Lawrence Berkeley National Laboratory
LBL Publications

Title
Surrogate modelling for urban building energy simulation based on the bidirectional long short-term memory model

Permalink
https://escholarship.org/uc/item/84z4t4f2

Journal
Journal of Building Performance Simulation, ahead-of-print(ahead-of-print)

ISSN
1940-1493

Authors
Pan, Xiyu
Xu, Yujie
Hong, Tianzhen

Publication Date
2024

DOI
10.1080/19401493.2024.2359985

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at https://creativecommons.org/licenses/by-nc/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Building Technologies & Urban Systems Division
Energy Technologies Area
Lawrence Berkeley National Laboratory

Surrogate Modeling for Urban Building Energy Simulation Based on the Bidirectional Long Short-Term Memory Model

Xiyu Pan[1] [2], Yujie Xu[1], Tianzhen Hong[1]

[1]Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, USA, [2]School of Civil and Environmental Engineering, Georgia Institute of Technology, USA

Energy Technologies Area
May 2024

doi.org/10.1080/19401493.2024.2359985

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the US Department of Energy under Contract No. DE-AC02-05CH11231.

Disclaimer:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Journal of Building Performance Simulation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tbps20

Surrogate modelling for urban building energy simulation based on the bidirectional long short-term memory model

Xiyu Pan, Yujie Xu & Tianzhen Hong

To cite this article: Xiyu Pan, Yujie Xu & Tianzhen Hong (28 May 2024): Surrogate modelling for urban building energy simulation based on the bidirectional long short-term memory model, Journal of Building Performance Simulation, DOI: 10.1080/19401493.2024.2359985

To link to this article: https://doi.org/10.1080/19401493.2024.2359985

Published online: 28 May 2024.

Submit your article to this journal

Article views: 87

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tbps20

JOURNAL OF BUILDING PERFORMANCE SIMULATION
https://doi.org/10.1080/19401493.2024.2359985

Taylor & Francis
Taylor & Francis Group

Check for updates

Surrogate modelling for urban building energy simulation based on the bidirectional long short-term memory model

Xiyu Pan[a,b], Yujie Xu [a] and Tianzhen Hong[a]

[a]Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [b]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT

The urban microclimate is essential for accurate simulation-based urban building energy modelling (UBEM). However, a high spatial-resolution microclimate can increase the computational resources demands of UBEM. Surrogate modelling is one of the promising approaches for fast UBEM. This study proposes a bidirectional Long Short-Term Memory (LSTM)-based approach for simulation-based UBEM surrogate modelling. The estimations are aggregated into census tracts using total building floor area. A case study using UBEM to estimate annual hourly building energy use and anthropogenic heat from all existing buildings in Los Angeles County found that most of the surrogate models can complete the annual hourly simulation within 90 minutes with a normalized mean absolute error lower than 10%, and that the bidirectional LSTM outperforms the standard LSTM in accuracy. This study demonstrates the advantages of bidirectional RNN architecture in building energy surrogate modelling and is expected to promote long-term and high-resolution UBEM with detailed microclimates.

ARTICLE HISTORY
Received 8 November 2023
Accepted 21 May 2024

KEYWORDS
Bidirectional LSTM model; anthropogenic heat; building stock; energy use; surrogate model; machine learning

1. Introduction

Urban Building Energy Modelling (UBEM) is the 'computational modelling and simulation of the performance of a group of buildings in the urban context,' and it includes not only the dynamics of individual buildings (e.g. energy use and demand) but also interactions between buildings (e.g. solar reflection) and between buildings and urban microclimates (e.g. building heat emission) (Hong et al. 2020). The spatiotemporal estimation of building performance by UBEM is a critical reference for urban-scale building energy performance evaluation, energy policy-making, building anthropogenic heating estimation, district planning and retrofitting, and more.

One of the most commonly used approaches for UBEM is to utilize simulation models to capture the physical dynamics of building performance at the individual building level and then scale the estimation up to the urban scale (Li et al. 2020). The simulation-based UBEM reduces the amount of historical data needed (Abbasabadi and Mehdi Ashayeri 2019), and is particularly important for estimating various types of building performance (e.g. building heat emission) that are difficult to measure or to collect data on. The simulation model-based approach also enables estimation under a 'what-if' scenario, which is useful for many use cases, such as UBEM-based design modifications and long-term future estimates under climate change.

Unlike modelling a single building in a fixed location, UBEM requires separate simulations for buildings or groups of buildings under a variety of microclimate conditions. The integration of microclimate in UBEM is essential; many studies have supported the idea that urban microclimate has significant impacts on the accuracy of thermal load estimation, and thus building energy performance (Hong et al. 2020). A typical example is the effect of urban heat island (UHI). A recent study in the city centre of Rome, Italy (Mediterranean climate), showed there could be a 35% – 50% underestimation of the cooling load if the climatic effect of UHI is not considered (Ciancio et al. 2018). Nevertheless, the consideration of microclimate dramatically increases the number of simulations needed for a UBEM, and the number of microclimatic zones grow quadratically with smaller sizes of microclimate zones and a higher resolution of UBEM. As will be shown in this paper, for a 12 × 12-kilometre microclimate zone, Los Angeles County's UBEM needs to model 62 separate microclimate zones. This number

CONTACT Tianzhen Hong thong@lbl.gov Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

© 2024 International Building Performance Simulation Association (IBPSA)

would increase to nearly 9,000 if the 1×1 kilometre microclimate zone size suggested by previous studies (Sezer et al. 2023) is adopted. The number of microclimates and corresponding simulations necessary for modelling can be even higher in the background of climate change and the need for long-term UBEMs.

Building surrogate models for simulation is a promising strategy to reduce the computational resources and time required for UBEMs. The idea of surrogate modelling for UBEM is to emulate computationally expensive building energy simulations using a statistical model. The statistical model is trained using the inputs and outputs from a smaller number of simulations. Once trained with satisfactory accuracy, the statistical model can estimate the outputs of simulations without actually running the simulation model, within a small amount of time, and with unseen inputs (Westermann and Evins 2019). In prior studies, building energy simulation (BES) surrogate modelling has been extensively used for model calibration (Herbinger, Vandenhof, and Kummert 2023), uncertainty analysis (Fu et al. 2020), sensitivity analysis (Tian et al. 2015), performance optimization (Yigit 2021), and building predictive control (Taheri, Hosseini, and Razban 2022). As an example, Magnier and Haghighat managed to speed up multi-objective building optimization using the surrogate model, which utilizes building properties (e.g. the thickness of concrete) and heating, ventilation, and air conditioning (HVAC) system behaviours (e.g. heating setpoint) as features (Magnier and Haghighat 2010). Despite its relevance to UBEM surrogate modelling, BES surrogate modelling usually does not consider varying microclimates (Liang et al. 2022). For instance, the BES for optimizing building design at a fixed location usually uses building design variables (e.g. external insulation) as features and uses weather conditions as merely fixed parameters (Prada, Gasparella, and Baggio 2018). However, as mentioned above, microclimates are essential for accurate UBEM surrogate modelling.

For simulation-based UBEMs with varying microclimate surrounding the buildings, one of the most important features for surrogate modelling is the weather feature (e.g. outdoor dry bulb temperature, relative humidity), which has a temporal sequential nature. The lack of modelling temporal pattern is unfavourable for accurate UBEM surrogate modelling (Yoo, Clayton, and Yan 2023). Therefore, although the deep feed-forward neural network is popular in BES surrogate models because of its simplicity (Zhang et al. 2021), it is not a good option for UBEMs that are affected by microclimates. One of the most effective approaches is using a one-dimensional convolutional neural network (1D-CNN) (Zhou et al. 2022) to encode the dynamic weather features in microclimates and to form multivariate time series forecasting models. In a typical study of utilizing 1D-CNN in building energy surrogate modelling with multiple climate zones, Westermann et al. used kernels with an eight-time-step length to encode the annual hourly weather features and decode the representation to the annual estimation for the target variable (Westermann, Welzel, and Evins 2020). One of the advantages of 1D-CNN over the basic RNNs, including LSTM and Gated recurrent unit (GRU), is the assumption that the target is largely influenced by its neighbourhood. The neighbourhood is not limited to the features at the previous time steps, and it also includes the features at the subsequent (i.e. the future) time steps if needed.

Nevertheless, RNNs are still very popular because their structure is specially designed for time-series inputs and can achieve good accuracy (Cohen et al. 2021; Pinto, Deltetto, and Capozzoli 2021; Li et al. 2022). In RNN-based surrogate modelling, features usually have a high temporal resolution (e.g. hours) and form time series (Ohta, Sasakawa, and Sato 2020; Li et al. 2022). The weather features and other dynamic features are processed by moving a sliding window over the time series. The sliding window stops at each time step (e.g. hour) and creates a sequence of features being used to estimate targets at one or more time steps. Previous studies comparing the accuracy of RNNs and 1D-CNNs on time series forecasting have demonstrated that RNNs can achieve better accuracy than 1D-CNNs (Chandra, Goyal, and Gupta 2021; Lara-Benítez, Carranza-García, and Riquelme 2021). Despite the controversy over this finding (Bai, Kolter, and Koltun 2018), RNNs remain an important approach for UBEM surrogate modelling, and this study is therefore focused on the RNN-based approaches. In the related literature, Ohta et al. used LSTM to estimate the thermal comfort level and the power consumption for optimizing the air conditioner setpoint (Ohta, Sasakawa, and Sato 2020). In another example, Li et al. used the time-lagged values of the outdoor air dry bulb temperature, relative humidity, etc. to estimate the hourly values of the heating and cooling energy load (Li et al. 2023).

It is important to note that weather feature time series are predicted in advance during both the training and estimation phases of UBEM surrogate modelling. Thus, both the features before (i.e. the past) and after (i.e. the future) the target time step are available when estimating the target and could therefore be utilized to train the surrogate model. Since correlations are bidirectional, not only past features, but also future features, are correlated with the target. Using both 'past' and 'future' dynamic features is thus potentially beneficial for model training. However, the vast majority of previous studies have only employed dynamic features from the 'past' and

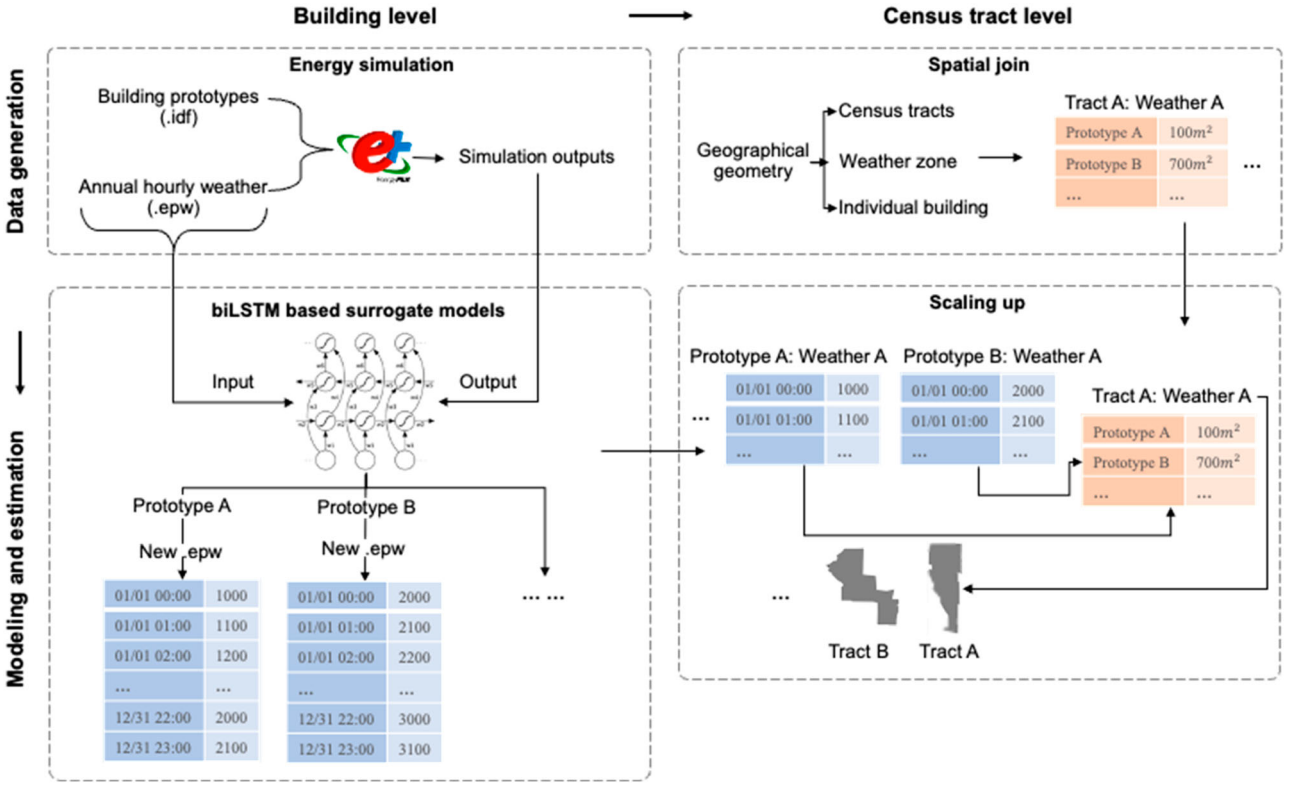

Figure 1. Proposed approach for UBEM surrogate modelling.

neglected (Ohta, Sasakawa, and Sato 2020; Pinto, Deltetto, and Capozzoli 2021; Li et al. 2022; Li et al. 2023) or underestimated (Cohen et al. 2021) the potential benefits of features in the 'future.'

