Data Driven Investigation of Faults in HVAC Systems with Model, Cluster and Compare (MCC)

Balakrishnan Narayanaswamy[†], Bharathan Balaji[†], Rajesh Gupta[†], Yuvraj Agarwal[†‡]

[†]University of California, San Diego   [‡]Carnegie Mellon University

[†]{muralib, bbalaji, rgupta}@cs.ucsd.edu,   [‡]yuvraj.agarwal@cs.cmu.edu

Abstract

The complexity of modern HVAC systems leads to device mis-configuration in about 40% of buildings, wasting upto 40% of the energy consumed. Fault detection methods generate excessive alarms leading to operator alert fatigue, faults left unfixed and energy wastage. Sophisticated fault detection techniques developed in the literature are seldom used in practice. We investigate this gap by applying various fault detection techniques on real data from a 145,000 sqft, five floor building. We first find that none of these algorithms are designed to capture control loop configuration faults. We develop a novel algorithm, Model, Cluster and Compare (MCC) that is able to detect anomalies by automatically modeling and clustering similar entities in an HVAC system, in an unsupervised manner, and comparing them. We implemented MCC to detect faults in Variable Air Volume boxes in our building, and demonstrate that it successfully detects non-obvious configuration faults. We propose a two stage approach, where we design *intelligent rules (iRules)* based on anomaly exemplars from a mix of data driven algorithms. iRules are successful in capturing a large fraction of faults in our building, with only one false alarm and 78 anomalies detected out of 237 zones. Thus, comparative data mining is useful in filtering the large amount of data generated in modern buildings, but that human in the loop systems are better still.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications-Data mining

General Terms

Algorithms, Measurement

Keywords

Anomaly detection, Smart Buildings

1 Introduction

Buildings are large and complex systems with thousands of moving parts including fans, water pumps, dampers and many more [2]. Improvements in sensing, communication and control infrastructure offer hope that increased instrumentation and fine grained control will improve operational efficiency [19]. However, this also increases the number of devices that are a part of the HVAC system and its complexity. This increase in complexity also increases susceptibility to faults and configuration errors. In total, it is estimated that 40% of buildings have mis-configured devices [20] and that there is a 40% energy savings potential from fixing building faults [22]. In addition to energy waste, Heating, Ventilation and Air Conditioning (HVAC) faults also lead to occupant discomfort, reduced indoor air quality and equipment deterioration [17]. As a result, Fault Detection and Diagnosis (FDD) is becoming important.

Modern buildings typically use a supervisory computer based Building Management System (BMS) system that coordinates its sensing and control infrastructure, primarily for managing the HVAC system [5]. Most commercial BMSes already provide fault detection features [7], but use simple detection techniques such as threshold monitoring of individual sensors. In discussions with our campus facility managers we learned that, typically, these BMSes generate more than a 10000 alarms *per day* on the UC San Diego campus. This deluge leads to operator alert fatigue where most alarms are ignored, and many faults are detected only via occupant complaints or during a separate, expensive, commissioning process. Many faults remain undetected for long periods of time leading to equipment deterioration and energy wastage, thus motivating automated FDD.

Many algorithms have been proposed in literature to capture various faults that occur in HVAC systems using available sensors [14]. These methods include domain knowledge based simple rules [18], physical modeling of HVAC system [4], data driven models using machine learning techniques [15] and Principal Component Analysis (PCA) [12]. Although there is considerable evidence that these algorithms detect anomalies successfully, they are not used for our university buildings, nor supported by popular vendors [7]. Our work is motivated by this gap between fault detection theory and practice, to help identify which techniques actually work in real deployments.

As a first step, we investigate different data driven fault

detection techniques on the sensor data collected from the HVAC system, focusing on the faults present in the Variable Air Volume (VAV) boxes. VAV boxes are among the most numerous components of the HVAC system, but have received relatively less attention in the FDD community due to a focus on the larger Air Handler Units (AHUs). Each VAV box covers a small physical area within a building, called a *zone* (see Section 2 and Figure 1). Our first finding is that (see Table 2, Section 5) for specific faults, rule based approaches offer excellent coverage with low false alarm rates for appropriate thresholds. However, if generic rules are written to cover many faults, which may not be applicable to a particular buildings situation, they result in a large number of false alarms. Writing specific rules for all possible faults that may occur in a particular building requires considerable knowledge about the building location, equipment and usage. Since this is often unavailable in practice we turn to data driven models which offer a reasonable substitute for expensive expert supervision. In particular we use well established subspace methods (PCA) [9] and recent work on correlation change detection [10] as applied to our set-up.

These data driven algorithms come with their own drawbacks. PCA flags zones that are extreme in their sensor measurements, such as large zones (Figure 2). Change detection methods unsuccessful in our deployment since they are sensitive to changes in confounding parameters like external temperature (Section 3.2). In addition, they cannot detect faults which were present from the start of data collection. Many of the VAV faults we discover are of exactly this form. To locate *long standing* faults *without* labeled anomalies or building models we that argue that *comparative* data mining is a useful approach in large commercial buildings. The intuition behind our algorithm is that in a large building there will be, for most zones, a reasonable number of similar zones. If these clusters of zones are identified, automatically from sensor data, then zones that are dissimilar from any cluster are probable anomalies. Based on this intuition, we propose *Model, Cluster and Compare* (MCC) - an unsupervised algorithm for automated fault detection. In particular we, (a) develop zone-centric black-box *models* using sensor data, (b) use *clustering* techniques to automatically identify zone categories within a building, and (c) use data driven methods for *comparing* zones to flag anomalous zones while limiting false alarms. We show that (i) each of these steps is useful in anomaly detection in practical systems, (ii) the clusters can indeed be learned automatically from minimal data and (iii) the anomalies detected by this procedure are meaningful.

We find that, in combination, data driven methods discover three classes of faults previously unaccounted for in the standard reference for rules for VAV fault detection[11]. While these are recognizable as faults by experts, and prevalent throughout our building, they are not commonly tested for. Thus, data mining is useful in filtering the large amount of data generated in modern buildings to generate fault exemplars with minimal human supervision. At the same time, we find that each algorithm is effective in detecting certain types of faults, but is unable to detect all the faults that occur on their own. In addition, none of the data driven algo-

rithms we study *group* similar faults well. That is, instances of the same fault (or even the same zone on different days) appear in different positions on the list of faults depending on (un-sensed) confounding parameters. There is thus a balance between the labor required, and specificity of, manual inspection of anomalies with rules and the different automatic anomaly detection algorithms. To get the best of both worlds, we propose a two stage approach where these faults are then presented to users who then come up with a set of *intelligent* rules (iRules) based on them.

