

Reliability Estimation Based on Condition Monitoring Data

by

Yang Liu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Mechanical Engineering

University of Alberta

© Yang Liu, 2016

ABSTRACT

Reliability estimation based on condition monitoring data contains two important parts: thresholding and probability density estimation. Thresholding is to determine a critical level of an indicator corresponding to the transition of system states. Probability density estimation is to estimate the probability density function (PDF) of the indicator value over a certain time period.

Existing reliability estimation methods usually require prior knowledge such as design criteria and past experiences or event data such as time-to-failure data to determine a threshold for estimating reliability; while the information as such might be costly and impractical to acquire for expensive or highly reliable systems. Therefore, there is a demand on the reliability estimation methods which can determine thresholds and estimate probability density relying on solely condition monitoring data. However, very few studies are reported in this respect.

A method falling into this type is recently reported. The reported method jointly uses one-class support vector machine (OC-SVM) solution path for thresholding and kernel density estimation (KDE) for probability density estimation to estimate system reliability based on only condition monitoring data. This thesis studies in-depth the reported method and finds that there are four aspects that are unclear and deficient. These four aspects are thus investigated and the suggestions are provided to address the concerns. The findings of this thesis are listed as follows:

- (1) The impact of the width parameter of OC-SVM on reliability estimates is investigated and an applicable range for width parameter selection is given to acquire reasonable reliability estimates;

- (2) The impact of the bandwidth parameter of KDE on reliability estimates is investigated and an applicable range for bandwidth selection is given to acquire reasonable reliability estimates;
- (3) The impact of the sliding window size of KDE on reliability estimates is investigated. A strategy of variable window size is developed to enable stationary data to be used for probability density estimation. The results show that the proposed strategy can provide not only the reliability estimates comparable to the fixed sliding window size but also the reliability estimates corresponding to the transition of system states;
- (4) The impact of outliers is investigated and a strategy of removing outliers is developed for KDE to ensure reasonable reliability estimates can be obtained.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Ming J. Zuo for his immeasurable guidance and support through all my study and research. I have gained valuable academic training and achieved this work. My sincere thanks also go to the members of my examining committee: Dr. Yongsheng Ma and Dr. Zhigang Tian.

I would like to thank all the members of Reliability Research Lab for their kind help. I would also like to thank all the other people who have helped me during my M.Sc. program.

This thesis would not have been possible without the help and support of my husband Jian Qu. I take this opportunity to express my thanks to him not only for his love and encouragement, but also for his proofreading which improved the language of this thesis. I would like to thank my mother and parents-in-law for their understanding and consistent encouragement. My thanks also go to my children for their quietly and nicely behaviors.

TABLE OF CONTENTS

Chapter 1	Introduction	1
1.1.	Reliability Estimation	1
1.2.	Reliability Estimation Based on Event Data	2
1.3.	Reliability Estimation Based on Condition Monitoring Data	4
1.4.	Research Objectives	6
1.5.	Thesis Organization	8
Chapter 2	Literature Review	9
2.1.	Thresholding	9
2.2.	Probability Density Estimation	13
2.3.	Reliability Estimation Using Condition Monitoring Data	18
2.4.	Summary	20
Chapter 3	Fundamentals of One-Class Support Vector Machine and Kernel Density Estimation	21
3.1.	Fundamentals of One-Class Support Vector Machine (OC-SVM)	21
3.1.1.	OC-SVM Solution Path	25
3.1.2.	Reported Methods for σ Selection	26
3.2.	Fundamentals of Kernel Density Estimation (KDE)	28
3.3.	Summary	33
Chapter 4	Reliability Estimation Using One-Class Support Vector Machine and Kernel Density Estimation	34
4.1.	Data Preparation	34
4.1.1.	Simulation Data	35
4.1.2.	Experiment Data	40
4.1.3.	Data Normalization	43
4.2.	Introduction to Hua's Method	44
4.3.	Selection of the Width Parameter of OC-SVM for Reliability Estimation	46
4.3.1.	Selection of the Width Parameter Using the Enumeration Method	46

4.3.2.	Testing with Simulation and Experiment Data.....	55
4.3.3.	Summary.....	60
4.4.	Selection of the Bandwidth Parameter of KDE for Reliability Estimation.....	60
4.4.1.	Selection of the Bandwidth Parameter Using the Enumeration Method.....	60
4.4.2.	Testing with Simulation and Experiment Data.....	67
4.4.3.	Summary.....	72
4.5.	Investigation on the Sliding Window Size for Reliability Estimation	72
4.5.1.	Impact of the Sliding Window Size for Reliability Estimation.....	73
4.5.2.	Variable Sliding Window Size Based Reliability Estimation	75
4.5.3.	Summary.....	79
4.6.	Investigation on Outlier Impact for Reliability Estimation.....	80
4.6.1.	Implementation of Outlier Removal.....	80
4.6.2.	Investigation on Outlier Impact.....	81
4.6.3.	Summary.....	86
4.7.	Comparisons with Hua’s Method.....	87
4.8.	Summary	89
Chapter 5	Summary and Future Work.....	91
5.1.	Summary	91
5.2.	Future Work	93
Bibliography	95

LIST OF TABLES

Table 3.1 Sample kernel functions.....	29
Table 4.1 Observations of σ values for different RNLs.....	55
Table 4.2 The σ values calculated using reported methods for simulation data	56
Table 4.3 The σ values calculated using reported methods for experiment data.....	57
Table 4.4 Selection of a for different combinations of N_{split} and h (RNL=0.2).....	66
Table 4.5 The h values calculated using reported methods for simulation data	68
Table 4.6 The h values calculated using reported methods for experiment data.....	69
Table 4.7 Parameter settings for comparisons	87

LIST OF FIGURES

Figure 2.1 Categorization of thresholding methods.....	13
Figure 2.2 Categorization of probability density estimation methods.....	17
Figure 3.1 Separation of data using OC-SVM.....	22
Figure 3.2 Plots of sample kernel functions.....	29
Figure 3.3 Illustration of KDE using Gaussian kernel.....	30
Figure 3.4 KDE using Gaussian kernel with different bandwidth values.....	31
Figure 4.1 The plot of true values of simulation data.....	36
Figure 4.2 The plots of simulation data with different RNLs.....	39
Figure 4.3 The plots of simulation data with 1 and 5 outliers.....	40
Figure 4.4 Degradation data of water pump.....	41
Figure 4.5 Degradation data of planetary gearbox.....	42
Figure 4.6 Degradation data of bearing.....	43
Figure 4.7 Frame work of Hua's method [28].....	44
Figure 4.8 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.2).....	49
Figure 4.9 Reliability estimates with σ values from 0.001 to 0.01 (RNL=0.2).....	49
Figure 4.10 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.3).....	50
Figure 4.11 Reliability estimates with σ values from 0.01 to 0.1 (RNL=0.3).....	51
Figure 4.12 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.5).....	52
Figure 4.13 Reliability estimates with σ values from 0.001 to 0.01 (RNL=0.5).....	53
Figure 4.14 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.7).....	54
Figure 4.15 Reliability estimates with σ values from 0.001 to 0.01 (RNL=0.7).....	54

Figure 4.16 Reliability estimates for simulation data using reported σ selection methods	57
Figure 4.17 Reliability estimates for water pump using reported σ selection methods	58
Figure 4.18 Reliability estimates for planetary gearbox using reported σ selection methods	59
Figure 4.19 Reliability estimates for bearing units using reported σ selection methods	59
Figure 4.20 Reliability estimates with h values from 0.0001 to 1000 (RNL=0.2)	61
Figure 4.21 The PDF for each sliding window with h values from 0.0001 to 1000 (RNL=0.2)..	62
Figure 4.22 The PDF for the first sliding window with $h=10$ (RNL=0.2)	63
Figure 4.23 Reliability estimates with different h and N_{split} (RNL=0.2)	66
Figure 4.24 Reliability estimates for simulation data using reported h selection methods	69
Figure 4.25 Reliability estimates for water pump using reported h selection methods	70
Figure 4.26 Reliability estimates for planetary gearbox using reported h selection methods	71
Figure 4.27 Reliability estimates for bearing units using reported h selection methods	71
Figure 4.28 Reliability estimates for simulation data with different sliding window sizes	74
Figure 4.29 Reliability estimates for simulation data with different sliding distances	75
Figure 4.30 Flow chart of reliability estimation using variable sliding window size	77
Figure 4.31 Water pump data with highlighted characteristic time points	78
Figure 4.32 Reliability estimates for water pump with fixed and variable sliding window sizes	78
Figure 4.33 Flow chart of implementation of outlier removal for KDE	81
Figure 4.34 Simulation data with 1 outlier (RNL=0.2)	82
Figure 4.35 PDFs with (left) and without (right) outlier removed (1 outlier)	83
Figure 4.36 Reliability estimates with and without outlier removed (1 outlier)	83
Figure 4.37 Simulation data with 5 outliers (RNL=0.2)	84
Figure 4.38 The PDFs with (left) and without (right) outlier removed (5 outliers)	84

Figure 4.39 Reliability estimates with and without outlier removed (5 outliers)	85
Figure 4.40 Water pump data with 1 additive outlier	85
Figure 4.41 Reliability estimates for water pump with and without outliers removed.....	86
Figure 4.42 Comparisons of reliability estimates using different methods	89

Chapter 1

Introduction

1.1. Reliability Estimation

Engineering systems are designed to perform sophisticated functions. For example, in the oil sands industry, heavy haulers need to carry tons of ore from one location to another and slurry pumps need to experience the medium containing high wear ingredients like sands. The components like gears and impellers in such systems are subject to cyclic stress and/or excessive wear which may lead to unexpected failures of the whole system and end up with the loss in production and profit.

To alleviate such loss, one can implement maintenance actions to prevent the occurrence of unexpected failure, in which reliability plays an important role in making maintenance decisions [1]. Reliability describes the ability of a system to work properly for a specified time period under a specified environment. The process of estimating reliability is referred to as reliability estimation; therefore, a proper execution of reliability estimation is crucial to the success of reliability based maintenance actions.

Reliability estimation methods are widely reported in existing literature [2][3][4][5][6][7] where the data used generally fall into two types, namely event data and condition monitoring data. Reliability estimation methods can also be differentiated by these two types of data. This thesis divides the existing reliability estimation methods into two classes, namely reliability estimation based on event data and reliability estimation based on condition monitoring data.

This section introduces particularly the two types of data and the relevant reliability estimation methods will be introduced in Sections 1.2 and 1.3, respectively.

Event data refer to the information of what had happened to the system such as installation, start-up, shutdown and failure, and also the information of what had been done to the system such as repair/replacement and overhaul [1]. One type of event data widely used in reliability estimation is time-to-failure data or lifetime data which refer to the length of time for an individual system from the start to the failure of operation. For example, the time-to-failure data of an LED light may be 50,000 hours from the first time use to the ultimate failure.

Condition monitoring data refer to the versatile data measured by monitoring devices such as thermometers, pressure gauges, vibration sensors and acoustic emission sensors, which include value-type data, waveform data, and multidimensional data [1]. All these types of data are time series data which are gathered at a specified time interval over a certain time period. Typical value-type data include oil analysis data, temperatures, pressures, moistures, etc. Waveform data display a pattern of waveform which typically include vibration signals and acoustic signals. Ultrasonic data and visual images which display images are typical multidimensional data.

1.2. Reliability Estimation Based on Event Data

Reliability is defined as the probability that a system will perform its intended functions satisfactorily for a specified time period under specified operating conditions [8]. Computationally, reliability is the probability that a system has a lifetime greater than a certain interested length of time. A probability density function (PDF) of system lifetime which is also called the lifetime distribution is usually adopted to estimate the reliability. Several popular lifetime distributions are lognormal, exponential and Weibull distribution [9].

In practical applications, lifetime distribution can be determined based on available information of event data among which time-to-failure data are the most widely reported [10]. First, assume that the lifetime of system be a random variable and follow a certain type of lifetime distribution, for example, Weibull distribution. Goodness of fit test may be used to select the best distribution type. Then, estimate the parameters of the selected lifetime distribution based on available time-to-failure data using parameter estimation methods [11]. Once the parameters of the lifetime distribution are determined, the reliability of system can be estimated.

Reliability estimation based on event data has been studied for decades [11][12][13][14][15][16][17]. The event data such as time-to-failure data are collected from a large group of systems that have the same characteristics (e.g. designed to achieve the same function, produced from the same batch and operated under similar working and operating conditions). The time-to-failure data of a large number of such systems are used to obtain the lifetime distribution for the entire group. The reliability of any individual system belonging to this group can thus be estimated.

However, there are some drawbacks. Reliability estimated using lifetime distribution based on event data cannot depict the specific condition of an individual system at a certain time point. For example, the reliability of a pump is estimated as 80% based on event data, but it is impossible to know which specific pumps fall into the 20% of failure probability and fail at the time when the reliability is estimated. Also, time-to-failure data are not always available especially for expensive or highly reliable systems. In the case where time-to-failure data are limited or even not available, reliability estimation based on event data is difficult to apply.

1.3. Reliability Estimation Based on Condition Monitoring Data

Due to the limitations of event data, condition monitoring data attract more and more interests in reliability estimation over the recent decade. With proper processing of the data, one could possibly obtain the reliability that reflects the ability of a specific system fulfilling its anticipated design function rather than a reliability estimated for a group of systems based on event data.

As mentioned in Section 1.1, there are three types of condition monitoring data. Usually, the value-type data could be directly used to represent system condition. For example, when the crack size on gear tooth is greater than a certain value, the gearbox could be treated as having a failure [18]. However, value-type data as such are difficult to obtain with non-intrusive means in practice which makes online reliability estimation impossible. Some other value-type data such as environmental temperatures may not be sensitive to the change of system health condition.

Waveform data and multidimensional data are usually not directly used for reliability estimation. Data processing needs to be implemented to extract quantitative measures, the so-called health indicator or indicator, as the representation of system health conditions. For example, the root mean square (RMS), standard deviation and Kurtosis are used as the indicator of the health of a gearbox [19]. A large number of indicators have been studied in the reported literature [20][21][22][23][24]. Existing methods for reliability estimation usually comprise two key parts: thresholding and probability density estimation which are introduced in the following.

Thresholding refers to the process of determining a critical level of the indicator corresponding to the transition of system states. The critical level of the indicator is also called threshold which is usually a pre-specified constant value [8][11][12][13]. The performance of the system is regarded as deviating from the expected normal state if the indicator value exceeds the threshold.

Thresholds are traditionally determined with prior knowledge [25][26], for example past experiences, design criteria, etc. The event data have also been used to determine the thresholds. For example, in the reported study [27], time-to-failure data are used to estimate the average availability of drill bits which is treated as the threshold for making maintenance decision for drill bits. In the case where prior knowledge or time-to-failure data are not available, thresholds can be determined directly from condition monitoring data. In the reported study [28], a boundary between normal and abnormal data was determined based on condition monitoring data and was treated as a threshold for reliability estimation. Thresholding methods will be reviewed in Section 2.1.

Probability density estimation refers to the process of estimating the PDF of indicator values over a certain time period. In the reported studies [29][30][31][32], it is assumed that the indicator values over a certain time period follow an identical statistical distribution described by a PDF. As reported in [33], probability density estimation methods are able to obtain the PDF of a given set of data. Therefore, it is applicable to use probability density estimation methods to obtain the PDF of indicator value for reliability estimation. Reported methods for probability density estimation fall into two categories: parametric methods and nonparametric methods [34]. Parametric methods estimate the PDF of data with the assumptions that the type of PDF is known and with unknown finite parameters. Nonparametric methods estimate the PDF of data based on the data themselves and regardless of the prior information of the type of PDF [35]. Probability density estimation methods will be reviewed in Section 2.2.

The studies in [36][37][38][39][40] reported the applications of reliability estimation by means of thresholding and probability density estimation. With the acquired threshold and PDF of indicator values, the reliability is estimated as the probability that the indicator value does not

exceed the threshold value. Compared to reliability estimation based on event data, the reliability estimation based on condition monitoring data can provide reliability estimates in accordance with the change of indicator value and reflect the degradation process of individual systems. Also, it can estimate the reliability for expensive or highly reliable systems for which the event data such as time-to-failure data are not always available [3][4].

However, reliability estimation based on condition monitoring data requires condition monitoring devices and the threshold of indicator value is difficult to determine when there are not sufficient data [3][4].

1.4. Research Objectives

As described above, condition monitoring data based reliability estimation has some significant advantages over the event data based one, but determining the threshold of indicator value is a challenging task. As a matter of fact, in many reported studies on condition monitoring data based reliability estimation [36][37][38][39][40][41][42], the threshold is determined using event data and condition monitoring data are used only to obtain the probability density function (PDF) of indicator value. Therefore, for the systems for which it is too costly or impractical to obtain event data, particularly the time-to-failure data, there is a demand for reliability estimation methods which use only condition monitoring data to perform both thresholding and probability density estimation. Unfortunately, very few studies are reported in the literature. One reported study [28] which estimated reliability based on condition monitoring data falls into this type and implies a possible direction for addressing this kind of problem.

The reported study [28] proposed a method jointly using one-class support vector machine (OC-SVM) solution path algorithm and kernel density estimation (KDE) to estimate the

reliability of a water pump. The OC-SVM is used to determine threshold of pump health indicator and the KDE is used for probability density estimation of pump health indicator. The details of the method will be given in Section 4.2. This thesis investigates in-depth this method and finds four aspects of the method that may be deficient and need further investigations:

- (1) The threshold determined by OC-SVM is sensitive to the width parameter of OC-SVM [43][44][45][46][47], but its selection was not given in the reported method.
- (2) Probability density function determined by KDE significantly relies on the selection of bandwidth parameter of KDE [48][49]. The reported method provided a formula to calculate the parameter value but without any explanations.
- (3) In the reported method [28], condition monitoring data were subjectively split into multiple time windows in each of which the data were assumed stationary and used for KDE. This may not be appropriate as time window may contain non-stationary data which will adversely affect the results of probability density estimation.
- (4) It is widely agreed that outliers have negative effects on the analysis of condition monitoring data and the assumption that data do not have outliers may provide the result of reliability estimation non conforming to the truth [50][51]. In the reported method, outliers were removed from the data for training OC-SVM but were kept for KDE. The effects of outlier need to be further investigated.

Accordingly, this thesis investigates the above four aspects and provides insights that could be utilized to enhance the robustness of the reported method for potential practical applications. The investigations and contributions are summarized as follows:

- (1) Investigate the impact of the width parameter of OC-SVM on the reliability estimates and provide suggestions on its selection.
- (2) Investigate the impact of the bandwidth parameter of KDE on the reliability estimates and provide suggestions on its selection.
- (3) Investigate the impact of the time window size for KDE on the reliability estimates and test the applicability of using variable time window size for KDE.
- (4) Investigate the impact of outliers in the data for OC-SVM and KDE on the reliability estimates.

1.5. Thesis Organization

The thesis is organized as follows. Chapter 2 presents the literature review of reliability estimation based on condition monitoring data. Chapter 3 presents the fundamentals of OC-SVM and KDE. Chapter 4 presents the investigations results and also introduces simulation data and experiment data used for investigations. Chapter 5 summarizes the results and lists future topics.

Chapter 2

Literature Review

As mentioned in Section 1.4, reliability estimation based on condition monitoring data is the focus of this thesis. When doing reliability estimation, there are two key parts that need to be considered which are thresholding and probability density estimation. This chapter first reviews the reported methods for thresholding and probability density estimation, respectively. The methods reviewed are not limited to the applications in the realm of reliability estimation. The recently reported reliability estimation methods based on thresholding and probability density estimation are reviewed at last.

2.1. Thresholding

In condition monitoring, health indicators are used to track the conditions of a system. As a competent indicator, its values should vary along with the change of system health conditions. When the indicator value reaches a certain level, the performance of the system could be regarded as deviating from the normal condition or having a fault or a failure. The certain level of indicator value is so-called the threshold of the indicator or the failure threshold [52]. Thresholding refers to the process of determining the threshold. Existing methods for thresholding fall into three categories [25][26]: (1) design standard based thresholding, (2) experience based thresholding, and (3) model based thresholding. The selection of an appropriate threshold value is a challenging task. An over-estimated threshold may cause excessive use of

the system and lead to unexpected system downtime; while an under-estimated threshold may increase repair or replacement rates and results in high maintenance costs [52].

Design standard based methods are usually provided by manufacturers [1][36][53]. For example, International Organization for Standardization (ISO) 3685 recommends the permissible wear level for cutting tools [54]. When the wear level reaches the specified permissible level, the cutting tool is considered worn out. The level of permissible tool wear is specified ranging between 0.15 mm and 1.00 mm. The choice needs to be made depending upon the operating conditions and materials of the tools [55].

Experience based methods use the experiences collected from the past use of similar systems to determine a threshold. A large number of experience based methods are reported in literature. Wang et al. [56] determined that a display panel had failure when the light intensity of the illumination falls to 50% of the original level based on past experiences. Lu et al. [57] used 1.6 inches as the failure threshold of metal fatigue crack length based on their experiences and this threshold value was also used by other researchers [25][52]. Experience based methods usually require significant amount of historical data about the operation of the system to determine a threshold and sometimes it may be subjective [56].

Model based methods treat the failure threshold as a random variable which follows a statistical distribution or can be obtained from available formulas. Nystad et al. [58] assumed the failure threshold of erosion level of choke valves to be a random variable following the Gamma distribution. The failure threshold is regarded as locating within a certain range that was estimated with the Gamma distribution. When the indicator value enters the range, the valve is considered having high probability of failure. When the indicator value is beyond the upper

bound, the valve is considered failure. Yu et al. [52] and Wang et al. [56] assumed the failure threshold of fatigue crack length to be a random variable following the normal distribution. The mean and the variance of the normal distribution were estimated based on historical data. Xiang et al. [59] assumed that the failure threshold of a repairable system could follow four different statistical distributions and found that the Weibull distribution outperformed its counterparts in terms of determining thresholds for the system.

Model based methods may also use available formulas. Wang et al. [60] proposed a stepwise function to estimate the threshold of an indicator for a gearbox system. The stepwise function includes formulas for calculating the thresholds for the scenarios of early fault, transition from fault to failure and final failure. Son et al. [10] detected the failure of automotive battery based on the threshold determined by the following formula:

$$D = \psi + \mu_{\theta} T_{0.5}, \quad (2.1)$$

where D represents the failure threshold, ψ represents the fixed effect parameter, μ_{θ} represents the mean of samples, and $T_{0.5}$ represents the median life of a population estimated by failure data. Guo et al. [61] used a pre-specified mathematical formula to estimate the threshold of the operating temperature signaling the failure of the gearbox in wind turbines.

To apply model based methods, one needs to know in advance which statistical distribution the random variable of threshold follows or the formula for estimating failure thresholds. The main flaw is that these methods are not applicable when the required data are not available. Recently, Hua et al. [28] proposed a model based method which is able to provide failure thresholds without knowing the prior information of the model. The reported method adopts one-class support vector machine (OC-SVM) as a universal model to determine the threshold. The

threshold is represented by a boundary which is internally created by OC-SVM for distinguishing normal data from abnormal data. In this way, determining whether an indicator value exceeds the threshold turns into determining whether the indicator value belongs to the abnormal data class. Hua et al. [28] used OC-SVM as a part of the reliability estimation method to identify abnormalities from a high pressure water descaling pump based on condition monitoring data. The applications of OC-SVM in thresholding are also reported by other researchers. Fernández-Francos et al. [62] used OC-SVM to determine the fault threshold for bearings of high speed rotating machinery based on vibration signals. Hu et al. [63] used OC-SVM to estimate failure threshold for turbo pumps of rocket engine based on vibration signals. The drawback of OC-SVM based thresholding is that the threshold is not in an explicit form as statistical distributions or formula based methods, because OC-SVM model is a black box [64]. Also, the performance of OC-SVM may be compromised if its parameters are not properly selected [44].

Figure 2.1 illustrates the categorization of thresholding methods and the reviewed studies of each category. In summary, the standard and experience based methods are favorable in thresholding for a group of many systems. Model based methods are favorable in thresholding for individual systems. For the systems which have a large amount of failure data available, statistical distribution or formula based thresholding is applicable; for expensive or highly reliable systems which have limited failure data, the universal model based thresholding is a good choice.