In response to the shortcomings of previous RNN-based UBEM surrogate modelling methods, a Long Short-Term Memory (LSTM) model with bidirectional RNN architecture was used for surrogate modelling-based UBEM in this study. First, the bidirectional LSTM (biLSTM) model took the features at both the 'past' and 'future' of the target as inputs and estimated the targets of interest, including electricity consumption and heat emissions, at the building prototype level. The biLSTM was supported to outperform the baselines in metrics, including the basic LSTM. As the second step, the prototype-level estimations for other microclimates in the UBEM were aggregated to a higher geographical scale, using the building metadata of the studied city. This study is expected to provide a less computationally expensive approach to simulation based UBEMs in the context of urban microclimates and climate change. It also highlights the advantage of bidirectional RNN architecture in UBEM surrogate modelling.

2. Method

Figure 1 demonstrates the workflow of the proposed approach. First, the annual hourly energy consumption and heat emission of building prototypes in the studied area were simulated in EnergyPlus. The simulated energy consumption and heat emission were the targets of the surrogate model training, and the weather variables were the features. The data were fed into a bidirectional LSTM model, to construct separate surrogate models for each prototype that can estimate the annual hourly building energy and heat emission in out-of-sample microclimates. Estimations at the prototype level may not provide sufficient urban-scale insights. Therefore, as a second step, the prototype level estimations were scaled up to the census tract resolution, a common spatial resolution in many urban-scale datasets. To obtain the census tract total energy and heat emissions, the building prototype level energy and heat emission per floor area was multiplied by the total building floor area in each census tract and weather grid. This section elaborates on the components of the proposed approach.

2.1. Simulation and datasets

The first step of the study is to generate training data for surrogate models by conducting simulations for UBEM. In building energy modelling, individual buildings should be characterized in depth, while for UBEM, building characteristics are usually simplified to prototypes. Almost all

existing UBEM tools recommend the use of pre-defined building prototypes that are categorized by construction year, usage (e.g. residential, office, commercial), and building type (tower, detached residential, liner, etc.) (Ferrando et al. 2020). Each prototype can represent a category of buildings, and the information is stored in EnergyPlus IDF files. The other major data source is microclimates. The microclimate is a group of weather features with an hourly resolution, including hourly longwave and shortwave flux at the ground surface, surface pressure, relative humidity, outdoor air dry bulb temperature, and other information. The geographical scope of a microclimate depends on the use case, but it is usually cells with side lengths of hundreds to thousands of metres. Each microclimate is represented by an EnergyPlus Weather File in the EPW format in this study. With B prototypes and D microclimates, there are no more than $B \times D$ prototype-microclimate pairs, as a microclimate zone does not necessarily have all prototypes. Then, the IDF-EPW pairs should be input into EnergyPlus for simulating the hourly building performance of a year. Note that in the real use case, only a portion of the IDF-EPW pairs is required by the simulation to generate surrogate modelling training data, and most of the simulations should be omitted. The target building performance could include electricity consumption, natural gas consumption, or building heat emissions to the surrounding outdoor environment.

In this study, surrogate models were built separately for different building prototypes. Thus, the simulation results of IDF-EPW pairs were grouped by prototypes, and in each group, there were multiple simulations of P microclimates but with the same prototype. Each simulation had Q sequences of weather features and a sequence of target values. Each sequence contained hourly values and had a length of T. To generate samples $(\boldsymbol{X_n}, y_n)(n = 1, 2, \ldots, N)$ for surrogate model training, a sliding window with length L moving with a stride of one was applied on weather features sequences to make feature matrices $\boldsymbol{X_n} \in \mathbb{R}^{L \times Q}$. $y_n \in \mathbb{R}$ was the value of the target at one time step after the last time step in $\boldsymbol{X_n}$. Following this computation method, the simulation data of different microclimates were processed separately, and the number of samples N were $P(T - L + 1)$ in each prototype group.

Weather features under microclimates are strong features for estimating building performance, they can also be regarded as variables that vary close to the TMY (Typical Meteorological Year) weather. Therefore, building performance under TMY is also a highly correlated variable with building performance under microclimates. For this reason, annual hourly building performance in TMY was also used as a surrogate modelling feature. Building prototype IDF files and the TMY EPW file were sent to EnergyPlus for simulation, and it produces sequences of hourly target values of length T. Adding this additional feature, the updated $\boldsymbol{X_n}$ belongs to $\mathbb{R}^{L \times (Q+1)}$. Note that this research practice results in additional simulations beyond those required by UBEM. However, considering that TMY is stable in the long term and is the same for all building prototypes, the additional simulation cost is very limited. Therefore, the energy simulation under TMY could be included in the training data generation process for surrogate modelling.

2.2. Bidirectional LSTM model

The samples generated in Section 2.1 were used to build the surrogate model. The surrogate model used in this study was built on the structure of RNN. Given input samples $(\boldsymbol{X_n}, y_n)(n = 1, 2, \ldots, N)$, the goal of the surrogate modelling is to find a function $f_\theta : \boldsymbol{X_n} \rightarrow y_n$, where θ is the set of parameters that map the features matrix $\boldsymbol{X_n} = [\boldsymbol{x_{k-L+1}}, \ldots, \boldsymbol{x_k}](n = 1, \ldots, N, k = L, \ldots, T-1, \boldsymbol{x} \in \mathbb{R}^{Q+1})$ to the building performance y_n at time $k + 1$, by iterating the equations below from time $t = k - L + 1$ to $t = k$.

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{1}$$

where $\mathcal{H}$ is the hidden layer function, W denotes the weight matrices, and b denotes the bias vector. h_k is fed into a dense layer to produce the predicted $\hat{y}_n$.

$\mathcal{H}$ is a simple activation function in the standard RNN, but it can be replaced with other functions to form more complex RNNs, including LSTM and the Gated Recurrent Unit model. In the case of LSTM, $\mathcal{H}$ is implemented using the following composite function:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{2}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{3}$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{4}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{5}$$

$$h_t = o_t tanh(c_t) \tag{6}$$

Where σ is the sigmoid function. i, f, o, c are the input gate, the forget gate, the output gate, and the cell vector, respectively, with the same size of the hidden states.

Although supported as effective in many cases by previous studies, standard RNNs and LSTMs only take the previous hidden state as input. However, in UBEM, both the previous and subsequent weather features are known during the training and testing. A bidirectional structure could utilize both the previous and subsequent information and obtain a higher accuracy. Thus, the model training in this study is to find a function $F_\Theta : \boldsymbol{X_n} \rightarrow y_n$, where

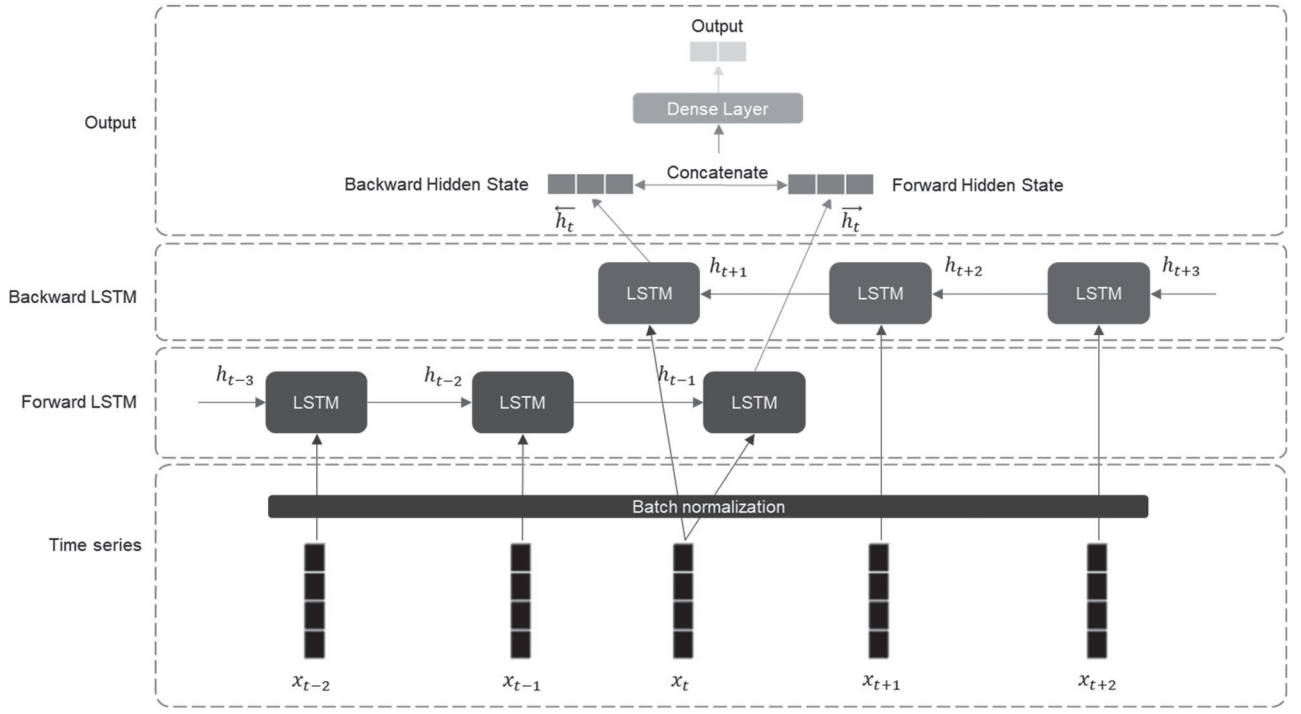

Figure 2. Architecture of the adapted bidirectional LSTM. h_t is the hidden state of the RNN cell at time t. x_t is the feature vector at time t. $\overrightarrow{h_t}$ is the hidden state of the RNN cell in the forward direction at time t, and $\overleftarrow{h_t}$ is the hidden state of the RNN cell in the backward direction at time t.

Θ is the set of parameters that map the features at time t and the hidden state at time $t-1$ and time $t+1$ to the building performance at time t, by iterating the equations below.

$$\overrightarrow{h_t} = \mathcal{H}(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}) \tag{7}$$

$$\overleftarrow{h_t} = \mathcal{H}\left(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t+1} + b_{\overleftarrow{h}}\right) \tag{8}$$

$$h_t = concatenate(\overrightarrow{h_t}, \overleftarrow{h_t}) \tag{9}$$

where $\overrightarrow{h_t}$ is the forward (i.e. from previous to subsequent time steps) hidden state and $\overleftarrow{h_t}$ is the backward (i.e. from subsequent to previous time steps) hidden state. The $\mathcal{H}$ used in this study was implemented using Equations (2–6).

Figure 2 shows the details of the biLSTM used in this study. The model follows the way a standard LSTM inputs features by sequentially feeding the features at each time step for all samples. Unlike a standard LSTM, in biLSTM the features in a sample are fed into two separate LSTMs in both forward and backward directions, after passing through a batch normalization layer.

The model computes the features and hidden states according to Equations (7–8). For a sample of a given length, depending on the length of previous time steps and subsequent time steps, the hidden states $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ of a particular time step could be concatenated and fed into a dense layer to produce the output of the model. For example, for a sample five time steps long, the $\overrightarrow{h_3}$ and $\overleftarrow{h_3}$ could be used for the output, with the assumption that the 'past' and 'future' are equally important. The loss function used in this study is the mean squared error.