We validate our algorithms on one year of data from a modern 145,000 square feet commercial building with over 4000 HVAC related sensors distributed across 237 zones. Using just one month of data from a single floor of our building test-bed we locate two new classes of faults with MCC (alternate heating-cooling cycles and excess airflow when unoccupied - Section 5) which were not found by rules or the baseline methods. A third class of faults (short cycling anomaly - Section 5) was discovered through correlation based analysis. Given these fault exemplars, building and fault specific iRules are then used to detect faults in the larger data, while limiting false positives. We believe that this two step process will narrow the gap between theoretical fault detection methods and their use them in practice. Based on the exemplar anomalies detected by MCC and other algorithms, we deployed iRules and showed that these faults are are surprisingly prevalent in our modern building testbed. We found 78 faults in 237 zones, with one false alarm.

2   Background and Related Work

Building HVAC systems: HVAC systems are designed to provide a comfortable, safe and productive environment to building occupants in an energy efficient manner. Centralized HVAC system with Variable Air Volume (VAV) have become prevalent since they are easier to maintain and allow independent control of thermal zones. To understand the scale of centralized HVAC systems, and the potential for faults we studied our own test-bed building – the 145,000 sq-ft CSE building at UCSD which was commissioned in 2004. CSE is supported by a Johnson Controls Metasys which uses the BACnet protocol to connect >4000 sensors acrss 237 HVAC zones, each served by an individual VAV box.

VAV terminal units: Figure 1 depicts the control inputs and sensor data points for one such VAV box. VAV boxes are zone specific flow control devices, with associated sensors (temperature and airflow), actuators (dampers, fans), set-points (cooling, heating, airflow) and control signals (occupied command, heating command, cooling command). In total, in our system we have 17 sensor points per zone. Our VAV box also returns a "Occupied Command", which is a setting which can be in Occupied, Standby or Unoccupied modes. Each level is associated with increasing larger guard bands around the set-point. The purpose of the VAV box is to maintain the zone temperature below the cooling set point and above the heating set-point subject to safety constraints. These safety constraints are usually in the form of setting of minimum airflow, which take different levels when occupied, not occupied and in standby. The amount of heating and cooling required is relayed to the central HVAC through
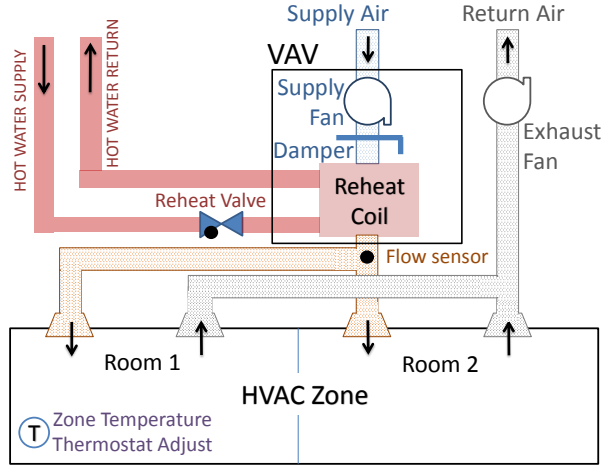
Figure 1: VAV with reheat system used for controlling the temperature of discharge air in each HVAC zone [3].

heating and cooling commands.

Faults in VAV units: Faults in the VAV controller can be divided into the hardware and software faults. Hardware faults include sensor failure (temperature, airflow, damper position), broken or stuck damper, or a leaking valve. Software mis-configuration cause issues such as over cooling, wind chill or undercooling in addition to hardware deterioration, discomfort to occupants and energy wastage. Detecting these faults automatically from large amounts of VAV sensor data becomes an important challenge.

Anomaly detection in HVAC systems: Traditionally, HVAC fault analysis have been focused on degradation of equipment such as heat exchangers, rusting of fans, mis-calibration of sensors or misconfiguration of the control system itself such as incorrect temperature setpoint, or a forgotten operator override to keep a pump running [13]. We focus on the faults in the VAV boxes themselves, rather than the AHUs and other central HVAC components for two reasons. First, our work shows that VAV systems have a large number of faults that are not fixed, due to both a historical focus on AHUs (which constitute the heavy equipment in the HVAC system) and due to the large number of false alarms generated by simple rule based systems when they are applied to VAVs. Second, there are typically a large number of VAV units in each building as exemplified by our building above. These are individually and manually programmed, leading to errors both in initial deployment and over time as changes are made based on occupant complaints.

2.1 Prior work in FDD

Anomaly detection, the process of detecting non-conformal patterns in data, has a rich history [6] and these can directly be used for Fault Detection and Diagnosis (FDD). Conventional methods for this are based on physical models, rules or process history [24, 14]. More broadly, depending on whether labeled examples of faults are available, FDD algorithms are *supervised* or *unsupervised*.

Model based supervised methods are usually based on physical model learned in a lab and are not adaptive to changes in building operation [15, 14]. Models need to be built to capture all the conditions faced by an operational building at different locations, e.g., different zones in a building can have vastly different model behavior. Rule based systems are among the most widely used supervised FDD methods in buildings [21, 11]. The 7 alarms from [11] (page 45) that are applicable to our VAV boxes form the Rules baseline in our work. We find that rules and thresholds need to be building specific, as otherwise they lead to many false alarms. While a rule can be written to catch any anomaly, a large number of rules is required to cover all possible faults. Rules that test the dynamics of zone controllers are not common.

Among unsupervised methods, sub-space based methods, exemplified by Principle Component Analysis (PCA) [23] form the basis for most process history based methods [24] and building systems [25, 9]. PCA, with variable selection, allows us to localize the directions in the space spanned by sensor measurements that contains most of the typical variation. A data point, or group of data points, that conforms with the normal correlation structure of the data will have most of its variation in these principle components. The residue when sensor measurements are projected on to these directions is then a measure of the a-typicality of those readings. If certain sensors, rooms or days have large residue then they can be labeled as anomalies.

Process history based methods, which attempt to replace manual models with learned models require lots of training data, often require labels of anomalies and are usually evaluated only on simulated data [12]. Most such methods largely have the same form and involve comparing a calculated statistic (like mean or variance) for a current window with the same statistic on a past window of the data. If these values don't agree then the models have changed and this is flagged as an anomaly. We found that change detection based methods are heavily affected by changes in confounding parameters such as external temperature and occupancy (Section 3.1). Recent data driven work of this form for buildings is Strip, Bind and Search (SBS) [10] which conducted a large scale study of anomaly detection in plug loads. SBS uses the correlation between energy consumption across plug loads to define features, followed by euclidean distance between correlation matrices as a measure of dis-similarity. We extend this to the total correlation between all normalized sensors in a zone, forming the correlation baseline. These methods miss faults that have existed since deployment as well for newly deployed buildings where commissioning artifacts make a zone mis-configured from the the start - which form the focus of our work. In contrast to these methods, we also compare zones with each other and not just zones with their own historical trends.

