Journal of Sound and Vibration (2002) **252**(3), 545–562
doi:10.1006/jsvi.2001.4148, available online at http://www.idealibrary.com on IDEAL®

QUALITATIVE AND QUANTITATIVE ANALYSIS OF STOCHASTIC PROCESSES BASED ON MEASURED DATA, I: THEORY AND APPLICATIONS TO SYNTHETIC DATA

J. GRADIŠEK AND I. GRABEC

Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva, 6, SI-1000 Ljubljana, Slovenia.
E-mail: janez.gradisek@fs.uni.lj.si

AND

S. SIEGERT AND R. FRIEDRICH

Institute for Theoretical Physics, University of Stuttgart, Pfaffenwaldring 57/5, D-70550 Stuttgart, Germany

(*Received* 30 *April* 2001, *and in final form* 1 *August* 2001)

Analysis of stochastic processes governed by the Langevin equation is discussed. The analysis is based on a general method for non-parametric estimation of deterministic and random terms of the Langevin equation directly from given data. Separate estimation of the terms corresponds to decomposition of process dynamics into deterministic and random components. Such decomposition provides a basis for qualitative and quantitative analysis of process dynamics. In Part I, the following analysis possibilities are described and illustrated using various synthetic datasets: (1) qualitative inspection of the estimated terms presented as fields, (2) reconstruction of the deterministic and stochastic evolution of the process and (3) approximation of the deterministic term by an analytical function and quantitative treatment of the equations obtained. In Part II, these analysis possibilities are applied to experimental datasets from metal cutting and laser-beam welding.
© 2002 Elsevier Science Ltd. All rights reserved.

1. INTRODUCTION

In recent decades, processes which generate non-periodic data have been studied intensively. Interest in these processes was fuelled mainly by the theory of deterministic chaos, which showed that non-periodic, even chaotic data can result from a non-linear deterministic process with only a few active degrees of freedom [1, 2]. Numerous analysis methods have been developed to extract meaningful information about the process from its chaotic data [3]. These methods require the data to be generated by a deterministic process, and allow only for negligible measurement noise uncorrelated to the process dynamics. Applicability of these methods to the analysis of data from a stochastic process of which noise is an integral part is limited. However, all experimental data are to some extent noisy, and it is usually difficult to distinguish between noisy chaotic data and stochastic data, which may also be corrupted by measurement noise. The problem is illustrated in Figure 1 using data from a forced oscillator. The phase portrait of the oscillations appears complicated, the time series of the displacement is non-periodic, and the associated power spectrum is broad. Since all these properties are also typical of chaotic data, one might assume that the oscillations are chaotic, and employ the analysis methods inspired by chaos theory [3]. In

0022-460X/02/$35.00 © 2002 Elsevier Science Ltd. All rights reserved.

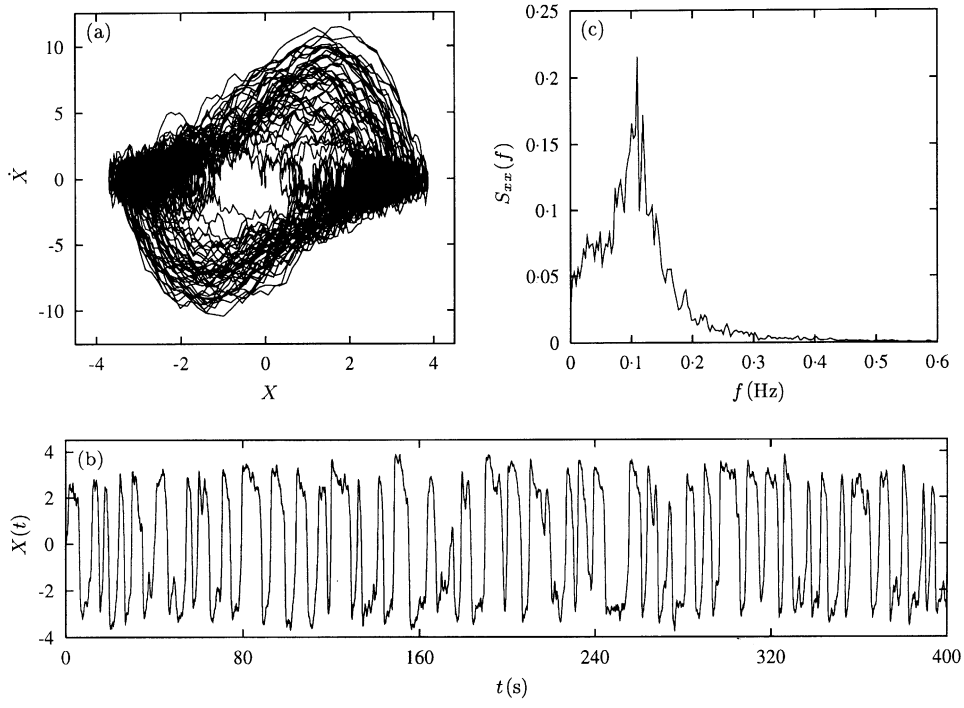

Figure 1. (a) Phase portrait, (b) time series of the displacement and (c) power spectrum of the displacement of a forced oscillator.

the present case, this assumption would be wrong because the data in fact result from a stochastic process, i.e., from the randomly forced van der Pol oscillator in a limit cycle regime (see section A.1 for details). Therefore, methods suited to stochastic data should be employed instead.

The example in Figure 1 belongs to a wide and important class of stochastic dynamic processes that can be described by the Langevin equation:

$$\mathrm{d}/\mathrm{d}t\ (\mathbf{X}(t)) = \mathbf{h}(\mathbf{X}(t)) + \mathbf{g}(\mathbf{X}(t))\mathbf{\Gamma}(t). \tag{1}$$

Here $\mathbf{X}(t)$ denotes the time-dependent d-dimensional stochastic variable which characterizes the process state completely. The evolution of $\mathbf{X}$ in time is governed by a sum of a deterministic term $\mathbf{h}$ and a random term $\mathbf{g} \cdot \mathbf{\Gamma}$. The random term consists of uncorrelated Gaussian white noise $\mathbf{\Gamma}$ and a $(d \times d)$ matrix of noise amplitudes $\mathbf{g}$. No restricting assumptions for $\mathbf{h}$ and $\mathbf{g}$ are necessary. $\mathbf{h}$ can be non-linear which means that deterministic chaos can also be modelled by equation (1). When noise amplitude $\mathbf{g}$ depends on the process state $\mathbf{X}(t)$, noise is of the multiplicative type, whereas constant $\mathbf{g}$ implies the additive type of noise. Note that noise in equation (1) does not affect the process parameters.

A general method for non-parametric estimation of the deterministic and random terms of the Langevin equation (1) has already been proposed [4, 5], and applied to synthetic and experimental datasets from medicine and engineering [6, 7]. The aim of this article is to present several possibilities the method offers for qualitative and quantitative analyses of stochastic data. For this purpose, the method is first reviewed briefly and shows how both the deterministic and random terms of equation (1) can be estimated from data, and inspected qualitatively. Since the terms in fact form a model of the process they can be employed to reconstruct either the deterministic or the stochastic evolution of the process. If equations are needed for the model, the deterministic term can be approximated by an

analytical function which can be further analyzed quantitatively. These analysis possibilities are illustrated by examples which include synthetic datasets from (1) the stochastic van der Pol oscillator, (2) a stochastic process exhibiting the sub-critical Hopf bifurcation and (3) the stochastic Lorenz system in a chaotic regime.

2. THE METHOD

As formulated in equation (1), the Langevin equation represents a model of a continuous stationary Markovian process. Stationarity is ensured by the fact that both the deterministic and random terms of equation (1), **h** and **g**, depend only on the state of the process $\mathbf{X}(t)$, and have no explicit time dependence. Due to the Markovian property, the state of the process at any time t depends only on the state at the preceding time $t - \tau$. The probability that the process trajectory visits location $\mathbf{x}_{n+1}$ at time $t + \tau$, given that it visits $\mathbf{x}_n$ at time t, is thus described by the conditional probability density $p(\mathbf{x}_{n+1}, t + \tau | \mathbf{x}_n, t)$, which is independent of the trajectory's path prior to time t.

