Impact of Occupant Behavior on Energy Prediction

Ki Uhn Ahn#1, Yong Se Kim*2, Young Min Kim*3, Cheol Soo Park#4

# School of Civil and Architectural Engineering, SungKyunKwan University
Suwon, 440-746, South Korea
1ahnkuiuhn@skkku.edu
4cheolspark@skkku.ac.kr

*Department of Convergence Engineering for Future City, SungKyunKwan University
Suwon, 440-746, South Korea
2kyongse@skku.edu
3yezzir@skku.edu

Abstract

A data-driven model is advantageous since it requires less inputs to mimic the dynamics of the reality than the first principles based simulation model. In this study, the authors developed the Gaussian Process (GP) model, one of the data-driven models, to predict the energy consumption of HVAC systems for two existing buildings. In developing such data-driven model, it is important to identify the correlation between inputs and outputs. The occupant behavior and presence were used as one of the inputs to the model. A Normalized Cumulative Periodogram (NCP) and a wavelet coherence were used respectively to investigate the predictability of occupant presence (behavior) and time-series correlation between occupants and energy consumption. It is shown that the GP model is good enough to predict energy consumption; however the model should be updated in real-time to capture the effect of non-stationary occupant behavior and presence.

Keywords - occupant; Gaussian process; random walk; wavelet coherence

1. Introduction

The data-driven model is advantageous since it requires less inputs than the first principles based model. In addition, it does not demand in-depth knowledge and expertise with regard to building physics and numerical methods.

The Gaussian Process (GP) model, one of the data-driven models, has been used for nonlinear dynamic systems [1]. It differs from other black-box approaches in that it does not try to approximate the modelled system by fitting model parameters of selected basis functions, but rather by searching for relationships among the training data [1]. Therefore, it is important to select appropriate inputs that are strongly correlated to the output.

This paper reports the development of the GP model to predict energy consumption. The wavelet coherence was used to identify the time-series
correlation between inputs and outputs. In particular, a non-stationary occupant presence was investigated based on the random walk theory.

2. Experiments

The data were gathered with real-time monitoring of room A for 7 people at building A, and room B for 2 people at building B. The sampling times were 1 min. (Building A) and 5 min. (Building B) respectively, and the list of monitored information is tabulated in Table 1. The monitoring periods are as follows: from 2 to 5 March in room A, and from 24 to 27 February in room B.

<table>
<thead>
<tr>
<th>Table 1. Monitored data</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of occupants*</td>
<td>Number / 0 or 1</td>
</tr>
<tr>
<td>Indoor air temperature</td>
<td>°C</td>
</tr>
<tr>
<td>Indoor air set-point temperature*</td>
<td>°C</td>
</tr>
<tr>
<td>Outdoor air temperature</td>
<td>°C</td>
</tr>
<tr>
<td>Window opening ratio* / status*</td>
<td>% / 0 or 1</td>
</tr>
<tr>
<td>Door opening status*</td>
<td>0 or 1</td>
</tr>
<tr>
<td>Occupants’ controlling an EHP#</td>
<td>0 or 1</td>
</tr>
<tr>
<td>Electricity of EHP# / FCU*</td>
<td>kW</td>
</tr>
<tr>
<td>Supply air temperature of FCU*</td>
<td>°C</td>
</tr>
</tbody>
</table>

# : room A only, *: room B only
1): 0 means no occupant. 1 means one or more than one occupant exist(s) in room B.
2): Once opened, the status is 1 (If fully closed, the status is 0.).
3): Once controlled, the status is 1 (in case of no change, the status is 0.).

The indoor temperature in room A was controlled by an electric heat pump (EHP). Occupants in room A can change the indoor temperature by their own wills. In contrast to room A, a fan coil unit (FCU) in room B was automatically operated to maintain the indoor set-point temperature. The indoor set-point temperature was determined by the operation policy of building B. For example, the set-point temperature in room B is dependent on occupant’s presence. If an occupant is detected in room B, the indoor temperature is set to 24 °C. If there is no occupant in the room, the room set-point temperature was maintained to 21°C. Thus, the set-point temperature in room B indicates the occupant presence.

