July 2018

A Data-driven Approach towards Integration of Microclimate Conditions for Predicting Building Energy Performance

Soumik Sarkar
Department of Mechanical Engineering, Iowa State University, United States of America, soumiks@iastate.edu

Linjiang Wu
Department of Mechanical Engineering, Iowa State University, United States of America, lwu26@iastate.edu

Venkatesh Chinde
Department of Mechanical Engineering, Iowa State University, United States of America, vchinde@iastate.edu

Himanshu Sharma
Department of Mechanical Engineering, Iowa State University, United States of America, hsharma@iastate.edu

Ulrike Passe
Department of Architecture, Iowa State University, United States of America, upasse@iastate.edu

A Data-driven Approach towards Integration of Microclimate Conditions for Predicting Building Energy Performance

Linjiang Wu[1], Venkatesh Chinde[1], Himanshu Sharma[1], Ulrike Passe[2], Soumik Sarkar[1*]

[1] Department of Mechanical Engineering, Iowa State University,
Ames, Iowa, USA
[2] Department of Architecture, Iowa State University,
Ames, Iowa, USA
{lwu26,vchinde,hsharma,upasse,soumiks}@iastate.edu

* Corresponding Author

ABSTRACT

Energy consumption in buildings is a major part of the overall energy usage in the United States and across the world. Energy performance of buildings is primarily affected by the heat exchange between the building outer skin and the surrounding environment. Building energy simulation (BES) tools are capable of predicting energy usage with variable degree of accuracy using the building geometry, construction and weather data. In this regard, it is a common practice in BES tools to include boundary conditions of the building shell based on the local weather station. However, to account for accurate building energy consumption, especially in urban environments with lots of anthropogenic heat source it is necessary to consider the microclimate around the building. These conditions are influenced by the immediate environment such as surrounding buildings, hard surfaces and trees. However, deployment of sensors to monitor microclimate of a building can be quite expensive and hence, not scalable. Therefore, a model to predict the microclimate information is essential to provide a more reasonable weather information for the BES tools, and hence, to predict energy consumption in buildings more accurately. In this work, we propose a scalable, computationally inexpensive data-driven approach for predicting microclimate information (e.g., temperature) under multiple weather conditions. We demonstrate that such a framework can be implemented based on machine learning techniques such as spatiotemporal pattern network (STPN) and neural networks (NNs).

1. INTRODUCTION

Buildings major energy consuming entities and have a total share of 40% energy usage in the United States in 2016, as reported by the U.S. Energy Information Administration (EIA). Ambient weather conditions are the main factors affecting the heat exchange from surroundings to the inside of a building. Commonly, the surrounding environment is considered to be from meteorological weather data (Meek and Hatfield, 1994) or local micro weather station and is used in various building energy simulation (BES) tools for energy consumption or energy costs analysis (Swan and Ugursal, 2009). BES usually uses TMY data sets, which are derived from airport weather stations, which do not resemble the specific microclimate of the actual location of a building. Given the fact that the microclimatic conditions around the building are affected by the neighboring buildings, vegetation, air flow patterns etc. These factors also affect the energy performance of the building. The local weather station climatic data, since recorded in the far-field, will therefore not account for the local climatic information around the building. Hence, it is important to include microclimate information to have a better understanding of the building energy performance. The significance and the use of microclimatic information are shown for various studies, such as in urban heat islands for peak building energy use and energy savings in (Bhiwapurkar, 2015). Further, microclimatic conditions under urban environment can also significantly affect the urban energy management. They are important in the development of cities or certification of buildings based on their energy performance such as United States Green Building Council's (USGBC) Leadership in Energy and Environmental Design (LEED) (Scofield and Oberlin, 2009), Energy Star (Boyd et al., 2008) etc. The major microclimatic variables are average surrounding temperature, wind, humidity, solar radiation, and vegetation. These variables are usually monitored around the building. For such monitoring, a spatiotemporal relationship with the local weather data can be constructed as a model to predict the microclimatic condition.

Various efforts are made in the area of integrating the micro climatic conditions for predicting the energy performance of buildings, a few are listed below: 1) A procedure for coupling microclimate model ENVI-met with EnergyPlus for quantitative analysis of building energy performance by (Yang et al., 2012). 2) A simulation platform relying on both computational fluid dynamics (CFD) and themoradiative simulation codes by (Bouyer et al., 2011) to measure the impact of urbanization and planning strategies on energy demand. In the work they also did senstivity analysis for finding important microclimatic variables affecting building energy consumption. 3) Coupling CFD and BES is shown by (Dorer et al., 2013) for analyzing the impact of urban microclimate (UMC) on space heating and cooling energy demand. In the work by Yuming et.al (Sun et al., 2014), a framework is proposed to compare the microclimate variables computed from standard model in BES programs to those with meso scale model which takes into account more detailed specifications of urban form. Further, the vegetation effect of the energy performance of the building is also studied in detail by (Huang et al., 1987). The occupancy which act as an important parameter in building design, in association with microclimate can be looked in the work by (Kalvelage et al., 2015). In the current study, we

