Dynamical Signatures of Gearbox Vibrations

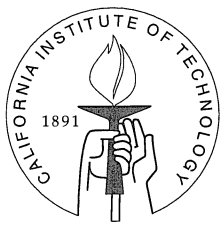
Thesis by

Slawomir M. Zaremba

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

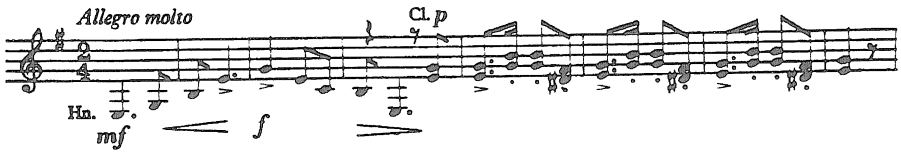


California Institute of Technology

Pasadena, California

1998

(Submitted September 10, 1997)



Symphony No. 9 in E Minor, "From the New World"

Antonin Dvořák

Acknowledgements

One of the benefits of having completed a Ph.D. thesis is an opportunity to acknowledge and thank the people who provided their help during this endless process. I would like to express a special word of thanks to my advisors, Prof. Thomas K. Caughey and Prof. James L. Beck. I appreciate all the precious time, loyal support, and tremendous patience Prof. Caughey had given me over these long years. I am indebted to Prof. Beck for his hearty support and for invaluable and stimulating suggestions, which not only resulted in a significantly improved quality of this project but also will certainly bear fruits in years to come.

I would like to thank Doug Perry and Jim Fanson from the Jet Propulsion Laboratory for providing equipment for the data acquisition, and two of my fellow students, Mike Vanik and Bob Uy, for their precious time and enthusiastic help with the experiment.

During most of my stay at Caltech, my work was funded by the H. & G. Bauer Fellowship at Caltech, whose generous support is greatly appreciated.

Last but not least, I would like to give special thanks to my wife, my inspiration and the most faithful supporter over these years, for raising my motivation and for proofreading the entire manuscript.

Abstract

The applicability of a dynamical systems approach to the analysis of gearbox vibration signatures is investigated. The signal acquired from a standard one-step helical gearbox is analyzed and the existence of the low-dimensional nonlinear and chaotic behavior is examined. For this purpose, the criteria of broadband spectrum, sensitivity to initial conditions, positive Lyapunov exponents, and short-term dynamical predictability are applied. The largest Lyapunov exponent is also used to quantify the predictability of the measured time series and a surrogate data test is performed to confirm that the analyzed signal is unlikely to correspond to a linear stochastic time-invariant model. To confirm the applicability of a dynamical systems approach, forecasts of the gearbox time series are produced, using various predictors in the state space. Several methods of dynamical signal separation are also discussed. Finally, a family of statistics is proposed, on the basis of geometric features of the system dynamics in the state space, to serve as dynamical monitoring tools for observing changes in the operating conditions. These are tested using several models of dynamical systems which exhibit behavior similar to that observed in the actual gearbox. The most promising monitoring statistics are recommended for additional studies of gearboxes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Why study gearbox vibration signatures?

A very large percentage of all industrial maintenance costs involves mechanical systems. Since gearboxes are a basic component of almost any power transmission system, their condition monitoring and fault detection is of primary interest to any system maintenance engineer. To monitor the condition of gears in an operating machine, a vibration signal is usually acquired from the gearbox casing. This signal is later analyzed in hope of finding information describing the gear condition and detecting early signs of possible damage.

Why nonlinear dynamical approach?

In gearboxes the existence of the modulation sidebands which appear in the spectrum near the meshing frequencies is often attributed to local tooth defects [34, 31, 6]. Similar sidebands occur also in the spectra of good gears due to nonlinearities and errors inherent in their construction, i.e., transmission errors, non-identical teeth, imperfect involute form of gear-tooth profiles, etc. The modulation is usually assumed in the form of amplitude and phase (or frequency) modulation functions approximated by their finite-length discrete Fourier series [34]. Since there are many sources of nonlinear behavior in gearbox vibrations (e.g., contact and stiffness forces, backlash), the hypothesis that the modulation sidebands result from a dissipative nonlinear dynamical (possibly chaotic) system is examined. An affirmative result would have some very important implications. It would suggest, for example, that different modulations of different harmonics of the meshing frequency result from (or rather, can be modeled by) a low-dimensional nonlinear dynamical system. In this case, in addition to looking for linear, spectral features, one should benefit from analyzing the problem using multivariate methods that utilize the knowledge of the system dimensionality and the structure of the underlying attractor. It is known that nonlinear, and particularly chaotic systems, are often extremely sensitive to even the slightest changes

in the operating conditions. There are many proven tools to quantify such behavior (e.g., Lyapunov exponents or entropy) and these can be particularly useful for the purposes of system diagnostics and condition monitoring.

The dynamical time-delay embedding approach differs from traditional univariate analysis in two fundamental aspects:

- It provides information about the degrees of freedom in addition to the one that is directly measured, although it may not capture all the degrees of freedom that are of importance in the underlying mechanical system [48].
- Since the dynamics has been modified by an unknown change of coordinates, only the quantities that are invariant under such transformation can be used to characterize the analyzed system.

The first aspect refers to perhaps the biggest advantage of this approach over all other methods of time series analysis. From the measurement of a one-dimensional time series, it is possible to infer information about the degrees of freedom other than the one directly observed. This method, however, also has its restrictions, which should be kept in mind, especially in those cases where a dynamical approach does not produce results that were expected. It should be also noted that the numerical implementation of the analysis in the reconstructed state space very often proves to be difficult. Therefore, one should not jump to an early conclusion that the dynamical approach is inapplicable only because a particular statistic failed one's expectations.

The second aspect pertains to a restriction which is vital if the results of performed analyses are to be meaningful. Fortunately, this restriction is not very difficult to satisfy. A significant number of *system invariants* is available and can be used to produce many powerful analytic tools.

Previous work

The characterization of irregular broadband signals, typical of nonlinear dynamical systems, and the extraction of useful information from such signals, has been a topic of extensive research over last fifteen years [2]. Much of the work in this field, known as dynamical time series analysis, has rested on powerful computational tools supported by some underlying rigorous mathematics. Most of the attention has been paid, however, to applications in

classical physics. There exist a relatively small number of references, [21, 24, 46, 35, 36], in which a possibility of chaotic behavior in some mechanical engineering devices (also gearboxes) is investigated. Unfortunately, in most of these contributions, the conclusions about the actual investigated devices were drawn from the mathematical models and not from the experimental data. One of the exceptions (and inspirations for this work) was a paper by Frison [18], in which he analyzed the vibration data from a spur-gear gearbox and concluded that it came from a chaotic system, whose embedding dimension was not smaller than 12. Despite the claim that the analysis presented there was just a prelude to gear fault prediction, further results have not yet been published.

In this contribution, we perform analyses using tools similar to those used by Frison but apply them to a one-step helical gearbox. We also go further than just to state our "beliefs" about the nature of the "underlying system." We present a model of system dynamics which is typical of nonlinear chaotic systems. We also propose a set of dynamical and probabilistic tools that can be applied to the analysis and condition monitoring of a class of gearbox signals.

Objective and overview

The main objective of this work is to present a combined dynamical and probabilistic methodology which can be applied to the analysis of a certain class of engineering problems. The main task of this contribution is to describe an alternative way to produce working models of the system dynamics in its state space. These models are used later for various purposes like forecasting, signal separation, and condition monitoring.

The main body of this work can be divided into two parts. The first part, composed of Chapters 2-4, discusses the dynamical approach to identification and verification of a general dynamical model of the gearbox system inferred from its one-dimensional vibration signal. The way in which the experimental time series was acquired from the gearbox is described in Chapter 2. In Chapter 3, we discuss methods of determining the embedding parameters of the reconstructed state space, which is similar to deciding the "model order" in system identification. Next we impose some structure on the proposed dynamical model in the state space. The notions of recurrent dynamics, periodic orbits, sensitivity of response to initial conditions, Lyapunov exponents, and probabilistic measures in the state space are introduced using the example of the gearbox data. In Chapter 4, the general

model of dynamics introduced thus far is described in more detail. The initial estimates of the embedding parameters, i.e., embedding dimension and time-delay, are refined using a measure of system predictability. Various types of prediction schemes are discussed and two nonlinear predictors are introduced.

The second part of this work discusses the applicability of the dynamical state space methodology to the problems of signal separation and system condition monitoring. Chapter 5 introduces three methods of signal separation which take advantage of the information provided in the reconstructed state space of the system. In Chapter 7, three new methods of condition monitoring are proposed and their performance investigated for the data obtained from the mathematical models introduced in Chapter 6.

Dynamical time series analysis is a young, vast, and vibrant field, which has already produced many important contributions to our understanding and treatment of complex signals. We are concerned, however, that further progress in this area is impossible without a significant input from engineering circles. It is our hope that this work, aimed mainly at these circles, will encourage engineers to experiment with the dynamical and probabilistic methodology presented here and to include it in their analytical toolboxes.

Chapter 2

Experiment

"Jacta est alea"

In order to investigate the applicability of a dynamical approach to the problem of gearbox condition monitoring, we conducted an experiment in which we tested a typical one-step helical gearbox with 49 teeth on the pinion and 109 teeth on the gear, giving a ratio 2.224. The gearbox was not modified for the purpose of the experiment and was tested under its typical operating conditions. The input shaft was driven by an electric motor at various speeds and under constant load provided by a water pump. A stroboscope was used to verify that no torsional oscillations were transmitted from the motor to the input shaft.

As shown in Fig. 2.1, we placed a measuring accelerometer on the top of the casing of the statically mounted gearbox and acquired its vibration signal at a sampling rate of 71 kHz. The transducer used was a piezoelectric accelerometer, *ENDEVCO 7701A–100,* with the amplitude response in the range of 1 Hz - 5 kHz and charge sensitivity of 100 pC/g. This signal was stored temporarily on a notebook PC with a data acquisition board and later analyzed on an IBM RS-600 workstation and a Pentium PC running Linux.

In order to observe the effect of vibration of the casing on the experiment, we identified its natural frequencies from impact-induced vibrations. As shown in Fig. 2.2, the dominant natural frequencies appear beyond 3 kHz. Since the range of the most stable performance of the controller of the driving motor was about 600 rpm (as measured on the output shaft), which corresponds to the meshing frequency of 490 Hz, the dominant natural frequencies of the casing were not excited (in our setup) by the meshing frequency or its dominant harmonics. This is important because we wanted to capture the dynamics of the meshing gear teeth rather than that of the more complex system including the casing. We managed to obtain a relatively weak contribution of the dominant modes of the casing to the

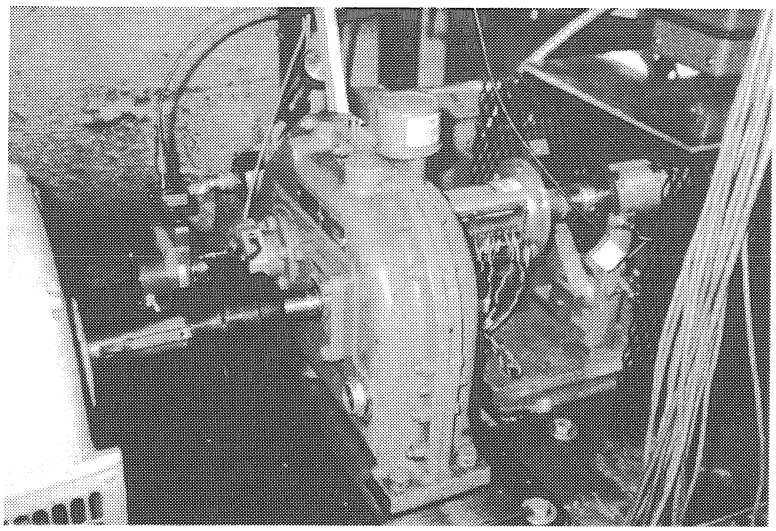


Figure 2.1: Photograph of the investigated one-step, helical gearbox.

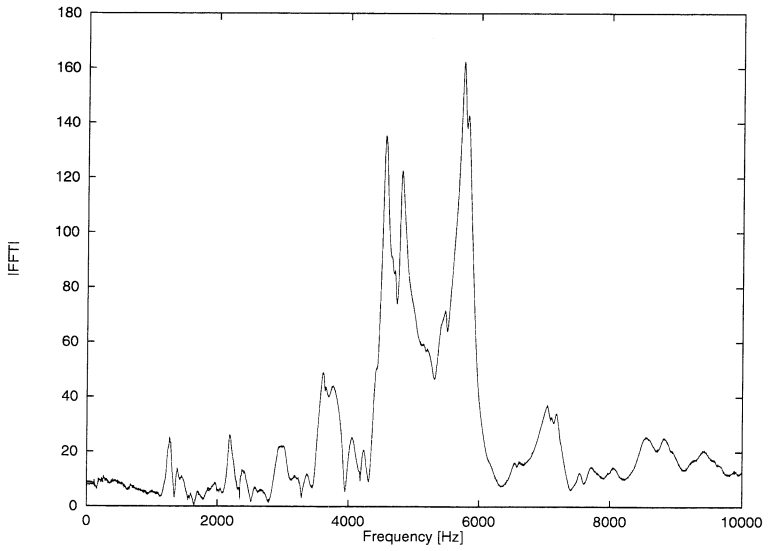


Figure 2.2: Natural frequencies as measured on the gearbox casing.

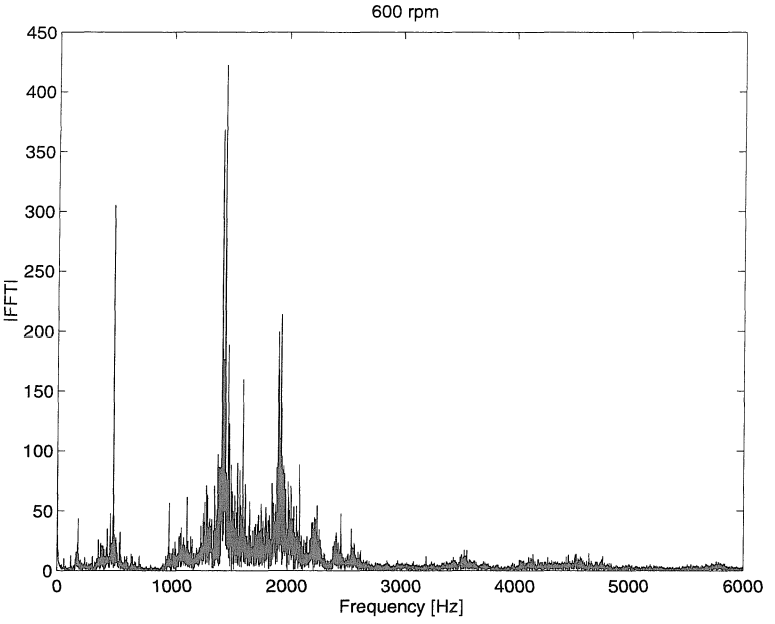


Figure 2.3: FFT computed for gearbox signal obtained at 600 rpm. The meshing frequency $f_m = 490$Hz.

acquired signal thanks to the appropriate location of the accelerometer. Even though, in principle, only one transducer is needed to recover all the active degrees of freedom of the analyzed system when the dynamical analysis is used, the quality of analysis depends on the positioning of the transducer.

An FFT of a typical gearbox vibration signal possesses certain characteristic features which can be observed in Fig. 2.3, created for the gearbox signal acquired in our experiment at 600 rpm. One can notice there the presence of dominant peaks corresponding to the meshing frequency and its harmonics. The other feature, which is of particular importance to us, is that these peaks are accompanied by modulation sidebands. In recent years, more and more researchers have expressed their beliefs that those sidebands contain more information about the state of the system than can be extracted from the linear discrete component[1]. In Chapter 6 we will suggest three models which have certain qualitative features resembling those exhibited by our experimental signal. One of the required features will be that the Fourier spectrum is qualitatively similar to that shown in Fig. 2.3.

[1]especially if the problem of condition monitoring is the main goal of analysis

Chapter 3

Nonlinear Analysis of Vibration Signature and Identification of Underlying Dynamics

"Res ipsa loquitur?"

Techniques of extracting information about the system from experimental data are commonly known in engineering as *system identification.* The equivalent methods in physical sciences are known under the name of *time series analysis* or, if applied to dynamical systems, as *dynamical time series analysis.* The technique which distinguishes dynamical analysis from others and forms its conceptual backbone is the phase space reconstruction. This method allows the experimenter to approximate the multi-dimensional, and directly unobservable, state of the physical system based only on scalar-valued measurements. The purpose of this reconstruction is not only to gain information about the number of active degrees of freedom of the system but also to look for its geometric features which can be used for noise reduction, forecasting, system condition monitoring and many others [2, 11]. Note that this concept is very general and in principle can also be applied to linear finite dimensional systems. However, in this case (and also for some weakly nonlinear systems) existing linear methods of analysis are usually very efficient and the phase space approach is often not worth its computationally intensive application. That is why dynamical time series analysis becomes a preferred approach only when the underlying system is strongly nonlinear, or high accuracy and/or sensitivity is required from the applied statistics. We will now explain what we mean by a dynamical system and try to clarify briefly the above mentioned concepts (for more details on theory of *embedology* see [48]).

3.1 State space embedding

In this work, by a dynamical system we understand a mechanical system that can be modeled by a following set of autonomous nonlinear differential equations:

$$\dot{u} = G(u) \tag{3.1}$$

with the initial condition $u(t_0) = u_0$, where $u = u(t) \in \mathbb{R}^d$ is a state vector and $G : U \to \mathbb{R}^d$ is a smooth function defined on some compact subset $U \subset \mathbb{R}^d$. Under assumption of a dissipative system, even if originally the system has a high (possibly infinite) dimensional phase space, its long term behavior will be eventually limited to some compact subset, U, of relatively low dimension. The trajectory of the differential equation (3.1) initiated at u_0 can be written as $u(t) = f^t(u_0)$. In order to make the physical process accessible to an observer, one needs to perform a measurement. Since the measurements are performed along an orbit f^t at discrete time intervals, $t = \tau_s$ (called later the sampling time or the time step), it is convenient to use a unit-time map to transform (3.1) into a discrete dynamical system synchronized with the observation process. Define a measurement function $g : \mathbb{R}^d \to \mathbb{R}$ that results in a scalar time series

$$y(k) = g(u(k\tau_s)) = g(f^{k\tau_s}(u_0)) = g(F^k(u_0)); \;\; y \in \mathbb{R}, \; k = 1, 2, \ldots, n \tag{3.2}$$

where F is the discrete dynamic map $F(u) \doteq f^{\tau_s}(u)$, $F^k(u) = f^{k\tau_s}(u)$. Hence one obtains the following discrete dynamical system

$$y(k) = g(F^k(u_0)) \tag{3.3}$$

describing a composite process which is not directly accessible to the experimenter. The scalar time series, y, is thus the result of an experiment. The classical system identification approach to non-parametric modeling would look for meaningful system features in the spectrum of y. If the vector field G is nonlinear, however, the spectral analysis will often fail to provide any meaningful information. This is because many nonlinear (in particular chaotic) systems exhibit a broadband spectrum and the spectral analysis is not sensitive enough to detect changes in the system behavior reliably. Dynamical analysis has proven

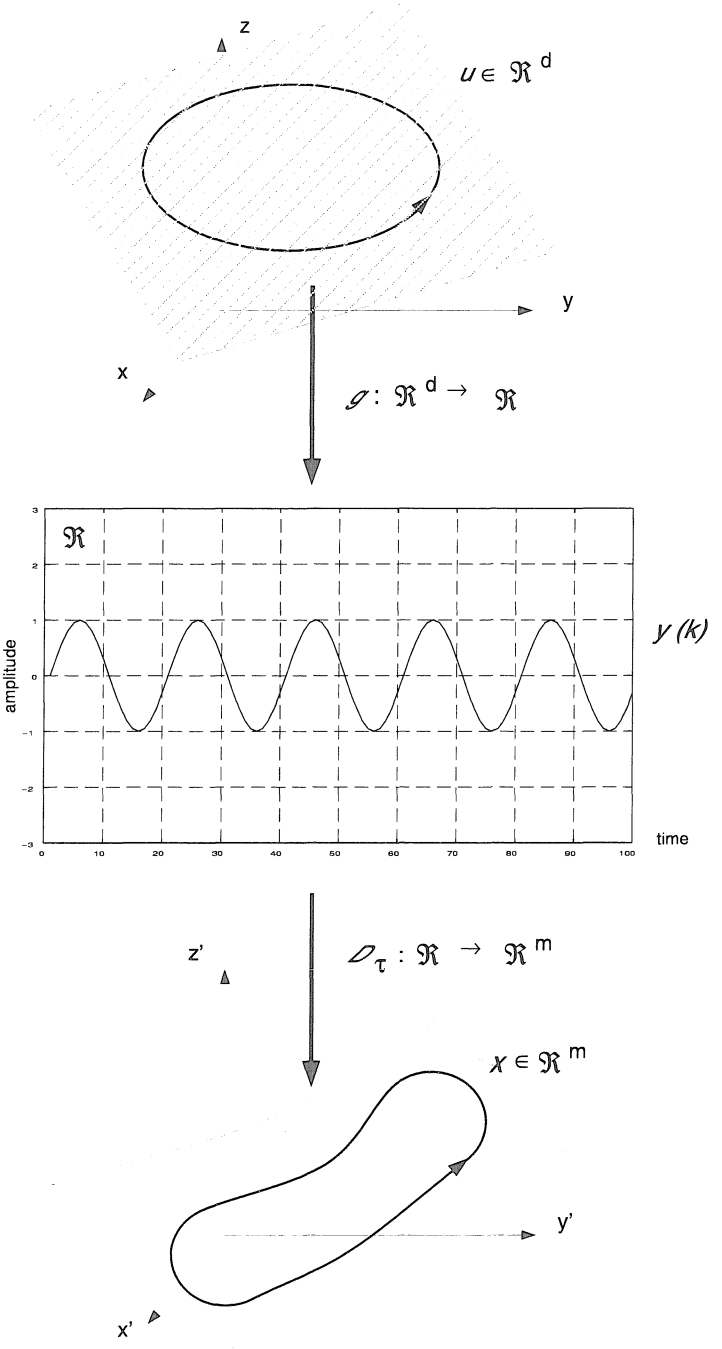


Figure 3.1: Embedding process showed for the example of a harmonic signal. In this example $d = 2$ is the dimension of the harmonic solution of the autonomous system of nonlinear differential equations (3.1), modeling the abstract physical system; $m = 3$ is the dimension of the *embedding model* described in (3.4) and (3.5).

to be particularly useful under these circumstances.

Time-delay embedding

The concept of the phase space reconstruction is based on Takens embedding theorem [50]. The theorem states that if one performs a "correct" time-delay embedding of the time series (3.3) one can construct an m-dimensional space which inherits many of the properties of the original d-dimensional space corresponding to the system (3.1) even though neither the dynamics of the original system nor the measurement function is known. For each state $u \in U$ of the system (3.1), one can define the m-dimensional vector

$$\begin{aligned} x(k) &= D_\tau(y) \\ &= (y(k), y(k+\tau), \ldots, y(k+(m-1)\tau)) \\ &= (g(F^k(u_0)), g(F^{k+\tau}(u_0)), \ldots, g(F^{k+(m-1)\tau}(u_0))) \end{aligned} \tag{3.4}$$

where $k = 1, 2, \ldots, n-(m-1)\tau$. This vector is called the *delay coordinate vector*, and τ is the *reconstruction time-delay*, which is an integer multiple of the sampling time τ_s. The dimension of the reconstructed phase space, m, needs to be chosen in such a way that the dynamical structure of the orbits in the original u-space is preserved in x-space, and the mapping from one space to the other is one-to-one. This can be achieved if m is greater than twice the "box-counting dimension" of U, and the measurement function g is chosen generically [48]. The process of mapping the time series into its embedding space is shown schematically in Fig. 3.1.

We can relate all the above to the general problem of modeling a mechanical input-output system:

$$\begin{aligned} x(k+1) &= \bar{f}(x(k)) + \Xi_k \\ y(k) &= \bar{g}(x(k)) + \varepsilon_k \end{aligned} \tag{3.5}$$

where $x \in \mathbb{R}^{d_E}$ is the state vector for the system, $\bar{f} : \mathbb{R}^{d_E} \to \mathbb{R}^{d_E}$ models the dynamics of the system, and $\bar{g} : \mathbb{R}^{d_E} \to \mathbb{R}$ models the relationship between the system state and output (in our case $\bar{g}$ is simply an approximation of the measurement function g in (3.2)). Ξ_k and ε_k correspond to the *embedding model* and measurement errors.

The process of reconstructing the embedding space and building local or global maps $x \to \hat{F}(x, a)$, which evolve every trajectory point $x(k) \mapsto x(k+1)$, is equivalent to defining a class of possible models of the system, known under the name of *specification* in system identification [5]. Using the information about how neighborhoods of trajectory points evolve in the state space, we can fit the parameters a of the chosen class of models. An example of this type of modeling would be using a variation of a local autoregressive model[1] in which instead of fitting the model parameters with sequences of points that are close in time, one uses neighborhoods of points that are close in the state space. This general geometric model of system dynamics will be described in more detail in Section 4.2.3 where it will be used to make predictions and to verify the applicability of the phase space approach to the analysis of the experimental gearbox signal.

The typical second step in system identification is *estimation* of the model parameters from available measurements or, in other words, determining the "best" model in the previously defined class of models. In dynamical time series analysis, this step corresponds to finding the appropriate embedding parameters, i.e., time delay and embedding dimension. The choice of the parameters of the predictive model, as defined above, will be postponed until Section 4.2. Before we proceed to estimate the embedding parameters, however, we will present two extensions of the simple time-delay embedding introduced above.

SVD embedding

The idea of filtering the delay coordinates was first introduced by Broomhead and King [9]. They suggested replacing raw delay coordinates with the dominant modes obtained from the singular value decomposition (SVD), which eliminates the need for an arbitrary choice of the time-delay parameter. The procedure starts with a *trajectory matrix*, X, whose rows are created from N trajectory points, $x \in \mathbb{R}^m$, initially embedded in a conventional fashion,

$$X = N^{-1/2}[x^T(1), x^T(2), \ldots, x^T(N)]^T \tag{3.6}$$

The dimension m is chosen high enough so that the Whitney embedding theorem is satisfied. The time-lag is chosen as $\tau = 1$. Next a real symmetric *covariance matrix*, $\Xi = X^T X$, is decomposed using the singular value decomposition into $\Xi C = C\Sigma^2$. Finally, the trajectory

[1] For a short discussion of linear prediction schemes, see Section 4.2.2.

matrix, X, is projected on the vectors of the orthogonal matrix C chosen so that they correspond to the singular values (entries of the diagonal matrix Σ) which are larger than an arbitrary noise level.

The main advantages of this method are that it eliminates the need for an arbitrary choice of the embedding parameters and that it provides initial filtering of the data. The main disadvantage is that the value of τ is not fixed across the trajectory, which complicates further analysis and forecasting of the data.

Filtered embedding

The idea behind this method is to merge the useful filtering techniques, known from conventional signal processing, with the geometrical approach of dynamical time series analysis [47]. Rather than using the conventional lag vectors $x(i) \in \mathbb{R}^d$, they are first multiplied by a "filtering" matrix, whose rank is larger than or equal to the desired embedding dimension. An interesting choice for that matrix is suggested by Sauer in [47], where it is produced by a composition of three linear operators: discrete Fourier transform, low-pass filter, and inverse Fourier transform. The vectors $x(i)$ are usually embedded in the space whose dimension is higher than necessary which allows to "intelligently" downsample the data. This type of filtered delay coordinate embedding, called by Sauer a "low-pass embedding," removes noise but preserves embedding of the reconstruction. In the absence of noise, there is no advantage of this approach over just downsampling the time series. There is, of course, a large number of possible linear filters that can be applied to this procedure [38].

3.1.1 Initial choice of time-delay

According to Takens embedding theorem, the quality of the phase space reconstruction does not depend on the choice of the time-lag τ. In practice, however, since all signals are contaminated with noise, selection of the "proper" time-delay becomes very important. For the first choice of the optimal time-delay, we use the method of the *average mutual information* [2]. The value of the average mutual information, $I(\tau)$, expressed in bits,

$$I(\tau) = \sum_{k=1}^{n-\tau} p(y(k), y(k+\tau)) \log_2 \left(\frac{p(y(k), y(k+\tau))}{p(y(k))p(y(k+\tau))} \right) \tag{3.7}$$

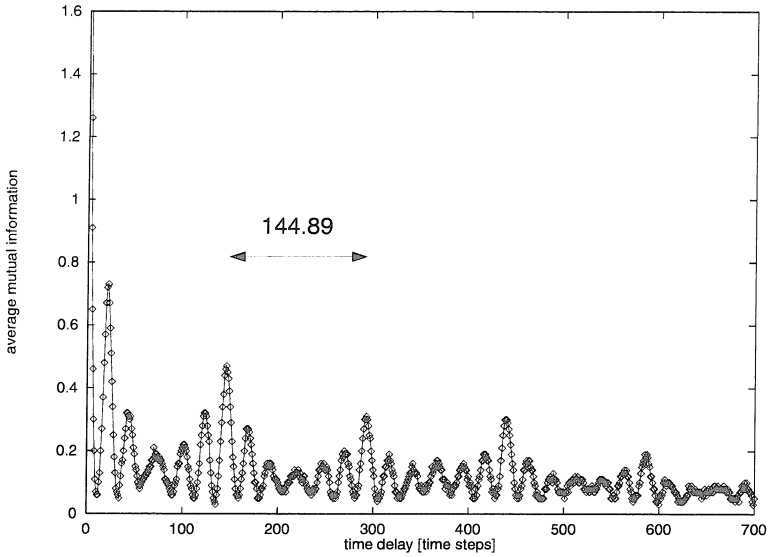


Figure 3.2: Average mutual information computed for a gearbox signal obtained at 600 rpm. The "optimal" value of time-lag was found to be $\tau = 11$.

is found for all possible pairs of measurements $y(k)$ and $y(k+\tau)$. $p(y(k)), p(y(k+\tau))$, and $p(y(k), y(k+\tau))$ are the probabilities of measuring $y(k)$, $y(k+\tau)$, and both $y(k)$ and $y(k+\tau)$, respectively. The probabilities are estimated as frequencies of occurrence as shown in [16]. We compute the values of $I(\tau)$ for some interval of τ. Then we plot the resulting points (I vs. τ) and choose the value of τ for which the average mutual information reaches its first minimum. This prescription for choosing τ reflects the case in which $y(k+\tau)$ is nearly independent, on average, from $y(k)$ ($I = 0$ if exactly statistically independent). Fig. 3.2 shows the mutual information function for the gearbox signal (compare with results for a spur pinion gearbox in [18]). Its first minimum occurs for the value of $\tau = 11$ which is temporarily chosen as the optimal time-delay and will be used as such until it is verified in the forthcoming sections. Another interesting feature visible in Fig. 3.2 is the presence of equidistant peaks in the average mutual information function. These local maxima are separated by the distance equal to the reciprocal of the meshing frequency (about 145 [*time steps*] at a sample interval of $\tau_s = 1.41 \cdot 10^{-5}$ sec).

3.1.2 Initial choice of embedding dimension

Both Takens theorem and Sauer prevalence theorem specify the lower bound on the value of the embedding dimension. Both theorems, however, establish only a sufficient condition for the correct embedding. In many cases it is possible to unfold the dynamics of (3.3) in a lower dimensional space. This means that one can find a d_E-dimensional space in which the trajectory will not intersect itself and whose dimension is lower than that given by both of the above mentioned theorems. This reduction in dimensionality leads to significant improvements in computation time for all numerical methods of analysis.

In our work we applied a method which uses the concept of *false nearest neighbors*[2], which are generated when the orbit is projected on a too low dimensional space. Two points are considered false neighbors if they are close to each other in $\mathbb{R}^m$ but not in $\mathbb{R}^{m+1}$ (for any m). To find false neighbors, one starts with a reasonably small m and examines the change in Euclidean distance for all pairs of nearest neighbors, as m is increased by 1. If for some pair this change is significant (we considered significant an increase in the initial distance by at least 15 times), those points are qualified as the false neighbors. One can plot the percentage of the false nearest neighbors with increasing dimension. The minimum embedding dimension, d_E, is chosen to correspond to the first point for which this percentage drops below the value of 1%. Fig. 3.3 shows the percentage of false nearest neighbors for various values of m used to embed the gearbox signal. From this graph one can conclude that the minimum embedding dimension can be as low as $d_E = 5$. As one can see, the number of false neighbors does not drop to zero. This may be caused by the noise contamination introduced by the motor controller and the measurement process. It is very common to observe such behavior of this statistic whenever the analyzed signal is contaminated with noise[3]. For a finite-dimensional deterministic system, the number of false nearest neighbors should become saturated as soon as the embedding dimension of the system is reached. However, if the analyzed data is contaminated with noise (which can be interpreted as a signal whose dimensionality is significantly higher than that of the signal of interest), then, after reaching an apparent plateau, the statistic starts to rise again. As it is demonstrated for the well known Lorenz system [32] in Fig. 3.4, the rate at which the curve drops or grows and the level of the plateau it reaches depend on the signal-to-noise ratio in

[2]For a comprehensive discussion of this method, see [2].

[3]For details about our interpretation and treatment of noise in dynamical analysis, refer to Chapter 5.

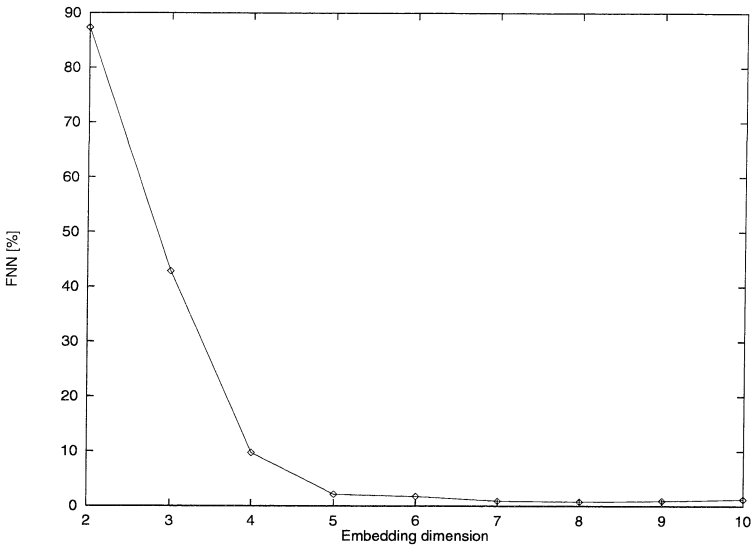


Figure 3.3: Embedding dimension estimated for gearbox data obtained at 600 rpm. The percentage of false nearest neighbors is computed for increasing embedding dimension, m. The lines connecting points have no physical meaning and are shown only to emphasize the trend.

the analyzed signal. The figure illustrates how the length of the plateau becomes shorter with the decreasing signal-to-noise ratio and how it finally disappears for the pure noise signal (uniformly distributed noise obtained from a pseudo-random number generator). It can also be seen how the noise contamination can be estimated with this statistic.

Note that the false nearest neighbor statistic was computed for the gearbox signal for the value of $\tau = 11$ estimated in the previous section. It is always a good idea to verify the stability of the embedding dimension estimation by computing the false nearest neighbors statistic for various values of time-delay. The result of such computation is shown in Fig. 3.5 from which it can be concluded that the obtained value of d_E is stable over a range of time-delays. We use $d_E = 5$ and $\tau = 11$ as the initial estimates of the embedding parameters until they are verified in Section 4.2.3.

3.2 Recurrent dynamics and periodic orbits

Characterizing time series through its periodic orbits is an old and well established procedure. It is commonly used to analyze linear systems, where peaks on the frequency spectrum imply system modes. In chaotic systems which result in broadband spectra it

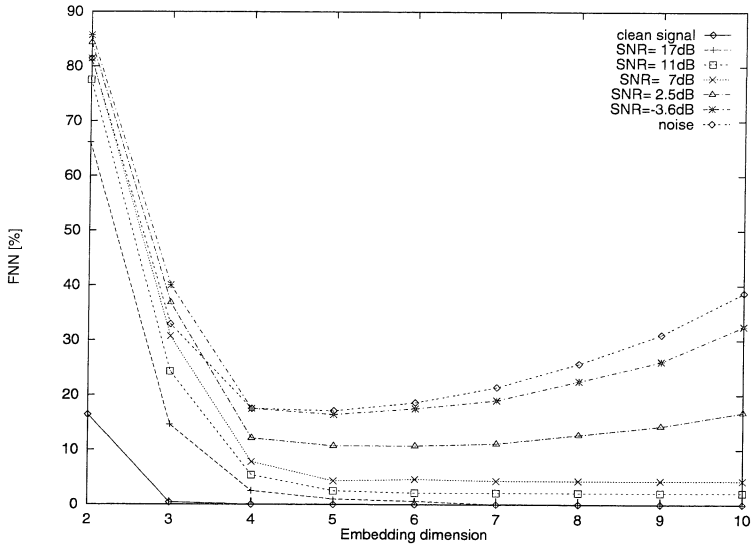


Figure 3.4: Percentage of false nearest neighbors for contaminated (to various degrees) signal from the Lorenz system.

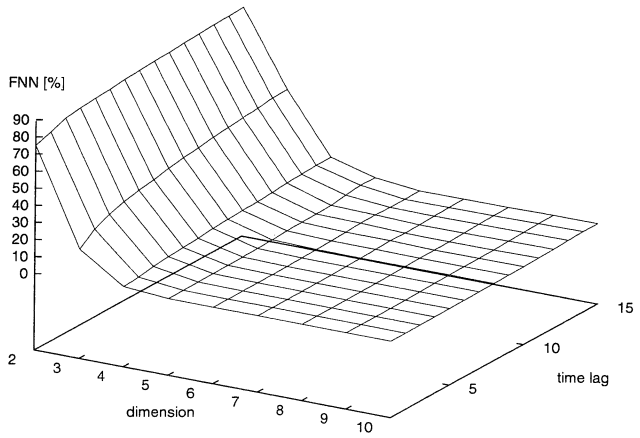


Figure 3.5: Embedding dimension estimated for gearbox data obtained at 600 rpm. The percentage of false nearest neighbors is computed for increasing embedding dimension, m, and time-lag, τ.

might be difficult to extract periodic orbits this way. Simply because clear peaks might not be present in the spectrum does not mean that periodic orbits are useless as a feature to characterize a nonlinear, chaotic, system. Poincaré was the first to point that out in 1899 [43]. An inverse is also true. As observed by Auerbach et al. [4], a chaotic trajectory can be successfully used to locate periodic orbits. Existence of the recurrent regions in the phase space allows one to define a global cross section of the flow and associate with it a return map, called a Poincaré map[4]. The dynamics of the Poincaré map is often easier to study than dynamics in the original phase space. It also reduces dimensionality of the problem from d_E to $d_E - 1$. For systems embedded in $\mathbb{R}^3$, it also provides an interesting graphical representation of the dynamics. Unfortunately, Poincaré maps are difficult to compute for experimental time series. Fortunately, there exist other, relatively new, methods which are as useful as Poincaré maps but much easier to implement.

3.2.1 Recurrence plots

In 1987, Eckmann et al. [14] introduced a very informative, graphical tool, called a *recurrence plot*[5], which can be easily applied to experimental data. Given a trajectory $\{x(i)\}_{i=1}^{N}$, let us define the following set of integers, $\mathbb{Z}_m = \{0, 1, 2, \ldots, m-1\}$, composed of indexes corresponding to the trajectory points, $\{x(0), x(1), x(2), \ldots, x(m-1)\}$. Consider the Euclidean distance, $\delta(i,j) : \mathbb{Z}_m \times \mathbb{Z}_n \to \mathbb{R}_+$, between two trajectory points $x(i), x(j) \in \mathbb{R}^{d_E}$

$$\delta(i,j) = \|x(i) - x(i+j)\| \tag{3.8}$$

where $m + n < N$. Now, we define a function $\rho : \mathbb{R} \to \mathbb{Z}_k$, as a linear function of δ with a constant k/ϵ, where k corresponds to the number of shades of grey chosen for the map, and ϵ is a suitably chosen cutoff value. In our work we used $k = 100$, i.e., $\rho(\delta(i,j)) = 0$ if $\delta(i,j) > \epsilon$, and $\rho(\delta(i,j)) = 1$ if $\delta(i,j) < \epsilon/100$, where 0 denotes white, and 1 black. The choice of the cutoff value, ϵ, is somewhat arbitrary. In our work, we used $\epsilon = \max ||x(i) - x(j)||$ for all possible values of $i \neq j$. Finally, a recurrence plot is defined as the graph of the composition [26]

$$\rho \circ \delta : \mathbb{Z}_m \times \mathbb{Z}_n \to \mathbb{Z}_k \tag{3.9}$$

[4]For a detailed discussion of Poincaré maps, see, e.g., [55].

[5]See also [26].

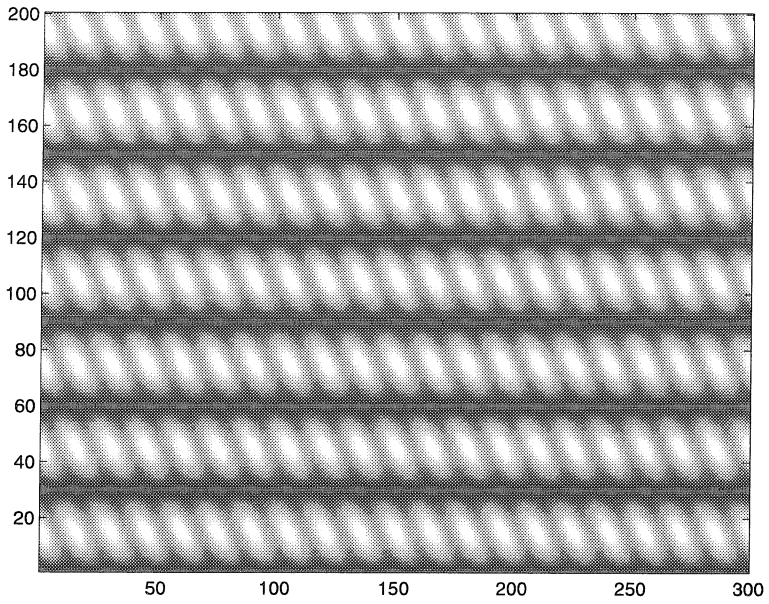


Figure 3.6: Recurrence plot for a harmonic signal embedded in $\mathbb{R}^2$ using time-lag value $\tau = 5$.

From the above definition, we can conclude that the $m \times n$ recurrence plot array is symmetric with respect to its diagonal $i = j$. Fig. 3.6 shows a recurrence plot produced for a harmonic signal, which was embedded in $\mathbb{R}^2$. There can be observed features of the recurrence plot which help identify certain types of system behavior [26]. A complete classification of line-type patterns within a recurrence plot can be summarized in seven possible values of line-segment slopes[6]:

- slope 0; these line segments correspond to trajectory points which become close and stay close for an interval of time. In particular, evenly spaced, horizontal line segments denote periodic recurrence (see Fig. 3.6.) Length of these segments quantifies the local rate of divergence.

- slope -2; denotes a segment of successive trajectory points which reappear in the state space reversed in time (see Fig. 3.6).

