



Using recurrent neural networks for localized weather prediction with combined use of public airport data and on-site measurements

Jung Min Han ^{a,b,*}, Yu Qian Ang ^c, Ali Malkawi ^{a,b}, Holly W. Samuelson ^{a,b}

^a Harvard University Graduate School of Design, Cambridge, MA, USA

^b Harvard Center for Green Buildings and Cities, Cambridge, MA, USA

^c Sustainable Design Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Keywords:

Weather file
Climate change
Building energy modeling
Deep learning
Recurrent neural network
Gated recurrent unit

ABSTRACT

Weather data is a crucial input for myriad applications in the built environment, including building energy modeling and daylight analysis. Building science practitioners and researchers have been able to select from a variety of weather files, such as Weather Year for Energy Calculation 2 (WYEC2) and the Typical Meteorological Year (TMY). However, commonly used weather files are typically synthesized to represent trends over a relatively longer periods of time, and are often unable to accurately depict climatic conditions that result from local contexts, such as the heat island effect, wind flow, even local temperature and relative humidity. This results in discrepancies in building performance simulations.

This study proposes a methodology using recurrent neural networks to generate synthetic localized weather data that are significantly more accurate and representative of local conditions than standard weather files. The predictions were validated against actual on-site measurements, and achieved a low mean square error of 2.96 and over 185% improvement in validation accuracy. Overall, the performance of selected models has shown over 100% improvements in test accuracy compared with standard weather files and weather station data at the nearest airport. The proposed methodology can be used to morph generic weather files to accurately represent localized conditions, or generate localized data for a longer time span with only a subset of data available/collected. This is useful for downstream built environment applications, especially building energy modeling, since representative weather data capturing trends of temperature and other variables will result in enhanced accuracies of the building energy models. The method can also be used in urban analysis pipelines to enhance resilience against climate change.

1. Introduction

1.1. Importance of weather files for building performance simulations (BPS)

Building performance simulations (BPS) are immensely impactful in a building's lifecycle especially in the early design phase. Multidisciplinary designers, consultants, builders, and clients rely on these simulations to develop initial concepts and iteratively improve the performances of building designs and their associated configurations. For decades, BPS software and applications have utilized a variety of weather data in different formats, and weather data is now an integral input for myriad applications and simulations in the built environment. These applications include urban building energy modeling [1] and

daylighting analysis, among others. Studies such as [2–4] have also found that the spatial variability of weather data is a dominating factor in BPS, and the selection of weather data significantly influences building energy simulation results [5].

In modern building projects, design and performance simulations require weather data for a variety of analysis at some point. One can set up a local weather station, but not without cost and time implications, since there is often a limited time to collect data before it is required (e.g., to start analyzing design decisions). Additionally, BPS analysis typically require at least a full year of data to adequately capture climatic and seasonal trends over time.

A common weather dataset, the EnergyPlus Weather Format (EPW) [6], is a widely used file format in BPS for the EnergyPlus simulation engine. The EPW database now covers over 2,100 locations around the

* Corresponding author. Harvard University Graduate School of Design, Cambridge MA, USA. Harvard Center for Green Buildings and Cities, Cambridge, MA, USA.
E-mail addresses: jhan2@gsd.harvard.edu, jhan2harvard@gmail.com (J.M. Han).

world. Another popular database is Climate. [EnergyPlus.net](#) provides Typical Meteorological Years (TMY) files published by different organizations to support BPS. These TMY files provide representations of the weather conditions in specific locations, and formats such as TMY3 avoid using only single-year information – such as the Test Reference Year (TRY) – in an attempt to capture typical rather than extreme conditions. TMY is the go-to standard set of weather files for use in building energy simulations. However, due to the relatively outdated and often averaged data in TMY files, opportunities exist for potential improvements especially to provide data that better represent local climatic conditions. Typical weather files (e.g., TMY3 or EPW) are often unable to accurately represent climate conditions that result from local phenomenon, such as the effect of urban heat island and wind, as well as contextual buildings. Additionally, TMY files typically come from certain locations – often airports – which may be remote, and have different weather conditions than the building site of interest. With the lack of localized data and the often laborious effort required to collect them over a period of time (to capture seasonality), BPS practitioners have no choice but to use standard weather files.

BPS practitioners also tend to avoid using single-year weather data such as the TRY, since a single-year range neither accurately represent typical long-term weather patterns nor representatively depict the local boundary conditions around the target building. Therefore, more comprehensive methods to produce a synthetic year representing temperature and other variables are necessary, and will result in enhanced accuracies in predicted energy consumption in BPS and analysis.

With the global imperative to address the issues of climate change and meet carbon reduction goals, one method that is gaining popularity is the concept of morphing available weather data to reflect desired climate and meteorological conditions. Weather file morphing focuses on predicting scenarios throughout a building's lifecycle, current and future. With increased urbanization, urban heat island (UHI) effect has become a common phenomenon in many major cities worldwide, where in many regions the air temperature of densely built urban area is higher than that of the surrounding rural area [7]. This has led to a growing concern that the use of standard weather data would result in inadequate decision-making regarding energy efficiency of buildings in metropolitan areas. The impact on climate change on buildings and preparation of the future weather files have also been highlighted in studies [8]. In this context, initiatives and applications in site-specific performance-driven design and analysis, as well as building operation and control, have emerged. For example, [9] developed a web-based Future Weather File Generator for four Intergovernmental Panel of Climate Change (IPCC) scenarios, to morph existing EPW and TMY files to reflect future climates, while [10] developed an approach to construct a modified weather file for Hong Kong, taking into account urban heat island effect in the summer season. [11] deployed a morphing technique to transform historical time series data to generate a projected weather file that reflects a plausible future climate.

1.2. Using data-driven methods to enhance weather modeling

To further circumvent or resolve the problems mentioned, it is also becoming increasingly common for researchers and/or practitioners to employ data-driven models incorporating statistical or modern machine learning toolsets, especially when physical, physics-based models or site-nonspecific averaged data may not fully describe the processes in operational situations. In the building science domain, modern challenges necessitate hybrid models to solve design, simulation, optimization and control problems. These models are useful tools for understanding the processes of physical properties that surround buildings and their associated impact, allowing researchers and/or practitioners to make inferences and predictions. The models also provide valuable feedback on design options and alternatives. With increasing computing power and the proliferation of available data resources on open data platforms, the combined use of physical and data-

driven modeling techniques will be essential in future BPS endeavors.

Such data-driven or integrative empirical models involve not just numerical mathematical equations from the physical BPS processes alone, but analysis of concurrent input and output data in time series data of higher fidelity. As more data become available – and more degrees of freedom are required and explored in these models – it is becoming increasingly challenging to incorporate all available data, properties, and parameters into a single predictor. One can further argue that the empirical parameterizations necessary for numerical models should be constructed using data-driven modeling methods, such as machine learning (ML) and deep learning (DL) techniques, because these are tools capable of operating on sparse, high-dimensional datasets. Examples of commonly used ML techniques in BPS include artificial neural networks (ANNs) [12,13], Bayesian inference [14], as well as evolutionary algorithms [15].

Studies such as [16,17] have reviewed ANNs models for building energy simulations. Specifically, ANNs research in the building science domain has been extended to myriad of areas such as building energy use [18], solar radiation, renewable energy systems such photovoltaic [19], and wind speed, etc. However, many applications in building science, such as heating systems [20], weather conditions [21], and daylighting [22] are done primarily with feed-forward ANNs. In view of the proliferation of more complex neural network architectures and open-source machine learning libraries from major corporations such as Google (TensorFlow) and Facebook (PyTorch), it is worthwhile exploring more modern, sophisticated networks able to model these complexities. In this context, the use of recurrent neural networks (RNNs) is receiving increased attention due to its potential to work with time series data. This research investigates the utilization of RNN models for localized weather predictions, and subsequent application of the predicted data in energy modeling and other BPS workflows.

2. Research problem

2.1. Literature Review

There are multiple ways to predict weather conditions and synthetic weather information such as temperature, relative humidity, precipitation, wind speed, and solar radiation. Most methods utilized in building performance simulations can be classified under four main categories (*viz.* Regression models including Auto Regressive Integrated Moving Average (ARIMA) [23,24], artificial neural network approaches [25,26], numerical weather prediction models [27,28], as well as other methods such as direct sensing and hybrid models).

ARIMA is a class of models that, given time-series data, seeks to make inferences and predictions based on the data's own past values. Prior studies in building science have used ARIMA methods to develop models for temperature and relative humidity [29], weather forecasting [30], global solar radiation [31], and land surface coverage as well as PM₁₀ concentrations in a high altitude city [32]. Classical ARIMA models have been conventionally popular, being the go-to method for researchers working with time-series weather-related data before the proliferation of more sophisticated machine learning approaches. These models typically generate relatively reliable results and predictions. ARIMA models are also relatively easy to implement and can be utilized for almost any type of time series data. However, while they can overcome some disadvantages associated with pure physics-based models, these classical regression models typically possess limitations with regard to assumptions of normality, linearity, and variable dependence. First, while ARIMA models work with short-term and univariate data, they do not scale well in the longer term, and with multiple variables (exceptions include models like seasonal ARIMA) [33]. These models also typically do not support missing values, and assume that the underlying data has a linear relationship [34].

