Complexity and Entropy Representation for Machine Component Diagnostics

Srinivasan Radhakrishnan[a], Yung-Tsun Tina Lee[b], Sudarsan Rachuri[c], Sagar Kamarthi[a,*]

[a] *Mechanical and Industrial Engineering, Northeastern University, Boston MA, USA*
[b] *Systems Integration Division, NIST*
[c] *Advanced Manufacturing Office, Department of Energy, Gaithersburg MD*

Abstract

The Complexity-entropy causality plane (CECP) is a parsimonious representation space for time series. It has only two dimensions: normalized permutation entropy ($\mathcal{H}_S$) and Jensen-Shannon complexity ($\mathcal{C}_{JS}$) of a time series. This two-dimensional representation allows for detecting slow or rapid drifts in the condition of mechanical components monitored through sensor measurements. The CECP representation can be used for both predictive analytics and visual monitoring of changes in component condition. This method requires minimal pre-processing of raw signals. Furthermore, it is insensitive to noise, stationarity, and trends. These desirable properties make CECP a good candidate for machine condition monitoring and fault diagnostics. In this work we demonstrate the effectiveness of CECP on three rotary

*sagar@coe.neu.edu

component condition assessment applications. We use CECP representation of vibration signals to differentiate various machine component health conditions for rotary machine components, namely roller bearing and gears. The results confirm that the CECP representation is able to detect, with high accuracy, changes in underlying dynamics of machine component degradation states. From class separability perspective, the CECP method is able to generate linearly separable classes for the classification of different fault states. The classification performance improves with increasing signal length. For signal length of 16,384 data points, the fault classification accuracy varies from 90% to 100% for bearing applications, and from 85% to 100% for gear applications. We observe that the optimum parameter for CECP depends on the application. For bearing applications we find that embedding dimension $D = 4, 5, 6$, and embedding delay $\tau = 1, 2, 3$ are suitable for fault classification. For gear applications we find that embedding dimension $D = 4, 5$, and embedding delay $\tau = 1, 5$ are suitable for fault classification.

Keywords: Complexity-Entropy Plane, Permutation Entropy, Support Vector Machine, Machine Fault Diagnostics Component Condition Visualization

1. Introduction

Prognostics and Health Management (PHM) and its functions, diagnostics and prognostics, are key areas of interest for smart manufacturing. While, diagnostics detects and identifies a failure mode within a component, prognostics predicts the degradation trends of a component and estimates the remaining useful life (RUL) of a component or system. Table 1 lists the common machine component problems addressed by PHM [1]. This paper investigates a novel sensor data representation method that allows both automatic and visual detection of health condition of bearings and gears, which are the most prevalent components among the ones listed in Table 1.

Table 1: Common machine component and machine subsystem problems in PHM [1]

Machine Component/Subsystem	**Problems**
Bearing	Outer-race, inner- race, roller, cage failures
Gear	Manufacturing error, missing tooth, tooth pitting, gear crack, gear fatigue/wear
Shaft	Unbalance, bend, crack, misalignment, rub
Pump (machine subsystem)	Valve impact,score, fracture,pistons lap, defective bearing and revolving crank, hydraulic problem
Alternator (machine subsystem)	Stator faults, rotor electrical faults, rotor mechanical faults

Bearings perform a key function of converting sliding friction into rolling friction in rotary machines. A bearing consists of an inner ring, an outer ring, a set of ball rollers or cylindrical rollers (usually termed as rolling elements), and a cage. The rolling elements are placed inside a cage which is then set

between the inner ring and outer ring. The bearing fault can occur in any of the aforementioned four components of a roller bearing. The causes for faults in bearing include, increase of operating loads, shaft imbalance or bent shaft, surface defects, surface roughness, surface contamination, and presence of particles on inner or outer raceway [2]. El-Thalji and Jantunen [3] present dynamics of wear progression of roller bearings.

Gears, like bearings, serve a critical function in a rotary machine. Like bearings, gears operate in tough operating conditions involving static and impact forces. Gears are subjected to wear in the form of cracking, pitting, and scaling which eventually culminates into a chipped tooth or broken tooth condition. When such a fault state is reached, gears do not perform as expected and hence they are, like bearings, ought to be continuously monitored for their health condition.

For both bearings and gears, sensor-based monitoring methods are viable options for fault detections and estimation of RUL. The common indirect measurements include vibration, acoustic emission, and power sensors. The general steps for implementing machine learning models for machine component fault detection are signal acquisition, signal processing, feature extractions, and building a machine learning model for classification and regression.

Most of the sensor generated structured data are time series data. Dimensionality reduction is an inherent challenge, particularly when real-time data processing is required and data streams are generated continuously at high volume and high rate [4]. Dimensionality reduction through parsimonious signal representation and feature extraction are necessary for diagnostics and prognostics of mechanical systems. Time domain features commonly include Root Mean Square (RMS), Average, Kurtosis, crest factor, autoregressive models [2]. Studies have found that RMS extracted from vibration measurements exhibits a strong correlation with bearing wear [2, 5–8]. Similarly, kurtosis and crest factor are sensitive to the signal shape; kurtosis is also sensitive to rotational speed and the frequency bandwidth of the signal [2]. Fast-Fourier Transform (FFT) is the most widely applied method for frequency domain analysis. However due to the non-stationary nature of the signals, time-frequency domain analysis is preferred. WignerVille distribution (WVD), wavelet transform (WT) [9], discrete wavelet transform (DWT) [10], and short time Fourier transform (STFT) are some of the methods of choice for bearing fault diagnosis. Several studies have used time domain analysis for gear fault detection [11–14]. RMS and kurtosis are found to be sensitive to gear faults, especially cracked gear [14]. Apart from time

domain, frequency domain has been explored to represent features for gear fault detection [15–18] and time-frequency domain [19–21] has been used for enhancing accuracy, robustness, and reducing sensitivity to noise.

This paper presents a different approach signal processing for diagnostics and prognostics of bearings and gears. It investigates a complexity-entropy causality plane (CECP) based sensor data representation that is parsimonious and effective in detection of faults in bearings and gears. In Section 1 we reviewed the different sensor representation methods which includes, time, frequency and time-frequency methods. In section 2 we review the different entropic methods used for characterizing time series signals. In Section 3 we present the formulation for the CECP representation and discuss the parameter selection for machine component monitoring applications. In Section 4 we present the results for roller bearing fault detection and helical gear fault detection and discuss the merits of the proposed approach. In the conclusion section we discuss the advantages of CECP method over existing time, frequency, and time-frequency based approaches commonly used in PHM applications.

2. Sensor Data Representation Methods

The representation of the sensor data is the process of transforming a raw time series vector $X = \{x_1, x_2, \ldots, x_n\}$ into a vector $F = \{f_1, f_2, \ldots, f_k\}$ in transformation space such that $k < n$, where n is the number of data points in a given time series, and k is the number of entities in the transformed space. By reducing the dimension, the computational complexity is reduced from $\mathcal{O}(n)$ to $\mathcal{O}(k)$. The desired properties of a good representation are preservation of local and global characteristics with low information loss and high robustness in the presence of noise and outliers.