The central idea of the method is the following [6]: every time t_i the process trajectory visits a point $\mathbf{x}$ in state space, the location of the trajectory at time $t_i + \tau$ is determined by a sum of the deterministic function $\mathbf{h}(\mathbf{x})$ and the stochastic function $\mathbf{g}(\mathbf{x})\,\boldsymbol{\Gamma}(t_i)$. Note that $\mathbf{h}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are constant for fixed $\mathbf{x}$, while $\boldsymbol{\Gamma}(t_i)$ is Gaussian distributed white noise. Based on this reasoning, it has been proven mathematically using Itô's definitions for stochastic integrals [8], that the terms **h** and **g** can be estimated by conditional averages as [4, 5]

$$\mathbf{h}(\mathbf{x}) = \lim_{\tau \to 0} \frac{1}{\tau} \langle \mathbf{X}(t + \tau) - \mathbf{x} \rangle |_{\mathbf{X}(t) = \mathbf{x}}, \tag{2a}$$

$$\mathbf{G}(\mathbf{x}) = \mathbf{g}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})^{\dagger} = \lim_{\tau \to 0} \frac{1}{\tau} \langle (\mathbf{X}(t + \tau) - \mathbf{x})(\mathbf{X}(t + \tau) - \mathbf{x})^{\dagger} \rangle |_{\mathbf{X}(t) = \mathbf{x}}, \tag{2b}$$

where †denotes the normal transpose. Given that the process trajectory $\mathbf{X}(t)$ visits a point $\mathbf{x}$ at time t, it follows from equation (2a) that the deterministic term **h** at $\mathbf{x}$ is obtained for small τ by a difference of the process states at times $t + \tau$ and t, averaged over an ensemble of process trajectories. In the ergodic case, the averaging can be performed over all times $t = t_i$ in a trajectory for which $\mathbf{X}(t_i) = \mathbf{x}$. The random term **G** is obtained in a similar way (2b). The deterministic and random components of the process dynamics can, therefore, be estimated for every point $\mathbf{x}$ in state space provided that the point is visited statistically often by the process trajectory $\mathbf{X}(t)$. Because of the limit $\tau \to 0$ in equation (2), it should be verified whether the estimated terms converge as the time step τ is decreased. Although the convergence rate in general depends on the properties of the process, one finds that estimates of the deterministic term **h** converge to their theoretical value at an order of magnitude larger time steps than estimates of the random term **G**. However, in practice, one is often faced with experimental data recorded with a sampling time too short to ensure convergence of the estimates. In this case, the estimated terms should be considered crude approximations and used cautiously.

The method is illustrated using the example in Figure 1, the stochastic van der Pol oscillator in a limit cycle regime (section A.1). Components of the estimated deterministic and random terms are shown in Figures 2 and 3, respectively, together with the corresponding theoretical values. In the region of state space visited by the process trajectory, the estimates agree well with theoretical values. In the rest of the state space, the terms **h** and **G** could not be estimated and were, therefore, set to zero.

For experimental data the theoretical values of **h** and **G** are usually not known, and estimates presented as in Figures 2 and 3 are not easy to interpret. More informative

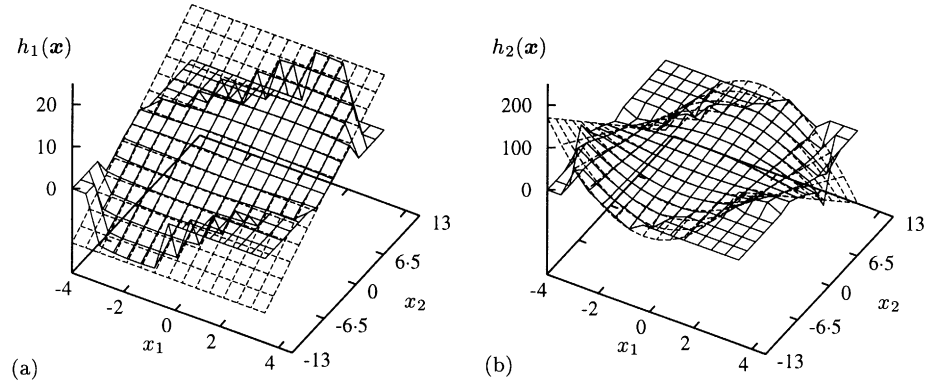

Figure 2. Comparison of estimated and theoretical deterministic terms for the randomly forced van der Pol oscillator. Solid grid, estimated values; dashed grid, theoretical values.

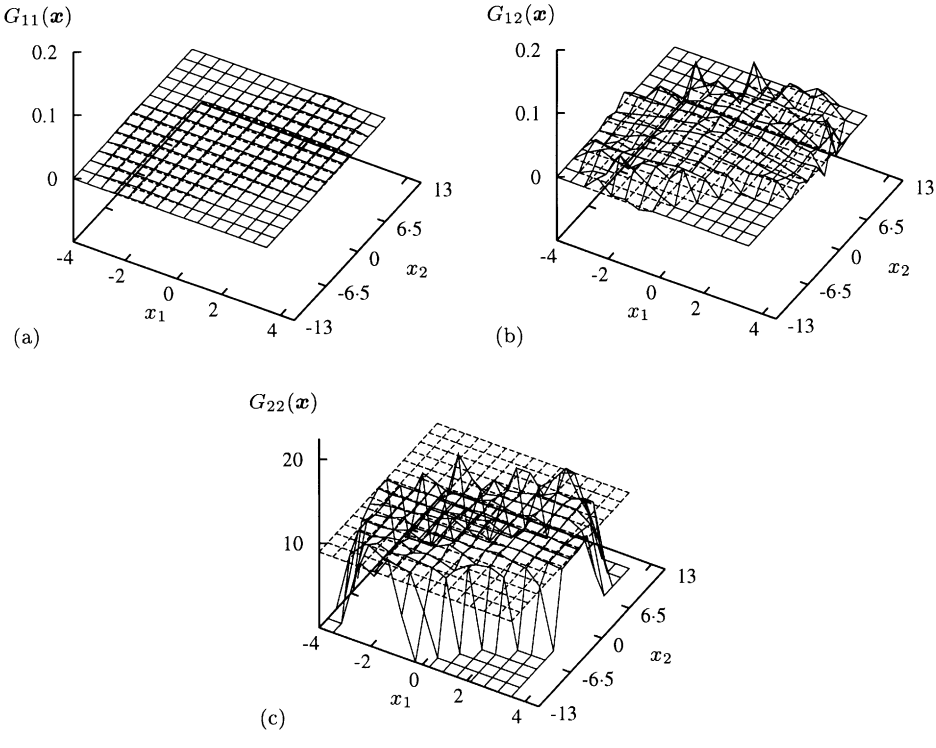

Figure 3. Comparison of estimated and theoretical random terms of the randomly forced van der Pol oscillator. Solid grid, estimated values; dashed grid, theoretical values.

presentation of the estimates, and possibilities for their qualitative and quantitative analysis are discussed below.

3. QUALITATIVE ANALYSIS

The qualitative analysis of the deterministic and random terms is presented below, and shows how the terms can be employed to numerically reconstruct the deterministic and stochastic trajectories of the process.

3.1. DETERMINISTIC TERM

The deterministic term **h** can be presented as a vector field in which an arrow at point **x** represents the value **h**(**x**). The length of the arrow is proportional to the average velocity of the deterministic flow, while the inclination of the arrow shows the average direction of the deterministic flow at point **x**. In the case of the van der Pol oscillator (Figure 4), the arrows in the field point on average in a clockwise direction and indicate motion in a stable non-symmetric limit cycle. The arrows outside the limit cycle run approximately in parallel to the limit cycle around most of the cycle, except for the right and the left corners where the arrows point towards the cycle. This suggests that the local dissipation is close to zero around most of the limit cycle, and strongly negative at the two corners. The arrows inside the limit cycle point outwards, indicating that an unstable fixed point is located at the center of the state space. These findings are confirmed by the trajectories superimposed on the vector field in Figure 4. The trajectories represent the deterministic motion of the oscillator reconstructed from the estimated term **h**. The reconstruction procedure is explained in section 3.3. At this point it is important to note that the trajectories which start at the edge or center of the state space both terminate on the limit cycle.

