Exploring the integration of simulation and deep learning models for urban building energy modelling and retrofit analysis

Alex Nutkiewicz¹, Rishee K. Jain¹

¹Urban Informatics Lab, Stanford University, Stanford, USA

Abstract

As the world rapidly urbanizes, cities will need to be created and expanded to accommodate their growing populations – putting immense pressure on engineers, scientists and policymakers to improve the efficiency of the energy intensive built environment. One of the key barriers to improving the energy efficiency of cities is the ability to accurately model and characterize the energy performance of their buildings. While simulation-based methods have been developed to help predict the energy consumption of urban buildings, they are limited in their ability to quickly evaluate the effects of various design or retrofit scenarios. New data-driven methods are emerging to model building energy usage, but their lack of an underlying physics-based engine limits applicability and interpretability for assessing design or retrofit scenarios. In this paper, we employ the use of an integrated simulation and data-driven method (i.e., Data-driven Urban Energy Simulation or DUE-S) to model a large-scale retrofit policy on a case study of 52 buildings in a Californian city. Our results indicate that the DUE-S model is able to capture the energy impacts that the urban context has on buildings that undergo retrofits as well as those that do not. Our primary contribution is to demonstrate the merits of combining physics-based building simulation methods with new data-driven machine learning methods (i.e., transfer learning) to assess the impact of various design and retrofit scenarios across a large urban area and in turn spawn future research at the intersection of simulation and data science. In the end, realizing deep energy savings from urban buildings will require new tools that are both accurate and interpretable enough to inform decision-making for a variety of urban sustainability stakeholders regarding early stage designs, energy efficiency retrofits and environmental policymaking.

Introduction

Urban populations are rapidly growing, placing increased pressures on cities to reduce their long-term greenhouse gas emissions. This strain is seen particularly in the buildings sector as the urban built environment currently accounts for 75% of total primary energy use and 50-75% of carbon emissions – making it the largest energy-consuming and pollution-emitting sector (EIA, 2016). As a result, hundreds of global cities have set ambitious targets for reducing their greenhouse gas emissions (e.g., San Francisco plans to reduce GHG emissions 40% below 1990 levels by 2030 and 80% by 2050) (SF Environment, 2017). The reality is that a significant amount of the building stock that will exist in 2050 has already been constructed and thus necessitates the need for large-scale retrofits to meet these ambitious emission reduction targets. However, there lacks a widely accepted and trusted method to conduct large-scale retrofit analyses within cities. Widespread academic efforts have been made to understand how both data-driven and simulation-based methods can be used to assess the effectiveness of various retrofits across multiple buildings in an urban area. While data-driven methods are effective in understanding and suggesting operational changes, they lack the interpretability to propose the implementation of various energy conservation measures (ECMs). And while simulation-based methods are capable of quantifying the change in building energy consumption resulting from an ECM, newly emerging urban building energy models (UBEM) require recalibration with additional ground truth data when changes are made to any existing building or when a new building is created in the study area (Reinhart and Cerezo, 2016).

Recent work has shown that integrating data-driven and simulation-based methods for urban energy modelling can yield accurate predictions of building energy consumption at multiple spatiotemporal scales (Nutkiewicz et al., 2018). This study extends a previously developed urban energy modelling framework – DUE-S: Data-driven Urban Energy Simulation – by utilizing a technique inspired by the deep learning concept of “transfer learning” to its workflow. Transfer learning is a method in deep learning that uses a pre-trained neural network, where, by holding its learned parameters constant, it can be repurposed with new data to predict a similar task. In the domain of energy efficiency retrofits, we postulate that by repurposing our pre-trained network, DUE-S can be used to assess the effects of various energy conservation measures (ECMs) without the need for re-training the model. We assess the implications of this extension by conducting a case study looking at a large-scale retrofit program of 52 densely co-located buildings in a Californian city. We expect that by introducing this addition to the DUE-S workflow we will be able to accurately characterize the energy consumption impact of various building retrofits across multiple spatial (individual building, block, urban) and temporal (daily, monthly) scales in a city.
Background

In order to support large-scale energy and greenhouse gas reductions, policymakers rely on various tools to evaluate and select energy conservation measures (ECMs) for widespread adoption. These tools include both data-driven and simulation-based methods, and each of them come with tradeoffs related to interpretability and computational efficiency. In this section, we provide a brief review of the methods used to conduct large-scale retrofit analyses.

