Two Similarity Measure Approaches to Whole Building Fault Diagnosis

Guanjing Lin[1] and David E. Claridge[2], Ph.D. P.E.

[1,2] *Energy Systems Laboratory, Texas A&M University, College Station, TX 77845*

Email: guanjinglin@tees.tamus.edu

Abstract:
This paper introduces two methods based on cosine similarity and Euclidean distance similarity respectively to diagnose possible causes of abnormal whole building energy consumption. The concepts of cosine similarity and Euclidean distance similarity are defined and the methodology for implementing the proposed whole building fault diagnosis approaches is presented. Cosine similarity and Euclidean distance similarity are applied to two field observed fault test cases, and both the cosine similarity and Euclidean distance similarity methods indicated that the most probable fault was the fault observed in the field survey.

Keywords:
Energy consumption; Cosine angle; Euclidean distance; Fault diagnosis; Similarity measure

1. Introduction

In the U.S., faults in HVAC systems can increase HVAC energy consumption by 30% (Brambley et al. 1998). Hence Fault Detection and Diagnosis (FDD) in building HVAC systems is of significant interest. There are two fundamental approaches to FDD in buildings: a component level (bottom-up) approach and a whole building (top-down) approach (Seem 2007). The component level approach focuses on faults in individual systems such as air-handling units, variable-air-volume boxes, meters, chillers, or boilers. The whole-building approach specializes in abnormal behavior in global-level measurements such as the whole building cooling, heating or electrical consumption. Most HVAC FDD studies focus on components, but several whole-building FDD studies are summarized below.

Haberl et al. (1989) introduced a three-parameter, steady–state model to predict the energy consumption. A fault is identified when the absolute residual between the actual and predicted consumption exceeds a specified deviation. Friedman and Piette (2001) illustrated a whole building FDD tool, which uses a daily energy consumption index to show if the actual energy consumption was higher than normal, normal or lower than normal. Seem (2007) described a method for detecting abnormal energy consumption in buildings. The method uses outlier detection to determine if the energy consumption for a particular day is significantly different than previous energy consumption.

Lee and Claridge (2007) examined the use of the ASHRAE Simplified Energy Analysis Procedure (SEAP) for fault detection at the whole-building level. The calibrated SEAP model is used to predict the cooling and heating consumption during a post-commissioning period. The model is established and calibrated based on the building chilled water (CHW) and hot water (HW) consumption in the baseline period chosen from a post-commissioning time period when the building's operation is considered to be optimal. Curtin (2007) developed a prototype of the Automated Building Commissioning Analysis Tool (ABCAT) following the system of Lee and Claridge (2007). The “Cumulative Cost Difference” plot is applied as the primary fault detection metric. A SDVAV w/Economizer Rules for Diagnostic Clarifier was proposed in the thesis for fault diagnosis.

The papers reviewed above show that most whole building FDD research or tools developed have focused on fault detection. They identified abnormal consumption and may estimate the cost of any abnormality identified. A general scheme to diagnose fault is seldom mentioned in the studies reviewed. In practice, it is meaningful to find a diagnostic method to indicate the possible cause(s) for the detected abnormal energy consumption. Narrowly classifying a fault into a subset of possible causes would help the operator or technician find the specific cause of a fault and correct it more quickly and efficiently.

Similarity measures are widely used in pattern matching. They quantitatively represent the degree of compliance within vectors. Similarity measures have shown effectiveness in FDD in many industries (Yoon and MacGregor 2001, Li and Dai 2005, Huang et al. 2007, Kabir 2009, Lee et al. 2009). According to McGill et al. (1979), there are more than 60 different similarity measures. Among them, the most popular are cosine similarity and Euclidean distance similarity.

Two approaches for whole building fault diagnosis based on cosine similarity and Euclidean distance similarity respectively are presented in this paper. The level of diagnostics proposed emphasizes limiting the possible causes to several options and ranking the options according to their probabilities. It will not attempt to “find a needle in a haystack”, but instead will attempt to effectively reduce the size of the haystack in which the operator must look. This paper first introduces the concepts of cosine similarity and Euclidean distance similarity, next presents the methodology of the proposed whole building fault diagnosis approaches, and then demonstrates the results of two field test cases.