2.3. Aggregation to the census tract level

After the simulation and the model training, the biLSTM is able to estimate the performance of a specific prototype in unseen microclimates. This estimation is at the prototype level, which cannot demonstrate urban building energy performance. To aggregate the estimates to the urban scale, three types of information in the studied region need to be collected: (1) urban sub-region (i.e. census tract in this study) geometries with geographical information, (2) microclimate geometries with geographical information, and (3) the metadata of all individual buildings (i.e. location, prototype, and building floor area) in the studied region. First, the microclimates of sub-regions are determined by matching the geographical location of sub-region geometries and microclimate zone geometries. Second, the individual buildings are matched with all sub-regions using the geographical locations. The building floor areas of the prototypes in each sub-region are obtained by grouping the individual buildings

using their prototypes. As the final step, the target estimations for prototypes are normalized using the building areas of the prototypes. The building areas of the prototypes in sub-regions are multiplied by the normalized estimates (e.g. electricity consumption per square metre) to derive the hourly estimates of performance for the studied region and with the resolution of the sub-region.

2.4. Evaluation metrics

Both relative and absolute metrics are used to evaluate the accuracy of the surrogate models. The absolute metrics include root mean squared error (RMSE) and mean absolute error (MAE), and the relative metrics used are normalized RMSE (nRMSE) and normalized MAE (nMAE). The metrics are defined below.

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N} (y_n - \hat{y}_n)^2}{N}} \tag{10}$$

$$MAE = \frac{\sum_{n=1}^{N} |y_n - \hat{y}_n|}{N} \tag{11}$$

$$nRMSE = \frac{RMSE}{\frac{1}{N}\sum_{n=1}^{N} |y_n|} \tag{12}$$

$$nMAE = \frac{MAE}{\frac{1}{N}\sum_{n=1}^{N} |y_n|} \tag{13}$$

where $\hat{y}_n$ is the estimation of the n^{th} sample, y_n is the ground truth of the n^{th} sample, and $n(n = 1, \ldots, N)$ is the index of the n^{th} sample. RMSE and MAE are always non-negative, and a low value indicates better accuracy. nRMSE and nMAE are non-negative, where a lower value indicates better accuracy and 0 indicates perfect estimation.

The commonly used mean absolute percentage error (MAPE) was replaced with nMAE in this study, as building performance targets are likely to be zero in some cases (e.g. zero cooling load in winter), and that is beyond the definition of MAPE. But even if some of the targets are zero, nMAE is still in the defined domain as long as the average of samples is non-zero. It should also be noted that the denominator of the nMAE, as well as nRMSE, is the average absolute value of the ground truth. This operation is conducted to avoid underestimation of the variance in the target when there are negative target values, which could exist in building heat emissions.

3. Case study

Los Angeles (LA) County is used as the testbed for the surrogate modelling-based UBEM. The County of Los Angeles aims to reduce greenhouse gases (GHGs) by 50% by 2035 and achieve carbon neutrality by 2045 (County of Los Angeles 2022), through decarbonization, efficiency improvement, and water conservation in the buildings sector, and by employing integrated strategies in other sectors including transportation, energy systems, and agriculture. In the City of Los Angeles (LA), the Los Angeles 100% Renewable Energy Study (LA100) outlines pathways for the city to achieve the goal of a 100% electricity power supply by 2045 (Cochran et al. 2021). With these ambitious goals, it is crucial to understand the energy consumption and heat emissions of the LA building stock under the current climate and future climate change scenarios. On the other hand, due to the city's large size and its long history, there are a large number of building prototypes in LA County. LA County also has very different microclimate zones because of its high-density urban centres and geographic characteristic of being by both the sea and the mountains. Therefore, performing a UEBS for LA County would incur a large number of separate simulation models and long computational time. In this study, surrogate models were built for LA County to estimate the annual hourly energy consumption and heat emissions.

Currently data sources like Energy Atlas (https://www.energyatlas.ucla.edu/) provide measured energy consumption data aggregated at the city and neighbourhood level. However, the temporal resolution is very low; only reporting an annual total for each city or neighbourhood and each building type. Furthermore, the building type classification is rather coarse, with all commercial buildings in one category. In this study, an 8760 energy and anthropogenic heat dataset was produced for all of LA County at three spatial resolutions (Xu et al. 2022), using EnergyPlus simulations of 54 prototype building models and 62 local climate data points from the Weather Research and Forecasting (WRF) TGW dataset (Jones et al. 2023). The surrogate models in this study were trained using this dataset.

The 54 prototypes include 18 building types (i.e. multifamily houses, single-family houses, heavy and light manufacturing facilities, nursing home, hospital, small hotel, large hotel, small office, medium office, large office, full-service restaurant, stand-alone retail, midrise apartment, primary school, warehouse, college, supermarket, and religious worship buildings), with three vintages: pre-1980, 2004, and 2013. The prototypes used in this study are relatively extensive compared to many previous UBEM studies that used 13–30 building prototypes (Zhou et al. 2012; Zheng and Weng 2018; Luo et al. 2020; Chen et al. 2022). On the other hand, microclimates were simulated for 2018. Each microclimate is represented by a group of weather features, including hourly downward longwave flux at the ground surface (GLW), downward

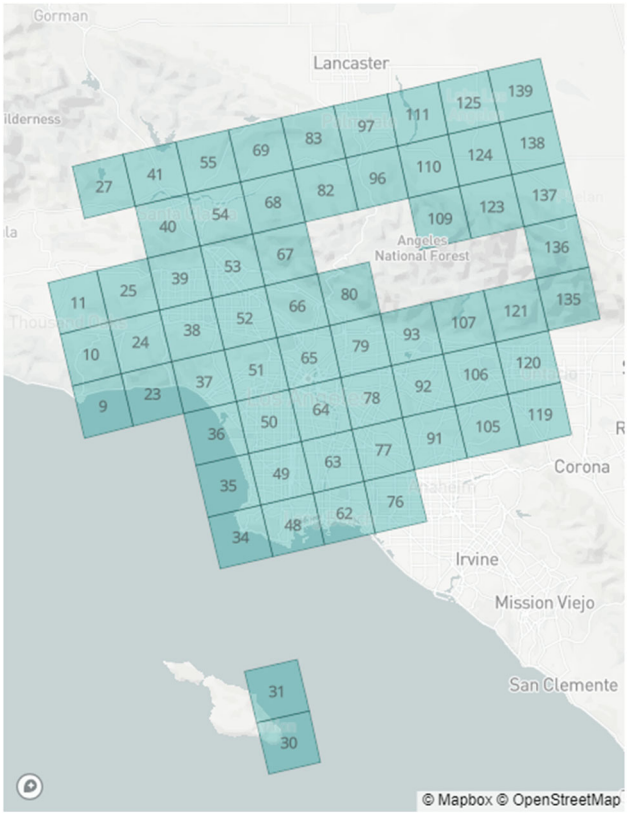

Figure 3. Microclimate zones with index numbers in the case study.

shortwave flux at the ground surface (SWDOWN), surface pressure (PSFC), water vapour mixing ratio at 2 metres (Q2), relative humidity (RH), outdoor air dry bulb temperature at 2 metres (T2), wind direction (WINDD), and wind speed (WINDS). The microclimate zones are 12 km × 12 km cells, and LA County is discretized into 62 cells (shown in Figure 3).

Matching 62 microclimates and 54 prototypes according to the spatial distribution of the prototypes resulted in 1,667 prototype-microclimate pairs. 1,667 EnergyPlus simulations were built and run by using different pairs of building prototypes and local climates, and each model simulated one year of building performance. The simulations were finished on a 2.8 gigahertz (GHz) Quad-Core Intel Core i7 central processing unit in approximately 45 hours. In this study, electricity consumption (ELEC), and anthropogenic heat components including the surface convection heat loss (E-SURF), exhaust air heat loss (E-EXH), and HVAC system heat rejection energy (E-REJ) were selected as the targets. The target values under TMY (Typical Meteorological Year) were obtained by running EnergyPlus simulations for each prototype with the TMY3 at LA airport.

The objective of the surrogate modelling is to estimate the outputs of UBEM. Therefore, before surrogate modelling, the accuracy of UBEM itself for estimating the real-world energy modelling in LA County was also validated. In the authors' precursor study to this study (Xu et al. 2024), the electricity consumption of LA County by UBEM was compared with that in Energy Atlas 2016 (UCLA California Center for Sustainable Communities (CCSC) 2020). The percentage estimation error regarding the electricity consumption of the whole LA County in Energy Atlas is 6.75%, and the average of percentage estimation error regarding the Energy Atlas electricity consumption on the neighbourhood level is 10%.

For the surrogate modelling, the building performance under TMY, microclimate weather features, and target values were time series with a length of 8,760 ($T = 8,760$). The length of the sliding window for generating samples was 25 ($L = 25$). For each sample (X, y), x belongs to $\mathbb{R}^{25\times 9}$ and y belongs to $\mathbb{R}$. 5% – 25% of the prototype-microclimate pairs in 2018 were randomly selected from each prototype for the model training. Note that at least two microclimates were selected for each prototype. In the model training, the learning rate was set as 0.001, and the size of the hidden states was 512. To avoid redundant hyperparameter searches, 15% of the training set of one prototype's (e.g. multi-family houses) surrogate model was split into a validation set. The validation set was used to determine the hyperparameters (i.e. learning rate, batch size, and the number of hidden layer units) based on the grid search method. The training was implemented on an NVIDIA V100 graphics processing unit (GPU) and can be finished within 1.5 hours in most cases. 75% – 95% of the prototype-microclimate pairs and their samples were used for the test. Note that in this study, a small portion of the data was used for training, and most of the data was used for testing. The reason for this is that the research objective is to reduce the number of simulations required for UBEM and maintain a satisfactory UBEM accuracy by means of surrogate modelling. The training data indicates the required simulation computation amount, and the test data indicates the reduced computation workload and is used to evaluate the accuracy.

The geometries with geographical information of the census tracts and the metadata of all individual buildings (i.e. location, prototype, and building area) were collected from the LA County open data portal and are also available in the work by Xu et al. (Xu et al. 2022). The aggregation followed the method in Section 2.3, which outputs the estimations on the census tract level for LA County. The datasets and codes of this study are available at the GitHub repository github.com/pppxiyu/SurUBEM. The reproducible procedures of preparing the datasets are available at the GitHub repository github.com/IMMM-SFA/xu_etal_2022_sdata.

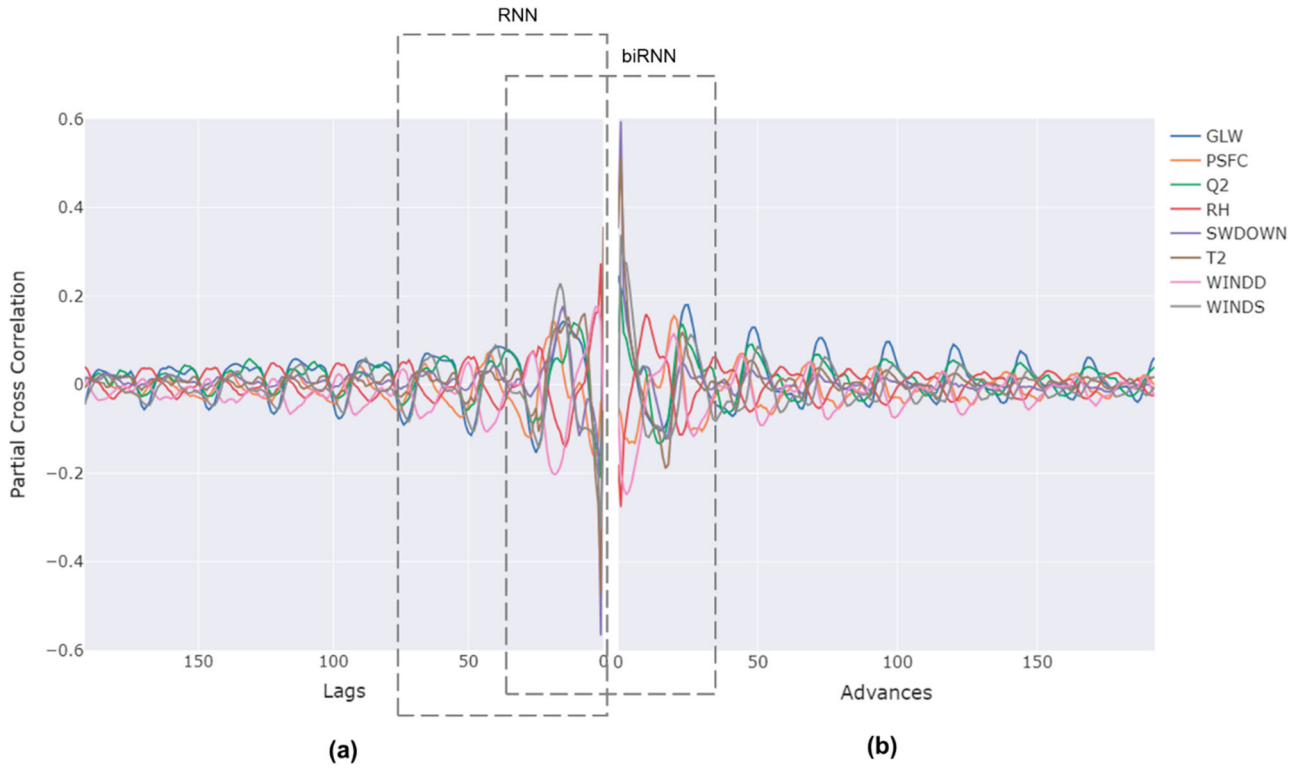

Figure 4. Partial correlation plots between the electricity consumption and the weather features. The duration of one lag or advance is one hour. The electricity consumption is from multi-family buildings with pre-1980 vintage at Microclimate 51. Weather features time series are from Microclimate 51: (a) partial correlations with weather features in previous time steps. (b) partial correlations with weather features in subsequent time steps.

