ESTIMATING OCCUPANCY STATES FROM BUILDING TEMPERATURE DATA USING WAVELET ANALYSIS

Michael Georgescu, and Igor Mezić
Department of Mechanical Engineering, University of California, Santa Barbara

ABSTRACT
Whole building energy models have found widespread use in estimating the energy consumption of building systems. Within these models, usage profiles are assumed when capturing the influence of processes such as occupancy, lighting, and equipment operation. Usage profiles are defined hourly, but are repetitive in the sense that their shapes are periodic, typically, from week to week. When evaluating a design, the use of periodic usage profiles is accepted since detailed knowledge of building operation is usually unknown. In the case of an existing building, however, this approach may not accurately capture building behavior and cause an error in prediction from the resulting mismatch in operation between the actual and modeled building. In this work, fluctuations in space occupation of a multi-use university building is studied. By decomposing building temperature data using wavelets, room occupancy states (i.e. the status of a room being occupied or vacant) are estimated. From this estimate, usage profiles are generated which better capture the actual behavior of the building. Predictions of energy usage from a model simulation which utilizes this implementation is compared to building utility data as well as a simulation where conventional usage profiles are assumed.

INTRODUCTION
A growing sector of the building systems community focuses on modeling existing buildings. In the past several years, applications using models of existing buildings have included retrofit studies (Jenkins 2008), failure mode analysis (Otto et al. 2012), and model predictive control (Corbin, Henze, and May-Ostendorp 2011). Despite current capabilities, a barrier encountered when modeling existing buildings is defining usage profiles which capture the energy consuming and thermal effects from processes such as occupancy, lighting, and equipment operation. Conventionally, building simulation software such as EnergyPlus and TRNSYS (TESS 2012; DOE 2012) expect usage profiles to be predefined hourly based on knowledge of how a building to be modeled should ideally operate in an assumed setting. In the case of an existing building, however, this approach can create mismatch in behavior between an actual and modeled building leading to reduced accuracy.

To better understand the complexities of how working buildings operate, research exists on methods for sensing building occupants and their behavior. The work of (Dong et al. 2010) describes the implementation of a large-scale environmental sensing network; using measurements of carbon dioxide and other volatile organic compounds, an algorithmic relationship is sought after between the number of occupants within a room and human expired contaminants for use in demand control ventilation. In (Benezeth et al. 2011), a computer vision based system is used to identify individual occupants, and characterize the activities performed by occupants through a network of video cameras. Other studies include the Time Use Survey (TUS) (ESDS 2012) and American Time Use Survey (ATUS) (BLS 2012) where participants are tasked with tracking their activities at sub-hourly intervals.

Using this data, occupancy models have been developed to capture the complexities of building operation. Although conventional usage profiles may not capture isolated occupant events, in (Davis III and Nutter 2010; Widn, Molin, and Ellegrd 2012) periodic usage profiles are defined from approximations of survey data and measurements of occupant behavior of building sites with good success. In (Richardson, Thomson, and Infield 2008; Wilke, Haldi, and Robinson 2011), stochastic occupancy models are created using Time Use Survey (TUS) and American Time Use Survey (ATUS) data previously mentioned. Statistical models have also been studied which use regression techniques to correlate building occupancy with building utility usage (Martani et al. 2012) or outdoor environmental conditions (Dong, Lam, and Neuman 2011). A number of works develop models which simulate the actions of a virtual community of building occupants by creating multiple agents, each with a persona, and predicting actions the agents (Hoes et al. 2009; Liao, Lin, and Barooah 2012).

In this work, a new approach for the generation of usage profiles for use in the modeling of existing buildings is examined. By decomposing building temperature data of individual rooms using wavelets, features can be identified
corresponding to heat loads generated by occupant presence and equipment operation. From this information, the occupancy state (i.e., the status of a room being occupied or vacant) is estimated. To the best of the author’s knowledge, this technique has not been used for analyzing building data however other spectral based decomposition methods have been explored in the previous works. In (Eisenhower et al. 2010), an EnergyPlus model of a building is partially calibrated by decomposing temperature data into oscillatory modes, and comparing the modes produced by the model to that from measured temperature data. Later in (Georgescu, Eisenhower, and Mezic 2012), these oscillatory modes are again used to systematically create reduced order building energy models. Because building temperature behavior is periodic in nature, a frequency based approach can illustrate characteristics of building data which may otherwise go unnoticed in a time-series analysis.

