Subsea pump vibration analysis for predictive maintenance

Mahdi Rouzbahaneh

Master Thesis in

Applied and Computational Mathematics

Professor Hans Z. Munthe-Kaas

Department of Mathematics

University of Bergen

Norway

November 2015

Acknowledgment

I would like to express my sincere gratitude to my supervisor Prof. Hans Z. Munthe-Kaas for his valuable advice and patient guidance during my studies. I extend special thanks to Oddvar Christiansen from NAXYS Company for his helpful assistance and providing historical information. I would like to take this opportunity to thank all of the members of the department of Mathematics especially Kristine Lysnes for her helpful assistance.

Special thanks go to my friends Sadaf Ghorbani, Novin Balafkan and Amin Nematbakhsh and other friends for their kind help and friendship. I am deeply indebted to my family and my family-in-law for their support, encouragement, dedication, patience and love.

My final words go to my wife. I would like to express my special gratitude to my wife, Zahra, who supported and motivated me during my studies.

Mahdi Rouzbahaneh
Bergen, November 2015.

Contents

Summary and Conclusions

The goal of this study is to compare vibration signals of subsea pumps from two different time periods. These signals are gathered for predictive maintenance reasons.

Vibration analysis is one of the most efficient tools for machine diagnostics. Vibration analysis has been an active area of research in the past few decades. In this thesis an attempt has been made to cover all the common vibration analysis methods to find more reliable results.

Since a pump includes many rotating components, presence of a kind of periodicity is expected in the vibration signal. Local defects in each cycle of rotation cause an impulse in the vibration signal. Therefore high impulsivity in the vibration signal is an indicator of fault.

This Impulsivity is not clearly observed in the vibration signal. The impulse response function of the transmission path deforms an impulse in the source into a spread segment in the vibration signal. Furthermore the impulses are more hidden when the period of impulses is shorter than the width of the impulse response function. Detection of impluses is more subtle due to the presence of background noise.

Our methods mainly focus on these two obstacles to achieve impulsivity of source signal. These Impulses also appear as amplitude modulation on the high resonant frequency of the pump. In general, there is low signal-to-noise ratio in low frequencies. Therefore we look into demodulation in high frequencies to find impulses.

Several methods were applied on the signals object to 1) Separation of deterministic and non-deterministic parts of signal. 2) Separation of source signal and transmission path function. 3) Determine the optimal resonant frequency band and envelope analysis. In addition a new method to estimate the shaft speed from the vibration signal is proposed.

Two segments were chosen from the old and new vibration signals for analysis. We used Minimum Entropy Deconvolution (MED) and Cepstrum Analysis to remove the transmission path effect from the vibration signal. We also used Time Synchronous Averaging (TSA), Autoregressive Filter (AR), Discrete/Randome Separation (DRS) and Cepstrum Analysis to separate deterministic components from non-deterministic components in the vibration signal. The latter resolves the first problem and former resolves the second one. Band-Pass Filtering, Wavelet Packet Transform (WPT), Hilbert-Huang Transform (HHT) and Spectral Kurtosis (SK) were used to detect excited resonant frequency for Demodulation and Enve-

lope Analysis. High kurtosis is a measure of high impulsivity in a signal. Therefore all of our methods rely on kurtosis.

The results demonstrate that the Minimum Entropy Deconvolution is a very effective method to remove the transmission path effect such that the defect impulses of the source can be clearly observed. A combination of wavelet and HHT as an Improved-HHT is a highly reliable method to detect the excited resonant frequencies. Spectral Kurtosis is an efficient and direct detector of the excited resonant frequencies. Cepstrum Editing or Liftering is a multi-purpose method for each stage. A new kind of cepstrum editing with a satisfactory output has been performed.

Order Tracking is performed before Time Synchronous Averaging and Discrete/Random Separation and in this study there is no tachometer signals to order tracking. Therefore these methods were not effective.

After examination on the old signal only the 5th and 27th harmonics of the shaft speed had a high amplitude. These frequencies are indeed the natural frequencies of the pump and do not indicate a fault.

On the other hand, the new signal led to many peaks. The strongest of these peaks were the shaft speed harmonics and $\frac{1}{2}$ harmonics. Lack of information on the pump specifications disables us to relate these peaks to specific faults. But these harmonics are generally presented in mechanical looseness.

1. Pump and Maintenance

1.1 Pump

A pump is a device that moves fluids (liquids or gases), or sometimes slurries, by mechanical action. A pump is a machine that moves a liquid or gas, or can propel a liquid or gas to a higher level or pressure. Pumps are the second most common machines in the world and are used in almost all industries including food and beverage, wastewater, pulp and paper, textiles, agriculture, electronics, steel, chemical, metals, and oil and gas.

1.1.1 Types of Pumps

The categories by which water-lifting equipment may be classified vary substantially both in terminology and type. One or a combination of the following general principles may be used to convey water:

- **Gravity systems:** Water flows downward under the influence of gravity but cannot be raised to a point higher than the source. The system can only be used to transfer water to a lower point.
- **Direct lift systems:** Fixed volume of water is physically raised in a single or a number of containers.
- **Displacement systems:** Water cannot be compressed (unlike air) and when it moves through a pump it draws further water behind it. The volume of water that is pumped is equal to the displacement of the piston when it is moved. This is effected immediately in a direct lift pump where the piston is generally in water, but in a suction pump where the piston is above the level of water, air must be evacuated before the system is able to be effective.
- **Velocity systems:** When water is propelled with sufficient momentum in the absence of air further water is drawn through the pump.

Pumps can be classified by their method of displacement in the following techniques:

- Positive displacement pumps
- Impulse pumps
- Velocity pumps
- Gravity pumps
- Steam pumps
- Valveless pumps

Mechanical pumps, in accordance the fluid they are pumping, may be **Submerged** or **Placed external**.

1.1.2 Subsea Pumps

Subsea water injection pumps are installed on the seabed without taking up space and weight capacity on the platforms topsides. The pumps use seawater from the surrounding ocean to inject into the reservoirs.

Different type of pumps in oil market:

- **Helico-Axial:** A rotodynamic pump with one single shaft that requires two mechanical seals, this pump uses an open-type axial impeller. It's often called a Poseidon pump, and can be described as a cross between an axial compressor and a centrifugal pump.
- **ESP (Electrical Submersible Pump):**
  Submersible pumps involve a submersible motor with a close coupled to single stage pump that allows the entire assembly to be operated submerged.
- **Centrifugal:** Effective in mature fields or in those with low reservoir drive energy and/or pressure. Single phase pumps facilitate water flooding to increase recovery and stimulate production and have the highest differential pressure capability among the pump types.
- **Twin-Screw:** Both twin screw and three screw pumps are successfully operating in multiphase, heavy crude oil and crude oil/water emulsion applications. The pump has two rotors, one drive and one driven, and relies on the pumped fluid to fill the clearances between the rotors and liner. The rotors are supported on both ends by bearings, and torque is transmitted from drive to driven rotor via timing gears.

- **Hybrid Pump:** Hybrid pump is a combination of helico-axial and centrifugal technologies, ideal to use downstream from a separator where the gas volume fraction is kept constantly low. High efficiency is a key characteristic of this pump.

1.2 Maintenance

All actions appropriate for retaining an item, part or equipment in a given condition or restoring it to a desired condition is called **maintenance**.

1.2.1 Why Maintenance?

At a closer look maintenance is important to gain some serious objects:

- Maximize availability
- Increase reliability
- Extend lifetime
- Minimize unplanned stops
- Keep the high safety standard
- Protect the environment
- Save the investment

1.2.2 Types of Maintenance

In General, there are three types of maintenance in use. After equipment break down, during using equipment and before equipment break down. The first one is often the most expensive one because worn equipment can damage other parts and cause multiple damages [22].

- Reactive maintenance
  - Failure or abnormal operation
- Preventive time based maintenance
  - Time based maintenance
  - Based on manufacturers experience
- Predictive / condition based maintenance
  - Maintenance based on actual measurements

The last one is more effective than the others.

1.2.3 Condition Based Maintenance

It has obvious advantages compared with either run-to-break or preventive maintenance, but requires to have an access to reliable condition monitoring techniques. Condition monitoring is based on being able to monitor the current condition and predicts the future condition of machines while it is in operation. Thus it means that information must be obtained externally about internal effects while the machines are in operation.

Vibration Analysis is one of the best techniques for obtaining information about internal conditions. A machine in standard condition has a certain vibration signature. Any damage to the machine changes that signature in a way that can be related to the fault. This has known as **Signature Analysis**.

2. Pump Vibration

2.1 Vibration Signals

Vibration signals belong to the general categories of different signals shown in Figure 2.1. In a basic view, signals are disparted to stationary and non-stationary.

Statistical properties in stationary signals does not vary with the time and in terms of signal processing, frequency content of stationary signals do not change in the time. The Fourier transform of a stationary signal can be used perfectly to describe the signal behavior.

For deterministic signals this basically means that they consist of discrete frequency sinusoids and thus their frequency spectrum consist of discrete lines at the frequencies of those sinusoids. Once the frequency, amplitude and initial phase of these components are known, the value of the signal can be predicted at any time in the future or past; hence it is called **deterministic signal**.

Random signals are somewhat more complex, as their value at any time cannot be predicted and for stationary random signals their statistical properties do not change by time.

Non-stationary means anything which does not meet the conditions for stationarity and once again it can be divided into two main classes, continuously varying and transient. There is no hard and fast rule to distinguish between these two types, but in general it can be said that transient signals only exist for a finite length of time and are typically analyzed as an entity. Once again, this requires a clarification, since a decaying exponential function, for example, theoretically decays to infinity, but in practical terms it only has a measurable value for a finite time. The terms energy and power are used to distinguish between transient and continuous signals.

A stationary random signal by definition has a constant power and therefore infinite energy.

Cyclostationary signals by definition have power (always positive) which varies periodi-

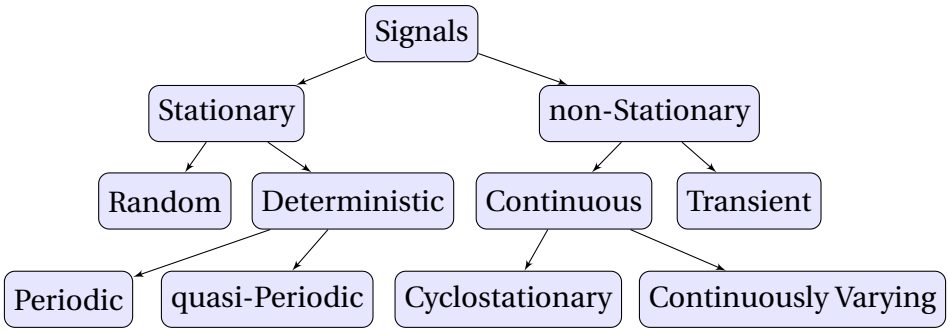

Figure 2.1: Signal types [22]

cally by time, also their total energy is infinite. Cyclostationary is an amplitude modulated white noise.

Continusely varying signal is a non-stationary signal with no regularity in time domain and frequency domain.

2.2 Pump Faults and Vibration

Most machine components produce specific vibration signals that characterize them and allow them to be separated from others, as well as distinguishing faulty from healthy condition. The distinguishing vibration features may be because of different repetition frequencies, for example a gearmesh frequency, which characterizes a particular pair of gears, and different sideband spacing which characterize the modulating effects of the two meshing gears on their common mesh frequency. Gear-generated signals are usually at harmonics (integer multiples) of the associated shaft rotation speeds, whereas the characteristic frequencies of rolling element bearings are generally not at harmonics of the associated shaft speeds. Some signals, typically associated with fluid flow, such as turbulence or cavitation, have a random nature, but may have a characteristic distribution with frequency. These signals are often stationary, but other random signals, characterized as cyclostationary, are often generated by machines and have statistical properties which vary periodically. The purpose of vibration analysis is mainly to categorize the various signals generated by machine components in healthy and faulty condition [19, 22].

Many rotating machines, may be considered as consisting of three major parts: the rotor, the bearings and the foundations. These rotating machines have a high capital cost and hence, the development of condition monitoring techniques is very important. Vibration

based identification of faults, such as rotor unbalance, rotor bends, cracks, rubs, misalignment, fluid induced instability are well developed based on the qualitative understanding of measured data and are widely used in practice. However the quantitative part, the estimation of the extent of faults and their locations, has been an active area of research for many years. Over the past three to four decades, theoretical models have played an increasing role in the rapid resolution of problems in rotating machinery.

In case of condition monitoring, changes in vibration signals are ascribed to changes in condition, so it is important that other factors which cause changes in vibration signals are considerably reduced or eliminated. Vibrations tend to change with the speed and load of a machine, so rotating machine will primarily be considered to be operated at constant speed and load, and the signals will typically be stationary and/or cyclostationary [28].

All pump faults that can be recognized by vibration analysis are as the following:

- Rotor unbalance (Imbalancing of mass distribution on rotor)
- Shaft bow or thermal bow
- Misalignment (Misalignment between driver and driven shaft)
- Shaft cracks
- Rubs
- Mechanical looseness of components
- Faults in gears
- Rolling element bearing faults

Unbalance, bent shaft and misalignment faults produce the low orders and low harmonics of shaft speed in vibration signals. Misalignment tends to even harmonics of shaft speed, in particular the second. Shortly saying, cracked shaft increases vibration at the first and second harmonics of shaft speed also gives an increase at the third harmonic, which gives a better chance of distinguishing it from unbalance and misalignment. Rubs are produced when the rotating shaft comes into contact with the stationary components of the machine. Rubs are said to be accompanied by a great deal of high-frequency spectral activities and it is generally transitory phenomenon.

Mechanical looseness can create a variety of patterns in a vibration signature. In some cases the fundamental frequency ($1X$) is excited. In others a frequency component at one-half multiples of shaft speed e.g. $0.5, 1.5, 2.5, \ldots$ is present. In almost all cases there are multiple harmonics, both full and half.

Gears typically generate a complex, broad vibration spectrum beginning with frequencies well below the shaft rotational speeds and extending to several multiples of gear mesh frequency. The product of the number of teeth on gear and the shaft rotational frequency is the characteristic frequency of a particular gear and is called the **gear mesh frequency**. Gear mesh frequency will be evident in the spectrum relating to any gearbox, in good condition or fault. When one gear becomes damaged the gear mesh frequency component of vibration may increase substantially as compared to the base line vibration measurements. Harmonics of gear mesh frequency may also become more apparent. Another frequency, which is often excited by gear defects, is the **resonant frequency** of geared shaft itself. This frequency can usually be measured by impulse testing. Both the natural frequency of the geared shaft and the gear mesh frequency may have accompanying side bands; sometimes side bands themselves can be the main indicator of a defective gearbox.

The frequencies that rolling element bearings generate are called **fundamental fault frequencies**. They are when rollers pass over an anomaly of roller surface or raceway surface. These frequencies are a function of the bearing geometry and the relative speed between the two raceways. When bearing geometry is known, the fundamental fault frequencies can be calculated using following equations.

Table 2.1: frequencies associated with defects in different bearing components

Ballpass frequency, outer race	BPFO	$\frac{nf_r}{2}\left(1-\frac{d}{D}cos\phi\right)$
Ballpass frequency, inner race	BPFI	$\frac{nf_r}{2}\left(1+\frac{d}{D}cos\phi\right)$
Fundamental train frequency (cage speed)	FTF	$\frac{f_r}{2}\left(1-\frac{d}{D}cos\phi\right)$
Ball (roller) spin frequency	BSF(RSF)	$\frac{D}{2d}\left(1-(\frac{d}{D}cos\phi)^2\right)$

where d is ball diameter, D is pitch diameter, f_r is the shaft speed, n is the number of rolling elements and ϕ is the angle of the load from the radial plane. This guidelines can be used as a quick reference: BPFO = $0.4nf_r$, BPFI = $0.6nf_r$, FTF = $0.4f_r$ and BSF = $1.6f_r$.

Signals produced by bearings with extended spalls are cyclostationary if the discrete carriers (such as gear-mesh harmonics) are modulated at a fixed cyclic frequency (shaft speed for an inner race fault). The roughness of the spall surface introduces randomness in the modulation and allows separation from deterministic gear faults [22, 28].

3. Vibration Analysis Techniques

3.1 Order Tracking

Frequencies that match to the machine's RPM or multiples of it are called **orders**. In analyzing rotating machine vibrations, it is often desired to have a frequency x-axis based on harmonics (orders) of shaft speed. Order tracking is a family of signal processing tools object to transform a vibration signal from time domain to angular (or order) domain. Order spectra in contrast of frequency spectra is often preferred for vibration or acoustic signals analysis of rotating machines. An order spectrum presents the amplitude and/or the phase of the signal as a function of order of the rotation frequency. The x-axis is based on harmonics or orders of the shaft speed; this means that a harmonic or subharmonic analysis remains in the same independent of the shaft speed. This technique is called **tracking**, as the rotation frequency is being tracked and used for analysis. Dynamic forces in a machine are almost related to the frequency of rotation so those can be interpreted simply by use of order tracking. The smearing of the frequency components caused by speed alteration in machine is solved by using order analysis. In cases where the frequency components from a normal frequency analysis are smeared together, proper diagnosis will be made easy via order analysis. In particular, order analysis is helpful for the analysis of the vibrations during a startup or a shutdown of a machine when the structural resonances are excited by the rotational frequencies or its harmonics. It is very important to distinguish the critical speeds for excitation of normal modes for the machines especially for large machines such as turbines and generators. In the normal mode all the parts of the system have free oscillation in the same frequency synchronously. The key words in the digital tracking technique are: Oversampling, Interpolation and Resampling [22, 27].

Three main families of computed order tracking techniques have been developed in the past [7]:

- **Computed order tracking:** It obtains by inserting many extra zeros between samples and then using a low-pass filter and resampling again, or by interpolating and then resampling. It can be also found by Spectrum zero padding, by padding the FFT spectrum with zeros in the center.
- **Vold-Kalman filter:** Vold-Kalman filter is a particular formulation of Kalman filter, able to estimate both instantaneous speed and amplitude of a series of harmonics of the shaft rotational velocity [11].
- **Order tracking transforms:** That is a combination of Order tracking and Fourier transform in a single step. The additional resampling does not need in this method. The Velocity Synchronous Discrete Fourier Transform is one of the formula for the order tracking transform.

Resampling by a factor of four is illustrated in Figure 3.1.

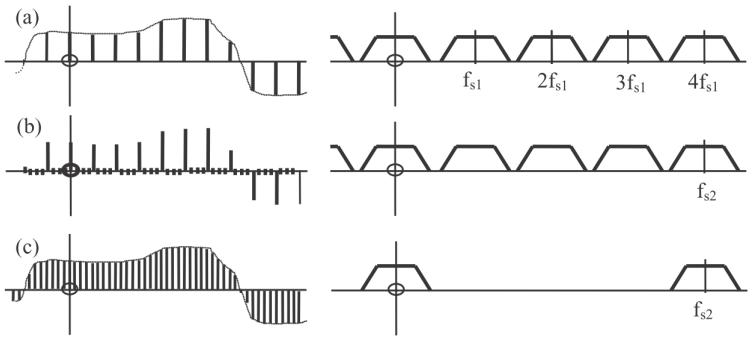

Figure 3.1: Digital resampling with four times higher sampling frequency: (a) signal sampled at f_{s1} and its spectrum; (b) addition of zeros which changes sampling frequency to f_{s1}; (c) low-pass filtration and rescaling [22]

.

3.2 Time Synchronous Averaging (TSA)

The most common and easiest method of discovering frequency information is to apply spectrum averaging. Averaging is performed in the frequency domain and is applied for any type of signal. But if the signal is periodic and we already know about that the averaging can be performed in the time domain. The averaging must be run synchronously. The random components are damped with TSA. Signal enhancement or time synchronous averaging is a classic way of separating periodic signals from background noise and each component which is not a subharmonic of the particular fundamental frequency. In practice, to perform

the time synchronous averaging the signal is splitted to a series of synchronous segments and then an averaging is done on.

$$y_a(t) = \frac{1}{N}\sum_{n=0}^{N-1} y(t+nT) \tag{3.1}$$

Before the averaging regardless in the time domain or in the frequency domain the signal must be order tracked. Spectrum averaging smooths the spectrum of the random noise in a spectrum and makes the discrete frequency components easier to see, but it does not actually reduce the noise level unlike the time domain averaging.