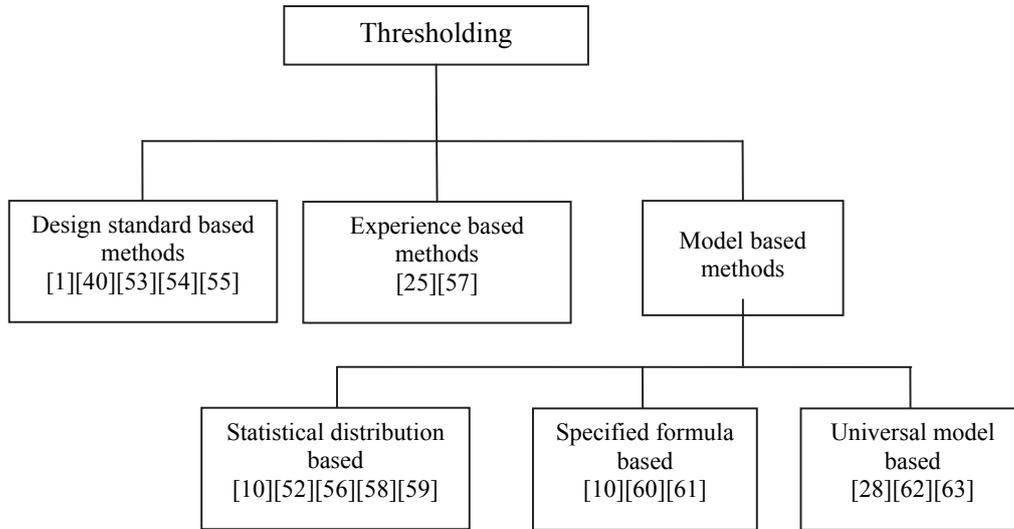


Figure 2.1 Categorization of thresholding methods

2.2. Probability Density Estimation

Existing methods for probability density estimation can be divided into two categories: parametric methods and nonparametric methods [34]. Parametric methods assume that data follow a statistical distribution which can be represented by a probability density function (PDF) but with unknown finite parameters. Once the parameters are estimated, the PDF is fully determined. When the parameters are not available, parameter estimation methods need to be used. Maximum likelihood estimation (MLE) [33] and least squares estimation (LSE) [29] are commonly used methods to estimate the parameters of PDF. The process of estimating the parameters of PDF using MLE is described as follows:

Given a set of measured data or samples x_1, x_2, \dots, x_n which are assumed to follow a certain type of statistical distribution with a PDF of $f(x)$. According to MLE, we have the following function:

$$L(x_i; \theta) = \prod_{i=1}^n f(x_i; \theta), \quad (2.2)$$

where $L(x_i; \theta)$ is the likelihood function and $f(x_i; \theta)$ is the PDF of the data with unknown parameter θ . Taking the partial derivative of the logarithm of the likelihood function with respect to parameter θ and setting the derivative to 0, the following equation is obtained:

$$\frac{\partial \log L(x_i; \theta)}{\partial \theta} = \frac{\partial \log \prod_{i=1}^n f(x_i; \theta)}{\partial \theta} \equiv 0. \quad (2.3)$$

By solving Eq. (2.3), the estimate of parameter θ can be obtained by:

$$\hat{\theta}(x) = \arg \max_{\theta} L(x_i; \theta). \quad (2.4)$$

The PDF of the samples can thus be determined and available for probability estimation.

Commonly used statistical distributions for probability density estimation include Gaussian distribution, Weibull distribution, exponential distribution and Gamma distribution. Raykar et al. [66] estimated the PDF of two dimensional data using multivariate Gaussian distribution of which parameters were estimated using the MLE method. Varanasi et al. [67] obtained the PDF of environmental noise data by using Gaussian distribution. The parameters of the PDF were estimated using the MLE method. Arenberg et al. [68] estimated the probability of the occurrence of defects on fused silica using Weibull distribution and the MLE method was used for estimating the parameters of the Weibull distribution. Parametric methods rely on good understanding of data to select proper type of statistical distribution and require a large amount of data to acquire appropriate parameters for the statistical distributions.

Nonparametric methods estimate the PDF of data based on the data themselves [35] and do not require assumptions on which statistical distributions the data should follow. Popular nonparametric methods are histogram, k-nearest neighbors (KNNs) and kernel density estimation (KDE). Reviews of nonparametric methods can be found in [33][34][69][70]. Histogram which can provide a quick visualization of probability density is the first widely used probability

density estimation method and is regarded as the simplest among its counterparts [33][65]. The histogram is determined by two parameters, the bin width and the starting position of the first bin. Given a set of data or samples x_1, x_2, \dots, x_n which are first divided into a certain number of bins. The PDF of the samples is represented by the fraction of samples in each bin and is described as follows [33]:

$$p(x) = \frac{1}{n} \frac{[\text{The number of samples in the same bin as } x]}{[\text{bin width}]}, \quad (2.5)$$

where $p(x)$ is the probability density of the sample x and n is the number of samples. Different bin widths and starting positions provide different shapes of probability density.

Several reported studies using histogram for probability density estimation are listed here. Wang et al. [71] used histogram to estimate the PDF of load samples for transmission gears of heavy duty excavators. Mazzuchi et al. [72] estimated the lifetime distribution of mechanical equipment using histogram. Buchta et al. [73] obtained the failure frequency of large power units based on the PDF of operation and repair times estimated by histogram.

The shape of PDF obtained by histogram is not smooth, so it is not easy to draw a contour to represent the probability density of data. Histogram requires a very large amount of data to solve high dimensional problems, so it is usually used for one or two dimensional data.

KNN can be used to estimate the PDF of data in high dimensional spaces [74]. For a set of samples x_1, x_2, \dots, x_n , the k th nearest neighbor density estimation in d dimensions is defined as follows [33]:

$$f(x) = \frac{k/n}{V_k(x)} = \frac{k/n}{c_d r_k(x)^d}, \quad (2.6)$$

where $f(x)$ represents the probability density at the sample x , n represents the sample size, $V_k(x)$ represents the volume of the d dimensional sphere with radius $r_k(x)$ and c_d represents the volume of unit sphere.

Raykar [66] used KNN to estimate the PDF of an unknown probability distribution based on the samples of two dimensional feature vectors. Zamini et al. [75] used KNN to estimate the PDF of the lifetime of brake pads. Kung et al. [76] reported the use of KNN to estimate the PDF of high dimensional samples. KNN is suitable for high dimensional samples and can provide simple and flexible estimation of probability density [76]. However, the estimated probability density tends to be affected by local noise [33]. Also, the PDF obtained by KNN might have a heavy tail. These drawbacks limit the application of KNN in estimating the entire probability density [33].

Kernel density estimation (KDE) has been developed to overcome the drawbacks of KNN in recent decades. Given a set of samples x_1, x_2, \dots, x_n , the PDF estimated by KDE can be described as follows [33]:

$$f(x) = \frac{1}{nh^D} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right), \quad x \in R^D, \quad (2.7)$$

where n is the number of samples x_i , h is the bandwidth, D is the number of dimensions, and $K(\bullet)$ is the kernel function. KDE is an essential part of this thesis. Its fundamentals will be further introduced in Section 3.2.

Reported studies of KDE for probability estimation are as follows. Hua et al. [28] used KDE to estimate the failure probability for a high-pressure descaling pump based on vibration signals. Ebrahim et al. [29] used KDE to obtain the PDF of lifetime for laser devices based on degradation data of operating current. Ferracuti et al. [30] used KDE to estimate the PDF of fault for induction motors based on current signals. Cho et al. [31] developed a method using KDE to

perform off-line and on-line probability density estimation based on vibration signals. KDE is capable of dealing with high dimensional samples and providing smooth PDFs [33]. The drawback is that the selection of bandwidth for kernel function may affect the performance of KDE. In the event that an inappropriate bandwidth is selected, an over or under smoothed PDF may be obtained which cause unreliable probability estimations.

Figure 2.2 illustrates the categorization of probability density estimation based on the reviewed literature. In terms of pros and cons of each category, Ebrahim et al. [29] provided good insights through the comparisons between parametric and non-parametric methods. In the tests, the exponential and Weibull distributions of parametric method and KDE of non-parametric method were chosen for comparisons. Simulation data generated from linear degradation model and experiment degradation data of laser operating current were used. The results implied that parametric methods tend to outperform KDE when statistical distributions are known; while KDE tends to perform better when the statistical distributions are not selected properly.

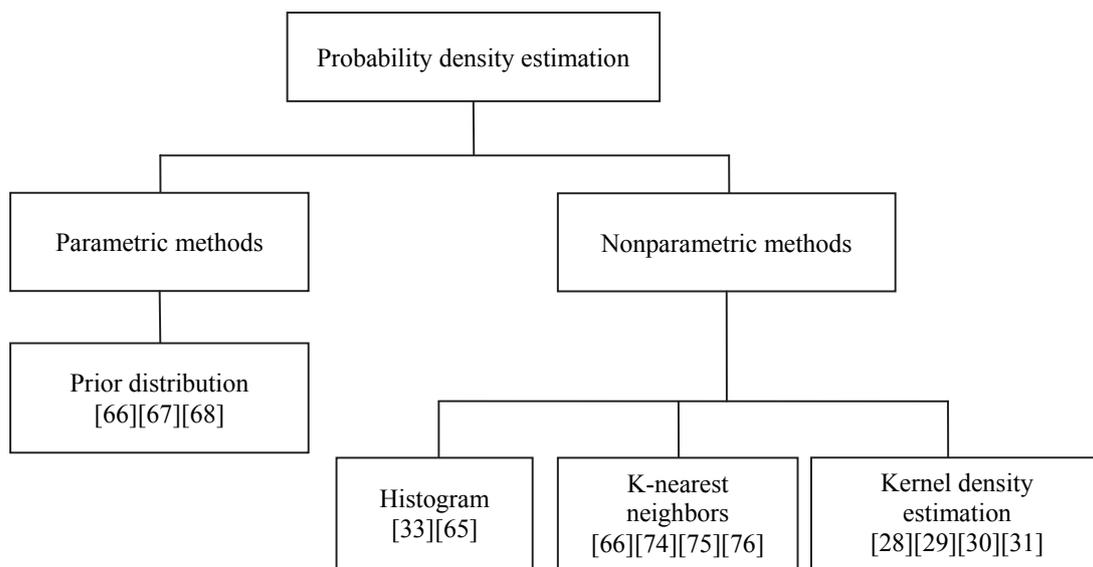


Figure 2.2 Categorization of probability density estimation methods

2.3. Reliability Estimation Using Condition Monitoring Data

As mentioned at the beginning of this chapter, reliability estimation methods consist of two parts, thresholding and probability density estimation. Sections 2.1 and 2.2 reviewed existing methods for thresholding and probability density estimation, respectively. These methods can be utilized for reliability estimation by using either event data or condition monitoring data. Since the focal point of this thesis is reliability estimation based on condition monitoring data, this section reviews specifically reported studies in this respect. The reported studies on reliability estimation based on condition monitoring data are limited. In the following we summarize the studies reported in recent years.

Feng et al. [41] estimated the reliability of the cylinder of diesel engine based on compression pressures in the cylinder. Design standard was adopted to specify the failure threshold of compression pressure. A parametric method was used to estimate the probability density where the compression pressure data are assumed to follow the normal distribution.

Hua et al. [42] proposed a method of reliability assessment based on dynamic probability model for a centrifugal pump. The failure threshold was a critical value of pressure based on design standard provided by manufacturer. Nonparametric method of KDE was used to obtain the probability density function for the monitored pressure measurements.

Ding et al. [36] estimated the reliability of the cutting tool by jointly using tool wear data and condition indicators extracted from vibration signals. A parametric method was used where tool wear data were assumed to follow the lognormal distribution. A threshold of wear level was chosen based on the standard on cutting tools. Two indicators, root mean square and peak in time domain, which showed the pattern of the wear of cutting tool were also used. A model which

relates the wear level to the two indicators was developed to estimate the reliability based on condition monitoring data.

Ebrahimi [5] proposed two models to indirectly estimate the reliability of ship hulls and airplane structures. A predetermined crack length was set as the failure threshold. Parametric methods of Gaussian distribution and Brownian motion were used to fit the fatigue crack length data.

Gao et al. [37] estimated reliability for aircraft engine based on mixed condition monitoring data. Multiple failure modes were considered. Failure threshold for each failure mode was determined based on prior experiences. Two reliability analysis models were used. One was Gamma process model for modeling degradation and the other was Wiener process model for modeling random disturbance effects. Reliability of aircraft engine was estimated based on a combination of these two models.

Hua et al. [28] proposed a method that used one-class support vector machine (OC-SVM) to determine failure threshold and kernel density estimation (KDE) to estimate failure probability density based on indicators extracted from vibration signals of a descaling pump. As mentioned in Section 2.1, the threshold was the boundary between normal and abnormal data of indicator. KDE estimated the probability density based on the indicator values windowed over different time periods. The reliability is estimated for each window and a trend is obtained to exhibit the change of reliability over different time periods. This method offers a promising direction for reliability estimation based on only condition monitoring data.

It is found that the reported methods mainly determine thresholds based on recommended standards or prior experiences; while few model based methods are reported. The reported

probability density estimation methods are versatile for reliability estimation. Parametric methods are dominant when the prior knowledge of data is available such as the distribution of condition monitoring data. Otherwise, nonparametric methods are utilized.

2.4. Summary

This chapter reviews condition monitoring data based reliability estimation methods in terms of thresholding and probability density estimation. Sections 2.1 and 2.2 categorized and reviewed existing methods for thresholding and probability density estimation, respectively. Section 2.3 reviewed existing methods specifically for reliability estimation based on condition monitoring data. Pros and cons of different categories of methods are commented and the suggestions are made for one to select appropriate options in possible scenarios.

It is noticed from the reviewed studies that the universal model based thresholding is rarely reported for reliability estimation; however, this is in demand for expensive or highly reliable systems for which the standards, experiences, known distributions and formulas for thresholding are unavailable. Hua's method [28] offered a promising direction, but there are some unclearness and deficiencies in this reported method in terms of parameter selection and data processing. Therefore, this thesis puts effort on addressing these concerns and the findings will be presented in Chapter 4.

Chapter 3

Fundamentals of One-Class Support Vector Machine and Kernel

Density Estimation

This chapter presents the fundamentals of one-class support vector machine (OC-SVM) and kernel density estimation (KDE) which are the two key parts of the reliability estimation utilizing condition monitoring data.

3.1. Fundamentals of One-Class Support Vector Machine (OC-SVM)

The OC-SVM was developed by Schölkopf [43] which aims to address the classification problem that only one class of data is available for classification. The one class of data available will be referred to as original data or training data as appropriate and the data to be determined whether belonging to the class of original data will be referred to as test data hereafter. The OC-SVM is first trained by the original data and a boundary is created that enables all original data to be on only one side of the boundary. Next, suppose the boundary is a straight line in the original space, the test data locating at the same side of the original data are considered belonging to the same class. In practice, the boundary is usually non-linear. OC-SVM introduces a mapping function ϕ which projects the original data to a high dimensional feature space in which a boundary is created in the form of hyperplane to ensure that all original data are on the same side of the boundary. Figure 3.1 illustrates how OC-SVM dissects the feature space. The solid straight line is the hyperplane (boundary) determined by OC-SVM. The distance from the

hyperplane to the origin which is maximized by OC-SVM is represented by $\frac{\rho}{\|\mathbf{w}\|}$. The distance between a data point lying on the false class side and the hyperplane is denoted by ξ . The dots on the hyperplane are the support vectors.

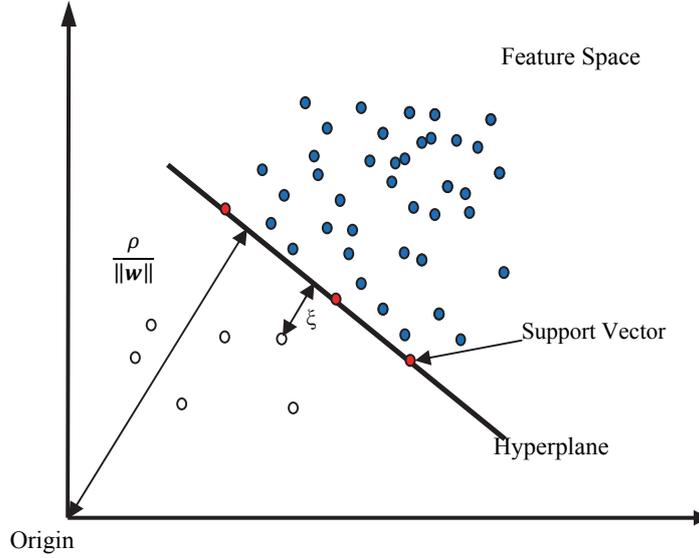


Figure 3.1 Separation of data using OC-SVM

Consider a dataset $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T \in \mathbb{R}^{N \times M}$ that contains N training data points. OC-SVM aims to learn a function $f(\mathbf{x})$ that returns $+1$ in a small region consisting of all the training data points and returns -1 for the other data points. Mathematically, the target is to learn the weight vector $\mathbf{w} = (w_1, \dots, w_M)^T$ and an offset ρ for the function of $f(\mathbf{x})$ with the following form:

$$f(\mathbf{x}) = \begin{cases} 1, & (\mathbf{w}^T \cdot \boldsymbol{\phi}(\mathbf{x})) - \rho \geq 0, \\ -1, & \text{otherwise,} \end{cases} \quad (3.1)$$

where $\boldsymbol{\phi}$ describes the non-linear mapping from the original space to the feature space. In practice, $\boldsymbol{\phi}$ cannot be directly calculated, so kernel function is used to represent the inner product of the mapping of two data points, \mathbf{x} and \mathbf{y} , as:

$$K(\mathbf{x}, \mathbf{y}) = (\boldsymbol{\phi}(\mathbf{x})^T \cdot \boldsymbol{\phi}(\mathbf{y})), \quad (3.2)$$

where $K(\bullet)$ represents the kernel function. For example, the Gaussian kernel has the following function form:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right), \quad (3.3)$$

where σ represents the kernel parameter and needs to be properly selected.

To find the hyperplane that separates the training data from the origin with the maximum margin, the following quadratic optimization problem is modeled:

$$\min_{\mathbf{w}, \xi, \rho} \quad \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{vN} \sum_{i=1}^N \xi_i - \rho, \quad (3.4)$$

$$\text{s. t.} \quad (\mathbf{w}^T \cdot \boldsymbol{\phi}(\mathbf{x}_i)) \geq \rho - \xi_i, \quad (3.5)$$

$$\xi_i \geq 0, \quad \forall i = 1, \dots, N, \quad (3.6)$$

where $v \in (0,1]$ is a given constant representing the upper bound on the fraction of outliers and the lower bound on the fraction of support vectors and ξ_i is a slack variable representing the distance between a data point, \mathbf{x} , lying on the false class side and the plane in its virtual class side, ρ is an offset. The \mathbf{w} , ξ_i and ρ are decision variables of the optimization model.

The Lagrangian function of the above optimization problem can be expressed as:

$$L(\mathbf{w}, \boldsymbol{\xi}, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{vN} \sum_{i=1}^N \xi_i - \rho - \sum_{i=1}^N \alpha_i (\mathbf{w}^T \cdot \boldsymbol{\phi}(\mathbf{x}_i) - \rho + \xi_i) - \sum_{i=1}^N \xi_i \beta_i, \quad (3.7)$$

where $\alpha_i, \beta_i \geq 0$ are the Lagrange multipliers. Setting the partial derivatives of Eq. (3.7) with respect to \mathbf{w} , $\boldsymbol{\xi}$ and ρ to 0, \mathbf{w} and $\boldsymbol{\alpha}$ can be expressed as:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i \boldsymbol{\phi}(\mathbf{x}_i), \quad (3.8)$$

$$\alpha_i = \frac{1}{\nu N} - \beta_i \leq \frac{1}{\nu N}, \sum_{i=1}^N \alpha_i = 1. \quad (3.9)$$

Substituting Eqs. (3.8) and (3.9) into Eq. (3.7), the terms with $\boldsymbol{\xi}$ and ρ are canceled. The Eq. (3.7) can be transformed to the following quadratic optimization problem:

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j), \quad (3.10)$$

$$\text{s. t. } 0 \leq \alpha_i \leq \frac{1}{\nu N}, \sum_{i=1}^N \alpha_i = 1. \quad (3.11)$$

where α_i which is the Lagrange multiplier is the new decision variable, ν is a given constant representing the upper bound on the fraction of outliers and the lower bound on the fraction of support vectors, N is number of data and K is kernel function.

According to Karush-Kuhn-Tucker conditions, the data points can be classified into three categories: (1) the data points with $\alpha_i = 0$ locate within the boundary, (2) the data points with $0 < \alpha_i < \frac{1}{\nu N}$ are on the boundary and the corresponding ξ_i is equal to 0, and (3) the data points which satisfy $\alpha_i = \frac{1}{\nu N}$ fall outside the boundary. The data points with $\alpha_i > 0$ are the so-called support vectors.

Solving Eqs. (3.10) and (3.11), we can express ρ as:

$$\rho = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \alpha_i, \quad (3.12)$$

where n is the number of support vectors which satisfy $\xi_i = 0$ and $0 < \alpha_i < \frac{1}{\nu N}$.

3.1.1.OC-SVM Solution Path

The OC-SVM model has several parameters which include ν , N and ρ that need to be specified properly to ensure a successful classification. Lee et al. [77] developed an algorithm to estimate the entire path of OC-SVM solutions to simplify the selection of parameters. This algorithm is facilitated by an alternative formula of OC-SVM in which parameter $\lambda = \nu N \rho$ is used to replace parameters ν , N and ρ . Using the OC-SVM Solution Path, the quadratic optimization problem of Eqs. (3.4), (3.5) and (3.6) is transformed as [77]:

$$\min_{\mathbf{w}, \xi} \quad \frac{\lambda}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^N \xi_i, \quad (3.13)$$

$$\text{s.t.} \quad \left(\mathbf{w}^T \cdot \boldsymbol{\phi}(\mathbf{x}_i) \right) \geq 1 - \xi_i, \quad (3.14)$$

$$\xi_i \geq 0, \quad \forall i = 1, \dots, N. \quad (3.15)$$

By using the Lagrange technique, the solution of \mathbf{w} can be expressed as [77]:

$$\mathbf{w} = \frac{1}{\lambda} \sum_{i=1}^N \alpha_i \boldsymbol{\phi}(\mathbf{x}_i). \quad (3.16)$$

The Karush-Khun-Tucker conditions for the constrained optimization problem also require that:

$$\alpha_i (f(\mathbf{x}_i) - 1 + \xi_i) = 0, \quad (3.17)$$

$$\beta_i \xi_i = 0, \quad (3.18)$$

$$\xi_i \geq 0, \quad (3.19)$$

where α_i and β_i are the Lagrange multipliers and $f(\mathbf{x}_i) = \left(\mathbf{w}^T \cdot \boldsymbol{\phi}(\mathbf{x}_i) \right)$. The $\boldsymbol{\phi}(\mathbf{x}_i)$ that enables $f(\mathbf{x}_i) = 1$ defines the hyperplane with a distance $\frac{1}{\|\mathbf{w}\|}$ from the origin.

At the beginning of this algorithm, a sufficiently large value of λ is set to ensure that all the data points fall inside the margin. The margin is defined as follows:

$$f(\mathbf{x}_i) = \frac{1}{\lambda} (\mathbf{w}^* \cdot \boldsymbol{\phi}(\mathbf{x}_i)) \leq 1, \quad \forall i = 1, \dots, N, \quad (3.20)$$

where $\mathbf{w}^* = \sum_{i=1}^N \boldsymbol{\phi}(\mathbf{x}_i)$. Finding the most extreme point \mathbf{x}_i from the origin, the initial value of λ can be obtained as:

$$\lambda_0 = \max_i \sum_{j=1}^N \boldsymbol{\phi}(\mathbf{x}_i) \boldsymbol{\phi}(\mathbf{x}_j). \quad (3.21)$$

To obtain the entire solution path, λ decreases from λ_0 to 0. Accordingly, $\|\mathbf{w}\|$ increases and the distance of the margin decreases. As the distance of margin decreases, the data points cross the margin and move from inside to outside of the margin; while the corresponding α_i changes from 1 to 0. During this process, the algorithm monitors the following three subsets:

$$R = \{i : f(\mathbf{x}_i) > 1, \alpha_i = 0\}, \quad (3.22)$$

$$E = \{i : f(\mathbf{x}_i) = 1, 0 < \alpha_i < 1\}, \quad (3.23)$$

$$L = \{i : f(\mathbf{x}_i) < 1, \alpha_i = 1\}. \quad (3.24)$$

When the λ changes to a value that all data points are entering region E in Eq. (3.23) from region R in Eq. (3.22) or leaving region E from region L in Eq. (3.24), the optimal λ can be determined. The detailed derivation of the algorithm can be found in [77].

3.1.2. Reported Methods for σ Selection

Kernel function is critical to the performance of OC-SVM. Many kernel functions are available in the literature in which the Gaussian kernel function is the most widely used one [44]. As presented in Eq. (3.3), the Gaussian kernel has a width parameter σ that needs to be specified. The impact of σ on reliability estimation will be investigated in Section 4.3. This section introduces three reported methods for estimating σ . The three methods will be used in verifying

the findings acquired from the investigations. The mathematical definition for each method is given, but its derivation is not provided since it is not the focus of this thesis.

Method 1: Automated Method Based on Variance and Mean (VM)

This method obtains an optimal value of σ by using the variance and mean of kernel matrix [46]:

$$\max_{\sigma} \frac{s^2}{\bar{K} + \epsilon} \quad \forall (i \neq j), \quad (3.25)$$

where $K(\mathbf{x}_i, \mathbf{y}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$ represents the kernel function, $\bar{K} = \frac{\sum_{i=1}^n \sum_{j=i+1}^n K(\mathbf{x}_i, \mathbf{y}_j)}{l}$ represents the mean of the kernel matrix, $s^2 = \frac{\sum_{i=1}^n \sum_{j=i+1}^n (K(\mathbf{x}_i, \mathbf{y}_j) - \bar{K})^2}{l-1}$ represents the variance of the kernel matrix where $l = \frac{(n^2 - n)}{2}$, n represents the number of data points, and ϵ represents a small number to make the denominator nonzero.

Method 2: Maximum Distance Method (MD)

This method uses the maximum distance between data points to select σ [47]:

$$\sigma = \frac{d_{\max}}{\sqrt{-\ln(\delta)}}, \quad (3.26)$$

where d_{\max} represents the maximum distance between data points and δ represents the width of boundary in kernel space.