Numerical weather prediction models [27], on the other hand, are based on the physical laws of weather and its associated boundaries and

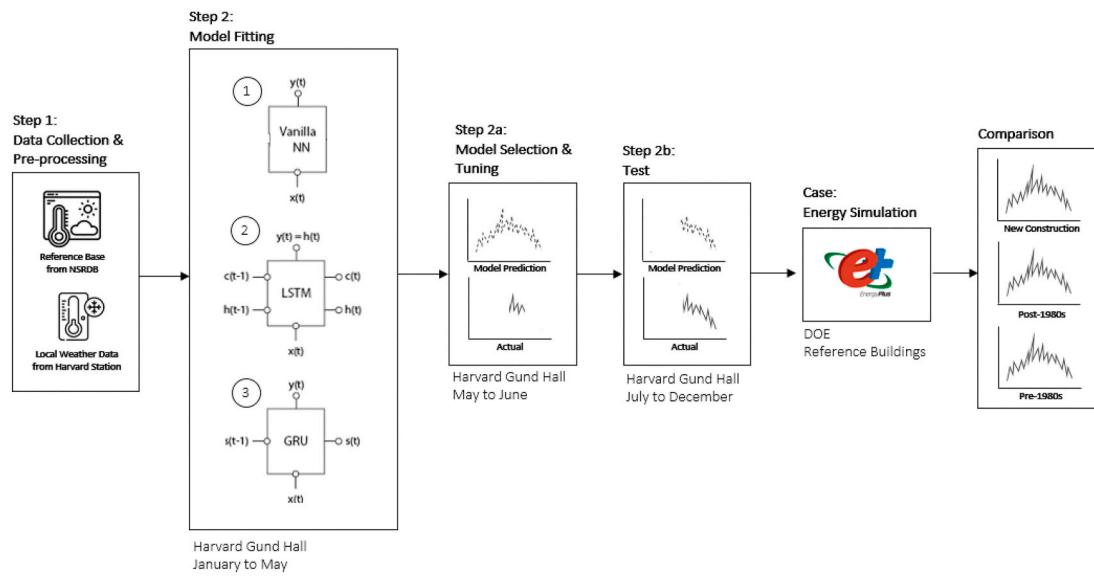


Fig. 1. Proposed workflow for the study.

environmental conditions. This method is normally used for generating regularly populated data throughout the year, or simulating parametric options with the freedom to compute all possible options of the output based on numerical inputs. For example, a physical sky model can be a representative numerical model in BPS, used to calculate annual daylighting values or solar power generation using different solar azimuth angles as numerical inputs [35].

Artificial neural network (ANNs) or feed-forward neural networks (FFNN) are a class of approach popularized by the proliferation of lower-cost computational resources in recent years, and modern open-source machine learning libraries. ANNs/FFNNs are data-driven predictive models able to represent non-linear behavior in high dimensional weather datasets. These models can approximate non-linear relationships in time series values such as irradiance, precipitation, relative humidity, etc. As ANNs/FFNNs allow fewer inputs, require less computation time, and offer superior performance, they have received significant attention for use in predicting weather and other time-series applications [36]. The benefits include reproducibility, time-efficiency, and scalability, ANNs model can be easily expand to different temporal resolutions such as daily or hourly resolution [37]. Deep neural networks (DNNs) are ANNs models with multiple hidden layers, and these models have proven to be adequate in approximating and solving complex problems with non-linearities [38]. These deep learning techniques have been used in a variety of works to infer trends and generate predictions from time-series data. For example [39,40] studied the utilization of neural networks to estimate building energy consumption, solar radiation, and local temperature, while [25] deployed a multi-layer perceptron (MLP) to derive daily predictions of global solar radiation on a horizontal surface. The method was validated and found to perform similar or better than conventional methods such as ARIMA. In other studies, [41,42] used ANNs for weather predictions and temperature forecasting. It is thus clear that these data-driven machine learning approaches using various architecture confer advantages and new functionalities to time series modeling and weather predictions.

Recurrent neural networks (RNNs) are a relatively more sophisticated class of neural networks especially suited for modeling complex relationships in time series data. RNNs can process multi-variate sequence data as inputs, perform temporal feature extraction, and generate multi-variate output (predictions). They are also robust to noise. However, research on utilizing RNNs for weather prediction and solar radiation studies is still relatively nascent in the field of BPS. Unlike conventional feed-forward neural networks, RNNs possess a “memory

state” to process temporal information. Although RNNs require increased computational resources compared with standard ANNs/FFNNs, advancements in computing hardware have democratized these methods, and made RNNs highly accessible.

This study analyzes the use of RNNs in BPS and proposes a methodology to generate highly accurate localized weather predictions, suitable for downstream BPS applications such as energy simulations. The study also compares the performance of vanilla ANN/FFNN approaches with RNN variants, validated against data measured on-site.

2.2. Research objectives

This study uses generic, public data collected from the National Solar Radiation Database (NSRDB) together with measurements from on-site sensors to generate improved weather data that more closely matches site-specific local conditions. Typical local weather stations are located on the rooftop of buildings, with time and cost implications to set-up, and collect target data at a certain resolution during limited periods. This renders full-year data unavailable for building performance simulations. In this context, the objectives of this study are two-fold:

- 1) To demonstrate that, by using public data with only a small subset of measured data, one can still generate accurate weather predictions over a longer/full timeframe, incorporating seasonal trends crucial for BPS. This will be useful since on-site measured data are usually intermittent, and may be collected for a much shorter duration than a full year. This method will enable researchers and practitioners to morph generic weather files such as airport data into representative localized data, with relatively lesser on-site measurements required, thereby saving time and cost.
- 2) To illustrate the impact of using accurate localized weather files in subsequent downstream BPS use cases, such as energy modeling, and to discuss implications for designers and modelers in the built environment especially during early design stages.

3. Methodology

The methodology employed in this study is a three-step empirical process replicable and generalizable across different regions. The first step comprises data collection and pre-processing, with the two main data streams being NSRDB airport data and on-site measurements collected specifically for this study. These time series data served as

Table 1

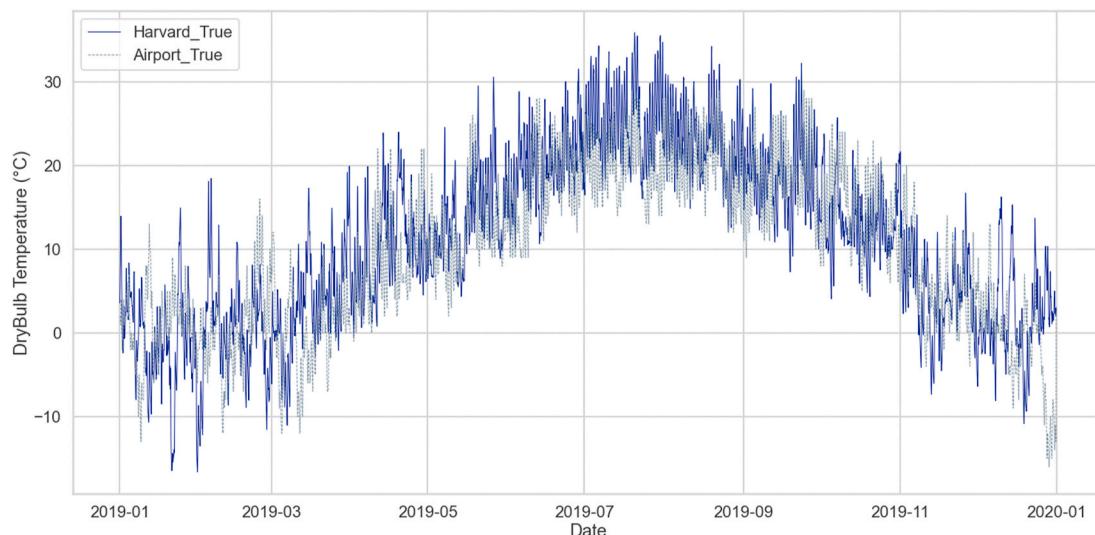
Datasets, labeling, and periods of utilizations.

	Full Data Collected	Training Data	Model Selection	Test Data
Data Period	Jan to Dec	Jan to May	May to June	July to Dec
Data Labeling	growhead Airport_True	–	–	–
Data Period	Jan to Dec	Jan to May	May to June	July to Dec
Data Labeling	Harvard_True	Harvard_Training	–	Harvard_Prediction (FFNN/LSTM/GRU)
Data Period	Jan to Dec	April to June	June to July	July to Dec
Data Labeling	HouseZero_True	HouseZero_Training	–	HouseZero_Prediction

Table 2

Comparison of the measurements (mean) from typical airport dataset and local weather station at the Harvard Gund Hall.

Mean (std)	Dry bulb Temperature	Dewpoint Temperature	Relative Humidity	Solar Radiation	Wind Direction	Windspeed
Airport	9.72 °C	6.40 °C	83.57%	186.89W/m2	208.07°(N = 0°)	0.66 m/s
Harvard	11.25 °C	5.02 °C	68.96%	151.05W/m2	211.55°(N = 0°)	1.67 m/s

**Fig. 2.** Temperature data plot for airport weather file vs data collected from localized sensors.

structured feature inputs fed into the different neural network models.