The sensor signal analysis using time, frequency, or time-frequency domain are well suited when the underlying system dynamics exhibit a linear behavior. In case of machine components like bearing, gears, or combination of coupled systems, sensor signals exhibit high complexity and nonlinear characteristics. In recent years, entropic measures have proven to be effective to deal with complex and nonlinear sensor signals for feature extraction in rotary component fault classification and estimation problems. Approximate Entropy (ApEn) [22–24] and Sample Entropy (SampEn) [25] were developed for characterizing nonlinear time series in biomedical applications. SampEn is an improvement over ApEn, where ApEn performance

depends on the signal length and yields lower estimation for shorter lengths [25]. For PHM applications, ApEn [26, 27] method has been adopted for monitoring machine health and SampEn [28–30] has been adopted for bearing fault diagnosis. The multi-scale entropy (MSE) method developed by Costa et al. [31, 32] proved effective in case of machine health monitoring where interaction between multiple components (bearing, gear, and shafts) generate vibration signals that contain multiple intrinsic oscillatory modes in which case single scale entropic methods such as ApEn and SampEn may be less effective in characterization of the measured signals [33, 34]. Unlike ApEn and SampEn, MSE analyzes signals in multiple time scales rather than in a single time scale, since entropy values do not necessarily capture complexity changes. Zhang et al. [33] applied MSE method for bearing fault application.

Permutation entropy, introduced by Bandt and Pompe [35] has been used for analyzing and characterizing nonlinear time series data. Studies have adopted Permutation entropy for fault classification of mechanical components [36, 37]. Using permutation entropy alone as a feature limits its ability to classify different types of faults within fault domain. Wu et al. [38] found that permutation entropy on its own as a feature does not fare well with

classification algorithms. They used a technique called as multi-scale permutation entropy for feature extraction and used support Vector Machine (SVM) to classify fault signals. However, using multi-scale permutation entropy, one may end up with as many features as scales used with redundant information, hampering accuracy, and causing further increase in analysis time. Li et al. [39] used Laplacian score [40] to select the important features generated by multi-scale permutation entropy technique.

In this work we investigate Complexity-Entropy Causality Plane (CECP) as a data representation and feature extraction technique for machine fault diagnostic applications. The key interest in CECP stems from the fact that it handles both stationary and non-stationary signals, reduces the number of features required for accurate classification and prediction, and allows for visualization of time series in two-dimensional space. We demonstrate the effectiveness of CECP method to detect machine component faults on publicly available sensor datasets on bearings and gears.

3. CECP Representation of Sensor Data

Shannon entropy, $S[P]$, is a popular measure to compute the information associated with a process described by a probability distribution $\{P = p_i : i = 1, 2, \ldots M\}$.

However, for differentiating simple processes from complex processes, which exhibit organizing patterns, $S[P]$, is ineffective [41, 42]. In addition, Shannon entropy on its own does not capture temporal relationship between values in a time series. It requires prior knowledge about the process in the form of an underlying probability distribution function [43], yet it poorly characterizes highly non-linear processes (e.g. chaotic systems).

To overcome the aforementioned limitations, Bandt and Pompe (BP) [35] provided a method to extract the underlying probability distribution from the time series. The BP method is effective since the emergence of different permutation patterns reflects the dynamics of the underlying process. The BP method is non-parametric, rank based, and the probability of the ordinal patterns is invariant to nonlinear monotonic changes [42, 44]. This renders a good quantifier which is robust against nonlinear drifts and scaling [42] with ability to handle non-stationary time series [44–46]. The BP method for generating probability distributions is a simple symbolization technique that incorporates causality in the evaluation of the probability distributions associated with a time series [43]. For a given time series $X = \{x_t : t = 1, 2, ...N\}$, at each time instance s, a sequence of values $X_s = \left\{x_s, x_{s+\tau(1)}, ..., x_{s+\tau(D-1)}\right\}$ is selected, where D represents the *embedding dimension* and τ represents

the *embedding delay*. The embedding dimension D captures the amount of information contained in X_s. An ordinal pattern (0 1 ... $D-1$) is assigned to X_s such that (0 1 ... $D-1$) $\mapsto \left\{x_s, x_{s+\tau(1)}, ..., x_{s+\tau(D-1)}\right\}$. The values of X_s are then sorted in ascending order and the corresponding ordinal pattern is shuffled in parallel. The shuffled ordinal patterns are represented as π. In case the elements of the time series segment are identical, the corresponding ordinal pattern is taken (without shuffling) as the resulting permutation pattern. When the embedded dimension is D, $D!$ permutation patterns are possible. The relative frequency p_i of each π_i is obtained by dividing the count of permutation pattern π in the signal by the total number of permutation patterns (of any type) in the signal. Thus, $p_i = |\pi_i| / \sum_{i=1}^{D!} |\pi_i|$ for $i = 1, 2, \ldots, D!$ and $P = \{p_i : i = 1, 2, \ldots, D!\}$ is the probability distribution of the permutation patterns in the signal. Here $|\pi_i|$ is the count of occurrence of permutation pattern π_i. The *permutation entropy* is computed as

$$S[P] = -\sum_{i=1}^{D!} p_i log(p_i) \tag{1}$$

The permutation entropy as defined above takes maximum value when $p_i = p = \frac{1}{D!}$ for all i. From this the max $S[P] = \log D!$. Thus the normalized

permutation entropy is given by

$$\mathcal{H}_S = \frac{S[P]}{\log D!} \tag{2}$$

where $\mathcal{H}_S \in (0, 1)$.

Entropic measures, including *permutation entropy*, are able to quantify information but they do not capture the structure or patterns in a process [43, 47]. To uncover organizational patterns of a process, several statistical complexity measures (SCM) have been developed [42]. Of the many SCM metrics, Jensen-Shannon complexity $\mathcal{C}_{JS}$ combines both information and disequilibrium measures has demonstrated effective detection of the underlying dynamics and is defined as [48, 49]

$$\mathcal{C}_{JS} = Q_J[P, P_e]\mathcal{H}_S \tag{3}$$

where $P_e = \{1/D!, 1/D!, \ldots, 1/D!\}$ is the uniform distribution and disequilibrium Q_J is Jensen-Shannon divergence $J[P, P_e]$ that links two probability distributions.

$$Q_J[P, P_e] = Q_0 J[P, P_e] \tag{4}$$

where

$$J[P, P_e] = S[(P + P_e)/2] - S[P]/2 - S[P_e]/2 \tag{5}$$

and Q_0 is a normalization constant equal to the inverse of the maximum possible value of $J[P, P_e]$, which happens when one of the values of P is 1 and all other values are 0 [42]. The Q_0 is computed using the following equation [50]

$$Q_o = \frac{1}{-(\frac{1}{2})\left[\frac{D!+1}{D!}\log(D!+1) - 2\log(2D!) + \log(D!)\right]}. \tag{6}$$

. The inclusion of factor Q_0 ensures that $0 \leq Q_J[P, P_e] \leq 1$. The Jensen-Shannon divergence quantifies the difference between two probability distributions and is a non-trivial function of entropy, which uses two different probability distributions; P is the probability distribution, which represents the state of the system and P_e is a uniform distribution, which serves as a reference [42]. As implied by the second law of thermodynamics, $\mathcal{H}_S$ can be regarded as the time dimension and the $\mathcal{H}_S$ versus $\mathcal{C}_{JS}$ mapping can be used to study the temporal evolution of the SCM [51]. Also, for a given value of $\mathcal{H}_S$, the values of complexity $\mathcal{C}_{JS}$ varies between a minimum and maximum boundary which are termed as limit curves [52]

Signal length N, and embedding dimension D effect the value of permutation entropy. A large value of N may result in a near constant features on CECP, thereby reducing the ability of CECP to capture the dynamic changes. On the other hand a small value of N may yield statistically insignificant results.

