Spring 2014

Data classification and forecasting using the Mahalanobis-Taguchi method

Adebolaji A. Jobi-Taiwo

DATA CLASSIFICATION AND FORECASTING USING THE MAHALANOBIS-TAGUCHI METHOD

by

ADEBOLAJI A. JOBI-TAIWO

A THESIS

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN SYSTEMS ENGINEERING

2014

Approved by

Dr. Elizabeth Cudney, Advisor
Dr. Steven Corns
Dr. Brian Smith

PUBLICATION THESIS OPTION

This thesis consists of the following three articles that have been prepared in the styles specified by the American Society for Engineering Management International Annual Conference, the Institute of Industrial Engineers Industrial and Systems Engineering Research Conference and the Institute of Electrical and Electronics Engineers Instrumentation and Measurement Journal respectively:

Pages 3-22 were accepted by the 2013 American Society for Engineering Management International Annual Conference and consist of the article – “A Review of Literature on Mahalanobis-Taguchi Strategy in Condition Monitoring”.

Pages 23-35 were accepted by the 2013 Institute of Industrial Engineers Industrial and Systems Engineering Research Conference and consist of the article – “Predicting Faults in Heavy Duty Vehicles Using the Mahalanobis-Taguchi Strategy”.

Pages 36-53 were submitted to the Institute of Electrical and Electronics Engineers Instrumentation and Measurement Journal and consist of the article – “Mahalanobis-Taguchi System for Multiclass Classification of Steel Plates Fault”.

ABSTRACT

Classification and forecasting are useful concepts in the field of condition monitoring. Condition monitoring refers to the analysis and monitoring of system characteristics to understand and identify deviations from normal operating conditions. This can be performed for prediction, diagnosis, or prognosis or a combination of any these purposes. Fault identification and diagnosis are usually achieved through data classification, while forecasting methods are usually used to accomplish the prediction objective. Data gathered from monitoring systems often consists of multiple multivariate time series and is fed into a model for data analysis using various techniques. One of the data analysis techniques used is the Mahalanobis-Taguchi strategy (MTS) because of its suitability for multivariate data analysis. MTS provides a means of extracting information in a multidimensional system by integrating information from different variables into a single composite metric. MTS is used to conduct analysis on the measurement parameters and seeks a correlation with the result while also seeking to optimize the analysis by identifying variables of importance strongly correlated with a defect or fault occurrence. This research presents the application of a MTS based system for predicting faults in heavy duty vehicles and the application of MTS in a multiclass classification problem. The benefits and practicality of the methodology in industrial applications are demonstrated through the use of real world data and discussion of results.

ACKNOWLEDGMENTS

Thank God for the gift of life through which all of this was possible. I express deep gratitude to Dr. Elizabeth Cudney, my advisor, for all the help and support throughout this research and my master's degree. Thank you for an invaluable learning experience and funding. I would like to thank my committee, Dr. Steven Corns and Dr. Brian Smith for their guidance and support. Thanks to Dr. Robert Woodley and Michael Gosnell for the opportunity to work with them. A special thanks to Linda Turner and Karen Swope for all their help throughout my time at the EMSE department.

I would like to thank my friends Olutosin Ajakaiye and Kayode Akinbobola for making my transition to life in Rolla seamless. Moyosore Onanuga and Adebayo Adewunmi, you guys made it fun as always. Where do we go from here? I would like to thank the Hassan and Adesanya families for providing me homes away from home. God bless you all.

Most importantly, I thank my parents and my family for their continued love, support, and prayers. Thank you for pushing me to do better always and for the values you instilled in me. Thank you for the opportunities you made available to me and my siblings, depriving yourselves so much. I am forever grateful. And to you, Fadekemi Adedayo, you made it all worth it.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

The Mahalanobis-Taguchi system (MTS) employs the Mahalanobis distance (MD) and principles of Taguchi Methods for pattern recognition in multidimensional analysis. The measure provided by MD accounts for the correlation within a group and this is vital to the MTS. The evaluation of homogeneity of a sample to a group cannot be complete without accounting for the interrelationship between the elements of the group. In the MTS, the Mahalanobis space (MS, the reference group) is created with the MD of the most representative sample of the homogenous state of interest. Once the MS is established, the required attributes of the MTS are optimized using orthogonal array (OA) and signal-to-noise ratio (SN) by evaluating the contribution of each attribute. The result is a simple yet robust metric for decision making in multivariate problems.

The focus of this research is on the application of the MTS to the data classification and prediction in condition monitoring problems. Paper I presents a review of available literature on the application of MTS to condition monitoring problems over the past decade. MTS has proven to be a valuable tool for cost effective condition monitoring in various fields. The historical information on the application of MTS to condition monitoring problems shows that high levels of accuracy are consistently achieved with fewer features of the system. It also shows MTS as a flexible tool which can be modified or combined with other tools and techniques as required by the peculiarities of a problem.

Paper II presents the application of MTS to condition monitoring in heavy duty vehicles. MTS is used to create a composite scale, which reduces the dimensionality of the problem space and forms the basis of the prediction model for the condition of the vehicles. Fifty-one attributes on the vehicles are monitored in real-time and the data fed into the system. The scale is used to measure the degree of abnormality of these measurements from the vehicles compared to “normal” measurements. MTS also reduced the dimensionality of the problem as 23 useful variables were identified from the original 51 attributes.

Condition monitoring for a single fault condition is an unbalanced one as processes and equipment are usually prone to failure due to multiple faults. Paper III presents MTS used in a multiclass classification for condition monitoring in a manufacturing process. The MTS methodology is applied to data collected on faults from the manufacturing of Steel Plates. The proposed scheme utilizes MD-based thresholds to classify faults into distinct fault groups by assessing the degree of abnormality in the variables being monitored relative to a reference group for each fault class. A classification threshold based on 1.5 sigma shift from the center of the measurement scales was utilized for each fault class.

The work in this thesis collates and presents the chronological progression of the available literature on the application of MTS to condition monitoring and other closely related problems. The work also presents the flexibility and robustness of MTS as it is combined with theory on normal process variation in a multiclass classification problem. MTS is also shown to reduce the dimensionality of a problem with successful results achieved with the reduced model. The research presented in the following papers all show practicality of MTS in the design of condition monitoring systems and demonstrate effectiveness in industrial application.

PAPER I

A REVIEW OF LITERATURE ON MAHALANOBIS-TAGUCHI STRATEGY IN CONDITION MONITORING

Adebolaji A. Jobi-Taiwo, Elizabeth A. Cudney
Missouri University of Science and Technology
Rolla, MO 65409

Robert S. Woodley
21st Century Systems, Inc.
Omaha, NE 68132

Abstract

This paper presents a review of literature on condition monitoring systems based on the Mahalanobis-Taguchi strategy (MTS). MTS is based on the Mahalanobis distance (MD), a distance measure which takes into account the correlation between variables in a data set. MD enables pattern recognition in multidimensional systems, which is one of the approaches used in the design of condition monitoring systems. MTS provides a means of extracting information in a multidimensional system, which has led to the use of the methodology in the development of fault detection, prediction, diagnosis, and prognosis systems. There are usually numerous parameters evaluated in condition monitoring. MTS significantly reduces the need to measure all parameters by identifying the variables that are strongly associated with a fault occurrence. This paper focuses on the applications of MTS to condition monitoring and other closely related problems.

Keywords

Mahalanobis-Taguchi Strategy, Mahalanobis Distance, Pattern Recognition, Condition Monitoring

Introduction

The Mahalanobis-Taguchi strategy (MTS) is a diagnosis and forecasting technique using multivariate data (Woodall, Koudelik, Tsui, Kim, Stoumbos & Carvounis, 2003). In a multivariate system, decision-making is typically based on analyzing information provided by more than one variable. Evaluation of each variable without considering the relationship to all other variables within the system would be incomplete. MTS bridges the relationships between variables using the Mahalanobis distance (MD). MD is a distance measure introduced by P.C. Mahalanobis in 1936; it is a generalized measure of a distance representing the degree of divergence in the mean values of different characteristics of a population considering the correlation between the variables (Taguchi & Jugulum, 2002). MD is a useful measure since it accounts for the correlation of the variables in a multidimensional system. For this reason, MTS is an ideal tool for the analysis of multivariate data and systems.

Condition monitoring refers to the analysis and monitoring of system characteristics to understand and identify deviations from normal operating conditions. This can be performed for prediction, diagnosis, or prognosis or a combination of any these purposes. Knowledge about the future timing of a fault occurrence and the fault cause can be used to improve the system. Data gathered from monitoring systems often consists of multiple multivariate time series (Xue, Williams, & Qiu, 2011) and is fed into a model for data analysis using various techniques. Condition monitoring approaches based on various data analysis techniques have been applied to different problems. One of the data analysis techniques used is MTS because of its suitability for multivariate data analysis. MTS provides an analysis on the measurement parameters and seeks a correlation with the result; as such, MTS helps in reducing the number of parameters being measured by identifying only the useful parameters contributing to the problem. MTS has been used in a number of condition monitoring systems for prediction, diagnosis, and prognosis or a combination of these. This study chronologically reviews the application of MTS to condition monitoring problems and other closely related research.

Mahalanobis Distance

MD is a distance measure derived from an analysis of the deviation in the mean values of different variables in multivariate analysis considering the correlation between the variables. MD, as a discriminant analysis method, is useful in determining the similarity of a set of values from an unknown sample to a set of values measured from a collection of known samples. MD proves to be superior to other multidimensional distance measures for the following reasons (Taguchi & Jugulum, 2002):

- Correlation between the variables is used in its calculation.
- It is very sensitive to intervariable changes in the reference data.
- It is not affected by the dimensionality of the dataset.

Assuming the dataset consists of k variables; i is the variable (i = 1, 2, …, k); n represents the number of samples in the dataset; and j is the sample number (j = 1, 2, …, n), the variables are standardized as defined in Equation (1).

$$z_{ij} = (x_{ij} - m_i)/s_i \tag{1}$$

where, m_i and s_i represent the mean and standard deviation of the ith variable, respectively; and z_{ij} is the standardized vector obtained from the standardized values of x_{ij}. MD values are calculated as defined in Equation (2).

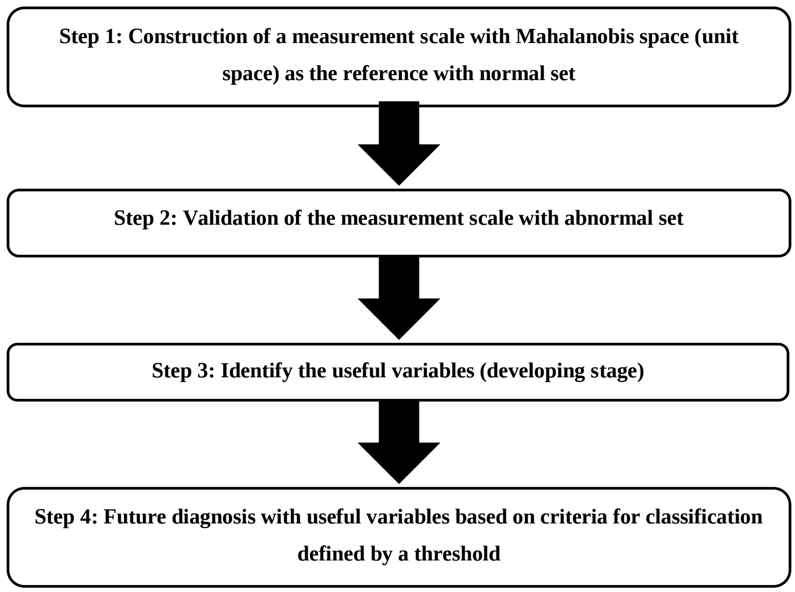
$$MD_j = \frac{1}{k}\boldsymbol{Z}'_{ij}\boldsymbol{C}^{-1}\boldsymbol{Z}_{ij} \tag{2}$$

where, MD_j is the Mahalanobis distance calculated for the jth case and $\boldsymbol{C}^{-1}$ represents the inverse of the correlation of the variables in the dataset.

Mahalanobis-Taguchi Strategy

The Mahalanobis-Taguchi strategy was developed by Genichi Taguchi as a diagnosis and forecasting method using multivariate data for robust engineering. It is a pattern recognition technology that assists in quantitative decision-making by constructing a multivariate measurement scale using data analytic procedures (Taguchi & Jugulum, 2002). MTS is used to develop a scale to measure the degree of abnormality of these measurements compared to “normal” measurements. The Mahalanobis distances for the attributes are calculated, then orthogonal arrays (OA) and signal-to-noise (S/N) ratio are used to identify attributes of importance. Exhibit 1 shows the steps for implementing MTS.

Exhibit 1. Steps for implementing MTS.



