

Health Assessment of Liquid Cooling System in Aircrafts: Data Visualization, Reduction, Clustering, and Classification

Nayeff Najjar, Chaitanya Sankavaram, and James Hare
University of Connecticut

Shalabh Gupta
University of Connecticut

Krishna Pattipati
University of Connecticut

Rhonda Walthall and Paul D'Orlando
Hamilton Sundstrand

ABSTRACT

This paper addresses the issues of data reduction, visualization, clustering and classification for fault diagnosis and prognosis of the Liquid Cooling System (LCS) in an aircraft. LCS is a cooling system that consists of a left and a right loop, where each loop is composed of a variety of components including a heat exchanger, source control units, a compressor, and a pump. The LCS data and the fault correlation analysis used in the paper are provided by Hamilton Sundstrand (HS) - A United Technologies Company (UTC). This data set includes a variety of sensor measurements for system parameters including temperatures and pressures of different components, along with liquid levels and valve positions of the pumps and controllers. A graphical user interface (GUI) is developed in Matlab that facilitates extensive plotting of the parameters versus each other, and/or time to observe the trends in the data. The parameters that reflect interesting information are selected by observing the correlations between different parameters. The data are analyzed using the wavelet transform to highlight the interclass separation and subdue the within class differences for more accurate classification. Subsequently, the number of parameters is significantly reduced using the *principal component analysis* (PCA). PCA-based data reduction resulted in different clusters when applied to the healthy data, and faulty data of the left and right loops. Several classification methods have been tested and their performance was evaluated. The results of this paper will be used for the purpose of fault diagnosis and prognosis in LCS. The methodology used in this paper can also be applied to other HS systems.

CITATION: Najjar, N., Sankavaram, C., Hare, J., Gupta, S. et al., "Health Assessment of Liquid Cooling System in Aircrafts: Data Visualization, Reduction, Clustering and Classification," *SAE Int. J. Aerosp.* 5(1):2012, doi: 10.4271/2012-01-2106.

INTRODUCTION

This paper addresses the issues of data visualization, reduction, clustering, and classification for fault diagnosis and prognosis of the Liquid Cooling System (LCS) in an aircraft. The LCS is a cooling system that consists of the left and the right loop, where each loop is composed of a variety of components, including a heat exchanger, source control units, a compressor, and a pump. These components gradually degrade over several cycles of operation and timely

diagnosis is critical for prevention of the loss of equipment, premature maintenance costs, and part failures. For the purpose of fault diagnosis and prognosis, several data sets are collected from a number of heterogeneous sensors (e.g., temperature, pressure, pump speed, etc.) onboard. These data are collected for different operating modes, different tail numbers, and for the nominal (i.e., healthy) as well as the faulty conditions. These data sets are noisy, voluminous, redundant, and are often difficult to visualize. Therefore, the objective of this paper is to extract useful information from

these numerous heterogeneous sensor parameters via data reduction to a set of few features and to identify clusters in the data to enable classification of the nominal and faulty conditions.

The technical literature abounds with diverse tools for data analysis [1], [2]. The art of machine learning and pattern recognition can be classified into three main categories: supervised, unsupervised, and semi-supervised. Supervised machine learning techniques, or classification, classify the patterns of the measured data based on well-defined labels. In contrast, unsupervised machine learning algorithms, or clustering techniques, are applied to classification problems where the class labels are unknown [1], [2]. Semi-supervised machine learning techniques combine the supervised and the unsupervised algorithms to enhance the efficiency of classification [3]. The following advantages pertain to clustering. Firstly, clustering eliminates the cost and time complexity associated with classification algorithms. Secondly, when data varies slowly with time, clustering techniques can be used to infer the drift in data ("the so-called concept drift"). Clustering can also be used for feature extraction and visualization of the complexity of the problem [2]. Finally, in many cases, data is gathered by a third party who may not know the reason behind data collection; clustering helps focus the data collection and analysis efforts to relevant data.