Simulation-based methods

Large commercial and institutional buildings have frequently relied on simulation-based assessments of various retrofit strategies to better understand their resulting financial and energy-related consequences. The results of these assessments are often based on building energy modelling (BEM) programs that develop and solve heat balance equations representing the physical energy processes occurring within a single building. Once the whole-building BEM is developed and calibrated, the model can be modified to explore a range of retrofit scenarios.

In more recent years, single-building BEMs have been applied to the neighbourhood and urban scales, where the bulk of research done in this domain has focused on simplifying the thousands of input parameters required for simulation modelling while preserving the accuracy of a fully-calibrated simulation model (Reinhart and Cerez, 2016). These urban building energy models (UBEM) typically divide the urban building stock into a set of archetypes based on factors such as construction age, building type and geometric shape. Using the pre-defined archetypes, each building is then assigned a set of standardized non-geometric properties (e.g., building and construction materials, operation schedules, HVAC systems) in order to reduce the number of input variables required for the full simulation (Cerez et al., 2016). Once calibrated, UBEMs have shown to be successful in projecting the aggregate annual energy consumption of a neighbourhood of residential buildings (Mastrucci et al., 2014), estimating the energy impact of building and campus-level ECMs (Nagpal and Reinhart, 2018) and analysing the energy and cost savings related to several different ECMs and ECM packages across individual buildings in a city (Chen et al., 2017). Subsequently, these models can provide policymakers or utilities with probabilistic predictions of building energy performance under various retrofit measures (Heo et al., 2012). These types of approaches have also been extended to web-based applications (Hong et al., 2016; Robinson et al., 2009) that draw from GIS and city-based open data initiatives to help support policymaking and university master planning (Coccolo et al., 2015).

However, physics-based simulation models also face several key limitations. While it is widely understood that a building’s energy usage is affected by its urban context, such as the shading and heat transfer impacts of neighbouring buildings (Pisello et al., 2012; Han and Taylor, 2016), urban microclimates (Samuelson et al., 2016) and other urban systems (Singh et al., 2013; Lv et al., 2015), many of these effects are not modelled in simulation-based models. Furthermore, many of these models are limited to accurately predicting energy consumption at a single spatial scale (individual building, block, urban scale). Without the flexibility to accurately characterize energy performance at multiple scales, these UBEMs have limited ability to estimate the impacts a retrofit made in one building could have on the energy use of surrounding ones. Finally, because of the hundreds of inputs and assumptions required to develop an accurate UBEM, creating and calibrating a reliable model can take a considerable amount of time and computational resources. While there are emerging methods used to help speed up the calibration process, retraining and recalibrating an urban-scale model for each retrofit scenario can be time and resource prohibitive (Coakley et al., 2014). For example, in a study comparing the spreadsheet and simulation-based approaches to retrofit analysis, the UBEM approach was shown to take three times as long – largely because of the high number of input parameters and calibration time required of the energy modeler (Nagpal and Reinhart, 2018). Therefore, while simulation-based approaches are effective in providing policymakers and planners an interpretable understanding of energy impacts related to various ECMs, the time and resources required may not be feasible in practice.

Data-driven methods

With the rapid surge in sensor-based data and open data initiatives in cities, combined with the ongoing development of new machine learning algorithms to utilize this data, data-driven methods are becoming increasingly popular to analyse building energy consumption. For the application of building retrofit analysis, these methods primarily consist of benchmarking and load shape analysis methods that rely on a combination of metered data and publicly accessible databases to provide quick estimates of building energy performance (Chung, 2011).

Across the United States, 24 cities have enacted city-wide benchmarking ordinances that aim to quantify and evaluate building energy use patterns, identify the poorest performing buildings and target the worst of them for energy efficiency improvements through various retrofit programs and competitions (Yang et al., 2018). Various building portfolios and large universities have also employed benchmarking techniques to target specific buildings for energy efficiency retrofiting (Chen et al., 2017), especially because these entities are often limited in capital available for this type of work (Gliedt and Hoicka, 2015). While benchmarking methods are able to compare the relative energy efficiency of buildings across an urban area, additional building audits are still necessary to target the specific ECM or set of them that would best optimize each building’s energy efficiency. To better understand an individual building’s energy demand over time, engineers may also conduct a load profile analysis, which can identify operational inefficiencies at...
different times during the day, month or year (Luo et al., 2017).

