



Master's in Industrial Engineering

Master's thesis

**Development of a Control Algorithm Based on Machine
Learning for PV-Driven Heat Pumps and Compression Chillers**

Author

Ana Elechiguerra Batlle

Supervised by

Prof. Dr.-Ing. Thomas Sattelmayer

Dr.-Ing. Markus Spinnler

Bernd Heithorst, M.Sc.

Madrid

June 2020

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título "Development of a Control Algorithm Based on Machine Learning for PV-Driven Heat Pumps and Compression Chillers" en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el curso académico 2019-2020 es de mi autoría, original e inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.



Fdo.: Ana Elechiguerra Batlle

Fecha: 10/06/2020

Autorizada la entrega del proyecto
EL DIRECTOR DEL PROYECTO



Fdo.: Bernd Heithorst

Fecha: 11/06/20

Resumen

DESARROLLO DE UN ALGORITMO DE CONTROL BASADO EN APRENDIZAJE AUTOMÁTICO PARA BOMBAS DE CALOR Y MÁQUINAS FRIGORÍFICAS ALIMENTADAS POR ENERGÍA SOLAR FOTOVOLTAICA

Autor: Elechiguerra Batlle, Ana.

Director: Heithorst, Bernd.

Entidad Colaboradora: Universidad Técnica de Múnich, Alemania

RESUMEN DEL PROYECTO

Introducción

El sector de la construcción consume aproximadamente el 40% del consumo mundial de energía y la mitad de este consumo se utiliza para calentar, enfriar y ventilar [1]. En Alemania, las viviendas son los mayores consumidores de calor (44%), seguidos de la industria (38%) [2]. Dos tercios de este consumo de calor están cubiertos por gas y petróleo [3]. Más del 60% de los sistemas de calefacción, ventilación y aire acondicionado (HVAC) instalados en todo el mundo se alimentan mediante combustibles no renovables. Dado que los sistemas de climatización representan una parte tan importante del consumo de combustibles fósiles y de las emisiones de gases de efecto invernadero, el uso de soluciones más respetuosas con el medio ambiente, como las bombas de calor, tienen un gran potencial en la descarbonización [1].

El acoplamiento entre sectores ofrece la oportunidad de aprovechar la potencia fluctuante producida por las fuentes de energía renovable, como la energía solar fotovoltaica. Concreta-

mente, el acoplamiento de los sectores de la electricidad y el calor permite utilizar el exceso de energía producida por estas fuentes y proporciona flexibilidad al sistema [4]. La combinación de los sectores del calor y la electricidad (conocida como "power-to-heat") se considera una de las estrategias de combinación de sectores más atractivas, ya que tanto la generación de calor a partir de la electricidad como el almacenamiento de calor son procesos baratos [5]. Las tecnologías más representativas de "power-to-heat" son las bombas de calor y las máquinas frigoríficas, que convierten la electricidad en calor y frío respectivamente [6].

La optimización del consumo de energía en los edificios es un constante compromiso entre confort, consumo y coste [7]. En este contexto, la investigación demuestra que la implantación de sistemas inteligentes, y en concreto de sistemas de control automático basados en técnicas inteligentes, permite atender los diferentes objetivos, a la vez que ahorrar energía y reducir las emisiones de CO₂ [8]. Por este motivo, tal sistema se desarrolla e integra al proyecto "PVCool" del Instituto de Termodinámica de la Universidad Técnica de Munich.

Objetivo

El Proyecto "PVCool" es un proyecto de la Universidad Técnica de Munich, en cooperación con el Instituto de Investigación del Medio Ambiente y la Energía de la Universidad Hamad bin Khalifa de Qatar. El proyecto investiga el ajuste óptimo del acoplamiento de un ciclo de vapor frío (bomba de calor o máquina frigorífica) y una unidad de generación fotovoltaica para diferentes aplicaciones [9]. En esta tesis, sólo se investiga la aplicación a una casa unifamiliar en Alemania. El sistema energético investigado se ilustra en la figura 1. En el centro

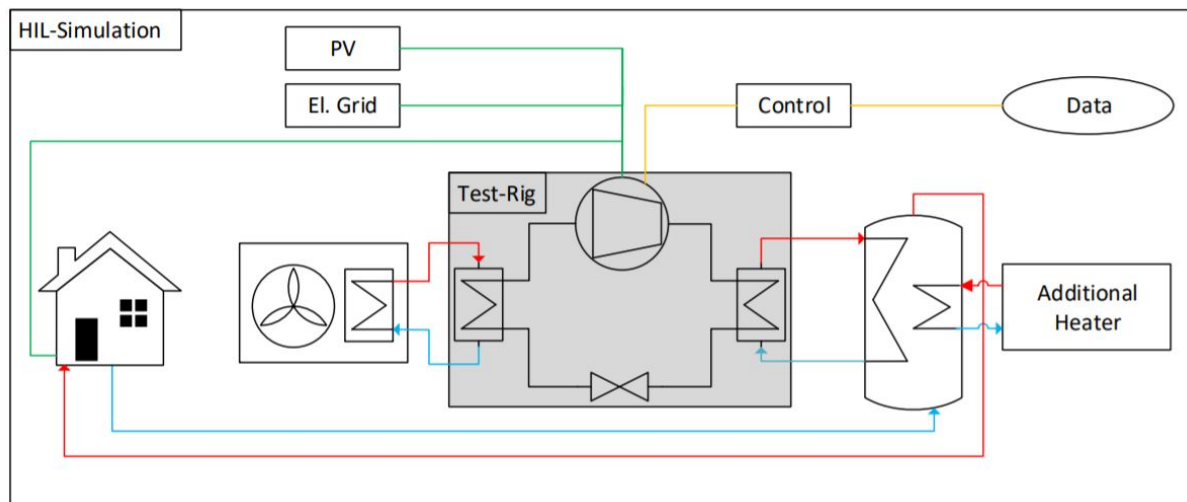


Figure 1 – Esquema del sistema energético estudiado [10]

del esquema en gris rodeado por un marco negro, se representa el ciclo de vapor. Un ciclo de vapor consiste en un evaporador, un compresor, un condensador y una válvula de expansión.

Las líneas azules y rojas representan los conductos de agua fría y caliente respectivamente. El ciclo de vapor puede funcionar como una bomba de calor o como una máquina frigorífica dependiendo de si el sistema está proporcionando calor o frío al edificio respectivamente. En la figura [1], el sistema está funcionando como una bomba de calor, dado que está recibiendo calor de un intercambiador de calor aire-agua (a la izquierda del banco de ensayos) y transfiriendo calor al almacenamiento térmico del edificio (a la derecha). Las líneas verdes y amarillas de la figura [1] representan los flujos eléctricos y de control respectivamente. La electricidad utilizada en el edificio puede ser suministrada por la red o por la unidad fotovoltaica. El exceso de producción fotovoltaica se inyecta en la red. La unidad de control recoge los datos del sistema y deriva la energía deseada que será suministrada por el ciclo de vapor [9].

El objetivo principal de esta tesis es desarrollar e implementar un control basado en el aprendizaje automático (machine learning) para el sistema energético del edificio descrito anteriormente. El control deberá minimizar el gasto energético y maximizar el auto-consumo. Otro objetivo es el desarrollo de un estudio de parámetros, para determinar la combinación de dichos parámetros que es económicamente óptima para el sistema. Por último, se comparan los diferentes controles desarrollados en el proyecto PVcool. Estos controles incluyen un control basado en reglas desarrollado por Achatz [11], un control predictivo por modelo diseñado por Rauscher [9] y el control basado en machine learning desarrollado en el marco de esta tesis.

Estado del arte

Esta tesis explora la implementación de un control inteligente en un sistema de calefacción, ventilación y aire acondicionado alimentado por energía fotovoltaica para optimizar el consumo de energía de un edificio unifamiliar. Por ello, la bibliografía examinada se centra en los progresos realizados en este ámbito, y más concretamente en los controles basados en machine learning centrados en las bombas de calor y la optimización del consumo de energía de edificios. Las técnicas encontradas en la literatura pertenecen principalmente a tres de las cuatro subáreas del aprendizaje automático. Estas incluyen el aprendizaje supervisado, el aprendizaje por refuerzo y el aprendizaje evolutivo [12]. A continuación se resume un ejemplo significativo de una técnica de cada categoría.

Han et al. [13] estudian los avances en los métodos de control de aprendizaje por refuerzo aplicados en la energía para edificios. Describen los diferentes algoritmos y explican el potencial y las ventajas de la implementación de este control sin modelo en el campo de la automatización de edificios inteligentes. Sin embargo, resaltan que la "maldición de la dimensionalidad" dificulta su implementación.

En el campo del aprendizaje supervisado, Mohammadzaheri et al. [14] estudian las estrategias de control más comunes basadas en las redes neuronales. Los autores concluyen que las redes neuronales primero se combinaron con controles tradicionales, luego se utilizaron como controladores únicos para finalmente volver como controladores auxiliares en el diseño de

controles híbridos.

En cuanto al aprendizaje evolutivo, Wright et al. [15] abordan el problema del diseño térmico de los edificios como una cuestión de optimización de criterios múltiples. Para determinar un compromiso entre el confort térmico y el coste de la energía, los autores aplican con éxito un algoritmo genético multicriterio.

Por último, Ahmad et al. [16] observan que técnicas híbridas que combinan las estrategias anteriormente mencionadas u otras, generalmente llevan a mejores soluciones y rendimiento. Se revisan diferentes estrategias híbridas en la literatura. Por ejemplo, Mocanu et al. [17] y Wei et al. [18] aplican con éxito el aprendizaje profundo por refuerzo (una combinación de redes neuronales y aprendizaje por refuerzo) para desarrollar un control de la energía de un edificio. Otro ejemplo es el de Matallanas et al. [19]. Los autores utilizan un algoritmo genético para ajustar la red neuronal de un sistema de gestión del lado de la demanda activa.

Methodología

Los procedimientos y la metodología de esta tesis son los siguientes:

- El primer paso es una extensa revisión de la literatura sobre controles de aprendizaje automático para sistemas energéticos para edificios. Se elige el enfoque más prometedor listando y analizando las ventajas e inconvenientes de las diferentes técnicas.
- A continuación, el enfoque elegido se programa utilizando Simulink y MATLAB. El software se desarrolla siguiendo un enfoque bottom-up, comenzando con las funciones básicas y probándolas para facilitar la corrección de posibles errores. Para ello, se introducen datos de prueba en las funciones básicas diseñadas y las salidas del software se observan y rastrean. Una vez en funcionamiento, todas las subunidades básicas se ensamblan para formar un sistema más complejo. Por último, el control de supervisión se integra en el modelo Simulink del proyecto PVCool.
- El control se verifica mediante simulaciones a corto plazo. La verificación incluye la conformidad con los límites y objetivos del control.
- El estudio de los parámetros tiene como objetivo determinar la combinación económicamente óptima de los parámetros del sistema. El estudio sólo considera los elementos susceptibles de tener el mayor impacto en los costes. Se consideran cuatro volúmenes diferentes de almacenamiento térmico, tres dimensiones distintas de la bomba de calor y tres tamaños diferentes de la unidad de generación fotovoltaica. Esto da lugar a 36 escenarios diferentes que se compararan a partir de simulaciones anuales. Con los resultados de las simulaciones se calculan valores de comparación que permiten orientar la decisión del escenario óptimo. Los valores de comparación incluyen: el coste normalizado de la energía, el coste normalizado del calor, el coste normalizado de la electricidad, el ratio de auto-consumo y las emisiones generadas.

-
- Finalmente, se comparan los tres controles desarrollados en el marco del proyecto PV-Cool. Los parámetros económicamente óptimos determinados en el estudio de parámetros se utilizan como condiciones límites comunes. El control basado en reglas desarrollado por Achatz [11], el control predictivo por modelo diseñado por Rauscher [9] y el control de supervisión basado en el aprendizaje automático desarrollado en esta tesis se comparan usando los resultados de simulaciones anuales de alta resolución temporal.

Resultados

Después de una extensa revisión del estado del arte y un análisis de las ventajas y desventajas de cada técnica, se concluye que el aprendizaje profundo por refuerzo es la técnica de aprendizaje automático más prometedora para el sistema investigado. El sistema se desarrolla, prueba e implementa en el modelo Simulink del proyecto PVCool. Las pruebas realizadas mediante simulaciones a corto plazo muestran que se respetan los límites de control y que el control también cumple con el objetivo deseado.

El estudio de parámetros concluye que la configuración óptima para el sistema de energía investigado consiste en una bomba de calor con una potencia eléctrica máxima de 3 kW, un acumulador térmico de 750 L y una unidad de generación fotovoltaica de 5 kWp. Este escenario logra el menor costo nivelado de energía y las menores emisiones, evitando elementos de tamaño sobre- o infradimensionados.

Finalmente, con la configuración elegida, se realizan simulaciones anuales con los diferentes controles desarrollados dentro del proyecto PVCool: el control basado en reglas desarrollado por Achatz [11], el control predictivo por modelo desarrollado por Rauscher [9] y el control de aprendizaje profundo por refuerzo desarrollado en el marco de esta tesis. Los resultados muestran que el control predictivo del modelo logra los menores costes energéticos. Concretamente, el control predictivo por modelo logra un coste de energía 2,2% menor que el coste logrado por el control basado en reglas y 6,7% menor que el coste obtenido con el control de aprendizaje profundo por refuerzo. El mayor ratio de autoconsumo se consigue con el control basado en reglas. El ratio de autoconsumo alcanzado cuando se utiliza el control basado en reglas es 24% y 37,8% mayor que el ratio de autoconsumo obtenido por el control predictivo por modelo y el control de aprendizaje profundo por refuerzo respectivamente.

Conclusión

Los resultados de la evaluación del control muestran resultados subóptimos del control de aprendizaje profundo por refuerzo desarrollado. Esto puede explicarse por diferentes razones. En primer lugar, el hecho de que la exploración por prueba y error realizada por el control de aprendizaje profundo por refuerzo a veces puede quedar atrapada en una solución subóp-

tima. Por otro lado, el proceso de aprendizaje tuvo que ser acortado con el fin de reducir el tiempo de computación, lo cual también disminuyó el rendimiento del control. Por último, los parámetros y el diseño del control podrían aún ser perfeccionados.

Aunque los costes de energía más bajos se logran mediante el control de predicción por modelo, se debe tener en cuenta que este control utiliza predicciones correctas al 100%, que son poco probables en la realidad. Teniendo en cuenta que los resultados en términos de costes energéticos entre el control basado en reglas y el control predictivo del modelo son muy bajos, es probable que el control basado en reglas sea el más adecuado en una aplicación real no simulada. Además, este control es más simple de implementar, más rápido y conduce al mayor ratio de auto-consumo y a las emisiones más bajas.

Summary

DEVELOPMENT OF A CONTROL ALGORITHM BASED ON MACHINE LEARNING FOR PV-DRIVEN HEAT PUMPS AND COMPRESSION CHILLERS

Author: Elechiguerra Batlle, Ana.

Director: Heithorst, Bernd.

Collaborating Institution: Technical University of Munich

SUMMARY OF THE PROJECT

Introduction

The building sector consumes approximately 40% of global energy consumption and half of this consumption is used for heating, cooling and ventilating [1]. In Germany, households use the largest share of heat (44%), followed by industry (38%) [2]. Two-thirds of this heat consumption are covered by gas and oil [3]. Over 60% of the installed heating, ventilation and air conditioning (HVAC) systems worldwide are driven by non-renewable fuels. Since HVAC systems are accountable for such an important share of the consumption of fossil fuels and the emissions of greenhouse gasses, the use of more environmentally friendly solutions such as heat pumps have a great potential in decarbonization [1].

The linking of sectors provides an opportunity of taking advantage of the fluctuating power produced by renewable energy sources such as photovoltaic (PV). Specifically, the coupling of the power and heat sectors allows using the excess power produced by these sources and provides flexibility to the system [4]. The linking of the heat and power sectors (known as

power-to-heat) is considered one of the most attractive sector coupling strategies since both generating heat from electricity and storing heat are inexpensive processes [5]. The most representative technologies of power-to-heat are heat pumps and compression chillers, which turn electricity into heat and cold, respectively [6].

Energy consumption optimization in buildings is a constant trade-off between comfort, consumption, and cost [7]. In this context, research shows that the implementation of intelligent systems, and specifically automatic control systems based on intelligent techniques, allows dealing with the different goals, while saving energy and reducing CO₂ emissions [8]. Therefore, such a system is developed and implemented to the PVCool project at the Institute of Thermodynamics of the Technical University of Munich.

Objective

The PVCool Project is a project of the Technical University of Munich, in cooperation with the Qatar Environment and Energy Research Institute of the Hamad bin Khalifa University in Qatar. The project investigates the optimal setting of the coupling of a cold vapor cycle (heat pump or compression chiller) and a photovoltaic generation unit for different applications [9]. In this thesis, only an application in a single-family house in Germany is investigated.

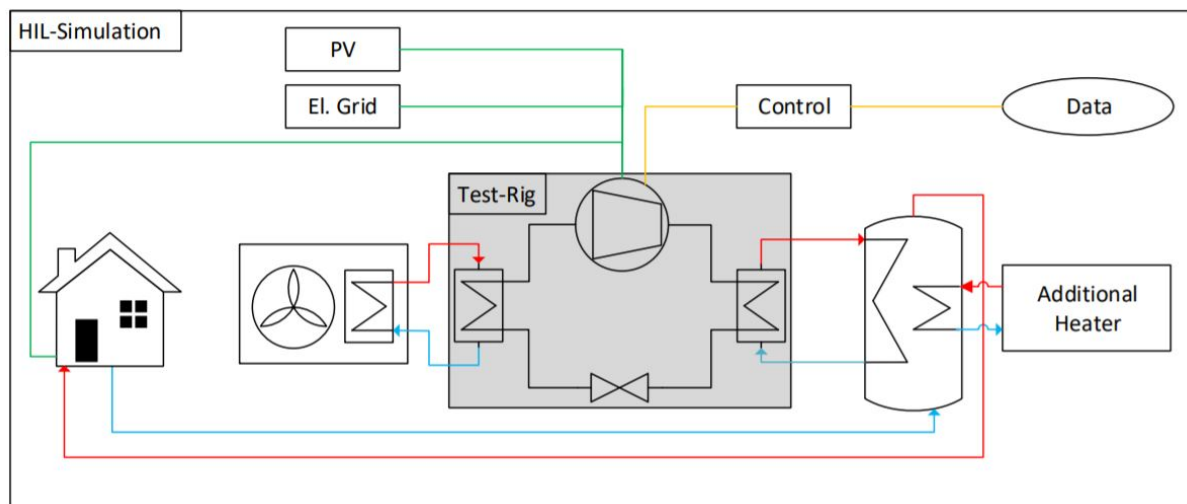


Figure 2 – Scheme of the studied building energy system [10]

The investigated energy system is illustrated in figure 2. In the center of the scheme surrounded by a black frame, the cold vapor cycle is represented. A cold vapor cycle consists of an evaporator, a compressor, a condenser, and an expansion valve [9]. Blue and red lines represent cold and hot-water pipes, respectively. The vapor cycle can work either as a heat pump or as a compressor chiller depending on whether the system is providing heat or cold to the building, respectively. In figure 2, the system is working as a heat pump, as it is get-

ting heat from an air heat exchanger (on the left of the test-rig) and transferring heat to the building's thermal storage (on the right). Green and yellow lines in figure 2 depict electric and control flows, respectively. The electricity used in the building energy system can be supplied by the grid or the PV unit. Excess PV production is fed to the grid. The control unit collects data from the system and derives the desired power to be provided by the vapor cycle [9].

The main objective of this thesis is to develop and implement a machine learning-based control for the described building energy system that minimizes energy cost and maximizes self-consumption. A further goal is the development of a parameter study, to determine the economically optimal combination of system parameters. Finally, the controls developed within the PVcool projects are compared. The controls include a rule-based control developed by Achatz [11], a model-predictive control designed by Rauscher [9] and the machine-learning control developed within this thesis.

State of the Art

This master's thesis explores the implementation of an intelligent control in PV-driven heating, ventilation, air conditioning system to optimize energy consumption. Therefore, the reviewed literature focuses on progress made on this area, and more specifically on controls based on machine learning focused on heat pumps and building energy consumption optimization. The techniques found in the literature research mainly belong to three of the four subareas of machine learning. These include supervised learning, reinforcement learning and evolutionary learning [12]. One significant example of technique of each category is summarized hereafter.

Han et al. [13] study the advances in reinforcement learning control methods applied in building energy. They describe the different algorithms and explain the potential and advantages of the implementation of this model-free control in the field of smart building automation. However, the curse of dimensionality difficults its implementation.

In the field of supervised learning, Mohammadzaheri et al. [14] develop a review of the most common control strategies based on neural networks. The authors conclude that neural networks were first mainly combined with traditional controllers, then used as sole controllers to finally return as auxiliary controllers in the design of hybrid ANN controls.

Regarding evolutionary learning, Wright et al. [15] approach the problem of building thermal design as a multi-criterion optimization issue. To determine a compromise between thermal comfort and energy cost, the authors successfully apply a multi-criterion genetic algorithm.

Finally, Ahmad et al. [16] notice hybrid technique combining the aforementioned or further strategies, generally lead to better solutions and performance. Different hybrid strategies are reviewed in the literature. For instance, Mocanu et al. [17] and Wei et al. [18] successfully apply

deep reinforcement learning (a combination of neural networks and reinforcement learning) to building energy control and Matallanas et al. [19] use a genetic algorithm to tune the neural network of an active-demand side management system.

Methodology

The steps and methodology of this thesis are as follow:

- The first step is an extensive literature review on machine learning controls for building energy systems. The most promising approach is chosen by listing and analyzing the advantages and disadvantages of the different techniques.
- The chosen approach is then programmed using Simulink and MATLAB. The software is developed following a bottom-up approach, starting with basic features and testing them to ease the correction of possible errors. To do so, test inputs are fed into the basic features and software outputs are observed and tracked. Once functioning, all the basic sub-units are assembled to form a more complex system. Finally, the supervisory control is integrated into the PVCool Simulink model.
- The control is verified using short-term simulations. The verification includes the compliance to control bounds and objectives.
- The parameter study aims at the determination of the economically optimal combination of system parameters. The study only considers the elements susceptible to having the highest impact on costs. These include four different volumes of thermal storage, three different sizes of the heat pump, and three different sizes of the PV generation unit. This results in 36 different scenarios to be compared with annual simulations. With the results of the simulations, comparative values are calculated to guide the decision of the optimal scenario. The comparative values include the levelized cost of energy, the levelized cost of heat, the levelized cost of electricity, the self-consumption ratio, and the generated emissions.
- Finally, the three controls developed within the PVCool projects are compared. The economically optimal parameters determined in the parameter study are used as common boundary conditions. The rule-based control developed by Achatz [11], the model predictive control designed by Rauscher [9], and the supervisory control based on machine learning developed within this thesis are compared using the results of high temporal resolution annual simulations.

Results

After an extensive literature review and analysis of the advantages and disadvantages of each technique, it is concluded that deep reinforcement learning is the most promising machine-

learning technique for the investigated system. The method is developed, tested, and implemented in the PVCool Simulink model. Tests carried using short-term simulations show that the control bounds are respected and that the control also complies with the desired objective.

The parameter study concludes that the optimal setup for the investigated energy system consists of a heat pump with a maximum electrical power of 3 kW, a thermal storage of 750 L and a PV generation unit of 5 kWp. This scenario achieves the lowest leveled cost of energy and the lowest emissions while avoiding undersized or oversized elements.

Finally, with the chosen setup, annual simulations are performed with the different controls developed within the PVCool project: the rule-based control developed by Achatz [11], the model predictive control developed by Rauscher [9] and the deep reinforcement learning control developed within this thesis. Results show that the model predictive control achieves the lowest energy costs. Indeed, the model predictive control achieved an energy cost that was 2,2% lower than the cost achieved by the rule-based control and 6,7% lower than the cost obtained with the deep reinforcement learning control. The highest self-consumption ratio is achieved by the rule-based control. The self-consumption ratio achieved when using the rule-based control is 24% and 37,8% higher than the self-consumption ratio obtained by the model predictive control and deep reinforcement learning control, respectively.

Conclusion

Results of the control evaluation show suboptimal results of the developed deep reinforcement learning control. This can be explained by different reasons. First, the fact that the trial and error exploration performed by the deep reinforcement learning control sometimes can get caught in a suboptimal solution. The shortening of the learning process to reduce computational time also diminished the performance of the control. Finally, the parameters and design of the control could still be improved.

Although the lowest energy costs are achieved by the model predictive control, this control used 100% correct predictions which are unlikely to be available in reality. Taking into account that the results in terms of energy costs between rule-based control and model predictive control are very low, it is likely that the rule-based control would be the most suitable in a non-simulated real application. Additionally, this control is simpler to implement, faster and leads to the highest self-consumption and the lowest emissions.

Bibliography

- [1] F. RUELENS, S. IACOVELLA, B. CLAESSENS AND R. BELMANS: Learning Agent for a Heat-Pump Thermostat with a Set-Back Strategy Using Model-Free Reinforcement Learning. In: *Energies* 8 (2015), Nr. 8, S. 8300–8318. <http://dx.doi.org/10.3390/en8088300>. – DOI 10.3390/en8088300
- [2] AG ENERGIEBILANZEN E.V.: *Anwendungsbilanzen für die Endenergiesektoren in Deutschland in den Jahren 2013 bis 2017*. Berlin,
- [3] UMWELT BUNDESAMT: *Erneuerbare Energien in Zahlen*. <https://www.umweltbundesamt.de/themen/klima-energie/erneuerbare-energien/erneuerbare-energien-in-zahlen#uberblick>
- [4] H. Ü. YILMAZ, R. HARTEL, D. KELLES, R. MCKENNA AND W. FICHTNER: *Analysis of the potential for Power-to-Heat/Cool applications to increase flexibility in the European electricity system until 2030* (Policy Report)
- [5] A. BLOESS, W.-P. SCHILL AND A. ZERRAHN: Power-to-heat for renewable energy integration: A review of technologies. In: *Applied Energy* 212 (2018), S. 1611–1626. <http://dx.doi.org/10.1016/j.apenergy.2017.12.073>. – DOI 10.1016/j.apenergy.2017.12.073.
- [6] BUNDESVERBAND WÄRMEPUMPE (BWP) E.V.: *Sektorkopplung*. <https://www.waermepumpe.de/politik/sectorkopplung/>. Version: 2017
- [7] W. YU, B. LI, H. JIA, M. ZHANG AND D. WANG: Application of multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design. In: *Energy and Buildings* 88 (2015), S. 135–143. <http://dx.doi.org/10.1016/j.enbuild.2014.11.063>. – DOI 10.1016/j.enbuild.2014.11.063.
- [8] P. SHAIKH, N. NOR, P. NALLAGOWNDEN, I. ELAMVAZUTHI AND T. IBRAHIM: A review on optimized control systems for building energy and comfort management of smart sustainable buildings. In: *Renewable and Sustainable Energy Reviews* 34 (2014), S. 409–429. <http://dx.doi.org/10.1016/j.rser.2014.03.027>. – DOI 10.1016/j.rser.2014.03.027.
- [9] B. RAUSCHER: *Development and Implementation of Control Strategies for PV-Driven Heat Pumps and Compression Chillers*, Technische Universität München, Diss., 2019

- [10] HEITHORST, Bernd ; BAUER, Gregor ; RAUSCHER, Benedikt ; IRRGANG, Ludwig ; SPINNLER, Markus ; SATTELMAYER, Thomas: Experimental Comparison of Scroll and Swash-Plate Compressors for PV Driven Compression Chillers and Heat Pumps. In: HÄBERLE, Andreas (Hrsg.): *EuroSun2018*. Freiburg : International Solar Energy Society, op. 2018. – ISBN 978-3-9820408-0-6, S. 1–12
- [11] H. ACHATZ: *Simulations on Covering the Heating and Cooling Demand of Buildings in Temperate and Desert Climate with PV Driven Heat Pumps and Compression Chillers*. Munich, Technical University of Munich, Bachelor's thesis, 2019
- [12] MARSLAND, Stephen: *Machine learning: An algorithmic perspective*. Second edition. Boca Raton, Fla. : CRC Press/Taylor & Francis, 2015 (A Chapman & Hall book). – ISBN 9781466583283
- [13] M. HAN, X. ZHANG, L. XU, R. MAY, S. PAN AND J. WU: A review of reinforcement learning methodologies on control systems for building energy. In: *Working papers in transport, tourism*, 2 (2018)
- [14] MOHAMMADZAHERI, Morteza ; CHEN, Lei ; GRAINGER, Steven: A critical review of the most popular types of neuro control. In: *Asian Journal of Control* 14 (2012), Nr. 1, S. 1–11. <http://dx.doi.org/10.1002/asjc.449>. – DOI 10.1002/asjc.449. – ISSN 15618625
- [15] J. WRIGHT, H. LOOSEMORE AND R. FARMANI: Optimization of building thermal design and control by multi-criterion genetic algorithm. In: *Energy and Buildings* 34 (2002), Nr. 9, S. 959–972. [http://dx.doi.org/10.1016/s0378-7788\(02\)00071-3](http://dx.doi.org/10.1016/s0378-7788(02)00071-3). – DOI 10.1016/s0378-7788(02)00071-3
- [16] AHMAD, Muhammad W. ; MOURSHED, Monjur ; YUCE, Baris ; REZGUI, Yacine: Computational intelligence techniques for HVAC systems: A review. In: *Building Simulation* 9 (2016), Nr. 4, S. 359–398. <http://dx.doi.org/10.1007/s12273-016-0285-4>. – DOI 10.1007/s12273-016-0285-4. – ISSN 1996-3599
- [17] MOCANU, Elena ; MOCANU, Decebal C. ; NGUYEN, Phuong H. ; LIOTTA, Antonio ; WEBBER, Michael E. ; GIBESCU, Madeleine ; SLOOTWEG, J. G.: On-Line Building Energy Optimization Using Deep Reinforcement Learning. In: *IEEE Transactions on Smart Grid* 10 (2019), Nr. 4, S. 3698–3708. <http://dx.doi.org/10.1109/TSG.2018.2834219>. – DOI 10.1109/TSG.2018.2834219. – ISSN 1949-3053
- [18] WEI, Tianshu ; WANG, Yanzhi ; ZHU, Qi: Deep Reinforcement Learning for Building HVAC Control. In: UNKNOWN (Hrsg.): *2017 54th ACM*. Piscataway, NJ : IEEE, 2017. – ISBN 9781450349277, S. 1–6
- [19] MATALLANAS, E. ; CASTILLO-CAGIGAL, M. ; GUTIÉRREZ, A. ; MONASTERIO-HUELIN, F. ; CAAMAÑO-MARTÍN, E. ; MASA, D. ; JIMÉNEZ-LEUBE, J.: Neural network controller for Active Demand-Side Management with PV energy in the residential sector. In: *Applied*

BIBLIOGRAPHY

Energy 91 (2012), Nr. 1, S. 90–97. <http://dx.doi.org/10.1016/j.apenergy.2011.09.004>, – DOI 10.1016/j.apenergy.2011.09.004



Master's in Industrial Engineering

Master's thesis

**Development of a Control Algorithm Based on Machine
Learning for PV-Driven Heat Pumps and Compression Chillers**

Author

Ana Elechiguerra Batlle

Supervised by

Prof. Dr.-Ing. Thomas Sattelmayer

Dr.-Ing. Markus Spinnler

Bernd Heithorst, M.Sc.

Madrid

June 2020

Abstract

As the European energy supply sector moves towards more sustainable energy systems, progressively replacing conventional generation by renewable energy plants, other sectors can benefit from the emissions-free electrical power generated. The coupling of the heating and power sectors is particularly interesting, as the building sector has a high decarbonization potential, and the prices of both storing heat and generating heat from electricity are low.

The purpose of this thesis is to contribute to the transition to a more sustainable building sector by developing an intelligent control that manages and optimizes the sector-coupling strategy of a building energy system. The system includes a cold vapor cycle, driven by a photovoltaic (PV) generation unit and connected to the grid, and heat and cold storage.

After conducting a literature review on machine learning control strategies for building energy systems, the deep reinforcement algorithm is identified as a promising method to deal with the problem. The control strategy developed uses weather and demand predictions and aims at minimizing energy costs. The control is verified using short-term simulations. The performance of the control is evaluated by comparing it to two additional supervisory controls developed on previous work within the same project: a rule-based control and a model-predictive control (MPC). The results of annual simulations show that the lowest annual electricity cost is achieved by the MPC control and the highest self-consumption ration and lowest annual operation emissions are achieved by the rule-based control.

Additionally, a parameter study is conducted to identify the economically optimal combination of system parameters, considering the case of a single-family household in Munich, Germany.

Key Words: Building energy system, heat pump, photovoltaic, thermal energy storage, supervisory control, machine learning, deep reinforcement learning control, economic study, comparison of control approaches.

