Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 142 (2017) 2114–2119

Energy Procedia

www.elsevier.com/locate/procedia

9th International Conference on Applied Energy, ICAE2017, 21-24 August 2017, Cardiff, UK

Data-driven Urban Energy Simulation (DUE-S): Integrating machine learning into an urban building energy simulation workflow

Alex Nutkiewicz[a], Zheng Yang[a], Rishee K. Jain[a,*]

[a]*Urban Informatics Lab, Dept. of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA 94305, USA*

Abstract

Urban building energy models are emerging tools meant to analyze and understand the energy performance of multiple buildings within a dense urban area. However, accurate performance prediction of these models remains a challenge because of their inability to account for the inter-building energy dynamics and interdependencies in an urban area. This paper analyzes the literature to highlight the limitations of current urban scale energy simulation models and proposes a new Data-driven Urban Energy Simulation (DUE-S) workflow capable of capturing inter-building effects on a building's energy usage. Specifically, DUE-S combines a data-driven machine learning model with a traditional physics-based energy simulation to enable more accurate simulation results on multiple scales (single building, community, urban). More accurate and robust energy performance characterization and simulations of urban buildings could provide valuable insight on early-stage building design, building conservation and portfolio management, and urban energy efficiency policy-making crucial to helping our cities transition to a more sustainable energy future.

Peer-review under responsibility of the scientific committee of the 9th International Conference on Applied Energy.

Keywords: building performance; energy efficiency; energy data; machine learning; simulation; urban sustainability

1. Introduction

The world is rapidly urbanizing. Over 50% of the 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

* Corresponding author. Tel.: 1-650-721-8559; fax: +1-650-723-4806.
E-mail address: rishee.jain@stanford.edu

1876-6102
Peer-review under responsibility of the scientific committee of the 9th International Conference on Applied Energy.
10.1016/j.egypro.2017.12.614

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]. Because 90% of urban buildings are estimated to be energy inefficient and up to 30% of a building's energy consumption is wasted, urban buildings represent a tremendous opportunity to enhance the sustainability of cities [3].

Extensive academic and industrial efforts have been undertaken to develop energy conservation measures within individual buildings (e.g., demand driven heating/cooling control). However, a key challenge to enhancing the energy efficiency of urban buildings is the inaccuracy in energy performance prediction models. Current performance models fail to account for the inter-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, resulting in further substantial energy, environmental, and monetary impacts.

Rapid development of new sensing technologies and emerging smart city initiatives have led to an explosion of structured and unstructured data streams describing buildings and their urban environments. The proliferation of such data provides an opportunity to enhance the accuracy and robustness of urban building energy models. This paper aims to understand the gaps of current urban scale energy simulation and proposes a new Data-driven Urban Energy Simulation (DUE-S) workflow capable of accurately predicting the energy performance of urban buildings across multiple spatial scales within a city.

2. Key Factors in Urban Energy Simulation

2.1. Urban context

Building energy models (e.g., EnergyPlus, DOE2, etc.) (1) take inputted building geometries and abstract them to a network of connected nodes, (2) create heat balance equations for all nodes across each time step of a virtual year, and (3) solve those equations simultaneously, using a large number of assumed non-geometric building parameters to calculate a building's energy consumption. However, because there are a large number of nodes and equations to solve, simulating the performance of hundreds of buildings across a city is too computationally expensive [4]. Efforts have been made to simplify this process by instead individually simulating a diverse subset of buildings and scaling up those results to an urban level by multiplying each individual output by the number of similar buildings [5] or by a floor area-weighted function [6]. However, in reducing the computational requirement of modeling an entire city, the model can no longer 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 urban heat island effects [7] - all of which can dramatically impact the energy demands of an urban building.