The criteria for classification are then defined as a threshold based on these useful attributes and the MD scale. In developing a multivariate measurement scale it is important to (1) have a reference point to the scale, (2) validate the scale, (3) select the important variables adequate for measuring abnormality, and (4) be able to carry out future diagnosis with the measurement scale. These form the basis of MTS implementation.

A Literature Review

In order to identify scientific literature about the use of MTS in the design of condition monitoring systems, a search of international papers published within the time interval of 2001 through 2013 was performed. What follows is a chronological review of these papers from the earliest to the most recent publications. The review is presented in two parts, research conducted within the first decade of the time interval and research conducted since 2010 follows.

Research Conducted from 2001 – 2010

In 2001, Asada used MTS to predict the yield in the production of wafers. The yield of a wafer refers to the ratio of non-defective chips compared to all chips in a wafer. Yield of wafers is determined by the variability of electrical characteristics and dust. MTS was applied to data collected on 20 electrical characteristics of wafers. With the application of the signal-to-noise ratio, Asada was able to identify six basic parameters of the 20 selected for prediction. Although the study was limited to one particular product during a certain period of time, the research confirmed the application of MTS in forecasting yield.

In 2004, Saraiva, Faísca, Costa, & Gonçalves applied a modified version of MTS, called the Modified Mahalanobis-Taguchi strategy (MMTS), to fault identification in chemical processes. Two approaches were identified for fault identification, model-based techniques and process history-based methods. MMTS was classified as a process history-based method. MMTS was applied to a continuous-stirred tank reactor (CSTR) simulated by Kano, Tanaka, Hasebe, Hashimoto & Ohno, in 2002, through a combined multivariate statistical process control (CMSPC) technique and the results were compared. CMSPC is an integration of a principal component analysis-based statistical process control (SPC) and an independent component analysis-based SPC. Saraiva et al. used the same monitored variables and data used by Kano et al., which covered normal operating conditions and eleven different abnormal conditions. The primary modification by Saraiva et al. to MTS was the application of multiple regression analysis (MRA). Steps one and two of the four step MTS implementation proposed by Taguchi and Jugulum (Taguchi & Jugulum, 2002) remained unchanged in the MMTS implementation. However, steps three and four were modified. For optimizing the system in step three, Saraiva et al. chose to use stepwise MRA for the selection of useful variables. Finally, for the fault diagnosis in step four, Saraiva et al. identified a threshold using the corresponding average run length (ARL) from Kano et al. Saraiva et al. stated that ARL was an accepted criterion more suited for evaluation using statistical process control (SPC) and fault diagnosis procedures. The results showed little or no difference between the ARL scores obtained from the CMSPC approach compared to the MMTS approach when all variables were used. More importantly, when a subset of the variables was used, the results showed similar performances. Saraiva et al. described MMTS as “a cheaper and efficient fault diagnosis system.” In addition, Saraiva et al. described MMTS as a promising data-based approach for on-line fault detection.

In 2005, Riho, Suzuki, Oro, Ohmi, & Tanaka applied a modified version of MTS, called MTS+, to diagnose the cause of invisible defects in order to enhance production yield in a wafer production process. The defects referred to as a white point (WP) were especially common in charge-coupled devices (CCD), which was the focus of the research. If a chip has a WP after the CCD process is completed then a failure has occurred; otherwise, the chip is said to be normal. In the implementation of MTS+, which combined several original techniques from the aspect of yield enhancement, Riho et al. tried to determine the degree of contribution from every process parameter to the WP failures observed. Riho et al. used MTS to identify the important variables contributing to the WP failures and, based on these variables, designed the experiments and carried out investigations. Riho et al. confirmed that the WP defects were connected with organic matter found on the chips after the CCD process and, subsequently, developed countermeasures to enhance the yield quicker and more accurately.

In 2006, Cudney, Paryani, & Ragsdell applied MTS to forecast consumer satisfaction ratings as related to vehicle handling. The research involved the application of MTS using the adjoint matrix approach (MDA) and the Gram-Schmidt approach, Mahalanobis-Taguchi Gram Schmidt (MTGS), on 72 data points collected over 21 vehicle handling parameters. MDA is used to signify that the Mahalanobis distances in the approach are obtained using the adjoint of the correlation matrix as opposed to the inverse the matrix. The adjoint matrix is used to address the issue of multicollinearity. MTGS applies the Gram-Schmidt orthogonalization process to increase the effectiveness of the forecasting process by identifying the direction of abnormality in an MTS implementation. Cudney et al. developed a breakthrough method for the identification of outliers before carrying out MTS analysis on the data. MD values were calculated for all the observed data and a threshold was determined using the quality loss function (QLF). This threshold was used prior to MTS implementation to identify four outliers in the data. MTS analysis as outlined by Taguchi and Jugulum (Taguchi & Jugulum, 2002) was then carried out on the 68 data points remaining. MDA identified 14 useful variables while

MTGS identified 15 useful variables. These variables were used separately in regression analysis and the resulting correlation between the actual and predicted consumer satisfaction ratings were 0.891 and 0.765 for MTGS and MDA, respectively.

In 2006, Cudney, Paryani, & Ragsdell applied MTS to forecast consumer satisfaction ratings as related to vehicle ride. In this research MTS was implemented on data from 67 vehicles over six ride parameters. Cudney et al. classified the data set into normal and test groups by calculating the MD values for the 67 data points. For the analysis, 61 data points were identified as normal and six data points were identified as abnormal. In line with MTS as described by Taguchi and Jugulum (Taguchi & Jugulum, 2002), OA and S/N ratio were used to optimize the system and five useful parameters were selected. The S/N ratio indicates the impact of each parameter to the system, the larger the ratio in the positive direction, the more important the parameter. However, regression analysis was carried out using the two most important parameters and a correlation coefficient of 0.864 was achieved. Cudney et al. also compared the results for MTS methods to those obtained from using neural networks and noted MTS has a higher accuracy with considerably less data.

In 2006, Miki & Okazawa applied chemical evaluation and MTS to the development of a technology for diagnosing the remaining service life for insulators, which in turn allows for determining the service life of power distribution equipment. The research methodology involved the application of MTS to determine the degree of deterioration in surface resistivity of an insulator which is obtained through chemical analysis. The threshold value used in MTS implementation was derived for the shape of the insulator and the deterioration over time for the insulator. Miki et al. used data from new insulators for the construction of the Mahalanobis space (MS). A trend in the deterioration over time in the surface resistivity of the insulator was determined by evaluating the relationship between elapsed years and the result of the deterioration diagnosis with MTS implementation. Miki et al. demonstrated that the remaining service life of an insulator can be predicted by determining the year in which the straight line connecting the surface resistivity for new insulators and old insulators intersect with the threshold value selected for the insulator. A correlation was also found between laboratory results of electrical discharge and findings from the research confirmed that the remaining service life can be successfully diagnosed using the proposed methodology.

In 2006, Itagaki, Takamiya, Watanabe, Nukaga, & Umemura applied a variant of MTS to the corrosion diagnosis of carbon steel in fresh water. By applying MTS to this problem, water quality affecting the corrosion of carbon steel was distinguished. For this research, data was collected on 17 environmental factors affecting corrosion of carbon steel in water as it is difficult to measure the actual corrosion rate of the material in a real world environment. First, data was collected in the normal “no corrosion” state to construct the MS. Then MTS was implemented as described by Taguchi and Jugulum (Taguchi & Jugulum, 2002) by applying distinction data for validation of the MS collected for both normal and abnormal cases. Three useful variables were identified for the diagnosis based on water quality. A threshold MD value of three was selected for MTS implementation and the MS was adjusted to compensate for the difficulty in collecting significant data for the normal case with no corrosion. Itagaki et al. changed the mean value for the calculation of the MS, which in effect created an artificial MS suitable to the nature of the data collected and the nature of the research. The standard deviation for the calculation of the MS was also changed; this corresponds to changing the size of the MS to further augment the artificial MS. The resulting MS was one, which was established on scientific corrosion theory. Itagaki et al. selected four useful variables based on knowledge of the factors that significantly influence corrosion and compared the diagnosis on the artificial MS, the original MS, and an experimental method. The results showed the original MTS to have wrongfully classified 15 out of 23 samples tested. However, the comparison of MTS with the artificial MS reduced the misclassifications to three samples. Itagaki et al. concluded the MTS method is useful for corrosion diagnosis but the MS must be customized depending on material and environment.

In 2007, Miki, Okazawa, Hasegawa, Tsunoda, & Inujima conducted further research to improve the accuracy of earlier research (Miki & Okazawa, 2006) and expand the range of insulators covered by researching insulators of circuit breakers. The methodology was also applied to phenol insulators (Miki, Hasegawa, Umemura, Okazawa, Otsuka, & Inujima, 2007) and phenol insulators for circuit breakers (Miki, Hasegawa, Umemura, Okazawa, Otsuka, Matsuki, Tsunoda & Inujima, 2008).

In 2008, Rai, Chinnam, & Singh utilized MTS analysis for online prediction of drill-bit failure (breakage) from two degradation signals, thrust force and torque, during a drilling operation. Rai et al. described the advantage of MTS to other online tool-condition monitoring methods as the flexibility of the methodology, as it allows monitoring multiple features simultaneously and the selection of useful features. Ten features were monitored for both degradation signals over 128 drilled holes. The data was collected from running nine drill-bits until breakage. Data collected from the last hole successfully drilled by a drill-bit was defined as belonging to the abnormal class. All data collected prior to that was classified as being in the normal operation set. Rai et al. acknowledged the fact that an on-set of the drill-bit degradation might occur well in advance to the last hole but stated that this method allowed for the maximum usage of tool life. MTS was implemented on the collected data as described by Taguchi and Jugulum (Taguchi & Jugulum, 2002). In the research, five useful features were identified and used to successfully predict drill bit failure based on a threshold value of the resulting MD values from these features.

In 2008, Mohan, Saygin, & Sarangapani developed an MTS-based real time diagnostics and root cause analysis tool to diagnose the quality of fastening operations for a hand-held pull-type pneumatic tool and specify the cause of the failure. For the research four characteristics were measured including peak strain, peak displacement, and depth and width of a bowl-shaped dip on the process signature. The data was collected wirelessly and fed to the system to make real-time decisions on the grip length of the fastening operation. In addition to being used to identify failures (deviations in grip length), MTS was used for root cause analysis. Mohan et al. reproduced each abnormality and calculated the MD values with respect to data from the normal operating condition. For instances where there were a number of similar cases of abnormalities, an MD range corresponding to this set was determined with respect to the ideal case. Signatures were analyzed from these ranges using a correlation matrix and the MD value was calculated. A fault was characterized by which range the MD value fell under and the type of abnormality was determined. The MTS tool selected two of the four characteristics as important and had a detection rate of 87.5%, 100%, and 96.8% on over grip, normal grip, and under grip, respectively.

In 2009, Hu, Zhang, Liang, & Wang developed an incipient mechanical fault detection method based on multifractal and MTS methods. Multifractal features of vibration signals were obtained from machine state monitoring that were extracted by multifractal spectrum analysis and generalized fractal dimensions. Through multifractal analysis generalized dimensions of the three mechanical running states were obtained. MTS was applied to optimize feature selection for different mechanical running states, based on which incipient faults were identified and diagnosed. The experiments covered fault detection in oil pumps and nine multifractal features were monitored. Hu et al. were able to reduce this to seven important features. The method was tested on Mahalanobis distances of 44 groups of observational signals over the three states identified including normal state, race wear of a rolling bearing, and air clearance in sliding. The MD thresholds were selected for each state and the results showed accuracy of 100%.

In 2009, Jeong, Park, Yang, Lee, & Oh applied MTS to fault diagnosis on rotating machinery by analyzing the vibration signals through signal processing. A rotor kit was used as a case study for the experiment. Vibration analysis was performed on the kit over steady and unbalanced states. The data obtained was analyzed using diagnosis techniques to determine seven representative variables. Application of MTS to the data identified two variables of importance, which were used to effectively diagnose steady state and unbalanced state operation of the machinery.

In 2010, Yang & Cheng applied MTS to improve inspection efficiency for the flip-chip bumping height of dies in chip manufacturing processes. Data was collected on two inspection positions over five areas on each for a total of ten features on each wafer. SPC enabled the selection of average bump heights of dies on wafers based on the normal condition which fell within two categories of 2σ and 3σ control limits. Data collected from these wafers in normal condition was used to create two Mahalanobis spaces, MS1 and MS2. MS1 was constructed using the average bump heights of ten dies that were 3σ. MS2 was constructed using the average bump heights of ten dies of a wafer within 2σ. Rather than use the loss function in determining the threshold as described by Taguchi and Jugulum (Taguchi & Jugulum, 2002), a bisection algorithm was used to determine the threshold value; which satisfies the requirement that the threshold value be the point where the losses due to the two types of mistakes are balanced (Taguchi, Chowdhury & Wu, 2001). The important features identified based on the scales from MS1 and MS2 were reduced to five and six, respectively. The results showed that the number of inspection positions on a wafer could be reduced from ten to two without significantly reducing the classification accuracy (99%) when using the MS constructed with MS1. As a result, the inspection time of the bump process was reduced from 100-150 to 20-30 seconds per wafer. Yang et al. also showed that using the MS constructed with MS2 reduced the inspection positions on a wafer from ten to six without affecting the classification accuracy (100%). In addition, the inspection time of the bump process was reduced from 100-150 to 60-90 seconds per wafer.

