Comparison of Data-driven Model-based and Model-free Approaches for Unlocking Building Energy Flexibility

Anjukan Kathirgamanathan\textsuperscript{1,2}, Eleni Mangina\textsuperscript{2,3}, Donal P. Finn\textsuperscript{1,2}
\textsuperscript{1}School of Mechanical and Materials Engineering, University College Dublin, Ireland
\textsuperscript{2}UCD Energy Institute, O’Brien Centre for Science, University College Dublin, Ireland
\textsuperscript{3}School of Computer Science, University College Dublin, Ireland

Abstract
Commercial buildings are significant end-use energy consumers and given their inherent thermal mass and use of advanced control infrastructure together with heating, ventilation and air conditioning systems, have the potential to offer significant load shifting opportunities. Data-driven control frameworks show promising results as a robust and scalable technique for the heterogeneous building stock that will enable automated demand response whilst ensuring occupant thermal comfort is maintained. This research compares two different approaches to the problem. First, a model based approach is considered and is denoted Data Predictive Control with Ensemble methods (DPC-En). This approach is based on the use of the random forest predictor with the ‘separation of variables’ technique to deliver an optimal control problem. Second, a model free approach is considered, which utilises a soft actor critic deep reinforcement learning (SAC DRL) algorithm. The DPC-En technique was able to minimise energy costs by 14.0\% compared to the baseline rule-based control. The SAC DRL similarly achieved 10.7\% savings compared to the baseline. Both techniques are able to respect occupant thermal comfort constraints.

Key Innovations
- Comparison of model-based data-driven control and model-free reinforcement learning for intelligent HVAC load control in a commercial building for automated demand response.
- Novel application of soft actor-critic deep reinforcement learning algorithm.

Practical Implications
Model-based optimal control approaches are recommended where dynamics are known (e.g., one-zone building or battery). Model-free approaches are recommended where dynamics are more complex and/or where historical data does not exist.

Introduction
This research investigates automated and intelligent data-driven control frameworks for commercial buildings to unlock building energy flexibility as part of a smart and sustainable electrical grid. This is motivated by a desire to increase penetration of variable renewable energy sources (RES) reducing the carbon footprint of generation, deferring infrastructure upgrades and minimising the increasing volatility of electricity prices. The flexibility to manage any mismatch in supply (from variable RES) and demand can be provided from the demand side (with buildings making up 40\% of the total consumption in Europe (Economidou et al., 2011)) through Demand Response (DR). DR is where consumers shift or curtail their electricity usage in response to financial or other incentives. Commercial buildings are significant end-use consumers and given their inherent thermal mass and use of advanced control infrastructure together with heating, ventilation and air conditioning systems, have the potential to offer significant load shifting opportunities (Aduda et al., 2017). Data-driven control frameworks show promise as a robust and scalable technique for the heterogeneous building stock (Kathirgamanathan et al., 2020). This research compares two different approaches to the problem; a model-based approach which uses historical data to train predictive models as part of an optimal control problem and a model-free approach using deep reinforcement learning where an agent learns an optimal control policy from experience of interacting with the environment.

Model-based - Data Predictive Control - Ensemble (DPC-En)
Data-driven models, together with representing the building as a cyber-physical system (Schmidt and Ahlund, 2018), show promising potential in harnessing energy flexibility in commercial buildings, where deriving a physics-based control oriented model is challenging. One such promising technique is ‘Separation of Variables’ where regression trees are built using training data ensuring that the control inputs (variables to be optimised) are excluded (Behl et al.,...
2016). This control input data is then used to fit affine convex models under each leaf of the regression trees leading to a convex optimisation problem. The initial study by Behl et al. (2016) only provided a one-step look ahead prediction and hence was not suitable for a receding horizon problem such as model predictive control (MPC). Jain et al. (2016) solved this problem and extended the initial work using multivariate regression trees with output corresponding to each step of the prediction horizon. They were able to show that their technique had only a small justifiable additional cost compared to a traditional linear MPC approach. The approach was further improved by Smarra et al. (2018) where each regression tree was replaced with a random forest (ensemble of regression trees) in order to reduce the variance in prediction and mitigate the tendency of regression trees to overfit. The authors termed this technique ‘Data-Driven Predictive Control with Ensemble Methods’ or DPC-En. The training data is composed of the disturbance variables \( d \) which includes both autoregressive terms \( \delta_d \), predicted disturbances \( d \) and autoregressive terms of the state variables \( x \). From this training data, random forest models are constructed over the prediction horizon \( j \) predicting the evolution of the state variables. Each forest is composed of \( n \) regression trees with an ensemble being taken of all the predictions. In every leaf node of each of the trees, a linear model is fitted from the affine combination of the control inputs to refine the state prediction. In this case, the ensemble of the regression model coefficients is used. Bunning et al. (2019) utilises a variation of this technique applied to a real-life residential apartment. Whilst the reader is referred to Smarra et al. (2018) for more details on the algorithm used to train the data-driven model and the mathematical definitions, the current paper employs the same approach as used by Bunning et al. (2019), where the modifications in training algorithm are outlined in Section Methods.

