**Forecasting Bearing Fault Signature diagnosis**

**ABSTRACT**

Bearing fault detection still remains a very challenging task especially when defects occur on rotating bearing components because the fault-related features could be non stationary in nature. In this paper, the recent development of bearing fault detection and the challenges facing reliable bearing health condition monitoring will be discussed. Specifically, the paper will discuss the bearing characteristic frequency analysis, denoising to improve the signal-to-noise ratio, and advanced signal processing techniques for nonstationary signal analysis and bearing fault detection.

As it is difficult to extract the fault characteristic frequency caused by nonlinear and nonstationary features of the rolling bearing fault signal, this paper presents a fault feature extraction method of rolling bearing based on Nonlinear Autoregressive Neural Network (NARNET) and wavelet threshold denoising method. First of all, the fault signal was pre-processed via wavelet threshold denoising. Case study results show that the proposed method provides accurate failure predictions across several system failures, and that the training approach can significantly reduce the time necessary to generate an effective, generalized model. The results indicate very promising performance in identifying various faults with virtually perfect accuracy, recall, and precision. In addition, the results demonstrate an outstanding prediction rate for the fault diameter of bearing defects.

**INTRODUCTION**

Rolling element bearings find widespread domestic and industrial application. It is one of the most widely used elements in rotating machinery. As a critical component, it carries most of the load during the running of rotating machinery. The faults arising in motors are often linked with bearing faults. If rolling bearing fails, serious problems arise, which will in turn result in the decrease of production efficiency and large economic loss. Defects in bearing unless detected in time may lead to malfunctioning of the machinery. Fault detection is the process of observing the measured system data and system status information and comparing them with a normal range of observed attributes to determine whether some measurements fall outside the range representing the healthy condition of the system.

In many instances, the accuracy of the instruments and devices used to monitor and control the motor system is highly dependent on the dynamic performance of bearings. Records show that faulty bearings contribute to about thirty percent of failures in rotating machinery. Thus, it is of great importance to study the effective fault diagnosis approaches for rolling bearings. When the machine is operated under the normal condition the vibration is small and constant but when fault is developed, the dynamic process of machine changes,

which alters the vibration pattern. The faulty vibration signal and normal vibration can be compared and analysed to determine the sign of failure. In practice, such comparisons are not effective with unprocessed (raw) data due to large variation in time series data, in such cases features extraction technique provides considerably good result as compare to the data itself. Extracted features are more stable and well behaved and they also provide a reduced data set.

Bearing vibration can generate noise and degrade the quality of a product line which is driven by a motor system. Heavy bearing vibration can even cause the entire motor system to function incorrectly, resulting in downtime for the system and economic loss to the customer. Proper monitoring of bearing vibration levels in a motor system is highly cost effective in minimizing maintenance downtime both by providing advance warning and lead time to prepare appropriate corrective actions, and by ensuring that the system does not deteriorate to a condition where emergency action is required. Thus, it is important to include bearing vibration diagnosis into the scheme of motor system fault diagnosis.

**RESEARCH OBJECTIVE**

This thesis is a report of a study of techniques for forecasting fault detection and signature bearing elements based narnet. The primary objective of this research thesis is to identify the fault in the bearings and wavelet denoising. The secondary objective is to extract the statistical features.

Some techniques include (Skewness, Kurtosis, Mean, Variance, Energy, Entropy). Because each of the presented techniques has shown to be successful in extracting features hidden within the signals, a proportional evaluation of these techniques provides a useful direction to select the most effective technique for the fault detection of the bearing. Thus, this thesis investigates these representative techniques in detail to diagnose faulty bearings. These will detect and predict the location of the fault.

The final objective is to evaluate a classification methodology to determine the bearing as healthy or defective and identify the type of defect(s) as well as the level of the defect severity. This classification was achieved to some extent by appropriate feature extraction from time-series signals and their respective channels.

**SOCIAL IMPACT**

Bearings have a high load carrying capacity and can operate under extreme conditions of performance and speed. Another advantage of bearings is that they have little operational wear, require simple methods of lubrication and are “inherently precision mechanisms.”

An early detection of an initial fault avoids difficult consequences and reduces financial loss, bringing about only short downtime for the working process. In fact, both correct diagnosis and early detection of incipient faults lead to fast unscheduled maintenance and short downtime for the process under consideration. They also prevent the harmful and sometimes devastating consequences of faults and failures. Generally, failure prevention can be identified as the process of fault detection, diagnosis, and prognosis. If a fault is detected, repairs are made fast and to restore full protective functionality.

Bearings are among the most critical mechanical components that have wide applications in many industries and have proven to be reliable and long-lived when properly applied. As a result of improvements in bearing materials, design, lubrication technology and service life, they have been gradually employed under more severe application requirements such as higher load, higher speed, and restricted lubrication. These requirements have made condition monitoring and fault diagnosis of bearings very important to ensure safe operation of rotary machines. Bearing failure prevention through health monitoring have been one the most researched areas in the past decades in the field of mechanical engineering. These research activities have resulted in a better understanding of bearing health analysis as well as advanced diagnostic methodologies.

As one of the most important fault sources of mechanical equipment, any fault of the bearing will seriously affect the performance of the entire machine. If the fault cannot be discovered and diagnosed in time, it will cause serious personal injury and unnecessary economic loss. Therefore, monitoring the running condition of the rolling bearing and finding out its early failure in time is of great significance to the safety of its operation. Condition monitoring of rolling element bearings has enabled cost saving of over 50% as compared with the old traditional methods.

**LITERATURE SURVEY**

1. International Journal of Engineering, Science and Technology Vol. 11, No. 2, 2019, pp. 33-47 Performance and predict the ball bearing faults using wavelet packet decomposition and ANFIS

Arun R. Pathiran1 , K. Erikiananda 2\* , T. Getachew 3 and Haftom G. Gziabher4

In this paper, a fault identification method has been proposed to identify the type of ball bearing fault. The proposed method is the combination of wavelet decomposition method, Principal component analysis and ANFIS or SVM. ANN and fuzzy based algorithm has been popularly used in the fault identification process. However, the use of SVM for machine condition monitoring and fault diagnosis is still rare. In this work, both conventional ANFIS and SVM techniques are implemented. It is found that SVM technique has higher accuracy compared to ANFIS. The proposed method has been simulated and validated using the vibration data collected from the ball bearing setup. The simulation result shows that the proposed method exactly predicts the type of fault and it has very good accuracy. Some improved version of SVM and ANFIS algorithm can be used to further improve accuracy of prediction is the future work. The proposed method may be extend to fault detection in other mechanical rotating elements.

1. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019- Identification of Bearing Faults using Wavelet Transform

Ajay Sharma, Prem Narayan Vishwakarma

The bearing fault simulator is an essential tool for analysis, development for vibration. Traditional time domain analysis or frequency domain approaches dictate the compromise of frequency resolution, but wavelet transform have no such limitations and therefore more suitable for the processing of vibration signature. In this study, wavelet packed decomposition (WPD) that is one of the many wavelet transforms, because it is giving a good difference among all the bearings and enables us for getting optimum results.

