Estimation of parameters and unobserved components for nonlinear systems from noisy time series

A. Sitz,[1] U. Schwarz,[1] J. Kurths,[1] and H. U. Voss[2]
[1]*Center for Dynamics of Complex Systems, University of Potsdam, 14469 Potsdam, Germany*
[2]*Freiburg Center for Data Analysis and Modeling, 79104 Freiburg, Germany*
(Received 8 November 2001; revised manuscript received 20 March 2002; published 19 July 2002)

We study the problem of simultaneous estimation of parameters and unobserved states from noisy data of nonlinear time-continuous systems, including the case of additive stochastic forcing. We propose a solution by adapting the recently developed statistical method of unscented Kalman filtering to this problem. Due to its recursive and derivative-free structure, this method minimizes the cost function in a computationally efficient and robust way. It is found that parameters as well as unobserved components can be estimated with high accuracy, including confidence bands, from heavily noise-corrupted data.

DOI: 10.1103/PhysRevE.66.016210 PACS number(s): 05.45.Tp, 07.05.Kf

I. INTRODUCTION

For a quantitative understanding of time varying phenomena from nature and technology, it is often desired and quite informative to fit coefficients of nonlinear models to time series of observations. These models may contain quantities that cannot be measured directly. Instead, only a rather small portion of noise-corrupted observations is available. Reliable reconstruction of all model components and parameters using such ''real-world data'' is one of the most challenging research topics in nonlinear data analysis. Applications of nonlinear data analysis tools range from physical problems (e.g., the identification of nonlinear electronic circuits and experiments from nonlinear optics, the analysis of pattern forming systems, and the analysis of granular media and astrophysical data) to engineering problems (such as determining nonlinear excitation responses and modeling of complex manufacturing processes), to mention just a few.

Several algorithms have been proposed and successfully applied for considerably restricted model classes for the system and observation process. The restrictions demand only weak nonlinearities or small amounts of noise, for example. Unfortunately, most real-world systems do not possess these properties. The general approach to treat the problem of estimating parametrized models from incomplete time series amounts in a *state space description*. For linear state space models with Gaussian process and observation noise, the well-known Kalman filter [1] is the method of choice for the consistent estimation of the indirectly observed or unobserved states. But for the estimation of parameters, even for linear models this inevitably leads to nonlinear state space equations, which prevent the direct use of the Kalman filter [3,4]. Very recently, Julier and Uhlmann developed a substantial extension of the Kalman filter for nonlinear models, the unscented Kalman filter (UKF) [5]. Compared with the widely used extended Kalman filter [6], nonlinearities are handled in a more superior way in the sense that a better quality of estimates is achieved with less computational expense.

The prediction-correction structure of the Kalman filter approach in general and the superior handling of nonlinearities in the UKF approach in particular become important especially for chaotic systems. Estimation of states and parameters in chaotic systems using only noisy and incomplete observations is a difficult task, because chaotic solutions depend sensitively on initial conditions, and sometimes also on initially chosen parameters. This leads to typically rather complex cost functions that have to be minimized, as was demonstrated for the Lorenz system in Ref. [7]. Unless the initial values are chosen in very close proximity to the (unknown) true values, the convergence of global methods such as initial value approaches [8,9] may not be guaranteed. For this reason, improved techniques such as multiple-shooting methods have been proposed and successfully applied [10,11], but these are expensive to implement and difficult to tune. For our perspective it is more important, that due to their dependence on numerical derivatives in the optimization procedure, they are restricted to models that are differentiable. This may prevent the application to models where the nonlinearities are computationally complex (like in problems of meteorological data assimilation), not differentiable (like discontinuous nonlinearities in many engineering problems), or where an explicit form of the nonlinearity is not known (like in models derived from artificial neural networks). In contrast, the UKF works without using numerical derivatives, which makes an application to these problems feasible. Due to the recursive structure of the Kalman filter, the probability of stopping in a local minimum of the cost function is greatly reduced, and it allows application to unevenly sampled data as well. All these facts motivate to study the potentials and limits of the UKF for state and parameter estimation of chaotic systems.

This paper is organized as follows. First, the mathematical framework of state space modeling is reviewed. Then, the Kalman filter and its extensions are briefly described with emphasis on the unscented Kalman filter. Finally, the unscented Kalman filter is applied to simultaneously estimate states and parameters from noisy data of the periodic Lotka-Volterra, the chaotic Lorenz, and the stochastic van der Pol systems.

II. STATE SPACE REPRESENTATION

We consider models with only few degrees of freedom. Influences of rapidly fluctuating subsystems or unknown dynamics are approximated by a stochastic term ϵ. This noise

term is assumed to be independently and identically distributed over time, and influences the system in such a way that the state becomes a random variable itself. The dynamics or time evolution of this state $\mathbf{X}(t) \in \mathbb{R}^n$ is modeled by a stochastic differential equation, the *state equation*,

$$\dot{\mathbf{X}}(t) = \mathbf{F}(\mathbf{X}(t), \boldsymbol{\lambda}, \boldsymbol{\epsilon}(t)). \tag{1}$$

In general, $\mathbf{F}$ is a nonlinear function of the state, the parameter vector $\boldsymbol{\lambda}$, and the noise $\boldsymbol{\epsilon}$. In a more general setting the parameter vector could also depend on time, but with respect to our applications to only autonomous systems, this dependence is omitted for ease of notation. The impact of the observation process is considered by adding an *observation equation*

$$\mathbf{Y}(t) = \mathbf{H}(\mathbf{X}(t)) + \boldsymbol{\eta}(t) \tag{2}$$

to the state equation, which maps the state $\mathbf{X}(t)$ to the observation $\mathbf{Y}(t) \in \mathbb{R}^m$. Here, $\mathbf{H}$ is the observation function and $\boldsymbol{\eta}$ again denotes independent white Gaussian noise that represents unpredictable distortions and influences occurring during the observation process. Both $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}$ are assumed to be mutually independent and independent from the state and observation. Therefore, $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ do not alter the statistical properties of the noises. The entity of Eqs. (1) and (2) is called *(time-continuous) state space model* [21].