However, these methods often rely too heavily on the mathematical patterns present in a building’s energy use data without considering the underlying thermodynamics of its systems. As a result, many of these data-driven models are unable to model the effects of physical changes made to a building (e.g., HVAC systems, building envelope, lighting) without having data on these ECMs included in their training dataset.

**Transfer learning**

Transfer learning is a method in deep learning that takes an existing, pre-trained model and holds its learned parameters constant to allow for new data to be passed through it with the goal of performing a different, but similar, function. Transfer learning is frequently done in deep learning when low level features learned in the initial network could be useful for learning a related task (Goodfellow et al., 2016).

One of the primary drawbacks of data-driven methods for retrofit analysis is that training data is required to understand how different ECMs will impact future building energy use. We aim to repurpose our neural network in a similar manner to transfer learning in order to overcome this limitation. By using a pre-trained neural network that already understands the uncertainty between simulated and metered energy consumption, DUE-S can take in a new simulated dataset based on a proposed energy efficiency retrofit and predict what its new metered energy usage would be under those conditions. We postulate that by comparing the initial and retrofitted conditions of an urban study region, we will be able to quantify the impact a large-scale retrofit scenario would have on not only an individual building’s energy usage, but the ones around it as well.

**Methods**

In this section, we describe the procedure to create the DUE-S model (see Figure 1 located at the end of the paper due to size): a two-part process that integrates baseline energy simulation models (Step 1) with a graph-based deep learning model (Step 2) to capture the spatiotemporal dynamics of urban building energy consumption. We expand upon this framework by leveraging an emerging technique inspired by deep learning called transfer learning (Step 3) that takes the pre-trained model used to predict the energy consumption of buildings in their current state, and, holding its pre-trained weights constant, repurposes it to assess the change in energy consumption under a proposed retrofit scenario. In this step, selected baseline building energy simulations from Step 1 are modified based on a proposed retrofit scenario. The new time series simulation outputs are then fed into the pre-trained deep learning model in order to produce a new set of predicted results, where the difference in pre- and post-retrofit results represents the energy impact of a proposed retrofit scenario. As discussed in the Introduction, the main contribution of our work is to demonstrate how repurposing our pre-trained DUE-S neural network can be used to quickly evaluate various retrofit scenarios in a given urban area.

**Step 1: Build baseline energy simulation models**

The first step of the DUE-S modelling framework is to create baseline building energy models that capture the first-order energy use dynamics of each building in the study area. To generate these models, we first create building geometries by drawing information from publicly accessible datasets: building footprints, heights and elevation. Simplified “2.5D” massing models can be constructed for each building by extruding building footprints by their heights. They can then be further divided into floors based on additional information or by calculating a standardized floor-to-floor height. Weather data is gathered from historical databases such as ones curated by NOAA (NOAA, 2019). Finally, non-geometric building characteristics, which include the building’s constructions, systems and schedules, are based on the U.S. Department of Energy’s Commercial Reference Building models (Deru et al., 2011). These models define the various plug and process loads, construction assemblies and operating systems and schedules based on national data from CBECS. Parameters from these models are allocated based on building type and construction age.

Once all inputs for building geometry, weather and non-geometric parameters are defined for each individual building, they are then simulated using EnergyPlus, where the resulting output is whole building energy consumption at 15-minute intervals. We reiterate that the goal of this step is to capture the basic energy use dynamics of each building. The outputs of these models serve as the inputs for the deep learning model, which aims to capture the uncertainties associated with energy modelling.

**Step 2: Develop graph-based deep learning model**

The goal of the graph-based deep learning model is to capture the sources of uncertainty that arise in building energy simulation: the assumptions made during modelling process and the hidden urban context impacts that cannot be captured by individual building simulations. The core of our spatiotemporal deep learning model is the integration of a graph and a recurrent neural network (RNN).

