In this research the CBS data and the data converted from the Netatmo data. With options such *Aggregator* and *ESRI shapefile writer* workable shapefiles can be made for further research in QGIS.

QuantumGIS (2021) is a professional GIS application built onto free and open source software (FOSS) and designed to be FOSS itself. QGIS has the ability to analyse and visualize geographical data. Multiple functions are available and, through the plug-in options, Different functions can be downloaded for personal use. For this research QGIS is mainly used to analyse and visualize the maps in this thesis. With options such as *Neareast join*, *Buffers* and *clips* the data is been correctly processed to visualize in the QGIS *Layout view*. Most of the steps are discribed in figure 3.9 and the results are shown in chapter 4.

Excel is a well-known software package that almost every computer possesses. Excel is mostly used to create tables and display data. Another function of Excel is the ability to perform statistical tests. Besides the usual data processing, Excel was used to perform the statistical validation of the created data in FME and QGIS.

Chapter 4

Results

This chapter investigates the sub-questions. The chapter is divided into four parts. In the first place the research will investigate whether there is a difference in the degree days in The Hague for the years 2018 and 2019. The scope of the thesis is the PC6 level, so the results will compare the degree days for almost 13,000 areas in The Hague. The will results indicate if there is a difference in degree days between these two years. The second part of this chapter will investigate the difference in gas use. The third part will combine these two results into correction of gas use per degree day and the last part of this chapter will discuss the climatic and internal factors that contribute to these local differences.

As stated in the methodology chapter, the degree days, the household gas use and eventually the gas use corrected with degree days will be compared for 2018 and 2019. As said in the methodology chapter the climatic factors can put in perspective if there is a difference or similarity between these two years. Both 2018 and 2019 are fully present in the database from Dr. Ir. Ohori. May the weather conditions in both year a very different has this an influence on the local climate and the factors that are investigated in this thesis. The differences or similarities will be supported by statistical tests. Similarities or differences are visible in the maps, but the significance of the differences/similarities is not. With the statistical evidence, the interpretation of the data can be scientifically substantiated.

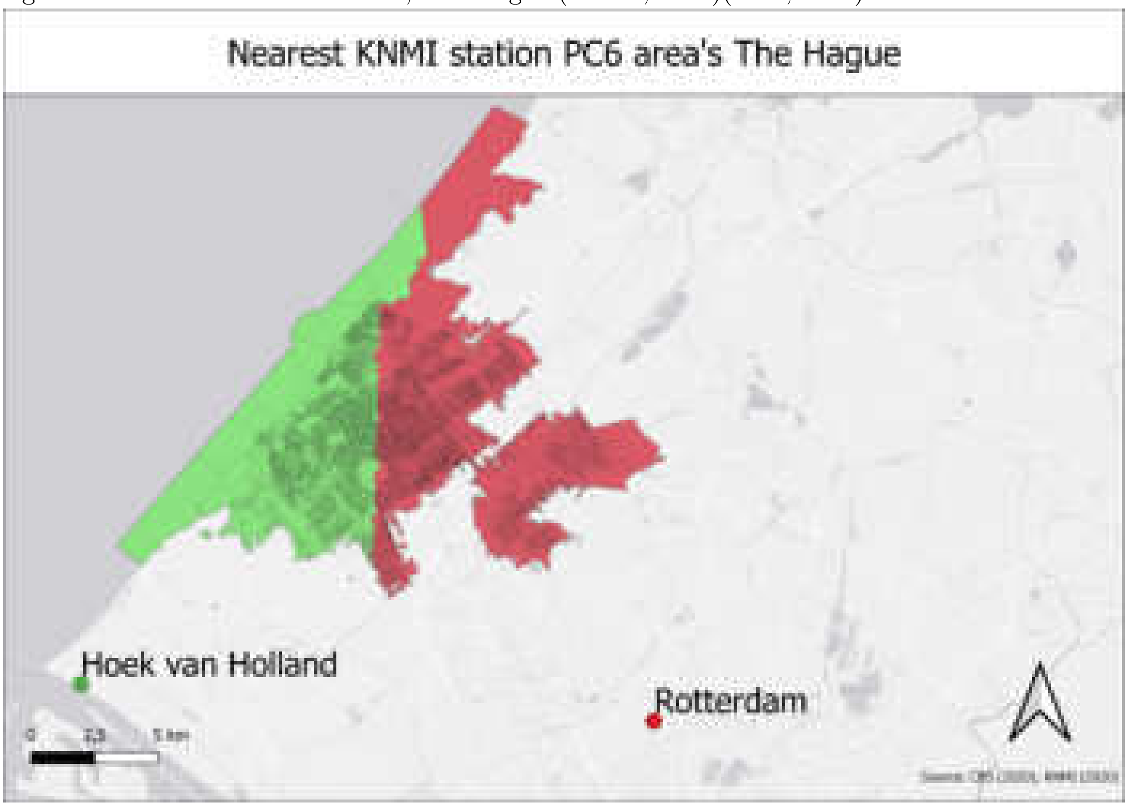
As stated in the literature study, five external factors will be investigated to see whether there is an influence on the gas consumption of households. If there is a difference in correction of gas consumption per DD between the two years, climatic factors besides temperature can also have an influence on the gas consumption. These factors can be climatic (wind, rain and irradiation), due the presence of greenery or there may be internal factors such as more sustainable houses. The last part of this chapter will investigate whether these factors are present and what their influence is.

4.1 KNMI and Netatmo stations

The first step to determine which Netatmo stations are of use for the further analysis is to select the validated stations. In the case region of The Hague, there are no KNMI stations within the borders of the municipality. As stated in the methodology chapter, the KNMI stations provide validated data. The Netamo database is a combination of different types of sensors. Some of these sensors are reliable but there are also weather stations with less reliable data. The cause of this unreliable data can be twofold. First the quality of the sensor is low and providing unreliable data. Second the quality of the sensor is high but the sensor is misplaced. A sensor placed in the sun can provide inaccurate data. For further investigation, the nearest KNMI stations will be used to validate the Netatmo amateur stations.

The nearest KNMI stations are Hoek van Holland and Rotterdam. According to Google Maps, the distance from the centre of The Hague to the KNMI station Hoek van Holland is approximately 15 km and the distance from the centre of The Hague to the Rotterdam station is approximately 15,5 km. The Hoek van Holland KNMI station is close to the North Sea and the Rotterdam station is more inland. Using a nearest-join analysis, the PC6 regions were assigned to the nearest KNMI stations. Figure 4.1 presents the division of the PC6 regions and clarifies which KNMI station will be used as validation point. The Hoek van Holland station is closest to the western part of The Hague and represents most of the PC6 levels closer to the sea. The Rotterdam station is closest to the eastern part of The Hague and represents the PC6 levels that are more inland.

Figure 4.1: Nearest KNMI stations, The Hague (KNMI, 2020)(CBS, 2020)



After the division of the PC6 levels in The Hague, the Netatmo prepared data was added to the map. As stated in the methodology chapter, the Netatmo data are averaged from every 15 minutes to data for every day and eventually added up to data for a whole year. The degree days were created as stated in the literature review. The Python script was used to create the degree days per Netatmo station for the years 2018 and 2019. In this case, we can compare the Netatmo stations and the KNMI stations with each other.

The total heating degree days for 2018 for the station Hoek van Holland were 2509 and for Rotterdam 2537. For 2019, the total heating degree days for Hoek van Holland were 2443 and for Rotterdam 2547. After assigning each Netatmo station to the nearest KNMI station, the validation could begin. Most of the Netatmo stations are amateur weather stations and not all of these stations report high-quality weather measurements. Both extremely low and extremely high measurements were discovered after a quick scan. Only the measurements with enough quality could be used for further research. The only stations that will be used in this research are the ones with a margin of measurement of 25% to the nearest validating KNMI station.

The margins of measurement for 2018:

- Hoek van Holland: 1882 <2509 <3136
- Rotterdam: 1902 <2537 <3171

The margins of measurment for 2019:

- Hoek van Holland: 1832 <2443 <3054
- Rotterdam 1910: <2547 <3183

The margins of measurement for both years are indicated in the enumeration above. Using the field calculator in QGIS, the simple calculations for the selection of Netatmo stations were performed. All the amateur weather stations within the allocated areas, within the minimum and maximum ranges of the margin of measurement, are useful; all other Netatmo stations were not used.

Figure 4.2 illustrate the range of spread of the degree days linked to the amateur stations. Both 2018 and 2019 are plotted in this chart. The line between the two boxes (quartiles) represent the median of each dataset, the cross in the boxes illustrates the average of that dataset and the whiskers show the variability outside the upper and lower quartiles. Both the Netatmo datset for 2018 and 2019 are showing a dot above the whiskers, these dots represent the highest value of degree days that is present in this dataset. What stands out in the boxplots is that the unfiltered Netatmo stations in 2019 appears to have a larger range than in 2018. Also the range in the selected stations in 2018 and 2019 are much closer to each other but it seems that there is a small difference in degree days.

From the total amount of amateur stations that are in and around The Hague about 35% is therefore useful for further research. This means that from the 475 available stations 307 stations are removed and 168 are useful. Figure 4.3 presents all the Netatmo stations available in the research area, also the invalidated stations in red. Figure 4.3 presents the validated stations with high-quality weather measurements in yellow. The figure clearly shows that not every PC6 area is represented with a Netatmo station. There are 13.458 PC6 areas in The Hague so it would not be reasonable to assume that there are also this amount of stations to represent the areas. The main idea of this thesis is to illustrate the differences in local climate through the PC6 areas. Although almost every PC6 area is in possession of the average gas consumption (discussed in the chapter: methodology). So for this research the closest (validated) station will be used to illustrate the degree days for every PC6 area. Table 4.1 shows how many stations are situated in the total amount of PC6 areas. It is clear to see that the biggest amount of PC6 areas are not provided with a weather station. Both the selected and total station in the case region are shown in the table. Hence there are stations missing in this table. **35** stations are missing for the selected stations and **89** are missing in the total amount of stations. Because of the nearest join method the stations just outside the municipality The Hague can be used as the closed station to a PC6 area instead of a station within the municipality. The assumption in this thesis is that the closest station represent the local climate at best. So the stations don't have to be situated within the case region itself. The same applies for the KNMI stations Hoek van Holland and Rotterdam. Also visible in table 4.1 is that some PC6 areas have more than one Netatmo station, in this case the average of both stations is used for this specific area. The ideal situation would be if for every PC6 area has it's own station to illustrate the local climate. In this case it is the second best method to illustrate the local climate.

Table 4.1: Count of stations in PC6 areas

PC6 count of stations	**PC6 areas with selected stations**	**PC6 areas with total stations**
4 stations	0	1
2 stations	4	22
1 station	125	338
0 stations	13.329	13.097

Figure 4.2: Spread of stations (Netatmo, 2020)

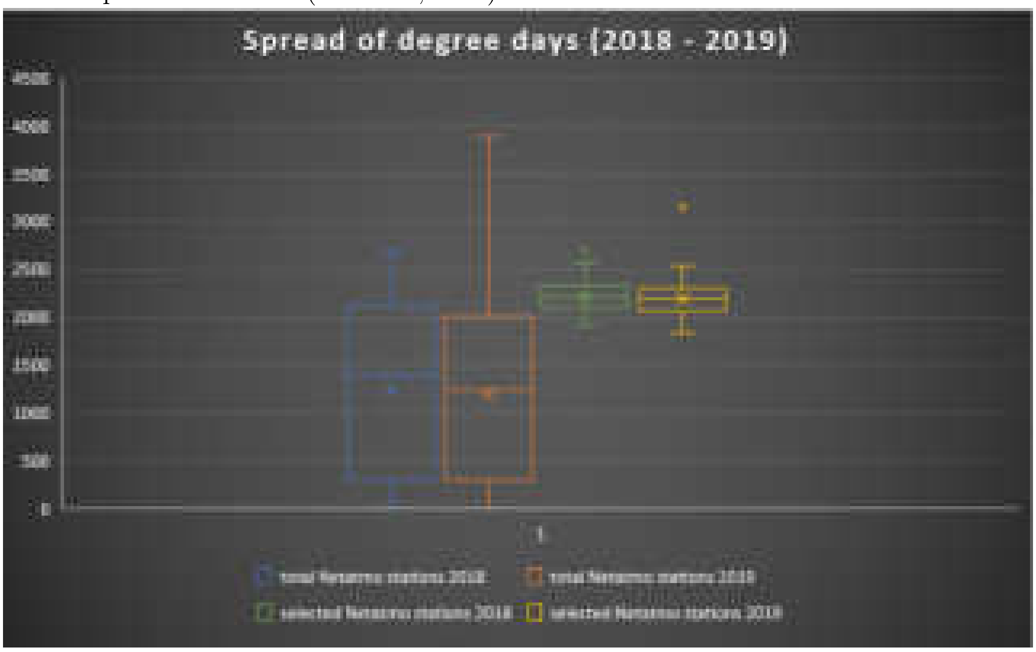
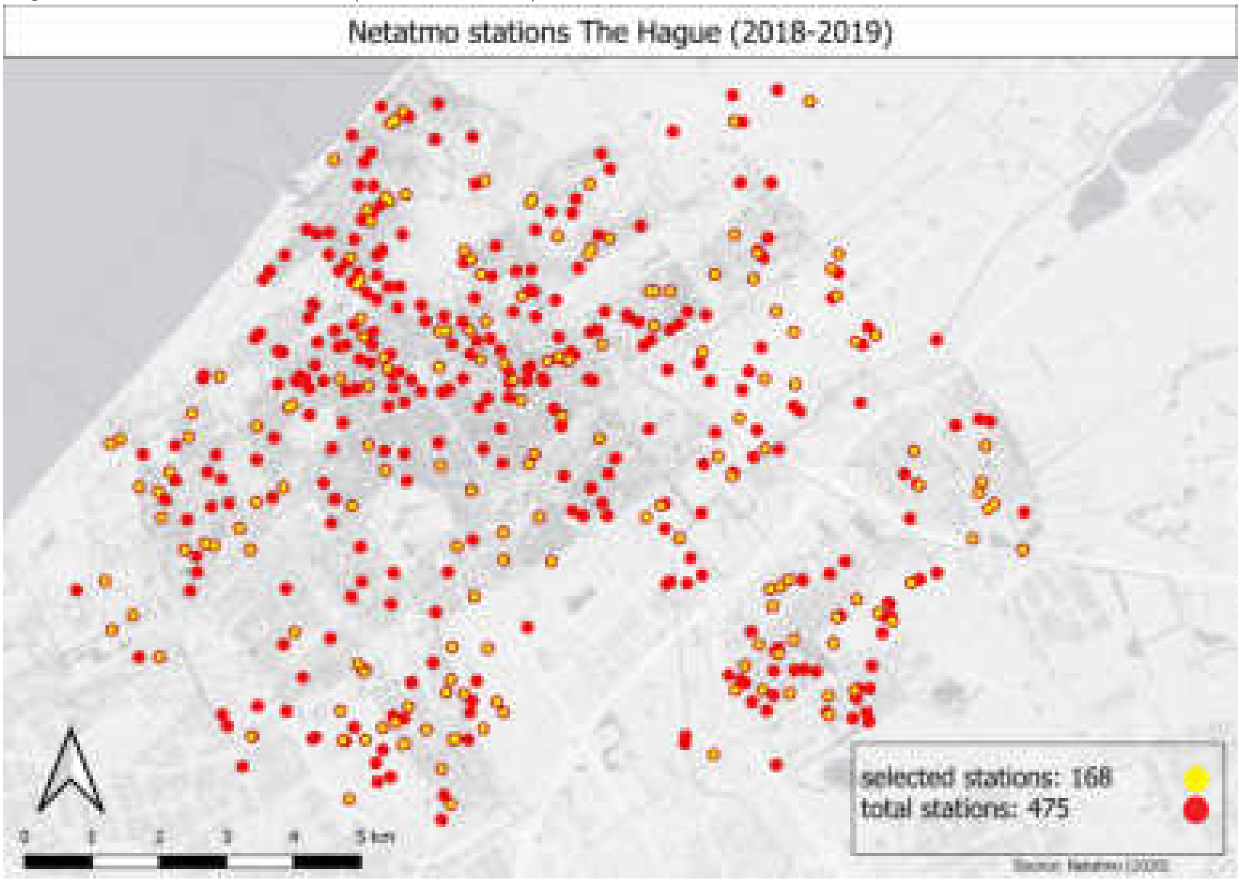


Figure 4.3: Netatmo stations (Netatmo, 2020)



4.2 Analysis of the degree days at PC6 level

4.2.1 Unfiltered weather stations for the degree days

In the section about the KNMI and the Netatmo stations is mentioned that only the validated Netatmo stations would be used for further research to eliminate the extreme values. This section will illustrate why it is not wise to use all the Netatmo stations in the case region of The Hague. In table 4.2 the mean temperatures from Hoek van Holland and Rotterdam are illustrating the mean temperature for 2018 required from the KNMI. With a simple math the corresponding degree days are calculated. This is the following equation, (The baseline for cooling degree days (18 degrees) - the mean temperature) x all the days in one year = are the (approximate) degree days. So, for Hoek van Holland it is (18 - 11,52) x 365 = 2366 degree days. Hence that these degree days are lower than the real degree days in section 5.1. This is because the this equation is a simplified method that assumes that every day in the year 2018 had a mean temperature of 11,52 degrees. Still the real degree days for Hoek van Holland (2509) and this approximate degree days of 2336 are not even that far from each other. The same has been done for Rotterdam. For the selected Netatmo stations and all the Netatmo stations the equation works the otherway around. The degree days are transformed in the mean temperature for 2018 with the mean of the degree days for these stations.