Because observations are sampled not continuously but in a finite amount of time, one has to adapt the problem of modeling accordingly. A discretization of the time-continuous state space model can be accomplished by transforming Eqs. (1) and (2) into corresponding difference equations of the form

$$\mathbf{X}_t = \mathbf{f}(\mathbf{X}_{t-\Delta t}, \boldsymbol{\lambda}, \boldsymbol{\epsilon}_t) \quad (\Delta t > 0), \tag{3}$$

$$\mathbf{Y}_t = \mathbf{h}(\mathbf{X}_t) + \boldsymbol{\eta}_t. \tag{4}$$

Here, $\mathbf{X}_{t-\Delta t}$ and $\mathbf{X}_t$ are the time-discrete states, and $\mathbf{Y}_t$ is the observation or measurement. Equations (3) and (4) are called *discrete state space model* and describe the evolution of the state $\mathbf{X}_t$ and its observation $\mathbf{Y}_t$ with time. The collection of these random variables at times $t_1 = 1, t_2 = 2, \ldots, t_N = t$ form time-discrete stochastic processes $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_t\}$ and $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_t\}$. The statistical properties of both processes, which are functionally related via Eqs. (3) and (4), are contained in the associated joint probability density $\rho_{\mathbf{XY}}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})$, which describes how likely joint realizations $\mathbf{x}$ and $\mathbf{y}$ of the respective processes are. In our applications for deterministic models (the Lotka-Volterra and the Lorenz system) one has $\boldsymbol{\epsilon}_t = 0$, and for each $\mathbf{X}_t$, the realization $\mathbf{x}_t$ is uniquely determined. The mapping $\mathbf{f}$ in Eq. (3) is then defined as

$$\mathbf{f}(\mathbf{x}_{t-\Delta t}, \boldsymbol{\lambda}) = \mathbf{x}_{t-\Delta t} + \int_{t-\Delta t}^{t} \mathbf{F}(\mathbf{x}(T), \boldsymbol{\lambda}) dT. \tag{5}$$

Since this equation generally cannot be solved analytically, a numerical approximation for the function $\mathbf{f}$ has to be applied. It should be mentioned explicitly here that although the state space model is discrete, time-continuous model functions are included in Eq. (3) as well. For the case of stochastic systems with additive noise the respective integral equation is given in Sec. V.

III. STATE AND PARAMETER ESTIMATION

A. Kalman filter

For notational convenience in the following we set $\Delta t = 1$, unless specified otherwise. The filtering problem in statistics in general consists of the determination of the *filter density* $\rho(\mathbf{X}_t | \{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_t\})$ and its evolution in time t. In the special case of linear state space dynamics with mutually uncorrelated Gaussian noises $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_\epsilon)$ and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_\eta)$,

$$\mathbf{X}_t = \mathbf{f}\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t, \tag{6}$$

$$\mathbf{Y}_t = \mathbf{h}\mathbf{X}_t + \boldsymbol{\eta}_t, \tag{7}$$

the filter density is a Gaussian conditioned on all observations up to time t [12]. Since a Gaussian density is described completely by its mean and covariance, it is sufficient to treat the evolution of only these two quantities, instead of the whole density. As has been shown first by Kalman and Bucy [1], there exists an analytical expression that describes exactly the time evolution of mean and variance, often called estimation and estimation error, of a Gaussian filter density. These recursive algebraic equations, the Kalman filter, are reviewed next.

The Kalman filter consists of a prediction and a correction step. Given the state space model Eqs. (6) and (7), the prediction step extrapolates the mean of the filter density, i.e., the prediction for the state estimation $\hat{\mathbf{x}}(t|t-1)$ and its associated observation $\hat{\mathbf{y}}(t|t-1)$, using information of all prior observations $\tilde{\mathbf{Y}} = \{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{t-1}\}$. The most likely values or optimal predictions are given by the conditional expectations [12]

$$\hat{\mathbf{x}}(t|t-1) = E[\mathbf{X}_t | \tilde{\mathbf{Y}}] = E[\mathbf{f}\mathbf{X}_{t-1} | \tilde{\mathbf{Y}}], \tag{8}$$

$$\hat{\mathbf{y}}(t|t-1) = E[\mathbf{Y}_t | \tilde{\mathbf{Y}}] = E[\mathbf{h}\mathbf{X}_t | \tilde{\mathbf{Y}}]. \tag{9}$$

A conditional expectation value of some random variables $\mathbf{A}$ and $\mathbf{B}$ with outcomes $\mathbf{a}$ and $\mathbf{b}$ is defined by $E[\mathbf{A}|\mathbf{B}] = \int \mathbf{a} \rho_{\mathbf{A}|\mathbf{B}}(\mathbf{A} = \mathbf{a} | \mathbf{B} = \mathbf{b}) d\mathbf{a}$ with $\rho_{\mathbf{A}|\mathbf{B}}$ denoting the respective conditional density. Due to linearity, Eqs. (8) and (9) can be solved analytically. The solution for the states is

$$\hat{\mathbf{x}}(t|t-1) = E[\mathbf{f}\mathbf{X}_{t-1} | \tilde{\mathbf{Y}}] = \mathbf{f}E[\mathbf{X}_{t-1} | \tilde{\mathbf{Y}}] = \mathbf{f}\hat{\mathbf{x}}(t-1|t-1), \tag{10}$$

where $\hat{\mathbf{x}}(t-1|t-1)$ denotes the mean of the filter density at time $t-1$. The associated covariances or prediction errors are defined by

$$\mathbf{P}(t|t-1) = E[(\mathbf{X}_t - \hat{\mathbf{x}}(t|t-1))(\mathbf{X}_t - \hat{\mathbf{x}}(t|t-1))^T | \tilde{\mathbf{Y}}], \tag{11}$$

$$\mathbf{P}_{\mathbf{YY}}(t|t-1)=E[(\mathbf{Y}_t-\hat{\mathbf{y}}(t|t-1))(\mathbf{Y}_t-\hat{\mathbf{y}}(t|t-1))^T|\tilde{\mathbf{Y}}], \tag{12}$$

$$\mathbf{P}_{\mathbf{XY}}(t|t-1)=E[(\mathbf{X}_t-\hat{\mathbf{x}}(t|t-1))(\mathbf{Y}_t-\hat{\mathbf{y}}(t|t-1))^T|\tilde{\mathbf{Y}}], \tag{13}$$

where for linear models the expectations can be computed analytically. The correction step then updates the predictions for the state and estimation error using the new observation $\mathbf{y}_t$:

$$\hat{\mathbf{x}}(t|t)=\hat{\mathbf{x}}(t|t-1)+\mathbf{K}_t(\mathbf{y}_t-\hat{\mathbf{y}}(t|t-1)), \tag{14}$$

$$\mathbf{P}(t|t)=\mathbf{P}(t|t-1)-\mathbf{K}_t\mathbf{P}_{\mathbf{YY}}(t|t-1)\mathbf{K}_t^T. \tag{15}$$

Uncertainties given by the errors of the previous prediction step are considered by the *Kalman gain matrix*,

$$\mathbf{K}_t=\mathbf{P}_{\mathbf{XY}}(t|t-1)\mathbf{P}_{\mathbf{YY}}^{-1}(t|t-1), \tag{16}$$

which acts as weight of the innovation $\mathbf{y}_t-\hat{\mathbf{y}}(t|t-1)$, used in Eq. (14). Starting with initial guesses for the state $\hat{\mathbf{x}}(1|1)$ and its covariance matrix $\mathbf{P}(1|1)$, the filter recursively improves the estimations $\hat{\mathbf{x}}(t-1|t-1)$ and $\mathbf{P}(t-1|t-1)$ to $\hat{\mathbf{x}}(t|t)$ and $\mathbf{P}(t|t)$ using information available by the measurement $\mathbf{y}_t$. The Kalman filter yields unbiased and consistent estimates for linear state space models only. In the case of nonlinear models non-Gaussian densities are approximated by Gaussian distributions. The particular choice of mean and covariance of these Gaussians leads both to the extended and unscented Kalman filter.