Model-free - Soft Actor Critic Deep Reinforcement Learning (SAC DRL)

The model free technique investigated in this research is deep reinforcement learning (DRL). Note that model-based reinforcement learning techniques also exist although they are not considered in this research. RL was chosen as the model-free control framework of interest given its infancy in the building energy management domain and theoretical suitability for DR applications (Wang and Hong, 2020). What makes RL particularly attractive is the potential of learning a control policy tabula rasa and ability to take user feedback into account whilst reacting to an external grid signal. RL has garnered significant interest in recent years in control applications in various domains but also in the building energy management domain as highlighted by the review of Vázquez-Canteli and Nagy (2019). Within RL applications, DRL has further garnered increased interest due to their increased effectiveness at learning control policies in complex non-linear environments and systems (Mason and Grijalva, 2019).

The nature of the DR engineering problem is such that it features continuous state and action spaces (e.g., cooling setpoint). Whilst discretising the action space is one option to allow these DRL approaches to be adapted for continuous action environments, this poses challenges such as the curse of dimensionality (Lillicrap et al., 2016). This is especially an issue in situations where fine control of actions is required. Another challenge with model-free DRL algorithms is that they are often notoriously expensive in terms of sample efficiency (Haarnoja et al., 2018), often requiring millions of steps of data collection before any meaningful control policy is learnt.

The current research utilises the soft actor-critic (SAC) algorithm, an off-policy maximum entropy actor-critic algorithm, as first proposed by Haarnoja et al. (2018). At their core, actor-critic methods are a type of policy gradient methods which have separate memory structures to explicitly represent the policy (Sutton, Richard S.; Barton, Andrew G., 2014). The policy structure is known as the actor and the estimated value function is known as the critic. The actor selects the actions whereas the critic evaluates the actions made by the actor. The reader is referred to Sutton, Richard S.; Barton, Andrew G. (2014) for a more detailed explanation of actor-critic methods. Haarnoja et al. (2018) suggest that the soft actor-critic algorithm provides for both sample-efficient learning and stability and hence extends readily to complex, high-dimensional tasks. They found the SAC algorithm showed substantial improvement in both performance and sample efficiency over both off-policy and on-policy prior methods. There have been limited applications of this algorithm to a high-fidelity environment such as EnergyPlus for applications in building energy management.

Methods

Boundary Conditions

The case study building is considered to be located in Rome, Italy. An EnergyPlus model was utilised and executed for a one year period using the typical meteorological year (TMY) weather file for Rome, Italy (ASHRAE climate zone 3). The real-time price used in this study was based on actual market data from Italy for 2017 (GME, 2018). Both controllers are tested on a work week in July (03/07-07/07).

Building Details

As a representative commercial building, a US-DOE (United States Department of Energy) commercial building archetype model was chosen as an initial model (Deru et al., 2011) (Figure 1). The DOE claim that the archetype models represent approximately
two-thirds of the commercial building stock in the USA (Deru et al., 2011). In this body of research, the ‘Large Office’ reference building model was chosen as it represented the largest buildings by floor area. The version with “new construction” was selected in particular with this building complying with the minimum building envelope and thermal properties of ASHRAE Standard 90.1-2004. Based on this standard, the building has a typical U-value of 0.857 W/m²·K. This building has a floor area of 46,320 m² over 12 floors. The building operates from 6.00 am to midnight on weekdays and 6.00 am to 5.00 pm on Saturdays (there is no occupancy on Sundays). This information is summarised in Figure 1 together with a 3D view and a zone plan of the ‘core mid’ zone that is considered in this research. In terms of HVAC systems, the building has a gas boiler for heating (1,766 kW), two water-cooled chillers (1,343 kW) for cooling and a multi-zone variable air volume systems for air distribution. Note that limitations of using synthetic data generated from an EnergyPlus model include “perfect” data which lacks stochasticity’s associated with real data as well as data-quality issues such as missing or erroneous data. These issues are considered beyond the scope of this paper.