3 Analysis

We now discuss the need to exploit similarity between HVAC components, which lead us to the design of our algorithm MCC. The analysis below of a variety of data driven FDD algorithms serves to justify both the need for a new algorithm (MCC) and for a combination mechanism (iRules).
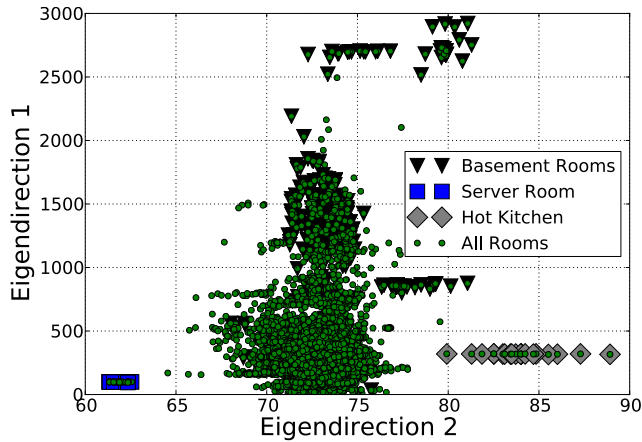
Figure 2: Scatter plot of first two eigen-directions of sensor data. X axis is essentially room temperature, and y axis is essentially Supply Flow

3.1 The need for models

Principal Component Analysis (PCA) based FDD is based on the intuition that anomalous sensor readings are very different from most normal data. Faults are detected by finding sensor data that are 'far away' from the center (mean/median) of the principle components of the sensor data. However, this approach has its shortcomings. In Figure 2 we show how points are scattered along the first two eigenvector directions of all 17 sensors we use. This would correspond to PCA with dimension 2, though the intuition extends to higher dimensions. The projection is for illustration and not as conclusive proof. Since among the sensors the largest variance is in airflow and zone temperature these essentially correspond to direction 1 and 2 (x and y axis) in the figures above.

An outlier detection method run on these rooms, flags three anomalous regions (i) **Basement rooms:** which are large rooms and require larger amounts of airflow to maintain temperature. These are not anomalous in configuration, but in size and location. (ii) **Server Room:** This is a zone that must be maintained at a lower temperature consistently. This is not a fault, but a design choice. (iii) **Hot Kitchen:** This is a kitchen which is open and consistently reaches temperatures beyond 85 degrees. The hot kitchen anomaly is flagged by all the algorithms. Investigation revealed that in this room the water cooler/heater was placed next to the thermostat. As a result, it always read as hot leading to constant cooling and energy wastage. Moving the water cooler provided an immediate fix. In conclusion, sensor readings directly are not good features for anomaly detection. We use models of sensor readings (Section 4) and show that the model parameters are more suitable for FDD (Figure 5). We see that looking at only the process history of a single zone, raises many alarms.

3.2 The need for comparisons

The intuition behind process history based methods (PCA and Strip-Bind-Search for example) is that anomalies represent changes in process history. In SBS the 'typical' correlation structure between sensors is learned and sensors that deviate substantially from their historical correlations
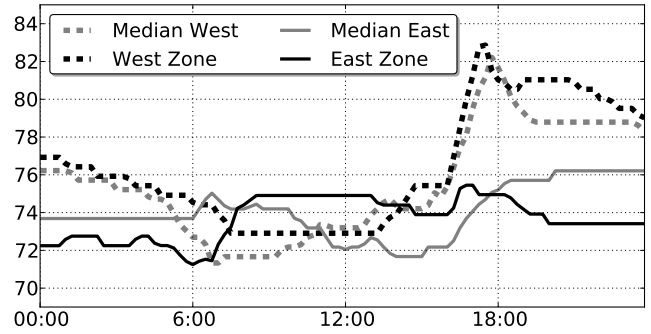

Figure 4: Difference between zone temperatures of east facing and west facing zones on different floors.

are flagged as anomalous. We extend this basic algorithm to multiple sensors (as described in Section 5). When run on our data, the top anomaly classes discovered (based on changes in correlation structure) are shown in Figure 3.

Figure 3a is a DR event which effected the entire building, and Figure 3b is a time interval when the sensor data collection was briefly offline. Figure 3c, a particularly hot day, indicates the sensitivity of these algorithms to confounding parameters (in this case external temperature). All these are anomalies when compared to historical sensor traces for the same zones, but not as anomalous when compared to sensor traces of other zones. Thus, comparing zones before flagging anomalies may be useful. The anomaly in Figure 3d is important, and not tested for in our rules, showing the benefit of change detection methods.

3.3 The need for clustering

In Figure 4 we see that west facing facing zones tend to get hotter faster and at different times when compared to east facing zones - naturally leading to different models and sensor readings. Thus, a direct application of anomaly detection methods (without clustering) would flag certain room types as anomalies because their models will be very different from a 'typical' room. As another example, consider again Figure 3c. Setting a single threshold (i.e. assuming all zones are the same) labeled rooms like Figure 3c as anomalies when used with correlation or PCA based algorithms. If multiple clusters are allowed, then it is likely that west facing rooms would form a single separate cluster, and such zones will not be flagged as anomalous.