In 2010 Soylemezoglu, Jagannathan & Saygin applied an MTS based fault prognostics system to rolling element bearing failures. The system detects a fault, diagnoses its root cause (fault isolation), and estimates the remaining useful life or time to failure which completes the prognosis. MD values were calculated for the variables being monitored and thresholds were set for the normal operation condition. The three types of faults considered in the experiment included cage defect, inner race defect, and outer race defect. Fault detection occurred when the MD exceeded the normal operation range. The specific fault could be determined by identifying the specific threshold band in which the MD value falls. To complete the fault prognosis, the MD values are calculated using a predetermined time window as the bearing was being monitored. Fault detection occurred once the MD crossed the specified threshold and the tracking of the MD trend is initiated. By tracking the direction of the transition of the MD trend and by calculating the angle between the MD point and the mean MD of the three known fault clusters, a root cause was identified. The smaller the angle the more likely it is that the fault is progressing towards one of the fault clusters. The prognosis of the time to failure was calculated via linear approximation. Initially ten features were selected to construct the MS which reduced to eight used in the prognosis. Soylemezoglu et al. achieved a 100% success rate in correct detection and isolation of bearing faults with this tool.

Research Conducted from 2011 – 2013

In 2011, Soylemezoglu, Jagannathan & Saygin applied a comprehensive fault monitoring tool based on MTS to centrifugal pump failures. The methodology was modified by including cluster analysis for better identification of threshold values and for identifying the optimum number of sensors required for the condition monitoring. MTS was used to fuse data from multiple sensors on the centrifugal pump into a single system level performance metric using MD. Cluster analysis was used to create fault clusters based on the MD values generated. Thresholds determined from the clustering analyses were used to detect and isolate faults. To complete the fault prognosis, the MD values were calculated using a predetermined time window and linear approximation was used to estimate the time to failure. The experiment used 18 parameters measured from a 1/2 HP centrifugal pump operated for 150 hours. The experiment investigated three types of failures including seal failure, impeller failure, and filter clog. A high success rate was achieved on all three faults in the research.

In 2011, Yoneda applied MTS to failure diagnosis on check-out of space system launch operation. The failure model of an electric actuator that drives a nozzle on a launch vehicle was used as a case study. Data was collected on the operation of the actuator at the nozzle from potentiometers that measure nozzle motion. Two types of failures were observed including Error 1 and Error 2. An MTS implementation was carried out on waveform data collected from the nozzle over the normal operating conditions and the two stages of failure to illustrate the effectiveness of the methodology.

In 2011 Kumano, Mikami, & Aoyama discuss the Mitsubishi Heavy Industries, Ltd. remote condition monitoring for gas turbines based on MTS. The system uses MTS to identify abnormalities in gas turbine operations and implement an artificial intelligence technique based on Bayesian network model for root cause analysis. The remote monitoring system (RMS) extracts plant operation data from gas turbines located around the globe, sampled at one minute intervals. The RMS collects up to 2000 data points on each monitoring cycle from a gas turbine. The key to abnormality diagnosis in such a complex system is early detection to protect the equipment from damage. However, detecting such small changes are difficult by monitoring a broad range of parameters, which are affected by atmospheric conditions and operation condition. For abnormality diagnosis the RMS evaluates two types of variations. The first type is a variation that exceeds a predetermined limit which the RMS applies using a standard alarm threshold method. The second type is a variation in the form of a deviation from the normal relationship between correlative parameters referred to as trend monitoring. The RMS applies MTS to trend monitoring. MTS enables the detection of small changes in the patterns of parameters being monitored long before an alarm is generated, thereby preventing severe equipment damage. MTS converts all the parameters being monitored into one index which is used in the diagnosis. For each gas turbine the correlation between 150 parameters are observed for an operation pattern. MTS combines all 150 parameters to an MD value. When a sample pattern is diagnosed as “abnormal”, the S/N ratios of each variable used for the calculation of the MD are estimated and the major parameters causing the large MD value are identified. A root cause analysis is determined by the artificial intelligence technique based on the Bayesian network model.

In 2011, Ren, Cai, & Xing applied MTS to data obtained from the Hilbert-Haung transform on vibrating signals for mechanical fault diagnostics. The Hilbert-Haung transformation was used to extract characteristics of the vibration signals relevant to the fault diagnosis. Data was collected on the three stages of operations of the gas turbines including normal operation, slight defect, and severe defect stages. Ren et al. applied MTS to the data and derived appropriate threshold values for each stage of operation. The research used real time data obtained from the gas turbines during operation for diagnosis.

In 2011, Lv & Gao applied a multifractal-MTS based system to predict the condition of a chemical industry complex system. Multifractal analysis was applied to the data collected from the numerous condition characteristics being monitored on a compressor system in a chemical plant to extract suitable features for the experiment. MTS was utilized to distinguish the important features of multiple variables used in the prediction task. The research utilized twelve multifractal features extracted from three variables which were collected from the air compressor system on the plant. MTS identified six as optimal features, which were used successfully to predict abnormal operating conditions of the compressor system.

In 2012, Wang, Wang, Tao, & Ma used MTS for fault diagnosis of a rolling bearing element after feature extraction by time/frequency domain analysis from vibration data. Data was collected over four states of the rolling bearing element including normal condition, inner race fault, outer race fault, and rolling element fault. Initially a set of eight characteristics were extracted via the time/frequency domain analysis. MTS was used to identify six useful variables used in the diagnosis.

In 2012, Okazawa & Miki applied MTS to the diagnosis of switch gear insulator degradation and estimation of remaining service life. This technique enabled the diagnosis of the state of the insulator itself as opposed to other methods which measured electrical effects of insulator degradation that varied greatly with the weather and time of measurements. Okazawa et al. measured chemical properties of the insulators and applied MTS to the data. By evaluating the correlation of the data with surface resistivity, the surface resistivity of degraded insulators under consistent weather conditions was determined. Using the humidity characteristics equation it was also possible to calculate the surface resistivity for any humidity value. Degradation to the level at which an electrical discharge occurs during an insulator service life defines the insulator lifetime. Based on this theory a threshold was established that enabled an estimate of the remaining service life of the insulator to be obtained.

In 2012, Liparas, Angelis, & Feldt combined MTS with a cluster analysis technique to determine the best training set for software defect diagnosis. MTS was used to build a measurement scale for detecting faulty software modules; and a cluster analysis was used to aid MTS implementation by selecting the most appropriate defect-free software modules for creating the MS. Liparas et al. observed that the MD values obtained for the normal (defect-free) and abnormal (defect) group had overlapping regions which would negatively affect the prognostic outcome of the MS and determining a threshold for the classification would be difficult. To overcome this, a two-step cluster analysis algorithm was applied to the data set to effectively separate the normal and abnormal groups. The cluster analysis minimized the problem of misclassification during prediction by ensuring that the data set used to construct the MS contained the best representation of the homogeneous defect-free software modules. Liparas et al. applied this methodology to ten data sets over four programming languages, including C, Java, C++, and Perl. A different number of features were monitored on each dataset. The accuracy of the prediction on each dataset increased after the application of the two-step cluster analysis. When compared to the performance of 22 other classifiers on the same ten datasets overall, the MTS-cluster analysis hybrid had a higher accuracy than all of the other methods on all but two datasets.

In 2013, Hu, Zhang, & Liang presented a dynamic degradation observer for identifying and assessing degradation of bearings using a hybrid of MTS and self-organization mapping (SOM). The MTS-SOM system proved useful in capturing the onset of bearing failure at the second stage of degradation, the incipient fault stage, where the fault information is very weak and difficult to discover. The methodology enables tracking of the dynamic degradation trend of a running bearing through real-time bearing observations. Hu et al. applied multifractal analysis to obtain the fractal features of the bearing over time. MTS was applied to the features obtained from the bearing experiment to select the optimum features contributing to the degradation identification. In an unsupervised learning implementation, an SOM algorithm via neural network was applied to enable the visualization of the degradation trajectory. Nine multifractal features were initially collected over 50 samples and MTS was used to optimize the system by reducing the number of features to seven. MTS was used to classify the bearing under four bearing degradation stages including normal condition, incipient fault, severe degradation, and complete failure. Data collected on the optimal features was used to train the SOM algorithm for the degradation assessment mechanism and the results created different clusters made up of the same degradation stage. Finally, by plotting the trajectory of the data in real time, the entire bearing degradation process can be followed over time. Hu et al. obtained an overall classification accuracy of 85.71% over the entire experiment.

Conclusion

This paper has reviewed literature on the application of MTS in condition monitoring systems over the past decade. Although the paper focuses on condition monitoring, there have various application of MTS in medical diagnosis, ranking systems, and manufacturing inspection, to name just a few. The literature reviewed has shown that

MTS is a valuable tool for cost effective condition monitoring in various fields. Without knowledge of the parameters related to the fault under observation the approach would be to collect data on all parameters and analyze this data. This is an expensive approach. MTS is a more effective methodology; it reduces the dimensionality of condition monitoring problems by identifying the variables correlated to the factor being monitored. By eliminating the need to monitor the entire set of parameters of the process or equipment, the cost of the condition monitoring system is significantly reduced. Selecting the threshold is an important aspect of MTS and this review has shown various approaches to determining the right threshold for effective condition monitoring. The literature reviewed has also shown that MTS consistently achieves a high level of accuracy with fewer features. MTS is a flexible tool which can be easily modified or combined with other tools and techniques as required by the peculiarities of the problem. It was demonstrated that knowledge of the process is important and can lead to more specialized applications by selecting specific useful variables based on scientific theory and knowledge; and by modifying the procedure for constructing the MS (Itagaki, Takamiya, Watanabe, Nukaga, & Umemura, 2006).

References

Asada, M. (2001). Wafer yield prediction by the mahalanobis-taguchi system. Paper presented at the *International Workshop on Statistical Metrology, Proceedings, IWSM,* 25-28.

Cudney, E. A., Paryani, K., & Ragsdell, K. M. (2006). Applying the mahalanobis-taguchi system to vehicle handling. Paper presented at the *Proceedings of the ASME Design Engineering Technical Conference.*

Cudney, E. A., Ragsdell, K. M., & Paryani, K. (2006). Applying the mahalanobis-taguchi system to vehicle ride. Paper presented at the *American Society of Mechanical Engineers, Design Engineering Division (Publication) DE.*

Hu, J., Zhang, L., Liang, W., & Wang, Z. (2009). Incipient mechanical fault detection based on multifractal and MTS methods. *Petroleum Science, 6*(2), 208-216.

Hu, J., Zhang, L., & Liang, W. (2013). Dynamic degradation observer for bearing fault by MTS-SOM system. *Mechanical Systems and Signal Processing, 36*(2), 385-400.

Itagaki, M., Takamiya, E., Watanabe, K., Nukaga, T., & Umemura, F. (2006). Diagnosis of water quality for carbon steel corrosion by mahalanobis-taguchi method. *Zairyo to Kankyo/ Corrosion Engineering,55*(7), 300-305.

Jeong, J.-E., Park, S.-G., Yang, I.-H., Lee, Y.-Y., & Oh, J.-E. (2009). A fault diagnosis on rotating machinery using the mahalanobis taguchi system. Paper presented at the 16th International Congress on Sound and Vibration 2009, ICSV 2009, 5 2967-2973.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., Ohno, H. (2002) *Proceedings of International Symposium on Design, Operation and Control of Chemical Plants (PSE Asia 2002).*

Kumano, S., Mikami, N., & Aoyama, K. (2011). Advanced gas turbine diagnostics using pattern recognition. Paper presented at the Proceedings of the ASME Turbo Expo, 3 179-187.

Liparas, D., Angelis, L., & Feldt, R. (2012). Applying the mahalanobis-taguchi strategy for software defect diagnosis. Automated Software Engineering, 19(2), 141-165.

Lv, Y., & Gao, J. (2011). Condition prediction of chemical complex systems based on multifractal and mahalanobis-taguchi system. Paper presented at the ICQR2MSE 2011 - Proceedings of 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, 536-539.

Miki, S., & Okazawa, H. (2006). Deterioration diagnosis of insulators for breakers using chemical evaluation and mahalanobis-taguchi (MT) method. Paper presented at the Annual Report - Conference on Electrical Insulation and Dielectric Phenomena, CEIDP, 473-476.

