

PAPER • OPEN ACCESS

Energy use prediction with information theory and machine learning technique

To cite this article: YW Tong *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **291** 012031

View the [article online](#) for updates and enhancements.

Energy use prediction with information theory and machine learning technique

YW Tong¹, WY Yang², and DL Zhan³

¹ School of economics and management, Nanchang University, Nanchang, China

² College of automation, Shenyang Aerospace University, Shenyang, Liaoning, China

³ Department of Mathematics, Sichuan University, Chengdu, Sichuan, China

E-mail: ellabc369@gmail.com

Abstract: Appliances energy consumption plays an increasingly important role in the overall building electric energy consumption and its temporal trending. However, predicting appliances energy consumption is complicated by lack of causal understanding of the appliances energy use as well as too many potential predictors that might be relevant to the appliances energy use. In this study, we apply information theory and advance machine learning neural network technique to first rank the importance of potential drivers that dominate appliances energy consumption and secondly model the temporal evolution of appliances energy consumption with a restricted set of environmental predictors. Our results showed that temperature and humidity were the two most important environmental drivers in the house appliances energy consumption modeling. Furthermore, using those environmental drivers, the machine learning model was able to accurately capture the temporal dynamics of appliances energy consumption.

1. Introduction

Appliances energy consumption takes up about 20% - 30% of the building electric energy consumption [1], thus significantly contribute to the daily dynamics as well as long term trend of regional energy consumption. For example, a 10% increase in the domestic building electricity use was resulted from appliances energy use in UK [2]. Hence, improving our understanding of the temporal changes, drivers, and effects of appliances energy consumption in buildings is critical important. Historically, a variety of approaches have been proposed to understand the appliances energy consumption. For example, EenergyPlus model simulates high frequency building energy use with the utility bills [3]. Data-driven machine learning techniques have also been widely employed. Markov Chain Monte Carlo technique was successfully used to estimate building occupancy and potential energy demand and consumption [4]. A multiple linear regression model was established to explain the variability of monthly energy consumption with environmental drivers [5] and showed that over 90% of energy consumption variability could be explained. Despite a vast amount of efforts have been allocated to predict appliances energy consumption, it is still challenging to accurately predict high-frequency dynamics of appliances energy use, given the complexity and non-linearity of the predictive variable itself.

In addition, it is still under debates about the dominant drivers of building appliances energy use. Previous work indicated that a wide spectrum of factors including socioeconomic factors, dwelling



characteristics, climate conditions, region of interest, presence of double pane windows, energy efficient light fixture, number of refrigerator and entertainment devices, and so on [6].

In this work, our objective is two-fold. We first aim to understand the causal relationship between environmental drivers and appliances energy consumption and to verify whether exterior or interior drivers can best predict future energy consumption. Secondly, we aim to model the high frequency of appliances energy use temporal dynamics with advanced machine learning technique.

2. Methodology

2.1 Data description

We used observational data of house appliances energy use as well as detailed interior and exterior environmental drivers from an energy efficient house located in Stambruges, Belgium [7]. Appliances energy consumption (in Wh) was continuously recorded from 1/11/16 to 5/27/16 about 137 days and the temporal resolution is ten minutes. Humidity and temperature within the house were recorded by wireless monitoring system and exterior temperature, humidity and pressure, window speed, visibility, dew point were observed at nearby Chièvres Airport.