- slope -1 and vertical; corresponds to a particular trajectory point which remains close to an earlier (in time) trajectory segment.

[6]For a detailed description and classification of possible features in recurrence plots, see [26].

- slope smaller than -2; these line segments correspond to trajectory sections which reappear reversed and dilated in time.
- slope between -2 and -1; indicates sequences of trajectory points which reappear reversed and contracted in time.
- slope between -1 and 0; denotes trajectory segments which reappear contracted in time.
- positive slope; corresponds to trajectory segments which reappear dilated in time.

The most clearly visible feature is usually a collection of black horizontal line segments, which correspond to periodic orbits. To see this, denote the horizontal axis by i and vertical axis by j. The horizontal segment of length l can be expressed as the following set

$$\{(i,j),(i+1,j),\ldots,(i+l-1,j)\} \tag{3.10}$$

which, in turn, corresponds to the set of distances

$$\{\|x(i)-x(i+j)\|,\|x(i+1)-x(i+j+1)\|,\ldots,\|x(i+l-1)-x(i+j+l-1)\|\} \tag{3.11}$$

Thus, these line segments correspond to the sequences of the data which result in trajectory points coming close to one another and remaining close for an interval of time. Fig. 3.7, which shows a recurrence plot made for the gearbox signal embedded in $\mathbb{R}^5$, clearly indicates recurrent sequences of successive trajectory points. Note that some of the horizontal line segments are longer than others. The short ones correspond to the areas of the phase space with relatively high rate of local divergence. The longest horizontal segments are vertically repeated about every 145 time steps, which corresponds to the gear meshing frequency. Harmonics of the meshing frequency and other recurrences can also be identified from Fig. 3.7. Thus, recurrence plots can also be used to extract periodic orbits from experimental data or to construct such orbits in certain regions of the state space.

A different type of the recurrence plot, suited particularly well to identify periodic orbits (in this case of period $T = 145$), is shown in Fig. 3.8. It is constructed by plotting $\delta(i,T)$ vs. i for a fixed value of T. Fig. 3.8 illustrates how a periodic orbit can be identified in the

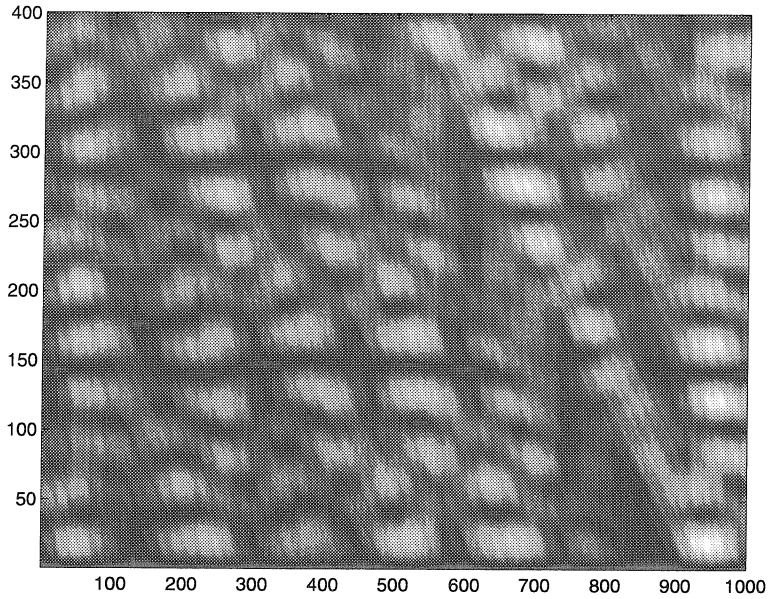


Figure 3.7: Recurrence plot for the gearbox signal embedded in $\mathbb{R}^5$ using time-lag value $\tau = 9$.

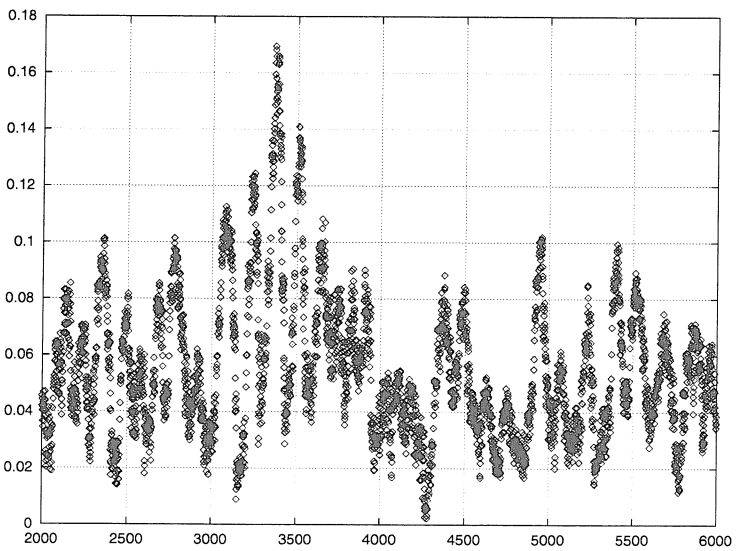


Figure 3.8: Recurrence plot $\|x(i) - x(i + 145)\|$ vs. i for the gearbox signal embedded in $\mathbb{R}^5$ using time-lag value $\tau = 9$.

vicinity of the 4270-th trajectory point. The period of the meshing frequency is also visible as a "pseudo-period" of the fluctuations in values of δ. Of course, this technique requires a separate plot for every value of T. Sometimes, for data obtained with a large number of samples per period, one can observe that some recurrent points[7] tend to become clustered, i.e., not only $x(i)$ is a recurrent point but also $x(i+1), x(i+2), \ldots$, etc. Such clusters signify the near periodicity over several complete periods in the corresponding trajectory segment.

Since various periodic orbits create a "topological skeleton" for strange attractors, one should be able (at least in principle) to recover the qualitative features of the underlying dynamics by superimposing extracted periodic orbits. The methodology which seeks to express averages over the regions of the state space in terms of short unstable periodic orbits is known as the *cycle expansions* [13]. If the dynamics is low-dimensional ($d_E \leq 3$), there exist techniques to extract a topological signature of the dynamical system by treating every periodic orbit as a knot, and quantifying the patterns in which they are interwoven [54]. Cycle expansions partition the trajectories into groups, where longer trajectories are approximated by the shorter ones. This leads to simpler representation of complex dynamics and to significantly more efficient computation. Fig. 3.9 shows an example of a periodic orbit decomposition for the Rössler system. A very deep analogy can be made between cycle expansions and the Riemann ζ-function from number theory: just as with the ζ-function, the prime numbers are used to create other numbers, in cycle expansions the "primitive orbits" are used to re-create other orbits. The main tool of this approach is the cycle expansion of the dynamical ζ-function [3]

$$1/\zeta = \prod_p (1 - t_p) \tag{3.12}$$

into p primitive cycles t_p. The main problem with this approach is that it is still unclear when, and if, the cycles suffice for the complete characterization of a dynamical system. Since this area has experienced very rapid progress in the last few years, it is a common belief that also this problem will be soon resolved. In Chapter 7, we put emphasis on practical applications of a small number of cycles (usually insufficient to reconstruct the

[7] A *recurrent point* is understood here as an approximation of a trajectory point belonging to a periodic orbit.

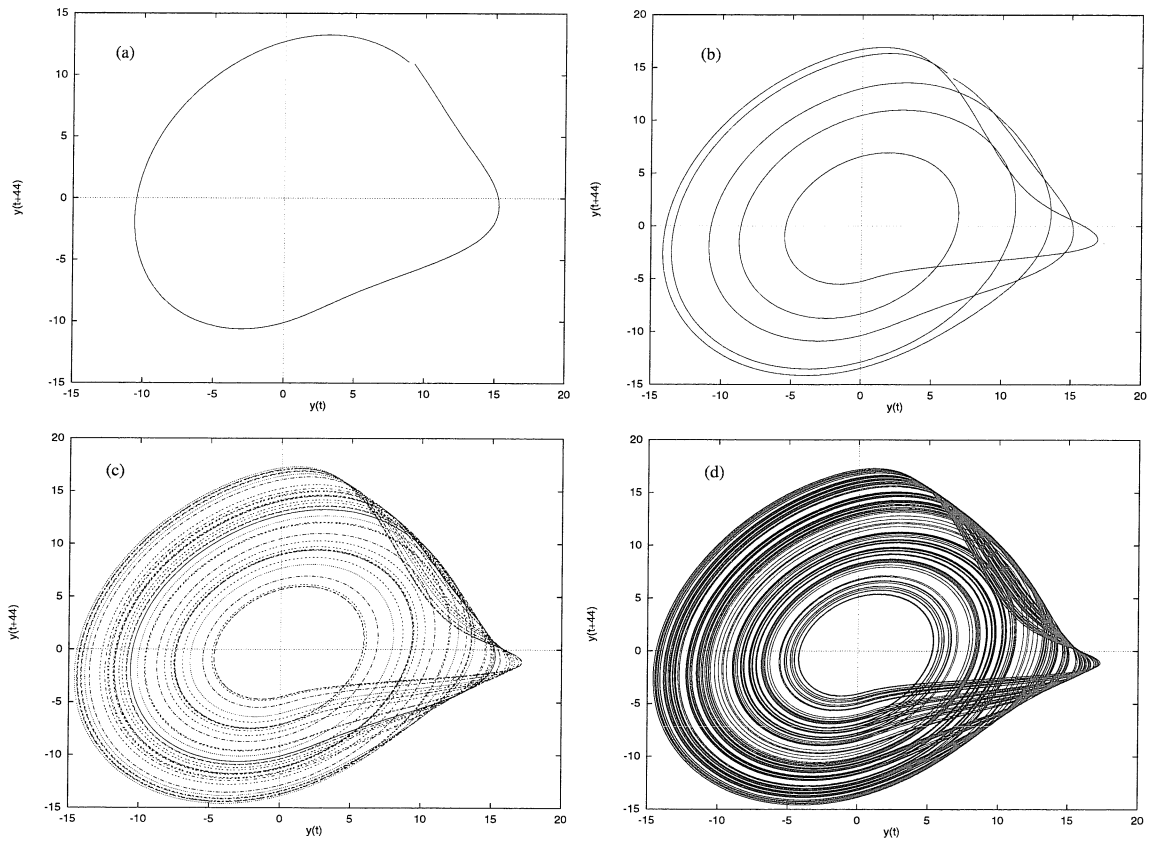


Figure 3.9: Periodic orbit decomposition of the Rössler attractor in its phase space. Geometry of the trajectory reconstructed from a time series of length 30,000 data points, (d), is well represented by a superposition of only eight of its periodic orbits, (c). (a) shows main periodic orbit, $T = 6.07s$ or 200 time steps, and (b), its harmonic at $T = 1000$ time steps.

system dynamics). That is why we shall refer those seeking more details and mathematical rigour in the treatment of cycle expansions to [4, 13, 3].

There is an interesting way in which unstable periodic orbits help to understand short term predictability in chaotic systems. One can achieve good short-term predictability for a segment of a chaotic trajectory in which it approaches an unstable periodic orbit along its stable manifold. Then the trajectory is ejected along the unstable manifold which results in a significant drop in its predictability. This behavior is responsible for a sudden diversion of the true and predicted signals shown in Fig. 4.7. The short-term predictability is again possible as soon as the trajectory is captured by the stable manifold of another periodic orbit.

3.2.2 Extraction of periodic orbits

As it was mentioned in the above section, periodic orbits (also called *cycles*) can be crudely located using recurrence plots. If there is interest, however, in using some quantitative measures which are defined for the periodic orbits, a more accurate method of orbit extraction is needed. Before we discuss the issue of extraction of periodic orbits, however, we shall briefly explain why these orbits are so interesting. The first reason for this is that they provide an invariant characterization of a deterministic dynamical system. The invariant character of a periodic orbit is reflected in the fact that it remains a periodic orbit under any smooth change of coordinates (in any representation). This is of paramount importance, because, as mentioned in Section 3.1, we cannot observe the system dynamics directly, but only through some measurement function, g, defined as in equation (3.2-3.3). Under the assumption of smoothness of g, the cycles are not only topologically invariant, but their stability is also invariant. In other words, eigenvalues of the Jacobian matrices computed for the system orbits are invariant, and vary slowly with smooth parameter changes, regardless of whether the time series is a long transient, a long cycle, or truly ergodic (for more details see [13]). Furthermore, only a finite, usually small number of periodic orbits are necessary for a good approximation of system dynamics. Short orbits contribute more to this "cycle expansion" than the long ones, and the error which comes from neglecting the long cycles can be bounded. The last statement is generally true for systems whose dynamics is restricted to a hyperbolic attracting set. It does not follow, however, that only such systems can be effectively approximated with periodic orbits. It just means that in a

nonhyperbolic case, the exponential decrease in importance of the long cycle contributions to the expansion needs to be verified. Note that this is very important for cases where one seeks to replace analysis of the experimental trajectory with analysis of its cycle expansion.

Since in this work we are only interested in certain measures defined for at most a few cycles, the exponential convergence of the cycle expansion is not necessary for successful application of our condition monitoring methodology. That is why we are particularly interested in two features of the representation by periodic orbits: first, short cycles can be accurately extracted from experimental data, and second, they are robust in terms of local dynamics in the neighborhood of the periodic orbits. The accurate extraction of periodic orbits is possible for relatively low dimensional systems, but only when the experimental time series is long enough to visit all neighborhoods of every periodic orbit of interest. Another important limitation on extraction of a cycle can be its stability. In general, the more unstable the orbit, the more difficult it is to observe and extract. Finally, contamination of data with noise[8] can become a serious limitation as well. All methods of extraction of periodic orbits require computation of distances in the phase space, and hence are susceptible to noise.

To summarize, the periodic orbits have a potential to become a very powerful tool in many areas of dynamical time series analysis. They can be effectively extracted from experimental data, although their accurate extraction can sometimes be difficult.

We will now describe how we extract periodic orbits from experimental data. Let us suppose that we are interested in finding a cycle of period T. We start with the trajectory $\{x(i)\}_{i=1}^{n} \in \mathbb{R}^d$ embedded in its phase space using the methods described in Section 3.1. We follow the time evolution $x(i+1), x(i+2), \ldots$, for every trajectory point $x(i)$, until we encounter such a point, $x(k)$, for which $\|x(k) - x(k+T)\| < \epsilon_1$. We call such a point a period-T cycle (or recurrent) point. All these points are then grouped into sequences representing periodic orbits. The recurrence threshold, ϵ_1, is chosen large enough to include several such sequences. In our work we found, that the value of $\epsilon_1 = 0.02\sigma$, where σ is the standard deviation of the time series, works well. For higher order, as well as for higher dimensional orbits, however, this value might need to be increased.

Notice that if a point $x(i)$ is a period-T cycle point, it is quite likely that $x(i+1)$ is also a

[8]Noise is understood throughout this work as a signal coming from a very high dimensional, possibly infinite dimensional, dynamical system.

period-T cycle point, as well as, possibly, $x(i+2), \ldots$, etc. All these points correspond to the same periodic orbit, as a result of the trajectory approaching the cycle from its attracting direction. One way of approximating a periodic orbit is to choose it to correspond to the point with the closest recurrence in the set. This will eliminate redundant cycle points from the interval $[x(i), x(i+T))$. There is another problem, however. How shall we classify the periodic point $x(i+T+1)$ or $x(i+T+2)$? There will also be other period-T cycle points in the neighborhood of $x(i)$. In order to decide whether two nearby nearly periodic orbits correspond to the same periodic orbit, the relative distances of all corresponding pairs of points of the two orbits need to be computed. Only if all such pairs of points are located within the distance ϵ_2, we conclude that they represent the same periodic orbit and put them into the same set. The position of a point representing the "best" estimation of the periodic orbit is chosen to be at the center of mass of all corresponding points in the above set. This procedure seems to be more robust, because it involves averaging in the phase space. The value of the second recurrence threshold, ϵ_2, was chosen to be $\epsilon_2 = 5\epsilon_1$, which seemed to work well for our data.

3.3 Lyapunov exponents and sensitivity of response to initial conditions

One of the most interesting features of the chaotic dynamics is its sensitive dependence on initial conditions. This phenomenon can be easily visualized by comparing the time evolution of two nearby passes of the trajectory in the phase space. To do that one needs to find a pair of nearest neighbors in the reconstructed phase space, called here $x(j), x^{NN}(j) \in \mathbb{R}^{d_E}$, that belong to two nearby trajectory segments. Then, while observing the time evolution of $x(j)$, one needs to update the search for its nearest neighbor (belonging to the same trajectory segment as $x^{NN}(j)$) to make sure that the evolution of distance between two trajectory segments is computed correctly. Sensitive dependence on initial conditions will be manifested by a rapid increase in the distance between initially close trajectories (averaged over the attractor). Such behavior can be seen in Fig. 3.10 created for the gearbox signal.

Another interesting way of identifying this phenomenon is to create a *scattergram*. Instead of computing a distance function, one can also project the time continuation of two initially nearby stretches of trajectory on a real line, thus creating two time series. Any

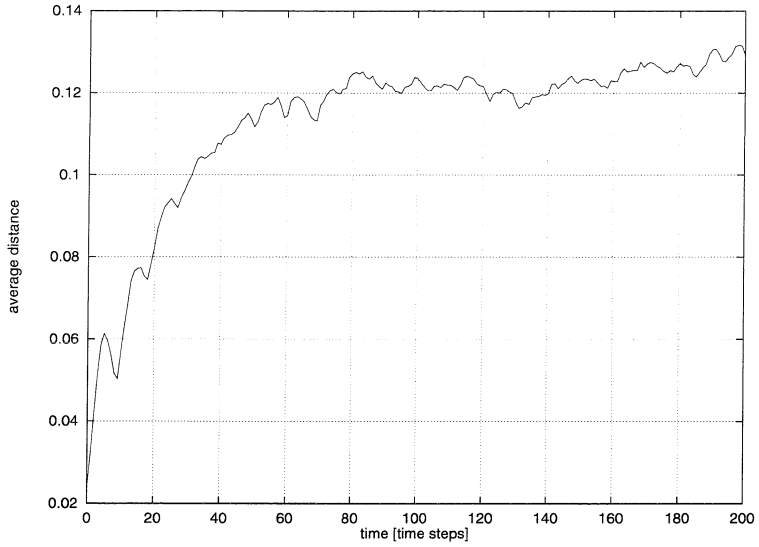


Figure 3.10: Time evolution of the average distance between two initially close trajectory points in the phase space. The distance was computed with respect to the estimated attractor size for the 600 rpm gearbox signal.

particular choice of projection should work. One could, for example, choose to project the trajectories on the direction corresponding to the largest singular value, or just take the first components of the two trajectories. The latter choice would result in two time series: $\{y(j), y(j+1), \dots\}$ and $\{y^{NN}(j), y^{NN}(j+1), \dots\}$, where $y(j)$ and $y^{NN}(j)$ denote the first coordinate of $x(j) \in \mathbb{R}^{d_E}$ and of the nearest neighbor to the point $x(j)$, respectively. The derived series can now be plotted, one against the other, which will result in the scattergram (see Fig. 3.11.) If the system is not sensitive to initial conditions, the two series will be highly correlated and the points on the scattergram will lie on a smeared line $y = \pi/4\, x$, as in Fig. 3.11(a). If the system exhibits sensitive dependence, however, one will notice a cloud of scattered points, as it is shown for the gearbox data in Fig. 3.11(b), resulting from two initially close segments of trajectory which rapidly diverged in the phase space.

Divergence of nearby orbits can also be quantified by Lyapunov exponents, which can be thought of as generalized eigenvalues averaged over the attractor. There are two commonly used definitions of the Lyapunov characteristic exponent (LCE). One definition describes the asymptotic behavior of the fundamental solution matrix, $X(t)$, of the analyzed dynamical

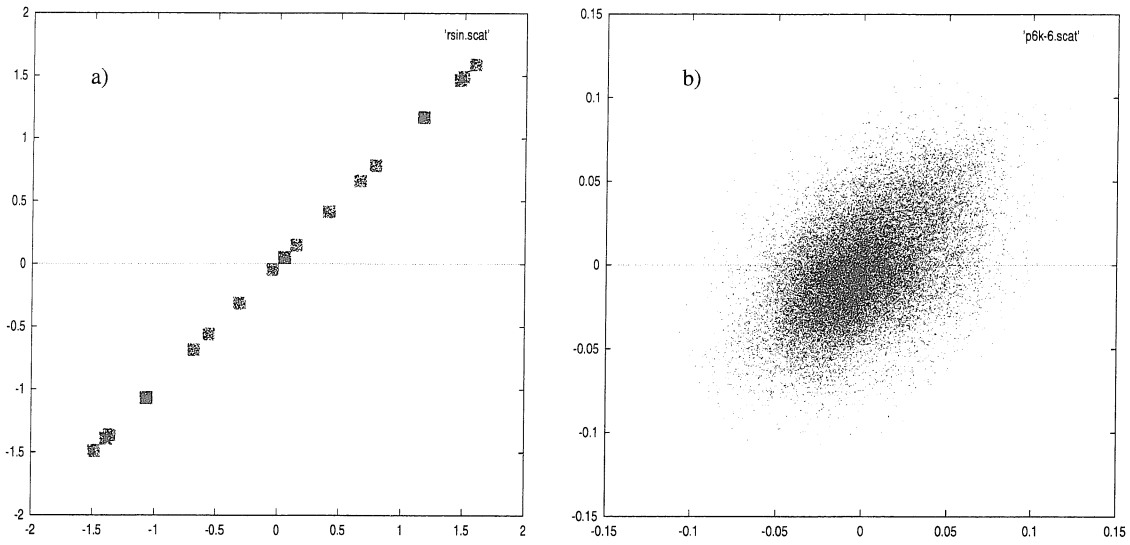


Figure 3.11: Scattergram plots for a) harmonic system response with added noise, and b) the 600 rpm gearbox signal. Horizontal axis corresponds to the time evolution of first components of points on one trajectory, and the vertical axis to the time evolution of the corresponding components of their nearest neighbors on the other trajectory.

system, in the direction specified by the vector w and as $t \to \infty$,

$$\lambda(x_0, w) \equiv \limsup_{t\to\infty} \frac{1}{t} \log_2 \frac{||X(t)w||}{||w||} \tag{3.13}$$

The other, better suited for computation, refers to the long-term evolution of an infinitesimal n-sphere (which becomes an n-ellipsoid due to the locally deforming flow in the phase space) of initial conditions,

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log_2 \frac{p_i(t)}{p_i(0)} \tag{3.14}$$

where $p_i(t)$ is the length of the principal axis of the n-ellipsoid at time t, and λ_i are numbered in descending order. In both formulas LCEs are expressed in *bits* per time unit.

Lyapunov exponents represent the average rate of divergence of initially close trajectory segments. Since the trajectory is embedded in the phase space, the number of Lyapunov exponents is equal to the dimension of the phase space. Negative exponents correspond to the directions along which two trajectory segments converge to each other and positive exponents to the directions along which they diverge exponentially. Every nontrivial, cor-

rectly embedded trajectory[9] should also possess a zero Lyapunov exponent. Since all of real world physical systems are dissipative and their dynamics confined to a globally finite region of the space, the exponential divergence has to be compensated by stretching and folding of the trajectory, and that is responsible for the Smale Horseshoe scenario leading to chaos. This is why in the dynamical time series analysis community, it has been agreed that the existence of a positive Lyapunov exponent, for a system whose trajectories are bounded to a finite attracting set, implies that the system dynamics is chaotic. There are many interesting theoretical and computational aspects involving Lyapunov exponents that are worth discussing (like the existence of the limits in equations 3.13 and 3.14 or how to compute all of the LCEs from a time series), but since our space is limited here, we will only refer the reader seeking more details to [2, 58] and references therein.

To verify the sensitive dependence on initial conditions and assess the predictability of the gearbox data, we computed a full set of global Lyapunov exponents. We obtained one zero, two positive, and two negative Lyapunov exponents. The largest exponent, $\lambda_1 = 0.0163$ $[bits/time\,step]$, quantifies the approximate average maximum length of a reasonable prediction horizon. In our case this length is about 1963 time steps, which means that after that many steps, prediction essentially looses all 32 bits of information. Note, however, that this is just an average number. Depending on the location of the predicted trajectory segment in the phase space, the resulting predictability can be significantly higher or lower.

This result can also be verified by fitting $y = 2^{\alpha x+\beta}$ to the increasing part of the curve in Fig. 3.10, and the value of α obtained this way can be considered a crude estimate of the largest Lyapunov exponent. In our case we obtained $\alpha = 0.0169$ $[bits/time\,step]$, which is in good agreement with the previously estimated value of λ_1, and hence confirms our conclusions about the system's sensitive dependence on initial conditions.

3.4 Kolmogorov-Smirnov test

We have already shown that the gearbox signal exhibits features that justify application of dynamical analysis. On the other hand, it is possible that a noisy signal which comes from a hypothetical linear system might also exhibit similar features. In recent years there has been considerable attention devoted to the problem of determining whether the nature of the

[9] By a nontrivial trajectory we understand an orbit which does not contain a fixed point.

"underlying system" is chaotic or "random" (e.g., [52, 25]). Since, in our understanding, the randomness is a measure of one's lack of knowledge about the system rather than a definition of an abstract physical system, the meaningfulness of such tests is somewhat questionable from this point of view. Most of the suggested algorithms, from which one is supposed to infer presence of nonlinear structure in time series, are based on statistical hypothesis testing. An alternative model (null hypothesis), that is used for that purpose, is usually defined in the form of artificially created (*surrogate*) data. Unfortunately, all of these methods suffer not only from problems of "philosophical" nature (typical of statistical hypothesis testing) but also from their limited reliability. Even if one believes that the underlying physical system, modeled by (3.1), can be reliably characterized that way, few would probably believe that its nature is indeed linear. Since there seems to be a strong tendency, within the community of dynamical time series analysis, to perform these tests anyway, we use here the Kolmogorov-Smirnov test with forecasting error as a discriminating statistic. This choice of statistic seems to have the highest relevance in our case.

The null hypothesis that we use states that the data comes from a linear "stochastic process" and there is no nonlinearity either in the dynamics or its observation. The surrogate data is created from the original time series by randomizing its phases but preserving its spectrum [52]. To do that we first need to compute the Discrete Fourier Transform, $Y(f)$, of the time series $y(k)$ and then randomize its phases $Y^j(f) = Y(f)e^{i\phi(f)}$, where $\phi(f)$ is uniformly distributed in $[0, 2\pi]$. Note that the created j-th surrogate data set has the same spectrum as the original time series and hence complies with our null hypothesis. For this analysis to be statistically meaningful, we need to create an ensemble of 10–100 surrogate data sets. Before inverting $Y(f)$, we also need to ensure that the resulting time series is purely real. This can be done by requiring the following equality of the imaginary components to hold: $\Im(Y^j(f)) = -\Im(Y^j(n-f))$. Now the actual testing can be applied to the set containing the original and surrogate time series. The prediction statistic we use is specifically designed to emphasize the deterministic character of the data. A prediction algorithm similar in spirit but a little simpler than the one described in Section 4.2.3 is employed. To eliminate the artificial increase in predictability that results from short-time correlations, a "decorrelation interval" of 50 time steps is used [25]. We compute the prediction 10 steps ahead in the reconstructed phase space for all data points whose true continuation we know. Since the Kolmogorov-Smirnov test requires the data samples to

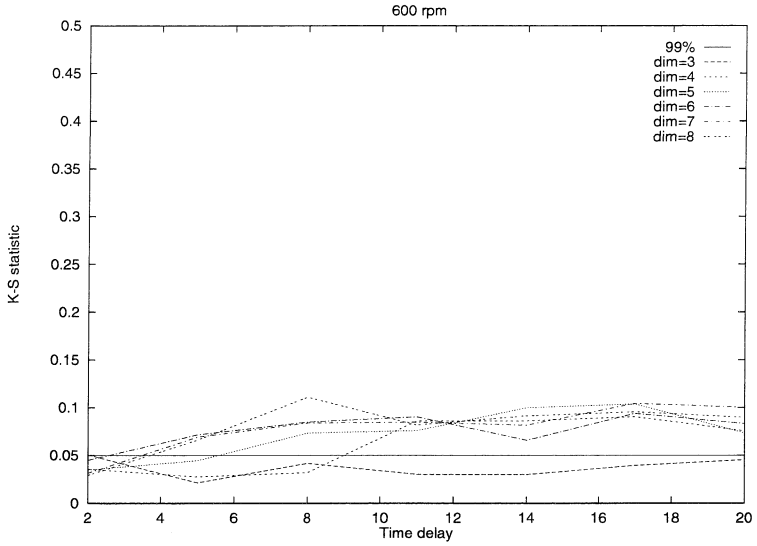


Figure 3.12: Kolmogorov-Smirnov test computed for gearbox data, 600 rpm, prediction horizon $T = 10$ sampling intervals. The 99% (solid) line corresponds to the 99% confidence level of the rejection of the hypothesis that the signal is a linearly correlated noise.

be independent, we decimate the set of prediction errors by a factor corresponding to the autocorrelation time of the predictor errors [25]. Next we save the results in two files, one corresponding to the original time series and one to all of the surrogate data sets. Now we can compute the Kolmogorov-Smirnov test and repeat the whole process for various values of the embedding parameters τ and d_E. The result of this procedure is shown in Fig. 3.12. It indicates superior predictability of the original time series (compared with the surrogate series) over a range of time-delays, provided the embedding dimension is higher than 4. Rejection of the null hypothesis implies that the dynamical features of the prediction model are unlikely to correspond to the surrogate data. Nonetheless, one still cannot assert from this test the evidence for a low-dimensional chaotic behavior.

3.5 Statistical measures of dynamics

The purpose of defining statistical measures in the system phase space is to identify quantities that are invariant under changes in initial conditions. These measures are particularly informative whenever a system state is altered by unknown perturbations and/or it is chaotic. In the latter case, different orbits differ almost everywhere in the system phase

space. Therefore, in order to identify the system that produced the observation, one has to implement measures which produce the same results for all, sufficiently long, observations of the same system.

3.5.1 Probability density function and its estimation

The *probability density function* (later also referred to as *pdf*) is one of the fundamental concepts in probability theory and statistics. It has proved to be very powerful in analyzing complex systems where it is used both as an independent tool and as an intermediate step in computation of its functionals, e.g., mutual information, characteristic function, expected value and other moments, etc. Given a data set of measurement results, $\{x_i \in \mathbb{R}^d\}$, $i = 1, \ldots, N$, its density function, $p(x)$, provides a natural description of the data distribution in $\mathbb{R}^d$, and is usually defined with the following relation

$$P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} p(x)dx \tag{3.15}$$

which quantifies the probability of finding an observation x inside an interval $(x_1, x_2]$, for any $x_1 < x_2$.

The purpose of density estimation is to construct an estimate, $\hat{p}$, of the density function, p, from the available observations. There are many known methods of density estimation (for more details about density estimation see Silverman [49].) The most commonly used density estimator is the *histogram*. Since it is the oldest known density estimator, all other such estimators are usually tested by comparing their performance to that of the histogram. The main drawback of the histogram is that its quality of estimation depends not only on a choice of the discretization bin width (which controls the amount of smoothing) but also on a choice of an origin and, in multi-dimensional spaces, the coordinate directions of the grid of cells. Moreover, the discontinuity of histograms make them inapplicable to cases where derivatives of the density function are required.

A different approach to density estimation is derived from an "empirical" model of the probability density function

$$p(x) = \sum_{i=1}^{N} p_i \delta(x - x_i) \tag{3.16}$$

where for every trajectory point x_i there is associated a probability p_i, and the Dirac delta function is applied in $\mathbb{R}^d$. In this work we assign equal probability to all N trajectory points; therefore, the p_i symbol will be replaced by the value $1/N$ in the forthcoming formulas. This definition is very useful if a smooth function needs to be integrated over the region weighted by the density function. Integration can then be replaced by a summation. In this case, however, eq. (3.16) is not very useful for *pdf* estimation at discrete points, because it results in infinite values at the observations and zero values everywhere else. This problem can be solved by replacing the delta function with a weight function which has a non-zero width.

In the *naive estimator* method, the estimate is constructed by placing a "box" of width $2h$ and height $(2Nh)^{-1}$ over each observable and then summing over the entire data set. The estimator is defined as

$$\hat{p}(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}w(\frac{x-x_i}{h}) \tag{3.17}$$

where the weight function, $w(x) = 1/2$ if $|x| < 1$ and $w(x) = 0$ otherwise. The main advantage of this method, over the histogram, is that it frees the estimation from a particular choice of bin positions. The estimated *pdf*, however, is still not continuous.

There is another parametric method of density estimation which is also often used in data analysis. In this method, one assumes an underlying model for the density function, e.g., a normal *pdf*, and then estimates the model parameters, e.g., mean and variance.

In the method we discuss in more detail below, we consider the generalization of the naive estimator, known as the *kernel estimator*. This method overcomes some of the difficulties associated with both histogram and naive estimators. The density computed with the kernel estimator is continuous and it also allows easy implementation of adaptive kernel width size. It is important, because fixed-bin methods often fail to approximate accurately highly non-uniform distributions, and the density functions obtained from analysis of chaotic data are often found to be discontinuous, multimodal, and nonuniform [23], as it is shown in Fig. 3.13 (these *pdf*s are discussed later). Even if the adaptive scheme is used to vary bin sizes across the attractor, the result can still be significantly improved (particularly in higher dimensions) when using kernel estimators.

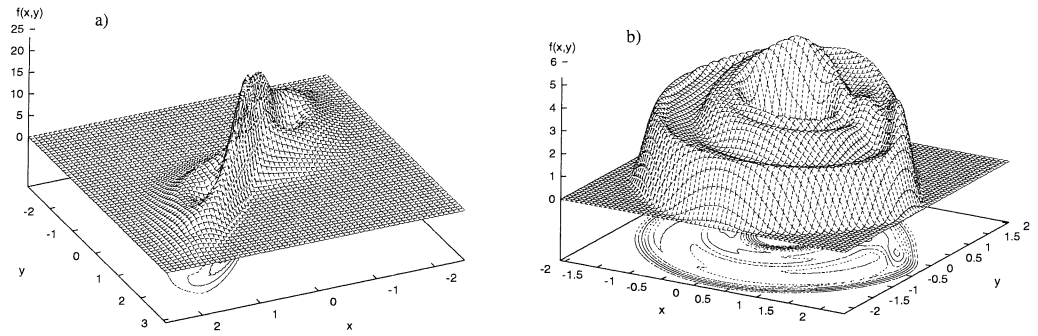


Figure 3.13: Kernel estimation of multimodal probability density function for: a) Lorenz system, and b) Rossler system. Note that since both of the mentioned systems are embedded in $\mathbb{R}^3$, the above figures represent two-dimensional projections of their respective three-dimensional density functions.

Kernel density estimation

For convenience, we will normalize the trajectory, for which the density function needs to be estimated, to have zero mean and unit RMS norm (or unit standard deviation, for one-dimensional systems.) The general multivariate kernel density estimator with kernel K and window width h is defined by

$$\hat{p}(x) = \frac{1}{Nh^d} \sum_{i=1}^{N} K(\frac{x - x_i}{h}) \tag{3.18}$$

The kernel function $K(x)$ is a function defined for any $x \in \mathbb{R}^d$, such that

$$\int_{\mathbb{R}^d} K(x)dx = 1 \tag{3.19}$$

and is usually chosen to be a radially symmetric unimodal *pdf*.

Kernel density estimation approximates a multi-dimensional *pdf* surface by superposing the kernel functions centered at every observed data point, as it is schematically shown in Fig. 3.14. The important issue now is the choice of a kernel function, K, and a window width, h.

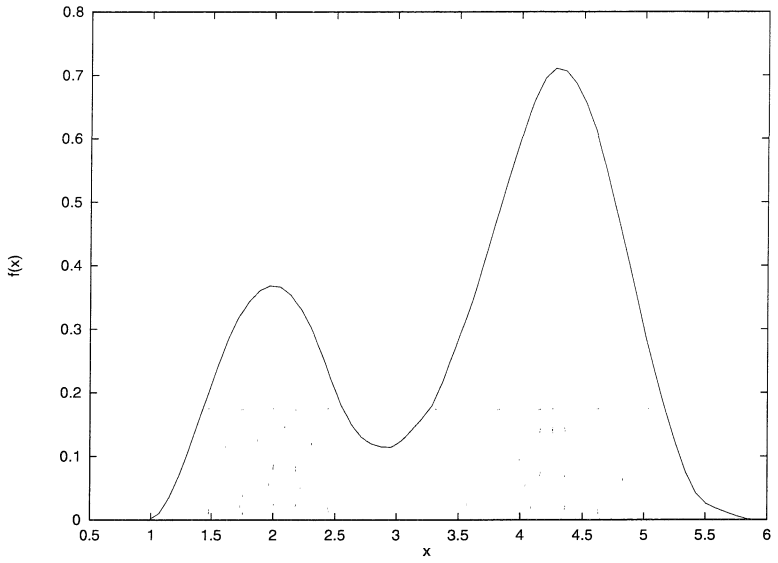


Figure 3.14: Kernel estimation of a one-dimensional bimodal density function. Estimate is computed by superposing the kernels placed on every observable. Height of the individual kernels, shown here to illustrate the principle of estimation, is exaggerated.

Choice of kernel

It turns out that the choice of the *shape* of the kernel function does not significantly effect the quality of the approximation (for more details see [49]). Only two of the most popular kernels will be shortly discussed here. One of them is *the normal kernel*, K_σ, which is just the standard multivariate normal density function, $\mathcal{N}(0,1)$,

$$K_\sigma(x) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}x^T x\right) \tag{3.20}$$

This kernel is important obviously because of the role the normal distribution plays in statistical analysis. Another possible kernel is the multivariate *Epanechnikov kernel*

$$K_e(x) = \begin{cases} \frac{d+2}{2V_d}(1 - x^T x) & \text{if } x^T x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.21}$$

where V_d is the volume of the unit d-dimensional sphere

$$V_d = \frac{\sqrt{\pi^d}}{\Gamma(\frac{d}{2}+1)} \tag{3.22}$$

e.g., $V_1 = 2$, $V_2 = \pi$, $V_3 = 4\pi/3$, etc., and Γ is the gamma function, $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$. This kernel has some major advantages over the normal kernel. First, it minimizes the "mean integrated square error" [49] of the approximation over the set of possible kernels (which is nice, but as it was already mentioned, not that important.) Secondly, it has a finite support, which turns out to be very useful for correct approximation of the tails of the distribution in higher dimensional spaces. Finally, the Epanechnikov kernel is significantly faster to compute.

Choice of smoothing parameter

The choice of the window width is as important to the problem of density estimation as it is difficult and there are still no rigorous methods of its selection. There are a few well tested rules-of-thumb, though. One of them suggests choosing a value of h which minimizes the "mean integrated square error" of the approximation. Unfortunately, the expression thus obtained contains the value of the *pdf* which one is trying to estimate. This problem can be resolved only by substitution of some *reference* density function to the formula. The most common choice for the reference *pdf* is the normal density function. Eq. (3.23) shows the optimal value of the window width evaluated with respect to the normal distribution

$$h_{opt} = A(K)/N^{1/(d+4)} \tag{3.23}$$

where $A(K)$ is a constant depending on type of the used kernel. For the Epanechnikov kernel

$$A(K_e) = (\frac{8}{V_d}(d+4)(2\sqrt{\pi})^d)^{1/(d+4)} \tag{3.24}$$

e.g., for $d = 2, A(K_e) = 2.40$; $d = 3, A(K_e) = 2.49$; etc. Because h_{opt} refers to normal distribution, one should be careful using the formula (3.23) for an unknown distribution. In most cases, it will be appropriate to use a smaller value, say σh_{opt}, for the window width. A possible value for σ might be

$$\sigma = \sqrt{\sum_i s_{ii}/d} \tag{3.25}$$

where s_{ij} are elements of the sample covariance matrix of the data which makes σ^2 the average marginal variance. For the purpose of analysis of chaotic distributions, however, it is safer to use about (0.05 to 0.5) of the value of h_{opt} given by equation (3.23).

In conclusion, the choice of the optimal value of the smoothing parameter is rather arbitrary and its effectiveness depends on the experience of a person performing the analysis. Fortunately, thanks to some of the adaptive methods, there are ways to make this process more automatic and efficient.

Adaptive kernel estimates

The adaptive kernel estimator is a modification of the standard kernel estimator which allows the smoothing parameter to vary at every center of the kernel function. The window width is chosen to be narrower in highly populated regions of the attractor, so as to capture the local character of the *pdf*, and wider in less populated regions to avoid undersmoothing. This approach is similar in spirit to the *nearest neighbor method*, but turns out to be much more accurate. The adaptive kernel approach applies a three-stage procedure. In the first step, an initial estimate of the *pdf* is obtained using one of the standard kernel methods. Estimated in this way, the density, called a *pilot estimate*, $\tilde{p}$, is then used to evaluate a set of *local bandwidth factors*, λ_i, for every observation. These factors are then used, in the third step, to construct the final estimator $\hat{p}$. This procedure can be outlined as follows:

Step 1 Compute a *pilot estimate* $\tilde{p}$ which satisfies $\tilde{p}(x_i) > 0$ for every observation x_i.

Step 2 Define *local bandwidth factors* λ_i by

$$\lambda_i = \sqrt{\frac{g}{\tilde{p}(x_i)}} \tag{3.26}$$

where g is the geometric mean of the $\tilde{p}(x_i)$

$$\log g = \frac{1}{N}\sum_{i=1}^{N} \log \tilde{p}(x_i) \tag{3.27}$$

Step 3 Define the *adaptive kernel estimate* $\hat{p}$ by

$$\hat{p}(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{K(\frac{x-x_i}{h\lambda_i})}{(h\lambda_i)^d} \tag{3.28}$$

Table 3.1: Sample size required to estimate a standard multivariate normal density function using a normal kernel.

Dimension	Required sample size
1	4
2	19
3	67
4	223
5	768
6	2,790
7	10,700
8	43,700
9	187,000
10	842,000

where K is the kernel function and h is the window width used to compute the pilot estimate.

This approach seems to be a very successful combination of features of the kernel and nearest neighbor methods.

Some computational considerations

Note that equations (3.18) and (3.26)-(3.28) require the kernel function to be evaluated N^2 times because $\hat{p}(x)$ must be evaluated at each x_i. There is a way, however, to significantly decrease this number. The use of a finite support kernel reduces the number of summed elements in (3.18) from N to N_b, where x_{N_b} is the most distant neighbor of x satisfying $(\frac{x-x_i}{h})^T(\frac{x-x_i}{h}) < 1$. Therefore, for a fixed width kernel, one needs to consider only these neighbors that are located within the circle of radius h from the point x. The most efficient way to do that is to construct a d-dimensional binary tree [17], in $\mathbb{R}^d$, which contains all elements of $\{x_i\}$. Search for nearest neighbors, using a d-dimensional tree, requires then only $\mathcal{O}(N \log N)$ operations and seems to be the most efficient choice for this purpose.

Another factor one has to consider is the sample size N required to ensure the assumed accuracy of estimation. It turns out that this size becomes very large in high dimensional spaces. To illustrate that, let us consider the case where the true density p is the unit multivariate normal *pdf* and that we use a normal kernel for the estimation. In Table 3.1 (taken from [49]) the displayed values of sample size N were computed so as to minimize the relative mean square error of the approximation at $x = 0$. This is obviously just an

example, but this kind of rapid increase of required sample size with dimension should be expected for any form of density function estimation. This observation reemphasizes the importance of dimension estimation before statistical analysis can be undertaken.

