Identification methods for nonlinear stochastic systems

Jose-Maria Fullana* and Maurice Rossi
Laboratoire de Modélisation en Mécanique UMR 7607, Université Paris VI, 4 Place Jussieu, 75005 Paris, France
(Received 25 July 2001; published 28 February 2002)

Model identifications based on orbit tracking methods are here extended to stochastic differential equations. In the present approach, deterministic and statistical features are introduced *via* the time evolution of ensemble averages and variances. The aforementioned quantities are shown to follow deterministic equations, which are explicitly written within a linear as well as a weakly nonlinear approximation. Based on such equations and the observed time series, a cost function is defined. Its minimization by simulated annealing or backpropagation algorithms then yields a set of best-fit parameters. This procedure is successfully applied for various sampling time intervals, on a stochastic Lorenz system.

DOI: 10.1103/PhysRevE.65.031107 PACS number(s): 05.40.−a, 05.45.−a

I. INTRODUCTION

In various areas, such as geophysics, engineering, computing as well as economics or biological sciences, identification techniques are used to build directly from experimental data, models which best reconstruct measurements [1–3]. For parametric models, this inverse problem generally reduces to the optimization of a cost function $\mathcal{J}$. This quantity, which determines the agreement between experiment and model, is selected according to the phenomenology and the available experimental data. For instance, the function $\mathcal{J}$ may be defined as the Euclidean distance between observations of a given vector field and its respective prediction. Note that the predicted quantity is, generally, an implicit function of model parameters since it is computed through a numerical integration of the parametric model.

On general grounds, predictions or characterization procedures are sensitive to the presence of noise. This effect is generally studied in the framework of the so-called observational (i.e., measurement) noise problem [2–4]. In this case, the dynamics itself is not altered but noise is added to measurements because of experimental inaccuracies. Various means [2–4] of removing this type of noise may be considered [19]. Within the context of parametric deterministic models, the backpropagation algorithm [5,6] allows, even in the presence of measurement noise, an efficient and accurate computation of gradients of the cost function with respect to model parameters. This technique may thus be reliably coupled to an improved Newton procedure [7] to search for the minimum of function $\mathcal{J}$. Contrary to the classical one time-step algorithm, this approach does not assume that sampling time and discretization time are equal and it is less sensitive to measurement noise [8]. The simulated annealing procedure [9,10], constitutes another strategy to find a minimum of $\mathcal{J}$. This latter method, which is more time consuming than the backpropagation algorithm, may be applied to cases in which local minima are present.

Model identification is much more intricated when noise is built in the dynamics itself [11,12], i.e., added to the governing equations. In this paper, we focus on this so-called dynamical noise or, otherwise stated, we are interested in the identification of stochastic differential equations [13–15] within the context of parametric models. In an earlier paper [16], we considered a purely statistical approach, which could identify the parameters even for significant noise amplitude and for sampling time interval greater than the correlation time. The data set, however, was necessarily quite large. By contrast, the method proposed here works with a smaller data set, but it is based on the assumption that measurement sampling time is small compared to the system characteristic time. In this context, statistical variances remain of weak amplitude and the classical backpropagation algorithm or simulated annealing method based on orbit tracking is extended to the case of stochastic systems. This requires several modifications since (i) the notion of deterministic true orbit is lost and (ii) noise directly affects the dynamics. Finally, an appropriate cost function, which contains both deterministic and statistical features, should be defined. This new cost function will be then introduced into an optimization procedure based on backpropagation or simulated annealing methods.

The paper is organized as follows. In Sec. II, we formalize the dynamical noise problem and we recall how it is related to the stochastic equation theory. In order to model the time evolution of ensemble averages and variances, linear and weakly nonlinear approximations are proposed in Sec. III. This approach, in particular the closure hypothesis of the weakly nonlinear case, is then numerically tested on a stochastic Lorenz system. In Sec. IV, an identification procedure is described based on a cost function $\mathcal{J}$ containing deterministic as well as statistical aspects. A simulated annealing technique finds a minimum of function $\mathcal{J}$ if ensemble averages and variances are assumed to follow the above weakly nonlinear equations. For the linear approximation instead a backpropagation algorithm is used. In Sec. V, results of the identification method are discussed on the specific case of a stochastic Lorenz system.

II. PROBLEM FORMULATION

Consider a model characterized by N variables x_i $(i=1,...,N)$, which satisfy a set of N nonlinear Langevin equations

*Present address: Service de Biophysique, Laboratoires Innothera, 7-9 Av. F. V. Raspail, 9411 Arcueil, France.

$$\frac{dx_i}{dt} = F_i(x_1,\ldots,x_N;\mu_1,\ldots,\mu_P) + \tau_i\theta_i(t), \tag{1}$$

depending on P parameters μ_j $(j=1,...,P)$. Functions $\theta_i(t)$ stand for uncorrelated normalized white noises

$$\langle \theta_i(t)\,\theta_j(t')\rangle = 2\,\delta_{ij}\delta(t-t'), \tag{2}$$

and τ_i for noise amplitudes. The quantity δ_{ij} denotes the Kronecker symbol, $\delta(t-t')$ the Dirac distribution and $\langle\rangle$ an ensemble average. The proposed identification procedure retrieves, from a data set

$$\mathcal{S} = \{x_i^{\rm obs}(t_k), i=1,...,N | t_k = kT, k=1...M+1\}, \tag{3}$$

produced by $M+1$ measurements performed at every sampling time interval T, a set of best-fit values for parameters $\overline{a} = (\mu_1,\ldots,\mu_P;\tau_1,\ldots,\tau_N)$. The data set $\mathcal{S}$ may be altered by measurement noise as well. In the following, this type of noise will be assumed weak compared to dynamical noise in order to focus on this particular case.

Since dynamical noise plays an intrinsic role in Langevin equations (1), stochastic and deterministic behaviors should be introduced in the reconstruction algorithm both at the level of cost function $\mathcal{J}$ and optimization procedure. More precisely, function $\mathcal{J}$ should rely not only on the distance between observed data $x^{\rm obs}(t)$ and predicted orbit, as in a purely deterministic evolution, but also on statistical deviations from this predicted orbit. It is thus, natural to consider the probability density function (PDF) $P(x,t)$ of finding the system at position x when measured at time t. This PDF is governed by the Fokker-Planck equation [13,14]

$$\frac{\partial P(x,t)}{\partial t} = -\sum_{i=1}^{N}\frac{\partial}{\partial x_i}[F_i(x,t)P(x,t)] + \sum_{i=1}^{N}\tau_i^2\frac{\partial^2 P(x,t)}{\partial x_i^2}, \tag{4}$$

which contains both a deterministic aspect [first right hand side (rhs) term] and a stochastic one (second rhs term). As a first attempt, one would like to minimize cost function

$$\mathcal{J}(\overline{a}) \equiv -\frac{1}{M}\sum_{k=1}^{M}\ln P^k(x^{\rm obs}(t_{k+1}),t_{k+1}) + \frac{1}{N}\sum_{i=1}^{N}\tau_i^2 \tag{5}$$

in which $P^k(x,t)$ denotes the PDF satisfying Eq. (4) within time interval $t_k < t < t_{k+1}$ and initial condition

$$P^k(x,t_k) = \delta(x - x^{\rm obs}(t_k)) \tag{6}$$

when no measurement noise is present, or else

$$P^k(x,t_k) = \frac{1}{\sqrt{2\pi D}}\exp\left(\frac{-[x-x^{\rm obs}(t_k)]^2}{2D}\right) \tag{7}$$

with measurement noise of variance D. This cost function is clearly related to the logarithm of the maximum likehood [12] with the proviso that an extra term is added to tentatively minimize noise amplitudes. Unfortunately, the optimization of Eq. (5) is practically impossible except in the simplest cases in which the probability density is a known analytical function of parameters $\overline{a}$. Numerically, PDF $P^k(x,t_k)$ may be evaluated by generating many realizations of the same initial condition $x^{\rm obs}$ [20]. Such a brute force method can hardly be put into practice, since it requires a considerable number of simulations of Eq. (1), an intensive computational effort even for a simple stochastic equation. However, when sampling time interval T is small compared to the characteristic correlation time of the stochastic system, it is possible to get the variances with respect to time. In such a case, an alternative cost function may be defined based on these ingredients.