Figure 1: Interlock house located near Ottumwa, Iowa, used as a testbed in the study. It is equipped with energy and microclimate monitoring sensors.

consider a real-time monitored and operated building at Honey Creek Resort State Park in Iowa. The availability of data pertaining to the key microclimatic variables for the building makes it possible to apply data driven approaches to have a deeper understanding of patterns and process in building energy systems. This motivates us to use data driven approaches for microclimatic modeling and analyze its impact on the energy performance. The main goal of this work is to present a proof of concept for estimating microclimate from an ambient weather information using data-driven approaches (Jiang et al., 2017; Sarkar and Srivastav, 2016).
Contributions: The major contributions of this work are: (i) the use of data-driven models in estimating microclimate variables from ambient weather data, (ii) demonstrating the efficacy of using estimated microclimate information for energy consumption prediction as compared to the ambient weather data, and (iii) validating the prediction and regression models using the real data from Interlock house test bed (CBER, 2016b) near Ottumwa, Iowa. Figure-1 shows the pictorial representation of the test bed considered for the data collection in the study.

2. DATA-DRIVEN METHODOLOGIES

2.1 Feature Extraction with Spatiotemporal Pattern Network

2.1.1 Spatiotemporal Pattern Network (STPN): STPN framework relies mainly on symbolic dynamic filtering (SDF) technique and is found to be extremely useful in extracting key features from time series data (Rao et al., 2009). The core idea lies on the construction of symbol sequence from the time series data and identifying patterns in these sequences through the construction of probabilistic finite-state machine. These transition matrices capture the patterns in the

underlying data. For more details on the approach refer to (Sarkar et al., 2014). The discretization or symbolization process is noted as partitioning (Rao et al., 2009). Let $\mathcal{X}$ denote a set of partitioning functions, $\mathcal{X} : X(t) \rightarrow S$, that transforms a general dynamic system (time series $X(t)$) into a symbol sequence S with an alphabet set Σ. This study uses simple maximum entropy partitioning (MEP). The D-Markov machine (Markov chain of order D (depth) is essentially a probabilistic finite state automaton (PFSA) that can be described by states (representing various parts of the data space) and probabilistic transitions among them that can be learn from data. Related definitions of deterministic finite state automaton (DFSA), PFSA, D-Markov machine, xD-Markov machine and the learning schemes can be found in (Sarkar et al., 2014; Liu et al., 2016). The definition of STPN is given as follows:
Definition: A PFSA based STPN is a 4-tuple $W_D \equiv (Q^a, \Sigma^b, \Pi^{ab}, \Lambda^{ab})$: (a,b denote nodes of the STPN)

1. $Q^a = \{q_1, q_2, \cdots, q_{|Q^a|}\}$ is the state set corresponding to symbol sequences S^a;
2. $\Sigma^b = \{\sigma_0, \cdots, \sigma_{|\Sigma^b|-1}\}$ is the alphabet set of symbol sequences S^b.
3. Π^{ab} is the symbol generation matrix of size $|Q^a| \times |\Sigma^b|$, the ij^{th} element of Π^{ab} denotes the probability of finding the symbol σ_j in the symbol string s^b while making a transition from the state q_i in the symbol sequence S^a; while self-symbol generation matrices are called atomic patterns (APs) i.e., when $a = b$, cross-symbol generation matrices are called relational patterns (RPs) i.e., when $a \neq b$.
4. Λ^{ab} denotes a metric that can represent the importance of the learnt pattern for $a \rightarrow b$ which is a function of Π^{ab}.

