Decomposing building system data for model validation and analysis using the Koopman operator

Article

4 authors:

Bryan Eisenhower
University of California, Santa Barbara

Tobias Maile
Maile Consulting, Fellbach, Germany

Martin Arthur Fischer
Stanford University

Igor Mezic
University of California, Santa Barbara

DECOMPOSING BUILDING SYSTEM DATA FOR MODEL VALIDATION AND ANALYSIS USING THE KOOPMAN OPERATOR

Bryan Eisenhower[1], Tobias Maile[2], Martin Fischer[2], and Igor Mezić[3]
[1] Center for Energy Efficient Design, UCSB
[2] Civil and Environmental Engineering, Stanford University
[3] Department of Mechanical Engineering, UCSB

ABSTRACT

Large amounts of sensor information is often captured from either real world building sensors, or virtual building models, for many purposes including control design, fault or aging analysis, and model calibration. Because of the large dimension of this data on both spatial and temporal scales, it is often challenging to come to quick conclusions about what information of engineering importance is in the data. In this paper we present an approach to quickly assess spatial information in data based on the spectral content of a certain projection operator. We use operator theoretic methods to capture Koopman modes that represent the spatial content of oscillations in thermal quantities. By investigating these modes for different physically significant time-scales (e.g. diurnal, or control system time-scales) we can quickly capture how different parts of a building are responding to load changes at these frequencies (”breathing”). This information helps us to understand anomalies in different aspects of the data, as well as out of phase behavior between zones which may highlight areas of poor control system performance. We present actual and EnergyPlus data from a real building (170K square foot building with approximately 2000 data points) and illustrate how this approach to data analysis and model validation highlights aspects of the data which may otherwise have been overlooked.

INTRODUCTION

As the need for high performance buildings increases, higher fidelity models are becoming more prevalent, and a large amount of data is being captured from buildings for both model calibration and other analysis including fault or problem detection. As the size of data from both sensors and models grows, new tools are needed to quickly isolate important information in this data.

In model verification, many variables are needed to accurately and uniquely identify whether a model is capturing pertinent physical processes correctly. In some studies, model validation has been performed on too few physical quantities, and this complicates the process of identifying sources of discrepancies between model and data (Jensen 1995). On the other hand, investigating data from each sensor one-by-one is too laborious to be practical, and some sort of decomposition or reduction technique is needed.

Data decomposition is a method to simplify data to its core or most insightful components (or modes). Modal analysis is such a method which is an excellent tool for coarse-graining global properties of dynamic data. In the context of analyzing a partial differential equation or spatial data (e.g. multiple sensor measurements from structural vibrations of a bridge), an array of sensor data/variables at specific locations can be regressed into several variables which contain information that is not confined to one sensor or position in space. Although traditionally the case, there is certainly no reason that modal analysis needs to be confined to the analysis of structural systems.

There are many decomposition methods which take large data sets and separate them into different and meaningful components. Principal component analysis (PCA) is an approach which decomposes data based on its energy/variance (also called Karhunen−Loeve transform (KLT), the Hotelling transform, or proper orthogonal decomposition (POD), depending on research field). In (Reginato et al. 2009) a prediction method for energy use in a commercial building based on solar conditions, sales, humidity, electricity consumption, and temperature was performed using PCA. In (Lam et al. 2010) PCA was performed on data over a very long time frame (29 years) to identify the impact of climate on energy usage. A similar method was used to identify/quantify the behavior of a building before and after renovation in (Ruch et al. 1993). PCA has also been used for data analysis and even fault detection (Du, Jin, and Wu 2007), while to our knowledge, there has been no attempt to use this decomposition technique for model validation/calibration in building systems.

Although the PCA methods are great examples of rigorous data analysis for building system data, the issue with these methods is that although insightful information is gained, the principal components represent only large variances in the data and information regarding temporal frequencies is lost. In particular, the temporal evolution of PCA modes still has complex variability at more than one frequency. Data from building management systems is nearly (or quasi-) periodic, and contains information at many frequencies. That is, the external forcing and internal operation of buildings excite its dynamics at a number of frequencies. The fastest of these frequencies are at minute time scales dominated by control system cycling (10-20 minutes). On the scale of hours, we have human behavior like academic class scheduling (in the case of a campus building) and normal workday hours, as well as diurnal forcing from daily weather patterns. At longer scales we have weekly work schedules and seasonal effects. All of these phenomena behave in a nearly periodic fashion. Because of all of this frequency content, it is reasonable to try to extract spatial modes that oscillate at a single frequency, and connect these to the forcing typically encountered in building operations.