In addition to the direction and velocity of the deterministic flow, the term **h** shown as a field also provides information on local stability properties in the regions of state space explored by the process trajectory. This may be of particular importance when there are multiple co-existing stable attractors which the process trajectory can visit. Such a situation is typical of the sub-critical bifurcation phenomena where two stable attractors co-exist within the sub-critical parameter region [2]. Sub-critical bifurcation phenomena also occur in various mechanical systems [9–12], where one of the two stable attractors is often considered an unfavorable operating regime. In stochastic processes, in which noise frequently drives the trajectory from one attractor to the other, it may not be trivial to recognize the co-existing attractors from measured data. As an example consider the time series in Figure 5. It was generated by a stochastic system exhibiting a sub-critical Hopf

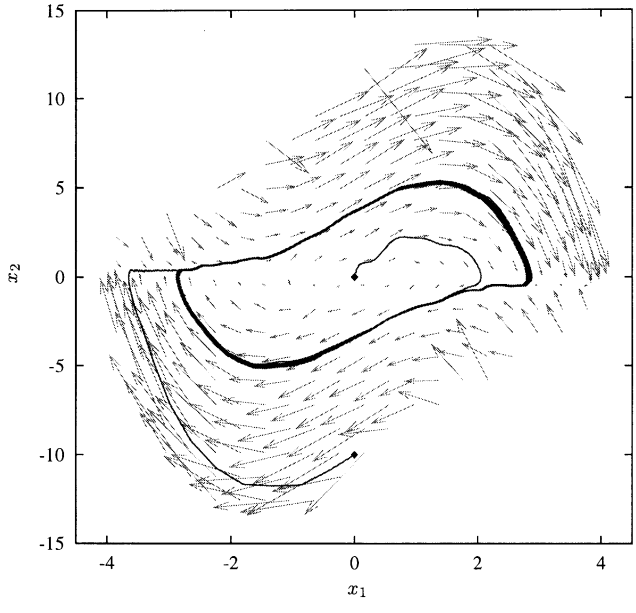

Figure 4. Estimated deterministic term **h** of the randomly forced van der Pol oscillator shown as a vector field with two reconstructed deterministic trajectories superimposed.

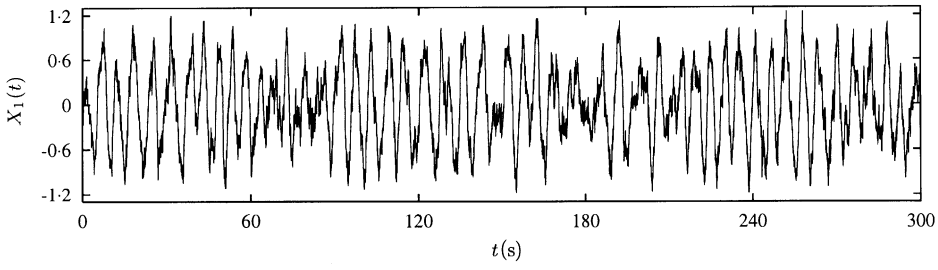

Figure 5. Time series from a stochastic system operating in the sub-critical region of a Hopf bifurcation.

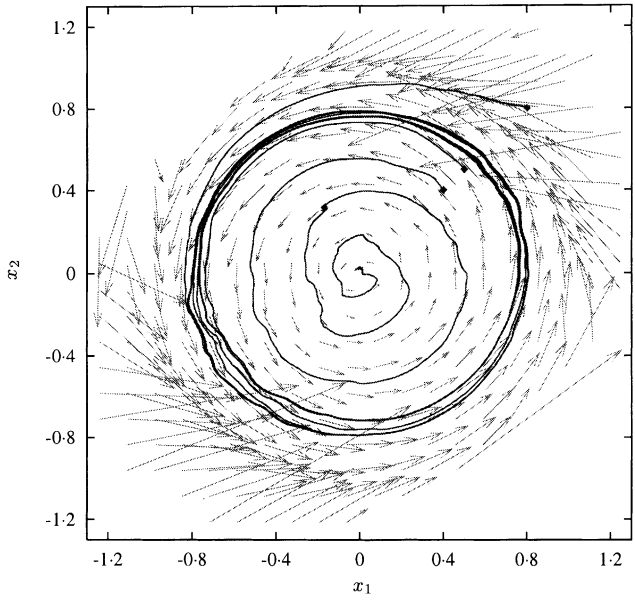

Figure 6. Estimated deterministic term **h** of the sub-critical Hopf bifurcation system shown as a vector field with three reconstructed deterministic trajectories superimposed.

bifurcation from a stable fixed point to a stable limit cycle, and operating in the sub-critical parameter region (see section A.2 for details).

While it is difficult to recognize from either the time series or the corresponding phase portrait (not shown) that the process trajectory visits two different attractors, both attractors can be found in the estimated deterministic field **h**(**x**) (Figure 6). Arrows in the inner region of the state space indicate a stable fixed point in the center, whereas arrows in the outer region indicate a stable limit cycle with motion in a counter clockwise direction. This is also established from the reconstructed deterministic trajectories superimposed on the vector field. Depending on the starting point, the trajectories terminate either on the limit cycle or at the fixed point. According to the Hopf bifurcation theory [2], there is an unstable limit cycle inside the stable limit cycle, separating the basins of attraction of the two stable attractors. Based on the vector field and the reconstructed trajectories one suspects that the unstable limit cycle in Figure 6 closely encircles the trajectory which spirals towards the fixed point.

To show that the method described in section 2 is not restricted to trivial attractors such as fixed points and limit cycles, it was applied to analyze data from the Lorenz system in

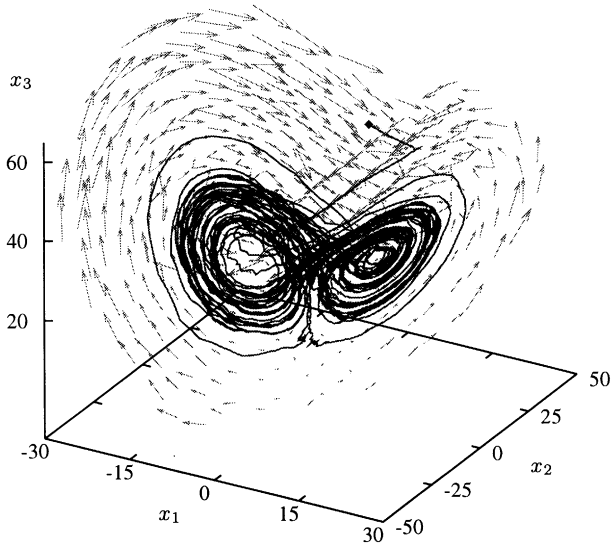

Figure 7. Estimated deterministic term **h** of the Lorenz system shown as a vector with a reconstructed deterministic trajectory superimposed.

a chaotic regime (section A.3). In the estimated deterministic field (Figure 7), a two-lobed structure is revealed, typical of the Lorenz system attractor. The inclination of the arrows in the field indicates that the trajectory spirals in a clockwise or counter clockwise direction on the left and the right lobe of the attractor respectively. The reconstructed deterministic trajectory, which is superimposed on the field in Figure 7, forms an attractor similar in size and shape to the original Lorenz attractor. These results confirm that the estimated term **h** indeed captures the main geometric and dynamic properties of the chaotic Lorenz system.

The Lorenz example also illustrates that three-dimensional fields can be difficult to inspect visually. Since visual inspection of the field can provide useful information about the process, one- or two-dimensional cross-sections can be used for analysis of high-dimensional fields [6].