Method 3: Distance from the Farthest and Nearest Neighbors Method (DFN)

This method utilizes the farthest and the nearest distance from data samples to their neighbors to obtain an optimal value of σ [44]. The farthest neighbors of data samples are treated as the boundary to separate the normal data and abnormal data. The nearest neighbors are considered

having the closest structures to data samples. An objective function is formulated to obtain the optimal σ by maximizing the following difference:

$$\max_{\sigma} f(\sigma) = \frac{1}{n} \sum_{i=1}^n \max_j \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2 - \frac{1}{n} \sum_{i=1}^n \min_{j \neq i} \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2, \quad (3.27)$$

where \mathbf{x} represents a data point, n represents the number of data points, and $\|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2$ represents the distance between two mappings in the feature space. The first term of the function is the mean of the farthest distances of data points and the second term is the mean of the nearest distances of data points. The mapping function from the original space to the feature space is not available, so Eq. (3.27) is simplified using the monotonicity of the exponential function. The optimal σ can thus be obtained by solving the following optimization problem:

$$\max_{\sigma} f(\sigma) = \frac{2}{n} \sum_{i=1}^n \exp\left(-\frac{\min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right) - \frac{2}{n} \sum_{i=1}^n \exp\left(-\frac{\max_j \|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right). \quad (3.28)$$

3.2. Fundamentals of Kernel Density Estimation (KDE)

KDE is a non-parametric probability density estimation method that has been studied for more than 50 years. Consider a random variable X which follows a particular statistical distribution and a dataset, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T \in \mathbb{R}^{N \times M}$, is observed from the same distribution. KDE is able to estimate the probability density function (PDF) of the random variable X using the dataset \mathbf{X} by the following expression:

$$f(\mathbf{x}) = \frac{1}{Nh^M} \sum_{i=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right), \quad (3.29)$$

where h represents the bandwidth, M represents the dimensions of a data point, N represents the number of data points, and $K(\bullet)$ represents the kernel function. The kernel function is usually assumed symmetric about 0 and satisfies the following condition [33]:

$$K(x) \geq 0, \int_{-\infty}^{+\infty} K(x) dx = 1. \quad (3.30)$$

Kernel function and the bandwidth h are the most important parameters for KDE. Once the kernel function K and the bandwidth h are specified, the probability density can be estimated for a set of data points \mathbf{x} using Eq. (3.29). Figure 3.2 shows the shapes of some kernel functions and their formulas are listed in Table 3.1.

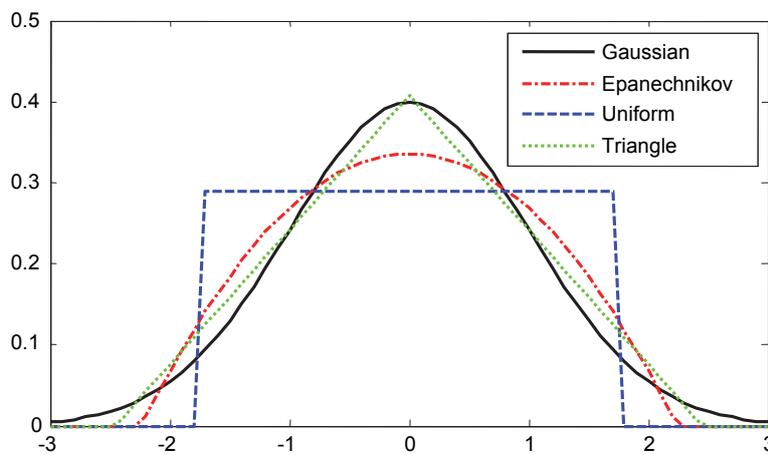


Figure 3.2 Plots of sample kernel functions

Table 3.1 Sample kernel functions

Name	Kernel Function
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
Uniform	$K(x) = \frac{1}{2}, x \leq 1$
Triangular	$K(x) = 1 - x , x \leq 1$
Epanechnikov	$K(x) = \frac{3}{4}(1 - x^2), x \leq 1$

The kernel function determines the shape of the PDF of the data points and the bandwidth determines the width of the shape. The PDF of all data points is the sum of the kernel functions

of each data point based on Eq. (3.29). Figure 3.3 uses the Gaussian kernel to illustrate how the kernel function of each data point constructs the PDF. The red dots represents the data points used for probability density estimation. Blue dashed lines represents the kernel function of each data point. The black solid line represents the PDF. The bandwidth parameter used is $h=4$.

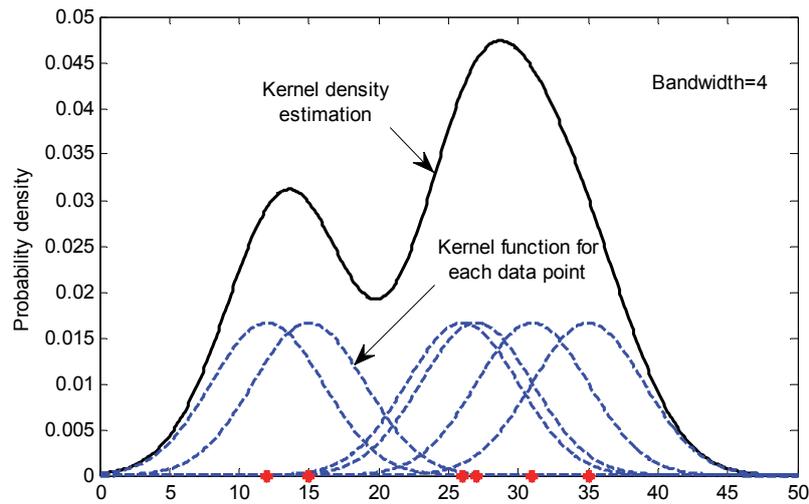


Figure 3.3 Illustration of KDE using Gaussian kernel

The selection of bandwidth h has impact on the PDF. It is reported in [33] that a small bandwidth may lead to under-smoothing of PDF and a large bandwidth may lead to over-smoothing of PDF. Figure 3.4 illustrates the effects of different bandwidth values on the shape of PDF. The Gaussian kernel and the same set of data points producing Figure 3.3 are used here. Two bandwidths of $h = 1$ and $h = 10$ are tested. Figure 3.4 shows that when $h=1$ a spiky shape of PDF is yielded which is difficult to describe and explain; while when $h=10$ a very smooth PDF is yielded but it may be too broad to represent the nature of the data. Therefore, the bandwidth should be carefully selected when using the Gaussian kernel for probability density estimation which is also suggested by [33][78][79].

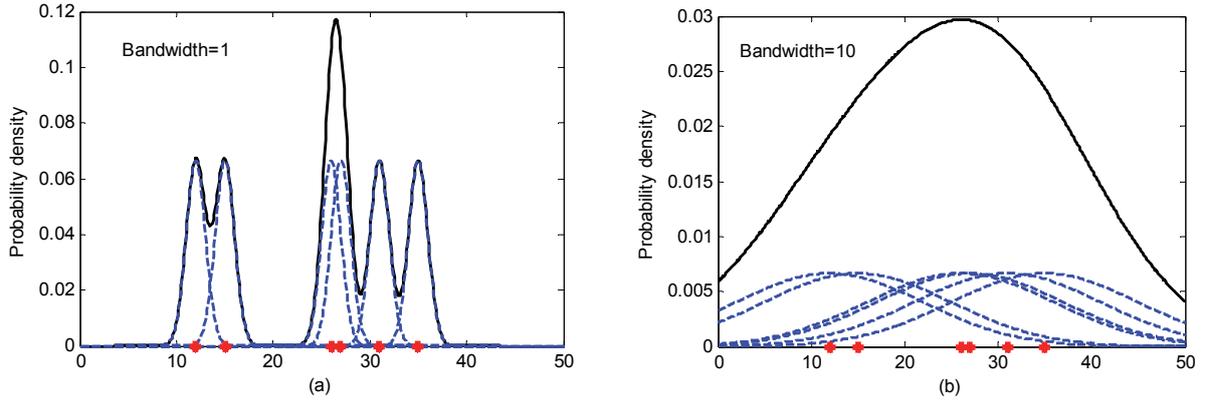


Figure 3.4 KDE using Gaussian kernel with different bandwidth values

The following four reported methods for selecting the bandwidth parameter h are summarized next. They will be used in verifying the findings acquired in Section 4.4. Mathematical formulas are given for each method; however, the detailed derivations are not provided since it is not the focus of this thesis. The Matlab programs for Methods 2, 3 and 4 below were available from the online resource [80].

Method 1: Hua's Method

Hua et al. [28] selects the bandwidth for KDE using the following expression:

$$h = \frac{d_k}{\sqrt{(2n)}}, \quad (3.31)$$

where d_k represents the maximum distance among data points in the k th time window and n represents the number of data points in each time window. When the data points are normalized to $[0, 1]$, h is approximately equal to $(2n)^{-1/2}$.

Method 2: Mean Integrated Squared Error (MISE)

Silverman [33] proposed a method to determine the bandwidth based on MISE. The MISE is defined as:

$$MISE = \int_{-\infty}^{\infty} E \left[\left(\hat{f}(x) - f(x) \right)^2 \right] dx, \quad (3.32)$$

where $\hat{f}(x)$ represents the kernel based PDF and $f(x)$ represents the true PDF. A formula is given to obtain the bandwidth h as follows:

$$h_{MISE} = \left(\frac{R(K)}{\mu_2^2(K)R(f'')} \right)^{1/5} n^{-1/5}, \quad (3.33)$$

where K represents the kernel function, R represents the integral for any squared function of K , f represents the density function, and n represents the number of data points. When the Gaussian kernel is used for the kernel function and the normal distribution is used for the true PDF, minimizing the MISE gives an optimal bandwidth, h^* , as follows:

$$h^* = 1.06\sigma n^{-\frac{1}{5}}, \quad (3.34)$$

where σ represents the sample standard deviation.

Method 3: Direct Plug-in Method (DPI)

Ruppert et al. [81] proposed the DPI method for bandwidth selection based on asymptotic mean squared error (AMSE) which is expressed as follows:

$$h_{DPI} = \left[35 \frac{\hat{\sigma}^2(\hat{\lambda}_{AMSE})(b-a)}{\hat{\theta}_{22}^{0.5}(\hat{g}_{AMSE})n} \right]^{1/5}, \quad (3.35)$$

where $\hat{\sigma}$ represents the prior variance, $\hat{\theta}_{22}^{0.5}$ represents the pre-estimator, \hat{g}_{AMSE} represents the prior bandwidth that minimizes the AMSE of an estimator for θ_{22} , $\hat{\lambda}_{AMSE}$ represents the prior bandwidth that minimizes the conditional AMSE of $\hat{\sigma}^2$, and a and b represent the boundary values which can be estimated based on the observed data points.

Method 4: Least Square Cross-Validation Method (LSCV)

The LSCV method [82] is an expansion of the integrated square error method [33] and estimates the bandwidth by solving the following optimization problem:

$$CV_{LS}(h) = \int \hat{f}_h^2(x) dx - \frac{2}{n} \sum \hat{f}_{h,-i}(x_i), \quad (3.36)$$

$$\hat{f}_{h,-i}(x_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{x-x_j}{h}\right), \quad (3.37)$$

where K represents the kernel function and n represents the number of data points. The optimal bandwidth can be obtained by the following minimization:

$$h^* = \operatorname{argmin}_h CV_{LS}(h). \quad (3.38)$$

3.3. Summary

This chapter presents the fundamentals of OC-SVM and KDE, respectively. The OC-SVM will be used for thresholding in Section 4.3 and KDE will be used for probability density estimation in Section 4.4. The reported methods for selecting width parameter of OC-SVM introduced in Section 3.1.2 will be used to test the observations obtained in Section 4.3. The reported methods for selecting bandwidth parameter of KDE introduced in Section 3.2 will be used to test the observations obtained in Section 4.4.

Chapter 4

Reliability Estimation Using One-Class Support Vector Machine and Kernel Density Estimation

This chapter investigates the four aspects of Hua's method which have vagueness and deficiencies. Simulation data with different noise levels are generated and used for the investigations. Three experiment data sets are also used for testing the validity of the obtained results. This chapter is organized as follows. Section 4.1 presents the simulation data and the experiment data. Section 4.2 introduces Hua's method and brings forward the four aspects to be investigated. Section 4.3 investigates the impact of the width parameter σ of one-class support vector machine (OC-SVM) on reliability estimation. Section 4.4 investigates the impact of the bandwidth parameter h of kernel density estimation (KDE) on reliability estimation. Section 4.5 investigates the impact of the window size on reliability estimation. Section 4.6 investigates the impact of outliers on reliability estimation. Section 4.7 summarizes the results of the investigations.

4.1. Data Preparation

This thesis focuses on a system whose working conditions can be split into two stages, the early stage and the late stage. Condition monitoring data and health indicators developed from condition monitoring data can both be used to represent the two stages as long as they show the desired pattern as described below. For simplicity, condition monitoring data and health

indicators will be referred to as indicators hereafter. At the early stage, the system works under normal condition and the indicator trend exhibits the pattern of a horizontal straight line with allowable fluctuations around a particular value. At the late stage, a fault occurs to the system and the indicator trend exhibits a monotonically ascending or descending pattern with the rate of change increasing along with time. Accordingly, at the early stage, the reliability has the value of 1 and at the late stage the reliability starts to decrease until reaches 0. This chapter uses both simulation data and experiment data which show the trend with the desired pattern.

4.1.1.Simulation Data

Simulation data are generated by the following equation [23]:

$$x(t)=x_{\text{True}}(t)+\varepsilon(t), \quad (4.1)$$

where $x(t)$ represents the indicator value at time t , $x_{\text{True}}(t)$ represents the true value of indicator at time t and $\varepsilon(t)$ represent the additive noise. To obtain a degradation pattern, exponential function is used to generate the true values of the indicator. The additive noise follows Gaussian distribution with the mean of 0 and the standard deviation of 1. The equation is given as:

$$x(t)=10+10^{-3}e^{t}+b\varepsilon(t), \quad b>0, \quad (4.2)$$

where b is the coefficient used to generate the desired noise effects. Figure 4.1 shows the trend of true values where no noise is added. The true values remain the same from the beginning to about time 650 which corresponds to the turning point indicating the transition of system conditions from normal to abnormal.

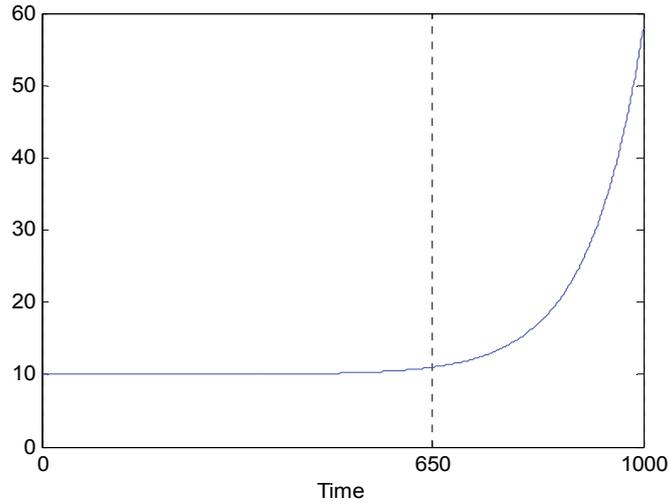


Figure 4.1 The plot of true values of simulation data

In practical scenarios, the degradation data collected may contain noise due to various random factors e.g. environmental conditions, instrument errors, human errors, etc. To generate the degradation data with different noise effects, the reported measure called relative noise level (RNL) [83] is used. The RNL is defined as follows:

$$\text{RNL} = \frac{\text{STD}(\text{Noise})}{\text{STD}(\text{Data})}, \quad (4.3)$$

where STD represents standard deviation. The standard deviation of noise is also referred to as noise level. The RNL values range within (0, 1) where 0 means there is no noise at all and 1 means noise dominates the true values of the data.

For simulation data, the “Data” in Eq. (4.3) is represented by the term, $x(t)$, of Eq. (4.2); the “Noise” in Eq. (4.3) is represented by the term, $b\varepsilon(t)$, of Eq. (4.2). When a set of time series data of $x(t)$ and $\varepsilon(t)$ are generated, the RNL value can be calculated for the time series data using Eq. (4.3). By varying the value of coefficient b , a desired RNL value for a given set of simulation data can be achieved. For experiment data, the “Data” in Eq. (4.3) is a set of time series data of

health indicator values extracted from collected condition monitoring data; the “Noise” in Eq. (4.3) is not available, so the reported k-nearest-neighbors regression method is used to estimate the noise level by the following expression [84]:

$$\varepsilon(t) = \sqrt{\frac{n^{1/5}k}{n^{1/5}k-1} \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2}, \quad (4.4)$$

where \hat{x}_i represents the estimate of the data point x_i by using k-nearest neighbors regression, k represents the number of the nearest data points from x_i , and n represents the sample size of the whole dataset. With the available “Data” and “Noise”, the RNL can be calculated for experiment data using Eq. (4.3).

The intent of introducing the RNL measure is to represent the noise effects in different datasets, so that we can investigate the performance of reliability estimation method is for different noise conditions. In terms of reliability estimation, the reliability estimates mainly remain at the value of 1 at the early stage since the systems are usually working normally during this period. One may be more interested in the period when reliability estimates decrease after a certain time point (turning point) at the late stage. For this reason, the noise effects for the late stage are attached more importance than those for the early stage. To measure the noise effects, we should calculate the RNL value using only the data for the late stage. However, in this thesis we compute the RNL value using the entire set of data instead. The reason is explained as follows.

For a given set of degradation data, the noise determined by the random factors is assumed to follow an identical distribution, so the noise level remains approximately unchanged over time. At the early stage, the true values of degradation data are relatively stable, so the variations of the degradation data are mainly governed by the noise. As a result, the RNL value will be large

at the early stage. At the late stage, the true values of degradation data increase or decrease as time passes, which causes the standard deviation of the degradation data increased. Because the noise level remains stable over the entire time period, the RNL value will decrease at the late stage.

For illustration purposes, in this thesis we use the simulation data and experiment data that have apparent changes (increases or decreases) in the true values of data at the late stage and do not have very long time length at the early stage. This enables the variations of the values of the entire dataset to be mainly caused by the variations of the values of the data at the late stage. Therefore, the RNL value computed using the entire dataset will approximate the one using the data at the late stage.

In practice, the RNL value may need to be computed using only the data at the late stage when the early stage is very long, because the long early stage will cause the standard deviation of the entire degradation data insensitive to the change of the values of the data at the late stage. As a result, the RNL value of the entire dataset will deviate much from the one using only the data at the late stage.

The RNLs used to simulate different environment conditions are 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7. The RNLs of 0.1, 0.2 and 0.3 represent the scenarios where the noise effects are small relative to the simulation data and the RNLs of 0.6 and 0.7 represent the scenarios where the noise effects are significant. The RNLs of 0.4 and 0.5 have noise effects in between. The seven RNLs are selected because the experiment data to be used are within this range. Figure 4.2 shows the plots of simulation data with different RNLs.

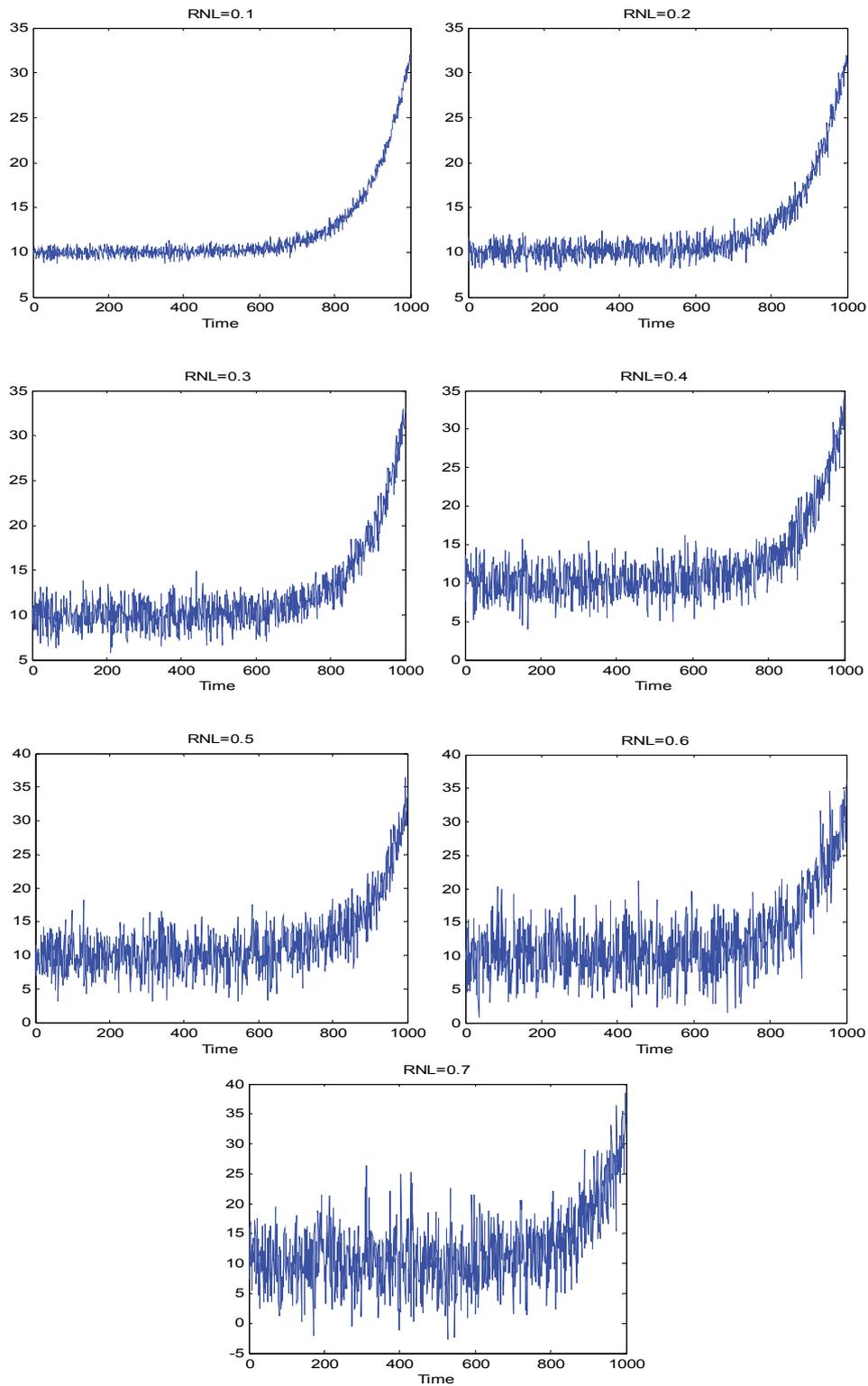


Figure 4.2 The plots of simulation data with different RNLs

To investigate the impact of outliers, outliers with higher values are added to the simulation data. The position at which the outlier is added is chosen from a uniform distribution within a specified time span. Figure 4.3 shows the simulation data with 1 and 5 outliers.

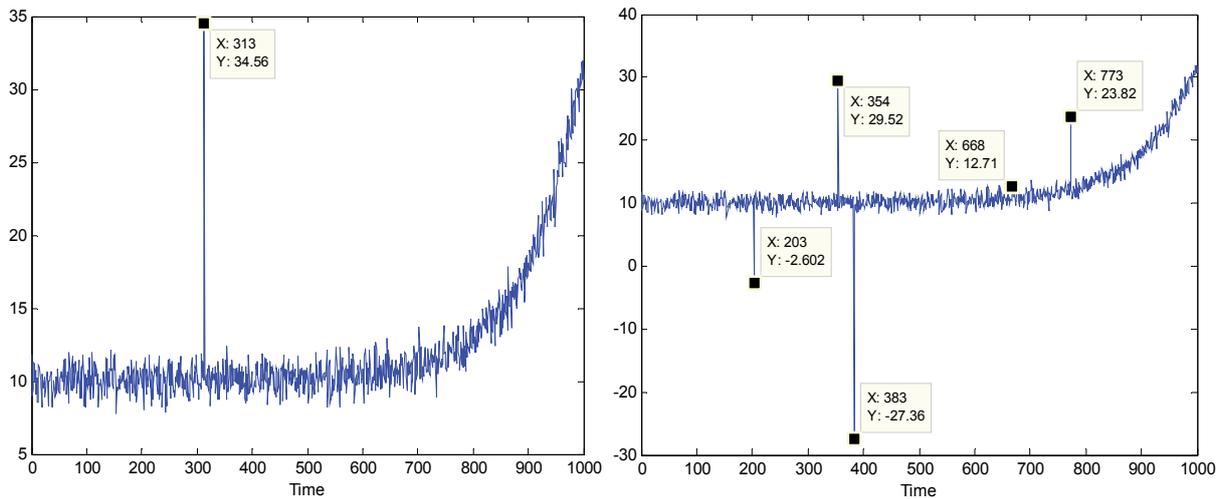


Figure 4.3 The plots of simulation data with 1 and 5 outliers

4.1.2. Experiment Data

(1) Water Pump Data

The experiment data of a water pump are used. The data were read from the figure provided in [28]. The pump worked for 2184 hours and then failed. Between time 900 and 1000 hours, the sealing rings of the water descaling pump were found worn out. Vibration signals were collected. The indicator of root mean square of the vibration data was computed. Figure 4.4 shows the indicator values of the first 1230 hours. The RNL is computed as 0.17.

