MASTER

Adoption of photovoltaic panels across neighbourhoods in the Netherlands
A spatial analysis

Tijsse-Klasen, S.J.J.

Award date:
2021

Eindhoven University of Technology
Department of the Built Environment
Architecture, Building and Planning
Track: Urban Systems and Real Estate

Adoption of photovoltaic panels across neighbourhoods in the Netherlands

A spatial analysis

Thesis (45 ECTS)
7Z45M0

S.J.J. Tijsse Klasen

studentnr: 1393901

Supervisors:
ir. A.W.J. Borgers
dr. G. Dane
dr. ing. P.J.H.J. van der Waerden

version 1.0

This Master thesis has been carried out in accordance with the rules of the TU/e Code of Scientific Integrity

Eindhoven,
Tuesday, September 30, 2021

Abstract

An important problem the world is facing these days are man-made greenhouse gases (GHG). Man-made GHG are produced, among others, by the use of energy that is produced with fossil fuels. Renewable energy sources, such as solar energy, play an important role in this transition from non-renewable sources to renewable sources. The spatial importance of energy generation from the sun has been identified by several researchers. However, in the existing literature, the spatial component of residential technology diffusion is mainly taken into account at the global scale, rather than the local scale. For that reason, the contribution of this research was to add a spatial dimension to the analysis and draw conclusions at the local scale. This research aimed to establish a spatial analysis of influential factors in the adoption of PV systems in the Netherlands because this is the energy source with the most potential compared to other renewable energy sources in the residential sector. The results of this research can be used for scientific purposes, policy-making and marketing.

First, an explorative literature review on sustainability, solar energy and adoption mechanisms introduced the subject. Then literature was reviewed to determine influential factors on solar panel adoption. These factors can be categorized into the following categories: sociodemographic, socioeconomic, built environment, environmental awareness and peer effects. Some important factors are household wealth, age, environmental concern and governmental incentives. The analysis has been conducted with data from Statistics Netherlands and SolarGIS. An Ordinary Least Square (OLS) regression, together with a Geographically Weighted Regression (GWR) has been used to explore relations between potential influential factors and the adoption of solar panels. A GWR creates a local regression for each neighbourhood in the dataset. From the GWR, it was observed that there is a cluster of relatively high coefficients in Groningen, Noord Holland and the southern part of Limburg for the people aged between 45 and 64. For the same variable, low coefficients are more dispersed over the country however Zeeland and parts of Noord-Brabant, Gelderland, Drenthe, Overijssel and Friesland have the lowest coefficients. Both the variables 'moderate' and 'strong degree of urbanity' had similar spatial distribution of coefficients. With a centering of low coefficients in Utrecht and surrounded by higher values with the highest coefficients in Zeeland and Groningen.

Some unexpected things came to light during this research. First, essential variables were missing, such as income, governmental incentives and environmental concern. Secondly, residuals were spatially clustered and local collinearity occurred, which makes the result biased. Thirdly, some contrary findings with the literature were found and fourthly the adjusted R^2 was relatively low in comparison with comparable studies.

Some recommendations were formulated for further research and other practical purposes. Firstly, policies and marketing should be more on a local level rather than a national one, because there are differences between areas when it comes to the determinants for the adoption of PV panels. Secondly, it is recommended to find ways to improve the quality of this data. Thirdly, future research needs to be done on these indirect and direct relations with the PV panel adoption. Fourthly, it is highly recommended to investigate other techniques.

Preface

Dear reader,

Before you lies the final product of my graduation project on influential factors for photovoltaic panels. This thesis was written as the final test for the master Architecture, Building and Planning, with a specialization in Urban Systems and Real Estate. During the master, I gained a lot of interest in sustainable energy supply, big data and relatively new methodologies. All these facets I tried to combine into one research.

At this point, I want to thank several individuals for participating and supporting me during this period. First of all, I want to thank Gamze Dane for her daily supervision and critical thinking. Secondly, I want to thank Aloys Borgers and Peter van der Waerden for their support and quick replies to emails. For the remainder, I want to thank my family and friends for their emotional support during the period of writing my thesis. It has been a strange experience, mostly due to the COVID-19 pandemic. Family and friends supported me in distraction and relaxation in these boring times.

I hope that you enjoy reading this thesis!

Wijchen, September 2021

Stan Tijsse Klasen

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and background information

An important problem the world is facing these days are man-made greenhouse gases (GHG). Man-made GHG are produced, among others, by the use of energy that is produced with fossil fuels. The supply of energy has to transform into renewable energy solutions. There are different kinds of renewable energy, such as hydropower-, solar-, wind-, biomass-, geothermal- and ocean energy (Bose, 2010). According to Hoogwijk and Graus (2008), solar energy is the largest energy source followed by wind and ocean energy. The production of energy from renewable energy sources is an important factor in the 'World Energy Trilemma' (Austin, 2016). According to several studies, energy consumption plays an important role in preventing global warming. Kaygusuz (2002) argues the importance of energy in improving the living standard and retaining economic growth. Bose (2010) states that energy is the lifeblood of the evolution of national prosperity and industrial civilization and Dincer (2000) argues the increase in energy demand because of the increase in world population, consumption and industrial activity. He also states that solar energy plays an important role in reducing the energy consumption of non-renewable resources (Dincer, 2000). However, according to Unruh (2002), the implementation of renewable energy proceeds slowly. This is due to the fact that it has to compete with an energy supply that is completely dependent on fossil fuels. Together with several years of experience, low costs, high efficiency and other factors that come with this experience (Unruh, 2002).

According to Georg et al. (2018), solar energy capacity can be divided into different categories. The largest category includes installations that produce less than 10 kWp, this category mainly includes private residences. This category produces approximately 80% of the total capacity, see table 1.1 for the other categories. According to Statistics Netherlands (2020b), in 2018 there are approximately 700,000 photovoltaic installations installed on residences that produce more than 2 billion kW. Statistics show that this was only 1.3 billion kW in 2016 with about 400,000 systems and in 2012 this was only about 180 thousand kW from 70,000 systems. Thus, from these statistics, it can be concluded that the adoption of PV panels has increased significantly (Statistics Netherlands, 2020b). In figure 1.1 the increase in installed PV panels and their capacity in the residential sector of the Netherlands can be seen. Due to this increase in the adoption of PV panels, the factors that influence this adoption gained attention from researchers to identify which factors have a strong influence and which less. From there, targeted interventions can then take place.

An important characteristic of PV panel adoption is the spatial component in it. From figure 1.2, it can be concluded that the distribution of PV panels in the Netherlands is not homogeneous, through this research, factors influencing this spatial diffusion will be tested. The first law of geography sounds as follows: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). This spatial importance has been identified by several researchers (Balta-Ozkan et al., 2015; Hofierka et al., 2014; Schaffer & Brun, 2015; Snape, 2016).

Table 1.1: Percentage of installation capacity per category (Georg et al., 2018)

Category	e.g.	Percentage of total capacity (%)
<10 kWp	Private residences	79.1
10-100 kWp	Multifamily, commercial and tertiary-building sector	3.1
100-500 kWp	Industrial buildings, commercial centers, schools, etc.	10.9
>500 kWp	Industries and ground-mounted solar PV systems	6.0

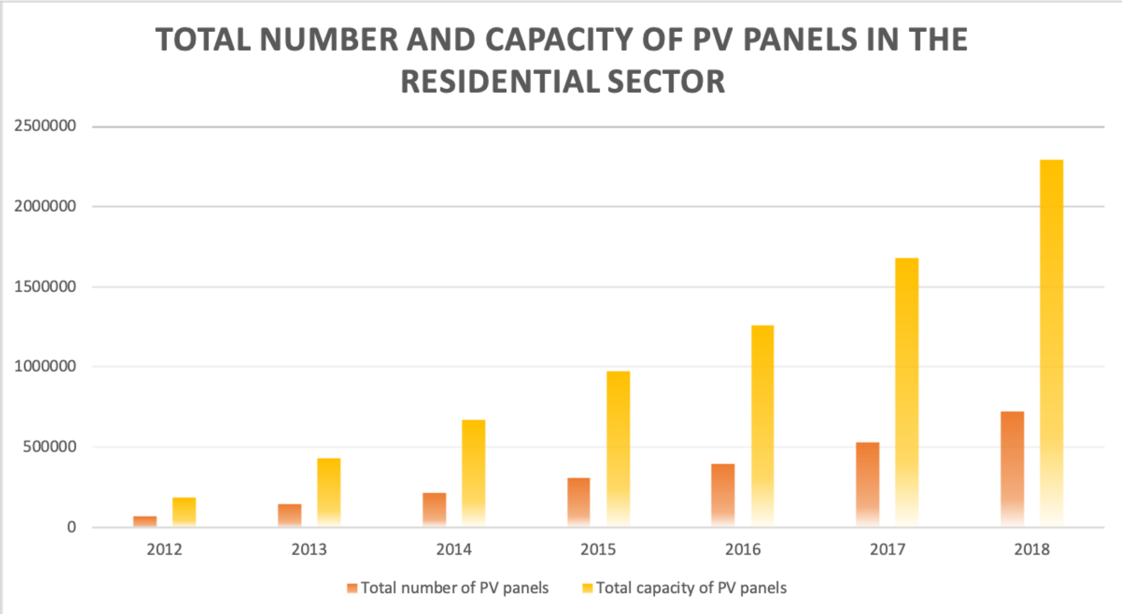

Figure 1.1: Installed PV panels in the residential sector (Statistics Netherlands, 2020b)

Another factor that shows the importance of the spatial component in PV adoption are the spatial neighbour effects (also known as 'peer effects'). Graziano and Gillingham (2015) found a strong significant positive relation between PV adoption and the number of earlier installed PV panels in the surrounding area. They also indicate the spatial importance of PV panel adoption by identifying spatial patterns and diffusion. These peer effects are influenced by social interaction and visibility and diminish over space and time. According to their study the relation between adoption and spatial neighbour effects can be explained as follows; adding an installation within 0.5 miles of adopters in the year prior to the adoption increases the number of installations by 0.44 PV panels on average (Graziano & Gillingham, 2015). Rode and Weber (2016) describe this phenomenon in their study and state that the visibility of PV panels might correlate with adoption in the surrounding. Jager (2006) also describes this as follows; "Consumers frequently feel satisfied when consuming the same as their neighbours (social needs) and often engage in social comparison and imitation when deciding what to consume" (Jager, 2006). Bollinger and Gillingham (2012) find almost the same results as Graziano and Gillingham (2015), they find that the probability of adoption increases by 0.78 percentage points when adding a PV installation within the zip code. However, in their study they address three issues when drawing conclusions for these peer effects, those are; homophily, simultaneity and correlated unobservables. Homophily means that results can be influenced by the fact that people with similar interests and characteristics live in the same environment. Thus, in the case of peer effects, this may also mean that these pre-installed installations are a consequence of similarities between people. Simultaneity means that

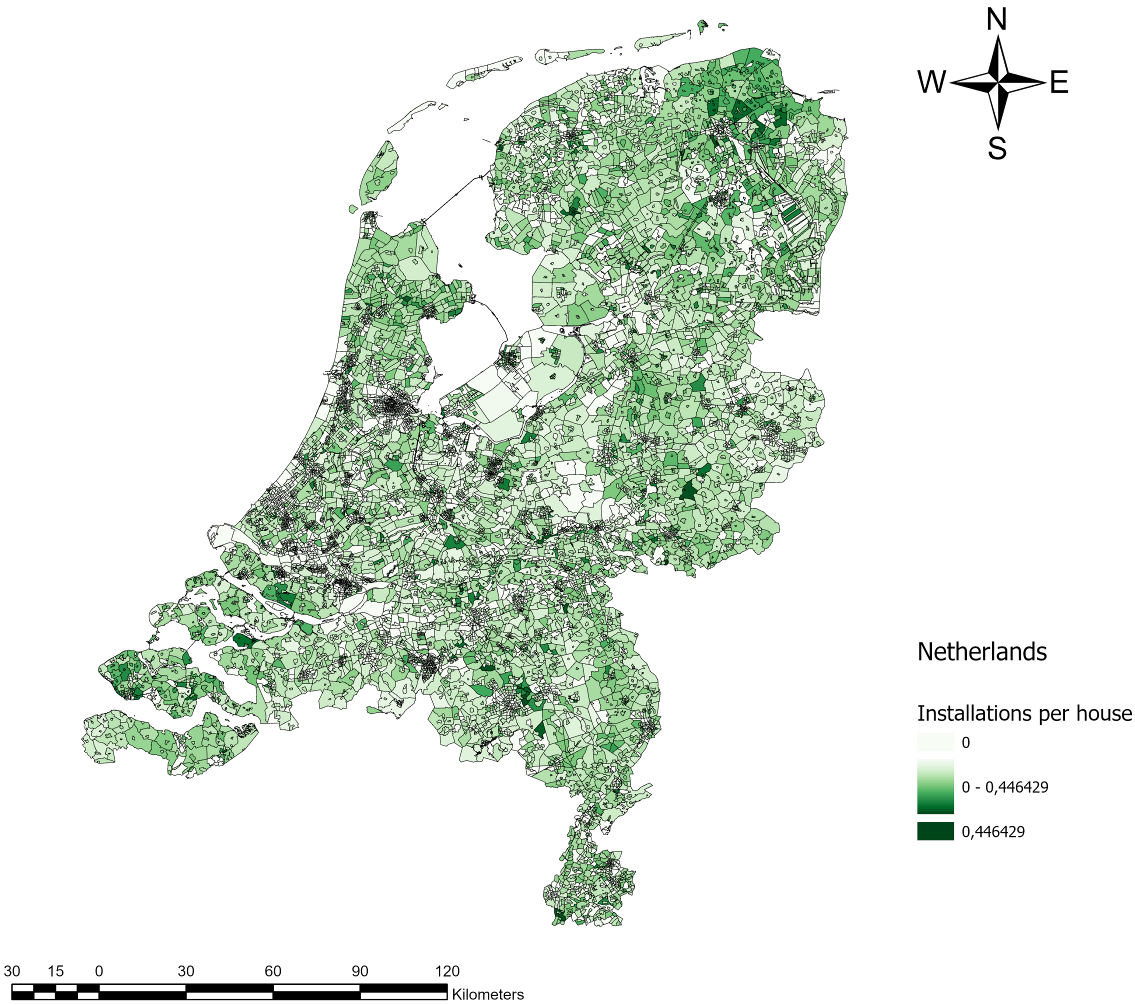

Figure 1.2: Number of PV installations per house in 2018 (Statistics Netherlands, 2020c)

people are influenced by peers, but at the same time, the people influence their peers as well. Correlated unobservable factors mean that, for example, local campaigns or promotions have been implemented that are not known (Bollinger & Gillingham, 2012). A third research of Richter (2013) also found a small, but significant positive relation between added installations within the postcode district and the adoption rate within this postcode district (Richter, 2013).

According to Statistics Netherlands (2020a) and Verhees et al. (2013) stimulating the use of renewable energy supply can support the adoption rates. Governmental institutions can do this by subsidies, tax rebates and other liabilities. Nowadays, the price people pay for energy from renewable sources is still higher than from non-renewable sources. Due to these reasons, the Dutch government has tried to stimulate the adoption of renewable energy sources. In 2001, the Dutch government started subsidizing PV panels by the Energie Premie Regeling (EPR). This subsidy was granted to residents for installing a PV system. In 2004, the Dutch government decided to terminate the subsidy, because of the large number of requests (Overheid.nl, 2003). In 2003 another incentive was introduced, namely the environmental quality of electricity production (MEP). This subsidy was terminated in 2006, because of the high costs and the target of 9% renewable energy for the year 2010 was almost reached. The incentive after the MEP was the Stimulus Policy Renewable Energy Production (SDE). The main difference between the MEP and the SDE was that the SDE also focused on green gases and renewable heath instead of only renewable energy.

The SDE was also dependent on the market value of energy and natural gas. The budget for this incentive has increased over the years from 1.5 billion in 2011 to 12 billion in 2018 and 10 billion in 2019 (Statistics Netherlands, 2020a). In 2011, an addition was made to the SDE, the so-called SDE+. The difference is that every technique has its tariff and maximum budget. Another important addition is the adding of incentives for renewable heath without taking into account the combination with electricity. From 2016 to 2021, the Sustainable Energy Investment Subsidy (ISDE) was applicable, this was mainly focused on fewer natural gases and sustainable heath (Negro et al., 2009; Statistics Netherlands, 2020a; Verhees et al., 2013). Due to these subsidies, the attractiveness of adopting a PV system increased a lot and the break-even period decreased. However, due to the increased attractiveness, some dwellings are equipped with PV systems that are not sufficient. This insufficiency could be because of the orientation towards the sun or other limiting factors. This could be a result of a lack of reliable and clear information supply (Jager, 2006).

1.2 Research goals and problem statement

This research aims to establish a spatial analysis of influential factors in the adoption of PV systems in the Netherlands. In this analysis, different influential factors will be taken into account. These factors can be divided into categories, those categories are sociodemographic, built environment, economical, environmental awareness and peer-effects.

Many researchers have studied the factors that influence the adoption of PV systems. Balcombe et al. (2013) has analyzed the motivations and barriers in the adoption of micro-generation adoption. Other researchers have focused on the sociodemographic factors that influence the adoption of PV systems (Diamantopoulos et al., 2003; Kurdgelashvili et al., 2019; Sigrin et al., 2015; Ugulu, 2019). Important sociodemographic characteristics are income, education level and age. Others have researched whether the built environment influences the adoption of PV systems (Graziano & Gillingham, 2015; Zahran et al., 2008). For this category, literature has looked at population density (Graziano & Gillingham, 2015; S. Müller & Rode, 2013), building density (Rode & Weber, 2016; Schaffer & Brun, 2015), degree of urbanization (Rode & Weber, 2016; Schaffer & Brun, 2015; Wallace & Wang, 2006; Zahran et al., 2008), owner-occupied homes and the ratio between single and multi-family homes (Graziano & Gillingham, 2015; S. Müller & Rode, 2013). Environmental awareness is found to be significant in the adoption of renewable energy systems (Bamberg, 2003; Bamberg & Möser, 2007). Rode and Weber (2016) and Schaffer and Brun (2015) found that solar radiation influences the diffusion of PV systems across different regions.

However, in the existing literature, the spatial component of residential technology adoption is not taken into account much. For that reason, the contribution of this research is to add a spatial dimension to the analysis. In prior studies, this spatial component was taken into account as an explanatory variable that relates to space (e.g. prior installed installations in the neighbourhood). However, these studies did not take into account that explanatory variables can differ from place to place. Several studies have found significant relations between PV adoption and peer-effects (Bollinger & Gillingham, 2012; Dastrup et al., 2012; Graziano & Gillingham, 2015; Richter, 2013; Robinson & Rai, 2015), which shows the relevance of the spatial element in adoption rates. However, these studies use global regression techniques instead of local regression techniques what can influence the insights. In one area could an explanatory variable have a higher influence than in another area. Also, local craftsmen and local solar incentives make this variable more spatial. Lately, the importance of space in this relation is getting more interest and several studies have focused on this. For example, Dharshing (2017) has investigated the regional differences between PV adoption on the county level in Germany. Graziano and Gillingham (2015) have looked at spatial patterns in PV adoption in Connecticut. Jayaweera et al. (2018) have researched the local factors that affect PV diffusion in the residential sector in Sri Lanka. For this research the following research question is formulated:

What are the influential factors in the spatial adoption of photovoltaic panels across neighbourhoods in the Netherlands?

The corresponding sub questions are:

1. To what extent do sociodemographic factors influence the adoption of photovoltaic panels in the neighbourhoods of the Netherlands?
2. To what extent do built environment factors influence the adoption of photovoltaic panels in the neighbourhoods of the Netherlands?
3. To what extent do economic factors influence the adoption of photovoltaic panels in the neighbourhoods of the Netherlands?
4. To what extent does environmental concern influence the adoption of photovoltaic panels in the neighbourhoods of the Netherlands?
5. What are the spillover effects between neighbouring neighbourhoods looking over time and space?

1.3 Relevance

The results of this research can be used for scientific purposes, policy-making and marketing (Graziano & Gillingham, 2015). With knowledge about spatial diffusion of PV adoption, governmental institutions can intervene in a more targeted and effective way, for example, by providing local incentives or by marketing on specific target groups. According to Best et al. (2019a), local incentives will positively influence the PV uptake, which indicates the importance of policies on the local level instead of policies on higher levels. As said at the beginning of the introduction, solar energy is an important renewable energy source. Existing literature shows that there are a lot of influential factors in PV adoption, however major differences exist between regions, which makes it necessary to look at a more detailed spatial scale. J. Palm and Eriksson (2018) identify the importance of information supply by different means. Each population group in their study requires a different way of information supply, for instance the quantity, the used channels and the reliability (J. Palm & Eriksson, 2018). This research confirms earlier determined influential factors and gives new insights on new influential factors.

1.4 Structure

The next chapter will review the existing literature and the possible determinants of the adoption of PV systems. In the chapter after that, the data will be described and the methods that will be used are explained. In the following chapter, the results of the analysis will be shown and the last chapter will present the conclusions and discussion, lastly, some recommendations will be given.

1. Introduction; in the introduction, the research topic will be supported with literature. Both literature on the background and the relevance together form the boundaries of the research. Additional to this, the research question and corresponding sub-questions will be formulated.
2. Literature review; together with the introduction this literature review forms the fundamental part of the research. Firstly, an introduction to sustainability and solar energy will be shown. This include also detailed information about PV panel usage in the past, mainly focused on the Netherlands, taking into account are the adoption rates, market developments and incentives. The second part is about the factors that influence the adoption rates of PV panels for residents. For this part, it is important to investigate the existing data on these influential factors and the possible methods.

3. Methodology; in this part different aspects will be discussed. These include the research design, the data description, variables determination, the discussion of the regression method and the expected results.
4. Analysis; in this part, the data is collected from different sources. Secondly, the data sets will be combined. Then, the different variables will be reviewed and some basic analysis will be done. In the last part, the regression analysis is conducted.
5. Results; in the results section, the results will be interpreted and analyzed.
6. Conclusions, discussion and recommendations; the last step in this research consist of the answers to the research questions, the discussion of the results and the conclusions and some recommendations for possible ways to use these results. Also included in this part of the research are the recommendations for future research.

Chapter 2

Literature Review

In this chapter, the literature review will be reviewed. This includes, first an introduction to sustainability and solar energy, after that the adoption mechanisms will be explained and lastly the factors that affect the adoption of PV panels.

2.1 Sustainability

In 1987, a definition of sustainability is formulated by the World Commission on Environment and Development (WCED). This definition is as follows: "Sustainable development consists of economic-development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (World Commission on Environment and Development, 1987). However, since the 1980s, several meanings have been assigned to sustainability. These meanings differ in definition, application, target group, etc. This wide range of meanings may indicate that sustainability is a growing concept of which the meaning is not fully explored. In general, the core of the idea is that the condition of the biophysical environment of the earth will contribute to the needs of economic growth and the human population without depletion of resources and health of living things (Kent, 2015). The same report of the World Commission on Environment and Development in 1987 introduced the three E's of sustainability, namely, Environment, Economy and Equity (World Commission on Environment and Development, 1987). The interpretation is as follows: protecting the environment, preserving economic development and growth and promoting equity. In figure 2.1, three representations of these three E's are shown. The figure on the left represents it as three intersecting circles, the bottom right corner as three pillars holding sustainability and the top right corner as a concentric circle approach (Purvis et al., 2019). As the figure already shows, there are different ways to represent sustainability, this confirms the idea of the different meanings associated with sustainability. In addition, solar energy plays an important role in this process, because it can solve the problem with the environment, it can also stimulate economic developments and it also can create more equity in the world since the sun is not as dispersed as other power sources.

Other ideas of sustainability were formulated after this. For example, Brown et al. (1987) and Kidd (1992) elaborated the concept of sustainability for clarification and different contexts. The concepts of sustainability according to Brown et al. (1987) are: sustainable biological resource use, sustainable agriculture, carrying capacity, sustainable energy, sustainable society and sustainable economy and sustainable development. The purpose of Brown et al. (1987) with this elaboration of concepts was to create a common understanding and meaning of sustainability. Those six different definitions together form two major aspects that describe sustainability, namely, ecology and economy. The concept of sustainable energy mainly consists of the high reliance on fossil fuels to generate electricity. However, the amount of carbon that is produced with the burning of fossil fuels is also important in this concept. Thirdly, this concept also includes the decrease in the demand side of energy. The implementation is through the use of renewable energy sources such as solar, wind, hydro, etc. The foundations of sustainability according to Kidd (1992) are ecological/carrying capacity, resource/environment, biosphere, critique of technology, no growth-slow growth and ecodevelopment. Both studies show similar meanings for the concept of sustainability as the definition of WCED.

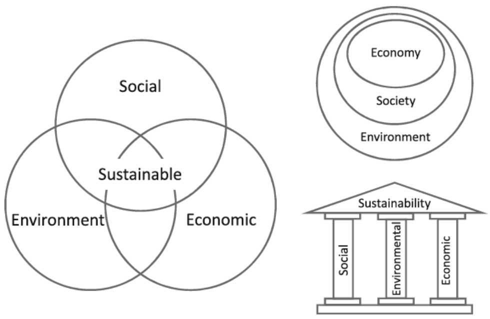

Figure 2.1: Representation of the three E's of sustainability (Purvis et al., 2019)

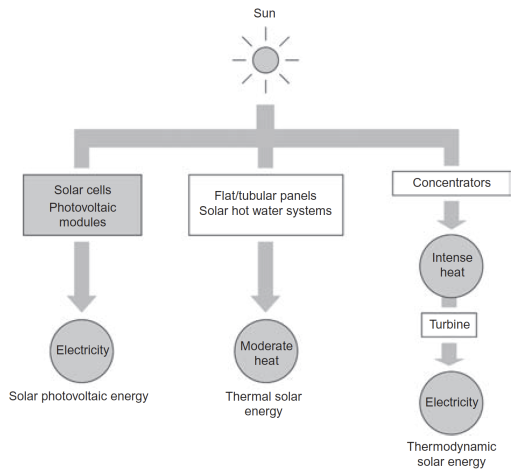

Figure 2.2: Three ways of exploiting solar energy (Labouret & Villoz, 2010)

2.2 Solar Energy

Solar energy, as the word already indicates is the production of energy from the sun. There are various applications to use the sun as an energy source, an active and a passive way. The passive way is characterized by the fact that buildings are designed that they make use of the warmth of the sun. The active application is characterized by the use of PV generators to generate electricity and thermal collectors for heating and cooling (Hammer et al., 2003; Phillips, 2019). The transformation of sunlight into energy is performed through PV modules that consist of PV cells. According to Labouret and Villoz (2010), there are three ways of exploiting solar energy. The first one is with PV modules that consist of PV cells, the second one is from the heath of the sun to thermal solar energy and the third one is through concentrators, steam and a turbine into electricity (figure 2.2). A common question that is asked is if the sunniness of a location influences the power generation. In fact this is correct, but PV cells transform light into energy, so in less sunny places it is still possible to generate energy from the sun. The amount of light received by a PV module will influence the amount of energy generated. It is possible to generate energy from artificial light, but the intensity of sunlight is much higher than that of artificial light even if the weather is not that sunny.

According to (Labouret & Villoz, 2010), there are two different categories for PV installations, namely, grid-connected installations and stand-alone installations. The difference between these two is that grid-connected installations add electricity to the collective grid and stand-alone installations use generated electricity directly. These stand-installations can either make use of a battery to store electricity or use the generated electricity immediately as a power source. For this last category, storing electricity is not possible, so when there is no light there is also no energy. When the demand for electricity is higher than the supply, it could happen that a user runs out of electricity. For that reason, there are hybrid stand-alone systems, that add a second generator to the system. The grid-connected installations deliver electricity to the grid. Two possible methods can be observed. Firstly, all the electricity produced is sent back to the grid (receive money) and the corresponding dwelling will use electricity from the grid (pay money). Secondly, only the surplus electricity is sent back to the collective grid (receive money) (Labouret & Villoz, 2010).

As sad in the section above, renewable energy supply plays an important role in sustainable development. Labouret and Villoz (2010) gave several arguments that solar energy is an important source for the environmental problems the world is facing these days. Firstly, the sun is an extremely powerful energy source (Ashok, 2021; Kabir et al., 2018). The International Energy Agency calculated that an area of 145,000 km km^2 is needed to meet the demand of planet earth. This is approximately 0.03% of the total surface of planet earth (510,000,000 km km^2) (Sharp, 2021). Secondly, PV panels use silicon as their resource and this is much present in the soil of the earth. Thirdly, the time it takes before a PV panel generates the amount of electricity that is required to produce a PV panel is much lower compared to other energy generators. Fourthly, the manufacturing of PV panels is mainly done with recycled materials. Fifthly, the generation of electricity using PV panels does not emit any greenhouse gases. Sixthly, PV panels have a long life expectancy and are easy to install and maintain. The use of solar energy has also an impact on the human being, it reduces the amount of toxic waste in the air, generates job opportunities and economic activities and it reduces the depopulation and urbanization since it increases the living standards for places that are less equipped with electricity (Labouret & Villoz, 2010).

2.3 History of solar energy in the Netherlands

According to Phillips (2019), the start of PV cells date back to 1883 when Charles Fritts invented the first one. In 1954, Bell Labs connected a silicon PV panel to a battery. The PV cell of Fritts had an efficiency of approximately 1%, where the PV cell of Bell had an efficiency of 6%. This new technique was first introduced for commercial purposes in the space industry. However, the low efficiency and high cost resulted in low adoption rates, even in the oil crisis of the 1970s. During this time, the demand for other energy sources was urgent. After this, research on efficiency continued and PV panels started to perform better when looking at adoption rates, efficiency and costs (Phillips, 2019).

In 1974, the first energy white paper was published in the Netherlands, mainly focused on the call for new alternative energy sources. The following energy paper was published in 1979, this paper was mainly focused on creating interest for research and an exploring phase was the result. This exploring phase had positive outcomes and the research fields were expanded. In 1984, the International Solar Energy Society was established, this society consist of a collaboration between the government and three universities. They focus on two research themes, namely, crystalline gallium arsenide (GaAs) with high efficiency of 35% to 40% and amorphous silicon (a-Si) with a lower efficiency of 15% to 20%, but lower production costs. The years after, PV energy was getting more interest and the budgets increased. In 1986, the national research program was launched and this had a budget of 2.7 million euros and the aim was to follow the international market and promote development and increase awareness. Right after this time, there were some doubts about the technology and its suitability in the Netherlands, because of the amount of solar radiation in the Netherlands and the implementation in a strongly centralized energy system. The doubts were not strong enough to stop the development and in 1988 a first incentive was introduced with the Support Regulation Energy Savings and Flow Energy (SES), this support regulation was renewed in 1991 and it compensated the costs of a system by 40%. In 1990, an energy white paper was published with for the first time, a measurable goal. The goal was to replace two Peta joule of fossil fuels with PV energy. The years after, there was a change of strategy, more focused on the efficiency and balance of PV systems. Shortly after the white paper of 1991, a new plan was introduced, the National Environmental Policy Plan (NMP), which stimulated energy companies to contribute to the sustainable and environmental friendly energy supply. For this an Environmental Action Plan (MAP-1) was established to reduce the CO_2 by 9 million tons by the year 2000, this is equal to 10%. In 1994, the Energy research Centre of the Netherlands (ECN) and Renewable Energy Systems (R&S) started with the production of a PV system with 16% efficiency. Just before that time, ECN started cooperation with foreign and Dutch parties, this cooperation consist of parties from the research sector but also from the industrial sector. Due to the new production line, the Netherlands acquired an important role in the research market.

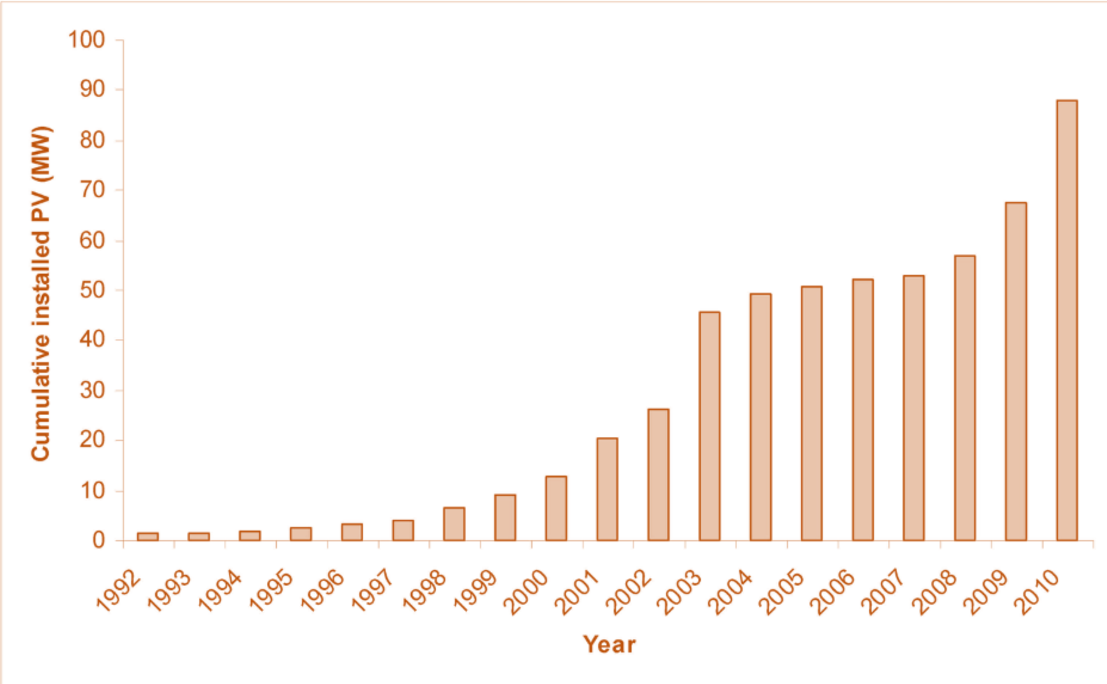

Figure 2.3: Installed PV capacity per year in the Netherlands (Negro et al., 2012)

Around this time some universities in the Netherlands set up some research projects related to residential PV placing. With these projects not only the knowledge was stimulated, but also shows the importance of PV panels.

In 1994, a turning point can be observed, this is due to the ending of the incentives on small projects. Initially, the government wanted to end this incentive abruptly. However, due to criticism the incentive gradually ended in 1995. In 1995, the third energy white paper was introduced, this included several programs and policies to achieve the goals.

From 1995 until 2003 the largest growth of renewable energy can be observed (figure 2.3). During this time, the Energy Premium Regulation, which was introduced in 1998, was the main subsidy for PV systems. It gained 1.59 euros per Wattpiek (Wp) for homeowners. In 1997, R&S turned into Shell Solar Energy BV. They develop a new production line for PV panels with a capacity of 2.5 MWp per year. Also around this time, energy distribution companies, such as NUON, PNEM and REMU, become more involved in the PV market and large-scale roof projects were set up. Due to the high interest of the government, the non-governmental organizations also became more active, for example, Greenpeace with their 'Solaris' project. This project is characterized by the fact that it offers complete PV modules for 500 euros. These kinds of projects are used to create a larger PV market by lowering prices for end-users.

At the beginning of the 21st century, a large political shift can be observed. An agreement is made that in the following years 500 million on energy and environmental subsidies has to be saved each year. These savings were needed, because of the large number of incentives. Also, there are some examples that for one project several different incentives are consulted. In 2001, the green energy market is liberalized, so from this moment, PV has to compete with other renewable energy sources. Still, the PV market is growing, due to the EPR and other incentives from non-governmental organizations. Also, municipalities boost the market for small-scale PV systems by providing local subsidies. For example, 'ZonZeker' in Groningen together with Essent and 'More roofs under the sun' from Novum.

After some time, a focus shift towards other countries can be observed. This shift towards other countries can be explained by the growth of the worldwide market. Due to the leading position of the Netherlands on knowledge and technology production, entrepreneurs concentrate on the export market in foreign countries such as Germany and Spain. These countries have more favourable financial benefits. In 2003, the government decided to abruptly end the EPR within a month after the announcement, which ended in a rapid growth in requests before the end of the subsidy. In the last month, 90,000 PV panels were sold to households. In the three years after 2003, the installed capacity is almost every year zero. In figure 2.3, the flattening of installed capacity is well visible in the years after 2003. To replace the EPR a new subsidy is introduced in the same year, the Environmental Quality of Electricity Production (MEP). This subsidy is introduced to promote renewable energy in the Netherlands. The tariff of this subsidy is lower than the tariff of the EPR, so it is unlikely that this will increase the adoption since the investment costs will remain high.

