Physica A 506 (2018) 248–253

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The mutual information based minimum spanning tree to detect and evaluate dependencies between aero-engine gas path system variables

Keqiang Dong [a,*], Linan Long [a], Hong Zhang [b], You Gao [a]

[a] College of Science, Civil Aviation University of China, Tianjin 300300, China
[b] Basic Courses Department, Tianjin Sino-German University of Applied Sciences, Tianjin 300300, China

HIGHLIGHTS

- We examine the dependence in gas path system by average mutual information method.
- We propose the minimum spanning tree test based mutual information distance.
- The mutual information based the MST demonstrate two observably parameter clusters.

ARTICLE INFO

Article history:
Received 14 October 2017
Received in revised form 20 February 2018
Available online 21 April 2018

Keywords:
Mutual information
Minimum spanning tree
Complex system
Statistical dependence

ABSTRACT

There is a great interest in studying statistical dependence characteristics of aero-engine gas path system time series. The mutual information is effective, mainly in quantifying the dependency of time series. By applying the mutual information and average mutual information method to aero-engine gas path system, the statistical dependence between two data steams from a finite number of samples are established. To better understand dependency of gas path system time series, we define the mutual information distance and propose the mutual information based minimum spanning tree to investigate the performance parameters and their interaction of gas path system. By examining the minimum spanning tree, we find that the exhaust gas temperature (EGT) and the low-spool rotor speed (N1) are confirmed as the predominant variables in fourteen gas path parameters. The results show that the proposed method is effective to detect the statistical dependence of gas path system parameters and has more valuable information.

1. Introduction

Aero-engine gas path system performance parameters are determined by the performance of its gas path components and their interaction. Performance-related problems can be detected by measuring and monitoring performance parameters and their interaction along the gas path such as pressures and temperatures. The measure and analysis of gas path system performance was subject to linear approximation methods in the late 1970's and early 1980's [1,2], which led to the accuracy and reliability was limited. However, the presence of high degree dependence or statistical correlation between the synchronous time evolutions of a set of aero-engine gas path system is a well-known empirical fact [3]. Therefore, to improve calculation accuracy has been noticed for developing better gas path analysis techniques [1–4]. Consequently, W.P.J.

* Corresponding author.
E-mail address: hongzhangdong@163.com (K. Dong).

https://doi.org/10.1016/j.physa.2018.04.059
0378-4371/

Visser used model based method to gas path system analysis [5]. M. Zedda and R. Singh applied machine learning techniques in combination with large volumes of performance data to performance parameters interaction analysis [6].

However, the complexity of aero-engine dynamic interaction among gas path components makes it difficult to quantify the statistical dependence among gas path system performances. The gas path component deterioration also affects component interaction, and the effect of deterioration is usually observed by simultaneous changes to several performance parameters [3,7]. So it is significant to estimate the statistical dependence among gas path system components. For considering the aero-engine dynamic interaction among gas path components, we attempt to quantify statistical correlation or statistical dependence among gas path performance parameters here. A development in the field of statistical correlation by Pearson is Pearson's correlation coefficient. However, Pearson's correlation coefficient measures only linear dependence by calculated and interpreted for positive or negative correlation [8,9]. Nevertheless, in aero-engine systems, the non-linear and non-stationary characteristics are present [10,11]. Fortunately for us, mutual information has the advantage of being the measurement of dependence that measures all (linear and non-linear) dependence between stochastic variables [12,13].

Mutual information, which is a powerful and well developed tool to detect if two data are independent or not, proposed by Shannon in 1948 has received wide attention since it was put forward [14,15]. Because the mutual information method allows us to detect the statistical dependence between noisy signals with non-linearity and non-stationary [12,13,16–19], the mutual information method has been successfully applied to research fields such as biology [20,21], graphics processing [22,23], high energy physics [24] and financial markets [25].

Here the aero-engine gas path performance parameters are investigated by applying the mutual information and average mutual information method firstly. The preliminary test results demonstrate that the dependency properties between the gas path performance parameter series exist. And then, the mutual information distance matrix is introduced to study the internal structure of the gas path system. Further, for investigating the predominant variables, the minimum spanning tree (MST) method is proposed to learn about the coreness of interactions in the gas path system [26,27]. The results indicate that the low-spool rotor speed (N1), and exhaust gas temperature (EGT), may be the predominant gas path elements. The possible predominant variables are in accord with that of previous document [28–30].

The organization of this paper is as follows. In the next section, we simply present the aero-engine gas path performance parameters employed. We show the mutual information method, the average mutual information method and their empirical results in Section 3. In Section 4, we introduce a new technique named mutual information based minimum spanning tree method enabling us to estimate the parameter dependency characteristic for aero-engine gas path system. In particular, the ability to identify the internal structure in gas path system is demonstrated by fourteen parameters. Finally, we draw some conclusions in Section 5.

2. The dataset

The aero-engine gas path performance parameters, offered by Aircraft Maintenance and Engineering Corporation, are civil aviation flight data from on-board flight data recorders, which are part of the Quick Access Recorder. The Quick Access Recorder data consist of an extensive list of flight parameters recorded at specific sampling intervals which are set by the manufacturer [31,32]. Fig. 1 shows a section of the Quick Access Recorder data. By means of the aero-engine gas path parameters, the object of mutual information analysis in this paper is to detect as many of the statistical dependence of the gas path system as possible.

In order to amplify the study of statistical dependence between aero-engine gas path variables, we here measure the value of mutual information for fourteen variables. The performance parameters such as exhaust the gas temperature (EGT), low-spool rotor speed (N1), high-spool rotor speed (N2), N1 tracked vibration channel A (N1TA), N1 tracked vibration channel B (N1TB), N2 tracked vibration channel A (N2TA), N2 tracked vibration channel B (N2TB), combustion inlet temperature (T3), other temperature (T2.5), inlet air pressure (P2), other pressure (P2.5, P5), fuel flow (WF), and wind speed (WIND) are employed in this paper.

3. Mutual information method and empirical results

3.1. Mutual information method

Mutual information is a measure of how much of one time series depends on another. Zero mutual information means two time series are independent or their distributions are separable. Mutual information increases as the dependency between two time series increases. The mutual information is a more general measure of correlation than the correlation coefficient which is only able to evaluate linear dependencies.

For two sets of variables X and Y, the joint distribution between is given by $p(x, y)$. If the two time series are independent, then their joint distribution will factorize into the product of their marginals $p(x, y) = p(x)p(y)$. If the two time series are not independent, some idea of whether they are close to independent can be obtained by considering the relative entropy (or Kullback–Leibler divergence) between the joint distribution and their marginals as follows [33]:

$$
\begin{aligned}
I(X, Y) &= KL(p(x, y)) \parallel p(x)p(y) \\
&= \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dxdy.
\end{aligned}
$$