Contents lists available at ScienceDirect

Energy and Buildings

Robust model-based fault diagnosis for air handling units

Timothy Mulumba [a], Afshin Afshari [a,*], Ke Yan [a], Wen Shen [a], Leslie K. Norford [b]

[a] Department of Engineering Systems and Management, Masdar Institute of Science and Technology, PO Box 54224, Abu Dhabi, United Arab Emirates
[b] Department of Architecture, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

A R T I C L E   I N F O

A B S T R A C T

Fault detection and diagnosis (FDD) for heating, ventilation and air conditioning (HVAC) equipment significantly impact energy consumption in both residential and commercial buildings. In most modern building management systems (BMS), HVAC historical data logs in large quantity and high resolution are recorded and available for further online or offline analysis, including automated FDD. In this paper, a model-based fault diagnosis method is developed by applying support vector machine (SVM) techniques to model parameters recursively calculated by an online estimator. The estimator presumes an autoregressive time series model with exogenous variables (ARX). A real-world air handling unit (AHU) dataset containing process variables measured at regular intervals is pre-processed by the online estimation algorithm. Each data vector in the original dataset (measured), together with a small number of appropriately selected lags, is converted to a parameter vector representing the state at the same instant. The set of parameter vectors is sub-divided into classes by SVM, enabling fault classification. Validation via experimental data demonstrates that the proposed hybrid approach produce superior performance measured by *F*-measure scores compared to alternative methods. Robustness to model uncertainty is also established.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Poorly maintained HVAC equipment represents up to 30% of total energy consumed in commercial buildings [1]. By improving control and maintenance, FDD techniques significantly impact HVAC energy efficiency. Despite the integration of increasingly sophisticated monitoring and control functions in modern BMS to meet the growing requirement for reliable and economic operation, these systems are usually not able to detect non-catastrophic HVAC malfunction before the scheduled maintenance nor do most malfunctions impact end-user thermal comfort to any noticeable degree. Meanwhile, building management systems monitor various sensors and store the data in large databases, which make it possible to save energy by mining the data and scheduling necessary preventive maintenance. In hot and humid climates, the potential impact of this function is compounded by the fact that the air-conditioning load may even represent up to 60% of the total grid electricity consumption. Detection methods based on energy/cost benchmarking of HVAC equipment can detect faults but they are best suited to offline (after the fact) analysis and do not account

for weather/schedule variability. Accuracy and detection speed are important from an operational point of view and require the use of high frequency process measurements.

The AHU is an important component of HVAC plants. Its main function is to control indoor air quality by continuously supplying conditioned outdoor air. The mixed air, carefully dosed mixture of outdoor air and return air from the indoor zone, is conditioned in the AHU. The conditioning is achieved by passing the mixed air flow over the heating and/or cooling coils, as necessary. Once the air has been conditioned to the prescribed temperature/humidity levels, it is supplied to the indoor environment. The prototypical AHU system considered in this study does not include an energy recovery system.

Categories of faults in AHU include mechanical failures, control problems, design errors and inappropriate operator intervention. A series of papers summarize up to 15 major faults for an AHU system including exhaust air damper stuck, return fan fault, cooling coil valve control fault, outside air damper stuck, cooling coil valve stuck, heating coil valve leakage, outside air damper leakage, heating coil fouling, heating coil capacity reduction, AHU duct leakage, outside air temperature sensor bias, air filter area blockage, mixed air damper unstable, supply fan control unstable and outside air damper stuck [2–5]. In this work, we investigate faults most likely to occur in summer operating conditions. Real world data provided

* Corresponding author. Tel.: +971 2 8109149; fax: +971 2 8109901.
E-mail address: aafshari@masdar.ac.ae (A. Afshari).

Nomenclature

E_{EX}	enthalpy rate of exhaust air (kW)
E_{MA}	enthalpy rate of mixed air (kW)
E_{OA}	enthalpy rate of outside air (kW)
E_{RA}	enthalpy rate of return air (kW)
E_{SA}	enthalpy rate of supply air (kW)
F	F-measure
FN	false negative
FP	false positive
$G(t)$	gain function
H_{MA}	mixed air humidity ratio (kg/kg)
H_{OA}	outside air humidity ratio (kg/kg)
H_{RA}	return air humidity ratio (kg/kg)
K	number of faults
M	number of measured process variables
$\dot{m}$	mass flow rate (kg/s)
$\dot{m}_{MA}$	mass flow rate of mixed air (kg/s)
$\dot{m}_{OA}$	mass flow rate of outside air (kg/s)
$\dot{m}_{SA}$	mass flow rate of supply air (kg/s)
N	number of data vectors
$P(t)$	variance-covariance matrix
Q_{coil}	heat transfer rate of the cooling coil (kW)
t	time step
T	temperature ($^\circ$F)
TP	true positive
$v_M(t)$	measured process variable
V	objective function
X	training dataset
$x(t)$	data vector
$y(t)$	observed stationary process
$z(t)$	class label
Z	set of class labels
Z_1	class labels training subset
Z_2	class labels testing subset
Δ	sampling interval
$\epsilon(t)$	prediction error
$\theta(t)$	parameter vector
Θ	set of parameters
Θ_1	parameter vector training subset
Θ_2	parameter vector testing subset
λ	forgetting factor
τ_f	exponential forgetting time constant
$\phi(t)$	vector of exogenous variables

directly to us by the authors of ASHRAE report 1312-RP [6–8] is utilized to train and validate our algorithm.

The proposed FDD algorithm implements a novel hybrid method combining an autoregressive model with exogenous inputs (ARX) and support vector machine (SVM) technique to detect and diagnose AHU faults. First, the original dataset is divided into two parts; the first part of data with known fault types constitutes the training dataset. The second part of data, where fault labels are initially hidden, is called the testing dataset and is used for validation purposes. Both training and testing subsets are pre-processed by an online model identification algorithm and converted into a set of parameter vectors, one per sampling instant. It is essential to retain the sequential disposition of the data at this stage, given the assumed time series nature of the process. Thereafter, the parameter dataset is randomly shuffled and subdivided into training (2/3 of the data) and testing (1/3 of the data) subsets. The random shuffling simulates a stationarized process. The resulting training and testing parameter subsets are then processed by SVM first for supervised learning, second for validation. In the validation stage, *F-measure*

scores are determined (see details in Section 3.4). High *F*-measure is achieved indicating superior prediction accuracy and exceptionally low false alarm rates in identifying different types of AHU faults. Alternative methods are tested and shown to be inferior. The proposed method is shown to be fairly robust to model parameter uncertainty.

1.1. Related work

Studies on FDD are numerous and various methods have been reported for AHUs in the literature. Venkatasubramanian et al. [9–11] present an overview of FDD methods applicable to building HVAC systems; popular methods for HVAC systems which process large databases include expert systems, neural network models, principal component analysis (PCA), support vector machines (SVM), and combinations of these techniques.