Clustering techniques have been widely used in the literature for fault diagnosis and prognosis [4] [5]. In [6], the authors combined the *Self-Organizing Feature Map* (SOFM) and the K-means clustering algorithm for automated determination of the number of neurons and iterations of the SOFM. Grey Clustering Analysis is suggested for early fault detection of oil-immersed transformers in [7]. In [8], the author applied SOFM and *Gaussian Mixture Models* (GMM) and K-means clustering to bearing test beds, proposing a *Log Likelihood probability* (LLP) for health assessment and the *Bayesian Inference-based failure Probability* (BIP) algorithm to calculate failure risk probability of the machine. A new clustering technique named *Compensation Distance Evaluation Technique* (CDET) was proposed in [9]. A novel method of clustering, called *adaptive resonance theory and the learning strategy of Kohonen neural network* (ART-KNN), was suggested in [10] for bearing vibration data and it achieved better accuracy than *SOFM*, *Learning Vector Quantization* (LVQ), and *Radial Basis Function* (RBF) methods. A method of fault detection in a semiconductor manufacturing process using the k-nearest neighbor rule was presented in [11]. The PCA method was used to reduce the amount of data from an ultrasonic sensor system managed by an *artificial neural network* (ANN) in [12]. Using clustering and PCA, [13] tested an Integrated Gasification and Combined Cycle section of an oil refinery plant to show their effectiveness in process fault detection and isolation. Fault detection and identification on railway infrastructure

components based on pattern recognition and data analysis algorithms (e.g., PCA) was discussed in [14]. Health monitoring of gas turbine engines was presented in [15] using PCA and ANN analysis for early detection of anomalies. An enhanced version of the *Clusot* algorithm to utilize the information contained in a trained SOFM and the automatic detection of clusters in the Clusot surface was described in [16]. The application of wavelet transform to detect and isolate an external leakage fault in valve-controlled hydraulic actuators was presented in [17]. A new method based on wavelet packet *Principal Component Analysis-Support Vector Machines* (PCA-SVM) was introduced in [18] for fault diagnosis of turbine generators. A new semi-supervised learning approach based on Gaussian random field model was proposed in [3]. The use of Laplace-Beltrami operator to generate the Laplacian Eigenmaps as method for semi-supervised learning was suggested in [19].

The wavelet transform is very effective in noise reduction on high-dimensional data without substantially affecting the essential information in the signal [20]. In addition, the wavelet transform of the data also helps in improving the results of the clustering algorithms by enhancing the interclass separation of clusters. After preprocessing the sensor data with the wavelet transform, this paper applied the PCA method for data reduction and to infer the clusters of the normal and faulty modes in LCS.

The main contributions of this paper are as follows:

- i. Design and development of a graphical user interface (GUI) that facilitates visualization as well as analysis of heterogeneous sensor data;
- ii. Data reduction and clustering via transformation of the sensor data using appropriate wavelet transform and using the PCA method for identification of clusters in the data that correspond to the faulty and the healthy data sets. Several classification methods have been tested and their performance was evaluated.

This paper is organized as follows. The second section presents the description of the LCS data that is used throughout the paper. The data visualization procedure is discussed next. Following this, different analysis techniques are described. This is followed by an application of the analysis techniques on the LCS data. The final section concludes the paper with a discussion of future research.

DATA DESCRIPTION

The Liquid Cooling System (LCS) is a cooling system in aircraft. The LCS data used in this paper was supplied by Hamilton Sundstrand and stored in three reports. The first one is the performance report, denoted as *report P* in this paper. This report contains sensor data collected for 5 tail numbers¹ at nominal conditions when the system is known to be healthy. The data were collected for a certain period for the five different tail numbers; however, quantities of data sets

¹Tail number is a unique identifier for each aircraft.

for each airplane are different. This data includes general parameters, such as latitude, longitude, total air temperature, and static pressures. In addition, some sensor readings such as loop-supply temperature, pressure, temperature valve position, pump speed, pump temperature, pump current, and isolation valve states are provided for various events. The events are ground start, main engine start, hydraulic start, and engine shutdown. The data is collected for different system modes, left and right pump 1 status, left and right pump 2 status, etc. The data file spanned 449 different sensor readings and the total number of data points for each sensor was 397.