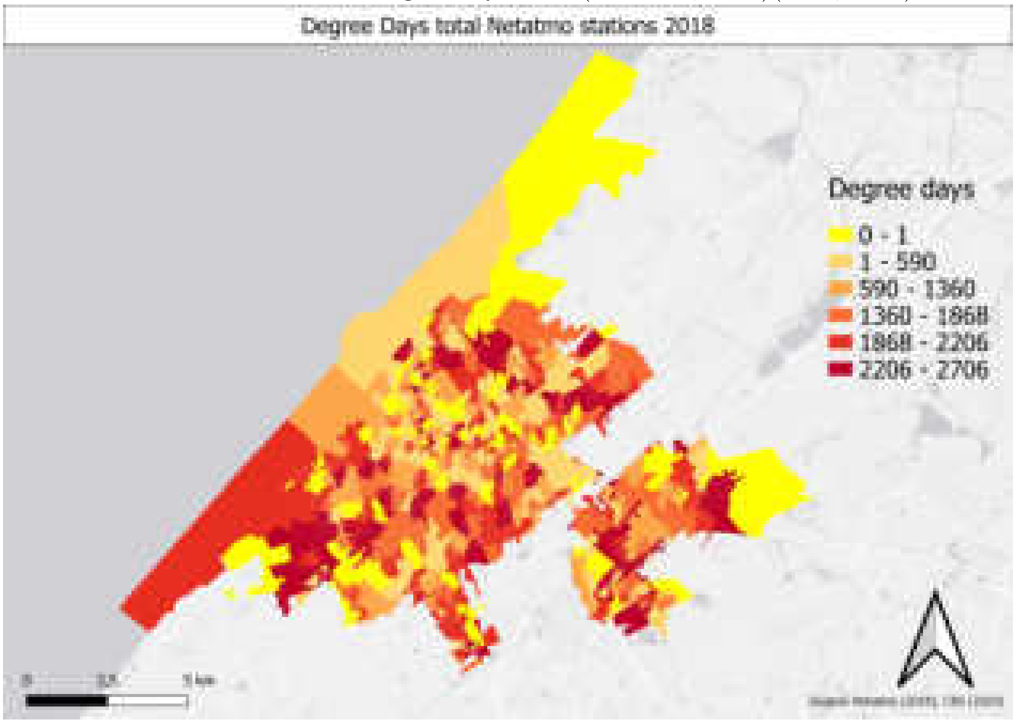
The main point of table 4.2 is to show how biased the outcome will be if all the netatmo stations are used. Figure 4.4 illustrates the the distribution of degree days per PC6 level if all the stations are used. A lot of the PC6 areas will show values between 0 and 590 degree days. These non validated weather stations will suggest that the mean temperature in 2018 around these stations are 16,5 degrees Celsius or higher.

As said in the previous section, more stations scattered around in the case region The Hague will illustrate a more complete idea of the local climate but this would only be the case if the stations are from a higher quality. In this case the more stations are illustrating a disturbed image of the real situation.

Table 4.2: Mean temperatures and degree days

Station or Netatmo data in 2018	**Mean temperature (in degrees Celcius)**	**Total degree days**
Hoek van Holland	11,52	2366
Rotterdam	11,54	2357
Selected points	11,90	2225
Total Points	14,61	1238

Figure 4.4: Total stations Netatmo - degree days 2018 (Netatmo, 2020)(CBS, 2020)



4.2.2 Degree days

After assigning the Netatmo stations and the validation of these stations, another nearest join was performed. In this case, the PC6 areas were the input layer and the validated weather stations from Figure 3.7 were the overlay layer. Every PC6 area was linked to the nearest Netatmo station and the data from this station was linked to this particular area. The result is a new layer, illustrated in Figure 4.5. All the PC6 levels of the case area The Hague were linked with the nearest amateur weather station to create a representation of the degree days distribution within the different areas.

For the representation in the map, the degree days per PC6 level were classified into five levels of more or less the same size. The equal-quantiles method was used to compare the degree days data from 2018 with the degree days data from 2019. The same was done for the degree days for 2019, presented in Figure 4.6.

By comparing the two years with each other, two phenomena stand out. First, the range from low to high values in 2019 is greater than in 2018. In other words: there are more geographical differences and more extreme measurements in 2019 than in 2018. Second, the degree days are differently distributed in the two maps. Also the degree days distribution for 2018 seems to be more diverse. It seems that more smaller groups of areas are at the same level in 2018 than in 2019. Also, more lower values of degree days are clustered together in 2019 than in 2018. If we compare the two maps with the naked eye, we can conclude that there were more degree days in 2018 than in 2019 and that they were more diverse in 2018.

Although the evidence suggests that the degree days in 2018 and 2019 are different, a statistical test needs to be done to verify this. As mentioned previously, maps can be deceiving. The borders of the different levels are arbitrary and the actual values of the PC6 levels can be closer to each other than they seem. For example, the degree days in a particular PC6 area can be 2068 in the year 2018 and 2071 in the year 2019. This is a difference of three degree days in a year of 365 days. This would be a negligible difference, but in the visualisation of the maps it appears as a significant difference.

Figure 4.5: Degree days 2018 (Netatmo, 2020)(CBS, 2020)

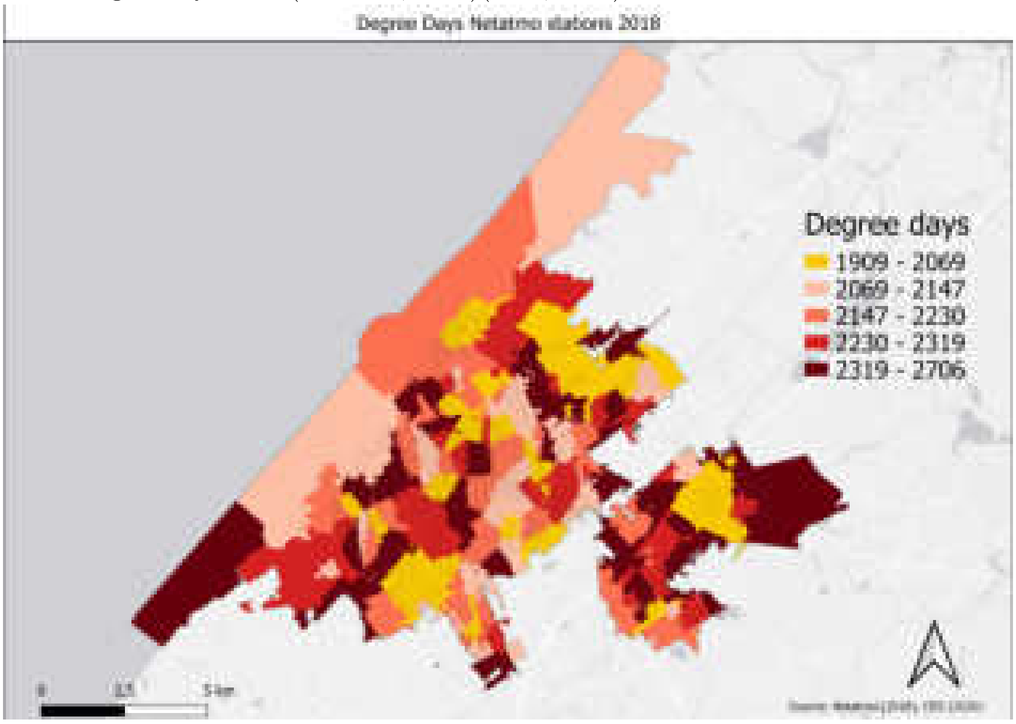
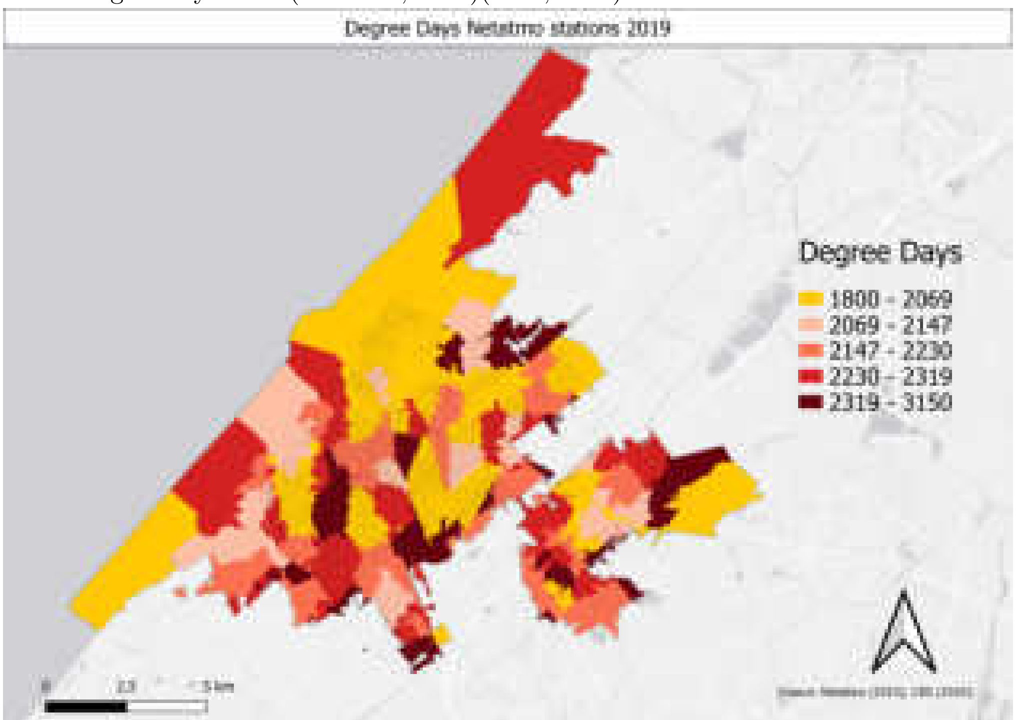


Figure 4.6: Degree days 2019 (Netatmo, 2020)(CBS, 2020)



The paired samples t-test for the degree days

To determine whether the degree days for 2018 and 2019 are statistically different, the paired samples t-test was used. As said in the methodology chapter, the paired samples t-test is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired samples t-test, each subject or entity is measured twice, resulting in pairs of observations.

The P-value of the test is 1,1453E-100. This means the P-value is incredibly low. This leads to the conclusion that the chance of coincidence for the difference between the degree days of 2018 and 2019 is almost zero and this is 99% certain. Although the differences in the dataset are small because of the large number of observations, these small differences can be significant. If there were a smaller number of observations, for example the comparison of municipalities, the differences would no longer be significant. This is not the case, so it can be stated that the degree days in 2018 are significantly different than the degree days in 2019.

Figure 4.7: Paired sample T-test degree days

Paired sample T-test degree days		
	DD_2018	*DD_2019*
Mean	2189.893672	2149.603
Variation	20196.12129	32252.4
Observations	13458	13458
Pearson-correlation	0.099953456	
Expected difference between means	0	
Degrees of freedom	13457	
T-static data	21.48077385	
P(T<=t) one-tailed	5.7266E-101	
Critical area T-test: one-tailed	1.644966867	
P(T<=t) two-tailed	1.1453E-100	
Critical area T-test: two-tailed	1.960140285	

Figure 4.8: Paired sample T-test degree days 2

mean difference	40,29058546
Stand. Dev. Of Difference	217,5924976
standard error of the difference	1,875658006
T alpha half 95% CI	1,960140285
lower confidence level	36,61403264
upper confidence level	43,96713828

4.2.3 Geographical differences

The previous section explained that there is a difference between the degree days in 2018 and 2019 in the municipality The Hague. This could be expected because the degree days have been invented because there are significant climatic differences between years. The other important question is if the PC6 levels also differ in degree days geographically. In other words, what is the relationship of PC6 level X in 2018 with PC6 level X in 2019? In the previous section the degree days of the PC6 levels have all ready been compared with each other. A different year with different climatic factors correspond with different degree days. To find out if there is a geographical trend between the local degree days and the KNMI degree days both 2018 and 2019 need to be investigated.

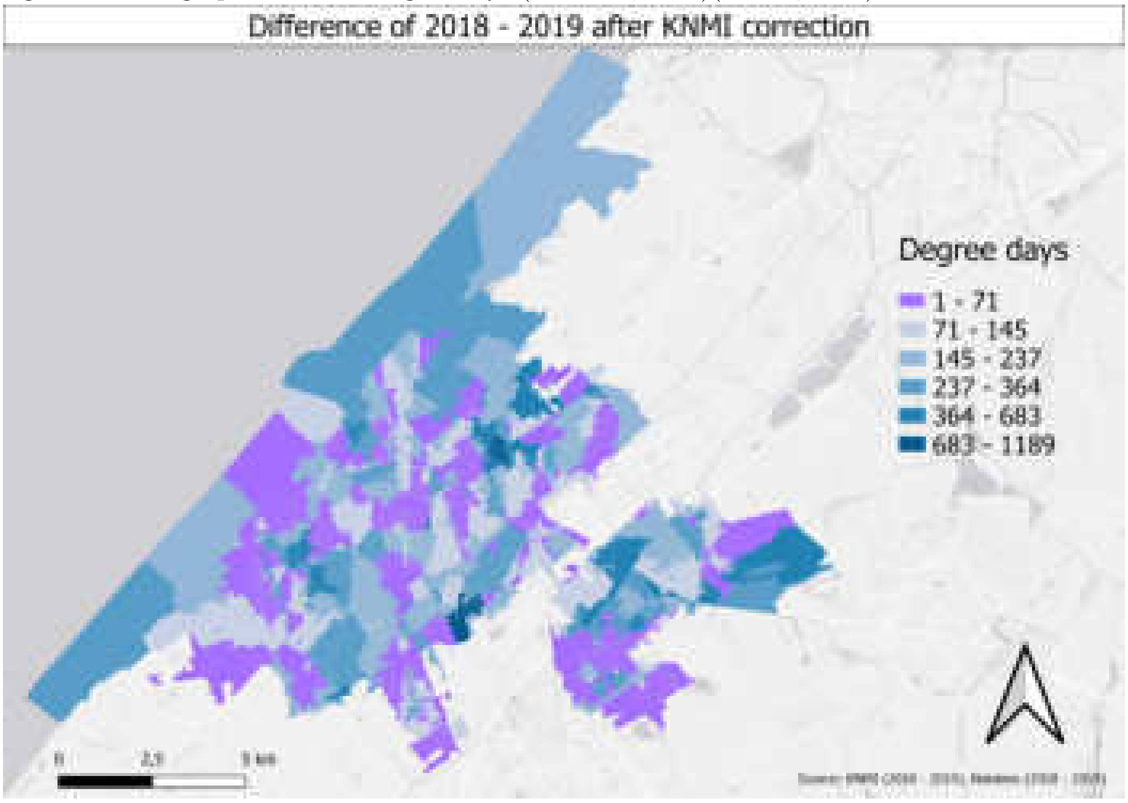
To investigate this trend the difference between the degree days from the KNMI and the difference from the degree days per PC6 level need to be calculated for both years. In this case the differences between PC6 levels can be investigated if they are identical or far apart from each other. If these differences are identical it means that there is a spatial component that can correlate with this difference.

So an example to show how this works: we take a PC6 level closest to the KNMI station Rotterdam. In 2018 in Rotterdam the degree days were 2537 and in 2019 the degree days were 2547. Lets say that in this specific PC6 level in 2018 the degree days were 2517 and in 2019 the degree days were 2527. The following equation will show the trend = ((DD-KNMI-2018) - (DD-PC6-2018)) - ((DD-KNMI-2019) - (DD-PC6-2019)). So (2537 - 2517) = 20 and (2547 - 2527) = 20, (20 - 20) = 0. In this fictional example it will mean that the difference of the difference is the same and there might be spatial factor involved.

Figure 4.9 shows a map of all the these differences of degree days where these equation is used. The labels are categorized with natural breaks (jenks) to show how these differences are spatially distributed. The purple to lighter blue areas represent the areas with the smaller differences and how more darker blue the colours become how bigger the differences are between the differences in degree days with the KNMI correction.

In this case it seems to be that the lighter blue areas are clustered with each other and and the darker blue areas clustered with each other, this could be an indicator that the spatial distribution of these PC6 levels and the differences in degree days from 2018 - 2019 are correlating.

Figure 4.9: Geographical trends degree days (Netatmo, 2020)(KNMI, 2020)



4.3 Analysis of the uncorrected gas use for postal code 6

4.3.1 Household gas use at PC6 level

Besides the degree days, another value needed to be calculated and compared to create the weighted degree days for The Hague, namely the gas consumption. In this thesis, household gas use at PC6 level was used. The CBS and the responsible grid operators provided data for PC6 level for household gas consumption, gas consumption of companies and offices, household electricity consumption and electricity consumption for offices and companies. The CBS dataset was used because they already prepared the data for PC6 level and it could easily be used with the already existing PC6 dataset that was used in the degree days map.

The gas use per household layer for 2018 and 2019 were created with the CBS dataset, which is in a CSV format and joined with the PC6 shapefile of The Hague. New layers of gas use for 2018 and 2019 were created. Figure 4.10 shows the average gas consumption for 2018 and figure 4.11 shows the average gas consumption for 2019. Both maps look similar but there are some differences. Noticeable is that it seems that the centre of The Hague consumes less gas than the south-west region and the north-east region. The literature study mentioned the urban heat islands. It is possible that the centre region uses less gas through the higher temperatures created by UHIs. Another possible explanation is that the climatic influences such as wind, rain and irradiation play a part in this difference. These topics will be discussed later on in this thesis.

Figure 4.12 illustrates the difference in gas use between 2018 and 2019. The negative values represent a decrease of gas use at PC6 level for 2019 and the positive values an increase of gas use. The map also indicates PC6 levels with 0 values, which reflects missing data from the (CBS, 2019). Gas use below PC6 level is privacy-sensitive information. PC6 levels with more than five private households were used in this map; PC6 levels with fewer than five households were not released by the CBS. Based on the map, it looks like many PC6 levels were omitted, but in reality 90% (CBS, 2019) of the PC6 levels were taken into account. The business parks and green areas represent the larger areas with 0 values.

One thing that stands out in figure 4.12 is the difference between south-west (almost all blue-green) and the city centre (orange red). This figure represents the gas use difference and no other factor such as degree days or climatic factors are involved. Still it maybe possible to see the climatic differences between the PC6 levels. One thing that could be noticed is that there seems to be a similar line between the centre region and the south-west region as it is visible in figure 4.1. This is not the case both these stations or the division of The Hague have nothing to do with the gas consumption. Figure 4.12 illustrates that the centre region had a higher gas use in 2019 than in 2018, hence the negative values that are represented by the red an orange colours. The south-west region shows the opposite, the gas use in 2018 was higher than in 2019. If both figure 4.5 and figure 4.6 are being compared with figure 4.12 it seems to be that the the south-west region has higher values of degree days in 2018 than in 2019. This could indicate the difference, because a higher value of degree days means a lower mean temperature through the year. If we take a look at figure 4.9 the spatial distribution of the differences in the south-west and the centre region of The Hague we can see that more regions in the centre the difference is small and in the south-west regions some of the parts seems to occur a bigger difference. Still it is an analysis with the naked eye and not all parts of the south-west region meet these requirements. But it could be an indicator for the differences in gas consumption between the years 2018 and 2019. Also non-climatic factors such as more sustainable houses could be the cause, section 5.5 would elaborate more on this topic.

