

Optimal Sensor Selection and Fusion for Heat Exchanger Fouling Diagnosis in Aerospace Systems

Nayeff Najjar, Shalabh Gupta, *Member, IEEE*, James Hare, Sherif Kandil, and Rhonda Walthall

Abstract—Heat exchangers are critical components of the environmental control system (ECS) of an aircraft. The ECS regulates temperature, pressure, and humidity of the cabin air. Fouling of the heat exchangers in an ECS may occur due to the deposition of external substances (e.g., debris) on the fins that obstruct the air flow, which increases the pressure drop across the heat exchanger and degrades its efficiency. Fouling is a critical issue, because it necessitates time consuming, periodic, and expensive maintenance. In this regard, this paper presents a two step process for fouling diagnosis of the heat exchanger: 1) optimal sensor set selection that contains the most relevant information for fault classification and 2) robust data analysis and sensor fusion in the presence of various uncertainties for the inference of fouling severity via different machine learning tools. This process of heat exchanger fouling diagnosis is implemented and tested on the data generated from an experimentally validated high-fidelity Simulink model of the ECS provided by an industry partner.

Index Terms—Heat exchanger fouling, optimal sensor selection, environmental control system, minimum redundancy maximum relevance (mRMR), fault diagnosis.

I. INTRODUCTION

HEAT exchangers are critical components of the Environmental Control System (ECS) of an aircraft. The ECS regulates temperature, pressure and humidity of the cabin air [1]. Typically, several plate fin heat exchangers are used in an ECS, which consist of plates and fins stacked over each other and brazed together. Plate fin heat exchangers are used in this application because of their compact design, light weight and high efficiency. Often, physical objects (e.g., debris) accumulate on the fins of the heat exchanger due to particulates and other contaminants present in the air stream. This phenomenon is known as *fouling* which obstructs the flow of the cooling medium through the heat exchanger and hence degrades its efficiency. Fouling [2] is a critical issue because in absence of a reliable fouling diagnosis methodology, it necessitates time consuming, periodic and expensive maintenance [3]. As such, early fouling diagnosis is of utmost importance to facilitate *Condition Based Maintenance* (CBM), i.e., to clean the heat exchanger only when necessary.

Manuscript received October 26, 2015; accepted March 17, 2016. Date of publication April 1, 2016; date of current version May 17, 2016. This work was supported by United Technologies Aerospace Systems. The associate editor coordinating the review of this paper and approving it for publication was Dr. Akshya Swain.

N. Najjar, S. Gupta, and J. Hare are with the University of Connecticut, Storrs, CT 06269 USA (e-mail: naifnajjar@gmail.com; shalabh.gupta@engr.uconn.edu; james.hare@uconn.edu).

S. Kandil and R. Walthall are with United Technologies Aerospace Systems, Windsor Locks, CT 06096-1000 USA (e-mail: sherif.kandil@hs.utc.com; rhonda.walthall@hs.utc.com).

Digital Object Identifier 10.1109/JSEN.2016.2549860

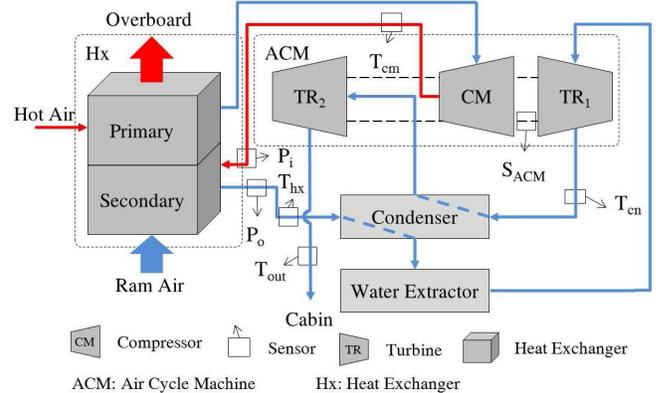


Fig. 1. System diagram of the environmental control system (ECS). Note: Not all sensors are shown in the figure.

The ECS is a complex system consisting of various components such as the primary and secondary heat exchangers, turbines, compressor, condenser, and water extractor [4]. These components are interconnected through various mechanical and pneumatic connections, as shown schematically in Fig. 1. In addition, various sensing devices such as temperature, pressure and flow sensors are mounted at different locations in the ECS. Due to interdependence of this complex system, fouling of a heat exchanger may influence conditions in other locations in the ECS. As a result, its effects might be observed in the readings of various sensors throughout the system.

In this regard, there are three critical issues that need to be addressed for heat exchanger fouling diagnosis. First, in a complex interconnected system such as the ECS, a large number of sensors are available for data collection. For applications considered in this study, more than 100 ECS parameters are recorded. Second, due to the complexity of the system, it is often possible that the sensors that carry the most pertinent information about the heat exchanger fouling may be located away from the close vicinity of the heat exchanger. Therefore, to address both problems, optimal sensor selection and data reduction are essential for efficient fault diagnosis, which is often infeasible by means of visual inspection only. Third, the sensor outputs are stochastic in nature and are susceptible to various sources of uncertainties that arise from variations in operational parameters such as occupant count, day type, and other factors such as mechanical vibrations, sensor biases, and measurement noise. Therefore, it is important to include uncertainties in the data analysis procedure for making robust and accurate classification decisions. Furthermore, it is desired

to fuse the information derived from selected sensors for improved diagnosis accuracy and robustness to sensor failures. The paper focuses on the above issues while primarily addressing the challenges of optimal sensor selection.

Recent literature has developed several feature selection algorithms that are categorized into two main types based on their evaluation criteria. The first in this category are the *Wrapper Algorithms* that depend on the evaluation of the *Correct Classification Rate (CCR)*¹ for each feature using a specified classifier [5]–[7]. Wrapper algorithms guarantee a high CCR but are computationally expensive if the number of features is large because they rely on the *cross-validation* algorithm to calculate the CCR. Besides that, the wrapper algorithms cannot be generalized to any classifier. The second type are the *Filter Algorithms* that evaluate the performance of each feature based on an evaluation function. Recently, many filter algorithms have been developed using the concepts of information theory [8], [9]. Filter algorithms do not depend on the classifier, are computationally less expensive, and may perform as good as the wrapper algorithms [6].

In addition, there exist the *Embedded Algorithms* that take advantage of both the wrapper and the filter algorithms. The embedded algorithms use a filter to select a candidate list of sensors and then apply a wrapper on this list to rank the optimal set of sensors [9]. Embedded algorithms are less expensive than wrappers and more accurate than filters; yet they are pertinent to the specified classifier [10]. Several search methods have been suggested for the above algorithms, such as the forward and backward search [8], [9]. Dash and Liu [11] compared different such search methods.

This paper utilizes the embedded algorithm with the *minimum Redundancy Maximum Relevance (mRMR)* criterion [9] as a filter, and applies it to the sensor selection problem in the ECS to find a candidate list. Subsequently, the paper presents a modified embedded algorithm, called the *Unsupervised Embedded Algorithm* for ranking the candidate list, that relies on the *K*-means clustering method instead of depending on a classifier. This method has low computational complexity, faster execution, and it does not depend on a specific classifier. Once the optimal set of sensors is selected, the paper uses different machine learning tools for data analysis and fusion to make classification decisions.

This paper uses an experimentally validated high fidelity Simulink model of the ECS provided by an industry partner to generate sensor data for nominal and different fouling conditions of the heat exchanger while considering various sources of uncertainties in the system. This data is then used for feature extraction using tools such as the Gaussian Mixture Models (GMM) and the Principal Component Analysis (PCA). The extracted features are then classified using the *k*-Nearest Neighbors (*k*-NN) classifier. The main contributions are below:

- Optimal sensor selection for fouling diagnosis using an existing embedded algorithm that uses the mRMR criteria as a filter and a classifier (*k*-NN) for ranking.
- Optimal sensor selection for fouling diagnosis using a

novel *Unsupervised Embedded Algorithm*, that uses the mRMR criteria as a filter and the *K*-means clustering method for ranking.

- Application of the *Maximum Entropy Principle* for data partitioning to compute probability distributions and to estimate the mutual information in the mRMR criteria.
- Application of different machine learning tools for classification of the heat exchanger fouling severity.
- Validation of the above methods on the data generated from a high fidelity Simulink model of the ECS.

The paper is organized into seven sections and two appendices. Section II presents the relevant background information while Section III describes the ECS system and the data generation process. Sections IV and V describe the optimal sensor selection methodology and the data analysis method for fouling diagnosis, respectively. Finally, results are discussed in Section VI and the paper is concluded in Section VII with recommendations for future work. Appendices A and B describe the maximum entropy distribution and the calculation of mutual information, respectively.

II. LITERATURE REVIEW

Several techniques have been proposed in recent literature for fault detection, diagnosis and prognosis (FDDP) of air-conditioning systems, in particular, *Heating, Ventilating and Air Conditioning (HVAC)* systems [12]–[14]. Katipamula and Brambley introduced a two-part survey of FDDP of HVAC systems [15], [16]. Buswell and Wright [17] accounted for uncertainties in model-based approaches to minimize false alarms in fault diagnosis of HVAC systems. Fault diagnosis of *Air Handling Units (AHU)* was presented in [18]–[20]. Pakanen and Sundquist [21] developed an *Online Diagnostic Test (ODT)* for fault detection of *Air Handling Units (AHU)*. Qin and Wang [22] performed a site survey on hybrid fault detection and isolation methods for *Variable Air Volume (VAV)* air conditioning systems. Rossi and Braun [23] designed a classifier that uses temperature and humidity measurements for fault diagnosis of the *Vapor Compression Air Conditioners*. Zhao *et al.* [24] utilized exponentially-weighted moving average control charts and support vector regression for fault detection and isolation in centrifugal chillers. Najjar *et al.* [25] developed a tool for data visualization, reduction, clustering, and classification of the actual data obtained from flight test reports of the *Liquid Cooling System (LCS)* in aircrafts. Shang and Liu [26] used the *Unscented Kalman Filter (UKF)* to diagnose sensor and actuator faults in the *Bleed Air Temperature Control System*. A model-based approach for fault isolation in *Aircraft Gas Turbine Engines* was presented by Gupta *et al.* [27] and Sarkar *et al.* [28]. Isermann [29] provided a review of model-based fault detection and diagnosis methods.

Heat exchanger fouling diagnosis has become a critical research issue in recent years. Najjar *et al.* [2] presented the fouling severity diagnosis of the *Plate Fin Heat Exchanger* using the principal component analysis (PCA) and the *k*-nearest neighbor classification (*k*-NN). Kaneko *et al.* [30] introduced a statistical approach to construct predictive models for thermal resistance based on operating conditions.

¹CCR is the ratio of correctly classified samples to the total number of testing samples.

TABLE I
LIST OF CRITICAL SENSORS IN THE ECS

S	Description	S	Description
\dot{m}_1	PD mass flow rate	\dot{m}_2	SD mass flow rate
T_1	PD air temperature	T_2	SD air temperature
P_1	PD air pressure	P_2	SD air pressure
P_i	SHX input pressure	P_o	SHX output pressure
T_{hx}	SHX output temperature	T_{cm}	Compressor output temperature
T_{out}	ECS output temperature	T_{cn}	Condenser output temperature
T_{FD}	Flight deck zone temperature	T_{Zj}	Zone Z_j temperature
		$\forall j = 1 \dots n_z$	

S: Sensor SHX: Secondary heat exchanger
 PD: Primary bleed air duct SD: Secondary bleed air duct
 n_z : The number of zones in the cabin

Riverol and Napolitano [31] used *Artificial Neural Networks* (ANN) to estimate the heat exchanger fouling. Garcia [32] used *Neural Networks* and *rule based* techniques to improve heat exchanger monitoring. Adili et al. [33] used genetic algorithms to estimate the thermophysical properties of fouling.

Sensor selection has also gained recent attention by a diverse research community. Han et al. [34] studied feature selection problem for chillers. Namburu et al. [35] used genetic algorithm for sensor selection and applied SVM, PCA, and Partial Least Squares (PLS) for fault classification in HVAC systems. Optimal sensor selection for discrete-event systems with partial observations was performed by Jiang and Kumar et al. [36]. Gupta et al. [37] discussed stochastic sensor selection with application to scheduling and sensor coverage. Joshi and Boyd [38] used *convex optimization* to perform sensor selection. Hero and Cochran [39] provided a review of the methods and applications of sensors management. Xu et al. [40] used sensor configuration, usage and reliability costs for sensor selection for PHM of aircraft engines. Shen et al. [41] considered the problem of multistage look-ahead sensor selection for nonlinear dynamic systems.