This research focuses on the building ‘core mid’ zone treating this zone as representative of the entire building. This is justified as this zone represents the majority of the zonal temperatures, being the largest zone per floor and representing 10 of the 12 floors through symmetry properties of the simulation (the EnergyPlus model only simulates one of these 10 ‘middle’ floors and assumes the other nine floors are identical). In this study, the cooling setpoint (indoor zone air dry bulb temperature for cooling seasons) is used as the decision variable, as it is easily interpretable and commonly used in building energy management systems as a feedback to end users/occupants. During occupied hours, the temperature constraints are set at ±1°C from a reference temperature of 22°C and this is relaxed during unoccupied hours to ±5°C.

Co-simulation Environment

In the absence of a real building, the same EnergyPlus model described earlier, acting as the surrogate model, is used to test the data-predictive control in a closed-loop simulation through the use of co-simulation. This is necessary, as EnergyPlus, whilst a detailed physics based simulation engine, lacks the capabilities for implementation of more advanced control strategies and only allows manually coded rule-based control strategies. In this case, the data-driven model or RL agent and the EnergyPlus model are digital twins.

The PyEp python module is used for communication between EnergyPlus and Python (where the data-driven control is implemented) through the use of an open platform communications (OPC or OLE for Process Control) bridge. The PyEp based controllers use the OpenOPC library to connect to an OPC server. The free Matrikon OPC Server Simulator is used as the default server. The reader is referred to Jain et al. (2018) for further information on PyEp and the EnergyPlus-OPC bridge.

Data-Driven Predictive Control with Ensemble Methods (DPC-En)

A variation of the DPC-En approach employed by Smarra et al. (2018) and as utilised in the previous work of Kathirgamanathan et al. (2020) is used in this study. The synthetic data generated from the EnergyPlus model was split with the months from January to June being used for training. Note that the length and season of the training period was considered in further detail in the previous work of Kathirgamanathan et al. (2020) and is beyond the scope of the current paper. A random forest with 1000 trees is used with a minimum amount of 200 samples in each leaf (Bunning et al., 2019). A prediction horizon of 5 hours (20 time-steps) is used.

The predicted electricity consumption at the \(j^{th}\) step (\(\hat{Y}_{\text{electricity},j}\)) is given by fitting a regression model on the samples in the leaf with the affine sum of the control inputs (\(u(k),...,u(k+j-1)\)) being the dependent variable as follows:

\[
\hat{Y}_{\text{electricity},j} = \hat{x}(k+j) = \gamma_j[1, u(k),...,u(k+j-1)]^T
\]  
(1)

Here, \(\hat{x}(k+j)\) is the endogenous and exogenous variables representing the autoregressive terms and predicted disturbances. Similarly, the predicted temperature at the \(j^{th}\) step (\(\hat{Y}_{\text{temperature},j}\)) is as follows:

\[
\hat{Y}_{\text{temperature},j} = \hat{x}(k+j) = \alpha_j[1, u(k),...,u(k+j-1)]^T
\]  
(2)

The \(\gamma_j\) and \(\alpha_j\) terms are calculated from the average of the regression coefficients from all the trees of...
the forest. The reader is referred to Smarra et al. (2018) for an in-depth description of the derivation of the above model. Similarly to what Bunning et al. (2019) found, the dimensionality of these coefficients is reduced to 2 for each \( j \) due to poor prediction performance for larger horizons resulting from the high dimensionality for the model fitting process. The control variable (in this case the indoor zonal cooling set point temperature) is standardised around 0 as this was found to improve the building model accuracy significantly.