III. LINEAR AND WEAKLY NONLINEAR APPROACH

A. Theoretical aspects

When sampling time T is small enough, variances associated with $P^k(x,t_k)$ remain of weak amplitude. During the interval $t_k \leqslant t < t_{k+1}$, the stochastic orbit, thus, stays close to the deterministic orbit x_i^{0k} defined by

$$\frac{dx_i^{0k}}{dt} = F_i(x^{0k};\mu_1,\ldots,\mu_P), \tag{8}$$

with $x_i^{0k}(t_k) = x_i^{\rm obs}(t_k)$. During the same interval, the stochastic variables $y \equiv x - x^{0k}$ are governed by the Langevin equations

$$\frac{dy_i}{dt} = G_i(y,x^{0k};\mu_1,\ldots,\mu_P) + \tau_i\theta_i(t), \tag{9}$$

where

$$G_i(y,x^{0k},\mu_1,\ldots,\mu_P) \equiv F_i(y+x^{0k};\mu_1,\ldots,\mu{\rm P}) - F_i(x^{0k};\mu_1,\ldots,\mu_P). \tag{10}$$

At time t_k, the stochastic variables $y \equiv x - x^{0k}$ verify, in the absence of the measurement noise, the initial condition $y(t_k) = 0$, or else satisfy a Gaussian probability distribution of variance D when noise is present. We assume that quantities y_i remain small during period $t_k < t < t_{k+1}$. At zeroth order, Eq. (9) may then be linearized as

$$\frac{dy_i}{dt} = \gamma_{ij}(t)y_j + \tau_i\theta_i(t), \tag{11}$$

where

$$\gamma_{ij}(t) = \frac{\partial F_i}{\partial x_j}(x_1^{0k}(t),\ldots,x_N^{0k}(t);\mu_1,\ldots,\mu_P). \tag{12}$$

Probability density $P_l^k(y,t)$ associated to the linearized Langevin Eq. (11) approximates the nonlinear one $P_{\rm nl}^k(y,t) \equiv P^k(x^{0k}+y,t)$ associated to Eq. (9). However, it is a much easier quantity to compute since it satisfies a Ornstein-Uhlenbeck equation [13,14]

$$\frac{\partial P_l^k(y,t)}{\partial t} = -\sum_{i,j}\gamma_{ij}(t)\frac{\partial}{\partial y_i}[y_j P_l^k(y,t)] + \sum_{i=1}^{N}\tau_i^2\frac{\partial^2 P_l^k(y,t)}{\partial y_i^2} \tag{13}$$

with a condition at time $t_k=kT$ similar to a δ function or a Gaussian function with zero average. Such an equation may be analytically integrated in $t_k<t<t_{k+1}$, the probability density function remains Gaussian [14] with zero average and variances $\sigma_{ij}(t)$ governed by

$$\frac{d\sigma_{ij}^k}{dt}=H_{ij}^k(\sigma_{pq}^k), \tag{14}$$

with initial conditions $\sigma_{ij}^k(t_k)=D\delta_{ij}$, where

$$H_{ij}^k(\sigma_{pq})=\sum_n \gamma_{\mathrm{in}}(t)\sigma_{nj}+\sum_n \gamma_{jn}(t)\sigma_{ni}+2\delta_{ij}\tau_i^2. \tag{15}$$

Variances that satisfy Eqs. (14) and (15) are expected to generically increase in an approximate exponential manner. When T becomes large, such an evolution clearly departs from the dynamics of the fully nonlinear problem. However, a better approximation may be found that reintroduces nonlinear contributions in a weakly nonlinear phase and thus remains valid for a longer time period. First, the governing equations for ensemble average $\langle y_i\rangle$ and variances $\sigma_{ij}\equiv\langle y_iy_j\rangle$ [21] are computed in the fully general problem (9). The computation of time derivatives of $\langle y_i\rangle$ and σ_{ij} leads to an integral in which appears the partial time derivative of the PDF $P_{\mathrm{nl}}^k(y,t)$. Replacing this time derivative by space derivatives using the Fokker-Planck equation, it is readily found, after an integration by parts, that

$$\frac{d\langle y_i\rangle}{dt}=\sum_p \gamma_{ip}(t)\langle y_p\rangle+\langle N_i\rangle, \tag{16}$$

$$\frac{d\sigma_{ij}}{dt}=\sum_p \gamma_{ip}(t)\sigma_{pj}+\sum_p \gamma_{jp}(t)\sigma_{pi}+2\delta_{ij}\tau_i^2+\langle N_iy_j\rangle + \langle N_jy_i\rangle, \tag{17}$$

where functions

$$N_i\equiv G_i(y,x^{0k};\mu_1,\ldots,\mu_P)-\sum_m \gamma_{im}(t)y_m \tag{18}$$

denote the purely nonlinear terms contained in functions $G_i(y)$. To get the above relation, the probability density function and its various gradients are assumed to vanish in phase space at infinity. Similar manipulations can be performed for higher-order moments, e.g., $\langle y_iy_jy_m\rangle$; an infinite hierarchy is hence defined for the time evolution of overall moments. This situation is akin the one found in isotropic turbulence. Similarly, a hypothesis should be used to close the hierarchy. An *ad hoc* Gaussian assumption is generally performed for higher moments in turbulence modeling. In this work, we follow this path though this hypothesis is better justified here. More precisely, for T small enough, we neglect the nonlinear terms N_i and the dynamics is reduced to Eq. (14) since the linear equation for $\langle y\rangle$ is trivially satisfied in that case, the average value $\langle y\rangle$ being initially zero. For larger T, one may relax this hypothesis by introducing the nonlinear contributions N_i in the dynamics, i.e., by using the exact Eqs. (16) and (17). However, the exact PDF is now no more Gaussian, but reads

$$P_{\mathrm{nl}}(y,t)=P_G(y,\langle y\rangle,\sigma_{ij})+\epsilon(t)P_{\mathrm{NG}}(y,t) \tag{19}$$