In this paper we pursue decomposition of spatially and temporally evolving temperature data into its components evolving in a purely periodic manner. This is akin to linear normal modes in vibration studies; however, the analysis leading to modes we extract here is nonlinear and based on the properties of the Koopman operator. Every nonlinear finite-dimensional process can be embedded into evolutions driven by an infinite-dimensional, linear operator, called the Koopman operator.

Bernard Koopman pioneered the use of linear transformations on Hilbert space to analyze Hamiltonian systems by introducing the so-called *Koopman operator*, and studying its spectrum (Koopman 1931); see (Peterson 1983; Lasota and Mackey 1994) for futher details. This linear, infinite-dimensional operator is defined for any nonlinear dynamical system (Peterson 1983; Lasota and Mackey 1994). Even if the governing dynamics of a system are finite-dimensional, the Koopman operator is infinite-dimensional, and does not rely on linearization, it captures the full information of the nonlinear dynamical system. In (Mezić and Banaszuk 2004) the authors identified a relationship between general Fourier analysis (Wiener and Wintner 1941) and eigenfunctions of the Koopman operator. Comparison of complex data is also facilitated by the use of the Koopman operator (Mehta and Vaidya 2005) and (Mezić and Banaszuk 2004). In (Mezić 2005) the author showed using spectral analysis of the Koopman operator that single-frequency modes can be embedded in highly nonlinear, spatiotemporal dynamics. These modes are later named *Koopman Modes* (KMs) in (Rowley et al. 2009) where the authors presented a technique for characterizing the global behavior of complex fluid flows by decomposing a flow profile into KMs. In a completely different context, Koopman modes were used to study extended power system grids (Susuki and Mezić 2010). To the best of our knowledge, this technique has not yet been used for analysis of either sensor of model data from building systems.

Koopman Operator and Modal Decomposition

To introduce the properties of the Koopman operator, and its use for data analysis, consider evolution of a dynamical system on a finite but multi-dimensional space of variables $x \in M$, which can be thought of as a state-space. The evolution of these variables in time is described by the nonlinear equation

$$x(k+1) = f(x(k)), \tag{1}$$

where f is a function that maps the variables at instant k to new variables at instant $k+1$. Now define a observation function $g : M \to \mathbb{R}$ which maps the arbitrary variables on the manifold M to real numbers (this can be thought of as sensors in a building). The Koopman operator $U : \mathbb{R} \to \mathbb{R}$ is then defined as

$$Ug(x) = g(f(x)). \tag{2}$$

Equation 2 can simply be thought of as transforming, or projecting the dynamics into a different domain. The benefit of using the Koopman operator is that it simplifies the original evolution function f, which may be highly nonlinear and complex (although evolving in a finite dimensional domain), into a linear operator which is infinite dimensional. This process is performed without linearizing or discarding any information in the original dynamics.

Since U is a linear operator, it satisfies the eigenvalue equation

$$U\psi_i(x) = \lambda_i \psi_i(x), \tag{3}$$

for $i = 1,2,3,\ldots$, where $\psi_i(x)$ is the i-th eigenfunction corresponding to the i-th eigenvalue λ_i. It is exactly this property that provides an approach to dissect the dynamics of building system data into useful and actionable information. As an aside, the Koopman operator is a unitary operator, and thus all of its eigenvalues lie on the unit circle in the complex domain.

As we mentioned, building system data is intertwined with elaborate and overlapping periodic content. Fortunately, using the Koopman approach, we can isolate different frequencies (proportional to λ_i) in this data. Once we have isolated the i-th frequency of interest, we can then investigate the i-th Koopman eigenfunction ψ_i to gather global information in the data.

