Estimation of noise parameters in dynamical system identification with Kalman filters

Frank Kwasniok
College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom
(Received 5 April 2012; published 26 September 2012)

A method is proposed for determining dynamical and observational noise parameters in state and parameter identification from time series using Kalman filters. The noise covariances are estimated in a secondary optimization by maximizing the predictive likelihood of the data. The approach is based on internal consistency; for the correct noise parameters, the uncertainty projected by the Kalman filter matches the actual predictive uncertainty. The method is able to disentangle dynamical and observational noise. The algorithm is demonstrated for the linear, extended, and unscented Kalman filters using an Ornstein-Uhlenbeck process, the noise-driven Lorenz system, and van der Pol oscillator as well as a paleoclimatic ice-core record as examples. The approach is also applicable to the ensemble Kalman filter and can be readily extended to non-Gaussian estimation frameworks such as Gaussian-sum filters and particle filters.

DOI: 10.1103/PhysRevE.86.036214 PACS number(s): 05.45.Tp

I. INTRODUCTION

Identification of system states and parameters from data is an ubiquitous task in many areas of science. Often a dynamical model with unknown parameters and time series of noisy observations are available; the problem of only partial observation of the system is common. Kalman filters [1,2] are a standard tool for state and parameter estimation in many scientific disciplines, for their easy implementation due to the recursive nature of the algorithm and their applicability even in high-dimensional state spaces [3,4].

The use of Kalman filters involves the specification of dynamical and observational noise covariances. State and parameter estimates are inherently dependent on both the dynamical and observational noise covariances. Even the linear Kalman filter is suboptimal when insufficient information on the noise statistics is available. Moreover, when identifying stochastic dynamical systems from data estimation of the dynamical noise level might be as important and interesting as estimation of the parameters of the deterministic part of the system [5].

The noise parameters in Kalman filtering are often assumed to be known *a priori* [6] or determined in an ad hoc manner by fitting some quantity of interest of the system [5]. Algorithms for estimating noise covariances based on lag-covariances of the innovations have been proposed in the engineering community [7–9]. Such approaches are fundamentally limited to Gaussian estimation.

The present paper discusses an approach based on predictive likelihood (cf. [10,11]) for estimating noise parameters in Kalman filtering, focusing on nonlinear systems, and nonlinear Kalman filters. All parameters are optimized simultaneously: The system states and parameters are estimated continuously directly in the Kalman filter using the augmented state approach; the noise parameters are found by maximizing an appropriate likelihood function. The method is able to disentangle dynamical and observational noise. The algorithm is explored using simulated data from known mathematical systems as well as a paleoclimatic ice-core record. The linear, the extended, and the unscented Kalman filters are considered. A basic advantage of the likelihood approach is that it can be transferred to non-Gaussian estimation frameworks.

II. STATE AND PARAMETER ESTIMATION WITH KALMAN FILTERS

The framework of a continuous-discrete nonlinear state space model and the corresponding versions of the Kalman filter is adopted here. The evolution of a state vector $\mathbf{z}$ of dimension n is assumed to be governed by a continuous-time, nonlinear stochastic dynamical system

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z};\boldsymbol{\lambda}) + \boldsymbol{\xi}, \tag{1}$$

where $\boldsymbol{\lambda}$ is a constant parameter vector of dimension p and $\boldsymbol{\xi}$ is a vector of Gaussian white noises with zero mean and covariance matrix $\mathbf{Q}$. An augmented state vector $\mathbf{x}$ of dimension $n_a = n + p$ is formed by merging the state vector $\mathbf{z}$ and the parameter vector $\boldsymbol{\lambda}$. Equation (1) together with a constant dynamics for the parameters

$$\dot{\boldsymbol{\lambda}} = 0 \tag{2}$$

form the dynamical or state equation of the state space model:

$$\dot{\mathbf{x}} = \mathbf{f}^a(\mathbf{x}) + \boldsymbol{\xi}^a. \tag{3}$$

$\boldsymbol{\xi}^a$ is the augmented noise process vector with corresponding covariance matrix $\mathbf{Q}^a$. At discrete times t_k, observations $\mathbf{y}_k$ of dimension m are available, which are linked to the state vector by the observation or measurement equation

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k, \tag{4}$$

where $\boldsymbol{\varepsilon}_k$ is a vector of white Gaussian observational noises with zero mean and covariance matrix $\mathbf{R}$. The setting includes the special case of state estimation only where the system parameters are assumed to be known, corresponding to an empty parameter vector with $p = 0$ and $\mathbf{x} = \mathbf{z}$.

Kalman filters are an algorithm for estimating the system states and parameters given only a time series of noisy observations $\{\mathbf{y}_k\}_{k=0}^N$. Let $\hat{\mathbf{x}}_{k-1|k-1}$ be the mean estimate of the augmented state vector and $\mathbf{P}_{k-1|k-1}$ its covariance matrix at time step $k-1$ having processed all data up to time step $k-1$. Means and covariances are propagated from time t_{k-1} to t_k to yield estimates $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$. Then the estimates of the states and the parameters as well as their uncertainties are updated using the new observation according to the Kalman

update equations

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \boldsymbol{\zeta}_k, \tag{5}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_{k|k-1}. \tag{6}$$

Here,

$$\boldsymbol{\zeta}_k = \mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) \tag{7}$$

is the vector of innovations or residuals and

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{S}_k^{-1} \tag{8}$$

is the Kalman gain matrix. $\mathbf{H}_k$ denotes the linearized observation function

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}}, \tag{9}$$

and

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R} \tag{10}$$

is the (predicted) residual covariance matrix.