Next, data are fitted to various neural network model architectures, including a base vanilla FFNN, a Long-Short Term Memory (LSTM) model, and a Gated Recurrent Unit (GRU) model. This study also describes an important feature engineer step – the *look-back* – which determines the number of previous time-steps or inputs to consider. The hyper-parameters were tuned for optimal predictive performance, and validated against a held-out test set.

In the final stage, impact of using accurate localized weather data was demonstrated by running energy simulations using the EnergyPlus engine on the United States (US) Department of Energy (DOE) reference building templates [43]. By controlling for other variables and parameters in the energy model, the study analyzes and compares the effects of using model predictions vis-à-vis using standard weather files. This illustrates the implications and importance of using accurate localized weather data for early-stage design analysis and studies, especially for energy modelers, architects, and sustainability consultants. Fig. 1 illustrates the workflow/framework for our study, while Table 1 summarizes the datasets and their utilizations.

3.1. Data collection and pre-processing

As described in Section Three, the study utilized two main data

sources – generic public airport data from NSRDB and on-site measured data. NSRDB data fundamentally tracks meteorological data as well as three common measurements of solar radiation – global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI) [44]. Older iterations of the NSRDB data were modeled using weather information collected at airports around United States, while current NSRDB data utilizes measurements from geostationary satellites. TMY weather files, introduced in Section One of this study, are typically derived from the NSRDB time-series data, where the weather and meteorological data are consolidated and condensed [45]. For example, in a typical TMY1 data file, twelve months that best represent median conditions are selected from a multi-year dataset. Since these TMY data represent typical conditions – with the values nuanced – they are unable to capture more extreme seasonal or diurnal variations as well as local conditions. The National Renewable Energy Laboratory (NREL), which manages the NSRDB database, has also stated that these TMY files are unsuited for use in systems design and predictive control, especially to meet extreme or localized conditions.

For this study, full sets of annual hourly temperature profiles with 8,760 datapoints were sourced from the NSRDB database (for Boston

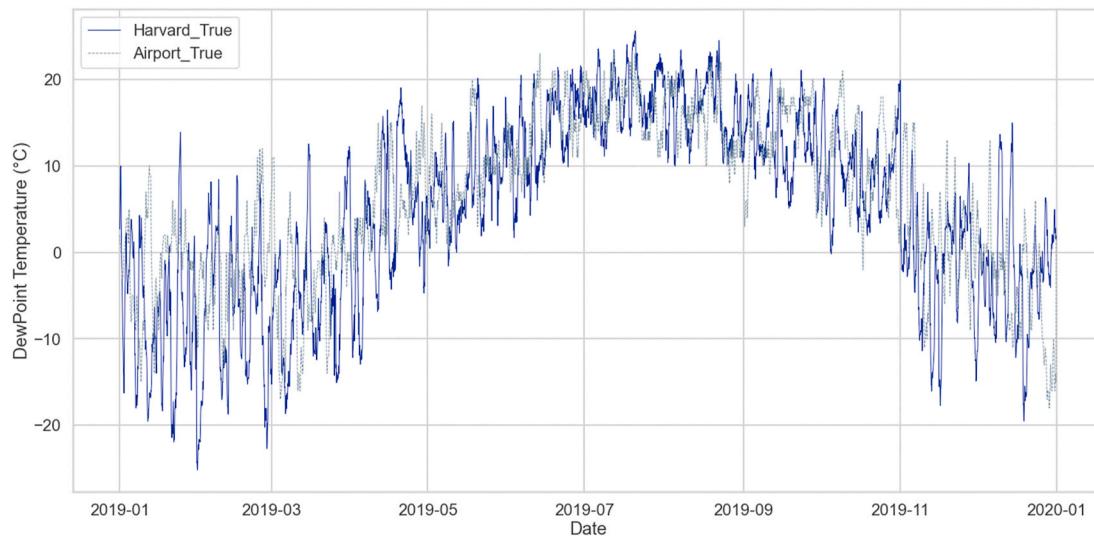


Fig. 3. Dewpoint temperature plot for airport weather file vs data collected from localized sensors.

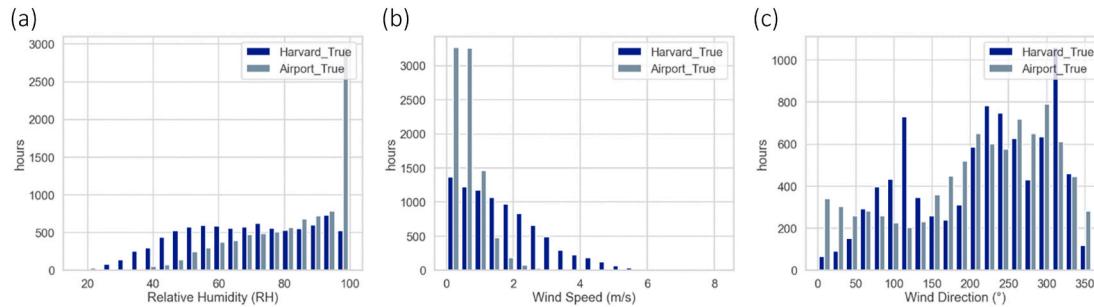


Fig. 4. a (left), b (middle), c (right). Histogram for relative humidity, wind speed and wind direction at the Harvard Gund Hall and the Boston Logan Airport.

Logan airport). Separately, on-site measurements were taken from the weather stations and sensors at the Harvard Graduate School of Design (GSD) Gund Hall¹ and the Harvard HouseZero.² Logan airport is directly adjacent to the Boston Harbor and close to the water body. It is approximately five miles away from the Harvard HouseZero and the Harvard Gund Hall, both of which are inland.

The NSRDB data (“Airport True data”) serves as a reference base, and nine distinctive features representing weather conditions such as temperature, solar radiation, relative humidity, wind speed, wind direction, and solar radiation information were gathered for a full year. The solar information varied in terms of several detailed values, such as the DHI, DNI, GHI, and the solar zenith angles were used as a reference. Data from the Harvard Gund Hall, on the other hand, was segmented into two half-sets of 4,044 datapoints each – from January to June and July to December. The first dataset was used for training and validation, and the second (unseen) set was used to validate the result of the model prediction. A separate set of weather data was collected at the Harvard HouseZero weather station. This weather dataset from HouseZero serves as an additional independent test set to further validate the validity and accuracy of the model.

Values for the local condition were calculated by using the physics-based models as detailed in [46] and the local weather file was prepared using data from sensors installed at Harvard Gund Hall. Table 2 shows the differences between the airport data and data collated from the localized weather station. The mean dry-bulb temperature difference

was approximately 2 °C and the mean difference in relative humidity was approximately 15%. The mean windspeeds are rather different for both areas as well – 0.66 m/s at Logan airport and 1.67 m/s at Harvard.

From these initial observations, it is evident that the use of weather data obtained directly from the nearest airport does not represent well the local boundary conditions around buildings, even if the locations are geographically not too far apart. This highlights the risks of directly using generic public weather files for downstream BPS applications, such as the potential of introducing inconsistencies and inaccuracies if building simulation practitioners simply use generic airport data for simulations. For example, Figs. 2 and 3 illustrate the difference between the dry bulb and dew point temperatures obtained using the generic airport weather file and the measured data collected via local sensors. Fig. 4a, b and c, on the other hand, show the histograms for the collated data. The temperature data from local sensors (at Harvard Gund Hall), although in general follows the trend depicted in the airport data, depicts a wider temperature range profile, and generally displays greater diurnal variations from January to June. Relatively humidity at the airport generally varies along greater ranges across the year compared to the local data, as reflected in Fig. 4a – the airport data contains a high number of hours with high relative humidity at or close to 100%. The airport data also has a higher concentration of low wind speed hours (Fig. 4b) compared with the local data, which has a relatively more even spread of wind speed. Both the airport and local weather are more similar in terms of wind direction (Fig. 4c).

3.1.1. Training, validation, and test data splits

This study assumed full resolution data collected from the airport and from sensors at the local weather station at the Harvard Gund Hall

¹ 48 Quincy St, Cambridge MA 02138, USA.

² 20 Summer Road, Cambridge MA 02138, USA.

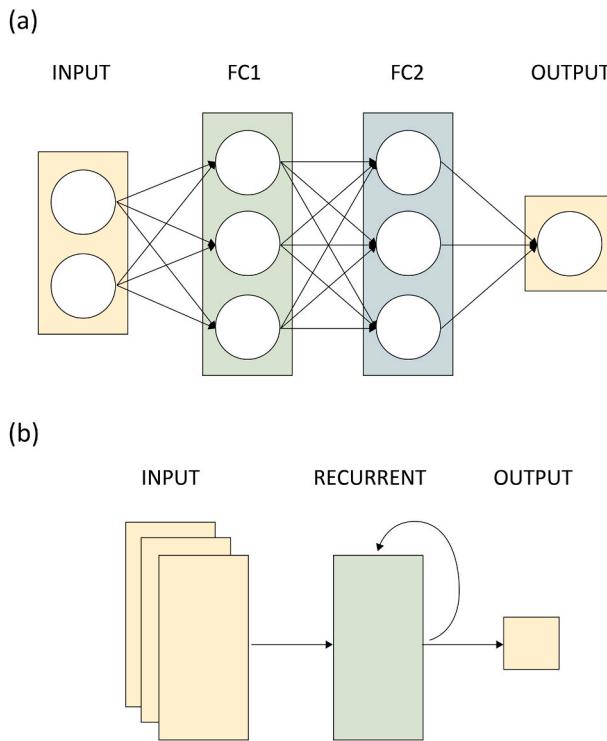


Fig. 5. a (top), b (bottom). Comparison of feed-forward neural network (top) with recurrent neural network (bottom).

over the duration of one year (2019). A subset of the Harvard Gund Hall data is used for training by the neural network – i.e. the fitted ANN models will be trained on the weather data from January to June, and the models will then be used to generate predictions for July to December. For validation and testing, these predictions will be compared against actual on-site measured data, to calculate accuracy improvements over the base data from the generic airport file.