3. Gaussian Process

The GP model is a probabilistic and non-parametric black-box model. A GP is a collection of random variables which have a joint multivariate Gaussian distribution. For every input $x$ there is an associated random
variable $f(x)$. A GP can be fully specified by its mean function (2) and covariance function (3) as follows [2]:

\[
\begin{align*}
    f(x) & \sim GP(m(x), k(x,x')) \\
    m(x) & = E[f(x)] \\
    k(x,x') & = E[(f(x)-m(x))(f(x')-m(x'))^T]
\end{align*}
\]

(1) \hspace{1cm} (2) \hspace{1cm} (3)

The mean function $m(x)$ in (2) defines the average value of $f(x)$. The covariance function $k(x,x')$ in (3) defines the correlation between the individual outputs $f(x)$ and $f(x')$ with respect to inputs, $x$ and $x'$.

If the training dataset (4) is given, the posterior distribution of the GP model (5) can be obtained by the posterior mean function (6) and posterior covariance function (7).

\[
\begin{align*}
    D & = \{(x_i,f_i), i=1:N\}, \text{ where } f_i = f(x_i) \\
    p(f^*|x^*,D) & = N(f^*|\mu^*,\Sigma^*) \\
    \mu^* & = \mu(x^*) + K^*K^{-1}(f-\mu(x)) \\
    \Sigma^* & = K^{**} - K^*K^{-1}K^*
\end{align*}
\]

(4) \hspace{1cm} (5) \hspace{1cm} (6) \hspace{1cm} (7)

where $K=k(x,x)$, $K^*=k(x^*,x)$, $K^{**}=k(x^*,x^*)$

4. Selection of inputs by wavelet coherence

The wavelet coherence analysis can be used to investigate the dependencies and correlations between two simultaneous time-series data. In this study, the authors used the wavelet coherence [3] to quantify the degree of correlation between energy consumption and other monitored data (Table 1).

Figs. 1-2 show the results of wavelet coherence analysis. In Figs. 1-2, the regions plotted in a warm color (close to 1.0 in the right Y-axis scale) indicate that two time-series data are significantly correlated. The cold color close to 0.0 in the right Y-axis scale means that they are not correlated. The arrows in Figs. 1-2 can be interpreted as follows [3]:

- arrows pointing to the right (left): two time-series are in-phase (anti-phase), or positively (negatively) correlated.
- arrows pointing upwards (downwards): an oscillation of first time-series occurs 1/4 of a period before(after) that of second time-series.

In case of room A, the occupant’s control of the EHP was most correlated to energy use by the EHP among the collected data (Fig. 1(a)), and the indoor air temperature was almost not correlated (Fig. 1(b)). In case of room B, the indoor air set-point temperature (Fig. 2(a)) and indoor air temperature (Fig. 2(b)) were significantly correlated to the energy use by the FCU.
(a) Correlation between occupants’ control of EHP and energy use by EHP

(b) Correlation between indoor air temperature and energy use by EHP

Fig. 1 Wavelet coherence results (room A)

(a) Correlation between indoor air set-point temperature (or occupant presence) and energy use by FCU

(b) Correlation between indoor air temperature and energy use by FCU

Fig. 2 Wavelet coherence results (room B)

For lack of space, the other results are not included in this paper, but the other data in Table 1 were not strongly correlated to the energy use by EHP or by the FCU. Thus, the inputs of GP model were determined as shown in Table 2.
Table 2. Inputs/outputs of the GP model

<table>
<thead>
<tr>
<th>Room #</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room A</td>
<td>Occupants’ control of EHP</td>
<td>energy use by EHP</td>
</tr>
<tr>
<td>Room B</td>
<td>Indoor air set-point temperature (or occupant presence), Indoor air temperature</td>
<td>energy use by FCU</td>
</tr>
</tbody>
</table>

5. **Non-stationary data related to the occupants**

In section 4, the occupant’s control of EHP, which is related to the occupant behavior, is selected as the input to predict energy use by the EHP in room A. In case of room B, the indoor air set-point temperature, which is related to the occupant presence, is selected as the input. Hence, it is important to test predictability of occupant’s control of EHP as well as indoor air set-point temperature. In other words, it is needed to test whether occupant’s control of EHP and indoor air set-point temperature follow a random walk or not [4][5]. If it is hard to predict occupant presence due to its randomness, it is required to update the GP model with regard to occupant presence.

