Reinforcement learning for occupant-centric operation of residential energy system: Evaluating the adaptation potential to the unusual occupants´ behavior during COVID-19 pandemic

Amirreza Heidari, *François* Maréchal, *Dolaana* Khovalyg

School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fdérale de Lausanne (EPFL), Lausanne, Switzerland, amirreza.heidari@epfl.ch

Abstract. Occupant behavior is a highly stochastic phenomenon, which is known as a key challenge for the optimal control of residential energy systems. With the increasing share of renewable energy in the building sector, the volatile nature of renewable energy is also another key challenge for optimal control. It is challenging and time-consuming to develop a rule-based or model-based control algorithm that can properly take into account these stochastic parameters and ensure an optimal operation. Rather, a learning ability can be provided for the controller to learn these parameters in each specific house, without the need for any model. This research aims to develop a model-free control framework, based on Reinforcement Learning, which takes into account the stochastic occupants' behavior and PV power production and tries to minimize energy use while ensuring occupants' comfort and water hygiene. This research, for the first time, integrates a model of Legionella growth to ensure that energy saving is not with the cost of occupants' health. Hot water use data of three different residential houses are measured to evaluate the performance of the proposed framework on realistic occupants' behavior. The measurement campaign was during the COVID-19 pandemic, which would further highlight the adaptability of the Reinforcement Learning framework to the unusual situation when the prediction of occupants' behavior is even more challenging. Results indicate that the proposed framework can successfully learn and predict occupants' behavior and PV power production, and significantly reduce energy use without violating comfort and hygiene aspects.

Keywords. Reinforcement Learning, Space heating, Hot water, Occupant behavior, Solar energy, Machine Learning
DOI: https://doi.org/10.34641/clima.2022.286

1. Introduction

Optimal operation of building energy systems is challenging as there are several stochastic and time-varying parameters that affect building energy use. One of these parameters is occupant behaviour, which is highly stochastic, can change from day to day, and therefore is very hard to predict [1]. The occupant behaviour of each building is unique, and thus there is no universal model which can be embedded in the control system of various buildings at their design phase. To cope with this highly stochastic parameter, current control approaches are usually too conservative to ensure the comfort of occupants regardless of their behaviour. An example is hot water production, where huge volume of hot water with high temperature is produced in advance and stored in a tank to make sure enough hot water is available whenever it is demanded [2,3].

Another stochastic parameter affecting building operation is renewable energy. The share of renewable energy in the building sector is increasing, and is expected to get doubled by 2030 [4]. Due to the volatile nature of renewable energy sources, it will also increase the complexity of optimal energy management in buildings [5]. There are several other stochastic parameters, such as weather condition or electric vehicles charging that all affect the building energy use. The control logic of buildings should properly take into account these stochastic parameters to guarantee an optimal operation.

To integrate these stochastic parameters, a possible option is Model Predictive Control (MPC) that uses a model of the system, together with the predictions of stochastic parameters, to determine the optimal control actions for an upcoming horizon. Despite its potential benefits, MPC relies on a model of the system. However, developing an accurate model for the building is extremely time-consuming, and therefore, not practical in most cases. Moreover, even if an accurate model of the system is developed, it can become fairly inaccurate over time due to, for instance, renovation or aging of the system. Furthermore, similar to the other model-based approaches, MPC requires a high computational power to optimize the model of the system. Last but not least, as MPC is based on a model of the building and the prediction of stochastic parameters, it is building-specific and not easily transferrable to other buildings [6].