Contents

Abstract	XIX
Contents	XXV
List of Figures	XXIX
List of Tables	XXXI
Nomenclature	XXXII
1 Introduction	1
1.1 Background	1
1.2 Aims of the Study	1
1.3 Outline	2
2 Fundamentals	4
2.1 Cold Vapor Cycle	4
2.2 Machine Learning	6
2.2.1 Supervised Learning	7
2.2.2 Reinforcement Learning	11
2.2.3 Evolutionary Learning	13

2.3 Control Theory	15
2.3.1 Basics	15
2.3.2 Conventional Controls	16
2.3.3 Intelligent Controls	17
2.4 Conclusion of the Fundamentals Chapter	18
3 State of the Art	20
3.1 PVCool Project	20
3.1.1 Hardware	21
3.1.2 Software	22
3.1.3 Existing Simulink Model	23
3.2 Literature Review on Control Strategies	24
3.2.1 Literature on Reinforcement Learning	24
3.2.2 Literature on Artificial Neural Networks	27
3.2.3 Literature on Evolutionary Algorithms	29
3.2.4 Literature Review on Hybrid Control Strategies	30
3.3 Conclusion of the State of the Art Chapter	33
4 Decision on the Supervisory Control Approach based on Machine Learning	36
4.1 Structure and Goals of the Supervisory Control	36
4.2 Decision for Control Strategy Approaches	37
4.3 Conclusion	39
5 Development of the deep reinforcement learning Control Strategy	41
5.1 General Architecture of the Supervisory Control	41

CONTENTS

5.1.1 Supervisory Control Architecture	42
5.1.2 Architecture and Elements of the DRL Control	44
5.2 Design of the DRL Supervisory Control	45
5.2.1 The Environment	45
5.2.2 The Observations	51
5.2.3 The Reward	51
5.2.4 Design of the DRL Agent	52
5.3 Conclusion	56
6 Verification of the Control Strategies with Short Term Simulations	58
6.1 Simulation Data	58
6.1.1 Simulation Inputs	58
6.1.2 Simulation Outputs	59
6.2 Verification of the DRL Control	60
6.2.1 Verification of the Control Bounds	60
6.2.2 Verification of the Control Objectives	62
6.2.3 Determination of the Number of Episodes	64
6.3 Verification of the Supervisory Control System	65
6.4 Conclusion	67
7 Parameter Study and Economic Analysis	69
7.1 Comparative Values	69
7.1.1 Levelized Cost of Electricity	70
7.1.2 Levelized Cost of Heat	70
7.1.3 Levelized Cost of Energy	71

7.1.4 Self-Consumption Ratio	71
7.1.5 CO ₂ Emissions	71
7.2 Study Parameters and Inputs	72
7.2.1 Simulation Data	72
7.2.2 Model Parameters	73
7.2.3 Economic Parameters	75
7.3 Identification of an Economically Optimal System	76
7.4 Conclusion	80
8 Evaluation of the Control Approach	82
8.1 Results of the DRL Long-Term Simulations	82
8.1.1 Monthly Cost	82
8.1.2 Monthly SCR	83
8.1.3 Monthly Average COP	84
8.1.4 Monthly Average SOC	84
8.2 Comparison with Previous Work	85
8.2.1 Comparison of the Monthly Results	86
8.2.2 Comparison of the Annual Results	88
8.3 Conclusion	89
9 Conclusion and Outlook	92
9.1 Summary of the Thesis	92
9.2 Outlook	93
Bibliography	105

A March Examples of the Verification Process	106
B Training Patters of the Verification Process	108
C Simulation Inputs	111
D Post-Processing of Parameter Study Simulations	114
E Further Results of the DRL Long-term Simulations	115
F Further Results of the Control Comparison	117
G Alignment of the Project with the Sustainable Development Goals	119

List of Figures

2.1 Simplified scheme of a cold vapor cycle	5
2.2 p-h diagram of a simplified cold vapor cycle	5
2.3 Simplified scheme of a neuron	9
2.4 Basic components of an artificial neural network	10
2.5 Schematic structure of a multi-layer feed-forward network	10
2.6 Schematic behavior of a reinforcement learning algorithm	11
2.7 Example of the effect of the mutation operator in a binary vector	14
2.8 Example of the effect of the crossover operator in a binary vector	14
2.9 Flowchart of a genetic algorithm	15
2.10 Flowchart of a closed-loop control system	16
3.1 Scheme of the studied building energy system	21
3.2 Scheme of the PVCool test rig	22
3.3 Snapshot of the Simulink software of the test rig	23
4.1 Scheme of the studied supervisory control	37
5.1 Communication scheme between deep reinforcement learning control, rule-based control, and Matlab script	42

LIST OF FIGURES

5.2 Matlab synchronizing script flowchart	43
5.3 Simulink model of the DRL control	44
5.4 Environment of the DRL control	46
5.5 Simulink block that calculates the amount of energy bought and sold to the grid	48
5.6 Simulink source and sink block	50
5.7 Energy values proposed by the DDPG and DQN agents for the 1st of November 2008	53
5.8 Representation of the critic of the DRL Algorithm	54
5.9 Flowchart of the training process of the DRL Algorithm	55
6.1 Bound compliance of the desired energies of the heat pump	61
6.2 Bound compliance of the state of charge of the thermal storage	62
6.3 Overview of the building thermal demand and the thermal power provided by the heat pump during the period 2nd to the 5th of November 2008	63
6.4 Overview of the available power, the heat pump power and the resulting bought and sold powers during three exemplary november days	63
6.5 Episode rewards for the 11th of November 2008	64
6.6 Number of episodes as a function of the relative increase of the energy cost	65
6.7 Comparison of the desired hourly energy for the heat pump and the energy actually applied by the supervisory control unit between 5 am and 5 pm of the 2nd of November 2008	66
6.8 Compliance of the supervisory control unit to heat pump power control bounds	67
7.1 Approximation of heat pump costs	73
7.2 Approximation of thermal storage costs	74
7.3 Cost structure of a PV system under 30kWp	75
7.4 LCOEnergy for a 3kW heat pump and different PV and storage sizes	78

7.5 LCOEnergy for different PV and storage sizes - the influence of the heat pump size	79
8.1 Monthly energy cost obtained when controlling the heat pump with the DRL control	83
8.2 Monthly SCR achieved when controlling the heat pump with the DRL control	83
8.3 Monthly average COP obtained when controlling the heat pump with the DRL control	84
8.4 Monthly average SOC of the thermal storage achieved when controlling the heat pump with the DRL control	85
8.5 Comparison of the monthly electricity cost when controlling the heat pump with the DRL, MPC and rule-based controls	86
8.6 Comparison of the monthly average COP when controlling the heat pump with the DRL, MPC and rule-based controls	87
8.7 Comparison of the monthly SCR when controlling the heat pump with the DRL, MPC and rule-based controls	87
8.8 Comparison of the monthly average SOC when controlling the heat pump with the DRL, MPC and rule-based controls	88
A.1 Bound compliance of the state of charge of the thermal storage	106
A.2 Overview of the building thermal demand and the thermal power provided by the heat pump during the period 2nd to the 5th of November 2008	107
A.3 Overview of the available power, the heat pump power and the resulting bought and sold powers during three exemplary november days	107
C.1 Annual electrical demand	111
C.2 Annual PV production	112
C.3 PV power for the heatpump : excess PV power after electrical demand coverage	112
C.4 Annual sum of thermal and hot water demand	113
C.5 Average ambient temperature in the center of Munich	113

LIST OF FIGURES

D.1 Post-processing spreadsheet of the parameter study	114
E.1 Monthly heat pump energy obtained with the DRL control	115
E.2 Monthly heat pump operation hours obtained with the DRL control	115
E.3 Monthly emissions produced with the DRL control	116
F.1 Comparison of the monthly heat pump energy when controlling the heat pump with the DRL, MPC and rule-based controls	117
F.2 Comparison of the monthly heat pump operation hours when controlling the heat pump with the DRL, MPC and rule-based controls	117
F.3 Comparison of the monthly emissions when controlling the heat pump with the DRL, MPC and rule-based controls	118

List of Tables

2.1 Overview of advantages and disadvantages of the considered control strategies	18
3.1 Summary of the literature review on reinforcement learning as control approach for HVAC, heat pumps and similar energy systems	24
3.2 Summary of the literature review on artificial neural networks as control approach for HVAC, heat pumps and similar energy systems	27
3.3 Summary of the literature review on evolutionary algorithms as control approach for HVAC, heat pumps and similar energy systems	29
3.4 Summary of the literature review on hybrid machine learning approaches to control HVAC, heat pumps and similar energy systems	31
4.1 Overview of advantages and disadvantages of the considered hybrid control strategies	39
7.1 Costs and sizes of the selected heat pumps for the parameter study	74
7.2 Costs and sizes of the selected thermal storages for the parameter study	74
7.3 Costs and sizes of the selected thermal storages for the parameter study	76
7.4 Summary of results of the parameter study	77
8.1 Overview of the results of the annual simulations with the three investigated controls	89

LIST OF TABLES

B.1 Training patterns of the month of March obtained during the DRL control verification	109
B.2 Training patterns of the month of November obtained during the DRL control verification	110

Nomenclature

Latin Letters

Symbol	Unit	Meaning
A		Action
a		Action
c	$\left[\frac{\text{J}}{\text{kgK}}\right]$	Heat capacity
E		Experience
E		Expectation
E	[kWh]	Electric energy
e		Error
d		Disturbance
h		Specific time
I		Investment
i		Real investment rate
m	[kg]	Mass
\dot{m}	$\left[\frac{\text{kg}}{\text{s}}\right]$	Mass flow
N		Upper limit
n	[years]	Economic life-time
P		Performance
P	[W]	Power
p	$\left[\frac{\text{€}}{\text{kWh}}\right]$	Electricity price
\dot{Q}	[W]	Heat flow
Q		Action-value function
Q	[kWh]	Heat production

Nomenclature

q	$\left[\frac{\text{J}}{\text{kg}}\right]$	Specific heat
r		Reward
R		Total reward
S		State
s		State
T		Task
T		Target vector
T	[K]	Temperature
t		Time step
u		Control signal
V		State-value function
w	$\left[\frac{\text{J}}{\text{kg}}\right]$	Specific work
X		Input vector
y		Output
y_r		Reference

Greek Letters

Symbol	Unit	Meaning
α	[-]	Learning rate
γ	[-]	Discount factor
η	[-]	Efficiency
π	[-]	Policy

Superscripts

Symbol	Meaning
*	Optimal
*	New
k	Step

π Policy

Subscripts

Symbol	Meaning
0	Reference condition
π	Policy
<i>bought</i>	Bought from the electricity grid
<i>bui</i>	Building
<i>buy</i>	Buy from the electricity grid
<i>comp</i>	Compressor
<i>des</i>	Desired
<i>el</i>	Electrical
<i>end</i>	End
<i>HP</i>	Heat pump
<i>i</i>	Counter variable
<i>in</i>	Input
<i>int, learn</i>	Learning interval
<i>k</i>	Step
<i>max</i>	Maximum
<i>mech</i>	Mechanical
<i>min</i>	Minimum
<i>out</i>	Output
<i>sell</i>	Sell to the electricity grid
<i>sim</i>	Simulation
<i>sink</i>	Heat sink
<i>source</i>	Heat source
<i>sold</i>	Sold to the electricity grid
<i>storage</i>	Thermal energy storage
<i>t</i>	Time step
<i>tot</i>	Total
<i>w</i>	Water

Abbreviations

Abbreviation	Meaning
ADSM	Active demand side management
AggCLC	Closed-loop control aggregate
AggIL	Input layer aggregate
AggMC	Main control aggregate
AggOL	Output layer aggregate
AggSupCtrl	Supervisory control aggregate
AggTrig	Trigger aggregate
ANN	Artificial neural network
BNMI	Best network after multiple iterations
BPIE	Building Performance Institute Europe
BST	Building simulation tool
CVC	Cold vapor cycle
COP	Coefficient of performance
DDPG	Deep deterministic policy gradient
DHW	Domestic hot water
DPG	Deep policy gradient
DQN	Deep Q-network
DR	Demand response
DRL	Deep reinforcement learning
EER	Energy efficiency ratio
EPBD	European performance for buildings directive
GA	Genetic algorithm
HiL	Hardware in the loop
HP	Heat pump
HVAC	Heating, ventilation and air-conditioning
LCOE	Levelized cost of electricity
LCOEnergy	Levelized cost of energy
LCOH	Levelized cost of heat
MDP	Markov decision process
MLFFN	Multi-layer feed-forward network
MOABC	Multi-objective artificial bee colony

MOGA	Multi-objective genetic algorithm
MPC	Model predictive control
NPV	Net present value
NSGA	Nondominated sorting genetic algorithm
PD	Proportional derivative
PID	Proportional integral derivative
PMV	Predicted mean vote
PPD	Percentage of people dissatisfied
PV	Photovoltaic
QEERI	Qatar Environment and Energy Research Institute
RBC	Rule-based control
SCR	Self-consumption ratio
SOC	State of charge
TUM	Technical University of Munich

1 Introduction

1.1 Background

"The greatest threat to our planet is the belief that someone else will save it." Robert Swan [1]

The building sector in the European Union accounts for 40% of the total energy consumption and produces 36% of greenhouse emissions, making it the largest energy consumer [2]. The European Commission has acknowledged that it is critical to improve the energy performance of the building sector, setting targets for its decarbonization. The European Performance for Buildings Directive (EPBD), approved in 2002 and revised in 2010, sets the path to strict requirements for energy efficiency in buildings, and more particularly for heating, ventilation, and air-conditioning (HVAC) systems [3].

One possibility for the decarbonization of heating and cooling in the building sector is the use of renewable energy sources and the linking of power and heat sectors, providing flexibility to the system and allowing the use of the surplus power [4]. Among the different sector coupling strategies, the coupling of heat and power (commonly known as power-to-heat) is seen as one of the most attractive combinations, since both storing heat and generating heat from electricity are inexpensive processes [5]. Heat pumps and compression chillers are two of the most common examples of power-to-heat technologies, which turn electricity into heat and cold respectively [6].

1.2 Aims of the Study

The Building Performance Institute Europe (BPIE) highlighted the importance of the implementation of intelligent energy management systems in buildings [3]. Smart and efficient energy management plays an essential role in decarbonization and sustainability, shaping the future generation of buildings. A survey carried out by Shaikh et al. [7] shows the growing interest of both developed and developing countries in intelligent building energy controls. The main objective of this thesis is to implement these findings to the PVCool Project. The PVCool Project is a cooperation project of the Technical University of Munich (TUM) and the and Qatar Environment and Energy Research Institute (QEERI) of the Hamad bin Khalifa

University in Doha in Qatar. The project investigates the optimized setting of the coupling of a photovoltaic installation and a cold vapor cycle that can work as a heat pump or a compression chiller. The project studies the interaction of the different elements of the system, as well as its control [8].

This thesis aims at the following objectives:

- A literature review on machine-learning-based control strategies for a building energy system is conducted. Some of the techniques reviewed include controls based on reinforcement learning, artificial neural networks (ANNs), and evolutionary algorithms. The control approach is chosen based on this information and the state of the project.
- The chosen supervisory control strategy is developed and implemented into the existing Simulink model of the PVCool project.
- The functionality of the control is verified using short-term simulations.
- The economically optimal system is determined by performing a parameter study. In the parameter study multiple scenarios, all with a different combination of parameters, are simulated. An economical analysis of the different scenarios allows choosing the optimal combination of parameters.
- The three control approaches developed within the PVCool project are evaluated and compared. These approaches include the rule-based control developed by Achatz [9], the Model Predictive Control (MPC) developed by Rauscher [8], and the machine learning-based control developed in this master thesis.

1.3 Outline

This section gives a short structure for a better overview of the thesis. The second chapter sets the theoretical basis for the understanding of the thesis. It introduces the fundamental concepts and physical correlations of cold vapor cycles (CVCs), the basics of machine learning, and the principles of control theory. The third chapter reviews the state of the art of the PVCool project, as well as the main machine learning techniques used in building energy control. Chapter 4 introduces the structure and goals of the implemented supervisory control and justifies the chosen machine learning approach. The design process and implementation of the supervisory control are detailed in Chapter 5. The sixth chapter discusses the verification process of the chosen approach. In Chapter 7, an economic analysis is conducted, and the economically optimal system is identified. Chapter 8 evaluates and compares the different control approaches developed in the context of the PVCool Project. These include the rule-based control proposed by Achatz [9], the MPC approach developed by Rauscher [8] and the approach presented hereafter. Finally, Chapter 9 concludes and summarizes the thesis.

2 Fundamentals

This chapter presents the theoretical concepts that are necessary for this thesis. The first section will examine the basics of cold vapor cycles. The second section presents the fundamental notions of machine learning and the main concepts of the algorithms that were considered in this thesis. Finally, the last section introduces the fundamentals of control theory.

2.1 Cold Vapor Cycle

Cold vapor cycles are counterclockwise thermodynamic cycles [10] applied in many industrial and residential heating and cooling processes [11]. Two important implementations of cold vapor cycles are heat pumps, and compression chillers or refrigerators. Both systems extract heat from a low-temperature source (Q_{source}) and release heat to a high-temperature sink (Q_{sink}) continuously. The main difference between heat pumps and compression chillers is their purpose: a heat pump is meant to supply heat to the high-temperature sink and a compression chiller aims at the removal of heat of the low-temperature source [12].

Cold vapor cycles typically include the following elements: evaporator, compressor, condenser and expansion valve. The heat transfer fluid is called a refrigerant [13]. Figure 2.1 presents a simplified scheme of a cold vapor cycle. In stage 1, a compressor takes in the evaporated vapor and ejects it in stage 2 at higher temperature and pressure. To do this, the compressor needs electrical energy. The increase in pressure leads to a higher condensation temperature. From stage 2 to 3 the refrigerant enters a condenser in which heat is transferred from the refrigerant to the high-temperature sink. There, the refrigerant changes its state from vapor to liquid. Then, from stage 3 to 4, the refrigerant experiences a pressure reduction in the expansion valve that also decreases its evaporation temperature. Finally, the refrigerant returns to stage 1 after going through the evaporator, in which the refrigerant takes heat from the low-temperature sink and evaporates [13].

Figure 2.2 shows the p-h diagram of a cold vapor cycle. The specific heat rejected in the condenser q_{41} , the specific heat absorbed in the evaporator q_{23} and the specific work of the compressor w_{12} are marked in red.

The coefficient of performance (COP) and the energy efficiency ratio (EER) are two measures

2.1 Cold Vapor Cycle

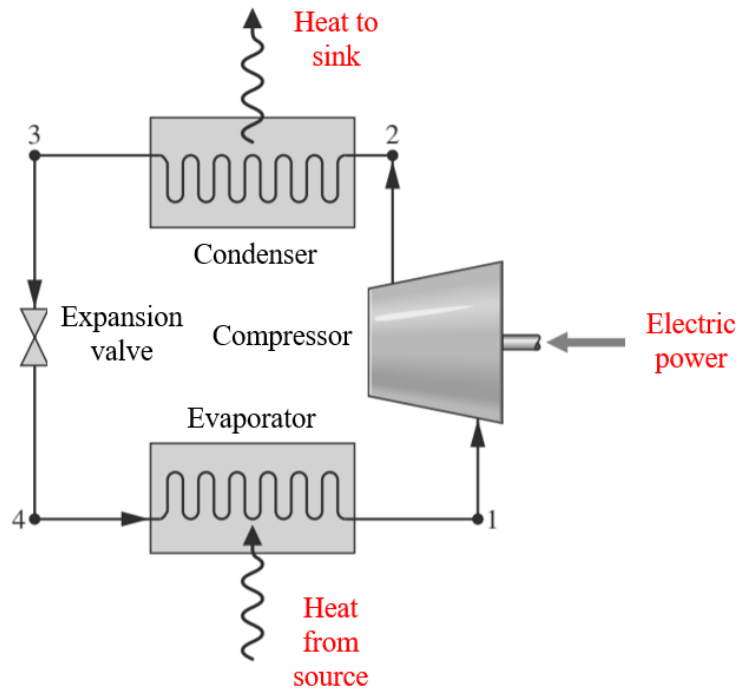


Figure 2.1 – Simplified scheme of a cold vapor cycle, based on [11]

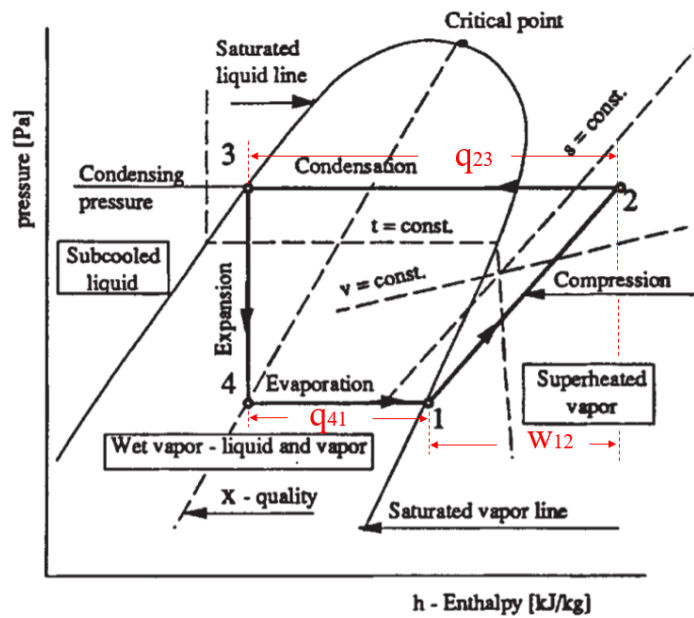


Figure 2.2 – p-h diagram of a simplified cold vapor cycle, based on [10]

of efficiency for cold vapor cycles. To calculate them, the three following variables are needed:

- The input power P_{in} calculated as show in equation (2.1) with the specific work of the compressor w_{12} , the mass flow of the refrigerant \dot{m} , and the mechanical and electrical efficiencies η_{mech} and η_{el} respectively [8]

$$P_{in} = \frac{1}{\eta_{mech}\eta_{el}} \dot{m}w_{12} \quad (2.1)$$

- The heat flow absorbed from the low-temperature source \dot{Q}_{source} is obtained as follows [14]:

$$\dot{Q}_{source} = \dot{m}q_{41} \quad (2.2)$$

- The heat flow transfered to the high-temperature sink \dot{Q}_{sink} is calculated as shown in the following equation [14]:

$$\dot{Q}_{sink} = \dot{m}q_{23} \quad (2.3)$$

In a cold vapor cycle, the efficiency is calculated as the ratio of the useful energy and the work needed to obtain it. In the case of the heat pump, as the purpose of the device is to provide the heat flow at the heat sink \dot{Q}_{sink} , the efficiency is calculated as the ratio of \dot{Q}_{sink} and the compressor power P_{in} as shown in equation (2.4). This relation is called the coefficient of performance. In the case of the compression chiller, as the objective of the device is to extract the heat flow of the heat source \dot{Q}_{source} , the efficiency is calculated as the ratio of \dot{Q}_{source} and the input power P_{in} as shown in equation (2.5) [14]. This efficiency is called the energy efficiency ratio.

$$COP = \frac{\dot{Q}_{sink}}{P_{in}} \quad (2.4)$$

$$EER = \frac{\dot{Q}_{source}}{P_{in}} \quad (2.5)$$

2.2 Machine Learning

Energy consumption optimization entails different challenges such as the constant trade-off between comfort, consumption, and cost [15] or the development of models, subject to non-linearities [3] and inaccuracies [16]. In this context, research has shown that intelligent control

techniques and more specifically those with learning capabilities can overcome these challenges and be successfully implemented in the field of HVAC [17]. As a supervisory control based on machine learning approaches is developed in this thesis, it is reasonable to briefly introduce the term, as well as the different approaches and algorithms.

Mitchell [18] proposed the following definition of machine learning: *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ."*

In other words, machine learning techniques are a subdivision of artificial intelligence that can predict, learn and find patterns from raw data [19].

The categorization of the machine learning approaches is not unique (e.g.: Marsland [20], Oladipupo [21], and Moubayed [22] propose different subdivisions). However, in this thesis, the approach of Marsland [20] is chosen because it includes evolutionary learning and its optimization algorithms. In Marsland [20], the categories of machine learning are established based on an algorithm perspective:

- **Supervised learning:** A training set containing examples with both inputs and outputs is available and the algorithm learns to generalize and map inputs to outputs.
- **Unsupervised learning:** Only the input data are provided, and the algorithm's goal is to find patterns and categories within the data.
- **Reinforcement learning:** the algorithm performs actions in an environment following a trial and error approach and receives a reward assessing the goodness of the action until the final policy is determined.
- **Evolutionary learning:** the algorithm uses the idea of biological evolution in order to find the optimal solution.

The most common approaches of supervised, reinforcement and evolutionary learning and their theoretical foundations will be presented hereafter. The application of the different algorithms in the context of control theory are introduced in greater detail in section 2.3.3 and Chapter 3.

2.2.1 Supervised Learning

In supervised learning problems, a set of training data (X_i, T_i) (with X_i being the input vector, T_i the target vector, and i going from 1 to a sufficiently high upper limit N) is available and provides the expected answer to each of the associated inputs. The goal of the supervised learning algorithm is to generalize. In other words, the algorithm should be able to find

reasonable outputs for inputs that are not included in the training data. Therefore, the supervised learning algorithm is able to deal with noisy data, or data containing inaccuracies [20]. However, the design of such algorithms is subject to a series of trade-offs described below [23]:

- **Bias-Variance dilemma:** Bias refers to the expected systematic error the learning algorithm makes when trained on a training set. Variance refers to the capture of random fluctuations in training sets [24]. Bias error comes from false assumptions made by the learning algorithm, and variance refers to the error that derives from sensitivity to small variations in the training data. A learning algorithm is said to have high bias (and low variance) when the algorithm omits important relations between inputs and outputs and hence generalizes poorly. This is commonly called underfitting. On the other hand, when an algorithm has high variance (and low bias), the algorithm adapts itself excessively to the training data, also modeling noise, and again performs a poor generalization. This is known as overfitting. A compromise must be found between bias and variance [25].
- **Function complexity and training data cost:** if the relationship the algorithm learns is simple, then the function can be approximated from a small training set, using a learning algorithm with low variance and high bias. If the function has a high degree of complexity, then a large amount of data and an algorithm with high variance and low bias would be most appropriate [26].
- **Input space dimensionality:** a large number of input features could confuse the learning algorithm, even if only a small number of those features is relevant. Here, the algorithm can fail to generalize appropriately, leading to high variance. Reduction of dimensionality, manually or via feature reduction algorithms, is therefore encouraged for large datasets [26].
- **Output value noise:** if the training output data contain errors or inaccuracies, the algorithm should avoid finding a function that maps inputs to outputs too carefully. Possible solutions to this problem include early stopping of the training process and the removing of noisy samples [26].

According to the "No Free Lunch Theorem" [27], there is no learning algorithm that is the best choice for all problems. There are many different algorithms to choose from, all with their advantages and disadvantages. The two most essential techniques in supervised learning are regression and classification techniques. Regression seeks to find relationships, forecast, and predict quantitative data. Classification methods analyze data and find patterns so as to predict a qualitative response. For these techniques, multiple algorithms can be used. Some of the most common examples include: support vector machines, decision trees, artificial neural networks, linear regression, logistic regression, and K-nearest [28]. ANNs are the most common supervised learning tool used in building energy control [3]. They are therefore explained in detail hereafter.

Artificial Neural Networks

The human brain is composed of more than 100 billion connected neurons [29]. These neurons, also called nerve cells, are formed from three main elements [30]: the soma, the axon and dendrites. The soma, or cell body, receives and processes incoming information sent by other neurons. The axon is the path through which information travels from neuron to neuron. The dendrites are branched extensions that surround the cell body and that can be seen as receiving terminals of the neuron.

Information travels through the nervous system by employing chemical transmitters in the synapse, the union of an axon with the dendrite of a different neuron. Incoming nerve impulses free the chemical transmitters [30]. Figure 2.3 illustrates the main parts of a neuron.

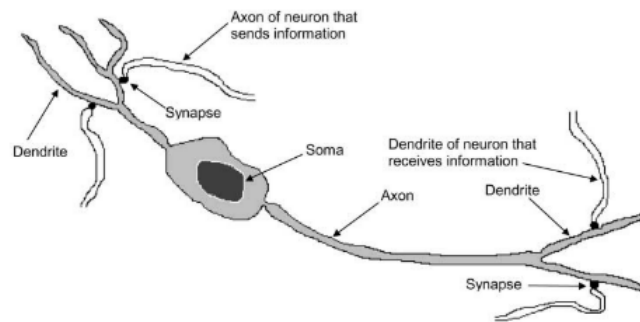


Figure 2.3 – Simplified scheme of a neuron [29]

Artificial neural networks are inspired by biological neurons. Haykin [31] defines an artificial neural network as a processor which is "massively parallel distributed", and can store and provide knowledge coming from experience. ANNs mirror the human brain in two ways. First, they collect knowledge following a learning process. Secondly, they store this knowledge in synaptic weights or interneuron connections [29]. Moreover, the learning process of artificial neural networks resembles the learning process of the brain inasmuch as both work as "black box" models and do not need specific knowledge about the system as a whole. The relationship between inputs and outputs can be learned though already recorded data [29].

The main components of a neuron in an artificial neural network are represented in figure 2.4. These are the following [30]: synapsis links, an adder and a nonlinear activation function. Each synapsis link is defined by a weight. Through these links, the weighed information of the different inputs arrives at the node. The adder sums all the weighed input signals. The result of the weighed addition is evaluated by the activation or transfer function of the node. The result is the output of the neuron.

In order to store and provide experimental knowledge, the ANNs must first undergo a training process. Once inputs and outputs are defined in an application, accurate data need to be gathered. Accuracy is crucial because erroneous data would lead to an incorrect model [33]. After

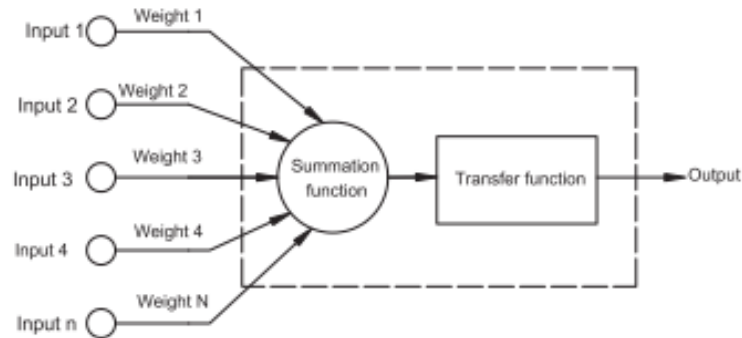


Figure 2.4 – Basic components of an artificial neural network [32]

the data are prepared, the appropriate architecture and the accompanying parameters (such as the learning rate, or the type of activation function of the ANN) must be chosen [29]. The most popular architecture of artificial neural networks is multi-layer feed-forward networks (MLFFN) [34]. An MLFFN has one input layer, one or multiple hidden layers, and one output layer. Figure 2.5 shows a simplified schematic of this architecture. Thanks to the presence of various layers with nonlinear activation functions, the ANN can approximate both linear and nonlinear functions [32]. Other common types of ANN architecture are single-layer feed-forward networks, radial basis networks, or recurrent neural networks [30].

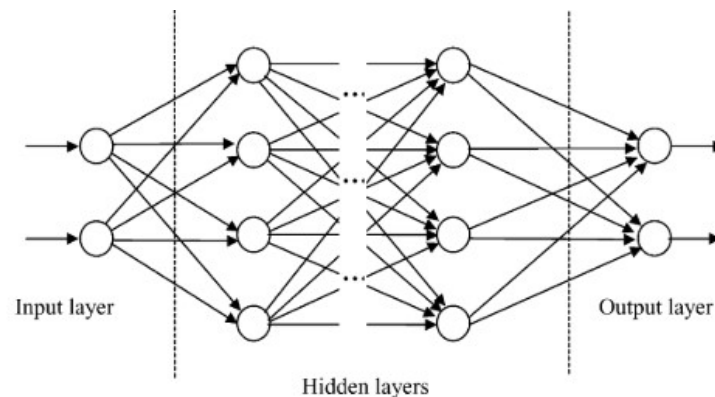


Figure 2.5 – Schematic structure of a multi-layer feed-forward network [29]

After the architecture and parameters of the network have been chosen, the network must be trained. The training consists of a systematic modification of synaptic weights [29]. The networks receive a set of training examples, each of them formed by the input and the expected output. The learning algorithm modifies the weights if the network's calculated solution differs from the expected output. The examples are presented to the network repeatedly until the error or the output attain acceptable levels [33]. After the training process, the network is validated using test data not previously employed in the training phase. This way, it becomes clear whether the developed ANN is well trained and performs well for previously unknown data. Finally, the network is optimized to enhance its overall performance [32].

In recent years, interest in artificial neural networks has substantially increased. Their ability to learn from examples, their robustness, and their error tolerance makes them useful tools for tasks like approximating functions, as well as finding, recalling or generating patterns. They have been applied to multiple different fields such as medicine, engineering, psychology, and economics [29].

2.2.2 Reinforcement Learning

A reinforcement learning algorithm tries to find the optimal sequence of actions needed to reach a long-term goal by repeatedly interacting with an environment [16]. An agent explores the environment by learning and interacting with it [35]. Figure 2.6 shows the flowchart of the learning process. The agent observes the current situation at a time t of the environment, described here as its state S_t , and then performs an action A_t . This action affects the environment, which changes its state from S_t to S_{t+1} . When this happens, the agent receives a reward, r_t , produced by the environment. This reward reflects the instant benefit of being in a given state. Cumulated rewards over a sequence of actions indicate whether or not the algorithm is properly progressing towards the long-term goal [16]. The goal of the reinforcement learning algorithm is to find a policy that maximizes long-term reward. A policy is a function that maps actions to states. Usually this function is described as a probability function where, given a state s , the probability of taking an action a is $\pi(s, a)$ [17].

The reinforcement learning problem can be seen as an optimization problem in which the concept of value is introduced. The value refers to the total expected reward from the current state projected into the future. By looking at the value, the agent can select actions that will maximize the long-term reward. The value function is also called "critic", since it looks at different action options and criticizes the agent's decisions [35].

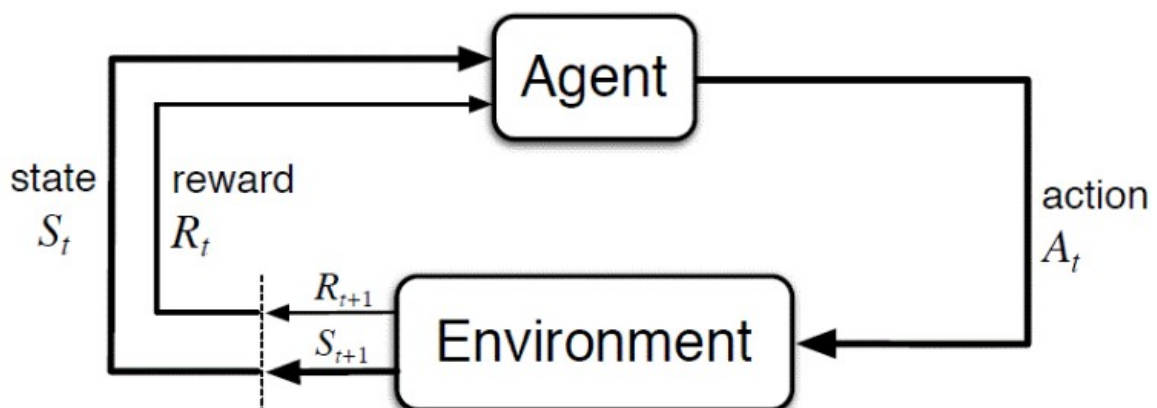


Figure 2.6 – Schematic behavior of a reinforcement learning algorithm [36]

In most cases, reinforcement learning problems can be described by a Markov Decision Process (MDP). A process is said to be Markovian when the future state only depends on the present, being independent of the past [36]. This property does not imply that the past events are irrelevant, but that the current state retains historical information. In an MDP the reward function depends on the current state, action, and the next state [16]. The total rewards, R_t , from the current state and onwards are shown in equation (2.6) [17]:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (2.6)$$

In equation (2.6), the discount factor γ , defined such that $0 \leq \gamma \leq 1$, allows the algorithm not to rely excessively on the promise of future rewards $\sum_{k=0}^{\infty} r_{t+k+1}$, discounting them to a greater extent the further they lay in the future. Given a state s and a policy π the state-value function $V^\pi(s)$ is defined as the expectation E_π to get a certain reward R_t given a state s_t [17]:

$$V^\pi(s) = E_\pi\{R_t \mid s_t = s\} = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\right\} \quad (2.7)$$

The action-value function $Q(s, a)$ that determines the value of choosing an action a in a state s following policy π is described in equation (2.8) [17]:

$$Q^\pi(s, a) = E_\pi\{R_t \mid s_t = s, a_t = a\} = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a\right\} \quad (2.8)$$

The state-value function $V^\pi(s)$ and the action-value function $Q^\pi(s, a)$ are measures of the expected return. One or the other is used depending on the type of problem. In a model-based application, the state-value function would most likely be applied, whereas in a model-free application, the action-value function would be preferred [36]. Because a model-free algorithm is used in this thesis, the focus from here on will be on the action-value function $Q^\pi(s, a)$, but both procedures are similar. As explained above, the reinforcement learning algorithm aims at identifying the optimal policy $\pi^*(s, a)$ (optimality being symbolized by an asterisk) that maximizes the expected long-term reward, or the value function [17]:

$$\pi^*(s, a) = \operatorname{argmax}_{a \in A_t} Q(s, a) \quad (2.9)$$

In order to optimize equation (2.9), the learning algorithm uses the recursive relationship between two successive actions chosen among an action set A_t . This algorithm, known as Q-learning algorithm, is considered one of the most relevant findings in reinforcement learning [16]. The Q-learning algorithm was first introduced by Watkins [37] and is still one of the most

popular and efficient model-free learning techniques [16]. The action-value function $Q(s, a)$ is updated to approximate the new value $Q^*(s, a)$ [17]:

$$Q^*(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \underbrace{[r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]}_{\text{error in Q estimate}} \quad (2.10)$$

In equation (2.10) the new value of the action-value function $Q^*(s_t, a_t)$ is calculated as the sum of the old value of $Q(s_t, a_t)$ and the error in Q estimate [35]. The error in Q estimate is weighted by the learning rate α ($0 \leq \alpha \leq 1$) and includes the following elements: the reward for taking a certain action r_{t+1} , the previously introduced discount factor γ ($0 \leq \gamma \leq 1$), the maximum expected value going forward $\max_{a_{t+1}} (Q(s_{t+1}, a_{t+1}))$, and the old value of $Q(s_t, a_t)$ [35].

Often, the algorithm stores the result of the action-value function $Q(s, a)$ in a lookup table, commonly called Q-table. The elements of the Q-table are the value estimations of each state-action pair [17]. However, this type of representation can result in practical problems when the number of state-action pairs increases [35]. This is called the "curse of dimensionality" in reinforcement learning [38]. Other types of policy representation are possible. An alternative to the Q-table would be to use a function to represent policies. However, that would require knowledge of the structure of such function beforehand. Another approach is to use a general function approximator such as an artificial neural network [35]. The combination of artificial neural networks and reinforcement learning is called Deep Reinforcement Learning (DRL). The Deep Q-Network (DQN) agent combines artificial neural networks and the Q-learning algorithm to approximate the Q-table [39]. The model-free DQN algorithm chooses among a discrete set of possible actions. Matlab also supports a model-free DRL algorithm that can choose among a continuous interval of actions: the Deep Deterministic Policy Gradient agent (DDPG) [40].

2.2.3 Evolutionary Learning

Evolutionary algorithms are metaheuristic- and population-based algorithms, inspired in natural evolution [41]. Natural evolution sets an "imaginary fitness landscape" [20] which favors the survival of living beings that are fitter. Fitter living beings are able to survive longer, find more mates and produce a larger number of healthier offsprings [20]. Similarly, evolutionary algorithms consider two key ingredients of natural evolution: superior selection and variational reproduction [42]. Although the implementations of evolutionary algorithms are varied, they usually undergo the following steps [42]:

1. Generation of an initial population or initial solution set

2. Generation of new solutions using the existing population (reproduction)
3. Removal of the worst individuals of the existing population
4. Repetition of this process from the second step until a certain stop condition is met

There are many different implementations of evolutionary algorithms [42], e.g: genetic algorithms (GAs), genetic programming, and evolutionary strategy. This section will detail the concepts of the genetic algorithm since it is the most popular algorithm [41] and it has been already applied to a broad range of building optimization problems [43].

Genetic Algorithms

Genetic algorithms are search algorithms inspired by Darwin’s natural selection [44]. A genetic algorithm starts with the generation of a set of potential solutions to a problem, or a population of solutions. Each solution is also called an individual. With each run of the optimization algorithm, a new generation (a new set of solutions) is created. This new set of individuals is generated by evolution operators [3]. The most common evolution operators are the mutation and crossover (also called recombination) [42]. The mutation operator randomly modifies one of the existing solutions to create a new one. Figure 2.7 shows an example of a mutation in a Boolean vector solution. In this example, one bit is randomly selected and its value is modified in the next generation [42].



Figure 2.7 – Example of the effect of the mutation operator in a binary vector, adapted from [42]

The crossover operator combines and mixes two or more solutions to generate new ones. Figure 2.8 depicts an example of a crossover in a Boolean vector solution. In this case, a position in the vectors of the two parents is selected. The new generation is created by flipping the elements of both parents after this position [42].

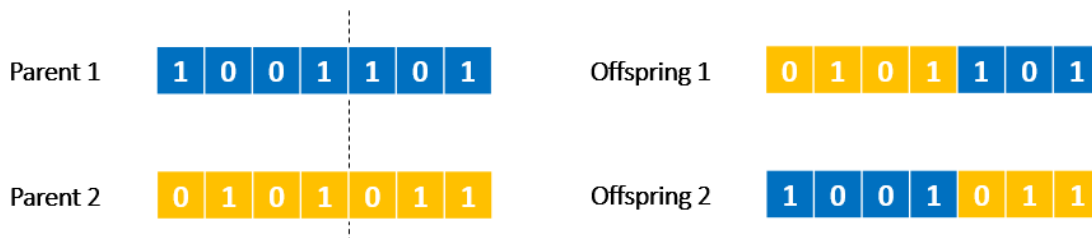


Figure 2.8 – Example of the effect of the crossover operator in a binary vector, adapted from [42]

A fitness function evaluates the goodness of the offsprings and parent solutions at each iteration and selects the best solutions. The iterations continue until a stopping criterion is met. Common examples of stopping criteria are:

2.3 Control Theory

- The surpass of the maximum computational time
- The meeting of a predefined quality of the solution
- The lack of improvement of the solutions.

Figure 2.9 describes the flowchart of a genetic algorithm.