where P_G denotes the Gaussian PDF with ensemble average and variances of P_{nl}, $P_{\mathrm{NG}}(y,t)$ the normalized non-Gaussian remaining part and $\epsilon(t)$ an amplitude that quantify how far the PDF if from being Gaussian. In the linear régime $\epsilon(t)$ is precisely zero. More generally an approximation for small time may be done to solve the nonlinear Fokker-Planck (see, for instance, Ref. [14]) in which the Gaussian part is shown to be still the leading order term. It is thus reasonable to assume that, for a larger time period, the Gaussian part is predominant in its contributions to terms $\langle N_i\rangle$ or $\langle N_iy_j\rangle$. Note that, in Eq. (19), variables $\langle y\rangle$, σ_{ij} obviously do not follow a linear evolution, in particular, $\langle y\rangle$ may be different from zero. In a way, this reasoning is quite similar to the one employed to derive amplitude equations in the theory of deterministic nonlinear extended systems; one assumes that the main term has the same space (here the phase space) dependence as in a general linear evolution but unsteady amplitudes [here $\langle y\rangle(t)$, $\sigma_{ij}(t)$] are governed by nonlinear evolution equations. These amplitude equations are easily derived here by computing $\langle N_i\rangle$, $\langle N_iy_j\rangle$ and $\langle N_jy_i\rangle$ with the approximate $P_G(y,\langle y\rangle,\sigma_{ij})$ rather than the true $P_{\mathrm{nl}}(y,t)$. Such contributions thus become only functions of $\langle y\rangle$, σ_{ij}. This closes the hierarchy since everything is now defined in terms of $\langle y\rangle$, σ_{ij}. When functions G_i are polynomials or power series in y, one gets

$$N_i\equiv\sum_{r,s}\Gamma_{irs}(t)y_ry_s+(\text{higher-order terms}), \tag{20}$$

where coefficients depend on the deterministic orbit $x^{0k}(t)$ [e.g., $\Gamma_{lrs}(t)$ in Eq. (20) stands here the Hessian of G_l]. Quantities $\langle N_i\rangle$, $\langle N_iy_j\rangle$, and $\langle N_jy_i\rangle$ may then be expressed in terms of various moments which, in the Gaussian approximation, can ultimately be written in terms of variances and ensemble averages. For instance, in the Lorenz system (see Sec. III B), one has the exact relations

$$\langle N_i\rangle=\sum_{r,s}\Gamma_{irs}(t)\sigma_{rs}, \tag{21}$$

$$\langle N_iy_j\rangle+\langle N_jy_i\rangle=\sum_{r,s}\Gamma_{irs}(t)\langle y_ry_sy_j\rangle+\sum_{r,s}\Gamma_{jrs}(t)\langle y_ry_sy_i\rangle. \tag{22}$$

In the weakly nonlinear phase, the contribution of the non-Gaussian part of the PDF of the three-point correlations is neglected, which reads

$$\langle(y_r-\langle y_r\rangle)(y_s-\langle y_s\rangle)(y_p-\langle y_p\rangle)\rangle=0, \tag{23}$$

or equivalently

$$\langle y_ry_sy_p\rangle=\langle y_r\rangle\sigma_{sp}+\langle y_s\rangle\sigma_{rp}+\langle y_p\rangle\sigma_{rs}-2\langle y_r\rangle\langle y_s\rangle\langle y_p\rangle. \tag{24}$$

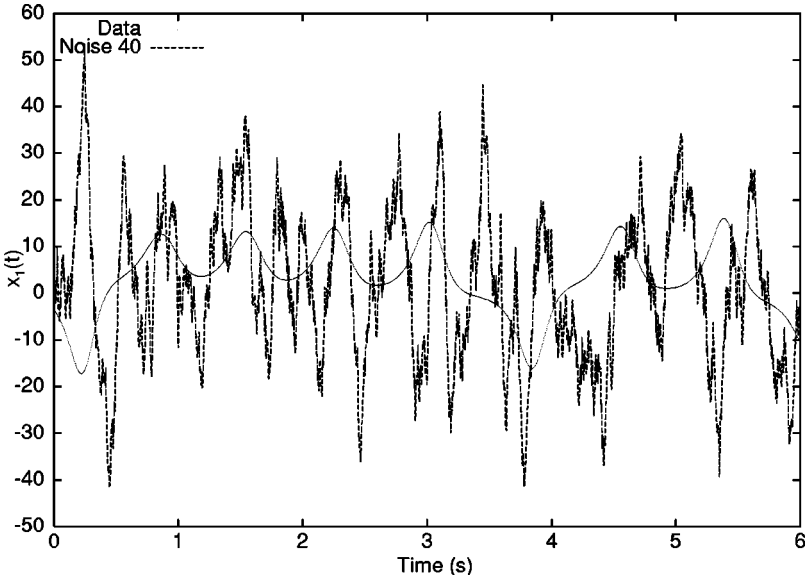

FIG. 1. Temporal evolution of variable $x_1(t)$ of the Lorenz system for $\sigma=10$, $r=28$, $b=2.666$, and $\tau=0$ (dots) or $\tau^2=40$ (solid line).

As for amplitude equations in bifurcation problems, the validity of this weakly nonlinear approximation depends on the nonlinear system and on the orbit. For supercritical bifurcations, it is quite appropriate for all times but for subcritical bifurcations, it is only locally valid. As in bifurcations problems, we do not attempt to define precise theoretical bounds for T. They are clearly dependent on the noise amplitudes as well as on the finite time Lyapunov exponents of the deterministic system related to the Langevin equation we study. However, a quantitative relation between the maximal sampling time T and the amplitude of the noise τ_i seems difficult, not to say impossible, to obtain. Practically, given a parametric model, one should play with it to determine how far one can push this assumption. In the following section, we did this search for the Lorenz system and we showed that nonlinear effects such as saturation can be obtained in this higher-order approximation.

B. Results of the closure assumption

The above linear and weakly nonlinear approximations have been tested on the particular data set produced by a stochastic Lorenz system [17]

$$\frac{dx_1}{dt}=\sigma(x_2-x_1)+\tau\theta_1(t),$$

$$\frac{dx_2}{dt}=rx_1+x_2-x_1x_3+\tau\theta_2(t),$$

$$\frac{dx_3}{dt}=-bx_3+x_2x_1+\tau\theta_3(t),$$

in which parameters σ, r, and b are constant coefficients. We used this system because it is a paradigmatic example of a nonlinear system. In the reconstruction community, or more generally in the nonlinear system community, much work and ideas have been worked out using this system earlier on. In this numerical check, coefficients are taken to be equal to $\sigma=10$, $r=28$, $b=2.666$ (case A) or $\sigma=19.03$, $r=7.63$, $b=3.87$ (case B) which, respectively, correspond to a chaotic and a fixed point régime for the Lorenz deterministic system. For most examples, the noise amplitude is taken to be $\tau^2=40$ (see Fig. 1). Moreover, in order to focus on the dynamical noise problem, data $\mathcal{S}$ is assumed free of observational noise ($D=0$). Note that, we tested our original algorithm on the Lorenz system because it is the paradigmatic example of nonlinear dynamics.