Uniqueness of occupant behaviour in each building makes it challenging to program a rule-based or model-based control logic that can be easily transferred to many other buildings. Rather than hard programming a rule-based or model-based control method, a learning ability can be provided to the controller such that it can learn and adapt to the specifications of that building and maintain an optimal operation. Reinforcement Learning (RL) is a method of Machine Learning that can provide this learning ability to the controller. In RL, a learning agent interacts with its environment, and uses feedback from the environment to determine the best possible action to maximize a pre-defined metric called reward [7]. RL provides two main benefits over the rule-based and model-based control methods. First of all, RL does not require a complex thermodynamic model of the system, as agent can learn the system model only by interacting with the system [8]. This is a great advantage over MPC, especially in case of complex systems that require a lot of time and effort for modelling [9]. Secondly, RL can continuously learn and adapt to the changes in system such as variating weather conditions, volatile renewable energy, or stochastic occupants behaviour [9]. These two benefits can ensure the transferability of RL to several buildings and provide the potential of a wide-spread implementation.

Recent studies evaluated RL on different aspects of buildings, such as joint control of thermal comfort and air quality [10], thermal comfort and electric vehicle charging [11] or lighting system [12]. Several studies have evaluated RL to provide a balance between occupants comfort and energy use in air conditioning systems. In these studies, RL agent is usually supposed to learn how to minimize the energy use while maintaining occupants comfort. For instance, Brandi et al. [13] investigated the application of RL to make a balance between energy use and comfort by a water-based space heating system in an office building. It was indicated that RL-based control provides 5% to 12% energy saving with an enhanced indoor temperature compared to the rule-based control. Considering that highly stochastic occupants behavior is a key challenge for efficient hot water production [3], few studies have evaluated model-based [14] and model-free RL [2] for occupant-centric hot water production. These studies indicated that while hot water use behavior of occupants is highly stochastic, RL can continuously learn and adapt the control strategy to the occupants' behavior and provide a significant energy saving. Although space heating and hot water production are usually combined, only few studies have evaluated RL for their combination. Lissa et al. [15] proposed a framework for optimal control of space heating and hot water system integrated to the PV solar panels. The proposed framework aimed to reduce the energy consumption by optimizing the operation of heat pump and maximizing the PV self-consumption, while keeping the occupants comfort. However, hot water use profiles were not used in this work and only random temperature drops for the tank were used to simulate a hot water demand.

Another key challenge in hot water systems is Legionella, which is a water-born bacteria that grows in the hot water with a temperature between 20 °C and 50 °C [16]. It be transferred to occupants by breathing in the contaminated water droplets and cause a respiratory disease [16]. To prevent the growth of this bacteria in the tank, conventional control methods usually maintain the tank temperature above 60 °C, which results in higher energy consumption [17]. To propose a realistic control method the hygiene aspect of hot water tank is very important to be considered. However, based on the literature review performed in this study, this aspect has been neglected in the previous studies on RL for hot water systems. Only a recent study by authors included the hygiene aspect in an RL framework for hot water production systems [2]. The hygiene aspect in this recent study was considered by following a simple rule stating that the hot water tank should be heated to 60 °C at least once a day [17].

The aim of this research is to develop an intelligent control framework that takes into account the stochastic hot water use behavior of occupants, and variating solar power production, and learns how to optimally operate the system to minimize energy usage while preserving the comfort and hygiene aspects. Case study energy system is the combined space heating and hot water production, assisted by Photovoltaic (PV) panels.

The main novelties of the proposed framework are:

Integration of water hygiene: While the pervious study by authors [2] followed a simple rule to respect hygiene aspect, this study for the first time integrates a temperature-based model that estimates the concentration of Legionella in hot water tank at each time step. Estimation of Legionella concentration in real-time enables the agent to spend as minimum energy as required for maintaining the hygiene aspect.

Investigation on real-world hot water use behavior: In this research, hot water demand of 3 residential houses is monitored to assess the performance of agent on real-world hot water use behavior of occupants.

