5. Fuzzy-Statistical Reasoning in Fault Diagnosis

Dan Stefanoiu and Florin Ionescu

When searching for faults threatening a system, the human expert is sometimes performing an amazingly accurate analysis of available information, frequently by using only elementary statistics. Such reasoning is referred to as "fuzzy reasoning," in the sense that the expert is able to extract and analyse the essential information of interest from a data set strongly affected by uncertainty. Automating the reasoning mechanisms that represent the foundation of such an analysis is, in general, a difficult attempt, but also a possible one, in some cases. The chapter introduces a nonconventional method of fault diagnosis, based upon some statistical and fuzzy concepts applied to vibrations, which intends to automate a part of human reasoning when performing the detection and classification of defects.

5.1. Introduction

Nowadays, the classical fault tolerant design paradigm is enriched by new methods and techniques (Wilsky, 1976; Reiter, 1987; Isermann, 1993; 1997). The trade-off between costs involved by ignoring fault prevention and costs of hyper-safety of systems is improved. The effort in designing satisfactory modules aptly to prevent failures is decreased, due to important technological advances. In a complete structure of fault detection and diagnosis, a module concerned with monitoring of system symptoms and anticipation for possible failures is included. In general, the symptoms are detected by using two kinds of methods: *analytical* and *heuristic*.

The *analytical methods* are involved with systems for which the characteristic parameters are measurable (or quantifiable). These parameters are determined by analysing either some signals or the system itself. For instance, the basic parameters of monitored signals are: the amplitude, the variance, the auto-correlation, the power spectral density, etc. Basically, the system analysis is founded on an identification model, in general parametric (Söderström and Stoica, 1989). Various models are used, such as: (auto)regressive, state representation, described by some parity equations, etc. The model parameters are deduced from measured input-output data by system identification techniques. In both cases, a quantitative expertise has to be performed. This consists mostly of comparisons between the measured values and a set of *tolerated* values assigned to normal behaviour of the system. The malfunction symptoms appear when the parameters start to systematically provide values beyond tolerances. Moreover, a classification of symptoms can be realized, depending on the difference between the measured and tolerated values.

Sometimes, the analytical approach is not sufficient or cannot be performed (especially because the characteristic parameters are not quantifiable). Moreover, the symptoms meaning is important for interpretation of associated faults. Often, this relies on the qualitative assessment of a human operator as expert. The expert experience plays an important role in symptoms investigation. For this reason, one says that the detection of symptoms is performed by using *heuristic methods* (from *heuriskein* (Greek) – to search, to investigate). The nonquantifiable information observed from the system could be reflected for example by: colours, smells, noise tones, etc. However, some quantifiable parameters, but with "fuzzy" values, represented by linguistic terms like: "small," "medium," "large," "about null," etc., belong to this category as well. The human operator integrates this information in a quasi-empirical history of system functioning. Qualitative comparisons are performed between the observed information and the information specified by the history. The history includes not only information about the normal functioning states, but also about the maintenance process, repairs, fault types, life-time, fatigue, etc. The decision concerning the symptoms and faults is based on operator's skills, experience or flair and is affected by uncertainty. However, the experience about the system can be improved through a learning mechanism.

Like in medicine, fault prevention remains a demanding task that requires both *self-anticipation* from the system and *intelligent approach* from the user. Usually, a self-anticipatory system transmits information about its behaviour through some *anticipating signals*. For example, human or animal muscles have different electrochemical activity just before they are damaged, due to high intensity and long effort (von Tscharner, 2000). Another example is issued from mechanical systems, for which the vibrations are anticipating signals (Angelo, 1987; Bedford and Drumheller, 1994; McConnell, 1995; Wowvk, 1995). Their intimate structure changes some time before a failure occurs (Braun, 1986). But this change is so fast and sometimes so difficult to distinguish that, without special detection and decoding techniques, it could be ignored. These techniques focus on the extraction of vibration main characteristics (features), in order to classify the possible faults. In general, the strategy adopted within a fault detection method starting from vibrations consists of the following stages: signal acquisition, signal analysis (in order to extract features), features grouping, faults classification (eventually adaptively, through a continuously learning mechanism), fault identification (if present).

Vibration acquired from mechanical systems is interesting mainly for its capacity to encode information about the defects or faults threatening them. Several distinct efforts in detection of machinery defects can be noticed, but only in the last few decades has vibration become crucial for automating this process. The earliest method, which dates back to the first days of machinery (and which is still in use today), is founded on a trained observer or listener referred to as (*expert*) *analyst*. A person with a great deal of experience in working with a particular machine or engine can detect flaws in operating machinery, by simply "watching" or "listening" to it. Very often, the resulting diagnosis, based on empirical observations and deductions, is amazingly accurate, but difficult to model. Other subsequent attempts became more systematic and used some parameters, such as: the lubricant temperature (which, unfortunately, provides too late a diagnosis, after

the defects are already severe), the oil cleanness (which requires an exhaustive and often inefficient analysis), the noise level of acoustic emission (which is often enabled only by already fatigued elements), etc.

The most efficient methods in early detection of defects are using signal processing (SP) techniques (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996). These methods differ from many typical SP applications where the noise attenuation is a fundamental requirement. When using vibrations, exactly the noise is the most concerned part in the analysis. This is due to the fact that not only the natural oscillations of machinery could encode the defective behaviour, but also the noise corrupting them. Moreover, the applications revealed that the signal-to-noise ratio (SNR) is extremely small for vibrations encoding information about defects. Therefore, the models of vibration used in fault detection and diagnosis (fdd) are, in fact, models of their noisy parts, encoding all the information about defect types and their severity degrees.

One of the most interesting applications in fdd is concerned with bearings, due to their simple structure and large integration within mechanical systems (Howard, 1994; FAG OEM and Handel AG, 1996; 1997). By inspecting the spectrum of vibration acquired from bearings, some researchers believed that its irregular shape is mainly due to the environmental noise and correlation between different components. Hence, they introduced techniques to "remove" the white noise and decorrelate the data, based on SP concepts such as: *autocorrelation*, *backstrum*, or *cepstrum*, but the irregularities are only slightly smoothed and the defect severity is difficult to derive. Perhaps the most popular method to extract information about defects in bearings (and geared coupling) is the (spectral) envelope analysis (EA). Some of these techniques (especially EA) are described in (Stefanoiu and Ionescu, 2002). They are poorly modelling the humanlike diagnosis, which probably requires nonconventional approaches. Actually, one can notice that experienced analysts perform a kind of fault classification, by simply inspecting the spectrum. Moreover, they are able to improve the accuracy of classification for every new case they analyse. It is by far not completely known what kind of reasoning lies behind their diagnosis, but one has assumed that the brain performs a qualitative statistical assessment inputting some pattern recognition mechanisms towards this goal. A very interesting approach combining statistics and pattern recognition has been introduced in (Xi *et al.*, 2000). This is in fact an attempt of automating human reasoning, which resulted in a quite efficient and simple fdd algorithm, though with unavoidable limitations.

In this research, one started from the largely accepted idea that *human reasoning is also fuzzy*. This means that a solution to a problem could be issued even from unclear, vague or ambiguous information, i.e., from information strongly affected by uncertainty. Usually, the analyst considers the solution the most "plausible" one, according to the available data. When an fdd or/and classification has to be performed from vibrations, the analyst's experience is crucial for the accuracy of subsequent analysis. Unfortunately, the analyst has to cope not only with external perturbations affecting the data, but also with his/her own subjectivism when performing such an analysis. Usually, this analysis is based on some simple statistical assessments aiming to increase its objectivity. Therefore, the reasoning hidden behind data analysis could be automated by performing a

combination between *spectral statistics* and *fuzzy clustering* (in entropy sense (Klir and Folger, 1988)), which should decrease both the subjectivism and the perturbations influence. Moreover, comparisons between the tested vibration and a *standard* (defect free) vibration could be performed, without specifying from the beginning the number of classes and/or their meaning, which has to be discovered later. In fact, this approach combines analytic and heuristic points of view, in order to build a model of human reasoning when performing fdd.

The chapter is structured as follows. The fuzzy-statistical reasoning method is presented in depth in the next section, which has two main parts: the first one is devoted to vibration acquisition and preprocessing, whereas within the second one, the fuzzy-statistical model is described. The resulting algorithm is practically listed in Section 5.2 as well, simultaneously with the method description. The simulation results and their interpretations are given in Section 5.3. The graphical simulations are presented in the Appendix. Some concluding remarks complete the chapter.

5.2. The Fuzzy-Statistical Reasoning Method

One (but probably not unique) way to overcome some fdd limitations when using spectral or envelope analysis is to combine the spectral representation with statistics and subsequently to use a fuzzy model aiming to minimize the diagnosis uncertainty. This approach is described next.

5.2.1. Method Overview

When measuring vibrations of a mechanical system, several signals are combined together within the resulting data, such as: natural oscillations, interference signals (due to interactions between its different parts); defect encoding noise, indicating that something is wrong with one or more of its parts and environmental noise. The *crude mechanical vibration* is converted into an *electrical vibration* signal (v) by means of a sensor connected to a transducer (which could induce slight distortions). For example, in the case of a bearing, if data v are rich enough (few thousands of rotations), the vibration spectrum $|V|$ looks like that in Figure 5.1. Two cases could be discussed here.

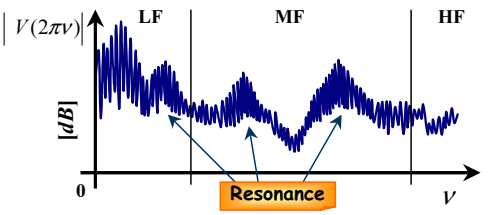


Figure 5.1. Overall vibration spectrum in case of defects.

When the bearing is defect free, the spectral energy is mainly concentrated inside the *low-frequency* band (LF) encoding information about oscillations and their natural frequencies (derived from bearing geometry, depending on shaft rotation speed). Few multiples of natural frequencies are replicated within the spectrum, but their power has an exponential decay (due to damping). In the case of defective bearing, the idea that the defect noise is basically generated by visible or microscopic quasi-random shocks has been largely accepted today. Shocks are modelled by trains of impulses and they put the sensor into resonance state. Usually, sensor resonance appears at (very) high frequency, but, by convolution with a train of impulses, it is replicated towards low frequency as well. In Figure 5.1, this is suggested by the energy concentration around some peaks located in *middle-frequency* band (MF). Usually, a resonance peak is mixed with basic LF spectrum as well, such that it could hardly be distinguished. The *high-frequency* band (HF) rather encodes information about resonance corrupted by environmental noises. The spectrum could change (even dramatically), depending on the applied load, sensor locations, shaft speed, bearing mounting, etc.

The EA principle is easy to explain now: select one of the resonance peaks, apply a bandpass filter on the vibration around the selected resonance, take the *envelope* of the resulting signal and zoom the LF part of the spectral envelope. If isolated, the defect appears now as distinctive peaks at locations depending again on natural frequencies. The higher the peaks are, the more severe the defect.

But the analyst just looks at the spectrum and provides the diagnosis by observing the changing parts relatively to the standard spectrum, though the latter has no constant shape. This means he/she is focusing on some spectral subbands that reveal significant shape and energy differences from the standard. Moreover, the similar differences are grouped in classes and each class points to a certain defect or combination of defects (with some confidence degree).

Therefore, when automating this kind of reasoning, the following operations could be involved: define a set of statistical parameters (sp) that quantify the information about shape and energy of a signal; split the spectrum into a number of subbands; compare the tested and standard subbands in terms of sp; group the results in similarity classes, by using a global fuzzy relation between them; select the best fault class, according to an entropy-based criterion aiming to minimize the information uncertainty. This constitutes the kernel of the method described hereafter. The presentation covers two main parts. The first one is concerned with vibration acquisition and preprocessing. The second one is devoted to the fuzzy-statistical model.

5.2.2. Vibration Data Acquisition and Pre-processing

Let us denote the raw vibration data by v. In practice, v is a finite length, finite bandwidth and discrete time signal encoding the information about defects that could exist within the tested component. In this case, the signal is acquired from bearings. The acquisition and preprocessing procedure encompasses several steps that are described next.

Step 1: Set the acquisition parameters.

The first parameter employed in data acquisition is the sampling rate, denoted by ν_s. The selection of ν_s is extremely important for the next analysis. On the one hand, ν_s should be large enough, in order to avoid aliasing (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996). On the other hand, large ν_s values involve expensive devices. Therefore, a suitable value should be selected, such that the resulting signal encode most part of the desired information about defects and the acquisition costs be affordable.

The sensor characteristic usually extends beyond 140–150 kHz. If defects exist, the sensor resonance is replicated towards LF and MF bands within the vibration spectrum (see Figure 5.1). At least 3 or 4 resonance peaks are located in the 0–20 kHz band and at least 2 of them lie inside the 0–10 kHz subband. In fact, the analyst focuses on this LF subband. Usually, the vibration spectrum extends beyond the limit of 20 kHz, but the band of interest remains 0–10 kHz (the SNR decreases rapidly beyond 10–12 kHz, because of HF noises that dominate the other fast decaying vibration components). All these arguments lead to the following trade-off in vibration acquisition:

a. Prefilter the sensor signal by using a *low-pass analogic anti-aliasing filter* (Proakis and Manolakis, 1996) that removes the HF components beyond 150 kHz;
b. Use the *sigma-delta modulation technique* (Proakis and Manolakis, 1996), in order to restrict the signal in the range 0–12 kHz, to attenuate the quantization noise and to avoid aliasing (a new low-pass analogic filter is applied in the end);
c. Sample the resulting analogic signal by setting a rate of at least 20–24 kHz (i.e., $\nu_s \geq 20$ kHz), according to Shannon-Nyquist Sampling Theorem (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996).

A standard sampling rate that has been employed for example in (Maness and Boerhout, 2001) is $\nu_s = 25.6$ kHz, which yields accurate vibration spectra in the range 0–12.8 kHz. Observe the powers of 2 hidden behind these values: $25600 = 2^8 \times 100$ and $12800 = 2^7 \times 100$, which avoids some computational errors due to division by multiples of 2.

Another parameter of interest is the vibration length, denoted by N. Normally, this is set according to the main rotation frequency ν_r and sampling rate ν_s. The vibration data should include a minimum number of complete rotations, n_r (usually, $n_r \geq 2000$). Then, obviously:

$$N = \lfloor n_r \nu_s / \nu_r \rfloor \tag{1}$$

For example, if $n_r = 2000$, $\nu_r = 50$ Hz (3000 rpm) and $\nu_s = 25.6$ kHz, the number of vibration data is: $N = 2^{10} \times 10^3 = 1{,}024{,}000$ samples, which takes 40 s. Usually, N is also set as a power of 2 multiple and this is the reason, in Eq. 1, ν_s is sometimes set with the same property. This setting is very useful in

evaluation of spectrum, when using a fast fourier transform (FFT) algorithm (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996).

Usually, the apparatus performing the vibration acquisition (connected to the sensor) could be tuned by only specifying these two parameters: ν_s and N or the duration of acquisition. The corresponding operations necessary to store the data in a memory are transparent for the user.

Step 2: Construct the raw vibration.

The sensor capacity of perception is determined by its bearing position. Different data could be obtained for different locations on the same bearing. When the bearing is under load, this variability is even more accentuated. This gives rise to the problem of appropriate sensor location, which is uncertain. The uncertainty could be attenuated if several sensors are located in different positions (instead of a single one). Unfortunately, in this case, other problems occur. For example, the acquired signals have to be mixed in a unique raw vibration, by synchronizing them appropriately. Another problem is that the number of sensors could increase the cost of acquisition solution. Sensors should be as light as possible, in order to introduce insignificant distortions into the genuine vibration. But, the lighter the sensor, the more expensive. Also, in general, sensors have slightly different characteristics. The bigger the sensor number, the more difficult to denoise the data. Hence, a suitable number of sensors should be employed, such that the acquired signals be easy to synchronize and the cost of acquisition be affordable.

An interesting and efficient solution is introduced in (Maness and Boerhout, 2001), as illustrated in Figure 5.2. Two sensors are employed to acquire the horizontal and the vertical vibrations, denoted by v_x and, v_y, respectively. These are, in fact, two quadrature signals easy to synchronize, by considering them the real and the imaginary part of raw vibration:

$$v \equiv v_x + jv_y \tag{2}$$

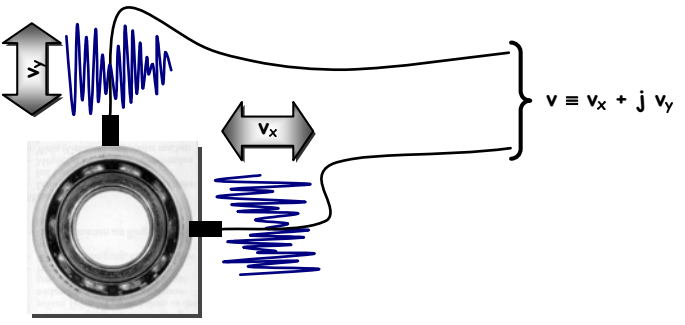


Figure 5.2. Construction of raw vibration from two quadrature signals.

The resulting signal is complex valued, but its sensitivity to sensors location is attenuated. In the absence of load, there are no significant differences between quadrature signals in terms of magnitude. If a load is applied, these

differences become important and should be accounted. In this case, Eq. 2 should be replaced by:

$$v \equiv av_x + jbv_y \tag{3}$$

where $a > 0$ and $b > 0$ are two constants selected such that av_x and bv_y have approximately the same range of variation. For example, in the case of vertical load, a could be set to 1 (no horizontal load), whereas b should be set inside (0,1) interval, since the load amplifies the defect noise of vibration.

Step 3: Vibration segmentation and windowing.
The vibration data set $\{v[n]\}_{n\in\overline{0,N-1}}$ is quite large. If the Fourier transform (FT) were to be applied on this set, the evaluation could be very slow. Moreover, the resulting spectrum is practically useless since the vibration signals are also *non-stationary* (Cohen, 1995). In other words, the spectrum is time varying. This involves the overall spectrum reflecting the intimate behaviour of vibration only on average, whereas, on the contrary, the spectrum variations are important for learning as much as possible about how the bearing runs. Therefore, the vibration segmentation becomes a necessity. In this context, one operates with two concepts: (vibration) *frames* and (vibration) *segments*.

A *frame* is a subset of successive samples that could not be further segmented. Frames could or could not be overlapped. In this approach, the frames are *nonoverlapping*, but the overlapping effect is hidden behind the concept of *segment*. One can denote by v_m the *m*-th frame of vibration (where $m \in \overline{0, M}$) and by $N_f \le N$ the frame length (constant for all frames). Obviously, the number of nonoverlapped frames is:

$$M + 1 = \lceil N / N_f \rceil \tag{4}$$

where $\lceil a \rceil$ is the smallest integer superior or equal to $a \in$ ☼. It is suitable that N_f be a divisor of N. For example, if N is a power of 2 multiple (as suggested within the previous step), then N_f could be 512, 1024, 2048, etc. For the model constructed next, one requires that $M \ge 2$ (i.e., at least 3 frames should be available). The frame length should be selected not only according to N, but also to the minimum resolution of frame spectrum (at least 400 rays for vibration in the range 0–10 kHz). The statistical part of the model constructed later is sensitive to N_f, since it determines the precision of corresponding sp.