Figure 4.10: Gas consumption in The Hague, 2018 (CBS, 2018)

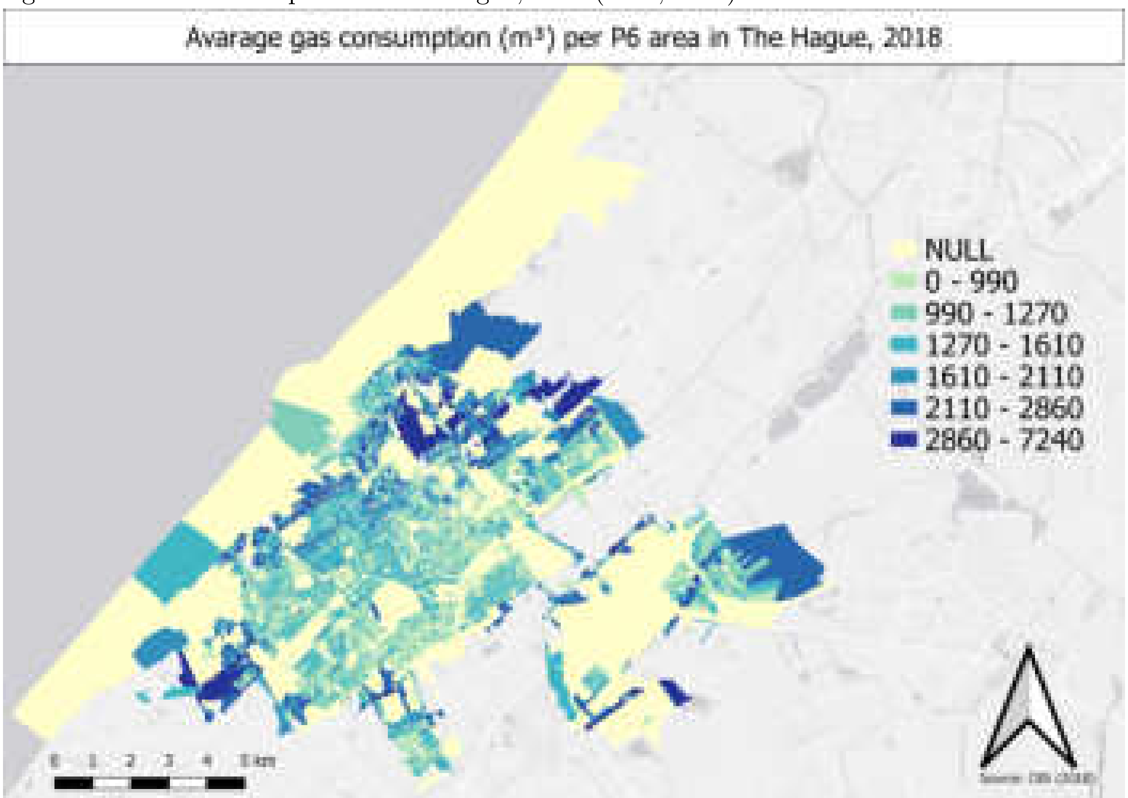


Figure 4.11: Gas consumption in The Hague, 2019 (CBS, 2019)

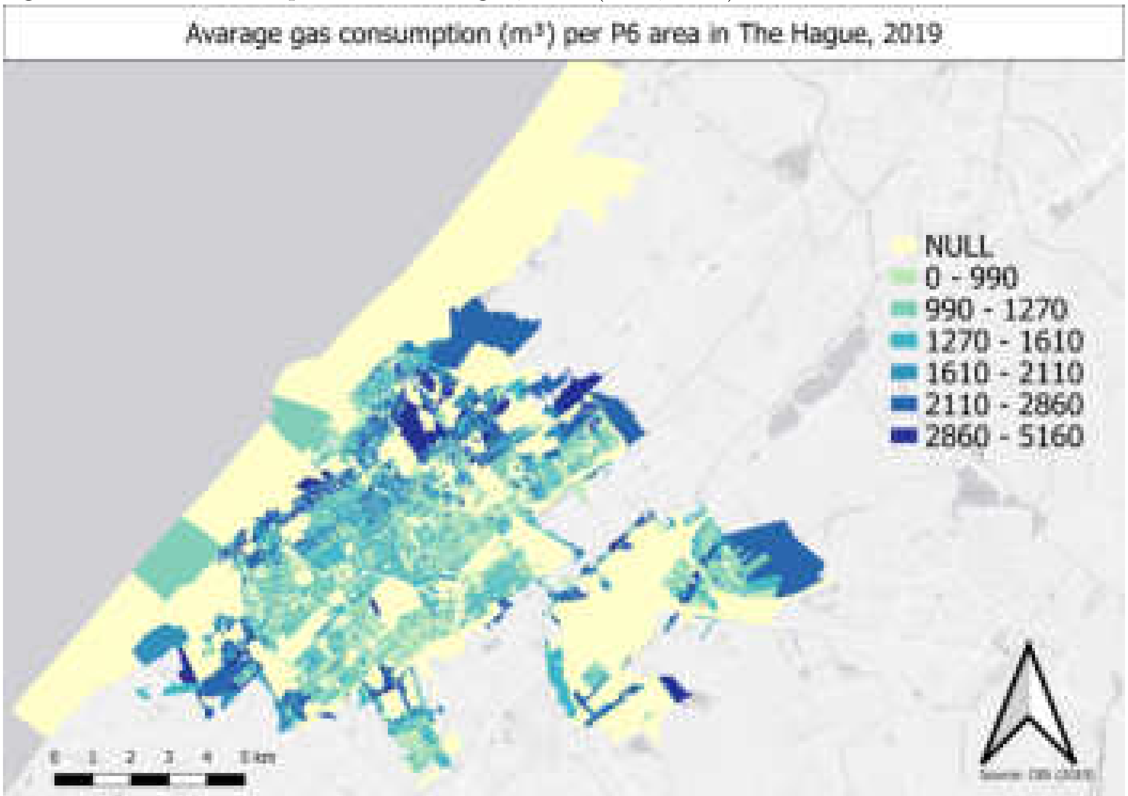
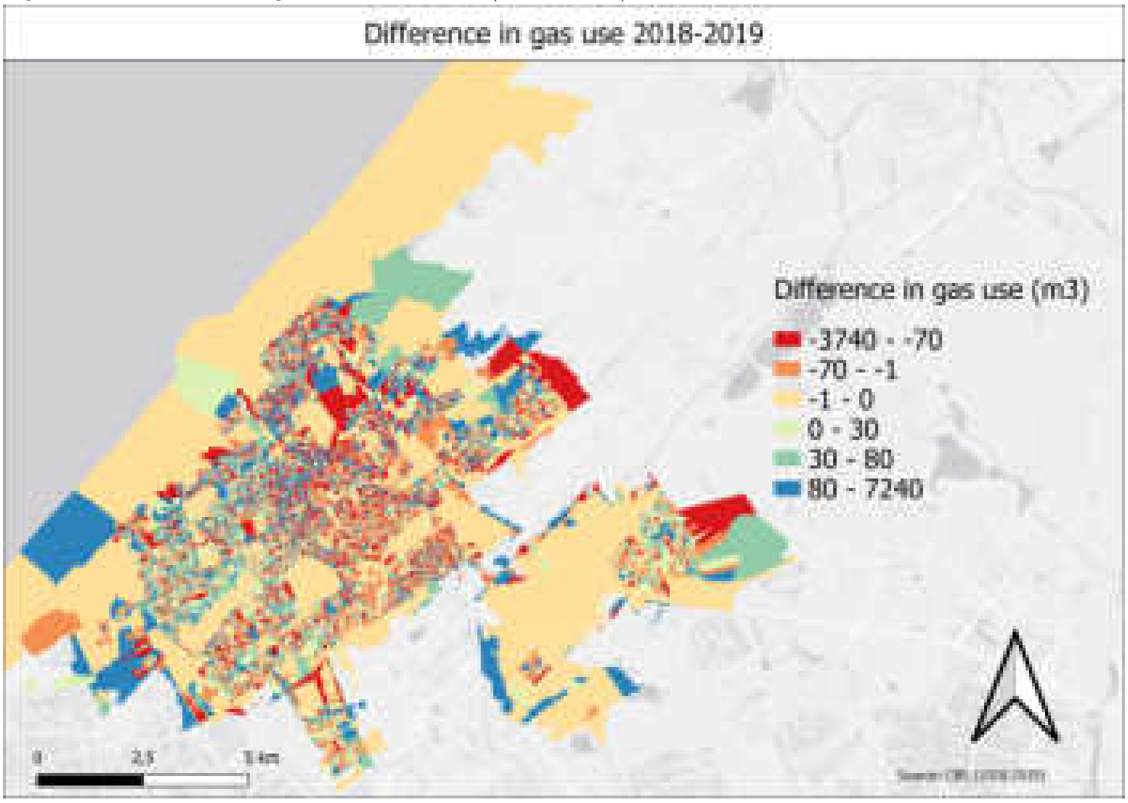


Figure 4.12: Difference gas use 2018-2019 (CBS, 2019)



The paired samples *t*-test for household gas consumption

After comparing household gas use in 2018 and 2019 the following results are shown in figure 4.12. Figure 4.12 indicates both an increase and a decrease in gas consumption after subtracting the gas consumption from 2018 with 2019. There are PC6 levels that used more gas in 2019 than they did in 2018 and vice versa. Based on these two datasets, it is not hard to conclude that gas use differed between 2018 and 2019 than it was for instance for the more arbitrary maps of the degree days. If we take a closer look at figure 4.12 we find that most of the differences in gas use are in the range of 20 to 70 m3 of gas. Nevertheless, to be sure that gas use in 2018 was different than in 2019, a paired samples *t*-test was executed.

The first thing that Figure 4.13 highlights is the number of observations. There were 13,458 observations in the statistical test for the degree days. The paired samples t-test has only 10,914 observations. As mentioned before, the CBS stated that 90% of the PC6 levels in their dataset are valid. For this dataset, only 81% of the PC6 levels are linked to an average gas use level. There are two possible explanations for this decrease in the percentage of validated PC6 levels. One reason could be that the CBS refers to a 90% coverage for the whole dataset. This dataset reflects gas consumption data for the whole of the Netherlands. Another explanation could be that the CBS refers to combined household gas consumption and offices and companies gas consumption.The last data were not taken into account for this thesis. In the dataset from the CBS households and companies are separated but it could be that the coverage is these two datasets merged together. For the further research towards the weighted degree days only 81% of the PC6 levels in The Hague were taken into account.

The P-value of the test is 1,91436E-38. This means the P-value is higher than the P-value of the degree days, but it is still low. The chance of coincidence in the difference between the results for household gas consumption for 2018 and 2019 is almost zero and this is 99% certain. It can be stated that household gas consumption in 2018 was significantly different from household gas consumption in 2019.

Figure 4.13: Paired sample T-test gas use

Paired sample T-test gas consumption		
	GAS_2018	*GAS_2019*
Mean	1321.925967	1300.205
Variation	272674.8755	259801.7
Observations	10914	10914
Pearson-correlation	0.943197549	
Expected difference between means	0	
Degrees of freedom	10913	
T-static data	13.01610888	
P(T<=t) one-tailed	9.57179E-39	
Critical area T-test: one-tailed	1.644993268	
P(T<=t) two-tailed	1.91436E-38	
Critical area T-test: two-tailed	1.960181389	

Figure 4.14: Paired sample T-test gas use 2

mean difference	22.38204306
Stand. Dev. Of Difference	187.5194857
standard error of the difference	1.794958568
T alpha half 95% CI	1.960181389
lower confidence level	18.86359868
upper confidence level	25.90048744

4.4 Analysis of the gas use corrected with degree days at PC6

4.4.1 The correction with degree days

The degree days and the energy use for The Hague at PC6 level are significantly different for 2018 and 2019. These are not surprising results, because the degree days are a representation of every degree Celsius below the borderline of 18 degree Celsius over 24 hours, summarized for a whole year. It would be more surprising if the degree days for 2018 and 2019 were the same. This is also the case for household gas consumption per PC6 level. It would be odd if the gas use in 2018 and 2019 had been exactly the same. In order to compare gas consumption in two different years, the correction of the gas consumption with the the degree days was introduced in the literature study, namely the gas consumption of year X / degree days of year X = gas consumption corrected with the degree day.

The assumption is that the heating degree days at PC6 level are the same for 2018 and 2019, climatic and internal factors do not play a role in the gas consumption at household level. If there is a difference, other factors than temperature play a role in gas consumption. Figure 4.15 and Figure 4.16 present the outcome of this calculation of the gas use correction with degree days for both years. The classification of both maps are the same, so the PC6 levels can be compared. In this case, the division of levels is again arbitrary. The correction with the degree days represents even levels. The first level is a correction of gas consumption of 0.5, displayed in orange, and the other levels have a range of 0.25. The last one, the blue level, displays all the values above 1,0. Again, the 0 value areas are not taken into account and are displayed in red. The first thing that stands out is that both maps look identical. Only for a few PC6 levels, the division into categories is different. It is harder to find these differences with the naked eye than for the maps indicating degree days and gas consumption.

In the case of the degree days map, a distortion of categories occurs when values from the same PC6 level in 2018 were under the borderline for one category and above this line in 2019. Values that were close to each other are visualized as not close. In the case of the correction of gas consumption with the DD, it is the other way around. Because of similarities in the map, it looks like values for the two different years are the same because they belong to the same category. Most of the values are in the range between 0,01 and 1,0. So if a certain PC6 level has a value of 0,76 in 2018 and a value of 0,99 in 2019, the categorisation for both maps leads to the same colour. However, if we take into account that the most common range is between 0,01 and 1,0, the difference between those two values is 23%, which is a much larger difference than it seems.

The paired samples t-test for the gas use corrected with degree days is the most important. The outcome determines whether factors other than climatic have an influence on gas use. With the naked eye, they look the same, but a the statistical test needed to be conducted to determine whether the two years are significantly the same.

As stated before, not all the PC6 levels are in possession of a local weather station. However almost all PC6 areas have an average gas use. The gas consumption corrected with the degree days is most local approach possible in this case. The chapter: discussion and conclusion will elaborate more on this topic.

Figure 4.15: Classification of the correction of DD 2018 (Netatmo, 2020)(CBS, 2019)(CBS, 2020)

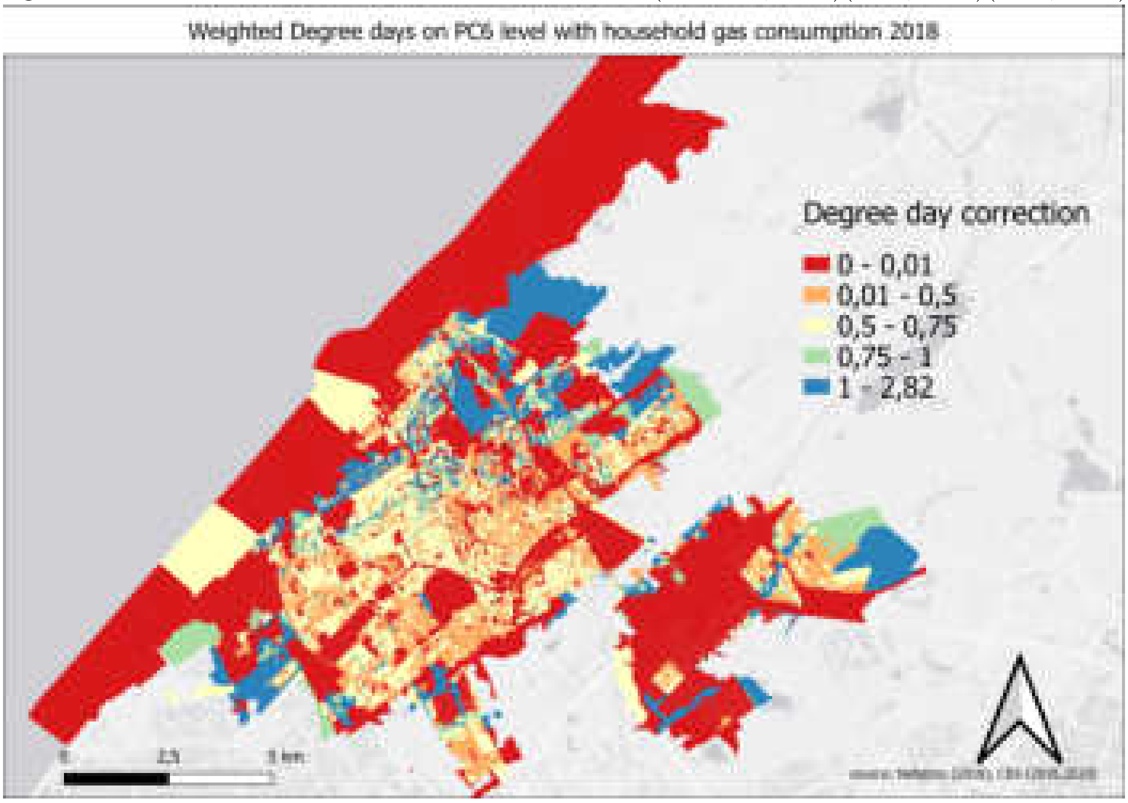
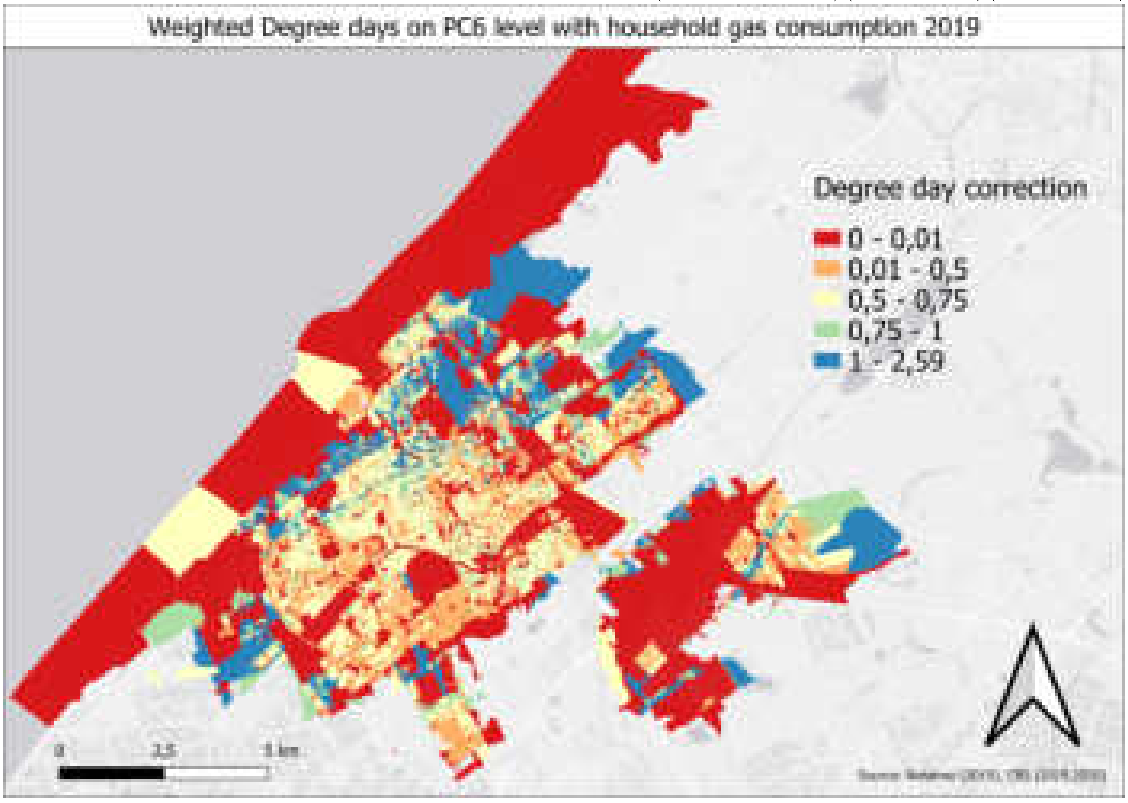


Figure 4.16: Classification of the correction of DD 2019 (Netatmo, 2020)(CBS, 2019)(CBS, 2020)



The paired samples *t*-test for the correction of gas use with degree days

For the paired samples *t*-test with the correction of gas use with degree days, the layers that were created for the correction with the DD 2018 and correction with the DD 2019 needed to be transformed into two CSV files that could be combined at the PC6 level. With CSV files, the *t*-test could be executed, as it was for degree days and gas consumption. Figure 4.17 is the result of the paired samples *t*-test. As stated in the subsection on the *t*-test for gas use, the number of observations is the same, because of the missing values for PC6 levels with fewer than five households. All 0 values are filtered out to avoid an influence on the statistical test.

