Spatial adoption patterns of residential heat pumps and their impact on the grid

Master thesis

Author:
J.M. de Waardt

Supervisors University of Twente:
dr. ir. M.R.K. Mes
dr. ir. J.M.J. Schutten

Supervisor Coteq:
J.W. Egberts

April 21, 2020

UNIVERSITY OF TWENTE.

Management summary

The energy transition has great consequences for the electricity network managed by Coteq. The energy transition accelerates the use of sustainable technologies such as heat pumps (HP), electrical vehicles, and photovoltaics (solar panels). These sustainable technologies are expected to cause larger peaks in the supply and demand of electricity. The current electricity network is not designed for these peaks. For this reason, Coteq needs to perform the necessary investments in the electricity network to prevent future bottlenecks. Investments are mainly done on electricity cables and transformers. Electricity cables transport/distribute electricity and transformers connect different electricity networks of possibly different voltage levels. To determine the needed investments, Coteq needs a forecast on the expected growth of sustainable technologies. Using such a forecast, Coteq can cost-efficiently perform investments, meaning that there is sufficient but not an excessive capacity for the expected future growth. In this research, we focused on forecasting the yearly growth of HPs in Almelo, Goor, and Oldenzaal up to 2050. As for the other sustainable technologies, the growth of HPs is expected to be dependent on socio-demographic characteristics such as income. For this reason, we studied the spatial diffusion of HPs, which is the growth of HPs over space and time. In this research, we assumed that the growth of HPs is mainly determined by individual decisions and minimal municipal involvement. Also, we focused on determining the expected impact expressed in the additional peak load on medium to low voltage (MV/LV) transformers to which households are connected. This is summarized in the following main research question.

What is the impact on medium to low voltage transformers due to the spatial diffusion of residential heat pumps?

To determine the growth of HPs over space and time, we combined a logistic regression model with S-curved growth patterns. The logistic regression model required a classification of HP adoption per household. Because no data were available on residential HP adoption, we used energy consumption data to infer residential HP adoption. A HP uses electricity to generate heat. As heat is generated by electricity, gas consumption decreases and electricity consumption increases. We used a method called change point detection to detect this change in energy consumption. Based on additional classification rules, we classified a household as having installed a hybrid HP or not. The activity of gas connections was used to infer full-electric HP adoption. Having obtained a classification per household, we were able to apply a logistic regression model. Independent variables were obtained from socio-demographic data and data on buildings and addresses. Using the logistic regression model, we extracted influencing factors for HP adoption that were used in a simulation study. The simulation study combined the logistic regression model with S-curved growth patterns based on literature. To dynamically assign HPs to households over space and time, we introduced an empirical distribution and a Fisher's noncentral hypergeometric distribution. Both distributions were able to include the found influencing factors of HP adoption on which we based the probability of adopting a HP for a household in a certain year. The HP growth was eventually transformed to electrical load on MV/LV transformers based on worst-case conditions (i.e., cold winter weekday). Using an algorithm based on expert knowledge, we determined the required investments to prevent bottlenecks and provided a cost indication.

Using change point analysis in combination with data on gas connection removal, we were able to classify 99 households. Performing the logistic regression model on the identified households in combination with socio-demographic and building data suggested that age, property value, and degree of urbanity are associated with residential HP adoption. Households located in residential areas with higher fractions of age categories 25-44 and 45-64 were found to have a higher probability of HP adoption (p-value < 0.05). This was also found for households located in urban residential areas with higher property values (p-value < 0.05). Finally, households located in residential areas with higher fractions of age category 15-24 and houses with a smaller size of living area were found to have a smaller probability of HP adoption (p-value < 0.05). The outcome of the logistic regression model was based on a small sample of households with uncertain HP adoption. Using literature, we validated these results and found that property value and degree of urbanity were the most important influencing factors of HP adoption. We combined these results with S-curves based on a 35% (low) and 85% (high) final HP market share and a current HP market size of 1,400. In 2050, approximately 5 and 80 overloaded MV/LV transformer substations are expected for the low and high scenarios, respectively. Costs are expected to be €0.07 million and €1.2 million for the low and high scenarios, respectively. Also, as the current number of HPs is highly uncertain, we added two scenarios based on a current market size of 400 HPs. Based on a current market size of 400 HPs, the estimated year in which the first bottlenecks occur shifts from 2033 to 2036 and 2028 to 2032 for scenarios low and high, respectively.

We presented an approach to model the spatial diffusion of HPs introducing methods that were not used before in this context. We found that change point detection can partly support in inferring HP usage but has shown to perform poorly. Energy consumption data was found to be too unstable for this method. For further research, it is advised to focus on smart meter data. These data are more detailed and can more closely detect HP energy consumption behavior. Also, as the HP growth is currently in its beginning phase, more research is needed on the influencing factors of HP adoption. A survey was not performed due to time management. For further research, a survey can result in valuable information for Almelo, Goor, and Oldenzaal specifically. The HP requirements were based on empirical data. Ideally, HP requirements should be focused on empirical data for households in Almelo, Goor, and Oldenzaal. Further research should be focussed on the differences between hybrid HPs and full-electric HPs. Differentiating between these HPs is valuable for Coteq as a full-electric HP makes the gas connection unnecessary if also electrical cooking is used.

In summary, the growth of residential HPs can result in large investment requirements. To completely prevent bottlenecks in the worst-case scenario, Coteq needs to expand the capacity of up to 80 overloaded MV/LV transformers. We estimated the investment costs in the worst-case scenario to be €1.2 million. Overloaded MV/LV transformers are likely to occur between 2025 and 2030 depending on the current market size of HPs. HP growth is expected to be most severe in residential areas in urban areas with high property values.

Acknowledgements

This thesis is written as part of the Master's degree in Industrial Engineering and Management. This thesis was the last step in completing the study. I would like to thank the people involved.

First, I would like to thank Nelly and Jan-Willem at Cogas Groep who made the assignment possible. After a few meetings, we found an interesting assignment. It was a challenging assignment that would not have been possible to complete without the help of my colleagues at Cogas Groep. I would like to thank all colleagues for their collaboration and input.

Second, I would like to thank Martijn as first supervisor for his guidance during this project. The meetings were very useful and provided me with new insights. Also, I want to thank Matthieu and Marco for their feedback on my work.

Finally, I would like to thank friends, family, and in particular my girlfriend for contributing in their own way.

Martin de Waardt
April 21, 2020
Hengelo

Contents

Acronyms

ACM Autoriteit Consument & Markt. 11

ADMD After Diversity Maximum Demand. 49

AGO Almelo, Goor, and Oldenzaal. 1

AIC Akaike information criterion. 41

AOR adjusted odds ratio. 41

ASHP air source heat pump. 17

BAG Basisregistratie Adressen en Gebouwen. 9

C-AR Centraal Aansluitregister. 9

CBS Centraal Bureau voor de Statistiek. 9

CI confidence interval. 41

COP coefficient of performance. 17

DHPA Dutch Heat Pump Association. 19

DSO Distribution System Operator. 1

EC electrical cooking. 2

EHPA European Heat Pump Association. 19

EV electrical vehicle. 2

GSHP ground source heat pump. 17

HP heat pump. 2

HV high voltage. 11

ISDE Investeringssubsidie Duurzame Energie. 18

kVA kilovolt ampere. 14

kW kilowatt. 14

kWh kilowatt-hour. 17

LV low voltage. 6

MV medium voltage. 6

NAC normalized annual consumption. 36, 73

PI prediction interval. 42

PV photovoltaic (solar panel). 2

VIF variance inflation factor. 79

List of Figures

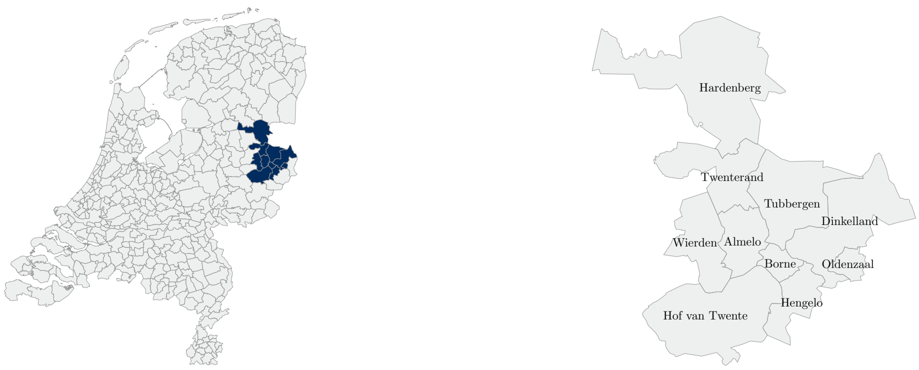
List of Tables

1 Introduction

Chapter 1 introduces this research and is organized as follows. Section 1.1 introduces the problem owner: Cogas Groep. Section 1.2 explains the motivation for this research. Section 1.3 introduces the problem. Section 1.4 describes the research goal and scope. Section 1.5 introduces the research questions. Section 1.6 describes the problem approach. Section 1.7 describes the deliverables. Finally, Section 1.8 describes the outline of this research.

1.1 Cogas Groep

This research is conducted at Cogas Groep in Almelo. Cogas Groep is a network company responsible for the development, construction, maintenance, and management of the network of gas, electricity, (sustainable) heat, and telecommunication in the service region shown in Figure 1.1. In Figure 1.1, Figure 1.1a shows the part of the Netherlands where Cogas Groep is active and Figure 1.1b shows the ten municipalities serviced by Cogas Groep. Cogas Groep is a holding with approximately 180 employees and consists of business-units Coteq, Cogas Infra, and Cogas Duurzaam. Coteq is a Distribution System Operator (DSO) and is responsible for the construction, operation, and maintenance of the gas and electricity network. The gas and electricity network connects approximately 143,000 gas and 54,000 electricity connections. Coteq distributes gas and electricity to houses located in Almelo, Goor, and Oldenzaal (AGO). Another DSO distributes electricity outside the AGO region where Coteq only distributes gas. Cogas Duurzaam is responsible for the development and realization of unregulated projects related to the energy transition and is also responsible for the telecommunication network. Cogas Infra provides activities on the gas and electricity network of Coteq but also provides activities for Cogas Duurzaam. Customers of Cogas Groep can count on an energy- and data- infrastructure that is safe, reliable, and affordable based on excellent network management. Also, the company's knowledge and experience, regional character, and position in society, enables an active involvement in the energy transition. Finally, Cogas Groep focuses on sustainable business operations concerning share capital, employees, and the environment. Although this research is relevant for all business-units of Cogas Groep, our focus lies mainly on the challenges experienced by business-unit Coteq.



(a) Service region Cogas Groep **(b)** Municipalities serviced by Cogas Groep

Figure 1.1: Layout of service region Cogas Groep including the municipalities serviced.

1.2 Research motivation

DSOs are facing new challenges due to the energy transition. As a result of the Paris Agreement, which is an agreement signed in 2016 to deal with global warming, the Dutch government has set the goal of a completely sustainable energy system by the end of 2050 (Rijksoverheid, 2019a). To achieve a sustainable energy system, renewable energy needs to replace fossil energy. The transition from fossil energy to renewable energy is called the energy transition. The energy transition accelerates the use of sustainable technologies. The acceleration of sustainable technologies is experienced by DSOs, as the number of electrical vehicles, photovoltaics (solar panels) and heat pumps is increasing. The growth of sustainable technology is expected to cause larger peaks in the supply and demand of electricity. Larger peaks in demand are mainly caused by heat pumps (HP), electrical vehicles (EV), and electrical cooking (EC). Larger peaks in supply are mainly caused by photovoltaics (PV). The current electricity network managed by DSOs is not designed for these large peaks in supply and demand. To facilitate the use of sustainable technologies, DSOs are faced with the task of performing the necessary investments in the electricity network such that enough capacity is available for the increase in supply and demand.

1.3 Problem description

Determining the required investments in the electricity network is a challenging task. Investments for capacity expansion are mainly performed on electricity cables and transformers (also called assets). Electricity cables transport/distribute electricity and transformers reduce the voltage levels between electricity networks. These assets, of which the lifetime can be more than 30 years, are generally either replaced or added to the electricity network to increase capacity. To determine the location and timing of the replacement or addition of assets, DSOs need to identify future bottlenecks in the electricity network. A bottleneck arises when demand (load) or supply (production) of electricity exceeds the capacity of the electricity network. Identifying these bottlenecks is challenging; demand and supply requirements differ per region and the exact future market share of sustainable technologies is unknown.

Traditionally, future requirements of the electricity network were determined by assuming a steady yearly growth of demand. A bottleneck was then identified by determining the year in which demand exceeded capacity. This approach is no longer suitable as the effect of upcoming sustainable technologies is not considered. To overcome this limitation, Coteq currently uses a forecast that estimates the number of PV installations per residential area[1] for multiple scenarios. This forecast provides Coteq with insight into the future requirements of the electricity network considering PV but does not consider other sustainable technologies. As a consequence, Coteq is not able to completely determine the future requirements of the electricity network. This makes it more difficult for Coteq to cost-efficiently facilitate the energy transition in its service region. We summarize this observation in the following problem statement:

It is unknown what effect sustainable technologies other than photovoltaics have on the requirements of the electricity network of Coteq.

[1] We often use the term *residential area* or *geographical area* to refer to a group of residential buildings demarcated as homogeneous based on characteristics such as income.

Solving the problem as stated in the problem statement is not feasible due to time management. For this reason, we focus on one sustainable technology. We choose to focus on HPs because the growth of this technology also affects the gas network of Coteq. Although we do not consider the gas network explicitly, the effect of HPs on the electricity network can be used by Coteq to determine the consequences for the gas network. Using Figure 1.2 we research the problem context of HPs and identify the core problem. We now briefly explain this problem context.

A HP is one of the sustainable technologies for heating and (partly) replaces the use of natural gas. The transition from heating systems that use natural gas to heating systems that use renewable energy is called the heat transition. The heat transition is part of the energy transition and is specifically focused on the phase-out of natural gas in the built environment. The heat transition contributes to the growth of HPs in two ways: municipal (Dutch: gemeentes) involvement and stimulation via subsidies and financial arrangements. Municipalities are given the responsibility to steer the phase-out of natural gas in the built environment. Municipalities have to make plans that describe which districts will no longer use natural gas and what the heating alternative (e.g., HPs) will be (Appendix A can be consulted for more information on the involvement of municipalities in the energy transition). Subsidies and financial arrangements from the Dutch government are used to stimulate households in taking energy saving measures. The support from the Dutch government in combination with financial and environmental considerations of households might drive households to purchase[2] a HP individually. Both the municipal plans and the stimulation via subsidies and financial arrangements are expected to affect the future share of HPs but the exact contribution is unknown. The unknown future share of HPs has as a consequence that Coteq is not able to determine the effect of HPs on the electricity network. This eventually leads to missing information for investment decisions and possibly inefficient facilitation of the energy transition in its service region.

In this research, we assume that the increase of HPs is mainly caused by individual decisions and that the municipal plans will be less important. We refer to the growth of HPs based on individual decisions as autonomous growth. There are several arguments for assuming autonomous growth. First, individuals can always purchase a HP although municipalities might have other plans. Second, there is possibly not enough support for policy goals and interventions (Vringer & Carabain, 2020). As we do not consider municipal plans and cannot influence the autonomous growth of HPs, we must focus on the uncertainty in the future share of HPs (illustrated as the grey node in Figure 1.2). In this research, we take an additional step by also focusing on the effect of HPs on the electricity network.

[2]We interchangeably use the terms *purchase*, *installation*, *usage*, *adoption*, and *deployment* to refer to the action of a household installing a technology.

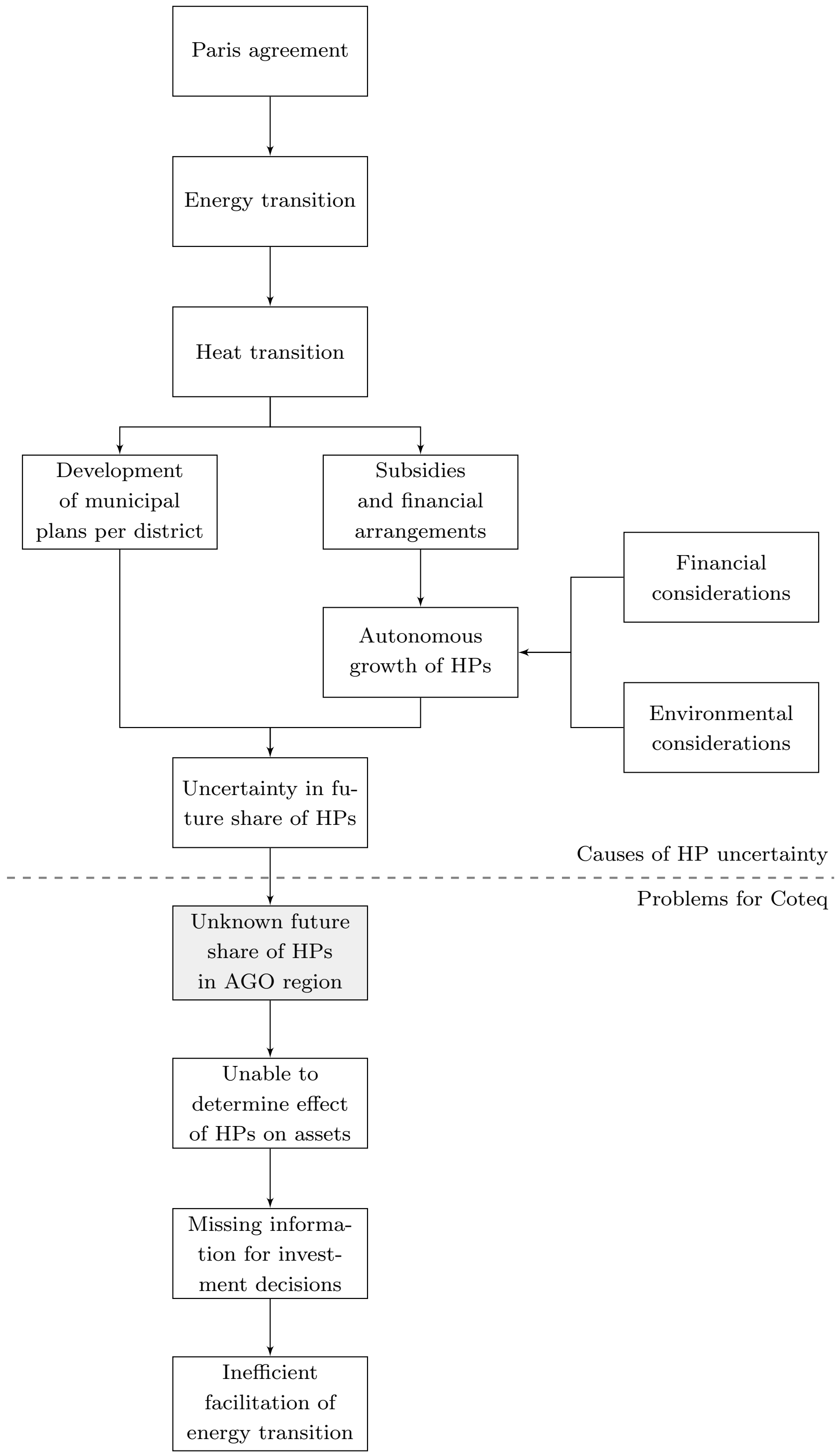


Figure 1.2: Overview of causes that lead to the problems experienced by Coteq.

1.4 Research goal

Given the problem description in Section 1.3, we focus on the following research goal.

Develop a model that quantifies the impact from increased residential heat pump installation on the electricity network of Coteq.

As this is a broad research goal we scope our research as follows.

Electricity network impact

We define impact as the additional peak load on the electricity network as a result of HP adoption. The peak load is the maximum load on the electricity network at a point in time by aggregating the demand for multiple connections (e.g., households) in a part of the electricity network. The additional peak load is determined by only focusing on the load resulting from HPs.

Space and time

Supply and demand differ per region as mentioned in Section 1.3. It is expected that the likelihood of HP installation for a household is related to certain characteristics of residential areas (e.g., income). For this reason, we determine the local impact on the electricity network per year up to 2050 for multiple scenarios. We focus on determining the growth of HPs over space and time (referred to as *spatial diffusion* or *spatial adoption patterns*). Determining the yearly impact is in line with the yearly market exploration performed by Coteq. The time horizon is based on the deadline resulting from the Paris agreement.

Consumer focus

All consumers that are not households are out of scope. Households are expected to have comparable energy consumption behavior. In addition, local characteristics that influence HP adoption are more easily determined for households than for companies.

Municipal plans

Municipal plans, which are currently in development, are not considered. This is based on the assumption that it will be difficult for municipalities to force households to adopt a certain technology. In addition, households can always purchase a HP, even if, for example, district heating is chosen as a heating alternative by a municipality.

Choice of energy transition technology

We focus on HPs. EVs, EC, and PV installations are not considered. We do describe heating alternatives other than HPs as this is required for a complete understanding of the heat transition. For example, deployment of district heating in a certain district can already indicate that HPs are less likely to be installed.

Data availability

To determine the characteristics of residential areas that might influence HP adoption, we need to know which residential areas already have installed HPs. In contrast to PV and EV, there is no publicly available database on the number of HP installations per residential area. We do not perform a survey as an approach to determine the current number of HP installations per residential area as this would not be feasible due to time management. The current number of HPs per residential area is determined by classifying HP usage per household using energy consumption data. By grouping the HP classifications per residential area we can estimate the current number of HPs per residential area. We validate this approach using literature. Other approaches, such as obtaining data from HP installers, are not considered because of privacy considerations.

Assets

We focus on the impact of residential HPs on medium to low voltage (MV/LV) transformers. The impact on electricity cables is more complex and is not considered due to time management. Also, smart grid approaches are not considered. Smart grid approaches are focused on shifting electricity demand in time and applying load control to reduce peak loads. Coteq manages multiple low voltage (LV) electricity networks to which households are connected. The LV electricity network is connected to the medium voltage (MV) electricity network through a MV/LV transformer substation. Because the capacity of a MV/LV transformer substation is determined by the capacity of the MV/LV transformer, we focus on the impact on MV/LV transformers.

Aggregation level

We do not research the characteristics of individual households but only of residential areas. For determining the impact on the electricity network, it is sufficient to differentiate between residential areas. We assume that households within a residential area are identical. Households connected to a MV/LV transformer substation are not necessarily located in the same residential area. This is illustrated with an example shown in Figure 1.3. The households connected to MV/LV transformer substation T are spread over three residential areas: A, B, and C. To determine the impact on the MV/LV transformer of substation T, we thus need to know the characteristics of the households in residential areas A, B, and C.

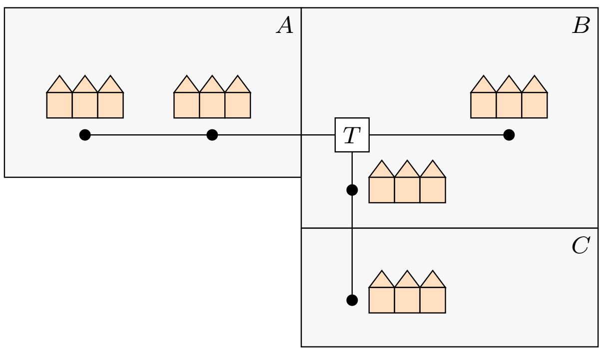


Figure 1.3: Example of a situation where houses connected to a MV/LV transformer substation (T) are located in multiple residential areas (A, B, and C).

Peer effects

Peer effects are not considered. The peer effect is the degree in which the action of one person affects the behavior of another person. Given the unavailability of the growth of HPs per household per year, it is not possible to determine peer effects statistically.

Infrastructure

Due to time management, researching possible developments related to infrastructure (e.g., demolish of houses) is out of scope. Also, changes in the electricity network (e.g., expansion) are not considered. This has as consequences that we mainly focus on existing houses.

Technological developments and economic considerations

To keep the research manageable, we do not explicitly research the relation between the technological development of HPs and HP adoption. This also holds for economic considerations such as price. Also, we do not research other possibly upcoming technologies that might become relevant in the future.

1.5 Problem approach

We approach the problem by following the four steps shown from Figure 1.4, where the dashed lines indicate needed input. First, we need to determine which households have installed a HP. We detect HP usage based on energy consumption data. A HP causes a decrease in gas usage and an increase in electricity usage. This change in energy consumption is detected using energy consumption data per household. When the decrease in gas usage and the increase in electricity usage meets pre-determined criteria, we classify a household as having installed a HP. Because we do not know exactly if a household has installed a HP, we denote this as an approximate classification. Second, we determine what characteristics influence HP adoption using logistic regression. Independent variables per residential area are extracted from socio-demographic data (e.g., income) and data on buildings and addresses (e.g., construction year). The dependent variable is obtained from the approximate classification. For each household, we thus have an indication of HP usage and characteristics that correspond to the residential area in which the household is located. As there are many variables available, we make a pre-selection using literature. The outcome of the logistic regression model is validated using literature. This is done because the use of energy consumption introduces additional uncertainty in the logistic regression model. Third, we estimate the expected growth of HPs per year for multiple scenarios up to 2050. Using the scenarios in combination with the results of the logistic regression model, we can determine the growth of HPs over space and time. We do this by applying a simulation study. This simulation study dynamically assigns HPs to households, meaning that households being assigned a HP in 2021 cannot be assigned a HP in 2022. In a year, HPs are assigned according to the output of the logistic regression model. For example, if we find that residential areas having a high average income are more likely to adopt HPs, then households within these residential areas have a higher probability of being assigned a HP in the simulation. Within a residential area, HPs are randomly assigned to households as we do not consider the characteristics per household. We can then determine the impact on MV/LV transformers since we know which households are connected to which MV/LV transformer substation. To determine the impact, we need the network topology (MV/LV transformer substation locations, MV/LV transformer capacities, current peak loads, and connections) and the required power of HPs.

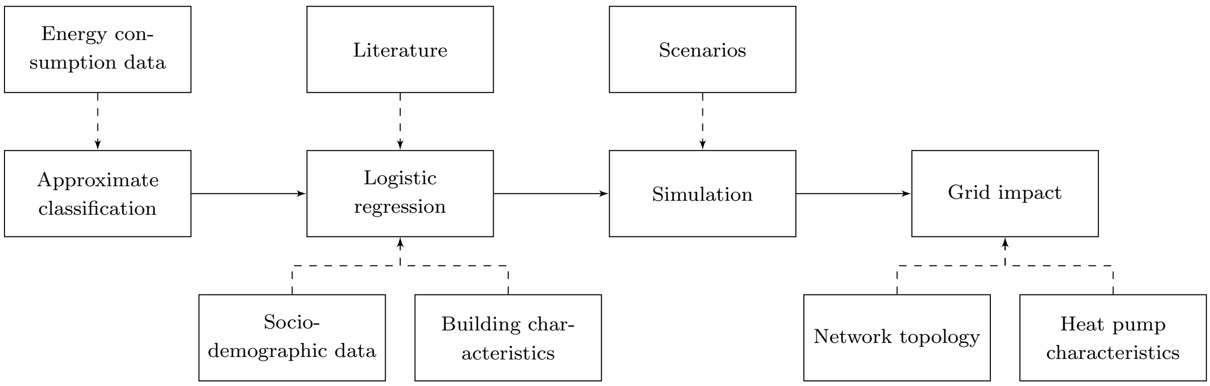


Figure 1.4: Overview of the problem approach, where dashed lines indicate needed input.

1.6 Research design

Given the scope, we address the following main research question.

What is the impact on medium to low voltage transformers due to the spatial diffusion of residential heat pumps?

We define the following research questions:

Research question 1: What is the role of Coteq in the heat transition?

1.1: What are the tasks of a DSO?
1.2: How is the electricity network designed?
1.3: What are the tasks of the asset management department of Coteq?
1.4: What is the heat transition?

Research question 1 increases our understanding of the context. In this stage, we mainly gather secondary data in the form of expert interviews, internal documents, and industry information from websites. Expert interviews are used to gather company-specific information such as network design, role in energy transition, and capacity planning process. Internal documents are an addition to these expert interviews. Industry information published by Netbeheer Nederland (branch organization for DSOs) is used to gather information on DSOs in general and the energy transition.

Research question 2: What literature is available that relates to our main research question?

2.1: What related work is available?
2.2: What are the influencing factors for residential HP installation?
2.3: What methods can be used to determine the influencing factors of residential HP adoption?
2.4: What methods can be used to simulate the electricity network impact for multiple scenarios?

We conduct a literature review to find approaches that are suitable for our main research question. Using literature, we determine the current knowledge on this topic. We base our approach on what is already known and possibly combine this with other methods. In this stage, we keep a broad orientation, meaning that we also consider other sustainable technologies for which the approach might be usable.

Research problem 3: How can the spatial adoption of HPs be modeled?

3.1: What data need to be collected?
3.2: What variables are relevant in determining the likelihood of residential HP adoption?
3.3: How can we infer household HP adoption?
3.4: How can we model the spatial adoption of HPs?
3.5: How can we determine which residential areas are more likely to install HPs?

Research question 3 is focused on modeling the spatial adoption of HPs in residential buildings. This is needed to determine the local impact of HP adoption. In this stage, we need primary data in the form of energy consumption data and secondary data in the form of literature and data from external organizations. Energy consumption data is collected from company databases based on the Centraal Aansluitregister (C-AR). External data is collected from the Centraal Bureau voor de Statistiek (CBS) and the Basisregistratie Adressen en Gebouwen (BAG). These organizations collect data that can be used as variables in our logistic regression model.

Research problem 4: How can we determine future HP growth?

4.1: What are the inputs for determining the future growth of HPs?
4.2: What scenarios can be developed?
4.3: How can we quantify the future growth of HPs?

Research question 4 is focused on quantifying the future growth of HPs. Because there is great uncertainty in what the exact future market share of HPs will be, it is necessary to approach this problem using scenarios. To do this, we collect already available studies on energy transition scenarios and will review its applicability for Coteq.

Research problem 5: How can we determine the impact on the electricity network resulting from the spatial diffusion of HPs?

5.1: How can the expected increase of HPs be expressed in electricity demand?
5.2: How can we determine the additional peak load on the electricity network?
5.3: How can we model the impact on MV/LV transformers?
5.4: How can we assess the effects of uncertainty in input variables?
5.5: What are the consequences of the expected impact resulting from HPs?

In this stage, we translate our findings on the spatial diffusion of HPs for multiple scenarios to electricity network impact. To do this, we mainly collect secondary data on HP power requirements using literature. The data are used in a simulation study that determines the yearly impact on MV/LV transformers. As we expect to have uncertainty in input variables, we additionally perform a sensitivity analysis. Finally, we translate the results to consequences for Coteq.

1.7 Deliverables

This research results in the following deliverables.

- Generalizable forecasting model indicating which residential areas are likely to install HPs.
- Generalizable simulation model indicating the impact on the electricity network resulting from the spatial diffusion of residential HPs.

1.8 Thesis outline

This research is organized as follows. Chapter 2 (Research question 1) discusses the context of this research. Chapter 3 (Research question 2) covers the literature review. Chapter 4 (Research question 3) covers the spatial adoption of HPs. Chapter 5 (research questions 4 and 5) describes the simulation of the impact of future residential HP adoption on MV/LV transformers and the consequences for Coteq. Finally, Chapter 6 covers the conclusion, recommendations, and limitations.

2 Context

This chapter addresses the context of this research (Research question 1) and provides a clear view of the current situation. This chapter is organized as follows. Section 2.1 describes the tasks of DSOs. Section 2.2 introduces the design of the electricity network. Section 2.3 describes the general and Coteq specific asset management process. Section 2.4 discusses the heat transition. Finally, we close this chapter with a conclusion in Section 2.5.

2.1 Distribution System Operators

DSOs are responsible for the construction, operation, and maintenance of the gas and electricity network in the Netherlands. DSOs are active in a regulated market supervised by the Autoriteit Consument & Markt (ACM). The ACM sets maximum rates for the transport of energy. Each of the seven DSOs has its service region. This prevents unnecessary costs resulting from maintaining multiple networks next to each other (Netbeheer Nederland, 2019b). DSOs have two legal tasks: managing the physical network infrastructure and facilitating the functioning of the market. No other commercial activities than managing the gas and electricity network are allowed for DSOs. There is a difference between a national DSO and a regional DSO. A national DSO manages the national high voltage (HV) network, while a regional DSO manages the MV and LV network. Because Coteq is a regional DSO, we mainly focus on the activities of regional DSOs.

2.2 Electricity network design

To understand the design of the electricity network, we first focus on a high-level overview of the electricity network. Next, we describe the design of the electricity network that Coteq manages. Finally, we focus on MV/LV transformer substations in the electricity network managed by Coteq.

Electricity network

The electricity network can be divided into transmission and distribution (Van Oirsouw, 2012). TenneT, the national DSO, manages the transmission network, and the regional DSO manages the distribution network in its service region (Netbeheer Nederland, 2019a). The transmission network consists of a main transport network that provides connections with other countries and a sub transport network that provides connections on the provincial level. Production-units on the transmission network are, for example, power plants and wind farms. The distribution network consists of a regional distribution network that connects large consumers and a local distribution network that connects small consumers. Production units on the distribution network are, for example, wind turbines and PV installations. The functions of the electricity network (transmission/distribution) are related to voltage levels. Higher voltage levels can transport larger electrical power. For this reason, the transmission network is mainly used for transport. Voltage levels are reduced by so-called transformers. The voltage levels eventually transported to the local distribution network are safe and practical for connecting small consumers. Because we focus on households (being small consumers), we are mostly interested in the distribution network.