1. MATEC Web of Conferences 255, 06005 (2019) -EAAI Conference 2018- An intelligent bearing fault diagnosis system: A review S.R.Saufi1,\* , Z.A.B Ahmad1 ,M.S Leong1 and M.H Lim1

In this paper, a machine learning model that was used on bearing fault diagnosis has been reviewed. Among the models, SVM achieved the highest usage in bearing fault diagnosis. However, the future trend shows that deep learning will receive a lot of attention by researcher due to its capability of providing automated feature extraction and feature selection.

4. An Effective Hybrid NARX-LSTM Model for Point and Interval PV Power Forecasting

This paper proposes an effective Photovoltaic (PV) Power Forecasting (PVPF) technique based on hierarchical learning combining Nonlinear Auto-Regressive Neural Networks with exogenous input (NARXNN) with Long Short-Term Memory (LSTM) model. First, the NARXNN model acquires the data to generate a residual error vector. Then, the stacked LSTM model, optimized by Tabu search algorithm, uses the residual error correction associated with the original data to produce a point and interval PVPF. The performance of the proposed PVPF technique was investigated using two real datasets with different scales and locations. The proposed NARX-LSTM technique has the following major achievements: 1) Improves the prediction performance of the original LSTM and NARXNN models; 2) Evaluates the uncertainties associated with point forecasts with high accuracy; 3) Provides a high generalization capability for PV systems with different scales. Numerical results of the comparison of the proposed NARX-LSTM method with two real-world PV systems in Australia and USA demonstrate its improved prediction accuracy, outperforming the benchmark approaches with an overall normalized Rooted Mean Squared Error (nRMSE) of 1.98% and 1.33% respectively.

5.Non-linear autoregressive neural network (NARNET) with SSA filtering for a university energy consumption forecast

This paper uses a non-linear autoregressive neural network (NARNET) for energy consumption forecast in a South African University with four campuses, using three-year daily energy consumption data. Singular Spectrum Analysis (SSA) technique was used for the data filtering. The study demonstrates the significance of data filtering in forecasting univariate autoregressive series**.** Singular spectrum analysis used in this study has proved to be a good tool for filtering data with noise. This study used three values of window length obtained from periodogram analysis (L=54, 103, 155). Window length of L=103 gave a more accurate result. Prediction accuracy in campus A model is highest (85.87%), and the lowest being campus B (75.62%). Up until now, the choice of window length for a series YYTT, still remains open. However, from this study, window length of TT 10< LL < TT 11 gave a better filtering. In using SSA for filtering time series data, it is recommended that different values of window length be chosen based on known metrics, one of which is periodogram analysis. Data filtering before forecast enhances data quality, most especially when the forecast is a step ahead using the same series as input, i.e. an autoregressive model.

6. AR-Net: A SIMPLE AUTO-REGRESSIVE NEURAL NETWORK

In this paper, we propose a AR-Net model, which uses stochastic gradient descent to estimate dynamics imposed by auto-regression. AR-Net makes it possible to learn a high order p model orders of magnitude faster than using least squares. We show that the resulting weights are as interpretable as those of AR. Further, by adding regularization, AR-Net reliably selects and learns sparse weights, even up to a sparsity of 3 : 1000. This eliminates the need to know the exact order of the AR-process and makes it possible to learn long-range dependencies on granular data without overfitting. We found the sparse model to be insensitive to the estimated sparsity s for estimates up to one magnitude off. However, as the model is trained with SGD, it is sensitive to learning rate and related hyperparameters. We hope to ease this sensitivity with a smart learning rate schedule such as the 1cycle-policy .In future work, we will demonstrate how AR-Net makes it possible to seamlessly include co-variate time-series and to expand the forecast horizon, all with the same model. This makes it far simpler for the practitioner to expand their analysis from univariate one-step forecasting to multivariate multi-step forecasting. Another part of our future work will be to extend AR-Net to have an MA component and eventually include further temporal components (e.g., custom trend or seasonality). Our long-term vision is to enable the practitioner with a simple but powerful time-series tool powered by neural networks.

**2.WAVELET DENOISING**

Wavelet-based denoising is a method of analysis that uses time-frequency to select an appropriate frequency band based on the characteristics of the signal. A signal describes various physical quantities over time. While noise is an unwanted signal which interferes with the signal carrying the original message. This causes a change in the parameters of the signal message. Denoising is the process of removing noise from the signal.In wavelet denoising, the thresholding algorithm is usually used in orthogonal decompositions: multi-resolution analysis and wavelet packet transform. Wavelet thresholding faces some questions in its application, for example, the selection of hard or soft threshold, fixed or level-dependent threshold. Proper selection of those items helps generating a better estimation. Wavelet analysis can be applied in daily life activities such as feature extraction, face recognition, data analysis and prediction, voice recognition, numerical analysis, and many more. The basic idea behind wavelet denoising, or wavelet thresholding, is that the wavelet transform leads to a sparse representation for many real-world signals and images. What this means is that the wavelet transform concentrates signal and image features in a few large-magnitude wavelet coefficients. Wavelet coefficients which are small in value are typically noise and you can "shrink" those coefficients or remove them without affecting the signal or image quality. After you threshold the coefficients, you reconstruct the data using the inverse wavelet transform.

2.1 Time Series Denoising Using Wavelet Transform

Conceptually, wavelet coefficients correspond to details and when details are small, they could be considered as noise and therefore omitted without compromising the sharp detail of the original series (unlike the Fourier transform). In other words, without going through a formal mathematical argument, wavelet transform compresses most of the energy of the original series into a small number of large wavelet coefficients. Thus, the few wavelet coefficients representing the original series stick up above the noise. Therefore, the thresholding has the effect that it kills the noise while not killing the series. This is the beauty of wavelet transform. Hence, the idea behind wavelet denoising is to threshold the wavelet coefficients at every multiresolution level (in some manner) to clean out unimportant details considered to be noise.

2.2Advantages of wavelet theory

One of the main advantages of wavelets is that they allow complex information such as images to be decomposed into elementary forms at different positions and scales and subsequently reconstructed with high precision. The second main advantage of wavelets is that using fast wavelet transform based on filter banks, it is computationally efficient. Wavelet transform provides sparse representation for a large class of signals , and it is capable of revealing aspects of data that other signal analysis techniques miss the aspects like trends, breakdown points, and discontinuities in higher derivatives and self-similarity. Wavelets have the great advantage of being able to capture the energy of a signal in few energy transform values, it does not change the number of pixels required to represent the image and separate the information in a way that resembles the human visual system.