2. Similarity Measures

2.1. Cosine Similarity

Cosine similarity is a fundamental angle-based measure of similarity between two vectors of n dimensions using the cosine of the angle between them (Candan and Sapino 2010). It measures the similarity between two vectors based only on the direction, ignoring the impact of the distance between them. Given two vectors of attributes, $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$, the cosine similarity, $\cos\theta$, is represented using a dot product and magnitude as

$$\cos\theta = \frac{X \cdot Y}{\|X\|\|Y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \quad (1)$$

The resulting similarity ranges from -1 meaning exactly opposite in direction, to 1 meaning exactly the same, with 0 indicating independence, and intermediate values indicating intermediate similarity or dissimilarity.

2.2. Euclidean Distance Similarity

Euclidean distance similarity is a common distance-based measure of similarity between two vectors of n dimensions using the distance between the vectors (Candan and Sapino 2010).

The distance-based similarity measure considers only the impact of the distance between vectors, regardless of the direction of the vectors. Given two vectors of attributes, $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$, the Euclidean distance d from vector X to Vector Y is

$$d(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2)$$

Shepard (1987) proposed as a universal law that Euclidean distance d and perceived Euclidean distance similarity s are related via an exponential function

$$s(X,Y) = e^{-d(X,Y)} \quad (3)$$

The resulting similarity ranges from 0 to 1 with 1 meaning the two vectors are identical.

3. Methodology

Fig.1 displays the major steps required to diagnose abnormal cooling and heating consumption in buildings using similarity measures. The method is referred to as the cosine similarity method if cosine similarity is adopted and is referred to as the Euclidean distance similarity method if Euclidean distance similarity is implemented.

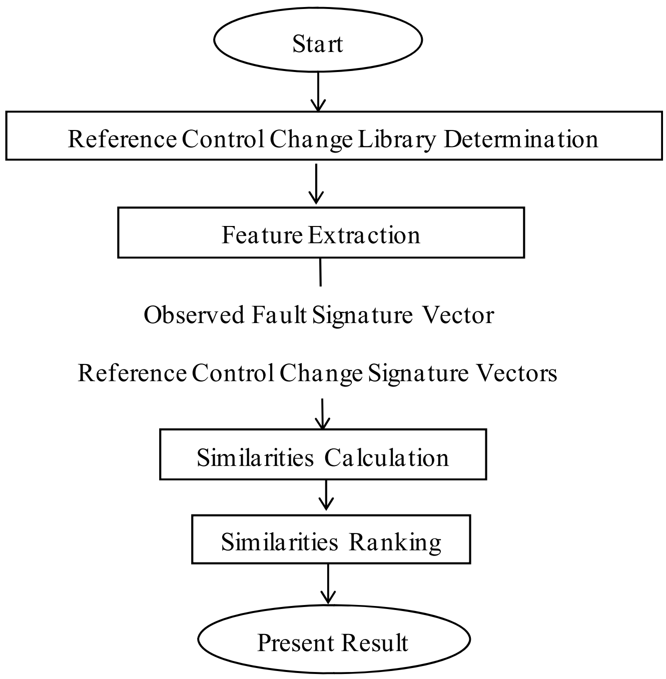

Fig. 1 Block diagram for diagnosing abnormal energy consumption

Step 1: Reference Control Change Library Determination

Whole building fault diagnosis is different from component level fault diagnosis. It can only give a general clue, for example, that there is excess outside air flow in the building, but can't tell which specific component, e.g. the fully closed outside air damper of AHU 3-1, is causing the problem. A technician still needs to investigate in the field to determine and correct the specific cause. The reference control change library collecting known whole building level faults is pre-determined initially. Table 1 gives ten whole building level fault examples. Each fault listed in Table 1 will be called a reference control change in the subsequent discussion.

Table 1: Whole building level fault examples

Outside air flow volume increase/decrease
Preheat/precool temperature increase/decrease
Preheat/precool coil valve leakage
Cooling coil (SD)/cold deck (DD) leaving temperature increase/decrease
Hot deck (DD) leaving temperature increase/decrease
Heating coil valve leakage (DD)
Minimum airflow volume increase/decrease
Maximum airflow volume (CV) increase/decrease
Room set-point temperature increase/decrease
Terminal box damper leakage (DD)

Note: SD – Single Duct System; DD – Dual Duct System; CV – Constant Volume System

The signatures of the reference control changes are used as the reference symptoms in fault diagnosis. The energy pattern may be different when the control change severity is different, so the number of levels of severity for a control change will be defined in advance in the library.