3.2. RANDOM TERM

Unlike the deterministic term **h**, which is a vector, the estimated random term $\mathbf{G} = \mathbf{g}\mathbf{g}^{\dagger}$ is a matrix and therefore cannot be visualized directly as a vector field. A different approach is needed in order to meaningfully present the information contained in **G**. One possibility is to plot **G** as a field of parallelograms or ellipses. The directions and lengths of the sides of a parallelogram (principal axes of an ellipse) centered at **x** are defined by the eigendirections and eigenvalues of **G**(**x**) respectively. The size and shape of the parallelograms are related to the amplitude and the direction of noise. For the van der Pol oscillator, which was randomly forced by noise with a non-symmetric amplitude matrix $\mathbf{g} = [0{\cdot}4\ 0{\cdot}1;\ 0\ 2]$, the estimated random term is shown in Figure 8. Almost all parallelograms in the field have a similar size and shape, indicating that the matrix of noise amplitude **g** is constant across the entire attractor. The slightly different parallelograms at the edge of the attractor result from the statistically poorer estimate of **G** in that region. Vertical orientation of the parallelograms indicates that the acceleration of the oscillator is disturbed more than its velocity. For the example in Figure 1, where only the acceleration was disturbed by noise

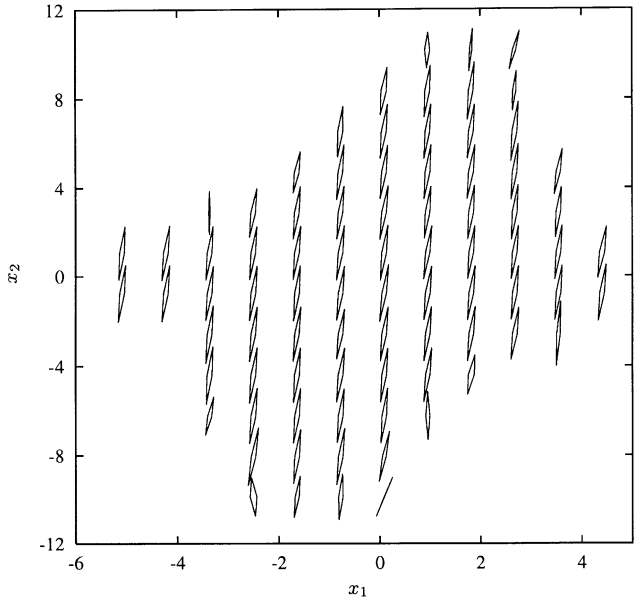

Figure 8. Estimated random term **G** of the randomly forced van der Pol oscillator shown as a field of parallelograms. The matrix of noise amplitudes **g** was non-symmetric.

with the amplitude matrix $\mathbf{g} = [0\ 0; 0\ 3]$, the parallelograms would be flattened to vertical lines.

3.3. RECONSTRUCTION OF PROCESS DYNAMICS

According to equation (1), the deterministic and random terms form a model of the process. Having extracted the two terms separately from data, one has therefore obtained a model which can be further used to numerically reconstruct the process dynamics. Adopting the Euler integration scheme, the process trajectory $\mathbf{X}(t)$ can be reconstructed as

$$X_i(t + \Delta t) = X_i(t) + h_i(\mathbf{X}(t))\Delta t + \sqrt{\Delta t} \sum_{j=1}^{i} g_{ij}(\mathbf{X}(t))\Gamma_j(t), \tag{3}$$

where **g** is obtained from the estimated **G** by the Cholesky decomposition [13, p. 96]. To fulfill the conditions for the decomposition, **g** is assumed to be of a lower triangular form. Consequently, the sum in equation (3) includes only the first i terms.

Both the deterministic and stochastic trajectories of the process can be reconstructed using equation (3). While the deterministic trajectories show how the process would evolve in the absence of random fluctuations, the stochastic trajectories represent the process evolution which can actually be observed. Examples of reconstructed deterministic trajectories are given in Figures 4–7. In Figure 9, the reconstructed stochastic trajectory of the displacement of the randomly forced van der Pol oscillator (bottom trace) is compared to the original stochastic trajectory (top trace) obtained by integrating the corresponding Langevin equation (section A.1). The trajectories are difficult to distinguish visually. They are similar both qualitatively and quantitatively, in terms of statistical characteristics (Figure 10). Note, however, that although they start from the same initial condition, the trajectories are not identical because they were generated using different noise time series $\Gamma(t)$.

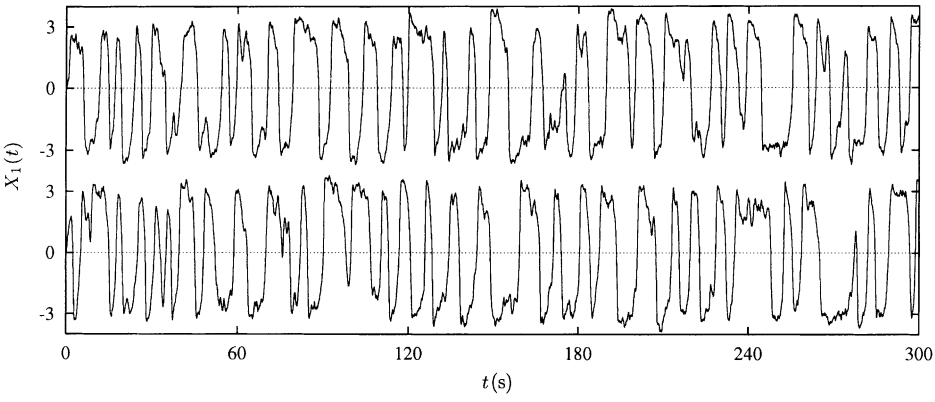

Figure 9. Original (top trace) and reconstructed stochastic trajectory (bottom trace) of the displacement X_1 of the randomly forced van der Pol oscillator.

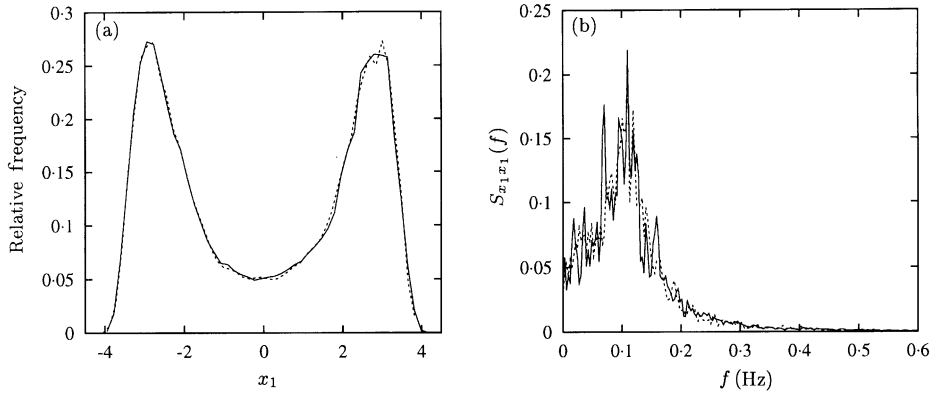

Figure 10. Comparison of (a) amplitude distributions and (b) power spectra of the reconstructed (—— line) and original trajectories (– – – line) of the displacement X_1 of the randomly forced van der Pol oscillator.

In Figure 11, the reconstructed deterministic and stochastic trajectories are shown, together with their original versions for the Lorenz system in a chaotic regime (section A.3). All trajectories start from the same initial condition. Comparison of the two original trajectories (top two traces) reveals the influence of noise on the process dynamics. One immediately observes that patterns of gradual increase of the oscillation maxima, which correspond to protracted spiralling of the process trajectory around one of the two lobes of the Lorenz attractor, occur frequently only in the deterministic trajectory (second trace from top). Despite the rare occurrence of these patterns in the stochastic trajectory, which are analyzed to extract the terms **h** and **G**, the patterns can be clearly observed in the reconstructed deterministic trajectory (third trace from top). The reconstructed deterministic trajectory thus preserves the main qualitative features of the original deterministic trajectory, although their paths are different. The differences can be attributed to the chaotic nature of the process and to imperfect estimates of the conditional moments in some regions of state space. The reconstructed stochastic trajectory (bottom trace) is similar to the original stochastic trajectory (top trace). They both exhibit similar spiky patterns and lack the pronounced spiralling feature.