There are multiple approaches to the actual calculation of the Koopman modes, such as using harmonic averages of the spatial field (as described in (Mezić 2005)), or by using the Arnoldi algorithm (Rowley et al. 2009) or (Sohn and Law 2001) (note that the Ritz values/vectors behave similar to the eigenvalues/vectors of a finite truncation of the Koopman operator). One should note in particular that

$$Ug^*_\omega(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{i2\pi j\omega} g(f(x)), \tag{4}$$

where $\omega \in [-0.5, 0.5)$ and $e^{i2\pi j\omega}$ are eigenvalues, and g^* is a harmonic average. These harmonic averages contain information about the spectral content of U. For further information about performing decomposition using the Koopman operator see the references above. The algorithm used in this paper has been adapted from the algorithm described in (Rowley et al. 2009).

The Jerry Yang and Akiko Yamazaki Environment and Energy (Y2E2) Building

The Jerry Yang and Akiko Yamazaki Environment and Energy (Y2E2) Building (Fig. 1) was constructed in early 2008 on the Stanford University campus. The building houses many different teams with a focus of research in sustainability, energy, water, and land use.

Figure 1: The Jerry Yang and Akiko Yamazaki Environment and Energy Building

The 166K square foot *L-shaped* building contains a basement and three above ground floors. The basement is predominately designed for laboratory use while the three upper floors contain academic offices and other meeting rooms.

The focus of the energy strategy for the building was to reduce usages, pacify ventilation, increase efficiencies in heating and cooling, recover wasted energy, generate portions of its needed energy (rooftop PV installations) and offset purchases from the community.

Since a low energy footprint was desired, four atria for both natural ventilation and smoke release were designed with interconnected meeting rooms, and common workspaces. The natural ventilation throughout the building was further enhanced by a combination of motorized and manual windows as well as auxiliary louvers. The building management system (BMS) actuates these louvers as needed when outside conditions are adequate as well as at night (a description of the building can be found in (Graffy et al. 2008)).

Not only is the building designed with cutting edge laboratory space, the building itself is a living laboratory. A total of 2370 HVAC system measuring points, each sampled at one minute intervals, were installed in the building to monitor its performance. Using this data, it was found that in the first year, the energy consumption was twice what was predicted at the design stage (Kunz, Maile, and Bazjanac 2009). The factors contributing to the discrepancy between predicted and actual performance were incorrect design assumptions (e.g., unexpected occupancy scheduling), control problems, and last minute changes to the design.

In the process of analyzing the data to assess its performance, it was found that even with the remarkable number of sensors in the building, many different problems existed to gathering pertinent information from the data. In particular, many of the data sensors were mis-calibrated, incorrectly mapped, or converted/scaled incorrectly. Similarly, upon looking in detail at the data, many different control issues were found including improper set-points, undesired periodic cycling, and other systematic issues including improper night purge. Many of these issues are important to efficient operability of the building, while unfortunately there is no automated approach to identifying these issues. In fact, it took the cooperation of eleven students an academic semester to pour over the data sets and identify these issues.

EnergyPlus Model of Y2E2

The thermodynamics of the Y2E2 building were thoroughly modeled using the EnergyPlus simulation program. The EnergyPlus model contains approximately 500 zones for the four levels of the building which corresponds mainly to the architectural spaces in the building. There are more data points in the model than what is sensed in the actual building, for the naturally ventilated corridors the zones in the EnergyPlus model are more detailed than the temperature measurements available. The weather files for this model were modified to use actual weather data and occupancy and other schedules were adjusted to best approximate what the building actually experiences.

The technique in this paper is a step in the direction of automating the process of identifying faulty and other global characteristics in either model or experimental data and to compare the model to data. The first part of the approach is to decompose data into global or modal characteristics and associated spectra. This is described in the analysis section below.

Koopman Analysis of the Y2E2 Building

In this section we will analyze the behavior of the Koopman operator for data collected from both sensor data and the EnergyPlus model outputs of the Y2E2 building. We will first illustrate how qualities of the spectra of the operator quickly highlights inconsistencies in controller function on the first floor of the building. We will then compare the model and data using the Koopman operator for the second floor of the building (because it has more sensor information there).