The averages of correction with the DD 18 and correction with the DD 19 are almost the same. After rounding off the numbers, they both have an average of 0,61. Figure 4.18 indicates a mean difference of -0,004. All the other statistical values in figure 4.18 indicate that the two datasets are incredibly close to each other. Nevertheless, the P-value indicates that both datasets are statistically different from each other.

The P-value of the test is 4.1456E-05. This means that the P-value is higher than the P-values of the degree days and the gas use, but is still low. The chance of coincidence in the difference of the results between the correction of degree days for 2018 and 2019 is extremely low and this is 99% certain. It can be stated that the correction with the DD in 2018 was significantly different from the correction with the DD in 2019.

Figure 4.17: Paired sample T-test gas use corrected with degree days

Paired sample T-test corrected degree days		
	CDD_2018	*CDD_2019*
Mean	0.60675714	0.610746418
Variation	0.059278588	0.062118557
Observations	10913	10913
Pearson-correlation	0.915181801	
Expected difference between means	0	
Degrees of freedom	10912	
T-static data	4.100883116	
P(T<=t) one-tailed	2.07281E-05	
Critical area T-test: one-tailed	1.64499328	
P(T<=t) two-tailed	4.14562E-05	
Critical area T-test: two-tailed	1.960181408	

Figure 4.18: Paired sample T-test corrected degree days 2

mean difference	-0.0033879
Stand. Dev. Of Difference	0.10162686
standard error of the difference	0.000972829
T alpha half 95% CI	1.960181408
lower confidence level	-0.00583482
upper confidence level	-0.00208098

Categorized gas use corrected with degree days

Because the statistical test with the values from the gas use corrected with the degree days were so similar and also the maps were similar it would be interesting to calculate the paired samples t-test for the gas use corrected with degree days again, but this time the values are placed in categories. The original values from the previous t-test were continuous values it the chance was obvious that there would a significant difference. By categorizing the values it could be that there will be no difference between the paired datasets

The datasets are categorized in 26 different categories. The 0 values are categorized in the category 0, the values bet between 0,01 and 0,1 in the category 0,1; the values between 0,11 between 0,2 in category 0,2 and so on. This categorisation is in fact rounding up the numbers. Sometimes the values will be rounded up to high and some times to low, but with datasets with over 10.000 values each the chances of a biased numbers will be low. The results of the paired samples t-test are shown in figure 4.19 and figure 4.20.

Figure 4.19 and figure 4.20 shows that there is no significant differences between the two datasets when they are categorized. Again the 0 values are left out of the dataset. the mean of both categorized datsets are almost the same and that is also visible in the mean difference. The critical area of the two-tailed t-test shows that there are no significant differences. The main goal of this specific paired samples t-test is to show that the differences in the t-test of the gas use corrected by degree days does seems to be significant different but with this test it shows that it needs to put in perspective. The differences are small and when you categorize the values the differences are not visible.

Figure 4.19: Paired sample T-test categorized gas use corrected with degree days

Paired sample T-test with catagorized gas use corrected with degree days		
	catagorized 2018	catagorized 2019
Mean	0.61747851	0.616027359
Variation	0.058070329	0.060271848
Observations	10819	10819
Pearson-correlation	0.892308415	
Expected difference between means	0	
Degrees of freedom	10818	
T-static data	1.336085637	
P(T<=t) one-tailed	0.090774694	
Area T-test: one-tailed	1.644994494	
P(T<=t) two-tailed	0.181549388	
Critical area T-test: to tailed	1.960183298	

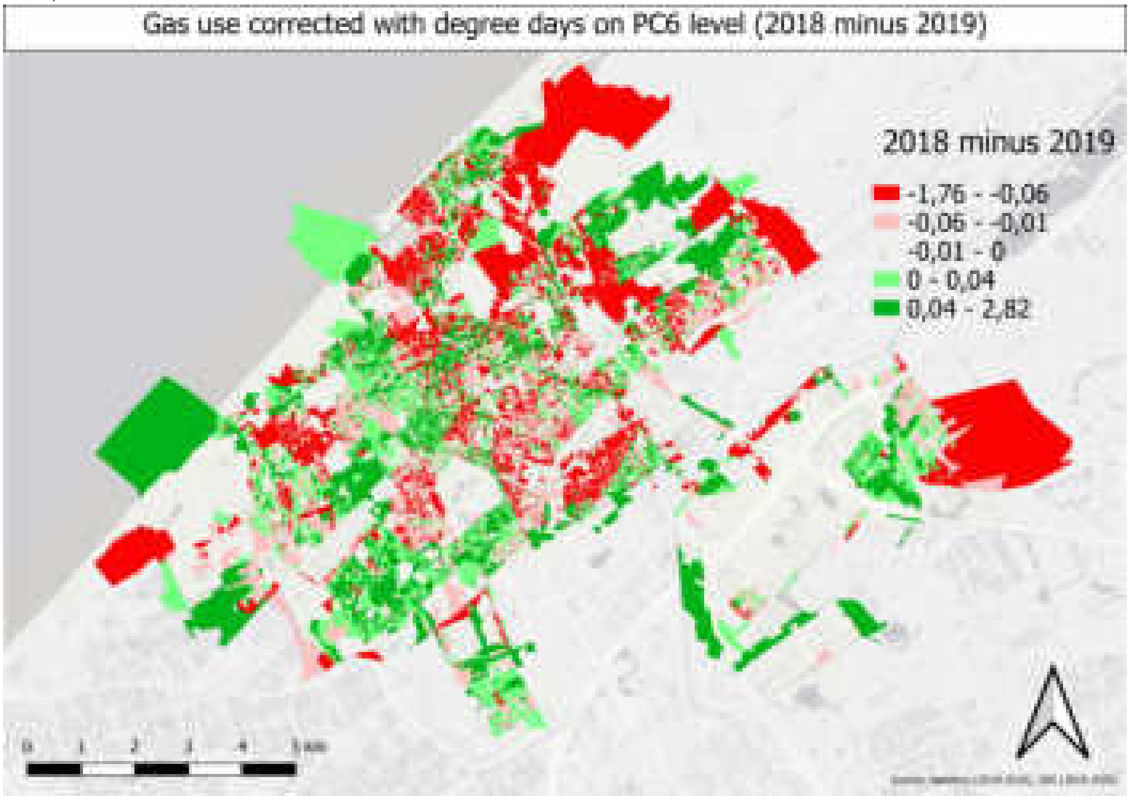
Figure 4.20: Paired sample T-test categorized gas use corrected with degree days 2

mean difference	0.001451
Stand. Dev. Of Difference	0.112972255
standard error of the difference	0.001086121
T alpha half 95% CI	1.336085637
lower confidence level	-1.50754E-07
upper confidence level	0.002902151

The difference in consumption per DD

The results of the paired samples t-test for the weighted degree days indicate that 2018 was significantly different from 2019. Although both maps for 2018 and 2019 seem the same, there is a difference. This means that other factors besides temperature had an influence on gas consumption at PC6 level. The statistical test concluded that there was a difference, but this difference was very small. Further investigation is needed to determine how this difference is geographically distributed. Figure 4.21 presents the result of the difference between the correction with the DD for 2018 and 2019.

Figure 4.21: Difference in gas use corrected with degree days 2018-2019 (Netatmo, 2020)(CBS, 2019)



If we take a closer look to Figure 4.21 two things stand out. First, again many PC6 levels are omitted. Second, there seems to be a kind of spatial distribution of positive and negative values. To start, the first phenomenon is that again the 0 values are omitted from the data. In this case, that can mean two things:

- The PC6 levels have fewer than five households; or
- the difference between correction with the DD 2018 and correction with the DD 2019 is 0.

The last reason is a new one. Because of the small differences between the two datasets, it is not impossible that the correction with the DD in 2018 is the same as in 2019, and there is no positive or negative difference.

It seems that the positive values are clustered together and the negative values are clustered together. This is not the case for all PC6 levels, but there seems to be some kind of spatial pattern in the distribution of the negative and positive values. The south-west shows more positive values and the center and north areas show more negative values.

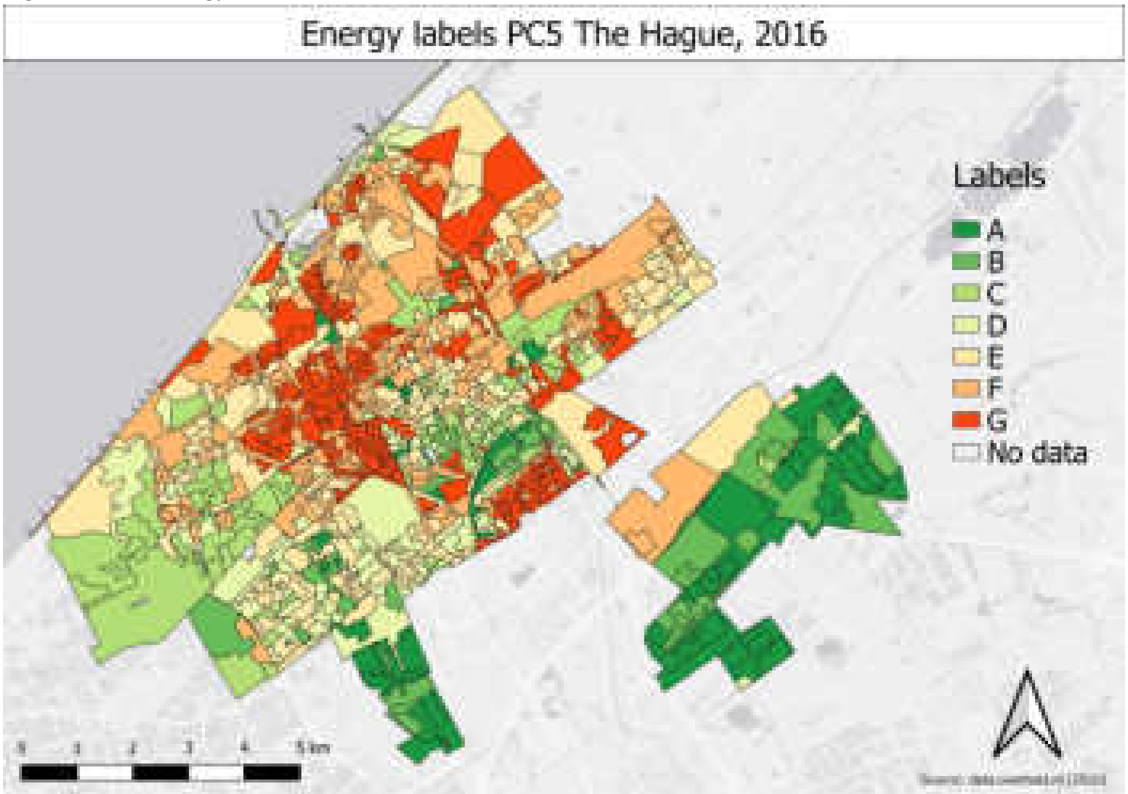
4.5 Climatic and internal factors

As stated at the beginning of this chapter, if there is a small significant difference between the degree days in 2018 and those in 2019, factors other than temperature are responsible for that. This could be climatic factors or internal factors, such as more sustainability of the PC6 levels. The following sections discuss which factors contribute to the differences in gas use corrected with degree days and to what extent. The first part of this section will discuss the non-climatic factor: sustainability where the energy labels will be discussed for The Hague in 2018 and 2019. The second part will elaborate the climatic factors such as irradiation, rain and wind.

4.5.1 Internal factors

It may be that some buildings have been renovated between 2018 and 2019, which would explain some of the variations between 2018 and 2019. The data centre for The Hague has many different datasets and shapefiles containing heat and cold storage, rooftops with solar panels and residential buildings that are disconnected from the gas network. Most of these datasets are not fully complete and are hard to use to compare different PC6 levels. There is one way to compare the PC6 levels in terms of their sustainability and that is via energy labels. The data centre for The Hague has a shapefile on PC5 level with the energy labels for 2016. Unfortunately, there are no shapefiles for the years 2018 and 2019. Figure 4.22 is the map that shows these energy labels on PC5 level. What stands out in this figure is that the south, south-west and south-east are provided with more sustainable labels, the city centre has less sustainable labels.

Figure 4.22: Energy labels 2016 PC5 level



Some datasets provide energy label data at district and neighbourhood level (Gemeente Den Haag, 2020) for the past two years (2019 and 2020) on the total number of houses per district or neighbourhood level per energy label. These data are open source and can be downloaded. Unfortunately, there are no data available in this dataset for 2018. The (CBS, 2018) has a dataset with information about the energy labels in 2018, in percentages per district and neighbourhood level.

The CBS (2018) only categorizes the energy labels as A, B, C, D and E-G. The data from The Hague (Gemeente Den Haag, 2020) categorizes the data as A, B, C, D, E, F and G. If we want to compare the labels from 2018 and 2019, the last three categories need to be merged. One dataset (2018) is in percentages and the other (2019) is in total houses per category per label. With an easy calculation in Excel, both years are provided, with the total numbers and percentages. Figure 4.23 and Figure 4.24 present the percentages for 2018 and 2019 for the whole municipality of The Hague.

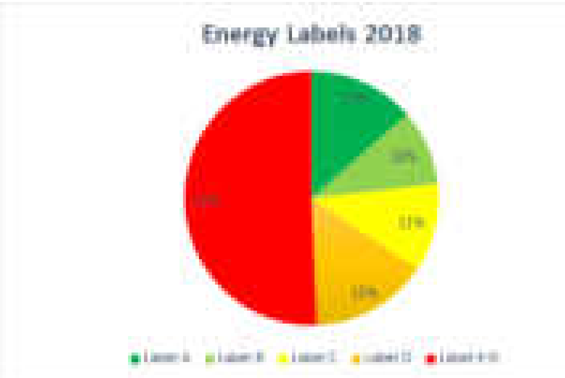


Figure 4.23: energy label 2018

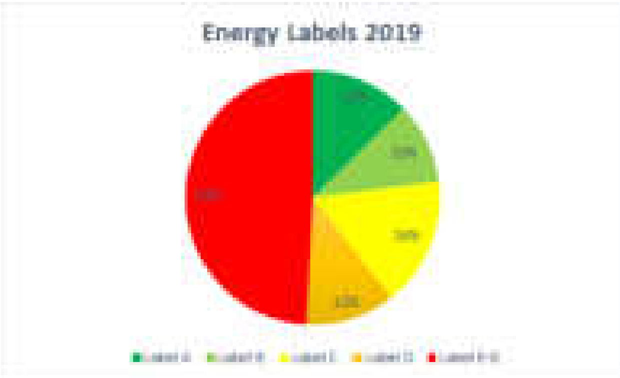


Figure 4.24: energy label 2019

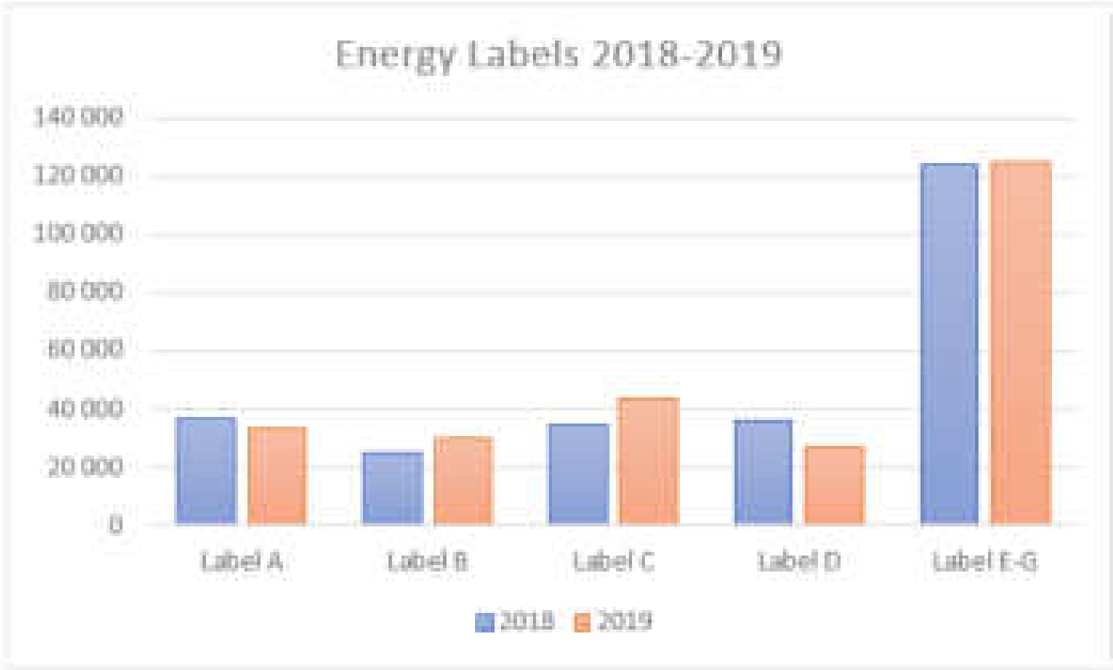
Figure 4.23 and figure 4.24 indicate that there are almost no differences per label category for 2018 and 2019. Label E-G decreased slightly, from 51% to 49%, and label D decreased in 2019 by 3% while label C increased by 5%. It is nice to see that the energy labels did not change that much in one year, but this does not give a good overview of how the PC6 levels relate to each other in the different years. To investigate that, the total numbers per PC6 level need to be compared with each other.