Wavelet denoising

In the context of wavelets, "denoising" means reducing the noise as much as possible without distorting the signal. Denoising makes use of the time-frequency-amplitude matrix created by the wavelet transform. It's based on the assumption that the undesired noise will be separated from the desired signal by their frequency ranges. Most commonly in scientific measurements, the desired signal components are located at relatively low frequencies and the noise is mostly at high frequencies. The process is controlled both by the selection of wavelet type and by a positive integer number called the wavelet "level"; the higher the level, the lower is the frequency divider between signal and noise. (To that extent, the wavelet level is similar to the effect of the smooth width of a smoothing operation). Again, Matlab's Wavelet Toolbox provides some useful tools. First, there is the GUI app called the "Wavelet Signal Denoiser". The selection of the wavelet type and level are all selectable manually in the Wavelet Signal Denoiser app. I used that app to analyze the "buried peaks" signal as previously, using the "sym4" wavelet at a relatively high level of 11, because lower levels allow too much of the interfering swept sine wave to come through and higher levels would damp out the Gaussian peaks too much. The "Approximation" result (the dotted line) is the low-frequency information in the data, and you can clearly see that this is a "denoised" version of the original signal (shown in blue). The two bumps at sample numbers 5000 and 10000 are the two Gaussian peaks.

WAVELET TRANSFORM

A wavelet transform (WT) is a decomposition of a signal into a set of basis functions consisting of contractions, expansions, and translations of a wavelet function. It can be computed by repeated convolution of the signal with the chosen wavelet as the wavelet is translated across the time dimension, in order to probe the time variation, and as the wavelet is stretched or compressed, in order to probe different frequencies. Because two dimensions are being probed, the result is naturally a 3D surface (time-frequency-amplitude) that can be conveniently displayed as a time-frequency contour plot with different colors representing the amplitudes at that time and frequency. Of course, one expects that such calculations will require more complex algorithms and greater execution times. That might have been a problem in the early days of computers, but with modern fast processors and great memory capacity, it's unlikely to be a problem.

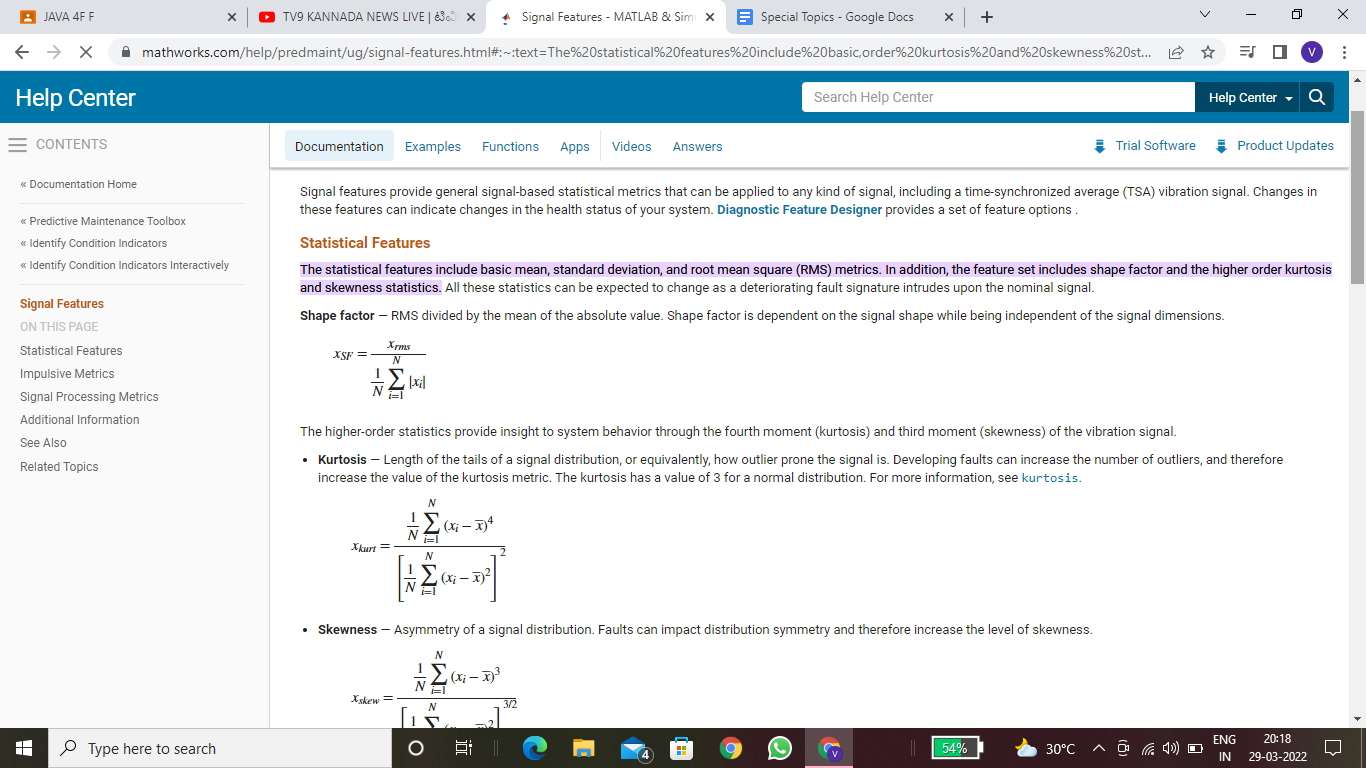
## Signal Features

Signal features provide general signal-based statistical metrics that can be applied to any kind of signal, including a time-synchronized average (TSA) vibration signal. Changes in these features can indicate changes in the health status of your system.

