# Identification of nonlinear spatiotemporal systems via partitioned filtering

A. Sitz,[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 2 October 2002; published 8 July 2003)

The problem of identifying continuous spatiotemporal nonlinear systems from noisy and indirect observations is determined by its computational complexity. We propose a solution by means of nonlinear state space filtering along with a state partition technique. The method is demonstrated to be computationally feasible for spatiotemporal data with properties that occur typically in experimental recordings. It is applied to one component of the simulated chaotic data of a two-component reaction diffusion system, yielding estimates of both the unobserved state component and the diffusion constant.

DOI: 10.1103/PhysRevE.68.016202 PACS number(s): 05.45.Tp, 07.05.Kf

## I. INTRODUCTION

Recent investigations into nonlinear systems with a spatiotemporal dynamical behavior initiated a fruitful interaction between experiments and theoretical modeling, especially in the physical and chemical sciences [1–8]. Sometimes, the model structure can be derived directly from data by using inverse approaches [9–14]. For an accurate quantitative description of spatiotemporal dynamics, the exact values of the model parameters need to be known, but these are often not available completely or determined only imprecisely by theory [15,16].

In this paper, the inverse problem of modeling, as a tool to estimate parameters and unobserved states from measurements on systems with spatiotemporal dynamics, is considered. Due to measurement influences, this identification task is not straightforward: Since the data samples provide only an incomplete amount of information, the results of the identification have to be regarded as estimates. These estimates should approximate the true parameters and the true dynamics in some optimal sense. The choice of an appropriate optimality criterion then leads to an associated cost function to be minimized. This minimization may be accomplished by initial value approaches with more or less sophisticated search strategies [17–21]. In case of strong nonlinearities, noise, and indirect observations, often the cost function is so complex that the applicability of the initial value approach is significantly diminished. On the other hand, recursive techniques, to evaluate the cost function in a sequential way, provide a promising solution to accomplish the minimization of complex cost functions of deterministic dynamical systems [22].

Here, recently introduced novel nonlinear filtering techniques [23–26] to estimate parameters and indirectly and unobserved states from spatiotemporal data are utilized. These methods are based on the state space concept of system dynamics. First, the spatiotemporal system is transformed by the method of lines to a system of coupled ordinary differential equations amenable to a treatment with the state space formalism. To solve the inverse problem, then, in principle, filtering could be applied to these high-dimensional ''states.'' Due to the high state dimension, the tracking of the whole filter density by means of Monte Carlo simulations is often out of scope. Even filters that maintain only mean and covariance suffer from limited computer memory capacity.

The method introduced here solves the problem at its root: The global but high-dimensional estimation problem is partitioned into several local low-dimensional ones. An ill-considered reduction of dimensionality would reduce the available information content, too. It follows that if the partition area is smaller than the spatial correlation length, one cannot expect accurate estimates any longer.

The outline of the paper is as follows. First, the state space formalism and its inverse problem, the filtering problem, is introduced briefly. Then, the state space formalism is adapted to deterministic spatiotemporal systems along with the partition method. The method will be elaborated on the estimation of an unobserved component and a parameter of a reaction-diffusion model. Rather than elaborating on mathematical details, we aim at giving a readable introduction focusing on the applicability to experiments.

## II. SPATIOTEMPORAL SYSTEMS AND THE STATE SPACE CONCEPT

### A. General state space modeling

Before considering the full spatiotemporal approach, in this section the state space formalism and the associated inverse problem of low-dimensional systems [27] is introduced for a dynamical system, the time evolution of its state $\mathbf{S}(t) \in \mathbb{R}^{D_s}$ is given by the *system equation*

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

In general, the function $\tilde{\mathbf{F}}$ is nonlinear with respect to the state $\mathbf{S}(t)$. The dynamics depends on a parameter vector $\boldsymbol{\lambda}(t) \in \mathbb{R}^{D_\lambda}$, some external but nonrandom input $\mathbf{u}(t) \in \mathbb{R}^{D_u}$, and a white noise process $\boldsymbol{\epsilon}(t) \in \mathbb{R}^{D_s}$. The stochastic term $\boldsymbol{\epsilon}(t)$ is introduced to approximate rapidly fluctuating subsystems or unknown dynamics, e.g., unpredictable environmental influences.

In view of the parameter estimation, it is convenient to treat $\boldsymbol{\lambda}$ as an additional state component. That is, an extended state $\mathbf{X}(t)$ is constructed by augmenting the state $\mathbf{S}(t)$ with the parameter $\boldsymbol{\lambda}(t)$:

 

$$\mathbf{X}(t)=\begin{pmatrix}\mathbf{S}(t)\\ \boldsymbol{\lambda}(t)\end{pmatrix}, \tag{2}$$

with $\mathbf{X}(t)\in\mathbb{R}^{D_x}$ $(D_x=D_s+D_\lambda)$. For a constant parameter vector, the associated evolution equation, called the *state equation* in the following, is

$$\dot{\mathbf{X}}(t)=\mathbf{F}(\mathbf{X}(t),\boldsymbol{\epsilon}(t),\mathbf{u}(t))=\begin{pmatrix}\tilde{\mathbf{F}}(\mathbf{S}(t),\boldsymbol{\lambda},\boldsymbol{\epsilon}(t),\mathbf{u}(t))\\ \mathbf{0}\end{pmatrix}. \tag{3}$$

Typically, one cannot measure $\mathbf{X}(t)$ directly but only through an observation process which is described by the *observation equation*

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

The observation equation maps the *unobserved state* $\mathbf{X}(t)$ to the *observation* $\mathbf{Y}(t)\in\mathbb{R}^{D_y}$ via the observation function $\mathbf{H}$. Unpredictable influences and distortions occurring during the process of observation are represented by the white noise process $\boldsymbol{\eta}(t)\in\mathbb{R}^{D_y}$. Both the dynamical noise $\boldsymbol{\epsilon}(t)$ and the observational noise $\boldsymbol{\eta}(t)$ are assumed to be mutually uncorrelated over time, i.e., their joint probability density factorizes for all times. The entity of Eqs. (3) and (4) is referred to as (*time-continuous*) *state space model*.

Because observations can be sampled at discrete times only, the problem of modeling has to be adapted accordingly by means of discretization with respect to the sampling time interval $\Delta t$. For notational convenience in the following, it is set to $\Delta t=1$. Discretization can be accomplished for the time-continuous state space model (3) and (4) by transforming it to the corresponding *discrete state space model*

$$\mathbf{X}_t=\mathbf{f}(\mathbf{X}_{t-1},\boldsymbol{\epsilon}_{t-1},\mathbf{u}_t), \tag{5}$$

$$\mathbf{Y}_t=\mathbf{H}(\mathbf{X}_t)+\boldsymbol{\eta}_t. \tag{6}$$

The mapping $\mathbf{f}$ is given by the integral equation

$$\mathbf{X}_t=\mathbf{X}_{t-1}+\int_{t-1}^{t}\mathbf{F}(\mathbf{X}(T),\boldsymbol{\epsilon}(T),\mathbf{u}_t)dT. \tag{7}$$

Since this stochastic functional, in general, cannot be solved analytically, numerical approximation schemes have to be employed.