A schematic of the procedure of STPN is shown in Fig. 2

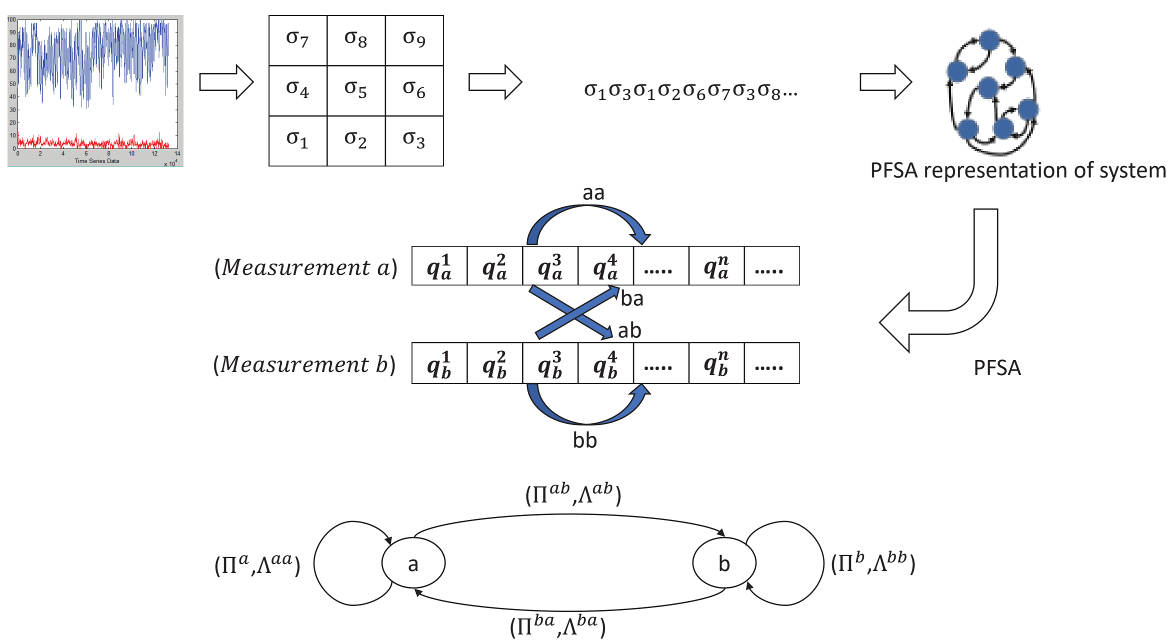


Figure 2: Formulation of STPN with multiple time-series (nodes in graphical modeling). It extracts atomic patterns (AP) and relational patterns (RP) with *D*-Markov machine and *xD*-Markov machine respectively and the depth *D*=1.

2.1.2 Information Based Metric: Based on the above definition of STPN, we can use the atomic/relational patterns to interpret the Granger-causal dependencies among the sensors. In this context, information theoretic criteria have been widely used, e.g., transfer entropy (Wibral et al., 2011) and mutual information (Sarkar et al., 2014; Solo, 2008). In this paper, the concept of mutual information is applied for representing Λ^{ab} of the patterns (APs & RPs). The definition of Λ^{ab} is as follows.

$$\Lambda^{ab} \triangleq I^{ab} = I(q^b_{k+1}; q^a_{k+1} = H(q^b_{k+1}) - H(q^b_{k+1}|q^a_k) \tag{1}$$

where, I^{ab} is the mutual information of pattern (a, b), H is the conditional entropy defined as follows,

$$H(q^b_{k+1}) = -\sum_{i=1}^{Q^b} P(q^b_{k+1} = i) \log_2 P(q^b_k = i) \tag{2}$$

$$H(q^b_{k+1}|q^a_k) = \sum_{i=1}^{Q^a} P(q^a_k = i) H(q^b_{k+1}|q^a_k = i) \tag{3}$$

$$H(q^b_{k+1}|q^a_k = i) = -\sum_{j=1}^{Q^b} P(q^b_{k+1} = j|q^a_k = i) \cdot \log_2 P(q^b_{k+1} = j|q^a_k = i) \tag{4}$$

Detailed description of mutual information theoretic causality metric in the context of APs and RPs can be found in (Sarkar et al., 2014).

2.2 STPN Framework for Multi-variables

In this paper we propose a model with multiple time-series inputs and a single prediction target as shown in Fig. 3. Suppose $X = \{X^{\mathbb{A}}(t), t \in \mathbb{N}, \mathbb{A} = 1, 2, .., n\}$ represents the multivariate time-series data from city weather station, n is the number of time-series and $Y = Y(t), t \in \mathbb{N}$ denotes the microclimatic information (here is temperature) collected by microclimate station. And for the time-series X, Y we have symbol sequences $\boldsymbol{\sigma} = \{\sigma^{\mathbb{A}}(t)\}$ and $\boldsymbol{\mu} = \{\mu(t)\}$ after data processing and partitioning. To generate the dependencies we form a joint symbol sequence for X, $\sigma^{\mathbb{J}} = \sigma^1(t) \oplus ... \oplus \sigma^n(t)$. Here, the joint symbol space is created by summing the individual symbol spaces directly. For example, $\sigma^a \oplus \sigma^b$ is the product space of σ^a and σ^b. Then based on the previous definition of STPN we can get the state sequences Φ and Ψ respect to $\sigma^{\mathbb{J}}$ and $\boldsymbol{\mu}$. The training step of STPN is to compute the transition matrix $\boldsymbol{\omega}(\Phi, \mu)$ from the states in Φ to the symbols in μ using a frequentist's technique(e.g. counting the number of occurrences). For example, we can compute the probability of the state Φ_m to the symbol μ_i by $Pr(\varphi_m, \mu_i) = N_{mi}/N_m$, where N_{mi} is the number of times that the symbol $\mu_i \in \boldsymbol{\mu}$ is emanated after the state $\varphi \in \Phi$, and here $N_m \triangleq \sum_{i=1}^{n} N_{mi}$. In the model, the true values represented by