Fig. 3.15 illustrates the convergence and the accuracy of the kernel estimation of the stationary *pdf* computed for the nonlinear Duffing oscillator excited with Gaussian noise,

$$\ddot{x} + 2\zeta\omega\dot{x} - \omega^2 x + \epsilon x^3 = \sqrt{2D}n(t) \tag{3.29}$$

with parameter values: $\omega = 1$, $\zeta = 0.25$, $\epsilon = 0.5$, $D = 1$, and $n(t)$ denoting Gaussian noise. The two-dimensional stationary *pdf* for the system (3.29) was computed analytically using the Fokker-Planck equation [10] and estimated with the Epanechnikov kernel from the samples of x and $\dot{x}$ obtained by numerical integration. The figure shows one-dimensional slices of the two-dimensional stationary *pdf* obtained for $\dot{x} = 0$. From the comparison between the estimated and the exact values of the *pdf* in Fig. 3.15(d) it can be concluded that a sample of at least $10,000$ data points should be used if stationary measures of the dynamics are to be observed. A good quality of the *pdf* estimation is also evident from the figure.

3.5.2 Characteristic function

The *characteristic function* for a probability distribution of a delayed vector $x \in \mathbb{R}^d$ is defined as [41]:

$$\phi(k) = \int p(x) e^{ikx(t)} dx \tag{3.30}$$

$\phi(k)$ can be interpreted as the Fourier transform of the density function $p(x)$. If the Laplace transform is used instead, $\phi(k)$ is called the *moment generating function* of x. Since $p(x) \geq 0$, the modulus of $\phi(k)$ has its maximum at the origin. If the characteristic function is known, each moment can be calculated by evaluating the corresponding coefficient in the Taylor expansion of $\phi(k)$.

This statistic has another useful property: the characteristic function for a sum of two independent signals is equal to the product of the separate characteristic functions. This factorization property of the characteristic function makes it a useful tool for signal

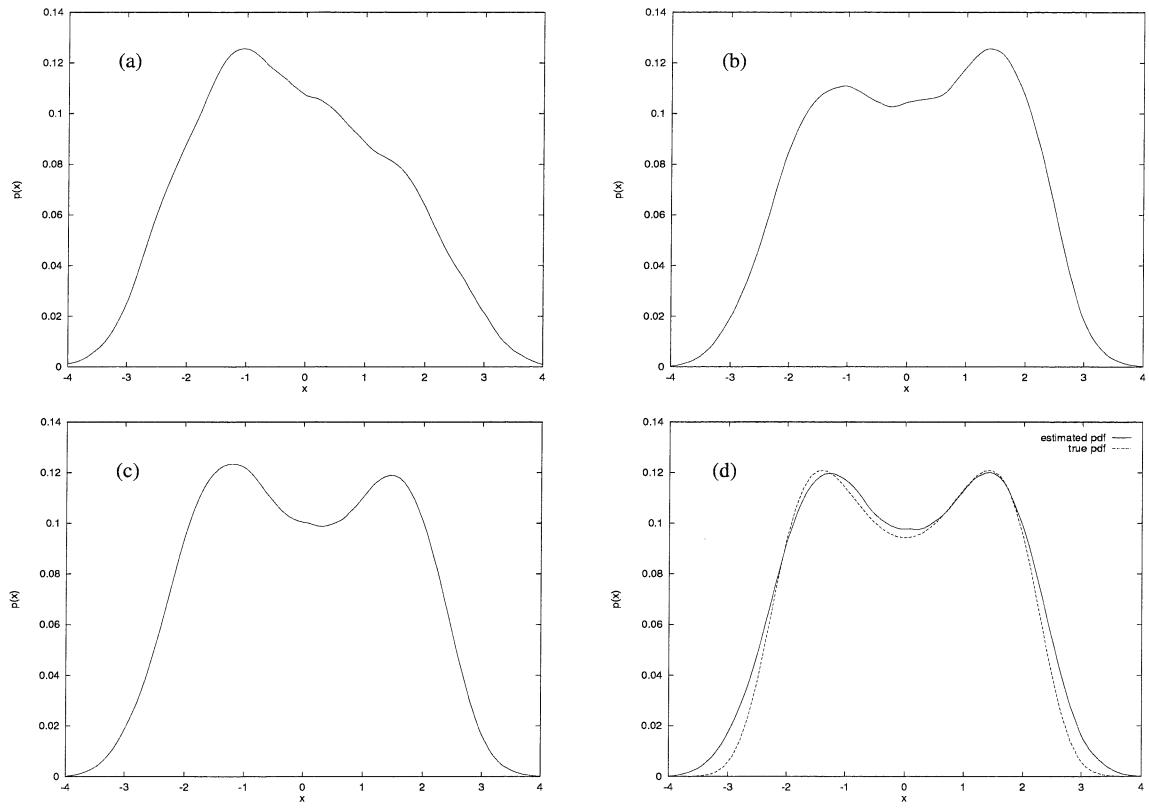


Figure 3.15: Convergence of the kernel estimation of the two-dimensional bimodal *pdf* for the Duffing oscillator (3.29). A one-dimensional slice of the estimated *pdf* is shown for (a) 1000-point sample, (b) 5000-point sample, (c) 8000-point sample, and (d) 10,000-point sample. The *pdf* shown with the dashed line in (d) represents the exact value obtained analytically.

separation and possibly also condition monitoring (for more details about both applications refer to [57, 56]).

3.5.3 Mutual information

Mutual information, $I(x, y)$, is one of the basic functionals introduced in information theory. It is usually defined through entropies. The entropy of the distribution of a single random variable, x, is defined as

$$H(x) = -\int p(x) \log_2(p(x))dx \tag{3.31}$$

For discrete-time observations, the integral in the above formula can be replaced by a sum over all admissible values of x, i.e., over the entire data set $\{x_i\}_{i=1}^N$. The entropy quantifies the average number of bits required to describe a single observation. If x is the only possible value, then $H(x) = 0$. If all observed values are equally probable, then the full resolution of $\{x_i\}_{i=1}^N$ is necessary to describe x, and hence $H(x) = \log_2 N$. The joint entropy of variables x and y is defined as

$$H(x, y) = -\int p(x, y) \log_2(p(x, y))dxdy \tag{3.32}$$

The average amount of information one can learn from the variable x about the variable y is quantified by the mutual information, $I(x, y)$,

$$I(x, y) = H(x) + H(y) - H(x, y) \tag{3.33}$$

This definition implies that $I(x, y) = 0$ only if $p(x, y) = p(x)p(y)$, i.e., if x and y are statistically independent. After substituting expressions (3.31) and (3.32) into (3.33), we obtain another form of $I(x, y)$, [19]

$$I(x, y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dxdy \tag{3.34}$$

From this expression, one can notice that the average mutual information can be interpreted as a "nonlinear extension" of the correlation function.

In nonlinear time series analysis, mutual information is often used to select an "appro-

priate" time delay necessary for the phase space reconstruction (as we did earlier in this chapter). In this work, another application of mutual information is suggested. It comes from an observation that there is predictive power in $I(x, y)$. This application will be fully developed in Chapter 4.

Chapter 4

Verification of Dynamical Model

"Fide, sed cui vide"

The last step in classical system identification is *evaluation* of whether the chosen model adequately serves its purpose, i.e., represents the underlying system for the purpose in hand. In dynamical time series analysis, this problem corresponds to determining whether the underlying system is best modeled as nonlinear, and maybe even chaotic. It is possible to verify applicability of nonlinear dynamical models to the given set of data and perform certain "sanity checks" which, when used with caution, can reduce the danger of making unnecessary or unreasonable assumptions about the system dynamics. Since we have confirmed that the behavior of the gearbox (at least for some set of operating conditions) is well described by a nonlinear dynamical system, we shall now use this knowledge for the purpose of forecasting and condition monitoring. It is worth noting that the "dynamical approach" can be successfully used for other purposes, e.g., noise reduction and signal separation [11]. A monitoring statistic proposed in Section 7.1, which seems to be "natural" for a nonlinear dynamical system, takes advantage of its short term predictability in the reconstructed state space. If the system has explored the state space, i.e., statistics computed from the time series converge, or in other words, the time series is long enough, and a reasonably good predictor has been constructed, then significantly and consistently higher prediction errors would indicate that there is a change in the system operating conditions, or an error occurred during the data acquisition process. As already mentioned above, this statistic can only be applied provided the acquired time series allows a faithful reconstruction of the system dynamics.

4.1 Stationarity of reconstructed system

Statistical analysis is yet another way of exploring dynamical systems quantitatively [57]. Since the long term predictability of chaotic signals is impossible, the best one can do is to find some statistical measure of the dynamics which is independent of the initial conditions and to which the system with an attractor always converges. For example, one can estimate the stationary probability density function, or its functionals, for the trajectory in the reconstructed state space. This information can later be used for diagnostic or prediction purposes. Before that, however, one should confirm that the dynamical properties of the system can be modeled as stationary over some reasonable time frame (they are not "truly" stationary because of slowly occurring deterioration). If both statistical and dynamical measures on the attractor converge within the sample size, one can assume that the sample adequately represents the behavior of the underlying system. If the system is less than three-dimensional, one can graphically display the time evolution of the *pdf* in the system state space as the length of the sample size is increased, as demonstrated on the example of the Rossler system, in Fig. 4.1. To verify stationarity for the gearbox signal, we computed a two-dimensional projection of the estimated five-dimensional probability density function in the reconstructed state space and confirmed its convergence as the number of data points was increased. A two-dimensional projection of the density function, approximated using kernel estimators (see Section 3.5.1), is shown in Fig. 4.2. We also doubled the length of the analyzed time series and repeated the calculations of the time-delay and embedding dimension. The results we obtained were the same as previously shown in Fig. 3.2 and Fig. 3.3, which means that the *pdf* derived from the acquired time series can be successfully used for the purpose of dynamical forecasting and condition monitoring.

4.2 Forecasting

In signal processing, forecasting is very often a goal in itself. In this work, however, we perform short-term prediction using the acquired time series not only to demonstrate the effectiveness of the dynamical approach but also to use the estimated predictor to quantify changes in the system's operating conditions.

The invention of the autoregressive technique by Yule in 1927 marks one of the milestones in the history of time series forecasting. Before that, prediction was attempted by

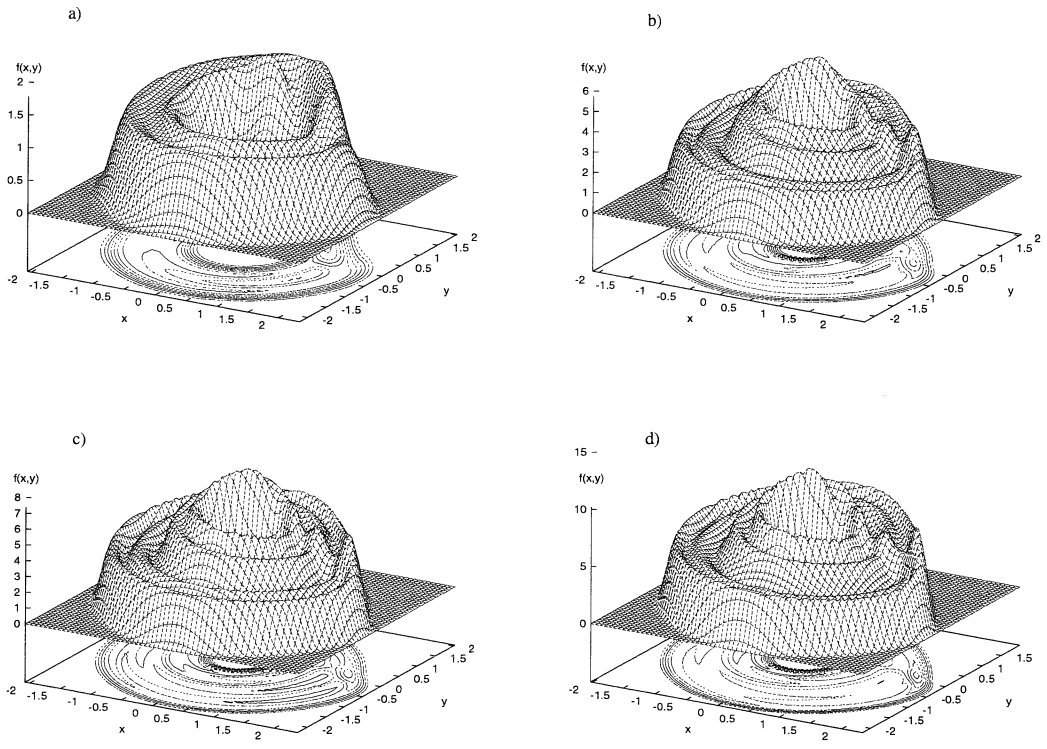


Figure 4.1: Convergence of 2-dim projection of the probability density function estimated for Rossler system. Length of data samples: a) $1,000$, b) $30,000$, c) $60,000$, and $150,000$ data points.

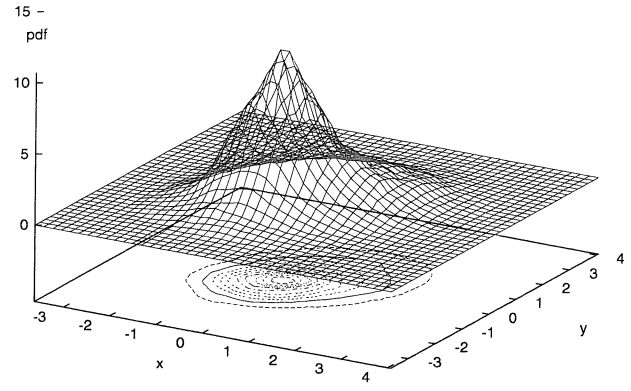


Figure 4.2: 2-dim projection of the probability density function estimated for a gearbox signal obtained at 600 rpm. The variables x and y correspond to the first two coordinates of the state vector.

extrapolating time series through a simple polynomial fit. Yule proposed a model in which the predicted value of the time series is computed as a weighted sum of previously observed values. This method and its variations are still the backbone of engineering time series prediction. Another milestone appeared in early 1980s, when time series generated by a chaotic system was first analyzed in its reconstructed phase space. The method which made such analysis possible, and became known as time-delay embedding, opened new possibilities, especially for investigation of apparently complicated signals whose broadband power spectrum caused a breakdown of linear methods. The main advantage of this nonlinear approach is that it uses information about the state space structure which, in turn, quantifies the underlying dynamics. This information can be used to identify the parameters for an applied local or global, linear or nonlinear, adaptive or non-adaptive input-output model, like the one in equation (3.5).

In local models, we are considering local dynamics of the system. This can be done by, for example, applying polynomials or other natural basis functions to the model (3.5) to represent the time evolution of neighborhoods of trajectory points in the phase space. Global models, on the other hand, provide a closed functional representation of the dynamics in the whole attracting set. They usually approximate the dynamical vector field $\bar{f}(x)$ by

expanding it in a set of basis functions in the phase space. The number of data points used to fit the model parameters is always very large in global models, which is both their strength and weakness. The latter is manifested in large computation time. The strength comes from the fact that the global approach does not rely on accurate determination of the geometry of local neighborhoods, which makes it more robust against contaminated data [2].

The choice between adaptive vs. non-adaptive prediction is strongly related to selection of the length of the prediction horizon. It seems that the choice of the short prediction horizon and an adaptive scheme is an obvious one, especially for chaotic systems where long time prediction may not even be possible. Note, however, that in the adaptive approach we use the predicted time series continuation as input data. That requires a lot of confidence in the applied prediction model and its results. The problem here is that we build a predictor which is based on the points which are known and some that are unknown, i.e., predicted. This may not be a major consideration for non-chaotic, or weakly chaotic[1] systems, but if the signal is contaminated with "noise" and its predictability decreases, using the earlier predicted continuation to fit the prediction model often proves inferior to limiting the training data set to known data. If the length of the required prediction horizon exceeds its maximum admissible value, as estimated using the largest Lyapunov exponent, then the only possibility is an adaptive scheme. One should, however, proceed with extreme caution and be very skeptical about predictions obtained in such a case.

4.2.1 Predictability and its measures

The "statistical pair" observation-prediction, $(y(t), \hat{y}(t))$, where $y(t)$ is the system output and $\hat{y}(t)$ is the model output, characterizes not only the properties of the chosen prediction model but also the properties of the observed physical process. Its time evolution can be observed and on the basis of this information, the system can be classified as predictable, unpredictable, or partially predictable [29]. Denoting the prediction error by $\varepsilon(t) = y(t) - \hat{y}(t)$, the general character of the behavior of the probability density function [5]

$$p = p(\varepsilon, t \mid \mathcal{D}(t_0), \mathcal{M}) \tag{4.1}$$

[1] By a weakly chaotic system, we understand a chaotic system with positive, but relatively small Lyapunov exponents.

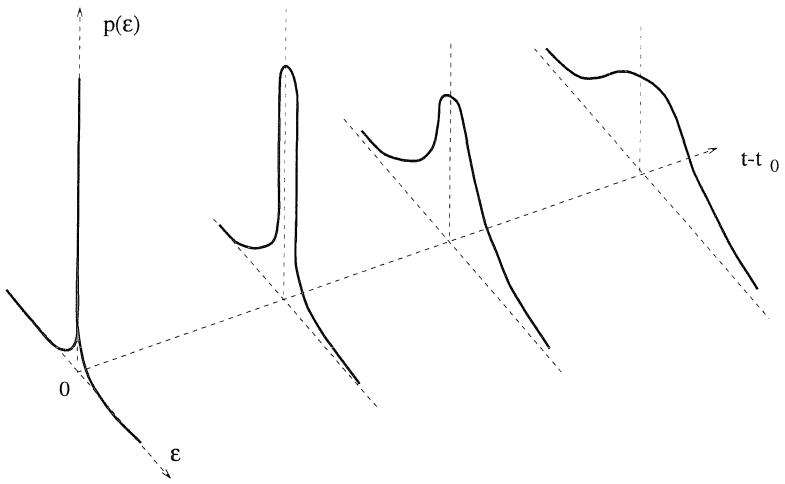


Figure 4.3: Evolution of the conditional probability density function (4.1) quantifying predictability of the system.

can be schematically depicted as in Fig. 4.3 for a one-dimensional case. As shown in the figure, in the limit $t \to t_0$ the density function degenerates into a Dirac delta function. As the length of the prediction horizon increases, the initial density function spreads out, which indicates that the quality of prediction deteriorates. Notice that the quality of prediction depends on the available data, $\mathcal{D}(t_0)$, and on the class of prediction models $\mathcal{M}$. The former dependency will be used in Section 7.1 for the purpose of monitoring changes in the system operating conditions.

The prediction accuracy of the chosen method can be evaluated in many ways. The most commonly used measure of the model accuracy is the variance of the prediction error, ε, averaged over the attractor. Its normalized version is known as the *normalized mean squared error*,

$$NMSE = \frac{1}{\sigma^2 T} \sum_{t=1}^{T} (y(t) - \hat{y}(t))^2 \tag{4.2}$$

where σ is the standard deviation of the known system output throughout the prediction horizon in the training set. Values of $NMSE$ close to zero correspond to high predictability, while $NMSE = 1$ to simply predicting the mean. We decided to use also a more interesting

measure, rich in information, called the *degree of predictability* [29, 51], which is defined as,

$$D(T) = \frac{< y(t+T)\hat{y}(t+T) >}{\sqrt{< y^2(t+T) >< \hat{y}^2(t+T) >}} \tag{4.3}$$

where the angle brackets denote the average over all values of t for which data is available. For this measure, values of D close to 1 correspond to high degree of predictability, while those close to zero to poor predictability. The case $D = 1$ can also be interpreted as a measure which qualifies the physical system as predictable or "deterministic," and $D = 0$ as unpredictable, or "random." In cases where $0 < D < 1$, we classify the system as a partially predictable. Notice that this classification is made purely from the point of view of an experimenter. On the basis of this approach, determinism or randomness of the analyzed system depend on the class of models $\mathcal{M}$ used for prediction of its response. According to this concept, one can define a threshold value of the degree of predictability, $D(T)$, say $1/2$, which can then be used as a quantitative measure of the system randomness [29]. One needs to be very careful with such conclusions, however, because of the conditional dependence of $D(T)$ on the class of selected prediction models, which is hidden in equation (4.3). For more details on this very interesting and important issue, see [5].

Another interesting feature of the degree of predictability, $D(T)$, is that it also provides information about the length of the prediction horizon for which good forecasting is possible (like Lyapunov exponents, but much easier to compute).

We will return to the notion of the degree of predictability, as defined above, in Section 7.1, where it will be used for condition monitoring of response of dynamical systems.

4.2.2 Linear prediction

Since there is vast literature available on linear prediction methods (e.g. [12, 7]), we will review very briefly only their most basic features and discuss why they usually fail when applied to signals from nonlinear or chaotic systems. The biggest advantages of linear models are that they are very well understood, easy to implement, and in many cases, prove to be very successful. The most popular family of globally linear models is known under the name of ARMA models. Two special cases within this family are the *autoregressive* and *moving average* models.

The m-th order autoregressive model, denoted by AR(m), for a time series $y(t)$ is defined

as

$$y(t) = \sum_{i=1}^{m} a_i y(t-i) + e_t \tag{4.4}$$

where t denotes discrete time and e_t can represent either controlled input or noise. Sometimes this model is also called an *infinite impulse response* (IIR) filter, because the output of a stable system to any impulsive input theoretically takes an infinite amount of time to decay away. Note that the input term e_t is crucial for the performance of the autoregressive model. Without it, the model is only capable of reproducing three basic types of linear behavior (periodic solution, or exponential decay or growth to stable or unstable fixed points, respectively).

The n-th order moving average model, denoted as MA(n), is defined as

$$y(t) = \sum_{i=0}^{n} b_i e_{t-i} \tag{4.5}$$

According to this model, the present value of time series is influenced only by the present and n past values of the input series e. It is also called a *finite impulse response* (FIR) filter, because the output to any impulsive input ceases $n+1$ time steps later.

The combination of these two models is called an ARMA(m,n) model

$$y(t) = \sum_{i=1}^{m} a_i y(t-i) + \sum_{i=0}^{n} b_i e_{t-i} \tag{4.6}$$

Application of an ARMA model is thus reduced to fitting its coefficients and deciding on the order of the model. A good model should be able to represent the signal faithfully, using a small number of coefficients and some residual random input. There exist standard techniques for fitting the coefficients for all of the above models (see, e.g., [7].) Unfortunately, there is no unique best choice for the order of an ARMA model. A commonly used strategy is to increase the order of the model as long as it significantly improves the forecasts beyond the training set.

From the early stages of research in the field of dynamical time series analysis, it has been noted that linear models often fail to represent nonlinear (and chaotic in particular) dynamics. In such cases, the errors of the estimation of the coefficients in a linear model (that are chosen on the basis of the best, in the least-squares sense, fit to the observed

data) are often comparable with the size of the attracting set. Even if one uses a more sophisticated approach, like linear neural networks, the results are usually poor. For time series whose FFT is composed of a discrete and a broadband component, linear methods can sometimes be used successfully. Even in those cases, however, nonlinear prediction schemes usually produce better results. It is impossible, of course, to compare efficiency of linear and nonlinear methods in general. Even when two particular classes of models, one linear and the other nonlinear, are chosen and their performance rated, one has to be very careful in concluding that the better predictions of, say, the nonlinear model imply that the underlying dynamical system is nonlinear. Given a prediction scheme one can, in most cases, present another scheme which suits the purpose on hand better. We are stating this apparently simple observation because in many papers published in recent years, claims of the nonlinear character of analyzed systems were made on the basis of observing superior performance of certain classes of nonlinear models. It is true that nonlinear systems can be better analyzed with nonlinear tools but so can some linear systems. In spite of this ambiguity, we decided to compare an average performance of a linear neural network and that of the nonlinear prediction schemes introduced in the next section. We employed a five-layer linear neural net with 100 delayed inputs to predict the future continuation of 400 [*time steps*] ahead for an ensemble of signals acquired from the analyzed gearbox. An example of the performance of the scheme is shown in Fig. 4.4. We compared the normalized mean squared error ($NMSE$) obtained by using the neural net with that obtained with the nonlinear methods and found it to be significantly ($50-80\%$) higher for the linear approach.

Another deficiency of a linear approach is also typical of the ARMA-family models. The coefficients of ARMA models can be determined from the signal power spectrum or its autocorrelation coefficients. This means that all of these three methods of characterization of the system features contain the same information (or lack of thereof.) This is the main reason why ARMA models fail when applied to systems and cases where the power spectrum is not a very useful measure. This problem is of primary importance to state monitoring. It is possible to observe two time series with very different properties but very similar broadband spectra. Also, if there exist changes in the system operating conditions which are not visible in its spectrum, diagnostic methods implementing linear models will not be able to detect these changes. This problem will be discussed in more detail in Chapter 7.

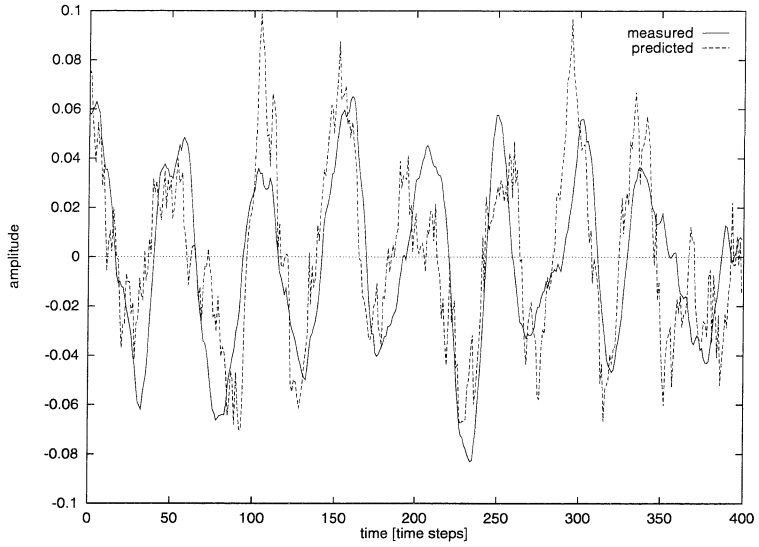


Figure 4.4: Prediction of the gearbox signal using a linear neural network.

4.2.3 State space prediction

The first nonlinear prediction method was suggested by E. Lorenz in 1969. It was also the first method which used information about local structure of a trajectory in its phase space. The proposed method assumed that two trajectory points which were initially close at time t_0 will stay close for at least some elapsed time, t_1. Then, the predicted evolution of the trajectory point $x(N)$ which is t time steps ahead, $x(N+t)$, is equal to the corresponding time evolution of the nearest neighbor of $x(N)$ in the state space.

Another historically important step beyond linear prediction models was the *threshold autoregressive model* (TAR) introduced by Tong and Lim in 1980 [53]. This globally nonlinear model was composed of two local linear autoregressive models evaluated for the system state. It opened the door to a new family of methods which use a set of local linear models for various regions of the state space.

Forecasting in the system state space can be interpreted as a special form of the Yule's original idea for predicting the future state by using previous values of the signal, as in equation (4.6). The order of the autoregressive model corresponds here to the dimension of the state space. The main benefit from using dynamical prediction methodology is that it replaces time series extrapolation with interpolation of the trajectory embedded in its phase space. The method of prediction that we apply here is similar to the scheme suggested by

Sauer in [47].

Let us refer now to the way we defined a mechanical system in eq. (3.5). This model requires specification of two functions, $\bar{f}$ and $\bar{g}$, and estimation of two types of errors, Ξ_k and ε_k. To simplify the prediction process, we use a composition function of $\bar{f}$ and $\bar{g}$, denoted by P, which requires the estimation of only one prediction error, ϵ. Note that ϵ contains both the model and the measurement errors, Ξ and ε. The general predictor, used to extrapolate the time series t-time steps ahead, can thus be expressed in the form:

$$\hat{y}(k+t) = P_t(x(k)) \tag{4.7}$$

where $\hat{y}(t)$ denotes a predicted value of $y(t)$.

We will now predict the short-term future behavior of the time series by fitting a linear model in the state space in the neighborhood of the last given point of the reconstructed trajectory. First, we need to find K nearest neighbors, $x^N_{k_1}, \ldots, x^N_{k_K}$, of the last known trajectory point $x(N)$, where $N = n - (d_E - 1)\tau$ and n is the length of the time series, subject to the condition that only one neighbor is chosen from each nearby trajectory segment. According to this notation, $x^N_{k_i} = x(k_i)$; $i = 1, \ldots, K$ corresponds to the i-th nearest neighbor of the point $x(N)$, and the search for neighbors means a search for values of k_i (see Fig. 4.5). Since we know the complete time evolution of all thus found neighbors, both in $\mathbb{R}^{d_E}$ and in $\mathbb{R}$, we can define the prediction function, $P_t : \mathbb{R}^{d_E} \to \mathbb{R}$, as the following mapping which predicts the time series value t-time steps ahead,

$$P_t : x(N) \mapsto Y_N \equiv \hat{y}(N + (d_E - 1)\tau + t) \tag{4.8}$$

Thus obtained values are used to estimate the future value $y(n+t)$. The calculation of the prediction function $P_t(x(N))$ can be summarized as follows. First, we find the "center of gravity," $x^N_c = 1/K \sum_{j=1}^{K} x^N_{k_j}$, of all the neighbors $x^N_{k_1}, \ldots, x^N_{k_K}$ in $\mathbb{R}^{d_E}$. Then, we find a linear projection space $\mathbb{R}^l$ passing through the point x^N_c, for $l \leq d_E$, which minimizes the squared distances to the neighbors. To do that one can use the singular value decomposition of the matrix whose rows are composed of vectors $x^N_{k_1} - x^N_c, \ldots, x^N_{k_K} - x^N_c$. If we call this matrix A, then its SVD will be of the form $A = U\Sigma V^T$, where U and V are orthogonal matrices and Σ is diagonal with non-increasing non-negative entries. The first l columns

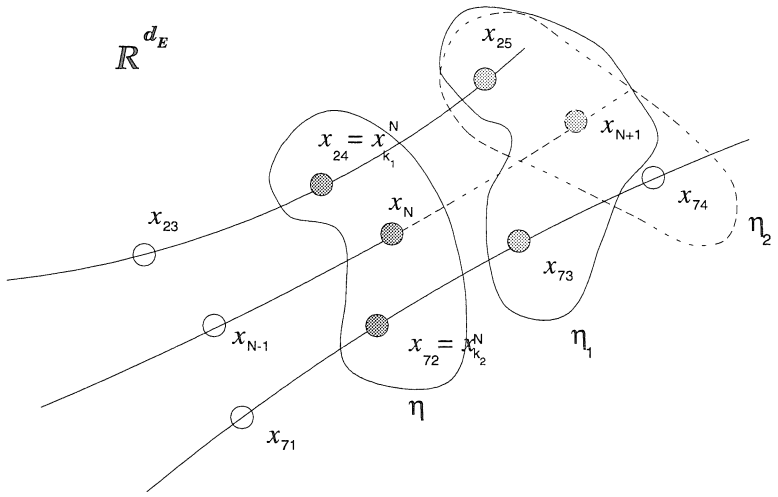


Figure 4.5: Time evolution of the neighborhood of the last given trajectory point x_N. Note that the flow in the phase space does not preserve the nearest neighbors. This phenomenon is responsible for serious limitations of all prediction methods based on the neighborhood to neighborhood mapping of nearest neighbors.

of V span the desired space $\mathbb{R}^l$. The choice of the value of l is somewhat arbitrary (for the gearbox data embedded in $\mathbb{R}^5$ we used $l = 3$ because it resulted in best predictions). In the next step, we project $x^N_{k_1} - x^N_c, \ldots, x^N_{k_K} - x^N_c$ to $\mathbb{R}^l$ (call it the projection function $\Pi : \mathbb{R}^{d_E} \to \mathbb{R}^l$) and form a model $L_N : \mathbb{R}^l \to \mathbb{R}$ which best fits the weighted data points $(\Pi(x^N_{k_1} - x^N_c), Y_{k_1}), \ldots, (\Pi(x^N_{k_K} - x^N_c), Y_{k_K})$. The weights we use are inversely proportional to the distance between the point $x(N)$ and its neighbors. The purpose of the weights is to emphasize the importance of those neighbors that are closer to $x(N)$ in the phase space. The model can be chosen as the simple affine model

$$L_N(\bar{x}) = a_N^T \bar{x} + b_N \tag{4.9}$$

where $a_N, \bar{x} \in \mathbb{R}^l$ and $b_N \in \mathbb{R}$. Finally, we project $x(N) - x^N_c$ on $\mathbb{R}^l$ and evaluate the above model there to obtain an estimate of $y(n+t)$,

$$\hat{y}(n+t) = P_t(x(N)) = L_N(\Pi(x(N) - x^N_c)) \tag{4.10}$$

The result of applying this procedure to a segment of the gearbox data is shown in Fig. 4.6. One can see that the obtained prediction is reasonably faithful. The above scheme was

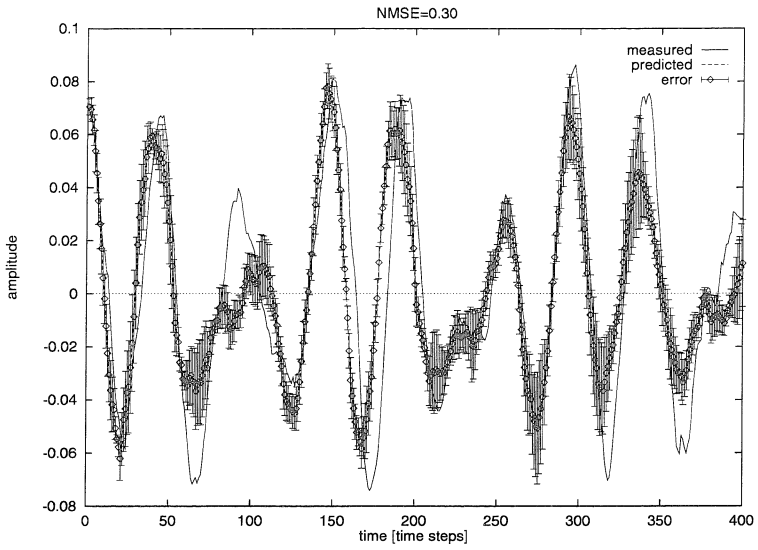


Figure 4.6: Prediction of the gearbox signal using a local linear model in the 5-dim embedding space reconstructed from the given 60,000 point data segment. The estimated prediction error bars are shown on the predicted signal.

applied for various values of embedding dimension and time-delay in order to determine their optimal values from the prediction viewpoint. The best prediction of the gearbox signal was obtained for $d_E = 5$ and $\tau = 9$ and hence these will be considered the final values of the embedding parameters.

4.2.4 Prediction with mutual information

As already mentioned earlier, mutual information among measurements u and v drawn from sets U and V of possible measurements is defined as [19]

$$I(u,v) = \log_2 \frac{p(u,v)}{p(u)p(v)} \tag{4.11}$$

$I(u,v)$ quantifies a number of bits of information one has about v having already observed u (or vice-versa because $I(u,v) = I(v,u)$.) If the observations are independent, then $p(u,v) = p(u)p(v)$, and as a result $I(u,v) = 0$, as expected.

The mutual information function can be used for forecasting the future evolution of a trajectory in its phase space. In this case, $u, v \in \mathbb{R}^{d_E}$ would be a pair of trajectory points, the latter being a future evolution of the former. Suppose that we are interested (as in the

previous section) in predicting a state of the system t time steps ahead of the last known measurement $x(N)$, or rather $y(n)$. We need to set $u = x(N)$ and $v = x(N+t)$, and estimate the required densities $p(x(N))$, $p(x(N+t))$, and $p(x(N), x(N+t))$. The most probable future system state can be found by maximizing the mutual information function (4.11) over all admissible values of v. This would be a very difficult and time consuming task which would involve looking for a global maximum of a function in a multi-dimensional space. Instead of considering a pair of trajectory points in $\mathbb{R}^{d_E}$, we take the second point of the pair to be a projection of $x(N+t)$ on $\mathbb{R}$, i.e., $v = \hat{y}(N + (d_E - 1)\tau + t)$ as in the dynamical scheme in (4.8). There are two immediate benefits from using this approach. First, it significantly simplifies the optimization process by reducing the range of the maximized function from a subset of $\mathbb{R}^{d_E}$ to a subset of $\mathbb{R}$. Second, it eliminates the need for projecting the final result on the measurement space.

Probability density functions entering eq. (4.11) are computed using the Epanechnikov kernel estimator, as described in Chapter 3. Note that $p(x(N))$ is just a constant and will remain unchanged during optimization. The prediction process can, therefore, be reduced to finding a value of y which maximizes the ratio $R(y; x(N)) = p(x(N), y)/p(y)$.

Before we attempt maximization, we initially bracket the maximum of R using our knowledge of the system dynamics in its state space. The initial bracket is chosen as an open interval $y(k_1+(d_E-1)\tau+t)\pm\sigma/4$, where k_1 is the number corresponding to the nearest neighbor of $x(N)$ in $\mathbb{R}^{d_E}$ and σ is the sample standard deviation. Choice of the length of the initial bracket is somewhat arbitrary and reflects a balance between computation time and the uncertainty in inferring the closeness between the future evolutions of initially nearest neighbors. We chose this particular length of the bracket because it worked most consistently in cases we considered. As in the dynamical prediction algorithm in the previous section, the nearest neighbor, $x^N_{k_1}$, should be chosen from a nearby trajectory segment, closest to the segment containing $x(N)$. In the next step, the length of the initial bracket is reduced by a commonly used bracketing method implementing the golden section search [44]. Now, we perform a function maximization (over all possible values of y in the final bracket) which utilizes Brent's method. Since this method is only capable of finding a local maximum, we partition the final bracket into a set of sub-brackets and then look for "all" local maxima of R. Finally, the value of the largest local maximum is chosen as the most probable value of $\hat{y}(n+t)$.

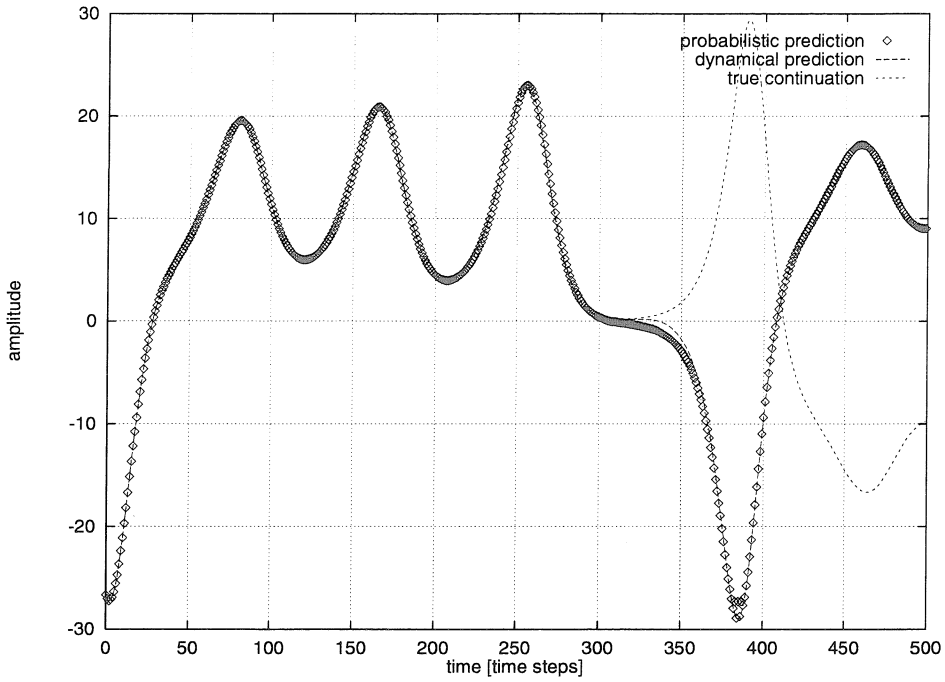


Figure 4.7: Comparison of prediction results obtained for a "clean" Lorenz signal, using the mutual information approach and a purely dynamical prediction scheme.

Because this approach implements both dynamical and statistical tools, it is more robust than the purely dynamical prediction scheme introduced in the previous section. Since it is hard to imagine any experimental data without noise, such robustness is a highly welcomed feature. To illustrate this, we predicted evolution of a chaotic signal, generated by numerical integration of a well known Lorenz dynamical system [2], using both prediction schemes. In both cases, the trajectory was embedded in the phase space using the same embedding parameters, $d_E = 3$ and $\tau = 20$. Fig. 4.7 shows almost identical effectiveness of both methods for the signal without noise. A big difference in performance appears when we contaminate the signal with a uniformly distributed noise, whose amplitude is 10% of the amplitude of the "clean" signal. From Fig. 4.8 and Fig. 4.9 we can see the benefit from using the mutual information prediction. Fig. 4.10 and Fig. 4.11 show results of prediction for the gearbox signal. Also, in this case, the combined dynamical and statistical approach (mutual information function in the state space) proved to be more efficient ($NMSE = 0.40$, as opposed to $NMSE = 0.46$ for the dynamical scheme with the same embedding parameters $d_E = 5$, and $\tau = 9$).

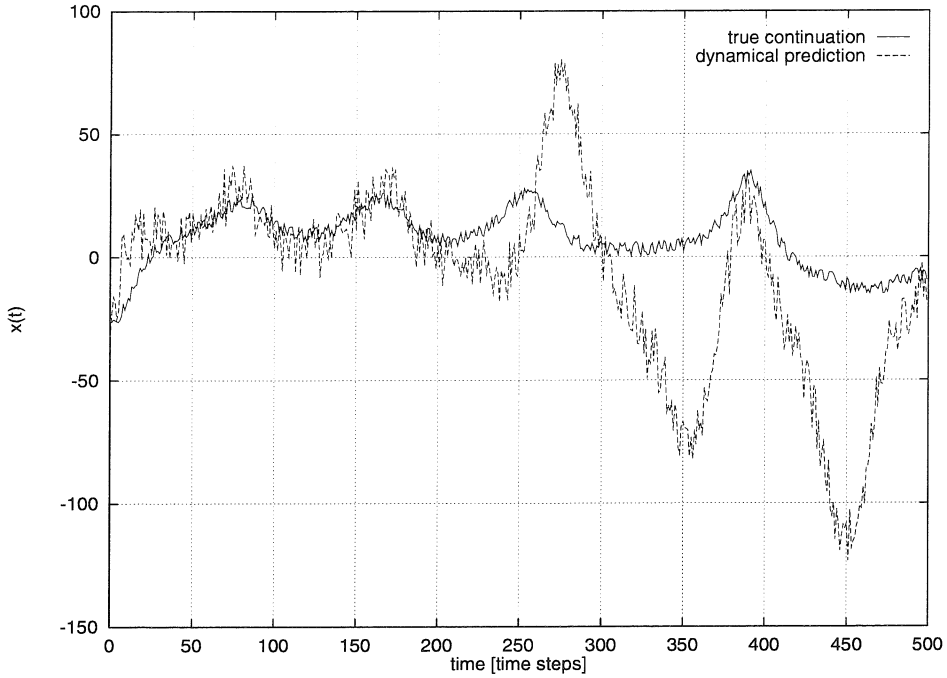


Figure 4.8: Prediction for a Lorenz signal contaminated with uniformly distributed random signal ($SNR = 10$), obtained using purely dynamical prediction scheme.