3.4 The need for intelligent rules

One problem with data driven methods, that we have not seen emphasized in the literature is one of *grouping*. Most FDD algorithms we use come with an associated Figure Of Merit (FoM) which measures how anomalous a measurement is. For PCA this is the residue after the selected principle components are removed and for correlation based clustering this is how far the correlation matrix for a zone-day is from the average. Faults are sorted based on this FoM and presented to the user. However, we found that the same kinds of faults (all the DR event rooms like Figure 3a, all the dynamic anomalies like Figure 7 or all rooms with large setpoint violations with rules) do not have similar FoMs. As a
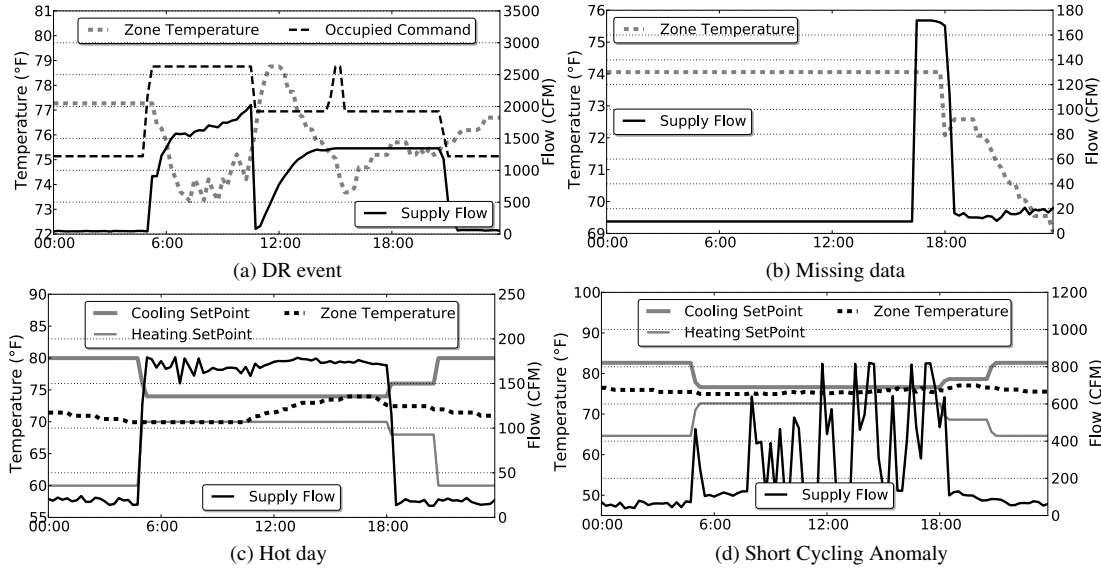
Figure 3: Anomalies detected by correlation based model with process history comparison. (3a) Due to a Demand Response event HVAC systems in the entire building were turned down. Each room was flagged as an anomaly. (3b) Due to problems with data collection engine, data for all sensors is missing for a period. Each room was flagged as an anomaly. (3c) On a hot evening all the west facing rooms heated up, though not much beyond their set-points. Not anomalous behavior, but all rooms were flagged. (3d) Rapid increases and decreases in air-flow near the set-point, with damper position cycling. This is a fault that reduces hardware life.

result, they are not grouped together. In addition, if the same zone is anomalous on different days (typically due to confounding parameters of external temperature or occupancy) these are not grouped together. Grouping is particularly useful (and in fact necessary) for building managers to identify classes of faults, their impacts and potential fixes. For example, in our larger experiments on 237, zones which are instances of Figure 7, when sorted by FoM, appear in positions 2 through 140 and the DR anomaly Figure 3a detected by correlation based methods, appears in positions 1 through 300). Not all were declared anomalous by the respective algorithms and thresholds.

This is related to but distinct from the problems of blame assignment or root cause detection [24]. While these are important directions of research, they are unsolved. We follow an alternative route. Presented with the anomalies detected by different data driven algorithms a user can quite easily write rules to find other similarly anomalous zones. Since FoMs are more meaningful for a single anomaly class rather than when comparing different anomalies, these can be sorted and fixed in order of severity.

4   Our Approach

Based on this intuition, we develop our MCC algorithm that *models* the inter-relationship of sensors in a VAV box, *clusters* these model parameters and *compares* them to identify anomalies. Based on the outputs of MCC (and other algorithms) we write intelligent rules (iRules) capable of detecting a specific class of anomalies in the entire data set.

The primary objective of MCC is to determine clusters of zones in a buildings and use these clusters to determine mis-

configured, anomalous or faulty zone controller configuration automatically. A single zone in a building has multiple sensors and actuators (temperature, air-flow, damper position etc.). These sensors and actuators observe and act upon the same spatio-temporal space. As a result they will have *analytical redundancy* [24]. We use black-box statistical *models* to learn the relationship structure induced by this redundancy, directly from sensor data. The structure and strengths of these relationships will be a function of both observed sensor values and *confounding* parameters including sensor noise and room usage, size and orientation.

While the confounding parameters are unobserved, MCC is based on the intuition that they will be similar for subsets of rooms in a building. For example, some rooms in a large building will be of approximately the same size and some sub-sets will have the same orientation and usage patterns. We show that these groupings or sub-sets can be automatically determined by *clustering* in an unsupervised setting with only unlabeled data. Zones whose learned models, by *comparison*, do not lie in any cluster consistently, are identified as faulty. From our initial analysis (Figure 2), we found that using measured sensor readings directly does not detect faults accurately because the properties are dominated by the highest variance sensors. This holds even when sensor data is normalized. This problem can be avoided if the sensor data are used to predict (or forecast) a common quantity. The quantity of interest that is predicted may be one of the sensor readings at one time step into the future or the energy consumption of the zone (1). The parameters of this model will naturally normalize different sensor readings based on their

predictive power of a quantity of interest.

Given sensor readings $\vec{x}_t^i \in \mathbb{R}^s$ from s sensors for room i at time step t, the outline of our approach is then (i) Build a model $x_{t+1}^i = f(\theta_{t-\Delta:t+\Delta}^i, x_{t-W:t}^i) + n_t^i$, parameterized by $\theta_{t-\Delta:t+\Delta}^i$ for a window W and a horizon Δ and with error (due to confounding parameters) of n_t^i, (ii) Cluster the parameters $\theta_{t-\Delta:t+\Delta}^i$ into k clusters, (iii) Identify anomalies as points that are not well represented by clustersm (iv) Based on the identified anomalies, write rules that can detect similar anomalies.

Model: Given sensor readings over a horizon (in our experiments we naturally use 1 day horizons) we construct, for each day n and room i a linear model of the form (1). We comment here that in prior work [10] occupancy was considered a confounding variable requiring careful signal processing, filtering and feature selection. However, in our VAV system Occupied Command is available as a sensor reading and this complication is largely eliminated. In addition, they were concerned with only power consumption sensors, whereas we have 17 complementary sensors per zone requiring some extension of their methods.

In our experiments we use a horizon of 1 day and window of 1 sample point. We choose $f()$ to be linear, thus

$$x_{t+1}^i = \Theta x_t^i + n_t^i \tag{1}$$

We use a more restricted choice, which works well in practice, which is to use the physical sensor and actuator readings to forecast a virtual sensor that corresponds to energy consumption of that particular zone. While HVAC energy consumption dis-aggregation is a hard problem, we use ZonePac [3] which estimates the heating, cooling and electrical power consumption of each zone in a Variable Air Volume (VAV) type system using existing infrastructure sensors installed as part of the Building Management System (BMS). In particular, the cooling thermal power $Q_{cooling} \propto q * (T_{zone} - T_{supply})$ where q is the airflow and T_{supply} is the supply water temperature. Similarly, $Q_{heating} \propto h$ where h is a PID signal called Heating Command. Thus, we learn a function that forecasts energy consumption given the sensor readings. Since we use a linear model, we add a virtual sensor that is the product $q * (T_{zone} - T_{supply})$. The parameters of this function are then used to detect anomalies.