For this reason, a new advocacy coalition is formed with professionals of energy research, industry and policy. This coalition was formed to stimulate the future of PV development and to increase the expertise level. The Ministry of Economic Affairs together with Senternovum developed a new set of energy research and technology development programs. These programs were not only focused on PV, but on all sorts of renewable energy. On the other hand, the priority was on PV expenditure. HollandSolar also proposed a vision for the future of PV and forms the basis of the transition path of PV. During this period research institutes and universities, including the Eindhoven University of Technology, continue to investigate the technology. These research programs rely mainly on European support. Another national programme that is introduced in 2004 is the Joint Solar Programme, this programme focused on exploratory research of efficiency. In 2003, a former employee of Shell started a new PV factory to prevent leaking knowledge to other countries. The PV was boosted by these kinds of activities and gained attraction from entrepreneurs again. The 'Roadmap Solar Energy' was introduced in 2005 and outlined the future of the PV market in the Netherlands. This roadmap was proposed to the government, however not included in the agenda. In 2006, the MEP subsidy is abruptly ended by the Dutch government, this created mistrust in the government in the PV sector. In 2007, a new cabinet is entered and together with this, a new coalition agreement is launched. This agreement consists of new goals for renewable energy and the reduction of energy from fossil fuels. Next to this, a platform is established for this transition towards sustainability. This platform activated some initiatives in the Dutch PV sector. This platform consist of the Dutch research en technology development institutes, the Ministry of Economic Affairs and the PV industry. In 2008, the Stimulation of Renewable Energy production (SDE) subsidy is introduced. However, this subsidy generated more interest than there was budget. The goal was to stimulate the PV sector with this subsidy, however, due to the long stagnated PV sector, experts think that it will take some time to get back on track. In 2011, the share of PV in the Netherlands is only small compared to other European countries. The total capacity installed is approximately 130 MW, which is equal to 0.3% of the total renewable energy and 0.02% of total energy demand (Huijben & Verbong, 2013). Because of the lack of implementation due to the SDE subsidy, in 2011, an addition was made to this subsidy, the so-called SDE+ subsidy. The focus of this subsidy was mainly on cheaper renewable energy sources. In 2012, no subsidies were applicable for small-scale PV systems, since these systems already profited from the fact that these systems provided energy to the grid and get a reduction on the energy bill. The same year the cabinet felt and a new agreement was introduced, in this agreement, a new subsidy was included. This subsidy ensured a reduction of 12% on the costs for a system with a maximum of 650 euro. Al these changing subsidies and programmes have ensured the position of the Netherlands on the global scale (Negro et al., 2009).

2.4 Adoption process

The adoption process of innovations and thus renewable energy sources is a well-studied subject in literature and dates back to the work of Gabriel Tarde, a French sociologist, at the beginning of the 19^{th} century (Tarde, 1903). However, it only got serious attention after the work of Everett Rogers in 1962. Rogers (1962) introduced the diffusion of innovations theory, this theory will be explained in more detail below. According to Dolowitz and Marsh (2000), adoption has different degrees, these degrees are copying (no change to innovation), emulation (adjustments to full fill needs), hybridization (combination of innovation or part of innovation) and inspiration (innovation as inspiration for new innovation). There are numerous studies that investigated the adoption of innovation and the process behind this (Davis, 1989; Fishbein & Ajzen, 1975; Rogers, 2003; Venkatesh et al., 2003). These theories all have their relationship with the theory of Rogers (2003).

The theory of reasoned action was developed by Fishbein and Ajzen (1975). A visual representation of the model is shown in appendix A figure A.1. The TRA model is about how different factors affect the social behaviour of people when adopting. The foundation of this model is based on the 3-component model of Rosenberg and Hovland (1960). However, Fishbein and Ajzen (1975) considered the component of the subjective norm. The theory assumes that the way people intend to behave affects their actual behaviour. This intention of behaviour is determined by the attitude towards behaviour and subjective norms (Macovei, 2015).

The Technology Acceptance Model was developed by Davis (1989). This model is an extension of the TRA model. The model explains the steps that are undertaken during the process of accepting new technologies and thus the actual use of the innovation. This model served as the basis for a number of refined models. In appendix A, figure A.2 a visualization of the model is shown.

The Unified Theory of Acceptance and Use of Technology was developed by Venkatesh et al. (2003) and is a combination of eight different models. In appendix A, figure A.3, a visualization of the model is shown. It shows that there are four determinants for behavioral intention, namely, performance expectancy, effort expectancy, social influence and facilitating conditions. The factors gender, age, experience and voluntariness of use affect on these four determinants.

2.4.1 Diffusion of Innovations Theory (DOI)

The theory of 'Diffusion of Innovations' was developed by Rogers (1962). In figure 2.4, a visualization of the theory is shown. It shows the different types of adopters and the probability that they will adopt an innovation. According to van der Kam et al. (2018), the Netherlands was among the category of early adopters in 2018. Looking at the increase in the number of solar panels in the years before 2018, the Netherlands will currently be around the transition from early adopters to early majority. Comparing this innovation curve with figure 2.3, it is visible that solar panel adoption still increases each year. Due to subsidies, figure 2.3 shows a somewhat distorted picture, however, it can be concluded that the Netherlands is still in front of the center of the innovation curve.

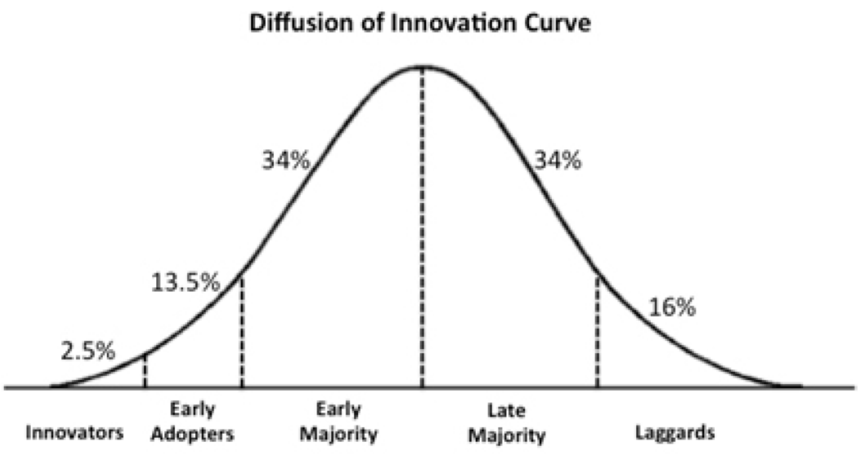

Figure 2.4: Diffusion of Innovations (Rogers, 1962)

Rogers (2003) also explains the decision process in combination with innovations. In 2003, he constructed the following definition for the process of adopting an innovation:

> *A process through which an individual or other decision-making unit passes from first knowledge of an innovation, to forming an attitude towards the innovation, to a decision to adopt or reject, to implementation of the new idea, and to confirmation of this decision."* Rogers (2003)

A visualization of this is shown in figure 2.5. This figure illustrates the process of an individual when making a decision. The stages are explained below:

- Knowledge: In this stage individuals get familiar with the innovation;
- Persuasion: In this stage individuals form an attitude towards the innovation;
- Decision: As the word indicates, in this stage individuals take a decision;
- Implementation: Individuals started to implement the innovation;
- Confirmation: In this stage, individuals tend to look for a confirmation of their decision in the decision stage.

All these different stages are important in the implementation of PV panels since each stage can influence an individual. The first stage is the knowledge stage, reliable and understandable information supply is important to introduce individuals to the power generation from the sun. This stage has an influence on the attitude of individuals towards this innovation and thus the decision to adopt it. In the implementation phase, the experience towards the innovation influences the way people recommend or discourage the innovation. The same goes for the confirmation phase.

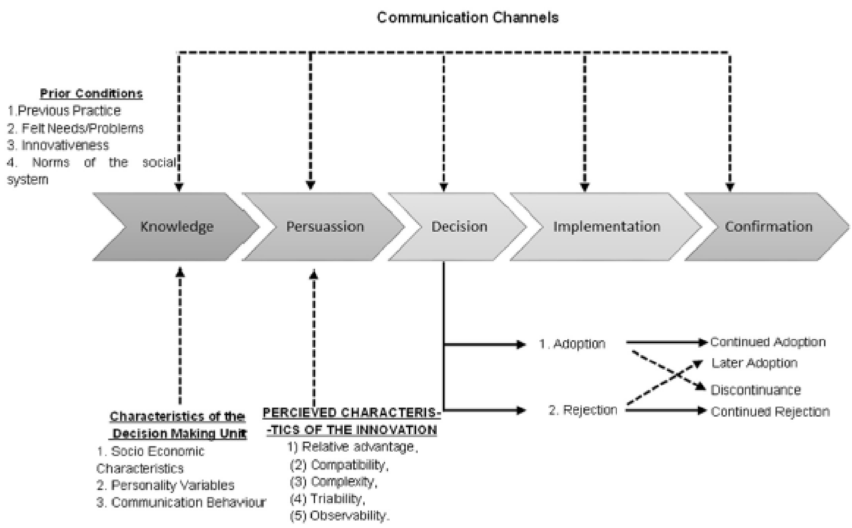

Figure 2.5: The innovation-decision process (Rogers, 2003)

2.5 Influential factors in the adoption of PV panels

In this section, the literature is reviewed on influential factors in the adoption of residential PV panels adoption. These factors are categorized into the following categories; sociodemographic, built environment, economic, environmental concern, peer effects and other factors.

2.5.1 Sociodemographic factors

Income

The influence of (household) income has already gained a lot of interest in literature. This is due to the number of positive significant relations between income and adoption rates. However, some studies found an insignificant relation and Balta-Ozkan et al. (2021) found a negative impact on PV adoption (Balta-Ozkan et al., 2021). Graziano and Gillingham (2015) have studied the spatial patterns of adoption of residential solar PV systems and have found that income is positively related to the adoption rates. However, they also found that these spatial patterns not only follow patterns of income but also of other variables (Graziano & Gillingham, 2015). Guta (2018) found that households with a higher income are more likely to adopt PV systems than poorer people in their study to driving factors of solar technology in rural Ethiopia. Also, Zahran et al. (2008) found a positive significant relation between the wealth and the number of solar households, this wealth was measured through the median house value. Vasseur and Kemp (2015) found in their descriptive analysis in general a higher income for adopters in comparison with non-adopters. This study was based on an analysis of four different groups: voluntary adopters, involuntary groups, potential adopters and rejecters (Vasseur & Kemp, 2015). These same results were found by Sigrin et al. (2015), they made a division between adopters and non-adopters and also found higher incomes for adopters (Sigrin et al., 2015). Several other studies found positive significant relations between income and PV adoption rates. Briguglio and Formosa (2017), Dharshing (2017), Jacksohn et al. (2019), S. Müller and Rode (2013), Rode and Müller (2016), Rode and Weber (2016), Sardianou and Genoudi (2013), Schaffer and Brun (2015) and Zhang et al. (2011) found this positive significant relation in their study. Some studies found an insignificant relation between income and adoption rates, such as Mundaca and Samahita (2020), Richter (2013) and Zhang et al. (2011).

Gender

The relation between gender and the adoption of PV panels has been less discussed in literature. Bollinger and Gillingham (2012) found higher adoption rates with the percentage of the population who are male (Bollinger & Gillingham, 2012). Similar findings were found in the research of Jacksohn et al. (2019) and Rahut et al. (2018). Jacksohn et al. (2019) found a decrease in choice probability for renewable technology adoption if the household head is a female (Jacksohn et al., 2019). Rahut et al. (2018) found a positive significant relation between a male as household head and the adoption of solar energy. Diamantopoulos et al. (2003) found a significant relation based on the following hypothesis: "Females are more likely to participate in green activities" (Diamantopoulos et al., 2003). Guta (2018) found also differences between males and females for adopters and non-adopters of PV panels. In the category adopters, the household head was 84% male and for the non-adopters, this was 93%. The study found indeed that males are less likely to adopt solar energy technology compared to females (Guta, 2018). Sardianou and Genoudi (2013) found an insignificant relation between gender and the willingness to adopt renewable energy (Sardianou & Genoudi, 2013).

Age

For the variable age, the results vary widely. Bollinger and Gillingham (2012) found less adoption for people aged between 20 and 45 and for the age group 65+. Diamantopoulos et al. (2003) found that age is related to knowledge about environmental issues. They also found that younger people are more concerned about the quality of the environment and younger and older people partly differ in participation in green activities (Diamantopoulos et al., 2003). Guta (2018) found a positive significant relation between the age of the household head and the adoption of PV panels in rural Ethiopia. A positive relation in this study means the older the age, the more likely to adopt solar energy technologies (Guta, 2018). In the research of Zahran et al. (2008), only people aged between 40 and 49 were taken into account, however, they found a significant positive relation between this variable and the number of households heated through solar energy by 9.1% (Zahran et al., 2008). Vasseur and Kemp (2015) found in their descriptive comparison between adopters and non-adopters a younger mean age for adopters in comparison to the group of non-adopters. Sardianou and Genoudi (2013) indicated in their empirical study that middle-aged people are more willing to adopt renewable energy sources than other age categories. The same results were found by Jayaweera et al. (2018) in their study to local factors influencing the spatial diffusion of PV panels in Sri Lanka. Sommerfeld et al. (2017) found that people aged over 55 are more likely to adopt PV panels due to their concerns about increasing electricity prices. Jacksohn et al. (2019) investigated the importance of factors in the adoption of renewable energy sources, in their research they found that the choice probability for PV panels decreases with the age (Jacksohn et al., 2019). Mundaca and Samahita (2020) found a negative effect of age on home PV uptake, which means the older people get, it is less likely to adopt PV panels (Mundaca & Samahita, 2020). So to conclude, different results can be observed in literature, however, most of the research found that middle-aged people are more likely to adopt.

Marital status

The relation between marital status has not been addressed a lot in the existing literature, one study found a significant relation between the adoption of renewable energy technologies and marital status, while two studies found an insignificant relation. Gezahegn et al. (2018) found that married people positively affect renewable energy adoption (Gezahegn et al., 2018). Diamantopoulos et al. (2003) found no differences between singles and married people in profiling green consumers. Sardianou and Genoudi (2013) also find an insignificant relation between marital status and the willingness of consumers to adopt renewable energy.

Unemployment

Dharshing (2017) found a significant negative relation between unemployment and PV panel adoption. In his study, he investigates household dynamics in PV adoption rates at the county level in Germany (Dharshing, 2017). Briguglio and Formosa (2017), Diaz-Rainey and Ashton (2011), Shi et al. (2013) and Welsch and Kühling (2009) found also a negative significant relation between unemployment and PV adoption, this result strengthens the result of Dharshing (2017).

Education

Diamantopoulos et al. (2003) investigated sociodemographics in profiling green consumers through testing of hypothesizes. From their study, it can be stated that higher educated people are more knowledgeable of environmental issues and are partly more likely to perform green activities (Diamantopoulos et al., 2003). Several other studies found significant positive relations between education level and PV adoption rates, such as Davidson et al. (2014), Dharshing (2017), Guta (2018), Jager (2006), Jayaweera et al. (2018), Keirstead (2007), Sardianou and Genoudi (2013), Sigrin et al. (2015) and Vasseur and Kemp (2015). Balta-Ozkan et al. (2015) identified a positive effect of technical and vocational qualifications as determining factors for PV deployment. Sommerfeld et al. (2017) found an insignificant relation between higher educated people and PV uptake. Contrary to all these studies, Jacksohn et al. (2019) identified a higher choice probability in adopting PV systems for medium educated households heads instead of higher educated ones.

Ethnicity

Bollinger and Gillingham (2012) found higher adoption rates for the percentage of people who are white, Graziano and Gillingham (2015) found the same result, however, the results were only weak (Bollinger & Gillingham, 2012; Graziano & Gillingham, 2015). The study of Sunter et al. (2019) is dedicated to the relation between PV panels adoption and the race and ethnicity of the population. In their research, they use data about rooftop PV panels and demographic information for the United States. The results from their research indicate a lower PV adoption for blackand Hispanic majorities. In the analysis household income and home-ownership are controlled for, since these variables are often attributed to this difference (Sunter et al., 2019).

2.5.2 Built environment factors

Population density

Balta-Ozkan et al. (2015) found a negative significant relation between population density and PV adoption and this means that the lower the population density, the more likely people adopt PV panels. Similar findings were identified by S. Müller and Rode (2013) and Rode and Müller (2016). Similar results were also found by Graziano and Gillingham (2015), who found also that housing density (in their research measured as population divided by land area) is more important than income and political affiliation in predicting PV adoption rates (Graziano & Gillingham, 2015).

Housing density

According to Schaffer and Brun (2015), housing density is an important factor in residential PV adoption rates. PV adoption increases when there is a higher housing density. Housing density is measured through the number of residential buildings per km^2 (Schaffer & Brun, 2015).

Commuting distance

The factor commuting distance is not addressed much in literature, however, Bollinger and Gillingham (2012) found a positive significant relation between the commuting distance and higher adoption rates. This is measured through the percentage of the population who have over 30-minute commuting time. In their study, a possible explanation for this significance could be due to the visibility of PV panels. The longer the commuting time, the more PV panels are observed by people on rooftops (Bollinger & Gillingham, 2012).

Availability of roof space

The available rooftop space is also a determining factor in higher rates of adoption. Briguglio and Formosa (2017) found that unshared roof space increases the adoption of PV panels and according to Mundaca and Samahita (2020) a possible barrier to adopt PV panels is that their roof is not optimal (Briguglio & Formosa, 2017; Mundaca & Samahita, 2020).

Detached homes

According to Balta-Ozkan et al. (2015), the share of detached homes has a significant positive effect on the PV adoption rate. The reason for this could be due to the available roof space or due to the ease of management of construction (Balta-Ozkan et al., 2015).

Home-ownership/share of renter-occupied dwellings

An important factor in determining PV uptake is home-ownership. Therefore, this has gained a lot of interest in literature. According to Sommerfeld et al. (2017), PV adoption is positively related to home-ownership with a direct relation, however, they also found a lower adoption rate for areas with a higher concentration of rented houses or apartments (Sommerfeld et al., 2017). Similar results were found by Graziano and Gillingham (2015), they found significant results for decreasing adoption with an increasing share of renters (Graziano & Gillingham, 2015). Briguglio and Formosa (2017), Keirstead (2007), Schaffer and Brun (2015) also found a positive relation between home-ownership and PV uptake. Balta-Ozkan et al. (2015) found a negative effect on PV uptake, they give as a possible reason that home-owners feel less inclined to reduce their energy costs (Balta-Ozkan et al., 2015). Vasseur and Kemp (2015) found that non-adopters with no home-ownership see this as a barrier for adoption.

Construction of new buildings

According to Dharshing (2017), the construction of new buildings is negatively related to regional PV adoption (Dharshing, 2017).

Urbanization

Zahran et al. (2008) found, in their study to the spatial distribution of solar-heated households in the United States, that urbanization has a positive effect on the number of solar-heated households. In this study, they measured urbanization of a county by dividing the people that are living in urban areas by the total population (Zahran et al., 2008). Comparable findings were found by Wallace and Wang (2006).

2.5.3 Economic factors

Solar radiation

The importance of solar radiation in the decision to adopt PV panels is through the effectiveness of PV panels. Šúri et al. (2007) identify the importance of solar radiation in relation to the power generation of PV panels. Their study aims to analyze differences in solar electricity generation from PV panels in the European Union (Šúri et al., 2007). Several other studies found a positive significant relation between solar radiation and PV uptake (Balta-Ozkan et al., 2015; Rode & Müller, 2016; Schaffer & Brun, 2015). Solar radiation is mostly measured through the yearly global radiation (kWp/m^2). Zahran et al. (2008) found indeed the same results for solar-heated households (Zahran et al., 2008).

High up-front costs

The negative relation of up-front costs is addressed a lot in the literature together with the other economic factors. Schelly and Letzelter (2020) found that up-front costs are an important decision factor in residential PV adoption under residences in upstate New York (Schelly & Letzelter, 2020). Balcombe et al. (2013), Vasseur and Kemp (2015) found that the costs of an installation are too high and acts as a barrier to the adoption. Zhang et al. (2011) identified the installation costs as a significant negative relation on the diffusion of PV panels (Zhang et al., 2011). Some research in less developed countries indicates the importance of these up-front costs. Qureshi et al. (2017) found that the costs of PV panels are the most important barrier in the adoption of PV panels. Their study was focused on Lahore, Pakistan (Qureshi et al., 2017). Comparable results were found by Parsad et al. (2020) for Kerala, India, Jayaweera et al. (2018) for Sri Lanka and Ugulu (2019) for urban Nigeria.

Governmental incentives

Governmental incentives are closely related to the high up-front costs of installations since these incentives are mostly focused on reducing the up-front costs. Jayaweera et al. (2018), Parsad et al. (2020), Qureshi et al. (2017), Ugulu (2019) mention the lack of financial support from governmental institutions as a barrier to the adoption of residential PV systems. Also, Bollinger and Gillingham (2012), Briguglio and Formosa (2017), Mundaca and Samahita (2020) mention governmental incentives as a motivator for adoption. Bauner and Crago (2015) found in their study that discounted value of benefits has to exceed the installation costs by 60%. Without any incentives, the average adoption time is eight years longer. Financial incentives play an important role in adoption time (Bauner & Crago, 2015). Vasseur and Kemp (2015) found that approximately one-third of their sample would adopt solar panels if there would be an attractive subsidy program (Vasseur & Kemp, 2015). According to Best et al. (2019a), local incentives and according to Zhang et al. (2011), regional incentives will positively influence PV adoption. Jager (2006) found that financial support is an important motive to adopt PV systems.

Electricity costs

Sigrin et al. (2015) identified that recent adopters adopt PV panels to protect against the increasing electricity costs. According to Best et al. (2019b) and J. Müller and Trutnevyte (2020) households with a higher average electricity price will be more likely to adopt PV panels. In relation to the expected electricity prices and actual electricity prices are the savings on the electricity bill. Several studies have pointed out the influence of this factor. According to Balcombe et al. (2013), Schelly and Letzelter (2020), Sigrin et al. (2015), Ugulu (2019) and Vasseur and Kemp (2015), reducing energy bills is an important decision factor for adoption. Also, the payback time could be seen as a barrier to adopting, this is discussed in the research of Balcombe et al. (2013).

2.5.4 Environmental awareness

Environmental awareness is identified as an important factor in the adoption and diffusion of PV panels. Some studies found that it is even more important than the economical factors. Zahran et al. (2008) found, in their study to influential factors for solar-heated households in the United States, that the percentage of Democrats has a positive effect on PV adoption (Zahran et al., 2008). According to Bollinger and Gillingham (2012) and Davidson et al. (2014), the share of hybrid cars increases the probability of the adoption of PV panels. This could be due to the environmental awareness of these households or due to the fact that these electric cars need to be fueled at home. Fueling these hybrid cars with solar energy will reduce energy bills (Bollinger & Gillingham, 2012; Davidson et al., 2014). Balta-Ozkan et al. (2015) found that more polluted areas do have a positive effect on the adoption rates within this area (Balta-Ozkan et al., 2015). Some other studies also indicate the importance of environmental awareness (Mundaca & Samahita, 2020; Zhang et al., 2011). Balcombe et al. (2013) identify that environmental benefit can be considered as motivation, they also identify that the image to others of being environmentally aware is a decision factor (Balcombe et al., 2013). According to Jager (2006), environmental problem awareness could also be identified as a decision factor in PV adoption. Dharshing (2017) found a significant positive relation between environmental attitude, which was measured through the share of green voters, and the adoption rates in their spatial lag panel model (Dharshing, 2017). However, Jacksohn et al. (2019) found that environmental concern has a weak, but significant relation, on the probability of opting for a PV system. In their study, they mention the difference compared to economic factors, which are strongly related (Jacksohn et al., 2019).

2.5.5 Peer-effects

Some researchers have investigated the relation between peer effects and PV adoption rates in the residential sector. For this particular case, the effects are mostly through social interaction or by the visibility of PV systems. Graziano and Gillingham (2015) found a strong positive significant relation between PV adoption and the number of earlier installed PV panels in the surrounding area. They also indicate the spatial importance of PV panel adoption by identifying spatial patterns and diffusion. These peer effects diminish over space and time. The relation between adoption and spatial neighbour effect can be explained as follows; adding an installation within 0.5 miles of adopters in the year before the adoption increases the number of installations by 0,44 PV panels on average (Graziano & Gillingham, 2015). Rode and Weber (2016) found that the visibility of PV panels might have a correlation with adoption in the surrounding. Their study is focused on imitation in the adoption of PV panels, they found that imitation is mainly relevant on a smaller scale (Rode & Weber, 2016). Jager (2006) also describes this as follows; "Consumers frequently feel satisfied when consuming the same as their neighbours (social needs) and often engage in social comparison and imitation when deciding what to consume" (Jager, 2006). Graziano et al. (2019) found also peer effects within 0.5 miles and effects are stronger within block groups than between neighbouring blocks. These small-scale peer effects were also found by Rode and Müller (2016), who found that peer effects only had influence within 200 meters and diminishes over time. S. Müller and Rode (2013) found a significant positive relation between preexisting systems nearby and the adoption rates in this area. The peer effect diminishes over distance, the further away, the less influence it has on the decision. Bollinger and Gillingham (2012) found almost the same results as Graziano and Gillingham (2015), they find that the probability of adoption increases by 0.78 percentage points when adding a PV installation within the zip code. However, in their study they address three issues when drawing conclusions for these peer-effects, those are; homophily, simultaneity and correlated unobservables. Homophily means that results can be influenced by the fact that people with similar interests and characteristics live in the same environment. Thus, in the case of peer effects, this may also mean that these pre-installed installations are a consequence of similarities between people. Simultaneity means that people are influenced by peers, but at the same time, the people influence their peers as well. Correlated unobservable factors mean that, for example, local campaigns or promotions have been implemented that are not known (Bollinger & Gillingham, 2012). A third research of Richter (2013) also found a small, but significant positive relation between added installation within the postcode district and the adoption rate within this postcode district. These effects are stronger in areas with higher educated people (Richter, 2013). According to Mundaca and Samahita (2020), peer effects are positive and significantly related to the adoption rates of PV panels. These peer effects are mostly through hearing, however, seeing is also important. Their study was based on a survey that determines the influence of peer effects (Mundaca & Samahita, 2020). Similar findings were identified by A. Palm (2017). His study focused on the inner workings of peer effects and concluded that adopters confirmed that peer effects have influenced their decision. This influence was mainly by confirming the functioning of PV systems. He also found that active peer effects are more important than passive peer effects (A. Palm, 2017). The study of J. Palm and Eriksson (2018) also mentions the importance of peer effects for a certain group of the population. In their comparison between different kinds of (non) adopters, they found that the accidental adopter group prefers face-to-face information. Since PV panel adoption is increasing, it will be a more important discussion point in daily life (J. Palm & Eriksson, 2018).

2.5.6 Other

Information supply

Several studies found that the lack of knowledge is considered as a barrier to the adoption of PV panels. J. Palm and Eriksson (2018) investigated differences between four different groups, namely, non-adopters, environmentally engaged, professional and accidental group. Each group experience information supply differently. The non-adopters see it as a barrier, while the environmentally engaged group wanted information from third neutral parties and the professional group had problems with the comparison of suppliers. The last group, the accidental group, preferred face-to-face information (J. Palm & Eriksson, 2018). Balcombe et al. (2013) found, in their study to motivations and barriers for adoption of microgeneration energy technologies, that the lack of information is experienced as a barrier for people. Ugulu (2019) mentions that knowledge of PV systems would be necessary for households to adopt a system. Also, Jayaweera et al. (2018) specify the importance of information supply for potential adopters. This lack of knowledge about PV panels is also determined as a barrier by Mundaca and Samahita (2020), Parsad et al. (2020). Jager (2006) found that organizing information and support meetings has a positive effect on the diffusion of PV systems. This could be due to the lack of knowledge provided over other sources (Jager, 2006).

Self-sufficient

Becoming self-sufficient in energy supply can be seen as a motivator for the adoption. Balta-Ozkan et al. (2015) mention the relation between increasing energy demand and becoming self-sufficient. Important in this relation is the increasing energy costs (Balta-Ozkan et al., 2015). Balcombe et al. (2013) also mentioned the importance of becoming self-sufficient as a motivator by the argument that not becoming self-sufficient can provide unforeseen circumstances (Balcombe et al., 2013). Vasseur and Kemp (2015) mention also this self-sufficiency as a motivator.

2.6 Conclusions

From this literature review, several conclusions can be drawn. There are different ideas and interpretations of the term sustainability, but a common thread is certainly visible. The core of the idea is that the condition of the biophysical environment of the earth will contribute to the needs of economic growth and the human population without depletion of resources and health of living things. PV cells make use of the infinite supply of solar energy. Solar energy dates back to the end of the 19^{th} century. After several occurrences, it attracted more and more interest. Governmental and non-governmental institutions have taken various steps to increase its adoption. However, the diffusion of solar panels is not yet at the intended level. Around 1960, researchers also gained more attention for the mechanism behind adoption. There are different models developed to predict people's behaviour. For example the diffusion of innovations theory from Rogers (1962). Following his theory, the Netherlands was among the category of early adopters in 2018. The increase in the number of solar panels in the years before 2018 shows that the Netherlands will currently be around the transition from early adopters to early majority.

Important influential factors that have consistent results over the literature are the household wealth, measured through different variables, such as income, housing value, unemployment and education level. These variables are positively related to PV adoption. Also, age is well investigated in literature and overall shows consistent results, middle-aged people are more likely to adopt PV panels, due to their capital and environmental concern. An important lifetime occurrence is the planning of retirement since the installations will reduce the costs for electricity. Variables that are less investigated are gender, marital status and ethnicity. However, males, married people and people with a western background are more likely to adopt PV panels. For the category built environment, the results vary across existing literature. For example, population density has a negative relation with PV adoption, this can be explained by the available roof space and shadowing from other buildings in urban areas, but also the amount of owner-occupied homes and share of single and double family homes in less dense areas. However, some positive relations can be observed too, for example, urbanization, housing density, commuting distance, availability of roof space, detached homes, home-ownership and the construction of new buildings. Most of these results are as expected, however contrary findings are found for population density in relation to housing density and urbanization. Longer commuting distance will increase the visibility of PV panels and thus the acceptance of the technology. For the economical factors, all the variables are related to the high upfront costs and the return on investment time of PV panels. Existing literature discusses the following variables: solar radiation, high upfront costs, governmental incentives and electricity costs. A high consistency can be observed for solar radiation since this positively influences the amount of electricity is generated. The high upfront costs of PV installations are often experienced as a barrier to the adoption of PV panels. On the other hand, governmental incentives are experienced as a motivator, since it reduces the upfront costs. Lastly, people see the reduction in electricity bills as a positive aspect of PV panels. An important factor that influences the adoption rates of PV panels is the environmental concern of people. Some studies indicate this as the most important factor. This factor is measured through different variables, however, the most common one is the political preference of people. The influence of peer effects is also consistent in the literature, several studies find significant results for peer effects on PV panel adoption, this variable was mainly measured through previously installed PV panels. Peer effects for PV adoption consist mostly of social interaction and visibility of systems in the neighbourhood and diminish over space and time. Another factor that influences the adoption of PV panels is the information supply, some people experience that there is too much information available and others found the information not complete or hard accessible. Also, becoming self-sufficient can also be experienced as a motivator for PV adoption. In table 2.1 the variables are shown and the expected results of these variables.

Table 2.1: Expected results

	Variable	Adoption
Sociodemographic		
Male	↑	↑
Female	↑	↓
Age	↑	↑↓
Married	↑	↑
Not married	↑	↓
Ethnicity	↑	↑↓
Income	↑	↑
Unemployment	↑	↓
Education	↑	↑
Built environment		
Population density	↑	↓
Housing value	↑	↑
Singe-family homes	↑	↑
Multi-family homes	↑	↓
Owner-occupied homes	↑	↑
Rented homes	↑	↓
Construction of new buildings	↑	↓
Housing density	↑	↓
Degree of urbanity	↑	↑
Commuting distance	↑	↑
Availability of roof space	↑	↑
Detached homes	↑	↑
Economical		
Solar radiation	↑	↑
High up-front costs	↑	↓
Governmental incentives	↑	↑
Electricity costs	↑	↑
Environmental awareness		
Environmental concern	↑	↑
Peer-effects		
Previously installed installations	↑	↑
Other		
Information supply	↑	↑↓
Self-sufficiency	↑	↑↓

Chapter 3

Methodology

In this chapter, the methodology will be presented. First, the datasets are described, then the different variables are explained with some basic descriptives to explore and get familiar with the data. The third part will explain the method that is used. Lastly, the conclusion with the steps for the analysis will be discussed.

3.1 Data description

For this research data is used from Statistics Netherlands (CBS), which is an institution that provides data about the Netherlands. The first data that is used gives information about the installed PV panels in the Netherlands for the years 2016, 2017 and 2018. The data set is divided into different spatial levels, namely, neighbourhood, district, municipality and country. For the year 2016, there are 397,390 systems installed spread over the Netherlands, in 2017 this is 529,005 and for 2018 this is 720,522. These numbers are based on registration at a certain location in a certain year. The number of installations in each year represents the total number of installations, so the installations in 2018 include also the installations of 2016 and 2017. For that reason this research continuous to use the data set of 2018. To determine which spatial level suits best the research, the missing values of each level are visualized in maps. These maps are shown in appendix B, next to these maps, some basic statistics of the data are shown to determine the level of focus. An overview of the missing cases for the number of installations is shown in table 3.1. The scope of this research has to do with spatial relations on small scale and for that reason, the lowest level will be taken into account. This is also because of the fact that the number of PV panels in a district is a cumulative of all the neighbourhoods in that district. So missing cases are applicable for all the different spatial levels. Two other variables are retrieved from this data set and these include the number of installations in a neighbourhood in the two years prior to 2018. These two variables will be used as the variable for peer effects.

The second data set that is used is also retrieved from the CBS, this data set includes variables connected to sociodemographics, socioeconomic and built environment characteristics. For the sociodemographic factors, the following variables are retrieved: gender, age, marital status and ethnicity. Secondly, for the socioeconomic variables, the following variables are used: income and unemployment. For the built environment characteristics, the following variables are retrieved: population density, housing value, the ratio between single and multi-family homes, ratio between owner-occupied and rented houses, date of construction, housing density and the degree of urbanity.

Thirdly, a shapefile for the solar radiation is used, this raster data is retrieved from SolarGIS (figure 3.1). This solar radiation is measured through the daily average of direct normal irradiation and measured in kWh/m^2.

Table 3.1: Missing values overview for the number of installations per neighbourhood (Statistics Netherlands, 2020c))

	Valid	Missing	Total	Percentage valid
Neighbourhood	**10,551**	**2,754**	**13,305**	**79.3**
District	**2,920**	**166**	**3,086**	**94.6**
Municipality	**380**	**0**	**380**	**100.0**

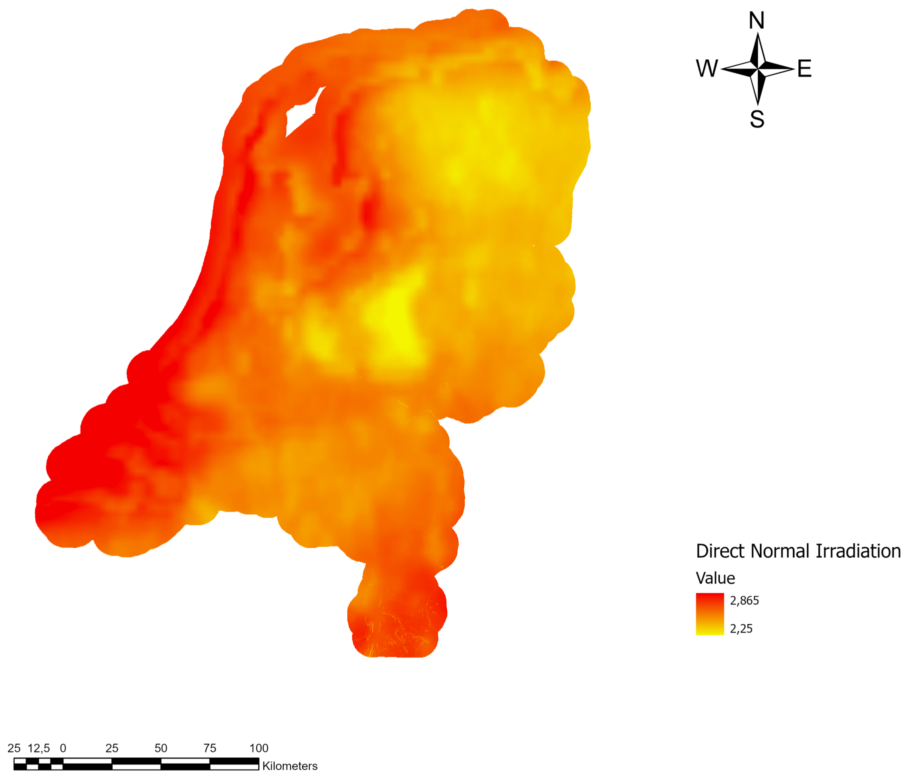

Figure 3.1: Direct Normal Irradiation Netherlands (SolarGIS, 2019)

3.2 Variables

In table 3.2, the dependent and independent variables are shown. In this table, some additional information is added, such as the variable name in the data set, the source and the level of measurement.