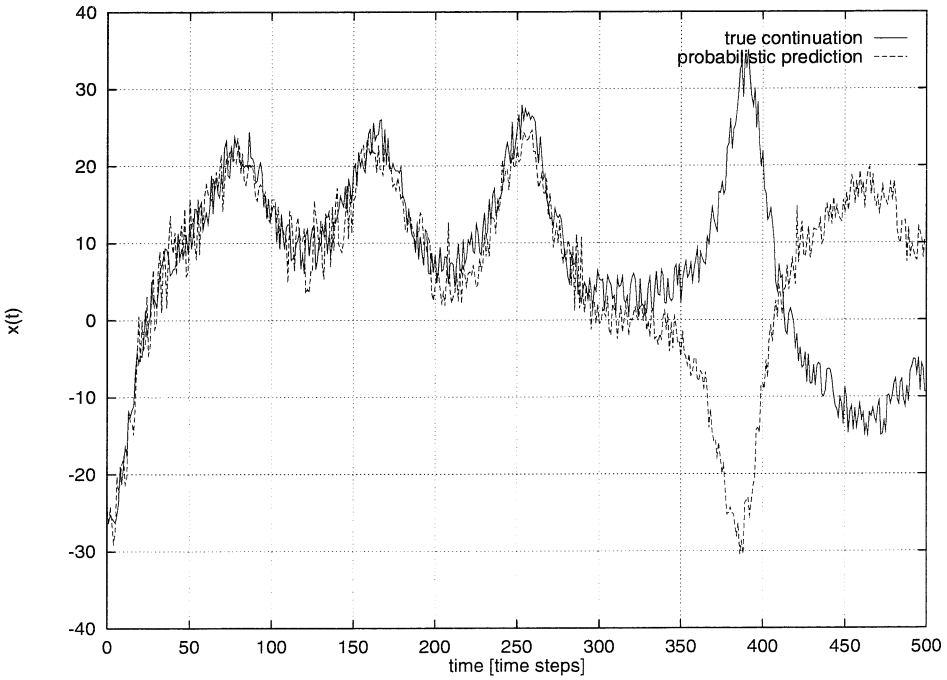


Figure 4.9: Prediction for a Lorenz signal contaminated with uniformly distributed random signal ($SNR = 10$), obtained using the mutual information prediction scheme.

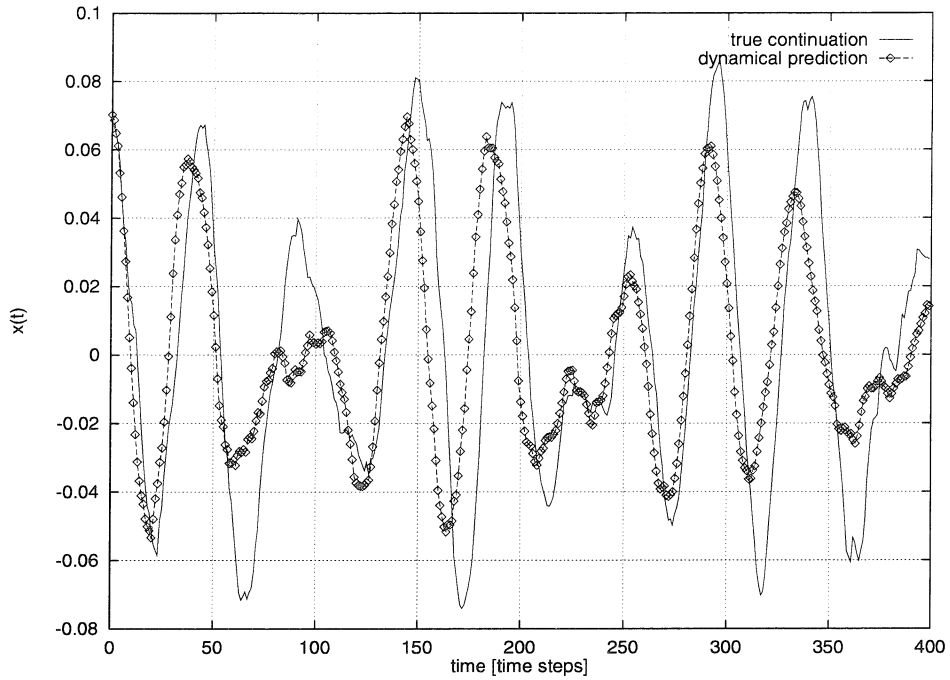


Figure 4.10: Prediction results obtained for a gearbox signal using a purely dynamical prediction scheme.

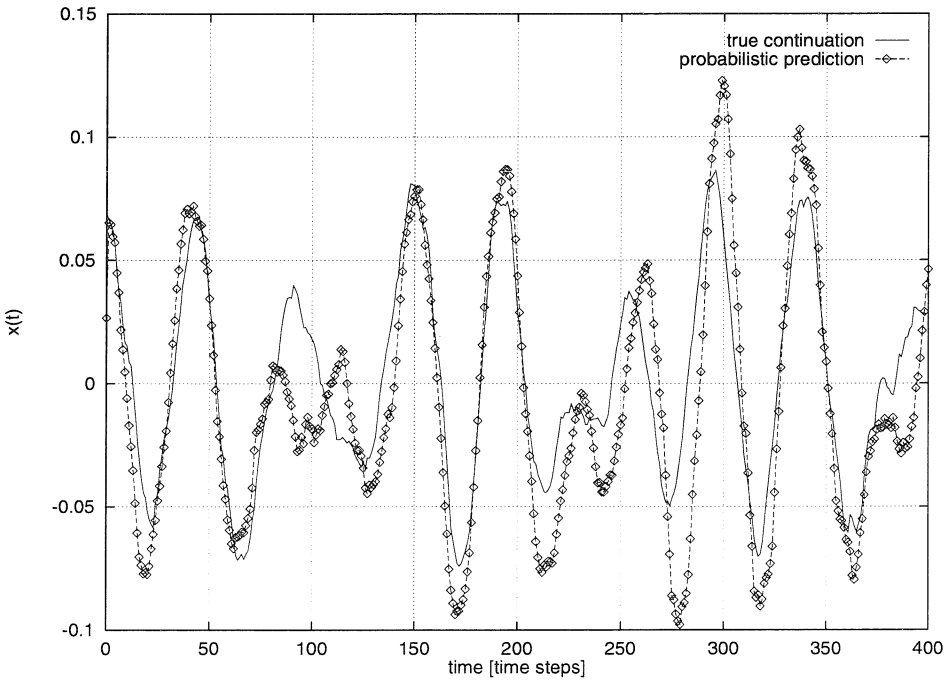


Figure 4.11: Prediction results obtained using a combined dynamical and probabilistic approach.

Chapter 5

Dynamical "Noise" Reduction

"Quis separabit?"

"Noise" reduction, or filtering, is a term commonly used for a class of problems in which a time series is "cleaned" or, in other words, separated into "noise" and signal of interest, on the basis of some objective criterion. From this description, one might conclude that the unwanted or uninteresting part of the analyzed signal is referred to as "noise" (at least this meaning of the term is most frequently adapted in engineering literature). Another definition of "noise" is employed in experimental statistical literature where it is understood as the unpredictable ("random") part of the analyzed signal, or sometimes just as a signal generated by a pseudo-random number generator. A different definition which has been reasonably well established in the field of dynamical time series analysis (and also will be used in this work), is that by "noise" we understand a signal resulting from a high-dimensional dynamical system. How high-dimensional the system needs to be in order to produce "noise" depends on the particular application. For example, if we wanted to separate a signal which is of interest to us and which can be embedded in $\mathbb{R}^5$ from another signal which can be embedded in $\mathbb{R}^{20}$, the latter signal would be considered "noise" and might be characterized in some conventional statistical sense. It is worth emphasizing that this distinction is caused strictly by practical (numerical) considerations. In most cases, the benefits from analyzing a high-dimensional signal in its state space, as opposed to using statistical methodology, do not justify time consuming computations required by dynamical analysis. That is why it is easier and faster to characterize noise using well established statistical tools.

Conventional noise reduction methods employ various types of linear filters. They all work on the assumption that the broadband components of the spectra are related to "noise"

while the sharp peaks relate to the "clean" signal. This assumption can be made for sources which result in periodic or quasiperiodic signals. However, if the broadband spectrum is an inherent feature of the pure analyzed system, as it is in chaotic systems, a linear filter used to remove noise will also result in distorting the signal. There has been a number of recently proposed dynamical noise reduction methods which use the fact that deterministic chaotic motion takes place on attractors which are smooth submanifolds in the state space, and therefore it can be analyzed as locally linear [15, 20].

When a deterministic signal is contaminated with noise, two cases of such contamination are usually considered: additive *measurement* (or *observational*) noise and additive *dynamic* noise [20]. We are presented with the first kind when the measured signal, $y(t)$, is corrupted by signal(s) independent of the system dynamics according to the formula

$$y(k) = g(F^k(u_0)) + \eta(k) \tag{5.1}$$

where g denotes a measurement function, as in equation (3.3), and η, the contamination. Dynamic noise, on the other hand, influences the measurement process indirectly by acting on the system dynamics, and can be represented as

$$y(k) = g(F(u(k-1)) + \eta(k-1)) \tag{5.2}$$

Most of the noise reduction schemes proposed in the literature on dynamical time series analysis consider the former case and very few, also the latter one. Note, however, that if the state embedding model

$$\begin{aligned} x(k+1) &= \bar{F}(x(k)) \\ y(k+1) &= G(x(k+1)) \end{aligned} \tag{5.3}$$

has a linear measurement function G (as it is in our case, where G is just a simple selection of the first coordinate of $x(k+1)$), both of the noise models (5.1)-(5.2) are essentially the same.

In dynamical analysis of time series, the problem of appropriate reconstruction of system dynamics in the presence of noise has attracted significant attention in the last few years. It is because every experimental time series is contaminated with one or the other type of

noise. If one is interested in accurate predictions or determination of system invariants, like embedding dimension, Lyapunov exponents, or the *pdf* in the state space, one should first decontaminate the analyzed data. On the other hand, whenever any type of noise reduction scheme is used, some information about the system is lost. In cases where the observer is not interested in very accurate estimation of the system invariants but, for example, only in detecting their change, any attempt to separate signal from noise carries a danger of potential decrease in the sensitivity of the applied condition monitoring method. Provided a reasonably high signal-to-noise ratio (SNR) is achieved[1], we found that it is safer for the purpose of the system condition monitoring to use the observed time series without any software filtering or preprocessing.

An important application where dynamical noise reduction schemes might play a significant role is in the field of acoustic noise control [11]. Many noise control problems can be reformulated as signal separation problems. Even if the signal of interest is high- (possibly infinite-) dimensional, the contaminating processes usually have a small number of active degrees of freedom. Therefore, they can be treated as deterministic nonlinear dynamical systems whose randomness is accounted for by chaotic dynamics. The noise control problem can thus be reduced to the separation of the low-dimensional signal and background noise.

There are several methods available for reducing noise that are built into possible phase space reconstruction schemes (like one proposed by Broomhead and King [9]). They usually involve singular value decomposition, which is locally or globally performed on the matrix constructed from the analyzed time series. More sophisticated methods are also available that exploit the local behavior of the system in the phase space. Some of them attempt to reduce noise by adjusting the analyzed data in the process of the estimation of local linear models of the system dynamics [27, 20, 2]. The adjustments are performed so that the resulting "data" is more consistent with the estimated dynamics than the original data. Some of these schemes are performed in just one step and some can be iterated.

We will now discuss three methods of noise reduction which we found interesting and useful for the purpose of noise reduction. The first method uses different predictability of the composite signals as a distinguishing characteristic. It requires certain knowledge about the prediction horizon of the signal one attempts to recover. The second method

[1] as it can be concluded on the basis of the false nearest neighbors statistic shown in Fig. 3.3 for the gearbox data

applies a simple iterative technique of trajectory smoothing in the phase space. The last method is also iterative and uses an approach in which signals are separated on the basis of the statistical properties of the observed dynamics. Both of these methods work best when a sample of clean data is available. Their efficiency very strongly depends, of course, on the length of the available reference time series. All three methods are superior to the standard linear filters in the cases where the signals involved possess broadband spectra. Even though they were designed primarily for nonlinear systems, they also perform well when one or more of the signals that need to be separated correspond to linear dynamics.

5.1 Separation through prediction

Since this approach is described in detail in [51] and [11], we will only briefly summarize it here. Suppose that the observed signal $y(k)$ is composed of the signal of interest, $y_1(k)$, and other signals, $y_2(k), y_3(k), \ldots$, etc. We need to identify some *distinguishing characteristic* for the above signals which will enable us to separate them. Then signal separation can be attempted using nonlinear prediction methods. We restrict our analysis to the case where at least some of the measurements, $y_i(k)$, are also available. These reference signals will be used to establish the geometrical and statistical properties of the dynamical systems they represent.

There are several possible ways of selecting the distinguishing characteristic (see for example [2]). In our work we used the approach similar in spirit to that proposed by Taylor [51] who applied the *degree of predictability*[2] criterion to separate the signals. The characteristic we employed could be called a *local predictability measure* and is based on computation of the *local Lyapunov exponents* [3]. This approach enables us to quantify the short term predictability by specifying the time horizon in which the accuracy of prediction is best for the signal of interest. The length of the prediction horizon is proportional to the rate at which the local Lyapunov exponents converge to their global values. Then, if we can establish the value of the prediction horizon, τ_p, for which either of the signals, say $y_p(k)$, is more predictable than the rest, this signal can be separated by iterating the nonlinear prediction scheme τ_p time steps into the future.

One needs to build a prediction model on the basis of the knowledge of $y_p(k)$ and by

[2] See Section 4.2.1.

[3] For detailed information about local Lyapunov exponents, refer to [1].

working in its embedding space, say $\mathbb{R}^{d_p}$. As discussed in the previous chapter, there is a large variety of prediction models that can be used here. From our experience, the local neighborhood-to-neighborhood models defined in a subspace $\mathbb{R}^l \subset \mathbb{R}^{d_p}$ work best also in this case. After applying one of the schemes introduced in Chapter 4.2, the estimated process $\bar{y}_p(k)$ can be now subtracted, within its embedding space, from the composite signal and the result projected back on the $y(k)$ axis. If needed, this scheme can be iterated in order to separate more signals and this way, hopefully, obtain the required noise reduction.

The biggest advantage of this method is that its application is not limited to any frequency bandwidth of the contaminating signals. Unfortunately, it has many limitations. The biggest limitation is that the length of the maximum prediction horizon has to be significantly different for composite signals if one hopes to separate them with this method. Even though theoretically this procedure can be used to separate more then two signals, practically it is very difficult to achieve. One of the reasons for that is the limitation mentioned above. It is not very likely that one is presented with a set of mixed signals with significant differences in their predictability. Another significant limitation is that this approach does not work for the case of dynamic noise. Finally, the efficiency of the method is limited by the quality of the applied prediction scheme.

5.2 Smoothing in state space

In 1989, Kostelich and York [28] proposed a method of noise reduction based on the idea that since dynamics on manifolds can be considered locally linear, noise in the signal can be reduced by performing local linear approximations of the system dynamics in the neighborhood of every trajectory point, and then finding perturbed orbits which best satisfy the derived linear maps. In the first step of that method, a set of neighbors was found for every trajectory point and used to compute a linear approximation of local system dynamics, $f : x(i) \mapsto x(i+1)$. In the second step, the stored approximations to $f(x)$ along a segment of trajectory were perturbed in order to find a ("clean") trajectory which was more consistent with the dynamics. There are some problems with this approach, however. The most significant difficulty is that in order to approximate Jacobians accurately, even in a relatively low-dimensional phase space, one needs to find many close neighbors to every trajectory point. This requirement is rarely satisfied. There can be at least two potential remedies for

this problem. One can project the dynamics on a lower-dimensional subspace, as we did in Section 4.2.3, or replace the estimation of the Jacobian with a simpler method of trajectory approximation. One way of performing such an approximation is to fit a multi-dimensional polynomial to the sequence of trajectory points. This is what we do in the examples shown below. We also replace the least squares problem proposed in [28] with a simpler one which minimizes the Euclidean distance between a trial (perturbed) trajectory, $\{\hat{x}(i)\}$, and the image of the trajectory under the polynomial approximation $\{f(x(i))\}$. We put a restriction on the distance a point can be moved. The points which require particularly large adjustments as a result of the minimization are, therefore, left unchanged. In what follows, we use a simple quadratic polynomial, $f(x):\ a_0+a_1x+a_2x^2$ (a_i for $i=0,1,2$ can be made vectors or scalars), which practically, in case when the coefficients are scalars, requires finding only six near neighbors for every trajectory point [11]. We found that using the scalar coefficients for low-dimensional systems lead to the same results as using the vector coefficients. The neighbors are found with the restriction that they are all close in the phase space but not close in time, as also required in Section 4.2.3 in the prediction scheme used there.

If a segment of clean data is available, it can be used to fit the local polynomials. If we are presented with already contaminated data, f is first estimated directly from the noisy data and then the procedure is iterated using the smoothed trajectory, until the results cease to improve.

To illustrate this method, we applied it to the well known Lorenz system [32], whose time series was contaminated with a uniformly distributed measurement noise obtained from a pseudo-random number generator. The resulting time series had the signal-to-noise ratio[4]: SNR= 14.54 dB. After one iteration of the noise reduction procedure, the SNR of the resulting signal was raised to 18.43 dB. Fig. 5.1 shows the results of this procedure after one, four, and five iterations were performed. After five iterations, we obtained a noise reduction of 5.74 dB. Further iterations of this scheme did not result in significant noise reduction.

The method of smoothing in the phase space is also applicable to linear systems. To

[4]We define SNR as

$$SNR = 10\log_{10}\left(\sum_{i=1}^{n} y_1^2(i)/\sum_{i=1}^{n}(y_1(i)-y(i))^2\right)$$

where y_1 is the clean signal and y, the observed contaminated signal.

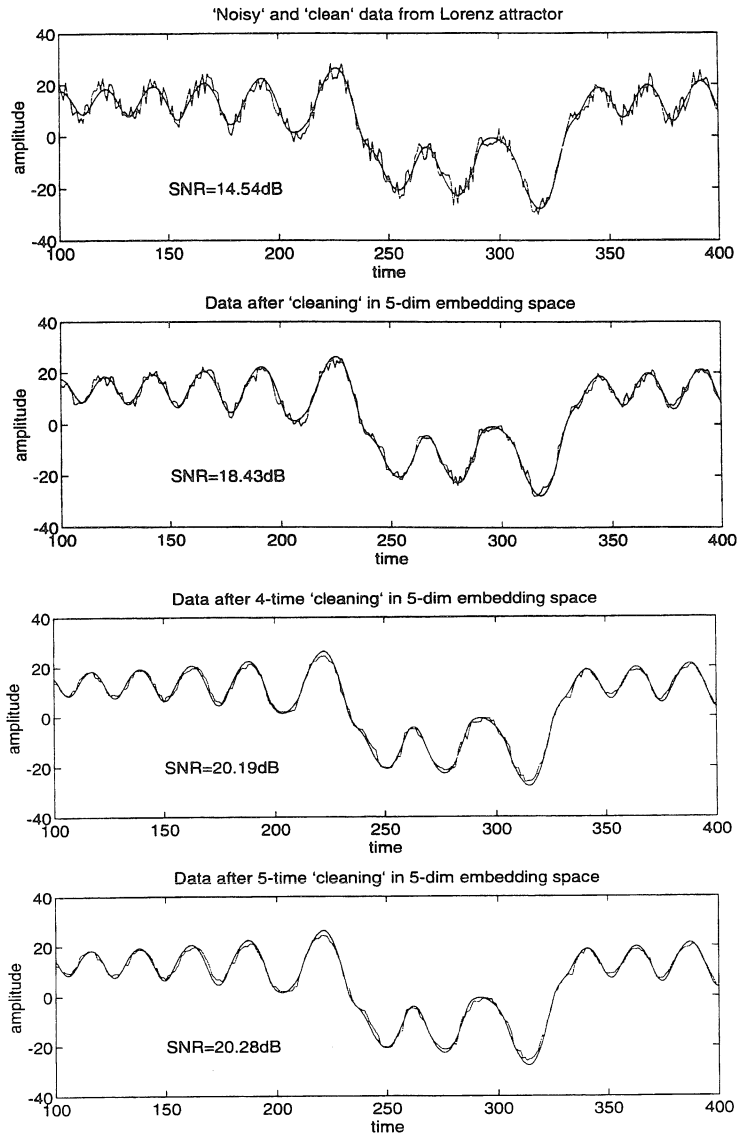


Figure 5.1: Contaminated and clean (shown as smooth lines) signals illustrating the iterative application of the trajectory smoothing technique in $\mathbb{R}^5$. The example used was the signal acquired from the Lorenz system contaminated with uniformly distributed noise obtained from a pseudo-random number generator.

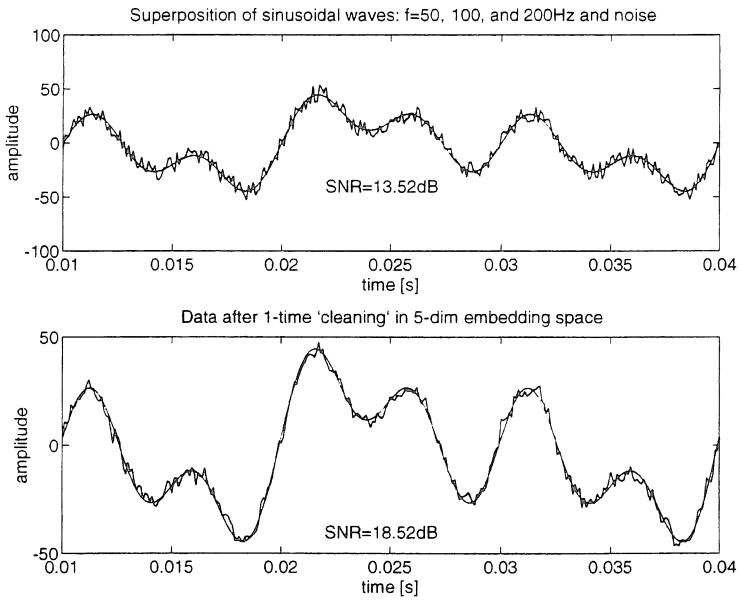


Figure 5.2: Noise reduction by applying the trajectory smoothing technique iteratively in $\mathbb{R}^5$. The example used was a periodic signal composed of three superposed harmonic signals that was contaminated with uniformly distributed noise obtained from a pseudo-random number generator.

demonstrate that, we will use a periodic signal composed of three superposed harmonic signals ($f = 50, 100$, and 200 Hz) and contaminated with uniformly distributed (measurement type) noise obtained from a pseudo-random number generator. The initial value of SNR= 13.52 dB was increased by 7.43 dB after the noise reduction scheme was iterated three times. Fig 5.2 shows the signals before and after single application of this procedure.

From our experience in the application of the method of smoothing in the reconstructed state space, we conclude that it works well for low-dimensional nonlinear and linear systems, which are derived from a one-dimensional time series contaminated with noise. The method seems to be suited particularly well for the purpose of reduction of high-frequency measurement noise. Its effectiveness depends on the signal-to-noise ratio and on the choice of dimension of the embedding space in which the smoothing is performed. For example, for the periodic signal analyzed above, the change in the embedding dimension from 4 to 5 resulted in improvement of the noise reduction by about 2.5 dB. Further increase in the

embedding dimension did not improve the results any further. This observation suggests that smoothing in the reconstructed phase space can also provide information about the appropriate choice of the embedding dimension for contaminated data. The power spectra of the analyzed signals suggest that most of the high-frequency noise was eliminated. It is impossible, however, to give a precise estimate of the noise reduction for a typical chaotic experimental data.

5.3 Probabilistic cleaning

Probabilistic signal cleaning refers to a group of methods which, instead of trying to approximate the dynamics of the system, take into account its statistical properties in the state space [15, 33]. These statistical properties are reflected in the distribution of data points in the state space and are evaluated using a clean or noisy reference signal, which was acquired during earlier measurement. The approach we describe in this section is essentially the same as that which was proposed by Marteau and Abarbanel [33]. The only difference is that instead of the *pdf* kernel suggested in [33], we used the Epanechnikov kernel (as described in Section 3.5.1). We will illustrate this method with an example involving measurement noise and assume that a sample of clean reference signal is available to the observer. Notice that, except for coming from the same source, the reference signal, y_r, and the clean signal, y_1, which is to be recovered, are uncorrelated, especially if they come from a chaotic system. We will also assume that the contaminating signal is independent of the clean signal and represents identical but statistically independent (*iid*) quantities at each time.

In the short description of the method below we will use the following notation. We will denote the clean, the reference, and the observed signals reconstructed in the phase space by x_1, x_r, and $x \in \mathbb{R}^d$ respectively. We assume that the sequence of points $\mathcal{D}_{1,N} \equiv \{x_1(k)\}_{k=1}^{N}$ evolves in the state space according to

$$x_1(k+1) = F(x_1(k)), \quad k = 1, 2, \ldots, N \tag{5.4}$$

even though we neither know the map F nor try to estimate it. The same map governs the dynamics of the system represented by the reference orbit $\{x_r(k)\}_{k=1}^{m}$. To simplify the notation further, we denote the set of observations by $\Omega_{1,N} \equiv \{x(k)\}_{k=1}^{N}$. We are inter-

ested in finding the most probable set of trajectory points that maximizes the conditional probability of the state of the system being $\mathcal{D}_{1,N}$ given the observations $\Omega_{1,N}$,

$$P_{max}(\mathcal{D}_{1,N}|\Omega_{1,N}) = \max_{\mathcal{D}_{1,N}} \frac{P(\Omega_{1,N}|\mathcal{D}_{1,N})P(\mathcal{D}_{1,N})}{P(\Omega_{1,N})} \tag{5.5}$$

Only the two probabilities in the numerator need to be estimated for that maximization. Using the fact that the clean data satisfies the deterministic relationship (5.4), the probability $P(\mathcal{D}_{1,N})$ can be expressed as

$$P(\mathcal{D}_{1,m}) = P(x_1(m)|x_1(m-1))P(\mathcal{D}_{1,m-1}) \tag{5.6}$$

Since we assumed the additive (measurement) and the *iid* type of noise, the conditional probability $P(\Omega_{1,m}|\mathcal{D}_{1,m})$ can be simplified to the following form

$$P(\Omega_{1,m}|\mathcal{D}_{1,m}) = P(x(m)|x_1(m))P(\Omega_{1,m-1}|\mathcal{D}_{1,m-1}) \tag{5.7}$$

This leads to the recursion relation for the conditional probability [33]

$$P(\mathcal{D}_{1,m}|\Omega_{1,m}) = P(x(m)|x_1(m))P(x_1(m)|x_1(m-1))P(\mathcal{D}_{1,m-1}|\Omega_{1,m-1}) \tag{5.8}$$

which needs to be maximized over all possible values of the state $\mathcal{D}_{1,m}$. Now, if we call the product of the two last probabilities in equation (5.8) the *forward probability*, the optimal forward probability ending at the state $x_1(m)$ can be recursively defined as

$$PF(x_1(m)) = \max_{\mathcal{D}_{1,m-1}} [P(x_1(m)|x_1(m-1))P(x(m-1)|x_1(m-1))PF(x_1(m-1))] \tag{5.9}$$

Finally, the conditional probability maximized over all admissible trajectory locations at time m is

$$P_{max}(\mathcal{D}_{1,m}|\Omega_{1,m}) = \max_{x_1(m)} [P(x(m)|x_1(m))PF(x_1(m))] \tag{5.10}$$

Similarly, one can define the optimal *backward probabilities* as a recursion which starts at time N and moves back to m. However, if one is presented with sufficiently long data,

the application of the forward, backward, or hybrid of both probabilities leads to the same results. In our work we used the forward probabilities only.

What now remains to be done is to estimate the probabilities entering the recursion formulas, using both the analyzed and the reference signals. As we already mentioned above, we estimate them in the same way as proposed by Marteau and Abarbanel, but use the Epanechnikov kernel instead. For more details about this process, refer to [33].

To demonstrate this method of state estimation and signal separation, we applied it to a mixture of two deterministic signals: one harmonic and the other obtained from the Lorenz system. We used a segment of trajectory from the Lorenz system as a reference time series. Another segment from the Lorenz system, obtained from a set of different initial conditions, was then contaminated with a harmonic signal. Both the reference and the observed signals were embedded in $\mathbb{R}^5$ using the method of time delays, where the probabilistic cleaning was attempted. The result is shown as the two-dimensional projection of the phase space portrait for the sinusoidal signal and the composite signal in Fig. 5.3. The upper part of the figure displays the phase space representation of the harmonic signal which, together with the Lorenz signal, form the dynamics represented by the middle figure. The bottom figure corresponds to the separated harmonic signal which was obtained after 30 iterations of this procedure. Even though the time series from the Lorenz system was used as reference data, we chose to illustrate the result by displaying the separated harmonic component. This was done to demonstrate the effectiveness of this method more clearly.

The main limitation of this method is its dependence on the availability of a clean reference signal, which is used to estimate the required probabilities. Even though, in principle, a part of the contaminated signal can be used iteratively as a reference orbit, in practice the performance of this scheme drops very significantly.

As demonstrated above, this method also allows the separation of two signals coming from low-dimensional sources. It works when the signal-to-noise ratio is as low as 0 dB and sometimes even lower [33]. It allows a gain in the SNR of even 30 to 40 dB, provided the initial signal-to-noise ratio is sufficiently high. As with the other methods presented in this chapter, this approach can be also applied to linear systems.

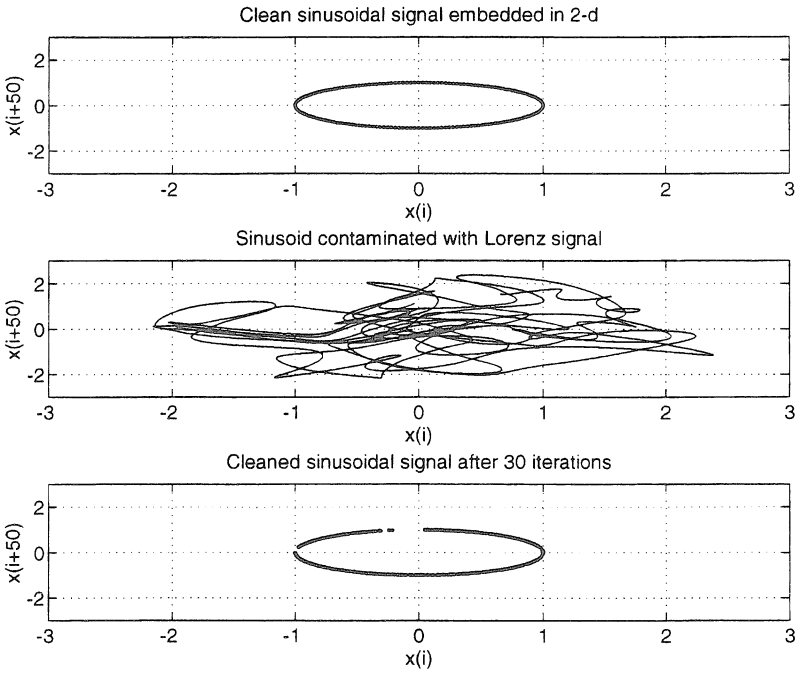


Figure 5.3: Iterative signal separation applying the probabilistic cleaning in $\mathbb{R}^5$. The example used was the signal obtained from the Lorenz system and a sinusoidal signal.

Chapter 6

Analytical and Numerical Models of Gear Dynamics

"Vox populi, vox Dei"

In recent years, a number of papers has been published whose goal was to explain chaotic vibrations in various types of machinery, e.g., tool chatter, gear transmission rattling, and contact problems [40]. In gearboxes, much attention has been paid to modeling and understanding the mechanisms causing their rattling behavior [22, 21, 24, 42, 35]. Among many factors that contribute to the nonlinear behavior in gearboxes are the backlash, the contact forces, and nonlinear and variable mesh stiffness.

Since vibration data for damaged gears was unavailable, we demonstrate the sensitivity of the proposed condition monitoring methods using "synthetic" data obtained from the models we present below. The first model is a nonlinear model of torsional vibrations of the gear system subjected to a variable, nonlinear mesh stiffness, backlash, and harmonic excitation. The second model describes the dynamics of the rolling motion of a smooth rigid cylinder driving another flexible cylinder with a surface waviness. This waviness is responsible for exciting the system vibrations. Nonlinearity in the model is introduced by assuming the Hertzian type of the contact force. The last model is a well known Rössler system, which constitutes a commonly used example of a mathematical model exhibiting chaotic behavior. All of these models have certain qualitative features similar to those typical of the gearbox vibration signal acquired during the experiment. The first feature is the Fourier spectrum possessing distinct peaks (corresponding to the harmonics of the meshing frequency for the experimental data) with modulation sidebands (Fig. 2.3). Secondly, dynamics of the underlying system has to be nonlinear and embedded in a relatively

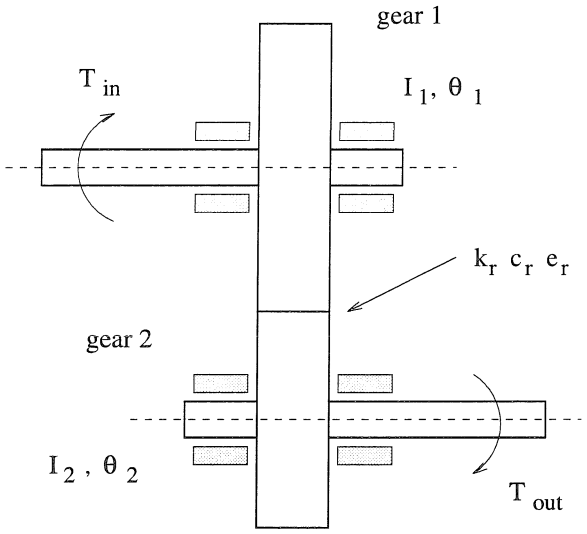


Figure 6.1: Schematic used to write equations of motion for the gear model (6.1).

low-dimensional phase space, $d_E \leq 5$. Next, at least one of the Lyapunov exponents computed for the time series obtained from the system has to be positive. Finally, the mutual information function has to exhibit features which are recurrent with the model equivalent of the meshing frequency.

6.1 Model with variable mesh stiffness and backlash

This model is based on a simple single degree of freedom torsional dynamic system with a harmonic excitation similar to that proposed by Sato et al. [46], who suggested occurrence of chaotically transitional phenomena in the gear system. The purpose of their model was to demonstrate the bifurcation scenario leading to chaotic behavior. Since in this work we are interested in a steady state chaotic response, reminiscent of that obtained from our experiment, we introduced two significant modifications to the model suggested in [46]. Firstly, we propose to use a cubic term in the expression representing the tooth stiffness. Secondly, we use a harmonic function which describes the variable mesh stiffness, $k(\tau)$. This choice seems to provide a better qualitative representation of the response of our experimental system.

The equations of motion for a single pair of gears, as schematically shown in Fig. 6.1,

in terms of their relative motion, can be written as

$$\ddot{\Theta} + 2\zeta\dot{\Theta} + k(\tau)g(\Theta + e_r(\tau), \eta) = \frac{T_1 + T_2 i^3}{1 + i^2} \tag{6.1}$$

where:

I_1, I_2 - moments of inertia of gears 1 and 2, respectively

i - gear ratio

I_c - equivalent moment of inertia of the gears in mesh, $I_c = 1/(\frac{1}{I_1} + \frac{1}{i^2 I_2})$

Θ_1, Θ_2 - angular gear displacements

T_{in} - input torque

T_{out} - output torque

k_r - variable mesh stiffness

θ_{st} - angular displacement of the meshing teeth, $\theta_{st} = \frac{\|T_{in}\|_{max}}{\|k_r\|_{max}}$

Θ - relative, normalized angular gear displacement, $\Theta = (\Theta_1 - \Theta_2/i)$

e_r - normalized transmission error

η - normalized backlash

c_r - damping between the meshing teeth

ζ - damping ratio, $\zeta = c_r/2\sqrt{I_c\|k_r\|_{max}}$

ω_n - natural frequency, $\omega_n = \sqrt{\|k_r\|_{max}/I_c}$

τ - dimensionless time, $\tau = \omega_n t$

T_1 - dimensionless input torque, $T_1 = T_{in}/\|T_{in}\|_{max}$

T_2 - dimensionless output torque, $T_2 = T_{out}/\|T_{in}\|_{max}$

ω - dimensionless meshing frequency

$k(t)$ - dimensionless variable mesh stiffness, $k(t) = 0.8 + 0.2\sin(\omega t)$

$g(x, \eta)$ - nonlinear function representing gear teeth backlash[1],

$$g(x,\eta) = \begin{cases} x + x^3, & \text{if } x \geq 0 \\ 0, & \text{if } -\eta < x < 0 \\ x + x^3 + \eta, & \text{if } x \leq -\eta \end{cases} \tag{6.2}$$

[1] One might be concerned that the backlash function $g(x, \eta)$ as defined in (6.2) suffers from a discontinuity. This problem can be alleviated by substituting $g(x, \eta) = x + \eta + (x + \eta)^3$ for $x \leq -\eta$. This change, however, does not result in qualitatively different behavior of the system for our values of parameters. Even though for all of the calculations performed for this model in Chapter 7, $g(x, \eta)$ was assumed in the form of (6.2), these results were also confirmed for the system with the continuous form of backlash function $g(x, \eta)$.

where $\| \bullet \|_{max}$ denotes a maximum norm, and all normalized values were obtained through dividing by the angular displacement of the meshing teeth θ_{st}. Assuming that the transmission error and the external torques act at an integer multiple of the meshing frequency, and substituting $x \equiv \Theta + e_r$, equation (6.1) can be written as

$$\frac{d^2x}{d\tau^2} + 2\zeta\frac{dx}{d\tau} + k(\tau)g(x,\eta) = B\cos(n\omega\tau + \theta) + B_0 \tag{6.3}$$

where:

x - relative angular displacement for meshing gears,

B - amplitude of harmonic excitation,

B_0 - static component of excitation.

We integrated the above system numerically, with the fixed time step of $\tau_s = \pi/40$ in the Runge-Kutta fourth-order algorithm, and used $x(t)$ to investigate the system (6.3). We found that the behavior of this model is richer than of that originally suggested in [46]. The response of the model changes from harmonic to poliharmonic, and finally chaotic, as the value of the backlash parameter η is varied. The chaotic response also exhibits various spectral characteristics, from possessing distinct, chaotically modulated peaks, to completely broadband with no dominating discrete frequency modes.

Fig. 6.2(a) shows the frequency spectrum of the signal obtained for the following values of parameters: $n = 1$, $\zeta = 0.01$, $\omega = 0.4$, $B_0 = 0.1$, $B = 4.0$, $\eta = 5.2$, and $\theta = 0$. We chose 60000 data points for each sample time series to ensure the stationarity of the probability density function and the stationarity of the eigenvalues computed for the dominant periodic orbit[2]. The signal was embedded in $\mathbb{R}^3$ using the time-lag corresponding to the first minimum, $\tau = 30\,[time\ steps]$, of the mutual information function, shown in Fig. 6.2(c). Notice the presence of the equidistant peaks, separated by the distance equal to the reciprocal of the meshing frequency, $T = 200[time\ steps]$.

The Lyapunov exponents computed for this model using the Brigg's method [8] are: $\lambda_1 = 0.381, \lambda_2 = 0.032, \lambda_3 = -0.388$ $[bits/time\ step]$. The value of λ_1 was also verified by fitting $y = 2^{ax+bx^2+c}$ to the curve illustrating sensitive dependence of the system dynamics to the initial conditions, Fig. 6.2(b).

[2]See Section 7.2.

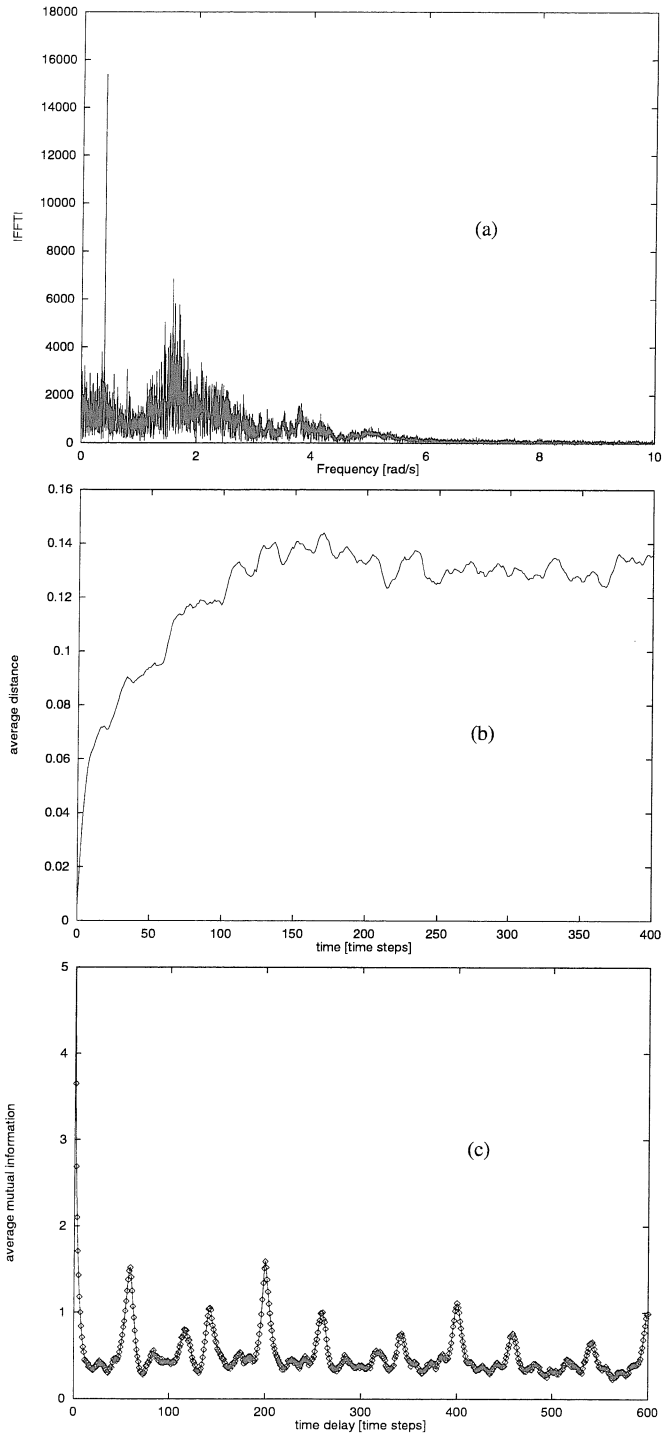


Figure 6.2: (a) FFT computed for the x signal, obtained by numerical integration of the model (6.3), (b) time evolution of the average distance between two initially close trajectory points in the phase space, (c) average mutual information function.

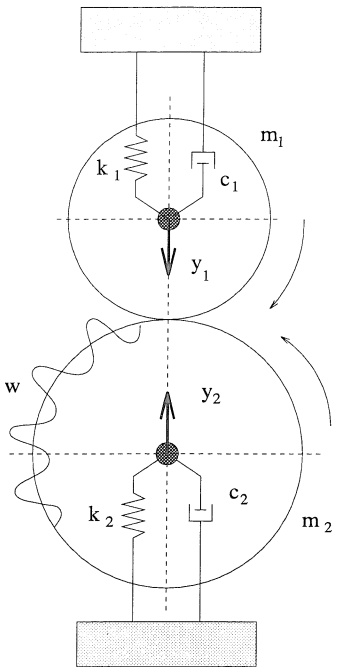


Figure 6.3: Two cylinders in rolling contact, subject to the Hertzian force.