III. SYSTEM DESCRIPTION

The Environmental Control System (ECS) is an air conditioning system that regulates temperature, pressure and humidity of the cabin air. In order to meet the health and comfort requirements of the passengers, the ECS supplies air to the cabin at moderate temperatures and pressures [1]. Figure 1 shows a simplified system diagram of the main ECS components, namely: i) *primary heat exchanger*, ii) *secondary heat exchanger*, iii) *air-cycle machine* (ACM), iv) *condenser*, and v) *water extractor*. The ACM in turn consists of a *compressor* and two *turbines*: a) *first stage turbine* and b) *second stage turbine*. The compressor and the turbines rotate on the same shaft [1], [42]–[44]. In addition, various sensing devices such as temperature and pressure sensors [1] are mounted at different locations of the ECS. Table I shows a list of critical sensors, as also shown in Fig. 1.

The primary heat exchanger is supplied with hot bleed air through two ducts, namely, the primary bleed air duct and the secondary bleed air duct, where air flow in each duct is controlled by a valve (not shown in Fig. 1). These ducts are then merged together to drive the bleed air to the primary

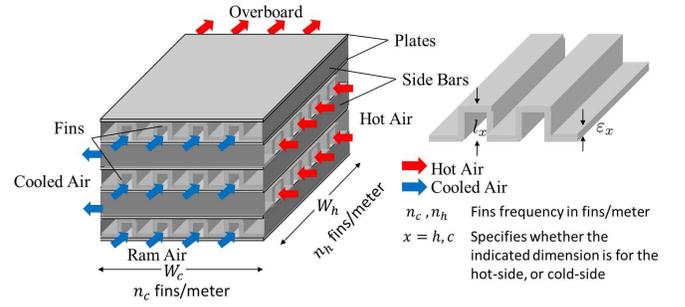


Fig. 2. An illustration of the plate fin heat exchanger.

heat exchanger. As shown in Fig. 1, hot bleed air is cooled in the primary heat exchanger, using ambient ram air as a sink, to a temperature below the auto-ignition temperature of fuel as a safety measure in case of a fuel leak. Air that comes out of the primary heat exchanger flows into the compressor section of the ACM where it gets compressed and thus heated. Air then flows out of the compressor into the secondary heat exchanger where it is cooled again using ram air as the sink. Air then flows through the hot side of the condenser heat exchanger where moisture is condensed out of the air-flow and collected by the water extractor. Air then flows into the first stage turbine where it gets expanded and cooled. Cold air out of the turbine flows through the cold side of the condenser heat exchanger into the second stage turbine where it gets further expanded and cooled providing the air at the desired cabin supply temperature and pressure [42].

A. Primary and Secondary Heat Exchangers

The heat exchangers used in the ECS under consideration are the *cross-flow plate fin heat exchangers* that are built from light weight plates and fins stacked over each other, as shown in Fig. 2. By definition, the direction through which the hot-air flows is called the *hot-side* while the direction through which the ram air flows is called the *cold-side* of the heat exchanger. The fins are placed alternatively in parallel to the hot air flow and the cold air flow, hence the name *cross-flow plate fin heat exchanger*. Plate fin heat exchangers are desirable for their compact sizes, high efficiency, and light weight.

The function of the heat exchanger is to transfer heat from the hot air to the ram air. The temperature can be set to the desired value by controlling the flow of the ram air in the cold-side of the heat exchanger. Debris accumulates on the fins of the heat exchangers due to several factors including chemical reactions, corrosion, biological multiplications and freezing. This phenomenon is known as *fouling* and it obstructs the ram air flow. Fouling lowers the heat efficiency of the heat exchanger because the deposited material has low thermal conductivity and hinders the transfer of heat [26], [33]. A detailed description of fouling substances and cleaning methods can be found in [26] and [45]. In this regard, this paper focuses on the fouling diagnosis of the secondary heat exchanger.

The heat transfer rate \dot{Q} (*Watts*) through the heat exchanger [46] is given by Eq. (1) as follows

$$\dot{Q} = \kappa \cdot A_h \cdot (T_{avg,h} - T_m) = \kappa \cdot A_c \cdot (T_m - T_{avg,c}) \quad (1)$$

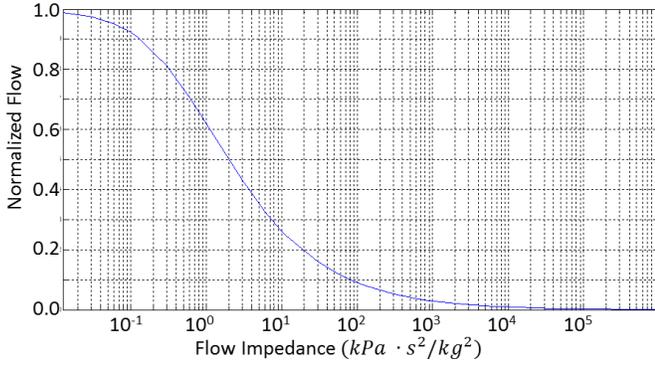


Fig. 3. Normalized flow vs secondary heat exchanger impedance.

where κ is the overall heat transfer coefficient ($W/(m^2K)$), T_m is the metal temperature (K), and $T_{avg,x}$ and A_x are the average air temperature (K) and the total heat transfer area (m^2) at the x -side, respectively. The subscript x is either h for the hot-side or c for the cold-side. The total heat transfer areas of the hot and cold sides are calculated as follows

$$A_x = W_h W_c N_x [1 + 2n_x(l_x - \varepsilon_x)] \quad (2)$$

where W_h , W_c , l_x , and ε_x are the fin dimensions (m) as shown in Fig. 2, and N_x and n_x are the number of fin layers and fin frequency per unit length at the x -side, respectively [47], [48].

The heat transfer is also calculated as a function of the input and the output temperatures of the heat exchanger, as follows

$$\dot{Q} = \dot{m}_h c_{p,h} (T_{i,h} - T_{o,h}) = \dot{m}_c c_{p,c} (T_{o,c} - T_{i,c}) \quad (3)$$

where \dot{m}_x , $c_{p,x}$, and $T_{i,x}$ and $T_{o,x}$ are the mass flow rate (kg/s), specific heat ($J/(kgK)$), and the input and output temperatures of the x -side, respectively [46].

The pressure drop between the input and the output pressures at the cold-side of the heat exchanger is modeled as

$$\begin{aligned} \Delta P &= P_{in,c} - P_{out,c} \\ &= \frac{1}{\beta} z_c \dot{m}_c^2 \end{aligned} \quad (4)$$

where $P_{in,c}$ and $P_{out,c}$ are the input and the output pressures (kPa) for the cold-side, β is a dimensionless correction factor and z_c is the flow impedance ($kPa \cdot s^2/kg^2$) which is varied in the simulation to represent different fouling conditions. The plot of the flow vs the flow impedance is shown in Fig. 3. A change in z_c affects \dot{m}_c and thus affects the heat transfer and the output temperatures of hot and cold air streams as computed using Eq. (1)-(3). This also affects the sensor readings of all other sensors in the ECS.

B. Data Generation Process

This paper utilizes an experimentally validated high-fidelity Simulink model of the ECS provided by an industry partner. The model is used to generate dynamic data for various sensor locations around the ECS system for fouling diagnosis. It is important to note that the model represents the ECS performance for a specific aircraft and has been validated to match experimental results from lab testing and flight data for

TABLE II
DAY TYPES

Day Type	Ambient Temperature ($^{\circ}F$)
Extremely cold	-30 – 0
Cold	0 – 30
Medium	30 – 60
Hot	60 – 90
Extremely hot	90 – 120

TABLE III
PASSENGER LOAD CATEGORIES

Load Type	Occupant Count (OCC)
Low Load	0% – 60%
Medium Load	60% – 75%
Heavy Load	75% – 95%
Very Heavy Load	95% – 100%

TABLE IV
DEFINITION OF FOULING CLASSES

Class	Flow	Flow Impedance ($kPa \cdot s^2/kg^2$)
Green (c_0)	80% – 100%	0.00 – 3.62e-01
Yellow (c_1)	60% – 80%	3.62e-01 – 1.09e+00
Orange (c_2)	40% – 60%	1.09e+00 – 3.62e+00
Red (c_3)	0% – 40%	3.62e+00 – 1.21e+04

this specific ECS. For this paper, the model is exercised to generate time series of sensor data for various ground operating conditions (e.g., ambient temperature, occupant count, etc.). Ground operating conditions are chosen because typically more debris exists in the aircraft vicinity while on the ground as opposed to in-flight operation. Data generated for this study includes a large number (> 100) of sensor outputs.

Figure 4 shows the stochastic time series data plots of three critical sensors under various uncertainties for different day types. The structure of the data is explained below. Let us denote the sensor suite by a set $\mathcal{S} = \{s_1, \dots, s_N\}$, where N is the total number of sensors. For each sensor s_i , the time series data are collected for 600 seconds at the sampling rate of 1 sample/sec, thus generating a data sequence $\mathbf{z}_i = [z_i(1), \dots, z_i(600)]$, $\forall i = 1, \dots, N$. The system reaches steady state after 300 seconds, thus the data from 301 to 600 seconds is used for analysis; however this interval could be reduced for higher sampling rates.

It is to be noted that the system behavior and the sensor data are affected by several input parameters, which affect the accuracy of fouling diagnosis. This paper considers variations in two main input parameters: the ambient temperature for different day types and the load corresponding to different occupant counts on the aircraft. Besides the heat exchanger fouling itself results in variations in sensor data. Thus the objective is to capture the effects of fouling under different input conditions. Specifically, the data is generated by varying the aforementioned parameters as described below.

- **Ambient Temperature (T_A):** As expected, the ambient temperature is the most critical external parameter that affects sensor readings. The effect of ambient temperature could be misinterpreted and could lead to false diagnosis of heat exchanger fouling. Thus, to incorporate the effect of ambient temperature, sensor data is categorized into five different day types: i) extremely cold, ii) cold, iii) medium, iv) hot and

