A method for discovering functional relationships between building components from sensor data

Marco Pritoni[1], Arka Bhattacharya[2], David Culler[2], Mark Modera[1]
[1]University of California, Davis [2] University of California, Berkeley
marco.pritoni@gmail.com
arkaaloke@berkeley.edu
culler@berkeley.edu
mpmodera@ucdavis.edu

ABSTRACT

In Building Automation Systems contextual information about sensors is frequently missing or hard-coded in the control code. Retrieving this data is time consuming and error-prone, but necessary to write any type of control application. Automating metadata acquisition is a new and active area of research. Methods to infer metadata from sensor labels or from recorded data have been previously proposed. However, these methods are ineffective in uncovering the functional relationship between HVAC components. In fact, measured variables (pressures, temperatures, flows, valve positions) have slow and attenuated responses to changes in input variables, thus impairing the efficacy of correlation methods. In addition, sensor readings are frequently constrained between physical limits and kept around setpoints by nested control loops. For this reason, pure statistical methods fail to capture the differences between sensor streams and are unable to classify them. In this article, we propose a new method for discovering functional relationships between building systems from sensor data. The method utilizes perturbations of subsystem variables, while guaranteeing that the building zones remain within comfort boundaries. The method is applied to an existing building and its results are compared with those obtained by other pre-existing methods.

Categories and Subject Descriptors

C.3 **[Special-Purpose and Application-Based Systems]**: Process Control Systems

General Terms

Algorithms, Measurement, Performance, Experimentation, Verification, Design.

Keywords

Correlation, Clustering, Perturbation, System Identification, Modeling.

1. INTRODUCTION AND MOTIVATION

Energy efficiency in the buildings sector offers great potential for cost-effective emissions reductions. In buildings we spend ~90% of our time, consume ~75% of total electricity, which represents nearly half of our primary energy consumption, and generate 45% of our CO_2 emissions [1]. In large commercial buildings, traditional digital control systems regulate the majority of the energy use, particularly HVAC systems, which we focus on here, and lighting. These large sensor deployments are cyber-physical systems with thousands nodes.

Software applications have been recently developed to optimize energy use, improve comfort, and identify faults for these systems [8, 9, 10, 11, 12]. For all these applications to be implemented detailed information about sensors' context is required. Unfortunately, such information is very difficult to obtain, because it is either hard-coded in the building automation system (BAS) or missing. Manual retrieval is time consuming and error-prone, but currently necessary to write any type of control application. Current efforts in automatic metadata acquisition include two different strategies: extrapolating metadata from labels (e.g. BACnet point names), and inferring them from sensor readings. Recent work has just started exploring the latter approach. Fontugne et al. [2] proposed a method to correlate inter-device user patterns by extracting traces of occupancy from electrical energy use. Koc et al. [3] compared correlation methods to infer spatial relationships between discharge and zone temperature sensors in different rooms of a building. Rajagopal et al. [6] developed a method for using LED frequency modulation and smartphones cameras to establish a relationship between fixture and occupant location.

Despite these efforts, many issues remain unresolved. In particular, we set out to devise a method for inferring functional relationships between HVAC components, as lack of this information precludes the adoption of common energy efficiency strategies (resets). This paper explores an instance of this problem (on purpose this is a very difficult case to solve). The example is specific to a particular building, but technique and considerations are generalizable to many other buildings. Indeed, the majority of the US large commercial buildings have some variation of these systems installed.

2. PRELIMINARY RESULTS WITH EXISTING METHODS

This article analyzes data from a large commercial building with 3 chillers, 4 air-handling units (AHU) and 179 thermal zones. A variable air volume (VAV) box modulates the airflow from the AHU to each zone. VAV modulation is achieved through a combination of two control actions: adjusting the air damper

BuildSys'15, November 3–4, 2015, Seoul, South Korea.

position (DMP) and regulating the local reheat valve (RVP) (Figure 1). The temperature in each zone (T_{zone}) is influenced also by the supply air temperature (T_{sa}) and flow (FLW_{sa}) coming from the AHU. All these points are monitored and recorded by the BAS, and represent the datasets used here. In addition, the room temperature is impacted by uncontrolled variables, such as weather, internal gains, and other thermal gains. The association between AHU and VAV is not stored in the BAS; however, for the purpose of the experiment, ground truth association was collected to verify the results of the proposed method.