The test of the closure assumption is performed as follows. First, the deterministic orbit $x^{0k}(t)$ [resp. variances $\sigma_{ij}(t)$] is obtained by integration of Eq. (8) [resp. linear Eq. (14) or weakly nonlinear Eqs. (21)–(24)] starting from a given initial condition $x^{\mathrm{obs}}(0)$. In the second stage, $\mathcal{R}\gg 1$ realizations of the stochastic system (1) are numerically simulated with the one time-step numerical algorithm explained in [18] with a time step $\Delta t=0.0001$. Using these $\mathcal{R}$ different values of $x^{\mathrm{obs}}(t)$, an histogram (i.e., a PDF) is obtained that approximates $P(x,t)$. It is checked that (a) the integration step Δt is small enough to not affect the PDF results (see Fig. 2), and (b) the histogram is well approximated with $\mathcal{R}=1000$ realizations (see Fig. 3). Once these numerical checks are done, such a histogram, computed with $\mathcal{R}=5000$ realizations is shown on Fig. 4, for two initial points $x^{\mathrm{obs}}(0)$ on the ''attractor'' and two different sampling times $T=0.1$ (solid line) and $T=0.5$ (dashed line). The time necessary for the loss of Gaussianity to occur, depends on the phase space region; it is not surprising and may be clearly related to finite time Lyapunov exponents. Variances $\sigma_{ij}(t)$ of $y(t)=x^{\mathrm{obs}}(t)-x^{0k}(t)$ can also be derived from the R different realizations of $x^{\mathrm{obs}}(t)$ and compared to the predicted values of the closure assumption. On Figs. 5–8, the variation of $\sigma_{11}(t)$ with respect to time is shown: solid lines represent the linear approximation results, dots correspond to the nonlinear approximation, averages over $\mathcal{R}$ realizations are indicated using $\mathcal{R}=1000$ (squares) and $\mathcal{R}=5000$ (circles). As seen in Figs. 5 and 6, the linear approximation behaves reasonably well for case A and $\tau^2=20$ (resp. $\tau^2=40$) until T

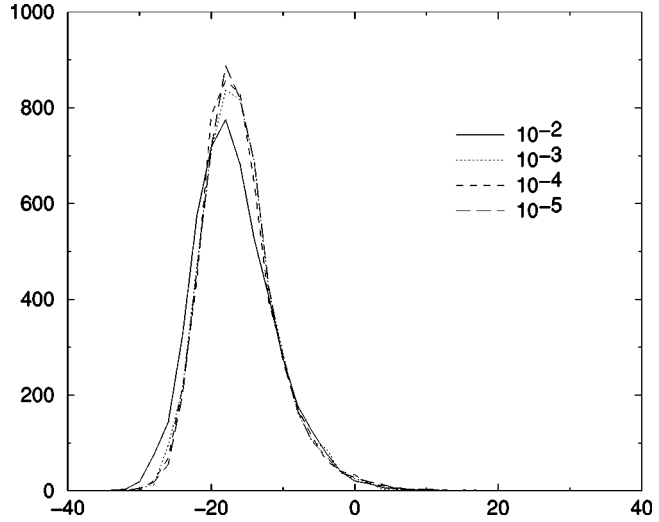

FIG. 2. Influence of the discretization time step $\Delta t(10^{-2},10^{-3},10^{-4},10^{-5})$, on the PDF of x_1 computed at time $T=1$. Computations were performed for $\mathcal{R}=5000$ and for case A, $\tau^2=10$ and initial condition $x_1=-3.6$, $x_2=-6.9$, $x_3=7.9$.

$\leqslant 0.1$ (resp. $T\leqslant 0.08$) while the nonlinear approximation may be used for larger time intervals, i.e., $T\leqslant 0.25$ (resp. $T\leqslant 0.12$). Even for larger times, the weakly nonlinear approach is acceptable. As previously mentioned, the quality of the approximation depends on the position of the initial conditions (see Fig. 7) as well as on the parameters (see Fig. 8). As a conclusion, it is possible to use the linear approximation for σ_{ij} if T is small enough or the weakly nonlinear one for larger values. Clearly, this latter possibility is also limited to a range of time intervals T.

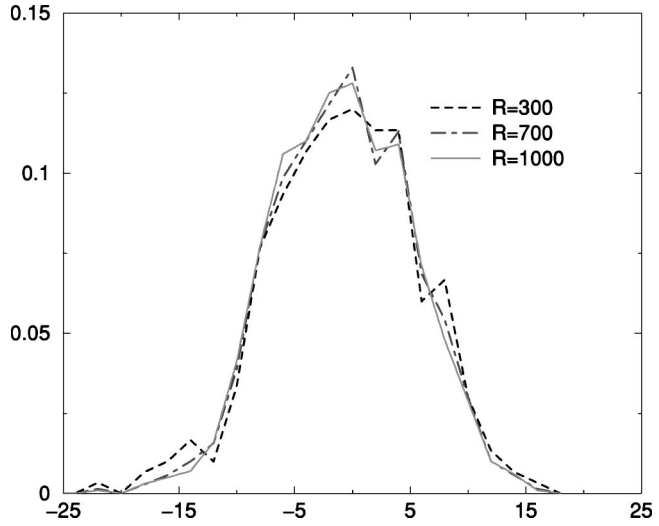

FIG. 3. Rescaled histogram of variable x_1 computed for $\sigma=10$, $r=28$, $b=2.666$, and $\tau^2=10$ and the initial condition $x_1=-3.2$, $x_2=1.5$, $x_3=28.2$; for $\mathcal{R}=300$ (dashed line) $\mathcal{R}=700$ (dot-dashed line), $\mathcal{R}=1000$ (solid line) realizations.

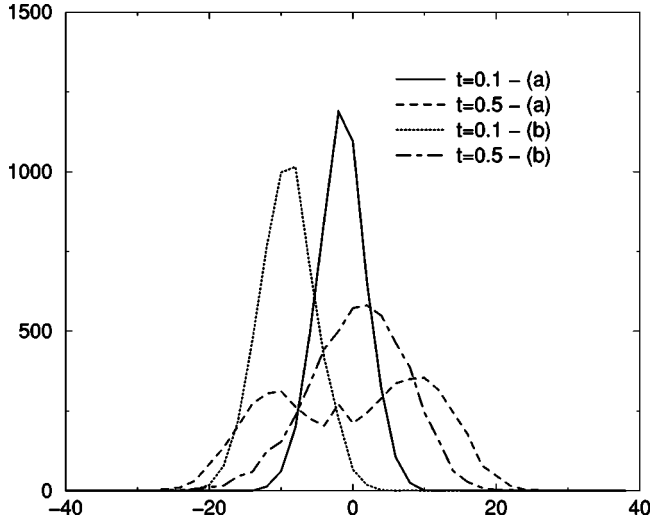

FIG. 4. Histogram of variable x_1 computed for $\sigma=10$, $r=28$, $b=2.666$, and $\tau^2=10$ using $\mathcal{R}=5000$ data points (a) For the initial condition $x_1=-3.2$, $x_2=1.5$, $x_3=28.2$ and times $T=0.1$ and $T=0.5$ (b) For the initial condition $x_1=-3.6$, $x_2=-6.9$, $x_3=7.9$ and times $T=0.1$ and $T=0.5$.