Step 2: Feature Extraction

The feature extraction block generates the observed fault signature vector and a number of reference control change signature vectors, each corresponding to a known control change in the reference control change library.

It is assumed some fault detection mechanism has already determined that an abnormal consumption fault is present and has persisted for a certain time. This period is referred as the fault period. In this block, first, the calibrated simulation model in ABCAT is used to produce the fault-free cooling and heating consumption in the fault period. Second, the calibrated simulation model in ABCAT is used to predict the cooling and heating consumption when there is a known control change from the reference library persisting during the fault period. For a specified control change, a specific input parameter of the calibrated simulation model will be changed. Since there are several levels of severity for a control change, the corresponding input parameter will be changed several times to simulate various fault sizes. Finally, the observed fault signature vector and reference control change signature vectors are generated using the following expression:

$$V = \begin{bmatrix} fs_{CHW} \\ fs_{HW} \end{bmatrix} \quad (4)$$

where $fs_{CHW,i} = \frac{CHW_{mea,i} - CHW_{sim,i}}{E_{AveBaseline}}$, $fs_{HW,i} = \frac{HW_{mea,i} - HW_{sim,i}}{E_{AveBaseline}}$ (Observed fault)

$fs_{CHW,i,j} = \frac{CHW_{ref\,C,i,j} - CHW_{sim,i}}{E_{AveBaseline}}$, $fs_{HW,i,j} = \frac{HW_{ref\,C,i,j} - HW_{sim,i}}{E_{AveBaseline}}$ (Reference control change)

A signature vector includes two parts: the CHW signature f_{sCHW} and the HW signature fs_{HW}. In this way, the similarity of both CHW and HW features can be considered. $CHW_{mea,i}$ and $HW_{mea,i}$ are the daily measured cooling and heating energy consumption values respectively on the i^{th} day of the fault period; $CHW_{sim,i}$ and $HW_{sim,i}$ are the daily fault-free cooling and heating energy consumption values respectively predicted by the calibrated simulation model on the i^{th} day of the fault period. $CHW_{ref\ C,i,j}$ and $HW_{ref\ C,i,j}$ are the daily cooling and heating energy consumption values respectively on the i^{th} day of the fault period when there is the j^{th} control change from the reference library persisting during the fault period. $E_{AveBaseline}$ is the average cooling plus heating energy consumption values in the baseline period. Assuming the fault severity has five levels in the reference library, there would be five reference control change signature vectors for a single reference control change.

Step 3: Similarities Calculation

In this block, cosine similarity and Euclidean distance similarity between the observed fault signature vector and each of the reference control change signature vectors are calculated. X in expressions (1-3) is the observed fault signature vector and Y is the reference control change signature vector. Substituting the expressions of observed and reference signatures, the expressions for cosine similarity cosθ and Euclidean distance similarity S (X, Y) become

$$\cos\theta = \frac{\sum_{i=1}^{n}[(CHW_{mea}-CHW_{sim})_i(CHW_{ref\ C}-CHW_{sim})_i+(HW_{mea}-CHW_{sim})_i(HW_{ref\ C}-HW_{sim})_i]}{\sqrt{[\sum_{i=1}^{n}[(CHW_{mea}-CHW_{sim})_i^2+(HW_{mea}-HW_{sim})_i^2]}\sqrt{[\sum_{i=1}^{n}[(CHW_{ref\ C}-CHW_{sim})_i^2+(HW_{ref\ C}-HW_{sim})_i^2]}} \quad (5)$$

$$S(X,Y) = e^{-\sqrt{\sum_{i=1}^{n}\frac{(CHW_{mea}-CHW_{ref\ C})_i^2}{E_{AveBaseline}}+\sum_{i=1}^{n}\frac{(HW_{mea}-HW_{ref\ C})_i^2}{E_{AveBaseline}}}} \quad (6)$$

where i is the i^{th} day in the fault period and n is the number of days in the fault period. If the reference control change doesn't cause any energy shift over the fault period, $CHW_{ref\ C}$ and $HW_{ref\ C}$ would be the same as CHW_{sim} and HW_{sim} respectively. In this context, the cosine similarity is defined as zero.