4. Results

4.1. Modelling and accuracy

In contrast to modelling using a standard RNN, this study employed a bidirectional RNN to estimate the target values. To demonstrate the advantages of the biLSTM in utilizing the data, the multi-family building model with pre-1980 vintage at Microclimate 51 (MultiFamily-pre-1980-51) was analyzed as an example, as MultiFamily-pre-1980-51 has the largest building floor area among all the prototype-microclimate pairs. In the analysis, the electricity consumption and microclimate feature time series were used to compute the partial correlations (P-Corr) (Stark, Drori, and Abeles 2006). P-Corr describes the associations between the electricity consumption and the time-lagged weather features, without the effect of previous time lags (Figure 4(a)). By simply reversing the time series, the same calculation will produce the P-Corr between the electricity consumption and the time-advanced weather features (Figure 4(b)). A standard RNN will use time-lagged weather features within a predefined sequence length. As shown in Figure 4(a), the P-Corr within the RNN time window decreases as the time away from the target grows. Using a time window of the same length, biRNN will use the features with higher P-Corr in the RNN time window and also use the feature with high P – Corr in the 'future' time steps. In short, biRNN utilizes the more correlated features in the data, which is favourable for the model training and higher accuracy.

In addition to microclimate features, the prototype's performance under TMY was also used as one of the features for the surrogate models. Electricity consumption data for the three prototypes was used to illustrate the benefit of this additional feature. Figure 5 shows the electricity consumption of heavy manufacturing facilities, large hotels, and single-family houses in February. It is clear that heavy manufacturing facilities consume a great deal of electricity and show a pattern of high electricity consumption on weekdays and low electricity consumption on weekends. This pattern is very regular and could be less related to local weather than to facility use. In contrast, the electricity consumption pattern of single-family houses is visually irregular and can potentially be modelled by weather features. The electricity consumption pattern for large hotels falls somewhere in between, showing stable seasonality and a declining trend, but still could be related to microclimates. Faced with the issue of

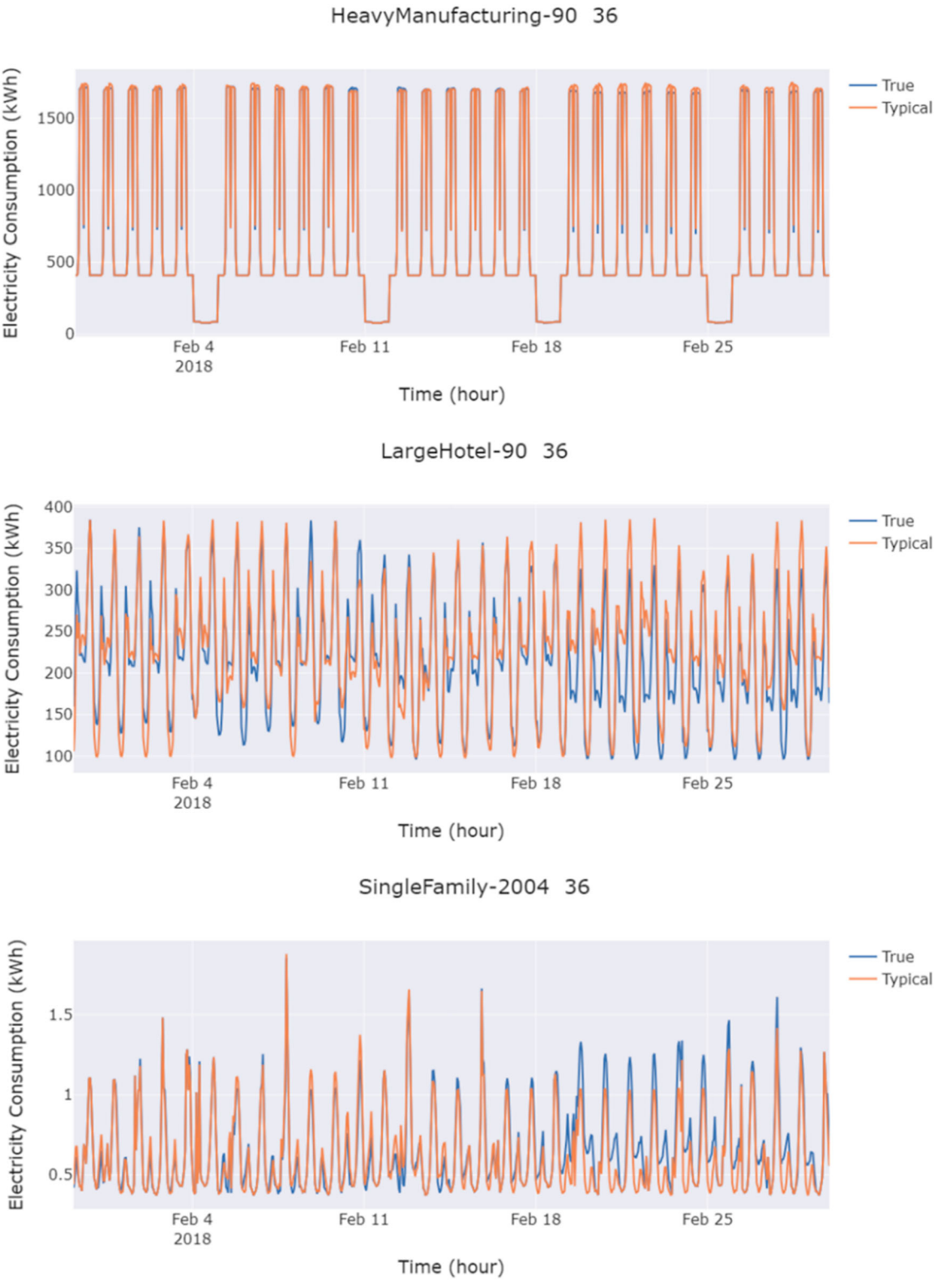

Figure 5. Hourly electricity consumption of heavy manufacturing facilities, large hotels, and single-family houses in February.

varying microclimate-target correlation, the prototype's performance under TMY can be used as a 'best guess' for the electricity consumption in the corresponding microclimate, which can maintain its correlation with the target even when the microclimate influence is weaker.

Table 1 summarizes the accuracy and computation time of the UBEM surrogate models. The metrics are at the census tract level, and the computation time includes both data preprocessing, model training, and estimating. The training of surrogate models estimating ELEC,

Table 1. Metrics of target (i.e. ELEC, E-SURF, E-EXH, and E-REJ) estimations at the census tract level.

	Data used in training (%)	Time (min)	RMSE (kWh)	MAE (kWh)P	nRMSE	nMAE
ELEC	5	71	279.59	108.75	0.1421	0.0553
	10	77	250.35	81.49	0.1440	0.0469
	15	87	122.78	57.80	0.0936	0.0441
E-SURF	5	75	3924.08	2100.85	0.2013	0.1078
	10	80	2963.46	1257.23	0.1545	0.0655
	15	90	2888.69	1211.64	0.1440	0.0604
E-EXH	5	57	135.71	84.52	0.2260	0.1407
	10	62	86.13	50.44	0.1656	0.0970
	15	69	77.41	41.44	0.1512	0.0810
E-REJ	10	79	337.31	163.94	0.2496	0.1213
	15	92	311.79	146.74	0.2413	0.1136
	20	108	274.45	122.54	0.1978	0.0883
	25	113	261.45	116.10	0.1976	0.0877

E-SURF, and E-EXH can be completed in up to 90 minutes using 5% to 15% of the simulation results. Most of the nMAE of these estimations are within 10%, and most of the nRMSE are within 15%. E-REJ is more difficult to estimate, and its nMAE can only be reduced to less than 10% when using at least 20% of the simulation results. It should be noted that the error rates could be lower if more simulation results were used to train the surrogate models. However, to balance accuracy and computational time, 10% of the data are used for training and all following evaluations. Note that the metrics in Table 1 measure the accuracy of the surrogate models with respect to the simulation outputs, not the accuracy regarding the real measurements (i.e. the ground truth). In prior research, the normalized RMSE of the energy demand UBEM with respect to the ground truth ranges from 5% – 40% for overall evaluation (Oraiopoulos and Howard 2022). As mentioned in the Case Study, the error rate of UBEM regarding ground truth in this study is 6.75%. Considering the nRMSE of surrogate modelling is 14.40% with 10% training set. The overall error rate of the surrogate modelling-based UBEM could be approximated as 21%. Comparing to the metric range in prior studies, the accuracy of the proposed method in the present study is within the lower half of the range.

The accuracy of the biLSTM surrogate model is compared with those of the baselines, including naive estimation, linear regression, multilayer perceptron, and standard LSTM (shown in Table 2). The naive estimation calculates the mean value of the 10% selected simulation results and uses the mean value as the estimation of targets in unseen microclimates. This is a good indicator of the sensitivity of the target to microclimates. The electricity consumption is not very microclimate-sensitive, whereas the nMAE for heat emission targets is much larger, indicating a higher microclimate sensitivity. For low microclimate-sensitive targets, the advantage of the surrogate modelling is not obvious. But as the sensitivity increases, the difference between the nMAE of surrogate models and the naive estimation is growing, and the surrogate modelling easily outperforms the naive estimation. On the other hand, biLSTM outperforms the standard LSTM model in all four metrics and targets. It echoes the analysis of Figure 4 that the bidirectional architecture is expected to produce better accuracy as the higher correlation between features and targets.

Table 2. Metrics of prediction target (i.e. ELEC, E-SURF, E-EXFIL, and E-REJ) estimations at the census tract level by biLSTM and the baselines. The metrics are from the surrogate models trained with 10% of the simulation results. Linear regression, multilayer perceptron, and standard LSTM use the past 24-hour lags data as features, while biLSTM uses data of the past 12-hour lags and subsequent 12-hour advances as features.