This can be modelled as the convolution of $y(t)$ with a train of N delta functions displaced by integer multiples of the time period T. This system is a filter similar to comb in the frequency domain [22]:

$$C(f) = \frac{1}{N}\frac{sin(N\pi Tf)}{sin(\pi Tf)} \tag{3.2}$$

3.3 Autoregressive Filter (AR)

Autoregressive filter is the most famous filter to capture the deterministic information from the noise corrupted signals. Autoregressive model (AR) is an all-pole linear IIR filter. When an AR filter is excited with white noise it produces a signal with the same statistics as the primary signal that we are trying to model with the AR. AR is numerically as the same as a linear prediction filter as shown in Figure 3.2 that predicts the deterministic or predictable part of the next value of signal by a certain number of samples in the immediate past. The residual part or the unpredictable part of the signal is then calculated by the subtraction from the actual signal value. In the case of ideal AR model the residual signal will be white noise [22, 26, 27].

$$\hat{x}(n) = \sum_{i=1}^{p} a_i x(n-i) \tag{3.3}$$

Choosing p as called the **model order** is the most challenging part of the AR modelling. When the application is to separate discrete frequency components from stationary white noise, a standard approach is to use the Akaike Information Criterion (AIC). Residual $e(n)$ is said to be **prewhitened** as it includes stationary white noise and impulses. $E(Z)$ is the z-domain counterpart of $e(n)$ and is named **forcing function** [22, 26].

$$e(n) = x(n) - \hat{x}(n) \tag{3.4}$$

$$x(n) + \sum_{i=1}^{p} a(i)x(n-i) = e(n) \tag{3.5}$$

$$X(Z)A(z) = E(Z) \qquad or \qquad X(Z) = (1/A(Z))E(Z) \tag{3.6}$$

The large kurtosis of AR residual signal $e(n)$ is a signature of local gear faults [22, 26].

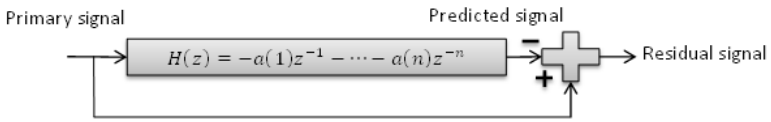

Figure 3.2: Autoregressive (AR) model used for Linear Prediction

The $\{a[k]\}$ are obtained using the matrix form Yule-Walker equations that is often solved using the so-called Levinson-Durban recursion (LDR) algorithm [22].

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & \dots & r_{xx}[-p+1] \\ r_{xx}[1] & r_{xx}[0] & \dots & r_{xx}[-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[p-1] & r_{xx}[p-2] & \dots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = \begin{bmatrix} r_{xx}[1] \\ r_{xx}[2] \\ \vdots \\ r_{xx}[p] \end{bmatrix} \tag{3.7}$$

In this equation the square matrix is a Toeplitz matrix and $r_{xx}[n]$ is simply autocorrelation function of time series $x[n]$.

$$r_{xx}[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[k][k-n], \quad 0 \le n \le p-1 \tag{3.8}$$

3.4 Self-Adaptive Noise Cancellation (SANC)

Separation of periodic vibrations from non-deterministic ones is very helpful in the vibration analysis. A typical application is for separation of gear and bearing signals [22].

Vibration signals measured on machines can be generally very complex when the numerous sub-systems are involved, especially for machines with a gearbox that several components are rotating in different speed and make effect in the vibration signal [27]. In many situations, for such a complex system, can be very difficult to use vibration analysis in a

simple condition monitoring scheme. However, when operating at constant speed, many of the contributions in the signals can be decomposed by Fourier transformation and thus be individually analyzed [22, 27].

This applies for all sub-systems that are mechanically phase-locked to the operating speed and then can be expected to match with specific spectral lines in the frequency domain. Gears obviously belong to this class of sub-systems as their vibrations cause to happen families of discrete and equi-spaced peaks in the frequency domain. However, there are other mechanical devices and processes same as bearing which do not produce discrete spectra because they are generally not exactly phase-locked to the machine speed. These can actually experience some degree of randomness in their operation, thus cause to happen smeared continuous spectra in the frequency domain. The rolling is unavoidably accompanied by some slipping with an intensity depending on the ratio between the axial and radial loads. This behavior has been shown to be typical of rolling element bearings [22, 27].

As a consequence, it will only be possible to return two sources, one containing all the periodic components and the other one all the broadband components. In other word, the issue is to decompose a frequency spectrum into a discrete and a continuous part.

3.4.1 Prediction Theory

The issue of decomposing a spectrum into discrete and continuous components is a recurrent one in signal processing. Regardless of the physical reasons, it is also an advantage from a statistical point of view since the two types of signals have different properties and should therefore be analyzed with different statistical tools. Before thinking of how to achieve that, it should first be questioned if such a decomposition is always feasible. Simply, any stationary process can be decomposed into a deterministic (or singular) and a non-deterministic (or regular or stochastic) part. There is a fundamental theorem, Wold's theorem, that says any stationary process $X(n)$ has the representation [18, 22]:

$$X(n) = p(n) + r(n) \tag{3.9}$$

$p(n)$ and $r(n)$ are respectively a deterministic process and zero mean stationary process that are uncorrelated with each other. The zero mean stationary process as a non-deterministic

process can be modeled by moving average

$$r(n) = \sum_{i=0}^{\infty} a_i \epsilon(n-i) \tag{3.10}$$

Here $\epsilon(n)$ is a white noise as random error that is called **innovation term**. The deterministic component can be modeled by determining its N previous samples without any innovation term. This presentation of stationary signal is ARMA (Autoregressive Moving Average) model presentation [1, 29].

Prediction theory also gives an elegant solution for the separation of the two signals. It was pointed out that the deterministic signal could be predicted from its past values with zero prediction error. In fact, since the correlation time of a periodic signal is infinite, the prediction can theoretically be made from any values arbitrarily far back in the past, still with zero prediction error. On the other hand, the autocorrelation of a non-deterministic signal decays towards zero with increasing time-lags; thus prediction from past values becomes less and less accurate as data become old. Ultimately, when the data used for prediction become older than the memory (effective correlation time) of the non-deterministic signal, the best predicted value is zero and the prediction error equals the signal itself. This forms the basis of the separation scheme, where the periodic signal is predicted from the far past values of the vibration signal, and the non-deterministic signal is estimated from the prediction error [1, 2].

Let us define $\hat{x}(n)$ as the predictor of $x(n)$ from a N finite number of past values from $x(n-d)$ to $x(n-d-N+1)$ or

$$\hat{x}(n) = \sum_{i=0}^{N-1} h_i x(n-d-i) \tag{3.11}$$

And suppose d be selected such that autocorrelation of prediction error for more far from d is zero

$$R_{rr}(k) = E(r(n)r(n-k)) = 0, \qquad k > d \tag{3.12}$$

R_{rr} is autocorrelation function of error prediction r. Then the best answer for $\hat{x}(n)$ that minimizes the mean squared error of $e^2 = E\{(x(n)-\hat{x}(n))^2\}$ is given by the conditional expectation [3, 29].

$$\hat{x}(n) = E(x(n)|x(n-d), x(n-d-1), ..., x(n-d-N+1)) \tag{3.13}$$

This is a projection of $x(n)$ to basis of order N with basis vectors $x(n-d), x(n-d-1), ..., x(n-d-N+1)$. If we select an appropriate N, then by minimizing e^2, the best result for deterministic component of $x(n)$, $p(n)$ will be $\hat{x}(n)$ and the non-deterministic component of $x(n)$ is $r(n) = x(n) - p(n)$.

A close-form solution for calculating $\{h_i\}$ is in the form of Weiner-Hopf equations [1]:

$$\sum_{i=0}^{N-1} h_i R_{xx}(n-d-i) = R_{xx}(n), \quad n \in Z \tag{3.14}$$

In this method, R_{xx} is **autocorrelation function** of x. The number of weights $\{h_i\}$ can be very large so the directly solving is not a good idea. Instead, there are some methods to optimally solve the equations:

- Wiener filtering
- Autoregressive estimation
- Kalman filtering
- Recursive least squares estimation

Autoregressive estimation is in fact the time average form of Wiener filtering and the recursive least squares estimation is the time average form of Kalman filter. For further refer to [4].

Adaptive Noise Cancellation algorithm (ANC) is an efficient recursive way to solve Weiner-Hopf equations [2].

3.4.2 Time-Domain SANC

Adaptive noise cancellation (ANC) is a class of general adaptive filters [18] and is a procedure where a primary signal containing two uncorrelated components can be separated into those components by reference differencing. When one of the components is deterministic and the other is random, the reference signal can be a delayed version of primary signal, so the method is called Self-Adaptive Noise Cancellation (SANC) or so called Adaptive Linear Enhancer (ALE) [22].

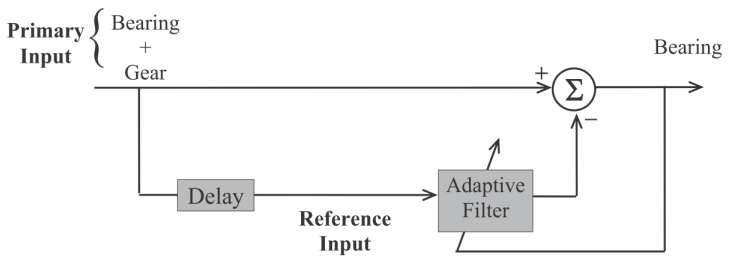

Figure 3.3: Self-Adaptive Noise Cancellation (SANC) [22]

The key idea used is the difference in correlation times or in other words memories of two types of contributions. For a sufficiently long time-lag, they become uncorrelated and can be separated on this basis. The idea has been known in the statistical and signal processing in terms of prediction theory as expressed in previous section.

Therefore, as long as we are concerned with stationary signals, the SANC algorithm is nothing other than a convenient tool for solving the Wiener-Hopf large system of equations.

The adaptive filter adjusts its parameters to minimize the variance of the error signal ϵ. Because the two components are uncorrelated, the variance of the total signal will be the sum of the variances of the two constituents, and thus the separation will be achieved when the variance of the difference signal is minimized, meaning that it then contains no part of the reference signal.

The time delay or prediction depth should be chosen large enough so as to exceed the memory of the noise in the input signal for uncorrelation. This time delay is achievable because of broadband spectrum (continues spectrum) of noise. In the theoretical aspect correlation time of periodic signal is infinite, but since the deterministic component may not be exactly periodic but rather pseudo-periodic then the correlation time is not infinite and so delay can not be too large. In other word, correlation length will be of the inverse order of the narrowest bandwidth of noise.

The filter length N must be sufficient to extract all important periodic factors from the original signal. In theory, when there is no noise, an order of 2k is required to extract k sinusoidal components. When noise is present, higher orders are required for its rejection. In general, the choice of N faces a compromise between sufficient selectivity of the frequency response on the one hand, and convergence and small estimation bias on the other hand.

The recursive algorithm so-called Least Mean Squares (LMS) algorithm is used to compute filter coefficients. It uses the gradient of the mean squared error to change the filter

coefficients in a recursive algorithm. If a component of the gradient is positive then the corresponding coefficient will be decreased for the next round and for a negative component the coefficient will be grown up [22, 30]. The formula 3.15 and 3.16 show the idea.

$$W_{k+1} = W_k - \mu\Delta_k \tag{3.15}$$

$$\Delta_k = \frac{\partial E[\epsilon_k^2]}{\partial W_k} \tag{3.16}$$

where μ is a convergence factor, which should be chosen carefully to avoid divergence on the one hand, but not give rise to excessive adaptation time on the other. Figure 3.4 shows LMS process.

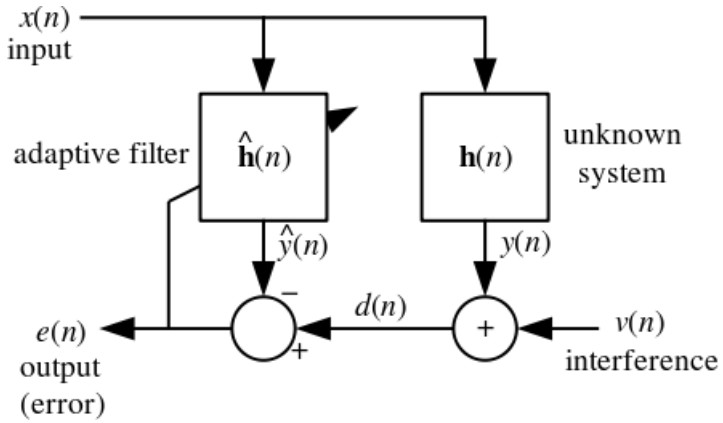

Figure 3.4: Least Mean Squares Adaptive Filter (LMS)

3.5 Discrete/Random Separation (DRS)

This method is virtually the same as SANC but more efficient, because it is based in the frequency domain and takes advantage of speedup of FFT. In contrast to SANC, it does not require adaptation phase, because the filter to remove discrete frequency components is first determined, possibly from the whole length of data. For this reason, it requires the discrete frequency components to be very stable, and so order tracking is recommended as a preprocessing step.

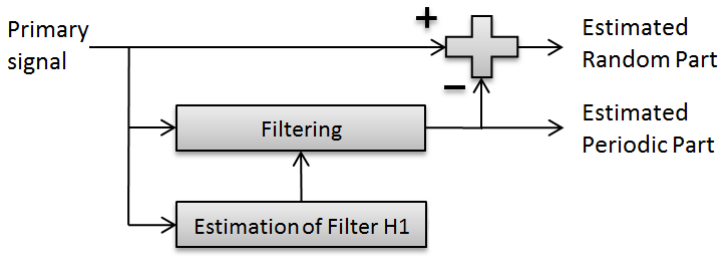

Figure 3.5: Discrete/Random Separation (DRS)

Weiner-Hopf equations (expressed in SANC section) describe a convolution of autocorrelation of the signal with a finite-length filter. In frequency domain, it can be simplified by product. For more simplification the autocorrelation signal is then estimated as an average of products of short-time Fourier transforms taken over the signal. To make this approach to be efficient, the short-time Fourier transforms should be run on the shortest sequences that at least cover the filter length. Of course, this derives unavoidable distortions due to windowing effects.

The basic principle is to obtain the transfer function between the signal and a delayed version. The cross-spectrum from input to output is divided by the auto-spectrum of the input. Let $X_k(n)$ be a windowed sequence of length N taken at time kN; i.e. $X_k(n) = X(n + kN)W_N(n)$, $n = 0, 1, ..., N-1$ with $W_N(n)$ a weighting window of length N. In the same way, define $X_k^d(n) = X(n + kN - d)W_N(n)$ as a delayed signal. The objective is to find the filter which best predicts $X_k(n)$ from $X_k^d(n)$. By choosing appropriate d and N, the deterministic components in the signal remain deterministic and the broadband noise will be rejected from the primary signal $X_k^d(n)$ [22].

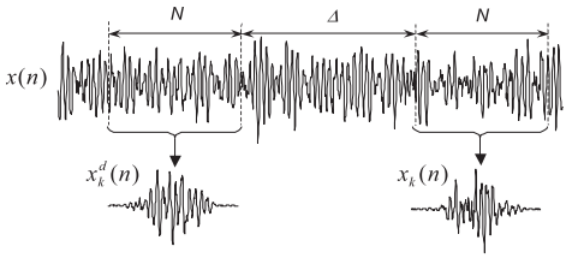

Figure 3.6: Short-time sequences used in the frequency-domain algorithm

Let us bring the formulation for this filter:

$$X(f) = R(f) + P(f), \quad X^d(f) = R^d(f) + P^d(f) \tag{3.17}$$

Input of the filter is $x^d(n)$, delayed of $x(n)$, and output is only deterministic part of $x(n)$

$$P(f) = H(f).X^d(f) \tag{3.18}$$

Multiplying both side by $X^{d*}(f)$:

$$P(f).X^{d*}(f) = H(f).X^d(f).X^{d*}(f) \quad \Rightarrow \quad S_{px^d}(f) = H(f).S_{x^dx^d}(f) \tag{3.19}$$

$S_{ab}(f)$ is called **cross-power spectrum** and is equal to spectrum of cross-correlation. For the stationary signals that the input noise and the output noise are not correlated the cross-spectral densities are:

$$S_{x^dx^d}(f) = F(\gamma_{x^dx^d}) = F(\gamma_{p^dp^d} + \gamma_{r^dr^d} + \gamma_{r^dp^d} + \gamma_{p^dr^d}) = S_{p^dp^d}(f) + S_{r^dr^d}(f) \tag{3.20}$$

$$S_{px^d}(f) = S_{xx^d}(f) = F(\gamma_{xx^d}) = F(\gamma_{pp^d} + \gamma_{rr^d} + \gamma_{rp^d} + \gamma_{pr^d}) = S_{pp^d}(f) \tag{3.21}$$

$$H(f) \approx \frac{S_{xx^d}(f)}{S_{x^dx^d}(f)} \tag{3.22}$$

When the Signal-to-Noise Ratio (SNR) is higher the method is more accurate. Ideally, where the signals are correlated, this formula provides a value of unity. This happens in the components with a discrete frequency. In contrast, the output will be zero at the frequencies that are noise. But in practice with a complex and noisy signal we expect the different values depend on the signal to noise ratio. This equation is called H_1 **filter** in system analysis. Several Frequency Response Function (FRF) estimation techniques have been proposed in literature in order to minimize the error effects; $H_1(f) = S_{yx}(f)/S_{xx}(f)$ and $H_2(f) = S_{yy}(f)/S_{xy}(f)$ are the estimates obtained with least square techniques. The **coherence** defines as a ratio of H_1 and H_2:

$$G_{xy} = \frac{H_1(f)}{H_2(f)} = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \tag{3.23}$$

and varies between zero and one. The coherence is a measure that shows the signal is noisy or noise-less. The value one for the coherence shows the signal is noise-less or high-quality.

Usually the spectral densities are averaged over a number of sampling intervals to reduce

the effect of noise [22]:

$$H(f) \approx \frac{\sum S_{xx^d}(f)}{\sum S_{x^d x^d}(f)} \tag{3.24}$$

Compared with the SANC algorithm, the frequency-domain (FD) algorithm has a much lower complexity. However, this introduces unavoidable distortions (leakage error) due to windowing effects, which must be care in choosing the proper window.

3.6 Cepstrum Analysis

The cepstrum analysis is used to highlight periodicities in the frequency domain, in the same way that the spectrum is used to highlight periodicities in the time domain. Harmonics and sidebands in the spectrum are summed into one peak in the cepstrum. With the cepstrum we can identify the modulation frequencies associated with a specific fault.

The complex cepstrum is defined as the inverse FFT of the logarithm of the FFT of a signal [22, 23].

$$C(f) = F^{-1}\{log(F\{x(t)\})\} \tag{3.25}$$

X(f) as FFT of $x(t)$ is complex with an even amplitude of $A(f)$ and an odd phase of $\phi(f)$. So, the complex cepstrum will be a real valued.

$$X(f) = A(f)e^{j\phi(f)} \tag{3.26}$$

$$log(X(f)) = log(A(f)) + j\phi(f) \tag{3.27}$$

The power cepstrum or real spectrum is the inverse FFT of the logarithm of the power spectrum of a signal.

$$C(f) = F^{-1}\{log|F\{x(t)\}|\} \tag{3.28}$$

Source and transmission path effects are additive in cepstrum:

$$y(t) = x(t) * h(t)$$

$$Y(f) = X(f).H(f)$$

$$logY = logX + logH$$

$$F^{-1}\{logY\} = F^{-1}\{logX\} + F^{-1}\{logH\}$$

Note that before calculating the complex cepstrum, the phase function $\phi(f)$ must be unwrapped to a continuous function of frequency, and this is often difficult, so it is better to make solutions with the real cepstrum.

The real cepstrum was originally proposed as a better alternative than the autocorrelation function for detecting echo delay times, specifically for seismic signals.

The auto-correlation function is the inverse Fourier transform of the power spectrum

$$R_{xx}(\tau) = F^{-1}\left\{|F\{x(t)\}|^2\right\} \tag{3.29}$$

It is also equally a spectrum of a spectrum, but the critical difference with cepstrum is the logarithmic conversion of the power spectrum, which converts source and transfer function effects to additions in the cepstrum, whereas they are related by a product and a convolution, respectively, in the spectrum and autocorrelation function. Inverse spectrum in the cepstrum formula is only useful where the signal in the time domain is comparable with the cepstrum.