A vibration *segment* includes three successive (nonoverlapping) frames: the *previous* frame (v_{m-1}), the *current* frame (v_m) and the *next* frame (v_{m+1}), for $m \in \overline{1, M-1}$. Thus, the vibration data could generate up to $M-1$ segments of length $3N_f$ each. Unlike frames, segments are overlapping (two of the three frames in a segment are identical within the next segment), in order to prevent marginal effects when filtering. Actually, the characteristic frame of a segment is the current one, located in the middle. Its left and right neighbours are only playing the role of

background signals, which avoids zero-padding and performs a smooth passage from a frame to another, when filtering.

The samples of neighbour signals could or could not be as important as the samples of current frame in a segment. This feature is controlled through *windowing*. The windowing technique is very simple, in fact. Let w be a $3N_f$-length window that slides along the vibration data with a step of N_f samples. Then the current segment is extracted from raw vibration by simply multiplying v and w in a certain position (0, N_f, $2N_f$, ..., $(M-2)N_f$). The sliding effect is suggested in Figure 5.3, where the window support is given by three successive frames (a segment, in fact). The window symmetry axis should be centred on the current frame middle point.

Several windows are usually employed in SP (Proakis and Manolakis, 1996). Some of them are weighting not only the neighbour frames but also some samples of central frame (like the window in Figure 5.3). The most utilized windows are the following nine, expressed next only for their N_w-length support $n \in \overline{0, N_w - 1}$, with $N_w \geq 2$.

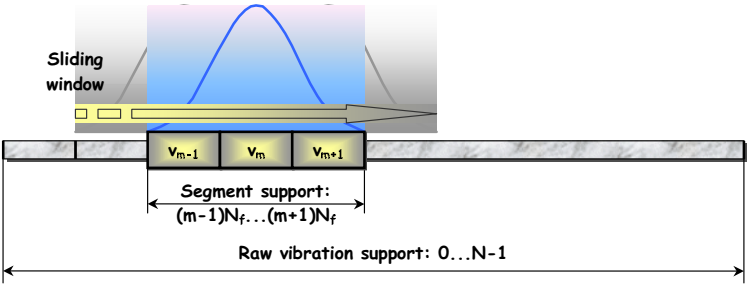


Figure 5.3. Windowing the raw vibration.

1. Rectangular (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996): $w[n] = 1$.
2. Bartlett (or triangular) (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996): $w[n] = 1 - \frac{2\left|n - \frac{N_w - 1}{2}\right|}{N_w - 1}$.
3. Blackman (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996): $w[n] = 0.42 - 0.5\cos\frac{2n\pi}{N_w - 1} + 0.8\cos\frac{4n\pi}{N_w - 1}$.
4. Chebyshev: recursive algorithm (see MATLAB function **`chebwin`**). Besides the support length (N_w), a second parameter is necessary: $r_w > 0$, which stands for the attenuation in decibels (dB) of the window spectrum side lobe with respect to the main lobe. As r_w increases, the window aperture decreases, but below 70 dB, significant marginal errors are

introduced. A good trade-off between the window aperture and its marginal errors is obtained for $r_w \in [80,100]$ dB.

5. Hamming (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996): $w[n] = 0.54 - 0.46\cos\frac{2n\pi}{N_w - 1}$.

6. Hanning (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996): $w[n] = \frac{1}{2}\left(1 - \cos\frac{2n\pi}{N_w - 1}\right)$.

7. Kaiser (Kaiser, 1974; Proakis and Manolakis, 1996): $w[n] = \frac{\sinh\left[\alpha\sqrt{\left(\frac{N_w-1}{2}\right)^2 - \left(n - \frac{N_w-1}{2}\right)^2}\right]}{\sinh\left[\alpha\frac{N_w-1}{2}\right]}$, where sinh stands for the *hyperbolic sine* ($\sinh x \overset{def}{=} \frac{e^x - e^{-x}}{2}$) and the parameter $\alpha > 0$ is the height in dB of the window spectrum side lobe. Sometimes (see MATLAB function **`kaiser`**), α is replaced by another parameter, β, defined as follows: $\beta = \begin{cases} 0.1102(\alpha - 8.7), & \alpha > 50 \\ 0.5842(\alpha - 21)^{0.4} + 0.07886(\alpha - 21), & \alpha \in [21,50] \\ 0, & \alpha < 21 \end{cases}$

As β increases, the window aperture decreases, but below $\beta = 6$, significant marginal errors are introduced. A good trade-off between the window aperture and its marginal errors is obtained for $\beta = 9$.

8. Lanczos (Proakis and Manolakis, 1996): $w[n] = \left[\frac{\sin 2\pi\left(\frac{2n - N_w + 1}{2(N_w - 1)}\right)}{2\pi\left(\frac{2n - N_w + 1}{2(N_w - 1)}\right)}\right]^L$, where the exponent $L > 0$ controls the window aperture. As L increases, the window aperture decreases, but below the unit value ($L < 1$), significant marginal errors are introduced. A good trade-off between the window aperture and its marginal errors is obtained for $L = 1$.

9. Tukey (Proakis and Manolakis, 1996):

$$w[n]=\begin{cases}1, & \left|n-\dfrac{N_w-1}{2}\right|\le\alpha\dfrac{N_w-1}{2}\\ \dfrac{1}{2}\left[1+\cos\left(\dfrac{n-(1+\alpha)\dfrac{N_w-1}{2}}{(1+\alpha)\dfrac{N_w-1}{2}}\pi\right)\right], & \alpha\dfrac{N_w-1}{2}<\left|n-\dfrac{N_w-1}{2}\right|\le\dfrac{N_w-1}{2}\end{cases}$$

where the parameter $\alpha\in(0,1)$ controls the percentage of rectangular window centred inside. For the vibration segment, a good choice is $\alpha=1/3$, since the central frame takes only one third of the whole segment.

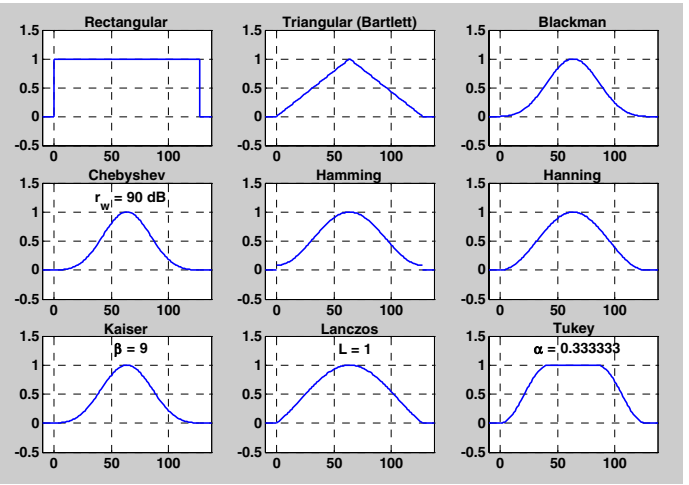


Figure 5.4. Nine of the most utilized signal processing windows.

All windows above are symmetric, as shown in Figure 5.4, where, beside the window shape, the parameter values are also depicted for Chebyshev, Kaiser, Lanczos and Tukey windows. But not all windows of this collection have the same performances when using them in SP applications. Their efficiency depends on the specific criteria that have to be matched. Although some windows seem to have the same shape, they are actually quite different. The differences are better emphasized by their spectra, as drawn in Figure 5.5. The graphics are plotted by using the spectral power expressed in dB and on all horizontal axes normalized frequencies are represented. *The main lobe* lies in LF subband, whereas the *side lobes* extend to MF and HF subbands. The main lobe is best emphasized for windows like Blackman, Chebyshev or Kaiser. (For the last two, the main lobe height relative to the first side lobe can be controlled.) One of the most employed criterions in selection of the appropriate window is the *attenuation performed by the side lobes.* Since the window multiplies the data, their corresponding FT are convoluted (according to the Inverse Convolution Theorem (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996)). Hence, the genuine data spectrum is distorted by

the window spectrum. Ideally, the window spectrum is not distorting the genuine one only if it is identical to the unit (or Dirac) impulse. In another words, only the main lobe should be present (not the side lobes) and its aperture should be null in spectral images below. But, as one can see from the windows' spectra, none of them verify this (ideal) property.

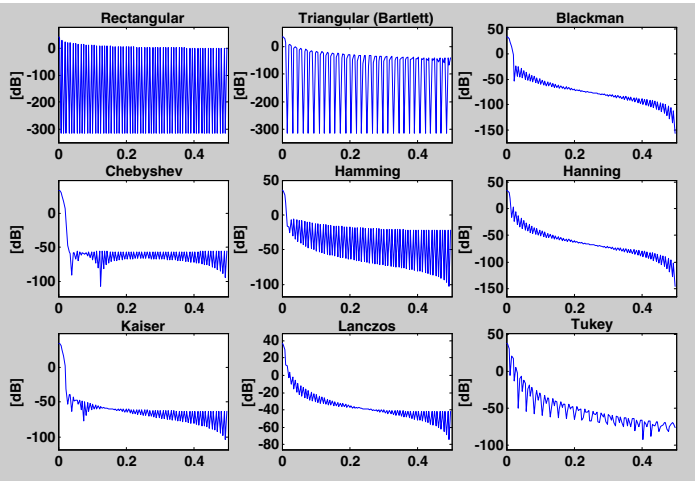


Figure 5.5. Spectra of the nine signal processing windows above.

Thus, one can say that a "good" window (in terms of attenuation criterion) should have a small aperture of the main lobe and a rapid attenuation over the side lobes. In this way, a minimal distortion is introduced into the genuine data. But one may easily guess that these two properties are opposite, as a direct consequence of the Gabor-Heisenberg Uncertainty Principle (Cohen, 1995; Proakis and Manolakis, 1996). Actually, except the rectangular window, all the other windows are performing a trade-off between the main lobe aperture and the side lobes attenuation.

The rectangular window, which anyone is tempted to select for its simplicity, is, in fact, the worst one in terms of side lobes attenuation, but probably one of the best in terms of main lobe aperture. The triangular window improves in some respect this trade-off, but not essentially. Among the other windows, Blackman, Hanning and Kaiser prove very good performances. (the Hanning window is actually employed in many filter design methods.)

But, for the purpose of our model, the Tukey window is very likely the most appropriate. As one can see, its shape in the time domain (Figure 5.4) is very well adapted to the manner in which the vibration segments are constructed: one important central frame and two lateral auxiliary frames (that should gradually be weighted). In frequency, a good trade-off between main lobe aperture and side lobes attenuation is realized (see again Figure 5.5). Therefore, the vibration segments are built by windowing the data with a Tukey window (for $\alpha = 1/3$). Note that all the other eight windows have been tested by simulation, but none of them could overtake the Tukey window in terms of final defect classifications

properties. But, actually, the method presented here is not very sensitive to the employed window, which constitutes an advantage.

Denote by s_m (for $m \in \overline{1, M-1}$) the current segment resulting after windowing the data by w. Then the windowing effect could be described by:

$$s_m \equiv [v_{m-1}\ v_m\ v_{m+1}] \cdot w \tag{5}$$

Step 4: Digital filtering of vibration data.

The vibration segments s_m are utilized next in a filtering procedure aiming to remove the LF oscillatory part and, eventually, some HF noise. The filters are digital. Unlike many approaches regarding vibrations filtering, here, one takes benefits from the modern and powerful finite impulse response (FIR) filters design procedures described, for example in (Proakis and Manolakis, 1996).

Two types of digital FIR filters could be employed: high-pass and band-pass. The first one just removes most of the harmonic natural oscillations. The second one could moreover remove the HF noise inherited by vibration data especially from environmental sources. For these filters, some parameters should be set, in order to perform the design: the filter length (N_h), the left cutoff frequency (ν_{lc}) and the right cutoff frequency (ν_{rc}, in case of high-pass filters).

The filter length should be large enough to yield good filters characteristics, but it should not overtake the segment length. A suitable choice is $N_h \in \{N_f, N_f + 1\}$, provided that the frame length is sufficiently large. (According to FIR procedure design, in the case of high-pass filters, the length must be odd. If N_f is even, then N_h should be set to $N_f + 1$.)

The left cutoff frequency ν_{lc} has to be set such that the decaying natural harmonics in raw vibration are strongly attenuated or removed. Thus, on the one hand, $\nu_{lc} \geq \nu_{lc,\min}$, where the inferior limit $\nu_{lc,\min}$ is set to 7-10 times the maximum natural frequency of oscillation. On the other hand, increasing the left cutoff frequency beyond a limit of 2 kHz may result in a loss of information about possible defects. Thus, ν_{in} should be set in the range $[\nu_{lc,\min}, 2000]$ [Hz].

Unlike within the EA method, here, the right cutoff frequency ν_{rc} should ensure a sufficiently wide pass band, in order to extract all information encoding defects. If the anti-aliasing analogic filters do not remove some HP noises, then ν_{rc} should be selected such that they are attenuated in subband $[\nu_{rc}, \nu_s/2]$. Normally, the width of this subband should not be larger than ν_{lc}, but this is not a requirement. Sometimes, the right cutoff is imposed by a central symmetry frequency, usually selected according to a resonance peak in vibration spectrum.

In Figure 5.6, the characteristics of two filters have been depicted: a high-pass one (to the left) and a band-pass one (to the right). For both filters, $N_f = 2048$, but the high-pass one must have an odd number of coefficients.

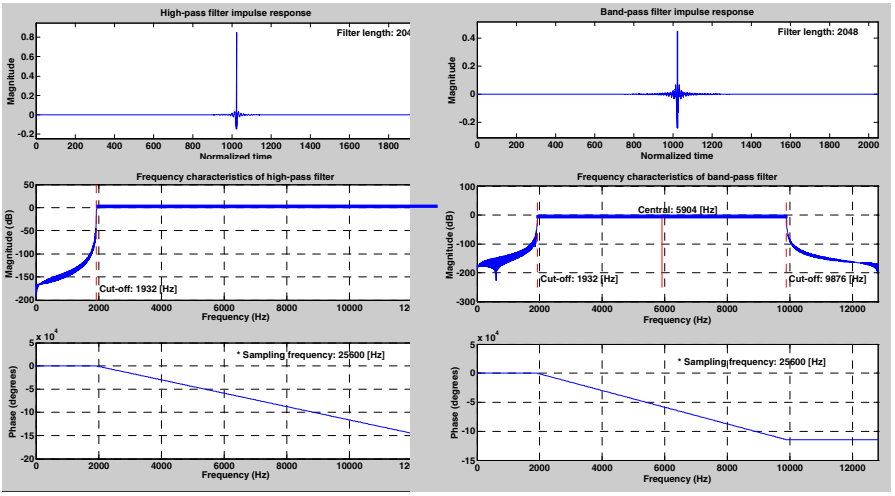


Figure 5.6. High-pass (left) and band-pass (right) filter characteristics.

The time domain characteristics (the impulse responses) are less suggestive than the frequency domain characteristics (magnitude and phase of frequency responses). "Good" filters should have an abrupt change at the cut-off frequency, a strong attenuation in stop band(s), no ripples on the main lobe and linear phase. In this figure, the left and right cutoff frequencies have been set to $\nu_{lc} = 1932$ Hz and $\nu_{rc} = 9876$ Hz, whereas the sampling frequency is $\nu_s = 25.6$ kHz. Actually, the band-pass filter was centred on 5.9 kHz. The attenuation in stop bands is quite strong, thanks to the large filter lengths.

Segments are one by one filtered. If h is the impulse response of the selected filter, then any filtered segment is simply obtained by convolution: $s_m * h$. Since $N_h \in \{N_f, N_f + 1\}$, its length is $3N_f$ as well. This involves $s_m * h$ could also be split into three frames with same length (N_f):

$$s_m * h \equiv [v_{h,m-1}\ v_{h,m}\ v_{h,m+1}] \tag{6}$$

The reason the filtered segment is split again into three frames in Eq. 6 is very simple. The filter was not actually applied to all frames in s_m but to its main frame, the central one. The lateral frames are only context signals that tell to the filter there are nonnull signal values before and after the main frame. Since filters are shift invariant linear systems (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996), the main frame in Eq. 6 is also the central one. Therefore, from the filtered segment, only one frame is extracted for the next step: $v_{h,m}$. Note that, in general, $v_{h,m}$ is different from $v_m * h$ and it is closer to the real behaviour of filtered vibration, due to the lateral frames. Also, the first and the last raw vibration frames (v_0 and v_M) are only involved as context signals aiming to avoid marginal errors. They are not furthermore transmitted.

The resulting filtered frames could be considered as nonoverlapping, since the main frame of the segment becomes the context (auxiliary) frame for the next segment. There are $M-1$ filtered frames $\{v_{h,m}\}_{m\in\overline{1,M-1}}$. These are inputs for the fuzzy-statistical model described next. Note that a set of standard (defect free) vibration preprocessed data $\{v^0_{h,m}\}_{m\in\overline{1,M-1}}$ is also provided by the same technique.

5.2.3. The Fuzzy-Statistical Model

The steps aiming to construct the fuzzy-statistical model are grouped into two categories: construction of the spectral statistic information about the filtered frames $\{v_{h,m}\}_{m\in\overline{1,M-1}}$ and utilization of this information in a fuzzy approach.

Step 1: Spectrum evaluation and segmentation.
The spectrum of each frame $v_{h,m}$ (or $v^0_{h,m}$) is evaluated by using one of the powerful existing FFT algorithms (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996). Denote by $\mathsf{V}_{h,m}$ (respectively by $\mathsf{V}_{h,m}^0$) the spectrum of current (filtered) frame ($m\in\overline{1,M-1}$), i.e., the magnitude of its FT. Since the spectrum is symmetric for real valued data sequences, it follows that only the first $N_f/2$ rays could be accounted, which corresponds to a bandwidth of $\nu_s/2$.

The main difference between spectra encoding information about defects $\mathsf{V}_{h,m}$ and defect-free spectra $\mathsf{V}_{h,m}^0$ is that the former have a bigger variability among frames, whereas the later vary within some minimum and maximum bounds, close to each other. The variability could be expressed in various ways, but, for this model, sp are employed to quantify the spectral behaviour.

By convention, let $\mathsf{V}^*_{h,m}$ stand for any of two spectra above ($\mathsf{V}_{h,m}$ or $\mathsf{V}_{h,m}^0$). The full frequency band of each spectrum $\mathsf{V}^*_{h,m}$ is uniformly segmented next into $K\geq 1$ subbands, in order to evaluate a set of local sp. Such a frequency segment (subband) should include between 5 and 10 rotations of main shaft, in order to construct a consistent set of sp. Thus, the segment bandwidth should be set between $5\nu_r$ and $10\nu_r$ (i.e., K should vary in the range $[1/10,1/20]\nu_s/\nu_r$). The minimum bound yields a good *frequency resolution* (i.e., narrow subbands), but a smaller sp accuracy than the maximum bound, where, however, the resolution is worst. Obviously, the sp consistency (accuracy) depends on the number of accounted data. In this case, the consistency depends on the number of rays included in a segment, that is, on the segment bandwidth. The bigger the bandwidth, the more consistent the sp, but the less focused on local spectral variation. A good compromise is realized for $8\nu_r$:

$$K=\lceil \nu_s/16\nu_r \rceil \tag{7}$$

The number of rays within each frequency subband (except possibly the one located at the highest frequency) is:

$$N_K = \lfloor N_f / 2K \rfloor \tag{8}$$

For example, $N_f = 2048$, $\nu_r = 50$ Hz (3000 rpm), and $\nu_s = 25.6$ kHz lead to: $K = 32$ subbands (of 400 Hz bandwidth each) and $N_K = 32$ rays/sub-band, according to Eqs. 7 and 8.