Stochastic offline training to ensure occupants comfort and health: An RL control framework always starts with a training process, in which the agent starts a learning curve by interacting with the environment. As during this phase the agent is not experienced enough, and even has to perform random actions to explore the environment, it is very probable that it perform non-optimal actions that violates the comfort of occupants, or endanger their health by not maintaining the hygiene aspect of hot water. In practice, a control strategy that endangers the health of occupants would never be acceptable. Frequent violations of comfort aspect at the training phase also can reduce the satisfaction of occupants and their willingness to use the intelligent controller. To ensure that agent would quickly learn the optimal behavior with a minimum risk of violating comfort and hygiene aspects, an offline training phase is designed in this study. This offline phase integrates a stochastic hot water use model to emulate the realistic occupants' behavior. Also it includes a variety of climatic conditions and system sizes to provide a comprehensive experience to the agent.

The remainder of this paper is organized into four sections: Section 2 describes the research methodology, ssection 3 presents the results, and Section 4 concludes the paper.

2. Methodology

Fig. 1 shows the interactions of agent and environment in an RL problem. In RL, agent observes the current state (s_t) of the environment. According to the observed state, it selects an action and performs it on the environment (a_t). Due to this action, environment would transit to the next state (s_{t+1}), and agent would receive a reward (r_{t+1}) that quantifies how good was the performed action. The goal of agent is then to maximize the cumulative rewards during an entire episode. Therefore, different from Supervised Learning where a labelled dataset is required to train a model, in RL an interaction between agent and environment should be provided to enable the agent to learn optimal control strategy. In this section, the design of environment and agent are first explained. Then, the design of state, action, and reward space is described. Finally the monitoring campaign and training procedure are presented.

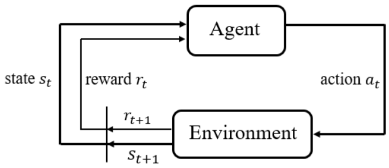

Fig. 1: Interaction of agent and environment in Reinforcement Learning [18]

2.1 Environment design

Layout of residential energy system in this study is shown in **Fig. 2**. This system uses an air-source heat pump to provide hot water in a tank, which is used for both hot water production and space heating through radiators. PV panels are also connected to the heat pump. PV panels are grid-connected, so the surplus power can be supplied to the grid. A dynamic model of the system is developed in TRNSYS.

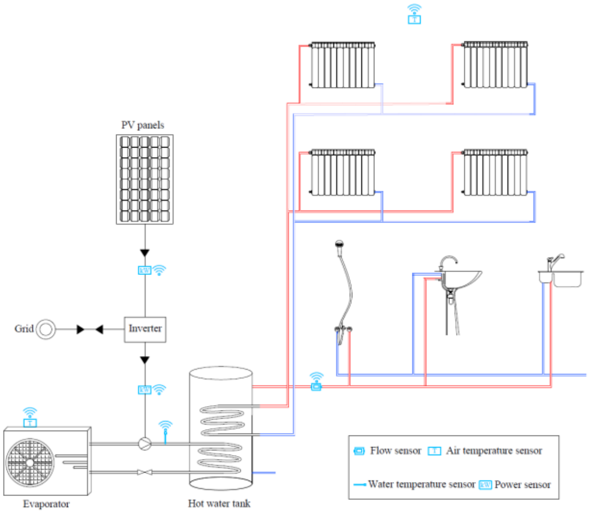

Fig. 2: Layout of solar-assisted space heating and hot water production system

2.2 Agent design

The agent is developed in Python using Tensorforce library [19]. An improved version of Deep Q-Network (DQN), known as Double DQN is used as it is proved to solve the issue of overestimation by typical DQN. Specifications of agent are provided in **Tab. 1**.

Tab. 1: Selected parameters for the agent

Parameter	Value
Learning rate	0.003
Batch size	24
Update frequency	4
Memory	48×168
Discount factor	0.9

2.3 State, action and reward space

Proper design of state, action and reward space is very important to obtain a good performance by RL framework. Parameters included in the state are presented in **Tab. 2**. Each parameter is a vector including the value of that parameter during one or multiple previous hours. The demand ratio is the ratio of total hot water demand of the current day until the current hour, to the total demand of the previous day. Hour of day is a value between 1-24 indicating what is the upcomming hour of day. Day of week, similarly, indicates the current day as a value between 1-7, where 1 represents Monday. The values are normalized to a vlue between 0 to 1.

