Comparative Analysis of Model-Free and Model-Based HVAC Control for Residential Demand Response

Kuldeep Kurte
Oak Ridge National Laboratory, USA
kurtekr@ornl.gov

Kadir Amasyali
Oak Ridge National Laboratory, USA
amasyalik@ornl.gov

Jeffrey Munk
Oak Ridge National Laboratory, USA
munkjd@ornl.gov

Helia Zandi
Oak Ridge National Laboratory, USA
zandih@ornl.gov

ABSTRACT

In this paper, we present a comparative analysis of model-free reinforcement learning (RL) and model predictive control (MPC) approaches for intelligent control of heating, ventilation, and air-conditioning (HVAC). Deep-Q-network (DQN) is used as a candidate for model-free RL algorithm. The two control strategies were developed for residential demand-response (DR) HVAC system. We considered MPC as our golden standard to compare DQN's performance. The question we tried to answer through this work was, ***What % of MPC's performance can be achieved by model-free RL approach for intelligent HVAC control?***. Based on our test result, RL achieved an average of ≈ 62% daily cost saving of MPC. Considering the pure optimization and model-based nature of MPC methods, the RL showed very promising performance. We believe that the interpretations derived from this comparative analysis provide useful insights to choose from various DR approaches and further enhance the performance of the RL-based methods for building energy managements.

CCS CONCEPTS

• **Computing methodologies → Reinforcement learning**; • **Hardware → Power and energy**.

KEYWORDS

deep reinforcement learning, model predictive control, intelligent HVAC control, demand response, building energy

ACM Reference Format:
Kuldeep Kurte, Kadir Amasyali, Jeffrey Munk, and Helia Zandi. 2021. Comparative Analysis of Model-Free and Model-Based HVAC Control for Residential Demand Response. In *The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '21), November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3486611.3488727

BuildSys '21, November 17–18, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9114-6/21/11…$15.00
https://doi.org/10.1145/3486611.3488727

1 INTRODUCTION

Buildings are one of the major consumers of energy in the world, consuming over 36% of the total global energy [9]. In the U.S., more than 50% of buildings' energy use is attributed to HVAC [7]. HVAC energy use is highly dependent on the weather, causing non-uniform power demand throughout the day. Efficiently generating power to match the load during peak demand periods is challenging. To help curtail power use during peak demand periods, many utilities offer time-of-use (TOU) price structures that incentivize power use during off-peak hours with a low electricity rate and penalize power use during on-peak hours with a high electricity rate. This type of rate structure can illicit a reduction in demand during peak power periods, decreasing the generation capacity required to meet the load.

Along with demand reduction, cost reduction and home owner's comfort are important objectives of intelligent HVAC control, making it a multi-objective problem. Various methods for intelligent control of the HVAC system can be broadly divided into the following three categories:

- **Rule-based HVAC control**: Rule-based approach uses rules encoded by experts to control HVAC [17]. A recent work used a rule-based approach to reduce the operating cost of a thermal storage tank [8]. Despite of the simplicity of the approach, it is difficult to attain the optimal performance. Moreover, the expert rules devised for a particular building cannot be directly used for the control in other buildings. Furthermore, the rule-based methods cannot solve the multi-objective problems.
- **Model-based HVAC control**: Model-based approaches for HVAC control requires a model of building's thermal dynamics and HVAC performance. MPC is one of the well studied model-based approaches for HVAC control [19]. MPC needs accurate building and HVAC model for predicting building's thermal behavior and energy performance over a finite horizon and performs optimization of the control variables. Additionally, running building models along with the optimization algorithms is a computationally expensive task which limits its practical applicability. Using simplified building models can help improving the computational efficiency of the model-based approaches. However, such models may suffer from lack of accuracy.
- **Data-driven approaches for HVAC control**: Recently, the data-driven model-free approaches such as RL have been explored for optimal control [1–3, 5, 6, 12]. In this approach,

an agent interacts with the environment and learns from its experience, which sometimes makes learning slow. Further, RL can not be deployed from scratch due to the risk of compromising the home owner's comfort. To overcome this issue, most of the work utilized simulated environment to pre-train the RL agent. Once trained, RL agent can be deployed in the real environment.