Distribution network

The distribution network distributes electricity to consumers. This network consists of a MV network and a LV network. Large consumers can directly be connected to the MV network via MV customer substations, while small consumers, connected to the LV network, are connected to the MV network via MV/LV transformer substations. Figure 2.1 shows the topology of a MV network. Electrical power is delivered to the MV network via a high to medium voltage (HV/MV) transformer substation. The electrical power is subsequently distributed to the consumers. There is an additional transformer between the MV network and the LV network to reduce the voltage to a level suitable for households. These transformers are the MV/LV transformers we focus on in this research.

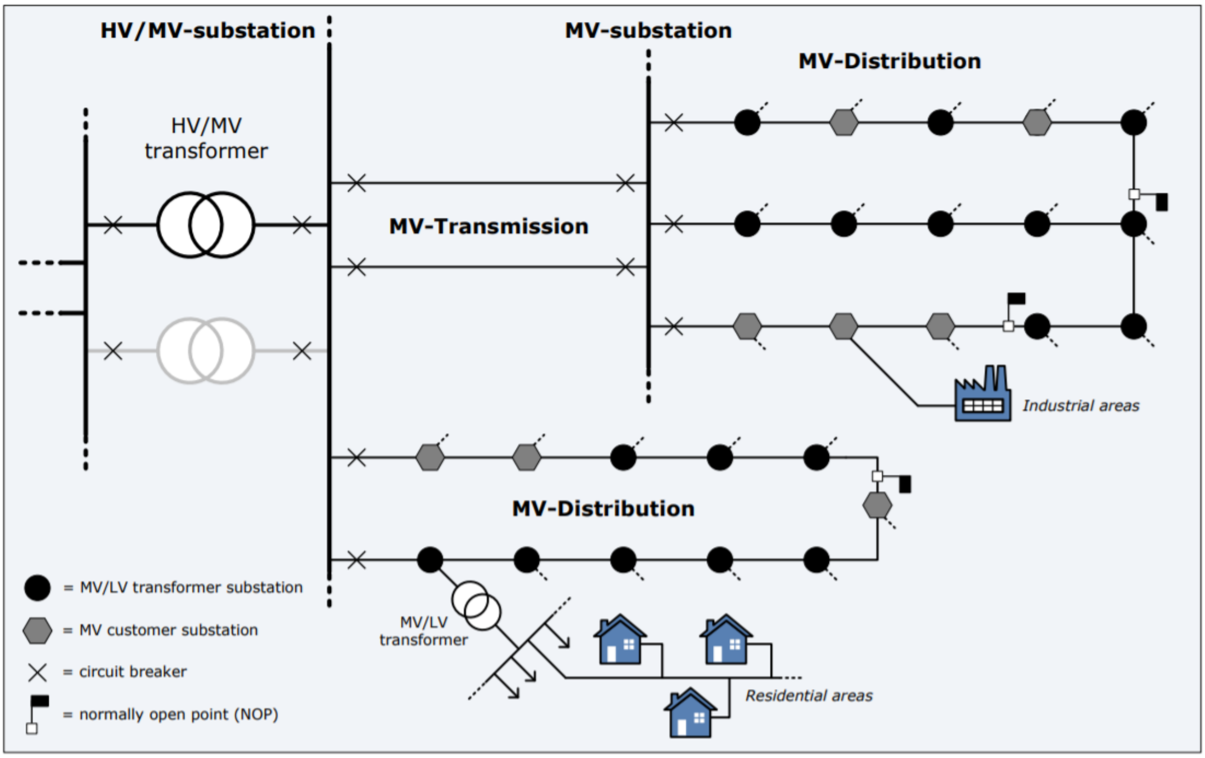


Figure 2.1: Structure of a MV network. Retrieved from: (Grond, 2016).

MV/LV transformer substations

A MV/LV transformer substation consists of three elements: MV installation, MV/LV transformer, and LV rack. The MV installation connects the MV cables. The MV/LV transformer connects the MV network with the LV network and reduces the voltage level. The LV rack connects all outgoing LV cables. A MV/LV transformer substation can contain one, two, or three MV/LV transformers and LV distribution racks. In this case, each MV/LV transformer connects a group of customers connected to the MV/LV transformer substation. Coteq has 403 MV/LV transformer substations that connect households. 390 MV/LV transformer substations contain one transformer, 12 two transformers, and 1 three transformers. On average, a MV/LV transformer substation has 123 household connections. The number of MV/LV transformer substations deployed is dependent on the size of the service region, power density, and uniformity of connections. LV electricity cables have a maximum length that restricts the size of the service region that a MV/LV transformer substation can feed. The size of a service region fed by a

MV/LV transformer substation is also affected by the power density (the density of power requirements in a geographical area), as the capacity of a MV/LV transformer substation is limited. If the power requirement is denser, then the available capacity of a MV/LV transformer substation is already utilized in a smaller service region. Finally, the uniformity of connections is important as each type of connection has its typical electrical load. For this reason, identical connections are often grouped and fed by one MV/LV transformer substation. Figure 2.2 shows the spread of MV/LV transformer substations (shown as dots) and the density (in terms of the number of households per km^2) per residential area in the AGO region. To illustrate the current condition of MV/LV transformer substations, we plot an orange dot if the total capacity if a MV/LV transformer substation, accounting for possibly multiple MV/LV transformers, is utilized for at least 75%. A MV/LV transformer substation feeds on average connections located in 1 to 2 different neighborhoods.

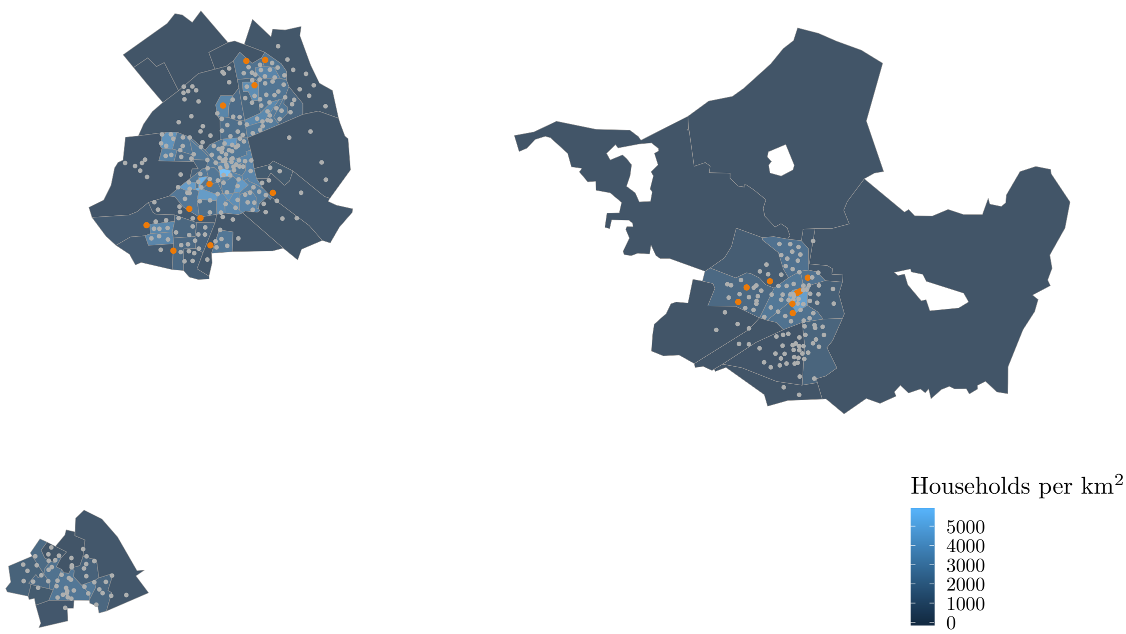


Figure 2.2: Density (households per km^2) per residential area and MV/LV transformer substations connecting households (grey dots $<$ 75% utilized, orange dots $>$ 75% utilized). Left to right: Goor, Almelo, Oldenzaal.

The size and expected power exchange in the service region determines the capacity needed for a MV/LV transformer substation. The power exchange is determined using residential load profiles that indicate the amount of power used by households over time. The sum of all individual loads at one point in time is generally not equal to the aggregated load measured at a MV/LV transformer substation. This is referred to as simultaneity or coincidence (Van Oirsouw, 2012). The coincidence is expressed in a coincidence factor ranging from 0 to 1 and is used to determine the peak load on the network. For example, most households have peaks in electricity demand around the evening when individuals come home from work. Because it is unlikely that the peak load of every household occurs at the same time, the coincidence factor is smaller than 1. The aggregated peak load gets smaller when more households are considered and eventually reaches a constant value. Equation 2.1 shows the relation between individual peak loads and the peak load observed by a MV/LV transformer station. In Equation 2.1, g_n is the coincidence factor for n households, $P_{max,n}$ is the aggregated peak load for n households and $P_{max}(i)$ is the peak load of one household. Knowing the coincidence factor of residential peak loads as a function of the number of households and assuming an equal peak load for all households, enables DSOs to estimate the peak load at a MV/LV transformer substation. This is shown in Equation 2.2. The peak load at a MV/LV transformer substation eventually determines the needed capacity.

$$g_n = \frac{P_{max,n}}{\sum_{i=1}^{n} P_{max}(i)} \tag{2.1}$$

$$p_{max,n} = n \cdot P_{max,1} \cdot g_n \tag{2.2}$$

The capacity of a MV/LV transformer in a MV/LV transformer substation is expressed in kilovolt ampere (kVA). Figure 2.3 gives an overview of the MV/LV transformer capacities used by Coteq. The measure kVA expresses the so-called apparent power. Next to apparent power, there is also the so-called active power. Electrical demand at households is expressed in kilowatt (kW). The power that needs to be transported to households is usually higher because power gets lost during transport. The measure kVA takes this loss into account and is therefore referred to as apparent power, while kW is referred to as active power. A so-called power factor ($\cos\varphi$) quantifies the fraction of kVA that can be used as kW. The capacity currently not used by MV/LV transformers determines the number of sustainable technologies that can be connected to a MV/LV transformer before getting overloaded. Figure 2.4 illustrates the current capacity utilization of MV/LV transformers managed by Coteq. According to Grond (2016), a MV/LV transformer is allowed to be overloaded for a short amount of time. Without introducing the details, a MV/LV transformer only experiences a peak load for a short time. Because of this short time, DSOs often allow a percentage of overloading.

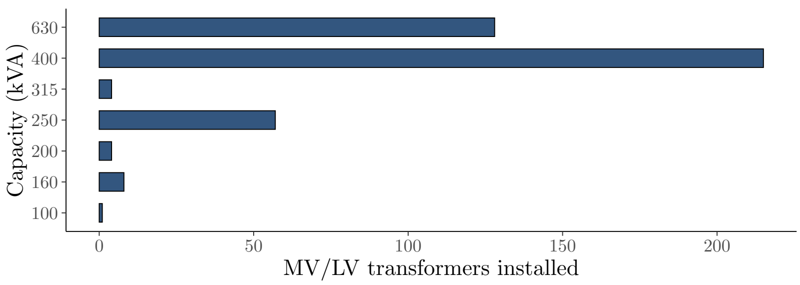


Figure 2.3: Number of MV/LV transformers installed by Coteq grouped by capacity (in terms of kVA).

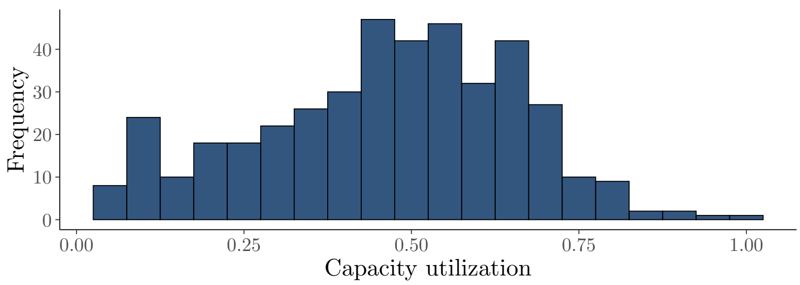


Figure 2.4: Histogram of capacity utilization of MV/LV transformers managed by Coteq.

2.3 Asset management

To understand the relation between asset management and the energy transition, we first focus on the challenges for asset management due to the energy transition. Next, we describe the expansion possibilities for DSOs to increase capacity. Finally, we focus on the capacity planning process of Coteq to identify the fit with our research.

Energy transition challenges

The energy transition introduces new challenges for asset management. To do long term investment planning, planners must have insight into the expected growth of connections and the changing behavior of consumers (Van Oirsouw, 2012). The behavior of consumers is expected to be influenced by the energy transition. Because there are multiple approaches possible to achieve a sustainable energy system in 2050, it is unclear what the future share of sustainable technologies will be. As each sustainable technology has its energy requirements, an often-used approach is scenario analysis. Each scenarios reflects a different outcome of the energy transition in terms of market share of sustainable technologies and is evaluated on possible future bottlenecks. The worst-case scenario is most interesting for DSOs as this scenario determines an upper bound for the required electricity network capacity. Figure 2.5 illustrates the impact of the increasing number of HPs and EVs for a worst-case day in a residential area in terms of additional electrical load. When considering a worst-case day (i.e., maximum peak load due to a cold winter day), no electricity is supplied by a PV installation. As a consequence, all electricity required by households must be distributed by the electricity network. If also EC and EV are used, then the peak load can highly exceed current capacity. Also, EC and EV are expected to be used around the same time, meaning that the coincidence factor is closer to 1. For example, EVs are mostly expected to be charged after work. For this reason, there is a large probability that many EVs are charged at the same time. The same holds for HPs, but then the coincidence is related to the outside temperature (i.e., many HPs will be active when the outside temperature is low).

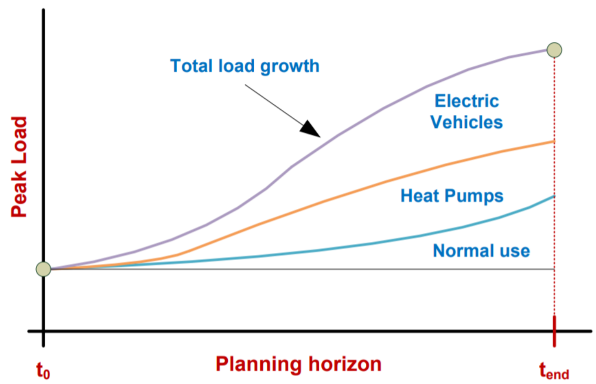


Figure 2.5: Peak load increase caused by the growth of sustainable technologies under worst-case conditions (i.e., winter conditions). Retrieved from: (Grond, 2016).

Expansion options

When plausible scenarios are determined, one can perform capacity planning. To goal is to determine where, how much, which and when new equipment needs to be installed such that future demand can be facilitated (Grond, 2016). Grond (2016) describes several options to increase capacity when considering the MV network. For example, when a MV/LV transformer gets overloaded, one can replace the current MV/LV transformer with one that has more capacity. If the maximum capacity of a MV/LV transformer is reached, then a new point of entry is required to increase capacity. Other options are the replacement or addition of cables in the MV network as shown in Figure 2.6. In Figure 2.6, option A is related to increasing the capacity of a MV/LV transformer, while the numbered options are related to increasing the capacity by replacing or adding cables in specific parts of the network.

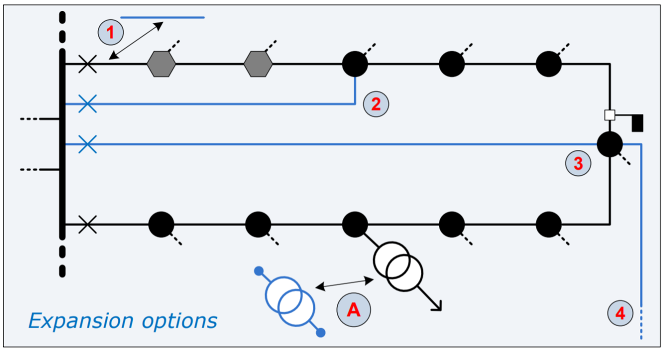


Figure 2.6: Capacity expansion options for a MV network. Retrieved from: (Grond, 2016).

Asset management Coteq

The asset management department of Coteq is responsible for keeping the gas and electricity network safe, reliable, and affordable. Also, Coteq has the task of complying with the expected future functionality of the network. To achieve this, Coteq uses the following stages in the capacity planning process:

1. Assessment and scenario drafting
2. Choice of growth scenario
3. Bottleneck assessment
4. Drafting of capacity planning
5. Approval of plans
6. Final capacity planning

In stage 1, data is used on the development and expansion of the electricity network, realized energy consumption, and company data. These data are used for market exploration to set up future scenarios. Based on historical data and plans from municipalities, the most realistic scenario is chosen in stage 2. Stage 3 uses the scenario to assess possible future bottlenecks by doing so-called network calculations. Network calculations determine the detailed loads on cables and transformers for a certain network configuration. Capacity plans are then drafted to prevent future bottlenecks in stage 4. These plans are approved by the manager of Coteq in stage 5. When approved, this becomes a definitive capacity plan in stage 6. In this research, we focus on stages 1, 2 and partly 3. We partly focus on stage 3 as a complete bottleneck assessment is not possible without knowing the impact of other sustainable technologies.

2.4 Heat transition

In this section, we focus on the heat transition to obtain an understanding of the alternatives for natural gas in the built environment. The heat transition is focused on reducing the use of natural gas in the built environment. Table 2.1 illustrates heating alternatives and how they affect the number of households than can be serviced by a MV/LV transformer (Netbeheer Nederland, 2019c). All-electric alternatives increase electricity demand substantially. As a consequence, fewer households can be connected to a MV/LV transformer. High-/low-temperature heat is the use of residual heat from, for example, industries. This requires no gas connection but increases electricity demand as a HP may be required for heating water. Finally, a hybrid alternative consists of a combination of electricity and gas by using, for example, a hybrid HP. In the following subsections, we describe the available technologies for the alternatives mentioned in Table 2.1 and their developments. First, we describe the HP technology. Next, we describe technologies other than HPs. Finally, we describe the current adoption of heating alternatives and close with an overview of our findings.

Table 2.1: Indication of impact on MV/LV transformers resulting from different heating alternatives.

Alternative	Connection	Households per MV/LV transformer
Current situation	Electricity & gas	400
All electric	Electricity	150
High-temperature heat	Electricity & heat	250
Low-temperature heat	Electricity & heat	200
Hybrid	Electricity & gas	200

Heat pumps

A HP extracts heat from the air, ground or water (Rijksoverheid, 2019b). There are two applications of HPs: full-electric and hybrid. A full-electric HP eliminates natural gas usage whereas a hybrid HP reduces natural gas usage. Full-electric HPs provide heat for space and water. To install a full-electric HP, houses require a high degree of insulation. For this reason, full-electric HPs are mainly installed in new-build houses. Full-electric HPs are installed as water/ground source HP (GSHP) or air source HP (ASHP). GSHPs require more space as construction in the ground is needed to extract heat from ground or water. ASHPs require less space as only a small unit is needed on the outside of the house to extract heat from the air. Hybrid HPs are used in combination with, for example, fossil energy. Most of the time the hybrid HP will provide heat to a household but on colder days the hr-boiler will provide heat. Hybrid HPs mainly use heat extracted from (ventilation) air.

A HP is a sustainable alternative as it is able to efficiently produce heat with little need for electricity. The efficiency of a HP is expressed in a coefficient of performance (COP). A COP of 4 means that 4 kW of heat is produced with 1 kW of electricity (Berenschot, 2017). The energy consumption measured at a DSO does not distinguish between heat needed for space and water. For this reason, we assume one COP where in reality the COP for space and water heating differ. We use an example to form our understanding of the effect of HPs on energy consumption. We consider a household that uses 1,600 m^3 gas for heating and 3,000 kilowatt-hour (kWh) electricity. We use that 1 m^3 gas approximately equals 10 kWh electricity (Berenschot, 2017). If this household would install a full-electric HP with a COP of 4, then 16,000 kWh heat should be produced by the HP to replace 1,600 m^3 gas. Given the COP of 4, the HP only uses 4,000 kWh electricity to produce 16,000 kWh of heat. The new energy consumption would thus become 0 m^3 gas and 7,000 kWh electricity. Considering a hybrid HP, this depends on the fraction of heat delivered by the HP. Berenschot (2017) examined two scenarios in which the hybrid HP delivered 50%/50% and 25%/75% of heat by gas and electricity, respectively. We illustrate this for the 50%/50% example. 50% of the heat is now supplied by the HP, meaning that 8,000 kWh of heat should be produced by the HP. Using a hybrid HP with a COP of 4 this means that the new energy consumption will be 800 m^3 gas and 5,000 kWh electricity.

Non-heat pumps

Beside HPs, there are multiple alternatives for heating with natural gas. Currently, households can, for example, be connected to district heating. District heating is only possible for households if there is a connection to district heating and if sustainable sources are available. An example of such a source is the use of residual heat from the industry. Almelo is the only municipality in the service region of Coteq having district heating (Autoriteit Consument & Markt, 2020). Households not connected to district heating can use other alternatives such as pellet stoves, biomass boilers, and solar boilers. Pellet stoves use pressed wood to generate heat. This technique is used to heat the room in which it is located but can also be connected to radiators and floor heating. Biomass boilers use biomass such as wood to heat water. This water is directed to a radiator or floor heating. Solar boilers use heat extracted from the sun. This heat is used for space and water heating. The solar boiler is the only technology that cannot eliminate the use of natural gas as not enough heat can be generated in winter. The use of hydrogen and green gas are also considered promising solutions. The use of these alternatives can then be used in combination with hybrid HPs to eliminate the use of natural gas.

Current development

There is little data available on the current development of heating alternatives. Natuur & Milieu (2019) studied the development of the most important gas-less techniques. The authors only considered the individual solutions hr-boilers, HPs, pellet stoves, biomass boilers, and solar boilers. Figure 2.7 shows the development of these gas-less techniques. These figures are based on sales data except for the biomass boiler as sales data were not available. Instead, subsidy requests from the Dutch government on biomass boilers were studied. The subsidy for this purpose is called the ISDE (Investeringssubsidie Duurzame Energie). Figure 2.7 shows that the HP is the most popular gas-less technique. According to Natuur & Milieu (2019), approximately half of the HPs were installed in residential buildings. The authors expect that most of these HPs are installed in new-build houses. A further increase in HPs is expected as it is no longer the legal task of DSOs to connect new-build houses to the gas network since the first of July 2018 (Autoriteit Consument & Markt, 2020). Therefore, new-build houses must be connected to district heating or have to install gas-less techniques. Of these gas-less techniques, the HP is the most likely option given the sales figures in Figure 2.7. For existing houses, subsidies can only be requested if the house was constructed before 30-06-2018. As the subsidies for biomass boilers and pellet stoves expired per 01-01-2020, only subsidies for solar boilers and HPs remain (Rijksoverheid voor Ondernemend Nederland, 2020b). Because only HPs can eliminate the use of natural gas, it is expected that this technique will be most important in the heat transition. Another possibility would be the combination of solar boilers and HPs.

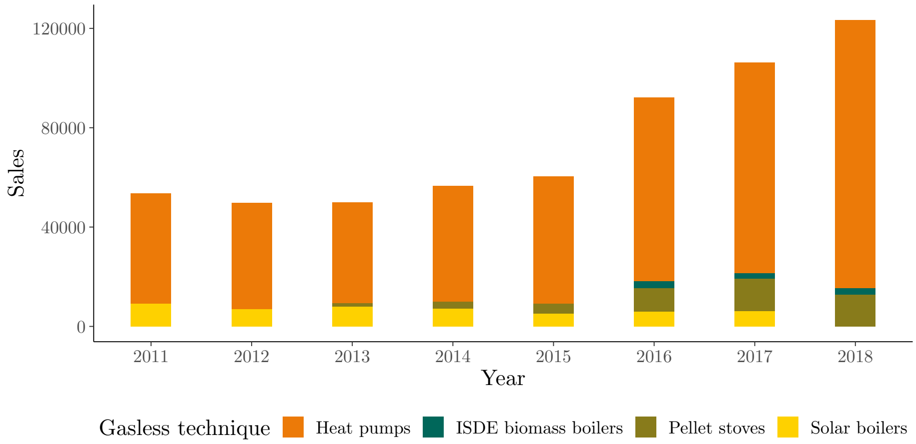


Figure 2.7: Sales of gas-less techniques per year in the Netherlands. Retrieved from: (Natuur & Milieu, 2019).

Multiple studies are focused on current HP development. According to the Dutch Heat Pump Association (2020), approximately 200,000 HPs should be installed in 2020. In 2018, there were in total 140,000 residential HPs according to the Dutch Heat Pump Association (2020). This study used data from the CBS and combined it with data from the Dutch Heat Pump Association (DHPA). Current numbers published by the CBS suggest that 200,000 HPs are already installed as shown in Figure 2.8. Next to national studies, there is also a European study conducted by the European Heat Pump Association (EHPA). The EHPA solely focuses on HPs with a heating function, which was not confirmed by the other studies. The European Heat Pump Association (2020) suggests that in 2016 there were 40,000 air/water HPs and 42,000 ground/water HPs in the Netherlands, whereas the CBS publishes 180,000 installed HPs in 2016. There are thus significant differences in what is expected to be the total number of HPs in the Netherlands. The results from the Dutch Heat Pump Association (2020) and the European Heat Pump Association (2020) show similar results. There is less data available on hybrid HPs. According to Berenschot (2017) approximately 20,000 hybrid HPs were installed in 2016.

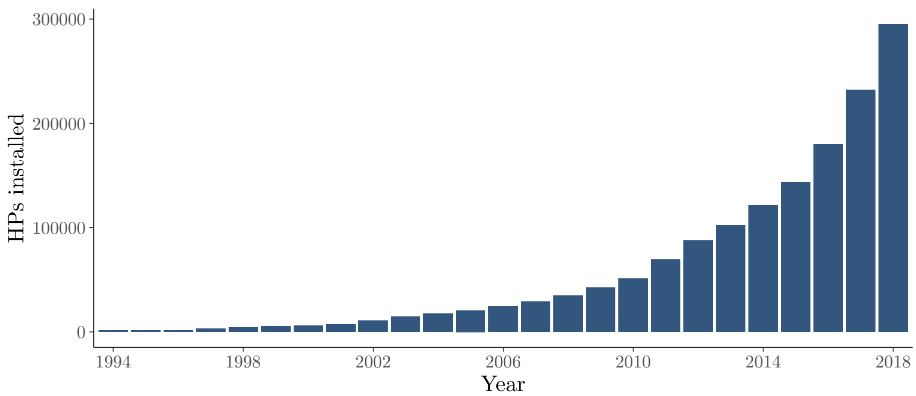


Figure 2.8: Total number of HPs installed in the Netherlands. Retrieved from: (Centraal Bureau voor de Statistiek, 2020).

There is no data on the spatial adoption of HPs in the Netherlands. One might assume that the HPs are evenly spread over the Netherlands using the number of houses per region. Assuming that the HPs are evenly spread, this would result in a fraction of 0.68% of all HPs in this region. To determine the number of HPs, we use data published by the DHPA as this is in concordance with the EHPA. The DHPA also involved installers of HPs which improves the reliability of the data. Given that there were roughly 150,000 HP in 2018 in the Netherlands, we would expect to have 1,000 to 1,500 HPs in the AGO region. Another option is to use data published by the Dutch government on requested subsidies on heating alternatives (Rijksoverheid voor Ondernemend Nederland, 2020a). These data contain the subsidy requests from the start of the ISDE in 2016 up to 2019 and is published on provincial and municipal level. Assuming that HPs were only purchased using the ISDE subsidy, then there would be in total 350 HPs in the AGO region (obtained by scaling the 414 requests in the municipalities Almelo, Oldenzaal and Hof van Twente to the AGO region).

Overview

Table 2.2 gives an overview of all discussed heating alternatives. These heating alternatives are important for understanding plausible future scenarios.

Table 2.2: Overview of introduced alternatives for natural gas.

Heating alternative	Effect on the electricity network
Full-electric HP	Increased electricity demand because a full-electric HP replaces the use of natural gas. The increase in electricity demand depends on the efficiency of the HP.
Hybrid HP	Increased electricity demand but less than full-electric HPs because natural gas is partly replaced. The increase in electricity demand depends on the efficiency of the HP.
District heating	Increased electricity usage because of electrical cooking and possibly a HP for water heating. An additional HP depends on the temperature of the received heat.
Pellet stove	Normal electricity usage as wood is used for heating. A pellet stove can completely replace the use of natural gas.
Biomass boiler	Normal electricity usage as biomass is used for heating. A biomass boiler can completely replace the use of natural gas.
Solar boiler	Normal electricity usage as electricity is generated by the sun to heat water. A solar boiler reduces the need for natural gas.
Hydrogen	Hydrogen replaces the use of natural gas. Because of limited supply, this would still require a heating alternative such as a HP.
Green gas	Green gas replaces or reduces the use of natural gas as it can be blended with natural gas. Because of limited supply, this would still require a heating alternative such as a HP when it fully needs to replace natural gas.

2.5 Conclusion

This chapter described the tasks of DSOs, the design of the electricity network, the asset management process and the development of the heat transition. Coteq manages 403 MV/LV transformer substations with possibly multiple MV/LV transformers. In Chapter 5, we use the data on MV/LV transformer substations to calculate the increase in peak load resulting from HP adoption. To determine the peak load on MV/LV transformers, we have to determine the number of currently installed HPs per residential area. Recent studies have shown that this current number is highly uncertain, especially per residential area. From the context analysis, we know that some residential areas already use district heating. This information can be used in our analysis as it is less likely that HPs are installed in these areas. Also, households connected to district heating might affect the analysis of energy consumption data as it may seem that a full-electric HP is used. Last, to determine the adoption of HPs over space and time, we must know the current development of HPs. Several studies are done on the current national development of HPs. By scaling these data based on, for example, the number of houses, we can estimate the development in the AGO region. Another option is to use the subsidy requests on HPs. These option are important for constructing plausible HP growth curves. In Chapter 3, we mostly focus on the methods available to determine HP adoption, model spatial adoption and simulate electricity network impact.

3 Literature review

This chapter discusses the literature related to our main research question (Research question 2). This chapter is organized as follows. In Section 2.1, we divide our main topic into subtopics and discuss for each subtopic related work. Section 2.2 discusses the literature on determinants for heating system selection including HPs. Section 2.3 focuses on methods that can be used to model the spatial adoption of HPs. Section 2.4 describes possible approaches to simulate future HP installation and the impact on the electricity network. Finally, we close this chapter with a conclusion in Section 2.5.

3.1 Related work

We divide the topic of our research into six subtopics: EV, PV, HP, spatial (diffusion), scenarios and impact. For each subtopic, we discuss related literature. This results in a structured overview from which we can synthesize performed literature. We determine if a study discussed a subtopic as follows. A sustainable technology (EV, PV, or HP) is considered addressed if a study is focused on forecasting the future growth of the sustainable technology. Subtopic spatial diffusion is considered addressed if the study considers spatial diffusion, meaning that differences in technology adoption between geographical areas are considered. Subtopic scenarios is considered addressed if the study discusses multiple future scenarios. Finally, subtopic impact is considered addressed if the study is additionally focused on determining the impact on the electricity network. Table 3.1 shows the results of this analysis. These studies were chosen for two reasons: the studies were recently conducted (within 8 years) and most studies were performed in the Netherlands in cooperation with a Dutch DSO. In the following subsections, we shortly address each of the mentioned subtopics related to the found literature.

Table 3.1: Overview of literature related to the subtopics of this research

Reference	EV	PV	HP	Spatial	Scenarios	Impact
Ahkamiraad and Wang (2018)	✓			✓	✓	
Bernards et al. (2016)		✓		✓	✓	
Bernards, Morren, and Slootweg (2018)	✓	✓	✓	✓	✓	✓
Dong, Sigrin, and Brinkman (2017)		✓		✓	✓	
Eising et al. (2014)	✓			✓	✓	✓
Graziano and Gillingham (2015)			✓	✓	✓	
Saarenpää, Kolehmainen, and Niska (2013)	✓			✓		
Van der Kam, Meelen, van Sark, and Alkemade (2018)		✓		✓	✓	
Van De Sande, Danes, and Dekker (2017)	✓	✓	✓	✓	✓	✓
Veldman et al. (2013)	✓	✓	✓	✓	✓	✓

Electrical vehicles

Ahkamiraad and Wang (2018), Bernards et al. (2018), Eising et al. (2014), Saarenpää et al. (2013), Van De Sande et al. (2017), and Veldman et al. (2013) all focused on forecasting the growth of EVs. All authors made use of the information on EVs that becomes increasingly available. An example is the use of registers that contain information on the number of EVs purchased per geographical area. Each of the studies, except for the study performed by Saarenpää et al. (2013), was focused on the growth of EVs in the Netherlands.

Photovoltaics

Bernards et al. (2016), Bernards et al. (2018), Dong et al. (2017), Van der Kam et al. (2018), Van De Sande et al. (2017), and Veldman et al. (2013) focused on forecasting PV installations. The studies on PV are more detailed than the studies on EV and HP. More detailed data are available such as locations of PV installations and power produced by PV installations per household. All studies are focused on the Netherlands, except for Dong et al. (2017).