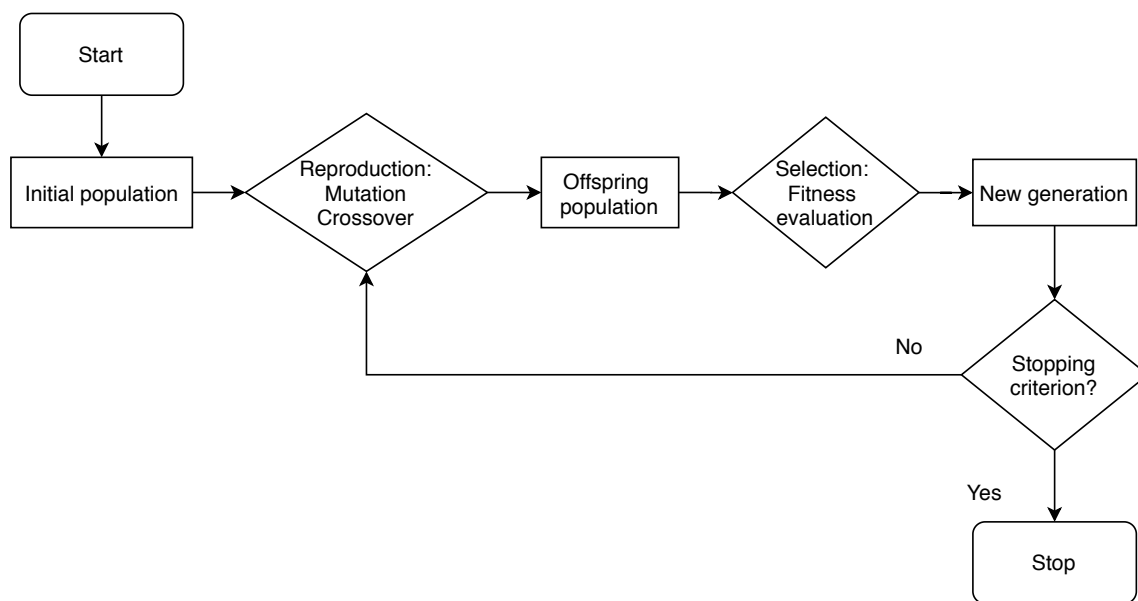


Figure 2.9 – Flowchart of a genetic algorithm

2.3 Control Theory

This section briefly introduces the basics of control engineering and traditional controls. Then, intelligent controls are presented. This thesis focuses on learning-based controls, a subset of intelligent controls.

2.3.1 Basics

Control engineers see a process as a dynamic system which has outputs, inputs, and internal variables that are called state variables. Systems can be described as open or closed loops. In the case of open-loops, unlike for close-loops, the output value is not used to rectify the

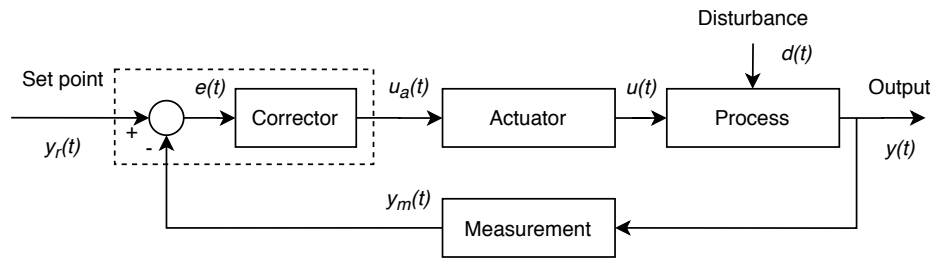


Figure 2.10 – Flowchart of a closed-loop control system [45]

error [45]. As in this thesis, a closed-loop approach is applied [8], the closed-loop process is described below.

Figure 2.10 describes the process of a closed-loop system. The output $y(t)$ tries to follow a certain desired reference $y_r(t)$ despite the unpredictable disturbances $d(t)$. To do so, the output $y(t)$ is measured and compared to the reference $y_r(t)$. The difference is the error ($e(t)$). After certain correction, this value is fed into the actuator which sends a control signal $u(t)$ to the process. The control signal $u(t)$ and the disturbances $d(t)$ influence the value of the new output of the process $y(t)$ [45].

Controls can also be classified into local and supervisory controls. Local controls are lower-level controls that ensure tracking of a reference and robust performance. Supervisory controls are higher-level controls that aim at the optimization of the process and control settings. The main focus of this thesis is the development of a supervisory control algorithm for the PVCool project [46].

2.3.2 Conventional Controls

Conventional controls mainly include on-off and proportional-integral-derivative (PID) controls. Both have been widely used in the field of building engineering [7]. The implementation of on-off controls is simple but subject to multiple drawbacks [36], such as overshooting or instability, leading to poor performance [7]. The implementation of PID controls is also simple. The proportional control (P) aims at reaching the reference as fast as possible, whereas the derivative control (D) prevents the control from adjusting too quickly or too slow. The integral control (I) aims at removing the residual error by regulating the aggregated error [36]. However, PID controls have no knowledge of the plant, have constant control parameters, and have limited performance against noise and non-linearities. Some PID architectures such as the cascade structure can enhance the performance of PID controls [7].

Traditionally the above-mentioned strategies only focus on energy consumption minimization and do not take into account thermal comfort or vice-versa. To solve this, and the other

beforementioned issues, researchers investigated other approaches in the 1980s. These approaches included predictive, adaptive and optimal controls [7].

2.3.3 Intelligent Controls

Multiple control approaches fall under the umbrella of intelligent controls, such as fuzzy logic or multi-agent systems [7]. Fuzzy systems are considered intelligent because of their use of human-like reasoning to create a set of control rules [36]. Multi-agent systems are coordinated, intelligent and autonomous agents or devices with local goals that work together to achieve complex tasks [3]. However, this thesis will only focus on those controls based on machine learning approaches, or those with learning capabilities. These machine learning approaches include artificial neural networks, evolutionary algorithms [3], reinforcement learning, and hybrid approaches. Below a brief overview of the implementation of these approaches in the field of building control is given. Additionally, the main advantages and disadvantages are mentioned.

- **Artificial Neural Networks:** In the literature, ANNs are mostly used because of their capacity to model non-linear problems and making predictions without knowing the system. However, ANNs rely heavily on the training data, and they are only reliable when the operation is within the training set range. The topology of the neural networks is also hard to determine and varies for each application [3].
- **Evolutionary algorithms:** Genetic and evolutionary algorithms are used as global optimization techniques in control [3]. Specifically, Genetic algorithms are the most acknowledged technique in building performance optimization [7]. They are easy to understand and configure and can handle complex problems and trade-offs [43]. However, they can present a slow convergence [47].
- **Reinforcement learning:** This model-free, data-driven control technique is also able to optimize building performance [36]. This technique is optimal for online applications [48] and can be adapted to changes in the environment during operation [49]. However, the implementation can be difficult because the performance of reinforcement learning algorithms depends on many factors [46], constraints are hard to handle [50] and the algorithm can suffer from the curse of dimensionality [51] as seen in section 2.2.2.
- **Hybrid techniques:** These techniques combine multiple intelligent approaches to achieve better performance and obtain better solutions. Each combination has its characteristics and should be discussed independently [3]. Examples of these combinations are ANNs with genetic algorithms, fuzzy logic with ANNs [3], or multi-agents and reinforcement learning [36].

2.4 Conclusion of the Fundamentals Chapter

In the first section, the basic principles of cold vapor cycles are introduced. Cold vapor cycles are the working principle of heat pumps and compression chillers. Then, the fundamental concepts of machine learning are presented. The principles of supervised learning, reinforcement learning, and evolutionary learning, as well as their most relevant algorithms, are introduced. Finally, the basics of control theory are described. The relevant intelligent control methods based on machine learning, as well as their advantages and disadvantages, are also listed. Table 2.1 summarizes the last main aspects.

Table 2.1 – Overview of advantages and disadvantages of the considered control strategies

Approach	Advantages	Disadvantages
Artificial Neural Networks	<ul style="list-style-type: none"> • effective real-time operation [3] • able to model complex relationships [34] • need no knowledge of the system [3] 	<ul style="list-style-type: none"> • only reliable on the range of the training data [3] • Topology hard to determine [3]
Evolutionary algorithms	<ul style="list-style-type: none"> • easy to understand and configure [43] • can handle complex problems and trade-offs [43] 	<ul style="list-style-type: none"> • slow convergence [47]
Reinforcement Learning	<ul style="list-style-type: none"> • model-free [3] • optimal for real-time operation [48] • adaptable during operation [49] 	<ul style="list-style-type: none"> • implementation depends on many factors [46] • difficult handling constraints [50] • curse of dimensionality [51]
Hybrid techniques	<ul style="list-style-type: none"> • better solutions [3] • better performance [3] 	<ul style="list-style-type: none"> • each combination is unique [3]

In the next chapter, the literature on machine-learning-based controls in the field of building energy optimization is introduced. Afterward, it is concluded which approach is the most suitable for the problem of this thesis.

3 State of the Art

This chapter describes the state of the art that envelopes the thesis. The supervisory control of this thesis is developed for the PVCool Test Rig. The PVCool Test Rig, as well as the work of previous student theses, are presented. The tests in the project can be performed in a physical test rig, or a hardware in the loop simulation (HiL Simulation). The physical test rig or hardware is the CVC. Section 3.1.1 details the physical test rig. Section 3.1.2 introduces the control software of the PVCool Test Rig. The Simulink simulation model of the CVC environment is presented in section 3.1.3.

Section 3.2 summarizes the state of the art in the field of controls based on machine learning techniques for building energy optimization. These publications use reinforcement learning, supervised learning, evolutionary learning, or hybrid approaches to manage energy systems that are similar to the PVCool project.

3.1 PVCool Project

Figure 3.1 represents the studied building energy system. The system consists of a PV plant and the electrical and thermal (domestic hot water and space heating) demand of a single-family house [52]. In the center of the scheme, the gray background highlights the cold vapor cycle. The CVC can work either as a heat pump or as a compressor chiller. When the device provides heat to the building, it is working as a heat pump. In the case of providing cold to the building, it works as a compression chiller. The figure shows only the operation of the heat pump. The system highlighted in gray represents the test-rig where experiments for the PV-Cool project are conducted. The test rig consists of an evaporator, a compressor, a condenser, and an expansion valve [8]. The design of the test-rig is described in the thesis of Böing [53]. The subordinate compressor control and hardware modifications are described in Bauer [54].

The thermal storage is shown on the right of the vapor cycle. It is connected to an additional ohmic heater. The latter heater can support the thermal storage in the coverage of the peak loads of winter days. On the left of the test rig, an air heat exchanger serves as a heat source [52]. Although the operation as a compression chiller is not represented in the picture, the process is analog: a cold water storage is used as source and an air heat exchanger serves as heat sink. In figure 3.1, red and blue lines represent hot and cold-water pipes, respectively.

3.1 PVCool Project

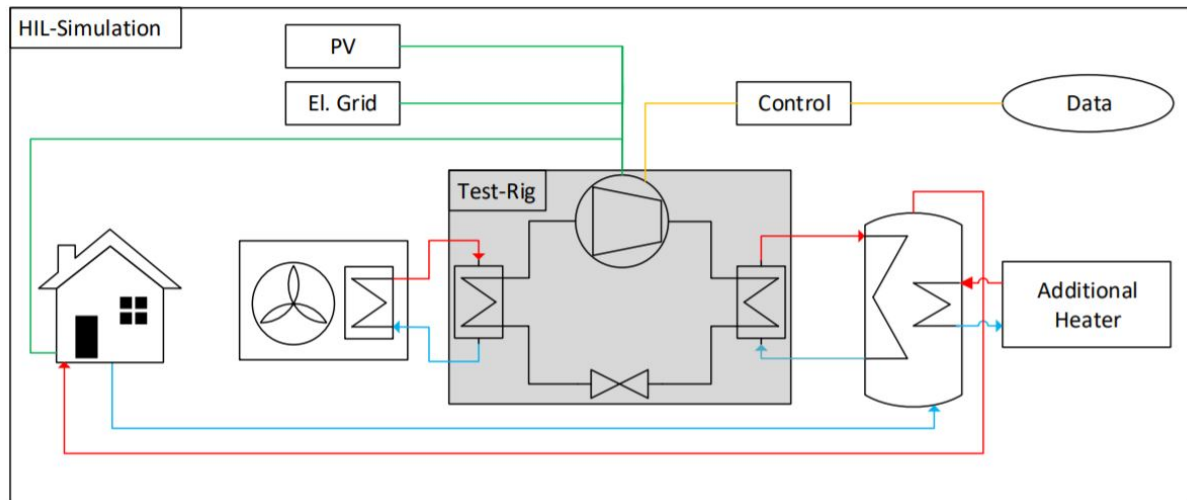


Figure 3.1 – Scheme of the studied building energy system [52]

Green and yellow lines represent electric and control flows, respectively. Electricity can be supplied by the grid or the PV panels and can be supplied to the vapor cycle or the building. Excess electricity is fed to the grid. The control unit collects data from the system and derives the desired power to be provided by the vapor cycle [8].

The HiL Simulation considers all the elements of the above-described system. Schmideder [55] developed the building energy system. The next sections focus on the hardware of the test rig, the software of the test rig, and the HiL simulation.

3.1.1 Hardware

Figure 3.2 is a scheme of the resulting test rig. In the thesis of Bauer [54] and Irrgang [56] the testing facilities were developed the testing facility further. The refrigerant evaporates by exchanging heat with the source. The fluid is then compressed, either by a swash-plate or a scroll compressor. Both compressors, driven by an AC motor, are installed in parallel so that the user can decide which one to use. In the next step, the refrigerant goes through a condenser. The fluid exchanges heat with the heat sink and condenses. Vapor and liquid phases are separated by a receiver. Finally, the refrigerant decreases its pressure in the expansion valve and returns to its original state [8].

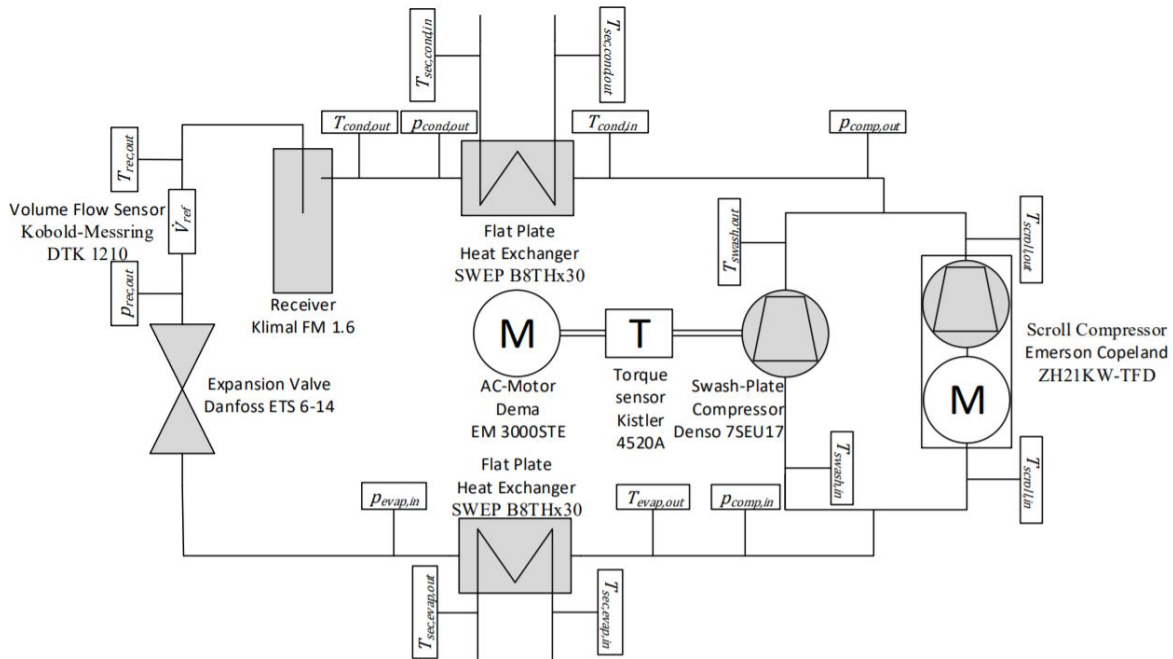


Figure 3.2 – Scheme of the PVCool test rig [52]

3.1.2 Software

Figure 3.3 is a snapshot of the software of the test rig, implemented Simulink. Multiple aggregates, each of them fulfilling different objectives, build the software [8]. The system is modified during the development of the thesis.

- The trigger aggregate (AggTrig) coordinates the execution of the aggregates. It schedules the calling of the different subsystems to allow real-time operation of the test rig software.
- The main control aggregate (AggMC) is the user interface of the testing facility. It defines the values of the parameters and the settings of the system. It allows the user to select the type of control to be tested. Its building blocks are taken from the SimulinkDashboard Library.
- The input layer aggregate (AggIL) prepares the data and operates calculations when needed. It allows communication between hardware and software. It receives, processes and prepares input data so that other aggregates can access it.
- The supervisory control aggregate (AggSupCtrl) includes the simulation models of the CVC and heating system and the supervisory controls that compute the desired power

3.1 PVCool Project

to be applied by the compressor. This aggregate already includes the rule-based supervisory control designed by Achatz [9] and the fuzzy and MPC supervisory controls developed by Rauscher [8]. The machine learning approach is added during this thesis.

- The closed-loop control aggregate (AggCLC) acts as a subordinate control. It transforms power values in signals of frequency for the compressor.
- The output layer aggregate (AggOL) saves the computed output values or sends them to the hardware.

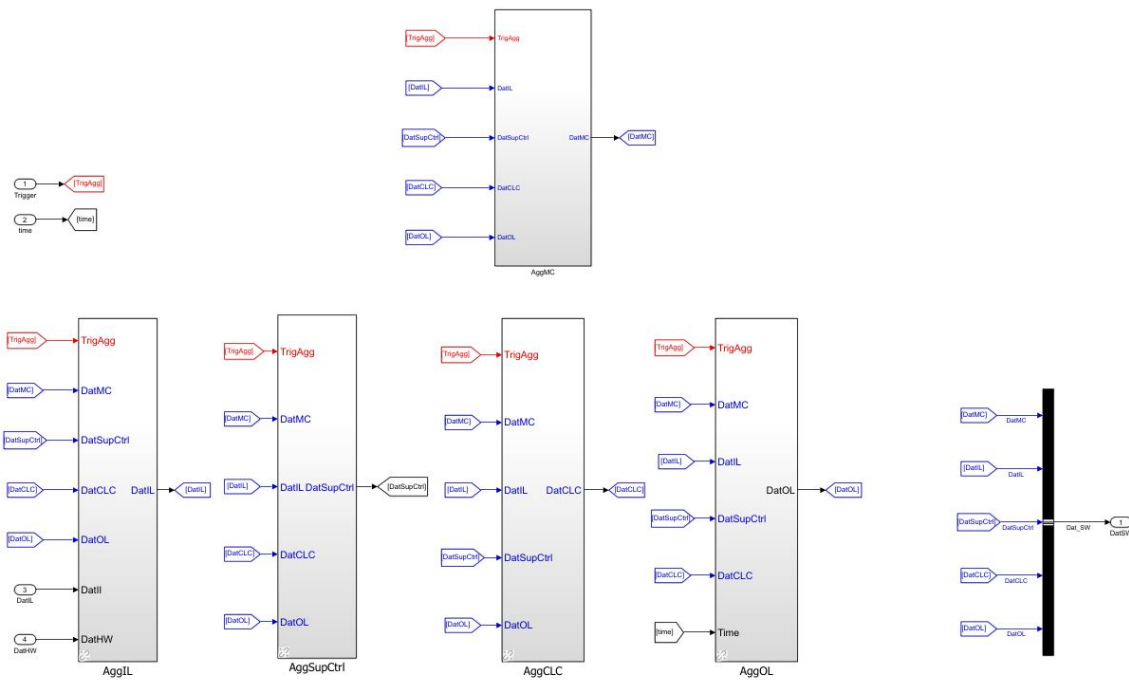


Figure 3.3 – Snapshot of the Simulink software of the test rig

3.1.3 Existing Simulink Model

Schmideder [55] designed the Simulink model of the building energy system that was developed further by Rauscher [8] and Achatz [9]. In this model, the PV plant and the electrical grid cover both the thermal and electrical demands of the building. The cold vapor cycle uses part of this electricity to cover the thermal demand, providing cold and heat to the building. The provision of cold or heat depends on the working model of the cold vapor cycle (heat pump or compression chiller). Cold and heat storage are also modeled. The building is equipped with an underfloor heating or cooling system that mixes the flows to meet the required temperature. A heating or cooling cycle distributes the heat or cold from the thermal storage to the building.

The Simulink model of the building energy system includes an artificial neural network to approximate the value of the COP. The inputs of the neural network are the heat sink temperature, the compressor power, and the heat source temperature. The model also contains elements of the Carnot library developed by Wemhöfer et al. [57] and the Solar Institute Jülich [58]. This library contains s-function blocks that model both traditional and renewable energy systems.

3.2 Literature Review on Control Strategies

This section offers a review of different control strategies for HVAC systems, PV-driven heat pumps, and similar energy systems. Sections 3.2.1, 3.2.2, and 3.2.3 present the literature on control strategies based on reinforcement learning, artificial neural networks, and metaheuristic algorithms, respectively. Section 3.2.4 summarizes publications where hybrid methods are applied. Most of the reviewed literature is presented in a tabular format from the most recent to the oldest. Relevant publications are detailed at the end of the sections.

3.2.1 Literature on Reinforcement Learning

Table 3.1 summarizes the reviewed publications in reinforcement learning control for energy systems that are similar to the PVCool project. The authors implement different types of algorithms. The broadly used Q-learning algorithm is often subjected to the "curse of dimensionality", influencing the computational cost.

Table 3.1 – Summary of the literature review on reinforcement learning as control approach for HVAC, heat pumps and similar energy systems

Author	Year	Publication
Peirelinck et al. [59]	2018	Using reinforcement learning for optimizing heat pump control in a building model in Modelica: the authors apply batch reinforcement learning, a method in which the policy is previously learned offline and used to derive the solution of an online application. More specifically, the authors use a neural fitted Q-iteration algorithm to a thermostatically controlled load in the context of demand response (DR) and flexibility exploitation. They study the arising property of partial observability. The learning algorithm is applied to a heat pump and building model realized in Modelica.

3.2 Literature Review on Control Strategies

Han et al. [36]	2018	A review of reinforcement learning methodologies on control systems for building energy: Han et al. study the advances in reinforcement learning control methods applied in building energy. They describe the different algorithms and explain the potential and advantages of the implementation of this model-free control in the field of smart building automation. However, a lack of attention in previous scientific research is also noticed by the authors.
Chen et al. [60]	2018	Optimal control of HVAC and window systems for natural ventilation through reinforcement learning: the article presents a control of a HVAC system in the context of natural ventilation, to reduce energy usage and maximize thermal comfort. The control is implemented using reinforcement learning and more specifically a Q-learning algorithm.
Görgeş [50]	2017	Relations between Model Predictive Control and Reinforcement Learning: this article establishes a comparison between reinforcement learning and model predictive control in the context of a linear, discrete and time invariant system. Both control methods are presented in a tutorial manner.
Beghi et al. [17]	2017	Reinforcement Learning Control of Transcritical Carbon Dioxide Supermarket Refrigeration Systems : the authors apply reinforcement learning to the supervisory control of a CO ₂ booster refrigeration system of a supermarket. Changing conditions of operations and system intricacy makes this model-free approach that learns from the interactions with the environment suitable.
Ruelens et al. [61]	2015	Learning Agent for a Heat-Pump Thermostat with a Set-Back Strategy Using Model-Free Reinforcement Learning: Ruelens et al. propose a model-free learning agent to control a heat pump that has an auxiliary heating element. They investigate the use of a set-back method for heat pumps that softens the set-point temperature during night or work hours. This method allows the reduction of energy consumption in comparison to the common constant temperature set-point strategy. The developed approach is based on a batch reinforcement learning technique combined with an auto-encoder, to lower the state vector dimension.

Wen et al. [38]	2015	Optimal Demand Response Using Device-Based Reinforcement Learning: Wen et al. propose a new energy management system formulation for DR issues in small buildings using reinforcement learning. To solve the limitations of this method, they suggest to divide the problems into device clusters. Results of simulation are also presented in the study.
Fazenda et al. [62]	2014	Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems: Fazenda et al. apply reinforcement learning to a supervisory control of an HVAC system. The goal of the control is to reduce the cost of energy and to maintain the occupant's comfort. The reinforcement learning approach is applied to both a continuous and discrete practical case.
Ernst et al. [51]	2009	Reinforcement learning versus model predictive control: a comparison on a power system problem: Ernst establishes a comparison between reinforcement learning and model predictive control using a unified framework and applies both approaches to an electrical power oscillations damping problem.
Dalamagkidis and Kolokotsa [49]	2008	Reinforcement Learning for Building Environmental Control: the authors propose a controller based on reinforcement learning for an HVAC system. A case of study is presented, in which results are compared to other control approaches (fuzzy-PD and ON/OFF).
Liu and Henze [16]	2006	Evaluation of Reinforcement Learning for Optimal Control of Building Active and Passive Thermal Storage Inventory: this paper presents a model-free supervisory control of an electrically driven chilled water system for commercial buildings. Various learning scenarios are discussed. The authors also investigate the influence of parameters, state, and action space configuration.

Han et al. [36] (2018): A review of reinforcement learning methodologies on control systems for building energy

Han et al. review the motivation for building energy optimization and present the fundamentals of the reinforcement learning approach. The authors establish a categorization of the different reinforcement learning algorithms dividing them into two main groups: model-based or model-free. They describe the basics of some of these algorithms, as well as their applications in building energy.

Han et al. review the existing control strategies in building energy. They divide these strategies

3.2 Literature Review on Control Strategies

into classic (e.g. PID or ON/OFF), hard (e.g. MPC), soft (e.g. ANN) and other control methods. The authors introduce the main attributes of these categories.

Overall, the authors notice scarcity in the number of publications about reinforcement learning for building energy control. Although reinforcement learning does not need a model, the "curse of dimensionality" is seen as an inherent problem of the approach.

3.2.2 Literature on Artificial Neural Networks

Table 3.2 reviews the publications of ANN-based controls of energy systems. Neural networks were used in different contexts, such as the approximation of models, or their combination with other control approaches.

Table 3.2 – Summary of the literature review on artificial neural networks as control approach for HVAC, heat pumps and similar energy systems

Author	Year	Publication
Huang et al. [63]	2015	A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings : Huang et al. design an ANN-based predictive control for HVAC systems in commercial buildings. They use ANN to model the interaction between different zones of the building. The control uses a start-stop rule and successfully maintains comfort temperature and reduces energy consumption.
Karabacak and Cetin [30]	2014	Artificial neural networks for controlling wind–PV power systems: A review : Karabacak and Cetin review the application of neural networks in PV and wind energy conversion systems. They conclude that for most cases, ANN show a better performance than traditional controls.
Mohammadzaheri et al. [34]	2012	A critical review of the most popular types of neuro control : in this paper, the main types and characteristics of neurocontrols, in other words of controls in which at least one artificial neural network is utilized, are described. The authors note that artificial neural networks went from auxiliary controllers, in combination with conventional control, to sole nonlinear controllers and back to auxiliary controllers in recent times, as hybrid ANN controllers.

Mohanraj et al. [32]	2012	Applications of artificial neural networks for refrigeration, air-conditioning and heat pump systems-A review : Mohanraj et al. reviewed more than 90 papers on the application of ANN for modeling, forecasting and controlling of refrigeration, air conditioning, and heat pump systems. The reviewed literature confirms that ANNs successfully and accurately serve in this field. Additionally, they state that the hybridization of artificial neural networks with other expert systems would perform better than a traditional ANN approach.
Ekren et al. [64]	2010	Comparison of different controllers for variable speed compressor and electronic expansion valve : in this publication a proportional integral differential, a fuzzy logic, and an ANN control algorithms are compared. The controls are applied to a variable speed compressor and an electronic expansion valve of a chiller. The ANN approach outperformed the other two, because of the achievement of lower power consumption and an improvement of stability.
Al-Alawi et al. [33]	2007	Predictive control of an integrated PV-diesel water and power supply system using an artificial neural network : the publication investigates the implementation of a control based on ANN to optimally operate an integrated hybrid PV-based water and power supply system. The control is designed to maximize the consumption of solar power when available and decide on the ON and OFF states of the diesel generator. The ANN model successfully works as a power usage and generator status predictor.
Kalogirou [29]	2001	Artificial neural networks in renewable energy systems applications: a review : Kalogirou reviews the implementation of artificial neural networks in different renewable energy systems, suggesting that this alternative approximation method is not to be underestimated. The authors also underline its modeling, mapping, and system identification potential.

Mohammadzaheri et al. [34] (2012): A critical review of the most popular types of neuro control

The publication reviews the most common neuro controls. Neuro controls appear when at least one ANN is involved in the control command generation.

First, the authors introduce the basics of the neural network architectures used in control. They go over the fundamentals of multi-layer perceptrons and radial basis function networks,

3.2 Literature Review on Control Strategies

as well as their advantages and disadvantages. Then, Mohammadzaheri et al. review the historical development of neuro controls. In the 1980s, the neural networks were used to map the output to the control command. This method is called inverse dynamics. In the early 1990s, researchers used two ANNs to model and control. This second approach is called by the authors "model reference control". Later neuro-predictive controllers were introduced, where ANN worked as a predictive model. In this case, the controller or optimizer could also be implemented with a neural network. The next approach was the so-called "feedback linearization control". In the 2000s, researchers used ANNs to model a complex part of the system's dynamics. The authors describe the characteristics of each control type.

Overall the authors notice that ANN was first used in combination with other conventional controllers. Then, researchers started to use them as sole controllers. In recent years, the trend among the researchers is to use them again as auxiliary controllers or designing hybrid ANN controllers.

3.2.3 Literature on Evolutionary Algorithms

Table 3.3 gathers and summarizes publications on different evolutionary algorithms used in control of energy systems.

Table 3.3 – Summary of the literature review on evolutionary algorithms as control approach for HVAC, heat pumps and similar energy systems

Author	Year	Publication
Delgarm et al. [65]	2016	Multi-objective optimization of building energy performance and indoor thermal comfort: A new method using artificial bee colony (ABC) : this paper approaches the problem of conflicting objectives of thermal comfort (measured by the percentage of people dissatisfied or PPD) and the energy consumption of buildings employing a multi-objective optimization process based on simulation. The algorithm used is a multi-objective artificial bee colony (MOABC) implemented on MATLAB and using an EnergyPlus simulation. The authors study the case of a single office room and notice a notable improvement in thermal comfort and a reduction of energy consumption.
Congradac and Kulic [66]	2009	HVAC system optimization with CO₂ concentration control using genetic algorithms : Congradac and Kulic propose a genetic-algorithm-based CO ₂ concentration control to operate an HVAC system with the objective of reducing energy performance. Through a simulation, the authors quantify the power saved with the implemented method.

Mossolly et al. [67]	2009	Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm : Mossolly et al. Investigate different control strategies for an HVAC system. The optimization problem is based on the cost of energy consumption and is solved with a genetic algorithm implemented in MATLAB.
Nassif et al. [68]	2004	Evolutionary algorithms for multi-objective optimization in HVAC system control strategy : this publication presents a supervisory control for an HVAC system that is optimized by a two-objective evolutionary algorithm, more specifically the NSGA-II algorithm. The two objectives of the optimization are the reduction of energy consumption and the increase in thermal comfort. The results show that substantial energy savings are achieved after implementation.
Caldas and Norford [43]	2003	Genetic algorithms for optimization of Building Envelopes and the Design and Control of HVAC Systems : Caldas and Norford review the notion of genetic algorithm and its applications in building energy topics. More specifically, they present previous applications of genetic algorithms in building design and HVAC configuration and operation.
Wright et al. [69]	2002	Optimization of building thermal design and control by multi-criterion genetic algorithm : Wright et al. recognize the building thermal design as a multi-criterion optimization problem because a compromise between resident thermal comfort and energy cost needs to be found. In that context, they study the implementation of a multi-criterion genetic algorithm (MOGA) as search strategy that proved fast progress towards the optimal solution. They underline the potential of this approach in building optimization problems.

3.2.4 Literature Review on Hybrid Control Strategies

Table 3.4 gathers different publications that implement hybrid approaches. All these publications use at least one machine learning technique. Each combination has different characteristics.

3.2 Literature Review on Control Strategies

Table 3.4 – Summary of the literature review on hybrid machine learning approaches to control HVAC, heat pumps and similar energy systems

Author	Year	Publication
Mocanu et al. [70]	2019	On-line Building Energy Optimization using Deep Reinforcement Learning : the authors apply two techniques of deep reinforcement learning: DQN and deep policy gradient (DPG) to a scheduling problem for energy management systems in buildings. The objective of the algorithm is to minimize energy cost.
Wei et al. [71]	2017	Deep Reinforcement Learning for Building HVAC Control : Wei et al. develop a Deep Reinforcement Learning (DRL) algorithm to reduce energy cost of HVAC systems in buildings. The proposed algorithm is compared to reinforcement learning and rule-based approaches. DRL approach achieves better results.
Afram et al. [47]	2017	Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system : Afram et al. apply an ANN-based MPC to the case of an HVAC system of a residential building in Canada. They use the algorithm best network after multiple iterations (BNMI) which outperformed the single iteration training implemented in MATLAB. They developed a dynamic set-point temperature strategy that performs better than the fixed set-point approach.
Shah and Gopal [72]	2016	Model-free predictive control of nonlinear processes based on reinforcement learning : Shah and Gopal solve the challenge of finding a suitable nonlinear model for MPC by using a model-free predictive control based on reinforcement learning. They apply the control to the practical case of a continuous stirred tank reactor.
Yu et al. [15]	2015	Application of multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design : The authors approach the trade-off of thermal comfort and energy consumption present in building design with a multi-objective optimization model, based on the non-dominated sorting genetic algorithm NSGA-II algorithm. Thanks to a back-propagation optimized model they model building behavior and predict both energy and thermal performance. A practical example of a building design problem in China is also presented.

Ferreira et al. [73]	2012	Neural networks based predictive control for thermal comfort and energy savings in public buildings : This paper presents an MPC to control an HVAC system, with the goals of improving thermal comfort (characterized by the predicted mean vote or PMV) and minimizing energy consumption. The predictive model is determined by means of neural networks, identified using a multi-objective genetic algorithm. Ferreira et al. use a branch and bound approach as an optimization method.
Matallanas et al. [74]	2012	Neural network controller for Active Demand-Side Management with PV energy in the residential sector : In this publication an active demand side management (ADSM) control based on ANNs is implemented. The control uses a genetic algorithm to set the parameters of the neural networks. The final goal of the control is to maximize self-consumption.
Huang and Lam [44]	1997	Using genetic algorithms to optimize controller parameters for HVAC systems : In order to enhance the performance of an HVAC system, the authors develop an adaptative learning algorithm that makes use of genetic algorithms to tune and select the parameters of a PID controller.

Wei et al. [71] (2017): Deep Reinforcement Learning for Building HVAC Control

The authors focus on the intelligent scheduling of HVAC systems in the residential sector. They note the inefficiency of rule-based and model-based approaches. These strategies depend heavily on the accuracy of the building model. To solve this, researchers proposed data-driven reinforcement learning strategies. However, traditional Q-learning algorithms are no match to large state spaces. Wei et al. apply the novel technique of deep reinforcement learning. This method was already applied in the field of gaming. This publication is (according to the authors) the first to implement deep reinforcement learning in the field of HVAC control. In this paper, the authors formulate the operation of the HVAC system as an MDP and propose a control based on DRL. To test and validate their work, they use the simulation tool EnergyPlus. The objective of the algorithm is to minimize energy costs and to maintain thermal comfort.

The DRL algorithm developed by the authors uses artificial neural networks as an approximation of the Q-value function. They also adapt the algorithm to work in a multiple zone building. They compare the performance of the DRL algorithms to the performance of a rule-based (ON-OFF) control as well as a traditional Q-learning algorithm. The study shows that the DRL algorithm can keep thermal comfort while achieving a significant reduction of energy cost (compared to the on-off, and Q-learning approaches).

Mocanu et al. [70] (2019): On-line Building Energy Optimization using Deep Reinforcement Learning

Mocanu et al. recognize reinforcement learning as the most suitable control algorithm to minimize the cost of energy in buildings. However, they also noticed in previous research the limitations of reinforcement learning when facing large-scale problems due to the "curse of dimensionality". Interest returned when reinforcement learning and artificial neural networks (Deep Learning) fused in a new approach called Deep Reinforcement Learning. In this paper, Mocanu et al. apply Deep Reinforcement Learning to an on-line scheduling and optimization problem for energy management systems in the residential sector. The authors describe the application of DRL in the context of smart grid and smart buildings. They implement two different algorithms: Deep Q-Learning and the Deep Policy Gradient method both at the aggregated and building level. Mocanu et al. intend to minimize cost, reduce peaks and meet the constraints of electrical devices. The study concludes that both DQN and DPG methods can be successfully implemented to flatten the energy profile or minimize energy. The authors notice a better performance of the DPG method to solve scheduling problems both at the aggregated and building level.

Matallanas et al. [74] (2012): Neural network controller for Active Demand-Side Management with PV energy in the residential sector

Matallanas et al. combine automatic control with a Demand-Side Management system. They call this concept active demand-side management. The implementation of the ADSM in households can reduce electrical network stress, increase self-consumption and reduce congestion situations.

The authors study the case of a household provided with PV generation and supported with batteries. The solar house is connected to the grid and has deferrable and non-deferrable loads. The objective of the ADSM controller developed by the authors is to maximize self-consumption by scheduling the activation of the loads when the PV generation is higher. To achieve this, the system needs the input from the user and a forecast of the generation, to schedule the desired tasks. The ADSM consists of a scheduler, a coordinator and an actuator. Scheduler and coordination are artificial neural networks. The parameters of the network are tuned by a genetic algorithm. The results of the study show that GA-tuned ANN control can be implemented as an ADSM system to increase self-consumption in PV-provided buildings.

