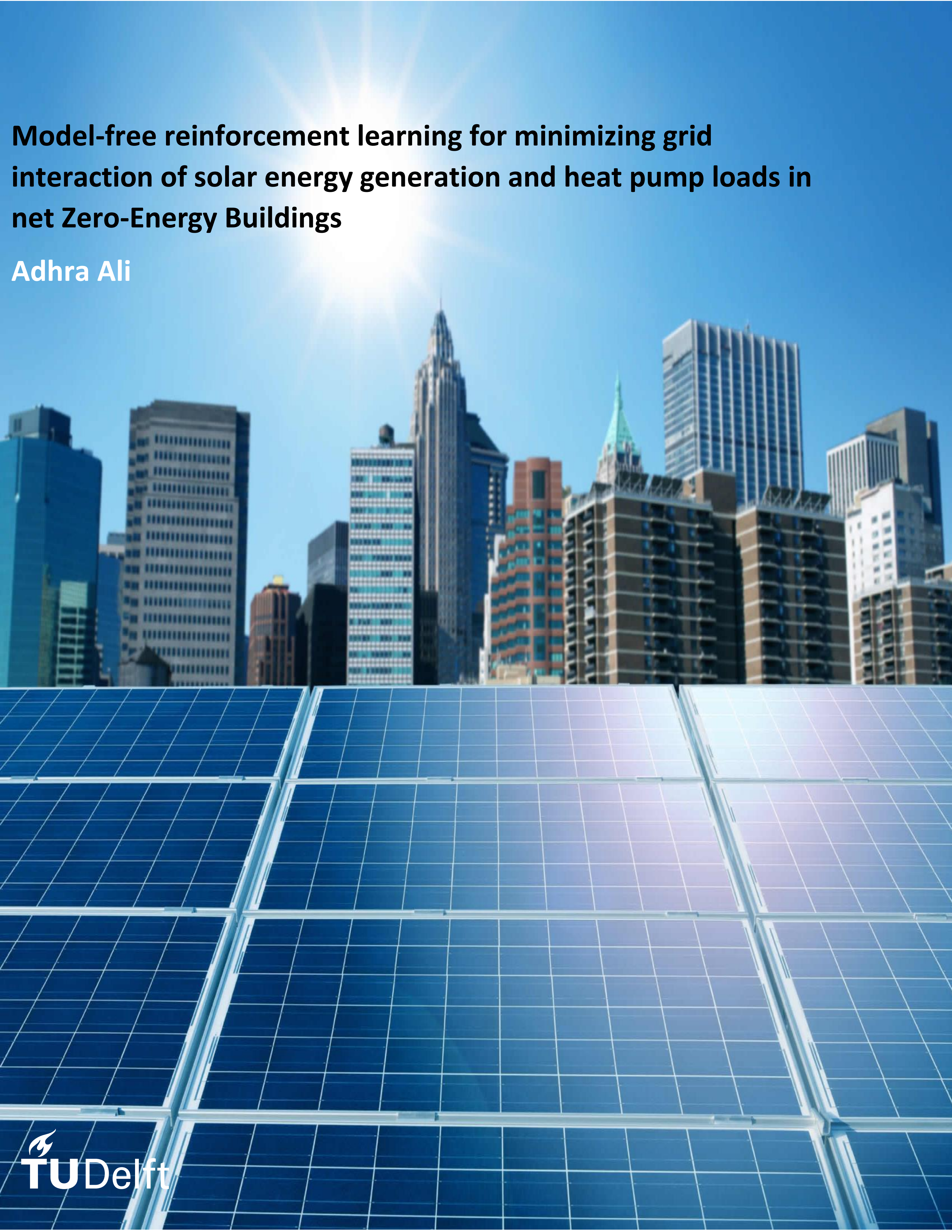
Model-free reinforcement learning for minimizing grid interaction of solar energy generation and heat pump loads in net Zero-Energy Buildings

Adhra Ali



TUDelft

Model-free reinforcement learning for minimizing grid interaction of solar energy generation and heat pump loads in net Zero-Energy Buildings

By

Adhra Ali

in partial fulfilment of the requirements for the degree of

Master of Science

in Sustainable Energy Technology

at the Delft University of Technology,

to be defended publicly on Tuesday May 9th, 2017 at 09:00 AM.

Supervisor:	Dr. ir. P.W. Heijnen	TU Delft
Company supervisor	MSc. H. Kazmi	Enervalis
Thesis committee:	Prof. dr. ir. K. Blok,	TU Delft
	Dr. ir. E. Mlecnik,	TU Delft
	Dr. ir. P.W. Heijnen	TU Delft

Abstract

This study applies model-free reinforcement learning (RL) on a case study based in Utrecht province in the Netherlands to optimize for on-site renewable energy. This aims at reducing the interaction of net zero-energy buildings with the grid as a result of an increase of heat pump installations and renewable energy systems (RES) integration. It is believed that this will become increasingly more important since the regulations regarding 2020 ascribe significant increase in energy efficiency of the built environment. On-site RES self-consumption is therefore a central lead in this research. The project data comprise air source heat pump and solar energy data of 6 different households for the months June to November 2016. The RL learning algorithm was applied to the different data sets to derive an optimized individual and generalized control strategy. Simulations were carried on, to acquire the resulting energy consumption, self-consumption, and self-sufficiency. The results show an increase of individual self-consumption between 17% and 348% and self-sufficiency between 18% and 72%. This results in an additional monetary benefit for the occupants based on the transition proposals of 2020 for the renewable energy generation net-metering abolishment in the Netherlands. Furthermore, reducing the grid interaction implies benefits for the grid operators in terms of investments required for grid reinforcement.

Acknowledgement

I enrolled in the master programme of Sustainable Energy Technology with the intention to learn as much as possible about the various fields of renewable energy and sustainable energy operations. Now at the end of the master I can say that it surpassed my expectations and I am glad to have been part of this programme. Especially this master thesis was by far the most instructive part of the master. Therefore, I would like to express a heartfelt word of thank to all my supervisors for their continuous support throughout the process of this research. I can now say that I am one of the lucky few to have had supervisors whom not only showed a lot of interest in the project but were also very much engaged in the process. Therefore, it is only fair to say that I am thankful to them of having given me the opportunity to learn from their expertise and their supervision throughout the process. A special word of thank for Hussain Kazmi who gave me the opportunity to be part of this amazing project and also for his support to make me master a small part of the machine learning theorem starting from zero.

Adhra Ali

Den Haag, April 2017

Summary

The renewable energy share is increasing significantly in the residential built environment energy sector which implies an expected increase in grid interaction. Much effort has been invested in building energy demand reduction through many efficient building design concepts and strategies such as zero energy buildings. However, research has shown that a significant performance gap is evident between predicted and actual building energy performance. Occupant behaviour is believed to play an important role in this matter. This has a negative influence on the objective of net Zero Energy Buildings (nZEB) for these concepts are formulated, amongst other, according to a simulated share of the renewable energy source (RES) of the total residential energy consumption. It is therefore crucial to predict the building energy consumption accurately in order to arrive at an accurate nZEB building energy performance. In addition, the inaccuracy of renewable energy consumption results in a larger than expected feed to the grid.

Moreover, the installation of electrical heating systems such as heat pumps is projected to increase substantially. With the objective of large scale implementation in nZEB concepts, this implies a significant increase in grid interaction as a result of both the electrification of nZEBs and the installation of renewables challenging the stability of the utility grid. Optimizing control strategies is therefore necessary to be applied to regulate the building energy loads in order to implement peak shaving to reduce the need for grid reinforcement. As a result, this will allow for larger integration of RES in the residential built environment. Additional advantages hold a reduction of required storage capacity.

The majority of the research in this field relies on optimization of load matching by means of rule-based control and model-predictive control. These methods have the disadvantage of being poorly generalizable for system specifications and lack the potential to incorporate accurate predictions for the occupant behaviour. It is argued that data-driven methodologies such as reinforcement learning algorithms (RL) are therefore more suitable for such applications. One advantage of model-free learning algorithms over model-based approaches, is that it provides generalisability whilst physical models are built based on specific cases. This makes them generalizable and not domain-specific.

In order to accurately incorporate the stochastic nature of occupant behaviour and enhance generalisability, model-free RL algorithms will be applied in order to find the optimal control strategy for an air-source heat pump (ASHP) for the domestic hot water (DHW) load for a case study in Soesterberg in the Netherlands. The available data for the case study comprises data of June to November of 2016 for 6 different houses.

The optimal control strategy is required to optimize for solar energy self-consumption and self-sufficiency, facilitated by the ASHP and hot water storage vessel. The occupant comfort and the efficiency of the ASHP are two limitations for the optimization. Hence, the self-consumption is not supposed to increase by means of dissipating solar energy.

The results show a significant reduction in the grid interaction as a result of the optimized solar energy consumption. The solar energy self-consumption increased by 91% on average compared to the baseline values. This leads to a rise in the self-sufficiency of around 36% on average reaching final an average self-sufficiency of 89% in summer. The ability of the algorithms to incorporate the occupant behaviour proofs to improve ASHP control strategies for maximizing on-site renewable energy consumption. Therefore, applying this algorithm on a cluster level provides promising results for the objectives of the study in terms of profitability for the grid operators and the occupants.

List of abbreviations

ASHP – Air Source Heat Pump

COP – Coefficient Of Performance

COP21 – Paris Climate Conference 2016

DG – Decentralized Generation

DP – Dynamic Programming

DHW – Domestic Hot Water

DSM – Demand Side Management

DSO – Distribution System Operator

EBC – Energy in Buildings and communities

HVAC – Heating Ventilation And Cooling

IEA – International Energy Agency

MDP – Markov Decision Process

MPC – Model Predictive Control

NREAP – National Renewable Energy Plans

nZEB – Net Zero Energy Buildings

PV – Photo-voltaics

RED – Renewable Energy Directive

RES – Renewable Energy Systems

RL – Reinforcement Learning

TD – Temporal Difference

TES – Thermal Energy Storage

List of symbols

$\mathbf{Q_0}$ - Action Value Function for action 0

$\mathbf{T_m}$ – Midpoint temperature

$\mathbf{T_{set}}$ – Maximum Set point to which the vessel is reheated

$\mathbf{T_{threshhold}}$ – Minimum temperature the vessel is allowed to reach

Table of contents

1. Introduction

Recent developments in the global policies regarding CO_2 and climate are evident from representations such as the Paris climate conference (COP21) held in 2016 [1]. The mandate stated in Paris portrays a global willingness for taking action in the context of shifting towards a low carbon economy [1]. Renewable energy technologies are being integrated on an increasingly larger scale as time progresses as shown for the European case in fig.1 [2]. The European statistics show a gradual increase for the individual National Renewable Energy Plans (NREAPs) of the European member states complying to the Renewable Energy Directive (RED) [2].

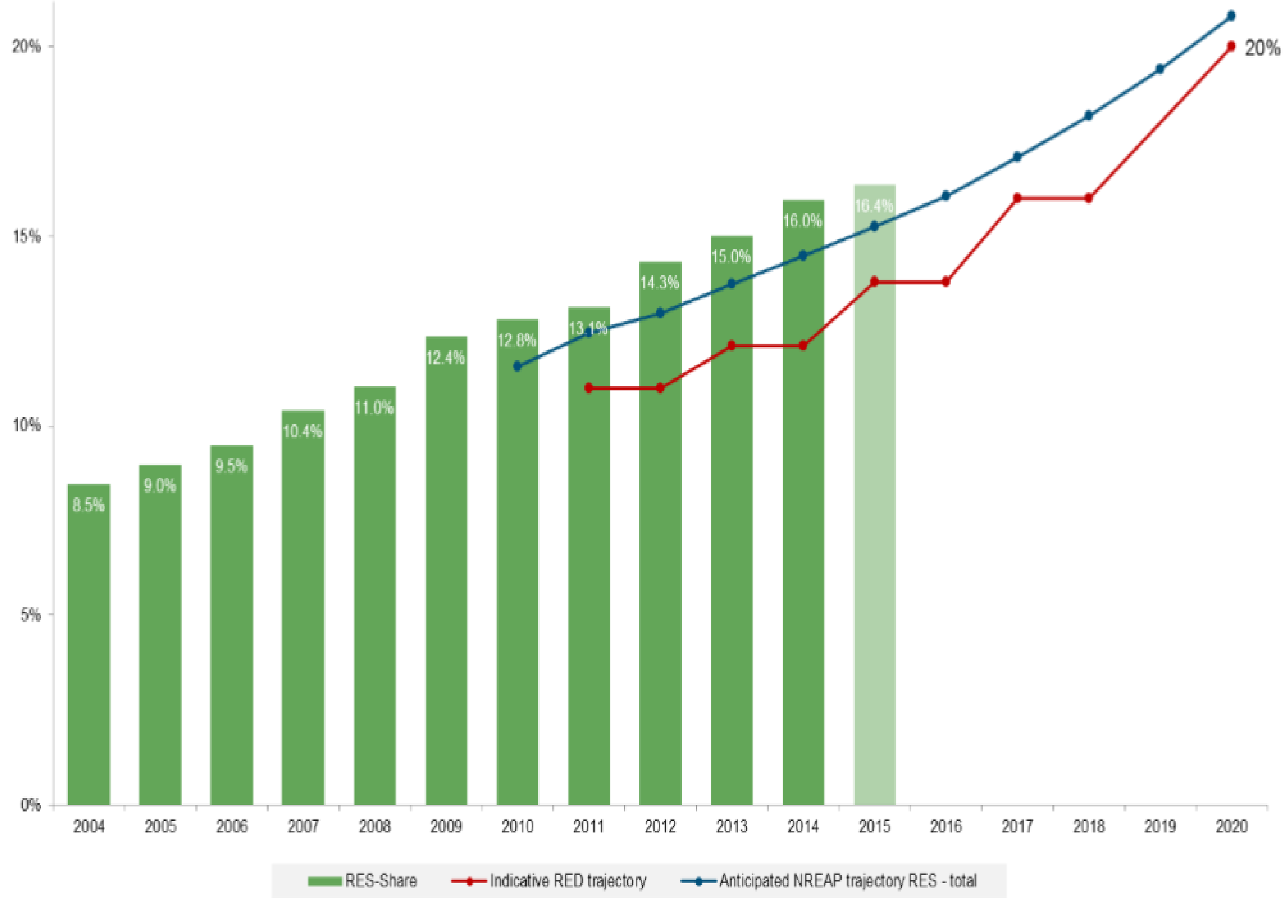


Figure 1. Energy mix 1990 – 2015 [2]

This progress corresponds with international and local policy development, such as those in the EU, prescribing binding directives carving a path for a larger integrated renewable energy share [3]. The development towards a carbon free energy market is realized by reforming the different energy sectors. The European energy market is divided in four main sectors comprising the residential built environment, industry, transportation and services [4]. Different policies are directed for each of these sectors according the current and projected production and consumption. The energy consumption of European households alone comprises around 26 % of the final energy consumption. Therefore, like for the other segments of the energy sector, policies have been stated regarding energy conservation in the residential built environment. The most determining directive indicates that all new building must be nearly zero energy buildings by 2020 [5]. Furthermore, when the nearly zero energy concept is developed, it could be further extended to a net zero energy buildings performance concept. A net Zero

Energy Building (nZEB) concept provides that the energy consumption beyond the high energy conservation, the remaining energy consumption of the building is compensated by onsite energy generation [6]. In the IEA SHC Task 40/ECBCS Annex 52, several criteria have been set for defining zero energy buildings concepts. This is due to the large variety of building energy sources and sinks making up the total energy balance. The focus of this thesis will rely on the energy segment of the nZEB that is provided by the renewable energy source at stake for that segment is highly dependent on load matching.

1.1. R&D and policy

The policies mentioned in the previous section give shape to subsequent measures aiming at achieving energy conservation objectives depending on the responsible institution and specific location. The International Energy Agency (IEA) is one of the prominent organisations promoting research and development serving as an advisory source for policy makers in 29 member countries [8]. Accordingly, the IEA has established several programmes initiating objectives for research in this field. A relevant programme for this paper is the Energy in Buildings and Communities (EBC) programme established as a result of its predecessor program, the Energy Conservation in Buildings & Community System (ECBCS) and the Solar Heating and Cooling Program (SHC) programme. These programmes are initiated in recognition of the significance of energy consumption in buildings executed through a series of annexes in the EBC programme and Tasks in the SHC program [9]. The projects extend the IEA objectives concerning renewable energy, to building related energy and building integrated solar power applications. Additionally, IEA sets up a heat pump program (IEA Heatpump Program) providing an international information service for heat pumping technologies, applications and markets [10]. A third interesting programme set up by the IEA correlated with building energy efficiency, is the IEA Demand Side Management Energy Efficiency Technology Collaboration Program (DSM TCP) aiming at developing and promoting opportunities for demand-side management [11].

Projects likes these provide the required R&D to serve as support and framework for policy change in the context of the residential built environment. These developments are especially crucial for enhancing a large-scale deployment of renewable energy and energy efficient technologies in the pursuit of reducing greenhouse gasses emissions. The 28 EU member states have increased the share of the final energy consumption generated by renewables from only 14.4 % in 2004 to 16 % in 2014 [12]. Currently, the EU has set their mission to increase the renewable energy share to be 20% of the final energy consumption in 2020 [3]. The requirement for faster transition to a carbon free energy mix indicates the unquestionable key influence of policies on the energy market. This total European RES share is effectuated by allocating specific energy from renewables targets to the different member states. The individual targets have been set according to the individual GDP, modulated to reflect the individual starting points, and by accounting in terms of gross final consumption of energy with account being taken of the individual past efforts with regards to the use of energy from renewable energy sources [13]. The targets vary between 10% to 49% of the final consumption.

1.1.1. The Dutch context

A target of 14% has been fixed for the Netherlands for 2020 given the limited current share of renewables of around 5.8% [14]. The Dutch government followed by initiating policies aiding this goal. The first regulation to mention is the SDE+ subsidy allocation of euro per unit of energy for enterprises, companies and institutions generating renewable energy. A budget of € 4 milliards is allocated in total to cover the subsidy applications comprising renewable energy generated from biomass, geothermal, water, wind, and solar. Different conditions apply to each of these different sources of energy [15]. The second stimulation measure, the EIA, is given in the form of a tax reduction to the revenue tax of 58% for investing in energy saving and renewable energy technology. On average this tax reduction could save up to 14 % of the cost resulting from energy consumption and taxes [16]. A total budget of €161 million is allocated for this scheme. An additional created incentive is a subsidy for individuals purchasing a solar water heater, heat pumps, biomass boiler, and pellet heaters [17]. Whereas the previously mentioned subsidies hold for enterprises and companies, the ISDE subsidy is open for individuals interested in investing in energy saving technologies such as for their own households or offices.

Other measures include the increase of taxes on oil and natural gas and covering the cost related to risks of drilling in unsuitable locations for geothermal applications [18]. All of these measures hold up until 2020 in order to meet the directive's goals.

Similar goals have been set for the built environment to enhance the improvement of energy efficiency and deployment of renewables. The main aim of the national policy regarding this issue is reach nearly zero energy buildings towards the end of 2018 to 2020. In order to reach this goal, the standards for the energy performance has been tightened since January 2015 [19]. Government buildings have to be transformed to nearly zero energy buildings by the end of 2018 and all other buildings will follow from 2020 and on.

1.2. nZEB grid interaction and the role of grid operators

Given the aforementioned directive concerning the nZEB implementation and the directive of a share of 20% renewable energy in the final energy consumption in 2020, a change in the energy infrastructure is necessary to accommodate this significant increase of volatile power to the grid. Generally, it is argued that in order to reach significantly large shares of renewables, the current infrastructure should be transformed from a centralised to a decentralised generation (DG) system that would utilise renewable energy technology sources [20]. This is due to the rise in locally generated energy compared to the traditional centralised power plants, adding new actors to the energy system. Moreover, aiming at increasing energy efficiency, energy consumption close to the source is favourable to avoid transmission and distribution losses as a result of necessity to transport the generated energy to the load over long distances.

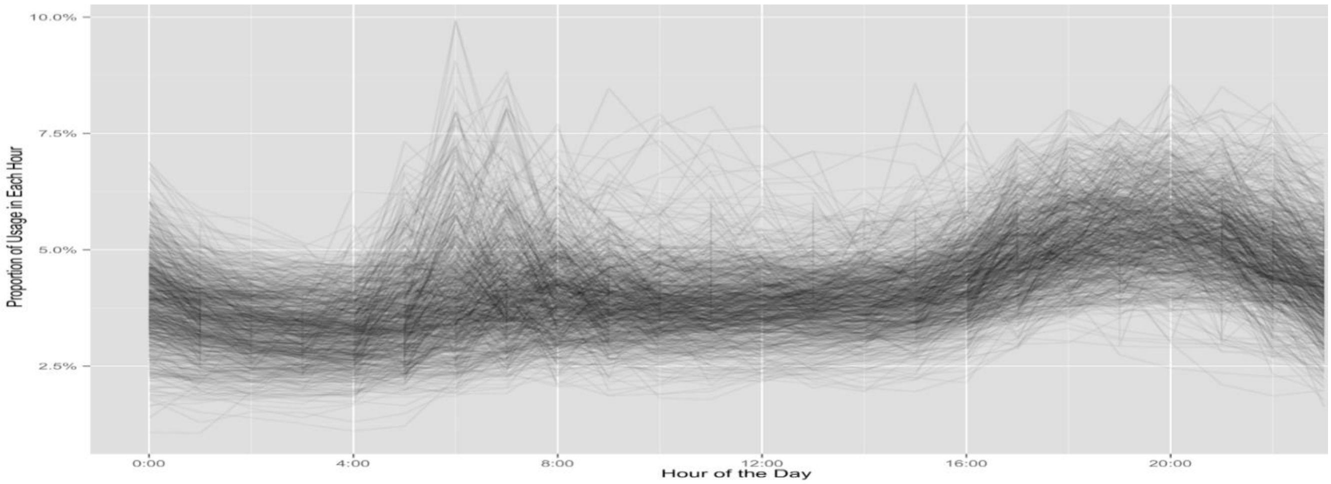


Figure 2. Hourly electricity consumption from a random sample of 1000 residential utility consumers, for a typical weekday [21]

The rise in decentralised energy generation imposes risk for the grid reliability as a result of power congestion caused by the volatile renewable energy resources. The high power fluctuation of renewable energy sources such as wind solar power pose several challenges on the power quality and have therefore to be carefully implemented [20]. Moreover, residential energy demand patterns are often not matching the power production patterns of the renewable energy sources. In fact, an average European load demand of a residential building is known for having two dominant peaks, one in the morning and one in the evening as shown in figure 2 [21]. These peaks are not synchronous with daily solar radiation patterns which has a peak around noon.

This could be addressed by integrating more storage facilities and reinforcing the grid components such as the transformers and cables which, given the required RES goals of 2020 and beyond, implies significant investments. The weak link in this area of renewable energy generation is the lack of large scale storage capacity due to its relatively high cost. Also, increasing the size of the grid components is a similarly costly procedure. However, it is believed that the necessity for a costly reform of the energy infrastructure could be reduced and substantially spread over a larger period of time if the variety of decentralised actors could be centrally monitored and managed [74]. The need for centrally monitoring and managing the variety of power sources and sinks in the grid has introduced the concept of Smart Grids [74]. Smart Grids are required to facilitate real-time monitoring and steering tools so as to allow the distribution and transportation capacity of the grid to progress in a more flexible fashion.

In the traditional top-down approach of the energy system, the Distribution Systems Operators (DSO) is the main responsible for the operation and maintenance of the distribution system. However, the need for Smart Grids imposes the need for a change in this role to a more pro-active grid management. This introduces new service opportunities for grid operators. Connecting a large number of nZEB buildings with fluctuating individual energy generation is one such example of locally generated energy that adds to the volatility of the grid power for which grid operators could monitor and manage. Smart grids enable buildings to respond and provide current operation data to the grid [76].

An nZEB is a technical concept introduced to reduce the energy required due to the conservative nature of the design. Nevertheless, studies have shown that nZEBs could even perform less efficiently than technically poorly performing building designs if the occupants fail to steer the concept to its full potential [22], [23], [24], [83]. This particular matter has been investigated by different approaches all aiming at increasing building energy flexibility and improving building energy performance. Research in this area ranges from fully automated building energy discarding occupant behaviour, to smart metering objectives aiming at increasing occupants' energy performance awareness, and control strategies based on electricity prices signals etc [26], [27], [28], [73].

Also, the integration of renewables and high efficiency HVAC devices is expected to amplify the interaction of residential built environment with the grid. Therefore, one of the main goals of the IEA EBC Annex 67 is to arrive at a common definition of building energy flexibility. This concept is currently defined as 'the ability of a building to manage its demand and generation according to local climate conditions, user needs and grid requirements' [25]. Considering these facts, the deployment of nZEB should be better monitored and analysed in order to achieve the standards aimed at initially. The managing role of grid operators allows it to provide services for the transforming of the residential built environment to a more economically and technically feasible energy sector.

1.2.1. Economic aspects nZEB grid interaction

Grid-tied renewable energy sources in the residential built environment influence not only the technical aspects of the power system but has also implications for the economics of it. Excess renewable generated power is fed to the grid and power deficit is replenished by power from the grid. Since currently the interaction between grid operators and the consumer is transforming, the economic aspects are changing accordingly. Concerning the consumer, feed-in tariffs and metering methods are utilised to compensate for the delivered power by on-site renewable sources to the grid. Depending on the local regulations regarding these two methods, power delivered to the grid might be more profitable compared to maximised onsite renewable energy consumption as will be illustrated in the next section. In contrast, a maximized onsite consumption allows for a reduction in the interaction with the grid and therefore enhances the goal of increasing the share of renewable energy. The same principle holds for optimizing on-site renewable energy consumption in the residential built environment. An nZEB concept reduces the net annual building operation energy significantly, however, does not take the matching between energy consumption and production into account as long as the consumption is compensated over a certain period of time. Nevertheless, increasing the share of renewables in the built environment, requires peak shaving strategies in order to reduce the volatile power surges from and to the grid. Consequently, reducing grid interaction delays the need for reformations necessary to the contemporary energy infrastructure given the development towards a fully renewable energy infrastructure. Hence, it is of interest to investigate the potential of maximal onsite renewable energy consumption in order to recommend for the future role of grid operators for this energy sector.

1.2.2. Dutch metering policy

Dutch policy regarding this issue provides a metering approach by which the cost of the excess energy provided behind the meter to the grid, is deducted from the cost of the total power delivered to the consumer. The cost per unit of power is equal for delivering and consuming power i.e. no additional taxes are applied for power delivered to the grid. However, this type of metering goes up until a delivered power quantity equal to the annual power consumption of the consumer [29]. Power delivery beyond the quantity of the total annual consumption is compensated according to a different regulation dictating the distribution system operators (DSOs) to compensate the producer with a fair feed-in tariff which is to be determined by the DSO. This whole regulation scheme could also be described as a *net* feed-in tariff compensation scheme; Only surplus electricity is sold to the grid for a fixed price which is determined by the DSO. This implies that for an individual household, it is very much profitable to generate electricity in the range of their own consumption and beyond. This regulation reduces the energy bill significantly and so it stimulates an increase in renewable energy installation. In turn, this reduces the need for power generated from conventional resources. In the long run, this regulation is expected to be abolished by 2020 [29]. By then, plans for new measures for maintaining the profitability of renewables installations will be established. The reason for this, is the increase in pressure on the grid utility as the penetration of on-site renewable energy is expected to increase causing congestions and large fluctuations.

In conclusion, grid interaction has not only an influence on the technical aspects but the economic factors related are equally important for policy shaping aiming at increasing the share of renewables. This illustrates the importance of grid interaction management in the field of the residential built environment and the enhancement of direct consumption of renewable energy. As grid interaction in this field originates from the balance between the building load and the onsite renewable energy generation, this balance has to be further investigated.

1.2.3. Heat pumps in nZEB concepts

An important starting point for building energy performance examination is to identify the individual influence of each of the different energy sinks and sources in residential buildings in order to optimize the energy performance effectively. Indeed, studies have shown that the largest portion of end energy consumption in residential buildings consists of energy related to heating, ventilation, and cooling (HVAC) operation ranging between 42% and 68% of the end use on average followed by Domestic Hot Water (DHW) appliances ranging between 14% and 30% for the European average, UK, USA, Australia and Spain [30][35]. The DHW load share of the total energy end use in nZEB concept is higher than less efficient constructions as a result of the reduction of the energy needed for space heating [75]. Moreover, an extensive research on HVAC operation and optimization has been conducted in the various fields of the subject [31][32][33], however, less focus has been paid to DHW appliances when especially this is becoming increasingly more important for grid interaction, given the increased electrification in the field of hot water technologies for residential purposes such as heat pumps [34].

Heat pumps and electric water heaters have been increasingly promoted by financial incentives and implemented in nZEBs as they operate with high efficiencies in proper conditions [35][36]. Amongst all type of heat pumps, Air Source Heat Pumps (ASHP) are implemented the most due to their relatively lower implementation cost [35]. Though, the performance of heat pumps is highly dependent on the site-specific conditions such as the heat source temperature and operation schemes, generally, ASHPs attain a higher coefficient of performance when the outside temperature is higher as there is more energy to exchange. Site-specific conditions were found to influence variation in domestic energy demand from ASHP by up to 13% [37]. Also, the carbon footprint has been found to be energy-related depending on the energy source. In order for heat pumps to be economically viable as domestic water heaters, their efficiency needs to be at least twice that of standard electric water heaters [35]. Typical heat pump efficiencies range between a coefficient of performance (COP) value of 2 and 4 [84]. Moreover, heat pumps increase the seasonal imbalance between PV power generation and load demand patterns resulting in higher grid interaction on seasonal level [38] as shown in figure 3.