In order to visualize the model parameters, we plot the projection of the model parameters into the first two dimensions of their eigenvector space in Figure 5, essentially doing PCA on the model parameters instead of the sensor space as was done in Figure 2. First of all we see a clearer separation of different kinds of rooms. Namely, kitchens, halls and restrooms which have very different PID settings than other rooms. Again, while this is not proof, in Sections 5.1 and 5.2, we discuss that the outliers (or anomalies) in this space are interesting faults. We will show that such a model is able to uncover more configuration anomalies, beyond those that are extreme in temperature or air-flow.Dynamical models for fault detection are not commonly applied to VAV systems, particularly in the comparative form we use here.

Cluster: We cluster the points using the k-means++ algorithm [1] and 4 clusters. A problem with vanilla k-means is
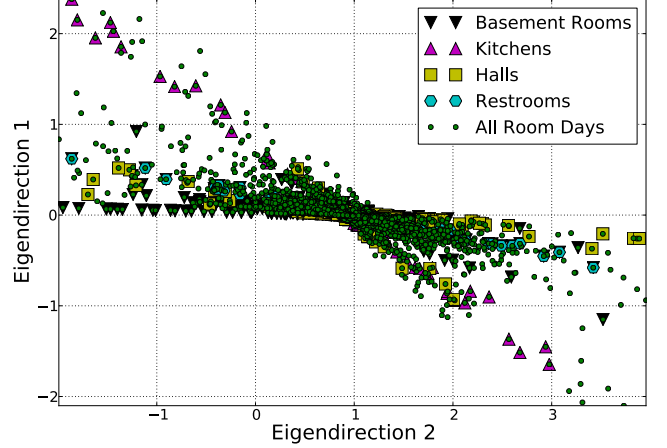

Figure 5: Scatter plot of first two eigen-directions of model parameters, for model built on daily data. Axis are uninterpretable, but clustering and separation are clearer

that different runs with different initializations lead to different clusters and anomalies. This is addressed by k-means++. Each cluster is expected to capture zones with similar confounding parameters.

Compare: Define r as the mean of radius r_i around each point i that has $k = 4$ neighbors. Based on the stability of good clusters intuition (above and Section 4.1), and prior work on anomaly detection, we define a point as anomalous if (i) there are no normal points within a radius r or (ii) the zone changes cluster membership often in consecutive days or (iii) whose points on different days have large span (we use greater than 3 times the variance in span as a threshold throughout). Here span is defined as the maximum (euclidean) distance between the model parameters. Thus, according to (i) anomalous points are 'far away' from normal points and according to (ii) and (iii) 'behave' differently. The motivation for (ii) is discussed in Section 4.1.

4.1 The stability of good clusters

In Figure 5 we have the projection of 30 days for 237 zones (7110 points). We see that most of the points are tightly clustered, with a small number forming outliers. We provide some justification for the spread of data points for rooms that are more heavily influenced by (correlated) confounding parameters. Consider the effect of confounding variables on the model in (1). Suppose the true model is

$$\vec{x}_{t+1}^i = \Theta_n^i \vec{x}_t + B_n^i \vec{d}_t^i + \vec{u}_t^i \tag{2}$$

where $\vec{d}_t^i$ represent confounding variables. If we fit the true model (2) with the assumed model (1), the estimated parameters $\hat{\Theta}_n^i$ will be of the form

$$\hat{A}_n^i = \Theta_n^i + \Pi_{\vec{x}^i}(\vec{d}_t^i) B_n^i \tag{3}$$

where $\Pi_x(d_t)$ is the projection of d_t into the space spanned by the x_ts. i.e., our estimate of the model parameters, will be shifted from the true Θ_n^i linearly by a quantity which depends on correlation between the sensors and the confounding parameters. This is seen clearly in Figure 5, where we see that the parameters corresponding to rooms that are heav-

55

ily effected by confounding parameters, zones like hallways and kitchens where usage and occupancy can be very different from day to day, form lines in 2D space. This indicates that these special zones are probably where additional sensing (e.g., occupancy detection) would be the most valuable - an observation requiring further exploration. While a linear shift is indicative of confounding parameters, a large spread is still a possible 'anomaly' or 'non-traditional' room.

5 Evaluation

We evaluate MCC and contrast it with PCA, correlation based change detection and rule based methods on data from our VAV boxes in our building test bed. We obtain ground-truth information by manual inspection of sensor data by an expert, verified by the building manager, together with physical verification for some faults. There are two properties of anomaly detection algorithms that are of interest - the number of missed faults and number of false alarms. For PCA we computed the first 4 principle components of the 17 sensor data readings for the entire data. The Figure Of Merit (FoM) of a data point was the Root Mean Square of the residue (i.e. projection into the remaining 13 directions). For the correlation based method, we computed the total correlation between all 17 sensors (after magnitude normalization) between each pair of days, resulting in a correlation matrix of size nZones x nZones for each day. nZones was 54 for the fourth floor and 237 for the entire building. Following [10] we defined the FoM for a room-day to be the euclidean distance between the median correlation matrix and the correlation matrix for a day. For all algorithms, zones where the Figure Of Merit (FoM) was greater than the median + 3 times the variance were flagged as anomalies.

For correlation based methods, we found that in our 30 day sample, a direct application of the algorithm generated many false or uninteresting alarms (as described in Section 3.2). To compensate for this, we ran the algorithm separately depending on whether the VAV box was in occupied mode (based on the Occupied Command sensor) and for weekdays and weekends after removing the missing data and DR event. As illustrated in Figure 2, PCA is biased toward rooms with high air-flow and temperature extremes so in addition to this we removed basement rooms, the hot-kitchen and the server rooms before generating the results below.

Based on prior work in the literature [11], and the rules currently in place in our system we derived an initial rule set. The rules we used test for the faults described in Table 1. These rules compare various combinations of set points (temperature and airflow) with their corresponding sensor readings. The rules were run on each zone for each day. The magnitude times the duration of rule violation was used as the FoM. The SFE, HCA and SCA faults were categorized after analysis of the results, as it was not clear a-priori that they may be present. i.e. while to building managers they are a-posteriori obvious classes of faults, and easy to identify manually, they are not commonly tested for. Since classifying and finding signatures of (and rules for) all possible fault types is difficult enough to be infeasible in practice, data mining is a critical tool in identifying and filtering fault exemplars for which rules can then be created.