By convention, sub-bands are indexed from 0 to $K-1$.

Splitting the spectrum in a number of equally spaced subbands may not be the best solution to focus on spectral power local variation. However, the trade-off between frequency resolution (or K) and sp accuracy (or N_K) determines the minimum bandwidth for carrying out the statistical analysis. Nonuniform segmentations could be realized by compacting together two or more adjacent subbands with minimum bandwidth. But the fdd method described here is independent on the type of frequency segmentation. Therefore, for the sake of simplicity, the segmentation is kept uniform hereafter.

To conclude this step, a final remark should be noted. Filtering the vibration segments involves a separation of frequency stop subbands and pass subbands. The statistical parameters might not be similarly employed for any of these 2 subband types, because the information encoded inside the stop subbands is probably extremely poor and noisy compared to the information inside the pass subbands. Since the whole band was practically quantified by K values, separation lines between stop and pass subbands have to be defined. Obviously, the cut-off frequencies ν_{lc} and ν_{rc} belong to some subbands as follows:

$$\nu_{lc} \in [K_{lc}, K_{lc}+1)\frac{\nu_s}{2K} \text{ and } \nu_{rc} \in (K_{rc}-1, K_{rc}]\frac{\nu_s}{2K} \tag{9}$$

where

$$K_{lc} = \lfloor 2K\nu_{lc} / \nu_s \rfloor \text{ and } K_{rc} = \lceil 2K\nu_{rc} / \nu_s \rceil \tag{10}$$

For example, if, like previously, $\nu_{lc} = 1932$ Hz and $\nu_{rc} = 9876$ Hz, whereas $\nu_s = 25.6$ kHz and $K = 32$, then: $K_{lc} = 4$ and $K_{rc} = 25$.

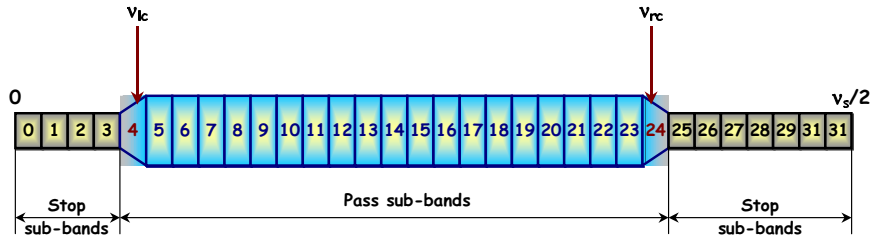


Figure 5.7. An example of frequency segmentation.

Normally, the transition subbands (i.e., including the cutoff frequencies) should be pass type, in order to avoid removing useful side information. Therefore, the stop subbands are: {0, 1, ..., $K_{lc}-1$, K_{rc}, ..., $K-1$}. Consequently, the pass subbands are: {K_{lc}, ..., $K_{rc}-1$}. For the previous example, the stop and pass sub-

bands are: {0, 1, 2, 3, 25, 26, ..., 31}, respectively {4, 5, ..., 23, 24}, as depicted in Figure 5.7.

<u>Step 2</u>: Definition of sp and construction of relative statistical vectors.
Using statistics to extract information about defects from raw vibration is not a new idea. Many analysts perform diagnosis with the help of some parameters such as the *root mean square* (RMS) or the *peak value* evaluated either from vibration data or their spectrum.

A quasi-complete statistical set of parameters includes the following 12 parameters: *peak (to valley)* (Δv); *average* ($\bar{v}$); *absolute average* ($\overline{|v|}$); *energy* (E_v); *normalized energy* (E_v^N); *root mean square* (RMS_v); *peak to average ratio* (PAR_v); *crest factor* (CF_v); *impulse factor* (IF_v); *shape factor* (SF_v); *clearance factor* (CLF_v); *Kurtosis* (K_v). Their definitions are listed in Eq. 11, for any N-length data series, $\{v[n]\}_{n\in\overline{0,N-1}}$ (such as vibrations or their spectra):

$$\Delta v \overset{def}{=} \frac{1}{2}\left[\max_{n\in\overline{0,N-1}}\{v[n]\} - \min_{n\in\overline{0,N-1}}\{v[n]\}\right];\ \bar{v} \overset{def}{=} \frac{1}{N}\sum_{n=0}^{N-1} v[n];\ \overline{|v|} \overset{def}{=} \frac{1}{N}\sum_{n=0}^{N-1}|v[n]|;$$

$$\mathsf{E}_v \overset{def}{=} \sum_{n=0}^{N-1}|v[n]|^2;\ \mathsf{E}_v^N \overset{def}{=} \frac{1}{N}\sum_{n=0}^{N-1}|v[n]|^2;\ RMS_v \overset{def}{=} \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left(v[n]-\bar{v}\right)^2};$$

$$PAR_v \overset{def}{=} \frac{1}{\overline{|v|}}\max_{n\in\overline{0,N-1}}\{|v[n]|\};\ CF_v \overset{def}{=} \frac{\Delta v}{RMS_v};\ IF_v \overset{def}{=} \frac{\Delta v}{\overline{|v|}}; \qquad (11)$$

$$SF_v \overset{def}{=} \frac{RMS_v}{\overline{|v|}};\ CLF_v \overset{def}{=} \frac{\Delta v}{\frac{1}{N}\left(\sum_{n=0}^{N-1}\sqrt{|v[n]|}\right)^2};\ \mathsf{K}_v \overset{def}{=} \frac{\frac{1}{N}\sum_{n=0}^{N-1}\left(v[n]-\bar{v}\right)^4}{RMS_v^4}$$

The first six parameters are concerned with energetic characteristics, whereas the other six quantify different shape properties. Obviously, the number of data, N, is a measure of sp accuracy. (The accuracy increases with N.)

Usually, the values of parameters defined in Eq. 11 are compared to standard values corresponding to defect-free systems. Their biases could indicate the desired information about defects (including estimations of severity degree). Though the number of parameters to account for is large enough, no one is able to extract all the necessary information about defects.

Once the frequency segmentation has been realized, some sp should be evaluated within every subband. Note that the set of 12 sp above is redundant. For example, in (Xi *et al.*, 2000), one states that peak-to-valley is similar to RMS, to energy and to absolute average; impulse factor is similar to shape factor; kurtosis is similar to crest factor. These similarities are not realized in sense of similarity measure from physics, but in terms of some features ad hoc defined in the context of that research. Therefore, a safe approach is to take into consideration as much sp as possible. An obvious remark is that, for nonnegative data (like spectral powers),

the absolute average is identical to the average. Also, it is better to evaluate the normalized energy instead of pure energy, especially in the case of nonuniform frequency segmentation (when subbands have different numbers of rays and, thus, their energy becomes noncomparable). Thus, only 10 sp are retained in this context. They are denoted according to time and frequency segmentations performed so far: *peak (to valley)* ($\Delta V^*_{h,m}$); *average* ($\overline{V^*_{h,m}}$); *normalized energy* ($E^{N,*}_{h,m}$); *root mean square* ($RMS^*_{h,m}$); *peak to average ratio* ($PAR^*_{h,m}$); *crest factor* ($CF^*_{h,m}$); *impulse factor* ($IF^*_{h,m}$); *shape factor* ($SF^*_{h,m}$); *clearance factor* ($CLF^*_{h,m}$); *kurtosis* ($K^*_{h,m}$). As usual, the $*$ employed in notations points to any of 2 vibration data types: acquired from the tested bearing (* vanishes) or from the standard (defect free) bearing (* is replaced by 0).

Any of the sp above takes K values for every frame $m \in \overline{1, M-1}$ (one value for each subband). The number of rays per subband determines their consistency, N_K. For example, $RMS^*_{h,m}$ could be evaluated as follows:

$$RMS^*_{h,m}[k] \overset{def}{=} \sqrt{\frac{1}{N_K}\sum_{n=0}^{N_K-1}\left(V^*_{h,m}[kN_K+n]-\overline{V^*_{h,m}}[k]\right)^2}\ ,\ \forall k \in \overline{0, K-1} \tag{12}$$

where the local average is:

$$\overline{V^*_{h,m}}[k] \overset{def}{=} \frac{1}{N_K}\sum_{n=0}^{N_K-1} V^*_{h,m}[kN_K+n]\ ,\ \forall k \in \overline{0, K-1} \tag{13}$$

A $10 \times K$ statistical matrix $S^*_{h,m}$ could be constructed for every spectral frame $V^*_{h,m}$, by stacking the sp values in successive row vectors, as enumerated above. Thus, for example, the RMS value in Eq. 13 is the element $[4, k+1]$ of matrix $S^*_{h,m}$, i.e., $S^*_{h,m}[4, k+1] = RMS^*_{h,m}[k]$, whereas the fourth row of the matrix packs all RMS values among subbands. The generic element of matrix $S^*_{h,m}$ is $S^*_{h,m}[i, j]$, where $i \in \overline{1,10}$, $j = k+1 \in \overline{1, K}$ and $m \in \overline{1, M-1}$.

When the tested bearing is defect-free (standard), the statistical values of matrices $S^0_{h,m}$ vary within some acceptable tolerances among frames. Thus, in this case, the values of every sp are located inside a min-max domain, whose bounds depend on the evaluation subband. More specifically, let $P_i{}^*$ be the *i*-th sp in the list above (for $i \in \overline{1,10}$). (For example, $P_2{}^* \equiv \overline{V}^*$, $P_4{}^* \equiv RMS^*$, etc.) Then its value for the *m*-th frame and the *k*-th sub-band is $P_i{}^*[m,k]$. For the standard vibration, $P_i{}^0[\bullet, k]$ could vary in the range $\left[P_i{}^{\min}[k], P_i{}^{\max}[k]\right]$ among frames, but within the same subband (*k*). A natural manner to evaluate the min-max bounds is to account for all frames:

$$\mathsf{P}_i^{\min}[k] = \min_{m\in\overline{1,M-1}}\left\{\mathsf{P}_i^{0}[m,k]\right\},\ \mathsf{P}_i^{\max}[k] = \max_{m\in\overline{1,M-1}}\left\{\mathsf{P}_i^{0}[m,k]\right\},\ \forall k\in\overline{0,K-1} \tag{14}$$

This involves that two remarkable matrices could be constructed, by gathering together all minimum or maximum values evaluated in Eq. 14:

$$\mathcal{S}^{\min}[i,j]=\mathsf{P}_i^{\min}[j-1],\ \mathcal{S}^{\max}[i,j]=\mathsf{P}_i^{\max}[j-1],\ \forall i\in\overline{1,10},\ \forall j\in\overline{1,K} \tag{15}$$

The same result is obtained if the min and max operators are applied elementwise on matrices $\left\{\mathcal{S}_{h,m}^{0}\right\}_{m\in\overline{1,M-1}}$. In practice, the min and max values are furthermore *corrected* by multiplication with constants $\sigma^{\min}<1$ and $\sigma^{\max}>1$, respectively, , in order to avoid diagnosing as defective the defect-free bearings. For example, $\sigma^{\min}=0.6$ and $\sigma^{\max}=1.1$. The lower bound is, however, less important than the upper bound and this is the reason the constant $\sigma^{\min}$ is not 0.9 (the symmetrical value of $\sigma^{\max}=1.1$), but 0.6. By convention, hereafter, one preserves the same notations $\mathsf{P}_i^{\min}[k]$ and $\mathsf{P}_i^{\max}[k]$ for corrected bounds as well.

Defective bearings provide vibrations that exceed some or all the (corrected) bounds in matrices defined by Eq. 15. The biases of sp P_i outside the standard range could indicate the desired information about defects, including estimations of severity degree. Note that defects could be detected not only when maximum bound is overtaken, but also if the minimum bound is undertaken. The second effect is especially induced by lubrication defects, excessive wear or multiple-point defects (when the phases of FT could lead to energy attenuation inside some subbands). In order to quantify the severity degree of defects, the sp are replaced by the relative statistical parameters (rsp), defined as explained next.

There are two types of assessments when performing the comparison between sp and their bounds: by accounting for both min and max limits or by considering only the max limit. Both limits should be accounted for pass subbands, whereas only the max limit is sufficient for the stop bands. In the first case, for each sp P_i ($i\in\overline{1,10}$) one defines a corresponding rsp R_i as follows:

$$\mathsf{R}_i[m,k]\overset{def}{=}\frac{1}{\sqrt{10}}\begin{cases}\mathsf{P}_i[m,k]/\mathsf{P}_i^{\max}[k], & \text{if } \mathsf{P}_i[m,k]>\mathsf{P}_i^{\max}[k]\\ 1, & \text{if } \mathsf{P}_i[m,k]\in\left[\mathsf{P}_i^{\min}[k],\mathsf{P}_i^{\max}[k]\right]\\ \mathsf{P}_i^{\min}[k]/\mathsf{P}_i[m,k], & \text{if } 0<\mathsf{P}_i[m,k]<\mathsf{P}_i^{\min}[k]\end{cases} \tag{16}$$
$$\forall m\in\overline{1,M-1},\ \forall k\in\overline{0,K-1}$$

Similarly, in the second case, the definition of rsp can be expressed as:

$$\mathsf{R}_i[m,k]\overset{def}{=}\frac{1}{\sqrt{10}}\begin{cases}\mathsf{P}_i[m,k]/\mathsf{P}_i^{\max}[k], & \text{if } \mathsf{P}_i[m,k]>\mathsf{P}_i^{\max}[k]\\ 1, & \text{if } \mathsf{P}_i[m,k]\le\mathsf{P}_i^{\max}[k]\end{cases}, \tag{17}$$
$$\forall m\in\overline{1,M-1},\ \forall k\in\overline{0,K-1}$$

The same philosophy was employed in both definitions of Eqs. 16 and 17: if the maximum bound is exceeded, evaluate how many times the parameter

overtakes the bound; if the minimum bound is exceeded, evaluate how many times the bound overtakes the parameter; set by 1 the rsp when the parameter stays within the tolerance limits.

Note that, independently of the sp type, the values of different rsp could now be compared, thanks to their relative nature. Thus, for example, although RMS (P_4) is not comparable with kurtosis (P_{10}), the relative RMS (R_4) has values varying in a similar range to the relative kurtosis (R_{10}). Therefore, the rsp values of the same frame within the same subband could be packed in a 10-length column vector $\mathsf{R} = [\mathsf{R}_1, \mathsf{R}_2, \ldots, \mathsf{R}_{10}]^T$. The purpose of the $1/\sqrt{10}$ factor employed in both definitions above is to normalize the vector R in the following sense:

$$\left\|\mathsf{R}[m,k]\right\| \geq 1, \ \forall m \in \overline{1, M-1}, \ \forall k \in \overline{0, K-1} \tag{18}$$

and $\left\|\mathsf{R}[m,k]\right\| = 1$ if the spectrum of the m-th frame behaves normally within the k-th subband (as for the defect-free bearing). Starting with the next step, Euclidean norms $\left\|\mathsf{R}[m,k]\right\|$ are actually employed. For a more general approach, other norms could be considered as well. For example, one can consider that not all sp have the same weight and thus a weighting matrix $\mathsf{Q} \in \mathbb{R}^{10\times 10}$ (eventually diagonal) has to multiply left the rsp vector R. The norm of the resulting vector QR is in fact a generalized Euclidean Q-norm.

Returning to Eqs. 16 and 17, a special case remains to be considered: the null parameter values, when both bounds have to be accounted. If one recalls the sp definitions in Eq. 11, it is easy to see that not all parameters could be null, even when the input data consists of a finite length null signal. This property is proven by those parameters quantifying the signal shape, since a part of the shape information is the signal length (denoted by N in Eq. 11). In fact, simple algebraic manipulations lead to the following interesting limits when the signal v tends to the null signal:

$$\begin{gathered}
\lim_{v\to 0} \Delta v = \frac{0}{2} = 0 \ ; \ \lim_{v\to 0} \bar{v} = \frac{0}{N} = 0 \ ; \ \lim_{v\to 0} \left|\bar{v}\right| = \frac{0}{N} = 0 \ ; \\
\lim_{v\to 0} \mathsf{E}_v = 0^2 = 0 \ ; \ \lim_{v\to 0} \mathsf{E}_v^N = \frac{0^2}{N} = 0 \ ; \ \lim_{v\to 0} RMS_v = \frac{\sqrt{N-1}}{N} 0 = 0 \ ; \\
\lim_{v\to 0} PAR_v = N \ ; \ \lim_{v\to 0} CF_v = \frac{N}{2\sqrt{N-1}} \ ; \ \lim_{v\to 0} IF_v = \frac{N}{2} \ ; \\
\lim_{v\to 0} SF_v = \sqrt{N-1} \ ; \ \lim_{v\to 0} CLF_v = \frac{N}{2} \ ; \ \lim_{v\to 0} \mathsf{K}_v = \frac{N^2 - 3N + 3}{N-1}
\end{gathered} \tag{19}$$

Thus, the shape parameters are null if and only if the signal is empty. Practically, in context of spectral frames, they are always nonnull. But the energetic parameters could be null inside some subbands, if and only if all corresponding rays are null. Usually, if in a pass subband all rays are null, either a severe defect is announcing or there are some important errors within the available data. The second hypothesis could be confirmed when the spectrum is null for many pass subbands. But, if only few isolated pass subbands provide null data, then the first hypothesis is

more plausible. In this case, the rsp should be set to a value equal to or more than $10\mathsf{P}_i^{\max}[k]$, for all subbands where $\mathsf{P}_i\ [m,k]=0$. The reason for this setting will become obvious in the next step.

An average set of norms $\left\{\left\|\mathbf{R}\ [i,k]\right\|\right\}_{i\in\overline{1,m}}$ is also evaluated for each subband, after every $m\in\overline{1,M-1}$ processed frames:

$$\overline{\left\|\mathbf{R}\right\|}[m,k]\overset{def}{=}\frac{1}{m}\sum_{i=1}^{m}\left\|\mathbf{R}\ [i,k]\right\|,\ \forall k\in\overline{0,K-1} \tag{20}$$

This entity is extremely useful for initialising the fuzzy model. In fact, one can consider that the processing starts from a *virtual frame* that provides the average information about rsp norms at any moment. Set the index of virtual frame by $m=0$ and change notation $\overline{\left\|\mathbf{R}\right\|}[m,k]$ by $\left\|\mathbf{R}\ [0,k]\right\|$. In the new notation, the current number of frames was omitted, in order to unify all notations regarding the rsp norms. But, hereafter, one can consider by convention that the set of rsp norms $\left\{\left\|\mathbf{R}\ [i,k]\right\|\right\}_{i\in\overline{0,m}}$ always starts with the average of currently processed frames $\left\{\left\|\mathbf{R}\ [i,k]\right\|\right\}_{i\in\overline{1,m}}$ in the first position. This average could recursively be upgraded, from a frame to another, according to the equation below:

$$\overline{\left\|\mathbf{R}\right\|}[m+1,k]=\frac{m\overline{\left\|\mathbf{R}\right\|}[m,k]+\left\|\mathbf{R}\ [m+1,k]\right\|}{m+1}\ ,\ \forall k\in\overline{0,K-1} \tag{21}$$

After processing the first frame, the average is identical to $\left\|\mathbf{R}\ [1,k]\right\|$, but starting from the second processed frame, the average and the other rsp norms are, in general, different. Therefore, within the next steps, one shall assume that the average starts to be evaluated after at least two frames have been processed.