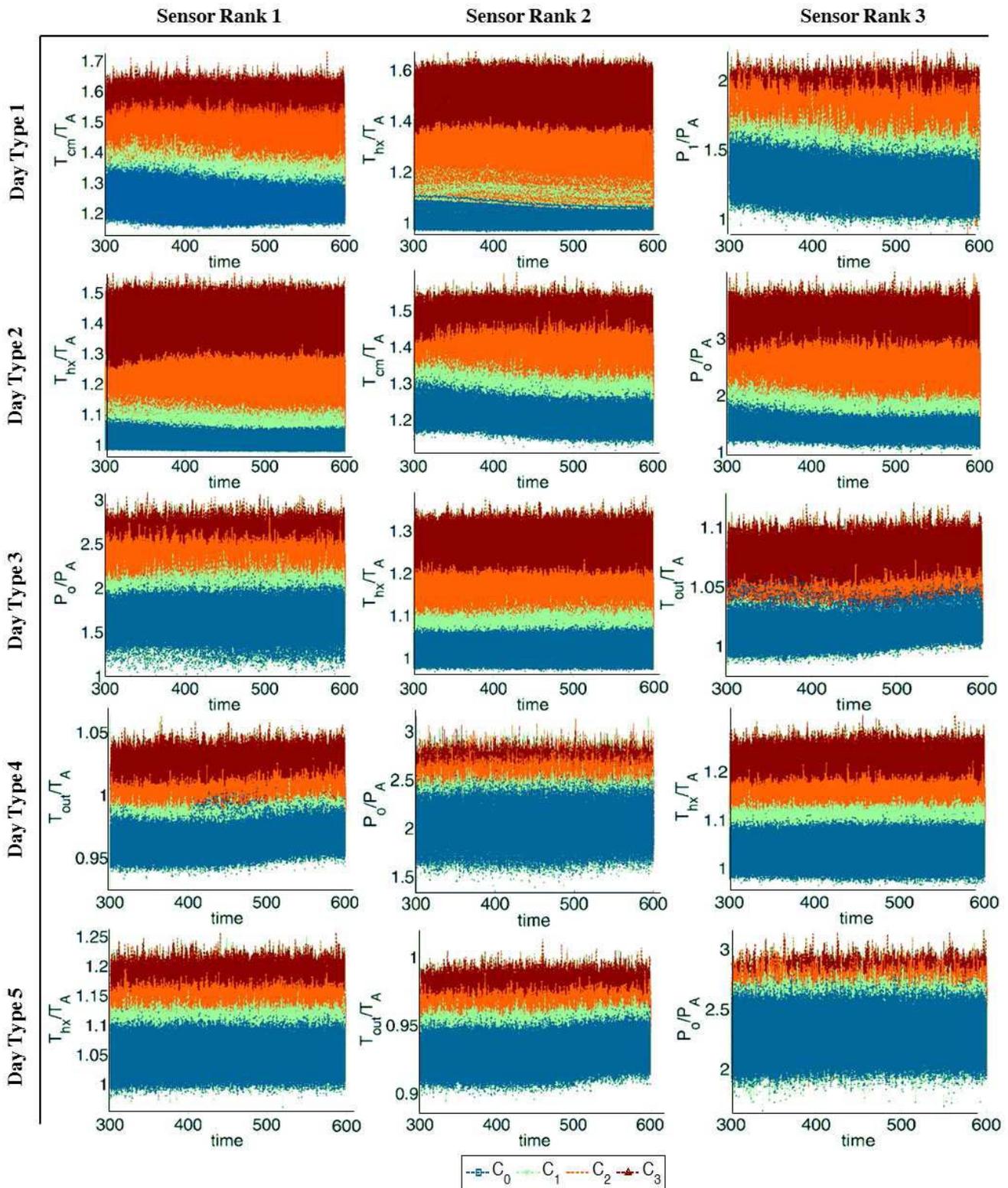


Fig. 4. Stochastic time series data of three critical sensors for five different day types.

v) extremely hot day types. The temperature ranges for each day type are shown in Table II. Each day type is further partitioned into eight uniformly spaced temperature values.

• *Occupant Count (OCC)*: The occupant count also affects ECS sensor readings due to passengers adding heat load that

the ECS reacts to in order to maintain the desired cabin conditions. The number of occupants is grouped into four categories: i) Low Load, ii) Medium Load, iii) Heavy Load, and iv) Very Heavy Load based on the percentage occupancy in the cabin. Table III shows these four load categories. Since

OCC has relatively less influence on the sensor readings, only the middle point of each category is used for data generation.

- *Heat Exchanger Fouling* (z_c): Fouling of the secondary heat exchanger is modeled as an increase in the ram air-flow impedance (z_c) on the cold side of the heat exchanger. When the flow impedance is increased, the air-flow decreases simulating blockage due to heat exchanger fouling. This lowers the effectiveness of the heat exchanger. For this paper four fouling classes have been defined based on the flow through the cold-side of the secondary heat exchanger as follows: i) *Green Class* (c_0)- i.e., 80-100 % flow, ii) *Yellow Class* (c_1)- i.e., 60-80 % flow, iii) *Orange Class* (c_2)- i.e., 40-60 % flow, and iv) *Red Class* (c_3)- i.e., 0-40 % flow. The reason to introduce Yellow and Orange classes is to avoid direct confusion between the Green and Red classes. The model is run for different values of flow impedance and the resulting flow through the heat exchanger is observed. The plot in Fig. 3 is used to determine the range of impedance values for each of the above classes that are defined based on the flow. Table IV shows the impedance intervals associated with each class. Each class is further partitioned into eight uniformly spaced flow values for data generation.

Thus, for each day type stochastic time series data are generated for various combinations of the above parameters to represent each fouling class. The model is run for different combinations of the values of ambient temperature (8) (within each day type), occupant counts (4), and impedance values (8) (within each fouling class), resulting in a set consisting of a total number of $8 \times 4 \times 8 = 256$ runs of time series data. Furthermore, for each day type, similar data sets are generated for all the fouling classes, thus leading to a total of $4 \times 256 = 1024$ runs of time series data. Subsequently, the above data sets are generated for all five day types. Let $\Gamma = \{\gamma_1, \dots, \gamma_{1024}\}$ denote the set of parametric combinations and let $t \in T = \{1, \dots, L = 600\}$ denote the set of discrete time indices. Then, for each day type the entire data for each sensor $s_i \in \mathcal{S}$ is arranged in a $|\Gamma| \times L$ matrix Z_{s_i} , where $Z_{s_i}(\gamma, t)$ denotes the sensor reading at time t for parametric combination γ . Thus, for any given $s = s_i$ and $\gamma = \gamma_j$, Z_{s_i} is a vector $\mathbf{z}_i(\gamma_j, \bullet) = [z_i(\gamma_j, 1), z_i(\gamma_j, 2), \dots, z_i(\gamma_j, L)]$, which is the time series data for sensor s_i for input condition $\gamma_j \in \Gamma$. Figure 4 shows the stochastic time series data plots of three critical sensors for each day type. For the purpose of data analysis, the fluctuations in OCC and the variations of impedance values within each class are considered as uncertainties. Other sources of uncertainties such as measurement noise, mechanical vibrations, and fluctuations in valve positions have been considered by adding white gaussian noise with 25 dB SNR to the data. The variations in ambient pressure have not been considered in this paper.

C. Heat Exchanger Fouling Diagnosis Architecture

Figure 5 shows the Heat exchanger fouling diagnosis architecture that consists of a training and a testing phase. The training phase consists of generating stochastic data for each sensor in the ECS (total 109 sensors) as described above. This sensor data is labeled with the fouling class information and is

used for optimal sensor selection for each day type separately, as described in Section IV. From the data of optimal sensors, some useful features are extracted using PCA and GMM methods and classifiers (k -NN) are trained to identify the fouling classes, as described in Section V.

In the testing phase, an unlabeled time series data is generated for an unknown parametric condition $\gamma \in \Gamma$ where the fouling severity is also considered as unknown. Subsequently, the optimal sensors identified in training phase are used for feature extraction and classification using trained classifiers. To further improve the classification accuracy, the results of the top three optimal sensors are fused using the majority vote.

IV. OPTIMAL SENSOR SELECTION METHODOLOGY

Since a large number of sensors are available in the ECS mounted at different locations, the underlying processes of data generation, storage, and analysis become cumbersome. Therefore, an optimal sensor selection methodology is needed to rank the most relevant sensors in terms of the best classification performance for heat exchanger fouling diagnosis. This is formally stated in the following problem statement.

Optimal Sensor Selection Problem: Given the sensor set $\mathcal{S} = \{s_1, \dots, s_N\}$, with N sensors, and the class set $\mathcal{C} = \{c_1 \dots c_M\}$, with M classes, the optimal sensor selection problem is to select a set $\mathcal{U}^* \subseteq \mathcal{S}$, where $|\mathcal{U}^*| = n$, $n < N$, that consists of sensors with maximum classification accuracy and are ranked accordingly in decreasing order.

As discussed in the introduction, two commonly used sensor selection methods are: i) the *wrapper method* and ii) the *filter method*. Since the wrapper algorithms rank the sensors based on their correct classification rate (CCR), a feature extractor and a classifier have to be designed, trained, and applied to all sensors in order to compute their CCRs, thus making the whole process computationally expensive. Furthermore, the wrapper algorithms cannot be generalized to any classifier [5]–[7]. On the other hand, the filter algorithms evaluate the performance of each sensor based on an information theoretic measure [8], [9]. Filter algorithms are computationally less expensive and do not depend on the choice of a classifier, but they may not perform as good as the wrapper algorithms [6].

To circumvent this difficulty, the *embedded algorithms* take advantage of both the wrapper and the filter algorithms by using a filter to select a candidate list of sensors and then applying a wrapper on this list to rank and select the optimal set of sensors [9]. Embedded algorithms are less expensive than wrappers and more accurate than filters; yet they are pertinent to the specified classifier [10]. In this regard, this section presents a detailed description of the optimal sensor selection methodology based on the *embedded algorithm*. In addition, a novel algorithm for sensor selection is presented, called the *unsupervised embedded algorithm*, that relies on the K -means clustering approach. This method has the advantage that it does not depend on the choice of a classifier and enables faster execution with very low computational complexity.

Both the embedded and the unsupervised embedded algorithms are based on the *minimum Redundancy Maximum Relevance* (mRMR) [9] criteria for the filter algorithm as a

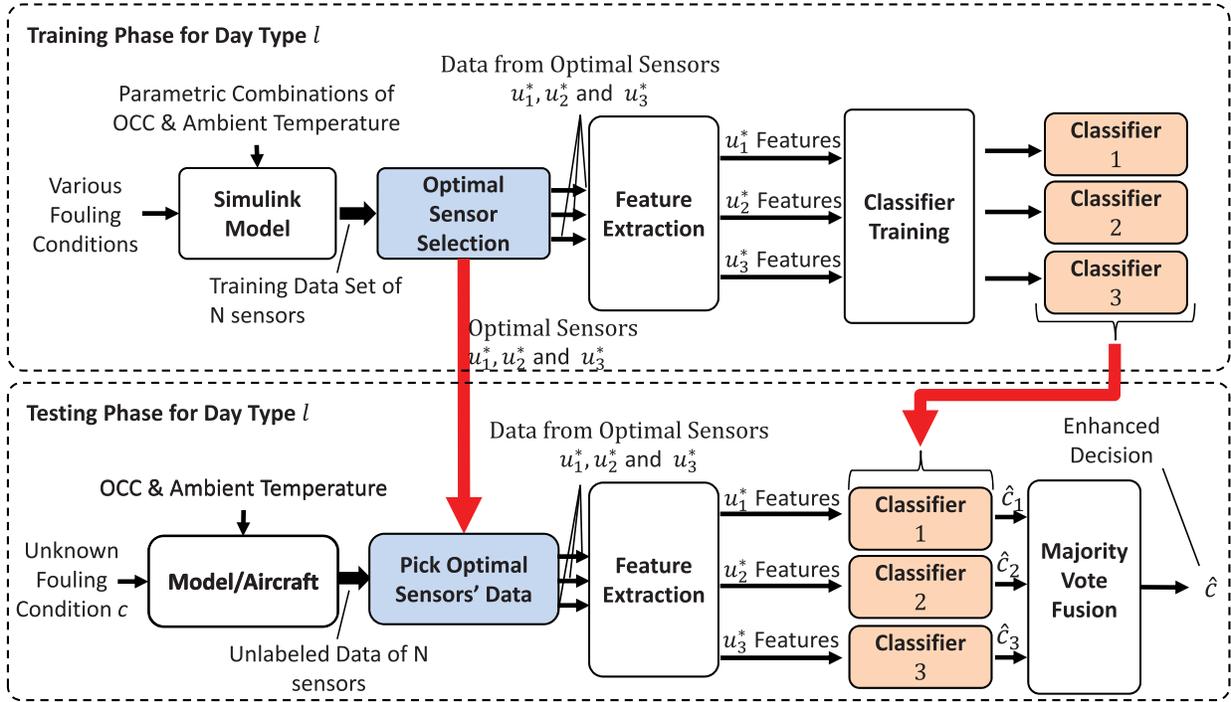


Fig. 5. Overall heat exchanger fouling diagnosis methodology with training and testing phases for each day type $\ell = 1 \dots 5$.

precursor step before applying the wrapper. The filter step facilitates fast execution of the first round of data reduction and produces a candidate list of top ranked sensors. Before describing the optimal sensor selection techniques, some useful information-theoretic quantities are defined below.

A. Information-Theoretic Measures

Definition 1 (Entropy): Entropy $H(X)$ is defined as a measure of uncertainty in a random variable X such that

$$H(X) = - \sum_{i=1}^r p^i \ln p^i \quad (5)$$

where X is a random variable whose outcomes belong to the set $\mathcal{X} = \{x^1, \dots, x^r\}$ with the associated probability distribution defined as $P(X = x^i) = p^i$ for $i = 1, \dots, r$.

