

# A Two-Step Parametric Method for Failure Prediction in Hard Disk Drives

Yu Wang, *Student Member, IEEE*, Eden W. M. Ma, *Member, IEEE*, Tommy W. S. Chow, *Senior Member, IEEE*, and Kwok-Leung Tsui

**Abstract**—Predicting the impending failure of hard disk drives (HDDs) is crucial for preventing essential data from losing. In this paper, a two-step parametric method was developed to predict the impending failure of HDDs using the aggregate of statistical models. This method deals with the problem of failure prediction in two steps: anomaly detection and failure prediction. First, Mahalanobis distance was used for aggregating all the monitored variables into one index, which was then transformed into Gaussian variables by Box–Cox transformation. By defining an appropriate threshold, anomalies in HDDs were detected as a result. Second, a sliding-window-based generalized likelihood ratio test was proposed to track the anomaly progression in an HDD. When the occurrence of anomalies in a time interval is found to be statistically significant, indicating the HDD is approaching failure. In this work, we also derived a new cost function to adjust the prediction rate. This is important in a way to balance the failure detection rate and false alarm rate as well as to provide an advanced warning of HDD failures to the users, whereby the users can back up their data in time. Then the developed method was applied on a synthetic data set showing its effectiveness on predicting failures. To demonstrate the practical usefulness, this method was also applied on a real-life HDD data set. The result shows that our method could achieve 68% failure detection rate with 0% false alarm rate. This is much better than the results achieved by the state-of-the-art methods, such as support vector machine and hidden Markov models.

**Index Terms**—Anomaly detection, failure prediction, generalized likelihood ratio test, hard disk drive (HDD).

## I. INTRODUCTION

THE technologies used in hard disk drives (HDDs) have been improved immensely in recent years making the storage capacity reach several terabits (Tb) per drive [1], [2]. Failure of an HDD may cause serious data loss and services' down time, which could result in causing tremendous financial and economic losses to users and communities. On the

Manuscript received November 05, 2012; revised February 19, 2013; accepted May 06, 2013. Date of publication May 20, 2013; date of current version December 12, 2014. This work was supported in part by the by the City University Strategic Research Grants (7002914), Innovation Technology Fund (GHX/003/12), HK RGC CRF (CityU8/CRF/09), and HK RGC GRF (121410). Paper no. TII-12-0756.

Y. Wang and K.-L. Tsui are with the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong (e-mail: ywang95@student.cityu.edu.hk; klttsui@cityu.edu.hk).

E. W. M. Ma is with the Center for Prognostics and System Health Management, City University of Hong Kong, Hong Kong (e-mail: eden.wm.ma@cityu.edu.hk).

T. W. S. Chow is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: eetchow@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TII.2013.2264060

other hand, as HDD has become one of the most competitive computer industries, designing and manufacturing highly reliable HDD is essential to maintain the market shares for manufacturers. Designing new algorithms for online anomalies detection and failure prediction is strategically important.

Currently, the techniques used by HDD experts for monitoring the HDD healthy status mainly fall into three categories. First, the experts utilize external sensors, such as accelerometers and acoustic emission sensors, to monitor the evolution of vibration signals during HDD operation [3]. But attaching these sensors outside the HDD cannot obtain the degradation signatures, while attaching them inside the HDD may damage its components and increase the physical size [3]. Second, the experts utilize the log files of storage systems, which collect the error events of the software and hardware deployed in the storage system [4], [5]. However, such techniques generally fail to provide the insightful information associated with HDD performance due to lack of close monitoring. Third, the experts utilize self-monitoring, analysis, and reporting technology (SMART), which is a built-in function of HDD [6]–[8]. It collects the values of attributes (performance parameters) that correspond to record counts or physical units, e.g., temperature and current, to determine an HDD's health state without any intrusion. Such an advantage inspired most HDD manufacturers to adopt this technology.

However, the SMART software used by HDD manufacturers only provides a basic evaluation. For example, “threshold not exceeded” and “threshold exceeded” are represented as “drive OK” and “drive fail,” respectively [7]. The threshold for every attribute is confidentially defined by manufacturers. In operation, when any attribute exceeds its predefined threshold value, the drive is considered a failure, which results in returning to the manufacturers for replacement. Apparently HDD manufacturers would like to minimize false alarm rate (FAR) by setting a higher threshold. This, however, makes the failure detection rates (FDRs) to be poor (the FDR of the SMART algorithm is 3–10 % [9]).

Several algorithms have been developed in the past decade to enhance the FDR. Naive Bayes expectation-maximization was proposed by Hamerly *et al.* [10] to predict HDD failure using the data from Quantum Company. Instead of using all attributes from the original data set, three attributes (grown defect count, read soft errors, and seek errors) provided better prediction performance. Hughes *et al.* [7] proposed a non-parametric method, the rank-sum statistical test, to replace the threshold method. Murray *et al.* [9], [11] compared different machine learning methods (multiple-instance naive Bayes, support vector machine (SVM), autotclass, and rank-sum test)

and found that SVM had the highest FDR among all tested methods under the same FAR. In particular, SVM gets 50.6% FDR at zero FAR. Tree augmented naive Bayesian (TAN) [12] and hidden Markov models (HMM) [13] were proposed to further improve the FDR. The results of TAN [12] showed that they could achieve over 80% FDR with about 3% FAR. But when pushing to 0% FAR, it can only deliver a 20%–30% FDR [12]. The HMM-based approach [13] is able to achieve 52% FDR at 0% FAR, but fails to get a higher FDR at higher FAR compared to TAN approach.

Despite SVM [9], TAN [12], and HMM [13] being able to achieve higher prediction performance than other methods mentioned above, they still have certain disadvantages. First, these methods do not thoroughly take the characteristics of the observed attributes over time in the failed drives underwent a process into account: from health to failure. Instead, they label the observations of the failed drives to be failed (positive) irrespective of the measurements' healthy or faulty state. Labeling an observation as a failure, however, requires much prior knowledge on the data set, and always to be a major challenge for the standard supervised learning schemes [14]. Arbitrarily labeling the observations in failed drives as fail makes the healthy classes confused with failed classes, resulting in a suboptimal model with poor prediction performance [9]. Second, some of these methods, such as SVM and TAN, tend to conduct failure prediction based on individual observation. Using individual anomaly, though can realize real-time prediction, may not be used consistently for providing a reliable estimate on impending HDD failure. This can be explained on the one hand by the fact that an individual measurement cannot give the confident information of an HDD's health due to the measurement noise. On the other hand, anomalies are not the evidence of HDD failure, but the evidence that the HDD system suffers certain deviation from normal condition to some extent. Third, the computations of SVM and HMM are too expensive, making the online application to be impracticable.

The objective of this study is to develop an effective failure prediction method, which can predict the impending failures of HDDs with high prediction accuracy, low FAR, and sufficient backup time, as well as feasibility for online application. To this end, a two-step parametric (TSP) method was developed. This method takes the advantage of the nature of the data in both healthy and failed drives instead of directly labeling the individual observations to health or fail. The aim is to deal with the uncertainty of the labels of individual observations that the standard supervised learning methods cannot handle. During model training, the individual observations in a drive were treated as a time series. Through tracking the evolution features, the deviation/degradation information of this drive can be extracted. Making use of such information in conjunction with the prior knowledge of the drives allows us to develop the decision boundary. Labels of individual observations can be inferred as a result. This manipulation avoids the confusion effect that exists in SVM and HMM, resulting in the training model more effective to distinguish the healthy and failed drives.

To obtain a reliable prediction with minimum false alarm, the failure prediction problem is divided into two steps for deeply making use of the information of the data. The first step is to

measure the HDD deviations from the health state such that each observation is indexed as the deviation/distance rather than the labels. The deviation information is, however, not adequate to demonstrate the degradation of an HDD due to the measurement error or some unknown system behaviors. Therefore, the second step tracks the anomaly progression via a temporal probabilistic model that can make use of not only the current observation but also its neighbors, and thus, a confident prediction with minimum false alarms can be obtained as a result.