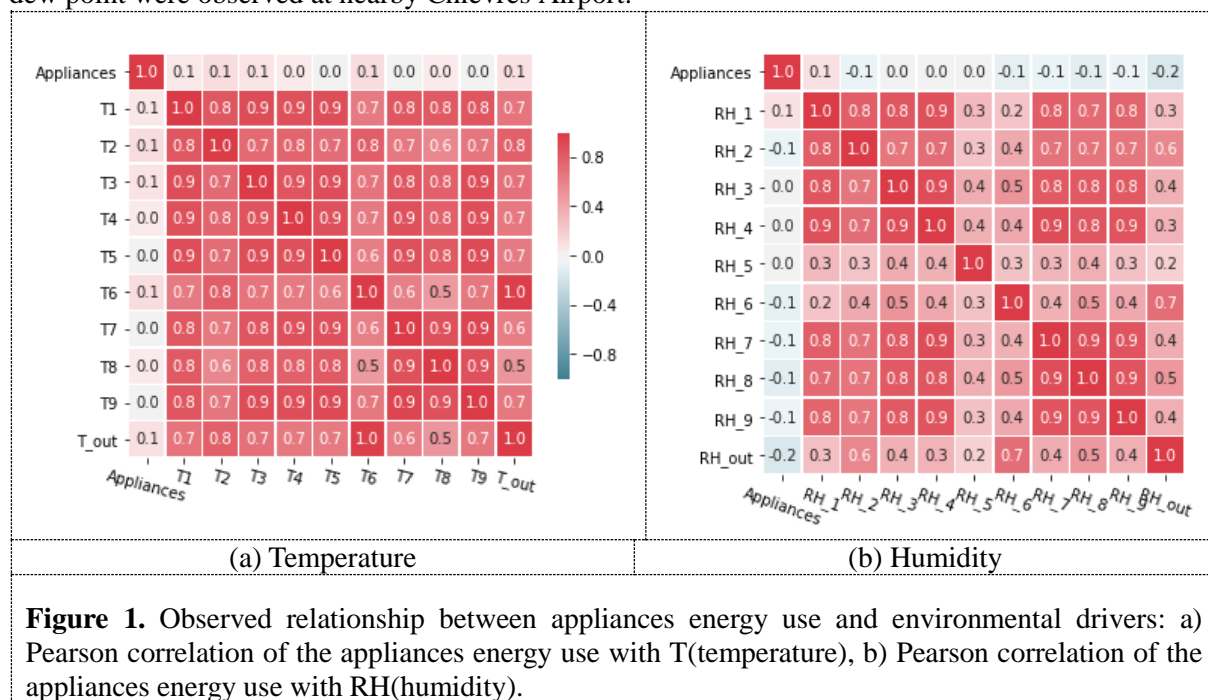


Figure 1. Observed relationship between appliances energy use and environmental drivers: a) Pearson correlation of the appliances energy use with T(temperature), b) Pearson correlation of the appliances energy use with RH(humidity).

Our data preprocessing showed that the linear correlation between appliances energy use and temperature or humidity were low; while temperature or humidity from some rooms were significantly correlated with temperature or humidity from other rooms, respectively (Figure 1). The density distribution of appliances energy use showed a long tail; the mean, median and standard deviation are about 98, 60 and 102 (Figure 2)