The remainder of this paper is organized as follows: in the following section, a brief overview of signal analysis using wavelets is given and the continuous wavelet decomposition is introduced. Then using the continuous wavelet decomposition, temperature data of an actual building is analyzed, and occupancy states of rooms are estimated based off of characteristics of their temperature response. The estimated occupancy states are then applied to a whole building energy model of the measured building. The paper is concluded with a comparison of the utility consumption of the building model, with and without this implementation, compared to actual utility data to illustrate the achieved improvements in model predictive ability.

REVIEW OF WAVELET ANALYSIS

Building heat transfer occurs at multiple time-scales and is driven by internal and external sources. A building experiences heat loads which change slowly over months, such as seasonal changes, or at a time scale of minutes in the case of HVAC control loops. Using wavelet analysis, a given temperature response can be decomposed into wavelets allowing time and scale specific features of the response to be studied.

Wavelet transformations are inner products of a square-integrable signal with a family of certain basis functions derived from what is known as a mother wavelet. To introduce the transformation, consider the time-domain signal \( x(t) \). The continuous wavelet transformation (CWT) constructs a time-scale representation of the signal, and is expressed by the following integral:

\[
X_{a,b}(t) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^* \left( \frac{t-b}{a} \right) dt.
\]  

In Equation 1, the term \( \psi(t) \) is a continuous function in the time and scale domain known as the mother wavelet and * represents complex conjugation. The basis functions of the transformation in Equation 1 are translated and scaled version of the mother wavelet \( \psi \) expressed as

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left( \frac{t-b}{a} \right).
\]

In Equation 2, \( a \) is a scale factor and \( b \) is a time location. The wavelet transformation, \( X_{a,b}(a,b) \), gives information of \( x(t) \) at scales relating to parameter \( a \) and temporal location given by \( b \). An alternate conceptualization is that \( X_{a,b}(a,b) \) is the convolution of \( x(t) \) with the wavelet function \( \psi_{a,b}(t) \). Because of this, the CWT is easily calculated using the Fast Fourier Transformation (FFT).

Many functions, \( \psi(t) \), have been designed for use in wavelet analysis. The decision to use a certain mother wavelet over another depends on the system being analyzed. For this paper, the so-called Morlet wavelet (Goupillaud, Grossmann, and Morlet 1984) is chosen as the mother wavelet. The general form of the Morlet wavelet is given by the function

\[
\psi_\sigma(t) = c_\sigma \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2} (e^{i\sigma t} - e^{-\frac{1}{2}\sigma^2})
\]

where

\[
c_\sigma = \left(1 + e^{-\sigma^2} - 2e^{-\frac{1}{2}\sigma^2}\right)^{-\frac{1}{2}}.
\]

The Morlet wavelet is a sinusoid with a gaussian window. It is common in literature to select the envelope factor, \( \sigma \), so Equation 3 becomes

\[
\psi_\sigma(t) = e^{-\frac{t^2}{2}} \cos(5t)
\]

The Morlet wavelet is chosen for the analysis performed because the contribution of different frequencies present in an input signal are kept reasonably separated in its decomposition. Since building temperature data is periodic in nature, this allows features at different scales to be distinguishable.

In this work, the continuous wavelet transformation is utilized to estimate occupancy states of measured rooms in an existing building. The occupancy state corresponds to whether a room has (or has not) been occupied on a particular day. When occupants are present in a room for sufficient length of time, effects are seen in the decomposed wavelets of the room temperature response which are unlikely due to other processes. Through estimating occupancy states using this approach, usage profiles can be generated and applied to a whole building energy model. This enhancement more realistically captures the complex behavior of occupants and process loads when modeling an existing building.
THE STUDENT RESOURCES BUILDING

The Student Resources Building (SRB) was constructed in 2007 at the University of California, Santa Barbara. The building is a hub for campus organizations which primarily serve student needs.