To capture the spatial interdependencies of building energy usage, we model their interactions using a graph where nodes are the buildings, and the edges represent the spatial relationships between them. We do not assume any prior knowledge of the relationship between these buildings aside from their relative proximity between one another.

Let graph $G$ be defined as $G = (E, V)$ where $E$ represents the edges and $V = \{v^1, v^2, ..., v^n\}$ denotes the nodes. Each node of the graph represents a building, denoted by $v^i$, and contains data $x$ (in our application being simulated energy consumption) at each time step $t$, represented as a scalar $x^k_t$, where each time step $t$ is a 15-minute interval of energy consumption. Furthermore, we represent the simulated energy consumption of a single building $k$ for the full time window $T$ as $x^k = \{x^k_1, x^k_2, ..., x^k_T\}$ and use...
Furthermore, we assume the vector \( \mathbf{a}_t = \{a_1^t, a_2^t, \ldots, a_n^t\} \) represents the actual energy consumption of all nodes at time \( t \), and the vector \( \mathbf{y}_t = \{y_1^t, y_2^t, \ldots, y_n^t\} \) denotes the prediction for energy consumption of all buildings at time \( t \). Therefore, using the current and previous values of the input (simulation) data as well as the current and previous values of the target (metered) data, our model aims to learn a nonlinear mapping (1) to predict the target series \( \mathbf{y}_T \):

\[
\hat{\mathbf{y}}_T = F(a_1, a_2, \ldots, a_{T-1}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)
\]

(1)

where \( F(\cdot) \) is the nonlinear mapping function the model is trying to learn.

To map the simulated to metered energy consumption in the buildings represented by nodes in our graph, we employ the use of a recurrent neural network (RNN) – a type of deep neural network that is commonly used in sequence-based applications. However, because RNNs suffer from the issue of vanishing gradients and subsequently have difficult capturing long-term dependencies within sequence data, we rely on long short-term memory (LSTM) units to overcome this limitation. LSTM networks are a specific type of sequence-based neural network characterized by a progression of memory cells. An exhaustive description of LSTM networks can be found in (Goodfellow et al., 2016). The model’s memory cells allow the network to “remember” long-term dependencies about the patterns in sequence data for future predictions, which makes it particularly effective in time series prediction and machine translation applications (Ghaderi et al., 2017; Qin et al., 2017). They are useful when predictions made during a certain time of day or year may be particularly important in predicting future energy consumption.

The goal of our deep learning model is to find a nonlinear function to map simulated data into metered data, which, in the deep learning domain, is similar to the idea of machine translation. Thus, the model’s architecture is based on combining two LSTM networks to create an encoder-decoder network (Sutskever et al., 2014) – a common deep learning architecture used in machine translation to translate one time series sequence into another (Figure 2). The key idea in our encoder-decoder network is to use the first LSTM to encode the input sequence (e.g., simulation data) as a set of fixed-length vectors and then use another LSTM to decode a translation of our simulation data into our desired output sequence (metered predictions).

**Figure 2: General encoder-decoder architecture.** The model is made up of a progression of LSTM cells which allow the network to remember or forget long-range dependencies to help make predictions deep into the sequence data. In this figure, \( \mathbf{x}_1 \) represents the input data originating from the energy simulation models for every building in the study area and \( \hat{\mathbf{y}}_T \) represents the predicted energy consumption for each building at time \( t \). The model is optimized to reduce the loss between metered consumption \( \mathbf{a}_t \) and the network’s prediction.

The outputs of our recurrent neural network are the final predictions of energy consumption for each building in the network, where, depending on how the data is structured, the model can predict at several different spatial and temporal scales. This model is now trained to map the relationship between simulated and metered energy consumption and can be repurposed to infer the effects of various retrofits made on buildings in the urban study area.

**Step 3: Conduct multi-scale retrofit analysis**

We expand on the DUE-S framework by using a technique inspired by transfer learning to infer the change in building energy consumption under a proposed retrofit scenario. Transfer learning takes the pre-trained RNN developed in Step 2 of the DUE-S workflow, and, holding all of its learned parameters constant (these parameters representing the uncertainty in the initial baseline energy simulations), runs a new input dataset through it in order to produce a new prediction of urban building energy consumption.