Literature points out that in order to use CECP to differentiate between chaotic and stochastic processes it is necessary to satisfy the condition that signal length is relatively much larger than $D!$, i.e., $N \gg D!$ [44, 53]. CECP approach is able to distinguish stochastic processes with different long range correlations when D is between 3 and 6 [44–46]. For practical purposes Bandt and Pompe [35] recommended $3 \leq D \leq 7$ and $\tau = 1$.

With respect to vibration signals for component fault diagnostics, Yan et al. [36] studied the relationship between N and $\mathcal{H}_S$. After analyzing signal of different lengths (N=32, 64, 128, 256, 512, 1024, and 2048), they reported that variation in $\mathcal{H}_S$ values for $N > 256$ is insignificant; they found a stable and consistent $\mathcal{H}_S$ value when $N = 128$ or $N = 256$. Yan et al. [36] observed that (1) when $D < 4$ permutation entropy is not able to detect the exact dynamic changes in the mechanical vibration signals; (2) a $D > 8$ is computationally expensive; and (3) time delay $\tau > 5$ is not conducive for detecting small changes in the signal. They used $D = 6$ and $\tau = 3$ for computing permutation entropy values for component fault diagnostics.

CECP analysis is extensively used in characterizing correlated stochastic processes [45, 54, 55], and distinguishing chaotic and stochastic processes [44, 51]. In econophysics, CECP is used to quantify stock market inefficiencies [56],

asses efficiency of bond markets [57], and perform commodity analysis [58]. In addition, CECP is used in river flow characterization [42], temperature variation in mountain streams [59], and song classification [50].

In this paper, we demonstrate the performance of CECP method for characterizing faults of rotary machine components in three different applications: (1) ball bearing dataset from Machinery Failure Prevention Technology (MFPT) Society, (2) bearing dataset from Case Western Reserve University, and (3) gear dataset from PHM Society.

4. Results

4.1. MFPT Ball Bearing Experiment

In this application we consider two classes of vibration signal samples from ball bearing dataset provided by the MFPT Society [60]. Bechhoefer [60] complied this dataset from a bearing test rig experiment. The dataset contains labeled signatures of faulty outer race and inner race of ball bearings. The signatures of faulty outer race and inner race were generated at a constant shaft rotational speed of 1500 rpm and at seven different load conditions: 11.33, 22.67, 45.35, 68.0388, 90.71, 113.39, and 136.07 kg (or 25, 50, 100, 150, 200, 250, and 300 lbs). Each vibration signal is recorded for 3

seconds at 48,828 Hz frequency; it resulted in a signal of length 146,484 data points. Figures 1(a) and 1(b) show the sample vibration signals representing the fault signatures of inner race and outer race of the ball bearing collected at 45.35 kg (100 lb) load and shaft rotational speed of 25 rps (1500 rpm). The vibration signals for both inner race fault and the outer race fault are periodic. The fault characteristics frequencies for inner race fault and outer race fault are outlined by Zhang et al. [61]. We performed Augmented Dickey-Fuller (ADF) test to verify the stationarity of the signals. For both the inner race and outer race fault we obtain a p value of $0.01 < 0.05$ confirming the stationary characteristics of the signals.

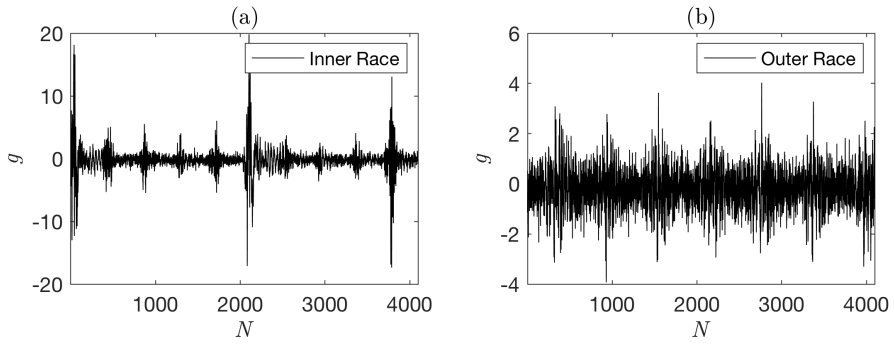

Figure 1: Sample vibration signals representing (a) inner race fault and (b) outer race fault

The raw data includes inner race fault and outer race fault signals of 146,484 data points each. For demonstration , we split each set of signals into

35 sub-signals of 4,096 data points. Figure 2(a) plots *Permutation Entropy* ($\mathcal{H}_S$) values of 35 outer race and 35 inner race sub-signals. The overlap of $\mathcal{H}_S$ values for inner race and outer race signals indicates that $\mathcal{H}_S$ is not effective parameter to distinguish inner race faults from outer race faults correctly. Figure 2(b) plots CECP map (i.e., scatter plot of $\mathcal{C}_{JS}$ vs. $\mathcal{H}_S$) of 35 outer race and 35 inner race sub-signals. We observe that the CECP representation of outer and inner race sub-signals is able to separate outer and inner race faults in the plot.

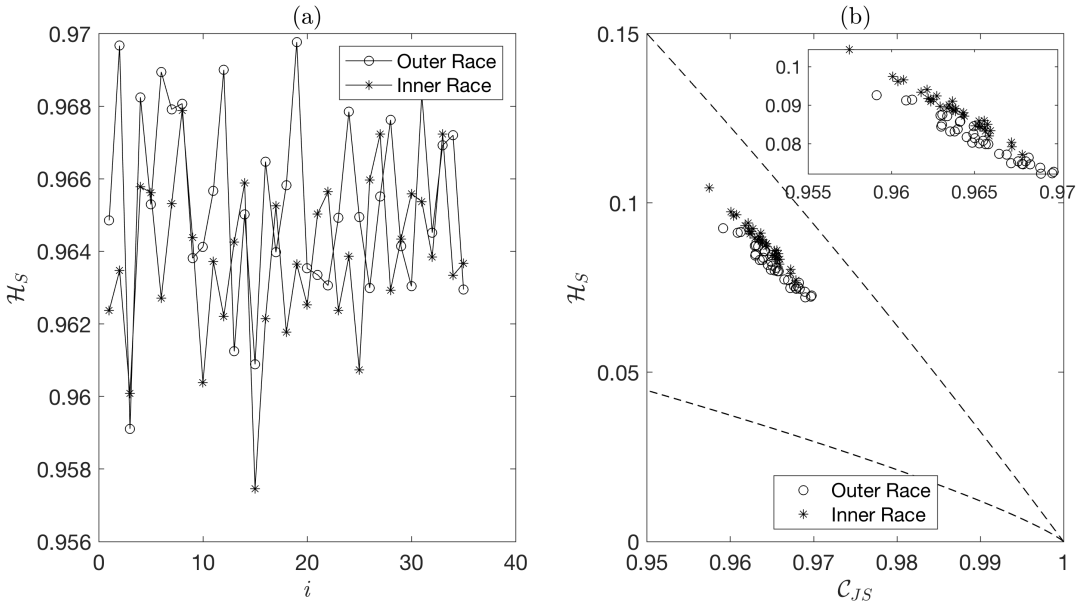

Figure 2: (a) Permutation entropy values of 35 inner race fault signals and 35 outer race fault signals, and (b) CECP values of 35 inner race fault signals and 35 outer race fault signals. The dashed lines represent the lower and upper limit curves. $N = 4096$, $D = 6$, and $\tau = 1$. Inset figure is a scaled version of subplot (b)