Better local weather data around buildings are useful in downstream applications, especially in predicting and calculating a building's energy consumptions as well as indoor environments. The following section describes the model architecture and generic workflow for predicting local weather data.

3.2. Model architecture design, fitting, and analysis

3.2.1. Feed-forward neural network (FFNN)

A typical ANN comprises a series of connected neurons, each producing a sequence of linear or non-linear activations [47]. Based on the complexity of the problem, the network can be parameterized with more layers representing different non-linearities or latent space. A shallow

FFNN has fewer such layers, while a FFNN can be characterized as a deep neural network if it contains many layers. Because non-linearity is often an issue when predicting local weather, vanilla NN models can be constructed with multiple, so-called hidden layers to model or approximate the non-linearities of high dimensional weather datasets. FFNNs also possess autoregressive properties that can be deemed more complex than typical ARIMA models; in ARIMA, the model generates the next prediction based on the moving average, while FFNNs attempt to approximate a non-linear function mapping previous steps (or inputs) to the next. Our base model in this study comprises a FFNN with 50 neurons with a linear output activation unit, optimized using the Adam [48] optimizer.

3.2.2. Recurrent neural network (RNN): long short-term memory (LSTM)

While theoretically, a deep FFNN or MLP can approximate any function to arbitrary precision by simply using more layers and neurons, the typical assumption is that the inputs-outputs are independent of each other. In contrast, a recurrent neural network [49] contains shared neuron layers (and weights) between inputs through time, providing it with the capacity to ‘memorize’/‘process’ previous time-steps. This makes RNNs suitable to model sequential data and its associated temporal dynamics to greater accuracies. RNNs have been used in domains such as stock prices and language modeling, and the nature of its architecture makes it suitable for weather data in the context of this study. This is especially important since weather data includes non-stationary temporal characteristics such as temperature and wind-related features. Fig. 5 shows a simplified comparison.

However, simple RNNs typically witness a vanishing gradient problem, where depending on the activation function, information “fades over time”, and the non-linearity term is often inadequate for longer-term memory. Specifically, once the weight of the neurons reaches a certain range, the previous states are no longer informative. To overcome this problem, LSTMs [50] and GRUs [51] were developed. LSTMs help preserve errors that can be backpropagated through time and layers. By maintaining these errors, LSTMs allow RNNs to continue to learn more effectively across many time-steps.

An example LSTM architecture used in this study is illustrated in Fig. 6. The key features of LSTM include its *cell state* and various *gates*, which act as a medium to transfer temporal information throughout the input-output sequences. As the cell state moves through the sequence, information is accepted or removed via the gates, which can “learn” what information is relevant as the neural network goes through training epochs. For this study, the LSTM model is parameterized with a layer of 50 LSTM units with hyperbolic tangent function connected to a final feed-forward layer of 30 units, followed by a linear output. Training and validation are monitored using the mean square error metric. The Adam optimizer was selected after hyperparameter tuning, and 3000 epochs were used (with an appropriate early-stopping callback), with batch sizes of 64.

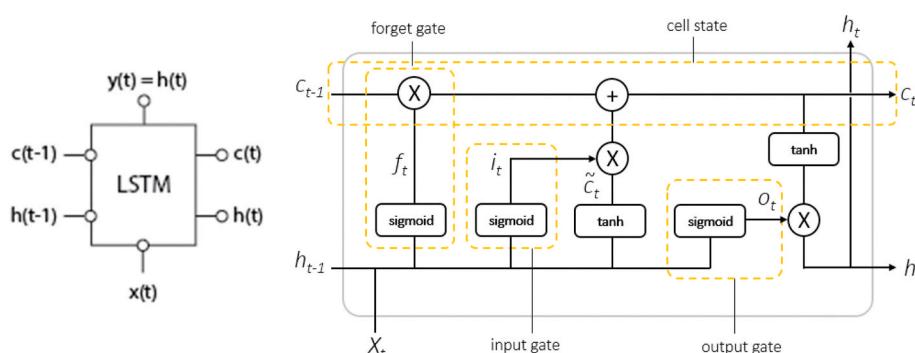


Fig. 6. Typical LSTM architecture including inputs and outputs.

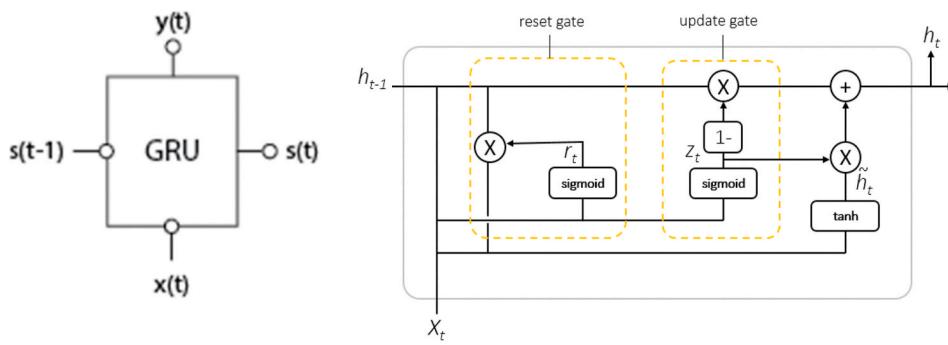


Fig. 7. Typical GRU architecture including inputs and outputs.

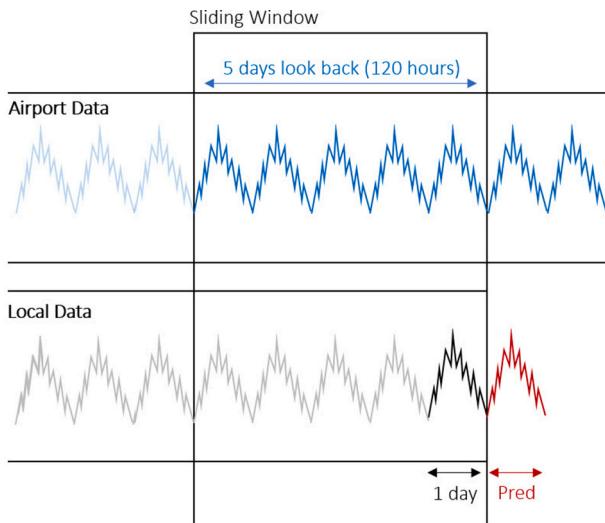


Fig. 8. Example illustration of the look back parametrization in this study.

3.2.3. Recurrent neural network (RNN): Gated Recurrent Unit (GRU)

In the GRU (Fig. 7), a newer but simplified version of LSTM, the cell state is removed, and the hidden state is utilized to transfer information instead. GRUs also have only two gates – a *reset gate* and an *update gate*. The update gate behaves similarly to the forget gate in LSTM by deciding what information to retain and what new information to add, while the reset gate is another mechanism to determine how much past temporal information to discard. GRUs are typically faster to train compared with LSTM since they have less tensor operations. For this study, the GRU is parameterized with a layer of 50 GRU units with hyperbolic tangent activation, connected to a final layer of 30 feed-forward units followed by a linear output. Training and validation are monitored using the mean square error metric. The Adam optimizer were used and 3,000 epochs were trained (with an appropriate early-stopping callback), with batch sizes of 64 for training sets.

3.2.4. RNN look-back

A key feature of RNNs is the ability to incorporate learned

Table 3

Comparison of model mean square error (MSE) for temperature against measured.

	Generic Airport Data	Simple FFNN	RNN: LSTM	RNN: GRU
MSE against measured data	19.66	31.28	4.72	2.96
% improvement over airport data	–	40.9%	176.0%	184.9%

parameters from previous inputs or time states/steps. A typical, vanilla RNN may only be able to incorporate data from a prior time-step, which may be insufficient to generate robust or accurate predictions, and also suffer from the potential vanishing-gradient problem. LSTM or GRU on the other hand, can incorporate a larger number of prior time-steps, in certain cases with different assigned weightages. These models can selectively ‘remember’ or ‘forget’ information, trends, or patterns for much longer durations, and simultaneously determine which data should be retained in the network going forward. They are thus more suited for time series modeling compared with conventional ANNs/FFNNs or vanilla RNNs. In RNNs, the *look-back* refers to and determine the number of prior time-steps of data (or data points) to feed to the model, typically in one data point. Fig. 8 illustrates the *look back* parameterization for our RNNs method.