3.2.1 Dependent variable

The dependent variable is built up by two different variables. First, the data about the number of PV installations per neighbourhood in the Netherlands is retrieved from CBS (Statistics Netherlands, 2020c). Second, the data on the number of houses is also retrieved from CBS, however from a different data set (Statistics Netherlands, 2018). To control for different housing stocks per neighbourhood the dependent variable is divided by the housing stock per neighbourhood. The dependent variable will be the number of PV installations per house in a neighbourhood.

Installed PV panels

The data consist of numbers about the installed PV installations and their capacity on houses in the Netherlands. This data is available on different spatial levels, namely on neighbourhood-, district-, municipal- and country-level. The number of installed panels is based on a registration of a panel at a certain location in a certain year. A registration in the Productie Installatie Register (PIR) or CertiQ or through a request for a grant scheme, VAT deduction or energy investment deduction scheme (Kramer & Segers, 2018). In appendix D (figure D.1), the data descriptives are shown. A missing case in this data set means that a value cannot occur based on logical reasons or a value is unknown, secret or unreliable.

Number of houses

The housing stock is the number of residences on January 1 of the corresponding year, retrieved from Statistics Netherlands (2018). A residence is characterized by an object with at least one living function. In appendix D, figure D.2, the data descriptives for this variable are shown.

3.2.2 Explanatory variables

For the explanatory variables, three different data sources are retrieved. Firstly, two variables are retrieved from CBS, which include the variables to determine the prior installed installations (Statistics Netherlands, 2019, 2020d). Secondly, a shapefile is used with statistical data about neighbourhoods in the Netherlands. According to ESRI (n.d.-i), the following definition is formulated: "A shapefile is an ESRI vector data storage format for storing the location, shape, and attributes of geographic features. It is stored as a set of related files and contains one feature class" (ESRI, n.d.-i). The borders of neighbourhoods and districts are based on information supplied by the municipalities. This data is retrieved from CBS (Statistics Netherlands, 2018). Thirdly, a raster data file for solar radiation is used (SolarGIS, 2019). The data from CBS is linked through an identical neighbourhood code. The raster data for solar radiation is linked through an overlay function in ArcGIS Pro. After this overlay tool, the mean direct normal irradiation is calculated for each neighbourhood and this created the variable for solar radiation.

Table 3.2: List of variables and their sources

Variable	Full name	Source	Level of Measurement
NEICODE	Neighbourhood code	CBS	Nominal
NEINAME	Neighbourhood name	CBS	Nominal
DISCODE	District code	CBS	Nominal
MUNCODE	Municipality code	CBS	Nominal
MUNNAME	Municipality name	CBS	Nominal
NUMPEOPLE	Number of people	CBS	Ratio
NUMHOUSE18	Number of houses in 2018	CBS	Ratio
NUMINSTAL18	Number of installations in 2018	CBS	Ratio
INS18	Number of installations per house in 2018	CBS	Ratio
CAPACITY18	Capacity in 2018	CBS	Ratio
NUMHOUSE16	Number of houses in 2016	CBS	Ratio
NUMINSTAL16	Number of installations in 2016	CBS	Ratio
INS16	Number of installations per house in 2016	CBS	Ratio
INCR16	Increase in number of installations per house in 2016	CBS	Ratio
NUMHOUSE17	Number of houses in 2017	CBS	Ratio
NUMINSTAL17	Number of installations in 2017	CBS	Ratio
INS17	Number of installations per house in 2017	CBS	Ratio
INCR17	Increase in number of installations per house in 2017	CBS	Ratio
NUMMALES	Number of males	CBS	Ratio
NUMFEMALES	Number of females	CBS	Ratio
NUM00_14	Number of people aged between 0 to 14	CBS	Ratio
NUM15_24	Number of people aged between 15 to 24	CBS	Ratio
NUM25_44	Number of people aged between 25 to 44	CBS	Ratio
NUM45_64	Number of people aged between 45 to 64	CBS	Ratio
NUM65	Number of people aged 65 and over	CBS	Ratio
NUMUNMAR	Number of unmarried people	CBS	Ratio
NUMMAR	Number of married people	CBS	Ratio
NUMDIV	Number of divorced people	CBS	Ratio
NUMWID	Number of widows	CBS	Ratio
NUMWEST	Number of western foreigners	CBS	Ratio
NUMNOTWEST	Number of non-western foreigners	CBS	Ratio
PERC1FAMH	Percentage single-family houses	CBS	Ratio
PERCMFAMH	Percentage multi-family houses	CBS	Ratio
PERCOWNED	Percentage owner-occupied houses	CBS	Ratio

continued on the next page

continued from the previous page			
PERCRENTED	Percentage rented houses	CBS	Ratio
PERCUNKNOWN	Percentage property unknown	CBS	Ratio
PERCBEFORE2000	Percentage of houses built before 2000	CBS	Ratio
PERCAFTER2000	Percentage of houses built after 2000	CBS	Ratio
POPDENS	Number of residents per km^2	CBS	Ratio
ADDRESSDENS	Neighbourhood address density	CBS	Ratio
URBAN	Addresses per km^2	CBS	Ratio
HOUSEVAL	Average housing value	CBS	Ratio
PERCUNDER40	Percentage of people that belong to the national 40% of persons with the lowest income	CBS	Ratio
PERCABOVE20	Percentage of people that belong to the national 20% of persons with the highest income	CBS	Ratio
NUMUNEMPLOY	Number of people with an unemployment benefit	CBS	Ratio
SOLAR	Amount of solar radiation	SolarGIS	Ratio

3.3 Methods

This section will show the method to find answers to the above-mentioned questions. In the introduction of this research, it was mentioned that the key contribution of this paper is to add a spatial viewpoint to the regression. For this reason, a spatial regression will be conducted.

3.3.1 Spatial regression

According to Anselin (2017), a spatial regression can be described as a regression with attributes that have a location factor included in the variable and this location factor matters. For a normal regression, it is possible to randomly place data on a map without influencing the result of the regression. If spatial data would be randomly placed on a map and a spatial regression would be conducted, this will influence the result of the regression. In the left image of figure 3.2 spatial data is randomly placed over a map (so this is a fake map), and on the right, the same data is placed as it should be. When conducting a spatial regression on this data the results will change, but doing a normal regression the result will be the same in both cases.

This spatial regression incorporates two spatial effects that have to be taken into account, namely, spatial dependence and spatial heterogeneity. Both spatial effects have similar characteristics, however, there are some differences. According to Basile et al. (2014), spatial dependence means that values measured on a certain place are related to values of neighbouring areas (Basile et al., 2014). On the other hand, spatial heterogeneity means that under-clustered and over-clustered areas are scattered in the study area (Anselin, 2010). There are some pitfalls in spatial analysis. The first one is about ecological fallacy, which means that individual behavior cannot be explained at an aggregate level of regression. For example, if the crime rate is high in a certain area, it cannot be stated that on an individual level someone is a criminal. The second one is about the scale of the analysis. For this, it is important to take into account which scale fits best to answer the research question. The third one is about data that is measured on different scales (Anselin, 2017), for example, country versus province data. Spatial regression does have some additional difficulties in contrast with normal, non-spatial regression techniques. These are spatial autocorrelations, which means that data tend to be similar to each other when it is closer to each

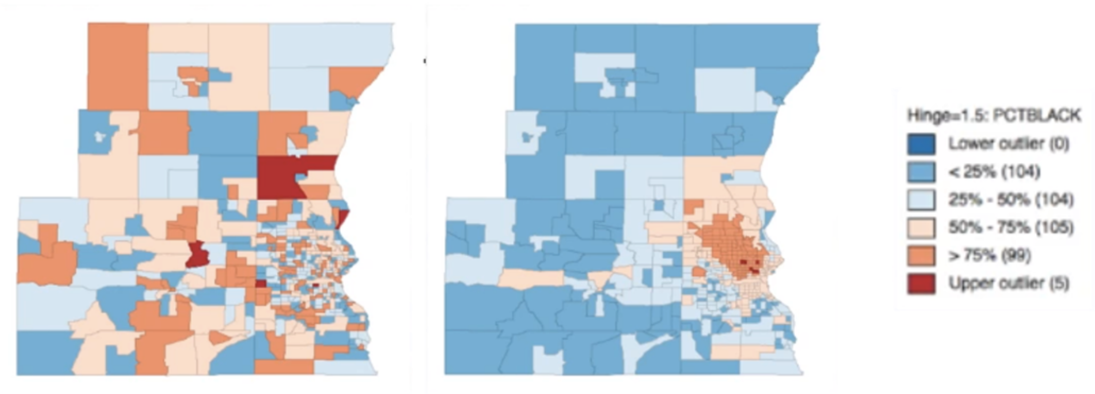

Figure 3.2: Spatial data randomly placed on a map and placed properly (Anselin, 2017)

other. The second one is nonstationarity, which means that relationships between explanatory variables and the dependent variable are inconsistent over the study area (ESRI, n.d.-h).

3.3.2 Geographically Weighted Regression (GWR)

Firstly, a brief explanation of an OLS is given. An OLS regression, also called a global regression model, predicts a vector of the dependent variable by means of a set of predictor or explanatory variables. This relationship between predictor variables and dependent variables can be modelled as follows:

$$y = \sum_j X_j \beta_j + \varepsilon \tag{3.1}$$

where y is the dependent variable, X_j means an explanatory variable, β_j means the corresponding regression coefficient for this explanatory variable and ε means a random error term. The corresponding coefficient can be calculated as follows:

$$\beta_j = (X^T X)^{-1} X^T y \tag{3.2}$$

Geographically Weighted Regression (GWR) is a spatial regression technique, which was developed by Brunsdon, Fotheringham and Charlton (Fotheringham et al., 2002). A GWR makes a local model of each feature in the dataset by fitting an equation to it. This regression is an expansion of the Ordinary Least Squares (OLS) regression, which makes a global model of the data. The added value is the locality factor in the regression and the incorporation of the detection and consideration of spatial nonstationarity. The size of these local models can be selected by the user (based on literature) or they can be determined by other statistical tools, such as the least-squares cross-validation, which will be explained below. Just like other regressions, a GWR predicts the dependent variable (y_i) with estimations for β_{ij}, for each variable j and corresponding location i.

$$y_i = \sum_j X_{ij} \beta_{ij} + \varepsilon_i \tag{3.3}$$

where y_i means the dependent variable at location i, X_{ij} means an explanatory variable at location i and variable j, β_{ij} means the corresponding regression coefficient for this explanatory variable at location i and variable j and ε_i means a random error term at location i. The corresponding coefficient of β_{ij} can be calculated as follow:

$$\beta_{ij} = (X^T W_i X)^{-1} X^T W_i y_i \tag{3.4}$$

where W means a matrix containing a set of weights for the regression points (equation 3.5) and y is a vector of the dependent variable.

$$W_i = \begin{pmatrix} \alpha_{i1} & 0 & \cdots & 0 \\ 0 & \alpha_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_{iN} \end{pmatrix} \tag{3.5}$$

In this formula, N stands for the number of neighbourhoods. This weight matrix above is calculated from the chosen kernel function (explained below). For a GWR this maximum likelihood estimate formula is not a single equation, however, it is a sequence of equations.

Circle of influence

According to Brunsdon et al. (1998), for each data point (for this research neighbourhood), a circle is drawn of a certain radius (r). As said above an OLS regression is conducted for each data point in the dataset. This means that only data is taken into account that geographically falls within this circle, also called regression points (see equation 3.6). An issue that occurs when dealing with this regression is the value that the user uses for the radius. If a large radius is chosen, this radius will be that large, that all the regression points fall within this circle and if this radius is too small, the standard errors will be large. A second issue could be the binary notation of including and excluding data related to this radius. This means that if data is close to the border of the radius, it is included or excluded from the regression. In spatial analysis this abruptly ending of an area is unusual.

Initially, the data that is included in each regression is determined by the following formula:

$$\alpha_{ik} = \begin{cases} 1 & \text{if } d_{ik} < \text{r}, \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

where α_{ik} is the weighting for a regression point that is included or excluded from the model and d_{ik} is the distance between i and k. However, to take into account the issue of this binary notation a continuous weighting function can be used:

$$\alpha_{ik} = \exp(-d_{ik}^2/2h^2) \tag{3.7}$$

or:

$$\alpha_{ik} = \exp(-d_{ik}/h) \tag{3.8}$$

or:

$$\alpha_{ik} = \begin{cases} \left\{1 - (d_{ik}/h)^2\right\}^2 & \text{if } d_{ik} < \text{r}, \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

where h provides some control of the weighting of a regression point. These weighting functions can be notated as kernels, the notation is $\alpha_{ik} = K(d_{ik})$. These functions ensure a gradual decrease in the influence of regression points. According to Wheeler and Páez (2010), there are two types of kernels, namely, adaptive and fixed kernels. Adaptive kernels are characterized by the fact that they are influenced by the density of regression points and fixed ones do not. For the above mentioned fixed kernel functions some preferred constraints are applicable, namely:

(a) $K(0) = 1$,
(b) $\lim_{d\to\infty} \{K(d)\} = 0$ and
(c) K is a monotone decreasing function for positive real numbers

In the above-mentioned formulas, h means the earlier mentioned bandwidth, also called the kernel bandwidth. This bandwidth can be determined based on literature, however, sometimes it is not possible to base it on existing literature. For these cases, an automatic data-led choice of

h can be the desired method. A possible method could be the least-squares cross-validation. The sum of squared errors can be written as follows:

$$\mathrm{SS}(h) = \sum_i \{y_i - \hat{y}_i(h)\}^2 \tag{3.10}$$

To find h in the above mentioned formula is by minimizing the above-mentioned formula. However, when doing this the answer will be zero, as $h \to 0, \hat{y}_i \to y_i$. To avoid this problem the kernel function K has to be replaced to a function K^* such that:

$$K^*(0) = 0, \tag{3.11}$$

$$K^*(d) = K(d) \quad \text{if } d \neq 0 \tag{3.12}$$

To determine h, the cross validated sum of squared error formula will look as follows and by minimizing this formula h can be found:

$$\mathrm{CVSS}(h) = \sum_i \{y_i - \hat{y}_i(h)\}^2 \tag{3.13}$$

3.4 Conclusions

In this chapter, the methodology for the analysis is defined. A GWR, which is a spatial regression technique, makes a local regression for each neighbourhood in the dataset by fitting an equation to it, so it is a sequence of OLS regressions for each neighbourhood in a dataset. The added value is the locality factor in the regression and the incorporation of the detection and consideration of spatial nonstationarity. An elaboration of the steps for the analysis is shown in figure 3.3. These phases consist of three main steps that can be subdivided into smaller steps. The phases are the data collection, the data preparation and the regression. The data collection is about collecting the number PV installations, neighborhood characteristics and solar radiation data. These different data sets will be linked to each other before moving on to the next phase. In the next phase, different steps have to be made concerning the preparation of the data. These steps have to do with missing values, outliers, multicollinearity and the final set of variables. Once this is determined a regression can be made. This regression consists of three parts, namely an OLS, a Global Moran's I and a GWR.

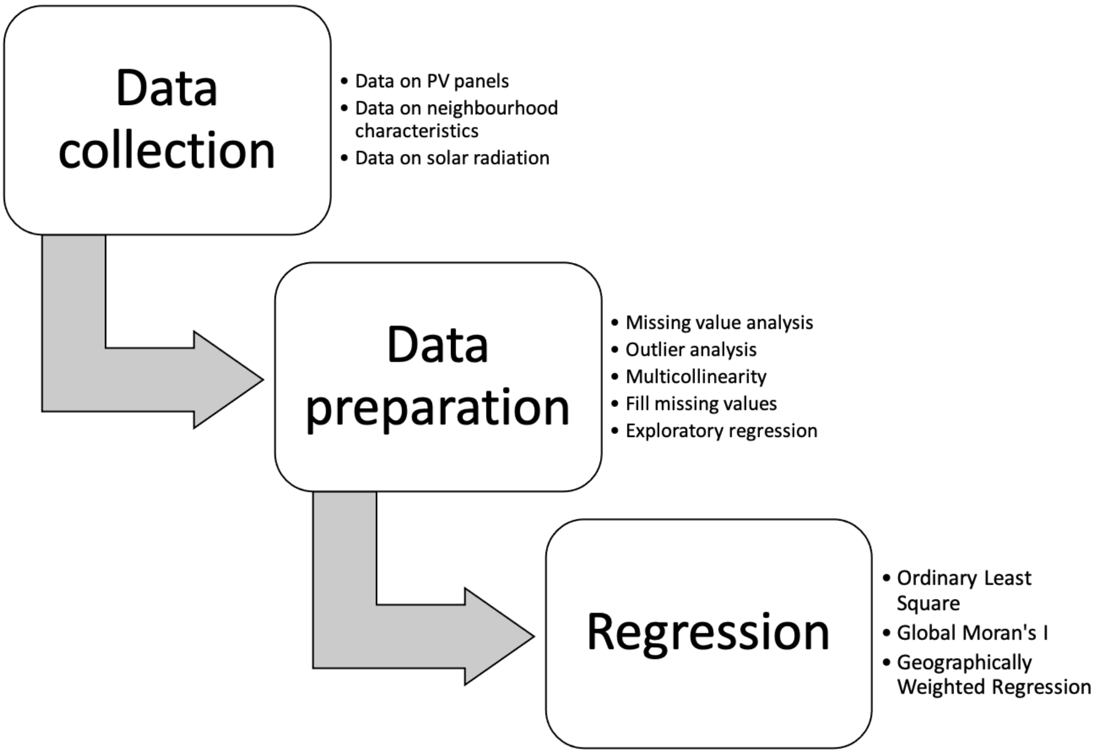

Figure 3.3: Procedure of analysis

Chapter 4

Results

In this chapter, the results of the analysis will be shown. In the first subchapter, the steps for the data preparation will be shown, this includes detection and processing of multicollinearity and missing values, an exploratory regression and data descriptives. Then the regression analysis will be discussed, this includes an Ordinary Least Squares regression, Global Moran's I statistic and a Geographically Weighted Regression.

4.1 Data preparation

The total sample size is 13,305 neighbourhoods, before any data preparation. The syntax for the data preparation is shown in appendix C. First, the variables are renamed to get better variable names. Secondly, the frequency table of the number of PV installations per neighbourhood is computed and this shows some basic statistics of the variable (see appendix D, table D.1). The neighbourhoods that do not have values for the dependent variable (NUMINSTAL18) are deleted from the data. A missing value, in this case, means that no value is available for one of the following reasons: the number cannot occur on logical grounds or the number is secret, unknown or insufficiently reliable. Some different options are considered, however, the number of system missing cases for this variable is too large to use a certain tool. ArcGIS provides a tool to fill in missing values based on values of neighbouring areas, however, it is recommended to not predict more than 5% of the values (ESRI, n.d.-a). The percentage of missing cases for the dependent variable is approximately 20% (see table 3.1). It is also considered to change the scope of the study to a different degree of urbanity, however, the percentage of missing values remains approximately 20 percent for all the different degrees and combinations of different degrees. To control for different neighbourhood sizes, the number of PV installations is divided by the number of houses in a neighbourhood. The descriptives of the number of houses are shown in appendix D, table D.2. For the same reason, some variables are operationalised. The variables that are influenced by the size of the neighbourhood are divided by the number of people in a certain neighbourhood (NUMPEOPLE), see appendix D table D.3. So the new values are percentages of the total population in a neighbourhood. It relates to the following variables: number of males, number of females, number of people aged between 0 to 14, number of people aged between 15 to 24, number of people aged between 25 to 44, number of people aged between 45 to 64, number of people aged 65 and over, number of unmarried people, number of married people, number of divorced people, number of widows, number of western foreigners, number of non-western foreigners and number of unemployed people. The following step was to specify missing values. The impossible values are specified as missing, these include values with more than 100%. After this modification, the statistics changed a bit and some other statistics are derived to further investigate the variable. These include a histogram, the interquartile range (IQR) and the upper- and lower quartiles. The boundaries are calculated with the following formula (equation 4.1 and 4.2) (Finn, 2017; Frost, n.d.; Purplemath, n.d.):

$$Lower \ \ boundary = Q1 - 3.0IQR \tag{4.1}$$

and

$$Upper \ \ boundary = Q3 + 3.0IQR \tag{4.2}$$

Where Q1 is the first quartile, Q3 is the third quartile and IQR is the interquartile range. The results of this calculation are used to specify outliers. These outliers are specified as missing. This process is followed for all the explanatory variables as well. After that, based on the distribution of histograms, some 'zero' values are specified as missing, this applies to the following variables: percentage of people aged between 0 and 14, percentage of people aged between 15 and 24, percentage of people aged between 25 and 44, percentage of widows, percentage of divorced people, percentage of foreign western people, percentage of non-western foreigners, percentage built before 2000, population density, address density and percentage of unemployed people. The corresponding histograms are shown in appendix D, figure D.1 to figure D.31.

4.1.1 Fill missing values

For this step, the data is exported from SPSS to ArcGIS Pro. This data is connected to a shapefile with polygons of neighbourhoods. The neighbourhoods with no value for the dependent variable are left out from the data. This connection is made through an identical neighbourhood code. After that, all the missing values for the variables that are taken into account are filled. According to ESRI (n.d.-d), this tool is different from other normal replacement techniques. Instead of normal replacements of missing values, this tool uses spatial neighbours to predict the value. The reason to fill in missing values is to prevent that during the analysis features with missing values are deleted from the analysis. With the deletion of these features also valuable data will be deleted. Firstly, ESRI advises to not predict more than 5% of missing values, these statistics are shown in appendix D, table D.4 (ESRI, n.d.-a, n.d.-d). Because of this criterion, the following variables are deleted from the model: increase in the number of installations in 2016, increase in the number of installations in 2017, percentage divorced people, percentage widows, percentage of western foreigners, percentage of non-western foreigners, percentage of property unknown houses, percentage houses built before 2000, percentage houses built after 2000, housing value, percentage of people that belong to the national 40% of persons with the lowest income, percentage of people that belong to the national 20% of persons with the highest income and percentage of unemployed people. Secondly, missing values need to be equally distributed over the study area, this is shown in figure E.1 to E.17 of appendix E. From the figures in the appendix, it can be seen that all the predictor variables are randomly distributed over the Netherlands, however, some clustering for the dependent variable in the North-East of the Netherlands can be observed. The tool has successfully filled all the missing values. The tool uses several input features. First of all, the fill method needs to be determined, this means on which statistic the value is estimated. For this research, the average of the neighbouring neighbourhoods is used. Secondly, the conceptualization of spatial relationships needs to be determined. For this input feature, the average number of neighbourhoods for each neighbourhood in the Netherlands is calculated. This calculation resulted in an average of 6.8 neighbourhoods per neighbourhood, so the values are estimated based on the seven nearest neighbourhoods (ESRI, n.d.-a).

4.1.2 Multicollinearity

The following step in the data preparation is the detection of multicollinearity. For this analysis two methods are used, the first one is a Pearson correlations table that detects the correlation between predictor variables. Secondly, a linear regression is conducted to detect the variance inflation factor (VIF). The correlations table is used to determine high correlations between variables, for this the following strength of the correlations factors is used (Health Knowledge, n.d.):

- $r \geq 0.8$ *very strong relationship*
- $r \geq 0.6$ *and* $r \leq 0.8$ *strong relationship*
- $r \geq 0.4$ *and* $r \leq 0.6$ *moderate relationship*
- $r \geq 0.2$ *and* $r \leq 0.4$ *weak relationship*
- $r \leq 0.2$ *very weak relationship*

According to Field (2018), a VIF of 10 or higher indicates multicollinearity. In appendix F, the correlation table is shown. The red marked correlations are very strong correlations and the yellow marked correlations are strong relationships. With the results of the correlation table and VIF values it is determined that the following variables are deleted from the model: percentage of females, percentage of unmarried people, percentage multi-family houses, percentage rented houses, population density and address density. The expectation was that these variables would have high correlations since most of these are the opposite of other variables.

4.1.3 Exploratory regression

The exploratory regression tool is used to find the best model for this analysis. The tool evaluates all the different combinations of predictor sets by means of some user input features. The tool is comparable with a step-wise regression, which is used in other statistical software packages. However, this tool is more comprehensive since it takes more into account than only the adjusted R^2. This tool uses several input features, these need to be determined by the user. First, the user needs to determine the maximum and the minimum number of explanatory variables (ESRI, n.d.-c). Before this regression, the variable people between 0 and 15 years old will be excluded, since the category age consists of five variables. The percentages of these five variables will accumulate to 100%, for that reason a regression will exclude one variable since it is redundant. To prevent that an important age category is excluded the above-mentioned variable will be excluded. The total possible explanatory variables are taken into account and this accumulates to fourteen different variables. Secondly, the minimum acceptable adjusted R^2, maximum p-value and the maximum VIF value need to be determined. The following values are selected, a minimum adjusted R^2 of 0.1, a maximum p-value of 0.05 and a maximum VIF value of 10. From the range of models, the model with the highest adjusted R^2 and lowest AIC is chosen. This model has an adjusted R^2 of 0.29 and an AIC of -28,044.15. This model includes the following explanatory variables: percentage of males, percentage of people aged between 15 to 24, percentage of people aged between 45 to 64, percentage of people aged 65 and over, percentage of married people, percentage of single-family houses, percentage of owner-occupied houses and degree of urbanity. In the next section, the data descriptives of these variables are shown.

4.1.4 Data descriptives

Dependent variable

In table 4.1, the statistics of the dependent variable are shown. The mean of this variable is 0.12, in other words, this means that on average approximately 12% of the houses in a neighbourhood are equipped with a PV installation. The observed maximum is approximately 42% and the minimum is less than 1%. In appendix D, figure D.1, the corresponding histogram is shown. The histogram shows a positive skewness and kurtosis, which means that the graph is asymmetrical and has pointy tails.

Explanatory variables

In table 4.2, the statistics of the explanatory variables are shown. These descriptive statistics show no peculiarities, the values seem to be logical. For the variable 'degree of urbanity', no mean is shown because this variable was measured on an ordinal scale. The corresponding histograms are shown in figure D.4 to figure D.31 of appendix D.

Table 4.1: Descriptives number of PV installations per house in a neighbourhood in 2018

Statistic		Value
Neighbourhoods	Valid	10445
	Missing	0
Mean		0.120
Median		0.110
Std. Deviation		0.075
Range		0.424
Minimum		0.001
Maximum		0.424

Table 4.2: Descriptives explanatory variables

Variable	N	Mean	Median	Std. Deviation	Min	Max
Percentage of males	10445	50.328	50.000	2.629	40.909	59.375
Percentage of people aged between 15 and 24	10445	11.930	11.628	3.223	1.149	23.853
Percentage of people aged between 45 and 64	10445	30.648	30.609	6.357	5.000	55.556
Percentage of people aged 65 and over	10445	19.536	19.048	7.973	0.000	51.634
Percentage of married people	10445	42.706	44.126	8.743	10.662	75.000
Percentage of single-family houses	10445	78.937	89.000	24.659	0.000	100.000
Percentage of owner-occupied houses	10445	68.991	74.000	20.836	0.000	100.000
Degree of urbanity	10445		4.000	1.475	1.000	5.000

4.2 Ordinary Least Squares and Global Moran's I

After the determination of the explanatory variables the OLS regression can be conducted. Before conducting this regression. the variable 'Degree of urbanity' has been merged into one with three categories instead of five and the variable is recoded into two dummy variables. The new variables are moderate degree of urbanity (MODURB) and strong degree of urbanity (STRURB). The distribution of residuals across the Netherlands is shown in appendix G, figure G.1. This figure shows the distribution of residuals across neighbourhoods in the Netherlands. Directly after this regression, the Global Moran's I is calculated to determine if residuals are spatially clustered over the study area. It turns out that there is a statistically significant spatial autocorrelation of regression residuals (see table 4.3. According to ESRI (n.d.-g), the reason for this spatial autocorrelation could be that there are key variables missing from the model. The consequence of spatially clustered residuals is that the model is biased.

In table 4.3, the model performance of this OLS regression is shown. From these statistics, it can be concluded that the model explains approximately 29 percent of the variation in the dependent variable. In the table, two R^2 values are listed, the difference between these two is that the adjusted R^2 takes into account the number of explanatory variables. The multiple R^2 will likely keep increasing when adding more explanatory variables to the model and the adjusted R^2 controls for that. So it could happen that the adjusted R^2 decreases when an extra explanatory variable is added. The Joint Wald Statistic tells something about the overall model significance, if this statistic is significant this indicates a significant model. The null hypothesis for this statistic is that the explanatory variables are effective in predicting the dependent variable. The Koenker statistic indicates whether the explanatory variables have a consistent relation with the dependent variables over all the neighbourhoods in the Netherlands. If this statistic is significant, a GWR analysis could most likely have an added value. The Jarque-Bera statistic is to test if residuals are spatially correlated, together with the Global Moran's I statistic, it can be concluded that the predictions are biased since both statistics show not normally distributed residuals (ESRI, n.d.-f).

Table 4.3: OLS model performance (model 1)

	Value	Probability
Akaike's Information Criterion	-28,020.66	
Multiple R-Squared	0.2940	
Adjusted R-Squared	0.2934	
Joint Wald Statistic	5756.278	0.00**
Koenker (BP) Statistic	205.46	0.00**
Jarque-Bera Statistic	7860.44	0.00**
Global Moran's I test for residuals	66.70	0.00**
p-values: ** p < 0.01, *p < 0.05		

In table 4.4, the OLS estimation results are shown. This includes the following statistics: coefficient, probability and VIF. The coefficient reflects the strength and type of relationship between the explanatory variable and the dependent variable. According to the results, the following variables have a significant negative relation with PV adoption: percentage of males, percentage of people aged between 15 and 24 and percentage of people aged 65 and over. The variables that have a significant positive relation are percentage of people aged between 45 and 64, percentage of married people, percentage of single-family houses, percentage of owner-occupied houses, moderate degree of urbanity and strong degree of urbanity. The VIF-values are all below 10, which means that there is little redundancy among explanatory variables. The coefficient for the percentage of males is negative, which is in conflict with the expectation. It was expected that when the percentage of males increases the adoption would also increase. The predictions of the different age categories are in line with the expectation, with a positive relation for people aged between 45 and 64. For all the other variables, the type of relation is in line with the expectation (shown in table 2.1). A strong degree of urbanity has a coefficient of 0.0302. Also, a moderate degree of urbanity has a positive and significant relation with the adoption of PV panels. This positive influence is in line with what was found in literature, however, some contrary findings were already found in literature since populations and housing density have negative relations. The relation between predictor variables and the dependent variable is shown in equation 4.3.

$$\begin{aligned} y = 0.0396 + (-0.0006 * PERCNUMMALES) + (-0.0014 * PERCNUM15\_24) + \\ (0.0004 * PERCNUM45\_64) + (-0.0008 * PERCNUM65) + (0.0004 * \\ PERCNUMMAR) + (0.0008 * PERC1FAMH) + (0.0005 * PERCOWNED) + \\ (0.0245 * MODURB) + (0.0302 * STRURB) + \varepsilon \end{aligned} \tag{4.3}$$

Table 4.4: OLS estimation results (model 1)

Variable	Coefficient	Probability	VIF
Intercept	0.0396	0.0149*	-
Percentage of males	-0.0006	0.0363**	1.4279
Percentage of people aged between 15 and 24	-0.0014	0.0000**	1.2240
Percentage of people aged between 45 and 64	0.0004	0.0061*	1.5889
Percentage of people aged 65 and over	-0.0008	0.0000**	1.6400
Percentage of married people	0.0004	0.0047**	3.0043
Percentage of single-family houses	0.0008	0.0000**	2.7453
Percentage of owner-occupied houses	0.0005	0.0000**	2.8645
Moderate degree of urbanity	0.0245	0.0000**	1.8478
Strong degree of urbanity	0.0302	0.0000**	2.7055
p-values: ** $p < 0.01$, *$p < 0.05$			

4.3 Geographically Weighted Regression

In this section, the GWR will be discussed. The input that this tool uses is shown below. First, the input features need to be completed, this includes the data with its dependent and explanatory variables. Secondly, the model type needs to be determined. ESRI (n.d.-e) states that for models with a dependent variable that can take a wide range of values, the continuous (Gaussian) model type is the best fitted. The other two possible model types are the binary model type and count model type. The binary model type is for models that have a binary dependent variable. The count model type is for models where the dependent variable represents an occurrence and is discrete. Thirdly, the bandwidth for the model needs to be identified, this defines the neighbourhood for the local regression. This is similar to the adaptive and fixed method (explained in chapter 3), where the option number of neighbours takes into account the density of neighbours and where the distance band does not. This second one incorporates neighbourhoods that fall within a radius. So this method does not take into account the density of neighbourhoods, whereas the first method does. For this research the number of neighbours is chosen, since the size of neighbourhoods can differ a lot in the Netherlands, e.g. smallest neighbourhood is 0.014 km^2 and the largest is 130.142 km^2 (Statistics Netherlands, 2018). The problem that occurs when using the distance band is that the regressions in less dense areas are conducted over a relatively low number of neighbourhoods and for more dense areas the number of neighbourhoods in the regression is large. The value for the number of neighbourhoods can be determined by minimizing the Akaike's Information Criterion or it can be determined by the user (ESRI, n.d.-b, n.d.-e). First, the automatic method is used, however, no results can be generated since the regression is unable to estimate at least one regression because of data redundancy. For that reason, the minimum number of neighbours is increased. This minimum number is step-wise increased by 50 neighbourhoods and the first model that gave results was with a minimum of 300 neighbourhoods. In table 4.5, the model performance of this GWR is shown, it shows that the number of neighbours for each local regression is 301 (this is determined by minimizing the Akaike's Information Criterion and with a minimum of 300 neighbourhoods). It also shows the value for this criterion, this value is used to compare different models with each other. The model fit can also be determined on basis of the adjusted R^2, for this model the value is 0.3894. This means that the model explains approximately 39 percent of the variation in the dependent variable. Comparing this value with the adjusted R^2 of the OLS (model 1), the value increased by approximately 10 percent points. In figure 4.1, the local adjusted

R^2 values are shown, the values range from 0.1555 to 0.5443. It can be observed that there are several clustered areas distributed over the Netherlands of adjusted R^2 values.