<u>Step 3</u>: Definition and construction of a *statistical network*.

Let $\left\|\mathbf{R}\ [m,k]\right\|_{\mathrm{dB}}$ be the value of $\left\|\mathbf{R}\ [m,k]\right\|$ expressed in dB (for $m\in\overline{0,M-1}$, i.e., including the average (Eq. 20). Then the severity degree of defect could be expressed in terms of a grid, in dB as well. Usually, there are 4 severity types: *normal* (when no defect seems to be detected), *incipient*, *medium* and *severe*. The separation values between severity types could be set as follows: 1, $\left[2\right]_{\mathrm{dB}}\approx 6\,\mathrm{dB}$ and $\left[10\right]_{\mathrm{dB}}=20\,\mathrm{dB}$. Thus, if $\left\|\mathbf{R}\ [m,k]\right\|$ varies in the range [1,1.22), no defect is present; for range [1.22,2), the defect is incipient; inside the range [2,10), the defect is medium and if $\left\|\mathbf{R}\ [m,k]\right\|$ is more than 10, the defect is severe. The grid could refine the severity levels for every type as follows: 0, 1, $\left[\sqrt{2}\right]_{\mathrm{dB}}\approx 3$, $\left[2\right]_{\mathrm{dB}}\approx 6$, $\left[3\right]_{\mathrm{dB}}$, $\left[4\right]_{\mathrm{dB}}$, ..., $\left[9\right]_{\mathrm{dB}}$, $\left[10\right]_{\mathrm{dB}}=20$ [dB] ($L=12$ levels). Let $\Lambda=[\lambda_l]_{l\in\overline{0,L-1}}\in\overset{\circ}{\mathbb{R}}_+^L$ be the L-length vector of all severity levels expressed in dB and set $\lambda_L=\infty$.

All settings above aimed to build a map like the one depicted in Figure 5.8 and referred to as the *statistical network* (sn). Thus, for each subband, a box cell is assigned to every severity degree. Each value $\|\mathbb{R}[m,k]\|_{\mathrm{dB}}$ is uniquely located inside such a box, as suggested by the diamonds in figure. In this example, the location of rsp norms of a frame is depicted. The maximum rsp norm is reached inside subband #5, where an incipient-medium defect is announced. Its severity degree is ≅5.89 dB (at least one sp is about 1.97 times out of standard min-max range). Note that the box cells corresponding to severe defects are *open*, in the sense that their height varies depending on maximum pointed severity degree (if applicable). On the contrary, the other box cells have fixed heights (but differ from one severity degree to another).

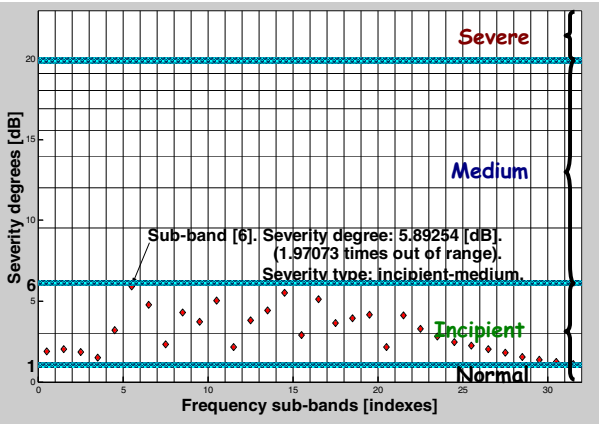


Figure 5.8. A statistical network example.

The sn provides in fact a *statistical map* of possible defects, simpler than the spectrum image. And yet, it is difficult to perform a good fdd by only inspecting this map. Therefore, a technique of grouping network cells in similarity classes could make this task easier.

Step 4: Covering the statistical network with clusters.
The previous steps prepared the fuzzy model construction. Starting from this step, a fuzzy approach is combined with statistics in order to provide defect classifications expressed as partitions of sn above. This approach is based on concepts of *fuzzy relations* and *fuzzy entropy* (Klir and Folger, 1988) and its kernel has already been integrated into another (but very different) method concerned with identification of main structures inside Multi-Agent Systems (Ulieru *et al.*, 2000).

In context of vibrations, the fuzzy model relies on the fact that every frame encodes the same information about existing defects (if the frame length is large enough to induce a good accuracy of sp). Consequently, the statistical maps resulting from every frame reveal about the same correlation between those box cells that actually encode the defect, whereas the remaining cells are less correlated. More specifically, the rsp norms from different frames "fall" more often into the

same boxes for those subbands that seem to be directly affected by the defect. One can say that rsp norms *occur* more often inside box cells that apparently encode the systematic spectrum biases caused by a specific defect.

Therefore, the basic idea is first to construct a similarity fuzzy relation between box cells within sn and then to unpack the result as different classifications comprising similarity classes. A *similarity class* is actually a group of box cells that seem to point to the same fault or group of faults (with some confidence degree).

In construction of a fuzzy relation between box cells, the first action is to specify how the sn could be covered by collections of box cells for every spectral frame. Any collection of box cells is referred to in this context as a *cluster*. Denote by $B_{l,k}$ the generic box-cell of sn, where $l \in \overline{0, L-1}$ is the severity level and $k \in \overline{0, K-1}$ is the frequency subband. A natural way to construct clusters is to consider two types of sn covers as follows:

a. a horizontal one, $\mathcal{H}$, with L clusters, each of which includes only constant severity level box cells: $C_l = \{B_{l,k}\}_{k \in \overline{0, K-1}}$ ($l \in \overline{0, L-1}$);
b. a vertical one, $\mathcal{G}$, with K clusters, each of which includes only box cells corresponding to the same frequency subband: $D_k = \{B_{l,k}\}_{l \in \overline{0, L-1}}$ ($k \in \overline{0, K-1}$).

Thus:

$$\mathcal{H} = \bigcup_{l=0}^{L-1} C_l \text{ and } \mathcal{G} = \bigcup_{k=0}^{K-1} D_k \tag{22}$$

Note that the covers in Eq. 22 are independent of frame index (they preserve the same structure for all frames), since, at this stage, one focuses only on the structural information about how the sn could be roughly organized. The information about defects encoded by rsp norms will be accounted for in a future stage.

An example of horizontal and vertical clusters is displayed in Figure 5.9.

Other structures of sn covering could be considered as well, for example, the one consisting of cross-clusters obtained by taking the union between horizontal and vertical clusters (also illustrated in Figure 5.9). But the main advantage of coverings above is that they lead to one of the simplest fuzzy relation construction algorithms.

The box cells that belong to the same cluster are in fact entities verifying the same elementary crisp (binary) relation. Two crisp relations could thus be stated: (a) two box cells are in the same relationship if they reveal the same severity level; (b) two box cells are in the same relationship if they point to the same frequency subband. The characteristic (index) functions describing these crisp relations are $KL \times KL$ binary matrices, where the element (i, j) is unitary only if the box cells i and j are in relation to each other (otherwise, the element (i, j) is null). These matrices could be expressed only after linearization of sn indices. Thus, the box cell $B_{l,k}$ located in plane by the indices (l, k) is equivalently located on a

lineal by the index $i_{l,k}$. There are two possibilities to derive the expression of index $i_{l,k}$: by enumerating all columns or by enumerating all rows of sn. In this approach, one selects to enumerate the sn rows, starting from bottom to top (see Figure 5.10). Thus, the first group of box cells is associated with normal behaviour. The incipient, medium and severe defect box cells follow (in this order).

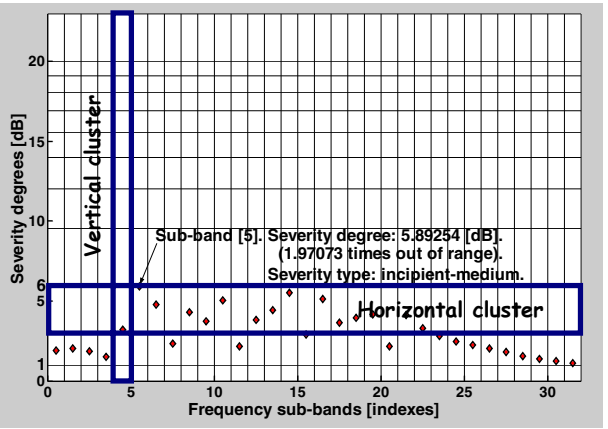


Figure 5.9. Horizontal and vertical clusters inside the statistical network.

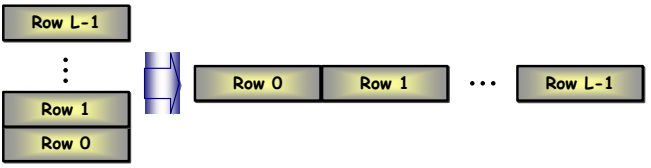


Figure 5.10. Linear enumeration of box cells in a statistical network.

The index $i_{l,k}$ is then: $i_{l,k} = lK + k$, $\forall l \in \overline{0, L-1}$, $\forall k \in \overline{0, K-1}$. Conversely: $l = \lfloor i_{l,k} / K \rfloor$ and $k = i_{l,k} \% K$, $\forall i_{l,k} \in \overline{0, LK-1}$, where $\lfloor a \rfloor$ is the integer part of $a \in \mathbb{R}$ and $n \% N$ is the rest of division between integers n and N.

Since any of the two sn covers provided by a frame is a union of its (disjoint) clusters, the associated global binary crisp relation is also a union of elementary crisp relations. Hence, the global characteristic matrix is obtained by summing together all corresponding elementary matrices. The specific form of the selected covers leads to the global characteristic matrices given in Eq. 23.

As a toy example, set $L = 2$ and $K = 3$. The corresponding sn looks as in Figure 5.11 and its covers are:

a. Horizontal: $\mathcal{H} = \{B_{00}, B_{01}, B_{02}\} \cup \{B_{10}, B_{11}, B_{12}\}$

b. Vertical: $\mathcal{G} = \{B_{00}, B_{10}\} \cup \{B_{01}, B_{11}\} \cup \{B_{02}, B_{12}\}$

$$\underset{\text{Horizontal cover}}{H=\begin{bmatrix} U_{K\times K} & O_{K\times K} & \cdots & O_{K\times K} \\ O_{K\times K} & U_{K\times K} & \cdots & O_{K\times K} \\ \vdots & \vdots & \ddots & \vdots \\ O_{K\times K} & O_{K\times K} & \cdots & U_{K\times K} \end{bmatrix}_{\substack{L\times L \\ \text{blocks}}}} \qquad \underset{\text{Vertical cover}}{G=\begin{bmatrix} I_{K\times K} & I_{K\times K} & \cdots & I_{K\times K} \\ I_{K\times K} & I_{K\times K} & \cdots & I_{K\times K} \\ \vdots & \vdots & \ddots & \vdots \\ I_{K\times K} & I_{K\times K} & \cdots & I_{K\times K} \end{bmatrix}_{\substack{L\times L \\ \text{blocks}}}} \tag{23}$$

where $U_{K\times K}$, $O_{K\times K}$ and $I_{K\times K}$ are the $K\times K$ all unit, all zero and identity matrices, respectively, expressed as:

$$U_{K\times K}=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{K\times K}, O_{K\times K}=\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{K\times K}$$
$$I_{K\times K}=\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{K\times K} \tag{24}$$

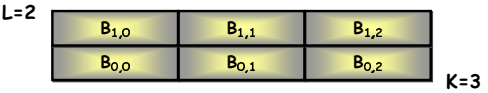


Figure 5.11. A toy statistical network.

Then Eqs. 23 and 24 imply:

$$U_{3\times 3}=\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, O_{3\times 3}=\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, I_{3\times 3}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underset{\text{Horizontal cover}}{H=\begin{bmatrix} U_{3\times 3} & O_{3\times 3} \\ O_{3\times 3} & U_{3\times 3} \end{bmatrix}=\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}} \qquad \underset{\text{Vertical cover}}{G=\begin{bmatrix} I_{3\times 3} & I_{3\times 3} \\ I_{3\times 3} & I_{3\times 3} \end{bmatrix}=\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}} \tag{25}$$

The matrices in Eq. 23 (or Eq. 25) are actually binary maps of the two crisp relations that every frame provides. The position of every unit value shows which couple of box cells are in relation to each other. These relations are in fact rough approximations of the following relation directly related to defects: *two box cells are in relation to each other if they point to the same fault*. Of course, at this time, we don't know exactly which box cells verify this property and this is the reason one operated with two approximations. Any horizontal or vertical cluster

could include abnormal box cells pointing to the subband affected by a specific defect or, respectively, to the severity degree proved by a specific defect. The approximations are refined next by using the rsp norms, until specific defect classes are obtained.

Step 5: Evaluating the occurrence degrees.
The covers constructed above do not partake in the fuzzy relation that one intends to construct, but their clusters do. Recall that, unlike within crisp relations, two entities are in a fuzzy relation to each other only if they belong to a crisp relation with some membership degree (Klir and Folger, 1988). The membership values express the uncertainty regarding the specific relationship between entities. Within the crisp approach, this relationship is either certainly existing or certainly not existing. There are no other possibilities. Within the fuzzy approach, two entities could be in a relationship, but this assertion has a degree of uncertainty varying from 0 (certainly not) to 1 (certainly yes).

The relationships between sn box cells should also be fuzzy, for two main reasons. Firstly, the horizontal and vertical clusters could not be totally reliable since, in general, they gather together boxes inside of which some rsp norms fall and boxes that are untouched by these norms, even for long strings of vibration data. These act in fact as different entities inside the sn. They were only roughly gathered together, according to structural criteria of same severity level or frequency subband, but without accounting for the information provided by the vibration itself. Secondly, the structure of selected clusters (horizontal, vertical) could not be certain, but only intuitively more plausible than another structure. Fortunately, the final fuzzy relation is not that sensitive to the initial clustering of box cells and refines these approximations.

The horizontal and vertical clusters encode no information about defects unless they are put into correspondence with the rsp norms. In reality, after processing $m+1 \in \overline{1,M}$ frames (including the virtual one naturally associated with the average information about rsp norms – see Eq. 20), inside every box cell $B_{l,k}$ a number of rsp norms could occur. Refer to this number as (*occurrences*) *counter* and denote it by $N_m[l,k]$. Obviously, since for each subband $k \in \overline{0,K-1}$ a unique severity level λ_l ($l \in \overline{0,L-1}$)exists such that:

$$\lambda_l \le \left\| \mathbf{R}\,[m,k] \right\|_{\mathrm{dB}} < \lambda_{l+1} \tag{26}$$

it follows that:

$$0 \le N_m[l,k] \le m+1,\ \forall m \in \overline{0,M-1},\ \forall l \in \overline{0,L-1},\ \forall k \in \overline{0,K-1} \tag{27}$$

Null counter values are associated with those box cells for which no rsp norms occurred so far. Furthermore, another obvious property holds:

$$\sum_{l=0}^{L-1}\sum_{k=0}^{K-1} N_m[l,k] = K(m+1),\ \forall m \in \overline{0,M-1} \tag{28}$$

i.e., the total amount of counters equals the number of subbands touched by all currently processed frames, including the virtual one.

After processing a new frame, the counters are upgraded following a rule given by Eq. 27:

$$N_{m+1}[l,k]=\begin{cases}N_m[l,k]+1, \text{ if } \lambda_l \le \|\mathbf{R}[m+1,k]\|_{\text{dB}} < \lambda_{l+1} \\ N_m[l,k], \quad \text{otherwise}\end{cases}, \begin{bmatrix}\forall m \in \overline{0,M-2} \\ \forall l \in \overline{0,L-1} \\ \forall k \in \overline{0,K-1}\end{bmatrix} \tag{29}$$

which means: increment by 1 only those counters corresponding to box cells where the rsp norms occurred. However, this rule is not that simple. The virtual frame gives the initial values of these counters and, thus, they could change depending on the number of currently processed frames, $m+1 \in \overline{1,M}$. So, Eq. 29 must be understood as a recursive recipe where the initial values are also dependent on the current step of upgrading. Consequently, a counter could even be incremented by 2 and not by 1, or decreased by 1, when the average moves its position.

A consistent set of *occurrence degrees* is constructed and one-by-one associated with the collection of sn box cells, by using counters. Denote by $\nu_m[l,k]$ the occurrence degree uniquely associated with box cell $B_{l,k}$, after processing $m+1 \in \overline{1,M}$ frames (starting from the virtual one). Two possible definitions could be used to set $\nu_m[l,k]$, according to Eqs. 27 and 28:

$$\nu_m[l,k] \overset{def}{=} \frac{N_m[l,k]}{m+1} \text{ or } \nu_m[l,k] \overset{def}{=} \frac{N_m[l,k]}{K(m+1)} \tag{30}$$

In both cases $\nu_m[l,k] \in [0,1]$, but for the first one:

$$\sum_{l=0}^{L-1} \nu_m[l,k]=1 \text{ and } \sum_{l=0}^{L-1}\sum_{k=0}^{K-1} \nu_m[l,k]=K \tag{31}$$

whereas for the second one:

$$\sum_{l=0}^{L-1} \nu_m[l,k]=\frac{1}{K} \text{ and } \sum_{l=0}^{L-1}\sum_{k=0}^{K-1} \nu_m[l,k]=1 \tag{32}$$

From a probabilistic point of view, Eqs. 31 and 32 show that only the second definition in Eq. 30 could be associated to the occurrence frequency of rsp norms inside box cells. But, in the context of fuzzy logic theory, requirements like the last one in Eq. 32 are often not necessary (Klir and Folger, 1988). The only requirement is to include the occurrence degree variation in range [0,1]. One of the main drawbacks of the second definition is the rapid decay towards null values of all occurrence degrees, due to product $K(m+1)$. No occurrence degree could increase. Even if a counter is upgraded, its value is only increased by maximum 2, whereas the corresponding occurrence degree is decreased about K times. In contrast, the first definition keeps the occurrence degrees more balanced and, furthermore, the occurrence degrees could increase. The last remark is due to a very simple algebraic property:

$$\frac{n}{m} > \frac{n}{m+1} > \frac{n-1}{m+1}, \ \forall n,m>0, \text{ but } \frac{n}{m} \le \frac{n+1}{m+1} < \frac{n+2}{m+1}, \text{ if } 0<n\le m \tag{33}$$

Since the first definition provides occurrence degrees that are more sensitive to counters upgrading than the second one, it will be selected for the next steps. In fact, the occurrence degree is only raw information about rsp norms distribution over the statistical map. More processing operations are necessary in order to derive the uncertainty degrees associated with the elementary crisp relations previously constructed.

An example of the two-dimension occurrence degrees distribution is displayed in Figure 5.12. The distribution is improved after every new processed frame.