6.2 Model with Hertzian force

Another important nonlinear phenomenon present in gearboxes is the contact force between the meshing gear teeth[3]. The model introduced in this section describes the dynamics of the rolling motion of a rigid cylinder driving another flexible cylinder. The equations of motion are a generalized version of those derived by Nayak [37].

Consider a rigid cylinder, with mass m_1, rolling on a flexible cylinder, with mass m_2, as in Fig. 6.3. The vibrations arise due to the surface waviness of one of the contacting cylinders. The other cylinder, m_1, is assumed to have a smooth surface. It is also assumed that the cylinders are precompressed to a static load P_0. The damping in the system is assumed to be of the viscous type and the contact force between the cylinders is assumed to be Hertzian, i.e.,

$$P_c = c_h(y_1 + y_2 + w)^{3/2} \tag{6.4}$$

[3]Apart from gearboxes, the problem of contact vibrations can be found in many other important engineering applications like ball bearings or wheel-rail contact problems.

where P_c is the contact force, w the surface waviness of the cylinder 2 (measured radially outwards), and c_h is the Hertzian material constant. The equations of motion for the two cylinders are given by

$$m_i\ddot{y}_i + c_i\dot{y}_i + k_i y_i = P_0 - P_c, \quad i = 1, 2 \tag{6.5}$$

Assuming that $c_1/m_1 = c_2/m_2 = c/m$, and $k_1/m_1 = k_2/m_2 = k/m$, where $m = m_1 m_2/(m_1 + m_2)$, the equations (6.5) can be combined, resulting in

$$m(\ddot{y}_1 + \ddot{y}_2) + c(\dot{y}_1 + \dot{y}_2) + k(y_1 + y_2) = P_0 - P_c \tag{6.6}$$

Introducing the Hertzian deflection $z = y_1 + y_2 + w$, equation (6.6) can be further transformed to

$$m\ddot{z} + c\dot{z} + kz + c_h z^{3/2} = P_0 + m\ddot{w} + c\dot{w} + kw \tag{6.7}$$

We will now convert equation (6.7) to its non-dimensional form by referring the displacement variable, z, to the static deflection $z_s = (P_0/c_h)^{2/3}$, defining this way a non-dimensional displacement $x = z/z_s$, and by using a non-dimensional time $\tau = \omega_c t$, where $\omega_c = (3P_0/2mz_s)^{1/2}$ is the contact resonant frequency. Furthermore, substituting $\xi = w/z_s$, $2\zeta = c/\omega_c m$, and $\alpha = k/\omega_c^2 m$ results in

$$x'' + 2\zeta x' + \alpha x + \frac{2}{3}[H(x)x^{3/2} - 1] = \xi'' + 2\zeta\xi' + \alpha\xi \tag{6.8}$$

where $H(x)$ is the Heaviside unit step function ($H(x \geq 0) = 1$ and 0 otherwise). If we approximate the surface waviness, w, of cylinder 2 by a harmonic wave,

$$\xi = \xi_0 \cos(\omega\tau) \tag{6.9}$$

the equation of motion takes its final form

$$x'' + 2\zeta x' + \alpha x + \frac{2}{3}[H(x)x^{3/2} - 1] = -\xi_0[(\omega^2 - \alpha)\cos(\omega\tau) + 2\zeta\omega\sin(\omega\tau)] \tag{6.10}$$

A special case of equation (6.10), with $\alpha = 0$, was investigated by Narayanan et al. [36], who found that for a certain set of parameter values the system undergoes bifurcations and exhibits chaos through a familiar period doubling scenario. We found that the behavior of the system (6.10) is a little richer, but essentially similar to that studied by Narayanan et al. In our work, we are interested in changes in system behavior as the parameter ξ_0 is varied, as opposed to varying frequency ω, as in [36]. Changes in ξ_0 can be thought of as a simple model for changes in gear teeth profiles, which are usually caused by normal wear mechanisms. If that kind of change lasts only for a relatively short time (short with respect to the main orbital period), it may model a local change in a tooth profile[4]. System (6.10) undergoes a series of bifurcations, for changing ξ_0, and varies from exhibiting a harmonic, to poliharmonic, and then chaotic response. The choice of $\alpha \neq 0$ seems to be more adequate for the purpose of qualitative modeling of the system response obtained from our experiment.

For our analysis, the model (6.10) was integrated numerically, using the fixed time step of ($\tau_s = \pi/40$) in the Runge-Kutta fourth-order algorithm. We chose 60,000 data points for each sample of $x(t)$ to ensure the stationarity of the probability density function and the stationarity of the eigenvalues computed for the dominant periodic orbit[5]. The following values of parameters were chosen for the forthcoming analysis: $\zeta = 0.05$, $\alpha = 0.045$, $\omega = 1.6$, and $\xi_0 = 1.1$. The frequency spectrum for the generated time series is shown in Fig. 6.4(a).

The chaotic character of the modulation sidebands, visible in the spectrum, was confirmed by quantifying the sensitivity of the system response to the initial conditions, Fig. 6.4 (b), and by the calculation of the Lyapunov exponents, using the Briggs' method [8]. Their values are $\lambda_1 = 0.198$, $\lambda_2 = -0.003$, and $\lambda_3 = -0.226$ [*bits/time step*]. The value of λ_1 was also verified by fitting the function $y = 2^{ax+bx^2+c}$ to the curve in Fig. 6.4(b).

Finally, the average mutual information function was computed for the system and plotted in Fig. 6.4(c). Its recurrent character is manifested in repetitive occurrence of significant peaks at the rate of $T = 50$ [*time steps*]. This recurrence rate corresponds to the period of the dominant cycle for the system ($T = 2\pi/(\omega\tau_s)$). The average mutual information function was also used to select the value of the time-delay for the purpose of phase space reconstruction. The time-lag corresponding to the first minimum of the mutual information function, $\tau = 10$ [*time steps*], was used to embed the time series in $\mathbb{R}^3$.

[4] A detailed discussion of the way in which we model tooth defects will be given in Section 6.4.

[5] See Section 7.2.

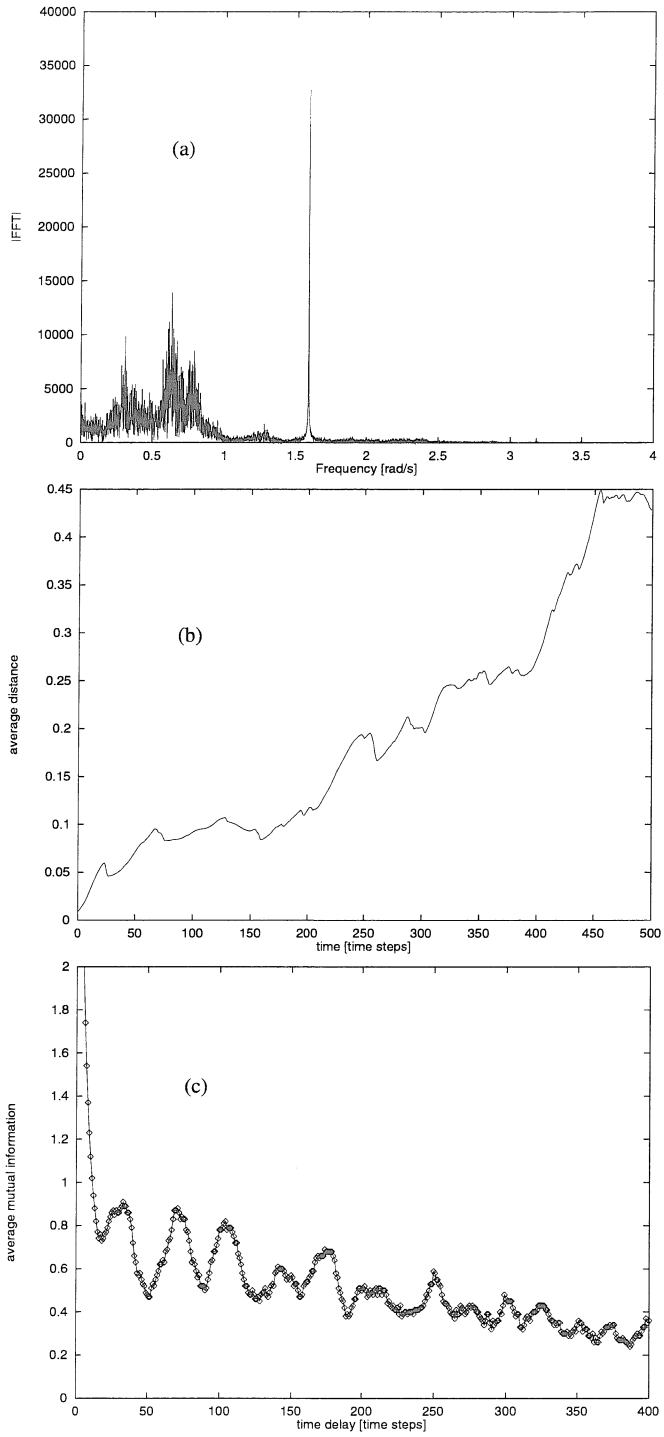


Figure 6.4: (a) FFT computed for the x signal, obtained by numerical integration of the Hertzian model (6.10), (b) time evolution of the average distance between two initially close trajectory points in the phase space, (c) average mutual information function.

6.3 Rössler model

At first sight, this model does not appear to have anything in common with gear dynamics. It was first introduced by Rössler in 1976 as a simple and well understood example of nonlinear dynamical system exhibiting chaotic behavior. It can be written as [45]

$$\dot{x} = -(y+z), \quad \dot{y} = x + ay, \quad \dot{z} = b + z(x-c) \tag{6.11}$$

where constants $a = 0.15$, $b = 0.2$, and $c = 10$ correspond to the chaotic range for the above system. The set of equations (6.11) is known to posses one positive Lyapunov exponent $\lambda_1 = 0.13\,[bits/s]$ (the other two are $\lambda_2 = 0.00$ and $\lambda_3 = -14.1\,[bits/s]$). This model was numerically integrated, using the Runge-Kutta fourth-order algorithm with the fixed time step of $\tau_s = 0.0304$, and the first coordinate, $x(t)$, of the solution vector was used as the time series. We chose $30,000$ data points for each sample in an ensemble. This choice was made again to ensure the stationarity of the probability density function and the eigenvalues estimated for the extracted dominant periodic orbit.

The response of this model is used here to illustrate applicability of our methodology to a general chaotic system with a broadband frequency spectrum, yet with dominant discrete features, as shown in Fig. 6.5(a). This spectrum is qualitatively similar to that obtained for an experimental gearbox signal (Fig. 2.3). The main spectral peak corresponds to the period-one orbit for the Rössler attractor, with period about $T = 200\,[time\ steps]$.

The signal was embedded in $\mathbb{R}^3$ using the time-lag corresponding to the first minimum of the average mutual information function, $\tau = 44\,[time steps]$, shown in Fig. 6.5(c). All of the above mentioned features, together with the recurrent character of the peaks in the mutual information function, qualify the Rössler system as a qualitative model for our analysis.

6.4 Modeling changes in operating conditions

On the basis of mathematical models described above, we will now introduce models of two specific kinds of defects or changes in an operating parameter of the system, which will be used in the next chapter to demonstrate the applicability of the condition monitoring methods proposed there. The first type of the modeled defects, called *type-I*, is modeled as a step function acting on one of the parameters and may be considered a qualitative

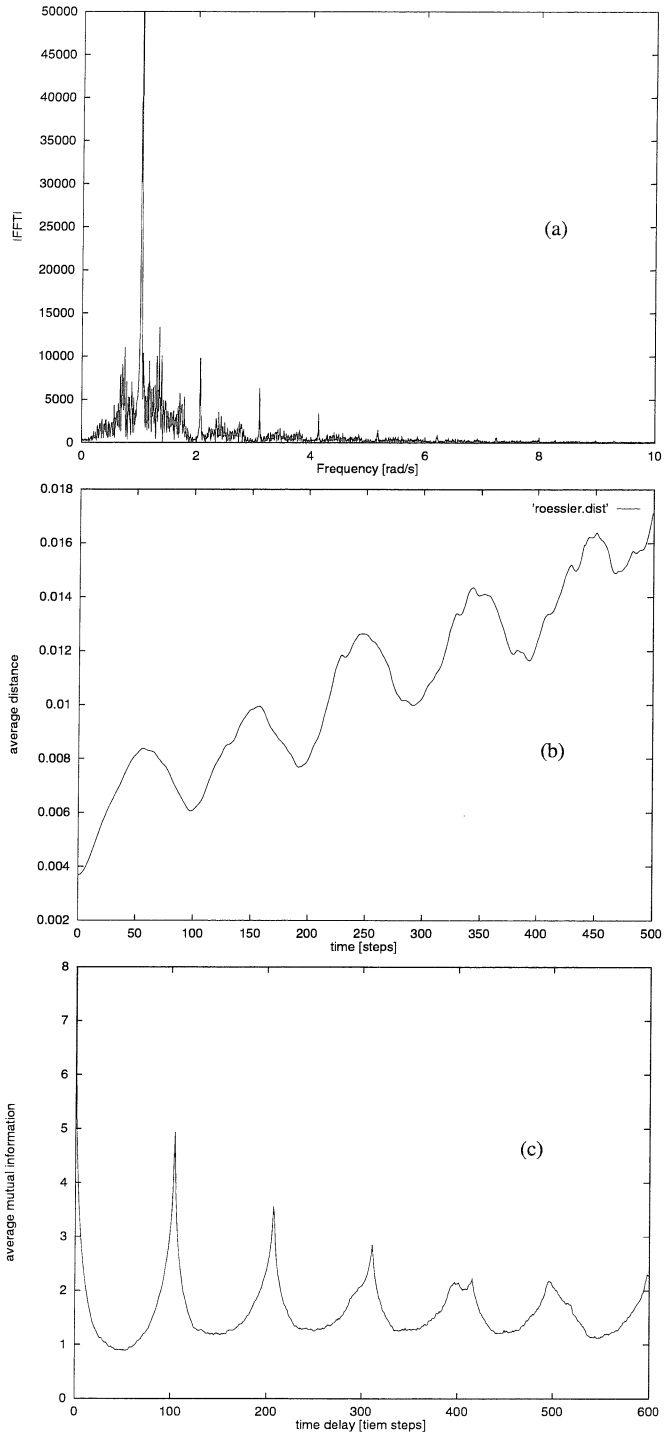


Figure 6.5: (a) FFT computed for the x signal, obtained by numerical integration of equation (6.11), (b) time evolution of the average distance between two initially close trajectory points in the phase space, (c) average mutual information function.

representation of "global" tooth defects which are evenly distributed over surfaces of all teeth and are caused by normal wear processes that are typical of kinematic mechanisms. We implement the *type-I* defect by integrating the analyzed system numerically from time t_0 to t_1, at which instant we change the value of the traced parameter, and then we integrate the "new" system from the time t_1 on.

For the model with backlash, we chose to monitor the parameter η, whose initial value of $\eta = 5.2$ defined the so called "good," or reference, operating state. The value of the parameter which defines the *type-I* condition for this model was chosen as $\eta = 5.6$. Similarly, for the Hertzian model (6.10) and the Rössler model (6.11), the "good" state was defined by the values of $\xi_0 = 1.0$ and $a = 0.15$, and the *type-I* state, by $\xi_0 = 1.1$ and $a = 0.16$, respectively.

The second, "local," type of change in the system operating conditions, denoted here by *type-II* state, is modeled as a short impulse acting periodically on a system parameter. It is numerically implemented in the following way. Suppose that the main orbital period for the analyzed system is equal to T. Two parameters need to be chosen: the impulse frequency of occurrence and its duration, τ_{II}. We take T as the impulse occurrence time and some small fraction of T as τ_{II}. While the impulse lasts, the system is integrated with one value of the traced parameter, while at any time before or after, the other value is used. If we think of the frequency of the impulse recurrence as the tooth meshing frequency in a gearbox, this type of the system change would correspond to local tooth defects, like a chipped tooth head or a crack at the tooth base.

For the model with backlash, we associated the recurrence time of the impulse with the meshing frequency $T = 200$ [*time steps*], and set its duration as $\tau_{II} = 10$ [*time steps*]. We took the same values for the Rössler model. The values of the corresponding parameters which were used during the impulse are: $\eta = 8.6$ for the model with backlash and $a = 0.18$ for the Rössler model. For the Hertzian model we used, respectively: $T = 50$, $\tau_{II} = 3$, and $\xi = 1.18$.

Obviously, detection of a change in the system state strongly depends on the values of parameters defining both "abnormal" stages. For example, the larger the value of τ_{II}, the easier the *type-II* condition will be detected. Notice, however, that this does not necessarily hold for the values of the observed system parameters. That is why a change of 10% in a parameter, for either *type-I* or *type-II*, might be harder to detect than a change of, say, 3%.

Therefore, analysis of nonlinear systems requires a different definition of "small" changes in the system operating conditions.

Changes in parameter values in both types of modeled defects investigated here are considered "small," by which we mean that simple measures like Fourier spectrum, auto-correlation function, phase space portrait, or one-dimensional probability density function are not sensitive enough to detect them. It does not mean that nonlinear methods of analysis would not work well where the "classical" linear methods can be successfully applied. It just seems difficult to justify application of often time consuming nonlinear methodology where faster and easier to implement linear methods, familiar to all engineers, can be effectively used. In order to verify that the analyzed data satisfies the above requirements, we compared the FFT for the "good" signals with those for the *type-I* and *type-II* signals and found them indistinguishable up to the value of statistical fluctuations in the ensemble. This observation can be made by comparing the Fourier spectra in Fig. 6.2(a) and Fig. 6.6(a)-(b) created for signals with both types of modeled errors obtained from the model with backlash. The same conclusions can be drawn from comparison of the spectra in Fig. 6.4(a) and Fig. 6.7(a)-(b) for the Hertzian model, and of the spectra in Fig. 6.5(a) and Fig. 6.8(a)-(b) for the Rössler model.

To further confirm the "smallness" of monitored changes, we computed the one-dimensional *pdf* signature for all analyzed cases. After the probability density functions were estimated, the alert zones were constructed for the ensemble of "good" signals. Each zone is centered at the median value of the estimated *pdf* and is $\pm 3\sigma$ wide, where σ denotes the standard deviation computed for a corresponding ensemble of reference data. This standard, global, statistic was estimated for all analyzed models and their respective parameter values. The result for the model with backlash is shown in Fig. 6.6(c)-(d) for the *type-I* and *type-II* signals respectively. Only one sample *pdf*, which is representative of the entire ensemble, is shown in each figure. The 3σ alert zone, shown between the solid lines, is not intersected by the *pdf* curve computed for the *type-I* signal and is only intersected in the tail[6] of the density function for the *type-II* signal. Therefore, we can conclude that no significant change in the state of the observed system was detected. Since the one-dimensional approach to the *pdf* estimation suffers from the loss of information due to disregarding the meaningful structure of the system in its state space, we expect that the methods which utilize the

[6]which is the region of *pdf* with the lowest significance for the purpose of condition monitoring

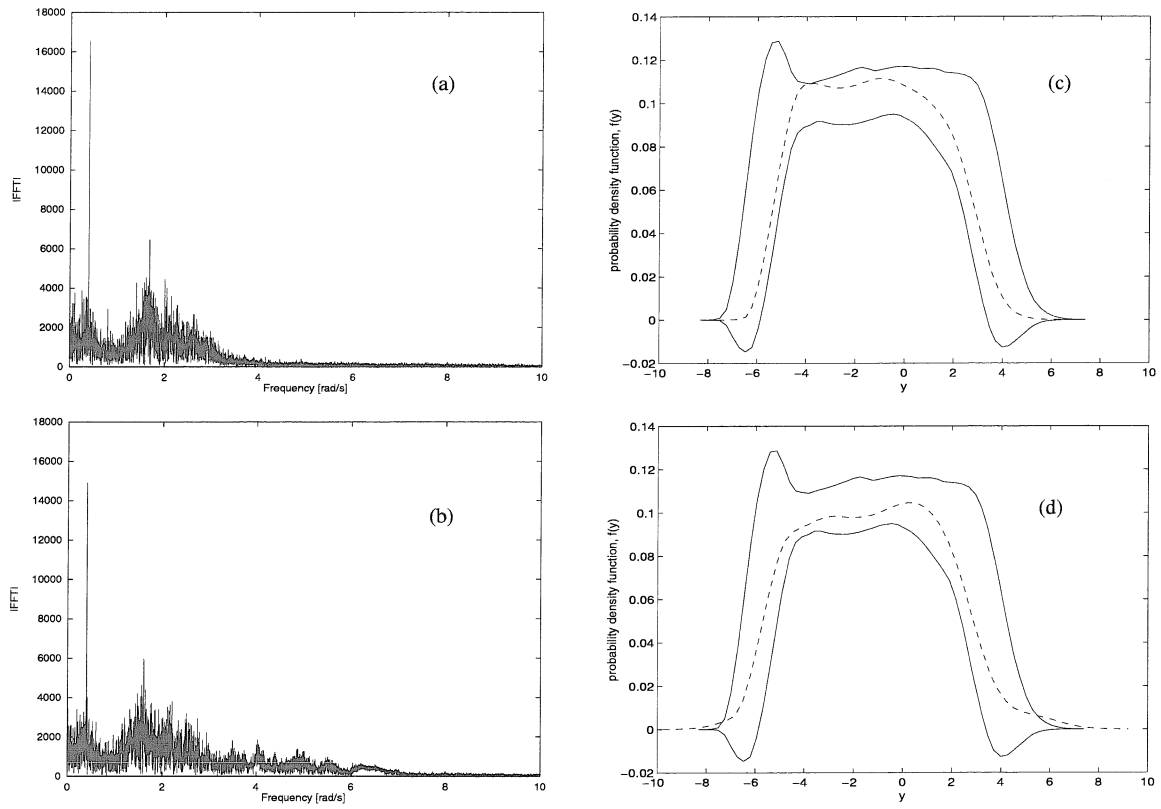


Figure 6.6: Most commonly used condition monitoring measures as computed for both types of defects in the model with backlash: (a) FFT for the *type-I* state, (b) FFT for the *type-II* state, (c) one-dimensional probability density function for the *type-I* state, and (d) one-dimensional probability density function for the *type-II* state. The 3σ alert zone, shown between solid lines, was created for the signal obtained for the reference value of the observed parameter $\eta = 5.2$. The dashed lines correspond to typical cases where the parameter value was changed according to the *type-I* or *type-II* model.

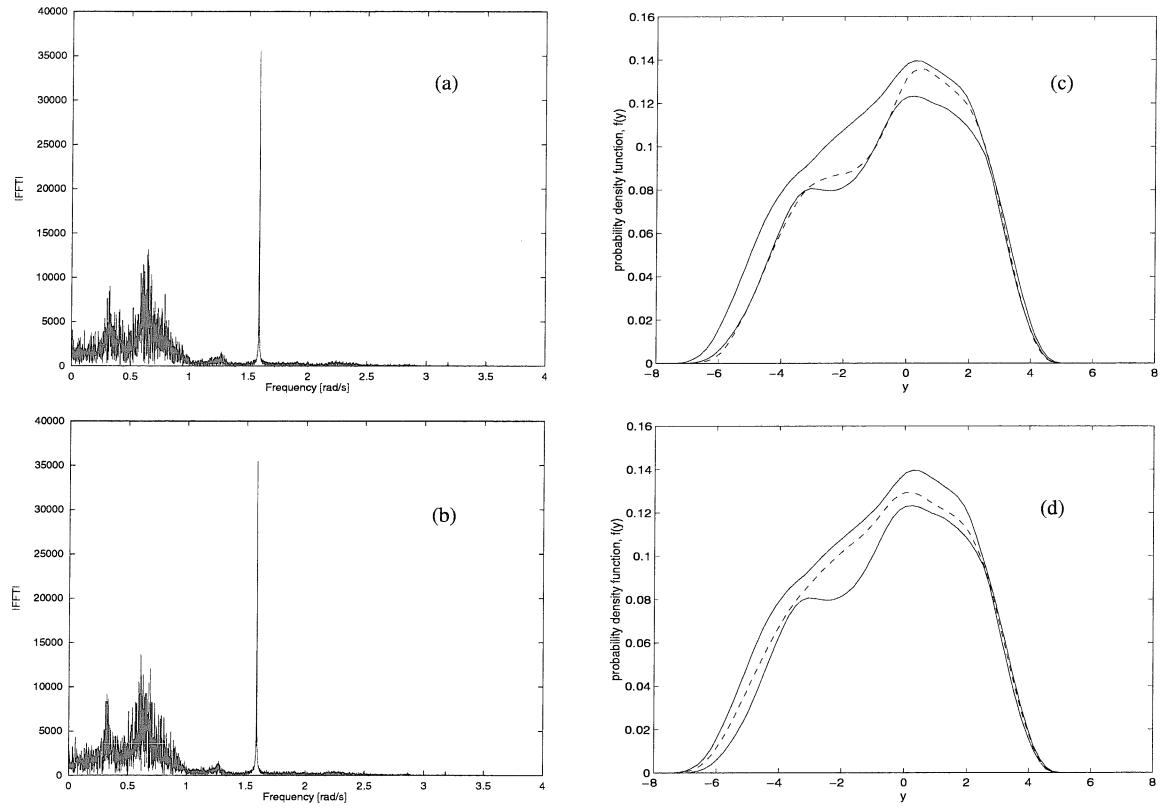


Figure 6.7: Most commonly used condition monitoring measures as computed for both types of defects in the model with Hertzian force: (a) FFT for the *type-I* state, (b) FFT for the *type-II* state, (c) one-dimensional probability density function for the *type-I* state, and (d) one-dimensional probability density function for the *type-II* state. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $\xi_0 = 1.0$.

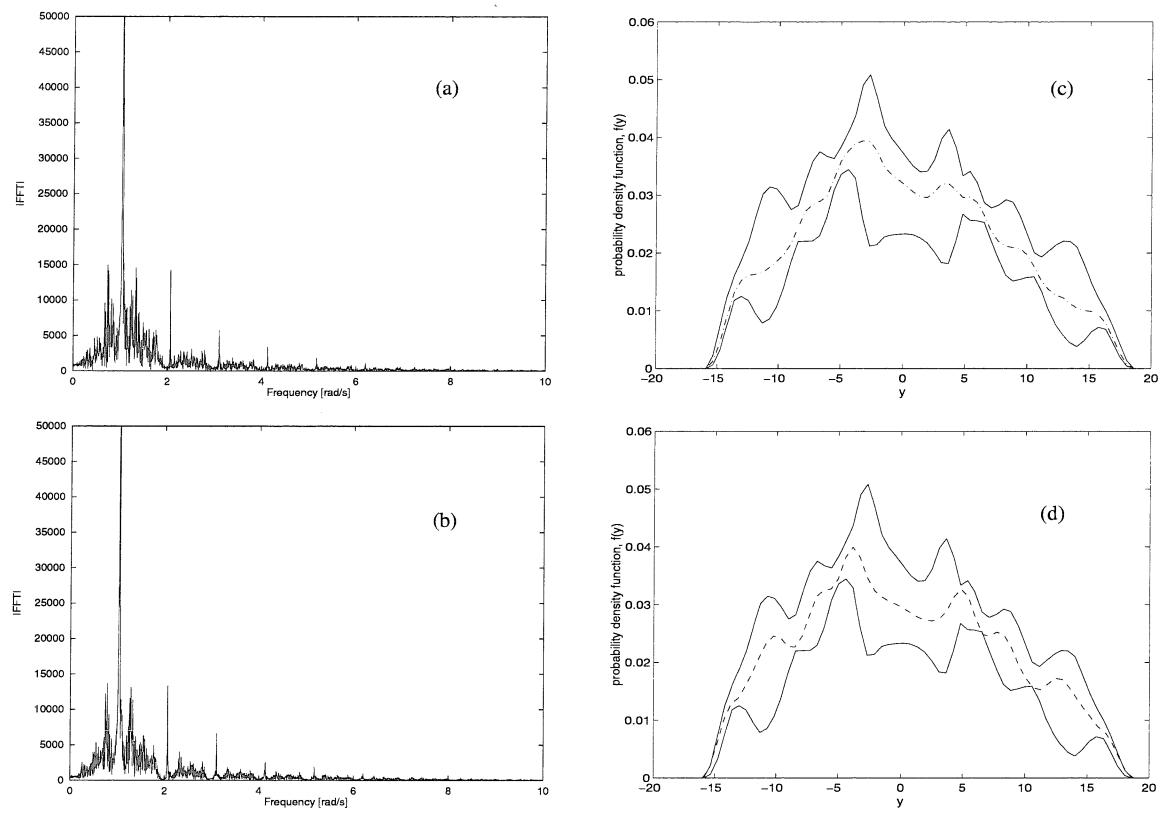


Figure 6.8: Most commonly used condition monitoring measures as computed for both types of defects in the Rössler model: (a) FFT for the *type-I* state, (b) FFT for the *type-II* state, (c) one-dimensional probability density function for the *type-I* state, and (d) one-dimensional probability density function for the *type-II* state. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$.

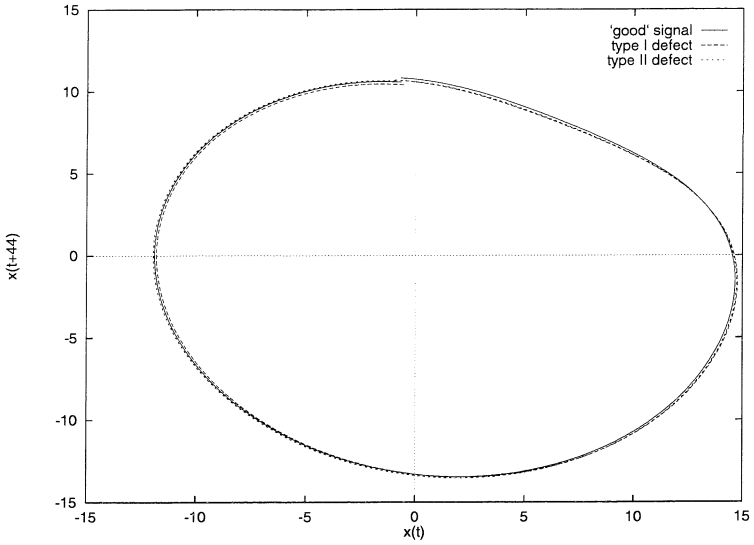


Figure 6.9: $T = 200$ periodic orbits extracted from the "good," *type-I*, and *type-II* signals obtained for the Rössler model.

phase space information (that will be proposed in the next chapter) should have a superior sensitivity than the standard one-dimensional *pdf* statistic.

The results of using the one-dimensional *pdf* signature to detect changes in the Hertzian model and the Rössler model are shown in Fig. 6.7(c)-(d) and Fig. 6.8(c)-(d), respectively. The same conclusions can be drawn from these figures as described for the model with backlash for the *type-I* and *type-II* signals.

We also reconstructed the phase space portraits and extracted the first three periodic orbits for all the analyzed signals. After comparing the resulting attracting sets and the cycles corresponding to all the cases of interest, we found these signatures are also not sensitive enough to detect the modeled defects imposed on the analyzed models. Fig. 6.9 illustrates how the extracted periodic orbits are almost identical for all the three cases of the Rössler signal. Similar conclusions were also made for the other two models.

To summarize, in this chapter we proposed a simple way to model local and global (or distributed) types of defects for three low-dimensional nonlinear dynamical systems, whose FFT, average mutual information function, and the exponential divergence of nearby orbits, qualitatively resemble those computed for the experimental data obtained from the gearbox test rig. Furthermore, we imposed a condition which limited the size of the modeled defects so that their detection was impossible using the standard, most commonly used techniques

of the system condition monitoring, i.e., FFT and the one-dimensional probability density function. This restriction was made not only to demonstrate the superiority of the proposed methods, but also to reflect many potential practical situations where the analyzed system is nonlinear and its behavior is manifested in the "noisy" broadband spectrum. In classical linear analysis of time series, especially if the broadband spectrum appears together with dominant discrete peaks, the common procedure is to treat the broadband component as "noise" and to filter it out. In the proposed approach, the broadband component is treated as an equally important source of information about the state of the system as the discrete component. Since there are many possible explanations of nonlinear features in the gearbox dynamics, we considered models with different types of nonlinearity. Even though they model physical phenomena in different ways, their response is qualitatively similar and can be used as a good testing ground for our purpose. In the next chapter, the detection of these defects will be used to judge the performance of dynamical signatures proposed there.

Chapter 7

Dynamical System Condition Monitoring

"Exitus acta probat"

Modern engineering puts very high demands on the quality of structures, devices, and machines, which have to operate with high accuracy and stability under heavy load and in a great variety of conditions. With constantly increasing precision and working speed of machines, an early detection of possibly catastrophic changes in their concealed operating parameters becomes a very important and difficult task. To achieve this, the alarm levels indicating a potential fault are set as close as possible to the "good" reference value of the measured diagnostic, corresponding to measurements taken when the machine was operating normally. There is an obvious trade-off between narrowing the alert zone and the number of indicated false alarms.

Various methods of technical diagnostics and monitoring are based on the assumption that the operating system is the best instrument for evaluating its own state, or in other words, that its indicators (diagnostics) can be observed and a correct diagnostic decision can be made based on these observations. To meet this assumption, the "correct" measurement location and the "correct" transducer need to be selected. These two requirements are often easily accommodated into the monitoring strategy. Another requirement is that the operating condition for the machine is stable during the measurement process. This sometimes can be more difficult to satisfy. A gearbox driving a pump, for example, may work in only one state of operation, defined by its rotational speed. An automobile gearbox, on the other hand, works at different running speeds and requires different alert zones for measurement comparison to ensure that faults are automatically detected at an early stage of development for all machine states and that this goal is achieved without triggering false alarms. Similarly, changes in the other operating conditions of the machine, such as pres-

sure, temperature, or flow of the lubricants, also affect its vibration characteristics. In this work, we assume that these changes in the operating conditions can be inferred using other monitoring systems, e.g., measuring temperature and the level of lubricants, monitoring the rotational speed of the input and the output gearbox shafts, etc.

There are many signal processing algorithms that are used to determine the condition of a gear pair in an operating machine. The most commonly used ones, however, are still based on spectral analysis. In this chapter, we propose three new methods for monitoring the state of a dynamical system. Instead of considering spectral features, we propose to focus on geometric and probabilistic features present in the system state space. Their main application is to an early detection of changes in system operating conditions, which correspond to various wear mechanisms, before these changes become catastrophic. This approach is expected to be most useful for dynamical systems that possess relatively broad-band spectrum with some dominant discrete peaks, like those usually observed in vibration signals acquired from gearboxes.

Data from the models introduced in Chapter 6 are used to investigate the effectiveness of the proposed methods of condition monitoring. All of them employ invariant features typical of nonlinear and/or chaotic dynamical systems, i.e., system predictability, stability of orbits, and multi-modal nature of the stationary probability density function estimated in the state space. The first method utilizes the fact that, even for chaotic systems, a short term predictability is always possible and the estimated predictor can be used to quantify changes in the system operating conditions. The other methods make use of periodic orbits which are extracted from the system trajectory in its re-created state space. The second proposed statistic uses stability of periodic orbits to characterize the system state and to monitor its change. Instability of orbits is one of the characteristic features of chaotic dynamics. It is usually quantified with Lyapunov exponents which are inherently difficult to compute for experimental data. Replacing their time-consuming estimation with calculation of eigenvalues corresponding to a set of chosen periodic orbits serves essentially the same goal, but is significantly more efficient computationally. The last method can essentially be classified as a new possible solution to the old statistical problem of projection pursuit. This problem is particularly interesting in the context of chaotic dynamics, whose trademark is a generically multi-modal probability distribution. The probability density function is estimated in the system state space, and then slices along chosen periodic orbits

are taken. The question whether this approach is superior to taking an arbitrary slice or to just evaluating the *pdf* for a one-dimensional time series and then projecting it to a one-dimensional orbit is also addressed for all investigated models.

Throughout this chapter, the decision whether the system state has changed or not is based on a simple statistical ensemble averaging. Every statistic is computed for an ensemble of samples corresponding to the same system state. The *alert zone* is created for the ensemble obtained for the system in the reference, "good," state. The zone is centered at the median value of the statistic and spans the area of $\pm 3\sigma$, where σ denotes the standard deviation of the statistic evaluated across the ensemble. If the statistic crosses the alert zone, it constitutes an alarm (true or false) which should correspond to a change in the analyzed system. From the persistence of such events, we determine the quality of the used condition monitoring tools and their applicability to the investigated type of system change.

7.1 Condition monitoring using system predictability

As already mentioned in Section 4.2, predictability of the system response is an important feature which characterizes its properties. A measure of system predictability can also be employed to identify the state of the system and to trace its possible changes [59]. Recall the *degree of predictability*, previously defined in equation (4.3),

$$D(T) = \frac{< y(t+T)\hat{y}(t+T) >}{\sqrt{< y^2(t+T) >< \hat{y}^2(t+T) >}} \tag{4.3}$$

Since $D(T)$ is such a valuable source of information about the system, it could be used directly, in principle, to identify the state of the system. One might, for example, define a diagnostic measure equal to the length of the prediction horizon corresponding to a certain value of D, say $D = 0.5$, and use that to monitor changes in the system state. Unfortunately, such a diagnostic tool is not sensitive enough to reliably detect small parameter changes (as they were defined earlier in this chapter).

Another monitoring method is also possible when we notice that there is a hidden conditional dependence of the degree of predictability on the available data $\mathcal{D}$, as it was denoted in equation (4.1). Having chosen a class of deterministic or probabilistic prediction models, $\mathcal{M}$, as in Section 4.2, we fit the model based on the available "training set" composed

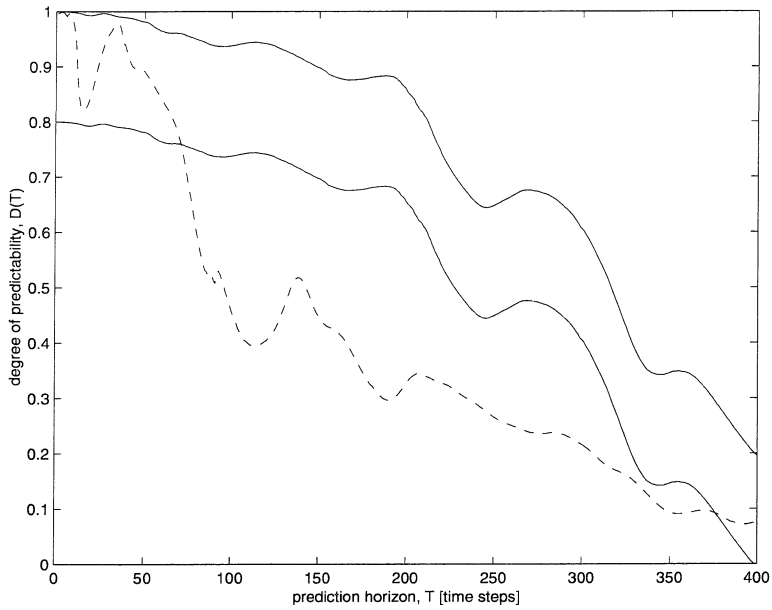


Figure 7.1: Change in the *degree of predictability* for two signals with a value of one parameter changed by a small amount. The measure is based on the predictor estimated for a training set obtained from the first set of operating conditions. The area between the solid lines defines the alert zone.

of the data acquired from the system in its reference state. The quality of the prediction made with such a model should be, at least in theory, as good for any sample time series as for the training data, provided the samples are large enough. Even small changes in a system parameter, however, should be manifested by a drop of predictability of the model. This idea is illustrated schematically in Fig. 7.1. It shows an alert zone which was created for the signal acquired from the "good" system. For the examples that follow, we employed 3σ alert zones, which were created for the ensembles consisting of 20 sample time series. The zones were centered at the median value of the degree of predictability, $D(T)$, computed for an ensemble corresponding to each value of the prediction horizon T. Every forecast was made using a *direct* prediction scheme, that was based on the same sample. The dashed line in Fig. 7.1 and in figures that follow in this section, corresponds to a typical result curve, obtained for the test runs. Since the prediction model is calculated only for the training data set, any change of the system's operating conditions is visible in a decrease in both the degree of predictability and the length of the possible prediction horizon. We expect to see the highest sensitivity of this method for the smaller values of the prediction horizon,

which is in agreement with the scenario described earlier in Fig. 4.3. This is why it is not always necessary to compute $D(T)$ until its value drops to zero, particularly if our goal is only to distinguish between different states of the system.

Rössler model

The geometry of the attractor for the Rössler system is simple enough to allow very good prediction of its response. It can be seen in the size of the 3σ alert zone in Fig. 7.2(a)-(b). Predictability of the samples for the "good" system is almost perfect, $D(0 < T \leq 1000) \approx 1$ which makes the width of the alert zone almost negligibly small. Fig. 7.2(a) shows the degree of predictability for a typical time signal of length $30,000$ data points, acquired from the Rössler model in its *type-I* state, as it was defined in Section 6.4. The first striking observation that can be made is how such a small change in the system parameter is clearly visible in the predictability of the observed signal. Another observation results from a very good short term predictability of the Rössler model – even though the quality of predictions made for the cases of changed system parameters is significantly lower than that for the same state of the system as in the training data set, the average quality of prediction is still relatively very high.

Altogether, we can conclude that both types of change in the system parameters are detected very clearly and reliably for this model.

Model with Hertzian force

This simple model also allows excellent predictions to be made, as revealed in $D(0 < T \leq 1000) \approx 1$, and as also seen in the very small size of the alert zone in Fig. 7.3. Notice that predictability of the model response in *type-I* and *type-II* states based on the training data for the reference state, although very good, is not as good as it was for the Rössler model. There is no significant difference in the degree of predictability between the cases shown in Fig. 7.3(a)-(b). It means that for this model, the proposed measure is equally effective for both distributed and local types of parameter changes. Both types of changes are detected very clearly and reliably. Since the detection of both types of changes in the state of the system was so clear, there was no need to investigate the behavior of the degree of predictability statistic for larger values of the prediction horizon, T.

This allows significant savings in time needed for computation and decision making.