In practice, the accuracy of discrimination is not all in failure prediction in HDDs. To provide the sufficient time to the users for backing up the data is also a critical task. The traditional methods, however, have not well designed for this problem. To resolve it, a new cost function incorporating an index called alarm distance, which measures the duration between alarm point and the real failure point, is developed in this paper. By minimizing this cost function, the optimal model parameters with potential capability of providing an early warning can be selected.

In addition, the algorithms used in the TSP method are not computationally complex, enabling the proposed method practical for online application. For example, Mahalanobis distance is used to transform the multivariate data into univariate data, which significantly improves the computational efficiency. The differential evolution algorithm is also applied for quickly finding the optimal parameter setting through minimizing the cost function.

For the purpose of performance evaluation, our TSP method was applied for failure prediction in both synthetic and real data set. The results show TSP method outperformed the state-of-the-art methods, such as SVM and HMM in both prediction accuracy and computational efficiency.

The rest of this paper is organized as follows. In Section II, some related works are briefly reviewed. In Section III, the developed approach and relevant algorithms are presented. A case study is conducted in Section IV. Section V gives the corresponding results, which are also compared with the state-of-the-art methods. Section VI makes some concluding remarks.

## II. RELATED WORK

Numerous algorithms have been developed for anomaly detection and failure prediction. Their applications range from image/video retrieval [15], [16], bearings' fault detection [17], [18], to the financial crisis prediction [19]. Most of these applications are focused on the short-term failure prediction. There is also another group of failure prediction focusing on the long-term prediction [20]–[22]. But its prediction accuracy highly relies on the physics-of-failure model (or called degradation model). In our application, an HDD failure may be caused by many mechanisms, making the difficulty in deriving a physics-of-failure model. On the other hand, the purpose of our HDD failure prediction is to provide a short-term (i.e., 24 h [7]) prediction that can provide a sufficient margin for the users to back up their data, long-term prediction seems not necessary. The thorough surveys of the prediction methods can be found in [14], [15], and [23]. In the following subsections, we discussed distance measure and generalized likelihood ratio test, which are related to our work.

### A. Distance Measure and Anomaly Threshold Determination

In literature, a useful way of measuring deviations from the normal condition with low computational cost is to use distance measurement. Among distance measurements, Mahalanobis distance (MD) is superior in the scaling effect and correlation among the variables compared to Manhattan distance, Euclidean distance, and Hamming distance [24]. Therefore, MD is widely used in the health monitoring of electronic products [25], [26]. The implementation of MD can also help to establish the useful anomaly measures for the prognostic evaluation [25].

Although MD is a useful index to measure the deviations of system, the threshold for distinguishing the anomalous and healthy events cannot be directly determined due to the nonparametric nature of MD values. A solution of this problem is to use the power transformation to transform the MD values into parametric variables. Box–Cox transformation is one of the most popular power transformations due to its competency to Gaussianize data [27]. In this paper, the commonly used limit, mean plus three standard deviations, in Gaussian distribution is used as the threshold to determine the anomalies after the transformation of MD values. A preliminary conference version related to this part has appeared in [28].

### B. Generalized Likelihood Ratio Test (GLRT)

To detect whether there is a significant change in the data stream, a likelihood ratio test (LRT) is generally utilized. LRT is based on a hypothesis testing that exploits prior distribution of data to decide the population of a candidate set [29]. For a binary case, there are two hypotheses:

$H_0$ : No change

$H_1$ : Presence of a change

The conditional probability density function can be written as  $p(x|H_i)$  when hypothesis  $H_i$  is true, where  $i = 0, 1$ . Given a new set that consists  $N$  independent and identically distributed observations  $x = \{x_1, \dots, x_N\}$ , the likelihood ratio  $L(x)$  is defined as [29]

$$L(\mathbf{x}) = \prod_{k=1}^N \frac{p(x_k|H_1, \Theta_1)}{p(x_k|H_0, \Theta_0)} \quad (1)$$

where  $\Theta_i$  are the model parameters of the conditional probability density function  $p(x|H_i)$  for  $i = 0, 1$ .

The classical LRT assumes the model parameters are known. However, in practice, these parameters are not always available, and thus, a generalized likelihood ratio test (GLRT) [29] is used in this paper. It is defined as

$$G(\mathbf{x}) = \prod_{k=1}^N \frac{p(x_k|H_1, \hat{\Theta}_1)}{p(x_k|H_0, \hat{\Theta}_0)} \quad (2)$$

where  $\hat{\Theta}_i$  are the maximum likelihood estimates of  $\Theta_i$ .

The GLRT is based on Neyman–Pearson theorem, which allows the maximization of the probability of detection. In Section III, a sliding-window-based GLRT is proposed to compare the observed and expected occurrence of anomalies within a fixed-width sliding window. When the GLRT of a window exceeds the prespecified failure threshold, a failure can be predicted.

## III. TWO-STEP PARAMETRIC (TSP) METHOD

In this section, the framework of failure prediction using TSP method (Fig. 1) is presented. In Subsection A, a failure modes, mechanisms, and effects analysis (FMMEA) is conducted on the HDD system to determine the relevant performance parameters, called features. In Subsection B, MD is used to compress the collected SMART data from multiple variants into one index to measure the system deviations. Box–Cox transformation is utilized to transform the MD values into normally distributed variables. Mean plus three standard deviations is used as the threshold for anomaly detection. After the anomalies are detected, an anomaly counter is triggered and a sliding window begins to track the progression of anomalies. The occurrence of the anomalies in the window is treated as a Bernoulli process. GLRT is used as a predictor to predict the impending failure of HDDs. When the impending failure is predicted, a warning will be issued. Further details on failure prediction are given in Subsection C. A parameter tuning process is used to determine the failure threshold and window size before the testing phase through minimizing a cost function, as discussed in Subsection D.

### A. Feature Selection by FMMEA

FMMEA is a systematic method for analyzing the physics-of-failure of a system [30]. During implementation of FMMEA, the potential failure mechanisms and models for potential failure modes can be identified using accelerated life test, numerical stress analysis, past experience, and engineering judgment. The criticality of the failure mechanisms can be quantified by their risk priority number (RPN). The RPN is determined by the product of the failure occurrence and severity of each mechanism. According to the values of RPNs, failure mechanisms can be prioritized. Then, the failure precursors/performance parameters (in this paper also called features) can be selected according to the critical failure modes and failure mechanisms.

In our early conference version [31], FMMEA was applied to identify potential failure mechanisms and failure modes, as well as the critical features of HDDs. The head-disk interface and head stack assembly were identified as the most vulnerable components. The features that are highly correlated with the health of head-disk interface and head stack assembly were selected as critical. The identified failure mechanisms are mainly focused on wear, overstress, and resonance. According to these mechanisms, the key features can be determined as: flying height, read/write errors, seek errors, and servo errors. The features may vary among the HDD manufacturers due to the different meanings and interpretations. For the SMART data used in our paper, the key features are listed in Section V.

### B. Anomaly Detection

This section is to present an anomaly detection method, which distinguishes anomalous behaviors from healthy events in each drive. Because a preliminary work has been reported in an earlier conference version [28], we only give a brief introduction to the anomaly detection.

The training data is a subset of healthy data, denoted as  $\mathbf{X}$ . The columns of  $\mathbf{X}$  are attributes (features) denoted as  $\mathbf{X}_j$ . To