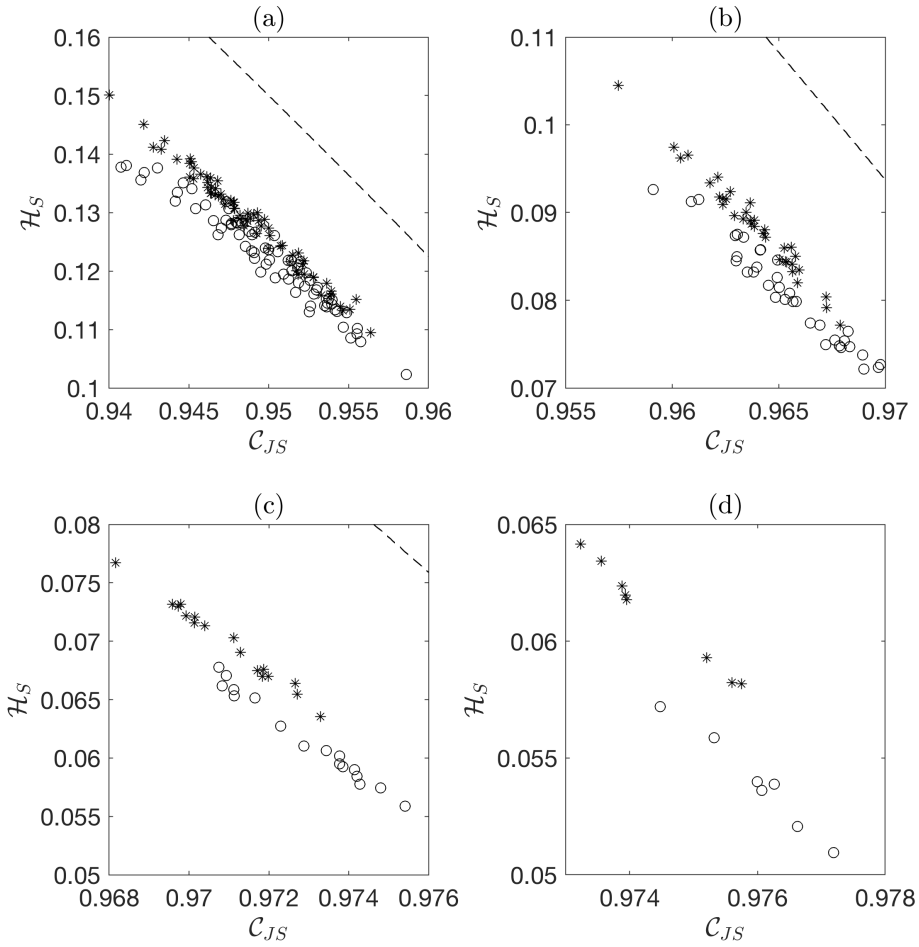

Figure 3: Complexity $\mathcal{C}_{JS}$ vs. Entropy $\mathcal{H}_S$ feature plane for different values of subsample length N for the case of 45.35 kg (100 lb) load and 25 rps (1500 rpm) rotational speed: (a) $N = 2048$, (b) $N = 4096$, (c) $N = 8192$,and (d) $N = 16384$; the CECP parameters are set to $D = 6$ and $\tau = 1$. The stars represent inner race fault and the circles represent outer race fault. The dashed lines represent the lower and upper limit curves. Some figures may not show the limit curves due to axis scale effects.

Figure 3 presents CECP maps for four different signal lengths. It is evident from the figure that separability between fault classes widens and cluster variance of $\mathcal{H}_S$ decreases with increasing signal length. This leads to improvement of Dunn index of $\mathcal{H}_S$ with increasing signal length (see Figure 4). Dunn index, which is the ratio of minimum inter-cluster distance to maximum intra-cluster distance [62], measures the quality of cluster formations. The higher the Dunn index the better the cluster quality. Dunn index for m number of clusters is defined as

$$\text{Dunn Index} = \frac{\min_{1\leq i<j\leq m} d(C_i, C_j)}{\max_{1\leq k\leq m} diam(C_k)} \tag{7}$$

where $d(C_i, C_j)$ is dissimilarity (inter-cluster) between clusters C_i and C_j; it is defined as $d(C_i, C_j) = \min_{\mathbf{a}\in C_i, \mathbf{b}\in C_j} E(\mathbf{a}, \mathbf{b})$, where $E(\mathbf{a}, \mathbf{b})$ is the Euclidean distance between points $\mathbf{a}$ and $\mathbf{b}$. $diam(C_k) = \max_{\mathbf{a},\mathbf{b}\in C_k} E(\mathbf{a}, \mathbf{b})$ is the diameter of cluster C_k.

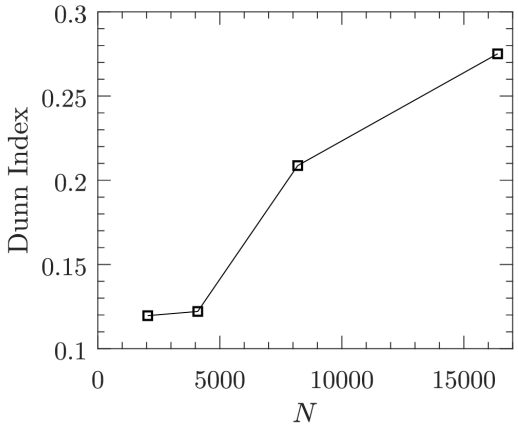

Figure 4: Effect of signal length N on Dunn Index. In this example, N=2048, 4096, 8192, and 16384. (Case: 45.35 kg (100 lb) load, 25 rps (1500 rpm) rotational speed)

As mentioned in the earlier section, $D < 4$ for mechanical vibration is not desired since the permutation entropy is not able to detect exact dynamic changes. On the other hand $D > 8$ is computationally expensive. Similarly, $\tau > 5$ is not recommended for vibration signals. We perform sensitivity analysis for all the load conditions by varying N, D, and τ. The results of the sensitivity analysis are given in the supporting information Figures S1 to S12. We set the parameter values as follows, N=2048, 4096, 8192, 16384, and 32768, D= 3, 4, 5, and 6, and τ= 1, 2, 3, 4, and 5. We did not increase the length of the signal beyond 16384 because it results in fewer training records. In general, CECP is used for characterizing time series

and the parameter choice effects the characterizations. Here we are more interested is selecting parameter values that enhances class separability (or classification accuracy). From the sensitivity analysis we find that for load conditions 11.33 kg and 22.67 kg, parameters D =4, 5, 6, and τ =1, 2, 3, 4 are suitable to achieve class separability. For load conditions 45.35 kg, 68.03 kg, and 90.71 kg, parameters D =4, 5, 6, and τ =1, 4 are suitable to achieve class separability. For load conditions 113.39 kg, and 136.07 kg, parameters D =4, 5, 6, and τ =1, 2, 3 are suitable to achieve class separability. To overcome the visual limitations of observing the separability, we use support vector machine (SVM) to see how the classification accuracy improves with respect to the increase in the signal length. For demonstration purpose we use $D = 6$ and $\tau = 1$ (for practical purposes Bandt and Pompe [35] recommended $3 \leq D \leq 7$ and $\tau = 1$). We use Receiver Operating Characteristic (ROC), Area Under Curve (AUC), and classification accuracy (ACC) to evaluate the performance of the SVM classifier. We use linear SVM model with 5 cross validation. The results of the SVM are given in supporting information Table ST1. Figures S13 to S15 (supporting information) plot the ROC, AUC and ACC values. We observe that for all load conditions the SVM classifier performance improves with increase in signal length which is consistent with

the earlier analysis using Dunn Index.