Using of words such as cepstrum, quefrency, rahmonics and lifter formed by reversing the first syllable of the corresponding originals, spectrum, frequency, harmonics and filter were proposed in the papers because of the cepstrum is a spectrum too, but of a spectrum.

There are three main areas of application of cepstrum analysis:

- As the cepstrum determines the periodicities in the frequency domain, it summarizes the all families of uniformly spaced harmonics and sidebands.
- As the cepstrum is a logarithm function or in other word converts the products to additions, it separates forcing functions (at the tooth-mesh) from transfer functions as some discrete points.
- Echoes as a convolution with some delta functions will be observed by the cepstrum as some delayed impulses as shown in Figure 3.7.

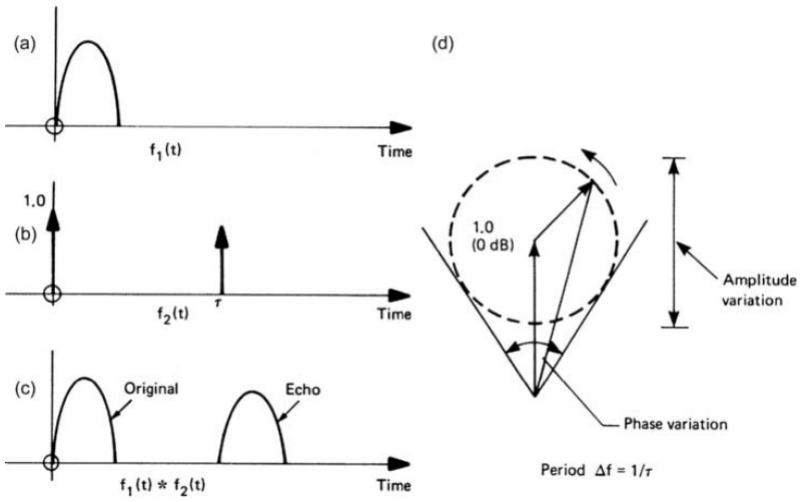

Figure 3.7: Showing that an echo gives a periodic component in the signal spectrum

3.6.1 Cepstrum Application in Machine Diagnostics

As mentioned above, cepstrum analysis detects periodic structure in the frequency domain and separates the source and channel effects. So, cepstrum can be applied for machine diagnostics as follows [6, 22]:

- Detection of periodic structure in spectrum
  - Harmonics (Faults in gears, bearings, blades)
  - Sidebands (Faults in gears, bearings, blades)
  - Echoes, reflections
- Separation of source and transmission path effects (for SIMO, single input multiple output systems)

Cepstrum is very good solution for gear fault diagnosis but it can only be used for bearing fault diagnosis when the fault generates discrete harmonics in the spectrum.

Local faults in gears produce some impulsive modulation of the gearmesh signals placed in a period correspond to gear speed. There are a large number of sidebands of both amplitude modulation and frequency modulation. The sidebands are weak and more visible in a log amplitude spectrum similar to cepstrum. Furthermore, cepstrum is effective representation because of summing all sidebands in the spread spectrum in one sideband. For slow-speed machines the excited resonance frequencies are usually high harmonics and they are

typically smeared together, so this technique is more useful for high-speed machine with resonances excited relatively low harmonic.

The noise level in the spectrum affects the detection of a series of harmonic components. It is obvious that if the harmonics are completely immersed in the noise, they will not be detected at all in the cepstrum. The techniques for noise reduction such as time synchronous averaging can be used to cause discrete frequency components to be more highlighted from background noise.

The cepstral coefficients are a very compact representation of the envelope. It also turns out that cepstral coefficients are uncorrelated. This is very useful for statistical modeling, because we do not need to store their covariances, which reduces exponentially required memory and process.

3.6.2 Cepstrum Editing

As the same filtering the signal in spectrum domain we can lifter the signal in cepstrum domain. For cepstrum editing, it previously was necessary to use complex cepstrum but it needs unwrapping the phase and it is not possible for response signals so, **real cepstrum** is used for liftering. Cepstrum editing is used to reveal system resonant frequencies or to remove selected harmonics and sideband families. In the new method, the amplitude of force or response signals are edited in real cepstrum and then the edited cepstrum combined with the original phase is applied together to produce the original signals. Figure 3.8 shows schematic of real cepstrum editing [22, 23].

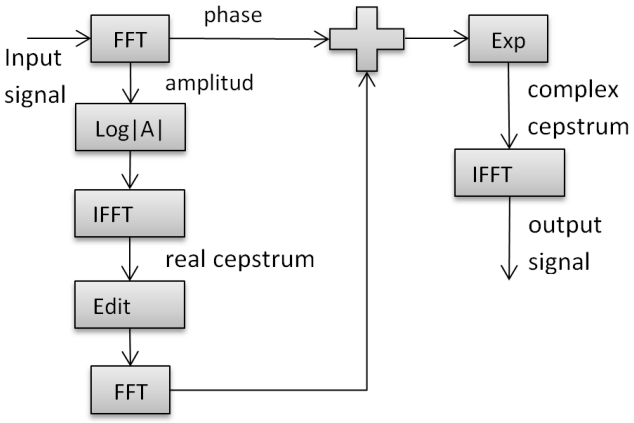

Figure 3.8: Real cepstrum editing

3.7 Empirical Mode Decomposition (EMD)

Empirical Mode Decomposition (EMD) is a time adaptive decomposition operation that iteratively separates high frequency components from the original data. The decomposed components are called Intrinsic Mode Function (IMF) which ideally are mono-component signals and follow IMF conditions [12].

Specifically, the core of EMD is the sifting process, which first defines envelopes of upper and lower extremes, and iteratively subtracts the mean of both envelopes from the previous data or residual, until it satisfies the necessary condition of IMF. The IMF satisfy the following two conditions [12]:

- The number of local extremes and the number of mode zero-crossings are equal or differ at most by one.
- The local average of upper and lower envelopes is zero.

Any oscillatory signal, regardless of stationarity or non-stationarity, can be decomposed into separate constituent functions that satisfy the above conditions.

The process of EMD is as follow [21, 24]:

- Given a data set, find all the local maxima, and connect those points to develop an upper envelope, using a cubic spline. Likewise, find the lower envelope from local minima.
- If the mean value of the upper and lower envelopes is denoted as m_{10}, the first component h_{10} is:

$$h_{10} \stackrel{\Delta}{=} x(t) - m_{10} \tag{3.30}$$

New maxima and minima shall again be identified, and if steps 1 and 2 are repeated k times to satisfy IMF conditions:

$$h_{1k} \stackrel{\Delta}{=} h_{1(k-1)} - m_{1k} \tag{3.31}$$

This repeated process is known as **sifting**. h_{1k} will be the first IMF c_1:

$$c_1 \stackrel{\Delta}{=} h_{1k} \tag{3.32}$$

We can separate c_1 from the original $x(t)$, to extract the second IMF:

$$r_1 \stackrel{\Delta}{=} x(t) - c_1 \tag{3.33}$$

Then, we assume r_1 to be the original signal, and repeat the previous steps to find the second IMF c_2. After the n-th iteration, the original signal $x(t)$ can be decomposed into n number of IMFs, as below:

$$r_2 \stackrel{\Delta}{=} r_1 - c_2 \tag{3.34}$$

$$\dots$$

$$r_n \stackrel{\Delta}{=} r_{n-1} - c_n \tag{3.35}$$

$$x(t) = \sum_{i=1}^{n} c_i + r_n \tag{3.36}$$

Decomposition process ends when r_n (n empirical modes) becomes a monotonic function (contains no more than two extrema), so that no more IMF can be deduced. r_n is residual component and the mean trend of $x(t)$. The individual IMFs $c_1, c_2, ..., c_n$ covers a broad range of the frequency band, from the highest to the lowest.

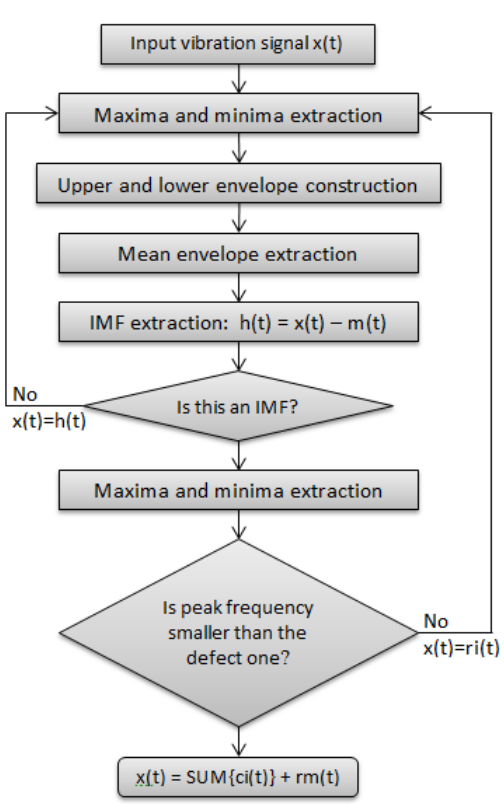

Figure 3.9: Empirical Mode Decomposition (EMD) algorithm

Genetic Algorithm seeks for the vectors of weight for each IMF that maximizes the kurtosis. After completing the GA, we will find a significant contrast of kurtosis between healthy and damaged bearings. A schematic block diagram for the process of sensitivity enhancement of a bearing defect through EMD-GA is following:

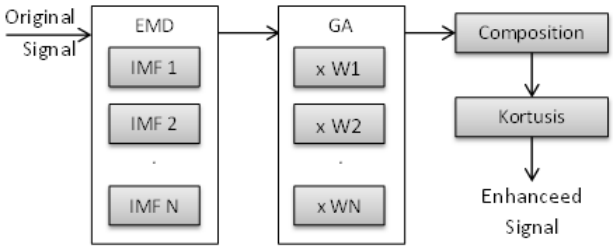

Figure 3.10: Using Genetic Algorithm applied to IMFs for choosing a appropriate weight vector subjected to high kortusis

3.8 Hilbert-Huang Transform (HHT)

3.8.1 Analytic Signal and Hilbert Transform

The Fourier transform of a real function is Hermitian:

$$s(t) = s(t)^* \Rightarrow S(-f) = S(f)^* \tag{3.37}$$

Define $S_a(f)$ contained only non-negative frequencies of S(f):

$$S_a(f) = \begin{cases} 2S(f) & f > 0 \\ S(f) & f = 0 \\ 0 & f < 0 \end{cases}$$

Or:

$$S_a(f) = 2u(f)S(f) = (1 + sgn(f))S(f) \tag{3.38}$$

Inversely, and according to $S(f)$ is Hermitian symmetric:

$$S(f) = \begin{cases} 1/2S_a(f) & f > 0 \\ S_a(f) & f = 0 \\ 1/2S_a(-f)^* & f < 0 \end{cases}$$

Or:

$$S(f) = 1/2[S_a(f) + S_a(-f)^*] \tag{3.39}$$

Then $s_a(t)$, the analytic signal of $s(t)$, is defined as:

$$s_a(t) = F^{-1}(S_a(f)) = F^{-1}[(1 + sgn(f))S(f)] = s(t) + s(t) * j/\pi t \tag{3.40}$$

$$s_a(t) = s(t) + j\hat{s}(t) \tag{3.41}$$

Hilbert transform of $s(t)$ is:

$$\hat{s}(t) = s(t) * 1/\pi t \tag{3.42}$$

Then instantaneous amplitude or envelope, instantaneous phase and instantaneous frequency are defined as:

$$s_m(t) = |\hat{s}(t)| \tag{3.43}$$

$$\phi(t) = arg[\hat{s}(t)] \tag{3.44}$$

$$\omega(t) = 1/2\pi d\phi(t)/dt \tag{3.45}$$

Also the Hilbert transform can be said to be the relationship between the real and imaginary parts of the Fourier transform of a one-sided function. Any natural impulse response function is causal and thus one sided in the time domain, and this means that the real and imaginary parts of the corresponding frequency function are related by a Hilbert transform. This becomes evident when it is considered that a causal function is made up of even and odd components which are identical for positive time, and canceled for negative time. Suppose $x(t) = (x(t) + (x-t))/2 + (x(t) - (x-t))/2 = x_e(t) + x_o(t)$ and because $x(t)$ is one-sided function the even and odd parts are related by $x_e(t) = x_o(t).sgn(t)$ or $X_e(f) = X_o(t) * F(sgn(t))$ where $sgn(t)$ is the sign function with Fourier transform of an imaginary hyperbolic function $1/j\pi t$. Since the real part and imaginary part of $X(f)$ are respectively the Fourier transform of the even part and odd part of $x(t)$, e.g. $X_R(f) = X_e(f)$ and $X_I(f) = X_o(f)/j$ we can write the Hilbert relationship between real and imaginary part as:

$$X_R(f) = \frac{1}{\pi}\sum_{-\infty}^{+\infty} X_I(\phi)\frac{1}{f-\phi}d\phi \tag{3.46}$$

The equivalent equation for the Hilbert transformation of a time function x(t) is

$$\hat{x}(t) = \frac{1}{\pi}\sum_{-\infty}^{+\infty} x(\tau)\frac{1}{f-\tau}d\tau \tag{3.47}$$

And in the frequency domain

$$\hat{X}(f) = X(f)(-jsgn(f)) \tag{3.48}$$

Which shows the Hilbert transform can be achieved more simply in the frequency domain just with a shifting the phase by $-\pi/2$ for positive frequency components and $\pi/2$ for negative frequency components.

3.8.2 Hilbert-Huang Transform (HHT)

HHT can decompose the data on the base of time characteristics auto-adaptively, which is called data driving [13]. It seems that HHT can be competent for the underwater acoustic signal detection and analyzing.

The signal has to be decomposed into some mono-component signals (IMF) by empirical mode decomposition (EMD).

$$x(t) = \sum_{i=0}^{N} c_i(t) + r_n(t) \tag{3.49}$$

c_i is IMF component and r_n is Gaussian noise as residual.
Hilbert transform for each IMF will be:

$$Hilbert[c_i(t)] = c_i(t) + jd_i(t) \tag{3.50}$$

Then the instantaneous amplitude, phase and frequency for each IMF are:

$$a_i(t) = \sqrt{c_i^2(t) + d_i^2(t)} \tag{3.51}$$

$$\phi_i(t) = arctan\left(\frac{d_i(t)}{c_i(t)}\right) \tag{3.52}$$

$$\omega_i(t) = 2\pi f_i(t) = \frac{d\phi_i(t)}{dt} \tag{3.53}$$

$c_i(t)$ is denoted the ith IMF, which has time-varying amplitude and time-varying frequency.

$$c_i(t) = Real\left[a_i(t)e^{j\int \omega_i(t)dt}\right] \tag{3.54}$$

And with eliminating residual term, $x(t)$ can be written as:

$$x(t) = Real\left[\sum_{i=1}^{n} a_i(t)e^{j\int \omega_i(t)dt}\right] \tag{3.55}$$

It is similar to Fourier representation of $x(t) = \sum_{i=1}^{n} a_i e^{j\int \omega_i dt}$ with a_i and ω_i constant.

So IMF represents a generalized of Fourier expansion. The variable amplitude and the instantaneous frequency have not only greatly improved the efficiency of the expansion, but also enabled the expansion to accommodate nonlinear and nonstationary data. With the

IMF expansion, the amplitude and the frequency modulations are also clearly separated. Thus, the restriction of the constant amplitude and frequency of the Fourier expansion has been overcome, with a variable amplitude and frequency representation. This frequency-time distribution of the amplitude is designated as the Hilbert Spectrum $H(\omega, t)$.

There are some suggestions to improve HHT. One of them considers each IMF as a time series:

$$c_i(n) = A(n)sin[2\pi f(n)n + \phi] + e_i(n) \tag{3.56}$$

It can be written as a time-varying autoregressive moving average model (TV-ARMA) model that is also called Adaptive ARMA or AARMA.

$$c_i(n) + \sum_{j=1}^{2} a_{ij}(n)c_i(n-j) = \sum_{j=1}^{2} b_{ik}(n)e_i(n-k) + e_i(n) \tag{3.57}$$

There are several methods to calculate the time-varying parameters A(n) and B(n) such as LMS, RLS, Kalman filtering and recursive AR algorithm. The model is AARMA(2,2) since at the moment n, with the theory of harmonic retrieval, the instantaneous frequency of IMF is unique.

The characteristic polynomial of the part AR(2), $1 + a_{i1}(n)z^{-1} + a_{i2}(n)z^{-2} = 0$, has two conjugate roots $[z_i(n), z_i^*(n)]$; Then the instantaneous frequency is:

$$\omega_i(n) = 2\pi f_i(n) = arctan[\frac{Im(z_i(n))}{Re(z_i(n))}] \tag{3.58}$$

Another approach to improve HHT is by eliminating the uncorrelated IMFs with original signal. A threshold of correlation will be help to select useful IMFs.

The Table 3.1 summarizes a comparative view between Fourier, wavelet and Hilbert-Huang transforms. HHT is as an adaptive method which is useful for non-stationary and non-linear signals and that is very efficient.

Tansform	**Fourier**	**Wavelet**	**Hilbert**
Basis	a priori	a priori	adaptive
Frequency	convolution: global, uncertainty	convolution: regional, uncertainty	differentiation: local, certainty
Presentation	energy-frequency	energy-time-frequency	energy-time-frequency
Nonlinear	no	no	yes
Non-stationary	no	yes	yes
Feature Extraction	no	discrete: no, continuous: yes	yes
Theoretical Base	theory complete	theory complete	empirical

Table 3.1: HHT vs Fourier and Wavelet

3.9 Minimum Entropy Deconvolution (MED)

Many vibration signals are considerably distorted by the transmission paths from the source to the transducer. This is particularly the case for impulsive type signals. The Minimum Entropy Deconvolution (MED) method is designed to reduce the spread of Impulse Response Functions (IRF), to obtain signals closer to the original. In fact, the MED technique deconvolves the effect of the transmission path [17, 22].

That is particularly useful to identify a train of response pulses arising from sharp impacts for example from local spalls in gears and bearing. In the next step we determine the frequency of repetition of impulses and that is only possible when if the IRFs are shorter than the spacing between them, and this is not always the case for high-speed machines. So, the MED is useful to enhance impulses arising from faults in gears and detect the bearing faults. [14].

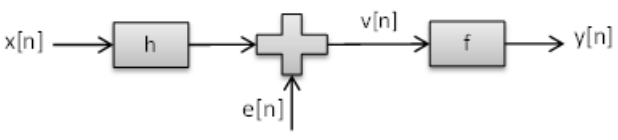

Figure 3.11: Inverse filtering process for MED

$$y[n] = \sum_{l=1}^{L} f[n]\, v[n-l] \tag{3.59}$$

$f[i]$ is a filter with L coefficients that has to invert the system IRF $h[i]$

$$f[i] * h[i] = \delta(i - l_m) \tag{3.60}$$

Filter f is a causal system so the delay l_m will not be zero. It will displace the signal by l_m but will not change pulse spacing in the signal.

This is an Objective Function Method (OFM) where the objective function to be maximized is impulsivity or kurtosis of the output signal $y[n]$.

$$Kurtosis[y] = k(f) = \sum_{i=1}^{N} y^4[i] \Big/ \Big[\sum_{i=1}^{N} y^2[i] \Big]^2 \tag{3.61}$$

the maximum of k is found by values of $f[l]$ as independent variable for which the derivative of k is zero

$$\partial(k(f[l])) \big/ \partial(f[l]) = 0 \tag{3.62}$$

the filter coefficients of $f[l]$ is achieved iteratively [9, 17].

In other word, the goal is simply recovering or deconvolving the source signal from noise-corrupted measurements, without a priori knowledge of IRF. That is similar to Blind Source Separation, but the constraint towards the coefficient of the deconvolution filter is different. This algorithm finds a solution to minimize the entropy of the signal. The entropy represents the amount of data content in the sequence of a given signal. Thus, a higher level of entropy is associated with the enforcement of randomness or uncertainty. In contrast, minimizing the entropy leads to a simple structured signal, such as sparse spikes.