Tab. 2: Parameters included in the state vector

Parameter	Length of look-back vector
Hot water demand	6
Demand ratio	-
Outdoor air temperature (°C)	1
Indoor air temperature (°C)	3
PV power (kW)	6
Heat pump outlet temperature (°C)	1
Legionella concentration (CFU/L)	1
Tank temperature (°C)	1
Hour of day	-
Day of week	-

Possible actions should allow the agent to exploit the possibility of energy storage in hot water tank and in thermal mass of building, by regulating the tank and indoor air temperatures. To this aim, at each hour agent can select between four possible actions: Turning ON the heat pump, Turning OFF the heat pump, selectig the indoor air temperature setpoint of 21 °C (as an energy-saving setpoint) or 23 °C (as an energy-storing setpoint). Based on the selected indoor air temperature setpoint by the agent, a two-point controller with a dead-band of 2 °C tries to maintain the specified setpoint during the next hour.

Reward function includes 4 different terms. An energy term to penalize the agent for net energy use, hot water comfort term to penalize the agent if a hot water demand is supplied with a temperature less than 40 °C, which is considered as the lower limit of comfort for hot water uses [2], space heating comfort term to penalize the agent if the indoor air temperature is out of the comfort region of 20 °C-24 °C, and a hygiene term if the estimated concentration of Legionella is above the maximum threshold of 500×10^3 CFU/L recommended for residential houses [20]. Equations 1-4 shows the formulation of energy, hot water comfort, space heating comfort, and hygiene terms.

$$R_{energy} = -a \times |HP_{power} - PV_{power}| \quad (1)$$

$$if\ T_{tank} \geq 40{:}\ R_{DHWcomfort} = 0\ else - b \quad (2)$$

$$if\ 20 \leq T_{indoor} \leq 24{:}\ R_{Indoorcomfort} = 0\ else - c \quad (3)$$

$$if\ Conc \leq Conc_{max}, R_{Hygiene} = 0\ else - d \quad (4)$$

Where HP_{power} and PV_{power} are the power use of heat pump and power production of PV panels (kW), T_{tank} and T_{indoor} are the tank and indoor air temperature, $Conc$ and $Conc_{max}$ are the current and maximum concentration of Legionella in the tank (CFU/L), R_{energy}, $R_{DHWcomfort}$, $R_{Indoorcomfort}$ and $R_{Hygiene}$. a, b, c and d are set to 1, 12, 10 and 10 determined by a sensitivity analysis. The total reward is therefore the summation of all these terms.

2.4 Monitoring campaign

It is challenging to directly test the performance of the proposed RL controller on a real-world residential building because if it fails to learn the behavior of occupants it can violate their comfort. On the other hand, it is important to investigate its performance on the realistic behavior of occupants. To perform a realistic test without disturbing the occupants, hot water use behavior of people is monitored and the collected data are used in TRNSYS model to emulate the real hot water demand. For the current framework, as shown in **Fig. 2**, only one single sensor at the tank outlet is enough to measure the hot water demand. In this study, to collect a comprehensive dataset which can be used also for future research, the hot and cold water demand is monitored at all the end uses in the case study buildings. The hourly hot water use data are then summed to represent the total hourly demand. **Fig. 3** shows the example of sensor installation on a faucet and a shower. LoRaWAN-based low power IoT flow sensors (with an integrated temperature sensor) are used to enable monitoring for a long duration (several months) with a single battery.