Different variants of the Kalman filter differ in how the means and covariances are propagated between times t_{k-1} and t_k. The linear Kalman filter is applicable if the augmented state dynamics are linear, that is, $\dot{\mathbf{x}} = \mathbf{F}\mathbf{x} + \boldsymbol{\xi}^a$ with a (possibly time-dependent) matrix $\mathbf{F}$. This is the case if the evolution of $\mathbf{z}$ is linear and we have state estimation only or the parameters enter only additively. Means and covariances are then propagated by solving the differential equations

$$\dot{\hat{\mathbf{x}}} = \mathbf{F}\hat{\mathbf{x}}, \tag{11}$$

$$\dot{\mathbf{P}} = \mathbf{F}\mathbf{P} + \mathbf{P}\mathbf{F}^T + \mathbf{Q}^a \tag{12}$$

on the time interval $[t_{k-1}, t_k]$ with initial conditions $\hat{\mathbf{x}}(t_{k-1}) = \hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}(t_{k-1}) = \mathbf{P}_{k-1|k-1}$, and then setting $\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}(t_k)$ and $\mathbf{P}_{k|k-1} = \mathbf{P}(t_k)$.

The extended Kalman filter linearizes a nonlinear dynamics in augmented state space about the current state. Means and covariances are evolved on the time interval $[t_{k-1}, t_k]$ according to the equations

$$\dot{\hat{\mathbf{x}}} = \mathbf{f}^a(\hat{\mathbf{x}}), \tag{13}$$

$$\dot{\mathbf{P}} = \mathbf{F}\mathbf{P} + \mathbf{P}\mathbf{F}^T + \mathbf{Q}^a, \tag{14}$$

with initial conditions $\hat{\mathbf{x}}(t_{k-1}) = \hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}(t_{k-1}) = \mathbf{P}_{k-1|k-1}$. We then set $\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}(t_k)$ and $\mathbf{P}_{k|k-1} = \mathbf{P}(t_k)$. $\mathbf{F}$ now denotes the Jacobian of $\mathbf{f}^a$:

$$\mathbf{F} = \left.\frac{\partial \mathbf{f}^a}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}}. \tag{15}$$

The unscented Kalman filter [12,13] keeps the full nonlinear system dynamics rather than linearizing it. The filter density is represented by a small number of so-called sigma points that are propagated through the full nonlinear dynamical equations. The interval $[t_{k-1}, t_k]$ is divided into L equal subintervals of size $h = (t_k - t_{k-1})/L$ and a sequence of estimates $\{\hat{\mathbf{x}}_l, \mathbf{P}_l\}_{l=0}^{L}$ is generated. Having arrived at $\hat{\mathbf{x}}_{l-1}$ and $\mathbf{P}_{l-1}$, we use $2n_a$ sigma points, $\{\mathbf{x}^i_{l-1|l-1}\}_{i=1}^{2n_a}$, each in augmented state space of dimension n_a, given as $\{\hat{\mathbf{x}}_{l-1} - \mathbf{w}^j_{l-1}, \hat{\mathbf{x}}_{l-1} + \mathbf{w}^j_{l-1}\}_{j=1}^{n_a}$. The vectors $\{\mathbf{w}^j_{l-1}\}_{j=1}^{n_a}$ are the columns of $\mathbf{A}$ where $\mathbf{A}$ can be any matrix satisfying $\mathbf{A}\mathbf{A}^T = n_a\mathbf{P}_{l-1}$. Here, we calculate $\mathbf{A}$ using the Cholesky decomposition of $\mathbf{P}_{l-1}$. The sigma points are transformed as

$$\mathbf{x}^i_{l|l-1} = \mathbf{x}^i_{l-1|l-1} + h\mathbf{f}^a\left(\mathbf{x}^i_{l-1|l-1}\right), \quad i = 1, \ldots, 2n_a \tag{16}$$

and new mean and covariance estimates are given by

$$\hat{\mathbf{x}}_l = \frac{1}{2n_a}\sum_{i=1}^{2n_a} \mathbf{x}^i_{l|l-1} \tag{17}$$

and

$$\mathbf{P}_l = \frac{1}{2n_a}\sum_{i=1}^{2n_a} \left(\mathbf{x}^i_{l|l-1} - \hat{\mathbf{x}}_l\right)\left(\mathbf{x}^i_{l|l-1} - \hat{\mathbf{x}}_l\right)^T + h\mathbf{Q}^a. \tag{18}$$

The sequence is initialized with $\hat{\mathbf{x}}_0 = \hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_0 = \mathbf{P}_{k-1|k-1}$. We set $\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}_L$ and $\mathbf{P}_{k|k-1} = \mathbf{P}_L$. The unscented Kalman filter has the capability of taking full account of a nonlinear observation function rather than linearizing it by propagating the sigma points also through the observation function and then using an alternative formulation of the Kalman update equations. But we drop this complication as the observation variables here are just linear projections of the state.