Figure 4.25 presents the total number of energy labels for 2018 and 2019. Instead of the percentage per label for the total number of labels that were indicated in Figure 4.23 and figure 4.24, the absolute numbers give a better overview to compare the labels. The differences between 2018 and 2019 are easier to see. The first thing that stands out is that labels E-G outnumber the other labels. This is also the main problem if we want to compare the different PC6 levels with each other. Almost all PC6 levels fall into this last category. If both years were mapped and we used the most common energy label per PC6 level, the whole map would be covered by the E-G label. Also the values in Figure 4.25 do not clearly indicate the energy labels. There are more label A counts in 2018 than there are in 2019. Using QGIS, is it possible compare the percentages for each region with the percentages of the same region for another year. Unfortunately, the data from both datasets are for neighbourhoods and districts, as said before. There is no comparison between the PC6 levels. If we want to visualise the percentages per neighbourhood or district in a pie-chart and display it on a map, it would be confusing and unclear.

However, it is not a bad idea to use the energy labels to compare PC6 levels to see whether they are getting more sustainable. There is only a need for more complete data. If 2018 had the individual labels E, F and G, it would be easier to divide the different districts and neighbourhoods into the most common energy labels. This thesis focused on PC6 levels, so comparing district or neighbourhood data is not ideal. If the CBS or the municipality of The Hague could provide the

PC6 data, it would be more useful. The energy labels do not change within in a timeframe of two years. For a really good overview of how the PC6 levels are becoming more sustainable, we need to check a longer time frame. With these data, it is not possible to say whether sustainability is a factor that influences the gas use for households. On the other hand energy labels are valid for 10 years, so it is unlikely that people register all the changes. When they are selling their house and did some changes like changing the glazing, they would probably ask for a new label.

Figure 4.25: Energy labels 2018-2019 (CBS, 2018)(Gemeente Den Haag, 2020)



4.5.2 Climatic factors

Irradiation

Besides the factor of sustainability, there are also climatic factors that can play a role in the different gas use. One of the factors could be irradiation, which is not measured by the Netatmo stations. The government database website (overheid.nl, 2010) provides a shapefile with the calculated irradiation per rooftop for the city The Hague. This was calculated in 2010 using a 3D model and takes the following factors into account:

- the rooftop surface;
- the slope of the rooftop; and
- the location of the rooftop.

One of the attributes of this shapefile is the attribute irradiation which values each BAG premise from 0 to 1100 (overheid.nl, 2010). This shapefile was originally created to indicate which rooftops are most suitable for solar panels. A value of 0 means it is not suitable and the more the value rises, the more suitable the rooftop is for solar panels. This thesis used the shapefile to determine how much irradiation each rooftop gets. The more suitable for solar panels, the more irradiation the BAG premise gets. Figure 4.26 illustrates the average irradiation per cell from 50 meters x 50 meters. The original dataset was divided in BAG premises. In this thesis normally such datsets will be translated to PC6 areas but because of the division of a low amount of premises in combination with big PC6 areas in some places the distortion of the visualisation of the map was considerable. So for figure 4.26 the decission was made to create a raster display to show the differences in irradiation in a more or less same scale as the PC6 areas. This map was created by adding all the irradiation values and dividing them by the number of BAG premises in the specific raster cell. Figure 4.26 therefore presents the 50x50 display of irradiation that have the most suitable rooftops for solar panels and therefore the most irradiation.

There are many flaws in this way of reasoning. First of all, the shapefile from overheid.nl (2010) is from 2010. The time frame this thesis investigates is eight and nine years later. Nevertheless, it is still possible to get some useful information from it. The Hague changed in these years, but most of the buildings from 2010 are still present in 2018–2019. Using Mindergas (mindergas, 2021), one can calculate the number of sun hours per year per KNMI station. According to the KNMI, in 2018 there were more sun hours than in 2019, approximately 5% more. So it is possible to compare Figure 4.26 with the gas use data of 2018 and 2019 and see whether in 2018 the more irradiated PC6 areas used less gas.

Figure 4.27 is the representation of this action. The following equation is created to compare the gas use of 2018 and 2019 with the irradation and amount of sun hours per PC6 area. First the average irradiation is calculated per PC6 are through sum up the irradiation of all the BAG premises in a PC6 area and divide it with the same amount of BAG premises. Second the irradation is multiplied with the amount of sun hours (PC6 areas x Hoek van Holland & PC6 areas x Rotterdam) both for 2018 and 2019. There are now two datasets, one for 2018 and one for 2019. Both datasets are divided with the corresponding gas use for both years. The last step was to subtract these to datasets (2018 minus 2019) and figure 4.27 is the result. A positive percentage means that the gas use corrected with the irradation in 2018 was that amount of percentages higher than 2019, and a negative percentage means the opposite. The yellow PC6 areas illustrate that there was no difference between 2018 and 2019. It is not so strange that there are more green PC6 areas than red ones. As said before there where more sun hours in 2018 than in 2019. The noticeable aspect of Figure 4.27 is that even with the correction of sun hours and irradiation there are still some PC6 areas that had a higher gas consumption in 2019 than in 2018. So there is another (climatic) factor that causes this phenomenon.

Figure 4.26: Irradiation The Hague 50 M x 50 M (overheid.nl, 2010)

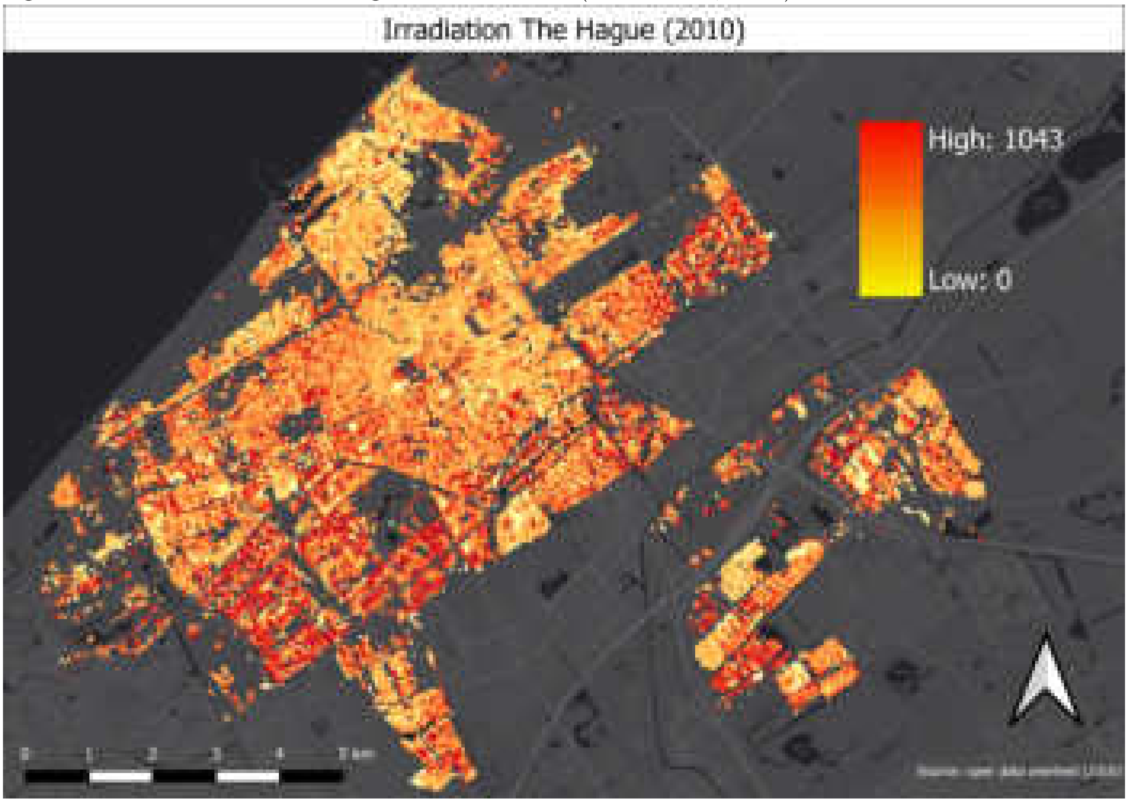
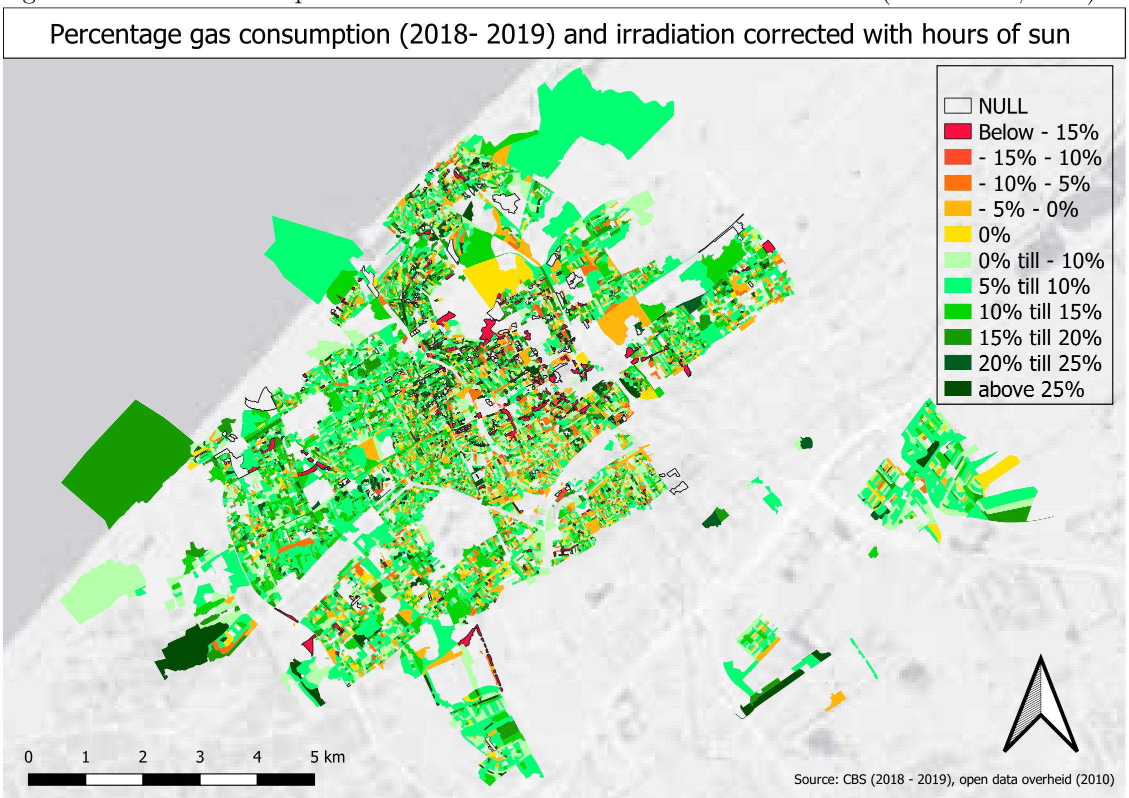


Figure 4.27: Gas use compared with irradation corrected with hours of sun (overheid.nl, 2010)



Rain

Contrary to the irradiation, hourly rain and 24-hour rain figures are collected in the Netatmo datasets. At least, some Netatmo stations can collect these data. As for temperature, validation of the rain data could have been done with the KNMI stations. The data could have been validated with the same margins of measurement as for temperature, as only the quality type of rain sensors could be used. There is only one problem: almost none of the weather stations collect rain data.

For temperature, 17% of the values were missing. After preparation with Python and Excel, a check was done on how much data was missing: 83% of the temperature data was collected and validated with the KNMI stations. This same check was done for the 24-hour rain data: 63% of the data was missing. This means only 37% of the data was available, and even these numbers were sometimes dubious. As an example, the 24-hour rain data were available for a specific station and indicated 0,1, but the column next to it displayed a value of 0,3. So there were some inaccurate numbers. If the validation with the KNMI had been done, almost no stations would have remained. A nearest join, as with the temperature (degree day) data, would have distorted the actual historical situation. The few Netatmo stations left with high-quality data could have been used for some PC6 levels, but not for the whole of The Hague.

Wind

As with irradiation, the Netatmo stations do not measure wind data. Unlike the irradiation data, there is no other open source data set that displays wind data. The only open data source for this kind of data is the KNMI. Unfortunately, it is not possible to collect wind data at PC6 level. There is a German website, Wetterzentrale (2021),that collects the KNMI data and visualizes the data in a proper way. The data are available for free and in case of wind data, it could be a good reference for data. Figure 4.28 is a representation of the yearly wind data for 2018 and 2019 both for Hoek van Holland and Rotterdam. There are several aspects that are noticeable for this illustration. First of all, the wind direction for 2018 is in Hoek van Holland more or less the same as in Rotterdam. This is not that astonishing because both stations are not that far apart from each other (approximately 23 kilometers). The same applies for the wind direction in 2019 for Hoek van Holland an Rotterdam. The second aspect that stands out is that the wind intensity in Hoek van Holland is bigger than in Rotterdam for both 2018 and 2019. This can be explained through the fact that station Hoek van Holland is situated nearby the coast and station Rotterdam is more inland. The last aspect that stands out is that the wind direction in 2018 was mostly south, south-west, west and in a smaller amount also east till north-east. In 2019 the wind direction was most of the time south and south-west.

If we would make a analysis with the naked eye to conclude if the wind is a noticeable factor in the gas consumption two things need to stand out on a gas consumption map. First the buildings that are directly influenced by the wind direction from the the south till south-west need a higher gas consumption than buildings that are hidden behind other buildings or other barricades from the same direction. The second factor is that buildings situated close to the coast need a higher gas consumption than the more inland buildings due to the differences in intensity shown in Figure 4.28. In this thesis the PC6 areas are investigated and the first factor (directly influenced by wind direction) is hard to measure with PC6 areas, because it can only be measured at individual building level, but the second factor (situated on the coast or more inland) can be investigated. In section 4.3 the differences in gas use were showed. This section showed that there are differences in gas use between 2018 and 2019, Overall the differences between 2018 and 2019 were small. A PC6 are in 2018 with a high value of gas consumption was still a PC6 are with a high gas consumption in 2019 and vise versa. In this example the year 2019 will be used an is shown in figure 4.29. This year showed a clear direction from south till south-west. It is noticeable that the coastal PC areas have a higher gas consumption than the more inland areas. Again it could indicate that the factor wind plays a role in differences between local gas consumption but is an analysis on the naked eye.

Figure 4.28: Wind intensity and direction (KNMI, 2020)

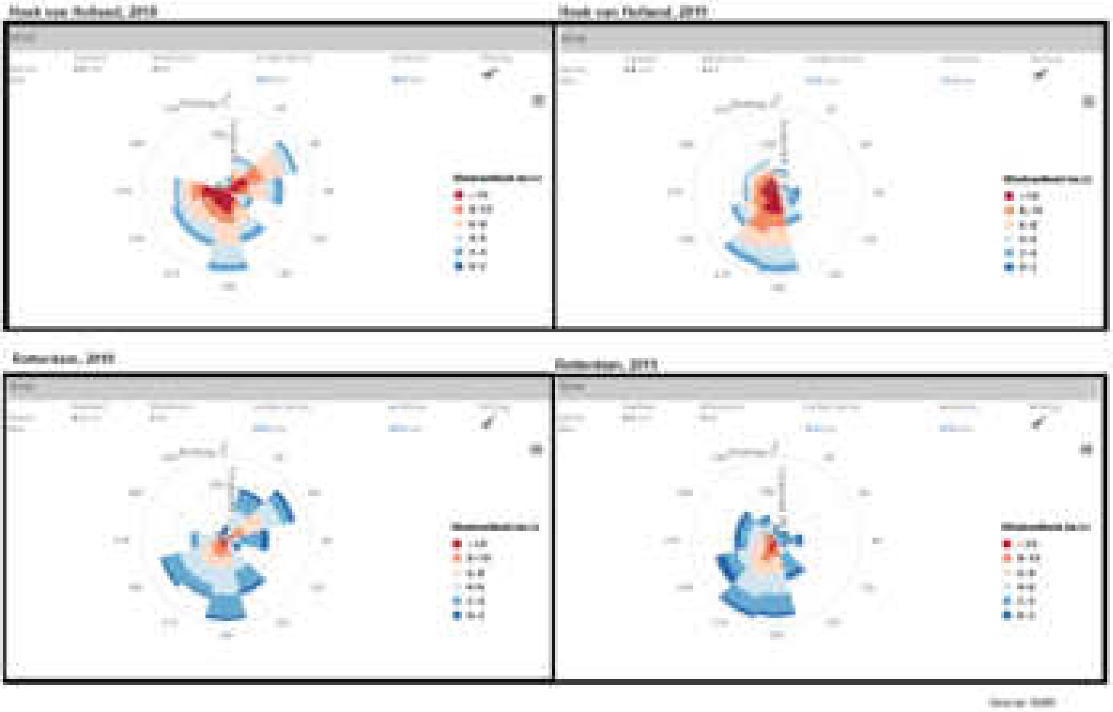
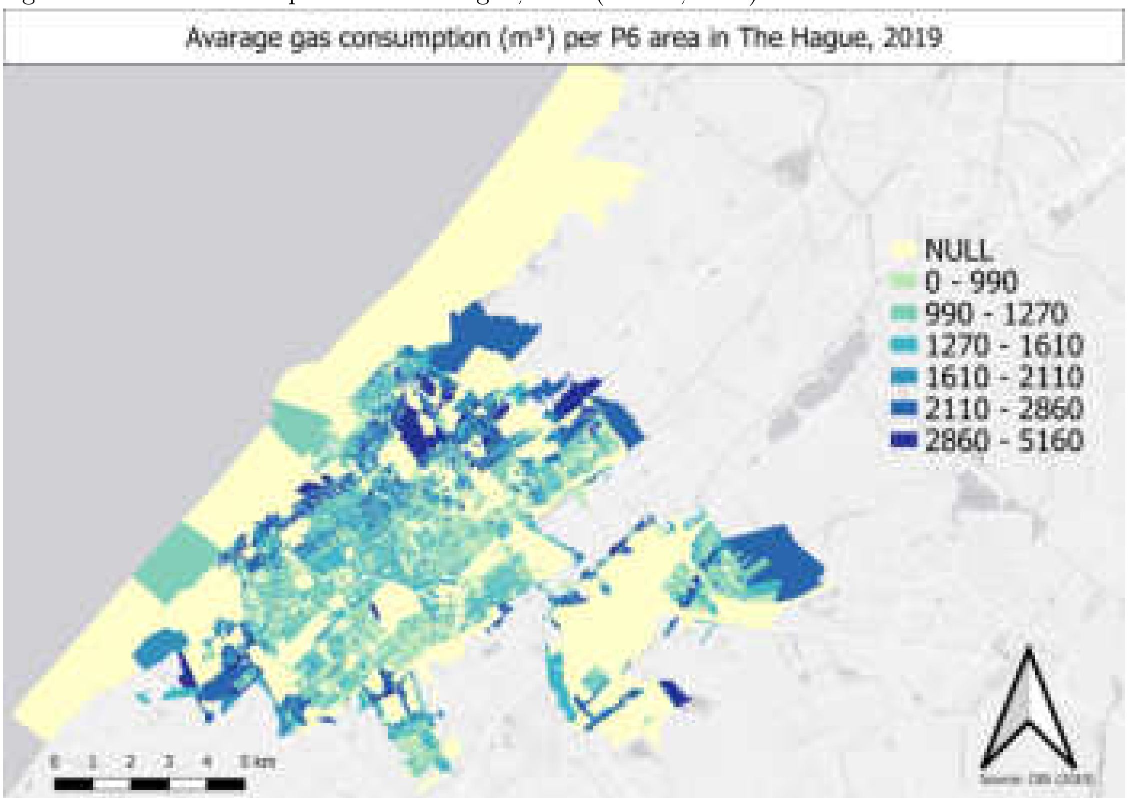