IV. THE IDENTIFICATION METHOD

A. The cost function

In a purely deterministic case, the cost function $\mathcal{J}$ is often based on the distance between data $x^{\mathrm{obs}}(t_{k+1})$ and predicted deterministic orbit $x^{0k}(t_{k+1})$, i.e.,

$$\mathcal{J}(\overline{a})\equiv\frac{1}{M}\sum_{k=1}^{M}[x^{\mathrm{obs}}(t_{k+1})-x^{0k}(t_{k+1})]^2. \tag{25}$$

Contrary to this classical choice, a cost function $\mathcal{J}$ is introduced here for the Langevin Eqs. (1), based on the numerical probability density of the stochastic variable $y_i(t_{k+1})=x(t_{k+1})-x^{0k}(t_{k+1})$. This function contains a deterministic part through the noise-free solution $x^{0k}(t)$ as well as a stochastic part through variables $y(t)$. If T is small enough, the linear approximation holds and the Gaussian PDF P_l^k of the Ornstein-Uhlenbeck process may be used

$$\mathcal{J}(\overline{a})\equiv-\frac{1}{M}\sum_{k=1}^{M}\ln P_l^k(x^{\mathrm{obs}}(t_{k+1})-x^{0k}(t_{k+1}),t_{k+1})+\frac{1}{N}\sum_{i=1}^{N}\tau_i^2. \tag{26}$$

A version that is easier to implement and that can be extended to the weakly nonlinear regime is considered here. Let us first divide the phase space region that contains the observed data $\mathcal{S}$ into U domains $\mathcal{V}_u$ $(u=1,...;U)$ and denote by $\sigma_{ij}(x,T)$ the variance computed starting at x and evolving during a time interval T. If the number N_u of data points in $\mathcal{V}_u$ is large enough, the averaged variance over the domain $\mathcal{V}_u$

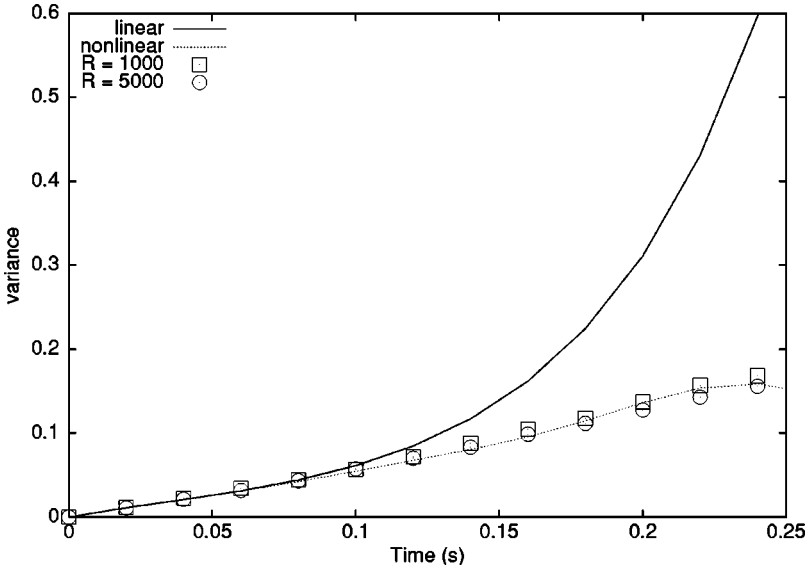

FIG. 5. Case A and $\tau^2=20$. Quantity $\sigma_{11}(t)$ computed, starting from the initial condition on the attactor $x_1=-3.6$, $x_2=-6.9$, $x_3=7.9$ averaging over $\mathcal{R}=1000$ realizations (squares) and $\mathcal{R}=5000$ (circles), using the linear approximation (solid), and the weakly nonlinear approximation (dots).

$$\frac{\int_{\mathcal{V}_u}\sigma_{ij}(x,T)P(x,t)dx}{\int_{\mathcal{V}_u}P(x,t)dx} \tag{27}$$

may be replaced by the following discrete term:

$$\frac{1}{N_u}\sum_{\mathcal{V}_u}\sigma_{ij}(t_{k+1}), \tag{28}$$

in which the symbol $\mathcal{V}_u$ means that sums are performed over all points $x_j^{0k}(t_k)$ located in $\mathcal{V}_u$. Moreover, when N_u is large enough, the following equality applies:

$$\frac{1}{N_u}\sum_{\mathcal{V}_u}\sigma_{ij}(t_{k+1})\sim\frac{1}{N_u}\sum_{\mathcal{V}_u}[x_i^{\mathrm{obs}}(t_{k+1})-x_i^{0k}(t_{k+1})]\times[x_j^{\mathrm{obs}}(t_{k+1})-x_j^{0k}(t_{k+1})]. \tag{29}$$

Using these constraints, the cost function $\mathcal{J}$

$$\mathcal{J}=\sum_{u=1}^{U}\sum_{i,j}\alpha_{ij}^u\mathcal{J}_{ij}^u \tag{30}$$

is introduced where the α_{ij}^u are weight coefficients and

$$\mathcal{J}_{ij}^u=\left(\frac{1}{N_u}\sum_{\mathcal{V}_u}\sigma_{ij}(t_{k+1})-\frac{1}{N_u}\sum_{\mathcal{V}_u}[x_i^{\mathrm{obs}}(t_{k+1})-x_i^{0k}(t_{k+1})]\times[x_j^{\mathrm{obs}}(t_{k+1})-x_j^{0k}(t_{k+1})]\right)^2. \tag{31}$$

When dynamical noise amplitudes τ_i are quite small, variances σ_{ij} become negligible and cost function (30) is similar to a least-squares function (25) used for deterministic equations. When noise amplitudes become large, statistical aspects predominate and a cost function, such as Eq. (25), is irrelevant; during time interval T, variances may significantly

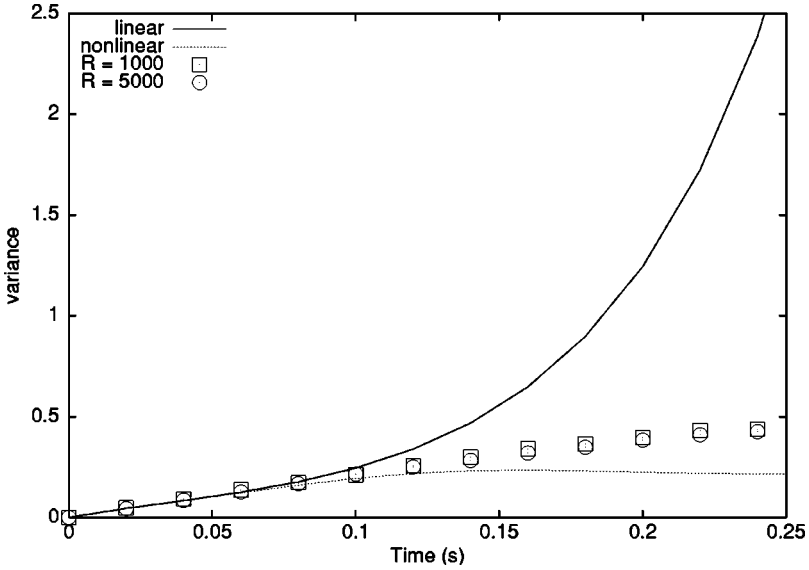

FIG. 6. Case A and $\tau^2=40$. Quantity $\sigma_{11}(t)$ computed, starting from the initial condition on the attactor $x_1=-3.6$, $x_2=-6.9$, $x_3=7.9$ averaging over $\mathcal{R}=1000$ realizations (squares) and $\mathcal{R}=5000$ (circles), using the linear approximation (solid), and the weakly nonlinear approximation (dots).