Figure 3.12 illustrates the conceptual flowchart of MED.

The MED process for a non-stationary signal can be combined with AR linear prediction filtering or any other techniques related to separating deterministic and non-deterministic parts. The total process of AR and MED shown in Figure 3.13 is called AR-MED [22, 27]. An example of applying AR-MED filtering is shown in Figure 3.14.

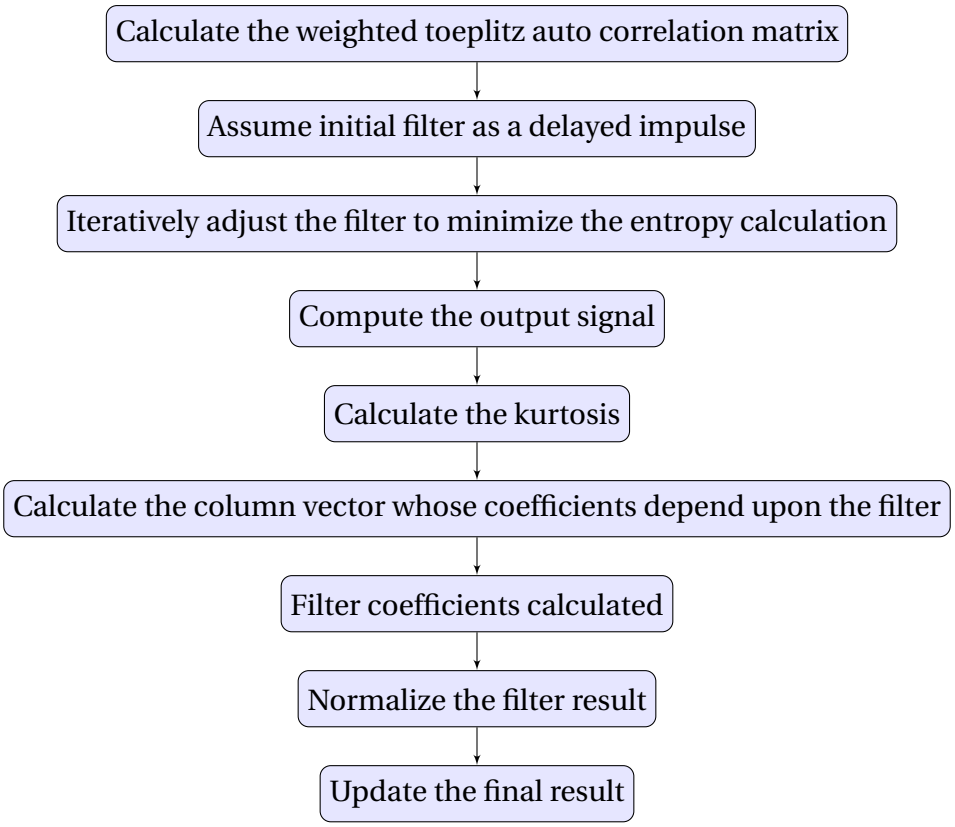

Figure 3.12: Minimum Entropy Deconvolution (MED) [17]

Figure 3.13: Autoregressive (AR) model used for Linear Prediction

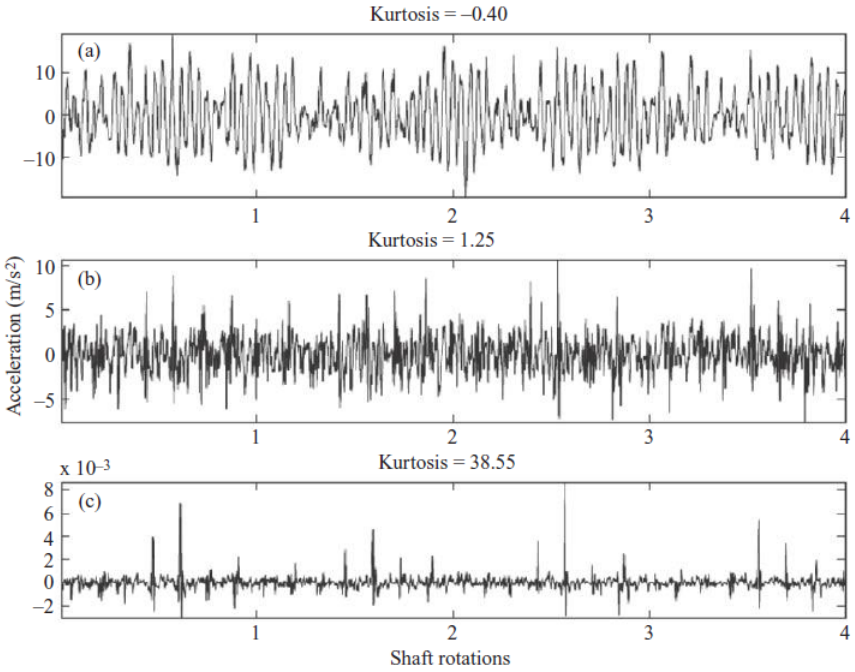

Figure 3.14: Example of applying both AR and MED filtering to bearing signals with an inner race fault in a high-speed bearing: (a) original time signal; (b) after application of AR filtering; (c) after additional MED filtering.

3.10 Teager-Kaiser Energy Operator (TKEO)

TKEO algorithm belongs to the category of nonlinear high-pass filters, which reduces the variation of low frequency background signals, while boosting transient components of a signal in the high frequency region. Eventually, transients and background signals can be easily separated through TKEO. The first-order discrete time model of TKEO is expressed as equation [14, 17, 22]:

$$\psi(x[n]) = x[n]^2 - x[n-1]x[n+1] \tag{3.63}$$

$\psi(x[n])$ is known as **Teager Energy** of the signal. TKEO detects a sudden change of the energy because it is based on differentiating technique and is useful for condition monitoring of a non-stationary signal. It can be applied to amplify impulse component caused by defects, and suppress the background noise, to increase the kurtosis sensitivity [16, 22].

3.11 The Short Time Fourier Transform

A simple approach is to move a short time window along the record and obtain the Fourier spectrum as a function of time shift. This is called the Short-Time Fourier Transform (STFT). However, the uncertainty principle means that the frequency resolution is the reciprocal of the effective time window length. STFT is sometimes useful for tracking changes in frequency with time, even with the restriction of resolution. It is described by the formula:

$$Y(f,t) = \int_{-\infty}^{\infty} x(\tau)\, w(\tau - t) e^{-j2\pi f t} dt \tag{3.64}$$

where $w(t)$ is a window which is moved along the signal. Normally, the amplitude squared $|Y(f,t)|^2$ is displayed on a timefrequency diagram known as a spectrogram. The window could be of finite length such as a Hanning window, or theoretically infinite such as a Gaussian window, but in practice of course it must be truncated.

3.12 Wavelet Analysis

An approach to time-frequency analysis is to decompose the signal in terms of a family of wavelets which have a fixed shape, but can be shifted and dilated in time. The formula for the wavelet transform is [5, 15, 20, 22]:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \tag{3.65}$$

where $\psi(t)$ is the mother wavelet, translated by b and dilated by factor a. Since this is a convolution, the wavelets can be considered as a set of impulse responses of filters, which because of the dilation factor have constant percentage bandwidth properties. In principle, they are not very different from $1/n$th octave filters, but with zero phase shift because the mother wavelet is normally centered on zero time. Wavelets give a better time localization at high frequencies, and for that reason can be useful for detecting local events in a signal. Wavelets can be orthogonal or non-orthogonal and continuous or discrete. Examples of orthogonal wavelets are the Daubechies dilation wavelets, which are compact in the time domain, but in principle infinite in the frequency domain. They tend to have irregular shapes in the time domain. There are also complex harmonic wavelets, which are compact in the

frequency domain, but infinite in the time domain. They have the appearance of windowed sinusoids and are typically of one-octave bandwidth. The advantage of complex wavelets is that the imaginary part of the wavelet is orthogonal to the real part and thus the overall result is not sensitive to the position of the event being transformed. The local sum of squares of the real and imaginary parts is a smooth function [21].

Wavelets have an important application in signal denoising. After wavelet decomposition, the high frequency sub-bands contain most of the noise information and little signal information. Most wavelet denoising is based on two types of thresholding to remove noise. In hard thresholding any components with amplitude less than a certain threshold is removed but in soft thresholding, the threshold is set to higher values for high frequency sub-bands and lower values for low frequency sub-bands [17, 22].

Wavelet is useful to localize the frequency band of faults in the vibration signal specially in connection with Envelope Analysis [8, 10, 15, 26, 27].

3.13 Envelope Analysis

The signals produced by faults in rolling element bearings are a series of high-frequency bursts as resonance frequencies are excited by near periodic impacts. The diagnostic information is contained in the repetition frequency, not in the resonance frequencies excited, but spectra obtained by direct Fourier analysis are dominated by the resonance frequencies, and the important information is disguised by smearing of the high-order harmonics. Such signals can be modelled as an amplitude modulation of a carrier signal at the resonance frequency by a near periodic series of exponential pulses. In so-called **envelope analysis**, envelope of the signal is extracted by amplitude demodulation to reveal the repetition frequencies. Vibration signal is bandpass filtered in a high-frequency band in which the fault impulses are amplified by structural resonances and then with a method as Hilbert transform, the envelope or instantaneous amplitude is derived. The spectrum of envelope signal contains the desired diagnostic information.

3.14 Spectral Kurtosis and Kurtogram

In a blind analysis, selecting a frequency band that is contained the diagnostic information is surprisingly possible. Simply, the frequency with maximum impulsivity is the best candidate

for bearing fault. This result is very important since without having the historical data from a healthy case the fault bearing information can be detect. Spectral kurtosis (SK) can be used as a method to calculate impulsivity in frequency domain. The kurtosis for each frequency f is defined by taking the fourth power of $H(t,f)$ as STFT at each time and averaging its value along the record, then normalizing it by the square of the mean square value. It can be shown that if two is subtracted from this ratio the result will be zero for a Gaussian signal.

$$K(f) = \frac{< H^4(t,f) >}{< H^2(t,f) >^2} - 2 \tag{3.66}$$

$< . >$ is mean or expected value. To obtain a maximum value of kurtosis, the window must be shorter than the spacing between the pulses but longer than the individual pulses.

Since SK is large in frequency bands where the impulsive bearing fault signal is dominant, and effectively zero where the spectrum is dominated by stationary components, it makes sense to use it as a filter function to filter out that part of the signal with the highest level of impulsiveness. However, the optimum result in a given case may vary with both the centre frequency and bandwidth of the filter. A display showing the optimum combination was called the **kurtogram**. Kurtogram is a 3 dimension graph of SK, center frequency and bandwidth. Computation of the full kurtogram covering all combinations of centre frequency and bandwidth is very costly and so a number of more efficient alternatives have been proposed. **Fast kurtogram** is based on a series of digital filters rather than STFT. The gains in computational speed are based on a dyadic decomposition similar to the discrete wavelet packet transform (DWPT) [22, 25, 26, 27, 31].

4. Experiments

4.1 Data-Set

The signals under study are the vibration signals from two subsea pumps A and B which gathered in two different years. All the collected signals are from a hydrophone array that is made up of four hydrophones placed in known locations close to the pumps. The sampling rate is 12500 Hz. We have some information about the pumps status and operations, specially the pumps shaft speed are recorded every two minutes. Usually at any moment only one of the two pumps is on. The signals have been collected in the files with the **ogg** format every 419 seconds. The control and status data are saved in the MySQL database.

At the first, we read the data (pump A speed, pump B speed, date and time) from MySQL table into two MATLAB memory files, one of them is for old data and the other is for new data. At a glance the signals seemed to have high level of noise, more specifically the old signal had a strong spread spectrum at the frequency domain.

4.2 Shaft Speed and Frequency Domain

The relation between pump speed and signal in the frequency domain shows the energy of signal which is almost directly proportional to pump speed. Reasons of the vibration of any kind, are repeated one or more time at one rotation of pump and is obviously occurred more at higher speeds. So, the energy of signal is increased as it is shown in Figure 4.1. The spiky data at the plot shows the resonant frequencies. Although this drastic change in energy can not be caused by just increasing the vibration; flow and turbulence increasing is the main reason of this energy.

Figure 4.2 shows at the higher speeds the energy of signal is shifted toward the high frequencies; and at the lower speeds is shifted toward the low frequencies. Furthermore, it can

be seen that two plots have negative high correlation and positive high correlation together for low speeds and high speeds respectively.

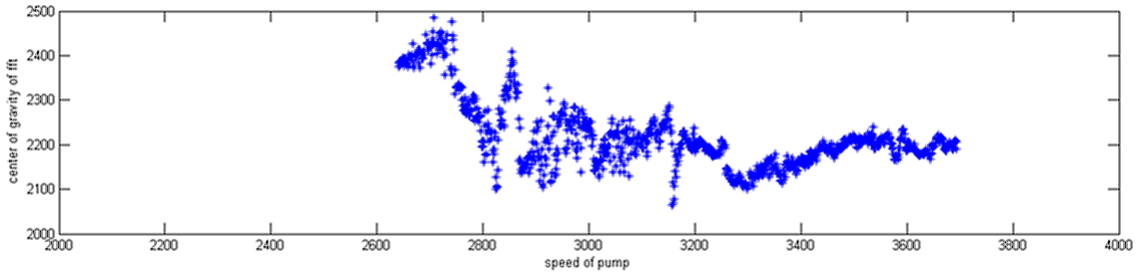

Figure 4.1: Energy of vibration signal related to pump speed

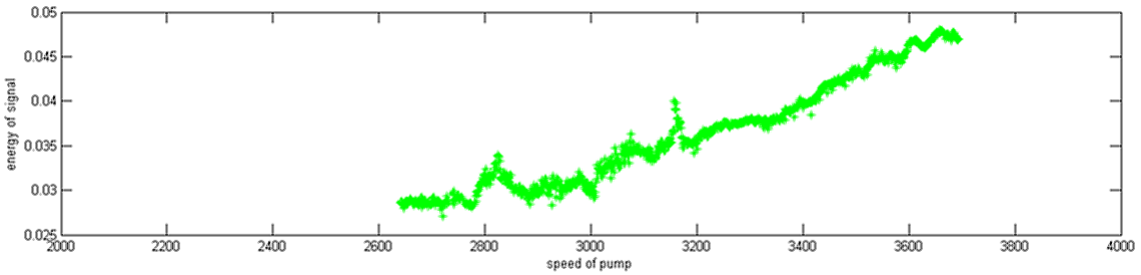

Figure 4.2: Central frequency of energy related to pump speed

4.3 Shaft Speed Estimation

It is interested to find a method to determine the shaft speed from the vibration signal as if the speed is unknown. We combined some empirical methods to calculate speed:

1. Using the 80 first strongest coefficients from FFT.
2. Using the 10 first strongest coefficients from FFT of |FFT|.
3. Replacing the frequencies in the neighborhood of ±5 step by zero except of the strongest one and the same way for FFT of |FFT|.
4. Calculating distance matrix for non-zero components for both collections as candidates for speed.
5. A statistical quasi-mode is used to final choosing.

FFT of |FFT| as a simple initiative way gives the periodicity at frequency domain and it is a collector of the all harmonics at just one line. A few month later, we were faced with

cepstrum analysis which is similar to my empirical method.

Statistical **Quasi-Mode** is a simple initiative way to find out a number as most frequent number and the mean simultaneously from a data set. It is like a combination of two statistical functions Mode and Mean . In this method we used distance matrix, then the column with minimum sum is the final selection as speed.

	56	78	79	85
56	0	22	23	29
78	22	0	1	7
79	23	1	0	6
85	29	7	6	0
sum	74	30	30	42

Table 4.1: Calculating the statistical quasi-mode with the distance matrix. 30 is the minimum sum, so 78 and 79 are final candidate for speed. choosing 78.5 is reasonable.

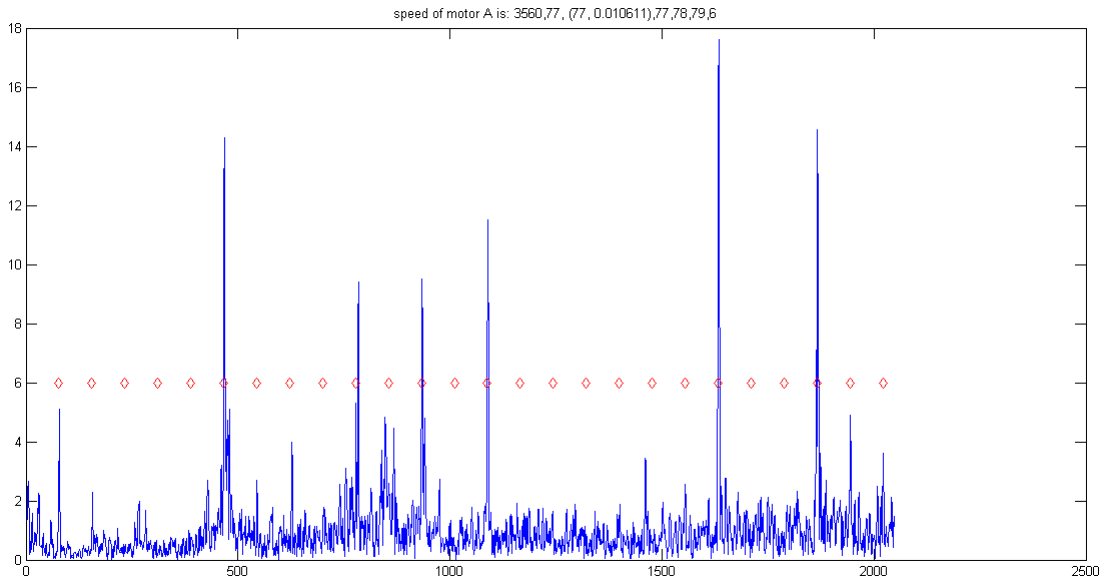

Figure 4.3: Using statistical Quasi-Mode as final selection among 77,78,79,6 is 77, exactly the same as the actual speed

4.4 Signal Selection

To minimize the effect of speed fluctuation and to eliminate order tracking process we used two segments with relatively stable speed from both old and new data. The order tracking

was not really an option with conventional methods because we did not have tachometer pulses to resampling data as angular sampling. We selected the frames with relatively long length 2^{14} with some overlapping and Blackman-Harris windowing to compensate the overall leakage effects. For overlapping, we chose the progression of one second or 12500 samples (equal to the sampling rate) between frames. We have four channels from a hydrophone array with high cross-correlation at low frequencies and very low cross-correlation at high frequencies. Since these correlations depend on acoustic velocity and the hydrophones distance we used the average of them as a kind of TSA for low frequency analysis. The frequency correspond to the shaft speed is shown by $1X$ and the i_{th} harmonics are shown by iX for some integer $i > 1$.

4.5 Time Synchronous Averaging

Time Synchronous Averaging is used to eliminate signal components that are not synchronous with the shaft rate of rotation. Eliminated components include electrical noise, bearing vibrations, and vibrations related to other shafts. We expected a damping effect at least to random parts after time synchronous averaging or in the other word, converting the signal more deterministic and a measure for that is the impulsivity increasing of signal in frequency domain. The increase of kurtosis in frequency domain is a good sign to show the damping effect. If the kurtosis remains almost unchanged this is a sign that shows there is no other vibration except of shaft speed. This is a well-known technique for gear and shaft fault diagnosis.

After averaging some frames of length correspond to 10 cycle rotations of pump and then perform FFT, the kurtosis was calculated on every FFT as a frame. Figure 4.29 shows the kurtosis of FFT for some frames before and after Time Synchronous Averaging.