### B. State space filtering

The estimation of states amounts to the estimation of probability densities; the estimated densities provide both the state estimates (e.g., by the mean value of the density estimates) and the estimation uncertainty (e.g., by the variance of the density estimate).

Therefore, the problem of retrieving the density of a state $\mathbf{X}_t$ from indirect and noisy observations is faced. Note that the state contains the parameter vector $\boldsymbol{\lambda}$ which will be estimated simultaneously with the trajectories. Since the density of $\mathbf{X}_t$ is conditioned on the given observations, it is often referred to as an *inverse density*. Finding the inverse density can be achieved in a recursive manner, i.e., the density at time $t$ is constructed from the density at time $t-1$ and the new incoming observation at time $t$. The theorem of Bayes and the theory of Markov chains provide the framework for recursive or sequential estimation of the indirectly observed state from data:

The collection of random variables $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_t$ forms a stochastic process $\mathbf{X}=\{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_t\}$ with outcomes $\mathbf{x}_{1:t}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_t\}$. The statistical properties of such a process are given by the joint probability density $\rho_{\mathbf{X}}(\mathbf{X}=\mathbf{x}_{1:t})$, or in short $\rho(\mathbf{x}_{1:t})$. In order to describe dependencies between the process variables, a conditional density is defined by

$$\rho(\mathbf{x}_t|\mathbf{x}_{1:t-1})=\frac{\rho(\mathbf{x}_{1:t})}{\rho(\mathbf{x}_{1:t-1})}, \tag{8}$$

which relates $\mathbf{X}_t$ to its predecessors $\mathbf{X}_{t-1},\mathbf{X}_{t-2},\ldots,\mathbf{X}_1$. A process is called a Markov chain if the conditional density in Eq. (8) simplifies to the density $\rho(\mathbf{x}_t|\mathbf{x}_{t-1})$. The state variable $\mathbf{X}_t$ in Eq. (5) depends only on its predecessor $\mathbf{X}_{t-1}$, which leads to the interpretation of the state equation (5) as a Markov chain with the conditional density $\rho(\mathbf{x}_t|\mathbf{x}_{t-1})$. The statistical properties of both processes $\mathbf{X}$ and $\mathbf{Y}=\{\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_t\}$, which are related by Eq. (6), are described by the associated joint probability density $\rho_{\mathbf{XY}}(\mathbf{X}=\mathbf{x}_{1:t},\mathbf{Y}=\mathbf{y}_{1:t})$, or in short $\rho(\mathbf{x}_{1:t},\mathbf{y}_{1:t})$.

In order to relate the probability densities of stochastic processes $\mathbf{X}$ and $\mathbf{Y}$ with the outcomes $\mathbf{x}_{1:t}$ and $\mathbf{y}_{1:t}=\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_t\}$, the Bayes theorem

$$\rho(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})=\frac{\rho(\mathbf{y}_{1:t}|\mathbf{x}_{1:t})\rho(\mathbf{x}_{1:t})}{\rho(\mathbf{y}_{1:t})} \tag{9}$$

is utilized. Since the density $\rho(\mathbf{x}_{1:t})$ contains the information about the state prior to the observation $\mathbf{y}_{1:t}$, it is called the *prior density* or simply *prior*. The information of the observation $\mathbf{y}_{1:t}$ given by the *likelihood density* $\rho(\mathbf{y}_{1:t}|\mathbf{x}_{1:t})$ is used to update the prior to the density $\rho(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ called the *posterior density* or *posterior*. Within the framework of the state space formalism, the state process is a Markov chain, i.e., $\rho(\mathbf{x}_t|\mathbf{x}_{1:t-1})=\rho(\mathbf{x}_t|\mathbf{x}_{t-1})$, and the observation density depends on the actual state $\mathbf{x}_t$ only, i.e., $\rho(\mathbf{y}_t|\mathbf{x}_{1:t},\mathbf{y}_{1:t-1})=\rho(\mathbf{y}_t|\mathbf{x}_t)$. A recursive scheme for the required inverse density, a marginal density of the posterior, is obtained by

$$\rho(\mathbf{x}_t|\mathbf{y}_{1:t})=\int\rho(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})d\mathbf{x}_{1:t-1}, \tag{10}$$

where the actual state $\mathbf{x}_t$ is conditioned on all observations up to time $t$. Applying the chain rule of Bayes and marginalization, the so-called *filter density* $\rho(\mathbf{x}_t|\mathbf{y}_{1:t})$ is decomposed into two equations. These relate the transition densities of the state and observation equation, $\rho(\mathbf{x}_t|\mathbf{x}_{t-1})$ and $\rho(\mathbf{y}_t|\mathbf{x}_t)$, to the prior

$$\rho(\mathbf{x}_t|\mathbf{y}_{1:t-1})=\int\rho(\mathbf{x}_t|\mathbf{x}_{t-1})\rho(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \tag{11}$$

and the filter density

$$\rho(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{\rho(\mathbf{y}_t|\mathbf{x}_t)\rho(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int \rho(\mathbf{y}_t|\mathbf{x}_t)\rho(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t}, \tag{12}$$

respectively. Equation (11) predicts the new state utilizing the employing information of all observations up to time $t-1$. The new incoming observation at time $t$ is used then to correct this prediction by Eq. (12).

If the state and observation are governed by a linear dynamics, with Gaussian white noises $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_\epsilon)$ and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_\eta)$, the filter equations can be solved analytically, as was shown first by Kalman and Bucy [28,29]. The case of linear dynamics is not of much interest for parameter estimation, because even for linear dynamics the state space model becomes nonlinear if the extended state method is used for parameter estimation.

Analytical solutions, in general, cannot be found, therefore approximations to the filter equations are needed: For weakly nonlinear models, the optimal filter equation may be approximated by means of linearization of the system and observation function. For this so-obtained extended Kalman filter (EKF), explicit system function derivatives have to be provided. This may become rather involved quite rapidly for higher-order approximations. Besides, the rather restrictive approximation used within the EKF sometimes leads to strongly biased and inconsistent estimates [30].

An alternative approximation approach is provided by Monte Carlo methods, e.g., particle filter or Monte Carlo-Markov chain algorithms. Although they may yield better estimation results, accuracy is increased for the prize of numerical expense due to cumbersome stochastic simulations. Moreover, the convergence rate is rather slow and must be accelerated by further approximations.

A hybrid approach that unites advantages of Monte Carlo methods (better accuracy) with the Kalman filter (easy to handle the update equations for the states) is the recently proposed unscented Kalman filter (UKF) [23–26]. Due to the technique of ''deterministic sampling,'' a considerable reduction of the sample size is achieved. Despite its numerical simplicity, the unscented Kalman filter estimates for mean $\mathbf{x}_t$ and covariance $\mathbf{P}_t$ are equally or more accurate than those of the EKF of first order [23,30]. The UKF is numerically easier to handle, compared with Monte Carlo methods, while giving better estimates, compared with the EKF. It has been successfully applied so far to the estimation of unobserved states and parameters in nonlinear models given by neural networks [31] and ordinary and stochastic differential equations [22].