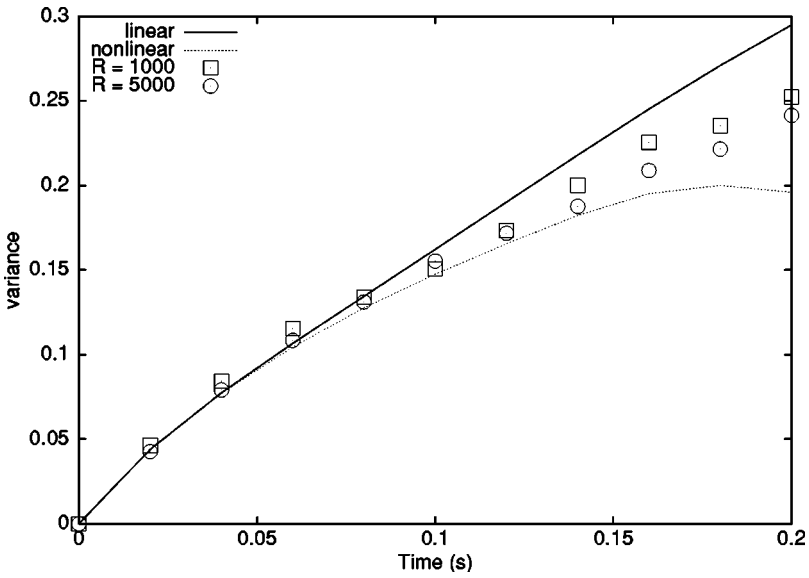

FIG. 7. Case A and $\tau^2=40$. Quantity $\sigma_{11}(t)$ computed, starting from the initial condition $x_1=-8.1$, $x_2=-10.7$, $x_3=22.9$, averaging over $\mathcal{R}=1000$ realizations (squares) and $\mathcal{R}=5000$ (circles), using the linear approximation (solid), and the weakly nonlinear approximation (dots).

change compared to the time variation of the deterministic orbit x^{0k}. However, cost function (30) takes this aspect into account. Note that, if the number N_u is too small—convergence must be checked case by case by modifying the total number of measurement points, i.e., by using part of the data set $\mathcal{S}$—the constraint on $\mathcal{V}_u$ is not considered and $\alpha_{ij}^u=0$ holds (on the reverse case, $\alpha_{ij}^u=1$).

B. Optimization algorithm: The simulated annealing

The method of simulated annealing is used to find the global minimum of a given functional just as the cost function $\mathcal{J}$, which implicitly depends on parameter $\overline{a}$ through a set of evolution equations acting as dynamical constraints. Here this technique is applied to minimize cost function (30) with the constraint that the deterministic orbit $x^{0k}(t)$ satisfies Eq. (8) and the weakly nonlinear variances satisfy Eqs. (21)–(24). This procedure is an extension of the classical Monte Carlo method [9,10] since it is based on a probabilistic search of the global minimum. After picking up a random value $\overline{a}_1$ and computing the associated cost function $\mathcal{J}_1$, the method consists in iterating three basic steps. Namely, at each iteration i, one picks up the specific value $\overline{a}_{i-1}$ of parameter $\overline{a}$, which was computed at the previous iteration $i-1$. A new guess value $\overline{a}$ is obtained stochastically using the Gaussian random distribution $P(\overline{a},i)=(2\pi\phi_i)^{-1/2}\{[-(\overline{a}-\overline{a}_{i-1})^2]/2\phi_i\}$ where ϕ_i is the so-called temperature parameter at iteration i. The simulation of Eqs. (8), (21)–(24) is thereafter performed with this new parameter $\overline{a}$ and the value of the cost function denoted by $\mathcal{J}_i$ is obtained as a byproduct. If $\mathcal{J}_i \leqslant \mathcal{J}_{i-1}$ then value $\overline{a}_i$ is set to $\overline{a}$ at iteration i. In the reverse case, $\overline{a}_i$ keeps the value at the previous iteration, i.e., $\overline{a}_i=\overline{a}_{i-1}$ with a probability $1-h_i$, or else $\overline{a}_i$ becomes equal to $\overline{a}$ with a probability h_i. The probability of acceptance h_i is a major ingredient in the simulated annealing procedure and is defined as $h_i=\exp[-(\mathcal{J}_i-\mathcal{J}_{i-1})/\phi_i]$. Clearly, the crucial step of the simulated annealing procedure lies in the choice of h_i or, more precisely, of the temperature ϕ_i at each iteration i. The simplest evolution has been considered here;

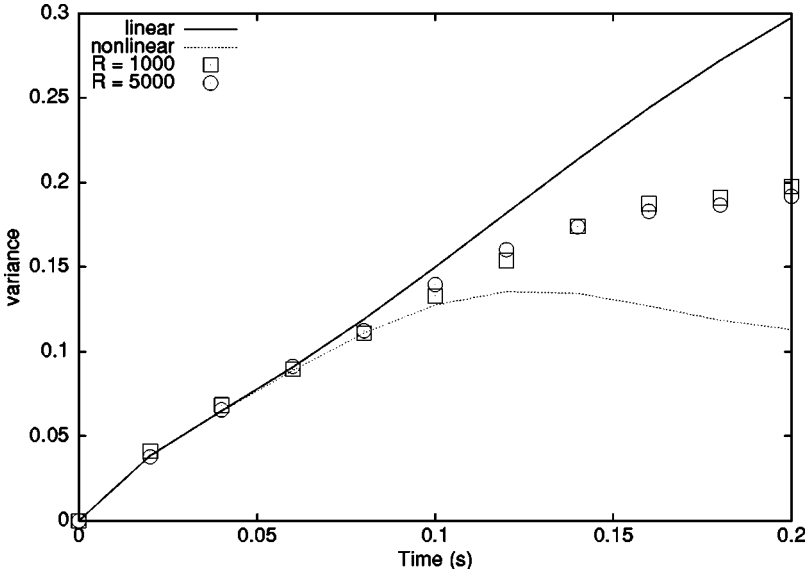

FIG. 8. Case B and $\tau^2=40$. Quantity $\sigma_{11}(t)$ computed, starting from initial conditions $x_1=-3.6$, $x_2=-6.9$, $x_3=7.9$, averaging over $\mathcal{R}=1000$ realizations (squares) and $\mathcal{R}=5000$ (circles), using the linear approximation (solid), and the weakly nonlinear approximation (dots).

a linear decrease with iteration index i. The decrease should be slow enough; ϕ must not evolve faster than $\phi_i = \phi_0/\ln(i)$ with ϕ_0. For details of the general method the reader may see Ref. [9]. Clearly we want here to demonstrate the accuracy of the overall method. As a consequence, we do not try to accelerate the simulated annealing convergence by using more sophisticated temperature evolutions. In this paper, this method is implemented only for the weakly nonlinear Eqs. (21) and (22). It could be *a priori* employed for the linear case as well. However a technique requiring less numerical efforts, is presented in the following section for the linear case; the backpropagation procedure.

C. Optimization algorithm: The backpropagation procedure

For deterministic equations, another iterative method known as preconditioned limit memory algorithm (PLMA) algorithm [7] was used to search for the minimum of a functional as in Eq. (25) with constraints as in Eq. (8). This approach, which searches for local minima, is less general than the simulated annealing. However it converges much faster. In this section, a similar method is introduced for the stochastic problem. As for the pure deterministic case, it is again a much faster algorithm than the simulated annealing.