B. Nonlinear extensions to the Kalman filter

The most widely used approach for filtering nonlinear state space models is to approximate the nonlinear functions **f** and **h** of Eqs. (6) and (7) by their Taylor series expansions in terms of the prediction error $\mathbf{X}_t-\hat{\mathbf{x}}(t|t)$ [12,13]. This is realized by the extended Kalman filter (EKF) of first and second order. The first-order EKF assumes that the state space model is linear within the scale of errors. Therefore, it suffers from second and higher-order linearization errors, the need for Jacobian matrices, and implementation difficulties. If nonlinearities cannot be approximated well by linearized terms (like in dry friction systems, for example), most EKF estimates are biased and inconsistent. The second-order EKF demands extensive implementation efforts due to the need for Hessian matrices. Another way for filtering is to treat the filter density directly by means of a representative set of samples. Applying the state space equations to these sample sets yields new prediction statistics, i.e., mean and covariance, that can be used with the Kalman filter update equations.

C. Unscented Kalman filtering

A novel procedure for dealing with estimation in nonlinear state space models has been proposed recently by Julier and Uhlmann [5]. This procedure belongs to the class of statistical linearization schemes [13,14] in which *densities* are truncated instead of the models **f** and **h**. Higher order moments of the filter density are neglected, i.e., mean and covariance are used only. A sample set with same mean and covariance is generated and propagated through the full (not approximated as with EKF) state space model. Unlike stochastic approaches, e.g., rejection schemes, where large amounts of samples are needed, the idea of Julier and Uhlmann was to use a set that is constructed in a deterministic way and therefore much smaller in size. This sample set $\{\boldsymbol{\mathcal{X}}_i\}_0^{2n}$ is given by the so-called sigma points

$$\boldsymbol{\mathcal{X}}_0(t-1|t-1)=\hat{\mathbf{x}}(t-1|t-1),$$

$$\boldsymbol{\mathcal{X}}_i(t-1|t-1)=\hat{\mathbf{x}}(t-1|t-1)+[\sqrt{(n+\kappa)\mathbf{P}(t-1|t-1)}]_i, \tag{17}$$

$$\boldsymbol{\mathcal{X}}_{i+n}(t-1|t-1)=\hat{\mathbf{x}}(t-1|t-1)-[\sqrt{(n+\kappa)\mathbf{P}(t-1|t-1)}]_i,$$

with $i=1,\ldots,n$. The value of κ is fixed (see below), and $(\sqrt{.})_i$ is the ith row or column of the matrix square root. The data set $\{\boldsymbol{\mathcal{X}}_i\}_0^{2n}$ is propagated through the full nonlinearities **f** and **h** using the prediction rules

$$\boldsymbol{\mathcal{X}}_i(t|t-1)=\mathbf{f}(\boldsymbol{\mathcal{X}}_i(t-1|t-1)), \tag{18}$$

$$\boldsymbol{\mathcal{Y}}_i(t|t-1)=\mathbf{h}(\boldsymbol{\mathcal{X}}_i(t|t-1)). \tag{19}$$

The resulting sample sets $\{\boldsymbol{\mathcal{X}}_i(t|t-1)\}_0^{2n}$ and $\{\boldsymbol{\mathcal{Y}}_i(t|t-1)\}_0^{2n}$ may represent a density with higher order moments but due to linearization only mean

$$\hat{\mathbf{x}}(t|t-1)=\sum_{i=0}^{2n} W_i\boldsymbol{\mathcal{X}}_i(t|t-1), \tag{20}$$

$$\hat{\mathbf{y}}(t|t-1)=\sum_{i=0}^{2n} W_i\boldsymbol{\mathcal{Y}}_i(t|t-1) \tag{21}$$

and covariance

$$\mathbf{P}_{\mathbf{YY}}(t|t-1)=\sum_{i=0}^{2n} W_i\{\boldsymbol{\mathcal{Y}}_i(t|t-1)-\hat{\mathbf{y}}(t|t-1)\}\{\boldsymbol{\mathcal{Y}}_i(t|t-1)-\hat{\mathbf{y}}(t|t-1)\}^T, \tag{22}$$

$$\mathbf{P}_{XY}(t|t-1)=\sum_{i=0}^{2n} W_i\{\boldsymbol{\mathcal{X}}_i(t|t-1)-\hat{\mathbf{x}}(t|t-1)\}\{\boldsymbol{\mathcal{Y}}_i(t|t-1)-\hat{\mathbf{y}}(t|t-1)\}^T, \tag{23}$$

$$\mathbf{P}(t|t-1)=\sum_{i=0}^{2n} W_i\{\boldsymbol{\mathcal{X}}_i(t|t-1)-\hat{\mathbf{x}}(t|t-1)\}\{\boldsymbol{\mathcal{X}}_i(t|t-1)-\hat{\mathbf{x}}(t|t-1)\}^T \tag{24}$$

are used. The weights are defined as

$$W_0 = \frac{\kappa}{n+\kappa},$$

$$W_i = \frac{1}{2(n+\kappa)} \quad (i=1,\ldots,2n). \tag{25}$$

The predicted mean $\hat{\mathbf{x}}(t|t-1)$ and covariance $\mathbf{P}(t|t-1)$ are updated by the Kalman equations (14)–(16) to yield the state estimation $\hat{\mathbf{x}}(t|t)$ and its estimation error $\mathbf{P}(t|t)$ for the nonlinear state space model. The set of Eqs. (17)–(25) is called *unscented transformation*.

In order to control estimation properties, a constant scaling parameter κ is introduced. Problems with scaling arise if $\kappa<0$, because then the predicted covariance may not be positive definite any longer, and the filter would yield inconsistent estimates. This disadvantage can be overcome with the scaled unscented transformation [15] or a square root implementation approach [16]. In our applications $\kappa=0$ is used to keep things simple.