According to Shannon [49], the entropy $H(X)$ qualifies to be a measure of uncertainty because it satisfies the following three conditions:

- $H(X)$ is a continuous function of p^i .
- If the random variable X is uniformly distributed (i.e., $p^i = \frac{1}{r}$, $\forall i = 1, \dots, r$), then $H(X)$ is a monotonically increasing function of r .
- If an event $X = x^i$ is split into two posterior sub-events, then the original entropy can be expressed as a weighted sum of the entropies of the sub-events.

The higher the entropy is, the higher is the uncertainty in the random variable. On the other hand, the entropy reaches its lowest value, $H(X) = 0$, when $P(X)$ is a delta distribution.

Suppose now that we have two random variables: X defined as above, and Y whose outcomes belong to the

set $Y = \{y^1, \dots, y^r\}$ with probabilities $P(Y = y^j) = q^j$ for all $j = 1, \dots, r$. Furthermore, suppose that the joint probability distribution is defined as $p^{i,j} = P(X = x^i, Y = y^j)$ for all $i, j = 1, \dots, r$. Then the joint and conditional entropies are defined as follows:

$$H(X, Y) = - \sum_{i=1}^r \sum_{j=1}^r p^{i,j} \ln p^{i,j} \quad (6)$$

$$\begin{aligned} H(X|Y) &= - \sum_{i=1}^r \sum_{j=1}^r p^{i,j} \ln p(X = x^i | Y = y^j) \\ &= H(X, Y) - H(Y). \end{aligned} \quad (7)$$

Definition 2 (Mutual Information): The mutual information between two random variables X and Y is defined as

$$\begin{aligned} I(X, Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y) \end{aligned} \quad (8)$$

The subtraction of $H(X|Y)$ from $H(X)$ represents the information gained about the random variable X given the information about the random variable Y [8]. The next section presents a partitioning approach for transformation of the continuous data to the symbolic domain for computation of the information-theoretic quantities as needed in the filter method.

B. Data Partitioning for Symbol Sequence Generation

Consider the data matrix Z_{s_i} of size $|\Gamma| \times L$ for any particular sensor s_i , $i = 1, \dots, N$, generated under different parametric conditions as described in Section III-B. The encoding of the underlying dynamics of this sensor data is achieved by partitioning [50] of the sensor observation space using an appropriate partitioning method. Let $\Omega_i \subset \mathbb{R}$ be the compact

(i.e., closed and bounded) region within which the observed sensor data Z_{s_i} is circumscribed. Let Σ be the *symbol alphabet* that labels the partition segments such that $|\Sigma| = r$, where $2 \leq r < \infty$. Then, the symbolic encoding of Ω_i is accomplished by introducing a partition $\{\varphi_i^1, \dots, \varphi_i^r\}$ consisting of $|\Sigma|$ mutually exclusive (i.e., $\varphi_i^j \cap \varphi_i^k = \emptyset, \forall j \neq k$), and exhaustive (i.e., $\bigcup_{j=1}^r \varphi_i^j = \Omega_i$) cells. Each cell is encoded with a symbol from the alphabet Σ . For each input condition, as the system evolves in time, the state trajectory (i.e., sensor readings) fall within different cells of the partition, accordingly the corresponding symbol is assigned to each point of the trajectory. Let $\mathbf{z}_i(\gamma, \bullet) \equiv [z_i(\gamma, 1), \dots, z_i(\gamma, L)]$ be a row of Z_{s_i} for a given $\gamma \in \Gamma$. Then, for each sensor $s_i \in \mathcal{S}$ and for each $\gamma \in \Gamma$, the time series data $\mathbf{z}_i(\gamma, \bullet)$ are transformed into a symbol sequence [51] $\sigma_i(\gamma, \bullet) \equiv [\sigma_i(\gamma, 1), \dots, \sigma_i(\gamma, L)]$ as

$$[z_i(\gamma, 1), \dots, z_i(\gamma, L)] \rightarrow [\sigma_i(\gamma, 1), \dots, \sigma_i(\gamma, L)] \quad (9)$$

where L is the data length and $\sigma_i(\gamma, t) \in \Sigma, \forall t = 1, \dots, L$. Note: As mentioned earlier, this paper makes use of only the steady state part of the data for fouling diagnosis analysis.

To do the above symbolization, this paper uses the *maximum entropy principle* [52] based partitioning to create a partition of the observed sensor data space, which is finer in the information dense regions and coarser in the low information regions as described below.

Definition 3 (Maximum Entropy Principle, Jaynes [52]): The maximum entropy principle states that the probability distribution that unbiasedly estimates the distribution of a random variable X under a given set of constraints is the distribution that maximizes the entropy $H(X)$.

Consider the following optimization problem:

$$\mathbf{P}^* = \begin{bmatrix} p^{1*} \\ \vdots \\ p^{r*} \end{bmatrix} = \arg \max_{\mathbf{P}} H(X); \quad H(X) = -\sum_{j=1}^r p^j \ln p^j$$

$$\text{subject to: } \sum_{j=1}^r p^j = 1 \quad (10)$$

The entropy is maximized for the uniform distribution (i.e., $p^{j*} = \frac{1}{r}, \forall j = 1, \dots, r$). For proof please see Appendix A.

The Maximum Entropy Principle generates the unbiased distribution for each of the sensor readings. Considering the data matrix $Z_{s_i} = [z_i(\gamma, 1), \dots, z_i(\gamma, L)]_{\gamma=\gamma_1, \dots, \gamma_{|\Gamma|}}$ for sensor i , the goal is to find the partition that results in maximum entropy distribution (i.e., the uniform distribution). The partition cells are defined by the partitioning levels $\{L_i^0 \dots L_i^r\}$, such that $\varphi_i^j = [L_i^{j-1}, L_i^j] \forall j = 1 \dots r$. To compute the maximum entropy partition for the sensor data Z_{s_i} , the first step is to calculate the number of samples in each cell (i.e., $\eta^* = \eta^{j*} = \text{floor}(|\Gamma| \times L / r), \forall j = 1, \dots, r$). The second step is to sort the entire data into a vector $y_i = [y_i(1), \dots, y_i(|\Gamma| \times L)]$, such that

$$y_i(\ell) \in Z_{s_i} \quad \forall \ell = 1, \dots, |\Gamma| \times L \quad (11)$$

$$\& y_i(1) \leq y_i(2) \dots \leq y_i(|\Gamma| \times L) \quad (12)$$

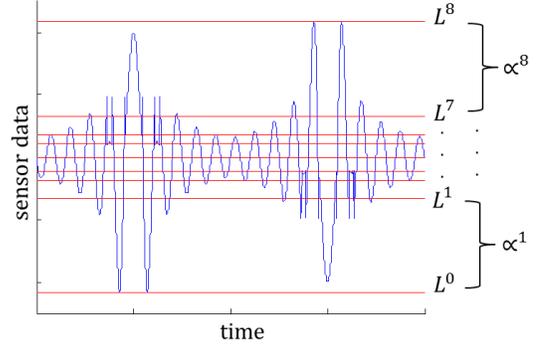


Fig. 6. An illustration of the maximum entropy partitioning.

Then the partitioning levels are defined as follows:

$$L_i^0 = y_i(1) \quad (13)$$

$$L_i^h = y_i(h \cdot \eta^*) \quad \forall h = 1, \dots, r-1, \quad \text{and} \quad (14)$$

$$L_i^r = y_i(|\Gamma| \times L) \quad (15)$$

The algorithm counts the samples from the bottom and defines the partitioning levels at the multiples of η^* while setting the first and the last levels at the min and the max of the original data. This procedure generates a partition that is finer in the regions of high data density and coarser in the regions of low data density, as shown by an illustrative example in Fig. 6. Subsequently, a unique symbol from the alphabet Σ is assigned to all the data points in each cell of the partitioning. This process transforms each data sequence in Z_{s_i} into a symbol sequence, as shown in Eq. (9).

In the above manner, the maximum entropy partitioning is constructed for all sensors and the corresponding data are transformed into symbol sequences. Subsequently, the candidate list of sensors is selected and ranked according to the filter criteria as described next.

C. Minimum Redundancy Maximum Relevance (mRMR)

Based on mutual information, the mRMR criterion [9] evaluates and ranks the sensors that best describe the classes and simultaneously avoid sensors that provide redundant information by means of the following two conditions: i) *Maximum Relevance* and ii) *Minimum Redundancy*, as described below.

Let us define the random variables C and $S_j, j = 1, \dots, N$, as follows:

- C : A random variable whose sample space is the set of all symbol sequences and its outcome belongs to the class set $\mathcal{C} = \{c_1, \dots, c_M\}$, and
- S_j : A random variable whose sample space is the symbolized data matrix σ_{s_j} for sensor s_j and its outcome belongs to Σ .

Then the Maximum Relevance criteria is defined as follows.

Definition 4 (Max Relevance): The Maximum Relevance criterion aims to find the set $\mathcal{U}_1^* \subseteq \mathcal{S}$, where $|\mathcal{U}_1^*| = n, n < N$, that has the maximum average mutual information between its

Algorithm 1 *The Forward Selection Search Algorithm*

Result: An optimal set of sensors $\mathcal{U}^* = \{u_1, \dots, u_n\}$.
Initialization: $\mathcal{S} = \{s_1 \dots s_N\}$, $\mathcal{U}^* = \emptyset$, $j = 1$
while $j \leq n$ **do**
 • Step 1: Find the sensor $u_j \in \mathcal{S}$ that maximizes the criterion in Eq. (18) for a single sensor
 • Step 2: Update $\mathcal{S} \rightarrow \mathcal{S} - u_j$
 • Step 3: Update $\mathcal{U}^* \rightarrow \mathcal{U}^* \cup u_j$, $j \rightarrow j + 1$
end

sensors and the random variable C , such that

$$\mathcal{U}_1^* = \arg \max_{\mathcal{U}_1 \subseteq \mathcal{S}, |\mathcal{U}_1|=n} \Upsilon(\mathcal{U}_1, C);$$

$$\Upsilon(\mathcal{U}_1, C) = \frac{1}{n} \sum_{s_j \in \mathcal{U}_1} I(\mathcal{S}_j, C) \quad (16)$$

The Maximum-Relevance criterion does not account for the information redundancy between sensors. Thus, the Minimum Redundancy criteria is defined as follows.

Definition 5 (Min Redundancy): The Minimum Redundancy criterion aims to find the set $\mathcal{U}_2^* \subseteq \mathcal{S}$, where $|\mathcal{U}_2^*| = n$, $n < N$, that has the minimum average mutual information between its sensor pairs, such that

$$\mathcal{U}_2^* = \arg \min_{\mathcal{U}_2 \subseteq \mathcal{S}, |\mathcal{U}_2|=n} \Psi(\mathcal{U}_2),$$

$$\Psi(\mathcal{U}_2) = \frac{1}{n^2} \sum_{s_i, s_j \in \mathcal{U}_2} I(\mathcal{S}_i, \mathcal{S}_j) \quad (17)$$

The minimum Redundancy Maximum Relevance (mRMR) criterion combines the above two criterion as follows.

Definition 6 (Minimum Redundancy Maximum Relevance): The minimum Redundancy Maximum Relevance (mRMR) criterion aims to find the set $\mathcal{U}^* \subseteq \mathcal{S}$, where $|\mathcal{U}^*| = n$, $n < N$, to optimize Υ and Ψ simultaneously, such that

$$\mathcal{U}^* = \arg \max_{\mathcal{U} \subseteq \mathcal{S}, |\mathcal{U}|=n} \Phi(\mathcal{U}, C),$$

$$\Phi(\mathcal{U}, C) = \Upsilon(\mathcal{U}, C) - \Psi(\mathcal{U}) \quad (18)$$

The evaluation of the mRMR criteria requires: a) computation of $I(\mathcal{S}_j, C)$ and $I(\mathcal{S}_i, \mathcal{S}_j)$, $\forall i, j = 1, \dots, N$, and b) finding the solution of the optimization function in Eq. (18). The mutual information quantities are computed from the symbol sequences of each sensor data, as described in Appendix B. The optimization problem based on the mRMR criterion is a combinatorial problem, which can be solved using the *Forward Selection* search method [9]. Note: this information-theoretic method of sensor selection is more efficient and several orders of magnitude faster as compared to the full wrapper method that requires computation of the CCRs for all sensors.