### **Statistical Features**

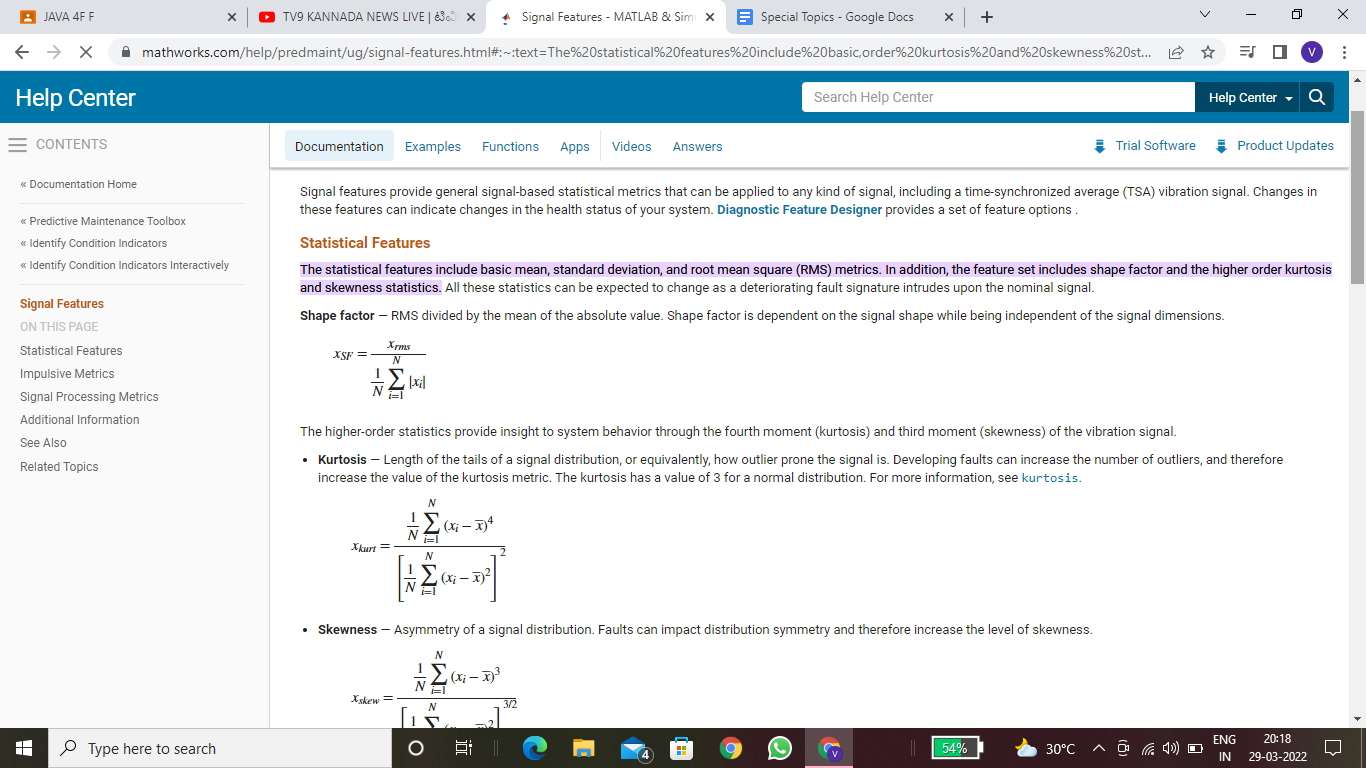
**The statistical features include basic mean, standard deviation, and root mean square (RMS) metrics. In addition, the feature set includes shape factor and the higher order kurtosis and skewness statistics. All these statistics can be expected to change as a deteriorating fault signature intrudes upon the nominal signal.**

**Shape factor — RMS divided by the mean of the absolute value. Shape factor is dependent on the signal shape while being independent of the signal dimensions.**

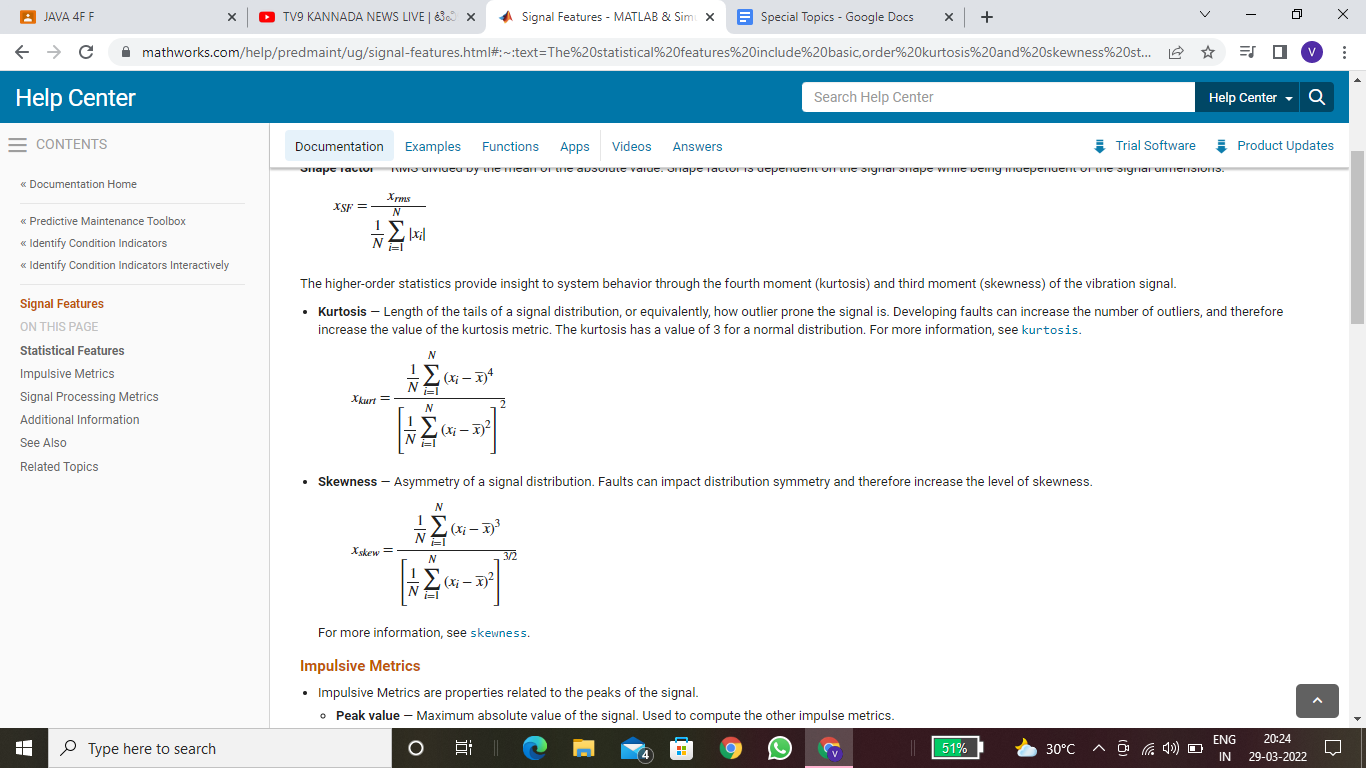
****

**The higher-order statistics provide insight to system behavior through the fourth moment (kurtosis) and third moment (skewness) of the vibration signal.**