The advantages of the UKF can be summarized as follows: (i) higher order accuracy as compared with the often used EKF of first order, (ii) derivative-free optimization, (iii) easy computational implementation (matrix algebra operations only). More details and investigations related to the UKF and some recently published improvement and theoretical framework are given in Refs. [15–17].

D. Joint state estimation

A possible way to encounter with parameter estimation is to treat the parameter vector as a dynamical variable itself. It is important to note that even if the state space model for $\mathbf{x}_t$ and $\boldsymbol{\lambda}_t$ is linear, i.e., $\mathbf{f}$ and $\mathbf{h}$ are matrices, the joint-model is bilinear in state and parameters. To this end, the parameter vector $\boldsymbol{\lambda}_t$ is modeled by the evolution equation

$$\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_{t-1}. \tag{26}$$

Although the parameter is constant within the state dynamics, it is modified in each recursion step by the measurement update equations (14)–(16), as long as the current value deviates from the true parameter. In order to deal with the noise realizations $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ in a simple way, they are treated as state variables as well [13]. A new, joint state is constructed by merging the signal $\mathbf{x}_t$ with the parameter $\boldsymbol{\lambda}_t$ and noise realizations $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ into the joint state vector $\mathbf{j}_t$ with the time evolution

$$\mathbf{j}_t = \begin{pmatrix} \mathbf{x}_t \\ \boldsymbol{\lambda}_t \\ \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{x}_{t-1},\boldsymbol{\lambda}_{t-1}) + \boldsymbol{\epsilon}_{t-1} \\ \boldsymbol{\lambda}_{t-1} \\ \boldsymbol{\epsilon}_{t-1} \\ \boldsymbol{\eta}_{t-1} \end{pmatrix} = \mathbf{f}^j(\mathbf{j}_{t-1}) \tag{27}$$

and the observation function

$$\mathbf{y}_t = \mathbf{h}^j(\mathbf{j}_t) = \mathbf{h}(\mathbf{x}_t) + \mathbf{h}^\eta(\boldsymbol{\eta}_t). \tag{28}$$

For our case of additive and uncorrelated measurement noise the function $\mathbf{h}^\eta$ reduces to the identity. But it should be mentioned here that measurement disturbances often have to be described in terms of correlated noise components; the function $\mathbf{h}^\eta$ then differs from the identity, e.g., it becomes a nondiagonal matrix. Note that the filter implementation must assure not to alter the noises by the incoming observation $\mathbf{y}_t$ during the update step, because the noises are by definition independent of this observation. Detailed information concerning this issue within the UKF framework can be found in Ref. [5].

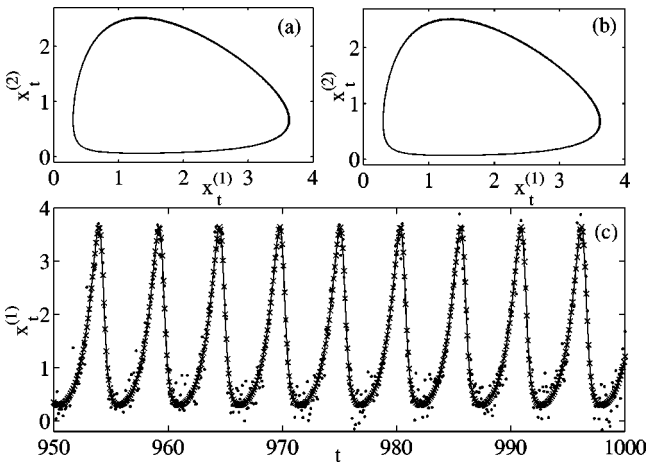

FIG. 1. (a) Limit cycle of the Lotka-Volterra system with parameters $\boldsymbol{\lambda}=(1,1.5,2)^T$ for $t\in[900,1000]$. (b) Estimated limit cycle for $t\in[900,1000]$. (c) Clean ($\times$), noisy ($\cdots$), and estimated (—) time series for $t\in[950,1000]$.

IV. PARAMETER ESTIMATION FOR NONLINEAR DETERMINISTIC SYSTEMS

We demonstrate the potentials of the UKF for two paradigmatic models in nonlinear science, the Lotka-Volterra system [18] and the chaotic Lorenz system [19,23].

A. Lotka-Volterra system

First we estimate the state and parameters for the Lotka-Volterra system,

$$\dot{x}^{(1)} = \lambda^{(1)}x^{(1)} - \lambda^{(2)}x^{(1)}x^{(2)},$$

$$\dot{x}^{(2)} = \lambda^{(2)}x^{(1)}x^{(2)} - \lambda^{(3)}x^{(2)}. \tag{29}$$

Superscripts denote the component of the state $\mathbf{x}_t = (x^{(1)},x^{(2)})^T$ and the parameter $\boldsymbol{\lambda}_t = (\lambda^{(1)},\lambda^{(2)},\lambda^{(3)})^T$. Data of length $N=10000$ are generated by numerically integrating Eq. (29) with the parameters $\lambda^{(1)}=1$, $\lambda^{(2)}=1.5$, and $\lambda^{(3)}=2$, using an embedded Runge-Kutta method [20] with automatic step size control and subsequent sampling with interval $\Delta t=0.1$ (note that the integration step size may differ from the sampling interval). This yields nonlinear periodic oscillations represented by a limit cycle in state space [Fig. 1(a)]. The observations are generated by corrupting $x_t^{(1)}$ with additive Gaussian noise $\boldsymbol{\eta}_t \sim \mathcal{N}(0,\mathbf{P}_\eta)$.