Heat pumps

Less research is done on HPs. Bernards et al. (2018), Graziano and Gillingham (2015), Van De Sande et al. (2017), and Veldman et al. (2013) focused on HPs. These authors confirmed the poor data availability on HPs as discussed in Chapter 2. Data collection is not discussed by Van De Sande et al. (2017). None of the authors had data on regional HP installation at its disposal. To determine which residential areas are more likely in adopting HPs, the authors used rules. Veldman et al. (2013) and Graziano and Gillingham (2015) both used a set of rules to indicate if a residential area is likely to adopt HPs. Veldman et al. (2013) used the degree of population density in combination with a distinction between existing and newly developed neighborhoods to determine the likelihood of HP adoption. The authors additionally assumed that ASHPs are mainly installed in existing houses and GSHPs in recently build houses. Graziano and Gillingham (2015) used total population, total floor area, dwelling type distribution (e.g., the fraction of detached houses) and age distribution of dwelling types to form areas. All studies, except for the study performed by Graziano and Gillingham (2015), focused on the Netherlands.

Spatial diffusion

All authors considered the spatial adoption/diffusion of technologies. All studies, except for the studies of Graziano and Gillingham (2015), Veldman et al. (2013), and Saarenpää et al. (2013), used some form of regression to extract important characteristics of technology adoption. The aggregation level differed per study. Bernards et al. (2016) and Bernards et al. (2018) focused on households because detailed data was available on individual PV installations. The studies on HPs showed that no study used a regression model to infer characteristics for HP adoption as no data was available on HPs.

Scenarios

All authors, except for Saarenpää et al. (2013), considered multiple scenarios. Some authors constructed new scenarios and others based their scenarios on other studies. Most studies used time buckets of one year with an end time of 2030, 2040, or 2050. The way the spatial diffusion was included in scenarios differed per study. Eising et al. (2014), Graziano and Gillingham (2015), Dong et al. (2017), and Van der Kam et al. (2018) constructed growth curves based on diffusion models for each area separately. Diffusion models describe how innovations are adopted in a population over time, which often follows an S-curve. Bernards et al. (2016), Bernards et al. (2018), and Veldman et al. (2013) did this for the total area and randomly assigned technologies to households based on the outcome of a regression model.

Impact

Bernards et al. (2018), Eising et al. (2014), Van De Sande et al. (2017), and Veldman et al. (2013) all considered the impact on the electricity network. Eising et al. (2014) only focused on (MV/LV) transformers whereas the other authors also considered cables. Only Veldman et al. (2013) utilized a commercial package to calculate the impact.

Conclusion

From the chosen studies we can draw three conclusions. First, no study had access to data containing the development of HPs per household/residential area. Second, to overcome the limited data availability on HPs, all studies used residential area characteristics to estimate the likelihood of HP adoption. Third, none of the studies used energy consumption data to infer technology adoption. Given these conclusions, we need to introduce methods that can be used to infer HP usage from energy consumption data.

3.2 Factors influencing residential heat pump installation

In this section, we consult literature to determine the factors that influence residential HP installation. First, we determine factors that influence residential heating system selection in general to get an overview of influencing factors. This can be useful as more data is available on heating systems in general than for HPs specifically. Second, we consult literature specific on HPs. The results can be used to determine what data to collect for modeling the spatial adoption of HPs. Also, we can use the results for validation purposes.

Residential heating selection

Karytsas and Theodoropoulou (2014) performed a literature review on the factors that influence residential heating system selection. Table 3.2 gives an overview of the most important factors related to socioeconomic/demographic, residential and spatial characteristics. Also, characteristics related to consumers' behavior, attitudes and preference are shown. We highlighted in italic the factors that were found significant in multiple studies according to Karytsas and Theodoropoulou (2014).

Table 3.2: Overview of factors influencing residential heating system selection. Based on: (Karytsas & Theodoropoulou, 2014).

Category	Factors
Socioeconomic & demographic	*income*, *age*, *educational level*, household size, presence of children, number of children, gender
Residential	size of house, type of residence, construction year, ownership, renovations, improving energy efficiency
Spatial	*region*, *area type (urban/rural)*, *climate*, *presence of green areas*
Behavior, attitude & preference	economic (e.g., investment cost), environmental (e.g., energy efficiency, environmental concerns), energy supply security (e.g., independence), comfort (e.g., functional reliability, air quality), general attitude (e.g., habits), social reasons (social subjective norms), supplier issues (e.g., guarantee).

Heat pumps

Of the studies shown in Table 3.1, none was able to statistically conclude which characteristics are significantly associated with HP adoption. All studies based the most important characteristics for the likelihood of HP installation on surveys or assumptions. In this section, we focus on studies that specifically studied the influencing factors of HP adoption. Table 3.3 gives an overview of the studies found. For comparing the found studies, we utilize the variable categories introduced in Table 3.2.

Table 3.3: Positive (+) and negative (-) significant factors influencing HP installation

Reference	Methodology	Operationalization	Key findings
Ameli and Brandt (2015)	Study of online questionnaires performed in 2011 over 12 countries, including the Netherlands, indicating investment in clean energy technologies. 7,645 participants were used for studying influencing factors of HP adoption. Use of logistic regression model.	Dependent variable: having invested in a HP or not. Participants were considered that could in principal have invested in GSHP. Renters were not included as renters are not allowed to invest.	Age (-) and ownership (+) were found to be significantly associated.
Karytsas and Theodoropoulou (2014)	Study of surveys performed between December 2011 and January 2012 in two municipalities in Attice, Greece with 201 respondents. Use of logistic regression model.	Dependent variable: possibility of considering installing a GSHP.	Age (+), income (-) and education (+) were found to be significantly associated.
Karytsas (2018)	Study of online questionnaires performed between June and October 2012 in Greece with 533 respondents. Use of logistic regression model.	Dependent variable: intention of installing a GSHP.	Income (+) and education (+) were found to be significantly associated.
Karytsas, Polyzou, and Karytsas (2019)	Study of online questionnaires performed between June 2016 and February 2017 in Greece, Spain and Portugal with 400 respondents. Use of logistic regression model.	Dependent variable: willingness to adopt a system combining GSHP, solar thermal panels, and thermal energy storage.	Income (+), education (+), size of living area (+) and urbanity (+/-, dependent on country) were found to be significantly associated.
Michelsen and Madlener (2012)	Study of online questionnaires performed between January 2009 and August 2010 in Germany with 2,985 respondents. Use of logistic regression model.	Dependent variable: GSHP/ASHP adoption.	Age (-), income (+), education (+), construction year (-), size of living area (+) and urbanity (+) were found to be significantly associated.

For the socioeconomic and demographic factors, authors are not fully in agreement. Ameli and Brandt (2015), Karytsas and Theodoropoulou (2014), and Michelsen and Madlener (2012) found different effects of age. Only Karytsas and Theodoropoulou (2014) found that older people are more likely to adopt HPs. A reason for this might be that older people have their own house and larger investment capital. Also, younger people might be less involved in decision making regarding the house. The study of Michelsen and Madlener (2012) suggests that younger people tend to be more open to innovative techniques. Income was overall found to have a positive effect on the likelihood of installing a HP. Only Karytsas and Theodoropoulou (2014) found a negative effect that could be because people with a lower income are more likely to consider cost-saving measures. A reason for the positive relation is the ease with which people can invest in sustainable technologies. All studies found that education had a positive effect on the likelihood of installing a HP. The reason might be that higher educated people are more aware of the long-term benefits.

More agreement is found for the residential and spatial factors. Construction year seems to have a negative effect on the likelihood of installing a HP, as shown by Michelsen and Madlener (2012) and Karytsas et al. (2019). A reason might be that individuals residing in new houses are less favorable in replacing their existing system or that the investment in the new house might not have yet paid back. The same authors agreed on the positive relation of the size of living area with HP adoption. This suggests that HPs are more likely to be installed in larger houses not being apartments for example. Ameli and Brandt (2015) were the only authors who found a positive influence of ownership on HP adoption. Finally, individuals living in urban areas tend to be more likely to install a HP than individuals living in rural areas as shown by Michelsen and Madlener (2012) and Karytsas et al. (2019). It should be noted that the study of Karytsas et al. (2019) shows both negative and positive effects on urban areas depending on country.

Besides the measurable factors listed in Table 3.3, there are important non-measurable factors. Michelsen and Madlener (2012) were the only authors that considered the effectiveness of subsidies. Michelsen and Madlener (2012) found that a subsidy was not an important factor for the decision of installing a HP. The other authors mentioned the importance of policy considerations following from their research, meaning that spatial differences should be taken into account in making policy.

Conclusion

In summary, there are a large number of factors influencing residential heating selection. For HP adoption, we conclude that age, income, educational level, construction year, size of living area, and urban area are the most important influencing factors. It should be noted that the type of HP differed per study as well as the country where the study was performed. The findings for one country might not transfer to another country, meaning that the results are not guaranteed to be applicable to the Netherlands.

3.3 Spatial adoption

This section covers the description of the models than can be used for modeling the spatial adoption of HPs. We focus on relatively simple regression models for a couple of reasons. First, we use classification rules to identify residential HP adoption based on energy consumption data. Although we expect specific patterns to be recognized in energy consumption data, there is still a significant probability of misclassifying households. For this reason, we do not want to use complex models for which there is a large probability of overfitting. Second, from Section 2.4, we know that the number of currently installed HPs is small. Given the small number of currently installed HPs, there is a possibility that neighborhoods will have no HPs. We are not interested in correctly classifying these neighborhoods as not having HPs, but rather we are interested in factors influencing HP installation. Third, Section 3.1 shows that regression models are often used for modeling spatial adoption of technologies.

Linear regression

One possibility to model the spatial adoption of HPs is by using linear regression. We only describe the way it can be applied as we assume that the theory is known. More information on linear regression can, for example, be found in the work of James, Witten, Hastie, and Tibshirani (2013). Linear regression can be used to determine the relation between characteristics of residential areas and the number of sustainable technologies found in these residential areas. There are two possibilities to structure the model. To illustrate this, we assume that we have an independent variable being the number of people with an age between 15-24 and a dependent variable being the number of HPs in a residential area. Using Figure 3.1, based on random number, we expect the characteristics of residential area Dinkelland to be important because a larger number of HPs are found there. We can regress the number of installed HPs in a residential area on the number of people with an age between 15-24 in the residential area. We should keep in mind that looking at the fraction might be more useful as the number of houses per residential area differs. Another method is to regress the fraction of HPs in a residential area on the fraction of people with an age between 15-24 in the residential area. Both methods have the possibility of obtaining a negative response when using the model for prediction. One solution might be to group the outcome from the regression model. Given that the responses are ordered, we can, for example, classify the two residential areas having the largest response as most innovative. Residential areas in a higher class are then more likely to install HPs than other residential areas. Dong et al. (2017), Eising et al. (2014), Van der Kam et al. (2018), and Van De Sande et al. (2017) all use one of the discussed methods where the dependent variable is possibly transformed such that it complies with assumptions needed for regression models.

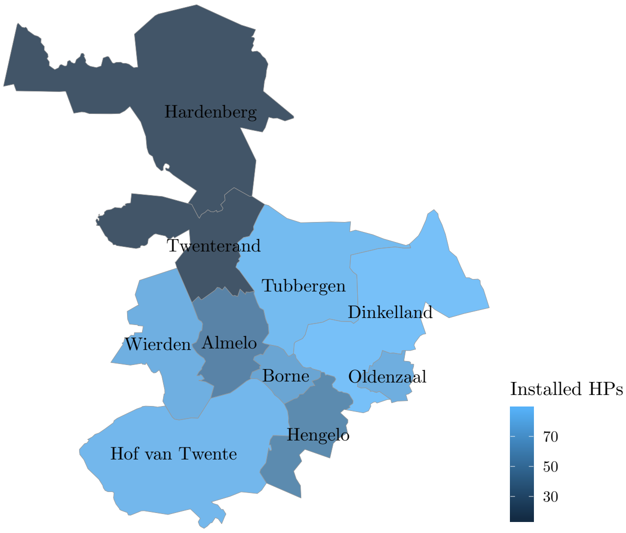


Figure 3.1: Illustration of how linear regression can be used in this research. The number of installed HPs in each residential area functions as a dependent variable to which a linear regression model can be fitted.

Logistic regression

Logistic regression might be a more appropriate method since the dependent variable is forced to [0, 1]. Logistic regression uses the logistic function shown in Equation 3.1. This equation generates an S-shaped curve whereas linear regression would generate a straight line as illustrated in Figure 3.2. Logistic regression directly models the so-called posterior probability from Bayes' theorem shown in Equation 3.2 (Hastie, Tibshirani, & Friedman, 2009). In this equation, π_k is the prior probability, which indicates the fraction of observations in a certain class. $f_k(x)$ indicates the density function of X for an observation that comes from the kth class. The outcome is then the posterior probability, or in other words, the probability that the observation with $X = x$ belongs to the kth class. If one would know the distribution of the variables it would be possible to apply Bayes' theorem. For example, if the variables would be approximately normally distributed, then one can apply linear discriminant analysis (LDA) (James et al., 2013).

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \tag{3.1}$$

$$Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)} \tag{3.2}$$

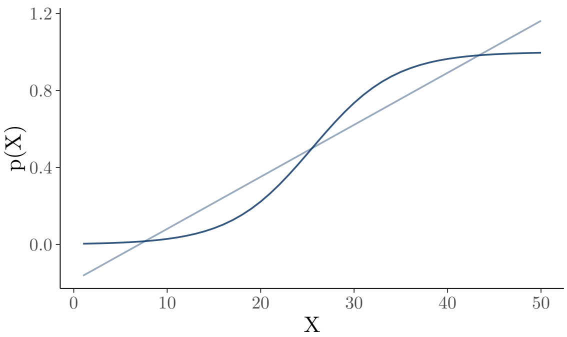


Figure 3.2: Illustration of difference in prediction using linear and logistic regression.

From the logistic regression model, the quantity $p(X)/[1 - p(X)]$ is called the odds. The odds is a value in the range $[0, \infty]$. A value < 1 indicates a probability smaller than 0.5 whereas a value > 1 indicates a probability larger than 0.5. The coefficients are often expressed in odds as these are easier to understand. An increase in X then either increases or decreases the odds. When fitting a logistic regression model, we try to find the coefficients expressed as β shown in Equation 3.1. This is usually done using maximum likelihood. By taking the derivative of the log-likelihood, equating it to zero and solving it, we obtain the coefficients of the logistic regression model. For each X we can then determine the probability of belonging to a certain class and which coefficients are most important.

We can use this method as follows. A household can either have a HP installed ($Y = 1$) or not ($Y = 0$). We can then use the logistic function to obtain the likelihood for a household of having a HP. First, we list all households in the service region of Coteq. Next, for each residential area, we assign the installed number of HPs randomly to households located in the residential area. It does not matter which households get assigned a HP because all households are assumed to have identical characteristics. Last, using logistic regression, we obtain the likelihood of having a HP per household which is identical for households in the same residential area.

Model selection

Model selection is done to improve prediction accuracy and model interpretability (James et al., 2013). When the number of events is small, there might be too much variance in the prediction. Austin and

Steyerberg (2017) suggests that having 20 events per variable is sufficient to obtain a performance estimate that is comparable to an out of sample estimate. Including only relevant variables reduces this variance and also increases model interpretability. There are several methods available to achieve this. We only discuss subset selection and shrinkage as these are easy to interpret. Subset selection tries to find the best model by adding or removing variables in a few steps. Each model is compared using a performance measure. Shrinkage uses a penalty to reduce the number of included variables. Coefficients that introduce much variance opposed to other coefficients are shrunk such that a model remains with only the most relevant variables.

Conclusion

Linear regression suffers from possibly negative predictions but the results can still be used as the predictions are ordered. Logistic regression does not suffer from negative predictions of penetration levels and more intuitively reflects the likelihood of HP adoption for a household. Logistic regression has also been proven to be useful in this context as shown by Bernards et al. (2018).

3.4 Simulating electricity network impact

This section discusses the methods that can be used to simulate the impact on the electricity network. First, we discuss the methods that can be used to model the diffusion of HPs. Second, we discuss the methods that can be used to combine the growth scenarios with the methods to model spatial adoption from Section 3.3. Third, we describe methods that can be used to determine the impact on MV/LV transformers.

Growth patterns

Growth patterns of technologies are often based on a theory called the *Diffusion of innovations* (Rogers, 1983). Diffusion is the process by which an innovation is communicated through certain channels over time among the members of a social system. The communication of these innovations is achieved by, for example, media or face-to-face conversations. As time progresses, more people adopt the innovation which often occurs in an S-shaped curve. Rogers (1983) defined five categories: innovators, early-adopters, early-majorities, late-majorities, and laggards. These groups indicate the types of individuals that can be identified based on the speed of technology adoption. These groups may be classified according to the time of technology adoption. Rogers (1983) describes that this is a tedious task and that individuals could also be classified according to their degree of innovativeness. Innovativeness can then be determined based on socioeconomic status, personality variables, or communication behavior. Multiple models can model this diffusion. We discuss the two most often used models in our selection of related work: the Fisher-Pry model and the Bass model.

Fisher-Pry model

The Fisher-Pry model is a substitution model for technological change, meaning that this model is based on the assumption that one technology replaces another (Fisher & Pry, 1971). The rate at which a new technology replaces an old one is dependent on the remaining number of old technologies in a population. Equation 3.3 illustrates the mathematical notation of the model. In Equation 3.3, $f(t)$ is the fraction of potential market substituted by the new technology in year t, α is half annual fractional growth (i.e., rate of substitution determining the steepness of the S-curve) and t_h is the time at which diffusion has reached 50 percent. Multiplying $f(t)$ by the market size gives the expected growth in a year. Figure 3.3 shows an application of the Fisher-Pry model for EVs in the service region of DSO Liander. Eising et al. (2014) and Ahkamiraad and Wang (2018) both used the Fisher-Pry model.

$$f(t) = \frac{1 + tanh[\alpha(t - t_h)]}{2} \tag{3.3}$$

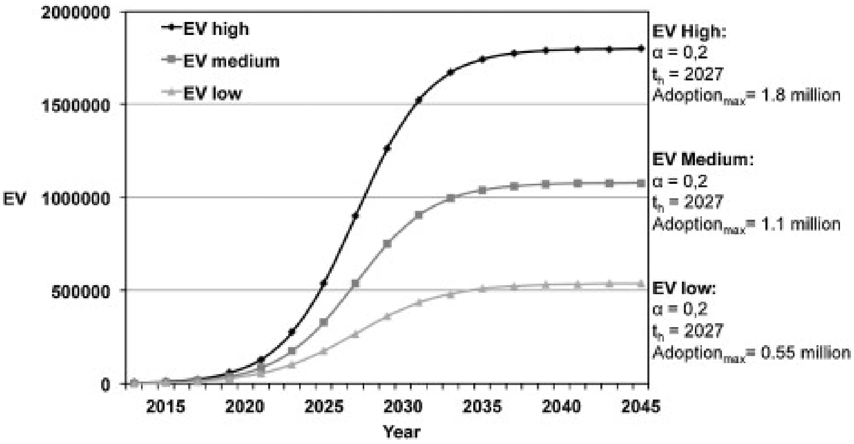


Figure 3.3: Illustration of a Fisher-Pry model for EV used by DSO Liander. Retrieved from: (Eising et al., 2014).

Bass model

The Bass model of diffusion is based on innovators and imitators (Bass, 1969). Innovators decide to adopt new technology independently whereas imitators are influenced by the decisions of others. Equation 3.4 shows the mathematical notation, where $f(t)$ is the change of installed base fraction and $F(t)$ is the installed base fraction. The parameters p and q are the coefficients of innovators and imitators, respectively. The cumulative number of adopters is then given by Equation 3.5. Dong et al. (2017) and Van der Kam et al. (2018) used the Bass model. A reason for using the Bass model was its widely tested validity and the usefulness for predicting clean energy technology growth (Van der Kam et al., 2018). The Bass model is however criticized as it uses some simplifying assumptions such as population homogeneity whereas people are heterogeneous (Kiesling, Günther, Stummer, & Wakolbinger, 2012). Kiesling et al. (2012) suggest the use of agent-based modeling that uses a set of decision rules related to the decision of technology adoption to overcome these assumptions. However, this is out of scope as peer effects are not considered.

$$\frac{f(t)}{1 - F(t)} = p + qF(t) \tag{3.4}$$

$$A(t) = m\frac{1 - \exp^{-(p+q)(t-t_0)}}{1 + \frac{p}{q}\exp^{-(p+q)(t-t_0)}} \tag{3.5}$$

Spatial adoption patterns

There are mainly two methods used in literature for modeling spatial diffusion. The first method uses multiple S-curve by constructing an S-curve for each geographical area. The second method uses one S-curve and statistically assigns HPs to residential areas/households. We discuss these methods in the following paragraphs. First, we discuss the method that uses multiple S-curves. Next, we introduce the method that uses one S-curve. The description that uses one S-curve is first explained assuming an equal spread of technology adoption. Next, we explain how diversity is included. Table 3.4 is used to support our explanation for the second method.

The first method constructs multiple S-curves. This is done in various ways. One way is to construct S-curves based on the results from a regression model. The output of the regression model determines the class of a geographical area. These classes are based on the *Diffusion of innovations* (Rogers, 1983). The predicted values, which are based on important characteristics for technology adoption, are assumed to describe how innovative a geographical area is. By classifying each area on innovativeness, one can determine area-specific parameters for an S-curved model such that the growth is area specific. Another way to construct multiple S-curve is to use already available data to fit S-curves to the growth of each residential area. Fitting is done either on just the available data or by additionally assuming a final market share such that the S-curve complies with a predetermined scenario. A drawback is that much data is required to fit multiple S-curves.

The second method, used by Bernards et al. (2016) and Bernards et al. (2016), is more complicated and uses individual (per household) predicted probabilities to achieve spatial diffusion. This method works as follows. A national S-shaped growth curve determines how many technology adoptions are expected in a year. If all households have an equal probability of technology adoption, then the probability of technology adoption p_t for all households is given by the growth in year t divided by the number of households not having adopted the technology. To assign the growth in year t, each household n is assigned a random number from the uniform distribution $U(0, 1)$. If the random number for household n is smaller than p_t, then a technology is assigned to household n. On average, the total number of technology adoptions now complies with the growth in year t. For example, if five households have not adopted the technology and two technology adoptions are expected, then each household has a 0.4 probability of technology adoption. If five random numbers are drawn (one for each household), then on average two are expected to be smaller than or equal to 0.4. This is illustrated in Table 3.4a.

In the second method, a logistic regression model is used to include the spatial adoption of a technology. The logistic regression model predicts per household the likelihood of technology adoption p_n (see Section 3.3). The predicted values are used to construct per household a scaling factor. This is done by dividing each predicted value p_n by the mean of the predicted values of all households $p_{n,mean}$. This scaling factor represents per household how far the predicted value is off from the mean of the predicted values. The scaling factor is multiplied by p_t using Equation 3.6 and results in $p_{n,t}$. This increases or decreases the likelihood of technology adoption for each household. Because the average of the scaling factors is equal to one, the average of $p_{n,t}$ over n is still equal to p_t. A technology is now assigned to households using the same procedure of random numbers. This is illustrated in Table 3.4b.

Table 3.4: Example of modeling the spatial diffusion of a technology using the method introduced by Bernards et al. (2016).

(a) Equal spread

Connection	p_t	$U(0,1)$	Adoption
a	0.4	0.12	Yes
b	0.4	0.71	No
c	0.4	0.39	Yes
d	0.4	0.89	No
e	0.4	0.52	No

(b) Diversified spread

Connection	p_t	p_n	Scaling factor	$p_{n,t}$	$U(0,1)$	Adoption
a	0.4	0.06	1.5	0.6	0.12	Yes
b	0.4	0.02	0.5	0.2	0.71	No
c	0.4	0.02	0.5	0.2	0.39	No
d	0.4	0.04	1.0	0.4	0.89	No
e	0.4	0.06	1.5	0.6	0.52	Yes

$$p_{n,t} = \frac{p_n}{p_{n,mean}} p_t \tag{3.6}$$

Bernards et al. (2018) evaluated the performance of the second method illustrated in Table 3.4 by considering the diversified adoption and the equally spread adoption using, among others, the Mean Average Percentage Error (MAPE). The MAPE measures the overall forecast error. The growth of PV per neighborhood between 2006 and 2016 was used for this analysis. The authors found that the MAPE was reduced from 103.42 to 51.61 indicating that their approach improved the long term forecast. We note that this approach might introduce bias. The approach makes it possible that the probability $p_{n,t}$ is outside the range [0, 1]. Imagine that only 100 households are still left to adopt a HP and the growth is expected to be 80, then p_t would be 0.8. The scaling factor would now need an upper bound of 1/0.8 to restrict $p_{n,t}$ to be within [0, 1]. To the best of our knowledge, this is not considered by Bernards et al. (2018) and could create bias when p_j is close to 1.

Grid impact

Determining grid impact can be done in various ways. Grond (2016) mentions that electricity networks are designed based on peak loads. The level of detail to determine these peak loads differs. One possibility is to only consider the peak load resulting from a technology based on the capacity of the technology. For example, Eising et al. (2014) assumed that charging an EV requires a certain capacity in terms of kVA. Also, the authors assumed a coincidence factor as discussed in Section 2.3. By multiplying the coincidence factor with the capacity needed for a certain number of expected EVs, the authors were able to determine the additional impact on a transformer. Using the already known peak load, one can determine the total load resulting from the number of households having adopted EV. A more detailed approach is to determine the so-called residential load profiles used by Bernards et al. (2018), Van De Sande et al. (2017), and Veldman et al. (2013). A load profile indicates the electricity demand pattern of a household over a day in kW. Figure 3.4 illustrates this for a household for different seasons. A load profile of HPs can then be constructed from empirical data based on a sample of households using a HP.

This load profile can then be added to the residential load profile to find the increased demand pattern for a household. This approach is more exact but requires empirical data retrieved from households that have installed a HP.

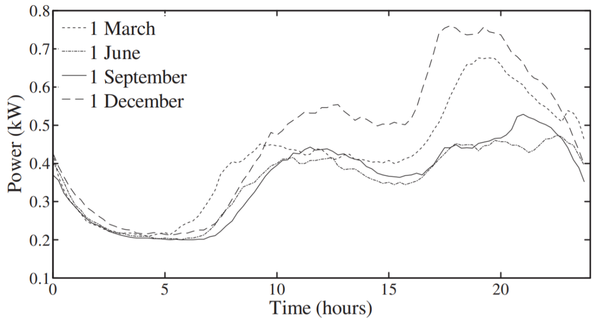


Figure 3.4: Daily load profile of a household in different seasons. Retrieved from: (Veldman et al., 2013).

Conclusion

For modeling the growth of a technology, authors use either the Fisher-Pry or the Bass model. The Bass model is widely validated and has shown to be useful for predicting the growth of clean energy technologies (Van der Kam et al., 2018). The Fisher-Pry model has also shown to be useful in this context but is used less often. To model the spatial diffusion of a technology, authors either construct an S-curve for each residential area or construct one S-curve and statistically assign growth to residential areas/households. Fitting multiple S-curves requires data on the development of a technology over space and time which is not available for HPs. Constructing multiple S-curve would also require the assumption that each residential area is considered a social system. Residential areas within the AGO region might be too small for this assumption. Finally, the approach introduced by Bernards et al. (2016) in Equation 3.6 might introduce bias. Given that multiple S-curves are not feasible and the used statistical method introduces bias, we need to find another statistical method to statistically assign HPs over the AGO region using one S-curve.

3.5 Conclusion

Multiple approaches for solving the problem at hand are discussed in the selected literature. Literature showed that all approaches to model the spatial diffusion of HPs are based on assumptions regarding the likelihood of HP adoption for a residential area. Approaches using energy consumption data are not discussed. To use energy consumption data for this purpose, we need to find methods to infer HP usage. Having obtained a sample of households using HPs, we can use multiple statistical models to construct a prediction model. Two approaches are used in the selected literature, namely: linear and logistic regression. Logistic regression does not suffer from negative predictions and better resembles the likelihood of HP adoption. To use the results of, for example, a logistic regression model, we can construct one or multiple S-curve(s) based on the *Diffusion of innovations*. Multiple S-curves are not feasible because of limited data availability on HP development and the small size of residential areas. A single S-curve is feasible but requires a statistical approach to assign HPs to residential areas/households. The proposed statistical approach suffers from possible bias and thus new approaches seem appropriate to be researched. In Chapter 4, we discuss Research question 3, namely: spatial adoption of HPs in the AGO region.

4 Spatial adoption

In this chapter, we model the spatial adoption of HPs in the service region of Coteq (Research question 3). Modeling the spatial adoption of HPs can help Coteq in identifying future local requirements of the electricity network. This way, Coteq can focus on areas in which HPs are expected to have a greater impact. This chapter is organized as follows. In Section 4.1, we discuss data collection and data pre-processing. Section 4.2 introduces the method that is used to detect abrupt changes in energy consumption that correspond to HP usage. Also, we discuss the application and the results of applying the method. In Section 4.3, we introduce logistic regression to model the spatial adoption of HPs in the AGO regions. Section 4.4 describes the validation of the results. Finally, we close with a conclusion in Section 4.5.

4.1 Data collection

In this research, we use socio-demographic data and data on buildings and addresses. These data are used as independent variables in the logistic regression model. Socio-demographic data is obtained from the CBS and data on buildings and addresses from the BAG. As discussed in Chapter 1, we focus on residential areas. The CBS publishes socio-demographic data in different residential area structures. The AGO region consists of 3 municipalities, 28 districts and 94 neighborhoods, where a neighborhood is part of a district and a district part of a municipality. In this research, we use the neighborhood structure because we can more detailed determine local requirements. A neighborhood is defined as an area that is demarcated as homogeneous because of the socio-economic structure. More information on residential area structures can be found in Appendix B. Next to data from the CBS, we collect data from the BAG. The BAG publishes data per building or address. Therefore, we group the data from the BAG such that it complies with the chosen neighborhood structure from the CBS. Data on connections, such as energy consumption, is obtained from the C-AR. The C-AR is a database that contains information on gas and electricity connections in the Netherlands. From this database, we obtain the normalized annual consumption (NAC), which is the annual energy consumption accounted for seasonal differences. These three data sources have some inconsistencies and need to be pre-processed before being used in the analysis. We pre-process the data according to the approach given by Kuhn and Johnson (2013). First, we pre-select variables based on literature. Second, we evaluate the variables based on completeness and variance. Finally, we evaluate the (multi-)correlation between the variables. The data pre-processing steps are described in Appendix C. Table 4.1 shows the resulting set of variables.

Table 4.1: Description of the variables used for the logistic regression model.

Variable	Description
PercMen	Percentage of men
Perc1524	Percentage of people in the age category 15-24
Perc2544	Percentage of people in the age category 25-44
Perc4564	Percentage of people in the age category 45-64
PercHouseholdNoChildren	Percentage of households without children consisting of married/unmarried couples and other households
PopulationDensity	Population density expressed in the number of inhabitants per km^2
PropertyValue	Property value for objects described as dwellings for main residence and dwellings with practice rooms
PercMultipleFamily	Percentage of multiple-family dwellings that share with multiple dwellings one building
DegreeOfUrbanity	Degree of urbanity ranging from 1 (>2500 addresses per km^2) to 5 (<500 addresses per km^2)
SizeLivingArea	Median of size of living area of dwellings per neighborhood
ConstructionYear	Median of construction year of dwellings per neighborhood

4.2 Change point detection

In the selected studies in Section 3.1, we found that no approaches are currently used to infer HP usage in households. In this section, we introduce a method that can detect change points in time series and can possibly capture the change in energy consumption resulting from HP usage. First, we introduce the method. Second, we discuss how it is applied. Last, we apply the method and discuss the results.

Method

As discussed in Section 2.4, the use of a HP increases electricity consumption and decreases/eliminates gas consumption. If this change is large enough and if the variability is limited, then this change can be detected by a method called change point detection. Figure 4.1 illustrates this. We assume that a household only experiences one change point. Because change point detection is not applied before in this context, we limit ourselves to a relatively simple method using statistical evidence given by Ross (2015). We focus on finding the point in time in which the difference in the means before and after the change point is maximum and significant. The method using statistical evidence works as follows. We have a fixed-length sequence of the form $x_1, \ldots, x_n$. A change point is then detected using hypothesis testing. The null hypothesis H_0 is that no change point exists in the sequence and that the observations are identically distributed according to some distribution F_0. The alternative hypothesis H_1 is that a change point does exist and that the observations before and after the change point are identically distributed and follow some distribution F_0 and F_1, where $F_0 \neq F_1$. This gives the two hypotheses shown in equations 4.1 and 4.2. The problem is solved by calculating a test statistic, which is dependent on the underlying assumption of the data used (e.g., normally distributed). For each possible change point this test statistic is calculated. The best estimate of the change point follows from finding the maximum value in the distribution of test statistics resulting from all possible change point locations.