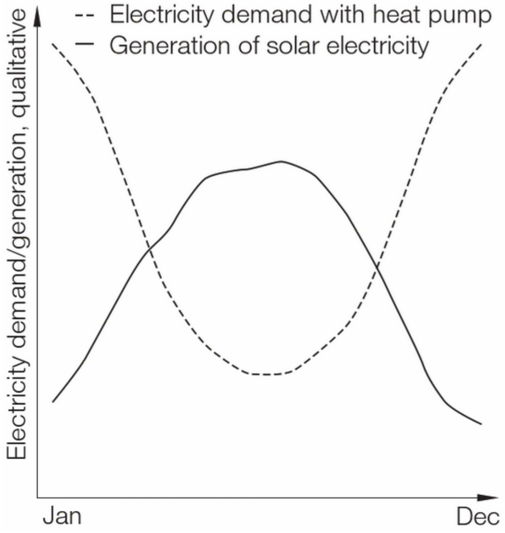


Figure 3. Monthly variation of energy produced by PV and required by an ASHP [38]

Overall, the increased electrification of DHW appliances in addition to the increasing trend of onsite power generation such as from PV in combination with the inefficient occupant behaviour is expected to increase the interaction with the grid and thereby affect the grid stability. It is therefore evident that the implementation of heat pumps requires maintaining ideal operation conditions as far as possible. Especially in the context of a country as the Netherlands where the outside temperatures are not predominantly high compared to lower latitudes, performance improvements need to come from controlling the operation patterns of the device.

It is estimated that a total of 500.000 heat pumps will be installed in the Netherlands by 2020, corresponding with 1 in 15 households [70]. Also, it corresponds with a total installed capacity of 22.5 GW for auxiliary heating whereas the maximum allowed installed capacity is equal to 15 GW for the current grid capacity [70]. The central weakness concerning a large-scale installation of heat pumps, is the timing of the operation scheme. In extreme scenarios in which all heat pumps are activated simultaneously, the peak demand surpasses the capacity limit of the grid. Black-outs will occur more often implying the need for grid reinforcement. Evidently, managing the operation of heat pumps on district level is necessary to guarantee a certain level of grid reliability and safety which emphasizes the importance of a centralised monitoring entity such as the DSO. This is to avoid a high monetary investment which is associated with large-scale grid reinforcement [69].

The majority of the heat pumps are installed with a heat buffer around one 100 litres for DHW purposes [70]. This allows for a larger potential for load shifting for the DHW. The thermal inertia of the dwelling is also often used as a thermal storage tool in nZEBs to manage the operation of thermostatically controlled loads [85]. Additionally, the evident increase of RES installation for nZEBs as mentioned previously will allow for less electricity uptake from the grid if the energy generation is matched sufficiently with the consumption profile. It is therefore interesting to investigate the potential of load shifting applied to heat pump installations backed up by renewable energy sources as a service applied by grid operators.

1.3. ASHP automatic control systems

Limited research has been conducted to simulate realistic consumption profiles for ASHPs domestic water heating purposes [36]. The majority of the research in this field is conducted by means of simplified rule-based models based on assumptions on the expected load profile [34], [67], [72]. Not only there is a lack of actual experimental data of the thermal performance of the accompanying storage tanks but also a lack of a developed field of research regarding occupant behaviour patterns [36]. As a result, the thermal performance of ASHPs is expected to be insufficiently correlated with real case actual performance given the results in [22], [23], [24]. Moreover, studies that make use of more sophisticated data-driven modelled occupant behaviour, apply control to thermodynamic models [71], [73]. However, thermodynamic models are built upon physical parameters that are not only site dependent but also depend on the physical characteristics of the whole installation. Generalising this type of simulations is therefore challenging and adaptation to individual cases is problematic. Instead, reinforcement learning (RL) based approaches could learn the occupant behaviour and the thermal response using data from individual houses [36], [71], [77]. Based on this, an optimal control policy for heating is learnt optimizing the energy consumption by the heat pump. Learning the optimal control policy for the heating device by this approach tackles most of the aforementioned challenges related to generalisability and inaccuracy due to the lack of accurate occupant behaviour simulations. It is therefore believed that this approach is very much suitable to improve DHW appliance efficiency and perform load shifting.

Therefore, the concept of reinforcement learning for optimizing for renewable energy consumption by means of an ASHP applied to DHW purposes will be central in this thesis. RL methods will be described in more detail in the methods section followed by the research problem and subsequent research questions. This will then be followed by a review of the literature on this field of research setting a context for the methodology and case study.

1.4. Summarized problem statement

The combined effect of the increased *onsite power generation* as a result of the expansion of nZEB concepts, increase of *electrification* in building energy installations and the *stochastic occupant behaviour* on the interaction of the residential built environment is expected to have a challenging impact on the grid. The significance of this issue is strongly emphasized by the notion that the implementation of this type of hybrid systems will have to increase as a result of the objective of reducing fossil fuel utilisation. Hence, the investigation of this matter is crucial for paving the path for increasing energy efficiency in the residential built environment by complying to the objective of increasing the share of renewable energy. An important indicator to be investigated is the coupling of actual occupant behaviour to the building performance and the resulting load pattern. To emphasize, applying rule-based control methods to building energy performance simulation do not accommodate accuracy in this field of research as described earlier [36]. This is illustrated by a significant performance gap which describes a large discrepancy between simulated energy performance and the measured energy performance [22], [23], [24]. This is attributed, for a great deal, to the lack of accurate occupant behaviour simulation. This emphasized that the occupant behaviour cannot be discarded from the energy performance analysis and this factor should be incorporated in the building energy systems simulations for more accuracy. Therefore, it is argued that the occupant behaviour could be optimally incorporated by acquiring real data of actual recorded behaviour. This will result in more accurate building energy performance simulation and the resulting control schemes for optimal utilisation of on-site renewable energy. Therefore, it is expected to have a positive effect on the grid interaction and hence, on the profitability of the nZEB concept and its renewable energy components.

This research will aim at addressing this interconnected matter in order to come up with a concept that integrates occupants' behaviour to DHW utilisation and renewable energy generation. This is considered to be essential to reduce nZEB interaction with the grid which is expected to support the increase a total share of onsite produced renewable energy in the grid. In order to analyse this particular issue, a framework will be constructed for the research question in the next section aiming at analysing the different aspects of the matter. In addition, the research question will be broken-down by several sub-questions aiming at supporting the different aspects involved in the research question.

1.5. Research framework

The aforementioned problem will be analysed by the following research question:

"How does reinforcement learning assist to match DHW loads to solar energy generation peaks to reduce the interaction of nZEBs with the grid and what are resulting benefits for the occupants and the grid operators?"

This research question will be covered by the following sub-questions:

1. What are the comfort standards that apply for DHW consumption?
2. What is the potential of shifting the ASHP heating cycles from low to high energy generation hours to maximize solar energy consumption without violating the comfort standards for DHW?
3. What is the effect of shifting the heating cycles on the grid interaction?
4. What is the effect of the optimization on the number heating cycles and the resulting energy efficiency of the heat pump and hence, the energy efficiency of the dwelling?
5. What is the influence of the individual DHW consumption on the performance of the algorithm?
6. How could this methodology be applied on an aggregated level?
7. What does this optimization imply for both the occupants and the grid operators?

Analysing these aspects of the research poses the necessity to conduct a literature research in order to identify the various studies conducted in this field. This thesis will focus the research in the fields of optimal control and reinforcement learning applied to the residential built environment. The following section will fulfil this role in presenting the literature review on the advances reached on this matter to give position to the significance of the research framework in this field.

2. Literature review

Energy efficiency in the built environment is a relatively mature field of science encompassing a wide range of methods for increasing energy efficiency. Generally, the different approaches are to be divided in three main categories of fields of improvement:

1. Energy demand reduction
2. Energy and material recycling
3. Renewable energy production

This field is made up of a large variety of research aiming at improving the building energy efficiency through building envelope improvements and smart and bioclimatic design in order to *reduce* the end use consumption related to the building energy performance [7], [48], [49], [50] and [51]. Reducing the energy required for maintaining a comfortable indoor climate is the first step to start with for increasing the energy efficiency. A subsequent aspect is to install high efficiency technologies used for building energy performance operation such as HVAC and DHW appliances. These technologies increase the energy efficiency due to their high efficiency operation mechanisms and/or their energy *recovery* ability. Lastly, the remaining required energy is compensated for by *producing* energy from on-site and/or off-site renewable energy sources.

Moreover, as stated previously, the share of renewable energy resources has to increase given the depletion of conventional energy resources and climate change mitigation. In addition to the drawback of RES of exhibiting large fluctuations in power generation, RES have lower energy density compared to conventional resources. The energy density is typically factor 100-1000 lower than that of fossil fuels [47]. This implies that a larger area for energy generation from RES is required. Limiting this effect in the residential built environment leads to another incentive for reducing the building energy consumption as much as possible. To serve this purpose, nZEBs have been studied in an extensive body of literature analysing its performance by means of a variety of indicators such as the heat exchange with the ambient, total energy end use etc [7], [48], [49], [50], [51].

Nevertheless, different studies have shown a dominant discrepancy between the simulated building performance and actual performance of this type of high efficient building concepts [22], [23], and [24]. This discrepancy is believed to be the direct result of the lack of occupant behaviour incorporations in the simulations [36], [22], [23], [24]. Therefore, building energy performance is increasingly more approached from a demand response perspective. Demand response has therefore been studied from a variety of angles to acquire more understanding in the characteristics and extent of this factor on the building energy performance and consequently, grid interaction.

The next sections will elaborate further on the influence of the occupant behaviour on the discrepancy between the simulated and measured energy performance of nZEB concepts and its implications of demand response.

2.1. Occupant behaviour

Branco et al (2004) [60] studied a low energy multi-family complex data of three years that indicated a discrepancy between a predicted yearly gas consumption of 160 MJ/m^2 and measured gas consumption of 246 MJ/m^2. Moreover, in the design phase a solar coverage of 31% of the thermal energy supply was predicted whereas in reality solar energy covered only 19% of the thermal load. Subsequent detailed calorimetric simulation indicated the large difference to be, to a large extent, due to the lack of incorporation of the real conditions of utilisation. The study of Santin et al, 2009, [24] finds that the influence of the occupant behaviour affect the energy use by 4.2%. Similarly, Haas et al (1997) [62] attributes the difference between predicted and measured energy demand to a so called rebound effect with respect to space heating. Taking the indoor temperature as an indicating parameter for the occupant behaviour, a linear relationship has been found between the indoor temperature settings and the thermal quality of the building. Hence, indoor temperature comfort requirements tend to be lower for occupant living in a poorly performing building which supports the theory of a rebound effect in high efficiency buildings. In addition, the authors conclude that this rebound is a direct effect of the lower prices of the services due to increasing energy efficiency. Another possible explanation for this phenomenon could lay in the frugal energy use for space heating in poorly insulated buildings. Since it is harder to capture the heat in all spaces in these buildings, the different spaces in these buildings are heated only when needed and not simultaneously which leads to less energy use.

In another study conducted by Zachary et al (2010) [22], investigated this phenomenon with focus on socio-psychological research to include the social character of occupant behaviour. The study found that the reason behind the demand for higher temperatures in high efficiency buildings, vary from lack of technical understanding of the high efficiency installations to different comfort standards to lack of adjustment to new habits that these installations impose. The influencing parameters could be further extended to whether the building was privately owned by the occupants or rented as found by Leth-Petersen and Togeby (2001) [62]. The energy consumption of rented dwellings is higher than for owned dwellings. This was linked to the energy cost that is usually included in the rent whilst homeowners have full insight in the periodic energy consumption. Both these studies indicate the relationship between the occupant engagement and the energy consumption. In addition, the studies discussed so far, state many different factors influencing the occupant behaviour which indicates that it is rather stochastic of nature and should ideally be predicted by data-driven methodologies based on a sufficiently large dataset of measured indicators.

2.2. Demand response in nZEBs

It is proposed here that the optimal method to achieve high building performance efficiency, is not by controlling occupant behaviour such as is aimed for in most Demand Side Management (DSM) methodologies but to construct a control strategy for the building energy systems that would optimize the energy systems operation efficiency based on occupant behaviour data. This methodology needs therefore to capture the diversity, variability and randomness of occupant behaviour. Since HVAC is the largest portion of building energy demand, the relationship between occupant behaviour and HVAC has been widely studied by different methodologies [24], [46], [63]. It is therefore interesting to focus on other controllable components of the building energy demand. After space heating, both the total electricity use (lighting and appliances) and the DHW follow in significance. Lighting and appliances are less suitable to be controlled for optimization for onsite renewable generation for the instantaneous nature of electricity does not provide means for buffering unless when supplied with electrical storage. DHW on the other hand, could be easily controlled due to the inertia associated with heat transfer which provides an option for energy buffering. This feature is utilised in storage tanks which therefore are often applied to DHW appliances to increase their efficiency. Hence, DHW control is considered as a suitable approach for improving self-consumption of onsite generated renewable energy. It is therefore evident that increasingly more electricity-to-heat appliances are used for this application [34], [70].

Recent studies have been conducted to analyse the potential of DHW controllability in demand response management. The study of Arteconi et al (2011) [34] applies a rule-based control system simulation to an air source heat pump coupled with thermal energy storage (TES) tank, based on thermal

models in TRANSYS. Although the control strategy used for this study does not simulate accurate occupant behaviour, this study is interesting for analysing the potential of heat pumps in reducing energy demand. The most notable results concern the difference between the control applied to a floor heating and a radiator. The large volume of the floor heating increases the thermal inertia of the system diminishing the need for a TES tank. The total energy consumed in the radiator model with a TES tank was 4% higher than for the floor heating model. However, the presence of a TES provides more controllability than the floor heating system. Thus, it is expected that utilising TES tank heat pump systems is a better strategy to perform load shifting. This emphasizes the importance of utilising a TES tank to improve thermostatic controllability of the DHW demand and optimize for the renewable energy self-consumption. Boait et al (2012) [64] conducted an empirical study to compare the operation efficiencies of different domestic water heating systems which amongst other include water immersion heaters and ground source heat pumps. These efficiencies are studied by means of applying different control strategies. Interestingly, the heat pump efficiencies showed a relatively strong dependency on the control strategy and the maximum power capacity. The maximum power capacity of the heat pump implemented in a building is determined by the size of the load demand of the building to increase the overall efficiency of the building. This implies that for a building with a high-energy performance, the maximum power capacity of the heat pump is minimized. Due to this low capacity, the heat pump will need more time heat up a unit water volume per unit of time. As a result, the water temperature set point hysteresis (which represents the threshold for the control system to be activated), is adjusted to a low value to maintain comfort. The low water temperature set point hysteresis results in an increase in heating cycles even when there is no active request for this hot water. This study points out the importance of an efficiency favourable control strategy of the heating appliance implemented. It is therefore a good practice to improve the control strategy by incorporating the occupant behaviour as the comfort is highly dependent on the activity of the occupant in the building; during the night, the occupants will have different comfort standards than during the day.
In the next section, the different types of control strategy methodologies used in literature will be analysed and discussed.

2.3. Optimization control strategies

The three most important demand response methodologies are listed below [71]:

1. Rule-based
2. Model-based
3. Model-free (learning-based)

Rule-based control strategies comprise strategies based on a set of simple controls in the form of *if, then, while* conditions. These sets of algorithms have the advantage of being relatively simple and therefore less computationally-intensive. Generally, it requires no forecasting but is rather based on domain and case specific parameters that have to be tuned manually. This leads to its disadvantage of being hard to generalise for systems with different parameter setting. Additionally, since it does not rely on optimization, it is expected to be less suitable to be applied to load shifting/matching purposes. Most importantly, the objective of the load shifting at stake is to incorporate the occupant behaviour in order to acquire more accurate results concerning the nZEB and heat pump energy performance. This is however not applied in rule-based control algorithms and thus is expected to reduce its effectivity compared to data-driven approaches. Nevertheless, this type of algorithms is applied to certain case studies that put less emphasize on the occupant behaviour.

The second group of control algorithms are made up of Model predictive control (MPC). MPC methods make explicit use of a mathematical model of the process dynamics [71]. Since these strategies rely on accurate models of the system dynamics, they provide more accurate results. On the downside of this characteristic, the model dynamics have to be modelled with great precision which could pose a limitation to its application as the model for the process dynamics could be only partially or completely not understood.

The main difference between the MPC and model-free methods is that model-free methods are completely based on data and do not require a model of the system dynamics. This gives it an advantage over MPC as it requires less work and is generalizable. The models in MPCs are case-specific and need to be adjusted when system environment changes. Renewable energy on-site self-consumption is regarded most effective if applied on a large scale. If the model parameters have to be adjusted from one case to another, MPCs will pose a limitation to large-scale implementation of the optimization [36]. Moreover, important for accounting for the stochastic nature of occupant behaviour, is to apply a control algorithm that would learn to improve its behaviour by getting more data rather than deploying a fixed control based on a physical model of the environment and the occupant behaviour. Examples of applications of these three types of control optimization methods on residential loads are presented in the next section.

2.4. Optimization control strategies application studies

The study of Ijaz Dar et al, 2014 [67] applies a rule-based control on a similar system setup with a storage vessel of 500 litre storage capacity. The results show an improvement of 6% for self-consumption and 11.5% for the self-sufficiency on annual basis. Another study conducted by De Coninck et al 2013 [72], applies a rule-based control strategy resulting in an electricity consumption reduction of 3.4% compared to the baseline control on cluster level. These results seem to be less effective than necessary to mitigate the expected increase of the grid interaction of nZEB concepts. More results of rule-based methods are discussed in the review paper of Luthander et al, 2015 [68]. The authors present figure 4 in which an overview of different DSM studies based on model-based and rule-based scheduling results is shown. The results show a wide range of outcomes depending on the location of the conducted studies, the type and size of the PV supported system, the type of building and the considered shiftable appliances. These studies have in common that they do not use battery capacity to store power and perform load shifting instead. The overall conclusion of these studies is that the DSM approach lacks effectivity in improving onsite self-consumption to a great extent, compared to systems in which battery storage and no DSM is incorporated.

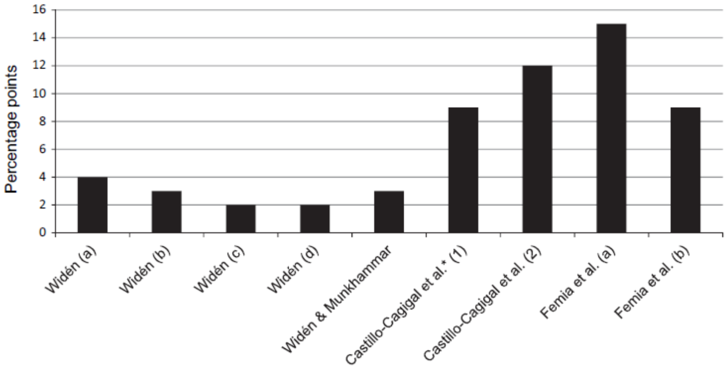


Figure 4. Different results of increase of self-consumption using load shifting found in other literature. [68]

Studies conducted by MPCs, on the other hand, show significant improvements. In a study conducted by Sossan et al, 2013[73], the research aim at optimizing on-site solar energy consumption for a single element electric heater by using a grey-box MPC strategy without including a model for the occupant behaviour. The results show a simulated increase for a spring day in May in Denmark of 293% of the baseline self-consumption. The study of van Houdt et al, 2014 [82] found an average rise in self-consumption of around 8-29% and self-sufficiency of 5-25% compared to the baseline simple control by using an MPC approach. Moreover, grid interaction reduced on average by 15% for the energy injected to the grid and 4.5% of the energy extracted from the grid. These results show that MPCs indeed perform better than simple rule-based control as it's expected given the theoretical differences. As mentioned before, MPCs have poor generalisability potential which gives them a disadvantage compared to learning-based methods.

Learning-based control strategies applied for load matching purposes are not a novel concept for building loads. These methods have been applied for a wide variety of objectives concerning the control of residential loads to facilitate the operation of smart grids. A majority of such studies focusses on learning control policies minimizing building energy operation cost for both generation to consumption [78], [79], [80], [81]. The results are directly reflected in the reduction of the energy bill ranging between 16% and 47% for these studies.

RL is also applied to improving the energy efficiency of thermostatically controlled loads such as heat pumps such as in the study of Peng and Morrison, 2016 [77]. This study compares MPC based controlled strategy to a RL based strategy showing a reduction in the energy consumption by 10% for the summer and 6% for the winter when using an RL control which performed 2% less optimal than the MPC. The study of [36] using RL show similar results of 15% reduction of the energy consumption compared to the baseline control. These results show promising results for applying RL for learning and optimization purposes. Yet, it is also interesting to focus on the long-term benefits of applying this type of control strategies such as the benefit of reducing grid interaction of heat pumps and RES in nZEBs for the utility grid. The generalisability potential of RL makes it suitable for large-scale applications, facilitating the two-way communication between grid operators and buildings which is believed to reduce immediate grid reinforcement cost.

The quantification of reducing grid interaction is, therefore, generally expressed in terms of the avoided grid reinforcement cost. This cost varies according to the location and the size of the system to be upgraded. A study by ECN in the Netherlands conducted by Pruisen and Kamphuis, 2010 [69] investigated the potential of a decentralized control of power consumption and generation technology (PowerMatcher) on reducing the grid reinforcement cost. The study is based on a case study of a neighbourhood with 1400 houses all provided with ground source heat pumps for HVAC and DHW purposes. The study investigates two extreme scenarios in which all heat pumps simultaneously require full capacity either as a result of a previous black-out or an extremely cold day. Such extreme events require the wiring to be 2-6 times larger in size with two distribution stations pf 630 kVA per 124 houses for that location. The projected costs of the location independent grid components are 20.000 euro per distribution station and a distribution cost of 5 Euro/kVA. The study assumes a total reduction of 50% of the total required capacity increase as a result of the PowerMatcher. This implies therefore a location independent cost mitigation of 250 thousand Euro for 1400 houses provided with heat pumps. These results are achieved by spreading the same energy consumption volume over the whole day rather than at specific moments. Nevertheless, the results illustrate the extent of the effect of the heat pumps on the grid reinforcement cost. Given the projected cost reduction for 1400 houses as in the study of ECN, a yearly reduction of the DHW load could therefore effectuate a substantial cost reduction for the grid given that the DHW load comprises around 14% to 30% [30] of the end use of an average residential building.

It is therefore interesting to investigate the potential of using RL to optimize for on-site self-consumption by means of the DHW load. RL methods are not only promising for aggregation purposes but allow the occupant behaviour to be incorporated in the assessment. The next section will introduce the case-study to which the research will be applied and elaborate more on the system setup.

3. Case study: nZEB neighbourhood in Soesterberg, the Netherlands

The housing market in the Netherlands exists of a relatively high percentage of around 30% of dwellings owned by housing corporations [42]. When considering district level energy efficiency improvement, this factor is of great help as the housing corporation could act as the project leader and integrate a large number of dwellings in the process. Governments can improve low energy performance neighbourhoods at once in a shorter amount of time compared to a more privatized building sector. One example of such projects in the Netherlands is conducted by housing corporation Portaal in the city of Soesterberg. "De Stroomversnelling" is a project implemented to refurbish low performing dwellings into nZEB concepts [43]. This is done by applying a high performing façade to the existing one and supplying the houses with each 28 solar panels. The houses are also fully disconnected from the gas network and operate therefore only on electricity. This project is therefore very interesting to investigate the effect of the increasing electrification of the built environment on the grid. Currently, BAM, Alklima/Mitsubishi Electric and Enervalis are collaborating in this project to accommodating these buildings with the necessary technology for a smart grid [44]. The heat operation schemes are investigated by BAM and Enervalis. This thesis research is conducted in this context in collaboration with Enervalis to investigate the possibilities for DHW demand response to optimize for onsite solar energy consumption by means of control applied to the ASHP.

3.1. System setup and data

An air source heat pump (ASHP) is installed to supply the hot water and special heating. The ASHP has a COP ranging between 3 and 5 for a temperature range of 45 ℃ and 50 ℃, depending on the outside temperature. The detailed specifications are presented in appendix B. Secondly, a hot water storage vessel of 200 litres is installed in the dwellings which is only used for DHW purposes. This vessel has a heat loss of 1.99 kWh per 24 hours in idle state which corresponds with -8.6 ℃/24hrs. The third system component are the 28 solar panels with each peak power capacity of between 4.5 and 5 kW_p.
The data is measured in uniform intervals of 5 minutes which could be subsampled if necessary. This data is gathered from four different sensors each measuring a different parameter:

1. A hot water flow meter registers how much hot water is flowing at a certain rate
2. A smart meter registers the power consumption of the heat pump at different modes and the power delivered by the PV panels
3. An ambient temperature sensor
4. A temperature sensor positioned midway in the storage vessel.

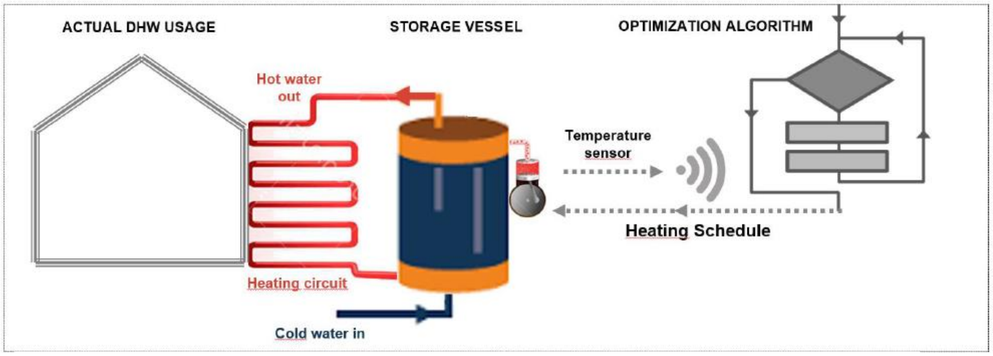


Figure 5. Schematic overview of the experimental setup [36].

Available data comprises output parameters from the sensors for 6 different houses starting from the first of June. The setup is presented in figure 5. as presented in Kazmi et al, 2016 [36].
The data is measured and sent to a central repository in which the baseline control and/or the optimization algorithm is embedded. The individual control action is calculated based on the embedded control strategy and subsequently sent as an action signal to the corresponding ASHP per house.

3.2. Baseline control and comfort standards

The default control strategy is a simple rule-based control algorithm given by eq. 1 [36]:

$$\mathbf{a}_t = \begin{cases} 1, if\ T_m < T_{th} \\ 0, if\ T_m \geq T_{th} \end{cases} \qquad \textbf{Eq. 1.}$$

In which $\mathbf{a}_t$ is the action taken by the heat pump represented by 1 when it is heating and 0 when it is off. The ASHP also controls the HVAC operation alongside heating the water in the storage vessel. Other modes (2 and 3) of the HP involve therefore HVAC operation modes. The dataset shows additional mode 6 indicating the Legionella heating cycle which happens regularly. Modes 2, 3, and 6 are therefore non-controllable actions for the DHW operation. Considering the temperature in the vessel, T_m is the temperature measured by the midway sensor in the vessel and T_{th} is the threshold temperature set point which is usually set to 45 – 50 °C [36].

The comfort standards applied by BAM dictate that if T_{th} is lower than 45 °C, the ASHP initiates a reheat cycle. Additionally, the upper limit for T_m was set threshold of 50 °C to which T_m increases during a reheat cycle. Additional comfort standards determine the water content of the vessel containing the threshold temperatures. A minimum of 50 litres with temperature T_m of at least 45 °C , was defined by BAM as a requirement to maintain occupant comfort.

The PV output is not considered in the default control strategy as the main goal of that control strategy is to comply to the comfort standards. The solar energy delivered is only considered for compensating the energy required from the grid in order to comply to the nZEB concept on annual basis [43].

3.3. Optimization objectives

The main objective for this study is to shift the DHW energy consumption towards high solar energy production hours in order to reduce the energy required during low production hours. It is also important to maintain a balanced ASHP consumption to not affect the nZEB concept by either violating the occupant comfort or a large rise in the energy consumption. Two main optimization indicators for this objective, are defined as the relative change in self-consumption and self-sufficiency compared to the baseline conditions. The following figure illustrates the two concepts of self-consumption and self-sufficiency:

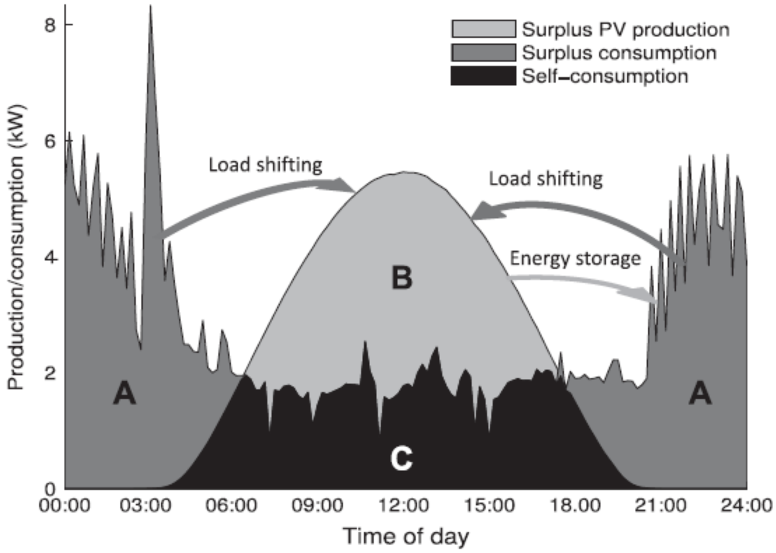


Figure 6. Schematic overview of the daily net load (A+C), net generation (B+C), and absolute self-consumption (C) [69].