Forward Selection Search: The forward selection search is a greedy search algorithm that is used to find a (sub)optimal solution of the mRMR optimization problem in Eq. (18). To be specific, the algorithm starts with an empty set of sensors, then keep adding sensors that maximize the mRMR criteria until the desired number of n sensors is obtained. The details [53] of the algorithm are shown in Algorithm 1.

D. Embedded Algorithm

As mentioned earlier, an *embedded wrapper and filter* algorithm is used to tradeoff between the low complexity of filter algorithms and the accuracy of wrapper algorithms in the optimal sensor set selection procedure. In other words, an embedded algorithm uses a filter algorithm first to select a *candidate list* (CL) of n sensors; subsequently, a wrapper algorithm (which uses a specific classifier) is deployed to select or rank the *optimal set of sensors* [9] from the candidate list. The embedded algorithms also have several deficiencies including being specific to a certain classifier and requiring tuning the classifier beforehand for each sensor separately. To circumvent these disadvantages, the paper proposes the unsupervised embedded algorithm as described next.

E. Unsupervised Embedded Algorithm

The unsupervised embedded algorithm also relies on a filter algorithm (e.g., the mRMR) to select the candidate list (CL) of n sensors. Then the data Z_{s_i} corresponding to each sensor $s_i \in \text{CL}$, which consists of the data of all classes, are clustered into M clusters using the K -means clustering algorithm [54], where M is equal to the number of fouling classes (for this paper $M = 4$). Lets call these clusters as $\{\mathcal{O}_1, \dots, \mathcal{O}_M\}$. Lets now define a random variable Ξ_j that is drawn on the cluster \mathcal{O}_j and whose outcome belongs to the set of classes $\mathcal{C} = \{c_1, \dots, c_M\}$. Subsequently, the entropy $H(\Xi_j)$, $j = 1, \dots, M$, of the class distribution within each cluster is computed using Eq. (5). Then, the weighted entropy for all clusters for a sensor s_i is calculated as

$$H_{s_i} = \sum_{j=1}^M \frac{|\mathcal{O}_j|}{\sum_{j'=1}^M |\mathcal{O}_{j'}|} \cdot H(\Xi_j), \quad \forall s_i \in \text{CL}. \quad (19)$$

Finally, the sensors are ranked according to their entropies such that the sensor that has the lowest entropy is ranked the highest and so on. In this fashion the candidate list is re-ranked and a possible list of top ranked sensors is selected for further analysis. This process ranks the sensors in the order such that the sensors that have the least uncertainty between classes in their data clusters are ranked the highest, thus facilitating a better classification decision.

V. DATA ANALYSIS FOR FOULING DIAGNOSIS

Once an optimal sensor set is obtained, different machine learning methods are applied for analysis of sensor data for fouling diagnosis. These methods consists of the feature extraction and the classification steps as described below.

A. Feature Extraction

This paper explores two methods of feature extraction from sensor data for heat exchanger fouling diagnosis; namely, the *Principal Component Analysis* (PCA) and the *Gaussian Mixture Model* (GMM) as described below.

1) *Principal Component Analysis (PCA)*: The Principal Component Analysis (PCA) is a data reduction method. Consider a data matrix \mathbf{X} of dimension $d \times m$, where $d > m > 0$, whose columns are data vectors (e.g., sensor data). The objective of PCA is to transform the data matrix \mathbf{X} into a matrix \mathbf{Y} of size $d \times m'$, where $m' < m$. The columns of \mathbf{Y} hold the *Score Vectors* (also known as the *Principal Components*). This transformation is accomplished using the *Karhunen-Loève* (KL) algorithm as summarized here. First, the $m \times m$ covariance matrix \mathbf{C}_X of \mathbf{X} is computed and the corresponding eigenvalues are obtained and sorted in descending order. Second, the eigenvectors associated with the m' largest eigenvalues are generated and arranged into an $m \times m'$ transformation matrix \mathbf{T} . Finally, the scores of \mathbf{X} are computed using the following linear transformation

$$\mathbf{Y} = \mathbf{X} \times \mathbf{T} \quad (20)$$

For implementation to the heat exchanger data, consider the sensor data Z_{s_i} for a specific sensor s_i for a specific day type. The nominal data corresponding to class c_0 is extracted from Z_{s_i} and averaged to get a time series data $\bar{\mathbf{z}}_i^0$ as follows:

$$\bar{\mathbf{z}}_i^0 = \frac{1}{256} \sum_{\gamma \in \{\gamma_1 \dots \gamma_{256}\}} \mathbf{z}_i(\gamma, \bullet) \quad (21)$$

Thereafter, the steady state part of $\bar{\mathbf{z}}_i^0$ is partitioned into $m = 10$ segments, each of length $d = 30$. These data segments are organized to form a $d \times m$ data matrix X^0 . Then, following the steps of the *KL* algorithm above, the $m \times m'$ transformation matrix \mathbf{T} is obtained, where $m' = 2$ and is kept fixed. Subsequently, the scores (or principal components) of any observation sequence \mathbf{z}_i are generated by reorganizing the sequence into a $d \times m$ matrix \mathbf{X} by breaking \mathbf{z}_i into m segments of length d each. The scores of this sequence are then computed using Eq. (20). These scores consist of m points each of which are $m' = 2$ dimensional. The scores for all sequences in Z_{s_i} are plotted on the m' -dimensional feature space. In the training phase these scores are labeled with fouling class and are sent to the classifier for training, while in the testing phase they are unlabeled and are sent to the trained classifier for decision on the fouling class.

2) *Gaussian Mixture Model (GMM)*: Consider sensor data $\mathbf{z}_i(\gamma, \bullet) = [z_i(\gamma, 1), \dots, z_i(\gamma, L)]$. The Gaussian Mixture Model (GMM) is a statistical model of $\mathbf{z}_i(\gamma, \bullet)$ represented as a sum of R different Gaussian distributions as

$$P(\mathbf{z}_i(\gamma, \bullet) | \mathcal{M}) = \sum_{j=1}^R w^j \cdot \mathcal{Q}(z; \mu^j, \rho^j) \quad (22)$$

where $\mathcal{M} = \{w^1, \dots, w^R, \mu^1, \dots, \mu^R, \rho^1, \dots, \rho^R\}$ is the set of weights w^j s, means μ^j s and variances ρ^j s, $j = 1, \dots, R$, and z is a random variable. The function $\mathcal{Q}(z, \mu^j, \rho^j)$ is a Gaussian distribution given as

$$\mathcal{Q}(z; \mu^j, \rho^j) = \frac{1}{(2\pi)^{1/2} \sqrt{\rho^j}} \times e^{-\frac{1}{2}(z-\mu^j)/(\rho^j)^{-1}(z-\mu^j)} \quad (23)$$

The parameters w^j, μ^j and $\rho^j, \forall j = 1, \dots, R$, are estimated from the data using the *Expectation Maximization* (EM)

algorithm [55]. Subsequently, \mathcal{M} is used as a feature for the classifier. For implementation to the heat exchanger data, a GMM is constructed from sensor data with $R = 2$. For each observation sequence $\mathbf{z}_i(\gamma, \bullet)$ in Z_{s_i} , the feature set $\mathcal{M}_i(\gamma)$ is computed as

$$\mathcal{M}_i(\gamma) = [w_i^1, w_i^2, \mu_i^1, \mu_i^2, \rho_i^1, \rho_i^2](\gamma). \quad (24)$$

B. Classification

Once the features are obtained as the principal components or the parameter set of the GMM, they are processed by a classifier to make a decision on the heat exchanger fouling severity. This section describes the *k-Nearest Neighbor* (*k-NN*) algorithm that is used as the classification technique.

k-Nearest Neighbor (k-NN): The *k-Nearest Neighbor* (*k-NN*) classification algorithm is popular for its simplicity, efficiency and low complexity. First, an odd value of k is chosen, that represents the number of nearest neighbors on the feature space. In the training phase, the optimal k is selected that results in the highest *Correct Classification Rate* (CCR) using the cross-validation algorithm for different values of $k = 1, 3, 5, \dots, 21$. In the testing phase, a new feature point is classified using the majority rule among the k -nearest neighbors as obtained from the training data. Thus the predicted class $\hat{c} \in \mathcal{C}$ is obtained as follows:

$$\hat{c} = \arg \max_{i=1 \dots M} \frac{n_i}{k} \quad (25)$$

where n_i is the number of feature points corresponding to the class $c_i \in \mathcal{C}$ among the k nearest neighbors of the testing point [2], [25]. The performance of the classifier is evaluated using the random subsampling hold-out cross-validation method. The results are summarized into a *confusion matrix* [56], whose columns contain the predicted classes while the rows contain the actual classes.

For implementation, the *k-NN* classifier is applied on the feature space generated by PCA and GMM for each sensor in the candidate list and each day type. For each observation sequence $\mathbf{z}_i(\gamma, \bullet)$, the PCA based features (i.e., principal components) are $d = 30$ dimensional vectors in $m' = 2$ dimensional feature space while the GMM based features (i.e. the parameter set of GMM) are 1 dimensional in the 6 dimensional feature space. Since the PCA based features are vectors, the *k-NN* classifier produces d decisions one for each point in the vector. Then, a single decision is obtained from these decisions using the simple majority rule. For the cross-validation method, 30 data sequences are hold-out from each class and are used for testing while the remaining are used for training. This process is repeated 50 times, where in each run 30 data sequences are randomly selected from each class. This generates a total of 1500 testing samples for each class.

C. Sensor Fusion

Sensor fusion is performed for further improvement in the classification performance. Suppose for a given day type the top three optimal sensors are $\{u_1^*, u_2^*, u_3^*\}$, where $u_i^* \in \mathcal{U}^* \subseteq \mathcal{S}, \forall i = 1, 2, 3$. Moreover, lets say that the classifier has generated the following three decisions $\{\hat{c}_1, \hat{c}_2, \hat{c}_3\}$,

TABLE V
DECREASINGLY SORTED CANDIDATE LIST AND OPTIMAL SENSOR SETS USING mRMR, UNSUPERVISED EMBEDDED, EMBEDDED mRMR+PCA+k-NN AND EMBEDDED mRMR+GMM+k-NN ALGORITHMS

Rank	Day Type 1				Day Type 2				Day Type 3				Day Type 4				Day Type 5			
	CL	UE	EP	EG	CL	UE	EP	EG	CL	UE	EP	EG	CL	UE	EP	EG	CL	UE	EP	EG
1	T_{cm}	T_{cm}	T_{hx}	P_o	T_{hx}	T_{hx}	T_{hx}	T_{hx}	T_{out}	T_{hx}	T_{hx}	T_{out}	T_{out}							
2	T_{out}	T_{hx}	T_{cm}	T_{cn}	T_{out}	T_{cm}	T_{cm}	P_o	T_{cn}	T_{hx}	T_{cm}	T_{cm}	T_{cn}	P_o	T_{out}	T_{hx}	T_{hx}	T_{hx}	T_{hx}	
3	T_{cn}	P_1	T_{cn}	P_2	P_2	P_o	P_o	P_2	T_{out}	T_{out}	P_o	P_o	T_{out}	T_{hx}	P_o	P_o	P_o	P_o	P_o	
4	\hat{m}_1	P_2	P_o	P_1	T_{cn}	P_1	P_2	P_1	P_2	T_{cm}	T_{out}	T_{out}	P_o	T_{cn}	T_{cm}	T_{cm}	T_{out}	P_2	T_{cm}	
5	P_2	P_o	P_1	T_{cm}	T_{cm}	P_2	T_{cn}	T_{cm}	\hat{m}_2	P_1	P_2	P_2	T_{cm}	T_{cm}	P_2	P_2	P_2	P_1	P_2	
6	T_{hx}	T_{cn}	P_2	P_o	\hat{m}_1	T_{out}	P_1	T_{cn}	P_1	P_2	P_1	P_1	T_{FD}	P_1	P_1	P_1	P_1	T_{cm}	P_1	
7	\hat{m}_2	T_{out}	\hat{m}_1	T_{out}	P_1	T_{cn}	T_{out}	\hat{m}_1	P_o	T_2	\hat{m}_2	T_2	P_1	P_2	T_{cn}	T_{cn}	T_{FD}	T_{cn}	T_{FD}	
8	P_1	\hat{m}_1	\hat{m}_2	\hat{m}_2	P_o	\hat{m}_1	\hat{m}_2	\hat{m}_2	T_2	T_1	T_2	T_1	P_2	T_{FD}	T_{FD}	T_{FD}	T_{cn}	T_{FD}	T_{FD}	
9	P_o	\hat{m}_2	T_{out}	\hat{m}_1	\hat{m}_2	\hat{m}_2	\hat{m}_1	T_{out}	T_{cm}	T_{cn}	T_1	\hat{m}_2	T_{Z1}	T_{Z1}	T_{Z1}	T_{Z1}	T_{Z2}	T_{Z2}	T_{Z1}	
10	T_{Z5}	T_1	\hat{m}_2	T_{cn}	T_{cn}	T_{Z5}	T_{Z5}	T_{Z5}	T_{Z5}	T_{Z1}	T_{Z1}	T_{Z2}								