Miki, S., Hasegawa, T., Umemura, S., Okazawa, H., Otsuka, Y., & Inujima, H. (2007). Remaining service life diagnostic technology of phenol insulators for power distribution equipment. Paper presented at the Annual Report - Conference on Electrical Insulation and Dielectric Phenomena, CEIDP, 41-44.

Miki, S., Okazawa, H., Hasegawa, T., Tsunoda, S., & Inujima, H. (2007). Remaining service life diagnostic technology of insulators for power distribution equipment. IEEJ Transactions on Power and Energy,127(7), 863-869+13.

Miki, S., Hasegawa, T., Umemura, S., Okazawa, H., Otsuka, Y., Matsuki, H., Tsunoda, S., & Inujima, H. (2008). Remaining service life diagnostic technology of phenol insulators for circuit breakers. IEEE Transactions on Dielectrics and Electrical Insulation, 15(2), 476-483.

Mohan, D., Saygin, C., & Sarangapani, J. (2008). Real-time detection of grip length deviation during pull-type fastening: A mahalanobis-taguchi system (MTS)-based approach. International Journal of Advanced Manufacturing Technology, 39(9-10), 995-1008.

Okazawa, H., & Miki, S. (2012). Estimating the remaining service life of switchgear insulators: Application of the mahalanobis-taguchi (MT) method to diagnosis of degradation and estimation of the remaining service life of insulators. Paper presented at the Proceedings of 2012 IEEE International Conference on Condition Monitoring and Diagnosis, CMD 2012, 317-319.

Rai, B.K., Chinnam, R.B. & Singh, N. (2008) ‘Prediction of drill-bit breakage from degradation signals using Mahalanobis-Taguchi system analysis’, Int. J. Industrial and Systems Engineering, Vol. 3, No. 2, pp.134-148.

Ren, J., Cai, Y., & Xing, X. (2011). Application of hilbert-huang transform and mahalanobis-taguchi system in mechanical fault diagnostics using vibration signals. Paper presented at the Proceedings - IEEE 2011 10th International Conference on Electronic Measurement and Instruments, ICEMI 2011, 4 299-303.

Riho, T., Suzuki, A., Oro, J., Ohmi, K., & Tanaka, H. (2005). The yield enhancement methodology for invisible defects using the MTS+ method. IEEE Transactions on Semiconductor Manufacturing, 18(4), 561-568.

Saraiva, P., Faísca, N., Costa, R., & Gonçalves, A. (2004). Fault identification in chemical processes through a modified mahalanobis-taguchi strategy.

Soylemezoglu, A., Sarangapani, J., & Saygin, C. (2010). Mahalanobis taguchi system (MTS) as a prognostics tool for rolling element bearing failures. Journal of Manufacturing Science and Engineering, Transactions of the ASME, 132(5)

Soylemezoglu, A., Sarangapani, J., & Saygin, C. (2011). Mahalanobis-taguchi system as a multi-sensor based decision making prognostics tool for centrifugal pump failures. IEEE Transactions on Reliability, 60(4), 864-878.

Taguchi G., Chowdhury S. & Wu Y. (2001). The Mahalanobis–Taguchi System. New York: McGraw-Hill; 2001.

Taguchi, G. & Jugulum, R. (2002). The Mahalanobis-Taguchi Strategy, New York: John Wiley & Sons, Inc.

Wang, Z., Wang, Z., Tao, L., & Ma, J. (2012). Fault diagnosis for bearing based on mahalanobis-taguchi system. Paper presented at the Proceedings of IEEE 2012 Prognostics and System Health Management Conference, PHM-2012.

Woodall, W. H., Koudelik, R., Tsui, K. –L., Kim, S. B., Stoumbos, Z. G., & Carvounis, C. P. (2003). A review and analysis of the mahalanobis-taguchi system. *Technometrics, 45*(1), 1-15.

Xue, Y., Williams, D. P., & Qiu, H. (2011). Classification with imperfect labels for fault prediction. Paper presented at the *Proceedings of the 1st International Workshop on Data Mining for Service and Maintenance, KDD4Service 2011 - Held in Conjunction with SIGKDD'11*, 12-17.

Yang, T., & Cheng, Y.-T. (2010). The use of mahalanobis-taguchi system to improve flip-chip bumping height inspection efficiency. *Microelectronics Reliability*, *50*(3), 407-414.

Yoneda, Y. (2011). Intelligent diagnostics based on the mahalanobis taguchi method for space systems. Paper presented at the *62nd International Astronautical Congress 2011*, *IAC 2011*, *9* 7064-7068.

PAPER II

Predicting faults in heavy duty vehicles using the Mahalanobis-Taguchi Strategy

Adebolaji A. Jobi-Taiwo, Elizabeth A. Cudney
Missouri University of Science and Technology
Rolla, MO 65409, US

Robert S. Woodley
21st Century Systems, Inc.
Omaha, NE 68132, US

Abstract

This paper presents a Mahalanobis-Taguchi Strategy (MTS) based system for predicting faults in heavy duty vehicles. Costs associated with heavy duty vehicle breakdown in a large fleet while in operation can be significantly reduced if faults leading to these breakdowns are predicted and prevented. Fifty-one attributes on the vehicles are monitored in real-time and the data fed into the system. MTS is used to develop a scale to measure the degree of abnormality of these measurements from the vehicles compared to "normal" measurements. The Mahalanobis distances (MD) for the attributes are calculated, orthogonal arrays (OA) and signal-to-noise (S/N) ratio are used to identify attributes of importance. By reducing the dimensionality, less attributes are tracked which reduces the cost of the system. Criteria for classifying fault measurements are defined based on these variables of importance and the MD scale. The result is a real-time monitoring system that predicts faults in the vehicles thereby preventing breakdowns during operation. The information obtained can also assist in creating an effective preventive maintenance schedule for the vehicles in the fleet.

Keywords: Mahalanobis-Taguchi Strategy, Orthogonal Arrays, Signal-to-Noise ratio

1. Introduction and Background

Fault prediction is a useful concept in condition monitoring of equipment. The premise of fault prediction is that knowledge about the timing of a fault or defect occurrence can be used to improve efficiency of machine operation and reduce maintenance cost. Fault prediction involves monitoring system operation conditions with the goal of predicting impending failure occurrences based on symptoms exhibited in sensor-measured process variables and is vital in preventing costly component damage or catastrophic failure of equipment [1]. Data gathered from equipment being monitored often consists of multiple multivariate time series [1] and is fed to a system or model that applies some data analysis technique to complete the prediction task. Fault prediction systems based on various data analysis techniques have been applied to different problems. One of the data analysis techniques used is the Mahalanobis-Taguchi Strategy (MTS). MTS has been used in a number of fault diagnosis and prognosis systems.

According to Cudney et al. [2], "The Mahalanobis-Taguchi System is a diagnosis and predictive method for analyzing patterns in multivariate cases." MTS is a pattern recognition tool that develops a multidimensional measurement scale through the integration of mathematical and statistical concepts such as Mahalanobis distance (MD) with the principles of robust engineering (Taguchi Methods) [3]. MTS allows for a reduction in dimensionality and the development of a scale based on the Mahalanobis distance by which abnormality can be measured. MD is a generalized measure of a distance representing the degree of divergence in the mean values of different characteristics of a population considering the correlation between the variables [3]. It is a measure for describing the distance of a data point from the mean of a multivariate population. In implementing MTS, the MD is calculated to create a scale for discriminating between the normal and abnormal measurements in a multivariate dataset, and then orthogonal arrays (OA) and signal-to-noise (S/N) ratios are used to evaluate the contribution of each variable and select the useful set of variables [4]. The MD scale created from the selected variable can then be used for diagnosis or prediction.

Lv and Gao [5] applied a Multifractal-MTS based system to predict the condition of a chemical industry complex system. Multifractal analysis was applied to the dataset to extract nonlinear features and MTS was utilized to distinguish important features of multiple variables used in the prediction task. Wang et al. [6] used MTS for fault diagnosis of a rolling bearing element after feature extraction by time/frequency domain analysis. A set of eight characteristics were extracted and MTS was used to identify six useful variables used in the diagnosis. Hu et al. [7] carried out incipient machine degradation assessment with a Multifractal-MTS system through a case study on vibration measurements of rolling bearings. This research proved to be a comprehensive tool for machine condition monitoring management in both current and predictive analysis of fault degradation behavior. Multifractal analysis was applied for the extraction of nine useful features from the vibration data. MTS was used to identify the optimal feature set which constituted the Mahalanobis space and were used for incipient fault diagnosis.

Soylemezoglu et al. [8] applied an MTS based fault prognostics system to rolling element bearing failures. The system detects a fault, identifies its root cause (fault isolation), and estimates the remaining useful life or time to failure. MD values were calculated for the variables being monitored. Fault detection occurred when the MD exceeds the normal operation range. By tracking the direction of the transition of the MD trend, a root cause was identified and a prognosis of the time to failure was calculated via linear approximation. Soylemezoglu et al. [9] then applied the MTS based fault prognostics system to centrifugal pump failures.

Rai et al. [10] utilized MTS analysis for online prediction of drill-bit failure (breakage) from two degradation signals, thrust force and torque, during a drilling operation. Ten features each were monitored for both degradation signals. The data was collected from running nine drill-bits until their breakage. Five useful features were identified and used to successfully predict drill bit failure.

The objective of this research is to develop an MTS based system for predicting faults in heavy duty vehicles, in real-time. MTS is used to extract the useful features from the fifty-one attributes monitored from the vehicles and accurately predict vehicle faults.

2. Methodology

In this paper, an MTS-based fault prediction system is presented. The proposed system utilizes an MD-based measurement scale to assess the degree of abnormality in the variables being monitored by measuring the distance of each observation from a reference group (normal group) also known as the Mahalanobis space (MS).

2.1 Mahalanobis Distance

MD was introduced by Prasanta Chandra Mahalanobis in 1936 [3]. MD is a distance measure derived from an analysis of the deviation in the mean values of different variables in multivariate analysis considering the correlation between the variables. MD, as a discriminant analysis method, is useful in determining the similarity of a set of values from an unknown sample to a set of values measured from a collection of known samples. MD proves to be superior to other multidimensional distance measures [3] since:

- Correlation between the variables is used in its calculation.
- It is very sensitive to intervariable changes in the reference data.
- It is not affected by the dimensionality of the dataset.

Assuming the dataset consists of k variables; i is the variable ($i = 1, 2, \ldots, k$); n represents the number of samples in the dataset; and j is the number of sample ($j = 1, 2, \ldots, n$). The variables are standardized as defined in Equation (1).

$$z_{ij} = (x_{ij} - m_i)/s_i \tag{1}$$

where, m_i and s_i represent the mean and standard deviation of the *ith* variable, respectively; and z_{ij} is the standardized vector obtained from the standardized values of x_{ij}.

MD values are calculated as defined in Equation (2).

$$MD_j = \frac{1}{k}\boldsymbol{Z}'_{ij}\boldsymbol{C}^{-1}\boldsymbol{Z}_{ij} \tag{2}$$

where, MDj is the Mahalanobis distance calculated for the *jth* case and $\mathbf{C}^{-1}$ represents the inverse of the correlation of the variables in the dataset.

2.2 Mahalanobis-Taguchi System

MTS was developed by Genichi Taguchi as a diagnosis and forecasting method using multivariate data for robust engineering. MTS is a pattern recognition technology that assists in quantitative decision-making by constructing a multivariate measurement scale using data analytic procedures [3]. In developing a multivariate measurement scale it is important to (1) have a reference point to the scale, (2) validate the scale, (3) select the important variables adequate for measuring abnormality, and (4) be able to carry out future diagnosis with the measurement scale. These form the basis of the MTS implementation. The steps of MTS are outlined below.

Step 1. Construction of the measurement scale

In this step variables that define the "normal" and "abnormal" conditions are identified. Data is then collected on these variables and constitutes the dataset. Values of the variables are standardized (Equation 1) and MD values are calculated for the healthy group using the inverse of the correlation matrix (Equation 2). These MDs are used to define the Mahalanobis space.

Step 2. Validation of the measurement scale

The next step is to calculate MD values for the observations that belong to the "abnormal" group The MD calculation for the "abnormal" group utilizes the mean and standard deviation of the corresponding variable from the "normal" group for standardization. Also, the correlation matrix of the "normal" group is used for the MD calculation. The premise behind this is that since the data from the abnormal group is from the same system then it should be evaluated along with the "normal" cases. If the scale is correct, the MD values for the "abnormal" group should have higher values compared to the "normal" group. This validates the measurement scale.