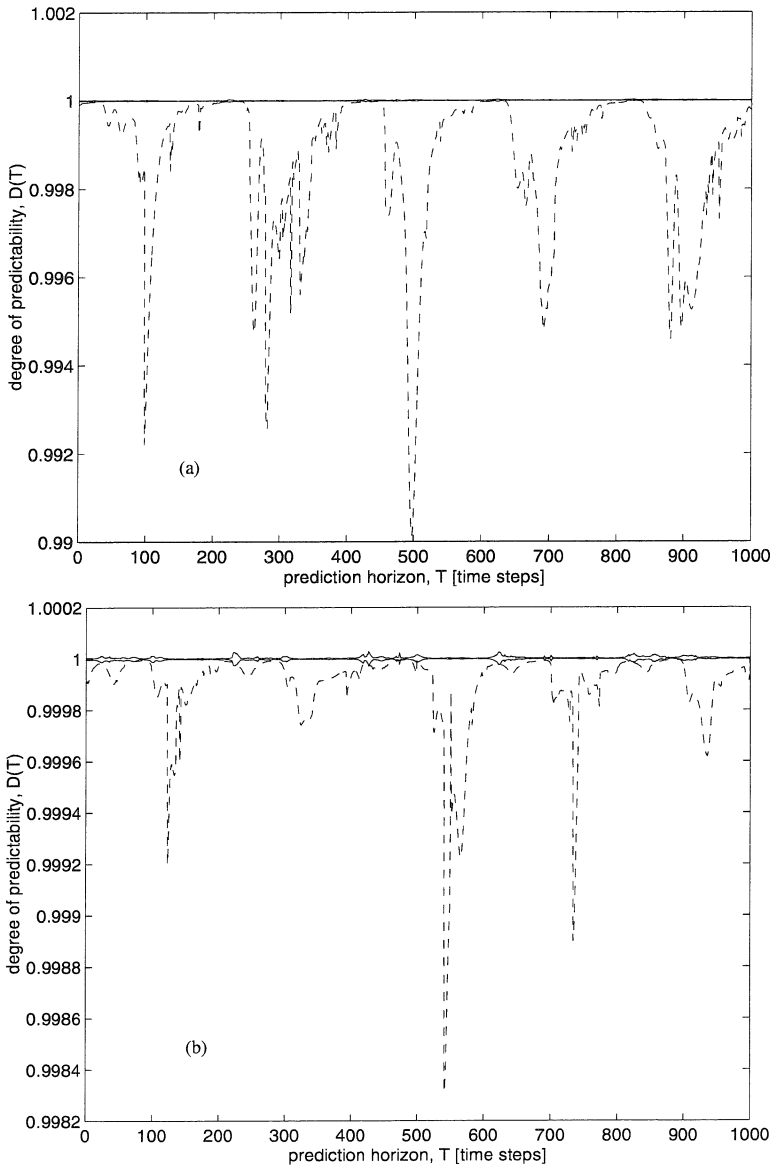


Figure 7.2: Change in the *degree of predictability* for the signals corresponding to the *type-I*, (a), and *type-II*, (b), states as defined for the Rössler model. The area between the solid lines corresponds to the 3σ alert zone.

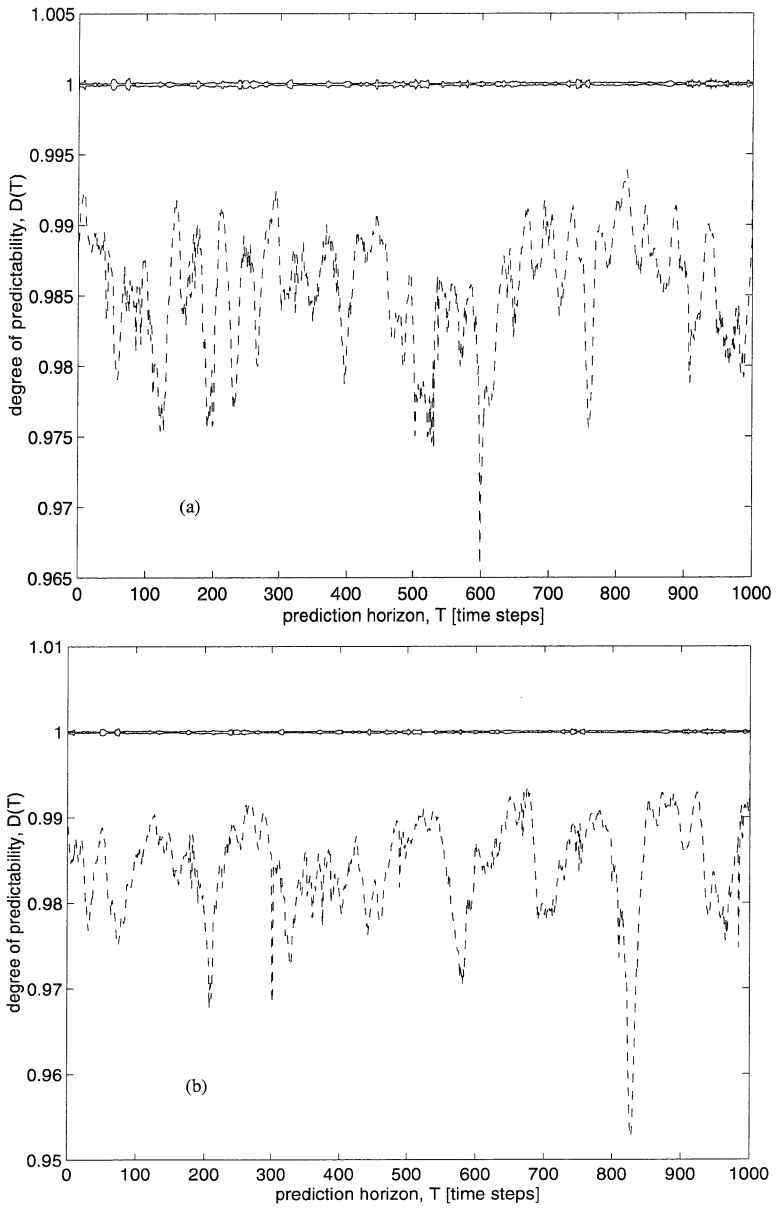


Figure 7.3: Change in the *degree of predictability* for the signals corresponding to the *type-I*, (a), and *type-II*, (b), states as defined for the model with Hertzian contact force. The area between the solid lines corresponds to the 3σ alert zone.

Model with backlash

One might ask: if this method is so very effective, why should we bother with any other condition monitoring methods? The reason is that this approach, as any other method, has its limitations. The main problem, which can be encountered when using system predictability as a condition monitoring tool, is a possible poor quality of prediction. It is usually caused by inadequacy of the prediction scheme used to model complex or highly contaminated signals. To produce Fig. 7.2-7.3, we used the local linear predictor defined in the system state space, as introduced in Section 4.2.3. We used the setup which we found to work well for most of the low-dimensional chaotic systems we investigated. For both systems, we initially embedded the trajectory in $\mathbb{R}^5$ using the time-delay corresponding to the first minimum of the average mutual information function. This reconstructed trajectory was then projected on $\mathbb{R}^3$ where the local linear predictors were estimated for every considered neighborhood. It is evident from the results obtained for both of the models analyzed above that this basic setup worked very well. To demonstrate that it does not always have to be true, we also used the same setup for the signals obtained from the model (6.3). Results of applying this procedure to detect the *type-I* and *type-II* states are shown in Fig. 7.4(a)-(b). The first striking conclusion which can be drawn from these figures is that the predictability of the analyzed system is very inconsistent. It is manifested by a very large alert zone (in some cases reaching even beyond the $[0, 1]$ interval) which makes it almost impossible to apply the predictability criterion to detect small changes in the system operating conditions. Notice that the inconsistency in predictability of the system response does not necessarily imply its poor predictability. A wide alert zone can be obtained even though the degree of predictability, $D(T)$, is higher than 0.5 for all of the samples in the investigated ensemble and throughout most of the prediction horizon. As a result, the degree of predictability signature was not sensitive enough to detect either type of modeled defects in this particular case.

As it has already been mentioned above, another case where this signature may not perform well is when data is highly contaminated with "noise." Since this method depends very strongly on accurate estimation of distances among neighbors in the phase space, it is very susceptible to contamination of data. Sometimes, one can alleviate this problem by using a better, more robust, prediction scheme, like the probabilistic scheme we proposed in

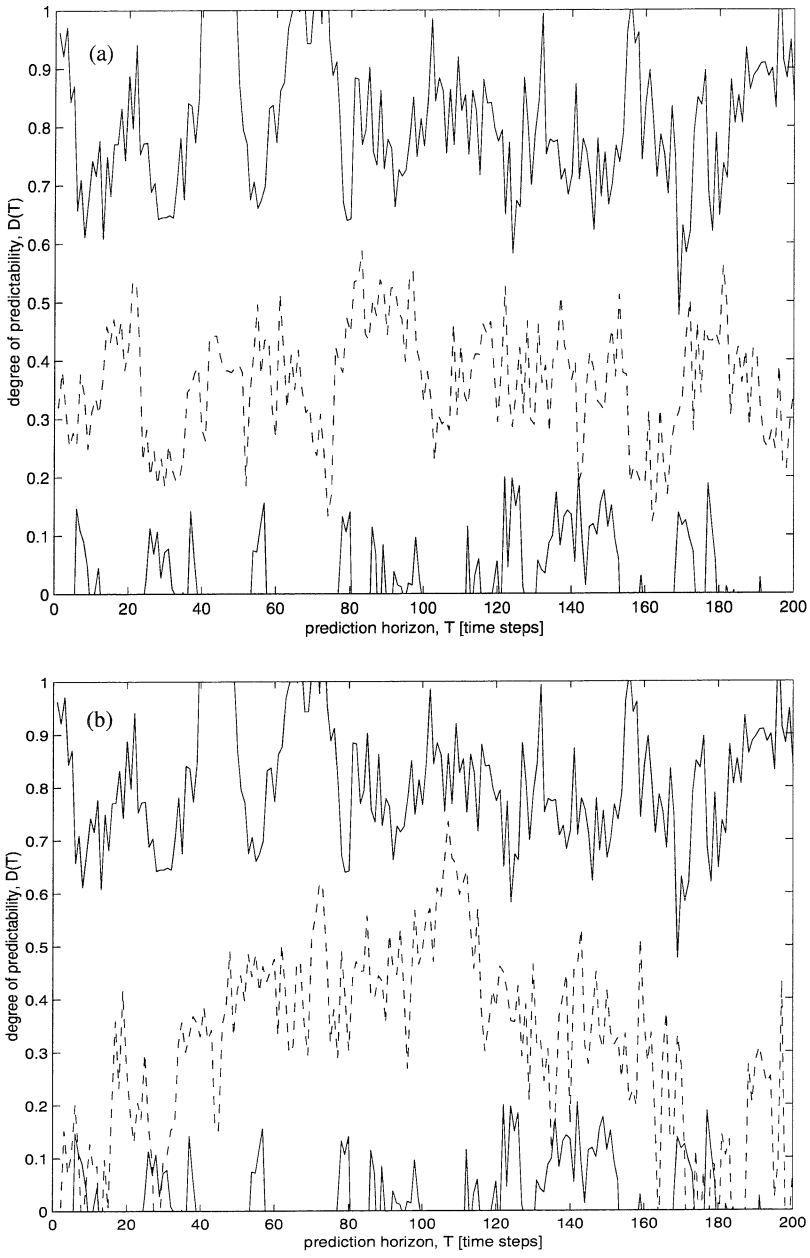


Figure 7.4: Change in the *degree of predictability* for the signals corresponding to the *type-I*, (a), and *type-II*, (b), states as defined for the model with backlash (6.3). The area between the solid lines corresponds to the 3σ alert zone.

Section 4.2.4. It is also possible that the analyzed system is strongly unpredictable, which is manifested in high divergence of nearby orbits and large positive Lyapunov exponent(s). In this case, the degree of predictability would drop very quickly, and the alert zone would be too wide (just like in Fig. 7.4), or positioned too low to allow any meaningful conclusions about changes in the state of the system. In the absence of "noise," however, the quality of the condition monitoring decisions that are made employing this statistic are not strongly influenced by the type of the predictor used, as in the cases of the model systems we studied. Using a statistical prediction scheme would not improve the quality of this method. In such cases, after realizing that the signature reached its regime of low sensitivity, one should attempt other possible methods of condition monitoring. Since the signatures proposed below do not use system predictability but employ other features of system dynamics, they constitute viable alternatives for the cases where the method proposed in this section might not work well.

7.2 Stability of periodic orbits as a diagnostic measure

We have already discussed why periodic orbits are an important tool in identifying system behavior. We also presented ways of extracting them from a given time series. We will now use their stability as a measure which characterizes the state of the system.

Consider a periodic orbit which has been extracted from the data, for example the one shown in Fig. 7.5. Condition monitoring of the dynamical system can be attempted by observing changes in the stability of its periodic orbits. The larger the number of orbits we observe, the more sensitive the monitoring process becomes. On the other hand not all of the system periodic orbits need to be considered since the topology of an attractor is usually well represented by only a relatively small number of them. We will now illustrate the monitoring procedure for one periodic orbit, $\{x(i)\}_j^{j+k}$.

Evolution of every trajectory point, $x(i)$ (and therefore of every point on a periodic orbit as well), is governed by a map, F, acting in the system phase space,

$$x(i+1) = F(x(i)) \tag{7.1}$$

Small perturbations to this orbit, $x(i) \rightarrow x(i) + \delta x(i)$, evolve according to the linearized

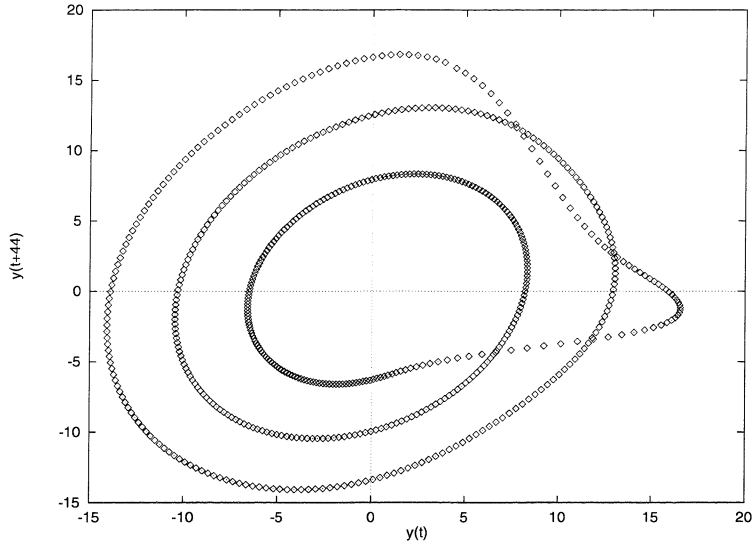


Figure 7.5: Periodic orbit (with period $T = 600$ [*time steps*]) extracted from a time series embedded in the reconstructed phase space and obtained by numerical integration of the Rössler system.

system

$$\delta x(i+1) = DF(x(i))\delta x(i) \tag{7.2}$$

where $DF(x(i))$ is the Jacobian matrix of the map F, evaluated at the point $x(i)$. Consider now the evolution of the perturbation throughout the length of the orbit $\{x(i)\}_j^{j+k}$,

$$\begin{aligned} \delta x(j+k) &= DF(x(j+k-1))DF(x(j+k-2))\ldots DF(x(j+1))DF(x(j))\delta x(j) \\ &= DF^k(x(j))\delta(x(j)) \end{aligned} \tag{7.3}$$

where $DF^k(x(j))$ denotes the product of k Jacobian matrices, starting at the point $x(j)$ on the periodic orbit.

Stability of the orbit is determined by computing the eigenvalues (which we denote $exp(\lambda_i)$) of the following real and symmetric matrix [39, 1]

$$[DF^k(x(j))^T DF^k(x(j))]^{1/(2k)} \tag{7.4}$$

In order to estimate the Jacobian matrix at a trajectory point, $x(i)$, the map $F(x(i))$

is expanded into a Taylor series in the vicinity of that point. We need not assume that the orbit is smooth in the state space. We must assume, however, that $F(x)$ is smooth. This allows us to compute the eigenvalues for periodic orbits which were estimated from an experimental trajectory. The Jacobian matrix is extracted from the linear term of the Taylor expansion. Since the numerical implementation of the procedure that we used in this work is taken from [1] (where it was used to compute local Lyapunov exponents), we will not describe it here in much detail. We start with the set of nearest neighbors found for every point on the orbit and then observe its time evolution, one time step ahead. This evolution defines a local neighborhood-to-neighborhood dynamics in the phase space, i.e., it corresponds to the map F. In order to make a Taylor expansion in the small deviations from the center of the neighborhood, the variables are centered at $x(i)$ and $x(i+1)$. A least squares fit is then used to determine the Taylor coefficients up to the second order terms in the expansion. Even though only first order terms contribute to the Jacobian matrix, the second (or even higher) order terms need to be computed to improve the quality of fit. For most of the practical situations we encountered, the improvement in the results from considering terms of the order higher than the second was so small that it did not justify a significant increase in the computation time. For a more detailed discussion of the Jacobian estimation and the application of the QR decomposition of the product of k matrices[1], see [1].

Since the spectrum of periodic orbits provides a measure of system dynamics and their stability is a system invariant, we will now monitor changes in the eigenvalues, corresponding to a set of extracted periodic orbits, in order to investigate their usefulness as a condition monitoring tool.

Rössler model

Since in all our calculations we refer to the trajectory representing the system in its reconstructed phase space, possible concerns may arise regarding the influence of the embedding procedure on the condition monitoring process. As long as we use features of the system which are invariant under smooth changes of coordinates, any change in the system parameters observed in the reconstructed state space represents a corresponding change in

[1]Direct multiplication of Jacobian matrices, even for a relatively short periodic orbit, usually results in a severely ill-conditioned matrix $DF^k(x)$.

the "real" system [48]. To demonstrate this, we perform condition monitoring using the $T = 200$ [*time steps*] periodic orbit extracted from a three-dimensional signal corresponding to the three coordinates, (x, y, z), of the Rössler model. Fig. 7.6 shows an absolute value of the largest eigenvalue, λ_1, as computed for two types of modeled defects. Fig. 7.6(a) corresponds to the *type-I* and Fig. 7.6(b) to the *type-II* defect model. The units on the horizontal axis correspond to the number of a subsequent time series sample, each $30,000$ points long, obtained from the model. The 3σ alert zone, shown between the solid straight lines, was created from samples of the signal which proceeded those shown in the figure, and were obtained for the parameter value $a = 0.15$ (corresponding to the "good" state). The zone is centered at the median value of λ_1. After the 20-th sample was taken, the parameter value in the model was changed according to the *type-I* (Fig. 7.6(a)) or *type-II* (Fig. 7.6(a)) of modeled system change. This change is reflected very well in the value of λ_1 leaving and remaining outside of the alert zone for all of the subsequent samples in Fig. 7.6(a). Although there seem to be some clear signs of a change in the system parameters visible in Fig. 7.6(b), they are not as clear as for the *type-I* case.

One can observe that the mean behavior of the largest eigenvalue is quite different from its counterpart before the 20-th sample time series was acquired. This ambiguity may have at least three possible causes. The first cause may be an insufficient length of the analyzed sample time series. The easiest way to verify that is to increase the length of the sample and compare the results. We doubled the size of each sample used to produce Fig. 7.6 and found that the resulting variations in λ_1 were not significantly smaller. From this observation, we concluded that the sample size of $30,000$ data points is sufficiently large for our analysis. The second possible reason might be a poor approximation of the periodic orbit, which commonly results in large variations in eigenvalues across the ensemble. This suspicion seems to be supported by an occurrence of a false alarm at the 17-th sample time series. We will return to the problem of treating such "outliers" later in this section. In this case, one can try to use different trajectory ensembles or more efficient methods for extraction of periodic orbits. An inability to approximate periodic orbits faithfully is the main limitation for this method. The third reason for an ambiguous monitoring result may be that the change in the system occurred in the part of the state space which was not visited (or was only locally visited) by the considered orbit, and hence remains difficult to detect. In such a case, one should verify the suspected change in the system by computing the same statistic

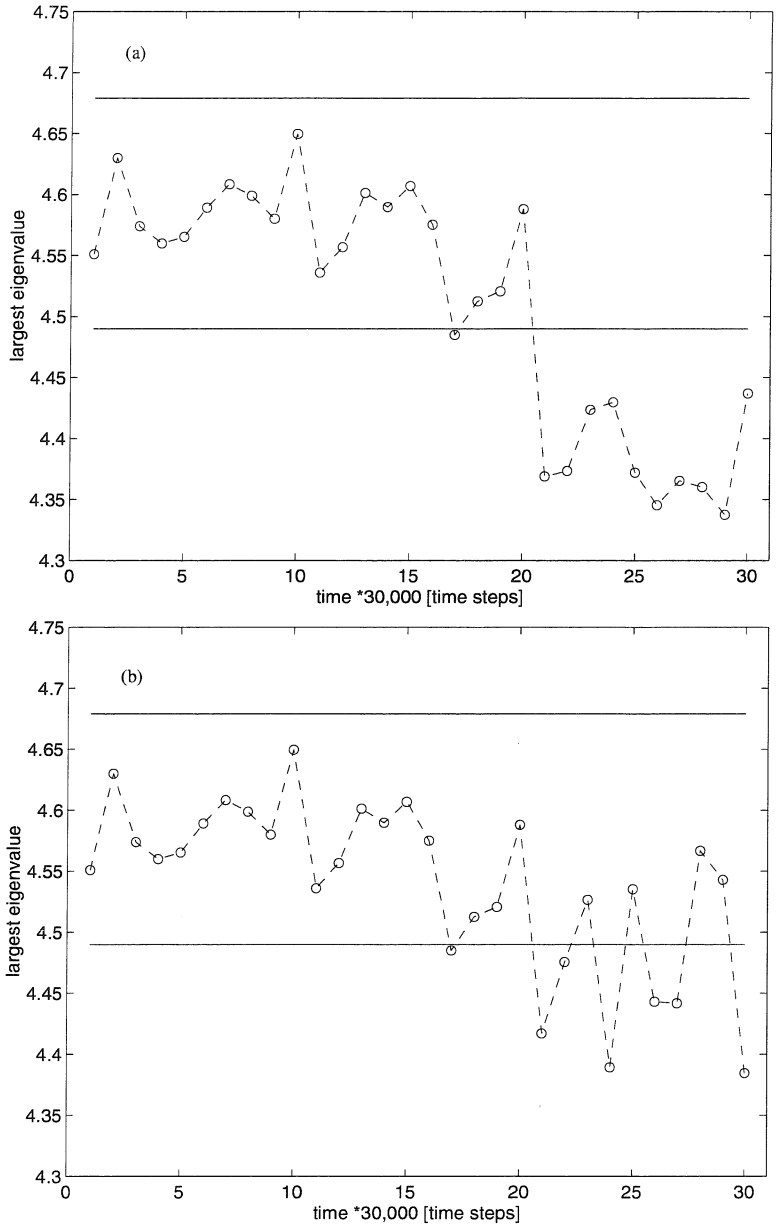


Figure 7.6: Absolute value of the largest eigenvalue computed for the period $T = 200$ [*time steps*] orbit extracted from the x, y, z coordinates of the Rössler system. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$ at times before those shown. (a) corresponds to the *type-I*, and (b) to the *type-II* model defect.

for different periodic orbits or by using other condition monitoring method(s) (for example the one suggested in Section 7.1) for the controversial sample.

Let us now return to the analysis of one-dimensional signals. We acquired our one-dimensional time series from the Rössler model, reconstructed the phase space, extracted periodic orbits and again computed their eigenvalues. Since a faithful approximation of periodic orbits is of paramount importance for this approach to condition monitoring, as we already mentioned in Section 3.2.2, we usually use two methods of their extraction. The simpler method looks for an orbit whose starting and ending points are closest in the phase space, for a given period T. Fig. 7.7(a) shows the largest-eigenvalue-statistic for the period $T = 200$ orbit extracted this way for the system in "good" and *type-I* states, as in Fig. 7.6. Fig. 7.7(b) shows the same statistic computed for the orbit of the same length but which is constructed by averaging the nearby orbits, as described in Section 3.2.2. From comparing these two figures, it is not immediately clear which method of extraction is superior in this case. Although the difference in the statistic is much larger for the simpler extraction method, it also suffers from producing a false alarm. A way to verify that this alarm is indeed false is to notice that λ_1 leaves the alert zone for only one, 16-th, sample, and then returns back for the next few samples. Since no such behavior for this sample is visible in Fig. 7.7(b), and because of a temporary character of this change, we can conclude that this event is caused by an imperfect estimation of the periodic orbit and therefore constitutes a false alarm. To support this conclusion, we also computed the orbital projection of the probability density function statistic for the sample in question, as described in the next section. As expected, it showed no sign of change in the system parameters. Obviously, this simple check does not guarantee that the observed event constitutes a false error. Together with the previous observations, however, it gives an observer a higher confidence in his decision making. Since one of the first priorities of any condition monitoring scheme is to minimize the number of false alarms and since they seemed to occur more often for the method used to produce Fig. 7.7(a), in all cases which follow and which required periodic orbits, we used the method of averaging as the method of their extraction.

As we already observed in Fig. 7.6(b), the local type of defects, modeled in our work as the *type-II* state, is generally more difficult to detect than the distributed, *type-I*, state. This observation can be confirmed by computing the orbital stability statistic for a one-dimensional time series for the *type-II* state, as it was done above for the *type-I* state. In

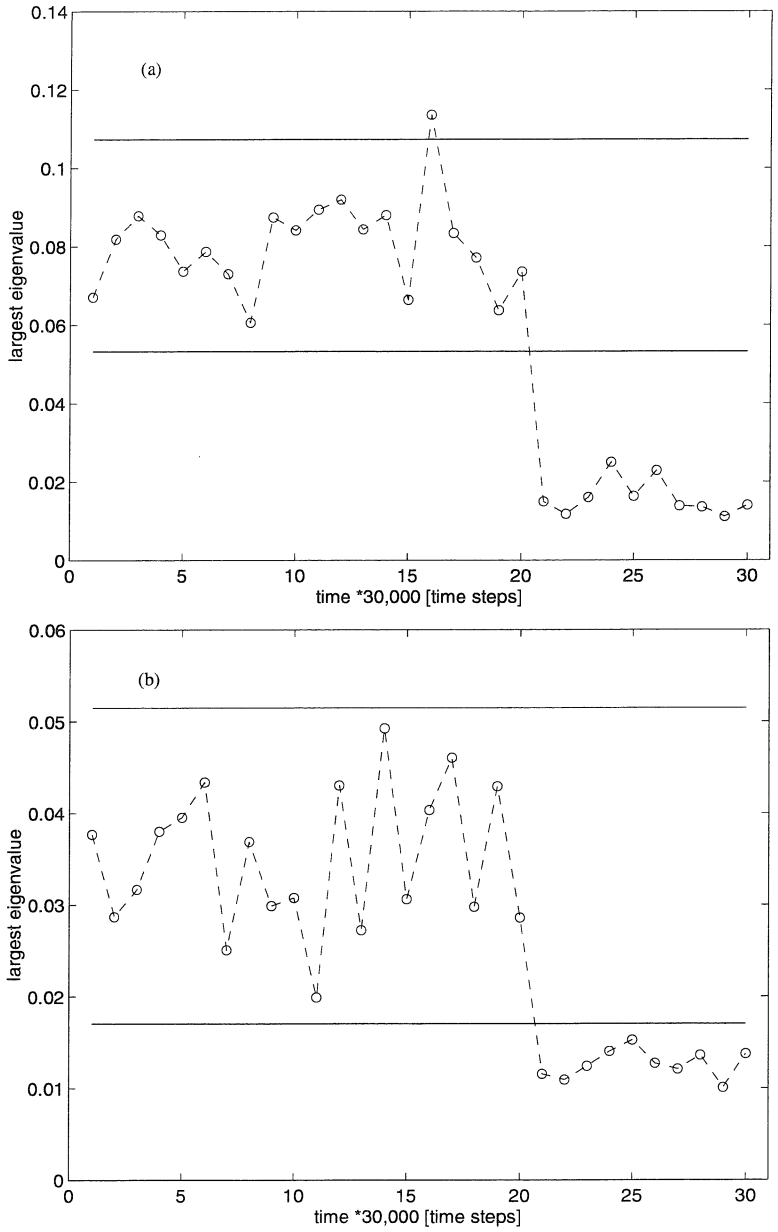


Figure 7.7: The largest eigenvalue, λ_1, computed for the period $T = 200$ [*time steps*] orbit extracted, as an orbit closest to being periodic (a), or the average of all nearby and almost periodic orbits (b), for the Rössler system. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$ at times before those shown.

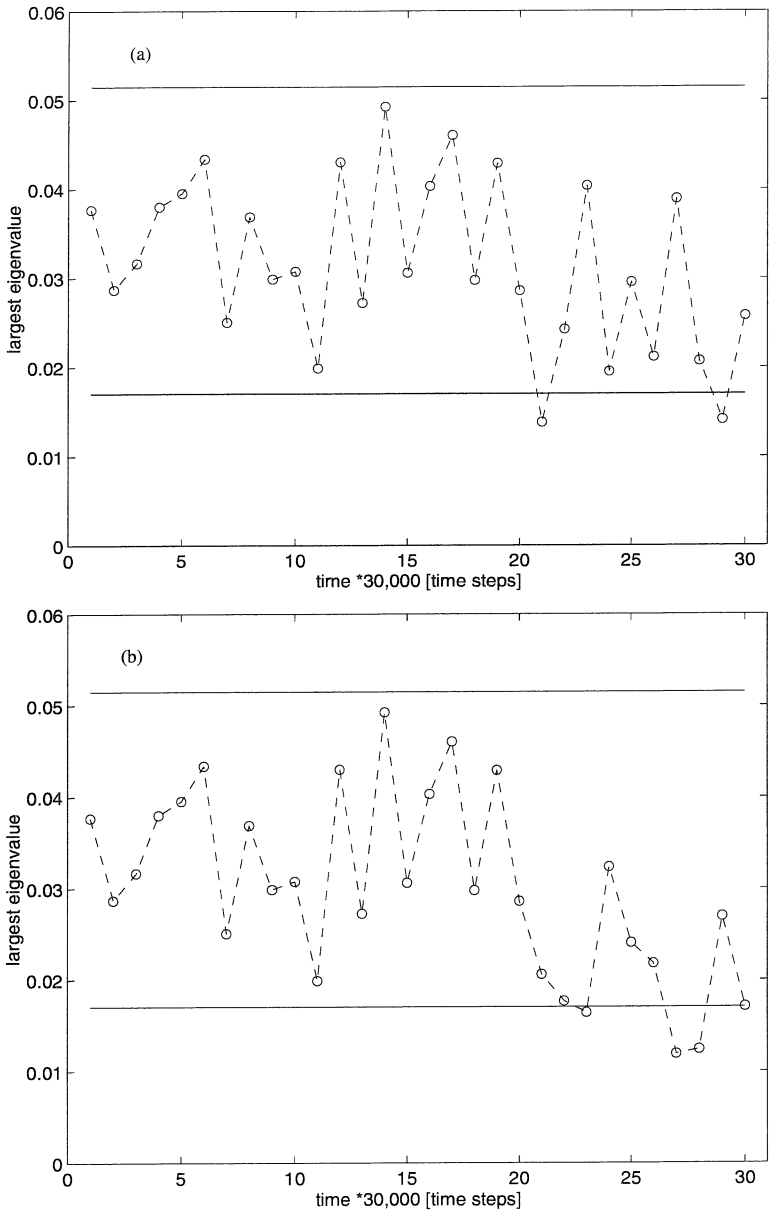


Figure 7.8: The largest eigenvalue, λ_1, computed for the period $T = 200$ [*time steps*] orbit extracted from the Rössler system, for the *type-II* model defect of the same duration but different "strength": (a) $a = 0.16$, (b) $a = 0.3$. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$ at times before those shown.

Fig. 7.8(a), we can see again large variations in the values of λ_1. As expected, the local, short lasting changes in system parameters depend very strongly on the quality of the estimation of periodic orbits. Observability of local defects depends on both their time duration and their "strength." In analogy to a local defect of a gear tooth surface, we could think of the defect duration as corresponding to the ratio of the area of the worn tooth surface to the area of the entire tooth surface in contact. "Strength" of the defect can be thought of as corresponding to its depth. When the defect "grows," it should be more visible, at least in principle. This simple effect can be seen in Fig. 7.8(b).

Until now we have analyzed only a single, $T = 200$ [*time steps*] periodic orbit. Let us consider a couple more orbits. We estimated eigenvalues for the orbit $T = 600$ [*time steps*] using the same ensembles as before. The distribution of the largest eigenvalue is shown in Fig. 7.9(a). Here again we see a persistent change in the stability of the orbit, occurring after the 20-th sample (after which the *type-I* state was inflicted), confirming good sensitivity of this measure and its applicability to this case of the system condition monitoring.

As one can suspect, not all periodic orbits are equally susceptible to changes in system operating conditions. If a change in the system dynamics is observable in a region of the phase space which is not visited by the periodic orbit of interest, this change cannot be detected by any measure computed for this orbit. Also, if a considered orbit is very long and the part of the phase space affected by the change is explored for a short time (relative to the length of the orbit), the averaging process used to estimate eigenvalues may cause such an orbit to provide a poor diagnostic signature. Such a case is shown in Fig. 7.9(b), created for the orbit with the $T = 1000$ [*time steps*] period. The signature in this case is not statistically clear and creates an illusion of a false alarm in a place where the actual change in the system occurred.

This raises a question of how many orbits one should observe. The answer is twofold. In the case when no changes are detected, one should monitor all significant orbits, where by "significant" we mean the periodic orbits which significantly contribute to the orbital decomposition of the system dynamics in its phase space[2]. However, in the case of detecting alarms, like these shown in Fig. 7.7-7.9, we do not need to go any further to conclude that a change in the operating conditions did occur.

[2]See Section 3.2.2.

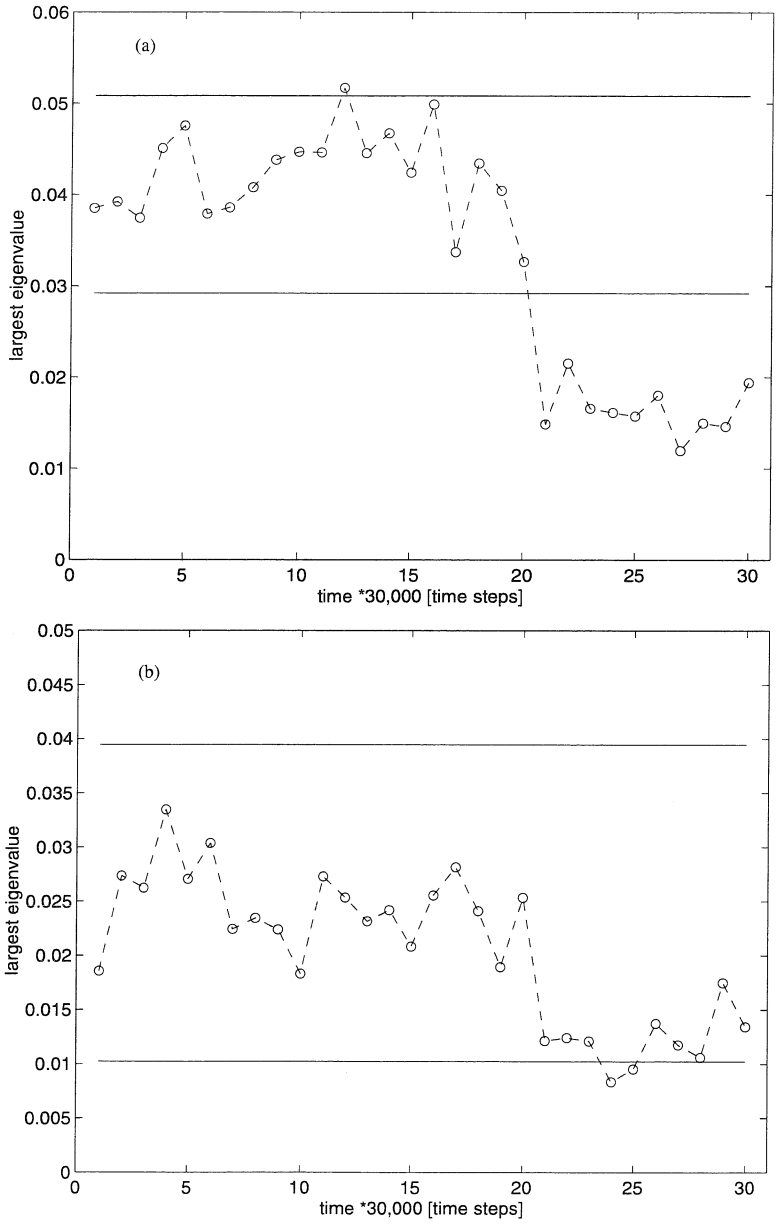


Figure 7.9: The largest eigenvalue, λ_1, computed for the period $T = 600$ [*time steps*], (a), and $T = 1000$ [*time steps*], (b) orbit extracted from the Rössler system. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$ at times before those shown.

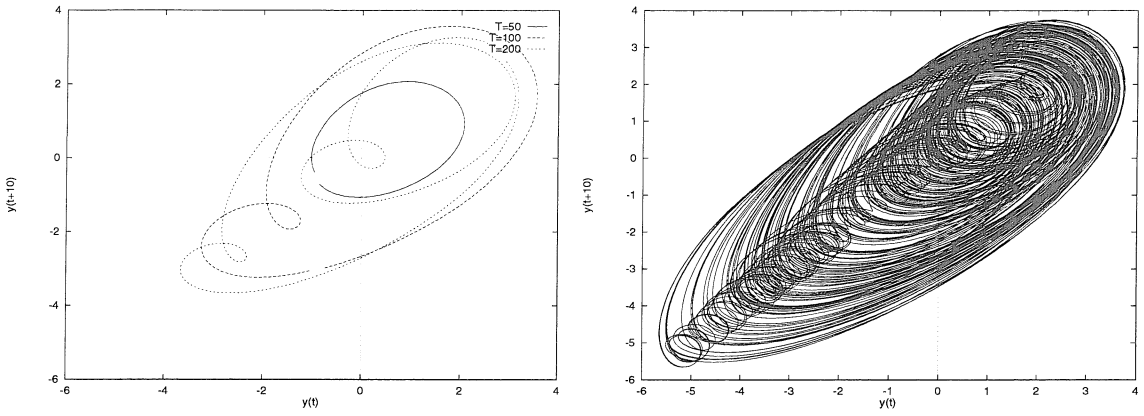


Figure 7.10: (a) Periodic orbits ($T = 50, 100$, and 200 [*time steps*]) extracted from the signal acquired from the model with Hertzian force. (b) Reconstructed phase space portrait, projected on $\mathbb{R}^2$.

Model with Hertzian force

For this model, the proposed signature will be illustrated for three periodic orbits, with periods $T = 50, 100$, and 200 [*time steps*]. These orbits are displayed in Fig. 7.10(a) next to the phase space portrait reconstructed for the system and shown in Fig. 7.10(b). The λ_1-signature computed for *type-I* and *type-II* states for the periodic orbit corresponding to the frequency of the harmonic excitation, i.e., $T = 50$ [*time steps*], is displayed in Fig. 7.11(a) and Fig. 7.11(b), respectively. There are two alert zones visible in both figures. Depending on the quality of the approximation of the periodic orbit, one can attempt to use the narrower, 2σ, alert zone. This choice can only be made, however, if no false alarms were generated for that zone during the test run. By a test run we understand the ensembles generated for the system in the steady state after the alert zone was established. In Fig. 7.11, the test run corresponds to the points up to the 20-th sample. Recall that the alert zones are always created using ensembles preceding those shown in all the figures. For example, the ensemble used to create the alert zones in Fig. 7.11 consisted of 20 samples, each 60,000 points long, which were acquired before the first shown eigenvalue point was computed. Since the variation in the value of λ_1 is relatively small for this orbit and no false alarms were indicated, we can try using the 2σ alert zone. Fig. 7.11(b) indicates good detection of the *type-II* state for the 2σ alert zone, and a statistically little less significant result for the 3σ zone.

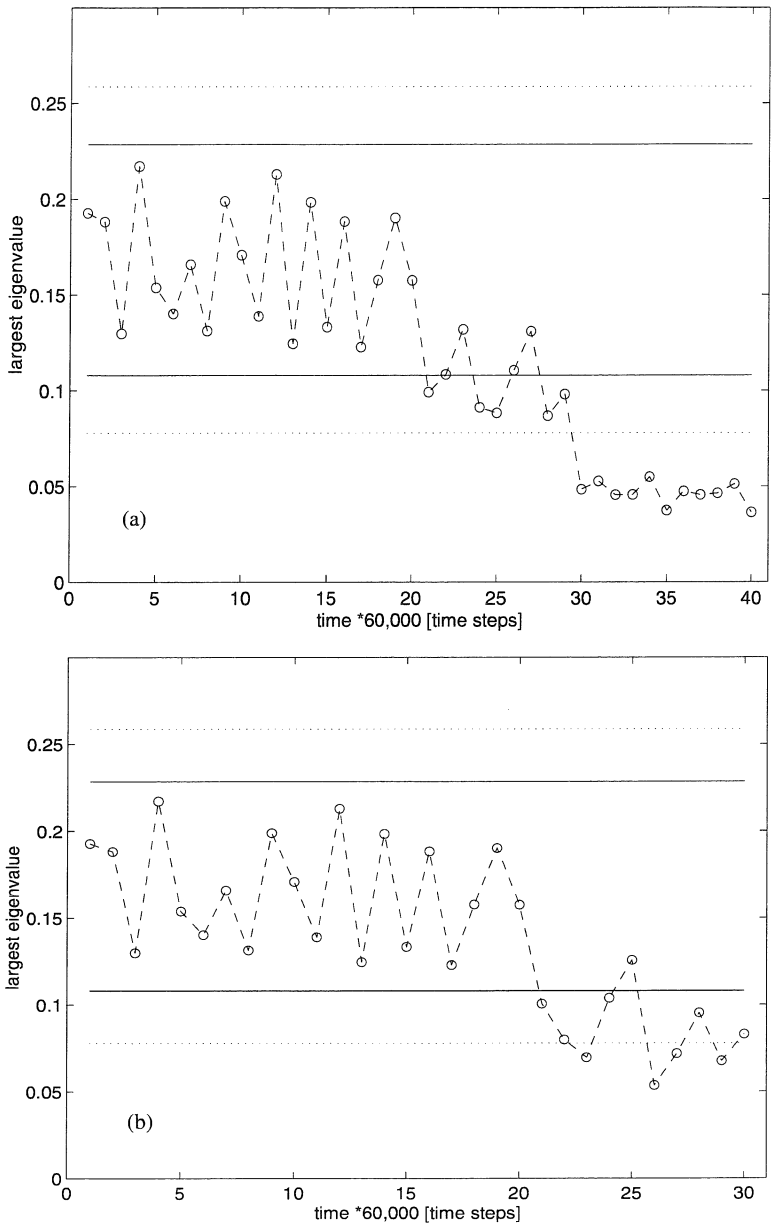


Figure 7.11: The largest eigenvalue, λ_1, computed for the period $T = 50$ orbit extracted from two different *type-I*, (a), and one *type-II*, (b), signals obtained from the model with Hertzian forces. The solid lines denote the 2σ alert zone, and the point lines the 3σ zone.

Fig. 7.11(a) shows the λ_1-signature for two ensembles corresponding to the *type-I* state, but generated with a different value of the parameter ξ_0. The samples from the interval $(20, 30]$ were generated for $\xi_0 = 1.1$ and the following samples for $\xi_0 = 1.18$, which can be interpreted as a "more significant" defect. The interpretation of the result in Fig. 7.11(a) is almost trivial: the larger the cause, the more significant the effect. In the language of this statistic, it means that the increase in value of ξ_0 results in decrease of instability of the employed periodic orbit. It is not hard to imagine cases where this relationship is not true. Whether stability of the analyzed orbit increases or decreases depends on the location of the parameter value in the system bifurcation tree.