Self-consumption is defined as the energy of the load that was covered by the solar energy expressed in either in kWh in absolute terms or as a percentage of the total generated energy. Self-sufficiency on the other hand, is defined as this same energy covered by the solar energy expressed either in kWh in absolute terms or a percentage of the total load. Additionally, the illustration in the figure shows the concept of load shifting which is defined as the total load allocated from low to high energy generation hours. In contrast, the concept of energy storage describes the excess generated energy by the renewable source that is stored to be used later than at the time of generation.

The optimization aims at optimizing the default control strategy of the case study by refining the control to a more sophisticated approach in which several thermodynamic factors as well as occupant behaviour related factors are accounted for by extracting the necessary data. One of the important factors to be optimized for, is the actual demand rather than a default threshold. At times when the occupants require less hot water, for example during the night, the threshold could be lowered. Another important factor is the coefficient of performance (COP) of the heat pump which is higher when the ambient temperature ($T_{ambient}$) is high. $T_{ambient}$ will then be taken as signal to determine the required energy to heat up T_m by 1 degree. Coupling of the ambient temperature might therefore provide optimization possibilities. Optimizing for solar energy consumption requires the PV production output to be an additional variable considered in the analysis in contrast to the baseline control.

These objectives will be approached by using RL methods which will be described in the methods section followed by the application to the problem described.

4. Methods

Machine learning is a subfield of computer science that deals with pattern recognition in data [39]. It gives computers the ability to learn without being explicitly programmed. This happens through the construction of algorithms that can learn and take decisions based on data rather than following programmed instructions such as in rule-based modelling. Different methods exist within this fields of study amongst which RL is one of these methods [40].

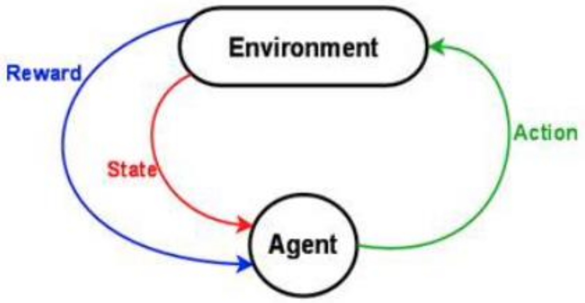


Figure 7. Agent – Environment interaction RL [41]

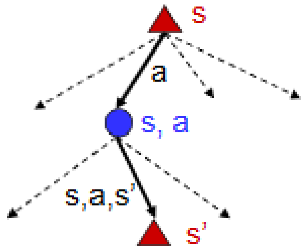


Figure 8. State transition from state s to the next state s'[41]

RL is a collective term for an area of machine learning methods that build upon algorithms that learn how to behave optimally by interacting with the environment. The learning algorithm is called the *agent* whereas the *environment* is represented by *states* which value is influences by external parameters governing it. learning mechanism is established through *rewards* (positive or negative) representing how good a particular state is in which the environment landed after the agent took a certain *action*. Figure 7 illustrates schematically the interaction between the environment and the agent.

Applying this theory to the heat pump and storage vessel issue, the state is represented by the temperature of the water in the storage vessel. This temperature is arrived at as a result of thermal and physical interaction of the water body with the external environment. When the heat pump takes the action to reheat, the environment is then interacted with and the temperature of the storage vessel increases i.e. the storage vessel reaches a different state. This state is a function of the amount of withdrawn water per unit time, the heat loss to the environment, and the action taken by the heat pump; either remain idle or reheat. The algorithm learns the optimal sequence of actions also referred to as *policy* (π) by observing the rewards it receives taking a certain policy. The reward is determined based on whether or not the state falls within the comfort levels of the occupants. Each cycle starting from the initial state to the state that marks the end of the cycle is called the episode.

4.1. Types of reinforcement learning algorithms

RL algorithms that satisfy the *Markov property* are called the Markov Decision Processes (MDP). The Markov property assumes that the current state is independent of the path that leads to that particular state. Hence, in Markovian problems a memoryless property of a stochastic process is assumed. In practice it means that the probability distribution of the future states depends only on the current state and not on the sequence of events that preceded. This is a useful property for stochastic processes as it allows for analysing the future by setting the present

An MDPs consist of state (**s**), action (**a**) sets and given any state and action to be taken, a transition probability function of each possible next state (**s'**) illustrated in figure 8. In addition, each taken action to arrive to the next state is rewarded giving each of all possible actions a reward value depending on the type of action. Each visited state is accredited by a value given to it according to a value function V(s) which represents how good it is for an agent to be in a given state. The value of a state ***s*** under a policy **π** is then denoted as $V^{\pi}(s)$ which in theory denotes the expected return when starting in state **s** and following a sequence of states to be visited according to the order defined in **π** thereafter. When this theorem is applied to a model-free control problem, the state-value function may not suffice as it does not show what action was taken for the state value to be acquired. Therefore, a similar function has been introduced representing an estimation of the value of each possible action in a state. This is described as the action-value function for policy **π** $Q^{\pi}(s,a)$. Figure 9 illustrates an example of the relationship between the action-value function and the state-value function. In 9.a. the action-values are shown for each direction of the propagation, North, East, South, and West respectively. The state-value function represents then the highest action-value possible in that state which is the action North in the example.

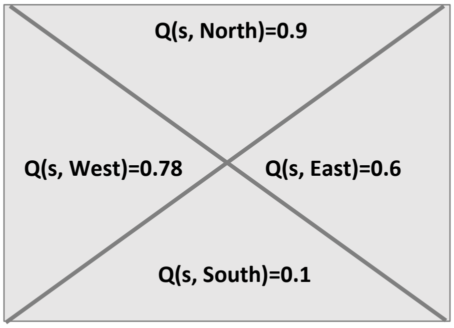


Figure 9.a. action-values of state s Q(s,a).

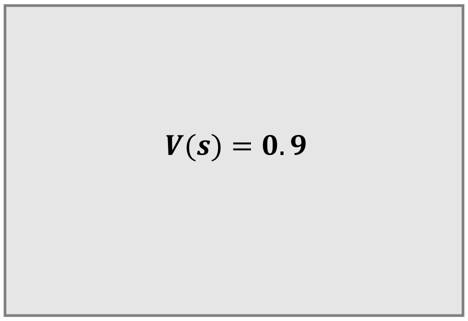


Figure 9.b. state-value of state s V(s).

The optimal policy is denoted as the superscript asterisk to the action-value-function Q(s,a) and state-value-function V(s). Formally, the optimal value function is then given by [40]:

$$V^*(s) = \underset{a}{max} Q^*(s,a) \qquad \textbf{Eq. 2}$$

Where Q*(s,a) is given by:

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')] \qquad \textbf{Eq. 3}$$

Herein, T(s, a, s') is the transition probability to the next state **s'** given state **s** and action **a**. **γ** presents the discount factor which is usually smaller than 1 and is used to discount for earlier values in order to assign higher values for sooner rewards. This is necessary to converge the algorithm.
Substituting equation 3 in 2 gives the Bellman equation [40]:

$$V^*(s) = \underset{a}{max} \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')] \qquad \textbf{Eq. 4}$$

These updates will be appended to the states that were visited resulting (after a significant number of iterations) in state values showing how good to be in that state. In order to be able to choose between the states to select a policy, as many states as possible need to be visited in order to converge to an accurate estimation of the state value. Acquiring the highest reward depends on these visited states and the reward accumulated. However, in order to discover more states and potentially higher rewards, the agent needs to take actions it has never taken before. This is referred to as the trade-off between exploitation and exploration. This trade-off could be achieved by setting a variable denoted as Epsilon (**ε**) which gives the extent of exploration versus exploitation. A fully exploiting policy is referred to as an epsilon-greedy policy and holds a value of 0 for **ε**. Correspondingly, a fully exploring policy gives a value of 1 to **ε** and is referred to as an epsilon-soft policy. The learning can therefore be tuned between these two extremes in order to allow for convergence towards an optimal value by occasionally exploring new states and actions.

4.2. Classes of RL algorithms

RL knows three fundamental classes of methods for solving these learning problems:
1. Dynamic Programming (DP)
2. Monte Carlo methods
3. Temporal-difference learning

Dependent on the problem at stake, each of these methods could be more suitable than the other. DP methods are model-based and require therefore a complete and accurate model of the environment i.e. all the aforementioned functions of the environment need to be known to initiate learning. However, the environment is not always defined prior to the learning process which poses a challenge to this method. This is where the two other model-free learning methods come in handy. The Monte Carlo algorithms only require an experience sample such as a data set in which the states, actions and rewards of the (simulated) interaction with the environment. In comparison with DP methods, no model of the transition probability function is required and neither the dynamics of the environment. Monte Carlo algorithms solve the RL problem by averaging sample return of each episode. Only after the termination of an episode, that the value estimation and policies are updated. Hence, it is based on averages of complete returns of the value functions of each state. This class of algorithms does not exploit Markov property described before and is therefore more efficient in non-Markovian environments [65]. On the other hand, Temporal-Difference methods do also not require a model of the environment but are like DP solving for incrementing step-by-step rather than episode-by-episode. Hence, TD methods exploit the Markovian property and perform usually better in Markovian environments.

The choice between these two classes of model-free RL algorithms very much depends on the type of data set available. For continuous processes in which there are no fixed episodic transitions, Monte Carlo methods may not be the optimal solution as they average the return only at the end of each episode. TD algorithms might then be a better solution as they assign a reward incrementally over each state. This allows them to converge faster towards an optimal policy for large data sets with a large number state spaces.

4.3. On-policy and off-policy TD control

TD algorithms comprise two important RL classes of algorithms divided in Off-Policy and On-Policy TD control algorithm classes. The difference between the two lays in the policy that is learned from the simulation or set of experiences (data). On-Policy TD control algorithms are often referred to as SARSA algorithms in which the letters refer to the sequence of State, Action, Reward associated with the state transition, next State, next Action. This sequence is followed in each time-step and is used to update the action-value of these two states according [40]:

$$Q_{(s_t,a_t)} \leftarrow Q_{(s_t,a_t)} + \alpha[r_{t+1} + \gamma Q_{(s_{t+1},a_{t+1})} - Q_{(s_t,a_t)}] \qquad \textbf{Eq. 5.}$$

Here, α represents the step-size parameter which functions as the exponentially moving average parameter. It is especially useful for non-stationary environments for weighting recent rewards more heavily than long-past ones. This could also be illustrated by rearranging the above equation to [40]:

$$Q_{(s_t,a_t)} \leftarrow (1-\alpha)Q_{(s_t,a_t)} + \alpha[r_{t+1} + \gamma Q_{(s_{t+1},a_{t+1})}] \qquad \textbf{Eq. 6.}$$

If α is a number smaller than one for non-stationary environments which indicates that recent updates weight more than previous ones. This transition happens after every nonterminal state. The $Q_{(s_{t+1},a_{t+1})}$ components of every terminal state is defined as zero. Hence, every terminal state has an update value of 0.

SARSA is called an on-policy algorithm because it updates the action-value-function according to the policy it is taking in every step. Therefore, it takes the epsilon-policy into account in order to arrive the optimal policy for a certain problem. Off-policy algorithms approximate the best possible policy even when that policy is not taken by the agent. Hence, Off-Policy algorithms base the update of the state-action-value function on the assumption of optimal behaviour without taking into account the epsilon-policy (the chance to take a negative action). Figure 10 shows a suitable example given by Sutton and Barto (1998) [40] and which illustrates the policy outcome differences between the two types of TD algorithms. The cliff represents states with high negative reward. Since SARSA takes the epsilon policy into account, it learns that at some instances a non-optimal action will be taken which results in a high negative reward. Hence, it will learn to take the safe path rather than the optimal path. Q-learning algorithms on the other hand, will take the optimal path by which the highest total reward could be achieved. This is because it does not take the epsilon probability into account of taking an extremely negative action. This class of algorithms is denoted by the following equation:

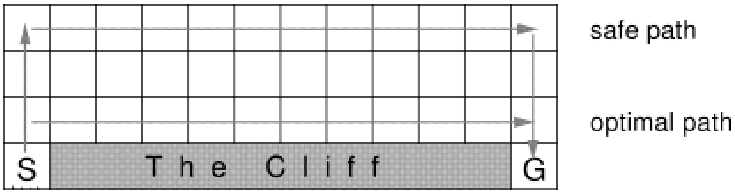


Figure 10. Cliff-walking example (Sutton and Barto, 1998)[40]

$$Q_{(s_t,a_t)} \leftarrow Q_{(s_t,a_t)} + \alpha[r_{t+1} + \gamma \overset{max}{a} Q_{(s_{t+1},a)} - Q_{(s_t,a_t)}] \qquad \textbf{Eq. 7.}$$

This difference will inevitably influence the suitability for the type of application. Applying these methods to the problem of this paper, in which occupant comfort is at stake, taking the stochasticity of occupant behaviour into account is crucial for complying to the comfort standards of the occupants. It is therefore argued that SARSA is more suitable for this type of application compared to Q-learning to avoid operating along a path with a probability of breaching occupant comfort even if that path would lead to the highest total reward.

4.4. Eligibility traces

In essence, the difference between MC and TD is the number of steps over which V^{π} is backed-up with the observed rewards. The *one-step* back-up TD class of algorithms could be seen as one extreme and the *episodic* MC class of algorithms as the other extreme. The back-up steps of the TD algorithms could be increased according to the most suitable number of back-up steps applicable for a certain data set type. Additionally, an update could be composed of an average of different n-step updates. This approach is represented by the forward view of TD(λ) class of algorithms which average the return of the different n-step backups according to:

$$R_t^{\lambda} = (1-\lambda)\sum_{n=1}^{T-t-1}\lambda^{n-1}R_t^{n} + \lambda^{T-t-1}R_t \qquad \textbf{Eq. 8.}$$

The backward view of the TD(λ) algorithms applies this theorem by introducing a new factor representing the frequency of occurrence of a certain state. This is referred to as the eligibility trace given by:

$$e_t(s) = \begin{cases} \gamma\lambda e_{t-1}(s) & if\ s \neq s_t \\ \gamma\lambda e_{t-1}(s) + 1 & if\ s = s_t \end{cases}$$

γ represents the discount factor as presented before and λ represents the trace decay parameter. Together the two parameters add a trace to the visited state-action pair to include how frequent and how recent a state-action pair is visited. This makes recently and frequently visited states attribute more to the total update. In contrast to the one-step TD methods in which only the last action taken in a sequence is emphasized, the eligibility traces emphasize the whole sequence of taken actions. The level of emphasis is determined by the $\gamma\lambda$-factor. This makes TD(λ) algorithms more efficient than one-step TD or Monte Carlo algorithms.

4.5. Function approximation

TD(λ)-algorithms solve for model-free RL problems for a variety of applications in different fields of theory. Model-free RL is very much suitable to be applied for highly stochastic problems such as the heating patterns of heat pump and hot water storage vessels. However, the previously described algorithms hold a disadvantage for being tabular methods which refers to the table look-up fashion by which the control policy is determined. The best action to take is then determined from the state-action pair values **Q-matrix**, as shown in example table 1 in which a set of possible temperatures is defined as the states and the actions as a binary set of 0 or 1. The action with the highest Q-value is then chosen as the optimal action for that particular state.

State-Action pair matrix		Action	
		0 (ON)	**1 (OFF)**
States (Temperature)	**57.00**	1.19	0.28
	56.00	0.64	0.76
	55.00	0.64	0.50
	54.00	0.92	0.92
	53.00	-2.27	0.38
	52.00	0.65	0.55

Table 1. State-Action pair matrix (Q-matrix) example

However, a drawback to this tabular method is evident when the state-space increases significantly resulting in many possible states. The probability of a state to occur decreases accordingly which results in an unknown policy for unseen states. In the case of the control strategy for the ASHP, the states are made up of a combination of a set of variables such as the temperature, water consumption and the solar energy. Hence, the number of possible states is then equal to temperature state-space multiplied by the water consumption state-space and the solar energy state space. This results in significantly large state-set of which sizes depends on the measured data. The more variables to compose the state, the more possible states which reduces the chance for all states to be seen significantly. Consequently, the uncertainty in the control strategy increases which indicates a reduced generalisation potential. This problem is solved by applying function approximation to the seen states in order to approximate for surrounding unseen states. A state is then represented in a feature based representation describing the properties of that particular state. The Q-function then becomes [40]:

$$Q(s,a)=w_0 + w_1f_1(s,a)+w_2f_2(s,a)+.....+w_nf_n(s,a) \qquad \textbf{Eq. 9.}$$

Here, w_n is the weight of the features f_1 of the states. Features should not be too specific as they then challenge the generalisability. Hence, features should be carefully selected.

These methods allow for learning the optimal heating policy without the need to build the thermodynamic model of the system which in turn offers a higher generalisability potential as the learning algorithm could be applied to systems with different physical parameters. Furthermore, learning the optimal policy for the heat pump to operate in, could provide energy savings by adjusting the heating cycles and enhance in this way the building energy performance to reach the goal of an nZEB. Lastly, when in a grid-tied nZEB, PV power generation is used to power the heat pump, this learning algorithm could be used to optimize for onsite PV power consumption. This allows for a reduction in grid interaction which helps to integrate a larger share of renewable energy. These methods will therefore be applied in this thesis paper to examine the extent of these methods to meet these objectives.

5. Application to case study

The control algorithm that is aimed at in this study, is an algorithm that optimized for onsite solar energy consumption. The aim is therefore to utilise the hot water storage vessel similarly to a storage capacity for the PV energy. Hence, it is important to use as much PV energy as possible during the high solar energy production hours of the day. To achieve this, a SARSA(λ) as described previously is applied with a reward function that takes several parameters regulating the vessel state, into consideration. The overall optimization strategy planning is presented in the flowchart shown in figure 11.

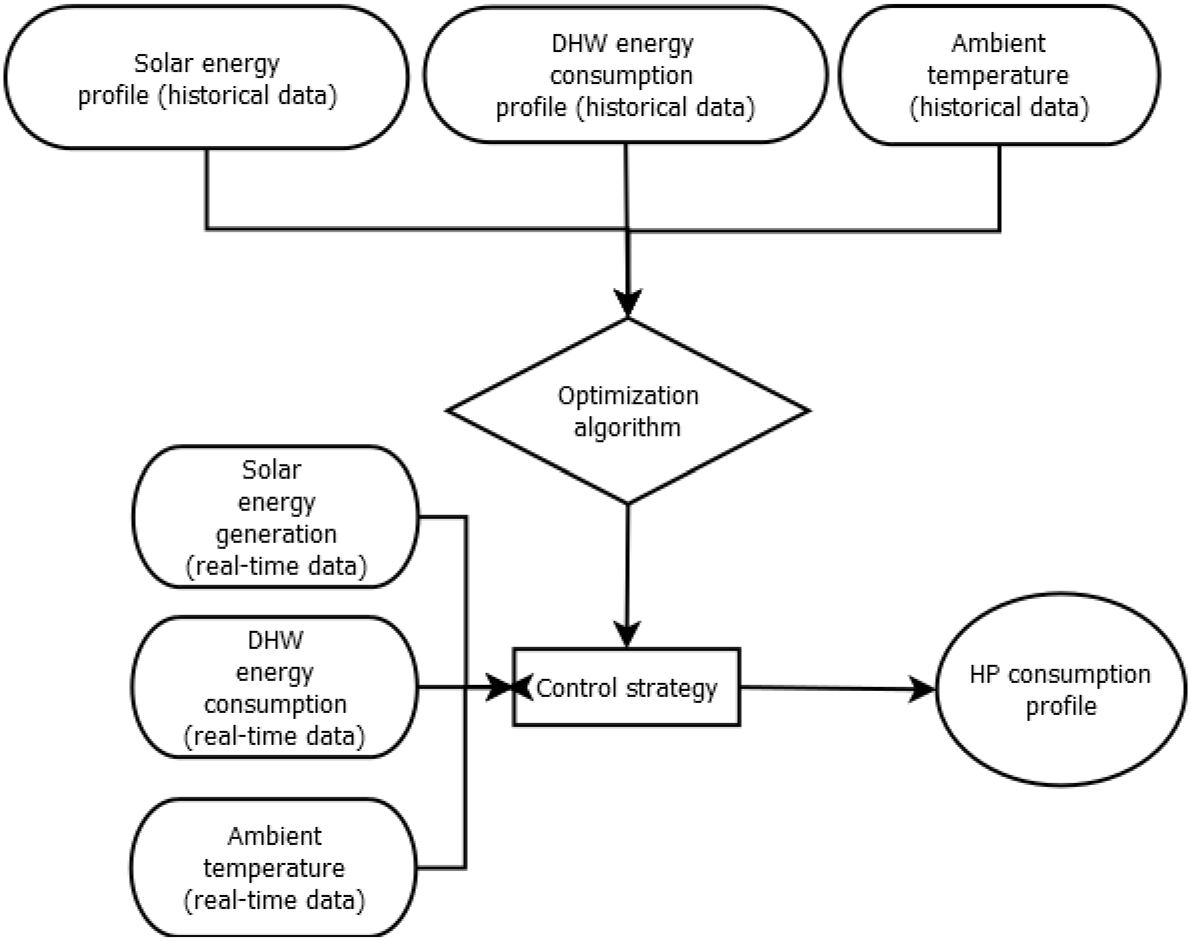

Solar energy profile (historical data)
DHW energy consumption profile (historical data)
Ambient temperature (historical data)
Optimization algorithm
Solar energy generation (real-time data)
DHW energy consumption (real-time data)
Ambient temperature (real-time data)
Control strategy
HP consumption profile

Figure 11. Optimization strategy flowchart

The next sections will present the pseudocode used to achieve this alongside a detailed description of the reward function and corresponding conditions applied to achieve the results.

5.1. Pseudo-code

```
Get:
Tm (°C), Energy (Wh), Mode (1,2,3,6),
Water consumption (Liter), and PV energy (Wh)

If 1< mode <6 (spatial heating):
        Mode=0

If vessel heating cycle is interrupted by spatial heating modes:
        Mode=1
If mode=6 (Legionella):
        Mode=1

Create episodes:
While mode=0:
        Episode is running
If mode=1:
        Episode ends

Get:

Cumulative water consumption in each episode

Create array with Wh/K/episode with linear regression model:

 $Y \leftarrow B + c_1 \Delta T_m + c_2 \sum L_t + c_3 T_{ambient}$
        $\frac{Wh}{°C} = Y$
```

```
Start Sarsa (λ):

  Initialise parameters:

  • $\alpha, \gamma, \lambda$, and $T_{limit}$
    (Temperature set point for a heating cycle)
  • maximum water consumption
    in episode ($L_{max}$)
  • action-space
  • state-space

  For all episodes:

  • $Q_{(s,a)} \leftarrow$ arbitrary
  • $e_{(s,a)} = 0$

  Repeat (for each episode):

  • $s_{(t)} \leftarrow T_{m(t)}, \ \sum Liter_{(t)}, Solar\ (Wh)_{(t)}$
  • $s_{(t+1)} \leftarrow T_{m(t+1)}, \ \sum Liter_{(t+1)}, Solar\ (Wh)_{(t+1)}$
  • $a_{(t)} \leftarrow Mode(t)$
  • $a_{(t+1)} \leftarrow Mode_{(t+1)}$

  Collect reward

  Perform update:

  • $\delta \leftarrow r + \gamma Q_{(s',a')} - Q_{(s,\ a)}$
  • $e_{(s,a)} \leftarrow e_{(s,a)} + 1$

  For all s and a:

  • $Q_{(s,\ a)} \leftarrow Q_{(s,\ a)} + \alpha \delta e_{(s,a)}$
  • $e_{(s,a)} \leftarrow \gamma \lambda e_{(s,a)}$
  • $s \leftarrow s'\ ; a \leftarrow a'$

  Until s is terminal

  For all Q(s,a) apply linear regression function
  approximation:

  • $Q(s,a) = w_0 + w_1 T_m + w_2 \sum L_t + w_3 Solar_t$
```

Figure 12. SARSA(λ) Pseudo-code applied to the case study

Firstly, the algorithm filters the controllable modes and their corresponding variables, T_m, Energy, Water consumption, and PV energy. Modes 2 and 3 have no influence on the vessel states and are therefore translated to mode 0. This results in representative episodes in most cases, however, this is not the case when the heating cycle (mode 1) is interrupted by these modes. The heating cycle is at times interrupted, giving HVAC modes priority when the indoor temperature comfort standards are at stake. Nevertheless, when this is completed, the heat pump will continue the water heating cycle until the upper threshold (51.5 ˚C) is reached. In these special cases, the interrupting modes are translated to action 1 so as to simulate the continuation of the heating cycle and avoid the compilation of false episodes. In contrast to the HVAC states, the legionella cycle which is indicated by mode 6, influences the vessel state and increases the water temperatures to around 66 °C, however, is not a controllable but rather a timed action. It has therefore been decided to replace these modes with action 1 to avoid the simulation of an increase in temperature and falsly appending a negative reward to certain states when action 0 is taken.

Once the correct modes and their corresponding variables have been filtered, the episodes are simulated accordingly. Each episode consists of the states during which the mode was 0 and ends with the state at which action 1 was taken. From this data, the features needed for the SARSA(λ) code, $T_{m_{(t)}}$, cumulative water consumption per episode $\sum \text{water}_{(\text{t})}$, maximum water consumption threshold L_{max}, solar energy $\text{Solar}_{\,(\text{t})}$, and energy required to heat up T_m by 1 degree, are compilated. Every state is made up of a combination of T_m, $\sum water$ and the solar energy production $\text{Solar}_{\,(\text{t})}$.

5.2. Reward function

The reward function applied in this algorithm is presented in a flowchart fashion in figure 13. This reward will be appended to each action 0 to be taken. Action 1 leads to a terminal state which will lead by definition to a standard update of 0 as described by section 4.3. Consequently, if the Q-value of taking action 0 is higher than 0, action 0 is more favourable than taking action 1 and vice versa. A drawback to the binary action-space, is that action 1 could not be punished or rewarded. However, this is solved by amplifying the negative or positive reward for not reheating (action 0) instead.

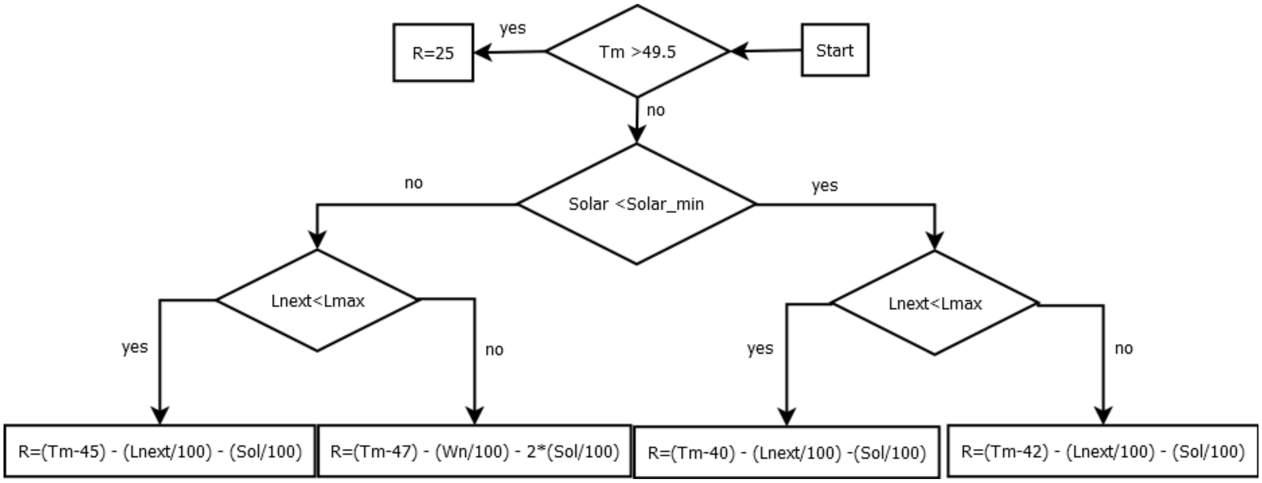

Start
Tm >49.5
yes
R=25
no
Solar <Solar_min
no
yes
Lnext<Lmax
yes
no
Lnext<Lmax
yes
no
R=(Tm-45) - (Lnext/100) - (Sol/100)
R=(Tm-47) - (Wn/100) - 2*(Sol/100)
R=(Tm-40) - (Lnext/100) -(Sol/100)
R=(Tm-42) - (Lnext/100) - (Sol/100)

Figure 13. Flowchart of reward function

The reward function is defined as a function of all three parameters, T_m, $\sum water$, and the solar energy Sol. The relative weight of these parameters on the reward is based on the scenario at stake as shown in figure 13. A drawback to the binary action-space, is that action 1 could not be punished or rewarded to avoid enhanced reheating events. However, this is solved by amplifying the negative or positive reward for not reheating (action 0) instead as shown in the functions in figure 13.