Paucity of comparative analysis: Most of the previous works discuss model-based and model-free approaches separately. Recently, Kou et al., 2021 [10] compared MPC and RL based HVAC control. There is still a need for such comparative analysis to evaluate the feasibility of model-based and model-free approaches in different scenarios.

Contributions: Our contributions in this paper are as follows: 1) We developed DQN (model-free) and MPC based intelligent HVAC control to minimize its energy cost and the occupant discomfort. 2) We compared and interpret their performance with the fixed set point baseline as well as with each other. The comparative analysis of these two methods will contribute to the current literature of the intelligent HVAC control and provide useful insights to evaluate their feasibility in different scenarios.

Workshop relevance: We used DQN as our model-free RL algorithm and compared it with the model-based MPC approach. The insights drawn from this comparative analysis will be beneficial to improve the RL's performance in building energy management and will provide guidelines to the decision makers on choosing appropriate approach for building energy management.

The rest of the paper is organized as follows: Section 2 and Section 3 describe the RL-based and MPC-based HVAC control development followed by their comparative analysis in Section 4. Finally, Section 5 concludes the paper by highlighting the insights gained from this comparative study.

2 RL-BASED HVAC CONTROL

We developed an end-to-end framework to train intelligent HVAC control using DQN [12] using the procedure described by Mnih et al., 2015 [13] and Wei et al., 2017 [18]. We trained DQN-agent in an offline mode using the building simulation developed from the data collected from a real research house located in Knoxville, TN. We employed a grey-box model that used an electrical network equivalent model to represent the thermal resistance (R) and capacitance (C) associated with the air, walls, and internal mass for each zone. The model includes the effect of outdoor air temperature, sol-air temperature, direct solar radiation, air conditioner cooling, and internal heat load. More details can be found in [4].

During offline training, DQN-agent interacts with the simulated environment at every 15 mins., collects the observations (i.e., state (S_t), performs some action (A_t), and receives reward (r_{t+1}). After executing the action A_t, the environment will transition to the next state (S_{t+1}). The current state S_t of the environment is defined by $S_t = \{t, T^{in}_t, T^{out}_{t+lookahead}, P_{t+lookahead}, P_d\}$. Where, t is the time of the day, T^{in}_t is the indoor temperature, $T^{out}_{t+lookahead}$ represents current outdoor temperature and six hours of a lookahead, $P_{t+lookahead}$ is the current price and price look ahead, and P_d represents few features derived from the price lookahead [11]. A set of discrete setpoints {70°F, 71°F, 72°F, 73°F, 74°F} were used. Effectively, there are a total of 25 actions in the action space as the building has two zones. This action space avoided DQN-agent setting setpoints outside the comfort range of occupants and allowed DQN-agent to focus on monetary savings. The reward function used for training DQN-agent include only the cumulative cost of running HVAC for 15 minutes, i.e. reward=-cost. Various parameters of the DQN training are given in the Table 1. Figure 2 shows the DQN training performance in terms of the total episodic reward and cost over the training session. We observed satisfactory training performance.

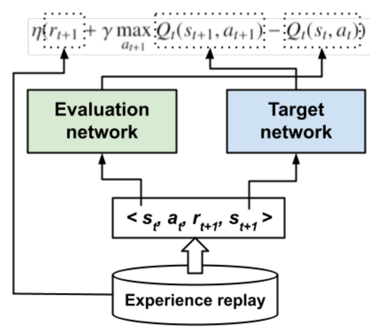


Figure 1: DQN training setup

Table 1: DQN training parameters

Parameter	Value
Episodes	100
Simulation step (Δt_s)	1 min
Control step (Δt_c)	15 min
Learning rate	0.0001
Optimizer	Adam
Reward decay (γ)	0.9
ϵ-greedy value	0.9
Target replacement iterations	200
Initial steps	1440
Batch size	64
Experience replay memory size	20,000
[LT, UT]	[70 °F, 74 °F]

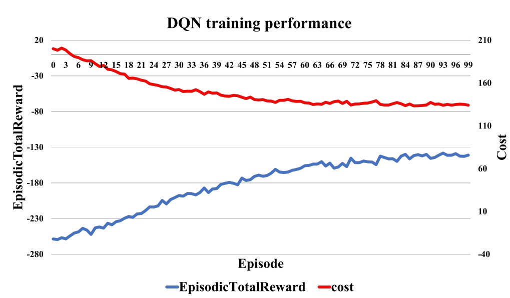


Figure 2: DQN training performance

3 MPC-BASED HVAC OPERATION

We compared the performance of the RL agent against an MPC-based controller's performance for a month. For MPC-based controller, we used the same building simulation as the RL. Figure 3 illustrates the concept of a model-based predictive controller which consists of three main steps.