Using the state space formalism, the system

$$\begin{pmatrix} \mathbf{x}_t \\ \boldsymbol{\lambda}_t \\ \boldsymbol{\eta}_t \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{\mathrm{LV}}(\mathbf{x}_{t-\Delta t}, \boldsymbol{\lambda}_{t-\Delta t}) \\ \boldsymbol{\lambda}_{t-\Delta t} \\ \boldsymbol{\eta}_{t-\Delta t} \end{pmatrix}, \tag{30}$$

$$y_t = x_t^{(1)} + \eta_t^{(1)} \tag{31}$$

is obtained, where $\mathbf{f}_{\mathrm{LV}}$ denotes the integral equation (5) for the Lotka-Volterra system with $\Delta t = 0.1$ and $\eta_t^{(1)}$ the first component of $\boldsymbol{\eta}_t$ with $\eta_t^{(1)} \sim \mathcal{N}(0,\sigma^2)$. The amount of observation noise is characterized by the percental signal-to-noise ratio of the standard deviations of the clean state $x_t^{(1)}$ and the noise. In this case σ is 10%. Next, we estimate the parameters and the clean state, under the assumption that the underlying dynamical model and statistics of the measurement noise is known. Initial guesses for parameters are chosen to be twice the true values; $\hat{\lambda}_1^{(1)} = 2$, $\hat{\lambda}_1^{(2)} = 3$, and $\hat{\lambda}_1^{(3)} = 4$. The initial state estimation $\hat{\mathbf{x}}(1|1) = \hat{\mathbf{x}}_1$ is chosen to be $\hat{\mathbf{x}}_1 = (y_1, y_1)^T$. As with every estimation, the error for the estimated quantities is required in order to state the reliability of the results. This information is provided by the covariance matrix. If the estimation errors $\mathbf{j}_t - \hat{\mathbf{j}}(t|t)$ are assumed to be Gaussian, then with probability p the true value of the parameter lies within the respective confidence interval around $\hat{\mathbf{j}}(t|t)$. The confidence interval for each parameter $\hat{\lambda}^{(i)}$ is given by

$$[\hat{\lambda}^{(i)} - c(p)\sqrt{P_{ii}},\ \hat{\lambda}^{(i)} + c(p)\sqrt{P_{ii}}], \tag{32}$$

with P_{ii} being the ith diagonal element of the associated estimated covariance matrix and $c(p)$ a distribution dependent value [21]. For non-Gaussian estimation errors, the value of $c(p)$ cannot necessarily be determined exactly and therefore is approximated by making the assumption of Gaussianity.

The result is

$$\hat{\lambda}^{(1)} = 1.0024, \quad \sqrt{\mathbf{P}_{11}} = 0.0057,$$

$$\hat{\lambda}^{(2)} = 1.4870, \quad \sqrt{\mathbf{P}_{22}} = 0.0119,$$

$$\hat{\lambda}^{(3)} = 1.9844, \quad \sqrt{\mathbf{P}_{33}} = 0.0160.$$

All estimated parameters differ from the true values by less than 1%. The estimated parameters with 95%-confidence intervals calculated from the covariance matrix are given in Fig. 2, the state estimates $x^{(1)}$ and $x^{(2)}$ in Fig. 1. The differences between estimated and clean states are negligible small if related to the difference between clean states and noisy data.

B. Lorenz system

Now we treat the, more complicated case with respect to convergence, of chaotic dynamics. A well known example is the Lorenz system [23]

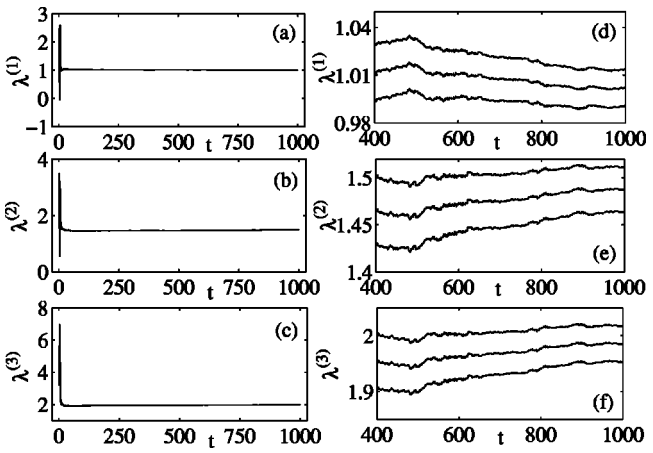

FIG. 2. (a)–(c) Estimated parameters vs time for the Lotka-Volterra system. (d)–(f) Enlargements taken from (a)–(c) show estimates with associated 95%-confidence intervals.

$$\dot{x}^{(1)} = -\lambda^{(1)} x^{(1)} + \lambda^{(1)} x^{(2)},$$

$$\dot{x}^{(2)} = \lambda^{(2)} x^{(1)} - x^{(2)} - x^{(1)} x^{(3)},$$

$$\dot{x}^{(3)} = -\lambda^{(3)} x^{(3)} + x^{(1)} x^{(2)}, \tag{33}$$

with parameters $\lambda^{(1)} = 10$, $\lambda^{(2)} = 46$, and $\lambda^{(3)} = 8/3$. We add a difficulty here by generating unevenly sampled data of length $N = 10\,000$ by integrating these equations with sampling time intervals Δt that vary in a random manner, uniformly distributed within the range $[0.01, 0.05]$. Figure 3(a) shows a projection of the three-dimensional clean chaotic attractor into a two-dimensional subspace spanned by $x_t^{(1)}$ and $x_t^{(2)}$. The observation is constructed by corrupting the component $x_t^{(1)}$ with Gaussian noise $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \mathbf{P}_\eta)$.

Again, in order to estimate states and parameters simultaneously, the problem is written in the state space form as

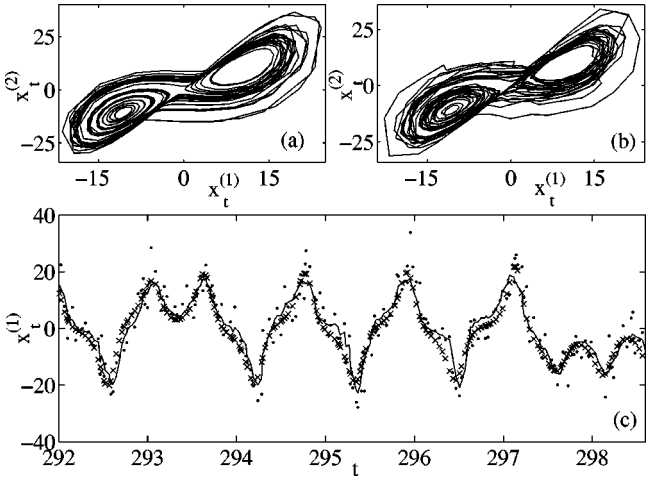

FIG. 3. (a) Projection of the clean attractor of the Lorenz system with parameters $\boldsymbol{\lambda} = (10, 46, 8/3)^T$ into the two-dimensional subspace for $t \in [268, 300]$. (b) Projection of the estimated attractor for $t \in [268, 300]$. (c) Clean ($\times$), noisy ($\cdots$), and estimated (—) time series for $t \in [292, 300]$.