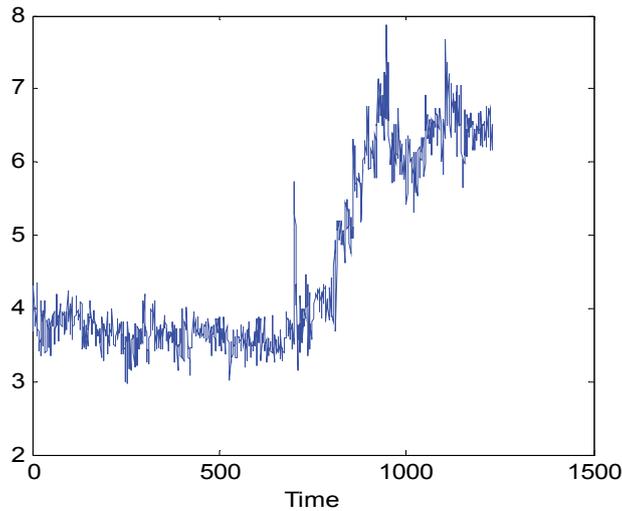


Figure 4.4 Degradation data of water pump

(2) Planetary Gearbox Data

The run-to-failure experiment data of a planetary gearbox test rig are used. The data were obtained by Reliability Research Lab, University of Alberta [85]. The experiment was operated for 762 hours. The experiment stopped when the gear teeth lost more than 60% materials. Two low sensitivity accelerometers and two high sensitivity accelerometers were used to collect the vibration signals. A 5-minute time span of vibration signals were collected every two hours.

The indicator used is the first order sideband with the larger amplitude in the frequency spectrum. This sideband indicator was reported being able to reflect the growth of gear tooth wear and used for prognosis of planetary gearbox condition in [23]. This thesis uses a subset of the indicator values computed by [23] to generate a degradation dataset which contains 1445 data points corresponding to the experiment time from the 320th hour to the 762th hour. From the start of experiment to the 319th hour, the sideband values generally remain stable.

Figure 4.5 shows the trend of sideband starting from the time point “0” corresponding to the experiment time at the 320th hour. At time point 485, obvious pits were found on gear teeth during the inspection. The pits caused the sideband values to increase until time point 604 as shown in the figure. After time point 604, the sideband values started to decrease until time point 927. A possible reason for this phenomenon is that the pits were removed due to the abrasive function of gear tooth engagement. However, the sideband values around time point 927 are even smaller than those in the stable stage. This is difficult to explain. After time point 927, the loss of a large volume of gear tooth materials caused the sideband values to increase rapidly to a higher level until the experiment terminated. The RNL is computed as 0.7 which implies very large noise effects.

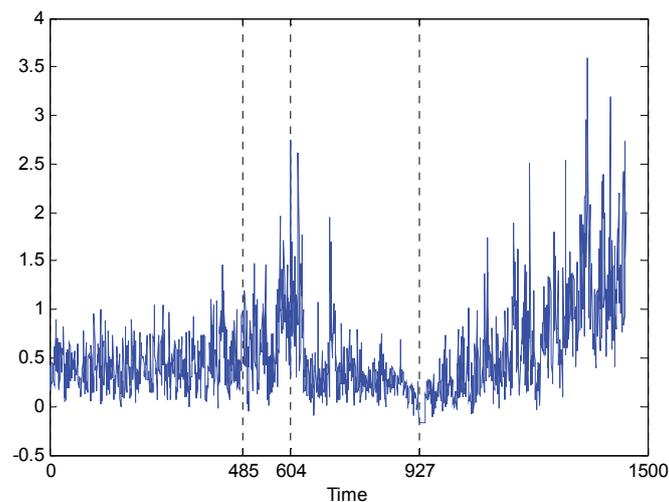


Figure 4.5 Degradation data of planetary gearbox

(3) Bearing Data

The run-to-failure data of a bearing test rig are used [86]. Four bearings on one shaft were tested. The bearings were working at 2000 rpm rotation speed under 6000 lbs radial load until the

system failed. Two high sensitivity accelerometers were used on each bearing to collect vibration signals. The vibration signals were collected every 10 minutes for 1 second snapshot. Details of the bearing test rig can be found in [87]. The bearing # 1 had an outer race failure around 700 time point. The indicator of kurtosis, the fourth standardized moment, was computed using the vibration signals. The equation of the kurtosis is given as follows [88]:

$$\text{Kurtosis} = \frac{\sum_{t=1}^T [x(t) - \mu]^4}{T(\sigma^2)^2}, \quad (4.5)$$

where $x(t)$, $t = 1, 2, \dots, T$ represents data series, T represents the length of the data series, μ and σ^2 represent the mean and the variance of the data series, respectively.

Figure 4.6 shows the trend of the indicator. The RNL is computed as 0.19.

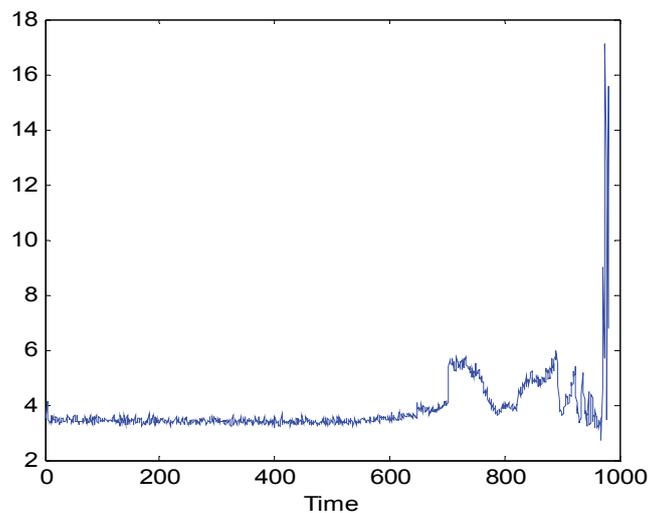


Figure 4.6 Degradation data of bearing

4.1.3. Data Normalization

Data normalization will be conducted to each dataset in order to ensure the results obtained are applicable to other cases without limitations. For reliability estimation, it is reasonable to assume

that a set of training data, $X_{tr}=(x_1, \dots, x_n)^T$, which correspond to the normal condition of the system are available. A new data point, x_i , can thus be normalized by following expression:

$$x'_i = \frac{x_i - \min(X_{tr})}{\max(X_{tr}) - \min(X_{tr})}, i = n + 1, n + 2, \dots \quad (4.6)$$

It should be noted that Eq. (4.6) is unable to guarantee the normalized data are within the range of $[0, 1]$, since the collection of condition monitoring data is an ongoing process. However, with the normalization, the values of the collected data can be restrained in a relatively small range as opposed to the original range, so the negative effects on OC-SVM and KDE due to the large differences in the original data values could be effectively mitigated.

4.2. Introduction to Hua's Method

Hua et al. [28] jointly used OC-SVM solution path and KDE for estimating system reliability.

Figure 4.7 shows the frame work of Hua's method.

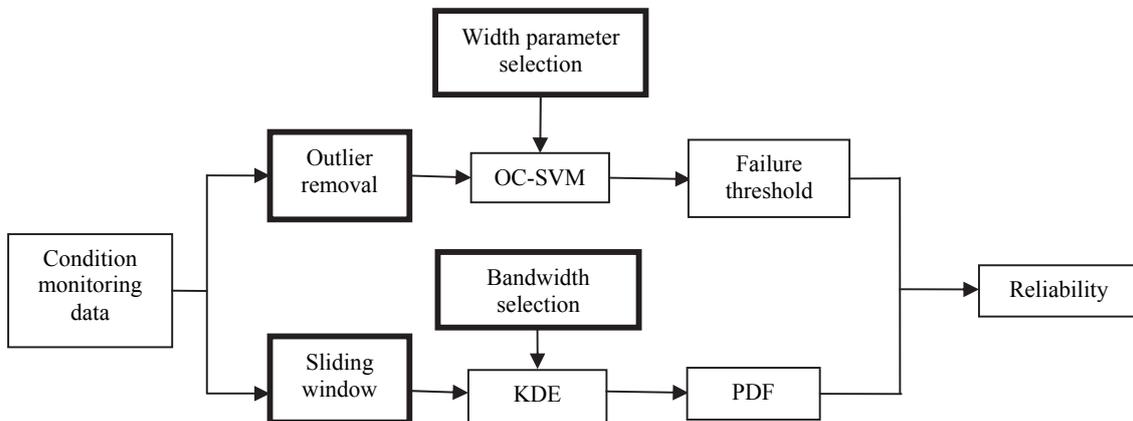


Figure 4.7 Frame work of Hua's method [28]

The method uses the sliding window technique to split the entire time span into multiple sliding windows. The size of the sliding window is a preset fixed value. The sliding windows could be overlapping depending on the selection of the sliding distance between sliding windows.

When the sliding distance is smaller than window size, the sliding windows will be overlapping and vice versa. For each sliding window, KDE uses all the collected data available in the sliding window to estimate the PDF of the data in the sliding window. Once it is done, the probability density of any data points in the range covered by the PDF is available.

Next, a group of data points with a preset number, N_{split} , are chosen from the range of the PDF and the interval, δ , between every two neighboring data points are identical. The product of δ and N_{split} should cover the range of the PDF. The selection of δ and N_{split} will be discussed in Section 4.4. The chosen data points then enter into the OC-SVM process and are labeled as normal or abnormal data by the OC-SVM solution path algorithm. Once all the chosen data points are labeled, the PDF obtained from KDE is used to estimate the reliability, namely the probability of the labeled normal data points, for the sliding window by the following expression:

$$R_i = \sum_{j=1}^{N_{\text{normal}}} p_j \delta, j = 1, \dots, N_{\text{normal}}, \quad (4.7)$$

where R_i represents the reliability for the i th sliding window, p_j represents the probability density of j th labeled normal data point and N_{normal} represents the number of labeled normal data points and is no larger than N_{split} . Such process is repeated until the reliability for each sliding window is estimated.

There are four important aspects that are not described clearly in Hua's method [28]: (1) the selection of the width parameter of OC-SVM, (2) the selection of the bandwidth parameter of KDE, (3) The selection of the window size of KDE, and (4) The removal of outliers for OC-SVM and KDE. The four aspects relate to thresholding and probability density estimation which may have negative impact on reliability estimation if not properly addressed.

The aspects (1), (2) and (3) are associated with parameter selection. The parameters are the width parameter, σ , of OC-SVM, the bandwidth parameter, h , of KDE, and the window size for KDE. As they are not correlated mathematically, it is assumed that they are independent to each other in terms of the influences on reliability estimates. However, for the investigations in the following sections, when one parameter is studied, the other parameters will be properly selected to avoid any adverse impact on the interested one.

4.3. Selection of the Width Parameter of OC-SVM for Reliability Estimation

As reported in [43][44][45][46][47], the width parameter, σ , of the Gaussian kernel is important to the performance of OC-SVM. The function of the Gaussian kernel is as follows:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right), \quad (4.8)$$

A small value of σ will yield an under-estimated threshold which may cause normal data to be misclassified as abnormal data and results in a smaller reliability value than the true one. On the contrary, a large value of σ will yield an over-estimated threshold which may cause abnormal data to be misclassified as normal data and results in a larger reliability value than the true one. This section investigates the impact of σ value on reliability estimates using the enumeration method.

4.3.1. Selection of the Width Parameter Using the Enumeration Method

Practically, the real reliability of a system being monitored is not accessible, so the optimal value of σ is difficult to determine. However, a reasonable trend of reliability estimates can be defined as per its relation to the trend of degradation data of the interested system. This thesis uses

incomplete enumeration method [89] to observe a wide range of σ values and find out a particular range of σ values that can offer a reliability trend satisfying the following Evaluation Criteria:

- (1) At the early stage, the reliability starts from 1 and remains at 1; then the reliability decreases after a turning point at which a fault occurs.
- (2) At the late stage, the reliability trend should be in line with the degradation trend of the data; it means that the decrease in reliability values should correspond to the change of degradation data values.
- (3) The reliability trend should change gradually without sudden spikes or drops; however, a small scale fluctuation is allowed.

To find out an applicable range of σ values, the incomplete enumeration method will be executed per the following steps:

Step 1: Create a set of σ values by an interval of the order of magnitude as: [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000].

Step 2: Estimate the reliability for each σ value and determine an applicable range of σ value based on the Evaluation Criteria.

Step 3: Create a subset of σ values with an appropriate interval based on the results of Step 2.

Step 4: Repeat Steps 2 and 3 until the obtained reliability estimates all satisfy the Evaluation Criteria.

The key parameters other than the σ value are fixed for the tests. The number of training data is 200. The window size is 200. The sliding distance of the window is 50. Hua's method is selected for bandwidth selection. These parameters remained the same, unless otherwise mentioned.

The simulation data with RNL=0.2 is first tested. Figure 4.8 shows the obtained reliability trends. The simulation data show the start of observable increases in value around time 650. It is seen that for σ values equal to 1, 10, 100 and 1000, the reliability trends are straight lines with reliability close to 1. The reliability estimates based on these σ values cannot reflect degradation pattern of simulation data; therefore, the four σ values are ruled out by the Evaluation Criterion (1). For $\sigma=0.1$, the reliability estimates start to decrease around time 850 and show a degrading trend. However, the turning point of the decrease of reliability occurs later than the one in simulation data; therefore, the σ value of 0.1 is ruled out by the Evaluation Criterion (2). For $\sigma=0.01$, the reliability estimates start to decrease around time 650 and show a reasonable degrading trend. The reliability estimates based on the σ value of 0.01 satisfy all Evaluation Criteria. For $\sigma=0.001$, the reliability estimates decrease to 0.9122 at time 200 while the simulation data are quite stable. For $\sigma=0.0001$, the reliability estimates drop to 0.34 and remain at this value until time 600. Both σ values of 0.001 and 0.0001 are ruled out by the Evaluation Criteria (1) and (2).

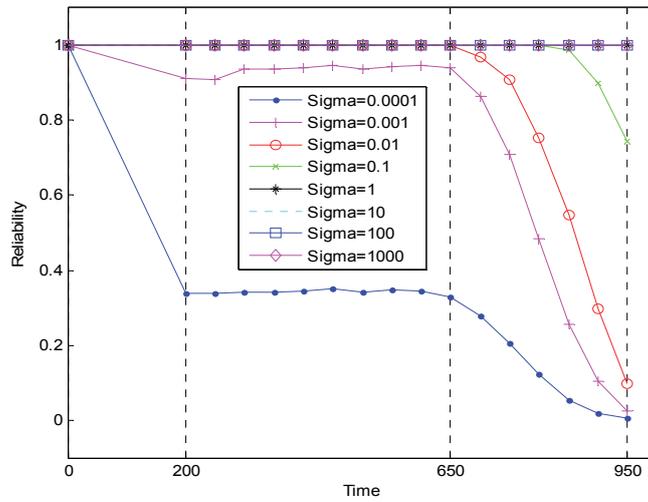


Figure 4.8 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.2)

The σ value of 0.01 is selected based on Step 2. A subset of σ values between 0.001 and 0.01 are chosen with an interval of 0.001. Then Steps 3 and 4 are implemented. Figure 4.9 shows that the reliability estimates for σ values from 0.002 to 0.01 are able to reflect a degradation pattern with the turning point at time 650. At time 950, the reliability estimates drop to around 0.1. The σ values from 0.002 to 0.01 offer the reliability trends satisfying the Evaluation Criteria.

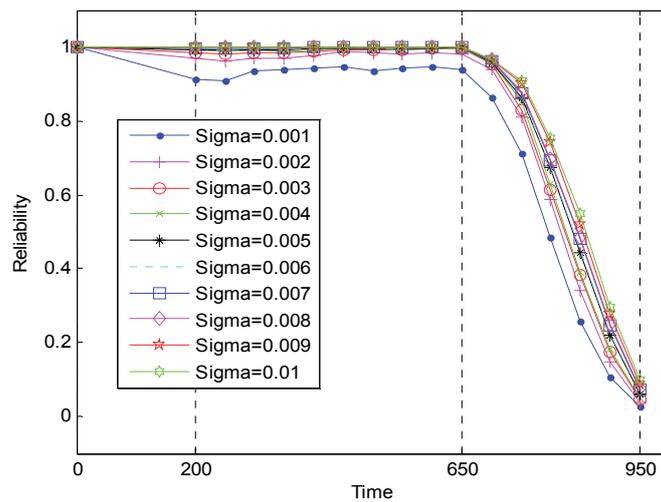


Figure 4.9 Reliability estimates with σ values from 0.001 to 0.01 (RNL=0.2)

Next, simulation data with RNL=0.3 is tested. The σ values for testing are the same as Case RNL=0.2. Figure 4.10 shows that the reliability estimates with σ values of 0.1, 1, 10, 100 and 1000 do not satisfy the Evaluation Criteria (1) or (2). Similar to Case RNL=0.2, the reliability estimates show a pattern that satisfies the Evaluation Criteria when σ value equals to 0.01. Also, the σ values equal to 0.0001 and 0.001 do not satisfy the Evaluation Criteria (1) or (2), as the reliability estimates show unexpected drops at the early stage.

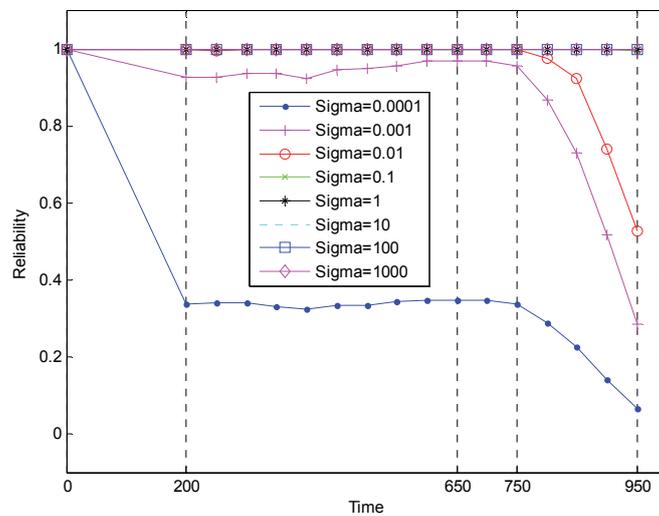


Figure 4.10 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.3)

Following the Steps 3 and 4, a subset of σ values between 0.001 to 0.01 are further tested. The selected σ values are from 0.001 to 0.01 with an interval of 0.001. Figure 4.11 shows that the σ values from 0.002 to 0.01 have similar degradation pattern and can reflect the degradation of the data. For the σ values of 0.002 and 0.003, the reliability estimates at time 200 decrease to 0.9680 and 0.9857, respectively. As per Evaluation Criterion (3), these small scale fluctuations are allowed. In summary, the σ values from 0.002 to 0.01 are all applicable for Case RNL=0.3.

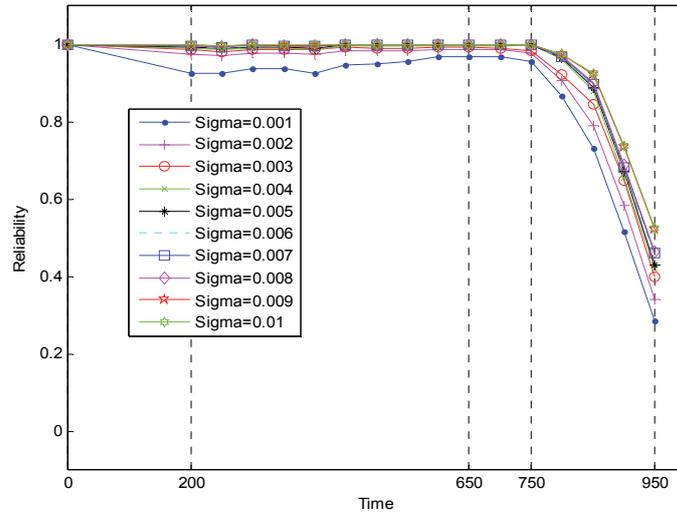


Figure 4.11 Reliability estimates with σ values from 0.01 to 0.1 (RNL=0.3)

It is noticed that the trends of reliability decrease slower for Case RNL=0.3 than for Case RNL= 0.2. The reliability estimates at time 950 are around 0.5 for RNL=0.3, while they are around 0.1 for RNL=0.2. Such differences are caused by the noise effects on the data. The first 200 data points are used as training data for the OC-SVM, so the noising data are regarded as the part of the normal data by OC-SVM. When the degradation starts, the data values are still within the range of noise levels, so the trend of data exhibits a stable pattern. When the data values exceed the range of noise levels, the trend of data starts to show degradation pattern. This means that the noise delays the data showing the degradation pattern. Since the trend of reliability reflects the degradation trend of the data, the reliability estimates also drop slower for Case RNL=0.3 than for Case RNL=0.2. Based on the observation, it is expected that as the RNL increase, the reliability estimates will decrease slower and slower; in other words, for the same time point, the reliability estimates for the larger RNL values will be greater than those for the smaller RNL values.

Next, simulation data with RNL=0.5 are tested. The results are similar to the Case RNL=0.3. Figure 4.12 shows that when $\sigma=0.001$, the obtained reliability estimates satisfy the Evaluation Criteria while other σ values do not.

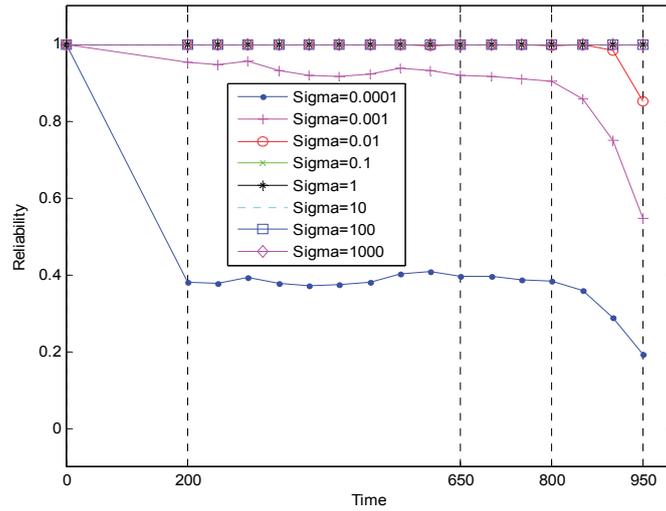


Figure 4.12 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.5)

The subset of σ values is selected between 0.001 and 0.01 and the results are shown in Figure 4.13. The results are quite similar to Case RNL=0.3 as shown in Figure 4.11. Also as expected, the reliability estimates decrease slower than Case RNL=0.3 due to the larger noise. Taking $\sigma=0.01$ as an example, the reliability estimates are 0.9762 at time 800 and 0.5270 at time 950 for Case RNL=0.3 and are 0.9977 at time 800 and 0.8522 at time 950 for Case RNL=0.5.

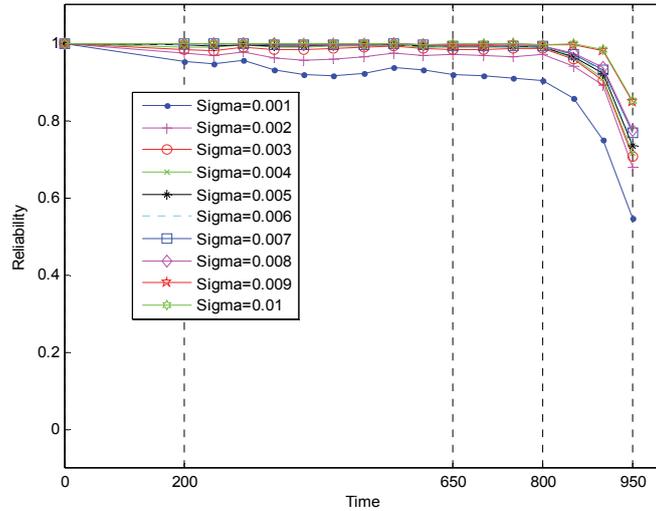


Figure 4.13 Reliability estimates with σ values from 0.001 to 0.01 (RNL=0.5)

In terms of data pattern (Figure 4.1), it is noticed that the degradation degree (the degree of the increase of data value after time 650) of RNL=0.5 is reduced as compared to the ones of RNL=0.2 and RNL=0.3 in Figure 4.2. Since the degradation degree of the simulation data without noise is known, it may be concluded that the reliability estimates obtained for Case RNL=0.5 are not able to fully represent the true degradation degree of the system. However, it is not caused by the deficiency of the method; this is due to the overwhelming effects of noise. The solution to improve the results is to conduct de-noising prior to using the reliability estimation method. Nevertheless, this is not the scope of this thesis and will not be further discussed.

At last, simulation data with RNL = 0.7 is tested. Figure 4.14 and Figure 4.15 show the results similar to Case RNL=0.5. This is not beyond our expectation, so no further descriptions are provided here. Again, the noise effects are overwhelming which makes the reliability estimates be unable to reflect the true degradation of the data.