To run the retrofit analysis, baseline energy simulations are modified and re-simulated based on proposed retrofits to one or more buildings in the study area. The new outputs of these simulations become the time series inputs to the same deep learning model developed in Step 2 of the DUE-S workflow. Models that undergo this sort of transfer learning do not require retraining, so, holding all of its learned parameters constant, the RNN will quickly output a new prediction representing the new energy consumption with the proposed retrofits included. As a result, one of the primary benefits of this approach to retrofit analysis is that this prediction model does not need to be retrained or recalibrated, whereas simulation-based methods may require a significant amount of additional time for each scenario evaluated. To find the change in energy consumption resulting from the large-scale retrofit, the difference in output results can be taken between the baseline and proposed retrofit scenario.

**Case Study**

To demonstrate the feasibility of our extension of the DUE-S framework, we drew from a study area encompassing 52 densely co-located buildings in a Californian city (Figure 3). The region is part of a university campus, where its buildings are primarily used as offices, classrooms and libraries. The study area is fully paved with some greenspaces as shown in Figure 3.
The resulting size of the input array to the RNN was \([52, 105216]\) – representing the 52 buildings and 105,216 observations in the data. This input data was then shaped into 3D tensors and separated into training, development and testing sets, where we selected a time period from August 1, 2017 to December 31, 2017 as our test set for evaluations. Hyperparameters, including the size of each hidden state for the encoder and decoder and the learning rate, were optimized using the Adam algorithm (Kingma and Ba, 2014) to minimize mean bias error. The accuracy of our prediction model using mean bias error (MBE) and coefficient of variation (CV) of RMSE, per standards set by ASHRAE Guideline 14 (ASHRAE, 2002). The results in Table 2 show the prediction accuracy for various spatial and temporal scales. Our results indicate that our model is able to predict building energy consumption to the level of accuracy required per ASHRAE Guideline 14 standards for individual buildings. Similar to our findings in (Nutkiewicz et al., 2018), we discover that our combined simulation-deep learning model has improved prediction accuracy when estimates are aggregated at larger spatial or temporal intervals. This is likely the result of reduced variability in overall electricity consumption which subsequently dampens the buildings or blocks with worse prediction accuracy.

Table 2: Results of initial baseline DUE-S predictions – evaluated using MBE and CV(RMSE).

<table>
<thead>
<tr>
<th>Building Scale</th>
<th>Daily Interval</th>
<th>Monthly Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBE (%)</td>
<td>CV(RMSE) (%)</td>
<td>MBE (%)</td>
</tr>
<tr>
<td>Building</td>
<td>24.2</td>
<td>16.5</td>
</tr>
<tr>
<td>Block</td>
<td>16.7</td>
<td>12.8</td>
</tr>
<tr>
<td>Urban</td>
<td>6.28</td>
<td>5.12</td>
</tr>
</tbody>
</table>

Using our now-trained RNN, we ran an updated simulation dataset containing the five buildings with the two proposed retrofits detailed in the Case Study section through the model to produce new monthly predictions of individual and block-scale energy consumption. Table 3 shows the average percentage reduction in monthly electricity consumption for each individual building in the study area; for all individual buildings; for the full urban area. It is important to note that the “Baseline + Retrofit” model only looks at the change in electricity consumption between the different runs of the energy simulation: the baseline simulation and the simulations with changes made to the selected retrofitted buildings. Therefore, the individual and block scale results are lower than the “DUE-S + Retrofit” model because many of the buildings are unaltered, meaning their change in consumption is zero. However, because the “DUE-S + Retrofit” model relies on the electricity use of surrounding building to make future predictions, we see more changes in electricity use for a greater number of buildings and blocks – indicating that retrofitted buildings have an impact on surrounding buildings and blocks of buildings. This is effect is further demonstrated at the urban scale. The reduction in energy usage for the “DUE-S + Retrofit” model is once again higher than that of the “Baseline + Retrofit” model and indicates that buildings beyond those physically retrofitted will have reductions to their energy usage. This result corroborates previous work highlighting the impact urban context can have on...
building energy dynamics (Pisello et al., 2012; Samuelson et al., 2016) and underscores the need for incorporating multi-building impacts into the assessment of retrofit programs.

