Large-scale data-driven predictive control of AHU and RTU systems for New York State in winter

Zhipeng Deng1,3, Xuezheng Wang1,3, Zixin Jiang1,3, Nianxin Zhou2,3, Haiwang Ge2,3, Bing Dong1,3

1Department of Mechanical & Aerospace Engineering, Syracuse University, Syracuse, NY 13244, US
2School of Information Studies, Syracuse University, Syracuse, NY 13244, US
3Built Environment Science and Technology (BEST) Lab, Syracuse University, Syracuse, NY 13244, US

Abstract
To meet the standards for reducing greenhouse gas (GHG) emissions, we need to create large-scale solutions for managing smart buildings. Using data-based models instead of physics-based models can make it easier to automatically create and implement these solutions on a large scale. However, it can be difficult to use data-driven control in cities, particularly for various heating, ventilation, and air conditioning systems and buildings.

This study aimed to assess the usefulness, strength, and feasibility of using data-driven predictive control (DDPC) for widespread use in the real world. To begin, we trained deep neural network models using data from 37 buildings in the RTEM database. We then used these models to improve energy efficiency in heating, ventilation, and air conditioning (HVAC) systems. We focused on two common types of HVAC systems: air handling units and rooftop units. The aim of the modeling is to use the collected data to establish the relationship between air temperature and HVAC load. We then use DDPC for large-scale optimal control. We simulated a month of winter weather to see how energy-efficient the DDPC method was. Finally, we looked at how much it reduced greenhouse gas emissions.

The DDPC method resulted in an average energy savings of 40% and a peak load reduction of 30% compared to current control systems. It also led to an average reduction of 2.6 kg of CO₂ emissions per square meter per month during the winter. We gathered valuable insights from implementing DDPC on a large scale. The smart control methods we proposed can easily be adopted by building owners for various types of buildings and HVAC systems. Cities and states can benefit from increased energy efficiency and reduced greenhouse gas emissions for a more sustainable future.

Highlights
- DDPC was evaluated for AHUs and RTUs in 37 buildings in one winter month
- We achieved 40% energy savings and 30% peak load reduction
- 2.6 kg CO₂ reduction per m²/month can be reached with DDPC
- DDPC can be easily adopted for promoting sustainable future

Introduction
New York State has set a target to lower GHG emissions to 60% of the 1990 levels by 2030 and 15% by 2050, in an effort to address climate change (New York State Department of Environmental Conservation, 2022a). According to a 2021 report from the Department of Environmental Conservation in New York (New York State Department of Environmental Conservation, 2022b), buildings are the largest contributor to greenhouse gas emissions, accounting for 32%. Also, space heating and cooling in the residential and commercial sectors account for 38% and 10% of building energy usage respectively (US EIA, 2021). To achieve these emission reduction goals, it is necessary to create a scalable smart control system that improves building energy efficiency and reduces greenhouse gas emissions.

The New York State Energy Research and Development Authority (NYSERDA) supported a program called the Real-Time Energy Management (RTEM) Incentive Program (https://www.nyserda.ny.gov/All-Programs/Real-Time-Energy-Management/Project-Dashboard), which was implemented throughout the state. By using this database, we can create and test smart building control systems.

Data-driven model predictive control
Model predictive control (MPC) is a control method that is commonly used to increase the energy efficiency of buildings, particularly in HVAC systems. MPC is based on advanced process control and can take into account the dynamics of building systems, making it a useful tool for managing sustainable building energy systems (Bazmi, 2011). However, traditional MPC methods that rely on building thermodynamic models require a significant amount of time and expert knowledge to create and calibrate. This makes it difficult to automatically develop and implement them in multiple buildings. As a result, physics-based controls can limit the large-scale deployment of decarbonization in buildings and the power grid.