Ongoing research to quantify urban context is primarily focused on modeling microclimatic factors alongside existing simulation software. Several studies, for example, are quantifying the impact of radiation exchange with building energy consumption - some also including the heat and mass transfer phenomena associated with it. ENVI-met, for example, a three-dimensional microclimate model, has been integrated into EnergyPlus in order to quantify the interactive effects of CFD and thermodynamics on building energy consumption [8]. Other architecturally based tools such as the Urban Modeling Interface (UMI) are focused on modeling conditions such as shading and urban heat island effects in the context of urban planning [9]. While emerging tools incorporate some aspects of building interdependencies in energy performance models, previous research is limited in its ability to comprehensively capture all the influences of the urban context due to the time- and resource-intensive process of simultaneously simulating each individual building and their associated mutual impacts.

2.2. Multi-scale calibration and performance

Current building energy simulation practices generally focus on single scale (e.g., individual building level) energy modeling [10] and multiple calibrations within a single building [11]. Little work to date simultaneously achieves accurate energy simulation results beyond a single spatial scale (e.g., community scale), which, however, is of great importance for evaluating and improving the energy performance of urban buildings. For example, building scale accuracy can represent the energy implications of building operations and subsequently guide facility management and retrofitting for energy efficiency; community scale accuracy can reveal the energy consequences of urban system interactions and assist the optimization of energy use across a building portfolio; urban scale accuracy can provide insights on the overall energy performance of an urban area and support decision-making regarding sustainable zoning and other planning guidelines that impact a city's morphology. In addition, different scales of simulations are closely associated with each other. Even if the focus of an analysis is on a single scale, energy simulation of other scales is necessary for comprehensive analysis that ensures design and energy management decision-making accounts for impacts from outside the particular building of study. However, achieving high accuracy at one scale does not necessarily guarantee high accuracy at another [12], especially when the analyzed buildings have heterogeneous thermal properties (e.g., different window-wall ratios), inter-building system connections (e.g., district heating system), and functional relations (e.g., an occupant network). Moreover, obtaining high accuracies at multiple scales becomes increasingly difficult as the complexity increases exponentially – often resulting from nonlinear and dynamic interactions and interdependencies of thermal balances, geometric impacts, fluid dynamics, occupant network, and system responses in urban buildings. Therefore, a method to accurately predict energy consumption on increasingly large geographic areas is critical to understanding these factors.

2.3. Model inputs and assumptions

As discussed previously, only if an energy performance model is accurate at different scales is it reliable and useful to evaluate and support the energy conservation measures for urban buildings. Simulation is a context-related process, and results from uncalibrated models tend to deviate significantly (up to 90%) from actual energy use [13], making calibration indispensable before any practical use can be made of the model. Energy model calibration is an inverse approximation because of the need for determining values of inputs to reconcile the simulation outputs to the measured energy data. It is also over-parameterized because of the large number of independent and interdependent input parameters and assumptions to be specified, given the relatively limited and variant information and evidence available to determine these inputs. The inverse and over-parameterized characteristics of calibration directly result in the uncertainty in these input parameters and misassumption of thermal processes, which are the main sources of building energy simulation discrepancies [14]. In order to overcome these issues, engineering methods have been developed to iteratively adjust input parameters and assumptions based on audit data [15] until the discrepancies between simulated results match the measurements are within the ranges regulated by standards [16,17]. Since the availability of audit data differs building by building, the hierarchy of audit data has been generated, by which the higher-level data is considered more reliable than the lower ones for determining the input parameters and assumptions [18]. However, such methods follow the trial-and-error principle and as a result the performance of an energy model depends largely on ad-hoc audit data and subjective engineering judgment [19].

Other research efforts have focused on statistical methods that apply multivariable analysis to find the relationships between audit data and actual energy use [20,21]. The audit data can either be static (e.g. window-wall ratio) or dynamic (e.g., zone temperature). The relationships are the mathematical projections that do not necessarily have physical explanations. Once the relationships are established, they can be used to ingest new audit data and output corresponding energy use. This type of method is computationally efficient and can provide for more accurate energy simulation [22]. However, accurately accounting for relationships requires retraining if any change is made to the building (e.g., additions) or its building systems (e.g., HVAC control strategies). More importantly, since the statistical relationships are constructed on a building-to-building basis, they are sometimes overspecialized and may not be generalized to other buildings, making it is impossible to mix the heterogeneous relationships for large-scale urban building energy models.