Table 4.5: GWR model performance (model 1)

	Value
Number of neighbours	301
Akaike's Information Criterion	-29,414.27
R-Squared	0.4064
Adjusted R-Squared	0.3894

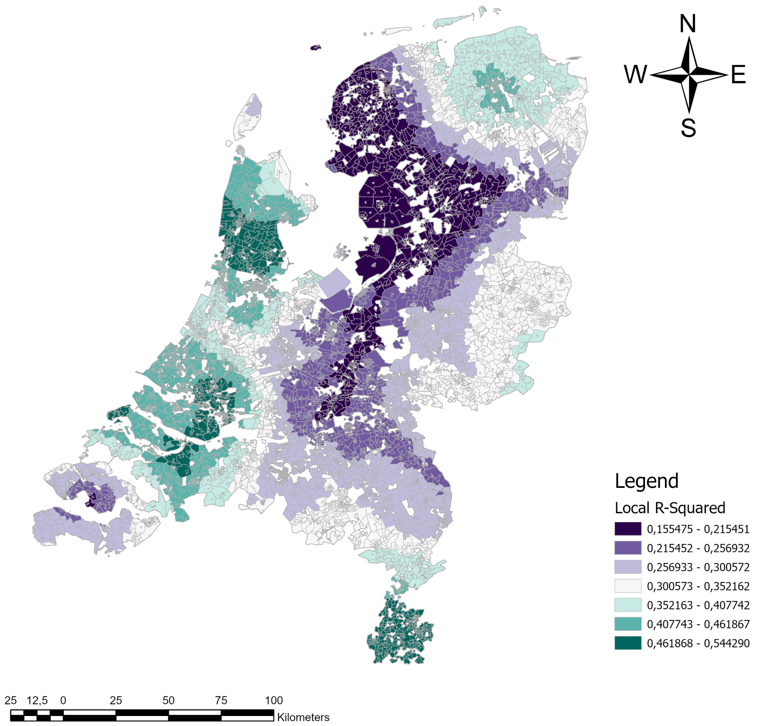

Figure 4.1: GWR Local Adjusted R^2

However, a GWR generates a condition number in the output table. According to ESRI (n.d.-b), This number indicates the amount of local collinearity in the model. Where in OLS regressions, results are biased when two or more variables are correlated with each other, in GWR, local collinearity occurs when a variable clusters spatially. If this condition number is greater than 30, results are biased because of local collinearity (ESRI, n.d.-b). In this model, the condition number ranges from approximately 2650 to 5150 with a mean of 3611.14, so this model exceeds the assumption of a maximum of 30. For this reason, a new model with fewer independent variables was generated. This resulted in a model with the following variables: percentage of people aged between 45 and 64, strong degree of urbanity and moderate degree of urbanity. These variables were chosen because the literature shows that they have a relatively strong influence. Since a new set of explanatory variables was constructed for the GWR, an OLS with the same set of variables was conducted as well. The model performances of these models are shown in table 4.6 and table 4.7. From the OLS (model 2) performance, it can be concluded that the model explains approximately 21 percent of the variation in the dependent variable. The Jarque-Bera statistic is to test if residuals are spatially correlated, together with the Global Moran's I statistic, it can be concluded that the predictions are biased since both statistics show not normally distributed residuals. The GWR model performance of model two shows that the statistics are a bit less in comparison with the first GWR model. The Akaike's Information Criterion has a value of -27,990.38 which is approximately 1500 more than the same criterion of the second model. The adjusted R^2 has decreased by approximately 10 percent points in comparison with the first GWR. However, the condition number of this model has decreased significantly, for this model this ranges from 27.74 to 93.96 with a mean of 42.93. Comparing the model performance of the second OLS model with the second GWR model. It stands out that this GWR model shows better overall performance because the adjusted R^2 is approximately 9 percent point higher and the Akaike's Information Criterion is approximately 1100 less. In the next part, these models will be explained in more detail. In table 4.6 and table 4.7, the Global Moran's I statistics is shown and indicated that there is less than 1% chance that the clustered pattern is a consequence of randomness. As stated earlier, this could be due to the fact that there are key variables missing from the model.

Table 4.6: OLS model performance (model 2)

	Value	Probability
Akaike's Information Criterion	-26,848.54	
Multiple R-Squared	0.2093	
Adjusted R-Squared	0.2090	
Joint Wald Statistic	3077.79	0.00**
Koenker (BP) Statistic	47.27	0.00**
Jarque-Bera Statistic	6009.80	0.00**
Global Moran's I test for residuals	63.20	0.00**
p-values: ** p < 0.01, *p < 0.05		

Table 4.7: GWR model performance (model 2)

	Value	Probability
Number of neighbours	301	
Akaike's Information Criterion	-27,990.38	
R-Squared	0.3024	
Adjusted R-Squared	0.2944	
Global Moran's I test for residuals	35.83	0.00**

In table 4.8, the estimation output of the OLS (model 2) is shown. All the explanatory variables have a significant positive relation with the dependent variable. The type of relations are in line with the previous OLS (model 1). The VIF-values are all below 10, which means that there is little redundancy among explanatory variables.

In table 4.9, the estimation output of the GWR (model 2) is shown. A side note needs to be made, since the condition number, that detects local collinearity in the model exceeds the boundary of 30 in this model as well, the results could be biased. Table 4.9 shows the statistics of the estimated coefficients, this includes the mean, median, standard deviation, minimum and maximum.

The spatial distribution of the variable 'percentage of people aged between 45 and 64' indicates that the coefficients range from -0.0003 to 0.0038, where the OLS had a coefficient of 0.0017. The coefficient of this OLS corresponds with the mean of the GWR model. Figure 4.3 shows the spatial distribution of coefficients for this variable. A cluster of relatively high coefficients can be observed in Groningen, Noord Holland and the southern part of Limburg. Low coefficients are more dispersed over the country, however, Zeeland and parts of Noord-Brabant, Gelderland, Drenthe, Overijssel and Friesland have the lowest coefficients. This indicates that there are differences in the strength of predictor variables over the Netherlands. In the regions with high coefficients, the particular variable has more influence on PV adoption. Thus, in these regions, people between the age of 45 and 64 have been found to adopt more solar panels.

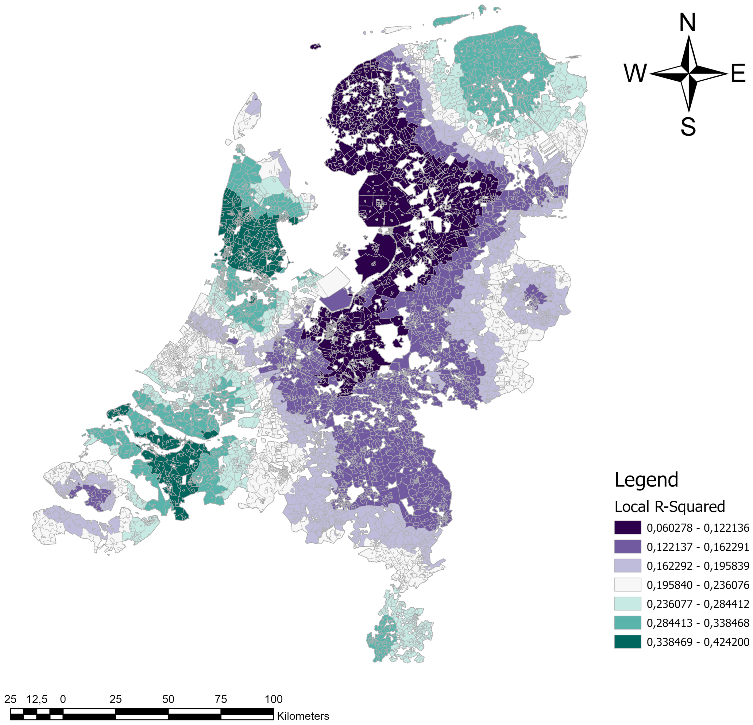

Figure 4.2: GWR Local Adjusted R^2 (model 2)

Both the variables 'moderate' and 'strong degree of urbanity' have a similar spatial distribution of coefficients, as shown in figure 4.4 and 4.5. With a centering of low coefficients in Utrecht and surrounded by higher values with the highest coefficients in Zeeland and Groningen. The values for the variable 'moderate degree of urbanity' range from 0.0233 to 0.0895 with a mean of 0.049. The mean of the coefficients is lower than the coefficient of the earlier conducted OLS. Secondly, the variable 'strong degree of urbanity' ranges from 0.0134 to 0.0960 with a mean of 0.055, which is also lower than the coefficient of the OLS. Some contrary findings were found in the literature review of the influence of population density, housing density and urbanisation on the adoption of PV panels. However, from the OLS regression and the GWR, the variables 'moderate' and 'strong degree of urbanity' show a positive influence on the dependent variable. Comparing the second OLS model with the second GWR model, both show similar predictions.

Table 4.8: OLS estimation results (model 2)

Variable	Coefficient	Probability	VIF
Intercept	0.0236	0.0000*	-
Percentage of people aged between 45 and 64	0.0017	0.0000*	1.2567
Moderate degree of urbanity	0.0525	0.0000**	1.4433
Strong degree of urbanity	0.0653	0.0000**	1.7380
p-values: ** p < 0.01, *p < 0.05			

Table 4.9: GWR estimation results (model 2)

Variable	N	Mean	Median	Std. Deviation	Min	Max
Intercept	10445	0.0285	0.0296	0.0256	-0.0355	0.0924
Percentage of people aged between 45 and 64	10445	0.0017	0.0016	0.0008	-0.0003	0.0038
Moderate degree of urbanity	10445	0.0486	0.0484	0.0118	0.0233	0.0895
Strong degree of urbanity	10445	0.0547	0.0547	0.0138	0.0134	0.0960
Condition number	10445	42.9340	40.4355	11.0102	27.7368	93.9592

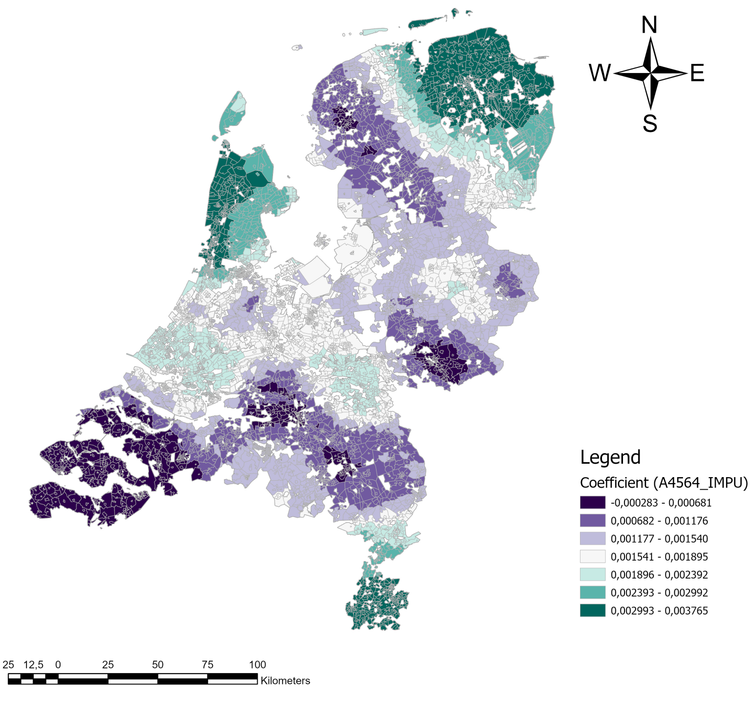

Figure 4.3: Coefficients of percentage of people aged between 45 to 64

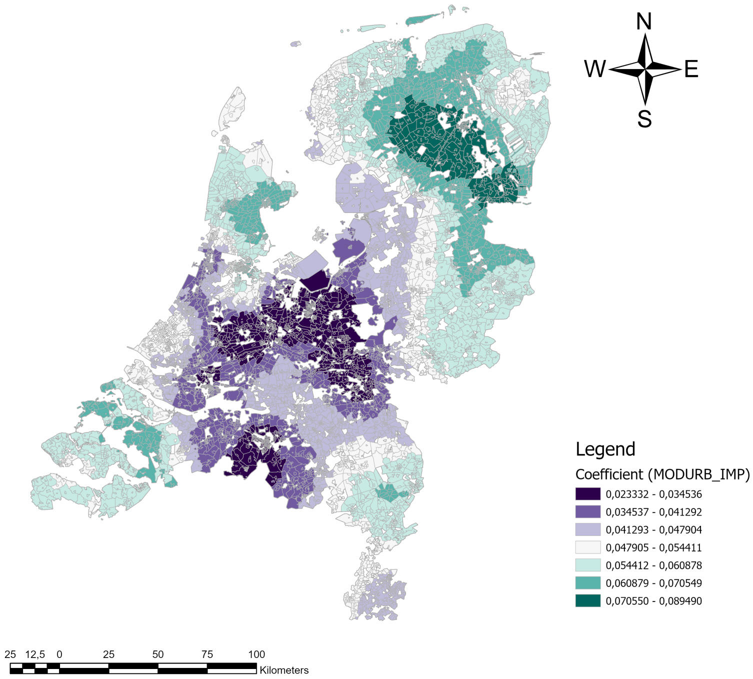

Figure 4.4: Coefficients of moderate degree of urbanity

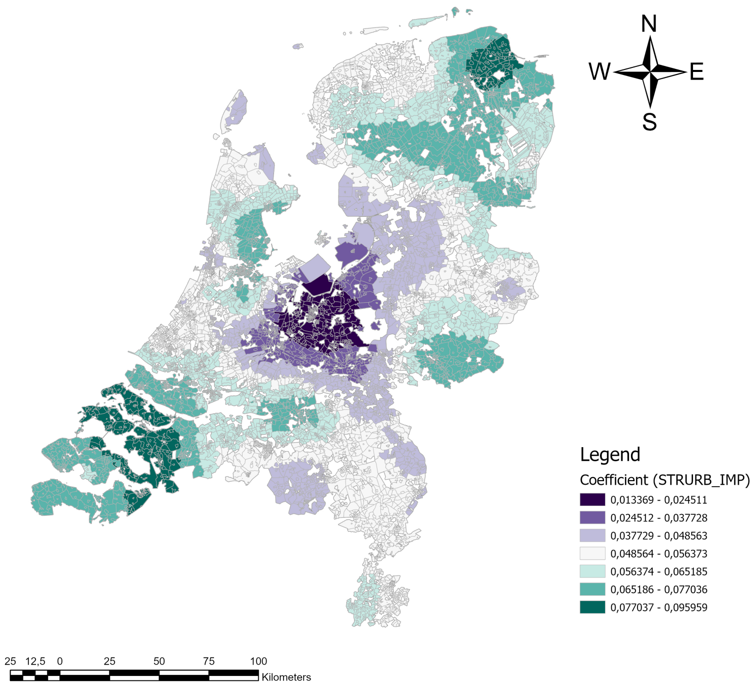

Figure 4.5: Coefficients of strong degree of urbanity

4.4 Conclusions

In table 4.10, the different models are compared with each other. The comparison includes some model fit statistics and a comparison of coefficients. The Akaike's Information Criterion for the first GWR model is the lowest, indicating that this model fits best the data. The adjusted R^2 is also an indicator for model fit and also for this value the first GWR scores best. For the coefficients, the different models show similar results, no unexpected type of relations can be observed looking at the mean values of the GWR and the coefficients of the OLS. Based on the OLS models, it can be concluded that the adoption of PV panels in the Dutch neighborhoods depends on the age of the inhabitants, the degree of urbanity, and according to the first OLS model also on the share of single-family houses and owner-occupied houses.

The second GWR model is considered the final result, due to the fact that it deviates the least from constraints, such as the Global Moran's I and the condition number. For this GWR with fewer explanatory variables, the following can be stated. For the age group 45 to 64, a cluster of relatively high coefficients can be observed in Groningen, Noord Holland and the southern part of Limburg. Low coefficients are more dispersed over the country however low values can be observed in Zeeland and parts of Noord-Brabant, Gelderland, Drenthe, Overijssel and Friesland. In the regions with high coefficients, the particular variable has more influence on PV adoption. Thus, in these regions, people between the age of 45 and 64 have been found to adopt more solar panels. Both the variables 'moderate' and 'strong degree of urbanity' have a similar spatial distribution of coefficients. With a centering of low coefficients in Utrecht and surrounded by higher values with the highest coefficients in Zeeland and Groningen.

Table 4.10: Comparing models

	Ordinary Least Squares (model 1)			Ordinary Least Squares (model 2)			Geographically Weighted Regression (model 1)			Geographically Weighted Regression (model 2)		
Akaike's Information Criterion	-28,020.66	-		-26,848.54	-		-29,414.27	-	-	-27,990.38	-	-
Adjusted R-Squared	0.29	-		0.21	-		0.39	-	-	0.29	-	-
Global Moran's I test for residuals	66.70	0.00**		63.20	0.00**		27.82	0.00**	-	35.83	0.00**	-
	b	**p**	**VIF**	**b**	**p**	**VIF**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
Intercept	0.0396	0.0149*	-	0.0236	0.00**		0.0452	-0.1764	0.2770	0.0285	-0.0355	0.0924
Percentage of males	-0.0006	0.0363**	1.4279				-0.0009	-0.0043	0.0026			
Percentage of people aged between 15 and 24	-0.0014	0.0000**	1.2240				-0.0006	-0.0037	0.0022			
Percentage of people aged between 45 and 64	0.0004	0.0061*	1.5889	0.0017	0.00**	1.2567	0.0001	-0.0018	0.0022	0.0017	-0.0003	0.0038
Percentage of people aged 65 and over	-0.0008	0.0000**	1.6400				-0.0010	-0.0022	0.0000			
Percentage of married people	0.0004	0.0047**	3.0043				0.0009	-0.0013	0.0028			
Percentage of single-family houses	0.0008	0.0000**	2.7453				0.0007	0.0001	0.0015			
Percentage of owner-occupied houses	0.0005	0.0000**	2.8645				0.0005	-0.0005	0.0013			
Moderate degree of urbanity	0.0245	0.0000**	1.8478	0.0525	0.00**	1.4433	0.0186	-0.0030	0.0493	0.0486	0.0233	0.0895
Strong degree of urbanity	0.0302	0.0000**	2.7055	0.0653	0.00**	1.7380	0.0181	-0.0187	0.0638	0.0547	0.0134	0.0960
Condition number	-	-	-	-	-	-	3611.14	2649.39	5155.55	42.93	27.74	93.96

p-values: ** $p < 0.01$. * $p < 0.05$

Chapter 5

Conclusion, discussion and recommendations

Firstly, this chapter discusses the main results of the research. Secondly, it discusses the process, limitations and the trustworthiness of the results and lastly some recommendations are given.

5.1 Conclusions

This research used different methods to investigate the influential factors in the spatial adoption of photovoltaic panels. Firstly, the existing literature is reviewed and a list of possible determinants is composed. The current literature provides a good basis for this research, as much research has already been done into factors that influence the adoption of solar panels. In general, these factors can be divided into categories. These categories are social demographic, social-economic, built environment, environmental concern and peer effects factors. The literature shows that certain factors have been studied more than others and that there are strong and less strong influences. Factors such as age, income and education level, unemployment, governmental incentives and environmental concern have a strong influence on the adoption of solar panels. Also, some built environment characteristics show strong relationships (e.g. urbanization). An important finding from this literature review is the contrary findings for some of the built environment characteristics. Population density, mostly measured through the number of people divided by land area, has a negative relation with adoption rates. This means that in less dense areas people are more likely to adopt PV panels. Possible reasons for this are the share of single and double family homes in these regions. Also, the shadowing is higher and available roof space is lower in urban areas, which also would support this result. On the other hand, urbanization, measured through people living in cities of a certain area divided by total people of that area, had a positive influence on PV adoption. The same goes for housing density, measured through the number of residential buildings per km^2. The result of this literature review was a list of possible determinants for PV panel adoption.

From the OLS performed in this study, the following conclusions can be drawn. It can be concluded that the model explains approximately 29 percent of the variation in the dependent variable. According to the results, the following variables have a significant negative relation with PV adoption: percentage of males, percentage of people aged between 15 and 24 and percentage of people aged 65 and over. The variables that have a significant positive relation are percentage of people aged between 45 and 64, percentage of married people, percentage of single-family houses, percentage of owner-occupied houses, moderate degree of urbanity and strong degree of urbanity. The degree of urbanity is measured through addresses per km^2. The coefficient for the percentage of males is negative, which conflicts with some of the previous findings. It was expected that when the percentage of males increases the adoption would also increase. However, Diamantopoulos et al. (2003) found a significant relation based on the following hypothesis: "Females are more likely to participate in green activities" (Diamantopoulos et al., 2003). Guta (2018) found also differences between males and females for adopters and non-adopters of PV panels. In the category adopters, the household head was 84% male and for the non-adopters, this was 93%. The study found indeed that males are less likely to adopt solar energy technology compared to females (Guta, 2018). It can be concluded from this that across literature there is no clear understanding of the influence of gender on PV adoption. The results for the different age categories are in line with the previous findings, with a positive relation for people aged between 45 and 64. For all the other variables, the type of relation is in line with the expectation. A strong degree of urbanity has a positive relation with the dependent variable. Also, a moderate degree of urbanity has a positive and significant relation with the adoption of PV panels. This positive influence is in line with what was found in literature, however, some contrary findings were already found in literature since population and housing density has negative relations.

From the GWR, some different results were found. Firstly, the model with the same explanatory variables as the OLS will be discussed. This model explains approximately 39 percent of the variation in the dependent variable. Comparing this value with the adjusted R^2 of the OLS, the value increased by approximately 10 percent points. However, this GWR exceeds the assumption of local collinearity. For this reason, a new model with fewer independent variables was sought. For this second GWR model, with fewer explanatory variables, the following can be stated. For the people aged between 45 and 64, a cluster of relatively high coefficients can be observed in Groningen, Noord Holland and the southern part of Limburg. Low coefficients are more dispersed over the country, however low values can be observed in Zeeland and parts of Noord-Brabant, Gelderland, Drenthe, Overijssel and Friesland. Both the variables 'moderate' and 'strong degree of urbanity' have a similar spatial distribution of coefficients. With a centering of low coefficients in Utrecht and surrounded by higher values with the highest coefficients in Zeeland and Groningen. Some contrary findings were found in the literature review of the influence of population density, housing density and urbanisation on the adoption of PV panels. However, from the OLS regression and the GWR, the variable 'degree of urbanity', measured through addresses per km^2, show a positive influence on the dependent variable.

5.2 Discussion

In this section, some issues will be discussed. First, the completeness of the data as published by Statistics Netherlands. The data set containing the information on the number of solar panels proved to be incomplete. About 20% of the neighbourhoods in the Netherlands have no value for the number of solar panels. For this reason, these neighbourhoods had to be removed from the data set for the analysis, what created gaps in the study area. In this study, a tool is used that fills in missing data based on values of neighbouring neighbourhoods. The number of neighbourhoods that it used to predict this value is calculated as the mean number of neighbourhoods per neighbourhood. The result was an average of seven neighbourhoods per neighbourhood, it could be that predictions are conducted on less than seven neighbourhoods, because of these missing neighbourhoods. After the use of this tool, the statistics of the filled variables are validated by looking at the basic descriptives, such as mean, standard deviation, minimum and maximum. The GWR uses the same method, so missing neighbourhoods could have an influence on the result.

Secondly, a number of essential independent variables were missing. These were missing for various reasons, for education level, there was no data available for the corresponding year, for income, there were too many missing values, the same goes for unemployment and peer effects, political preference was measured on a different scale and data on country and regional subsidies was unknown. Also, the number of missing values for the different variables is relatively high. For that reason, the set of explanatory variables that remained was not very large and some of these variables already showed inconsistency in literature.

Thirdly, during the analysis, the Global Moran's I showed that the residuals were correlated with each other. According to ESRI (n.d.-g), this could be a result of the fact that essential variables are missing from the model. This was the case for the global regression, but also for the local regression. The effect of this is that results are biased and that they are not fully confidential. Next to this issue in the analysis, the GWR showed high local collinearity. Where in OLS regressions results are biased when two or more variables are correlated with each other, in GWR, local collinearity occurs when a variable clusters spatially. If this condition number is greater than 30, which was true for all neighbourhoods, the results are biased because of local collinearity (ESRI, n.d.-b).

Fourthly, in the OLS regression, a significant negative relation was found for the percentage of males, this is in line with the findings of Diamantopoulos et al. (2003), Guta (2018). However, some other studies found positive relations between males and adoption rates (Bollinger & Gillingham, 2012; Jacksohn et al., 2019; Rahut et al., 2018). Also, some contrary findings were found for the variable 'degree of urbanity'. As discussed, the findings in the literature about populations density, housings density and urbanization differ a lot. In this study, a significant positive relation was found between a moderate and strong degree of urbanity and the adoption of PV panels. However, Balta-Ozkan et al. (2015), Balta-Ozkan et al. (2021), Graziano and Gillingham (2015), S. Müller and Rode (2013), Rode and Müller (2016) found negative relations between population density and PV adoption, while Schaffer and Brun (2015) found a positive relation between housing density and PV adoption. Also, Wallace and Wang (2006), Zahran et al. (2008) found a positive effect of urbanization on the number of solar-heated households. In conclusion, the results of this research are in line with some of the previous findings.

Fifthly, in comparison to other studies that investigated the same influence on PV panel adoption, higher adjusted R^2 values are observed. For example, Dharshing (2017) had in their study about household dynamics of technology adoption, an adjusted R^2 value of 0.57. This is significantly higher than the GWR model explained above. Another study, that uses the same method and dependent variable, had an adjusted R^2 value of 0.79, which is also a lot higher than an adjusted R^2 of 0.29 (Balta-Ozkan et al., 2021).

5.3 Recommendations

Since from literature it can be stated that the diffusion of PV panels is related to space, it is expected that a GWR gives new insights in comparison to OLS since this does not take into account the spatial component. The results of this research can determine more detailed information on the diffusion of PV panels across the Netherlands. Together with the information on the adoption factors, more targeted policies can be developed to increase the adoption of PV panels since it is an electricity source with a lot of potential in the residential sector. Also, information supply can be sharpened up since information supply is a determining factor and is related to the adoption stage of the population. Because of the diffusion of PV panels, the adoption stages of the population will also differ across the different regions.

Firstly, adoption can be improved by policies on more targeted population groups and on different spatial scales. The analysis showed that variables have different patterns when looking at the Netherlands, thus proving that there are differences in factors influencing the adoption of solar panels in different parts of the country. With this information, municipalities and private companies can focus on a segment of the population that are more sensitive to policies regarding PV panel adoption. This can be done through marketing or subsidies for certain groups or places. In Groningen, the influence of people aged 45-64 is much higher than in Zeeland. So, it showed that this group of people in Groningen has a higher influence on the adoption of PV panels. Triggering this group in areas with less influence could increase the total adoption rate.

The second recommendation is about the availability of reliable data. As said above, the quality of the data from Statistics Netherlands can be improved. This includes the number of missing values, but also the quality and reliability of the data itself. For example, unemployment in a neighbourhood had already 12% missing values when retrieving the data from the source. It is recommended to find ways to improve the quality of this data.

Thirdly, from this research, several significant relations have been determined. However, some doubts are there since these significant relations could also be a result of an indirect relation. Future research needs to be done on these indirect and direct relations with the PV panel adoption.

Fourthly, as the use of GWR was decided upon at the beginning of this project, it is highly recommended to investigate other techniques. An alternative approach might be to use a regression technique including interactions between the independent variables and province indicators or apply tree analysis with provincial indicators in addition to the independent variables. In addition, the large difference between neighbourhoods can also have an influence on the result, so investigating relations on smaller scales or less differentiating areas could also improve the model. This research showed that the influential factors of PV adoption on a global scale are well explored. However, on the local scale, there is a lot more research needed. From literature and from this research, it has been found that variables differ spatially. To investigate the relations on the local scale, the use of different spatial regression techniques need to be explored and investigated.

Bibliography

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, *89*(1), 3–25. https://doi.org/10.1111/j.1435-5957.2010.00279.x

Anselin, L. (2017). Intro & Review. https://www.youtube.com/watch?v=50kc3F0aE6Y

Ashok, S. (2021). Solar energy. https://www.britannica.com/science/solar-energy

Austin, F. (2016). *The World Energy Trilemma: Progress Toward Balanced, Sustainable Energy Remains Slow* (tech. rep.). https://www.oliverwyman.com/content/dam/oliver-wyman/global/en/2016/mar/Energy-Journal/Energy%20Trilemma-online.pdf

Balcombe, P., Rigby, D. & Azapagic, A. (2013). Motivations and barriers associated with adopting microgeneration energy technologies in the UK. *Renewable and Sustainable Energy Reviews*, *22*, 655–666. https://doi.org/10.1016/j.rser.2013.02.012

Balta-Ozkan, N., Yildirim, J. & Connor, P. M. (2015). Regional distribution of photovoltaic deployment in the UK and its determinants: A spatial econometric approach. *Energy Economics*, *51*, 417–429. https://doi.org/10.1016/j.eneco.2015.08.003

Balta-Ozkan, N., Yildirim, J., Connor, P. M., Truckell, I. & Hart, P. (2021). Energy transition at local level: Analyzing the role of peer effects and socio-economic factors on UK solar photovoltaic deployment. *Energy Policy*, *148*, 1–15. https://doi.org/10.1016/j.enpol.2020.112004

Bamberg, S. (2003). How does environmental concern influence specific environmentally related behaviors? A new answer to an old question. *Journal of Environmental Psychology*, *23*, 21–32.

Bamberg, S. & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. *Journal of Environmental Psychology*, *27*, 14–25. https://doi.org/10.1016/j.jenvp.2006.12.002

Basile, R., Durbán, M., Mínguez, R., María Montero, J. & Mur, J. (2014). Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control*, *48*, 229–245. https://doi.org/10.1016/j.jedc.2014.06.011

Bauner, C. & Crago, C. L. (2015). Adoption of residential solar power under uncertainty: Implications for renewable energy incentives. *Energy Policy*, *86*, 27–35. https://doi.org/10.1016/j.enpol.2015.06.009

Best, R., Burke, P. J. & Nishitateno, S. (2019a). Evaluating the effectiveness of Australia's Small-scale Renewable Energy Scheme for rooftop solar. *Energy Economics*, *84*, 1–11. https://doi.org/10.1016/j.eneco.2019.104475

Best, R., Burke, P. J. & Nishitateno, S. (2019b). Understanding the determinants of rooftop solar installation: evidence from household surveys in Australia. *Australian Journal of Agricultural and Resource Economics*, *63*, 922–939. https://doi.org/10.1111/1467-8489.12319

Bollinger, B. & Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, *31*(6), 900–912. https://doi.org/10.1287/mksc.1120.0727

Bose, B. K. (2010). Global warming: Energy, environmental pollution, and the impact of power electronics. *IEEE Industrial Electronics Magazine*, *4*(1), 6–17. https://doi.org/10.1109/MIE.2010.935860

Briguglio, M. & Formosa, G. (2017). When households go solar: Determinants of uptake of a Photovoltaic Scheme and policy insights. *Energy Policy*, *108*, 154–162. https://doi.org/10.1016/j.enpol.2017.05.039

Brown, B. J., Hanson, M. E., Liverman, D. M. & Merideth, R. W. (1987). Global sustainability: Toward definition. *Environmental Management*, *11*(6), 713–719. https://doi.org/10.1007/BF01867238

Brunsdon, C., Fotheringham, S. & Chariton, M. (1998). Geographically Weighted Regression-Modelling Spatial Non-Stationarity. *Journal of the Royal Statistical Society*, *47*(3), 431–443.

Dastrup, S. R., Graff Zivin, J., Costa, D. L. & Kahn, M. E. (2012). Understanding the Solar Home price premium: Electricity generation and "Green" social status. *European Economic Review*, *56*(5). https://doi.org/10.1016/j.euroecorev.2012.02.006

Davidson, C., Drury, E., Lopez, A., Elmore, R. & Margolis, R. (2014). Modeling photovoltaic diffusion: An analysis of geospatial datasets. *Environmental Research Letters*, *9*, 1–15. https://doi.org/10.1088/1748-9326/9/7/074009

Davis, F. D. (1985). *A technolgy acceptance model for empirically testing new end-user information systems: theory and results* (Doctoral dissertation). Wayne State.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–339. https://doi.org/10.2307/249008

Dharshing, S. (2017). Household dynamics of technology adoption: A spatial econometric analysis of residential solar photovoltaic (PV) systems in Germany. *Energy Research and Social Science*, *23*, 113–124. https://doi.org/10.1016/j.erss.2016.10.012

Diamantopoulos, A., Schlegelmilch, B. B., Sinkovics, R. R. & Bohlen, G. M. (2003). Can socio-demographics still play a role in profiling green consumers? A review of the evidence and an empirical investigation. *Journal of Business Research*, *56*, 465–480. https://doi.org/10.1016/S0148-2963(01)00241-7

Diaz-Rainey, I. & Ashton, J. K. (2011). Profiling potential green electricity tariff adopters: Green consumerism as an environmental policy tool? *Business Strategy and the Environment*, *20*, 456–470. https://doi.org/10.1002/bse.699

Dincer, I. (2000). Renewable energy and sustainable development: A crucial review. *Renewable & Sustainable Energy Reviews*, *4*(2), 157–175. https://doi.org/10.1016/S1364-0321(99)00011-8

Dolowitz, D. P. & Marsh, D. (2000). Learning From Abroad: The Role of Policy Transfer in Contemporary Policy-Making. *Governance*, *13*(1), 5–23. https://doi.org/10.1111/0952-1895.00121

ESRI. (n.d.-a). Fill Missing Values (Space Time Pattern Mining). https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/fillmissingvalues.htm

ESRI. (n.d.-b). Geographically Weighted Regression (GWR) (Spatial Statistics). https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-geographicallyweightedregression-works.htm

ESRI. (n.d.-c). How Exploratory Regression works. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-exploratory-regression-works.htm

ESRI. (n.d.-d). How Fill Missing Values works. https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/learnmorefillmissingvalues.htm

ESRI. (n.d.-e). How Geographically Weighted Regression (GWR) works. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-geographicallyweightedregression-works.htm

ESRI. (n.d.-f). How OLS Regression Works. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-ols-regression-works.htm

ESRI. (n.d.-g). Ordinary Least Squares (OLS)(Spatial Statistics). https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/ordinary-least-squares.htm

ESRI. (n.d.-h). Regression analysis basics. https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/regression-analysis-basics.htm

ESRI. (n.d.-i). Shapefiles. https://doc.arcgis.com/en/arcgis-online/reference/shapefiles.htm

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th). London, SAGE Publications Ltd.

Finn, K. (2017). How to calculate outliers. https://sciencing.com/calculate-outliers-5201412.html

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behaviour: An introduction to theory and research*. Reading, Addison-Wesley.

Fotheringham, A. S., Brunsdon, C. & Chariton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons Inc.

Frost, J. (n.d.). 5 ways to find outliers in your data. https://statisticsbyjim.com/basics/outliers/

Georg, L., Auer, H., Fleischhacker, A., Schwabeneder, D., Dallinger, B. & Moisl, F. (2018). *Existing and future PV prosumer concepts* (tech. rep.). Vienna.

Gezahegn, T. W., Gebregiorgis, G., Gebrehiwet, T. & Tesfamariam, K. (2018). Adoption of renewable energy technologies in rural Tigray, Ethiopia: An analysis of the impact of cooperatives. *Energy Policy*, *114*, 108–113. https://doi.org/10.1016/j.enpol.2017.11.056

Graziano, M., Fiaschetti, M. & Atkinson-Palombo, C. (2019). Peer effects in the adoption of solar energy technologies in the United States: An urban case study. *Energy Research and Social Science*, *48*, 75–84. https://doi.org/10.1016/j.erss.2018.09.002

Graziano, M. & Gillingham, K. (2015). Spatial patterns of solar photovoltaic system adoption: The influence of neighbors and the built environment. *Journal of Economic Geography*, *15*, 815–839. https://doi.org/10.1093/jeg/lbu036

Guta, D. D. (2018). Determinants of household adoption of solar energy technology in rural Ethiopia. *Journal of Cleaner Production*, *204*, 193–204. https://doi.org/10.1016/j.jclepro.2018.09.016

Hammer, A., Heinemann, D., Hoyer, C., Kuhlemann, R., Lorenz, E., Müller, R. & Beyer, H. G. (2003). Solar energy assessment using remote sensing technologies. *Remote Sensing of Environment*, *86*, 423–432. https://doi.org/10.1016/S0034-4257(03)00083-X

Health Knowledge. (n.d.). Linear regression and correlation. https://www.healthknowledge.org.uk/e-learning/statistical-methods/specialists/linear-regression-correlation

Hofierka, J., Kaòuk, J. & Gallay, M. (2014). The spatial distribution of photovoltaic power plants in relation to solar resource potential: the case of Czech Republic and Slovakia. *Moravian Geographical Reports*, *22*(2), 26–33. https://doi.org/10.2478/mgr-2014-0009

Hoogwijk, M. & Graus, W. (2008). *Global potential of renewable energy sources: A Literature assessment* (tech. rep. March).

Huijben, J. C. & Verbong, G. P. (2013). Breakthrough without subsidies? PV business model experiments in the Netherlands. *Energy Policy*, *56*, 362–370. https://doi.org/10.1016/j.enpol.2012.12.073

Jacksohn, A., Grösche, P., Rehdanz, K. & Schröder, C. (2019). Drivers of renewable technology adoption in the household sector. *Energy Economics*, *81*, 216–226. https://doi.org/10.1016/j.eneco.2019.04.001

Jager, W. (2006). Stimulating the diffusion of photovoltaic systems: A behavioural perspective. *Energy Policy*, *34*, 1935–1943. https://doi.org/10.1016/j.enpol.2004.12.022

Jayaweera, N., Jayasinghe, C. L. & Weerasinghe, S. N. (2018). Local factors affecting the spatial diffusion of residential photovoltaic adoption in Sri Lanka. *Energy Policy*, *119*(April 2017), 59–67. https://doi.org/10.1016/j.enpol.2018.04.017

Kabir, E., Kumar, P., Kumar, S., Adelodun, A. A. & Kim, K. H. (2018). Solar energy: Potential and future prospects. *Renewable and Sustainable Energy Reviews*, *82*, 894–900. https://doi.org/10.1016/j.rser.2017.09.094

Kaygusuz, K. (2002). Renewable and sustainable energy use in Turkey: A review. *Renewable and Sustainable Energy Reviews*, *6*(4), 339–366. https://doi.org/10.1016/S1364-0321(01)00007-7

Keirstead, J. (2007). Behavioural responses to photovoltaic systems in the UK domestic sector. *Energy Policy*, *35*, 4128–4141. https://doi.org/10.1016/j.enpol.2007.02.019

Kent, E. P. (2015). The concepts of Sustainability (1st ed.), In *Sustainability* (1st ed.). Cambridge, Mit Press Ltd.