Henley [12] first applied the expert system to diagnose faults. Henley's inference algorithm used a set of if-then rules to determine the state of a system and provide a possible source of the fault. Ramesh et al. [13] introduced a hierarchical classification to improve Henley's approach by focusing on the information processing tasks underlying diagnostic reasoning. Their framework was successfully applied to the development of knowledge-based diagnostic systems for several processes. Schein et al. [14] developed a rule-based fault detection method derived from mass and energy balances for AHUs. Their rule-based system was computationally simple enough that it could be embedded in commercial building automation and control systems. Norford et al. [15] developed two methods for FDD in HVAC equipment. One of the methods used first-principles-based models of system components. The data used by this approach was obtained from sensors typically installed for control purposes. The second method was based on semi-empirical correlations of sub-metered electrical power with flow rates or process control signals. As a result, the first-principles-based models require a larger number of sensors than the electrical power correlation models, although the latter method requires power meters that are not typically installed. Li and Braun [16] formulated model-based FDD for vapor-compression air conditioning equipment in a generic way by developing a physical decoupling methodology as an alternative to mathematical decoupling.

Farell and Roat [17] modified the traditional back-propagation neural network for fault diagnosis by incorporating a small amount of process knowledge which minimized data limitations (since neural networks perform only as robustly as the data from which they are trained) and maximized the neural network's performance. Lee et al. [18] developed a general regression neural network (GRNN) model at the subsystem level of AHU. Their results showed that the GRNN model is an accurate and reliable estimator for highly non-linear and complex AHU processes. Magoulès et al. [19] developed a recursive deterministic perceptron (RDP) neural network for FDD at the building level. Based on their results which showed a higher than 97% generalization performance, they proposed a new diagnostic architecture that reported both the source of the faults and their degradation likelihood.

Wang et al. [20,21] and Xiao [22] applied the PCA method and diagnosed sensor faults for both chiller and AHU systems. They developed two models based on PCA analysis of sensor readings which could group sets of correlated variables and capture the systematic trends of the system. The models were also used to estimate the bias magnitudes of the sensors. Du et al. [23] applied a PCA model to detect sensor faults including fixed bias, drifting bias and complete failure in air dampers and VAV terminals.

In order to maximize the inherent advantages of each technique, many researchers use various combinations of the aforementioned techniques in the last two decades; Becraft and Lee [24] combined neural networks and expert systems to develop a diagnostic

strategy that could handle novel, previously unreported or multiple faults, and noisy process sensory measurements. Chen et al. [25] presented an integrated framework for FDD which combined wavelet analysis and neural networks. Their model, which used multiscale wavelet analysis to determine the singularities of transient signals, proved to be effective in dealing with the noisy transient signals. Zhao et al. [26–28] compared machine learning techniques, such as support vector data description (SVDD) and Bayesian belief network, to PCA based methods on real-world chiller data. Their experimental results showed that the machine learning methods outperformed the PCA based method. Han et al. [29] proposed a hybrid model that combined a support vector machine (SVM) with a genetic algorithm (GA). Their method showed an improvement in performance with respect to the true positive and false alarm rates for the refrigerant leak and refrigerant overcharge faults. Li et al. [30] developed a hybrid fault detection strategy for AHUs based on principal component analysis (PCA) method and pattern matching method. They used the pattern matching method as a form of pre-processing to locate periods of operation from a historical data set whose operational conditions are similar to the target operating condition and then built PCA models for these identified periods. Wang et al. [31] also developed a hybrid approach integrating model-based and rule-based FDD. They used a genetic algorithm-based optimization method to adjust model parameters and a simple threshold method to detect faults. For the diagnosis task, they used three rule-based fault classifiers.

In more recent work, online parameter-based FDD tools are developed with the ability to integrate with a modern BMS in order to leverage the capabilities of these BMS. Various approaches are utilized to estimate parameters online, such as ARX [32,33], recursive least squares [34–36] and Kalman Filtering [37,38]. After the construction of these models, measured data is converted into a set of residuals, parameters or performance indices to enhance FDD accuracy.

Kim et al. [39] used sliding windows to detect and diagnose faults from residuals for a vapor compression chiller. Bonvini et al. [37] proposed a nonlinear state-space model using unscented Kalman Filtering (UKF) and performed FDD based on parameter variations.

Additionally, the residuals or parameter vectors can then also be used as input features for various machine learning techniques. Liang and Du [40] transformed measured data to residuals using a regression model and used the residuals as inputs to the SVM classifier. Yan et al. [41] proposed an ARX model estimator to convert chiller data into parameters which were classified by SVM.

1.2. Approach

The proposed FDD approach combines an ARX model with SVM technique to detect and diagnose AHU faults. Given a set X of process data measured at equidistant time intervals (time step: 1 minute) including both normal and faulty operational periods, each data vector $x(t)$ belonging to X consists of M "features" (measured process variables): $v_1(t), v_2(t), \ldots, v_M(t)$. Furthermore, each $x(t)$ belongs to a class $z(t)$, where $z(t)$ can take any value in $Z = \{0, 1, \ldots, K\}$. A value of 0 indicates normal operation and a non-zero index corresponds to one of K faults. We select the M features from the initial set of features using a feature selection algorithm called *ReliefF* [42] and construct a time series ARX model which relates one of the features (called "dependent variable") to the remaining features (called "exogenous variables"). For each data vector $x(t)$ and a limited number of its lags (representing process dynamic behavior), we estimate a parameter vector $\theta(t)$ uniquely specifying the state of the ARX model at time t. By gathering all θ, the original dataset X is effectively converted into a parameter set Θ. Although

it is essential to retain the sequential disposition of the original data during the parameter estimation stage (due to the dynamic nature of the process), the parameter dataset can be randomized prior to the classification stage without loss of information in order to obtain a more reliable indication of the performance of the FDD method. The ARX model is a dynamic time series model in nature; however, it is reasonable to assume that the system reaches its stationary state before the occurrence of each fault. Additionally, it is assumed that the transition from normal to faulty operation does not affect the basic structure of the system; only the parameters differ. A multi-class SVM model is subsequently trained on the parameter subset Θ_1 with the knowledge of the associated label subset Z_1. In the testing phase, for each parameter vector in subset Θ_2 estimated, a label (0 or one of the K faults) is assigned by the SVM classifier, which completes the fault detection and diagnosis process. Performance is evaluated by comparing estimated to actual labels Z_2.

1.3. Contribution

Utilizing a minimum number of sensors to achieve satisfactory F-measure rate is always a primary objective for application-oriented FDD strategies. In this study, we utilize only three features from the original list of over one hundred features. Both fault detection and diagnosis are completed in one phase, using a multi-class SVM. In order to obtain a realistic assessment of the algorithm's performance, since the normal AHU data do not originate from the same AHU as the faulty data, we calculate the average F-measure over faulty data only (the diagnosis of the normal state is artificially enhanced by the slight differences between the two AHUs). The resulting average F-measure value is equal to 0.923. The high F-measure value indicates superior prediction accuracy and negligible false positive rates. The method is shown to be robust in presence of parameter uncertainty.