First, the MPC-based controller gets the measurements from the system and collects the electricity price and weather data for the corresponding prediction horizon. In our case, these measurements include indoor temperatures for both zones of the building. The prediction horizon is set to 6 hours, and therefore electricity price and weather data for the next 6 hours are gathered.

Second, the control variables are optimized for the prediction horizon using the initial indoor temperatures taken from the system, the electricity price, and the weather data for 6 hours. Here, the objective of the optimization is to minimize electricity cost and the control variables included the thermostat setpoints for both zones in 15 minute intervals. In order to ensure the thermal comfort of the occupant, the control variables were constrained between 70°F and 74°F. Because the building model that we utilize to simulate indoor temperature and power consumption is nonlinear and non-convex, the optimization problem was solved using genetic algorithms.

Third, the optimal control variables resulting from the optimization are implemented as much as the control horizon. We set the control horizon to 1 hour and therefore the first hour of the optimal variables are implemented and the control proceeds for an hour. After that the procedure is repeated by taking the new measurements from the system until the end of the control period.

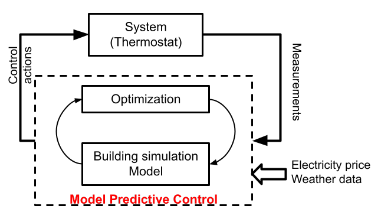


Figure 3: Overview of the MPC implementation

4 COMPARATIVE ANALYSIS OF RL AND MPC FOR HVAC CONTROL

The main objective of this work is to compare the MPC-based and RL-based HVAC control and derive the insights that are helpful to the control optimization community in evaluating the feasibility of these approaches in different scenarios.

4.1 Dataset used

We used weather data from a Typical Meteorological Year (TMY) data for the cities of Knoxville, TN, USA and Memphis, TN, USA [14, 15]. We used TMY weather data of Memphis, TN for the months of May–August for training RL. We compared the performance of RL and MPC using Knoxville, TN TMY data. We used a TOU price signal with a peak price of \$0.25/kWh during 2pm–7pm and an off-peak price of \$0.05/kWh.

4.2 Indoor temperature variations

Figure 4 shows the indoor temperature variations in both the zones for MPC and RL on July 10, 11, and 12 with Knoxville, TN weather data. We can visually see the commonalities and differences in the indoor temperature of both the zones due to the actions performed by MPC and RL. We observed that the indoor temperature variations due to both MPC and RL matched closely when the outdoor temperature is not too high, e.g. on July 11 and 12. We also observed some mismatch in the indoor temperature variations, specifically, during hot days, e.g. on July 10. In fact, RL tried to pull the indoor temperature down when the outdoor temperature was high but could not match the MPC's performance. The potential reason for this behaviour could be the scarcity of such hot days in the training data. We found that the average outdoor temperature in the training data (Memphis, TN) was ≈ 81°F. One potential way to overcome this scarcity of the less frequent samples is to prioritize such tuples in the replay memory to increase their chance to get selected during the training process [16].