Once the models for describing the building dynamics are found for the pertinent prediction horizon, they can be integrated into a traditional MPC-like receding horizon problem. Given a grid signal \( e(t) \), e.g., real-time pricing, the corresponding linear optimisation problem is:

\[
\min_{u, \epsilon} \quad \sum_{j=1}^{N} e(t) p_{k+j}^{grid} + \lambda \epsilon_{k+j} \tag{3a}
\]

subject to
\[
x_{k+j} = \gamma_j [1, u(k), \ldots, u(k + j - 1)]^T, \tag{3b}
\]
\[
t_{\min} - \epsilon_{k+j} \leq x_{k+j} \leq t_{\max} + \epsilon_{k+j}, \tag{3c}
\]
\[
u \in U, \tag{3d}
\]
\[
\epsilon \geq 0, \tag{3e}
\]
\[
j = 1, \ldots, N. \tag{3f}
\]

where \( \lambda \) is the weighting term used to adjust the relative cost of comfort constraint violations, \( \epsilon \) is a slack variable for the comfort constraint, \( t_{\min} \) and \( t_{\max} \) are the time-varying zonal temperature constraints, \( p_{k+j}^{grid} \) defines the power drawn from the grid at time-step \( k \), and \( N \) is the prediction horizon. A sensitivity study was carried out for the value of \( \lambda \) (weighting term for relative cost of meeting thermal comfort constraints) and it was set to 100 for this study.

The Sklearn package (Pedregosa et al., 2011) was used in Python to create the random forest models. The objective functions are linear, thereby guaranteeing a convex program and a tractable solution. The linear program is solved using the COIN-CBC solver (Forest et al., 2018).

**Soft Actor Critic Deep Reinforcement Learning**

For reinforcement learning, the question of state space selection is considered to be a key part of the development and tuning of the RL controller. The state is what the RL agent observes for each control step. The RL problem is initially framed such that the DRL agent is required to select the value of the cooling set point for the VAV3 system feeding the ‘core mid’ zone (see Figure 1). A range of 21 °C to 28 °C is set as the allowed range with the agent outputting normalised values between the range of 0 and 1. The variables were selected aiming to ensure that they provide the agent with all the necessary information to predict the immediate future rewards (aiming to satisfy the Markov property of the RL problem) and given that they are feasible to be collected in a real-world implementation. All variables were scaled to be in the \((0,1)\) range according to min-max normalisation. The final selected set of states are listed below:

1. Electricity price (€/kWh)
2. Electricity price (€/kWh) (up to 4 hours ahead)
3. Total HVAC Demand (kW)
4. Total Power Demand (kW)
5. Core mid zone temperature (°C)
6. Day of week (1-7)
7. Hour of the day (1-24)
8. Outside drybulb temperature (°C)
9. Outside wetbulb temperature (°C)
10. Outside wind speed (m/s)
11. Outside wind direction (degrees)
12. Outdoor air relative humidity (%)
13. Direct solar radiation (W/m²)
14. Chilled water storage tank 1 temperature (°C)
15. Building occupancy schedule
16. Total HVAC Demand lag terms (up to 1 hour lag) (kW)

As with the previous section on the model-based technique, for a fair comparison to be made, initially the reward function is based on an automated price-based DR scheme with dynamic pricing. The reward function is split into two competing terms, a cost term reflecting the price of electricity purchased from the grid and a comfort related term (Equation 4). The cost term is proportional to the component of power demand that arises from the HVAC equipment (and hence arising from the cooling needs of the building) multiplied by the price of electricity. The comfort term is proportional to the deviation of zonal temperature outside of pre-specified and time-varying temperature bounds \( (t_{\min}, t_{\max}) \). Weighting parameters \( \beta \) and \( \lambda \) are introduced to weight the importance of the two terms of the reward function based on stakeholder requirements. \( \beta \) is fixed to a value of \( 1 \times 10^{-5} \) and the value of \( \lambda \) is set to 100 based on a sensitivity study.

\[
r = -\beta \times e_{HVAC} \times e_1 - \lambda \times |t_{zone} - t_{lim}| \tag{4}
\]

where:

\[
t_{lim} = \begin{cases} 
  t_{\min}, & \text{if } t_{zone} < t_{\min} \\
  t_{\max}, & \text{if } t_{zone} > t_{\max} \\
  t_{\lim}, & \text{otherwise}
\end{cases}
\]

There are numerous hyperparameters that are found in the implementation of the SAC DRL algorithm.
which influence the behaviour of the agent. A sensitivity study was performed to select the value of certain hyperparameters. All hyperparameters selected are summarised in Table 1. The reader is referred to Haarnoja et al. (2018) for a detailed explanation of the different hyperparameters.