Kidd, C. V. (1992). The evolution of sustainability. *Journal of Agricultural and Environmental Ethics*, *5*(1), 1–26. https://doi.org/10.1007/BF01965413

Kramer, A. & Segers, R. (2018). *Zonnestroom naar regio* (tech. rep.). Statistics Netherlands. Den Haag.

Kurdgelashvili, L., Shih, C. H., Yang, F. & Garg, M. (2019). An empirical analysis of county-level residential PV adoption in California. *Technological Forecasting and Social Change*, *139*, 321–333. https://doi.org/10.1016/j.techfore.2018.11.021

Labouret, A. & Villoz, M. (2010). Some basic questions on photovoltaics, In *Solar photovoltaic energy*. Stevenage, The Institution of Engineering; Technology. https://doi.org/10.1049/pbrn009e{_}ch1

Macovei, O.-I. (2015). Applying the Theory of Planned Behavior in Predicting Pro-environmental Behavior: The Case of Energy Conservation. *Acta Universitatis Danubius. Œconomica*, *11*(4), 15–32.

Müller, J. & Trutnevyte, E. (2020). Spatial projections of solar PV installations at subnational level: Accuracy testing of regression models. *Applied Energy*, *265*, 1–13. https://doi.org/10.1016/j.apenergy.2020.114747

Müller, S. & Rode, J. (2013). The adoption of photovoltaic systems in Wiesbaden, Germany. *Economics of Innovation and New Technology*, *22*(5). https://doi.org/10.1080/10438599.2013.804333

Mundaca, L. & Samahita, M. (2020). What drives home solar PV uptake? Subsidies, peer effects and visibility in Sweden. *Energy Research and Social Science*, *60*, 1–14. https://doi.org/10.1016/j.erss.2019.101319

Negro, S. O., Vasseur, V., Sark, W. G. J. H. M. V. & Hekkert, M. P. (2009). *Understanding innovation system build up : the rise and fall of the Dutch PV Innovation System*. http://igitur-archive.library.uu.nl/milieu/2010-0324-200149/UUindex.html

Negro, S. O., Vasseur, V., Van Sark, W. G. & Hekkert, M. P. (2012). Solar eclipse: The rise and 'dusk' of the Dutch PV innovation system. *International Journal of Technology, Policy and Management*, *12*(2-3), 135–157. https://doi.org/10.1504/IJTPM.2012.046923

Overheid.nl. (2003). Tijdelijke Regeling Energiepremie. https://wetten.overheid.nl/BWBR0014481/2003-01-01/0/Bijlage2

Palm, A. (2017). Peer effects in residential solar photovoltaics adoption—A mixed methods study of Swedish users. *Energy Research and Social Science*, *26*, 1–10. https://doi.org/10.1016/j.erss.2017.01.008

Palm, J. & Eriksson, E. (2018). Residential solar electricity adoption: how households in Sweden search for and use information. *Energy, Sustainability and Society*, *8*(14), 1–9. https://doi.org/10.1186/s13705-018-0156-1

Parsad, C., Mittal, S. & Krishnankutty, R. (2020). A study on the factors affecting household solar adoption in Kerala, India. *International Journal of Productivity and Performance Management*, *69*(8), 1695–1720. https://doi.org/10.1108/IJPPM-11-2019-0544

Phillips, L. (2019). Solar energy, In *Managing global warming: An interface of technology and human issues*. McLean, Academic Press. https://doi.org/10.1016/B978-0-12-814104-5.00009-0

Purplemath. (n.d.). Interquartile Ranges & Outliers. https://www.purplemath.com/modules/boxwhisk3.htm

Purvis, B., Mao, Y. & Robinson, D. (2019). Three pillars of sustainability: in search of conceptual origins. *Sustainability Science*, *14*(3), 681–695. https://doi.org/10.1007/s11625-018-0627-5

Qureshi, T. M., Ullah, K. & Arentsen, M. J. (2017). Factors responsible for solar PV adoption at household level: A case of Lahore, Pakistan. *Renewable and Sustainable Energy Reviews*, *78*, 754–763. https://doi.org/10.1016/j.rser.2017.04.020

Rahut, D. B., Mottaleb, K. A., Ali, A. & Aryal, J. (2018). The use and determinants of solar energy by Sub-Saharan African households. *International Journal of Sustainable Energy*, *37*(8), 718–735. https://doi.org/10.1080/14786451.2017.1323897

Richter, L. L. (2013). *Social Effects in the Diffusion of Solar Photovoltaic Technology in the UK* (tech. rep.).

Robinson, S. A. & Rai, V. (2015). Determinants of spatio-temporal patterns of energy technology adoption: An agent-based modeling approach. *Applied Energy*, *151*. https://doi.org/10.1016/j.apenergy.2015.04.071

Rode, J. & Müller, S. (2016). *Spatio-temporal variation in peer effects - The case of rooftop photovoltaic systems in Germany* (tech. rep.). Vienna.

Rode, J. & Weber, A. (2016). Does localized imitation drive technology adoption? A case study on rooftop photovoltaic systems in Germany. *Journal of Environmental Economics and Management*, *78*, 38–48. https://doi.org/10.1016/j.jeem.2016.02.001

Rogers, E. (1962). *Diffusion of Innovations* (1st ed.). New Yotk, Free Press.

Rogers, E. (2003). *Diffusion of Innovations* (5th ed.). New York, Free Press.

Rosenberg, M. & Hovland, C. (1960). *Cognitive, affective, and behavioral components of attitudes.* New Haven, Yale University Press.

Sardianou, E. & Genoudi, P. (2013). Which factors affect the willingness of consumers to adopt renewable energies? *Renewable Energy*, *57*, 1–4. https://doi.org/10.1016/j.renene.2013.01.031

Schaffer, A. J. & Brun, S. (2015). Beyond the sun - Socioeconomic drivers of the adoption of small-scale photovoltaic installations in Germany. *Energy Research and Social Science*, *10*, 220–227. https://doi.org/10.1016/j.erss.2015.06.010

Schelly, C. & Letzelter, J. C. (2020). Examining the key drivers of residential solar adoption in upstate New York. *Sustainability (Switzerland)*, *12*, 1–13. https://doi.org/10.3390/su12062552

Sharp, T. (2021). How big is Earth. https://www.space.com/17638-how-big-is-earth.html

Shi, L., Zhou, W. & Kristrom, B. (2013). Residential demand for electricity. *Environmental Economics*, *4*(1), 51–62.

Sigrin, B., Pless, J. & Drury, E. (2015). Diffusion into new markets: Evolving customer segments in the solar photovoltaics market. *Environmental Research Letters*, *10*, 1–8. https://doi.org/10.1088/1748-9326/10/8/084001

Snape, J. R. (2016). Spatial and temporal characteristics of PV adoption in the UK and their implications for the smart grid. *Energies*, *9*, 1–18. https://doi.org/10.3390/en9030210

SolarGIS. (2019). Solar resource maps of Netherlands. https://solargis.com/maps-and-gis-data/download/netherlands

Sommerfeld, J., Buys, L., Mengersen, K. & Vine, D. (2017). Influence of demographic variables on uptake of domestic solar photovoltaic technology. *Renewable and Sustainable Energy Reviews*, *67*, 315–323. https://doi.org/10.1016/j.rser.2016.09.009

Statistics Netherlands. (2018). Kerncijfers wijken en buurten 2018. https://www.cbs.nl/nl-nl/maatwerk/2018/30/kerncijfers-wijken-en-buurten-2018

Statistics Netherlands. (2019). Zonnestroom; vermogen zonnepanelen woningen, wijken en buurten, 2016. https://www.cbs.nl/nl-nl/cijfers/detail/84131NED?q=Veenendaal

Statistics Netherlands. (2020a). *Hernieuwbare energie in Nederland 2019* (tech. rep.). Den Haag. https://www.cbs.nl/nl-nl/publicatie/2015/40/hernieuwbare-energie-in-nederland-2014

Statistics Netherlands. (2020b). Zonnestroom; vermogen bedrijven en woningen, regio (indeling 2018), 2012-2018. https://www.cbs.nl/nl-nl/cijfers/detail/84518NED

Statistics Netherlands. (2020c). Zonnestroom; vermogen zonnepanelen woningen, wijken en buurten, 2018. https://www.cbs.nl/nl-nl/cijfers/detail/84772NED?q=zonnestroom

Statistics Netherlands. (2020d). Zonnestroom;vermogen zonnepanelen woningen, wijken en buurten, 2017. https://www.cbs.nl/nl-nl/cijfers/detail/84517NED?q=bouwen

Sunter, D. A., Castellanos, S. & Kammen, D. M. (2019). Disparities in rooftop photovoltaics deployment in the United States by race and ethnicity. *Nature Sustainability*, *2*, 71–76. https://doi.org/10.1038/s41893-018-0204-z

Šúri, M., Huld, T. A., Dunlop, E. D. & Ossenbrink, H. A. (2007). Potential of solar electricity generation in the European Union member states and candidate countries. *Solar Energy*, *81*, 1295–1305. https://doi.org/10.1016/j.solener.2006.12.007

Tarde, G. (1903). *The laws of imitation*. New York, Henry Holt; Company.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234–240.

Ugulu, A. I. (2019). Motives for solar photovoltaic (PV) adoption in urban Nigeria. *International Journal of Sustainable Energy Planning and Management*, *21*, 19–34. https://doi.org/10.1088/1755-1315/385/1/012012

Unruh, G. C. (2002). Escaping carbon lock-in. *Energy Policy*, *30*(4), 317–325. https://doi.org/10.1016/S0301-4215(01)00098-2

van der Kam, M. J., Meelen, A. A., van Sark, W. G. & Alkemade, F. (2018). Diffusion of solar photovoltaic systems and electric vehicles among Dutch consumers: Implications for the energy transition. *Energy Research and Social Science*, *46*, 68–85. https://doi.org/10.1016/j.erss.2018.06.003

Vasseur, V. & Kemp, R. (2015). The adoption of PV in the Netherlands: A statistical analysis of adoption factors. *Renewable and Sustainable Energy Reviews*, *41*, 483–494. https://doi.org/10.1016/j.rser.2014.08.020

Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478.

Verhees, B., Raven, R., Veraart, F., Smith, A. & Kern, F. (2013). The development of solar PV in the Netherlands: A case of survival in unfriendly contexts. *Renewable and Sustainable Energy Reviews*, *19*, 275–289. https://doi.org/10.1016/j.rser.2012.11.011

Wallace, W. & Wang, Z. (2006). Solar Energy in China: Development Trends for Solar Water Heaters and Photovoltaics in the Urban Environment. *Bulletin of Science, Technology & Society*, *26*(2), 135–140. https://doi.org/10.1177/0270467606286948

Welsch, H. & Kühling, J. (2009). Determinants of pro-environmental consumption: The role of reference groups and routine behavior. *Ecological Economics*, *69*, 166–176. https://doi.org/10.1016/j.ecolecon.2009.08.009

Wheeler, D. C. & Páez, A. (2010). Geographically Weighted Regression, In *Handbook of applied spatial analysis*. Springer.

World Commission on Environment and Development. (1987). *Our Common Future: Report of the World Commission on Environment and Development* (tech. rep.). https://doi.org/10.1590/S0104-42302009000600016

Zahran, S., Brody, S., Vedlitz, A., Lacy, M. & Schelly, C. L. (2008). Greening local energy: Explaining the geographic distribution of household solar energy use in the United States. *Journal of the American Planning Association*, *74*(4), 419–434. https://doi.org/10.1080/01944360802310594

Zhang, Y., Song, J. & Hamori, S. (2011). Impact of subsidy policies on diffusion of photovoltaic power generation. *Energy Policy*, *39*(4). https://doi.org/10.1016/j.enpol.2011.01.021

Appendix A

Adoption mechanisms

In this appendix, several visual representations of adoption mechanisms will be shown.

A.1 Theory of Reasoned Action

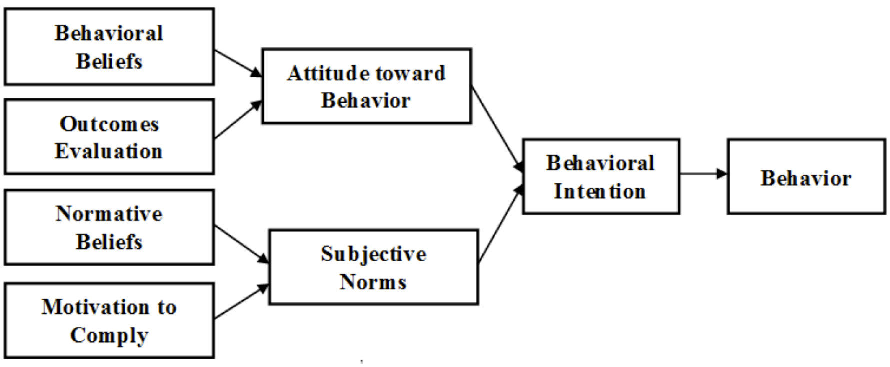

Figure A.1: Theory of Reasoned Action (Fishbein & Ajzen, 1975)

A.2 Technology Acceptance Model

In figure A.2, x1 to x3 are external variables that influence the perceived usefulness and ease of use.

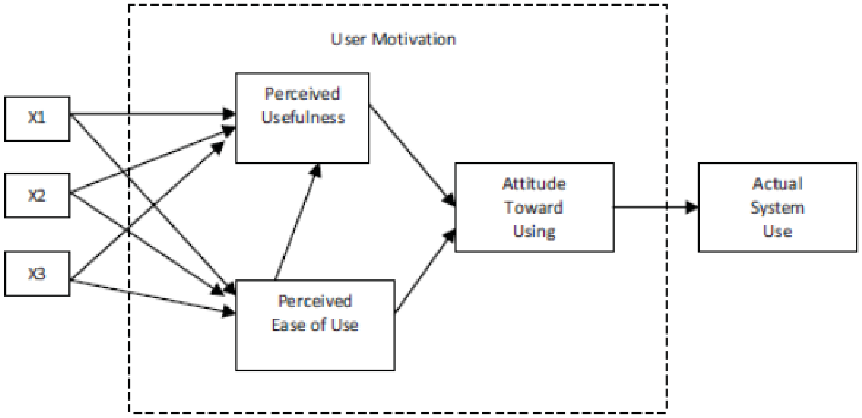

Figure A.2: Technology Acceptance Model (Davis, 1985)

A.3 Unified Theory of Acceptance and Use of Technology

The Unified Theory of Acceptance and Use of Technology is a combination of the following models:

- Theory of Reasoned Action
- Technology Acceptance Model
- Motivational Model
- Theory of Planned Behavior
- Combined TAM and TPB
- Model of PC Utilization
- Diffusion Of Innovation Theory
- Social Cognitive Theory

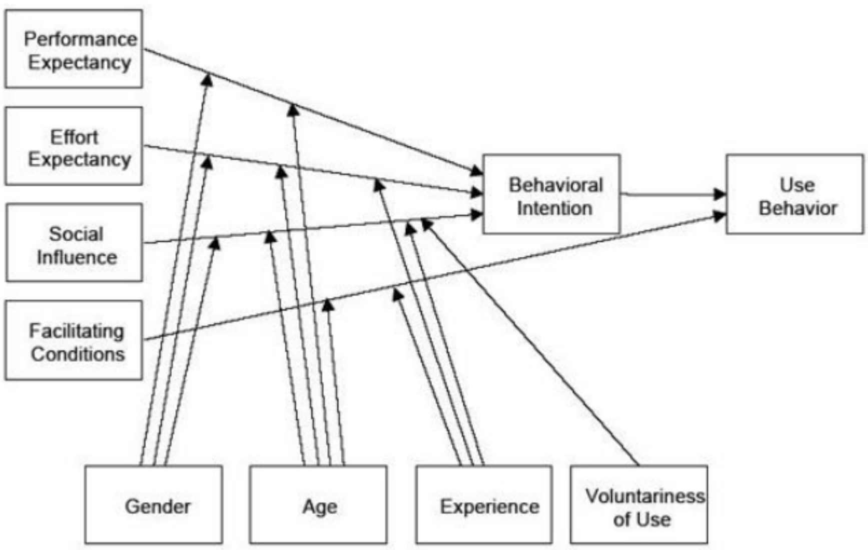

Figure A.3: Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003)

Appendix B

Missing values PV panel data

In this appendix, the missing values of the PV panel data are shown in maps of the Netherlands. This gives a proper indication of how these missing values are distributed across the Netherlands. Remarkable is that the number missing values for municipalities is zero and for the neighbourhoods and districts is higher. However, this can be explained by the fact that municipalities consist of a group of districts and districts consist of a group of neighbourhoods. So when there is a missing on neighbourhoods level, this missing value is still present looking at districts, only it may be that adjacent neighborhoods do have a value so there is no missing value for districts in the data.

B.1 2018

Table B.1: Missing values neighbourhoods (2018)

Neighbourhood	N	Valid	**10,552**
		Missing	**2,754**
		Total	**13,306**

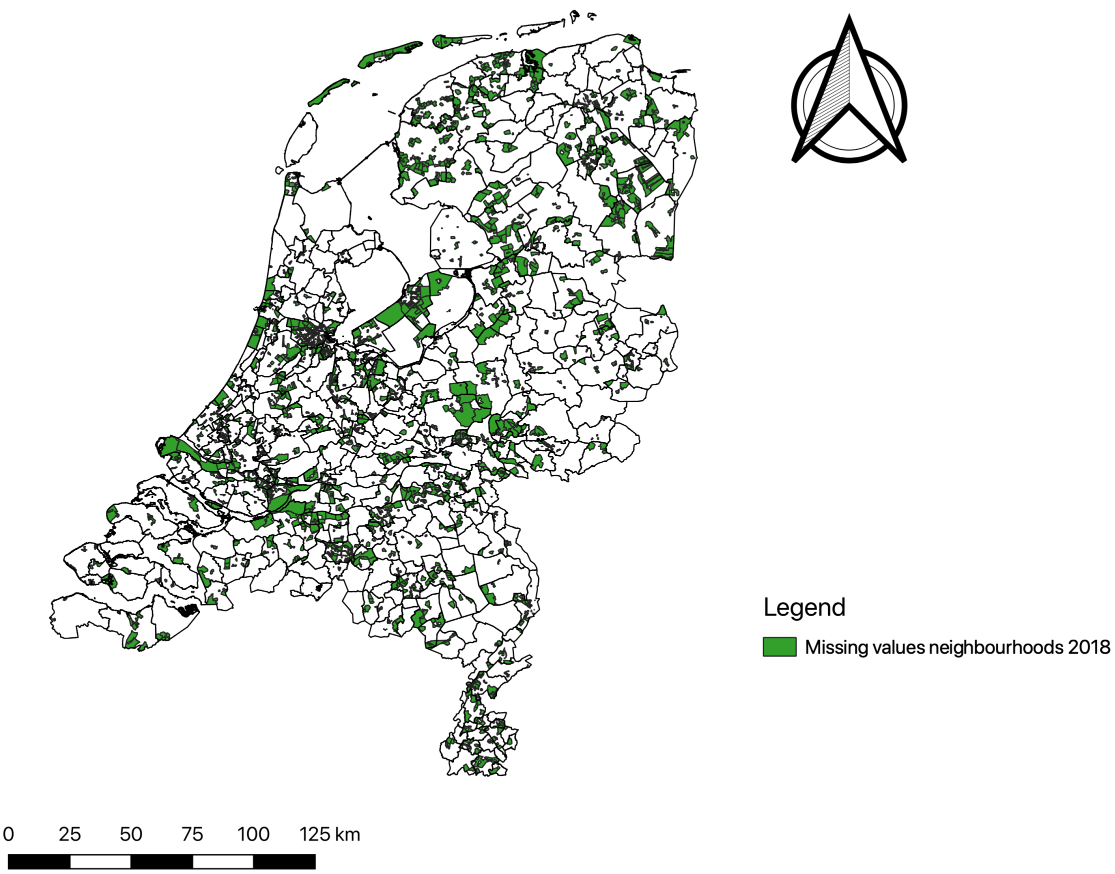

Figure B.1: Missing values in PV panels data across neighbourhoods in the Netherlands (2018)

Table B.2: Missing values districts (2018)

Neighbourhood	N	Valid	**2,921**
		Missing	**166**
		Total	**3,087**

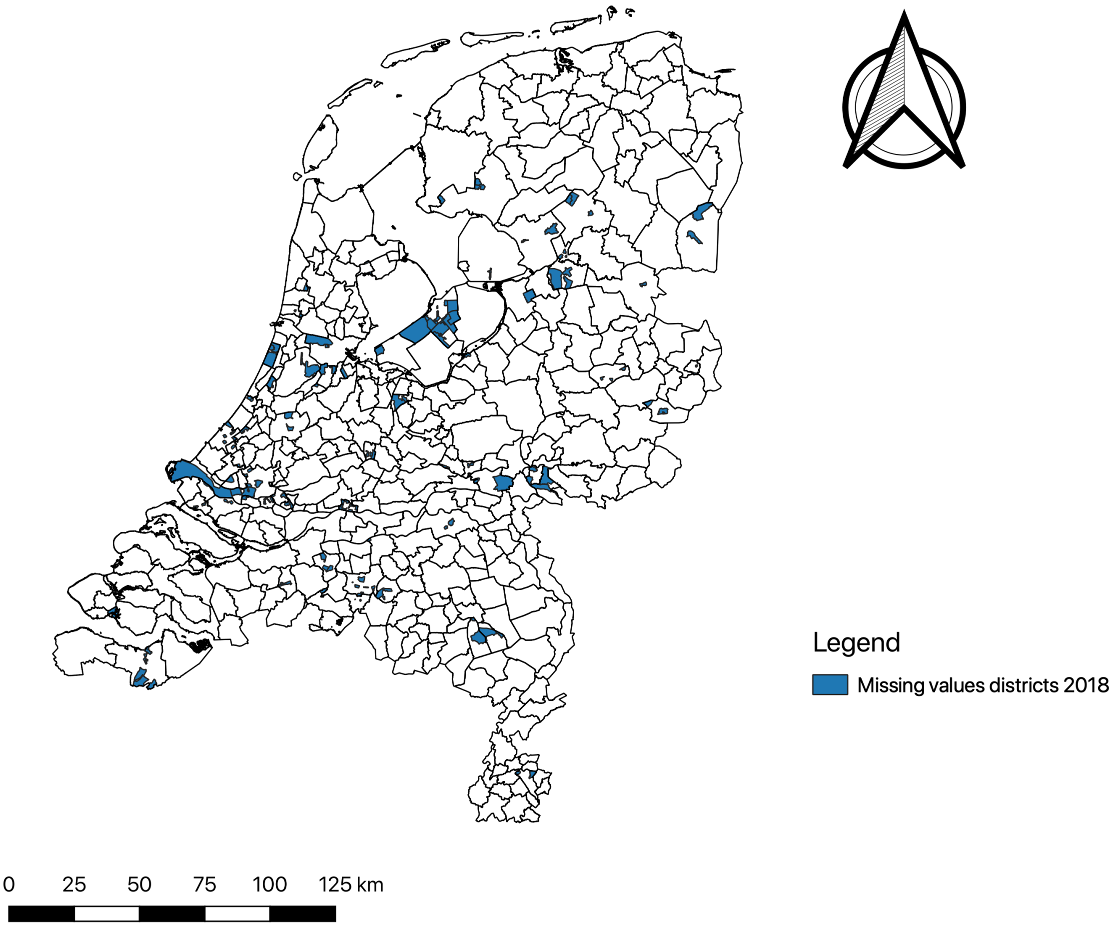

Figure B.2: Missing values in PV panels data across districts in the Netherlands (2018)

Table B.3: Missing values municipalities (2018)

Neighbourhood	N	Valid	**381**
		Missing	**0**
		Total	**381**

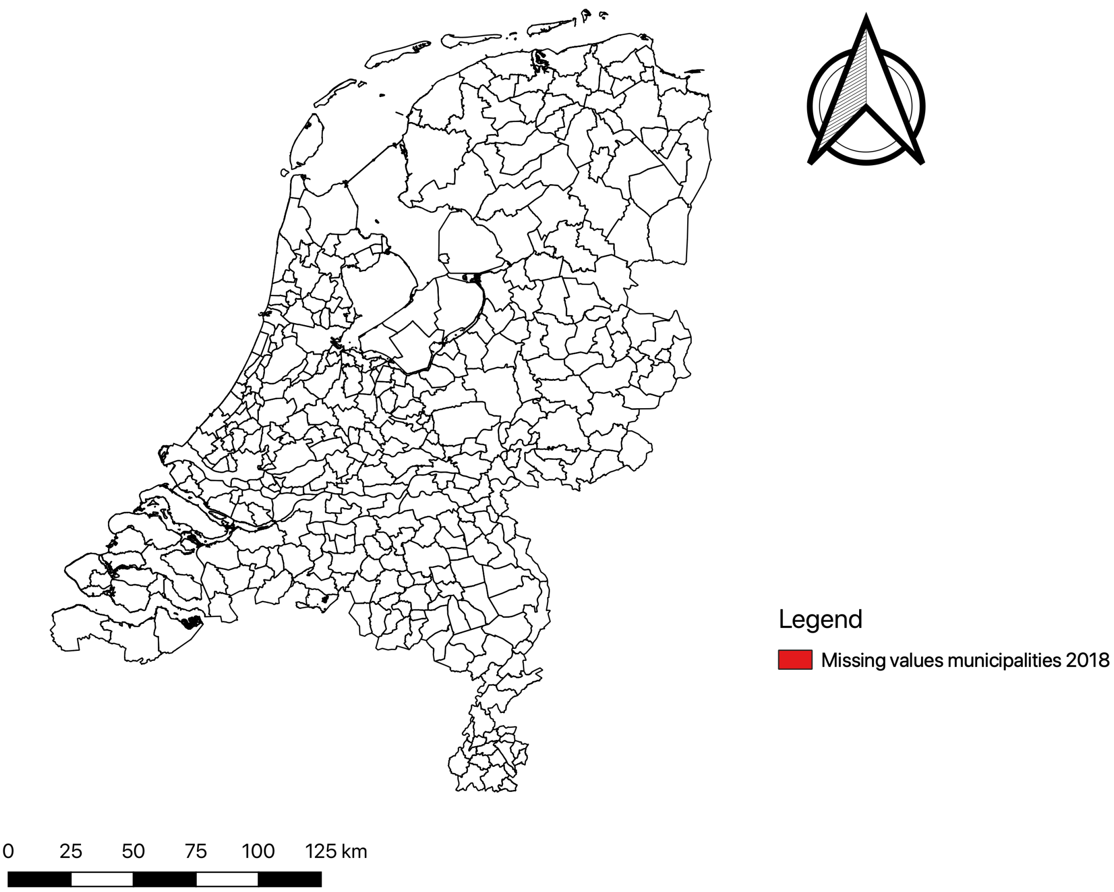

Figure B.3: Missing values in PV panels data across municipalities in the Netherlands (2018)

Appendix C

Syntax of SPSS

In this appendix the SPSS syntax for the data preparation is shown, it consists of the data adjustments to make the data more understandable, as well as computations of new variables or a combination of variables. Some other parts are frequencies, correlations and other aspects related to data preparation.

```
* Encoding: UTF-8.
RENAME VARIABLES (RegioaanduidingCoderingcode BevolkingAantalinwonersaantal
    BevolkingGeslachtMannenaantal BevolkingGeslachtVrouwenaantal
    BevolkingLeeftijdsgroepen0tot15jaaraantal
    BevolkingLeeftijdsgroepen15tot25jaaraantal
        BevolkingLeeftijdsgroepen25tot45jaaraantal
        BevolkingLeeftijdsgroepen45tot65jaaraantal
        BevolkingLeeftijdsgroepen65jaarofouderaantal
        BevolkingBurgerlijkestaatOngehuwdaantal
    BevolkingBurgerlijkestaatGehuwdaantal BevolkingBurgerlijkestaatGescheidenaantal
         BevolkingBurgerlijkestaatVerweduwdaantal
        BevolkingPersonenmeteenmigratieachtergrondWesterstotaala
        BevolkingPersonenmeteenmigratieachtergrondNietwestersNiet
    BevolkingParticulierehuishoudensHuishoudenstotaalaantal
        BevolkingBevolkingsdichtheidaantalinwonersperkm WonenWoningvoorraadaantal
        WonenGemiddeldewoningwaardex1000euro
        WonenWoningennaartypePercentageeengezinswoning
    WonenWoningennaartypePercentagemeergezinswoning
        WonenWoningennaareigendomKoopwoningen
        WonenWoningennaareigendomHuurwoningenHuurwoningentotaal
        WonenWoningennaareigendomEigendomonbekend
        WonenWoningennaarbouwjaarBouwjaarvoor2000
    WonenWoningennaarbouwjaarBouwjaarvanaf2000
        SocialezekerheidPersonenpersoortuitkeringWWaantal
        StedelijkheidMatevanstedelijkheidcode
        StedelijkheidOmgevingsadressendichtheidperkm
    InkomenInkomenvanpersonenAantalinkomensontvangersaantal
        InkomenInkomenvanpersonenGemiddeldinkomenperinkomensontva
        InkomenInkomenvanpersonenGemiddeldinkomenperinwonerx1
    InkomenInkomenvanpersonen40personenmetlaagsteinkomen
        InkomenInkomenvanpersonen20personenmethoogsteinkomen
        AantalInstallatiesBijWoningen_5 OpgesteldVermogenVanZonnepanelen_6 BU_CODE
        BU_NAME WK_CODE GM_CODE GM_NAAM POSTCODE MEAN
    =CODE NUMPEOPLE NUMMALES NUMFEMALES NUM00_14 NUM15_24 NUM25_44 NUM45_64 NUM65
        NUMUNMAR NUMMAR NUMDIV NUMWID NUMWEST NUMNOTWEST NUMHH POPDENS NUMHOUSE
        HOUSEVAL PERC1FAMH PERCMFAMH
    PERCOWNED PERCRENTED PERCUNKNOWN PERCBEFORE2000 PERCAFTER2000 NUMUNEMPLOY URBAN
         ADDRESSDENS NUMINCOME AVEINCOMEPERREC AVEINCOMEPERIN PERCUNDER40
        PERCABOVE20 NUMINSTAL CAPACITY NEICODE NEINAME DISCODE MUNCODE MUNNAME
        ZIPCODE SOLAR).

SELECT IF WATER="NEE".
```

```
EXECUTE.

FREQUENCIES VARIABLES=WATER
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

FREQUENCIES VARIABLES=NUMINSTAL18 CAPACITY18
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

COMPUTE INSTALPHOU18=NUMINSTAL18 / NUMHOUSE18.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (NVALID(NUMINSTAL18)).
EXECUTE.

FREQUENCIES VARIABLES=INSTALPHOU18
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /ORDER=ANALYSIS.

COMPUTE PERCNUMMALES=NUMMALES/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMFEMALES=NUMFEMALES/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUM00_14=NUM00_14/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUM15_24=NUM15_24/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUM25_44=NUM25_44/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUM45_64=NUM45_64/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUM65=NUM65/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMUNMAR=NUMUNMAR/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMMAR=NUMMAR/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMDIV=NUMDIV/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMWID=NUMWID/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMWEST=NUMWEST/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMNOTWEST=NUMNOTWEST/NUMPEOPLE*100.
EXECUTE.

COMPUTE PERCNUMUNEMPLOY=NUMUNEMPLOY/NUMPEOPLE*100.
EXECUTE.

DELETE VARIABLES NUMMALES NUMFEMALES NUM00_14 NUM15_24 NUM25_44 NUM45_64 NUM65
```

```
    NUMUNMAR NUMMAR NUMDIV NUMWID NUMWEST NUMNOTWEST.

MISSING VALUES INSTALPHOU18 INSTALPHOU16 INSTALPHOU17 (1.001 THRU HIGHEST).

FREQUENCIES VARIABLES=INSTALPHOU18 INSTALPHOU16 INSTALPHOU17
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES=INSTALPHOU18 INSTALPHOU16 INSTALPHOU17
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=INSTALPHOU18 INSTALPHOU16 INSTALPHOU17
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES PERCNUMMALES PERCNUMFEMALES (100.001 THRU HIGHEST).

FREQUENCIES VARIABLES=PERCNUMMALES PERCNUMFEMALES
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES=PERCNUMMALES PERCNUMFEMALES
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=PERCNUMMALES PERCNUMFEMALES
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES PERCNUM00_14 PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65
     (100.001 THRU HIGHEST).

FREQUENCIES VARIABLES=PERCNUM00_14 PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES=PERCNUM00_14 PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=PERCNUM00_14 PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65
  /PLOT BOXPLOT HISTOGRAM NPPLOT
```

```
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES PERCNUMUNMAR PERCNUMMAR PERCNUMDIV PERCNUMWID (100.001 THRU HIGHEST)
    .

FREQUENCIES VARIABLES=PERCNUMUNMAR PERCNUMMAR PERCNUMDIV PERCNUMWID
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES=PERCNUMUNMAR PERCNUMMAR PERCNUMDIV PERCNUMWID
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=PERCNUMUNMAR PERCNUMMAR PERCNUMDIV PERCNUMWID
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES PERCNUMWEST PERCNUMNOTWEST (100.001 THRU HIGHEST).

FREQUENCIES VARIABLES=PERCNUMWEST PERCNUMNOTWEST
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES=PERCNUMWEST PERCNUMNOTWEST
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=PERCNUMWEST PERCNUMNOTWEST
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES PERCUNDER40 PERCABOVE20 PERCNUMUNEMPLOY (100.001 THRU HIGHEST).

FREQUENCIES VARIABLES= PERCUNDER40 PERCABOVE20 PERCNUMUNEMPLOY
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES= PERCUNDER40 PERCABOVE20 PERCNUMUNEMPLOY
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
```

```
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=PERCUNDER40 PERCABOVE20 PERCNUMUNEMPLOY
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

DELETE VARIABLES AVEINCOMEPERREC AVEINCOMEPERIN.

MISSING VALUES ADDRESSDENS POPDENS ().

MISSING VALUES URBAN (6 THRU HIGHEST).

FREQUENCIES VARIABLES= ADDRESSDENS URBAN POPDENS
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES= ADDRESSDENS URBAN POPDENS
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=ADDRESSDENS URBAN POPDENS
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES HOUSEVAL ().

FREQUENCIES VARIABLES= HOUSEVAL
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES= HOUSEVAL
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=HOUSEVAL
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES PERC1FAMH PERCMFAMH (100.001 THRU HIGHEST).
```

```
FREQUENCIES VARIABLES= PERC1FAMH PERCMFAMH
   /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
   /HISTOGRAM NORMAL
   /ORDER=ANALYSIS.

EXAMINE VARIABLES= PERC1FAMH PERCMFAMH
   /COMPARE VARIABLE
   /PLOT=BOXPLOT
   /STATISTICS=NONE
   /NOTOTAL
   /MISSING=LISTWISE.

EXAMINE VARIABLES=PERC1FAMH PERCMFAMH
   /PLOT BOXPLOT HISTOGRAM NPPLOT
   /COMPARE GROUPS
   /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
   /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
   /STATISTICS DESCRIPTIVES EXTREME
   /CINTERVAL 95
   /MISSING LISTWISE
   /NOTOTAL.

MISSING VALUES PERCOWNED PERCRENTED PERCUNKNOWN (100.001 THRU HIGHEST).

FREQUENCIES VARIABLES= PERCOWNED PERCRENTED PERCUNKNOWN
   /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
   /HISTOGRAM NORMAL
   /ORDER=ANALYSIS.

EXAMINE VARIABLES= PERCOWNED PERCRENTED PERCUNKNOWN
   /COMPARE VARIABLE
   /PLOT=BOXPLOT
   /STATISTICS=NONE
   /NOTOTAL
   /MISSING=LISTWISE.

EXAMINE VARIABLES=PERCOWNED PERCRENTED PERCUNKNOWN
   /PLOT BOXPLOT HISTOGRAM NPPLOT
   /COMPARE GROUPS
   /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
   /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
   /STATISTICS DESCRIPTIVES EXTREME
   /CINTERVAL 95
   /MISSING LISTWISE
   /NOTOTAL.

MISSING VALUES PERCBEFORE2000 PERCAFTER2000 (100.001 THRU HIGHEST).

FREQUENCIES VARIABLES= PERCBEFORE2000 PERCAFTER2000
   /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
   /HISTOGRAM NORMAL
   /ORDER=ANALYSIS.

EXAMINE VARIABLES= PERCBEFORE2000 PERCAFTER2000
   /COMPARE VARIABLE
   /PLOT=BOXPLOT
   /STATISTICS=NONE
   /NOTOTAL
   /MISSING=LISTWISE.

EXAMINE VARIABLES=PERCBEFORE2000 PERCAFTER2000
   /PLOT BOXPLOT HISTOGRAM NPPLOT
   /COMPARE GROUPS
   /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
   /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
   /STATISTICS DESCRIPTIVES EXTREME
```

```
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES SOLAR (3 THRU HIGHEST, −99999999.000000000000000
).

FREQUENCIES VARIABLES= SOLAR
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

EXAMINE VARIABLES= SOLAR
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

EXAMINE VARIABLES=SOLAR
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

EXAMINE VARIABLES=INSTALPHOU18 INSTALPHOU16 INSTALPHOU17 PERCNUMMALES
    PERCNUMFEMALES PERCNUM00_14
    PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65 PERCNUMUNMAR PERCNUMMAR
        PERCNUMDIV PERCNUMWID
    PERCNUMWEST PERCNUMNOTWEST PERC1FAMH PERCMFAMH PERCOWNED PERCRENTED PERCUNKNOWN
         PERCBEFORE2000
    PERCAFTER2000 POPDENS ADDRESSDENS URBAN HOUSEVAL PERCUNDER40 PERCABOVE20
    PERCNUMUNEMPLOY SOLAR
  /PLOT BOXPLOT STEMLEAF
  /COMPARE GROUPS
  /MESTIMATORS HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5) TUKEY(4.685)
  /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

MISSING VALUES INSTALPHOU18 (0.4260564 THRU HIGHEST).
MISSING VALUES INSTALPHOU16 (0.2687012 THRU HIGHEST).
MISSING VALUES INSTALPHOU17 (0.332821 THRU HIGHEST).


RECODE PERCNUMMALES (Lowest thru 40.7193=9999).
EXECUTE.
RECODE PERCNUMMALES (59.4107 THRU HIGHEST=9999).
EXECUTE.
RECODE PERCNUMFEMALES (Lowest thru 40.5822=9999).
EXECUTE.
RECODE PERCNUMFEMALES (59.1353 THRU HIGHEST=9999).
EXECUTE.


MISSING VALUES PERCNUM00_14 (33.2059 THRU HIGHEST).
MISSING VALUES PERCNUM15_24 (23.8903 THRU HIGHEST).
MISSING VALUES PERCNUM25_44 (47.0224 THRU HIGHEST).
RECODE PERCNUM45_64 (Lowest thru 4.3596=9999).
EXECUTE.
RECODE PERCNUM45_64 (56.23 thru highest=9999).
```

```
EXECUTE.
MISSING VALUES PERCNUM65 (51.8062 THRU HIGHEST).


RECODE PERCNUMUNMAR (Lowest thru 14.3773=9999).
EXECUTE.
RECODE PERCNUMUNMAR (75.0666 thru highest=9999).
EXECUTE.
RECODE PERCNUMMAR (Lowest thru 10.6242=9999).
EXECUTE.
RECODE PERCNUMMAR (75.7953 thru highest=9999).
EXECUTE.
MISSING VALUES PERCNUMDIV (19.9579 THRU HIGHEST).
MISSING VALUES PERCNUMWID (15.03 THRU HIGHEST).


MISSING VALUES PERCNUMWEST (28.5616 THRU HIGHEST).
MISSING VALUES PERCNUMNOTWEST (34.6174 THRU HIGHEST).


MISSING VALUES PERC1FAMH (100.001 THRU HIGHEST).
MISSING VALUES PERCMFAMH (100.001 THRU HIGHEST).


MISSING VALUES PERCOWNED (100.001 THRU HIGHEST).
MISSING VALUES PERCRENTED (100.001 THRU HIGHEST).
MISSING VALUES PERCUNKNOWN (4 THRU HIGHEST).


RECODE PERCBEFORE2000 (Lowest thru 41=9999).
EXECUTE.
RECODE PERCBEFORE2000 (100.001 thru highest=9999).
EXECUTE.
MISSING VALUES PERCAFTER2000 (59 THRU HIGHEST).


MISSING VALUES POPDENS (20587 THRU HIGHEST).
MISSING VALUES ADDRESSDENS (6726 THRU HIGHEST).
MISSING VALUES URBAN (6 THRU HIGHEST).
MISSING VALUES HOUSEVAL (657 THRU HIGHEST).


RECODE PERCUNDER40 (Lowest thru 11.4=9999).
EXECUTE.
RECODE PERCUNDER40 (66 thru highest=9999).
EXECUTE.
MISSING VALUES PERCABOVE20 (56.7 THRU HIGHEST).
MISSING VALUES PERCNUMUNEMPLOY (4.7502 THRU HIGHEST).


RECODE SOLAR (Lowest thru 2.0201289=9999).
EXECUTE.
RECODE SOLAR (2.9988541 thru highest=9999).
EXECUTE.

FREQUENCIES VARIABLES=INSTALPHOU18 INSTALPHOU16 INSTALPHOU17 PERCNUMMALES
    PERCNUMFEMALES PERCNUM00_14
    PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65 PERCNUMUNMAR PERCNUMMAR
        PERCNUMDIV PERCNUMWID
    PERCNUMWEST PERCNUMNOTWEST PERC1FAMH PERCMFAMH PERCOWNED PERCRENTED PERCUNKNOWN
         PERCBEFORE2000
    PERCAFTER2000 POPDENS ADDRESSDENS MODURB STRURB HOUSEVAL PERCUNDER40
        PERCABOVE20
    PERCNUMUNEMPLOY SOLAR
  /FORMAT=NOTABLE
  /STATISTICS=STDDEV RANGE MINIMUM MAXIMUM MEAN MEDIAN
  /HISTOGRAM NORMAL
```

```
  /ORDER=ANALYSIS.

COMPUTE INCR161=INS17 - INS16.
EXECUTE.

COMPUTE INCR17=INS18 - INS17.
EXECUTE.

CORRELATIONS
  /VARIABLES=INSTALPHOU18 INCR16 INCR17 PERCNUMMALES PERCNUMFEMALES PERCNUM00_14
    PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65 PERCNUMUNMAR PERCNUMMAR
        PERCNUMDIV PERCNUMWID
    PERCNUMWEST PERCNUMNOTWEST PERC1FAMH PERCMFAMH PERCOWNED PERCRENTED PERCUNKNOWN
        PERCBEFORE2000
    PERCAFTER2000 POPDENS ADDRESSDENS URBAN HOUSEVAL PERCUNDER40 PERCABOVE20
    PERCNUMUNEMPLOY SOLAR
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.

REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT INSTALPHOU18
  /METHOD=ENTER PERCNUMMALES PERCNUM00_14 PERCNUM15_24 PERCNUM25_44 PERCNUM45_64
      PERCNUM65
    PERCNUMUNMAR PERCNUMDIV PERCNUMWID PERCNUMWEST PERC1FAMH PERCOWNED URBAN
        PERCNUMUNEMPLOY SOLAR.

COUNT MISSINGS = INSTALPHOU18 PERCNUMMALES PERCNUMFEMALES PERCNUM00_14
    PERCNUM15_24 PERCNUM25_44 PERCNUM45_64 PERCNUM65 PERCNUMUNMAR PERCNUMMAR
        PERCNUMDIV PERCNUMWID
    PERCNUMWEST PERC1FAMH PERCMFAMH PERCOWNED POPDENS ADDRESSDENS URBAN
        PERCNUMUNEMPLOY SOLAR (MISSING).

FREQUENCIES VARIABLES=missings
  /ORDER=ANALYSIS.

REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT INSTALPHOU18
  /METHOD=ENTER PERCNUMMALES PERCNUM00_14 PERCNUM15_24 PERCNUM25_44 PERCNUM45_64
      PERCNUM65
    PERCNUMUNMAR PERCNUMDIV PERCNUMWID PERCNUMWEST PERC1FAMH PERCOWNED URBAN
        PERCNUMUNEMPLOY SOLAR.

RECODE URBAN (3=1) (4=1) (ELSE=0) INTO MODURB.
EXECUTE.

RECODE URBAN (5=1) (ELSE=0) INTO STRURB.
EXECUTE.

RENAME VARIABLES (PERCNUMMALES PERCNUMFEMALES PERCNUM00_14 PERCNUM15_24
    PERCNUM25_44 PERCNUM45_64 PERCNUM65 PERCNUMUNMAR PERCNUMMAR PERCNUMDIV
    PERCNUMWID PERCNUMWEST PERCNUMNOTWEST PERC1FAMH
    PERCMFAMH PERCOWNED PERCRENTED PERCUNKNOWN PERCBEFORE2000 PERCAFTER2000
        ADDRESSDENS HOUSEVAL PERCUNDER40 PERCABOVE20 PERCNUMUNEMPLOY = MALES
        FEMALES A0014 A1524 A2544 A4564 A65 UNMAR MAR DIV WID
    WEST NOTWEST SINFAM MULFAM OWN RENT UNK BE2000 AF2000 ADDENS HVAL U40 A20
        UNEMPL).
```

```
RENAME VARIABLES (INSTALPHOU18 = INS18).
```