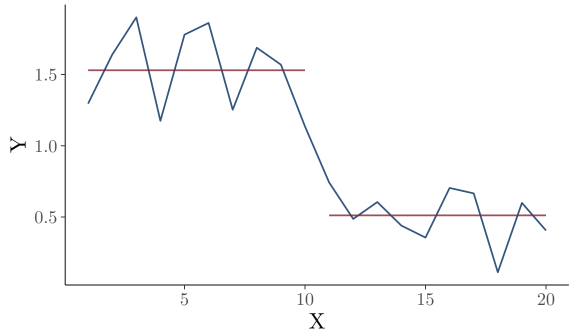


Figure 4.1: Illustration of finding a significant change point that maximizes the difference in the means before and after a change point.

$$H_0 : X_i \sim F_0(x; \theta_0), i = 1, \ldots, k \tag{4.1}$$

$$H_1 : X_i \sim \begin{cases} F_0(x; \theta_0), i = 1, 2, \ldots, k \\ F_1(x; \theta_1), i = k+1, k+2, \ldots, n \end{cases} \tag{4.2}$$

Application

The change point detection method introduced is only applied for hybrid HPs. Full-electric HP adoption is determined by analyzing the activity of gas connections. A gas connection removal indicates, for example, that a household has taken energy saving measures. To extract HP usage we use the following classification rules:

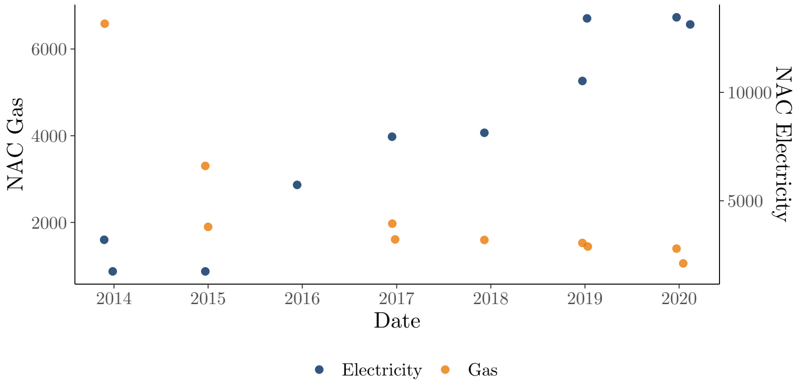
1. If a household has an active electricity meter reading, requested a gas connection removal and is not connected to district heating, then we assume that this household has adopted a full electric HP.
2. If a household has at least a 50 percent gas reduction, an electricity increase according to a hybrid HP having a COP between 3 and 5 (accounting for possible self-sufficiency due to PV) and is not connected to district heating, then we assume that this household has adopted a hybrid HP.

Classification rule 1 is based on the assumption that a household needs heating. Since households connected to district heating are not considered in this analysis, these households are likely to have installed HPs based on the context analysis in Section 2.4. Rule 2 is based on the findings of Berenschot (2017). Berenschot (2017) described the expected change in energy consumption as a result of the installation of hybrid HPs. We take into account that some households have installed PV installations, meaning that a percentage of the total electricity demand of a household can be supplied by electricity generated by PV installations (referred to as self-sufficiency). We assume that the percentage of self-sufficiency is between 25%-35% for households without a HP and between 20%-30% for households with a HP based on the findings of Obinna, Joore, Wauben, and Reinders (2017, p. 11) and Litjens (2018, p. 132). Self-sufficiency is lower for households having a HP as more electricity is needed in winter that a PV installation cannot provide. To reduce misclassification, we only consider households for which holds that the change point in gas and electricity occurred at most one year apart. This reflects the situation that changes in gas and electricity consumption are expected to occur around the same time. Finally, we use the Mann-Whitney test statistic. The Mann-Whitney test statistic is used for testing differences in means without assuming normality. Classifying households according to these classification rules has the following limitations:

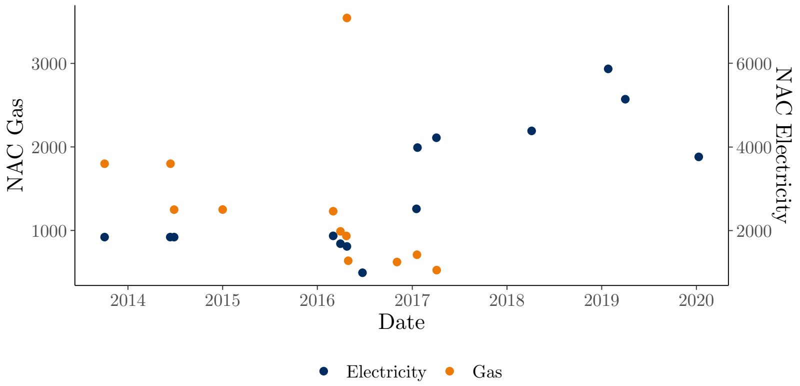
- Households that have installed pellet stoves, biomass boilers, or solar boilers might be detected. We assume that these households are environmentally oriented and will eventually make the transition towards HPs, which makes them usable for this analysis.
- Households that have installed PV installations can distort the approximate classification as Coteq does not know how much electricity produced by a PV installation is used for self-sufficiency. Although we made an assumption on the percentage of self-sufficiency, this still decreases the reliability of the approach.
- To detect a change point, we need sufficient measurements before and after the change point. For this reason, most recently installed HPs might not be detected in the current data set.
- There is a possibility that a house is inhabited for a while, which creates an additional change point that cannot be detected and can distort the analysis.

Results

Applying the classification rules gives the following results. 93 connections have requested a gas connection removal while having an active electricity connection. In addition, we applied change point analysis resulting in only 6 connections of which there were 2 households having a PV installation. Combining the results amounts to a total number of 99 connections, while Coteq has approximately 50,000 connections in the AGO region. In Section 4.4, we discuss what the consequences are of this small number of found households. Figure 4.2 shows two examples of the energy consumption data of classified households. Figure 4.2a shows for a household classified as having installed a hybrid HP the energy consumption pattern. Figure 4.2a additionally shows the energy consumption pattern for a household having requested a gas connection removal. Figures 4.2a and 4.2b are both examples that the analysis produces the desired results. It should be noted that households that were classified as having a full-electric HP could not be verified with change point detection. Most households classified as having a full-electric HP did not have enough measurements to detect a change point. In total, there are 35 neighborhoods having an approximate classification of a HP. Given that there are 94 neighborhoods in the AGO region, we thus have 59 neighborhoods not having HPs. Figure 4.3 shows the penetration level per neighborhood based on the connections found from the approximate classification. The outcome clearly shows hot spots. These hot spots have a greater effect in the following analysis and thus the results must be carefully analyzed. These hot spots might be related to autonomous growth, but might also be related to housing corporations that renovated a block of houses.



(a) Hybrid HP



(b) Full-electric HP

Figure 4.2: Energy consumption behavior of classified households.

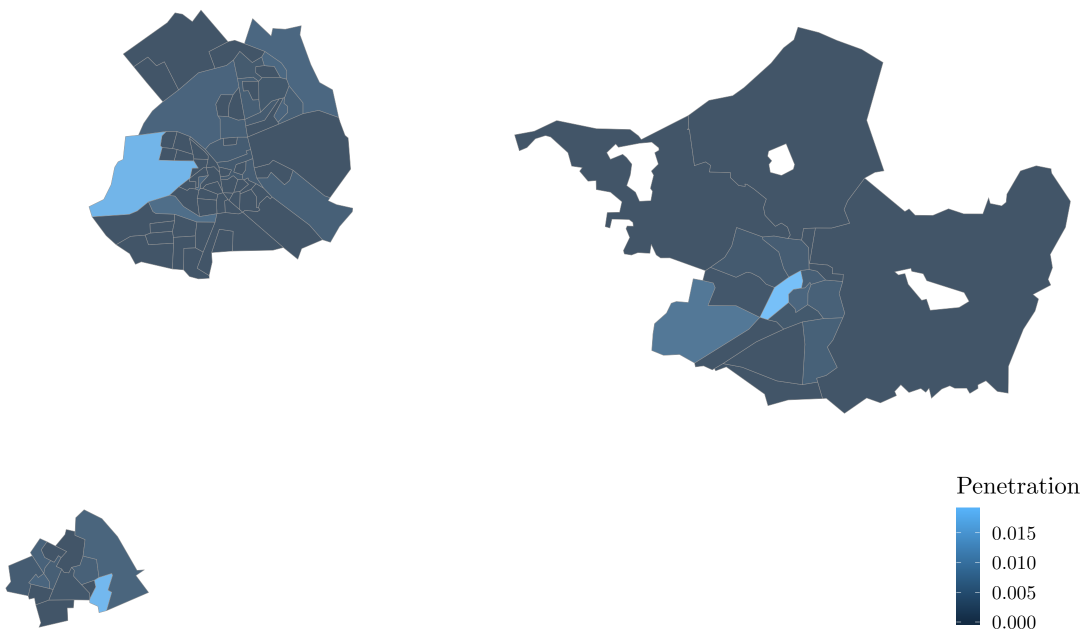


Figure 4.3: Heat map of HP penetration level per neighborhood. Left to right: Goor, Almelo, Oldenzaal.

4.3 Logistic regression

We choose to use a logistic regression model because it models a probability and thus remains in the range [0, 1]. This probability resembles the likelihood of HP adoption. Related work discussed in Section 3.1 has shown that logistic regression is useful for modeling the spatial adoption of a technology. In the logistic regression model, we use the variables listed in Table 4.1 as independent variables. We apply the logistic regression model as follows. In Section 4.2 we found 99 households that correspond to the introduced classification rules for HP usage. We do not distinguish between hybrid and full-electric HPs as the number of hybrid HPs found is too small. As discussed in Section 3.3, we need observations with either a 0 (not having a HP) or 1 (having a HP). Since we know the estimated number of HPs per neighborhood, we can randomly assign a 1 to a number of households indicating HP installation. The exact locations do not matter as we assume that each household in a neighborhood has identical characteristics. We note that the total number of households is roughly 50,000, while the classified number of HPs is 99. The ratio of events and observations is extremely low and thus the outcome of the logistic regression model is possibly unstable (i.e., large confidence intervals for the estimated coefficients). We first discuss simple logistic regression, meaning that we consider each variable individually. Next, we apply a multiple logistic regression model including all variables. Next, we apply a stepwise logistic regression model to increase model stability and improve interpretability. Finally, we discuss the results.

Simple logistic regression

First, we apply a simple logistic regression model to each variable separately. Figure 4.4 visualizes the results and shows the fitted line including the confidence interval. Often one would plot a graph with on the x-axis the value of the variable and on the y-axis the occurrence of an event being either 0 or 1 as illustrated in Figure 3.2. Since every household in a neighborhood has identical characteristics, we can also plot the fraction of installed HPs per neighborhood against the fitted logistic regression line. If a perfect fit would be found, then this fit would exactly describe the current fraction of HPs per neighborhood. The variable *DegreeOfUrbanity* is not plotted since this is a categorical variable. Individually significant variables with p-value less than 0.10 are the continuous variables *Perc1524*, *Perc4564*, *PopulationDensity*, *PropertyValue*, and *ConstructionYear* and the categorical variable *DegreeOfUrbanity* for values 2 and 4. The variables *Perc1524* and *PopulationDensity* have a negative effect on the likelihood of HP adoption, while the other variables have a positive effect.

Multiple logistic regression

We apply a multiple logistic regression model to study the relation between multiple variables and residential HP adoption. Table 4.2 shows the results. In Table 4.2, we show the adjusted odds ratio (AOR) variable since these are more easily interpreted than the log-odds as discussed in Section 3.3. The AOR is the odds of a variable adjusted for the other variables. It is the odds while keeping the other variables constant. Variables having an AOR greater than 1 have a positive effect on the likelihood of HP adoption. An increase in the value of a variable is not associated with a linear increase in the AOR as the steepness of the increase is dependent on the current value of variables. In addition, the confidence intervals (CI) are given for the AOR per variable indicating the strength of the estimate. If a 1 is contained in the confidence interval, then this variable is not considered significant as can be seen from the p-values. From Table 4.2, we observe that the variables *Perc1524*, *Perc2544*, *Perc4564*, *PropertyValue*, *SizeOfLivingArea*, and *DegreeOfUrbanity* for values 2 and 4 are significant with p-value 0.10.

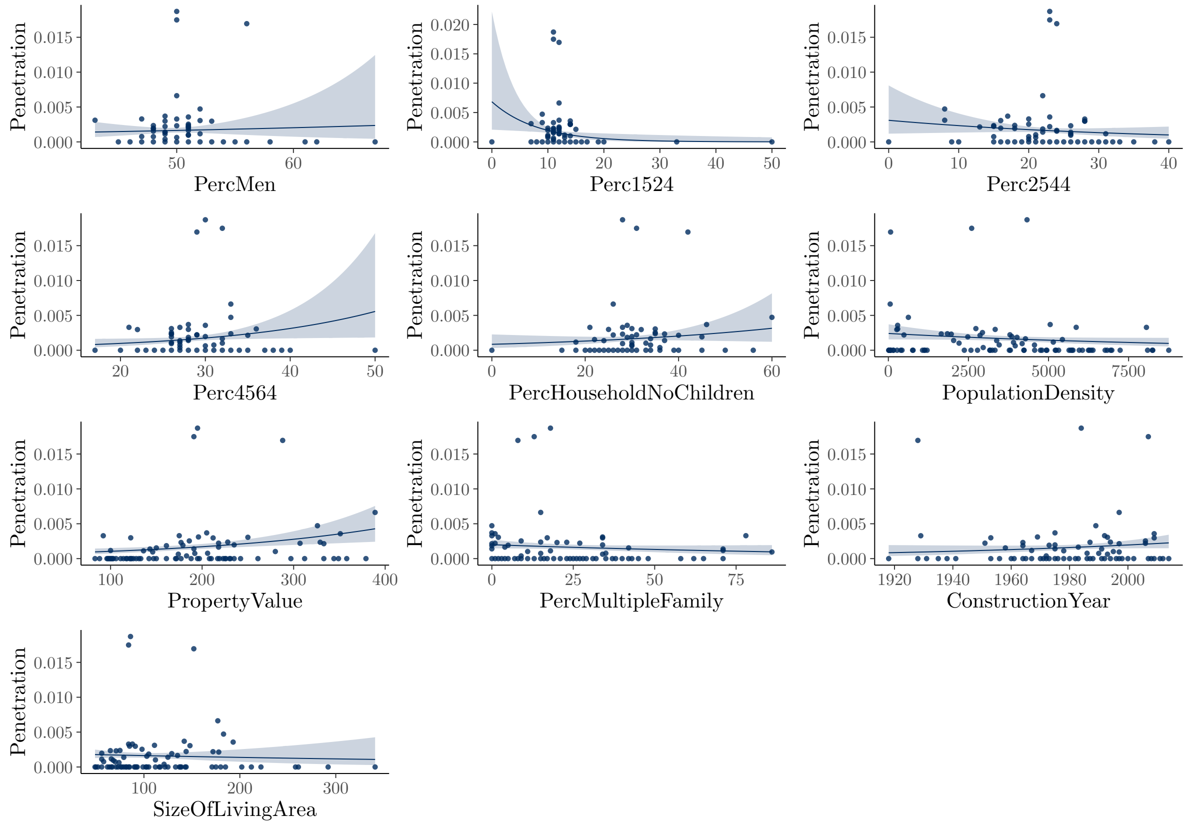


Figure 4.4: Fitted simple logistic regression models including confidence intervals for all pre-selected continuous variables.

Table 4.2: Results of the multiple logistic regression model.

	AOR	Lower 0.95 CI	Upper 0.95 CI	p-value
PercMen	1.14	0.94	1.36	0.159
Perc1524	0.71	0.61	0.83	0.000
Perc2544	1.09	0.99	1.20	0.096
Perc4564	1.16	1.06	1.28	0.002
PercHouseholdNoChildren	0.99	0.92	1.06	0.724
PopulationDensity	1.00	1.00	1.00	0.982
PropertyValue	1.01	1.01	1.02	0.000
PercMultipleFamily	0.99	0.98	1.01	0.473
ConstructionYear	1.01	1.00	1.03	0.139
SizeOfLivingArea	0.99	0.98	1.00	0.005
DegreeOfUrbanity2	18.43	3.66	336.86	0.005
DegreeOfUrbanity3	2.27	0.37	44.18	0.460
DegreeOfUrbanity4	7.16	1.24	136.29	0.070
DegreeOfUrbanity5	2.84	0.26	67.80	0.422

Stepwise logistic regression

To maintain a simple model that is more easily interpreted, we apply a stepwise variable selection procedure based on the Akaike information criterion (AIC). The AIC is a measure that penalizes models that contain many variables. We start with a full model containing all variables and repeatedly remove one variable until no improvement is found. We evaluate the improvement based on the Chi-Square test. Applying this method yields the final model shown in Table 4.3. The variables *Perc1524*, *Perc2544*, *Perc4564*, *PropertyValue*, *SizeOfLivingArea*, and *DegreeOfUrbanity* for values 2 and 4 are significant with p-value 0.05.

Table 4.3: Results of the stepwise logistic regression model.

	AOR	**Lower 0.95 CI**	**Upper 0.95 CI**	**p-value**
Perc1524	0.73	0.64	0.84	0.000
Perc2544	1.14	1.07	1.22	0.000
Perc4564	1.21	1.12	1.32	0.000
PropertyValue	1.02	1.01	1.02	0.000
SizeOfLivingArea	0.99	0.98	0.99	0.001
DegreeOfUrbanity2	16.47	3.45	295.51	0.006
DegreeOfUrbanity3	2.34	0.42	43.93	0.428
DegreeOfUrbanity4	8.59	1.63	158.58	0.041
DegreeOfUrbanity5	2.98	0.28	70.25	0.399

Results

We use the stepwise multiple logistic regression model as our final model because all variables in the model are significant. We summarize this model as follows:

- The probability of HP installation is associated with lower percentages of age category 15-24 (AOR = 0.73, CI = [0.64, 0.84], p = 0.000)
- The probability of HP installation is associated with higher percentages of age category 25-44 (AOR = 1.14, CI = [1.07, 1.22], p = 0.000)
- The probability of HP installation is associated with higher percentages of age category 45-64 (AOR = 1.21, CI = [0.12, 1.32], p = 0.000)
- The probability of HP installation is associated with higher property values (AOR = 1.02, CI = [1.01, 1.02], p = 0.000)
- The probability of HP installation is associated with smaller sizes of living area (AOR = 0.99, CI = [0.98, 0.99], p = 0.001)
- The probability of HP installation is larger in areas with 1500-2500 addresses per km^2 (AOR = 16.47 , CI = [3.45, 295.51], p = 0.006)
- The probability of HP installation is larger in areas with 500-1000 addresses per km^2 (AOR = 8.59, CI = [1.63, 158.58], p = 0.041)

Visualizing the final model results in Figure 4.5. In Figure 4.5, we observe that most outer regions have larger likelihoods of HP adoption. Using the confidence intervals of the odds, we determine the prediction intervals (PI) of the likelihood of HP adoption. Figure 4.6 shows per neighborhood the width of the prediction interval. Some estimates seem to be uncertain which should be taken into account in Chapter 4. Figure 4.7 compares the outcome of the logistic regression model with the currently found penetration levels. Figure 4.7 shows that we do not consider the current adoption of HPs (penetration), but rather we consider the likelihood of HP adoption based on influencing factors (prediction). It is thus possible that even if a neighborhood currently has no HPs, that this neighborhood is still likely to install HPs based on our findings from the logistic regression model. Figure 4.8 illustrates the expected effect on MV/LV transformers. Assuming an equal spread of HPs, then each neighborhood has an equal HP penetration level. Figure 4.8 illustrates that assuming a diversified likelihood of HP adoption reduces the pressure of most of the MV/LV transformers with high capacity utilization. This can be concluded from the blue points (diversified HP adoption) that are below the orange points (equal HP adoption) for a capacity utilization greater than 0.6.

4.4 Validation

The following items can affect the validity of the results from Section 4.3.

- The number of events (i.e., classified households) is extremely small, which makes the estimates of the logistic regression model highly unstable.
- The households classified as having installed a HP might not be representative. The estimates can be focused on a few events, rather than on a complete overview of HP adoption.
- The input was obtained by using an approximate classification. The approximate classification introduces additional uncertainty since there might be some error in the selected households (i.e., it remains uncertain if these households actually adopted a HP even though it is suggested by energy consumption data).

Checking the validity of the results from Section 4.3 is difficult because data on HP adoption per household is unavailable. Privacy of households needs to be guaranteed as documented in the AVG (Dutch: Algemene verordening gegevensbescherming), which is a European regulation for standardizing the use of personal data. In this research, we have chosen not to perform a survey because of time management. Also, a survey is not guaranteed to give good results as it might be difficult to obtain a high response rate given the small number of HPs in the service region. Nonetheless, we perform an external validation based on the conducted literature review on influencing factors of HP adoption in Section 3.2.

The positive association of income and urbanity with HP adoption found from literature corresponds with our findings; HPs are more likely to be located in urban areas with higher property values (suggesting higher income). For age and size of living area, we find a contradiction between our results and the findings from literature. Our results suggest a positive association of age and a negative association of size of living area with HP adoption, whereas literature often suggests otherwise. In literature, age is often found to have a negative effect on HP adoption; older people are possibly less willing to use new technologies. Our results suggest that older people possibly have more capital to invest in HPs. The difference in findings on size of living area might be explained as follows. Our results suggest that large houses are less easily heated by HPs, whereas literature suggests that small houses might not be suitable for HPs (i.e., apartments have no space for GSHPs). Given these findings, there might be a nonlinear relation; small houses do not have enough space, while large houses are not easily heated.

We conclude that our results are mostly in concordance with literature. The coefficients of the estimates are, however, highly uncertain given the confidence intervals. Also, we only validated the positive/negative association with HP adoption but not the strength of the coefficients. For these reasons, a grouping of the data seems reasonable; the likelihood of HP adoption remains ordered but without excessive outliers. We study the effect of grouping in Chapter 5. Due to the high uncertainty, we also perform a sensitivity analysis in Chapter 5 to study the impact of the spatial adoption of HPs on the electricity network. The current network topology might be insensitive to a diversified spread of HPs.

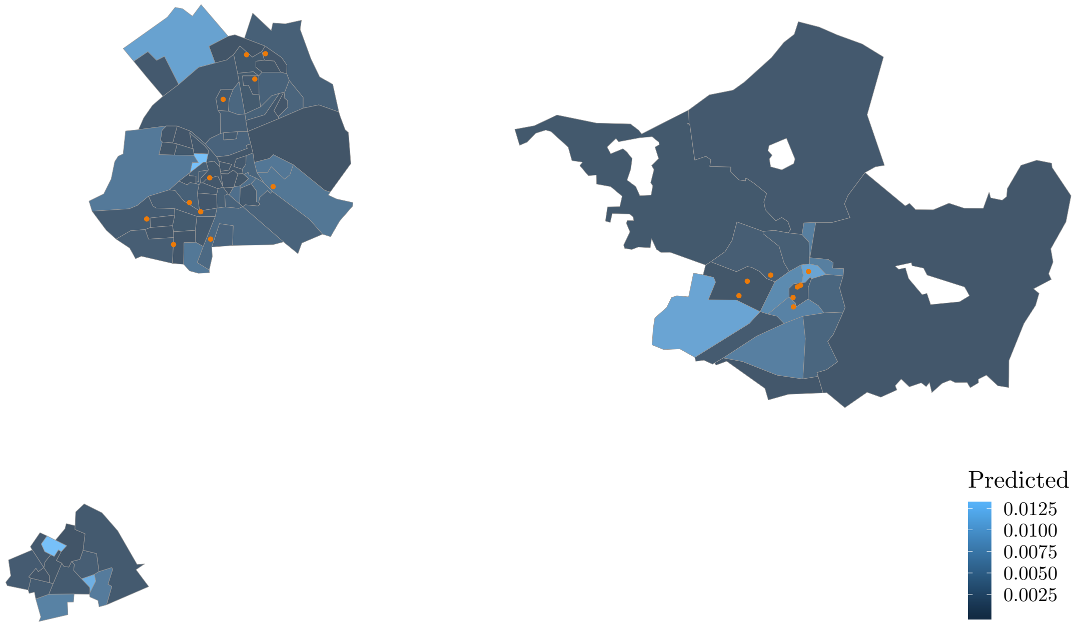


Figure 4.5: Heat map of predicted likelihood of HP adoption including MV/LV transformer substations having a capacity utilization greater than 120%. Left to right: Goor, Almelo, Oldenzaal.

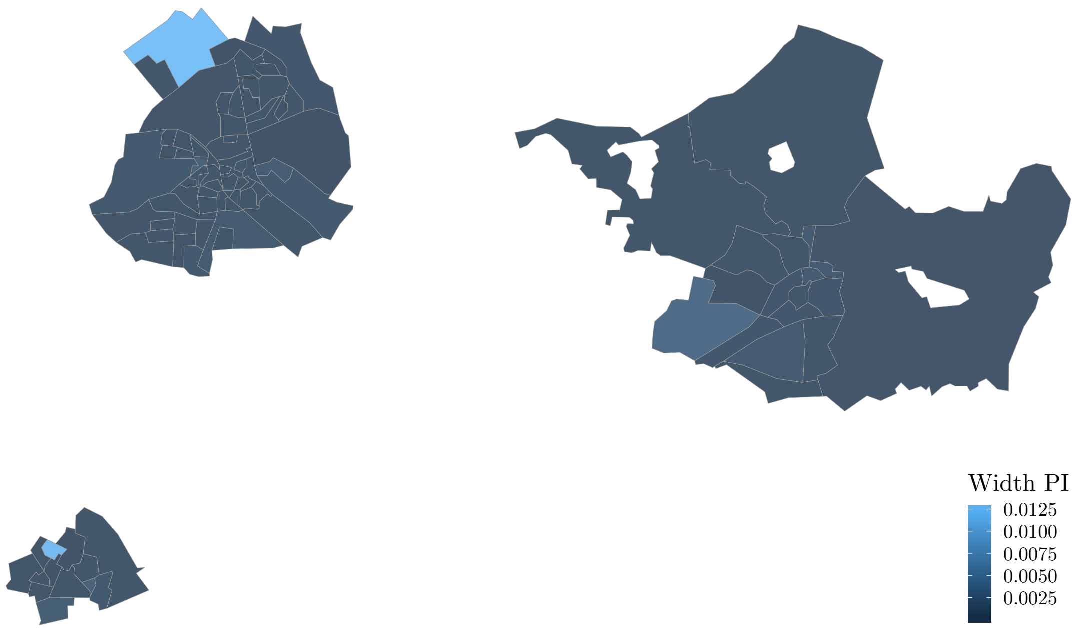


Figure 4.6: Heat map of widths prediction intervals (PI). Left to right: Goor, Almelo, Oldenzaal.

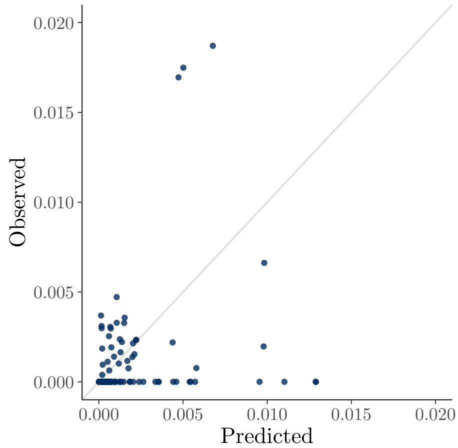


Figure 4.7: Comparison of observed penetration levels versus predicted penetration levels.

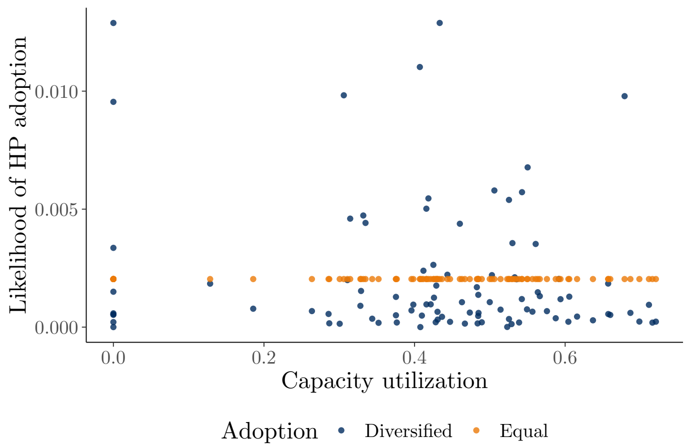


Figure 4.8: Relation between average MV/LV transformer capacity utilization per neighborhood and diversified/equal likelihood of HP adoption.

4.5 Conclusion

In this chapter, we found that the increase of residential HPs is expected to occur mostly in urban regions with higher property values. Less certain influencing factors are age and size of living area. Using literature, we found that our results are mostly valid. Given that our results are mostly valid, we found that the growth of HPs is expected to reduce the pressure on MV/LV transformers having a capacity utilization greater than 0.6. This is further studied in Chapter 5. There remains uncertainty in the strength of the coefficients. We choose the perform a sensitivity analysis to further study the effect of a diversified growth of HPs. This can help us to determine if it is feasible to include the found influencing factors although there is considerable uncertainty involved. In Chapter 5, we combine the results of Chapter 4 to simulate the spatial diffusion of HPs.

5 Simulation

In Chapter 5, we determine the impact on MV/LV transformers due to the expected spatial diffusion of heat pumps (research questions 4 and 5). We combine the findings from Chapter 4 with scenarios to diversify the growth of HPs across the AGO region for multiple scenarios. By translating the result to peak loads, we can determine the impact on MV/LV transformers. This can help Coteq in determining the necessary investments to facilitate HP usage. This chapter is organized as follows. Section 5.1 discusses the HP growth scenarios. Also, we discuss the Bass model that is used to model the HP growth scenarios. Section 5.2 discusses the simulation design for modeling the electricity network impact over space and time. In the simulation design, we introduce two approaches for modeling the spatial diffusion of HPs. Section 5.3 discusses the experiments that help us determine what approach to use and how to use the input from Chapter 4. Also, we discuss the experiments for simulating the electricity network impact and sensitivity analysis. In Section 5.4, we discuss the results of the experiments. Section 5.5 discusses the consequences for Coteq, which includes capacity expansion decisions and a cost indication. Finally, we close with a conclusion in Section 5.6.

5.1 Heat pump scenarios

In this section, we choose multiple scenarios for the future share of HPs in the AGO region. As the exact future growth of HPs is unknown, Coteq needs to evaluate the impact of multiple outcomes such that Coteq can anticipate the outcome of these scenarios. We choose to consider HPs in general, meaning that we do not distinguish between hybrid and full-electric HPs. This decision is made because of two reasons. First, in Chapter 4, we were not able to distinguish between hybrid and full-electric HPs given the small number of classified households. Second, we did not find literature suggesting that the characteristics of hybrid and full-electric HPs differ. This section is organized as follows. First, we introduce the HP scenarios on which we base the S-curves. Second, we explain how the S-curves are constructed using the Bass model and introduce four S-curves for the simulation study. Finally, we close Section 5.1 section with a conclusion.

Scenarios

Section 5.1 is based on the background information given in Appendix D. The HP is expected to be an important heating alternative for natural gas. The HP is an important technology because it can completely replace the use of natural gas but can also be combined with natural gas, hydrogen, and green gas. From the information given in Appendix D, we know that the energy system in 2050 will most likely be a combination of heating alternatives. Besides HPs, district heating is also expected to be an important heating alternative. Hydrogen, biomass, and green gas are also options but suffer from limited availability. If hydrogen, biomass, and green gas are used, then it is expected that this will be in combination with a hybrid HP. Depending on the involvement of the Dutch government, we focus on two situations. There is either a focus on using district heating, hydrogen, and green or a focus on individual solutions, which are mainly HPs. Using these situations, we introduce two scenarios described in Table 5.1.

Table 5.1: Chosen scenarios for the simulation study.

Scenario	Description
Low	This scenario is related to the *Heat* scenario introduced in the study by Berenschot (2018a) described in Appendix D. The Dutch government imports hydrogen, green gas, and biomass. Large projects are executed on hydrogen networks and heat networks (district heating). Also, a large part of the gas network remains intact such that green gas can be used. Because the Dutch government stimulates the use of these heating alternatives, fewer households are expected to install HPs. Total market share of HPs in the AGO region in 2050 is expected to be 30%.
High	This scenario is related to the *Generic* scenario introduced in the study by CE Delft (2017) described in Appendix D. Biomass and green gas are still imported, which requires an intact gas network. No hydrogen is used in the housing sector. Households mainly install hybrid HPs and are in winter provided with green gas. The increase of hybrid HPs in combination with the already installed full-electric HPs results in an expected 85% total market share of HPs in the AGO region in 2050. The remaining share consists mostly of district heating.

S-curves

We assume that the growth of HPs follows an S-shape curve and includes aspects such as governmental stimulation, HP/energy price development, municipal decisions for creating momentum, and technological development of HPs. To model the S-curves, we use the Bass model. In Chapter 3, we found that the Bass model is often used to model the diffusion of sustainable technologies and also widely validated. The Bass model requires four parameters: starting time t_0, coefficient of innovators p, coefficient of imitators q, and total market size m. The parameters t_0 and m are retrieved from the data set from the CBS and the scenarios mentioned in Table 5.1, respectively. Based on the data from the CBS, we assume that $t_0 = 1994$. We determine m by multiplying the expected market share of HPs for each scenario by the total number of households in the AGO region. The parameters p and q are found by minimizing the error between the desired S-curve and the known demand of HPs scaled to the AGO region including the final market size. In other words, we try to find the parameters p and q that follows the already known development of HPs in combination with the expected market share best. Minimizing an objective function can be done using various mathematical optimization methods. To find the desired S-curves, we assume that the current market size of HPs is 1,400. This is based on a situation in which HPs are equally spread over the Netherlands as explained in Section 2.4. Figure 5.1 visualizes the method to find the desired S-curve. In Figure 5.1, the dark line and points illustrate the known market share including the final market share and the bright line and points illustrate the S-curve based on the Bass model. The found S-curve uses what is already known on HP development and combines this with a final market share obtained from a scenario.

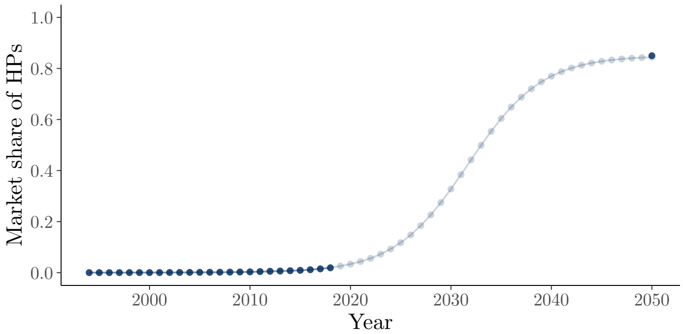


Figure 5.1: Illustration of findings the desired S-curve for the Bass model by minimizing the error between the Bass model and known demand including an estimated final market share.