In recent years, researchers have developed a method called data-driven model predictive control (DDPC) to...
address the limitations of traditional physics-based MPC (Kathirgamanathan, 2021; Rosolia, 2018). DDPC uses deep neural networks (DNNs) to model building systems, which can be trained using historical data, even with limited knowledge of building physics (Zhan, 2021). This allows a single model architecture to be applied to multiple cases to improve building energy efficiency (Schmidt, 2018). DNNs can learn complex and nonlinear building properties, which is difficult for physics-based models (Maddalena, 2020). DNN-based building control has been applied in both commercial and residential buildings (Kusiak et al. 2011). There have also been experiments and field implementations to evaluate the performance of DDPC in various buildings (Yang et al. 2021; Maddalena et al. 2022). Creating detailed physics-based building models for entire communities or cities can be difficult or time-consuming. Data-driven models are a better option for large-scale deployment. Some research has used data-driven methods to study electricity demand at the community (Li, 2018) and district level (Dagdougui, 2019) based on historical data. However, data-driven models are often opaque, and it can be difficult to understand the physical meaning of the model parameters, which can lead to errors in prediction results. Therefore, large-scale testing and verification is needed to ensure the scalability and robustness of data-driven models. There are currently no examples of the implementation of DDPC in a large number of building HVAC systems at the urban scale.

The purpose of this study was to see how well DDPC works in real-world, large-scale situations. To do this, we used data from the RTEM database to create models using DNNs to predict the temperature of the air in a building. The aim of the modeling is to use the collected data to establish the relationship between air temperature and HVAC load. We then used these models to improve energy efficiency in the building’s control system and evaluated the potential energy savings and reduction in greenhouse gas emissions of the proposed method. Finally, we analyzed the robustness and scalability of the models.

Methods

Fig. 1 shows the overall approach for this paper. Initially, we extracted and cleaned data from the RTEM database using only the periods where data was complete. We chose to use a 15-minute interval for our analysis because it was appropriate for the DDPC optimization problem and the frequency of control operations. Additionally, we included outdoor weather data from the nearest airport in each city to account for the influence of outdoor temperature on building energy usage and prediction. The climate region in New York State was cold according to International Energy Conservation Code. Therefore, in this study, we focused on heating season.

Data preprocessing and descriptive statistics of data

The RTEM database’s metadata included information about each building in the database, such as the building’s ID, area, customer type, location, number of equipment, number of data points, type and description of data points, logging time, and tags. The most common tags for HVAC systems were for air handling units (AHUs), fan coil units (FCUs), rooftop units (RTUs), unit ventilators (UVs), and variable air volume (VAV) systems. In this study, we focused on AHU and RTU and applied the DDPC to them, which were two kinds of system commonly used in many buildings. Data cleaning included selecting the building space with complete HVAC energy-related data, and then selecting the frequency of data recording for 15 minutes. We also used linear interpolation to supplement missing data points, zero drifts, and outliers. After cleaning the data, we identified a total of 148 HVAC units in 37
buildings with complete data on air temperature and energy usage, such as supply air temperature and airflow rate. We used this data from the 37 buildings to train and test the DDPC, and Table 1 shows the number of units and buildings we used for each type of HVAC system. We also collected information about the buildings used in our analysis of DDPC. The majority of the buildings for which we developed DDPC were commercial retail and commercial office buildings, making up about half of all the buildings.

<table>
<thead>
<tr>
<th>HVAC systems</th>
<th>Number of units</th>
<th>Number of buildings</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHU</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>RTU</td>
<td>98</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 1. Number of units and buildings for testing the DDPC in this study