Figure 4.29: Gas consumption in The Hague, 2019 (KNMI, 2020)



4.6 Non-climatic factor: green areas

Green areas are considered to have a positive effect on the reduction of urban heat islands, mentioned in the literature study. Therefore green areas have an effect on the local climate of a city. This factor is not considered a climatic factor but the literature study showed the essence of green areas on the local climate. Figure 4.30 presents the green areas for The Hague. In contrast to the climatic factors that are discussed in this thesis, the data for green areas and trees are easy to find. *Publieke Dienstverlening Op de Kaart (2020)* provides green areas and trees in its *Basisregistratie Grootschalige Topografie* (BGT) and these are open sources and up to date. The dataset BGT contains all the elements of the outside world in the Netherlands. For example, the green areas, roads, public benches, street lights, etc. are included in this dataset. Figure 4.30 shows the different type of green areas in The Hague. Most of them are different type of forest and grassland but also the dunes are taken into account in this dataset. The forest areas are the most important because they provide shadow and evaporation of water as stated in the literature study. The dunes are typical for the coastal part of The Hague but are not considered as green areas in this research. Hence the yellow colour in the figure.

In figure 4.31 shows the percentage of green area for each PC6 level. In this map the greenery type: dunes from figure 4.30 are left out. What stands out is that the PC6 areas in the south-west, north and the north-east have the highest percentages of green areas. The centre of the city has the lowest percentages of green areas. The most important question in this section is if the green areas have a noticeable effect on the local climate?

In the previous section there was an investigation of the climatic factor wind and the effects on the gas consumption in figure 4.11. The literature study showed that the green areas have an cooling effect on the local climate in a urban heat island. This thesis has a focus on heating residential buildings when it is cold. It is not that obvious to compare the gas consumption with the green areas. Nevertheless even during colder days and wintertime the UHIs could occur. In this case a lower gas use in a PC6 area could be correlate with a higher temperature that is caused by a UHI due less vegetation. If we compare figure 4.11 with figure 4.31 an unusual phenomena occurs. It seems that the centre region of The Hague has a lower gas consumption than the parts in the south-west, north and the north-east. These are the parts the percentages of green areas are the highest. So from this analysis on the naked eye it doesn't seems that the green areas effect the gas use. The areas with more greenery should use more gas because the UHI doesn't affect these areas as much as the centre region, this doesn't seems the case but could also be compensated due to the better energy labels.

As said before the gas use is effected by the local temperature. In this thesis the local temperature is represented due the degree days. Besides the comparison of the gas use with the green areas it is also necessary to compare the degree days with the green areas. Figure 4.5 and figure 4.6 are the representation of the degree days in 2018 and 2019 that all ready have been discussed. These maps illustrate different degree days in the same PC6 areas if both years are compared. In 2018 the degree days are higher than in 2019 centre region and the northern and southern part of The Hague. The east and west are more or less the same. Both Figure 4.5 and figure 4.6 don't seems to have a relation with the percentage of green areas in the PC6 levels. Again, the literature study showed that green areas have an effect especially when it is warmer. It could be that the effect of the green areas are much more noticeable is when the cooling degree days are being investigated.

Figure 4.30: Green areas The Hague (Publieke Dienstverlening Op de Kaart, 2020)

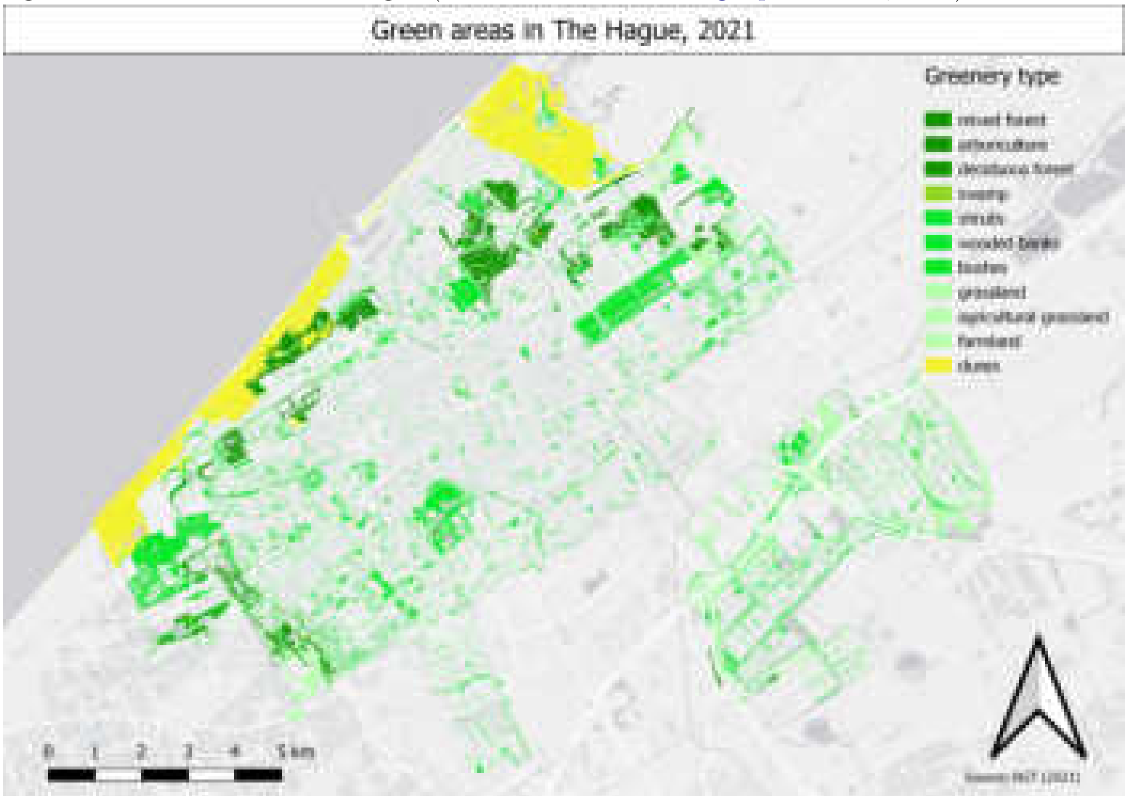
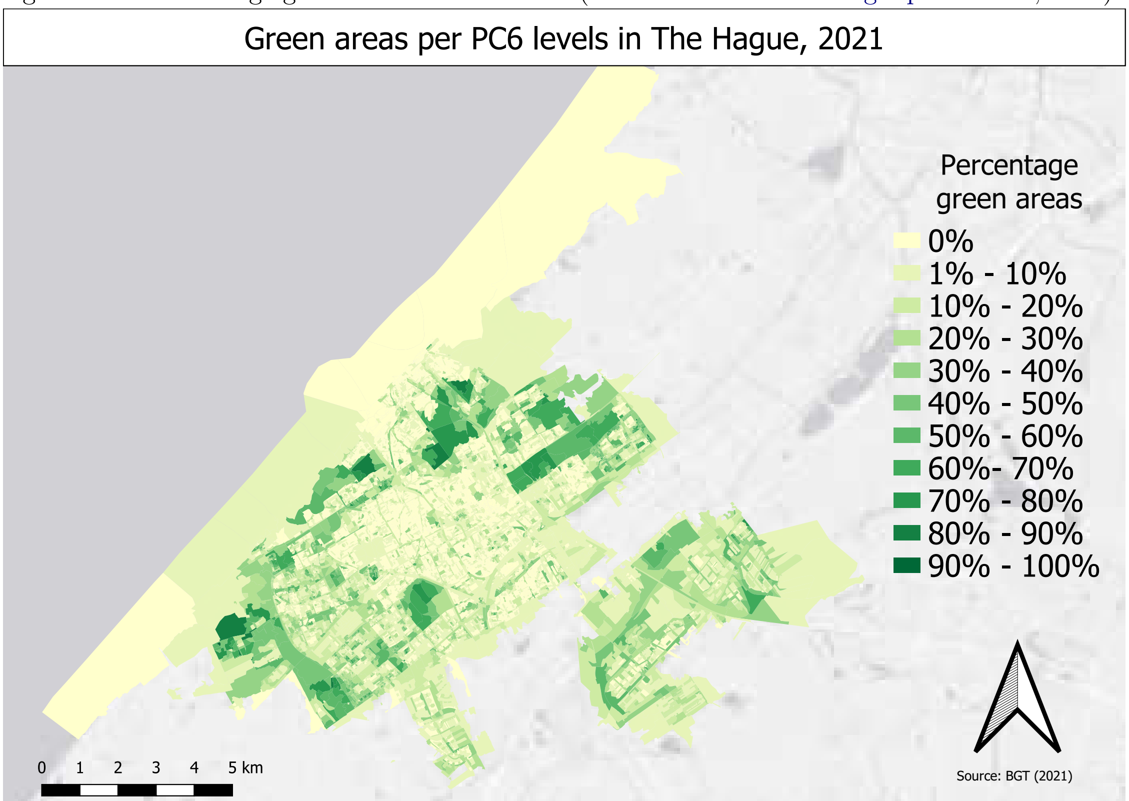


Figure 4.31: Percentage green areas on PC6 level (Publieke Dienstverlening Op de Kaart, 2020)



Chapter 5

Discussion and conclusion

This last chapter will elaborate the discussion and conclusion. First the limitations of the research will be discussed. The Netatmo dataset provided this thesis with the local climatic data. Unfortunately not all the required data was available and some of the data didn't match with the expectations. Besides the Netatmo data the other limitations will be discussed. The second section will answer the sub-questions and the main question that where investigated. The last section will give options and recommendations about further research on this topic.

5.1 Limitations of the research

5.1.1 Netatmo dataset

Most of the results were obtained from Netatmo data. Without the amateur weather stations of Netatmo, the local degree days could not have been calculated. In the methodology chapter and the results chapter, the processing of the data and the Netatmo dataset were discussed. This dataset would not have existed if Dr. Ir. Ohori had not created this dataset. For further research on this topic of local climatic influences, a dataset such as the one used for this thesis should be created. For this research study, the case region of The Hague was used partly because of the existence of this dataset. If other researchers would conduct similar research, it is recommended that a database such as this one should be created, because it is not that straightforward to get the historical information from the Netatmo website.

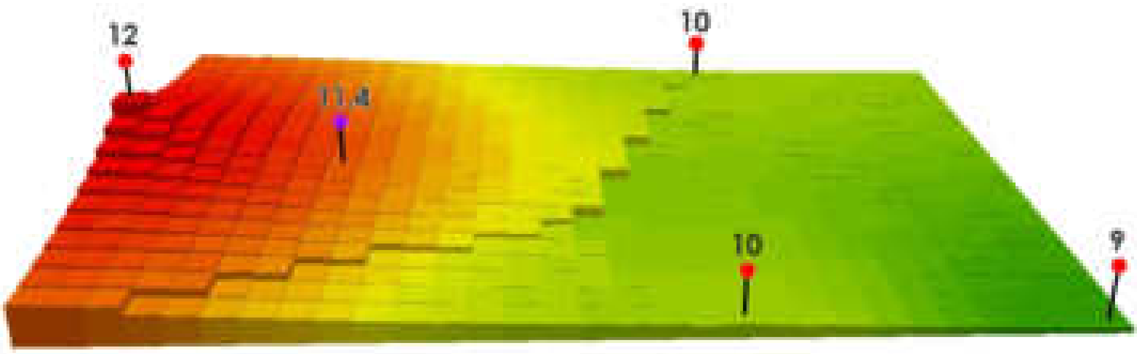
As stated in the chapter: results, not all the Netatmo stations where from a high quality or stations are misplaced causing odd measurements. A section in this chapter explained how the stations where validated. Figure 4.3 illustrated that only 168 of the 475 stations are useful after validation. There are other possibilities to collect amateur weather stations and to illustrate how the local climate differs within regions. Besides the Netatmo dataset there is also another network of amateur weather stations namely from the website *Wunderground.com* (Wunderground, 2021). This is a similar network of amateur stations. Although these stations will probably from the same quality as the from the Netatmo dataset a combination of both could create a network from more high quality stations than the separated datasets. When there will be a combination of both datasets a more refined network of sensors will occur and the differences between the local climatic effects can be illustrated better. Another option is to create your one network of sensors. This is a more expensive and time consuming exercise but you can place higher quality stations on the places of you choosing. The use of open data is easier and less expensive but it could also lead to inaccuracy of the data. The last option is a combination of amateur weather stations with high quality sensors. It works the same way as with the KNMI stations but in this case you place some high quality sensors in a case region and validate the amateur stations with these high quality stations. The distance between the KNMI stations and the amateur stations in this research could be more than five kilometres. With a own grid of high quality stations and amateur stations the local climate and validation could be even better put in perspective.The time-series of the Netatmo data is good. This means the relative values are good but the absolute values are not.

5.1.2 Other limitations and explanation

In this research the margins of measurement is 25%. This margins of measurements comes from the increase of temperature the UHIs can produce. In the literature study Lund et al. (2014) described that in Tel Aviv the UHIs can produce an anthropogenic heat between 1–5°C in summer times and that heavily trafficked streets accounted for up to 2°C warmer temperatures. The main temperature in the Netherlands was 11,3°C in 2018 and 11,2°C in 2019 according to the KNMI. Because of the more moderate climate in The Hague compared with Tel Aviv the decision was made to take a temperature increase of 3°C that correspond with 25% margin of Measurement. Because this thesis researched the heating degree days that mostly are due winter the margin of measurement is 25% above and beneath the validated stations of Rotterdam and Hoek van Holland.

Besides this explanation of the margin of measurement there is also an limitation mentioned in the section about the gas consumption corrected by the degree days. A problem in figure 4.15 and Figure 4.16 is that the gas consumption is at PC6 level, but the degree days are not. There are only 138 validated stations in this thesis. So there is an inaccuracy in gas consumption per degree day. The section about unfiltered weather stations all ready explained why not all the stations where used. But even with all the Netatmo stations not all the PC6 areas could be provided with a amateur station. This is the next best thing. In this thesis the nearest neighborhood method is used, there are other ways to modify the data to divide the degree day data over the PC6 areas. Inverse Distance Weighting (IDW) is an option. The IDW method is an distance interpolation with a set of points. The points get values on a weighted average on their position determination. Figure 5.1 illustrate how the IDW interpolation works. Imagine that the red pinpoints are the Netatmo stations and the purple pinpoint is a PC6 area. Instead of using the closest Netatmo station through the nearest neighbor (in this case pinpoint with value 12) it gets a weighted average from the surrounding stations. It still isn't a measurement in this specific PC6 area. The benefit of this method is that the margin of measurement could be bigger, outliers will be averaged with the surrounding measurements.

Figure 5.1: Inverse Distance Weighting (GISGeography, 2020)



5.2 Conclusion

This section discusses the outcomes of the results and attempts to answer the questions as fully as possible. In the section: Research questions the research gaps that where dissembled by the literature study where transformed into sub-questions. The chapter: results investigated these sub-questions. First the four sub-questions will be answered and then the main research question will be discussed and answered. To start this the first sub-question was:

- **Are there local differences within a city between the degree days at postal code level 6?**

As stated in the chapter on results, there is a significant difference between the gas consumption in 2018 and in 2019 in the municipality of The Hague. Figure 4.12 indicated that the difference between 2018 and 2019 could be both positive and negative, meaning that there are PC6 levels that used more gas in 2018 than in 2019 and vice versa. The maps were revealing enough to conclude that there was a difference between 2018 and 2019, but to make sure, as explained in the results chapter, a statistical test was performed. The outcome indicated that there was a significant difference between gas use in 2018 and 2019.

What can be concluded from these results? First, the differences between 2018 and 2019 were both positive and negative. If the temperature differences in 2018 were much lower than in 2019, one would expect more gas use in 2018. The timeframe 2018–2019 is very short, so the results indicating positive and negative amounts of gas use compared to the year before are not that strange. However, there are differences between two consecutive years, so there must be a reason why the gas consumption was not the same. The main conclusion is that there is a difference, so apparently there are factors that influence gas use. The results show that the degree days is one of these factors.