Using the scenarios from Table 5.1, we obtain two S-curves. From Chapter 2, we found that the current number of HPs is highly uncertain. For this reason, we also introduce two scenarios that have a current market size of 400 HPs. This market size of HPs is based on the total number of HP subsidy requests between 2016 and 2019 in the AGO region given in Section 2.4. Given our results from Chapter 4, this might be more realistic. We thus assume that 100 HPs are located in existing houses, given our results from Chapter 4, and 300 in new-build houses. In summary, we have the following four scenarios:

- High_A: A 85% final market share of HPs with 400 HPs currently installed
- High_B: A 85% final market share of HPs with 1,400 HPs currently installed
- Low_A: A 30% final market share of HPs with 400 HPs currently installed
- Low_B: A 30% final market share of HPs with 1,400 HPs currently installed

We obtain the S-curves based on 400 HPs currently installed by shifting the already fitted S-curves in time. This means that the steepness of the S-curves related to scenarios High_A and High_B is identical. This also holds for the S-curves related to scenarios Low_A and Low_B. Figure 5.2 visualizes the S-curves. The chosen scenarios are not exact predictions, but rather an illustration of plausible outcomes. The chosen S-curves reflect both the urgency of needed investments and the degree of electrification needed resulting from the share of HPs.

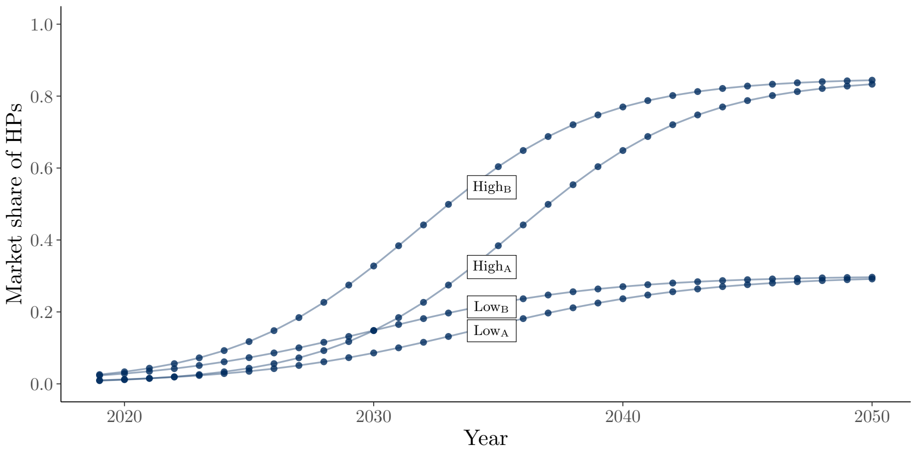


Figure 5.2: Constructed HP S-curves based on scenarios High_A, High_B, Low_A, and Low_B.

5.2 Simulation design

The goal of this section is to design a simulation model that is able to determine the increased peak load on the electricity network due to the spatial diffusion of residential HPs for multiple scenarios. The steps discussed in this section are based on the work of Law (2015). In this section, we discuss type of simulation, input, assumptions, process-flow, uncertainty, model validation, and sensitivity analysis. In the description of the process-flow, we introduce two approaches that can be used to assign HPs to households.

Type of simulation

We dynamically model the spatial diffusion of HPs, meaning that the result of time t is dependent on the result of time $t-1$ (i.e., households that have installed a HP in year $t-1$ cannot install a HP in the coming years $t, \ldots, T$). We evaluate the system once per year. As discussed in Chapter 2, it is most interesting to study the system under worst-case conditions. Because the electricity demand of HPs is largest in winter, we evaluate the system based on HP requirements in winter. The output of the simulation model is expressed in the capacity utilization of MV/LV transformers. We define the capacity utilization by the fraction of a MV/LV transformer that is utilized due to the peak load of electricity demand. For example, a MV/LV transformer with a capacity of 400 kVA and a peak load of 350 kVA is utilized for 87.5%. The reason for this output measure is that we can easily determine if a MV/LV transformer is overloaded.

Input

The simulation model requires the following input: likelihood of HP adoption, HP S-curves, network topology, and HP requirements in winter. In this paragraph, we first discuss the likelihood of HP adoption, HP S-curves, and network topology. In the next paragraph, we discuss the HP requirements. The likelihood of HP adoption is used to diversify the growth of HPs in a year and is obtained from the logistic regression model from Chapter 4. In Chapter 4, we introduced a grouping the likelihood of HP adoption. This setting is discussed in Section 5.3. The S-curves determined in Section 5.1 are used to determine the yearly growth of HPs per scenario, which we assume to be deterministic. Using the network topology, we can determine which households are connected to which MV/LV transformer. As discussed in Section 2.2, of the 403 MV/LV transformer substations connecting households, 13 have more than one MV/LV transformer. Unfortunately, only data is obtained on which households are connected to which MV/LV transformer substation. To deal with this issue, we aggregate the capacity and peak load if a MV/LV transformer substation has multiple MV/LV transformers.

Finally, data on HP requirements are used to make the translation to the electrical load. Based on empirical studies discussed in Appendix E, we construct a piecewise linear function. The constructed piecewise linear function is shown in Figure 5.3. In Figure 5.3, ADMD is the After Diversity Maximum Demand. The ADMD refers to the situation that not every HP is used at the same time. The average HP peak load measured at a MV/LV transformer decreases as more HPs are connected to the MV/LV transformer. The total additional peak load of HPs is thus not a linear function in the number of HPs. This is shown in Equation 2.2 in Section 2.2. We constructed the piecewise linear function shown from Figure 5.3 based on empirical data provided by Love et al. (2017). This function is based on a cold winter weekday with an external temperature of -0.3°C (illustrating worst-case conditions). Although this is not the exact curve found by Love et al. (2017), it covers the main characteristics. From Appendix E, we know that the peak load of HPs occurs in the morning whereas the normal residential peak load occurs in the evening. Based on the data from Love et al. (2017), we estimate that the peak of HPs is 25% lower during the normal residential peak load. For this reason, we have multiplied the ADMD by 0.75, which is included in Figure 5.3. We assume that the power factor ($\cos \varphi$), as discussed in Section 2.3, is 0.95. This is based on a study of Akmal, Fox, Morrow, and Littler (2014) who performed a detailed study on the HP impact in terms of kVA.

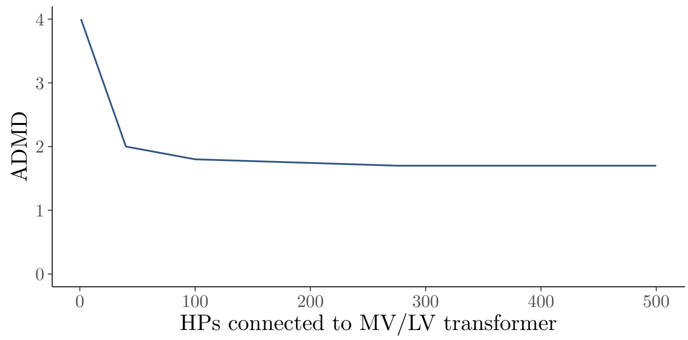


Figure 5.3: Relation between the average peak load of HPs and the number of households connected to a MV/LV transformer.

Assumptions

In the simulation study, we use the following simplifying assumptions.

- The currently found likelihood of HP adoption is representative for the future.
- Neighborhood characteristics will remain the same throughout the simulation study.
- There is no difference between households within a neighborhood.
- As we only consider one HP, we assume that the differences in needed power for different HPs are reflected in the ADMD given in Figure 5.3.
- Up to 2050, there is only one point in time in which a household installs a HP. There is no further upgrade of a HP.

Process-flow

Figure 5.4 shows the process-flow of the simulation study for one scenario. The dashed lines indicate the required input. For each year, we extract the HP growth from the constructed S-curve, assign it to households and determine the capacity utilization. Knowing which households installed a HP, according to our assumptions, we can calculate the increased peak load per MV/LV transformer substation. This is done by using the piecewise linear function shown from Figure 5.3 and a power factor ($\cos\varphi$). The HP peak load is added to the known peak load of a MV/LV transformer substation. The total load is divided by the available capacity to obtain the capacity utilization. For assigning HPs to households, we introduce two approaches that use different probability distributions for determining which households/residential areas get assigned HPs. To make a decision on what approach to use, we perform experiments. This is explained in Section 5.3. We now address these approaches.

1. Arbitrary discrete distribution
2. Fisher's noncentral hypergeometric distribution

Approach 1 is based on an arbitrary discrete distribution. Using the logistic regression model, we predict per household the likelihood of HP adoption. These likelihoods have a ratio scale; a household having a likelihood of 0.4 is twice more likely to adopt a HP than a household having a likelihood of 0.2. When assigning HPs to households, we want to use this ratio scale. The probability of assigning a HP a household having a likelihood of 0.4 should be two times larger than assigning a HP to a household with a likelihood of 0.2. This can be achieved by constructing an empirical discrete distribution. A detailed description of how this is achieved is given in Appendix F. After having assigned a HP, we remove the chosen household and reconstruct the empirical discrete distribution. This procedure is repeated until the total HP growth in year t is assigned to households. Approach 1 directly assigns HPs to households. This is slightly different for method 2.

Approach 2 is based on a Fisher's noncentral hypergeometric distribution. Instead of directly assigning HPs to households, we first allocate the growth of HPs in year t to the residential areas. We assume that the number of HPs per neighborhood given a certain yearly growth is distributed according to the Fisher's non-central hypergeometric distribution. The yearly allocation of the number of HPs to residential areas is based on four inputs: number of households per residential area, the likelihood of HP adoption per residential area (one value per residential areas suffices because all households are assumed to have identical characteristics), number of households not having adopted a HP per residential area, and the HP growth. The number of HPs allocated to a residential area is now based on the number of households in the area in combination with the likelihood of HP adoption in that residential area. Also, no more HPs can be allocated than the number of households not having adopted a HP yet. Knowing how many HPs are allocated to a residential area, we can now randomly assign the number of HPs to households within each residential area. The detailed procedure is given in Appendix F. Approach 2 is only used once per year in the simulation, as opposed to approach 1, which is used each year for each HP that needs to be assigned.

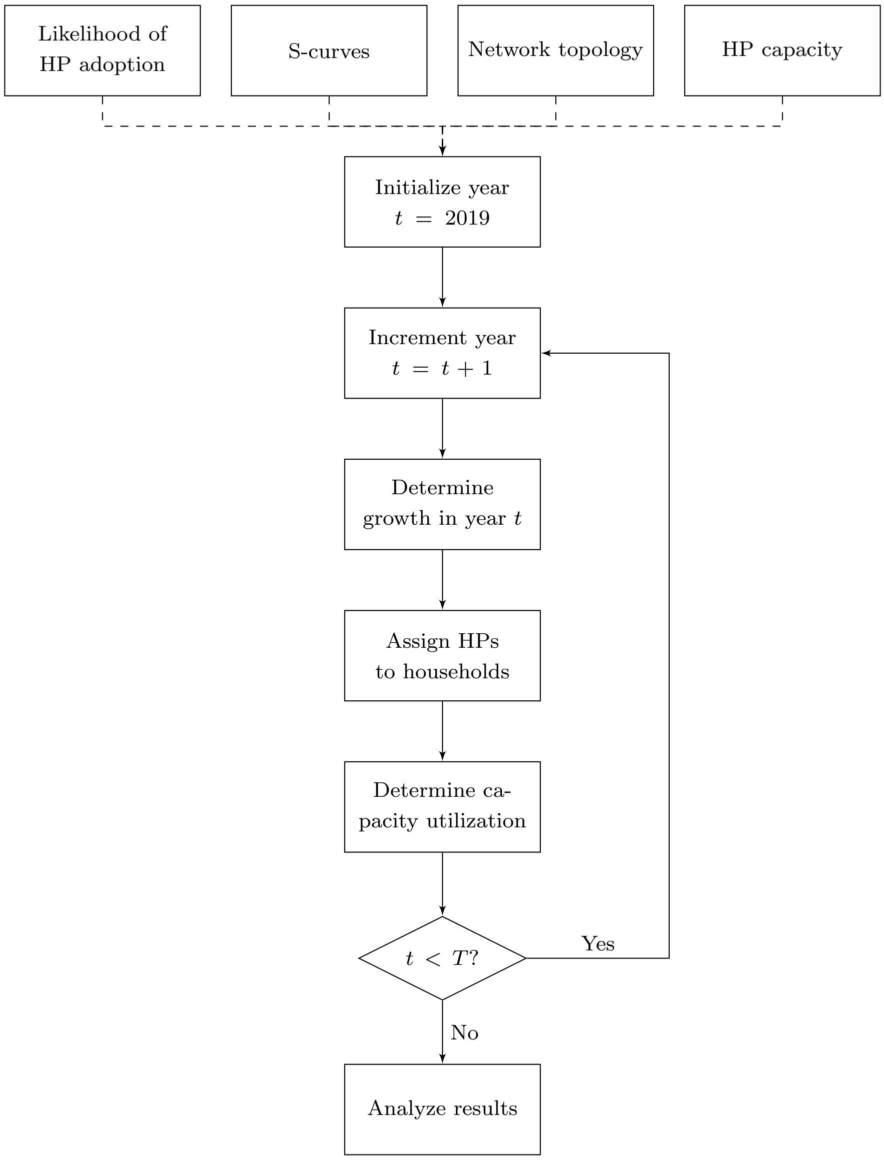


Figure 5.4: Process-flow used in the simulation study.

Uncertainty

We include uncertainty in this simulation study in various ways. First, the uncertainty of the logistic regression model in Chapter 4 is reflected in a confidence interval for each coefficient. The prediction intervals shown from Figure 4.6 illustrate the uncertainty in the predictions resulting from these coefficient confidence intervals. We want the simulation model to take this uncertainty into account since it eventually affects how HPs are assigned to neighborhoods. We do this as follows. We assume that the estimates of the odds are normally distributed. We then construct a multivariate normal distribution of the coefficients using the means and covariance matrix obtained from the logistic regression model. By repeatedly drawing random variates from this distribution, we essentially create a prediction interval of the response per neighborhood as we did in Chapter 4. This way we include the uncertainty in the likelihood of a neighborhood. This method is based on the work of Van Horssen, Pebesma, and Schot

(2002). Another way we include uncertainty is the way we assign HPs to households. We essentially want households in a residential area to have an equal probability of installing a HP. To achieve this, we randomly assign HPs to households using the uniform distribution (detailed information can be found in Appendix F).

Validation and sensitivity analysis

We want to evaluate the strength of the conclusions that can be drawn from the simulation study. The strength of our conclusions resulting from the simulation model is tested using model validation and sensitivity analysis (MIT Critical Data, 2016). The strength can be determined using internal and external validation. Internal validation is used in the logistic regression by using a stepwise procedure. The stepwise procedure reduced overfitting by focusing only on significant variables. We then externally validated the significant variables with literature. Because most agreement was found on the variables *PropertyValue* and *DegreeOfUrbanity*, we test the sensitivity of the simulation using only these two variables. Also, we want to test the sensitivity of not using spatial adoption, meaning that HPs are equally spread over the AGO region. Finally, we also consider a situation in which extreme winters occur. We introduce the settings for evaluating the sensitivity in Section 5.3.

5.3 Experimental design

In this section, we introduce the experimental design. First, we discuss the experiments used to make a decision between the two introduced approaches to model the spatial diffusion of HPs. Second, we introduce how grouping can be applied to the results of the logistic regression model. Also, we discuss the experiments on which we base the decision of using grouping or not. Third, we describe the experiments used for the complete simulation study. Finally, we describe the experiments for the sensitivity analysis.

Approaches

To determine what approach to use to model the spatial diffusion of HPs, we perform the following experiment. We simulate the spatial diffusion of HPs for years $t = (2020, \ldots, 2050)$ for 10 replications. We do this using scenarios High_B and Low_B. Scenarios High_A and Low_A are shifted versions of scenarios High_B and Low_B, meaning that the results on scenarios High_B and Low_B give sufficient evidence on which we can base our decision. We do not consider the uncertainty in coefficients of the logistic regression model as this would require a large number of replications to make a good comparison. The goal of this experiment is to assess the comparability and computation time of the approaches. Both approaches are expected to give comparable outcomes. We verify this and make a decision in Section 5.4.

Grouping

As found in Chapter 4, the spatial adoption of HPs was found to suffer from outliers and large prediction intervals for some neighborhoods. There are several elements that reduce the reliability of the spatial adoption of HPs found from Chapter 4. First, the results from Chapter 4 are based on soft labels (i.e., we do not know with 100% certainty which households have HPs). Even if the results from Chapter 4 were based on hard labels (i.e., knowing exactly which households have HPs), then this still may only describe a snapshot of what we currently see. Second, shown from the context analysis in Chapter 2, the HP is still in the beginning phase of its product life cycle. To reduce the effect of extreme likelihoods for some neighborhoods, we consider grouping the neighborhoods. To do this, we choose to group the neighborhoods according to their likelihood of HP adoption. To classify the neighborhoods, we use the adopter categories as given by Rogers (1983). We make the assumption that the results from Chapter 4 function as a measure of innovativeness (e.g., households with larger capital can more easily innovate). More innovative neighborhoods experience more rapid growth in the fraction of HP installed. Using this logic, we divide the neighborhoods according to the classes shown from Table 5.2 given by Rogers (1983). When all the neighborhoods are sorted on the likelihood of adoption, the most right 2.5% are the innovators, followed by the other categories. Instead of using the likelihood, we now use a score as the basis for the approaches. To study the effect of grouping, we study the spatial diffusion of HPs for years $t = (2020, \dots, 2050)$ for scenarios High_B and Low_B. In this analysis, we use only one approach, which is found using the previously mentioned experimental design. Finally, we replicate this analysis 10 times.

Table 5.2: Distribution of adopter categories based on the *Diffusion of innovations*. Retrieved from: (Rogers, 1983).

Adopter category	Percentage	Score
Innovators	2.5%	5
Early adopters	13.5%	4
Early majority	34%	3
Late majority	34%	2
Laggards	16%	1

Simulation

Using the results from the previous experiments, we determine what approach to use and if grouping should be applied. Based on these decisions we simulate the complete spatial diffusion of HPs. We use the following experimental design. We perform 100 replications for scenarios High_B, High_A, Low_B, and Low_A. We evaluate the model for the years $t = (2020, 2030, 2040, and 2050)$. For these experiments, we include the uncertainty given in the logistic regression model.

Sensitivity analysis

For the sensitivity analysis, we consider the following three input changes.

1. Spatial adoption based only on variables *DegreeOfUrbanity* and *PropertyValue*
2. No spatial adoption (random spread of HPs over AGO region)
3. Extreme winter conditions (ADMD increases with 150%)

Because we found a contradiction between literature and our results on the association of age and size of living area, we consider two input changes on the likelihood of HP adoption. Input change 1 refers to the situation where the likelihood of HP adoption is only based on the variables *DegreeOfUrbanity* and *PropertyValue*. Input change 2 refers to the situation in which HPs are randomly spread over the AGO region without preference for a residential area. The effect of the first two adjustments is evaluated by focusing on the time period $t = (2020, \dots, 2030)$. The reason that we focus on this time period is that we can provide a more detailed analysis. For example, we can more closely determine which neighborhoods will experience overloaded MV/LV transformers substations and in what year. We can more easily compare this with the original model to determine the effect. Finally, we consider the extreme winter case (input change 3). The power requirement of HPs is related to the outside temperature. When the outside temperature decreases, HPs operate less efficient and require more electrical power. This can have severe consequences on the assets of Coteq. To determine the HP requirements in winter, we multiple the complete piecewise linear function shown from Figure 5.3 by 1.5 (Miller, 2015). This situation has a low probability of occurring but illustrates the expected impact on the electricity network. We evaluate the effect of extreme winter conditions for the years $t = (2020, 2030, 2040, 2050)$. All experiments for the sensitivity analysis are based on hundred replications including uncertainty.

5.4 Simulation results

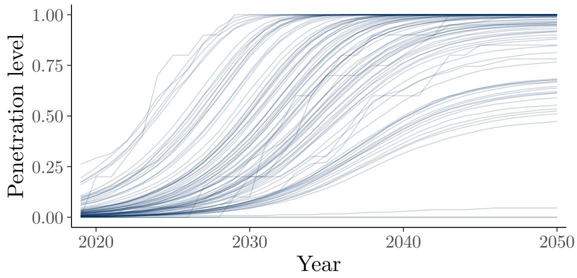
This section discusses the results of the simulation study. First, we discuss the two approaches introduced in the simulation design. Second, we study the effect of grouping on the spatial diffusion of HPs. Third, we perform the complete simulation using one approach with or without grouping including the uncertainty considerations and discuss the results. Finally, we perform a sensitivity analysis to study the robustness of the simulation study.

Approaches

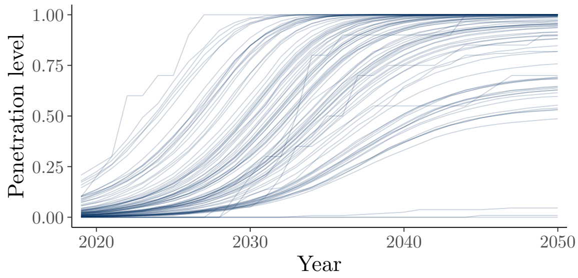
Figure 5.5 visualizes the results using the two approaches for two scenarios. In Figure 5.5, each line corresponds to one neighborhood. The penetration level increases over time as HPs are assigned to households according to the approaches. Because we have only run 10 replications, we can observe rough lines due to random numbers. These rough lines are a consequence of neighborhoods with few households. If a HP is assigned to a neighborhood with few households, then the penetration level increases abruptly. Figures 5.5a and 5.5b show that for scenario High_B the results are similar. We average the absolute difference in HP growth over the years and neighborhoods. This results in an average of 3. Based on this number we assume that both approaches produce similar results. Similar results are seen for scenario Low_B shown in figures 5.5c and 5.5d, where the average is 2. In terms of performance, the approach using the Fisher's non-central hypergeometric distribution is much faster. The time needed to simulate one scenario for one replication was 0.06 seconds, opposed to 46.45 seconds needed for the approach using the empirical distribution. We expect that this is because the values needed for the empirical distribution need to be updated every time a heat pump is assigned. This is needed since the probabilities of the other households changes when one household is removed (i.e., the sum must be equal to 1). Given these results, we use approach 2 in the upcoming experiments.

Grouping

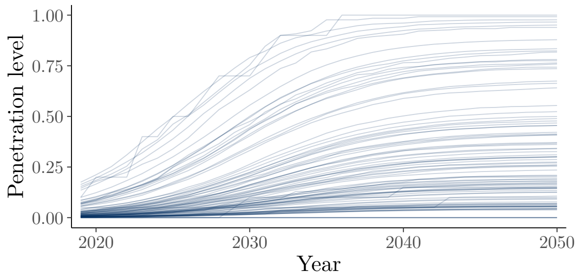
Figure 5.6 shows the effect of grouping the neighborhoods based on their likelihood of HP adoption. We can observe that grouping produces mainly five S-curves. This is a direct consequence of grouping. Figures 5.6a and 5.6b show that for scenario High_B grouping reduces the diversity between neighborhood penetration levels. Without grouping, some neighborhoods already achieve a 100 percent HP penetration level before 2030 in scenario High_A. By grouping, this effect is smaller and thus reduces the effect logistic regression has on the output in the simulation study. Figures 5.6c and 5.6d give similar results. Without grouping, a large part of the neighborhoods remains under 20 percent HP penetration level. Given our results in Chapter 4, we choose to use grouping because, in our opinion, there is not enough evidence to use the individual likelihood of HP adoption.



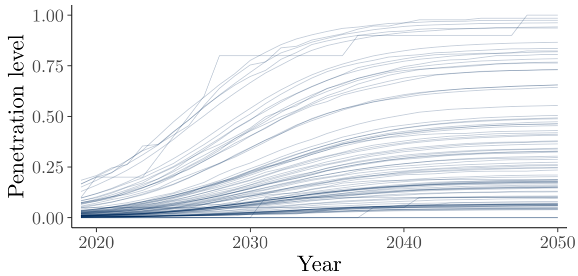
(a) Scenario High$_{\mathrm{B}}$, approach 1



(b) Scenario High$_{\mathrm{B}}$, approach 2

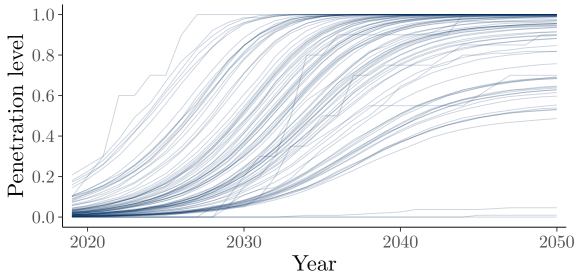


(c) Scenario Low$_{\mathrm{B}}$, approach 1

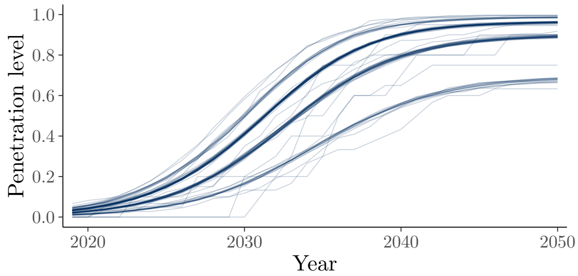


(d) Scenario Low$_{\mathrm{B}}$, approach 2

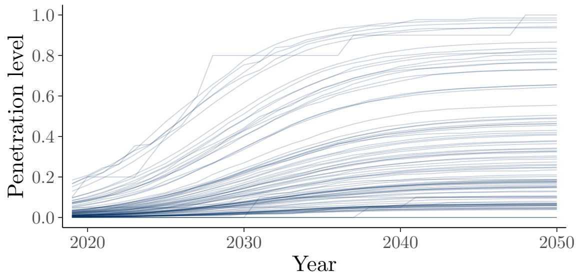
Figure 5.5: Comparison of results obtained for both approaches for scenarios High$_{\mathrm{B}}$ and Low$_{\mathrm{B}}$. Each line corresponds to the increase of HP penetration level for one neighborhood.



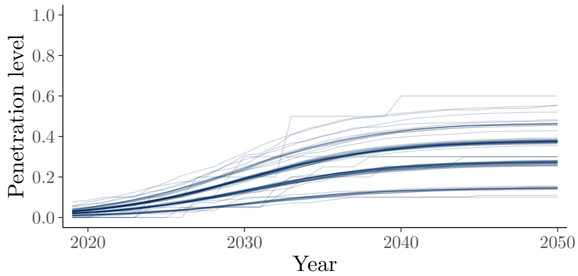
(a) Scenario High$_{\mathrm{B}}$, ungrouped



(b) Scenario High$_{\mathrm{B}}$ grouped



(c) Scenario Low$_{\mathrm{B}}$, ungrouped

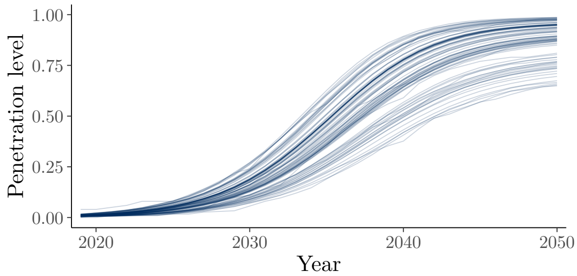


(d) Scenario Low$_{\mathrm{B}}$ grouped

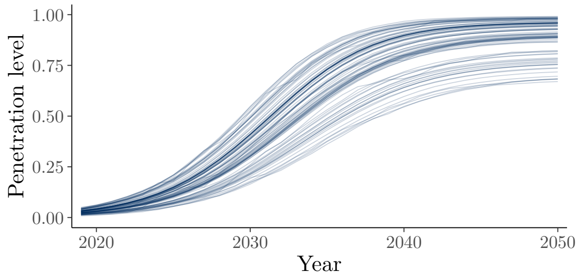
Figure 5.6: Comparison of results obtained by grouping the likelihood of HP adoption for scenarios High$_{\mathrm{B}}$ and Low$_{\mathrm{B}}$. Each line corresponds to the increase of HP penetration level for one neighborhood.

Results

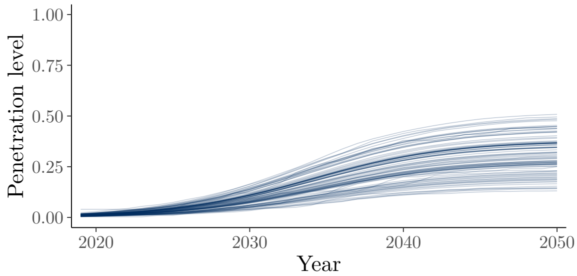
We now discuss the results of the simulation study based on using approach 2 (Fisher's non-central hypergeometric distribution) in combination with grouping the likelihood of HP adoption. Figure 5.7 visualizes the results of the growth per neighborhood. Figures 5.7a and 5.7b are very similar. There is a mild differentiation between neighborhoods; some neighborhoods achieve a 60% penetration level in 2050, while other neighborhoods achieve a 100% penetration level. The only difference is seen in how fast the penetration level increases. Figures 5.7c and 5.7d show similar results. We can observe that as opposed to Figure 5.6, we cannot distinguish between five S-curves. We visualize the impact on MV/LV transformer substations for the years 2030 and 2050 for the worst-case scenario $\mathrm{High_B}$ in Figure 5.8. Overloaded MV/LV transformer substations (i.e., more than 120% capacity utilization) are visualized as red dots. Shown from Figure 5.8a, there are five overloaded MV/LV transformer substations in 2030. In 2050, approximately 80 MV/LV transformer substations are overloaded, as shown in Figure 5.8b. Table 5.3 shows the average number of overloaded MV/LV transformer substations for each scenario for the years 2020, 2030, 2040 and 2050 including the confidence interval. We note that the overloaded MV/LV transformer substations are a consequence of HP adoption in general but not necessarily because of spatial adoption. Already highly utilized MV/LV transformer substations might rapidly become overloaded without being caused by the spatial adoption of HPs. This is further analyzed in the sensitivity analysis.



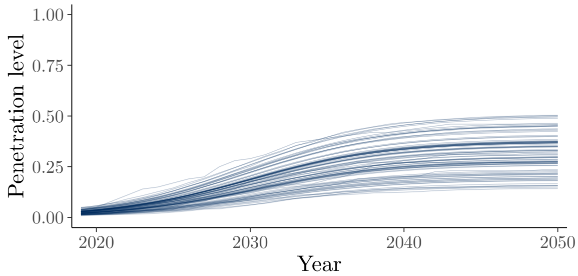
(a) Scenario $\mathrm{High_A}$



(b) Scenario $\mathrm{High_B}$

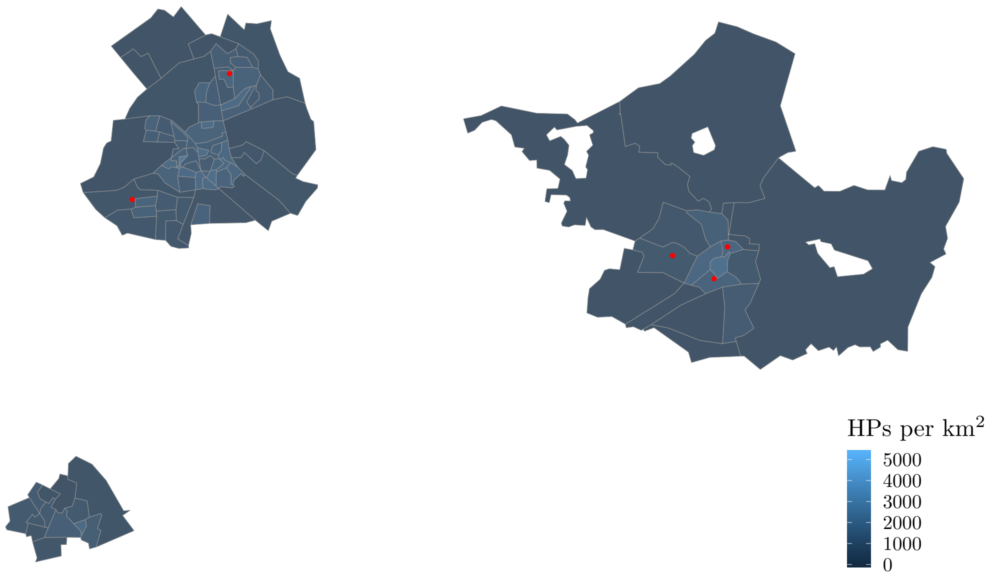


(c) Scenario $\mathrm{Low_A}$

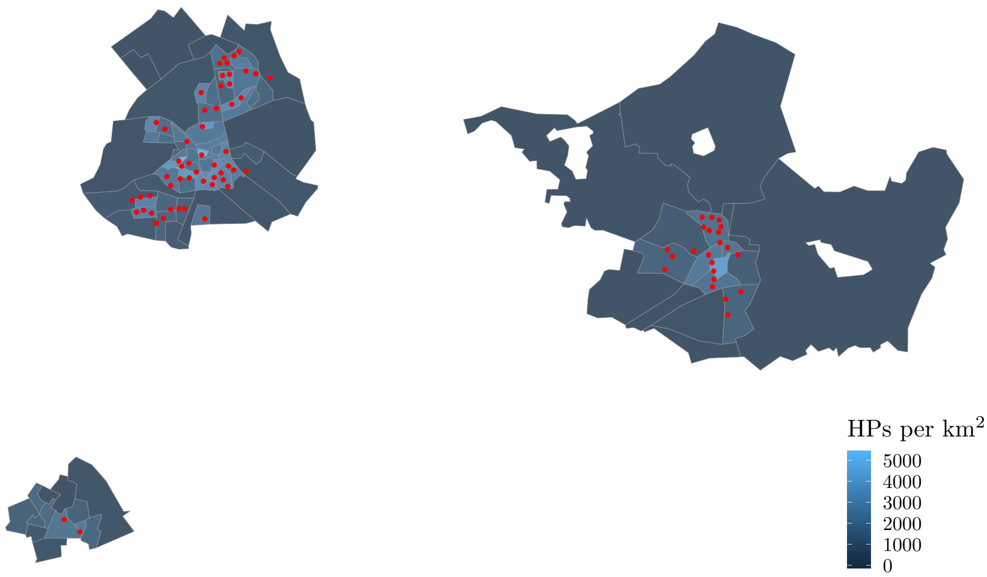


(d) Scenario $\mathrm{Low_B}$

Figure 5.7: Comparison of results for scenarios $\mathrm{High_B}$, $\mathrm{Low_B}$, $\mathrm{High_A}$, and $\mathrm{Low_A}$. Each line corresponds to the increase of HP penetration level for one neighborhood.