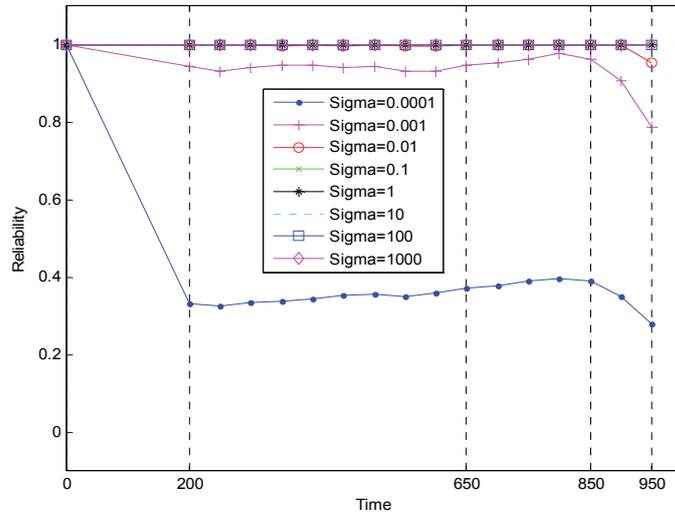


Figure 4.14 Reliability estimates with σ values from 0.0001 to 1000 (RNL=0.7)

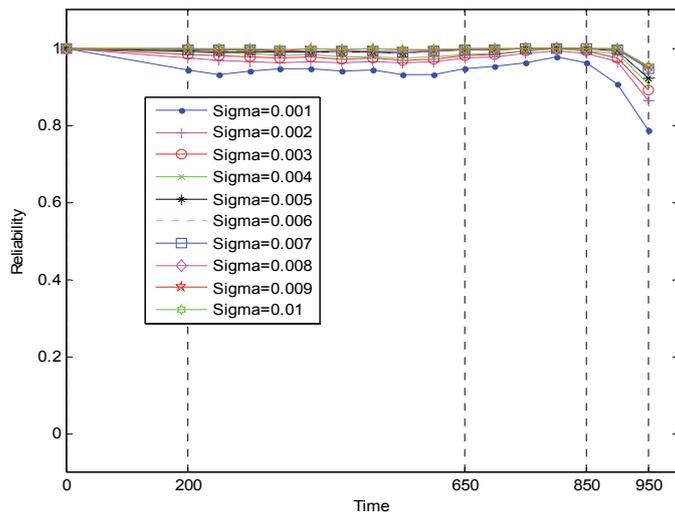


Figure 4.15 Reliability estimates with σ values from 0.001 to 0.01 (RNL=0.7)

In summary, based on the investigations under the cases with different RNLs, it is found that when σ values locate at the range of [0.002, 0.01], reliability estimates that satisfy the Evaluation Criteria can be expected. In terms of the cases with RNLs greater than 0.3, de-noising of data prior to reliability estimation are recommended in order to access applicable reliability estimates. Be noted that the investigations on the cases with RNLs of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7 are all conducted, but for illustration purposes only the cases with RNLs of 0.2, 0.3 0.5 and 0.7 are

provided with figures. Table 4.1 provides a summary of the observations from all cases conducted on which it can be concluded that a reasonable trend of reliability estimates could be expected if the σ value of OC-SVM is selected between 0.002 and 0.01. Of course, one could further conduct enumeration starting from this range to access more favorable results of reliability estimation. One thing to be strongly recommended is that for the data with large noise it is better to carry out a de-noising process prior to reliability estimation.

Table 4.1 Observations of σ values for different RNLs

RNL	Applicable range for σ values	Observations	Remarks	
0.1 0.2 0.3	$0.002 \leq \sigma \leq 0.01$	$\sigma < 0.002$	Reliability estimates have drops at early stage and are unacceptable to Evaluation Criteria (1) and (2).	De-nosing of data is not required
0.4 0.5	$0.002 \leq \sigma \leq 0.01$	$0.002 \leq \sigma \leq 0.01$	Reliability estimates are acceptable to Evaluation Criteria.	De-nosing of data is recommended
0.6 0.7	$0.002 \leq \sigma \leq 0.01$	$\sigma > 0.01$	Reliability estimates show a flat pattern and are unacceptable to Evaluation Criteria (1) and (2).	De-nosing of data is required

4.3.2. Testing with Simulation and Experiment Data

This section tests the observations of σ values obtained with the simulation data. Three reported methods, automated method based on variance and mean (VM) [46], minimum distance method (MD) [47] and distance from the farthest and nearest neighbors method (DFN) [44] mentioned in Section 3.1.2, are also tested using the observations. At first, the σ value is calculated using each reported method. Next, the calculated values are compared with the obtained applicable range of [0.002, 0.01]. If the values are within the applicable range, it is expected to provide favorably

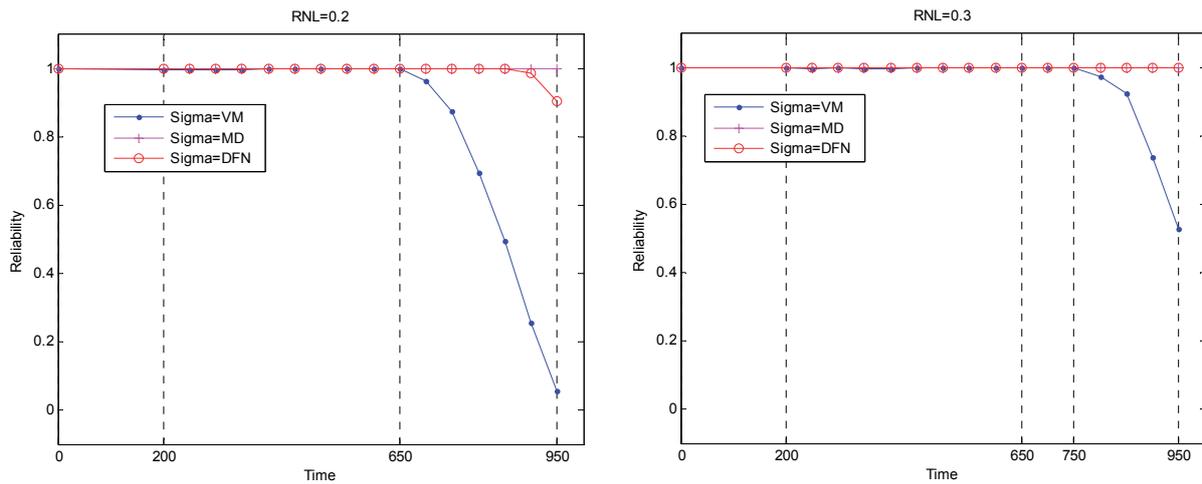
reliability estimates. At last, figures are plotted to verify the expectations. Both simulation and experiment data are used for testing.

(1) Results of simulation data

Table 4.2 shows the results of σ values calculated by the three reported methods. It is seen that only the VM method provides the σ values within the applicable range of [0.002, 0.01]; the σ values of other methods are all above 0.01. Based on Table 4.1, it is expected a flat pattern of reliability estimates for the MD and DFN methods. Figure 4.16 shows the results which are in agreement with our expectation.

Table 4.2 The σ values calculated using reported methods for simulation data

Method	σ values of simulation data						
	RNL=0.1	RNL=0.2	RNL=0.3	RNL=0.4	RNL=0.5	RNL=0.6	RNL=0.7
VM [46]	0.0023	0.0076	0.0094	0.0021	0.0020	0.0082	0.0083
MD [47]	0.4363	0.4363	0.4363	0.4363	0.4363	0.4363	0.4363
DFN [44]	0.1752	0.1823	0.1762	0.1865	0.1698	0.1747	0.1901



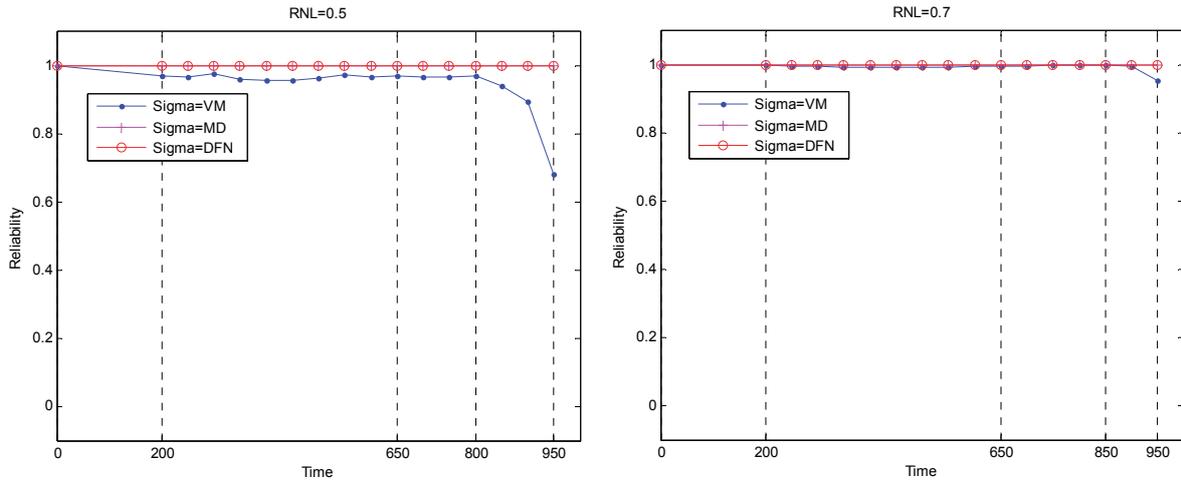


Figure 4.16 Reliability estimates for simulation data using reported σ selection methods

(2) Results of experiment data

Three sets of experiment data described in Section 4.1.2 are used for testing. Table 4.3 shows the σ values obtained using the three reported method. It is seen that only the VM method provides the σ values within the applicable range for all three data sets while other methods do not. It is noted that gearbox data has RNL=0.7. Based on Table 4.1, the reliability estimates may not reflect the degradation pattern of the data if the de-noising of data is not performed.

Table 4.3 The σ values calculated using reported methods for experiment data

Method	σ values of experiment data		
	Water pump data (RNL=0.17)	Gearbox data (RNL=0.7)	Bearing data (RNL=0.19)
VM [46]	0.0037	0.0034	0.0055
MD [47]	0.4363	0.4363	0.4363
DFN [44]	0.1799	0.1826	0.1799

Figure 4.17 shows the reliability estimates using water pump data with RNL=0.17. It is seen that only the VM method provides a reasonable reliability trend. The data of water pump start degrading around time 800 hour. The reliability estimates slightly decrease at early stage and

start degrading around time 800 hour until time 1000 hour at which the reliability reach 0. The MD and DFN methods provide a flat pattern of reliability estimates which is in agreement with our expectation.

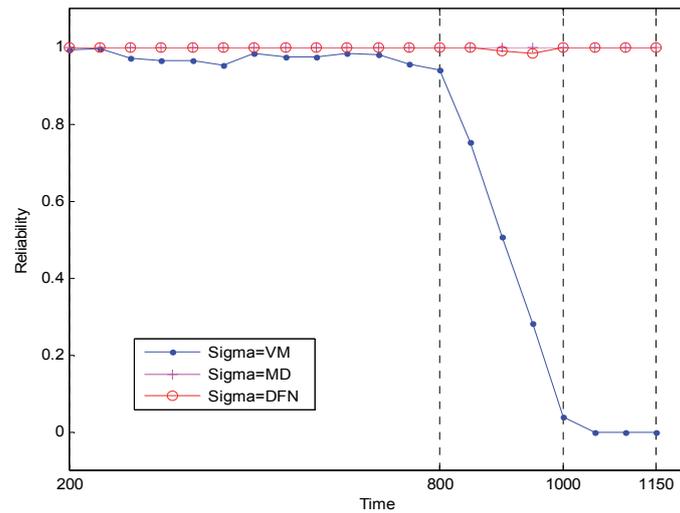


Figure 4.17 Reliability estimates for water pump using reported σ selection methods

Figure 4.18 shows the reliability estimates using gearbox data. The gearbox data has RNL=0.7, so the degradation pattern is not clear. The MD and DFN methods provide a flat pattern of the reliability estimates. The reliability estimates based on the VM method reflect the changes of gearbox data. However, the reliability estimates are affected by the large noise. De-noising methods are required before applying the reliability estimation.

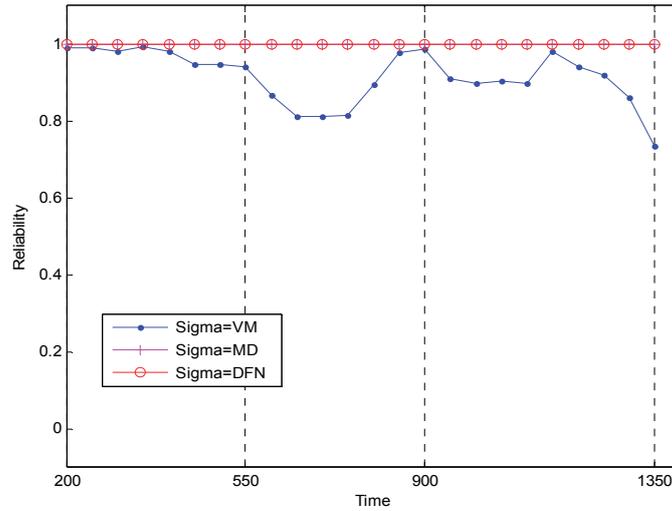


Figure 4.18 Reliability estimates for planetary gearbox using reported σ selection methods

Figure 4.19 shows the reliability estimates using bearing data with RNL=0.19. The VM method provides a reliability trend reflecting the degradation trend of bearing data while other methods do not. The reliability estimates at the early stage are close to 1 and start to decrease at time 650 until reach 0.1 around time 900.

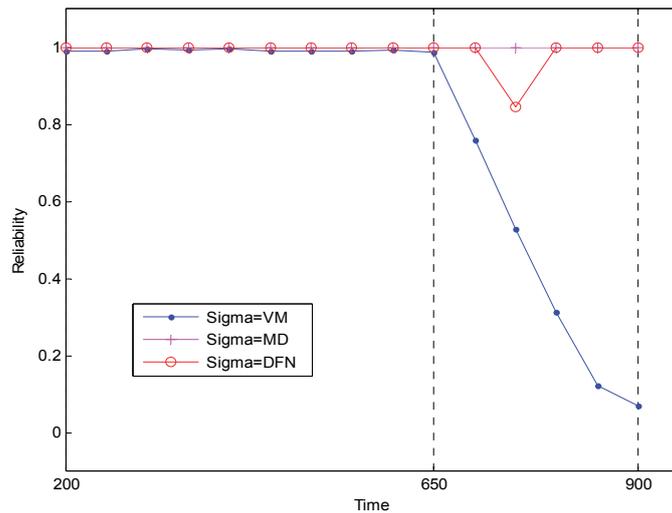


Figure 4.19 Reliability estimates for bearing units using reported σ selection methods

4.3.3. Summary

In Section 4.3, we have investigated the impact of σ values of OC-SVM on reliability estimation. Based on the tests using simulation data with various noise effects, the σ values within the range of [0.002, 0.01] are recommended for reliability estimation. Three reported methods for σ selection are employed to test the validity of the observations. The results show that the VM method which consistently provides σ values lying in the recommended range outperformed its counterparts for all tested simulation data and experiment data. This shows the good applicability of our recommendations on the selection of σ in practical applications. The results also suggest that it need to carry out de-noising process for the data with RNL values greater than 0.3 in order to obtain reasonable reliability estimates.

4.4. Selection of the Bandwidth Parameter of KDE for Reliability Estimation

Bandwidth, h , selection is critical for KDE [78][79][90][91][92][93][94][95][96][97]. A small value of h will yield a spiky shape of PDF which is difficult to interpret [33]. A large value of h will yield an over-smooth PDF and hide the structure of data [33]. This section investigates the impact of h value on the reliability estimates using the enumeration method.

4.4.1. Selection of the Bandwidth Parameter Using the Enumeration Method

This section will use the Evaluation Criteria introduced in Section 4.3 to evaluate the obtained reliability estimates. The incomplete enumeration method is used to find out the proper range of bandwidth, h , in accordance with the same steps introduced in Section 4.3. The σ value of OC-SVM is selected to be 0.003 based on the observations of Section 4.3. Other parameters remain the same as Section 4.3.

The simulation data with $RNL=0.2$ are first tested and the results are shown in Figure 4.20. It is found that when h equals to 10, 100 and 1000, the reliability estimates drop to 0 at time 200 which does not satisfy the Evaluation Criteria (1) and (2) stated in Section 4.3. When h equals to 1, the reliability estimates drop to 0.4 at time 200 and remain at this value until time 800. After time 800, the reliability estimates decrease to 0.1 at time 950. The trend does not satisfy the Evaluation Criteria (1) and (2) stated in Section 4.3. When h equals to 0.0001, 0.001, 0.01 and 0.1, the reliability estimates reflect the degradation pattern of the simulation data and satisfy all the Evaluation Criteria stated in Section 4.3. Based on Figure 4.20, h values greater than 1 are ruled out by the Evaluation Criteria and h values less than and equal to 0.1 can yield reasonable reliability estimates.

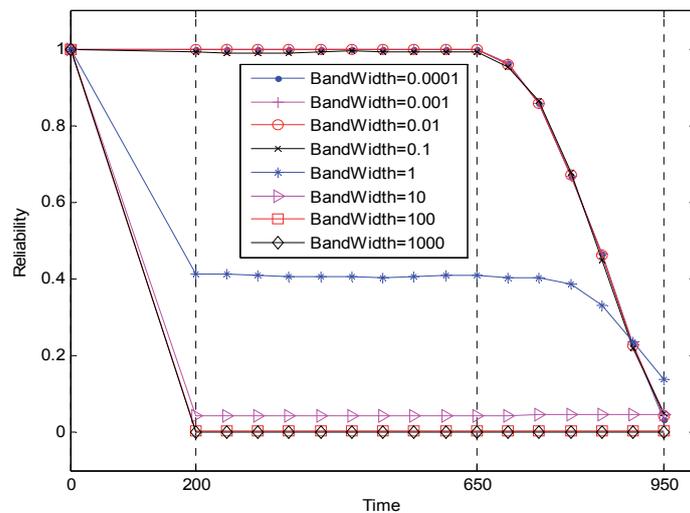


Figure 4.20 Reliability estimates with h values from 0.0001 to 1000 ($RNL=0.2$)

Figure 4.21 shows the PDFs of the data for each time window with different h values. It is found that the PDFs with h values equal to 1, 10, 100 and 1000 are overly smooth and obscure the structure of data. Such PDFs have larger ranges of data values than the true one. For example, Figure 4.22 shows the PDF for the first sliding window from time 0 to 200 when $h=10$. The

normalized data points within $[0,1]$ are used to determine the PDF while the PDF obtained by KDE covers the range from -30 to 30 approximately. Because the data points within $[0,1]$ are also used for training OC-SVM, those out of this range are considered abnormal. As a result, a small reliability value is yielded. The PDFs with h equal to 1 , 100 and 1000 have similar shapes and small reliability values are yielded for each time window. In the following, the scenarios with $h \leq 0.1$ are further investigated.

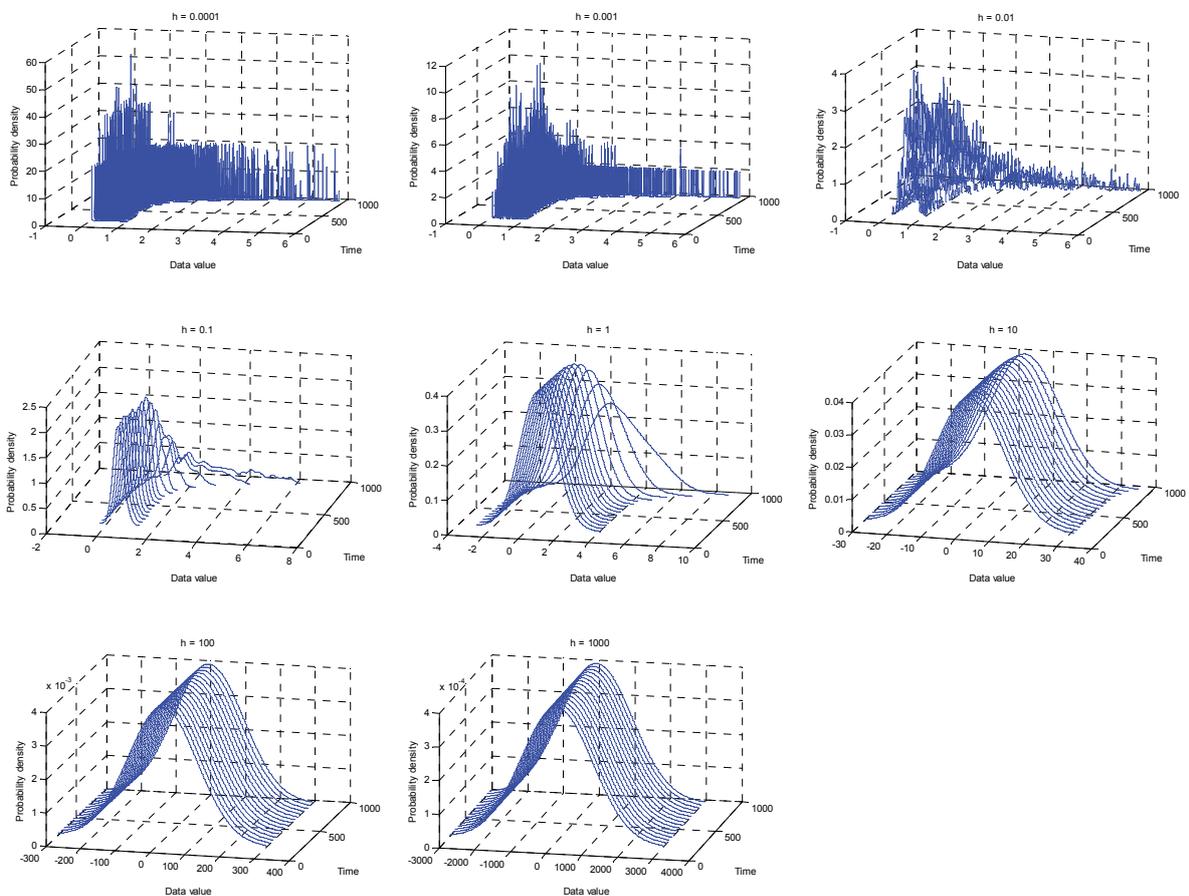


Figure 4.21 The PDF for each sliding window with h values from 0.0001 to 1000 (RNL=0.2)

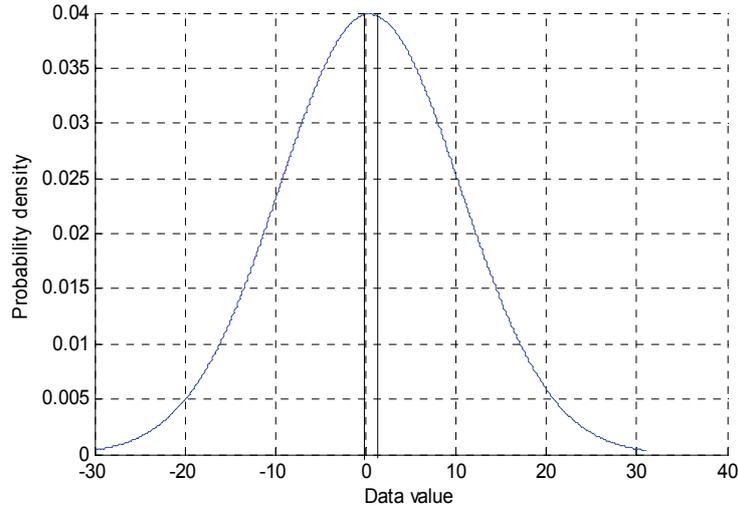


Figure 4.22 The PDF for the first sliding window with $h=10$ (RNL=0.2)

Figure 4.20 shows that h of 0.0001, 0.001, 0.01 and 0.1 can provide reasonable reliability estimates; however, the shapes of the PDFs are quite different based on Figure 4.21. For $h=0.1$, the shape of PDF is reasonably smooth as desired; while for $h=0.0001$, 0.001 and 0.01, the shapes of PDF are spiky. It is reported that such spiky PDF is not suitable for probability density estimation [33]. However, the trends of reliability estimates with h equal to 0.0001, 0.001, 0.01 and 0.1 satisfy the Evaluation Criteria stated in Section 4.3 and show the same favorable pattern as the one with $h=0.1$. The results we have obtained are not in agreement with the reported study.

The study in [33] is interested in the probability of unknown data points which depends on the shape of the PDF, so the selected h value must show the proper shape of PDF. Based on Eq. (4.7) in Section 4.2, reliability is determined by the sum of the products of the interval, δ , between adjacent data points and the probability density, p , of the normal data points. The normal data points are identified by OC-SVM and are obtained by splitting the range of PDF based on a pre-selected number of splits, N_{split} . When h is too large the range of PDF is stretched

making more split data points locate outside of the range of the original data points and resulting in smaller reliability estimates than the true ones. A small value of h does not have this disadvantage. Nevertheless, the reliability estimates are also influenced by other parameters such as N_{split} and δ .

The interval, δ , is determined by the range of the PDF and the number, N_{split} . For example, given that the range of the PDF is $[0,1]$, when $N_{\text{split}}=10000$, $\delta = \frac{1-0}{10000} = 10^{-4}$. The interval, δ , performs as a sampling frequency such that if a sufficient small interval value is chosen, the split data points will be able to capture the true shape of the PDF and the reliability estimates will thus meet the Evaluation Criteria. On the contrary, if the interval is not small enough, the true shape of the PDF may not be captured. As a result, the reliability estimates may be larger or smaller than the true ones and thus cause the trend of reliability estimates not satisfying the Evaluation Criteria.