		ELEC	E-SURF	E-EXH	E-REJ
Naive	RMSE (kWh)	209.53	4035.47	196.02	671.90
	MAE (kWh)	87.05	2311.86	109.43	340.34
	nRMSE	0.1205	0.2106	0.3768	0.4977
	nMAE	0.0500	0.1206	0.2103	0.2521
Linear regression	RMSE (kWh)	757.2	7255.17	258.27	1237.36
	MAE (kWh)	270.34	3614.21	164.36	753.54
	nRMSE	0.4356	0.3782	0.4968	0.9154
	nMAE	0.1555	0.1884	0.3162	0.5575
Multilayer perception	RMSE (kWh)	3873.80	13875.34	12644.90	16309.52
	MAE (kWh)	1978.86	8441.37	7433.49	9824.85
	nRMSE	2.2285	0.7232	24.3269	12.0671
	nMAE	1.1384	0.4400	14.3009	7.2692
LSTM	RMSE (kWh)	665.05	5353.69	188.50	752.51
	MAE (kWh)	180.32	2755.81	114.03	346.82
	nRMSE	0.3825	0.2790	0.3626	0.5567
	nMAE	0.1037	0.1436	0.2194	0.2566
biLSTM	RMSE (kWh)	181.23	2963.46	86.13	337.31
	MAE (kWh)	54.79	1257.23	50.44	163.94
	nRMSE	0.1031	0.1545	0.1656	0.2496
	nMAE	0.0311	0.0655	0.0970	0.1213

Table 3 provides a more detailed evaluation of biLSTM-based surrogate modelling at the prototype level. On the one hand, the nMAE of the biLSTM varies across prototypes. It is within the range of 1.5% (small hotel with 2004 vintage) and 9.61% (hospital with 2013 vintage). Single-family houses and multi-family houses, which have the largest floor area in LA among all prototypes, have an nMAE between 5.4% and 8.24%. Their corresponding accuracy is closer to the nMAE at the census tract level after the aggregation step (i.e. 4.69%). As can be seen in Table 3, although some of the prototypes have a lower estimation accuracy, the nMAE is still below 10%. On the other hand, biLSTM outperforms other neural network models (i.e. multilayer perceptron and LSTM) on the vast majority of prototypes. For example, MLP fails on the surrogate modelling of single-family houses with 2013 vintage, but biLSTM maintains a low error rate.

4.2. Error distributions

As shown in Section 4.1, the trained biLSTM surrogate models have nMAE lower than 10% in most cases.

Table 3. Prototype-level nMAE for multilayer perceptron, LSTM, and biLSTM. The metrics are from the surrogate models trained with 10% of the simulation results.

Building type	Vintage	MLP	LSTM	biLSTM
College	pre 1980	0.3406	0.0523	0.0119
	2004	0.2661	0.0518	0.0264
	2013	0.2595	0.0409	0.0173
Full Service Restaurant	pre 1980	0.1882	0.0806	0.0194
	2004	0.4078	0.0713	0.0512
	2013	0.3605	0.1031	0.0464
Heavy Manufacturing	pre 1980	0.0773	0.0679	0.0126
	2004	0.0784	0.6990	0.0314
	2013	0.2044	0.3079	0.0181
Hospital	pre 1980	0.0658	0.0288	0.0208
	2004	0.1064	0.1093	0.0195
	2013	0.1033	0.0640	0.0961
Large Hotel	pre 1980	0.1200	0.2741	0.0576
	2004	0.1384	0.2772	0.0435
	2013	0.1335	0.2786	0.0324
Large Office	pre 1980	0.1334	0.5805	0.0910
	2004	0.0747	0.1807	0.0237
	2013	0.0932	0.3060	0.0284
Light Manufacturing	pre 1980	0.2142	0.1946	0.0138
	2004	0.2169	0.0438	0.0122
	2013	0.2340	0.0565	0.0354
Medium Office	pre 1980	0.2550	0.0837	0.0499
	2004	0.3293	0.0760	0.0480
	2013	0.3375	0.0928	0.0758
Midrise Apartment	pre 1980	0.3888	0.0801	0.0394
	2004	0.2103	0.0949	0.0238
	2013	0.2349	0.0840	0.0277
Multi Family House	pre 1980	1.3097	0.0942	0.0623
	2004	1.5002	0.1003	0.0562
	2013	1.7590	0.0868	0.0540
Nursing Home	N/A	0.3150	0.0906	0.0479
Primary School	pre 1980	0.3066	0.0688	0.0281
	2004	0.2400	0.0575	0.0394
	2013	0.3118	0.0856	0.0564
Religious	pre 1980	0.3301	0.0463	0.0247
	post 1980	0.5473	0.0459	0.0218
Retail Standalone	pre 1980	0.4231	0.0470	0.0265
	2004	0.3177	0.0804	0.0477
	2013	0.4908	0.0700	0.0464
Single Family House	pre 1980	2.2233	0.1359	0.0699
	2004	8.8825	0.1620	0.0824
	2013	22.1449	0.1180	0.0572
Small Hotel	pre 1980	0.1604	0.0788	0.0376
	2004	0.1147	0.0399	0.0150
	2013	0.2829	0.0561	0.0363
Small Office	pre 1980	0.6976	0.0474	0.0186
	2004	1.3106	0.0602	0.0378
	2013	1.0747	0.0671	0.0314
Supermarket	pre 1980	0.1219	0.0653	0.0320
	2004	0.1074	0.0960	0.0906
	2013	0.1468	0.0754	0.0345
Warehouse	pre 1980	0.7478	0.0305	0.0213
	2004	0.4526	0.0568	0.0290
	2013	0.8191	0.0946	0.0431

However, the distribution of the error rate needs to be further investigated to demonstrate the potential bias of UBEM surrogate models in this study. Figure 6 shows the distribution of nMAE for different building types and targets. In terms of electricity consumption, residential buildings have significantly higher error rates than manufacturing facilities and colleges, which have regular operations schedules. The reason for that is the electricity consumption pattern of some non-residential buildings (e.g. manufacturing facilities) is more regular, which is more easily captured by the surrogate model, especially when one of the features is the target values of the prototype under TMY. Figure 6(b) shows the distribution of nMAE for estimating E-SURF. Most of the building types with low nMAE are small buildings, such as residential buildings, retail buildings, small offices, etc., while large buildings, such as manufacturing facilities, large office buildings, and large hotels, have higher error rates. The reason for this may be that larger buildings with a greater surface area have higher surface heat emissions, and the surrogate model performs better at estimating low heat emissions.

For illustrating the spatial distribution of the estimation errors, census tracts in four typical climate zones were selected (as shown in Figure 7). The nMAE of estimations are comparable in the inland residential area (i.e. the second column in the figure) and the seaside residential area (i.e. the fourth column in the figure), and both are relatively low. This indicates that the trained surrogate models perform well and can handle the difference between seaside and inland microclimates. The nMAE of the estimation for the downtown area census tract (i.e. the first column in the figure) is also low. The surrogate models are therefore considered to be adaptive to the microclimate difference caused by the UHI as well. However, the model's accuracy declines significantly in the mountainous area (i.e. the third column in Figure 7). A similar finding can be seen in Figure 8(a). Among the census tracts randomly selected for the evaluation, the lower left Area 1 (i.e. Portuguese Bend Reserve) has an obviously higher error rate. Areas 2 (i.e. Indian Springs) and 3 (i.e. Los Angeles National Forest) in the upper right section also have higher nMAE; they are all mountainous areas or have large areas of green space. The higher nMAE could be explained by the spatial distribution of the residential building (i.e. single-family and multi-family buildings) area percentage regarding total building area for each census tract. As shown in Figure 6(a), the error rate of estimating ELEC is the highest for residential building prototypes. Therefore, the error rate of estimating ELEC for a census tract is likely to be high if the census tract has a high percentage of residential building areas (shown in Figure 8(b)). As evidence, Figure 8(c – d) show that the spatial distribution of the residential building area percentage and the nMAE share a similar pattern. In particular, the mountainous or green space area (i.e. areas ① – ③ in Figure 8(d)) has a higher percentage of residential buildings, and its ELEC estimation error is also higher (shown in Figure 8(c)).

In addition to the spatial differences in the accuracy of the surrogate models, it is also essential to evaluate its variations in the temporal dimension. Figure 9 illustrates

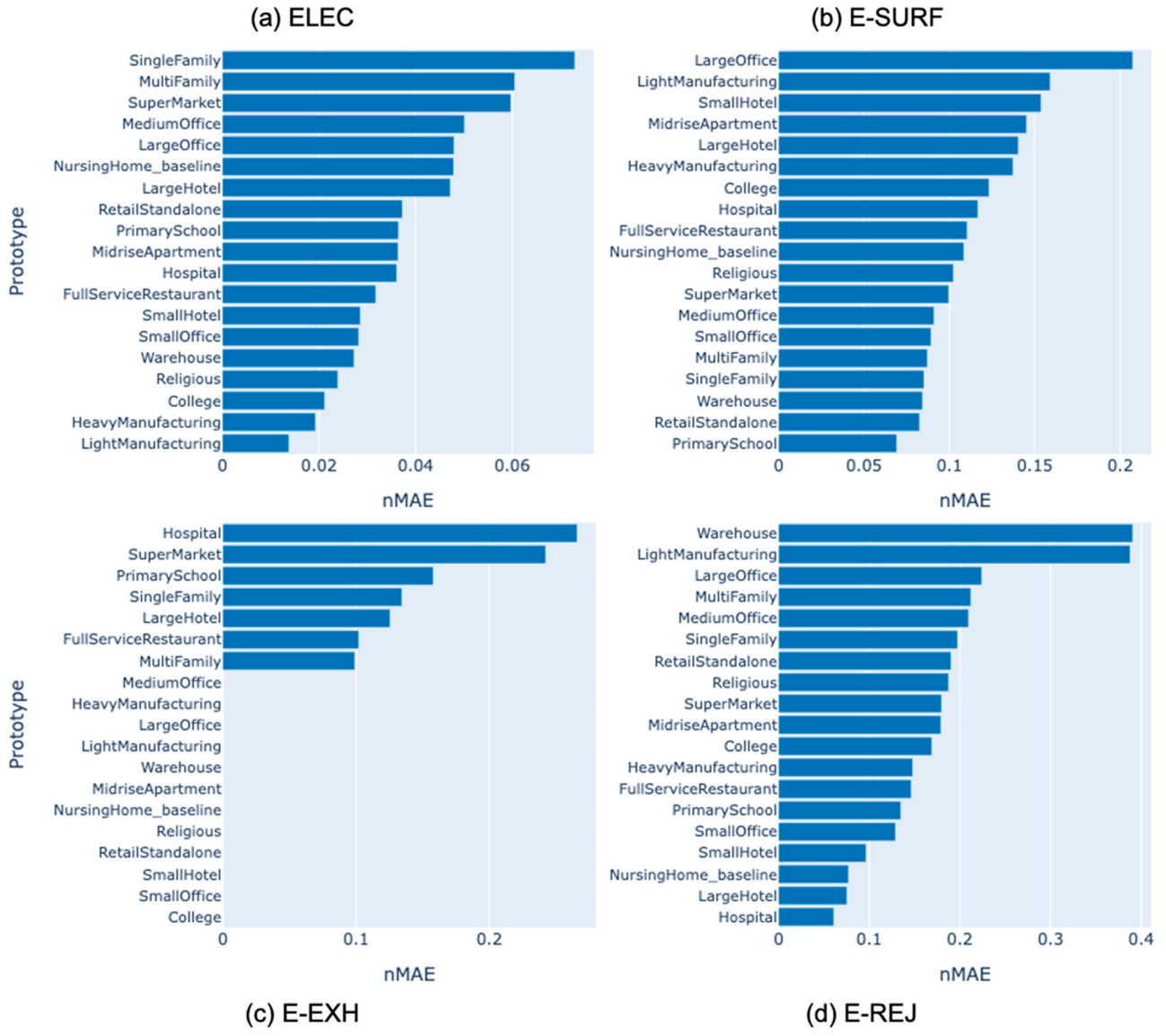

Figure 6. nMAE of the target (i.e. ELEC, E-SURF, E-EXH, and E-REJ) estimations at the prototype level. The experiments used 10% of the simulation data. It should be noted that the prototypes used in this study are not limited to the ones shown in this figure. For example, SingleFamily-2004 (SF2004), SingleFamily-2013 (SF2013), and SingleFamily-pre-1980 (SF1980) were used as single-family house prototypes. The nMAE of single-family houses is the weighted nMAE of SF2004, SF2013, and SF1980 using their building areas.

the average electricity consumption estimation nMAE of census tracts over the selected months, and one month from each season was chosen. As shown in the figure, there is no clear pattern of the change in nMAE from one month to another. However, comparing the nMAE between seasons, it is clear that the accuracy of the estimates is similar for April (i.e. spring) and July (i.e. summer), but the nMAE for October (i.e. fall) increases significantly, with more spikes, compared to April and July. In addition, the overall nMAE for December (i.e. winter) is significantly lower than the other months. Overall, the temporal variation in model accuracy is acceptable, but it should be scrutinized when using the proposed method on other datasets.