Appendix D

Data descriptives

In this appendix. some basic data descriptives are shown at neighbourhood level for the year 2018.

D.1 Dependent variable

D.1.1 Number of PV installations per neighbourhood (NUMINSTAL) and their capacity (CAPACITY)

Table D.1: PV panels and their capacity

		Number of installations (#)	Capacity (kWh)
N	Valid	10551	10550
	Missing	2754	2755
Mean		63.71	204.83
Median.		32	118.
Std. Deviation		85.373	254.857
Range		1300	3378
Minimum		1	1
Maximum		1301	3379
Sum		672231	2160943

D.1.2 Number of houses (NUMHOUSE)

Table D.2: Number of houses

N	Valid	13305
	Missing	0
Mean		581.81
Median.		293
Std. Deviation		795.997
Range		14234
Minimum		0
Maximum		14234
Sum		7740984

D.1.3 Number of people (NUMPEOPLE)

Table D.3: Number of people

N	Valid	13305
	Missing	0
Mean		1290.55
Median.		685
Std. Deviation		1712.680
Range		28450
Minimum		0
Maximum		28450
Sum		17170810

D.2 Missing values

Table D.4: Missing values at neighbourhood level

#	Variable	Valid (#)	Missing (#)	Missing (%)
1	INSTALPHOU18	10269	176	1,71
2	INCR16	7961	2484	31,20
3	INCR17	8976	1469	16,37
4	PERCNUMMALES	10273	172	1,67
5	PERCNUMFEMALES	10274	171	1,66
6	PERCNUM00_14	10323	122	1,18
7	PERCNUM15_24	10171	274	2,69
8	PERCNUM25_44	10303	142	1,38
9	PERCNUM45_64	10411	34	0,33
10	PERCNUM65	10369	76	0,73
11	PERCNUMUNMAR	10303	142	1,38
12	PERCNUMMAR	10407	38	0,37
13	PERCNUMDIV	9947	498	5,01
14	PERCNUMWID	9522	923	9,69
15	PERCNUMWEST	9904	541	5,46
16	PERCNUMNOTWEST	8214	2231	27,16
17	PERC1FAMH	10287	158	1,54
18	PERCMFAMH	10287	158	1,54
19	PERCOWNED	10276	169	1,64
20	PERCRENTED	10276	169	1,64
21	PERCUNKNOWN	9566	879	9,19
22	PERCBEFORE2000	9627	818	8,50
23	PERCAFTER2000	9627	818	8,50
24	POPDENS	10358	87	0,84
25	ADDRESSDENS	10323	122	1,18
26	URBAN	10444	1	0,01
27	HOUSEVAL	8958	1487	16,60
28	PERCUNDER40	9490	955	10,06
29	PERCOABOVE20	9497	948	9,98
30	PERCNUMUNEMPLOY	7455	2990	40,11
31	SOLAR	10445	0	0