Overall, the Constructions retrofit, where each’s building’s building envelope constructions are modernized to the inputs defined by the ASHRAE 90.1-2010 Commercial Reference Buildings, has greater electricity consumption savings in both the Baseline and DUE-S retrofit cases. This is further detailed in Figure 4, which visualizes the monthly difference in predicted electricity savings for the full urban study area. Generally, the most savings are seen in the later summer months, where the temperature, combined with strong, hot winds, in this particular Californian city is highest.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Baseline + Retrofit</th>
<th>DUE-S + Retrofit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>WWR (%): 1.24, Con-</td>
<td>WWR (%): 3.28,</td>
</tr>
<tr>
<td>Block</td>
<td>structs. (%): 1.32</td>
<td>Con- structs. (%)</td>
</tr>
<tr>
<td>Urban</td>
<td>WWR (%): 6.55, Con-</td>
<td>WWR (%): 7.63,</td>
</tr>
<tr>
<td>Scale</td>
<td>structs. (%): 6.99</td>
<td>Con- structs. (%)</td>
</tr>
</tbody>
</table>

Table 3: Results showing percentage reduction in monthly electricity consumption.

Figure 4: Monthly plot of projected urban electricity savings between the Baseline and DUE-S cases.

Each of the models used to predict the impact of these proposed retrofits have different underlying assumptions. The “Baseline + Retrofit” scenario relies strictly on the underlying thermodynamic processes and assumptions associated with EnergyPlus – the modelling program we used for this analysis. However, the architecture proposed in our deep learning model allows electricity consumption for each building to be predicted based on not only the new simulation data but also on its surrounding neighbours. As a result, the predicted retrofit impacts in the “DUE-S + Retrofit” case are also based on the same interdependent relationships between a building and its surrounding urban context learned during the initial training of the DUE-S deep learning model.

Limitations and Future Work

This study was limited to a large neighbourhood of buildings in a Californian city. While this work was able to predict the energy implications of retrofits on this study region, our future work aims to validate our transfer-learning-based method by using a real case study of an urban-scale building retrofit. By leveraging data from a retrofit program currently being deployed in a community, we can look to better understand the model’s performance before and after a retrofit has taken place and see how various urban forms may respond differently to a specific type of large-scale retrofit. We are also interested in conducting a sensitivity analysis of various types of retrofits to see if DUE-S can quantify how various urban context factors (e.g., heat transfer, mutual shading) affect building energy performance. Finally, one of the primary limitations in repeating this work is the difficulty in collecting highly granular energy consumption data – largely due to privacy concerns. As a result, future work is needed to assess the required spatial and temporal granularity required to create accurate data-driven urban energy models.

Conclusion

Overall, this paper aimed to assess the feasibility in using an integrated Data-driven Urban Energy Simulation (DUE-S) model to quickly evaluate various large-scale retrofits in an urban environment. Our findings indicate that our integrated simulation and deep learning approach was able to predict future building energy consumption under two different retrofit scenarios without relying on a time-intensive recalibration process. Our results also show that energy savings projections under these different retrofit scenarios using our “DUE-S + Retrofit” model are to some extent able to consider the urban context. The aim of this paper was to demonstrate the merits of integrating data-driven and physics-based approaches to model urban energy consumption under various retrofit scenarios. Purely data-driven approaches lack underlying thermodynamic modelling to ascertain how energy retrofits might affect future energy performance. On the other hand, simulation-based approaches are able to model effects of retrofits but at large (e.g., urban) scales require time and resource intensive recalibration and thus limit their applicability in real-world scenarios. By demonstrating the merits of an integrated approach, this work aims to catalyse research at the intersection of building simulation and data science to overcome the limitations posed by each approach individually. In the end, tools capable of accurately predicting and characterizing urban building energy usage under various retrofit scenarios will be of utmost importance as we begin the challenging task of transitioning our cities to a more sustainable energy future.

Acknowledgements

The material presented in this manuscript is based in part upon work supported by the Precourt Institute for Energy, Center for Integrated Facility Engineering (CIFE), a Terman Faculty Fellowship and the US National Science
Foundation (NSF) under Grant No. 1461549. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF.

References


---

**Figure 1:** Expansion of DUE-S modelling framework with transfer learning.