

# A Probability Box Modeling Method of Dimensionless Mechanical Fault Features in Time Domain

Du Yi<sup>1</sup>

City College  
Kunming University of Science and Technology  
Kunming, China  
e-mail: duy215@hotmail.com

Liu An<sup>3</sup>

Faculty of Mechanical and Electrical Engineering  
Kunming University of Science and Technology  
Kunming, China  
e-mail: liuanmike@sina.com

Ding Jiaman<sup>2\*</sup>

College of Information Engineering and Automation  
Kunming University of Science and Technology  
Kunming, China  
e-mail: tjom2008@163.com

**Abstract**—The simply multi-segment averages of mechanical fault features lead to the loss of signal uncertainty probability statistics which may contribute to fault diagnosis. To solve the problem, a probability box modeling method of mechanical fault features is proposed. Three similar dimensionless factors (peak factor, form factor, margin factor) in time domain were used to obtain probability boxes. The algorithm of modeling was then proposed. Firstly, the mechanical fault raw data was divided into several segments according to the sampling frequency. Secondly, the dimensionless features in time domain were obtained for every segment. After verifying the normal distribution probability distribution type, the row vector of features was divided into several sets. From these sets, the range of the mean and variance were obtained. Finally, the DSS of mean and variance were integrated into the probability modeling algorithm of normal distribution to get the probability box. The rolling bearing's faults raw data was used to verify this modeling method. The results show that the probability box modeling method can effectively reflect the uncertainties of fault state, which can be used for fault diagnosis.

**Keywords**- Probability box theory; modeling; Uncertainty; Mechanical Fault; fault feature

## I. INTRODUCTION

The treatment of fault signals at the same time in mechanical fault diagnosis is always a simple multi-segment average, which leads to the loss of signal uncertainty probability statistics. The uncertainties of mechanical fault signals are always omitted by many researchers or considered as harmful information for fault diagnosis. However, this is a prejudice of uncertainties. Many mechanical fault signs may reveal by signal's uncertainties. The probability box theory has great advantages in the expressing of uncertainties. It can

provide a new way to solve the above problem. It can envelope the time uncertainty and sampling uncertainty with its upper and down bounds. It can express almost all types of uncertainty, including: a precise point estimate, a precise CDF, imprecisely specified distributions, poorly known or even unknown dependencies, Non-negligible measurement uncertainty, non-detects or other censoring in measurements, small sample size, inconsistency in the quality of input data, model uncertainty, and non-stationary (non-constant distributions), etc.

Along with the development of the era, the probability box theory is also constantly updating its theory and application. Some stochastic algorithms such as the Bayesian reasoning, evidence theory are integrated into probability box theory. It also fuses with some artificial intelligence algorithms such as set theory, neural network, expert system etc. The USA Sandia National Laboratory in cooperation with the America Energy Department released an uncertainty research report<sup>[1]</sup> in 2003, which means that the probability box has entered a relatively mature stage in theory.

Its application has been the trend of diffusion in recent years. Its application areas include: the failure probability evaluation of fault system<sup>[2]</sup>, the uncertainty evaluation of vibration system dynamic responses<sup>[3]</sup>, The uncertainty of climate change<sup>[4]</sup>, Seawall risk modeling and reliability assessment<sup>[5-6]</sup>, Automobile gearbox reliability design<sup>[7]</sup>, finite element modeling and parameter optimization of rocket shell structure<sup>[8]</sup>, parameter uncertainty of damped oscillator<sup>[9]</sup>, multi-parameters uncertainty mathematical modeling<sup>[10]</sup>, Mechanical reliability system architecture and evaluation<sup>[11]</sup>, The flood control evaluation of water conservancy system<sup>[12]</sup>, The error accumulated expression and evaluation of the measuring system<sup>[13]</sup>, the sea level estimation in the further considering climate change<sup>[14]</sup>, etc. as its huge advantages in the express of uncertainties,

the application of its theory may explore to many other fields.

## II. THE BASIC PRINCIPLE AND CHARACTERISTICS OF THE PROBABILITY BOX THEORY

For a random variable X, when its estimated value is not a precise point estimate, and its cumulative probability function (Referred as the CDF) cannot be expressed by a single curve, the upper bound and lower bound are defined as following:

$$\bar{F}(x) = 1 - \underline{P}(X > x) \quad (1)$$

$$\underline{F}(x) = \underline{P}(X < x) \quad (2)$$

Where:  $\underline{P}(X < x)$  refers to the low probability of X,  $\bar{F}(x)$  refers to the upper bound of CDF, and  $\underline{F}(x)$  refers to the lower bound of CDF.