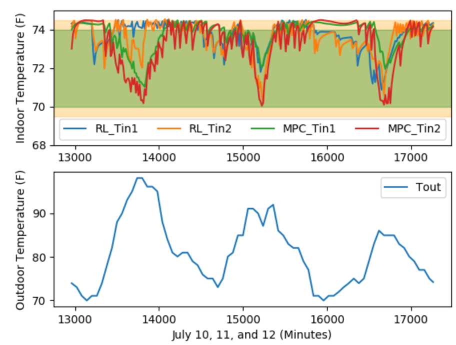
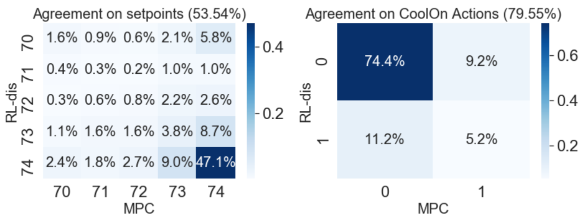


Figure 4: Indoor temp. for RL and MPC for Jul. 10–12. Comfort band [70–74°F] (green) + half deg. of deadband (orange)

4.3 Agreement of MPC's and RL's actions

We investigated the agreement of MPC's and RL's actions in terms of setpoints (Figure 5a) and AC's cooling status (Figure 5b). For brevity, we present here in Figure 5 the combined agreement on actions taken by these algorithms in both zones. We found that MPC and RL showed ≈ 53.4% agreement on setpoints and most of the agreement was on setting higher setpoints. The high agreement at the highest set point is expected, since it will yield the lowest AC energy use. Optimally, the system should only be set to lower set points when pre-cooling prior to peak price periods to limit energy use during that period. The relatively low level of agreement between setpoints used by RL and MPC can be attributed to the fact that there are potentially several different setpoint combinations that can illicit the same response from the HVAC system, (e.g., setting the thermostat to 70°F or 71°F when the indoor temperature is 74°F will illicit the same "on" response from the HVAC system). Accordingly, we compared AC's cooling status for the RL and MPC controls, which showed better agreement than the setpoint comparison. We found that MPC and RL showed a combined agreement of ≈ 79.5% on AC's cooling status out of which ≈ 74.4% agreement was on keeping AC off. This behavior can be attributed to the cost savings by shutting off the AC. These results show that a better agreement on the actions could be achieved if fewer discrete setpoints, e.g. [70°F, 74°F] were used.

We further interpreted the % agreement of MPC and RL's actions with respect to varying outdoor temperature and electricity price. Referring to Table 2, we found the highest agreement in the setpoints set by MPC and RL when the outdoor temperature was



(a) % agreement on setpoints **(b) % agreement on AC's status**

Figure 5: % agreement on actions by MPC and RL.

in the range of 71°F–80°F. The high % agreement in AC's cooling status was found when the outdoor temperature was less than 80°F. Similarly, % agreement of setpoint and AC status were found to be ≈ 53% and ≈ 79% respectively, which are consistent with the overall agreement shown in Figure 5.

Table 2: % agreement on actions by MPC and RL for varying outdoor temperature (T_{out}) and electricity price

		Setpoint	AC status
$T_{out}(F)$	**51-60**	6.9%	**93.1%**
	61-70	46.7%	92.9%
	71-80	**69.9%**	87.8%
	81-90	45.5%	65.3%
	91-100	26.5%	57.2%
Price	**Low ($0.05/kWh)**	53.6%	79.7%
	High ($0.25/kWh)	53.2%	79.1%

As mentioned earlier, the main cost saving is achieved by pre-cooling before the peak price period and shutting off the AC during the peak period. We further compared the energy consumption by MPC and RL, specifically during the peak period, as an indicator of the agreement of the actions during peak period. We can see that in Table 3, MPC consumed ≈ 20kWh less energy than RL which contributes to the lower cost consumption by MPC. This shows that there is still a scope for RL agent to reduce the cost savings by saving more during the peak period. This observation can be used to tweak the parameters in the training setting. However, the energy consumption of both MPC and RL are still comparable.

Table 3: Comparison of energy consumption

Energy (kWh)	Low price 0.05/kWh	Peak price 0.25/kWh	Total
RL	238.9	101.9	340.4
MPC	237.4	81.6	319.0