Step 4: Similarities Ranking

Similarities ranking block would sort different types of reference control changes by the similarity in descending order. As mentioned above, a reference control change would have more than one fault signature vector. Thus, two steps will be taken when ranking the similarities. First, choose the largest cosine similarity/Euclidean distance similarity from the cases with the same reference control change to be representative of that control change. Next, compare the representative similarities of all the reference control changes and sort them by descending order to create a rank-ordered list of control changes.

Step 5: Present Results

The rank-ordered list of control changes basically ranks the probability that the reference control change is the cause of the observed fault. The similarity measures compare the symptoms of the current fault against the symptoms of the reference control changes. A larger similarity corresponds to a higher probability that the known control change is the cause of the observed abnormal consumption.

4. Field Test Case 1 – Sbisa Dining Hall

4.1. Building Information and Field Data Sets

Sbisa Dining Hall is an 82,000 ft^2 single story building with a partial basement on the campus of Texas A&M University in College Station, TX. Its primary function is as a dining facility. The main AHUs are single duct constant volume (SDCV) AHUs with terminal reheat boxes. Three constant volume dedicated outside air handling units (OAHUs) provide pretreated makeup air for the majority of the AHUs. The ABCAT simulation was calibrated to the baseline consumption period of February 2, 2004 to December 31, 2004.

The field investigation discovered exceptionally low discharge air temperature in two of the three OAHUs in the building in 2006 (Curtin 2007). The investigated fault period is January 1-June 4, 2006. The maximum monthly average cooling and heating consumption increase in the fault period is 15% of the average energy consumption in the baseline period (76.8MMBtu/day) (Fig.2).

Monthly average energy (cooling and heating) use change index in Fig.2 is defined as

$$\text{Monthly average energy use change index} = \frac{\text{Monthly average energy use change}}{\text{Daily average energy use in the baseline period}} \quad (7)$$

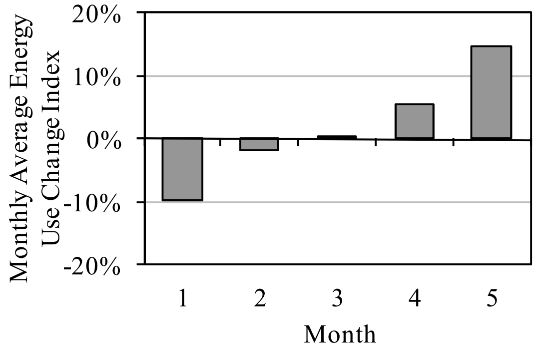

Fig.2 Monthly average energy (cooling plus heating) consumption changes in the period of 1/1/2006-6/4/2006 for the Sbisa Dining Hall

Reference Control Change Library**Error! Reference source not found.**Table 2 defines 12 different types of reference control change and there are five levels (I - VI) of fault severity for each control change. Each row shows a type of reference control change. Column one indicates the ID of the reference control change. Column two provides the key words describing the control change. The remaining columns present the different magnitudes of the control change. For example, "-10%" in the first row means outside airflow ratio decreased 10% and "10%" in the second row means outside airflow ratio increased 10%. The magnitudes III, IV, and V of control change "X_{oa} decrease" are blank; they would have negative values since the original input parameter was 28% in the calibrated simulation model.

Table 2: Reference control change library for the Sbisa Dining Hall

ID	Reference Control Change	Magnitude					Units
		I	II	III	IV	V	
1	X_{oa} decrease	-10%	-20%				
2	X_{oa} increase	10%	20%	30%	40%	50%	
3	T_{prec} decrease	-2	-4	-6	-8	-10	°F
4	T_{prec} increase	2	4	6	8	10	°F
5	T_{cl} decrease	-2	-4	-6	-8	-10	°F
6	T_{cl} increase	2	4	6	8	10	°F
7	X_{max} decrease	-10%	-20%	-30%	-40%	-50%	
8	X_{max} increase	10%	20%	30%	40%	50%	
9	T_{rc} decrease	-2	-4	-6	-8	-10	°F
10	T_{rc} increase	2	4	6	8	10	°F
11	T_{rh} decrease	-2	-4	-6	-8	-10	°F
12	T_{rh} increase	2	4	6	8	10	°F

Note: X_{oa} – Outside airflow ratio; T_{prec} – Outside air precooling temperature; T_{cl} – Cooling coil air leaving temperature; X_{max} – Maximum airflow ratio; T_{rc} – Room cooling set-point temperature; T_{rh} – Room heating set-point temperature.