3.3 Conclusion of the State of the Art Chapter

The beginning of the chapter presents the state of the art of the PVCool project. The test rig and HiL Simulation environments are described. The hardware, software and the Simulink model of the PVCool project are briefly introduced.

Then, the chapter reviews the relevant publications of controls for building energy systems based on machine learning techniques. The most widely used approaches in control are reinforcement learning, artificial neural networks, and evolutionary algorithms. The model-free approach of reinforcement learning is frequently limited by the "curse of dimensionality". The artificial neural networks are often used in control as a modeling tool and are also often used to design hybrid controllers. Genetic and evolutionary algorithms are optimization tools in control. Many authors use combinations of these three methods or a combination of one of these techniques with conventional controls to improve the performance of the building energy control. Section [3.2.4](#) presents different publications that use these hybrid approaches.

4 Decision on the Supervisory Control Approach based on Machine Learning

This section focuses on the structure, goals, and approach of the designed supervisory control. Section 4.1 presents the architecture of the control, as well as its main objectives. Section 4.2 details the considered approaches and justifies the final approach choice.

4.1 Structure and Goals of the Supervisory Control

As briefly introduced in section 2.3.1, the control functions in building energy systems can be local or supervisory. Supervisory controls are often referred to as energy management systems [46]. Local controls are lower-level controls that ensure a robust functioning of the dynamic process and aim at tracking a given setpoint or reference. Common examples of local controls are PID or rule-based controllers. Instead of receiving a reference, supervisory controls receive measurements from the environment and consider the overall system characteristics and components interactions. With this input, the objective of supervisory controls for HVAC and building energy systems is to determine the optimal setpoints that will optimize operation. Common objectives among HVAC supervisory control systems are to minimize energy cost and/or maintain thermal comfort [46]. The obtained setpoint trajectory can then be used by subordinate local controllers. The latter interact directly with the cooling or heating devices of the building.

This thesis develops a supervisory control system, whose flowchart is shown in figure 4.1. The supervisory control takes in environmental measurements and predictions such as the average ambient temperature or the available PV power deduced from solar irradiation. It also receives feedback from the state of the building energy system, for instance, the temperature of the thermal storage. The building energy system also suffers disturbances that affect the process such as heat losses [8].

The objective of the supervisory control, in this case, is to minimize the electricity cost of the system. As the price of buying electricity from the grid is higher than the price of selling unused PV energy to the grid, self-consumption is naturally being maximized. Self-consumption refers to the action of consuming or storing the generated PV instead of selling it to the grid.

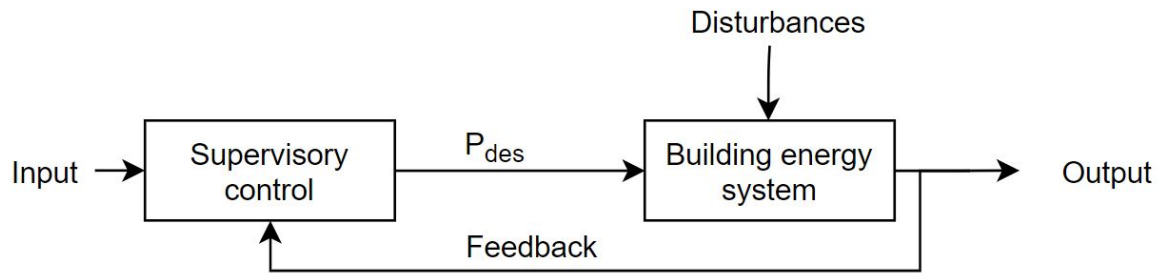


Figure 4.1 – Scheme of the studied supervisory control [8]

A further benefit of increasing self-consumption is the reduction of emissions. While minimizing energy costs and maximizing self-consumption, the boundaries of the thermal energy storage cannot be exceeded, and the building demand should be covered. The supervisory control calculates the setpoint trajectory of the desired compressor power P_{des} of the cold vapor cycle that minimizes energy cost.

4.2 Decision for Control Strategy Approaches

In light of the results of the literature review in section 3.2, it is clear that reinforcement learning, artificial neural networks, and evolutionary algorithms all have advantages and disadvantages. Hybrid approaches can combine the advantages of different methods to compensate for their disadvantages. According to Ahmad et al. [3], hybrid techniques lead to better performance and solutions by combining the different strategies thoughtfully. For this reason, in this thesis, a hybrid approach is chosen. Different combinations all lead to different characteristics. Below a description of the considered approaches, as well as their advantages and disadvantages are given.

- **Genetic algorithms and artificial neural networks:** Genetic algorithms can optimize a neural network. Ferreira et al. [73] use a multi-objective genetic algorithm to identify the structure of the ANN model. Matallanas et al. [74] use a GA to tune the parameters of their ANN-based control. The combination of evolutionary algorithms like GA with ANNs gives acceptable approximation and computational speed. However, this hybrid technique needs a large amount of training and simulation data [15].
- **Multi-agent systems and reinforcement learning:** A multi-agent system is an agent-based technique that splits complex problems into smaller sub-problems. Each agent is

in charge of a particular sub-problem. These intelligent agents can work autonomously towards a specific goal and can interact, negotiate and cooperate with the other agents [36]. The implementation of reinforcement learning algorithms into multi-agent systems enables learning abilities among agents. If one agent fails, the others can react and learn from the new environment. However, agents could also work independently rather than in cooperation. This would lead to suboptimal results. Another downside of the technique is that the complexity increases with the number of agents [36].

- **Reinforcement learning and artificial neural networks (DRL):** Wei et al. [71] and Mocanu et al. [70] already successfully implemented DRL online scheduling controls to online scheduling of energy systems to minimize energy costs. DRL algorithms can solve challenging problems and deal with the intrinsic variabilities and uncertainties of resident's electrical needs or renewable energy generation [70]. The DRL algorithm must, however, learn using a trial and error approach, which can lead to suboptimal results at the beginning of the learning phase. To solve this, DRL agents can be pre-trained [75].

In table 4.1, an overview of the main advantages and disadvantages of the control approaches considered is given. Some of the publications on hybrid techniques listed in section 3.2.4 consider the combination of MPC, ANN, and GA. A similar approach is already developed by Rauscher [8]. Rauscher [8] implements an MPC control using an ANN to estimate the COP of the heat pump. In his thesis, he considers using a GA to solve the MPC optimization but finally chooses a pattern search algorithm. As this approach has already been investigated, it is not considered in this thesis.

Considering the information in 4.1, the DRL approach seems more suited for the design of a supervisory control for the PVCool project, because of its successful previous application in scheduling and energy cost minimization problems. Additionally, considering the balance between advantages and disadvantages in table 4.1, DRL has a higher amount of advantages compared to the number of disadvantages. Therefore, the deep reinforcement learning approach is chosen in this thesis.

4.3 Conclusion

Table 4.1 – Overview of advantages and disadvantages of the considered hybrid control strategies

Approach	Advantages	Disadvantages
Genetic algorithms and artificial neural networks	<ul style="list-style-type: none"> • acceptable approximation [15] • speed in evaluation [15] 	<ul style="list-style-type: none"> • needs extensive training and simulation data [15]
Multi-agent systems and reinforcement learning	<ul style="list-style-type: none"> • agents can share and exchange information [36] • if one agents fails, the others can react and learn from the new environment [36] 	<ul style="list-style-type: none"> • complexity increases with the number of agents [36] • independent work of agents can lead to suboptimal results [36]
Reinforcement learning and artificial neural networks (DRL)	<ul style="list-style-type: none"> • can solve challenging tasks [70] • effective in reducing energy costs [71] • suited for online scheduling [70] • able to deal with uncertainty [70] 	<ul style="list-style-type: none"> • trial and error approach [75]

4.3 Conclusion

This chapter presents the goals and general structure of the supervisory control system. The control receives input from the environment and feedback from the energy system. With this information, the supervisory control chooses the setpoint trajectory of the compressor power of the cold vapor cycle to minimize energy costs. Different machine learning approaches can be implemented to perform this task. However, as seen in section 3.2 hybrid strategies tend to perform better. Different hybrid approaches are compared. Among the considered hybrid strategies, deep reinforcement learning is the most promising and has already been successfully implemented in similar past studies ([70], [71]). Therefore, a deep reinforcement learning supervisory control is implemented in this thesis.

5 Development of the deep reinforcement learning Control Strategy

This chapter focuses on the implementation process of the DRL supervisory control system. The first section focuses on the resulting architecture of the supervisory control and the elements of the DRL control. Section 5.2 reviews some design features and choices of the control.

The supervisory control is developed using a bottom-up approach: starting with basic features and building from there a more complex system. This work does not detail all the steps of the programming and development process, only its most important resulting characteristics. A brief overview of the process is given below:

1. The first the basic features of a DRL control are examined and tested through already programmed examples of the Mathworks reinforcement learning toolbox [76]. Both DQN and DDPG algorithms are investigated.
2. In a second step, a simplified version of the supervisory control is created. It is not integrated into the PVCool Simulink model. The goal of this version is to shape the reinforcement learning Simulink environment and to get familiar with the elements and inputs of the control. This version does not consider some elements such as the thermal storage and is therefore incomplete. It investigates both DQN and DDPG algorithms.
3. Then, a complete version of the supervisory control is developed. This version considers all the necessary elements, but like the steps above, it is not integrated into the PVCool Simulink model. This version is used to test the functionality of the control and chose some of the parameter values. The DQN algorithm is chosen.
4. Finally, the supervisory control is integrated into the PVCool Simulink model. The control is tested and verified through short-term simulations.

5.1 General Architecture of the Supervisory Control

This section presents the architecture of the supervisory control. Section 5.1.1 describes the general architecture of the supervisory control and the workflow of the control process. Sec-

tion 5.1.2 details the main elements of the DRL control.

5.1.1 Supervisory Control Architecture

As shown in section 4.2, real-time learning via trial and error from an unknown environment take a lot of time and can lead to suboptimal results at the beginning of the learning phase. To overcome this issue, DRL agents can be pre-trained. In this thesis, the DRL agent is pre-trained using predictive input data. The available input data, which can be accessed by the controller and used as predictions, as well as the compressor control signal have a temporal resolution of one second. However, control systems for building energy have usually a horizon of 24 to 48h. As the DRL agent needs to establish the sequence of actions to be performed over the control horizon, a resolution of one second for the DRL agent would mean that the agent has to optimize via trial and error a sequence of 86.400 to 172.800 actions. This would lead to high computational time and effort. To avoid this, it has been decided that the supervisory control should follow the architecture already developed by Rauscher [8].

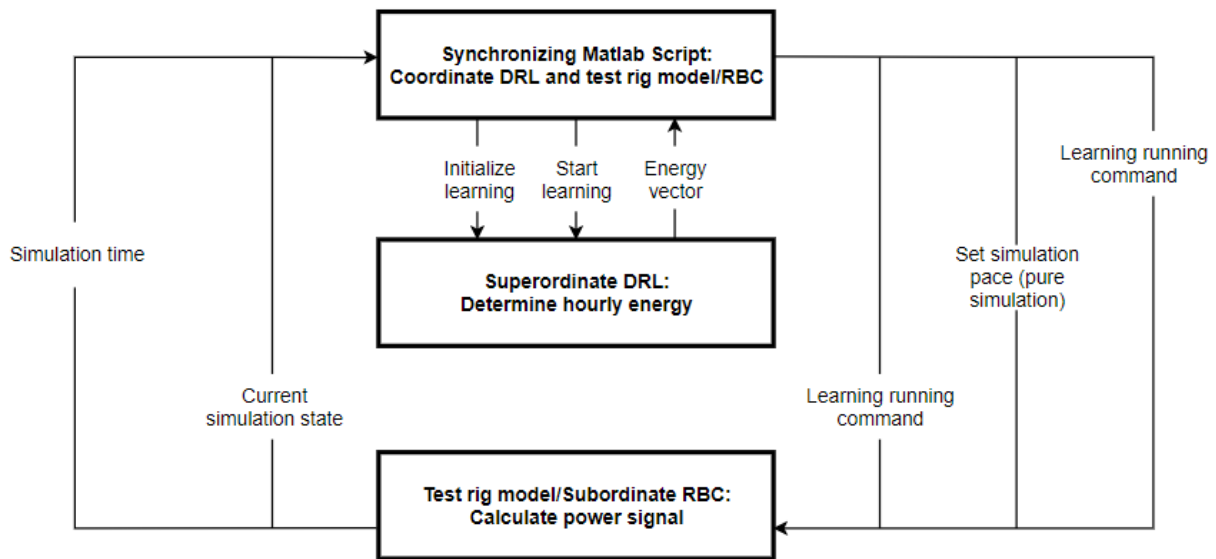


Figure 5.1 – Communication scheme between deep reinforcement learning control, rule-based control, and Matlab script, based on [8]

The supervisory control is divided into a superordinate controller (in this case a DRL controller) with a resolution of one hour for the long-term optimization, and a subordinate control (rule-based control or RBC) that has a temporal resolution of one second. The coordination of both controls, as well as the synchronization of the supervisory control and the test rig, is realized by a Matlab script. Figure 5.1 illustrates the communication process between

5.1 General Architecture of the Supervisory Control

the superordinate DRL, the subordinate RBC and test rig model, and the coordinating Matlab script.

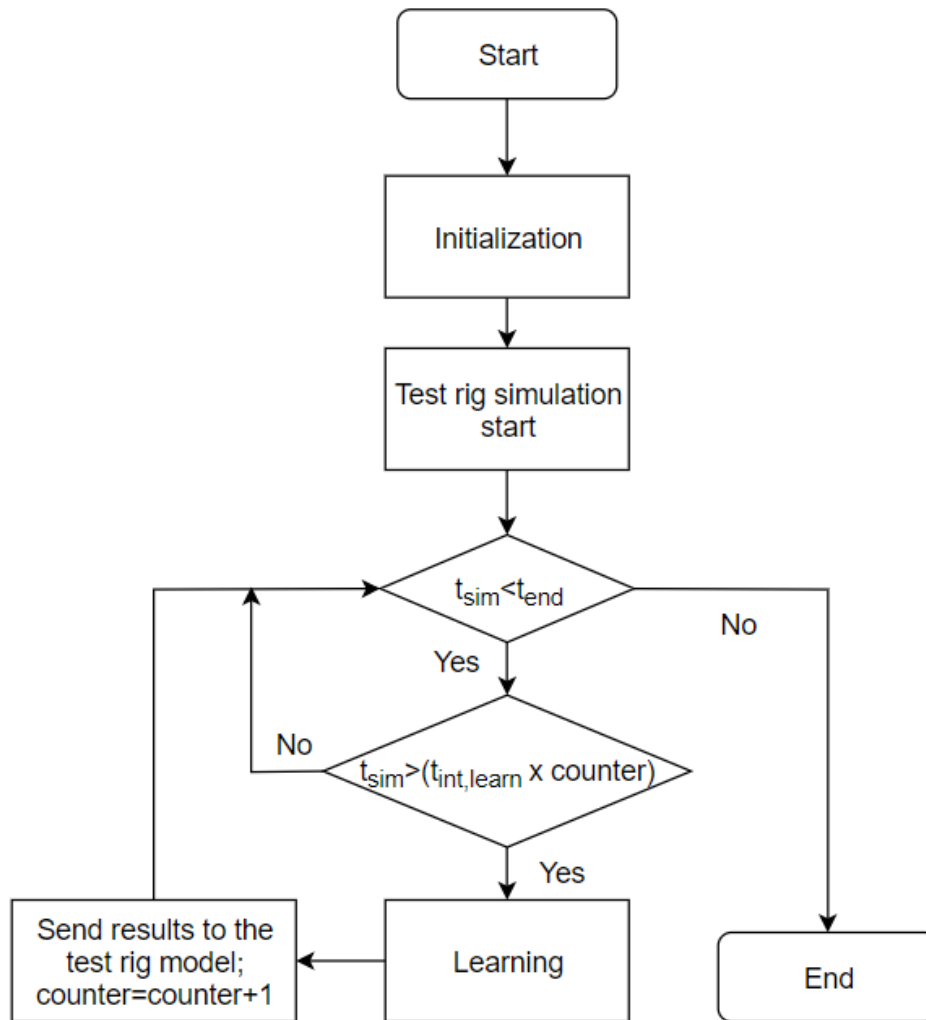


Figure 5.2 – Matlab synchronizing script flowchart, based on [8]

In the HiL simulation mode, the test rig model is simulated at a high pace. Every simulation hour the DRL algorithm is evoked, the simulation pace is slowed down to one simulation second per real-time second, and the learning running flag is set to 0. The role of the flag is to signal when the simulation starts and ends. The goal of the DRL is to find the optimal sequence of hourly compressor electrical energy for the next 48h. In other words, it provides a vector of 48 values, each one corresponding to the amount of electrical energy that the compressor supplies for one hour. Once the learning is finished, the learning running flag is set to 1 and the high simulation pace is restored. The optimized energy vector is sent to the RBC. The RBC takes the energy for the first hour and determines the power signal for the compres-

sor with a resolution of one second. The reader can refer to Rauscher’s thesis [8] to understand the details of the hour-to-second transformation. When the next simulation hour starts, the DRL algorithm is evoked again. Inputs and horizon of the control are shifted one hour. While the agent is trained, the RBC uses the second hour from the last received compressor energy vector to provide the power signal. Once the agent is trained, the RBC uses the first value of the new energy vector to calculate the power signal for the compressor.

All the coordination, as well as the synchronization with the test rig, is done by the Matlab script. Figure 5.2 shows the flow chart of the synchronizing script. The script is initialized, and the test rig simulation starts. Then, the scripts’ loop examines until the simulation end t_{end} is reached, whether the simulation time t_{sim} exceeds the learning starting time $t_{int,learn}$. The learning starting time is determined through the multiplication of the learning starting interval $t_{int,learn}$ and a counter variable. When the learning phase is finished, the results are transmitted to the test rig model and the counter variable is updated and increased. This process continues until the simulation time t_{sim} reaches the simulation end t_{end} .

5.1.2 Architecture and Elements of the DRL Control

As described in section 2.2.2, the objective of the reinforcement learning algorithm is to train an **agent** to perform a task within an **environment**. The agent recognizes the *state* of the environment through *observations*. It decides of an **action** to perform in the environment. The agent then collects a **reward**, indicating how successfully the last action was with respect to reaching the objective. This workflow is implemented in Simulink using the Reinforcement

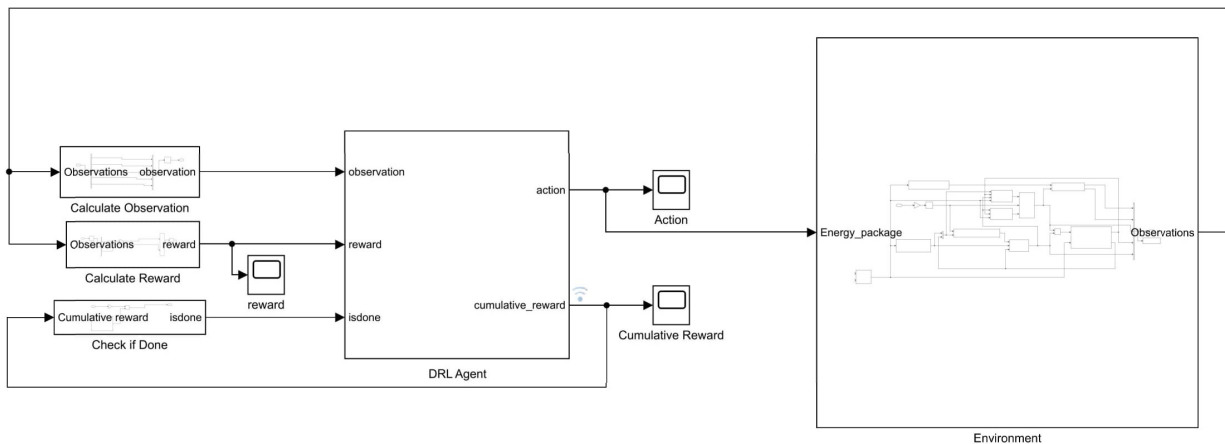


Figure 5.3 – Simulink model of the DRL control

Learning Toolbox [76] and is commanded by a Matlab script. Figure 5.3 shows a screenshot of the resulting model. In this case, the objective of the DRL control (**DRL agent**) is to find the

optimal sequence of compressor energy quantities for the next 48 simulation hours that will minimize energy cost and maximize self-consumption. To do so, the agent uses the information it gets from the environment (**observations**) to generate a series of 48 values of energy quantities (**action**). These are sent to the environment and a simulation is run. The new observations are used to calculate the resulting energy cost of the last simulation (**reward**). The reward guides the agent's trial and error learning process towards the final goal. The agent stores and continuously updates its mapping of actions to observations or environment states (**policy**). Policies in reinforcement learning can take different forms, as seen in section 2.2.2. However, a DRL agent uses artificial neural networks to represent its policy. The individual elements will be described in more detail in section 5.2.

5.2 Design of the DRL Supervisory Control

After reviewing the general architecture and workflow of the control, the design of the main elements (environment, observations, reward and agent) is described in detail in the following sections.

5.2.1 The Environment

The environment is a modified version of the predictive model realized by Rauscher [8]. The predictive model is an approximate version of the building energy system, with a temporal resolution of one hour instead of one second. The predictive model receives prediction inputs of the future energy system behavior and input from the DRL agent. Overall, the environment's 48h-horizon-inputs include:

- The available PV energy (prediction input)
- The ambient temperature (prediction input)
- Building's electrical demand (prediction input)
- Building's thermal demand (prediction input)
- Hourly compressor energy vector (from DRL agent)

The resulting environment or prediction model is shown in figure 5.4. The following blocks, developed by Rauscher [8], are only briefly introduced here.

- **Read PV electrical demand energy block:** This block subtracts the available PV power to the electrical demand.
- **Heat Ambient block:** This block calculates the average ambient temperature. This can be the source or sink temperature, depending on if the building tenants demand heat or cold, respectively.
- **Estimate COP block:** This block contains the same ANN that is used to estimate the COP in the test rig model. The ANN's inputs are the temperature of the heat sink, the temperature of the heat source and the compressor power estimated from the following block.
- **Estimate average compressor power block:** This block uses the temperatures of the heat sink and heat source, as well as the proposed energy vector to approximate the average value of the compressor power.

The most significant modifications introduced to Rauscher's model [8] aim to integrate the optimization constraints into the Simulink model and to be able to deal with both heating and cooling demand. The optimization constraints include the following:

- **Demand coverage:** The thermal and electrical demand of the building must be covered.
- **Compressor energy bounds:** The maximum power that can be provided by the heat pump is $P_{HP,max}$. If the compressor supplies its maximum power for one hour, that means the compressor energy supplied per hour must be below $E_{HP,max}$. The constraint is shown in equation (5.1).

$$0 \leq E_{comp} \leq E_{HP,max} \quad (5.1)$$

- **Thermal storage boundaries:** The state of the thermal storage is defined by the state of charge (SOC). The definition of the state of charge is presented in (5.2).

$$SOC = \frac{T_{storage} - T_{empty}}{T_{fullycharged} - T_{empty}} \quad (5.2)$$

The state of charge of the thermal storage temperature must remain between 0 and 1, as shown in equation (5.3).

$$0 \leq SOC_{storage} \leq 1 \quad (5.3)$$

As the DRL agent proposes a vector of compressor energy values, the compressor energy constraint is implemented when defining the range of actions of the DRL control. The demand coverage and thermal storage temperature boundaries constraints are implemented within

the prediction model. New blocks are created to deal with constraints: the energy bought and energy sold block, the upper bound constraint, and the lower bound constraint. Other blocks, such as the source and sink block or the COP or EER block aim to make the prediction control valid both for cooling and heating purposes. First, the function of the blocks aiming to deal with the optimization is detailed:

- Energy bought and energy sold block:** This block calculates how much energy needs to be bought and sold to the grid to meet the building's demand and maintain the energy balance. The priority of the supervisory control system, before minimizing costs, is to cover the building's thermal and electrical demand. Considering the electrical demand, if the difference between PV energy and electrical consumption (compressor's energy and electrical demand) is negative, missing electrical energy has to be bought from the grid. On the other hand, if that difference is positive, the excess of PV energy is sold to the grid. The block follows the logic described in figure 5.5.

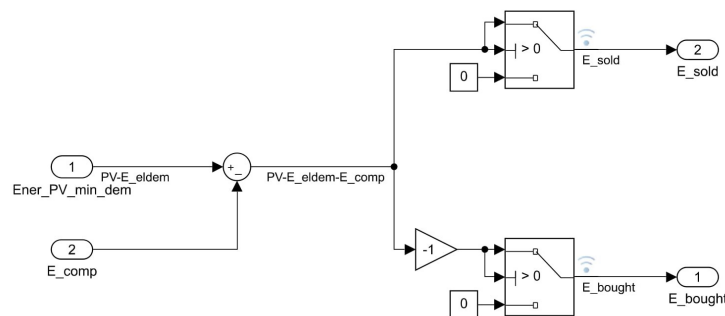


Figure 5.5 – Simulink block that calculates the amount of energy bought and sold to the grid: the signal *Ener_PV_min_dem* refers to the available PV energy and *E_comp* to the energy planned for the heat pump. The outputs of the block are *E_sold* and *E_bought*, the energy sold and bought, respectively

- Upper bound constraint:** This block corrects the values of energy quantities proposed by the DRL algorithm to ensure that the values make physical sense. The upper bound constraint block ensures the heat pump does not transfer more heat to the hot storage than it can accept (because the state of charge would otherwise surpass one, failing to meet constraint (5.3)). In a similar way in the case of the compressor chiller, it ensures no more heat is withdrawn from the cold storage than it is available so that the state of charge remains below one. Considering the example of a case of heating demand, equation (5.4) shows that when the heat pump provides more heat Q_{HP} than the thermal demand of the building Q_{bui} , the hot storage increases its temperature. Equation (5.4) T_0 being the initial temperature of the storage, m_w the mass of water in the tank

and c_p the specific capacity of water.

$$T_{storage}(h) = T_0 + \sum_{k=0}^{t=h} \frac{1}{c_p m_w} (Q_{HP}(t) - Q_{bui}(t)) \quad (5.4)$$

That means, that when the thermal storage already is at its maximum temperature and the state of charge is one, the maximum heat pump energy is the electrical energy needed to cover thermal demand. Otherwise, the thermal storage would exceed its temperature limits. When the hot storage has not reached its maximum temperature, the maximum compressor energy is the minimum value between the maximum energy that can be supplied by the heat pump and the electrical equivalent to the available heat capacity of the thermal storage. This way, the prediction model limits the proposed values of the DRL algorithm if they were to cause a transgression of the temperature limits of the thermal storage.

- Lower bound constraint:** This block corrects the values of energy quantities proposed by the DRL algorithm to ensure that the values make physical sense. The lower bound constraint ensures that no more heat is extracted by the building from the storage than it can provide. Similarly, in the case of cooling demand it ensures no more heat is transferred to the cold storage than it can accept. In other words, this constraint guarantees that the system does not rely on the thermal storage to provide heat or cold to the building when the state of charge of the thermal storage is 0. Considering the example of a case of heating demand, equation (5.4) shows that when the heat pump provides less heat than the thermal demand of the building, the thermal storage decreases its temperature. Therefore, to avoid a transgression of the temperature limits of the heat storage and to ensure thermal demand coverage, the minimum compressor energy value when the thermal storage is at its lower temperature limit is the electrical equivalent of the thermal demand. If the state of charge of the thermal storage is between 0 and 1, the block ensures the system does not rely on the storage to provide more than the available heat stored to cover the demand.
- Thermal storage block:** This block outputs the temperature of the thermal storage and its state of charge, considering the effect of both the building's thermal demand and the heat pump. The state of charge of the thermal storage is calculated using equation (5.5). Considering the case of heating demand, the heat that is withdrawn from the thermal storage to heat the building Q_{bui} decreases the hot storage temperature and state of charge. The heat provided by the heat pump Q_{HP} increases the heat storage temperature and state of charge. The difference between the two heat quantities is divided by the mass of the water in the water tank m_w and the specific heat capacity c_p . This leads to the temperature variation of the water in the heat storage. The result is divided by the temperature range of the thermal storage. The variations over all time steps are added and the result is summed to the starting state of charge SOC_0 to obtain the state

of charge of the thermal storage at an hour h .

$$SOC(h) = SOC_0 + \frac{1}{T_{st,full} - T_{st,empty}} \sum_{k=0}^{t=h} \frac{1}{c_p m_w} (Q_{HP}(t) - Q_{bui}(t)) \quad (5.5)$$

The following two blocks are included in the prediction model in order to make it applicable both for heating and cooling applications.

- **Source and sink block:** This block observes whether the building is in heating or cooling mode. The current mode is determined by a signal. If the signal is one, the CVC acts like a heat pump. The cooling mode is signaled by a minus one. If the heating mode is activated, the ambient air is selected as the source and the thermal storage as the sink. On the other hand, if the cooling mode is activated, the thermal storage is selected as the source and the ambient air as the sink. The block is shown in figure 5.6.

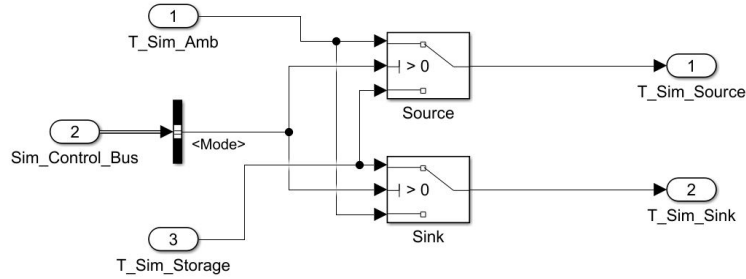


Figure 5.6 – Simulink source and sink block: the variable T_Sim_Amb corresponds to the average ambient temperature and the variable T_Sim_Storage to the temperature of the thermal storage. The block determines which of these temperatures is the source (T_Sim_Source) and which the sink (T_Sim_Sink)

- **COP or EER block:** This block decides whether the COP or the EER is used. In case of a heating application, the COP is passed to the next blocks. In case of a cooling application and considering no losses, the EER is calculated as shown in equation (5.6).

$$EER = \frac{Q_{source}}{P_{in}} = \frac{Q_{sink} - P_{in}}{P_{in}} = COP - 1 \quad (5.6)$$

The results of the prediction model are sent to the DRL agent. These outputs are called observations in the reinforcement learning environment.

5.2.2 The Observations

The DRL agent senses the state of the environment through observations. The observations and their accompanying reward guide the algorithm's learning towards the optimal solution. In the DRL supervisory control of the PVCool project, the following variables are taken as observations and sent from the environment to the DRL algorithm:

1. Energy bought from the grid
2. Energy sold to the grid
3. Compressor energy
4. COP of the heat pump
5. State of charge of the thermal storage
6. Corrected compressor energy (after going through the optimization constraints blocks in the environment)

In every case, these observations are vectors of 48 values, one for each hour of the control horizon. Some of these observations are used to calculate the reward resulting from the 48 simulation hours trial.

5.2.3 The Reward

The goal of the supervisory control is to minimize the energy cost. A reinforcement learning algorithm tries to maximize the reward [35]. The reward function R shown in equation (5.7) minimizes energy cost insofar as it maximizes the energy profits.

$$R = \sum(E_{sold} \cdot p_{sell}) - \sum(E_{bought} \cdot p_{buy}) \quad (5.7)$$

The energy sold to the grid E_{sold} has a positive impact on the total reward R , and the energy bought from the grid E_{bought} decreases the reward value. In this thesis, the price of selling electricity to the grid p_{sell} is taken as the German feed-in tariff in 2019 and set to $0.11 \frac{\text{€}}{\text{kWh}}$ [77]. The price of buying electricity from the grid is taken from the Stadtwerke München GmbH [78] tarif and is set to $0.30 \frac{\text{€}}{\text{kWh}}$. This price situation is typical in the current German electricity market. As the buying price is significantly higher than the selling price, the DRL agent is

guided towards minimizing the amount of energy bought from the grid. It is also naturally guided towards increasing self-consumption.

The reward guides the learning process of the DRL agent. The design of the DRL agent is detailed in the next section.

5.2.4 Design of the DRL Agent

The agent is at the center of a control based on reinforcement learning. The agent observes the state of the environment and receives the corresponding rewards, takes actions in this environment, and progressively learns via trial and error the path towards the maximization of long-term reward [16]. As explained in section 2.2.2, the policy or the mapping of states (observations) and actions can be represented by a neural network. The agent is then a deep reinforcement learning agent. This section will first focus on the DRL algorithm chosen for this thesis. Then, some design characteristics of the agent are presented. Finally, the training process is introduced.

DRL Algorithm

Matlab supports two types of deep reinforcement learning agents: DQN or DDPG [40]. Both agents can observe continuous or discrete states. However, the DQN algorithm can only perform actions in the discrete action state and the DDPG algorithm only in the continuous action state.

The DQN agent trains a critic or value function that can estimate future rewards using state observations and actions as input [35]. The critic is an artificial neural network. During the training, the agent explores the action space and updates the critic in every step of the learning process. The agent also stores past experience in a buffer [40].

The DDPG does not only involve a critic but also an actor, which is a second neural network. The function of the actor is to output actions based on state observations. Then, the critic reads the state observations and the actions chosen by the actor and outputs the value of the state-action pair [40]. Actor and critic are both updated during the learning process. Past experience is stored in a buffer [35].

In the developed supervisory control, the action is the sequence of 48 energy quantities to be supplied by the heat pump in the next 48 hours. Considering the example of a heat pump with a maximum electrical power of 3.000 W, the energy quantities can go from 0 to 3.000 Wh, as seen in section 5.2. Both DQN and DDPG agents are initially investigated. In the case of the discrete action space supported by the DQN agent, the agent is designed in a way that actions can vary between 0 and 3.000 Wh with a resolution of 100 Wh. The discrete action space for the DQN agent consists then of 31 possible actions. In the case of the continuous action space

5.2 Design of the DRL Supervisory Control

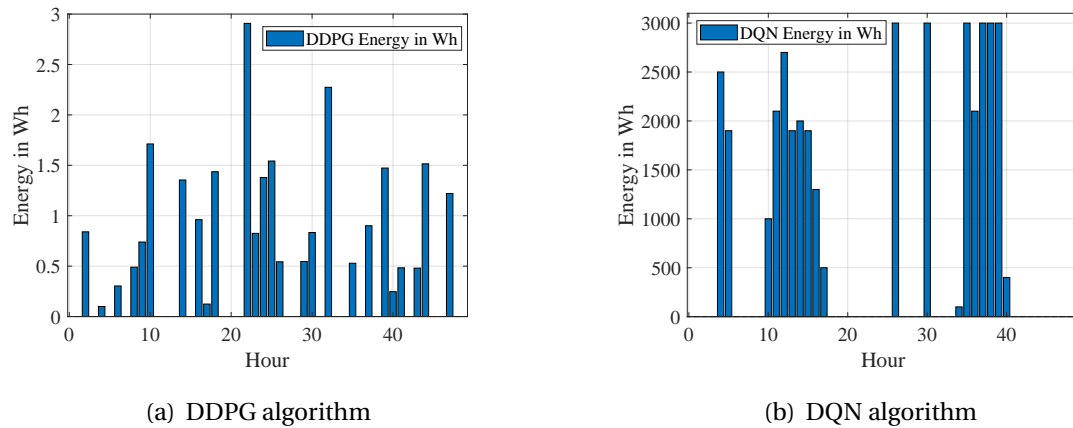


Figure 5.7 – Energy values proposed by the DDPG and DQN agents for the 1st of November 2008

supported by the DDPG agent, the agent can choose all the possibilities between 0 and 3.000 Wh.

Tests performed demonstrate better behavior and performance of the DQN agent in this application. Figure 5.7 shows the comparison of the proposed energy quantities of the two algorithm. The DQN agent explores energy values within the whole interval (from 0 to 3.000 Wh) whereas the DDPG agent only proposes low energy quantities (from 0 to 2,9 Wh), close to the lower bound, with a sensible variation. Overall, lower energy prices are achieved by the DQN algorithm. For these reasons, the DQN algorithm is selected moving forward.