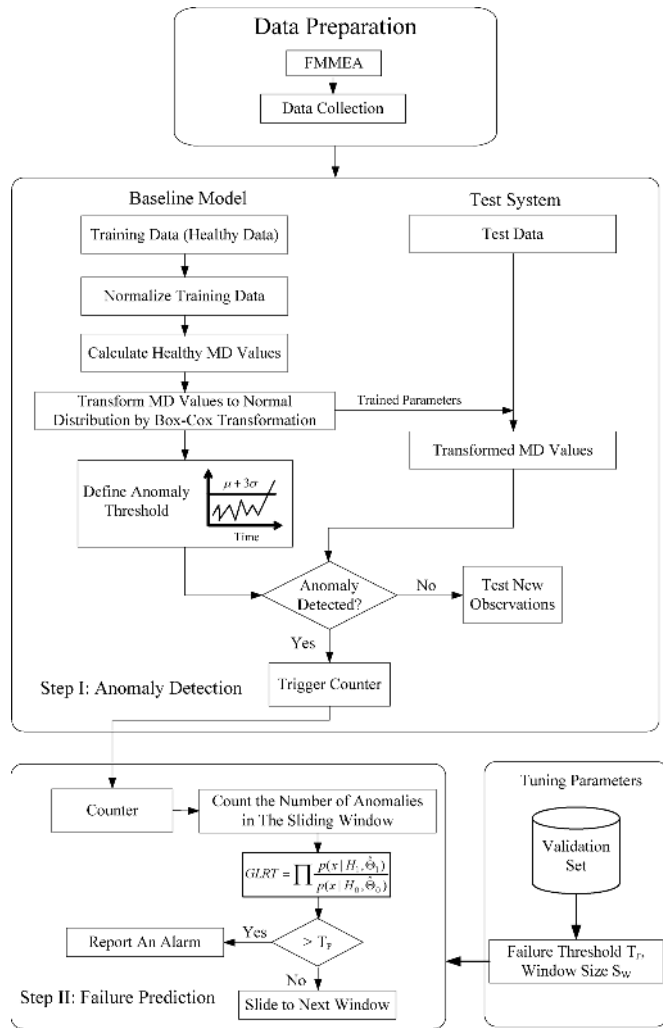


Fig. 1. Flowchart of failure prediction using TSP.

eliminate the scale effect among the attributes, each individual attribute is normalized by

$$z_{ij} = \frac{(x_{ij} - \bar{X}_j)}{S_j} \quad (3)$$

where  $x_{ij}$  denotes the  $i$ th observation at the  $j$ th attribute  $i = 1, 2, \dots, m$ ,  $m$  is the number of observations;  $j = 1, 2, \dots, n$ ,  $n$  is the number of attributes.  $\bar{X}_j$  and  $S_j$  are the mean and standard deviation of the vector  $\mathbf{X}_j$ , respectively.

The MD value of each observation is given by

$$MD_i = \frac{1}{n} \mathbf{z}_i \mathbf{C}^{-1} \mathbf{z}_i^T \quad (4)$$

where  $\mathbf{z}_i = [z_{i1}, z_{i2}, \dots, z_{in}]$ ,  $\mathbf{z}_i^T$  is the transpose of  $\mathbf{z}_i$ , and  $\mathbf{C}$  is the correlation matrix of the attributes.

To transform the MD values into normally distributed variables, the Box-Cox transformation [27] is utilized

$$y(\text{MD}, \lambda) = \begin{cases} \frac{\text{MD}^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln \text{MD}, & \lambda = 0. \end{cases} \quad (5)$$

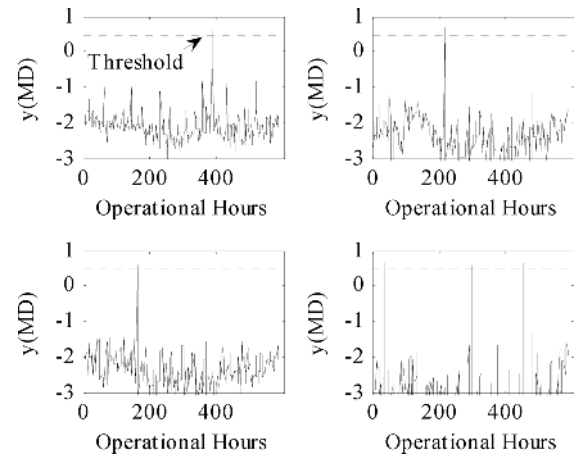


Fig. 2. Observations (transformed MD) in healthy drives.

where  $\lambda$  is the transformation parameter that can be optimized by maximizing the logarithmic likelihood function

$$f(\text{MD}, \lambda) = -\frac{N}{2} \ln \left[ \frac{\sum_{i=1}^N (y(\text{MD}_i, \lambda) - \bar{y}(\text{MD}, \lambda))^2}{N} \right] + (\lambda - 1) \sum_{i=1}^N \ln(\text{MD}_i)$$

where  $N$  is the sample size of the training data and  $\bar{y}(\text{MD}, \lambda)$  is mean value of  $y(\text{MD}, \lambda)$ .

To detect the anomalies, we use the mean plus three standard deviations ( $\mu + 3\sigma$ ), which is commonly used to identify the “out-of-control” points with a 99.7% confidence level [32], of the transformed MD values to be the anomaly threshold. During the test process, if the transformed MD values exceed this threshold, the observations will be determined as anomalies.

### C. Failure Prediction

When anomalies occur in an HDD, failure may not happen immediately. This can be explained by two folds. First, the anomalies are indeed the symptoms that the HDD has deviated from the healthy state to some extent rather than the symptoms of failure. The failure does not happen until the anomalies propagate to a certain level. Second, the measured anomalies may be resulted from measurement noises. These noises are generally randomly distributed no matter the HDD is in a healthy state or faulty state. Fig. 2 shows the transformed MD values in the healthy drives during 600 operational hours. It can be seen from this figure, although the anomalies occur, they are randomly and sparsely distributed. In contrast, the anomalies in failed drives (Fig. 3) are intensely distributed throughout the operating hours.

This fact indicates that the drive approaches failure when the anomalies frequently occur. Motivated by this fact, we use a sliding-window-based GLRT method to track the anomaly progression and predict the HDD failure. The main idea is that when the occurrence of anomalies within a certain time interval (sliding window) demonstrates a significantly increase

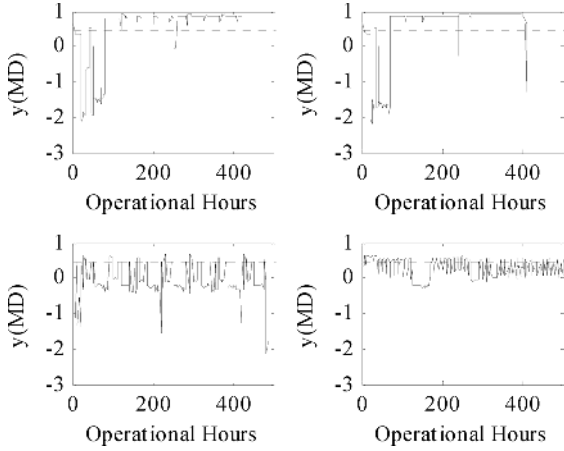


Fig. 3. Observations (transformed MD) in failed drives.

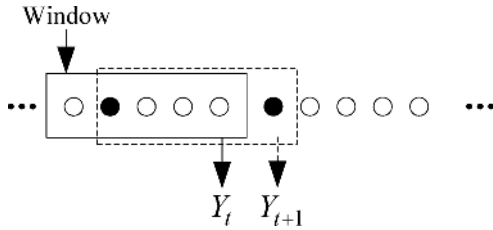


Fig. 4. Principle of sliding window. Open circle stands for normal event, solid circle stands for anomaly event.

compared to the expected value, implying the drive approaches failure.

The procedure of the sliding-window-based GLRT method is proposed as follows:

- 1) Through comparing with the anomaly threshold,  $\mu + 3\sigma$ , the transformed MD values in the drives can be converted into binary variables by (6). Let  $X_1, X_2, \dots, X_T$  be the sequence of the binary variables in a drive.

$$X_i = \begin{cases} 1, & y(\text{MD}, \lambda) \geq \mu + 3\sigma; \\ 0, & y(\text{MD}, \lambda) < \mu + 3\sigma. \end{cases} \quad (6)$$

where  $i = 1, 2, \dots, T$ , “0”, denotes normal event and “1” denotes anomaly happens.

- 2) A sliding window (Fig. 4) with window size  $S_w$  is used to scan all binary variables in the drives. The number of anomalies in the window is denoted as  $Y_t$ .

$$Y_t = \sum_{j=t-S_w+1}^t X_j \quad (7)$$

where  $t$  denotes each time in the drive’s life  $T$ ,  $S_w \leq t \leq T$ . For example, at  $t$ , there is one anomaly in the window, denoted as  $Y_t = 1$ . At  $t+1$ , there are two anomalies in the window, denoted as  $Y_{t+1} = 2$ .