Step 3. Identify useful variables

For this step, the aim is to identify the optimum number of variables required for the measurement scale. Variables that do not significantly contribute to the measurement scale are discarded. Orthogonal arrays and signal-to-noise ratios are used for this segregation. An orthogonal array is an experimental design matrix used to list the combination of characteristics and enables testing the effects of the absence or presence of a characteristic in experimentation. The size of the OA to be used is determined by the number of variables. For each run of the OA, the MD values for the "abnormal" cases are calculated using only the included variables as specified by the OA. The resulting MD values are then used to calculate a dynamic signal-to-noise ratio as shown in Equation (3).

$$S/N = 10log_{10}\frac{\frac{1}{r}(S_\beta - V_e)}{V_e} \tag{3}$$

where,

S_T = total sum of squares ($\sum_{i=1}^{t} y_i^2$)

r = sum of squares due to the input signal ($2\sum_{i=1}^{t} M_i^2$)

S_β = sum of squares due to slope ($\frac{1}{r}\sum_{i=1}^{t}(M_i y_i)^2$)

S_e = error sum of squares ($S_T - S_\beta$)

V_e = error variance ($\frac{S_e}{t-1}$)

The gain for each variable is then calculated by subtracting the average S/N ratio when the variable was excluded from the average S/N ratio when the variable was included as shown in Equation (4).

$$Gain = (Avg.S/N\ Ratio)_{variable\ present} - (Avg.S/N\ Ratio)_{variable\ absent} \tag{4}$$

The variables with positive gains are selected for the prediction and the rest are discarded.

Step 4. Reconstruction of Scale

In the final step, the measurement scale is then reconstructed with the useful set of variables identified in Step 3. Subsequently, the reconstructed scale is used to monitor the conditions of the system in question after a threshold for the boundary between the normal and abnormal cases has been specified.

3. Fault Prediction in Vehicles using MTS

MTS implementation begins with data collection on normal observations. The data collected from the vehicles during normal operation was presented in a dataset containing 51 variables obtained through sensors on the vehicles.

Data collection is always with the inclusion of outliers. Outliers are observations that are inconsistent with the general trend of the data set and may skew the final result of the experiment. Outliers in the data set were identified and removed by calculation of the MD for all the variables in the dataset. The first variable had a negative MD and was identified as an outlier and removed prior to carrying out the MTS analysis. Variables 22, 23, 24, 25, and 26 were also eliminated as they recorded information on time and date, which was irrelevant to the current analysis. Variable 27 was also eliminated as it was recorded once every 10 minutes and was considered inconsequential to the analysis. Variable 32 was also eliminated as it recorded vehicle data that returned the same value over each sample instance. The remaining 43 variables were used for the MTS analysis.

The variables were obtained at different frequencies as shown in Table 1. In order to carry out a correct implementation of MTS the sample rate of the variables had to be made uniform. After the removal of nine variables, the highest frequency for any of the variables left was 50Hz and in effect these variables had the highest number of samples. Data from all the other variables were upsampled to 50Hz by a factor calculated as shown in Equation (5).

$$Upsampling\ factor = \frac{Output\ frequency}{Input\ frequency} \tag{5}$$

where, output frequency represents the desired sample rate and the input frequency, the variable's original frequency.

The resulting dataset was truncated at the length of the variables with the shortest number of samples which was 1,3200,000 instances. This was used to create a matrix of size 1,320,000 by 43 for the MTS calculation.

To create a normal and abnormal set for the MTS calculation, the MD values for each sample in the dataset were calculated. The maximum MD value was 2874.405; and 1,313,221 instances (99.49%) had MD values less than 5.0. After analyzing the MD values cluster, a threshold of 30 was selected to divide the data into normal and abnormal sets. If the MD value of an instance was less than the threshold, it was classified in the normal group. If the MD value was greater than the threshold it was classified in the abnormal group. 650 instances were classified as abnormal according to this threshold.

The next step is to create the Mahalanobis space with the samples from the normal group. The MD value for each sample of the normal group was calculated and this formed the MS. The MD values for the abnormal group were also calculated to validate the MS. The maximum MD value for the normal group was 60.60. The average MD value of the abnormal group was 18,906,000, with a maximum value of 1,150,600,000 and a minimum value of 87.16. Thus, the scale was validated as the MD values corresponding to the abnormal group had higher values than the MD values for the normal group.

Next, the system is optimized by identifying the useful variables. An L64 orthogonal array was used for the experimentation. The useful variables were identified by calculating the gain associated with each factor as outlined in Equations (3) and (4). The gains for the 43 variables are shown in Figure 1. Twenty-three variables with positive gains were selected as useful variables.

A Mahalanobis scale is then constructed with the 23 variables identified as useful variables. This MS is to be used for diagnosis and prediction.

Table1: List of attributes obtained from the vehicles and the frequencies

Serial Number	Variable	Frequency (Hz)
1	2	1
2	3	1
3	4	1
4	5	1
5	6	1
6	7	50
7	8	50
8	9	50
9	10	10
10	11	10
11	12	10
12	13	10
13	14	10
14	15	50
15	16	50
16	17	50
17	18	50
18	19	50
19	20	50
20	21	50
21	28	1
22	29	1
23	30	1
24	31	1
25	33	5
26	34	2
27	35	1
28	36	5
29	37	0.01667 (1/60)
30	38	0.01667 (1/60)
31	39	0.01667 (1/60)
32	40	0.01667 (1/60)
33	41	0.01667 (1/60)
34	42	0.01667 (1/60)
35	43	0.01667 (1/60)
36	44	0.01667 (1/60)
37	45	0.01667 (1/60)
38	46	0.01667 (1/60)
39	47	0.01667 (1/60)
40	48	0.01667 (1/60)
41	49	0.01667 (1/60)
42	50	0.01667 (1/60)

Table1: List of attributes obtained from the vehicles and the frequencies (cont.)

Serial Number	Variable	Frequency (Hz)
43	51	0.01667 (1/60)

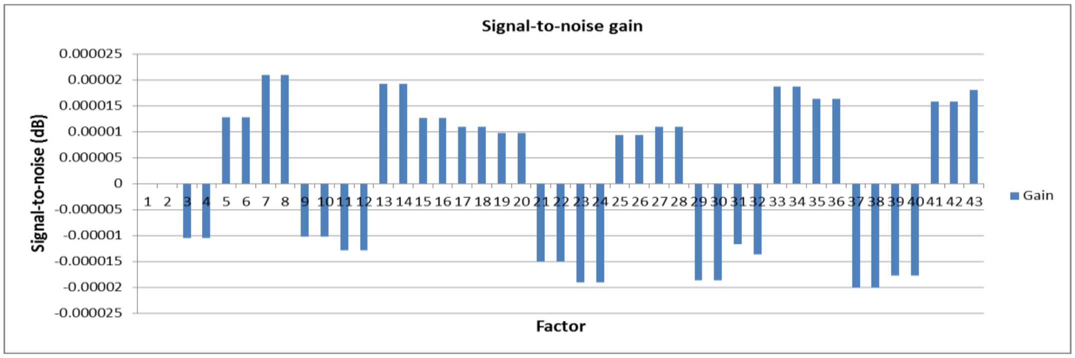


Figure 1: Gains on all variables

The new MS constructed with the useful variables successfully reduced the dimensionality of the problem. A scaled down (1:10000) plot of both the MS with all variables and the MS with the useful variable is shown in Figure 2. The correlation coefficient between the MS scale constructed with all the variables and the MS constructed with only the useful variables is 0.9985.

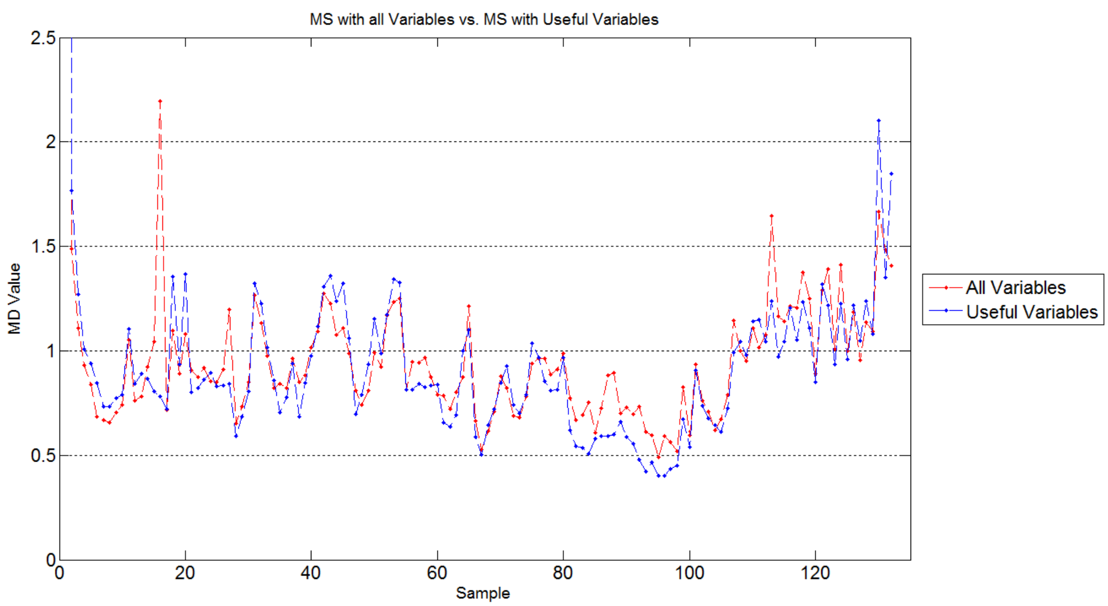


Figure 2: Plot of MS with all Variables vs. MS with useful variables.

4. Conclusions

In this paper, MTS was implemented to reduce the dimensionality of the problem and develop a scale based on MD values from the dataset. MTS identified a set of useful variables from the original dataset and these variables were used to create an MS with significant correlation to the scale created with all variables in the dataset.

Forty-three factors from the dataset were considered and 23 were determined to be useful variables with MTS. Using the 23 useful variables identified an MS was created. The correlation coefficient between the MS created with useful variables and that created with all variables was 0.9985.

5. Future Work

Future research will further apply more data sets to validate the MS created with the useful variables. Furthermore, the scale will be used to design a real time fault prediction system. The MS based on the identified normal group will be used to select a threshold value that enables appropriate classification of abnormal operating conditions of the vehicles. An MD value below the threshold value would indicate normal operating condition and a MD value greater than the threshold value would indicate a cross over to the failure zone and impending breakdown.

Future research should be conducted to utilize the methodology for fault or variation prediction in other contexts, particularly manufacturing. This can be applied to a notification system for impending faults in a manufacturing setup and immediate and clear actions integrated. Further work should also seek ways to classify and identify the different cause of failures along with the prediction.

6. Acknowledgements

This work was supported by the U.S. Army Tank Automotive Research, Development and Engineering Center (TARDEC).

constitute or imply its endorsement, recommendation, or favoring by the United States Government or the Department of the Army (DoA). The opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or the DoA, and shall not be used for advertising or product endorsement purposes.

References

1. Xue, Y., Williams, D. P., & Qiu, H. (2011). Classification with imperfect labels for fault prediction. Paper presented at the *Proceedings of the 1st International Workshop on Data Mining for Service and Maintenance, KDD4Service 2011 - Held in Conjunction with SIGKDD'11*, 12-17.
2. Cudney, E. A., Hong, J., Jugulum, R., Paryani, K., Ragsdel, K. M., & Taguchi, G. (2007). An evaluation of mahalanobis-taguchi system and neural network for multivariate pattern recognition. *Journal of Industrial and Systems Engineering, 1*(2), 139-150.
3. Taguchi, G. and Jugulum, R. (2002). *The Mahalanobis-Taguchi Strategy,* New York: John Wiley & Sons, Inc.
4. Woodall, W. H., Koudelik, R., Tsui, K. -., Kim, S. B., Stoumbos, Z. G., & Carvounis, C. P. (2003). A review and analysis of the mahalanobis-taguchi system. *Technometrics, 45*(1), 1-15.
5. Lv, Y., & Gao, J. (2011). Condition prediction of chemical complex systems based on multifractal and mahalanobis-taguchi system. Paper presented at the *ICQR2MSE 2011 - Proceedings of 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, 536-539.
6. Wang, Z., Wang, Z., Tao, L., & Ma, J. (2012). Fault diagnosis for bearing based on mahalanobis-taguchi system. Paper presented at the *Proceedings of IEEE 2012 Prognostics and System Health Management Conference, PHM-2012.*
7. Hu, J., Zhang, L., & Liang, W. (2012). Incipient machine degradation assessment by MF-MTS system. Paper presented at the *Proceedings of 2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, ICQR2MSE 2012*, 635-640.

8. Soylemezoglu, A., Sarangapani, J., & Saygin, C. (2010). Mahalanobis taguchi system (MTS) as a prognostics tool for rolling element bearing failures. *Journal of Manufacturing Science and Engineering, Transactions of the ASME, 132*(5)
9. Soylemezoglu, A., Sarangapani, J., & Saygin, C. (2011). Mahalanobis-taguchi system as a multi-sensor based decision making prognostics tool for centrifugal pump failures. *IEEE Transactions on Reliability*, *60*(4), 864-878.
10. Rai, B.K., Chinnam, R.B. and Singh, N. (2008) 'Prediction of drill-bit breakage from degradation signals using Mahalanobis-Taguchi system analysis', *Int. J. Industrial and Systems Engineering,* Vol. 3, No. 2, pp.134-148.