The basic steps for investigating the Koopman modes is to compute the spectra of the Koopman operator and select a particular frequency of interest. Once this frequency is selected, the Koopman mode associated with that frequency is obtained and investigated. For clarity, when viewing the Koopman mode, it is overlayed on architectural drawings of the building for the particular floor (compare this to the image in Fig. 1). For each image, a black dot depicts locations of sensor or data points from the model, and interpolation is performed between these data points.

The presentation of the information in the Koopman mode is always relative to the outdoor air temperature (OAT) in this paper. That is, for the magnitude data, each index of the Koompan mode is divided by the magnitude of the outdoor air temperature and then presented in decibels. A similar approach is used to obtain the relative phase to OAT:

$$|KM(i)| = 20\log_{10}\left|\frac{\psi_i}{\psi_{OAT}}\right| \tag{5}$$

$$\angle KM(i) = \angle\psi_i - \angle\psi_{OAT}. \tag{6}$$

Controller Tuning

As a first example, we investigate sensor data from the first floor of Y2E2 alone, and we notice that in April of 2009, there exists a high frequency oscillation with a short period of about 15 minutes in some of the sensor data (see the circled region in Fig. 2 [1]). When plotting the Koopman mode associated with this frequency (Fig. 3) we find that indeed three of the rooms are experiencing fluctuations in their temperatures at frequencies that other rooms are not experiencing.

The original time history can be viewed in Fig. 4 where we present data from three sensors (20,21,24) which are cycling at a very high frequency. Data from sensor 23 is also presented as a comparison of what data from other rooms/zones looks like. As it turns out, the controller was eventually re-tuned, and this high frequency oscillation is not evident in data from later months. The discovery and fix for this problem was found before the methods in this paper were developed by investigating time histories of the controller performance. It could be argued that this approach is adequate and complex mathematics is not needed to tune control systems, but we suggest that using the approach in paper accelerates this type of analysis.

A few other issues were found in the sensor data from the building including time instances when sensor communication is lost, which can be quickly observed in the Koopman modes. Another topic that is being investigated currently is efficient operation of the HVAC equipment.

[1]It should also be noted that the data has not been rearranged so that the variable number between the building and model coincide because the different number of sensors between the two (e.g. variable 24 from the model may not be anywhere near sensor 24 from building as in Fig. 5).

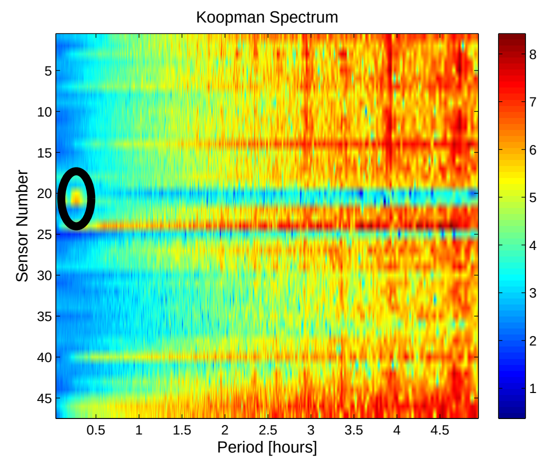

Figure 2: Koopman spectrum for the first floor of Y2E2 sensor data in April 2009. The circled region highlights an area where spectral energy is not expected.

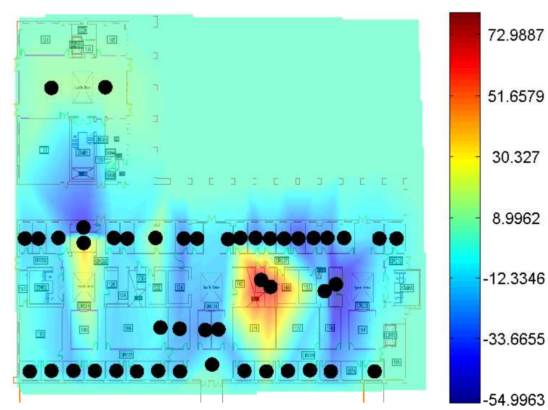

Figure 3: Koopman mode associated with the high frequency spectral energy in Fig. 2

We have found that at some frequencies, the phasing between different regions of the building contain some peculiarity which may be due to HVAC control systems operating partially or completely out of phase, indicating large inefficiencies in its operation. The investigation of this behavior is currently underway and therefore not reported in this paper.