4.2. Bearing Dataset from CWRU

Case Western Reserve University (CWRU) bearing dataset [63] includes high quality signals collected at normal and faulty working conditions of bearings. Figure 5 shows the setup of the experiment. The testbed consists of a 2-hp motor (left side), a torque transducer/encoder (center), and a dynamometer (right side).

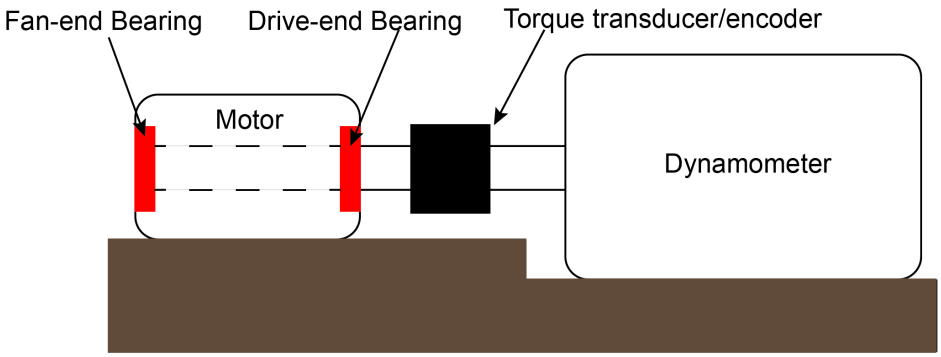

Figure 5: Schematic of CWRU experimental setup for bearing fault data collection [63]

The setup has test bearings located at the drive-end and the fan-end of the motor. The faults were introduced in inner race, outer race and ball for both drive-end bearings and the fan-end bearings using electro-discharge machining.

The accelerometers attached to the motor housing using magnetic bases were used to measure vibration signals from the setup. One accelerometer was attached on the drive-end of the motor and another on the fan-end of the motor. For some experiments, an additional accelerometer was attached to base plate supporting the motor. Vibration signals were collected using a 16-channel DAT recorder. Sensor signals were collected at a frequency of 12,000 Hz. The length of each fault-related signals was varied between 120,000 and 130,000 data points while the length of the baseline signals was varied between 200,000 and 500,000 data points. Figure 6 shows the sample vibration signals representing the baseline condition, ball fault and inner race fault. The fault characteristics of the CWRU dataset are exhaustively studied in time and frequency domains by Smith et al. [64]. We performed Augmented Dickey-Fuller (ADF) test to verify the stationarity of the signals. For all the three cases of baseline, inner race fault and ball fault we obtain a p value of $0.01 < 0.05$ confirming the stationary characteristics of the signals.

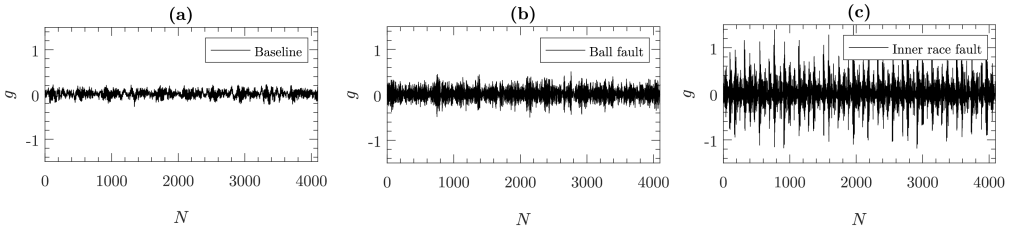

Figure 6: Sample vibration signals representing (a) baseline condition, (b) ball fault, and (c) inner race fault. In this case, the fault diameter for both the ball and inner race is 0.17 mm (0.007 inches), load is 0 kW (0 hp), and rotational speed is 29.95 rps (1797 rpm). The bearing considered is drive-end bearing

Speed and horsepower data were hand-recorded from the torque transducer/encoder. In this case we analyze only baseline signals (normal working condition), inner race fault signals, and ball fault signals. The experimental parameters are outlined in Table 2. For all the parameter variations the fault depth was maintained at 0.2794 mm (0.011 inches).

Table 2: Drive-end bearing and fan-end bearing data for analysis [63]

Fault Diameter (mm)	**Motor Load (kW)**	**Approx. Motor Speed (rps)**
0.17 mm (0.007")	0	29.95 rps (1797 rpm)
	0.73 kW (1 hp)	29.53 rps (1772 rpm)
	1.47 kW (2 hp)	29.16 rps (1750 rpm)
	2.20 kW (3 hp)	28.83 rps (1730 rpm)
0.35 mm (0.014")	0	29.95 rps (1797 rpm)
	0.73 kW (1 hp)	29.53 rps (1772 rpm)
	1.47 kW (2 hp)	29.16 rps (1750 rpm)
	2.20 kW (3 hp)	28.83 rps (1730 rpm)
0.53 mm (0.021")	0	29.95 rps (1797 rpm)
	0.73 kW (1 hp)	29.53 rps (1772 rpm)
	1.47 kW (2 hp)	29.16 rps (1750 rpm)
	2.20 kW (3 hp)	28.83 rps (1730 rpm)

We perform sensitivity analysis for all the load conditions by varying N, D, and τ. The results of the sensitivity analysis for selected operating conditions are given in the supporting information Figures S16 to S23. We set the parameter values as follows, N=2048, 4096, 8192, 16384, and 32768, D= 3, 4, 5, and 6, and τ= 1, 2, 3, 4, and 5. e did not increase the length of the signal beyond 16384 because it results in fewer training records. From the sensitivity analysis we find that for all operating conditions given in Table 2, for fan-end and drive-end bearing D=4, 5, 6 and τ= 1, 2 are suitable parameters to obtain class separability. For both drive-end bearing and fan-end bearing the baseline condition exhibit higher complexity and lower permu-

tation entropy compared to inner race and ball fault condition. We observe that the fan-end bearing exhibits higher class separation between baseline and faulty conditions while in case of drive-end bearing the class separability are almost equal between baseline, ball fault and inner race fault conditions.

To overcome the visual limitations of observing the separability, we use support vector machine (SVM) to see how the classification accuracy improves with respect to the increase in the signal length. For demonstration purpose we use $D = 6$ and $\tau = 1$ (or practical purposes Bandt and Pompe [35] recommended $3 \leq D \leq 7$ and $\tau = 1$). We use linear SVM model with 5 cross validation. The results of the SVM for fan-end bearing are given in supporting information Table ST3 and the results of the SVM for drive-end bearing are given in supporting information Table ST4. Figures S24 and S25 (supporting information) plot the ROC, AUC and ACC values. Similar to MFPT experiment, we observe that for almost all operating conditions the SVM classifier performance improves with increase in signal length.

4.3. Gear Dataset from the PHM Society

We consider a dataset provided by the PHM society that contains labeled data on different types of gear degradation [65]. The experiments were conducted using spur gears and helical gears. For CECP application we con-

sider experiments with helical gears. Figure 7 shows the experimental setup. The setup is common for both spur gears and helical gears. Figure 7 shows the details of number of gear teeth for both spur and helical gear. For our analysis we will consider only the helical gear case.

