By combining all measures we can thus identify dependencies between the features and locate errors which are caused by a coaction of several processes. The SVM classifier, which we use in a cross validation scheme to validate the ranking, guarantees reliable results. We avoid the phenomenon of over-fitting, often observed in setups with a small ratio of observations and features, and therefore our method requires only a small number of observations.

The results on our benchmark data sets show that the proposed scheme is a powerful tool to complete the existing process control in semiconductor manufacturing. We believe that the results also hold for other serial-group assembly lines, where the lot history is recorded in a database. The proposed ranking and validation methods are relatively easy to implement and give a fast overview on the impact of different machines or stages.

REFERENCES

[1] G. Taguchi, R. Jugulum, and S. Taguchi, *Computer-Based Robust Engineering*. Milwaukee, WI: ASQ Quality Press, 2004.

[2] J. Weston, F. Pérez-Cruz, O. Bousquet, O. Chapelle, A. Elisseeff, and B. Schölkopf, "Feature selection and transduction for prediction of molecular bioactivity for drug design," *Bioinformatics*, vol. 19, no. 6, pp. 764–771, 2003.

[3] R. Isermann, "Model-based fault-detection and diagnosis--status and applications," *Annu. Rev. Contr.*, vol. 29, no. 1, pp. 71–85, 2005.

[4] ——, "Process fault detection based on modeling and estimation methods—a survey," *Automatica*, vol. 20, no. 4, pp. 387–404, 1984.

[5] J. M. Agosta and T. Gardos, "Bayes network "smart" diagnostics," *Intel. Technol. J.*, vol. 8, no. 4, pp. 361–372, 2004.

[6] D. Braha and A. Shmilovici, "Data mining for improving a cleaning process in the semiconductor industry," *IEEE Trans. Semicond. Manuf.*, vol. 15, no. 1, pp. 91–101, 2002.

[7] M. Bensch, M. Schröder, M. Bogdan, and W. Rosenstiel, "Feature selection for high-dimensional industrial data," in *Proc. ESANN*, 2005.

[8] R. Webbink and S. Hu, "Automated generation of assembly system-design solutions," *IEEE Tran. Autom. Sci. Eng.*, vol. 2, no. 1, pp. 32–39, Jan. 2005.

[9] Q. Huang and J. Shi, "Stream of variation modeling and analysis of serial-parallel multistage manufacturing systems," *J. Manuf. Sci. Eng.*, vol. 126, no. 3, pp. 611–618, 2004.

[10] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Proc. Int. Conf. Mach. Learn.*, 1997, pp. 412–420.

[11] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, Eds., *Feature Extraction, Foundations and Applications*. New York: Springer, 2006.

[12] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artif. Intell.*, vol. 97, no. 1–2, pp. 245–271, 1997.

[13] R. Kohavi and G. John, *Feature Selection for Knowledge Discovery and Data Mining*. Norwell, MA: Kluwer, 1998, pp. 33–50.

[14] T. Lal, O. Chapelle, J. Weston, and A. Elisseeff, "Embedded methods," in *Feature Extraction, Foundations and Applications*, I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, Eds. New York: Springer, 2006.

[15] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn.*, vol. 5, pp. 1531–1555, Nov. 2004.

[16] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines 2001 [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[17] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge University Press, 2000.

[18] B. Schölkopf and J. Smola, *Lerning With Kernels*. Cambridge, MA: MIT Press, 2002.

[19] G. D. Nicolao, E. Pasquinetti, G. Miraglia, and F. Piccinini, "Unsupervised spatial pattern classification of electrical failures in semiconductor manufacturing," presented at the IAPR—TC3, Florence, Italy, Sep. 2003.