- 3) In the healthy drives, the probability of each sample to be an anomaly is denoted by  $P(X_i = 1) = p_i$ , the  $X_s$  can be thought as a sequence of Bernoulli trials, and this process as a Bernoulli process [33]. The null is the probability of an anomaly in a window in the healthy drives

$$P(X_i = 1) = p_i \quad \text{for } i = t - S_w + 1, \dots, t \quad (8)$$

For some unknown  $t$ , the pulse odds-ratio alternative is [34]

$$P_A(X_i = 1)/P_A(X_i = 0) = \begin{cases} \theta p_i / (1 - p_i), & i = t - S_w + 1, \dots, t; \\ p_i / (1 - p_i), & \text{otherwise.} \end{cases} \quad (9)$$

The logarithm of generalized likelihood ratio for the hypothesis test can be written as

$$G_t = \max_{\theta_t} \left\{ Y_t \ln \theta_t - \sum_{i=t-S_w+1}^t \ln(1 - p_i + \theta_t p_i) \right\} \quad (10)$$

Take the derivative of the expression of  $G_t$ ,  $\hat{\theta}_t$  can obtain by solving

$$Y_t / \hat{\theta}_t - \sum_{i=t-S_w+1}^t p_i / (1 - p_i + \hat{\theta}_t p_i) = 0.$$

- 4) Given a training/validation set consisting of  $l$  drives, the prior knowledge of the drives are the final status, denoted as  $\Omega_i \in \{0, 1\}$ ,  $i = 1, \dots, l$ . The training/validation drives can be written as  $\{(\mathbf{G}_1, \Omega_1), \dots, (\mathbf{G}_l, \Omega_l)\}$ , where drive  $i$  consists of  $m$  generalized likelihood ratios  $\mathbf{G}_i = \{G_{i1}, \dots, G_{im}\}$ . Given a failure threshold  $T_F$ , the labels  $\psi_{ij} \in \{0, 1\}$  can be assigned to  $G_{ij}$  in drive  $i$  by

$$\psi_{ij} = \begin{cases} 1, & G_{ij} \geq T_F \\ 0, & G_{ij} < T_F. \end{cases} \quad (11)$$

The health status of this drive can be estimated by  $\Omega_i = \max_j(\psi_{ij})$ . The generalized likelihood ratio in a drive first hits the  $T_F$  determines the alarm time  $F_T$ . The optimal  $S_w$  and  $T_F$  were obtained by minimizing the cost function, which will be discussed in detail in Subsection D.

- 5) After the optimal  $S_w$  and  $T_F$  are obtained, the incoming new drives will be tested. If  $G_{ij}$  in a testing drive is greater than  $T_F$ , a failure warning is reported to the users. Otherwise, anomaly tracking continues.

#### D. Cost Function and Parameters Tuning

For manufacturers, cost saving is a very important factor that needs to consider. Apparently, the prediction errors can lead to monetary cost. For example, if a failed drive is not detected, the data in the drive may be lost. If a healthy drive is false alarmed, the warranty issues will be caused. Different parameter (i.e.,  $S_w$  and  $T_F$  in this paper) settings may incur different prediction errors, and hence different monetary costs. Therefore, minimizing the cost function is a way to optimize the relevant model parameters.

A commonly used cost function can be expressed as [35]

$$E(\text{Cost}) = w_{\text{FN}} \cdot \text{FN} \cdot p^+ + w_{\text{FP}} \cdot \text{FP} \cdot p^- \quad (12)$$

where  $p^+$  is the ratio of the number of failed products to the number of total products,  $p^-$  is the ratio of the number of healthy products to the number of total products. FN is miss detection, also known as false negative. It is a measure that a product is determined to be healthy when it is not. FP is false alarm, also known as false positive. It predicts a product as failed when it is in fact healthy.  $w_{\text{FN}}$  is the weight that associates with miss

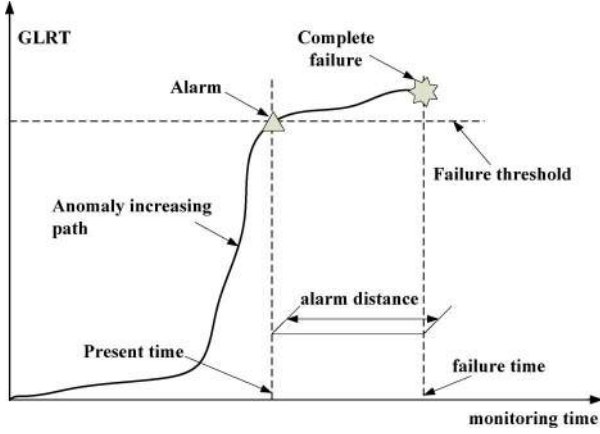


Fig. 5. Alarm distance.

detection,  $w_{FP}$  is the weight that associates with false alarm. As previously discussed in this paper, HDD manufacturers are very concerned with FAR due to the warranty issues. Therefore, HDD manufacturers can place a larger  $w_{FP}$  to penalize false alarms.

For our HDD problem, the expected cost can be minimized by adjusting the parameters ( $S_w, T_F$ )

$$\begin{aligned} & \underset{T_F, S_w}{\text{minimize}} \{E(\text{Cost})\} \\ & \text{subject to: } w_{FP} + w_{FN} = 1 \\ & \quad 0 \leq T_F \leq \max_i \{G_i\}, i = 1, \dots, l \\ & \quad 1 \leq S_w \leq N_{\text{sample}} \end{aligned} \quad (13)$$

where  $E(\text{Cost})$  is the implicit functions of  $S_w$  and  $T_F$ ,  $G_i$  is the likelihood ratios in  $i$ th drive,  $l$  is the number of validation drives.  $N_{\text{sample}}$  is the number of samples in a HDD. Because only the asymmetry of the weights have an influence on the final results, the weights,  $w_{FP}, w_{FN}$ , were constrained by  $w_{FP} + w_{FN} = 1$ . Manufacturers can use these weights to multiply the unit cost to get the total costs. For simplicity, this paper assumes the unit cost to be one, which does not affect the analysis.

In failure prediction problem, the prediction accuracy is not all. To provide the sufficient time to the users for backing up their data is also a critical task. To cope with this problem, we introduce an index called alarm distance that measures the duration between alarm point and the real failure point (Fig. 5). The alarm time is determined by the time when the likelihood ratio  $G_t$  in a window hit the failure threshold. The real failure time is the prior knowledge of the drives. At the real failure time, the drive cannot work at all. A larger alarm distance means the users have more time to back up their data.

To obtain the parameters with the potential capability of providing an earlier warning (a larger alarm distance), we incorporate the averaged alarm distance into the cost function (12). To our best knowledge, this is the first time to incorporate the alarm distance into the cost function for optimizing the model parameters. The optimal alarm distance may, however, conflict with the optimal prediction accuracy. For example, a lower threshold may provide longer alarm distance but may increase the false alarms as well. To avoid damaging the prediction accuracy that is the primary objective of this study, we placed a small weight

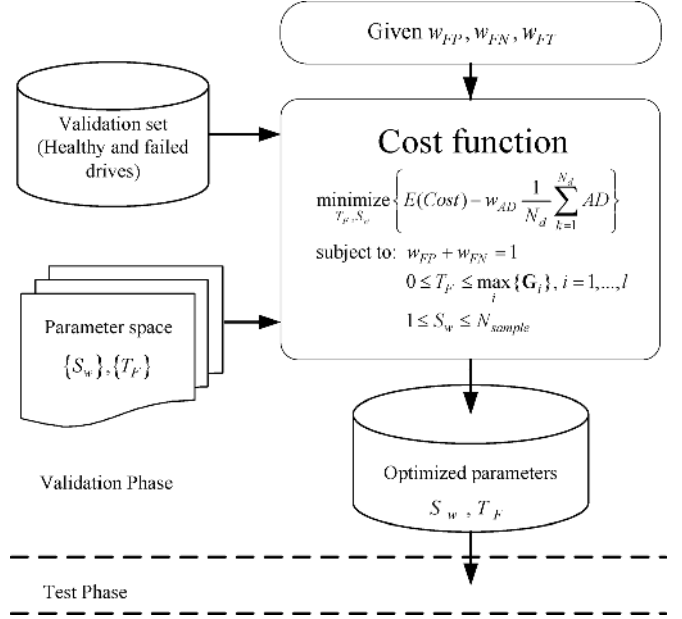


Fig. 6. Illustration of the parameter tuning in the validation process.

on the averaged alarm distance. After some manipulations, (13) becomes

$$\begin{aligned} & \underset{T_F, S_w}{\text{minimize}} \left\{ E(\text{Cost}) - w_{AD} \frac{1}{N_d} \sum_{k=1}^{N_d} AD \right\} \\ & \text{subject to: } w_{FP} + w_{FN} = 1 \\ & \quad 0 \leq T_F \leq \max_i \{G_i\}, i = 1, \dots, l \\ & \quad 1 \leq S_w \leq N_{\text{sample}} \end{aligned} \quad (14)$$

where AD is the measured alarm distance,  $w_{AD}$  is the weight of  $1/N_d \sum_{k=1}^{N_d} AD$ , and  $N_d$  is the number of failed drives in the validation set.