PAPER III

Mahalanobis-Taguchi System for Multiclass Classification of Steel Plates Fault

Adebolaji A. Jobi-Taiwo and Elizabeth A. Cudney

***Abstract*– Fault identification is fundamental to condition monitoring. An identification method for a single fault is unbalanced as there are usually multiple possible failures involved when considering a system. This paper presents a method for applying the Mahalanobis-Taguchi system (MTS) in a multiclass problem space. MTS provides a means of extracting information in a multidimensional system and integrating information from different variables into a single composite metric. MTS is used to construct reference scales by creating individual measurement scales for each class. These measurement scales are based on the Mahalanobis distance (MD) for each sample. Orthogonal arrays (OA) and signal-to-noise (SN) ratio are used to identify variables of importance and these variables are used to construct a reduced model of the measurement scale. By reducing the dimensionality of the problem, less variables are tracked which reduces the cost of the system monitoring. A classification threshold based on 1.5 sigma shift from the center of the measurement scales was utilized for each class. In order to evaluate the effectiveness of the method presented, a case on multiple fault class of manufacturing a steel plate is studied, and results indicate the practicality of the method in industrial applications.**

***Index Terms*– Classification, multiclass problem, multivariate analysis, Mahalanobis-Taguchi system (MTS)**

I. Introduction

Pattern analysis is a critical part of the human learning process and perception. Human intelligence is directly related to our ability to recognize and classify patterns [1] [2]. Pattern recognition involves making inferences based on available information for different purposes including classification, prediction, and estimation. While humans conduct pattern recognition by combining information gathered on patterns from past experiences and intuition, artificial systems carry out pattern recognition through algorithms that conduct comparisons based on statistical information extracted from features of a data set. The objective of pattern recognition is to arrive at accurate, robust, and reasonable decisions by identifying observed or perceived patterns and make associations or dissociations based on these patterns. The Mahalanobis-Taguchi System (MTS) is a pattern information technology that has been used in different applications for quantitative decision-making in multivariate problems [3].

MTS relies on multivariate statistical analysis and, as a result, is well suited for multidimensional problems. In a multivariate system, decision-making is typically based on analyzing information provided by more than one variable. The evaluation of each variable without considering the relationship to all other variables within the system would be incomplete. MTS evaluates the relationships between variables using the Mahalanobis distance (MD). MD is different from other distance measure such as the Euclidean distance because it presents a generalized measure of distance that represents the degree of divergence in the mean values of different characteristics of a population by considering the correlation among the variables [4] [5]. The MTS methodology involves identifying variables for an ideal reference group, calculating MD values for observations which represent this group, and using orthogonal arrays (OA) and signal-to-noise (SN) ratio to identify attributes of importance thereby reducing the dimensionality in multivariate problems [6]. By evaluating the correlation among variables along with other statistical measures, MTS presents a single compound value that can be used to judge the similarity of an observation to the reference group.

Data gathered from monitoring systems often consists of multiple multivariate time series from multiple sensing devices [7] and is fed into a model for data analysis using various techniques. One of the data analysis techniques used is MTS because of its suitability for multivariate data analysis. MTS provides an analysis on the measurement parameters and seeks a correlation with the result, thereby reducing the number of parameters being measured by identifying the useful variables contributing to the problem. Fault identification is fundamental to condition monitoring. Condition monitoring with a single fault identification method is unbalanced as equipment are typically prone to failure due to multiple faults [8]. Hence, multiclass classification of faults is important to effective condition monitoring. The purpose of this paper is to apply MTS to multiclass classification problem in fault identification. The MTS methodology is applied to the Steel Plates Fault Data Set obtained from the University of California, Irvine Machine Learning Repository. The objective is to correctly identify the type of fault recorded on a steel plate based on the data presented on the plate attributes.

II. BACKGROUND

At the core of the Mahalanobis-Taguchi System is the Mahalanobis distance. MD is used to create a reference frame from a "normal" sample known as the Mahalanobis space (MS) [4], which is based on the relationship between feature variables and is used to measure a sample's homogeneity to the reference group as the basis for classification. Orthogonal arrays and signal-to-noise ratios are then used to select an optimal set of variables for the classification. There is considerable research available on the application of MD based fault monitoring systems as well as solutions based on MTS. Based on the literature survey presented below, two broad categories are observed with the application of MTS or MD to fault monitoring problems. Researchers have taken these approaches: 1) direct application of MTS or MD, or a methodology based on MTS or MD, and 2) application of a modification or variant of MTS suited to the nature of the problem.

Kumar et al. [9] present an MD based diagnostic technique for condition monitoring by detecting trends and biasness in system health through a control chart for the MD values of observations across several parameters. The method employs a probabilistic approach to establish thresholds for classification of products as healthy or unhealthy. Using a Box-Cox transformation Kumar et al. obtained a normally distributed transformed variable for the MD values and an optimized MD value was selected to qualify a product against a particular fault using an error function. The distribution of the residual, which is the difference between a parameter's estimated and observed values, was used to isolate parameters that exhibit faults and establish reasons for the fault. Two fault thresholds were developed, a generic threshold MD value for detecting any type of fault in a product; and a fault specific threshold for detecting a particular fault based on historical data related to that fault. The authors presented a case study on notebook computers and demonstrated the approach was able to detect faults in a product with 99% accuracy.

Saraiva et al. [10] applied a modified version of MTS, called the Modified Mahalanobis-Taguchi strategy (MMTS), to fault identification in chemical processes. MMTS was classified as a process history-based method. MMTS was applied to a continuous-stirred tank reactor (CSTR) simulated by Kano et al. [11], through a combined multivariate statistical process control (CMSPC) technique and the results were compared. CMSPC is an integration of a principal component analysis-based statistical process control (SPC) and an independent component analysis-based SPC. Saraiva et al. used the same monitored variables and data used by Kano et al., which covered normal operating conditions and eleven different abnormal conditions. The primary modification by Saraiva et al. to MTS was the application of multiple regression analysis (MRA). To optimize the MTS system Saraiva et al. chose to use stepwise MRA for the selection of useful variables and identified a threshold using the corresponding average run length (ARL) from Kano et al. According to Saraiva et al., ARL is an accepted criterion more suited for evaluation using SPC and fault diagnosis procedures. The results showed little or no difference between the ARL scores obtained from the CMSPC approach compared to the MMTS approach when all variables were used. More importantly, when a subset of the variables was used, the results showed similar performances. Saraiva et al. described MMTS as "a cheaper and efficient fault diagnosis system." In addition, Saraiva et al. described MMTS as a promising data-based approach for on-line fault detection.

Jin et al. [12] present an approach for condition monitoring of cooling fans in electronic products based on MD with feature selection using the minimum redundancy maximum relevance (mRMR) criteria. A subset of available features on the cooling fans was selected using the mRMR criteria, hence reducing the dimensionality of the problem and redundancy within the data set. This also helped avoid multicollinearity with the MD values based on the features selected by the mRMR criteria. A case study based on a data set tracking the degradation process of cooling fans due to a loss of lubricant in the ball bearings was presented. Jin et al. showed their method to be more effective than principal component analysis by detecting anomalies in the cooling fan upon operations and clearly showing the degradation trend of the ball bearings.

Soylemezoglu et al. [13] applied an MTS based fault prognostics system to rolling element bearing failures. The system detects a fault, diagnoses its root cause (fault isolation), and estimates the remaining useful life or time to failure which completes the prognosis. MD values were calculated for the variables being monitored and thresholds were set for the normal operation condition. The three types of faults considered in the experiment included cage defect, inner race defect, and outer race defect. The specific fault was determined by identifying the specific threshold band in which the MD value falls. To complete the fault prognosis, the MD values were calculated using a predetermined time window as the bearing was being monitored. Fault detection occurred once the MD crossed the specified threshold and the tracking of the MD trend was initiated. By tracking the direction of the transition of the MD trend and calculating the angle between the MD point and the mean MD of the three known fault clusters, a root cause was identified. The smaller the angle the more likely it is that the fault is progressing towards one of the fault clusters. The prognosis of the time to failure was calculated via linear approximation. Initially ten features were selected to construct the MS which reduced to eight used in the prognosis. Soylemezoglu et al. achieved a 100% success rate in correct detection and isolation of bearing faults with this tool.

In subsequent research, Soylemezoglu et al. [14] applied a comprehensive fault monitoring tool based on MTS to centrifugal pump failures. The methodology was modified by including cluster analysis for better identification of threshold values and for identifying the optimum number of sensors required for the condition monitoring. MTS was used to fuse data from multiple sensors on the centrifugal pump into a single system level performance metric using MD. Cluster analysis was used to create fault clusters based on the MD values generated. Thresholds determined from the clustering analyses were used to detect and isolate faults. To complete the fault prognosis, the MD values were calculated using a predetermined time window and linear approximation was used to estimate the time to failure. The experiment used 18 parameters measured from a 1/2 HP centrifugal pump operated for 150 hours. The experiment investigated three types of failures including seal failure, impeller failure, and filter clog. A high success rate was achieved on all three faults in the research.

Yu [15] presents methods for testing parametric and catastrophic faults in analog and mixed signal circuits based on the wavelet transform of measured signals. The wavelet analysis uses algorithms with two different metrics, one based on the discrimination factor of Euclidean distances and the other utilizes the Mahalanobis distance. Experimental results from covering individual circuits testing and production line testing are presented, with the methods compared with other mixed-signal fault detection methods. The results showed that the test algorithm with the MD-based metric performed better than the one based on the discrimination factor due to more information obtained by capturing signal correlations.

Rai et al. [16] used MTS for online prediction of drill-bit failure (breakage) from two degradation signals, thrust force and torque, during a drilling operation. Rai et al. described the advantage of MTS to other online tool-condition monitoring methods as the flexibility of the methodology, as it allows monitoring multiple features simultaneously, and the selection of useful features. Ten features were monitored for both degradation signals over 128 drilled holes. The data was collected from running nine drill-bits until breakage. Data collected from the last hole successfully drilled by a drill-bit was defined as belonging to the abnormal class. All data collected prior to that were classified as being in the normal operation set. Rai et al. acknowledged the fact that an on-set of the drill-bit degradation might occur well in advance to the last hole but stated that this method allowed for the maximum usage of tool life. In the research, five useful features were identified and used to successfully predict drill bit failure based on a threshold value of the resulting MD values from these features.

Wang et al. [17] used MTS for fault diagnosis of a rolling bearing element after feature extraction by time/frequency domain analysis from vibration data. Data was collected over four states of the rolling bearing element including normal condition, inner race fault, outer race fault, and rolling element fault. Initially a set of eight characteristics was extracted via the time/frequency domain analysis. MTS was used to identify six useful variables used in the diagnosis.

Riho et al. [18] applied a modified version of MTS, called MTS+, to diagnose the cause of invisible defects in order to enhance production yield in a wafer production process. The defects, referred to as a white point (WP), were especially common in charge-coupled devices (CCD), which was the focus of the research. If a chip has a WP after the CCD process is completed then a failure has occurred; otherwise, the chip is said to be normal. In the implementation of MTS+, which combined several original techniques from the aspect of yield enhancement, Riho et al. tried to determine the degree of contribution from every process parameter to the WP failures observed. Riho et al. used MTS to identify the important variables contributing to the WP failures and, based on these variables, designed the experiments and carried out investigations. Riho et al. confirmed that the WP defects were connected with organic matter found on the chips after the CCD process and, subsequently, developed countermeasures to enhance the yield quicker and more accurately.

Mohan et al. [19] developed an MTS-based real time diagnostics and root cause analysis tool to diagnose the quality of fastening operations for a hand-held pull-type pneumatic tool and specify the cause of the failure. For the research four characteristics were measured including peak strain, peak displacement, and depth and width of a bowl-shaped dip on the process signature. The data was collected wirelessly and fed to the system to make real-time decisions on the grip length of the fastening operation. In addition to being used to identify failures (deviations in grip length), MTS was used for root cause analysis. Mohan et al. reproduced each abnormality and calculated the MD values with respect to data from the normal operating condition. For instances where there were a number of similar cases of abnormalities, an MD range corresponding to this set was determined with respect to the ideal case. Signatures were analyzed from these ranges using a correlation matrix and the MD value was calculated. A fault was characterized by which range the MD value fell under and the type of abnormality was determined. The MTS tool selected two of the four characteristics as important and had a detection rate of 87.5%, 100%, and 96.8% for over grip, normal grip, and under grip, respectively.

From the literature survey it is clear that the applications of MTS and MD to condition monitoring and fault identification research efforts have largely been successful.