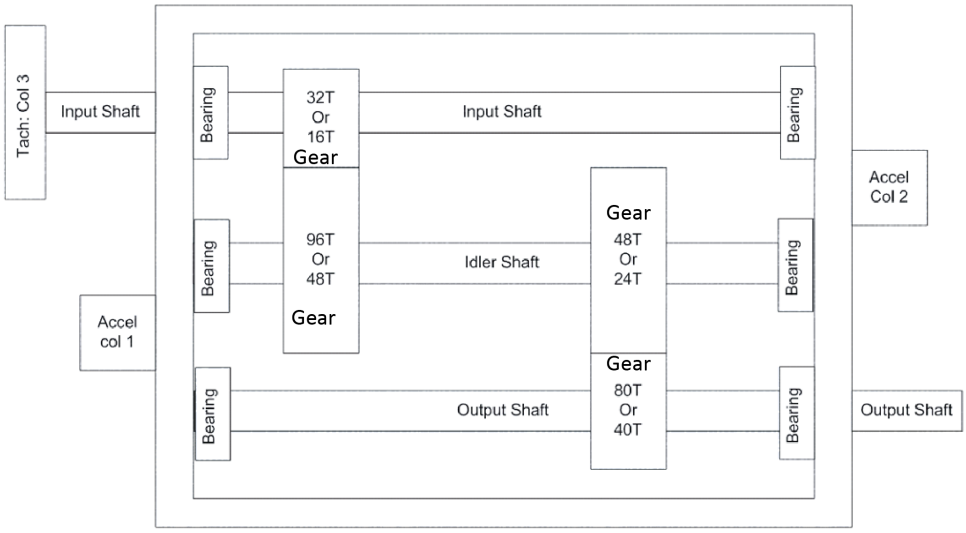

Figure 7: Experimental setup for gear fault detection [65]

The setup has an input shaft, an idler shaft and an output shaft on which the gears are mounted. The input side is on the left and the output side is on the right. Two accelerometers are mounted, one on the input side and the other on the output side. The helical gear on the input shaft towards left has 16 teeth and the helical gear on the idler shaft towards left has 48 teeth.

The helical gear on idler shaft on the right has 24 teeth and the helical gear on the output shaft on the right has 40 teeth. Experiments with helical gears were performed six times. All in all, there are two fault categories (chipped tooth and broken tooth) and one baseline category (no known faults).

From the dataset we consider the case titled helical 1 (has no known gear defects were found) as the *baseline* case. We consider helical 2 (has a chipped tooth in helical gear with 24 teeth) as a *chipped tooth* gear category and helical 5 (has a broken tooth in helical gear with 24 teeth) as a *broken tooth* gear category. In all the three cases, we use the vibration signals recorded from accelerometer 2 (placed on the output side). The signals were recorded under two different load conditions (labeled as Low and High) and five different rotational speeds (i.e., 30 rps (1800 rpm), 35 rps (2100 rpm), 40 rps (2400 rpm), 45 rps (2700 rpm), and 50 rps (3000 rpm)). For each of these settings, two signals were recorded for four seconds each. Thus for one fault signal, 533,312 data points were generated for eight-second recording at a sampling rate of 66,666.67 samples per second . Figure 8(a), 8(b), and 8(c) show the sample signals of length 4096 data points representing the baseline condition and the fault condition. We performed Augmented Dickey-Fuller (ADF) test to verify the stationarity of the signals. For all the three cases of

baseline, chipped tooth and broken tooth we obtain a p value of $0.01 < 0.05$ confirming the stationary characteristics of the signals.

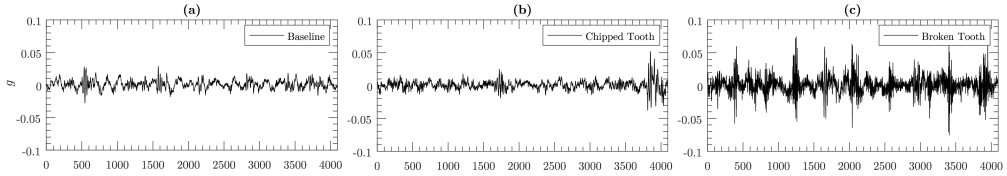

Figure 8: Sample signal representing **(a)** baseline condition, **(b)** chipped tooth condition, and **(c)** broken tooth condition. The sample signals shown here are taken from experimental condition of High load and 50 rps (3000 rpm) rotational speed

We perform sensitivity analysis for all the load conditions by varying N, D, and τ. The results of the sensitivity analysis for selected operating conditions are given in the supporting information Figures S26 to S41. We set the parameter values as follows, N=2048, 4096, 8192, 16384, and 32768, D= 3, 4, 5, and 6, and τ= 1, 2, 3, 4, and 5. We did not increase the length of the signal beyond 16384 because it results in fewer training records. We observe that D=4, 5 and τ =1, 5 are suitable for obtaining class separability. To overcome the visual limitations of observing the separability, we use support vector machine (SVM) to see how the classification accuracy improves with respect to the increase in the signal length. For demonstration purpose we use $D = 6$ and $\tau = 1$ (or practical purposes Bandt and Pompe [35] recom-

mended $3 \leq D \leq 7$ and $\tau = 1$). We use linear SVM model with 5 cross validation. The results of the SVM for low and high load conditions are given in supporting information Table ST5. Figures S42 to S45 (supporting information) plot the ROC, AUC and ACC values. Similar to MFPT experiment and CWRU experiment, we observe that for almost all operating conditions the SVM classifier performance improves with increase in signal length. The same observation is confirmed using the Dunn Index values that increase with increase in signal length across all operating conditions (see Figure 9).

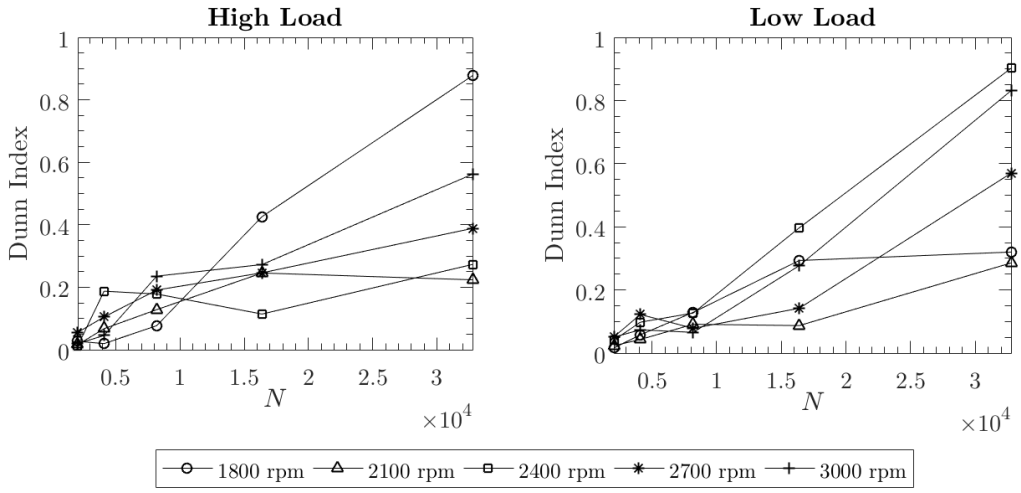

Figure 9: Variation of Dunn Index values with respect to varying signal length. N takes values 2048, 4096, 8192, 16384, and 32768. The plots give Dunn Index values for two Low and High load conditions and different rotational speeds (1800, 2100, 2400, 2700, and 3000 rpm)

