Short Paper: A Method for Discovering Functional Relationships Between Air Handling Units and Variable-Air-Volume Boxes 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 association 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 Air Handling Units and Variable-Air-Volume Boxes from sensor data. The method utilizes perturbations of subsystem variables, while guaranteeing that the building zones remain within comfort boundaries. When applied to an existing building, our proposed method reveals correct associations in ~80% of the cases, and outperforms other 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. We spend ~90% of our time in buildings, consume ~75% of total electricity, representing nearly half of our primary energy consumption, and generating 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 embedded in the building automation system (BAS) or non-existent. 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 (e.g., 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.
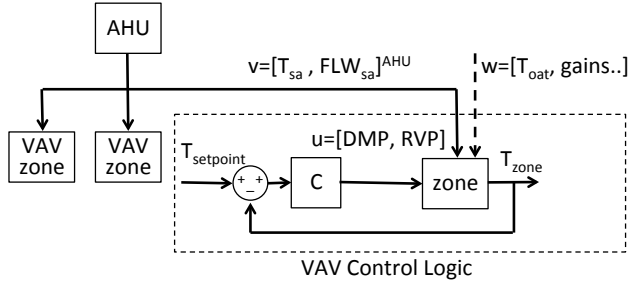
2. TESTBED

We analyze 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. One AHU is associated to multiple VAVs. VAV modulation is achieved through a combination of two control actions: adjusting the air damper 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 (T_{oat} is the outdoor air temperature), internal gains, and other thermal gains. This configuration is typical of a large number of buildings. 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. This paper explores different methods to discover the

functional relationship (i.e., the association) between the VAV box and the AHU that provides air to it.

Figure 1. AHU and VAV configuration. On the right, the VAV control logic is represented. T_{zone} is kept around the $T_{setpoint}$ by the control outputs u=[DMP, RVP]. T_{zone} is also influenced by parameters controlled by the AHU (v=[T_{sa}, FLW_{sa}]^{AHU}) and external disturbances w=[T_{oat}, internal gains …].

3. RESULTS WITH EXISTING METHODS

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. Table 1 shows an example of these coefficients. Results show very poor correlation. Desired values inside the bold boxes should be higher, in absolute value, than the corresponding values in the other rows. This method is able to correctly classify only 38% of the zones (Table 3). 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.

		Example VAV$_{AHU2}$			Example VAV$_{AHU4}$		
		T_{zone}	DMP	RVP	T_{zone}	DMP	RVP
AHU 2	T_{sa}	-0.06	0.06	0.11	-0.11	0.24	0.29
AHU 4	T_{sa}	-0.20	-0.19	-0.06	-0.04	-0.01	0.02

In contrast to prior research monitoring energy consumption [2], the variables 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 disturbances (Figure 1). For this reason, sensor values show a small signal to noise ratio. In addition, sensor readings are frequently constrained between physical limits (e.g., max damper position) and kept around setpoints by nested control loops. 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.

Next, we constructed feature vectors for the data of each VAV. Features included were: T_{zone}, DMP, RVP, $T_{setpoint}$, 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 and 3), yielding to 32% correct classifications.

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

		Example VAV$_{AHU2}$		Example VAV$_{AHU4}$	
		Eig1	Eig2	Eig1	Eig2
AHU 2	T_{sa}	0.12	0.12	0.20	0.09
AHU 4	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 input and output 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:

$$T_{zone,t} = \beta_1\, T_{zone,t-1} + FLW_t \cdot \sum_{t}^{t-k} \beta_{2i}\, RVP_i + \beta_3\, FLW_t \cdot T_{sa,t}^{AHU} + \beta_4 \cdot T_{oat,t}$$

where variable names are defined above, β_i are the statistical coefficients, and t stands for time. The equation is in the form of an autoregressive time series model with exogenous inputs and interactive effects (some variables are multiplied). The term with the sum represents the lag in the effect of the reheat valve. The model is based on detailed physical knowledge of the heat transfer processes in VAV boxes. Note that all the variables in this equation belong to the VAV with the exception of ($T_{sa,t}^{AHU}$), that represents the supply air temperature controlled by the AHU at time t. 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 [13] 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. Classification results are presented in Table 3. 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.

4. NEW 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 (reheat 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 (T_{zone}) will be affected and change. The AHU supply air temperature was chosen

Table 3. Percentage of correctly classified VAVs using different methods

	Random	**Correlation Raw**	**Correlation PCA**	**SID**	**Perturbation + Voting**
Correctly classified VAVs	25%	38%	32%	32%	79%

over the AHU flow rate as it is practically easier to tune. Also significantly perturbing flow rates might provide insufficient ventilation or damage the ducts due to high pressure. During the experiment, 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 (result found while using the other methods), we resampled the data in daily intervals. Average daily values for RVP, DAM and T_{zone} are still meaningful indicator of system behavior.

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 (e.g., ColdAHU2 means cold period for AHU 2). Each VAV data stream will have periods for all AHU perturbations.*
5. *Standardize each stream of data to have mean = 0 and standard deviation = 1.*
6. *For each VAV and each period combine data into average daily vectors of the form [RVP, DAM, T_{zone}].*
7. *Filter out vector elements that show a response value with opposite sign compared to physical intuition (e.g., a cold perturbation that reduced reheat).*
8. *For each VAV 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 perturbation difference (hot – cold).*
9. *For each VAV take a <u>vote</u> 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 that VAV box.*

One important intuition here was that evaluating a metric for each VAV box across all the perturbation periods (voting) would provide better results than assigning the zone to an AHU for each perturbation (independent classification), as discussed below.

5. RESULTS AND DISCUSSION

The proposed methodology was applied to the building explored with the other methods. The method correctly identified the relationship between VAV box and AHU in 79% of the cases.