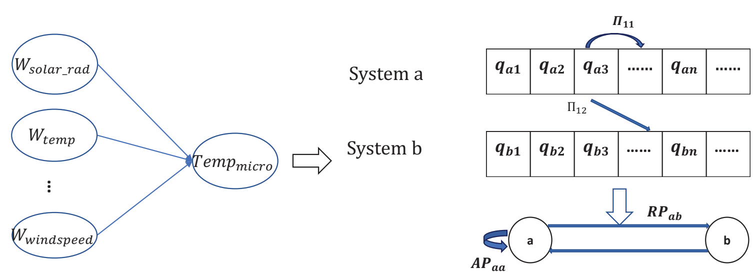


Figure 3: STPN framework for microclimate temperature prediction based on local weather station data: solar radiation, temperature, relative humidity, and wind speed.

$\boldsymbol{\mu}$ can be computed from the training data, and noted as $E(X|\mu = n), n = 1, 2, ...$ In the prediction model, we only have the time-series X and the microclimatic temperature time-series $\hat{Y}$ is assumed unknown and used for the verification. And similarly, we can get the symbol sequences $\hat{\sigma}$ and joint symbol sequence $\sigma^{\mathbb{J}}$ as well as state sequences. Thus, the microclimatic temperature Y_t is obtained as:

$$Y_t = \sum_{i=1}^{n} Pr(\mu|\hat{\varphi}(t)) \times E(Y|\mu = i) \tag{5}$$

2.3 Neural Networks

Neural networks (NNs) are inspired by the biological operation of specialized cells called neurons. In a human body a neuron is a cell which receive several inputs from outside process to get activated. Based on the activation level the neuron gets stimulated and produces the outputs. Further, each input and output path of the neuron can have there own

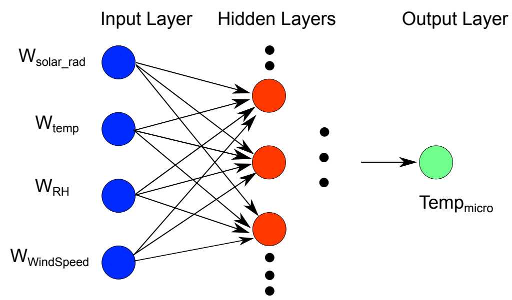


Figure 4: Schematic of a Neural network trained in the current study for predicting the microclimate variables based on the ambient weather data

strength and can differ from one another. A neural network tries to replicate this biological system by a set of weighted graphs where a neuron is represented as a node. The node receive inputs, and processes their sum with its *activation function* φ and passes the results to the nodes further in the graph. Mathematically this can be written as

$$\varphi(\mathbf{w}^{\mathbf{T}}\mathbf{a}) = \frac{1}{1 + \mathbf{exp}(-\mathbf{w}^{\mathbf{T}}\mathbf{a})} \tag{6}$$

The sigmoid activation function is used at each node in the network. A schematic of the neural network is shown in Fig. 4. The nodes are chained together to form the hidden layer in the network, there can be multiple set of hidden layer which passes the output as a input to the next hidden layer to reach the final output. The goals is to train a neural network based on the data so that it can be used for prediction to unseen input data.

Recent success of deep learning has demonstrated the enormous possibilities of wider and deeper neural networks for learning complex functions without feature hand-crafting (Lore et al., 2015). However, in the present work it is difficult to build a large neural network due to the small volume of data. A regular neural network developed here was trained using time-series data of the microclimate and weather station data, sampled at the same time-interval. The network can be used for classification as well as regression and in the current study we develop regression model, for which we define the loss function as follows,

$$L(\mathbf{w}) = \sum_{\mathbf{i}} \mathrm{f}(\mathbf{x_i}, \mathbf{w}) - \mathbf{y_i})^2 \tag{7}$$

where, $\mathrm{f}(\mathbf{x_i}, \mathbf{w})$ is the regression function and we must minimize $L(\mathbf{w})$. This equation has the closed form solution for general least squares, but here we use the gradient based method, ADAM (Kingma and Ba, 2014) to minimize loss. In general the gradient of the loss gradient $\nabla_{\mathbf{w}} L(\mathbf{w})$ is to minimize the overall error on the training data. The first step is to derive the loss with respect to a particular weight $w_{j\to k}$ (This is the weight of the edge connecting node j to node k) for the general setting we write this as follows:

$$\begin{aligned}\frac{\partial}{\partial w_{j\to k}} L(\mathbf{w}) &= \frac{\partial}{\partial w_{j\to k}} \sum_i (\mathrm{f}(\mathbf{x_i}, \mathbf{w}) - \mathbf{y_i})^2 \\ &= \sum_i \frac{\partial}{\partial w_{j\to k}} (\mathrm{f}(\mathbf{x_i}, \mathbf{w}) - \mathbf{y_i})^2 \\ &= \sum_i 2(\mathrm{f}(\mathbf{x_i}, \mathbf{w}) - y_i) \frac{\partial}{\partial w_{j\to k}} \mathrm{f}(\mathbf{x_i}, \mathbf{w})\end{aligned} \tag{8}$$