Data-Driven Modeling, Fault Diagnosis and Optimal Sensor Selection for HVAC Chillers

Setu Madhavi Namburu, Mohammad S. Azam, Jianhui Luo, Kihoon Choi, and Krishna R. Pattipati

***Abstract*—Chillers constitute a significant portion of energy consumption equipment in heating, ventilating and air-conditioning (HVAC) systems. The growing complexity of building systems has become a major challenge for field technicians to troubleshoot the problems manually; this calls for automated "smart-service systems" for performing fault detection and diagnosis (FDD). The focus of this paper is to develop a generic FDD scheme for centrifugal chillers and also to develop a nominal data-driven ("black-box") model of the chiller that can predict the system response under new loading conditions. In this vein, support vector machines, principal component analysis, and partial least squares are the candidate fault classification techniques in our approach. We present a genetic algorithm-based approach to select a sensor suite for maximum diagnosabilty and also evaluated the performance of selected classification procedures with the optimized sensor suite. The responses of these selected sensors are predicted under new loading conditions using the nominal model developed via the black-box modeling approach. We used the benchmark data on a 90-t real centrifugal chiller test equipment, provided by the American Society of Heating, Refrigerating and Air-Conditioning Engineers, to demonstrate and validate our proposed diagnostic procedure. The database consists of data from sixty four monitored variables of the chiller under 27 different modes of operation during nominal and eight faulty conditions with different severities.**

***Note to Practitioners*—Heating, ventilating and air-conditioning (HVAC) systems constitute largest portion of energy consumption equipment. Even though safety is not a critical issue in HVAC industry, the complexity of modern HVAC systems, the operational and maintenance costs associated with the equipment are calling for sophisticated automatic fault diagnosis tools. Among the HVAC components chillers are primarily known for significant energy consumption. Presently, very little of the existing research on fault diagnosis of HVAC systems is aimed at chillers and suffers from certain drawbacks. The primary goal of this paper is to address these issues and develop a generic fault diagnosis tool applicable to any HVAC component. In this vein, we proposed a data-driven approach based on neural network and statistical tools for fault diagnosis, and nominal model development. Since chillers are an example of a data-rich environment we adopted a genetic algorithm based approach for optimal sensor selection for maximizing the diagnosibility. The approach is also validated on an experimental data from 90-t centrifugal chiller provided by the American Society of Heating, Refrigerating and Air-Conditioning Engineers and also has a potential for practical application.**

***Index Terms*—Chillers, data-driven modeling, fault diagnosis, heating, sensor selection, ventilation and air-conditioning systems.**

I. INTRODUCTION

Heating, ventilating and air-conditioning (HVAC) systems constitute the largest portion of commercial and industrial energy consumption equipment. The research on automated fault detection and diagnosis (FDD) of HVAC systems was initiated in the late 80's, and has become a major area of interest over the past decade due to the increased complexity of building systems. The main goal of automated

Manuscript received December, 6, 2005; revised July 12, 2006. This paper was recommended for publication by Associate Editor S. Keerthi and Editor N. Viswanadham upon evaluation of the reviewers' comments. This work was supported in part by the NASA-ARC under Contract NAG-1635.

The authors are with the Electrical and Computer Engineering Department, University of Connecticut, Storrs, CT 06268 USA (e-mail: krishna@engr.uconn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2006.888053

FDD in HVAC systems is to reduce the operating costs. Degraded equipment, failed sensors, improper installation, poor maintenance, and improperly implemented controls are some of the operational problems that plague many commercial buildings [1]. Currently, the problems associated with building systems are discovered from occupant complaints, or alarms provided by the equipment. Due to the complexity of current HVAC systems, it has become difficult for the maintenance personnel to detect and diagnose faults using manual troubleshooting techniques alone. An automated real-time FDD scheme for HVAC equipment can ensure building occupants' comfort, better equipment efficiency, lower energy bills, and fast and accurate recovery from interrupted operations.