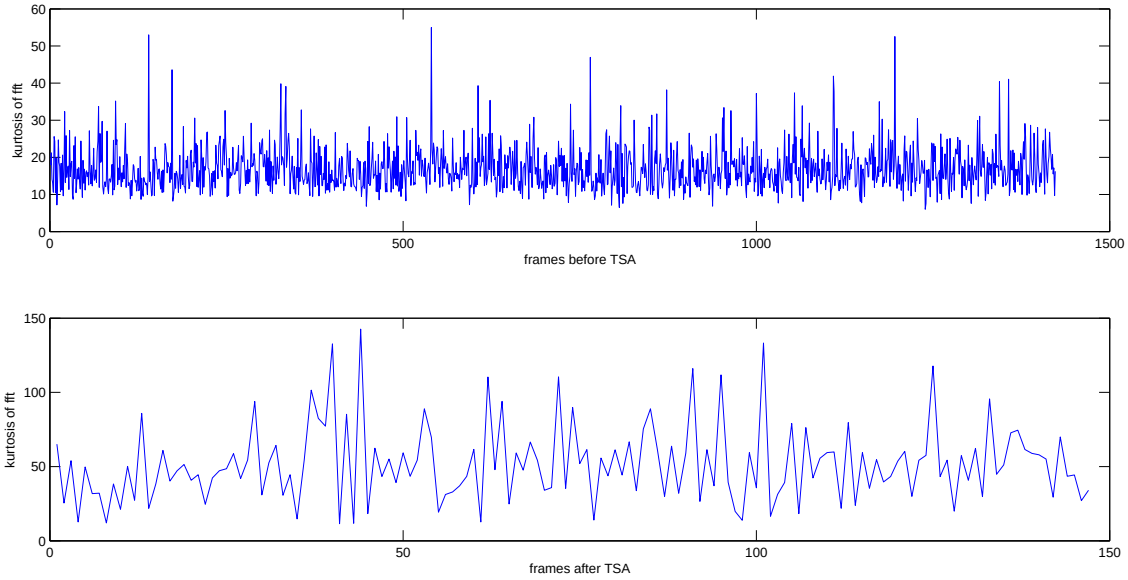

Figure 4.4: Kurtisis of FFT before and after TSA

On the contrary, the residual signal is possible to obtain using synchronous signals minus the TSA signals. The TSA residual is a step to bearing fault diagnostic.

4.6 Separating Deterministic and Random Parts, AR and DRS

Auto-regressive filter is used to separate random and deterministic components from the signal. Choosing order of the AR filter is the important consideration. In our application the impulsivity or in other word, the kurtosis in the random part is a criterion of having faults, so we choose the order of filter so that to maximize the kurtosis. In both old and new raw signals the kurtosis of the frames is very close to three as the same as Gaussian white noise. By using Yule-Walker method to estimate auto-regressive all-pole model, the kurtosis after applying AR filter is changed but not so much and still is about three and always less than 3.3.

Discrete Random Separation (DRS) is an alternative method for Self Adaptive Noise Cancellation (SANC) and is more efficient because of using FFT. In both methods, the basic principle is to obtain the transfer function between the signal and a delayed version. The high correlation between the deterministic components of the data and the delayed version and on the other hand, zero-correlation between random components of those is the key of this technique. The zero-correlation is yielded when the memory of the system is shorter than

the selected time as delay. But in practice we do not have the possibility to select a wide range of delay time because of fluctuation of the speed. As already mentioned there is no a tachometer signal to run for order tracking; then there is no big segment of data with a fixed speed in hand. Due to this issue we do not count on DRS in this study.

4.7 Removing Channel Effects, MED

Minimum Entropy Deconvolution filter is used to compensate the transmission path effect as the impulses in the source will be observable in the recorded vibration signal. By trial and error we used a proper length for filter MED to maximize the kurtosis. The MED as mentioned in the previous chapter is a complement to AR and DRS, so in this study the MED method applied after AR and DRS. The AR-MED significantly increases the kurtosis of both old and new data up to 14.

Specially with connection to envelope analysis we found that the MED has a very important rule to recognize the sidebands. Because the MED filter makes the smeared defect impulses at the raw signal very similar to the defect impulses at the source with a narrow body, we can even easily observe the narrow sidebands before applying amplitude demodulator. Furthermore, amplitude demodulator will be more efficient.

Figures 4.5,4.6 and 4.7,4.8 show the dissimilarity of the old and new data. With AR-MED filtering the kurtosis in the time domain is significantly increased. Particularly, in the new data some $1X$ periodic impulses are observed in the time domain that can be related to some defects. In contrast, for the old data there is no observable high-frequency repetitive impulses. In the frequency domain, it is more clear that the new data includes almost all the harmonics of the shaft speed. In both of the signals, there are some peaks along the sidebands that after envelope analysis are specified what those are about.

Although not as common as harmonics, it is critically important to learn how to recognize sidebands because amplitude modulation is always a sign of abnormality. Except of electrical malfunction, sidebands are generated by gear or bearing related problems. Since the sidebands can be at very low amplitude it is better to be analyze them on a logarithmic scale to be more observable.

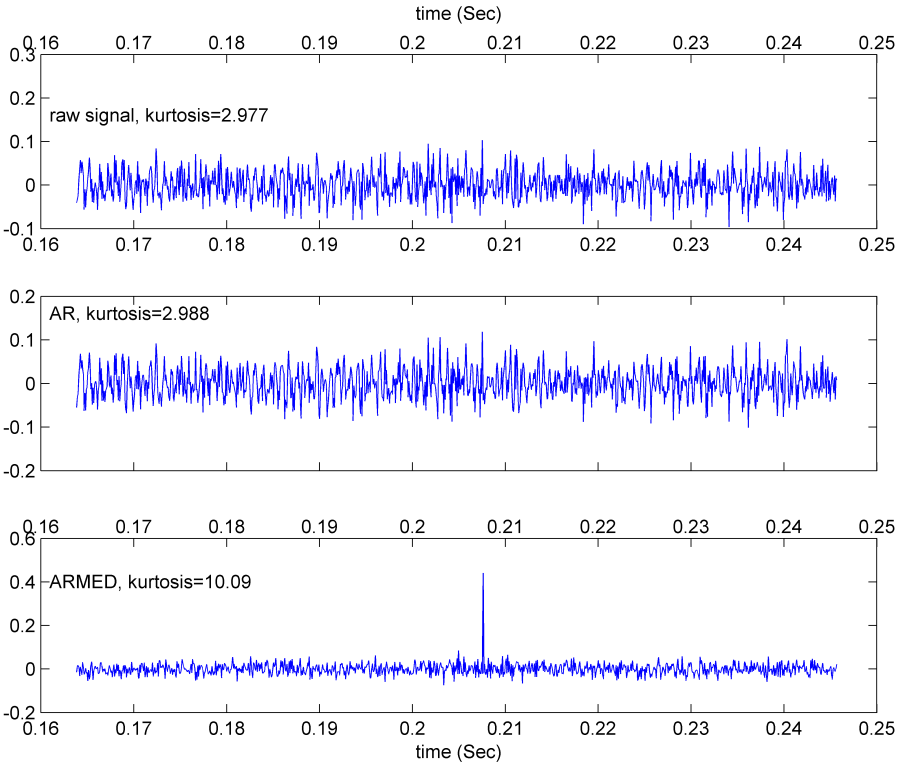

Figure 4.5: AR-MED filtering and Kurtosis for the old data: raw signal; after applying AR filter; after applying AR-MED filter

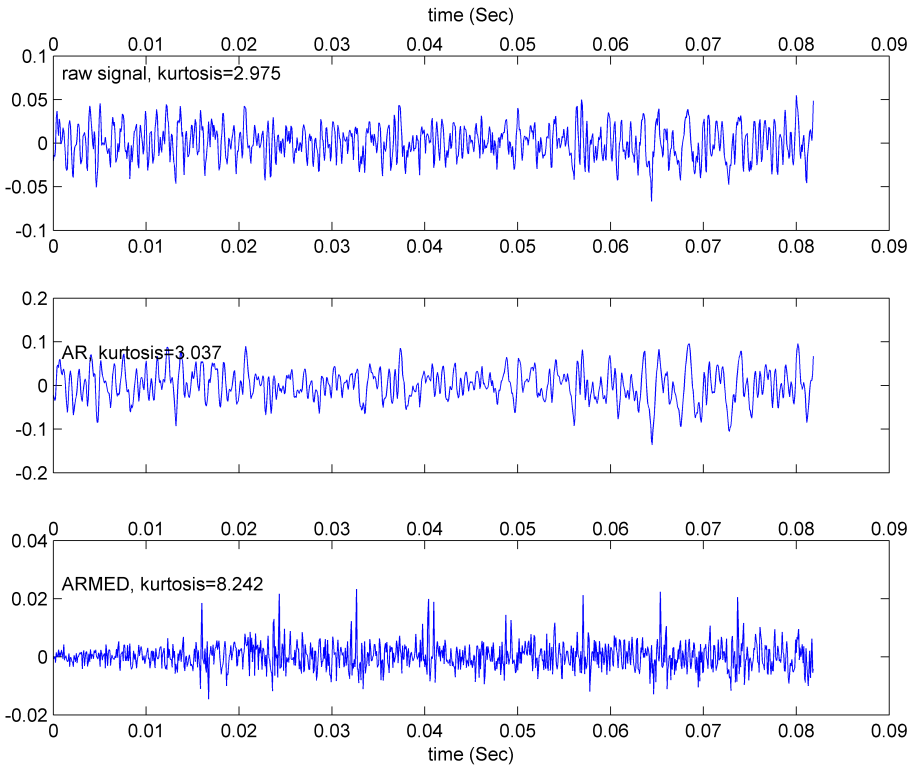

Figure 4.6: AR-MED filtering and Kurtosis for the new data: raw signal; after applying AR filter; after applying AR-MED filter

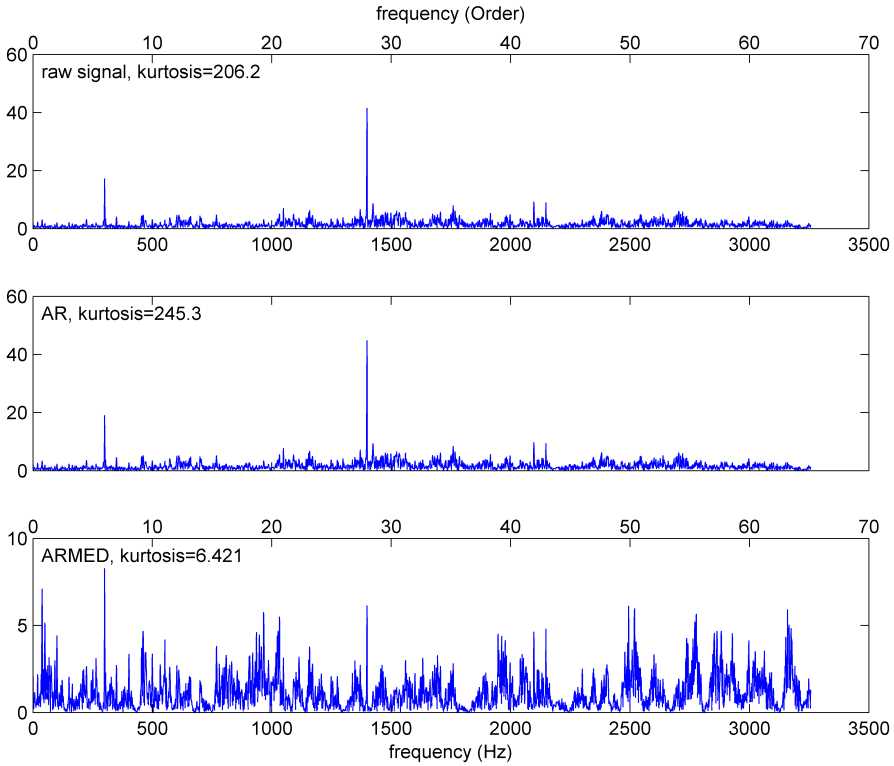

Figure 4.7: AR-MED filtering and Kurtosis for the old data: FFT of raw signal; FFT after applying AR filter; FFT after applying AR-MED filter

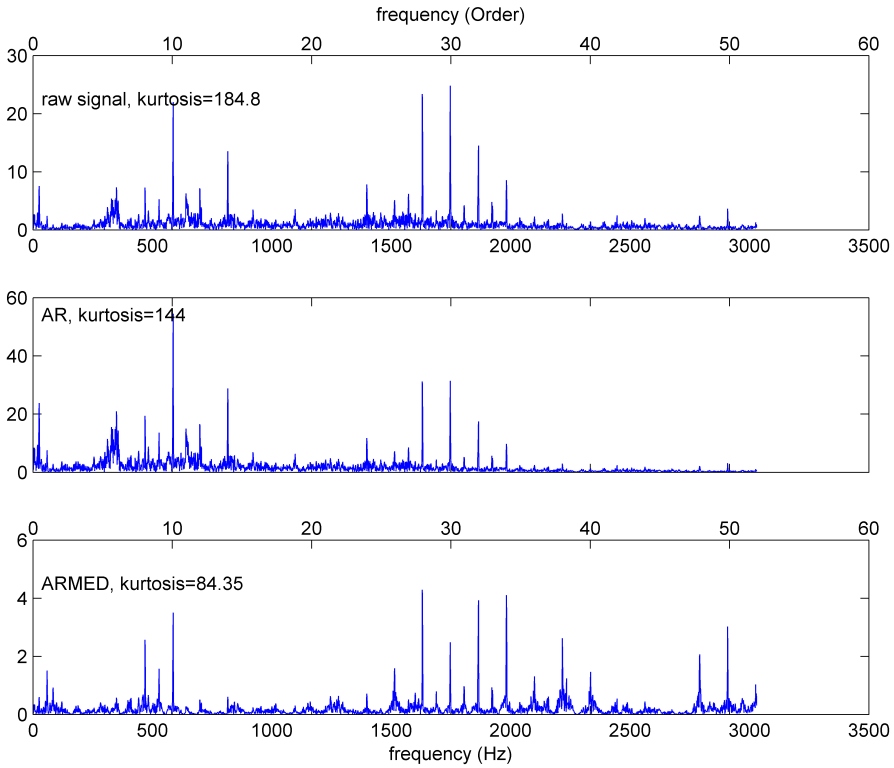

Figure 4.8: AR-MED filtering and Kurtosis for the new data: FFT of raw signal; FFT after applying AR filter; FFT after applying AR-MED filter

TSA is the first step to vibration analysis so the AR-MED filtering is also applied on the residual signal of TSA. The results have a larger kurtosis as expected. Figures 4.9,4.10 and 4.11,4.12 show the AR-MED filtering on the TSA residual signal that represents a more random signal respect to the raw signal.

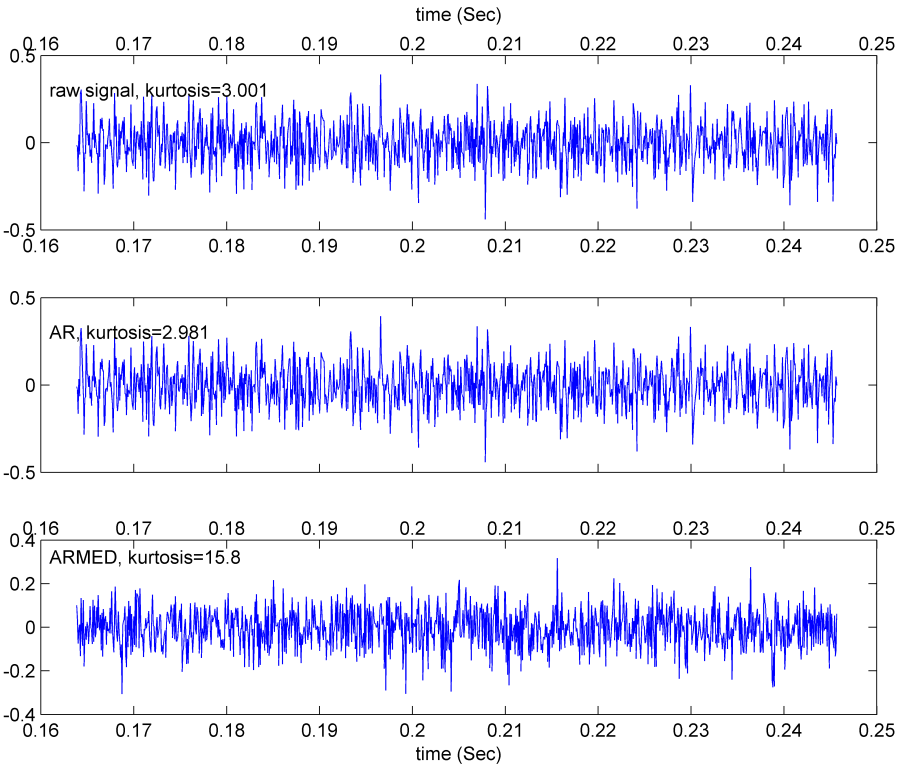

Figure 4.9: AR-MED filtering and kutosis for TSA residual of the old data: TSA residual; after applying AR filter on TSA; after applying AR-MED filter on TSA

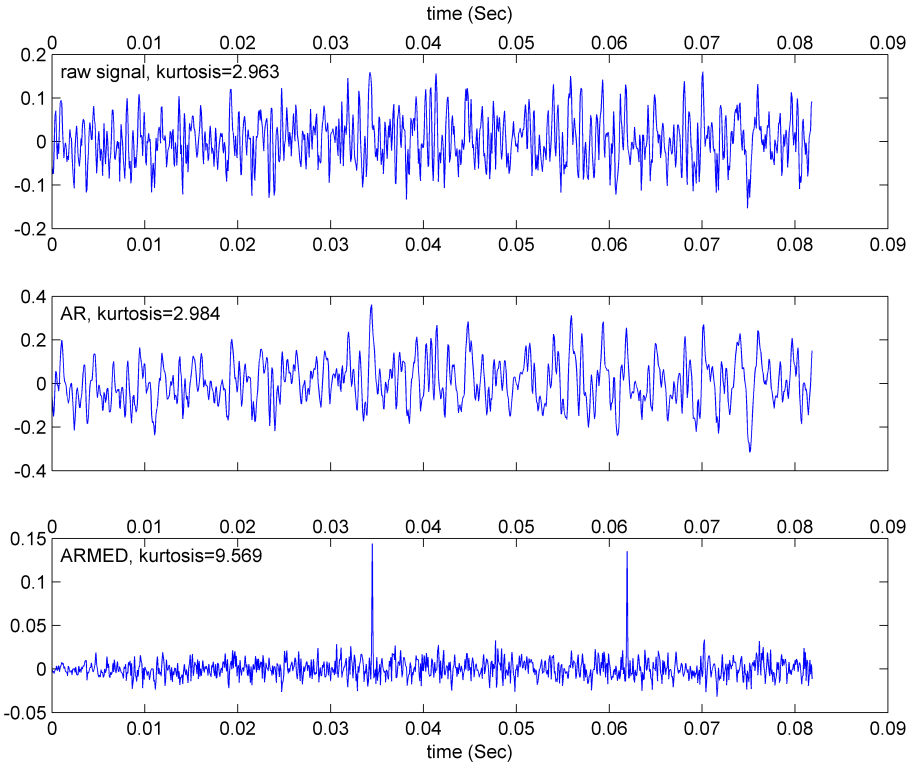

Figure 4.10: AR-MED filtering and kutosis for TSA residual of the new data: TSA residual; after applying AR filter on TSA residual; after applying AR-MED filter on TSA residual

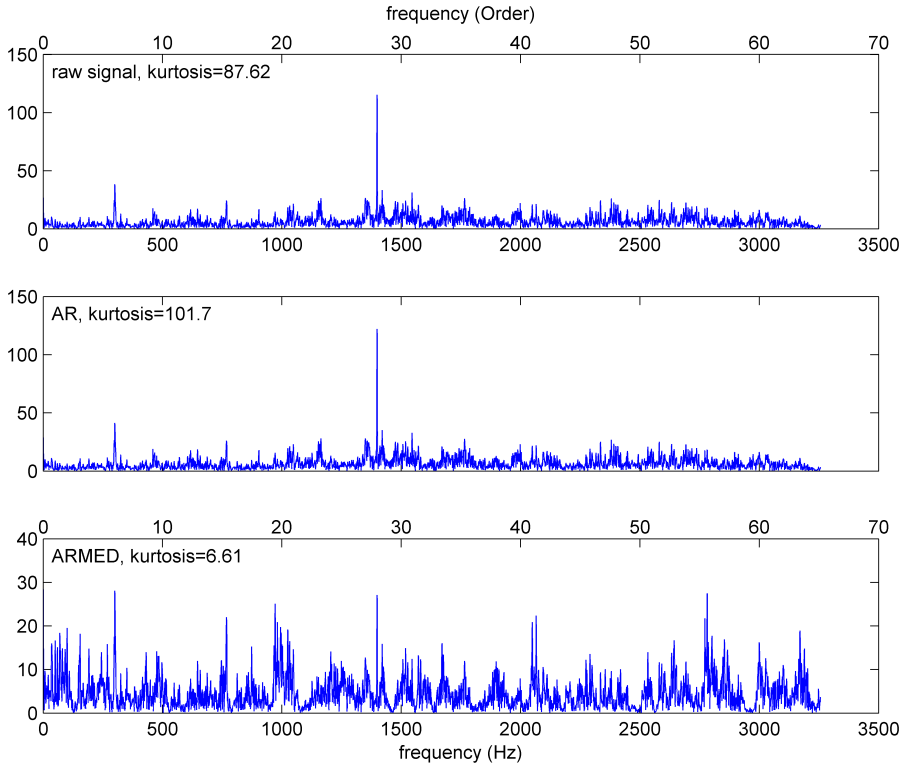

Figure 4.11: FFT of TSA residual for the old data: FFT of TSA residual; FFT after applying AR filter on TSA residual; FFT after applying AR-MED filter on TSA residual