Design of the DQN Agent

The critic trained by the DQN agent takes actions and state observations as inputs and gives the value of the state-action pair as an output. The critic is defined in a DRL Matlab Script. The training is performed using the Simulink reinforcement learning model depicted in figure 5.3. The critic network is illustrated in figure 5.8. The critic receives observations and actions as inputs. Observations and actions result from a 48h-horizon trial or episode in the environment. The action corresponds to the desired compressor energy. The observations include the energy bought from the grid, the energy sold to the grid, the corrected compressor energy, the COP of the heat pump, and the state of charge the heat storage. The corrected compressor energy refers to the compressor energy vector proposed by the agent that has been corrected when needed in the prediction model to ensure meeting the system's constraints. The critic only has one output, providing the value of the actions in that state. Between input and output layers the number of hidden layers has to be selected. As networks of three hidden layers and more can learn and model complex representations [79], it is decided that the network should have at least three hidden layers. To keep the computational complexity as low as possible, the network has three hidden layers. The number of neurons in each of the hidden layers is chosen using the following rules of thumb [80]:

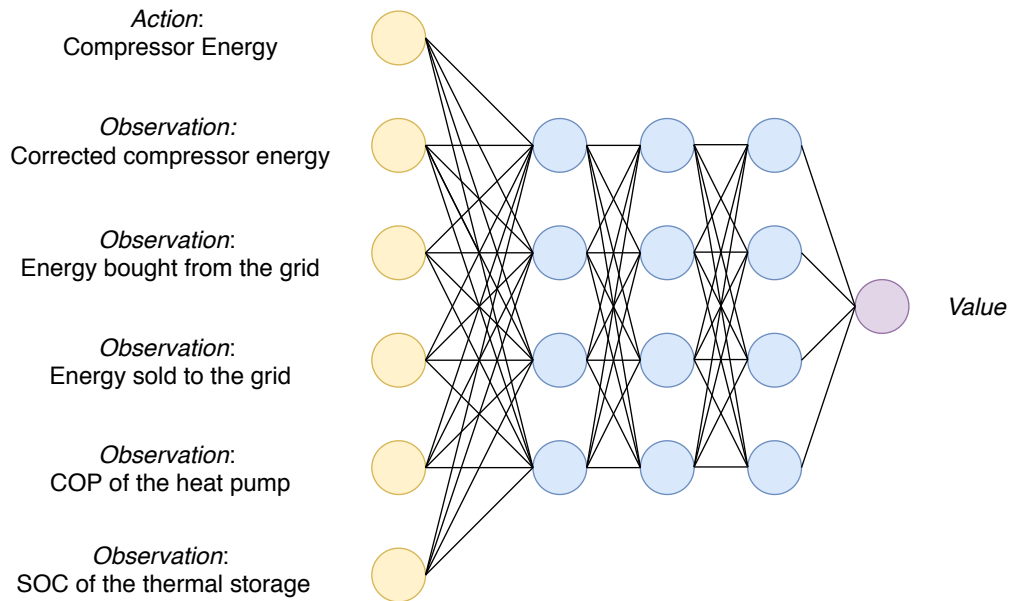


Figure 5.8 – Representation of the critic of the DRL Algorithm

1. The size of a hidden layer must be between the size of the output layer and the size of the input layer
2. The number of neurons in the hidden layer should be two thirds the size of the sum of output and input layer sizes
3. The size of a hidden layer should be less than twice the number of input neurons

As the input layer has six neurons and the output layer one neuron, the number of hidden neurons according to the first rule should be between two and five. Concerning the second rule, the sum of input and output layer neurons is seven, and two-thirds of six is a number between four and five. Tests with both sizes of hidden layers are performed, resulting in faster learning for four hidden neurons. Using four neurons per hidden layer also agrees with the third rule as four is less than twice the number of input neurons (twelve). The other characteristics of the neural network, such as the use of a fully connected layer or the choice of the relu activation function are inspired by exemplifying projects of the Reinforcement Learning Toolbox [76].

Training Process

Every simulation hour, the DRL control is evoked to learn the optimal behavior of the compressor for the next 48h. The learning is conducted through the training of the deep reinforcement learning agent. The training is controlled by the Matlab function `trainingStats`. The

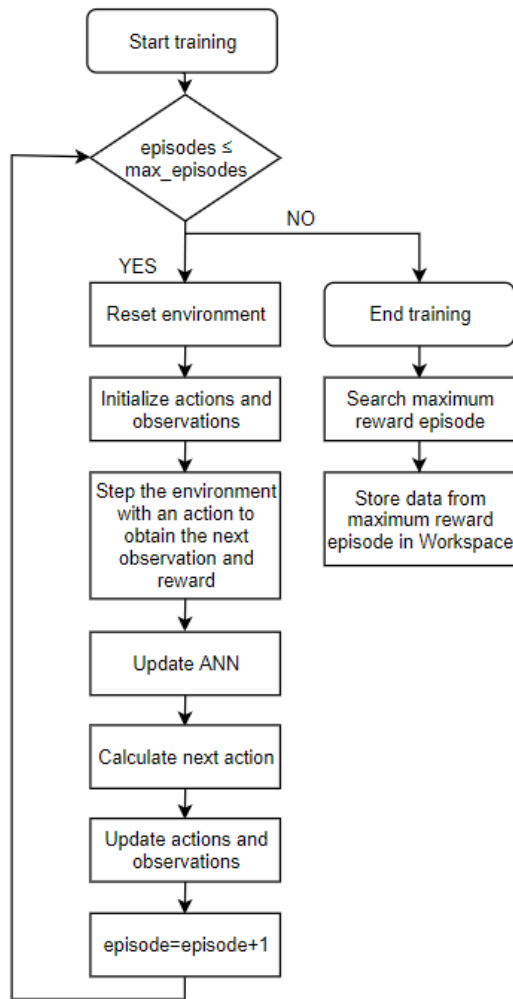


Figure 5.9 – Flowchart of the training process of the DRL Algorithm

function trains the DRL agent in the environment depicted in figure 5.4. The function checks the episode number. In this case, an episode refers to a 48 steps simulation, each step corresponding to one hour. At the beginning of every training episode, the function resets the environment and initializes actions and observations. The agent sends a sequence of 48 compressor energy quantities to the environment and collects the resulting reward and observations. The agent learns from this experience and updates actions and observations. This process continues until the algorithm reaches the maximum number of episodes. When the learning process ends, the information collected during all the episodes is processed. The episode with the maximum reward is selected and the compressor energy vector of that episode is stored and sent to the RBC. Figure 5.9 shows the workflow of the training process.

5.3 Conclusion

This section has introduced the overall structure of the supervisory control of the PVCool building energy system. The three main elements of the supervisory control are the superordinate DRL control, the subordinate rule-based controller and the synchronizing script. The goal of the DRL control is to calculate the sequence of compressor energy quantities for the next hours that will minimize energy costs while still covering the demand and maintaining the storage temperature within its operation bounds. The vector of energies is then sent to the rule-based controller that transforms the hourly energy quantities into a power signal with a resolution of one second for the compressor. The coordination is ensured by a synchronizing Matlab script.

Then, the architecture of the DRL control has been introduced. The DRL control includes four main elements: the DRL agent, the environment, the observations, and the reward. The environment consists of a predictive model that approximates the building energy system. The observations are relevant outcomes from every trial of the agent. They allow the calculation of the reward function (in this case, the energy cost). Although at first two algorithms have been investigated for the DRL agent, the DQN algorithm is chosen at the early stages, showing better performance. The DQN agent trains a neural network that outputs the value of a state-action pair using actions and states as inputs. The agent can choose between a discrete number of energy values within the operation range of the compressor.

6 Verification of the Control Strategies with Short Term Simulations

This chapter examines the verification process of the DRL control. First, section [6.1](#) describes the inputs and outputs used for the verification, as well as the control evaluation measures. Section [6.2](#) describes the testing of the code before its integration in the PVCool Simulink model. The objective of this first testing is to verify the compliance to control bounds and the effectiveness of the self-consumption strategy. Additionally, the number of episodes for the learning phase is determined. Then, section [6.3](#) discusses the conducted simulations after the integration of the DRL control in the HiL Simulation to verify the outputs of the supervisory control.

6.1 Simulation Data

This section gives an overview of the setup of the conducted simulations. Mainly, the inputs and outputs of the simulation are presented.

6.1.1 Simulation Inputs

The input data used for verification purposes within the DRL environment corresponds to the data used by Heithorst et al. [\[52\]](#) and Rauscher [\[8\]](#). The input data used for the verification of the code include:

- **Electric Demand Data:** The electric demand data are extracted from a dataset supplied by Tjaden et al. [\[81\]](#). The dataset includes the electrical demand data of 74 households at two measurement periods. The data of one of these households were selected by Möhrstedt [\[82\]](#) considering the annual and characteristic electricity consumption of a typical family household. This same profile is used in this thesis.
- **Thermal Demand Data:** The thermal demand data first used by Schmiederer [\[55\]](#) derive from the building simulation tool (BST) data. The BST software aims at the simulation

6.1 Simulation Data

of a building's thermal behavior. The building floor area of 120 m² and a medium insulation standard are taken. Oscillations in the data are flattened.

- **PV Data:** The data are an adaptation of the Solarenergieverein Bayern e.V. measurements of a 1MW peak PV plant. The data is scaled to the building and compressor size to a peak power of 5 kW.
- **Ambient Air Temperature Data:** The ambient air temperature data also derive from the BST data. The data are processed to obtain average ambient temperature values for each hour.

The original data cover one year from the 28th of September 2008 to the 28th of September 2009 with a resolution of one second. As memory usage for annual simulations is excessively high, the data are restructured by Rauscher [8] to be able to comfortably perform monthly simulations. The input data for the DRL controller are stored as a cell array where each cell contains the simulation data for each month starting January 1st with a resolution of one hour. Additionally, the user can also choose to use annual data starting September, 28th.

6.1.2 Simulation Outputs

The outputs collected to verify the control in the first stage of the verification process include actions, observations, computational time, and rewards. Specifically, the following outputs are gathered:

- **Action** (best episode)
 - Compressor energy
- **Observations** (best episode)
 - Corrected compressor energy
 - COP of the heat pump
 - Heat provided by the compressor
 - Energy bought from the grid
 - Energy sold to the grid
 - Thermal storage temperature
- **Reward**
 - Value and episode number of the maximum reward
 - Value and episode number of the maximum average reward
 - TrainingStats episode reward information
 - TrainingStats average reward information
- **Control performance** (full learning process: all episodes)
 - Computational time

The computational time is calculated in the DRL Matlab Script. Actions and observations are exported from Simulink using the To Workspace blocks. Episode rewards are stored in trainingStats. The episode with the highest reward is selected and the corresponding information is extracted. Observations, actions and rewards of the best control episode, along with computational time measures, are stored in a file.

The collected outputs for the second phase of the verification process are the outputs of the supervisory control system. These consist of:

- Applied compressor power
- Applied energy sum applied of the RBC for the current hour

6.2 Verification of the DRL Control

This section describes the verification process of the developed DRL control. Here, the control is verified before its integration in the PVCool Simulink model to confirm its optimal behavior using short-term simulations of the months of March and November. These months are interesting because they present a significant amount of PV and thermal demand. The training phase is launched every 24h with a horizon of 48h. The DRL algorithm runs 200 episodes. This number of episodes per training launch leads to high computational time. A compromise is found between accuracy and computational effort in section [6.2.3](#).

The first section describes the verification of the control bounds. Section [6.2.2](#) confirms the control meets the objectives. Finally, section [6.2.3](#) justifies the decision of the maximum number of episodes needed.

6.2.1 Verification of the Control Bounds

The control should keep the state of charge of the thermal storage between 0 and 1. The desired compressor energies sent to the RBC should stay between $E_{HP,min}$ equals 0 Wh and $E_{HP,max}$, wherein, in the base case, $E_{HP,max}$ equals 3.000 Wh. The compliance to this bounds is ensured throughout the simulations of March and November.

Verification of Energy Limits

Figure [6.1](#) presents the desired energies generated by the DRL control during an exemplary week. The dashed red lines at 0 Wh and 3000 Wh signalize the energy bounds. The energy bounds are ensured by the DRL action limits in the Matlab Script and in the Simulink DRL environment, where the compressor energy is limited by its maximum depending on the tem-

6.2 Verification of the DRL Control

perature conditions. The compliance with the compressor energy bounds is ensured throughout the simulation.

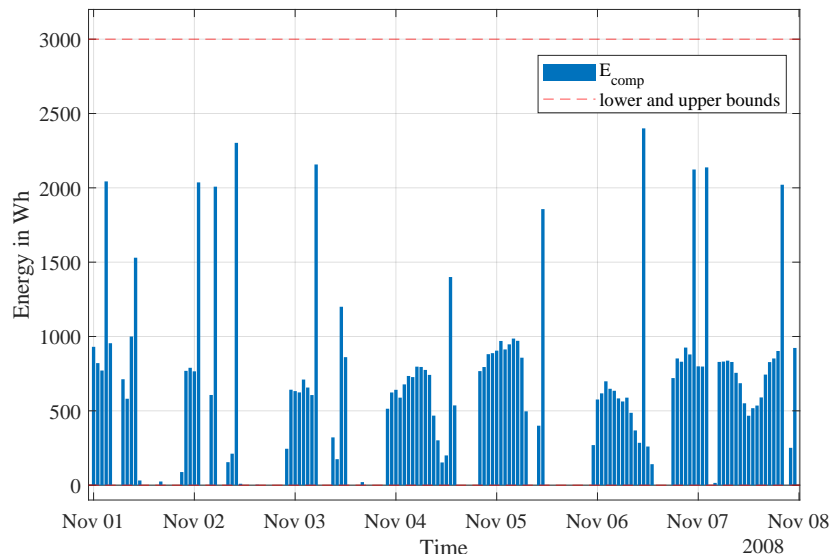


Figure 6.1 – Bound compliance of the desired energies of the heat pump

Verification of State of Charge Limits

Figure 6.2 shows the state of charge trajectory over an exemplary period going from the 2nd to the 5th of November 2008. The thermal storage reaches both its upper and lower thresholds signaled with a dashed red line but does not surpass them. If the upper threshold of the state of charge is reached, the DRL environment constraints the heat pump to input at most enough heat to cover the thermal demand, so the limits of the thermal storage are not exceeded. On the other hand, if the lower limit is reached, the system DRL environment prevents the control from relying on the storage to provide heat or cold. The compliance to the thermal storage bounds is therefore guaranteed.

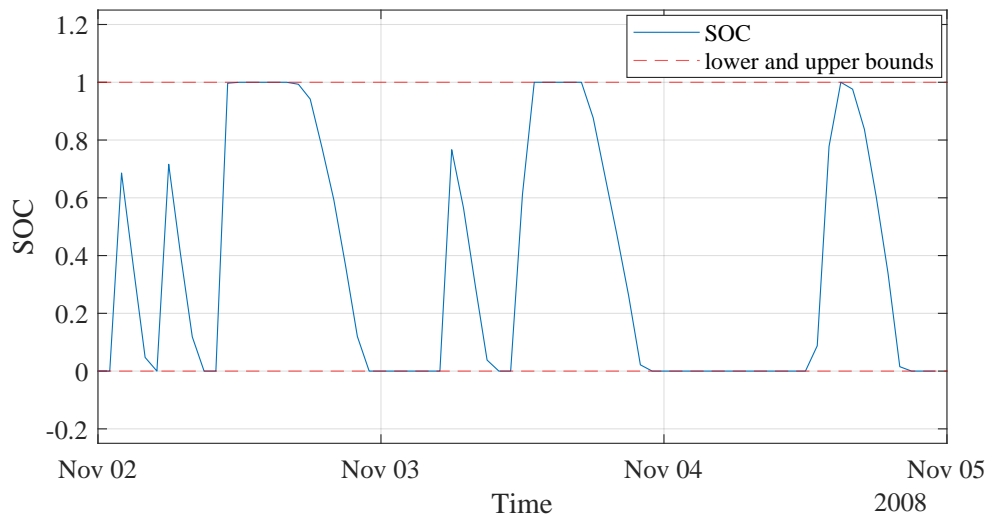


Figure 6.2 – Bound compliance of the state of charge of the thermal storage

6.2.2 Verification of the Control Objectives

The DRL control should ensure thermal and electrical demand coverage while increasing the self-consumption ratio of the system. Figures 6.2, 6.3, and 6.4 show the results of the SOC, thermal and electrical power of the period going from the 2nd to the 5th of November 2008 as an example of the correct functionality of the control. Analogue examples of a period of the month of March are presented in appendix A.

In figures 6.3 and 6.4, the blue lines represent the thermal and electrical energy, respectively, provided and used by the heat pump. The orange dashed line in figure 6.3 represents the building’s thermal demand. The yellow line in 6.4 represent the power coming from PV that is available to the heat pump, after the electric demand coverage, and the orange dashed lines represent the sold to the grid minus the power bought from the grid (if positive, power is sold to the grid, if negative, power is bought from the grid).

In these three figures (6.2, 6.3, and 6.4), both the self-consumption strategy and the demand coverage are clear. From the 2nd to the 5th of November the heat pump is operated during the hours of excess PV power, charging the heat storage completely until the state of charge reaches one, increasing self-consumption. The thermal storage covers the thermal demand until it reaches its minimum state of charge. When the thermal storage is empty, the heat pump takes over in the coverage of the thermal demand and re-charging the thermal storage. The thermal demand is covered therefore at every moment: when it is not covered by the thermal storage (when the thermal storage is not discharging), it is covered by the heat pump and vice-versa.

6.2 Verification of the DRL Control

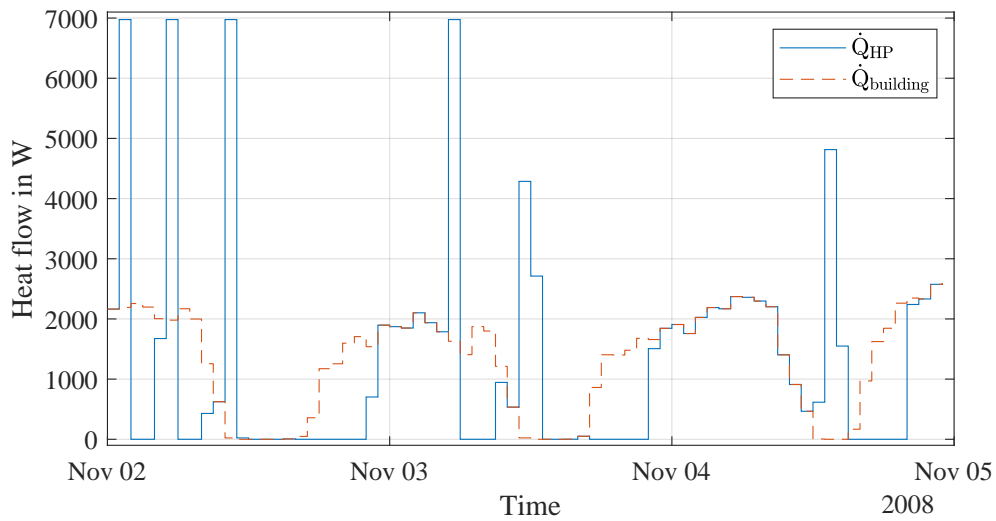


Figure 6.3 – Overview of the building thermal demand and the thermal power provided by the heat pump during the period 2nd to the 5th of November 2008

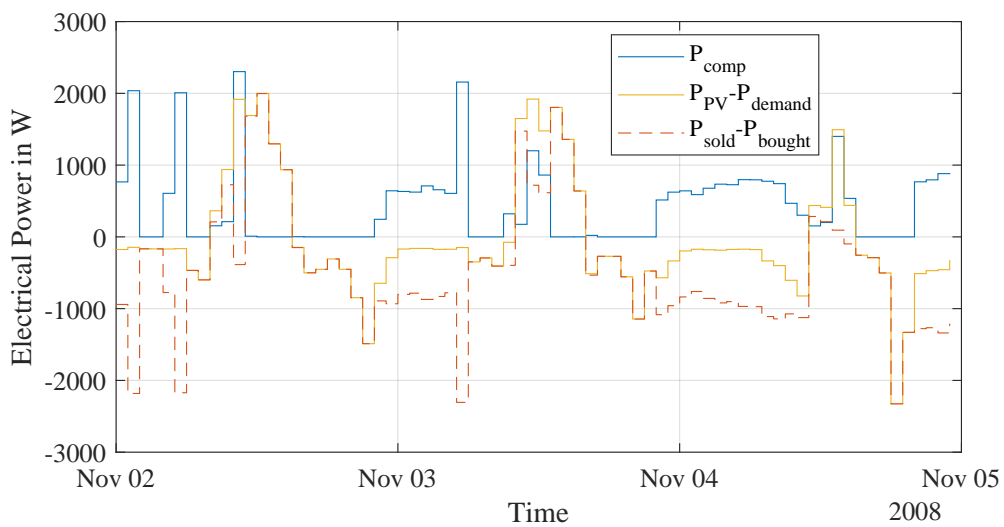


Figure 6.4 – Overview of the available power, the heat pump power and the resulting bought and sold powers during three exemplary november days

With 200 episodes per training phase, the average computational time per optimization launch is 1.124 seconds. The number of episodes was set to 200 to later find a compromise between computational time and accuracy. The details of the trade-off are described in the next section.

6.2.3 Determination of the Number of Episodes

The number of episodes is set in this verification phase to 200. At an average of 5,6 seconds per episode, this leads to high computational time. It was observed that in more than 60% of the cases, the training pattern was similar to the one of figure 6.5 (to see all training patterns refer to appendix B). The figure describes the training progress of the reinforcement learning agent. The blue circled line shows the reward obtained in each episode, the stated red line depicts the heretofore average reward, and the green line with crosses shows the estimate of the discounted long-term reward. Note that the rewards are multiplied by 1.000 to allow for a better visualisation of the training process in the MATLAB Episode Manager.

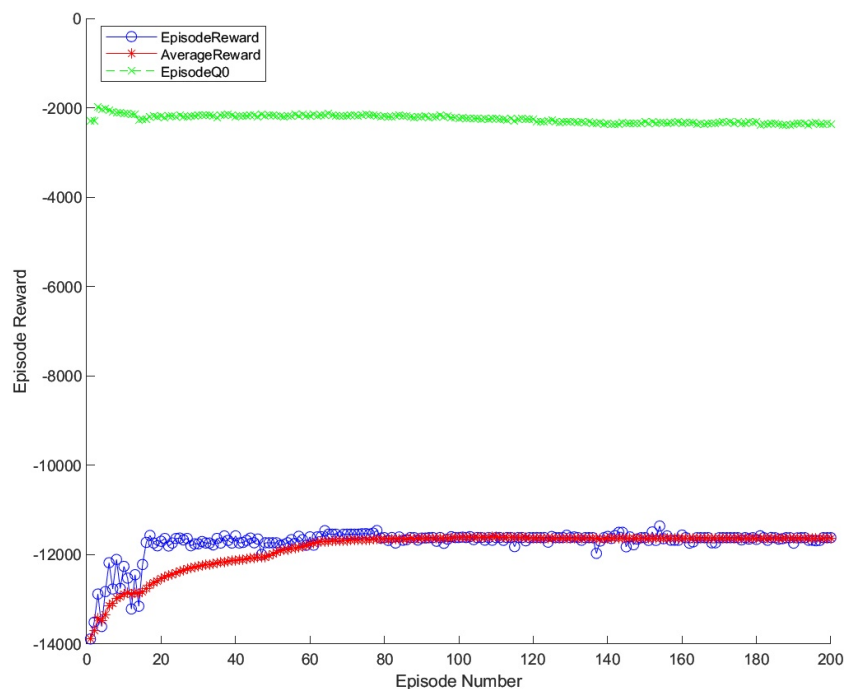


Figure 6.5 – Episode rewards for the 11th of November 2008

As can be seen in figure 6.5, the average reward reaches its maximum value around the 60th episode. Although the maximum reward episode might not have yet been reached (in this example, the maximum reward happens between the episodes 150 and 160), a sufficiently good solution has already been found. Therefore, the number of episodes can be reduced without significantly affecting the results to decrease the computational cost.

To evaluate the effect of reducing the number of episodes to achieve acceptable results and computation time, it was investigated how many episodes were needed to reach either the maximum reward episode or the maximum average reward for every day of March and November allowing a certain increase of the energy cost with respect to the best result achieved in 200

6.3 Verification of the Supervisory Control System

episodes. To exclude outliers the quantile 95 is taken, therefore including 95% of the data [83]. The results of the investigation are shown in figure 6.6.

It can be seen that the relative energy cost increases with the decrease of the maximum number of episodes chosen. As the time needed for each training phase is too high using 200 episodes, a compromise is to be found between accuracy and computation time. As can be seen in figure 6.5 and appendix B, the average reward generally reaches its saturation value after 60 episodes. This means that after 60 episodes an acceptable solution has probably already been found. A training process of 60 episodes corresponds to a 2% relative increase of the energy costs. It is decided that 2% is an acceptable relative loss of accuracy. For the following simulations, this maximum number of episodes is selected as a stopping condition for the learning process.

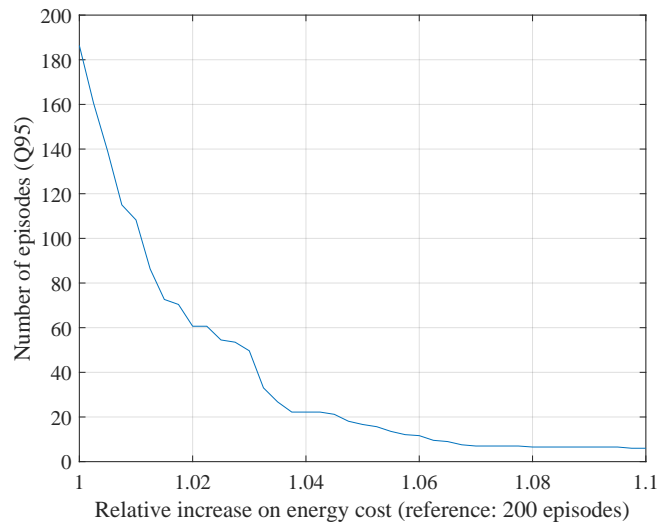


Figure 6.6 – Number of episodes as a function of the relative increase of the energy cost

6.3 Verification of the Supervisory Control System

In the second phase of the verification, the DRL control is integrated into the PVCool Simulink control. The objective of this verification is to confirm the correct coordination between scripts and to check the outputs of the overall supervisory control system. The simulation of the first week of November is run. The supervisory control launches the DRL agent training every hour with a prediction horizon of 48h.

This section only describes the verification of the supervisory control outputs, to avoid repeating information. The self-consumption strategy of the supervisory control is maintained,

showing a similar behaviour to the one presented in figure 6.4.

As described in chapter 5, the DRL controller determines the optimal sequence of energies to be applied by the heat pump during the following 48 hours. The hourly energies are then transformed into a power signal with a resolution of one second for the heat pump in the RBC developed by Rauscher 8. The goal of this subunit of the supervisory control is to match the desired hourly energy E_{des} determined by the DRL control and the energy E_{des} that is applied by the rule-based subordinate controller. Figure 6.7 compares the desired heat pump energy and the energy that is applied by the supervisory control during an exemplary timelapse of the first week of November 2008. Figure 6.8 proves the supervisory control does not exceed the control bounds, keeping the heat pump power within limits.

In figure 6.7, it can be seen that the applied energy generally matches the desired energy determined by the DRL controller. Some differences, particularly between 5 am and 7 am can be observed. These differences result from the fact that the sequence of desired energies is determined using hourly averages as inputs. Differences between the predicted operation range of the heat pump and the actual range of operation in real-time can appear, causing a certain divergence between the applied and the desired energy.

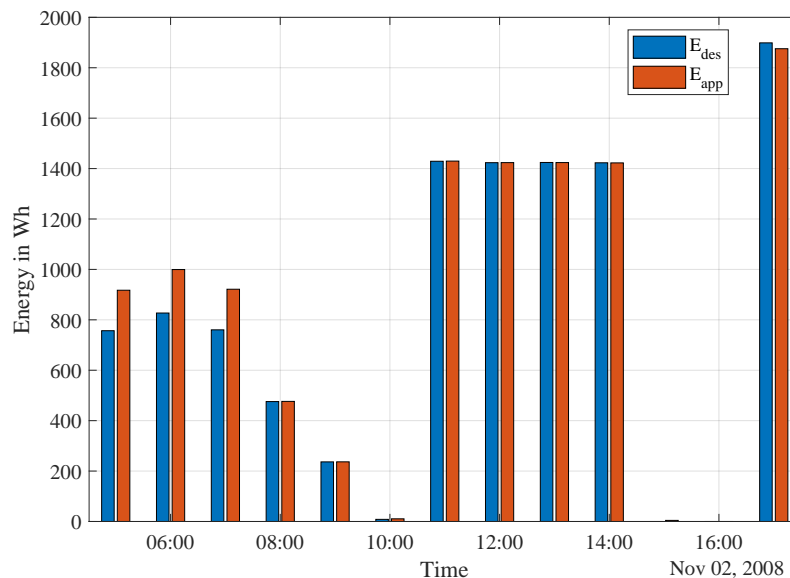


Figure 6.7 – Comparison of the desired hourly energy for the heat pump and the energy actually applied by the supervisory control unit between 5 am and 5 pm of the 2nd of November 2008

6.4 Conclusion

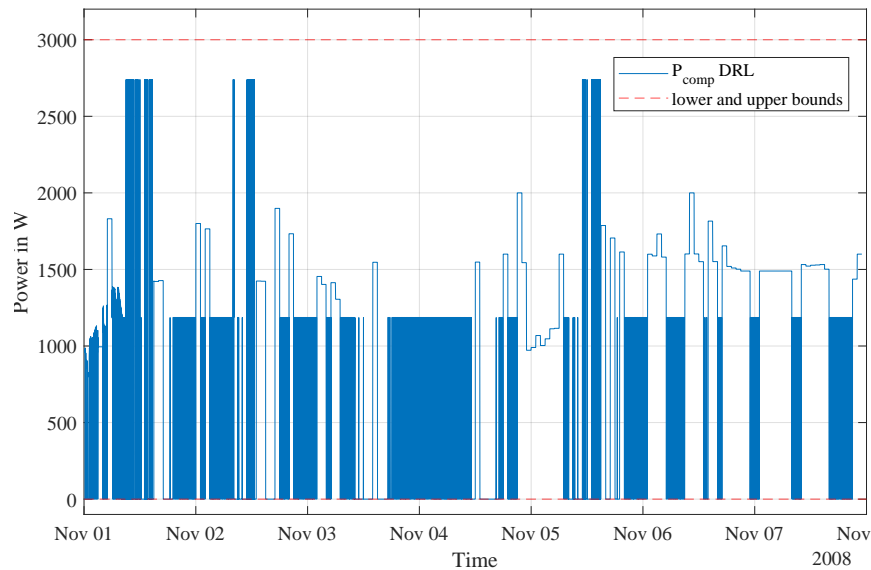


Figure 6.8 – Compliance of the supervisory control unit to heat pump power control bounds

6.4 Conclusion

The verification process undergoes two phases. First, the control is tested before its integration in the PVCool Simulink model. The stopping conditions of the learning process are determined. After its integration, the supervisory control is tested again using the definitive parameters.

In both cases, the compliance to the control bounds has been proven. The energy values proposed by the control do not exceed the heat pump energy limits. Similarly, the state of charge is kept within bounds. The control satisfies its objectives, accounting for demand coverage and operating the heat pump when excess PV is available to increase self-consumption.

The correct communication between scripts has been proven. The subunit of the supervisory control applies the desired energy quantities decided by the DRL control. Small differences between the applied and desired energy can appear since the DRL control uses hourly averages as input and hence does not take into account variations within these hours.

7 Parameter Study and Economic Analysis

This chapter focuses on the development of a parameter study for the PVCool project. The objective of the study is to determine the optimal combination of parameters that achieves the minimal cost of the complete PVCool system. In this study, only the parameters susceptible to have the highest impact on costs are considered. These parameters include:

- The volume of the thermal storage
- The maximum power of the heat pump
- The size of the PV generation unit

To limit the scope of the study, three different sizes of PV, three heat pumps and four thermal storages are tested. Therefore, a total of 36 annual simulations are performed. As the time resources to perform this study are limited, the rule-based control developed by Achatz [9] is chosen to perform these simulations due to its speed. A list of comparative values is then calculated with the result of simulations to determine the optimal system.

Section 7.1 describes the comparative values used to evaluate the results of the parameter study. Section 7.2 gives an overview of the parameters and inputs that are chosen to conduct the study. Finally, the results are presented in section 7.3 and the economically optimal system is determined.

7.1 Comparative Values

To compare the different scenarios, levelized costs are calculated. Levelized costs allow the comparison between energy systems with different capacities, costs and sizes [84]. The levelized cost of energy (LCOEnergy) is used as the main comparative value for the economic evaluation. To obtain the LCOEnergy, the levelized cost of electricity (LCOE) and the levelized cost of heat (LCOH) are calculated.

The objective of the supervisory control is not only to achieve the minimum energy cost but also to maximize self-consumption and reduce emissions. Therefore, the self-consumption

ratio (SCR) and the emissions obtained are additionally calculated as comparative values. In the following sections, the aforementioned comparative values are described in more detail.

7.1.1 Levelized Cost of Electricity

The levelized cost of electricity allows the comparison of energy systems with different generation and cost configurations. It takes into account all investment and operation costs of an energy system as well as the energy generated over its entire life cycle [84]. In this thesis, the net present value (NPV) method is used to calculate levelized costs. With the NPV method, the investment, as well as all costs and revenues occurring over the system's lifetime, are discounted and referred to a shared date. The result of this is then divided by the present value of the electricity yield. The LCOE is calculated as shown in equation (7.1) [84]:

$$LCOE = \frac{I_{PV} + \sum_{t=1}^n \frac{C_t}{(1+i)^t}}{\sum_{t=1}^n \frac{E_t}{(1+i)^t}} \quad (7.1)$$

where I_{PV} refers to the PV investment, C_t to the electricity and PV maintenance costs, and E_t to the electricity consumption of the building. The costs of each year t are discounted at rate i and summed from year one to the economic life of the system n . If the resulting LCOE is lower than the buying price of electricity, the energy system is economically beneficial [85].

7.1.2 Levelized Cost of Heat

The calculation of the levelized cost of heat is similar to the LCOE calculation. To calculate the costs of heat, Angenendt et al. [86] propose the following method: first, the overall costs and revenues of the system $C_{t,OverallSystem}$ are calculated taking into account both the electrical and heating system. Then, the costs and revenues of the system without the heating system $C_{t,noHeatSystem}$ are subtracted. The difference results in the balance of revenues and costs of the heating system. The result of the difference is finally divided by the present value of the total heat production of Q_t . Equation (7.2) illustrates the aforementioned method:

$$LCOH = \frac{I_{HP} + I_{TS} + \sum_{t=1}^n \frac{C_{t,OverallSystem} - C_{t,noHeatSystem}}{(1+i)^t}}{\sum_{t=1}^n \frac{Q_t}{(1+i)^t}} \quad (7.2)$$

where the resulting investment consists of the heat pump I_{HP} and thermal storage I_{TS} and i is the real interest rate.

7.1.3 Levelized Cost of Energy

The levelized cost of energy (LCOEnergy) proposed by Angenendt et al. [86] allows a summarized comparison of heat and electricity costs. The LCOEnergy is defined as the weighted sum of the LCOE and the LCOH. The thermal and electrical demand are taken in this work to define the weights. Equation (7.3) describes the relationship between the three levelized costs [87].

$$LCOEnergy = LCOE \cdot \frac{E_{elec}}{E_{total}} + LCOH \cdot \frac{Q_{heat}}{E_{total}} \quad (7.3)$$

7.1.4 Self-Consumption Ratio

The self-consumption ratio defines the amount of the PV production that is consumed in the building. As explained in section 5.2.3, the DRL algorithm uses the price difference between buying and selling electricity from and to the grid to guide its learning, thereby encouraging self-consumption. The SCR is obtained as shown in equation (7.4) by dividing the amount of electricity that is produced by the PV generator and consumed by the building itself $E_{PV,consumed}$ by the total PV generation $E_{PV,generated}$.

$$SCR = \frac{E_{PV,consumed}}{E_{PV,generated}} \quad (7.4)$$

7.1.5 CO₂ Emissions

The calculation of emissions in this thesis considers only the emissions produced during operation and not during the entire lifecycle of the system elements, therefore neglecting the emissions that arise from the production, transport, and disposal of the components. The two sources powering the building are the PV installed and the electricity bought from the grid. Since PV operation does not generate CO₂ emissions, only the energy bought from the grid is taken into account when calculating the total emissions. As the case of a building in Germany is considered in this study, the CO₂-equivalent of the German electricity mix is taken as a reference to calculate the operation emissions. According to the International Institute for Sustainability Analysis and Strategy [88], the CO₂-equivalent of the country in 2018 was $0,502 \frac{\text{kg}}{\text{kWh}}$. Therefore, the yearly emissions arising from the operation of the building m_{CO_2} are calculated as shown in equation (7.5).

$$m_{CO_2} = E_{bought} \cdot 0,502 \quad (7.5)$$

7.2 Study Parameters and Inputs

This section details the data, parameters, equipment sizes and prices used for the parameter study. First, the input data used to perform the simulations are described. Then, the heat pump, thermal storage, and PV parameters and price structures are presented. Finally, the choices of economic parameters are explained.

7.2.1 Simulation Data

Part of the simulation data are modified in an attempt to consolidate the dataset. The following points summarize the origin of the data.

- **Electric Demand Data:** The electric demand dataset are the same as the data used in chapter 6. It was developed by Tjaden et al. [81] and processed by Möhrstedt [82].
- **Thermal Demand Data:** The thermal demand is obtained using the BST software and was generated by Möhrstedt [82]. The dataset was developed by Kastl [89] and is the same dataset used for verification in chapter 6.
- **Hot Water Demand Data:** The thermal demand is originally extracted from a dissertation from Pflugradt [90]. The data used in this thesis was processed by Möhrstedt [82].
- **PV Data and model:** The radiation data come from the processing of the radiation data of the center of Munich collected by the *Meteorological Institute of Munich* and the data provided by the *Solarenergieförderverein Bayern e.V.* The PV model was developed by Schelze [91].
- **Weather Data:** The data are provided by the *Meteorologisches Institut München* and contain historical values of temperature, wind, humidity, radiation, and pressure of the center of Munich with a resolution of one minute.

The data are available in the time range from the 28th of September 2008 to the 27th of September 2009. They are adapted and processed to be used in the parameter study and the evaluation to perform the simulations. The hourly vectors are calculated and restructured in a similar way as described in chapter 6 to be used with the MPC and DRL controls. The resulting annual data are represented in Appendix C. The only difference is that in this case, the hot water demand is included in the thermal energy demand prediction vector.

7.2.2 Model Parameters

The parameter study aims to find the optimal combination of thermal storage volume, PV plant size, and heat pump maximal power for the PVCool project energy system. Three different sizes for the PV plant and heat pump, and four different thermal storage sizes are considered. In the following, the models and cost structures chosen for each of the elements are detailed.

Heat Pump

In Germany, the investment in heat pump systems can be supported by state subsidies. A list of the commercially available heat pumps that are eligible for such subsidies is published by the German Federal Office of Economics and Export Control [92]. The products of brand Viessmann, an international manufacturer of heating, refrigeration, and industrial systems based in Germany, are included in this list and are selected as product price references in this thesis. More specifically, the air-water heat pump model Vitocal 200-A is selected as a reference [93].