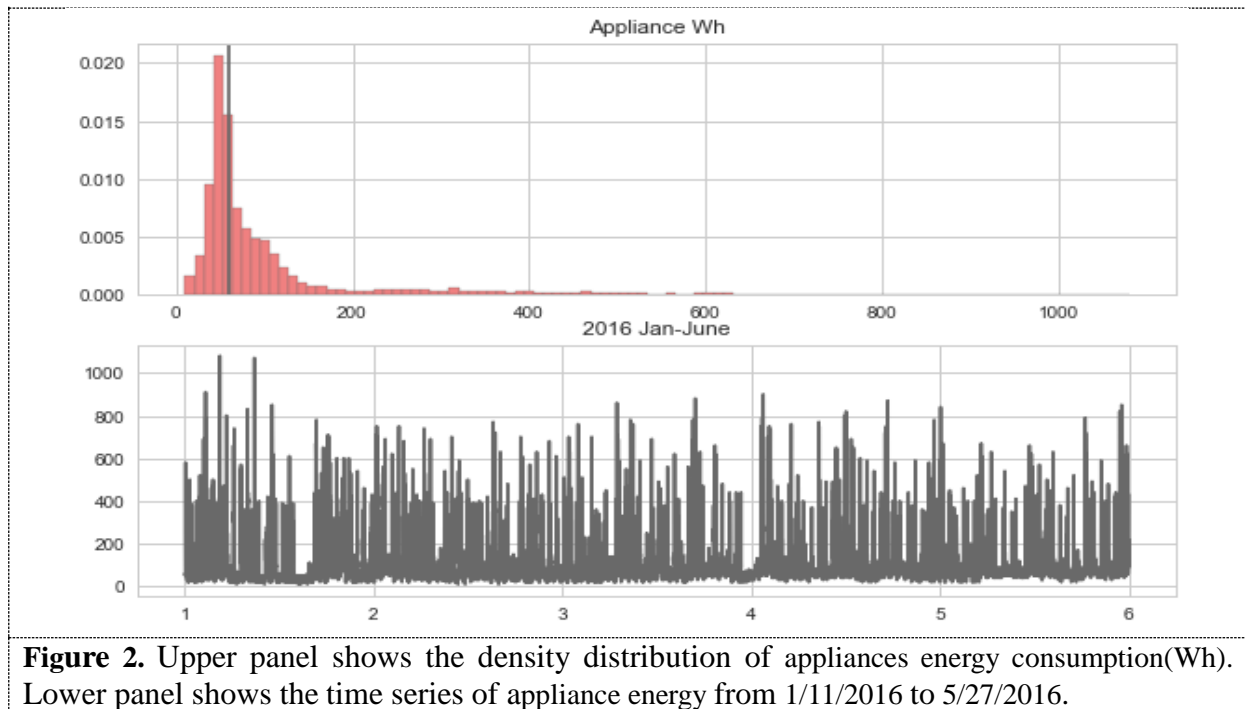


Figure 2. Upper panel shows the density distribution of appliances energy consumption(Wh). Lower panel shows the time series of appliance energy from 1/11/2016 to 5/27/2016.

2.2 House energy consumption modeling

2.2.1 Information transfer modeling. $X=\{X_t:t=1,2,\dots,n\}$ is a discretized time series with distribution $P(x)$. In our data set, X presents the variables including inside temperature, inside humidity, outside temperature, outside humidity, outside visibility and dew point. The Shannon Information Entropy H describes the uncertainty in realization of X , The target Y is a time series and presents electricity consumption in this study. Transfer Entropy measure the reduction in Shannon Information Entropy of current Y , due to the acknowledgement of the past X at some time lag [8] (Figure 3), which is between current Y and past X :

$$T(X \rightarrow Y, \tau) = H(X_{t-\tau}, Y_{t-\omega}) + H(Y_t, Y_{t-\omega}) - H(Y_{t-\omega}) - H(X_{t-\tau}, Y_t, Y_{t-\omega}) \quad (1)$$

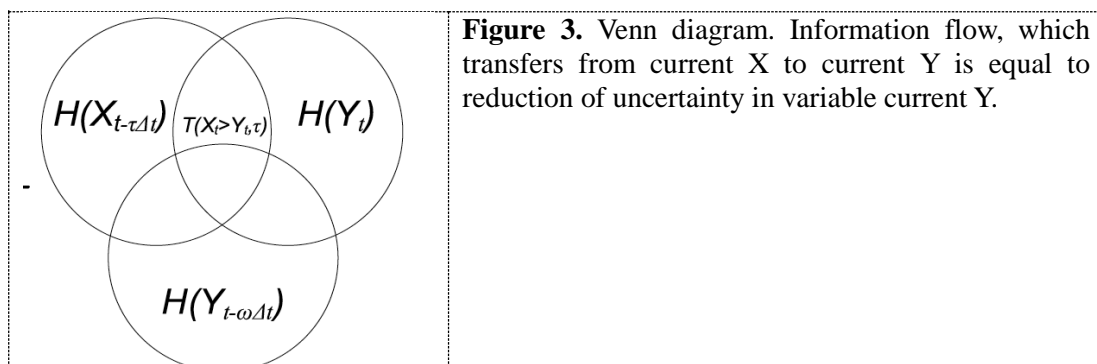


Figure 3. Venn diagram. Information flow, which transfers from current X to current Y is equal to reduction of uncertainty in variable current Y .

2.3 Recurrent Neural Networking

Recurrent Neural Networking (RNN) [9] is an artificial neural network, which could keep apart of historical information as memory to interact with the new input and generate effective outputs:

$$\begin{aligned}
f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_c) \\
C'_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\
C_t &= f_t * C_{t-1} + i_t * C'_t \\
o_t &= \sigma(W_o [h_{t-1}, x_t] + b_o) \\
h_t &= o_t * \tanh(C_t)
\end{aligned} \tag{2}$$

where $\sigma()$ is the sigmoid activation function. W_f, W_i, W_o are the weight parameters, which are optimized by fitting the model to observations during training. RNN outputs h_t given inputs x_t and h_{t-1} .

This dataset contained 25 features and the environmental data from Chièvres Airport weather station are not totally the same as the data in the building. In addition, some of them had a strong linear relationship (Figure 1, Pearson correlation > 0.7). That mean that some features were redundant in this data set. In RNN model, the filtered data set was split to 3 data sets. Dataset 1 contained all filtered variables. Dataset 2 only contained appliances and interior environmental variables, and dataset 3 only contained appliances and exterior environmental variables.