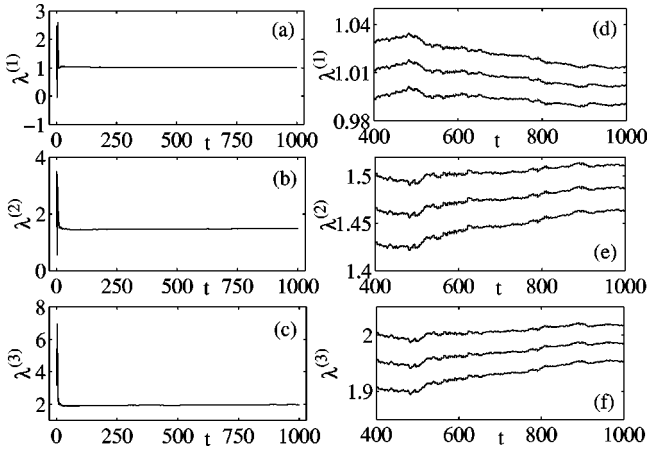

FIG. 4. (a)–(c) Estimated parameters vs time for the chaotic Lorenz system. Enlargements taken from (a)–(c) show estimates with associated 95%-confidence intervals (d)–(f).

$$\begin{pmatrix} \mathbf{x}_t \\ \boldsymbol{\lambda}_t \\ \boldsymbol{\eta}_t \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{\mathrm{L}}(\mathbf{x}_{t-\Delta t(t)}, \boldsymbol{\lambda}_{t-\Delta t(t)}) \\ \boldsymbol{\lambda}_{t-\Delta t(t)} \\ \boldsymbol{\eta}_{t-\Delta t(t)} \end{pmatrix}, \tag{34}$$

$$y_t = x_t^{(1)} + \eta_t^{(1)}. \tag{35}$$

The signal-to-noise ratio is set to 50% of the standard deviation of the clean signal. The integral $\mathbf{f}_{\mathrm{L}}$ in Eq. (34) is solved numerically by the same method as in the previous example. Initial guesses for the joint state are chosen to be half the true values. All components of the initial state $\hat{\mathbf{x}}_1$ are set to the only observed value $(y_1, y_1, y_1)^T$.

Figure 4 shows the parameter estimation, as it develops in time, with associated confidence bands. Convergence is reached for times larger than $t=275$. The final estimates (at $t=300$) are

$$\hat{\lambda}^{(1)} = 10.2588, \quad \sqrt{\mathbf{P}_{11}} = 0.1574,$$

$$\hat{\lambda}^{(2)} = 45.6972, \quad \sqrt{\mathbf{P}_{22}} = 0.4811,$$

$$\hat{\lambda}^{(3)} = 2.6399, \quad \sqrt{\mathbf{P}_{33}} = 0.0437.$$

These estimates differ from the true values $\lambda^{(1)}=10$, $\lambda^{(2)}=46$, and $\lambda^{(3)}=2.6666$ by a few percent only. The true values are located all within the estimated 95%-confidence bounds given by $\hat{\lambda}^{(i)} \pm 1.96\sqrt{\mathbf{P}_{ii}}$. The reconstructed attractor is shown in Fig. 3(b), and a visual comparison of noisy, estimated, and clean state variable $x_t^{(1)}$ is given in Fig. 3(c). One observes that the measurement disturbances have been compensated considerably.

V. PARAMETER ESTIMATION FOR LANGEVIN SYSTEMS

In the preceding section the UKF was used for state and parameter estimation of deterministic state equations. Now we treat the more general case of systems with additive dynamical noise, i.e., systems described by Langevin equations.

A. General setup

We consider systems of the form

$$\dot{\mathbf{X}}(t) = \mathbf{F}(\mathbf{X}(t), \boldsymbol{\lambda}) + \boldsymbol{\epsilon}(t), \tag{36}$$

with constant parameter $\boldsymbol{\lambda}$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{\epsilon})$. Again, due to time discrete observations, one needs to integrate Eq. (36) over the sampling time interval Δt. Then the function $\mathbf{f}$ in Eq. (3) reads

$$\mathbf{f}(\mathbf{X}_{t-\Delta t}, \boldsymbol{\lambda}, \boldsymbol{\epsilon}_t) = \mathbf{X}_{t-\Delta t} + \int_{t-\Delta t}^{t} \mathbf{F}(\mathbf{X}(T), \boldsymbol{\lambda}) + \boldsymbol{\epsilon}(T), dT. \tag{37}$$

Contrary to deterministic systems, for a numerical integration of this equation only low order integration methods, like the explicit Euler scheme, can be applied [2,21,22]. This is due to intractable stochastic integrals that would occur on the right-hand side of Eq. (37) when constructing higher order integration schemes like the Runge-Kutta method of fourth order. Using a low order scheme enforces an integration step δt that often is considerably smaller than a reasonable sampling time interval Δt. Therefore, several integration steps from time $t-\Delta t$ onwards are needed in order to predict the statistics of the state at the update time *t*. The Euler scheme for Langevin equations of the type Eq. (37) reads

$$\mathbf{X}_{t-\Delta t+\delta t} = \mathbf{X}_{t-\Delta t} + \delta t \mathbf{F}(\mathbf{X}_{t-\Delta t}, \boldsymbol{\lambda}) + \sqrt{\delta t}\, \boldsymbol{\epsilon}_{t-\Delta t}. \tag{38}$$

Since a random variable like the state $\mathbf{X}_{t-\Delta t}$ is completely characterized by its probability density, it is sufficient to look for the evolution of this density. This can be done using the respective discrete Fokker-Planck equation or simulations utilizing a discretization of the density by a finite, but representative, set of samples $\{\mathbf{x}_{i,t-\Delta t}\}$. The discrete density at time $t-\Delta t+\delta t$ then is obtained by propagating the samples according to

$$\mathbf{x}_{i,t-\Delta t+\delta t} = \mathbf{x}_{i,t-\Delta t} + \delta t \mathbf{F}(\mathbf{x}_{i,t-\Delta t}, \boldsymbol{\lambda}) + \sqrt{\delta t}\, \boldsymbol{\epsilon}_{i,t-\Delta t}. \tag{39}$$

Here, $\boldsymbol{\epsilon}_{i,t-\Delta t}$ denotes the ith sample of the stationary Gaussian white noise $\boldsymbol{\epsilon}_{t-\Delta t}$. With respect to the UKF, the sets $\{\mathbf{x}_{i,t-\Delta t}\}$ and $\{\boldsymbol{\epsilon}_{i,t-\Delta t}\}$ are chosen as the sigma points $\{\boldsymbol{\mathcal{X}}_i(t-\Delta t|t-\Delta t)\}_{i=0}^{2n}$ and $\{\boldsymbol{\mathcal{E}}_i(t-\Delta t|t-\Delta t)\}_{i=0}^{2n}$ of the respective state and noise density. Thus, only information provided by the mean and covariance of the full state density is utilized during each integration step. The numerical integration step over δt is performed by propagating the sigma points through the Euler scheme of the respective stochastic differential equation given by Eq. (39), in shorthand notation written as

$$\boldsymbol{\mathcal{X}}_i(t-\Delta t+\delta t|t-\Delta t)=\mathbf{f}_{\mathrm{EULER}(\delta t)}(\boldsymbol{\mathcal{X}}_i(t-\Delta t|t-\Delta t),\boldsymbol{\lambda},\boldsymbol{\mathcal{E}}_i(t-\Delta t|t-\Delta t)). \tag{40}$$