Chiller is an important component of HVAC equipment that consumes a significant portion of energy. Katipamula and Brambley [16], [17] provided an exhaustive literature review on fault diagnostics and prognostics of building systems. Several researchers [11]–[15] investigated various approaches for FDD in HVAC systems. However, there exists a relatively limited literature on the FDD of chillers. Due to the increased complexity of building systems, the importance of FDD in chillers and vapor compression equipment is growing rapidly. The ongoing research on the FDD of chillers suffers from the following limitations: testing at a single operating condition, testing at only one fault level (often corresponding to a catastrophic condition), testing on only a few types of faults, or developing procedures applicable to a particular system [1]. The purpose of the work presented here is the development of a generic real-time FDD strategy with an optimal sensor selection, and the development of a nominal data-driven model for centrifugal chillers.

In this paper, based on our extensive experience with system diagnosis in aerospace, automotive applications and information retrieval [4]–[7], [7], [9], we employed three well-known pattern recognition and statistical inference techniques, viz., support vector machines (SVMs), principal component analysis (PCA), and partial least squares (PLS), for fault classification. PLS-based techniques are employed for estimating the severity of faults. We validated our approach by testing these classification techniques on experimental data from a 90-t centrifugal chiller provided by ASHRAE [2]. We employed a genetic algorithm to solve the optimal sensor selection problem of maximizing diagnosability, while minimizing the number of sensors for reducing equipment costs. We also developed a nominal model of the system to estimate the responses of the optimal sensor suite under new operating conditions via data-driven modeling.

II. Target System Description

The target system considered here is a McQuay PEH048J 90-t centrifugal chiller of an HVAC system. The system consists of a shell-and-tube evaporator, a shell-and-tube condenser, a pilot-driven expansion valve, and a centrifugal compressor. For designing the FDD scheme, we utilized the database provided by ASHRAE as part of their research project on centrifugal chillers fault diagnosis. Comstock and Braun [1] provided a detailed description of the chiller considered here and also of the test facility.

A. Faults in Chillers

The faults in chillers occur due to normal wear and tear, poor installation, and human errors during servicing. Degradation faults in chillers lead to loss in performance, and are difficult to detect and diagnose. It was reported that these faults cumulatively account for 42% of the service calls made, and 26% of the repair costs [1]. In this paper, we considered the data from eight single faults provided by ASHRAE. Table I provides information on these faults and the levels of fault severities considered for FDD.

TABLE I
FAULT LIST FOR THE CHILLER

Fault Notation	Description	Severity Levels
RCWF	Reduced condenser water flow	10%, 20%, 30%, 40% reduction
REWF	Reduced evaporator water flow	10%, 20%, 30%, 40% reduction
RL	Refrigerant leak/undercharge	10%, 20%, 30%, 40% reduction
RO	Refrigerant overcharge	10%, 20%, 30%, 40% increase
EO	Excess oil	14%, 32%, 50%, 68% increase
CF	Condenser fouling	12%, 20%, 30%, 45% increase
NCR	Non-condensable in the refrigerant	1%, 2%, 3%, 5% nitrogen addition
DEV	Defective expansion valve	---

TABLE II
MONITORED VARIABLES

Type of Measurement	Number of Monitored Variables	Units
Temperature	29	°F
Pressure	5	PSIG
Power	1	kW
Current	1	Amps
% of maximum rated load Amps	1	%
Water flow rate	2	GPM
Valve position	7	% Open
Unit Status	2	---

B. Sensors and Monitored Variables

When a fault occurs, typically it cannot be diagnosed by data from a single sensor. To continuously monitor the health of the system, a set of measurements is needed. There are 64 monitored variables available in the current system, of which 16 are derived measurements (from VisSim [1] software application). We performed our FDD experiments on data obtained from a set of 48 variables, after eliminating the 16 derived measurements. Table II presents the number of variables of each type along with their units for the 48 monitored variables.

The experimental dataset consists of nominal and faulty data of chiller under 27 predefined operating conditions. The input variables to the chiller are: evaporator water leaving temperature (TEO), condenser water entering temperature (TCI), and the chiller cooling load (Capacity %).