III. ESTIMATION OF NOISE PARAMETERS

We propose to estimate the noise parameters according to the maximum likelihood principle. At time step k, the predictive probability density of the Kalman filter for the residual $\boldsymbol{\zeta}_k$ is a Gaussian with zero mean and covariance matrix $\mathbf{S}_k$. Thus, the predictive ln-likelihood function of the data set is

$$\ln \Lambda(\mathbf{Q}, \mathbf{R}) = -\frac{Nm}{2}\ln 2\pi - \frac{1}{2}\sum_{k=1}^{N}\left(\ln|\mathbf{S}_k| + \boldsymbol{\zeta}_k^T\mathbf{S}_k^{-1}\boldsymbol{\zeta}_k\right), \tag{19}$$

which is inversely proportional to the ignorance score [14] of the predictions. The Kalman filter is run with different sets of noise parameters and the likelihood maximized (the predictive ignorance minimized) with respect to the noise parameters $\mathbf{Q}$ and $\mathbf{R}$. We here focus on the case of only few noise parameters. No advanced optimization procedure is used; the ln likelihood is just calculated on a fine enough mesh in noise parameter space and the maximum found.

The present approach for identifying the noise parameters is based on internal consistency. For the correct noise parameters, the predictive uncertainty estimated by the Kalman filter matches the true predictive uncertainty of the underlying system reflected in the data. The sharpness or information content of the predictions is then neither over- nor underconfident but in line with the actual predictability of the system. This holds true at least in the perfect model scenario, that is, if there is a true model having generated the data which is contained in the model class prescribed by the system and noise parameters.

For real-world data the situation is less clear, as there is usually no perfect model known and any assumed model class has some structural model error. But the likelihood approach can be expected to be still useful as it picks the best predictive model within a given model class. Moreover, the likelihood

function (or some information criterion based on it) then provides a tool for model selection between different model classes.

IV. ORNSTEIN-UHLENBECK PROCESS

We start exemplifying the method with the Ornstein-Uhlenbeck (O-U) process:

$$\dot{z} = -\gamma z + \sigma \eta. \tag{20}$$

η denotes Gaussian white noise with zero mean and unit variance. The parameter setting $\gamma = 1$ and $\sigma = 1$ is chosen. A time series of length $N = 5000$ with sampling interval $\delta t = 0.1$ is generated by numerical integration. Here and in the following the Euler-Maruyama scheme [15] with step size 10^{-5} is used for the numerical integration of stochastic systems. Observation of the state z is corrupted by Gaussian white noise of zero mean and standard deviation $\tau = 0.25$. Figure 1 displays a piece of the system trajectory and the corresponding observations. The dimensions are $n = 1$, $p = 1$, $n_a = 2$, and $m = 1$. State estimation only is done with the linear Kalman filter (LKF); combined state and parameter estimation is performed with the extended Kalman filter (EKF). The filter equations [Eqs. (11) and (12) or (13) and (14), respectively] are integrated using the Euler scheme with step size $\delta t/100 = 0.001$.

Figure 2 shows the ln likelihood as a function of σ for both state estimation only and combined state and parameter estimation. The observational noise level is fixed at the true value $\tau = 0.25$. On a mesh of size 0.005 the maximum of the likelihood is in both cases located at $\sigma = 0.995$, very close to the true value. The estimate for γ is displayed in Fig. 3 together with the standard deviation of the estimation error. The obtained parameter value depends markedly on the assumed value of σ. At the identified noise level $\sigma = 0.995$ the parameter estimate is $\gamma = 0.993 \pm 0.032$; at the true noise level $\sigma = 1$ it is $\gamma = 0.996 \pm 0.032$. Figure 4 illustrates the ln likelihood as a function of both dynamical and observational noise levels for state estimation only. On a mesh of size 0.005 for σ and 0.0025 for τ the maximum is at $\sigma = 0.995$ and $\tau = 0.25$. Both noise levels are accurately identified. For combined state and parameter estimation the ln likelihood looks very similar (not shown). The maximum is at the same

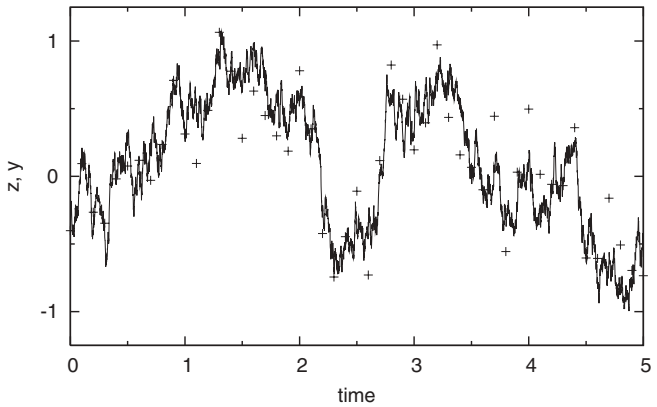

FIG. 1. Ornstein-Uhlenbeck process: System trajectory (solid) and observations (+).