(a) 2030



(b) 2050

Figure 5.8: Number of HPs per km^2 and MV/LV transformer substations loaded above 120% (shown as red dots) in 2030 and 2050 for scenario High$_\mathrm{B}$.

Table 5.3: Mean and confidence interval (CI) of the number of overloaded MV/LV transformer substations per scenario in 2020, 2030, 2040, and 2050 (L = lower CI, U = upper CI).

	2020			**2030**			**2040**			**2050**		
		CI			CI			CI			CI	
	Mean	L	U	Mean	L	U	Mean	L	U	Mean	L	U
$High_A$	0	0	0	0.69	0.57	0.81	47.87	47.3	48.44	80.48	80.12	80.84
$High_B$	0	0	0	5.1	4.93	5.27	68.01	67.47	68.55	79.24	78.83	79.65
Low_A	0	0	0	0.04	0	0.08	3.01	2.87	3.15	4.65	4.5	4.8
Low_B	0	0	0	0.42	0.32	0.52	3.67	3.53	3.81	4.16	4.02	4.3

Figure 5.9 illustrates the impact on the capacity utilization of MV/LV transformer substations for each scenario in 2050 compared to the current situation. There is not much difference between scenarios High_A and High_B as the final market share is identical. This also holds for scenarios Low_A and Low_B. The difference between these scenarios is more related to the urgency of doing network investments instead of the total investments needed. In the scenarios with an 85% market share of HPs in 2050, there is a considerable increase in capacity utilization. Scenario High_B, for example, shows that 50% of the MV/LV transformers substations have a capacity utilization between 52% and 109%, whereas in the current situation this is between 35% and 61%.

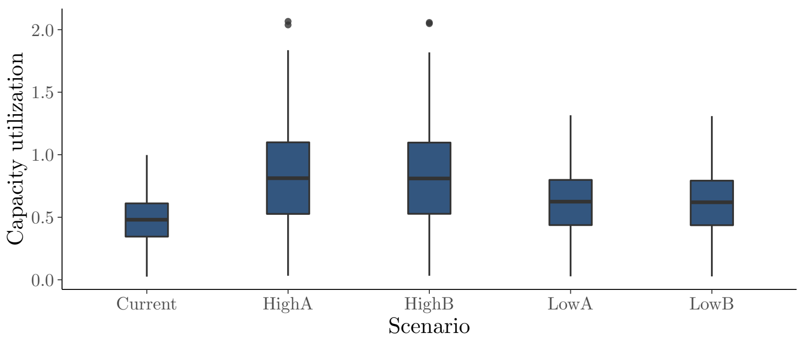
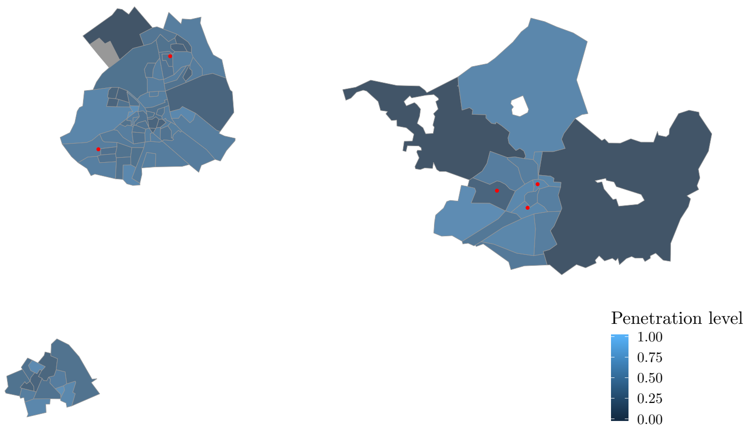


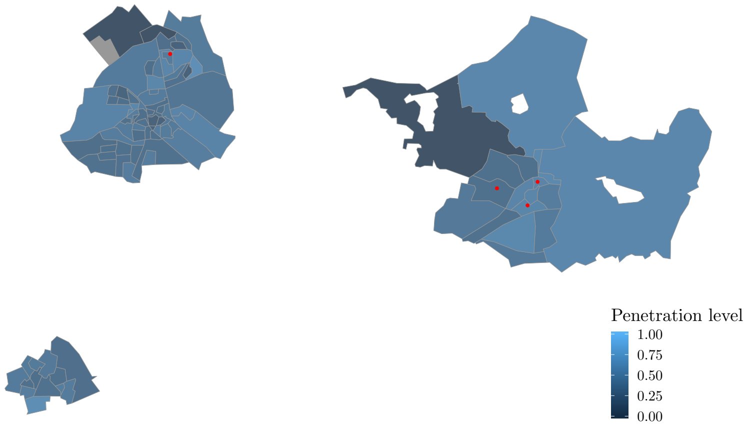
Figure 5.9: Capacity utilization boxplots of MV/LV transformers substations in 2050 for scenarios High_B, Low_B, High_A, and Low_A averaged over all runs.

Sensitivity analysis

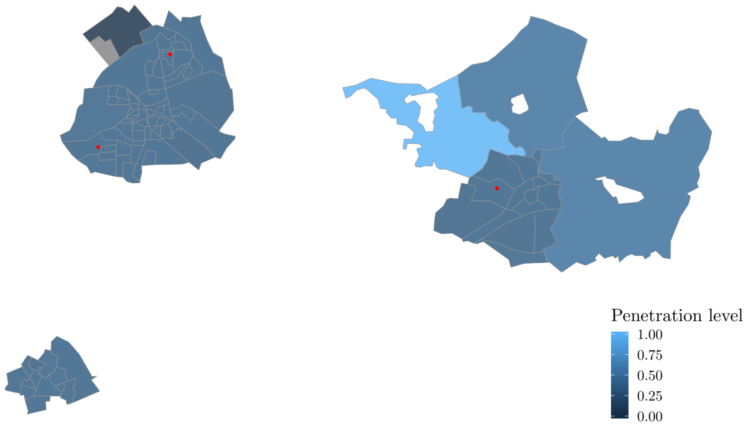
We now discuss the effects of altering the assumptions. Figure 5.10 visualizes the overloaded MV/LV transformer substations in 2030 for three types of input. Figures 5.10a and 5.10b both have similar overloaded MV/LV transformer substations. The model using only the variables *PropertyValue* and *DegreeOfUrbanity* shows one more overloaded MV/LV transformer substation compared to the model including all significant variables. Figure 5.10c clearly shows that assuming no spatial adoption results in equal penetration level for all neighborhoods, except for the outer regions where Coteq only manages a few electricity connections. To more clearly show the differences between the models, we also tabulated the number of overloaded MV/LV transformer substations per neighborhood. Table 5.4 shows the result. Tables 5.4a and 5.4b show that the spatial adoption causes a more diversified impact. Up to 2030, the HP growth affects MV/LV transformer substations in 4 to 5 neighborhoods. No spatial adoption of HPs, meaning that HPs are equally spread, results in fewer affected neighborhoods. This is shown in Table 5.4c. This is in concordance with our findings from Chapter 4. The spatial adoption of HPs reduces the pressure of MV/LV transformer with high capacity utilization.



(a) Original model



(b) Model containing variables *PropertyValue* and *DegreeOfUrbanity*



(c) No spatial adoption

Figure 5.10: Results of the sensitivity analysis on logistic regression model used as input for the simulation study based on scenario $\mathrm{High_B}$ in 2030.

Table 5.4: Cumulative sum of overloaded MV/LV transformer substations per neighborhood having at least one overloaded MV/LV transformer per model.

(a) Original model

	2020	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**
BU01411801									1	1	1
BU01411901											1
BU01730200										1	1
BU01730400										1	1
BU01731100								1	1	1	1

(b) Model containing variables *PropertyValue* and *DegreeOfUrbanity*

	2020	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**
BU01411801								1	1	1	1
BU01730200										1	1
BU01730400											1
BU01731100							1	1	1	1	1

(c) No spatial adoption

	2020	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**
BU01411801								1	1	1	1
BU01411901											1
BU01731100						1	1	1	1	1	1

Considering the effect of severe winters, the number of overloaded MV/LV transformer substations increases as shown in Table 5.5. Figure 5.11 illustrates the effect of severe winters on MV/LV transformer substations in 2050 for the scenarios introduced in Section 5.1. The box plots in Figure 5.11 show that the distribution of capacity utilization further increases. For scenario High_B, 50% of the MV/LV transformers have a capacity utilization between 59% and 137%.

Conclusion

In this section, we presented the results of the experiments introduced in Section 5.3. We found that for modeling the spatial diffusion of HPs both the empirical discrete distribution and Fisher's noncentral hypergeometric distribution are suitable. The Fisher's noncentral hypergeometric distribution has a preference because of its fast calculations. Also, we evaluated the effect of grouping the neighborhoods based on the degree of innovativeness. We chose to use grouping as we do not have sufficient evidence to believe that some residential areas already achieve a 100% penetration level in 2030. Next, we evaluated the impact on MV/LV transformer substations. In 2030, approximately 5 MV/LV transformers substations are expected to be overloaded for the scenarios having a high final market share of HPs. This rapidly increases to approximately 80 in 2050. The scenarios based on a low final market share of HPs only have 5 overloaded MV/LV transformer substations in 2050. The spatial adoption patterns of HPs reduce the of MV/LV transformers with capacity utilization. This gives that bottlenecks are expected to be postponed assuming the spatial adoption of HPs.

Table 5.5: Mean and confidence interval (CI) of the number of overloaded MV/LV transformer substations per scenario in 2020, 2030, 2040, and 2050 under extreme winter conditions (L = lower CI, U = upper CI).

	2020			**2030**			**2040**			**2050**		
		CI			CI			CI			CI	
	Mean	L	U	Mean	L	U	Mean	L	U	Mean	L	U
$High_A$	0	0	0	3.26	3.13	3.39	100.76	100.16	101.36	147.47	147.06	147.88
$High_B$	0	0	0	16.76	16.23	17.29	124.12	123.61	124.63	145.13	144.71	145.55
Low_A	0	0	0	0.9	0.74	1.06	7.82	7.56	8.08	13.85	13.44	14.26
Low_B	0	0	0	2.82	2.69	2.95	9.5	9.22	9.78	12.35	11.96	12.74

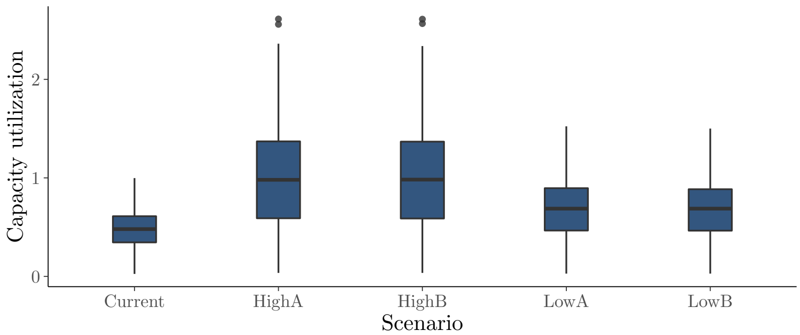


Figure 5.11: Capacity utilization boxplots of MV/LV transformers in 2050 under extreme winter conditions for scenarios High_B, Low_B, High_A, and Low_A averaged over all runs.

5.5 Consequences Coteq

We apply Algorithm 1 to determine what investments are at least needed on MV/LV transformer substations for one scenario to prevent bottlenecks resulting from HPs. When a MV/LV transformer substation is overloaded, we expand capacity such that the capacity utilization in 2050 is below 120%. As discussed in Section 5.2, we only know which households are located to which MV/LV transformer substation. Therefore, in Algorithm 1, we consider the capacity utilization of MV/LV transformer substations. In Algorithm 1, we have two options: upgrade or add a MV/LV transformer. We first try to upgrade MV/LV transformers to either 400 kVA, or if necessary, 630 kVA. If all MV/LV transformers at a MV/LV transformer substation have reached 630 kVA and if the capacity utilization is still above 120%, then we add a new MV/LV transformer. This automatically results in a new MV/LV transformer substation with a capacity of 630 kVA to support the LV network. For simplicity, we assume one mixed price to indicate the investment costs. We assume that the chosen action for a MV/LV transformer to increase capacity, which is at most one action per MV/LV transformer, has a cost of €15,000 based on expert knowledge at Coteq. This is in concordance with Grond (2016) and Veldman et al. (2013) who use €15,000 and €12,000, respectively. These actions can be the upgrade of a MV/LV transformer by one from stock, the upgrade of a MV/LV transformer by purchasing a new MV/LV transformer, or the addition of a new MV/LV transformer substation. We approach the problem using the following assumptions.

- Only MV/LV transformers with a capacity of 630 kVA are purchased.
- A MV/LV transformer with a capacity of 400 kVA is only used when it is available from stock.
- MV/LV transformers have a maximum allowed capacity utilization of 120%.
- There is currently no stock of MV/LV transformers.
- The lifetime of MV/LV transformers is sufficient up to 2050.

Algorithm 1: MV/LV transformer expansion routine

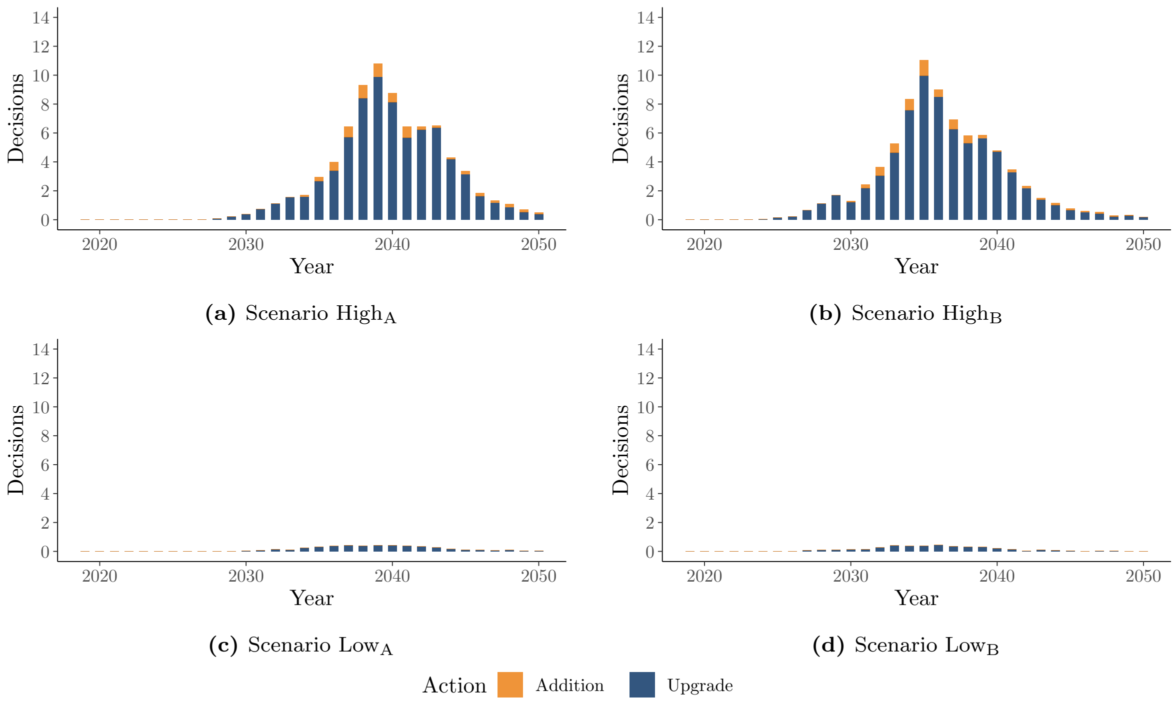
input : $load_{m,t}$ = peak load of MV/LV transformer substation m in year t
$capacity_{m,i}$ = current capacity of MV/LV transformer substation m transformer i
max = maximum allowed capacity utilization

output : $costs_t$ = vector including costs per year t
$upgrade_t$ = vector including the number of MV/LV transformer upgrades per year t
$addition_t$ = vector including the number of MV/LV transformer substation additions per year t

variables : $utilization$ = capacity utilization of a MV/LV transformer substation
$initial$ = initial capacity of a MV/LV transformer
$switch$ = 1 if a 400 kVA MV/LV transformer is taken from stock, 0 otherwise
$decision$ = 1 if there are costs involved for a MV/LV transformer investment decision, 0 otherwise
$entry$ = 1 for a new MV/LV transformer substation, 0 for a MV/LV transformer upgrade
$stock$ = stock of 400 kVA MV/LV transformers

```
foreach year t ∈ (2020, 2021, ..., 2049, 2050) do
    foreach MV/LV transformer substation m do
        utilization = load_{m,t} / Σ_i capacity_{m,i} ;
        if utilization > max then
            utilization = load_{m,2050} / Σ_i capacity_{m,i} ;
            i = 1 ;
            while utilization > max do
                initial = capacity_{m,i} ;
                decision = 0 ;
                entry = 0 ;
                while capacity_{m,i} < 630 and utilization > max do
                    switch = 0 ;
                    decision = 1 ;
                    if capacity_{m,i} = 0 then
                        capacity_{m,i} = 630 ;
                        entry = 1 ;
                    else if 0 > capacity_{m,i} < 400 and stock_{400} > 0 then
                        capacity_{m,i} = 400 ;
                        switch = 1 ;
                    else
                        capacity_{m,i} = 630 ;
                    end
                    utilization = load_{m,t} / Σ_i capacity_{m,i} ;
                end
                if decision = 1 then
                    costs_t = costs_t + 15,000 ;
                    if entry = 1 then
                        addition_t = addition_t + 1;
                    end
                    else
                        upgrade_t = upgrade_t + 1 ;
                    end
                    if initial = 400 then
                        stock = stock + 1;
                    end
                    if switch = 1 then
                        stock = stock - 1;
                    end
                end
                i = i + 1 ;
            end
        end
    end
end
```

Applying the algorithm to the results of the simulation study gives the results shown in Figure 5.12. Figure 5.12 shows that most capacity expansions are achieved by replacing a MV/LV transformer by one that has a larger capacity. In some situations, MV/LV transformer capacity cannot be expanded anymore and a new entry is needed. Figure 5.13 illustrates the costs associated with the capacity expansion of MV/LV transformer substations based on expert knowledge. For scenario Low_A and Low_B, the costs are limited, while the costs for scenarios High_A and High_B are much higher. Figure 5.14 illustrates which MV/LV transformer first need to be replaced based on capacity. Figure 5.14 shows that first the MV/LV transformer having a capacity of 250 kVA will experience bottlenecks followed by 400 kVA.



(a) Scenario High_A

(b) Scenario High_B

(c) Scenario Low_A

(d) Scenario Low_B

Figure 5.12: Investment decisions (upgrade/addition) for scenarios High_B, Low_B, High_A, and Low_A.

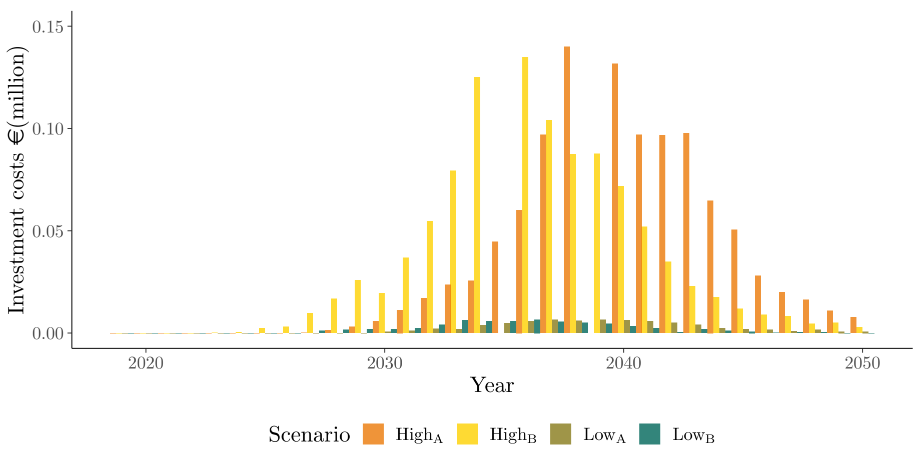


Figure 5.13: Investment costs for capacity expansion MV/LV transformers per scenario.

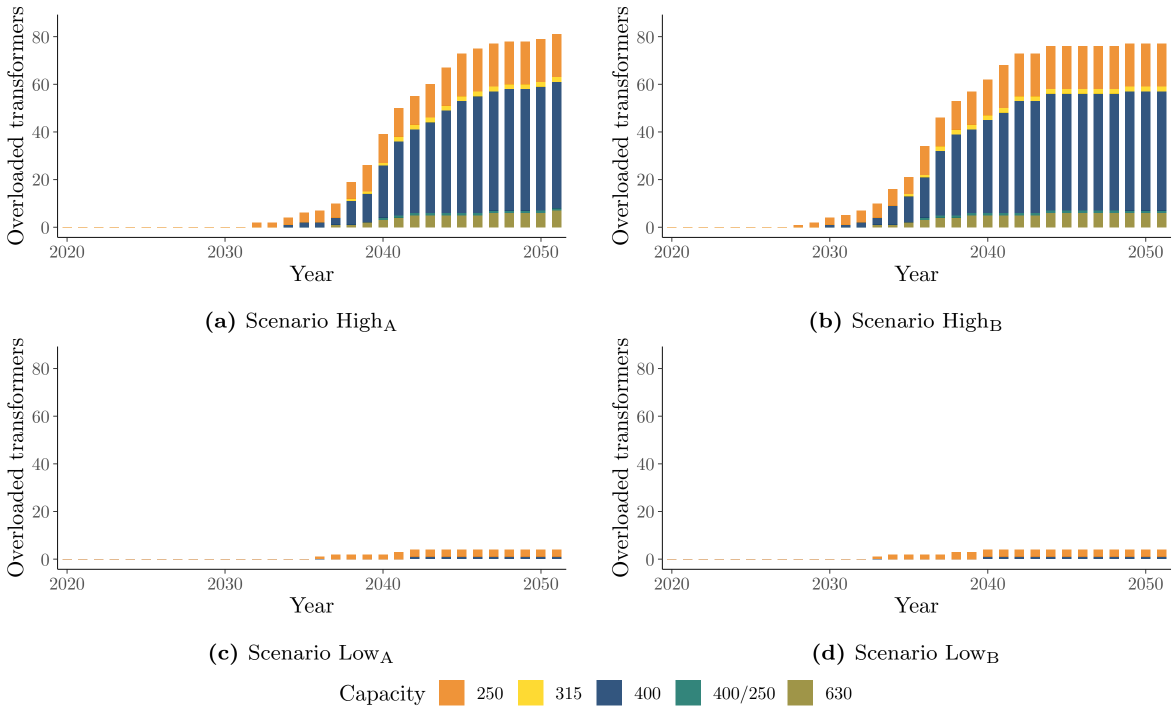


Figure 5.14: Yearly number of overloaded MV/LV transformers grouped by capacity for scenarios High$_B$, Low$_B$, High$_A$, and Low$_A$.

Conclusion

Preventing bottlenecks resulting from HP adoption requires substantial investments for the scenarios with a 85% final market share. Costs can reach up to €1.2 million. Approximately 80 additions/upgrades are needed to cope with the bottlenecks up to 2050. We only considered the investments needed to prevent bottlenecks resulting from heat pumps. A complete bottleneck assessments should include all sustainable technologies to prevent that the capacity resulting from investments related to only heat pump becomes a limiting factor in the future. For the scenarios having a 35% final market share of HPs, the costs are significantly lower and reach up to €75,000 referring to 5 upgrades. Of the MV/LV transformer substations, first the stations with a capacity of 250 kVA become overloaded, followed by stations with 400 kVA.

5.6 Conclusion

We find that the median capacity utilization of MV/LV transformers increases from 43% in the current situation, to 81% in 2050 assuming a final market share of 85%. Assuming a final market share of 35%, the median capacity utilization is 62% in 2050. The investments costs prevent bottlenecks can reach up to €1.2 million. Given our assumptions, Coteq can expect a steep increase in the yearly number of overloaded MV/LV transformer substations in 2030 for the scenarios based on a current market size of 1,400 HPs. For the scenarios based on a current market size of 400 HPs, this steep increase is expected in 2035. The spatial diffusion reduces the pressure on MV/LV transformers having a high capacity utilization and postpones the required investments. The first bottlenecks are expected in neighborhoods BU01411801 and BU01731100.

6 Conclusions and recommendations

This chapter finalizes this research. First, we present the main conclusions of this research. Second, we discuss the results. Third, we discuss the limitations of this research. Fourth, we discuss the contributions to theory and practice. Finally, we present our recommendations for Coteq.

6.1 Conclusions

In this research, we focused on the following main research question:

What is the impact on medium to low voltage transformers due to the spatial diffusion of residential heat pumps?

In this research, we found that the expected number of currently installed HPs in Almelo, Goor, and Oldenzaal is between 400 and 1,400. Unfortunately, no data were available on the locations of these HPs. Using change point detection, we classified 99 households as having a HP. Therefore, there is a discrepancy between what we found using change point detection and what we expected based on the context analysis. Nonetheless, we used logistic regression to extract the influencing factors for residential HP adoption using the results from change point detection. We found that residential HP adoption is associated with residential areas having higher property values, older people and smaller sizes of living area. Also, we found that residential HP adoption is associated with urban areas with 1500-2500 and 500-1000 addresses per km^2. All the independent variables were significant with p-value < 0.05. Using literature, we validated the results from the logistics regression model. We found that the independent variables property value and degree of urbanity are in concordance with literature. Contradictions were found for the independent variables age and size of living.

Using conducted studies on scenario analysis, we constructed four HP growth S-curves representing a low or high final market share of HPs in combination with a current HP market size of 400 or 1,400 HPs. The S-curves include governmental stimulation, HP/energy price developments, and municipal decisions to create momentum. The spatial diffusion was then determined by combining the S-curves with the results from the logistic regression model using a Fisher's noncentral hypergeometric distribution. We found a moderate differentiation between neighborhood HP penetration levels. For the low and high scenarios, HP penetration levels have a range of 0.15–0.50 and 0.65–1.00, respectively. We based the HP power requirements on an empirical study obtained from literature. This study was based on a cold winter weekday with an external temperature of -0.3°C. Also, the study showed that the HP peak load does not coincide with the base residential peak load. To account for this, we estimated that the additional peak load added to the base residential peak load is 75% of the actual HP peak load. The total number of overloaded MV/LV transformer substations for the low and high scenarios is 5 and 80, respectively. We found that the spatial adoption of HPs reduces the pressure on MV/LV transformer substations having a high capacity utilization. This postpones the required investments to prevent bottlenecks.

We expect that bottlenecks already occur in 2025 for the scenario with a current HP market size of 1,400 and a final HP market share of 85%. For the scenario with a current market share of 400 HPs, this is postponed to 2028. To prevent these bottlenecks, mainly upgrades of MV/LV transformers are needed. For the scenario with a HP market share of 85%, approximately 8% of the investments are related to

MV/LV transformer substation additions and 92% to MV/LV transformer upgrades. For the scenario with a HP market share of 35%, 100% of the investments are related to MV/LV transformer upgrades. For the MV/LV transformers substations, it is expected that first the MV/LV transformer substations with 250 kVA MV/LV transformers are affected followed by the MV/LV transformer substations with 400 kVA MV/LV transformers. Estimated investment costs based on expert knowledge for the low and high scenarios were found to be €0.07 million and €1.2 million, respectively.

6.2 Discussion

Identifying influencing factors for residential HP adoption remains a challenging problem because of limited data availability. In this research, we used energy consumption data to infer HP usage. To our best knowledge, our approach for finding the influencing factors of residential HP adoption is not considered before. Unfortunately, the results were found to be inconclusive. The HP is currently in the starting phase of its life cycle which resulted in an extremely small sample of households classified as having installed a HP. As a consequence, the logistic regression model has a poor fit in terms of large confidence intervals for the coefficients. Other studies mainly used a set of rules to distinguish between residential areas or used surveys to extract influencing factors. Modeling the adoption of HPs, or sustainable technologies in general, is done in various ways. Most studies constructed S-curves for each residential area individually. This was not feasible for our research as not enough data on the development per neighborhood were available. Also, the use of a diffusion model for such small residential areas is often not justified as an area with only 50 households cannot be considered a social system. Therefore, our research discusses a statistical approach. The use of an empirical discrete distribution or a Fisher's non-central hypergeometric distribution was not considered before to the best of our knowledge. We found that the use of an empirical distribution is an effective but relatively slow approach. We found that using a Fisher's non-central hypergeometric distribution gave similar results while being much faster.

6.3 Limitations

This research contains several limitations. First, HP requirements expressed in kW were based on empirical results from a study performed in Great Britain. Ideally, data on HP load profiles in Almelo, Goor, and Oldenzaal should be used to extract HP requirements. Second, an important input for this research was the use of energy consumption data. The validation of this approach is limited due to time management. Literature however partly confirmed the results of the logistic regression model. Third, we found a small number of households that corresponded to our classification rules. It is possible that we captured, for example, pellet stove users. When the sales of HPs keep rising while the sales of other gasless techniques remain constant or decrease, we can assume that the reliability increases as the probability of misclassifying HP decreases. Fourth, we assumed that all households in a neighborhood have identical characteristics. Results for neighborhoods might not transfer to individuals. Fifth, the investment decisions are based on only the impact of HPs. Ideally, the load profiles of multiple sustainable technologies should be combined to obtain the peak load on assets. The actual electricity network requirements are most likely larger considering also, for example, EV. This should be incorporated to prevent that the capacity becomes a limiting factor in the future. Last, we did not consider technological development, smart grid approaches, or infrastructure change. The outcome of this research should thus be used keeping these limitations in mind.

6.4 Contributions

Theory

Our research contributes to theory by introducing additional approaches to model the spatial diffusion of sustainable technologies. Also, the use of non-smart meter energy consumption data to infer HP usage was not considered before. Other researchers can use our results to further study the use of energy consumption data to infer HP usage. Given the results, we suggest that researchers focus on the use of smart meter data as our findings illustrate the many challenges of using non-smart meter data. Our comparison of the empirical discrete distribution and the Fisher's non-central hypergeometric distribution expands the choices researchers have in choosing suitable approaches to combine statistical analysis with a diffusion model such as the Bass model. The Fisher's noncentral hypergeometric distribution, however, makes use of several categories that each requires odds. In this research, we applied the Fisher's noncentral hypergeometric distribution to 94 neighborhoods. Additional categories were not considered and thus the performance for multiple categories should be researched. Also, when the characteristics of individual households are researched, the Fisher's noncentral hypergeometric distribution might be less suitable because this would require an odds for every household.

Practice

For Coteq, our research contributes to an increased understanding of the energy transition impact on the electricity network. Although the results of the logistic regression model were found to be inconclusive, it still shows the possibilities of using energy consumption data in the energy transition. Also, Coteq can obtain a fast bottleneck assessment on MV/LV transformer substations caused by the growth of residential HPs and a capacity expansion indication to prevent these bottlenecks. Further research on the exact costs is needed to obtain an improved cost indication. Finally, Coteq can use this research to study the impact of HPs on the gas network.

6.5 Recommendations

In this section, we present our recommendations based on our research.

- In this research, we utilized energy consumption data to infer HP usage. Using change point analysis we were able to classify a small number of households as having HPs. Assuming that HP growth follows an S-curve, the technology is only in its beginning phase. It is worthwhile to redo the analysis in the coming years as more HPs are expected. When more HPs are installed, there can occur a more reliable diversified growth of HPs over the region. If so, a grouping would not be needed anymore as the penetration levels of HPs are more certain.

- The S-curves are based on the current knowledge of energy transition scenarios in combination with diffusion models. Data from the CBS show that a steep increase in HPs is expected. Depending on the ongoing developments within the energy transition and this steep increase of HPs, the S-curves might need adaption. It is advised to monitor these developments using data from the CBS, DHPA, EHPA, and RVO. Ideally, more information can be shared on these developments. An important initiative for this purpose is the project by VIVET (Dutch: Verbetering van de Informatievoorziening Voor de Energie Transitie). Multiple stakeholders urge to increase the amount of information to be shared to improve the facilitation of the energy transition.