5. Conclusion

PHM enables effective preventative maintenance (PM), reliability centered maintenance (RCM), and condition based maintenance (CBM) of mechanical systems. Advancement in sensor technology, machine learning algorithms, and computing technology have contributed to the advancement of PHM. In a PHM framework, both diagnostics and prognostics use data-driven models for fault detection of machine components. In this approach

sensor signals need a good representation that enables accurate detection of faults and health condition of machine components. The results confirm that CECP representation enables accurate classification of different faults in bearings and gears. It is observed that CECP approach is able to handle signal lengths larger than those demonstrated in existing studies. In case of CECP representation, as the signal length increases, so does the Dunn index and SVM accuracy, which quantifies the quality clusters representing different component health conditions in the CECP plane. We observe that the optimum parameter values for CECP depends on the application. For bearing applications we find that embedding dimension $D = 4, 5, 6$, and embedding delay $\tau = 1, 2, 3$ are suitable for fault classification. For gear applications we find that embedding dimension $D = 4, 5$, and embedding delay $\tau = 1, 5$ are suitable for fault classification. For signal length of 16,384 data points, the fault classification accuracy varies from 90% to 100% for bearing applications, and from 85% to 100% for gear applications. Given that CECP representation has only two parameters, not only can it be used for predictive analytics but also for visualization of sensor signals in a 2-dimensional plane. While predictive models can used for optimizing maintenance decisions, visualization can be used for creating dashboards for monitoring health

condition of machine components. From class separability perspective, the CECP method is able to generate linearly separable classes for the classification of different fault states. Beyond separability characteristics, there are several statistical tests that can be performed for permutation entropy and complexity [66, 67]. Users can explore these statistical tests for their applications as needed.

Real manufacturing PHM applications often involve unreliable connectivity in cloud computing, require high bandwidth and cost for transferring data to the cloud, suffer from high latency which is not desirable for closedloop interaction between machine state and actuation, and subject to compliance, regulation, and cyber security constraints. These constraints create a need for localized edge computing, which pushes the intelligence, processing power, and communication capabilities of an edge gateway directly into devices like PACs (programmable automation controllers). Intelligent PACs collect, analyze, and process data from the physical assets they are connected to, at the same time they run the control system program. In such an edge computing environment, a signal representation like CECP is highly desirable because of its compactness, lean computation complexity, and good predictive performance.

Acknowledgement

Funding for this research was provided by the National Institute of Standards and Technology under sponsor award number 70NANB15H028.

[1] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systemsreviews, methodology and applications, Mechanical systems and signal processing 42 (2014) 314–334.

[2] I. El-Thalji, E. Jantunen, A summary of fault modelling and predictive health monitoring of rolling element bearings, Mechanical Systems and Signal Processing 60 (2015) 252–272.

[3] I. El-Thalji, E. Jantunen, A descriptive model of wear evolution in rolling bearings, Engineering Failure Analysis 45 (2014) 204–224.

[4] M. M. Gaber, A. Zaslavsky, S. Krishnaswamy, Mining data streams: a review, ACM Sigmod Record 34 (2005) 18–26.

[5] E. Jantunen, How to diagnose the wear of rolling element bearings based on indirect condition monitoring methods, International Journal of COMADEM 9 (2006) 24.

[6] Z. Zhi-qiang, L. Guo-lu, W. Hai-dou, X. Bin-shi, P. Zhong-yu, Z. Li-na, Investigation of rolling contact fatigue damage process of the coating by acoustics emission and vibration signals, Tribology International 47 (2012) 25–31.

[7] T. Harvey, R. Wood, H. Powrie, Electrostatic wear monitoring of rolling element bearings, Wear 263 (2007) 1492–1501.

[8] T. Yoshioka, S. Shimizu, Monitoring of ball bearing operation under grease lubrication using a new compound diagnostic system detecting vibration and acoustic emission, Tribology Transactions 52 (2009) 725–730.

[9] K. Shibata, A. Takahashi, T. Shirai, Fault diagnosis of rotating machinery through visualisation of sound signals, Mechanical Systems and Signal Processing 14 (2000) 229–241.

[10] K. Mori, N. Kasashima, T. Yoshioka, Y. Ueno, Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals, Wear 195 (1996) 162–168.

[11] A. Parey, M. El Badaoui, F. Guillet, N. Tandon, Dynamic modelling of spur gear pair and application of empirical mode decomposition-based statistical analysis for early detection of localized tooth defect, Journal of sound and vibration 294 (2006) 547–561.

[12] A. Parey, N. Tandon, Impact velocity modelling and signal processing of

spur gear vibration for the estimation of defect size, Mechanical Systems and Signal Processing 21 (2007) 234–243.

[13] W. Wang, Early detection of gear tooth cracking using the resonance demodulation technique, Mechanical Systems and Signal Processing 15 (2001) 887–903.

[14] H. Ma, X. Pang, R. Feng, R. Song, B. Wen, Fault features analysis of cracked gear considering the effects of the extended tooth contact, Engineering Failure Analysis 48 (2015) 105–120.

[15] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mechanical Systems and Signal Processing 21 (2007) 108–124.

[16] F. Combet, L. Gelman, Optimal filtering of gear signals for early damage detection based on the spectral kurtosis, Mechanical Systems and Signal Processing 23 (2009) 652–668.

[17] T. Barszcz, R. B. Randall, Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine, Mechanical Systems and Signal Processing 23 (2009) 1352–1365.

[18] T. Loutas, G. Sotiriades, I. Kalaitzoglou, V. Kostopoulos, Condition monitoring of a single-stage gearbox with artificially induced gear cracks utilizing on-line vibration and acoustic emission measurements, Applied Acoustics 70 (2009) 1148–1159.

[19] Y. Lei, M. J. Zuo, Z. He, Y. Zi, A multidimensional hybrid intelligent method for gear fault diagnosis, Expert Systems with Applications 37 (2010) 1419–1430.

[20] X. Zhang, N. Feng, Y. Wang, Y. Shen, Acoustic emission detection of rail defect based on wavelet transform and shannon entropy, Journal of Sound and Vibration 339 (2015) 419–432.

[21] R. Yan, R. X. Gao, Base wavelet selection for bearing vibration signal analysis, International Journal of Wavelets, Multiresolution and Information Processing 7 (2009) 411–426.

[22] S. M. Pincus, Approximate entropy as a measure of system complexity., Proceedings of the National Academy of Sciences 88 (1991) 2297–2301.

[23] S. Pincus, Approximate entropy (apen) as a complexity measure, Chaos: An Interdisciplinary Journal of Nonlinear Science 5 (1995) 110–117.

[24] S. M. Pincus, Assessing serial irregularity and its implications for health, Annals of the New York Academy of Sciences 954 (2001) 245–267.

[25] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, American Journal of Physiology-Heart and Circulatory Physiology 278 (2000) H2039–H2049.

[26] R. Yan, R. Gao, Machine health diagnosis based on approximate entropy, in: Instrumentation and Measurement Technology Conference, 2004. IMTC 04. Proceedings of the 21st IEEE, volume 3, IEEE, pp. 2054–2059.