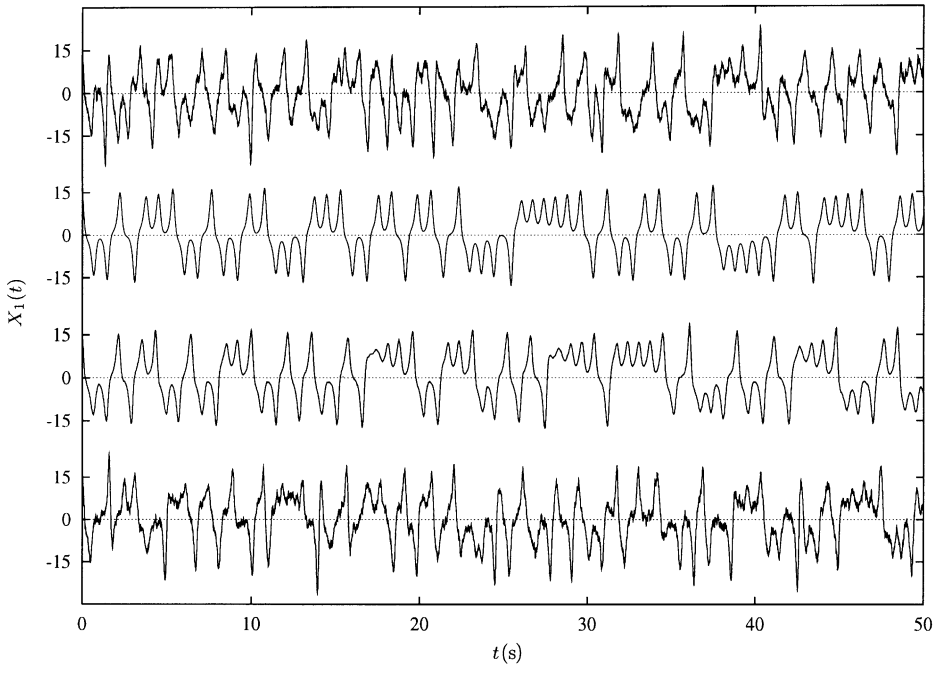

Figure 11. From top to bottom: original stochastic, original deterministic, reconstructed deterministic, and reconstructed stochastic time series of variable X_1 of the Lorenz system in a chaotic regime.

In order to verify agreement between the original and reconstructed deterministic attractors of the Lorenz system, their geometry and dynamics are compared quantitatively, by means of the correlation dimension and the maximal Lyapunov exponent. The correlation dimension ν measures dimensionality of a fractal set [1, 3]. The dimension of a set is determined from the $\nu(e)$ curve at the interval of lengths e where $\nu(e)$ is approximately constant. The $\nu(e)$ curve for the stochastic Lorenz attractor exhibits no constant plateau, whereas plateaus close to the theoretical value are found for both deterministic attractors (Figure 12(a)). In addition to the small difference between the plateau heights, the lengths of the plateaus are also different. The $\nu(e)$ plateau of the original deterministic attractor extends to shorter lengths e than the plateau of the reconstructed attractor. This indicates that the fine, fractal-like details of the Lorenz deterministic attractor could not be reconstructed from the stochastic data.

The Lyapunov exponents λ_i measure the average rate of exponential divergence of the nearby trajectories in time [1, 3]. Positive maximal Lyapunov exponent implies chaotic dynamics of the process. Figure 12(b) shows the average growth of distances $\Delta\mathbf{X}$ in time t within a cluster of nearby trajectories. The growth rate corresponds to the maximal Lyapunov exponent λ_1. In the stochastic Lorenz system, the distances $\Delta\mathbf{X}$ grow rapidly and reach the attractor extent within less than 1 s. In both deterministic systems, the distances $\Delta\mathbf{X}$ grow exponentially, with a rate similar to the theoretical one, and reach the attractor extent only after 4·5 s. These quantitative results confirm the geometric and dynamic similarity of the original and reconstructed deterministic attractors.

The reconstructed trajectories can be used for various purposes. The deterministic trajectories can be used to study the process as if it were not subject to random influences. The stochastic trajectories can be used as surrogate process trajectories. They can be employed when the amount of measured data available is insufficient for the task at hand,

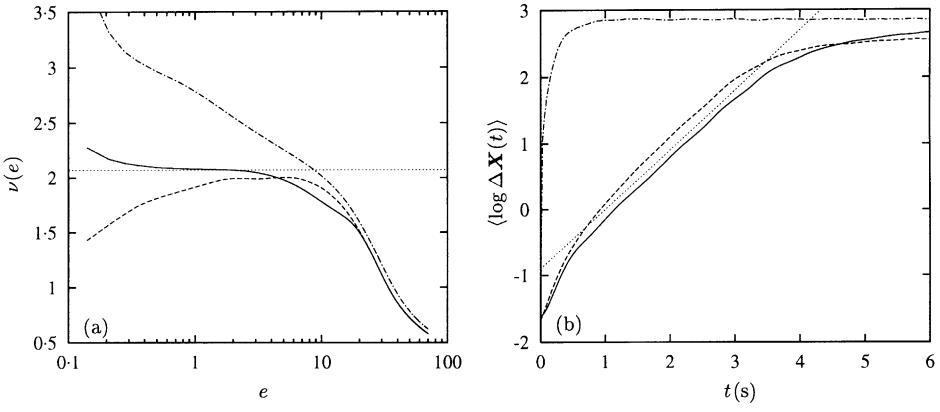

Figure 12. Comparison of estimated (a) correlation dimensions and (b) the maximal Lyapunov exponents for original deterministic (——), original stochastic (– · –), and reconstructed deterministic attractors (- - -). The dotted lines denote the theoretical values $\nu = 2{\cdot}06$ and $\lambda_1 = 0{\cdot}902\ \mathrm{s}^{-1}$.

and additional data is needed possessing deterministic and random properties similar to the original measured data. An example of such a task is the estimation of the mean first passage time when there are only few passages of interest observed in the measured data [7]. A similar situation can be encountered in experiments in which the long-term operation of a machine or a machine part is simulated based on short recorded signals of actual time-varying operating conditions.

In addition to qualitative analysis and description of a stochastic process, there is often also a need to analyze and characterize the process quantitatively. Some possibilities for quantitative analysis of processes using the method presented in section 2 are now discussed.

4. QUANTITATIVE ANALYSIS

Quantitative analysis is based on approximation of the estimated terms by a selected analytical function. In this study one analyzes the deterministic term **h** and approximate it by a polynomial

$$\frac{\mathrm{d}}{\mathrm{d}t} X_i = a_i^{(0)} + \sum_j a_{i,j}^{(1)} X_j + \sum_j \sum_k a_{i,jk}^{(2)} X_j X_k + \cdots \tag{4}$$

in which the coefficients $\mathbf{a}^{(i)}$ are obtained by a least-squares fit. While the coefficients themselves can serve as a quantitative measure for process characterization, the polynomial can be used as a model to generate approximate deterministic solutions of the Langevin equation (1), and to study the stability of these solutions. Examples of these applications are given below.