Unfortunately, no generally accepted method is able to effectively determine the input parameters and assumptions at the individual building scale because each building is unique, and building energy simulation has variant simulation requirements, variant available building information, variant human-building interactions, and variant levels of knowledge of modelers. Urban scale building energy simulation – the extension of individual building simulation – is significantly influenced by inter-building interactions and interconnections and thus contains much more complicated energy dynamics and uncertainties. As a result, previous work has been limited in its ability to incorporate such inter-building effects into the process of determining energy model input parameters and assumptions.

3. Data-driven Urban Energy Simulation (DUE-S)

Data has yet to be well integrated in generating urban scale building energy models. Simulation engines largely rely on thermodynamic calculations rather than real data to predict hourly energy usage and rarely captures the influence of urban context in the final result. As a result, the addition of machine learning to existing energy simulation workflows could result in more accurate, robust and scalable urban energy simulations by employing computational intelligence to learn the hidden and complex dynamics between urban buildings. Our proposed modeling workflow, DUE-S (Fig. 1), leverages emerging machine learning capabilities to capture the influence of the urban context on energy consumption on multiple spatial scales.

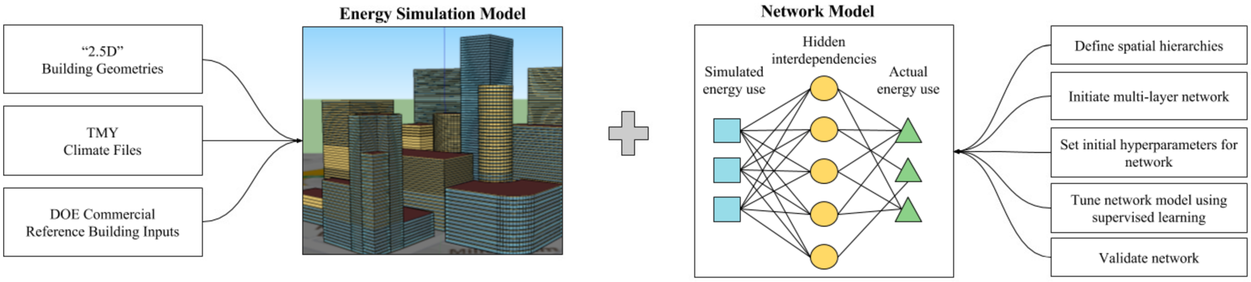

Fig. 1. Proposed DUE-S modeling workflow. Phase 1 uses information from existing data sources to create an energy simulation. In Phase 2, machine learning is employed to learn the energy use dynamics between buildings and their surrounding urban environment.

3.1. Thermal Models

Traditional building energy models are comprised of three primary inputs [23]:

- weather data
- geometric inputs
- non-geometric parameters

As mentioned prior, energy simulation is often over-parameterized and requires an infeasible amount of time and resources to complete on a larger, urban scale. The first phase of this hybrid simulation model is generating simplified energy simulation models for individual buildings in the urban study area.

In our proposed thermal model, we merge tax parcel and building GIS shapefiles, commonplace in many citywide GIS databases, with associated building heights information (e.g., satellite LiDAR data) to construct the "2.5D" massing model via SketchUp - Google's 3D modeling software. Using the SketchUp plugin, OpenStudio, we import these geometries into the EnergyPlus thermal simulation along with climate data corresponding to the geographic area of study. To generate the non-geometric properties, the building stock is classified by use type and age of construction. Using these classes, we define input parameters based on the U.S. Department of Energy's Commercial Reference Building Models [24]. 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. Other input parameters that are not specified by the reference models will be given default values. Once all inputs for each building are defined in EnergyPlus, the energy usage for each building in our urban study area will be simulated. We stress here that the goal of these thermal models is to capture the basic energy usage dynamics of each building and the output (i.e., simulated energy usage) will be utilized as an input in subsequent steps of our proposed modeling workflow. As a result, high accuracy and calibration is not required at this stage.