5.1 Misclassifications

There were three main groups of zones that were incorrectly classified: (1) VAVs that showed no response to any perturbation, (2) VAVs that had an unexpected behavior, such as counter-intuitive responses to perturbations, (3) VAVs that responded more strongly to the perturbation of the 'wrong' AHU.

As an example of group (1), zone 308 (not shown here for brevity) 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 cycles per day) of RVP ranging between 0% and 40%. It is very likely that the control loop is out of tune for this zone. This might be the reason why the perturbation has no visible effect on the data.

Among the zone exhibiting counterintuitive responses to perturbations (group (2)), zone 461 shows a very wide distribution of points (Figure 2). The zone seems subjected to large unmeasured disturbances. Some baseline weekdays reach max RVP (~100%), while others use only a fraction of reheat (20-35%). These unknown factors probably had a larger effect compared to the perturbation, causing the zone to be misclassified.

5.2 Independent classification *vs.* voting

Adopting a voting system compared to scoring each perturbation individually increases accuracy and precision (with one exception), as illustrated in Table 4. In addition, the classification of each AHU individually, based on its corresponding perturbation (P_i is the perturbation to AHU_i), requires defining a threshold to compare the calculated metric to. Choosing the right threshold is challenging and hard to generalize. In Figure 3 each line represents a VAV box sorted by ground-truth AHU, while each column represents a perturbation of an AHU. The color in each cell is shaded proportionally to the value of the calculated metric, normalized by column. Darker colors represent higher metric values that correspond to stronger relationships between the VAV box and the perturbed AHU. It is clear from the figure that each perturbation produces responses in multiple VAV boxes, some of which are not associated with the correct AHU. Choosing a threshold corresponds to choosing a minimum shade of grey that triggers the association with the AHU currently perturbed. Further, Figure 3 shows that perturbations 2 and 4 (in column) require a very different threshold to correctly classify the corresponding VAVs. In contrast, Figure 4 illustrates the clearer heat map generated with voting. Dark areas are more clearly defined, and it is also visually easier to identify which VAV boxes are associated to the AHUs. Furthermore, if a zone is classified in different ways under separate perturbations (e.g., belonging to two AHUs) then the classification conflict needs to be resolved. When a voting system is adopted, a zone is classified univocally.

Table 4. Comparison of accuracy / precision of the voting method and the independent scoring method for each AHU

Accuracy/ Precision	Independent classification		Voting
	Threshold = 0	**Threshold = 1 SD**	**Max Value**
AHU 1	39% / 31%	78% / 60 %	85% / 77%
AHU 2	35% / 16%	92% / 63%	94% / 77%
AHU 3	22% / 17%	83% / 48%	92% / 69 %
AHU 4	48% / 45%	88% / 85%	86% / 85%
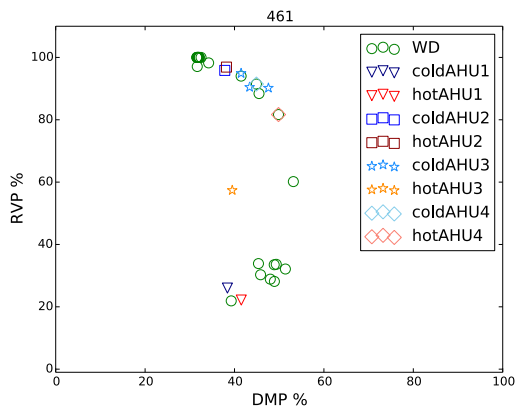
Figure 2. Zone 461 incorrectly classified by the new methodology. Results projected in 2D.

Figure 3. Treating each perturbation as a stand-alone experiment.

Figure 4. Advantages of using a voting method.

6. CONCLUSION AND FUTURE WORK

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 could identify the relationship correctly in 79% of the cases. There are three main directions of research we would like to explore in the future. First, we would like to explore how our technique generalizes to other buildings, where zones may have different sensors, setpoints, and control loops. Second, we would like to explore whether we can discover relationships between subsystems like AHUs and VAV boxes without having to introduce a large perturbation to the AHU. We may be able to utilize milder perturbations like daily occupant cycles, and techniques from statistical process control to infer the same relationships. Finally, a useful future direction of research is to quantify the amount of cross-talk between zones. The heat exchange between zones with common doors, passageways, etc. introduces errors into any single-zone heat exchange analysis, and its identification and quantification could lead to more efficient and better designed control loops.

7. ACKNOWLEDGMENTS

8. 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
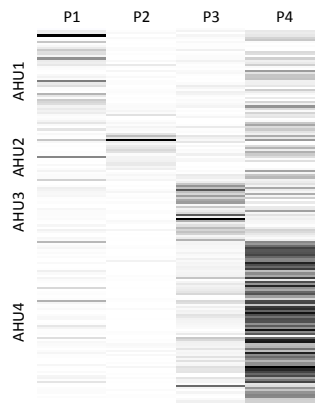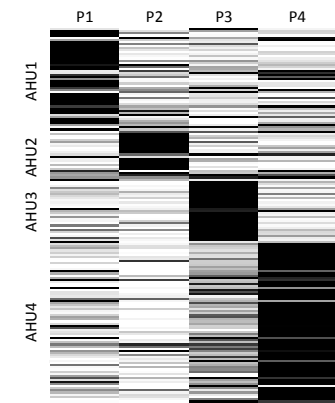
[3] M. Koc, B. Akinci, M. Bergés. 2014. Comparison of linear correlation and a statistical dependency measure for inferring spatial relation of temperature sensors in buildings. (BuildSys '14). ACM, DOI=10.1145/2674061.2674075

[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.

[13] Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B, 58, 267–288.