Data-driven models of HVAC systems for predicting air temperature

After obtaining the data, we used it to develop data-driven models. Heat transfer and airflow occur between different zones within buildings. A state-space model is commonly used to describe building thermodynamics in physics-based MPC. The state of the building usually includes air temperature and wall temperature. Data-driven models are based on historical time-series data, so after cleaning the data, we built data-driven DNN models for zone air temperature prediction. DNN models are a type of machine learning method that use multiple layers in a neural network model to learn the relationship between input parameters and output (Deng, 2018). The input parameters were space air temperature, outdoor air temperature, room occupancy, and the heating or cooling load of the HVAC system. These parameters are typically recorded by the BMS. Unlike white-box and gray-box physics-based models, data-driven models can be built without this information. Building thermodynamics involves several important factors, such as air temperature, wall temperature, solar radiation, number of occupants, internal heat gain, and heat transfer between different zones. Physics-based models typically require data for these factors, but collecting them can be difficult using sensors in existing buildings. The RTEM database also does not have this information. While it would be difficult to include these parameters in data-driven control for a large number of buildings, it is possible to learn their relationship from historical data using a DNN model. This model can predict the zone air temperature for the next time step. The DNN model can be written as:

\[
T_{\text{act}}(t+1) = f[T_{\text{act}}(t), T_{\text{amb}}(t), \text{Occ}(t), P_{\text{HVAC}}(t)]
\]

(1)

Where \( f \) is the trained DNN model. We assumed that the load of the HVAC system for the space was proportional to the supply airflow rate and the temperature difference between supply and return air as

\[
P_{\text{HVAC}}(t) \propto Q(t) [T_{\text{supply}}(t) - T_{\text{return}}(t)]
\]

(2)

To account for the variations in building envelopes and heat transfer among different buildings, we created separate models for each HVAC system using the collected data. We only used data from the specific HVAC system to train the DNN model, so that it could learn the relationship from the data. Each HVAC system was assumed to serve a single thermal zone. For the buildings in New York State, we trained two models for each HVAC system, one for the heating season (October to March) and one for the cooling season (June to August). We used historical data from 3 consecutive days for training and data from 30 consecutive days (one month) to evaluate the model’s performance, energy efficiency, and reduction of greenhouse gas emissions.

Model training and control development

To prepare the data for training, we first applied min-max normalization to all the input data. Then, we used a technique called grid search to find the best values for the hyperparameters of the DNN models. We found that for optimal performance, the best number of neurons was 50, the best number of hidden layers was 4, the best learning rate was 0.001, the best training method was the ADAM optimization algorithm, and the best number of training episodes was 10000. We also used a type of activation function called rectified linear unit (ReLU) and a batch size of 64. We split the training data randomly and used 20% of it as a validation set during the training process. To evaluate the model’s accuracy, we used a metric called mean absolute percentage error (MAPE):

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right|
\]

(3)

Where \( A_i \) and \( F_i \) were the actual values and predicted values, respectively. The training time of the DNN model for each space was less than 5 minutes.

Once the DNN models were developed, we used them for a technique called smart data-driven predictive control (DDPC). The goal of DDPC was to use the least amount of energy while keeping the room air temperature comfortable over a set time period, called the prediction horizon. The factors we controlled were the heating or cooling load of the HVAC systems. We aimed to keep the air temperature within 0.5°C of the desired temperature or set point for each space. For all the buildings in this study, we set the prediction horizon as 3 hours and the control time step as 15 minutes in order to balance computing time of iterative optimization calculation and the control operation. The DDPC system is written as

\[
\min \sum_{t=0}^{n-1} P_{\text{HVAC}}(t)
\]

s.t.

\[
T_{\text{act}}(t+1) = f[T_{\text{act}}(t), T_{\text{amb}}(t), \text{Occ}(t), P_{\text{HVAC}}(t)]
\]

\[
T_{\text{actual}}(t) - 0.5 \leq T_{\text{act}}(t) \leq T_{\text{actual}}(t) + 0.5
\]
Where \(T_{\text{air}}(t)\) was the predicted space air temperature in each time step, and \(T_{\text{actual}}(t)\) was the collected air temperature.

We needed a powerful computer to handle the large number of HVAC units we were studying, so we used a high-performance computer with 80 cores and 176GB memory for model training and DDPC validation. In actual deployment, the calculations would be distributed to local computers connected to the Building Management System (BMS) in each building.