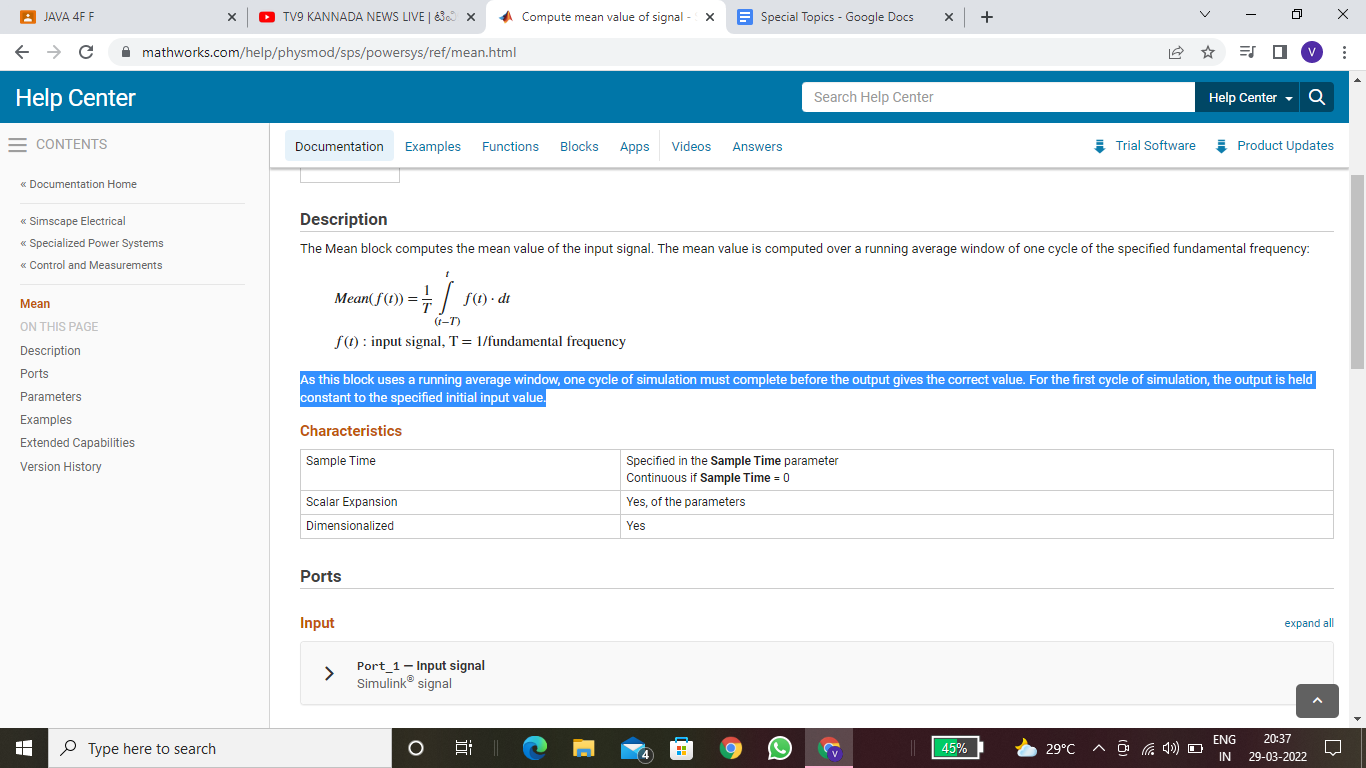
* **Kurtosis — Length of the tails of a signal distribution, or equivalently, how outlier prone the signal is. Developing faults can increase the number of outliers, and therefore increase the value of the kurtosis metric. The kurtosis has a value of 3 for a normal distribution.**

****

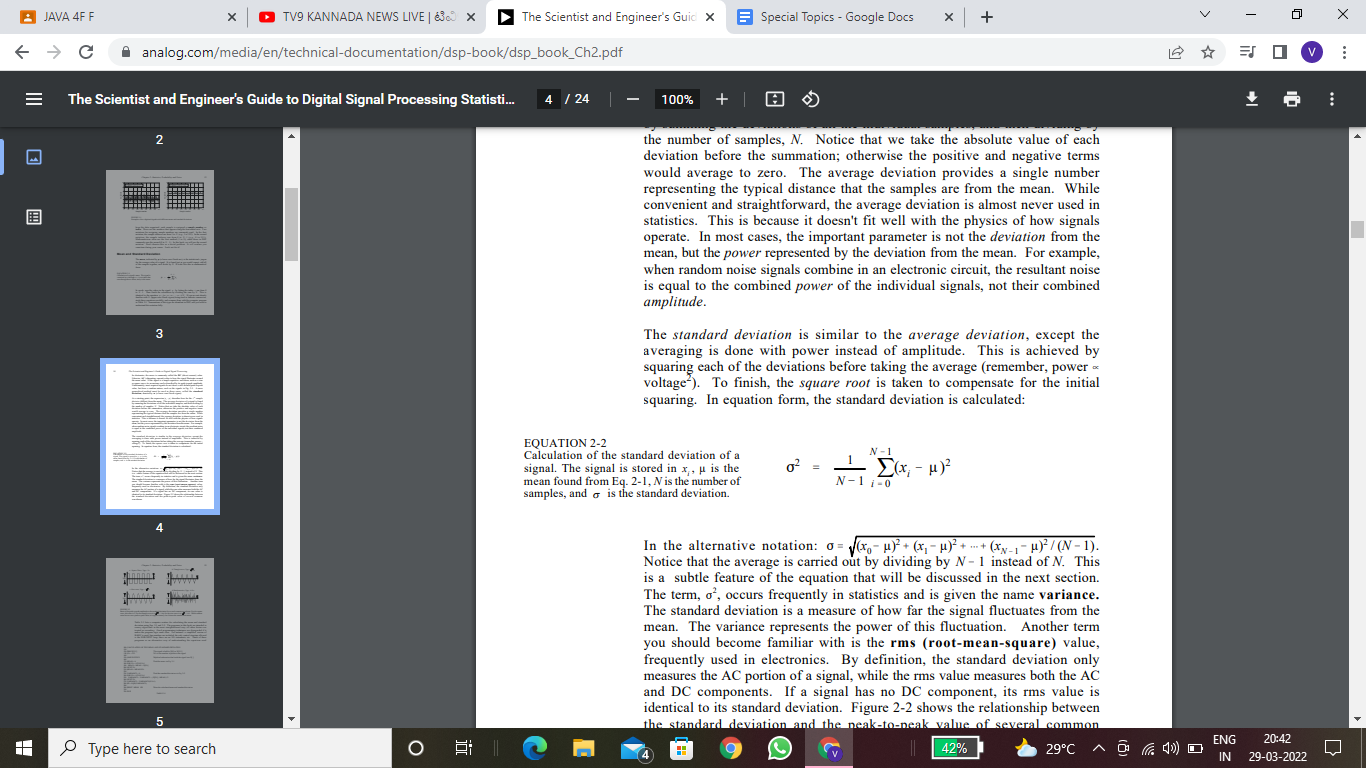
* **Skewness — Asymmetry of a signal distribution. Faults can impact distribution symmetry and therefore increase the level of skewness.**

****

* **Mean – The Mean block computes the mean value of the input signal. The mean value is computed over a running average window of one cycle of the specified fundamental frequency:**