The other two reports are the left and right event reports, denoted as *report L* and *report R*, respectively. These reports record sensor measurements for events that are triggered when a certain parameter exceeds its thresholds indicating a fault. The readings are recorded on the aircraft. Reports L and R were collected over almost the same period. All the three reports are collected for the same tail numbers. Reports L and R do not provide any information about the type of fault that was triggered when the data were being recorded, nor do they provide any information on how the system was judged to be faulty. Data in reports L and R, however, provide sensor measurements during the period from 5 seconds before the occurrence of an event to 5 seconds after the event. Some of the sensor measurements are not provided for some of the events. Report L contains 336 events, while Report R contains 507 events with different time stamps. Each of the event reports L and R contains 162 sensor readings. The goal of this paper is to extract clusters from the LCS data for classification of the faulty and the nominal data sets.

DATA VISUALIZATION PROCEDURE

A Matlab Graphical User Interface (GUI) is developed to facilitate plotting and visualization of large data sets consisting of diverse heterogeneous sensors' readings. These data sets, as described in the previous section, are collected over a period of time over different modes of operation, different tail numbers, and for healthy as well as faulty systems. The GUI provides the capability to plot the sensor parameters vs. time and against each other to observe their time evolution and the correlations between different parameters, respectively.

The GUI is also equipped with several data analysis tools such as wavelet transforms, statistical analysis methods, and several feature extraction and classification methods. For the LCS data sets, the entropy-based thresholds are applied to the data as a simple preliminary step for data reduction. Subsequently, the data are transformed using the wavelet transform and the PCA clustering algorithm is applied to visualize the clusters in the data. The procedure is summarized in [Figure 1](#) and elaborated below.

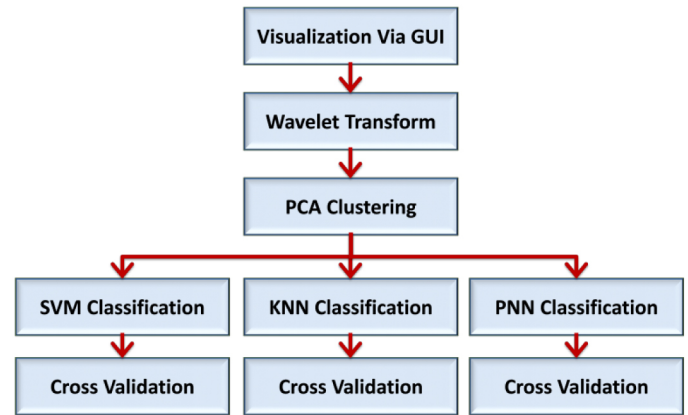


Figure 1. Data Analysis Procedure

Graphical User Interface (GUI)

The graphical user interface allows the user to visualize and assess correlations among sensed parameters. [Figure 2](#) shows the main window of the GUI; the users can select and load the data source file (e.g., report L, report R, or report P) and choose the category to group the data (e.g., by tail number or by different system modes). The GUI enables plotting the data vs. time to observe the chronological trends in data along with their upper and lower limits for the nominal conditions, as described in the report files.

For convenience, the users may choose to plot as scatter or as continuous lines. Moreover, the users may apply different types of wavelet transforms (e.g., Haar, Symlets, Coiflets, Meyer, and Gaussian) by choosing different scales for data preprocessing and visualize the outcomes. The GUI facilitates statistical analysis of the data to generate the histograms for different parameters. The plots can be generated in two or three dimensions. The plots provide the *sample* mean, standard deviation, and skewness of the parameters based on the following equations:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2} \quad (2)$$

$$\text{Skewness} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^3}{S^3} \quad (3)$$

where x_i is a data point. The mean of the data represents where the data are concentrated; the standard deviation is a measure of the dispersion of the data, while the skewness of the data measures how much the distribution deviates from the normal distribution. After plotting the histograms, the user can select to export the data to Excel sheets.