**Evaluate the performance of energy saving and GHG emissions**

To see how much energy our DDPC system was saving, we simulated its energy usage for each space in all the buildings for one month in both the heating and cooling seasons. We then compared these results to the current control strategies and energy use data for all the buildings, which we called the baseline control. We found that most of the buildings were using simple temperature set point or schedule controls.

To evaluate the reduction in GHG emissions, we used the GHG Emissions Calculator from the United States Environmental Protection Agency (US EPA, 2022). This calculator is a tool that helps estimate the energy and GHG emissions of various energy conservation measures for commercial buildings. The emissions factors vary depending on location, so we used the information for Upstate New York, New York City, and Long Island, since the data in this study were collected in New York State. We looked at the emissions of \(\text{CO}_2\), \(\text{CH}_4\), and \(\text{N}_2\text{O}\), which are the main contributors to GHG. Table 2 shows the emission factors for these gases in New York State.

**Table 2. Total emission factors in New York State (US EPA, 2022).**

<table>
<thead>
<tr>
<th>Location</th>
<th>(\text{CO}_2) (lb/MWh)</th>
<th>(\text{CH}_4) (lb/MWh)</th>
<th>(\text{N}_2\text{O}) (lb/MWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upstate</td>
<td>232.3</td>
<td>0.017</td>
<td>0.002</td>
</tr>
<tr>
<td>New York City</td>
<td>553.8</td>
<td>0.021</td>
<td>0.002</td>
</tr>
<tr>
<td>Long Island</td>
<td>1209.0</td>
<td>0.157</td>
<td>0.020</td>
</tr>
</tbody>
</table>

**Results**

**Training and testing of DNN models**

Fig 2 shows the results of training and testing the DNN models for different HVAC systems. The mean absolute percentage error (MAPE) of the predictions made by the DNN model for 98 RTUs was 1.1% for training and 2.8% for testing. The training results were very good, with a training error of less than 5% for almost all the RTUs. The testing results were slightly worse than the training results. Similar results were found for the other systems. The MAPE for training and testing of 50 AHUs was 1.0% and 2.3%, respectively. The DNN model achieved great results for predicting indoor temperature, which is one of the key performance metrics, for different HVAC systems and buildings. This indicates that the DNN model has good scalability. We can use the trained DNN model to predict air temperature and apply the trained model to DDPC to reduce energy use in each zone.

**Results of load reduction**

Once we built and trained the DNN models, we used them for data-driven predictive control. Fig 3 shows the results of how well the temperature and energy were controlled for an AHU in one building over 30 heating days. DDPC was able to control the predicted temperature to match the actual collected data within 0.5°C most of the time, ensuring that the thermal comfort in this zone was almost the same as the actual conditions. In the winter, the indoor air temperature during the day could be controlled around 22°C (72°F). At night, when the space was unoccupied,
Fig 3. Energy and air temperature results of the DDPC for an AHU in one building: (a) heating load in one month; (b) air temperature in one month.

Fig 4. Energy and air temperature results of the DDPC for a RTU in one building: (a) heating load in one month; (b) air temperature in one month.
The HVAC system did not provide heating to save energy. During the weekend, there was no load, and the air temperature was allowed to fluctuate, possibly dropping to 18-19°C. The low heating load may be because this building space was in interior space, so less load was required. Fig 4 also shows that the heating load could be reduced by DDPC compared to the current baseline control for RTU. The energy saving for heating load was 41% over the 30 days. The peak load reduction was 6% and 28% for AHU and RTU in these two buildings in the winter.