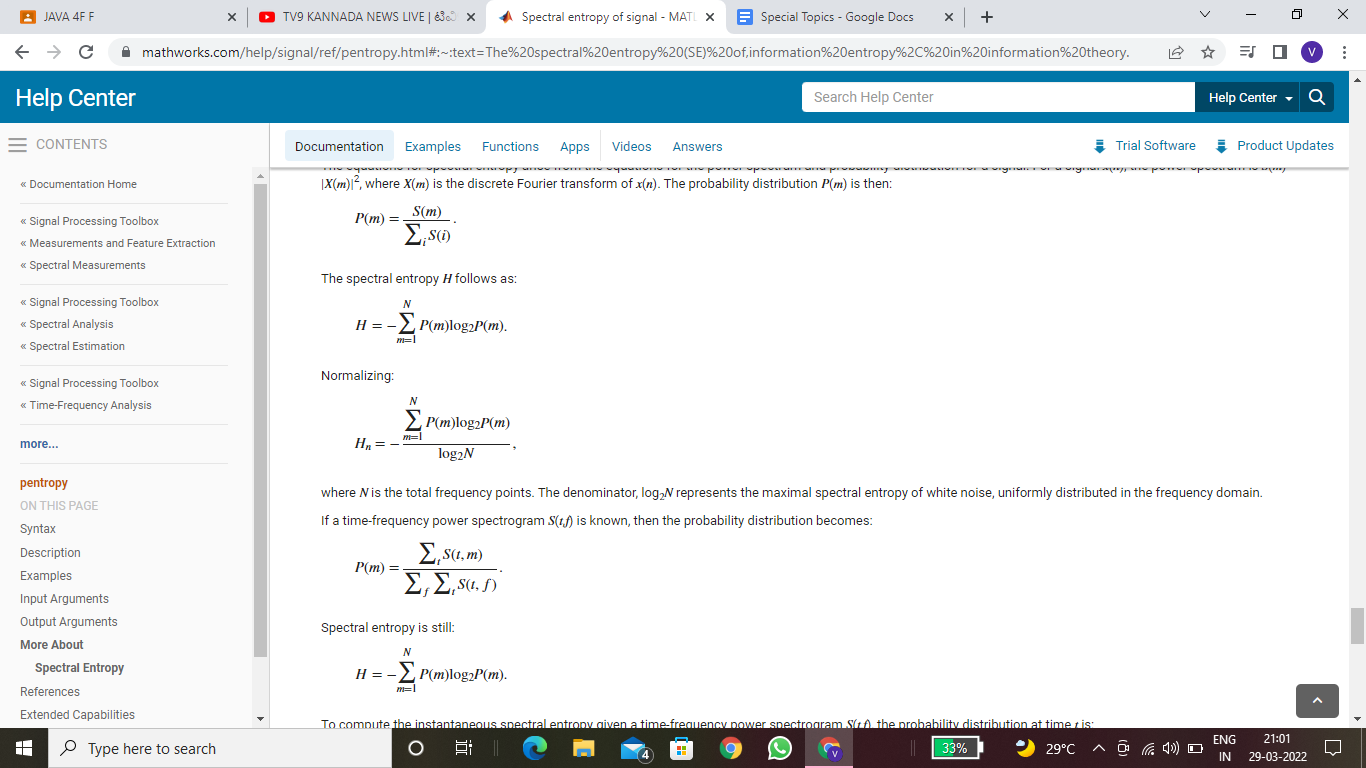
****

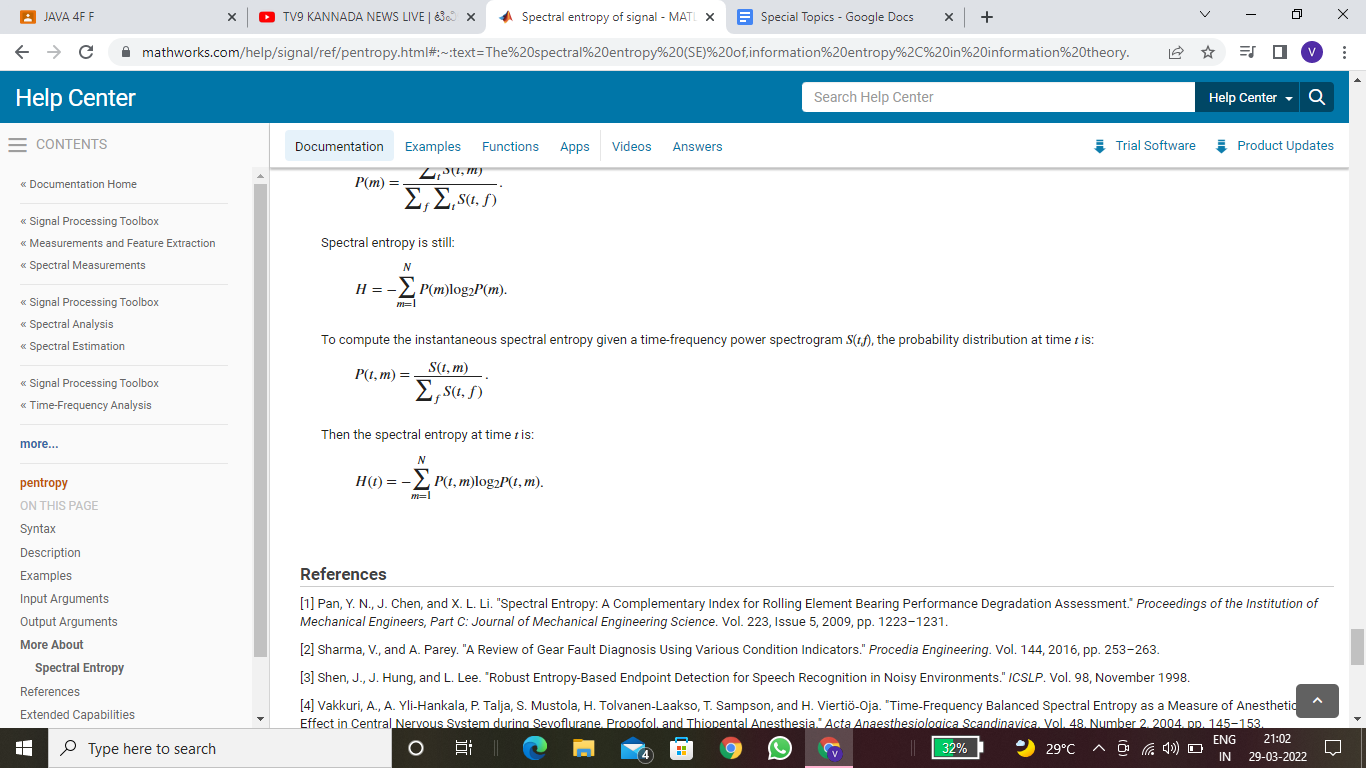
**As this block uses a running average window, one cycle of simulation must complete before the output gives the correct value. For the first cycle of simulation, the output is held constant to the specified initial input value.**

* **Standard Deviation – The standard deviation is similar to the average deviation, except the averaging is done with power instead of amplitude. This is achieved by squaring each of the deviations before taking the average. To finish, the square root is taken to compensate for the initial squaring. In equation form, the standard deviation is calculated as:**
* **Variance – The Variance block computes the unbiased variance of each row or column of the input, or along vectors of a specified dimension of the input. It can also compute the variance of the entire input. You can specify the dimension using the Find the variance value over parameter. The Variance block can also track the variance in a sequence of inputs over a period of time.**