CL Candidate List UE Unsupervised Embedded EP Embedded with PCA and k -NN EG Embedded with GMM and k -NN

TABLE VI
CONFUSION MATRICES FOR THE OPTIMAL SENSORS WHEN USING PCA FOR FEATURE EXTRACTION AND k -NN FOR CLASSIFICATION

Sensor	Day Type 1				Sensor	Day Type 2				Sensor	Day Type 3				Sensor	Day Type 4				Sensor	Day Type 5			
	Predicted					Predicted					Predicted					Predicted					Predicted			
	c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3
T_{cm}	1378	122	0	0	T_{hx}	1352	148	0	0	P_o	1223	277	0	0	T_{out}	1269	231	0	0	T_{hx}	1413	87	0	0
$k = 21$	129	1261	110	0	$k = 13$	146	1286	68	0	$k = 9$	287	1119	94	0	$k = 19$	218	1271	11	0	$k = 1$	84	1414	2	0
	6	145	1303	46		0	118	1330	52		0	99	1374	27		0	1	1499	0		0	11	1489	0
	0	0	60	1440		0	0	47	1453		0	0	9	1491		0	0	0	1500		0	0	18	1482
T_{hx}	1354	146	0	0	T_{cm}	1325	175	0	0	T_{hx}	1326	174	0	0	P_o	1332	168	0	0	T_{out}	1433	67	0	0
$k = 7$	105	1295	100	0	$k = 15$	213	1196	91	0	$k = 17$	162	1248	90	0	$k = 19$	205	1223	72	0	$k = 7$	94	1392	14	0
	0	108	1373	19		0	131	1278	91		0	89	1411	0		0	102	1356	42		0	17	1483	0
	0	0	41	1459		0	4	80	1416		0	0	1	1499		0	0	46	1454		0	0	0	1500
P_1	1284	216	0	0	P_o	1314	186	0	0	T_{out}	1065	412	7	16	T_{hx}	1359	141	0	0	P_o	1396	104	0	0
$k = 15$	311	1112	77	0	$k = 9$	401	1015	84	0	$k = 9$	400	1023	77	0	$k = 15$	119	1360	21	0	$k = 19$	106	1330	64	0
	0	151	1300	49		0	112	1357	31		25	107	1368	0		0	41	1459	0		0	73	1351	76
	0	0	50	1450		0	0	33	1467		6	0	0	1494		0	0	2	1498		0	0	73	1427
MVF	1439	61	0	0	MVF	1435	65	0	0	MVF	1349	151	0	0	MVF	1419	81	0	0	MVF	1480	20	0	0
	130	1332	38	0		173	1305	22	0		229	1259	12	0		97	1399	4	0		40	1460	0	0
	2	65	1419	14		0	79	1414	7		1	34	1465	0		0	9	1491	0		0	4	1496	0
	0	0	12	1488		0	0	5	1495		0	0	0	1500		0	0	0	1500		0	0	2	1498

k Number of nearest neighbors for the k -NN algorithm MVF Majority Vote Fusion

corresponding to the above three sensors, where $\hat{c}_i \in \mathcal{C}$, $\forall i = 1, 2, 3$. Then, the fusion decision \hat{c}^F is obtained as follows:

$$\hat{c}^F = \arg \max_{c_i \in \mathcal{C}} P(C^d = c_i) \quad (26)$$

where C^d is a random variable drawn on the set $\{\hat{c}_1, \hat{c}_2, \hat{c}_3\}$.

VI. RESULTS AND DISCUSSION

This section presents the results for optimal sensor selection and the classification for Heat Exchanger fouling diagnosis. As shown in the methodology in Fig. 5, the first step is training of the classifiers. Therefore, sensor data is generated using a Simulink model that has been experimentally validated by our industry partner. For each sensor $s_i \in \mathcal{S}$ the data is stored in a $|\Gamma| \times L$ matrix Z_{s_i} which includes the data for all four fouling classes as described in Section III-B. The rows of the matrix consist of various parametric combinations of the ambient temperature (T_A) within a specific day type, the occupant count (OCC), and the impedance values (z_c) within each class, as shown in Tables II-IV. As described in Section III-B, for each class there are 256 data sequences each of length $L = 600$, thus resulting in a total of $4 \times 256 = 1024$ sequences for each day type. Similar data are generated for each day type. To include the effect of other uncertainties beyond the above parametric uncertainties, the data is corrupted by 25dB Additive White Gaussian Noise (AWGN),

and the noisy data are plotted in Fig. 4. The noise is filtered out using 5-levels wavelet denoising technique [57] using 6-taps Daubechies wavelet with *soft-thresholding* on the detail coefficients of the wavelet. The thresholds are determined using *Stein Unbiased Risk Estimator* (SURE). The paper uses only the last 300 seconds of steady state data for analysis.

Subsequently, the data of all sensors for one day type are taken and the mRMR technique is applied to find the candidate list of top 10 sensors. To compute the information theoretic quantities in the mRMR criterion, the sensor data is transformed into symbol sequences using the Maximum Entropy Partitioning as described in Section IV-B. The above process is repeated for each day type. The resultant candidate lists (CL) are shown in Table V.

Once the candidate list is generated for each day type using the mRMR criteria, the unsupervised embedded algorithm is applied on the candidate list, which includes a two-step process: i) K -means clustering of each sensor data, ii) computation of weighted average cluster entropies and sensor ranking as described in Section IV-E. The resultant sensor ranking is presented in Table V. For comparison, the embedded algorithm is used to rank the sensors which used two different wrappers: i) PCA as the feature extractor and k -NN as the classifier, and ii) GMM as the feature extractor and k -NN as the classifier. As seen in Table V, the results of unsupervised embedded algorithm improved the candidate list rankings which were purely based on information-theoretic quantities, such that

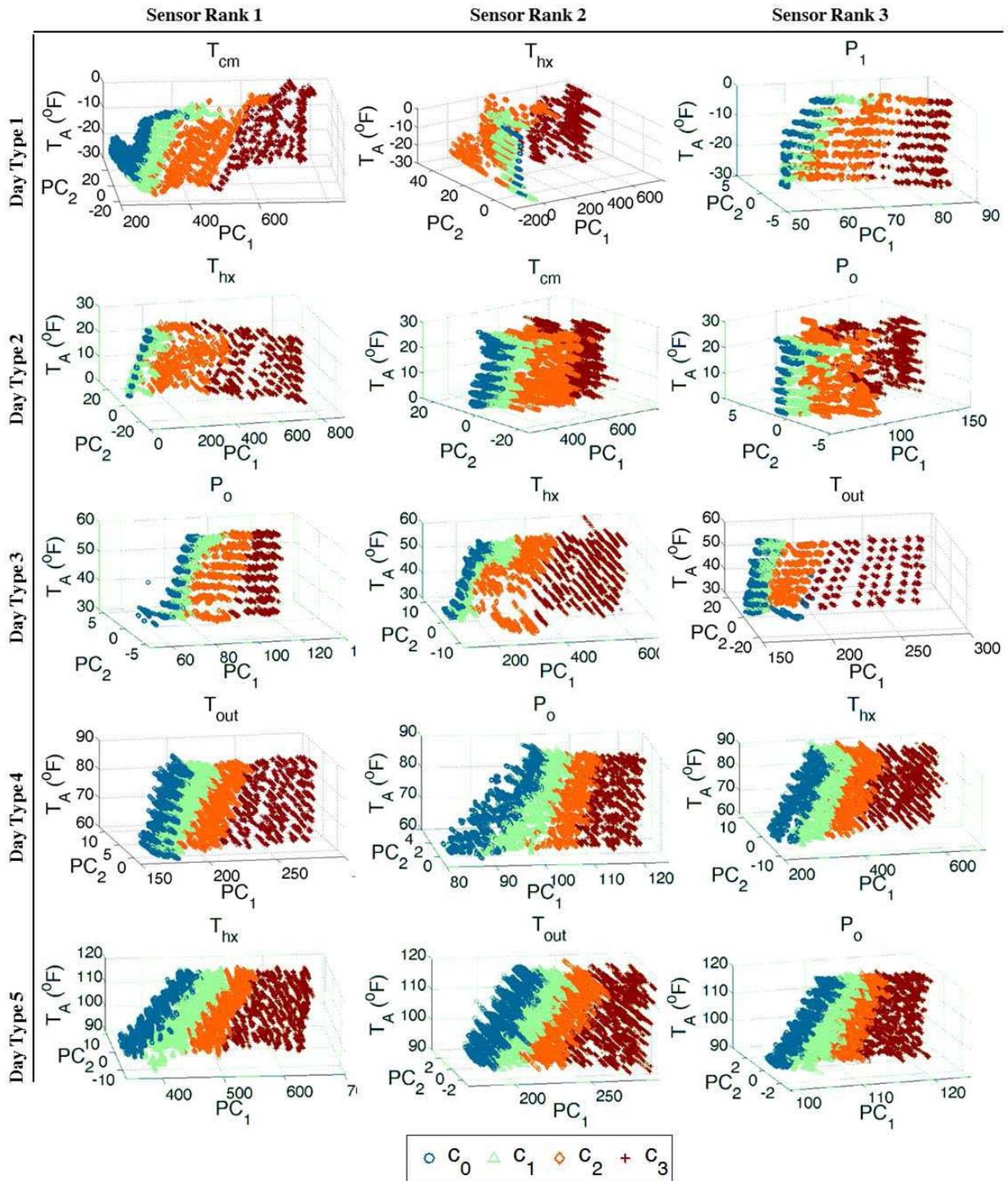


Fig. 7. Top three unsupervised embedded optimal sensors' principal components 1 and 2 vs the ambient temperature for day types 1, 2...5.

the updated rankings are similar to the embedded wrapper results. However, the unsupervised embedded algorithm does not depend on the feature extraction and classification methods and is computationally much more efficient.

Using the PCA procedure for feature extraction as described in Section V-A1, a feature vector consisting of 30 feature points are extracted from each data sequence in the data matrix Z_{s_i} . Each feature point is composed of the elements of the first and second principal components. These principal components are computed from the eigenvectors whose

corresponding eigenvalues contain more than 90% of the energy. Since ambient temperature can be measured and known, it is used to augment the feature space as the third axis where the other two axis are formed by the two principal components. As a result, we have 30 feature points for each data sequence plotted in a 3 – D feature space. This resulted in excellent clustering of the classes as shown in Fig. 7.