4.2. Diagnostic Results with Field Test Data

The proposed cosine similarity and Euclidean distance similarity methods were applied in the fault period. The observed fault signature vector components are plotted versus outside air temperature in Fig.3.

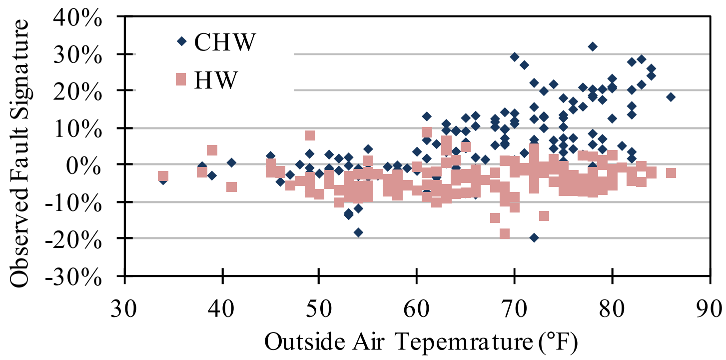

Fig.3 The observed fault signature vector components plotted as a function of outside air temperature in the period of 1/1/2006-6/4/2006 for the Sbisa Dining Hall

Fig.4 shows control changes ordered in descending order of the representative cosine similarity of the type of reference control change. In Fig.4, the first bar on the left has the highest cosine similarity, and the last bar on the right has the lowest cosine similarity. The reference control change IDs on the X axis correspond to the IDs in Table 2. Fig.4 shows that the control changes "T_{prec} decrease" and "X_{oa} decrease" have the largest and smallest cosine similarity respectively among the 12 kinds of reference control change. Therefore, the energy pattern of the control change "T_{prec} decrease" is most similar to the energy pattern of the observed fault. The observed abnormal consumption is most likely due to a decrease of the outside air precool temperature and is least likely to be caused by a decrease of outside airflow ratio. Similarly, the ranking of reference control changes based on the results of Euclidean distance similarity concludes that the decrease of outside air precool temperature is the most probable reason for the observed fault (Fig.5).

Curtin (2007) reported that investigation into trended control data points had led to the discovery of exceptionally low discharge air temperature in two of the three OAHUs in the building. It is obvious that the diagnosis result with either the cosine similarity or Euclidean distance similarity method is consistent with the field investigation conclusion.

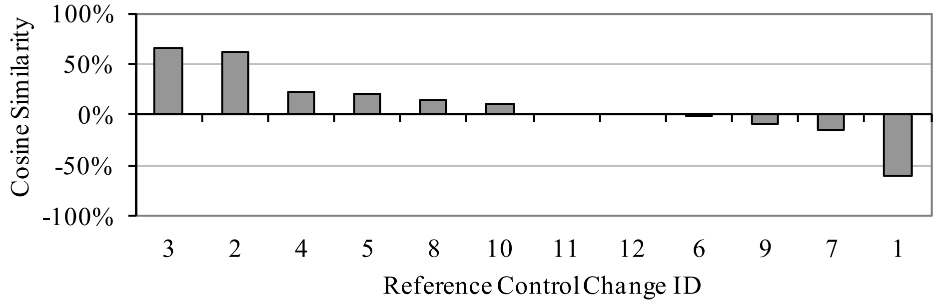

Fig.4 Representative cosine similarity for different reference control changes sorted in descending order for the Sbisa Dining Hall

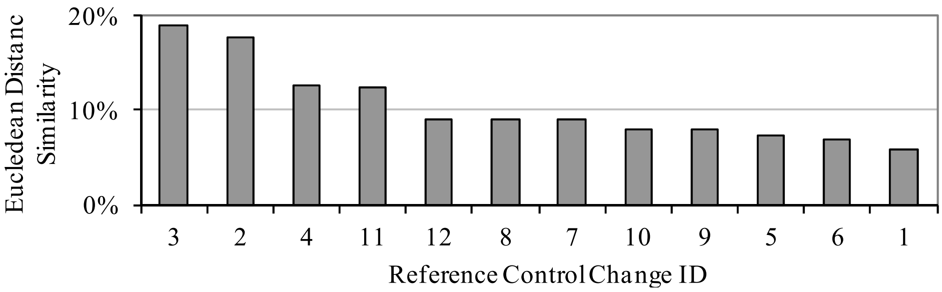

Fig.5 Representative Euclidean distance similarity for different reference control changes sorted in descending order for the Sbisa Dining Hall