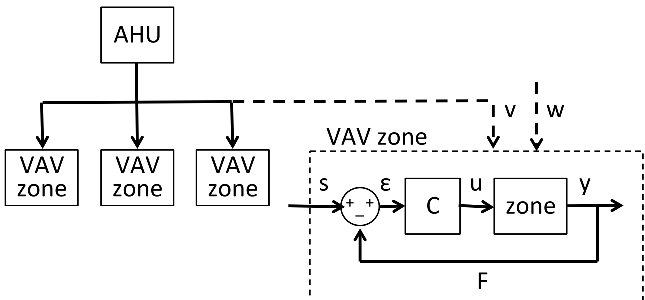

Figure 1. AHU and VAV configuration. On the right, the control logic is represented. In this case: $s=T_{setpoint}$, $F=y=T_{zone}$, $u=[DMP, RVP]$, $e=T_{setpoint}-T_{zone}$, $v=[T_{sa}, FLW_{sa}]$, w=external disturbances (T_{oat}, wind …).

Techniques commonly employed to detect relationships in data streams include: correlation of raw data [3], correlation of transformed data [2], principal component analysis (PCA) and clustering [4,7], statistical process control [12], and model-based system identification (i.e., building a model by looking for the best fit from a single VAV and alternative AHU data). Preliminary analysis of the building dataset tested conventional correlation methods to find relationships between AHU and the corresponding VAV boxes. Results show very low correlation between variables from AHU and corresponding VAV boxes. Thus, correlation does not allow identifying which VAV boxes are connected to which AHU. Table 1 shows an example of these coefficients. Desired values inside the bold boxes should be higher, in absolute value, than the corresponding values in the other rows. The same test was repeated with data resampled at 5 min, 15 min and daily, yielding similar results.

Table 1. Correlation matrix (showing raw data from two AHU and two VAV boxes). Data from May 28 to July 14 2015, resampled at daily rate.

		VAV 251 (under AHU3)			VAV 108 (under AHU5)		
		T_{zone}	DMP	RVP	T_{zone}	DMP	RVP
AHU 3	T_{sa}	-0.06	0.06	0.11	-0.11	0.24	0.29
AHU 5	T_{sa}	-0.20	-0.19	-0.06	-0.04	-0.01	0.02

In contrast to prior research monitoring energy consumption [2], the variable measured in this test are pressures, temperatures, flows and actuator positions, which show delayed and attenuated responses to changes in input variables, thus reducing correlation. Further, AHU and VAV boxes/zones are physically distant (differently from [3]), and variables in the latter are significantly influenced by additional measured and non-measured inputs (Figure 1). For this reason, sensor values show a small signal to noise ratio (Figure 2). In addition, sensor readings are frequently constrained between physical limits (e.g. max damper position) and kept around setpoints by nested control loops (Figure 3).

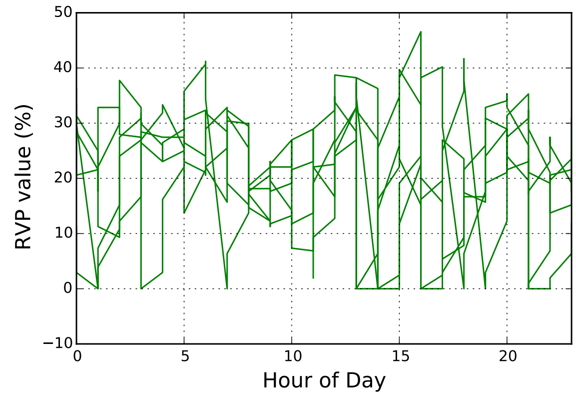

Figure 2. Raw data for RVP in zone 378 against time of the day for multiple days.

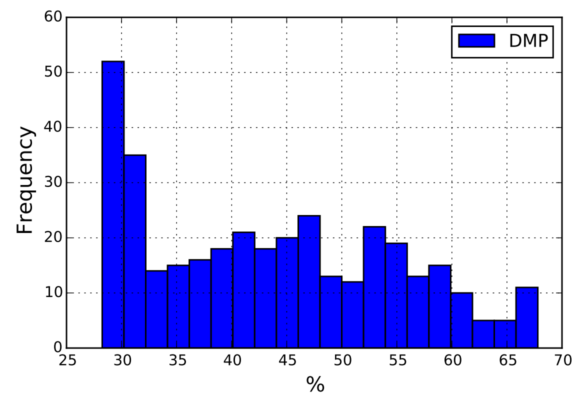

Figure 3. Raw data for DMP in zone 378 against time of the day for multiple days. High frequency of data at minimum damper position causes a bi-modal distribution.