This general process is modified in ADAM by the stochastic gradient finding method (more details in (Kingma and Ba, 2014)). Following the discussed procedure, the network is trained to predict the microclimate data as a dependent variable by using the weather station data as input. The trained model is then used to input the conditions to predict the energy performance of the building. We use the mean square error (MSE) to calculate the accuracy of the model, as a widely accepted metric for regression problems. The quality and quantity of the data affect the model performance and with more time-points the MSE can be reduced further for the current problem. Next, we use the model for predicting the microclimatic data, which then can be used for energy predictions.

3. TEST SCHEMATIC, DATA PREPARATION, AND MODEL SETUP

3.1 Test Schematic

The testing building, Interlock House, with the nearest local weather station in Ottumwa, Iowa is shown in Fig. 1. The building is equipped with sensors for thermal performance and energy consumption monitoring and collecting the surrounding microclimatic information (i.e., temperature, humidity, wind speed etc.). A schematic of the house (CBER, 2016a) is shown in Fig 5. The house has been designed as a passive and active solar powered house and is oriented south for maximum solar exposure. The house is heated by evacuated solar tubes and a radiant floor system. Based on these systems setting, the heating load is calculated using the heat gain model 9 for the radiant floor.

$$H_{load} = (T_{in} - T_{out}) \cdot C_w \cdot V_w \cdot \rho_w \tag{9}$$

where H_{load} is the heating load (energy consumption in winter), T_{in} is the temperature of the flowing in water, T_{out} is the temperature of the flowing out water, C_w is the specific heat of water, V_w is the volume and ρ_w is the density of water. give the microclimate data of the Interlock House and the associated ambient weather data, a data driven

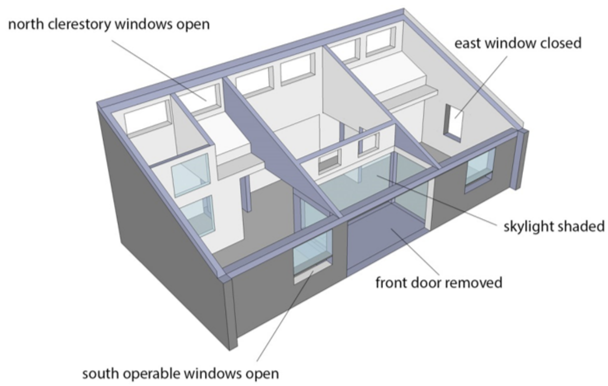


Figure 5: The building used in the study (Interlock house), the yearly energy data and microclimatic data is recorded for the building.

approach is proposed to predict the microclimatic conditions (temperature). The predicted temperature then in turn is used for energy consumption prediction based on a regression analysis.

3.2 Data Preparation and Model Setup

The workflow chart for the proposed data-driven energy analysis is shown in Fig.6. Typical meteorological data (TMY3) from Ottumwa city is considered for the weather station data. The key variables used for the regression analysis are outside air temperature (T_a), relative humidity (RH), global horizontal irradiation (E_e), and wind speed (V_w). For the simulations, the weather station data X is divided into two parts, summer X_1 (July-September 2015) and winter X_2 (December2015-Februaury2016). The corresponding microclimate data is also stored as dependent variable Y. The data sets are taken at 1-min interval resulting into 129,500 data points for each season. The complete data set is then further split into training ($X_{tr} = 85\%X$, $Y_{tr} = 85\%Y$) and validation ($X_{ts} = 15\%X$, $Y_{ts} = 15\%Y$) for the use in the data-driven approaches. The simplified model for both data-driven methodologies are:

$$Y_{temp} = f(X(t)|X(\tau : t-1), Y_{temp}(\tau : t-1)) \tag{10}$$

or:

$$\begin{aligned} T_{micro}(t) = {} & f(T_a(t), RH(t), E_e(t), V_w(t)|T_a(\tau : t-1), \\ & RH(\tau : t-1), E_e(\tau : t-1), V_w(\tau : t-1), \\ & T_{micro}(\tau : t-1)) \end{aligned} \tag{11}$$

where t is the target time and τ is the training start time.
The predictions of microclimate variables obtained using the data driven approaches (STPN and NN) are used for energy prediction analysis. To make the energy prediction, we use autoregressive–moving-average (ARMAX) (Yang et al., 1996) regression model using the four regression variables (RH,T_a,E_e,V_w). These models build the relationship between inputs and the outputs of the system using measurements and also help in understanding the energy characteristics of the building.