Because the features were recorded with different measurement units, it was necessary to scale them to a same numerical range. Here, we ranged all feature into the range zero and one. In this dataset, the temporal resolution was 10 mins, which mean the data vary too quickly and the random process was not stationary. To make sure that the random process was stationary, the time-series data were reorganized by averaging to 1 hour temporal resolution.

3. Results & Discussion

3.1 Causality modeling

The transfer entropy quantitatively estimated the magnitude of influence from each predictor to the target variable appliances energy consumption. All environmental drivers were considered, and we grouped the results into significant exterior drivers (T_out, RH_out, Tdewpoint, visibility, press, windspeed) and interior drivers (T1, T6, RH_1, RH_8, RH_5, RH_6). It is important to notice that the Pearson correlation between appliances energy consumption and any of these identified driving factors were very small (Figure 1), which means the causality relationships confirmed by transfer entropy were high non-linear.

Among the identified exterior environmental drivers temperature was the most important driving factors that control appliances energy consumption (Figure 4). Similarly, among the identified interior environmental drivers room temperature T1 was selected to be the most influential one. Consistently, relative humidity from both inside and outside the building were identified to be the secondary important drivers that controlled the appliances energy consumption.

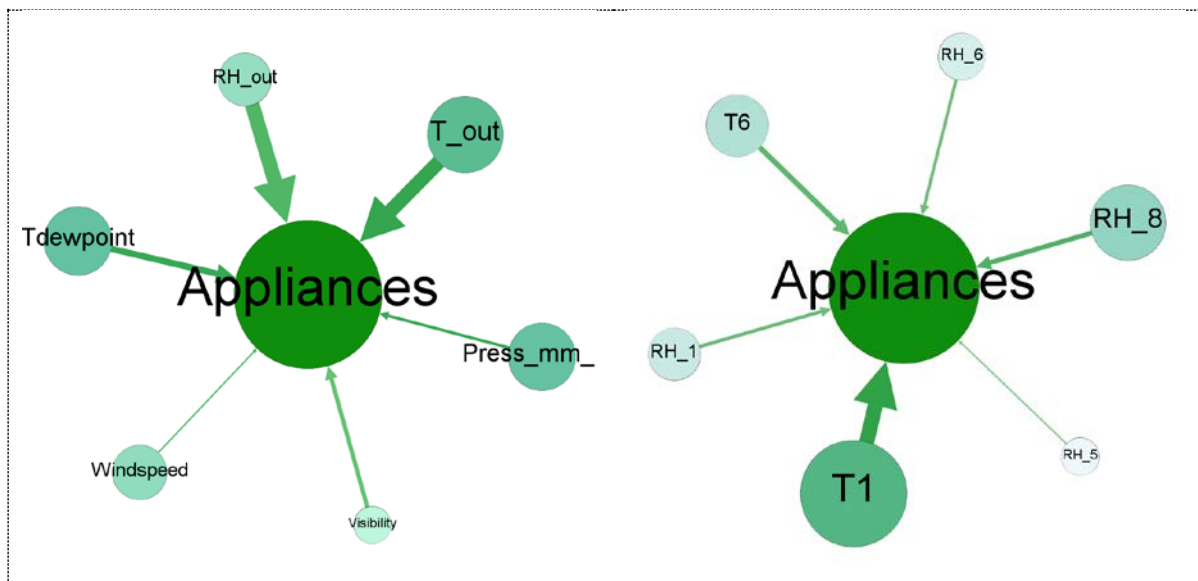


Figure 4. Causal relationship between appliances energy use (center node) and exterior environmental drivers (left panel) or interior environmental drivers (right panel). The size of node and the thickness of arrows are scaled with the amount of information transfer from one variable to Appliances energy consumption.

3.2 predictive modeling with recurrent neural network

We used 80% of observed appliances energy use to train the recurrent neural network model, and tested the trained model with the rest 20% data. Modeling experiments were carried out for three cases using (1) all environmental drivers; (2) only exterior environmental drivers (Figure 4 left panel); (3) only interior environmental drivers (Figure 4 right panel). Our modeling results showed that the memory based recurrent neural network was able to reproduce the observed temporal dynamics of appliances energy use under all three circumstances. It demonstrated that both interior and exterior environmental drivers are informative.