3.2. Network-based Machine Learning Model

The simulated energy usage signatures and ground truth (measured) energy data for all buildings in the urban study area will be fed into a network-based machine learning model to learn the hidden energy connections and interdependencies between buildings at multiple scales (e.g., individual building scale, community scale, and urban scale). There are five steps to the network-based machine learning model:

1. Define the hierarchies of scales that satisfy the requirements of the final urban energy model. For example, the *Urban* scale is made up of *Communities 1*, *2*, and *3*, where each community (e.g., Community 2) is made up individual *Buildings A*, *B*, *C*, and *D* (Fig. 2). The hierarchies determine the structure of inputs (simulated energy use) and outputs (measured energy use);
2. Initiate a network with multiple layers to build an improved feature space for connecting the inputs and outputs. The different layers will learn different orders of features that represent the non-linear and joint influences as well as the interconnections of inter-building energy use;
3. Set initial values to the hyper-parameters 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 hyper-parameters including network structure, size of filters, weightings, connectivity between layers, and activation functions through iterative performance evaluation;
5. Validate the network to avoid overfitting. Cross-validation is applied by separating the data into segments. Some segments will be used to train the network, and the remaining ones will be used to verify if the network can successfully process the simulated results. The network's ability to quantify the influences of changes in one building on the energy usage of adjacent buildings will also be tested.

Various types of network-based machine learning methods (e.g., convolutional neural network, belief network, reinforcement network, recurrent neural network, etc.) will all be evaluated, and the pros and cons will be compared to determine the most appropriate network-based method.

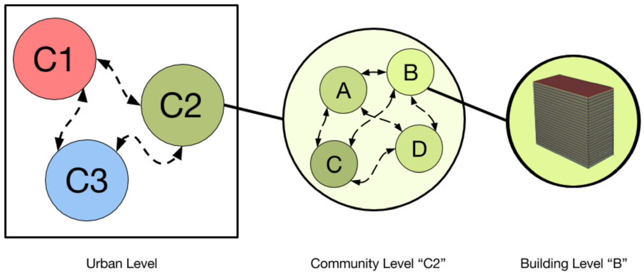

Fig. 2. Hierarchy for accurately quantifying energy use on different spatial scales. Urban scale energy use will analyze the relationships (hashed arrows) between Communities C1, C2, and C3 – each of which are comprised of relationships (hashed arrows) between individual buildings.

4. Conclusion and future work

As a result of shortcomings in urban scale energy simulation, this paper proposes a data-driven urban energy simulation (DUE-S) workflow that combines the computational power of a data-driven machine learning network model with the advanced thermodynamic modeling of traditional building energy simulation. Current models can only simulate buildings on an individual level and, therefore, their ability to evaluate the performance of a larger urban area is limited. Our review of the literature highlights the significant opportunity that exists to create more accurate multi-scale urban energy models if emerging machine learning models are integrated into the current energy simulation workflow. Future work aims to further develop, test and validate using empirical data from a case study city and compare our results to existing simulation methods.

By accurately simulating the energy usage of buildings across a city, policymakers and utility companies can better forecast energy loads and identify opportunities for energy efficiency and energy management (e.g., retrofits, load shifting) across a community or city. Designers and building operators can understand the effects of energy consumption and indoor environmental quality on not only their building, but surrounding ones as well, resulting in collective energy, emissions, and monetary savings. And 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 that enable our cities to transition to a more sustainable energy future.