Table 1: Hyperparameters for training for SAC DRL applied to ‘Large Office’ building

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Training Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ</td>
<td>Discount factor</td>
<td>0.99</td>
</tr>
<tr>
<td>α</td>
<td>Reward temperature parameter</td>
<td>0.05</td>
</tr>
<tr>
<td>λ</td>
<td>Comfort weighting term</td>
<td>100</td>
</tr>
<tr>
<td>τ</td>
<td>Target smoothing coefficient</td>
<td>3 × 10⁻³</td>
</tr>
<tr>
<td></td>
<td>Learning rate</td>
<td>1 × 10⁻³</td>
</tr>
<tr>
<td></td>
<td>Replay buffer size</td>
<td>2 × 10⁸</td>
</tr>
<tr>
<td></td>
<td>Minibatch size</td>
<td>2048</td>
</tr>
<tr>
<td></td>
<td>Policy</td>
<td>Gaussian</td>
</tr>
<tr>
<td></td>
<td>Update interval</td>
<td>96 (once per day)</td>
</tr>
<tr>
<td></td>
<td>Hidden layer size</td>
<td>64</td>
</tr>
</tbody>
</table>

The DRL agent is trained purely on summer data to focus on the cooling behaviour. One training episode (or one simulation) is taken to be three months of simulation from the start of April to the end of June. As training is conducted offline on a simulation environment, many episodes can be simulated repeatedly to allow the agent to explore many different trajectories. 50 episodes were used for training (selected after some experimentation and being conservative to ensure as much learning was captured as possible). One key difference that was found to be required to facilitate training of the SAC DRL agent was increasing the time between successive control actions from 15 minutes (used by DPC-En) to 1 hour. Without this, during early episodes of training when the agent is predominantly exploring, large and frequent changes in the control set point cause the agent difficulty in learning the dynamics and relationship between the set point and the power and zonal temperature evolution.

The SAC DRL agent was developed in Python and using the PyTorch library (Paszke et al., 2017) based off the implementation of Pranjal Tandon (https://github.com/pranz24/pytorch-soft-actor-critic) created for a different RL environment (MuJoCo).

Results and Discussion

The results, when the DPC-En and SAC DRL controllers are deployed for a 5-day working week in July for the test office building in a closed-loop simulation, are presented in this section. Comparisons are provided to the default rule-based controller (RBC) for which the cooling setpoint is based purely on time of the day. Figure 2(a) describes the real-time electricity price used as input by the controllers which shows the general trend of lower prices prevailing during night hours (midnight to 06.00 am) and higher prices during the evening peak (05.00 pm to 08.00 pm). Figure 2(b) illustrates the ambient environmental conditions (i.e., dry bulb temperature and direct (beam) solar radiation rate). Figure 2(c) shows the control variable, in this case, the optimal scheduled indoor zonal cooling set point, as output by the different controllers. It can be seen that the data-driven controllers are able to pre-cool the building during the early morning period taking advantage of lower real-time prices and hence requiring less cooling over the work day when higher prices are prevalent.

The controllers are also able to reduce the cooling demands during the hours prior to office closure (evening peak for the grid) and set-back by taking advantage of the thermal mass compared to the existing RBC. This is illustrated in Figure 3(b) showing the total power demand of both control strategies. Figure 3(c) illustrates how the controllers are able to respect the temperature bounds and (c) illustrates the cumulative objective function (which takes both economic cost and costed thermal comfort violations into account as defined in Eq 3a) showing the superiority of both data-driven techniques compared to the default RBC. Table 2 provides summaries of the controller performance specified in terms of energy consumption, energy spend (cost of electricity purchased from the grid in Euros) and discomfort hours (degree hours that zonal temperature is outside comfort bounds). Both Figure 2 and Figure 3 show that whilst the controllers are able to shift the cooling load, they do this in an inefficient manner with a considerable amount of over-actuation and hysteresis behaviour. This can cause actuation fatigue and is undesirable. One potential solution is to add an extra term to the objective function (in the case of the DPC-En controller) and the reward function (in the case of DRL) to penalise actuation in a proportional manner.

The DPC-En technique minimised energy costs (by 12.3% compared to RBC) through activation of building energy flexibility whilst also improving thermal comfort for occupants for a work week. The approach developed can be easily modified for minimum carbon emissions or energy consumption objectives. The DPC-En approach has a low computational burden with an average run time of only around five seconds per time step allowing the approach to be applied in real-time.