4.4 MPC and RL's cost saving over baseline

We collected the daily cost of operating MPC and calculated its cost reduction over a fixed setpoint baseline (74°F), and plotted along with the cost reduction of RL models over the baseline (refer to the bottom plots of the Figures 6). We can see that the daily cost savings of RL model is comparable with MPC. Further, we computed a quantity representing the % of MPC's cost saving over baseline achieved by RL. The tops bar-charts in the Figure 6 represent this quantity for each day in the month of July. The results are encouraging except three days where RL showed high cost than baseline and MPC. We observed this behavior of RL on July 2, 3, and 21. To analyze this behavior we compared the daily average outdoor temperature (shown in the bottom graph). We found that, the outdoor temperature on July 2–3, and July 21 was slightly lower than the average summer temperature (shown by the dips in the TOUT line graph in Figures 6). Additionally, the building simulation model used for training RL has only cooling capability so RL learnt only the cooling behavior. This could be a possible reason for the RL's poor performance on these four days. The behaviour is reflected as negative % of cost saving as shown by downwards bars in Figure 6.

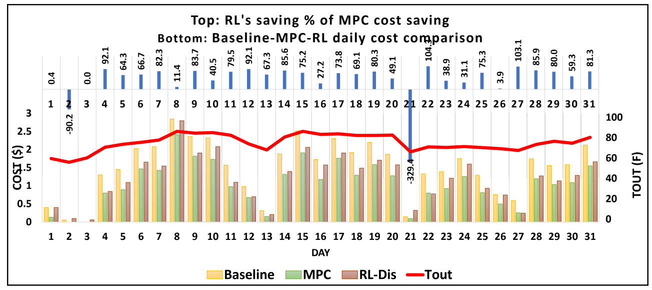


Figure 6: Daily cost saving by MPC and RL over baseline and MPC's % cost saving over baseline achieved by RL

We further calculated the mean % cost saving of MPC that was achieved by RL. We eliminated these four days from the calculation. We found that, RL achieved an average of ≈ 62% daily cost saving of MPC. Further, the total monthly cost of operation by MPC was $32.27 and showed 30.40% of cost savings over baseline. Whereas RL's monthly cost was $37.33 (19.48% cost saving over baseline). Considering the pure optimization and model-based nature of the MPC method, the RL showed very promising performance.

4.5 Run-time analysis

We optimized MPC model for 31 days of July in a simulation on a desktop with Intel i7-8700 CPU @3.20GHz and 16GB of main memory. MPC took ≈ 7 days to complete the optimization of 31 days. The RL offline training was done on desktop with Intel i7-7820HQ CPU @2.90GHz and 16 GB of main memory which took ≈3 hours. Once trained, the pre-trained RL model was used to make decisions which took less than ≈1 minute to simulate the 31 days.

5 CONCLUSION

We presented a comparative analysis of MPC (i.e. model-based) and RL based (i.e. data-driven) HVAC control. Model-based method needs an accurate model of the environment whereas data-driven approach interacts with the environment and learns from the experiences. Although in our experiment MPC showed more cost reduction than RL, considering a pure optimization and model-based nature of the MPC approach, RL showed very promising performance. From our comparative analysis we derived following conclusions:

- RL's performance can be improved by prioritizing the tuples that are less frequent. For instance, extreme hot or cold days.
- The fewer setpoints that are sufficient to change the AC's status must be used to formulate the action space.
- The runtime comparison of MPC and RL clearly shows the advantages of the data-driven approaches over model-based methods for optimal control. (refer Section 4.5)

ACKNOWLEDGMENTS

This work was funded by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Building Technology Office under contract number DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

[1] Ki Uhn Ahn and Cheol Soo Park. 2020. Application of deep Q-networks for model-free optimal control balancing between different HVAC systems. *Science and Technology for the Built Environment* 26, 1 (2020), 61–74.

[2] Enda Barrett and Stephen Linder. 2015. Autonomous HVAC Control, A Reinforcement Learning Approach. In *Machine Learning and Knowledge Discovery in Databases*, Albert Bifet, Michael May, Bianca Zadrozny, Ricard Gavalda, Dino Pedreschi, Francesco Bonchi, Jaime Cardoso, and Myra Spiliopoulou (Eds.). Springer International Publishing, Cham, 3–19.

[3] Yujiao Chen, Leslie K Norford, Holly W Samuelson, and Ali Malkawi. 2018. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings* 169 (2018), 195–205.

[4] Borui Cui, Cheng Fan, Jeffrey Munk, Ning Mao, Fu Xiao, Jin Dong, and Teja Kuruganti. 2019. A hybrid building thermal modeling approach for predicting temperatures in typical, detached, two-story houses. *Applied energy* 236 (2019), 101–116.