Comparison of Model and Data

In the previous section, we presented the Koopman decomposition technique on original sensor data. We now use similar method to compare the dynamics of the building data with that of the EnergyPlus model. We perform this analysis on the second floor because it contains more

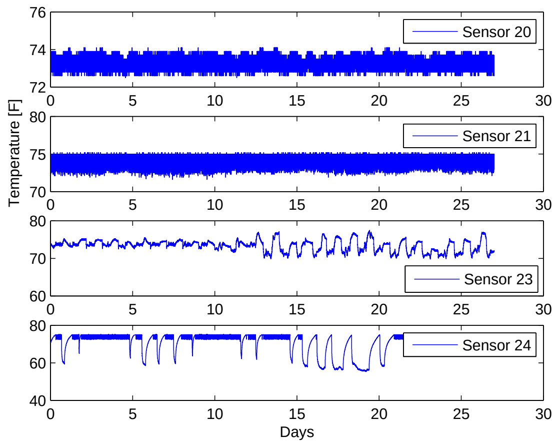

Figure 4: Time history of the sensors which are highlighted in Fig. 3 as having high frequency and unexpect oscillations.

sensor information than the first floor (the study of the other floors of the building is currently underway). In general, the EnergyPlus model behaves very well when compared to data, while we do find areas of slight discrepancy which we discuss below.

Below we study the spectra and modal behavior for the entire month of April 2009, this spectra is presented in Fig. 5 where it can be seen that in both the upper and lower figures, high spectral energy exists at the 24 hour period which coincides with the forcing from outdoor diurnal conditions. In addition to this, we find horizontal lines which are very low in amplitude. These lines depict very low spectral energy at all frequencies which corresponds to nearly constant temperature signals. These constant temperature signals are coming from data rooms which are served by their own fan coil units and therefore have very tightly controlled temperatures.

The most obvious mode to investigate in the data is the mode that contains the most spectral energy. Due to the strong response of the building dynamics from the outside ambient conditions, this frequency is at a period of 24 hours. We present the Koopman mode for this spectral bin in Figs. 6 and 7.

In the magnitude plot (Fig. 6), we find that in the model, the temperatures inside the building respond at a larger amplitude than what is observed in data. This can be seen by the similar color shading for most of the building in the model response. This just means that the temperatures in the model oscillate at about the same amplitude as the OAT, while for actual data, the oscillation amplitude at this frequency is much smaller (as shown by cooler colors). The very cold spots in both plots are just the equipment rooms where temperature barely fluctuates.

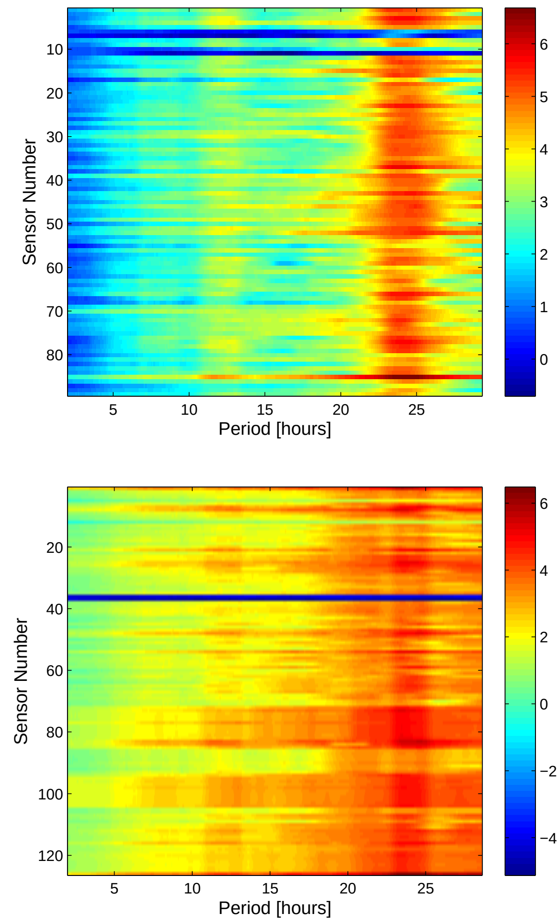

Figure 5: Koopman spectra for the sensor data (top) and EnergyPlus model (bottom)