- Though considered in this research on neighborhood level, it can be worthwhile to study the individual building characteristics in-depth. A connection between the data from the C-AR and the BAG is required for this purpose. This way, instead of assigning a median value of construction year to each household in a neighborhood, individual values can be assigned which increases the accuracy of the model. But as stated in our contributions to theory, the Fisher's noncentral hypergeometric distribution might be less suitable for this purpose.

- We validated the logistic regression model using literature. Although literature confirmed the largest part of our logistic regression model, it is worthwhile to do a separate survey on the influencing factors of HP adoption in Almelo, Goor, and Oldenzaal. Such a survey can provide insight into the unique influencing factors of HP adoption in Almelo, Goor, and Oldenzaal.

- Coteq is advised to more closely monitor the cancellations of gas connections. Ideally, the reason for the cancellations should be registered. The possibility to do this depends on privacy considerations. Cancellations can only be used to monitor the adoption of full-electric HPs. For this reason, as more smart meters are deployed, it is advised to use smart meter data as soon as it is allowed to monitor these data. When this becomes available, we recommend consulting the master thesis by Francica (2019). The author studied machine learning approaches to detect the use of HPs based on load profiles and temperature readings.

- The proposed simulation model is generalizable, meaning that it can be applied to other sustainable technologies. To determine the impact of, for example, EV, data on the influencing factors of EV adoption and EV electrical requirements are needed. This simulation model can be a first step in determining the local impact on the electricity network of Coteq including sustainable technologies. For a complete calculation of the impact, it is advised to study and combine the individual load profiles of sustainable technologies as the peaks of each type of sustainable technology can occur on different parts of the day.

- The simulation model is based on R software. The code developed for the simulation model can be incorporated in data visualization software used at Coteq. This enables a direct connection with the databases of Coteq and thus direct bottleneck assessments.

References

Ahkamiraad, A., & Wang, Y. (2018). An agent-based model for zip-code level diffusion of electric vehicles and electricity consumption in New York City. *Energies*. doi: 10.3390/en11030640

Akmal, M., Fox, B., Morrow, J. D., & Littler, T. (2014). Impact of heat pump load on distribution networks. *IET Generation, Transmission and Distribution*. doi: 10.1049/iet-gtd.2014.0056

Ameli, N., & Brandt, N. (2015). Determinants of households' investment in energy efficiency and renewables: Evidence from the OECD survey on household environmental behaviour and attitudes. *Environmental Research Letters*. doi: 10.1088/1748-9326/10/4/044015

Austin, P. C., & Steyerberg, E. W. (2017). Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Statistical Methods in Medical Research*. doi: 10.1177/0962280214558972

Autoriteit Consument & Markt. (2016). *Regelgeving.* Retrieved from `https://www.acm.nl/nl/publicaties/publicatie/14457/Informatiecode-Elektriciteit-en-Gas`

Autoriteit Consument & Markt. (2020). *Gasregister.* Retrieved from `https://www.acm.nl/nl/onderwerpen/energie/netbeheerders/gasaansluitplicht/gasregister`

Barteczko-Hibbert, C. (2015). *After Diversity Maximum Demand (ADMD) report.* Durham University: Customer Led Network Revolution.

Bass, F. M. (1969). A New Product Growth for Model Consumer Durables. *Management Science*. doi: 10.1287/mnsc.15.5.215

Berenschot. (2017). *Routekaart hybride warmtepomp.* Retrieved from `https://www.berenschot.nl/actueel/2017/maart/hybride-warmtepomp-maakt/`

Berenschot. (2018a). *Elektronen en/of Moleculen.* Retrieved from `https://www.berenschot.nl/actueel/2018/april/elektronen-moleculen-transitie/`

Berenschot. (2018b). *Het warmtescenario: Beelden van een op warmte gerichte energievoorziening in 2030 en 2050.* Retrieved from `https://www.berenschot.nl/actueel/2018/september/nieuw-warmtescenario/`

Bernards, R., Morren, J., & Slootweg, H. (2016). Evaluating impact of new technologies on low voltage grids using probabilistic data enriched scenarios. *EEEIC 2016 - International Conference on Environment and Electrical Engineering*. doi: 10.1109/EEEIC.2016.7555570

Bernards, R., Morren, J., & Slootweg, H. (2018). Development and Implementation of Statistical Models for Estimating Diversified Adoption of Energy Transition Technologies. *IEEE Transactions on Sustainable Energy*. doi: 10.1109/TSTE.2018.2794579

CE Delft. (2017). *Net voor de toekomst.* Retrieved from `https://www.netbeheernederland.nl/nieuws/resultaten-onderzoek-net-voor-de-toekomst-1204`

Centraal Bureau voor de Statistiek. (2019). *Kerncijfers wijken en buurten 2019.* Retrieved from https://www.cbs.nl/nl-nl/maatwerk/2019/31/kerncijfers-wijken-en-buurten-2019

Centraal Bureau voor de Statistiek. (2020). *Warmtepompen; aantallen, thermisch vermogen en energiestromen (82380NED).* Retrieved from https://opendata.cbs.nl/statline

Dilekli, N., Janitz, A. E., Campbell, J. E., & de Beurs, K. M. (2018). Evaluation of geoimputation strategies in a large case study. *International Journal of Health Geographics*. doi: 10.1186/s12942-018-0151-y

Dong, C., Sigrin, B., & Brinkman, G. (2017). Forecasting residential solar photovoltaic deployment in California. *Technological Forecasting and Social Change*. doi: 10.1016/j.techfore.2016.11.021

Dutch Heat Pump Association. (2020). *Nationaal Warmtepomp Trendrapport 2020.* Retrieved from https://www.duurzaamverwarmd.nl/trendrapport/

Ecofys. (2017). *Overige hernieuwbare energie in Nederland.* Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2017/10/19/overige-hernieuwbare-energie-in-nederland

Eising, J. W., van Onna, T., & Alkemade, F. (2014). Towards smart grids: Identifying the risks that arise from the integration of energy and transport supply chains. *Applied Energy*. doi: 10.1016/j.apenergy.2013.12.017

European Heat Pump Association. (2020). *Heat pump sales overview.* Retrieved from https://www.ehpa.org/market-data/

Fisher, J. C., & Pry, R. H. (1971). A simple substitution model of technological change. *Technological Forecasting and Social Change*. doi: 10.1016/S0040-1625(71)80005-7

Francica, R. (2019). *Assessment of machine learning algorithms for the purpose of heat pump detection based on load profiles and temperature readings.* Retrieved from https://repository.tudelft.nl/islandora/object/uuid%3A8b1d99a0-eda0-4efc-b519-04ae1e695079

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

Grond, M. O. W. (2016). *Computational Capacity Planning in Medium Voltage Distribution Networks* (Doctoral dissertation, Eindhoven University of Technology). Retrieved from https://research.tue.nl/en/publications/computational-capacity-planning-in-medium-voltage-distribution-ne

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: data mining, inference, and prediction.* New York: Springer.

Het Nationaal Programma Regionale Energie Strategieën. (2019). *Regionale Energie Stragie.* Retrieved from https://www.regionale-energiestrategie.nl/default.aspx

Hibbert, J. D., Liese, A. D., Lawson, A., Porter, D. E., Puett, R. C., Standiford, D., ... Dabelea, D. (2009). Evaluating geographic imputation approaches for zip code level data: An application to a study of pediatric diabetes. *International Journal of Health Geographics*. doi: 10.1186/1476-072X-8-54

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R.* New York: Springer.

Karytsas, S. (2018). An empirical analysis on awareness and intention adoption of residential ground source heat pump systems in Greece. *Energy Policy*. doi: 10.1016/j.enpol.2018.08.001

Karytsas, S., Polyzou, O., & Karytsas, C. (2019). Factors affecting willingness to adopt and willingness to pay for a residential hybrid system that provides heating/cooling and domestic hot water. *Renewable Energy*. doi: 10.1016/j.renene.2019.04.108

Karytsas, S., & Theodoropoulou, H. (2014). *Public awareness and willingness to adopt ground source heat pumps for domestic heating and cooling.* doi: 10.1016/j.rser.2014.02.008

Kiesling, E., Günther, M., Stummer, C., & Wakolbinger, L. M. (2012). Agent-based simulation of innovation diffusion: A review. *Central European Journal of Operations Research*. doi: 10.1007/s10100-011-0210-y

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling.* doi: 10.1007/978-1-4614-6849-3

Law, A. M. (2015). *Simulation Modeling and Analysis.* New York: McGraw-Hill.

Litjens, G. (2018). *Improving local use of electricity generated by rooftop photovoltaic systems* (Doctoral dissertation, Utrecht University). Retrieved from `https://dspace.library.uu.nl/handle/1874/373082`

Loertscher, S., Muir, E. V., & Taylor, P. G. (2017). A general non-central hypergeometric distribution. *Communications in Statistics - Theory and Methods*. doi: 10.1080/03610926.2015.1088033

Love, J., Smith, A. Z., Watson, S., Oikonomou, E., Summerfield, A., Gleeson, C., ... Lowe, R. (2017). The addition of heat pump electricity load profiles to GB electricity demand: Evidence from a heat pump field trial. *Applied Energy*. doi: 10.1016/j.apenergy.2017.07.026

Michelsen, C. C., & Madlener, R. (2012). Homeowners' preferences for adopting innovative residential heating systems: A discrete choice analysis for Germany. *Energy Economics*. doi: 10.1016/j.eneco.2012.06.009

Miller, D. (2015). *Developing the smarter grid.* Retrieved from `http://www.networkrevolution.co.uk/wp-content/uploads/2015/04/OSR-Final.pdf`

MIT Critical Data. (2016). *Secondary Analysis of Electronic Health Records.* Cambridge: Springer. doi: 10.1007/978-3-319-43742-2

Moraga, J. L., & Mulder, M. (2018). *Electrification of heating and transport.* Retrieved from `https://www.rug.nl/news/2018/nieuwsberichten/ceer-policy-paper-electrification-in-heating-and-transport.pdf`

Natuur & Milieu. (2019). *Gasmonitor 2019.* Retrieved from https://www.natuurenmilieu.nl/wp-content/uploads/2019/08/Gasmonitor-2019.pdf

Netbeheer Nederland. (2019a). *Kerngegevens energienetten.* Retrieved from https://energiecijfers.info/hoofdstuk-1/

Netbeheer Nederland. (2019b). *Regulering.* Retrieved from https://www.netbeheernederland.nl/dossiers/regulering-20

Netbeheer Nederland. (2019c). *Transitievisie warmte.* Retrieved from https://www.netbeheernederland.nl/transitievisie-warmte

Obinna, U., Joore, P., Wauben, L., & Reinders, A. (2017). Comparison of two residential Smart Grid pilots in the Netherlands and in the USA, focusing on energy performance and user experiences. *Applied Energy*. doi: 10.1016/j.apenergy.2017.01.086

Rijksoverheid. (2019a). *Duurzame energie.* Retrieved from https://www.rijksoverheid.nl/onderwerpen/duurzame-energie

Rijksoverheid. (2019b). *Warmtepomp.* Retrieved from https://www.rvo.nl/onderwerpen/duurzaam-ondernemen/duurzame-energie-opwekken/verduurzaming-warmtevoorziening/warmtepompen

Rijksoverheid voor Ondernemend Nederland. (2020a). *Investeringssubsidie duurzame energie (ISDE).* Retrieved from https://www.rvo.nl/subsidie-en-financieringswijzer/investeringssubsidie-duurzame-energie-isde

Rijksoverheid voor Ondernemend Nederland. (2020b). *Investeringssubsidie duurzame energie (ISDE) verandert in 2020.* Retrieved from https://www.rvo.nl/actueel/nieuws/investeringssubsidie-duurzame-energie-isde-verandert-2020

Rogers, E. M. (1983). *Diffusion of innovations-Third Edition.* doi: citeulike-article-id:126680

Ross, G. J. (2015). Parametric and nonparametric sequential change detection in R: The cpm package. *Journal of Statistical Software*. doi: 10.18637/jss.v066.i03

Saarenpää, J., Kolehmainen, M., & Niska, H. (2013). Geodemographic analysis and estimation of early plug-in hybrid electric vehicle adoption. *Applied Energy*. doi: 10.1016/j.apenergy.2013.02.066

Universiteit Utrecht. (2015). *Warmtepompen in de bestaande bouw in Nederland.* Retrieved from https://www.bdho.nl/wp-content/uploads/2015/06/20150322-Rapport_STEM-3.pdf

Van De Sande, P., Danes, M., & Dekker, T. (2017). ANDES: Grid capacity planning using a bottom-up, profile-based load forecasting approach. *CIRED - Open Access Proceedings Journal*. doi: 10.1049/oap-cired.2017.1071

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*. doi: 10.1016/j.erss.2018.06.003

Van Horssen, P. W., Pebesma, E. J., & Schot, P. P. (2002). Uncertainties in spatially aggregated predictions from a logistic regression model. *Ecological Modelling*. doi: 10.1016/S0304-3800(02)00060-1

Van Leeuwen, R. P., de Wit, J. B., & Smit, G. J. (2017). Review of urban energy transition in the Netherlands and the role of smart energy management. *Energy Conversion and Management*. doi: 10.1016/j.enconman.2017.05.081

Van Oirsouw, P. (2012). *Netten voor distributie van elektriciteit.* Arnhem: Phase to phase.

Veldman, E., Gibescu, M., Slootweg, H., & Kling, W. L. (2011). Impact of electrification of residential heating on loading of distribution networks. In *2011 ieee pes trondheim powertech: The power of technology for a sustainable society, powertech 2011.* doi: 10.1109/PTC.2011.6019179

Veldman, E., Gibescu, M., Slootweg, H. J., & Kling, W. L. (2013). Scenario-based modelling of future residential electricity demands and assessing their impact on distribution grids. *Energy Policy*. doi: 10.1016/j.enpol.2012.12.078

Vringer, K., & Carabain, C. L. (2020). Measuring the legitimacy of energy transition policy in the Netherlands. *Energy Policy*. doi: 10.1016/j.enpol.2019.111229

A Municipal involvement

In this research, we assume that the municipal plans play a less important role in the increase of residential HPs. We acknowledge that a situation in which individual decisions determine for 100% the increase of HPs is not likely to happen. Municipal plans might still be needed to create momentum in the adoption of HPs. For a complete understanding of the problem cluster introduced in Chapter 1, it is thus of importance to understand the involvement of municipalities.

Regionale Energie Strategie

Achieving the national goal on the energy transition requires cooperation between multiple stakeholders (Het Nationaal Programma Regionale Energie Strategieën, 2019). For this reason, it was decided in 2018 to divide the national approach over the 30 regions shown in Figure A.1. Each region develops a Regionale Energie Strategie (RES) with stakeholders such as businesses, inhabitants, authorities, and social partners. The RES provides insight into the spatial integration of the energy transition by considering issues such as available heat sources, consequences for infrastructure, and regional electricity generation. It is a product that describes per region what energy goals will be achieved and in what time frame. Also, it specifies the strategy to determine and achieve these goals.

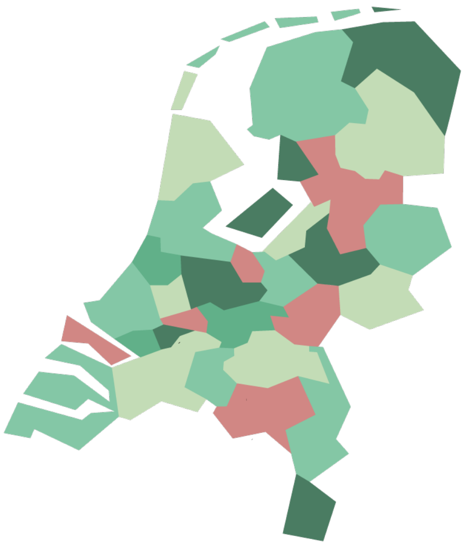


Figure A.1: Division of Regionale Energie Strategie (RES) regions in the Netherlands.

Regionale Structuur Warmte

An element of the RES is the so-called Regionale Structuur Warmte (RSW). This document provides insight into the supply and demand for heat in the region as well as the possibilities for new heat infrastructures that can be developed. Provinces, municipalities, and regional water authorities do this in cooperation with heat companies and DSOs.

Transitievisie Warmte

The Transitievisie Warmte (TVW) is a document that describes for a municipality the plans for insulation and phase-out of natural gas for the built environment up to 2030. Each municipality must complete their TVW by the end of 2021. Alternatives for natural gas must be known for districts that will be free of natural gas by the end of 2030. In short, the TVW contains for a municipality the alternatives for natural gas per district and a time plan for making districts free of natural gas.

Wijkuitvoeringsplan

The Wijkuitvoeringsplan (WUP) is a plan per district that contains how the phase-out of natural gas per district is achieved. It specifies when the change will take place and what alternative for natural gas will be used. This document is the basis on which investment decisions are made by DSOs, building owners, heat companies, providers of heat alternatives, and other stakeholders.

Relation between products

The RSW is part of the RES and is mostly concerned with identifying the supply and demand for heat in the region. The TVW is more detailed and is created by municipalities. The RSW is checked against the TVW of multiple municipalities to make sure that heat sources are used properly. Multiple iterations will be needed to come to the efficient use of heat sources. Finally, the WUP is a detailed product that contains the final plan for realizing the phase-out of natural gas.

B Structure of Dutch residential areas

The Centraal Bureau voor de Statistiek (CBS) publishes data on energy, social security, the proximity of facilities, land use, crime, and motor vehicles (Centraal Bureau voor de Statistiek, 2019). These data are published on multiple aggregation levels. Within these aggregation levels, one can distinguish two spatial distributions. The first distribution considers the regional levels municipality, district (part of a municipality), and neighborhood (part of a district). A district is defined as an area in which certain land use or type of building appears often. A neighborhood is defined as an area which is, because of socio-economic structure or building point of view, demarcated as homogeneous. The second distribution considers postcode levels ranging from 4-digit up to 6-digit. Table B.1 gives an overview of the degree of granularity for each level within each distribution. Comparing tables B.1a and B.1b shows that the second method is more granular than the first distribution. Figure B.1 shows the degrees of granularity of the regional distribution to get a visual idea of the level of detail available. Where the granularity increases for figures B.1a, B.1b, and B.1c. All data sets, except the 5 and 6-digit postcode data sets, are freely available.

Table B.1: Illustration of the degree of granularity for each distribution.

(a) Regional

Granularity	Entries
Municipal	355
District	3,160
Neighborhood	13,517

(b) Postcode

Granularity	Entries
PC4	4,067
PC5	33,253
PC6	460,289

(a) Municipalities **(b)** Districts **(c)** Neighborhoods

Figure B.1: Illustration of the granularity of CBS data for three regional levels.

C Data pre-processing

In this appendix, we discuss the data pre-processing steps that are needed in order to obtain usable data sets for the approximate classification and logistic regression analysis. First, we discuss the data pre-processing steps needed on the energy consumption data. Second, we discuss the data pre-processing steps needed on the data from the CBS and BAG.

C.1 Energy consumption data

To perform the approximate classification, we first pre-process the raw data. We use data from the C-AR (Centraal Aansluit Register), which is a database that contains information on gas and electricity connections. This data set is used to identify connections that possibly installed a HP. To do this, we need to know what the energy consumption is over time per connection as illustrated in Table C.1. The energy consumption is measured in normalized annual consumption (NAC), which is the annual energy consumption accounting for seasonal differences. In addition, we need to know the status of each connection in terms of the activity as illustrated in Table C.2.

Table C.1: Illustration of desired data for change point detection.

Connection	Product	Date	NAC
A	Electricity	01-01-2014	2000
A	Electricity	01-06-2014	2500
A	Gas	01-01-2014	1500
A	Gas	01-06-2014	1300

Table C.2: Illustration of desired data set for detecting households with full-electric HP.

Connection	Product	Delivery status	Physical status	Reason out of business
A	Electricity	Active	In business	
A	Gas	Inactive	Disconnected	Consumer request

Data description

Coteq collects information on the energy generation and consumption of connections, where energy generation is only possible for households having installed PV. Data is available from 2012 because this year the C-AR was used for the first time. The C-AR is a central database for information on energy connections for DSOs. The energy consumption and generation data are based on meter readings at the connections. It should be noted that the self-consumption of electricity for households with PV is unknown. The DSO only knows the energy consumption actually used that was not generated by the household itself. Consumption data is either expressed in the actual amount of energy used or in normalized annual consumption (NAC). DSOs translate the actual consumption over a certain period to a NAC by using so-called profile fractions (Autoriteit Consument & Markt, 2016). Profile fractions account for the differences in energy consumption over time. Actual consumption data within a certain time period is divided by the profile fractions belonging to this time period to obtain a NAC. It holds that the sum of the profile fractions equals 1. The profile fractions for electricity contain the differences in electricity consumption depending on the time of the year. The profile fractions for gas additionally take into account the temperature by multiplying it with a temperature coefficient. This way it is possible to compare the NAC over several years since the values are translated to a normal year with normal temperatures. Following the rules set by the ACM, a standardized yearly value of electricity is calculated based on the smallest consumption time series with at least 300 days. The same holds for gas, but now the months January and February need to be included in the 300 days since these are the most important months for gas consumption.

Data quality

Meter readings can be done multiple times per year, but are done at a minimum of once a year if the household is inhabited. It should be noted that the quality of the meter readings differs across the years. The upswing of smart meters has increased the reliability of data since actual energy consumption is directly registered. In some situations, values have to be approximated if the time period of energy consumption is too small or if no energy consumption is known (Autoriteit Consument & Markt, 2016). An approximation is done by, for example, taking the average of comparable connections. The connections contained in the data set are large and small consumers. Most small consumers are households while it is possible for a small consumer to be a small firm. Unfortunately, this cannot be deduced from the data set from the C-AR. Data from the BAG is therefore used as the BAG contains for each building the function of the building (e.g., resident). In addition, to prevent possible mismatches between the BAG and the C-AR, we only focus on connections in the C-AR with the following physical capacities: gas connections with a physical capacity of G4 and G6 and electricity connections with a physical capacity of 1x25, 1x35 or 3x25. For gas connections, these physical capacities indicate the amount of gas that can be delivered in m^3 while for electricity it indicates the amperage and thus the power that the households can use. Also, households connected to district heating are excluded from the analysis since these are likely to use district heating in the future and might distort the analysis: a gas connection removal might be requested for a house connected to district heating which might suggest HP usage. The district where most registered district heating connections are located is Windmolenbroek (Autoriteit Consument & Markt, 2020). We exclude these connections based on 6-digit postcodes published by the ACM.

Privacy considerations

To comply with the Algemene verordening gegevensbescherming (AVG), we code all privacy-sensitive information. A part of this research is to find the fraction of households that have installed a HPs in a residential area. As we are not interested in individual households, we also group the found HPs. This completely eliminates the use of privacy-sensitive information.

C.2 Socio-demographic and building data

Pre-selection

The CBS data set consists of 102 variables related to the categories energy, social security, the proximity of facilities, land use, crime, and motor vehicles. The CBS publishes these data each year. The most recent publication does not contain all data. For example, data on housing stock is available in the 2018 publication but not in the 2019 publication. Therefore, we compile a final data set by combining the publications over the three most recent years. Since no literature was found on the determinants of hybrid HP adoption, we assume that the determinants of adoption for full-electric HPs are applicable. From the literature review on influencing factors of HP adoption in Section 2.2, we find that gender, age, urbanity, income, education, household occupation, ownership, construction year, and size of living area are important variables. In the context analysis, we concluded that the type of house is also important. HPs are less suited for multiple-family dwellings than for single-family dwellings. In addition, existing housing stock often has to be renovated to be able to adopt HPs. These characteristics will be measured by using variables on the type of dwelling and construction year. From the CBS, we select of the category population: number of inhabitants, gender, age, household composition (single person, family, average household size), average household size, and population density. Of the category living we select: number of dwellings, property value, type of dwelling (single-family & multiple-family), ownership (owner-occupied, rental, housing corporation), and construction year. From the category energy, we select the percentage of households connected to district heating. From the category income, we select the percentage of low/high income. Finally, from the urbanity category, we select the degree of urbanity. The BAG only contains two variables that were found to be important: construction year and size of living area. To comply with the structure of the CBS, we take the median of the size of living area per neighborhood and the median of construction year per neighborhood. The median was chosen as it is less sensitive to outliers. The final set of variables, including their explanation, is shown in Table C.3. The descriptive statistics of the data set are shown in tables C.4 and C.5. From the descriptive statistics, we can already see that the variable *DistrictHeating* has too many zero entries to be valuable for our analysis. For this reason, we do not use this variable in the logistic regression analysis.

Completeness

We now determine the completeness of the variables listed in Table C.3. Table C.6 shows the patterns of missing data. The first row shows the variables that are missing in some of the patterns. The left column indicates the frequency in which the pattern of that row occurs. The bottom row indicates per variable how often it is missing. The right column indicates how many variables are missing in each pattern. Within the table, a 0 indicates that the variable is missing. We mostly observe missing values related to financial variables. When the number of houses in a region is less than 50, no information is given on property value (WOZ, Dutch: Waardering Onroerende Zaken). For less than 20 houses, no information is given on the type of property. Average income is only available for regions with at least 2,500 income recipients. Because the data set is small, we want to utilize each neighborhood in the AGO region. As we do not remove rows, we need to find ways to still use the data given the missing entries. Geographic imputation is used to overcome the problem of missing spatial data (Hibbert et al., 2009). A lot of studies try to impute missing data to individual person level, while we are trying to impute data on neighborhood level. Hibbert et al. (2009) and Dilekli, Janitz, Campbell, and de Beurs (2018) both use some sort of weighted metric based on distance, population, or another measure to impute the value. Based on these studies, we replace the missing values by taking the weighted average of neighboring polygons. For example, Figure C.1 shows the spatial dispersion of the variable *PropertyValue*. Visually, we can observe a spatial correlation between the variable *PropertyValue* and residential area. Properties in city centres tend to be cheaper than properties in outer regions. By averaging the values of neighboring neighborhoods we still maintain this spatial correlation.

Table C.3: Description of the pre-selected variables based on literature.

Variable	Description	Source
PercMen	Percentage of men	CBS
PercWomen	Percentage of women	CBS
Perc0014	Percentage of people in the age category 0-14	CBS
Perc1524	Percentage of people in the age category 15-24	CBS
Perc2544	Percentage of people in the age category 25-44	CBS
Perc4564	Percentage of people in the age category 45-64	CBS
Perc65oo	Percentage of people in the age category 64+	CBS
PercSinglePerson	Percentage of single-person households	CBS
PercHouseholdNoChildren	Percentage of households without children consisting of married/unmarried couples and other households	CBS
PercHouseholdChildren	Percentage of households with children consisting of married/unmarried couples and single-parent households	CBS
AverageHouseholdSize	Average household size calculated as the number of persons living in a household divided by the number of households	CBS
PopulationDensity	Population density expressed as the number of inhabitants per km^2	CBS
PropertyValue	Property value for objects described as dwellings for main residence and dwellings with practice rooms	CBS
PercSingleFamily	Percentage of single-family dwellings that is also an entire building	CBS
PercMultipleFamily	Percentage of multiple-family dwellings that share with multiple dwellings one building	CBS
PercOwnerOccupied	Percentage of owner-occupied dwellings	CBS
PercRental	Percentage of rental dwellings not inhabited by the owner	CBS
PercDistrictHeating	Percentage of dwellings connected to district heating	CBS
PercLowestIncome	Percentage of households belonging to the national 40 percent of households with the lowest household incomes	CBS
PercHighestIncome	Percentage of households belonging to the national 20 percent of households with the highest household incomes	CBS
DegreeOfUrbanity	Degree of urbanity ranging from 1 (>2500 addresses per km^2) to 5 (<500 addresses per km^2)	CBS
SizeLivingArea	Median of size of living area of dwellings per neighborhood	BAG
ConstructionYear	Median of construction year of dwellings per neighborhood	BAG

Table C.4: Descriptive statistics of pre-selected continuous variables.

	Q0	Q25	Q50	Q75	Q100	Mean	SD	Zero	NA
PercMen	43.0	49.0	50.0	52.0	67.0	50.7	3.4	0	0
PercWomen	33.0	48.0	50.0	51.0	57.0	49.3	3.4	0	0
Perc0014	0.0	14.0	16.0	18.0	40.0	15.5	5.5	3	0
Perc1524	0.0	10.0	12.0	13.0	50.0	12.3	5.4	2	0
Perc2544	0.0	18.0	22.0	26.0	40.0	22.3	6.6	1	0
Perc4564	17.0	26.0	28.0	32.8	50.0	29.3	5.3	0	0
Perc65oo	0.0	14.0	20.0	24.8	53.0	20.6	9.6	2	0
PercSinglePerson	0.0	21.0	33.0	41.0	62.0	31.9	14.4	3	0
PercHouseholdNoChildren	0.0	25.0	29.5	34.0	60.0	30.1	8.5	1	0
PercHouseholdChildren	9.0	30.2	36.0	44.8	100.0	38.2	14.2	0	0
AverageHouseholdSize	1.5	2.0	2.3	2.6	3.8	2.3	0.4	0	0
PopulationDensity	10.0	845.8	3358.0	5227.5	8754.0	3377.5	2483.1	0	0
PropertyValue	83.0	122.0	177.0	229.0	389.0	190.4	79.9	0	7
PercSingleFamily	14.0	66.0	83.0	96.8	100.0	78.1	20.8	0	4
PercMultipleFamily	0.0	3.2	17.0	34.0	86.0	21.9	20.8	11	4
PercOwnerOccupied	9.0	43.5	64.0	79.0	98.0	60.0	24.5	0	4
PercRental	2.0	20.0	34.5	56.5	91.0	39.0	24.7	0	4
PercHousingCorporation	0.0	2.2	23.0	44.0	89.0	28.0	26.0	22	4
PercDistrictHeating	0.0	0.0	0.0	0.0	78.4	1.0	8.3	92	0
PercLowestIncome	7.1	25.0	42.1	57.9	79.6	40.7	20.2	0	11
PercHighestIncome	1.4	5.5	12.3	28.2	46.5	17.1	13.3	0	11
ConstructionYear	1918.0	1964.0	1982.0	1994.0	2014.0	1977.2	23.1	0	1
SizeOfLivingArea	49.0	77.0	102.0	142.0	341.0	116.2	55.6	0	1

Table C.5: Descriptive statistics of pre-selected categorical variable.

	1	2	3	4	5
DegreeOfUrbanity	8	28	25	21	12

Table C.6: Missing data patterns of pre-selected variables.

	ConstructionYear	SizeOfLivingArea	PercSingleFamily	PercMultipleFamily	PercOwnerOccupied	PercRental	PercHousingCorporation	PropertyValue	PercLowestIncome	PercHighestIncome	
82	1	1	1	1	1	1	1	1	1	1	**0**
4	1	1	1	1	1	1	1	1	0	0	**2**
3	1	1	1	1	1	1	1	0	0	0	**3**
4	1	1	0	0	0	0	0	0	0	0	**8**
1	0	0	1	1	1	1	1	1	1	1	**2**
	1	**1**	**4**	**4**	**4**	**4**	**4**	**7**	**11**	**11**	**51**

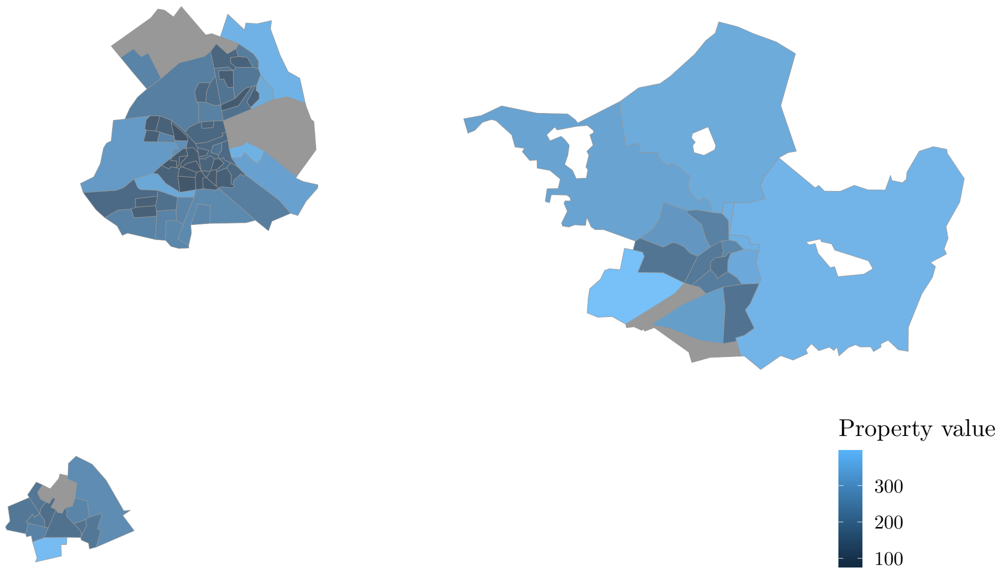


Figure C.1: Spatial correlation of variable *PropertyValue*. Left to right: Goor, Almelo, Oldenzaal.

Collinearity

Collinearity means that two variables are closely related to each other. Figure C.2 illustrates this using a correlation plot. There is a considerable correlation between the variables. There are also some aliased variables, meaning that the value of one variable is fully dependent on the other. This is logical: the percentage of men is fully dependent on the percentage of women. Using a correlation plot, we can only observe correlations between two variables and no multicollinearity. Multicollinearity can occur between more than two variables although no correlation was found between pairs of variables (James et al., 2013). When too much correlation is present in regression, it can be difficult to observe the individual effects of variables on the outcome. It reduces the accuracy of the estimates estimated by regression and thus affects the hypothesis test. To prevent this, we use the variance inflation factor (VIF). The VIF is the ratio between the variance of an estimate $\hat{\beta}_j$ when fitting a full model divided by the variance of estimate $\hat{\beta}_j$ when fitted on its own. A VIF of 1 illustrates no multicollinearity. We use a VIF of 5 as a rule of thumb mentioned by James et al. (2013). The remaining variables that comply with this rule of thumb are *PercMen*, *Perc1524*, *Perc2544*, *Perc4564*, *PopulationDensity*, *PropertyValue*, *PercMultipleFamily*, *PercOwnerOccupied*, *DegreeOfUrbanity*, *SizeLivingArea*, and *ConstructionYear*.