III. Overview of the FDD Process

Our FDD process consists of an offline training phase and an online implementation phase. Fig. 1 depicts the block diagram of our real-time FDD scheme for centrifugal chillers. As the techniques used in our FDD process are suitable for steady-state data, detecting when the system has reached a steady state is an inevitable part of the FDD process. The following sub-sections briefly describe each of the processing blocks in Fig. 1.

A. Steady-State Detector

The chiller's operating envelope changes frequently depending on the load. Consequently, the system responses require time to reach a steady state, which can be substantial. We implemented a steady-state detector based on two approaches: the exponentially weighted variance method of Glass *et al.* [10], and time derivatives of measurements with a moving window.

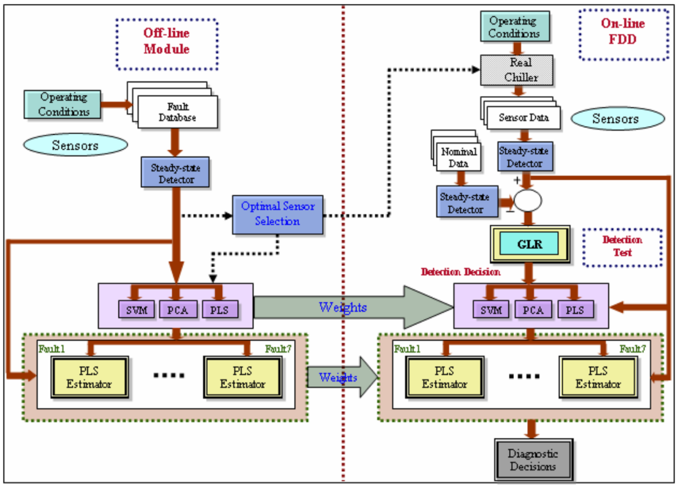

Fig. 1. Block diagram of real-time FDD for centrifugal chillers.

B. *Offline (Training) FDD Module*

In the offline phase, steady-state sensor data from different fault classes is used to train three fault classifiers (SVM, PCA, PLS). We also train a PLS estimator to assess fault severities after fault isolation. The trained classifiers are exported to the online module for real-time FDD. The readers are referred to [6]–[8] for detailed description of these three techniques and their usage. An optimal sensor selection block is used to select the significant sensor suite for maximum diagnosability.

C. *Online (Deployed) FDD Module*

The online FDD phase consists of three steps: fault detection, fault isolation or classification, and fault identification or severity estimation. In the fault detection step, a generalized likelihood ratio test (GLRT) [6] is performed on residuals generated from the steady-state measurements of faulty and nominal systems. Upon detection of a fault, trained classifiers are used for online categorization of the faults. In the next step, a PLS estimator corresponding to the isolated fault is used to determine its severity. We used this phase as a testing phase on the real-chiller data.

IV. Optimal Sensor Selection via a Genetic Algorithm

Sensor measurements play a vital role in determining system performance. Inaccurate measurements due to improper sensor placement or insufficient measurements can significantly deteriorate system performance. For complex systems, number of possible sensors, and number of failure modes are usually large. Existing chillers in building systems are equipped with a comprehensive suite of sensors. It is infeasible to perform an exhaustive search to evaluate the observable discrepancies in all the sensors resulting from all failure modes. Hence, optimal sensor selection is an important issue in complex system monitoring problems. Determining an optimal sensor suite will not only reduce the computational complexity associated with fault diagnosis, but also results in reduced equipment cost, redundancy, weight, volume, and increased reliability/availability.

Choosing the best combination of sensors from a pool of sensors in order to maximize the diagnostic accuracy is an NP-hard optimization problem. Genetic algorithms (GAs) have proven to be powerful search strategies for evolving near-optimal solutions from a complex solution space. Here, we apply a genetic algorithm-based approach to optimal sensor selection in the chiller systems.