[5] Yan Du, Fangxing Li, Jeffrey Munk, Kuldeep Kurte, Olivera Kotevska, Kadir Amasyali, and Helia Zandi. 2021. Multi-task deep reinforcement learning for intelligent multi-zone residential HVAC control. *Electric Power Systems Research* 192 (2021), 106959.

[6] Yan Du, Helia Zandi, Olivera Kotevska, Kuldeep Kurte, Jeffery Munk, Kadir Amasyali, Evan Mckee, and Fangxing Li. 2021. Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. *Applied Energy* 281 (2021), 116117.

[7] EIA. 2015. Use of energy explained Energy use in homes (U.S. Energy Information Administration (EIA), 2015). https://www.eia.gov/energyexplained/use-of-energy/homes.php

[8] Jingfan Hu, Wandong Zheng, Sirui Zhang, Hao Li, Zijian Liu, Guo Zhang, and Xu Yang. 2021. Thermal load prediction and operation optimization of office building with a zone-level artificial neural network and rule-based control. *Applied Energy* 300 (2021), 117429. https://doi.org/10.1016/j.apenergy.2021.117429

[9] IEA. 2020. IEA. Global Status Report for Buildings and Construction 2020. https://globalabc.org/sites/default/files/inline-files/2020BuildingsGSR_FULLREPORT.pdf

[10] Xiao Kou, Yan Du, Fangxing Li, Hector Pulgar-Painemal, Helia Zandi, Jin Dong, and Mohammed M Olama. 2021. Model-Based and Data-Driven HVAC Control Strategies for Residential Demand Response. *IEEE Open Access Journal of Power and Energy* 8 (2021), 186–197.

[11] Kuldeep Kurte, Jeffrey Munk, Kadir Amasyali, Olivera Kotevska, Borui Cui, Teja Kuruganti, and Helia Zandi. 2020. Electricity Pricing aware Deep Reinforcement Learning based Intelligent HVAC Control. In *Proceedings of the 1st International Workshop on Reinforcement Learning for Energy Management in Buildings & Cities*. 6–10.

[12] Kuldeep Kurte, Jeffrey Munk, Olivera Kotevska, Kadir Amasyali, Robert Smith, Evan McKee, Yan Du, Borui Cui, Teja Kuruganti, and Helia Zandi. 2020. Evaluating the Adaptability of Reinforcement Learning Based HVAC Control for Residential Houses. *Sustainability* 12, 18 (2020). https://doi.org/10.3390/su12187727

[13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.

[14] National Renewable Energy Laboratory. 2015 (accessed May 17, 2019). *National Solar Radiation Data Base, 1991–2005 Update: Typical Meteorological Year 3.* https://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/data/tmy3/723260TYA.CSV.

[15] National Renewable Energy Laboratory. 2015 (accessed May 17, 2019). *National Solar Radiation Data Base, 1991–2005 Update: Typical Meteorological Year 3.* https://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/data/tmy3/723260TYA.CSV.

[16] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. 2015. Prioritized experience replay. *arXiv preprint arXiv:1511.05952* (2015).

[17] Pervez Hameed Shaikh, Nursyarizal Bin Mohd Nor, Perumal Nallagownden, Irraivan Elamvazuthi, and Taib Ibrahim. 2014. A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable and Sustainable Energy Reviews* 34 (2014), 409 – 429. https://doi.org/10.1016/j.rser.2014.03.027

[18] Tianshu Wei, Yanzhi Wang, and Qi Zhu. 2017. Deep Reinforcement Learning for Building HVAC Control. In *Proceedings of the 54th Annual Design Automation Conference 2017* (Austin, TX, USA) *(DAC âĂŹ17)*. Association for Computing Machinery, New York, NY, USA, Article 22, 6 pages. https://doi.org/10.1145/3061639.3062224

[19] Ye Yao and Divyanshu Kumar Shekhar. 2021. State of the art review on model predictive control (MPC) in Heating Ventilation and Air-conditioning (HVAC) field. *Building and Environment* (2021), 107952.