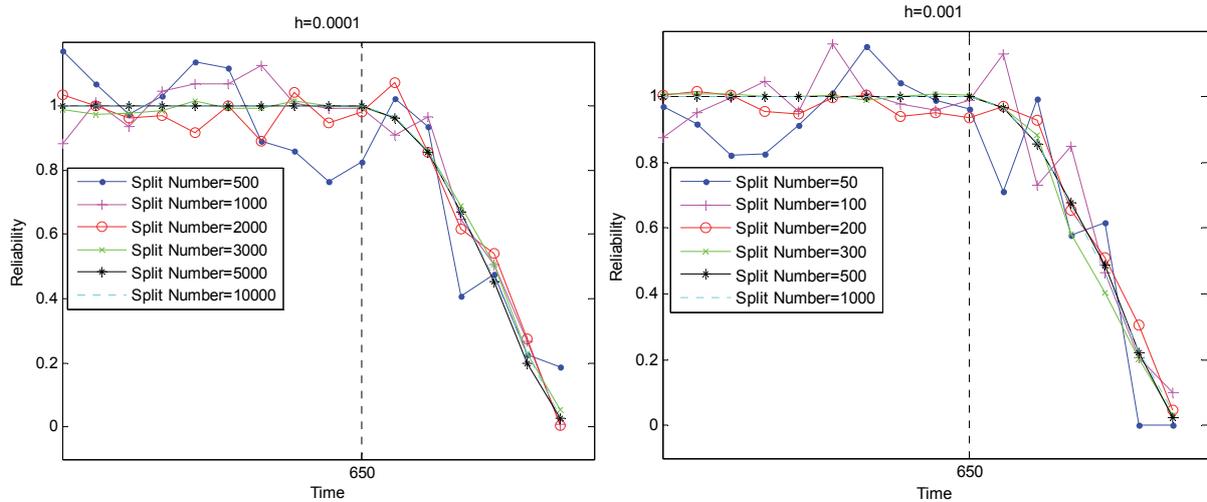
Based on above discussions, one can simply select a very large N_{split} value for a given h to expect good reliability estimates since N_{split} is inversely proportional to δ . The negative effect of this is the loss of computational efficiency. As mentioned in Section 3.2, the overall PDF of all original data points is the sum of the kernel function of each data point and the width of kernel function is determined by h . For this reason, it would be reasonable to select a value of N_{split} based on h so that the split data points are able to capture the main shape of the kernel function of each data point, and thus the overall PDF. Inspired by this idea, given a set of training data, $\mathbf{X}_{\text{tr}}=(x_1, \dots, x_n)^T$, the following expression is formulated:

$$\frac{\max(\mathbf{X}_{\text{tr}}) - \min(\mathbf{X}_{\text{tr}})}{N_{\text{split}}} = \delta \leq ah, \quad (4.9)$$

where a is a coefficient relating h to N_{split} and needs to be determined. The maximum difference among the training data points is used to approximate the range of overall PDF. For the normalized data, Eq. (4.9) can be simplified as:

$$N_{\text{split}} \geq \frac{1}{ah}. \quad (4.10)$$

To look for a proper value of a , we have conducted tests for various combinations of h and N_{split} . The intent is to find a value of a such that when a certain h value is picked an appropriate N_{split} value can be automatically determined. The simulation data with $\text{RNL}=0.2$ are adopted and are normalized to the range of $[0,1]$. Figure 4.23 shows the plots of reliability estimates under different N_{split} values with h equals to 0.0001, 0.001, 0.01 and 0.1. The Evaluation Criteria are adopted to assess the reliability estimates. As discussed above, larger N_{split} values can provide better reliability estimates. Figure 4.23 shows that the results are in agreement with this observation. Take $h=0.0001$ as an example, when $N_{\text{split}} \geq 3000$ the reliability estimates satisfy the Evaluation Criteria. When $N_{\text{split}} \leq 2000$, the reliability estimates show large scale fluctuations which are not reasonable.



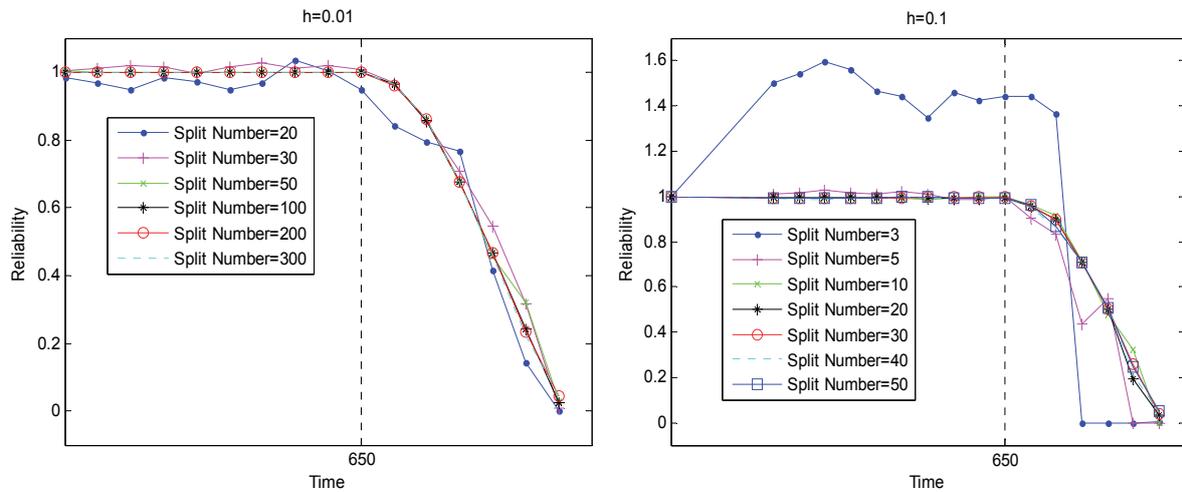


Figure 4.23 Reliability estimates with different h and N_{split} (RNL=0.2)

More h values between 0.01 and 0.1 are investigated in order to get proper a values. Only the quantitative results are given here. By plugging the pairs of h and N_{split} values into Eq. (4.10), the values of a are obtained. Table 4.4 lists the smallest applicable N_{split} values, the corresponding CPU times, and the calculated a values based on the tests. The obtained a values are rounded to the nearest integer.

Table 4.4 Selection of a for different combinations of N_{split} and h (RNL=0.2)

Bandwidth h	Split Number N_{split}	CPU time (s)	Coefficient a
0.0001	≥ 3000	7.6	≤ 3
0.001	≥ 300	1.65	≤ 3
0.01	≥ 30	0.437	≤ 3
0.02	≥ 30	0.350	≤ 2
0.03	≥ 20	0.234	≤ 2
0.04	≥ 16	0.406	≤ 2
0.05	≥ 16	0.468	≤ 2
0.06	≥ 16	0.347	≤ 2
0.07	≥ 10	0.343	≤ 1
0.08	≥ 10	0.359	≤ 1
0.09	≥ 10	0.359	≤ 1
0.1	≥ 10	0.328	≤ 1

Based on the results in Table 4.4.1, we can choose a as follows: when $h \leq 0.01$, $a=3$; when $0.02 \leq h \leq 0.06$, $a=2$; when $0.06 \leq h \leq 0.1$, $a=1$. It is also found that when $h \geq 0.01$ the consumed CPU times are relatively comparable. With above observations, a rule of thumb for the selection of h and N_{split} is given as:

$$N_{\text{split}} = \begin{cases} \frac{1}{3h}, & h \leq 0.01 \\ 30, & 0.01 < h < 0.1 \end{cases} . \quad (4.11)$$

In summary, the investigations using simulation data with RNL=0.2 suggest that the values of h no greater than 0.1 should be used for reliability estimation and N_{split} should be selected in accordance with Eq. (4.11). These observations are also applicable for the cases with RNL=0.1, 0.3, 0.4, 0.5, 0.6, and 0.7. Similar to the observations in Section 4.3, when RNL=0.6 and 0.7, de-noising method needs to be conducted to obtain reasonable reliability estimates. Due to page limit, these results are not presented here.

4.4.2. Testing with Simulation and Experiment Data

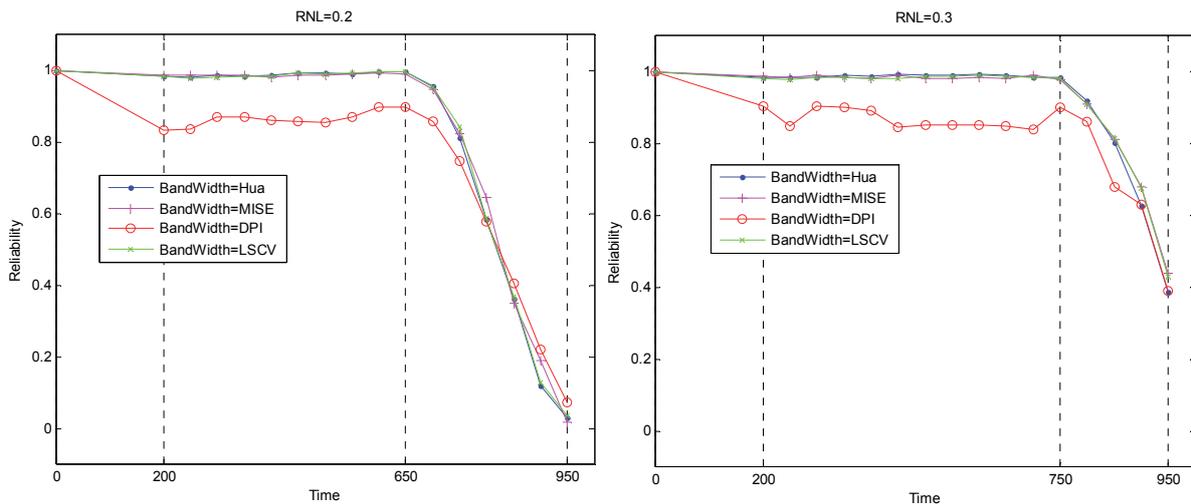
This section tests the observations of h and N_{split} values obtained in the preceding section. Four reported methods for selecting h value mentioned in Section 3.2 are used for testing. For each testing data set, h is estimated using the reported methods and compared with the range of [0.01, 0.1] with $N_{\text{split}}=30$ which is obtained based on the investigations in Section 4.4.1. For the h values within the range, the reliability estimates are expected to satisfy the Evaluation Criteria.

(1) Results of simulation data

Simulation data with the RNLs of 0.1 to 0.7 as introduced in Section 4.1 are used for calculating h values. Table 4.5 lists the calculated h values based on the four reported methods. Only the DPI method [81] provides the h values greater than 0.1 which is not in the range recommended by this thesis. Figure 4.24 shows the results using simulation data with RNLs of 0.2, 0.3, 0.4 and 0.5. The Hua [28], MISE [33] and LSCV [82] methods provide reasonable reliability estimates satisfying the Evaluation Criteria while the DPI method does not. These results are in agreement with our expectations.

Table 4.5 The h values calculated using reported methods for simulation data

Method	h values using simulation data						
	RNL=0.1	RNL=0.2	RNL=0.3	RNL=0.4	RNL=0.5	RNL=0.6	RNL=0.7
Hua [28]	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500
MISE [33]	0.0699	0.0668	0.0667	0.0663	0.0591	0.0655	0.0656
DPI [81]	0.2567	0.2546	0.2560	0.2551	0.2495	0.2561	0.2551
LSCV [82]	0.0586	0.0439	0.0793	0.0755	0.0589	0.0829	0.0730



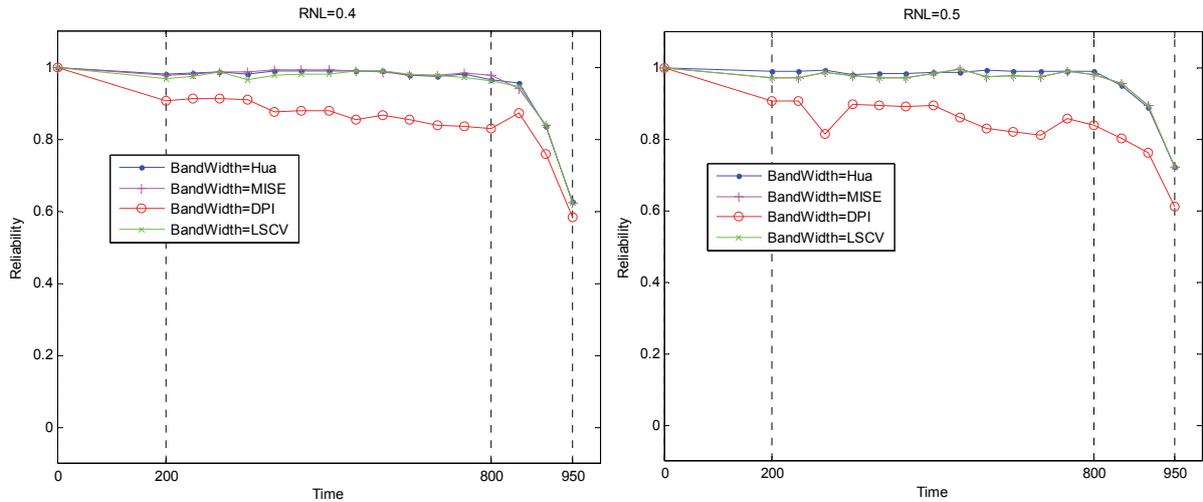


Figure 4.24 Reliability estimates for simulation data using reported h selection methods

(2) Results of experiment data

Three sets of experiment data described in Section 4.1.2 are used for testing. Table 4.6 lists the calculated h values using the four reported methods for each experiment data. It is seen that the DPI method provides the h values greater than 0.1 for all the experiment data while its counterparts provide values all within the range of $[0.01, 0.1]$, as recommended by this thesis research.

Table 4.6 The h values calculated using reported methods for experiment data

Method	h values using experiment data		
	Water pump data (RNL=0.17)	Gearbox data (RNL=0.7)	Bearing data (RNL=0.19)
Hua [28]	0.0500	0.0500	0.0500
MISE [33]	0.0653	0.0724	0.0431
DPI [81]	0.2532	0.2596	0.2405
LSCV [82]	0.0656	0.0930	0.0459

Figure 4.25 shows the result of the reliability estimates using water pump data [28]. It is seen that the DPI method provides the reliability estimates with unexpected declines at early stage which does not satisfy the Evaluation Criteria while its counterparts perform well as expected.

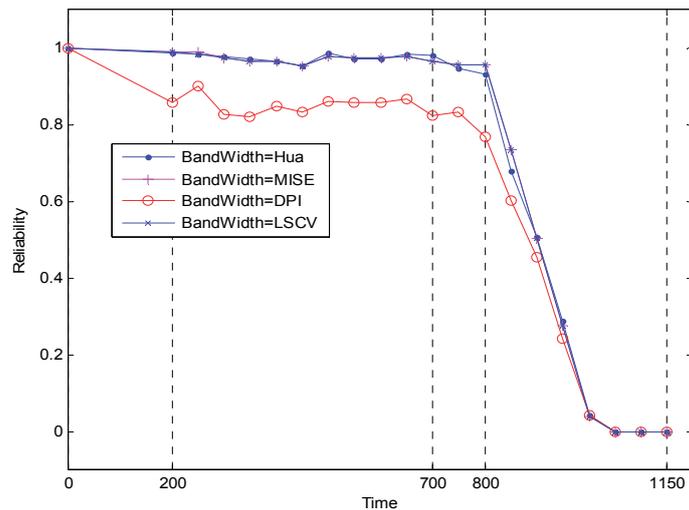


Figure 4.25 Reliability estimates for water pump using reported h selection methods

Figure 4.26 shows the reliability estimates using gearbox data [85]. The reliability estimates using the DPI method are lower than the ones using other three methods. The reliability estimates are affected too much by the large noise for all four methods. De-noising methods are required before implementing reliability estimation.

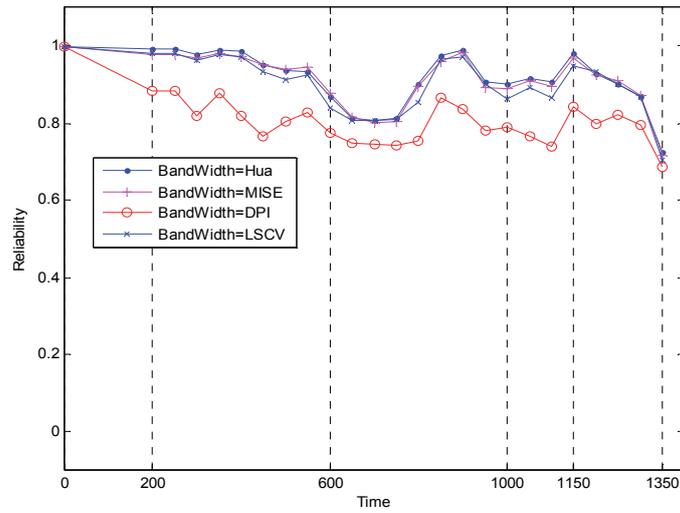


Figure 4.26 Reliability estimates for planetary gearbox using reported h selection methods

Figure 4.27 shows the reliability estimates using the bearing data [86]. Again, the DPI method does not perform as well as its counterparts due to the large h value it provides.

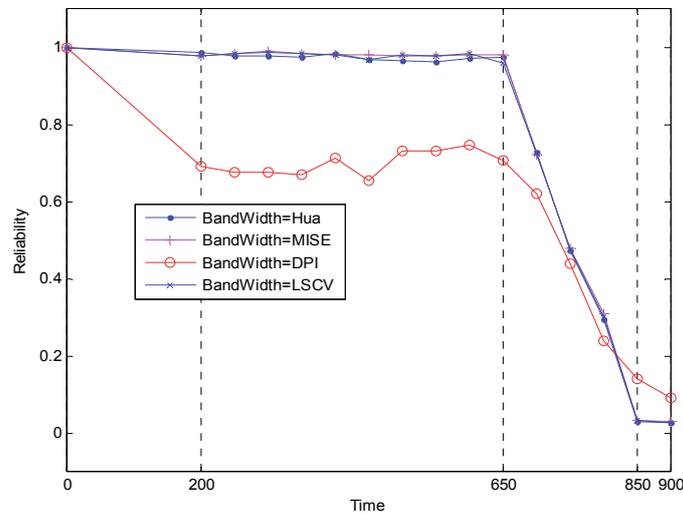


Figure 4.27 Reliability estimates for bearing units using reported h selection methods

4.4.3. Summary

Section 4.4 has investigated the impact of h values of KDE on reliability estimation. Based on the tests using simulation data with different noise effects, the h values within the range of [0.01, 0.1] are recommended for reliability estimation. Also, the formula for selecting the value of N_{split} which is paired with h value is provided. Four reported methods for h selection are used to test the validity of our observations. The Hua [28], MISE [33] and LSCV [82] methods are found able to provide h values within the range of [0.01, 0.1] and can provide reasonable reliability estimation for the tested simulation and experiment data. This shows the good applicability of our recommendations on the selection of h and N_{split} in practical applications.

4.5. Investigation on the Sliding Window Size for Reliability Estimation

Degradation data are non-stationary data as they usually exhibit a trend wherein the mean and the variance change over time. In Hua's method [28], the entire time span in which degradation data are collected is split into multiple overlapping sliding windows. It is assumed that the data in each sliding window are stationary as long as the size of the sliding window is sufficiently small. With this assumption, a fixed size of sliding window is specified and the data in each split sliding window are used for probability density estimation. However, in Hua's method, it did not mention how to determine the sliding window size to ensure the data in the sliding window stationary. This section investigates the impact of different window sizes on reliability estimates and attempt to develop a method to estimate reliability using the window sizes varied upon the change of the stationarity of data.

4.5.1. Impact of the Sliding Window Size for Reliability Estimation

Hua's method [28] adopts two parameters to control the sliding windows, window size and sliding distance. Refer to Section 4.2 for details on how these two parameters work. This section investigates the impact of these two parameters on reliability estimates. The simulation data with $RNL=0.2$ are used (see Figure 4.2) and the number of training data is 200. The strategy is to fix one of the two parameters and vary the other to visualize the impact on reliability estimation.

Figure 4.28 shows the results of varying the window size. The sliding distance is fixed to 50 for all testing cases. It can be seen that different values of window size do not affect the reliability estimates for the early stage up to around time 650; then the differences appear that the reliability curves of five window sizes are apart from each other as time passes. The reliability values with the smallest window size of 50 drop the quickest; while the ones with the largest window size of 300 drop the slowest. The reason is that in each sliding window KDE estimates the PDF based on only the data in the window; with the same sliding distance, a small window size means that the sliding window tends to contain a larger fraction of the newcomers of data points than a large window size. The newcomers possess larger/smaller values of data points due to the degradation of system conditions. This scenario does not happen to the sliding windows before time 650 because the newcomers correspond to the normal working condition the same as their predecessors. After time 650, the newcomers with higher values gradually increase their fractions in the window and have more influences on the PDF for small window than for large window. As a result, the reliability estimates of the small window drop quicker than the large window. As different window sizes give distinctive reliability estimates, its selection is important to the success of reliability estimation.

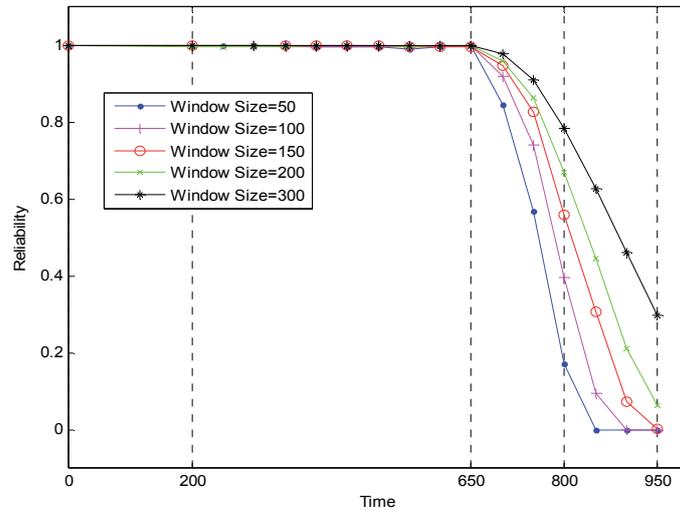


Figure 4.28 Reliability estimates for simulation data with different sliding window sizes

Next, the impact of different sliding distance on reliability estimation is investigated. Figure 4.29 shows the results of varying sliding distances. The window size is fixed to 200 for all tested cases. It is seen that all the reliability curves almost merge together except for the one with a sliding distance of 200. Due to this large sliding distance, only four sliding windows are resultant for reliability estimation which is not practical. Though it still provides the reliability estimates for each window comparable to others. We believe that the sliding distance has minor impact on reliability estimates, but its selection depends on the total number of observed data points and the users' preferences on reliability monitoring intervals.

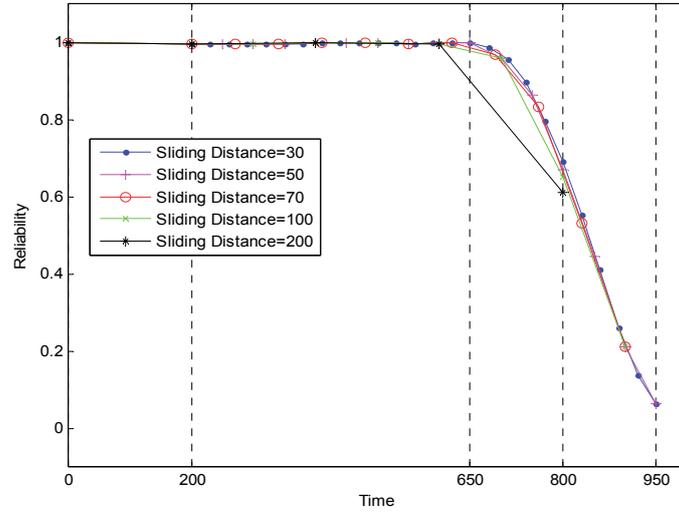


Figure 4.29 Reliability estimates for simulation data with different sliding distances

4.5.2. Variable Sliding Window Size Based Reliability Estimation

Since window sizes have non-negligible influences on reliability estimation, one should carefully select its value. To provide the sliding window with stationary data, the cumulative sum (CUSUM) technique is adopted. Stationary data do not change their mean and variance over time [98]. The CUSUM technique takes the advantage of this property by monitoring the change on mean and standard deviation of time series data [99]. Suppose that a random variable Z is drawn from a normal distribution with mean μ and standard deviation σ . CUSUM detects the deviation of a certain observation from its mean by:

$$d_i = \frac{x_i - \mu}{\sigma}, \quad (4.12)$$

where d_i represents the multiple of σ that a certain observation x_i deviates from μ . The two parameters, μ and σ , can be replaced with sample mean and sample standard deviation when a

number of samples are available. CUSUM adopts two sums to detect the unallowable deviation which are given by [99]:

$$\begin{aligned}
 UB_i &= \max(0, (d_i - m) + UB_{i-1}), \\
 LB_i &= \max(0, (-d_i - m) + LB_{i-1}),
 \end{aligned}
 \tag{4.13}$$

where UB_i detects positive deviation and LB_i detects negative deviation. The initial values of UB_0 and LB_0 are zeros. The value of m is usually selected to be 0.5 which is appropriate for detecting $1-\sigma$ deviation. There is a threshold, l , such that when it is exceeded by either sum, an unallowable deviation is detected. The l is usually selected to be 3 as suggested in [99].