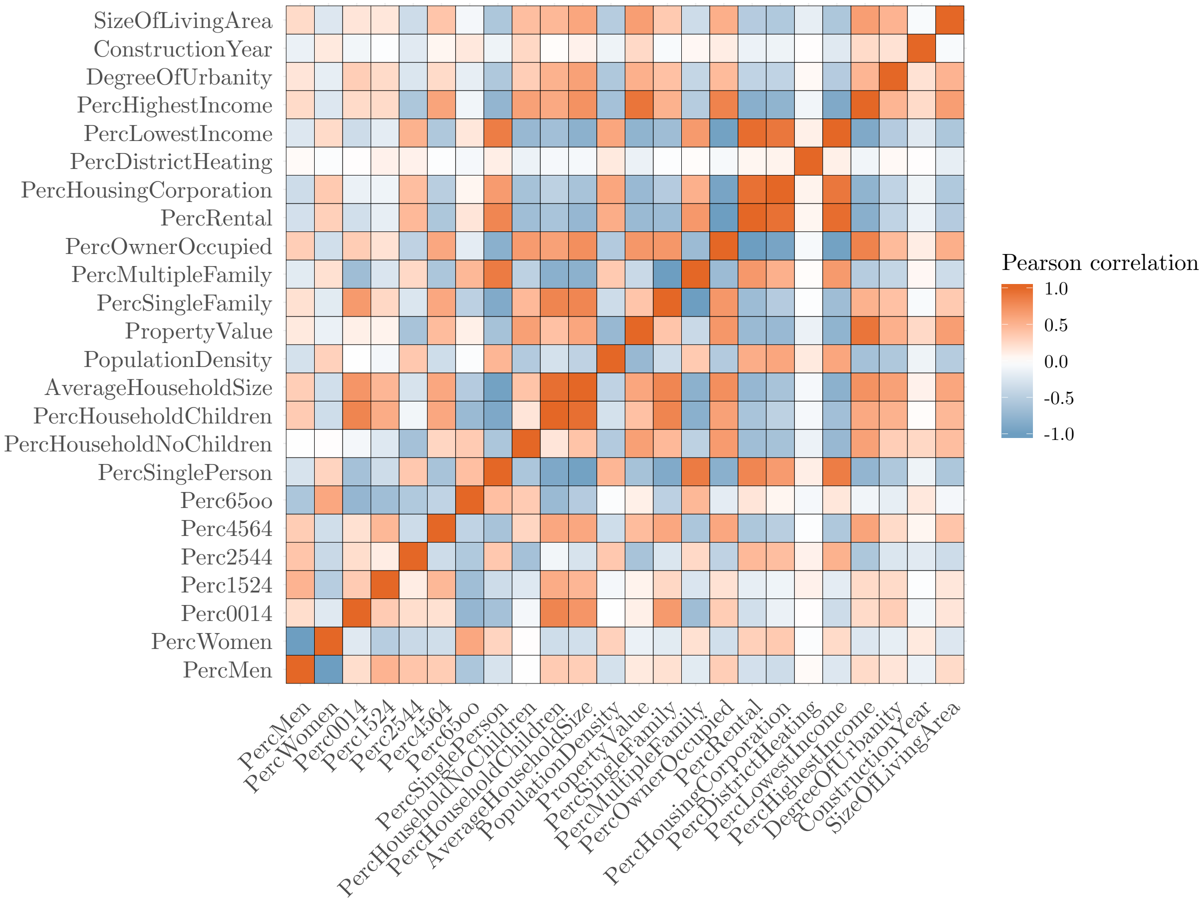


Figure C.2: Correlation matrix of pre-selected variables.

C.3 Conclusion

The data sets used in Chapter 4 require data pre-processing in order to use the data. The energy consumption data mainly needed to be transformed into a form that is suitable for change point detection. In addition, using several rules, we excluded small consumers not being small firms. The variables used from the CBS are based on literature. In addition, data from the CBS needed to be imputed to prevent data removal because of missing values. Also, there was found to be a considerable correlation between variables. For this reason, highly correlated variables were removed from the set. The final data set can be used for the analysis in Chapter 4.

D Background heat pump scenarios

In this research, we determine the impact on the electricity network for various scenarios. In this appendix, we provide a background that led to the final scenarios described in Chapter 5. This appendix is organized as follows. First, we briefly describe the current developments within the heating sector. Second, we describe the challenges for HP growth. Finally, we describe different views on the future share of HPs.

D.1 Heat transition development

The energy transition is a complex process that involves multiple transition areas. Renewable sources need to replace fossil fuel as a source and other technologies will be applied using other forms of energy (Van Leeuwen, de Wit, & Smit, 2017, p. 943). Citizens must become aware and need to be involved. More space for biomass production is needed and energy taxes should shift in favor of renewable energy consumption. To reduce energy demand and the use of carbon-based fuel in the coming years, several policies are already developed. First, new buildings are required to have a lower EPC-level (Energy Performance Coefficient), meaning that heat loss is reduced. Second, large scale energy renovations schemes are introduced for existing houses (Dutch policy called: *De Stroomversnelling*). Approximately 4 million of the 7.7 million houses in the Netherlands are poorly insulated and older than 40 years. These renovations improve insulation and energy efficiency by using, for example, low-temperature underfloor heating and HPs. Third, subsidy schemes are introduced for heat from renewable sources such as district heating and HPs. Many houses ideally first need to be renovated such that low-temperature heating can be applied, which increases the efficiency of, for example, HPs. Moraga and Mulder (2018, p. 19) mention an additional policy, namely the introduction of energy labels. Labeling buildings was believed to incentivise households because it would create a market for energy efficient buildings. This was later confirmed as energy efficient houses sold in 2008 and 2009 had a 3.5% price premium. About 3 million households have an energy label. Further electrification was stimulated by subsidies, as mentioned by Van Leeuwen et al. (2017, p. 943), and that new houses cannot be connected to the gas network anymore.

Van Leeuwen et al. (2017, p. 943) describe that wind turbines, PV, biomass conversion, and geothermal energy are most suitable for countries with shallow coasts and a flat landscape such as the Netherlands. Figure D.1 shows for these most suitable sources how these are converted to electricity or heat. Wind energy is converted via turbines to either electricity or heat using an ASHP. Solar energy is converted to electricity using PV or high-temperature heat using solar collectors. Biomass is either transformed into electricity or low-temperature heat using Combined Heat and Power (CHP) (Dutch: HRe boiler) or high-temperature heat using fermentation to obtain green gas. Finally, shallow geothermal energy is used by GSHPs to generate low-temperature heat and deep geothermal for high-temperature heat or electricity via a thermodynamic cycle.

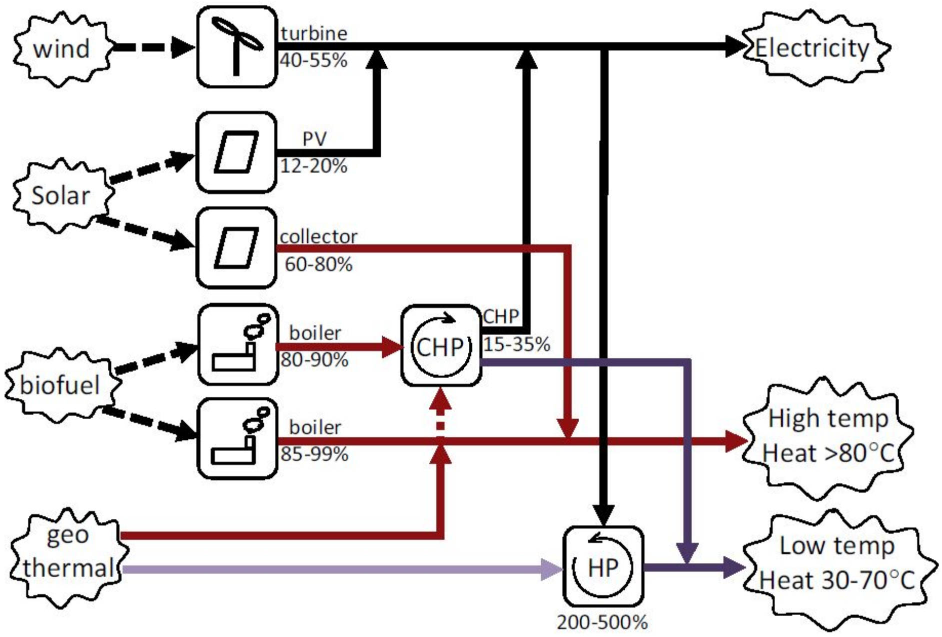


Figure D.1: Schematic overview of possible paths from renewable source to heat or electricity.

As energy demand, as well as energy supply, fluctuates over time, it is difficult to balance supply and demand. Solar energy is obtained during daytime hours. Considering seasonal fluctuations, there is more wind on cloudy days but less solar energy (Van Leeuwen et al., 2017, p. 944). These two complement each other to some extent but not completely. Biomass conversion requires a constant operation for longer times, which makes it less suitable for catching peaks in energy demand. As the fermentation process cannot be stopped, it is thus needed to store the gas. These aspects show the pros and cons of heating alternatives. To use a combination of these sources, a way to level out the mismatch between supply and demand is needed.

Currently, mostly two approaches towards integrating renewable energy options are seen (Van Leeuwen et al., 2017). First, an individual approach is applied in which the natural gas boiler is replaced by a HP. Ideally, electric power needed for a HP is supplied by PV or wind turbines. Second, a collective approach is seen in which district heating is used on a variety of renewable sources such as waste, biomass, bio-fuel conversion, solar thermal energy, and seasonal thermal storage. Mainly waste and biomass conversion are currently used. District heating has advantages for existing buildings in densely populated areas, although better insulation of existing buildings increases the possibility of using HPs. Van Leeuwen et al. (2017) argues that current Dutch policy is mainly aimed at reducing heat loss and increasing the share of renewable energy within the power system. Less attention is given to stimulate district heating. Therefore the positive role of district heating is overlooked by national and regional policymakers.

D.2 Heat pump challenges

The HP technology has some challenges that can prevent purchase by households (Dutch Heat Pump Association, 2020). First, the sound that ASHPs make can be experienced as noisy. Second, the temperature currently supplied by HPs is too low for a complete gas replacement. Third, HPs take a certain amount of space in the household. Fourth, HPs are currently expensive. Other challenges are the insulation of households, directions of the government, and the manpower to realize the installations. Research by the Dutch Heat Pump Association shows that installers are confident that the heat temperature produced by HPs will increase and that the noise will decrease over time. Universiteit Utrecht (2015) studied the HP from a technological innovation system (TIS) perspective. A TIS is a system that develops around, in this case, the HP. A TIS consists of five elements: actors, interactions, institutions, technology, and infrastructure. The speed and direction of technological development are influenced by a TIS. The authors put greater emphasis on the suitability of a HP per type of house (e.g., insulation), requirements from the user (e.g., capacity), release temperature (e.g., floor heating), and source (air, water or ground). GSHPs are mostly expected in new build houses and ASHPs for all other houses. The lower the release temperature, the higher the efficiency of the HP (i.e., COP). Ecofys (2017) estimated the suitability for terraced houses 75%, multiple-family households 50%, and detached & two under one roof as 90%. Given that there are approximately 14,000 multiple family households and 40,000 single-family households in the AGO region, then we expect 38,000 households to be suitable. Ecofys (2017) describes that the HP is most efficient for delivering low-temperature heat. The heat source from the surrounding is between the -10°C and 25°C. This temperature can more easily be increased to low-temperature heat than high-temperature heat as this requires fewer degrees to be heated. The authors mention that the costs associated with improving insulation must also be considered as this is often necessary to make a HP feasible. In summary, the success of HP adoption depends on technological development, price, household renovations, and manpower of installers. Also, governmental stimulation will affect success.

D.3 Performed studies

Veldman et al. (2013, pp. 234-235) studied the main drivers that lead to variations of supply and demand between energy transition scenarios and concluded that there are mainly three drivers.

- A more nationally or internationally focused policy
- A more economically or environmentally oriented society
- Economic growth

Most studies that included scenario analysis of the Dutch heat transition included a combination of these three drivers. We review two performed studies by CE Delft and Berenschot as these provided sufficient detail on the heat transition. Table D.1 shows these studies.

Table D.1: An overview of key findings from two studies on energy transition scenario analysis.

Reference	Key findings for heat transition
CE Delft (2017)	*Regional*: Focus on regional steering and the involvement of individuals. In 2050, mostly local sources are used that are currently already available resulting in 45% full-electric HP, 20% hybrid HP (green gas) and 35% HR boiler (hydrogen), CV boiler (biomass), residual heat & geothermal energy.
	National: Focus on being self-sustaining, individuals and companies allow national steering. In 2050, only nationally available sources are used resulting in 65% hybrid HP (hydrogen/green gas), 10% full-electric HP and 25% HR boiler (hydrogen), CV boiler (biomass), residual heat & geothermal energy.
	International: Focus on international cooperation supported by individuals and companies. In 2050, hydrogen, green gas, and biomass are imported resulting in 70% hybrid HP (hydrogen/green gas), 15% HR boiler (hydrogen) and 15% CV boiler (biomass), residual heat & geothermal energy.
	Generic: Focus on CO^2 reduction via taxes. The transition is an organic process and all alternatives are possible. In 2050, no collective projects are done on hydrogen resulting in 85% hybrid HP (green gas), 10% HR boiler (green gas) and 5% CV boiler boiler (biomass).
Berenschot (2018a) & Berenschot (2018b)	*Electrons*: Focus on full electrification using solar and wind energy. In 2050, 90% full-electric HP and 10% district heating.
	Molecules: Focus on the use of green gas and hydrogen extracted from natural gas. The market share in 2050 is 15% full-electric HP, 75% hybrid HP and 10% district heating.
	Heat: Focus on using geothermal energy, residual heat, and solar thermal. The market share in 2050 is 30% (hybrid/full-electric) HP, 25% solar thermal and 45% district heating.

Table D.1 shows that the HP technology plays an important role in each scenario. The studies performed by Berenschot (2018a) and Berenschot (2018b) note that there is some uncertainty in the future import of biomass and hydrogen. The use of these renewable sources would reduce the need for full electrification. The study performed by CE Delft (2017) shows the expected influence from the degree of steering. Both studies emphasize that these scenarios are extremes and that none will exactly reflect the energy system in 2050. Therefore, these scenarios function in this research to clarify the possibilities.

D.4 Conclusion

In this appendix, we researched the current development of the heat transition, HP challenges, and possible HP scenarios. Currently, mostly district heating and individual HPs are used as renewable energy options. Besides these options, there are more promising alternatives but can often not be used as a 100% replacement for natural gas. Therefore, it is likely that a combination of options will be needed. The HP is an important option as shown in the performed studies but has multiple challenges related to needed manpower for HP installation, noise, price, space requirements, and temperature produced. A research conducted by the Dutch Heat Pump Association shows however that these challenges can be overcome. In summary, the HP is a promising alternative for (partly) replacing the use of natural gas and is expected to have a high market share in 2050.

E Heat pump load profiles

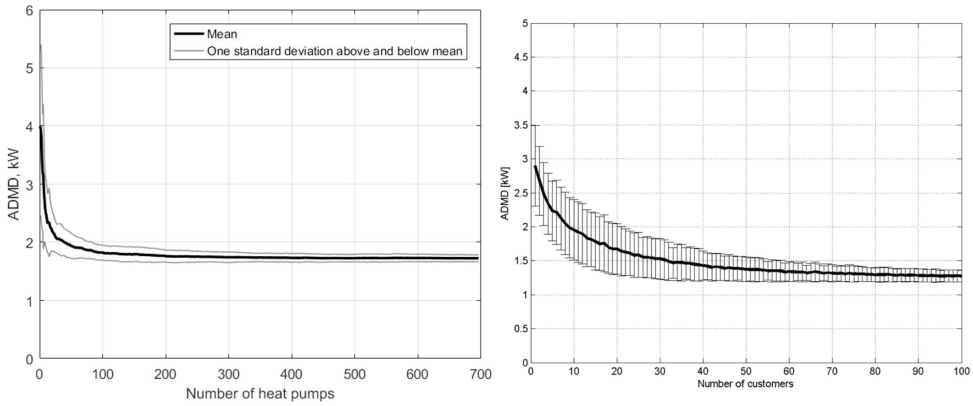
To determine the impact on MV/LV transformers, we need to know what the required HP power is. In this appendix, we study HP load profiles. We do this based on studies that analyzed the electricity requirements of HPs using empirical data. We first research performed studies based on Table E.1. Next, we determine the approach used in our research.

E.1 Performed studies

Table E.1: An overview of studies focused on HP load profiles addressing data, methods, and key findings.

Reference	**Data**	**Methods**	**Key findings**
Love et al. (2017)	Data set of 2–minute energy consumption behavior of 703 domestic HP installations in varying types of houses and construction years. The majority of the HPs were installed in existing houses with basic energy saving measures. Two-thirds of the HPs were installed in social housing. The 703 HPs covered 120 different models from 24 manufacturers. 75% were AHSPs and 25% GSHPs.	Construction of 2–minute load profiles of HPs accounting for coincidence. The impact on the electricity network was evaluated by aggregating the HP load profile with the residential load profile for several penetration levels of HPs.	HP load profiles show two peaks: between 06:00 – 09:00 and between 16:00 – 21:00. Peaks of HPs coincide with peaks of residential load profiles but the peak of HPs is more severe in the morning as shown in Figure E.2. The average peak load of aggregated HPs rapidly decreases when the number of HPs increases because of coincidence.
Barteczko-Hibbert (2015)	Study of 15–minute energy consumption behavior in one year for 15 customer types (e.g., rural & elderly). The number of observations was not given.	Bootstrapping of customers to compute the average aggregated peak load of HPs (ADMD).	The ADMD rapidly decreases when the number of HPs increases. This was studied for 0 to 100 customers.
Veldman, Gibescu, Slootweg, and Kling (2011)	Measurements of 15–minute time series of 7 different types of HPs at 6 locations. These HPs were either combined with a gas boiler or not.	Obtained heat demand is translated to electricity demand (kW) using the COP and characteristics of the HP. An aggregate load profile is constructed for n HPs using randomized heating moments of the HPs.	Peak loads of individual HPs were between 2 kW and 6 kW depending on tap water heating. The measured peak load of 183 HPs was 28.3 kW.
Obinna et al. (2017)	Multiple data sets obtained from pilots of sustainable technologies were used of which two included HPs. Data were obtained in 15–minute time series.	Comparison between the performances of the pilots in terms of monthly energy consumption/production, annual imported/exported energy, peak load & simultaneity (coincidence), and self–sufficiency.	The simultaneity of HPs was found to be between 65%–70%. This is because of outside temperature which is the same for households in a certain region. The ADMD was found to be 3 kW for both pilots with the number of households 21 and 26.

To model the increased peak load on MV/LV transformers, we need to obtain the HP requirements. Love et al. (2017) studied 275 HP load profiles. The authors mentioned that this is the first empirical study on the ADMD (After Diversity Maximum Demand) of HPs. The results of Love et al. (2017) were confirmed by Barteczko-Hibbert (2015). Figure E.1 shows the plots generated from empirical data of the studies by Love et al. (2017) and Barteczko-Hibbert (2015). Figure E.1a refers to the results of Love et al. (2017) and Figure E.1b refers to the results of Barteczko-Hibbert (2015). Love et al. (2017) additionally studied the HP load profiles under different winter conditions, which was not studied by Barteczko-Hibbert (2015). According to Love et al. (2017), for a cold winter weekday with an external temperature of –0.3°C, 1 HP has a peak of 4.0 kW, 40 HPs 2.0 kW, 100 HPs 1.8 kW and 275 HPs 1.7 kW. The uncertainty in this estimate decreases when the number of HP increases (first HP 1.5 kW standard deviation and 275 HPs 0.1 kW standard deviation). Barteczko-Hibbert (2015) instead provided a formula for the relation between the number of HPs and the average aggregated peak load as shown in Equation E.1, where n is the number of HPs. The findings of Veldman et al. (2011) and Obinna et al. (2017) were considered less usable as these studies were based on a smaller number of HPs.



(a) Retrieved from: Love et al. (2017). **(b)** Retrieved from: Barteczko-Hibbert (2015).

Figure E.1: Comparison of two studies on the ADMD of HPs.

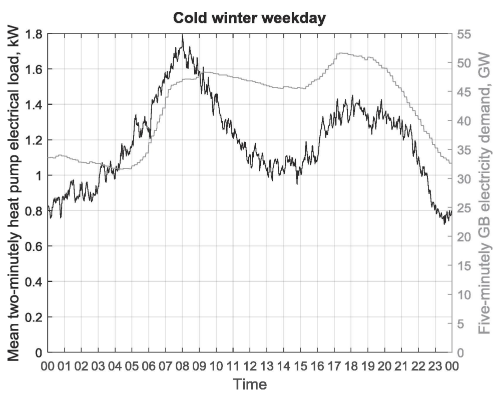


Figure E.2: HP load and the base residential load of households on a cold winter weekday. Retrieved from: Love et al. (2017).

$$ADMD = n \cdot 3.012998 \cdot n^{-0.195356} \tag{E.1}$$

E.2 Conclusion

In this research, we use the results given by Love et al. (2017) based on the following arguments. First, HPs were monitored mostly in existing houses with basic energy saving measures. Given the description of the current developments of the heat transition in Section D.1, we assume that currently no major energy saving measures are taken in the AGO region. Second, in this research, we assume that there are no changes in infrastructure which justifies the use of empirical results in existing houses. The study by Love et al. (2017) did include a large number of social houses. Of the building stock in the Netherlands, 28% is social housing. The results by Love et al. (2017) might underestimate the situation in the Netherlands since the study contained two-third social housing and one-third private housing. Third, varying types of HPs were used, mostly being ASHPs. Given that GSHPs are less likely to be constructed in existing houses because of space requirements, this is also a plausible reason to use the results of Love et al. (2017). Fourth, the empirical study was done in Great Britain. We assume that the weather conditions are similar in the Netherlands. Finally, the study was based on a large number of HPs compared to other studies which increases the reliability.

F Simulation approaches

In Chapter 3, we introduced two approaches to model the spatial diffusion of HPs. In this appendix, we give a detailed description of these approaches. In the description of these approaches, we only consider one year in the simulation. For each approach, we first describe the theory and then how the approach can be applied.

F.1 Empirical distribution

Theory

An approach not discussed in our selection of related work is the use of an arbitrary discrete distribution. Suppose we have a discrete distribution with $X = (x_1, \ldots, x_n)$ and $P(X = x_j) = p_j$, where $\sum_j p_j = 1$. We can divide the interval $[0, 1]$ into n intervals where interval j has length p_j. The order of the intervals in terms of p_j is not important as we only need to know the length of each interval. We could now generate a random variate X by using the following direct inverse-transform method (Law, 2015).

1. Generate a random variate from the uniform distribution $U(0, 1)$
2. Return the non-negative integer $X = I$ satisfying Equation F.1

$$\sum_{j=0}^{I-1} p(j) \leq U < \sum_{j=0}^{I} p(j) \tag{F.1}$$

Application

We can use this approach as follows. Resulting from the logistic regression model, each household has a predicted likelihood of HP adoption. With the likelihood of all households together we can generate a discrete distribution where each household x_j has likelihood p_j. Drawing a random number via the direct inverse-transform method now returns on average the household having the largest interval length and thus the largest likelihood of HP adoption. This method is then repeated each time a HP is assigned to a household (i.e., sampling without replacement). This is illustrated with an example given in Figure F.1. The choice for household a is based on the calculation given in Figure F.2. We have 5 households with each having a likelihood of HP adoption or a score obtained by grouping the likelihood of HP adoption (step 1). If we divide each likelihood by the sum of all likelihoods, we obtain the fraction in which each household contributes to the sum. This can be used in the empirical discrete distribution since each fraction can be used as a probability $p(x)$ (step 2). Summing these probabilities results in an empirical distribution function $F(x)$ from which we can sample using the direct inverse-transform method (steps 3 and 4). Having sampled a household, we now remove this household from the set and redo the sampling until the growth in year t is assigned.

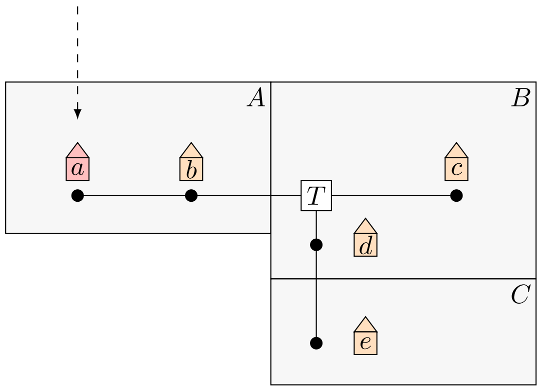


Figure F.1: Visualization of assigning a HP to a household located in a residential area A, B, or C connected to a MV/LV transformer substation T using an empirical discrete distribution.

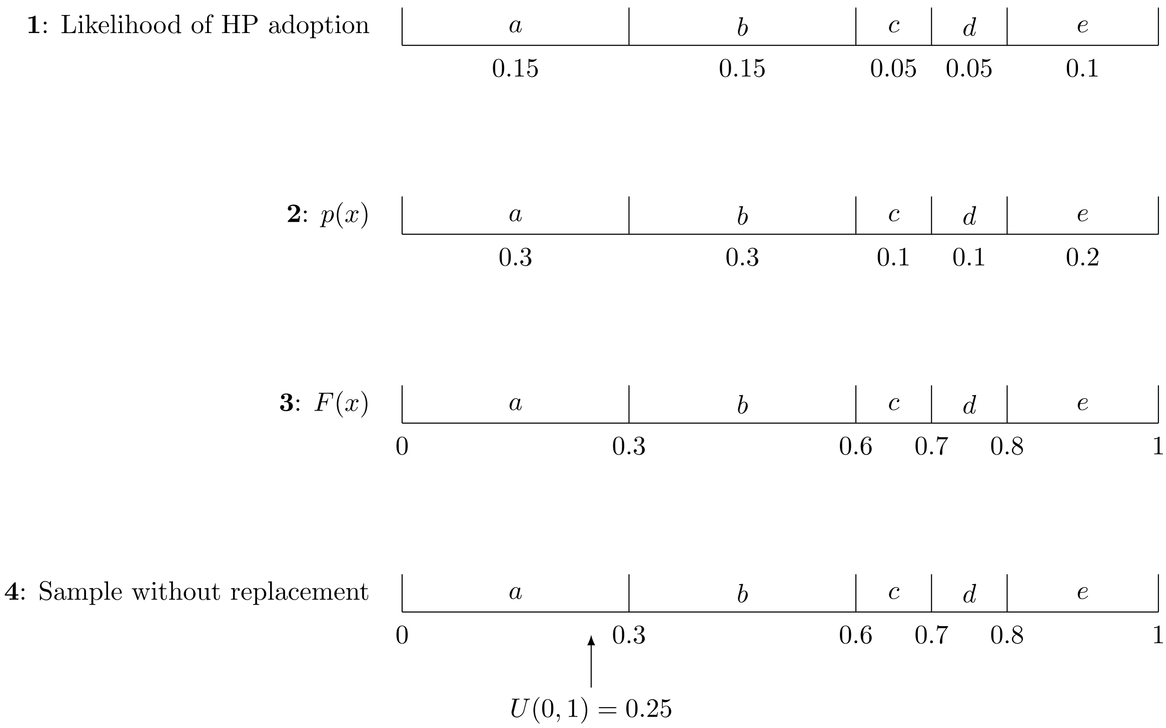


Figure F.2: Example of using the direct-inverse method to draw a household to which a HP is assigned.

F.2 Fisher's non-central hypergeometric distribution

Theory

Although not considered in literature, we can model the yearly spatial diffusion also by assuming a hypergeometric distribution. We first discuss the standard hypergeometric distribution using so-called urn theory. Given that we have two types of balls with colors a and b in an urn with a total of N balls. If there are D balls with color a and $N - D$ balls with color b and we draw n balls, then the number of balls with color a has a standard hypergeometric distribution. The probability mass function for two colors is shown in Equation F.2, where X is the range of the number of balls with color a that can be contained in a drawn sample (Loertscher, Muir, & Taylor, 2017).

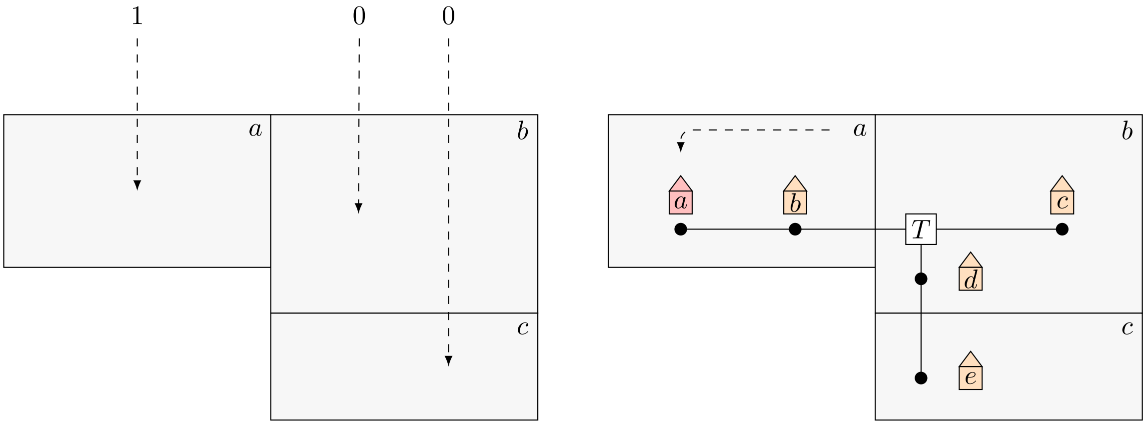
$$p_X(x; n, N, D) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \tag{F.2}$$

This distribution can be extended to a situation in which balls a and b have weights w_A and w_B. If we now draw n balls, then the number of balls with color a has a Fisher's non-central hypergeometric distribution. The probability mass function for two colors is given in Equation F.3. The sampling is done as follows. First, we draw Y_A balls of color a by including each ball with probability $w_A/(w_A + w_B)$. Next, we draw Y_B balls of color b by including each ball with probability $w_B/(w_A + w_B)$. Drawing one ball is independent of the inclusion of other balls in this way. The number of balls Y_A drew has a Fisher's non-central hypergeometric distribution, conditional on the sum $Y_A + Y_B = n$.

$$p_X(x; n, N, D, w) = \frac{\binom{D}{x}\binom{N-D}{n-x} w^x}{P_0}, \quad where \quad P_0 = \sum_{y=max\{0,n+D-N\}}^{min\{n,D\}} \binom{D}{y}\binom{N-D}{n-y} w^y \tag{F.3}$$

Application

We illustrate the application using the exampling given in Figure F.3. First, we determine the number of HPs per neighborhood by drawing random variates from the Fisher's non-central hypergeometric distribution. This results in a vector containing the number of drew HPs per neighborhood. As input, we need the total number of households per neighborhood, the number of households not having adopted a HP per neighborhood, the growth in year t and the odds of drawing households in a certain neighborhood. The odds create the differentiation in the number of HPs picked per neighborhood. If the odds for all neighborhoods are equal, then HPs are assigned according to the number of households per neighborhood (i.e., standard hypergeometric distribution). We can use the results from Chapter 4, the likelihood of HP adoption per household, to function as these odds, even if the likelihood of HP adoption is grouped. Because the likelihood of HP adoption for each household in a neighborhood is identical, one value per neighborhood suffices (i.e., the likelihood of HP adoption for one household). We then draw random variates from the Fisher's non-central hypergeometric distribution to obtain the number of HPs per neighborhood to be assigned. This is illustrated in Figure F.3a as step 1. At this point, we do not know which households are assigned a HP. To do this, we determine for each household the probability of being assigned a HP. This is determined by dividing the number of HPs to be assigned in a neighborhood by the total number of households not having adopted a HP in a neighborhood. This approach is identical to that introduced by Bernards et al. (2016) given in Section 3.4, Table 3.4a. We then assign a HP to a household if a random number from the uniform distribution $U(0, 1)$ is smaller than the calculated probability. This refers to step 2 in Figure F.3b. An example is given in Table F.1. We could have assigned HPs to households using the approach from Figure F.2 but this would require a large number of extra calculations to normalize the empirical distribution multiple times.



(a) Sample from Fisher's noncentral hypergeometric distribution.

(b) Random assignment of HPs.

Figure F.3: Visualization of assigning HPs to households located in residential areas A, B, or C connected to MV/LV transformer T using a Fisher's noncentral hypergeometric distribution.

Table F.1: Example illustrating how HPs are assigned to households in the approach using a Fisher's noncentral hypergeometric distribution. HPs are assigned to a household if a random number is smaller than the determined HP adoption probability.

Connection	**Residential area**	**HP adoption probability**	**Random number**	**HP assigned**
a	A	0.5	0.21	Yes
b	A	0.5	0.81	No
c	B	0	0.34	No
d	B	0	0.12	No
e	C	0	0.78	No