A. *Problem Formulation*

We consider two problem formulations: (1) determination of the optimal sensor suite without considering sensor costs, and (2) optimal sensor suite with uniform sensor costs. The sensor optimization problem for maximizing the accuracy is outlined below.

Case 1: The objective function in this case is as follows:

$$\begin{aligned} &\min \left(-\sum_{i=1}^{N_{oc}} D_{a_i} \right) \\ &s.t. \sum_{i=1}^{N_{oc}} D_{a_i} \geq N_{oc} T \\ &\quad N_s \leq N_{\max} \end{aligned} \tag{1}$$

where

D_{a_i}	diagnostic accuracy in each operating condition;
T	diagnostic accuracy threshold;
N_{oc}	number of operating conditions;
N_i	number of sensors;
$N_{\max}$	constraint on number of sensors to be used.

Case 2: When the change in accuracy with a reduction in the number of sensors is insignificant, there may be substantial savings in equipment cost and/or weight. The objective function in this case is defined as follows:

$$\begin{aligned} &\min \left(-\sum_{i=1}^{N_{oc}} D_{a_i} + \sum_{s_j} C_{s_j} I_j \right) \\ &s.t. \sum_{i=1}^{N_{oc}} D_{a_i} \geq N_{oc} T \\ &\quad N_s \leq N_{\max} \end{aligned} \tag{2}$$

where

C_{S_j}	cost of sensor s_j;
I_j	$\text{Indicator Variable} = \begin{cases} 1, & \text{if sensor } S_j \text{ is used} \\ 0, & \text{otherwise.} \end{cases}$

For simplicity, we assumed that the sensors have uniform costs in our numerical experiments.

V. Nominal Model Development

For fault detection, we need to compute the residuals of measurement variables. Computation of residuals requires the nominal data. However, a physics-based generic chiller model that is capable of simulating the widely varying operating conditions is unavailable (yet to be established) [1], [3]. Hence, to generate the nominal data with a reasonable accuracy, we developed a black-box model of the HVAC chiller. Black-box models have proven to be promising in capturing the behavioral characteristics of HVAC systems [14]. In addition to fault detection, this model can be used as the nominal system simulator for the FDD approaches, where fault isolation necessitates the computation of residuals.

Fig. 2 shows the black-box model for estimating the measurements of the chiller under new operating conditions. The inputs to the model are as follows: The first three inputs $(u1, u2, u3)$ are the measurements of the three input variables (the chilled water leaving temperature, the condenser water entering temperature and the evaporator heat transfer rate, respectively) to the system under each operating condition. The other three variables $(u4, u5, u6)$ are obtained by transforming the first three variables $(u1, u2, u3)$ in the following way: $u4 : (u3)^2$; $u5 : u1 \times u3$; $u6 : u2 \times u3$. These three derived variables are added

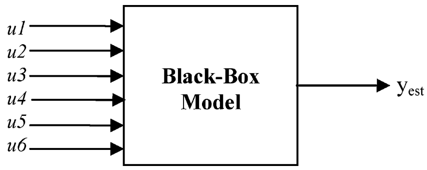

Fig. 2. Black-box model for estimating the response variables.

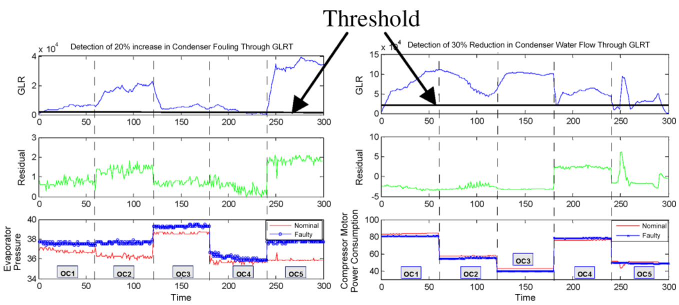

Fig. 3. Detection of RCWF and CF faults through GLRT on evaporator pressure (PSIG) and compressor power (Kw) residuals.

to improve the model's accuracy. The transformations are determined based on the estimation errors.