**y = var(X,W) computes the variance using the weight vector W . The length of W must equal the length of the dimension over which var operates, and its elements must be nonnegative. var normalizes W to sum to 1 . Use a value of 0 for W to use the default normalization by N – 1 , or use a value of 1 to use N .**

* **Entropy – The *spectral entropy* (SE) of a signal is a measure of its spectral power distribution. The concept is based on the Shannon entropy, or information entropy, in information theory. The SE treats the signal's normalized power distribution in the frequency domain as a probability distribution, and calculates the Shannon entropy of it. The Shannon entropy in this context is the spectral entropy of the signal. This property can be useful for feature extraction in fault detection and diagnosis . SE is also widely used as a feature in speech recognition and biomedical signal processing.**
* **The equations for spectral entropy arise from the equations for the power spectrum and probability distribution for a signal. For a signal *x*(*n*), the power spectrum is *S*(*m*) = |*X*(*m*)|2, where *X*(*m*) is the discrete Fourier transform of *x*(*n*). The probability distribution *P*(*m*) is then:**

****

****

* **Entropy – If the signal energy over one period is larger than zero but finite, then the total energy is infinite and the signal power is finite. Therefore, the signal is a power signal. If the signal energy in one period is infinite, then both the power and the total energy are infinite.Entropy of a signal is given by:**

**f = matlabFunction(f);**

**y(N)=symsum(f, -N , N);**

**energy=limit(y(N),N,inf);**

**z(N)=y(N)/(2\*N+1);**

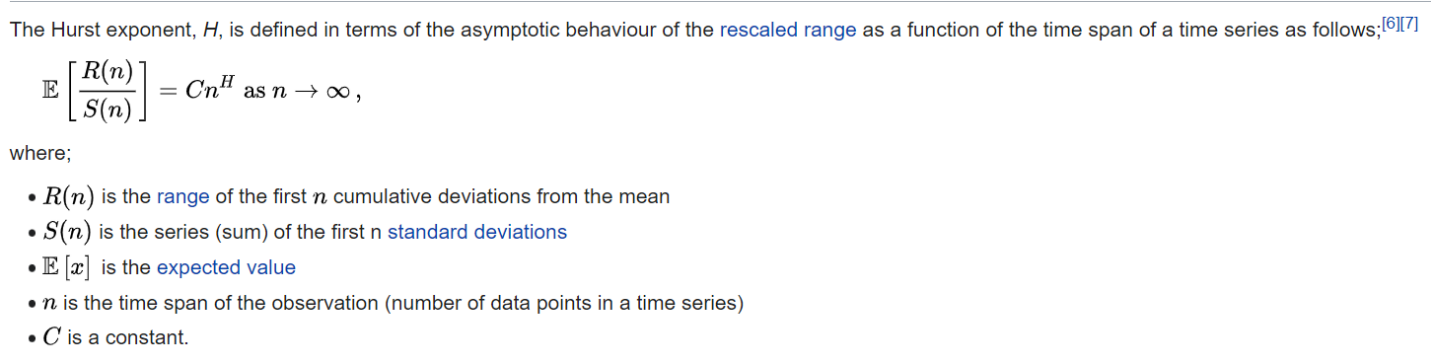
* **RMS – RMS is the root-mean-square value of a signal. For a digitised signal, you can calculate it by squaring each value, finding the arithmetic mean of those squared values, and taking the square root of the result.**

[**y**](https://in.mathworks.com/help/matlab/ref/rms.html#d123e1280555) **= rms(**[**x**](https://in.mathworks.com/help/matlab/ref/rms.html#d123e1280377)**) returns the root-mean-square (RMS) value of the input, x.**

* **If x is a row or column vector, then y is a real-valued scalar.**
* **If x is a matrix, then y is a row vector containing the RMS value for each column.**
* **If x is a multidimensional array, then y contains the RMS values computed along the first array dimension of size greater than 1. The size of this dimension is 1 while the sizes of all other dimensions remain the same as x.**

**Non-Statistical Features**

* **Hurst Exponent-** **The Hurst exponent is used as a measure of long-term memory of time series. It relates to the autocorrelations of the time series, and the rate at which these decrease as the lag between pairs of values increases.**



* **Correlation Dimension-The correlation dimension (denoted by ν) is a measure of the dimensionality of the space occupied by a set of random points, often referred to as a type of fractal dimension.**

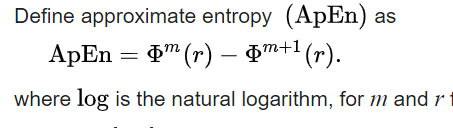


**where g is the total number of pairs of points which have a distance between them that is less than distance ε (a graphical representation of such close pairs is the recurrence plot).**

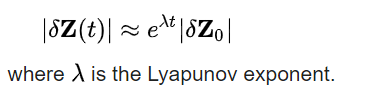
**As the number of points tends to infinity, and the distance between them tends to zero, the correlation integral, for small values of ε, will take the form:**



* **Approximate Entropy-In statistics, an approximate entropy (ApEn) is a technique used to quantify the amount of regularity and the unpredictability of fluctuations over time-series data.**



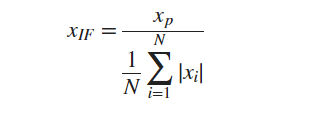
* **Lyapunov exponent-In mathematics, the Lyapunov exponent or Lyapunov characteristic exponent of a dynamical system is a quantity that characterizes the rate of separation of infinitesimally close trajectories.**



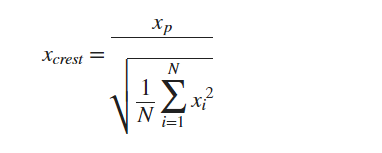
* **Fractal dimension-** **In mathematics, more specifically in fractal geometry, a fractal dimension is a ratio providing a statistical index of complexity comparing how detail in a pattern (strictly speaking, a fractal pattern) changes with the scale at which it is measured.**
* **Shape factor-Shape factor refers to a value that is affected by an object's shape but is independent of its dimensions. It may refer to one of number of values in physics, engineering, image analysis, or statistics.**

**SF = (P2/A).**

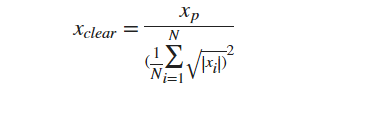
* **Impulse Factor - Compare the height of a peak to the mean level of the signal.**



* **Crest Factor — Peak value divided by the RMS. Faults often first manifest themselves in changes in the peakiness of a signal before they manifest in the energy represented by the signal root mean squared. The crest factor can provide an early warning for faults when they first develop. For more information, see peak2rms.**