To prevent the vessel to keep constantly reheating when there is high generation of solar energy, a temperature of 2 °C below T_{start} is defined as the threshold temperature for the ASHP to reheat the vessel. This is equal to 49.5 °C when T_{start} is set to 51.5 °C such as in the case in the default strategy. Therefore, two main scenarios are created this way, one describing a midpoint temperature of higher than 49.5 °C and one lower. Temperatures higher than 49.5°C receive a constant high reward of 25 points to ensure an idle state. On the other hand, to fully exploit the solar energy when available, if the temperature falls below this threshold and the solar energy available is higher than that is needed for rising the temperature of the vessel to 51.5 °C again, negative rewards are favoured to enhance reheating.

To accomplish this, the energy needed to reheat from 49.5 °C to 51.5 °C for each individual device has to be determined and compared to the available solar energy. The unit of energy required to raise T_m by 1 degrees is determined through a linear model based on variables influencing the energy depletion and efficiency of the heat pump. These features comprise the total required temperature rise to reach the upper limit of 51.5 °C, ΔT, the cumulative water consumed during the reheat cycle $\sum water$, and the COP of the ASHP which strongly correlates with the ambient temperature T_a. The correlation between the energy required for a reheat cycle and the corresponding features is given in table 2.

Variable	**Pearson product-moment correlation coefficient (P-values)**
$\boldsymbol{\Delta T}$ **(°C)**	0.3380
$\boldsymbol{T_a}$ **(weather station)**	-0.073642
$\boldsymbol{\sum water}$ **(L)**	0.7713

Table 2. Correlation coefficient of selected variables with required energy per reheat cycle

This shows that DHW consumed is the dominant variable controlling the energy required to reheat T_m by 1 degree. The second dependent variable is the temperature increase between the end of one episode and the start of the subsequent one (ΔT). T_a is the ambient temperature as measured from a weather station close by Soesterberg. It is evident that the ambient temperature has a small negative correlation with the required energy. Hence, when the ambient temperature increases, the total required energy decreases slightly. This corresponds with the effect of the ambient temperature on the COP of an ASHP. However, in this case, this effect is small relative to the effect of the required ΔT and $\sum water$. These three features have been selected for to represent the states of the vessel. Linear regression has been applied to these features to determine the resulting energy unit required for the reheat cycles per:

Wh/˚C/episode = $w_0 + w_1 * \boldsymbol{\sum water} + w_2 * \boldsymbol{\Delta T} + w_3 * \boldsymbol{T_a}$ **Eq. 12**

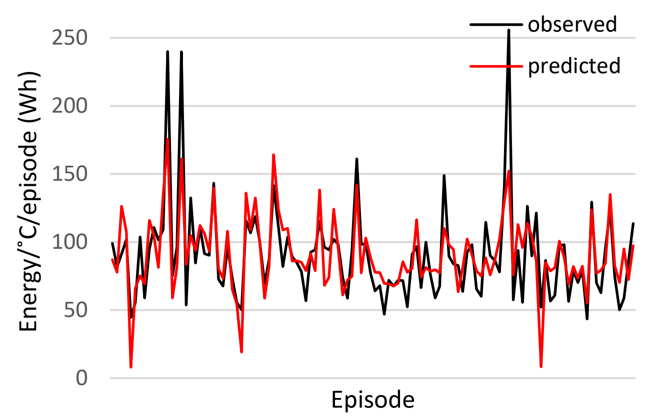


Figure 14. Observed vs. predicted Wh/˚C

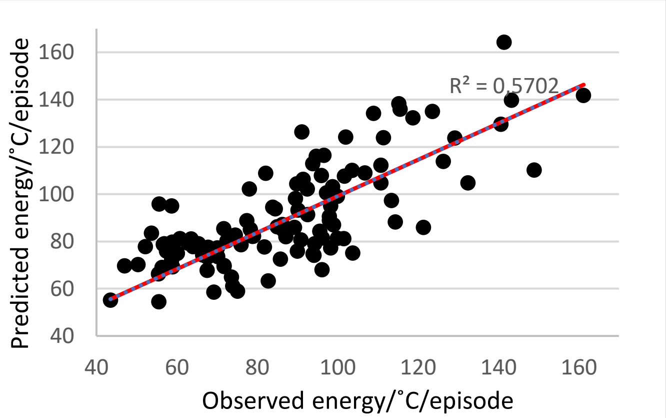


Figure 15. Linear correlation of observed vs. predicted Wh/˚C

The results are compared in figures 14 and 15. With a root square value of 57%, it is evident that the energy unit required to raise T_m by 1 degree is linearly correlated with the combined effect of $\sum water$, ΔT, and T_a with a. Hence, the linear model performs closely to the measured values and therefore, could estimate the episodic values of the required energy unit sufficiently adequate. This unit will allow for determining the difference between energy consumption and the solar energy production during the day to apply a corresponding reward. This model could be improved by adding the water temperature of the inlet of the storage vessel. This data is however not available in the current system setup.

When the temperature falls below 49.5 °C and the solar energy available is higher than what is needed to increase T_m by 2°C, the algorithm will append an increasingly negative reward to temperatures below 49.5 °C of which negativity depends on the amount of water consumed. For this purpose, the water consumption threshold is set according to the comfort standards determined by BAM of at least 50 litres of water of at least 45 °C at all times in the vessel. Assuming a constant temperature profile between the midpoint and the outlet of the storage vessel with a content of 200 litres in total, it is assumed that at least 100 litres of water with the same temperature as the midpoint is available in the tank. Therefore, the water consumption threshold is set to be equal to 100 litres. If the water consumed at time t is higher than this threshold L_{max}, a reward as a function of both T_m and the water consumption is appended according to:

$$R = (T_{next} - T_{min}) - \left(\frac{L_{next}}{100}\right) - (\frac{Sol}{100}) \qquad \textbf{Eq. 10.}$$

In which L_{next} represents the cumulative volume of water consumed and T_{min} represents the minimum midpoint temperature for that scenario. The solar and water parameters are divided by 100 to be normalized to the temperature range.

In each of the scenarios, the temperature is punished to a different extent by adjusting T_{min} so as to create a linear decrease in the reward for decreasing temperatures and increasing solar energy and DHW consumption. Also, in case of high solar energy generation in combination with a high DHW consumption, the weight of the solar parameter is increased to enhance reheating for these scenarios. Hence, reheat cycles are preferred when there is a sufficient amount of solar energy and T_m is more than 2 °C below T_{start}. In the scenario of temperatures above 49.5 °C, the appended reward is always positive.

This reward function is then applied to SARSA(λ) update which results in state-action pair values determining how favourable it is to reach a certain state at a particular action.

5.3. Simulation

The optimization algorithm will be applied to the vessel state model learned by the study of [66]. The vessel state model takes T_m to which the vessel was reheated, $\sum water$ since reheat, and the time passes since reheat in as parameters and returns the T_m for that state. This model will therefore serve to simulate the output of the algorithm. The comparison between the baseline control output and the optimized control output will be based on the same solar energy and water consumption profile derived from the data of June to September for the houses. This serves to compare the potential of reducing grid interaction with the Sarsa(λ) control in the same context. Henceforth, the optimization control strategy will be applied to three other months with lower solar energy output and a different water consumption profiles to test its generalisability. The simulation setup is shown schematically in figure 16.

The output of the vessel state model will be used as an input for the function derived by linear regression from the Sarsa(λ) algorithm along with the solar energy production and water consumed. If the Q-value is positive, the idle state is favoured and the simulation of the next time step will continue incrementally over time. However, if the Q-value reaches a negative value, then a reheat cycle is initiated. The vessel state model simulated only the idle state of the vessel which results in lack of knowledge of the duration and energy of the required reheat cycle. Two linear models derived from the data will therefore be modelled for determining the required time for the reheat cycle and the required energy. T_a, ΔT, and $\sum water$ at the start of the reheat cycle, will serve as input variables for both models. It is important to bear in mind that the accuracy of the simulation is dependent on the accuracy of the linear regression models presented.

Once the time required is determined, the total amount of water consumed during the reheat cycle could be determined along with the total amount of solar energy produced and the average T_a

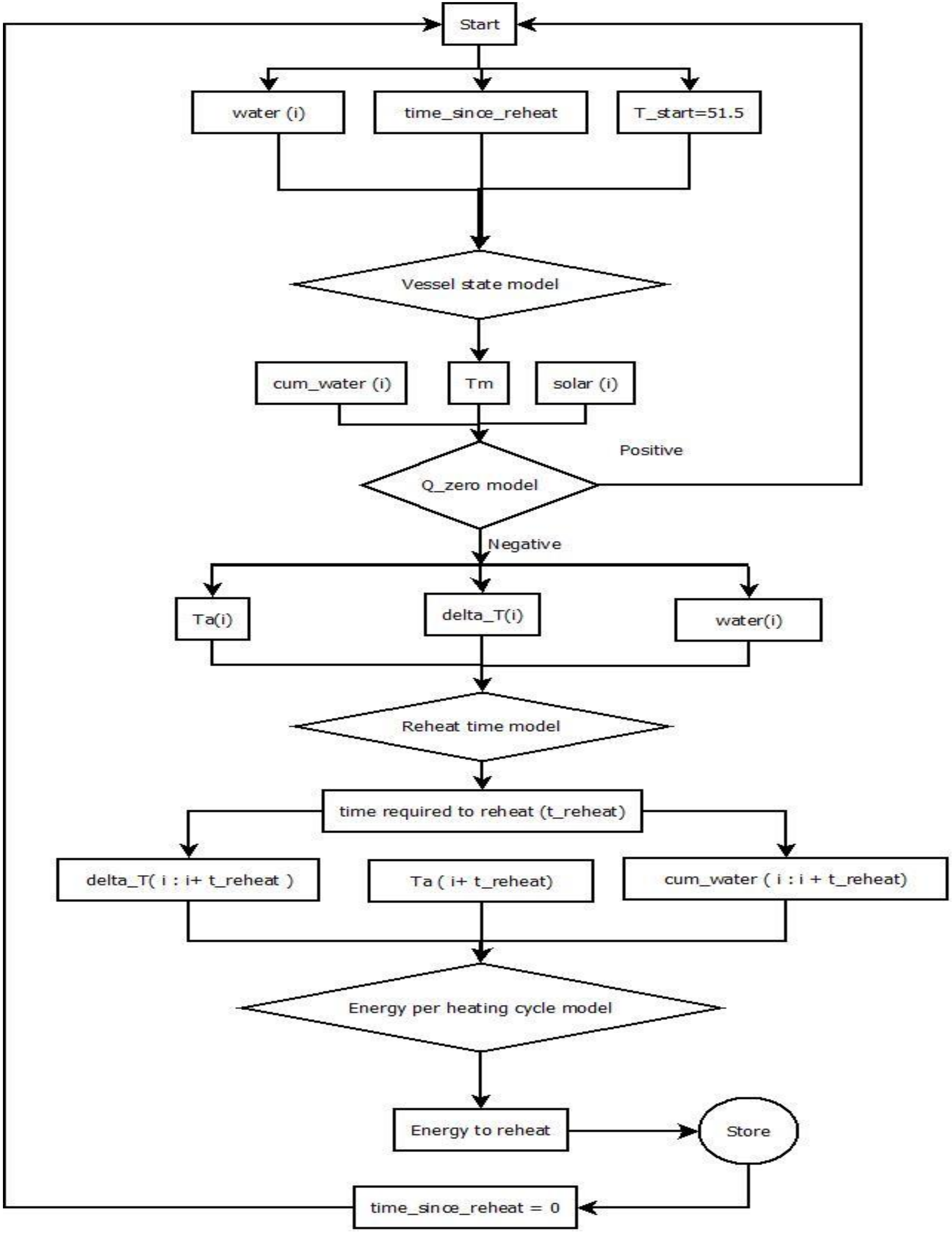


Figure 16. Simulation setup flowchart

during that period. T_a, ΔT, and $\sum water$, of the reheat cycle will then be used as input parameters for energy model to estimate the required energy for that reheat cycle. This could then be used to determine the extent of optimization.

5.4. Limitations

Perfect optimization is not assumed in this research as RL algorithms rely strongly on a substantial number of experiences. The available data set comprises data of 6 months which is relatively small to assume perfect optimization. The algorithm is therefore expected to be limited to a certain extent as a result. Additionally, in theory, optimizing for energy consumption during certain periods of the day could be achieved by dissipating power by substantially increasing the energy consumption of the ASHP. However, the energy efficiency of the nZEB is central to the optimization and is therefore taken as a limitation for the optimization. In other words, it is sought to find a delicate balance between the energy consumption of the ASHP and an optimized solar energy consumption. Another limitation posed by the performance of the nZEB is the occupant comfort. It is important to obtain optimal results without compromising the comfort of the occupants. Moreover, the former leads to perhaps the most important limitations which are posed by the uncertainties revolving around the type of occupant behaviour. As the occupant behaviour determines the amount and timing of the energy consumption, the optimization will be subject to this influence. Similarly, the amount of harvested solar energy is subject to seasonal and location related changes. As the case study is conducted with data derived from a location in the Netherlands, the results might vary when the algorithm would be applied to a location with a significantly different climate.

6. Results

6.1. Algorithm and model setup analysis

This section will provide a validation of the previously described optimization setup. Firstly, the Sarsa(λ) will be analyzed to validate its performance. Following this analysis, the linear models derived from the data will be analyzed and validated. This will allow for a performance analysis of the total model setup for analyzing the results of the optimization.

6.1.1. Algorithm analysis

The data-set applied to the algorithm comprises data from the first of June to the first of September as an example of a period of high solar production. The comparison will then be made with its performance during lower solar energy production months of September to December on cluster level. The data of house 1 will be applied to the Sarsa(λ) algorithm as an example for the validation since it shows an already significant concentration of consumption during high solar energy production hours. This allows for investigation of the potential of further optimization in such cases. Figures 17 shows the average daily total ASHP energy consumption profile including the legionella cycle scheduled every two weeks in the early mornings and the HVAC operation, against the average daily ASHP consumption profile of solely the DHW consumption for house 1. These profiles show that the ASHP is used primarily for DHW purposes in summer. Other slight energy use during the day may include floor heating purposes on cold summer days. Furthermore, the legionella cycle is an energy intensive process consuming significantly more energy than the other ASHP processes planned between 3 and 7AM.

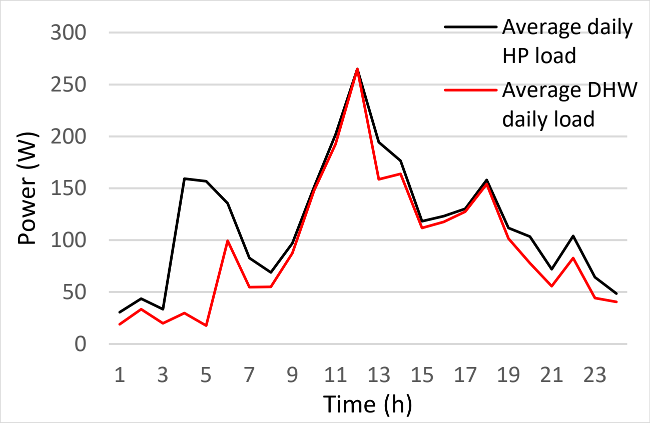


Figure 17. Average daily ASHP profiles June - September

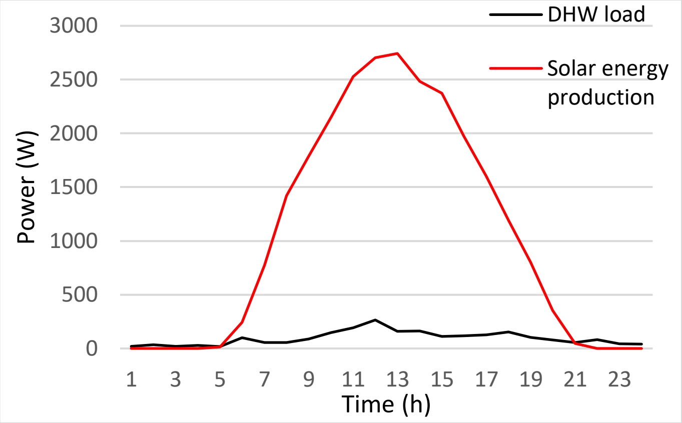


Figure 18. Average daily solar energy profile vs. DHW load profile June – September

Figure 18 shows that the solar energy profile is around 10 times larger than the DHW profile in the summer. Therefore, if the DHW profile is optimized such that the reheat cycles would fall as much as possible during solar energy production hours, the electrical energy required from the grid could reduce significantly. The excess solar energy injected to the grid will reduce accordingly, providing additional flexibility.

6.1.1.1. Q-value analysis

The Q-value function (indicating how good action 0 is; §4.5) results for the summer months for house 1 are shown in figure 19. A negative value (blue) suggests the start of a reheat cycle whereas a positive (green/orange) value favors the idle state. It is evident from the figure that temperatures higher than 51.5 ˚C have always favor the idle state. Temperatures between 45 ˚C and 51.5 ˚C favor the idle state when there is no to low solar energy production and decreases in value as a function of the total water consumed and the solar energy produced.

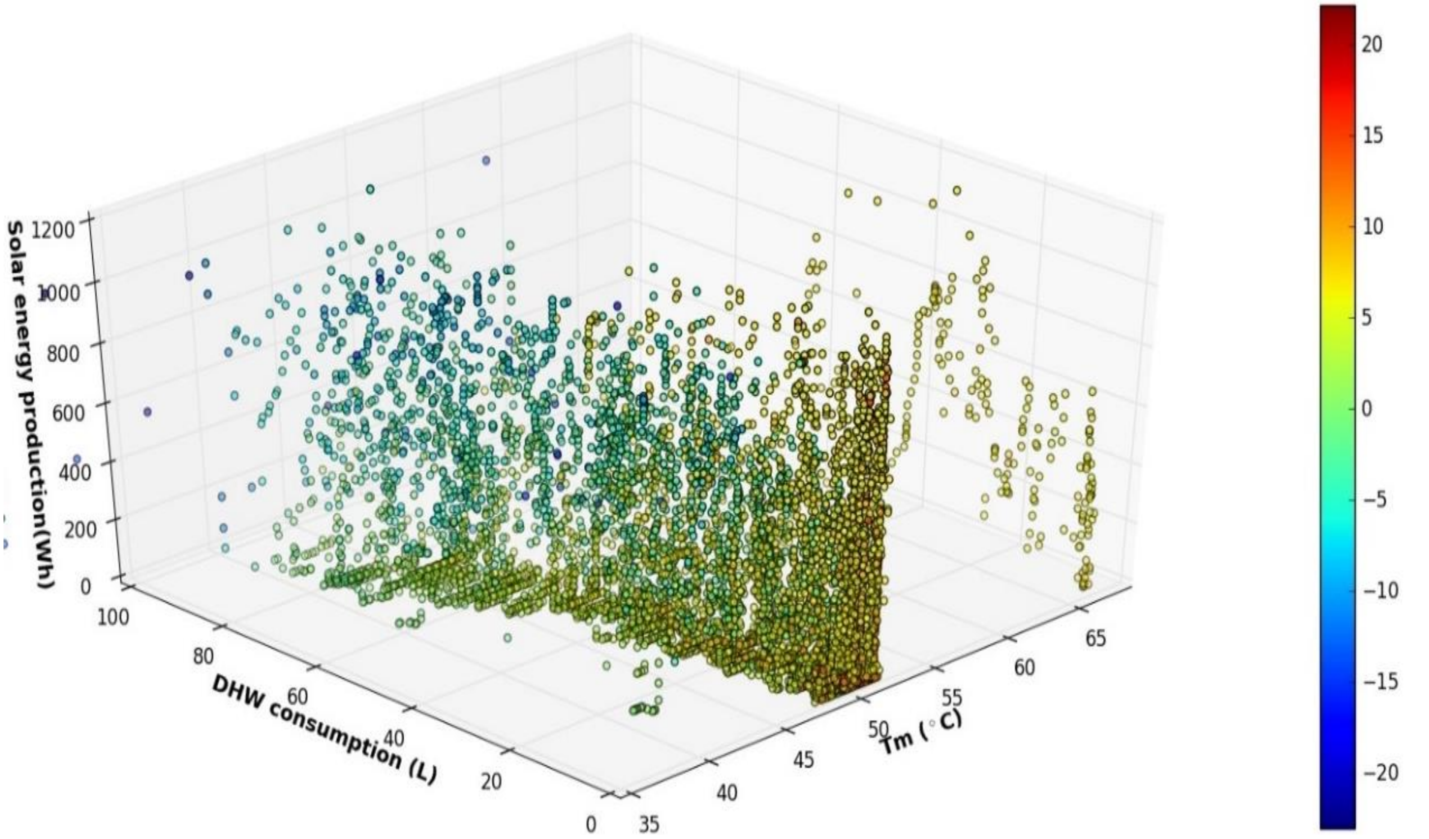


Figure 19. 3D representation of the Q-values of house 1 summer months' data

The most negative Q-values (most left corner) represent the least favourable states for the idle state. This shows that the algorithm learns to favour reheating when there is high solar energy generation. This effect is amplified when there is DHW consumption in the dwelling that causes the temperature to drop significantly. The Q-values in the most right corner of the plot show the most stable states which represent the states with high temperatures above 51.5℃.

The influence of the reward is visible this plot as the Q-values follow the strategy set in the reward function. Deriving from these results, it can be concluded that the algorithm learns the desired behaviour to enhance the heating cycles during high solar energy generation events while considering the occupant DHW use. This could also be shown by analysing the relationships between the Q-values and the individual variables. The individual relationships of each of the variables $\sum water$, T_m, and solar energy output with the Q-function are shown in figures 20a-c. The midpoint temperature shows the highest linear correlation with the Q-value as it is weighted more than the other two variables in the reward function. It is also evident from the three plots that two different relationships with the Q-value are prominent in the algorithm. This is also to be traced back to the reward function as it was defined with two main conditions representing the solar energy production higher and below the minimum required energy to raise T_m by 2 degrees Celsius. This number fluctuates around 300 Wh as shown in figure 22. This is also shown in figure 20.c. as the values under 300 Wh are primarily aligned around the first linear trend. Values above this 300 Wh fluctuate around the second linear trend which represents the second condition in the reward function. Whereas the Q-value increases with increasing temperature, it decreases with both increasing water consumption and solar energy production as assigned in the reward function. This is to aid the optimization for high solar energy production and water consumption hours.

Concerning the cumulative water consumption profile, figure 20.b. shows two linear trends until a consumption of 60 L after which the Q-value decreases at a higher rate. This is to be attributed to the co-linearity between the temperature and the water consumed in the vessel i.e. the higher the water consumption, the lower the temperature. Additionally, the higher the water consumption per time unit, the higher the mixing affects the state of the vessel which results in a non-linear relation as shown figure 21. In turn, this causes the relation between $\sum water$ and the Q-value to become non-linear at higher water consumption rates. Hence, the midpoint temperature is weighted higher than the $\sum water$ with regards to the Q-value. These results could alternatively be visualised as shown in figures 23.a-c. The algorithm shows an overall linear and positive correlation with the temperature on one hand governed by two different trends, one for T_m higher than 49.5 ℃ and one for temperatures below this value. The influence of T_m is weighted as the base trend to which the other variables contribute additionally.

The correlation with the cumulative water consumption $\sum water$ is a negative sub-linear trend based on its influence on T_m. Furthermore, the fluctuations in the solar energy production cause similar fluctuations in the Q-value of the algorithm additional to the overall trend based on the temperature. A trend of increasing/decreasing solar energy production results in a decrease/increase in the Q-values on top of the persisting trend of T_m's influence. This lowers the threshold for reheating during high

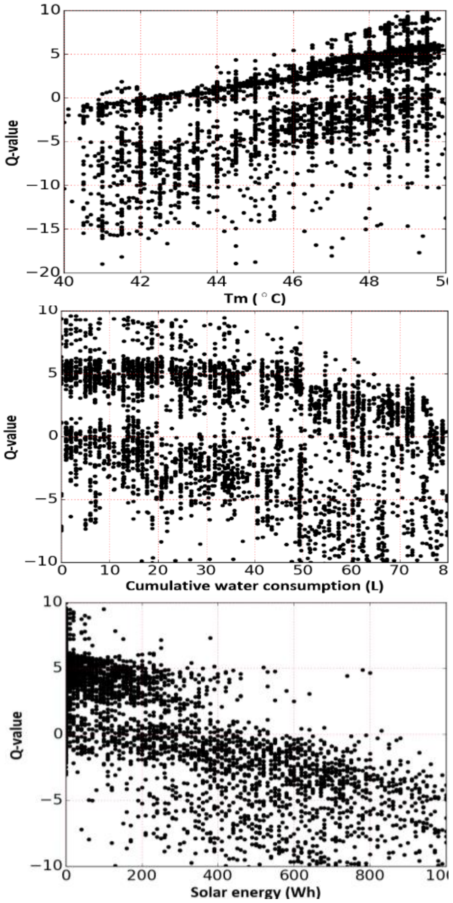


Figure 20. (a) Relation of $T_m$with Q-value. **(b)** Relation of $\sum water$ with Q-value. **(c)** Relation of Solar energy

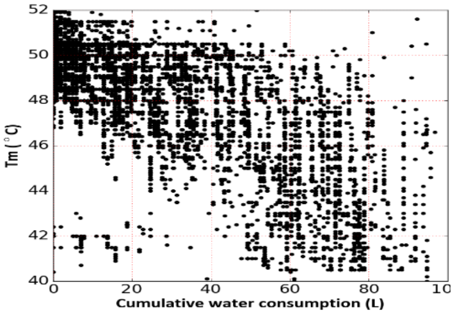


Figure 21. Relation of T_m vs. $\sum water$

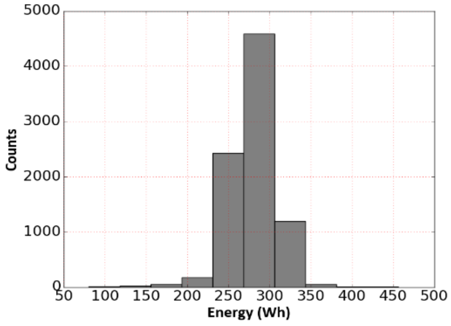


Figure 22. Histogram of minimal required energy to raise T_m by 2°C.

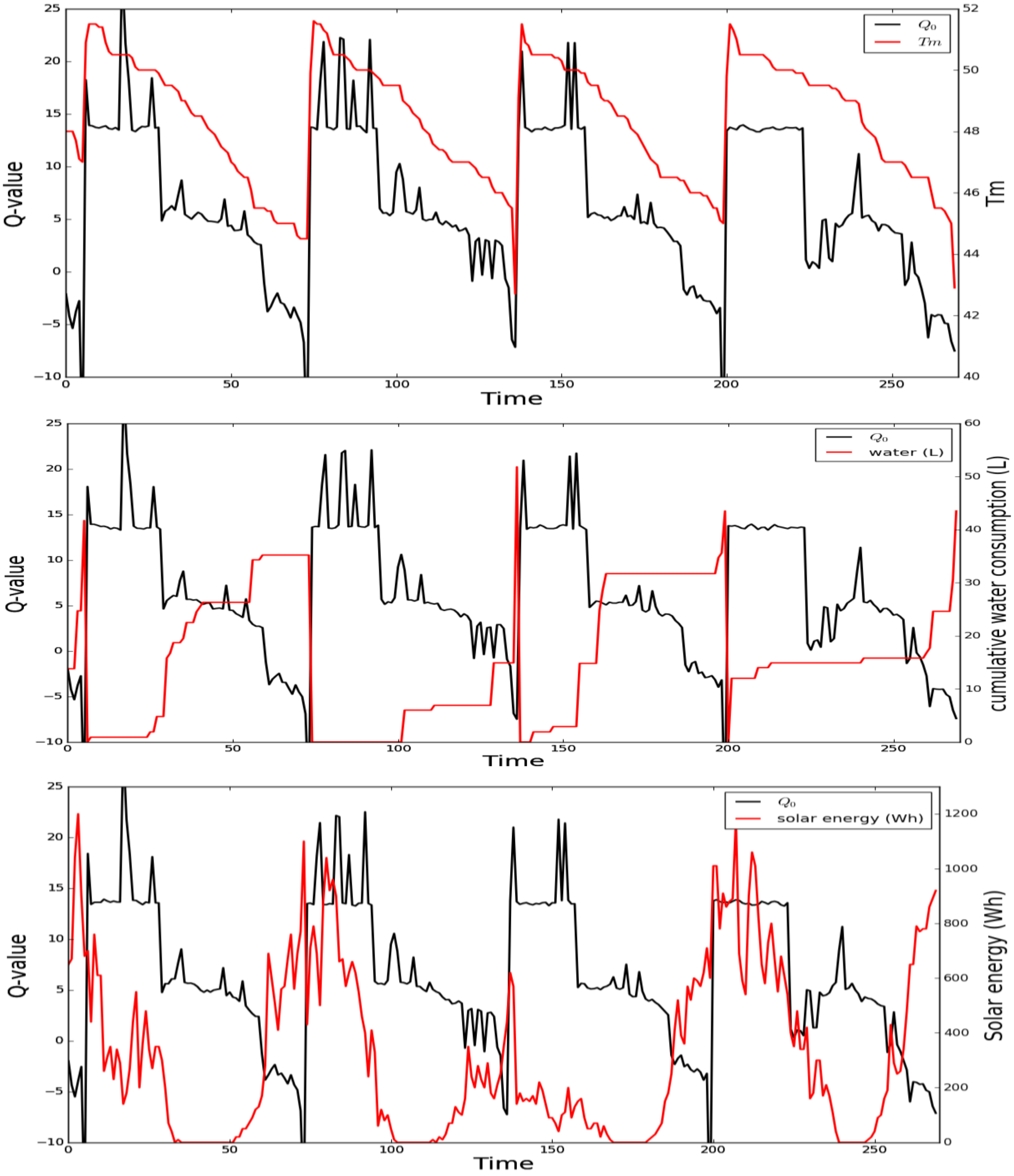


Figure 23. **(a)** T_m vs. Q_0 **(b)** $\sum water$ vs. Q_0 **(c)** Solar energy vs. Q_0

production hours and vice versa. The influence of the solar energy production is therefore larger than the water consumption and smaller than T_m.