D.3 Histograms

D.3.1 Number of installations per house

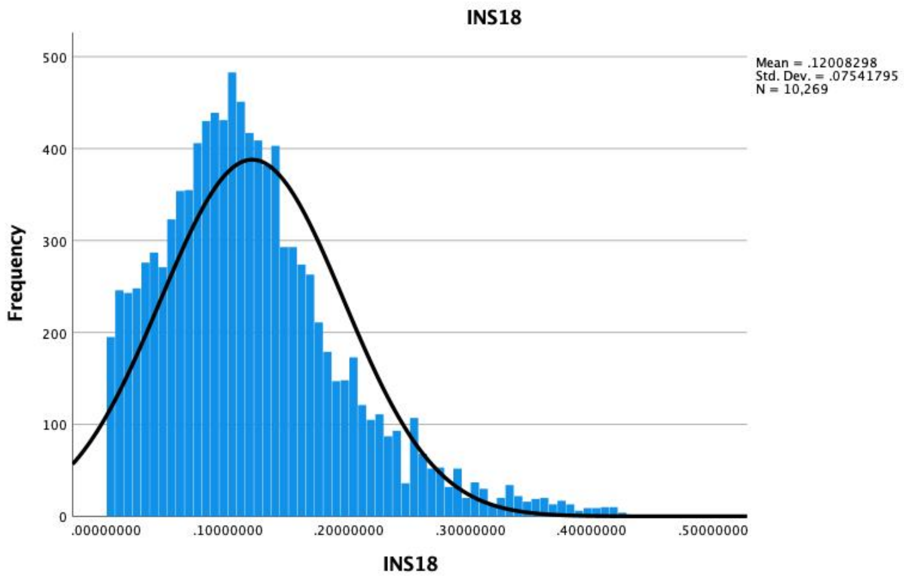

Figure D.1: Histogram number of installations per house

D.3.2 Increase in number of installations per house in 2016

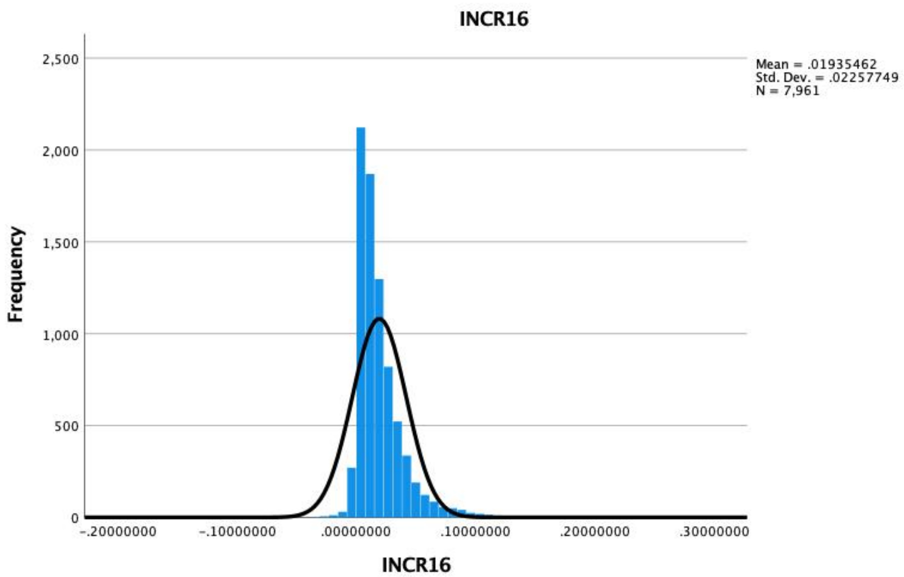

Figure D.2: Histogram increase in the number of installations per house in 2016

D.3.3 Increase in number of installations per house in 2017

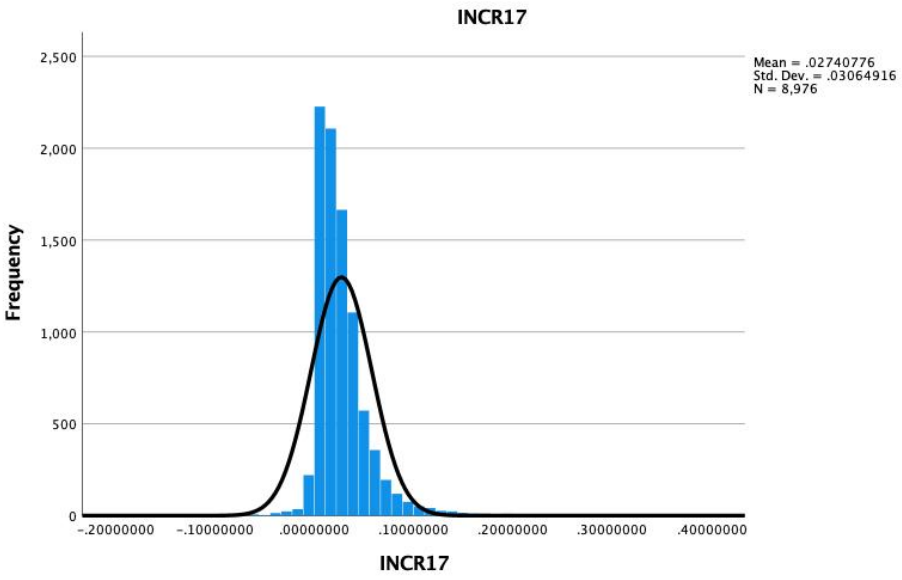

Figure D.3: Histogram increase in the number of installations per house in 2017

D.3.4 Percentage of males

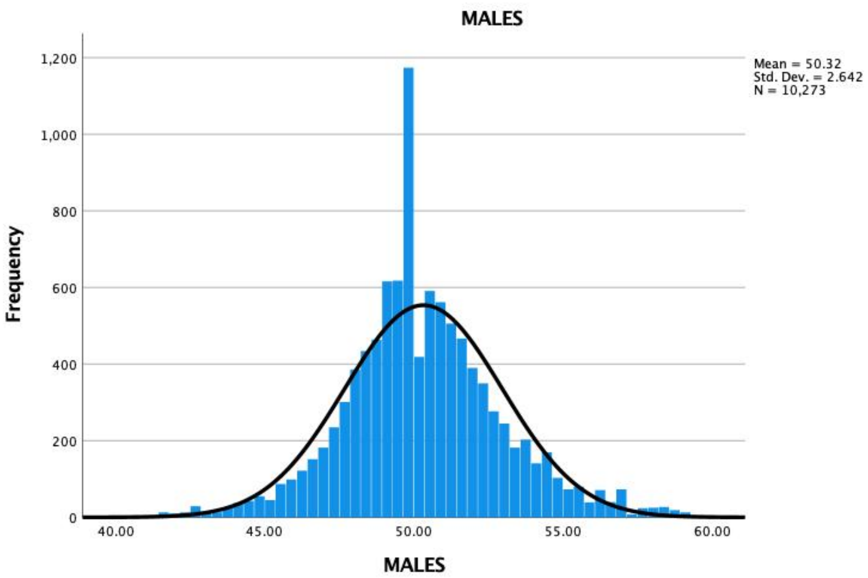

Figure D.4: Histogram percentage of males

D.3.5 Percentage of females

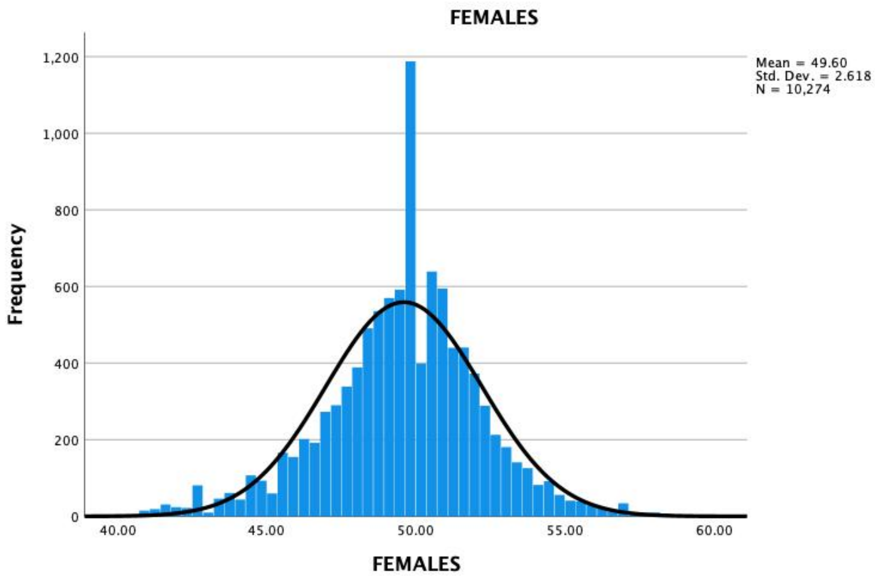

Figure D.5: Histogram percentage of females

D.3.6 Percentage of people aged between 0 to 14

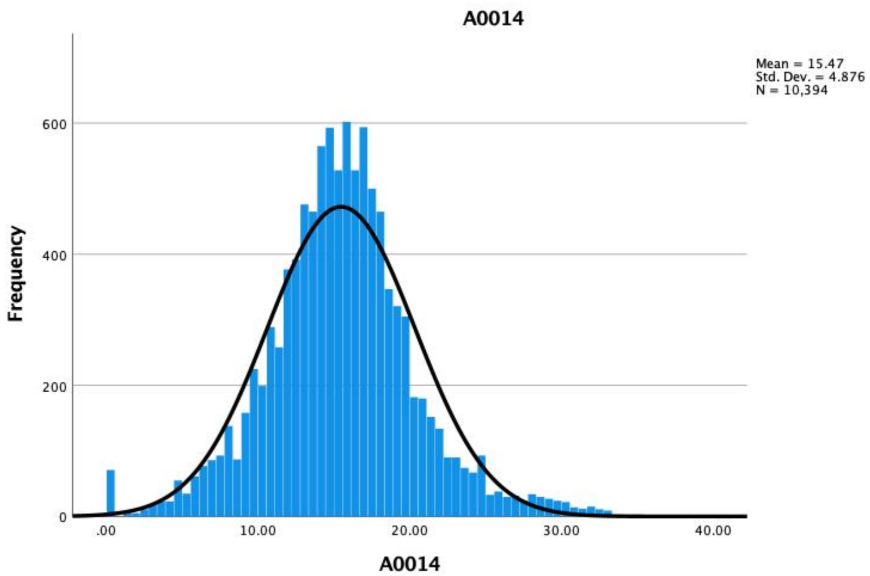

Figure D.6: Histogram percentage of people aged between 0 to 14

D.3.7 Percentage of people aged between 15 to 24

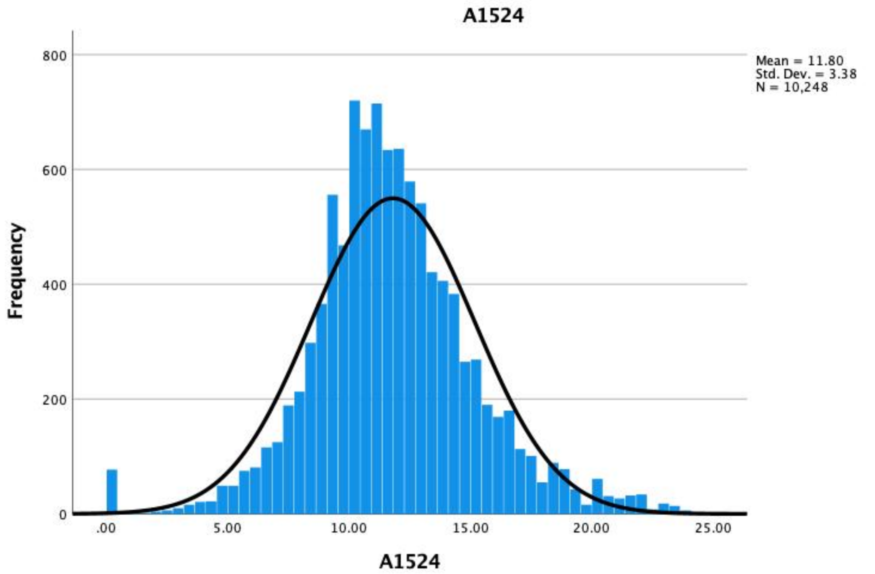

Figure D.7: Histogram percentage of people aged between 15 to 24

D.3.8 Percentage of people aged between 25 to 44

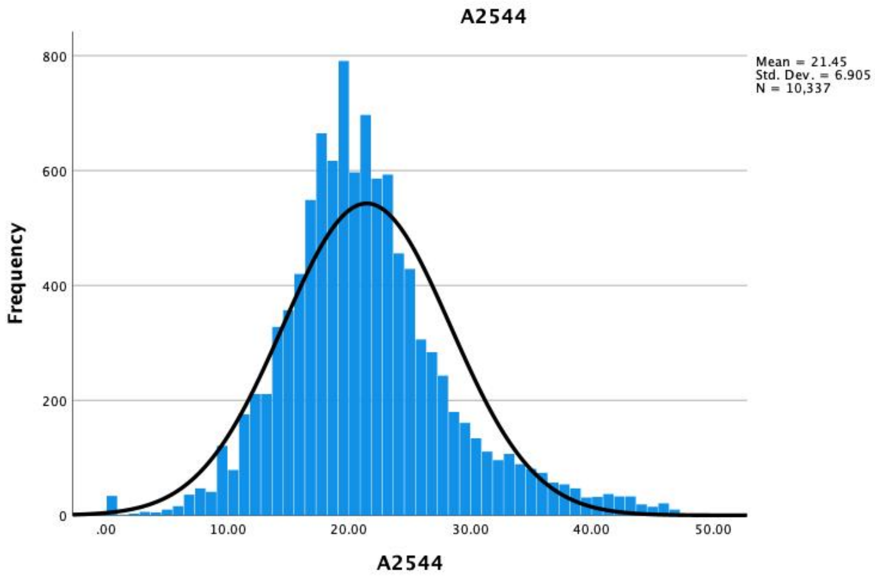

Figure D.8: Histogram percentage of people aged between 25 to 44

D.3.9 Percentage of people aged between 45 to 64

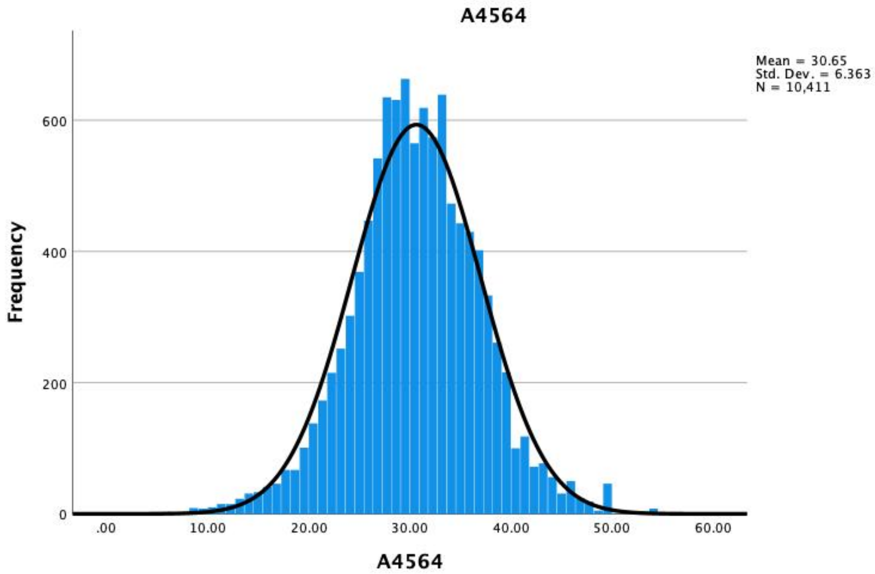

Figure D.9: Histogram percentage of people aged between 45 to 64

D.3.10 Percentage of people aged 65 and over

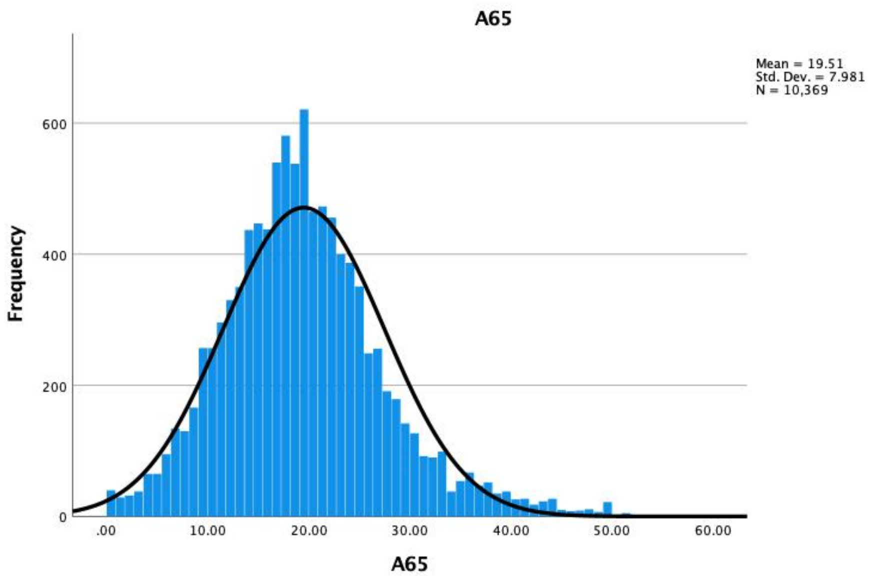

Figure D.10: Histogram percentage of people aged 65 and over

D.3.11 Percentage of unmarried people

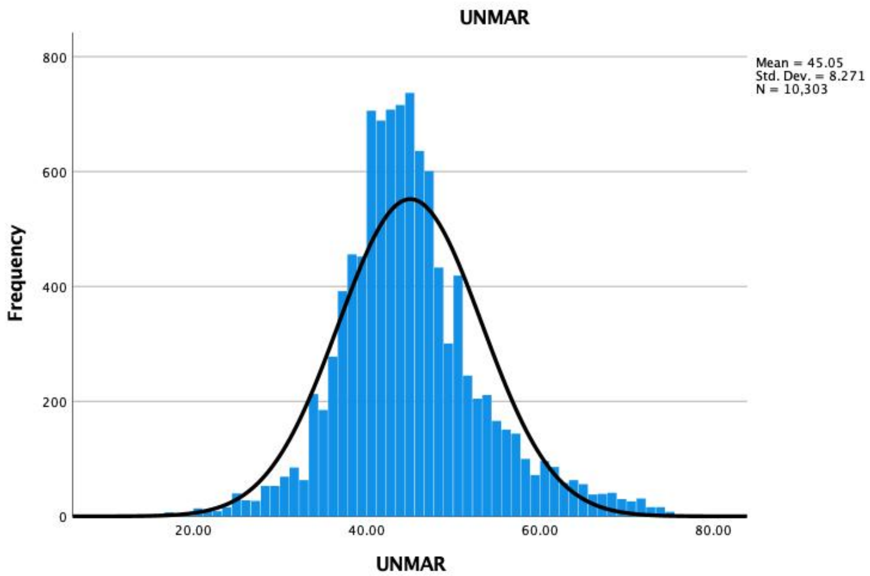

Figure D.11: Histogram percentage of unmarried people

D.3.12 Percentage of married people

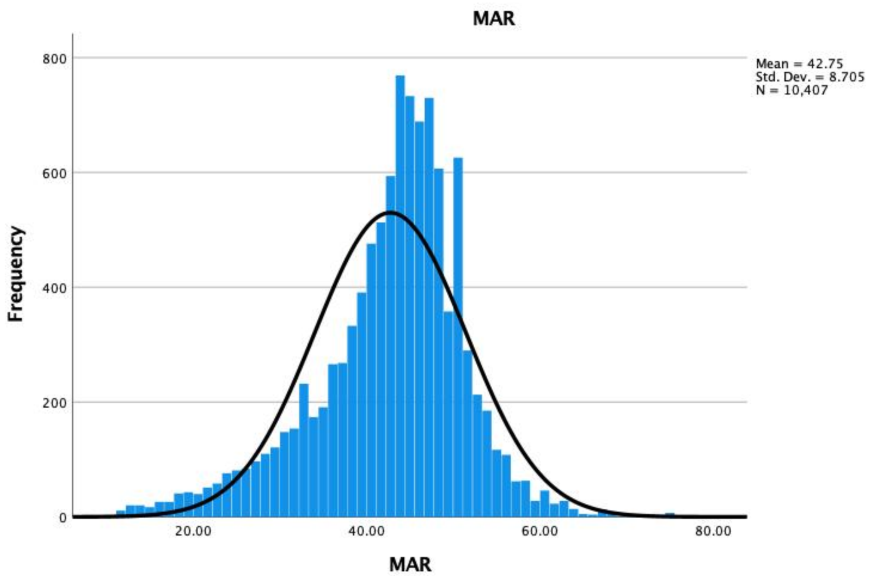

Figure D.12: Histogram percentage of married people

D.3.13 Percentage of divorced people

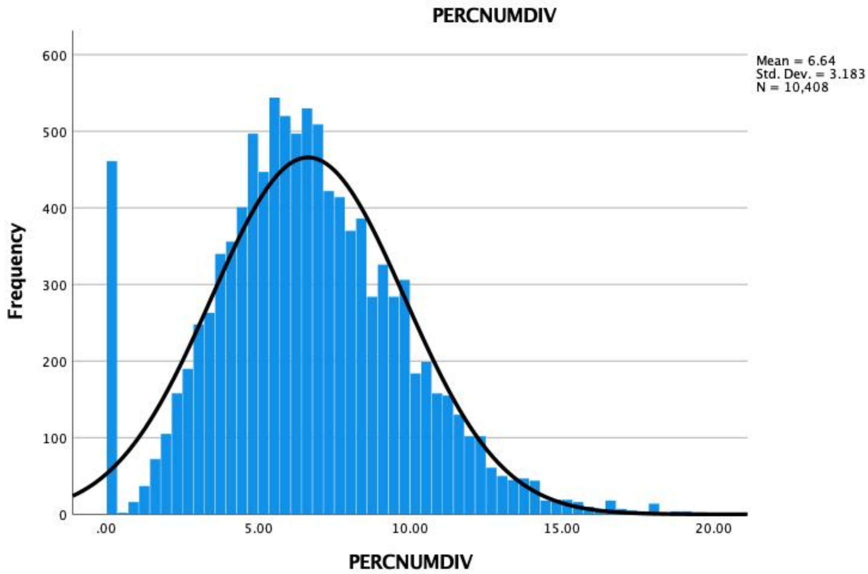

Figure D.13: Histogram percentage of divorced people

D.3.14 Percentage of widows

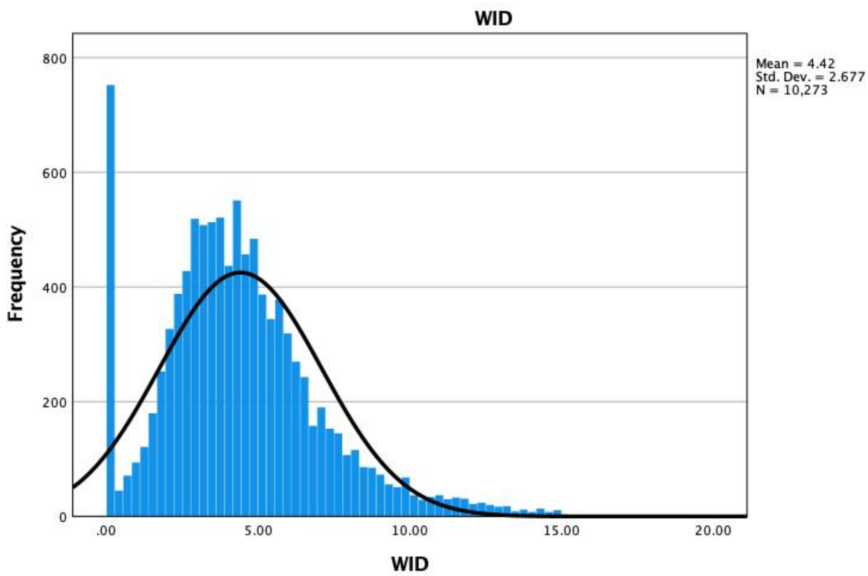

Figure D.14: Histogram percentage of widows

D.3.15 Percentage of foreign western people

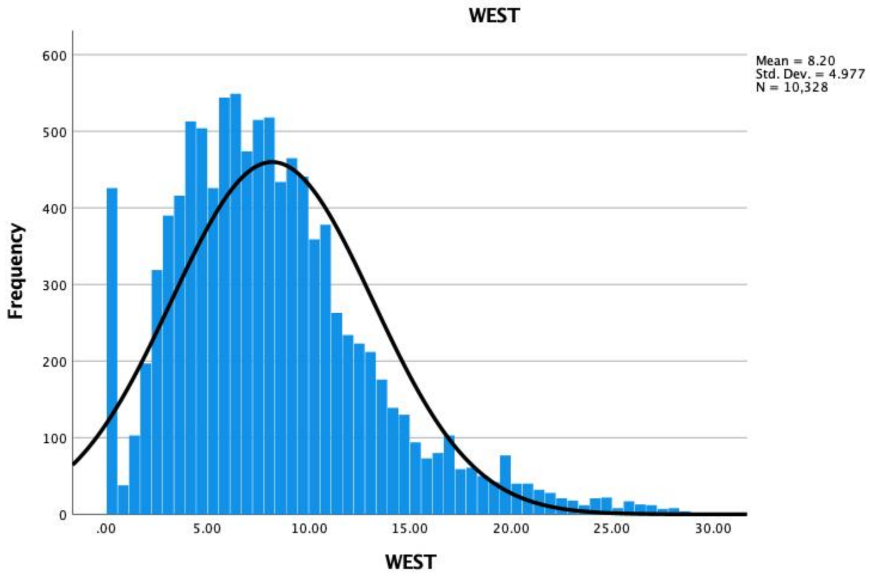

Figure D.15: Histogram percentage of foreign western people

D.3.16 Percentage of foreign not western people

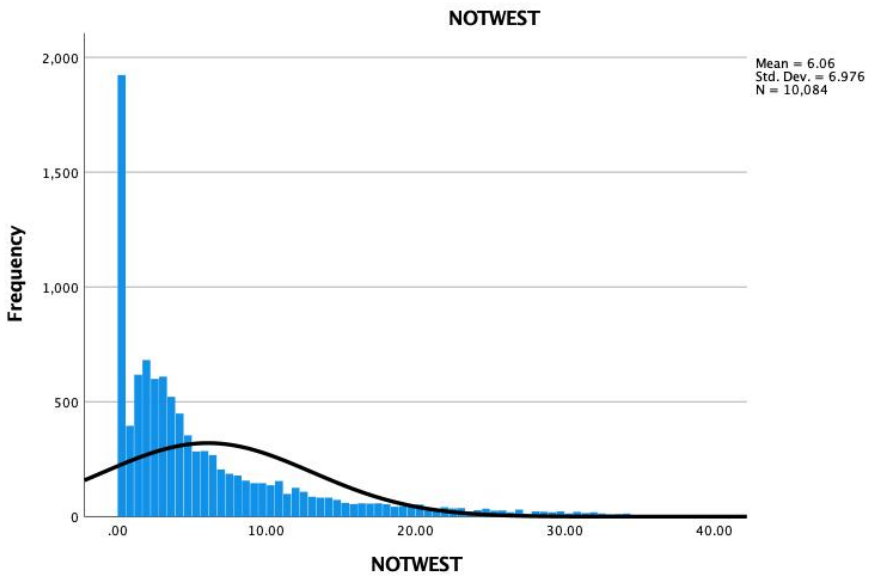

Figure D.16: Histogram percentage of foreign not western people

D.3.17 Percentage of single-family houses

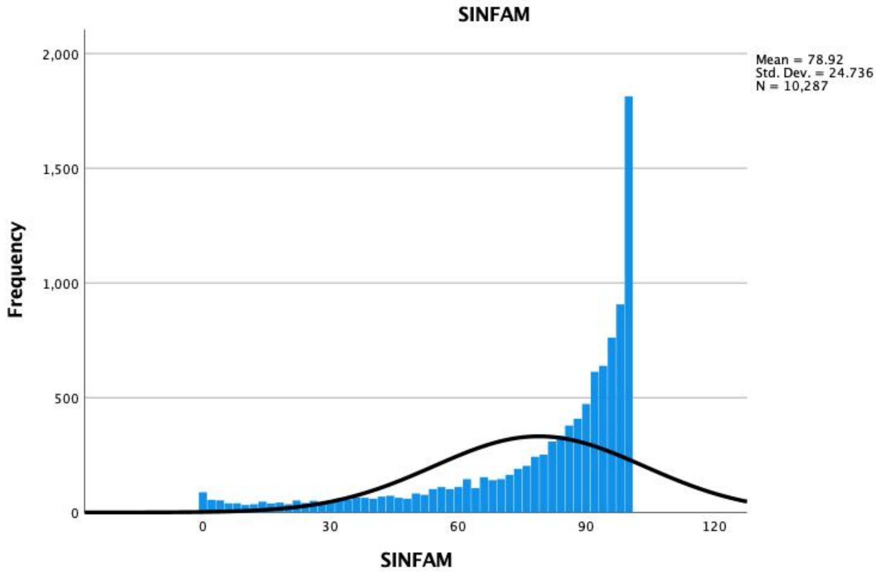

Figure D.17: Histogram percentage of single-family houses

D.3.18 Percentage of multi-family houses

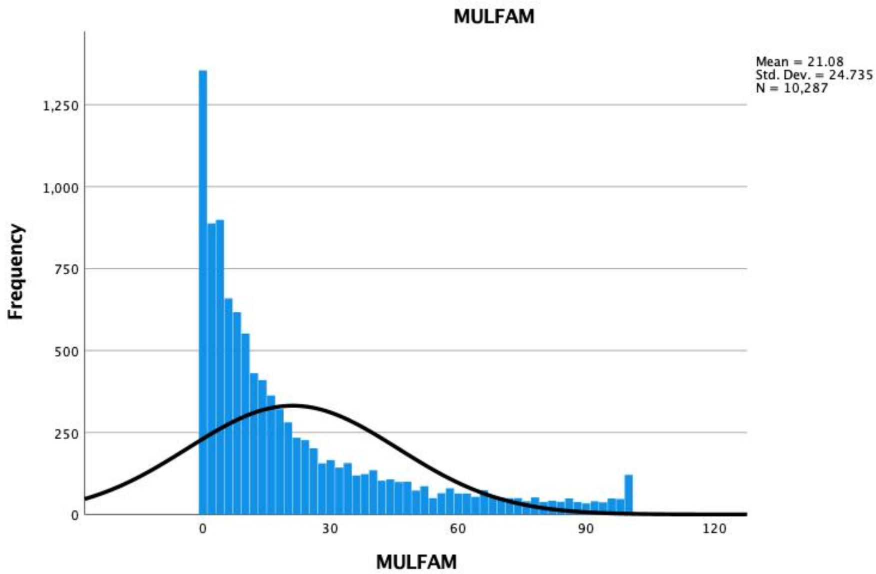

Figure D.18: Histogram percentage of multi-family houses

D.3.19 Percentage of owner-occupied houses

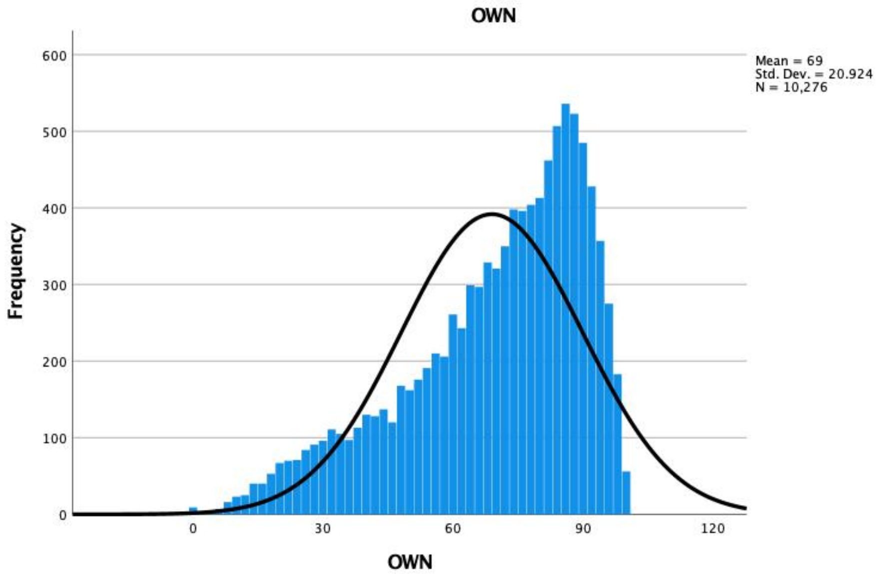

Figure D.19: Histogram percentage of owner-occupied houses

D.3.20 Percentage of rented houses

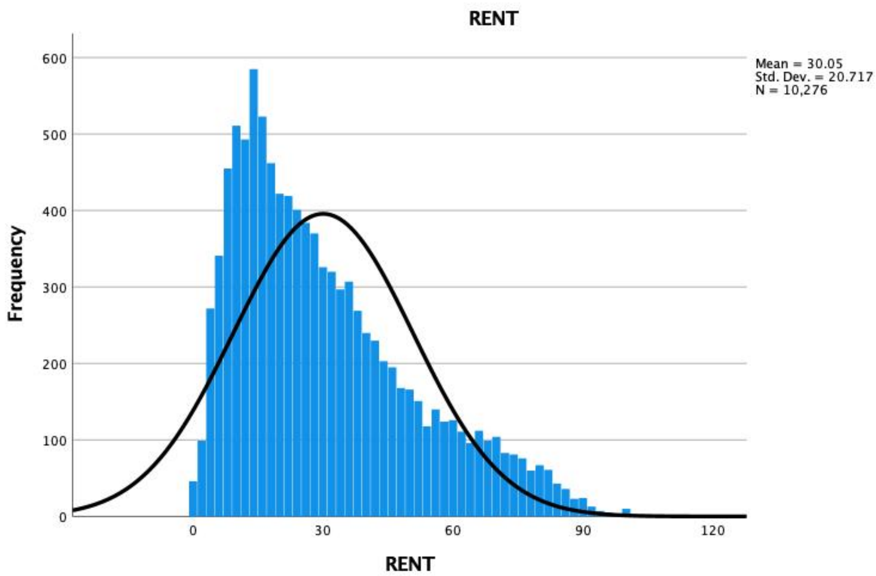

Figure D.20: Histogram percentage of rented houses

D.3.21 Percentage of property unknown houses

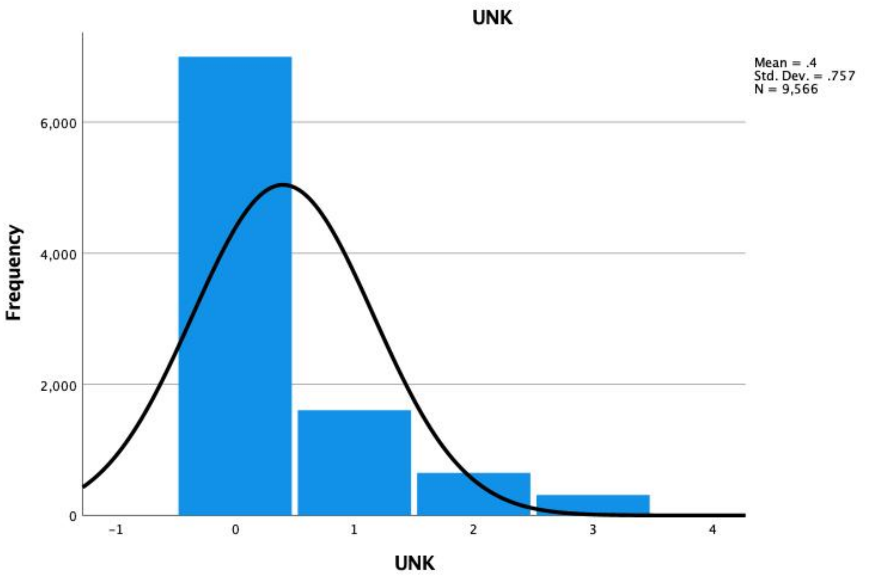

Figure D.21: Histogram percentage of property unknown houses

D.3.22 Percentage houses built before 2000

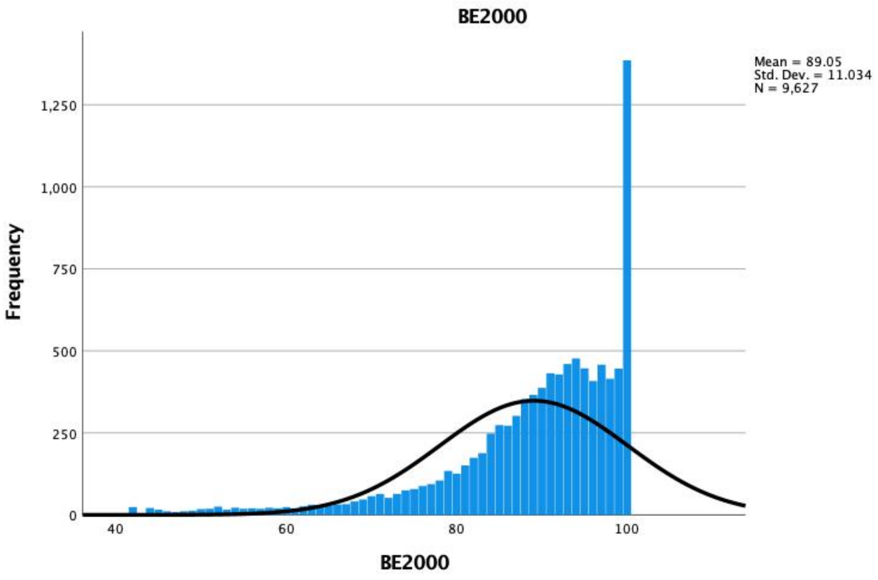

Figure D.22: Histogram percentage houses built before 2000

D.3.23 Percentage houses built after 2000

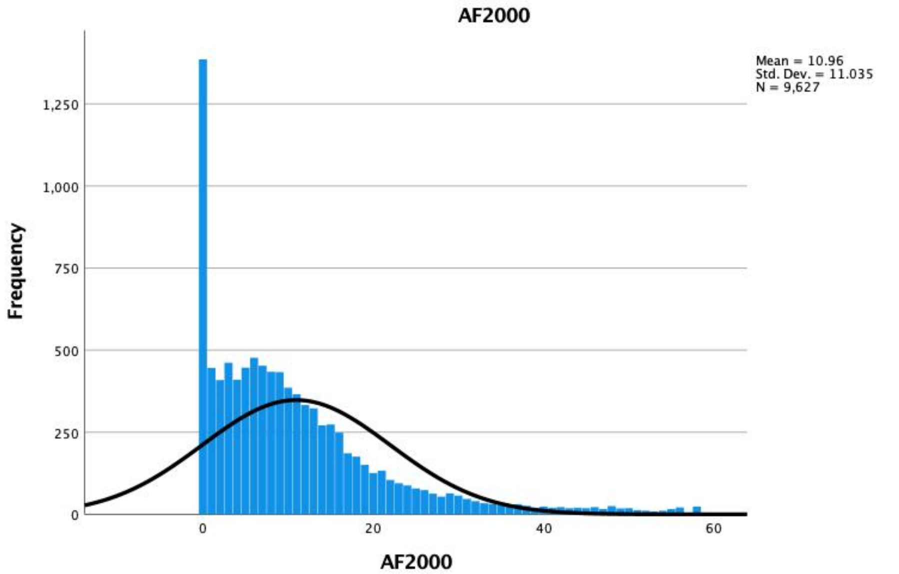

Figure D.23: Histogram percentage houses built after 2000

D.3.24 Populations density (residents per km^2)

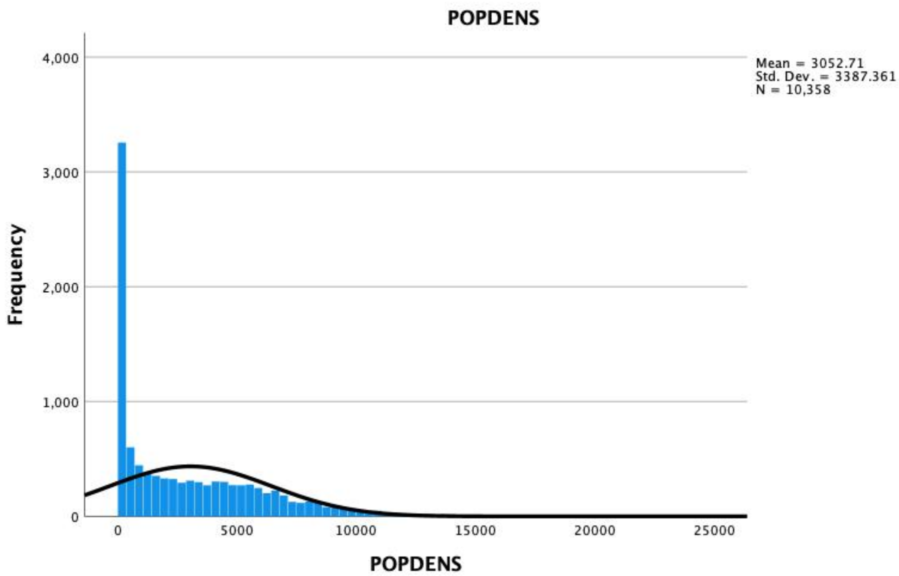

Figure D.24: Histogram populations density (addresses per km^2)

D.3.25 Neighbourhood address density

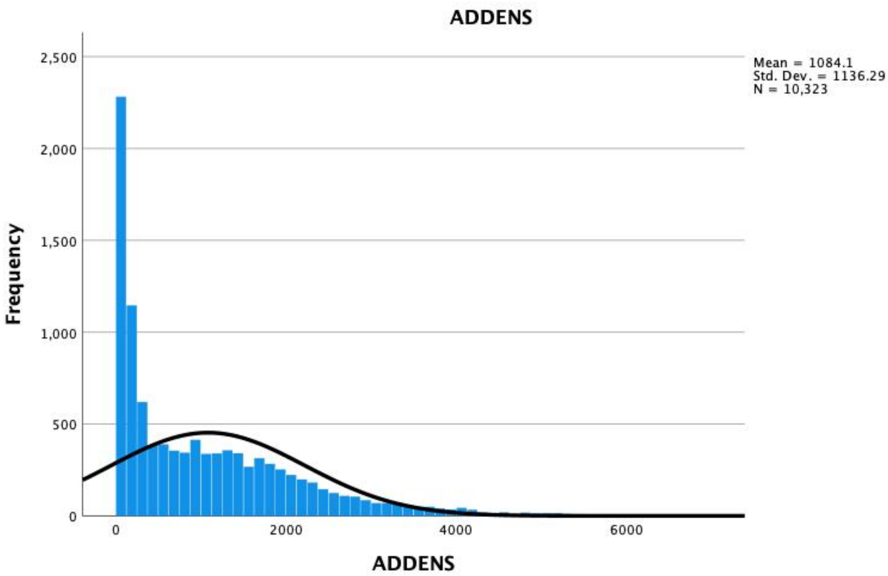

Figure D.25: Histogram neighbourhood address density

D.3.26 Degree of urbanity

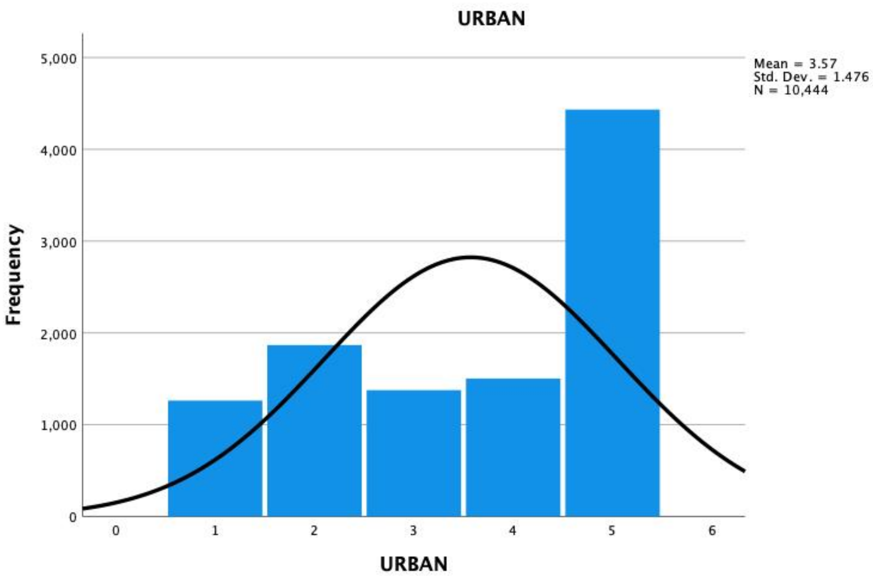

Figure D.26: Histogram degree of urbanity

D.3.27 Average housing value (x1000)

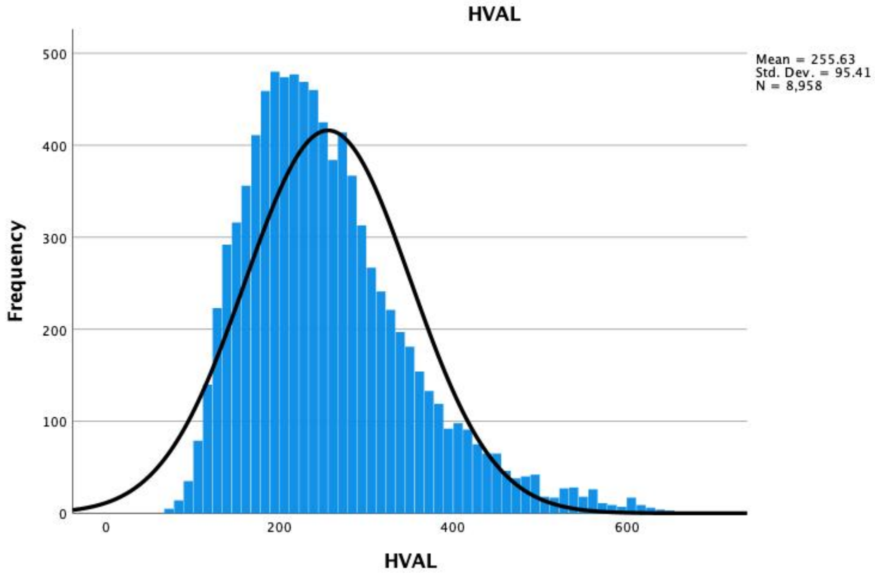

Figure D.27: Histogram average housing value (x1000)

D.3.28 40% persons with the lowest income

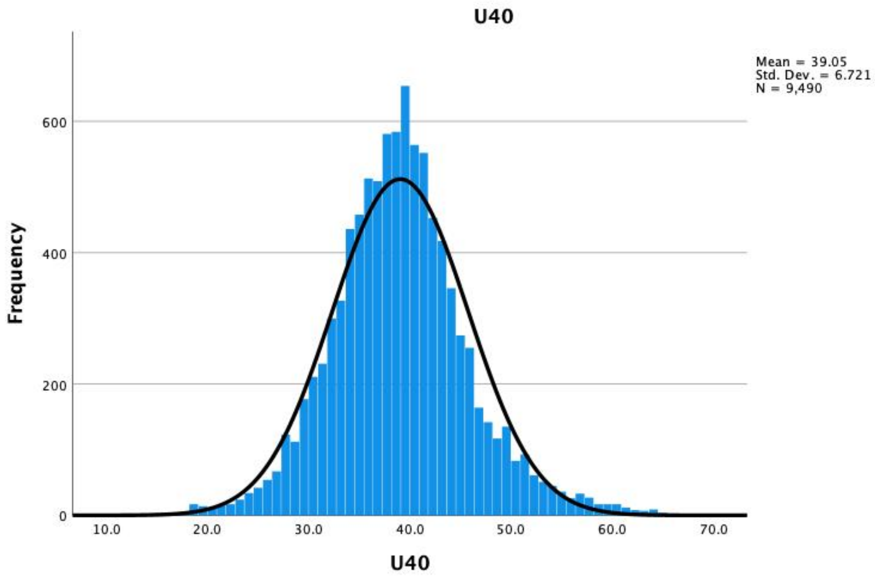

Figure D.28: Histogram 40% persons with the lowest income

D.3.29 20% persons with the highest income

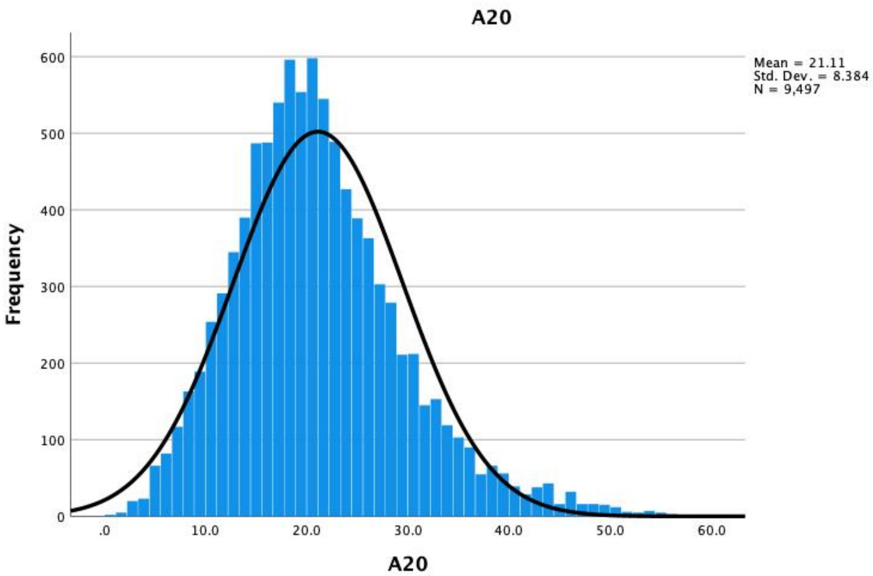

Figure D.29: Histogram 20% persons with the highest income

D.3.30 Number of people with an unemployment benefit

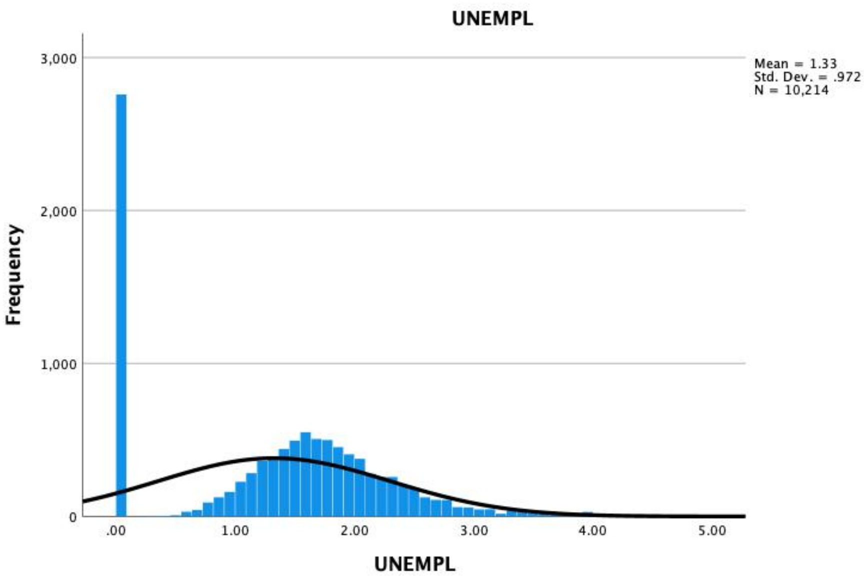

Figure D.30: Histogram number of people with an unemployment benefit

D.3.31 Amount of solar radiation

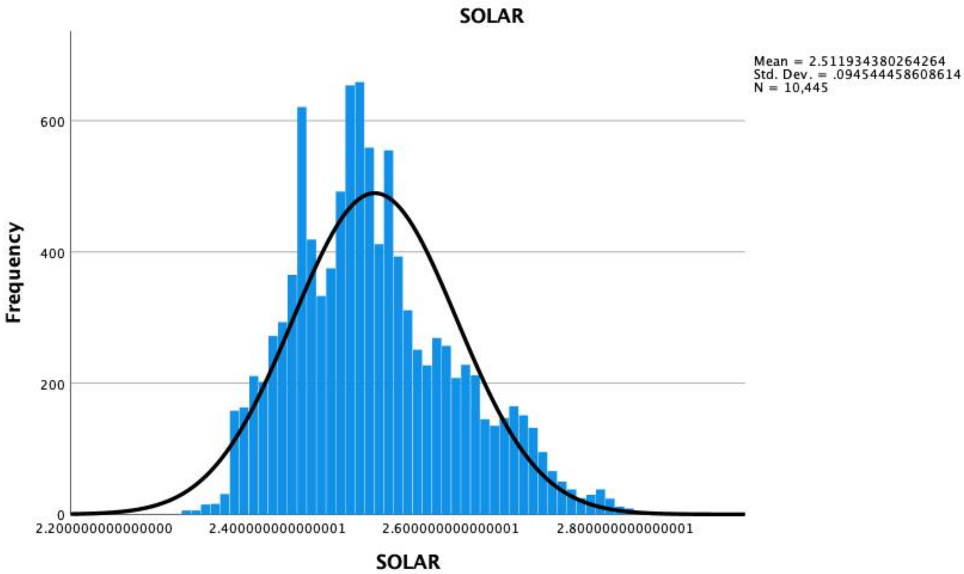

Figure D.31: Histogram amount of solar radiation

Appendix E

Spatial distribution of missing values per variable

Map of missing values per filled variable.