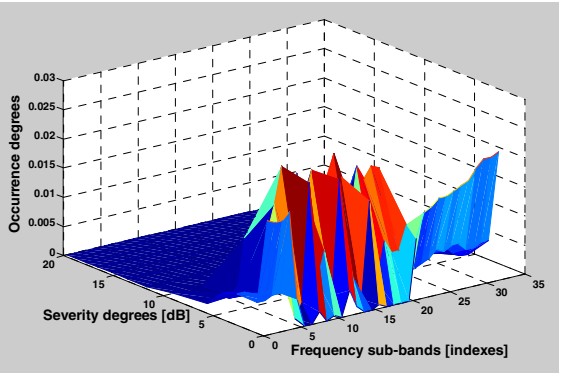


Figure 5.12. An occurrence distribution over the statistical network.

At this point, one can say that box cells supporting the biggest occurrence degrees are very likely directly associated to the defect type. But it is not that simple to build a group of such box-cells, based only on a set of occurrence degrees, because the attempt is rather empirical and affected by uncertainty. A systematic method to construct similarity classes by using statistical information is then necessary. Also, as already mentioned, it is desirable that every class be associated with some confidence degree.

The occurrence degrees are in fact values of some membership functions that change the nature of clusters from crisp to fuzzy. More specifically, consider the generic horizontal and vertical clusters, C_l and D_k, respectively. Then their associated membership functions are: $\mu_{m,l}(B_{l,k}) = \nu_m[l,k]$ for any box cell $B_{l,k} \in C_l$ and $\eta_{m,k}(B_{l,k}) = \nu_m[l,k]$ for any box cell $B_{l,k} \in D_k$.

Thus, $(C_l, \mu_{m,l})$ and $(D_k, \eta_{m,k})$ are now fuzzy sets. The new definitions are superior to the former ones, since the rsp norms have been accounted. Now, if the box cell $B_{l,k}$ that belongs to a crisp cluster C_l has a null occurrence degree, it cannot belong to the fuzzy cluster $(C_l, \mu_{m,l})$. For simplicity, denote the values of membership functions by $\mu_{m,l}[k]$ and $\eta_{m,k}[l]$, respectively (i.e., $\mu_{m,l}$ and $\eta_{m,k}$ could also be treated as vectors from $[0,1]^K$ and $[0,1]^L$, respectively).

Note that the membership functions change from a frame to another, though the crisp clusters are independent of frames. Hence, the fuzzy clusters have variable structure depending on the number of processed frames, which is closer to the real behaviour of sn (variable) structure.

Step 6: Associating certainty degrees with elementary crisp relations.
A unique *certainty degree* should be associated with every cluster C_l or B_k. This is a number that expresses, on the one hand, the certainty in considering the corresponding cluster and, on the other hand, the degree of box cells affiliation with the elementary fuzzy relation the cluster naturally generates. The membership matrix of elementary fuzzy relation is simply derived by multiplication between the cluster certainty degree and its characteristic matrix. This idea is developed next, but, first, the certainty degrees have to be evaluated.

The evaluation of certainty degrees is based on the concepts of *fuzzy* and *uncertainty measures* (Klir and Folger, 1988). Obviously, certainty is opposite to uncertainty. An interesting fuzzy/uncertainty measure is the Shannon Fuzzy Entropy (SFE). Its definition relies on the *multidimensional Shannon function* below:

$$S(x) = -\sum_{n=1}^{N}\left[x_n \log_2 x_n + (1-x_n)\log_2(1-x_n)\right],\ \forall x = [x_1\ \ldots\ x_n]^T \in [0,1]^N \qquad (34)$$

The Shannon function originated from the concept of *entropy*, first utilized in physics. Thus, if one restricts the sum in Eq. 34 to the first half, replaces "$\log_2$" by "$\ln$" (John Nepper's natural logarithm) and sets $x \in [0,1]^N$ as a discrete probability density (i.e., verifying $\sum_{n=1}^{N} x_n = 1$), then the entropy is obtained:

$$H(x) = -\sum_{n=1}^{N} x_n \ln x_n \qquad (35)$$

When $N = 1$, the entropy from Eq. 35 is associated with the event for which the probability was considered. The opposite event is described by the opposite probability: $1-x$. Hence, the second half of the sum in Eq. 34 becomes the entropy of the opposite event. The Shannon function thus expresses the total entropy of an entity, by accounting for not only its classical entropy, but also the entropy of its opposite. Note that, in Eq. 34, no restriction (like the one verified by probability densities) is imposed. The Shannon function is an instrument utilized in many domains, but was defined in the context of information theory, as a concept quantifying the information encoded or transported by an entity. Its unit is the bit. This is the reason the natural logarithm was replaced by $\log_2$ in the original definition of entropy.

Several interesting properties of the Shannon function could be noted. For this approach, the following two are of the most concern. Firstly, the function is bounded and reaches several null minima, but only one maximum. No other minima are possible, but the null ones are reached on the border of definition domain (the

hypercube $[0,1]^N$). Secondly, the maximum value is exactly the dimension of input argument, i.e., N. It is reached for the middle point argument, the function being symmetrical. For example, in Figure 5.13, the graphics of the only two Shannon functions that could be viewed are drawn.

When the argument in Eq. 34 is provided by values of the membership function describing a fuzzy set, the SFE is obtained. In this case, SFE has several interpretations. As a general fuzzy measure, SFE quantifies how close to the crisp state is the fuzzy set (or its *fuzziness*). The bigger the SFE value is, the less crisp the set (i.e., the fuzzier). But SFE could also play the role of *uncertainty measure*. Uncertainty has two major facets: *vagueness* and *ambiguity* (Klir and Folger, 1988; Ulieru *et al.*, 2000).

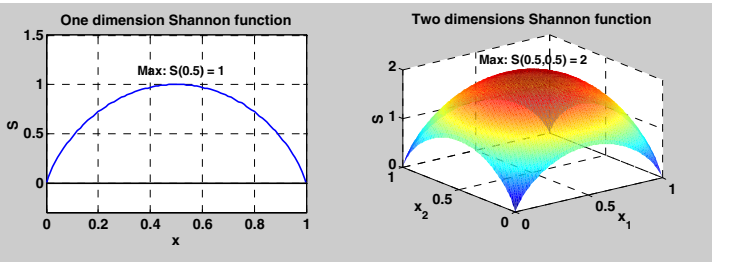


Figure 5.13. One- and two-dimension Shannon functions

The SFE is a *vagueness measure*. The bigger the SFE value is, the more vague the fuzzy set description, i.e., the more uncertain (or unreliable) the information about that set. Thus, maximum entropy means maximum uncertainty and fuzziness. The smaller the SFE values, the better.

Let us now get back into the context of previous steps. The certainty degree of a cluster should be opposite to its entropy (uncertain (vague) clusters should have small certainty degrees). Also, another property should be verified: the bigger the occurrence degrees of its box cells, the smaller its entropy. Since SFE has one maximum and several null minima (pointing to the lack of uncertainty), the values of membership functions $\mu_{m,l}$ and $\eta_{m,k}$ must be translated from [0,1] to [0.5,1] by a simple affine transformation, before using them subsequently:

$$v_m[l,k] \leftarrow \frac{v_m[l,k]+1}{2} \tag{36}$$

(By convention, one preserves the same notation for the translated values.)

Denote by $\alpha_{m,l}$ the certainty degree of horizontal cluster C_l, after processing $m+1$ frames (where $l \in \overline{0,L-1}$ and $m \in \overline{0,M-1}$). Similarly, $\beta_{m,l}$ stands for the certainty degree of vertical cluster D_k, after processing $m+1$ frames (where $k \in \overline{0,K-1}$ and $m \in \overline{0,M-1}$). The values of $\alpha_{m,l}$ and $\beta_{m,l}$ are then evaluated in three steps (by accounting for all previous remarks): compute the SFE of fuzzy clusters $(C_l, \mu_{m,l})$ and $(D_k, \eta_{m,k})$; normalize the SFE by K and,

respectively, by L; subtract the result from 1. This is summarized in Eqs. 37 and 38.

The normalization applied in Eqs. 37 and 38 is necessary because the certainty degrees have to vary only in the range [0,1], as well. This restriction is imposed by another meaning of a certainty degree, regarding the covers: any cluster belongs to an sn fuzzy cover with some membership degree. Actually, the crisp covers $\mathcal{H}$ and $\mathcal{G}$ are transformed into fuzzy covers, by a similar mechanism employed to transform crisp clusters into fuzzy clusters. Their membership functions are the following: $\mu_m(C_l)=\alpha_{m,l}$ for any cluster $C_l \in \mathcal{H}$ and $\eta_m(D_k)=\beta_{m,k}$ for any cluster $D_k \in \mathcal{G}$. Thus, $(\mathcal{H},\mu_m)$ and $(\mathcal{G},\eta_m)$ are now fuzzy sets, but their elements are other fuzzy sets (the fuzzy clusters, in fact). Like for fuzzy clusters, covers membership functions depend on the number of processed frames ($m+1$) (where $m \in \overline{0,M-1}$).

$$\begin{cases} \alpha_{m,l}=1-\dfrac{\mathcal{S}(\mu_{m,l})}{K} \\ \mathcal{S}(\mu_{m,l})=-\displaystyle\sum_{k=0}^{K-1}\left[\mu_{m,l}[k]\log_2\mu_{m,l}[k]+(1-\mu_{m,l}[k])\log_2(1-\mu_{m,l}[k])\right] \end{cases} \tag{37}$$

$$\begin{cases} \beta_{m,k}=1-\dfrac{\mathcal{S}(\eta_{m,k})}{L} \\ \mathcal{S}(\eta_{m,k})=-\displaystyle\sum_{l=0}^{L-1}\left[\eta_{m,k}[l]\log_2\eta_{m,k}[l]+(1-\eta_{m,k}[l])\log_2(1-\eta_{m,k}[l])\right] \end{cases} \tag{38}$$

Step 7: Constructing the α-sharp cuts of fuzzy relation.

Every fuzzy cluster generates, in association with its certainty degree, an elementary fuzzy relation between the box cells it includes. The membership matrix describing this relation is simply obtained by multiplication between the characteristic matrix of crisp cluster and the corresponding certainty degree. More specifically, if C_l and D_k are the generic horizontal and vertical clusters (as usual), then, after processing $m+1$ frames, their corresponding certainty degrees are $\alpha_{m,l}$ and $\beta_{m,k}$, respectively. One can denote by H_l and G_k the characteristic matrices of C_l and D_k, respectively. Then, obviously:

Horizontal cluster

$$H_l=\underbrace{\begin{bmatrix} O_{K\times K} & \cdots & O_{K\times K} & \cdots & O_{K\times K} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ O_{K\times K} & \cdots & U_{K\times K} & \cdots & O_{K\times K} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ O_{K\times K} & \cdots & O_{K\times K} & \cdots & O_{K\times K} \end{bmatrix}}_{L\times L\ \text{blocks}}$$

Vertical cluster

$$G_k=\underbrace{\begin{bmatrix} I^k_{K\times K} & I^k_{K\times K} & \cdots & I^k_{K\times K} \\ I^k_{K\times K} & I^k_{K\times K} & \cdots & I^k_{K\times K} \\ \vdots & \vdots & \ddots & \vdots \\ I^k_{K\times K} & I^k_{K\times K} & \cdots & I^k_{K\times K} \end{bmatrix}}_{L\times L\ \text{blocks}} \tag{39}$$

where the block $U_{K\times K}$ is located on the main diagonal in position (l,l) of matrix H_l, whereas the block $I^k_{K\times K}$ consists of one unit value on the main diagonal in position (k,k) of matrix G_k (all remaining values being null).

The elementary fuzzy relations are described by the following membership matrices: $\alpha_{m,l}H_l$ (horizontal) and $\beta_{m,k}G_k$ (vertical). The corresponding relations are α-sharp cuts of fuzzy relation after processing $m+1$ frames (where $m \in \overline{0, M-1}$). (See the definition of α-sharp cut in (Ulieru *et al.*, 2000.) In fact, this definition is similar to the definition of α-cut (Klir and Folger, 1988), but the inequality sign was replaced by the equality one.)

For example, recall the toy sn in Figure 5.11. For that structure, two horizontal and three vertical elementary fuzzy relations are available after every processed frame:

$$\alpha_{m,0}H_0 = \begin{bmatrix} \alpha_{m,0} & \alpha_{m,0} & \alpha_{m,0} & 0 & 0 & 0 \\ \alpha_{m,0} & \alpha_{m,0} & \alpha_{m,0} & 0 & 0 & 0 \\ \alpha_{m,0} & \alpha_{m,0} & \alpha_{m,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \alpha_{m,1}H_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{m,1} & \alpha_{m,1} & \alpha_{m,1} \\ 0 & 0 & 0 & \alpha_{m,1} & \alpha_{m,1} & \alpha_{m,1} \\ 0 & 0 & 0 & \alpha_{m,1} & \alpha_{m,1} & \alpha_{m,1} \end{bmatrix}$$

$$\beta_{m,0}G_0 = \begin{bmatrix} \beta_{m,0} & 0 & 0 & \beta_{m,0} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{m,0} & 0 & 0 & \beta_{m,0} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \beta_{m,1}G_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_{m,1} & 0 & 0 & \beta_{m,1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_{m,1} & 0 & 0 & \beta_{m,1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (40)$$

$$\beta_{m,2}G_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{m,2} & 0 & 0 & \beta_{m,2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{m,2} & 0 & 0 & \beta_{m,2} \end{bmatrix}$$

Equation 40 reveals another interesting property: the box cells that are very far from each other could not be in the same relation, even in the case of fuzzy relations. This is the case, for example, of box cells located at different severity levels and opposite subbands, such as $B_{0,0}$ and $B_{1,2}$ or $B_{1,0}$ and $B_{0,2}$. Practically, it is very unlikely that these box cells could associate together to reveal the same defect. But this property could be cancelled for the global fuzzy relation providing defect classifications, since such limitations are only intuitive.

Step 8: Constructing the fuzzy relation.

Two operations are applied in order to build the final fuzzy relation between sn box cells: aggregation of the (elementary) α-sharp cuts and evaluation of the transitive cover. The aggregation is simply performed through the max fuzzy union (Klir and Folger, 1988):

$$\mathcal{R}_m = \left[\bigcup_{l=0}^{L-1} \alpha_{m,l} H_l\right] \cup \left[\bigcup_{k=0}^{K-1} \beta_{m,k} G_k\right] \tag{41}$$

Thus, the membership matrix describing the crude fuzzy relation $\mathcal{R}_m$ is constructed by means of the elementwise max operator ("max•") applied on all matrices corresponding to fuzzy relation of the right term in Eq. 41:

$$\mathcal{M}_m = \max\bullet\left\{\max_{l\in 0,L-1}\bullet\left\{\alpha_{m,l}H_l\right\}, \max_{k\in 0,K-1}\bullet\left\{\beta_{m,k}G_k\right\}\right\} \tag{42}$$

Note that the same max operations like in Eq. 42 have been applied to obtain the characteristic matrices in Eq. 23, but by using unit certainty grades (since the relations were crisp). Obviously, the dimension of matrix $\mathcal{M}_m$ is $KL \times KL$.

For the toy example above, the membership matrix $\mathcal{M}_m$ is:

$$\mathcal{M}_m = \begin{bmatrix} \max\{\alpha_{m,0},\beta_{m,0}\} & \alpha_{m,0} & \alpha_{m,0} & \beta_{m,0} & 0 & 0 \\ \alpha_{m,0} & \max\{\alpha_{m,0},\beta_{m,1}\} & \alpha_{m,0} & 0 & \beta_{m,1} & 0 \\ \alpha_{m,0} & \alpha_{m,0} & \max\{\alpha_{m,0},\beta_{m,2}\} & 0 & 0 & \beta_{m,2} \\ \beta_{m,0} & 0 & 0 & \max\{\alpha_{m,1},\beta_{m,0}\} & \alpha_{m,1} & \alpha_{m,1} \\ 0 & \beta_{m,1} & 0 & \alpha_{m,1} & \max\{\alpha_{m,1},\beta_{m,1}\} & \alpha_{m,1} \\ 0 & 0 & \beta_{m,2} & \alpha_{m,1} & \alpha_{m,1} & \max\{\alpha_{m,1},\beta_{m,2}\} \end{bmatrix}$$

As one can see, some box cells are (co)related with various (un)certainty degrees, but between some other box cells no relationship seems to exist. The null values inside matrix $\mathcal{M}_m$ are always the same, independently of how many frames are processed (because of the horizontal and vertical crisp clusters), whereas the nonnull values vary from a frame to another (because of the occurrence degrees). Denote the generic element of $\mathcal{M}_m$ (i.e., the membership degree) by $\mathcal{M}_m[i,j]$ (where $i, j \in \overline{1,KL}$).

The resulting matrix $\mathcal{M}_m$ is *symmetric* and *reflexive* (since the elementary matrices H_l and G_k verify these two properties). Thus $\mathcal{R}_m$ is a *proximity* relation, but it is not necessarily *fuzzy transitive*. (See (Klir and Folger, 1988) for definitions.) Even though all elementary matrices H_l and G_k would describe (crisp) equivalence relations (i.e., all of them would be transitive as well), it is possible that $\mathcal{R}_m$ is nontransitive. This means, in general, $\mathcal{R}_m$ is not a *similarity* (fuzzy) relation. However, the similarity is a very important property, because the defect classes should also be (nonoverlapped) similarity classes. The direct involvement of similarity property in the construction of defect classes is revealed at the next step. Let us focus now on the transitivity property.

Actually, the transitivity property is the most difficult to insure in the case of fuzzy relations, because it is expressed (for example) as follows, differently from the crisp case (Klir and Folger, 1988; Ulieru *et al.*, 2000):

$$\mathcal{M}_m[i,j] \geq \max_{n\in 1,KL} \min\left\{\mathcal{M}_m[i,n], \mathcal{M}_m[n,j]\right\}, \ \forall i,j \in \overline{1,KL} \tag{43}$$

This is the *max-min (fuzzy) transitivity*. An equivalent matrix form of Eq. 43 can straightforwardly be derived:

$$\mathcal{M}_m \geq \bullet \left(\mathcal{M}_m \circ \mathcal{M}_m\right) \tag{44}$$

where "$\circ$" points to *fuzzy multiplication* (*product*) between matrices with compatible dimensions (involving the composition of the corresponding fuzzy relations). This multiplication is expressed starting from classical matrix multiplication, where max operator is used instead of summation and min operator is used instead of product. Also, "$\geq \bullet$"in Eq. 44 means that the ordering relation focuses on matrix elements and not globally, on matrices.

The lack of transitivity can be corrected by generating the *transitive closure* of $\mathcal{R}_m$, which is defined as the smallest transitive fuzzy relation including $\mathcal{R}_m$ (according to fuzzy inclusion) (Klir and Folger, 1988). A simple procedure allows us to compute this closure for any fuzzy relation $\mathcal{R}$:

Step 1. Compute the following fuzzy relation: $\overline{\mathcal{R}} = \mathcal{R} \bigcup (\mathcal{R} \circ \mathcal{R})$.

Step 2. If $\overline{\mathcal{R}} \neq \mathcal{R}$, replace $\mathcal{R}$ by $\overline{\mathcal{R}}$, i.e., $\mathcal{R} \leftarrow \overline{\mathcal{R}}$ and go to Step 3. Otherwise, $\overline{\mathcal{R}} = \mathcal{R}$ is the transitive closure of the initial $\mathcal{R}$.