Figure 1: (Color Online) Exterior views of the university building studied. Northern entrance of the SRB (left), and southern entrance (right).

The 68,000 sqft floor plan contains three above ground floors and is separated into northern and southern halves by a 4 story 5500 sqft. atrium acting as a divider between these halves. The building includes several areas of different space utilization including offices, classrooms, child daycare, and auditorium / theater, as well as common areas were student groups can assemble.

An energy saving feature of the building design includes extensive use of natural ventilation throughout the building floor space. The atrium and offices in the perimeter of the building are naturally ventilated with baseboard heaters while rooms in the core of the buildings are conditioned by one of six air handling units. Mechanically ventilated rooms are conditioned by either variable air volume units with terminal reheat or constant air volume units. Fan coil units condition unoccupied rooms that contain electrical or telecommunications equipment which require continuous space cooling. HVAC chilled water is provided by district cooling from a campus chilled water loop while heating for hot water is generated by a gas fired boiler located in the plant of the building.

Because of the coastal setting, cool offshore winds flow perpendicular to the length of the atrium. Using motors along the perimeter of the atrium, vents can be opened allowing hot air to advect from the atrium ceiling while cool outdoor air is drawn through vents located on the ground floor of the building.

THE ENERGYPLUS MODEL

The whole building energy model of the SRB was created using DesignBuilder (DesignBuilder 2012) and simulated using EnergyPlus v6.0.0.023. Weather input for the simulation was generated using local measurements of outdoor temperature, humidity, wind speed, wind direction, and solar radiation recorded from a local weather station. In this model, each room is represented as its own thermal zone. A rendering of the building model is shown in Figure 2. The model contains 215 zones. With a time-step duration of 30 minutes, the model takes 15 minutes to simulate one month of operation on a computer with 2.53 GHz CPU.

Figure 2: Exterior views of the EnergyPlus model of the university building studied. Northern exterior (top), and southern exterior (bottom).

Architectural data for the model was obtained from CAD drawings, and thermal properties of materials used in construction were estimated using reasonable assumptions from literature. Internal heat gains include occupancy, lighting, and equipment operation. In the case of lighting, heat generation was calculated from specifications available in building drawings. Heat from equipment operation and occupancy were estimated by an audit conducted during building walkthroughs. Initial usage profiles were modeled after the operating hours of the building and observational data of occupant activity. The building is closed during weekends, so it is assumed to be unoccupied during this time.

WAVELET ANALYSIS OF BUILDING DATA

In this section, temperature data collected from room sensors will be analyzed in the time-scale domain using the continuous wavelet transformation. For this study, building data was collected over a four week period from September 2, 2012 to September 29, 2012. Although...
some rooms do not contain temperature sensors, measurements were taken from both naturally ventilated and mechanically ventilated rooms. A total of 65 rooms contained temperature sensors whose data was collected. From this data, the CWT of the temperature response of each room was calculated. Figure 3 shows a spectrogram of the CWT calculated from temperature data of a single room during the four week period. A spectrogram illustrates magnitude, scale, and position in time of wavelets used to decompose the signal.

Figure 3 compares the CWT of a naturally ventilated room which contains manually operated baseboard heating, but otherwise has no control of temperature, to the CWT of outdoor air temperature for the same time period. Because wavelets are of finite duration, the concept of frequency is not directly applicable, however since the Morlet wavelet used in the decomposition has a frequency response which is constrained to a narrow range of frequencies, the concept of a pseudo-frequency can be applied. The pseudo-frequency and pseudo-period of a wavelet scale are defined as

\[ f_a = \frac{f_c}{a\Delta} \quad T_a = \frac{1}{f_c} \]  

where \( f_a \) and \( T_a \) are the pseudo-frequency and pseudo-period, \( \Delta \) is the sampling period of the signal, \( a \) is the scale of the wavelet as defined in Equation 2, and \( f_c \) is the center frequency in the wavelet’s frequency response.

Looking at the spectrogram of both temperature responses, features can be identified which relate the decomposed wavelets to attributes of the time-series data. Features such as daily temperature oscillations are represented by wavelets at pseudo-periods of 24 hours and greater, while fast irregular temperature changes are seen at pseudo-periods of smaller duration. At pseudo-periods smaller than 6 hours, differences in the CWT between weekdays and weekends can be seen in the room temperature response.