The bounds between  $\underline{F}(x)$  and  $\bar{F}(x)$  are then defined as a probability box (Referred as the p-box). The most important component of a p-box is the Dempster Shafer Structure (Referred as the DSS), which can be defined as following:

$$\{([x_1, y_1], m_1), ([x_2, y_2], m_2), \dots, ([x_n, y_n], m_n)\} \quad (3)$$

Where:  $[x_i, y_i]$  refers to a focus unit,  $m_i$  refers to the basic probability assignment of a focus unit.

The DSS can be converted to p-box. The relationship can be expressed as following:

$$\bar{F}(z) = \sum_{x_i \leq z} m_i, \quad \underline{F}(z) = \sum_{y_i < z} m_i \quad (4)$$

The characteristics of p-box can be summarized as follows:

- (1) It is evolved by the cumulative probability distribution function.
- (2) It can express almost all types of uncertainty.
- (3) It can accurately express the quantity of uncertainty.
- (4) It has very broad applicability.

## III. THE PROBABILITY BOX MODELING METHOD OF MECHANICAL FAULT FEATURES

Many features can be extracted from mechanical fault signals including different features in time domain, frequency domain and time-frequency domain. These features can be divided into the dimensional features and dimensionless features. The values of dimensionless features have nothing to do with the selection of units. It can be represented by a pure number. Compared with dimensional features, dimensionless feature can more effectively reflect the nature of mechanical faults. To illustrate the probability box modeling process, three types of dimensionless parameters are selected as follows:

$$C = \frac{X_{peak}}{X_{rms}} \quad (5)$$

$$S = \frac{X_{peak}}{X_{arv}} \quad (6)$$

$$L = \frac{X_{peak}}{X_r} \quad (7)$$

where

$$X_{peak} = \max(y) - \min(y), \quad \text{where } y = f(x) \quad (8)$$

$$X_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (9)$$

$$X_{arv} = \frac{1}{n} \sum_{i=1}^n |x_i| \quad (10)$$

$$X_r = \left[ \frac{1}{n} \sum_{i=1}^n \sqrt{|x_i|} \right]^2 \quad (11)$$

C is the peak factor (crest factor, also known as peak-to-average ratio, referred to as PAR). S is the form factor. L is the margin factor. These three factors are the different descriptions of the same parameters  $X_{peak}$  from three different angles respectively. Selection of the three parameters can compare similarities and differences of p-box modeling methods.

The probability box modeling algorithms of these three factors are described as follows:

- 1) Read the raw data of mechanical fault.
- 2) Divide the raw data into m segments according to the sampling frequency, which means there are n (the sampling frequency) pieces of data in each segment.
- 3) Obtain the peak factor  $c_i$  of each segment and constitute the row vector  $C = [c_1, c_2, \dots, c_m]$
- 4) Verify whether the data of crest factor C follow a normal distribution or not.
- 5) Divide the row vector C into t sets, which means there are m/t pieces of data  $c_i$  in every subset.
- 6) Obtain mean  $\mu_i$  and variance  $\sigma_i$  of each subset. The forms are shown as follows:

$$\mu = [\mu_1, \mu_2, \dots, \mu_t], \quad \sigma = [\sigma_1, \sigma_2, \dots, \sigma_t] \quad (12)$$

- 7) Obtain the maximum value ( $\mu_{max}, \sigma_{max}$ ) and the minimum value ( $\mu_{min}, \sigma_{min}$ ) of the means and variances respectively. Obtain the DSS forms of two parameters. The forms are shown as follows:

$$DSS_{\mu} = [\mu_{\min}, \mu_{\max}, 1]$$

$$DSS_{\sigma} = [\sigma_{\min}, \sigma_{\max}, 1] \quad (13)$$

8) Define the value of the basic probability assignment. For instance, the value is given as 0.001. And then obtain the discretized DSS of the above parameters. The forms are shown as follows:

$$DSS_{\mu} = \left\{ \begin{array}{l} ([\underline{\mu}_1, \bar{\mu}_1], 0.001), \\ ([\underline{\mu}_2, \bar{\mu}_2], 0.001), \\ \dots, \\ ([\underline{\mu}_{1000}, \bar{\mu}_{1000}], 0.001) \end{array} \right\} \quad (14)$$

$$DSS_{\sigma} = \left\{ \begin{array}{l} ([\underline{\sigma}_1, \bar{\sigma}_1], 0.001), \\ ([\underline{\sigma}_2, \bar{\sigma}_2], 0.001), \\ \dots, \\ ([\underline{\sigma}_{1000}, \bar{\sigma}_{1000}], 0.001) \end{array} \right\} \quad (15)$$

9) Obtain the p-box of the peak factor C based on the p-box algorithm as the condition of known the probability distribution.