It is not so difficult to prove that this procedure preserves the reflexivity and symmetry of $\mathcal{R}_m$ (Ulieru *et al.*, 2000), so that the transitive closure $\overline{\mathcal{R}}_m$ is a similarity relation. Also, in terms of membership matrices, $\mathcal{M}_m$ is replaced by $\overline{\mathcal{M}}_m$, derived according to the procedure above (but with max instead of union operator and with (max-min) fuzzy multiplication instead of composition operator).

The procedure is very efficient. The only limitation in terms of network granularity is here the dimension of $\mathcal{M}_m$ (i.e., $KL \times KL$), which could be very large. But, nowadays, the existing computing performances could yield reasonable running time for matrices with more than one million elements.

The main difference between $\mathcal{R}_m$ and $\overline{\mathcal{R}}_m$ is that $\overline{\mathcal{R}}_m$ is defined by means of a smaller number of membership degrees than $\mathcal{R}_m$. In general, small grades vanish. This is very suitable, since, probably, small membership degrees are mostly due to various noises still affecting the vibration data, even after filtering. In other words, by computing the transitive closure, *the statistical data have been denoised*. Another difference between the two fuzzy relations is that box cells previously unrelated (according to $\mathcal{R}_m$) could now be related (according to $\overline{\mathcal{R}}_m$). This means the nonnull values in $\mathcal{M}_m$ could overwrite the null ones. In general, inside the matrix $\overline{\mathcal{M}}_m$, null values could seldom appear. This effect is correcting the initial rough assumption that some box cells could never be related to each other.

Step 9: Generating the defect classifications.

The values in $\overline{\mathcal{M}}_m$ are referred to as (*fuzzy*) *confidence degrees*. The number of distinctive confidence degrees is $P_m \leq KL(KL+1)/2$, for each $m \in \overline{1, M-1}$ (due to symmetry). They could decreasingly be sorted: $\gamma_{m,0} > \gamma_{m,1} > \cdots > \gamma_{m,P_m-1}$ (by using natural new notations instead of $\mathcal{M}_m[i,j]$). For each confidence degree $\gamma_{m,p}$

($p \in \overline{0, P_m - 1}$), a partition of statistical network is generated, by evaluating the corresponding α *-cut* of fuzzy relation $\bar{\mathcal{R}}_m$ (Klir and Folger, 1988). Every α -cut plays the role of *defect classification* and is actually a partition of sn. Any class in such a partition gathers the cells with similar statistical properties and, therefore, is a *similarity class*. Obviously, all box cells with null occurrence degrees (see, for example, some high severity box cells in Figure 5.8) are grouped in an *inactive cluster* and do not actually partake in the classification. The inactive cluster is the same for any classification, if the number of processed frames, $m+1$, is constant, but its topology could change as m varies.

Let $\mathsf{C}_{m,p} = \left\{ \mathsf{F}_{m,p,q} \right\}_{q \in \overline{0, Q_{m,p} - 1}}$ be the defect classification corresponding to confidence degree $\gamma_{m,p}$ ($p \in \overline{0, P_m - 1}$). Inside, there are $Q_{m,p}$ defect classes generically denoted by $\mathsf{F}_{m,p,q}$. Usually, the classifications are listed in decreasing order of their confidence degrees. Moreover, it is well known that such an arrangement reveals a *holonic* behaviour (Ulieru *et al.*, 2000). That is, the confidence is also a measure of classifications granularity: as confidence decreases, a larger number of classes group more and more together. For maximum confidence, every cell is also a class, which means maximum of granularity as well ($Q_{m,0}$ equals the number of box cells with nonnull occurrence degrees). For minimum confidence, all cells are grouped in a single class, the granularity being also minimum ($Q_{m,P_m - 1} = 1$). Thus the trend of finite string $\left\{ Q_{m,p} \right\}_{p \in \overline{0, P_m - 1}}$ is decreasing when the confidence degree is decreasing. Only one classification shall be selected from this collection, as described in the next step.

Some examples of defect classes together with their confidence degrees are described in the section devoted to simulation results.

<u>Step 10</u>: Selecting the optimum classification.

Besides the confidence degree, the SFE of every class could also be evaluated. Actually, like in case of covers $\mathcal{H}$ and $\mathcal{G}$, every classification (an sn partition, in fact) is a fuzzy set with fuzzy sets (the defect classes) as elements. The membership functions associated with defect classifications are denoted by $\rho_{m,p}$ (where $m \in \overline{1, M-1}$ and $p \in \overline{0, P_m - 1}$). Thus, $\left(\mathsf{C}_{m,p}, \rho_{m,p} \right)$ is a fuzzy set and the membership function $\rho_{m,p}$ could be derived by means of a similar argument like in Step 6. There is, however, an important difference here. The entropy of a fuzzy set comprising fuzzy sets as elements should depend on the entropy of every element. If all elements would have small/large entropy values, then the set should also have small/large entropy. Consequently, the membership function $\rho_{m,p}$ has to reflect the normalized entropy of each defect class:

$$\begin{cases} \rho_{m,p} : \mathsf{C}_{m,p} \rightarrow [0,1] \\ \quad\quad \mathsf{F}_{m,p,q} \mapsto \rho_{m,p}\left(\mathsf{F}_{m,p,q}\right) \overset{def}{=} \rho_{m,p,q} = \frac{1}{2}\mathcal{S}_N\left(\mathsf{F}_{m,p,q}\right) \end{cases} \tag{45}$$

In Eq. 45, $\mathcal{S}_N\left(\mathsf{F}_{m,p,q}\right)$ is the normalized entropy of defect class $\mathsf{F}_{m,p,q}$ (where $q \in \overline{0, Q_{m,p}-1}$). To evaluate $\mathcal{S}_N\left(\mathsf{F}_{m,p,q}\right)$, first identify all the box cells that belong to $\mathsf{F}_{m,p,q}$ (together with their translated occurrence degrees – see Eq. 36), then use the definition in Eq. 34 and finally divide the result by the number of box cells. For example, consider that the following classification has been obtained inside the toy sn in Figure 5.11: $\mathsf{C}_{m,p} = \left\{\mathsf{F}_{m,p,q}\right\}_{q \in \overline{0,1}}$, where the defect classes are $\mathsf{F}_{m,p,0} = \left\{B_{0,0}, B_{1,1}\right\}$ and $\mathsf{F}_{m,p,1} = \left\{B_{0,1}, B_{0,2}, B_{1,2}\right\}$. (The box cell $B_{1,0}$ belongs to the inactive cluster.) Then:

$$\begin{aligned} \mathcal{S}_N\left(\mathsf{F}_{m,p,0}\right) = -\frac{1}{2}\big[&\nu_m[0,0]\log_2 \nu_m[0,0] + \left(1-\nu_m[0,0]\right)\log_2\left(1-\nu_m[0,0]\right) \\ &+ \nu_m[1,1]\log_2 \nu_m[1,1] + \left(1-\nu_m[1,1]\right)\log_2\left(1-\nu_m[1,1]\right)\big]; \end{aligned} \tag{46}$$

$$\begin{aligned} \mathcal{S}_N\left(\mathsf{F}_{m,p,1}\right) = -\frac{1}{3}\big[&\nu_m[0,1]\log_2 \nu_m[0,1] + \left(1-\nu_m[0,1]\right)\log_2\left(1-\nu_m[0,1]\right) \\ &+ \nu_m[0,2]\log_2 \nu_m[0,2] + \left(1-\nu_m[0,2]\right)\log_2\left(1-\nu_m[0,2]\right) \\ &+ \nu_m[1,2]\log_2 \nu_m[1,2] + \left(1-\nu_m[1,2]\right)\log_2\left(1-\nu_m[1,2]\right)\big] \end{aligned} \tag{47}$$

The division by 2 in Eq. 45 is required because SFE is nonmonotonic (recall Figure 5.13). The values of $\mathcal{S}_N\left(\mathsf{F}_{m,p,q}\right)$ varying in the range [0,1] are now restricted to the range [0,1/2], which involves the final entropy increases when the (translated) occurrence degrees decrease.

After the membership function $\rho_{m,p}$ has been evaluated, the entropy of classification $\mathsf{C}_{m,p}$ is computed by using again the definition stated in Eq. 34:

$$\mathcal{S}(\rho_{m,p}) = -\sum_{q=0}^{Q_{m,p}-1}\left[\rho_{m,p,q}\log_2 \rho_{m,p,q} + (1-\rho_{m,p,q})\log_2(1-\rho_{m,p,q})\right] \tag{48}$$

Note that the normalization is meaningless in Eq. 48, since the entropy also encodes information about the number of defect classes (clusters). Therefore, in general, the entropy values $\left\{\mathcal{S}(\rho_{m,p})\right\}_{p \in \overline{0,P_m-1}}$ prove a decreasing trend, since the number of defect classes (i.e., the maximum of entropy) decreases when the confidence degree decreases. This involves the entropy values $\left\{\mathcal{S}(\rho_{m,p})\right\}_{p \in \overline{0,P_m-1}}$ and the confidence degrees $\left\{\gamma_{m,p}\right\}_{p \in \overline{0,P_m-1}}$ are opposite.

A "good" classification should have high confidence degree and low entropy. This could be selected by means of a cost function that encodes the opposite behaviour of entropy and confidence degree. In order to define such a

function, it is first necessary to transform the entropy and the confidence degrees into maps comparable to each other. Before this operation, the comparison between them is impossible, because they vary in different ranges. The transformation is affine:

$$\gamma_{m,p}^{01} \overset{def}{=} \frac{\gamma_{m,p} - \min\limits_{p\in\overline{0,P_m-1}}\left\{\gamma_{m,p}\right\}}{\max\limits_{p\in\overline{0,P_m-1}}\left\{\gamma_{m,p}\right\} - \min\limits_{p\in\overline{0,P_m-1}}\left\{\gamma_{m,p}\right\}} \tag{49}$$

$$\sigma^{01}(\rho_{m,p}) \overset{def}{=} \frac{\sigma(\rho_{m,p}) - \min\limits_{p\in\overline{0,P_m-1}}\left\{\sigma(\rho_{m,p})\right\}}{\max\limits_{p\in\overline{0,P_m-1}}\left\{\sigma(\rho_{m,p})\right\} - \min\limits_{p\in\overline{0,P_m-1}}\left\{\sigma(\rho_{m,p})\right\}} \tag{50}$$

Obviously, both normalized maps $\gamma_{m,p}^{01}$ and $\sigma^{01}(\rho_{m,p})$ vary in the range [0,1] and, moreover, they are reaching the extreme values 0 and 1.

Define the cost function S_m as the geometric mean between the values of the map defined by Eq. 49 and the opposite values of the map defined by Eq. 50, over the classification indexes set:

$$\begin{bmatrix} \mathsf{S}_m & : & \overline{0,P_m-1} & \to & [0,1] \\ & & p & \mapsto & \mathsf{S}_m[p] \overset{def}{=} \sqrt{\gamma_{m,p}^{01}\left[1-\sigma^{01}(\rho_{m,p})\right]} \end{bmatrix} \tag{51}$$

In this context, S_m expresses the opposite entropy weighted by confidence degrees. Other cost functions could also be employed in this aim (such as the arithmetic mean or another algebraic combination between $\gamma_{m,p}^{01}$ and $\sigma^{01}(\rho_{m,p})$). But, in any case, this function could only have a finite number of maxima (or minima) that realize the trade-off between entropy and confidence degree. In the case of cost function S_m, the best compromise is reached for its global maximum. Thus, the best classification $\mathsf{C}_m^{opt} = \mathsf{C}_{m,p_m^{opt}}$ is selected by solving the following simple optimisation problem:

$$p_m^{opt} = \arg\max_{p\in\overline{0,P_m-1}} \mathsf{S}_m[p] \tag{52}$$

An example of how the optimum classification is selected by solving the problem stated by Eqs. 51 and 52 is displayed in Figure 5.14, where only 51 frames have been processed (including the virtual one). The opposite variation between the confidence degree and the (opposite) entropy, as well as the shape of their geometric mean are clearly drawn. In this example, 32 classifications are available and the optimum resulting index is $p_m^{opt} = 14$, which points to the 15th classification as being the optimum one. Note that the 19th classification is a sub-optimal one, though its entropy–confidence compromise is also maximum, but locally (and close to the global maximum). The number of classes inside the optimum classification is 82 (most of them being singletons). As one shall see in the section devoted to simulation results, the optimum classification constitutes an

image map about the specific fault(s) distorting the standard spectrum. Also, the classification confidence is $\gamma_{50,14} \cong 0.7$ and its entropy is $\mathcal{S}(\rho_{50,14}) \cong 81.06$. The entropy is quite high (close to its maximum, 82), since the number of processed frames is modest (only 51) and thus the occurrence degrees are inaccurate. As the number of processed frames increases, the entropy goes down, farther from its maximum.

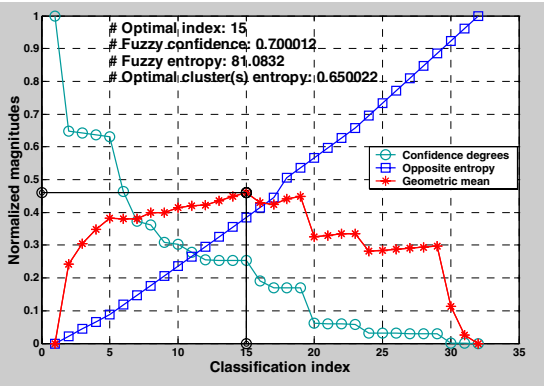


Figure 5.14. Selecting the optimum defect classification.

The most difficult part of the fuzzy model is the classification map interpretation (or analysis). This means that specific defects should be put into direct correspondence with map topologies. The subsequent analysis is more rigorous and simpler to perform than by inspecting the vibration spectrum, since a part of analyst reasoning has already been automated. Accounting for all classes in a map is sometimes sufficient to perform an accurate diagnosis. But, sometimes, this attempt leads to a rather complicated analysis. Therefore, some specific class (or a reduced number of classes) should be emphasized as representing the defect(s). One option is to consider the biggest class as revealing all subbands affected by the defect(s). A different option is to extract the minimum entropy class, which, in general, is smaller than the biggest class and, therefore, more focused on few subbands. These are very likely the most affected by defect(s). (Recall that minimum entropy means maximum occurrence degree of rsp norms.) Other representing classes could also be selected.

In order to complete the method, it is perhaps useful to show how an optimum cluster (or group of clusters) could be selected inside the best classification $\mathbf{C}_m^{opt}$ by using the normalized SFE as cost function. Thus, the index of optimum defect class(es) is (are) evaluated by solving the following optimisation problem:

$$q_m^{opt} = \underset{q \in \overline{0, Q_{m,p_m^{opt}} - 1}}{\arg\min} \mathcal{S}_N\left(\mathbf{F}_{m,p_m^{opt},q}\right) \tag{53}$$

For the example in Figure 5.14, the minimum (normalized) entropy of the optimum defect class inside the best classification is about 0.65. All the other

classes have entropy values at least as large as this value. Therefore, the corresponding cluster is representing in the best manner a specific defect.

But selecting an optimum defect class is less important than selecting the optimum classification. As already mentioned, sometimes, the classification configuration is itself a good image about defects, provided that its interpretation is not too difficult to perform. A very desirable property of such an interpretation is to reveal multiple defects by simple combinations of single defect maps. In general, this property is difficult to achieve. But the interpretation principle could be the same, independently of single or multiple-point defects generating the maps.

Note that, in this approach, the number of processed frames was considered variable. Though the notations are more complicated (the index m is omnipresent), one can clearly see how the concepts utilized inside are varying depending on this variable. The main reason the method was presented in terms of processed frames number is to show that its implementation could be performed by following either an on-line or an off-line strategy. For the on-line implementation the best classification should be provided after every processed frame (or group of frames). Step 8 is the critical one, since the evaluation of transitive cover could be time consuming when the product KL is too big (over 1500, with the actual computing performances). In this case, the best solution is to perform the defect classification only after several frames have been processed. This means the strategy is quasi-off-line (or even off-line). In general, the number of processed frames improves the method accuracy, since the estimation of occurrence degrees is more and more precise.

5.3. Simulation Results and Discussion

The two algorithms previously described constitute the kernel of a simulator designed to test the fuzzy-statistical reasoning method. The testing platform and the simulation results are described next.

5.3.1. The Testing Platform

The vibration data are acquired from bearings through a platform designed on purpose. Three main systems are connected, as illustrated by the pictures in Figure 5.15: a mechanical stand, a vibration data acquisition and pre-processing apparatus and a personal computer (PC).

The mechanical stand consists of the following elements:

1. A three-phase electrical engine, Siemens type, with maximum rotation speed of 2740 rot/min (about 45.67 Hz), working at 380 V and with a power of 370 W.
2. A couple of bearings mounted into mechanical seats, appropriately designed to fit to their geometry. The seats are easy to dismount in order to change the bearings, when necessary. The bearing near the engine is a standard high-quality one, without defects. The other bearing could also be standard (identical to the first one, in order to acquire the standard

vibration data) or a tested one, with possible defects (for raw vibration acquisition). All bearings are provided by Romanian and German industries. (See their geometry in Figure 5.16.).

3. A couple of metallic discs mounted between bearings, on the same axis, which play a double role. On the one hand, they produce a load of about 200 N applied in a radial-axial manner on bearings. This leads to a contact angle of 40° inside the bearings. On the other hand, they are creating an inertial momentum that rejects some external perturbations and keeps the rotation speed constant.

4. An elastic coupling between engine axis and load axis, aiming to attenuate the engine self-sustained vibrations or shaft wobbling that could corrupt the data.

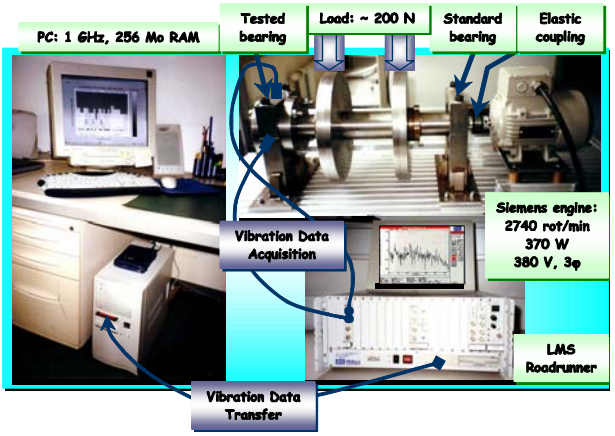
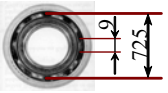


Figure 5.15. The bearings testing platform.



- # of balls: 13
- contact angle: 40°

Figure 5.16. Geometrical characteristics of tested bearings.

The geometrical characteristics of tested bearings (illustrated in Figure 5.16) lead in fact to a very small variation of natural frequencies, depending on contact angle. Thus, even the contact angle is not accurately set, and its influence over the natural frequencies is not decisive. The biggest natural frequency is about 325 Hz.