The PLMA algorithm is a generalized Newton method that looks for minima of $\mathcal{J}$ by computing the gradient and Hessian of $\mathcal{J}$ with respect to parameter $\overline{a}$. An efficient computation of this latter quantity is, thus, crucial to apply the PLMA algorithm especially for identification problems that are generally ill conditioned. For a deterministic system as in Eq. (8), this quantity can be obtained by a backpropagation method. This procedure, known to be quite robust to observational noise, consists in defining a companion deterministic problem (the adjoint problem) to system (8). A unique simulation of the N so-called backpropagated equations is sufficient to compute the gradient over all parameters. A similar procedure proposed here for stochastic equations imposes several assumptions. First time interval T is assumed small enough so that the linear Eq. (14) is appropriate for the evolution of variances and the explicit relation

$$x_j^{0k}(t) = x_j^{\mathrm{obs}}(t_k) + F_i(x^{\mathrm{obs}}(t_k);\mu_1,\ldots,\mu_P)(t-t_k), \quad (32)$$

instead of Eq. (8), holds for the deterministic part x_j^{0k} during $t_k \leqslant t < t_{k+1}$. This is understood as follows. For the case of large dynamical noise, variances are changing much more than the deterministic orbit x^{0k} and hence, Eq. (32) is a good approximation for x^{0k}. In the reverse case of small dynamical noise, our procedure is strictly equivalent to the one time-step method. The case of an almost deterministic system where Eq. (32) is not valid, can be treated by a classical backpropagation method [5]. As a second approximation, the first sum in Eq. (31) is approximated by $\sigma_{ij}(x_u^c,T)$, where x_u^c stands for a point in the neighborhood $\mathcal{V}_u$. This means that the variance is not changing much with position in the neighborhood $\mathcal{V}_u$. Function $\mathcal{J}_{ij}^u$ in Eq. (31) is now reduced to

$$\mathcal{J}_{ij}^u(\overline{a}) = [\sigma_{ij}(x_u^c,T) - G_{ij}^u(T)]^2, \quad (33)$$

where function

$$G_{ij}^u(T) \equiv \frac{1}{N_u}\sum_{\mathcal{V}_u} [x_i^{\mathrm{obs}}(t_{k+1}) - x_i^{0k}(t_{k+1})][x_j^{\mathrm{obs}}(t_{k+1}) - x_j^{0k}(t_{k+1})] \quad (34)$$

is explicit in terms of parameter $\overline{a}$ because of Eq. (32). On the contrary, quantity $\sigma_{ij}(x_u^c,T)$ is an implicit function of parameters via Eq. (14). All previous approximations may be removed. In particular, the weakly nonlinear problem may be introduced within the backpropagation method but the simplicity and efficiency is then lost since the problem would then contain nonlocal dependences not only in variables $x_i^{0k}(t)$ as for the pure deterministic case, but also dependences on variances $\sigma_{ij}(t)$. In that instance, the simulated annealing performs much better. In the simplified setting described above, the number of direct and backpropagated equations are $(N+1)N \times U$. The obvious drawback is clearly the reduced range of time interval T where it can be applied. In this framework, we briefly describe the method when it is extended to the case of dynamical noise.

The gradient of $\mathcal{J}(\overline{a})$ with respect to the various parameters in $\overline{a}$ is actually the sum of the contribution of each neighborhood $\mathcal{V}_u$ separately. As a consequence, below we consider only one neighborhood. The cost function thus reads

$$\mathcal{J}_u = \sum_{i,j} \alpha_{ij}^u [\sigma_{ij}(x_u^c,T) - G_{ij}^u(T)]^2. \quad (35)$$

The dependence with respect to parameter $\overline{a}$ is partly explicit through $G_{ij}^u(T)$, partly implicit since $\sigma_{ij}(x_u^c,T)$ depends on these parameters through the direct problem (14) and (32). The backpropagation procedure handles this difficulty by introducing a generalized Lagrangian

$$\mathcal{L} = \mathcal{J} + \sum_{i,j} \int_{t=0}^{T} dt \left[\frac{d\sigma_{ij}}{dt} - H_{ij}(\sigma_{pq}) \right] Y_{ij}(t), \quad (36)$$

where the multipliers $Y_{ij}(t)$ are introduced for $0 \leqslant t < T$ and $i, j = 1,\ldots,N$. When $x(t)$ and $\sigma_{ij}(t)$ satisfy the direct problem (14) and (32), one gets $\mathcal{J} = \mathcal{L}$ for any choice of multipliers $Y_{ij}(t)$, for $0 \leqslant t < T$. Moreover, $\delta\mathcal{J}$ the first-order variation of the cost function with respect to the variation of parameters a_l, then reads

$$\delta\mathcal{J} = \sum_l \frac{\partial\mathcal{L}}{\partial a_l}\delta a_l + \int_{t=0}^{T} dt \sum_{i,j} \frac{\partial\mathcal{L}}{\partial\sigma_{ij}(t)}\delta\sigma_{ij}(t), \quad (37)$$

where δa_l represents the variation of parameter a_l, and $\delta\sigma_{ij}(t)$ the first-order variations of variables $\sigma_{ij}(t)$ with δa_l. The first term in Eq. (37) denotes the explicit dependence on a_l while the second concerns the implicit dependence that is difficult to compute. Note that a term, such as $\partial\mathcal{L}/\partial Y_{ij}(t)$, is identically zero when $\sigma_{ij}(t)$ verifies the direct problem. At this stage the multipliers $Y_{ij}(t)$ are still arbitrary. The backpropagation procedure consists in choosing these free variables in such a way that

$$\sum_{i,j} \frac{\partial \mathcal{L}}{\partial \sigma_{ij}(t)} \delta\sigma_{ij}(t) = 0 \tag{38}$$

holds for any allowed variations $\delta\sigma_{ij}(t)$ [22]. This constraint defines the backpropagated problem, which is a dynamical system for multipliers $Y_{ij}(t)$ solvable backward in time. For details, the reader may see Refs. [5–8]. Integrating this new dynamical system thus provides the Lagrangian multipliers $Y_{ij}(t)$. When such choice is made for $Y_{ij}(t)$, the variation $\delta\mathcal{J}$ is then easily computed. Apart from the explicit dependence, the implicit part becomes

$$\delta\mathcal{J} = \sum_l \frac{\partial \mathcal{L}}{\partial a_l} \delta a_l \tag{39}$$

with

$$\frac{\partial \mathcal{L}}{\partial a_l} = -\int_{t=0}^{T} dt \sum_{i,j} \frac{\partial H_{i,j}}{\partial a_l} Y_{ij}(t). \tag{40}$$

The method is thus identical to the one used for the deterministic case, but it is extended to the stochastic problem by using the deterministic linear Eq. (14) for variances and avoiding direct use of governing equations. This trick then allows to use the usual procedure.

V. A TEST CASE: STOCHASTIC LORENZ SYSTEM

The two algorithms have been tested on a stochastic Lorenz system. Note that the equivalent deterministic system may be chaotic or may correspond to a fixed point regime.