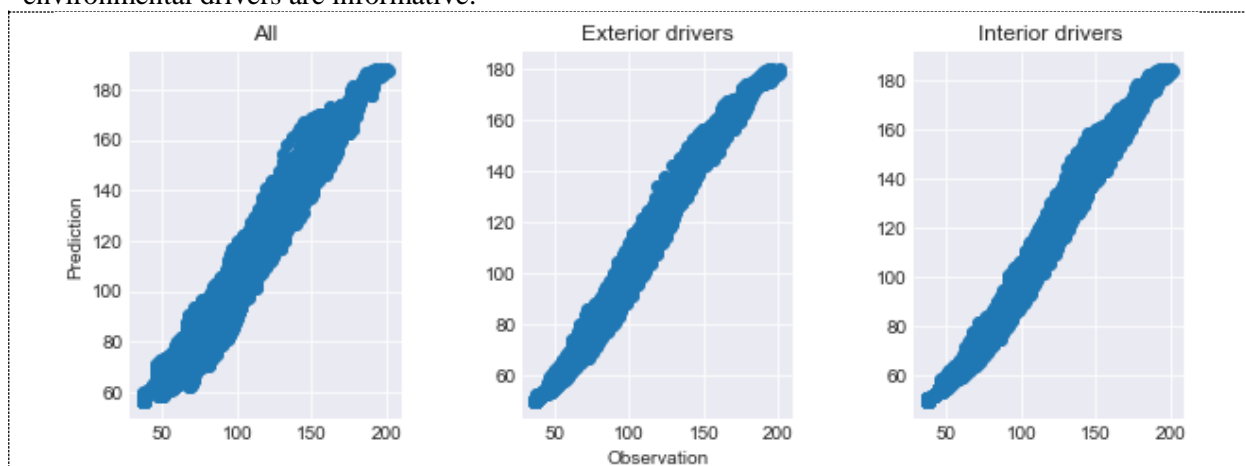


Figure 5. Observed versus predicted appliances energy use with recurrent neural network. Left, middle, and right panels correspond to using all, only exterior, and only interior environmental drivers, respectively.

3.3 Limitation and future work

This study is limited to availability of appliances energy consumption data. We focused one well developed house, but in order to build a more generic energy use predictive model, different type of

house with different energy use preference are needed. In this study, the data set only contains limited features and many other potentially important features are not discussed here such as building's geometry or floor plan. Ideally external environmental driver observation is supposed to be the condition right outside the house. We used Chièvres Airport observed conditions (12 km away from the house) as a proxy of the house external conditions, which may inevitably bring in some uncertainties.

4. Conclusion

The appliances energy consumption plays an important role in the overall electric energy consumption. Understanding the relationship between appliances energy use and relative variables is critically important. This study focused on understanding the causal relationship between potentially important drivers and the house appliances energy use as well as predicting the energy use with a limited number of environmental drivers. Via statistical analysis based on information theory, we identified exterior temperature and room one temperature as the most important drivers, and exterior relative humidity and the room eight relative humidity as secondary important drivers. Our predictive modeling using recurrent neural network successfully reproduce the observed appliances energy consumption time series with those identified important environmental drivers, thus further confirmed the important and predictable relevance between appliances energy use consumption and the environmental drivers.

5. Reference

- [1] A. Kavousian, R. Rajagopal, M. Fischer, Ranking appliances energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliances energy efficiency in residential buildings, *Energy Build.* 99 (2015) 220–230
- [2] S. Firth, K. Lomas, A. Wright, R. Wall, Identifying trends in the use of domestic appliances from household electricity consumption measurements, *Energy Build.* 40 (5) (2008) 926–936
- [3] N. Fumo, P. Mago, R. Luck, Methodology to estimate building energy consumption using EnergyPlus Benchmark Models, *Energy Build.* 42 (12) (2010) 2331–2337.
- [4] I. Richardson, M. Thomson, D. Infield, A high-resolution domestic building occupancy model for energy demand simulations, *Energy Build.* 40 (8) (2008) 1560–1566.]
- [5] C.-L. Hor, S.J. Watson, S. Majithia, Analyzing the impact of weather variables on monthly electricity demand, *IEEE Trans. Power Syst.* 20 (4) (2005) 2078–2085.
- [6] R.V. Jones, A. Fuertes, K.J. Lomas, The socio-economic, dwelling and appliances related factors affecting electricity consumption in domestic buildings, *Renew. Sustain. Energy Rev.* 43 (2015) 901–917.
- [7] Knuth, K. H. (2005), Lattice duality: The origin of probability and entropy, *Neurocomputing*, 67, 245– 274, doi:10.1016/j.neucom.2004.11.039.
- [8] Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30(1), 45-67.
- [9] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.