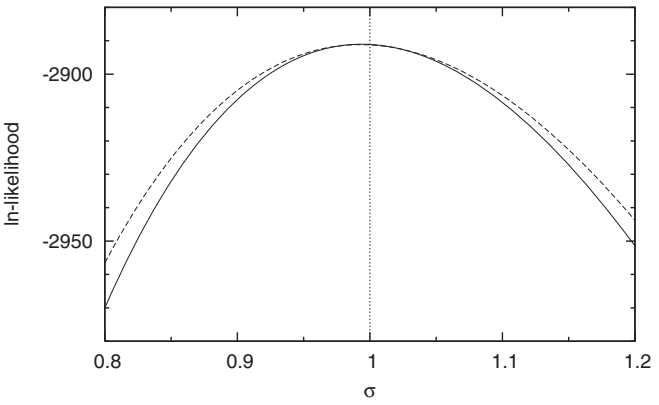

FIG. 2. Ornstein-Uhlenbeck process: ln likelihood as a function of dynamical noise level σ for state estimation only (solid) as well as combined state and parameter estimation (dashed). The true value $\sigma = 1$ is indicated by the dotted vertical line. Observational noise level is $\tau = 0.25$.

position. It is slightly less sharp, thus the identification problem is slightly more ill-conditioned. This is also visible in the cross-section at $\tau = 0.25$ shown in Fig. 2.

V. NOISE-DRIVEN LORENZ SYSTEM

Next, we consider the classical Lorenz system [16] augmented with stochastic noise [17]. The governing equations are:

$$\dot{z}_1 = -s z_1 + s z_2 + \sigma \eta_1, \tag{21}$$

$$\dot{z}_2 = -z_1 z_3 + r z_1 - z_2 + \sigma \eta_2, \tag{22}$$

$$\dot{z}_3 = z_1 z_2 - b z_3 + \sigma \eta_3. \tag{23}$$

The parameter setting $s = 10$, $r = 28$, and $b = 8/3$ is used for which the deterministic system exhibits chaotic dynamics. η_1, η_2, and η_3 are pairwise independent Gaussian white noises with zero mean and unit variance. The dynamical noise level is $\sigma = 1$. A post-transient data set of length $N = 5000$ with sampling interval $\delta t = 0.05$ is archived. Only the variable

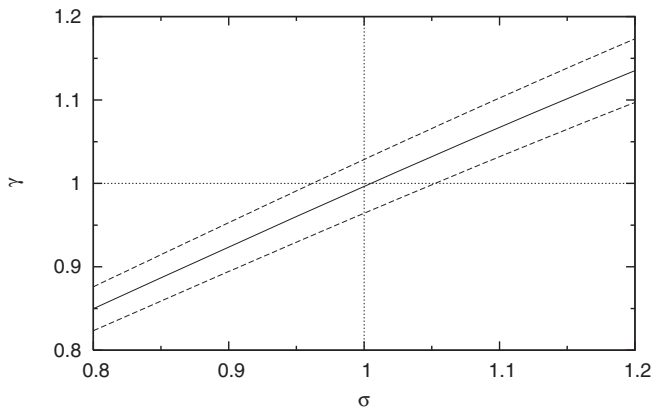

FIG. 3. Ornstein-Uhlenbeck process: Estimate of the parameter γ (solid) with error standard deviation (dashed) as a function of dynamical noise level σ. The dotted horizontal and vertical lines indicate the true values $\gamma = 1$ and $\sigma = 1$. Observational noise level is $\tau = 0.25$.

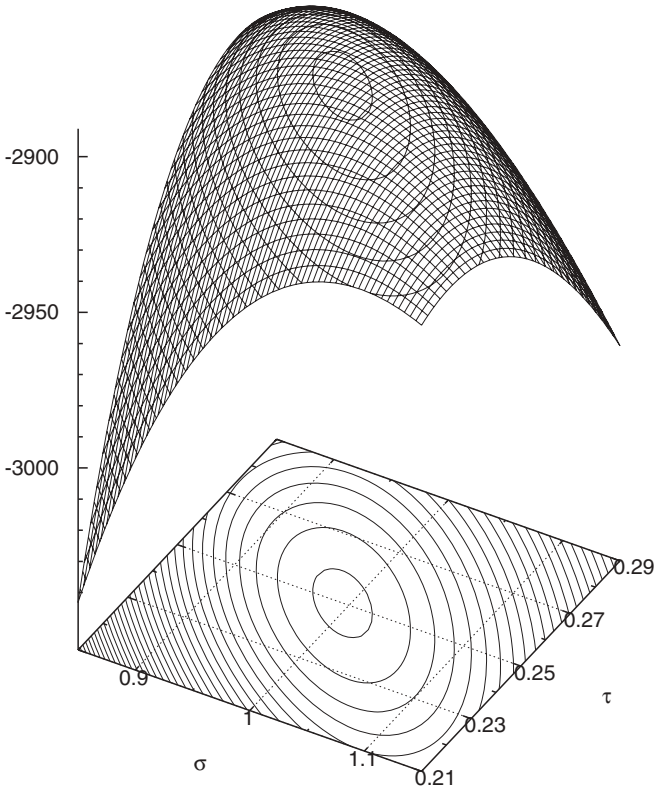

FIG. 4. Ornstein-Uhlenbeck process: ln likelihood as a function of dynamical and observational noise levels σ and τ for state estimation only. The true values are $\sigma = 1$ and $\tau = 0.25$. Neighboring contours differ by a factor 100 in likelihood.

z_1 is observed subject to Gaussian white noise of zero mean and standard deviation $\tau = 0.5$. Figure 5 shows the trajectory and the observations. Here, we have $n = 3$, $p = 3$, $n_a = 6$, and $m = 1$. State estimation only as well as combined state and parameter estimation are performed with the extended Kalman filter. Equations (13) and (14) are solved numerically using the Euler scheme with step size $\delta t/100 = 0.0005$.