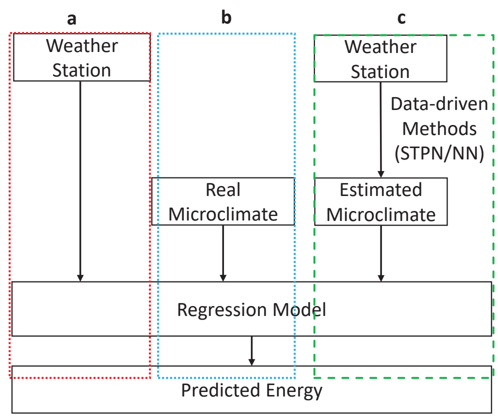


Figure 6: Three different procedures of predicting energy consumption in buildings based on regression model a) weather station data collected in Ottumwa b) real microclimate data collected around the testbed c) estimated microclimate data from weather station in Ottumwa.

4. RESULTS AND DISCUSSION

In this study, we validate our model for two different seasons, summer and winter for analysing the affect of climatic conditions on the prediction performance.

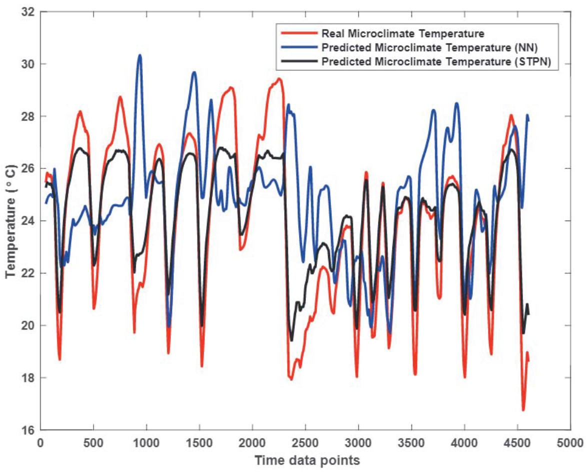
4.1 Summer Season

The procedure for using STPN and NN methods are detailed in Section 2. For the summer season (mainly focus on cooling), the data from the months of July, August, September of 2015 are considered. Key variables from the ambient weather data are used for prediction of microclimate temperature. The testbed has outdoor AC units which are used for providing cooling during the summer season. The data has been preprocessed to filter any outliers (i.e. manually shut off) in the data set which affects the prediction results. The temperature prediction results using both the methods in comparison to actual temperature is shown in left in Fig. 7 (a). The prediction results using both the methods captures the trend associated with the data and STPN performs better compared to NN. For the case of STPN, the error in the peaks can be improved by partitioning the data into more finer bins. Typically, NNs perform well while trained with a large data set (by avoiding overfitting). Therefore, in this case the slightly lesser accuracy compared to STPN can be attributed to the data quantity as well as lack of hyperparameter tuning.

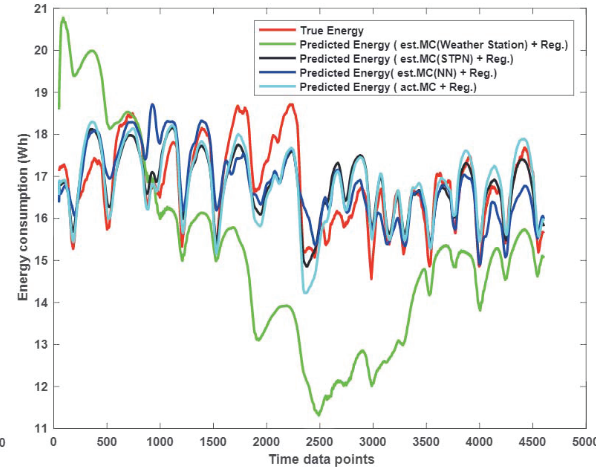
These temperature prediction results are used in the regression model for energy prediction. A comparison plot of energy prediction using the prediction results from the proposed methods along with energy predicted from weather station data is shown in right in Fig. 7 (b). The results indicate that the energy prediction from the data-driven methods has higher accuracy (i.e. fit better) compared to the energy predicted from ambient weather station data. This clearly signifies the importance of microclimate information on the building energy consumption prediction analysis. This study will be further extended in future to include building specific variables (e.g., building architecture, materials, vegetation around the building) for improving the prediction performance.