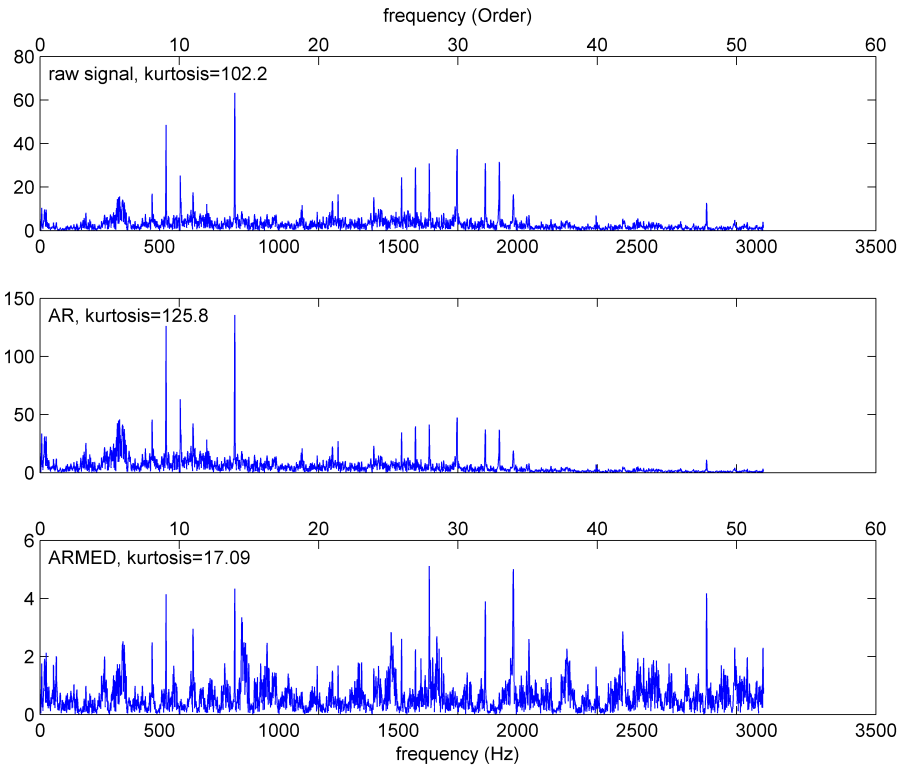

Figure 4.12: FFT of TSA residual for the new data: FFT of TSA residual; FFT after applying AR filter on TSA residual; FFT after applying AR-MED filter on TSA residual

4.8 Cepstrum

Cepstrum editing can be used as an alternative of TSA by setting zero the rahmonic correspond to shaft speed in the cepstra. In fact, the strongest terms in the cepstrum domain (qufrencies) represent the most spread harmonics in the frequency domain. By removing such a strong qufrencies we will actually eliminate the discrete frequencies from the spectrum. By applying this technique and removing a different number of strong terms, the result spectrum is more impulsive and non-deterministic. The residual signal shows many dominant integer and half harmonics of shaft speed. It should be mentioned that the word **liftering** as a dual of filtering refers to cepstrum editing.

The logarithm term helps the spectrum to be flat in the cepstrum formula and it makes all the harmonics to be almost with an equivalent effect and to be more collected just on one line. However in practice, a comb lifter rather than one line lifter is used for eliminating the shaft speed harmonics or other harmonics from the cepstrum. In the old data, there is no spread harmonics of shaft speed and in the new data, the sidebands and other components around the shaft speed make a triangle shape and almost hide the central frequency. So, by

using this method we do not see a significant change in the spectrum.

An exponential short-pass lifter makes the harmonics more removed. By applying this method after eliminating the strong qufrencies, the signal becomes more impulsive and non-stationary.

Also, a rectangular short-pass lifter (a dual term for low-pass filter in spectrum) is a choice to make signal more deterministic and in contrast a proper long-pass lifter makes the signal more random. In practice, this method does not separate random and discrete components from our data.

By checking all the techniques mentioned above, the best results show that the new data has strong multiple harmonics of the shaft speed, also it includes low power half harmonics of shaft speed. FFT of the strongest terms of cepstrum have almost full harmonics of shaft speed with a low amplitude sideband around the all. In contrast to the residual signal, this FFT seems similar to spectral envelope of FFT of the raw signal.

The old data just includes two vibration at about $6X$ and $28X$.

We do not know any information about the structure of the machine, ball bearing or gearbox under test, so we can not specify what the frequencies are about; but we know the complex of $1X, 1.5X, 2X, 2.5X, ...$ can be in connection with mechanical looseness.

Figures 4.13,4.14 and 4.15,4.16 show cepstrum and FFT of the old and new data after cepstrum editing. After keeping the strongest terms of cepstrum and removing the other terms the signal will have more repetitive form in the frequency domain. The spectrum of the residual will be more random in frequency domain. We observe much difference between the old and new data after keeping just the strongest terms of cepstrum.

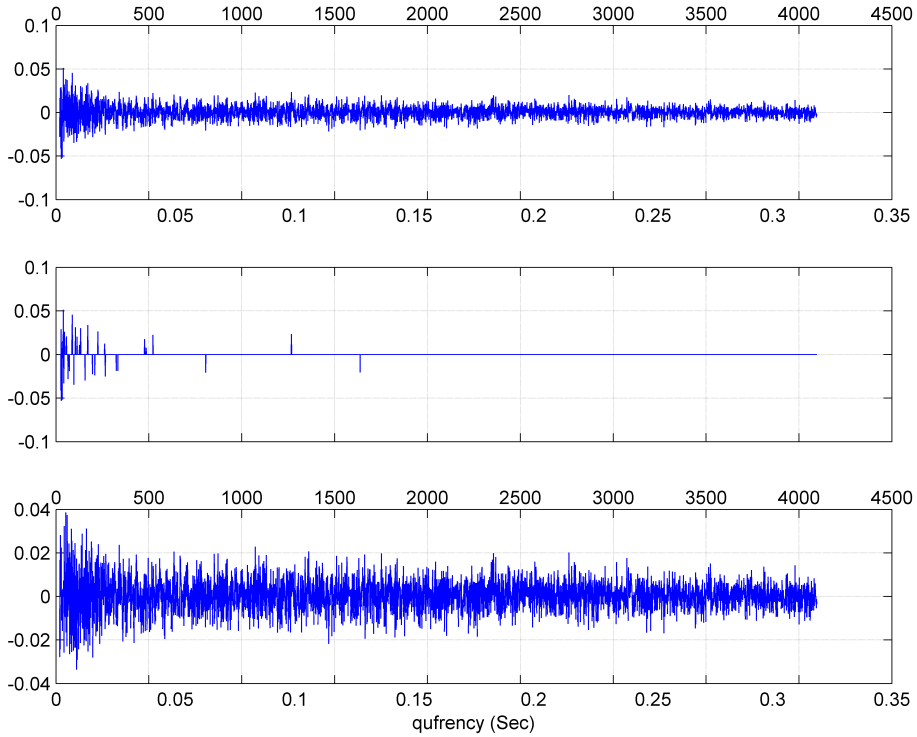

Figure 4.13: Cepstrum of the old data: raw signal; the strongest terms of cepstrum; the rersidual

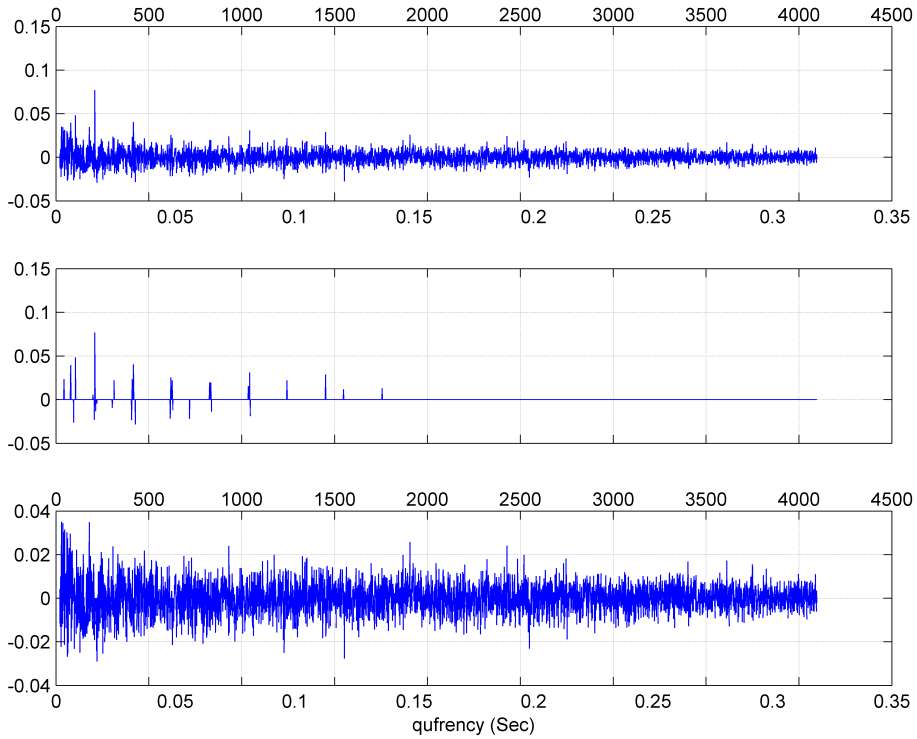

Figure 4.14: Cepstrum of the new data: raw signal; the strongest terms of cepstrum; the rersidual

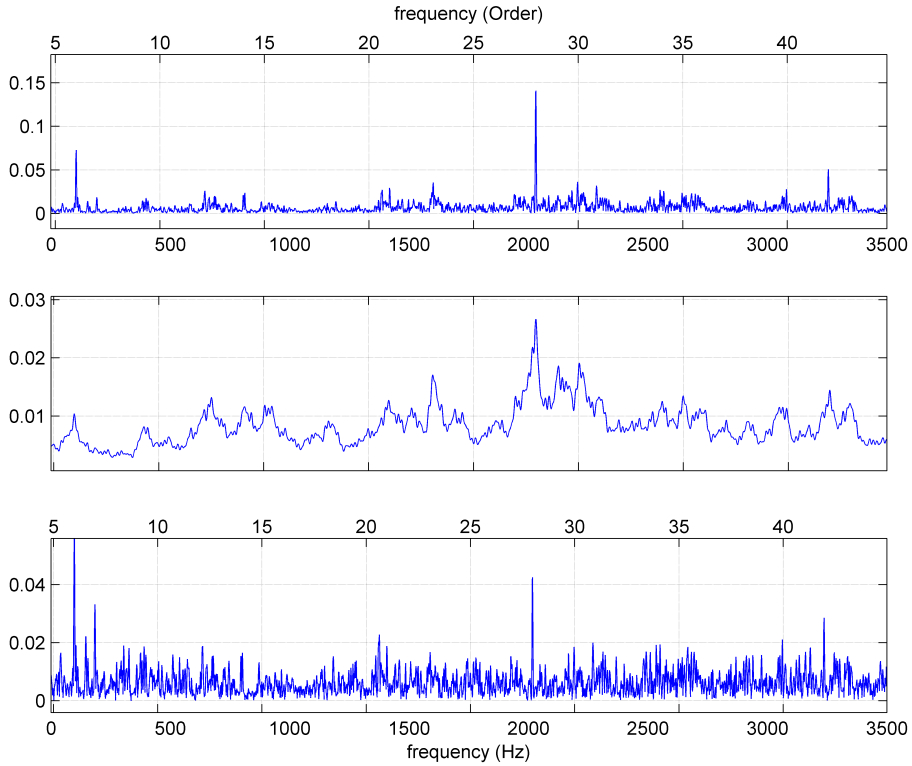

Figure 4.15: FFT of the Cepstrum of the old data: raw signal; FFT of the strongest terms of cepstrum; FFT of the rersidual

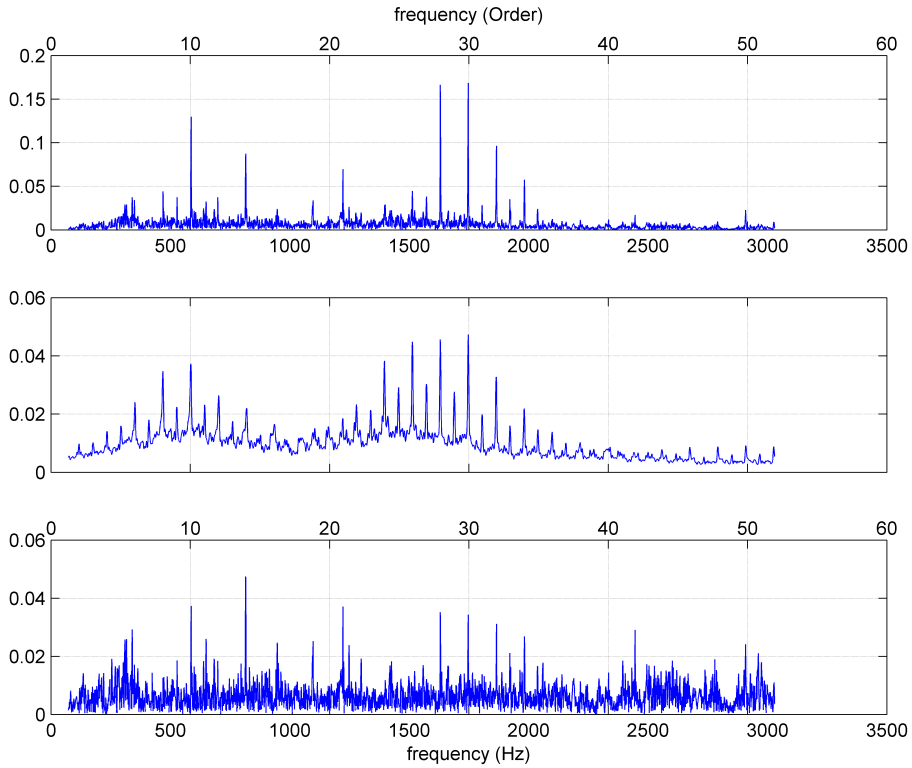

Figure 4.16: FFT of the Cepstrum of the new data: raw signal; FFT of the strongest terms of cepstrum; FFT of the residual

The results from AR-MED filtering are confirmed by cepstrum analysis. In both of the

methods we tried to separate the deterministic and non-deterministic components of the signal in order to detect the impulses produced by the faults as amplitude modulated in the resonance frequencies.

MED filter is an effective way which tries to minimize the width of path response function as much as an impulse function. On the other hand, in cepstrum analysis we know the cepstrum of the path transmission function is added to source cepstrum and cepstrum editing is an easy way to separate them from each other. So, MED and some type of liftering can have a similar effect. However, by applying cepstrum editing on the MED signal is achieved more randomness in the residual signal and more clarity at the deterministic part as can be seen in Figures 4.17,4.18.

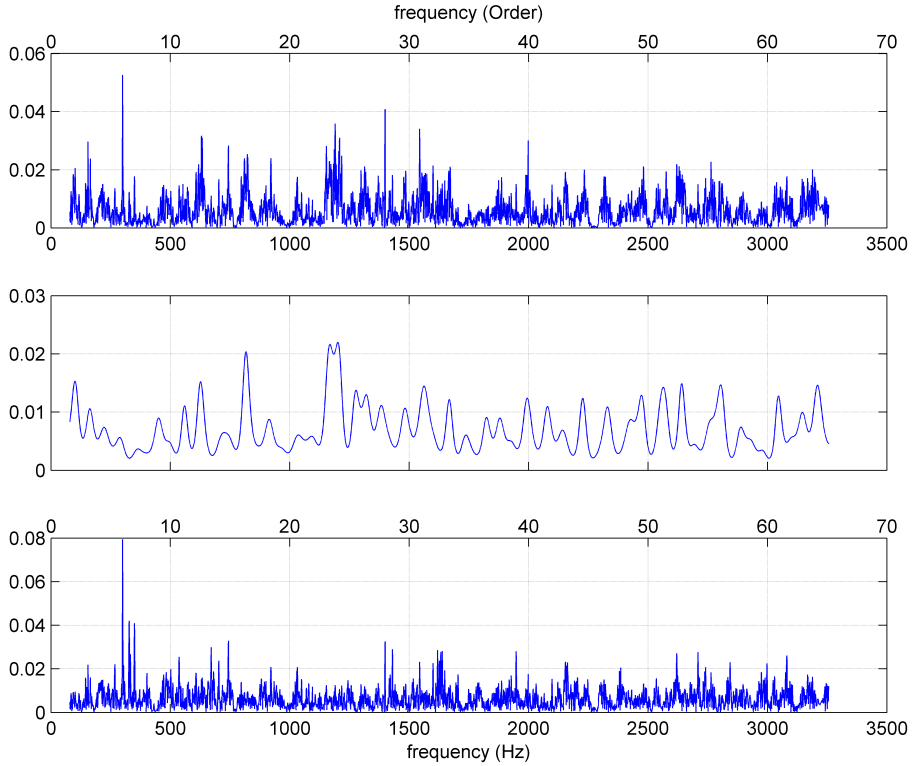

Figure 4.17: FFT of the MED signal for the old data: FFT of the strongest terms of the cepstrum of the MED signal; FFT of the residual after the cepstrum editing of the MED signal

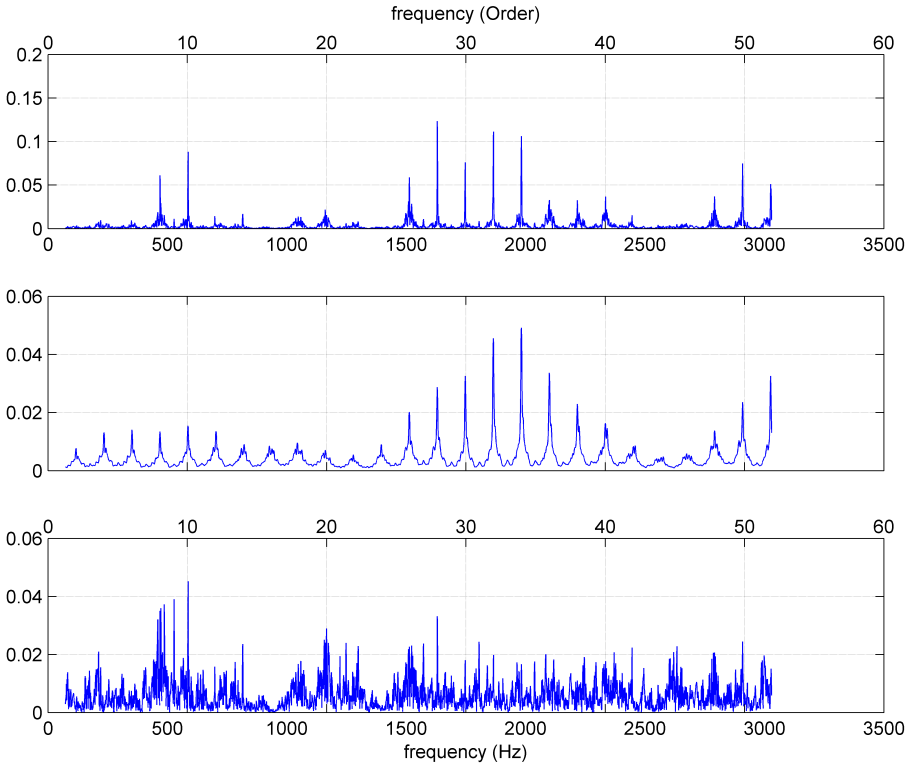

Figure 4.18: FFT of the MED signal for the new data: FFT of the strongest terms of the cepstrum of the MED signal; FFT of the residual after the cepstrum editing of the MED signal

4.9 Time-Frequency Band Selection and Envelope Analysis

Envelope analysis and high frequency resonance technique is a fundamental and effective method for fault detection. However, first step is to find the high frequency band that contains strong impulse signals as sidebands produced by faults. The conventional methods use a series of band-pass filters with equal bandwidth or octave-bands which are arranged next together. For non-stationary signals the Short Time Fourier Transform (STFT), Wavelet Packet Transform (WPT) and Empirical Mode Decomposition (EMD) can be more appropriate than just band-pass filtering. Finding some strong peaks in the frequency domain is desired as a target. Kurtosis Spectrum (SK) and Kurtogram are the methods that lead us directly into a short time-frequency segment with maximum kurtosis.