The vibration is acquired by using two light accelerometers. The definition in Eq. 2 is adopted to provide the complex valued vibration data, because both sensors are far enough from the direction of applied load. A very powerful apparatus has been employed to acquire vibration data: an *LMS Roadrunner* (LMS International, 1999). Its capabilities extend far beyond the minimal ones required by this method: accurate prefiltering of data, simultaneous acquisition on at least two channels and selectable recording format. The Roadrunner integrates a

microcomputer with a user-friendly interface that allows the user to work as comfortably as with any PC. It is also endowed with at least four channels (their number could be extended), compatible with a large number of sensors. The maximum allowed sampling frequency is 100 kHz. In this application, the sampling frequency has been set to $\nu_s = 25.6$ kHz. Data are saved in ASCII format, with 22 digits of representation. From Roadrunner, data are transferred to a PC, via floppy discs. The PC has the following main characteristics: 1 GHz (frequency), 256 Mb RAM (memory), 40 Gb (hard disk capacity). They rate the PC at the average of actual (public) technological level (years 2001, 2002). A laptop could also be successfully employed to implement the method.

5.3.2. Initial Simulation Parameters

In the description of the platform above, the shaft rotation speed and the sampling frequency were given: $\nu_r \cong 45.67$ Hz and $\nu_s = 25\,600$ Hz. Thus, a complete rotation takes about 21.9 ms, encoded by 560 vibration data samples. The vibration data length is set to $N = 2^{22} = 4{,}194{,}304$ samples, which takes 163.84 s in 7482 full rotations (see Eq. 1).

The vibration frame length is set to $N_f = 2^{13} = 8192$ samples (320 ms, ~15 full rotations). The number of nonoverlapped frames is then 512 (see Eq. 4), whereas every data segment includes three successive frames, as explained in the previous section. The frame length involves a frequency resolution of 3.125 Hz.

The window selected to smooth the overlapping between segments is Tuckey type, with 33.33% rectangular shape (see Figures 5.4 and 5.5). A high-pass filter will be applied to windowed segments. The LF cutoff frequency is set 7 times the largest natural frequency: $\nu_{lc} = 7 \times 325 = 2275$ Hz.

The vibration spectrum is segmented into $K = 32$ subbands. Every subband includes 128 rays for a bandwidth of 400 Hz. This setting realizes a good compromise between sp accuracy (each one is computed by using 128 spectral values) and bandwidth. The severity levels are set as already explained ($L = 12$ levels).

5.3.3. Comparative Discussion on Simulation Results

The experiments have been organized according to the following scenario:

1. Collect raw vibration data from four tested bearings: a standard (defect free) one (labelled **S720913**, according to its geometry); one with a chop on the inner race (**I720913**); one with a spall on the outer race (**O720913**); one with chops on both inner and outer races (**M720913**).
2. Apply EA to detect the severity degree of defects and to check if multiple defects on bearing **M720913** are visible or not. The following settings are performed in this aim: consider vibration segments of more than 1 s length; operate with 1/3-octave filters appropriately designed (as described in (Barkov *et al.*,1995a,b)); take full rectified envelope; focus on

the LF sub-band of envelope spectrum, for a bandwidth at least equal to 2 kHz.

3. Apply the fuzzy reasoning method.

A. Envelope analysis results

A standard horizontal vibration data segment of about 1.3 s (32,768 samples, 4 frames) and a zoom on the portion between 0.2 s and 0.25s are represented Figure 5.17a. The shape is almost harmonic, as expected. In Figure 5.17b, the corresponding spectrum is represented in dB, with a resolution of 0.78125 Hz (16,384 rays on half band 0-12.8 kHz). The energy of vibration is practically concentrated in LF-MF subband 0-5 kHz. The sensor resonance is insignificant. The peaks into the LF band are due to bearing natural frequencies. The envelope of standard signal, as well as a similar zoom as before, is drawn in Figure 5.18a. Signals appear very close to the white noise. Actually, the LF part of the envelope spectrum in Figure 5.18b reveals a quasi-constant variation on all frequencies around the spectral acceleration of 102.7 cm/s^2 (the spectrum average).

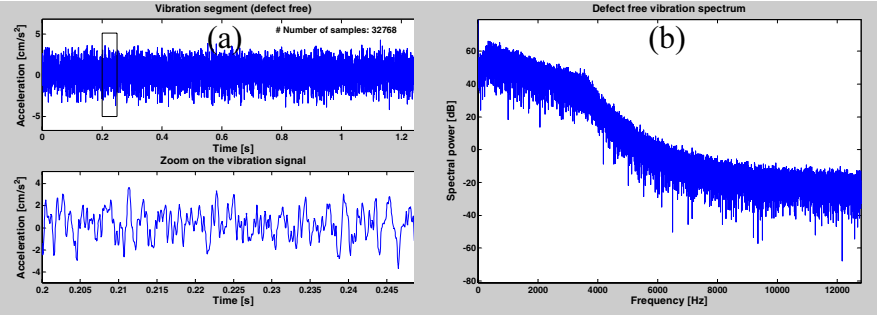


Figure 5.17. Standard vibration (a) and its spectrum (b) (bearing **S720913**).

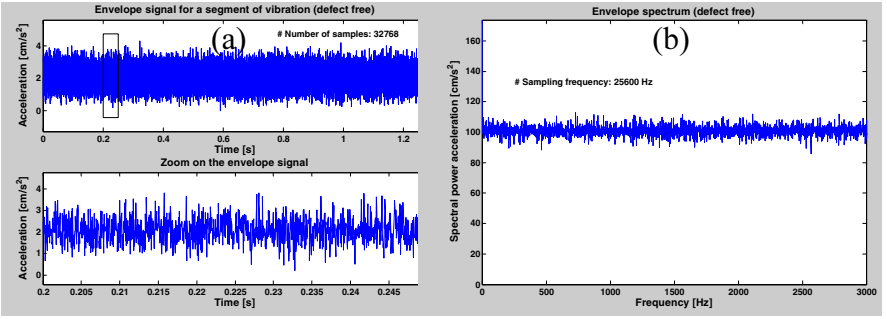


Figure 5.18. Standard envelope vibration (a) and its spectrum (b) (bearing **S720913**).

For the next three cases, the length of vibration data segments is identical to the standard set above. The vibration segments are represented in Figures 5.19a, 5.20a and 5.21a, while their corresponding spectra are found in Figures 5.19b, 5.20b and 5.21b (see Appendix). The time variations appear to be more irregular than previously. The harmonic behaviour is distorted by a noise encoding the defect

type. The effect of modulation could be seen in the zoomed portions of these signals, especially for the outer race defect. The sensors are forced to resonate and this effect is replicated towards LF and MF bands in all spectra. A significant resonance peak is located in band 6-10 kHz for all defective bearings, while it is missing for standard bearing. Also, peaks are more emphasized within the LF spectral zone for defective bearings than for standard.

The 1/3-octave filter (specific to EA) has been designed such that its central frequency is located somewhere in the median spectral valley between 4 and 6 kHz. Actually, it is selected as the minimum point of the spectral median in subband 4-6 kHz. This corresponds to the selection performed in (Barkov *et al.*,1995a,b) where the central frequency is located in a subband corresponding to the flattest zone of spectrum. In this specific case, the bandwidth is determined by the resonance peak flanking the valley to the right (in subband 6-10 kHz). The bandwidth is set as 3/2 times the difference between the location of this peak and the central frequency. The filter length is set to 2048 coefficients, in order to preserve high accuracy of filtering.

Figures 5.19, 5.20 and 5.21 are also concerned with the envelope signals (c) and the corresponding (envelope) spectra (d). In the case of single-point defects (Figures 5.19c and 5.20c), the abnormal behaviour is illustrated by the spectral envelope prominent peaks located around the multiples of natural frequency corresponding to the defective part: BPFI = 325.061 Hz (*Ball Pass Frequency on the Inner race*) or BPFO = 268.606 Hz (*Ball Pass Frequency on the Outer race*). The peaks decay exponentially, such that starting from the 9th multiple, they are practically sunk into the noisy part of spectrum. The severity degree is quite easy to estimate from these graphics, if the height of the largest peak is compared to the average standard envelope spectrum: about 3.5 (i.e., 10.88 dB) for inner race defect and about 4.5 (i.e., 13.06 dB) for outer race defect. This rates the defects as medium ones. Note, however, that the estimation could not be extremely accurate, since the vibration segments lengths are small (only 1.3 s, i.e., about 59 full rotations). An accurate estimation requires at least 100 rotations, but this increases the noisy part in all spectra, such that spectral estimation techniques should be employed (Oppenheim and Schafer, 1985; Proakis and Manolakis, 1996), in order to provide readable spectra.

Refer now to the multiple-point defect (Figure 5.21c). The envelope spectrum is so noisy that, practically, it is impossible to isolate some characteristics related to the defect type, though the spectrum in Figure 5.21b does not look very different from the spectra in Figures 5.19b and 5.20b. The energy increase revealed by the envelope spectrum is mainly due to the vibration signal itself (see Figure 5.21a), which has a larger energy level than in the case of single-point defects (Figures 5.19a and 5.20a). But the general level of noise is also increased. The EA failure in this case could have some plausible explanations. Besides the 1/3-octave filter selection (note that EA is very sensitive to this filter), perhaps the vibration model considered here cannot match the interpretation principle that worked well in the case of single-point defects (i.e., associate the natural frequencies directly to defect nature and location).

B. Fuzzy-statistical reasoning results

The three vibration signals have been entered into two MATLAB programs implementing the method described in the previous section. Thus, after collecting all information about rsp norms occurrences in box cells of sn, three occurrence degrees distributions have been obtained. For single point defects, there are two main rsp norms concentrations: one for MF and one for HF zones, but the most rsp norms seem to occur in the MF zone. On the contrary, for multiple defects, they occur rather in the HF zone. Thus, a first criterion for discriminating between single- and multiple-point defects is revealed.

After constructing the fuzzy model, a number of faults classifications resulted, for each tested bearing: 30 for **I720913**, 32 for **O720913** and 27 for **M720913**. The selection of an optimum faults classification is automatically performed (as described). The trade-off between the confidence degree (Eq. 49) and the classification entropy (Eq. 50) is quantified by means of geometric mean criterion (Eq. 51) that points to the optimal classification index. The variation of confidence degree and entropy among classifications as well as the shape of the geometric mean are illustrated in Figures 5.22, 5.25 and 5.28, for each bearing. The optimum classification indexes are: #20 for **I720913**, #20 for **O720913** and #17 for **M720913**. The corresponding optimum classification maps are pictured in Figure 5.23 (inner race defect), Figure 5.26 (outer race defect) and Figure 5.29 (multiple defects). For each classification, the representation is illustrated by using the grey levels scale to the right. One recognizes the sn by looking at the grid of each map. Thus, box cells that belong to the same class (cluster) have the same colour. Moreover, inside every box, the index of class the box belongs to is written, except the boxes that do not partake in the classification and belong to the inactive cluster. Besides the numerical parameters describing the classification minimum entropy (optimal) cluster, the average of rsp norms is represented as a curve passing through the map. Obviously, clusters are more or less grouped around this curve for all classifications.

As already mentioned, the most difficult part of the fuzzy model is the interpretation (or analysis) of classification maps. This means specific defects should be put into direct correspondence with map topologies. Such an analysis is more rigorous and simpler to perform than by inspecting the vibration spectrum, since a part of analyst reasoning has already been automated.

The shape of inactive cluster or of the rsp norms average could already constitute an image of defect types. For the three optimum classifications described above, the inactive clusters are all different, though their shapes are closer to each other for single-point defects. But this effect is noticed in EA as well: Figures 5.19d and 5.20d are not very different, since the values of the two corresponding natural frequencies are close to each other (BPFI=325.061 Hz and BPFO=268.606 Hz). The inactive cluster for multiple defects seems to be quite different, but the same interpretation principle or rules as for single-point defects could be used. In the case of EA, the interpretation rule that worked very well for single-point defects is useless in the case of multiple-defect spectrum (Figure 5.21d).

Another entry yielding map interpretation is to focus not on the inactive cluster, but rather on the active ones. Of course, one could consider all classes in a map (optimal or suboptimal). But this involves a complicated analysis. Therefore, some specific class (or a reduced number of classes) should be emphasized as

representing the defect(s). An option is to consider the biggest class as revealing all subbands affected by the defect(s). A different option is to extract the minimum entropy class (optimal cluster), which, in general, is smaller than the biggest class and, therefore, more focused on a few subbands. These are very likely the most affected by defect(s). (Recall that minimum entropy means maximum occurrence degree of rsp norms.) Other representing classes could also be selected.

The optimal detected clusters are the following:

a. for bearing **I720913** (inner race defect): cluster #13, with normalized entropy 0.516168, focusing on subband 5200-5600 Hz (MF);
b. for bearing **O720913** (outer race defect): cluster #20, with normalized entropy 0.711234, focusing on subband 4800-5200 Hz (MF);
c. for bearing **M720913** (inner and outer race defects): cluster #27, with normalized entropy 0.709225, focusing on subband 12.4-12.8 kHz (very HF).

That the optimal clusters #13 and #20 are located in adjacent box cells is not coincidental, but is due to the fact that the corresponding natural frequencies have values close to each other. The extreme HF subband pointed by the multiple defects is somehow surprising. A better interpretation could be given by considering other sub-optimal classifications (see the next discussion). But, in any case, a good insight concerning the "full optimality" (optimal clusters into optimal classifications) is the following: single-point defects are indicated by optimal clusters around the LF or MF peaks of rsp norms average (and there is a correlation between natural frequencies and focused subbands), while the optimal clusters of multiple-point defects seem to be located around the HF peak of average. A more refined frequency segmentation, with a larger number of subbands than here ($K \geq 64$) could probably help the user to make a sharper distinction between focused subbands in the case of single-point defects. Practically, the EA results are obtained by the fuzzy reasoning method as well. Concerning the multiple-point defects, it is possible that a frequency interpretation in terms of natural frequencies cannot be performed, but increasing K should lead to the same effect: the distinction between different defects should be easier to achieve. Unfortunately, the number of subbands (K) can only be increased at the expense of running time, especially due to the procedure evaluating the fuzzy transitive closure, which is the most time-consuming part of the algorithm (exponential type).

The severity degree estimated here is located on the 4th level (between 6 and 9.54 dB) – the first medium severity one – for single-point defects and on the 5th level (between 9.54 and 12.04 dB) for multiple defects. The first location is close to the severity degree estimated by EA for inner race defect (10.88 dB), but quite different from the outer race defect estimated severity (13.06 dB). For multiple defects, EA offers no severity degree estimation, but in this case the location of multiple defects optimal cluster is closer to the outer race severity (13.06 dB). Both estimations here are below the estimations proposed by EA. Since the severity degrees are conventionally set and in both methods the raw vibrations have been affected by filtering, the comparison in terms of severity degree is probably irrelevant. One could only note that, for the fuzzy-statistical method, the estimated

severity degree for multiple defects seems to be plausible, because the general level of vibration noise has been increased. This effect is proven by Figures 5.17a, 5.19a, 5.20a and 5.21a, where the amplitude of corresponding raw vibrations is about 3 cm/s^2 for standard and inner race defect, 2 cm/s^2 for outer race defect, but 5 cm/s^2 for multiple defects.

In order to extract more insights concerning classification map interpretation, several classifications should be depicted around the optimal ones. Their confidence and granularity are decreasing with classification index (according to the holonic phenomenon). In this context, some suboptimal classifications have been represented in Figure 5.24 (inner race defect), Figure 5.27 (outer race defect) and Figure 5.30 (multiple defects). They are selected according to the geometric mean values of Figures 5.22, 5.25 and 5.28. Thus, the suboptimal classifications have the best geometric mean values under the maximum one in every case. Sometimes, this requirement is fulfilled by local maxima, as in the case of bearings **O720913** and **M720913**. One could notice how box cells are more and more grouped together as the classification index increases.

An interesting observation could be noted with regard to all these maps: the optimal cluster (indicated by the optimal classification) is also optimum (with minimum entropy) for a large number of suboptimal classifications surrounding the optimal one, in the case of single-point defects. Though its index is changing (due to holonic phenomenon), its location is identical. The optimal cluster persistence among faults classifications is another good insight about the single-point defect nature, because, for multiple defects, the optimum cluster changes among classifications. However, in the case of multiple defects, it seems that another optimal cluster could also be considered, but extracted from suboptimal configurations. This is in fact the cluster #11 in classification #16 (as well as in classifications #13, #14, and #15, although not shown here). If one revisits Figure 5.27, one could notice that all these classifications, though suboptimal, prove a good compromise between confidence and entropy (they are only slightly below the optimal classification). Their unique optimum cluster focuses on the subband 7200-7600 Hz (still on the HF peak), but points to a lower severity degree (on level 3-6 dB, incipient).

One can infer from this analysis that selecting the cluster detected as optimal for the maximum number of classifications could be a good hint about the defect nature. But a reliable diagnosis requires a whole set of inference rules (and not isolated ones), in order to associate classification maps with specific defects and their severity degrees. A good achievement is that, by fuzzy-statistical reasoning, defects could be classified regardless of their nature as single- or multiple-point ones.

5.4. Concluding Remarks

Although with some obvious limitations, the method presented above aims to automate a part of human reasoning when detecting and classifying defects and to improve the multiple defect diagnosis. The main advantage of this method is that the defect classification maps could allow the user to perform a reliable detection

and diagnosis of defects, independently of their nature. Another advantage is its generality. On the one hand, the natural oscillation frequencies of the tested component play only a secondary role. On the other hand, gears, belt transmissions, or other vibration sources could replace bearings, provided that at least a good description of possible defects is a priori known in each case. Note that prefiltering is not mandatory: the fuzzy model could work with the whole raw vibration as well as with prefiltered data. The method's main drawbacks are the complexity (slightly bigger than EA complexity) and the difficulties in finding appropriate interpretations for classification maps.

Approaching the human reasoning in fault diagnosis is a demanding task. Not only because human reasoning is a complex mechanism (far to be completely understood nowadays), but also because such an attempt is mostly concerned with the inexplicable part of reasoning.

5.5. Appendix. Graphical Simulation Results

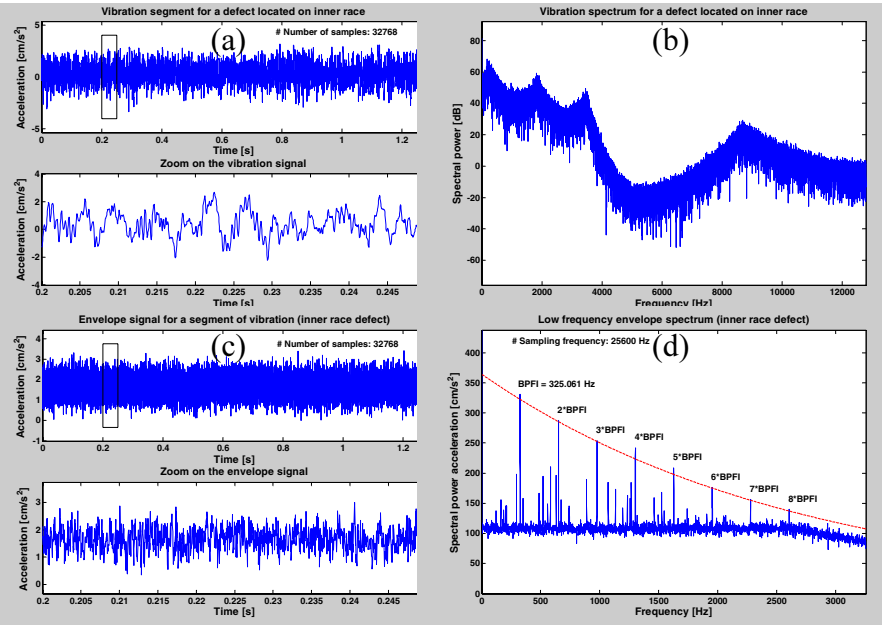


Figure 5.19. Envelope analysis for bearing `I720913`.

