<table>
<thead>
<tr>
<th>Proposal number:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposal title:</td>
<td>DUES: Data-driven Urban Energy Simulation</td>
</tr>
<tr>
<td>Principal investigator(s) and department(s):</td>
<td>Rishee Jain, Civil &amp; Environmental Engineering</td>
</tr>
<tr>
<td>Research staff:</td>
<td>Alex Nutkiewicz, PhD student, CEE Zheng Yang, Post-doctoral Fellow, CEE</td>
</tr>
<tr>
<td>Total funds requested:</td>
<td>$78,370</td>
</tr>
<tr>
<td>Project URL for continuation proposals</td>
<td>http://</td>
</tr>
<tr>
<td>Project objectives addressed by proposal</td>
<td>Usable, Operable, Sustainable</td>
</tr>
<tr>
<td>Expected time horizon</td>
<td>2 to 5 years</td>
</tr>
<tr>
<td>Type of innovation</td>
<td>Incremental, Breakthrough</td>
</tr>
</tbody>
</table>
| Abstract (up to 150 words) | **The problem:** Prediction performance of urban scale energy models remains a challenge because of the inability to quantify the influences urban context has on a building’s energy consumption. This inaccuracy can result in improper design or retrofit decisions and proliferate into unforeseen additional costs, emissions and/or energy usage.  

**The proposed solution:** A Data-driven Urban Energy Simulation (DUES) that integrates machine learning into a simulation-based workflow would improve energy prediction accuracy and would effectively support data-informed decision making for energy efficient design of buildings.  

**The proposed research approach:** By classifying output results provided by an energy simulation engine into spatial hierarchies (e.g., individual, community, urban scales), we will use machine learning models to quantify the impact of urban context on building energy use. This method will be developed and validated with a case study using city utility data or building portfolio data, in collaboration with a CIFE industry member. |
1 Engineering or Business Problem

The world is rapidly urbanizing. Over 50% of world population now resides in cities and the number is expected to increase to 67% (90% for the United States) by 2050 [1]. Cities account for over 75% of all primary energy usage and over 80% of greenhouse gas emissions, with the largest portion of such consumption (more than 40%) and emissions coming from the built environment [2, 3].

Urban buildings represent a tremendous opportunity to enhance the sustainability of cities as 90% of urban buildings are estimated to be energy inefficient and up to 30% of building energy consumption is wasted [4].

Extensive academic and industrial efforts have been undertaken to develop energy conservation measures within individual buildings (e.g., demand driven heating/cooling control). However, the key challenge to enhancing the energy efficiency of urban buildings is the lack of accurate energy performance prediction models. Current performance models fail to account for the intra-building energy dynamics and interdependencies that can have a substantial impact on the energy use of urban buildings. Without accurate performance characterization and prediction, designers and engineers struggle to assess the energy, environmental, and economic implications of their early-stage design and retrofit decisions thus failing to shape a building’s energy usage for its entire lifecycle. This challenge is further exacerbated as adjacent buildings and the overall urban area become increasingly energy intensive. For example, as shown in Figure 1, the performance of Building A, a large mixed-use office building, in a metropolitan city is significantly impacted by the diverse and convoluted urban context due to effects such as shading, thermal transfer and balance, fluid dynamics, urban heat island effect, and district level system service. Not taking into account such urban impacts could result in substantial energy, environmental, and economic impacts on both Building A and surrounding buildings. Inaccurately predicting performance by 10% could result in 2.8+ M kWh in extra energy usage [5], nearly 1.5 M lbs of CO₂ emissions [6], and $670k in additional costs per year [7] in Building A and have additional second-order impacts on surrounding buildings exceeding $10+ M. As a result, it is imperative that urban dynamics and context are accounted for in the performance prediction of urban buildings.

Figure 1. Problem: Buildings in urban environments are heavily impacted by their urban context (heat transfer, shading, urban heat island effect). However, current models do not account for such impacts resulting in inaccurate performance prediction. Photo adapted from San Francisco Office of Community Investment and Infrastructure [8]
Rapid development of new sensing technologies and emerging smart city initiatives have led to an explosion of data streams for urban buildings. The proliferation of data provides a basis for constructing building energy performance prediction models in an urban context. However, to date, limited work has been done to harness such data and translate it into insights for better building performance prediction. This proposal aims to develop a Data-driven Urban Energy Simulation (DUES) that fuses novel machine learning models from data science with traditional building energy performance simulations to enable accurate energy performance characterization of urban buildings at multiple scales.