Figure 4.30 illustrates the strategy of using CUSUM for reliability estimation. In the figure, $x_{i,j}$ represents the i th indicator value (data point) in the j th sliding window, t represents the time label of indicator value over the entire time span and R_{Th} represents the threshold of reliability pre-specified by users. CUSUM determines when the current sliding window ends and all the indicator values in the sliding window are used by KDE for probability density estimation. The obtained PDF of the indicator value combined with the threshold determined by OC-SVM are then utilized for estimating the reliability for the sliding window. For simplicity, the processes associated with probability density estimation using KDE and thresholding using OC-SVM are not shown in the flow chart, The two processes should be included in the highlighted box in Figure 4.30.

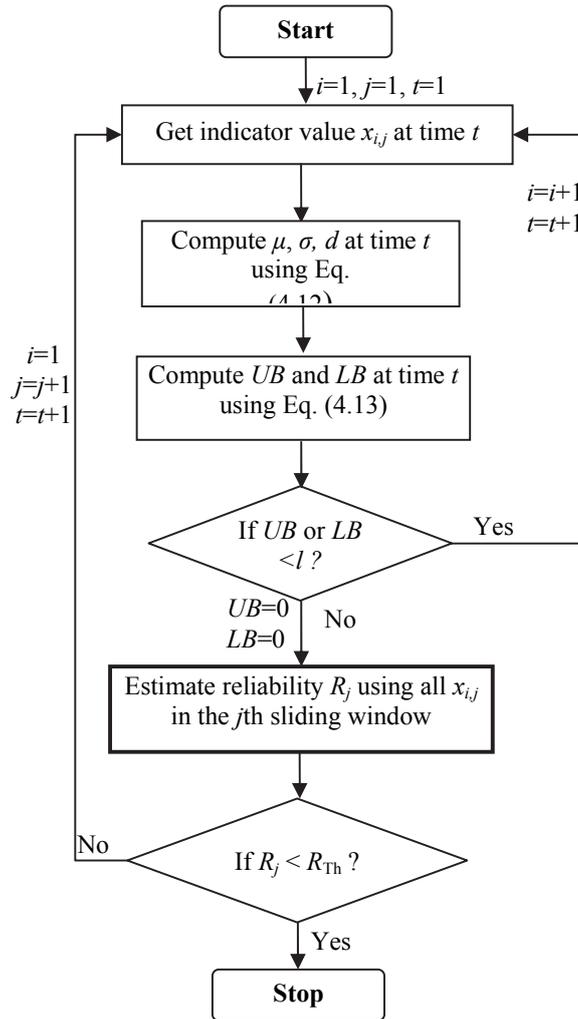


Figure 4.30 Flow chart of reliability estimation using variable sliding window size

The performance of the proposed strategy is tested with Hua's water pump data [28]. The parameters for the tests are set as reported in Hua's method. The number of training data is 200; the fixed window size is 200 and the sliding distance is 50. For comparison purposes, the window sizes of 50, 100 and 300 are also tested. The parameters of CUSUM is set to $l=3$ and $m=1.5$. The width parameter of OC-SVM is selected by the VM method per testing results obtained from Section 4.3 and the bandwidth of KDE is selected by Hua's bandwidth selection method.

Figure 4.31 shows the original water pump data with several time points labeled for illustration purposes. Figure 4.32 shows reliability estimates using three fixed window sizes and variable window sizes.

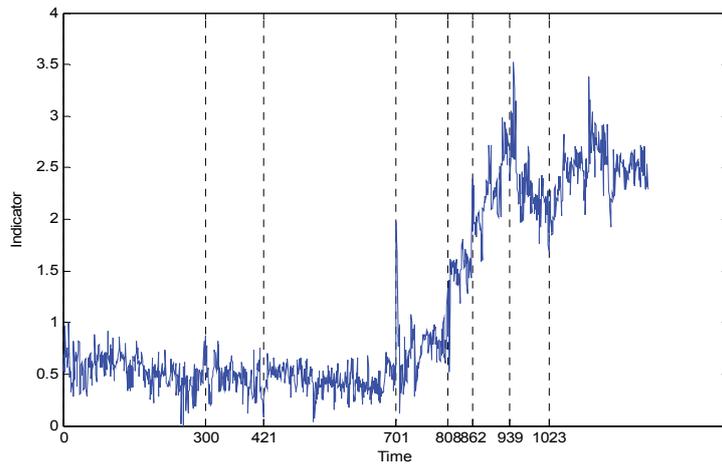


Figure 4.31 Water pump data with highlighted characteristic time points

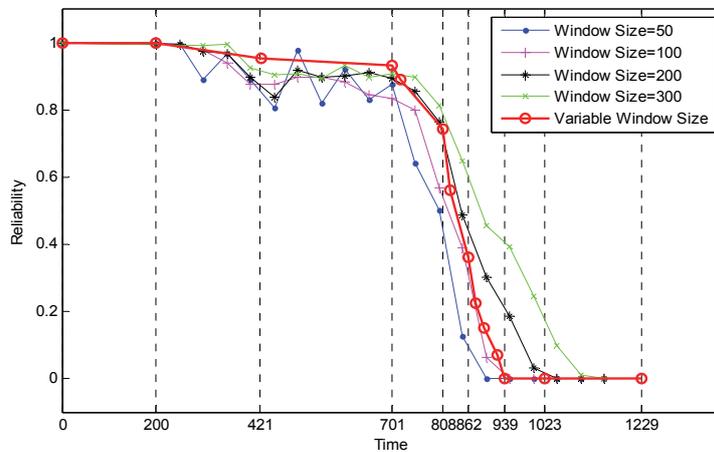


Figure 4.32 Reliability estimates for water pump with fixed and variable sliding window sizes

For the reliability curve obtained using variable window size, the first sliding window ends at time 421 and the corresponding reliability shows a little drop from 1 to 0.986. It is observed a

small increase of indicator value at this time point, but not significant. System is still in the normal condition. The second sliding window ends at time 701 when it is observed a sudden jump of indicator value. The reliability drops to 0.9252. After that the indicator values keep increasing and reliability keeps decreasing until reaches 0 at time 939. Similarly, the sliding window ends at the time points, 808, 862, 878 and 939 which all correspond to the significant increase in indicator values.

As compared to the curves of fixed window sizes, the curve obtained using variable window size is generally above the other curves up to time 701. This is preferred since the original water pump data do not show obvious value increase before this time point. After time 701, the reliability curve crosses the ones with window sizes of 200 and 300 and lies in between the ones with window sizes of 100 and 200.

Since the real reliability values are not available, it cannot be concluded that the variable window size is better than the fixed ones. Nevertheless, per the observations we believe that the method using variable window size is able to capture the degradation of the pump system and is potentially able to estimate the reliability at the interested time point when the abnormality is happening to the system. This is an advantage over the fixed window sizes which can only estimate the reliability at a fixed time interval and cannot be adjusted to capture what is happening dynamically.

4.5.3. Summary

This section investigated the impact of the sliding window size and the sliding distance for reliability estimation. It is found that the reliability estimates are quite sensitive to the selection of sliding window size, but insensitive to that of sliding distance. To provide the sliding window

with stationary data, a strategy of variable window size for reliability estimation was developed by using the cumulative sum technique. The results show that the proposed strategy can provide not only the reliability estimates comparable to the fixed sliding window size but also the reliability estimates at the time points when the abnormality occurs to the system.

4.6. Investigation on Outlier Impact for Reliability Estimation

An outlier is an observed data point that is distant from other observations [100]. It may be due to measurement errors or calculating errors [101]. Outliers can be the maximum or minimum value of the data. If the outliers are not removed, they may affect the results of reliability estimates. Hua's method removes outliers for OC-SVM but does not remove outliers for KDE. This section investigates the impact of outliers on reliability estimation. The simulation data to be used are introduced in Section 4.1.1.

4.6.1. Implementation of Outlier Removal

Hua's method [28] used the Pauta Criterion to remove outliers for OC-SVM. When applying the Pauta Criterion, the data are considered following the normal distribution. The mean, μ , and the standard deviation, σ , of the data are calculated to obtain an upper bound, $\mu+3\sigma$, and a lower bound, $\mu-3\sigma$. When a new data point is beyond these two boundaries, it is treated as an outlier and is not considered for updating the training data.

Hua's method did not conduct outlier removal for KDE. Since the value of an outlier will be way different from the values of other data points, it will potentially influence the PDF of the data in the form of unwanted long tails. As a result, the reliability estimates will be adversely affected.

To test the impact of outlier, a strategy of removing outliers from the data for KDE is developed. Figure 4.33 gives the flow chart of this proposed strategy. Each data point in a sliding window is tested using Pauta Criterion. Outliers are all removed from the sliding window and the remaining data points are used to estimate the PDF. This process is repeated for each sliding window.

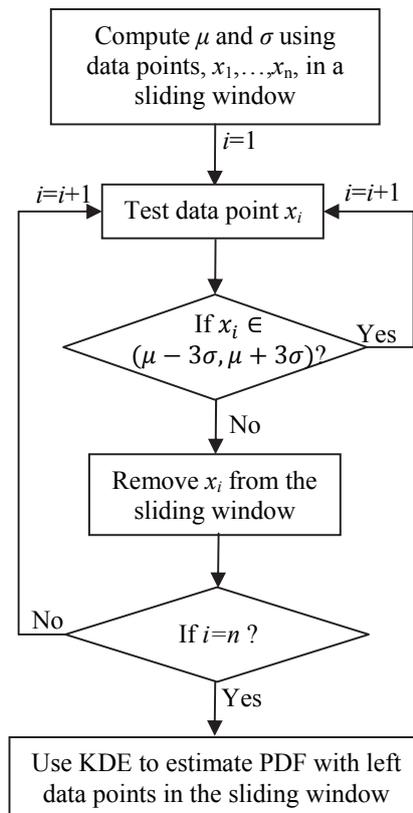


Figure 4.33 Flow chart of implementation of outlier removal for KDE

4.6.2. Investigation on Outlier Impact

In this section, the impact of outliers on reliability estimates is investigated. Two sets of simulation data which respectively contain 1 and 5 outliers are considered for investigation. The experiment data of water pump which contain 1 outlier are also used for investigation.

(1) Simulation data with 1 outlier

The simulation data with 1 outlier as introduced in Section 4.1.1 are used for testing the impact of the outliers. Figure 4.34 shows that the outlier locates at early stage at time 313 with the value of 34.56. The left panel of Figure 4.35 shows the PDF of data in each sliding window with the outlier removed. The right panel shows the PDFs without the outlier removed. It is seen that the PDFs of sliding windows 4, 5, 6 and 7 have long tails on the right hand side because of the outlier. The long tails of the PDFs cause a lower reliability estimate. Figure 4.36 shows the comparison of reliability estimates with and without outlier removed. It is seen that reliability estimates without removing the outlier have an apparent reliability drop for sliding windows 4, 5, 6 and 7. For the sliding windows that do not have the outlier, reliability estimates are not affected.

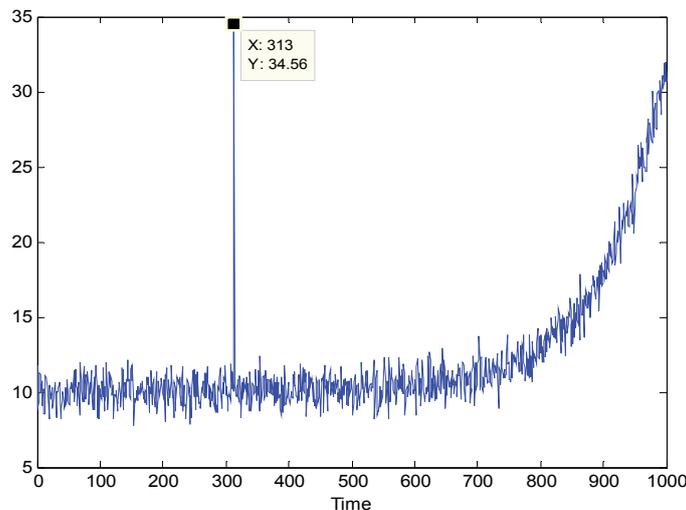


Figure 4.34 Simulation data with 1 outlier (RNL=0.2)

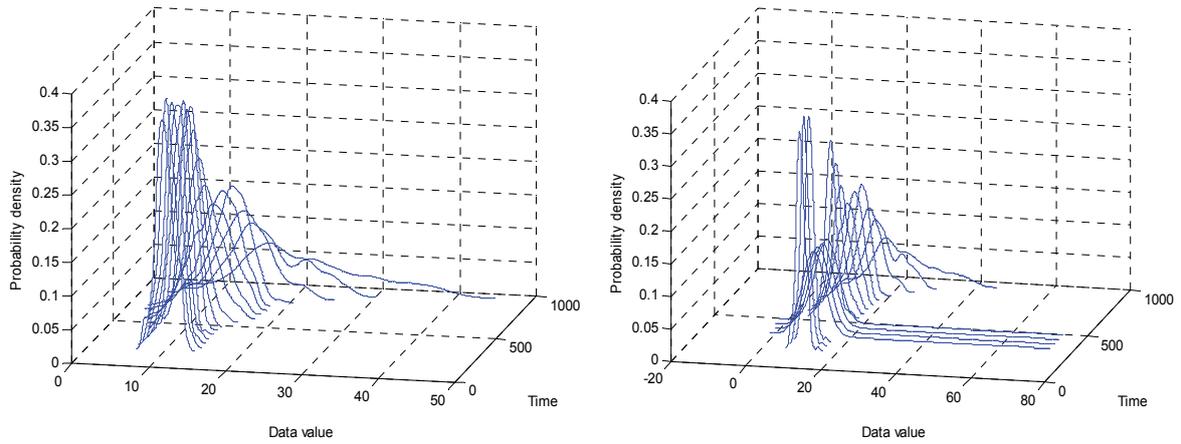


Figure 4.35 PDFs with (left) and without (right) outlier removed (1 outlier)

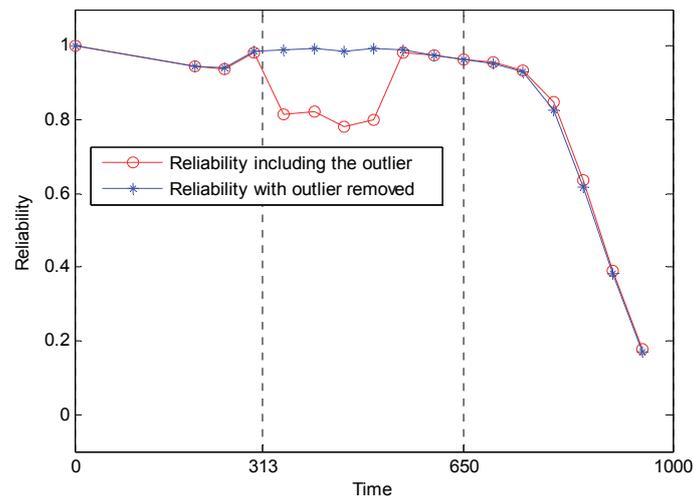


Figure 4.36 Reliability estimates with and without outlier removed (1 outlier)

(2) Simulation data with 5 outliers

The simulation data with 5 outliers as introduced in Section 4.1.1 are also used for testing the impact of the outliers. Figure 4.37 shows the plot of the simulation data with 5 outliers. Three of them locate at early stage with large amplitudes and two of them locate at the late stage with small amplitude at time 668 and large amplitude at time 773. Figure 4.38 shows the PDFs with

and without removing the 5 outliers. For the outliers located at the early stage, reliability estimates are similar to the case with 1 outlier. For the outliers at the late stage, the one with small amplitude is not detected because the values are comparable to the neighbor data points. Figure 4.39 shows the comparison of reliability estimates with and without the outliers removed.

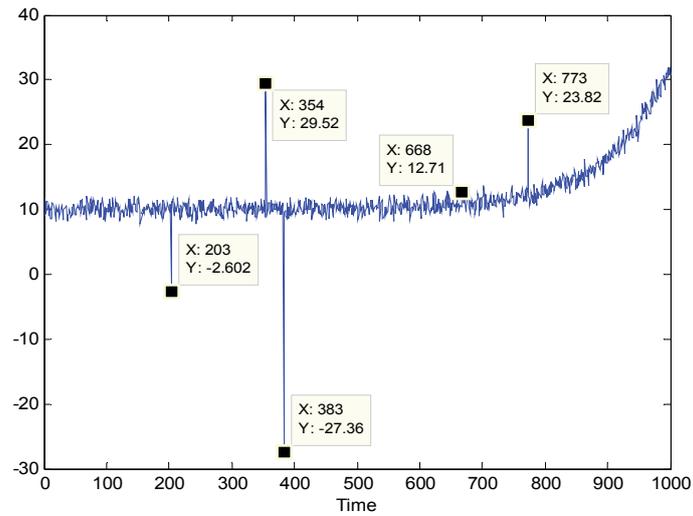


Figure 4.37 Simulation data with 5 outliers (RNL=0.2)

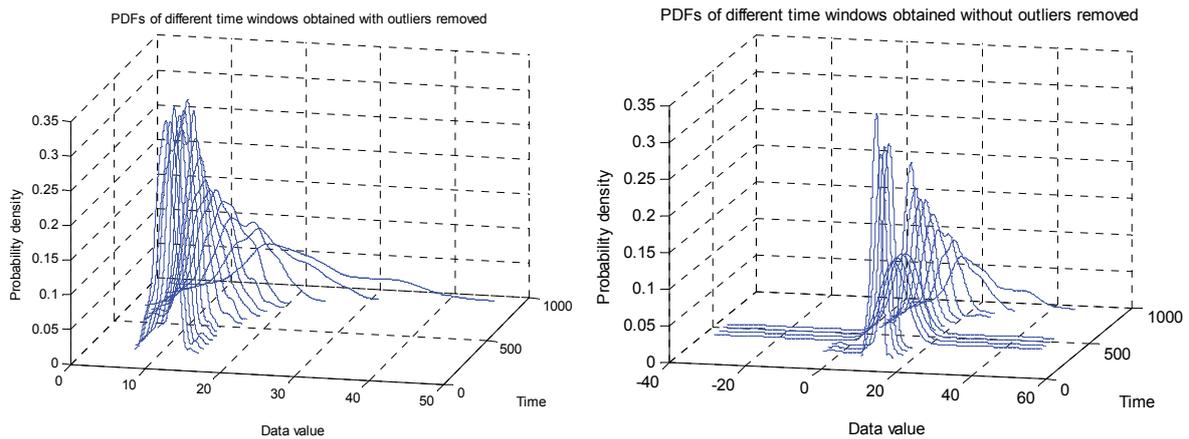


Figure 4.38 The PDFs with (left) and without (right) outlier removed (5 outliers)

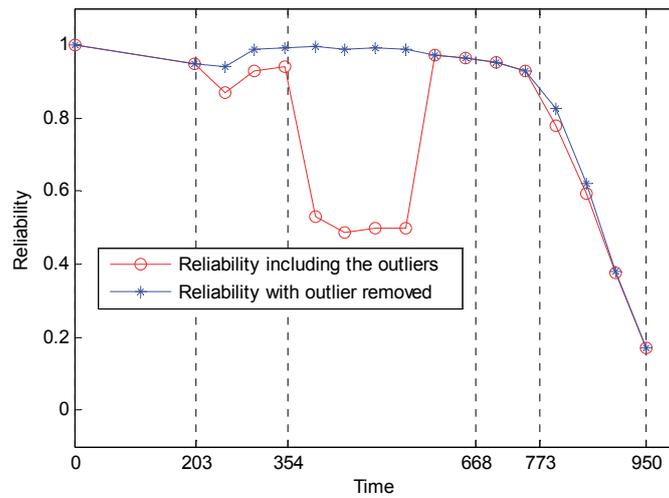


Figure 4.39 Reliability estimates with and without outlier removed (5 outliers)

(3) Water pump data with 1 outlier

The water pump data [28] with 1 outlier as introduced in Section 4.1 are used for testing the impact of outlier. Figure 4.40 shows that 1 outlier is randomly added to the water pump data at time 477. The data point at time 701 also has a large value which exists in the original data.

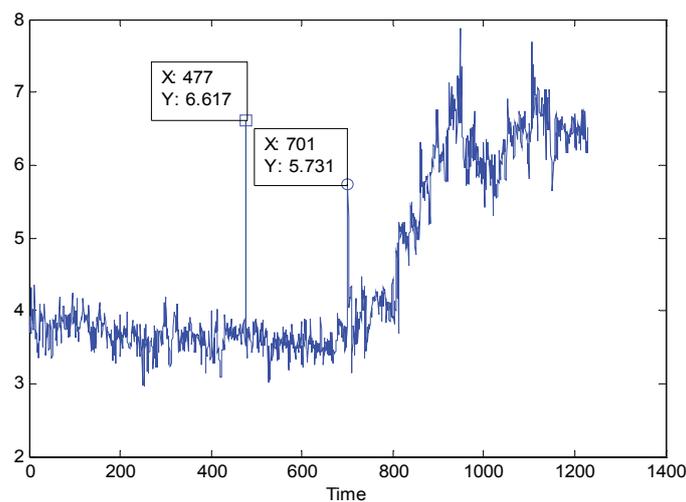


Figure 4.40 Water pump data with 1 additive outlier

The reliability with and without removing the outlier are estimated and compared in Figure 4.41. The number of training data is 300. The window size is 100 and the sliding distance is 50. It is found that Pauta Criterion removed the outlier at time 477 hours and also the non-outlier data point at time 701 hours. The obtained reliability estimates are able to reflect the degradation pattern of the data. For the case without removing the outlier, reliability estimates drop around times 477 and 701 hours. For the data point at time 701, it is close to the late stage, so it may not be an outlier but a sign of fault for the system. Further study may be required to determine whether a detected data point at the late stage is an outlier.

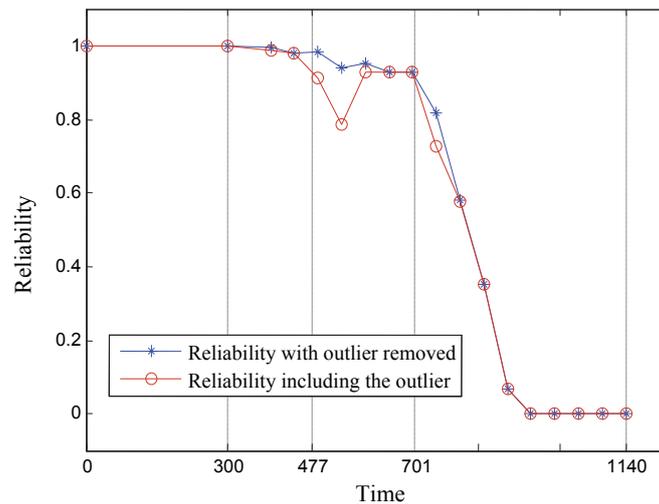


Figure 4.41 Reliability estimates for water pump with and without outliers removed

4.6.3. Summary

Based on the tests using simulation data and experiment data, the impact of outliers on reliability estimation is investigated. It is found that an outlier can cause a long tail of the PDF and leads to unexpected decrease in reliability estimates. Therefore, removing outliers from the data for KDE is a necessary step to ensure the success of reliability estimation. One thing to be noted is that the

data points with large values at the late stage may not be outliers. Such data points may be a sign of fault. Further study needs to be conducted to address this problem.

4.7. Comparisons with Hua’s Method

With the observations obtained from previous sections, this section compares the performance of using the parameters reported in [28] and the parameters recommended by this thesis. In [28], the water pump data are used for testing Hua’s method and the parameters adopted are as follows. The number of training data is 200, the fixed window size is 200, the sliding distance is 30, and the split number, N_{split} , is 300. The selection of width parameter, σ , is not provided in [28]. We choose $\sigma = 0.004$ which is obtained by VM method as presented in Table 4.3. The value of bandwidth, h , is given as 0.054 in [28]. It is seen that the parameters for Hua’s method are in compliance with our suggestions where σ lies in the applicable range of [0.002, 0.01], h lies in the applicable range of [0.01, 0.1] and N_{split} is equal to 300 which is more than enough per our suggestions. Based on these selected parameters, Hua’s method is expected to provide reasonable reliability estimates.

Table 4.7 Parameter settings for comparisons

Method	σ	h	N_{split}	CPU time (s)	# of Windows
Hua’s method	0.004	0.054	300	2.26	35
Option 1	0.004	0.1	30	0.3120	35
Option 2	0.004	0.01	30	0.4368	35
Option 3	0.004	0.001	330	1.9188	35
Variable Window Size	0.004	0.054	300	0.5616	12

To conduct a fair comparison, all the parameters are selected the same as those reported in [28] except for the parameters of σ , h , and N_{split} which are selected based on our observations. Table 4.7 shows the three selected options of the parameters. The σ is chosen as 0.004 for all three options the same as that for Hua's method. The h is chosen based on Eq. (4.11) where the boundary values of 0.01 and 0.1 are considered. Another small h value of 0.001 is also tested for interest. The N_{split} is chosen according to the selected h values based on Eq. (4.11). The method of variable window size is also included for the comparisons, wherein parameters of σ , h , and N_{split} are selected exactly the same as those for Hua's method.