5.1 Quantitative Evaluation of MCC

We first discuss some representative anomalies, that we found interesting or prevalent. The examples in Figure 6 also serve to exemplify the kinds of anomalies detected by MCC. While some of these anomalies are identifiable from a single days worth of data, since we have 237 zones to compare, we report on anomalies identified during a 30 day period. This is because the conditions for an anomaly to manifest itself do not always arise (e.g. room not occupied, external temperature too high). In the zone depicted in Figure 6a the actual supply flow and set points follow a complemented behavior where comfortable conditions are maintained when the zone is in unoccupied mode. This fault wastes excessive amount of energy, and existed from the time the building was operational. Rule based methods did not detect this anomaly, since the temperature and airflow is typically within set-points. This fault is missed by process history based algorithms, since there is no change from normal to faulty. However, since only a single room in our data set has this control structure it is flagged by MCC.

In Figure 6c, the zone temperature is not outside the set-points. So it is not flagged by rules. However, at fixed times the airflow increases, as verified by the decreasing temperature despite sunlight hours, cooling the room even when it is within set-points. In addition, we see that the set points never change. In Figure 6b the air flow is higher than similar rooms, though the difference when occupied and unoccupied is small, causing temperature to decrease below the cooling set-point causing discomfort. These anomalies (possibly obvious in hindsight) were found to be prevalent in our building and are a large energy waste.

The anomaly in Figure 6d was detected by MCC as well by rule and PCA based methods and is the 'most' anomalous zone in our dataset. This is the 'Hot Kitchen' example discussed in Section 3.1. Despite maximum airflow the zone temperature is measured as too high due to the placement of a water cooler near the thermostat. This is an example of a real, large and continuous energy waste not detected by change detection methods. The anomaly in Figure 6e is one where the Supply Flow set point does not change with the occupied command. While the room functions properly when unoccupied and the zone temperature is within limits, the non zero Supply Flow when unoccupied results in energy waste and over-cooling. This anomaly was also prevalent in many zones in our building. Finally, we turn to the anomaly in Figure 6f. This is actually the same day as the false alarm raised by the correlation based method in Figure 3c (observe the sudden raise in zone temperature at approximately 14 : 30). But we see that this room *is* actually anomalous since its temperature is below the heating set point for most of the day, but the fault lost in the false alarms. MCC detected this anomaly but did not mark the other zones on the same day as the correlation based methods did, thus, reducing false alarms and alert fatigue.

5.2 Dynamic anomaly detected by MCC

Rule based systems are simpler to set up for domain experts and catch a number of faults that are commonly seen in practice. However, rules based systems are also not explicitly comparative, and as a result require deep domain

Fault	Explanation
Heating or Cooling is Ineffective (HCI)	Large airflow not sufficient to bring temperature back to setpoint (Figure 6d)
Supply Flow Excess (SFE)	High flow when temperature between set-points and unoccupied (Figure 6c)
Heating-Cooling Anomaly (HCA)	Excessive cooling causes heating requirement (Section 5.2)
Temperature and Air Flow not within set points (SP)	Self-explanatory, (most Figures in 6 particularly Figure 6f)
Simultaneous Heating and Cooling (SHC)	self-explanatory
Too Many Mode Switches (MS)	Set-points switching rapidly between levels
Short Cycling Anomaly (SCA)	Rapid and unnecessary changes in air flow (Figure 3d and 6e)

Table 1: Fault classes of interest. SFE, HCA and SCA are based on the faults discovered by the data driven algorithms described here, the remaining 4 are from [11].
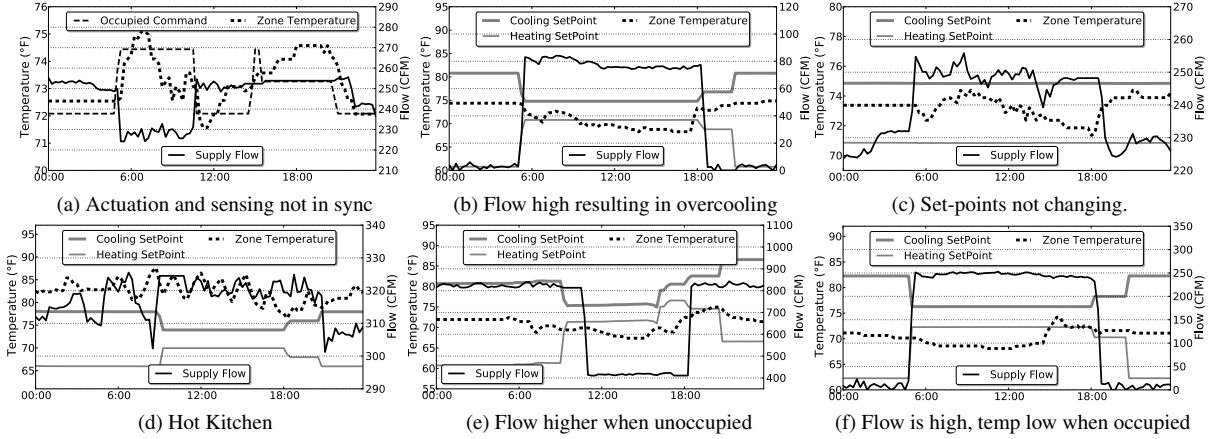

Figure 6: Selected anomalies discovered by MCC.

knowledge to set rules and thresholds. Rules are often not *dynamic* in that they usually compare sensor and actuator values to set points but may miss dynamics or proportional-integral-derivative (PID) controller anomalies. An interesting example of the kinds of dynamic faults that are present and detected by MCC and models is shown in Figure 7. Unnecessary cooling, caused because of large airflow set point, caused the temperature to fall below heating set point though external temperature was high and rising. This triggered heating (as seen by the heating command). Heating is a large waste in our building since it is local. We located 5 other rooms in just one floor of our building, and more than 15 overall, with this problem. The cause for this fault is that the designed air-flow specifications are based on the maximum occupancy of a zone. When these zones are not fully occupied, the minimum airflow is too high. The heating and cooling (in both summer and winter) cause this to be a large energy waste. However, this is by design and can violate safety and comfort requirements at maximum occupancy. This indicates the need for occupancy count based control. We are in the process of deploying occupancy sensors in zones where this type of fault was found.