2. Background

2.1. Experimental data

The AHU experimental data is provided by ASHRAE project 1312-RP entitled "Tools for evaluating fault detection and diagnostic methods for air-handling units" [6–8]. Two sets of measured data are produced by two AHU systems running simultaneously. The two systems are named AHU-A and AHU-B, where AHU-A runs normally and AHU-B simulates different fault conditions. The resulting data is recorded every minute, in three different seasons: spring, summer and winter. Each fault condition is grouped in daily sets (1440 data points per set). The daily sets corresponding to different process conditions are not necessarily sequential.

Based on the results of the AHU fault survey in Ref. [8] and our understanding of the most common faults in our region of interest, the dataset from the summer of 2007 is chosen in our study, which includes 11 typical faults:

1. Exhaust air damper: stuck ($F1$)
2. Return fan: stuck at fixed speed ($F2$)
3. Return fan: complete failure ($F3$)
4. Cooling coil valve control: unstable ($F4$)
5. Outside air damper: fully closed stuck ($F5$)
6. Cooling coil valve: fully open stuck ($F6$)
7. Cooling coil valve: fully closed stuck ($F7$)
8. Cooling coil valve: partially open stuck ($F8$)
9. Outside air damper: leaking ($F9$)
10. AHU duct: leaking before supply fan ($F10$)
11. AHU duct: leaking after supply fan ($F11$)

Table 1
Three most important features/variables selected from the original dataset.

Feature	Description
SA_TEMP	Supply air temperature
MA_TEMP	Mixed air temperature
SA_HUMD	Supply air humidity

The data comprises 11 AHU fault data files and 11 normal operation data files. Each file includes 1440 data points (24 h, sampling period 1 minute). In total, there are 12 possible labels for sample: Normal, $F1, F2, \ldots, F11$.

Before the experimental data is processed for parameter estimation, each fault dataset generated from AHU-A is matched with the normal dataset generated from AHU-B during the same time period. Therefore, corresponding to datasets $F1, F2, \ldots, F11$, there are normal datasets $N1, N2, \ldots, N11$, and the whole dataset is arranged as $\{N1, F1\}, \{N2, F2\}, \ldots, \{N11, F11\}$. This arrangement simulates a situation where different faults occur after running a period of normal operation, even though the dataset pairs actually correspond to different (albeit similar) machines. This artificial arrangement and the underlying assumption of continuity of data when transiting from normal operation in AHU-B to faulty operation in AHU-A does not undermine the validity of our results since the focus of this study is on diagnosis (ability to distinguish among different system conditions) rather than detection. Furthermore, the two AHUs are identical and daily schedules and boundary conditions do not differ in any significant way from day to day, so it can be safely assumed that any discontinuity in the measured data will be substantially due to a change in process condition.

2.2. Feature selection by ReliefF

Over one hundred monitored variables in the original dataset create heavy computational load for models and classifiers if all were to be analyzed. In real implementation, chances are that some sensors are quite difficult or rather expensive to retrofitted. It may be not practical or economical to capture all the variables. Therefore, a feature selection technique is crucial to select the most relevant and practical variables.

ReliefF was proved to be a strong and successful attribute estimator [43,44]. For each attribute A, the ReliefF algorithm assigns a weight $W(A)$ according to its influence on output. The most crucial attributes are selected as model features.

The top three most significant variables selected by the ReliefF algorithm are listed in Table 1. Best results are obtained by choosing SA_HUMD (supply air humidity) as the dependent variable to be modeled as ARX process. The other two features serve as independent or exogenous variables in the ARX models. The significance of the three selected variables is verified in Section 2.3.
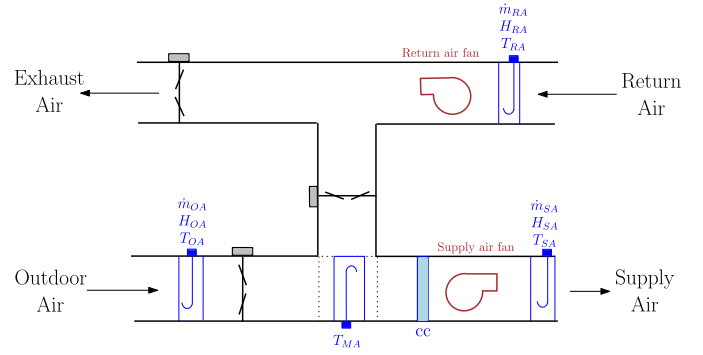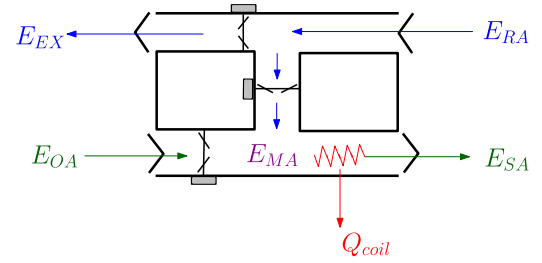
2.3. Energy balance of AHU

A schematic of AHU and sensors is shown in Fig. 1. The energy balance of an AHU is represented in Fig. 2 and can be written as:

$$E_{OA} + E_{RA} - Q_{coil} - E_{SA} - E_{EA} = 0,$$

and

$$E_{MA} - Q_{coil} - E_{SA} = 0, \tag{1}$$

where E_{OA}, E_{RA}, E_{SA}, E_{EA} and E_{MA} are the enthalpy rates of the outside air, return air, supply air, exhaust air and mixed air, respectively. Q_{coil} is the heat transfer rate of the cooling coil. The enthalpy depends on flow rate $\dot{m}$, specific humidity H and temperature T. To

Fig. 1. Schematic of AHU and measurement instrumentation.

Fig. 2. Energy balance of AHU.

simplify the expression of Q_{coil} in Eq. (1), we express $\dot{m}_{MA}$ and H_{MA} as follows:

$$\dot{m}_{MA} = \dot{m}_{SA}, \tag{2}$$

$$H_{MA} = \frac{H_{RA} \cdot (\dot{m}_{SA} - \dot{m}_{OA}) + H_{OA} \cdot \dot{m}_{OA}}{\dot{m}_{SA}}, \tag{3}$$

Eq. (3) is derived from the water vapor mass balance:

$$H_{MA} \cdot \dot{m}_{MA} = H_{RA} \cdot \dot{m}_{RA} - H_{EA} \cdot \dot{m}_{EA} + H_{OA} \cdot \dot{m}_{OA},$$

From Eqs. (2) and 3, we derive:

$$H_{MA} = \frac{H_{RA} \cdot \dot{m}_{RA} - H_{EA} \cdot \dot{m}_{EA} + H_{OA} \cdot \dot{m}_{OA}}{\dot{m}_{MA}}. \tag{4}$$

Q_{coil}, presumably the most important variable for our application, is often not monitored in legacy BMS and the retrofit of a thermal power meter is intrusive and expensive. As shown above, Q_{coil} is a function of three temperature measures: T_{SA}, T_{MA} and T_{RA} (or SA_TEMP, MA_TEMP, RA_TEMP); two humidity measures: H_{SA} and H_{RA} (or, indirectly, relative humidities SA_HUMD and RA_HUMD) and two air flow rate measures: $\dot{m}_{SA}$ and $\dot{m}_{RA}$ (or, indirectly, fan speeds SF_SPD and RF_SPD). To simplify the ARX model and practical implementation, we drop RA_TEMP, RA_HUMD and the two air flow rate measures; and retain only SA_HUMD, SA_TEMP and MA_TEMP. Best results are obtained when SA_HUMD is the dependent variable while SA_TEMP and MA_TEMP are exogenous variables.