Also, cross-talk between systems (i.e., zones might influence each other) and similarity in the way different AHU are controlled (setpoints and daily behavior are similar) make correlation of raw data ineffective. Figure 2 and 3 show the behavior of some of the VAV variables for a zone over several days. RVP in Figure 2 has no correlation with time of day and oscillates between 0% and 45%. DMP in Figure 3 shows a clear bi-modal distribution, as it is physically bounded to a minimum damper position and controlled to that value by the controller for some modes of operation.

Next, we constructed feature vectors for the data of each VAV. Features included were: T_{zone}, DMP, RVP, $T_{setpoint}$, measured flow (FLW), flow setpoint (FWS), day of the week, time of the day. PCA was applied to these feature vectors to identify the two principal components and correlate them to the T_{sa}. Unfortunately, results of this analysis are not very different from what obtained with raw data (Table 2). Again, this method was ineffective in inferring the relationship investigated.

Table 2. Correlation matrix (showing 2 principal components of VAV data and two AHU). Data from May 28 to July 14 2015, resampled at daily rate.

		VAV 251 (under AHU3)		VAV 108 (under AHU5)	
		Eig1	Eig2	Eig1	Eig2
AHU 3	T_{sa}	0.12	0.12	0.20	0.09
AHU 5	T_{sa}	-0.12	-0.15	-0.04	-0.05

Finally, we tested a completely different and novel approach involving system identification (SID) techniques, which are used in control engineering to find a mathematical relationship (model) between inputs and outputs variables in an observed system [5]. A physics-inspired black-box dynamical model was constructed to predict T_{zone}, based on the available sensor data:

$$T_{zone,t} = \beta_1 * T_{zone,t-1} + \beta_2 * FLW_t * RVP_t + \beta_3 FLW_t * T_{sa,t}^{AHU}$$

where variable names are defined above and t stands for time. Note that all the variables in this equation belong to the VAV with the exception of ($T_{sa,t}^{AHU}$). The idea is that using the $T_{sa,t}^{AHU}$ from the AHU actually connected to each VAV would improve the model fit. Both linear regression and Lasso were used to fit the model over 15-min resampled data. While the model fit the data very well (R^2=76-95% depending on the zone), it failed to capture the difference in AHU. Plugging in different $T_{sa,t}^{AHU}$ did not change the fit of the model as we had expected. We reasoned that the main cause of this is that the majority of the variation in the output variable $T_{zone,t}$ is captured by the first term of the model (zone temperature at the previous time step) and the remaining β_i coefficients are relatively small. With the calculated coefficients, the input variable $T_{sa,t}^{AHU}$ would have to change by more than 20 °F to produce measurable effects in the output variable. Such large temperature differential never occurs spontaneously in our recorded data, and if artificially produced would seriously compromise occupants' comfort. Nevertheless, this prompted us to explore the idea of actively perturbing the building to measure system response.

3. METHODOLOGY

The concept underlying the proposed methodology is that by arbitrarily perturbing an AHU we can generate a distinguishable signal in the connected VAV boxes. In practice, by changing dramatically the supply air temperature in an AHU (input), the VAV box will respond by changing some controlled variables (reaheat and damper position) to maintain the temperature setpoint. However, if despite saturation of actuators (damper or reheat valve completely open or close) the system cannot keep up with zone cooling/heating load, the zone temperature will be affected and change. Care was taken in ensuring that the perturbation had no impact on occupant comfort. To do so, while creating an overall large perturbation, the temperature setpoint for the supply air in the AHU was set one day to 52 °F (cold mode) and the following day to 60 °F (hot mode), whereby the normal setpoint for a week-day was 57 °F. Since thermal systems have a significant response lag, the perturbation was sustained for a full day in each mode. To reduce the noise in the data, knowing that VAV variables have little correlation with time of the day (Figure 2), we resampled the data in daily intervals. Average daily values for RVP, DAM and T_{zone} are still meaningful indicator of system behavior.

Another intuition was that evaluating a metric for each zone across all the perturbation periods would provide better results than assigning the zone to an AHU for each perturbation (using a threshold).