First, the observational noise level is assumed to be known and the dynamical noise level estimated. The ln-likelihood function is displayed in Fig. 6. A mesh of size 0.01 is used.

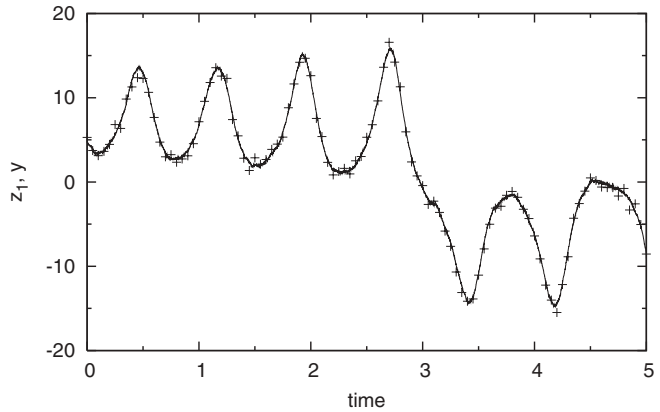

FIG. 5. Noise-driven Lorenz system: Trajectory of z_1 (solid) and observations (+).

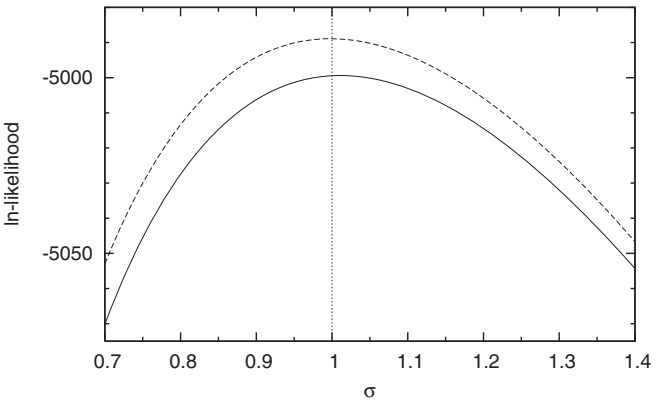

FIG. 6. Noise-driven Lorenz system: ln likelihood as a function of dynamical noise level σ for state estimation only (solid) as well as combined state and parameter estimation (dashed). The true value $\sigma = 1$ is indicated by the dotted vertical line. Observational noise level is $\tau = 0.5$.

For state estimation only the maximum of the likelihood is at $\sigma = 1.01$, an error of only 1%. For combined state and parameter estimation the maximum is at the correct value $\sigma = 1$; the parameter estimates and their error standard deviations then are $s = 9.81 \pm 0.03$, $r = 27.78 \pm 0.07$, and $b = 2.68 \pm 0.01$. The value of the likelihood is somewhat higher than with state estimation only; this is a combined effect of the finiteness of the data sample as well as the errors of the linear and Gaussian approximations in the extended Kalman filter. Figure 7 shows the ln likelihood as a function of both dynamical and observational noise levels for combined state and parameter estimation. On a mesh of size 0.01 for σ and 0.0025 for τ, the maximum is at $\sigma = 0.99$ and $\tau = 0.5025$. The parameter estimates then are the same as quoted before. For state estimation only the likelihood looks very similar (not shown); the maximum is at $\sigma = 1.01$ and $\tau = 0.5025$.

The results reported here for the Lorenz system are stronger than those in Ref. [11]. The dynamical noise level is considerable, giving the system a truly stochastic character, and all the system and noise parameters are estimated simultaneously, whereas in Ref. [11] only one-dimensional cross-sections of the likelihood are considered.

The unscented Kalman filter here yields virtually the same estimates for the noise levels with slightly more accurate parameter estimates; however, the extended Kalman filter performs very well and would be considered sufficient. For higher noise levels the advantage of the unscented Kalman filter over the extended Kalman filter due to superior covariance propagation becomes a bit more prominent. For example, with $\sigma = 2$ and $\tau = 1$ the extended Kalman filter has an error of about 5% in the estimate of σ, whereas the unscented Kalman filter still identifies σ almost perfectly.