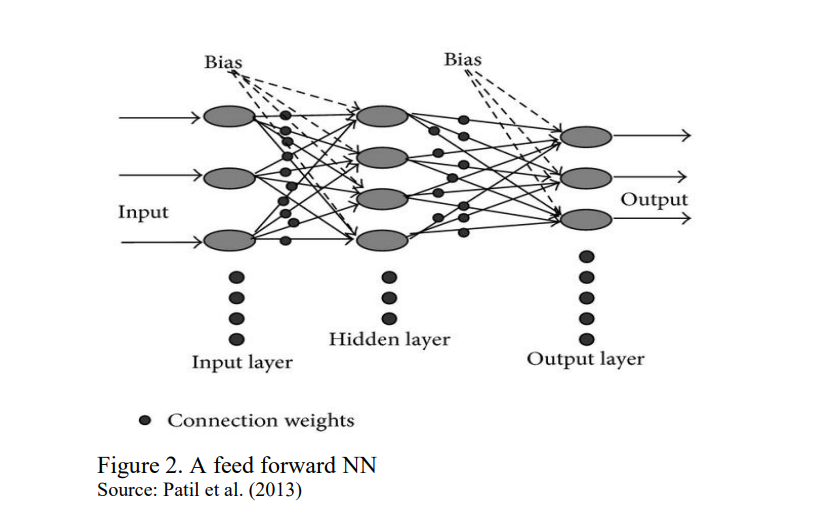


* **Clearance Factor — Peak value divided by the squared mean value of the square roots of the absolute amplitudes. For rotating machinery, this feature is maximum for healthy bearings and goes on decreasing for defective ball, defective outer race, and defective inner race respectively. The clearance factor has the highest separation ability for defective inner race faults.**



**NARNET**

NARNET stands for the non-linear autoregressive dynamic network, and it is suitable for forecasting financial instruments, without the use of companion time series. NARNET can be trained to forecast a time series using past values and it is therefore employed here to predict BPI, given the availability of this index’s past values. NARNET is a recurrent dynamic network with feedback arrangements. Output is fed back to the input of the feedforward network. Consequently, the feedback of the true output is used, instead of the estimated output, which makes the feedforward architecture more accurate (Patil et al., 2013). The classic architecture of a feedforward network s presented in Figure 1.



The NARNET can be written as in Eq (2), where the future values of the time series 𝑦(𝑡) are predicted from its past values 𝑦(𝑡 − 𝑖) (𝑖 = 1, 2, … 𝑑) (in this case, BPI data). The output of the NAR network is fed back to the input of the network, through delays.

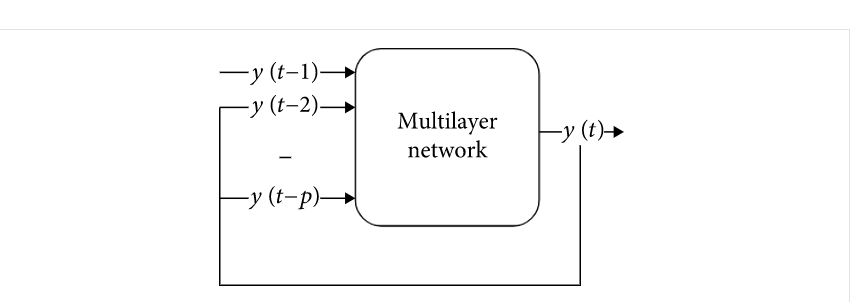
𝑦(𝑡) = 𝑓(𝑦(𝑡 − 1), … , 𝑦(𝑡 − 𝑑)) (2)

Network learning takes place as the weights are adjusted along the layers, according to the relationship between the inputs and the desired outputs. One of the most basic models is multilayer rerceptron (MLP) network, which is widely used in the approximation of nonlinear functions that describe complex relationships between independent and dependent variables in many applications.

Multilayer perceptron (MLP) was first introduced to solve complex classification problems. But because of their universal approximation property [9], they were quickly used as nonlinear regression models and then for time series modelling and forecasting.

However, the estimation and identification of these models use sophisticated techniques and it is not easy to determine the correct architecture. Indeed, these models are by definition overparametrized, the error functions to be minimized have many local minima, and the implementation is often difficult.

The nonlinear autoregressive neural network (NAR) as shown in Figure [1](https://www.hindawi.com/journals/jam/2020/5057801/fig1/) can be trained to predict a time series from that series past values 𝑦(𝑡) = 𝑓(𝑦(𝑡 − 1), … , 𝑦(𝑡 − 𝑑)) called feedback delays, with  is the time delay parameter.



**Figure 1**

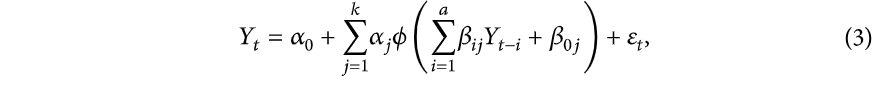
Architecture of the nonlinear autoregressive neural network.

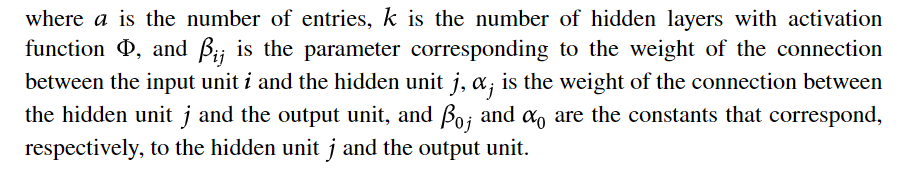
The network is created and trained in an open loop, using the real target values as a response and making sure of greater quality being very close to the true number in training. After training, the network is converted into a closed loop and the predicted values are used to supply new response inputs to the network. A nonlinear autoregressive neural network applied to time series forecasting, describe a discrete, nonlinear autoregressive model that can be written in this form:



The function h(.)  is unknown in advance, and the training of the neural network is aimed at approximating the function by means of the optimization of the network weights and neuron bias.

So a model (NAR) is defined precisely by an equation of the type





The optimization of the architecture is aimed at reducing as much as possible the number of synapses (weights) and neurons in order to reduce the complexity of the network, improve computing times, and maintain the generalization capabilities. Concerning the optimization of the network architecture, two main approaches have been proposed in the literature:

•Selection approach: consists of starting with the construction of a complex network that contains a large number of neurons, then this approach is to try to reduce the number of unnecessary neurons and remove redundant connections during or at the end of learning

•Incremental approach: we start with the simplest possible network, then we add neurons or layers, until we have an optimal architecture

An effective approach is to estimate the prediction error using a set of data that was not used to construct the predictor, i.e., not used for learning. This dataset is called a test set.

Divide the dataset into three kinds of target timesteps as follows:

•training: these datasets are presented to the network during training and the network is adjusted according to its error

•Validation: these datasets are used to measure network generalization and to halt training when generalization stops improving

•Testing: these datasets have no effect on training and so provide an independent measure of network performance during and after training.

References:

<https://ieeexplore.ieee.org/document/8560579>

<https://lost-contact.mit.edu/afs/inf.ed.ac.uk/group/teaching/matlab-help/R2016b/nnet/ref/narnet.html>

<https://www.researchgate.net/publication/333450728_Non-Linear_Autoregressive_Neural_Network_NARNET_with_SSA_filtering_for_a_university_Campus_Energy_Consumption_Forecast>