#### IV. THE P-BOX MODELING RESULTS OF BEARING FAULT SIGNALS

To illustrate the results of p-box modeling, the rolling bearing fault signal was used. Experimental bearings were NU205 cylindrical roller bearings. The motor speed of 800 r/min. Sampling frequency is 1024 Hz. The fault characteristic frequency of rolling bearing by the theoretical formula are shown in Tab.1

TABLE I. ROLLING BEARING FAULT CHARACTERISTIC FREQUENCY (HZ)

Inner ring	Outer ring	Rolling element	Cage
95.36	64.6	33.38	5.38

Deep grooves in line were cut out to simulate the outer ring and inner ring fault. One element of the rolling body was cut in line to get a cutting groove, which was used to simulate the rolling fault. Bearing failure entities are shown in Fig. 1



Figure 1. Bearing failure entities

The experimental raw data of normal bearing, inner ring fault, the outer ring fault and rolling element fault were obtained respectively.

Fig .1 shows the peak factor's p-box modeling results of rolling bearing in four different states, according to the above modeling algorithm. Abscissa represents the interval number of uncertainty. The ordinate represents the cumulative probability distribution function. In order to articulate the horizontal axis, the data outside the scope [-4000, 4000] was truncated.

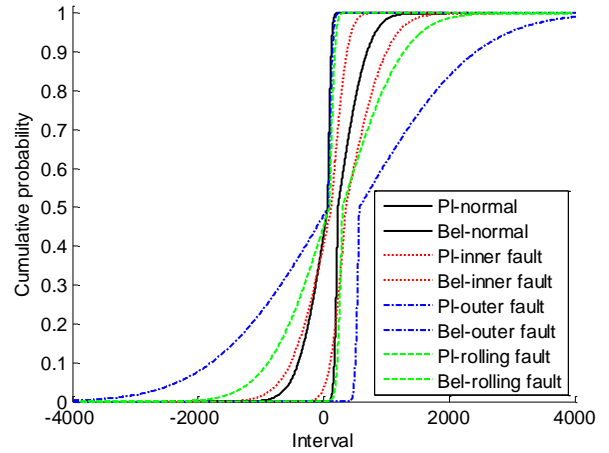


Figure 2. The peak factor's p-box modeling results of rolling bearing in four different states

The following points can analyzed from Fig .1:

- (1) The p-boxes in different fault states have different compactness due to their different uncertainty.
- (2) The p-box of normal bearing was the most compact, followed by the inner ring fault, rolling element fault, and finally the outer ring fault.
- (3) Different faults reflect different uncertainties.
- (4) The uncertainty characteristics expressed by p-boxes can contribute to the pattern recognition and fault diagnosis.

Fig .2 shows the form factor's p-box modeling results. In order to articulate the horizontal axis, the data outside the scope [-50000,50000] was truncated.

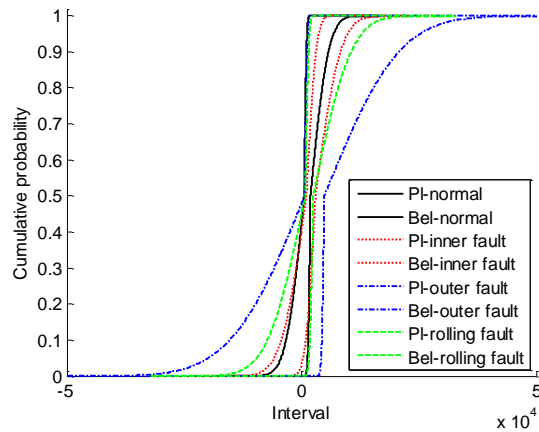


Figure 3. The form factor's p-box modeling results of rolling bearing in four different states

Figure .3 shows the margin factor's p-box modeling

results. In order to articulate the horizontal axis, the data outside the scope  $[-3 \times 10^6, 3 \times 10^6]$  was truncated.

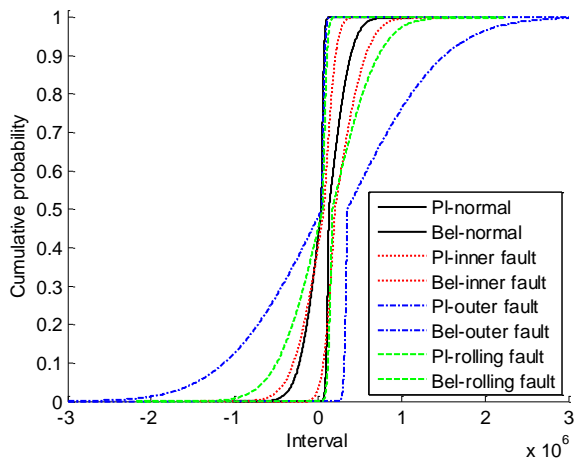


Figure 4. The margin factor’s p-box modeling results of rolling bearing in four different states

Compared with Fig .1, Fig .2 and Fig .3, they are very similar in graphics. Tab.1 shows the magnitude orders difference of three factors’ p-boxes. The data ranges of three type’s p-boxes are different in the magnitude orders. The uncertainty characteristics expressed by p-boxes can contribute to the pattern recognition and fault diagnosis.