4.2 Winter Season

For the winter season the months of December 2015, Januray & February of 2016 are considered. The microclimate

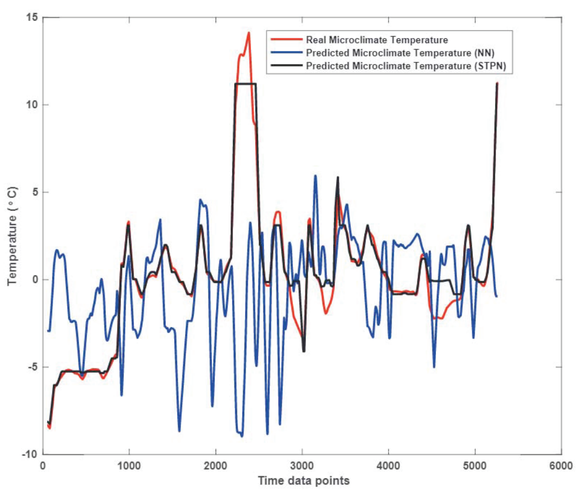


a: Comparisons of real microclimate temperature and the predicted microclimate temperature using STPN and Neural Networks methods for summer season

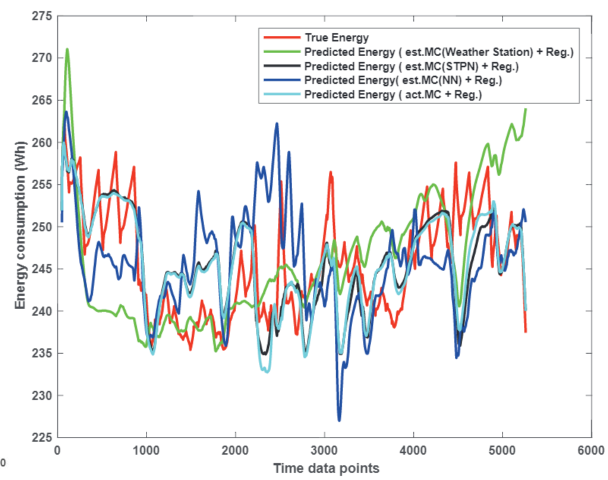


b: Comparison of energy consumption prediction using predictions from data-driven approaches, ambient weather data, actual microclimate data and true energy consumption. Here, est.MC represents predicted microclimate data, Reg. means the regression model and act.MC means the actual microclimate data

Figure 7: Predicted microclimate temperature and energy consumption, and comparisons with the real data in Summer



a: Comparisons of real microclimate temperature and the predicted microclimate temperature using STPN and Neural Networks methods for winter season



b: Comparison of energy consumption prediction using predictions from data-driven approaches, ambient weather data, actual microclimate data and true energy consumption.

Figure 8: Predicted microclimate temperature and energy consumption, and comparisons with the real data in winter

temperature prediction results are shown in Fig. 8 (a). The prediction result of STPN method has improved performance over the NN method. Similarly the energy comparison result is shown in Fig. 8 (b). The energy consumption results also suggests that microclimate information is crucial in predicting the building heating energy consumption. The heating energy consumption is mainly affected by the heat transfer between inside and outside of the building. From the Fig. 8 (a) we can see the temperature in winter changes in a larger range, which is more complicated for data-driven model to learn the patterns. The results can be further improved by including additional variables which captures the heat loss from the building which in turn affects the heating energy consumption.

For a multivariate time series, the proposed STPN framework captures the informative patterns, which results in predicting microclimate data. The case studies in this paper show that, although microclimatic information is not always available for every building, data-driven techniques help in obtaining a better energy consumption prediction. Also, the data quality and quantity can significantly improve the microclimatic prediction and hence the energy prediction results. Overall, both data-driven methodologies contribute great effects on the prediction of building energy consumption, especially, STPN method performs better in comparison with NN method due to the limited data points. Using variants of NN for prediction analysis needs to be investigated further. Energy prediction comparison results are shown in Table 1. The mean square error using predictions based on STPN method is close to the real microclimate data which suggests that the data-driven models are sufficient for predicting microclimate information which replaces the need for sensing cost in obtaining such information.

Table 1: Comparison of Energy Prediction Using Different Weather Conditions

Mean square error with different types of weather data				
Season	Ambient weather station data	Real microclimatic data	Predicted microclimatic data using data-driven models	
Summer	7.33	0.54	STPN	0.56
			NNs	0.70
Winter	88.07	41.45	STPN	44.04
			NNs	71.64

5. CONCLUSION

This paper presents a proof of concept study to demonstrate the efficacy of using microclimate information for building energy consumption prediction as compared to ambient weather data. In this regard, data-driven approaches can be extremely effective for prediction of microclimate variables and hence, can help avoid expensive microclimate sensor deployment. Quantitative results (using energy prediction accuracy metrics such as mean square error) show that the energy prediction from the data-driven approaches are sufficiently close to the actual microclimate data. These results illustrate that the predicted microclimate information based on data-driven approaches is a reliable and low-cost way to predict building energy consumption. Future work will focus on: (1) identifying the building specific key variables (i.e., building architecture, materials, vegetation around the building etc.) that affect the building microclimate and (2) applying the proposed concept to different types of buildings (i.e., urban and rural buildings).