Using the GMM procedure for feature extraction as described in Section V-A2, the second order GMM was used to extract a feature vector of size 1×6 from each data sequence

TABLE VII

CONFUSION MATRICES FOR THE OPTIMAL SENSORS WHEN USING GMM FOR FEATURE EXTRACTION AND k -NN FOR CLASSIFICATION

Sensor	Day Type 1 Predicted				Sensor	Day Type 2 Predicted				Sensor	Day Type 3 Predicted				Sensor	Day Type 4 Predicted				Sensor	Day Type 5 Predicted				Actual
	c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3		c_0	c_1	c_2	c_3	
T_{cm} $k=1$	1289	210	1	0	T_{hx} $k=9$	1371	129	0	0	P_o $k=13$	1317	183	0	0	T_{out} $k=3$	1344	127	29	0	T_{hx} $k=1$	1412	88	0	0	c_0
	194	1195	111	0		158	1286	56	0		339	1122	39	0		141	1349	10	0		93	1392	15	0	c_1
	11	131	1322	36		0	121	1352	27		0	97	1387	16		0	40	1454	6		0	31	1464	5	c_2
	0	0	89	1411	0	0	64	1436	0	0	35	1465	0	0	11	1489	0	0	36	1464	c_3				
T_{hx} $k=5$	1385	115	0	0	T_{cm} $k=3$	1243	257	0	0	T_{hx} $k=5$	1366	134	0	0	P_o $k=5$	1398	102	0	0	T_{out} $k=3$	1411	89	0	0	c_0
	113	1334	53	0		254	1140	106	0		118	1363	19	0		203	1249	48	0		82	1402	16	0	c_1
	0	129	1326	45		1	126	1279	94		0	89	1411	0		0	99	1366	35		0	50	1443	7	c_2
	0	0	81	1419	0	0	100	1400	0	0	5	1495	0	0	59	1441	0	0	7	1493	c_3				
P_1 $k=9$	1303	197	0	0	P_o $k=15$	1354	146	0	0	T_{out} $k=5$	962	493	21	24	T_{hx} $k=11$	1447	53	0	0	P_o $k=13$	1413	87	0	0	c_0
	288	1205	7	0		385	1080	35	0		437	1028	35	0		158	1335	7	0		122	1342	36	0	c_1
	0	170	1303	27		0	103	1353	44		14	141	1345	0		0	98	1396	6		0	100	1378	22	c_2
	0	0	63	1437	0	0	41	1459	0	0	27	1473	0	0	50	1450	0	0	112	1388	c_3				
MVF	1438	62	0	0	MVF	1438	62	0	0	MVF	1384	116	0	0	MVF	1479	21	0	0	MVF	1475	25	0	0	c_0
	111	1373	16	0		183	1303	14	0		187	1312	1	0		90	1409	1	0		30	1470	0	0	c_1
	2	60	1419	19		0	53	1436	11		4	20	1476	0		0	18	1478	4		0	10	1490	0	c_2
	0	0	14	1486	0	0	14	1486	0	0	1	1499	0	0	3	1497	0	0	4	1496	c_3				

k Number of nearest neighbors for the k -NN algorithm **MVF** Majority Vote Fusion

TABLE VIII

CLASSIFICATION RESULTS FOR THE TOP THREE OPTIMAL SENSORS OBTAINED USING THE UNSUPERVISED EMBEDDED ALGORITHM

UE	Sensor	Day Type 1				Day Type 2				Day Type 3				Day Type 4				Day Type 5			
		T_{cm}	T_{hx}	P_1	MVF	T_{hx}	T_{cm}	P_o	MVF	P_o	T_{hx}	T_{out}	MVF	T_{out}	P_o	T_{hx}	MVF	T_{hx}	T_{out}	P_o	MVF
	NE	0.50	0.51	0.52	-	0.42	0.45	0.45	-	0.44	0.46	0.50	-	0.44	0.46	0.47	-	0.49	0.49	0.51	-
PCA + k -NN	CCR	89.7	91.4	85.8	94.6	90.4	86.9	85.9	94.2	86.8	91.4	82.5	92.9	92.3	89.4	94.6	96.8	96.6	96.8	91.7	98.9
	FA	8.1	9.7	14.4	4.1	9.9	11.7	12.4	4.3	18.5	11.6	29.0	10.1	15.4	11.2	9.4	5.4	5.8	4.5	6.9	1.3
	MD	3.0	2.3	6.9	2.9	3.2	4.7	8.9	3.8	6.4	3.6	9.6	5.1	4.8	4.6	2.6	2.2	1.9	2.1	2.4	0.9
GMM + k -NN	CCR	87.0	91.1	87.5	95.3	90.8	84.4	87.4	94.4	88.2	93.9	80.1	94.5	93.9	90.9	93.8	97.7	95.5	95.8	92.0	98.9
	FA	14.1	7.7	13.1	4.1	8.6	17.1	9.7	4.1	12.2	8.9	35.9	7.7	10.4	6.8	3.5	1.4	5.9	5.9	5.8	1.7
	MD	4.6	2.5	6.4	2.5	3.5	5.7	8.6	4.1	7.5	2.6	10.0	4.2	3.1	4.5	3.5	2.0	2.1	1.8	2.7	0.7
UE	Unsupervised Embedded				MVF				Majority Vote Fusion				NE				Normalized Entropy				
CCR	Correct Classification Rate (%)				FA				False Alarm Rate (%)				MD				Miss-Detection Rate (%)				

TABLE IX

COMPUTATION TIMES FOR THE SENSOR SELECTION ALGORITHMS

Sensor Selection Technique	Procedure	Computation Time Per Day Type
Unsupervised Embedded	Maximum Entropy Partitioning	1 s
	mRMR	5.2 s
	Unsupervised Embedded	4.8 s
	Total	11 s
Embedded Wrapper (PCA + k -NN)	Maximum Entropy Partitioning	1 s
	mRMR	5.2 s
	PCA	1.06 s
	k -NN Tuning	2.9 hours
	Total	≈ 2.9 hours
Embedded Wrapper (GMM + k -NN)	Maximum Entropy Partitioning	1 s
	mRMR	5.2 s
	GMM	7.55 min
	k -NN Tuning	36.1 s
	Total	8.26 min

which is composed of the weights, means, and variances of the 2 mixtures as explicitly stated in Eq. (24). Similar to the PCA procedure, the GMM feature space is also augmented with the ambient temperature. Due to its high dimension, GMM feature vectors are difficult to visualize; however, it leads to high CCR in conjunction with k -NN classifier as discussed below.

The confusion matrices of the PCA + k -NN and the GMM + k -NN methods are presented in Tables VI and VII, respectively. An interesting observation on the confusion matrices is that false-alarms and miss-detections mostly occur between adjacent fouling classes. Table VIII presents the CCR results which indicate that the performances of PCA and GMM are comparable, with a slight lead for the GMM.

TABLE X

COMPUTATION TIMES FOR THE TESTING PHASE

Procedure	Computation Time
PCA + k -NN	50 ms/data sequence
GMM + k -NN	0.7 ms/data sequence
PCA + k -NN + Majority Vote	150 ms/data sequence
GMM + k -NN + Majority Vote	2.1 ms/data sequence

Table VIII shows that not only do optimal sensors selected using the unsupervised embedded algorithm lead to high CCRs (above 85% in most of the cases), but also the majority vote fusion of resultant sensors lead to superior results, above 94% for most of the day types. Besides that, testing the PCA requires majority vote among the 30 feature vectors (as described in Section V-A1) unlike the GMM; this makes the GMM much faster to train and test. Nonetheless, training the GMM requires the use of the Expectation Maximization (EM) algorithm, which is more expensive than the KL algorithm. The computation times on 32-bit MatLab running on a 3.10 GHz Intel(R) Core(TM) i5 – 2400 processor, 16 GB ram and Windows 7 Operating System are shown in Tables IX and X for training and testing phases, respectively.

VII. CONCLUSIONS AND FUTURE WORK

The paper presented a methodology for fouling diagnosis of the Secondary Heat Exchanger in the *Environmental Control System* of an aircraft that regulates temperature, pressure and humidity of the cabin air as well as the air used to cool

electronics onboard the aircraft. Since the ECS contains a large number of sensors, an optimal sensor selection methodology is presented to select the most useful sensors that provide the best diagnosis results. The results of unsupervised embedded (UE) algorithm for sensor selection are compared with embedded wrapper algorithms. It is shown that the sensors ranked by UE algorithm yield excellent classification results with significant improvement in computational complexity. Subsequently, the data of the top ranked sensors are analysed using the k -NN classifier in combination with either PCA or GMM as feature extractors and results are compared. The data is generated from an experimentally validated high-fidelity Simulink model of the ECS provided by our industry partner and included various uncertainties generated by parametric fluctuations in ambient temperature, occupant count, and flow impedance. Finally, the majority vote algorithm is applied as a simple fusion technique for further enhancement of the diagnosis results.

The following areas are envisioned for future research:

- Real-time implementation of the proposed heat exchanger fouling diagnosis methodology on actual aircraft data
- Testing and validation of different sensor fusion methods
- Utilization of different machine learning tools for improving the classification performance
- Development of a similar fouling diagnosis methodology for aircraft cruise operating conditions.

APPENDIX A MAXIMUM ENTROPY DISTRIBUTION

Consider the optimization problem in Eq. (10). Using the method of Lagrange multipliers, we define:

$$\mathcal{J} = H(X) - (\lambda - 1) \left(\sum_{j=1}^r p^j - 1 \right) \quad (27)$$

where $\lambda \geq 0$ is a real number. Taking the partial derivative with respect to p^j , $\forall j = 1, \dots, r$, and equating to zero, we get

$$\frac{\partial \mathcal{J}}{\partial p^j} = -\ln p^j - 1 - \lambda + 1 = 0 \quad (28)$$

Therefore we get:

$$p^{j*} = e^{-\lambda} \quad (29)$$

Summing Eq. (29) for $j = 1, \dots, r$, we get:

$$\lambda = \ln r \quad (30)$$

Using Eq. (30) into Eq. (29), we get the maximum entropy distribution as

$$p^{j*} = \frac{1}{r} \quad \forall j = 1 \dots r \quad (31)$$

which is the uniform distribution.

APPENDIX B CALCULATION OF MUTUAL INFORMATION

This section describes the calculation of the mutual information quantities $I(S_i, C)$ and $I(S_i, S_j)$. Let the symbol alphabet be equal to $\Sigma = \{\alpha^1 \dots \alpha^r\}$ such that

$\alpha^\omega \in \Sigma, \omega \in \{1, \dots, r\}$. Let $\gamma \in \Gamma, |\Gamma| = 1024$ denote a particular parametric combination for a simulation run. Let $\sigma_i(\gamma, \bullet) = [\sigma_i(\gamma, 1), \dots, \sigma_i(\gamma, L)]$ denote the symbol sequence for sensor s_i and simulation run γ , where $\sigma_i(\gamma, t) \in \Sigma, \forall t = 1, \dots, L$. Then, lets define a $|\Gamma| \times L$ matrix of all symbol sequences generated from the sensor $s_i, i = 1, \dots, N$, under different simulation runs as:

$$\mathcal{G}_i = [\sigma_i(\gamma, \bullet)]_{\gamma=1, \dots, |\Gamma|} \quad (32)$$

Lets now associate a label $a_i(\gamma)$ to each row $\sigma_i(\gamma, \bullet)$ of the matrix \mathcal{G}_i , such that $a_i(\gamma) = m$ if and only if $\sigma_i(\gamma, \bullet)$ belongs to the class $c_m, m \in \{1 \dots M\}$. Now, lets define the following:

- $\mu_i(\alpha^\omega)$: the number of occurrences of the event when $\sigma_i(\gamma, t) = \alpha^\omega$ in the matrix \mathcal{G}_i .
- $\nu_i(\alpha^\omega, c_m)$: the number of occurrences of the joint event when $\sigma_i(\gamma, t) = \alpha^\omega$ and $a_i(\gamma) = m$ in the matrix \mathcal{G}_i .
- $\chi_{i,j}(\alpha^\omega, \alpha^v)$: the number of occurrences of the joint events $\sigma_i(\gamma, t) = \alpha^\omega$ and $\sigma_j(\gamma, t) = \alpha^v$.

Then the following entropies are calculated:

$$H(S_i) = - \sum_{\omega=1}^r \frac{\mu_i(\alpha^\omega)}{|\Gamma| \cdot L} \ln \frac{\mu_i(\alpha^\omega)}{|\Gamma| \cdot L} \quad (33)$$

$$H(S_i, C) = - \sum_{m=1}^M \sum_{\omega=1}^r \frac{\nu_i(\alpha^\omega, c_m)}{|\Gamma| \cdot L} \ln \frac{\nu_i(\alpha^\omega, c_m)}{|\Gamma| \cdot L} \quad (34)$$

$$H(S_i, S_j) = - \sum_{\omega=1}^r \sum_{v=1}^r \frac{\chi_{i,j}(\alpha^\omega, \alpha^v)}{|\Gamma|^2 \cdot L} \ln \frac{\chi_{i,j}(\alpha^\omega, \alpha^v)}{|\Gamma|^2 \cdot L} \quad (35)$$

The mutual information quantities required in Eq. (18) are then calculated using Eq. (8).