5.3 Testing coverage

Testing coverage requires that we locate all rooms with anomalies. This is difficult in the case of all 237 zones, so we focus on only the fourth floor, where we manually inspected all zones. This floor has 54 zones in which we located 14 as anomalous. The performance of different algorithms is shown in Table 2, indicating the types of faults detected, and the number of false alarms. When rooms have two or more
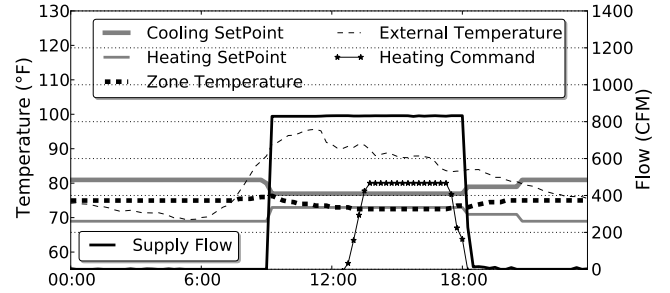

Figure 7: Example of a dynamic anomaly detected by MCC. Unnecessary cooling, caused the temperature to fall below heating set point though external temperature was increasing. This triggered heating which is a much larger waste.

faults we assign them to one of the categories based on which anomaly is more fundamental. For our purposes we do not consider the loss of data for all rooms and Demand Response events as anomalies, since detecting them is of limited practical value - this is of course a point for discussion.

We make some observations. The correlation based method raises a large number of false alarms because, despite removing and separating the data, they are sensitive to external conditions - in this case particularly to changes in external temperature. One possibility is that the threshold was set too liberally. However, the detected anomalies were in position 2 and 10 when sorted by FoM, so a lower threshold would also have decreased coverage. This implies that the problem is more fundamental than setting the appropriate thresholds. MCC on its own cannot detect SP violations

57

Method	Data	Ineffective	Excess Flow	Heating-Cooling	Set-Point	Cycling	False Alarms
Manual	Fourth floor, 30 days	7	1	3	2	1	0
Rules	Fourth floor, 30 days	4	1	2	2	0	2
Correlation	Fourth floor, 30 days	1	1	0	0	1	8
PCA	Fourth floor, 30 days	1	0	1	2	1	5
MCC	Fourth floor, 30 days	5	1	3	1	0	2
Rules (top 10 verified)	All Floors, 1 year	0	0	0	10	0	0
Correlation (top 10 verified)	All Floors, 1 year	0	1	0	1	0	8
PCA (top 10 verified)	All Floors, 1 year	1	0	0	1	0	8
MCC (top 10 verified)	All Floors, 1 year	0	2	6	1	0	1
iRules	All Floors, 1 year	18	12	14	27	7	1

Table 2: Coverage and Specificity of different algorithms. The top half of the table shows all anomalies found and verified for only the fourth floor, while the bottom half considers the top ten faults from each algorithm for the entire building. We see that Rules gives higher FoMs to set-point anomalies, while MCC gives higher FoM to dynamic anomalies. The only false alarm for iRules was a cycling anomaly due to user initiated changes in set-points, and is likely normal. Thus, specific rules with conservative thresholds have almost no false alarms.

directly as a linear model cannot learn that the 'temperature should not be outside set points'. However, the one SP violation it did detect was also a heating-cooling anomaly, and as a result it was detected. The rule based algorithms miss the heating-cooling anomalies, since the zone temperatures are in a relatively small region around the set point (see Figure 7) - thus they have low FoM. This highlights two important issues. The fact that multiple faults occur per zone indicates that some of these fault classes (or algorithms) do not get to the root cause of the problem in the PID controller. Thus, identifying and fixing these faults still requires manual interpretation. This, together with the fact that even basic rules that are not captured by linear (or simple non linear models) indicates that better ensemble models are required.

5.4   Testing specificity

We run the algorithms on all 237 zones for one year of data. Since evaluating all zones as anomalous or not is infeasible, we looked at the top 10 faults detected by each of the algorithms and classified them as true anomalies or false alarms. The results are in the lower half of Table 2. These results also allow us to study what are the kinds of anomalies each kind of algorithm favors i.e. which faults are given higher FoMs. Most of the top faults detected by MCC are of the Heating Cooling type - which is a large energy waste as well. All top 10 faults that had the largest FoM with the rules were of the 'temperature should not be outside set points' - these were clear faults and thus these rules have no FAs in the top 10 results. MCC finds many instances of the Heating-Cooling Anomaly and Supply Flow Excess that are not detected by the other methods. One interesting and practically important anomaly found only by correlation based methods is shown in Figure 3d. We call this a Short Cycling Anomaly (SCA) where the airflow changes rapidly due to faults in the PID controller. This thrashing is a particularly important fault since it can lead to damage of dampers and other equipment. Finally, we see PCA has a large number of false alarms. These were caused by rooms where occupants frequently changed set-points. This causes variation in temperature and airflow resulting in high FoMs from PCA. This also supports the insight behind models in MCC, that the relationship between sensor readings is as critical as (if

not more than) the absolute values.

5.5   Designing Rules

While data driven algorithms are able to find new faults, (HCA, SCA and SFE), they suffer from false alarms. Rules with a conservative threshold have low false alarm rates for specific faults. To get the best of both worlds, we suggest a two stage system where exemplar anomalies from different algorithms are presented to a user, who suggests rules for them. This allows us to obtain the coverage that is possible by the coverage of an ensemble of different algorithms and the low false alarms possible with very specific rules. This leads to the design of intelligent rules (iRules) to detect specific fault classes based on the output of the data mining algorithms. We continue to use a threshold of median value + thrice the variance for all our experiments. We wrote simple rules to detect heating-cooling, excess flow and short cycling faults for all 237 zones on the entire one year data. We manually inspected the detected faults of these iRules. The results in the last row of Table 2 show that we detect 78 anomalies with one false alarm in 237 zones. This illustrates the value of intelligent rules, showing that specific rules raise few false alarms. The live results of iRules for our test-bed are available and being used by our building managers.

6   Discussion

Conclusions: We motivate data driven FDD for VAVs in large commercial buildings. We study a selection of FDD methods, namely PCA, correlation based methods and rules. We develop a new comparative data mining algorithm, Model Cluster and Compare (MCC), which overcomes specific challenges of these methods. Unlike *model-based* methods it does not require detailed physical models of the building. In contrast to *process history* based methods it does not require large amounts of historical data (in-situ or in a lab) of either faulty and non-faulty operation. Finally, in an improvement over both these classes of methods, MCC does not use any deployment specific knowledge. While there is no perfect anomaly detection algorithm, we show that in addition to model, rule and process history based methods, MCC provides an alternate view that can localize an interesting set of faults in practical deployments. iRules based on these exemplars, find many faults with few false alarms.