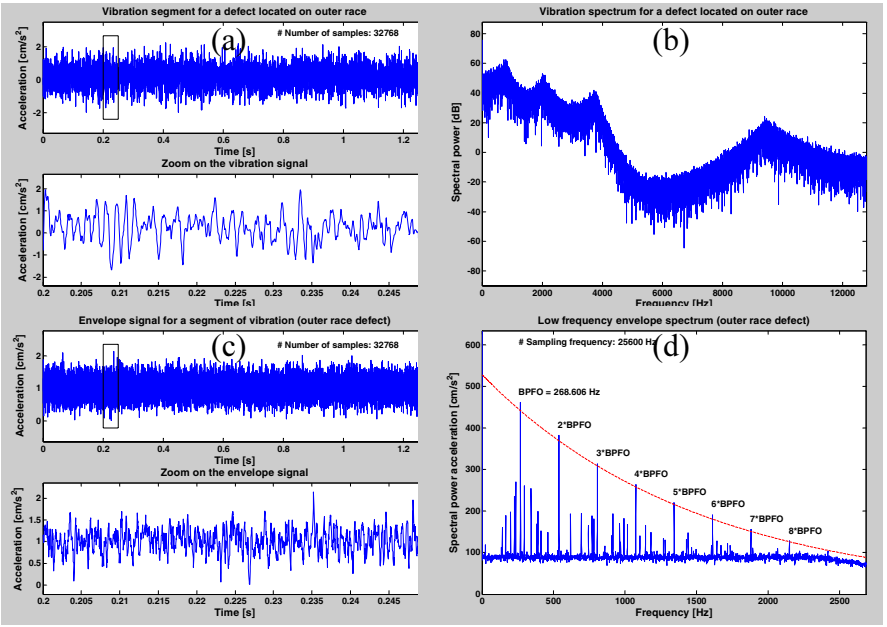


Figure 5.20. Envelope analysis for bearing **O720913**.

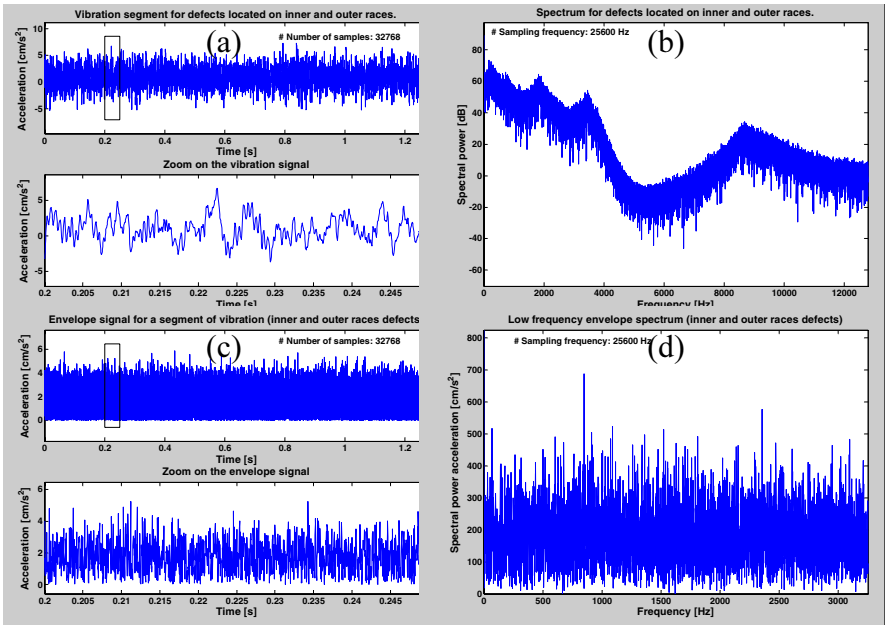


Figure 5.21. Envelope analysis for bearing **M720913**.

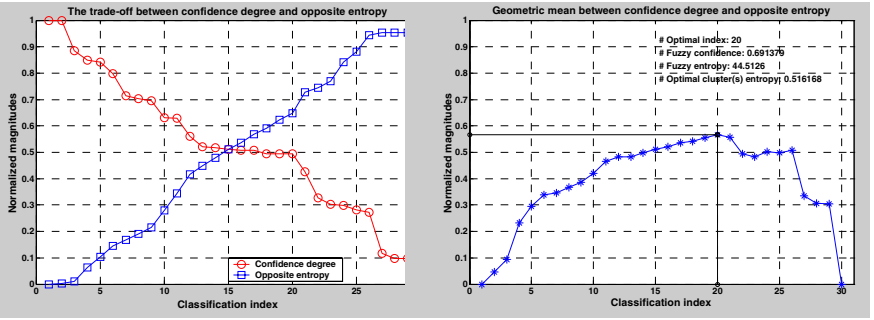


Figure 5.22. Selecting the optimum defect classification for bearing **I720913**.

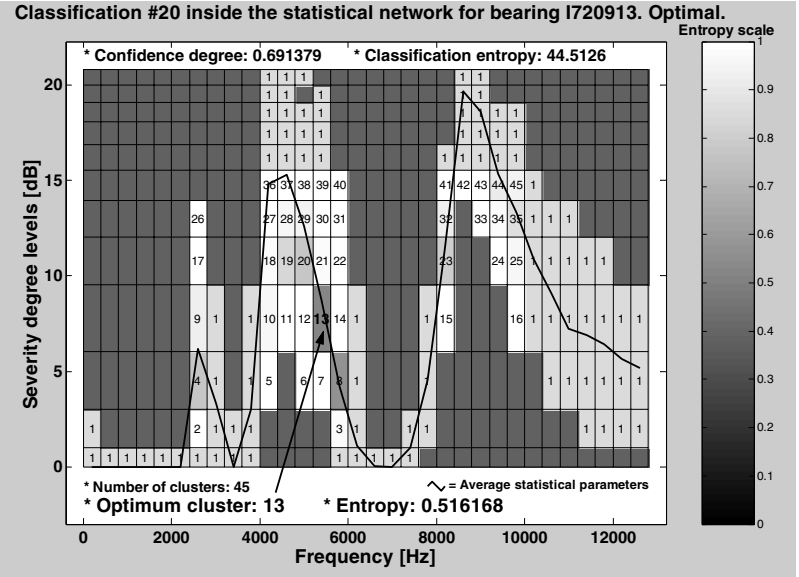


Figure 5.23. Optimum defect classification # 20 for bearing **I720913**.

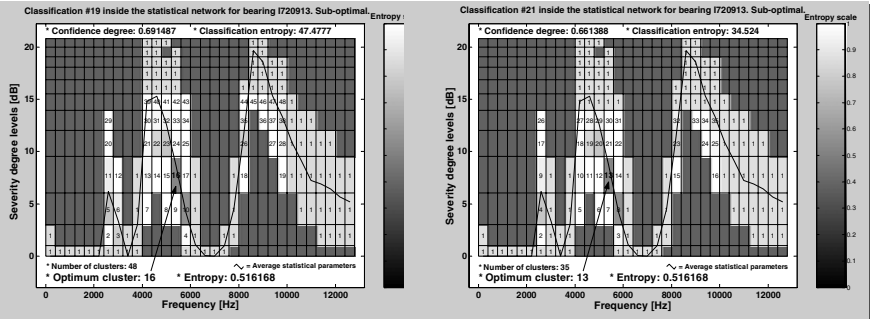


Figure 5.24. Suboptimal defect classifications #19 and #21 for bearing **I720913**.

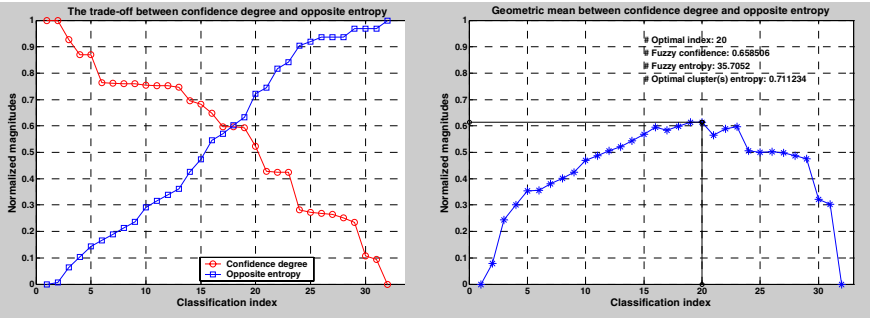


Figure 5.25. Selecting the optimum defect classification for bearing O720913.

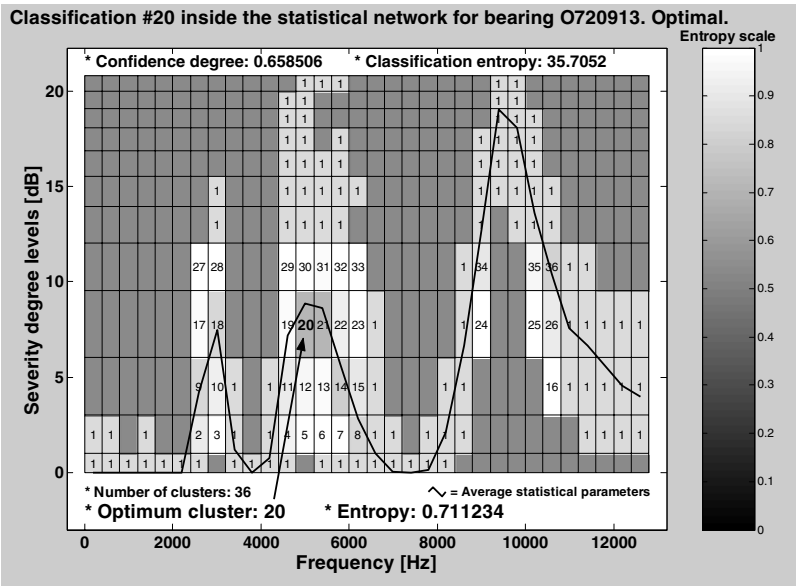


Figure 5.26. Optimum defect classification # 20 for bearing O720913.

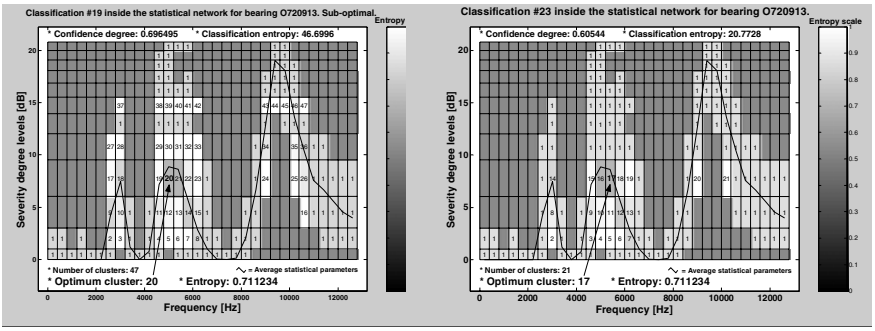


Figure 5.27. Suboptimal defect classifications #19 and #23 for bearing O720913.

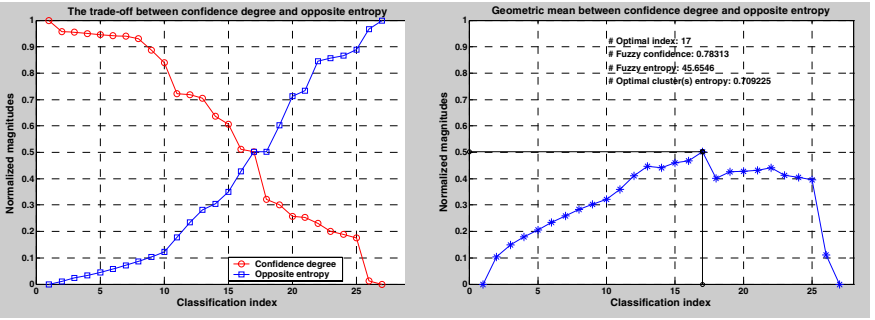


Figure 5.28. Selecting the optimum defect classification for bearing M720913.

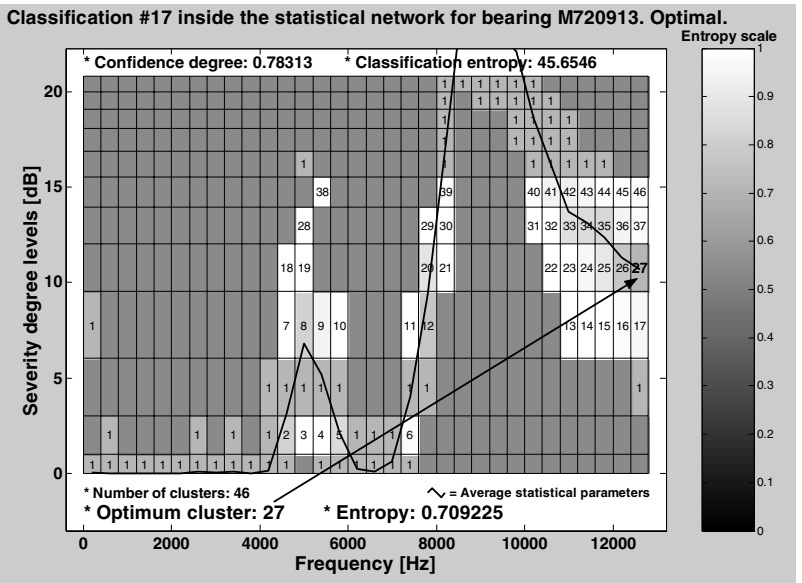


Figure 5.29. Optimum defect classification # 17 for bearing M720913.

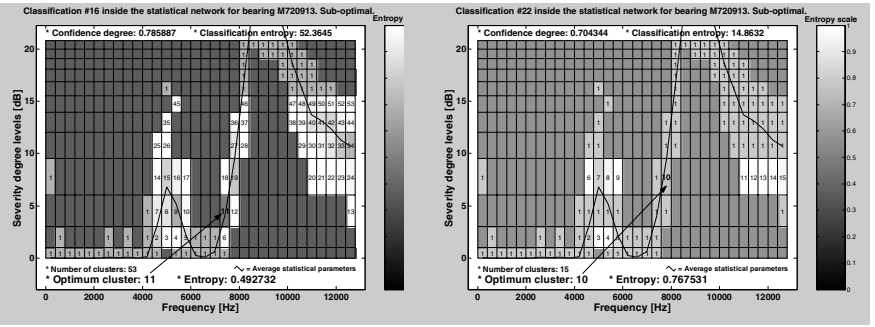


Figure 5.30. Suboptimal defect classifications #16 and #22 for bearing M720913.

References

1. Angelo M (1987) Vibration Monitoring of Machines. Bruel & Kjiaer Technical Review 1:1–36
2. Barkov AV, Barkova NA, Mitchell JS (1995a) Condition Assessment and Life Prediction of Rolling Element Bearings – Part 1. Journal of Sound and Vibration 6:10–17, June 1995 (http://www.inteltek.com/articles/sv95/part1/index.htm)
3. Barkov AV, Barkova NA, Mitchell JS (1995b) Condition Assessment and Life Prediction of Rolling Element Bearings – Part 2. Journal of Sound and Vibration 9:27–31, September 1995 (http://www.inteltek.com/articles/sv95/part2/index.htm)
4. Bedford A, Drumheller DS (1994) Introduction to Elastic Wave Propagation. John Wiley & Sons, Chichester, UK
5. Braun S (1986) Mechanical Signature Analysis. Academic Press, London, UK
6. Cohen L (1995) Time-Frequency Analysis. Prentice Hall, New Jersey, USA
7. FAG OEM & Handel AG (1996) Wälzlagerschäden – Schadenserkennung und Begutachtung gelaufener Wälzlager. Technical Report WL 82 102/2 DA
8. FAG OEM & Handel AG (1997) Rolling Bearings – State-of-the-Art, Condition-Related Monitoring of Plants and Machines with Digital FAG Vibration Monitors. Technical Report WL 80-65 E
9. Howard I (1994) A Review of Rolling Element Bearing Vibration: Detection, Diagnosis and Prognosis. Report of Defense Science and Technology Organization, Australia
10. Isermann R (1993) Fault Diagnosis of Machines via Parameter Estimation and Knowledge Processing. Automatica 29(4):161-170
11. Isermann R (1997) Knowledge-Based Structures for Fault Diagnosis and its Applications. In: Proceedings of the 4th IFAC Conference on System, Structure and Control, SSC'97, Bucharest, Romania, pp.15-32
12. Kaiser JF (1974) Nonrecursive Digital Filter Design Using the I_0–sinh Window Function. In: Proceedings of the IEEE Symposium on Circuits and Systems, pp.20-23
13. Klir GJ, Folger TA (1988) Fuzzy sets, Uncertainty, and Information. Prentice Hall, New York, USA
14. LMS International (1999) LMS Scalar Instruments Roadrunner. User Guide. LMS Scalar Instruments Printing House, Leuven, Belgium
15. Maness PhL, Boerhout JI (2001) Vibration Data Processor and Processing Method. United States Patent No. US 6,275,781 B1 (http://www.uspto.gov/go/ptdl/)
16. McConnell KG (1995) Vibration Testing. Theory and Practice. John Wiley & Sons, New York, USA
17. Oppenheim AV, Schafer R (1985) Digital Signal Processing. Prentice Hall, New York, USA
18. Proakis JG, Manolakis DG (1996) Digital Signal Processing. Principles, Algorithms and Applications (third edition). Prentice Hall, Upper Saddle River, New Jersey, USA
19. Reiter R (1987) A Theory of Diagnosis from First Principles. Artificial Intelligence 32: 57-95
20. Söderström T, Stoica P (1989) System Identification. Prentice Hall, London, UK

21. Stefanoiu D, Ionescu F (2002) Mathematical Models of Defect Encoding Vibrations. A Tutorial. Journal of the American-Romanian Academy (ARA), Montréal, Canada, Vol. 2001-2002
22. von Tscharner V (2000) Intensity Analysis in Time-Frequency Space of Modelled Surface Myoelectric Signals by Wavelets of Specified Resolution, preprint
23. Ulieru M, Stefanoiu D, Norrie D (2000) Identifying Holonic Structures in Multi-Agent Systems by Fuzzy Modeling. In: Kusiak A & Wang J (eds) Art for Computational Intelligence in Manufacturing, CRC Press, Boca Raton, Florida, USA
24. Willsky AS (1976) A Survey of Design Methods for Failure Detection Systems. Automatica 12:601-61
25. Wowk V (1995) Machinery Vibration. Balancing. McGraw-Hill, Upper Saddle River, New York, USA
26. Xi F, Sun Q, Krishnappa G (2000) Bearing Diagnostics Based on Pattern Recognition of Statistical Parameters. Journal of Vibration and Control 6:375–392