Fig. 6 shows the illustrations of the parameter tuning. The core of this tuning process is to solve (14). We used the differential evolution algorithm [36], which is a population-based algorithm similar to genetic algorithm, to do this work. This algorithm includes three operators: mutation, crossover and selection. Due to the high efficiency, effectiveness, and robustness in optimization, it has been widely used in many practical applications: fault diagnosis, digital filter design, and neural network training [37]. Alternative algorithms including genetic algorithm, particle swarm optimization, and ant colony optimization may also be able to solve this problem. The key parameters in implementing differential evolution are scale vector  $F$ , crossover rate CR, and the number of population NP. Storn *et al.* [36] suggested some choices: (1)  $F \in [0.5, 1]$ ; (2)  $CR \in [0.8, 1]$ ; (3)  $NP = 10 D$ . Where  $D$  is the dimension of the parameters that need to optimize. In this paper, we choose the parameters as follows:  $F = 0.8$ ,  $CR = 0.8$ ,  $NP = 20$ , and a maximum generation (iteration) number  $G_{\text{max}} = 10$ .

#### IV. PERFORMANCE EVALUATION

To evaluate the failure prediction performance provided by our method, two performance metrics, receiver operating characteristic (ROC) curve and alarm distance, are introduced.

### A. Receiver Operating Characteristic (ROC)

ROC curve is to evaluate the prediction accuracy. It is a tradeoff plot between FDR and FAR [38]. FAR is defined by the ratio of the false alarmed drives to the total healthy drives.

$$\text{FAR} = \frac{n_{\text{FP}}}{n_H} \quad (15)$$

where  $n_{\text{FP}}$  is the number of false alarmed drives and  $n_H$  is the number of total healthy drives.

FDR is defined by the ratio of detected failed drives to the total failed drives.

$$\text{FDR} = \frac{n_{\text{CP}}}{n_F}. \quad (16)$$

where  $n_{\text{CP}}$  is the number of detected failed drives and  $n_F$  is the number of total failed drives.

### B. Alarm Distance

This metric is used to guarantee the users have enough time to back up their data. The goal of SMART designed by HDD manufacturers is to provide a warning at least 24 h before drive's failure [7].

## V. RESULTS AND DISCUSSION

In this section, the experimental results are presented. The state-of-the-art methods, including SVM and HMM, are compared with our developed method using both synthetic data and real-life data.

### A. Synthetic Data Set

The synthetic data is designed to simulate the data in reality, which consists of the time series corresponding to the product evolves from health to fail. The data may be also non-Gaussian distributed and highly noised. To simulate the non-Gaussian characteristics, we used the Weibull distribution to generate the data. The alternative distributions include exponential distribution and lognormal distribution. The synthetic data set contains 600 non-Gaussian time series, where 300 time series stand for healthy products, 300 time series stand for the failed products. Each time series contain 500 observations. For ease of understanding, we defined the time interval between two observations is one hour. The parameters of the two classes are given in Table I. The parameters— $k$  and  $\lambda$ —are the shape and scale parameter of Weibull distribution respectively. For the healthy products, we used a mixture of several sets of Weibull distributions to simulate the products suffering different states and with high measurement noises. For the failed products, we used two sets of Weibull distributions to simulate the failed products evolving from health to fail.

We compared with SVM and HMM using the synthetic data first. All experiments were repeated ten times, the results were then averaged over these ten runs. In TSP method, we randomly

TABLE I  
PARAMETERS OF SYNTHETIC DATA

Data sampled from Weibull distribution	
$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0; \\ 0 & x < 0. \end{cases}$	
Parameters of healthy data	Parameters of failed data
$\lambda = 0.5, k = 0.75$ $\lambda = 1, k = 0.75$ $\lambda = 0.5, k = 1$	Healthy status: $\lambda = 0.5, k = 1$ Anomalous status: $\lambda = 3,$ $k = 1$

selected 60% of the healthy time series as the training set, and used the remaining data as the testing set. To obtain the ROC curve, the failure threshold was varied to obtain the different pairs of FAR and FDR.

For implementation of SVM, we used mySVM to perform the experiments, in which the SVM model is similar to [9]. Radial kernel function was used in the model. We randomly selected 60% of the total failed and healthy time series as the training sets for positive and negative models, respectively, and treated the remaining data as the testing set. A time series is predicted to be failed if any of its observation is predicted as positive. The parameters of the model were varied to obtain the ROC curve.

For implementation of HMM, we used discrete HMM model which is similar to [13]. The selection of training and testing set was similar to SVM. Binning method was used to discretize the data. The number of bins was set to 10. More bins did not improve the prediction performance. The number of states was set to be 30. The time series were segmented to different sequence segments by a fixed-size window. The positive model was trained with sequence segments, which were randomly sampled from the failed time series. The negative model was trained with sequence segments of healthy time series, also randomly sampled. A time series is predicted as failed if any of its sequence segments of certain consecutive observations over time is predicted as failed. The threshold to determine failure is the difference of the sequence log-likelihoods that were calculated over positive and negative models. We also varied this threshold to obtain the ROC curve.

The comparison among three methods is shown in Fig. 7. For overall performance, SVM performed worst, TSP performed best. In the lower FAR region, SVM and HMM with window size 10 show nearly no detected failures. Conversely, TSP obtained much better results in this region, especially when the window size was set to 50 (84% FDR). It is also interesting that at 0% FAR, TSP with window size of 100 could not perform better than that when the window size was set to 50, indicating the prediction accuracy did not monotonically increase with window size.

In terms of alarm distance, we used the averaged alarm distance to compare the performance. TSP still got the best performance. In particular, TSP can achieve more than 100 h of averaged alarm distance when the window size is set to 50. SVM



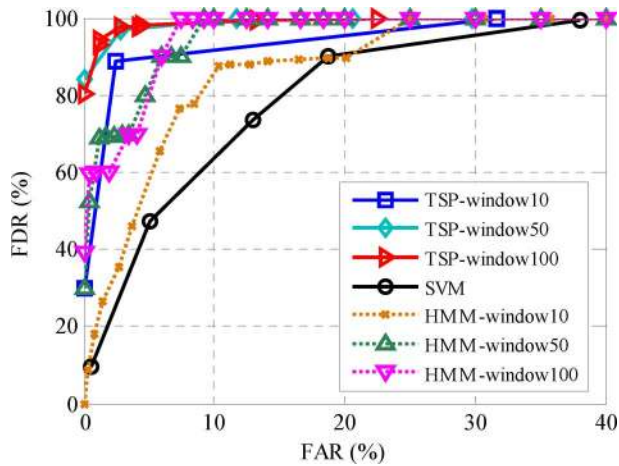


Fig. 7. ROC performance comparison of TSP, SVM, and HMM.

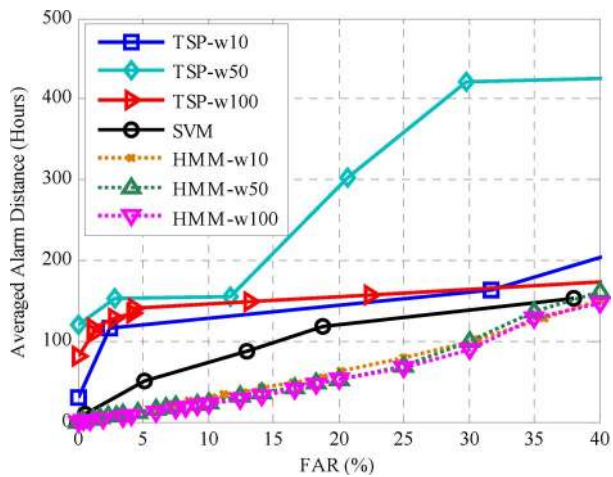


Fig. 8. Alarm distance performance comparison of TSP, SVM, and HMM.

shows better performance than HMM. Changing the window size shows nearly no effect on the HMM results.