However, there is little research on the application of MTS to multiclass classification problems. Hsiao et al. [20] applied an automatic multiclass classification system based on MTS to the inspection of saxophone timbre quality. Ren et al. [21] proposed an improved MTS based fault diagnosis scheme that used vibration signals from rotating machinery by means of a seeded-fault-test. Su et al. [22] present a multiclass MS for simultaneous feature selection and classification that was applied to a case study for the identification and prediction of stages of diabetes mellitus in pregnant women.

III. METHODOLOGY

In this paper, an approach to multiclass classification based on MTS is presented. The proposed scheme utilizes MD-based thresholds to classify faults into distinct fault groups by assessing the degree of abnormality in the variables being monitored relative to a reference group for each fault class, also known as the Mahalanobis space (MS). MTS is used to identify the most important features for the classification. Details on the approach are presented in the rest of this section.

A. Mahalanobis Distance

MD is a distance measure derived from an analysis of the deviation in the mean values of variables in multivariate data considering the correlation between the variables. This distance measure was introduced by Prasanta Chandra Mahalanobis in 1936. MD, as a discriminant analysis method, is useful in determining the similarity between a set of values from an unknown sample and a set of values from known samples. It provides a measure of homogeneity or heterogeneity between groups of data. MD proves to be superior to other multidimensional distance measures for the following reasons [4]: 1) it integrates correlation between the variables into its calculation, 2) it is very sensitive to inter-variable changes in the reference data, and 3) it is not affected by dimensionality of the dataset.

Assuming a dataset consists of k variables; i represents the variables (i = 1, 2, …, k); n represents the number of samples in the dataset; and j is the sample number (j = 1, 2, …, n), the variables are standardized as defined in (1).

$$z_{ij} = \frac{(x_{ij} - m_i)}{s_i} \tag{1}$$

Where, m_i and s_i represent the mean and standard deviation of the ith variable, respectively; x_{ij} is the value of the ith characteristic of the jth observation. MD values are calculated as defined in (2).

$$MD_j = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \tag{2}$$

Where MD_j is the Mahalanobis distance calculated for the jth observation, Z_{ij} is the standardized vector obtained by the standardized values of x_{ij} (z_{1j}, $z_{2j}, \ldots,$ z_{kj}), ^{T} represents the transpose of the vector, and C^{-1} the inverse of the correlation matrix. MD provides an indicator for the "nearness" of a sample to the mean point of a known group of samples taking into account the correlation between the variables [4].

B. Mahalanobis-Taguchi System

MTS was developed by Genichi Taguchi as a diagnosis and forecasting technique using multivariate data [23]. It is a pattern recognition technology that assists in quantitative decision-making through the construction of a multivariate measurement scale using data analytic procedures. In a multivariate system, decision-making is typically based on analyzing information provided by more than one variable. Evaluation of each variable without considering the relationship to all other variables within the system would be incomplete. As shown in the previous section, MD is a unique measure as it accounts for the correlation of variables in a multidimensional system. MTS bridges the relationships between variables using the MD. For this reason, it is an ideal tool for the analysis of multivariate data and systems.

In developing a multivariate measurement scale it is important to 1) have a reference point to the scale, 2) validate the scale, 3) select the important variables adequate for measuring abnormality, and 4) be able to carry out future diagnosis with the measurement scale. These form the basis of the MTS implementation. MTS is used to develop a scale to evaluate the degree of abnormality of measurements compared to the "normal" measurements. The Mahalanobis distances for the attributes are calculated to generate an ideal reference for the normal data. Taguchi's robust engineering is used to determine the variables of importance, thereby optimizing the system. Taguchi's robust engineering has two major tools, OA and SN ratios. The steps for implementing MTS are shown in Fig. 1.

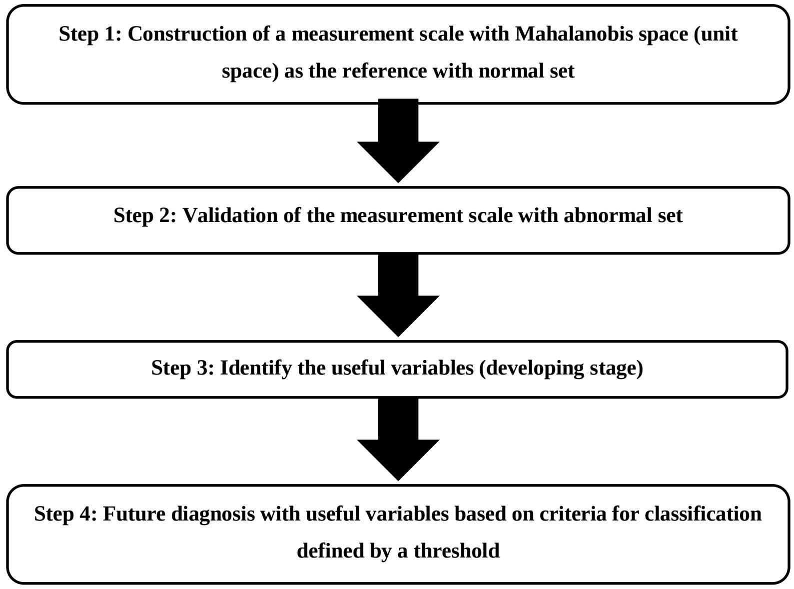


Fig. 1. Steps for implementing MTS.

Step 1: Construction of the measurement scale: In the first step of MTS, variables that define the “normal” state are identified and data is collected on all these variables for the normal state to create a reference. This reference is referred to as the Mahalanobis space, also known as the unit space as the average value of the MDs for all the observations constituting the reference is one [4]. Samples of the normal state for these variables are standardized as in (1) and MD values are calculated using the inverse of the correlation matrix as in (2). These MD values constitute the MS.

Step 2: Validation of the measurement scale: In order to validate the MS, MD values for the observations that belong to the “abnormal” group are calculated. The MD calculation for the abnormal group utilizes the mean and standard deviation of the corresponding variable from the normal group for standardization. Also, the correlation matrix of the normal group is used for the MD calculation. The premise behind this is that since the data from the abnormal group is from the same system then it should be evaluated along with the normal cases. If the scale is correct, the MD values for the “abnormal” group should have higher values compared to the “normal” group. This validates the measurement scale.

Step 3: Identify useful variables: The aim of this step is to identify the optimum number of variables required for the measurement scale. Variables that do not significantly contribute to the measurement scale are discarded. Taguchi's robust engineering techniques are employed to achieve this segregation through the use of OA and SN ratio. OA is an experimental design matrix used to list the combinations of characteristics and enables testing the effects of the absence or presence of a characteristic in experimentation. The size of the OA to be used is determined by the number of variables. Each variable is assigned to a column in the OA and it set to only two levels. Level 1 corresponds to the presence of a variable while level 2 corresponds to the absence of a variable. For each run of the OA taken across the rows, the MD values for the "abnormal" cases are calculated using only the included variables as specified by the OA. OA minimizes the number of experiments and reduces the impact of noise factors. The resulting MD values are used to calculate the SN ratio, which measures the system functionality. The signal to noise ratio can be calculated in two ways, dynamic SN ratio shown in (3) and larger-the-better shown in (4).

$$\eta_q = 10log_{10}\left(\frac{1}{r}\frac{S_\beta - V_e}{V_e}\right) \tag{3}$$

Where, η_q represents the dynamic SN ratio for the qth run of the OA and t is the number of abnormal conditions.

$y_i = \sqrt{MD_i}$, i=1, …, t and β is the slope

S_T = total sum of squares ($\sum_{i=1}^{t} y_i^2$)

r = sum of squares due to the input signal ($2\sum_{i=1}^{t} M_i^2$)

S_β = sum of squares due to slope ($\frac{1}{r}\sum_{i=1}^{t}(M_i y_i)^2$)

S_e = error sum of squares ($S_T - S_\beta$)

V_e = error variance ($\frac{S_e}{t-1}$)

$$\eta_q = -10\log\left[\frac{1}{t}\sum_{j=1}^{t}\frac{1}{MD_j}\right] \tag{4}$$

Where, t represents the number of abnormalities under consideration and η_q represents the larger-the-better SN ratio for the qth run of the OA. The gain for each variable is then calculated by subtracting the average SN ratio when the variable was excluded (SN^-) from the average SN ratio when the variable was included (SN^+) as shown in (5).

$$Gain = SN^+ - SN^- \tag{5}$$

The variables with positive gains are selected for the classification, while the rest are discarded.

Step 4: Reconstruction of Scale: In the final step, the measurement scale is reconstructed with the useful set of variables identified in Step 3. Subsequently, the reconstructed scale is used to monitor the conditions of the system in question after a threshold for the boundary between the normal and abnormal cases has been specified.

IV. CASE STUDY

The steel plate fault dataset was obtained from the University of California, Irvine Machine Learning Repository website [24]. The dataset contains data on 27 variables collected for 1,941 faulty steel plates over seven faults. Six of these fault classes were distinct faults identified as "Pastry," "Z_Scratch," "K_Scratch," "Stains," "Dirtiness," and "Bumps." The final class was a collection of several other possible faults identified as "Other_Faults." For this research, only data on the six distinct fault classes were used as there was insufficient information regarding the faults grouped into the "Other_Fault" category. MTS is employed to develop a multiclass classification for the six distinct faults in the steel plate fault dataset.

The first step is to define reference frames for each of the six fault classes. A Mahalanobis space is created for each fault class. Defining the reference point is critical to the MTS method as the MS is the base of the measurement scale. An abnormal condition would be outside of the MS and the degree of abnormality is measured in reference to the MS. Each MS defines a normal space for faults belonging to that respective fault class. Every fault outside the normal space is considered abnormal to that fault class. The MD values for faults in each of the of the six fault class were calculated as shown in (1) and (2). These MD values were used to constitute the MS for each fault class. The average values for the MS for each fault class are presented in Table I.

TABLE I
AVERAGE MD VALUE FOR FAULT CLASS MS

Class No.	Fault Name	Average MD Value for Class MS
1	Pastry	0.9998
2	Z_Scratch	1.0000
3	K_Scratch	1.0000
4	Stains	0.9998
5	Dirtiness	1.0001
6	Bumps	1.0000

In order to validate the MS scale created for each fault class, MD values were calculated for all other fault classes using the mean, standard deviation, and correlation of corresponding variables from the "normal" fault class. The MD values for the "abnormal" fault classes were significantly higher than the MD for the normal fault class, thereby validating the measurement scale. A summary of the validation results is shown in Table II.

TABLE II
AVERAGE MD VALUE FOR "ABNORMAL" FAULT CLASSES

"Normal" Fault Class	"Abnormal" Fault Classes	Average MD Value for Abnormal Fault Classes
1	2, 3, 4, 5, 6	1,060
2	1, 3, 4, 5, 6	325
3	1, 2, 4, 5, 6	134
4	1, 2, 3, 5, 6	1,288,246
5	1, 2, 3, 4, 6	3163
6	1, 2, 3, 4, 5	737

The next step is the optimization of the MTS implementation using OA and SN ratio. The objective is to reduce the dimensionality of the multivariate system and still obtain meaningful results by maintaining the discriminating power of the system. An L32 orthogonal array was used for the variable analysis. For each run of the OA an SN ratio is calculated using the larger-the-better calculation shown in (4) as it is desired that the MDs of the abnormal be as high as possible. The average gain for each variable when included and excluded from the experiments is calculated as shown in (5). This is performed for each measurement scale created for the six fault classes and the important variables identified. Eighteen important variables were identified for fault classes one, two, and three. Nineteen variables were identified for fault class four, twenty-three for fault class five, and fifteen for fault class six. The MTS implementation eliminated one variable across all six measurement scales. The outcome of the analysis for fault class 1 and fault class 6 are shown in Fig. 2 and Fig. 3 respectively.

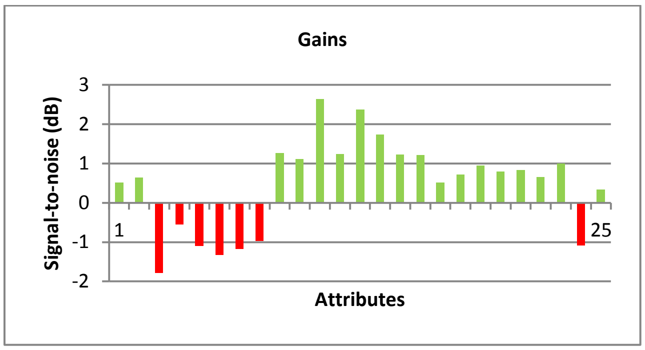


Fig. 2. OA analysis for variable gains for fault class 1 – Pastry.