These plots show that the control algorithm is learning the desired behaviour concerning the optimization for the onsite solar energy consumption and reduction of the grid interaction. However, as mentioned earlier, the tabular output of the algorithm does not generate Q-values for unseen states and reduces generalisability with increasing number of states. This problem is solved by applying function approximation to generate a function that follows the behaviour of the algorithm and estimates the Q-values for unseen states. The type of function is chosen based on the reward function. Since the reward function is a linear function, a linear model is generated to determine the Q-values output of the algorithm for all possible states. The linear model and its output will be further discussed in the next section.

6.1.2. Linear model analysis

Achieving generalisability with reinforcement learning methods requires the support of function approximation. That way, a general function is deduced by the algorithm's output as an example and thereby approximating the values of unseen states. The reward function built in the algorithm is based on a linear function which suggests that the output of the algorithm should vary accordingly. Therefore, linear regression is applied to approximate the function of the algorithm. The same features as the algorithm were taken as the variables of the linear function per:

$$Q(s,a) = w_0 + w_1 * T_m(s,a) + w_2 * \sum water\ (s,a) + w_3 * \text{Solar}\ (s, a) \qquad \textbf{Eq. 13}$$

Herein, w_1, w_2, and w_3 present the three coefficients to be approximated along with the intercept, w_0. The comparison between the Sarsa Q-values and the linear Q-values are shown in figure 24 and 25. The 3D representation of the approximated Q-values shows large similarities with the values of the algorithm shown in figure 19. The clearest differences lay in the transition from positive to negative values. The values of the Sarsa algorithm show sharp incline and decline in certain points in time whereas the transition of the linear model progresses in a smoother fashion. This difference is also visible in figure 25. Like the algorithm, the main course of the Q-values is primarily based on the progress of T_m.

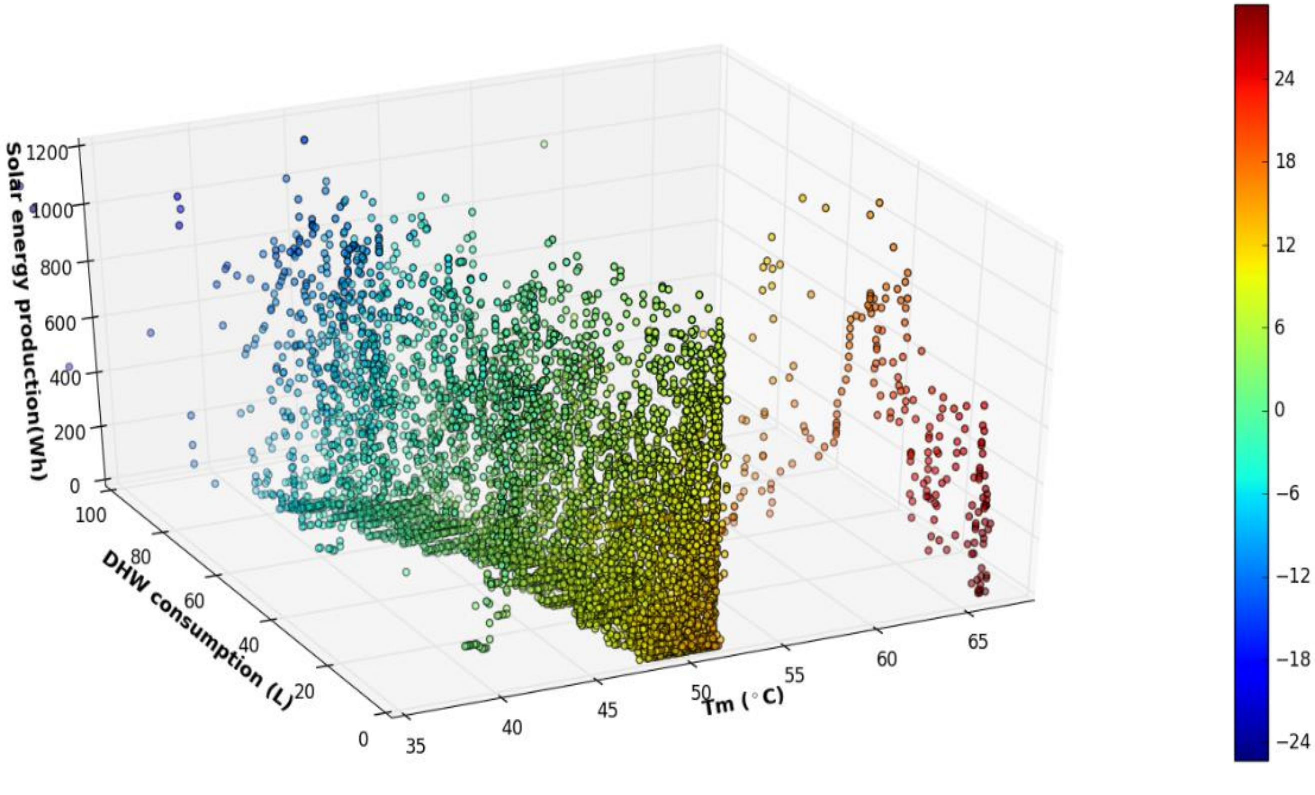


Figure 24. 3D representation of the linearly approximated Q-values of summer months' data of house 1.

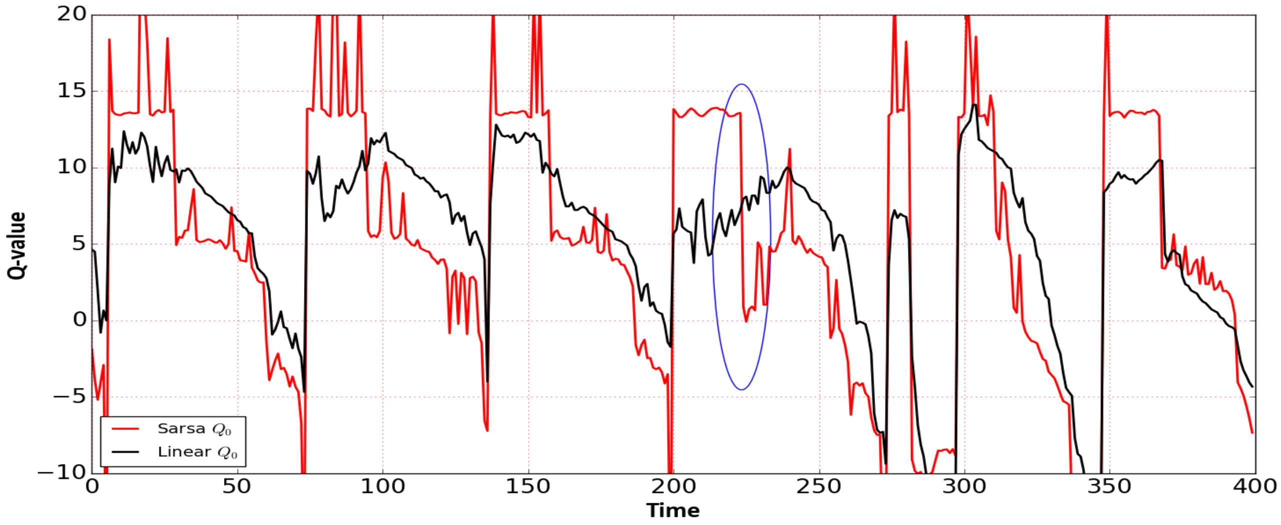


Figure 25. Sarsa Q-values vs. linearly approximated Q-values

The algorithm awards temperatures of higher than 49 °C with high stable rewards after which it declines to another policy based primarily on the solar energy production. As long as T_m is higher than around 45 °C and the solar energy production is too low, the Q-values remain positive and decline when the temperature decreases and solar energy increases. The linear model on the other hand, interpolates between these different policies to compose an overall linear policy. Therefore, the sharp declines are less evident in the linear model than in the algorithm. Hence, the Q-values of temperatures are less constant than in the linear model as they decrease and increase according to the weight of the solar energy and water consumed. This explains the increase of the difference shown, as an example, in the encircled area in which it clear that when the algorithm's output declined sharply, the linear output continued to increase steadily following the increase in solar energy production. Additionally, the fluctuations due to the solar energy fluctuation are damped significantly in comparison with the algorithm. This results also in a slight delay in the transition to negative values. However, this does not pose a drawback to the model as the delay remains in the comfort level temperatures of higher than 45℃ as shown in figure 26. In fact, the damping effect of the model could be considered as an improvement to the algorithm since it is less responsive to sudden fluctuations in the solar energy production.

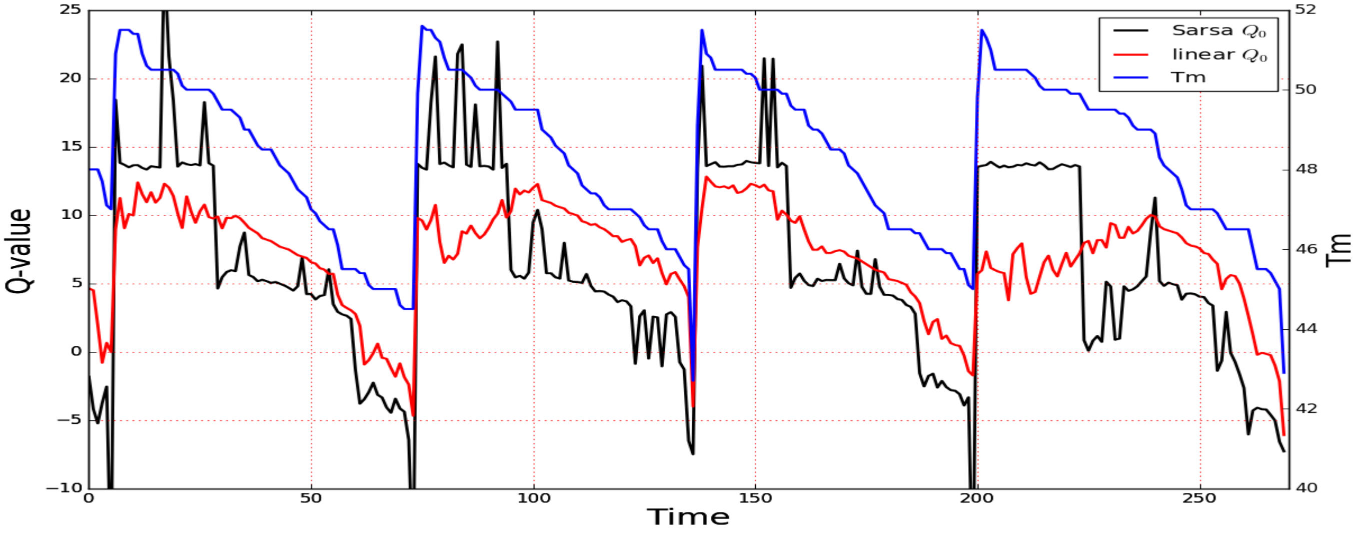


Figure 26. Comparison of T_m, Sarsa Q-values, and linear Q-values

These results show that the linear model for approximating the Q-values performs as aimed for and could even improve the robustness of the algorithm. This model will therefore be used for the simulation of the vessel state and the effect of the optimization on the timing and energy consumption of the reheat cycles.

6.1.3. Simulation setup analysis

Having established the approximated function of the Q-values, this function will be used as the optimization control strategy. This will be conducted according to the previously described setup (sec. 5.3). In order to accomplish this, two additional linear models have been derived from the data. The first model is required to determine the duration of the reheat cycle in order to determine the total water consumed during the reheat cycle. This will be required as one of the features required for the second model which is derived to determine the required energy for the reheat cycle. The comparison between the linear model for the reheat cycle time and the observed required time is shown in figure 27.

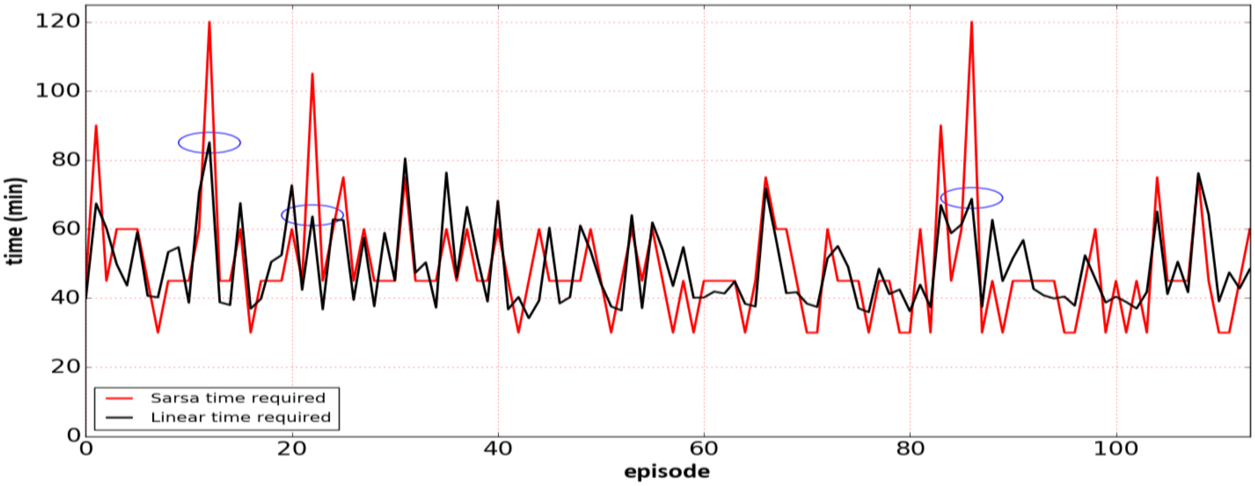


Figure 27. Time required for reheat cycles as observed vs. predicted

This model takes in ΔT, $\sum water$, and the ambient temperatures as input features and the time required per reheat cycle as output. Figure 27 shows that the model performs accurate for most of the cycles, however, underestimates high value events as shown in the encircled areas. This model underestimates for the same reheat cycles as the Wh/℃ model (fig. 14) underestimates for. Therefore, it could be stated that a feature other than ΔT, $\sum water$, and the ambient temperatures, is responsible for these high-energy cycle events. A possible additional variable of influence on the time and energy required for the reheat cycles could be the inlet water temperature. The inlet water temperature is somewhat dependent on the ambient temperature, however, is also influenced by the indoor temperature amongst other factors. No data is available for the water inlet temperature to improve these models, nevertheless, as for the energy model, the model performs sufficiently adequate on average to approximate the time and energy required for the reheat cycles.

The required time for the reheat cycle is used to sum the water consumed during every reheat event. This is then used as input feature for the required energy model to determine the total required energy based on the ΔT required to reach 50 ℃ and the average ambient temperature during that event. By this means, the energy to and from the grid could be modelled in order to be compared with the data for the same months of June, July, and August with the optimization control strategy. The results of this optimization will be presented in the next section.

6.2. Baseline versus optimization control

The results of the energy consumption and the average daily number of reheat cycles for house 1 are shown in figures 28.a. and b. respectively.

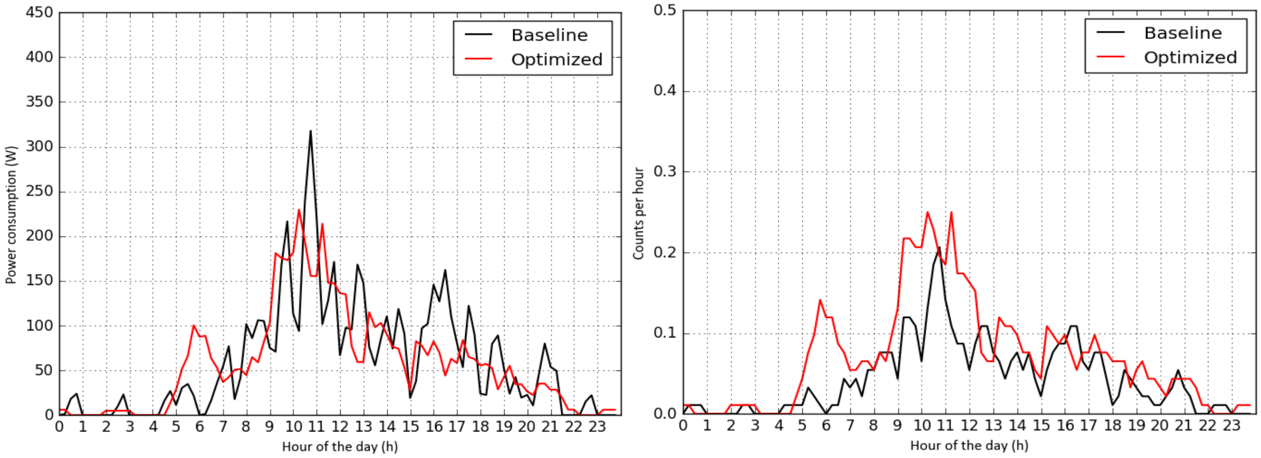


Figure 28. (a). Comparison of energy consumption baseline vs. optimized, house 1. **(b)** Comparison of average daily number of reheat cycles per hour, house 1.

House 1 has the highest daily average water consumption, with most of its daily average energy consumption covered by the produced solar energy (assuming all solar energy is available for DHW). The optimization algorithm improves this further by increasing the number of reheat cycles around noon and thereby reducing the energy required later in the day. It is clear from both graphs that the algorithm optimizes for onsite solar energy consumption and self-sufficiency. The number of reheat cycles is increased in general. A significant rise of the reheating cycles increases the energy consumption of the ASHP accordingly due to enhanced heat loss to the environment. Moreover, enhanced heating cycles increase the wear and tear of the ASHP. However, the reheat cycles in the last third of the day occur at higher temperatures due to the reheating earlier in the day, which reduces the energy required for the later reheat cycles. This behaviour could also be seen in the simulation of the other houses, house numbers 2-6. The following graphs present the energy and reheat cycles comparison.

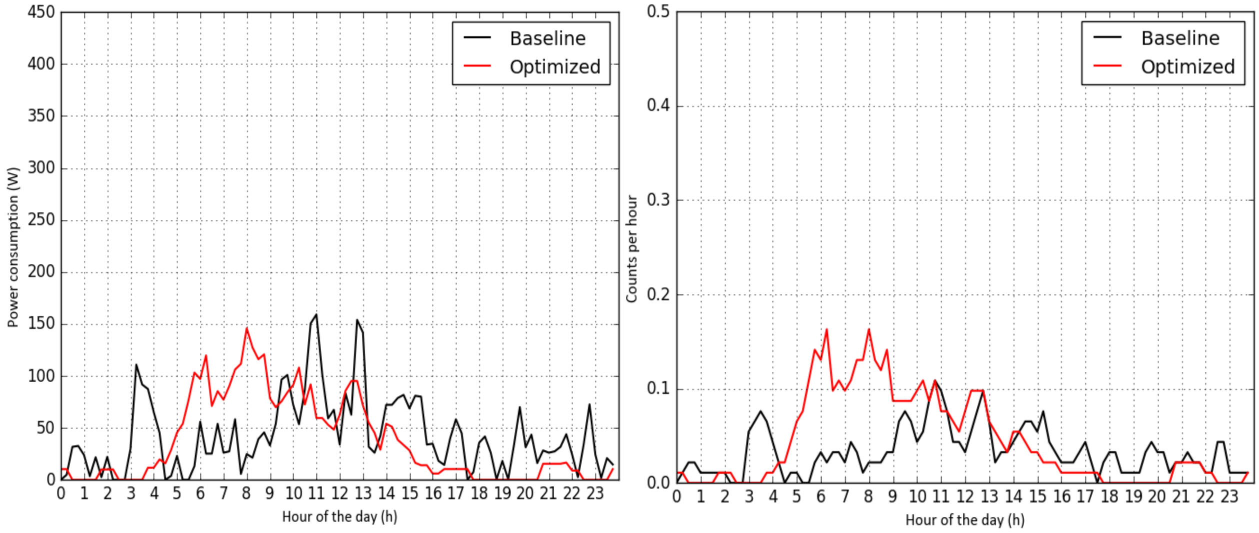


Figure 29. (a). Comparison of energy consumption baseline vs. optimized, house 2. **(b)** Comparison of average daily number of reheat cycles per hour, house 2.

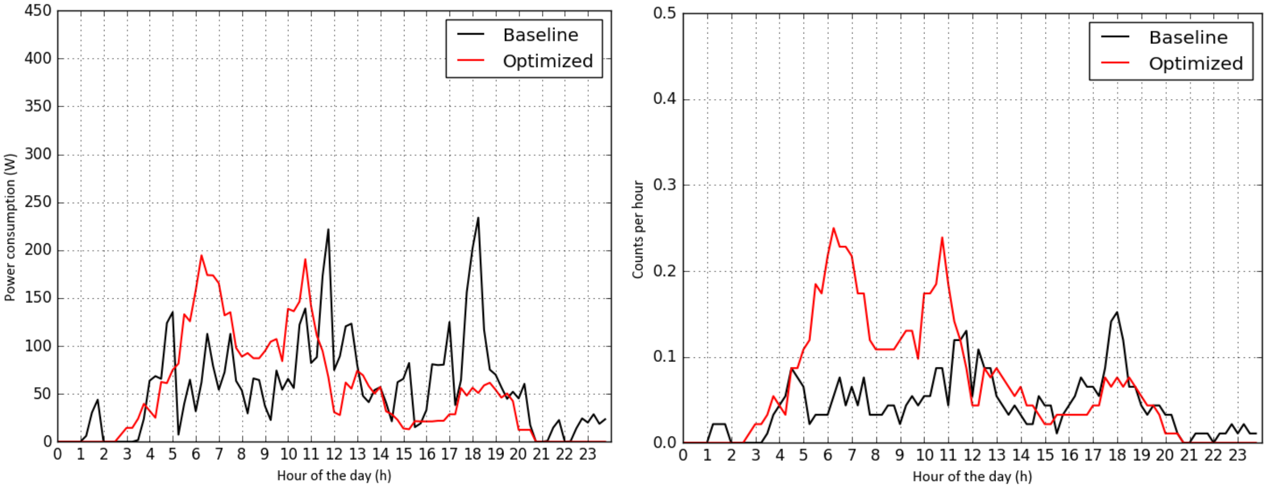


Figure 30. (a). Comparison of energy consumption baseline vs. optimized, house 3. **(b)** Comparison of average daily number of reheat cycles per hour, house 3.

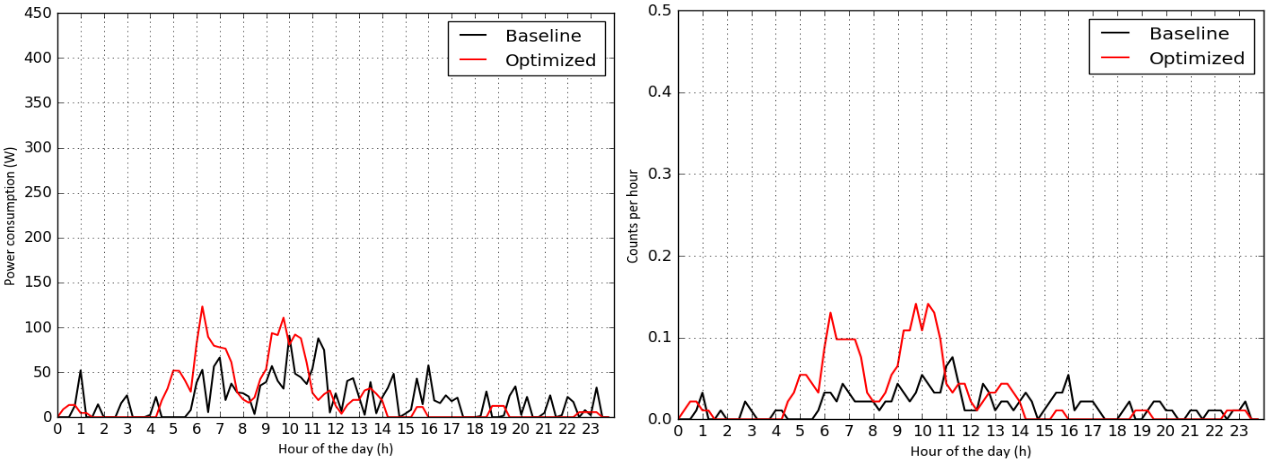


Figure 31. (a). Comparison of energy consumption baseline vs. optimized, house 4. **(b)** Comparison of average daily number of reheat cycles per hour, house 4.

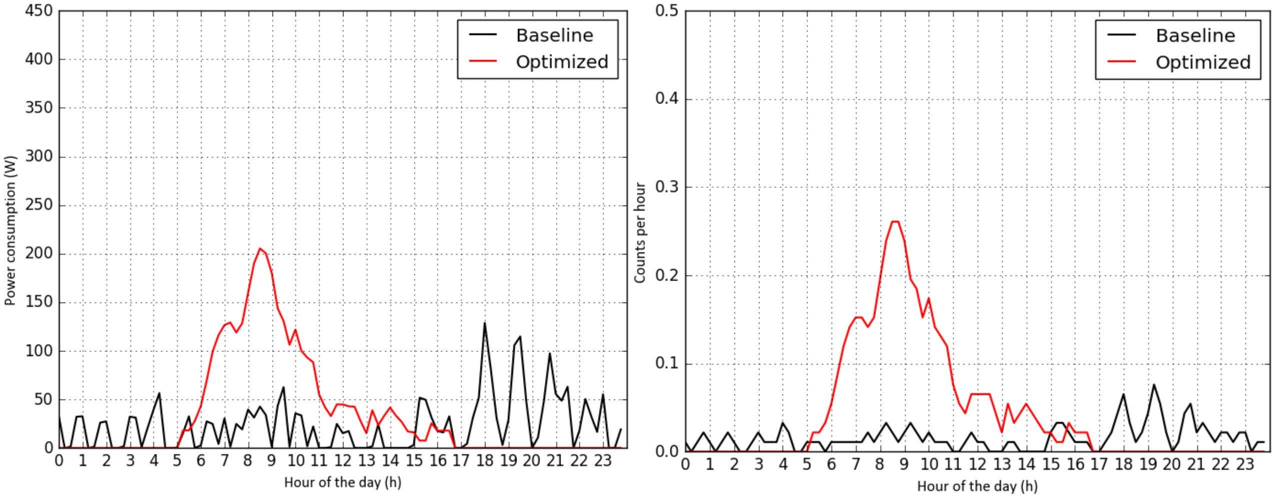


Figure 32. (a). Comparison of energy consumption baseline vs. optimized, house 5. **(b)** Comparison of average daily number of reheat cycles per hour, house 5.

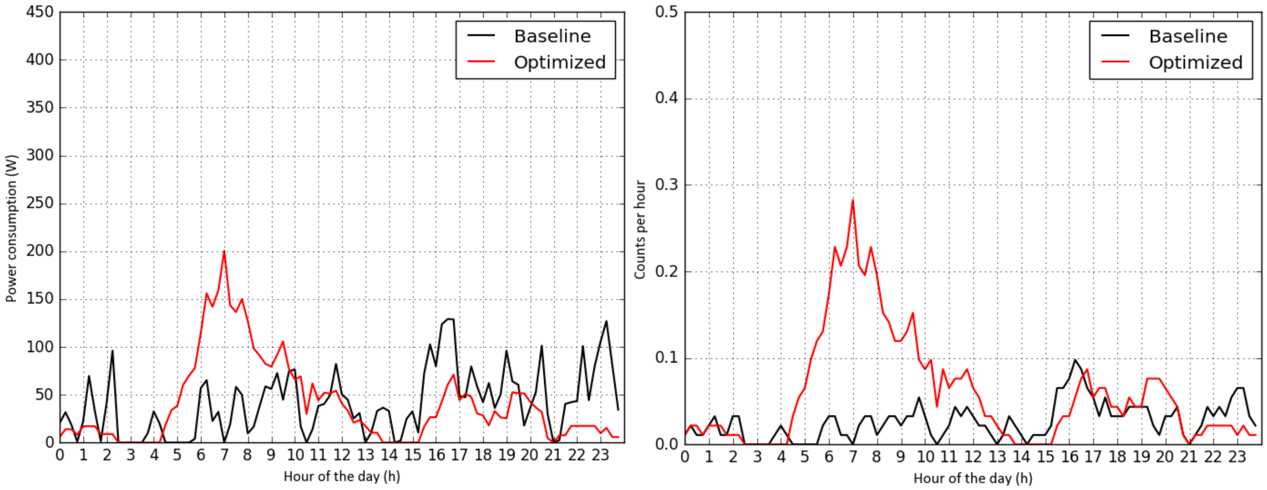


Figure 33. (a). Comparison of energy consumption baseline vs. optimized, house 6. **(b)** Comparison of average daily number of reheat cycles per hour, house 6.