While performing the regression, the X matrix corresponds to measurements of the six inputs to the model, and Y corresponds to the sensor measurements (target variables). PLS regression models are developed, one for each sensor measurement. We performed a leave-one-operating condition-out cross validation to test the viability of the estimated nominal model. The expected nominal measurements for the optimal sensor suite are estimated using the black-box model.

VI. Results and Discussion

We performed several FDD experiments on the target chiller data. The sampling interval for sensor measurements was 10 s. Each sample measurement is considered as one pattern. Steady-state faulty data from all 48 sensors provided 1740 patterns for each operating condition. Fault detection and diagnosis is performed for each operating condition separately. A simple GA (SGA) [18] is implemented to determine the optimal sensor suite. The values of these measurements are predicted for new operating conditions via the black-box models. The results are presented in the following subsections.

A. Fault Detection

Faults are detected via GLRT on the residuals of measurements. Fig. 3. shows the detection of condenser fouling and reduced condenser water flow faults with 20% and 30% severity levels, respectively, by way of GLRT on evaporator pressure and compressor motor power sensor measurement residuals, respectively. The thick lines on GLR plot indicate the predefined thresholds to detect faults. Here, the plots shown are from the first five operating conditions. The GLRT detected faults in each operating condition with 100% accuracy.

B. Fault Diagnosis

Fault diagnosis involves the classification of a fault (isolation) and its severity estimation (identification). The following sections elaborate on these two aspects of fault diagnosis.

1) Fault Isolation: After a fault is detected, the sensor measurements are presented to the fault classifiers. We used alternative patterns for training and testing from the 1740 patterns in each operating condition. Table III provides a partial view of classification results of all

TABLE III
Classification Rates via SVM, PCA, and PLS in Each Operating Conditions

Operating Condition	1	2	3	...	25	26	27	Average
SVM	100	99.4	100	...	99.9	99.5	99.8	99.81
PCA	99.7	99.7	100	...	99.2	99.2	99.7	99.54
PLS	99.8	99.8	100	...	97.1	96.6	97.1	98.77

TABLE IV
Fault Severity Estimation Results (Average % Errors)

Faults/ Operating Condition	RCWF	REWF	RL	RO	EO	CF	NCR
OC 1	0.98	1.3	3.33	0.09	4.31	9.75	2.45
	0.56	0.68	4.21	0.21	1.45	4.51	3.31
	0.31	0.40	2.66	0.11	1.25	2.62	1.32
	0.51	1.57	1.85	0.13	0.60	1.73	0.40
OC 9	2.18	1.24	10.83	7.46	5.32	6.73	5.02
	0.59	0.95	6.40	3.93	2.58	4.46	3.56
	0.37	0.63	3.93	2.36	0.85	2.74	1.34
	0.53	0.26	3.86	2.28	1.49	8.34	0.66
OC 25	2.57	1.85	9.98	6.79	2.91	10.59	17.58
	1.78	0.33	10.92	4.37	1.61	7.42	4.24
	0.53	0.39	8.25	3.77	2.18	2.20	4.62
	0.43	0.36	11.24	6.09	0.72	2.53	1.68

faults in different operating conditions. It is observed that the classification accuracy of each classifier is always above 95% in any operating condition.

2) Fault Severity Estimation: We estimated the severities of the first seven faults listed in Table I to assess the size of the fault. A PLS estimator is trained for each fault separately. In each operating condition, there were 60 patterns for each fault severity. Of the 240 patterns from four severity levels, we used 200 patterns for training, and tested on the remaining 40 patterns in each operating condition. Table IV presents the severity estimation results for each fault at all four severity levels for some of the operating conditions. The results show that the average estimation errors are within 20% for faults with severity above the first level, in most of the cases. The errors exceeded 20% for faults with the lowest severity levels.