4.1. DETERMINISTIC MODEL

For the randomly forced van der Pol oscillator, a third order polynomial was used to approximate the estimated deterministic term **h**. The fitted coefficients $\mathbf{a}^{(i)}$ are listed in Table 1, together with their theoretical values. Agreement between the fitted (fit.) and the

TABLE 1

Coefficients $\mathbf{a}^{(i)}$ *of the third order polynomial for the van der Pol oscillator*

i		$a_i^{(0)}$	$a_{i,1}^{(1)}$	$a_{i,2}^{(1)}$	$a_{i,11}^{(2)}$	$a_{i,12}^{(2)}$	$a_{i,22}^{(2)}$	$a_{i,111}^{(3)}$	$a_{i,112}^{(3)}$	$a_{i,122}^{(3)}$	$a_{i,222}^{(3)}$
1	Theo.	0	0	1	0	0	0	0	0	0	0
	Fit.	10^{-4}	-10^{-4}	1.000	10^{-5}	-10^{-6}	-10^{-6}	10^{-5}	10^{-4}	10^{-5}	-10^{-6}
2	Theo.	0	-1	2	0	0	0	0	-1	0	0
	Fit.	$-0{\cdot}010$	$-0{\cdot}827$	$2{\cdot}039$	$0{\cdot}012$	$-0{\cdot}010$	$0{\cdot}002$	$-0{\cdot}033$	$-1{\cdot}011$	-10^{-4}	-10^{-4}

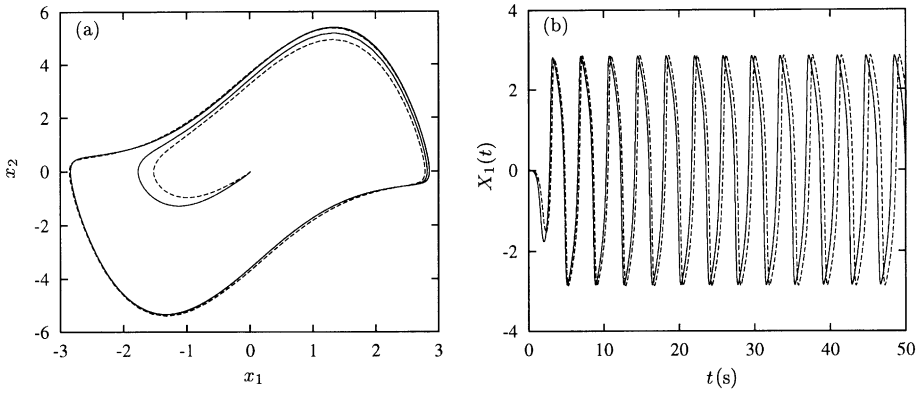

Figure 13. Comparison of the original (– – – line) and the approximate deterministic trajectories (—— line) of the van der Pol oscillator (a) in the state space, and (b) versus time.

theoretical (theo.) coefficients is very good, since **h** of the van der Pol oscillator is in fact described by a third order polynomial (section A.1).

In Figure 13, the original deterministic trajectory is compared to the approximate deterministic trajectory generated by numerically integrating the polynomial obtained (4). Notable differences between the two trajectories in state space are observed only during the transient time, when the trajectories, which start at the center of the state space, approach the corresponding limit cycle attractors (Figure 13(a)). The shapes of the two attractors are very similar. However, when the trajectories are compared as time series, a slight mismatch in the frequency of the oscillations is found (Figure 13(b)).

For the stochastic Lorenz system in a chaotic regime, the coefficients $\mathbf{a}^{(i)}$ of the third order polynomial approximating the estimated **h** are given in Table 2. Except for $i = 3$, the fitted coefficients do not agree with their theoretical values. The main reason for the disagreement is the fact that **h** of the Lorenz system is described by a second, and not third, order polynomial (section A.3). When a second order polynomial is used to approximate **h**, the agreement between the fitted and theoretical values is indeed much better (Table 3).

Despite significant differences between the theoretical and fitted coefficients of the third order polynomial, the attractor formed by the approximate deterministic trajectory is similar to the original deterministic Lorenz attractor (Figure 14(a)). Similar geometry of the original and the two approximate attractors was confirmed using the correlation dimension ν. The correlation dimension curves $\nu(e)$ of the two approximate attractors closely follow the $\nu(e)$ curve of the original Lorenz attractor (Figure 14(b)). As expected, slightly better agreement with the theoretical $\nu(e)$ curve is achieved by the second order approximation, although all three curves give a correct estimate of the dimension of the Lorenz attractor,

TABLE 2

Coefficients $\mathbf{a}^{(i)}$ *of the third order polynomial for the Lorenz system*

i		$a_i^{(0)}$	$a_{i,1}^{(1)}$	$a_{i,2}^{(1)}$	$a_{i,3}^{(1)}$	$a_{i,11}^{(2)}$	$a_{i,12}^{(2)}$	$a_{i,13}^{(2)}$	$a_{i,22}^{(2)}$	$a_{i,23}^{(2)}$	$a_{i,33}^{(2)}$
1	Theo.	0	−10	10	0	0	0	0	0	0	0
	Fit.	−0·006	1·947	3·709	0·313	−0·070	0·072	−0·554	−0·017	0·273	−0·014
2	Theo.	0	28	−1	0	0	0	−1	0	0	0
	Fit.	−0·190	5·920	10·972	−0·721	0·071	−0·094	0·025	0·049	−0·530	0·033
3	Theo.	0	0	0	−2·667	0	1	0	0	0	0
	Fit.	−0·290	0·101	0·047	−2·717	−0·018	1·002	−0·001	0·013	−0·009	0·003

i		$a_{i,111}^{(3)}$	$a_{i,112}^{(3)}$	$a_{i,113}^{(3)}$	$a_{i,122}^{(3)}$	$a_{i,123}^{(3)}$	$a_{i,133}^{(3)}$	$a_{i,222}^{(3)}$	$a_{i,223}^{(3)}$	$a_{i,233}^{(3)}$	$a_{i,333}^{(3)}$
1	Theo.	0	0	0	0	0	0	0	0	0	0
	Fit.	−0·001	-10^{-4}	0·002	0·001	−0·002	0·006	10^{-6}	10^{-4}	−0·003	10^{-4}
2	Theo.	0	0	0	0	0	0	0	0	0	0
	Fit.	0·001	10^{-4}	−0·002	−0·001	0·002	−0·012	10^{-4}	−0·001	0·006	-10^{-4}
3	Theo.	0	0	0	0	0	0	0	0	0	0
	Fit.	0·001	−0·001	0·001	10^{-4}	-10^{-4}	-10^{-4}	-10^{-4}	-10^{-4}	-10^{-4}	-10^{-4}

TABLE 3

Coefficients $\mathbf{a}^{(i)}$ *of the second order polynomial for the Lorenz system*

i		$a_i^{(0)}$	$a_{i,1}^{(1)}$	$a_{i,2}^{(1)}$	$a_{i,3}^{(1)}$	$a_{i,11}^{(2)}$	$a_{i,12}^{(2)}$	$a_{i,13}^{(2)}$	$a_{i,22}^{(2)}$	$a_{i,23}^{(2)}$	$a_{i,33}^{(2)}$
1	Theo.	0	−10	10	0	0	0	0	0	0	0
	Fit.	0·306	−9·932	9·975	−0·022	0·001	−0·002	−0·002	0·001	0·001	10^{-4}
2	Theo.	0	28	−1	0	0	0	−1	0	0	0
	Fit.	0·277	27·994	−0·968	−0·027	0·002	−0·005	−1	0·002	-10^{-4}	10^{-4}
3	Theo.	0	0	0	−2·667	0	1	0	0	0	0
	Fit.	0·055	0·038	−0·006	−2·678	−0·003	0·992	−0·001	0·004	−0·001	-10^{-4}

$\nu \approx 2{\cdot}06$. Similarity of the dynamics of the three deterministic systems was verified using the Lyapunov exponents λ_i. The exponents calculated directly from the corresponding equations of motion [1] are listed in Table 4. Excellent agreement with all theoretical λ_i is achieved by the second order approximation. For the third order approximation, the agreement of λ_1 and λ_2 is still very good, while λ_3 differs considerably from its theoretical value.