In the study, the *look-back* was set to 5 days (120 hours) of the generic airport data, and together with one day’s (24 hours) measured on-site data, the model will generate a prediction for the next 24 hours. The model then utilizes the new data points for further prediction, as illustrated in the ‘sliding window’ mechanism in Fig. 8. This will allow the model to capture seasonal variations and patterns throughout the year, even with only a subset of localized data collected at the Harvard Gund Hall, and morph/transform the airport data to represent and reflect local conditions more accurately. This capturing of seasonal trends and variations is crucial for BPS and its downstream applications, especially in regions with significant variations in climate conditions over the course of the year, with different cooling and heating demands.

4. Results and discussion

4.1. Model training – using airport data and Harvard Gund Hall data (Jan to May)

As described in earlier sections of the paper, the data is segmented into training, validation/model selection, and test sets. The training data comprises five months of data from both airport and the Harvard Gund Hall, from January to May. Subsequent data from May to June was utilized for model selection and tuning, and finally, data from July to December serves as the held-out test set against which the model prediction was validated. All three models – FFNN, LSTM, GRU – were trained with the same training dataset, and tuned similarly with grid search across hyperparameters. This ensures consistency across the three models, with the only variable being the neural network model architecture. The optimal hyperparameters for the FFNN, LSTM, and GRU differ slightly. Thus, in this study, the best performing score for each model was compared.

The next subsections below discuss the results and findings of the study.

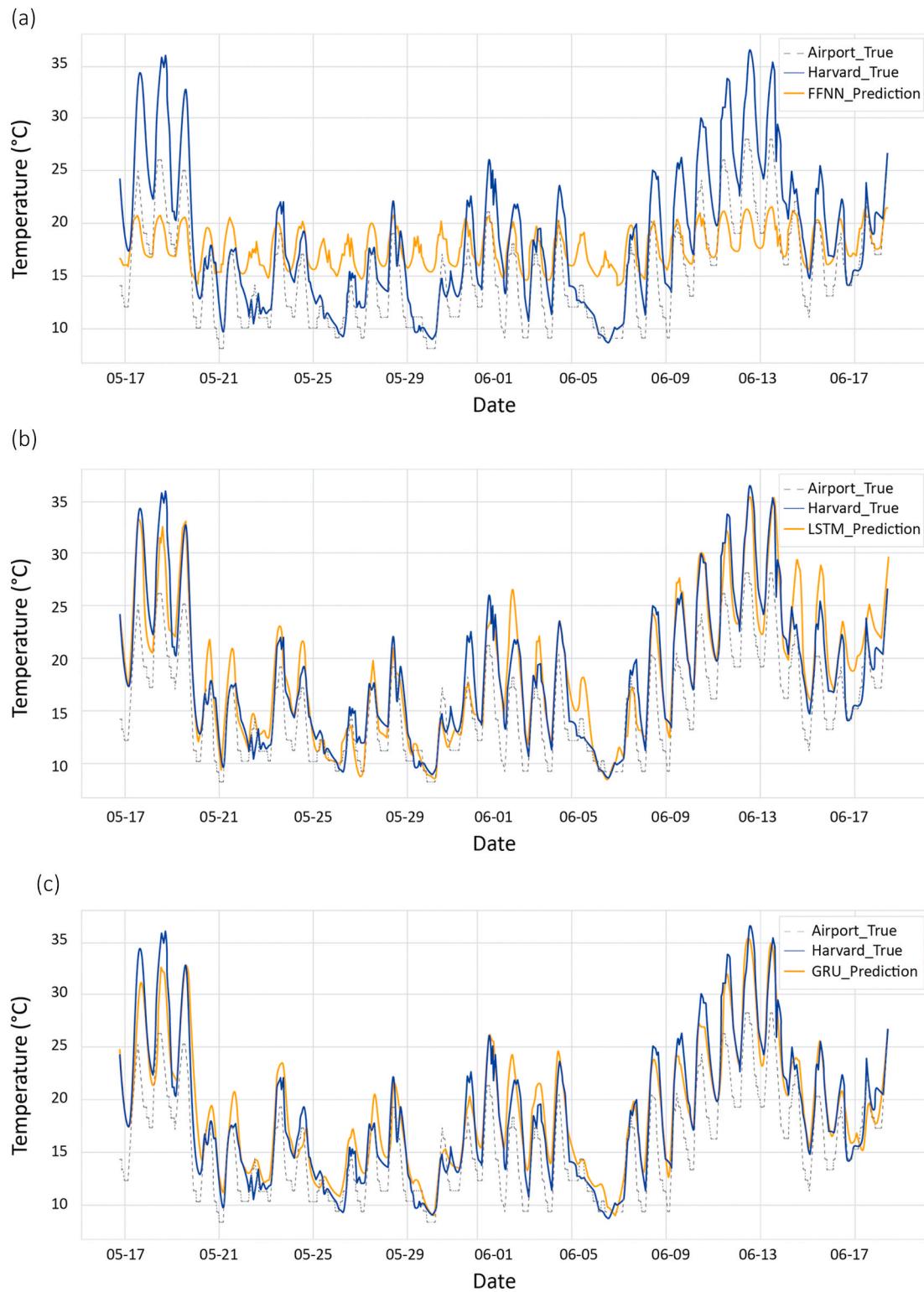


Fig. 9. a (top), b (middle), c (bottom). Model validation plots for FFNN, LSTM and GRU.

4.2. Model selection and tuning – using airport data and Harvard Gund Hall data (May to June)

The GRU model performed best during model selection and tuning, yielding a low mean square error (MSE) of 2.96, significantly better than the MSE of 4.72 for the LSTM and 32.28 for the simple FFNN. It is also over 185% more accurate than the airport data when validated against measured data at on-site (i.e. prediction from the GRU and the general

airport data were compared against data measured on-site via using sensors and apparatuses. Table 3 summarizes the MSE comparison between the three model and the airport data against ground truth measurements taken on-site.

Fig. 9 shows the plots for model selection using the Harvard Gund Hall data for May to June. In all three plots, Harvard_True (Navy line) represents the actual measured data on-site, Airport_True (dotted grey line) represents generic airport data, and the final orange line represents

Table 4

Comparison of MSE for four parameters against measured data.

MSE	Airport Data	LSTM	GRU
Against measured Dewpoint Temperature ($^{\circ}\text{C}$)	5.04	8.72 (27.0% improvement)	3.38 (132.9% improvement)
Against measured Relative Humidity (%)	670.51	170.70 (174.5% improvement)	181.42 (172.9% improvement)
Against measured Wind Speed (m/s)	2.40	1.34 (144.2% improvement)	0.99 (158.8% improvement)
Against measured Wind Direction (North = 0°)	9515	3515 (163.1% improvement)	3025 (168.2% improvement)

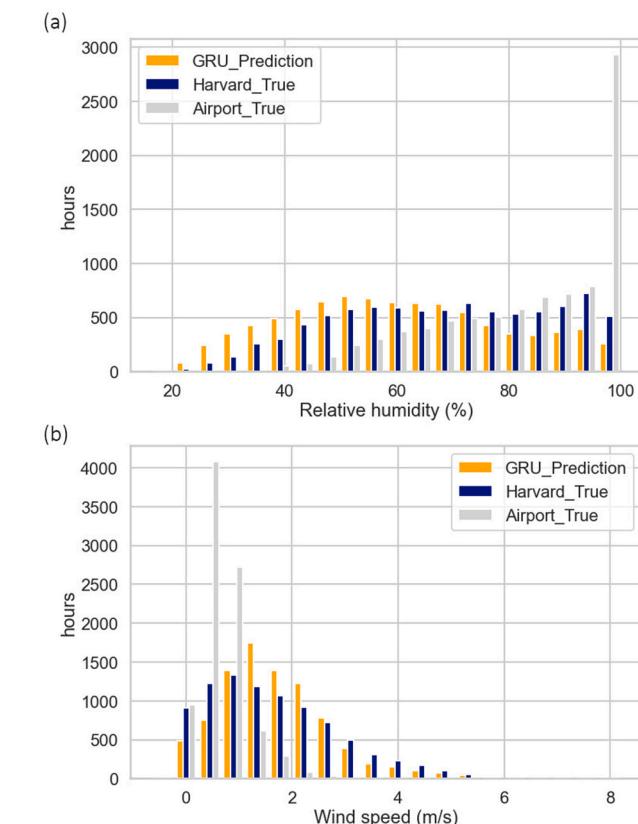


Fig. 10. a (top), b (bottom). Comparison plots for relative humidity and wind speed.

the model predictions for each model (*viz.* FFNN, LSTM and GRU). As observed, the simple FFNN is unable to model the non-linearities in the temperature data – while some of the general trends are captured, the more extreme variations especially the peaks and troughs are excluded. The FFNN also performs worse than generic airport data. In contrast, both RNN models exhibits much better alignment with the true measured values, and can accurately capture the diurnal variations. The GRU model presents the best fit out of all models, being much better at capturing diurnal variations and irregularities compared with simple FFNN model or the base generic airport data.