Figure 4.42 shows the results of reliability estimates. It is seen that the methods of Option 2, Option 3 and variable window size provide reasonable reliability estimates satisfying the evaluation criteria. Hua's method also provides reasonable reliability estimates but with small drops at the early stage. Compared to Hua's method, Option 2 provides better results and consumes less CPU time as listed in Table 4.7. This is due to the smaller N_{split} selected based on our observations. The method of variable window size provides smoothest reliability estimates at the early stage as compared to all its counterparts and consumes smaller CPU time than Hua's method due to less number of sliding windows being required as listed in Table 4.7.

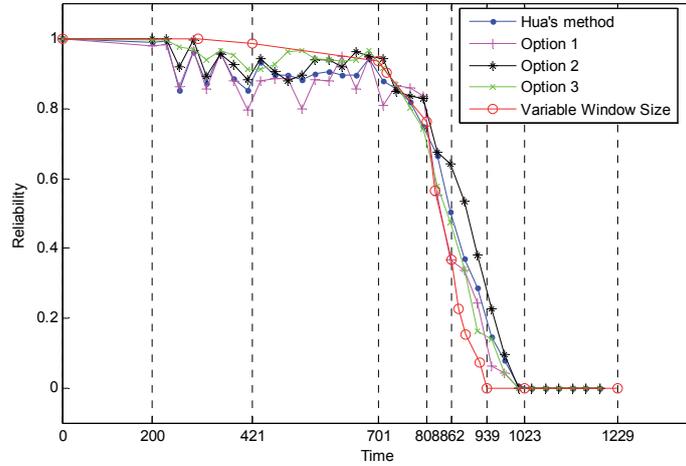


Figure 4.42 Comparisons of reliability estimates using different methods

Option 1 takes $h=0.1$ which is the boundary value of h based on our observations. It is seen from Figure 4.42 that the method of Option 1 provides reliability estimates with larger drops than Hua's method at the early stage. Users may need to determine its acceptability upon individual cases. Therefore, it is recommended to choose $h=0.01$ with $N_{\text{split}}=30$ since as compared to $h=0.1$ it requires comparable CPU time but provides better results of reliability estimates.

4.8. Summary

This chapter investigates four aspects of reliability estimation using OC-SVM for thresholding and KDE for probability density estimation based on condition monitoring data. The first aspect is the impact of the width parameter, σ , of OC-SVM on reliability estimation. An applicable range of $[0.002, 0.01]$ is found for the σ , which can provide reasonable reliability estimates for the data with various noise effects. The second aspect is the impact of the bandwidth, h , of KDE on reliability estimation. An applicable range of $[0.01, 0.1]$ for h and a formula for the number of splits are found able to provide reasonable reliability estimates for the data with various noise

effects. The third aspect is the impact of the sliding window size and the sliding distance of KDE on reliability estimation. A method is developed to use variable window sizes for reliability estimation. The method is found able to not only provide applicable reliability estimates but also compute the reliability estimates for the sliding window corresponding to the transition of system conditions. The fourth aspect is the impact of outliers for reliability estimation. It is confirmed that outliers can adversely affect the outcomes of reliability estimation and need to be removed before the implementation of thresholding and probability density estimation.

Chapter 5

Summary and Future Work

5.1. Summary

System reliability estimation using condition monitoring data includes two important parts: thresholding and probability density estimation. This thesis studied a reliability estimation method which uses one-class support vector machine (OC-SVM) for thresholding and kernel density estimation (KDE) for probability density estimation. Four aspects were investigated and the work of this thesis is summarized as follows:

- (1) The impact of the width parameter, σ , of OC-SVM on reliability estimation was investigated by using the enumeration method. Simulation data with different noise effects were used for investigation. An applicable range of [0.002, 0.01] for σ was found able to provide reasonable reliability estimates and reflect the degradation pattern of simulation data. For $\sigma > 0.01$, the reliability estimates tend to show a flat pattern at late stage, and for $\sigma < 0.002$, the reliability estimates tend to drop at early stage. Neither agrees with the degradation pattern of simulation data nor satisfies the Evaluation Criterion. The effects of different noise levels were investigated and the results showed that de-noising method is needed for the data with relative noise level (RNL) greater than 0.3. Three sets of experiment data were tested using the σ values obtained with three reported methods. Only one method provided σ values within the range recommended by this thesis and yielded reasonable reliability estimates. It is concluded that the range of [0.002, 0.01] may be a good option for selecting the width

parameter, σ , of OC-SVM for reliability estimation. One can select a σ value within the range as the first step when using the reported method for reliability estimation.

(2) The impact of the bandwidth, h , of KDE on reliability estimation was investigated by the enumeration method. The data used are the same as those for the cases of width parameter selection. A range of $[0.01, 0.1]$ for h with $N_{\text{split}}=30$ are recommended for KDE to provide reasonable reliability estimates. For $h>0.1$, the reliability estimates tend to drop at early stage and do not satisfy the Evaluation Criteria. For $h<0.01$, it requires a large value of N_{split} which is computationally inefficient. The effects of different noise levels were investigated and the results show that de-noising method is needed for the data with $\text{RNL}>0.3$. Three sets of experiment data were tested using h values obtained from four reported methods. Three methods provided h values within the range recommended by this thesis and yielded reasonable reliability estimates; while one provided the values of h greater than 0.1 and failed to provide reliability estimates satisfying the Evaluation Criteria. It is concluded that the range of $[0.01, 0.1]$ may be a good option for selecting the bandwidth, h , of KDE for reliability estimation with a recommended split number equal to 30. One can select an h value within the range and the recommended split number as the first step when using the reported method for reliability estimation.

(3) The impact of sliding window size and sliding distance for KDE on reliability estimation was investigated. The results showed that the sliding distance of window has minor effects on reliability estimation. A smaller sliding distance provides a larger number of reliability estimates than a larger sliding distance does. Different sliding window sizes do not affect the performance of reliability estimates at the early stage, but they yield different reliability estimates at the late stage. Since the true values of the reliability estimates are inaccessible, it

is difficult to choose an optimal size of time window. A strategy with variable window size was developed and compared with the one with fixed window sizes. Water pump data were used to test the performance of the variable window size. The results showed that as compared to fixed window sizes, the variable window size can provide more stable reliability estimates close to 1 at the early stage and can capture the transition of system condition at the time points when abnormality occurs.

- (4) The impact of outliers on reliability estimation was tested based on simulation data and experiment data. The reliability estimates and the probability density functions (PDFs) of each sliding window with and without removing the outliers were compared. It is found that the outliers may cause the PDFs of the data having long tails which results in unexpected decreases in reliability estimates for sliding windows containing the outliers. A strategy of outlier removal for KDE is developed based on Pauta Criterion. The results showed that removing outliers for KDE was a necessary step to ensure the success of reliability estimation. It is also observed that the data points with large values at the late stage may not be outliers but a sign of fault. The method for identifying true outliers for the late stage needs to be developed in future.

5.2. Future Work

The objectives of this thesis are to investigate four aspects of a reported method in order to simplify the use and enhance the performance of the reported method. However, in the study several other issues are found and need to be addressed in the future:

- (1) Data with large noise effects are not appropriate to be used in reliability estimation. Such data will result in a flat trend of reliability estimates which are unable to reflect the

degradation pattern of system health condition. In order to reduce the effects of noise, proper de-noising methods should be developed and used in the future.

- (2) Outliers need to be removed for reliability estimation. However, unusual data values appearing at the late stage may not be necessarily the true outliers. They could also be the sign of faults. Further studies need to be conducted to identify whether a data point with its value different from its neighbors is a true outlier or not.
- (3) This thesis studied the reliability estimation using a one dimensional time series of a health indicator. In practice, one health indicator may not be able to fully reflect the degradation of the system of interest, so multiple health indicators may be required for reliability estimation. Both OC-SVM and KDE are able to deal with multidimensional data, so the reported method can be extended to accommodate the case with multiple health indicators. As more information of system health is available from multiple indicators, more useful results are expected. This will be a study topic in the future.

Bibliography

- [1] Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.
- [2] Chinnam, R. B., & Baruah, P. (2004). A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems. *International Journal of Materials and Product Technology*, 20(1-3), 166-179.
- [3] Zhang, S., Ma, L., Sun, Y., & Mathew, J. (2007). Asset health reliability estimation based on condition data. *Proceedings World Congress on Engineering Asset Management*, Harrogate, England, pp. 2195-2204.
- [4] Heng, A., Zhang, S., Tan, A. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, Challenges and opportunities. *Mechanical Systems and Signal Processing*, 23(3), 724-739.
- [5] Ebrahimi, N. (2005). System reliability based on diffusion models for fatigue crack growth. *Naval Research Logistics*, 52(1), 46-57.
- [6] Dolas, D. R., Jaybhaye, M. D., & Deshmukh, S. D. (2014). Estimation the system reliability using Weibull distribution. *International Proceedings of Economics Development and Research*, 75(29), 144-148.
- [7] Montgomery, N., Lindquist, T., Garnero, M., Chevalier, R., & Jardine, A. K. S. (2006). Reliability functions and optimal decisions using condition data for EDF primary pumps. *Probabilistic Methods Applied to Power Systems, International Conference on IEEE*, pp. 1-6.

- [8] Kuo, W., & Zuo, M. J. (2003). *Optimal Reliability Modeling: Principles and Applications*. John Wiley & Sons, New York.
- [9] Ribrant, J. (2006). Reliability performance and maintenance - A survey of failures in wind power systems. *Master Thesis written at KTH School of Electrical Engineering*.
- [10] Son, J., Zhou, Q., Zhou, S., Mao, X., & Salman, M. (2013). Evaluation and comparison of mixed effects model based prognosis for hard failure. *IEEE Transactions on Reliability*, 62(2), 379-394.
- [11] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. 2nd ed., vol. 362. Hoboken, N. J.: Wiley-Interscience.
- [12] Kapur, K. C., & Lamberson, L. R. (1977). *Reliability in Engineering Design*. New York: Wiley.
- [13] Bazovsky, I. (1961). *Reliability Theory and Practice*. Englewood Cliffs, N. J.: Prentice-Hall.
- [14] Keller, A. Z., Perera, U. D., & Kamath, A. R. R. (1982). Reliability analysis of CNC machine tools. *Reliability Engineering*, 3(6), 449–473.
- [15] Crowder, M. J., Kimber, A., Sweeting, T., & Smith, R. (1994). *Statistical Analysis of Reliability Data*. vol. 27. London: Chapman & Hall/CRC.
- [16] Schömig, A. K., & Rose, O. (2003). On the suitability of the Weibull distribution for the approximation of machine failures. *Proceedings of the Industrial Engineering Research Conference*, Portland, OR. pp. 1.
- [17] Elsayed, E. A. (1996). *Reliability Engineering*. Englewood Cliffs, N. J.: Prentice-Hall.
- [18] Xu, X., Yu, Z., Gao, Y., & Wang, T. (2008). Crack failure of gears used in generating electricity equipment by wind power. *Engineering Failure Analysis*, 15(7), 938-945.

- [19] Qu, J., Liu, Z., Zuo, M. J., & Huang, H. (2011). Feature selection for damage degree classification of planetary gearboxes using support vector machine. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 225(9), 2250-2264.
- [20] Zuo, M. J., Li, W., & Fan, X. F. (2005). *Statistical methods for low speed planetary gearbox monitoring*. Technical Report, Department of Mechanical Engineering, University of Alberta, Edmonton.
- [21] Samuel, P. D., Pines, D. J., & Lewicki, D. G. (2000). A comparison of stationary and non-stationary metrics for detecting faults in helicopter gearboxes. *Journal of the American Helicopter Society*, 45, 125-136.
- [22] Pöyhönen, S., Jover, P., & Hyötyniemi, H. (2004). Signal processing of vibrations for condition monitoring of an induction motor. *Control, Communications and Signal Processing*, First International Symposium on. IEEE, pp. 499-502.
- [23] Qu, J. (2012). *Diagnostics and Prognostics: SVM-Based Diagnostics and Prognostics for Rotating Systems*. LAP LAMBERT Academic Publishing.
- [24] Večeř, P., Kreidl, M., & Šmíd, R. (2005). Condition indicators for gearbox condition monitoring systems. *Acta Polytechnica*, 45(6), 35-43.
- [25] Si, X., Wang, W., Hu, C., & Zhou, D. (2011). Remaining useful life estimation-A review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1), 1-14.
- [26] Gebraeel, N. Z., Lawley, M. A., Li, R., & Ryan, J. K. (2005). Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, 37(6), 543-557.

- [27] Li, L., You, M., & Ni, J. (2009). Reliability-based dynamic maintenance threshold for failure prevention of continuously monitored degrading systems. *Journal of Manufacturing Science and Engineering*, 131(3), 031010.1-031010.9.
- [28] Hua, C., Zhang, Q., Xu, G., Zhang, Y., & Xu, T. (2013). Performance reliability estimation method based on adaptive failure threshold. *Mechanical Systems and Signal Processing*, 36(2), 505-519.
- [29] Ebrahim, M. A., Eidous, O., & Kmail, G. (2009). Estimating percentiles of time-to-failure distribution obtained from a linear degradation model using kernel density method. *Communications in Statistics: Simulation and Computation*, 38(9), 1811-1822.
- [30] Ferracuti, F., Giantomassi, A., Iarlori, S., Ippoliti, G., & Longhi, S. (2015). Electric motor defects diagnosis based on kernel density estimation and Kullback–Leibler divergence in quality control scenario. *Engineering Applications of Artificial Intelligence*, 44, 25-32.
- [31] Cho, H. C., Fadali, M. S., & Lee, K. S. (2008). Online probability density estimation of nonstationary random signal using dynamic Bayesian networks. *International Journal of Control, Automation, and Systems*, 6(1), 109-118.
- [32] Joshi, N., Kadir, T., & Brady, M. (2011). Simplified computation for nonparametric windows method of probability density function estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8), 1673-1680.
- [33] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. CRC press.
- [34] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualisation*. New York: Wiley.

- [35] Fahmy, S. A., & Mohan, A. R. (2013). Architecture for real-time nonparametric probability density function estimation. *IEEE Transactions on Very Large Scale Integration Systems*, 21(5), 910-920.
- [36] Ding, F., & He, Z. (2011). Cutting tool wear monitoring for reliability analysis using proportional hazards model. *International Journal of Advanced Manufacturing Technology*, 57(5-8), 565-574.
- [37] Gao, J., & Wang, H. (2015). Operation reliability analysis based on fuzzy support vector machine for aircraft engines. *Journal of Maintenance and Reliability*, 17(2), 186-193.
- [38] Cai, G., Chen, X., Li, B., Chen, B., & He, Z. (2012). Operation reliability assessment for cutting tools by applying a proportional covariate model to condition monitoring information. *Sensors*, 12(10), 12964-12987.
- [39] Chen, W., Liu, J., Gao, L., Pan, J., & Zhou, S. (2011). Accelerated degradation reliability modeling and test data statistical analysis of aerospace electrical connector. *Chinese Journal of Mechanical Engineering*, 24(6), 957-962.
- [40] Chen, B., Chen, X., Li, B., He, Z., Cao, H., & Cai, G. (2011). Reliability estimation for cutting tools based on logistic regression model using vibration signals. *Mechanical Systems and Signal Processing*, 25(7), 2526-2537.
- [41] Feng, F., Xing, W., Si, A., & Wu, G. (2011). Reliability assessment for a certain diesel engine based on performance degradation data. *Advanced Materials Research*, 199, 481-486.
- [42] Hua, C., Zhang, Q., Xu, G., & Xie, J. (2010). Real-time performance reliability assessment method based on dynamic probability model. *Lecture Notes in Computer*

- Science*, 6319, 88-96. *British Library Document Supply Centre inside Serials and Conference Proceedings*.
- [43] Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., & Williamson, R. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), 1443–1471.
- [44] Xiao, Y., Wang, H., Zhang, L., & Xu, W. (2014). Two methods of selecting Gaussian kernel parameters for one-class SVM and their application to fault detection. *Knowledge-Based Systems*, 59, 75–84.
- [45] Xu, Z., Dai, M., & Meng, D. (2009). Fast and efficient strategies for model selection of Gaussian support vector machine. *IEEE Transactions on Systems, Man, and Cybernetics: Part B*: 39(5), 1292–1307.
- [46] Evangelista, P. F., Embrechts, M. J., & Szymanski, B. K. (2007). Some properties of the Gaussian kernel for one class learning. *Lecture Notes in Computer Science*, 4668, 269–278.
- [47] Khazai, S., Homayouni, S., Safari, A., & Mojaradi, B. (2011). Anomaly detection in hyperspectral images based on an adaptive support vector method. *IEEE Geoscience and Remote Sensing Letters*, 8(4), 646–650.
- [48] Heidenreich, N., Schindler, A., & Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *Asta: Advances in Statistical Analysis*, 97(4), 403–433.
- [49] Guidoum, A. C. (2015). *Kernel Estimator and Bandwidth Selection for Density and its Derivatives*. University of Science and Technology, Houari Boumediene, BP 32 El-Alia, U.S.T.H.B, Algeria. <https://cran.r-project.org/web/packages/kedd/vignettes/kedd.pdf>.

- [50] Glendinning, R. H. (2001). Selecting sub-set autoregressions from outlier contaminated data. *Computational Statistics and Data Analysis*, 36(2), 179–207.
- [51] Shieh, A. D., & Hung, Y. S. (2009). Detecting Outlier Samples in Microarray Data. *Statistical Applications in Genetics and Molecular Biology*, 8(1), 1-24.
- [52] Yu, I. T., & Fuh, C. D. (2010). Estimation of time to hard failure distributions using a three-stage method. *IEEE Transactions on Reliability*, 59(2), 405–412.
- [53] Sick, B. (2002). On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. *Mechanical System and Signal Processing*, 16(4), 487-546.
- [54] ISO 3685:1993. *Tool-Life Testing with Single-Point Turning Tools*.
- [55] Astakhov, V. P., & Davim, J. P. (2008). Tools (geometry and material) and tool wear. *Machining*, Springer London, pp. 29-57.
- [56] Wang, P., & Coit, D. W. (2007). Reliability and degradation modeling with random or uncertain failure threshold. *Reliability and Maintainability Symposium*, pp.392-397.
- [57] Lu, C. J., & Meeker, W. Q. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2), 161–174.
- [58] Nystad, B. H., Gola, G., & Hulsund, J. E. (2012). Lifetime models for remaining useful life estimation with randomly distributed failure thresholds. *European Conference of Prognostics and Health Management Society*, vol. 3.
- [59] Xiang, Y., Cassady, C. R., & Pohl, E. A. (2012). Optimal maintenance policies for systems subject to a Markovian operating environment. *Computers and Industrial Engineering*, 62(1), 190-197.

- [60] Wang, D., & Miao, Q. (2010). On-line automatic early fault detection of rotating machinery. *Prognostics and System Health Management Conference*, pp. 1-5.
- [61] Guo, P., & Bai, N. (2011). Wind turbine gearbox condition monitoring with AAKR and moving window statistic methods. *Energies*, 4(11), 2077-2093.
- [62] Fernández-Francos, D., Martínez-Rego, D., Fontenla-Romero, O., & Alonso-Betanzos, A. (2013). Automatic bearing fault diagnosis based on one-class v-SVM. *Computers and Industrial Engineering*, 64(1), 357-365.
- [63] Hu, L., Hu, N., Zhang, X., Gu, F., & Gao, M. (2013). Novelty detection methods for online health monitoring and post data analysis of turbo pumps. *Journal of Mechanical Science and Technology*, 27(7), 1933-1942.
- [64] Zhu, P., & Hu, Q. (2013). Rule extraction from support vector machines based on consistent region covering reduction. *Knowledge-based systems*, 42, 1-8.
- [65] Zambom, A. Z. & Dias, R. (2013). A review of Kernel density estimation with applications to econometrics. *International Econometric Review*, 5(1), 20-42.
- [66] Raykar, V. C. (2002). Probability density function estimation by different methods. *ENEE 739Q Spring*, pp. 1-8.
- [67] Varanasi, M. K., & Aazhang, B. (1989). Parametric generalized Gaussian density estimation. *Journal of the Acoustical Society of America*, 86(4), 1404-1415.
- [68] Arenberg, J. W., & Thomas, M. D. (2014). Sensitivity test method for the characterization of laser damage behavior. *Optical Engineering*, 53(12), 122517.1-122517.6.
- [69] Izenman, A. J. (1991). Review papers: recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413), 205-224.

- [70] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification*. 2nd ed. New York: Wiley.
- [71] Wang, J., Wang, Z., Yu, X., Yao, M., Yao, Z., & Zhang, E. (2012). Establishment method of a mixture model and its practical application for transmission gears in an engineering vehicle. *Chinese Journal of Mechanical Engineering*, 25(5), 1001-1010.
- [72] Mazzuchi, T. A., & Dorp, J. R. (2012). A Bayesian expert judgement model to determine lifetime distributions for maintenance optimisation. *Structure and Infrastructure Engineering*, 8(4), 307-315.
- [73] Buchta, J. & Oziemski, A. (2007). Reliability of large power units in probabilistic approach. *Electrical Power Quality and Utilisation, 9th International Conference*, pp.1-6.
- [74] Cheng, M. Y., & Hoang, N. D. (2014). Slope collapse prediction using Bayesian framework with k-nearest neighbor density estimation: case study in Taiwan. *Journal of Computing in Civil Engineering*, 30(1), 04014116.1-04014116.8.
- [75] Zamini, R., Fakoor, V., & Sarmad, M. (2014). Asymptotic behaviors of nearest neighbor kernel density estimator in left-truncated data. *Journal of Sciences, Islamic Republic of Iran*, 25(1), 56-67.
- [76] Kung, Y. H., Lin, P. S., & Kao, C. H. (2012). An optimal k-nearest neighbor for density estimation. *Statistics and Probability Letters*, 82(10), 1786-1791.
- [77] Lee, G., & Scott, C. D. (2007). The one class support vector machine solution path. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2, 521-524.
- [78] Ahmad, I. A., & Ran, I. S. (2004). Data based bandwidth selection in kernel density estimation with parametric start via kernel contrasts. *Journal of Nonparametric Statistics*, 16(6), 841-877.

- [79] Köhler, M., Schindler, A., & Sperlich, S. (2014). A review and comparison of bandwidth selection methods for kernel regression. *International Statistical Review*, 82(2), 243-274.
- [80] Ruan, D., Chen, G., Kerre, E., & Wets, G. (Eds.). (2005). *Intelligent Data Mining: Techniques and Applications. Studies in Computational Intelligence*. Berlin, Springer, vol. 5.
- [81] Ruppert, D., Sheather, S. J., & Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 432, 1257–1270.
- [82] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2), 353–360.
- [83] Qu, J., & Zuo, M. J. (2010). An LSSVR-based machine condition prognostics algorithm for slurry pump systems. *Proceedings of the Canadian Society for Mechanical Engineering Forum 2010*, Victoria, British Columbia, Canada.
- [84] Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks: The Official Journal of the International Neural Network Society*, 17(1), 113–126.
- [85] Lei, Y., Hoseini, M., Qu, J., Zhou, X., Pandey, M., & Zuo, M. J. (2009). *Planetary gearbox test rig data collection and analysis report for the repeatability test*. Technical report, Reliability Research Lab, Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta.
- [86] Prognostics Data Repository - Intelligent Systems Division - NASA. http://ti.arc.nasa.gov/projects/data_prognostics.

- [87] Qiu, H., Lee, J., & Lin, J. (2006). Wavelet filter-based weak signature detection method and its application on roller bearing prognostics. *Journal of Sound and Vibration*, 289(4), 1066-1090.
- [88] Lebold, M., McClintic, K., Campbell, R., Byington, C., & Maynard, K. (2000). Review of vibration analysis methods for gearbox diagnostics and prognostics. *Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology*, Virginia Beach, VA, pp. 623-634.
- [89] Muller-Merbach, H. (1974). Modelling techniques and heuristics for combinatorial problems. *Combinatorial Programming: Methods and Applications*, Springer Netherlands, pp. 3-27.
- [90] Stute, W. (1992). Modified cross validation in density estimation. *Journal of Statistical Planning and Inference*, 30(3), 293–305.
- [91] Martinez-Miranda, M. D., Nielsen, J. P., & Sperlich, S. A. (2009). One sided cross validation for density estimation with an application to operational risk. *Operational Risk toward Basel III: Best Practices and Issues in Modeling, Management and Regulation*, Wiley, Finance Series, pp. 177-195.
- [92] Savchuk, O. Y., Hart, J. D., & Sheather, S. J. (2010). Indirect cross-validation for density estimation. *Journal of the American Statistical Association*, 105(489), 415–423.
- [93] Park, B. U., & Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409), 66–72.
- [94] Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B. Methodological*, 53(3), 683–690.

- [95] Hardle, W., Muller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Berlin: Springer Series in Statistics.
- [96] Hall, P. (1990). Using the bootstrap to estimate mean square error and select smoothing parameters in nonparametric problems. *Journal of Multivariate Analysis*, 32(2), 177–203.
- [97] Mammen, E., Martínez-Miranda, M. D., Nielsen, J. P., & Sperlich, S. (2011). Do-validation for kernel density estimation. *Journal of the American Statistical Association*, 106(494), 651–660.
- [98] Challis, R. E., & Kitney, R. I. (1991). Biomedical signal processing (in four parts) - Part 1 time-domain methods. *Medical and Biological Engineering and Computing*, 28(6), 509-524.
- [99] Ryan, T. P. (1989). *Statistical Methods for Quality Improvement*. New York, Wiley.
- [100] Maddala, G. S. (1992). *Outliers, Introduction to Econometrics*. New York: MacMillan. 2, 88-96.
- [101] Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11(1), 1-21.