### B. Real-Life HDD Data

The real-life data set was acquired from the real HDDs, which includes 369 drives from one model, where 178 drives were labeled as healthy and 191 drives were labeled as failed [39]. The healthy drives were from the manufacturer. The failed drives were field returned. Each drive contains 300 samples (observations) at most, and the interval between two samples is 2 h. In other words, only the last 600 h of data were recorded. When the time exceeded 600 h, the data were overwritten. There are some failed drives with less than 300 samples because they were not able to survive 600 h of operation. Each sample contains 60 performance-monitoring attributes, in addition to other attributes such as the drive's serial number and total power-on-hours. Not all attributes were monitored in every drive. The neglected attributes were set to constant.

1) *Data Processing and Selected Features:* A preliminary examination was executed before using the data. Attributes with

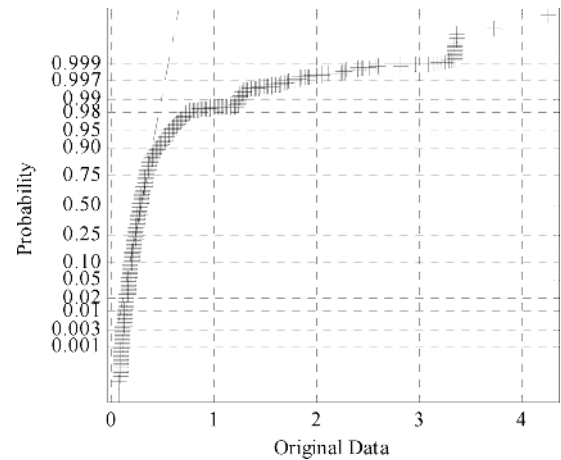


Fig. 9. Normal probability plot for the training data (MD).

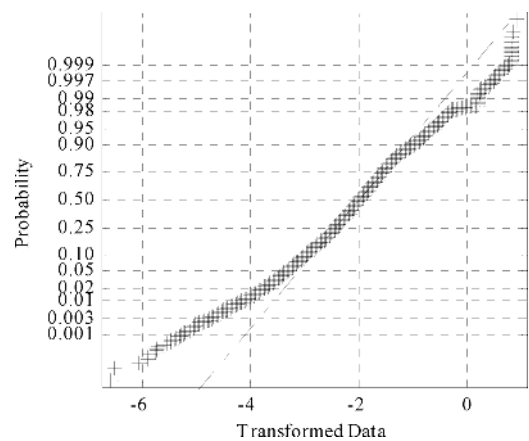


Fig. 10. Normal probability plot for the transformed values of the training data (MD).

all zeros were abandoned. The 47 critical features<sup>1</sup> were selected by FMMEA. Detailed information can be found in [31].

2) *Implementation of TSP:* During implementation of TSP, 60% of the healthy HDDs were randomly selected to build the baseline Mahalanobis space. The normal probability plot was used to check whether the MD values are subject to the normal distribution or not. The original MD values and the data transformed by Box-Cox transformation are shown in Figs. 9 and 10, respectively. The original data (Fig. 9) are far from the normal distribution, while the transformed data are nearly normally distributed.

We investigated the influence of different parameter settings on FDR, FAR and alarm distance using the rest 40% healthy drives and all failed drives, as illustrated in Figs. 11–13. With the increase of failure threshold, FDR almost linearly decreases (Fig. 11). Lower threshold can predict more failures, but can increase FAR as well. The linear decrease of FDR indicates that the failures are nearly uniform distributed. It was also found that FDR is insensitive to window size (the FDR difference between window size 10 and window size 50 is less than 5%). In contrast, FAR is sensitive to window size (Fig. 12). It can

<sup>1</sup>FlyHeight 1–16, GList 1–3, PList, Reads, Writes, ReadError 1–12, ReadError 18–20, Servo 1–3, Servo 5–10, and WriteError.



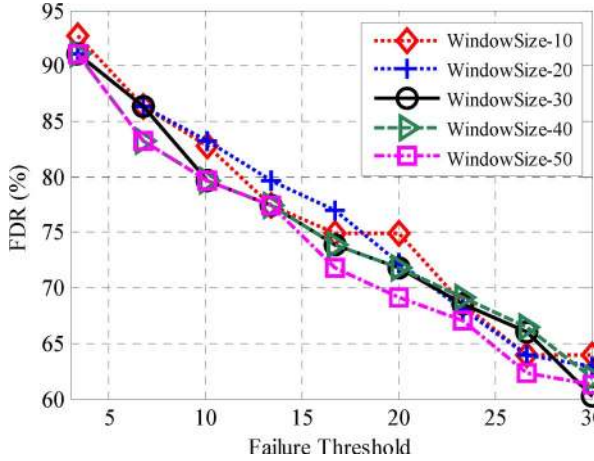


Fig. 11. FDR as a function of failure thresholds for different window size.

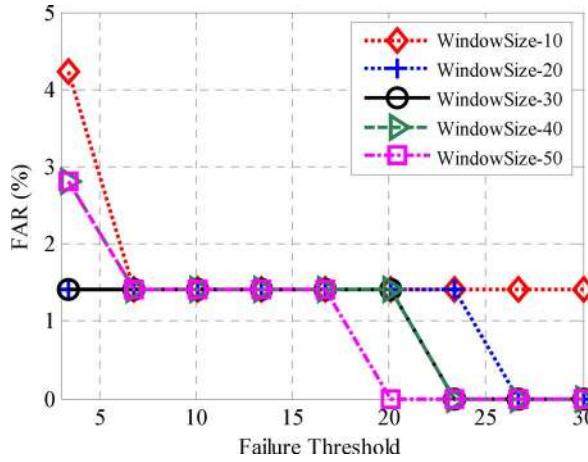


Fig. 12. FAR as a function of failure thresholds for different window size.

be explained that for a short window, only few data points at a time are considered, small differences in these data points, such as measurement noise could result in misclassifying these data as an instance of failure. A larger window can capture the stable change in the data points, and thus can prevent the false alarms effectively. Alarm distance also shows to be sensitive to window size (Fig. 13). Either larger window size or higher failure threshold could result in a time delay on prediction, but better FAR. The tradeoff between real-time prediction and the accuracy of prediction leads to the difficulty in determining a good set of the parameters ( $S_w$  and  $T_F$ ), and therefore, a tuning process is needed.

3) *Parameter Optimization*: To tune the parameters, 60% of the failed drives were randomly selected, which together with the 60% training drives (healthy) were used as the validation set. The rest 40% healthy drives and 40% failed drives were used as the testing set. By varying the weights ( $w_{FP}$  and  $w_{FN}$ ) in cost function, the different FAR, FDR, and alarm distance can obtain. The weight  $w_{AD}$  of  $1/N_d \sum_{k=1}^{N_d} AD$  in this experiment was set at  $e^{-5}$ . The experiment was also repeated 10 times to get an average value. Three sets of weights in the cost function were compared in Table II. It is found that at 0% FAR, TSP method can achieve 68.42% FDR. Table III shows the prediction performance in terms of alarm distance at different  $w_{FP}$ . Concerning

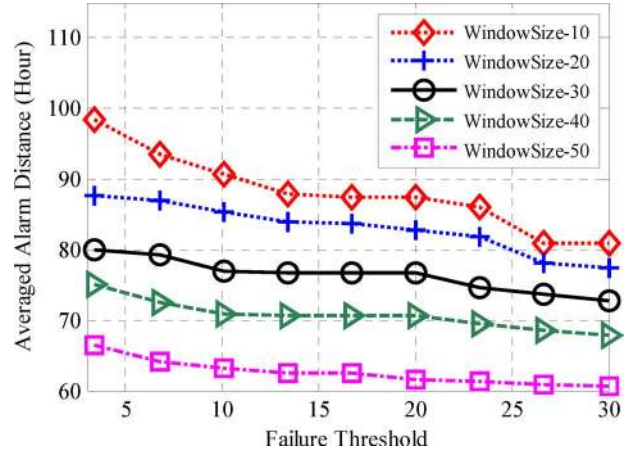


Fig. 13. Averaged alarm distance as a function of failure thresholds for different window size.