Figure 10 compares the distribution of the nMAE over the electricity consumption peak time, valley time, and the rest of the time. Both the peak and valley time were identified using the default algorithm in the SciPy *find_peaks* method and are shown in Figure 10(a) (Virtanen et al. 2020). Unlike MAE or RMSE, which could be much higher at peaks, nMAE has a significantly lower mean value at the peak time of electricity consumption, and its variance is also smaller. There is no obvious difference between the nMAE distribution of valley time and the rest of the time. Furthermore, For the evaluation at energy consumption peak time, the normalized RMSE regarding measured values ranges from about 10% to 35% in De Jaeger's work (De Jaeger et al.

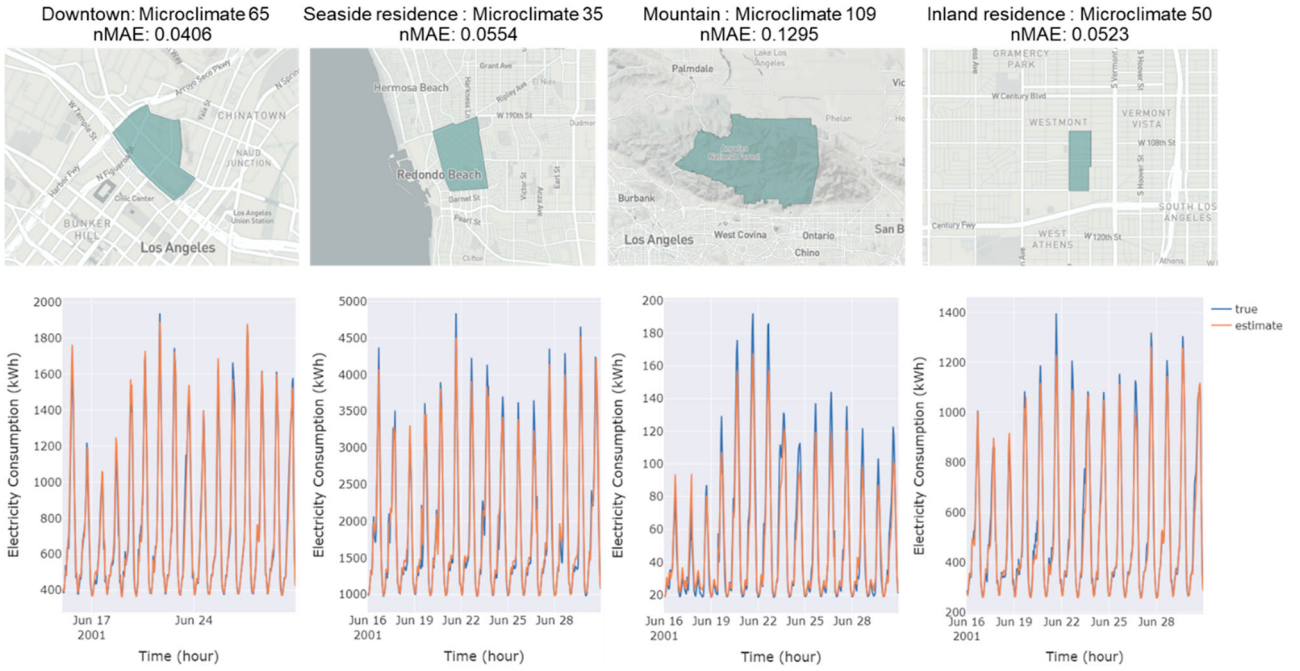

Figure 7. ELEC estimations for June 15 to June 30 in selected census tracts. Census tract 6037207101 was selected for Microclimate 65. Census tract 6037621201 was selected for Microclimate 35. Census tract 6037930301 was selected for Microclimate 109. Census tract 6037600303 was selected for Microclimate 50. The experiment generating the data for this figure used 10% of the simulation data to build the surrogate model.

2020), and ranges from 84% to 129% in Prataviera et al's work (Prataviera et al. 2022) with comparable model configuration. Compared to the range in the prior studies, the 4.98% nRMSE for the surrogate modelling in this study (shown in the caption of Figure 10) is modest. It suggests that the surrogate modelling in this study did not result in obvious extra errors on top of the base errors from UBEM, and even if the peak time accuracy of UBEM is low, the proposed surrogate modelling method could possibly remain a useful tool for reducing simulation time while maintaining the reasonable UBEM accuracy.

5. Discussion and conclusions

5.1. Contributions

In this study, surrogate modelling is used to shorten the computation time of simulation based UBEM with a large number of microclimates. In this novel approach, biLSTM is used to build surrogate models for building prototypes using a small portion of simulation results. The trained surrogate models can estimate the building performance of the prototype in new microclimates. The estimations are aggregated to the census tracts of LA County. The evaluation results show that the proposed approach can reduce computation time with an nMAE lower than 10% in most cases, and biLSTM can outperform baseline surrogate models in accuracy. This study has two major contributions.

First, this study reduces the computation time of UBEM with microclimates and therefore decreases the difficulty in long-term and high-resolution UBEM with microclimates. It makes related analyses and applications more convenient and easier to promote. On the one hand, some long-term urban designs and retrofits need to consider climate and microclimate in future decades. For example, energy retrofits across the entire building stock would require decades, during which the urban climate will change (Streicher et al. 2021; Ma et al. 2023). In the case where the building prototype is unchanged, the surrogate model can utilize the simulation outputs based on a small number of microclimates to estimate the performances of the corresponding building prototype under the remaining microclimates, which reduces the amount of simulation computation for long-term UBEM. In the case where the building prototypes change in the long term, although the surrogate models need to be retrained after each building prototype update, they would remain valid until the next update. Take the electricity consumption UBEM as an example, even if the building prototype is updated annually, the proposed surrogate modelling approach still reduces the simulation effort by 90% with an estimation accuracy higher than 90% (shown in Table 1). Therefore, by using the approach proposed in this study, the UBEM for energy retrofits can

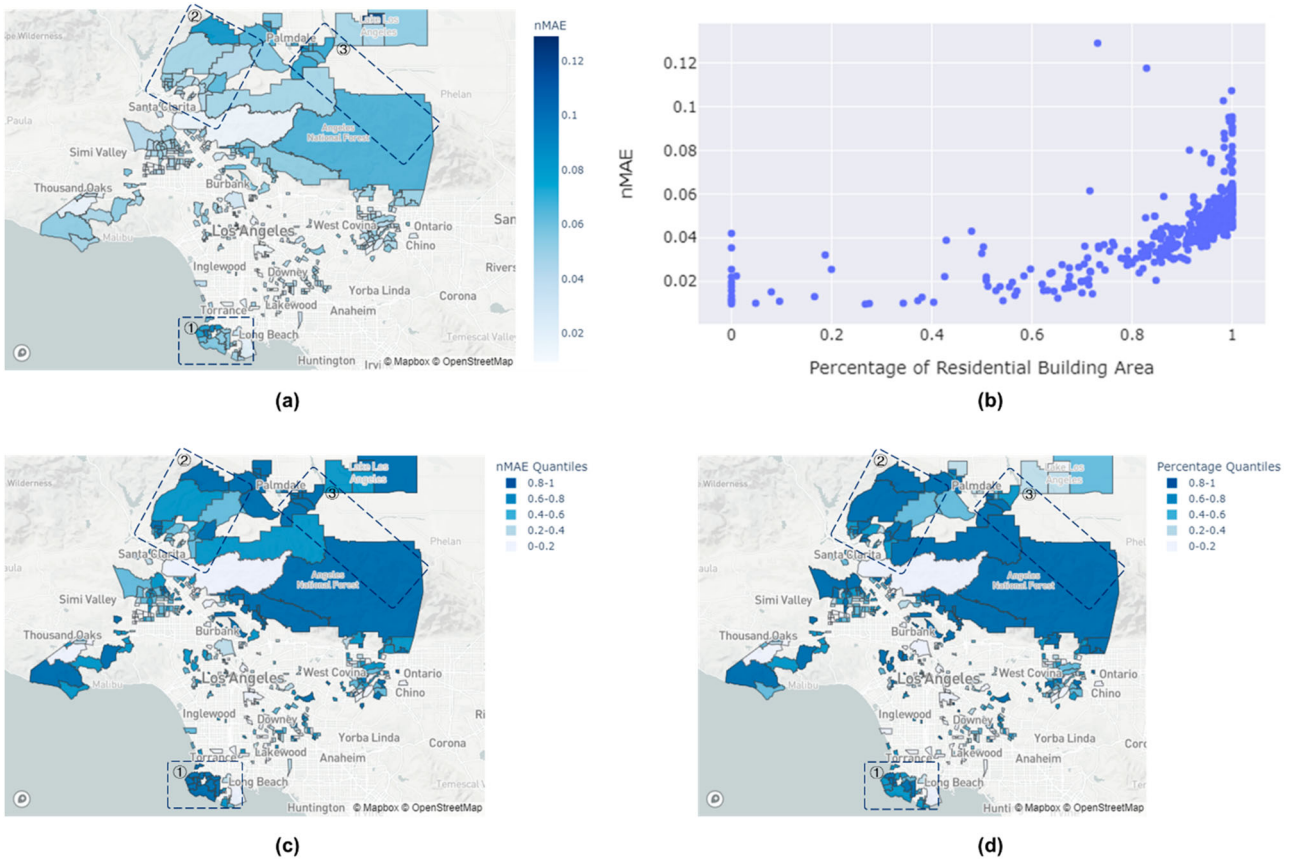

Figure 8. The relationship between the nMAE and the percentages of residential building (i.e. single-family and multi-family buildings) area over total building area. In sub-figures, Area ① is the Portuguese Bend Reserve area. Area ② is the Indian Springs area. Area ③ is the Los Angeles National Forest area. The experiment generating the data for this figure used 10% of the simulation data to build the surrogate model. **(a)** nMAE of ELEC estimations at randomly selected census tracts. **(b)** Scatter plot of the nMAE regarding the percentage of residential building area. Each point represents a census tract. **(c)** nMAE quantiles for each census tract. For example, the 0.8–1 category contains the census tract with an nMAE ranking top 20% among all. Each category has the same amount of census tracts. **(d)** Percentage of residential building area quantiles for each census tract.

be completed in a much shorter time or be completed with a higher temporal resolution in the same amount of time.

On the other hand, in the spatial dimension, some urban designs and retrofits are sensitive to local microclimates. For instance, UHI has a significant impact on building energy consumption, heat emissions, and environmental thermal comfort, and that influences the necessity and effectiveness of implementing some building retrofit strategies, such as solar reflective materials, shading trees around buildings, and high albedo surfaces (Bande et al. 2019). Therefore, urban designs and retrofits rely on a high spatial resolution UBEM. With the proposed approach, design and retrofit strategies can be tested and optimized in a short time, or in the same amount of time but with a higher spatial resolution.

Second, this study demonstrates the advantages of bidirectional LSTM architecture over LSTM in UBEM surrogate modelling. Although a small number of previous studies have used bidirectional RNN architectures in UBEM surrogate modelling, they concluded that standard RNNs can outperform bidirectional RNN architectures (Cohen et al. 2021). This study shows the advantage of biLSTM over LSTM in terms of accuracy and draws the opposite conclusion. This study also highlights the suitability of biLSTM in BES surrogate modelling for applications beyond UBEM, including building design optimization, simulation analysis acceleration, high-resolution building energy modelling, and others (Ohta, Sasakawa, and Sato 2020; Pinto, Deltetto, and Capozzoli 2021; Li et al. 2022; Li et al. 2023). In these applications, if one or more dynamic features of the target can be predicted prior to estimation and used as a 'known future,' biLSTM models should be prioritized over the standard LSTM.