2.4. ARX model

2.4.1. Pre-whitening
Pre-whitening is required to determine the significant lags for each exogenous variable [45]. After stationarization, the exogenous data is first differentiated and then fitted into an approximate AR model with the order determined by minimizing the Akaike Information Criterion (AIC) [46].

2.4.2. Lag selection
The candidate lags are selected by the pre-whitening process, the maximum lag for all independent variables is 2. Similarly, for

the dependent variable, the most informative lags are 1 and 2. The general ARX model can therefore be expressed as:

$$SA\_HUMD_t = \theta_1 + \theta_2 \cdot SA\_HUMD_{t-1} + \theta_3 \cdot SA\_HUMD_{t-2} + a_t$$
$$+ \theta_4 \cdot SA\_TEMP_t + \theta_5 \cdot SA\_TEMP_{t-1} + \theta_6 \cdot SA\_TEMP_{t-2}$$
$$+ \theta_7 \cdot MA\_TEMP_t + \theta_8 \cdot MA\_TEMP_{t-1}$$
$$+ \theta_9 \cdot MA\_TEMP_{t-2}$$

Each data vector, in combination with its 2 previous lags, corresponds to a set of 9 parameters at any given time step.

3. Fault detection and diagnosis of AHU faults

3.1. Recursive least squares

A procedure of generating parameters of an ARIMA (p,d,q) model that is based on the use of the recursive least square method with exponential weighting and constant forgetting factor is described in this section. We assume that a model structure has already been specified using the pre-whitening process. With regard to non-stationarity, since an appropriate dth difference of the observed series is assumed to be a stationary ARMA(p,q) process, we only need to consider the problem of estimating the parameters in such stationary models. In practice, we then treat the dth difference of the original time series as the time series from which the parameters of the complete model are estimated [47]. For simplicity, let $y(t)$ denote our observed stationary process even though it may be an appropriate difference of the original series.

In least square estimation, unknown parameters of a linear model are chosen such that the sum of the squares of the difference between the actually observed and the computed values, is minimized [48].

The ARX process can be represented in the following linear parametric form,

$$y(t) = \phi^T(t)\theta(t) \tag{5}$$

where $y(t)$ is the observed variable, $\theta(t)$ is the vector of model parameters to be determined and $\phi(t)$ is the vector of exogenous variables including appropriate lags. Least Squares estimation over an observation period spanning n data samples translates into finding the parameter $\hat{\theta}$ that minimizes the following objective function:

$$V(\theta, n) = \frac{1}{2} \sum_{i=1}^{n} \left(y(i) - \phi^T(i)\theta \right)^2 \tag{6}$$

Minimizing Eq. (6), we get the closed form Ordinary Least Squares (OLS) solution as follows:

$$\hat{\theta} = \left(\sum_{i=1}^{n} \phi(i)\phi^T(i) \right)^{-1} \cdot \left(\sum_{i=1}^{n} \phi(i)y(i) \right) \tag{7}$$

3.1.1. Forgetting factor

In our application (real time FDD), we are interested in online parameter estimation. Therefore it is computationally more efficient if we update the estimates in Eq. (7) recursively as new data becomes available.

This requires forgetting measurements that are too old, because they correspond to an out-of-date situation and would distort estimation when the "true" system parameters evolve (e.g. due to a fault condition). A particularly simple technique for this purpose is exponential forgetting, which weights prediction errors in the

cost function exponentially, decreasing with time elapsed according to a certain time constant, called "forgetting time constant". The recursive algorithm with forgetting is represented by the equations below:

$$\epsilon(t) = y(t) - \phi^T(t)\theta(t-1) \tag{8}$$

$$r(t) = \phi^T(t)P(t-1)\phi(t) \tag{9}$$

$$G(t) = \frac{P(t-1)\phi(t)}{1+r(t)} \tag{10}$$

$$\theta(t) = \theta(t-1) + G(t)\epsilon(t) \tag{11}$$

$$P(t) = \frac{1}{\lambda} \left[P(t-1) - \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{1+r(t)} \right] \tag{12}$$

where P, ϵ and λ are, respectively the parameter variance-covariance matrix, the prediction error and forgetting factor. G represents the algorithm's gain function which is an intermediate step in the update phase of the algorithm.

The corresponding time constant for the forgetting factor λ, is obtained from Ref. [49]:

$$\lambda = e^{-\Delta/\tau_f}, \tag{13}$$

where τ_f is the exponential forgetting time constant, λ is the forgetting factor and Δ is the sampling interval.

3.1.2. Model validation

The ARX model was identified using the dataset covering normal operation ($N1$, $N2$, ..., $N11$). The identified model was then validated using out-of-sample data (1/3 of the original dataset not used during the estimation process). Fig. 3 shows that the model's prediction performance is excellent with a Root-Mean-Square Deviation (RMSE) of 0.26 %RH (less than 1% of the peak).

3.2. Support vector machine

Support vector machine is a supervised machine learning technique which separates and classifies available data points by hyper-planes in high dimensional space [50]. A binary SVM library deals with only two classes (normal and faulty). Key settings of the binary SVM include the kernel function and SVM formulation. In this study, we used the radial basis function (RBF) kernel function and C-support vector classification (C-SVC) to divide the dimensional space.

During the FDD process, a multi-class SVM library is implemented following the one-against-all approach [51]. The idea behind the multi-class SVM library is to convert the multi-class classification problem into a set of binary classification problems. Therefore, in the one-against-all approach, a series of binary SVM libraries are combined, where each of the binary libraries separates one class from all other classes.

3.3. An overview of the proposed hybrid method

A hybrid approach combining ARX model identification and SVM classification is proposed for highly accurate FDD analysis. The proposed method includes the following stages:

(1) Using ReliefF to select most important features. The dependent and independent variables are chosen, respectively. The dependent variable is expressed as a function of independent variables and lags thereof.
(2) According to the data arrangement in Section 2.1, the original dataset is divided into eleven parts. Each part contains a pair of normal and faulty datasets. Using RLS, an ARX model with optimized forgetting factor is estimated with selected features for
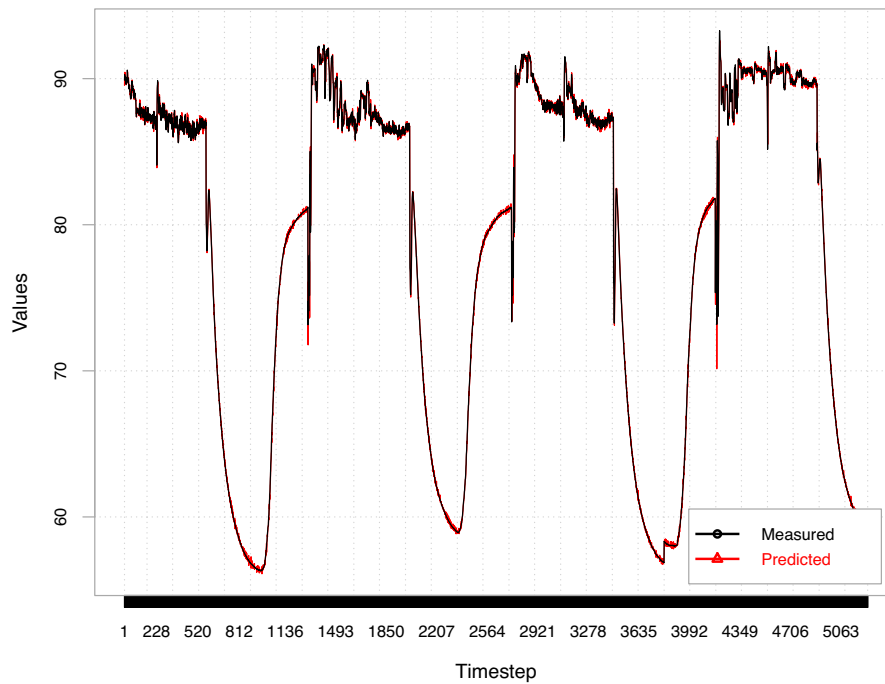
Fig. 3. Out-of-sample fit for supply air humidity model.

all dataset pairs. The model estimation procedure generates a vector of nine parameters at each time step. The whole original data set is transformed into a set of parameter vectors representing the state of the ARX model at the same time intervals as the original data. We plot the parameter vectors for the pairs $\{N1, F1\}$ and $\{N11, F11\}$ in Figs. 4 and 5 as examples.

(3) Fault detection and diagnosis:

(3.1) Labels are simplified by removing any distinction among normal files: $N = \{N1, N2, \ldots, N11\}$. Since each parameter vector at a given time instant is assumed to fully characterize the dynamic properties of the system, we do not need to retain the original order of the parameter dataset. The order of all parameter vectors is randomized regardless of their labels ($N, F1, F2, \ldots, F11$).

(3.2) The resulting parameter vector set is divided into two subsets. The first subset (2/3 of the data) is used for training.

The second subset (1/3 of the data) is used for testing. The labels of testing parameter vectors are initially hidden.

(3.3) A multi-class SVM library is trained using the training subset.

(3.4) The testing subset is input into the multi-class SVM library. A class label ($N, F1, F2, \ldots, F11$) is assigned to each parameter vector in the subset.

Finally, the average F-measure [52] over the faulty data is calculated. The overall flowchart is depicted in Fig. 6.

3.4. Testing result evaluation

The F-measure is used to evaluate the detection and diagnosis performance in this study. The F-measure is directly derived from the confusion matrix (shown in Table 2) given by trained model. In
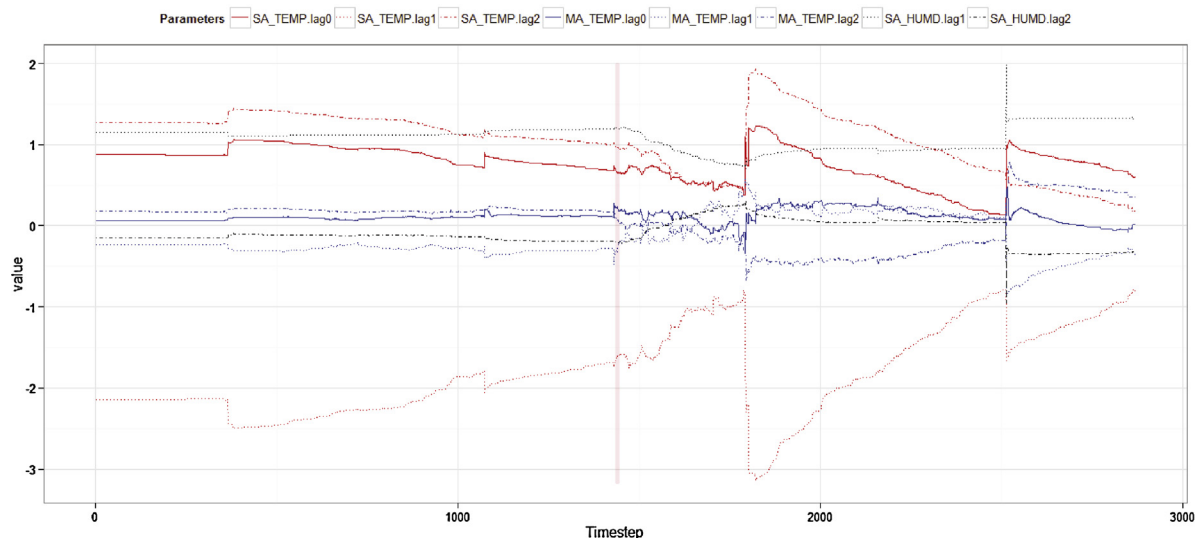

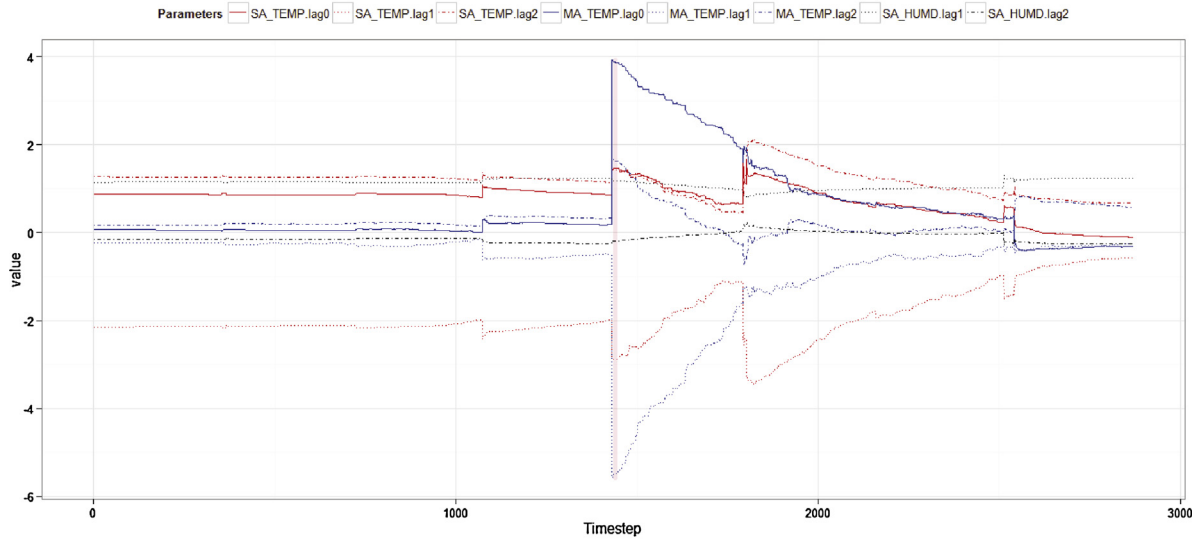**Fig. 4.** Plot of estimated parameters for pair $\{N1, F1\}$.