TABLE II  
PREDICTION PERFORMANCE WITH DIFFERENT  
WEIGHTS IN COST FUNCTION

$w_{FN}$	$w_{FP}$	FAR (%)	FDR (%)
0.05	0.95	0	68.42
0.10	0.90	1.41	90.00
0.50	0.50	4.23	95.53

TABLE III  
ALARM DISTANCE DISTRIBUTIONS OF FAILED DRIVES  
WITH DIFFERENT WEIGHTS

AD* (hours)	$w_{FP} = 0.95$	$w_{FP} = 0.90$	$w_{FP} = 0.50$
	Percentage (%)		
Non-detect	31.58	10.00	4.47
>0	68.42	90.00	95.53
>12	50.00	72.76	75.39
>24	46.05	63.95	65.26
>36	43.42	61.97	63.29
>48	32.89	49.34	52.63

\* AD denotes alarm distance

the goal of SMART is to provide at least 24-h alarm distance, 46.05% drives can meet this requirement at  $w_{FP} = 0.95$  (0% FAR). If placing a lower weight to false alarm, more drives can reach that goal, but the FAR increases.

4) *Comparative Analysis*: In this section, we compare our results to a set of methods that were mentioned in Section I. First, we focused on the SVM and rank-sum schemes, which were proposed by Murray *et al.* in [9]. The SVM-based method using the radial kernel function has the best performance (50.6% FDR with 0% FAR) among those studies; rank-sum-based method has second-best performance (52.8% FDR with 0.7% FAR). Second, we considered TAN and HMM, which were developed by Tan *et al.* [12] and Zhao *et al.* [13] to improve the FDR using the same data set.

Fig. 14 shows the ROC curves obtained by different methods. The overall performance of TSP method is much better than

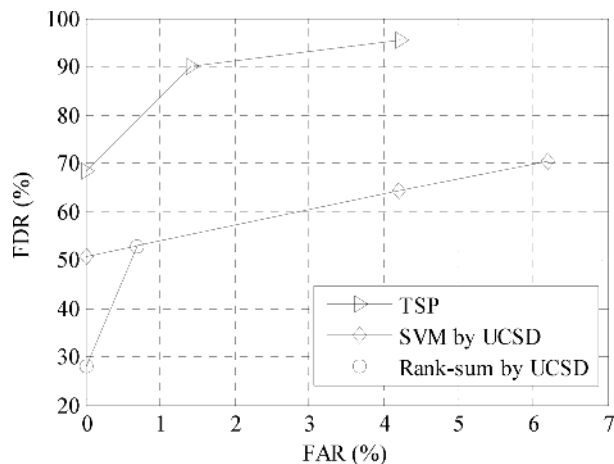


Fig. 14. ROC comparison for different approaches.

TABLE IV  
PREDICTION RESULTS COMPARISON AT 0% FAR

Scheme	FDR (%)	Drawback
TSP	68.4	Cannot realize real-time prediction.
SVM [9]	50.6	Exists confusion effect and requires longer training time.
Rank-sum [9]	28	Cannot fully use the data information.
TAN [12]	20–30	Exists confusion effect.
HMM [13]	52	Exists confusion effect and requires longer training time.

other methods. In particular, TSP got 68.4% FDR with 0% FAR. This result is about 18% higher than that of SVM in [9].

Table IV shows the prediction results at 0% FAR and the drawbacks of different methods. SVM, TAN, and HMM are supervised classification methods, which assume the observations in the failed drive are failed even those observations are measured in healthy state. This leads to the confusion effect and poor prediction performance. In contrast, TSP method takes advantage of the nature of the data in both healthy and failed drives rather than labeling individual observation, which avoids the healthy data being labeled as failed, thus improving the FAR while maintaining a high FDR. Rank-sum method does not require labeling the observations in failed drives, but it is a non-parametric method, whose disadvantage is that the data information cannot be thoroughly utilized, especially in the case that the underlying distribution of the data follows some parametric distribution. In terms of computational efficiency, TSP uses MD to compress the multidimensional data into univariate data, which requires much less computation than the multivariate classifiers, e.g., SVM.

Murray *et al.* [9] summarized the alarm distances of failed drives at 0% FAR. The results show that 48% of the failed drives could be detected with an alarm distance of more than 24 h. TSP shows comparable results. While at the region of lower than 24 h (Table III), TSP performed much better, which indicates that these failed drives were detected at the end of their lives. The FAR obtained by TAN method [12] is bigger than 5%, thus we did not compare with it. HMM [13] did not measure the alarm distance, therefore, we also did not conduct the comparison.

TABLE V  
COMPUTATION TIME

Approaches	Time consumption (Mins.)
TSP	12.6
SVM by [9]	17983 (3560 <sup>+</sup> )
HMM by [13]	128.3 (38.5 <sup>+</sup> )

\* The result is reproduced.

We also compared the computation efficiency among the methods (Table V). The computational times of SVM and HMM were also reproduced in our computational platform (Matlab 7.1 environment on an Intel Dual Core2 Processor running at 3.00 GHz, and 3.25 GB RAM). SVM shows computationally expensive, about 2.5 days. HMM shows much fast, but the result is only based on one window size (window size equals 50 observations). If includes other window sizes for optimization, the computation time will increase to many times of the reported one.

## VI. CONCLUSION

In this paper, we presented a novel way of predicting failures in HDDs using a two-step parametric (TSP) method. This method takes the advantage of the nature of the data in both healthy and failed drives by dealing with the uncertainty of the labels of individual observations. Such a manipulation avoids the confusion effect that exists in the standard supervised learning methods, such as SVM and HMM. By dividing the failure prediction into two steps, the degradation information in each HDD can be reliably tracked by a temporal probabilistic model that can make use of not only the current observation but also its neighbors, and thus, a confident prediction with minimum false alarms can be obtained as a result. In addition, to provide sufficient time to the users for backing up their data, a new cost function incorporating the alarm distance was developed to ensure that the model parameters with potential capability of providing an early warning were selected as optimal. Both synthetic and real-life data sets were used to evaluate the developed methodology.

Based on the evaluation results, TSP method outperforms previous works in predicting the HDD failures. In particular, TSP method could achieve 68.42% FDR at 0% FAR. It is about 18% improvement compared to the SVM-based approach proposed by [9]. Considering the alarm distance, TSP was able to provide a notification of 24 h in advance for 46.05% of the failed drives without any false alarm, which is comparable to the result of SVM. The computation of TSP method is also much less than SVM-based and HMM-based approach, making it a practicable way of predicting the HDD failures for online application.

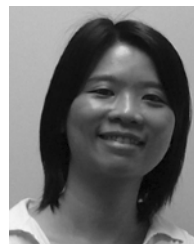
Finally, although our methodology was proposed to deal with HDD problems, it could be useful in a wide industrial area, where the products may suffer some deviations before they fail and require certain margin to take actions to prevent the products' damage.