Thus, the Normalized Cumulative Periodogram (NCP) was introduced to identify a dominant periodicity of a given time-series in a frequency domain. If the NCP line falls within the 95% confidence interval (the two parallel dotted lines in Fig. 3), the time-series follows the random walk.

As shown in Fig. 3(left), the occupant behavior (or occupant’s control of EHP) does not follow the random walk, but the frequencies were almost uniformly distributed. The occupant presence (or indoor air set-point temperature) follows the random walk (Fig. 3(right)). This means that it is difficult to predict occupant behavior (or occupant’s control of EHP) and occupant presence (or indoor air set-point temperature). Therefore, it is needed to update the GP model to reflect the non-stationary impact of occupants.

![Fig. 3 Random walk results; occupants’ control of the EHP in room A (left); indoor air set-point temperature in room B (right)](image-url)
6. Results

The two GP models (room A: EHP energy use prediction, room B: FCU energy use prediction) were developed using a non-linear autoregressive with exogenous external inputs (NARX) method. The NARX model is defined as follows [6]:

\[
y(t+1)=f(y(t), y(t-1),\ldots, y(t-n), u(t), u(t-1),\ldots, u(t-n))
\]  

(8)

where the prediction value at time \( t+1 \) \( (y(t+1)) \) is a function of previous values of itself \( (y(t), \ldots, y(t-n)) \) and of other inputs \( (u(t), \ldots, u(t-n)) \). The authors also use a recursive prediction strategy [7] for the NARX method.

Fig. 4 (a) and Fig. 5 (a) are the results of one-step ahead prediction. The model predictions were in good agreement with measurements (room A: MBE 4.5%, CVRMSE 17.3%, room B: MBE 0.1%, CVRMSE 27.4%, Table 3). However, the errors between model prediction and measurement were increased with respect to the time horizon (Table 3, Fig. 4, 5). This is due to the fact that the non-stationary inputs related to occupants (room A: occupants’ control of the EHP [occupant behavior], room B: indoor air set-point temperature [occupant presence]) influenced the model prediction.

![Graph](attachment:image.png)

Fig. 4 Comparison of GP model prediction and measurement for EHP energy use in room A
Fig. 5 Comparison of GP model prediction and measurement for FCU energy use in room B

Table 3. Prediction of the GP model

<table>
<thead>
<tr>
<th></th>
<th>Room A</th>
<th>One-step ahead (5 min.)</th>
<th>Three-step ahead (5 min.)</th>
<th>Twelve-step ahead (5 min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBE</td>
<td>4.7</td>
<td>10.9</td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>CVRMSE</td>
<td>17.3</td>
<td>22.9</td>
<td>31.2</td>
<td></td>
</tr>
<tr>
<td>Room B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBE</td>
<td>0.13</td>
<td>41.2</td>
<td>82.9</td>
<td></td>
</tr>
<tr>
<td>CVRMSE</td>
<td>27.5</td>
<td>272.2</td>
<td>359.9</td>
<td></td>
</tr>
</tbody>
</table>

7. Conclusion

A data-driven model has been highlighted as a surrogate model to the whole building dynamic simulation model. In this study, two GP models were developed to predict the energy use of EHP and FCU. The wavelet coherence was used to select appropriate inputs for prediction of the outputs. The inputs related to the occupants’ behavior and presence were significantly correlated to the energy consumption. However, such data follow the random walk or are close to the random walk. In other words, the occupant behavior and presence have randomness and are of non-stationary feature.

The GP models were constructed by a NARX method with a recursive prediction strategy. One-step ahead prediction models for room A and B
perform well. But the errors between the model prediction and measured data were increased with respect to the time horizon. It means that it is hard to provide multi-step ahead prediction of occupant behavior and presence. Thus, such non-stationary data must be updated to the model for better prediction.

**Acknowledgment**

This work was supported by the Energy Efficiency & Resources Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea. (2015202105550)

**References**