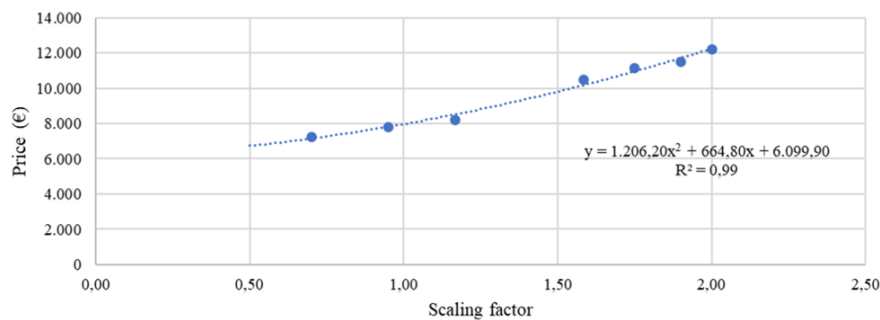


Figure 7.1 – Approximation of heat pump costs

The prices of the Vitocal 200-A are plotted as a function of the heat pump scaling factor. In this thesis, the scaling factor compares the maximal electrical power of the heat pump to the selected reference of 3000 W. Three different scaling factors are considered in the parameter study: 2/3, 1 and 4/3. That corresponds to maximal powers of 2.000 W, 3.000 W and 4.000 W respectively. As these values of maximum power are not exactly available in the Viessmann portfolio, the catalog is only used to give a realist approximation of investment costs. In figure [7.1], a tendency line with a R^2 factor of 0,99 is selected to approximate the prices of the chosen heat pump sizes. The factor R^2 ranges from zero to one and defines the accuracy of a regression model, with one corresponding to a good model fit to the data [83]. The heat pump prices and sizes that are used in the parameter study are summarized in table [?]. The operation and maintenance costs of the heat pump are taken as 2.8% of the investment costs [94]

Table 7.1 – Costs and sizes of the selected heat pumps for the parameter study

Maximum Power [kW]	Scaling Factor	Approximate Price [€]
2.000	2/3	7.079
3.000	1	7.970
4.000	4/3	9.131

and the installation and delivery costs of both the thermal storage and the heat pump add up to 6.500€ [95].

Thermal Storage

The process of finding approximate prices for thermal storage is similar to the procedure described above. The prices of the Vitocell-V are used to approximate the sizes of thermal storage tested. Storage volumes of 250, 475, 750 and 1500 liters are tested in the parameter study. Since not all of these storage capacities are available in the Viessmann portfolio [93], an equation is determined after plotting the storage prices as a function of their volume. Figure 7.2 shows the resulting equation and table 7.2 summarizes the chosen thermal storage volumes and their approximate prices. The costs of operation and maintenance are 1% of the investment costs.

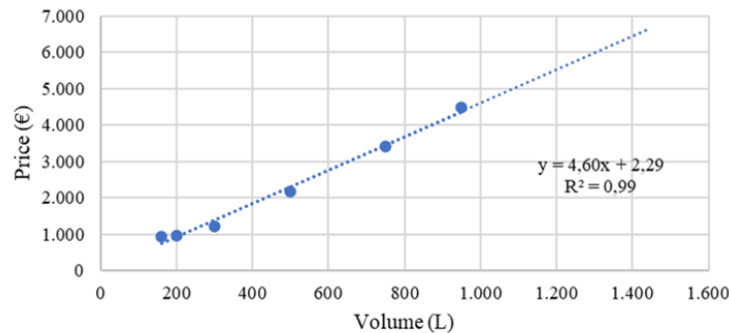


Figure 7.2 – Approximation of thermal storage costs

Table 7.2 – Costs and sizes of the selected thermal storages for the parameter study

Volume [L]	Approximate Price [€]
250	1.151
475	2.185
750	3.449
1.500	6.895

Solar Panels

This work considers the installation of rooftop PV in the investigated building. Three different sizes are investigated: 2,5, 5 and 10 kWp. According to the German Solar Industry Association [96], the price of such a PV system in 2016 was 1.456 €/kWp, including transportation and installation costs. The distribution of costs is shown in figure 7.3. The operation and maintenance costs were taken as 1% of the investment [97].

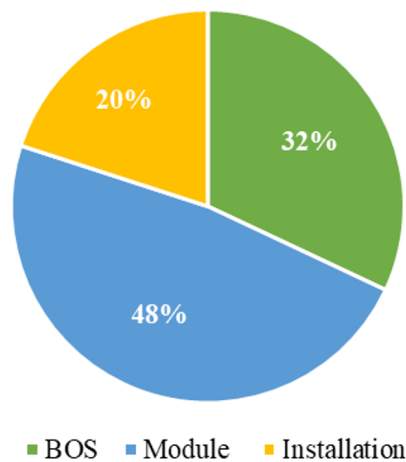


Figure 7.3 – Cost structure of a PV system under 30 kWp [96]

7.2.3 Economic Parameters

As mentioned in the last section, the Federal Office of Economics and Export Control rewards the installation of heat pumps and the associated equipment (including thermal storage) in Germany with attractive subsidies [98]. The subsidy for both existing and new buildings amounts to 35% of the investment costs. The value-added tax (VAT) applied to the purchase of the heat pump, thermal storage, and PV in Germany is 19% [99]. The life-time for this type of project is considered to be 20 years [94] and the applied real interest rate of 3% is taken following the European recommendations [85]. Table 7.3 summarizes the economic parameters that are used in the parameter study.

Table 7.3 – Costs and sizes of the selected thermal storages for the parameter study

Parameter	Unit	Value
Real interest rate		3%
Mean electricity price	$\frac{\text{€}}{\text{kWh}}$	0,30
Feed-in tarif	$\frac{\text{€}}{\text{kWh}}$	0,11
VAT		19%
Subsidies (heat pumps, thermal storage)		35%
Life-time of the project	years	20

7.3 Identification of an Economically Optimal System

To complete the parameter study 36 yearly simulations are run using the control developed by Achatz [9], chosen for its speed. The following annual data is collected or calculated from the simulations:

- Electrical Demand
- Heating Demand
- Heat Pump Electrical Consumption
- Heat Pump Thermal Yield
- Electricity cost without the heating system
- Energy bought from the grid
- Energy sold to the grid
- Self-consumed energy
- PV electricity production
- SCR

The analysis of each case is performed using Excel. An example of the spreadsheet used to calculate the results of each case of study is presented in appendix D. Table 7.4 summarizes the results of the parameter study.

Although not shown in table 7.4 the results of the simulations show that the annual thermal yield of the heat pump in most of the scenarios including thermal storages of 250 L and 475 L

7.3 Identification of an Economically Optimal System

Table 7.4 – Summary of results of the parameter study

Storage (L)	PV (kWp)	Heat Pump (kW)	SCR	CO2 (kg)	LCOE (€/kWh)	LCOH (€/kWh)	LCOEnergy (€/kWh)
250	2,5	2	0,71	3.009	0,27	0,20	0,22
		3	0,67	2.932	0,26	0,19	0,21
		4	0,65	3.248	0,27	0,23	0,24
	5	2	0,50	2.704	0,26	0,19	0,21
		3	0,49	2.659	0,26	0,19	0,21
		4	0,48	2.956	0,26	0,22	0,23
	10	2	0,31	2.454	0,26	0,19	0,20
		3	0,31	2.427	0,26	0,19	0,21
		4	0,30	2.714	0,26	0,21	0,22
475	2,5	2	0,75	3.095	0,27	0,20	0,22
		3	0,72	2.959	0,27	0,20	0,21
		4	0,69	3.301	0,27	0,22	0,23
	5	2	0,56	2.718	0,25	0,19	0,20
		3	0,56	2.634	0,25	0,19	0,20
		4	0,57	2.913	0,26	0,21	0,22
	10	2	0,35	2.419	0,25	0,18	0,20
		3	0,35	2.373	0,25	0,18	0,20
		4	0,30	2.508	0,26	0,22	0,23
750	2,5	2	0,77	3.201	0,27	0,20	0,22
		3	0,75	2.973	0,27	0,19	0,21
		4	0,72	3.378	0,27	0,22	0,23
	5	2	0,60	2.753	0,25	0,19	0,20
		3	0,62	2.611	0,25	0,18	0,20
		4	0,60	2.974	0,25	0,21	0,22
	10	2	0,38	2.397	0,25	0,18	0,19
		3	0,40	2.309	0,24	0,17	0,19
		4	0,41	2.594	0,25	0,20	0,21
1.500	2,5	2	0,75	3.649	0,27	0,23	0,24
		3	0,61	3.215	0,27	0,21	0,23
		4	0,56	3.737	0,28	0,24	0,25
	5	2	0,55	3.282	0,26	0,22	0,23
		3	0,49	2.871	0,26	0,20	0,22
		4	0,47	3.345	0,27	0,23	0,24
	10	2	0,35	2.932	0,26	0,21	0,22
		3	0,34	2.510	0,26	0,19	0,21
		4	0,34	2.897	0,26	0,22	0,23

is lower than the annual thermal demand of the building. This suggests the thermal storage is not big enough to deal with the thermal demand, sometimes failing to cover it. Only the cases with thermal storage of 475 L, PV of 10 kWp, and a heat pump of 3 or 4 kW can cover the full thermal demand. However, it seems unreasonable to install an oversized PV generation unit to compensate for undersized thermal storage. Therefore, it is hereafter considered that the thermal storages of 250 L and 475 L are undersized for this problem and are therefore no longer considered when seeking the optimal combination of parameters. The thermal storage sizes of 750 L and 1.500 L cover the demand completely in all cases. The results show that the LCOH and LCOEnergy prices for the 1.500 L thermal storage are higher compared to the analog cases of the 750 L thermal storage. The most evident reason for this is the doubled investment cost. A further reason for the cost increase is the increase in energy needed to heat a larger volume of water in the storage.

When looking at PV sizes, results show that SCR and LCOE decrease when the PV capacity increases. As a result, for the same heat pump and thermal storage sizes, the LCOEnergy decreases when the PV capacity increases. Figure 7.4 illustrates this phenomenon. The reason for this is that the amount of PV energy that can be self-consumed in the building is not unlimited. Larger capacities lead to a higher amount of energy sold to the grid and higher PV remuneration without a substantial increase in the amount of energy that is consumed on-site. Overall, the installation of a PV is beneficial since the LCOE is in every case lower than the electricity buying price.

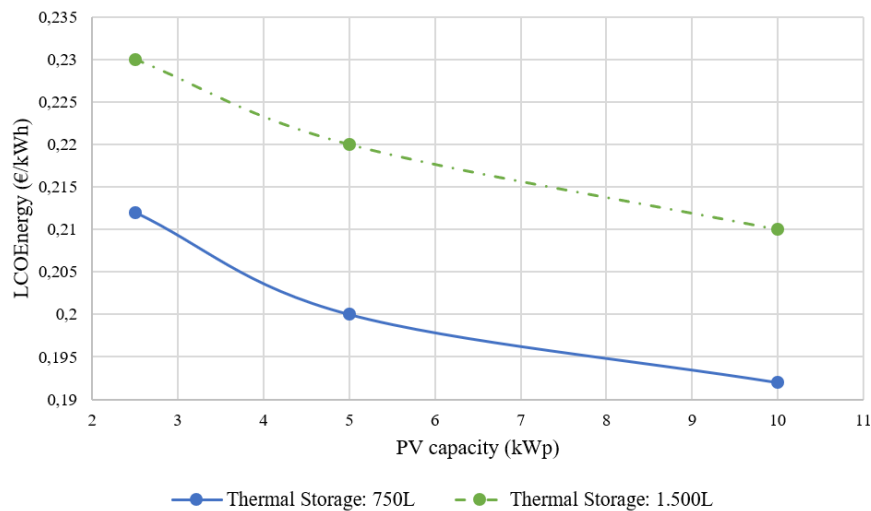


Figure 7.4 – LCOEnergy for a 3kW heat pump and different PV and storage sizes

Considering the different heat pump sizes, if the other parameters are kept constant, the heat pump of 3 kW generally achieves lower LCOH and lower LCOEnergy. This can be seen in figure 7.5. The 2 kW heat pump runs on average at lower average power in comparison to the 3 kW heat pump but runs more hours during the year. On the other hand, the 4 kW heat pump runs fewer hours than the 3 kW heat pump but at higher average power. The 3 kW heat pump

7.3 Identification of an Economically Optimal System

consumes in this context less electricity, achieving lower LCOEnergy values. In the end, for the considered application, the optimal strategy in terms of cost is to install a 3 kW heat pump.

After the analysis of the results of the parameter study, a decision needs to be made regarding the criteria to choose the optimal combination of parameters. The following points summarize the decision criteria:

- The system should be able to cover both thermal and electrical demand of the building, therefore the undersized thermal storages of 250 and 475 liters are not considered eligible.
- The objective of the investigated control systems is to maximize the SCR ratio. To be able to see the effects of the different control systems, a higher value of SCR is desirable. Therefore, 2,5 and 5 kWp installations are preferred for this thesis.
- Among the remaining choices, the lowest LCOEnergy is selected.
- In the case of having two scenarios with the lowest LCOEnergy, the scenario resulting in lower emissions is chosen.

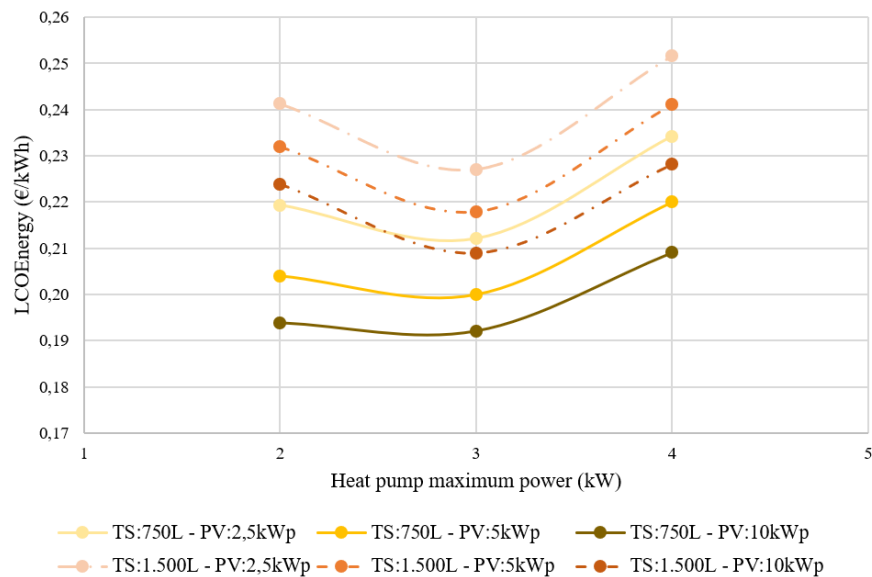


Figure 7.5 – LCOEnergy for different PV and storage sizes - the influence of the heat pump size

The application of the aforementioned criteria leads to the following combination of parameters: a thermal storage of 750 L, a heat pump of 3 kW and a PV capacity of 5 kWp.

7.4 Conclusion

This section investigates the optimal combination of parameters that allows achieving the lower levelized costs of energy. Different sizes of PV, thermal storages and heat pumps have been combined and tested using annual simulations. Further comparative values such as the LCOE, the LCOH, SCR, and the total of emissions have been calculated to guide the choice of the optimal system. The parameters and methods used in the study have been presented. Finally, the results of the study have been summarized. The results show it is cost-optimal in this application to chose a 3 kW heat pump and thermal storage of at least 750L. The costs of electricity decrease with the increase of the PV capacity, however, the SCR decreases as well.

A list of four criteria has been decided to chose the combination of parameters considered optimal. The scenario consisting of a 3 kW heat pump, a 750L thermal storage, and a 5 kWp PV plant is selected. This combination of parameters is used in the following to compare the different control strategies developed within the PVCool project.

8 Evaluation of the Control Approach

This chapter evaluates the performance of the control using long-term simulations. In section 8.1, the monthly and annual outputs are presented and commented. In section 8.2, the DRL control developed in this thesis is compared to the rule-based control programed by Achatz [9] and the MPC designed by Rauscher [8]. The inputs and comparative values used to assess the results of the simulations are presented in chapter 7.

As the simulations have high memory requirement and the available time to perform the simulations is limited, the annual simulations are split in quarters in the case of the rule-based control and in months in the case of the MPC and DRL. The simulations are run in parallel and therefore the state of the last day of the past period cannot be used as a starting point for the simulation of the next period.

8.1 Results of the DRL Long-Term Simulations

This section discusses the results obtained by the simulation of the PVCool model using the DRL control. The results are presented on a monthly basis. Appendix E includes further graphs and results not included in this section. These include the operation hours of the heat pump, the average heat pump power, and the monthly CO₂-equivalent emissions.

8.1.1 Monthly Cost

Figure 8.1 represents the monthly electricity costs of the system using the DRL control developed in this thesis to control the heat pump. It can be seen that electricity costs are higher in the coldest winter months and lower in the summer months. This can be explained by the fact that the amount of heat that the heat pump needs to provide to the building during the winter months is higher. More electricity needs to be bought from the grid, increasing the electricity bill. Additionally, more PV is available during the summer, when the thermal demand is lower, allowing to sell higher amounts of excess PV energy to the grid and therefore reducing the electricity costs. The minimum electricity cost is attained in July, where the amount of energy bought from the grid reaches a minimum and the amount of energy sold to the grid is

8.1 Results of the DRL Long-Term Simulations

at its maximum point (the available PV power for the heat pump is also maximum). January and December have the highest electricity bill: the PV power available is at the lowest while the thermal demand is at its highest point. A similar pattern to the one shown in figure 8.1 is seen in the figures of appendix E. The monthly compressor energy, the monthly operation hours of the compressor, and the monthly emissions are higher during the winter since the heating demand is higher. Minimums are reached during the summer, where the heat pump only needs to provide enough heat to cover the domestic hot water (DHW) demand.

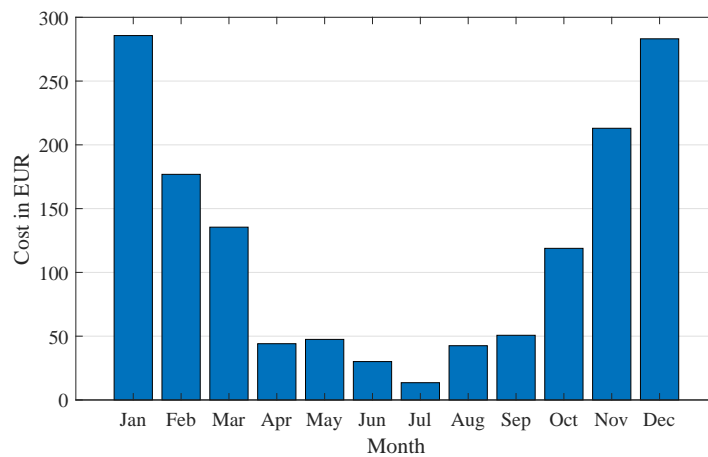


Figure 8.1 – Monthly energy cost obtained when controlling the heat pump with the DRL control

8.1.2 Monthly SCR

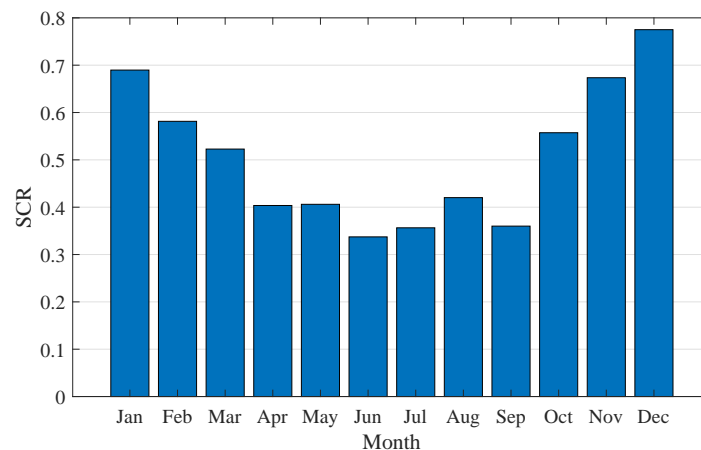


Figure 8.2 – Monthly SCR achieved when controlling the heat pump with the DRL control

The monthly self-consumption ratio is shown in figure 8.2. Similarly to the energy cost, the SCR is higher in winter and lower in summer. The PV production is higher in summer when the demand is lower, so a larger portion of the PV production is sold to the grid. On the other hand, the PV production is lower in winter, when the demand is higher, so a larger portion of the PV production can be consumed on-site, increasing the SCR.

8.1.3 Monthly Average COP

Figure 8.3 represents the monthly average COP of the heat pump when controlled by the DRL supervisory control. The COP is a measure of the efficiency of the heat pump. As shown, the COP is higher in the summer and lower in the winter. This can be explained by the fact that a lower difference between sink and source temperature increases the value of the COP, and vice-versa. The temperature of the thermal storage (heat sink in the investigated case of heating) approximately ranges from 40°C to 55°C. As the temperature is higher in the summer season, the difference between the ambient air (source) and thermal storage (sink) temperature is lower, so the value of the COP increases.

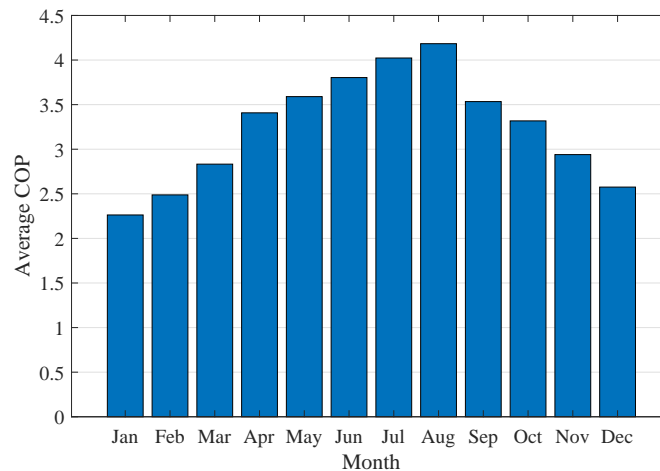


Figure 8.3 – Monthly average COP obtained when controlling the heat pump with the DRL control

8.1.4 Monthly Average SOC

Figure 8.4 shows the monthly averages of the state of charge of the hot thermal storage when controlling the heat pump with the DRL control. It can be seen that the monthly averages are similar year-round. A striking exception is the month of June when the state of charge of the thermal storage sinks considerably. This is because the thermal demand is zero between 13th

8.2 Comparison with Previous Work

to the 19th of June, to simulate the fact that the tenants are on holiday [82]. During that period the state of charge of the storage nears zero, lowering the monthly average. Another exception is July. During this month, the demand is low, and the PV production maximum. In this case, using predictions implies that the DRL agent identifies that the thermal storage might not need to be charged as often to fulfill the demand constraints.

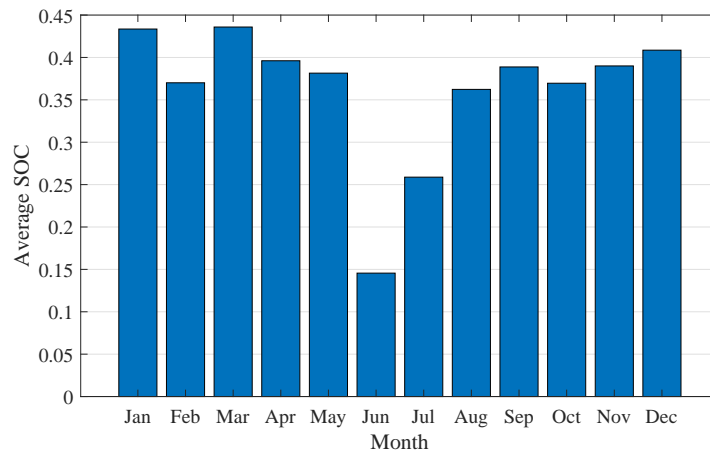


Figure 8.4 – Monthly average SOC of the thermal storage achieved when controlling the heat pump with the DRL control

8.2 Comparison with Previous Work

Different controls have been developed within the PVCool project. The rule-based control developed by Achatz [9] aims at the maximization of self-consumption by systematically using the available PV power on-site when the thermal storage allows it (whenever its state of charge is below one). The rule-based control works with the real-time state of the energy system and does not use weather or demand predictions. On the other hand, the MPC developed by Rauscher [8] and the DRL control developed in this thesis both consider 24 and 48 hours predictions, respectively, to determine the optimal behavior of the heat pump. The MPC uses a pattern search optimization algorithm to determine the trajectory of the heat pump energy yield. This section compares the performance of the three controls using the comparative measures described in section 7.1. First, the monthly results are discussed to understand the different behavior of the controls throughout different seasons. Afterward, the yearly results are examined to give a general overview of the control performances.

8.2.1 Comparison of the Monthly Results

This section summarizes the monthly results of the three different controls. The results of the monthly electricity costs, monthly SCR, and average state of charge of the thermal storage are presented hereafter. Further figures such as the results concerning the heat pump energy, hours of operation of the heat pump and monthly emissions can be found in appendix [F](#).

• Monthly Electricity Cost

The comparison of the monthly results for the electricity costs shown in figure [8.5](#) reveals that the MPC control in most cases achieves the lowest energy cost, and the DRL control the highest cost. Two exceptions to this are the month of August and the month of November. During the month of August, the MPC resulting costs are higher than the DRL and the rule-based control by 17% and 15% respectively. During the month of November, the energy cost obtained using the rule-based control is 1,5% lower than the MPC, achieving the lowest cost of the three investigated controls.

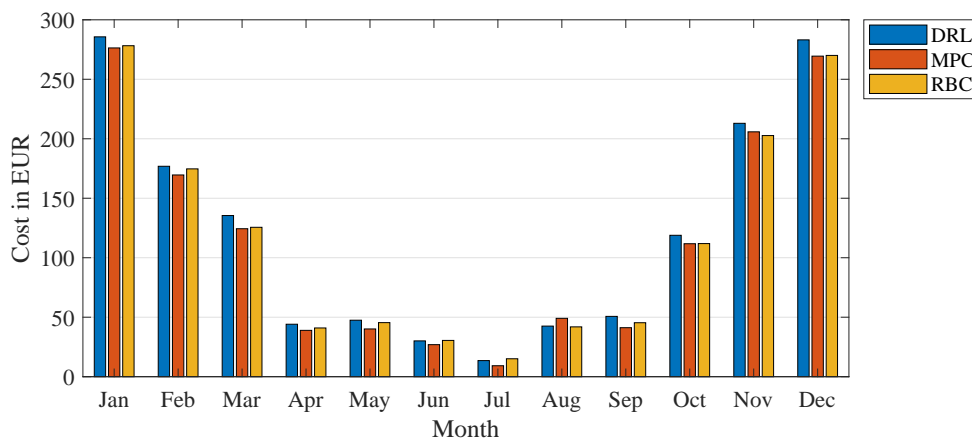


Figure 8.5 – Comparison of the monthly electricity cost when controlling the heat pump with the DRL, MPC and rule-based controls

• Average COP

Figure [8.6](#) summarizes the monthly results regarding the average COP achieved when using the three different controls. It can be seen that the three controls achieve similar average COPs in winter. For instance, in January, the differences between the average COP achieved by the rule-based control and MPC control in comparison to the DRL control are of 2% and 3%, respectively. In summer, however, the rule-based control leads to a substantially lower average COP. For example in the month of July, the COP achieved by the rule-based control is 16,7% and 18,6% lower than the values obtained by the DRL and MPC, respectively. This fact explains the higher heat pump energy use when using the rule-based control in summer that can be seen in appendix [F](#).

8.2 Comparison with Previous Work

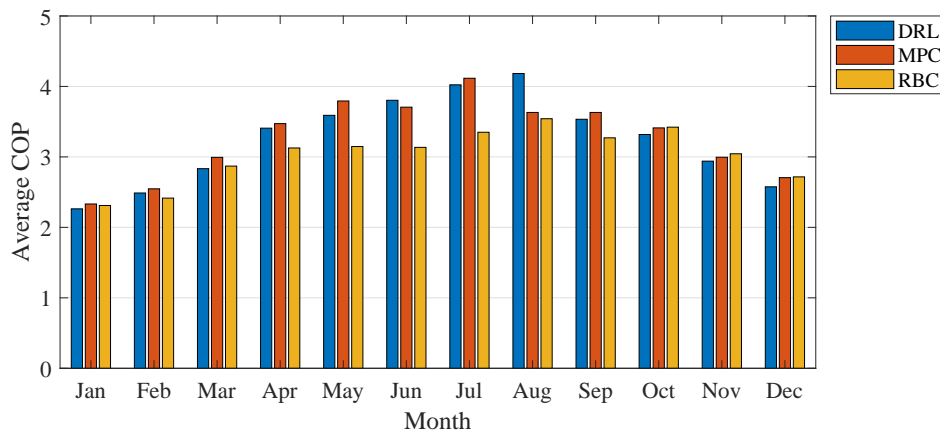


Figure 8.6 – Comparison of the monthly average COP when controlling the heat pump with the DRL, MPC and rule-based controls

- **Monthly SCR**

Figure 8.7 reflects the strategy difference between the rule-based control and the two predictive controls. The objective of the rule-based control is oriented to maximize real-time self-consumption, whereas the MPC and DRL controls aim at the minimization of the long-term energy costs. The SCR of the system when using the rule-based control is substantially higher than in the case of two other controls, achieving an increase in self-consumption that ranges from 11% to 22% in comparison to the MPC. The MPC achieves better results in terms of self-consumption than the DRL control, achieving a relative increase in the SCR going from 0% to 12%.

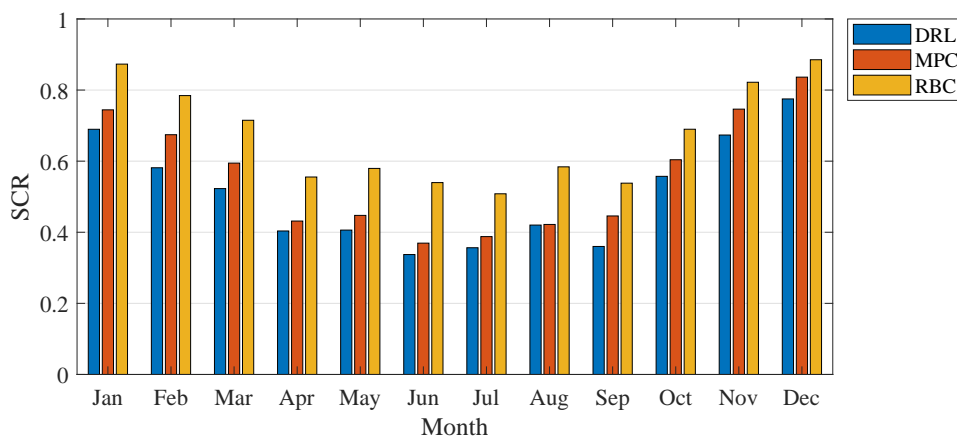


Figure 8.7 – Comparison of the monthly SCR when controlling the heat pump with the DRL, MPC and rule-based controls

- **Average SOC**

Figure 8.8 represents the monthly average state of charge of the thermal storage using the different controls. Again, a substantial difference can be seen between the real-time rule-based control and the two other controls. When comparing the SOC of the rule-based control to the two predictive controls, it can be seen that the SOC is lower in winter and considerably higher in summer. For instance, the SOC in the case of the rule-based control is 43% lower in January and 95% higher in July than the SOC in the case of the MPC. The behavior of the MPC and the DRL control in terms of the state of charge of the thermal storage is similar. The difference between the rule-based control and the other two controls arises from the fact that the rule-based control systematically charges the thermal storage when excess PV is available, independently from future thermal demand or future weather conditions. This way, in the months where the excess PV is higher (in summer, for instance), the heat pump charges the thermal storage to a higher extent, although the thermal demand is low. On the other hand, the DRL and MPC controls receive future predictions and can charge the thermal storage to meet the needs and sell a higher amount of energy to the grid.

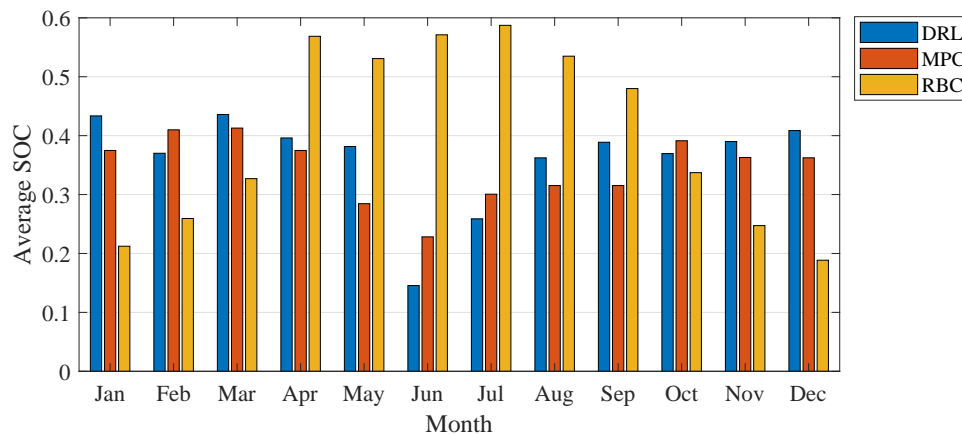


Figure 8.8 – Comparison of the monthly average SOC when controlling the heat pump with the DRL, MPC and rule-based controls

8.2.2 Comparison of the Annual Results

Table 8.1 summarizes the annual results obtained when controlling the heat pump with the three different controls.

As can be seen in table 8.1, the lowest annual cost is achieved using the MPC control. The electricity cost is 2,2% and 6,7% higher when using the rule-based control and DRL control, respectively. However, the rule-based control achieves the highest self-consumption ratio and

8.3 Conclusion

Table 8.1 – Overview of the results of the annual simulations with the three investigated controls

Category	DRL	MPC	RBC
Annual electricity cost	1.442 €	1.352 €	1.382 €
Annual electricity demand for compressor operation	3.783 kWh	3.618 kWh	4.035 kWh
Annual runtime of the heat pump	2.246 h	2.308 h	3.207 h
Average COP of the heat pump	2,94	3,04	2,85
Average self-consumption ratio	0,45	0,50	0,62
Average state of charge of the thermal storage	0,36	0,35	0,40
Average CO ₂ equivalent emissions	2.842 kg	2.652 kg	2.611 kg

lowest emissions. The SCR achieved when using the rule-based control is 24% and 37,8% higher than the SCR obtained by the MPC and DRL, respectively. The rule-based control also leads to higher operation hours and average heat pump electricity consumption, being 39% and 11,5% higher than the values obtained with the MPC, respectively. This means that the economic life of the heat pump using the MPC control is likely to be longer than the technical lifetime of the heat pump when using the rule-based control.

Economically speaking, the MPC control is therefore preferred. From an environmental point of view, however, the rule-based control should be chosen. The DRL control achieves generally the worst results of the three controls. When comparing the two predictive controls, the pattern search algorithm used by the MPC leads to better optimization results. The DRL algorithm can lead to suboptimal optimization results. Two reasons for this are the reduced number of training episodes established and the fact that the learning algorithm can get "stuck" in a suboptimal solution (see training profiles in appendix B). Additionally, some optimization accuracy has been traded for computational speed in section 6.2.3.

As the simulations of the DRL and MPC controls are performed in parallel and different computers, a study of the computation time is not possible. However, it is observed that the rule-based control is considerably faster than the DRL and MPC control. Whereas the simulation of one month takes between one and two hours for the rule-based control, the simulation of one month with the predictive controls takes approximately three days. The MPC is slower than the DRL control, taking on average more time to complete one month of simulation.

8.3 Conclusion

The monthly results obtained by the simulation of the model when using the DRL control have been presented and discussed. The DRL control is compared to the controls developed by Achatz [9] and Rauscher [8]. Monthly and annual results of the three control are presented.

Results show that the MPC performs better economically speaking, achieving the lowest an-

nual electricity cost. From an environmental point of view, the rule-based control achieves the lowest annual operation emissions of the three investigated controls. The suboptimal results obtained by the DRL algorithm can be explained by the reduced number of training episodes and the fact that the learning algorithm can get caught in a suboptimal result.

9 Conclusion and Outlook

This chapter concludes this work. Section 9.1 outlines the content and findings of this thesis. An outlook on the possible enhancements and future steps is provided in section 9.2.

9.1 Summary of the Thesis

The share of renewable energy generation in Germany is increasing, while the contribution of renewables for heating and cooling is still low [100]. However, the building sector plays an important role in decarbonization [3]. The development of concepts involving sector-coupling strategies and renewable energy technologies in the building sector is, therefore, important to reduce greenhouse emissions. The use of intelligent controls has been encouraged by the Building Performance Institute Europe to manage such energy systems.

The purpose of this thesis is the development of a supervisory control based on machine learning for PV-driven heat pumps and compression chillers. Firstly, the basic concepts of the cold vapor cycles driving heat pumps and compression chillers are introduced. Afterward, the relevant branches of machine learning and their application in control theory are presented.

From a literature review including controls based on reinforcement learning, evolutionary algorithms, artificial neural networks, and combinations of the latter methods, it is concluded that hybrid strategies are the most promising machine learning strategy for the PVCool application. Different hybrid strategies are considered. The combination of artificial neural network and reinforcement learning in what is commonly called deep reinforcement learning is identified as the best candidate for controlling the investigated building energy system.

The deep reinforcement learning control is developed using a bottom-up approach, starting with the basic features and progressively building to a more complex system. The control is based on the supervisory control architecture developed by Rauscher [8]. The deep reinforcement learning control uses 48h predictions and a simplified model of the building energy system to learn the optimal hourly energy trajectory of the heat pump. The predictive model is modified to include the optimization constraints and adapted for both heating and cooling applications. The hourly values proposed by the deep reinforcement learning algorithm are then transformed in signals with a temporal resolution of one second by a subordinate rule-

based controller. The correct functionality and behavior of the deep reinforcement learning control are verified through short-term simulations.

Then, a parameter study is conducted to find the economically optimal combination of system parameters. The study consists of the performance of 36 annual simulations, each of them testing a different combination of parameters. The simulations are conducted using the rule-based control developed by Achatz [9], chosen for its speed. Levelized costs, self-consumption ratio and CO₂ emissions are calculated to determine the optimal scenario.