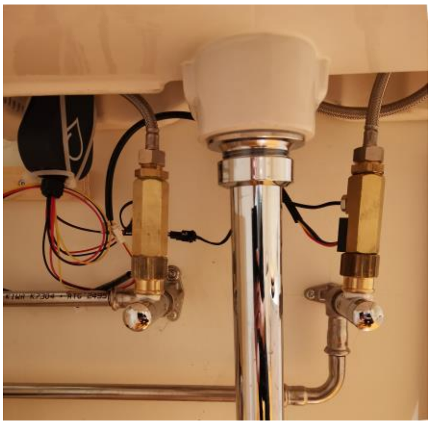

Fig. 3: Example of IoT flow and temperature sensor installation on a faucet

2.5 Training procedure

Interaction of agent and environment is provided by integrating Python and TRNSYS software. A Python code calls the TRNSYS simulation with the desired control actions, runs the simulation for a timestep, and reads the desired outputs to form the state and reward functions. Training and deployment stages are shown in **Fig. 4**. To ensure occupants' comfort and health, first, the agent is trained on an offline training process. In this stage, a virtual environment is provided to enable the agent to gain enough experience before being implemented on the target house. In this stage, a hot water use model [21] is used to emulate the hot water use behavior of occupants. This model is developed based on data from 77 residential buildings, and therefore is a great tool to provide a prior experience to the agent. In this stage, it is desired that the agent also experience a good variety which can help it to generalize its knowledge and quickly adapt to new cases. To this aim, the offline training phase is repeated for 10 years, and at each year the agent is interacting with a different size of the system (heat pump, tank, building area, etc), the weather condition of a different city, and a different hot water use profile generated by the stochastic model. After these 10 years, the agent is then trained on the target house for 16 weeks. The aim of training on the target house is to let the agent adapt to the specific characteristics of the target house, such as occupants' behavior, systems sizes, or weather conditions. To simulate the target house, in online training stage the collected hot water use data, and also the weather data collected from a weather station near the case study is used. After the online training on the target house, the training process can be stopped and the agent starts the deployment stage, in which agent is no longer learning but only controlling the system. Although to take the full power of RL the training process should be always continued, it makes it computationally expensive and agent should be always on the cloud. But once the training is stopped, the saved agent can be uploaded on a cheap hardware and control the system locally. Duration of deployment phase is 4 weeks.

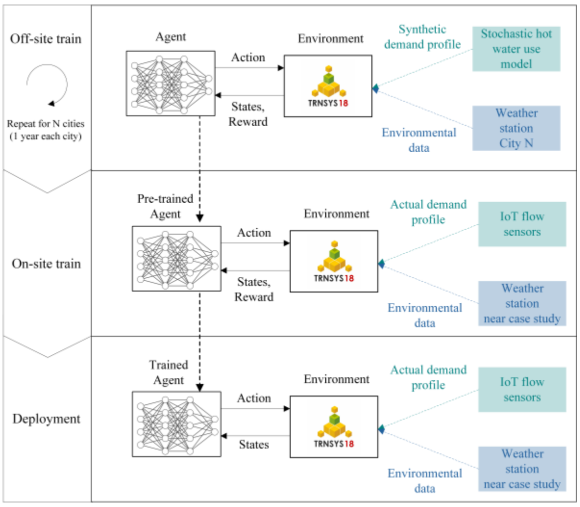

Fig. 4: Training and deployment process

3. Results

The stability of the reward function indicates that the agent is converged to an optimal control policy. **Fig. 5** shows the evolution of reward function during the offline training phase, and also online training on each of the case studies. $\varepsilon - greedy$ method with a linear decay is used to impose exploration at the 12 first weeks of the offline training phase. During the last 90 weeks of the offline training stage, the reward function is almost stable. During the online training on the target houses, the reward function is stable since the very first week. The existing variations compared to the potential variations of reward function are very small and mainly due to the energy use, which is not avoidable. It shows that the offline training stage has provided a generalizable experience for the agent, and since the beginning of implementation on the target houses, the agent can provide energy saving while maintaining comfort and hygiene aspects.