4.2. LINEAR STABILITY ANALYSIS

The obtained coefficients $\mathbf{a}^{(i)}$ can be used to assess the stability of a particular deterministic solution $\mathbf{X}^*$ of equation (1). Here, the analysis is restricted to that of linear stability, which is determined by the eigenvalues s_i of the Jacobian matrix evaluated at $\mathbf{X}^*$. Analysis of linear stability is illustrated by two examples in which the fixed point $\mathbf{X}^* = \mathbf{0}$ loses stability via a Hopf bifurcation as the control parameter becomes positive. For the van der Pol oscillator (section A.1), the dependence of the pair $s_{1,2}$ on the control parameter ε is shown in Figure 15. As expected, the real part of $s_{1,2}$, which determines the stability of the fixed point, is negative for $\varepsilon < 0$ and positive for $\varepsilon > 0$ (Figure 15(a)). This means that the

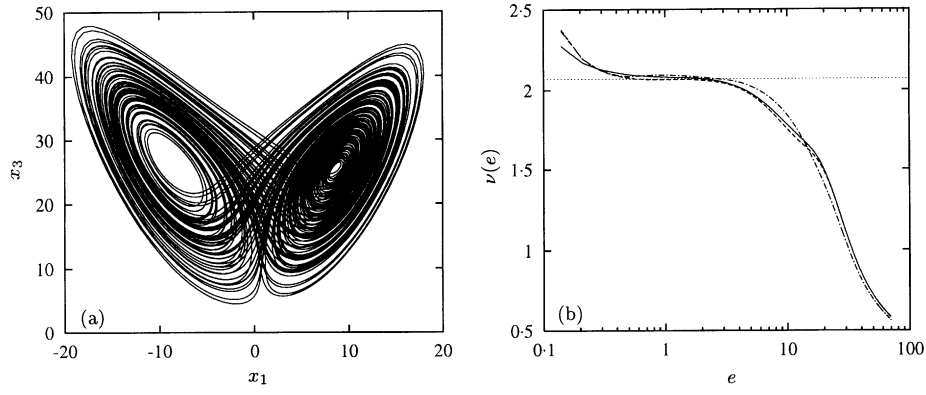

Figure 14. (a) Projection of the approximate deterministic attractor of the Lorenz system using a third order polynomial. (b) Comparison of the correlation dimension estimates for the original deterministic attractor (——), and for the two approximate attractors: second order approximation (– – – – line); third order approximation (– · –). The dotted line denotes the theoretical value of ν.

TABLE 4

Lyapunov exponents of original and approximate deterministic Lorenz systems

System	λ_1	λ_2	λ_3
Original	0·902	− 0·001	− 14·57
Second order approx.	0·901	− 0·002	− 14·53
Third order approx.	0·886	0·002	− 8·995

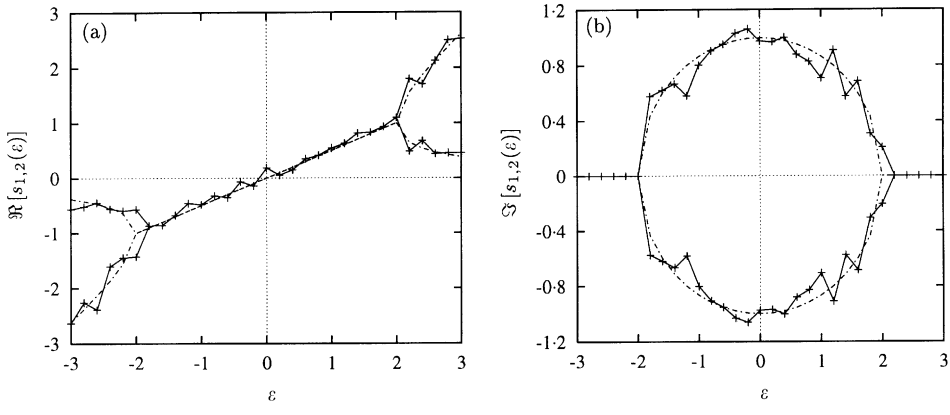

Figure 15. Dependence of the estimated eigenvalues $s_{1,2}$ on the parameter ε for the fixed point of the randomly forced van der Pol oscillator. (a) real part and (b) imaginary part of the eigenvalues. The deterministic term was approximated by third order polynomials. The theoretical dependences are shown by – · – lines.

fixed point is stable for $\varepsilon < 0$ and unstable for $\varepsilon > 0$. The dependences of real and imaginary parts of $s_{1,2}(\varepsilon)$ are both in good agreement with the corresponding theoretical dependences calculated from equation (A.1).

The situation is different for the stochastic system exhibiting a sub-critical Hopf bifurcation (Appendix A.2, Figure 16). The dependence of $\Re[s_{1,2}(\varepsilon)]$, obtained by a third

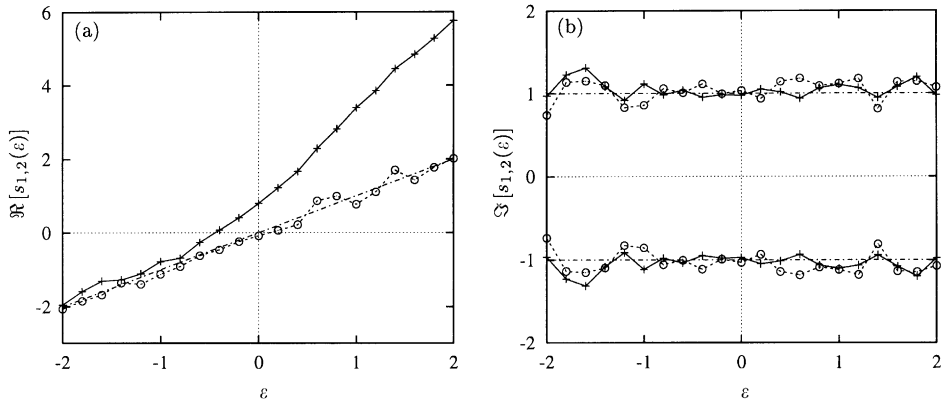

Figure 16. Dependence of the estimated eigenvalues $s_{1,2}$ on the parameter ε for the fixed point of the sub-critical Hopf bifurcation system. (a) Real part and (b) imaginary part of the eigenvalues. The deterministic term was approximated by third order polynomials (——, +) and fifth order polynomials (dashed line, ◦). The theoretical dependences are shown – · – lines.

order approximation (solid line in Figure 16(a)), indicates that the fixed point becomes unstable at $\varepsilon \approx -0{\cdot}4$. This result is incorrect since it follows from the theory that the fixed point loses stability at $\varepsilon = 0$. The error stems from poor approximation of the estimated **h**, which is in fact described by a fifth order polynomial (A.3). When the estimated **h** is approximated by a fifth order polynomial, the dependence of $\Re[s_{1,2}(\varepsilon)]$ agrees well with the theoretical one (dash–dotted line in Figure 16(a)). Such information about **h** is usually not available for experimental data. In this case, one should examine different orders of the approximating polynomial to find the one which fits the estimated deterministic term best and generates trajectories which follow closely those reconstructed from the deterministic term.

5. DISCUSSION AND CONCLUSIONS

Stochastic processes governed by the Langevin equation are analyzed. Analysis is based on estimation of the deterministic and random terms of the Langevin equation directly from data. The estimation method is non-parametric in the sense that no functional form need be assumed in advance for the estimated terms. The main disadvantage of the estimation method is that it requires rather densely sampled data [14]. There are approaches to parameter estimation in stochastic differential equations which are less demanding in terms of the sampling rate of the data than the method discussed in this article, but they are parametric and depend crucially on the assumed functional form [15, 16]. It, therefore, seems that oversampling is the price to be paid for a non-parametric estimate.

The estimated deterministic and random terms are presented as fields and inspected visually. Inspection of the deterministic term can yield information about the average direction and velocity of the deterministic flow, whereas information about the amplitude and the direction of noise can be extracted from the random term. In the examples presented, the terms were estimated in the state space spanned by all state variables. In experimental situations, when not all state variables are measurable, the terms can be estimated in space reconstructed from the time series of measured variables related to the process dynamics [3]. While most of the information provided by the estimated terms is preserved in the reconstructed space, information about the direction of noise is lost, since it is not possible to determine which state variable is originally influenced by noise.