2 Theoretical and Practical Points of Departure

The current workflow for creating an urban building energy model consists of (1) collecting and entering the required input parameters for each building being studied, (2) running the model in an energy simulation program, and (3) validating the results, typically by aggregating the outputs from each individual building into a single sum. Ongoing research in this field is primarily focused on the tradeoff between improving accuracy with increased model inputs and reducing the time required to generate a model of this size. However, there has been little work done to create these models on a multi-scale level, capture each building’s “urban context,” and integrate these physics-based models with data-driven models from machine learning to achieve improved prediction and characterization of urban building energy performance.

2.1 Data input

There are five primary specifications required in making a simulated, physics-based energy model: geometry, loads, construction, schedules, and assemblies. These five aspects have been condensed into three types of input data required to measure a building’s annual energy performance:

- **Geometric data** - tax parcel GIS shapefiles, merged with building footprints and heights
- **Non-geometric data** - building properties not associated with building footprints or location, e.g., construction assemblies, HVAC systems, occupancy schedules
- **Climate data** - the average weather conditions of a geographical area, divided into a single 8760-hour year, i.e., TMY files

The National Renewable Energy Laboratory (NREL) developed a database of standardized annual climate files called Typical Meteorological Files (TMY), which represent the most average weather conditions for a typical “virtual” year and serve as the environmental variables for simulation [9]. To generate a physical model of each building, tax assessment files, tax parcel GIS shapefiles, and associated building footprints and heights are merged to create “2.5D” massing models [10]. Lastly, each building must be further characterized with non-geometric data - the largest source of discrepancy between simulated and measured energy use [11]. To determine the values of these input parameters, energy engineers often rely on building codes, building audits, or general expertise and assumptions [12]. Because little to no data is used as evidence in generating the model, the simulated building energy use is often inaccurate at inception - a common characteristic for building energy models [13].

2.2 Thermal simulation

Energy simulators (e.g., EnergyPlus, DOE2, IES-VE, etc.) (1) take provided building geometries and abstract them to a network of connected nodes, (2) create heat balance equations for all
nodes across all 8760 hourly timesteps in the year, and (3) solve those equations, using the assumed non-geometric parameters to calculate a building’s energy consumption. As there are a large number of nodes and equations to solve, simulating the energy performance of hundreds of buildings across a city becomes too computationally expensive [14].

To overcome this issue, current approaches to urban energy simulation individually model a diverse subset of buildings and scale up those results to a city level by multiplying each individual output by the number of similar buildings [15] or by a floor area-weighted function [16]. However, in reducing the computational requirement for modeling an entire city, the model is no longer able to capture the “urban context” of surrounding buildings. Urban context refers to the considerations an energy engineer must make regarding a building’s external surroundings from both the physical and built environments. This includes shading, thermal transfer and balance, fluid dynamics and the urban heat island effect [17] - all of which can dramatically impact the heating and cooling demands of an urban building.

While some emerging tools aim to capture the urban context in the modeling process (e.g. Urban Modeling Interface (UMI) [18]), previous research has been limited to capturing simple dynamics such as shading due to the onerous process of specifying and co-simulating each individual building. We aim to overcome this limitation by employing emerging data-driven models from machine learning to infer the complex relationships of the urban context and improve prediction accuracy without increasing the modeling burden.

2.3 Integration of data-driven methods
Data has yet to be well integrated in generating urban building energy models. Determining non-geometric input parameters to an energy simulator is an ad-hoc process, requiring initial expertise and repeated tweaking to validate a model [19]. Simulation engines rely on thermodynamic equations rather than real data to predict hourly energy usage and rarely capture the influence of urban context on the final result. Because existing energy simulation tools are not capable of analyzing energy data, we see that addition of machine learning to existing workflows that can quantify hidden energy relationships in urban environments with computational ease. The goal of our proposed study is to leverage emerging machine learning models to capture the influence of urban context on energy usage on multiple spatial scales.

3 Research Methods and Work Plan
Our approach to generating a Data-driven Urban Energy Simulation (DUES) consists of two phases: building an energy simulation model and creating a complementary machine learning network model, as shown in Figure 2.

3.1 Energy Simulation Model
Phase 1 will generate an energy simulation using machine learning that can be applied to understanding the relationships of energy consumption between buildings. Tax parcel and building GIS shapefiles, commonplace in many citywide GIS databases, will be merged with associated building heights to construct the “2.5D” massing model via SketchUp - Google’s 3D modeling software [20]. Using the SketchUp plugin, OpenStudio [21], these geometries will be imported into the EnergyPlus thermal simulation [22] along with climate data corresponding to the geographic area of study. To generate the non-geometric properties, the building stock will be classified by use type and age of construction. Using these classes, input parameters can be
defined based on the U.S. Department of Energy’s Commercial Reference Building Models [23]. The DOE and several of its national laboratories used national data from the 2003 Commercial Building Energy Consumption Survey (CBECS) to determine an average mix of representative buildings. These reference models include 16 building types for 3 different construction periods and represent about 70% of the US commercial building stock. Each of these models is available as a ZIP file with detailed spreadsheets of plug and process loads, construction assemblies, operating schedules, and systems. Once all inputs for each building are defined in EnergyPlus, the model will be simulated and results exported for further network analysis.