5. Field Test Case 2 – Bush Academic Building

5.1. Building Information and Field Data Sets

Bush Academic Building is located on the west campus of Texas A&M University in College Station, TX. It consists primarily of offices and classrooms. The building has three floors for a total area of 133,326 ft^2. It is generally occupied on weekdays during the day. The HVAC system in the building is a DDVAV system. The ABCAT simulation was calibrated to the baseline consumption period of weekdays from June 01, 2007 to April 20, 2008.

The targeted fault period is weekdays from November 1, 2008 to June 30, 2009. The field inquiry indicates that there was a preheat valve leaking by on a pre-treat unit during the fault period (Claridge et al. 2009). The maximum monthly average cooling and heating consumption increase in the fault period is 30% of the average energy consumption in the baseline period (21.7MMBtu/day) (Fig.6).

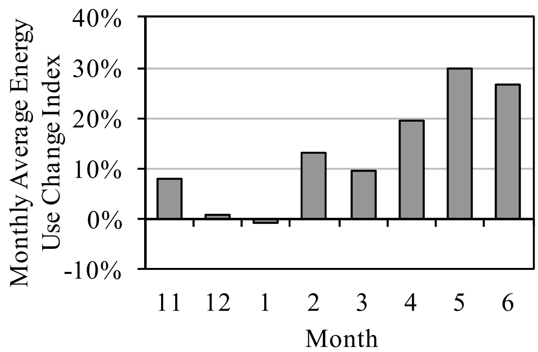

Fig.6 Monthly average energy (cooling plus heating) consumption changes for the weekday period of 11/1/2008-6/30/2009 for the Bush Academic Building

5.2. Reference Control Change Library

Seventeen different types of reference control change with five levels of magnitude are presented in Table 3.

Table 3: Reference control change library for the Bush Academic Building

ID	Reference Control Change	Magnitude					Units
		I	II	III	IV	V	
1	X_{oa} decrease	-2%	-4%	-6%	-8%	-10%	
2	X_{oa} increase	2%	4%	6%	8%	10%	
3	T_{preh} decrease	-3	-6	-9	-12	-15	°F
4	T_{preh} increase	3	6	9	12	15	°F
5	PreHL increase	10	20	30	40	50	kBtu/hr
6	T_{cl} decrease	-2	-4	-6	-8	-10	°F
7	T_{cl} increase	2	4	6	8	10	°F
8	T_{hl} decrease	-2	-4	-6	-8	-10	°F
9	T_{hl} increase	2	4	6	8	10	°F
10	HL increase	10	20	30	40	50	kBtu/hr
11	X_{min} decrease	-2%	-4%	-6%	-8%	-10%	
12	X_{min} increase	2%	4%	6%	8%	10%	
13	T_{rc} decrease	-2	-4	-6	-8	-10	°F
14	T_{rc} increase	2	4	6	8	10	°F
15	T_{rh} decrease	-2	-4	-6	-8	-10	°F
16	T_{rh} increase	2	4	6	8	10	°F
17	TDL increase	2%	4%	6%	8%	10%	

Note: X_{oa} – Outside airflow ratio; T_{preh} – Outside air preheating temperature; PreHL – Heat leakage of preheat coil; T_{cl} – Cold deck air leaving temperature; T_{hl} – Hot deck air leaving temperature; HL – Heat leakage of heating coil; X_{min} – Minimum airflow ratio; T_{rc} – Room cooling set-point temperature; T_{rh} – Room heating set-point temperature; TDL – Terminal box damper leakage.

5.3. Diagnostic Results with Field Test Data

The proposed cosine similarity and Euclidean distance similarity methods are implemented in the fault period. The observed fault signature vector components are plotted versus outside air temperature in Fig.7.