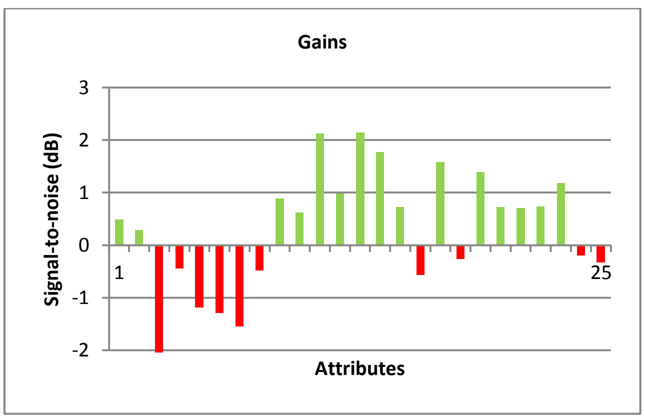


Fig. 3. OA analysis for variable gains for fault class 6 – Bumps.

Finally the measurement scales for the fault classes are reconstructed with the most significant variables and classification thresholds are defined for each fault class. Likewise, only the significant variables identified for each fault class were taken into consideration for the classification. The classification thresholds for each fault class were specified at 1.5 sigma shifts away from the center of the MS for that class based on the work of Bothe [25]. Bothe presents the statistical reason for the 1.5 sigma shift in most processes. The classification of a fault into a specific fault class is accomplished through a comparative analysis of the proximity of the MD value to each MS created for the fault classes. The MD value for each steel plate is calculated with the important variables identified for each fault class MS.

V. RESULTS

MTS was used to successfully create measurement scales for the different fault classes of the steel plate fault data set. The results for the multiclass classification were obtained using a subset of the variables for each class from the original data set. A sample of the results for the classification with the MTS multiclass approach using a classification threshold of 1.5 sigma shift is presented in Table III.

TABLE III
MTS MULTICLASS CLASSIFICATION RESULT WITH 1.5 SIGMA SHIFT THRESHOLD

		MTS Multiclass Classification Result					
S/N	**Actual Fault Class**	**1**	**2**	**3**	**4**	**5**	**6**
1	**1**	1	0	0	0	0	0
2	**1**	1	0	0	0	0	1
3	**1**	1	0	0	0	0	1
4	**1**	1	0	0	0	0	0
5	**1**	1	0	0	0	0	0
6	**1**	0	0	0	0	0	0
7	**1**	1	0	0	0	0	0
8	**1**	1	0	0	0	0	1
9	**1**	1	0	0	0	0	1
10	**1**	1	0	0	0	0	0
11	**1**	1	0	0	0	0	0
12	**1**	1	0	0	0	0	0
13	**1**	1	0	0	0	0	0
14	**1**	1	0	0	0	0	0
15	**1**	1	0	0	0	0	1

The first two columns show the serial number and actual fault class for the steel plate respectively. Columns three to eight present the binary values representing the classification results for each fault relative to fault class specified in the header and each data has a binary value in each of the six columns. A value of “1” signifies the measurement scale for the fault class classifies the fault as belonging to the corresponding fault class and a value of "0" signifies the measurement scale classifies the fault as not belonging to the corresponding fault class. The overall classification accuracy of the multiclass classification approach is 83.89%. This classification accuracy represents the correct classification of the steel plate faults on all the six measurements scales created for the fault classes. The confusion matrix for the multiclass classification is presented in Table IV.

TABLE IV
CONFUSION MATRIX FOR CLASSIFICATION RESULT WITH 1.5 SIGMA SHIFT THRESHOLD

		Classified As :	
		Not MS	**MS**
Actual Value:	**MS**	56 **(FP)**	1212 **(TP)**
	Not MS	5170 **(TN)**	1170 **(FN)**

With the 1.5 sigma shift threshold, 84.21% of the faults were narrowed down to two possible faults with one of the faults corresponding to the actual fault class. With the same threshold, 94.40% of the faults were narrowed down to three possible faults with one of the faults corresponding to the actual fault class.

VI. CONCLUSION

In this paper, the application of MTS to a multiclass classification problem is presented. MTS is a pattern technology tool that was developed for multivariate data analysis. Using MD to integrate the information from different variables in a single composite metric, MTS is able to create reference frames and make classifications based on these reference frames. It reduces the dimensionality of condition monitoring problems by identifying the variables correlated to the effect being monitored with the most significant contribution in a multivariate system. By reducing the number parameters to be monitored in a process or on equipment, the cost of the condition monitoring is significantly reduced. MTS is a flexible tool that can be easily modified or combined with other tools and techniques as required by the peculiarities of the problem.

The results show that MTS can be successfully used in the multiclass classification problems while also reducing the dimensionality of the problem. With the steel plate fault dataset MTS was able to completely eliminate one variable from the measurement scale created for all six fault classes. It also significantly reduced variables for all six fault classes without compromising the information available from the data as can be verified from the high level of accuracy with fewer variables. This has practical application in industry for diminishing inspection and production expenses related to sensing and measuring in condition monitoring.

Future research work should involve the application of the approach to other multiclass problems with larger data samples. In addition, methods of integrating a dynamic threshold to the classifications for several groups should also be explored.

REFERENCES

[1] J. R. Flynn, What is intelligence?: Beyond the Flynn effect. Cambridge: Cambridge University Press, 2007, pp. 6.

[2] S. K. Das, and T. De, “Methodology of the heuristic based hybrid clustering technique for pattern classification and recognition,” in *International Conference on Advanced Computing and Communication Technologies*, 2013, pp. 28-35.

[3] E. Cudney, J. Hong, R. Jugulum, K. Paryani, K. Ragsdell, and G. Taguchi, “An Evaluation of Mahalanobis-Taguchi System and Neural Network for Multivariate Pattern Recognition,” *Journal of Industrial and Systems Engineering*, vol.1, no.2, 2007, pp. 139-150.

[4] G. Taguchi, and R. Jugulum, *The Mahalanobis-Taguchi Strategy: A Pattern Technology System*, John Wiley & Sons, Inc., 2002.

[5] E. Cudney, D. Drain, K. Paryani, and N. Sharma, “Forecasting Consumer Satisfaction for Vehicle Ride Using a Multivariate Measurement System,” *Journal of Industrial and Systems Engineering*, vol.4, no.6, 2009, pp. 638-696.

[6] E. Cudney, D. Drain, K. Paryani, and N. Sharma, “A Comparison of the Mahalanobis-Taguchi System to a Standard Statistical Method for Defect Detection,” *Journal of Industrial and Systems Engineering*, vol.2, no.4, 2009, pp. 250-258.

[7] Y. Xue, D. P. Williams, and H.Qiu, “Classification with imperfect labels for fault prediction,” in *International Workshop on Data Mining for Service and Maintenance*, 2011, pp. 12-17.

[8] N. Bachschmid, P. Pennacchi, and A. Vania, "Identification of multiple faults in rotor systems," in Journal of Sound and Vibration, vol.254, no.2, 2002, pp. 327-366.

[9] S. Kumar, T. W. S. Chow, and M. Pecht, "Approach to Fault Identification for Electronic Products Using Mahalanobis Distance," *Instrumentation and Measurement, IEEE Transactions on*, vol.59, no.8, Aug. 2010, pp.2055-2064.

[10] P. Saraiva, N. Faísca, R. Costa, and A. Gonçalves, “Fault identification in chemical processes through a modified Mahalanobis-Taguchi strategy,” in *Computer Aided Chemical Engineering*, vol.18, no.121, May 2004, pp.799-804.

[11] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, “Combination of independent component analysis and principal component analysis for multivariate statistical process control,” in *Proceedings of International Symposium on Design, Operation and Control of Chemical Plants, Taipei*, Dec. 2002, pp. 319.

[12] X. Jin, E. W. M. Ma, L. L. Cheng, M. Pecht, "Health Monitoring of Cooling Fans Based on Mahalanobis Distance With mRMR Feature Selection," *Instrumentation and Measurement, IEEE Transactions on* , vol.61, no.8, Aug. 2012, pp.2222-2229.

[13] A. Soylemezoglu, J. Sarangapani, and C. Saygin, “Mahalanobis Taguchi system (MTS) as a prognostics tool for rolling element bearing failures,” in *Journal of Manufacturing Science and Engineering,* vol.132, no.5, 2010.

[14] A. Soylemezoglu, J. Sarangapani, and C. Saygin, "Mahalanobis-Taguchi System as a Multi-Sensor Based Decision Making Prognostics Tool for Centrifugal Pump Failures," *Reliability, IEEE Transactions on*, vol.60, no.4, Dec. 2011, pp.864-878.

[15] J. Yu, "Health Condition Monitoring of Machines Based on Hidden Markov Model and Contribution Analysis," *Instrumentation and Measurement, IEEE Transactions on*, vol.61, no.8, Aug. 2012, pp.2200-2211.

[16] K. Rai, Bharatendra, R. B. Chinnam, and N. Singh, "Prediction of drill-bit breakage from degradation signals using Mahalanobis-Taguchi system analysis," in *International Journal of Industrial and Systems Engineering,* vol.3, no.2, 2008, pp.134-148.

[17] Z. Wang, Z. Wang, L. Tao, and J. Ma, "Fault diagnosis for bearing based on Mahalanobis-Taguchi system," in *Prognostics and System Health Management (PHM), 2012 IEEE Conference on,* 2012, pp. 1-5.

[18] T. Riho, A. Suzuki, J. Oro, K. Ohmi, and H. Tanaka, "The yield enhancement methodology for invisible defects using the MTS+ method," *Semiconductor Manufacturing, IEEE Transactions on* , vol.18, no.4, Nov. 2005, pp.561-568.

[19] D. Mohan, C. Saygin, and J. Sarangapani, "Real-time detection of grip length deviation during pull-type fastening: a Mahalanobis–Taguchi System (MTS)-based approach," in *The International Journal of Advanced Manufacturing Technology,* vol.39, no.9-10, 2008, pp. 995-1008.

[20] Yu-Hsiang Hsiao; Chao-Ton Su, "Multiclass MTS for Saxophone Timbre Quality Inspection Using Waveform-shape-based Features," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol.39, no.3, June 2009, pp.690-704.

[21] Jiangtao Ren; Xiaochen Xing; Yuanwen Cai; Jing Chen, "A method of multi-class faults classification based-on Mahalanobis-Taguchi system using vibration signals," *Reliability, Maintainability and Safety (ICRMS), 2011 9th International Conference on*, no. 5979414, 12-15 June 2011, pp.1015-1020.

[22] Chao-Ton Su; Yu-Hsiang Hsiao, "Multiclass MTS for Simultaneous Feature Selection and Classification," *Knowledge and Data Engineering, IEEE Transactions on*, vol.21, no.2, Feb. 2009, pp.192-205.

[23] W. H. Woodall, R. Koudelik, K. L. Tsui, S. B. Kim, Z. G. Stoumbos, and C. P. Carvounis, “A review and analysis of the Mahalanobis-Taguchi system,” in *Technometrics*, vol. 45, no.1, 2003, pp. 1-15.

[24] Dataset provided by Semeion, Research Center of Sciences of Communication, Via Sersale 117, 00128, Rome, Italy. Available: https://archive.ics.uci.edu/ml/datasets/Steel+Plates+Faults

[25] D. R. Bothe, “Statistical reason for the 1.5σ shift,” in *Quality Engineering,* vol. 14, no.3, 2002, pp. 479-487.

SECTION

2. CONCLUSION

MTS is a viable tool for classification and forecasting tasks. MTS is also useful for the identification of important variables of a multivariate system and the construction of a reduced model of the measurement scale used for the classification or forecasting. This directly translates to cost savings in condition monitoring problems as it is used to reduce the dimensionality of the problem, thereby eliminating the need to monitor the entire set of parameters of the process or equipment. The work presented in this thesis also shows how MTS can be applied to multiclass classification problems. This has practical application in industry for diminishing inspection and production expenses related to sensing and measuring in condition monitoring.

Future work should explore the application of MTS to problems with larger data and attribute size to further demonstrate the ability of MTS. With the application of MTS to other multiclass classification problems, methods of defining thresholds for the class groups should be explored. Integration of a dynamic thresholding method to the MTS would also increase the versatility of the MTS methodology. Other areas might include the application of MTS or a variation of MTS to root cause analysis in condition monitoring problems.

VITA

Adebolaji A. Jobi-Taiwo was born in Lagos, Nigeria on April 21, 1986. He received his B.S. in Systems Engineering with a concentration in Operations Research and Manufacturing from the University of Lagos in 2009. He worked for Lafarge Cement from June 2010 until August 2012 first as a Project Engineer on the Operation Preparation and Commissioning team of the Lakatabu Expansion Project and as a Manufacturing Control Engineer on the commissioned plant. In May, 2014, He received his M.S. in Systems Engineering from the Missouri University of Science and Technology, Rolla, Missouri.

Adebolaji Jobi-Taiwo holds a Project Management Professional (PMP) certification. He is a member of the Project Management Institute (PMI), the International Council on Systems Engineering (INCOSE) and the American Society for Quality (ASQ).