The same combination of parameters is used to examine the performance of the controls developed within the PVCool project. The rule-based control developed by Achatz [9], the model predictive control designed by Rauscher [8] and the deep reinforcement learning control developed in this thesis are compared on a monthly and annual basis. Results show that the model predictive control achieves an annual electricity cost that is 6,4% and 2,2% lower than the annual cost obtained by the deep reinforcement learning and rule-based control, respectively. The best self-consumption ratio and the lower emissions are however achieved by the rule-based control. When using the rule-based control, the self-consumption ratio is 24% higher than the self-consumption ratio obtained with the model predictive control control, and the emissions 1,5% lower.

9.2 Outlook

There are different alternatives to enhance the performance of the deep reinforcement learning control. First, the control could potentially be improved by optimization of its learning parameters and a further investigation of the different deep reinforcement learning algorithms. Mocanu et al. [70] test both deep deterministic policy gradient and deep Q-network algorithms and observe a better performance of the deep deterministic policy gradient algorithm. In the tests performed in this work, however, the heat pump energies proposed by the deep deterministic policy gradient near the lower bound (0 W). The deep deterministic policy gradient algorithm could be adapted to suit the PVCool application. Additionally, the optimal architecture of the critic neural network could be investigated. Then, the predictive model could be improved, for example, by the modelling of a back-up heater or the storage losses.

The deep reinforcement learning control is adapted to deal with heating and cooling applications. However, in this thesis, only its heating application in Germany is investigated. The reward function and the parameter study should be adapted to other applications, weather conditions, and socio-economic contexts. Further possibilities of reward functions that could be more suitable for other applications are the minimization of the CO₂ emissions or the maximization of self-consumption.

The monthly prediction inputs for the predictive controls can be adapted to include at the

end of the monthly data two additional prediction days corresponding to the beginning of the next month to ensure the correct determination of the heat pump hourly energy for the last day of the investigated month. An extra day is already added to the model predictive control prediction data, without which the last day of each month could not have been simulated.

The different controls developed within the PVCool projects have been evaluated using annual simulations. The lowest annual costs are achieved by the model predictive control, showing the benefit of using prediction inputs. The highest self-consumption ratio is achieved by the rule-based control. However, this control also leads to higher operation hours, which could lead to a shorter technical lifetime. Both controls outperform the deep reinforcement learning control. Its performance could be enhanced by increasing the number of training episodes or optimizing the training parameters.

The control to be applied depends on the goal to be achieved (i.e. self-consumption or costs) and the possibility of obtaining reliable demand and weather prediction inputs. It is worth mentioning that the predictions used in this thesis are 100% correct and the difference between the electricity costs obtained by the rule-based and model predictive controls is relatively low (2,2%). Therefore, the rule-based control will likely achieve the lower annual energy costs in real applications, considering that 100% correct predictions will certainly not be available in reality. For this reason, and because of its speed and simplicity, the rule-based control might be the best solution.

Finally, the performance of the controls could be validated with experiments in the test rig.

Bibliography

- [1] ACTIVESUSTAINABILITY.COM: *Robert Swan and our planet*. <https://www.activesustainability.com/environment/robert-swan-and-our-planet/>.
Version: 2020
- [2] ENERGY - EUROPEAN COMMISSION: *Energy performance of buildings directive - European Commission*. <https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-performance-of-buildings/energy-performance-buildings-directive>. Version: 2020
- [3] AHMAD, Muhammad W. ; MOURSHED, Monjur ; YUCE, Baris ; REZGUI, Yacine: Computational intelligence techniques for HVAC systems: A review. In: *Building Simulation 9* (2016), Nr. 4, S. 359–398. <http://dx.doi.org/10.1007/s12273-016-0285-4>. – DOI 10.1007/s12273–016–0285–4. – ISSN 1996–3599
- [4] H. Ü. YILMAZ, R. HARTEL, D. KELLES, R. MCKENNA AND W. FICHTNER: *Analysis of the potential for Power-to-Heat/Cool applications to increase flexibility in the European electricity system until 2030* (Policy Report)
- [5] A. BLOESS, W.-P. SCHILL AND A. ZERRAHN: Power-to-heat for renewable energy integration: A review of technologies. In: *Applied Energy* 212 (2018), S. 1611–1626. <http://dx.doi.org/10.1016/j.apenergy.2017.12.073>. – DOI 10.1016/j.apenergy.2017.12.073.
- [6] BUNDESVERBAND WÄRMEPUMPE (BWP) E.V.: *Sektorkopplung*. <https://www.waermepumpe.de/politik/sectorkopplung/>. Version: 2017
- [7] P. SHAIKH, N. NOR, P. NALLAGOWNDEN, I. ELAMVAZUTHI AND T. IBRAHIM: A review on optimized control systems for building energy and comfort management of smart sustainable buildings. In: *Renewable and Sustainable Energy Reviews* 34 (2014), S. 409–429. <http://dx.doi.org/10.1016/j.rser.2014.03.027>. – DOI 10.1016/j.rser.2014.03.027.
- [8] B. RAUSCHER: *Development and Implementation of Control Strategies for PV-Driven Heat Pumps and Compression Chillers*, Technische Universität München, Diss., 2019

- [9] H. ACHATZ: *Simulations on Covering the Heating and Cooling Demand of Buildings in Temperate and Desert Climate with PV Driven Heat Pumps and Compression Chillers*. Munich, Technical University of Munich, Bachelor's thesis, 2019
- [10] PERRY, Robert H. ; GREEN, Don W.: *Perry's chemical engineers' handbook*. 8th ed. / prepared by a staff of specialists under the editorial direction of editor-in-chief, Don W. Green, late editor, Robert H. Perry. New York : McGraw-Hill Professional and London : McGraw-Hill [distributor], 2008 <http://www.loc.gov/catdir/enhancements/fy0729/2007026072-d.html>. – ISBN 0071422943
- [11] MORAN, Michael J. ; SHAPIRO, Howard N. ; BOETTNER, Daisie D. ; BAILEY, Margaret B.: *Moran's principles of engineering thermodynamics*. Global edition. Singapore : Wiley, 2017. – ISBN 9781119454069
- [12] REISEL, John: *Principles of engineering thermodynamics, si edition*. Cengage Learning, Inc, 2015. – ISBN 1285056485
- [13] PAL, Animesh ; UDDIN, Kutub ; THU, Kyaw ; SAHA, Bidyut B.: Environmental Assessment and Characteristics of Next Generation Refrigerants. In: *Evergreen* 5 (2018), Nr. 2, S. 58–66. <http://dx.doi.org/10.5109/1936218>. – DOI 10.5109/1936218. – ISSN 2189–0420
- [14] STRUCHTRUP, Henning: *Thermodynamics and energy conversion*. Berlin : Springer, 2014. – ISBN 9783662437155
- [15] W. YU, B. LI, H. JIA, M. ZHANG AND D. WANG: Application of multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design. In: *Energy and Buildings* 88 (2015), S. 135–143. <http://dx.doi.org/10.1016/j.enbuild.2014.11.063>. – DOI 10.1016/j.enbuild.2014.11.063.
- [16] LIU, Simeng ; HENZE, Gregor P.: Evaluation of Reinforcement Learning for Optimal Control of Building Active and Passive Thermal Storage Inventory. In: *Journal of Solar Energy Engineering* 129 (2006), Nr. 2, S. 215–225. <http://dx.doi.org/10.1115/1.2710491>. – DOI 10.1115/1.2710491. – ISSN 1532–4435
- [17] BEGHI, Alessandro ; RAMPAZZO, Mirco ; ZORZI, Stefano: Reinforcement Learning Control of Transcritical Carbon Dioxide Supermarket Refrigeration Systems. In: *IFAC-PapersOnLine* 50 (2017), Nr. 1, S. 13754–13759. <http://dx.doi.org/10.1016/j.ifacol.2017.08.2565>. – DOI 10.1016/j.ifacol.2017.08.2565. – ISSN 24058963
- [18] MITCHELL, Tom M.: *Machine learning*. International ed. New York, NY : McGraw-Hill, 1997 (McGraw-Hill series in computer science). – ISBN 0071154671
- [19] REBALA, Gopinath ; RAVI, Ajay ; CHURIWALA, Sanjay: *An Introduction to Machine Learning*. Cham : Springer International Publishing and Springer, 2019. – ISBN 3030157296

BIBLIOGRAPHY

- [20] MARSLAND, Stephen: *Machine learning: An algorithmic perspective*. Second edition. Boca Raton, Fla. : CRC Press/Taylor & Francis, 2015 (A Chapman & Hall book). – ISBN 9781466583283
- [21] OLADIPUPO, Taiwo: Types of Machine Learning Algorithms. Version: 2010. <http://dx.doi.org/10.5772/9385>. In: ZHANG, Yagang (Hrsg.): *Ant Colony Optimization*. INTECH Open Access Publisher, 2010. – DOI 10.5772/9385. – ISBN 978-953-307-034-6, S. 19–48
- [22] MOUBAYED, Abdallah ; INJADAT, Mohammadnoor ; NASSIF, Ali B. ; LUTFIYYA, Hanan ; SHAMI, Abdallah: E-Learning: Challenges and Research Opportunities Using Machine Learning & Data Analytics. In: *IEEE Access* 6 (2018), S. 39117–39138. <http://dx.doi.org/10.1109/ACCESS.2018.2851790>. – DOI 10.1109/ACCESS.2018.2851790
- [23] GUPTA, Manish ; GANAPATHI, Padmavathi ; SHANMUGAPRIYA, D.: *Handbook of Research on Machine and Deep Learning Applications for Cyber Security*. IGI Global, 2019. <http://dx.doi.org/10.4018/978-1-5225-9611-0>. <http://dx.doi.org/10.4018/978-1-5225-9611-0>. – ISBN 9781522596110
- [24] DIETTERICH, T. ; & KONG, E.: *Machine learning bias, statistical bias, and statistical variance of decision*. 1995
- [25] HASTIE, Trevor ; TIBSHIRANI, Robert ; FRIEDMAN, Jerome H.: *The elements of statistical learning: Data mining, inference, and prediction*. 2. ed. New York, NY : Springer, 2009 (Springer series in statistics). <http://dx.doi.org/10.1007/b94608>. <http://dx.doi.org/10.1007/b94608>. – ISBN 9780387848570
- [26] CHANG, Ni-Bin ; BAI, Kaixu: *Multisensor data fusion and machine learning for environmental remote sensing*. Boca Raton, FL and London and New York NY : CRC Press Taylor & Francis Group, 2018. – ISBN 9781498774338
- [27] WOLPERT, David H.: The Lack of A Priori Distinctions Between Learning Algorithms. In: *Neural Computation* 8 (1996), Nr. 7, S. 1341–1390. <http://dx.doi.org/10.1162/neco.1996.8.7.1341>. – DOI 10.1162/neco.1996.8.7.1341. – ISSN 0899-7667
- [28] MARK, Talabis (Hrsg.) ; KAYE, D. (Hrsg.): *Information security analytics: Finding security insights, patterns and anomalies in big data*. Waltham, Massachusetts : Syngress, 2015. – ISBN 9780128002070
- [29] KALOGIROU, Soteris A.: Artificial neural networks in renewable energy systems applications: a review. In: *Renewable and Sustainable Energy Reviews* 5 (2001), Nr. 4, S. 373–401. [http://dx.doi.org/10.1016/s1364-0321\(01\)00006-5](http://dx.doi.org/10.1016/s1364-0321(01)00006-5). – DOI 10.1016/s1364-0321(01)00006-5. – ISSN 13640321
- [30] KARABACAK, Kerim ; CETIN, Numan: Artificial neural networks for controlling wind-PV power systems: A review. In: *Renewable and Sustainable Energy Reviews* 29

- (2014), S. 804–827. <http://dx.doi.org/10.1016/j.rser.2013.08.070>. – DOI 10.1016/j.rser.2013.08.070. – ISSN 13640321
- [31] HAYKIN, Simon S.: *Neural networks: A comprehensive foundation / Simon Haykin*. New York : Macmillan and London : Maxwell Macmillan International, 1994. – ISBN 0023527617
- [32] MOHANRAJ, M. ; JAYARAJ, S. ; MURALEEDHARAN, C.: Applications of artificial neural networks for refrigeration, air-conditioning and heat pump systems—A review. In: *Renewable and Sustainable Energy Reviews* 16 (2012), Nr. 2, S. 1340–1358. <http://dx.doi.org/10.1016/j.rser.2011.10.015>. – DOI 10.1016/j.rser.2011.10.015. – ISSN 13640321
- [33] AL-ALAWI, Ali ; M AL-ALAWI, Saleh ; M ISLAM, Syed: Predictive control of an integrated PV-diesel water and power supply system using an artificial neural network. In: *Renewable Energy* 32 (2007), Nr. 8, S. 1426–1439. <http://dx.doi.org/10.1016/j.renene.2006.05.003>. – DOI 10.1016/j.renene.2006.05.003. – ISSN 09601481
- [34] MOHAMMADZAHERI, Morteza ; CHEN, Lei ; GRAINGER, Steven: A critical review of the most popular types of neuro control. In: *Asian Journal of Control* 14 (2012), Nr. 1, S. 1–11. <http://dx.doi.org/10.1002/asjc.449>. – DOI 10.1002/asjc.449. – ISSN 15618625
- [35] THE MATHWORKS, Inc.: *Reinforcement Learning with MATLAB*. <https://www.mathworks.com/campaigns/offers/reinforcement-learning-with-matlab-intro-ebook.html>. Version: 2019
- [36] M. HAN, X. ZHANG, L. XU, R. MAY, S. PAN AND J. WU: A review of reinforcement learning methodologies on control systems for building energy. In: *Working papers in transport, tourism*, 2 (2018)
- [37] WATKINS, C. ; DAYAN, P: Q-Learning. In: *Machine Learning* 8 (1992), S. 279–292
- [38] WEN, Zheng ; O’NEILL, Daniel ; MAEI, Hamid: Optimal Demand Response Using Device-Based Reinforcement Learning. In: *IEEE Transactions on Smart Grid* 6 (2015), Nr. 5, S. 2312–2324. <http://dx.doi.org/10.1109/TSG.2015.2396993>. – DOI 10.1109/TSG.2015.2396993. – ISSN 1949–3053
- [39] AHN, Ki U. ; PARK, Cheol S.: Application of deep Q-networks for model-free optimal control balancing between different HVAC systems. In: *Science and Technology for the Built Environment* 116 (2019), Nr. 10, S. 1–14. <http://dx.doi.org/10.1080/23744731.2019.1680234>. – DOI 10.1080/23744731.2019.1680234. – ISSN 2374–4731
- [40] MATHWORKS.COM: *Reinforcement Learning Agents- MATLAB & Simulink*. <https://www.mathworks.com/help/reinforcement-learning/ug/create-agents-for-reinforcement-learning.html>. Version: 2019

BIBLIOGRAPHY

- [41] CORNE, David ; LONES, Michael A.: Evolutionary Algorithms. Version: 2018. http://dx.doi.org/10.1007/978-3-319-07153-4_{_}27-1. In: MARTÍ, Rafael (Hrsg.) ; PANOS, Pardalos (Hrsg.) ; RESENDE, Mauricio G. C. (Hrsg.): *Handbook of Heuristics* Bd. 1. [S.l.] : Springer International Publishing, 2018. – DOI 10.1007/978-3-319-07153-4_27-1. – ISBN 978-3-319-07153-4, S. 1–22
- [42] ZHOU, Zhi-Hua. ; YU, Yang. ; QIAN, Chao.: *Evolutionary Learning: Advances in Theories and Algorithms*. Singapore : Springer Singapore and Imprint and Springer, 2019. – ISBN 9811359563
- [43] CALDAS, L. G. ; NORFORD, L. K.: Genetic Algorithms for Optimization of Building Envelopes and the Design and Control of HVAC Systems. In: *Journal of Solar Energy Engineering* 125 (2003), Nr. 3, S. 343–351. <http://dx.doi.org/10.1115/1.1591803>. – DOI 10.1115/1.1591803. – ISSN 1532–4435
- [44] HUANG, W. ; LAM, H. N.: Using genetic algorithms to optimize controller parameters for HVAC systems. In: *Energy and Buildings* 26 (1997), Nr. 3, S. 277–282. [http://dx.doi.org/10.1016/S0378-7788\(97\)00008-X](http://dx.doi.org/10.1016/S0378-7788(97)00008-X). – DOI 10.1016/S0378-7788(97)00008-X
- [45] CORRIOU, Jean-Pierre: *Process Control: Theory and Applications*. London and s.l. : Springer London, 2004. <http://dx.doi.org/10.1007/978-1-4471-3848-8>. <http://dx.doi.org/10.1007/978-1-4471-3848-8>. – ISBN 9781849969116
- [46] WANG, Shengwei ; MA, Zhenjun: Supervisory and Optimal Control of Building HVAC Systems: A Review. In: *HVAC&R Research* 14 (2008), Nr. 1, S. 3–32. <http://dx.doi.org/10.1080/10789669.2008.10390991>. – DOI 10.1080/10789669.2008.10390991. – ISSN 1078–9669
- [47] AFRAM, Abdul ; JANABI-SHARIFI, Farrokh ; FUNG, Alan S. ; RAAHEMIFAR, Kaamran: Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. In: *Energy and Buildings* 141 (2017), S. 96–113. <http://dx.doi.org/10.1016/j.enbuild.2017.02.012>. – DOI 10.1016/j.enbuild.2017.02.012
- [48] DOUNIS, A. I. ; CARASCOS, C.: Advanced control systems engineering for energy and comfort management in a building environment—A review. In: *Renewable and Sustainable Energy Reviews* 13 (2009), Nr. 6-7, S. 1246–1261. <http://dx.doi.org/10.1016/j.rser.2008.09.015>. – DOI 10.1016/j.rser.2008.09.015. – ISSN 13640321
- [49] DALAMAGKIDIS, Konstantinos ; KOLOKOTSA, Dionysia: *Reinforcement Learning for Building Environmental Control*. Vienna, Austria : I-Tech Education and Publishing, 2008 (Reinforcement Learning: Theory and Applications). – ISBN 978-3-902613-14-1
- [50] GÖRGES, Daniel: Relations between Model Predictive Control and Reinforcement Learning. In: *IFAC-PapersOnLine* 50 (2017), Nr. 1, S. 4920–4928. <http://dx.doi.org>

- [org/10.1016/j.ifacol.2017.08.747](https://doi.org/10.1016/j.ifacol.2017.08.747). – DOI 10.1016/j.ifacol.2017.08.747. – ISSN 24058963
- [51] ERNST, Damien ; GLAVIC, Mevludin ; CAPITANESCU, Florin ; WEHENKEL, Louis: Reinforcement learning versus model predictive control: a comparison on a power system problem. In: *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* 39 (2009), Nr. 2, S. 517–529. <http://dx.doi.org/10.1109/TSMCB.2008.2007630>. – DOI 10.1109/TSMCB.2008.2007630
- [52] HEITHORST, Bernd ; BAUER, Gregor ; RAUSCHER, Benedikt ; IRRGANG, Ludwig ; SPINNLER, Markus ; SATTELMAYER, Thomas: Experimental Comparison of Scroll and Swash-Plate Compressors for PV Driven Compression Chillers and Heat Pumps. In: HÄBERLE, Andreas (Hrsg.): *EuroSun2018*. Freiburg : International Solar Energy Society, op. 2018. – ISBN 978–3–9820408–0–6, S. 1–12
- [53] BÖING, Felix: *Auslegung und Aufbau eines Prüfstands zur Vermessung einer Photovoltaikbetriebenen Wärmepumpe*. Munich, Technical University of Munich, Term Thesis, 2014
- [54] BAUER, Gregor: *Development and Evaluation of Control Algorithms for a Heat Pump Operated in PV Mode*. Munich, Technische Universität München, Master's Thesis, 2018
- [55] SCHMIDEDER, Franziska: *Simulationen zur Deckung von Heiz- und Kühlbedarf von Gebäuden in gemäßigten und Wüsten- Klima mit PV-betriebenen Wärmepumpen / Klimageräten*. Munich, Technical University of Munich, Term Thesis, 2018
- [56] IRRGANG, Ludwig: *Experimental Comparison of Different Compressor-Refrigerant Combinations for PV Driven Heat Pumps and compression Chillers*. Munich, Technical University of Munich, Bachelor's Thesis, 2018
- [57] WEMHÖFER, Carsten ; HAFNER, Bernd ; SCHWARZER, Klemens: Simulation of Solar Thermal Systems with Carnot Blockset in the Environment Matlab Simulink. (2000). <https://pdfs.semanticscholar.org/c4c1/dcb296d1b4708f32b43984e22414842dd87c.pdf>
- [58] *Conventional And Renewable Energy Systems Optimization (Carnot): Matlab Extension*. Version: 2018
- [59] PEIRELINCK, Thijs ; RUELENS, Frederik ; DECNONINCK, Geert: Using reinforcement learning for optimizing heat pump control in a building model in Modelica. In: CONFERENCE, IEEE International E. (Hrsg.): *2018 IEEE International Energy Conference (ENERGYCON)*. [Piscataway, NJ] : IEEE, 2018. – ISBN 978–1–5386–3669–5, S. 1–6

BIBLIOGRAPHY

- [60] CHEN, Yujiao ; NORFORD, Leslie K. ; SAMUELSON, Holly W. ; MALKAWI, Ali: Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. In: *Energy and Buildings* 169 (2018), S. 195–205. <http://dx.doi.org/10.1016/j.enbuild.2018.03.051>. – DOI 10.1016/j.enbuild.2018.03.051
- [61] F. RUELENS, S. IACOVELLA, B. CLAESSENS AND R. BELMANS: Learning Agent for a Heat-Pump Thermostat with a Set-Back Strategy Using Model-Free Reinforcement Learning. In: *Energies* 8 (2015), Nr. 8, S. 8300–8318. <http://dx.doi.org/10.3390/en8088300>. – DOI 10.3390/en8088300
- [62] FAZENDA, Pedro ; VEERAMACHANENI, Kalyan ; LIMA, Pedro ; O'REILLY, Una-May: Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems. In: *Journal of Ambient Intelligence and Smart Environments* 6 (2014), Nr. 6, S. 675–690. <http://dx.doi.org/10.3233/AIS-140288>. – DOI 10.3233/AIS-140288
- [63] HUANG, Hao ; CHEN, Lei ; HU, Eric: A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. In: *Energy and Buildings* 97 (2015), S. 86–97. <http://dx.doi.org/10.1016/j.enbuild.2015.03.045>. – DOI 10.1016/j.enbuild.2015.03.045
- [64] EKREN, Orhan ; SAHIN, Savas ; ISLER, Yalcin: Comparison of different controllers for variable speed compressor and electronic expansion valve. In: *International Journal of Refrigeration* 33 (2010), Nr. 6, S. 1161–1168. <http://dx.doi.org/10.1016/j.ijrefrig.2010.05.005>. – DOI 10.1016/j.ijrefrig.2010.05.005. – ISSN 01407007
- [65] DELGARM, Navid ; SAJADI, Behrang ; DELGARM, Saeed: Multi-objective optimization of building energy performance and indoor thermal comfort: A new method using artificial bee colony (ABC). In: *Energy and Buildings* 131 (2016), S. 42–53. <http://dx.doi.org/10.1016/j.enbuild.2016.09.003>. – DOI 10.1016/j.enbuild.2016.09.003
- [66] CONGRADAC, Velimir ; KULIC, Filip: HVAC system optimization with CO2 concentration control using genetic algorithms. In: *Energy and Buildings* 41 (2009), Nr. 5, S. 571–577. <http://dx.doi.org/10.1016/j.enbuild.2008.12.004>. – DOI 10.1016/j.enbuild.2008.12.004
- [67] MOSSOLLY, M. ; GHALI, K. ; GHADDAR, N.: Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm. In: *Energy* 34 (2009), Nr. 1, S. 58–66. <http://dx.doi.org/10.1016/j.energy.2008.10.001>. – DOI 10.1016/j.energy.2008.10.001. – ISSN 03605442
- [68] NASSIF, N. ; KAJL, S. ; SABOURIN, R.: Evolutionary algorithms for multi-objective optimization in HVAC system control strategy. In: *IEEE Annual Meeting of the Fuzzy Information, 2004. Processing NAFIPS '04* (2004), S. 51–56 Vol.1. <http://dx.doi.org/10.1109/NAFIPS.2004.1336248>. – DOI 10.1109/NAFIPS.2004.1336248

- [69] J. WRIGHT, H. LOOSEMORE AND R. FARMANI: Optimization of building thermal design and control by multi-criterion genetic algorithm. In: *Energy and Buildings* 34 (2002), Nr. 9, S. 959–972. [http://dx.doi.org/10.1016/s0378-7788\(02\)00071-3](http://dx.doi.org/10.1016/s0378-7788(02)00071-3). – DOI 10.1016/s0378-7788(02)00071-3
- [70] MOCANU, Elena ; MOCANU, Decebal C. ; NGUYEN, Phuong H. ; LIOTTA, Antonio ; WEBBER, Michael E. ; GIBESCU, Madeleine ; SLOOTWEG, J. G.: On-Line Building Energy Optimization Using Deep Reinforcement Learning. In: *IEEE Transactions on Smart Grid* 10 (2019), Nr. 4, S. 3698–3708. <http://dx.doi.org/10.1109/TSG.2018.2834219>. – DOI 10.1109/TSG.2018.2834219. – ISSN 1949–3053
- [71] WEI, Tianshu ; WANG, Yanzhi ; ZHU, Qi: Deep Reinforcement Learning for Building HVAC Control. In: UNKNOWN (Hrsg.): *2017 54th ACM*. Piscataway, NJ : IEEE, 2017. – ISBN 9781450349277, S. 1–6
- [72] SHAH, Hitesh ; GOPAL, M.: Model-Free Predictive Control of Nonlinear Processes Based on Reinforcement Learning. In: *IFAC-PapersOnLine* 49 (2016), Nr. 1, S. 89–94. <http://dx.doi.org/10.1016/j.ifacol.2016.03.034>. – DOI 10.1016/j.ifacol.2016.03.034. – ISSN 24058963
- [73] FERREIRA, P. M. ; RUANO, A. E. ; SILVA, S. ; CONCEIÇÃO, E.Z.E.: Neural networks based predictive control for thermal comfort and energy savings in public buildings. In: *Energy and Buildings* 55 (2012), S. 238–251. <http://dx.doi.org/10.1016/j.enbuild.2012.08.002>. – DOI 10.1016/j.enbuild.2012.08.002
- [74] MATELLANAS, E. ; CASTILLO-CAGIGAL, M. ; GUTIÉRREZ, A. ; MONASTERIO-HUELIN, F. ; CAAMAÑO-MARTÍN, E. ; MASA, D. ; JIMÉNEZ-LEUBE, J.: Neural network controller for Active Demand-Side Management with PV energy in the residential sector. In: *Applied Energy* 91 (2012), Nr. 1, S. 90–97. <http://dx.doi.org/10.1016/j.apenergy.2011.09.004>. – DOI 10.1016/j.apenergy.2011.09.004
- [75] MASON, Karl ; GRIJALVA, Santiago: A review of reinforcement learning for autonomous building energy management. In: *Computers & Electrical Engineering* 78 (2019), S. 300–312. <http://dx.doi.org/10.1016/j.compeleceng.2019.07.019>. – DOI 10.1016/j.compeleceng.2019.07.019. – ISSN 00457906
- [76] MATHWORKS.COM: *Reinforcement Learning Toolbox*. https://www.mathworks.com/help/reinforcement-learning/index.html?s_tid=CRUX_lftnav
- [77] CHRISTIAN MÜNCH GMBH: *Einspeisevergütung*. <https://www.solaranlage.eu/photovoltaik/wirtschaftlichkeit/einspeiseverguetung>
- [78] STADTWERKE MÜNCHEN GMBH: *M-Strom*. <https://www.swm.de/privatkunden/mstrom>

BIBLIOGRAPHY

- [79] PÉREZ CASTAÑO, Arnaldo: *Practical Artificial Intelligence: Machine Learning, Bots, and Agent Solutions Using C#*. Berkeley, CA : Apress, 2018. <http://dx.doi.org/10.1007/978-1-4842-3357-3>. <http://dx.doi.org/10.1007/978-1-4842-3357-3>. – ISBN 9781484233566
- [80] NGUYEN, Ngoc T. ; KIM, Chong-Gun ; JANIÁK, Adam: *Lecture notes in computer science Lecture notes in artificial intelligence*. Bd. 6591: *Intelligent information and database systems: Third international conference, ACIIDS 2011, Daegu, Korea, April 20 - 22, 2011 ; proceedings, part I*. Berlin : Springer, 2011. <http://dx.doi.org/10.1007/978-3-642-20039-7>. <http://dx.doi.org/10.1007/978-3-642-20039-7>. – ISBN 9783642200380
- [81] *Repräsentative elektrische Lastprofile für Einfamilienhäuser in Deutschland auf 1-sekündiger Datenbasis: Datensatz*. Berlin, 2015
- [82] MÖHRSTEDT, Julius: *Fallstudie zu einem wirtschaftlichen Vergleich verschiedener Heizkonzepte für ein Einfamilienhaus: PV-betriebene Wärmepumpe vs. solarthermisch unterstützte Gastherme*. Munich, Technical University of Munich, Diploma thesis, 2018
- [83] ROSS, Sheldon M.: *Introductory statistics*. Fourth edition. Amsterdam and Boston and Heidelberg and London and New York and Oxford and Paris and San Diego and San Fransisco and Singapore and Sydney and Tokyo : Academic Press, 2017. – ISBN 9780128043172
- [84] SCHLEGL, Thomas ; KOST, Christoph: *Study: Levelized Cost of Electricity- Renewable Energy Technologies*
- [85] HIRVONEN, Janne ; KAYO, Genku ; HASAN, Ala ; SIRÉN, Kai: Zero energy level and economic potential of small-scale building-integrated PV with different heating systems in Nordic conditions. In: *Applied Energy* 167 (2016), S. 255–269. <http://dx.doi.org/10.1016/j.apenergy.2015.12.037>. – DOI 10.1016/j.apenergy.2015.12.037
- [86] ANGENENDT, Georg ; ZURMÜHLEN, Sebastian ; RÜCKER, Fabian ; AXELSEN, Hendrik ; SAUER, Dirk U.: Optimization and operation of integrated homes with photovoltaic battery energy storage systems and power-to-heat coupling. In: *Energy Conversion and Management: X* 1 (2019). <http://dx.doi.org/10.1016/j.ecmx.2019.100005>. – DOI 10.1016/j.ecmx.2019.100005. – ISSN 25901745
- [87] ANGENENDT, Georg ; ZURMÜHLEN, Sebastian ; AXELSEN, Hendrik ; SAUER, Dirk U.: Prognosis-Based Operating Strategies for Smart Homes with Power-to-Heat Applications. In: *Energy Procedia* 155 (2018), S. 136–148. <http://dx.doi.org/10.1016/j.egypro.2018.11.061>. – DOI 10.1016/j.egypro.2018.11.061. – ISSN 18766102
- [88] FRITSCHÉ, Uwe ; GRESS, Hans-Werner: *Der nichterneuerbare kumulierte Energieverbrauch und THG-Emissionen des deutschen Strom-mix im Jahr 2018 sowie Ausblicke*

- auf 2020 bis 2050: Bericht für die HEA – Fachgemeinschaft für effiziente Energieanwendung e.V.*
- [89] KASTL, A.: *Entwicklung eines thermischen Modells zur Gebäudesimulation*. Munich, Technical University of Munich, Term thesis, 2015
- [90] PFLUGRADT, N.: *Modellierung von Wasser und Energieverbräuchen in Haushalten: Loadprofil Generator*, TU Chemnitz, Dissertation, 2017
- [91] SCHELZE, E.: *Development of a PV model based on the Sandia Model*, Technical University of Munich, Reserach internship report (under review), 2020
- [92] BUNDESAMT FÜR WIRTSCHAFT UND AUSFUHRKONTROLLE: *Liste der Wärmepumpen mit Prüfnachweis*. https://www.bafa.de/SharedDocs/Downloads/DE/Energie/ee_waermepumpen_anlagenliste.html
- [93] VIESSMANN: *Preisliste 2018 DE Heizsysteme*
- [94] HANSEN, Kenneth: Decision-making based on energy costs: Comparing levelized cost of energy and energy system costs. In: *Energy Strategy Reviews* 24 (2019), S. 68–82. <http://dx.doi.org/10.1016/j.esr.2019.02.003>. – DOI 10.1016/j.esr.2019.02.003. – ISSN 2211467X
- [95] EXPERTEN.ORG energie: *Wärmepumpeheizung: Kosten*. <https://www.energie-experten.org/heizung/waermepumpe/waermepumpenheizung/kosten.html#c25164>. Version: 2020
- [96] EUPD RESEARCH: *Photovoltaik-Preismonitor Deutschland: German PV Module Price-Monitor 2016: Ergebnisse 1. Quartal*
- [97] BAKELLI, Yahia ; HADJ ARAB, Amar ; AZOUI, Boubekeur: Optimal sizing of photovoltaic pumping system with water tank storage using LPSP concept. In: *Solar Energy* 85 (2011), Nr. 2, S. 288–294. <http://dx.doi.org/10.1016/j.solener.2010.11.023>. – DOI 10.1016/j.solener.2010.11.023. – ISSN 0038092X
- [98] BUNDESVERBAND WÄRMEPUMPE (BWP) E.V.: *BAFA-Förderung für Wärmepumpen*. <https://www.waermepumpe.de/waermepumpe/foerderung/bafa-foerderung-fuer-waermepumpen/>
- [99] BUNDESMINISTERIUM DER FINANZEN: *Steuern im Einkaufswagen: Was wird wie hoch besteuert?* https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Service/Einfach_erklaert/2019-05-02-steuern-im-einkaufswagen-flash-infografik.html. Version: 2019
- [100] UMWELT BUNDESAMT: *Erneuerbare Energien in Zahlen*. <https://www.umweltbundesamt.de/themen/klima-energie/erneuerbare-energien/erneuerbare-energien-in-zahlen#uberblick>

BIBLIOGRAPHY

- [101] UNITED NATIONS: *Sustainable Development Goals*. <https://www.un.org/sustainabledevelopment/>
- [102] UNITED NATIONS: *Goal 11: Make cities inclusive, safe, resilient and sustainable: 11 Sustainable Cities and Communities*
- [103] GERMAN ASSOCIATION OF ENERGY AND WATER INDUSTRIES: *Wie heizt Deutschland 2019? BDEW-Studie zum Heizungsmarkt*. https://www.bdew.de/media/documents/Pub_20191031_Wie-heizt-Deutschland-2019.pdf
- [104] RATS INFORMATIONEN SYSTEM MÜNCHEN: *Emissionsfaktoren die von der Landeshauptstadt München verwendet werden*. <https://www.ris-muenchen.de/RII/RII/DOK/SITZUNGSVORLAGE/2291004.pdf>
- [105] UNITED NATIONS: *Ensure access to affordable, reliable, sustainable and modern energy: 7 Affordable and Clean Energy*. <https://www.un.org/sustainabledevelopment/energy/>

A March Examples of the Verification Process

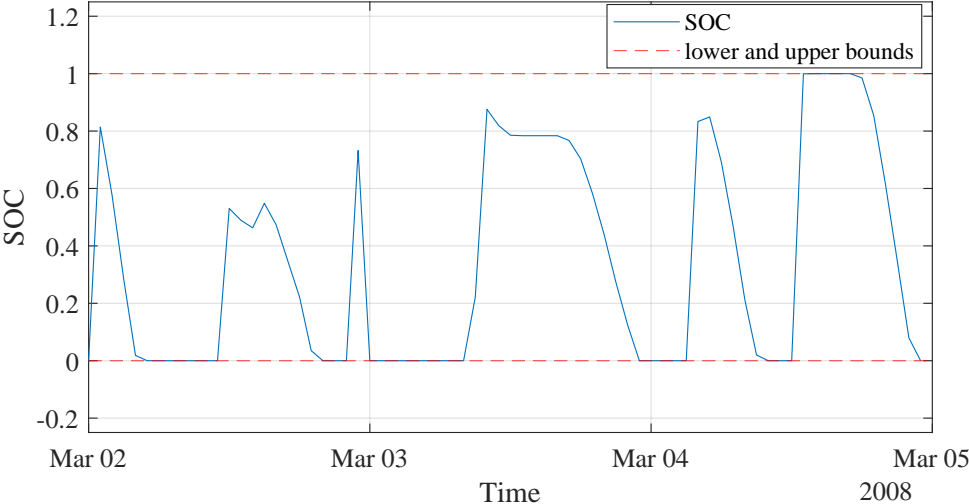


Figure A.1 – Bound compliance of the state of charge of the thermal storage during the period 2nd to the 5th of March 2009

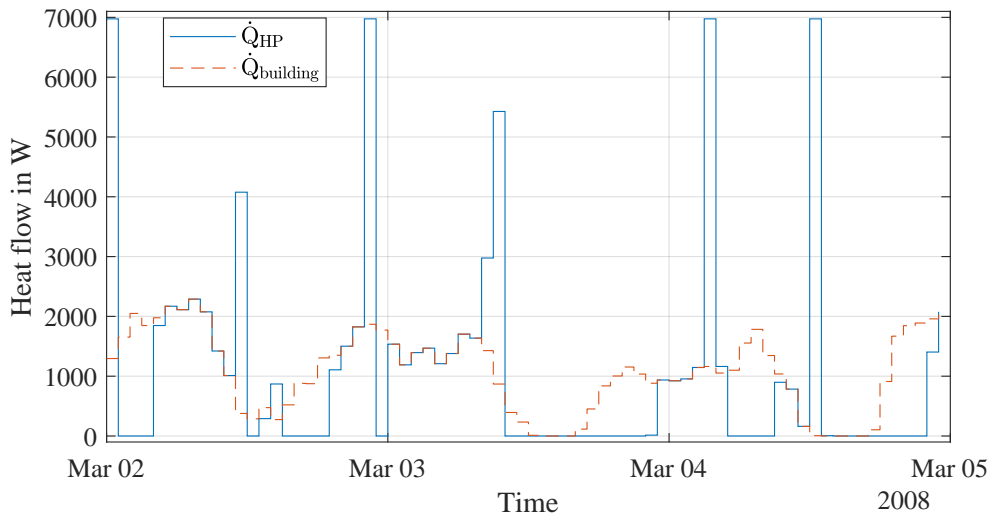


Figure A.2 – Overview of the building thermal demand and the thermal power provided by the heat pump during the period 2nd to the 5th of March 2009

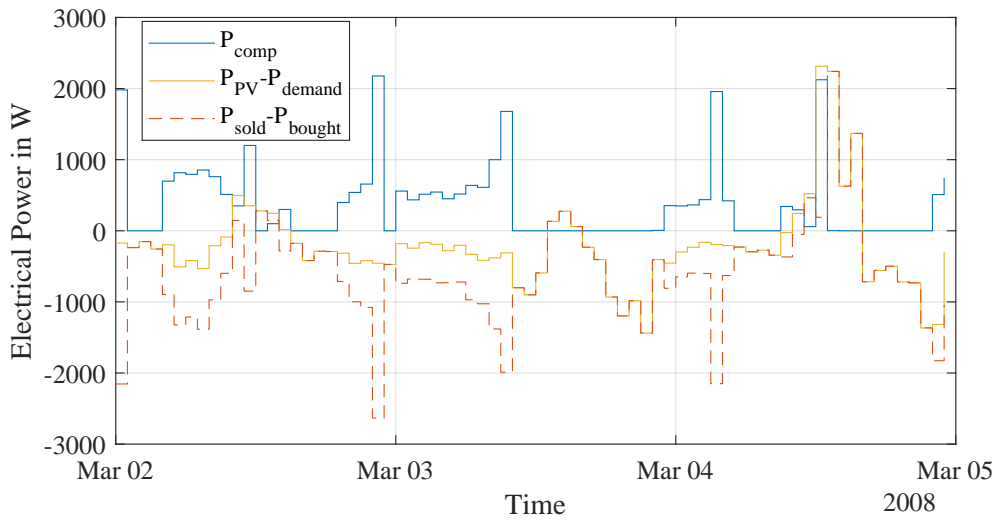


Figure A.3 – Overview of the available power, the heat pump power and the resulting bought and sold powers during three exemplary march days

B Training Patters of the Verification Process

This appendix gathers the training patterns obtained in the verification of the DRL control. The number of episodes is set to 200. In this training study and the observations of the DRL agent are the corrected compressor energy, temperature of thermal storage, energy bought from the grid, energy sold to the grid, and the COP of the heat pump. The action is the compressor energy vector for the next 48h.