### C. The unscented Kalman filter

The UKF algorithm for state space models with mutually uncorrelated additive Gaussian noises consists of the following steps [32,33]:

(i) Select an initial value $\hat{\mathbf{X}}_1 \sim \mathcal{N}(\hat{\mathbf{x}}_1, \hat{\mathbf{P}}_1)$. In the case of deterministic state dynamics, the covariance $\hat{\mathbf{P}}_1$ may be interpreted as an uncertainty about the chosen initial guess $\hat{\mathbf{x}}_1$ for the true state.

(ii) Sampling: Calculate a set of $K := 2D_x + 1$ sample points, the $\sigma$ *points*, from the given mean and covariance of the filter density $\mathcal{N}(\hat{\mathbf{x}}_{t-1}, \hat{\mathbf{P}}_{t-1})$ for $t \geqslant 2$:

$$\mathbf{x}_{t-1}^{(i)} = \hat{\mathbf{x}}_{t-1} + \alpha(\sqrt{D_x \hat{\mathbf{P}}_{t-1}})_i \quad (i = 1, \ldots, D_x), \tag{13}$$

$$\mathbf{x}_{t-1}^{(i)} = \hat{\mathbf{x}}_{t-1} - \alpha(\sqrt{D_x \hat{\mathbf{P}}_{t-1}})_i \quad (i = D_x + 1, \ldots, 2D_x), \tag{14}$$

$$\mathbf{x}_{t-1}^{(K)} = \hat{\mathbf{x}}_{t-1}, \tag{15}$$

where $(\sqrt{.})_i$ denotes the $i$th row of the matrix square root. The parameter $\alpha$ $(0 < \alpha \leqslant 1)$ determines the ''spread'' of the $\sigma$ points around $\hat{\mathbf{x}}_{t-1}$.

(iii) Prediction: Propagate the $\sigma$ points through the state and observation equations. This means, compute $\mathbf{f}(\mathbf{x}_{t-1}^{(i)})$ and $\mathbf{h}(\mathbf{f}(\mathbf{x}_{t-1}^{(i)}))$ for all $\sigma$ points $\mathbf{x}_{t-1}^{(i)}$ in order to estimate the mean and covariance of the predicted prior and likelihood by

$$\hat{\mathbf{x}}_{t|t-1} = \sum_{i=1}^{K} w_i \mathbf{f}(\mathbf{x}_{t-1}^{(i)}), \tag{16}$$

$$\hat{\mathbf{P}}_{x;t|t-1} = \sum_{i=1}^{K} w_i [\mathbf{f}(\mathbf{x}_{t-1}^{(i)}) - \hat{\mathbf{x}}_{t|t-1}][\mathbf{f}(\mathbf{x}_{t-1}^{(i)}) - \hat{\mathbf{x}}_{t|t-1}]' + (1 - \alpha^2)[\mathbf{f}(\mathbf{x}_{t-1}^{(K)}) - \hat{\mathbf{x}}_{t|t-1}][\mathbf{f}(\mathbf{x}_{t-1}^{(K)}) - \hat{\mathbf{x}}_{t|t-1}]' + \mathbf{P}_\epsilon, \tag{17}$$

$$\hat{\mathbf{y}}_{t|t-1} = \sum_{i=1}^{K} w_i \mathbf{h}[\mathbf{f}(\mathbf{x}_{t-1}^{(i)})], \tag{18}$$

$$\hat{\mathbf{P}}_{y;t|t-1} = \sum_{i=1}^{K} w_i [\mathbf{h}(\mathbf{f}(\mathbf{x}_{t-1}^{(i)})) - \hat{\mathbf{y}}_{t|t-1}][\mathbf{h}(\mathbf{f}(\mathbf{x}_{t-1}^{(i)})) - \hat{\mathbf{y}}_{t|t-1}]' + (1 - \alpha^2)[\mathbf{h}(\mathbf{f}(\mathbf{x}_{t-1}^{(K)})) - \hat{\mathbf{y}}_{t|t-1}][\mathbf{h}(\mathbf{f}(\mathbf{x}_{t-1}^{(K)})) - \hat{\mathbf{y}}_{t|t-1}]' + \mathbf{P}_\eta. \tag{19}$$

The vector products are outer products and the prime denotes the transpose of a vector. The weights are defined to be $w_i = 1/2\alpha^2 D_x$ $(i = 1, \ldots, 2D_x)$ and $w_K = 1 - 1/\alpha^2$. Assume that $\mathbf{f}$ and $\mathbf{h}$ are polynomials of order less than three and noises are additive and Gaussian, then the estimates of the moments given in Eqs. (16)–(19) are exact. The state dimension $D_x$ should be canceled out from the $\sigma$ point computation by setting $\alpha = 1/\sqrt{D_x}$. This ensures, for example, comparability of estimation results stemming from different state dimensions.

(iv) Update: Finally, correct the predicted moments by the data $\mathbf{y}_t$ using the Kalman update equations

$$\hat{x}_t = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}), \tag{20}$$

$$\hat{\mathbf{P}}_t = \hat{\mathbf{P}}_{x;t|t-1} - \mathbf{K}_t \hat{\mathbf{P}}_{y;t|t-1} \mathbf{K}_t', \tag{21}$$

with

$$\mathbf{K}_t = \hat{\mathbf{P}}_{xy}\hat{\mathbf{P}}_{y;t|t-1}^{-1} \tag{22}$$

and

$$\hat{\mathbf{P}}_{xy} = \sum_{i=1}^{K} w_i[\mathbf{f}(\mathbf{x}_{t-1}^{(i)}) - \hat{\mathbf{x}}_{t|t-1}][\mathbf{h}(\mathbf{f}(\mathbf{x}_{t-1}^{(i)})] - \hat{\mathbf{y}}_{t|t-1}]' + (1 - \alpha^2)[\mathbf{f}(\mathbf{x}_{t-1}^{(K)}) - \hat{\mathbf{x}}_{t|t-1}][\mathbf{h}(\mathbf{f}(\mathbf{x}_{t-1}^{(K)})] - \hat{\mathbf{y}}_{t|t-1}]'. \tag{23}$$

(v) Proceed recursively with steps (ii)–(iv), using the filter density estimates made one time step before, until convergence for the state estimate is reached.