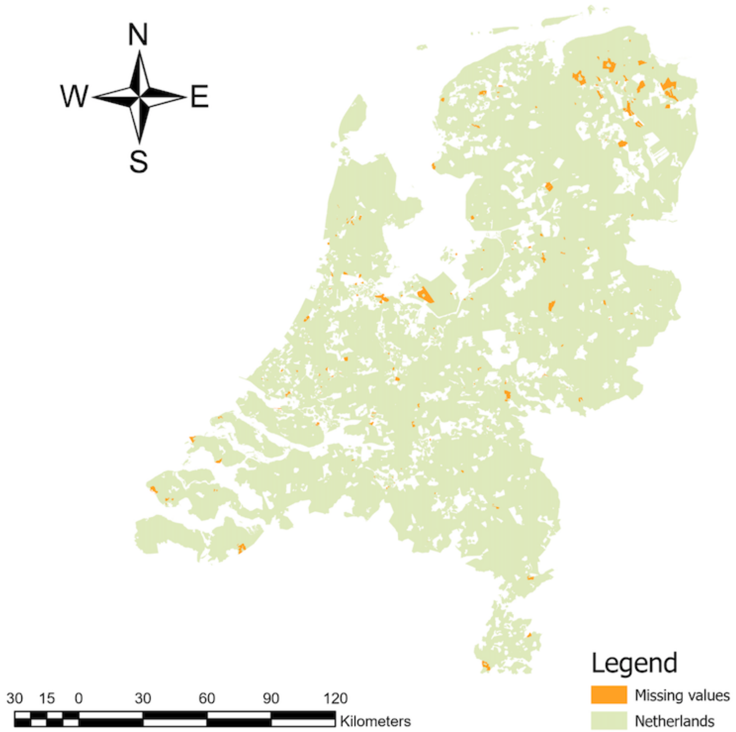

Figure E.1: Missing values of installations per house

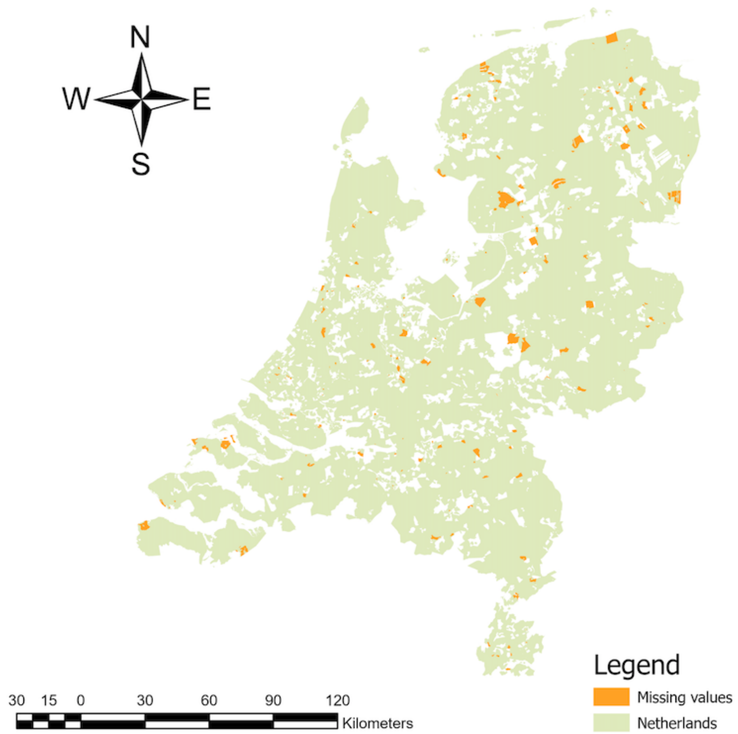

Figure E.2: Missing values of percentage of males

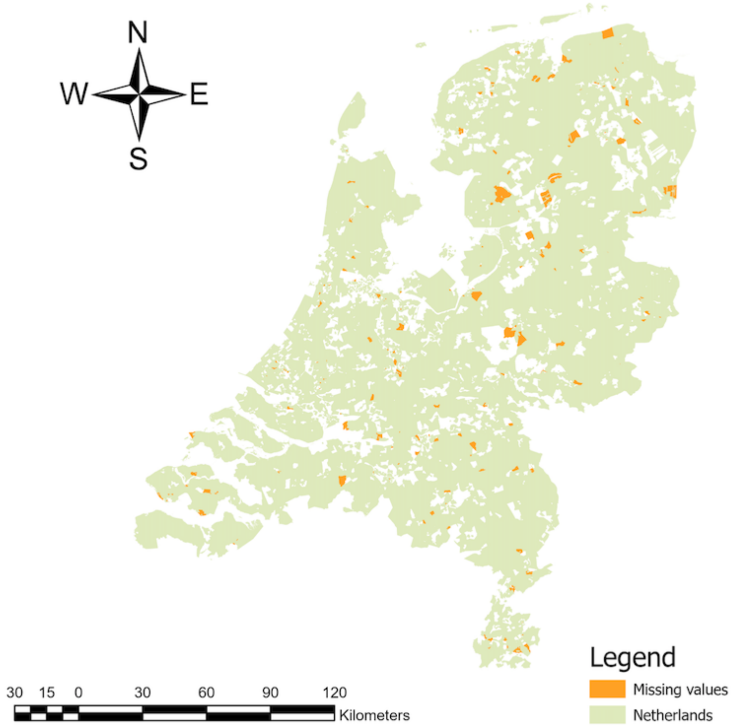

Figure E.3: Missing values of percentage of females

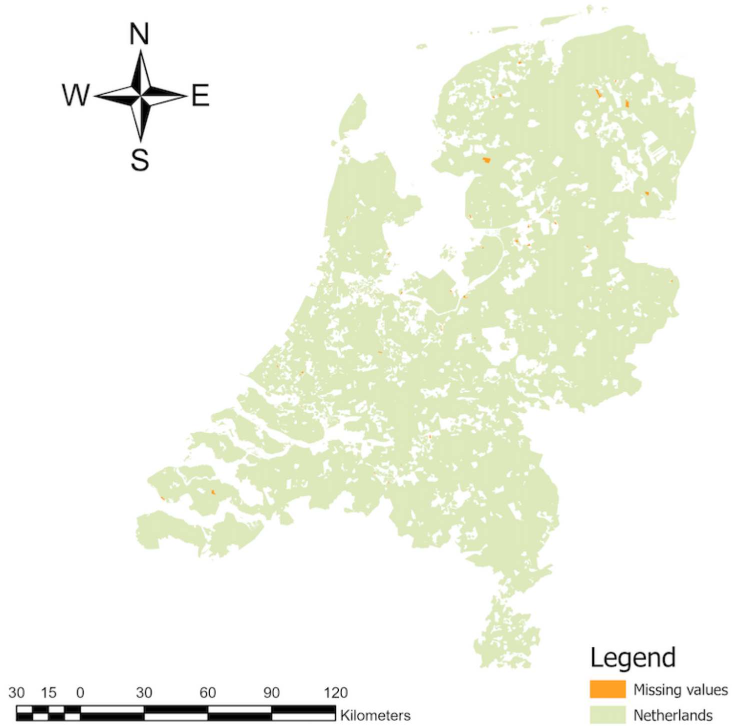

Figure E.4: Missing values of percentage of people between 0 and 14 years old

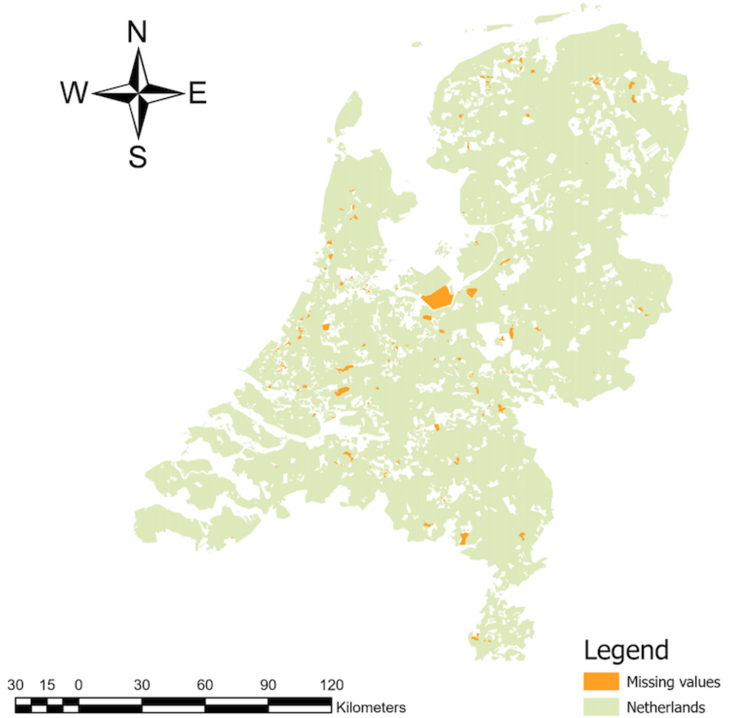

Figure E.5: Missing values of percentage of people between 15 and 24 years old

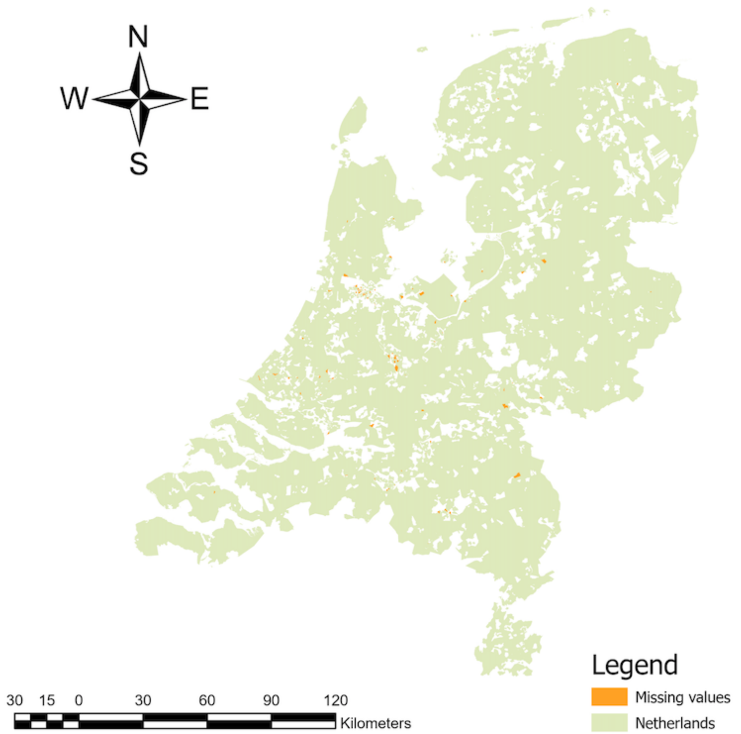

Figure E.6: Missing values of percentage of people between 25 and 44 years old

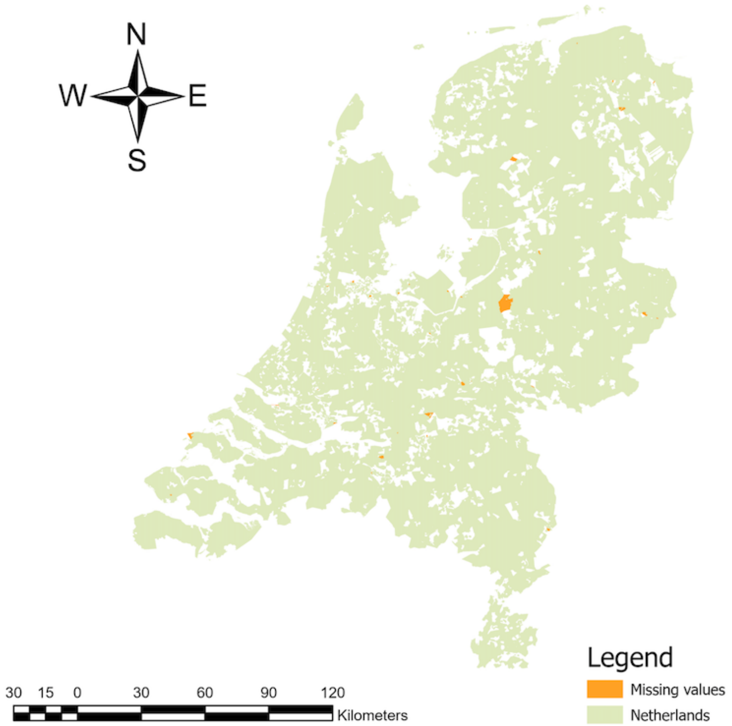

Figure E.7: Missing values of percentage of people between 45 and 64 years old

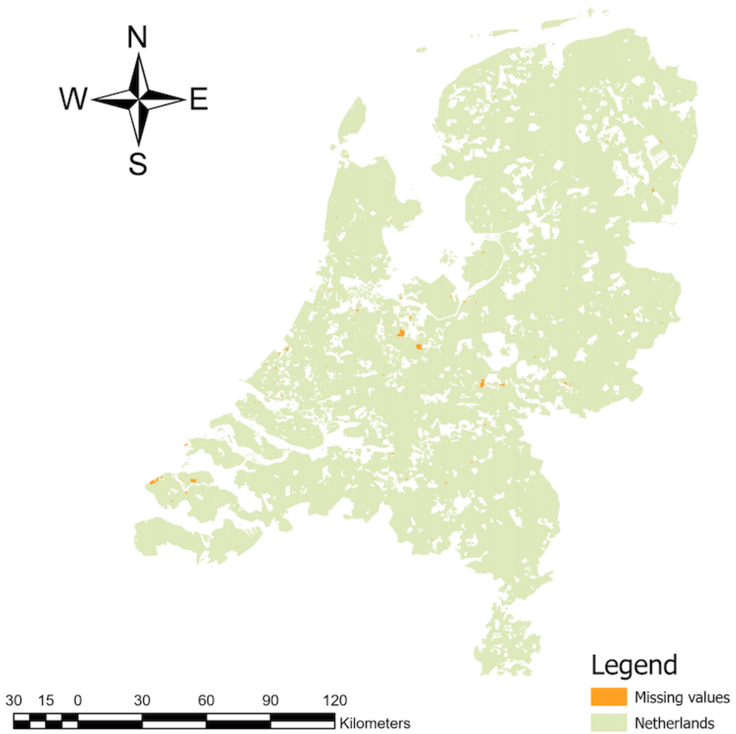

Figure E.8: Missing values of percentage of people 65 years old or older

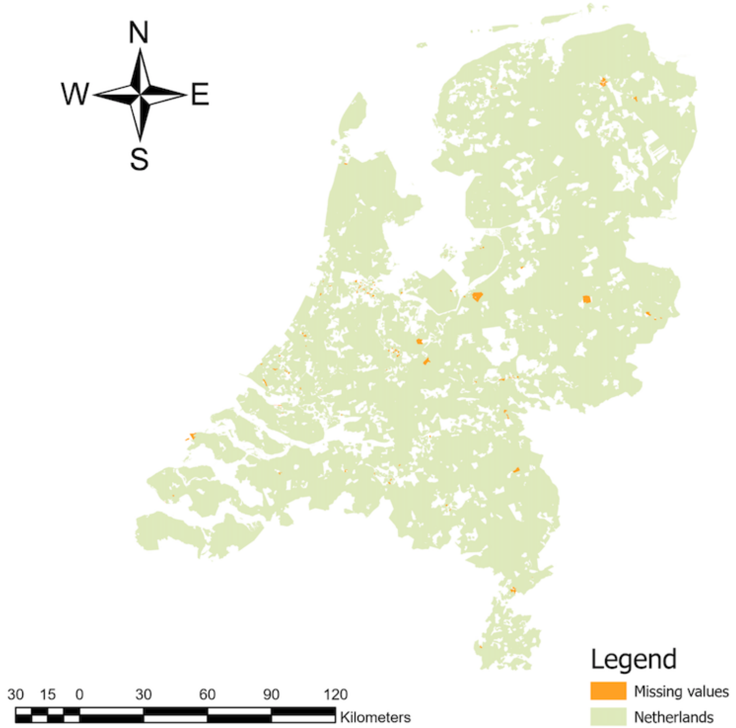

Figure E.9: Missing values of percentage of unmarried people

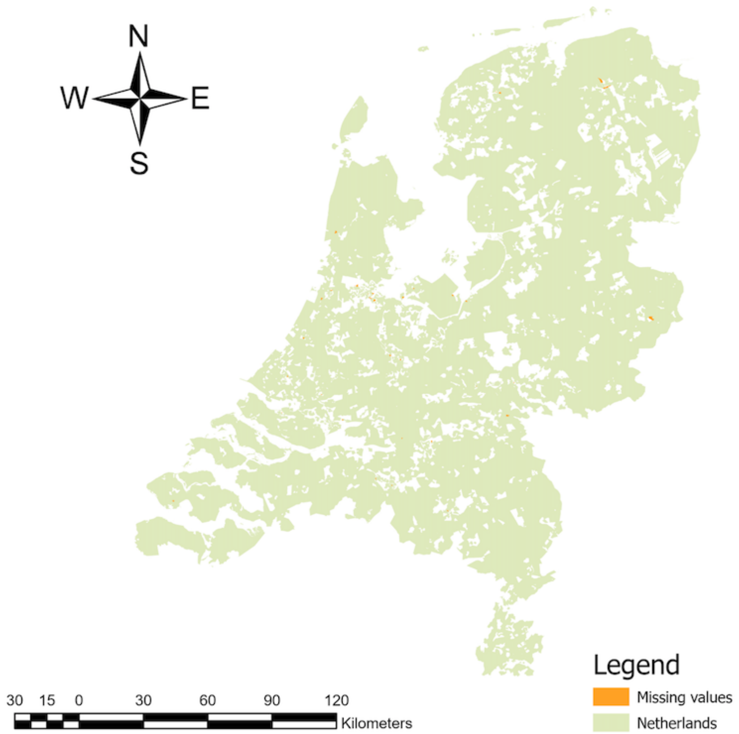

Figure E.10: Missing values of percentage married people

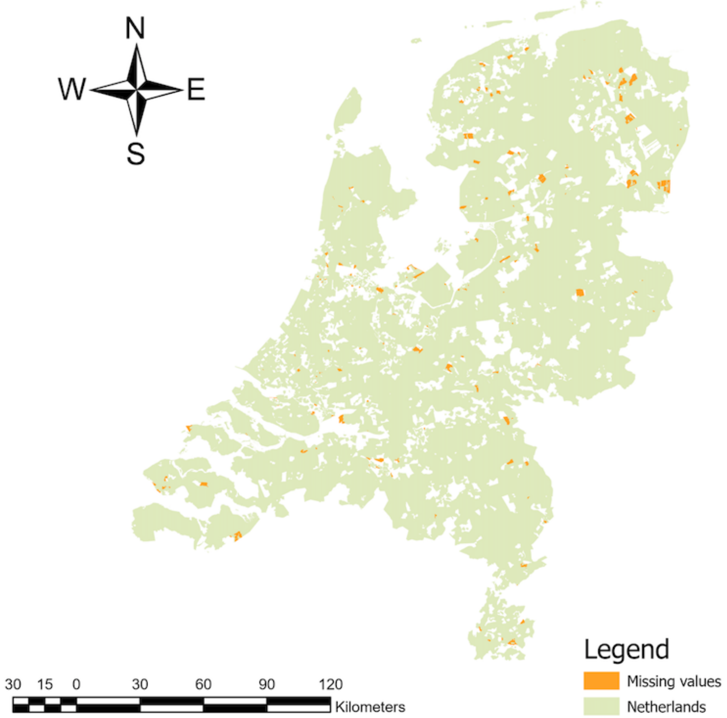

Figure E.11: Missing values of percentage of single-family houses

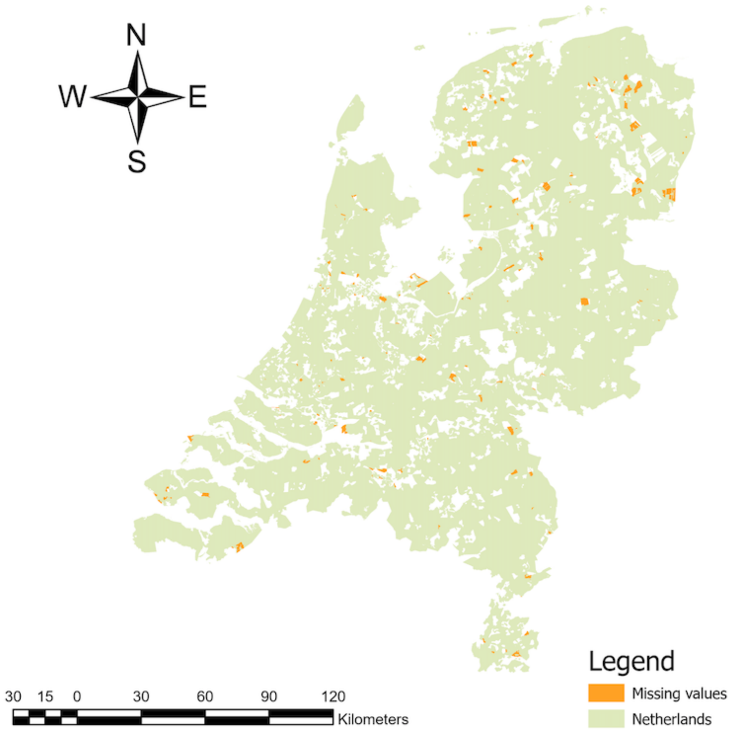

Figure E.12: Missing values of percentage of multi-family houses

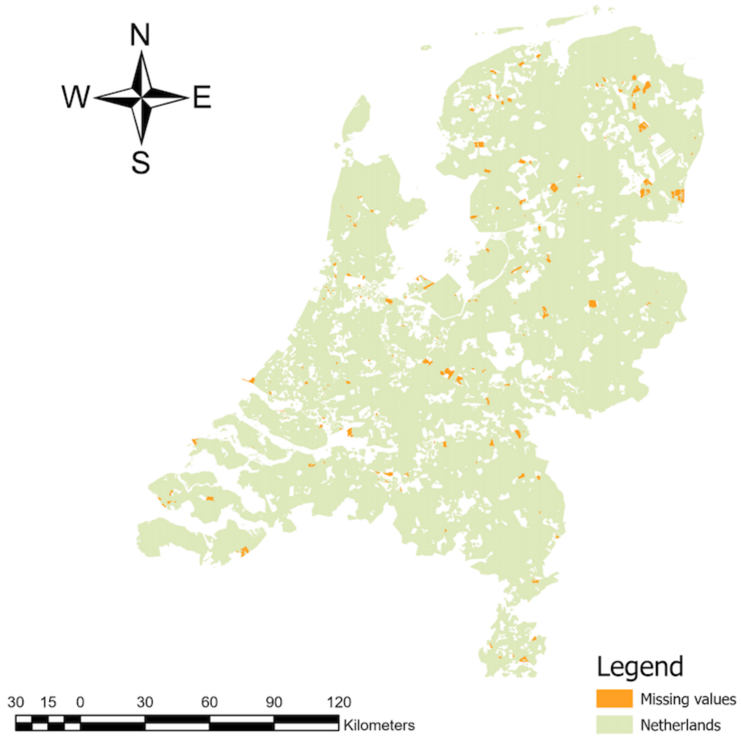

Figure E.13: Missing values of percentage of owner-occupied houses

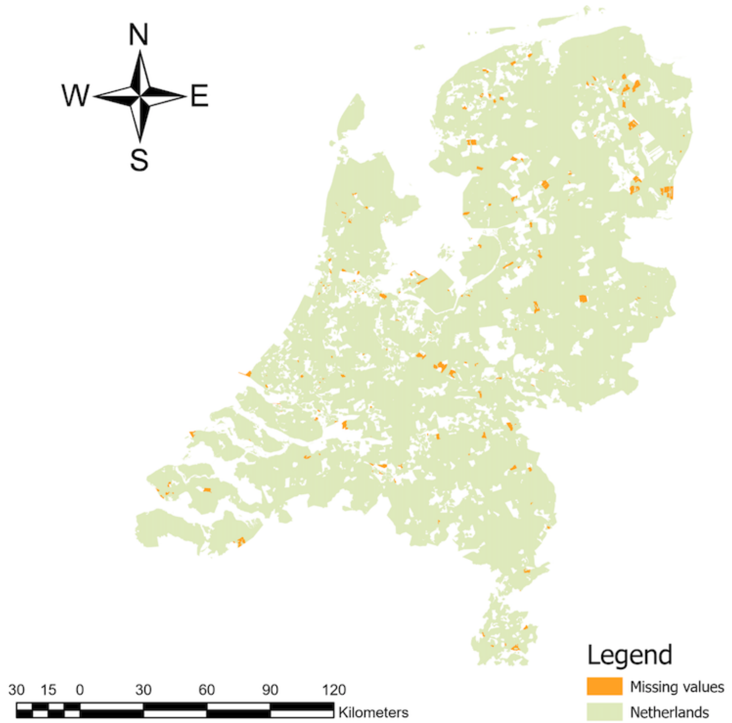

Figure E.14: Missing values of percentage of rented houses

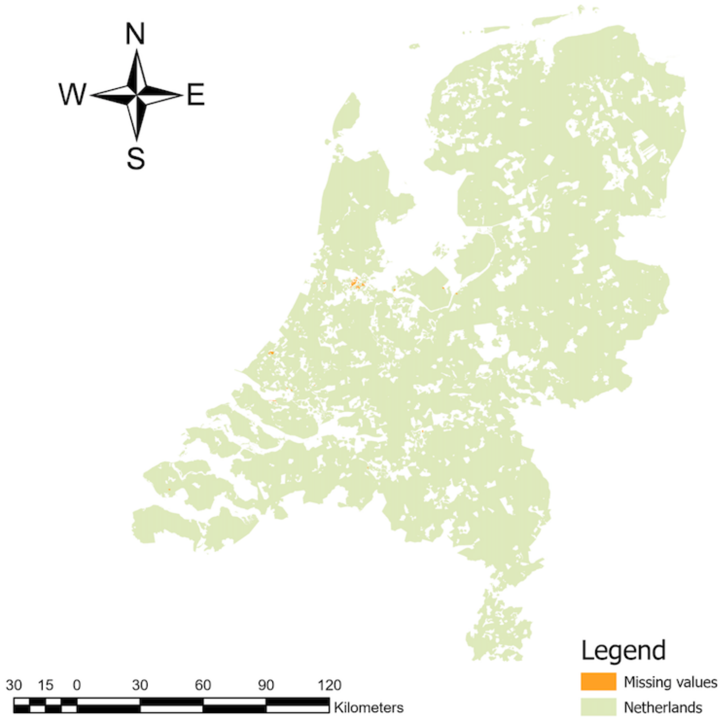

Figure E.15: Missing values of population density

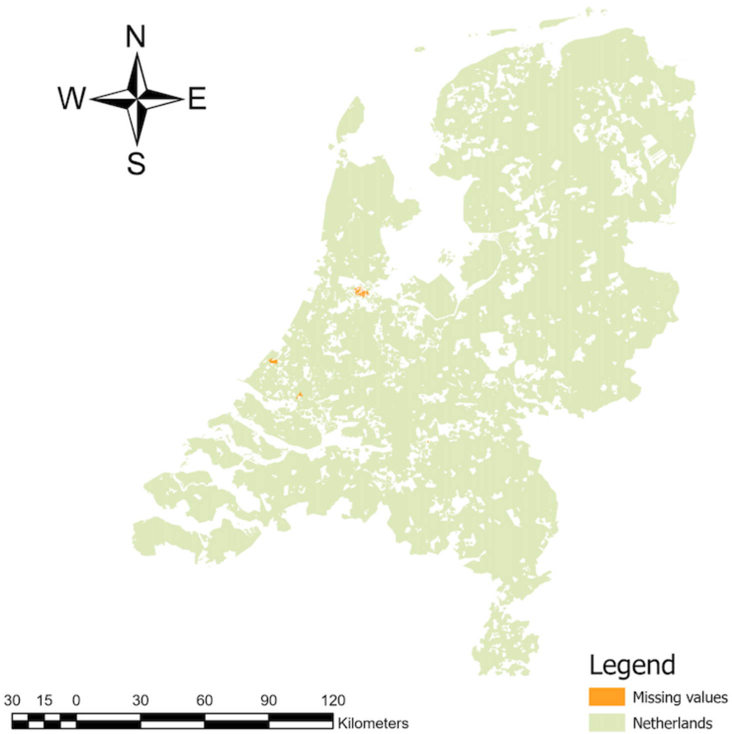

Figure E.16: Missing values of address density

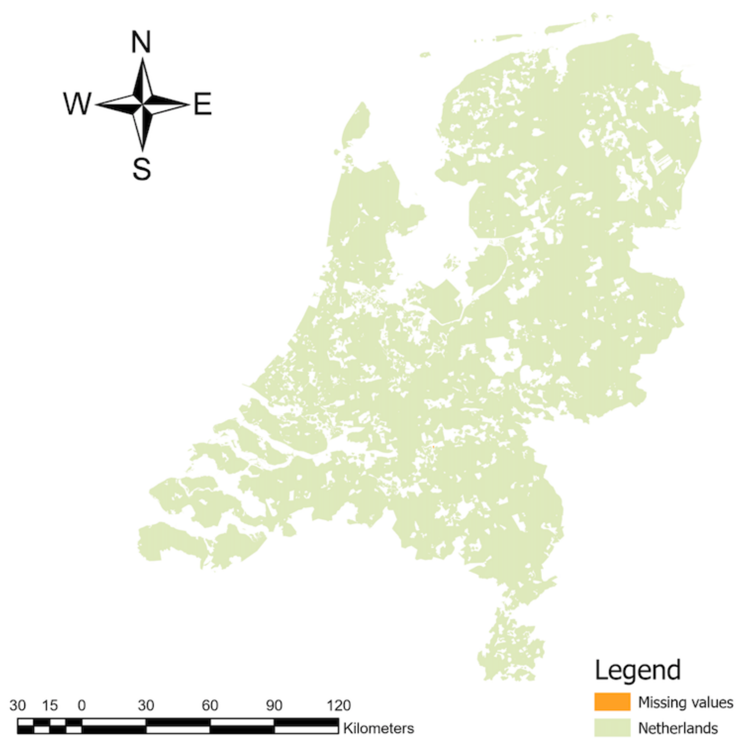

Figure E.17: Missing values of degree of urbanity

Appendix F

Correlation

In this appendix, the correlation table will be shown.

Table F.1: Correlation table

		MALES	FEMALES	A0014	A1524	A2544	A4564	A65	UNMAR	MAR	SINFAM
MALES	Pearson Correlation	1	-.889**	.021*	.240**	-.037**	.306**	-.289**	.082**	.132**	.310**
	Sig. (2-tailed)		0,000	0,031	0,000	0,000	0,000	0,000	0,000	0,000	0,000
FEMALES	Pearson Correlation	-.889**	1	-0,013	-.249**	.038**	-.303**	.293**	-.092**	-.148**	-.305**
	Sig. (2-tailed)	0,000		0,173	0,000	0,000	0,000	0,000	0,000	0,000	0,000
A0014	Pearson Correlation	.021*	-0,013	1	.048**	.305**	-.235**	-.477**	.355**	-.034**	.220**
	Sig. (2-tailed)	0,031	0,173		0,000	0,000	0,000	0,000	0,000	0,001	0,000
A1524	Pearson Correlation	.240**	-.249**	.048**	1	-.022*	.171**	-.382**	.285**	-.079**	.087**
	Sig. (2-tailed)	0,000	0,000	0,000		0,029	0,000	0,000	0,000	0,000	0,000
A2544	Pearson Correlation	-.037**	.038**	.305**	-.022*	1	-.589**	-.503**	.729**	-.672**	-.466**
	Sig. (2-tailed)	0,000	0,000	0,000	0,029		0,000	0,000	0,000	0,000	0,000
A4564	Pearson Correlation	.306**	-.303**	-.235**	.171**	-.589**	1	.032**	-.391**	.504**	.502**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000		0,001	0,000	0,000	0,000
A65	Pearson Correlation	-.289**	.293**	-.477**	-.382**	-.503**	.032**	1	-.658**	.344**	-.038**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,001		0,000	0,000	0,000
UNMAR	Pearson Correlation	.082**	-.092**	.355**	.285**	.729**	-.391**	-.658**	1	-.798**	-.396**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000
MAR	Pearson Correlation	.132**	-.148**	-.034**	-.079**	-.672**	.504**	.344**	-.798**	1	.644**
	Sig. (2-tailed)	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000		0,000
SINFAM	Pearson Correlation	.310**	-.305**	.220**	.087**	-.466**	.502**	-.038**	-.396**	.644**	1
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
MULFAM	Pearson Correlation	-.310**	.305**	-.220**	-.087**	.466**	-.502**	.038**	.396**	-.644**	-1.000**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
OWN	Pearson Correlation	.356**	-.361**	.142**	.088**	-.485**	.520**	-.025*	-.380**	.688**	.722**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,012	0,000	0,000	0,000
RENT	Pearson Correlation	-.354**	.359**	-.128**	-.084**	.494**	-.515**	0,012	.385**	-.682**	-.710**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,239	0,000	0,000	0,000
POPDENS	Pearson Correlation	-.324**	.326**	.083**	-0,017	.545**	-.421**	-.182**	.427**	-.534**	-.600**
	Sig. (2-tailed)	0,000	0,000	0,000	0,090	0,000	0,000	0,000	0,000	0,000	0,000
ADDENS	Pearson Correlation	-.336**	.323**	-.048**	-.020*	.527**	-.459**	-.091**	.439**	-.588**	-.706**
	Sig. (2-tailed)	0,000	0,000	0,000	0,047	0,000	0,000	0,000	0,000	0,000	0,000
URBAN	Pearson Correlation	.386**	-.371**	0,012	.068**	-.499**	.461**	.067**	-.394**	.551**	.671**
	Sig. (2-tailed)	0,000	0,000	0,215	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SOLAR	Pearson Correlation	-.068**	.050**	-0,011	-.038**	0,019	0,010	.062**	-.047**	.051**	-.048**
	Sig. (2-tailed)	0,000	0,000	0,262	0,000	0,055	0,300	0,000	0,000	0,000	0,000

continued on the next page

Table F.2: Correlation table

continued from the previous page

		MULFAM	OWN	RENT	POPDENS	ADDENS	URBAN	SOLAR
MALES	Pearson Correlation	-.310**	.356**	-.354**	-.324**	-.336**	.386**	-.068**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
FEMALES	Pearson Correlation	.305**	-.361**	.359**	.326**	.323**	-.371**	.050**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
A0014	Pearson Correlation	-.220**	.142**	-.128**	.083**	-.048**	0,012	-0,011
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,215	0,262
A1524	Pearson Correlation	-.087**	.088**	-.084**	-0,017	-.020*	.068**	-.038**
	Sig. (2-tailed)	0,000	0,000	0,000	0,090	0,047	0,000	0,000
A2544	Pearson Correlation	.466**	-.485**	.494**	.545**	.527**	-.499**	0,019
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,055
A4564	Pearson Correlation	-.502**	.520**	-.515**	-.421**	-.459**	.461**	0,010
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,300
A65	Pearson Correlation	.038**	-.025*	0,012	-.182**	-.091**	.067**	.062**
	Sig. (2-tailed)	0,000	0,012	0,239	0,000	0,000	0,000	0,000
UNMAR	Pearson Correlation	.396**	-.380**	.385**	.427**	.439**	-.394**	-.047**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MAR	Pearson Correlation	-.644**	.688**	-.682**	-.534**	-.588**	.551**	.051**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SINFAM	Pearson Correlation	-1.000**	.722**	-.710**	-.600**	-.706**	.671**	-.048**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MULFAM	Pearson Correlation	1	-.722**	.710**	.600**	.706**	-.671**	.048**
	Sig. (2-tailed)		0,000	0,000	0,000	0,000	0,000	0,000
OWN	Pearson Correlation	-.722**	1	-.991**	-.563**	-.602**	.612**	-.023*
	Sig. (2-tailed)	0,000		0,000	0,000	0,000	0,000	0,020
RENT	Pearson Correlation	.710**	-.991**	1	.577**	.608**	-.620**	.025*
	Sig. (2-tailed)	0,000	0,000		0,000	0,000	0,000	0,012
POPDENS	Pearson Correlation	.600**	-.563**	.577**	1	.816**	-.785**	.095**
	Sig. (2-tailed)	0,000	0,000	0,000		0,000	0,000	0,000
ADDENS	Pearson Correlation	.706**	-.602**	.608**	.816**	1	-.927**	.092**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000		0,000	0,000
URBAN	Pearson Correlation	-.671**	.612**	-.620**	-.785**	-.927**	1	-.090**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000		0,000
SOLAR	Pearson Correlation	.048**	-.023*	.025*	.095**	.092**	-.090**	1
	Sig. (2-tailed)	0,000	0,020	0,012	0,000	0,000	0,000	

Appendix G

Spatial distribution of residuals for OLS

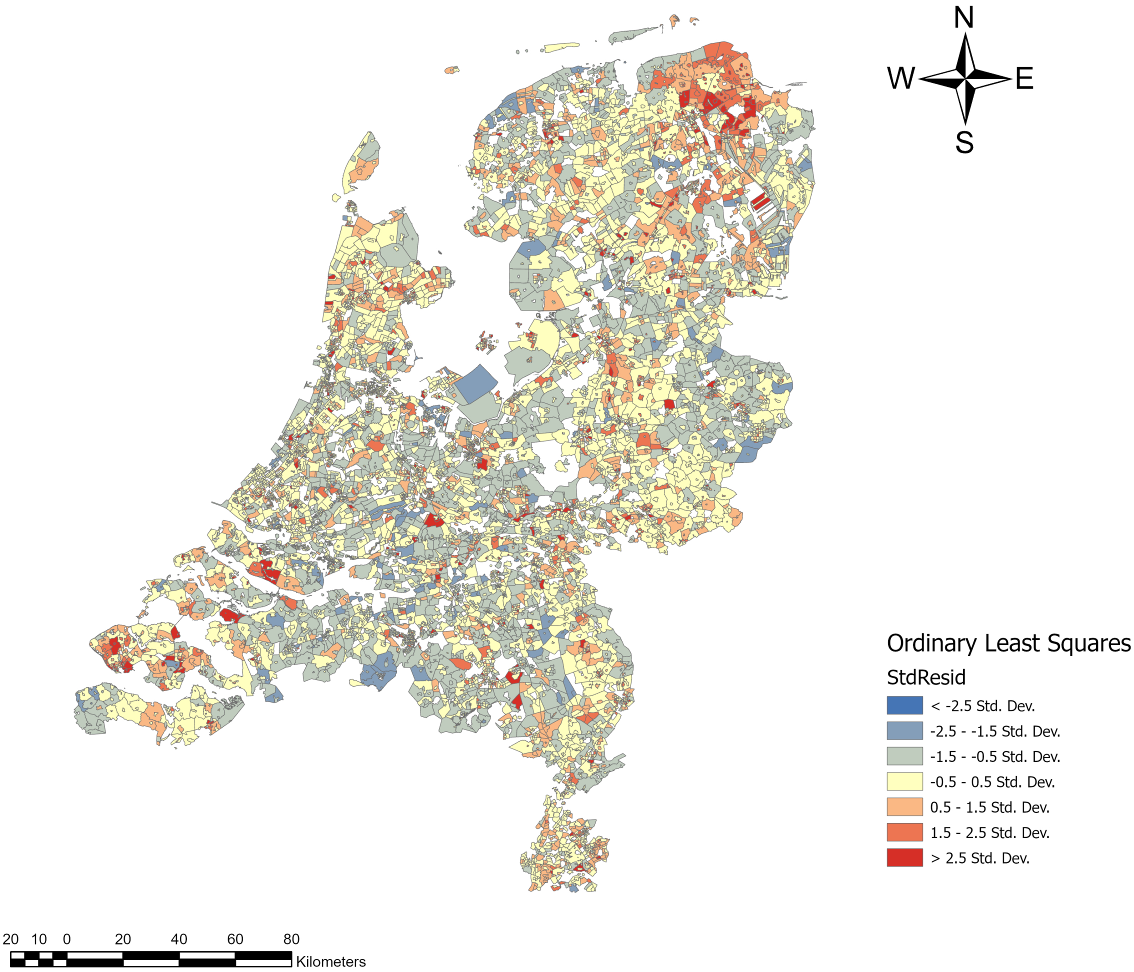

Figure G.1: The distribution of residuals for OLS model 1

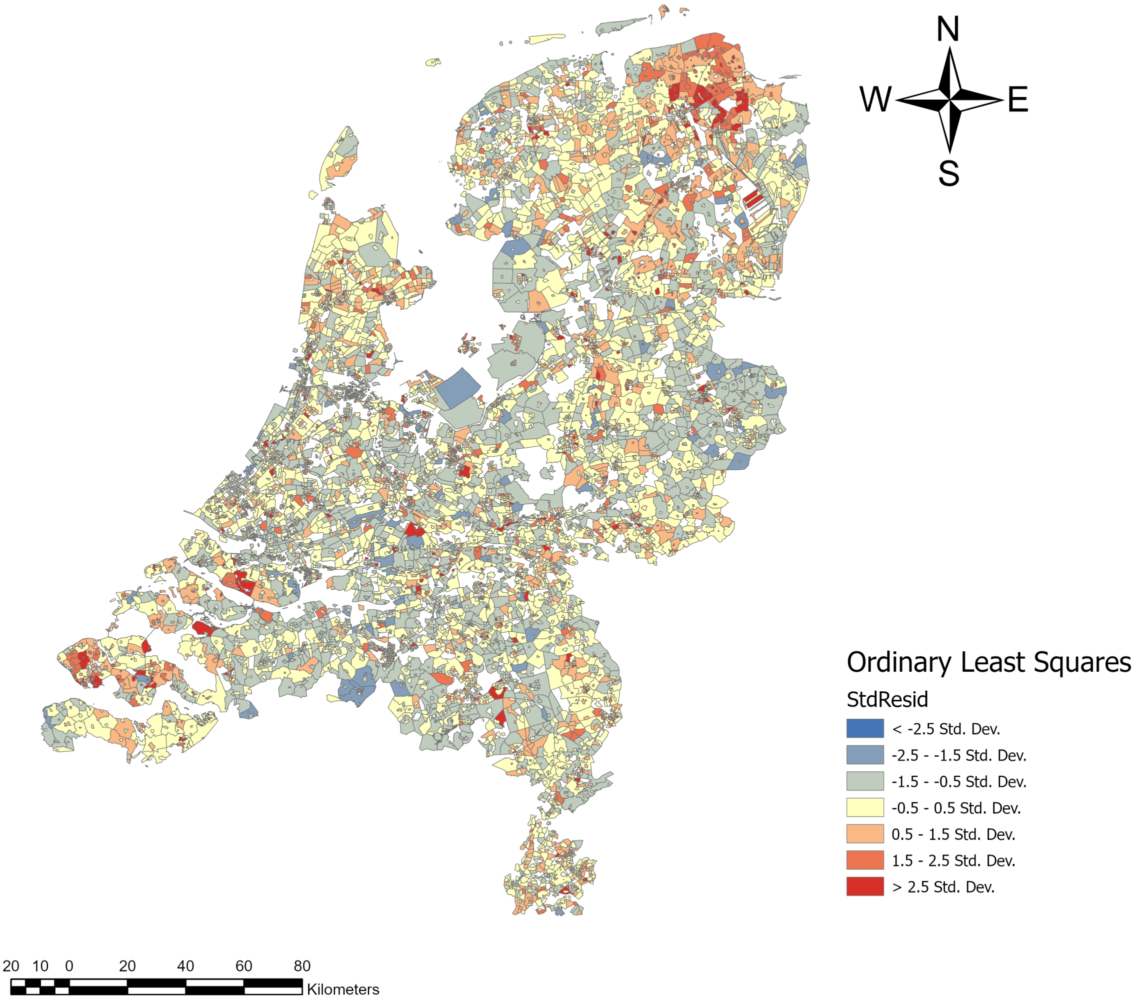

Figure G.2: The distribution of residuals for OLS model 2