The algorithm took the following steps:

1) Perturb one AHU at the time for two consecutive weekdays, one day in cold mode and one day in hot mode.
2) Collect data for each VAV for the following sensors: RVP, T_{zone}, DAM.
3) Re-sample data to obtain daily averages.
4) For each daily average, label the data as belonging to a baseline weekday (WD), cold day, or hot day, for each perturbed AHU (ColdAHU2 would mean cold period for AHU #2). Each VAV data stream will have periods for all AHU perturbations.
5) De-trend data using outdoor air temperature in the baseline. This step makes sure that all perturbations are "normalized" by outdoor conditions.
6) Standardize each stream of data to have mean =0 and standard deviation = 1.
7) For each zone and each period combine data into daily vectors of the form [RVP, T_{zone}, DAM].
8) For each zone and each couple of AHU perturbations, calculate the Euclidian distance between hot and cold vectors. At the end of this step each zone will have a metric for each AHU hot-cold perturbation.
9) For each zone take a vote to select the AHU whose perturbation has a highest metric (produced a larger difference in sensor vectors).

The AHU selected with this method is expected to be associated with the VAV box.

4. RESULTS AND DISCUSSION

The proposed methodology was applied to the building explored with the other methods. The results were encouraging, with a recall rate (TP/(TP+FN)) of ~80%. Figure 4 depicts a room that was correctly classified. The 2D scatter plot (average room temperature variations are small) clearly shows that the distance between hot and cold points during perturbation of AHU 2 is larger than the others. Indeed, AHU 2 was correctly associated to this room. Baseline weekdays (WD) are also plotted for comparison in Figure 4, but they were used in the algorithm only for standardization (step 6).

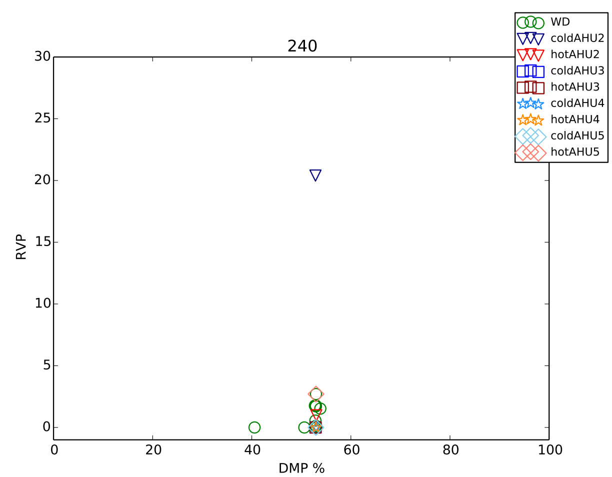

Figure 4. Room correctly associated by the new methodology.

Figures 5 and 6 show rooms that were not correctly classified. Zone 308 (Figure 5) presents a very tight distribution of variables, as DAM is at its minimum, and daily average RVP changes very little during all the periods. Further analysis of this case reveals very frequent oscillations (about 20/days) of RVP ranging between 0% and 40% (not shown). It is very likely that the control loop is out of tune for this room. This is likely the reason why the perturbation has no visible effect on the data. On the other hand, zone 461 (Figure 6) shows a very wide distribution of points. The room seem subjected to large unmeasured disturbances. Some baseline weekdays reach max RVP (~100%), while others use only a fraction of reheat (10-30%). These unknown factors probably had a larger effect compared to the perturbation, causing the room was misclassified.

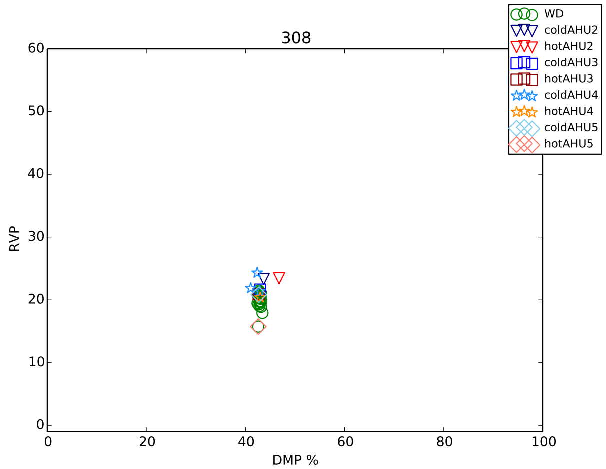

Figure 5. Zone 308 incorrectly classified by the new methodology.