The fact that a few sensors show very little fluctuation at all blurs out the detail in the rest of the magnitude plot and more information may be gained by removing these data points and comparing only those which have a reasonable response. On the other hand, the phase information from the same data set, which is presented in Fig. 7, contains useful information as it stands.

When we plot the phase information from the Koopman mode associated with the 24 hour oscillation we find there to be interesting differences with respect to phase delay in the model compared to actual sensor data. In the sensor data, we find that the temperatures in the perimeter offices lead the OAT signal (primarily the upper most in the figure which are all hot in color). This may be explained by the fact that these north facing offices have heating systems only, compared to the offices facing south and west. The other possible explanation is that the internal load schedules in the model do not coincide with the actual usage which is currently being investigated. The upper most of-

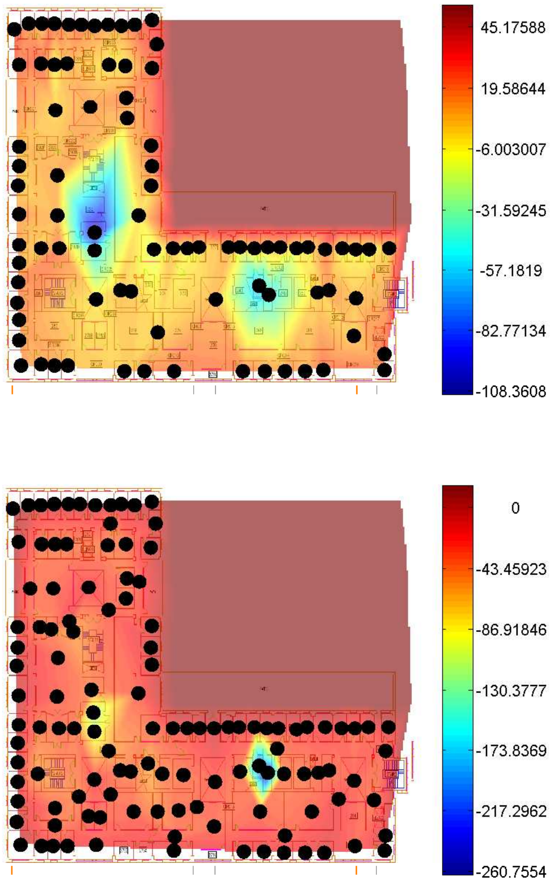

Figure 6: Magnitude of the Koopman mode for the 24 hour oscillation. Sensor data (top) EnergyPlus (bottom).

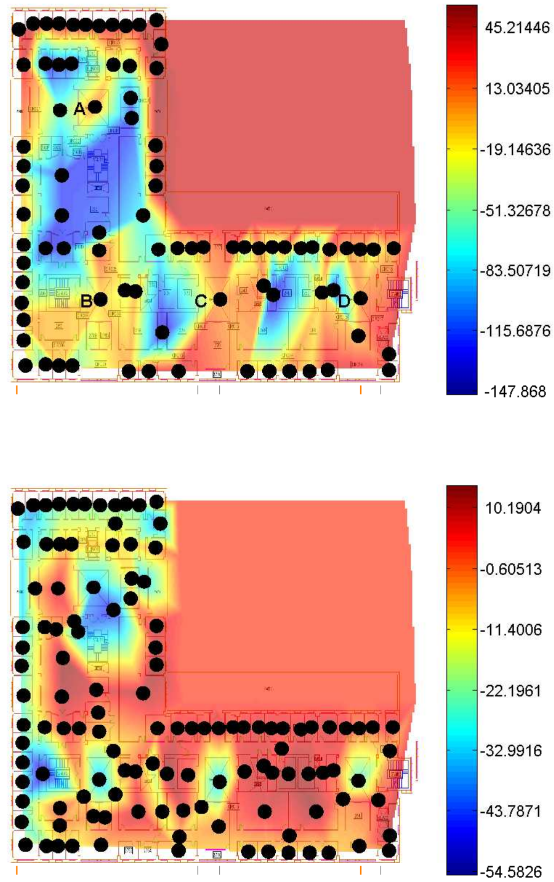

Figure 7: Phase of the Koopman mode for the 24 hour oscillation. Sensor data (top) EnergyPlus (bottom).