It is mentioned there are two main research directions during 21_{th} century in bringing bearing fault frequencies. The first one is to enhance the impulse signals produced by faults that is achieved by using the MED technique and the second one is achieved by using the Spectral Kurtosis (SK).

4.9.1 Band-Pass Filtering

In this study after some trial and error, the bandwidth $300 Hz$ is used for the band-pass filters. It is observed that the new data includes some shaft speed harmonics modulated in higher frequencies and in the old data there is no such evident modulation sidebands. The results are shown in Figure 4.19.

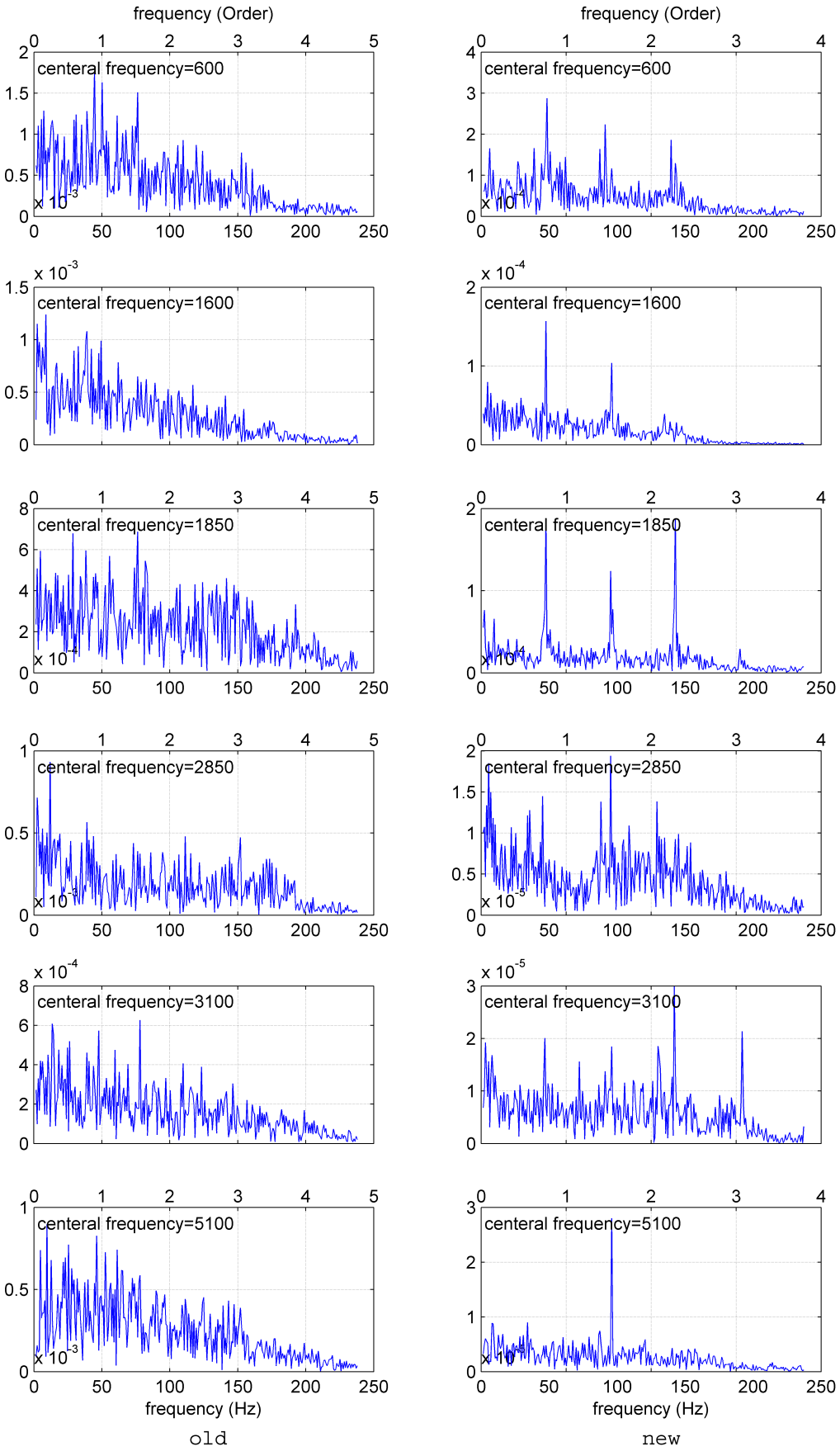

Figure 4.19: Bandpass filtering and envelope analysis. In the new data there is some electrical sidebands on the carriers of 600, 1600, 1850, ...

4.9.2 Wavelet

Although wavelet transform is capable of analysing nonlinear and non-stationary signals and considered suitable for vibration-based fault diagnosis, there are many deficiencies in the use of wavelet transform including the interference terms, border distortion and energy leakage. These deficiencies may generate a lot of undesired small spikes all over the frequency scales and make the results confusing and difficult to be interpreted. A large difference between the new and old data can be seen with respect to scalogram of them shown in Figures 4.20 and 4.21. In the new data, energy has been concentrated in two frequency bands that certainly have been generated by defects. While on the contrary, there is no focusing of energy in the old data.

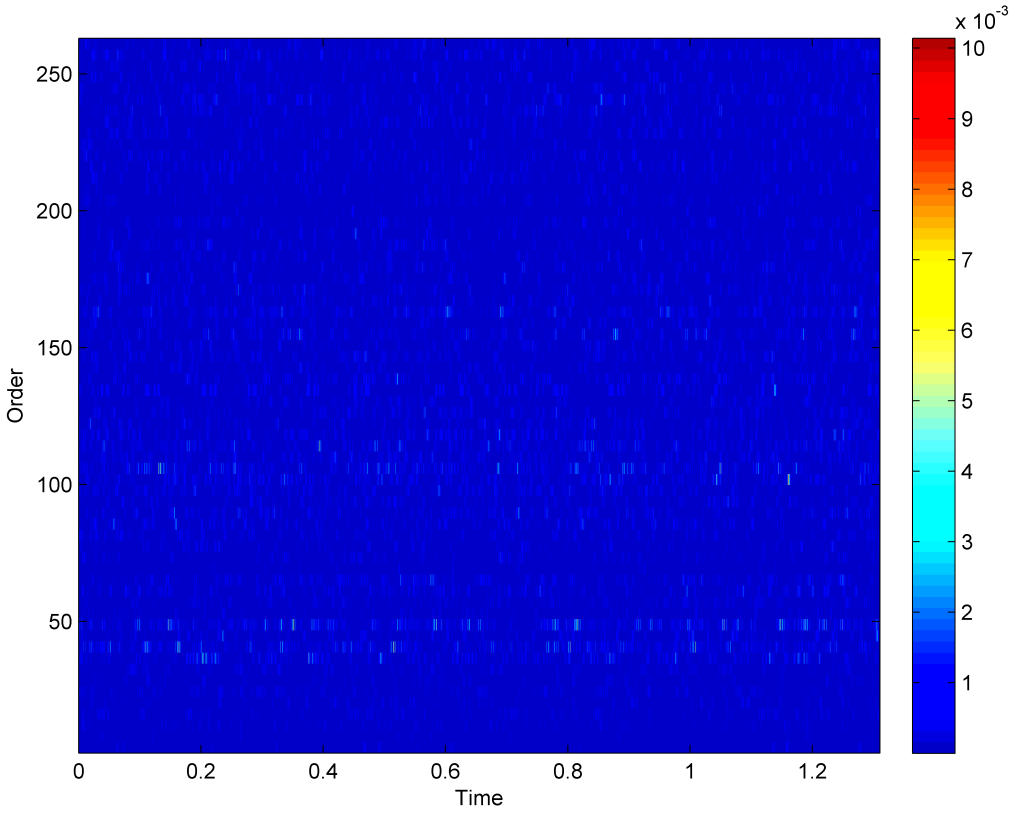

Figure 4.20: Scalogram of the old data

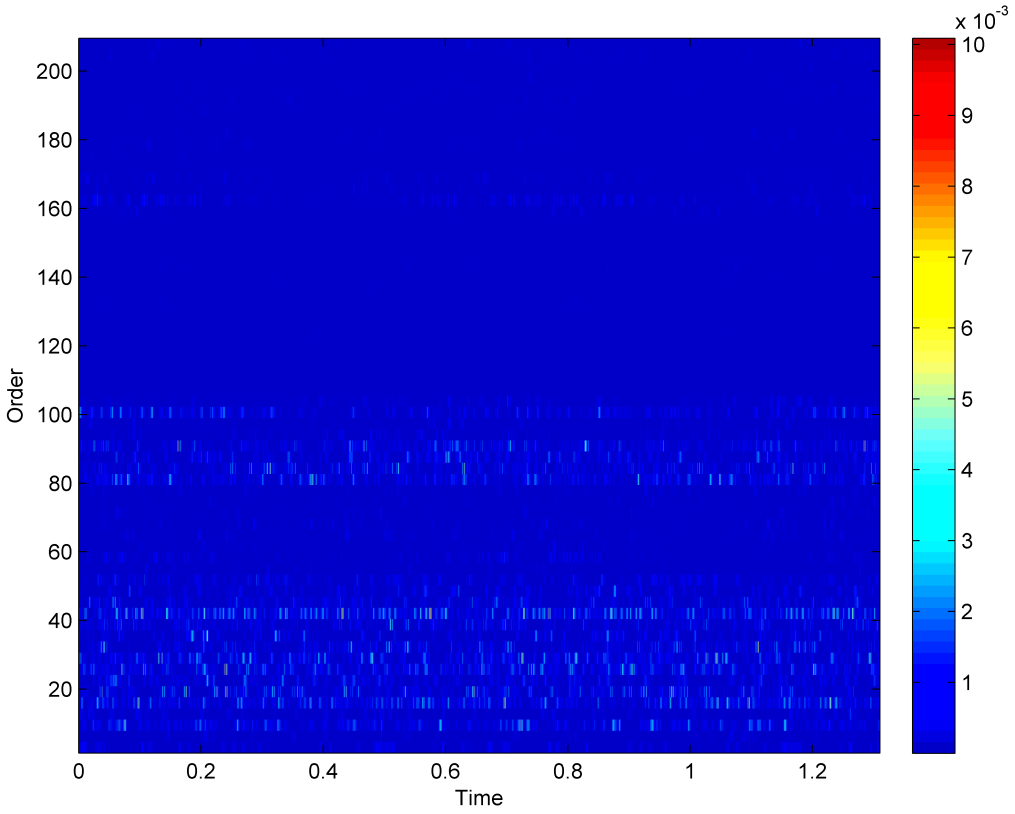

Figure 4.21: Scalogram of the new data

4.9.3 HHT

As the same as wavelet, HHT may become a successful method to extract the properties of non-stationary signal, but HHT also suffers from a number of shortcomings. First, the EMD will generate undesirable IMFs at the low-frequency region that may cause misinterpretation to the result. Second, the (first) obtained IMF may cover a very wide frequency range that the property of monocomponent cannot be achieved. Third, the EMD operation cannot separate signals that contain low-energy components. Figure 4.22 shows the IMFs for a frame of old data.

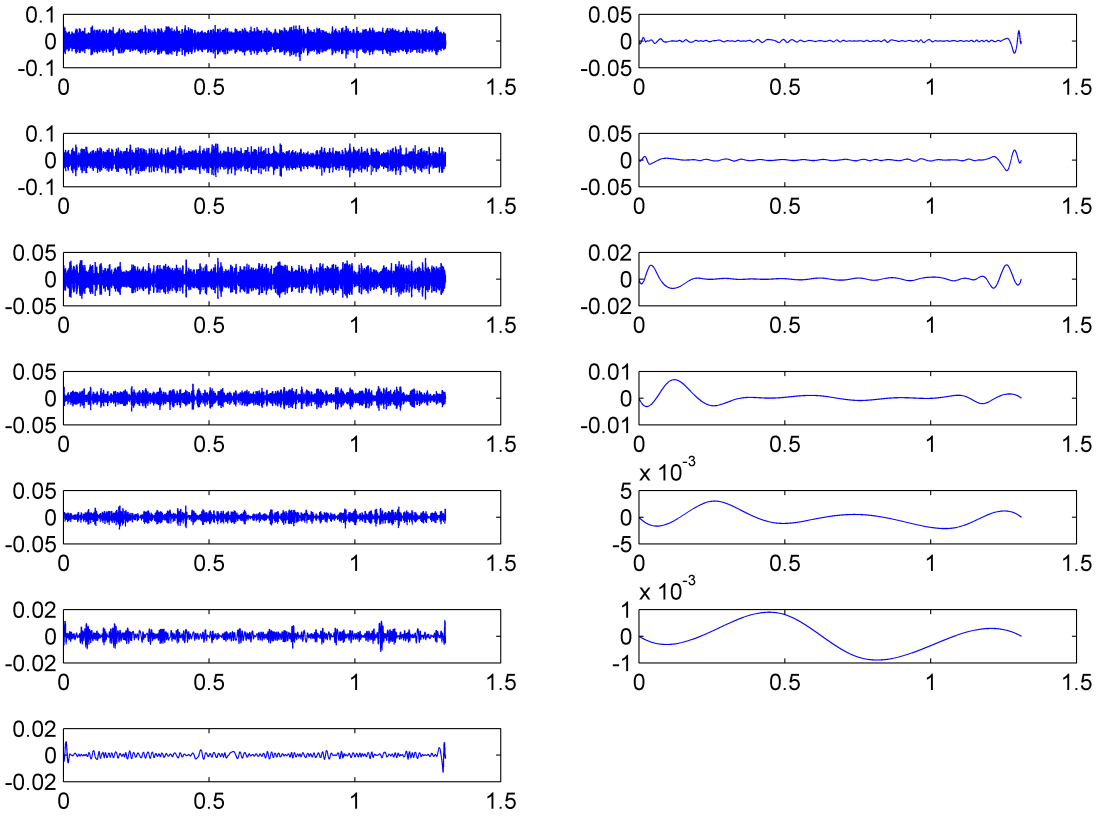

Figure 4.22: Empirical mode decomposition and 13 IMFs of the raw signal from old data set

4.9.4 Improved-HHT

As an improved HHT method, the wavelet packet transform (WPT) is used as preprocessing to decompose the signal into a set of narrow band signals prior to the application of EMD. With the help of WPT, each IMF derived from the EMD can truly become monocomponent. Then, a screening process is conducted to remove unrelated IMFs from the result. The screening can be achieved by calculating the correlation of the IMFs with the inspected raw signal. Unrelated IMFs that may cause distortion to the results, particularly in low-frequency range as mentioned in the first shortcoming, can be minimized.

In this study, firstly, the raw signal is decomposed to eight components by running a 3-levels db6 wavelet packet transform. Then empirical mode decomposition is performed for each components and then Hilbert spectrum is calculated for each one that is shown in Figures 4.23,4.24 and 4.25,4.26 . Totally it is obtained more than 100 IMFs for the old and the new data (more than 12 IMFs for each wavelet component), but by considering only the IMFs

with a correlation bigger than one tenth of the maximum, the total acceptable IMFs for the old data and the new data will be typically 17 and 22 respectively. Instantaneous amplitude, instantaneous frequency and Hilbert transform of each IMF are combined to make Hilbert spectrum in the number of 100 frequency bands for each wavelet components. The beautiful discrete lines in all figures show exactly the present frequencies in the time slot. Specially, the components with a red line are considered to perform other signal analysis for a detailed inspection. The meaning of the lines is almost quite dependent to the machine structure. The fluctuation of lines can come from the error in the phase and angle calculating method, inaccuracy computation, and presence of frequency modulation in the data. Discrete lines show that the corresponding IMF is fine mono-component. In contrast, the dots are related to the multi-component IMFs.

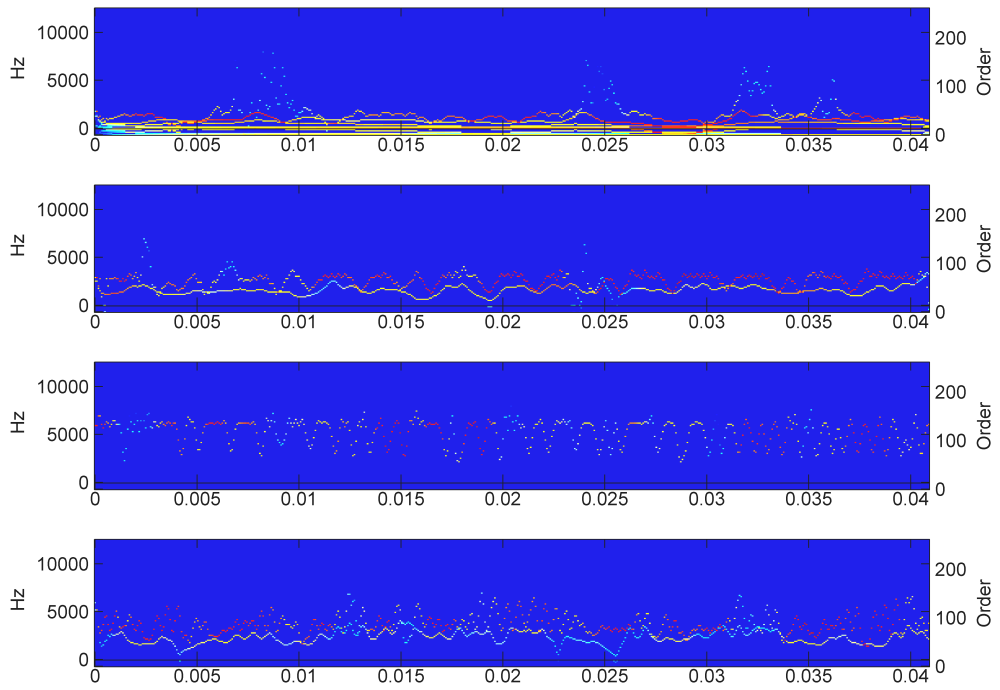

Figure 4.23: Improved HHT on the old data: applying 3-levels WPT on the raw signal and then performing HHT on the componenets 1-4

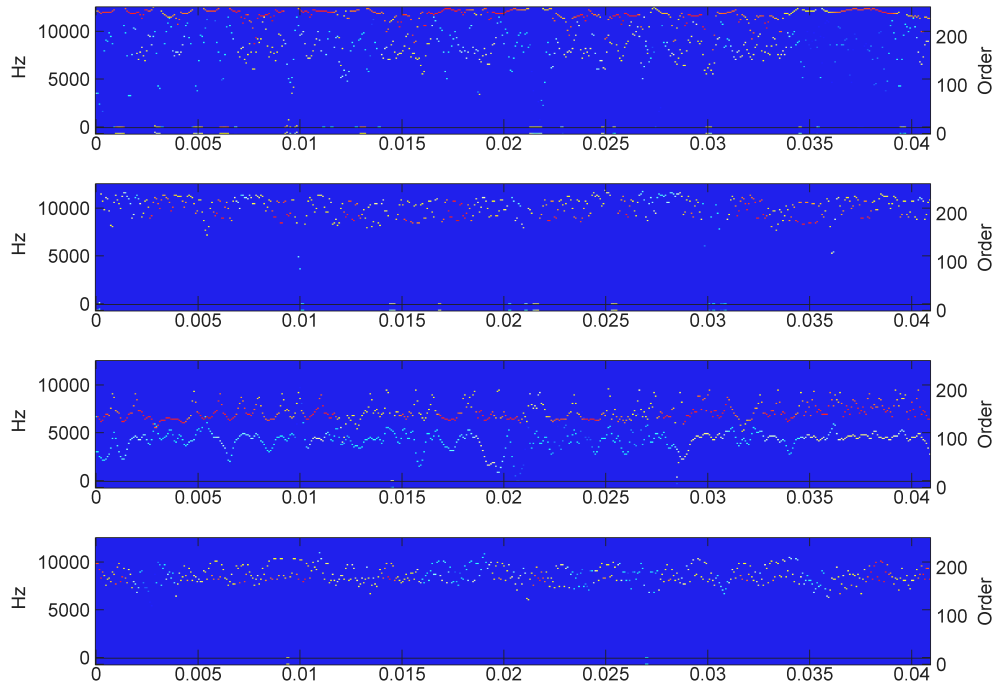

Figure 4.24: Improved HHT on the old data: applying 3-levels WPT on the raw signal and then performing HHT on the componenets 5-8