The difference in wavelet magnitude between weekdays and weekends of the room temperature response is likely due to the presence of occupants. Such patterns are only observed at low pseudo-periods. These patterns are unlikely due to heat transfer through conduction from outdoor air as any oscillatory response at this time scale does not have a large enough magnitude to exceed the thermal penetration depth of the wall construction. This assumption is tested by comparing both wavelet decompositions and noticing that the spectrogram of the outdoor air temperature lacks the weekday versus weekend pattern. Similar to outdoor air, heat transfer through conduction from adjacent rooms would experience the same lack of thermal penetration. The heat transfer, and its effect on wavelet magnitude, can however occur from internal heat loads due to occupancy.

From this assumption, it is believed that durations of high spectral energy in the wavelet decomposition at low pseudo-periods of the room temperature response are caused by occupant presence. Spectral energy at low pseudo-periods would be caused either through heat generation within the room, or the transport of air into and out of the room when occupants are present. During the weekend, little spectral energy would be measured since the room is unoccupied, equipment is turned off, and the air of the room is more quiescent.

Similar observations are made for mechanically ventilated rooms. In these rooms, the presence of occupants creates disturbances in the room’s temperature response. When an HVAC controlled room is vacant, little spectral energy at low pseudo-periods is measured much like the naturally ventilated spaces.

Because this approach relies on occupants to be the main source of heat at low pseudo-periods, this approach is not applicable to rooms whose temperature sensors are exposed to direct solar radiation. At times, these sensors may measure temperature which is not indicative of the space. Due to this, these sensors were excluded from the analysis.

Using this approach, the CWT of all available sensors were calculated. In the wavelet decomposition of each sensor, if a room contains more spectral energy on a specific day then what is measured, on average, during a weekend, the room is assumed to be in an occupied state. The day-by-day occupancy state of each measured room was tracked. Figure 4 illustrates the measured occupancy states for the 65 sensors monitored over the 4 week period. Some trends can be seen in Figure 4, which when compared to the university calendar, provide insight into occupancy trends of the building studied. In the four week period monitored, the second week of monitoring corresponds to the final week of the summer academic quarter. This is a time which associated with high building usage. In contrast, the third week of data contains a greater number of rooms in an unoccupied state the academic quarter has completed and students are on a week-long vacation. From the estimated occupancy states, usage profiles where created which capture the observed trends. The creation of these usage profiles, and their effect on model accuracy, will now be discussed.

**SIMULATING ESTIMATED OCCUPANCY STATES**

Using estimated occupancy states shown in Figure 4, a probability mass function (PMF) are calculated for each measured day. The PMF is used to calculate an occupancy state for rooms which do not contain temperature sensors. Figure 5 illustrates the PMFs calculated from sensors during the time period monitored.
Figure 3: (Color Online) (A) Time-series temperature trend from a naturally ventilated room building sensor. Time-scale representation of room measured temperature data calculated using the continuous wavelet transformation shown over all pseudo-periods (B) and high pseudo-periods (C). (D) Time-series temperature trend of outdoor air temperature. Time-scale representation of outdoor air temperature calculated using the continuous wavelet transformation shown over all pseudo-periods (E) and high pseudo-periods (F).
Usage profiles were originally created based on the operating hours of the building and observational data of occupant activity. These profiles were then modified using the calculated occupancy states. The occupancy state of rooms are determined by the PMFs shown in Figure 5. If a room is at an occupied state for a given day, its normal usage profile is simulated, otherwise, if the room is at an unoccupied state, the behavior of the room is overwritten to operate at unoccupied conditions.

The model was simulated with both original usage profiles as well as modified usage profiles which incorporate the calculated occupancy states. In Figure 6, the electric consumption of both simulations are compared with measured building electric consumption. In the building studied, electric consumption is strongly influenced by occupancy, so calculated occupancy states produce a simulation which more accurately predicts the fluctuating trends in building electric consumption. Compared to using modified usage profiles, electric consumption does not vary considerably from day to day due to the periodic nature of the original usage profiles.