5.2. Limitations and future directions

Despite the promising contributions, this study has limitations. First, the prototype-microclimate pairs used for

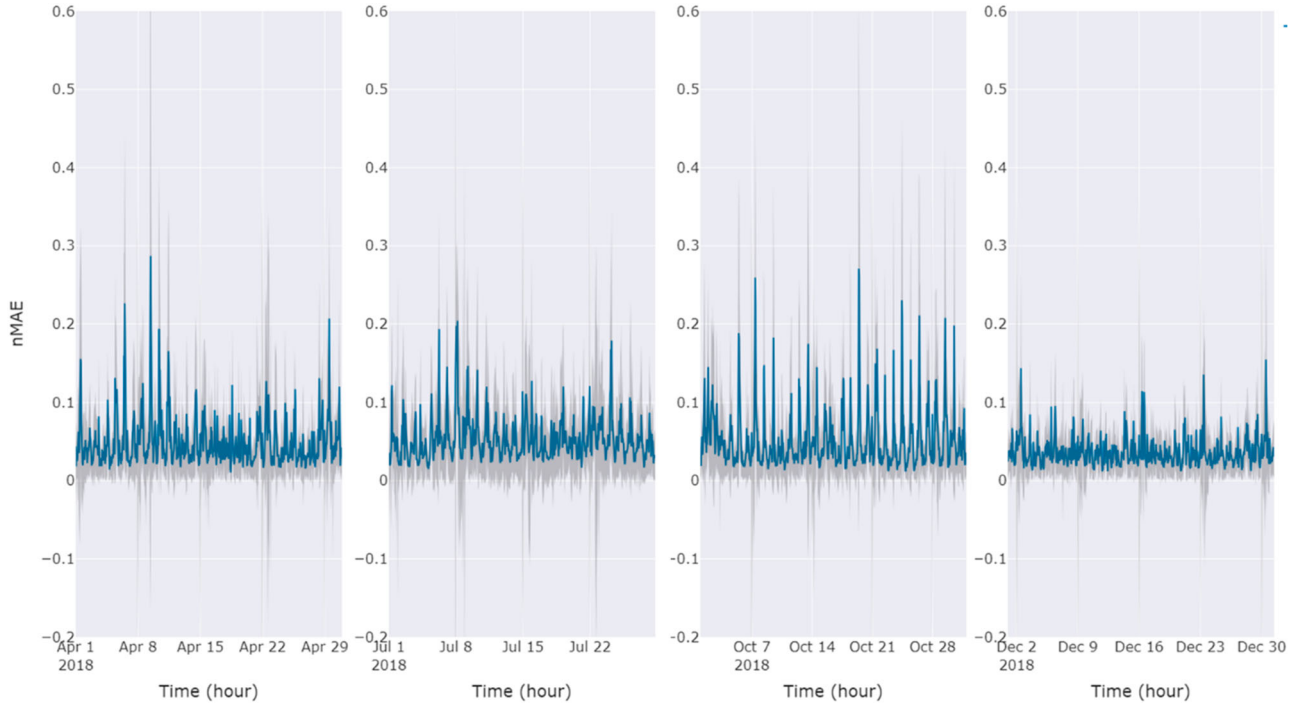

Figure 9. Average nMAE of ELEC estimation among all census tracts with error bar determined by standard deviations. The experiment generating the data for this figure used 10% of the simulation data to build the surrogate model.

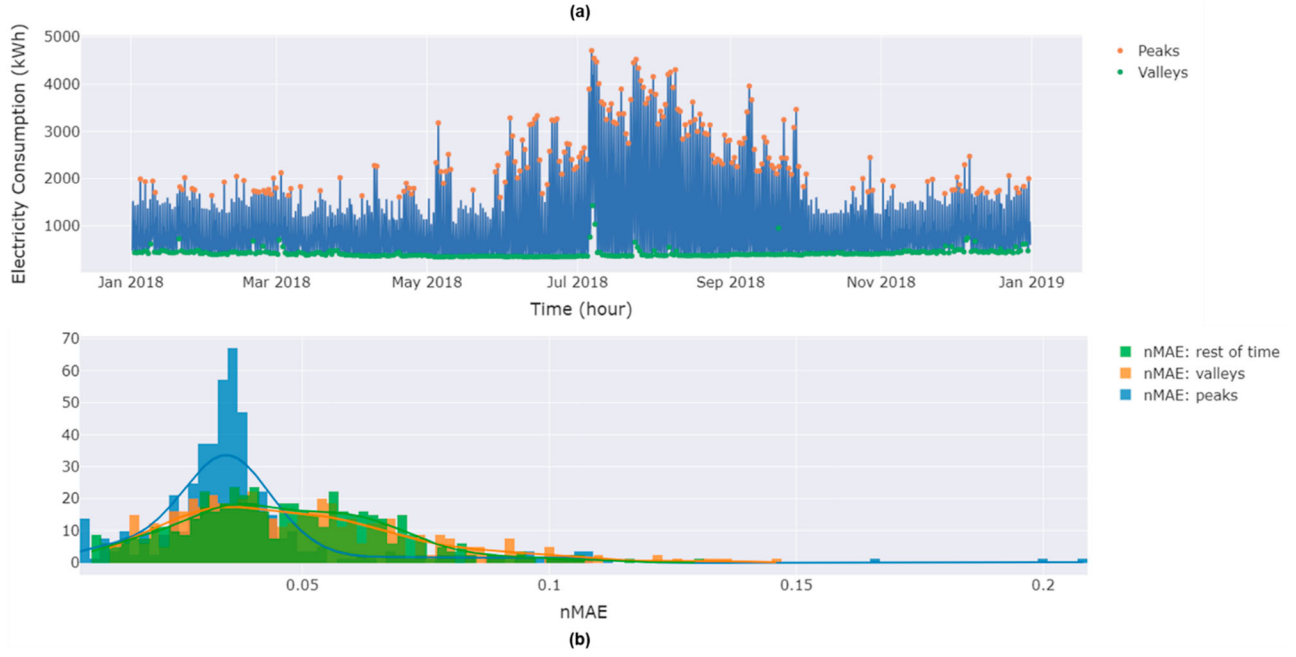

Figure 10. nMAE distribution at the time series peak, valley, and rest of the time. (a) true electricity consumption peaks and valleys of Census Tract 6037101122. (b) nMAE probability distribution of Census Tract 6037101122 at peaks, valleys, and the rest of the time. The mean value of the peak nMAE is 3.89%, and the mean value of the peak nRMSE is 4.98%. The experiment generating the data for this figure used 10% of simulation data to build the surrogate model.

EnergyPlus simulations and training data generation are randomly selected for each prototype. As only a small number of microclimates is selected, the chosen microclimates may be concentrated in certain areas, which could result in biased sampling. In this case, the surrogate models may be overfitted, and their estimation accuracy would be limited. Second, this study selected the same percentage of simulation microclimates for all prototypes

and did not adjust the number of microclimates in different surrogate model training. However, the surrogate model training for some prototypes (e.g. residential prototypes) is more difficult, and therefore requires more microclimates for simulations, and vice versa. On the other hand, the time needed to complete a simulation differs among prototypes. For small-sized buildings (e.g. single-family houses), the simulation model runs faster. It may be worthwhile to increase the number of microclimates and simulations for these prototypes in exchange for higher accuracy.

Third, although the proposed approach in this paper can reduce the one-year UBEM simulation time to about 10% of the original time in most cases, it is not refined in multi-year UBEM surrogate modelling. In this study, patterns learned by the surrogate models in one year are not utilized by the model training in the following years. Thus, the following model training still requires about 10% of the original simulation results, rather than making use of the patterns that are already learned and require fewer simulations. Fourth, the proposed UBES surrogate modelling method does not apply to all cities. In this study, LA County was used in the case study to highlight the significance of the UBES surrogate modelling. However, in small U.S. cities with populations of approximately tens of thousands of people, the number of building prototypes is limited. Moreover, the number of microclimates may also be limited because of the limited city size and potentially homogeneous geographical and building stock characteristics. In this case, it may be more efficient to conduct UBES more conventionally.

As responses to the limitations, some promising future research directions are highlighted. First, when selecting simulation microclimates for generating training data for surrogate models, the microclimates could be clustered. The sampling of microclimates should consider the distribution of microclimates and balance the contribution of all clusters. It could reduce the sampling bias, narrow the difference between the training and test data distributions, and decrease the possibility of model overfitting. Second, instead of setting a fixed microclimate sampling proportion for all prototypes, prototypes with low prediction accuracy and short simulation time could be assigned a higher sampling proportion, and vice versa. This practice may require multiple trials and errors, and the obtained rule-of-thumb sampling proportions and the simulation counts may also apply to other related studies.

Third, transfer learning could be used to transfer the patterns learned by the surrogate models from one year to another year in multi-year UBEM and reduce the simulation time for the latter. There could be two scenarios for this future work. In the case of transferring between the same prototype but different microclimates, a smaller proportion (i.e. below 10%, if 10% is used in the first year) of microclimates and simulations would still be required for the second year. Then, the fine-tuning method could tune the parameters of the trained surrogate models in the first year to adapt to the microclimates in the second year (Li et al. 2021). As an alternative to conducting the transfer learning approach, the domain adaptation method can force the optimizer in the training to look for the common patterns in both years, which makes the surrogate models in the first year also generalizable to microclimates in the second year without additional training (Zhu et al. 2021). In the case of transferring between different prototypes (e.g. transferring between cities), the transfer learning approach is still applicable. In addition, the similarity analysis can calculate the correlations between a new prototype in the second year and the prototypes in the first year. The correlations can be then used to weigh the outputs of the surrogate models in the first year and estimate the target for a new prototype. This approach is simple and free from further model training (Dwivedi and Roig 2019).

Last but not least, to further reduce the model training time of surrogate models, improvements could be made in terms of data preprocessing, model structure, and model training.

- In this study, all usable samples (i.e. sequences obtained from feature time series) were used to train the model. That is not necessarily needed to obtain high accuracy. Randomly selecting part of samples for model training or setting the sliding window stride to be greater than one when preparing training data may be beneficial for computation time, and that can also maintain good accuracy.
- In this study, the surrogate model for each prototype is built separately. One strategy as an alternative is to use a single model for estimating the target for all prototypes. Because the electricity consumption or heat emission mechanism of different prototypes may share some patterns, this strategy could work as multitask learning. It is expected to converge faster and reduce computation time while maintaining accuracy (Dong et al. 2015).
- RNN models process each time step sequentially, which limits the model's compatibility for parallel computation. The Transformer model structure does not rely on sequential computation and enables a more parallel computation (Vaswani et al. 2017). Future research could apply Transformer models and their variants to the surrogate modelling for UBEM and evaluate their performance in terms of accuracy and computation time.

5.3. Conclusions

In this study, biLSTM was used to build surrogate models for simulation based UBEMs. The inputs of surrogate models are weather features and building TMY performances before and after the target time step. The training data are about 10% of the simulation inputs and outputs for the UBEM, and the remaining 90% of the simulations are used for generating estimations for the test, aggregating to census tracts, and evaluating model accuracies. This study found that the surrogate modelling could produce nMAE lower than 10% using 10% of the simulation results in estimating ELEC, E-SURF, and E-EXH, and it requires 20% of the simulations to estimate E-REJ with an nMAE lower than 10%. The nMAE of the surrogate modelling based on biLSTM is lower than that based on LSTM by more than 7%. In addition, the total training time for all surrogate models is generally less than 90 minutes with an NVIDIA V100 GPU. Regarding the distributions of the estimation error, the nMAE are different among the prototypes, which shows the potential bias of the UBEM based on the proposed surrogate modelling approach. In the spatial dimension, the estimation errors are likely to be high in census tracts with a high residential building area percentage. In the temporal dimension, the mean nMAE of the aggregated estimation among census tracts reaches its lowest values in winter and highest in fall. In addition, the nMAE is lower at the peak of the target time series than at the rest of the time. In summary, the novel biLSTM-based UBEM surrogate modelling method could reduce the computation time of UBEM while maintaining a certain level of accuracy. The overall proposed approach is expected to reduce the difficulty in long-term and high-resolution UBEM with widely varying microclimates and promote the corresponding applications and decision-making.

Acknowledgments

Authors thank Dr. Pouya Vahmani for the microclimate data produced using the WRF programme.

Data availability statement

The data that support the findings of this study are openly available in Github at https://github.com/pppxiyu/SurUBEM.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the U.S. Department of Energy: [grant no DE-AC02-05CH11231].

ORCID

Yujie Xu http://orcid.org/0000-0002-1805-1872

References

Abbasabadi, N., and J. K. Mehdi Ashayeri. 2019. "Urban Energy use Modeling Methods and Tools: A Review and an Outlook." *Building and Environment* 161: 106270. https://doi.org/10.1016/j.buildenv.2019.106270.

Bai, S., J. Z. Kolter, and V. Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.

Bande, L., A. G. Cabrera, Y. K. Kim, A. Afshari, M. F. Ragusini, and M. G. Cooke. 2019. "A Building Retrofit and Sensitivity Analysis in an Automatically Calibrated Model Considering the Urban Heat Island Effect in Abu Dhabi, UAE." *Sustainability* 11: 6905. https://doi.org/10.3390/su11246905.