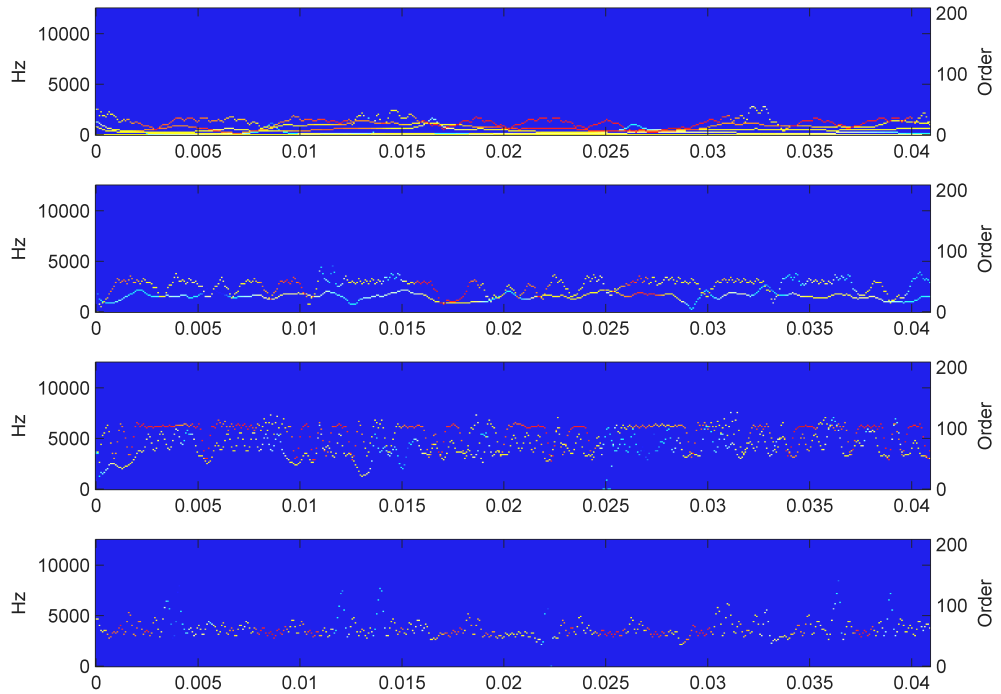

Figure 4.25: Improved HHT on the new data: applying 3-levels WPT on the raw signal and then performing HHT on the componenets 1-4

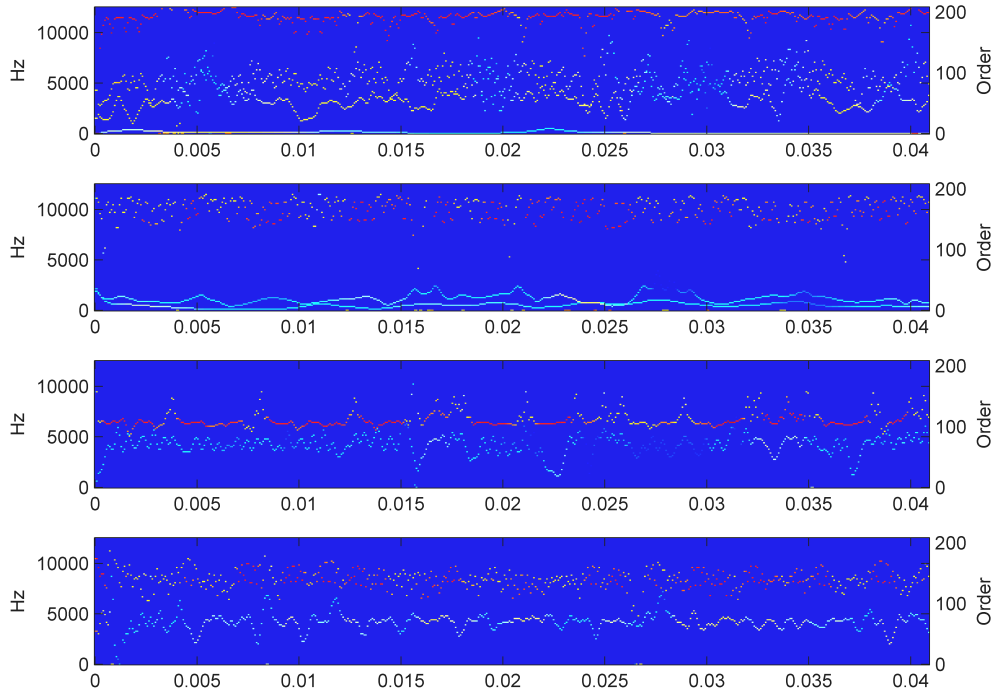

Figure 4.26: Improved HHT on the new data: applying 3-levels WPT on the raw signal and then performing HHT on the componenets 5-8

In the following figures very much impulses are visible, but clearly the impulses in the new data is much more. In the old data we still observe $6X$ and $28X$ impulses but a wide range of impulses can be seen in the new data.

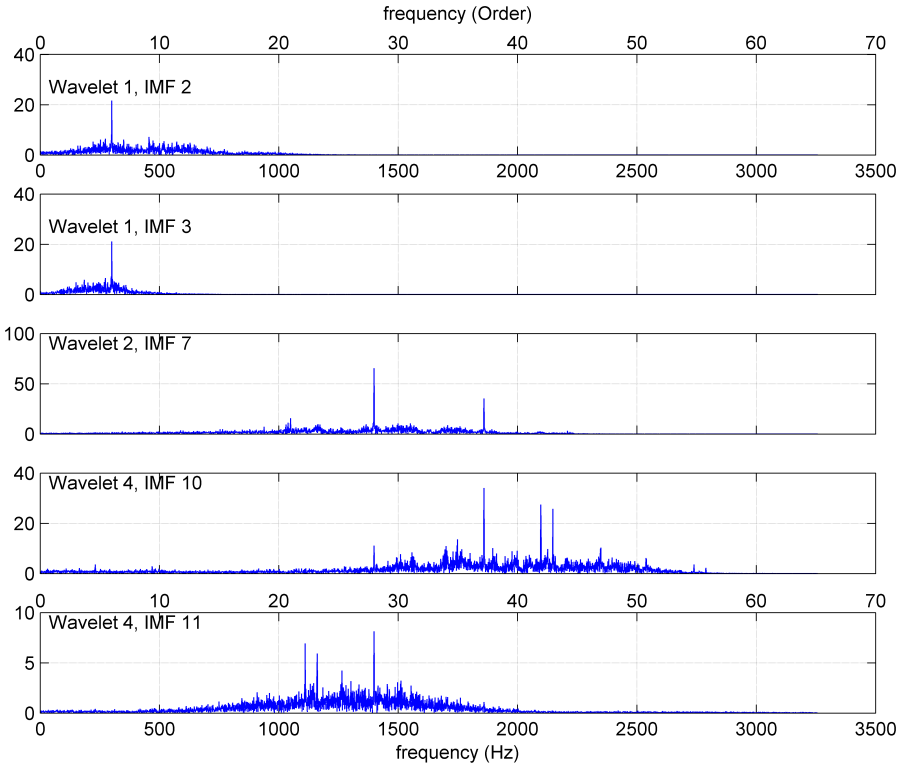

Figure 4.27: Some selected IMFs in improved HHT for the old data

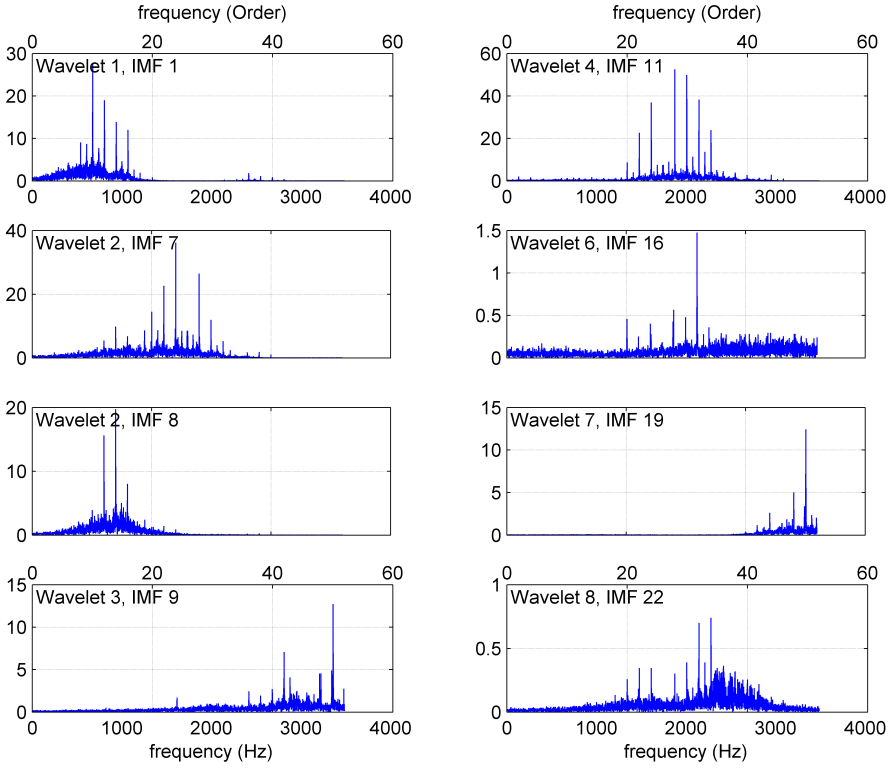

Figure 4.28: Some selected IMFs in improved HHT for the new data

4.9.5 Spectral Kurtosis

Spectral kurtosis was used as a method to find the best candidate for band-pass filtering and envelope analysis. Figure 4.29 shows spectral kurtosis of a MED filtered signal and the frequency band with maximum kurtosis that is achieved by band-pass filtering. The third subplot shows the spectrum of envelope signal. At the last plot the band-pass filtered signal along the envelope is shown. By applying the method for ten largest kurtosises we found out several carrier frequencies along the discrete line as sideband.

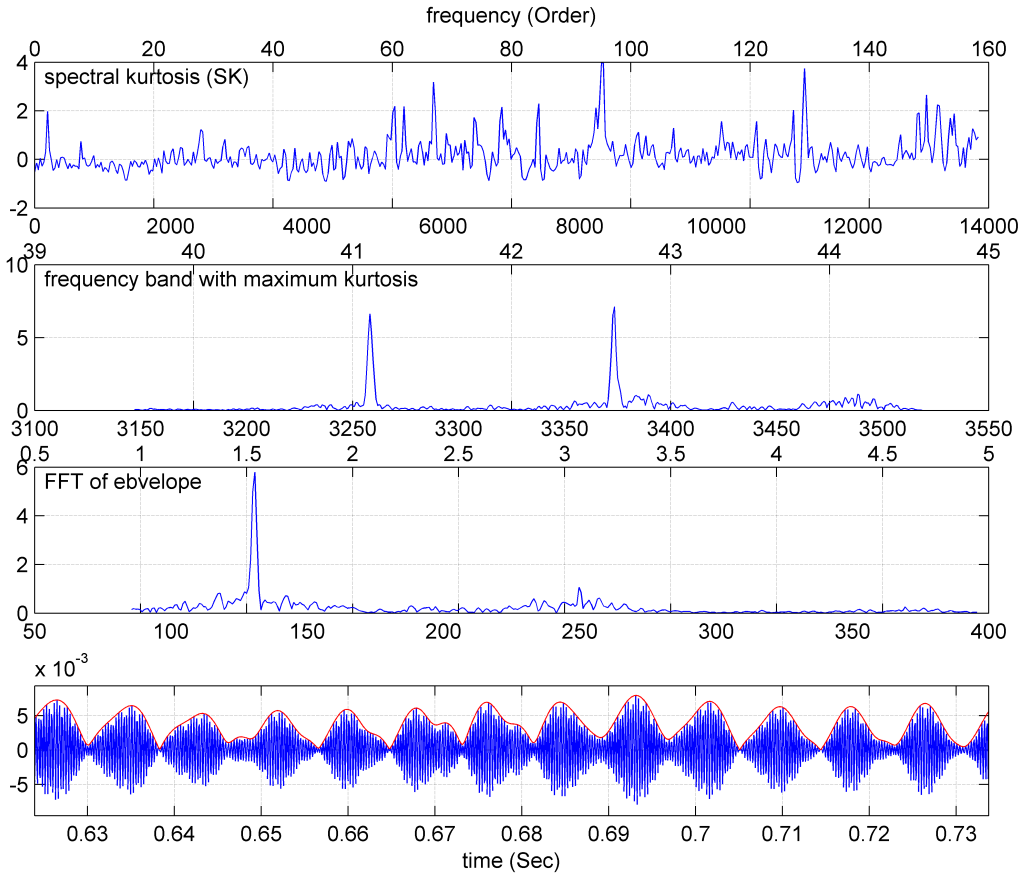

Figure 4.29: Spectral kurtosis and envelope analysis for a MED filtered signal from the new data

Bibliography

[1] J. Antoni. Unsupervised noise cancellation for vibration signals : part I — evaluation of adaptive algorithms Unsupervised noise cancellation for vibration signals :. (JANUARY), 2004. doi: 10.1016/S0888-3270(03)00012-8.

[2] J. Antoni and R. B. Randall. Unsupervised noise cancellation for vibration signals: Part II - A novel frequency-domain algorithm. *Mechanical Systems and Signal Processing*, 18 (1):103–117, 2004. ISSN 08883270. doi: 10.1016/S0888-3270(03)00013-X.

[3] T. Barszcz. Decomposition of Vibration Signals into Deterministic and Nondeterministic Components and its Capabilities of Fault Detection and Identification. *International Journal of Applied Mathematics and Computer Science*, 19(2):327–335, 2009. ISSN 1641-876X. doi: 10.2478/v10006-009-0028-0.

[4] A. J. Berkhout and P. R. Zaanen. A comparison between wiener filtering, kalman filtering, and deterministic least squares estimation*. *Geophysical Prospecting*, 24(1): 141–197, 1976. ISSN 1365-2478. doi: 10.1111/j.1365-2478.1976.tb00390.x. URL http://dx.doi.org/10.1111/j.1365-2478.1976.tb00390.x.

[5] A. Boggess and F. J. Narcowich. *A First Course in Wavelets with Fourier Analysis*. Prentice Hall, 2001. ISBN 9780130228093. URL https://books.google.no/books?id=jF8nAQAAIAAJ.

[6] P. Borghesani, P. Pennacchi, R. B. Randall, N. Sawalhi, and R. Ricci. Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. *Mechanical Systems and Signal Processing*, 36(2):370–384, 2013. ISSN 0888-3270. doi: 10.1016/j.ymssp.2012.11.001. URL http://dx.doi.org/10.1016/j.ymssp.2012.11.001.

[7] A. Brandt and Ab. Axiom Edutech. Main Principles and Limitations of Current Order Tracking Methods. (March):19–22, 2005.

[8] L. Gelman, I. Petrunin, I. K. Jennions, and M. Walters. Diagnostics of Local Tooth Damage in Gears by the Wavelet Technology. pages 1–7, 2012.

[9] Western Geophysical. Minimum entropy deconvolution. 16:21–35, 1978.

[10] D. S. Gu and B. K. Choi. Machinery Faults Detection Using Acoustic Emission Signal. *Acoustic Waves - From Microdevices to Helioseismology*, pages 171–190, 2011. doi: 10.5772/22892. URL http://www.intechopen.com/books/acoustic-waves-from-microdevices-to-helioseismology/machinery-faults-detection-using-acoustic-emission-signal.

[11] H. Herlufsen and S. Gade. Characteristics of the Vold / Kalman Order Tracking Filter. (April):2–8, 1999.

[12] H. Huang. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. 1998.

[13] N. E. Huang, M.-L. C. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, and K. L. Fan. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 459 (2037):2317–2345, 2003. ISSN 1364-5021. doi: 10.1098/rspa.2003.1123.

[14] M. Kedadouche, M. Thomas, and A. Tahan. Monitoring machines by using a hybrid method combining MED, EMD, and TKEO. *Advances in Acoustics and Vibration*, 2014, 2014. ISSN 1687627X. doi: 10.1155/2014/592080.

[15] A. Klepka. Wavelet Based Signal Demodulation Technique for Bearing Fault Detection. 15(4):63–71, 2012.

[16] E. Kvedalen. University of Oslo Department of Informatics. *Signal Processing*, 9120(May):121, 2003. URL http://folk.uio.no/eivindkv/ek-thesis-2003-05-12-final-2.pdf.

[17] D. H. Kwak, D. H. Lee, J. H. Ahn, and B. H. Koh. Fault Detection of Roller-Bearings Using Signal Processing and Optimization Algorithms. pages 283–298, 2014. doi: 10.3390/s140100283.

[18] E. V. K. Madisetti, D. B. Williams, and S. C. Douglas. Douglas , s . c .introduction to adaptive filters. 1999.

[19] C. Nalysis, S. Urrent, and A. Ignature. Advanced Spectral Analysis. 1985.

[20] G. K. Nilsen. Recursive time-frequency reassignment. *UiB, Master Thesis*, 2007.

[21] Z. K. Peng, Peter W. Tse, and F. L. Chu. An improved Hilbert-Huang transform and its application in vibration signal analysis. *Journal of Sound and Vibration*, 286(1-2):187–205, 2005. ISSN 0022460X. doi: 10.1016/j.jsv.2004.10.005.

[22] R. B. Randall. *Vibration-based Condition Monitoring*. 2011. ISBN 9780470747858.

[23] R. B. Randall, B. Peeters, J. Antoni, and S. Manzato. New cepstral methods of signal preprocessing for operational modal analysis. *Proceedings of the International Conference on Noise and Vibration Engineering ISMA 2012*, pages 755–764, 2012.

[24] G. Rilling, P. Flandrin, P. Gon, and D. Lyon. on Empirical Mode Decomposition and Its Algorithms. *IEEEEURASIP Workshop on Nonlinear Signal and Image Processing NSIP*, 3:8–11, 2003. URL http://perso.ens-lyon.fr/patrick.flandrin/NSIP03.pdf.

[25] N. Sawalhi and R. B. Randall. THE APPLICATION OF SPECTRAL KURTOSIS TO BEARING DIAGNOSTICS. (September 2015).

[26] N. Sawalhi and R. B. Randall. Semi-Automated Bearing Diagnostics - Three Case Studies. pages 545–555, 2007.

[27] N. Sawalhi and R. B. Randall. Localized fault detection and diagnosis in rolling element bearings : A collection of the state of art processing algorithms. (Hums), 2013.

[28] C. Scheffer and P. Girdhar. Machinery Vibration Analysis & Predictive Maintenance. *First published, Elsevier*, 2004. URL http://medcontent.metapress.com/index/A65RM03P4874243N.pdf$\delimiter"026E3B2$nhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Machinery+Vibration+Analysis+&+Predictive+Maitenance#1.

[29] R. H. Shumway and D. S. Stoffer. Time Series Analysis and Its Applications.

[30] B. Widrow and S. D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1985. ISBN 0-13-004029-0.

[31] X. Zhang, J. Kang, L. Xiao, J. Zhao, and H. Teng. A New Improved Kurtogram and Its Application to Bearing Fault Diagnosis. 2015.