Observations: We find that many faults occur in special purpose zones like common areas (kitchens and conference rooms). We suspect that this is because they are rare and procedures used for determining settings are not well developed, along with complex safety regulations. This parallels findings in software and embedded systems [16]. These zones also tend to be large consumers of energy. This suggests an iterative method where faults are detected and then settings are changed on the fly, i.e. continuous building commissioning. Situations where safety requirements cause faults (like the heating-cooling anomaly, Section 5.2) indicate the need for occupancy count based control.

Automatic rule generation: An alternative to the manual iRules is to learn classifiers based on anomaly exemplars. However, when we tried to learn a Support Vector Machine (SVM) based on exemplar anomalies, even for simple situations like temperature not within set-point the end result classified all rooms as not anomalous. This is because, even within an anomalous day, there are many normal samples and learning algorithms are not robust to this kind of noisy data. Thus, we require a higher level of supervision, where anomalous samples are marked and not just anomalous days.

Future Work: Finding the balance between automation and human intervention is of primary importance. A unified framework is required that can (i) support a large class of algorithms on live data, (ii) provide information to building mangers that will allow them to diagnose faults and easily determine a fix and (iii) provide estimates of the energy savings if a fault or class of faults is fixed. None of the choices we make are perfect. We use $k = 4$ clusters, with the intuition that they may correspond to West, not-West, interior and common areas but setting the number of clusters automatically is required. Non-linear, higher (≥ 1) order models are often preferred for buildings [8]. Once the model parameters have been estimated, a number of sensible definitions of anomalous exist [6]. Finally, even though our dataset is large, these and other methods must be evaluated in more detail in different parts of the HVAC system, with different HVAC systems and in different buildings in different climates.

We thank the UC San Diego Facilities team, particularly Anna Levitt, for their immense help. We thank the reviewers and apologize for the fact that a lack of space limited our inclusion of their detailed and insightful suggestions in full.

7 References

[1] D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. In *Proc. of ACM-SIAM Symposium on Discrete Algorithms*, pages 1027–1035, 2007.

[2] ASHARE. Handbook of HVAC systems and equipment. Technical report, 1996.

[3] B. Balaji, H. Teraoka, R. Gupta, and Y. Agarwal. ZonePAC: Zonal Power Estimation and Control via HVAC Metering and Occupant Feedback. In *Proc. of BuildSys*. ACM, 2013.

[4] J.-P. Bourdouxhe, J. Lebrun, and M. Grodent. *Reference guide for dynamic models of HVAC equipment*. ASHRAE, Inc., 1998.

[5] M. R. Brambley, P. Haves, S. C. McDonald, P. A. Torcellini, D. Hansen, D. Holmberg, and K. Roth. *Advanced sensors and controls for building applications: Market assessment and potential R & D pathways.* Prepared for US DOE by PNNL, 2005.

[6] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

[7] A. Costa, R. Sterling, L. M. Blanes, M. Howley, and M. M. Keane. A SWOT framework to investigate the integration between building management systems and fault detection and diagnosis tools. *Proc. of Applied Energy*, 2013.

[8] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, et al. EnergyPlus: creating a new-generation building energy simulation program. *Energy and buildings*, 33(4):319–331, 2001.

[9] Z. Du, X. Jin, and L. Wu. Fault detection and diagnosis based on improved PCA with JAA method in VAV systems. *Building and Environment*, 42(9):3221–3232, 2007.

[10] R. Fontugne, J. Ortiz, N. Tremblay, P. Borgnat, P. Flandrin, K. Fukuda, D. Culler, and H. Esaki. Strip, bind, and search: a method for identifying abnormal energy consumption in buildings. In *IPSN*, 2013.

[11] D. Frey and V. Smith. Advanced automated hvac fault detection and diagnostics commercialization program. Technical report, Architectural Energy Corporation for the California Energy Commission, 2008. Publication number: CEC-500-2013-054.

[12] H. Han, Z. Cao, B. Gu, and N. Ren. PCA-SVM-based automated fault detection and diagnosis (AFDD) for vapor-compression refrigeration systems. *HVAC&R Research*, 16(3):295–313, 2010.

[13] J. Hyvarinen and S. Karki. Building optimization and fault diagnosis source book. *IEA Annex*, 25(1996.8), 1996.

[14] S. Katipamula and M. R. Brambley. Review article: methods for fault detection, diagnostics, and prognostics for building systems - a review, part I. *HVAC&R Research*, 11(1):3–25, 2005.

[15] J. Liang and R. Du. Model-based fault detection and diagnosis of HVAC systems using support vector machine method. *International Journal of refrigeration*, 30(6):1104–1114, 2007.

[16] R. R. Lutz. Analyzing software requirements errors in safety-critical, embedded systems. In *Proc. of IEEE International Symposium on Requirements Engineering*, pages 126–133. IEEE, 1993.

[17] E. Mills. Building commissioning: A golden opportunity for reducing energy costs and greenhouse-gas emissions. *Lawrence Berkeley National Laboratory*, 2010.

[18] PECI (Portland Energy Conservation Inc.) and Battelle Northwest Division. *Methods for automated and continuous commissioning of building systems*. Arlington, Virginia: Air Conditioning and Refrigeration Technology Institute, 2003.

[19] M. A. Piette, S. K. Kinney, and P. Haves. Analysis of an information monitoring and diagnostic system to improve building operations. *Energy and Buildings*, 33(8):783–791, 2001.

[20] M. A. Piette, B. Nordman, and S. Greenberg. Quantifying energy savings from commissioning: preliminary results from the pacific northwest. In *Proc. of National Conference on Building Commissioning*, 1994.

[21] T. M. Rossi and J. E. Braun. A statistical, rule-based fault detection and diagnostic method for vapor compression air conditioners. *HVAC&R Research*, 3(1):19–37, 1997.

[22] J. Schein, S. T. Bushby, N. S. Castro, and J. M. House. A rule-based fault detection method for air handling units. *Energy and Buildings*, 38(12):1485–1492, 2006.

[23] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang. A novel anomaly detection scheme based on principal component classifier. Technical report, DTIC Document, 2003.

[24] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Parts I,II and III. *Computers & chemical engineering*, 27(3), 2003.

[25] S. Wang and F. Xiao. AHU sensor fault diagnosis using principal component analysis method. *Energy and Buildings*, 36(2):147–160, 2004.