The number of reheat cycles for houses 1 - 4 increased the least compared to the other houses. This could be explained by the concentration of the DHW consumption of these houses around the high energy generation hours and the additional enhancement of the shift of reheat cycles to the high generation hours. Hence, the reheat cycles coincide more often with the events of high DHW consumption. Moreover, the correlation between the timing of the reheat cycles and DHW consumption leaves little room for heat loss to the ambient. Additionally, the concentration of the reheat cycles around noon has the benefit of a higher COP given the ambient temperature profile which correlates with the solar energy profile. Consequently, the energy required to reheat is lower for high energy generation hours. As a result, the average daily required energy is reduced for all houses compared to the baseline control. The numerical representation of the optimization is given in table 3.

House nr.	**Baseline total energy consumption (kWh)**	**Δ Total consumed energy (kWh)**	**Δ Times turned on**	**Δ Total energy withdrawn from grid (kWh)**	**Δ self-consumption (kWh)**	**Δ self-sufficiency (%)**	**Δ Comfort violation (nr of times)**	**Baseline control reheat with no consumption**
1	133	-6 *-4%*	168 *(+37%)*	-20 *(-52%)*	14 *15%*	71% *85%*	-325 *(-79%)*	32/129
2	94	-13 *-13%*	42 *(+12%)*	-30 *(-75%)*	17 *31%*	58% *88%*	-13 *(-15%)*	60/120
3	124	-8 *-7%*	152 *(+34%)*	-34 *(-65%)*	25 *35%*	58% *84%*	-241 *(-71%)*	33/128
4	44	13 *3%*	50 *(26%)*	-9 *(-63%)*	12 *41%*	67% *88%*	-11 *(-61%)*	72/89
5	53	22 *42%*	194 *(+97%)*	-37 *(-97%)*	60 *410%*	28% *98%*	-130 *(-70%)*	40/65
6	91	-1 *-1%*	230 *(+78%)*	-38 *(-69%)*	37 *106%*	39% *81%*	-147 *(-65%)*	28/100

Table 3. Complete numerical results of the optimization control strategy for Jun-Sep

The interaction of the ASHP load with the grid decreases significantly as a result of the optimization control strategy. The total required energy from the grid is reduced by 52% to 97% for house 1 and 5 respectively. The contrast between the consumption profiles of these houses and the resulting optimization, reflects the influence of the individual DHW consumption on the grid interaction reduction potential.

Houses 1 - 4 have high energy consumption concentrated around the high solar energy generation hours. The optimized pattern for these houses changed less than for the other houses; enhancing the energy consumption during high solar energy generation hours and reducing when there is insufficient amount of solar energy. As can be read from the table, house 4 has the lowest energy consumption which is primarily controlled by the comfort threshold temperature of 45 ℃. House 4 has the highest percentage of reheating due to this control mechanism rather than the water consumption as can be read from the last column which indicates the number of times the ASHP turned on, solely for reaching the threshold temperature whilst there was no hot water consumption. The optimized energy profile (fig 31.a.) for this house did not change much during the high-energy generation hours for there is no significant DHW consumption in general. However, it has reduced energy consumption during low energy generation hours due to the reduction in reheating events triggered by reaching the threshold temperature. Hence, the optimization control strategy not only optimizes for maximizing on-site solar energy consumption and self-sufficiency, but also for reduction of reheat cycles during low generation hours and low DHW consumption.

As for houses 5 and 6, these houses have a relatively low energy consumption during the day and a peak in energy consumption in the evening hours. The optimization for these houses shows a reversed pattern; the energy consumption for house 5 reduces to nearly none and significantly lower for house 6 in the evening and early morning. This indicates that the energy consumption for these houses significantly shifted from low generation hours to high generation hours without posing a threat to the comfort standards as can be read from table 3. This is also reflected in the change of the self-sufficiency percentage. The highest rise in the self-sufficiency reflects the highest shift from low generation hours to high generation hours which can be seen for houses 5 and 6.

Overall, it is evident from the graphs and table 3, that the energy withdrawn from the grid is reduced primarily due to the optimized timing of the reheat cycles during high energy generation hours. The total energy injected to the grid by the solar panels (indicated by the self-consumption column) is therefore reduced accordingly. The highest increase in self-consumed energy is 410% larger than for the baseline control strategy for house number 5 and the lowest for 15% for house number 1. This supports the significance of the consumption pattern on the optimization potential furthermore.

The combined total amount of enhanced onsite solar energy consumption for these houses accounts for 165 kWh in the period of June to September. This would have alternatively been injected to the grid causing large peaks in the grid during high generation hours. Moreover, the number of comfort violation (T_m lower than 45℃ and water consumption is greater than 0L) is decreased significantly as a result of the optimization for all houses. This could be attributed to significant rise in the number of reheat cycles in order to maintain vessel midpoint temperatures of higher 49.5 ℃ during high generation hours as set in the reward function. Therefore, the reduction of the comfort violation is considered to account for part of the increase in total average daily required energy.

Perhaps the most notable effect of the optimization is visible in the self-sufficiency parameter. The optimization resulted in a remarkable increase in the self-sufficiency for all houses. House 5 even reached an increase of 70% in self-sufficiency to reach a total self-sufficiency of 99% of the load. The lowest increase is evident for house number 1 with an increase of 14% which resulted in a total self-sufficiency of 85%. The large difference could partly be explained by the high baseline self-sufficiency of house number 1 of 71% against that of only 28% for house 5. The lower the initial self-sufficiency, the larger the potential for optimization. Secondly, house 1 has significantly higher DHW consumption than the other houses which depletes the DHW storage tank much faster forcing it to reheat additionally during the low energy generation hours. Once more, the results emphasize the strong influence of the individual consumption pattern on the optimization potential.

These results show that the algorithm learns the desired behaviour and thereby presents a suitable method for reducing power consumption from the grid and thereby enhancing onsite renewable energy consumption. This covers both objectives of the optimization, for improving the energy efficient building behaviour in order to comply to the nZEB label and reducing the pressure on the national grid. However, there is a large increase in the number of reheat cycles as a consequence of the optimization for solar energy consumption. The increase in number of cycles is required for maintaining a high temperature in the hot water storage vessel during high solar energy generation hours. This way, the storage vessel serves as an energy storage facility for the solar energy. However, this trade-off could be potentially decreased by increasing the comfort margin from a baseline temperature range of 43-50 ℃, to a larger range during high energy production hours. The next section will present the result of an analysis serving to investigate the potential of reducing this trade-off.

6.3. Number of reheat cycles versus grid interaction

For the analysis of the trade-off between the number of reheat cycles and the grid interaction, the upper limit temperature to which the vessel reheats is adjusted to a higher temperature. This is meant to exploit the ability of the storage vessel to buffer energy and thereby reduce the number of heating cycles. Since the heat loss increases with increasing temperature, it is expected that his effect is limited at a certain vessel state in which the effect of the heat loss is balanced by the effect of the reduction of heating cycles. However, this investigation is limited to a certain extent for the DHW storage vessel model is limited for unseen states. In order to identify the limits of the states, the frequency distribution of the temperature data is analysed in figure 34.

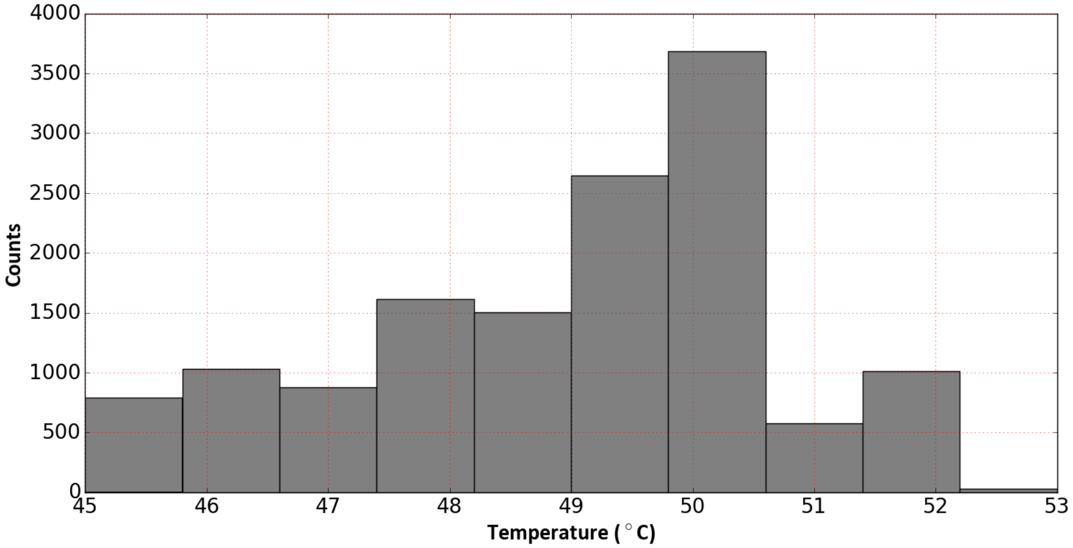


Figure 34. Frequency distribution of T_m for house 1

The frequency distribution shows that temperatures above 52 ℃ are highly underrepresented. The uncertainty for these states is inevitably high and the performance of the vessel model around these states is expected therefore to be underperforming. Therefore, a maximum of 52 ℃ is taken as a limit for this analysis.

The effect of the maximum temperature set point on both the number of heating cycles and the corresponding average daily energy consumption is shown in figures 35 a and b. It is evident that both

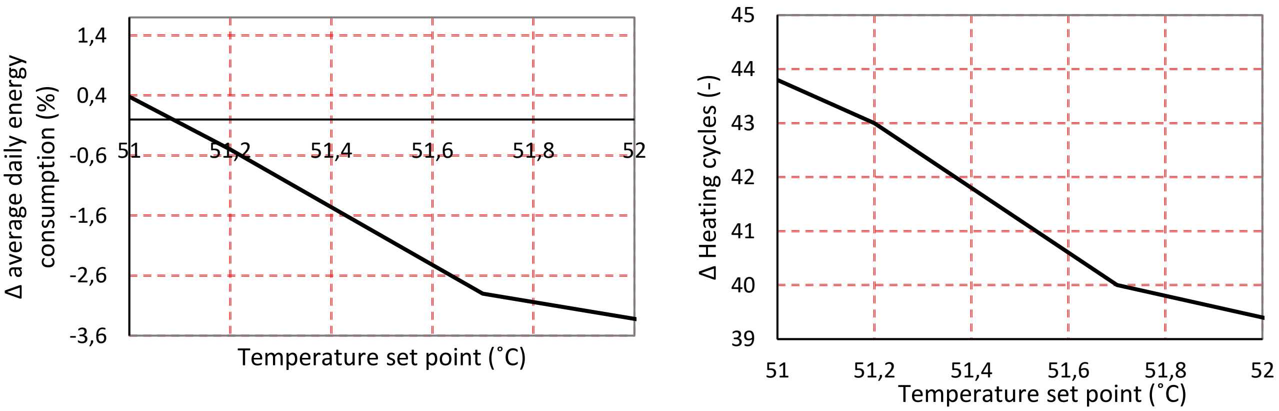


Figure 35 (a) Relationship between Tset and average daily consumed energy (%). **(b)** Relationship between Tset and number of heating cycles (%).

parameters decrease with increasing temperatures of this range. Therefore, to optimize for these parameters, a temperature of 52 ℃ is set as an upper limit for the heating cycles for this analysis. The resulting energy profiles are shown in figures A.1 to A.6 in appendix A. The main visible difference between the plots in the appendix and the plots of figures 28 – 33, is the width of the reheating events which are wider for an upper limit for T_m of 52 ℃. This results in smoother graphs and lower peaks during low solar energy events. Hence, more energy is stored during the high solar energy generation hours leading to lower energy consumption in the evenings and mornings. This is numerically represented in table 4.

2 House nr.	**Baseline total energy consumption (kWh)**	**Δ Total consumed energy (kWh)**	**Δ Times turned on**	**Δ Total energy withdrawn from grid (kWh)**	**Δ self-consumption (kWh)**	**Δ self-sufficiency (%)**	**Δ Comfort violation (nr of times)**	**Baseline control reheat with no consumption**
1	133	-8 *(-6%)*	162 *(+36%)*	-24 *(-63%)*	16 *17%*	71% *89%*	-363 *(-88%)*	32/129
2	94	-16 *(-17 %)*	19 *(+6%)*	-33 *(-82%)*	17 *31%*	58% *91%*	-45 *(-53%)*	60/120
3	124	-16 *(-13%)*	123 *(+28%)*	-37 *(-70%)*	20 *28%*	58% *85%*	-287 *(-84%)*	33/128
4	44	-0.2 *(-0.5%)*	44 *(23%)*	-11 *(-73%)*	10 *35%*	67% *91%*	-13 *(-72%)*	72/89
5	53	13 *(24%)*	135 *(+67%)*	-38 *(-99%)*	51 *348%*	28% *100%*	-132 *(-70%)*	40/65
6	91	-6 *(-6%)*	217 *(+74%)*	-37 *(-67%)*	32 *89%*	39% *79%*	-164 *(-73%)*	28/100

Table 4. Complete numerical results of the improved optimization control strategy for June-Sep.

With this control strategy, the energy required from the grid reduced substantially for most of the houses. The reduction for house 1, 3 and 6 are amongst the lowest compared to the other houses similar to the previous control strategy. This is to be expected since the energy consumption profile is primarily controlled by the water consumption and less by the threshold temperature of the baseline control as presented in the last column. Therefore, reheat cycles at low energy generation hours are inevitable given the relatively higher DHW consumption.

Furthermore, the number of total heating cycles reduced for all houses compared to the previous optimization algorithm. This was expected since the maximum temperature is set higher which increases the comfort level range by 2 ℃ allowing the DHW tank to store more energy. As a result, the consumed energy is reduced for all houses compared to the previous optimization control. This reduction results in lower solar energy consumption compared to the previous optimized control which is reflected in the reduced self-consumption. Nevertheless, the reduction of the average daily required energy without violating the comfort standards is an improvement to the energy efficiency. The increased self-sufficiency shows that the improved algorithm performs better in terms of shifting the energy consumption from low generation hours to high generation hours complying thereby to the objectives of the optimization. This optimized control strategy could, therefore, find the balance between the energy efficiency and the onsite solar energy consumption. Most importantly, the number of reheat cycles reduced significantly compared to the previous optimized control. Too many reheat cycles could increase the wear and tear in the ASHP reducing its lifetime and increase unnecessary heat loss. It is therefore important to achieve an optimization with as less number of reheat cycles as possible. Nevertheless, optimizing on-site solar energy consumption inevitability requires an increase in the reheat cycles around the high energy generation hours. A balance between these two objectives is therefore crucial as done in the second case by increasing the temperature upper limit of the reheating cycles. The second control strategy will therefore be applied rather than the first for further analysis.

The presented results show that the RL algorithm improves the self-consumption and self-sufficiency significantly resulting in simulated self-sufficiencies between 81% and 100%. As a result, the energy withdrawn from the grid is reduced significantly. The energy injected to the grid is similarly reduced between 10 and 51 kWh (depending on the initial electricity use) for the months of June to September. These improvements are compared to the default control strategy which does not take the solar energy profile or the occupant behaviour into account. However, as stated previously in the introduction and literature review, studies show a significant performance gap between measured and modelled nZEB performance as result of inaccurate occupant behaviour modelling or incorporation. Due to the increasing electrification of the building energy services, the interaction of these buildings with the grid will therefore be larger than estimated by these models. This challenges the role of grid operators in the current top-down energy infrastructure as there is no control of the building services loads. RL was therefore proposed as a control strategy to reduce the grid interaction. The main advantage of using RL for domestic load control is the introduction of occupant behaviour in the assessment. It is therefore important to assess the influence of this factor on the optimization. The following analysis will therefore focus on the role of the occupant behaviour on the performance of the algorithm.

6.4. Occupant behaviour profile influence

An important challenge for applying a fixed control strategy on a large scale, is the difference in the individual DHW consumption per household. The previous results have shown more than once that the consumption pattern plays an important role for the optimization potential. The Sarsa(λ) applies the rewards based on the states that are comprised of a combination of the midpoint temperature, water consumption and solar energy production. The linear model applied to reproduce the Q-values adapts its coefficients according to the algorithm's output. Table 5 presents the coefficients reproduced by the linear model of the Q-values for the six houses individually. The values show clear differences from one house to the other.

House	Intercept	Temperature coefficient	Water coefficient	Solar coefficient	Percentage T_m> 45℃
1.00	-42.02	1.08	-0.07	-0.0128	81%
2.00	-59.75	1.42	-0.03	-0.0118	97%
3.00	-49.27	1.21	-0.04	-0.0131	82%
4.00	-83.47	1.93	-0.09	-0.0123	99%
5.00	-45.49	1.12	-0.02	-0.0117	69%
6.00	-47.65	1.19	-0.05	-0.0126	78%

Table 5. Linear model coefficients and intercepts of each individual house

The solar energy production varies the least since the solar energy generation for the houses is nearly the same. The largest variation is evident in the value of the intercept and the temperature coefficient. Houses 1 and 4 show the greatest contrast for these two values, correlating with the greatest contrast in the DHW consumption of the houses. The significantly low water consumption of house 4 results in a temperature drop mainly due to heat loss to the environment rather than water consumption. Therefore, the storage vessel is at high temperatures for the majority of the states (as indicated in the last column of table 5), allowing for the lowest intercept and highest weight for T_m. This emphasizes the influence of the consumption pattern on the resulting control strategy. Discarding this effect will therefore lead to limited optimization.

This effect will be examined by applying average solar and DHW consumption profiles to the algorithm in order to compare the results with the occupant behaviour customized results. Conducting this analysis allows for investigating the influence of the occupant behaviour on the learning mechanism of the algorithm and its importance in control strategy optimization.

6.5. Average profile states simulation

An average monthly solar energy generation and water consumption profile are created from the available data-set in order to simulate the resulting states by using the vessel state model [66]. These states are then applied to the algorithm to produce the Q-values and compose the linear model. The linear model is then to be used to assess the potential reduction of required power from the grid and the injected energy on monthly basis. The optimization potential will be assessed on a monthly time-scale allowing for the comparison between the high and low solar energy generation months. The solar energy profiles show large differences between the warm and cold months for the latitude of the project location of Soesterberg as illustrated in figure 36.

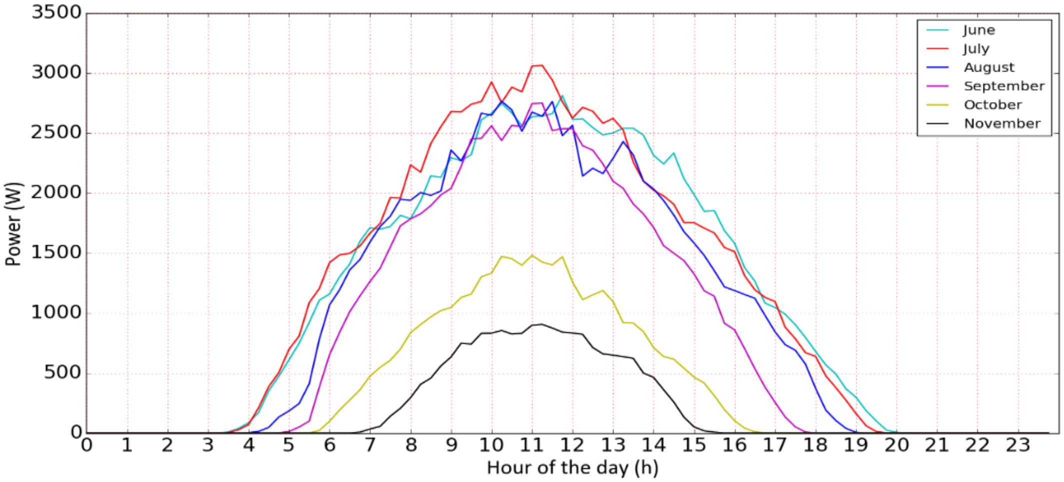


Figure 36. Average monthly solar energy generation

It is therefore natural that the optimization potential would vary significantly from summer to winter. A monthly averaged profile for the solar energy is therefore justified. Averaging the six different water profiles of the houses result in the following average daily profile for the 6 months:

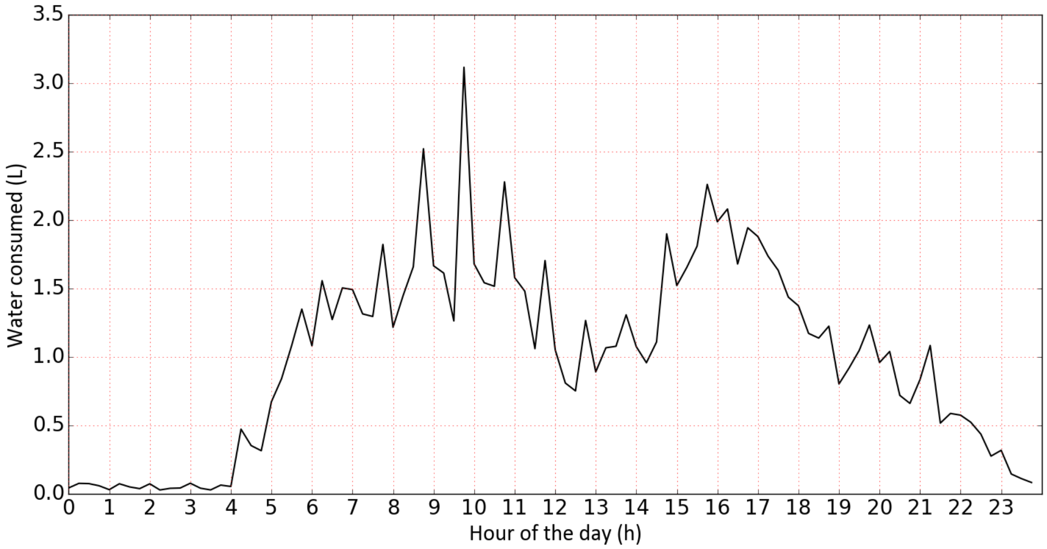


Figure 37. Average daily solar energy generation of all six houses for 6 months

This profile and the average monthly solar energy profile are applied to the vessel state model [66] to generate the resulting midpoint temperatures by applying the optimization control strategy. The resulting state-action-pairs were applied to the Sarsa(λ) algorithm in order to generate an average linear regression model. This serves the purpose of providing a generalised control strategy that could be applied to all houses for different months. The generalisation potential of this approach is examined by applying the resulting linear regression model to the individual houses on monthly basis.

Comparing the individual with the averaged profiles results of the summer months, results in the following for the total energy withdrawn and injected from/to the grid:

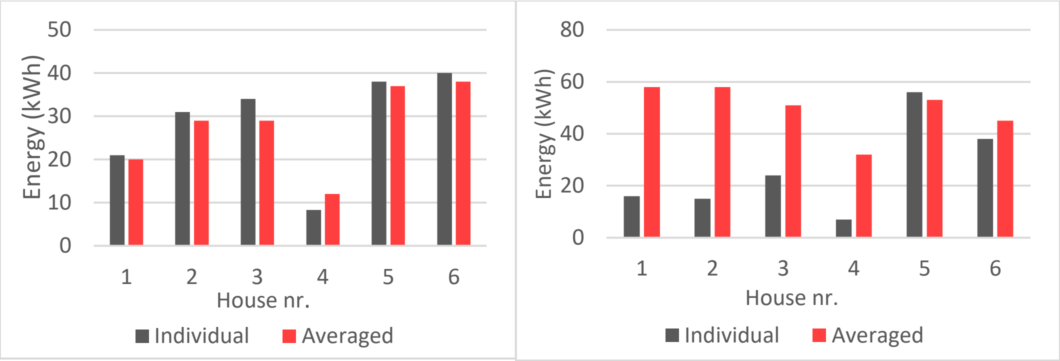


Figure 38. (a) Total reduction of energy uptake from the grid for the summer months' data, individual vs. averaged optimization. (b) Total change in self-consumption for the summer months' data, individual vs. averaged optimization.

The results show a consistent increase in the reduction of the grid interaction dominated by a large increase in self-consumption. The averaged profiles optimization policy thus keeps improving the interaction with the grid by reducing the total energy required from the grid and solar energy injected to the grid. However, the large increase in self-consumption is partly the result of the increase in the number of times the ASHP is turned on. Therefore, it could be said that the average profiles policy is effective in decreasing the grid interaction as aimed for, however increases the number of reheat cycles for most houses. This has the largest impact on houses with a relatively low energy consumption profile compared to the average consumption profile such as for house numbers 2 - 4 as can be seen in figure 39.

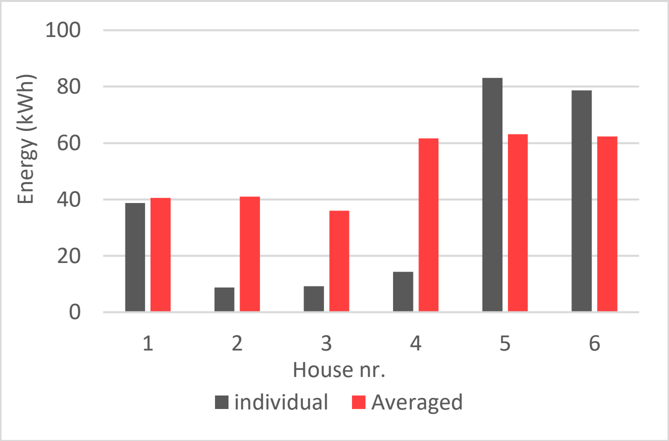


Figure 39. Percentage of change of reheat cycles compared to the baseline for the individual and averaged optimization strategies for June-Sep

As stated before, a large increase in the number of heating cycles can have negative effects on the lifetime of the device and also on the energy efficiency. Houses with a consumption profile below the average will be affected the most, leading to overconsumption of solar energy. Therefore, including the occupant behaviour aids to achieve a more effective and customized control strategy performance. This proofs the effectivity of the learning mechanism of RL algorithms and illustrates its role in tackling the challenge of incorporating the stochastic nature of occupant behaviour.

6.6. Practical implementation

6.6.1. Implementation setup

In the system setup of the case-study, the data of all different houses is sent to one central repository in which the control statements are integrated. The control action is then determined based on the embedded control strategy and the individually received data and sent as a control signal to the concerned heat pump. This setup is alternatively shown in figure 40.

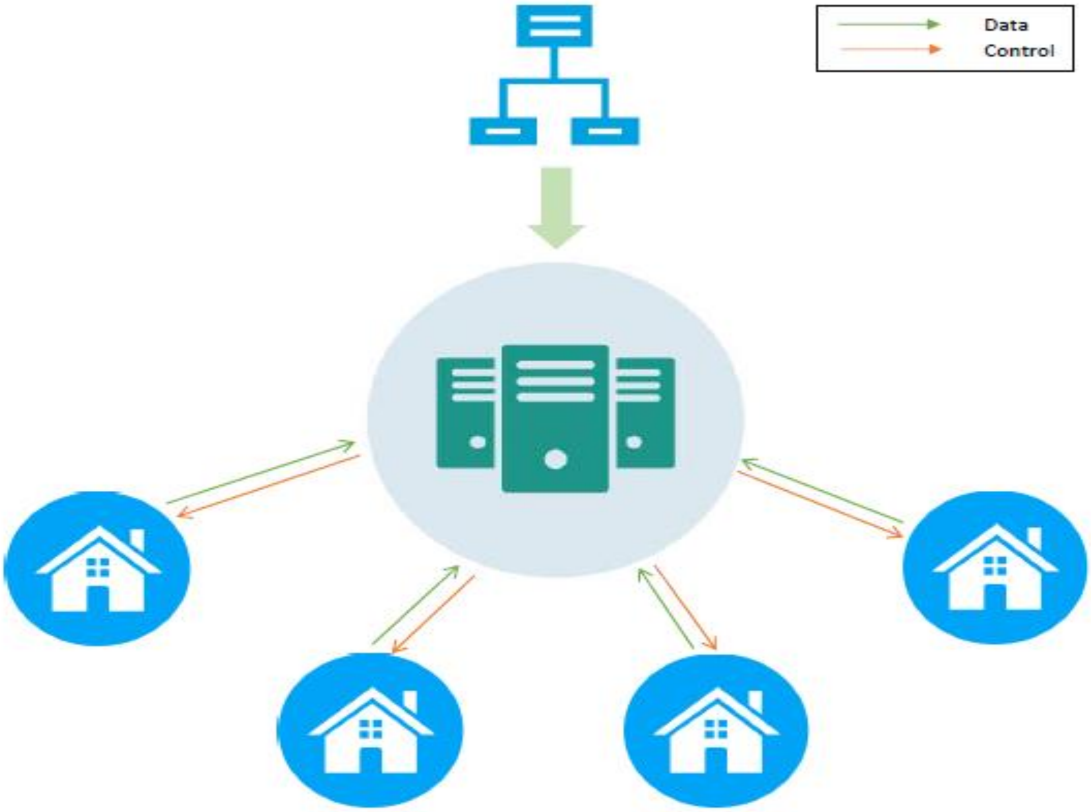


Fig. 40. Current centralized control and computing architecture

The central gathering of the data is a useful tool for applying this method on cluster level. The derived control algorithm could then be integrated centrally to determine the control action for each individual heat pump. This approach makes it possible to learn the individually customized optimal control for each house. As a result, overconsumption or comfort violation of the individual households is avoided. The time-scale for the algorithm update could be performed on 24-hour basis to update for the gradual change in solar energy production and to include the changes in the DHW consumption profile. In order to maintain timely adaptation to the weather, it is most ideal if the update time-scale does not exceed a month. This is also important to perform to deal with large changes in the occupant consumption in cases like changing the occupants and/ or periods of absence.