fices, and the offices facing the courtyard also experience different phase patterns in the sensor data. This may be due to the proximity of a neighboring building on the upper row while the lower offices face the courtyard. How this impacts the model is currently being investigated as well.

There is a similar contrast in the phase behavior in the atria of the building (the four atria are marked by the letters *A,B,C,D*). In the data, the response of the temperature in the atria each follows OAT very closely with very little phase lag. On the other hand, in the model, the phase lag is much greater (about 30 degrees which is about 2 hours). This indicates that the natural ventilation in the model does not have enough air flow through the building, and this was adjusted in the model to better capture this phasing.

The model was re-calibrated in attempt to better represent the atria temperatures. In order to do this, the active beam induction ratio was changed (from 50% to 75% recirculated air), the infiltration values were increased, and natural ventilation was introduced between corridors and offices on second floor (before the natural ventilation for the offices was ignored). These changes are still being further tuned while in Fig. 8 we illustrate the change in the phasing for the 24-hour Koopman mode for the second floor with these initial changes.

We find that with the changes mentioned above, the phase behavior changes floor-wide as well as in the Atria. The phase delay in the Atria is less than it was in the first version of the simulated model. An example of this can be seen in Fig. 9 where we can see that the peaks and valleys of the Atria data move left indicating less delay.

The phase response of the building is predominately influenced by either HVAC scheduling or thermal mass. We know that thermal mass stores heat and releases it slowly, and similarly, HVAC controls respond at different times compared to OAT. What information in analysis like this allows us to do is to quickly quantify whether these very

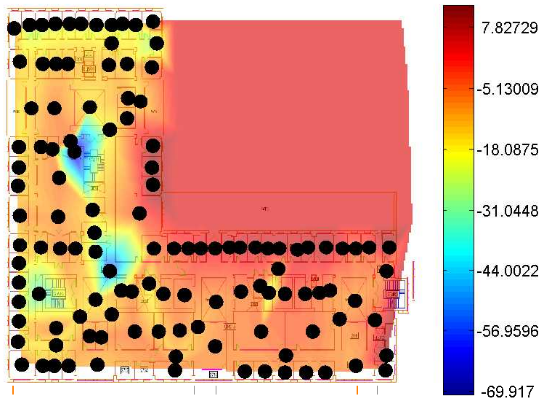

Figure 8: Phase of the Koopman mode for the 24 hour oscillation in the re-calibrated EnergyPlus model.

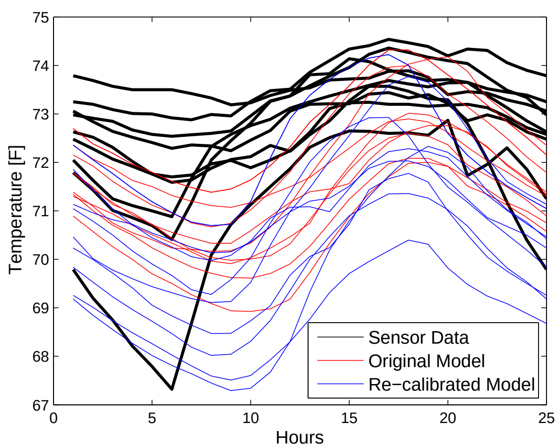

Figure 9: Data from Atria A from both the model and sensors. Nine days in April are overlapped for this figure.

important parts of the building dynamics are properly set up in the model.

The particular model in question was generated to find differences between predictions and measurements and thus to highlight performance issues with the building. We used the specific insights based on this phase information to improve the corresponding simulation model. On the other hand, in future studies, if this comparison was being made for a model that was intended for control design, getting this phase information correct would be essential for the design of a stable control system.