The SAC DRL agent learnt an optimal control policy to activate building energy flexibility given a dynamic price signal. For the same test summer work week, the DRL agent minimised the energy costs (by 9.7%) and maintained or even improved thermal comfort for occupants by respecting comfort constraints. This also highlights the future possibility of this technique
Table 2: Comparison of performance of DPC-En and SAC DRL compared to reference RBC control

<table>
<thead>
<tr>
<th>Control Type</th>
<th>Energy Purchased (MWh)</th>
<th>Energy Cost (€)</th>
<th>Discomfort Degree-Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBC (Ref)</td>
<td>115.34</td>
<td>5702</td>
<td>0.60</td>
</tr>
<tr>
<td>DPC-En</td>
<td>102.43</td>
<td>5002</td>
<td>0.16 73.3</td>
</tr>
<tr>
<td>SAC DRL</td>
<td>105.18</td>
<td>5150</td>
<td>0.13 78.3</td>
</tr>
</tbody>
</table>

Figure 2: DPC-En and SAC DRL controller results (compared to the baseline RBC) over work week in July for test climate of Rome. (c) plots the actuation of the control variable for the different controllers.

to take human feedback for occupant thermal comfort rather than temperature constraints. This was all possible without the need for training data-driven models for a given building and climate. The results show that the agent is able to outperform the RBC with a very minimal state space consisting of readily available variables. Note that the DRL controller achieved lower savings compared to the model-based technique as in the case of this particular building, the optimal control problem was better able to harness the upper and lower bounds of the control variable.

Conclusion

This research aimed to identify the most suitable data-driven control approach for automated demand response in commercial buildings. Two control approaches for unlocking building energy flexibility in the passive thermal mass were investigated. The model-based technique considered was the ‘separation of variables’ approach utilising random forest ensemble predictors integrated within a Model Predictive Control framework (called DPC-En). The model-free technique studies was deep reinforcement learning (DRL) with the ‘Soft Actor Critic’ (SAC) algorithm employed. Both techniques were evaluated in a surrogate simulation model of the same ‘large office’ building through the use of co-simulation. The novel SAC algorithm is able to handle continuous action/state spaces while most previous RL work in this field has been limited to discrete spaces. A further novelty
is the use of the SAC algorithm in conjunction with an EnergyPlus based environment which has seen limited applications to date.

Both model-based and model-free techniques outlined in this research appear as suitable candidates for harnessing energy flexibility from the passive thermal mass. These two techniques are able to react to a signal from the grid-side (or aggregator) whilst respecting occupant comfort constraints lending their suitability to automated demand response. Both techniques were able to achieve comparable (14.0% and 10.7% for DPC-En and SAC respectively) cost savings when subjected to a dynamic pricing scheme compared to a reference RBC. Both control techniques have a minimal computational burden for online deployment suggesting their potential for harnessing energy flexibility at sub-hourly resolutions. However, one caveat is that in the case of DRL, given the slower dynamics of a large building, the control timestep needed to be increased to a resolution of one hour for the agent to be able to learn the dynamics. In the case of simple 1-zone buildings or energy systems such as batteries, with known dynamics, the model-based approach is expected to be superior. There are techniques in which both model-based and model-free approaches can be combined aiming to leverage the benefits of the individual approaches although this is considered as a candidate for future research.

Limitations for the model-based technique include the use of synthetic data for training and perfect forecasts being used. There have been a very few examples using this technique on a real case study building. Given that the efficacy of the data-driven controller depends upon the quality and quantity of the training data available, this is an issue needing to be addressed. In terms of the DRL, further work is required to understand whether the small additional cost compared to the DPC-En technique can be minimised. Limitations include the significant training time considered in the development of the agent which would

Figure 3: DPC-En and SAC DRL controller results (compared to the baseline RBC) over work week in July for test climate of Rome
not be practical for a real building. A further unknown is how the DRL agent performs in buildings with different dynamics to the ‘large office’ building, especially smaller buildings with lower thermal mass and a greater proportion of the cooling load.

Acknowledgment

This work has emanated from research conducted with the financial support of Science Foundation Ireland under the SFI Strategic Partnership Programme Grant Number SFI/15/SPP/E3125.

References