The estimated terms of the Langevin equation constitute a model which is used to reconstruct the deterministic and stochastic trajectories of the process. The deterministic trajectories show the hypothetical process evolution in the absence of random noise. Although remotely similar, the deterministic reconstruction is not equivalent to filtering dynamic noise, since the process would evolve differently under the same deterministic laws if noise was present. The stochastic trajectories show a realistic process evolution which could actually be observed. They can be employed as surrogates in various situations when the length of the recorded original trajectories is insufficient for a particular task [7].

Quantitative analysis of stochastic processes is based on approximation of the estimated deterministic term by an analytical function. In the present study a polynomial is used, and the corresponding coefficients are obtained by a least-squares fit. The polynomial is employed to generate the approximate deterministic trajectories of the process and to assess their linear stability. The order of the approximating polynomial should be selected to be consistent with the estimated deterministic term. Consistence can be checked by comparing the deterministic trajectories reconstructed using the estimated deterministic term with those obtained by integrating the approximating polynomial. Several orders should be examined to find the most appropriate one. The Lorenz example reveals that overestimating the order of approximation preserves qualitative and, to a large extent, also quantitative properties of the system, although the polynomial coefficients do not agree with the theoretical ones. Moreover, sensitivity of the approximation and the amount of data required both increase with the order of approximating polynomial. On the other hand, the sub-critical bifurcation example shows that underestimating the order of approximation does not entirely capture the process properties. Note, finally, that the polynomial may not necessarily be the correct choice for the approximating function.

In all examples presented, noise amplitude $\mathbf{g}$ was constant, which means that noise was of the additive type. As shown in reference [17], the same formulae (2) for estimating the deterministic and random terms apply also for the multiplicative type of noise caused by dependence of the noise amplitude $\mathbf{g}$ on the process state $\mathbf{X}(t)$.

Some of the analysis possibilities discussed in this paper have already been applied to various experimental datasets from medicine [6, 18] and engineering [6, 19]. In Part 2 of the paper [20] these analysis possibilities are used to analyze experimental data from metal cutting and laser-beam welding.

ACKNOWLEDGMENTS

J.G. and I.G. gratefully acknowledge the support of the Volkswagen Foundation, the Ministry of Science and Technology of Slovenia, and the EU COST Action P4.

REFERENCES

1. E. OTT 1993 *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press.
2. S. H. STROGATZ 1994 *Nonlinear Dynamics and Chaos*. Reading Massachusetts: Addison-Wesley.
3. H. KANTZ and T. SCHREIBER 1997 *Nonlinear Time Series Analysis*, Vol. 7 *of Cambridge Nonlinear Science Series*, Cambridge: Cambridge University Press.
4. R. FRIEDRICH and J. PEINKE 1997 *Physica* **D 102**, 147–155. Statistical properties of a turbulent cascade.
5. S. SIEGERT, R. FRIEDRICH and J. PEINKE 1998 *Physics Letters* **A 243**, 275–280. Analysis of data sets of stochastic systems.
6. R. FRIEDRICH, S. SIEGERT, J. PEINKE, S. LÜCK, M. SIEFERT, M. LINDEMANN, J. RAETHJEN, G. DEUSCHL and G. PFISTER 2000 *Physics Letters* **A 271**, 217–222. Extracting model equations from experimental data.

7. J. GRADIŠEK, S. SIEGERT, R. FRIEDRICH and I. GRABEC 2000 *Physical Review* **E 62**, 3146–3155. Analysis of time series from stochastic processes.
8. H. RISKEN 1989 *The Fokker–Planck Equation*. Berlin: Springer-Verlag.
9. T. KALMÁR-NAGY, J. R. PRATT, M. A. DAVIES and M. D. KENNEDY 1999 *Proceedings of the* 17*th ASME Biennial Conference on Mechanical Vibration and Noise* 1999. *ASME Design and Technical Conferences*, 1–9 *Las Vegas, Nevada*. Experimental and analytical investigation of the subcritical instability in metal cutting.
10. J. GRADIŠEK, E. GOVEKAR and I. GRABEC 2001 *Journal of Sound and Vibration* **242**, 829–838. Chatter onset in non-regenerative cutting: a numerical study.
11. M. G. GOMAN, G. I. ZAGANOV and A. V. KHRAMTSOVSKY 1997 *Progress in Aerospace Science* **33**, 539–586. Application of bifurcation methods to nonlinear flight dynamics problems.
12. L. E. CHRISTIANSEN, T. LEHN-SCHIØLER, E. MOSEKILDE, P. GRANASY and H. MATSUSHITA 2002 *Mathematics and Computers in Simulation* **58**, 385–405. Nonlinear characteristics of randomly excited transonic flutter.
13. W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING and B. P. FLANNERY 1994 *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge: Cambridge University Press; second edition.
14. J. TIMMER 2000 *Chaos, Solitons and Fractals* **11**, 2571–2578. Parameter estimation in nonlinear stochastic differential equations.
15. B. M. BIBBY and M. P. SØRENSEN 1995 *Bernoulli* **1**, 17–39. Martingale estimation functions for discretely observed diffusion processes.
16. R. MEYER and N. CHRISTENSEN 2000 *Physical Review* **E 62**, 3535–3542. Bayesian reconstruction of chaotic dynamical systems.
17. S. SIEGERT and R. FRIEDRICH 2001 *Physical Review* **E 64**. To appear. Modeling of nonlinear Lévy processes by data analysis.
18. A. S. PIKOVSKY and M. G. ROSENBLUM 2001 Private communication.
19. J. GRADIŠEK, S. SIEGERT, R. FRIEDRICH and I. GRABEC 2001 *Mechanical Systems and Signal Processing*. To appear. Stochastic dynamics of metal cutting: bifurcation phenomena in turning.
20. J. GRADIŠEK, E. GOVEKAR and I. GRABEC 2002 *Journal of Sound and Vibration* **252**, 563–570. Submitted. Qualitative and quantitative analysis of stochastic processes based on measured data—II. Applications to experimental data.

APPENDIX A: EQUATIONS AND PARAMETERS OF THE SYSTEMS

All systems considered in the present study can be modelled by a Langevin equation (1). For the sake of brevity, only the deterministic terms **h** and the amplitudes **g** of the random terms of the corresponding Langevin equations are listed.

A.1. VAN DER POL OSCILLATOR

Dynamics of the van der Pol oscillator are governed by

$$\dot{X}_1 = X_2, \qquad \dot{X}_2 = (\varepsilon - X_1^2)X_2 - X_1, \tag{A.1}$$

$\varepsilon = 2$ was chosen. Unless stated otherwise, the noise amplitudes were

$$\mathbf{g} = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} \tag{A.2}$$

which corresponds to randomly disturbing the acceleration of the oscillator. The integration time step was 0·005 s.

A.2. SUB-CRITICAL HOPF BIFURCATION

A process exhibiting a sub-critical Hopf bifurcation is modelled by

$$\dot{Z} = (\varepsilon + \mathrm{i}\omega)Z + \delta|Z|^2 Z - |Z|^4 Z, \tag{A.3}$$

with $Z = X_1 + iX_2$. $\varepsilon = -0{\cdot}24$ and $\omega = \delta = 1$ were chosen. The noise amplitudes were

$$\mathbf{g} = \begin{bmatrix} 0{\cdot}3 & 0{\cdot}2 \\ 0{\cdot}2 & 0{\cdot}1 \end{bmatrix} \tag{A.4}$$

and the integration time step was 0·01 s.

A.3. LORENZ SYSTEM

The Lorenz system is governed by the system of equations

$$\dot{X}_1 = \sigma(X_2 - X_1), \qquad \dot{X}_2 = X_1(r - X_3) - X_2, \qquad \dot{X}_3 = X_1X_2 - bX_3. \tag{A.5}$$

The parameters $\sigma = 10$, $r = 28$, and $b = \frac{8}{3}$ were chosen to ensure a chaotic regime of the deterministic process. The noise amplitudes were

$$\mathbf{g} = \begin{bmatrix} 4 & 5 & 3 \\ 5 & 5 & 6 \\ 3 & 6 & 10 \end{bmatrix} \tag{A.6}$$

and the integration time step was 0·01 s.