6.6.2. Generalisability

The cluster of houses of this case-study comprise identical refurbished social housing dwellings. Therefore, the system size and parameters are all identical which will naturally result in control coefficients of which values lay relatively not far from each other. The only highly divergent variable on daily basis is the occupant behaviour. Nevertheless, like presented in §6.5, the algorithm learns the corresponding control strategy based on the individual data input. Therefore, this control strategy could also be applied to clusters with different and also divergent system parameters provided that the three parameters T_m, $\sum Water$, and the solar energy generation are available. In other words, the algorithm is generalizable and could be applied to any system that could provide these three parameters. The optimization potential will vary from one system to another based on the system parameters such as the insulation and the size of the water vessel, solar energy generation, type of dwelling, and the number of occupants of the dwelling. The aggregated results for the houses used in the case-study throughout the months of June to December will be illustrated in the next section. This is to illustrate the optimization potential for the specific net zero energy dwellings used in the case-study for the climate of the Netherlands and how this information could benefit grid operators and the occupants.

6.7. Monthly aggregated optimization potential

When applied to the study case, the acquired monthly grid interaction reduction potential differs from one month to the other following the changing consumption and generation patterns shown in figure 41. This figure shows the baseline spread of all six houses of energy injected to the grid (a) and withdrawn from it (b).

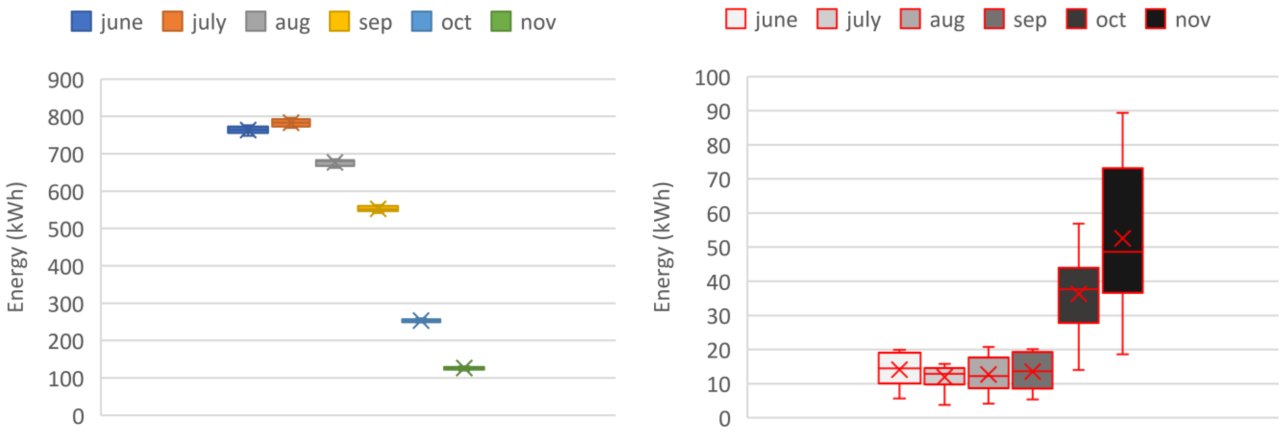


Figure 41. (a) Monthly spread of baseline total absolute energy injected to the grid of all six houses. (b) Monthly spread of baseline total absolute energy withdrawn from the grid of all six houses.

The change of the energy consumption shown in fig. 41.b. is a result of the decreasing ambient temperatures throughout the months, leading to a lower COP of the ASHP and therefore a higher energy consumption. Since the ASHP load increases significantly in the colder months, the amount of required energy from the grid increases accordingly. Similarly, the self-consumption potential imposed by the algorithm decreases since the produced solar energy is decreased (fig. 41.a.). This could be seen in figure 41 which shows the spread of the absolute amount of reduced energy injected to the grid (41.a.) and withdrawn from it after applying the algorithm on monthly basis.

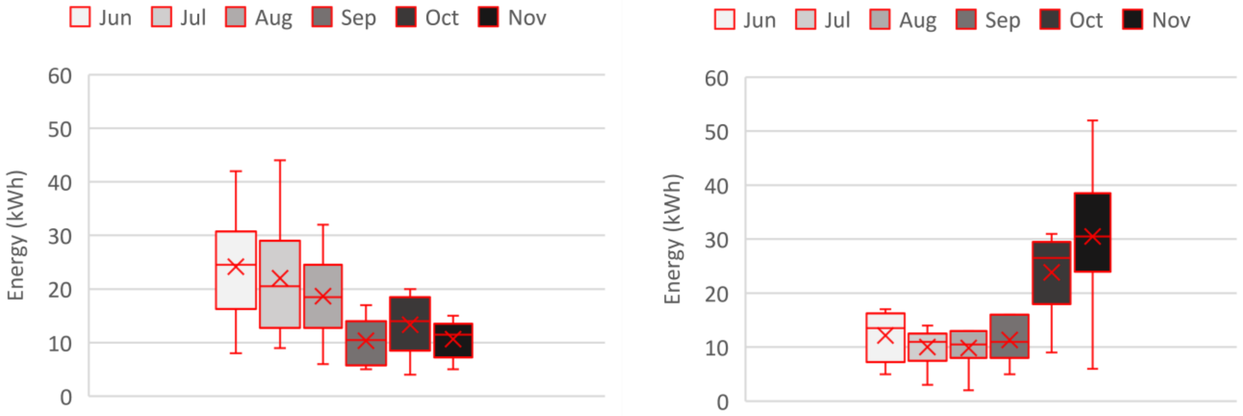


Figure 42. (a) Monthly spread of optimized total mitigated energy injection to the grid of all six houses. (b) Monthly spread of optimized total mitigated energy uptake from the grid.

These results show that the grid interaction reduction potential created by the algorithm persists throughout the different months. For this case study, the optimization implies a minimum energy uptake reduction of around 10 kWh at the peak of summer which increases gradually following the energy consumption profile. The change in energy injected to the grid exhibits similar results for which the energy injection reduces from summer to winter with maximum average of around 25 kWh. When expressed in terms of self-consumption and self-sufficiency, the results show that the potential for both decreases towards winter (fig. 43(a) and (b)).

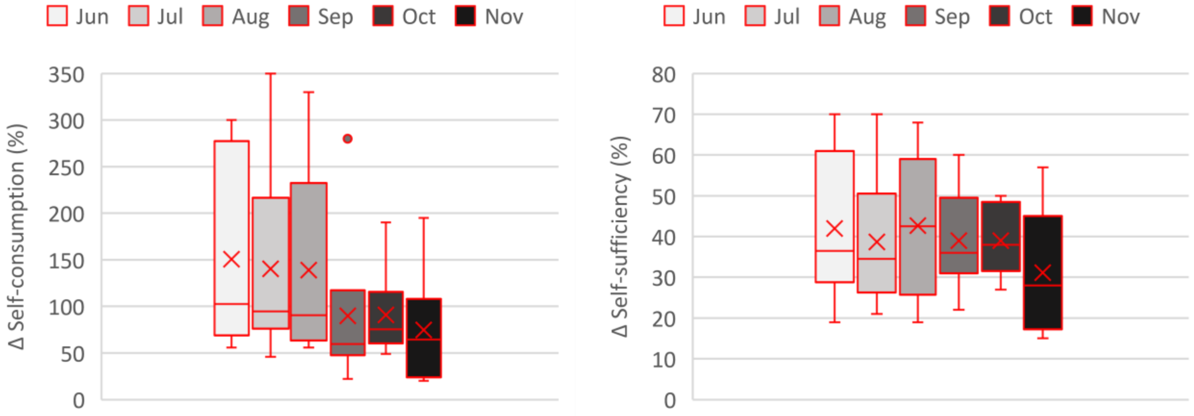


Figure 43. (a) Monthly spread of optimized self-consumption of all six houses. (b) Monthly spread of self-sufficiency of all six houses.

The reduction in the solar energy generation during the colder months reduces the potential for peak shaving from low generation hours to high generation hours. Hence, the optimization potential (%) for the energy uptake from the grid reduces since there is less solar energy available to supplement the consumption. Likewise, since the solar energy reduces, self-consumption becomes harder to achieve as can be seen in 43.a. Nevertheless, the optimization potential of the algorithm persists to be effective in reaching the goal of reducing grid interaction by shifting reheat cycle timings.

6.8. Implications for the occupants

Recalling the current Dutch net metering compensation scheme described in 1.2.2, selling excess solar energy is beneficial in all cases before 2020. Power sold to the grid is compensated against the total consumed power on annual basis. The excess power sold to the grid beyond the range of the total consumed power, is also compensated for a fair price determined by the DSO. This scheme is therefore stimulating consumers to invest in solar energy. As result, the consumers do not have to pay tax over their energy bill. This tax is estimated to account for 40 million Euros annually [29] which is funded by the Dutch tax system. It is therefore argued that this tax is paid for by tax-payers without solar panels [29]. Moreover, the scheme is not beneficial for grid operators on a long-term scale as a result of the need for grid reinforcement. It is therefore planned to diminish this scheme by 2020. Whilst currently reducing the power injected to the grid is not beneficial, it will become so according to the plans for the transition plan in 2020 [86]. Two main transition proposals are stated in the report state a reduced value for the excess sold power to the grid. The first proposal states a buying value of €0.22/kWh against a selling value of €0.06/kWh. The second proposal includes an additional feed-in tariff of €0.075 to the selling price. In both cases, on-site solar energy consumption will become more beneficial than selling to the grid. It is estimated by [86] that the monthly cost for nZEBs will increase, as a result of diminishing the net-metering scheme, by €40 given the current prices of solar technology and heat pumps and €30 considering the expected price reduction of these installations. Load matching to increase on-site solar energy consumption increases therefore the feasibility of these building concepts for the occupants. An increase of 150% in summer and 75% in the colder months on average as a result of implementing the algorithm, therefore, is a cost-efficient way to promote the concept of nZEBs even without the net-metering scheme. Table 6 presents the total reduction of the summed energy injection and uptake for all six houses in the period of June to December, against the monetary value in both scenario 1 and 2 of the proposal of [86]. Deducting the money missed due to the reduction of solar energy selling from the gained money due to the reduction of electricity uptake, results in a total benefit between € 4 - 24 in the first scenario and €1 - 9 in the second for the six houses for the period of June to December.

House nr	Total reduced injection (kWh)	Total reduced uptake (kWh)	Δ Monetary benefit scenario 1 (€)	Δ Monetary benefit scenario 2 (€)
1	79	89	14.84	8.915
2	98	118	20.08	12.73
3	102	137	24.02	16.37
4	37	30	4.38	1.605
5	164	104	13.04	0.74
6	115	108	16.86	8.235

Table 6. Overview of reduction of energy uptake and injection and the corresponding monetary benefit for the two scenarios proposed in [86] to replace the net-metering policy in the Netherlands in 2020.

6.9. Implications for grid operators

The optimization results discussed in section 6.7 show a decrease of energy injection to the grid of a maximum of 25 kWh on average for the houses in summer and a minimum of 10 kWh for the period of September to December. The reduction of injected power to the grid translates directly to a reduction in the necessary cost for grid reinforcement. The required grid reinforcement cost comes from the size of the transformers and cables required to accommodate a large increase solar energy generation due to an increase in nZEB concepts. In addition, the installed heat pumps in these buildings add to the electrical load required from the grid, thereby increasing the pressure on the grid components. A decrease in this effect will therefore reduce the necessary grid reinforcement cost which depends on the size of cluster of houses, the grid components size already in place, and the location. The study of ECN [69] discussed previously in the literature review, presents the grid reinforcement cost mitigation as a result of a peak shaving technique of around € 250 thousand for a cluster of 1400 houses. This illustrates the large cost associated with a large increase in the grid interaction as a result of the nZEB concepts.

Additional benefits for the grid operators relate to the real-time grid monitoring and steering role of the grid operators. Acquiring the data for the centralised control gives grid operators more information about the distributed generation and the consumption fluctuations of nZEBs. This provides the grid operators with the monitoring tools necessary to maintain grid resilience and avoid black-outs given the necessary changes in the top-down energy infrastructure.

7. Discussion

7.1. Relevance of study

Enhancing onsite RES energy consumption and self-sufficiency is an important optimization objective for the built environment for it allows a larger penetration of RES energy in the grid. In addition, the building sector is increasingly more being transformed into nearly zero energy concepts as described in the literature review. This transformation is aided by RES to complement the energy deficit for which the conservative architectural design cannot cover. The interaction between the built environment and the grid is therefore likely to increase posing grid reliability and performance challenges.

The built environment is likely to host PV energy technologies compared to other RES for its relatively low cost and easy integration in the architecture. Optimizing for maximum onsite solar energy consumption is therefore significant for reducing the pressure of the intermittent energy penetration in the grid from this sector. An additional benefit is relevant for the concept of nZEBs as the energy performance is improved even more to comply to the energy label standards. The analysis described in this study, examines the potential of reinforcement learning as an optimization tool for the control strategy for peak shaving by matching the load to the energy generation to reach these objectives. Reinforcement learning omits the need for a physical model of the system, saving thereby time and effort. Another benefit is that it provides generalisability whilst physical models are built based on specific cases. This entails the need to adapt the model from one case to another, reducing thereby its generalisability. In contrast, reinforcement learning algorithms are data-driven which makes them generalizable and not domain-specific.

7.2. Technical assessment

The studied indicators include the onsite self-consumption, self-sufficiency, and energy uptake from the grid. Two important limitations for the study comprise the number of reheat cycles and the comfort standards for the occupants. The number of reheat cycles has a direct influence on the total amount of energy required for the ASHP operation and the resulting efficiency. It is therefore important to maintain a balance between the on-site self-consumption and the energy consumption of the ASHP. As for the comfort standards, the optimization should not compromise the occupant comfort. The comfort standards as defined by BAM, entail a DHW temperature of at least 45 °C. The temperature to which the ASHP reheats each cycle is set at 51.5 °C.

The first optimization strategy shows significant optimization potential regarding the timing of the heating cycles. Figures 28 to 33 and A.1 – A.6. show that the algorithm reduces energy consumption during low energy generation hours whilst increasing the energy consumption during the afternoon. As a result, the self-consumption and the self-sufficiency improved substantially for the period of June to September. The self-consumption increased by 12% for house 1 compared to the initial value and 410% for house 5 (table 3). The self-sufficiency shows similar optimization as it increases to between 84% and 98% for houses 1 and 5 respectively. This is achieved whilst the comfort standards were not violated additionally. In fact, the comfort violations were reduced by 15% to 79%. On the other hand, the results show an increase in the number of reheat cycles between 12% and 97% compared to the baseline results. In general, it holds that the higher the increase in the number of reheat cycles and the bigger the time gap between reheating and DHW consumption, the higher the energy loss to the environment. This effect is pronounced strongest in the results of house 5 in which an increase of 97% in the number of reheat cycles resulted in an increase of 42% of the energy consumption.

Moreover, the discrepancy between the rise in self-consumption and the reduction of the energy uptake from the grid indicates that the rise in self-consumption is not solely the result of shifting the reheat cycles from low to high generation hours. It becomes therefore evident that the optimization algorithm results in overconsumption for houses 4 and 5 due to enhanced heat loss as a result of the increased number of reheat cycles. An improved optimization control strategy has therefore been deployed to reduce the number of reheat cycles whilst maintaining an optimal rise in self-consumption. This second optimization differs from the first in the temperature set point of 52 ℃ (as opposed to baseline 50 ℃) to which the vessel is reheated in each cycle.

The results given in table 3 show an improved behaviour. The number of reheat cycles and the total energy consumption of houses 4 and 5 reduced significantly compared to the first optimization strategy. This does not only provide an advantage for the life-time of the heat pump but improves the energy efficiency of the building to meet the nZEB labelling objectives and beyond. The self-consumption, however, reduced to between 17% and 348% due to the reduction in total consumed energy. However, the further improvement of the self-sufficiency indicates that the algorithm improved in allocating the heating cycles from low to high energy hours which resulted in self-sufficiencies of between 85% to 100% compared to baseline 28% to 71%. Additionally, the comfort violations reduced furthermore. As a result of this optimization, the grid interaction was reduced substantially. As the DHW load is only around a tenth of the size of the solar energy generation in summer, the reduction in the energy injected to the grid is rather sparse. Nevertheless, the energy uptake from the grid reduced by a minimum of 54% to a maximum of 99%, complying thereby to the objectives to reduce grid interaction.

All these results combined show that the second optimization strategy could find the balance between the energy efficiency and the objective of peak shaving without compromising the occupant comfort. By increasing the temperature set point of the DHW storage vessel, the vessel utilises more of its ability to act as buffer for the solar energy, reducing thereby the need for additional heating cycles. This is believed to be the reason for the improvement of the second optimization strategy over the first. The reduction of the average daily required energy is an indication for a significant potential to meet both objectives of maintaining and improving energy efficiency and enhancing on-site energy consumption. The required energy was reduced by 11% on average. Kazmi and D'Oca 2016 [36], conducted a study with model-based reinforcement learning to examine the potential for reducing the energy consumption of the same system setup used in this paper. The research finds a total simulated reduction of around 15% and real-world reduction of 27% compared to the baseline results. This research was solely focussed on the reduction of energy consumption of the ASHP with no regards to maximizing on-site energy consumption. The similarity between the results shows that both objectives could go hand in hand without having to compromise one of the two goals.

7.3. Results compared to outcomes found in literature

Analysing the effectivity of the method in reducing grid interaction, the results are compared to the outcomes found in the literature discussed in the literature review. The different outcomes are provided in table 7. It is good practice to note that the outcomes depend not only on the methods used but also on the system size and components. The studies presented here have system components specifications of around the same size. The only exception is the study of [69] using the PowerMatcher. This study applies scheduling a methodology to control a cluster of 1400 houses to spread the loads over the whole day and avoid peaks.

Evident from the rule-based results found in both [67] and [72], that the resulting optimization potential is evident but less effective than MPC and RL methodologies. This emphasizes the drawback of fixed control strategies as they lack the ability to adapt the control to changing patterns. MPCs on the other hand, show significant optimization potential. As discussed previously, MPCs utilise physical models of the system. This makes them more accurate in predicting the loads. Their optimization potential is therefore promising. One important drawback to MPCs concerns the need to apply these objectives to large scales. Since they require models of the system, they do not provide generalisability so they could be applied to systems with different system specifications. In comparison, the RL methodologies perform closely to the outcomes of the MPC methods. The study of [77] compares the results of both an MPC and RL algorithms and concludes that RL algorithms perform closely to the performance of MPCs. Their study results in RL performance of 2% lower than that of the MPC outcomes. This advantage of RL is also showed in the results of this thesis also indicate a significant optimization potential for the months of June to December on average (fig. 43). This potential is expected to decrease to a certain extent if the winter months will be included in the assessment. Nevertheless, this effect is expected to be limited since the results show a persistent grid interaction reduction throughout the analysed months. The optimization potential falls in the range of the results of [73], [82], and [69]. This shows significant potential of RL algorithms to be applied to case-studies with load matching purposes.

Study	Method	Δ Self-consumption (%)	Δ Self-sufficiency (%)
Ijaz Dar et al, 2014 [67]	Rule-based	6	11.5
De Coninck et al 2010 [72]	Rule-based	-	3.4
Sossan et al, 2013 [73]	MPC	293	-
Van Houdt et al, 2014 [82]	MPC	8 – 29	5 – 25
Peng and Morrison [77]	RL	-	12
Kazmi and D'Oca, 2016 [36]	RL	-	15
Pruisen and Kamphuis [69]	PowerMatcher	50	-
Thesis results	RL	17 – 348	18 – 72

Table 7. Overview of optimization outcomes found in literature [36],[67],[69],[72],[73],[77],[82], compared to the thesis results

7.4. Learning mechanism of RL

The main argument to use RL for this optimization is the learning aspect and generalisability of the method. The constantly changing solar and occupant behaviour patterns create a disadvantage for fixed optimization approaches such as MPCs and rule-based control methodologies. To test and proof the learning mechanism of the RL algorithm used, average occupant behaviour and monthly solar energy profiles were created from the individual profiles of the six houses. These profiles were then applied to the algorithm to simulate the results for each of the houses. The number of reheat cycles as a result of the averaged profiles optimization show remarkable results for being substantially amplified for houses with a low energy consumption profile. The large difference of the number of heating cycles between the individual and averaged profiles indicates the strong influence of the individual occupant behaviour on the performance of the algorithm. This emphasizes the significance of the learning behaviour of the algorithm as opposed to rigid approach of rule-based control optimization algorithms. The occupant behaviour determines the transition between the states which influences the occurrence of the states. As seen from the tables 3 and 4, higher states (higher T_m) dominate the state-space of the data set of the houses that have a low energy consumption profile and vice versa. This contrast is best evident from the results for house 1 and 4. This mechanism determines the relative weight of the three different parameters, T_m, $\sum water$, and the available solar energy in the resulting linear model. When these three parameters are incorporated in a rule-based control algorithm, the relative weight of each of them would be fixed at a certain value. The efficiency of the heat pump could therefore be compromised when aiming at load shifting. Nevertheless, rule-based algorithms provide a solution when historic sensor data are not available to build a learning algorithm.

7.5. Implementation analysis

The current control strategy is organised through a centralized data gathering and computing architecture. This allows for applying the algorithm for each house individually to reduce the grid interaction for a cluster of houses. This way not only could the algorithm learn the individual consumption pattern of each house but also continue learning and adjusting the control strategy accordingly. Additionally, the centralised data gathering assists the grid operators in their grid monitoring role to ensure grid reliability by providing information about the consumption and generation patterns.

The optimization objectives of reducing the grid interaction are most effective when applied on a large-scale. Reducing the grid interaction on a large scale provides the tools to reduce the effect of the increasing electrification of the residential built environment as a result of implementing nZEB concepts. Henceforward, more nZEBs could be implemented to assist to improve the energy efficiency of the built environment and enhance the penetration of renewable energy without having to invest significant amount of money to reinforce the grid components.

Concerning the occupants, the Dutch government will abolish the current net-metering scheme for solar energy generation in 2020. Two transition plans are proposed instead to deal with injected solar energy to the grid. Both scenarios describe a lower selling price for generated electricity from solar energy compared to the buying price electricity per unit of energy (kWh). As shown in the results, the reduction of the grid interaction of the buildings, therefore, reduces the annual electricity bill in both scenarios. This adds to the incentive to reduce the grid interaction.

Overall, it could be concluded that the method proofed to be effective in complying to the objectives of reducing the grid interaction of the buildings, improved solar energy self-consumption and the efficiency of the ASHP with no negative consequences for the occupant behaviour.

8. Conclusions and recommendations

Deriving from the results and discussion, using RL could provide means for improving the performance of nZEB to comply to the label standards and in the same time reduce the pressure on the grid as a result of the increasing electrification of the built environment without posing a threat to the occupant comfort. Providing these means for the utility grid not only reduces the need and cost for grid reinforcement, but allows for more renewable energy technologies to be deployed in the residential built environment.
The results show an increase of individual self-consumption between 17% and 348% and self-sufficiency between 18% and 72% by applying DHW load shifting. This indicates a significant potential to reduce grid interaction to boost the implementation of RES in the residential built environment for the sake of increasing the share of renewables and also to improve the efficiency of the built environment by implementing more nZEBs. Additionally, considering the planned abolishment of the net-metering policy for solar power in 2020 in the Netherlands, reducing the grid interaction increases the monetary benefit for the residents. Moreover, this approach improves the monitoring potential of grid operators to avoid grid curtailment as a result of the increased electrification of the building loads and locally generated renewable energy.

In this study the algorithm was trained on a data set with a length of 6 months. This provided promising results, however, the working principle of RL is that the algorithms improve behaviour when they acquire more experience. With changing occupant behaviour and weather patterns, the algorithm will fully exploit it potential and comply to its goal when provided with more experience on continuous or regular basis. Although in this current study only the DHW is considered, the other possible states of the ASHP could be incorporated in the algorithm to include the HVAC in the optimization. Additionally, the adaptable reward function provides an option to increase the weight of the available solar energy to provide additional flexibility when needed. In this study the total energy consumption of the ASHP is taken as a limitation for the optimization to maintain a high efficiency. However, when the goal is to provide positive flexibility to avoid grid curtailment, the algorithm could easily be modified to increase energy consumption. Negative flexibility, on the other hand, is represented by the acquired significant reduction in energy uptake from the grid as a result of the optimization. Hence, when applied to the complete electric load of the building, optimizing for on-site renewable energy consumption will inherently reduce the required negative flexibility.

9. References

[1] McInerney, C., Johannsdottir, L., 2016. Lima Paris Action Agenda: Focus on Private Finance-note from COP21. Journal of Cleaner Production, Elsevier, 126: pp. 707-710.

[2] European Commission, 2017. Report from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Renewable Energy Progress Report, Brussels. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017DC0057&qid=1488449105433&from=EN [15-03-2017]

[3] European Commission, 2015. Energy Efficiency Directive. Available from: https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-efficiency-directive [07-05-2016]

[4] Eurostat, 2016. Consumption of energy. Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Final_energy_consumption,_EU-28,_2013_(%C2%B9)_(%25_of_total,_based_on_tonnes_of_oil_equivalent)_YB15.png [21-06-2016]

[5] European Commission, 2010. Directive 2010/31/EU of the European parliament and of the council of 19 May 2010 on the energy performance of buildings. Official Journal of the Europian Union. Available from:
http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0031&from=EN [21-06-2016]

[6] Voss, K., Sartori, I., Lollini, R., 2012. *Nearly-zero, Net zero and Plus Energy Buildings – How definitions & regulations affect the solutions*. REHVA journal: Energy efficient renovation. Volume 49, issue 6 pp. 23-27. Available from: https://www.iea-shc.org/data/sites/1/publications/Task40-A-Nearly-zero-Net-zero-and-Plus-Energy-Buildings.pdf [09-05-2016]

[7] Marszal, A.J., Heiselberg, P., Bourelle, J.S., Musall, E., Voss, K., Sartori, I., Napolitano, A., 2011. Zero Energy Buildings – A review of definitions and calculation methodologies. Energy and Buildings, Elsevier 43: 971-979.

[8] IEA, 2016. *Our mission.* Available from: http://www.iea.org/aboutus/ [14-05-2016]

[9] IEA-EBC, n.d. *About EBC.* Available from: http://www.iea-ebc.org/ebc/about/ [14-05-2016]

[10] IEA Heat pump centre, 2005. About HPT. Available from:
http://www.heatpumpcentre.org/EN/ABOUTHPT/Sidor/default.aspx [23-06-2016]

[11] EIA DSM, 2016. Welcome to IEA Demand Side Management Energy Efficiency Technology Collaboration Program. Available from:
http://www.ieadsm.org/ [23-06-2016]

[12] Eurostat, 2015. Electricity generated from renewable sources. Available from:
http://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_from_renewable_sources [21—6-2016]

[13] Eurolex, 2009. Directive 2009/28/EC of the European Parliament and of the council of 23 April 2009. Official Journal of the European Union. Available from: http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009L0028 [21-06-2016]

[14] CBS, 2016. Verbruik hernieuwbare energie toegenomen naar 5.8%. Available from: http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009L0028 [21-06-2016]

[15] Rijksdienst voor Ondernemend Nederland, 2016. Stimulering Duurzame Energieproductie (SDE+). Available from: http://www.rvo.nl/subsidies-regelingen/stimulering-duurzame-energieproductie-sde [21-06-2016]

[16] Rijksdienst voor Ondernemend Nederland, 2016. Energie-investeringsaftrek (EIA). Available from: http://www.rvo.nl/sites/default/files/Energie%20Investeringsaftrek%20-%20Energielijst%202016.pdf [21-06-2016]

[17] Rijkstdienst voor Ondernemend Nederland, 2016. Investeringssubsidie duurzame energie (ISDE). Available from: http://www.rvo.nl/subsidies-regelingen/investeringssubsidie-duurzame-energie [22-06-2016]

[18] Rijskoverheid, 2016. 14% duurzame energie in 2020. Available from: https://www.rijksoverheid.nl/onderwerpen/duurzame-energie/inhoud/meer-duurzame-energie-in-de-toekomst [21-06-2016]

[19] Rijksdienst voor Ondernemend Nederland, n.d. Beleid overheid (bijna) energieneutraal bouwen. Available from: http://www.rvo.nl/onderwerpen/duurzaam-ondernemen/gebouwen/energieneutraal-bouwen/beleid-overheid [23-06-2016]

[20] Provoost, F., 2009. Intelligent Distribution Network design: Technische Universiteit Eindhoven 10.6100/IR651978

[21] Fischer, B., Shilts, E. We plotted 812,000 energy usage curves on top of each other. This is the powerful insight we discovered. Oracle Power Blog. Available from: https://blog.opower.com/2014/10/load-curve-archetypes/
[10-03-2017]

[22] Zachary M. Gill, Michael J. Tierney, Ian M. Pegg & Neil Allan (2010) Low energy dwellings: the contribution of behaviours to actual performance, Building Research & Information, 38:5, 491-508, DOI: 10.1080/09613218.2010.505371

[23] Zhun, Y., Fung, B.C.M., Haghighat, F., Yoshino, H., Morofsky, E., 2011. A systematic procedure to study the influence of occupant behaviour on building energy consumption. Energy and Buildings, Elsevier 43: 1409 – 1417.