So there is a difference in gas consumption by households. The results also indicated that there is a difference in degree days. As stated in the section on KNMI and Netatmo stations, the degree days for 2018 and 2019 did not differ that much from each other. Both stations recorded around 2500 degree days for both years. The margin of measurement was more or less the same for both types of stations and both years. Nevertheless the results showed that there where local differences within a city between the degree days at postal code level 6. The remark on this conclusion is that not all the PC6 areas had a local station. The PC6 areas where clusters of local degree days. This is due the characteristics and placement of the Netatmo stations. So the following sub-question was:

- **Does a correction with local HDD helps to understand the local heating energy use?**

The expectation for the heating degree days maps for both 2018 and 2019 was that these maps would be similar, but the results indicated that the maps were not similar and the statistical test confirmed this. How is it possible that the same stations indicated different results for 2018 and 2019? The literature study discussed urban heat islands. This could be one of the explanations for the different heating degree days for the same location for different years. The Netatmo stations measured temperature and it could be that these UHIs influenced the temperature locally through the release of anthropogenic heat by combustion processes, such as traffic. The assumption is that most outdoor objects, such as buildings, streets and industrial activities, do not change that much in one year. Shortwave radiation from the sun or the amount of traffic or turbulent heat transport from within streets, caused by a reduction of wind, can be different in different years. This could be one of the explanations for the difference in heating degree days for 2018 and 2019.

With a difference in degree days and gas consumption, one would assume a difference in gas use corrected with degree days for the two years, but the results indicated two maps that were almost identical. A difference in gas use and a difference in degree days resulted in almost identical difference in gas use corrected with degree days. With only the two maps as evidence, one would say that the correction of gas use corrected with degree days is the correct way to calculate and compare differences in gas use for different years. The statistical test concluded otherwise. Although the differences between 2018 and 2019 were small and almost not visible on the map, there was a difference between the datasets for 2018 and 2019.

With the paired samples t-test as evidence, we can conclude that factors other than temperature are responsible for the difference in gas use, although the difference may be small and the other factors do not contribute that much to gas use differences. Notice the paired samples t-test with the categorized gas use corrected with degree days

However, there are still a few question marks. This thesis only calculated the differences for two years. Because of the limited dataset, there was no other option, but it would be better to compare more years using difference in gas use corrected with degree days. Also, the difference in temperature between 2018 and 2019 was not great. The evidence is that there is a small difference between the degree days in 2018 and 2019. If we compare two years with a greater difference in degree days, the difference in gas use corrected with degree days could also perhaps differ more. This could also mean that the effects of the other factors are more visible. the third sub-qestion was:

- **Is the local influence of wind, solar irradiation and rainfall well accounted for in the weighted HDD?**

In the last part of the results the climatic factors came across. Missing key data sources to conclude if the local influence of wind, solar irradiation and rainfall is well accounted for in the weighted heating degree days is hard to say. The results show that there are some indicators that the climatic factors wind and irradiation are accounted for in the weighted degree days. Hence the differences in gas use corrected with degree days. On the other hand it could also be that non-climatic factors are responsible for these differences. The last sub-question tried to capture these factors:

- **Are there other factors like the ones captured by the energy label explaining differences?**

As said in the section about energy labels in the chapter: results there is not a direct connection visible between energy labels and gas use differences. The assumption was that the energy labels could explain the differences in gas use if the climatic factors couldn't. Besides that there is no up to date dataset of energy labels on PC6 level for the years 2018 and 2019 it is also the case that not all changes according to energy labels are (directly) registered. In this case the energy labels couldn't explain the differences in gas use. It is a brought factor that shows the more energy friendly residential buildings or energy friendly neighborhoods.

Besides the energy labels also the green areas in The Hague where investigated if they could be an indicator for explaining differences besides the local climate. In contrary with the energy labels the green areas are influencing the local climate as described in the literature study about UHIs. With these conclusions about the sub-questions the main question remains:

- **Are there outdoor climatic differences at postal code 6 level within large cities and what are the effects of these variations on heating energy consumption?**

So one of the findings is that factors other than temperature have an effect on gas use. The main research question is: Are there outdoor climatic differences at postal code 6 level within large cities and what are the effects of these variations on heating energy consumption? In the last part of the results factors other than temperature are discussed. Unfortunately, because of a lack of open data the part of the question dealing with to what extent could not be answered.

Besides temperature, other factors have an effect on heating energy use, in this case household gas consumption. There are two types of factors that could play a role in this difference, climatic factors and non-climatic factors such as changes in buildings and green areas. Climatic factors include wind, rain and irradiation, while changes to buildings could mean that they are becoming more sustainable. If we could determine that the PC6 levels were not more sustainable in 2018 that in 2019, then we could conclude that a climatic factor is the cause of the difference in gas use corrected with degree days.

Also, the results of the climatic research did not provide a define answer. If these results were very clear, we could conclude that climatic factors played a role in the difference in gas use corrected with degree days. These results could be researched at the level of KNMI stations, but not at PC6 level, simply because the data on PC6 levels are not suitable.

Are there outdoor climatic differences at postal code 6 level within large cities and what are the effects of these variations on heating energy consumption? Yes there are, and these do not only concern temperature. Can we prove which climatic factors have an influence and to what extent? Probably, but not with the data that is freely available.

5.3 Expanding the research

5.3.1 The scope of the research

This thesis focused on just one city, The Hague. It would be interesting to compare different cities and see whether the gas consumption corrected with degree days in other cities also differ from each other. As stated in the last section, if one wants to compare PC6 levels, a similar dataset to the Netatmo dataset would be advisable. However, it is also possible to compare different cities with each other at municipal level. How would a city such as Enschede, situated in the east of the Netherlands, compare with The Hague? The geographical placement of cities has an effect on the climatic effects within this city, as stated in the literature study. It would be advisable to use cities nearby a KNMI station. Further research could explain these differences in geographical placement and can conduct a study towards the differences in gas use corrected with degree days with these variety of placements taken into consideration. An example could be: are there differences in gas consumption corrected with degree days within city X and city Y and what is the variety in between cities.

Besides researching other cities there is also an option to research The Hague in another way. Figure 4.1 shows that this thesis divided The Hague between the two different KNMI stations, namely: Hoek van Holland and Rotterdam. It would also be interesting to divide The Hague in another way, namely: coastal and inland. In the chapter: results, the section about wind all ready did this. Instead of using the division of these two KNMI stations one would also could divide The Hague in a coastal part (2 kilometers from the coast for instance) and inland (the rest of The Hague). In this way the influence from the sea and wind could be investigated in a different way.

The previous paragraphs focus on geographical additions to expand the research. It is also advisable to compare more years with each other. As stated in the chapter: methodology, the Netatmo dataset that was collected had only all the data for the years 2018 and 2019. The other years in the dataset had gaps between certain days and sometimes between months. In this case the heating degree days for a whole year couldn't be calculated. When this data is available more years could be investigated. Trends in energy labels and (heating) degree days could be researched better.

5.3.2 Other climatic and non-climatic factors

This thesis discussed the three most important climatic factors after temperature, namely irradiation, wind and rain. According to Pijpers-van Esch (2015), other factors also influence UHIs and therefore the outside temperature, for instance: traffic and industrial activities are some of these factors that have an effect. These factors are not climatic, but influence local climatic circumstances. Industrial activities, traffic and also the number of stone buildings and roads have a negative effect by keeping temperatures high. This has an effect on the health of the inhabitants of local areas, it is assumed that this also has an effect on the reduction of gas use. Overall, UHIs are considered to be a negative development (Pijpers-van Esch, 2015).

The green areas are discussed in the chapter: results. The coverage of buildings is also not discussed in this thesis. In the theoretical chapter Britter and Hanna (2003) discussed the effects of higher buildings on circulation of air circulation. Also Stone Jr (2012) explained the effects of different laduse/land cover types. In further research these non-climatic effects can be taken into account. A simple solution and an effective dataset to conduct this research would be to make us of the *Basisregistratie Adressen en Gebouwen* (BAG). The dataset BAG contains all the buildings in the Netherlands. A combination with green areas, buildings and water cover can explain some non-climatic effects of differences in degree days corrected by gas consumption on a local level. The 3D geoinformation research group from the TU Delft (TUDelft3D, 2021) created a 3D BAG. Besides the normal BAG it would be interesting to use this dataset. This dataset could be used as a JSON, PostgreSQL or via a Web Feature Service (WFS) or Web Map Service (WMS). With this 3D dataset the factors wind and irradiation could be investigated further.

Besides researching new factors that may contribute to the local climate, it is also advisable to investigate the factors mentioned previously: wind, rain and irradiation. At the moment it is not possible to measure their exact influence on the local climate as this is not measured in a relative manner. It is possible to design a new research and focus specifically on these subjects. It would involve more extensive and more expensive research than this thesis, but with the weather measurements of different local climates, it would be possible to answer the last part of the research question for this thesis in a better way: what are the effects and extent of outdoor climatic factors? With the collected data from such a research study, multiple regression analyses could be executed. These data could be analysed to see whether there is a specific relationship. This is just one of the options to create a more in-depth research study on the relationship between climatic factors.

In the chapter: results, three main climatic factors are discussed: wind, irradiation and rain. Wind and irradiation are more or less discussed on a more brother level with the data available. The influence of the factor rain was much harder to investigate without the local data. Even with the data from the KNMI it is not easy to draw conclusions of the effects of this factor on the local climate. The literature study shortly described another factor that may can explain the differences of gas use through the local climate and that is the factor humidity. As discussed in the section: processing the data, most of the Netatmo stations also measure the humidity. In a follow up research the factor humidity could be investigated and can yet explain the influence of the factor rain/humidity.

5.3.3 Cooling energy and alternative energy sources

In the literature study from this thesis several scientific papers came across where besides the gas consumption also the electricity use is mentioned in combination with degree days. This thesis only focused on the heating degree days and how this effected the gas consumption, because the use of gas is still the most common way to heat residential buildings in the Netherlands. The type of energy consumption used to cool a residential buildings is by the use of air conditioning systems. In this case the cooling degree days will be taken into account. To create a complete image of the energy consumption and the effects of the local climate the electricity used for climate control in residential buildings need to take into account. In a continuation of this research both gas and electricity consumption need to be investigated. The thesis also focused on the residential buildings on a PC6 level. This created null-values in maps and didn't create the complete picture of the whole energy consumption on the scale of the local climate. It would also advisable to research the office buildings and business sites in a follow up research. Data from gas and electricity for both residential en businesses are provided by the CBS on a PC6 level.

This thesis has discussed gas use consumption per household at the PC6 level. Gas consumption is still the most significant type of heating energy in the Netherlands (CBS, 2019). But with the energy transition, it is also a hot topic in the Netherlands and the percentage of gas consumption will decrease over time. Because gas consumption still represents a big part of energy consumption for heating, this thesis is still relevant, but in the future other energy sources need to be taken into account. In the chapter; results, the irradiation all ready came across. The original use for the dataset from overheid.nl (2010) was created to show the potential for solar panels on rooftops. Besides solar energy the use of geothermal energy is upcoming. with this technique energy is gained from the temperature differences between the earths surface and heat reservoirs located deeper in the earth. It could be intersting to investigaye how the local climate or UHI's influence these types of energy generation. In the future more of these types of energy sources will be used. Also the thesis is focuses on household heating energy. In the theory chapter, cooling energy was discussed. With cooling energy, cooling degree days are significant. In further research, this type of energy and cooling degree days should be investigated.

References

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*, 1192–1205.

Aydin, E., & Brounen, D. (2019). The impact of policy on residential energy consumption. *Energy*, *169*, 115–129.

Britter, R., & Hanna, S. (2003). Flow and dispersion in urban areas. *Annual review of fluid mechanics*, *35*(1), 469–496.

CBS. (2017). *Aardgasbalans; aanbod en verbruik.* Retrieved from https://opendata.cbs.nl/#/CBS/nl/dataset/00372/table?ts=1622211594873

CBS. (2018). *Energiematrix den haag.* Retrieved from https://www.cbs.nl/nl-nl/maatwerk/2020/51/energiematrix-den-haag-2018

CBS. (2019). *aardgas verbruik.* Retrieved from https://www.cbs.nl/nl-nl/nieuws/2020/20/minder-steenkool-en-meer-aardgas-verbruikt-in-2019

CBS. (2020). *Energielevering aan woningen en bedrijven naar postcode.* Retrieved from https://www.cbs.nl/nl-nl/maatwerk/2020/33/energielevering-aan-woningen-en-bedrijven-naar-postcode

Clarke, J. A., & Clarke, J. A. (2001). *Energy simulation in building design.* Routledge.

Coseo, P., & Larsen, L. (2014). How factors of land use/land cover, building configuration, and adjacent heat sources and sinks explain urban heat islands in chicago. *Landscape and Urban Planning*, *125*, 117–129.

D'Amico, A., Ciulla, G., Panno, D., & Ferrari, S. (2019). Building energy demand assessment through heating degree days: The importance of a climatic dataset. *Applied energy*, *242*, 1285–1306.

Day, A., Knight, I., Dunn, G., & Gaddas, R. (2003). Improved methods for evaluating base temperature for use in building energy performance lines. *Building Services Engineering Research and Technology*, *24*(4), 221–228.

de Nijs, T., Bosch, P., Brand, E., Heusinkveld, B., van der Hoeven, F., Jacobs, C., . . . others (2019). *Ontwikkeling standaard stresstest hitte* (Tech. Rep.). Rijksinstituut voor Volksgezondheid en Milieu.

Energie in Nederland. (2018). *Infographic energie in nederland.* Retrieved from https://www.ebn.nl/wp-content/uploads/2018/01/EBN-Infographic-2018-pdf.pdf

Energie in Nederland. (2019). *Infographic energie in nederland.* Retrieved from https://www.ebn.nl/wp-content/uploads/2019/01/EBN_Infographic2019_14JAN19.pdf

energieatlas. (2020). *De rijksoverheid.* Retrieved from https://www.nationaleenergieatlas.nl/kaarten

energieleveranciers.nl. (n.d.). https://www.energieleveranciers.nl/netbeheerders/overzicht-netbeheerders. (Accessed: 2020-11-01)

EPA. (2008). *United states environmental protection agency.* Retrieved from https://www.epa.gov/heatislands/heat-island-compendium

Gemeente Den Haag. (2020). *Energietransitie, den haag in cijfers.* Retrieved from https://denhaag.incijfers.nl/dashboard/Overzichten/Energietransitie/

GISGeography. (2020). *Inverse distance weighting (idw) interpolation.* Retrieved from https://gisgeography.com/inverse-distance-weighting-idw-interpolation/

Guerra-Santin, O., & Itard, L. (2010). Occupants' behaviour: determinants and effects on residential heating consumption. *Building Research & Information*, *38*(3), 318–338.

Holmes, C., Tett, S., & Butler, A. (2017). What is the uncertainty in degree-day projections due to different calibration methodologies? *Journal of Climate*, *30*(22), 9059–9075.

Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., & Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, *45*, 141–151.

Huang, Y., Ritschard, R., Bull, J., & Chang, L. (1986). *Climatic indicators for estimating residential heating and cooling loads* (Tech. Rep.). Lawrence Berkeley Lab., CA (USA).

Ihara, T., Genchi, Y., Sato, T., Yamaguchi, K., & Endo, Y. (2008). City-block-scale sensitivity of electricity consumption to air temperature and air humidity in business districts of tokyo, japan. *Energy*, *33*(11), 1634–1645.

IPPC. (2021). *Intergovernmental panel on climate change.* Retrieved from https://www.ipcc.ch/

Jenerette, G. D., Harlan, S. L., Brazel, A., Jones, N., Larsen, L., & Stefanov, W. L. (2007). Regional relationships between surface temperature, vegetation, and human settlement in a rapidly urbanizing ecosystem. *Landscape ecology*, *22*(3), 353–365.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., & Djurovic-Petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and environment*, *45*(7), 1683–1697.

KNMI. (2010). *Warmte-eilandeffect van de stad utrecht.* Retrieved from https://www.knmi.nl/kennis-en-datacentrum/achtergrond/warmte-eilandeffect-van-de-stad-utrecht

KNMI. (2014). *Waarnemingen klimaatveranderingen.* Retrieved from https://www.knmi.nl/kennis-en-datacentrum/achtergrond/waarnemingen-klimaatveranderingen

KNMI. (2020). *Daggegevens van het weer in nederland.* Retrieved from https://www.knmi.nl/nederland-nu/klimatologie/daggegevens

Kohler, M., Blond, N., & Clappier, A. (2016). A city scale degree-day method to assess building space heating energy demands in strasbourg eurometropolis (france). *Applied energy*, *184*, 40–54.

Krese, G., Prek, M., & Butala, V. (2012). Analysis of building electric energy consumption data using an improved cooling degree day method. *Strojniški vestnik-Journal of Mechanical Engineering*, *58*(2), 107–114.

Lazos, D., Sproul, A. B., & Kay, M. (2014). Optimisation of energy management in commercial buildings with weather forecasting inputs: A review. *Renewable and Sustainable Energy Reviews*, *39*, 587–603.

Li, D. H., Yang, L., & Lam, J. C. (2012). Impact of climate change on energy use in the built environment in different climate zones–a review. *Energy*, *42*(1), 103–112.

Lindelöf, D. (2017). Bayesian estimation of a building's base temperature for the calculation of heating degree-days. *Energy and buildings*, *134*, 154–161.

Lo, C., & Quattrochi, D. A. (2003). Land-use and land-cover change, urban heat island phenomenon, and health implications. *Photogrammetric Engineering & Remote Sensing*, *69*(9), 1053–1063.

Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, B. V. (2014). 4th generation district heating (4gdh): Integrating smart thermal grids into future sustainable energy systems. *Energy*, *68*, 1–11.

Majcen, D., Itard, L., & Visscher, H. (2013). Theoretical vs. actual energy consumption of labelled dwellings in the netherlands: Discrepancies and policy implications. *Energy policy*, *54*, 125–136.

Marjaba, G., Chidiac, S., & Kubursi, A. (2020). Sustainability framework for buildings via data analytics. *Building and Environment*, *172*, 106730.

Mashhoodi, B. (2019). The spatial dimension of household energy consumption. *A+ BE—Architecture and the Built Environment*(5), 1–244.

Mashhoodi, B., Stead, D., & van Timmeren, A. (2019). Local and national determinants of household energy consumption in the netherlands. *GeoJournal*, 1–14.