Acknowledgements

The material presented in this manuscript is based in part upon work supported by the UPS Foundation and the US National Science Foundation (NSF) under Grants No. 1461549, 1642315. 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

[1] United Nations. World's population increasingly urban with more than half living in urban areas. 2014. <http://www.un.org/en/development/desa/news/population/world- urbanization-prospects-2014.html>.
[2] United States Energy Information Administration. Annual Energy Outlook 2017. 2017. <https://www.eia.gov/outlooks/aeo/>.
[3] Breaking Energy. The top ten ways we waste energy and water in buildings. 2011. <http://breakingenergy.com/2011/07/26/the-top-ten-ways-we-waste-energy-and-water-in-buildings/>.
[4] Davila, CC, Reinhart, CF, and Bemis, JL. Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. Energy 2016; 117: 237–250.
[5] Firth, SK., Lomas, KJ, and Wright AJ. Investigating CO2 emission reductions in existing urban housing using a community domestic energy model." Building Simulation 2009; 9: 2098-2105.
[6] Galante, A, and Torri, M. A methodology for the energy performance classification of residential building stock on an urban scale. Energy and Buildings 2012; 48: 211-219.
[7] Bouyer, J., Inard, C., and Musy, M. Microclimatic coupling as a solution to improve building energy simulation in an urban context. Energy and Buildings 2011 ; 43(7): 1549–1559.
[8] Morakinyo, TE, Dahanayake, KKC, Adegun, OB, and Balogun, AA. Modelling the effect of tree-shading on summer indoor and outdoor thermal condition of two similar buildings in a Nigerian university. Energy and Buildings 2016; 130: 721–732
[9] Reinhart, CF, Dogan, T, Jakubiec, JA, Rakha, T, and Sang, A. Umi - an urban simulation environment for building energy use , daylighting and walkability. Proceedings of Conference of International Building Performance Simulation Association. 2013: 476–483.
[10] Kim, YS, Heidarinejad, M, Dahlhausen, M, and Srebric, J. Building energy model calibration with schedules derived from electricity use data. Applied Energy 2017; 190: 997-1007.
[11] Yang, Z, Becerik-Gerber, B. A model calibration framework for simultaneous multi-level building energy simulation. Applied Energy 2015; 149: 415-431.
[12] Booth, AT, Choudhary, R, and Spiegelhalter, DJ. A hierarchical Bayesian framework for calibrating micro-level models with macro-level data. Journal of Building Performance Simulation 2013; 6(4): 293-318.
[13] Li, N, Yang, Z, Becerik-Gerber, B, Tang, C, Chen, NL. Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?. Applied Energy 2015; 159: 196-205.
[14] Manfren, M, Aste, N, and Moshksar, R. Calibration and uncertainty analysis for computer models–a meta-model based approach for integrated building energy simulation. Applied Energy 2013; 103: 627-641.
[15] Monfet, D, Charneux, R, Zmeureanu, R, and Lemire, N. Calibration of a building energy model using measured data. ASHRAE Transactions 2009; 115: 348-360.
[16] ASHRAE. ASHRAE guideline 14-2002: measurement of energy and demand savings. 2002.
[17] United States Department of Energy. M&V guidelines: measurement and verification for federal energy projects version 3.0. 2008.
[18] Coakley, D, Raftery, P, Molloy, P, and White, G. Calibration of a detailed BES model to measured data using an evidence-based analytical optimisation approach. Proceedings of Conference of International Building Performance Simulation Association. 2011: 374-381.
[19] Reddy TA, Maor I, and Panjapornpon C. Calibrating detailed building energy simulation programs with measured data – Part I: general methodology (RP- 1051). HVAC & R Research 2007; 13: 221-241.
[20] Magnier, L, and Haghighat, F. Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. Building and Environment 2010; 45(3): 739-746.
[21] Karunakaran, R, Iniyan, S, and Goic, R. Energy efficient fuzzy based combined variable refrigerant volume and variable air volume air conditioning system for buildings. Applied Energy 2010; 87(4): 1158-1175.
[22] Sanyal, J, New, J, and Edwards, R. Supercomputer assisted generation of machine learning agents for the calibration of building energy models." Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery. 2013.
[23] Fabrizio, E, and Monetti, V. Methodologies and advancements in the calibration of building energy models. Energies 2015; 8(4): 2548–2574.
[24] U.S. Department of Energy. "Commercial Reference Buildings. <https://energy.gov/eere/buildings/commercial-reference-buildings>