From the predicted points $\{\boldsymbol{\mathcal{X}}_i(t-\Delta t+\delta t|t-\Delta t)\}_{i=0}^{2n}$ the mean and covariance according to Eqs. (20) and (24) are computed. Using this mean and covariance, a new set of sigma points $\{\boldsymbol{\mathcal{X}}_i(t-\Delta t+\delta t|t-\Delta t+\delta t)\}_{i=0}^{2n}$ is constructed for the next Euler step to be performed at $t-\Delta t+\delta t$. Since the dynamical noise is stationary, the predictions for the sigma points, which represent the density of the noise read $\boldsymbol{\mathcal{E}}_i(t-\Delta t+\delta t|t-\Delta t)=\boldsymbol{\mathcal{E}}_i(t-\Delta t|t-\Delta t)$ for every *i*. It follows the joint state space representation

$$\begin{pmatrix} \boldsymbol{\mathcal{X}}_i(t-\Delta t+\delta t|t-\Delta t) \\ \boldsymbol{\lambda}_i(t-\Delta t+\delta t|t-\Delta t) \\ \boldsymbol{\mathcal{E}}_i(t-\Delta t+\delta t|t-\Delta t) \\ \boldsymbol{\eta}_i(t-\Delta t+\delta t|t-\Delta t) \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{\mathrm{EULER}(\delta t)}(\boldsymbol{\mathcal{X}}_i(t-\Delta t|t-\Delta t),\boldsymbol{\lambda}_i(t-\Delta t|t-\Delta t),\boldsymbol{\mathcal{E}}_i(t-\Delta t|t-\Delta t)) \\ \boldsymbol{\lambda}_i(t-\Delta t|t-\Delta t) \\ \boldsymbol{\mathcal{E}}_i(t-\Delta t|t-\Delta t) \\ \boldsymbol{\eta}_i(t-\Delta t|t-\Delta t) \end{pmatrix}, \tag{41}$$

where a constant parameter $\boldsymbol{\lambda}$ and stationary measurement noise $\boldsymbol{\eta}$ is assumed.

Repeating this unscented transformation for all subsequent integration steps until the sampling time *t* is reached, gives a predicted mean and covariance for the joint state that is updated finally by the observation $\mathbf{y}_t$ according to the UKF rules Eqs. (14) and (15). Next this ''unscented Euler integration scheme'' is used within the UKF framework in order to estimate states and the parameter of the stochastic van der Pol system.

B. Stochastic van der Pol system

The stochastic van der Pol system in first order writing is given by

$$\dot{X}^{(1)}=X^{(2)}, \tag{42}$$

$$\dot{X}^{(2)}=\lambda^{(1)}((1-X^{(1)}X^{(1)})X^{(2)}-X^{(1)})+\epsilon, \tag{43}$$

where the second component is driven stochastically by uncorrelated noise ϵ. This system has been studied extensively, and for the sake of comparability we use the setup proposed in Ref. [2] by fixing $\epsilon\sim\mathcal{N}(0,1)$, $\lambda^{(1)}=3$, and taking an integration step width of $\delta t=0.001$ for the respective explicit Euler scheme $\mathbf{f}_{\mathrm{EULER}(\delta t)}$. This is given by

$$x^{(1)}_{t-\Delta t+\delta t}=x^{(1)}_{t-\Delta t}+\delta t x^{(2)}_{t-\Delta t}, \tag{44}$$

$$x^{(2)}_{t-\Delta t+\delta t}=x^{(2)}_{t-\Delta t}+\delta t\lambda^{(1)}((1-x^{(1)}_{t-\Delta t}x^{(1)}_{t-\Delta t})x^{(2)}_{t-\Delta t}-x^{(1)}_{t-\Delta t}) + \epsilon_{t-\Delta t}. \tag{45}$$

The sampling time is set to $\Delta t=0.1$. Applying the Euler scheme to the stochastic van der Pol system yields a nonlinear periodic oscillation that is slightly disturbed by the dynamical noise. Observations $\{y_t\}_{t=0.1}^{1,000}$ are generated by taking the first state component only and corrupting it with 10% measurement noise $\eta\sim\mathcal{N}(0,\sigma_\eta^2)$. Applying the UKF with starting guesses $\hat{\mathbf{x}}(1|1)=(y_1,y_1)^T$, $\hat{\lambda}^{(1)}(1|1)=6$, and the unscented Euler scheme we obtain for the parameter at $t=1000$,

$$\hat{\lambda}^{(1)}=2.9600, \sqrt{\mathbf{P}_\lambda}=0.0432. \tag{46}$$

Figure 5 shows the convergence behavior of the estimated parameter over time. The final estimated parameter at $t=1000$ differs from the true one by less than 2%. Finally we can state that the UKF in conjunction with the unscented Euler scheme enables reliable parameter estimation from incomplete and noise corrupted data of the stochastic van der Pol system.

VI. DISCUSSION

We have demonstrated that the technique of unscented Kalman filtering enables simultaneous state and parameter

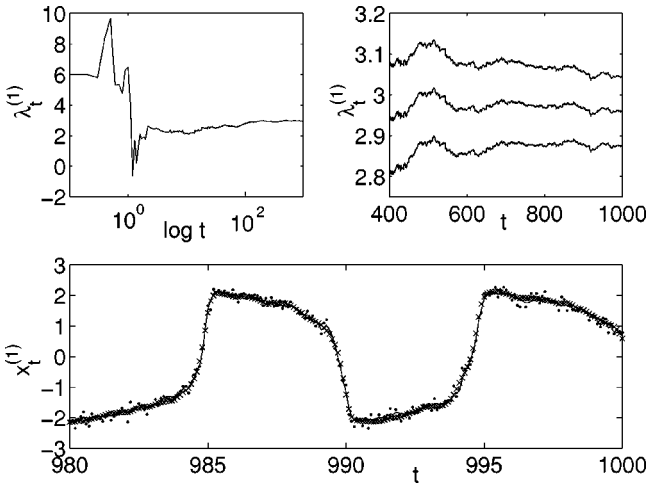

FIG. 5. (a) Estimated parameter vs time for the stochastic van der Pol system. Due to rapid convergence, a logarithmic axis is used for the time. (b) A zoom of the estimated parameter with associated 95% confidence intervals taken from (a). (c) Clean (×), noisy (···), and estimated (–) time series for $t\in[980,1000]$.

estimation from incomplete data of a variety of nonlinear dynamical systems with relatively large amounts of measurement noise. Contrary to methods that yield similar estimation results for deterministic systems [24], the unscented Kalman filter has smaller inplementational efforts and can be used with stochastic systems as well. Unlike in many other approaches, the model nonlinearities are taken as they are and are *not* approximated by a Taylor series expansion. Furthermore, time expensive stochastic simulations are not necessary. This makes this approach very flexible, and it can also be applied to systems where the explicit form of the nonlinearities are not known or derivatives are difficult to compute. Therefore the numerical solver of the model can be used identically also in the model estimation procedure, therefore.