For deterministic systems, the state $\mathbf{x}_t$ is uniquely defined and will be called the *true state*. Since initially only a guess $\hat{\mathbf{x}}_1$ with some uncertainty $\hat{\mathbf{P}}_1$ is given, the state can be interpreted formally as a random variable with Gaussian density: $\hat{\mathbf{X}}_t \sim \mathcal{N}(\hat{\mathbf{x}}_t, \hat{\mathbf{P}}_t)$. Its estimate is given by the mean value $\hat{\mathbf{x}}_t$. Note that the dynamics itself remains deterministic, i.e., $\boldsymbol{\epsilon}_t = \mathbf{0}$. With an increasing amount of data, the estimated filter density becomes narrower. The limit is determined by the approximation errors of the UKF. Such errors arise, for example, in the case of polynomial order higher than two in $\mathbf{f}$ or $\mathbf{h}$, for which the UKF estimates become biased.

The state and observation equations are not affected by the UKF in any way, unlike as with the EKF (truncated Taylor expansion) and the DD2 filter (truncated Stirling expansion) [30]. The only approximation here is to neglect cumulants of the order higher than two for the densities considered. If higher-order cumulants become significant, e.g., for long-tailed or multimodal distributions, the UKF estimates are inconsistent and therefore do not sufficiently reflect the main properties of true dynamics any longer. One of the main advantages of the UKF is that there is no need for a computation of derivatives with respect to the state. This allows for the straightforward use of state space models that contain nondifferentiable terms or models where the Jacobian cannot be computed easily. This is often the case for high-dimensional systems occurring with partial differential equations. The problem of adapting partial differential equations to the state space formalism is treated in the following section.

## III. PARAMETER AND STATE ESTIMATION IN SPATIOTEMPORAL SYSTEMS

### A. State space modeling of spatiotemporal systems

The evolution equation of a deterministic and time-homogeneous spatiotemporal system with only local interactions is given by

$$\partial_t \mathbf{V}(\mathbf{r},t) = \tilde{\mathbf{F}}(\mathbf{V}, \mathbf{r}, \boldsymbol{\lambda}, \partial_{\mathbf{r}}), \tag{24}$$

where $\mathbf{V}(\mathbf{r},t) \in \mathbb{R}^{D_v}$ denotes the system's state variable, $\mathbf{r} = (r_1, r_2, \ldots, r_{D_r})' \in \mathbb{R}^{D_r}$ the spatial coordinate, $\boldsymbol{\lambda} \in \mathbb{R}^{D_\lambda}$ a parameter vector, $t \geqslant t_1$ the time, and $\tilde{\mathbf{F}}$ a nonlinear functional. Equation (24), in conjunction with initial conditions $\mathbf{V}(\mathbf{r},t_1)$ and boundary conditions $\overline{\mathbf{V}}(t) = \mathbf{V}(\overline{\mathbf{r}},t)$ ($\overline{\mathbf{r}}$ being the boundary coordinates of the system), completely governs the spatiotemporal evolution of the variable $\mathbf{V}(\mathbf{r},t)$. This defines the system's spatiotemporal trajectory.

Continuous spatiotemporal trajectories cannot be observed, rather $\mathbf{V}(\mathbf{r},t)$ can be taken at discrete sampling points. A discrete spatiotemporal trajectory is then a set $\{\mathbf{V}_{\mathbf{r},t}\}$ with $\mathbf{r} = \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_r}$ and $t = t_1, t_2, \ldots, t_{N_t}$, where $N_r$ denotes the number of spatial sampling points. For notational convenience and with respect to the examples to come along, the description is reduced to the case of two spatial dimensions, $D_r = 2$, i.e., $\mathbf{r}_i = (r_{1,i}, r_{2,i})'$. Note that the procedure introduced next is valid for three-dimensional systems, for example, as well. Instead of indexing each point directly, the more compact writing $\{\mathbf{V}_{\mathbf{r},t}\}$ enables us to save indices and to define the method for arbitrarily shaped regions in space. For example, in some experiments there are obstacles, e.g., measurement devices, that would otherwise prevent a straightforward indexing of points. What is lost here is that the neighborhood relationships cannot be seen immediately anymore, but this will not turn out to be of much importance.

The elements of the *pattern* at some fixed time *t*, i.e., the set $\{\mathbf{V}_{\mathbf{r},t}\}$ with $\mathbf{r} = \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_r}$, may be interpreted as representing the components of a state vector $\mathbf{S}_t$ with dimension $D_s = N_r D_v$. In the following, the words ''pattern'' and ''state vector of a pattern'' are used synonymously.

In order to get the pattern $\mathbf{S}_t$ from the pattern $\mathbf{S}_{t-1}$, a numerical integration over time according to Eq. (24) has to be carried out. If the method of lines (MOL) is used, the state $\mathbf{S}_t$ is the solution of a system of coupled ordinary differential equations in time,

$$\dot{\mathbf{S}}(t) = \tilde{\mathbf{F}}_{\mathrm{MOL}}(\mathbf{S}(t), \boldsymbol{\lambda}). \tag{25}$$

With $\mathbf{S}_{t-1}$ being the initial condition, the discrete state equation thus reads

$$\mathbf{S}_t = \mathbf{S}_{t-1} + \int_{t-1}^{t} \tilde{\mathbf{F}}_{\mathrm{MOL}}(\mathbf{S}(T), \boldsymbol{\lambda}) dT =: \mathbf{f}_s(\mathbf{S}_{t-1}, \boldsymbol{\lambda}). \tag{26}$$

Considering observations of spatiotemporal trajectories $\mathbf{S}_t$, an observation equation

$$\mathbf{Y}_t = \mathbf{h}_s(\mathbf{S}_t) + \boldsymbol{\eta}_t \tag{27}$$

is added, accounting for indirect and imprecise observations. Here, $\boldsymbol{\eta}_t$ represents a white noise process, $\mathbf{h}_s : \mathbb{R}^{D_s} \rightarrow \mathbb{R}^{D_y}$ is the observation mapping, where $D_y$ again denotes the number of components that are observed (usually $D_y \leqslant D_s$). Equations (26) and (27) then correspond to the state space model given in Eqs. (5) and (6).

### B. Partitioned filtering

The state vector of a spatiotemporal problem of even low dimension $D_v$ turns out to be high dimensional for realistic grid sizes. For example, taking sample points lying on a two-dimensional square grid with $N_{r_1} = N_{r_2} = 60$, one has $N_r = 3600$, and with $D_v = 2$ one obtains $D_s = 7200$. Therefore, the covariance matrix has $52 \times 10^6$ entries. This turns out to be computationally infeasible, since matrix inversions,

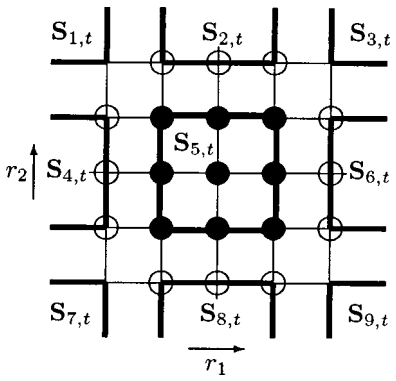


FIG. 1. Piece of a partitioned pattern: Thin lines correspond to the grid lines and thick lines mark the boundaries of the respective patch $\mathbf{S}_{i,t}$. Grid points needed with the estimation of the patch $\mathbf{S}_{5,t}$ are marked by filled and open circles, the latter being exterior points.

for example, are needed in the UKF update equations. To handle this problem, the estimation procedure is reformulated by an approximative decoupled estimation of an ensemble of *patches*. Patches are defined to be arbitrarily shaped and disjunct sets that constitute the pattern, if patched together in a way to be specified further. The state of each patch is denoted by $\mathbf{S}_{\omega,t}$ with $\omega=1,\ldots,N_p$, where $N_p$ is the number of patches. In analogy to the pattern, the phrases ''patch'' and ''state vector of a patch'' are used synonymously in the following. All points that lie outside a patch are called its *exterior points*.