C. Statistical Test for Comparing Classifiers

In the literature, several classification techniques exist for fault diagnosis. In order to customize the best technique for real-time application, it is necessary to compare these techniques in terms of their accuracy. Many statistical methods are available to compare classification techniques. We performed McNemar's test [6] to compare the classification performance of the three techniques in our approach, and found that SVM's performance is better than the PCA and PLS based on the chi-square statistic with one degree of freedom.

D. Optimal Sensor Selection Results

The objective function considered here uses the diagnostic accuracy computed from SVM as the performance criterion. The main reasons for considering only the SVM were that the SVM proved to be the most suitable technique for chiller application in Choi *et al.* [6] and that it also provided the best classification accuracy on the real-chiller data based on the McNemar's test.

Case 1: Here, a run of the GA is performed using all the 48 sensors without considering any sensor costs. The following constraints were imposed on the solution: the total accuracy (T) should be more than

TABLE V
AVERAGE % ERRORS OF THE MEASUREMENT PREDICTIONS IN 27 OPERATING CONDITIONS

Sensors / Operating Condition	TCO	TSO	TBO	T_suc	Unit Status	TO_sump	TWI	THO
1	0.08	1.29	0.18	0.34	6.06	2.06	6.15	0.92
2	0.05	0.60	0.09	0.14	3.24	2.08	1.55	1.78
3	0.22	0.88	0.17	0.18	7.28	3.86	5.75	0.89
4	0.14	0.65	0.06	0.28	2.53	1.97	1.72	0.34
5	0.05	0.69	0.08	0.23	1.10	1.87	2.31	2.14
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20	0.13	0.86	0.16	0.52	0.34	0.59	0.76	0.89
21	0.09	2.47	0.33	0.55	0.67	1.90	0.85	2.32
22	0.22	0.10	0.27	0.43	2.08	1.92	4.57	0.65
23	0.11	0.70	0.17	0.67	3.22	0.36	2.66	0.32
24	0.09	1.10	0.27	0.33	3.33	3.57	4.14	0.82
25	0.16	1.07	0.15	0.38	2.71	2.31	1.93	1.03
26	0.20	0.79	0.23	0.47	3.04	1.09	7.11	0.38
27	0.22	1.63	0.27	0.64	3.50	3.07	7.42	2.35

99% and the number of sensors $(N_{\max})$ should be less than or equal to 48. The GA converged in 17 iterations, showing the number of optimal sensors to be 24 with a maximum attainable accuracy of 99.898%.

Case 2: Here, we assumed equal costs for all 48 sensors. The GA is run with a constraint on total accuracy (T) to be more than 95% and the number of sensors $(N_{\max})$ to be less than or equal to 48. The GA converged in 39 generations resulting in 8 as the optimal number of sensors with a best accuracy of 99.187%. It is evident that the accuracy falls within 7-sigma region of the maximum attainable accuracy. We can see that this is a reasonable amount of accuracy compared to the best attainable accuracy, since the best accuracy is already very high.

E. Model Prediction Results

A nominal model of the chiller is developed via black-box modeling. A PLS regression model is developed for estimating the expected nominal measurement for each sensor. While performing the leave-one-operating condition-out cross-validation, we predicted the measurements for the optimal sensor suite of Case 2 presented above in the remaining operating condition, treating it as a new condition. Table V presents a partial view of the average % errors of the measurement predictions for each sensor in each operating condition. It is observed that the average error for each sensor is below 10% in any operating condition, which makes it suitable for predicting the expected sensor measurements online for use in FDD.

VII. CONCLUSION AND FUTURE WORK

In this paper, a generic real-time fault diagnosis scheme is developed for HVAC chillers. The GLRT is used to detect the faults. Three techniques, viz., SVM, PCA, and PLS, are used to isolate the faults; fault severities are estimated using PLS. A genetic algorithm is implemented for selecting the optimal sensor suite to maximize diagnosability. A black-box model of the chiller is developed using PLS, and was used to estimate the sensor measurements under new operating conditions.