4.3. Model test – using airport data and Harvard Gund Hall data (June to December)

Table 4 shows the prediction MSE for the LSTM and GRU models and the generic airport data against the on-site measured data, for dewpoint temperature, relative humidity, wind speed and wind direction. The

Table 5

Comparison of model errors and accuracy for different subsets of temperature profile for training.

		1-mth	2-mth	3-mth	4-mth	5-mth	6-mth
MSE	Prediction	50.55	36.84	27.87	22.94	17.79	12.12
	Generic airport	33.17	18.06	21.34	24.50	28.68	29.86

GRU performed best across all parameters except for relative humidity, where the LSTM mode prediction has a slightly lower MSE. As observed, the public airport data has high errors with compared against actual on-site measurements, especially for relative humidity and wind direction. The results validate our hypothesis that using public data obtained online may not best reflect on-site conditions, even if absolute distance between the weather station and the site of interest is not too far off. The GRU model predictions are much more representative of local site conditions compared with data taken off public databases. However, it was also noted that the MSE for wind direction is still relatively high, even after significant improvements. This could likely be due to wind direction being more unpredictable and less reliant on temporal and previous-state information, which the RNNs relies on and excels in utilizing.

Fig. 10a and **10b** shows the histograms for the model prediction for relative humidity and wind speed with the airport data and measured data on-site (ground true). In both cases, the model predictions are more closely aligned with the measured data. The model predictions also do not appear to be affected by the airport data's extreme peak points for both relatively humidity and wind speed. This shows that the RNN model methodology proposed, together with the “sliding window”/look-back in the model architecture, is able to capture seasonal trends while concurrently mimicking actual on-site conditions.

4.4. Discussion – minimum required data

The results presented in the previous sections attest to the effectiveness of the methodology, and illustrates the risks of simply utilizing public airport data for downstream BPS applications – with the potential high variations and uncertainties involved. However, the modeling parameterization and results raise the question of what the required data length is – or rather, how would the duration of the available data affect the predicted results?

The minimum required measured dataset to accurately and model and predict (or morph) generic airport data to obtain representative localized weather is an important factor to consider, especially since consultants, designers or practitioners may not have the luxury of time and resources to continuous track and measure on-site data over a long period.

To further analyze this aspect, six varying training datasets were prepared and modeled. These include training datasets comprising one-month, two-months, three-months, four-months, five months, and six-months of measured data. **Table 5** shows the prediction MSE for each training dataset of various periods. As observed, while a longer period of measured training data naturally leads to better prediction accuracy and lower errors, the inflection point in this study lies around four month, where the error starts to dip lower than the base generic airport file.

Fig. 11a, b and 11c show the model predicted temperature using one-month, three-months, and five-months measured training data length using the methodology in this study. The mode prediction using the one-month measured dataset, while able to capture general seasonal trend over the year, does not accurately align with diurnal variations, especially between June and September. The five-month and six-month models performed significantly better, as observed in greater alignment with ground truth data measured on-site.

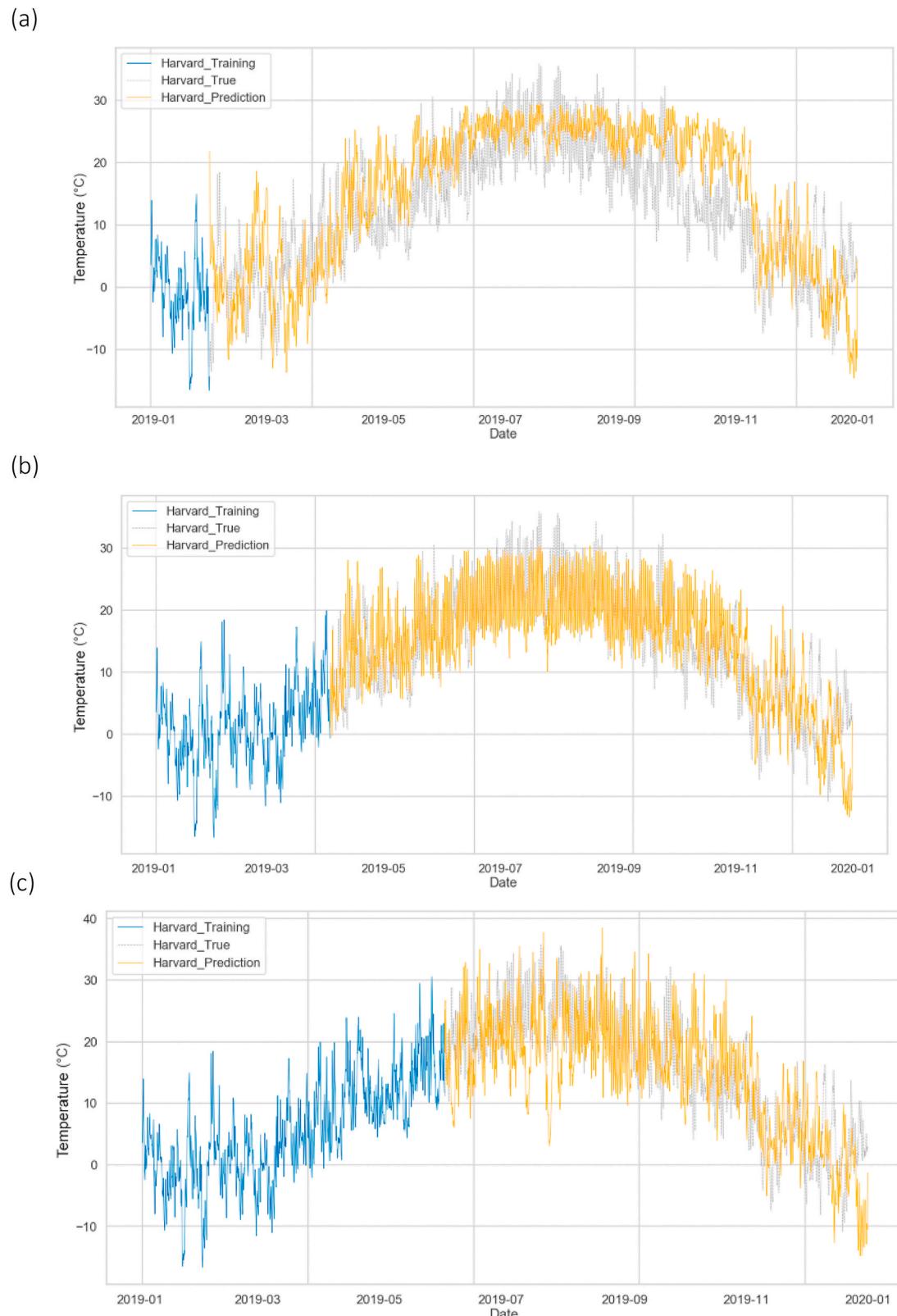


Fig. 11. a (top), b (middle) and c (bottom). Prediction using one-month (top) vs three-month training resolution (middle) vs five-month training resolution (bottom).

4.5. Discussion – impact of proximity (Harvard HouseZero test set)

The next important question is the impact of proximity – i.e., if there are generic/public weather data located geographically closer in distance to the site of interest, what would be the resultant impact? In a

second, independent test, our best performing GRU model is validated against a different dataset obtained from HouseZero, the office of Harvard Center for Green Buildings and Cities. HouseZero was selected as a more localized weather station situated closer to the Harvard Gund Hall. In this part of the study, local on-site measurements were collected using

Table 6

Errors at using airport data (top) and Harvard data (bottom) for HouseZero datasets.

MSE	HouseZero	Harvard	Airport
Airport training	19.43	31.28	54.05
Harvard training	3.03	9.44	53.05

sensors and equipment installed in HouseZero, to train and further validate the model. Two separate sets of training data were used to study the effect of proximity. The first training set comprises the generic airport data used in the previous sections, together with a subset of the measured HouseZero data. The second training set comprises data from the Harvard weather station, which is closer to the site of interest, together with a subset of the measured HouseZero data.

Table 6 shows the prediction MSE for the two different training datasets. The model utilizing the weather data at the Harvard weather station as base training data performed significantly better than the model using the generic airport data for predictions. This exemplifies the hypothesis that one should always use data from weather stations closer in proximity to the actual site/location of interest, if possible. Fig. 12

illustrates the predicted temperature profiles using predict HouseZero's local conditions, where model prediction using training data closer in proximity performed better – as in the case of using data from the Harvard weather station and a subset of the HouseZero data to model subsequent months for HouseZero.

5. Case study on downstream impact on building energy simulation

To understand the impact of utilizing different weather files representing generic or local site-specific conditions, energy simulations using the EnergyPlus engine on the US DOE reference buildings were conducted and compared [52]. In particular, a DOE small office reference building (5000 square feet) was selected as it more closely resembles the Harvard HouseZero's small office/lab environment.