Fig. 6 shows the control signal, PV power production, and tank temperature over the deployment stage on three case studies. The deployment stage of houses 1 and 2 is during December, while the deployment stage of house 3 is during July. Therefore the PV power production of the third case study is higher than others. In all of the case studies, it can be seen that the agent is trying to adapt the control signal to the PV power production and reduce the power use from the grid, by turning ON the heat pump more frequently during the hours of PV power production. This adaptation can be seen very well on house 3, where PV power production is significantly higher and the agent tries to turn ON heat pump only when there is a PV power production. In all of the case studies, the agent has learned how to keep tank temperature above 40 °C to respect the comfort of occupants. It shows that agent could successfully learn and adapt to the occupants behavior, because none of the demands reduced the tank temperature below 40 °C.

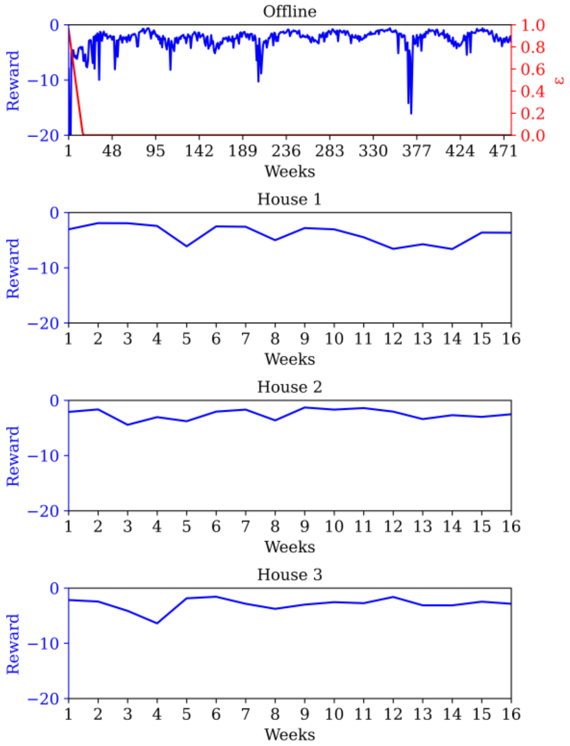

Fig. 5: Evolution of reward function over the offline training phase and case studies

To compare the performance of the proposed framework with the conventional methods, two conventional scenarios are modeled. The first method is Rule-based Conventional control (RC), where the tank temperature setpoint is 60 °C with a dead-band of 10 °C. The second scenario is Rule-based Energy-saving control (RE), in which with the aim of reducing energy usage the tank temperature setpoint is considered as 50 °C. This method might not be practical because a constant setpoint of 50 °C for the tank can still impose the risk of Legionella growth. But this method is included in comparison to represent an extreme case of energy-saving by a rule-based method and to prove that the better performance of RL is not only due to a lower tank temperature. Main performance metrics over the deployment stage are shown in **Tab. 3**. In all the houses, RL has provided energy saving compared to the RC and RE methods. As expected, energy saving compared to the RC is higher than RE, because heat pump COP is RC is lower than RE. The energy saving in the houses 2 and 3 are around 7% and 8%, which represent the order of potential energy saving by RL method duirng the cold season. Energy savings in house 3 are much more, because PV power production in the hot season is much more an agent learns how to get the best use of PV power production to cover space heating and hot water energy use. This is a great example to highlight the importance of a controller that can learn and adapt to the changes, over a static and rigid controller. As expected, the comfort of occupants in case of space heating is always preserved because the possible choices for agent has been in the comfort range, and the tank temperature has been always high enough to provide the specified setpoint. The hygiene aspect is also preserved in all the houses because the maximum Legionella concentration in the tank in all the cases is below 5×10^5 CFU/L.