First, an appropriate state equation is constructed for each patch by the method of lines, i.e., the spatial differential operator in Eq. (26) is substituted by a difference scheme. As a result, one gets a system of coupled ordinary differential equations called, in the following, the ''global state equation.'' In order to get $\mathbf{S}_{\omega,t}$, its initial conditions $\mathbf{S}_{\omega,t-1}$ and some exterior points are needed.

For example, let us assume $D_r=2$, $D_v=1$, $N_{r_1}=N_{r_2}=9$ and $\omega=1,2,\ldots,9$, i.e., a pattern $\mathbf{S}_t$ with $D_s=81$ at some time $t$ is decomposed into nine patches $\mathbf{S}_{1,t},\ldots,\mathbf{S}_{9,t}$ of equal sizes and shapes. Each $\mathbf{S}_{i,t}$ again consists of nine elements $S_{r_{1,j},r_{2,k},t}$, where $j$ and $k$ indicate their positions. Let us assume, furthermore, that only immediate spatial neighboring points are coupled. In order to get, for example, $\mathbf{S}_{5,t}$ from $\mathbf{S}_{5,t-1}$, using a central difference scheme for the approximation of spatial differential operators, the relevant part of $\mathbf{S}_t$ results as is shown in Fig. 1.

The evolution equations for the elements of $\mathbf{S}_{5,t}$ are provided by a subset of the global state equation that may utilize components of the neighboring patches $\mathbf{S}_{2,t}$, $\mathbf{S}_{4,t}$, $\mathbf{S}_{6,t}$, and $\mathbf{S}_{8,t}$. To summarize, the question of a dynamical partition corresponds to the issue of how to proceed with these exterior points.

The simplest partition is given by the exclusion of the exterior points. This can be accomplished by using noncentral spatial difference schemes in the method of lines to solve the partial differential equation. In this case, it is impossible to benefit from the exterior points as a source of information, though. It should be better to incorporate exterior information into each local estimation problem. In particular, a possible solution is to require the exterior points of each patch to be constant *numerical* values rather than dynamical variables during the integration from $t-1$ to $t$. In this way, dynamical decoupling of the patch is achieved while still maintaining exterior information. The constants are taken to be the corresponding initial values of the neighboring patches, i.e., the respective components of $\mathbf{S}_{2,t-1}$, $\mathbf{S}_{4,t-1}$, $\mathbf{S}_{6,t-1}$, and $\mathbf{S}_{8,t-1}$ in Fig. 1. If the integration from $t-1$ to $t$ has been performed for each patch, then the same pattern is obtained as one would get by solving the global state equation. Consequently, in this case not only exterior, but global information is utilized.

Formally, the exterior points are interpreted as introduced by an external input $\mathbf{u}_{\omega,t}$ to the state equation of each patch. One could think of ''controlling'' the local prediction by some additional information of exterior points provided through $\mathbf{u}_{\omega,t}$. Thus, the state equation for $\mathbf{S}_{\omega,t}$ of each patch reads

$$\mathbf{S}_{\omega,t}=\mathbf{f}_{\omega,s}(\mathbf{S}_{\omega,t-1},\boldsymbol{\lambda},\mathbf{u}_{\omega,t}). \tag{28}$$

Analogously, the observation equation for each patch reads

$$\mathbf{Y}_{\omega,t}=\mathbf{h}_{\omega,s}(\mathbf{S}_{\omega,t})+\boldsymbol{\eta}_{\omega,t}. \tag{29}$$

The predictions for state and observations can then be updated using the UKF rules (20) and (21). This procedure has to be repeated for all patches at time $t-1$ in order to get the estimated pattern $\hat{\mathbf{S}}_t=\{\hat{\mathbf{S}}_{1,t},\ldots,\hat{\mathbf{S}}_{N_p,t}\}$. For patterns produced by a deterministic system, the estimate of the true pattern $\mathbf{v}_t$ is given by the mean $\hat{\mathbf{s}}_t$, while $\hat{\mathbf{P}}_t$ represents the associated estimation errors.

The question that arises is: how the patch sizes should be chosen to get good estimation results? It stands to reason that the best estimate would result from solving the global problem, but as mentioned earlier this may be computationally intractable. Fortunately, due to decreasing spatial coupling strengths with increasing distances of the respective spatial points, certain covariance entries will not contribute much to the signal's variance, and therefore may be neglected during the estimation procedure. A statistical measure to quantify the dynamical importance of spatial neighbors can be formulated in terms of the correlation length. If the spatial sampling points of the pattern lie on a quadratic and regular grid, the corresponding spatial correlations can be estimated from given data $\mathbf{y}=\{y_{t,i,j},1\leqslant i,j\leqslant\sqrt{N_r},1\leqslant t\leqslant N\}$ by

$$\hat{C}_{k,l}:=\frac{1}{N(N_r-1)}\sum_{t=1}^{N}\sum_{i,j=1}^{\sqrt{N_r}}(y_{t,i,j}-\overline{y_t})(y_{t,i+k-1,j+l-1}-\overline{y_t}), \tag{30}$$

where $k,l=1,2,\ldots,\sqrt{N_r}$, forming the components of the correlation matrix $\hat{\mathbf{C}}$. The correlation lengths with respect to each spatial coordinate are then defined as the lags for which the correlations have decreased by their $e$th part. Taking into account the points that lie outside this range of correlation gives small covariance entries and thus has only minor effects on the estimation result. Consequently, in order to re-

duce the dimensionality while achieving good estimations, the patch sizes should correspond to the correlation range spanned by the correlation lengths.

The *partitioned filtering* allows for an approximation by a considerable reduction of the overall system's dimensionality. For the previous example, the hardly treatable estimation of a system with dimension 81 is reduced to nine estimation problems, each of dimension nine. The partitioned filtering is an approximation to the global filtering problem, since one does not use the global covariance matrix. Due to the nonlinear interactions and the partitioning, an explicit error analysis seems to be, in general, impossible. Therefore, to gain confidence into the results, the applicability of this technique should be validated on simulated data before an application to experimental data is considered.

### C. Parameter estimation in partitioned systems

So far a partition technique for state estimation was considered, i.e., the parameter values were known. A possible way for estimating unknown parameter values is treating the parameter as an additional state space component, as was described for low-dimensional systems in Sec. II A. The extended state vector of dimension $D_x = N_r D_v + D_\lambda$ now is

$$\mathbf{X}_{\omega,t} = \begin{pmatrix} \mathbf{S}_{\omega,t} \\ \boldsymbol{\lambda}_{\omega,t} \end{pmatrix}. \tag{31}$$

It will be called just *state* in the following.