ACKNOWLEDGMENTS

The authors would like to thank Leaper, Orlando, and Pattipati for their valuable contributions in this research.

REFERENCES

- [1] I. Pérez-Grande and T. J. Leo, "Optimization of a commercial aircraft environmental control system," *Appl. Thermal Eng.*, vol. 22, no. 17, pp. 1885–1904, 2002.
- [2] N. Najjar *et al.*, "Heat exchanger fouling diagnosis for an aircraft air-conditioning system," SAE Tech. Paper 2013-01-2250, Sep. 2013.
- [3] T. A. Pogiatis, D. I. Wilson, and V. S. Vassiliadis, "Scheduling the cleaning actions for a fouled heat exchanger subject to ageing: MINLP formulation," *Comput. Chem. Eng.*, vol. 39, pp. 179–185, Apr. 2012.
- [4] J. Hare, S. Gupta, N. Najjar, P. D'Orlando, and R. Walthall, "System-level fault diagnosis with application to the environmental control system of an aircraft," SAE Tech. Paper 2015-01-2583, Sep. 2015.
- [5] T. W. S. Chow and D. Huang, "Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 213–224, Jan. 2005.
- [6] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997.
- [7] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 3, pp. 131–156, 1997.
- [8] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.
- [9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [10] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

- [11] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artif. Intell.*, vol. 151, nos. 1–2, pp. 155–176, 2003.
- [12] C. H. Lo, P. T. Chan, Y. K. Wong, A. B. Rad, and K. L. Cheung, "Fuzzy-genetic algorithm for automatic fault detection in HVAC systems," *Appl. Soft Comput.*, vol. 7, no. 2, pp. 554–560, 2007.
- [13] K. Choi, S. M. Namburu, M. S. Azam, J. Luo, K. R. Pattipati, and A. Patterson-Hine, "Fault diagnosis in HVAC chillers," *IEEE Instrum. Meas. Mag.*, vol. 8, no. 3, pp. 24–32, Aug. 2005.
- [14] Q. Zhou, S. Wang, and Z. Ma, "A model-based fault detection and diagnosis strategy for HVAC systems," *Int. J. Energy Res.*, vol. 33, no. 10, pp. 903–918, 2009.
- [15] S. Katipamula and M. R. Brambley, "Review article: 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.
- [16] S. Katipamula and M. R. Brambley, "Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, Part II," *HVAC&R Res.*, vol. 11, no. 2, pp. 169–187, 2005.
- [17] R. A. Buswell and J. A. Wright, "Uncertainty in model-based condition monitoring," *Building Services Eng. Res. Technol.*, vol. 25, no. 1, pp. 65–75, 2004.
- [18] R. M. Kelso and J. A. Wright, "Application of fault detection and diagnosis techniques to automated functional testing," *ASHRAE Trans.*, vol. 111, no. 1, pp. 964–970, 2005.
- [19] H. Yang, S. Cho, C.-S. Tae, and M. Zaheeruddin, "Sequential rule based algorithms for temperature sensor fault detection in air handling units," *Energy Convers. Manage.*, vol. 49, no. 8, pp. 2291–2306, 2008.
- [20] S. Wang and F. Xiao, "AHU sensor fault diagnosis using principal component analysis method," *Energy Buildings*, vol. 36, no. 2, pp. 147–160, 2004.
- [21] J. E. Pakanen and T. Sundquist, "Automation-assisted fault detection of an air-handling unit; implementing the method in a real building," *Energy Buildings*, vol. 35, no. 2, pp. 193–202, 2003.
- [22] J. Qin and S. Wang, "A fault detection and diagnosis strategy of VAV air-conditioning systems for improved energy and control performances," *Energy Buildings*, vol. 37, no. 10, pp. 1035–1048, 2005.
- [23] 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.
- [24] Y. Zhao, S. Wang, and F. Xiao, "A statistical fault detection and diagnosis method for centrifugal chillers based on exponentially-weighted moving average control charts and support vector regression," *Appl. Thermal Eng.*, vol. 51, nos. 1–2, pp. 560–572, 2013.
- [25] N. Najjar *et al.*, "Health assessment of liquid cooling system in aircrafts: Data visualization, reduction, clustering and classification," *SAE Int. J. Aerosp.*, vol. 5, no. 1, pp. 119–127, 2012.
- [26] L. Shang and G. Liu, "Sensor and actuator fault detection and isolation for a high performance aircraft engine bleed air temperature control system," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 5, pp. 1260–1268, Sep. 2011.
- [27] S. Gupta, A. Ray, S. Sarkar, and M. Yasar, "Fault detection and isolation in aircraft gas turbine engines: Part 1: Underlying concept," *Proc. Inst. Mech. Eng. G, J. Aerosp. Eng.*, vol. 222, no. 3, pp. 307–318, 2008.
- [28] S. Sarkar, M. Yasar, S. Gupta, A. Ray, and K. Mukherjee, "Fault detection and isolation in aircraft gas turbine engines: Part 2: Validation on a simulation test bed," *Proc. Inst. Mech. Eng. G, J. Aerosp. Eng.*, vol. 222, no. 3, pp. 319–330, 2008.
- [29] R. Isermann, "Model-based fault-detection and diagnosis—Status and applications," *Annu. Rev. Control*, vol. 29, no. 1, pp. 71–85, 2005.
- [30] H. Kaneko *et al.*, "Statistical approach to constructing predictive models for thermal resistance based on operating conditions," *Ind. Eng. Chem. Res.*, vol. 51, no. 29, pp. 9906–9912, 2012.
- [31] C. Riverol and V. Napolitano, "Estimation of fouling in a plate heat exchanger through the application of neural networks," *J. Chem. Technol. Biotechnol.*, vol. 80, no. 5, pp. 594–600, 2005.
- [32] R. F. Garcia, "Improving heat exchanger supervision using neural networks and rule based techniques," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3012–3021, 2012.
- [33] A. Adili, N. Hasni, C. Kerkeni, and S. B. Nasrallah, "An inverse problem based on genetic algorithm to estimate thermophysical properties of fouling," *Int. J. Thermal Sci.*, vol. 49, no. 6, pp. 889–900, 2010.
- [34] H. Han, B. Gu, T. Wang, and Z. Li, "Important sensors for chiller fault detection and diagnosis (FDD) from the perspective of feature selection and machine learning," *Int. J. Refrig.*, vol. 34, no. 2, pp. 586–599, 2011.
- [35] S. M. Namburu, M. S. Azam, J. Luo, K. Choi, and K. R. Pattipati, "Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers," *IEEE Trans. Autom. Sci. Eng.*, vol. 4, no. 3, pp. 469–473, Jul. 2007.
- [36] S. Jiang, R. Kumar, and H. E. Garcia, "Optimal sensor selection for discrete-event systems with partial observation," *IEEE Trans. Autom. Control*, vol. 48, no. 3, pp. 369–381, Mar. 2003.
- [37] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, 2006.
- [38] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.
- [39] A. O. Hero and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors J.*, vol. 11, no. 12, pp. 3064–3075, Dec. 2011.
- [40] J. Xu, Y. Wang, and L. Xu, "PHM-oriented sensor optimization selection based on multiobjective model for aircraft engines," *IEEE Sensors J.*, vol. 15, no. 9, pp. 4836–4844, Sep. 2015.
- [41] X. Shen, S. Liu, and P. K. Varshney, "Sensor selection for nonlinear systems in large sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2664–2678, Oct. 2014.
- [42] H. Zhao, Y. Hou, Y. Zhu, L. Chen, and S. Chen, "Experimental study on the performance of an aircraft environmental control system," *Appl. Thermal Eng.*, vol. 29, no. 16, pp. 3284–3288, 2009.
- [43] J. V. C. Vargas and A. Bejan, "Thermodynamic optimization of finned crossflow heat exchangers for aircraft environmental control systems," *Int. J. Heat Fluid Flow*, vol. 22, no. 6, pp. 657–665, 2001.
- [44] A. Alebrahim and A. Bejan, "Thermodynamic optimization of heat-transfer equipment configuration in an environmental control system," *Int. J. Energy Res.*, vol. 25, no. 13, pp. 1127–1150, 2001.
- [45] S. Wright, G. Andrews, and H. Sabir, "A review of heat exchanger fouling in the context of aircraft air-conditioning systems, and the potential for electrostatic filtering," *Appl. Thermal Eng.*, vol. 29, no. 13, pp. 2596–2609, 2009.
- [46] H. Peng and X. Ling, "Optimal design approach for the plate-fin heat exchangers using neural networks cooperated with genetic algorithms," *Appl. Thermal Eng.*, vol. 28, nos. 5–6, pp. 642–650, 2008.
- [47] M. Mishra, P. K. Das, and S. Sarangi, "Second law based optimisation of crossflow plate-fin heat exchanger design using genetic algorithm," *Appl. Thermal Eng.*, vol. 29, nos. 14–15, pp. 2983–2989, 2009.
- [48] K. Thulukkanam, *Heat Exchanger Design Handbook*, 2nd ed. CRC Press, 2013.
- [49] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
- [50] S. Gupta and A. Ray, "Statistical mechanics of complex systems for pattern identification," *J. Statist. Phys.*, vol. 134, no. 2, pp. 337–364, 2009.
- [51] D. Lind and M. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
- [52] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, 1957.
- [53] D. Huang and T. W. S. Chow, "Effective feature selection scheme using mutual information," *Neurocomputing*, vol. 63, pp. 325–343, Jan. 2005.
- [54] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained K-means clustering with background knowledge," in *Proc. 18th ICML*, vol. 1. 2001, pp. 577–584.
- [55] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B (Methodol.)*, vol. 39, no. 1, pp. 1–38, 1977.
- [56] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [57] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. New York, NY, USA: Elsevier, 2008.



Nayeff Najjar received the B.S. degree in electrical engineering from the King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 2005, and the M.E. degree from Widener University, Chester, PA, USA, in 2008. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT. His research interests include data analysis in complex networked systems, fault diagnosis and prognosis, machine learning, sensor/feature selection, feature extraction, and information fusion.



Shalabh Gupta (M'07) received the M.S. degrees in mechanical and electrical engineering and the Ph.D. degree in mechanical engineering in 2006 from Pennsylvania State University, University Park. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Connecticut. His research interests include the science of autonomy, cyber-physical systems, swarm robotics, intelligent systems, data analysis, network science, and fault diagnosis and prognosis in complex systems. He is a member of the American

Society of Mechanical Engineers.



Sherif Kandil received the B.S. degree in mechanical engineering from America University in Cairo, the M.S.M.E. degree in mechanical engineering from West Virginia University, and the Ph.D. degree in mechanical engineering from the University of Florida. Currently, he is responsible for the development of advanced technologies for current and future Air Management Systems at UTC. He began his UTC career at United Technologies Research Center in 2010, where he held roles of increasing responsibility as a Staff Scientist leading multi-year

research and development projects in the HVAC&R and thermal management fields. Prior to joining UTC, he spent four years at Caterpillar as a Cooling Systems Design Engineer, performing component and system thermal CFD analysis.



James Hare received the B.S. degree in electrical engineering from the University of Connecticut, Storrs, CT, USA, in 2012, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. His research interests include decision making in distributed sensor networks, intelligent network control systems, data analysis in complex-networked systems, and fault detection and diagnosis.



Rhonda Walthall received the B.S. degree in aeronautical and astronautical engineering from Purdue University, and the master's degree in business administration from Pepperdine University. Currently, she is the Manager of Prognostics and Health Management at UTC Aerospace Systems. Prior to joining UTC in 2003, she worked for Northwest Airlines and McDonnell Douglas Aircraft Company. She is also an active member of SAE International, where she is currently serving as the Chair of the Integrated Vehicle Health Management (IVHM)

Steering Group and the Chair of the 2016 Fellows Selection Committee. She was an author and contributor to the SAE publications *Integrated Vehicle Health Management: Implementation and Lessons Learned* and *Integrated Vehicle Health Management: Business Case Theory and Practice*. She was a presenter in the SAE webcast "Taking Data to New Heights: How Airlines, Plane Manufacturers, and Suppliers are Shaping the Future of Integrated Vehicle Health Management." She is an active member of the Prognostics Health Management Society, Toastmasters International, and Women in Aviation International. She holds one patent and two invention disclosures.