Fig. 5. Plot of estimated parameters for pair {*N11*, *F11*}.

Table 2
Confusion matrix example.

	Actual classes	
	True	False
Testing result		
Positive	True positive (*TP*)	False positive (*FP*)
Negative	False negative (*FN*)	True negative (*TN*)

a binary classification problem, e.g. fault detection, the *F*-measure is defined in Eq. (14) as,

$$F_b = 2 \cdot \frac{precision \cdot recall}{precision + recall}, \tag{14}$$

where the precision and recall are defined as:

$$precision = \frac{TP}{TP + FP}, \tag{15}$$
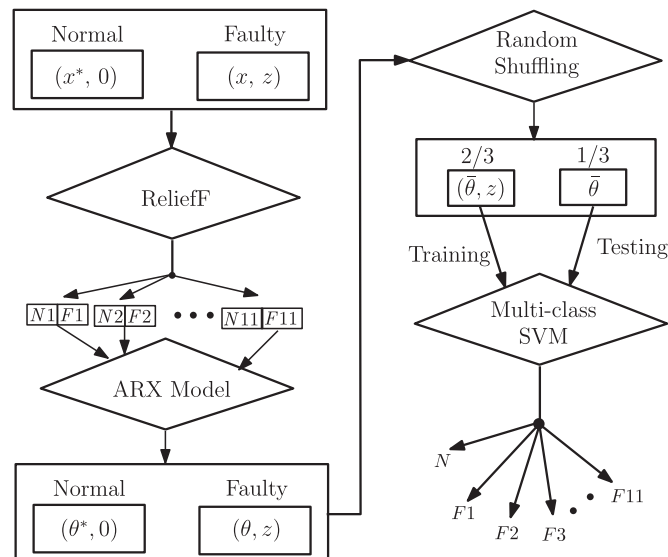
$$recall = \frac{TP}{TP + FN}, \tag{16}$$

Fig. 6. Flow chart of hybrid approach for FDD.

where *TP*, *FP* and *FN* indicate *true positive, false positive* and *false negative*, respectively.

In multiple classification problems, we take the average value of F_b in all different classes as the final *F*-measure.

$$F\text{-}measure = \frac{\sum F_b}{number\ of\ classes} \quad (0 \leq F\text{-}measure \leq 1) \tag{17}$$

4. Validation using real world data

4.1. Data pre-processing

4.1.1. Initial state generation

Each ARX model starts with an initial state, which is the vector of parameters θ_0 and there corresponding variance-covariance matrix P_0. θ_0 is generated applying OLS to all normal datasets, i.e. {*N1*, *N2*,...,*N11*}. The parameter vector θ_0, which represents the average normal parameter vector in the stationarized process, is therefore used as the initial state for all ARX models in different Normal/Faulty dataset pairs.

4.1.2. Random shuffling

Random shuffling of the training parameter set reduces the ambiguity during the transient period (i.e. when the AHU system switches from normal to faulty). It simulates a stationarized process. Therefore, the final result is more reliable if the machine learning classification methods are trained with shuffled data.

4.1.3. Optimizing the forgetting factor

Different choices of forgetting factors lead to different *F*-measures for fault diagnosis since there exists a trade-off between using only the latest samples (small window) for precise parameter estimation and a larger window for stable parameter estimation. Fig. 7 depicts the relationship between the *F*-measure and forgetting factor with an apparent *F*-measure maximum on the forgetting factor interval [0.998, 1.000]. The forgetting factor corresponding to the absolute *F*-measure maximum, λ = 0.9964, was determined via the golden section search algorithm [53] and it corresponds to a forgetting time constant of approximately 277 samples.

4.2. Result and discussion

The final performance result is shown in Table 3. The overall *F*-measure of the proposed hybrid approach is 0.923, which is
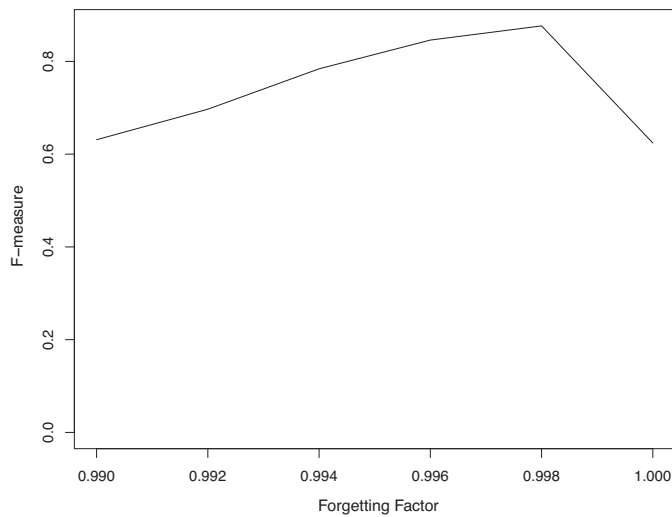
Fig. 7. Validation *F*-measure values versus forgetting factors for fault diagnosis.

significantly better than applying LibSVM [54] on the original "raw" data (*F*-measure 0.784). The same is true for the other machine learning methods, such as: Naïve Bayes [55], Bayesian Network [56], RBF Network [57], multilayer perceptron neural network [58] and random forest decision tree [59] (see Table 4).

4.2.1. Model uncertainty sensitivity test

A sensitivity analysis is conducted to test the robustness of the proposed method to model uncertainty. Table 5 shows the variation of the *F*-measure when applying SVM and random forest decision tree on the parameter vectors after adding noise (with the same standard deviation as the initial parameter vector θ_0) to the testing parameter dataset. The test is repeated one hundred times with

Table 3
The performance of the proposed method using different metrics: precision, recall and *F*-measure.

	Metrics		
	Precision	Recall	*F*-measure
*F*1	0.960	0.888	0.923
*F*2	0.684	0.949	0.795
*F*3	0.977	0.812	0.887
*F*4	0.972	0.924	0.948
*F*5	0.994	0.994	0.994
*F*6	0.941	0.673	0.785
*F*7	0.871	0.994	0.928
*F*8	0.902	0.935	0.918
*F*9	0.991	0.998	0.994
*F*10	0.998	0.966	0.981
*F*11	1.000	1.000	1.000
Overall	0.935	0.921	0.923

Table 4
A comparison of the overall performance obtained from different methods using different metrics: precision, recall and *F*-measure.

	Metrics		
	Precision	Recall	*F*-measure
Proposed	0.935	0.921	0.923
Naive Bayes	0.249	0.197	0.172
Bayes network	0.556	0.546	0.547
RBF network	0.629	0.668	0.617
Multilayer perceptron	0.506	0.481	0.477
LibSVM	0.787	0.785	0.784
Random tree	0.795	0.794	0.795
Random forest	0.817	0.817	0.817

Table 5
Noise-sensitivity test: applying SVM and random forest decision tree on the parameter vector set. The testing parameter vector set is augmented with noise. The test runs for one hundred times.