A. Results of the simulated annealing

We compute the optimal parameters using the simulated annealing procedure within the weakly nonlinear approximation. The input data set $\mathcal{S}$ is provided by the integration of case A with a noise amplitude $\tau^2=40$ and initial conditions $(-3.6, -6.9, 7.9)$, which is sampled at $T=0.1$. This sampling time corresponds to the interval for which the linear approximation fails while the nonlinear one is still valid (Fig. 6). When using such a data set of 100 points (respectively, 1000), the algorithm provides the following best estimate (10.7, 27.1, 2.06) with $\tau^2=37.54$, and respectively (9.8, 28.2, 2.54) with $\tau^2=40.51$. For a larger sampling time $T=0.2$ and 1000 measurement points, we still get a rather acceptable result: (8.0, 25.0, 2.84) with $\tau^2=43.9$. We checked that these values are slightly dependent on (a) the initial conditions for the data set [23] and (b) the initial guess values for the parameters. These results are satisfactory and robust against noise, which is pretty strong in this case. Furthermore, this is to be put in parallel with the results obtained from the identical data set by a simulated annealing procedure that uses the deterministic cost function (25) and the deterministic equations. In that case the optimal set is found to be (26.75, 39.8, 7.9) with 1000 points, which is far from the correct values. The method works also for small noise levels: For instance, we used a data set produced by case A with $\tau^2=1$ and obtained with 4000 points and sampling time $T=1$ (9.88, 27.72, 2.74) and $\tau^2=0.99$. The cells $\mathcal{V}_u$ are obtained by dividing the phase space $-20<x<20$, $-30<y<30$ in five intervals and $0<y<40$ in three intervals.

TABLE I. Deterministic/Stochastic method: identified values for $M=1000$ and 4000 measurement points. Exact values ($\sigma=10, R=28, b=2.66, \tau=1$).

	Deterministic method $T=0.1$ ($M=1000/M=4000$)	Deterministic method $T=0.4$ ($M=4000$)
σ	7.384/6.986	0.7096
R	26.41/26.46	29.36
b	3.896/3.908	4.253
τ		
	Stochastic method	Stochastic method
σ	10.12/9.990	9.818
R	28.57/27.45	27.85
b	2.257/3.222	2.678
τ	0.979/1.029	1.091

B. Results of backpropagation procedure

Our extended backpropagation approach is first tested using a data set $\mathcal{S}$ produced by a stochastic Lorenz system corresponding to case A with a noise level $\tau^2=1$. Table I presents a comparison between the identification results obtained by (i) a purely deterministic backpropagation approach, which is a rather robust technique with respect to measurement noise, and (ii) our stochastic backpropagation method. When the data set $\mathcal{S}$ is sampled every time period $T=0.1$, the parameters values are satisfactorily evaluated with the stochastic method for both $M=1000$ and 4000 measurement points. Indeed, the identified values fluctuate around the correct ones. On the contrary, the deterministic method does not converge towards the exact quantities when increasing the number of measurement points.

When the data set $\mathcal{S}$ is sampled with $T=0.4$, the deterministic method is far away from the correct value even with $M=4000$ points. Paradoxically, our method behaves even better for $T=0.4$ than for $T=0.1$. This can be accounted for as follows: for $T=0.1$ and noise amplitude $\tau^2=1$, variances dynamics is dominated by the noise term in Eq. (14). The part of the identification method devoted to variances is hence essentially affected by noise amplitude. For a larger sampling time $T=0.4$, the cost function becomes more sensitive to the other terms on the rhs of Eq. (14). Consequently, the other parameters directly influence the computation of variances. Thus, this improves predictions.

For values of parameters σ, R, and b corresponding to a nonchaotic regime, the two methods provide quite different results for the time asymptotic dynamics. For a case in which ($\sigma=10, R=20, b=2.66$) and a noise level $\tau^2=5$, the dynamics may be described as a noisy ''attractor,'' which has nothing to do with a fixed point dynamics observed for vanishing dynamical noise. Using $M=1000$ measurement points and sampling time $T=0.1$, the classical deterministic method gives a fixed point dynamics ($\sigma=7.69, R=19.03, b=3.87, \tau$

$=0$). In the same instance, the stochastic approach provides a realistic behavior ($\sigma=10.20, R=19.93, b=2.55, \tau=4.95$) with a noisy Lorenz attractor.

For larger noise levels, e.g., $\tau=40$ and a chaotic regime ($\sigma=10$, $R=28$, and $b=2.66$), the deterministic method fails. For $M=2000$ measurement points and a smaller sampling time ($T=0.001$), this deterministic method identifies $\sigma=20.76$, $R=16.94$, and $b=8.887$, which simulates a fixed point asymptotic dynamics. On the contrary, the stochastic approach provides the value (9.715, 28.27, 2.587) and $\tau=40.92$. For large dynamical noise, the stochastic method is the only one capable to identify and then reproduce the correct observations. In these computations, the phase space $-20<x<20$, in which the dynamics is observed, has been divided in three intervals, $-30<y<30$ in five and finally $0<y<40$ in three. We imposed $\alpha_{ij}=0$ for $i\neq j$ and for neighborhoods $\mathcal{V}_u$ containing less than 50 points.

C. Conclusion

We have proposed an identification method for stochastic equations. This procedure was tested with a data set produced with large dynamical noise; correct values were obtained while deterministic based approaches systematically failed. Note that, most of this discussion needs only a slight modification to include a multiplicative noise term in Eq. (1). A more stringent constraint concerns the assumption that the N variables are his condition is appropriate as a first step and will be in future work.

[1] E. Ott, T. Sauer, and J. G. Yorke, *Coping with Chaos* (Wiley, New York, 1994).

[2] H. Kantz and Th. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).

[3] G. Gouesbet and C. Letellier, Phys. Rev. E **49**, 4955 (1994).

[4] E. J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).

[5] G. Chavent, in *IFAC Symposium on Identification and System Parameters Estimations*, edited by R. Isermann (Pergamon, Oxford, 1979), Vol. 1, pp. 85–97.

[6] J. M. Fullana, P. Le Gal, M. Rossi, and S. Zaleski, Physica D **102**, 102 (1997).

[7] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic, New York, 1981).

[8] J. M. Fullana, Thèse de l'Université de Paris VI, 1997.

[9] A. L. Ingber, Math. Comput. Modell. **18**, 29 (1993).

[10] A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[11] J. Von Hardenberg *et al.*, Ann. N.Y. Acad. Sci. **808**, 79 (1997).

[12] J. Timmer, *Parameter Estimation in Nonlinear Stochastic Differential Equations* (Wiley, New York, 1994).

[13] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1983).

[14] H. Risken, *The Fokker-Planck Equation* (Springer-Verlag, Berlin, 1989).

[15] F. Moss and P. V. McClintock, *Noise in Nonlinear Dynamical Systems* (Cambridge University Press, Cambridge, 1989), Vol. 3.

[16] L. Battiston and M. Rossi, Int. J. Chaos Theor. Appl. **5**, 3 (2000).

[17] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[18] R. Mannella, in *Noise in Nonlinear Dynamical Systems*, edited by F. Moss and P. V. E. McClintock (Cambridge University Press, Cambridge, 1989), Vol. 3.

[19] Once a clean deterministic series is obtained, it is subsequently introduced into standard algorithms that compute correlation dimensions or make predictions.

[20] This is strictly valid only for the simplest case of no measurement noise. When measurement noise is present, the initial condition must be chosen according to the Gaussian probability (7).

[21] By definition, the ensemble average of quantity $q(y_1,\ldots,y_N)$ reads as the integral over phase space $\langle q\rangle(t) \equiv \int_{-\infty}^{+\infty} q(y_1,\ldots,y_N) P_{nl}(y,t)\Pi_p dy_p$.

[22] Because of initial and boundary conditions such variations should be such that $\delta\sigma_{ij}(0)=0$.

[23] We used as starting conditions (-8.1, -10.75, 22.9) to generate the data set, the estimates obtained for 1000 measurement points are (10.49, 27.02, 2.23) and $\tau^2=39.9$ for sampling time $T=0.1$ and (8.12, 27.12, 2.85) and $\tau^2=42.5$.