As can be seen, the Episode Manager encountered a graphic error in one day of November (12th) and one day of March (7th). In the following such graphic errors are avoided by choosing not to display the Episode Manager graphically. The evolution of the training process can also be displayed as text in the command window of MATLAB.

Each of the figures represents the training progress of the deep reinforcement learning agent. The stared red line represents the average of the hitherto obtained rewards marked by a circled-blue line. The crossed-green line refers to the discounted estimation of the long-term reward. In table [B.1](#) and table [B.2](#), each of the graphs correspond to one day of the respective month. The days are ordered from left to right and top to bottom, i.e. the first five days of the month are displayed in the first line of the table.

Table B.1 – Training patterns of the month of March obtained during the DRL control verification

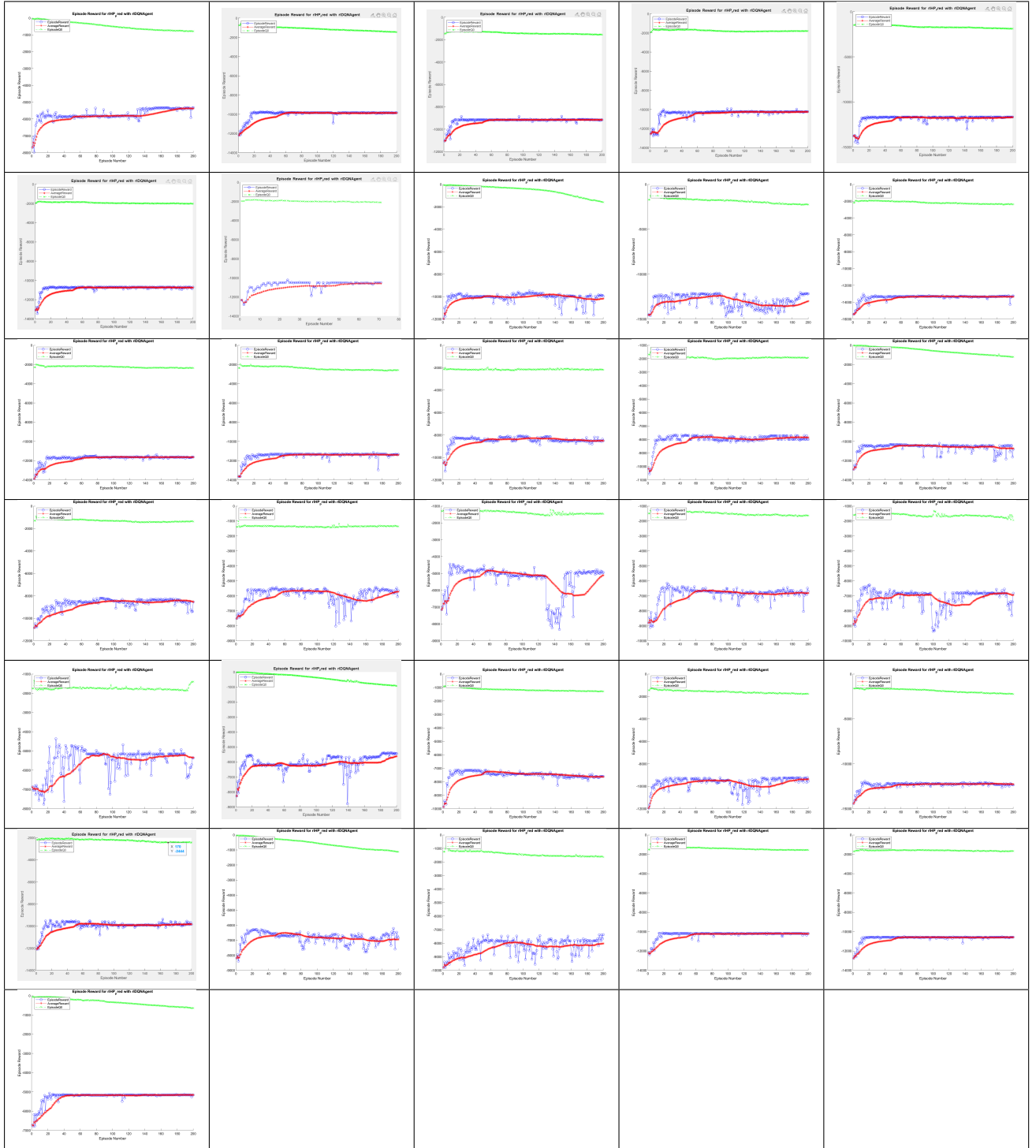
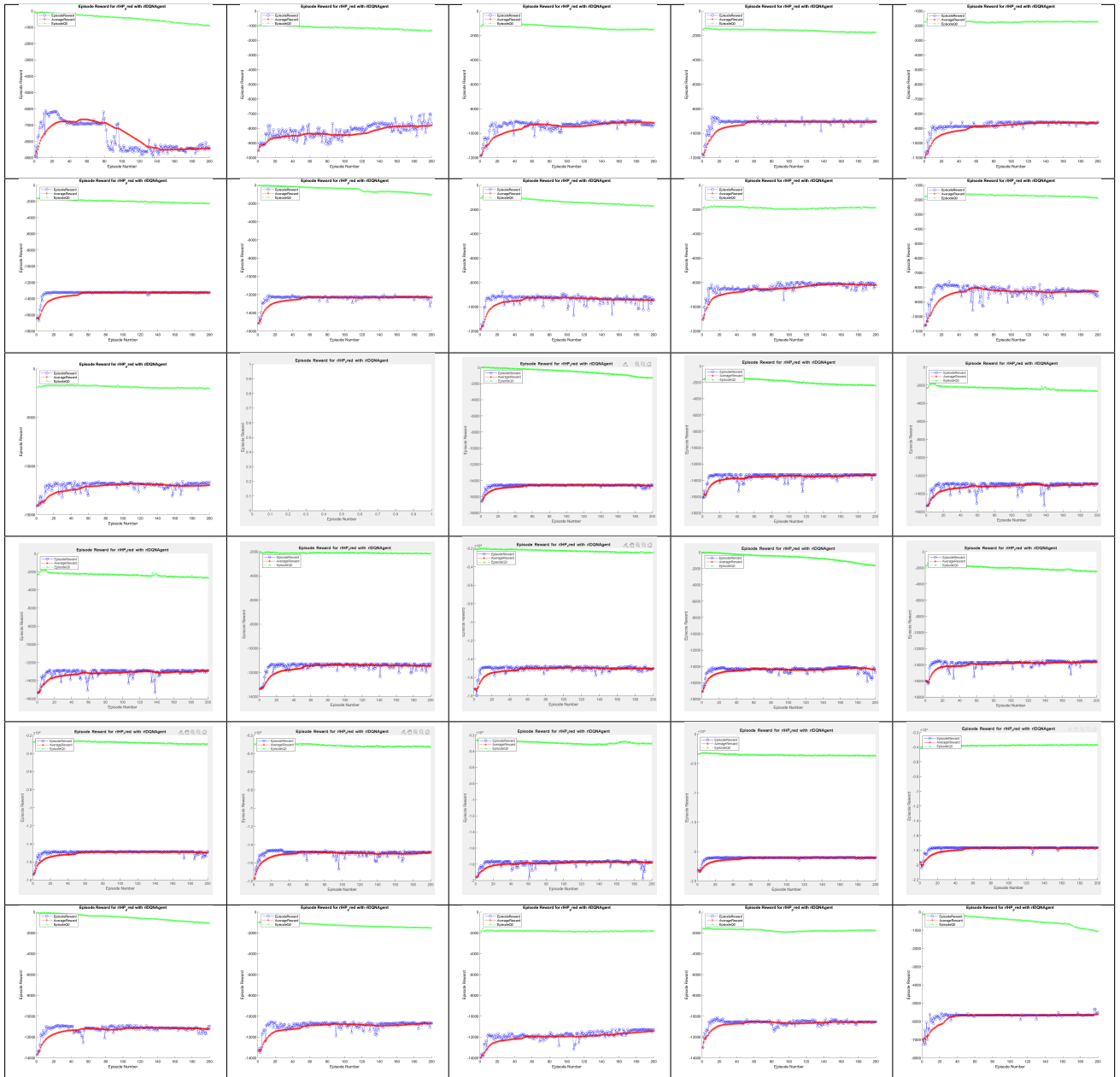


Table B.2 – Training patterns of the month of November obtained during the DRL control verification



C Simulation Inputs

The following images represent the annual prediction hourly vectors used to simulate the building energy system using Rauscher's MPC [8] and the DRL control developed in this thesis. Note that, although the annual data runs from the 28th of September 2008 to the 27th of September 2009, it has been here reorganized for clarity reasons. The following graphs represent the data starting January, 1st.

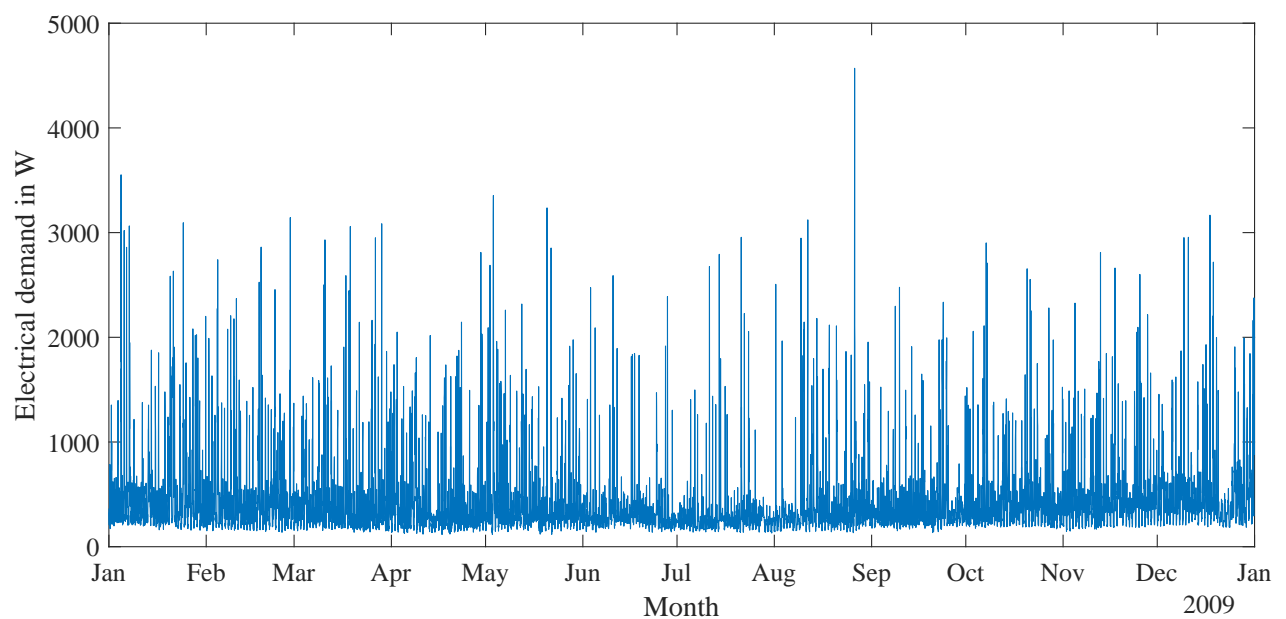


Figure C.1 – Annual electrical demand

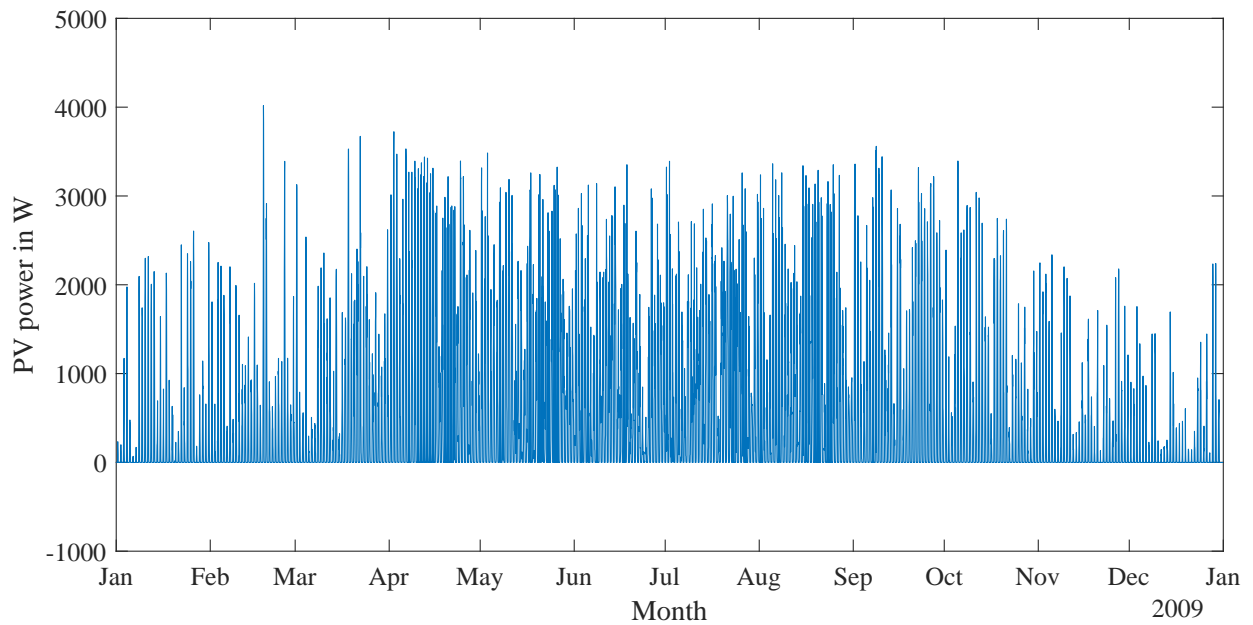


Figure C.2 – Annual PV production

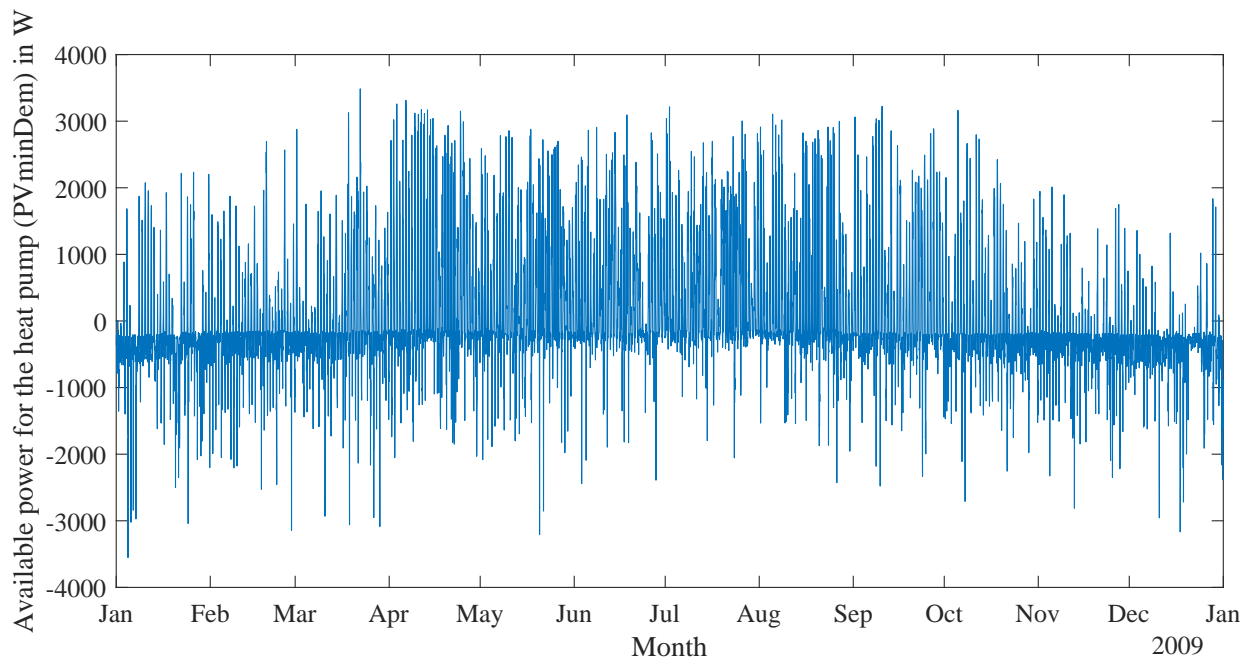


Figure C.3 – Available PV for the heatpump : excess PV after electrical demand coverage

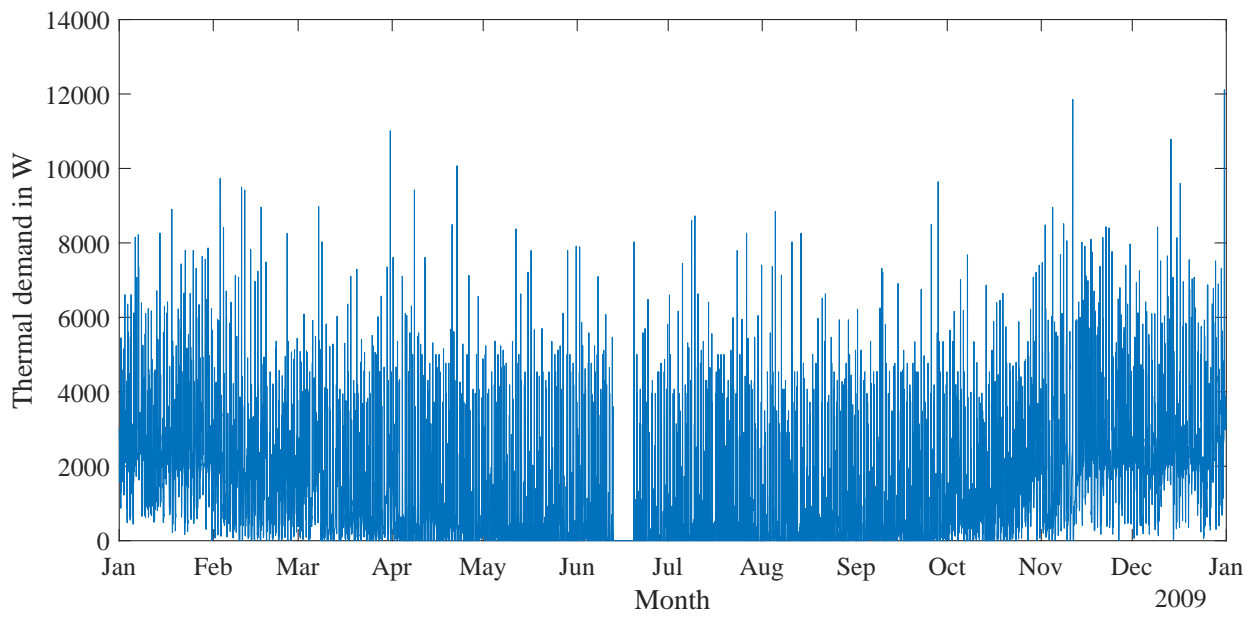


Figure C.4 – Annual sum of thermal and hot water demand

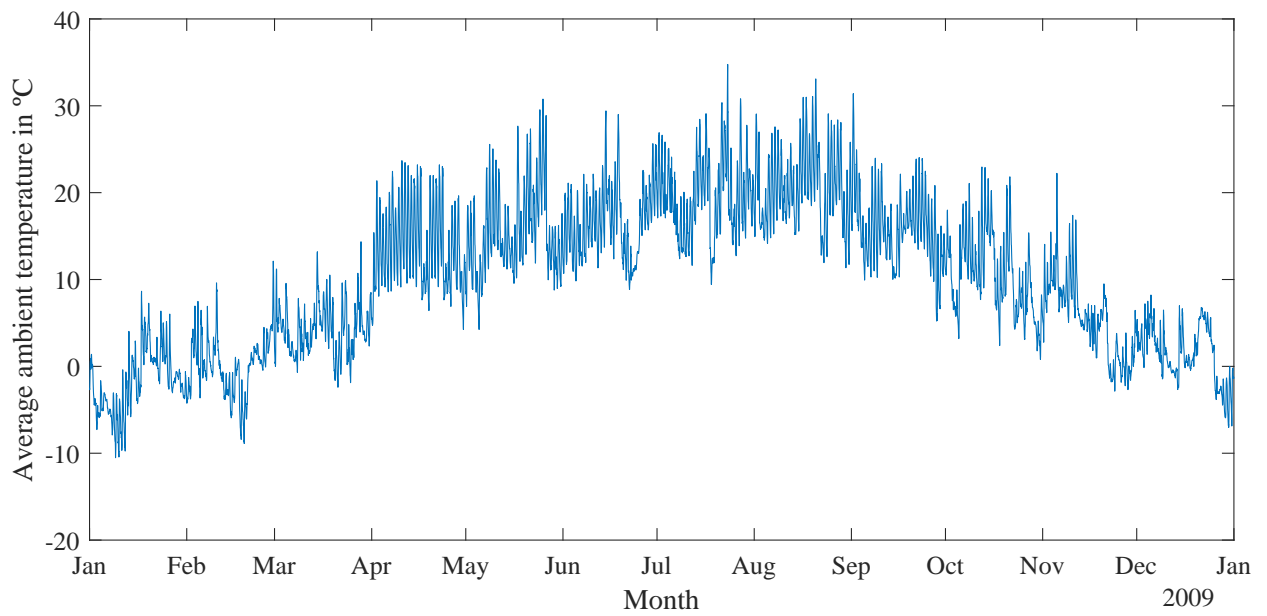


Figure C.5 – Average ambient temperature in the center of Munich

D Post-Processing of Parameter Study Simulations

Model Parameters		
Heat Pump		
Max. Power	KW	3
Cost	€	7.971
Installation with storage	€	6.500
Total cost	€	14.471
Taxes	€	2.749
National Subsidies	€	6.027
Total investment	€	11.193
Thermal Storage		
Capacity	L	750
Cost	€	3.449
Taxes	€	655
National Subsidies	€	1.437
Total investment	€	2.668
Solar Panels		
Power	KWp	5,0
Net Cost	€/KWp	1.456
BOS	%	32%
Module	%	48%
Installation	%	20%
Total investment	€	7.280

Economic Parameters		
Interest Rate	%	3%
Mean Electricity Price	€/kWh	0,30
Feed-in Tarif	€/kWh	0,11
Value Added Tax	% of total cost	19%
National Subsidies for Heat Pumps	% of total cost	35%
Life-time of the project	years	20

Operation and Maintenance		
Heat Pump	% of cost	2,80%
Solar Panel	% of cost	1%
Thermal Storage	% of cost	1%

Simulation Results		
Electrical Demand	kWh	3.775
Heating Demand	KWh	11.046
Heat Pump Electrical Consumption	kWh	4.035
Heat Pump Thermal Yield	kWh	11.614
Electricity cost without heating system	€	444
Energy bought from the grid	kWh	5.202
Energy sold to the grid	kWh	1.619
Self-consumed energy	kWh	2.608
PV Production	kWh	4.227
SCR		0,62

System Economic Performance		
<i>Investment Cost</i>		
Heat Pump	€	11.193
Thermal Storage	€	2.668
Solar Panels	€	7.280
Total Investment	€	21.141
<i>Operation Costs</i>		
Operation and Maintenance	€	296
Price of Imported Electricity	€	1.560
PV Remuneration	€	178
Annual Variable Cost	€	1.678

Discounted Sum of Results Over a 20 year Period		
Variable costs of Electricity	€	21.650
Electricity demand	kWh	116.189
Heating Production	KWh	172.787
Variable costs of Heat	€	17.791

Levelized Cost of Energy		
LCOE	€/kWh	0,25
LCOH	€/kWhth	0,18
LCOEnergy	€/kWh	0,20

Figure D.1 – Post-processing spreadsheet of the parameter study

E Further Results of the DRL Long-term Simulations

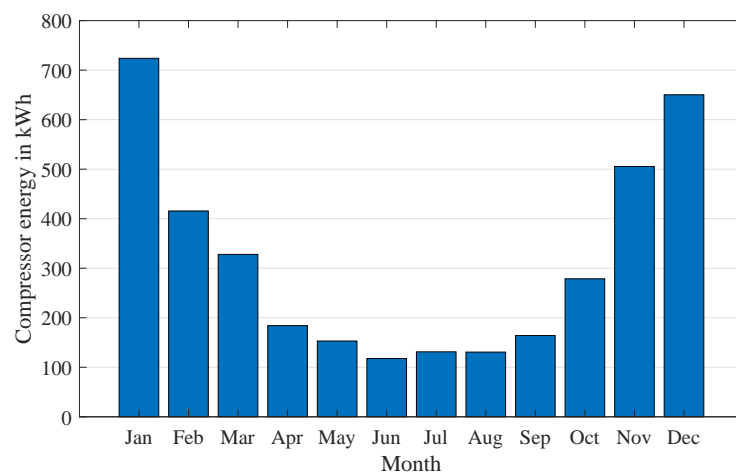


Figure E.1 – Monthly heat pump energy obtained with the DRL control

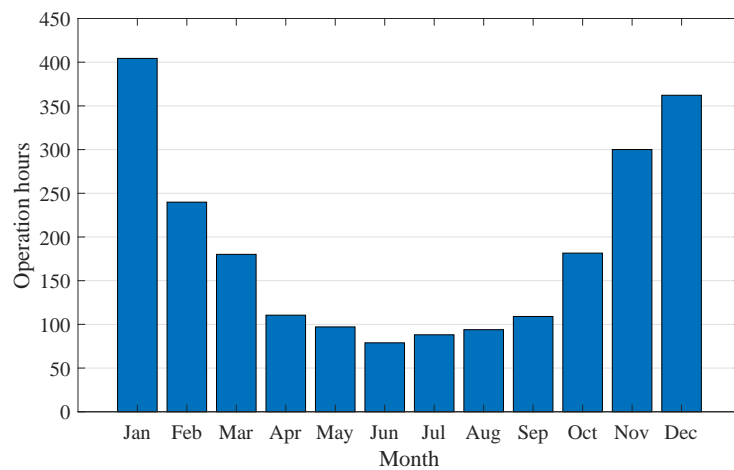


Figure E.2 – Monthly heat pump operation hours obtained with the DRL control

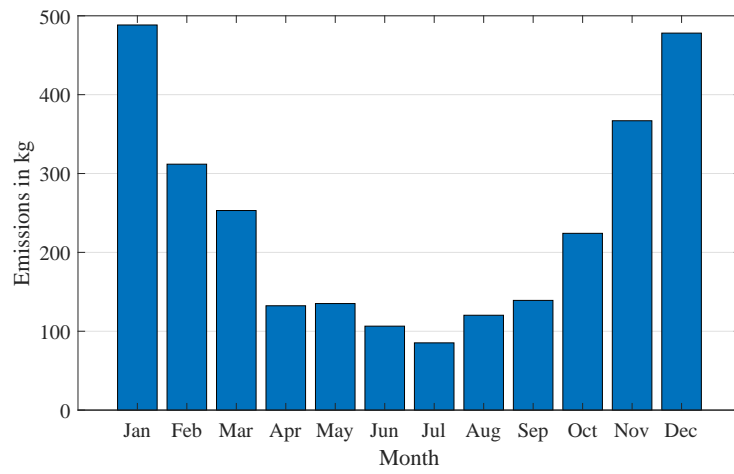


Figure E.3 – Monthly emissions produced with the DRL control

F Further Results of the Control Comparison

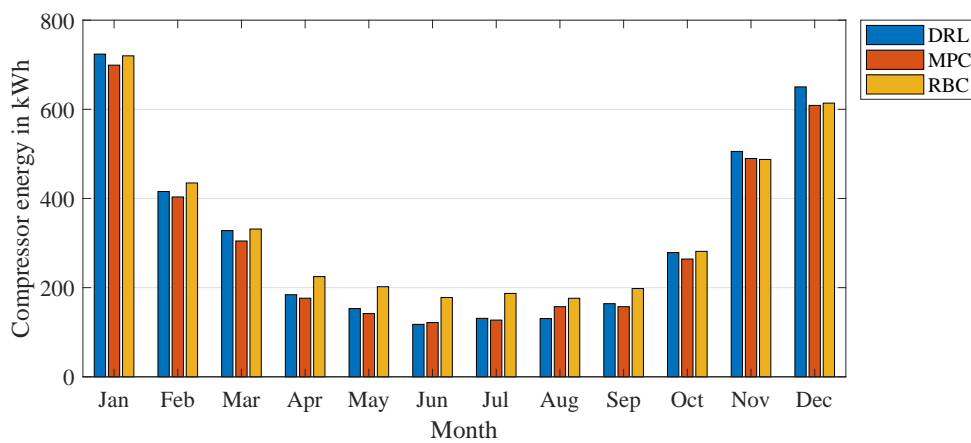


Figure F.1 – Comparison of the monthly heat pump energy when controlling the heat pump with the DRL, MPC and rule-based controls

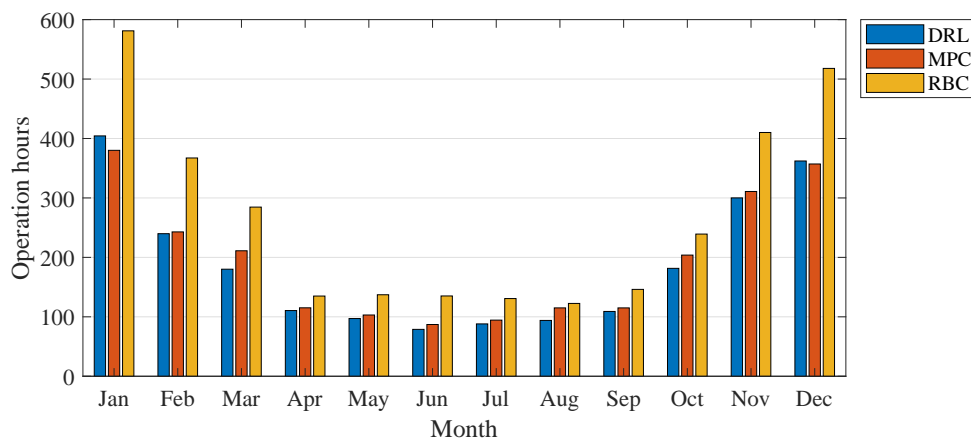


Figure F.2 – Comparison of the monthly heat pump operation hours when controlling the heat pump with the DRL, MPC and rule-based controls

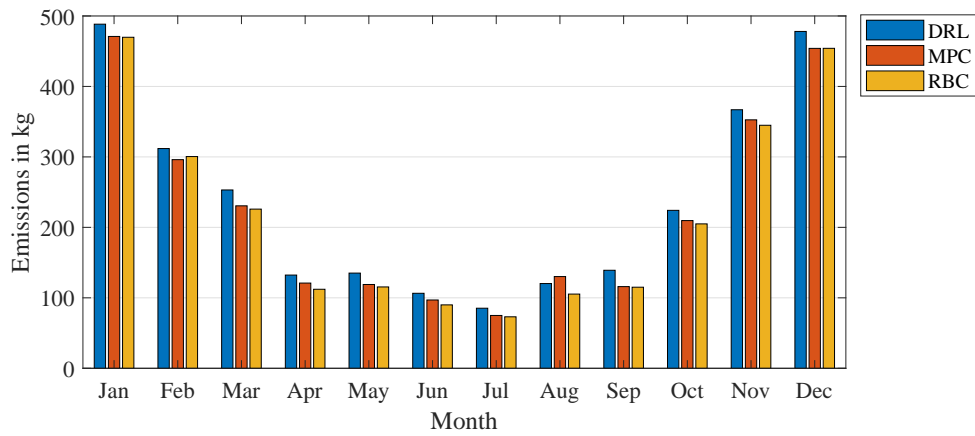


Figure E.3 – Comparison of the monthly emissions when controlling the heat pump with the DRL, MPC and rule-based controls

G Alignment of the Project with the Sustainable Development Goals

The United Nations decided in 2015 of a 15-year plan of action, the "2030 Agenda for Sustainable Development" [101]. The agenda aimed at the achievement of a list of 17 broad and interdependent Goals to create a more sustainable future for everyone all over the world. In an attempt to be a part of this global, highly necessary, and urgent initiative, this project is framed in at least one of the Sustainable Development Goals set by the United Nations.

The main objective of this thesis is to develop an intelligent control for a building energy system that aims at the reduction of the energy costs and the maximization of self-consumption and therefore, the minimization of CO₂ emissions. The data used in this thesis simulates a single-family household in the city of Munich (Germany). However, this application is only used as an example. With different input data and the pertinent alterations, the systems could be used in different climates, areas, or countries. Because of the aforementioned information, this project aligns with the eleventh Sustainable Development Goal: "Sustainable Cities and Communities". As mentioned, the project is not exclusively designed for urban areas, but can be applied to any rural or urban building.

The eleventh goal of the United Nations aims at making urban areas safer, more resilient, and more sustainable. The motivation behind this goal is the increasing importance of cities and their impact on energy consumption and CO₂ emissions. In fact, according to the United Nations [102], cities are responsible for 60 to 80% of global energy consumption and 75% of total CO₂ emissions. Additionally, nine out of ten urban inhabitants are exposed to polluted and unsafe air. Objective 11.6 strives for the reduction of the environmental impact of urban areas and especially the improvement of air quality [102].

To quantify to what extent this thesis has fulfilled this decarbonization goal the single-family household used to test and evaluate the control is now considered again. The household's annual electric demand is 3.774,5 kWh and its annual heating demand is 11.046 kWh. The CO₂ equivalent of the German electricity mix reaches 0,502 kg/kWh [88]. Therefore, if the electricity would entirely be supplied by the electrical grid, this would entail 1.894,8 kg of CO₂ emitted per year to cover the electrical demand. Regarding the thermal demand, 36% of the households in Germany use natural gas to cover the heating demand, making natural gas the leading heating provider for the residential sector [103]. The carbon footprint of natural gas

according to the Council Information System of the city of Munich [104] is of 0,255 kg per kWh. For the thermal demand of the investigated building, that leads to 2.816,7 kg of annual CO₂ emissions. In total, considering both thermal and electrical demand without PV and electric heating, the annual carbon emissions would reach 4.711,5 kg. The proposed system uses PV-driven electric heating and a deep reinforcement learning control. The system generates a total of 2.842 kg of emissions, which is 40% less than what a "traditional" system would produce. The decarbonization potential of the investigated system is striking and could highly help cities and other areas to become cleaner and more sustainable.

A further Sustainable Development Goal that is tackled in this thesis is the seventh objective: "Affordable and clean energy". This objective aims at ensuring more sustainable, affordable, and reliable access to energy [105]. More specifically, by installing a PV generation unit that is able to feed excess electricity into the electrical grid, this project is helping with objective number 7.2 that encourages the increase of the share of renewable energy sources in the energy mix [105]. The simulation year going from the 28th of September 2008 to the 27th of September 2009, the deep reinforcement learning control developed within this thesis led to 2.330 kWh of PV energy being fed to the German electrical grid.

To conclude, this thesis tries to contribute to the Sustainable Development Goals set by the United Nations in 2015. Specifically, the thesis tackles the issue of making cities and communities more sustainable, by proposing a cleaner energy and control system for buildings that leads to lower emissions and ultimately reduces air pollution. Additionally, the thesis also contributes to the objective of providing cleaner energy by being able to feed renewable energy in the electrical grid.