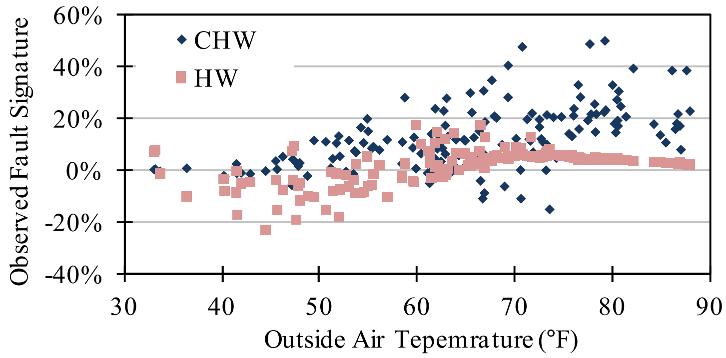

Fig. 7 The observed fault signature vector components plotted as a function of outside air temperature in the weekday period of 11/1/2008-6/30/2009 for the Bush Academic Building

Fig.8 indicates that the cosine similarity values for the observed fault and reference control changes “Heat leakage of preheat coil increase” (ID 5) and “Heat leakage of heating coil increase” (ID 10) are almost identical and rank in the top two places among the 17 reference control changes. This suggests that control changes “Heat leakage of preheat coil increase” and “Heat leakage of heating coil increase” are the two most similar energy change patterns to the energy change pattern of the observed fault. They are the two most probable causes of the observed abnormal energy consumption.

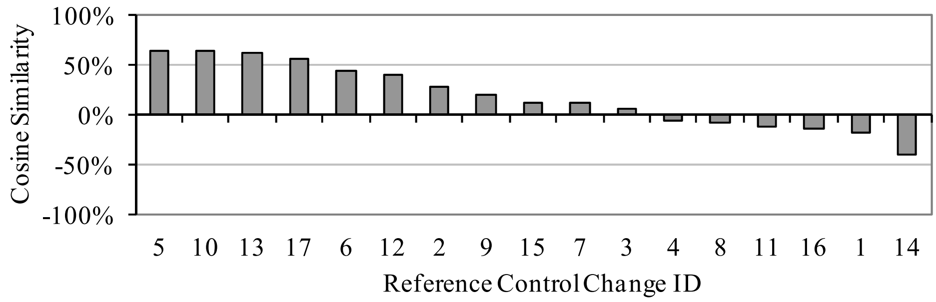

Fig.8 Representative cosine similarity for different reference control changes sorted in descending order for the Bush Academic Building

The difference between the Euclidean distance similarity values of the different reference control changes range from 3% to 7% (Fig.9). The control change “Heat leakage of preheat coil increase” (ID 5) has the largest Euclidean distance similarity value. The small value of Euclidean distance similarity is rooted in its definition. Fig.10 demonstrates that the Euclidean distance similarity exponentially falls with the increase of Euclidean distance. When Euclidean distance is 0.1, Euclidean distance similarity is 90%, and when Euclidean distance is 3, the similarity drops to only 5%. The Euclidean distance among the observed fault vector and all pre-determined reference control change vectors are above 2; thus the corresponding Euclidean distance similarities based on expression (3) are all below 10%.

Both the cosine similarity and Euclidean distance similarity methods indicate the control change “Heat leakage of preheat coil increase” has the highest similarity and thus is considered to be the most probable reason for the observed abnormal energy consumption. The field inquiry indicates that there was a preheat valve leaking by on a pre-treat unit during the fault period (Claridge et al. 2009). The fault diagnosis results are consistent with the field inspection conclusion.

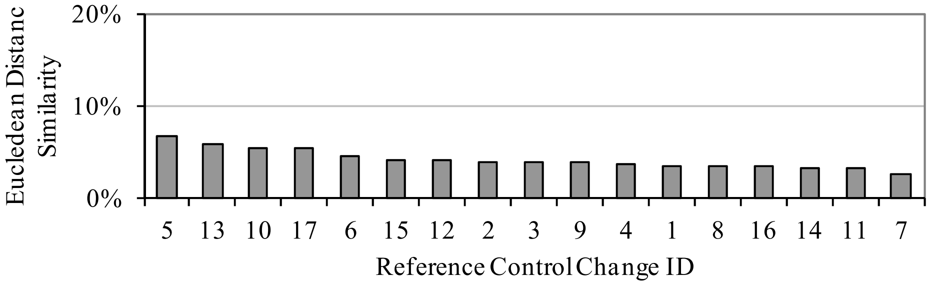

Fig.9 Representative Euclidean distance similarity for different reference control changes sorted in descending order for the Bush Academic Building