VI. NOISE-DRIVEN VAN DER POL OSCILLATOR

Finally, the noise-driven van der Pol oscillator is considered. The equations of motion are:

$$\dot{z}_1 = z_2, \tag{24}$$

$$\dot{z}_2 = \mu\left(1 - z_1^2\right)z_2 - z_1 + \sigma\eta. \tag{25}$$

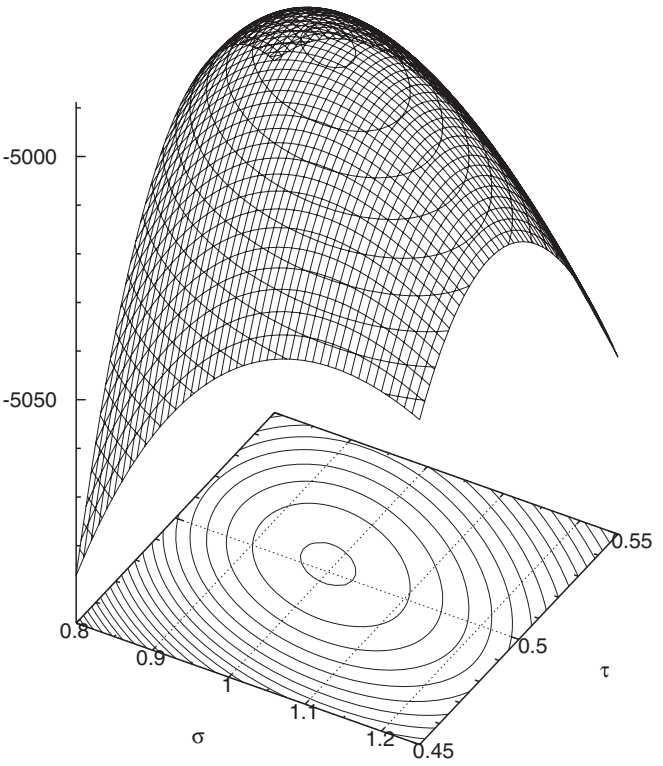

FIG. 7. Noise-driven Lorenz system: ln likelihood as a function of dynamical and observational noise levels σ and τ for combined state and parameter estimation. The true values are $\sigma = 1$ and $\tau = 0.5$. Neighboring contours differ by a factor 100 in likelihood.

The parameter setting $\mu = 3$ is adopted. η is a white Gaussian noise with zero mean and unit variance. The dynamical noise level is $\sigma = 0.5$. The motion of the system evolves on a noisy limit cycle. A time series of length $N = 5000$ with sampling interval $\delta t = 0.1$ is generated. Only the variable z_1 is observed with Gaussian white measurement noise of zero mean and standard deviation $\tau = 0.15$. Figure 8 displays a piece of the trajectory together with the observations. The dimensions here are $n = 2$, $p = 1$, $n_a = 3$, and $m = 1$. The data are processed with the unscented Kalman filter (UKF) as the dynamics in the augmented state space are quite strongly nonlinear. The

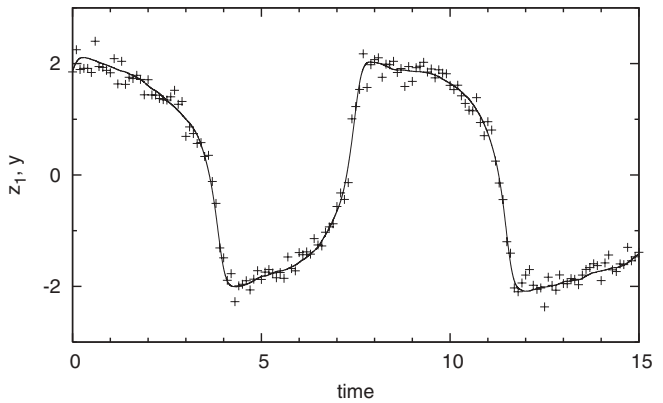

FIG. 8. Noise-driven van der Pol oscillator: Trajectory of z_1 (solid) and observations (+).

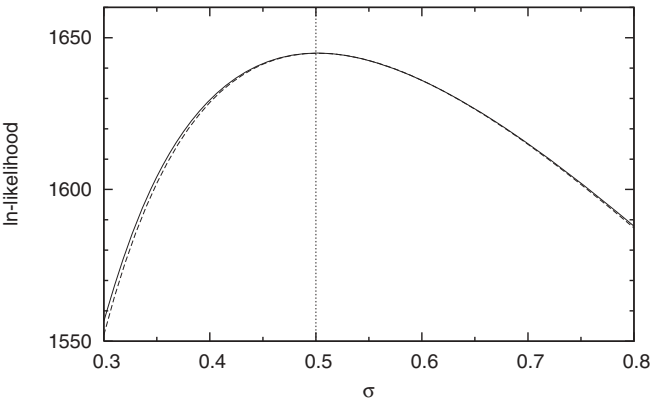

FIG. 9. Noise-driven van der Pol oscillator: ln likelihood as a function of dynamical noise level σ for state estimation only (solid) as well as state and parameter estimation (dashed). The true value $\sigma = 0.5$ is indicated by the dotted vertical line. Observational noise level is $\tau = 0.15$.

number of subintervals for the propagation of means and covariances is chosen as $L = 100$.