We note that the choice of the initial covariance matrices may be crucial for the convergence rate of the filter. The estimation of the variances, especially of the dynamical noise term, for the case that they are unknown, is an ongoing topic of research. The ability to tune the scaling parameter and the absence of derivatives in the unscented Kalman filter allows for parameter estimation in discontinuous models as well, but also this needs some further treatment [25].

In the examples studied here, the functional form of the state space model must be known beforehand. To our knowledge, there are no nonparametric methods of state estimation which would also allow for the estimation of the functional form of the model. This is possible so far only if all components of the model are observed, and only for small amounts of noise [26–28], problems of much less complexity than those considered here. (These approaches differ from other nonparametric approaches of nonlinear time series analysis based on reconstruction methods [29,30], in that the former yield estimates of equations of motion.) Our future research will concentrate on a combination of parametric modeling and nonparametric approaches [31].

We consider the approach proposed here to be promising in particular for the analysis of physical experiments in which the parametrization cannot be derived from first principles, like in many problems of pattern formation in liquids and granular media [27,32–34], in the modeling of severely nonstationary data that are abundant in life sciences [35–37], or in materials science [38,39]. Finally, we mention that the accurate estimation of a model can help in estimating invariant quantities like Lyapunov exponents from observations [40,41].

ACKNOWLEDGMENTS

A.S., U.S., and J.K. acknowledge financial support from the Volkswagen-Stiftung, H.U.V. thanks Jens Timmer for stimulating discussions.

[1] R. E. Kalman and R. S. Bucy, J. Basic Eng. **95**, 108 (1960).
[2] J. Timmer, Chaos, Solitons Fractals **11**, 2571 (2000).
[3] A. Dempster, N. M. Laird, and D. B. Rubin, J. R. Stat. Soc. Ser. B Methodal. **39**, 1 (1977).
[4] H. Cox, IEEE Trans. Autom. Control **AC-9**, 5 (1964).
[5] S. Julier and J. Uhlmann, in Proceedings of Aerosense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation, and Controls, 1997 (unpublished).
[6] L. Ljung, IEEE Trans. Autom. Control **AC-24**, 36 (1979).
[7] R. N. Miller, M. Ghil, and F. Gauthiez, J. Atmos. Sci. **51**, 1037 (1994).
[8] L. Edsberg and P. A. Wedin, Optim. Methods Software **6**, 193 (1995).
[9] M. Dobmann and K. Schittkowski, Technical report, Department of Mathematics, University of Bayreuth, 1995 (unpublished).
[10] R. Bulirsch and J. Stoer, Numer. Math. **8**, 1 (1966).
[11] H. G. Bock, Ph.D. thesis, Bonner Mathematische Schriften, Nr. 183, Bonn, 1987.
[12] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association* (Academic Press, Orlando, 1988).
[13] A. Gelb, *Applied Optimal Estimation* (MIT Press, Cambridge, MA, 1974).
[14] T. Lefebvre, H. Bruyninckx, and J. De Schutter, IEEE Trans. Autom. Control (to be published).
[15] S. Julier, Automatica (to be published).
[16] R. van der Merwe, in *The Square-Root Unscented Kalman Filter for State and Parameter-Estimation*. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (IEEE Signal Processing Society, Piscataway, NJ, 2001).
[17] M. Nørgaard, N. K. Poulsen, and O. Ravn, Technical Report No. IMM-REP-1998-15, 2000 (unpublished).
[18] J. Hofbauer and K. Sigmund, *Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection* (Cambridge University Press, Cambridge, MA, 1988).
[19] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors* (Springer, New York, 1982).
[20] J. R. Dormand and P. J. Prince, J. Comput. Appl. Math. **6**, 19 (1980).
[21] J. Honerkamp, *Stochastic Dynamical Systems* (VCH, New York, 1993).
[22] P. E. Kloeden, E. Platen, and H. Schurz, Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 277 (1991).
[23] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[24] J. Bröcker and U. Parlitz, Chaos **11**, 319 (2001).
[25] A. T. Nelson, Ph.D. thesis, Oregon Graduate Institute of Science and Technology, 2000.
[26] H. Voss and J. Kurths, Phys. Lett. A **234**, 336 (1997).
[27] H. U. Voss, P. Kolodner, M. Abel, and J. Kurths, Phys. Rev. Lett. **83**, 3422 (1999).
[28] H. U. Voss, in *Nonlinear Dynamics and Statistics*, edited by A. I. Mees (Birkhäuser, Boston, 2001), pp. 413–434.
[29] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
[30] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).
[31] J. Timmer, H. Rust, W. Horbelt, and H. U. Voss, Phys. Lett. A **274**, 123 (2000).

[32] R. Gonzáles-García, R. Rico-Martínez, and I. G. Kevrekidis, Comput. Chem. Eng. **22**, S965 (1998).
[33] K. Krischer *et al.*, AIChE J. **39**, 89 (1998).
[34] M. Bär, R. Hegger, and H. Kantz, Phys. Rev. E **59**, 337 (1999).
[35] M. Kremliovsky *et al.*, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 813 (1998).
[36] N. B. Yanson *et al.*, Tech. Phys. Lett. **25**, 412 (1999).
[37] D. Gribkov and V. Gribkova, Phys. Rev. E **61**, 6538 (2000).
[38] R. Hegger *et al.*, Chaos **8**, 1 (1998).
[39] S. Mandelj *et al.*, Int. J. Bifurcation Chaos Appl. Sci. Eng. **11**, 2731 (2001).
[40] U. Parlitz, Int. J. Bifurcation Chaos Appl. Sci. Eng. **2**, 155 (1992).
[41] R. Carretero-González, S. Ørstavik, and J. Stark, Physica D **147**, 204 (2000).