A total of twelve simulations were conducted using three different DOE small office reference building templates (viz. new construction, existing building constructed in or post-1980, and existing building constructed pre-1980). Each reference building template was simulated for energy use over a year with an Airport True weather data file (generic public data from NSRDB taken at Boston Logan Airport), a

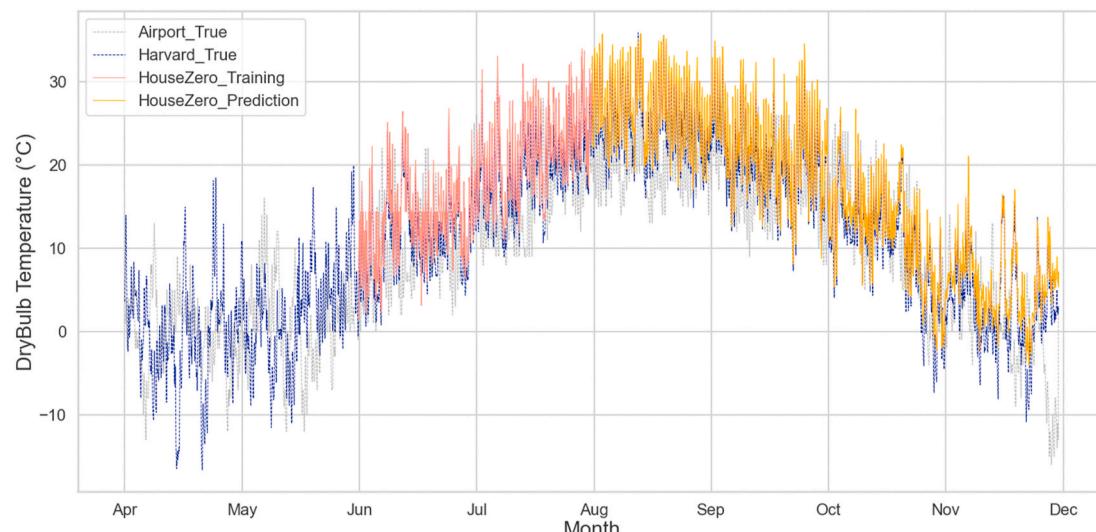


Fig. 12. Predicted temperature profile on HouseZero.

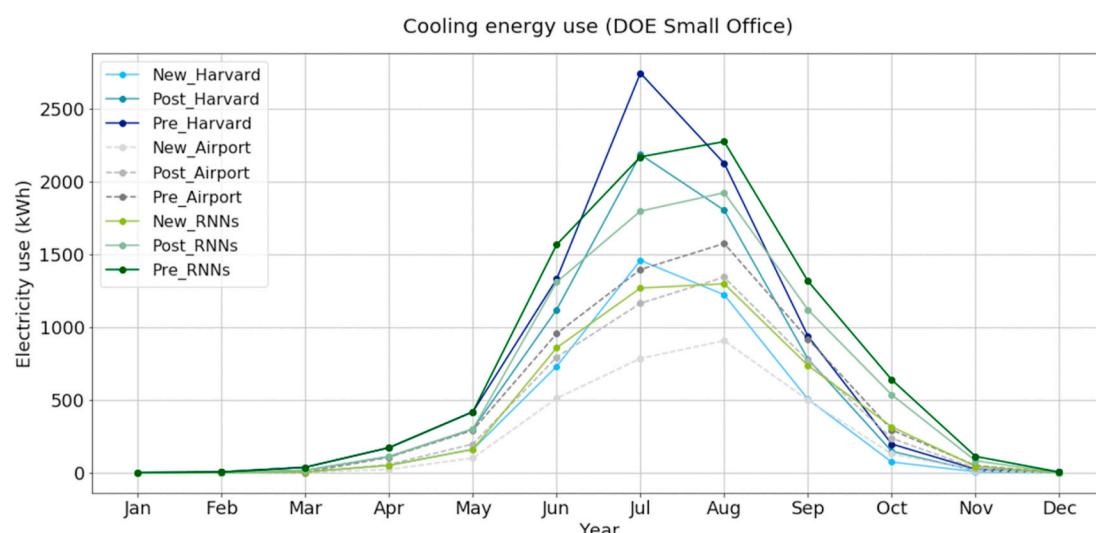


Fig. 13. Cooling energy use for the DOE small office simulation.

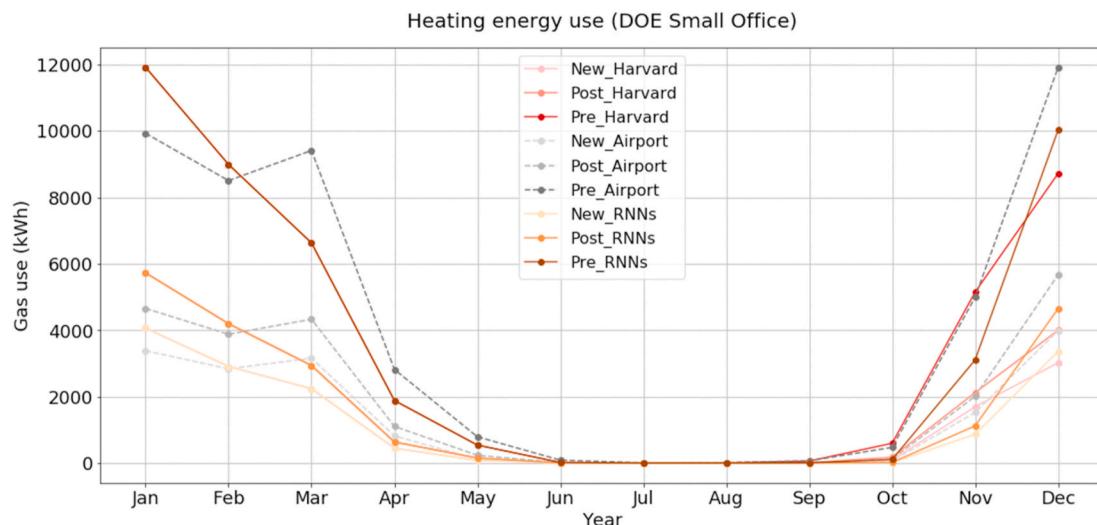


Fig. 14. Heating energy use for the DOE small office simulation.

Harvard True weather data file (from Harvard weather station) and a morphed weather file with predictions from the best RNN model in this study (in this case, the GRU).

The cooling energy use and heating energy use were extracted for each simulation and shown in Figs. 13 and 14. Space conditioning such as cooling and heating in buildings are most likely to be impacted by weather conditions, in contrast with other energy end-uses such as plug-loads, which are less dependent on weather. From the figures, differences in energy uses between the simulation runs with varying weather files can be observed. For example, the pre-1980 small office witnessed over 500 kWh difference in the (peak) month of July for cooling energy use, with different weather files being the only variable in the simulations. Across simulations for cooling energy demand, the RNN models generally exhibit higher energy use. More importantly, it was noted that the modeled energy use using the RNN prediction more closely matched the modeled energy use using the ground truth weather data, for example, in Fig. 14, the modeled energy use for the pre-1980 vintage using RNN prediction data completely overlaps with the modeled energy use using the Harvard measured data, over several months.

Another factor to note is that older buildings typically witnessed higher energy use in these simulations, with new construction being most energy efficient in terms of cooling energy demand. This is consistent with general intuition and the definitions in the DOE reference building templates, since new constructions have better insulation and weatherization properties as compared with buildings pre-1980s. On the other hand, while the pre-1980 reference building has higher heating demands than post-1980 and new construction, simulations done using the Airport True generally exhibit higher heating energy demands than the RNNs morphed weather files. The results from this exercise show that it is important to utilize localized weather data for BPS and energy modeling, especially at the early design stage, since the results will significantly impact and inform not just passive design strategies but active systems like as HVAC sizing and provision of mechanical conditioning systems, etc.

6. Conclusion

In this study, the authors developed a method using RNN to generate accurate weather predictions from generic, publicly available files. The model was subsequently validated, and the predictions and case study illustrated the importance of using localized weather data in building performance simulations. RNN models such as the GRU outperformed standard FFNNs, and the difference in predicted energy use in energy simulation models was substantial which will impact real-world

decision-making as well as early design iteration and systems selection decisions.

In terms of data length, the data length inflection point was found to be around four-months – when the predictions start to outperform generic data. The study also found that proximity was an important factor influencing accuracy, and practitioners should generally use data as close as possible to the actual site of interest.

This method will enhance the ability of architects, designers, and engineers to incorporate representative localized data at the early design stages, and informed performance driven design and decision-making processes. The proposed method also has potential to be utilized in workflows and pipelines studying urban heat island, local conditions, model prediction control, building maintenance, built environment resilience, and climate change. Advantages of the method includes scalability, time-efficiency, and reproducibility using public data.

There are limitations to the proposed model. For example, direct applications in modeling practices are still somewhat constrained by data availability and its associated concerns, such as imputing missing values, incorporating various weather variables etc. Future works can study workflows to impute missing data and/or methods to tackle sparse/low-resolution time series training datasets. Although in this study, hyperparameters of the networks were not determining factors affecting model prediction performance, it will also be worthwhile for future studies to analyze in greater details neural network model architecture and hyperparameters such as the optimizers, loss functions, and activation functions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The material presented in this paper is based in part upon work supported by the Harvard Center for Green Buildings and Cities. The authors declare that there is no conflict of interest in the work.

References

- [1] Y.Q. Ang, Z. Berzolla, C. Reinhart, From concept to application: a review of use cases in urban building energy modeling, *Appl. Energy* 279 (2020).