Meng, Q., & Mourshed, M. (2017). Degree-day based non-domestic building energy analytics and modelling should use building and type specific base temperatures. *Energy and Buildings*, *155*, 260–268.

Milieudefensie. (2019). *duurzaam wonen in vier stappen.* Retrieved from https://milieudefensie.nl/

Mindergas. (2020). *Gewogen graaddagen calculator.* Retrieved from https://www.mindergas.nl/degree_days_calculation/new

mindergas. (2021). *zoncalculator.* Retrieved from https://www.zonurencalculator.nl/sun_hours_calculation

Netatmo. (2020). *Netatmo.weathermap.* Retrieved from https://weathermap.netatmo.com/

NGR. (2020). *Nationaal georegister, de rijksoverheid.* Retrieved from https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search;jsessionid=1E3E44AACB574D8092882B2F8241242A#/home

OpenEnergyMonitor. (2019). *Building energy model.* Retrieved from https://learn.openenergymonitor.org/sustainable-energy/building-energy-model/readme

overheid.nl. (2010). *Zonne instraling per dakoppervlak den haag.* Retrieved from https://data.overheid.nl/dataset/zonne-instraling-per-dakoppervlak-den-haag

Pijpers-van Esch, M. (2015). Designing the urban microclimate. *A+ BE— Architecture and the Built Environment*(6), 1–308.

Popovski, E., Fleiter, T., Santos, H., Leal, V., & Fernandes, E. O. (2018). Technical and economic feasibility of sustainable heating and cooling supply options in southern european municipalities-a case study for matosinhos, portugal. *Energy*, *153*, 311–323.

Publieke Dienstverlening Op de Kaart. (2020). *De rijksoverheid.* Retrieved from https://www.pdok.nl/

QuantumGIS. (2021). *Qgis.* Retrieved from https://qgis.org/nl/site/

Rijksoverheid. (2019). *Nationaal programma regionale energiestrategie.* Retrieved from https://regionale-energiestrategie.nl/default.aspx

Rijksoverheid. (2020). *energielabel voor woningen.* Retrieved from https://www.energielabelvoorwoningen.nl/?cookieCheck=true

Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A., & Pendyala, R. M. (2017). Machine learning approaches for estimating commercial building energy consumption. *Applied energy*, *208*, 889–904.

SafeSoftware. (2021). *Fme.* Retrieved from https://www.safe.com/fme/

Santamouris, M., Cartalis, C., Synnefa, A., & Kolokotsa, D. (2015). On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—a review. *Energy and Buildings*, *98*, 119–124.

Schroedter-Homscheidt, M., et al. (2016). User's guide to the cams radiation service-status december 2016. *Shinfield Park: ECMWF*.

Sen, S., & Roesler, J. (2020). Wind direction and cool surface strategies on microscale urban heat island. *Urban Climate*, *31*, 100548.

Shin, M., & Do, S. L. (2016). Prediction of cooling energy use in buildings using an enthalpy-based cooling degree days method in a hot and humid climate. *Energy and Buildings*, *110*, 57–70.

StatisticsSolution. (2021). *Paired sample t-test.* Retrieved from https://www.statisticssolutions.com/manova-analysis-paired-sample-t-test/

Statline, C. (2020). *statline.* Retrieved from https://opendata.cbs.nl/#/CBS/nl/dataset/84575NED/table?ts=1613592182731

Stedin. (2020). *verbruiksgegevens.* Retrieved from https://www.stedin.net/zakelijk/open-data/verbruiksgegevens

Stone, B., & Norman, J. M. (2006). Land use planning and surface heat island formation: A parcel-based radiation flux approach. *Atmospheric Environment*, *40*(19), 3561–3573.

Stone Jr, B. (2012). *The city and the coming climate: Climate change in the places we live.* Cambridge University Press.

Storey, J. (2012). Sustainable urban environments: An ecosystem approach. *Smart and Sustainable Built Environment*.

TUDelft3D. (2021). *3dbag.* Retrieved from https://3dbag.nl/en/viewer

UnitedNations. (2015). *The paris agreement.*

van den Brom, P., Meijer, A., & Visscher, H. (2018). Performance gaps in energy consumption:

household groups and building characteristics. *Building Research & Information*, *46*(1), 54–70.

van het Schip, E. (2010). *zonnestralingsintensiteit.* Retrieved from http://wiki.bk.tudelft.nl/mw_bk-wiki/index.php?title=Zonnestralingsintensiteit&direction=prev&oldid=3260

Visscher, H. (2012). *Sustainable urban environments: An ecosystem approach* (E. van Bueren, H. van Bohemen, & L. Itard, Eds.). Springer. doi: 10.1007/978-94-007-1294-2

Wetterzentrale. (2021). *Knmi data gevisualiseerd.* Retrieved from https://www.wetterzentrale.de/averages.php?station=344&jaar=2018&maand=1&dag=1&jaar2=2018&maand2=12&dag2=31

Wever, N. (2008). Effectieve temperatuur en graaddagen. *Klimatologie en klimaatscenario's. KNMI publicatie*, *219*.

Wunderground. (2021). *Weather underground.* Retrieved from https://www.wunderground.com/

Zhang, W., Robinson, C., Guhathakurta, S., Garikapati, V. M., Dilkina, B., Brown, M. A., & Pendyala, R. M. (2018). Estimating residential energy consumption in metropolitan areas: A microsimulation approach. *Energy*, *155*, 162–173.

5.4 Appendix 1: Main script

This Python script is explained in section 3.8: Processing of the data. In this script the Netatmo txt files are transformed in CSV files per day for the years 2018 and 2019.

```
#[illegible]

import os
import csv
import pandas as pd
from datetime import datetime
import calendar
import sys

# [illegible]
class weatherStationData:
    def __init__(self,id,longitude,latitude,altitude,temperature,humidity,rain_60min,rain_24h,rain_live,pressure,datetime):
        self.id = id
        self.longitude = longitude
        self.latitude = latitude
        self.altitude = altitude
        self.temperature = temperature
        self.humidity = humidity
        self.rain_60min = rain_60min
        self.rain_24h = rain_24h
        self.rain_live = rain_live
        self.pressure = pressure
        self.datetime = datetime

    def to_dict(self):
        return {
            'id': self.id,
            'longitude': self.longitude,
            'latitude': self.latitude,
            'altitude': self.altitude,
            'temperature': self.temperature,
            'humidity': self.humidity,
            'rain_60min': self.rain_60min,
            'rain_24h': self.rain_24h,
            'rain_live': self.rain_live,
            'pressure': self.pressure,
            'datetime': self.datetime,
        }
def csvWriter(inputDataFrame, OutputDirectory, loopYear, loopMonth, loopDay):
    if not inputDataFrame.empty:
        inputDataFrame.to_csv(OutputDirectory + "[illegible]" + str(loopYear) + "[illegible]" + str(loopMonth)
                              + "[illegible]" + str(loopDay) + "[illegible]" + str(loopYear) + ".csv", index=False, sep=';')

def writeDataToCsv(loopYear, inputMonths2020, inputMonths, weatherstationDataDF):
    #[illegible]
    if loopYear == 2020:
        #[illegible]
        lastDayOfOctober = 21
        for loopMonth in inputMonths2020:
            if loopMonth == 10:
                for loopDay in range(1, lastDayOfOctober):
                    csvWriter(weatherstationDataDF.loc[(weatherstationDataDF['datetime'].dt.day==loopDay) &
                                                       (weatherstationDataDF['datetime'].dt.month==loopMonth) &
                                                       (weatherstationDataDF['datetime'].dt.year==loopYear)],
                              OutputDirectory, loopYear, loopMonth, loopDay)
            else:
                for loopDay in range(1, calendar.monthrange(loopYear, loopMonth)[1] + 1):
                    csvWriter(weatherstationDataDF.loc[(weatherstationDataDF['datetime'].dt.day==loopDay) &
```

```
                                                                                (weatherStationDataDF['datetime'].dt.month==loopMonth) &
                                                                                (weatherStationDataDF['datetime'].dt.year==loopYear)],
                                                  OutputDirectory, loopYear, loopMonth, loopDay)

        else:
            for loopMonth in inputMonths:
                for loopDay in range(1, calendar.monthrange(loopYear, loopMonth)[1] + 1):
                    csvWriter(weatherStationDataDF.loc[(weatherStationDataDF['datetime'].dt.day==loopDay) &
                                                        (weatherStationDataDF['datetime'].dt.month==loopMonth) &
                                                        (weatherStationDataDF['datetime'].dt.year==loopYear)],
                                              OutputDirectory, loopYear, loopMonth, loopDay)


inputYears = [2015, 2016, 2017, 2018, 2019, 2020]
inputMonths = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
inputMonths2020 = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

InputDirectory = "E:\\temp\\d-m\\weather-station_data"
OutputDirectory = "E:\\temp\\d-m\\output"
Errors = []

#Loop through the selected years
for loopYear in inputYears:

    #Reinitialize the array of objects for each year to save memory
    weatherStationDataArray = []

    for subdir, dirs, files in os.walk(InputDirectory):
        for file in files:
            if (file.endswith(".txt") and  file[0:4] == str(loopYear)):
                pathName = subdir + "\\" + file
                fileDatetime = datetime.strptime(file[0:20].strip(), '%Y-%m-%d %H_%M_%S')
                file = open(os.path.join(pathName), "r")
                reader = csv.reader(file, delimiter=',')
                counter = 0
                print(pathName)
                try:
                    for row in reader:
                        counter = counter + 1
                        if counter == 1:
                            continue
                        else:
                            try:
                                weatherStationDataArray.append(weatherStationData(row[0], row[1], row[2], row[3],
                                                                                  row[4], row[5], row[6], row[7], row[8], row[9],
                                                                                  fileDatetime))

                            except:
                                Errors.append(str(file) + " Column Error") #Add file with the error to the errors list
                except:
                    Errors.append(str(file) + " Row Error") #Add file with the error to the errors list

    weatherStationDataDF = pd.DataFrame.from_records([weatherData.to_dict() for weatherData in weatherStationDataArray])
    writeDataToCsv(loopYear, inputMonths2020, inputMonths, weatherStationDataDF)


for error in Errors:
    print(error)
sys.exit(0)
```

```
sys.exit(0)

for subdir, dirs, files in os.walk(InputDirectory):
    for file in files:
        if file.endswith(".txt"):
            pathName = subdir + "\\" + file
            fileDatetime = datetime.strptime(file[0:20].strip(), '%Y-%m-%d %H_%M_%S')
            file = open(os.path.join(pathName), "r")
            reader = csv.reader(file, delimiter=',')
            counter = 0
            print(pathName)
            try:
                for row in reader:
                    counter = counter + 1
                    if counter == 1:
                        continue
                    else:
                        try:
                            weatherStationDataArray.append(weatherStationData(row[0], row[1], row[2], row[3], row[4], row[5],
                                                                              row[6], row[7], row[8], row[9], fileDatetime))
                        except:
                            Errors.append(str(file) + " Column Error") #Add file with the error to the errors list
            except:
                Errors.append(str(file) + " Row Error") #Add file with the error to the errors list
        else:
            continue
```

5.5 Appendix 2: Analyse script

This Python script is explained in section 3.8: Processing of the data. In this script the Netatmo CSV files are analysed and the degree days per station per day are calculated.

```
# [illegible]

print("Starting script...")

import os
import csv
import pandas as pd
from datetime import datetime
import calendar

InputDirectory = "D:\\Temp\\Input"
OutputDirectory = "D:\\Temp\\Output"


# [illegible]
class weatherStationData:
    def __init__(self, id, longitude, latitude, altitude, temperature, humidity, rain_60min,
                 rain_24h, rain_live, pressure, datetime):
        self.id = id
        self.longitude = longitude
        self.latitude = latitude
        self.altitude = altitude
        self.temperature = temperature
        self.humidity = humidity
        self.rain_60min = rain_60min
        self.rain_24h = rain_24h
        self.rain_live = rain_live
        self.pressure = pressure
        self.datetime = datetime

    def to_dict(self):
        return {
            'id' : self.id,
            'longitude' : self.longitude,
            'latitude' : self.latitude,
            'altitude' : self.altitude,
            'temperature' : self.temperature,
            'humidity' : self.humidity,
            'rain_60min' : self.rain_60min,
            'rain_24h' : self.rain_24h,
            'rain_live' : self.rain_live,
            'pressure' : self.pressure,
            'datetime' : self.datetime,
        }
def csvWriter(InputDataFrame, OutputDirectory, loopYear, loopMonth, loopDay):
    if not InputDataFrame.empty:
        InputDataFrame.to_csv(OutputDirectory + "\\" + str(loopYear) + "\\" + str(loopMonth) + "\\" +
                              str(loopDay) + "-" + str(loopYear) + ".csv", sep=';', decimal=",")

def degreedays(temperatureMean):
    if temperatureMean >= 18: return 0
    elif temperatureMean < 18: return 18 - temperatureMean
    else: return 0

InputYears = [2018, 2019, 2020]
InputMonths = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
InputMonths2020 = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

```
Errors = []

# loop through the specified years
for loopYear in inputYears:
    print(loopYear)

    # recreate the array of objects for each year to save memory
    weatherStationDataArray = []

    for subdir, dirs, files in os.walk(inputDirectory):
        for file in files:
            print(file)
            if (file.endswith(".csv") and ((file[4:8] == str(loopYear)) or (file[5:9] == str(loopYear)) or
                                           (file[6:10] == str(loopYear)))):
                pathName = subdir + "\\" + file
                file = open(os.path.join(pathName), "r")
                reader = csv.reader(file, delimiter=";")
                counter = 0
                print(pathName)
                try:
                    for row in reader:
                        counter = counter + 1
                        if counter == 1:
                            continue
                        else:
                            try:
                                weatherStationDataArray.append(
                                    weatherStationData(row[0], row[1], row[2], row[3], row[4], row[5],
                                                       row[6], row[7], row[8], row[9], row[10]))
                            except:
                                Errors.append(str(file) + " Column Error")  # add file with the error to the errors list
                except:
                    Errors.append(str(file) + " Row Error")  # add file with the error to the errors list

    weatherStationDataDf = pd.DataFrame.from_records([weatherData.to_dict() for weatherData
                                                     in weatherStationDataArray])
    weatherStationDataDf['datetime'] = pd.to_datetime(weatherStationDataDf['datetime'])
    weatherStationDataDf['longitude'] = pd.to_numeric(weatherStationDataDf['longitude'])
    weatherStationDataDf['latitude'] = pd.to_numeric(weatherStationDataDf['latitude'])
    weatherStationDataDf['altitude'] = pd.to_numeric(weatherStationDataDf['altitude'])
    weatherStationDataDf['temperature'] = pd.to_numeric(weatherStationDataDf['temperature'])
    weatherStationDataDf['humidity'] = pd.to_numeric(weatherStationDataDf['humidity'])
    weatherStationDataDf['rain_##max'] = pd.to_numeric(weatherStationDataDf['rain_##max'])
    weatherStationDataDf['rain_24h'] = pd.to_numeric(weatherStationDataDf['rain_24h'])
    weatherStationDataDf['rain_live'] = pd.to_numeric(weatherStationDataDf['rain_live'])
    weatherStationDataDf['pressure'] = pd.to_numeric(weatherStationDataDf['pressure'])

    # since we have data until 2020 we need to cap the loop
    if loopYear == 2020:
        # since we have data until 22 10 we need to cap the loop
        lastDayOfOctober = 22
        for loopMonth in inputMonths2020:
            if loopMonth == 10:
                for loopDay in range(1, lastDayOfOctober):
                    dfToCheck = weatherStationDataDf.loc[
                        (weatherStationDataDf['datetime'].dt.day == loopDay) &
                        (weatherStationDataDf['datetime'].dt.month == loopMonth) &
                        (weatherStationDataDf['datetime'].dt.year == loopYear)
```

```
                    ].groupby('id').agg({'longitude': 'max',
                                         'latitude': 'max',
                                         'altitude': 'max',
                                         'temperature': 'mean',
                                         'humidity': 'mean',
                                         'rain_60min': 'mean',
                                         'rain_24h': 'mean',
                                         'rain_1live': 'mean',
                                         'pressure': 'mean'})

                if not dfToCheck.empty:
                    dfToCheck['degree_days'] = dfToCheck.apply(lambda x: degreedays(x['temperature']), axis=1)
                    dfToCheck.round(2)
                    csvwriter(dfToCheck, OutputDirectory, loopYear, loopMonth, loopDay)
        else:
            for loopDay in range(1, calendar.monthrange(loopYear, loopMonth)[1] + 1):
                dfToCheck = weatherstationDataDf.loc[
                    (weatherstationDataDf['datetime'].dt.day == loopDay) &
                    (weatherstationDataDf['datetime'].dt.month == loopMonth) &
                    (weatherstationDataDf['datetime'].dt.year == loopYear)
                    ].groupby('id').agg({'longitude': 'max',
                                         'latitude': 'max',
                                         'altitude': 'max',
                                         'temperature': 'mean',
                                         'humidity': 'mean',
                                         'rain_60min': 'mean',
                                         'rain_24h': 'mean',
                                         'rain_1live': 'mean',
                                         'pressure': 'mean'})
                if not dfToCheck.empty:
                    dfToCheck['degree_days'] = dfToCheck.apply(lambda x: degreedays(x['temperature']), axis=1)
                    dfToCheck.round(2)
                    csvwriter(dfToCheck, OutputDirectory, loopYear, loopMonth, loopDay)
    else:
        for loopMonth in inputMonths:
            for loopDay in range(1, calendar.monthrange(loopYear, loopMonth)[1] + 1):
                dfToCheck = weatherstationDataDf.loc[
                    (weatherstationDataDf['datetime'].dt.day == loopDay) &
                    (weatherstationDataDf['datetime'].dt.month == loopMonth) &
                    (weatherstationDataDf['datetime'].dt.year == loopYear)
                    ].groupby('id').agg({'longitude': 'max',
                                         'latitude': 'max',
                                         'altitude': 'max',
                                         'temperature': 'mean',
                                         'humidity': 'mean',
                                         'rain_60min': 'mean',
                                         'rain_24h': 'mean',
                                         'rain_1live': 'mean',
                                         'pressure': 'mean'})
                if not dfToCheck.empty:
                    dfToCheck['degree_days'] = dfToCheck.apply(lambda x: degreedays(x['temperature']), axis=1)
                    dfToCheck.round(2)
                    csvwriter(dfToCheck, OutputDirectory, loopYear, loopMonth, loopDay)
```