CONCLUSION AND FUTURE WORK

In this paper we discussed a spectral decomposition approach to analyze building system data. This approach is well suited for dynamics which contain significant amount of periodic content. We find that using this approach, quick conclusions can be made about sensor function and comparisons between models and data are accelerated. In this paper, this is done by visual inspection of modes while future efforts are to investigate the normative properties of the modes and how they vary between models and data or as data ages for fault-type analysis is planned.

REFERENCES

Du, Zhimin, Xinqiao Jin, and Lizhou Wu. 2007. "PCA-FDA-Based Fault Diagnosis for Sensors in VAV Systems." *HVAC&R Research* 13 (2): 349 (March).

Graffy, Kurt, Janette Lidstone, Cole Roberts, Brandon G Sprague, Jake Wayne, and Armin Wolski. 2008, March. "Y2E2: The Jerry Yang and Akiko Yamazaki Environment and Energy Building, Stanford University, California." Technical Report, Arup.

Jensen, Søren Ostergaard. 1995. "Validation of Building Energy Simulation Programs: A Methodology." *Energy and Buildings* 22:133–144.

Koopman, B. O. 1931. "Hamiltonian systems and transformations in Hilbert space." *Proceedings of the National Academy of Sciences of the USA* 17 (5): 315–318 (May).

Kunz, John, Tobias Maile, and Vlado Bazjanac. 2009, October. "Summary of the Energy Analysis of the First year of the Stanford Jerry Yang and Akiko Yamazaki Environment and Energy (Y2E2) Building." Technical Report TR183, v2, Stanford University.

Lam, Joseph, Kevin Wan, S. Wong, and N. Lam. 2010. "Principle Component Analysis and Long-term Building Energy Simulation Correlation." *Energy Conversion and Management* 51:135–139.

Lasota, A., and M. C. Mackey. 1994. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. New York: Springer-Verlag.

Mehta, Prashant, and Umesh Vaidya. 2005. "On Stochastic Analysis Approaches for Comparing Complex Systems." *Proceedings of the 44th Conference on Decision and Control*, no. ThC11.5:8082–8087.

Mezić, Igor. 2005. "Spectral properties of dynamical systems, model reduction and decompositions." *Nonlinear Dynamics* 41 (August): 309–325.

Mezić, Igor, and Andrzej Banaszuk. 2004. "Comparison of systems with complex behavior." *Physica D* 197:101–133.

Peterson, K. 1983. *Ergodic Theory*. Cambridge: Cambridge University Press.

Reginato, Bruno, Roberto Freire, Gustavo Oliveira, Nathan Mendes, and Marc Abadie. 2009. "Predicting the Temperature Profile of Indoor Buildings by using Orthonormal Basis Functions." *Eleventh International IBPSA Conference*, July 27-30, 1773–1780.

Rowley, Clarence, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan Henningson. 2009. "Spectral Analysis of Nonlinear Flows." *Journal of Fluid Mechanics* 641 (December): 115–127.

Ruch, D., Lu Chen, J. Haberl, and D. Claridge. 1993. "A Change-Point Principle Analysis (CP/PCA) Method for Predicting Energy Usage in Commercial Buildings: The PCA Model." *Journal of Solar Energy Engineering*, vol. 115 (May).

Sohn, Hoon, and Kincho Law. 2001. "Extraction of Ritz Vectors from Vibration Test Data." *Mechanical Systems and Signal Processing* 15 (1): 213–226.

Susuki, Yoshihiko, and Igor Mezić. 2010. "Nonlinear Koopman Modes and Coherency Identification of Coupled Swing Dynamics." *IEEE Transactions on Power Systems*.

Wiener, N., and A. Wintner. 1941. "Harmonic analysis and ergodic theory." *American Journal of Mathematics* 63 (2): 415–426 (April).

NOMENCLATURE

x	state space variable
M	high dimensional manifold
k	sampling instance
f	nonlinear function
$\mathbb{R}$	set of real numbers
g	observation function
U	Koopman operator
ψ_i	i-th Koopman eigenfunction (mode)
λ_i	i-th Koopman eigenvalue
KM	Koompan mode as presented in the Figures