A different scenario can be observed in Fig. 7.12(a) created for the period $T = 100$ [*time steps*] cycle. Here, the periodic orbit, after an increase in stability for $\xi_0 = 1.1$, undergoes an apparent increase in instability as the observed parameter is changed to $\xi_0 = 1.18$. Since the variation in the value of λ_1 is relatively large, only the 3σ (or possibly larger) alert zone can be used for this orbit. Another interesting conclusion from this figure is that the smaller change in the parameter is detected less ambiguously than the bigger one. This is not a surprising result for nonlinear systems. Fig. 7.12(b) is a good example of how a longer orbit can reveal more structure in data and as a result provide a better condition monitoring signature than a shorter orbit. The *type-II* system is detected very clearly in this case, especially when compared with the Fig. 7.11(b).

Finally, Fig. 7.13(a)-(b) constitutes another case where making a condition monitoring decision is somewhat difficult, this time for the cycle of the period $T = 200$ [*time steps*]. Ambiguous results were obtained for both types of modeled parameter changes. For the value of $\xi_0 = 1.18$, the λ_1-signature failed to detect any change in the system, as shown in Fig. 7.13(a) for the values of the statistic following the 30-th sample.

Model with backlash

As for the Hertzian model in the previous section, we investigate the λ_1-signature using three periodic orbits, with periods $T = 50, 200$, and 400 [*time steps*], which were extracted from the time series obtained from the model (6.3). Recall that the periodic orbit $T = 200$ corresponds to the gear meshing frequency in this model. All three orbits are shown in Fig. 7.14(a) next to the phase space portrait of the system that was reconstructed using the method of time delays and is shown in Fig. 7.14(b). The λ_1-signature computed for *type-I*

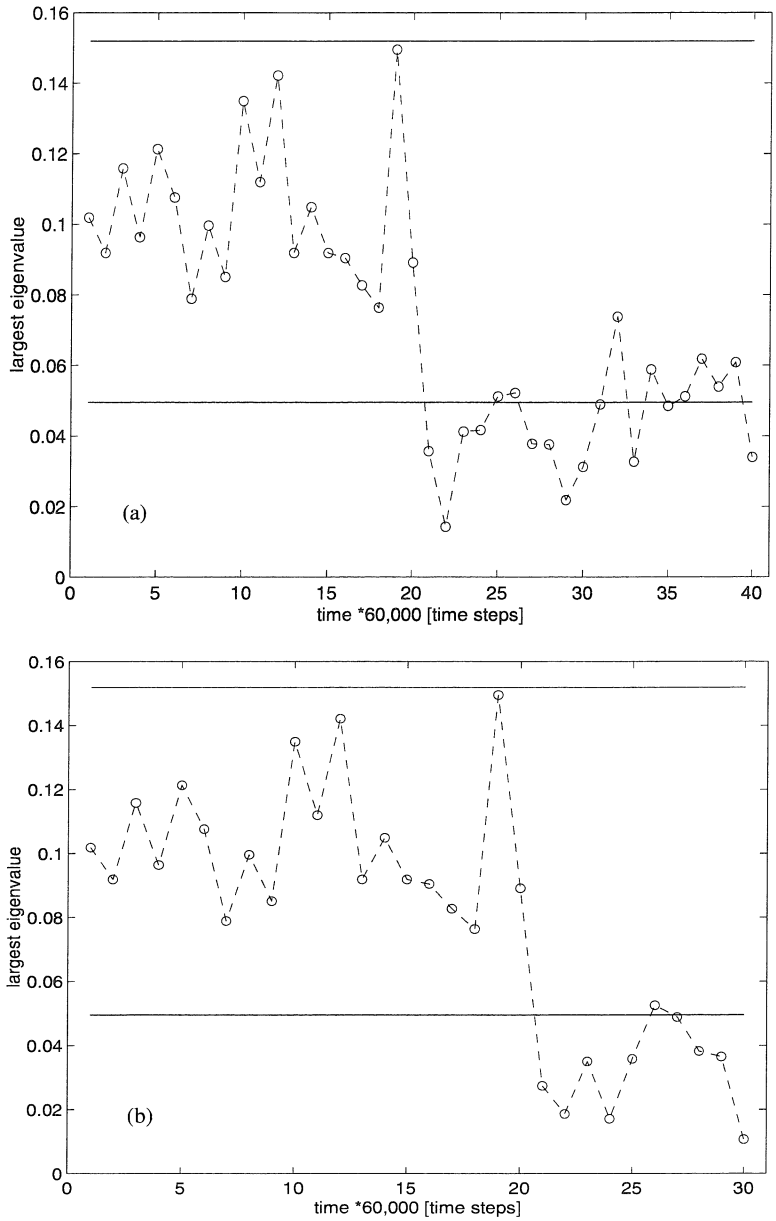


Figure 7.12: The largest eigenvalue, λ_1, computed for the period $T = 100$ orbit extracted from two *type-I*, (a), and one *type-II*, (b), signals obtained from the model with Hertzian force. The solid lines denote the 3σ alert zone.

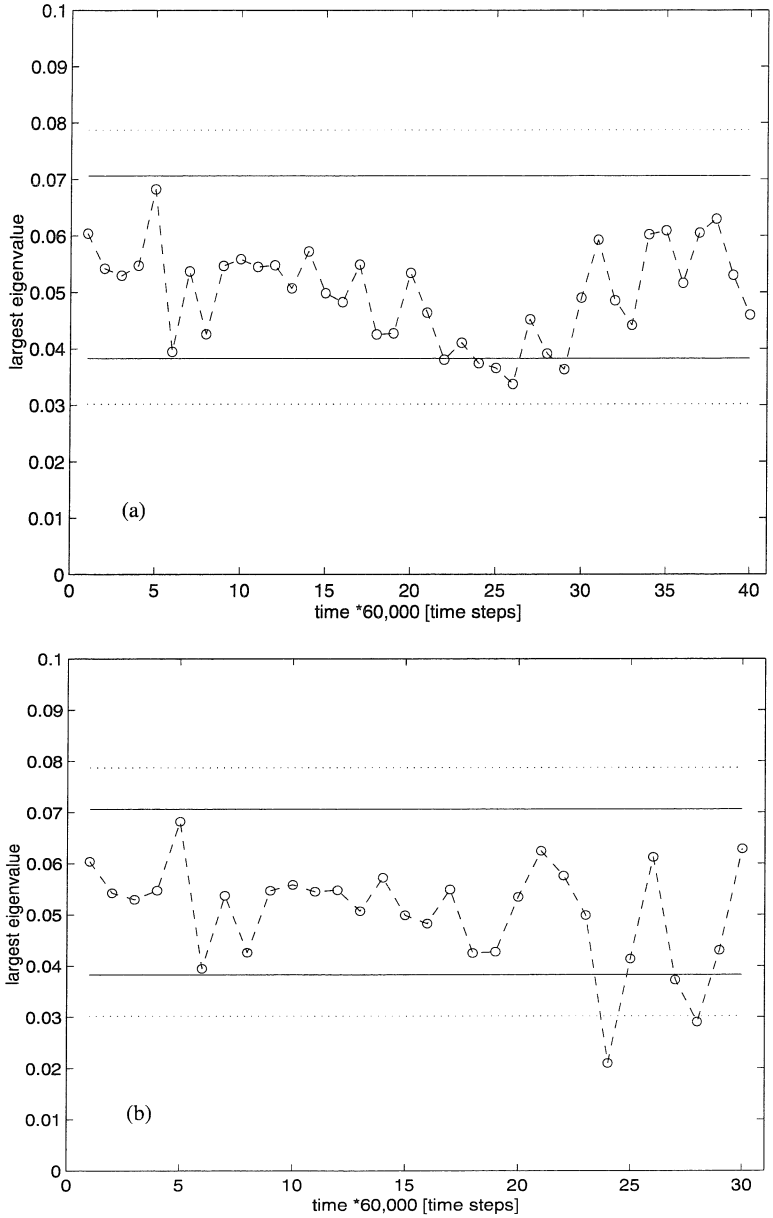


Figure 7.13: The largest eigenvalue, λ_1, computed for the period $T = 200$ orbit extracted from two *type-I*, (a), and one *type-II*, (b), signals obtained from the model with Hertzian forces. The solid lines denote the 2σ alert zone, and the point lines the 3σ zone.

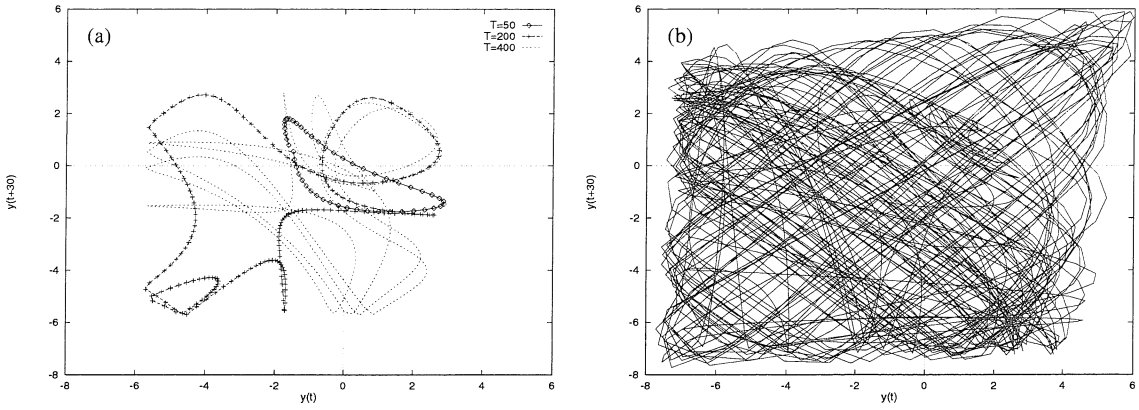


Figure 7.14: (a) Periodic orbits ($T = 50, 200$, and 400 [*time steps*]) extracted from the signal acquired from the model with backlash. (b) Reconstructed phase space portrait, projected on $\mathbb{R}^2$.

and *type-II* states using the $T = 50$ periodic orbit is shown in Fig. 7.15. It can be concluded from Fig. 7.15(a) that the distributed type of modeled defect, *type-I*, was not detected in this case. It was also undetected using the cycle corresponding to the meshing frequency, as shown in Fig. 7.16(a). Since essentially the same result was obtained for the other analyzed periodic orbits, we will not present them here. This is the first case in which this signature completely failed to give any warning about the *type-I* system change. As it has turned out, this case seems to be a difficult one since none of the techniques used so far have succeeded. Not only linear methods failed here but also both of the deterministic signatures employing the predictability of the system and stability of its periodic orbits failed too. There remains one more viable alternative for detection of this defect: the nonlinear statistical signature, which will be explored and applied to this case in the next section.

Significantly better results were obtained when the λ_1-signature was applied to detect the local, *type-II*, system change. As can be seen in Fig. 7.15(b), the detection is quite ambiguous using the period $T = 50$ cycle, but it becomes clear when the period $T = 200$ orbit is employed (Fig. 7.16(b)).

Let us make a few final comments on the λ_1-signature before we end this section. The eigenvalues computed according to the algorithm described here correspond to Lyapunov characteristic exponents evaluated along periodic orbits. An estimate of the global Lyapunov exponents for the system can be obtained from a weighted average of the eigenvalues corresponding to all the significant (for this representation) periodic orbits [30]. This process

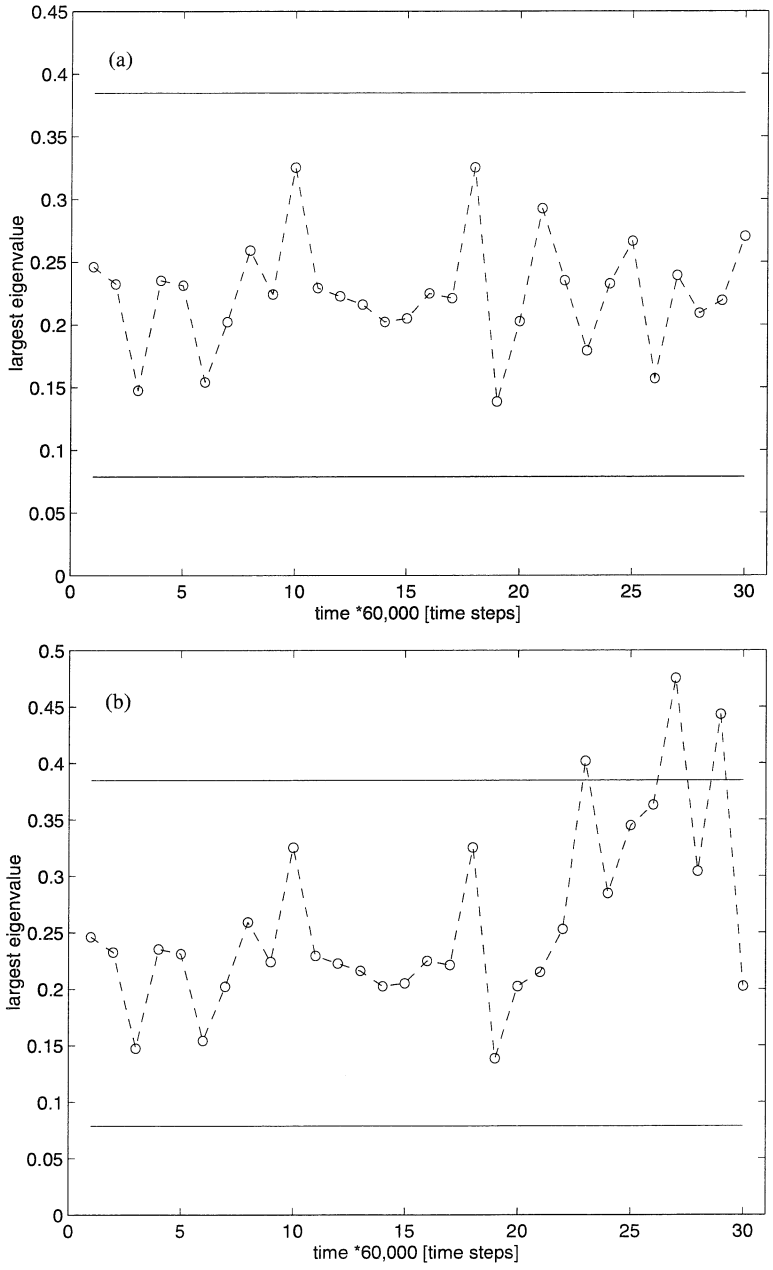


Figure 7.15: The largest eigenvalue, λ_1, computed for the period $T = 50$ orbit extracted from two *type-I*, (a), and one *type-II*, (b), signals obtained from the model with backlash. The solid lines denote the 3σ alert zones.

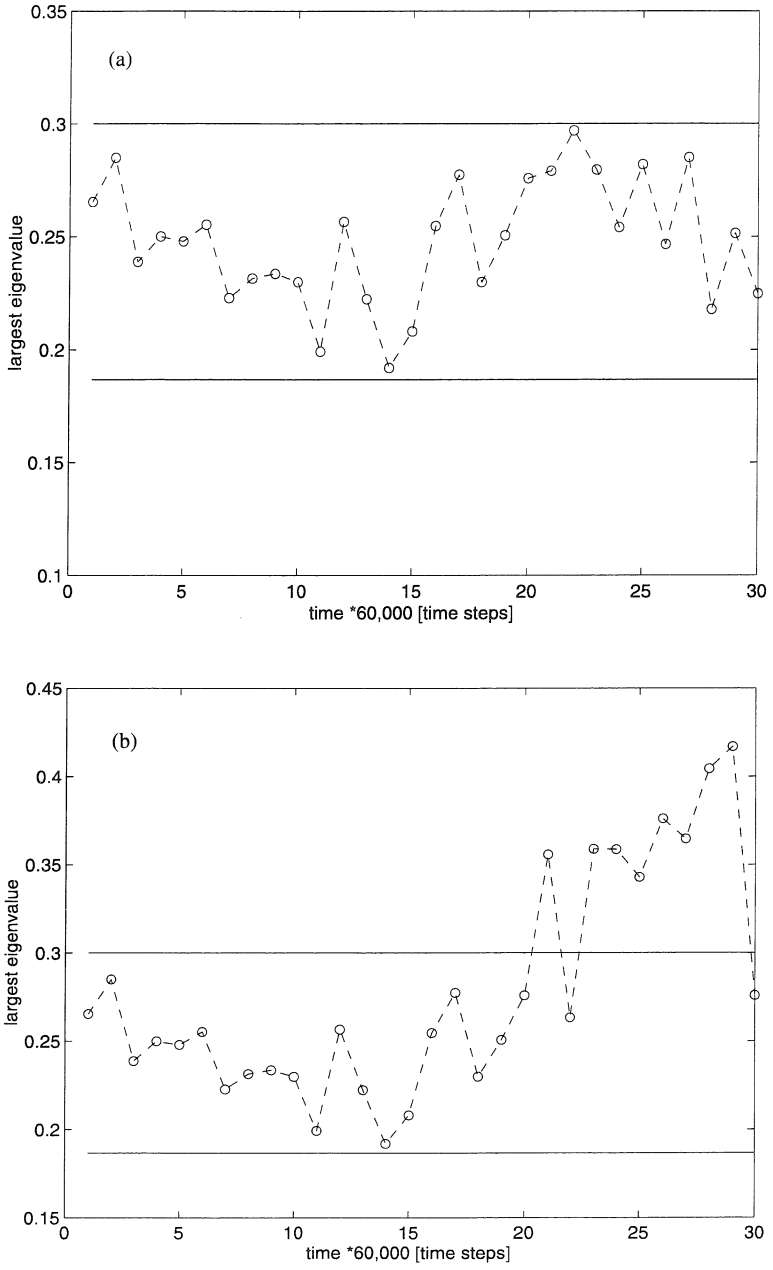


Figure 7.16: The largest eigenvalue, λ_1, computed for the period $T = 200$ orbit extracted from two *type-I*, (a), and one *type-II*, (b), signals obtained from the model with backlash. The solid lines denote the 3σ alert zones.

suggests a way to find a number of periodic orbits in the cyclic expansion which faithfully represents the system dynamics. The values of the Lyapunov exponents need not be known. It is enough to observe the last considered orbit which significantly contributed to the estimated value of the exponent. Such procedure should be undertaken when one builds a "cyclic" model of a dynamical system. In a qualitative model like this, one uses extracted periodic orbits, their stability and location in the state space in place of equations of motion.

As we already mentioned above, there is no need to observe all periodic orbits in order to use their stability as a condition monitoring tool to detect defects. Provided the regions of the state space where the change in dynamics can be observed are visited, only a small number of orbits is necessary (as in Fig. 7.16(b)). On the other hand, if no statistically significant change has been detected by a signature employing a number of orbits smaller than that for which the estimate of the Lyapunov exponent converges, we cannot conclude that a change did not occur (as it was shown in Fig. 7.15(a)-7.16(a)). In many practical applications, however, engineers know certain characteristics of the failure processes they are monitoring; for example, the frequency mode in which a catastrophic defect usually occurs. Such information can limit the set of monitored orbits to only a few.

As it has been clear throughout this section, the main limitation of the λ_1-signature is the accuracy with which periodic orbits are extracted from the experimental trajectory. Among the most important reasons for this is a small "visibility" of an orbit. Strongly unstable periodic orbits occupy areas of the state space which are infrequently visited by other orbits, and this makes them hard to observe. A few more conclusions can also be made. Shorter orbits are easier to extract than the long ones. The largest eigenvalue, i.e., the one representing the repelling direction of an orbit, is estimated most accurately. That is why in all figures in this section we used only the largest eigenvalue as the condition monitoring statistic (hence we also call this method the λ_1-signature). The quality of the approximation of a cycle can be assessed by observing the variation in its λ_1-signature. Even for the cases where this method is not very effective as a condition monitoring tool, it provides very important information about the system. In our work, we found that the small variations in the distribution of the orbital eigenvalues can be used as a good indicator of a sufficient length of the analyzed sample. To decide the sample size, one of the measures we employed for all analyzed signals was the λ_1-signature (together with stationarity of the probability density function and of the FFT).

7.3 Orbital density function and state monitoring

It was observed during early stages of work in the field of the analysis of experimental chaotic signals that a statistic referring to the probability of finding a trajectory in a certain area of the phase space may provide a very useful assessment of system dynamics. It is clear that the probability density estimated in a multi-dimensional phase space is richer in information than its one-dimensional counterpart. As pointed out by Kapitaniak [23], density functions are inherently multi-modal for complex nonlinear and chaotic systems. Observing a function in a multi-dimensional space can be troublesome, however. For many purposes, like exploration and presentation of data, it is useful to reduce the dimensionality of the analyzed signal by projecting it on a one- or two-dimensional subspace. The general idea behind the *projection pursuit* is to define a criterion for such a projection which can be then optimized in order to find the projection of most interest to the observer. A most commonly used example of the projection pursuit is the *principal component analysis*, in which the optimal projection is chosen as a subspace (usually spanned by the system eigenvectors) that maximizes the variance of the projected data. Another possible solution to the projection pursuit problem is a marginal distribution.

In their recent papers, Wright and Schult [57, 56] suggested that arbitrary one-dimensional slices of the probability density function and of its Fourier transform (characteristic function) can be successfully used for classification and monitoring of nonlinear chaotic signals. In this work, we propose a new approach in which the multivariate probability density function is projected on a periodic orbit extracted from the system dynamics. The projection is performed by estimating the values of the *pdf* for all the points that belong to the periodic orbit in the state space. This approach is especially informative if applied to periodic orbits which are significant to the observer. For example, a slice of the density function, created along the cycle corresponding to the gear meshing frequency, provides information not only about this particular spectral component but also about its modulation sidebands. That is why, since certain periodic orbits carry very important information for a condition monitoring engineer, they constitute a preferred projection that can be easily interpreted.

All of the comments we made about periodic orbits and their extraction from an experimental trajectory in the previous section are, of course, valid here as well. Since this

measure is statistical, we expect it to be a little less sensitive to the quality of the estimation of periodic orbits. This should result in better effectiveness and robustness of this method compared to the λ_1-signature.

Rössler model

The probability density function used throughout this section to evaluate the condition monitoring signature was computed in the system phase space employing the kernel estimators described in Section 3.5.1. We use the same periodic orbits here as we did for the λ_1-signature. For the sake of brevity, we will illustrate our results on the examples of $T = 200$ and $T = 600$ [*time steps*] cycles shown in Fig. 3.9(a) and Fig. 7.5, respectively. The probability density function (*pdf*) was estimated for signals obtained from the Rössler model. In all the figures in this section, ensembles of 20 samples (30,000 points long each) were used to compute the 3σ alert zones at every point along the considered cycle by fitting a normal distribution to each ensemble and finding its median and standard deviation σ. Similarly, as for the previously defined condition monitoring methods, the alert zone is centered at the median value and its width is taken as $\pm 3\sigma$ of the *pdf* at every point along the orbit. Since the value of the *pdf* usually varies along the orbit, the lines bordering the alert zones are no longer straight, as they were for the λ_1-signature. The width of the zone reflects the lack of stationarity of the estimated *pdf*.

The result of the projection of the *pdf* estimated in $\mathbb{R}^3$ on the period $T = 200$ and $T = 600$ orbits for the Rössler system, undergoing the *type-I* change in the values of the parameter a (as specified in Section 6.4), is shown in Fig. 7.17. Two solid lines shown in the figure correspond to the 3σ alert zones created for the system with the parameter corresponding to its reference ("good") state. There are two important points that need to be made here. First is that the proposed projection results in clear crossings of the alert zone and, as a consequence, a clear detection of a change in the system. Secondly, it can be noticed that for certain regions along the orbit, the probability density function curve stays within the alert zone. It means that if an arbitrarily chosen one-dimensional slice of the multivariate *pdf* happened to go through these regions, this change of the operating condition would not be detected. This is a very important observation because it is a widespread belief that an arbitrary slice, or a one-dimensional projection, of a density function defined in the phase space should work. As it will be shown for other models, in many

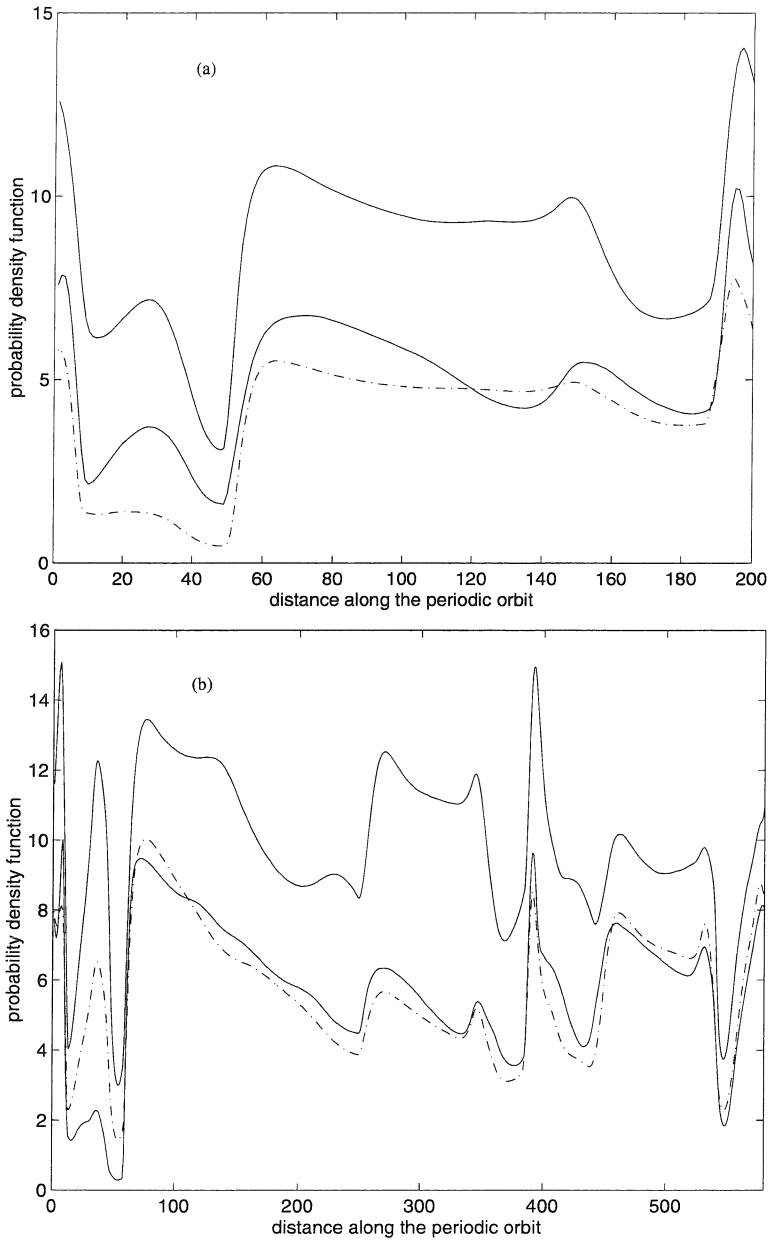


Figure 7.17: Projection of the probability density function on the period $T = 200$ [*timesteps*] orbit, (a), and the period $T = 600$ [*time steps*] orbit, (b), obtained from the Rössler system. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$. The point lines correspond to the case where the parameter value was changed to $a = 0.16$ (*type-I change*).

cases the regions of the poor or zero detectability are more common than those for which a reliable fault detection can be made. In this light, it seems that taking slices of the *pdf* along the system cycles is a safer and more sensitive monitoring method.

As can be noticed from Fig. 7.18(a), sensitivity of this method is unfortunately not good for this particular case of the local type of defect, at least as monitored along this particular orbit. It seems to be slightly better, even though still not acceptable, when the "strength" of the local defect increases. Fig. 7.18(b) was created for the *type-II* defect of the same duration, of 10 time steps, as in Fig. 7.18(a) but for the parameter value $a = 0.30$. As we recall from Fig. 7.8, the λ_1-statistic did not do well in this particular case either. In all of the results presented so far, there was only one line representing a typical behavior throughout the entire ensemble. The reason why there is a variety of lines in Fig. 7.18 is that some of the samples in the ensemble intersected the alert zone, indicating a change in the system, and some did not. We set a rule that it is only in the cases where all of the samples intersect the alert zone in the same way that the result is presented as one line.

The reasons why this particular statistic failed in this case are similar to those already mentioned in Section 7.2. Except for the problems with accurate estimation of periodic orbits and with the choice of the informative periodic orbit, there is also the problem of the estimation of the *pdf*. There is an interesting way in which the *pdf* estimation process can be optimized for the purpose of this condition monitoring technique. In Section 3.5.1, we devoted some time to the importance of the appropriate choice of the kernel smoothing parameters. Another possibility for their selection is to choose those values which minimize the width of the alert zone for a particular orbit. This local approach would allow different values of the smoothing parameters across the attracting set and would have an advantage of being streamlined for this particular application.

It is interesting to find out how much of the difficulties with the detection of the *type-II* state is caused by the process of reconstruction of the state space, similar to what we did in Fig. 7.6 for the λ_1-signature. Fig. 7.19(a) confirms the result obtained in Fig. 7.17 for the *type-I* change. It also confirms our earlier conclusions that a change in the signature observed in the reconstructed state space corresponds to the change in the original space and is not a result of the trajectory projection or the embedding process. From Fig. 7.19(b) and Fig. 7.20, we can conclude that when this signature is used in the original phase space, the system change is also not clear even though it is much better than that observed for

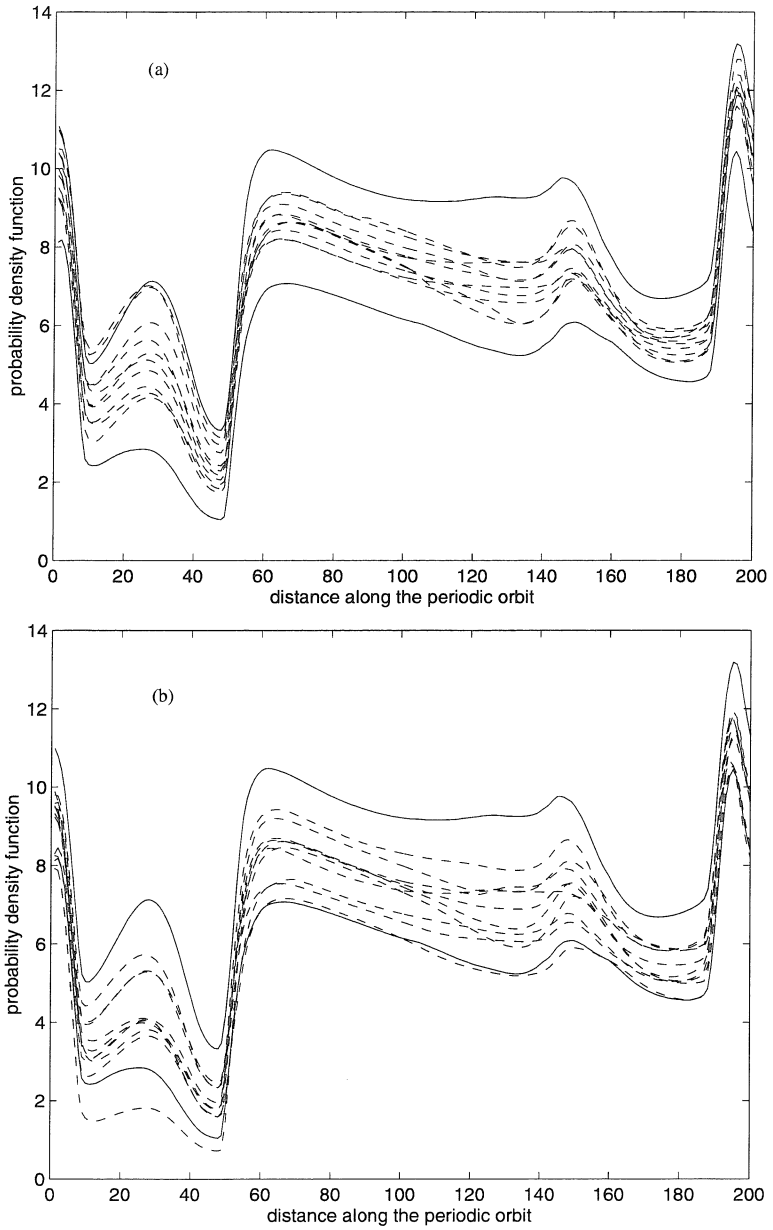


Figure 7.18: Projection of the probability density function on the period $T = 200$ [*timesteps*] orbit, extracted from the Rössler system, for the *type-II* model defect of the same duration but different "strength": (a) $a = 0.16$, (b) $a = 0.3$. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$.

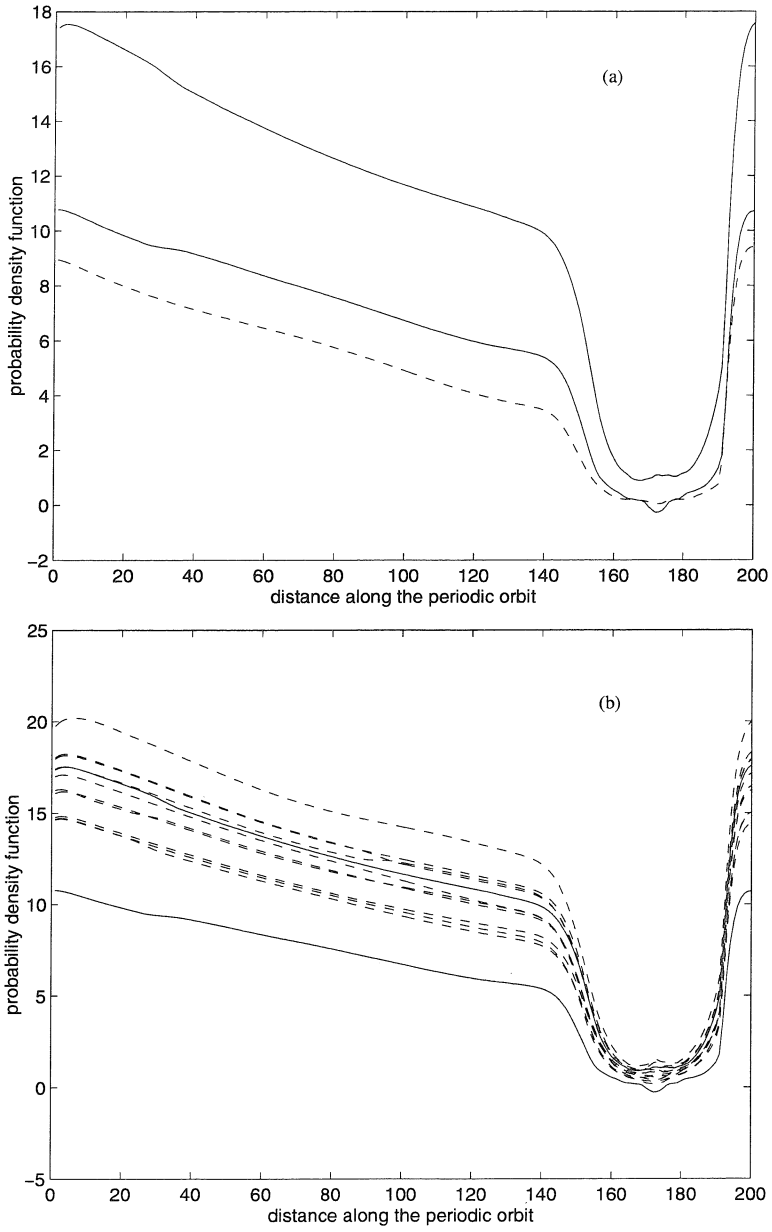


Figure 7.19: Projection of the probability density function on the period $T = 200$ [*timesteps*] orbit extracted from the x, y, z coordinates of the Rössler system. The 3σ alert zone, shown between the solid lines, was created from the signal obtained for the parameter value $a = 0.15$. The point line corresponds to the case where the parameter value was changed according to (a) *type-I* and (b) *type-II* defect model.

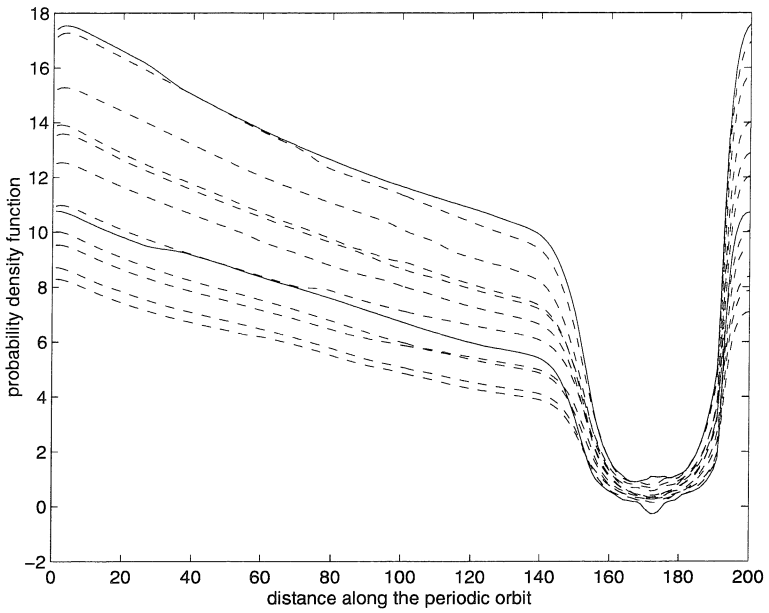


Figure 7.20: Projection of the probability density function on the period $T = 200$ [*timesteps*] orbit extracted from the x, y, z coordinates of the Rössler system. The point line corresponds to the case where the parameter value was changed according to the *type-II* defect model, $(a = 0.30)$.

the reconstructed trajectory. We can conclude from this figure that the difficulties we encountered are not only caused by the *pdf* estimation but also by the approximation of the employed periodic orbit.

One can wonder how the proposed method would compare with its one-dimensional counterpart. In order to find this out, we extracted (using the same method as before) the $T = 200, 400$, and 600 cycles directly from the one-dimensional time series, as opposed to extracting it in the phase space. For the sake of brevity, we illustrate the results for only one cycle, with the period $T = 600$ [*time steps*]. The time trace for this one-dimensional approximation to the periodic orbit is shown in Fig. 7.21(a). We estimated one-dimensional *pdf* for the time series along this orbit. The solid lines shown in Fig. 7.21(b) correspond to the 3σ alert zone obtained for the system with the parameter $a = 0.15$. The point line, which corresponds to the new, *type-I*, state, does not cross the alert zone, and as a result fails to identify the change in the system operating conditions. The same conclusions were made for the other periodic orbits. These observations confirm the superiority of the multi-dimensional approach in this method of system condition monitoring.

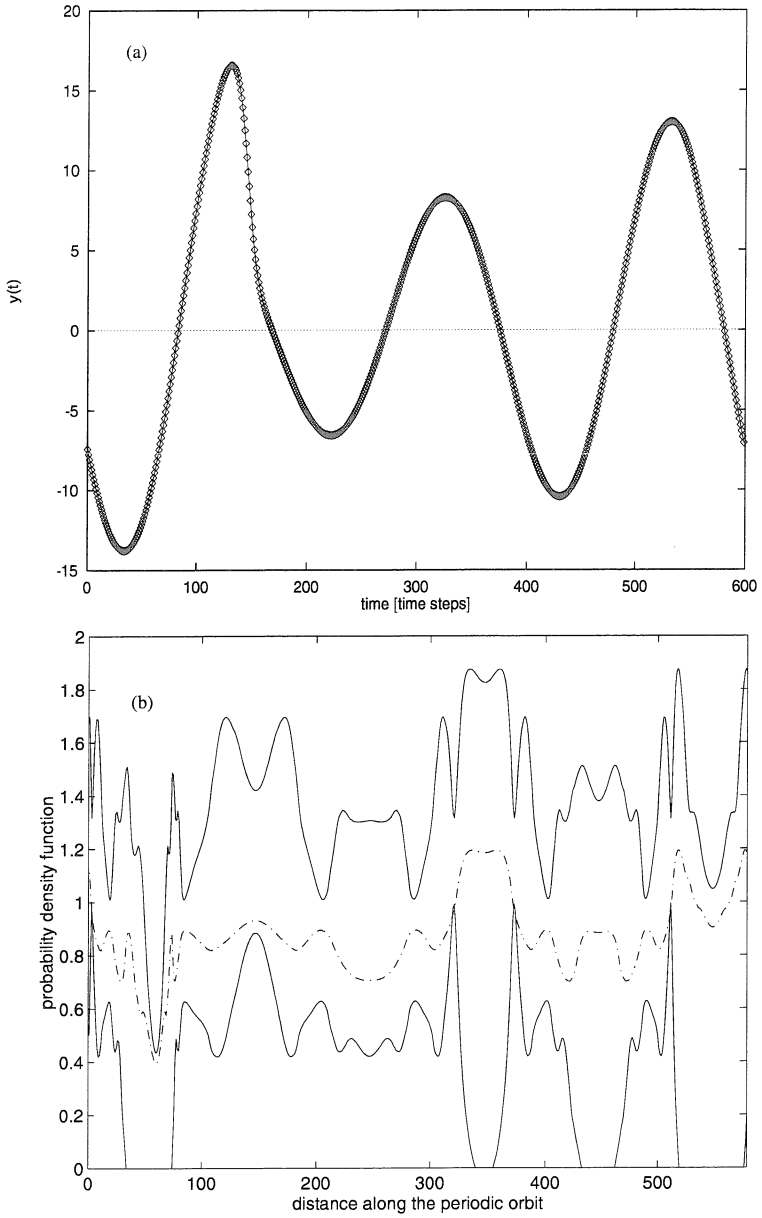


Figure 7.21: One-dimensional periodic orbit (period $T = 600$ [time steps]) extracted from a time series obtained by numerical integration of the Rössler system, (a). Probability density function computed along this orbit is shown in (b). The 3σ alert zone, limited by the solid lines, was created from the "good" signal, and the point line for to the *type-I* signal.

Model with Hertzian force

We will now investigate the orbital-*pdf*-signature for data from the Hertzian model for three cycles, $T = 50, 100$, and 200 [*time steps*], as we did in Section 7.2. The result for the case when the *pdf* is projected on the $T = 50$ periodic orbit is shown in Fig. 7.22. Both global and local modeled defects are detected very well by all samples in the analyzed ensembles. The sample results, representative of their respective ensembles, are shown in the form of single dashed lines for both cases. The first interesting conclusion that can be drawn from Fig. 7.22 is a local (in the state space) character of both types of changes. This is a very good example of the case (mentioned in Section 7.2), where it would be very difficult to find a simple one-dimensional cut through the state space which would intersect this particular region where the change in the *pdf* is observable. Another interesting observation can be made when the sensitivity of this signature is compared with the sensitivity of the λ_1-signature: from the comparison of Fig. 7.22 and Fig. 7.11 it is clear that for this case, both types of defects are detected better by the orbital-*pdf*-signature.

The equivalent results for the period $T = 100$ cycle are presented in Fig. 7.23. Here again we notice that both types of changes were detected clearly. As it is also evident from Fig. 7.12, this periodic orbit seems to be a better choice for monitoring of these particular changes in the system parameters than the orbit with period $T = 50$.

Obviously, there are orbits which only partially explore the regions of the system phase space where the change in the behavior can be detected. An example of such a case can be seen in Fig. 7.24 created for the orbit $T = 200$. For the *type-I* change, presented in Fig. 7.24(a), there are only two short segments of the cycle (near the 155-th and 180-th samples) where the alert zone is crossed by all of the signatures in the observed ensemble. Even though these events appeared consistently throughout the ensemble and therefore imply a change in the system state, they are not as impressive as those obtained for the shorter ($T = 50$ and $T = 100$) orbits. For the *type-II* state, only a small, and therefore statistically insignificant, number of samples shown in Fig. 7.24(b) intersect the alert zone. From these samples we cannot conclude that the system changed, and as a result this particular signature fails to detect the inflicted *type-II* change.