Since the parameter is treated as a state component, some ''dynamics'' has to be defined. If all patches are of the same size and shape, they may be treated as local solutions of the same system but with different initial and boundary conditions. The parameters estimated within a patch then may vary and represent an $N_p$-dimensional ensemble of proposals from which the parameter should be chosen. The most likely parameters with respect to the data follow from the maximum of the density estimated by the frequency distribution of the proposals. For the sake of simplicity, the ensemble density is assumed to be symmetric, such that the location of the maximum is given by its average value. The predicted parameter at time $t$, therefore, is set to be the average of the parameters of all patches at time $t-1$. In state space notation, this reads

$$\mathbf{X}_{\omega,t} = \begin{pmatrix} \mathbf{f}_{\omega,s}(\mathbf{S}_{\omega,t-1}, \boldsymbol{\lambda}_{\omega,t-1}, \mathbf{u}_{\omega,t}) \\ \dfrac{1}{N_p} \displaystyle\sum_{\tilde{\omega}=1}^{N_p} \boldsymbol{\lambda}_{\tilde{\omega},t-1} \end{pmatrix}, \tag{32}$$

$$\mathbf{Y}_{\omega,t} = \mathbf{h}_\omega(\mathbf{X}_{\omega,t}) + \boldsymbol{\eta}_{\omega,t}. \tag{33}$$

This section ends with a summary of the proposed method, given as the following.

(1) Construct a partition.

(2) Construct the extended state vector for each patch.

(3) Define the control inputs for the state by taking the numerical values of the estimates of the respective exterior points one time step before.

(4) Compute the prediction of each patch. The predicted parameter is obtained by averaging over the ensemble of estimated parameters one time step before.

(5) Compute the update for the extended state of each patch.

(6) Proceed with step 3 until convergence for the parameter estimate is reached.

(7) Construct the estimated patterns using the estimated patches. The final parameter estimate is taken to be the ensemble average over the parameter estimates.

## IV. EXAMPLE

The partitioned filtering method is applied to a two-dimensional reaction-diffusion system [34–36]. This system is chosen due to its rich spatiotemporal dynamical behavior and its ability to accurately describe observations in catalytic surface reactions [37,38]. If excited, the underlying chemical system produces concentration differences propagating through the chemical reactor. Target and spiral patterns are observed as well as spiral defect chaos. The underlying complex chemical reaction is simplified to a system of two variables, an activator $v_1$ and an inhibitor $v_2$.

The model is given by

$$\frac{\partial v_1}{\partial t} = -\frac{1}{c_4} v_1 (v_1 - 1)\left( v_1 - \frac{c_1 + v_2}{c_2} \right) + c_3 \left( \frac{\partial^2 v_1}{\partial r_1^2} + \frac{\partial^2 v_1}{\partial r_2^2} \right), \tag{34}$$

$$\frac{\partial v_2}{\partial t} = g(v_1) - v_2, \tag{35}$$

with the piecewise defined function

$$g(v_1) = \begin{cases} 0, & 0 \leqslant v_1 < 1/3 \\ 1 - 6.75 v_1 (v_1 - 1)^2, & 1/3 \leqslant v_1 \leqslant 1 \\ 1, & 1 < v_1. \end{cases} \tag{36}$$

To solve Eqs. (34) and (35) on a rectangular grid with spatial sampling steps $\Delta r = 0.5$ and with periodic boundary conditions numerically, the method of lines is used. The Laplacian is approximated by a difference scheme of second-order accuracy. The spatial resolution is chosen to be $N_{r_1} = N_{r_2} = 60$, while the parameters are set to $c_1 = 0.07$, $c_2 = 0.84$, $c_3 = 1$, and $c_4 = 0.08$. After the excitation of the system and omitting a transient phase, 100 consecutive patterns are recorded with a time step of $\Delta t = 0.04$, an example being shown in Fig. 2: Spiral waves are broken up and form a chaotic pattern for both components of the system.

The inhibitor concentration cannot be observed directly, but its theoretical dynamical behavior including the exact values of some parameters is known. (Parameter estimation for the *completely* observed system has been performed in Refs. [10,17].) Therefore, in this example the typical situation that only the first component $v_1$ is observed is considered. To simulate possible experimental conditions, it is corrupted by an additive noise with $P_\eta = 0.01$ (Fig. 3). The

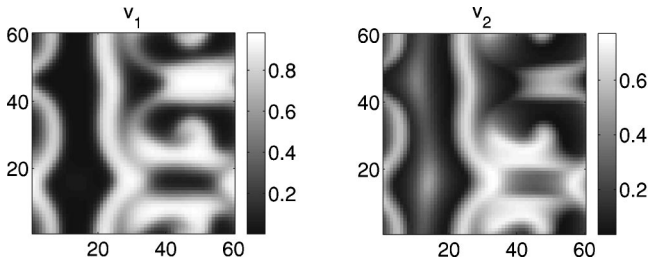


FIG. 2. The true patterns $v_{1,t_N}$ and $v_{2,t_N}$. Abscissa and ordinate values denote indices of grid points. The gray-scale units are dimensionless.

diffusioncoefficient $\lambda = c_3$ is assumed to be unknown. In order to provide a partition size, the correlation matrix has been estimated from data by

$$\hat{\mathbf{C}}_{sym} = \tfrac{1}{2}(\hat{\mathbf{C}} + \hat{\mathbf{C}}') \tag{37}$$

giving a correlation length of 6 (in units of the spatial sampling interval). Compared with Eq. (30), the estimator given in Eq. (37) was slightly modified in order to retain the system's isotropic structure, i.e., its invariance against permutation of the spatial coordinates.

Therefore, the problem is the following. Estimate the true quantities $v_1$, $v_2$, and $\lambda$ from the indirect and noisy observations $\mathbf{y}$. In terms of state space modeling, this is the same as the problem of estimating $\hat{\mathbf{X}}_t$, which has the dimension of $D_x = D_v N_r + D_\lambda = 7201$.

The estimated states $v_{1,t_N}$ and $v_{2,t_N}$ are shown for a partition into patches of size $6\times 6$ (Figs. 3 and 4), corresponding to the data correlation length of 6. Besides this, this partition provides the best trade-off between estimation accuracy and computational time. Results concerning other partitions are presented later in this section. The initial state is chosen to be $\hat{\mathbf{S}}_1 \sim \mathcal{N}(\overline{y_1}\cdot \mathbf{1}_{D_s}, \hat{\mathbf{P}}_{s,1})$, where $\overline{y_1} \in \mathbb{R}$ denotes the spatial mean value of the observed pattern $\mathbf{y}_1 \in \mathbb{R}^{N_r}$, $\mathbf{1}_{D_s}$ being the unit column vector of dimension $D_s = D_v N_r$. The entries of the diagonal matrix $\hat{\mathbf{P}}_{s,1} \in \mathbb{R}^{D_s \times D_s}$ are set to 10. The initial parameter estimate is chosen to be twice the true value, with the same covariance as the states. The estimated parameter $\hat{\lambda}_t$ is the average over the parameter ensemble at time $t$. The uncertainty of the final parameter estimate $\hat{\lambda}_{t_N}$, i.e., its covariance, is taken to be $\hat{\mathbf{P}}_{t_N} = \max(\{\hat{\mathbf{P}}_{\omega,t_N}\}_\omega)$ a rather conservative estimation. Additionally, the convergence of the parameter estimate is shown in Fig. 4 for two different initial guesses.