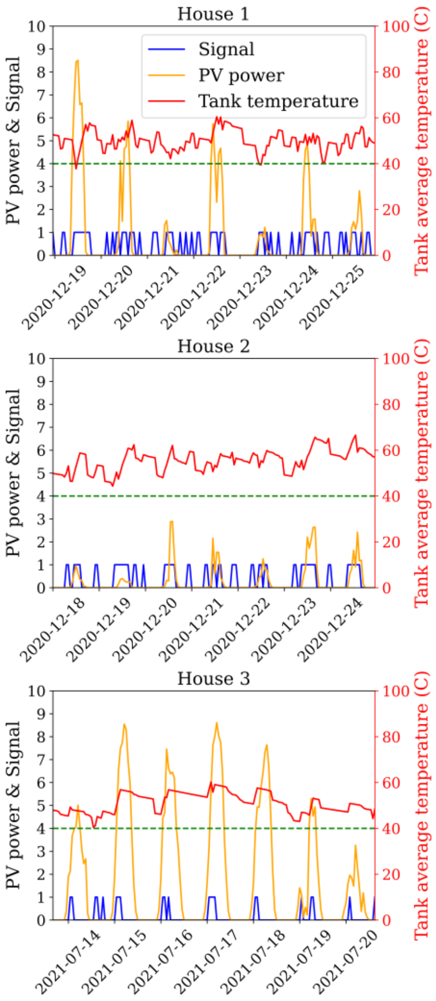

Fig. 6: Control signal versus PV power production and tank temperature

Tab. 3: Main performance metrics over different case studies

	House 1	House 2	House 3
Energy saving to RC (%)	28.9	40.4	75.7
Energy saving to RE (%)	7.2	8.7	61.6
Violated DHW comfort (%)	8.1	5	1.7
Average temperature of DHW comfort violations (°C)	38.8	39	37.98
Maximum Legionella concentration (CFU/L)	2060	49704	6764

To better highlight how RL could better exploit solar power production, the contribution of PV power production in the total power use of the heat pump is shown in **Fig. 7**. As can be seen, in all the case studies RL has used a higher contribution of PV power, compared to the RC and RE. In case of house 3, the contribution of PV power production is much higher than RC and RE, which is the why in this house the energy saving is much higher than other houses. It shows that a significant advantage of the proposed RL framework is to learn how to adapt the operation to the PV power production, and therefore potential energy-saving increases in regions with higher solar radiation.

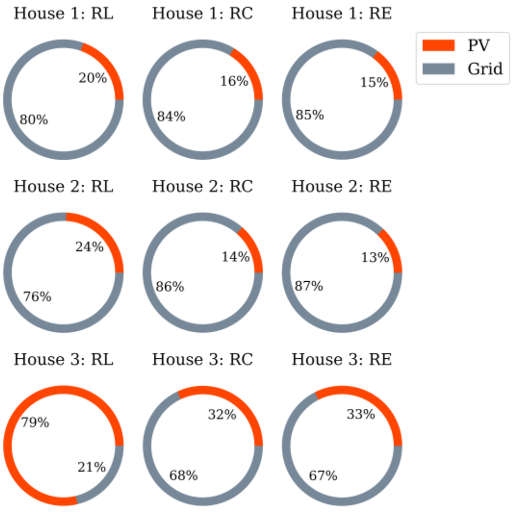

Fig. 7: Contribution of PV power production in power consumption of heat pump

4. Conclusion

Optimal energy management in the buildings is affected by several stochastic parameters, such as weather conditions, occupants' behavior, or solar energy, that vary by time and are hard to predict. It is therefore challenging and time-consuming to develop a rule-based or model-based control logic that can properly take into account all these parameters and maintain an optimal operation. Rather, a learning ability can be provided to the controller, so in each specific building, it can learn these stochastic parameters and continuously adapt the system operation to their variations.

This research proposed a model-free RL control framework that can learn the hot water use behavior of occupants and PV power production, and accordingly adapt the system operation to meet the comfort requirements with minimum energy use. Different from previous studies, where RL is supposed to make a balance between energy use and comfort, in this study RL tries to make a balance between energy use, comfort, and hygiene. Inclusion of hygiene aspect is very crucial to ensure the health of occupants. Real-world hot water use data is monitored in three residential case studies and used to evaluate the performance of the proposed framework over the realistic behavior of occupants. The RL framework is compared with two rule-based scenarios of RC and RE.