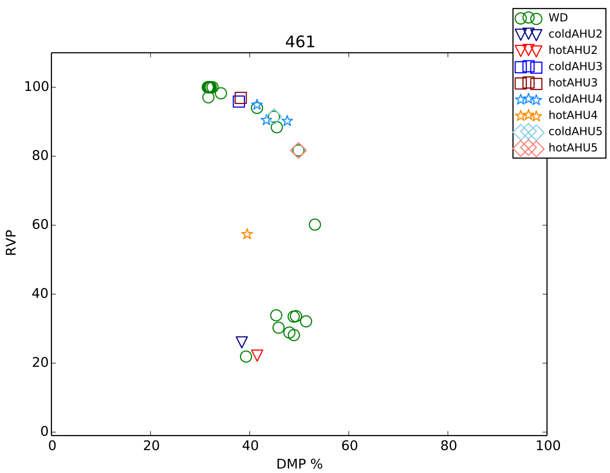

Figure 6. Zone 461 incorrectly classified by the new methodology.

5. CONCLUSION

In this paper we present a novel algorithm to infer relationships between HVAC components of large commercial buildings. We show that due to the characteristics of the data (response lags, nested control loops, tight variable boundaries) other common techniques are not effective in this context. The new algorithm utilizes perturbations of AHU variables and guarantees that the building zones remain within comfort. The method was applied to an existing building and its results, leading to ~80% recall rate, were compared with those obtained by other pre-existing methods. Future work should test alternative metrics to evaluate the distance between points, quantify uncertainty in the results, and applying this method to other buildings to verify whether it is generalizable.

6. ACKNOWLEDGMENTS

We would like to thank Miguel Sanchez, Ian Hurd, and Jacob Cabrera for their help coding some of these analyses.

7. REFERENCES

[1] Deparment of Energy (DOE), 2010a, Buildings Data Book Chapter 1, http://buildingsdatabook.eren.doe.gov/ChapterIntro1.aspx.

[2] R. Fontugne, J. Ortiz, N. Tremblay, P. Borgnat, P.Flandrin, K. Fukuda, D. Culler, and H. Esaki. 2013. Strip, bind, and search: a method for identifying abnormal energy consumption in buildings. (IPSN '13). DOI=10.1145/2461381.2461399

[3] M. Koc, B. Akinci, M. Bergés. 2014. Comparison of linear correlation and a statistical dependency measure for inferring spatial relation of temperature sensors in buildings. (BuildSys '14). ACM, DOI=10.1145/2674061.2674075

[4] B. Narayanaswamy, B. Balaji, R. Gupta, and Y.Agarwal. 2014. Data driven investigation of faults in HVAC systems with model, cluster and compare (MCC). (BuildSys '14). DOI=10.1145/2674061.2674067

[5] L. Ljung. 1999 System Identification — Theory For the User, 2nd ed, PTR Prentice Hall, Upper Saddle River, N.J., 1999

[6] N. Rajagopal, P. Lazik, and A. Rowe. 2014. Visual light landmarks for mobile devices. (IPSN '14). IEEE Press, Piscataway, NJ, USA, 249-260.

[7] D. Hong, J. Ortiz, K. Whitehouse, and D. Culler. 2013. Towards Automatic Spatial Verification of Sensor Placement in Buildings. (BuildSys'13). ACM, New York, NY, USA, Article 13 , 8 pages. DOI=10.1145/2528282.2528302

[8] A. Krioukov, G. Fierro, N. Kitaev, and D. Culler. 2012. Building application stack (BAS). (BuildSys '12). ACM, New York, NY, USA, 72-79.

[9] S. Dawson-Haggerty, A. Krioukov, J. Taneja, S. Karandikar, G Fierro, N. Kitaev, and D. Culler. 2013. BOSS: building operating system services. (nsdi'13), Nick Feamster and Jeff Mogul (Eds.). USENIX Association, Berkeley, CA, USA, 443-458.

[10] T. Weng, A. Nwokafor, and Y.j Agarwal. 2013. BuildingDepot 2.0: An Integrated Management System for Building Analysis and Control. (BuildSys'13). ACM, New York, NY, USA, , Article 7 , 8 pages. DOI=10.1145/2528282.2528285

[11] P. Arjunan, N. Batra, H. Choi, A. Singh, P. Singh, and M. B. Srivastava. 2012. SensorAct: a privacy and security aware federated middleware for building management. (BuildSys '12). ACM, New York, NY, USA, 80-87. DOI=10.1145/2422531.2422547

[12] D. Wheeler, Understanding Statistical Process Control Paperback. 2010. SPC Press.