Runs	Ave.	Std. dev.
SVM	0.920	0.107
Random forest	0.905	2.174

Table 6
Model comparison: we compare *F*-measures by replacing the ARX model by a regression model and a regression model with one additional feature.

Model	Forg. fac. (λ)	*F*-measure
Proposed	0.9964	0.923
Regression 1	0.9691	0.274
Regression 2 (with Q_{coil})	0.9742	0.482

different seeds. While the random forest decision tree can be highly accurate when the real-time system is exactly the same as the one used in training, lower average *F*-measure and higher standard deviation are shown in presence of parameter uncertainty. This result demonstrates the distinctly superior robustness of the proposed SVM method.

Robustness of the classification method to model uncertainty is of particular importance in our work. The proposed approach consists in testing, in lab conditions, a standard piece of HVAC equipment (an AHU in this case). First, we run the equipment in normal mode, for at least a day. Then, we artificially introduce a fault. Faulty condition is stabilized and remains in effect for at least one day. Then we switch back to normal operation and the procedure is repeated for different faults. This data is used to train the model-based classifier. Once the algorithm is trained, it is implemented in a multitude of buildings having the same (or similar) piece of equipment. The on-line diagnosis procedure consists both continuously estimating system parameters via RLS and sending the estimated parameter vector to the trained classifier in real time. There will be possibly small differences between the equipment used for training in the lab and the one installed in the building, which is the purpose of performing the sensitivity assessment of the selected classifier to small variations in system characteristics. The ultimate classifier is selected based not only on accuracy but also on robustness, i.e., stability to model uncertainty.

4.2.2. Model comparison

We replace the ARX model in our method with a stationary regression model built using the same number of features (without lags). We also added one more feature (Q_{coil}: heat transfer rate of the cooling coil) to the regression model. All forgetting factors and *F*-measure are depicted in Table 6. The *F*-measure values of using the two regression models is significantly lower than using the ARX model.

The superiority of the proposed method is attributed to two factors: first, the dynamic nature of the system is more consistent with the time series formulation which provides a more informative feature set to the classifier; second, the support vector machine as a strong machine learning technique demonstrates its predictive power and robustness in identifying various faults in the current application.

We develop our approach based on the fact that faults and symptoms can be studied using changes in the parameters of a dynamic model continuously tuned to the process. When the process operates under normal conditions, model parameters do not deviate significantly from their normal values. In case of a defect in the equipment or the sensor(s), the process parameters significantly deviate from their normal values and often, after a transient

period of less than 24 h, converge to values corresponding to the new (faulty) condition, as depicted in Figs. 4 and 5.

We limited ourselves to diagnosing abrupt faults and 'Level 1' faults among the gradual faults (for example, several positions for a partially open cooling coil valve) because we did not find a significant difference in the diagnosis of higher level faults. Most higher level faults were generally easier to diagnose than Level 1 faults.

5. Conclusion

In this paper, we proposed a robust parameter-based approach incorporating ARX model and SVM classification techniques to diagnose the faults commonly observed in air handling units. The proposed method utilizes a mathematical model of the monitored system to generate secondary features for the classification problem. The model parameters, which define the "state" of the system, represent the unobservable marginal effect of the exogenous variables on the dependent variable. We implemented the proposed method and evaluated its performance using the ASHRAE project 1312-RP dataset. We also compared the performance of our method with other machine learning approaches such as naive Bayes, RBF network, LibSVM, multilayer perceptron neural network and random forest decision tree applied to the raw data. Results show that the proposed method overwhelmingly outperforms all other approaches in terms of Precision, Recall and *F*-measure. The proposed method produced a high *F*-measure of about 0.923 on average, while only three features (variables) were required. This will significantly reduce the deployment cost of sensors, without compromising FDD performance. We attribute this superiority to the accurate representation of the process dynamics using an ARX model and the high predictive power and robustness of support vector machines.

Future work includes devising an FDD strategy that utilizes a sliding window of fixed size with variable forgetting applied only in the pre-determined window. This window size would be determined by the optimal forgetting factor, as described in Section 4. Future work also includes exploring the dynamic behavior of the training data using a Kalman filter. The Kalman filter overcomes the issue of selecting an optimal forgetting factor and the memory-less characteristic is particularly convenient for online applications.