In the future, we plan to perform FDD based on extracted features (e.g., wavelets, kurtosis, etc.). This will further reduce the computational complexity, and will also facilitate FDD using transient data. We also plan to develop a more robust model of the chiller via hybrid modeling (analytic and data driven), as well as methods for combining classifiers to aid in future FDD of HVAC systems.

REFERENCES

[1] M. C. Comstock and J. E. Braun, Development of analysis tools for the evaluation of fault detection and diagnostics for chillers Purdue Univ., West Lafayette, IN, Rep. 4036-3, 1999.

[2] ——, Experimental data from fault detection diagnostic studies on a centrifugal chiller Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN, Rep. HL99-18, 1999.

[3] ——, Literature review for application of fault detection and diagnostic methods to vapor compression equipment Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN, Rep. HL99-19, 1999.

[4] M. Azam, K. Pattipati, J. Allanach, S. Poll, and A. Patterson-Hine, "In-flight fault detection and isolation in aircraft flight control systems," presented at the IEEE Aerospace Conf., Big Sky, MT, Mar. 2005.

[5] J. Luo, K. Pattipati, L. Qiao, and S. Chigusa, "An integrated diagnostic development process for automotive engine control systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, to be published.

[6] K. Choi, S. M. Namburu, M. Azam, J. Luo, K. Pattipati, and A. Patterson-Hine, "Fault diagnosis in HVAC chillers: Adaptability of a data-driven fault detection and isolation approach," *IEEE Instrum. Meas. Mag.*, vol. 8, no. 3, pp. 24–32, Aug. 2005.

[7] S. M. Namburu, H. Tu, J. Luo, and K. R. Pattipati, "Experiments on supervised learning algorithms for text categorization," presented at the IEEE Aerospace Conf., Big Sky, MT, 2005.

[8] S. M. Namburu, J. Luo, M. Azam, and K. R. Pattipati, "Fault detection, diagnosis and data-driven modeling in HVAC chillers," *Proc. SPIE*, Mar. 2005.

[9] Qualtech Systems, Inc. [Online]. Available: http://teamqsi.com.

[10] A. S. Glass, P. Gruber, M. Roos, and J. Todtli, "Qualitative model-based fault detection in air-handling units," *IEEE Control Syst. Mag.*, vol. 15, no. 4, pp. 11–22, Aug. 1995.

[11] M. Stylianou and D. Nikanpour, "Performance monitoring, fault detection, and diagnosis of reciprocating chillers," *ASHRAE Trans.*, vol. 102, no. 1, pp. 615–625, 1996.

[12] M. Stylianou, "Application of classification functions to chiller fault detection and diagnosis," *ASHRAE Trans.*, vol. 103, no. 1, pp. 645–656, 1997.

[13] M. Breuker and J. E. Braun, "Evaluating the performance of a fault detection and diagnostic system for vapor compression equipment," *HVAC R. Res.*, vol. 4, no. 4, pp. 401–425, 1998.

[14] P. Sreedharan and P. Haves, "Comparison of chiller models for use in model-based fault detection," in *Proc. Int. Conf. Enhanced Building Operations*, Austin, TX, 2001.

[15] T. M. Rossi and J. E. Braun, "A statistical, rule-based fault detection and diagnostic method for vapor compression air conditioners," *HVAC R. Res.*, vol. 3, no. 1, pp. 19–37, 1997.

[16] S. Katipamula and M. R. Brambley, "Methods for fault detection, diagnostics, and prognostics for building systems—a review, part I," *HVAC R. Res.*, vol. 11, no. 1, pp. 3–25, 2005.

[17] S. Katipamula and M. R. Brambley, "Methods for fault detection, diagnostics, and prognostics for building systems—a reviews, part II," *HVAC R. Res.*, vol. 11, no. 2, pp. 169–187, 2005.

[18] GA Toolbox for Use With Matlab [Online]. Available: http://www.shef.ac.uk/acse/research/ecrg/gat.html.