To quantify the improvement in accuracy, the adequacy of each model is evaluated using ASHRAE Guideline 14-2002 (ASHRAE 2002). The ASHRAE guideline defines limits on the mean bias error (MBE) and coefficient of variation of root mean squared error (CV(RMSE)) when comparing a model prediction to measured data. The mean bias error (MBE) and coefficient of variation of root mean squared error (CV(RMSE)) are defined by

\[
\text{MBE} = \frac{\sum (M - S)_{hr}}{\sum M_{hr}} \quad (7)
\]

\[
\text{CV(RMSE)} = \frac{\sqrt{\sum (M - S)_{hr}^2} \times N_{hr}}{\sum M_{hr}} \quad (8)
\]

where \(M_{hr}\) is hourly measured data, \(S_{hr}\) is hourly simulated data, and \(N_{hr}\) is the number of hours in the interval being compared. For a model to be declared calibrated, according to the guideline, the model must have an MBE below ±5% and CV(RMSE) below ±15%. Using these metrics, the accuracy of each simulation to measured data was calculated. The results are shown in Table 1.

When incorporating occupancy states to the model usage profiles, the mean bias error and coefficient of variation of root mean squared error both reduce and reach a level considered acceptable by the guideline.

**CONCLUSIONS**

In this work, temperature data from rooms of a mixed-use university building are analyzed using the continuous

---

**Table 1: Model Accuracy Estimate**

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBE</td>
<td>4%</td>
<td>-0.8%</td>
</tr>
<tr>
<td>CV(RMSE)</td>
<td>18.5%</td>
<td>10.5%</td>
</tr>
</tbody>
</table>

---
Figure 6: Percentage error of EnergyPlus simulations in predicting building energy usage. The unmodified simulation uses conventional hour-by-hour occupancy schedules created from best estimates of building operation. The modified simulation incorporates predicted fluctuations in occupancy inferred from building temperature data.

wavelet decomposition. Utilizing the decomposition, features of the temperature behavior of rooms are extracted and used to estimate periods of occupancy and vacancy. From this information, probability mass functions are calculated to statistically capture the observed occupant behavior. This statistical measure of occupancy is incorporated into an EnergyPlus simulation of the building studied to gauge its effectiveness. From this approach, trends in the measured building energy consumption are captured well in the building simulation. Over the 4 week period studied, the prediction of energy consumption is within 10 percent of the measured consumption on a daily basis and captures fluctuations in energy consumption which are present in measured data.

Several key questions remain for further exploration in future work. One question is whether the resolution of wavelet analysis can be improved to estimate occupancy states at shorter time scales (i.e. determining daily initial entrance and/or final exit times of occupants). As more data becomes available, more sophisticated occupancy models can be created taking into consideration factors such as space usage or calendar trends for forming a more complete predictions. Because temperature data is available in many modern buildings, future work will aim at extending this approach, and study occupancy patterns, of other building types. With a building specific estimate of occupancy patterns, tasks such as model calibration can be simplified by reducing the amount of uncertainty from occupancy based model inputs.

ACKNOWLEDGEMENTS

The authors would like to thank Erika Eskenazi and Valerie Eacret for their invaluable time and effort in modeling the Student Resources Building, as well as Kazimir Gaslejevic and the employees from UC Santa Barbara facilities management for their help with capturing building data. This work was partially funded by Army Research Office Grant W911NF-11-1-0511, with Program Manager Dr. Sam Stanton.

NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x(t)$</td>
<td>time-series signal</td>
</tr>
<tr>
<td>$X_{W}(a,b)$</td>
<td>transformed signal</td>
</tr>
<tr>
<td>$\psi(t)$</td>
<td>mother wavelet</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>normalization constant</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>envelope factor</td>
</tr>
<tr>
<td>$a$</td>
<td>scale factor</td>
</tr>
<tr>
<td>$b$</td>
<td>time factor</td>
</tr>
<tr>
<td>$f_a$</td>
<td>pseudo-frequency</td>
</tr>
<tr>
<td>$T_a$</td>
<td>pseudo-period</td>
</tr>
<tr>
<td>$f_c$</td>
<td>central frequency</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>sampling period</td>
</tr>
</tbody>
</table>

REFERENCES