References

[1] S. Katimapula, M.R. Brambley, Methods for fault detection, diagnostics, and prognostics for building systems – a review, Part I, HVAC&R Res. 11 (1) (2005) 3–25.
[2] J.M. House, W.Y. Lee, D.R. Shin, Classification techniques for fault detection and diagnosis of an air-handling unit, ASHRAE Trans. 105 (1) (1999).
[3] W.Y. Lee, J.M. House, C. Park, G.E. Kelly, Fault diagnosis of an air-handling unit using artificial neural networks, ASHRAE Trans. 102 (1) (1996) 540–549.
[4] W.Y. Lee, C. Park, G.E. Kelly, Fault detection in an air-handling unit using residual and recursive parameter identification methods, ASHRAE Trans. 102 (1) (1996) 528–539.
[5] W.Y. Lee, J.M. House, D.R. Shin, Fault diagnosis and temperature sensor recovery for an air-handling unit, ASHRAE Trans. 103 (1) (1997) 540–549.
[6] S. Li, J. Wen, Development and validation of a dynamic air handling unit model – Part I, ASHRAE Trans. 116 (1) (2010) 45–56.
[7] S. Li, J. Wen, X. Zhou, C.J. Klaassen, Development and validation of a dynamic air fandling unit model – Part II, ASHRAE Trans. 116 (1) (2010) 57–73.
[8] J. Wen, S. Li, Tools for evaluating fault detection and diagnostic methods for air-handling units, ASHRAE 1312-RP (2011).
[9] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, A review of process fault detection and diagnosis, Part I: Quantitative model-based methods, Comput. Chem. Eng. 27 (3) (2003) 293–311.
[10] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, A review of process fault detection and diagnosis, Part II: Qualitative models and search strategies, Comput. Chem. Eng. 27 (3) (2003) 313–326.
[11] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, A review of process fault detection and diagnosis, Part III: Process history based methods, Comput. Chem. Eng. 27 (3) (2003) 327–346.
[12] E.J. Henley, Application of expert systems to fault diagnosis, in: AIChE Annual Meeting, San Francisco, CA, 1984.
[13] T.S. Ramesh, J.F. Davis, G.M. Schwenzer, Knowledge-based diagnostic systems for continuous process operations based upon the task framework, Comput. Chem. Eng. 16 (2) (1992) 109–127.
[14] J. Schein, S.T. Bushby, N.S. Castro, J.M. House, A rule-based fault detection method for air handling units, Energy Build. 38 (12) (2006) 1485–1492.
[15] L.K. Norford, J.A. Wright, R.A. Buswell, D. Luo, C.J. Klaassen, A. Suby, Demonstration of fault detection and diagnosis methods for air-handling units, HVAC&R Res. 8 (1) (2002) 41–71.
[16] Haorong Li, E. James, Braun, A methodology for diagnosing multiple simultaneous faults in vapor-compression air conditioners, HVAC&R Res. 13 (2) (2007) 369–395.
[17] A.E. Farell, S.D. Roat, Framework for enhancing fault diagnosis capabilities of artificial neural networks, Comput. Chem. Eng. 18 (7) (1994) 613–635.
[18] W.Y. Lee, J.M. House, J.M. Kyong, Subsystem level fault diagnosis of a building's air-handling unit using general regression neural network, Appl. Energy 77 (2) (2004) 153–170.
[19] F. Magoulès, H. Zhao, D. Elizondo, Development of an rdp neural network for building energy consumption fault detection diagnosis, Energy Build. 62 (2013) 133–138.
[20] S. Wang, J. Cui, Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method, Appl. Energy 82 (3) (2005) 197–213.
[21] S. Wang, F. Xiao, AHU sensor fault diagnosis using principal component analysis method, Energy Build. 36 (2) (2004) 147–160.
[22] X. Fu, Sensor Fault Detection and Diagnosis of Air Handling Units, 2004 (Ph.D. thesis).
[23] Z. Du, X. Jin, Y. Yang, Wavelet neural network-based fault diagnosis in air handling unit, HVAC&R Res. 14 (6) (2008) 1709–1720.
[24] W. Becraft, P. Lee, An integrated neural network/expert system approach for fault diagnosis, Comput. Chem. Eng. 16 (2) (1993) 1001–1014.
[25] B.H. Chen, X.Z. Wang, S.H. Yang, C. Mcgreavy, Application of wavelets and neural networks to diagnostic system development. I. Feature extraction, Comput. Chem. Eng. 23 (7) (1999) 899–906.
[26] Y. Zhao, S. Wang, F. Xiao, Pattern recognition-based chillers fault detection method using support vector data description (SVDD), Appl. Energy 112 (2013) 1041–1048.
[27] Y. Zhao, F. Xiao, S. Wang, An intelligent chiller fault detection and diagnosis methodology using Bayesian belief network, Energy Build. 57 (2013) 278–288.
[28] Y. Zhao, F. Xiao, J. Wen, Y. Lu, S. Wang, A robust pattern recognition-based fault detection and diagnosis (FDD) method for chillers, HVAC&R Res. (2014) (in press) Taylor & Francis.
[29] H. Han, B. Gua, J. Kang, Z.R. Li, Study on a hybrid SVM model for chiller FDD applications, Appl. Therm. Eng. 31 (4) (2011) 582–592.
[30] Shun Li, Jin Wen, Application of pattern matching method for detecting faults in air handling unit system, Autom. Constr. 43 (2014) 49–58.
[31] Haitao Wang, Youming Chen, CaryW.H. Chan, Jianying Qin, Wang Jinhua, Online model-based fault detection and diagnosis strategy for VAV air handling units, Energy Build. 55 (2012) 252–263.
[32] H. Yoshida, S. Kumar, ARX and AFMM model based online real time database diagnosis of sudden fault in AHU of VAV system, Energy Convers. Manage. 40 (11) (1999) 1191–1206.
[33] H. Yoshida, S. Kumar, Development of ARX model based off-line FDD technique for energy efficient buildings, Renew. Energy 22 (1–3) (2001) 53–69.
[34] L.B. Palma, F.V. Coito, R.N. Silva, Diagnosis of parametric faults based on identification and statistical methods, in: 44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05, 2005, pp. 3838–3843.
[35] E.M. Cimpoesu, B.D. Ciubotaru, D. Stefanoiu, Fault detection and diagnosis using parameter estimation with recursive least squares, Energy Build. 62 (2013) 133–138.
[36] T. Escobet, L. Travé-Massuyès, Parameter Estimation Methods for Fault Detection and Isolation, 2001.
[37] M. Bonvini, M.D. Sohn, J. Granderson, M. Wetter, M.A. Piette, Robust on-line fault detection diagnosis for HVAC components based on nonlinear state estimation techniques, Appl. Energy 124 (2014) 156–166.
[38] N. Tudoroiu, M. Zaheeruddin, E.-R. Tudoroiu, V. Jeflea, Fault Detection and Diagnosis FDD in Heating Ventilation Air Conditioning Systems HVAC Using an Interactive Multiple Model Augmented Unscented Kalman Filter IMMAUKF, 2008, pp. 334–339.
[39] M. Kim, S.H. Yoon, P.A. Domanskib, W.V. Payne, Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner, Int. J. Refrig. 31 (5) (2008) 790–799.
[40] J. Liang, R. Du, Model-based fault detection and diagnosis of HVAC systems using support vector machine method, Int. J. Refrig. 30 (6) (2007) 1104–1114.

[41] K. Yan, W. Shen, T. Mulumba, A. Afshari, ARX model based fault detection and diagnosis for chillers using support vector machines, Energy Build. 81 (2014) 287–295.

[42] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. 53 (1–2) (2003) 23–69.

[43] R.S. Marko, K. Igor, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. J. 53 (1–2) (2003) 23–69.

[44] Y. Sun, Iterative RELIEF for feature weighting: algorithms, theories, and applications, IEEE Trans. Pattern Anal. Mach. Intel. 29 (6) (2007).

[45] E.M. Douglas, R.M. Vogel, C.N. Kroll, Trends in floods and low flows in the united states: impact of spatial correlation, J. Hydrol. 240 (1) (2000) 90–105.

[46] Y. Sakamoto, M. Ishiguro, G. Kitagawa, Akaike Information Criterion Statistics, D. Reidel, Dordrecht, The Netherlands, 1986.

[47] J.D. Cryer, K.S. Chan, Time Series Analysis with Applications in R, Springer, New York, 2006.

[48] D. Weitz, Counting independent sets up to the tree threshold, in: Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing, 2006, pp. 140–149.

[49] S.J. Pollard, Recursive Parameter Identification for Estimating and Displaying Maneuvering Vessel Path, 2003 (Master thesis).

[50] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000.

[51] Y. Liu, Y.F. Zheng, One-against-all multi-class SVM classification using reliability measures, in: IEEE International Joint Conference on Neural Networks, volume 2, IEEE, 2005, pp. 849–854.

[52] M.W. Powers, Evaluation: from precision, recall and *F*-measure to ROC, informedness, markedness and correlation, J. Mach. Learn. Technol. 2 (1) (2011) 37–63.

[53] Jack, Kiefer, Sequential minimax search for a maximum, Proc. Am. Math. Soc. 4 (3) (1953) 502–506.

[54] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intel. Syst. Technol. (TIST) 2 (3) (2011) 27.

[55] I. Rish, An empirical study of the naive Bayes classifier, in: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3(22), 2001, pp. 41–46.

[56] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Mach. Learn. 29 (2–3) (1997) 131–163.

[57] D.J. Swider, A comparison of empirically based steady-state models for vapor-compression liquid chillers, Appl. Therm. Eng. 23 (5) (2003) 539–556.

[58] R. Collobert, S. Bengio, Links between perceptrons, MLPs and SVMs, Proc. ICML (2004).

[59] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.