In order to compare the sensitivity of this method of condition monitoring to just monitoring the one-dimensional *pdf*, as before, we extracted the one-dimensional periodic orbits

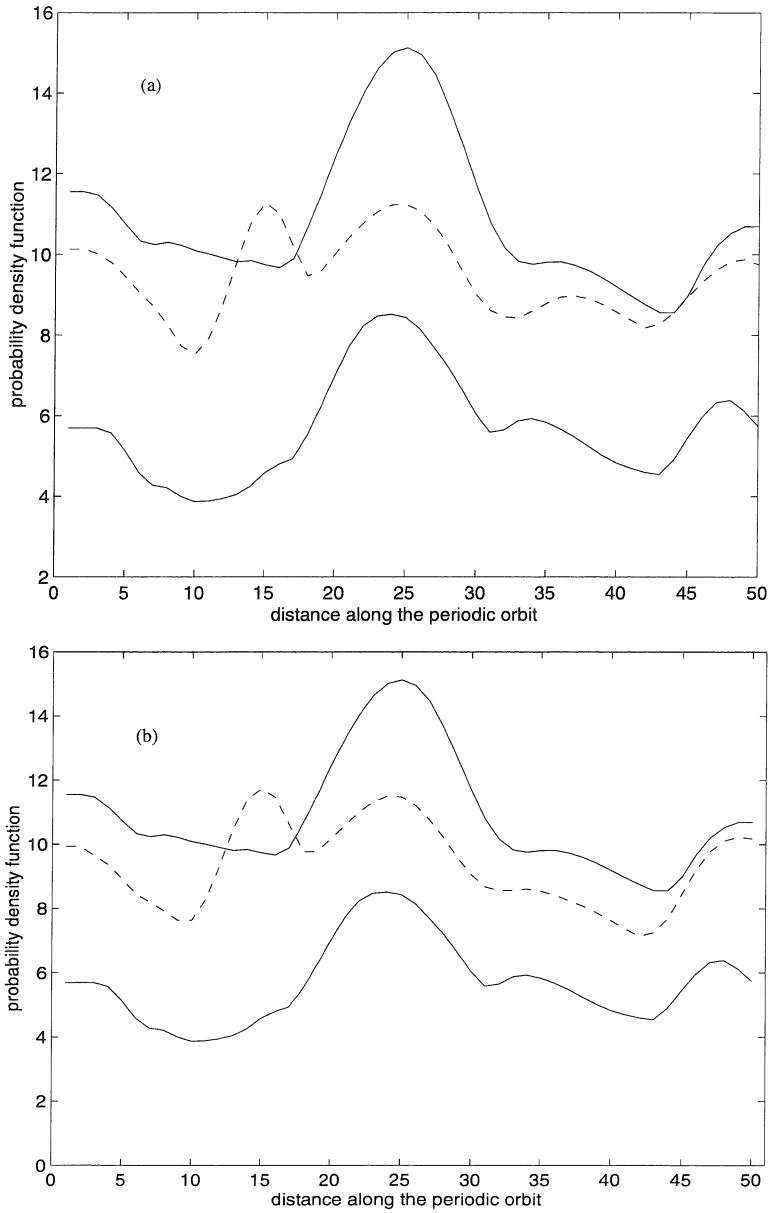


Figure 7.22: Projection of the probability density function on the period $T = 50$ orbit for the *type-I*, (a), and *type-II* signals obtained from the model with Hertzian force.

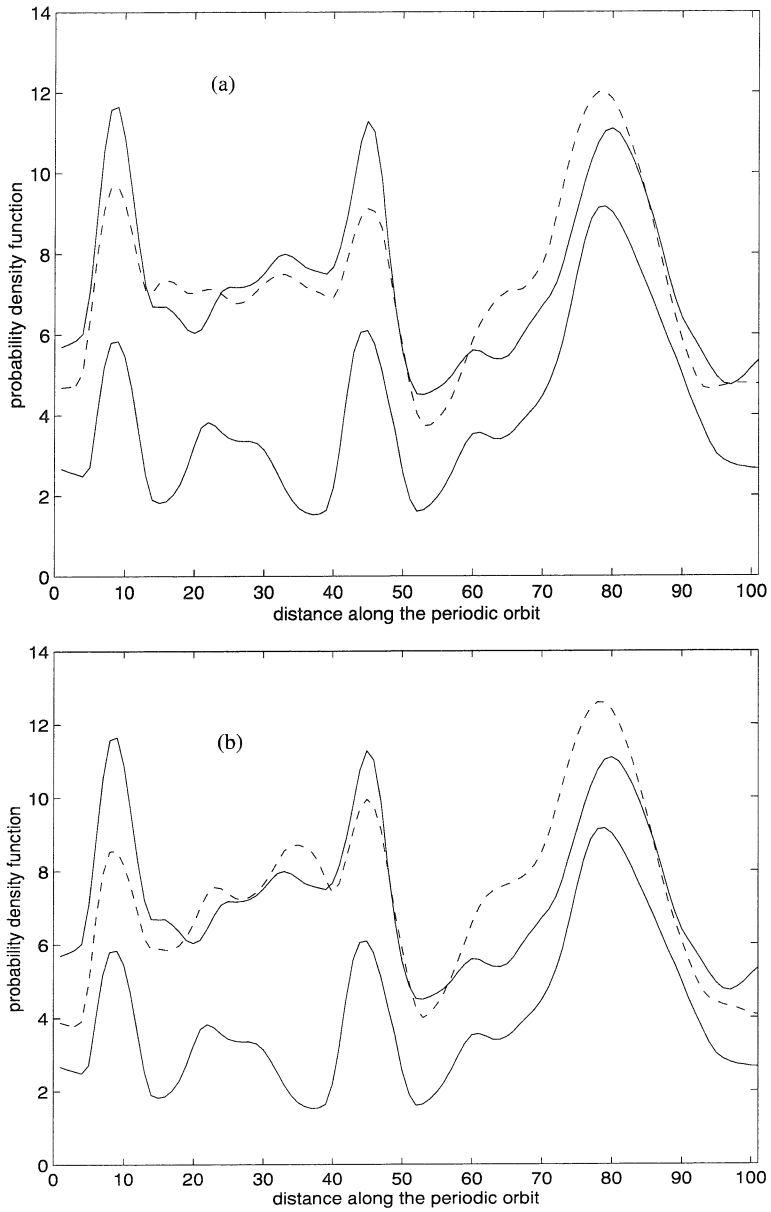


Figure 7.23: Projection of the probability density function on the period $T = 100$ orbit for the *type-I*, (a), and *type-II* signals obtained from the model with Hertzian force.

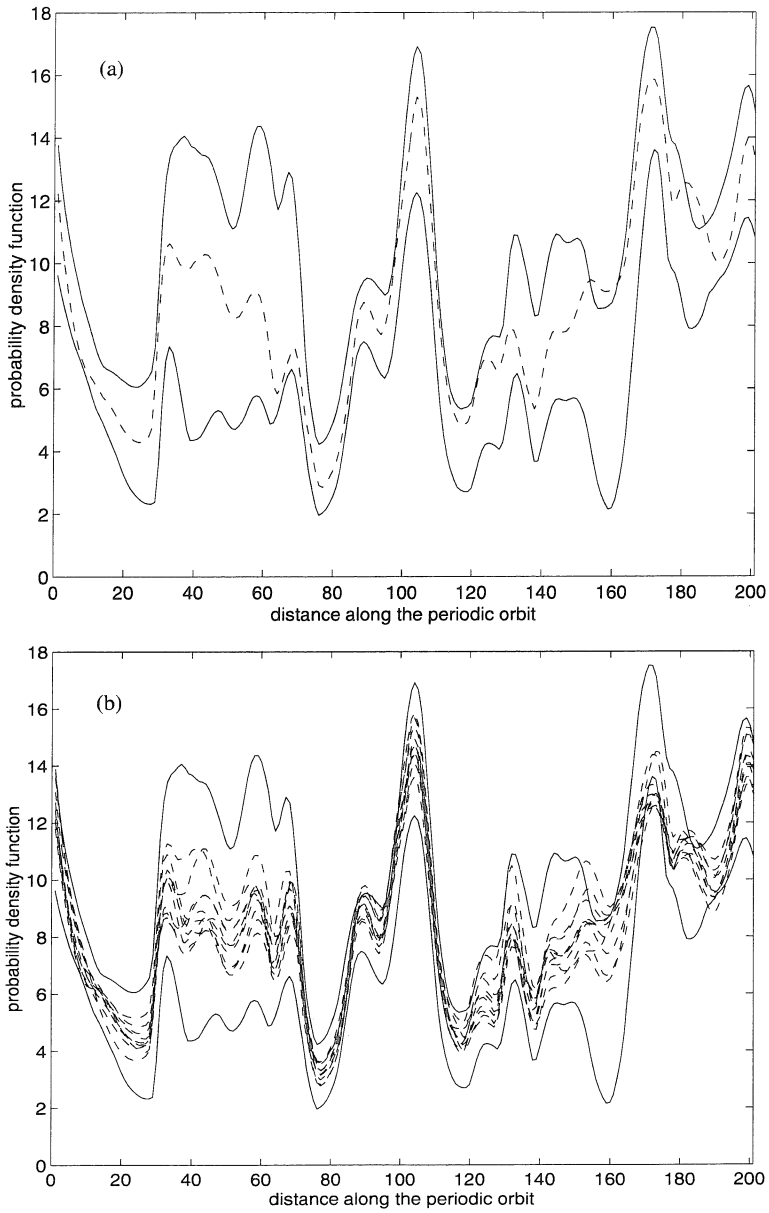


Figure 7.24: Projection of the probability density function on the period $T = 200$ orbit for the *type-I*, (a), and *type-II* signals obtained from the model with Hertzian force.

and computed their signatures. The resulting signatures did not show any intersections of the alert zones and therefore failed to detect all of the changes detected above by the proposed orbital projection method. We demonstrate that on the example of the $T = 100$ orbit, shown in Fig. 7.25(a). The signature obtained for this cycle is presented in Fig. 7.25(b). The lower bound for the alert zone is not visible in the figure because it falls below the zero probability level.

Model with backlash

Finally, we investigate the most difficult monitoring case introduced with the example of the model with backlash defined by (6.3). Recall that the *type-I* change for this model was not detected by any of the linear or nonlinear methods used thus far. We will now see how the orbital-*pdf*-signature performs in this case. Fig. 7.26 shows the result of projecting the probability density function, estimated in the phase space of this model, on the period $T = 50$ periodic orbit. As we can see, both types of inflicted parameter changes are clearly detected in this case. The same is true for the period $T = 200$ orbit. As can be seen in Fig. 7.27(a)-(b), multiple intersections of the 3σ alert zone clearly indicate the modeled defects.

Once again, we compare the sensitivity of this approach to just monitoring the one-dimensional *pdf* projected on one-dimensional periodic orbits. The resulting signatures did not show any intersections of the alert zones and therefore failed to detect any of the changes detected above by the proposed orbital projection method. We illustrate that in Fig. 7.28 on the example of the $T = 50$ orbit. The signatures obtained for the *type-I* and *type-II* cases are presented in Fig. 7.28(b)-(c).

In conclusion, the condition monitoring method proposed in this section allowed correct detection of the global, *type-I*, changes inflicted on all analyzed models. In some of the cases, its effectiveness was better than that of the purely "deterministic" λ_1-signature and the degree-of-predictability-signature. The localized types of defects, modeled in our analysis as the *type-II* change, were detected in most of the cases and the quality of result strongly depended on the choice of a periodic orbit used for the projection of the *pdf*. In the cases where the modeled defects were not detected by this method, they were also not detected by the λ_1-signature. This observation seems to indicate that the inadequacy in estimation of the system periodic orbits is mainly responsible for failure in those cases. It is also possible

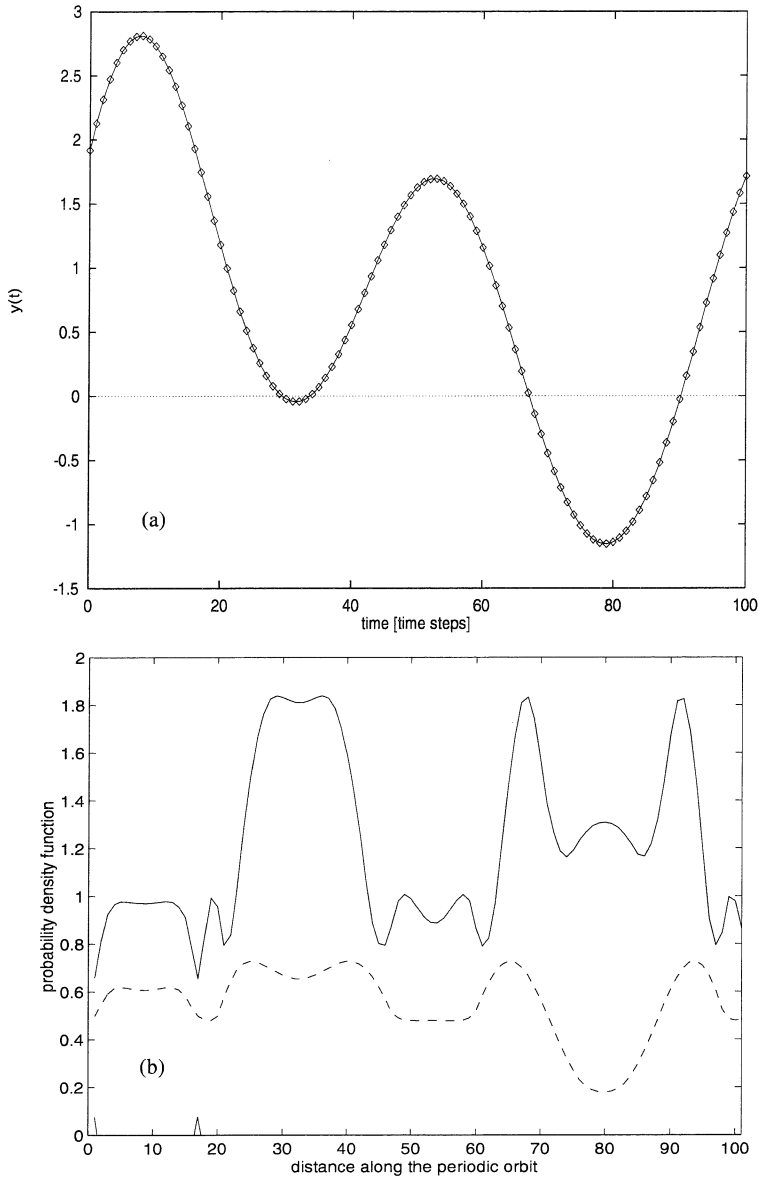


Figure 7.25: One-dimensional periodic orbit (period $T = 100$ [time steps] extracted from a time series obtained by numerical integration of the model with Hertzian force, (a). Probability density function computed along this orbit is shown in (b). The 3σ alert zone, limited by the solid lines, was created from the "good" signal, and the point line for the *type-I* signal.

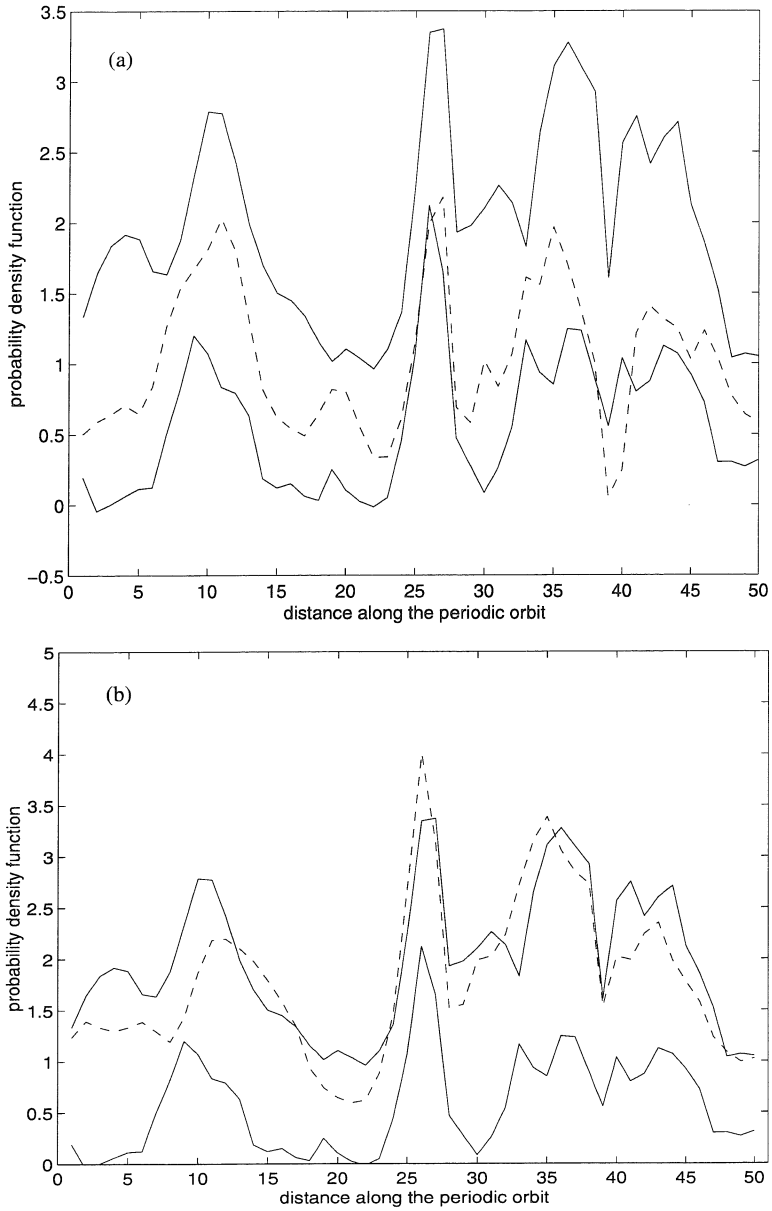


Figure 7.26: Projection of the probability density function on the period $T = 50$ orbit for the *type-I*, (a), and *type-II* signals obtained from the model with backlash.

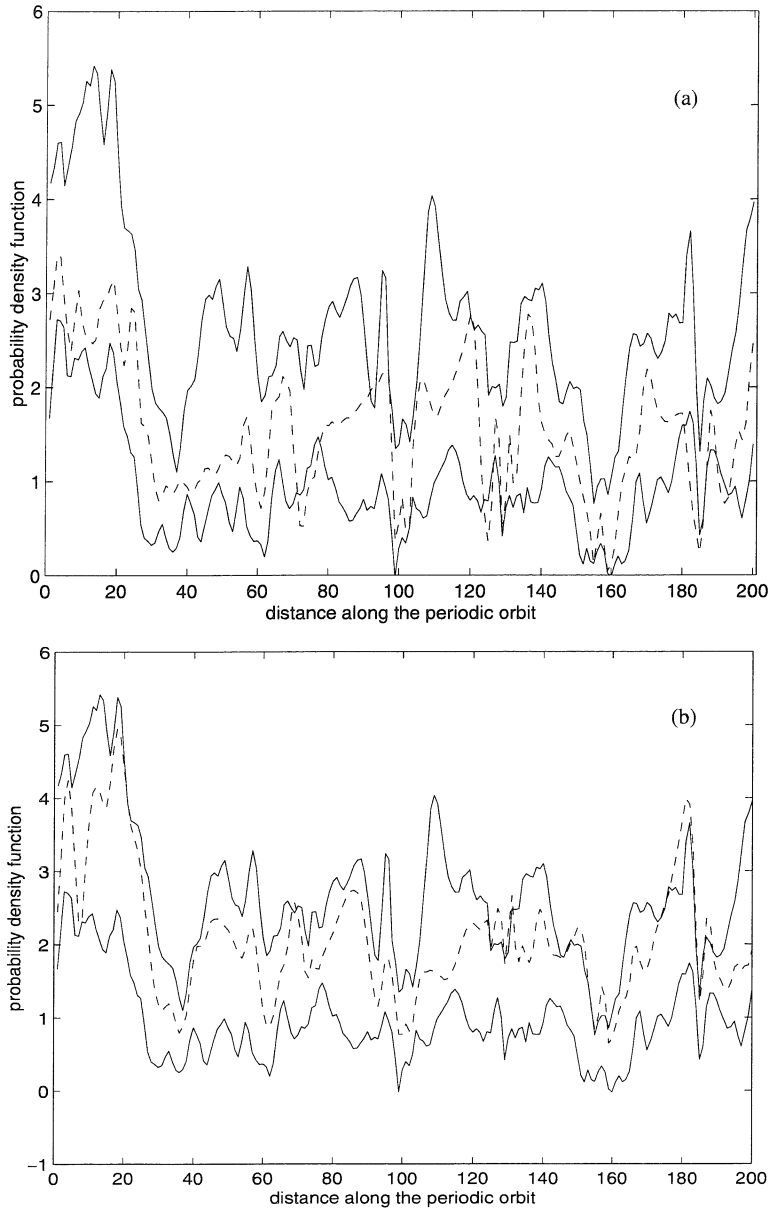


Figure 7.27: Projection of the probability density function on the period $T = 200$ orbit for the *type-I*, (a), and *type-II* signals obtained from the model with backlash.

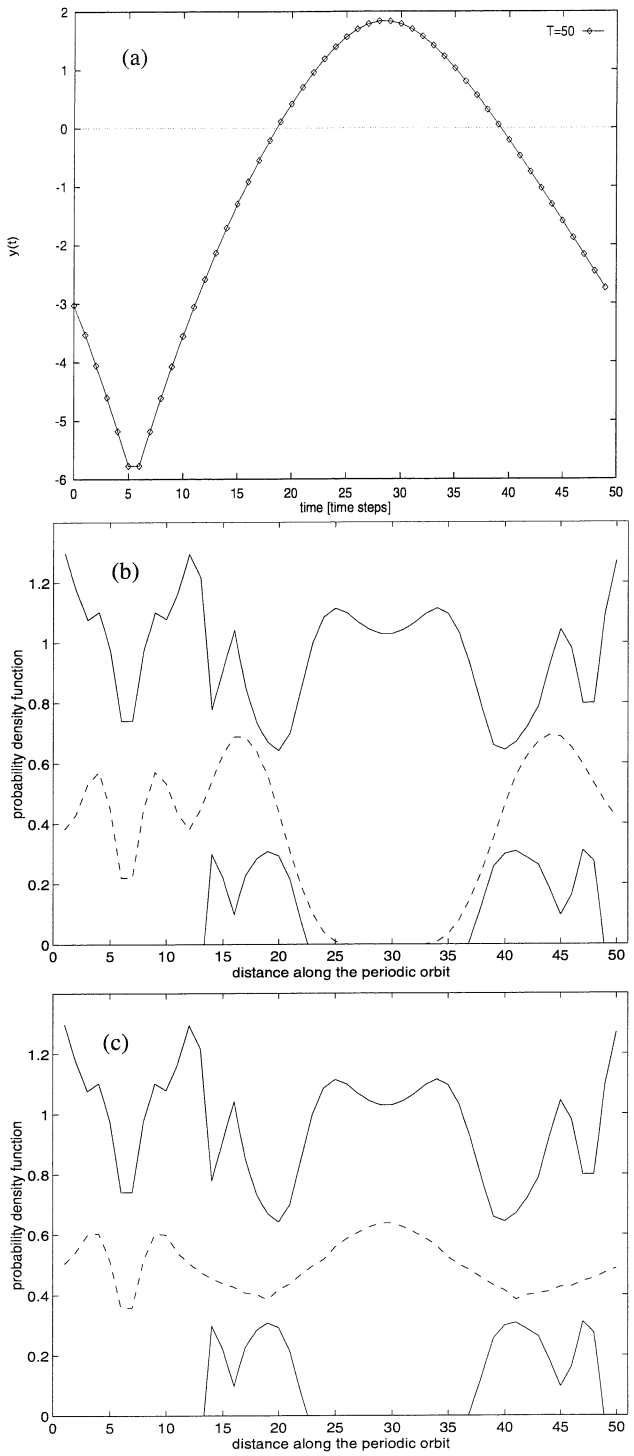


Figure 7.28: (a) One-dimensional periodic orbit (period $T = 50$ [*time steps*] extracted from a time series obtained by numerical integration of the model with backlash, (b) probability density function computed along this orbit for *type-I* state, and (c) *type-II* state.

that the changes in the geometry of the system behavior were so small, that the methods employing geometrical features of dynamics were unable to observe them. Only the method which quantifies predictability of the system was able to pass the test in all analyzed cases.

The other significant limitation of the orbital-*pdf*-signature is the estimation of the probability density function in a multi-dimensional phase space. From our observations we concluded that the obtained results depended on the choice of the smoothing parameters used to fit a selected kernel. A possible way to alleviate this problem is to choose the smoothing parameters in such a way that they minimize the size of the alert zones for a set of significant periodic orbits.

The biggest strength of this method is that since its nature is statistical, its dependence on the quality of the extraction of periodic orbits seems to be less important than for its purely deterministic counterpart (the λ_1-signature). This is one of the reasons why this statistic was significantly better at avoiding false alarms than the λ_1-signature. Another useful feature of this approach is that it allows localization of specific regions of the state space (and to a much higher extent than the λ_1-signature) where the defect influences the behavior of the trajectory in the system phase space. Finally, since this method is a combination of statistical and dynamical approaches, it should be more robust than the other two proposed methods of system condition monitoring.

Chapter 8

Conclusions

"Veni, vidi, vici?"

In 1963, Lorenz observed aperiodic behavior in a three-dimensional dynamical system which represented a crude approximation to the infinite-dimensional atmospheric model described with the Navier-Stokes equations. Even though the simplifications in the system proposed by Lorenz disqualified it from the point of view of quantitative analysis of fluid flow, the main contribution of that work was that it suggested a possibility that the "random"-looking behavior observed in an infinite-dimensional system might be caused by finite-dimensional dynamics.

One of the features typical of the Fourier spectrum obtained from a gearbox signal are modulation sidebands, which appear in the vicinity of the gear meshing frequency and its harmonics. The "random"-looking sidebands are commonly attributed to various inherent material and kinematic nonlinearities and imperfections in gearboxes. In our work, we postulate that this apparently high-dimensional behavior can be modeled using a family of deterministic and probabilistic analytical tools, which are employed in a relatively low-dimensional state space reconstructed for the gearbox system from its one-dimensional vibration signal.

In this contribution, our approach to signal analysis is different from what one would typically find in literature on dynamical time series analysis. Our goal is similar to that of conventional system identification and is focused on producing a working model of the state of the system rather than on making claims about the "inherent features of the underlying system." This difference can be well illustrated, for example, in the problem of identification of chaotic dynamics. The conclusion that most of the physicists would make from Chapter 3 about the analyzed gearbox signal is that it comes from a system which is "inherently

chaotic". We concluded, more humbly, that it can be modeled as such. In our opinion, the approach we undertook here not only suffers from fewer "philosophical" difficulties, but since it is closer to conventional system identification methodology, it should also make its implementation more intuitive for the engineering audience (to which this work is primarily addressed).

After the experimental vibration signal was acquired from the gearbox test rig, we used a family of analytical tools to identify a dynamical model of the system dynamics. These tools consisted of: mutual information function, false nearest neighbors statistic, recurrence plots, and Lyapunov characteristic exponents. The analysis of the experimental time series showed that the gearbox dynamics can be modeled as a nonlinear system embedded in $\mathbb{R}^5$. The first choice of the embedding parameters, i.e., embedding dimension and time-delay, was later verified (and in the case of time-delay adjusted from the initial value of $\tau = 11$ to $\tau = 9$) using the criterion of the predictability of the system response. The identified embedding parameters of the model state space suggest that the complex behavior and the modulation sidebands that vary for different harmonics of the meshing frequency may be interpreted as a result of a relatively simple low dimensional system. The analysis also showed that the system behavior is sensitive to initial conditions, which was confirmed by the existence of the positive Lyapunov exponents.

Because of a strong tendency within the community of dynamical time series analysis to use the surrogate data tests, we performed the Kolmogorov-Smirnov test of the prediction errors, obtained for the acquired time series and the artificially produced surrogate data. This test confirmed that the hypothesis that the observed time series comes from a linear time-invariant model is unlikely.

Further, to confirm the applicability of a dynamical systems approach to the analysis of gearbox vibrations, we attempted to predict the acquired time series using a nonlinear dynamical predictor and a probabilistic predictor in the reconstructed state space. The results presented in Section 4.2 confirmed that the short-term prediction of the gearbox signal is possible. Since modeling of experimental data always involves uncertainties, the probabilistic prediction scheme resulted in better forecasts of gearbox signal than its purely deterministic counterpart, as expected. We also compared these predictions with those obtained from a linear predictor employing a neural network. Both nonlinear predictors produced superior results in this comparison. We do not conclude from this, however, that

every nonlinear scheme would beat any linear scheme. It should also be pointed out that there are ways in which all of the predictors we used can be improved. Besides, making predictions is not the main goal of our analysis. It is a tool which we applied in Chapter 7 to perform system condition monitoring.

In many practical situations, engineers face the problem of signal separation. This problem is the primary concern, for example, in the field known as acoustic noise control. A large number of powerful linear filtering methods is available. They all work on the assumption that the broadband component of the Fourier spectrum is related to "noise" and therefore should be discarded. For many practical applications, this approach works fine. However, if the broadband component caries information about the analyzed system (as it does for chaotic systems), removing the "noise" also results in distorting the signal of interest. We discussed three signal separation schemes which we argued can be successfully used even when one (or more) of the composite signals can be modeled as chaotic. Their common feature is that they all explore dynamics of the system in its reconstructed state space and that is why they are superior to linear filters whenever the analyzed signals possess broadband spectra. The main limitation of the discussed methods of signal separation is that for high quality results they depend on the availability of a clean reference signal. Fortunately, in many engineering applications, this requirement is easily met. From our experience with the presented signal separation schemes, we concluded that the probabilistic method applied in the system state space was most effective and provided the broadest range of possible applications.

Since vibration data for damaged gears was unavailable, we presented three models whose response resembles qualitatively that obtained from our experiment. We required that all the models possess certain qualitative features typical of the analyzed gearbox vibration signal. One of these features was the Fourier spectrum with dominant discrete peaks and modulation sidebands. Another was that the model system had to be nonlinear, low-dimensional, and possess at least one positive Lyapunov exponent. Finally, the mutual information function had to exhibit features which were recurrent with the model equivalent of the meshing frequency. We also proposed two models for local and distributed gear tooth defects.

Data from these models introduced in Chapter 6 were later used to investigate the effectiveness of the condition monitoring methods proposed in Chapter 7. All of the methods

employ various features which characterize nonlinear dynamical systems, i.e., predictability, stability of periodic orbits, and probability density function estimated in the state space. The first of the proposed methods uses short term predictions to quantify changes in the system operating conditions. If the acquired time series faithfully represents the dynamics of the system, if a good predictor was constructed, and if the data acquisition process or the external excitation has not changed, then consistently larger prediction errors obtained for identical measurements acquired at later time suggest that the operating conditions have changed. The main limitation of this method is that it depends on consistency of produced forecasts. Even if the obtained degree of predictability is relatively high, inconsistencies in predictions can result in a wide alert zone and, as a consequence, in poor sensitivity of this signature. In cases like this it is recommended that a different approach be used, for example, one of the methods utilizing periodic orbits.

The second proposed method, referred to as the λ_1-signature, monitors changes in the stability of cycles which were extracted from the trajectory in the reconstructed state space. Performance of this method strongly depends on the accuracy of the estimation of the considered periodic orbits. From the comparison of the λ_1-signature with its one-dimensional implementation, the superior sensitivity of the former was concluded. In the last proposed method of condition monitoring, called orbital-*pdf*-signature, slices of the multivariate probability density function are taken along a set of extracted cycles. Even though its performance also depends on the accuracy of the estimation of the used cycles, this dependency seems to be less important than it was for the λ_1-signature. Another useful feature of this approach is that it allows localization of specific regions of the state space where the defect results in significant changes in the system behavior. This method also proved to be more sensitive in detecting both of the modeled types of defects than its one-dimensional implementation and than the conventional one-dimensional *pdf*-signature. It is important to emphasize that neither of the signatures employing periodic orbits require the system dynamics to be fully represented by its *cycle expansion* [3]. From our experience, we concluded that a small number of cycles usually suffices to make correct decisions about a possible change in the system operating conditions.

Even though most of the results presented throughout this work seem to be promising, they need to be verified for a large number of various types of gearboxes. Since we do not have the data obtained throughout the entire life of the gearbox until its failure, the above

results should also be verified by generating such data. Apart from gearboxes, there are other potential applications for this methodology, like the analysis of bearing vibrations. Since this approach is very general, it can be also used for any type of observed signal, for example an acoustic signal. One of the strengths of the state space reconstruction is that it provides information about the degrees of freedom other than the one measured. It does not mean, however, that it is restricted to analysis of univariate time series. One can think of a possible situation when both vibration and acoustic signals are acquired. Provided the multivariate mutual information function is used to reconstruct the state space, all of the analysis presented in this work can be directly applied to such a case as well.

The interesting new field of cycle expansions seems to provide many new opportunities for analysis of chaotic systems. Progress in this area should result in faster and more reliable methods of estimation of system invariants. An interesting and important application might be to use the cycle expansions for the purpose of probability density function estimation. To make this possible, the development of more reliable methods of extraction of periodic orbits from experimental time series is also necessary.

One of the conclusions that can be drawn from this work is that in many cases a hybrid of dynamical and probabilistic methods produced best results. It seems that future research in this area, including the application of Bayesian probabilistic methodology, should result in further improvements of the existing tools and development of new tools of analysis, forecasting, and system condition monitoring.

Last, but not least, a nonlinear state space approach clearly cannot replace the existing, often very powerful, methods of analysis of gearbox signals, but it offers some new possibilities to engineers and can prove to be of benefit in some difficult cases where nonlinear or chaotic system behavior is present, as inferred in the presented analysis of gearbox vibration signatures.

Bibliography

[1] H. D. I. Abarbanel, R. Brown, and M. B. Kennel. Local Lyapunov Exponents Computed form Observed Data. *J. Nonlinear Sci.*, 2:343–365, 1992.

[2] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. Sh. Tsimring. The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, 65:1331–1392, 1993.

[3] R. Artuso, E. Aurell, and P. Cvitanović. Recycling of Strange Sets: I. Cycle Expansions. *Nonlinearity*, 3:325–359, 1990.

[4] D. Auerbach, P. Cvitanović, J. P. Eckmann, G. Gunarante, and I. Procaccia. Exploring Chaotic Motion Through Periodic Orbits. *Physical Review Letters*, 58(23):2387–2389, 1987.

[5] J. L. Beck. System Identification Methods Applied to Measured Seismic Response. In *Proceedings 11th World Conference on Earthquake Engineering*, Acapulco, Mexico, June 1996.

[6] G. W. Blankenship and R. Singh. Analytical Solution for Modulation Sidebands Associated with a Class of Mechanical Oscillators. *Journal of Sound and Vibration*, 179(1):13–36, 1995.

[7] G. E. P. Box and F. M. Jenkins. *Time Series Analysis: Forecasting and Control.* Holden-Day, 1976.

[8] K. Briggs. An Improved Method for Estimating Liapunov Exponents of Chaotic Time Series. *Physics Letters A*, 151(1,2):27–32, 1990.

[9] D. S. Broomhead and G. P. King. Extracting Qualitative Dynamics from Experimental Data. *Physica D*, 20:217–236, 1986.

[10] T. K. Caughey. Derivation and Application of the Fokker-Planck Equation to Discrete Nonlinear Dynamic Systems Subjected to White Random Excitation. *Journal of the Acoustical Society of America*, 35:1683–1692, 1963.

[11] T. K. Caughey and S. M. Zaremba. Application of Dynamical Analysis to Noise Control. In R. J. Bernhard and J. S. Bolton, editors, *Proceedings of Inter-Noise 95 the 1995 International Congress on Noise Control Engineering*, pages 1191–1194, Newport Beach, California, July 1995. Noise Control Foundation.

[12] C. Chatfield. *The Analysis of Time Series.* Chapman and Hall, 1989.

[13] P. Cvitanović. Invariant Measurement of Strange Sets in Terms of Cycles. *Physical Review Letters*, 61(24):2729–2732, 1988.

[14] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle. Recurrence Plots of Dynamical Systems. *Europhysics Letters*, 4(9):973–977, 1987.

[15] J. D. Farmer and J. J. Sidorowich. Optimal Shadowing and Noise Reduction. *Physica D*, 47:373–392, 1991.

[16] A. M. Fraser and H. L. Swinney. Independent Coordinates for Strange Attractors from Mutual Information. *Phys. Rev. A*, 33:1134–1140, 1986.

[17] J. H. Friedman, J. L. Bentley, and R. A. Finkel. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Trans. on Math. Soft.*, 3(3):209–226, 1977.

[18] T. Frison. Chaos in Gearbox Vibrations. In *AIP Conference Proceedings 296*, pages 43–51. American Institute of Physics, 1993.

[19] R. G. Gallager. *Information Theory and Reliable Communication.* Wiley, 1968.

[20] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber. On Noise Reduction Methods for Chaotic Data. *Chaos*, 3(2):127–141, 1993.

[21] M. O. Hongler. Chaos in Mechanical Engineering Devices. *Helvetica Physica Acta*, 62:573–594, 1989.

[22] M. O. Hongler and L. Streit. On the Origin of Chaos in Gearbox Models. *Physica*, 29D:402–408, 1988.

[23] T. Kapitaniak. Quantifying Chaos with Amplitude Probability Density Function. *Journal of Sound and Vibration*, 114(3):588–592, 1987.

[24] K. Karagiannis and F. Pfeiffer. Theoretical and Experimental Investigations of Gear-Rattling. *Nonlinear Dynamics*, 2:367–387, 1991.

[25] M. B. Kennel and S. Isabelle. Method to distinguish possible chaos from colored noise and to determine embedding parameters. *Physical Review A*, 46(6):311–318, 1992.

[26] M. Koebbe and G. Mayer-Kress. Use of Recurrence Plots in the Analysis if Time Series Data. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, pages 361–378, Santa Fe, New Mexico, September 1990. Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publishing Company.

[27] E. J. Kostelich and T. Schreiber. Noise Reduction in Chaotic Time-Series Data: A Survey of Common Methods. *Physical Review E*, 48(3):1752–1763, 1993.

[28] E. J. Kostelich and J. A. Yorke. Noise Reduction: Finding the Simplest Dynamical System Consistent with the Data. *Physica D*, 41:183–196, 1990.

[29] Yu. A. Kravtsov. Randomness, Determinateness, and Predictability. *Sov. Phys. Usp.*, 32(5):434–449, 1989.

[30] D. P. Lathrop and E. J. Kostelich. Characterization of an Experimental Strange Attractor by Periodic Orbits. *Physical Review A*, 40(7):4028–4031, 1989.

[31] C. J. Li, J. Limmer, and J. Yoo. A Comparative Study of a Model-Based Demodulation Algorithm and Other Algorithms for Detection of Gear Defects. In R. J. Bernhard and J. S. Bolton, editors, *Proceedings of Inter-Noise 95 the 1995 International Congress on Noise Control Engineering*, pages 1143–1148, Newport Beach, California, July 1995. Noise Control Foundation.

[32] E. N. Lorenz. Deterministic Non-Periodic Flow. *J. Atmos. Sci.*, 20:130–141, 1963.

[33] P. F. Marteau and H. D. I. Abarbanel. Noise Reduction in Chaotic Time Series Using Scaled Probabilistic Methods. *Journal of Nonlinear Science*, 1:313–343, 1991.

[34] P. D. McFadden. Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration. *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, 108:165–170, 1986.

[35] F. C. Moon and T. Broschart. Chaotic Sources of Noise in Machine Acoustics. In *AIP Conference Proceedings 296*, pages 27–42. American Institute of Physics, 1993.

[36] S. Narayanan and P. Sekar. Bifurcation and Chaos in Contact Vibrations. In J. Awrejcewicz, editor, *Nonlinear Dynamics: New Theoretical and Applied Results*, pages 376–392, Berlin, 1995. Akademie Verlag.

[37] P. R. Nayak. Contact Vibrations. *Journal of Sound and Vibration*, 22(3):297–322, 1972.

[38] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing.* Prentice Hall, Englewood Cliffs, NJ, 1989.

[39] V. I. Oseledec. A Multiplicative Ergodic Theorem. Lyapunov Characteristic Numbers for Dynamical Systems. *Trudy Mosk. Mat. Obsc.*, 19(5):197–231, 1968.

[40] H. N. Özgüren and D. R. Houser. Mathematical Models Used in Gear Dynamics – a Review. *Journal of Sound and Vibration*, 121(3):383–411, 1988.

[41] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, 1991.

[42] F. Pfeiffer and A. Kunert. Rattling Models from Deterministic to Stochastic Processes. *Nonlinear Dynamics*, 1:63–74, 1990.

[43] H. Poincaré. *Les méthodes nouvelles de la mécanique céleste.* Gauthier-Villars, Paris, 1899.

[44] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C. The Art of Scientific Computing.* Cambridge University Press, 1992.

[45] O. E. Rössler. An Equation for Continuous Chaos. *Physics Letters*, 57 A(5):397–398, 1976.

[46] K. Sato, S. Yamamoto, and T. Kawakami. Bifurcation Sets and Chaotic States of a Gear System Subjected to Harmonic Excitation. *Computational Mechanics*, 7:173–182, 1991.

[47] T. Sauer. Time Series Prediction by Using Delay Coordinate Embedding. In A.S. Weigend and N.A. Gershenfeld, editors, *Time Series Prediction: Forecasting the Future and Understanding the Past*, pages 175–193, Santa Fe, New Mexico, May 1992. Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publishing Company.

[48] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics*, 65:579–616, 1991.

[49] B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 1993.

[50] F. Takens. Detecting Strange Attractors in Turbulence. In D. A. Rand and L. S. Young, editors, *Dynamical Systems and Turbulence*, pages 336–381, Berlin, 1981. Lecture Notes in Mathematics, Vol. 898, Springer-Verlag.

[51] W. W. Taylor. Application of Nonlinear Prediction to Signal Separation. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, pages 455–465, Santa Fe, New Mexico, September 1990. Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publishing Company.

[52] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer. Using Surrogate Data to Detect Nonlinearity in Time Series. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, pages 163–185, Santa Fe, New Mexico, September 1990. Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publishing Company.

[53] H. Tong and K. S. Lim. Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of Royal Statistical Society*, B 42:245–292, 1980.

[54] N. B. Tufillaro, T. Abbot, and J. P. Reilly. *An Experimental Approach to Nonlinear Dynamics and Chaos.* Addison-Wesley, 1991.

[55] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* Springer-Verlag, New York, 1990.

[56] J. Wright. Monitoring Changes in Time of Chaotic Nonlinear Systems. *Chaos*, 5(3):356–366, 1995.

[57] J. Wright and R. L. Schult. Recognition and Classification of Nonlinear Chaotic Signals. *Chaos*, 3:295–304, 1993.

[58] S. M. Zaremba. Theory of Liapunov Characteristic Exponents. Caltech memo, 1992.

[59] S. M. Zaremba, T. K. Caughey, and J. L. Beck. Dynamical Signature of Gearbox Vibration. submitted to: *Mechanical Systems and Signal Processing*, 1997.