Chandra, R., S. Goyal, and R. Gupta. 2021. "Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction." *Ieee Access* 9, https://doi.org/10.1109/ACCESS.2021.3085085.

Chen, W., Y. Zhou, Y. Xie, G. Chen, K. J. Ding, and D. Li. 2022. "Estimating Spatial and Temporal Patterns of Urban Building Anthropogenic Heat Using a Bottom-up City Building Heat Emission Model." *Resources, Conservation and Recycling* 177: 105996. https://doi.org/10.1016/j.resconrec.2021.105996.

Ciancio, V., S. Falasca, I. Golasi, G. Curci, M. Coppi, and F. Salata. 2018. "Influence of Input Climatic Data on Simulations of Annual Energy Needs of a Building: EnergyPlus and WRF Modeling for a Case Study in Rome (Italy)." *Energies (Basel)* 11: 2835. https://doi.org/10.3390/en11102835.

Cochran, J., P. Denholm, M. Mooney, D. Steinberg, E. Hale, G. Heath, B. Palmintier, et al. 2021. "The Los Angeles 100% Renewable Energy Study (LA100): Executive Summary".

Cohen, M., S. Le Corff, M. Charbit, A. Champagne, G. Nozière, and M. Preda. 2021. "End-to-end Deep Meta Modelling to Calibrate and Optimize Energy Consumption and Comfort." *Energy and Buildings* 250: 111218. https://doi.org/10.1016/j.enbuild.2021.111218.

County of Los Angeles. 2022. Draft 2045 Climate Action Plan.

De Jaeger, I., G. Reynders, C. Callebaut, and D. Saelens. 2020. "A Building Clustering Approach for Urban Energy Simulations." *Energy and Buildings* 208: 109671. https://doi.org/10.1016/j.enbuild.2019.109671.

Dong, D., H. Wu, W. He, D. Yu, and H. Wang. 2015. "Multi-Task Learning for Multiple Language Translation." Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: long papers), association for computational linguistics, Stroudsburg, PA, USA, pp. 1723–1732. https://doi.org/10.3115/v1/P15-1166.

Dwivedi, K., and G. Roig. 2019. "Representation Similarity Analysis for Efficient Task Taxonomy & Transfer Learning." 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR).

Ferrando, M., F. Causone, T. Hong, and Y. Chen. 2020. "Urban Building Energy Modeling (UBEM) Tools: A State-of-the-art Review of Bottom-up Physics-Based Approaches." *Sustain*

Cities Soc 62: 102408. https://doi.org/10.1016/j.scs.2020.102408.

Fu, X., W. Tian, Y. Sun, C. Zhu, and B. Yin. 2020. "Uncertainty Analysis of Urban Building Energy Based on Two-Dimensional Monte Carlo Method." Proceedings of the 11th international symposium on heating, ventilation and Air conditioning (ISHVAC 2019), pp. 1315–1323. https://doi.org/10.1007/978-981-13-9528-4_133.

Herbinger, F., C. Vandenhof, and M. Kummert. 2023. "Building Energy Model Calibration Using a Surrogate Neural Network." *Energy and Buildings* 289, https://doi.org/10.1016/j.enbuild.2023.113057.

Hong, T., Y. Chen, X. Luo, N. Luo, and S. H. Lee. 2020. "Ten Questions on Urban Building Energy Modeling." *Building and Environment* 168: 106508. https://doi.org/10.1016/j.buildenv.2019.106508.

Jones, A. D., D. Rastogi, P. Vahmani, A. M. Stansfield, K. A. Reed, T. Thurber, P. A. Ullrich, and J. S. Rice. 2023. "Continental United States Climate Projections Based on Thermodynamic Modification of Historical Weather." *Scientific Data* 10: 664. https://doi.org/10.1038/s41597-023-02485-5.

Lara-Benítez, P., M. Carranza-García, and J. C. Riquelme. 2021. "An Experimental Review on Deep Learning Architectures for Time Series Forecasting." *International Journal of Neural Systems* 31 (03): 2130001. https://doi.org/10.1142/S0129065721300011.

Li, G., W. Tian, H. Zhang, and X. Fu. 2023. "A Novel Method of Creating Machine Learning-Based Time Series Meta-Models for Building Energy Analysis." *Energy and Buildings* 281: 112752. https://doi.org/10.1016/j.enbuild.2022.112752.

Li, Y., C. Wang, S. Zhu, J. Yang, S. Wei, X. Zhang, and X. Shi. 2020. "A Comparison of Various Bottom-up Urban Energy Simulation Methods Using a Case Study in Hangzhou, China." *Energies (Basel)* 13, https://doi.org/10.3390/en13184781.

Li, A., F. Xiao, C. Fan, and M. Hu. 2021. "Development of an ANN-Based Building Energy Model for Information-Poor Buildings Using Transfer Learning." *Building Simulation* 14: 89–101. https://doi.org/10.1007/s12273-020-0711-5.

Li, Z., N. Zhang, J. Zhang, X. Lin, and W. Zhong. 2022. "Load Forecasting of Large-Space Stadium Based on Surrogate Modeling." 2022 international conference on high performance Big data and intelligent systems, HDIS 2022, 299–303. https://doi.org/10.1109/HDIS56859.2022.9991695.

Liang, Y., Y. Pan, X. Yuan, W. Jia, and Z. Huang. 2022. "Surrogate Modeling for Long-Term and High-Resolution Prediction of Building Thermal Load with a Metric-Optimized KNN Algorithm." *Energy and Built Environment* 4: 709–724. https://doi.org/10.1016/j.enbenv.2022.06.008.

Luo, X., P. Vahmani, T. Hong, and A. Jones. 2020. "City-scale Building Anthropogenic Heating During Heat Waves." *Atmosphere (Basel)* 11, https://doi.org/10.3390/atmos11111206.

Ma, D., X. Li, B. Lin, Y. Zhu, and S. Yue. 2023. "A Dynamic Intelligent Building Retrofit Decision-Making Model in Response to Climate Change." *Energy and Buildings* 284: 112832. https://doi.org/10.1016/j.enbuild.2023.112832.

Magnier, L., and F. Haghighat. 2010. "Multiobjective Optimization of Building Design Using TRNSYS Simulations, Genetic Algorithm, and Artificial Neural Network." *Building and Environment* 45: 739–746. https://doi.org/10.1016/j.buildenv.2009.08.016.

Ohta, Y., T. Sasakawa, and H. Sato. 2020. "Evolutionary Air-Conditioning Optimization Using an LSTM-Based Surrogate Evaluator." 2020 IEEE congress on evolutionary computation, CEC 2020 - conference proceedings.

Oraiopoulos, A., and B. Howard. 2022. "On the Accuracy of Urban Building Energy Modelling." *Renewable and Sustainable Energy Reviews* 158: 111976. https://doi.org/10.1016/j.rser.2021.111976.

Pinto, G., D. Deltetto, and A. Capozzoli. 2021. "Data-driven District Energy Management with Surrogate Models and Deep Reinforcement Learning." *Applied Energy* 304: 117642. https://doi.org/10.1016/j.apenergy.2021.117642.

Prada, A., A. Gasparella, and P. Baggio. 2018. "On the Performance of Meta-Models in Building Design Optimization." *Applied Energy* 225: 814–826. https://doi.org/10.1016/j.apenergy.2018.04.129.

Prataviera, E., J. Vivian, G. Lombardo, and A. Zarrella. 2022. "Evaluation of the Impact of Input Uncertainty on Urban Building Energy Simulations Using Uncertainty and Sensitivity Analysis." *Applied Energy* 311: 118691. https://doi.org/10.1016/j.apenergy.2022.118691.

Sezer, N., H. Yoonus, D. Zhan, L. (Leon) Wang, I. G. Hassan, and M. A. Rahman. 2023. "Urban Microclimate and Building Energy Models: A Review of the Latest Progress in Coupling Strategies." *Renewable and Sustainable Energy Reviews* 184: 113577. https://doi.org/10.1016/j.rser.2023.113577.

Stark, E., R. Drori, and M. Abeles. 2006. "Partial Cross-Correlation Analysis Resolves Ambiguity in the Encoding of Multiple Movement Features." *Journal of Neurophysiology* 95: 1966–1975. https://doi.org/10.1152/jn.00981.2005.

Streicher, K. N., M. Berger, E. Panos, K. Narula, M. C. Soini, and M. K. Patel. 2021. "Optimal Building Retrofit Pathways Considering Stock Dynamics and Climate Change Impacts." *Energy Policy* 152: 112220. https://doi.org/10.1016/j.enpol.2021.112220.

Taheri, S., P. Hosseini, and A. Razban. 2022. "Model Predictive Control of Heating, Ventilation, and air Conditioning (HVAC) Systems: A State-of-the-art Review." *Journal of Building Engineering* 60: 105067. https://doi.org/10.1016/j.jobe.2022.105067.

Tian, W., R. Choudhary, G. Augenbroe, and S. H. Lee. 2015. "Importance Analysis and Meta-Model Construction with Correlated Variables in Evaluation of Thermal Performance of Campus Buildings." *Building and Environment* 92: 61–74. https://doi.org/10.1016/j.buildenv.2015.04.021.

UCLA California Center for Sustainable Communities (CCSC). 2020. "Presenting the Energy Atlas 2.0", https://www.energyatlas.ucla.edu/about.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al. 2020. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." *Nature Methods* 17: 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Westermann, P., and R. Evins. 2019. "Surrogate Modelling for Sustainable Building Design – A Review." *Energy and Buildings* 198: 170–186. https://doi.org/10.1016/j.enbuild.2019.05.057.

Westermann, P., M. Welzel, and R. Evins. 2020. "Using a Deep Temporal Convolutional Network as a Building Energy Surrogate Model That Spans Multiple Climate Zones." *Applied Energy* 278, https://doi.org/10.1016/j.apenergy.2020.115563.

Xu, Y., P. Vahmani, A. Jones, and T. Hong. 2022. *A multi-scale time-series dataset of anthropogenic heat from buildings*

in Los Angeles County, MultiSector Dynamics-Living, Intuitive, Value-adding, Environment. https://doi.org/10.57931/1892041.

Xu, Y., P. Vahmani, A. Jones, and T. Hong. 2024. "Anthropogenic Heat from Buildings in Los Angeles County: A Simulation Framework and Assessment." *Sustainable Cities and Society* In Press.

Yigit, S. 2021. "A Machine-Learning-Based Method for Thermal Design Optimization of Residential Buildings in Highly Urbanized Areas of Turkey." *Journal of Building Engineering* 38: 102225. https://doi.org/10.1016/j.jobe.2021.102225.

Yoo, W., M. J. Clayton, and W. Yan. 2023. "ESMUST: EnergyPlus-Driven Surrogate Model for Urban Surface Temperature Prediction." *Building and Environment* 229, https://doi.org/10.1016/j.buildenv.2022.109935.

Zhang, L., S. Plathottam, J. Reyna, N. Merket, K. Sayers, X. Yang, M. Reynolds, et al. 2021. "High-resolution Hourly Surrogate Modeling Framework for Physics-Based Large-Scale Building Stock Modeling." *Sustain Cities Soc* 75: 103292. https://doi.org/10.1016/j.scs.2021.103292.

Zheng, Y., and Q. Weng. 2018. "High Spatial- and Temporal-Resolution Anthropogenic Heat Discharge Estimation in Los Angeles County, California." *Journal of Environmental Management* 206: 1274–1286. https://doi.org/10.1016/j.jenvman.2017.07.047.

Zhou, Y., Y. Liang, Y. Pan, X. Yuan, Y. Xie, and W. Jia. 2022. "A Deep-Learning-Based Meta-Modeling Workflow for Thermal Load Forecasting in Buildings: Method and a Case Study." *Buildings* 12, https://doi.org/10.3390/buildings12020177.

Zhou, Y., Q. Weng, K. R. Gurney, Y. Shuai, and X. Hu. 2012. "Estimation of the Relationship Between Remotely Sensed Anthropogenic Heat Discharge and Building Energy use." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 65–72. https://doi.org/10.1016/j.isprsjprs.2011.10.007.

Zhu, X., K. Chen, B. Anduv, X. Jin, and Z. Du. 2021. "Transfer Learning Based Methodology for Migration and Application of Fault Detection and Diagnosis Between Building Chillers for Improving Energy Efficiency." *Building and Environment* 200: 107957. https://doi.org/10.1016/j.buildenv.2021.107957.