Again, we first fix the observational noise level at the true value and estimate only the dynamical noise level. Figure 9 shows the ln-likelihood function. Both for state estimation only and combined state and parameter estimation there is a maximum at the correct value $\sigma = 0.5$, using a mesh of size 0.01. The parameter estimate then is $\mu = 2.98 \pm 0.02$. Figure 10 displays the ln likelihood as a function of both σ

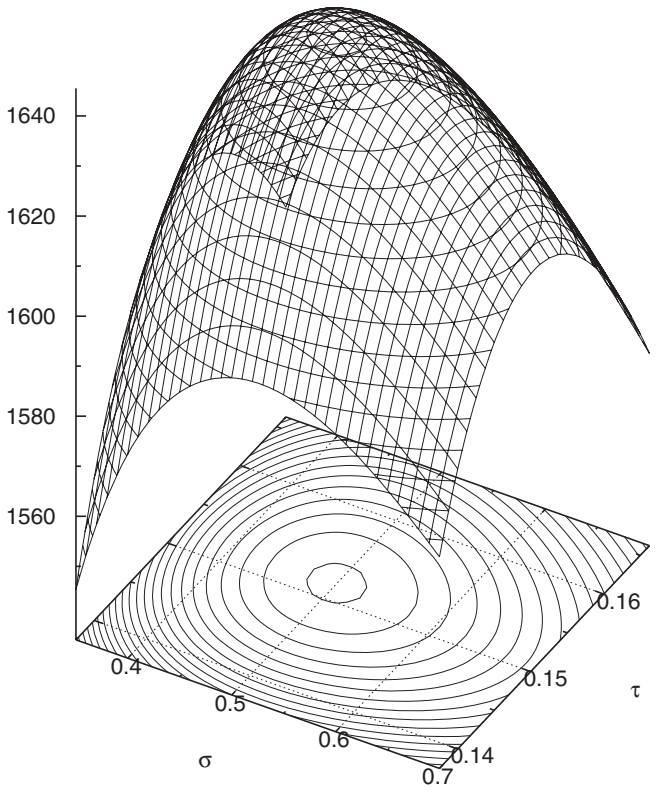

FIG. 10. Noise-driven van der Pol oscillator: ln likelihood as a function of dynamical and observational noise levels σ and τ for combined state and parameter estimation. The true values are $\sigma = 0.5$ and $\tau = 0.15$. Neighboring contours differ by a factor of 100 in likelihood.

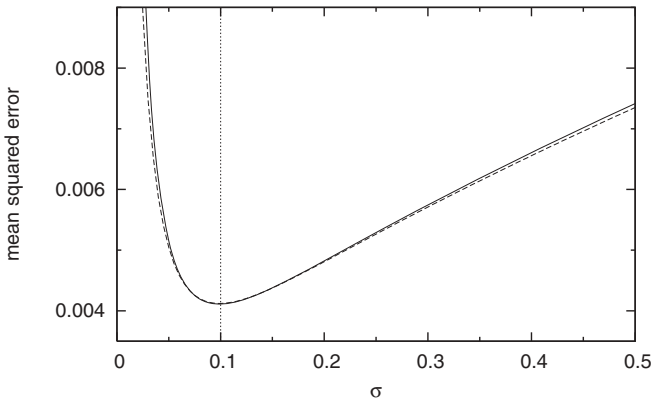

FIG. 11. Noise-driven van der Pol oscillator: Mean-squared error in the estimate of the state z_1 as a function of the dynamical noise level σ for state estimation only (solid) as well as combined state and parameter estimation (dashed). The true value $\sigma = 0.1$ is indicated by the dotted vertical line. Observational noise level is $\tau = 0.25$.

and τ for combined state and parameter estimation. There is a maximum at $\sigma = 0.5$ and $\tau = 0.151\,25$, using a mesh of size 0.01 for σ and $0.001\,25$ for τ. The parameter estimate then is $\mu = 2.99 \pm 0.02$. For state estimation only, the maximum of the likelihood is at $\sigma = 0.5$ and $\tau = 0.1525$ (not shown).

State estimation is inherently linked to parameter and noise-level estimation and can be expected to depend on model parameters and noise levels. This will be most apparent when the model is relatively accurate (small dynamical noise level) and observational uncertainty is large. We, therefore, adopt the setting $\sigma = 0.1$ and $\tau = 0.25$. Figure 11 displays the mean squared difference between the estimate of z_1 and its true value as a function of σ for fixed observational noise level. Both for state estimation only and combined state and parameter estimation the state estimate is optimal for the correct value $\sigma = 0.1$ (on a mesh of size 0.005). The likelihood has a maximum at $\sigma = 0.105$ in both cases (not shown). The parameter estimate both at $\sigma = 0.1$ and $\sigma = 0.105$ is $\mu = 3.00 \pm 0.01$.

A summary of the estimation results for the various systems is given in Table I.

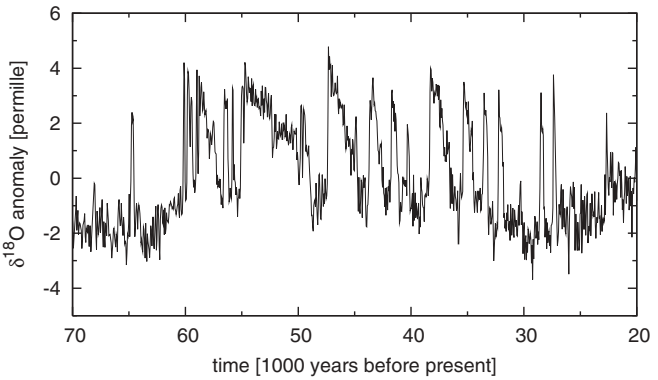

FIG. 12. δ^{18}O record from the NGRIP ice core during the last glacial period with the mean value removed.