We then evaluated the energy savings of DDPC for all HVAC systems in the 37 buildings. Fig 5 shows the reduction in heating and cooling load by DDPC for all AHUs and RTUs in buildings in one heating month. We found that it could save an average of 39% on the heating load of AHUs. For RTUs, an average of 43% on the heating load could be saved. The overall energy saving was 41% on heating and cooling load. DDPC has achieved similar energy-saving goals for different HVAC systems and buildings, indicating good scalability. Figure 6 shows the distribution of peak load reduction by DDPC for all buildings. The average peak load reduction in winter was 36% for AHUs and 42% for RTUs in all 37 buildings. Therefore, data-driven predictive control has significant potential for energy savings and peak load reduction in New York State.

3.3 Reduction of GHG emission for DDPC

Finally, we analyzed the reduction in GHG emissions resulting from DDPC in the 37 buildings. Fig 7 shows the distribution of CO₂ emission reduction among all the 37 buildings. We found that DDPC could reduce CO₂ emissions by an average of 2.58 kg per square meter per year, with a range of 1.3e-2 to 13.39 kg per square meter per year. The distribution of other GHGs was similar, as the GHG emissions were calculated based on energy reduction and emission factors. The results for the reduction of CH₄ and N₂O emissions were 9.8e-5 (ranging from 1.2e-8 to 4.6e-4) and 9.3e-6 (ranging from 1.2e-9 to 4.4e-5) kg per square meter per year, respectively. The results varied a lot for different buildings, as shown in Fig.
7. This may be due to the fact that the HVAC systems analyzed in different buildings may not represent all the systems inside the building, as data for some systems were missing or not accessible. Additionally, the building area listed in the database may differ from the conditioned area. Taking these possible reasons into account, the resulting GHG emission reduction may be higher in some buildings.

Discussions
For the scalability of DDPC, in this study, we developed data-driven models for energy-efficient control using data collected from 37 buildings in the RTEM database. Overall, the DDPC showed great robustness and provided reasonable operation to the HVAC systems, which is important for scalability. However, there were still some cases where the DDPC did not work properly. We categorized these cases into three stages: failure in model training, failure in model validation, and failure in control. Possible reasons for failure in model training include data outliers, constant load recording, discrete variables, and data abnormal variation and disturbance. Possible reasons for failure in model validation include underfitting and overfitting. Possible reasons for failure in control include impossible to find optimal solution, conditions beyond training set, or too large cumulative error. To address these issues, we recommend calibrating the model with actual data corrected every day or every few days.

As for the obstacles during the development of DDPC, the RTEM database had a well-organized data structure that made it easy for researchers to develop and validate different models. However, we still faced some obstacles during the development of data-driven models and the control algorithm. One obstacle was ensuring that data was uniformly named, labeled, and had consistent units. Another obstacle was synchronizing time and control steps. To avoid these issues in future databases designed for data-driven modeling purposes, it is important to pay attention to these obstacles.

Conclusion
In this study, we explored the scalability and energy efficiency of deploying data-driven predictive control at a large scale for AHUs and RTUs in 37 buildings in one month. This investigation led to the following conclusions:

1. We trained DNN models using data from 148 HVAC systems in 37 buildings in New York State and then deployed DDPC on a large scale to evaluate its performance. The results showed that it could save more than 40% of heating and cooling load on average, and also reduce peak load by 30% in winter in one month.

2. We also found that DDPC could reduce the emission of CO₂ by 2.58 kg per square meter per year in buildings. If all buildings in New York State used DDPC, it could reduce GHG emissions by over 3%.

This study has limitations and there is a need for future work. It is currently not practical to conduct field tests to validate the DDPC for different HVAC systems in a large number of buildings across different cities. In this study, we only focused on the air system and building heating/cooling load. To further develop and validate the DDPC, it is important to expand to more complex building energy systems, such as water systems (boiler, chiller, pump) and renewable energy systems (PV panel and wind turbine). Exploring data-driven predictive control for sustainable building and city could be a future direction of research.

Acknowledgment
The research work presented in this paper is supported by the U.S. National Science Foundation Award No. 1949372. We obtained the API for the RTEM database through the RTEM Hackathon by NYSERDA.

References