TABLE II. THE MAGNITUDE ORDERS DIFFERENCE OF THREE FACTORS’ P-BOXES

The Type of P-Boxes	Peak Factor	Waveform Factor	Margin Factor
	$C$	$S$	$L$
Orders of magnitude	$10^3$	$10^4$	$10^6$

## V. CONCLUSIONS

In formula respects, the peak factor, waveform factor and margin factor are similar. They are all associated with  $X_{peak}$ . But their denominators represent different statistical characteristic parameters. Selecting these three factors to obtain p-boxes can compare the similarities and differences. From the similarities, the common can be reflected to reveal the nature. From the differences, the differences of p-boxes can reveal the influences of uncertainties.

The results show that the p-boxes shapes of these three factors are similar, which means the modeling method is verified by the three similar dimensionless fault features in time domain. Analyzed from all the figures of different p-boxes, the compactness of different types of p-boxes is different. The normal state has the most compact, which means its uncertainty is the least. Other fault states have different compactness. They are various in shapes and

scopes. All this information may contribute to the pattern recognition and fault diagnosis.

## ACKNOWLEDGMENT

The corresponding author of this paper is Ding Jiaman. This project is supported by National Natural Science Foundation of China (Grant No.51365020, 51369012), Yunnan Natural Science Foundation (Grant No.2013FZ020).

## REFERENCES

- [1] Scott Ferson, Vladik Kreinovich, Lev Ginzburg, et al. Constructing Probability Boxes and Dempster-Shafer Structures [R]. California : Sandia National Laboratories, 2003.
- [2] Berleant, Bounding. The Times to Failure of 2-Components Systems. IEEE Transaction on Reliability, 2004, 53(4):542-550
- [3] Fulvio Tonon. Using random set theory to propagate epistemic uncertainty through a mechanical system. Reliability Engineering and System Safety, 2004, 85:169-181
- [4] [10] Kriegler, Held. Utilizing belief functions for the estimation of future climate change. International Journal of Approximate Reasoning, 2005, 39:185-209
- [5] Scott Ferson, Troy Tucker. Sensitivity in risk analyses with uncertain numbers. Applied Biomathematics, Setauket, New York, 2005
- [6] Ferson, Tucker. Sensitivity analysis using probability bounding. Reliability engineering and system safety, 2006, 91(10): 1435–1442
- [7] Limbourg, Savic. Fault tree analysis in an early design stage using the Dempster-Shafer theory of evidence. Risk, Reliability and Societal Safety. London : Aven & Vinnem (eds) Taylor & Francis Group, 2007
- [8] Michael Oberguggenberger, Julian King, Bernhard Schmelzer. Imprecise probability methods for sensitivity analysis in engineering. In 5th International Symposium on Imprecise Probability: Theories and Applications, Prague, Czech Republic, 2007, 6: 1130-1138
- [9] Destercke, Dubois, Chojnacki, Unifying practical uncertainty representations – I: Generalized p-boxes [J], International Journal of Approximate Reasoning, 2008,49(3):649-663
- [10] Matthias Troffaes, Sebastien Destercke. Probability boxes on totally preordered spaces for multivariate modelling, International Journal of Approximate Reasoning, 2011,52(6):767-791
- [11] Luis G. Crespo, Sean P. Kenny, Daniel P. Giesy. Reliability analysis of polynomial systems subject to p-box uncertainties, Mechanical Systems and Signal Processing, 2012,9:111-124
- [12] Siao Sun, Guangtao Fu, Slobodan Djordjević, Soon-Thiam Khu. Separating aleatory and epistemic uncertainties: Probabilistic sewer flooding evaluation using probability box[J], Journal of Hydrology, 2012,420(2):360-372
- [13] Matthias C.M. Troffaes, Enrique Miranda, Sebastien Destercke. On the connection between probability boxes and possibility measures Original Research Article, Information Sciences, 2013,224 (3): 88-108
- [14] N. Ben Abdallah, N. Mouhous-Voyneau, T. Denoeux. Combining statistical and expert evidence using belief functions: Application to centennial sea level estimation taking into account climate change[J]. International Journal of Approximate Reasoning, 2014, 55(1):341-354