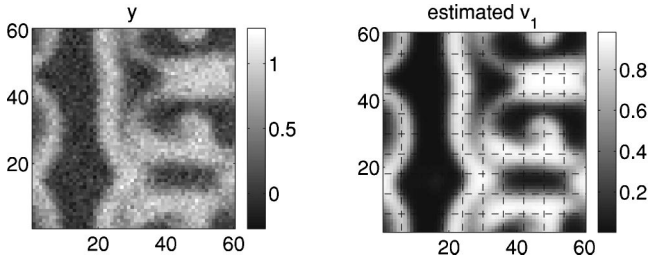


FIG. 3. Left: observation $y_{t_N}$. Right: estimated pattern $\hat{v}_{1,t_N}$. The partition into patches of size $6\times 6$ is marked by dashed lines. The gray-scale units are dimensionless.

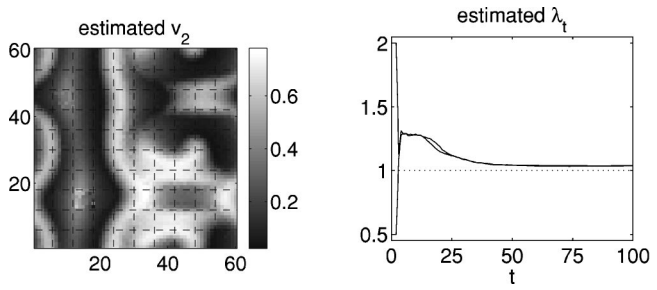


FIG. 4. Left: estimated pattern $\hat{v}_{2,t_N}$. The partition into patches of size $6\times 6$ is marked by dashed lines. The gray-scale units are dimensionless. Right: estimated parameter $\hat{\lambda}_t$ for two different initial guesses. The true parameter $\lambda = 1$ is indicated by a dotted line.

Results obtained for all partitions are summarized in Table I. One observes that the size of the resulting patches considerably influences the computational time and the estimation quality. On one hand, small patches worsen the quality of pattern estimations, since the global covariance matrix is approximated more crudely. The advantage is that the computational time may be lower. On the other hand, large patches give better estimates but suffer from increased computing time. For the example used here, the quality of pattern and parameter estimates do not differ much for sizes larger than patches with size $6\times 6$ (Table I), while the computation time grows drastically (Fig. 5). This confirms the $6\times 6$ partition as an optimal choice. Note that the dependence between the patch size and the computational time is nonmonotonic; a minimum at intermediate values occurs. The optimal partition with respect to computational time or estimation accuracy may differ from system to system and the numerical setup as well.

TABLE I. Estimation and performance results for different partitions. The root mean square error $E_{\mathrm{rms}} = \sqrt{[1/(D_s-1)](\hat{\mathbf{S}}_{t_N}-\mathbf{S}_{t_N})'(\hat{\mathbf{S}}_{t_N}-\mathbf{S}_{t_N})}$ is a measure for the quality of the estimated state $\hat{\mathbf{S}}_{t_N}$. Computation time $t_{comp}$ as measured for a MATLAB implementation on a 1.5-GHz computer.

| Patch size | No. of patches | $t_{\mathrm{comp}}$ (s) | $\hat{\lambda}_{t_N}$ | $\sqrt{\mathbf{P}_{\lambda,t_N}}$ | $E_{\mathrm{rms},v_1}$ | $E_{\mathrm{rms},v_2}$ |
|---|---|---|---|---|---|---|
| $1\times 1$ | 3600 | 1549 | 1.1173 | 0.3719 | 0.0139 | 0.0215 |
| $2\times 2$ | 900 | 270 | 1.0544 | 0.0948 | 0.0118 | 0.0247 |
| $3\times 3$ | 400 | 162 | 1.0424 | 0.0583 | 0.0114 | 0.0210 |
| $4\times 4$ | 225 | 179 | 1.0380 | 0.0371 | 0.0123 | 0.0224 |
| $5\times 5$ | 144 | 289 | 1.0336 | 0.0222 | 0.0145 | 0.0220 |
| $6\times 6$ | 100 | 593 | 1.0390 | 0.0146 | 0.0110 | 0.0155 |
| $10\times 10$ | 36 | 6636 | 1.0585 | 0.076 | 0.0107 | 0.0121 |
| $12\times 12$ | 25 | 12319 | 1.0531 | 0.0064 | 0.0107 | 0.0122 |
| $15\times 15$ | 16 | 26250 | 1.0546 | 0.0037 | 0.0112 | 0.0121 |
| $20\times 20$ | 9 | 68957 | 1.0558 | 0.0026 | 0.0133 | 0.0122 |

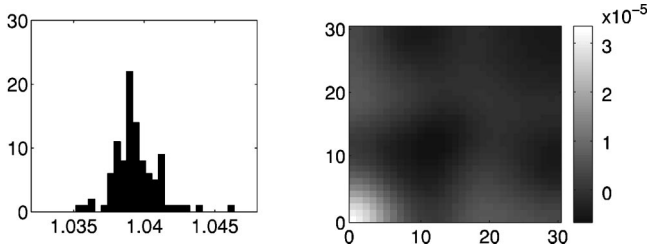


FIG. 5. Left: histogram of the parameter ensemble for the 6 $\times 6$ partition at $t=100$. The frequency distribution is approximately symmetric, justifying the use of the average in order to characterize the most likely parameter. Right: estimated correlation matrix (only the relevant part is shown). Axes denote spatial lags (in units of the spatial sampling step) along the respective spatial coordinates. Gray-scaled units are dimensionless.

## V. DISCUSSION

The method of partitioned filtering for estimating parameters and unobserved components from partial and noisy observations of spatiotemporal dynamical systems has been introduced. It has been exemplified on simulations of a reaction-diffusion system. Due to the method's recursive structure and the newly introduced partition technique, the method works efficiently for systems with several numbers of variables and spatial coordinates. A criterion for the choice of the partition is provided by means of the correlation length of the data, although there may exist systems, where the accuracy may also be good for patches smaller than those determined by the correlation length.