[24] Santin, O.G., Itard, L., Visscher, H., 2009. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. Energy and Buildings, Elsevier 41: 1223 – 1232.

[25] EBC, 2015. Energy Flexible Buildings: IEA EBC Annex 67. Available from: http://www.managenergy.net/lib/documents/1454/original_Soren_Jensen_DTI_7_DEC.pdf?1449567551 [24-06-2016].

[26] Yang, R., Wang, L., 2012. Multi-objective optimization for decision-making of energy management in building automation and control. Sustainable Cities and Society, Elsevier, 2: 1-7.

[27] Klein, L., Kwak, J., Kavulya, G., Jazizadeh, F., Becerik-Gerber, B., Varakantham, P., Tambe, M, 2012. Coordinating occupant behavior for building energy and comfort management using multi-agent systems. Automation in Constructions, Elsevier. 22: 525-536.

[28] Lund, P., Lindgren, J., Mikkola, J., Salpakari, J., 2015. Review of energy system flexibility measures to enable high levels of variable renewable electricity. Renewable and Systainable Energy Reviews, Elsevier. 45: 785-807.

[29] Rijksdienst voor Ondernemend Nederland, n.d. Saldering, zelflevering en tariedfkorting. Available from: http://www.rvo.nl/onderwerpen/duurzaam-ondernemen/duurzame-energie-opwekken/duurzame-energie/saldering-en-zelflevering [26-06-2016]

[30] Pérez-Lombard, L., Ortiz, J., Pout, C., 2008. A review on buildings energy consumption information. Elsevier, Energy and buildings. 40: 394-398.

[31] Kolokotsa, D., Rovas, D., Kosmatopoulos, E., Kalaitzakis, K., 2011. A roadmap towards intelligent net zero- and positive-energy buildings. Solar Energy, Elsevier. 85: 3067-3084.

[32] Guo, W., Zhou, M. Technologies toward Thermal Comfort-based and Energy-efficient HVAC Systems: A review. Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA – October 2009.

[33] Lee, Y.M., Horesh, R., Liberti, L., 2015. Optimal HVAC control as demand response with on-site energy storage and generation system. Energy Procedia, Elsevier. 78: 2106 – 2111.

[34] A. Arteconi, N.J. Hewitt, F. Polonara Domestic demand-side management (DSM): Role of heat pumps and thermal energy storage (TES) systems Applied Thermal Engineering, 51 (1–2) (2013), pp. 155–165.

[35] Vieira, A.S., Stewart, R.A., Beal, C.D., 2015. Air source heat pump water heaters in residential buildings in Australia: Identification of key performance parameters. Energy and Buildings, Elsevier. 15: 148 – 162.

[36] Kazmi, H., D'Oca, S., Delmastro C., Lodeweyckx, S., Corgnati, S.P., 2016. Generalizable occupant-driven optimization model for domestic hot water production in NZEB. Applied Energy, Elsevier. 175: 1-15.

[37] E.P. Johnson, 2011. Air-source heat pump carbon footprints: HFC impacts and comparison to other heat sources, Energy Policy. 39: 1369–1381.

[38] Voss, K., Musall, E., 2013. Net Zero Energy Buildings: International projects of carbon neutrality in buildings. IEA SHC Task 40/ECBCS Annex 62.

[39] Andrew Ng, 2016. Machine Learning. Course lectures, Stanford University. Available from: https://www.coursera.org/learn/machine-learning [28-06-2016]

[40] Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. The MIT Press Cambridge, Massauchusetts London, England.

[41] Dan Klein, 2011. CS 188: Artificial Intelligence. Lecture 10: Reinforcement Learning. UC Berkley. Available from:
https://inst.eecs.berkeley.edu/~cs188/fa11/slides/FA11%20cs188%20lecture%2010%20--%20reinforcement%20learning%20(2PP).pdf [28-06-2016]

[42] Centraal Bureau voor de Statistiek, 2014. Woningvoorraad naar eigendom; regio, 2006-2012. Available from:
http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71446ned [01-07-2016]

[43] Portaal, 2014. Informatieboekje duurzame renovatie Soesterberg – augustus 2014 : Warme jas voor uw huis. Available from:
http://www.portaal.nl/stroomversnellingsoesterberg.aspx [02-07-2016]

[44] Bam international, 2015. BAM takes a step towards Smart Districts with 'Smart Grid Ready'-homes in Soesterberg. Available from:
http://www.baminternational.com/articles/bam-takes-the-a-step-towards-smart-districs-with-smart-grid-ready-homes-in-soesterberg [02-07-2016]

[45] Wood, R.J., Al-Muslahi, S.M., O'Callaghan, P.W., Probert, S.D., 1981. Thermally stratified hot water storage systems. Applied Energy, Elsevier. 9: 231 – 242.

[46] Valentina Fabi, Rune Vinther Andersen & Stefano Paolo Corgnati (2013). Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in residential buildings, HVAC&R Research, 19:5, 635-645.

[47] Morren, J., 2015. Lecture 12: Summary of impacts of Decentralized Generation on Distribution Networks. Lecture notes Decentral Power Generation and active networks. Eindhoven University of Technology.

[48] IEA SHC, 2010. *Towards Net Zero Energy Solar buildings*. Available from:
http://www.iea-shc.org/data/sites/1/publications/T40A52Flyer3a.pdf [19-05-2016]

[49] Hermelink, A., Schimschar, S., Boermans, T., Pagliano, L., Zangheri, P., Armani, R., Voss, K., Musall, E., 2013. Towards nearly zero-energy buildings: Definition of common principles under the EPBD. Available from:
https://ec.europa.eu/energy/sites/ener/files/documents/nzeb_full_report.pdf [19-05-2016]

[50] Attia, S., Gratia, E., De Herde, A., Hensen, J.L.M, 2012. Simulation-based decision support tool for early stages of zero-energy building design. Energy and Buildings, Elsevier 49: 2-15.

[51] Kristinsson, J., 2012. Integrated Sustainable Design. Delft Digital Press. ISBN-13: 9789052694078.

[52] Schulze, T., Eicker, U., 2013. Controlled natural ventilation for energy efficient buildings. Energy and Buildings, Elsevier 56: 221 – 232.

[53] Gan, G., 2010. Simulation of buoyancy-driven natural ventilation of buildings – Impact of computational domain. Energy and Buildings, Elsevier 42: 8: 1290 – 1300.

[54] NREL, 2001. Passive Solar Design for Home. Available from:
www.nrel.gov/docs/fy01osti/27954.pdf [19-05-2016]

[55] Schiavoni, S., D'Alessandro, Bianchi, F., Asdrubali, F., 2016. Insulation materials for the building sector: A review and comparative analysis. Renewable and Sustainable Energy Reviews, Elsevier 62: 988-1011.

[56] Su, B., 2008. Building Passive Design and Housing Energy Efficiency. Architectural Science Reviews, 51:3, 277-286.

[57] Rodriguez-Ubinas, E., Montero, C., Porteros, M., Vega, S., Navarro, I., Castillo-Cagigal, M., Matallanas, E., Gutiérrez, A., 2014. Passive design strategies and performance of Net Energy Plus Houses. Energy and Buildings, Elsevier 83: 10-22.

[58] Lotfabadi, P., 2015. Analysing passive solar strategies in the case of high-rise building. Renewable and Sustainable Energy Reviews, Elsevier 52: 1340-1352.

[59] Sachs, H., Lin, W., Lowenberger, A., 2009. Emerging Energy-Saving HVAC Technologies and Practices for the Buildings Sector. American Council for an Energy-Efficient Economy. Available from:
http://aceee.org/sites/default/files/publications/researchreports/A092.pdf [22-05-2016]

[60] Branco, G., Lachal, B., Gallinelli, P., Weber, W., 2004. Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data. Energy and Buildings, Elsevier: 36: 543 – 555.

[61] Haas, R., Auer, H., Biermayr, P., 1997. The impact of consumer behaviour on residential energy demand for space heating. Energy and Buildings, Elsevier 27: 195 – 205.

[62] S. Leth-Petersen, M. Togeby, Demand for space heating in apartment blocks: measuring effect of policy measures aiming at reducing energy consumption, Energy Economics 23 (2001) 387–403.

[63] Andersen, R.V., Toftum, J., Andersen, K.K., Olesen, B.W., 2009. Survey of occupant behaviour and control of indoor environment in Danish dwellings. Energy and Buildings, Elsevier 41: 11 – 16.

[64] Boait, P.J., Dixon, D., Fan, D., Stafford, A., 2012. Production efficiency of hot water for domestic use. Energy and Buildings, Elsevier. 54: 160 – 168.

[65] Silver, D., 2016. Lecture 4: Model-free prediction. University College London. Available from:
http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching_files/MC-TD.pdf [02-08-2016]

[66] Kazmi, H., Cheaib, F., Driesen, J., 2017. Developing synergies for automated optimal control of residential heat pumps. 12th EIA Heat Pump Conference 2017.

[67] Ijaz Dar, U., Sartori, I., Georges, L., Novakovic, V., 2014. Advanced control of heat pumps for improved flexibility of Net-ZEB towards the grid. Energy and Buildings, Elsevier. 69: 74-84.

[68] Luthander, R., Widén, J., Nilsson, D., Palm, J., 2015. Photovoltaic self-consumption in buildings: A review. Applied Energy, Elsevier. 142: 80-94.

[69] Pruissen, O.P., Kamphuis, I.G., 2010. Grote concentraties warmtepompen in een woonwijk en gevolgen elektriciteitsnetwerk. Energy research Centre of the Netherlands (ECN).

[70] Dutch Heat Pump Association DHPA, 2015. Heat pumps in domestic housing and demand management. Available from: http://www.dhpa-online.nl/wp-content/uploads/2011/03/DHPA-32-pag.ENG_.LR_.pdf [23-02-2017]

[71] Ruelens, F., 2016. Residential Demand Response Using Reinforcement Learning. Arenberg Doctoral School, Faculty of engineering science, KU Leuven.

[72] De Coninck, R., Baetens, D., Saelens, D., Woyte, A., Helsen, L., 2013. Rule-based demand side management of domestic hot water production with heat pumps in zero energy neighbourhoods. Journal of Buildiing Performance Simulation.

[73] Sossan, F., Kosek, A.M., Martinenas, S., Marinelli, M., Bindner, H.W., 2013. Scheduling of domestic water heater power demand for maximizing PV self-consumption using model predictive control. Technical university of Denmark.

[74] van den Oosterkamp, P., Koutstaal, O., van der Welle, A., de Joode, J., Lenstra, J., van Hussen, K., Haffner, R., 2014. The role of DSOs in a Smart Grid environment. European Commission.

[75] Ulseth, R., Alonso, M.J., Haugerud, L.P., 2014. Measured load profiles for domestic hot water in buildings heat supply from district heating. The 14th International Symposium on District Heating and Cooling, Sweden.

[76] Lawrence, T.M., Boudreau, M., Helsen, L., Henze, G., Mohammadpour, J., Noonan, D., Pateeuw, D., Pless, S., Watson, R.T. Ten questions concerning integration smart buildings into the smart grid. Building and Environment, Elsevier 108: 273-283.

[77] Peng, K.S., Clayton, T.M., 2016. Model Predictive Prior Reinforcement Learning for a Heat Pump Thermostat. Electrical and Computer Engineering University of Arizona.

[78] O'Neill, D., Levorato, D., Goldsmith, A., Mitra, U., 2010. Residential demand response using reinforcement learning in Smart Grid Communications. First IEEE International Conference. Pp. 409–414, Oct 2010.

[79] Kim, T.T., Poor, H.V., 2011. Scheduling power consumption with price uncertainty. Smart Grid, IEEE Transactions. 2: 519-527.

[80] Chen, Z., Wu, L., Fun, Y., 2012. Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization. Smart Grid, IEEE Transactions. 3: 1822-1831.

[81] Berlink, H., Costa, A.H.R., 2015. Batch reinforcement Learning for Smart Home Energy Management. Proceedings of the Twenty-Fouth International Joint Conference on Artificial Intelligence.

[82] Van Houdt, D., Geysen, D., Claessens, B., Leemans, F., Jespers, L., van Bael, J., 2014. An actively controlled residential heat pump: Potential on peak shaving and maximization of self-consumption of renewable energy. Renewable Energy, Elsevier, 63: 531- 543.

[83] Majcen, D., 2016. Predicting energy consumption and savings in the housing stock: A performance gap analysis in the Netherlands. Architecture and the Built Environment, Delft Technical University.

[84] Power knot, 2011. COPs, EERs, and SEERs; How Efficienct is Your Air Conditioning System? Available from:
http://powerknot.com/wp-content/uploads/sites/6/2011/03/Power_Knot_about_COP_EER_SEER.pdf [21-03-2017].

[85] Arcuri, N., Carpino, C., De Simone, M., 2016. The role of the thermal mass in nZEB with different energy systems. Energy Procedia, Elsevier 101: 121 – 128.

[86] Merosch, 2015. De effecten van en oplossingen voor aanpassing van salderingsregeling op NOM-woningen in 2020. Available from:
http://www.merosch.nl/download/CAwdEAwUUkBFXw==&inline=0 [07-04-2017].

Appendix A

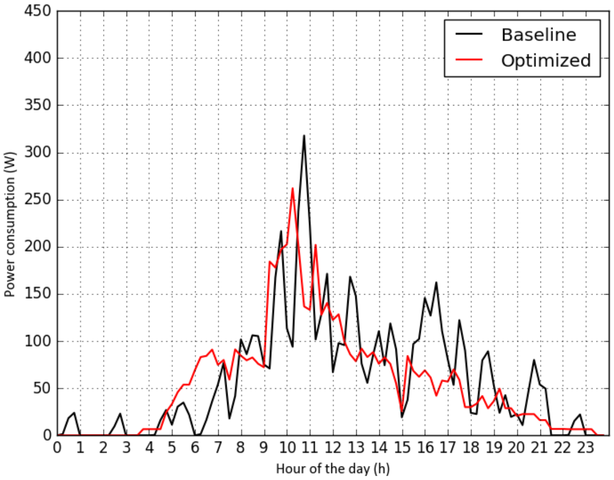


Figure A.1. House 1 power profiles baseline vs. optimized to T_{max} of 55 °C.

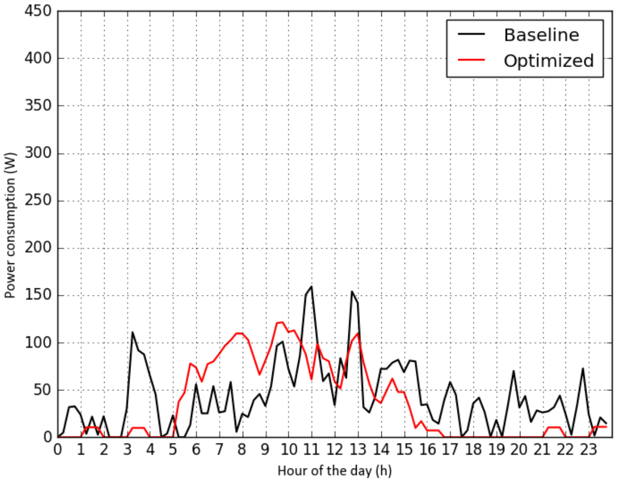


Figure A.2. House 2 power profiles baseline vs. optimized to T_{max} of 55 °C.

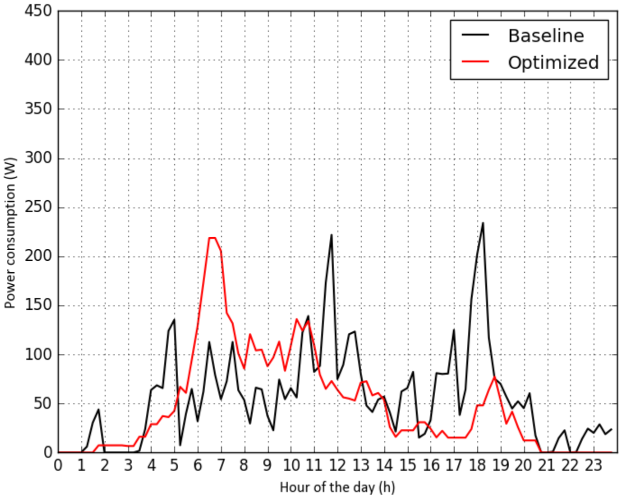


Figure A.3. House 3 power profiles baseline vs. optimized to T_{max} of 55 °C.

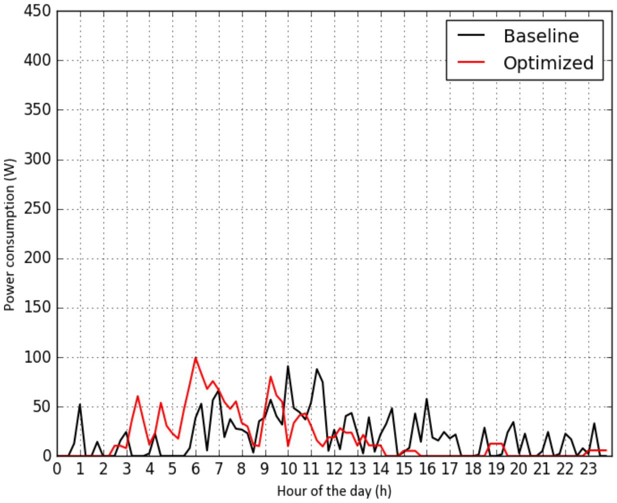


Figure A.4. House 4 power profiles baseline vs. optimized to T_{max} of 55 °C.

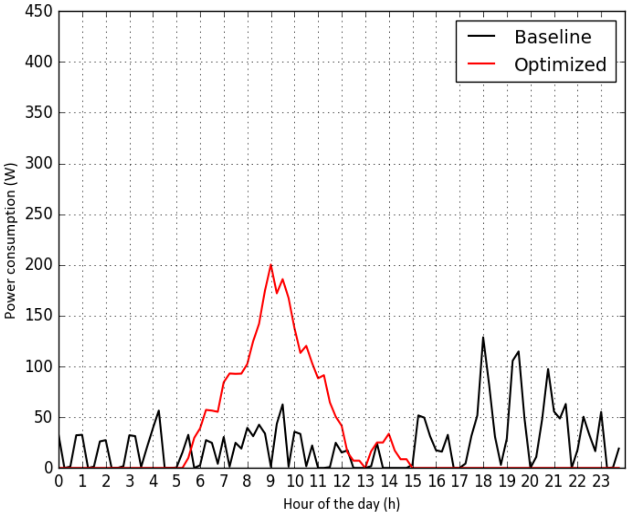


Figure A.5. House 5 power profiles baseline vs. optimized to T_{max} of 55 °C.

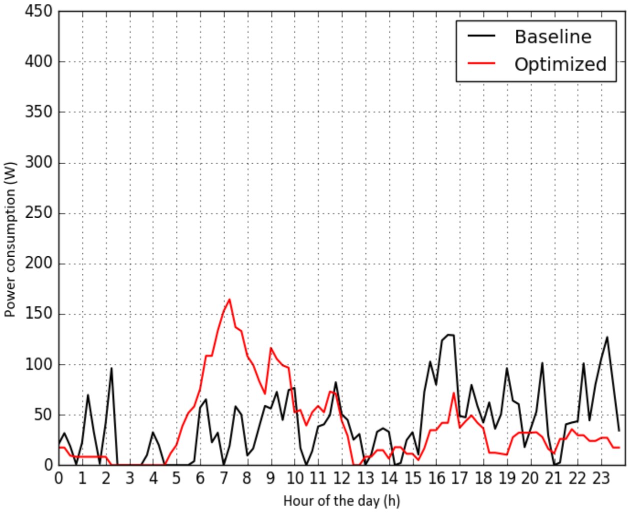
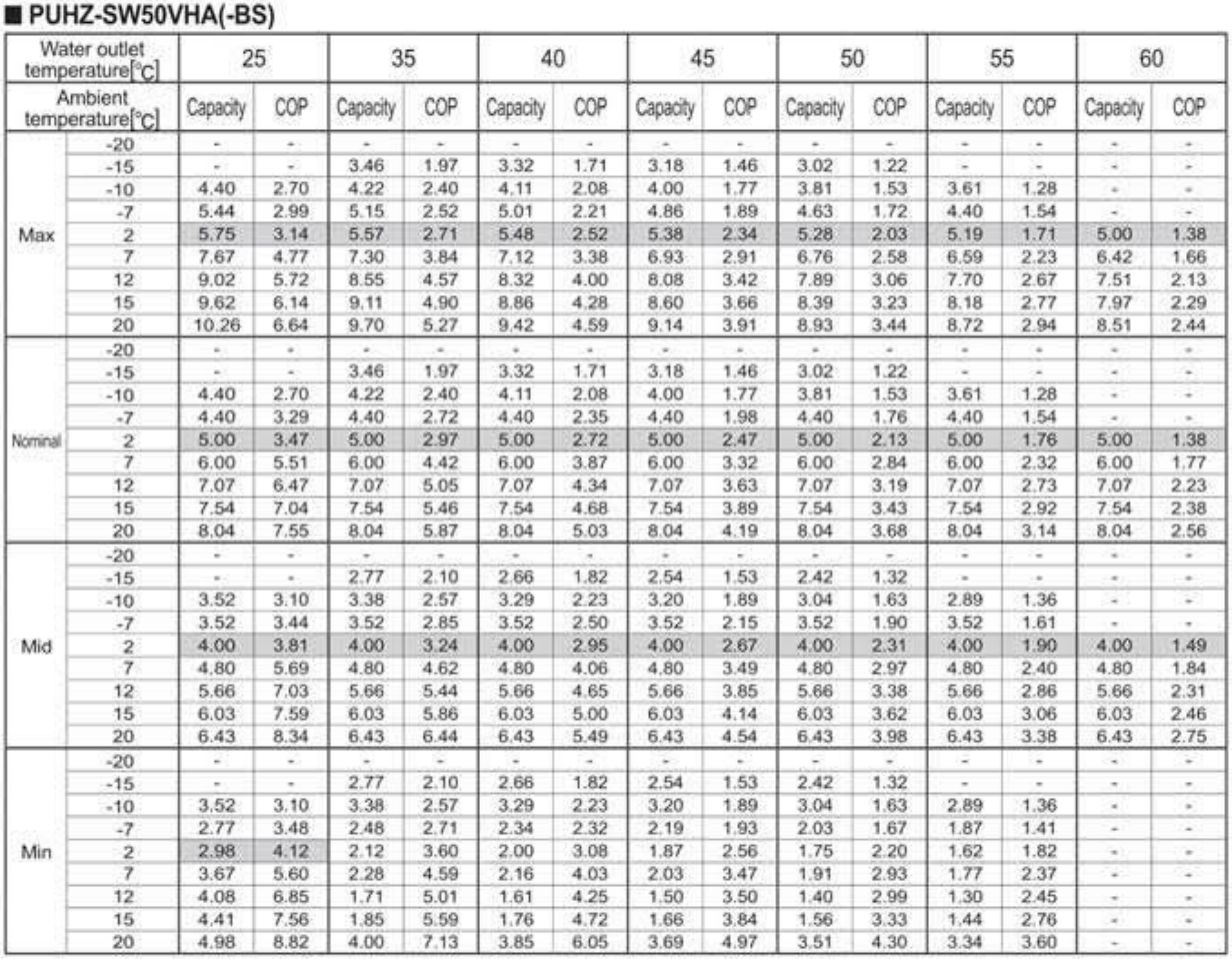


Figure A.6. House 6 power profiles baseline vs. optimized to T_{max} of 55 °C.

Appendix B



■ PUHZ-SW50VHA(-BS)

Water outlet temperature[°C]		25		35		40		45		50		55		60	
Ambient temperature[°C]		Capacity	COP	Capacity	COP	Capacity	COP	Capacity	COP	Capacity	COP	Capacity	COP	Capacity	COP
Max	-20	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-15	-	-	3.46	1.97	3.32	1.71	3.18	1.46	3.02	1.22	-	-	-	-
	-10	4.40	2.70	4.22	2.40	4.11	2.08	4.00	1.77	3.81	1.53	3.61	1.28	-	-
	-7	5.44	2.99	5.15	2.52	5.01	2.21	4.86	1.89	4.63	1.72	4.40	1.54	-	-
	2	5.75	3.14	5.57	2.71	5.48	2.52	5.38	2.34	5.28	2.03	5.19	1.71	5.00	1.38
	7	7.67	4.77	7.30	3.84	7.12	3.38	6.93	2.91	6.76	2.58	6.59	2.23	6.42	1.66
	12	9.02	5.72	8.55	4.57	8.32	4.00	8.08	3.42	7.89	3.06	7.70	2.67	7.51	2.13
	15	9.62	6.14	9.11	4.90	8.86	4.28	8.60	3.66	8.39	3.23	8.18	2.77	7.97	2.29
	20	10.26	6.64	9.70	5.27	9.42	4.59	9.14	3.91	8.93	3.44	8.72	2.94	8.51	2.44
Nominal	-20	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-15	-	-	3.46	1.97	3.32	1.71	3.18	1.46	3.02	1.22	-	-	-	-
	-10	4.40	2.70	4.22	2.40	4.11	2.08	4.00	1.77	3.81	1.53	3.61	1.28	-	-
	-7	4.40	3.29	4.40	2.72	4.40	2.35	4.40	1.98	4.40	1.76	4.40	1.54	-	-
	2	5.00	3.47	5.00	2.97	5.00	2.72	5.00	2.47	5.00	2.13	5.00	1.76	5.00	1.38
	7	6.00	5.51	6.00	4.42	6.00	3.87	6.00	3.32	6.00	2.84	6.00	2.32	6.00	1.77
	12	7.07	6.47	7.07	5.05	7.07	4.34	7.07	3.63	7.07	3.19	7.07	2.73	7.07	2.23
	15	7.54	7.04	7.54	5.46	7.54	4.68	7.54	3.89	7.54	3.43	7.54	2.92	7.54	2.38
	20	8.04	7.55	8.04	5.87	8.04	5.03	8.04	4.19	8.04	3.68	8.04	3.14	8.04	2.56
Mid	-20	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-15	-	-	2.77	2.10	2.66	1.82	2.54	1.53	2.42	1.32	-	-	-	-
	-10	3.52	3.10	3.38	2.57	3.29	2.23	3.20	1.89	3.04	1.63	2.89	1.36	-	-
	-7	3.52	3.44	3.52	2.85	3.52	2.50	3.52	2.15	3.52	1.90	3.52	1.61	-	-
	2	4.00	3.81	4.00	3.24	4.00	2.95	4.00	2.67	4.00	2.31	4.00	1.90	4.00	1.49
	7	4.80	5.69	4.80	4.62	4.80	4.06	4.80	3.49	4.80	2.97	4.80	2.40	4.80	1.84
	12	5.66	7.03	5.66	5.44	5.66	4.65	5.66	3.85	5.66	3.38	5.66	2.86	5.66	2.31
	15	6.03	7.59	6.03	5.86	6.03	5.00	6.03	4.14	6.03	3.62	6.03	3.06	6.03	2.46
	20	6.43	8.34	6.43	6.44	6.43	5.49	6.43	4.54	6.43	3.98	6.43	3.38	6.43	2.75
Min	-20	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-15	-	-	2.77	2.10	2.66	1.82	2.54	1.53	2.42	1.32	-	-	-	-
	-10	3.52	3.10	3.38	2.57	3.29	2.23	3.20	1.89	3.04	1.63	2.89	1.36	-	-
	-7	2.77	3.48	2.48	2.71	2.34	2.32	2.19	1.93	2.03	1.67	1.87	1.41	-	-
	2	2.98	4.12	2.12	3.60	2.00	3.08	1.87	2.56	1.75	2.20	1.62	1.82	-	-
	7	3.67	5.60	2.28	4.59	2.16	4.03	2.03	3.47	1.91	2.93	1.77	2.37	-	-
	12	4.08	6.85	1.71	5.01	1.61	4.25	1.50	3.50	1.40	2.99	1.30	2.45	-	-
	15	4.41	7.56	1.85	5.59	1.76	4.72	1.66	3.84	1.56	3.33	1.44	2.76	-	-
	20	4.98	8.82	4.00	7.13	3.85	6.05	3.69	4.97	3.51	4.30	3.34	3.60	-	-

Fig. B1. Heat pump COP specifications