- [2] M. Eames, T. Kershaw, D. Coley, The appropriate spatial resolution of future weather files for building simulation, *Journal of Building Performance Simulation* 5 (6) (2011) 347–358.
- [3] J. Taylor, M. Davies, A. Mavrogianni, Z. Chalabi, P. Biddulph, E. Oikonomou, P. Das, B. Jones, The relative importance of input weather data for indoor overheating risk assessment in dwellings, *Build. Environ.* 76 (2014) 81–91.
- [4] N. Bell, The Impact of Climate Change on Weather Data Applied to BPS, Sydney, 2018.
- [5] L. Merlier, L. Frayssinet, K. Johannes, F. Kuznik, On the impact of local microclimate on building performance simulation. Part 1: prediction of building external conditions, *Building Simulation*, 2019, pp. 735–746.
- [6] National Renewable Energy Laboratory, EnergyPlus weather data [Online]. Available, <https://www.energypplus.net/weather>, 2020 [Accessed 30 June 2020].
- [7] N.H. Wong, S.K. Jusuf, N.I. Syafii, Y. Chen, H. Norwin, H. Sathyaranayanan, Y. V. Manickavasagam, Evaluation of the impact of the surrounding urban morphology on building energy consumption, *Sol. Energy* 85 (2011) 57–71.
- [8] L. Guan, Preparation of future weather data to study the impact of climate change on buildings, *Build. Environ.* 44 (2009) 793–800.
- [9] A. Jiang, X. Liu, E. Czarnecki, C. Zhang, Hourly weather data project due to climate change for impact assessment on building and infrastructure, *Sustainable Cities and Society* 50 (2019).
- [10] A. Chan, Developing a modified typical meteorological year weather file for Hong Kong taking into account the urban heat island effect, *Build. Environ.* 46 (12) (2011).
- [11] R. Dickinson, B. Brannon, Generating Future Weather Files for Resilience, in: PLEA2016, Los Angeles, 2016.
- [12] X. Dai, J. Liu, X. Zhang, W. Chen, An artificial neural network model using outdoor environmental parameters and residential building characteristics for predicting the nighttime natural ventilation effect, *Build. Environ.* 159 (2019).
- [13] Y. El Mghouchi, E. Chham, E. Zemmouri, A. El Bouardi, Assessment of different combinations of meteorological parameters for predicting daily global solar radiation using artificial neural networks, *Build. Environ.* 149 (2019) 607–622.
- [14] J. Sousa, C. Gorle, Computational urban flow predictions with Bayesian inference: validation with field data, *Build. Environ.* 154 (2019) 13–22.
- [15] Z. Kaseb, M. Hafezi, M. Tahbaz, S. Delfani, A framework for pedestrian-level wind conditions improvement in urban areas: CFD simulation and optimization, *Build. Environ.* 184 (2020).
- [16] S.R. Mohandes, X. Zhang, A. Mahdiyar, A comprehensive review on the application of artificial neural networks in building energy analysis, *Neurocomputing* 340 (2019) 55–75.
- [17] J. Moon, S. Park, S. Rho, A comparative analysis of artificial neural network architectures for building energy consumption forecasting, *Int. J. Distributed Sens. Netw.* 15 (9) (2019).
- [18] S.A. Kalogirou, Artificial neural networks in energy applications in buildings, *Int. J. Low Carbon Technol.* (2020).
- [19] S.A. Kalogirou, Artificial neural networks in renewable energy systems applications: a review, *Renew. Sustain. Energy Rev.* 5 (2001) 373–401.
- [20] I. Yang, M. Yeo, K. Kim, Application of artificial neural network to predict the optimal start time for heating system in building, *Energy Convers. Manag.* (2003) 2791–2809.
- [21] S.A. Kalogirou, M. Eftekhari, D.J. Pinnock, Prediction of Air Flow in a Single-Sided Naturally Ventilated Test Room Using Artificial Neural Networks, Cyprus, 1999.
- [22] G. Georgiou, P. Christodoulides, S. Kalogirou, Implementing artificial neural networks in energy building applications - a review, in: IEEE International Energy Conference (ENERGYCON), 2018.
- [23] A. Conejo, M. Plazas, R. Espinola, A. Molina, Day-ahead electricity price forecasting using the wavelet transform and ARIMA models, *IEEE Trans. Power Syst.* 20 (2) (2005) 1035–1042.
- [24] K. Yurekli, H. Simsek, B. Cemek, S. Karaman, Simulating climatic variables by using stochastic approach, *Build. Environ.* 42 (10) (2007) 3493–3499.
- [25] C. Paoli, C. Voyant, M. Muselli, M.-L. Nivet, Forecasting of preprocessed daily solar radiation time series using neural networks, *Sol. Energy* 84 (12) (2010) 2146–2160.
- [26] Z. Zhang, L. Yeb, H. Qina, Y. Liua, C. Wangc, X. Yud, X. Yina, J. Lia, Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression, *Appl. Energy* 247 (2019) 270–284.
- [27] P. Mathiesen, J. Kleissi, Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States, *Sol. Energy* 85 (5) (2011) 967–977.
- [28] S. Al-Yahyai, Y. Charabi, A. Gastli, Review of the use of numerical weather prediction (NWP) models for wind energy assessment, *Renew. Sustain. Energy Rev.* 14 (9) (2010) 3192–3198.
- [29] S.A. Shamsnia, N. Shahidi, A. Liaghat, A. Sarraf, S.F. Vahdat, Modeling of Weather Parameters Using Stochastic Methods (ARIMA Model)(Case Study: Abadeh Region, Iran), in: 2011 International Conference on Environment and Industrial Innovation, Singapore, 2011.
- [30] M. Rahman, A.H.M. Saiful Islam, S.Y.M. Nadvi, R.M. Rahman, Comparative Study of ANFIS and ARIMA Model for Weather Forecasting in Dhaka, 2013 International Conference on Informatics, Dhaka, Bangladesh, 2013.
- [31] M.H. Alsharif, M.K. Younes, J. Kim, Time series ARIMA model for prediction of daily and monthly average global solar radiation: the case study of seoul, South Korea, *Symmetry* 11 (2) (2019).
- [32] C. Zafra, Y. Angel, E. Torres, ARIMA analysis of the effect of land surface coverage on PM10 concentrations in a high-altitude megacity, *Atmospheric Pollution Research* 8 (4) (2017) 660–668.
- [33] A. Adebiyi, A. Adewumi, C. Ayo, Comparison of ARIMA and artificial neural networks models for stock price prediction, *J. Appl. Math.* (2014).
- [34] M. Siraj-Ud-Doula, Time Series Forecasting: a comparative study of VAR ANN and SVM models, *J. Stat. Econom. Methods* (2019) 21–34.
- [35] J.M. Han, A. Malkawi, K.Z. Gajos, Eabbit 1.0: new environmental analysis software for solar energy representation, in: 16th IBPSA International Conference and Exhibition, 2019.
- [36] S. Siami-Namini, N. Tavakoli, A.S. Namin, A comparison of ARIMA and LSTM in forecasting time series, in: 17th IEEE International Conference on Machine Learning and Applications, 2018.
- [37] D.C. Mocanu, E. Mocanu, P. Stone, P.H. Nguyen, M. Gibescu, A. Liotta, Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science, *Nat. Commun.* (2018) 1–12.
- [38] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: a survey, *Heliyon* 11 (4) (2018).
- [39] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, *Energy Build.* 40 (12) (2008) 2169–2176.
- [40] A.K. Yadav, S.S. Chandell, Solar radiation prediction using Artificial Neural Network techniques: a review, *Renew. Sustain. Energy Rev.* 33 (2014) 772–781.
- [41] J. Taylor, R. Buizza, Neural network load forecasting with weather ensemble predictions, *IEEE Trans. Power Syst.* 17 (3) (2002) 626–632.
- [42] S.S. Baboo, I.K. Shereef, An efficient weather forecasting system using artificial neural network, *Int. J. Environ. Sustain. Dev.* 1 (4) (2010).
- [43] United States Department of Energy, New Construction - Commercial Reference Buildings, US Department of Energy, 2020 [Online]. Available, <https://www.energ.gov/eere/buildings/new-construction-commercial-reference-buildings>. Accessed 9 December 2020.
- [44] National Renewable Energy Laboratory, NSRDB: national solar radiation database [Online]. Available, <https://nsrdb.nrel.gov/about/what-is-the-nsrdb.html>, 2020. Accessed 23 June 2020.
- [45] National Renewable Energy Laboratory, Typical meteorological year (TMY) [Online]. Available, <https://nsrdb.nrel.gov/about/tmy>, 2020. Accessed 9 June 2020.
- [46] R. Bird, R. Hulstrom, Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces, Solar Energy Research Institute, 1981.
- [47] J. Schmidhuber, Deep learning in neural networks: an overview, *Neural Network* 61 (2015) 85–117.
- [48] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, in: The 3rd International Conference for Learning Representations, San Diego, 2015.
- [49] J.L. Elman, Finding structure in time, *Cognit. Sci.* 14 (1990) 179–211.
- [50] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [51] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, in: NIPS 2014 Deep Learning and Representation Learning, 2014.
- [52] US Department of Energy, US department of energy commercial reference Buildings' [Online]. Available, <https://www.energy.gov/eere/buildings/commercial-reference-buildings>, 2020. Accessed 9 May 2020.