REFERENCES

Bhiwapurkar, P. (2015). Urban climate change impacts on building heating and cooling energy demand. *of Architectural Research*, page 270.

Bouyer, J., Inard, C., and Musy, M. (2011). Microclimatic coupling as a solution to improve building energy simulation in an urban context. *Energy and Buildings*, 43(7):1549–1559.

Boyd, G., Dutrow, E., and Tunnessen, W. (2008). The evolution of the energy star® energy performance indicator for benchmarking industrial plant manufacturing energy use. *Journal of cleaner production*, 16(6):709–715.

CBER (Center for Building Energy Research, Interlock house 2013-2016b). *https://www.cber.iastate.edu/projects/interlock-house*.

CBER (Center for Building Energy Research, Interlock house feature 2016a). *http://interlockhousedata.com/sample-page/*.

Dorer, V., Allegrini, J., Orehounig, K., Moonen, P., Upadhyay, G., Kämpf, J., and Carmeliet, J. (2013). Modelling the urban microclimate and its impact on the energy demand of buildings and building clusters. *Proceedings of BS*, 2013:3483–3489.

Huang, Y., Akbari, H., Taha, H., and Rosenfeld, A. H. (1987). The potential of vegetation in reducing summer cooling loads in residential buildings. *Journal of climate and Applied Meteorology*, 26(9):1103–1116.

Jiang, Z., Liu, C., Akintayo, A., Henze, G., and Sarkar, S. (2017). Energy prediction using spatiotemporal pattern networks. *arXiv preprint arXiv:1702.01125*.

Kalvelage, K., Passe, U., Krejci, C., and Dorneich, M. C. (2015). Incorporation of future building operating conditions into the modeling of building–microclimate interaction: A feasibility approach. In *Sustainable Human–Building Ecosystems*, pages 150–158.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Liu, C., Ghosal, S., Jiang, Z., and Sarkar, S. (2016). An unsupervised spatiotemporal graphical modeling approach to anomaly detection in distributed cps. In *Cyber-Physical Systems (ICCPS), 2016 ACM/IEEE 7th International Conference on*, pages 1–10. IEEE.

Lore, K. G., Stoecklein, D., Davies, M., Ganapathysubramanian, B., and Sarkar, S. (2015). Hierarchical feature extraction for efficient design of microfluidic flow patterns. In *NIPS workshop on Feature Extraction: Modern Questions and Challenges*, pages 213–225.

Meek, D. and Hatfield, J. (1994). Data quality checking for single station meteorological databases. *Agricultural and Forest Meteorology*, 69(1-2):85–109.

Rao, C., Ray, A., Sarkar, S., and Yasar, M. (2009). Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal, Image and Video Processing*, 3(2):101–114.

Sarkar, S., Sarkar, S., Virani, N., Ray, A., and Yasar, M. (2014). Sensor fusion for fault detection and classification in distributed physical processes. *Frontiers in Robotics and AI*, 1:16.

Sarkar, S. and Srivastav, A. (2016). A composite discretization scheme for symbolic identification of complex systems. *Signal Processing*, 125:156–170.

Scofield, J. H. and Oberlin, O. (2009). A re-examination of the nbi leed building energy consumption study. In *International energy program evaluation conference, Portland, OR*.

Solo, V. (2008). On causality and mutual information. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 4939–4944. IEEE.

Sun, Y., Heo, Y., Tan, M., Xie, H., Jeff Wu, C., and Augenbroe, G. (2014). Uncertainty quantification of microclimate variables in building energy models. *Journal of Building Performance Simulation*, 7(1):17–32.

Swan, L. G. and Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and sustainable energy reviews*, 13(8):1819–1835.

Wibral, M., Rahm, B., Rieder, M., Lindner, M., Vicente, R., and Kaiser, J. (2011). Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. *Progress in biophysics and molecular biology*, 105(1):80–97.

Yang, H.-T., Huang, C.-M., and Huang, C.-L. (1996). Identification of armax model for short term load forecasting: an evolutionary programming approach. *IEEE Transactions on Power Systems*, 11(1):403–408.

Yang, X., Zhao, L., Bruse, M., and Meng, Q. (2012). An integrated simulation method for building energy performance assessment in urban environments. *Energy and Buildings*, 54:243–251.

6. ACKNOWLEDGEMENT

This Interlock House community lab and data set is based upon work supported by the National Science Foundation under Grant Number EPS-1101284. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.