Results indicate the proposed framework could provide a significant energy saving, mainly by learning how to get the best use of PV power production. Therefore the energy-saving potential is expected to be even more in regions with higher solar radiation than Switzerland. Also, the agent has successfully learned how to respect the comfort of occupants and water hygiene, so the potential energy saving is not with the cost of violating occupants' comfort or health.

Data statement

The datasets measured during the current study are not publicly available but can be shared based on request.

5. References

[1] Esrafilian-Najafabadi M, Haghighat F. Occupancy based HVAC control systems in buildings: A state-of-the-art review. Build Environ2021;197:107810.

[2] Heidari A, Marechal F, Khovalyg D. An adaptive control framework based on Reinforcement learning to balance energy, comfort and hygiene in heat pump water heating systems. J Phys Conf Ser 2021;2042:012006.

[3] Heidari A, Olsen N, Mermod P, Alahi A, Khovalyg D. Adaptive hot water production based on Supervised Learning. Sustain Cities Soc2021;66:102625.

[4] International Renewable Energy Agency. REmap 2030 Full Report. 2014.

[5] Mohammed NA, Baghdad-Iraq M. Modelling and Optimisation Planning of the Dynamic System of Energy Supply-Integrating Demand-Side Management and Forecasting. Opus4KobvDe 2019:xvi, 188 Seiten.

[6] Haji Hosseinloo A, Ryzhov A, Bischi A, Ouerdane H, Turitsyn K, Dahleh MA. Data-driven control of micro-climate in buildings: An event-triggered reinforcement learning approach. Appl Energy 2020;277:115451.

[7] Sutton R, Brato A. Reinforcement learning: an introduction. Robotica 1999;17.

[8] Zou Z, Yu X, Ergan S. Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network. Build Environ 2020;168:106535.

[9] Schreiber T, Eschweiler S, Baranski M, Müller D. Application of two promising Reinforcement Learning algorithms for load shifting in a cooling supply system. Energy Build 2020;229:110490.

[10] Valladares W, Galindo M, Gutiérrez J, Wu WC, Liao KK, Liao JC, et al. Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm. Build Environ 2019;155:105–17.

[11] Svetozarevic B, Baumann C, Muntwiler S, Di Natale L, Zeilinger MN, Heer P. Data-driven control of room temperature and bidirectional EV charging using deep reinforcement learning: Simulations and experiments. Appl Energy 2021:118127.

[12] Park JY, Dougherty T, Fritz H, Nagy Z. LightLearn: An adaptive and occupant centered controller for lighting based on reinforcement learning. Build Environ 2019;147:397–414.

[13] Brandi S, Piscitelli MS, Martellacci M, Capozzoli A. Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings. Energy Build 2020;224:110225.

[14] Kazmi H, Mehmood F, Lodeweyckx S, Driesen J. Gigawatt-hour scale savings on a budget of zero: Deep reinforcement learning based optimal control of hot water systems. Energy 2018;144:159–68.

[15] Lissa P, Deane C, Schukat M, Seri F, Keane M, Barrett E. Deep reinforcement learning for home energy management system control. Energy AI 2021;3:100043.

[16] Springston JP, Yocavitch L. Existence and control of Legionella bacteria in building water systems: A review. J Occup Environ Hyg2017;14:124–34.

[17] Booysen MJ, Engelbrecht JAA, Ritchie MJ, Apperley M, Cloete AH. How much energy can optimal control of domestic water heating save? Energy Sustain Dev 2019;51:73–85..

[18] QU X, Chang Q, Arinez J, Zou J. Knowledge-guided Reinforcement Learning for Gantry Work Cell Scheduling. IEEE Access 2018.

[19] Tensorforce: a TensorFlow library for applied reinforcement learning n.d.

[20] Standard SIA 385/1. n.d.

[21] Ritchie MJ, Engelbrecht JAA, Booysen MJ. A probabilistic hot water usage model and simulator for use in residential energy management. Energy Build 2021;235:110727. 10727.