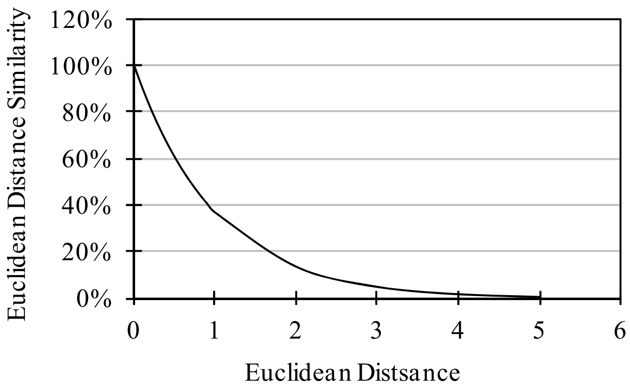

Fig.10 Euclidean distance similarity versus Euclidean distance

6. CONCLUSIONS

Two approaches called the cosine similarity method and the Euclidean distance similarity method proposed to diagnose abnormal whole building cooling or heating energy consumption faults are described in this paper. In these two approaches, a reference control change library collection of known whole building faults is determined in advance. The cosine similarity/Euclidean distance similarity within the observed fault signature vectors and reference control change signature vectors are calculated. Larger similarity values suggest a higher probability that the corresponding reference control change is the cause of the observed fault.

The proposed approaches were used to investigate the reasons for two abnormal energy consumption faults in two real buildings. In the two field test cases, the fault diagnosis results for both the cosine similarity and the Euclidean distance similarity method match the field survey results. This suggests that the cosine similarity method and the Euclidean distance similarity method are promising techniques for whole building fault diagnosis.

References:

Brambley, Michael R., R. G. Pratt, D. P. Chassin, and S. Katipamula. (1998), *Diagnostics for outdoor air ventilation and economizers*. ASHRAE Journal, **40**(10) 49-55.

Candan, K. S. and M. L. Sapino. (2010), *Data management for multimedia retrieval*, Cambridge University Press.

Claridge, D. E., B. John, and G. Lin. (2009), *Final report on retrospective testing and application of an automated building commissioning analysis tool*, Texas A&M University, College Station, TX.

Curtin. J. M. (2007), *Development and testing of an Automated Building Commissioning Analysis Tool (ABCAT)*, MS Thesis, Texas A&M University, College Station, TX.

Friedman, H. and M. A. Piette. (2001), *Comparison of emerging diagnostic tools for large commercial HVAC systems*. Proceedings of the 9th National Conference on Building Commissioning, Cherry Hill, NJ.

Haberl J. S., L.K. Norford, and J. V. Spadaro. (1989), *Diagnosing building operating problems*, ASHRAE Journal, 20-30, June.

Huang H.,C. Li, and J. Jeng. (2007), *Multiple multiplicative fault diagnosis for dynamic processes via parameter similarity measures*, Ind. Eng. Chem. Res. **46** 4517-4530.

Kabir. M. (2009), *Similarity matching techniques for fault diagnosis in automotive infotainment electronics*. International Journal of Computer Science, **3** 14-19.

Lee. S. U., F. L. Painter, D. E. Claridge. (2007), *Whole-building commercial HVAC system simulation for use in energy consumption fault detection*, ASHRAE Transactions, **113**(2) 52-61.

Lee W., H.W. Shen, and G. Zhang. (2009), *Research on fault diagnosis of turbine based on similarity measures between interval-valued intuitionistic fuzzy sets*, International Conference on Measuring Technology and Mechatronics Automation, China.

Li, J. and W. Dai. (2005), *Research on fault diagnosis for oil-immersed transformer based on included angle cosine*, China Academic Journal Electronic Publishing House, **26**(12) 1302-1304.

McGill. M., M. Koll, and T. Noreault. (1979), *An evaluation of factors affecting document ranking by information retrieval systems*, Syracuse, NY: School of Information Studies, Syracuse University.

Seem, J.E. 2007. *Using intelligent data analysis to detect abnormal energy consumption in buildings,* Energy and Buildings, **39** 52-58.

Yoon, S. and J. F. MacGregor. (2001), *Fault diagnosis with multivariate statistical models part I: using steady state fault signatures*, Journal of Process Control, **11** 387-400.