[27] R. Yan, R. X. Gao, Approximate entropy as a diagnostic tool for machine health monitoring, Mechanical Systems and Signal Processing 21 (2007) 824–839.

[28] Q. Ni, K. Feng, K. Wang, B. Yang, Y. Wang, A case study of sample entropy analysis to the fault detection of bearing in wind turbine, Case studies in engineering failure analysis 9 (2017) 99–111.

[29] Z. Zhao, S. Yang, Sample entropy-based roller bearing fault diagnosis method, Journal of Vibration and Shock 31 (2012) 136–140.

[30] M. Han, J. Pan, A fault diagnosis method combined with lmd, sample entropy and energy ratio for roller bearings, Measurement 76 (2015) 7–19.

[31] M. Costa, A. L. Goldberger, C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, Physical review letters 89 (2002) 068102.

[32] M. Costa, A. L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Physical review E 71 (2005) 021906.

[33] L. Zhang, G. Xiong, H. Liu, H. Zou, W. Guo, Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference, Expert Systems with Applications 37 (2010) 6077–6085.

[34] K. Zhu, X. Song, D. Xue, A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm, Measurement 47 (2014) 669–675.

[35] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Physical review letters 88 (2002) 174102.

[36] R. Yan, Y. Liu, R. X. Gao, Permutation entropy: a nonlinear statisti-

cal measure for status characterization of rotary machines, Mechanical Systems and Signal Processing 29 (2012) 474–484.

[37] X. Zhang, Y. Liang, J. Zhou, et al., A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized svm, Measurement 69 (2015) 164–179.

[38] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, C.-C. Wang, Bearing fault diagnosis based on multiscale permutation entropy and support vector machine, Entropy 14 (2012) 1343–1356.

[39] Y. Li, M. Xu, Y. Wei, W. Huang, A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree, Measurement 77 (2016) 80–94.

[40] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: Advances in neural information processing systems, pp. 507–514.

[41] R. López-Ruiz, J. Sañudo, E. Romera, X. Calbet, Statistical complexity and fisher-shannon information: Applications, in: Statistical complexity, Springer, 2011, pp. 65–127.

[42] F. Serinaldi, L. Zunino, O. A. Rosso, Complexity–entropy analysis of

daily stream flow time series in the continental united states, Stochastic Environmental Research and Risk Assessment 28 (2014) 1685–1708.

[43] M. Zanin, L. Zunino, O. A. Rosso, D. Papo, Permutation entropy and its main biomedical and econophysics applications: a review, Entropy 14 (2012) 1553–1577.

[44] O. Rosso, H. Larrondo, M. Martin, A. Plastino, M. Fuentes, Distinguishing noise from chaos, Physical review letters 99 (2007) 154102.

[45] L. Zunino, D. Pérez, M. Martín, M. Garavaglia, A. Plastino, O. Rosso, Permutation entropy of fractional brownian motion and fractional gaussian noise, Physics Letters A 372 (2008) 4768–4774.

[46] C. Bandt, F. Shiha, Order patterns in time series, Journal of Time Series Analysis 28 (2007) 646–665.

[47] D. P. Feldman, J. P. Crutchfield, Measures of statistical complexity: Why?, Physics Letters A 238 (1998) 244–252.

[48] R. Lopez-Ruiz, H. L. Mancini, X. Calbet, A statistical measure of complexity, Physics Letters A 209 (1995) 321–326.

[49] P. Lamberti, M. Martin, A. Plastino, O. Rosso, Intensive entropic non-triviality measure, Physica A: Statistical Mechanics and its Applications 334 (2004) 119–131.

[50] H. V. Ribeiro, L. Zunino, R. S. Mendes, E. K. Lenzi, Complexity–entropy causality plane: A useful approach for distinguishing songs, Physica A: Statistical Mechanics and its Applications 391 (2012) 2421–2428.

[51] O. A. Rosso, F. Olivares, L. Zunino, L. De Micco, A. L. Aquino, A. Plastino, H. A. Larrondo, Characterization of chaotic maps using the permutation bandt-pompe probability distribution, The European Physical Journal B 86 (2013) 1–13.

[52] M. Martin, A. Plastino, O. Rosso, Generalized statistical complexity measures: Geometrical and analytical properties, Physica A: Statistical Mechanics and its Applications 369 (2006) 439–462.

[53] A. Kowalski, M. Martín, A. Plastino, O. Rosso, Bandt–pompe approach to the classical-quantum transition, Physica D: Nonlinear Phenomena 233 (2007) 21–31.

[54] O. Rosso, L. Zunino, D. Pérez, A. Figliola, H. Larrondo, M. Garavaglia,

M. Martín, A. Plastino, Extracting features of gaussian self-similar stochastic processes via the bandt-pompe approach, Physical Review E 76 (2007) 061114.

[55] L. Zunino, D. Pérez, A. Kowalski, M. Martín, M. Garavaglia, A. Plastino, O. Rosso, Fractional brownian motion, fractional gaussian noise, and tsallis permutation entropy, Physica A: Statistical Mechanics and its Applications 387 (2008) 6057–6068.

[56] L. Zunino, M. Zanin, B. M. Tabak, D. G. Pérez, O. A. Rosso, Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency, Physica A: Statistical Mechanics and its Applications 389 (2010) 1891–1901.

[57] L. Zunino, A. F. Bariviera, M. B. Guercio, L. B. Martinez, O. A. Rosso, On the efficiency of sovereign bond markets, Physica A: Statistical Mechanics and its Applications 391 (2012) 4342–4349.

[58] L. Zunino, B. M. Tabak, F. Serinaldi, M. Zanin, D. G. Pérez, O. A. Rosso, Commodity predictability analysis with a permutation information theory approach, Physica A: Statistical Mechanics and its Applications 390 (2011) 876–890.

[59] C. S. Thaxton, W. Anderson, C. Gu, B. Stosic, T. Stosic, Detrended fluctuation analysis and entropy-complexity causality analysis of temperatures in an urbanized mountain stream, Stochastic environmental research and risk assessment 32 (2018) 843–858.

[60] MSFT bearing dataset, `http://mfpt.org/fault-data-sets/`, 2013. Accessed: 2016-09-20.

[61] X. Zhang, J. Kang, E. Bechhoefer, H. Teng, Enhanced bearing fault detection and degradation analysis based on narrowband interference cancellation, International Journal of System Assurance Engineering and Management 5 (2014) 645–650.

[62] J. C. Dunn, Well-separated clusters and optimal fuzzy partitions, Journal of cybernetics 4 (1974) 95–104.

[63] Cwru bearing dataset, http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website, 2015. Accessed: 2018-03-20.

[64] W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the case western reserve university data: A benchmark study, Mechanical Systems and Signal Processing 64 (2015) 100–131.

[65] Phm gear dataset, http://www.phmsociety.org/competition/PHM/09/apparatus, 2009. Accessed: 2016-09-20.

[66] F. Traversaro, F. O. Redelico, Confidence intervals and hypothesis testing for the permutation entropy with an application to epilepsy, Communications in Nonlinear Science and Numerical Simulation 57 (2018) 388–401.

[67] F. Pennekamp, A. C. Iles, J. Garland, G. Brennan, U. Brose, U. Gaedke, U. Jacob, P. Kratina, B. Matthews, S. Munch, et al., The intrinsic predictability of ecological time series and its potential to guide forecasting, Ecological Monographs (2019) e01359.