The example has been restricted to a system with a known structure and only one unknown parameter in order to prevent problems with structural identifiability issues, which may arise when increasing the number of unknown parameters. For example, even for linear state space models with known parameters, it may be possible that parts of the state components are not identifiable; in other words, it is impossible to estimate the complete state from partial observations. This is not a question of the amount of noise but mainly of the system's structure. For linear systems without spatial dynamics, criteria can be formulated that allow for a test for identifiability [39–42], but there is no concise theory up to now for nonlinear systems, except for some special classes, such as polynomials or compartment models of certain types [43–45]. To the knowledge of the authors, the identifiability problem for partial differential equations is yet unsolved.

The here-used particular approach of unscented Kalman filtering gives minimum mean squared error estimates for the state. When estimating the state and the parameter simultaneously, the resulting estimates for the parameter are optimal for the minimum mean squared error state, which may lead to some bias in the parameter estimates. Furthermore, the unscented Kalman filter can produce biased results for the states where strong nonlinearities are apparent. Only for polynomial nonlinearities up to the second order, it is guaranteed to yield optimal estimates.

It turns out in the example that the unscented Kalman filter gives excellent results for even higher-order nonlinearities. The reason is that the nonlinearities that arise from the temporal discretization of the dynamics of a time-continuous system are only weak, unless the time steps are too large; each numerical integration scheme is based on a decomposition of the dynamics into an identity mapping to which a comparably small correction is added. For large sampling time steps, where the nonlinearities become more significant, more accurate methods based on Monte Carlo approximations could be used in the here-introduced framework of partitioned filtering. This will be a topic of further research.

## ACKNOWLEDGMENTS

A.S. and J.K. acknowledge financial support from the Volkswagen-Stiftung. H.U.V. benefited from stimulating discussions with J. Timmer.

[1] P. Manneville, *Dissipative Structures and Weak Turbulence* (Academic Press, New York, 1990).
[2] M.C. Cross and P.C. Hohenberg, Rev. Mod. Phys. **65**, 851 (1993).
[3] G. Nicolis, *Introduction to Nonlinear Science* (Cambridge University Press, Cambridge, 1995).
[4] D. Walgraef, *Spatio-Temporal Pattern Formation* (Springer, New York, 1997).
[5] *Pattern Formation in the Physical and Biological Sciences,* Santa Fe Institute Studies in the Sciences of Complexity, edited by H. F. Nijhout, L. Nadel, and D. L. Stein, Lecture Notes, Vol. 5 (Addison-Wesley, Reading, MA, 1997).
[6] *Evolution of Spontaneous Structures in Dissipative Continuous Systems,* edited by F. H. Busse and S. C. Müller (Springer, Berlin, 1998).
[7] M. I. Rabinovich, A. B. Ezersky, and P. D. Weidman, *The Dynamics of Patterns* (World Scientific, Singapore, 2000).
[8] I.S. Aranson and L. Kramer, Rev. Mod. Phys. **74**, 99 (2002).
[9] R.D. De Veaux and J.M. Steele, Technometrics **31**, 91 (1989).
[10] H. Voss, M.J. Bünner, and M. Abel, Phys. Rev. E **57**, 2820 (1998).
[11] H.U. Voss, P. Kolodner, M. Abel, and J. Kurths, Phys. Rev. Lett. **83**, 3422 (1999).
[12] U. Parlitz and C. Merkwirth, Phys. Rev. Lett. **84**, 1890 (2000).
[13] S. Mandelj *et al.,* Int. J. Bifurcation Chaos Appl. Sci. Eng. **11**, 2731 (2001).
[14] K.H. Andersen *et al.,* Phys. Rev. Lett. **88**, 234302 (2002).
[15] D.P. Vallette, G. Jacobs, and J.P. Gollub, Phys. Rev. E **55**, 4274 (1997).
[16] P. Kolodner, S. Slimani, N. Aubry, and R. Lima, Physica D **85**, 165 (1995).
[17] R. Hegger, M. Bär, and H. Kantz, Phys. Rev. E **59**, 337 (1999).
[18] J.M. Fullana, P. Le Gal, M. Rossi, and S. Zaleski, Physica D **2**, 37 (1997).
[19] K. Schittkowski, Optimization Methods Software **7**, 165 (1997).

[20] S.C. van der Linde *et al.*, Appl. Catal. A **151**, 27 (1997).
[21] T.G. Müller and J. Timmer, Physica D **171**, 1 (2002).
[22] A. Sitz, U. Schwarz, J. Kurths, and H.U. Voss, Phys. Rev. E **66**, 016210 (2002).
[23] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, in Proceedings of the 1995 American Control Conference, Seattle, 1995 (unpublished), pp. 1628–1632.
[24] S. J. Julier and J. K. Uhlmann, in Proceedings of the International Symposium on Aerospace/Defense Sensing, Simulation and Controls, 1997 (unpublished).
[25] S. Julier and J. K. Uhlmann, in The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando, Florida, 2000 (unpublished).
[26] S. Julier, J. Uhlmann, and H.F. Durrant-Whyte, IEEE Trans. Autom. Control **45**, 477 (2000).
[27] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation* (Wiley, New York, 2001).
[28] R. Kalman, ASME J. Basic Eng. **82**, 35 (1960).
[29] R.E. Kalman and R.S. Bucy, ASME J. Basic Eng. **83**, 95 (1960).
[30] M. Nørgaard, N. K. Poulsen, and O. Ravn, Technical Report No. IMM-REP-1998-15, 2000 (unpublished).
[31] R. van der Merwe, in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 2001 (unpublished).
[32] S. J. Julier, Technical report, 1999 (unpublished).
[33] E. A. Wan, R. van der Merwe, and A. T. Nelson, in *Advances in Neural Information Processing Systems*, edited by S. Solla, T. Leen, and K.-R. Müller (MIT Press, Cambridge, MA, 2000), Vol. 12, pp. 666–672.
[34] M. Bär and M. Eiswirth, Phys. Rev. E **48**, R1635 (1993).
[35] M. Bär, M. Hildebrand, M. Eiswirth, M. Falcke, H. Engel, and M. Neufeld, Chaos **4**, 499 (1994).
[36] M. Hildebrand, M. Bär, and M. Eiswirth, Phys. Rev. Lett. **75**, 1503 (1995).
[37] M.D. Graham *et al.*, Phys. Rev. E **52**, 76 (1995).
[38] M. Bär *et al.*, Phys. Rev. Lett. **74**, 1246 (1995).
[39] A. Gelb, *Applied Optimal Estimation* (MIT Press, Cambridge, MA, 1974).
[40] R. Conti, *Linear Differential Equations and Control* (Academic Press, London, 1976).
[41] S. Barnett, *Polynomials and Linear Control Systems* (Marcel Dekker, New York, 1983).
[42] C. K. Chui and G. Chen, *Linear Systems and Optimal Control* (Springer, Berlin, 1989).
[43] S. Vajda, K.R. Godfrey, and H. Rabitz, Math. Biosci. **93**, 217 (1989).
[44] H. Pohjanpalo, Math. Biosci. **41**, 21 (1978).
[45] J.F. Ritt, *Differential Equantions from the Algebraic Standpoint of View* (AMS, New York, 1932).