3.2 Network Model

Given the newly generated energy models for individual buildings and the ground truth energy data for all buildings under certain urban context, network analysis can be implemented to recognize the hidden energy connections and interdependencies of urban buildings. Machine learning uses massive computational power to effectively model the nonlinear and varying relationships between the simulated and measured energy use for a network of buildings at multiple scales (e.g., individual building scale, community scale, and urban scale) simultaneously. It will be applied in the network analysis to ingest the simulated results of individual buildings and produce energy use information for buildings with urban context taken into consideration. There are five steps: (1) Define the hierarchies of scales that will satisfy the accuracy of the final urban energy model. A tree is built to specify the buildings in each hierarchy. For example, Building A, Building B, and Building C are at Urban Scale 1, while Building A and Building B are at Community Scale 1 (see Figure 3). The hierarchies then determine the structure of input (simulated energy use) and output (measured energy use); (2) Initiate a network with multiple layers to build an improved feature space for connecting the input and output. The different layers will learn different orders of features that represent the non-linear and joint influences as well as the interconnections of intra-building energy use; (3) Set initial values to the hyperparameters of the network and renormalize activations at each layer in order to keep the balanced learning across all layers; (4) Tune the entire network using supervised training, and determine the
hyperparameters including network structure, size of filters, weightings, connectivity between layers, activation functions through iterative performance evaluation; (5) Validate the network to avoid overfitting. Cross-validation will be applied by which the data are separated into different segments. Some segments will be used to train the network, and the remaining ones will be used to verify whether the network can successfully process the simulated results. The ability of the network to quantify the influences of changes in one building can be also tested. The different types of networks (e.g., convolutional neural network, belief network, reinforcement network, recurrent neural network, etc.) representing different forms and emphasis of the energy connections and interdependencies of urban buildings such as shading, thermal transfer, fluid dynamics, district system service, and heat island effect will be all evaluated, and the pros and cons will be compared to determine the most appropriate network.

4 Expected Results: Findings, Contributions, and Impact on Practice
The most significant anticipated outcome from this project is a Data-driven Urban Energy Simulation (DUES) that combines the computational power of a data-driven machine learning network model and interpretability of physics-based energy simulation. We expect this model and its associated workflow will be generalizable for any city or dense building portfolio data and will be able to simulate energy consumption on an individual, community, and urban scale. The research team is unaware of an existing model that integrates machine learning and simulation into a detailed urban energy model.

By visualizing the energy usage of buildings across a city, policymakers will have a better awareness of the effects of new citywide interventions. Designers and building operators will understand the effects of energy consumption and indoor environmental quality on not only their building, but surrounding ones as well. Finally, as data is being increasingly used in the planning of newer, smarter cities, this model can employ data from existing cities to optimize building energy use and help inform key decisions related to energy efficiency early on in the design process and as part of retrofit programs.

5 Industry Involvement
The research team plans to engage multiple members of CIFE to assist us in producing this model. Designers such as HOK would provide us with insight on how our tool would integrate with the design of urban spaces, especially considering their planning group’s specific focus on data-driven approaches to master planning [24]. We expect to interact with members who manage large campuses and building portfolios (e.g., Apple, Walt Disney Imagineering). Additionally, in working with producers of industry-standardized modeling tools (e.g., Autodesk), we hope to create a tool that could be easily adopted by cities across the United States.

6 Research Milestones and Risks
Four milestones correspond to the research methods and work plan above and will be used to measure progress of the proposed project (see Table 1). A multi-layer network model capable of understanding building energy use on multiple spatial scales will require the initial creation of a standard urban energy model, as detailed in Section 3.1. Using the outputs from the initial simulation, supervised learning techniques will allow us to calibrate the network model and complete the hybrid simulation. Validation will consist of a case study using real city or building portfolio data to ensure acceptable accuracy and interpretability for future application. Finally,
the results of the theoretical model and validation study will be documented, reported, and disseminated through publication in a top-tier peer-reviewed journal (e.g. *Applied Energy, Energy and Buildings, Building and Environment*) as well as through our lab website (for external CIFE partners).

Table 1. Proposed project schedule

<table>
<thead>
<tr>
<th>Network Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid Simulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation - Case Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper Submission</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Risks:** The largest risk to our proposed research is the availability of geographically dense energy use data in a city or other building portfolio. This type of data is critical to our model being able to learn the relationships between urban context and building energy consumption. While many cities across the United States are passing laws mandating the disclosure and collection of energy usage data, this is only required of buildings meeting a minimum square footage threshold. We plan to mitigate this issue through existing partnerships with local municipal governments and developing new partnerships with cities such as San Francisco, Chicago, and Washington DC, as well as working with a CIFE owner/operator that manages a campus of buildings within close proximity of one another.