## REFERENCES

- [1] M. Hilbert and P. Lopez, "The world's technological capacity to store, communicate, and compute information," *Science*, vol. 332, no. 60, pp. 60–65, 2011.
- [2] D. O'Connor and A. V. Zayats, "Data storage: The third plasmonic revolution," *Nat. Nanotechnol.*, vol. 5, pp. 482–483, 2010.
- [3] S. Kamarthi, A. Zeid, and Y. Bagul, "Assessment of current health of hard disk drives," in *Proc. IEEE Int. Conf. Automat. Sci. Eng.*, Bangalore, India, Aug. 2009, pp. 246–249.
- [4] B. Schroeder and G. A. Gibson, "Understanding disk failure rates: What does an MTTF of 1 000 000 hours mean to you?," *ACM Trans. Storage*, vol. 3, no. 3, pp. 8.1–8.31, Oct. 2007, article 8.
- [5] W. Jiang, C. Hu, Y. Zhou, and A. Kanevsky, "Are disks the dominant contributor for storage failures? A comprehensive study of storage sub-system failure characteristics," *ACM Trans. Storage*, vol. 4, no. 3, pp. 7:1–7:25, Oct. 2008, article 7.
- [6] R. K. Henry, "Monitoring PC hardware sounds in Linux systems using the Daubechies D4 wavelet," M.A. thesis, Dept. Comput. Sci., East Tennessee State Univ., TN, USA, 2005.
- [7] G. F. Hughes, J. F. Murray, K. K. Kreutz-Delgado, and C. Elkan, "Improved disk-drive failure warnings," *IEEE Trans. Reliab.*, vol. 51, no. 3, pp. 350–357, Sep. 2002.
- [8] E. Pinheiro, W. Weber, and L. A. Barroso, "Failure trends in a large disk drive population," in *Proc. Fifth Usenix Conf. File Storage Technol. (FAST)*, San Jose, CA, USA, 2007, pp. 17–29.
- [9] J. F. Murray, G. F. Hughes, and K. Kreutz-Delgado, "Machine learning methods for predicting failures in hard drives: A multiple instance application," *J. Mach. Learn. Res.*, vol. 6, pp. 783–816, 2005.
- [10] G. Hamerly and C. Elkan, "Bayesian approaches to failure prediction for disk drives," presented at the Eighteenth Int. Conf. Machine Learning (ICML), San Francisco, CA, USA, 2001.
- [11] J. F. Murray, G. F. Hughes, and K. Kreutz-Delgado, "Hard drive failure prediction using non-parametric statistical methods," presented at the ICANN/ICONIP, Istanbul, Turkey, 2003.
- [12] Y. Tan and X. Gu, "On predictability of system anomalies in real world," presented at the 18th Annual IEEE/ACM Int. Symp. Modeling, Analysis Simul. Comput. Telecommun. Syst., Miami Beach, FL, USA, Aug. 2010.
- [13] Y. Zhao, X. Liu, S. Gan, and W. Zheng, "Predicting disk failure with HMM- and HSMM-based approaches," presented at the 10th Ind. Conf. Adv. Data Mining. Appl. Theoretical Aspects (ICDM), Berlin, Germany, 2010.
- [14] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, 2009.
- [15] F. Salfner, M. Lenk, and M. Malek, "A survey of online failure prediction methods," *ACM Comput. Surv.*, vol. 42, no. 3, pp. 10:1–10:42, 2010.
- [16] M. Ma, D. S. Wong, S. Jang, and S. Tseng, "Fault detection based on statistical multivariate analysis and microarray visualization," *IEEE Trans. Ind. Inf.*, vol. 6, no. 1, pp. 18–24, Feb. 2010.
- [17] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
- [18] A. S. Raj and N. Murali, "Early classification of bearing faults using morphological operators and fuzzy inference," *IEEE Trans. Ind. Electron.*, vol. 60, no. 2, pp. 567–574, Feb. 2013.
- [19] W. Lin, Y. Hu, and C. Tsai, "Machine learning in financial crisis prediction: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 421–436, Jul. 2012.
- [20] W. Wang and M. Carr, "A stochastic filtering based data driven approach for residual life prediction and condition based maintenance decision making support," presented at the IEEE 2010 Prognostics Health Manage. Conf., Macau, China, 2010.
- [21] P. Lall, R. Lowe, and K. Goebel, "Extended Kalman filter models and resistance spectroscopy for prognostication and health monitoring of leadfree electronics under vibration," presented at the Proc. IEEE Int. Conf. Prognostics Health Manage., Denver, CO, Jun. 2011.
- [22] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on dempster-shafer and Bayesian Monte Carlo," *J. Power Sources*, vol. 196, no. 23, pp. 10314–10321, 2011.
- [23] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, Aug. 2011.
- [24] G. Taguchi and R. Jugulum, *The Mahalanobis-Taguchi Strategy: A Pattern Technology System*. New York, NY, USA: Wiley, May 2002.
- [25] S. Kumar, T. W. S. Chow, and M. Pecht, "Approach to fault identification for electronic products using Mahalanobis distance," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 8, pp. 2055–2064, Aug. 2010.
- [26] X. Jin, E. W. M. Ma, L. L. Cheng, and M. Pecht, "Health monitoring of cooling fans based on Mahalanobis distance with mRMR feature selection," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2222–2229, Aug. 2012.
- [27] G. E. P. Box and D. R. Cox, "An analysis of transformations," *J. R. Statist. Soc. B*, vol. 26, pp. 211–243, 1964.
- [28] Y. Wang, E. W. M. Ma, K. L. Tsui, and M. Pecht, "A fusion approach for anomaly detection in hard disk drives," presented at the IEEE 2012 Prognostics Health Manage. Conf., Beijing, China, 2012.
- [29] H. V. Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York, NY, USA: Wiley, 1968.
- [30] M. Pecht, *Prognostics and Health Management of Electronics*. New York, NY, USA: Wiley-Interscience, 2008.
- [31] Y. Wang, Q. Miao, and M. Pecht, "Health monitoring of hard disk drive based on Mahalanobis distance," presented at the IEEE 2011 Prognostics Health Manage. Conf., Shenzhen, China, 2011.
- [32] L. S. Nelson, "Technical aids," *J. Qual. Technol.*, vol. 16, no. 4, pp. 238–239, 1984.
- [33] J. Glaz, J. Naus, and S. Wallenstein, *Scan Statistics*. New York: Springer-Verlag, 2001.
- [34] J. Naus and S. Wallenstein, "Temporal surveillance using scan statistics," *Stat. Med.*, vol. 25, no. 2, pp. 311–324, Jan. 2006.
- [35] F. R. Bach, D. Heckerman, and E. Horvitz, "Considering cost asymmetry in learning classifiers," *J. Mach. Learn. Res.*, vol. 7, pp. 1713–1741, 2006.
- [36] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Opt.*, vol. 11, no. 4, pp. 341–359, Dec. 1997.
- [37] W. He, Z. Jiang, and K. Feng, "Bearing fault detection based on optimal wavelet filter and sparse code shrinkage," *Measurement*, vol. 42, no. 7, pp. 1092–1102, 2009.
- [38] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [39] G. F. Hughes, (2005, Apr.), *S.M.A.R.T. Data Set* [Online]. Available: <http://cmr.ucsd.edu/people/hughes/smart/dataset/harddrive1.zip>



**Yu Wang** (S'12) received the B.S. degree in mechanical design and manufacturing automation from the Xi'an University of Technology, Xi'an, China, and the M.S. degree in manufacturing engineering and automation from Xi'an Jiaotong University, Xi'an, China. He is currently working toward the Ph.D. degree in the Department of Systems Engineering and Engineering Management at City University of Hong Kong, Hong Kong.



**Eden W. M. Ma** (M'08) received the B.Eng. and Ph.D. degrees in electronics engineering at City University of Hong Kong, Hong Kong.

Her research interests are in data mining, clustering, feature selection and extraction, quality control, reliability, and prognostics and system health management.



**Tommy W. S. Chow** (M'93–SM'03) received the B.Sc. (First Hons.) and Ph.D. degrees from the University of Sunderland, Sunderland, U.K.

He is currently a Professor in the Electronic Engineering Department, City University of Hong Kong. He has authored or coauthored of over 200 technical papers. He is an Associate Editor of *Pattern Analysis and Applications* and the *International Journal of Information Technology*.



**Kwok-Leung Tsui** received the Ph.D. degree in statistics from the University of Wisconsin at Madison, WI, USA.

He is head and chair Professor in the Department of Systems Engineering and Engineering Management, City University of Hong Kong. Prior to the current position, he was Professor/Associate Professor in the School of Industrial and Systems Engineering, Georgia Institute of Technology from 1990 to 2011, and a research member in the Quality Assurance Center at AT&T Bell Labs from 1986 to

1990. His current research interests include health informatics, data mining and surveillance in healthcare and public health, prognostics and health management, calibration and validation of computer models, bioinformatics, process control and monitoring, and robust design and Taguchi method.

Dr. Tsui is a Fellow of American Statistical Association, a U.S. representative in the ISO Technical Committee on Statistical Methods, and a department Editor of the *IIE Transactions*.