**7 Next Steps**

This SEED grant supports a pilot study looking to develop a DUES, a novel data-driven urban energy simulation capable of understanding the relationships between building energy usage and the surrounding urban environment. We aim to leverage the results of this SEED grant to pursue subsequent funding through the National Science Foundation’s new initiative on Simulated and Synthetic Data for Infrastructure Modeling as well as the existing industry partnership programs in the Department of Energy’s Building Technologies Office (BTO).
References


Sponsor: CIFE
Submission Type: New
Budget Preparation Date: 4/11/2017
Budget Start Date: 10/1/2017
Project Name: CIFE AY2017-18
Department: Civil Engineering
Principal Investigator: Rishee Jain
Administrator: Roosmery Yang

<table>
<thead>
<tr>
<th>Period</th>
<th>From</th>
<th>All Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10/1/2017</td>
<td>10/1/2017</td>
</tr>
<tr>
<td></td>
<td>9/30/2018</td>
<td>9/30/2018</td>
</tr>
</tbody>
</table>

Personnel Salaries

<table>
<thead>
<tr>
<th>Faculty</th>
<th>Summer</th>
<th>%</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jain, Rishee</td>
<td></td>
<td>10.0%</td>
<td>3,828</td>
<td>3,828</td>
</tr>
<tr>
<td></td>
<td>Total Faculty Salaries</td>
<td>3,828</td>
<td>3,828</td>
<td></td>
</tr>
</tbody>
</table>

Graduate Students

<table>
<thead>
<tr>
<th>TBA</th>
<th>Academic</th>
<th>50.0%</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summer</td>
<td>50.0%</td>
<td>10,018</td>
<td>10,018</td>
</tr>
<tr>
<td></td>
<td>Total Graduate Student Salaries</td>
<td>40,071</td>
<td>40,071</td>
<td></td>
</tr>
</tbody>
</table>

Total Salaries

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Salaries</td>
<td>43,899</td>
<td>43,899</td>
</tr>
</tbody>
</table>

Benefits

<table>
<thead>
<tr>
<th>Faculty</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,242</td>
<td>1,242</td>
</tr>
<tr>
<td>Graduate</td>
<td>2,164</td>
<td>2,164</td>
</tr>
<tr>
<td>Total Benefits</td>
<td>3,406</td>
<td>3,406</td>
</tr>
</tbody>
</table>

Total Salaries and Benefits

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Salaries and Benefits</td>
<td>47,305</td>
<td>47,305</td>
</tr>
</tbody>
</table>

Other Direct Costs

<table>
<thead>
<tr>
<th>Tuition</th>
<th>Academic</th>
<th>50.0%</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBA</td>
<td>Summer</td>
<td>50.0%</td>
<td>6,820</td>
<td>6,820</td>
</tr>
<tr>
<td>Total Tuition</td>
<td>27,281</td>
<td>27,281</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total Other Direct Costs

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Other Direct Costs</td>
<td>27,281</td>
<td>27,281</td>
</tr>
</tbody>
</table>

Total Direct Costs

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Direct Costs</td>
<td>74,586</td>
<td>74,586</td>
</tr>
</tbody>
</table>

Less

<table>
<thead>
<tr>
<th>Tuition</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(27,281)</td>
<td>(27,281)</td>
</tr>
</tbody>
</table>

Modified Total Direct Costs

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Total Direct Costs</td>
<td>47,305</td>
<td>47,305</td>
</tr>
</tbody>
</table>

Indirect Costs

<table>
<thead>
<tr>
<th>IDC on MTDC</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total IDC Costs</td>
<td>3,784</td>
<td>3,784</td>
</tr>
</tbody>
</table>

Total Amount Requested

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Amount Requested</td>
<td>78,370</td>
<td>78,370</td>
</tr>
</tbody>
</table>

Benefits Rates

<table>
<thead>
<tr>
<th>Faculty</th>
<th>FY 1</th>
<th>FY 2</th>
<th>FY 3+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>32.45%</td>
<td>32.45%</td>
<td>32.45%</td>
</tr>
<tr>
<td>Graduate</td>
<td>FY 1</td>
<td>FY 2</td>
<td>FY 3+</td>
</tr>
<tr>
<td></td>
<td>05.40%</td>
<td>05.40%</td>
<td>05.40%</td>
</tr>
</tbody>
</table>

Indirect Cost Rates

<table>
<thead>
<tr>
<th>Campus</th>
<th>FY 1</th>
<th>FY 2</th>
<th>FY 3+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.00%</td>
<td>8.00%</td>
<td>8.00%</td>
</tr>
</tbody>
</table>