VII. ICE-CORE RECORD

We now consider a real-world data set where no true model is available. The record of δ^{18}O as a proxy for northern hemisphere temperatures from the North Greenland Ice Core Project (NGRIP) ice core [18] is studied. Figure 12 shows the record for the last glacial period, ranging from 70 to 20 ky before present (1 ky = 1000 years); the mean value over that period is removed as the dynamical model is formulated as an anomaly model. The sampling interval is $\delta t = 0.05$ ky resulting in 1000 data points. The time series displays switches between the cold stadial and the warm interstadial state.

Brownian motion in a potential landscape [5,19,20] is adopted as a simple dynamical model:

$$\dot{z} = -U'(z) + \sigma\eta. \tag{26}$$

η denotes a white Gaussian noise with zero mean and unit variance. The potential is assumed to have the polynomial form [5,20]

$$U(z) = \sum_{i=1}^{4} a_i z^i, \tag{27}$$

where the coefficients a_i are the parameters to be determined. Noisy observations are available (identified with the isotope record) with Gaussian white observational noise of zero mean and standard deviation τ. Here, we have $n = 1$, $p = 4$, $n_a = 5$,

TABLE I. Summary of the parameter estimation results for the different systems.

System	True parameters	Estim. quantities	Algorithm	Estimated parameter values
O-U process	$\gamma = 1$,	σ	LKF	$\sigma = 0.995$
	$\sigma = 1, \tau = 0.25$	γ, σ	EKF	$\gamma = 0.993, \sigma = 0.995$
		σ, τ	LKF	$\sigma = 0.995, \tau = 0.25$
		γ, σ, τ	EKF	$\gamma = 0.993, \sigma = 0.995, \tau = 0.25$
Lorenz	$s = 10, r = 28, b = 8/3,$	σ	EKF	$\sigma = 1.01$
	$\sigma = 1, \tau = 0.5$	s, r, b, σ	EKF	$s = 9.81, r = 27.78, b = 2.68, \sigma = 1$
		σ, τ	EKF	$\sigma = 1.01, \tau = 0.5025$
		s, r, b, σ, τ	EKF	$s = 9.81, r = 27.78, b = 2.68, \sigma = 0.99, \tau = 0.5025$
Van der Pol	$\mu = 3$,	σ	UKF	$\sigma = 0.5$
	$\sigma = 0.5, \tau = 0.15$	μ, σ	UKF	$\mu = 2.98, \sigma = 0.5$
		σ, τ	UKF	$\sigma = 0.5, \tau = 0.1525$
		μ, σ, τ	UKF	$\mu = 2.99, \sigma = 0.5, \tau = 0.151\,25$

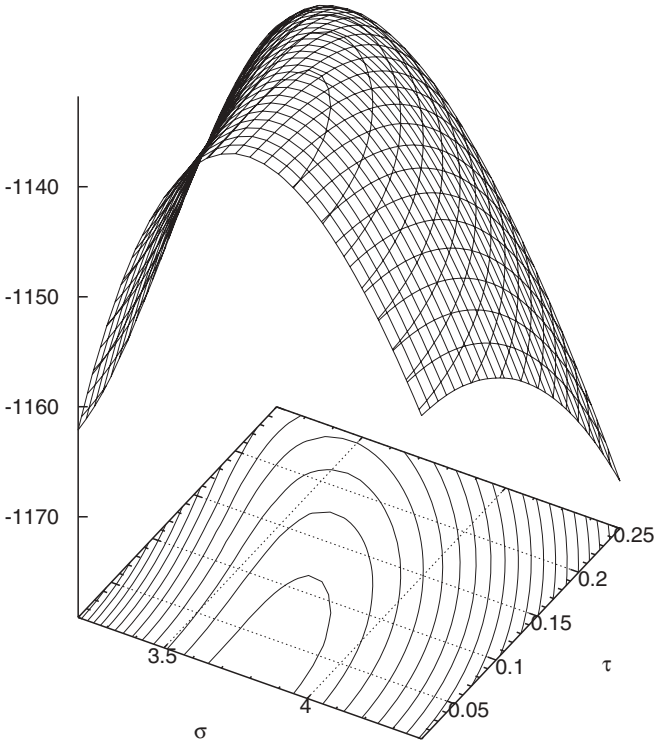

FIG. 13. Ice-core record: ln likelihood as a function of dynamical and observational noise levels σ and τ for combined state and parameter estimation. Neighboring contours differ by a factor of 10 in likelihood.

and $m = 1$. Exactly the same time series under the same dynamical model was processed with the unscented Kalman filter in Ref. [5]. The observational noise level was set to zero based on the assumption that the measurement error is small and negligible compared to the high dynamical noise level. The dynamical noise level was then found by running the Kalman filter for different values, integrating the resulting model in time and searching for a match in certain characteristic statistical quantities of the system. Matching the variance of the system yields $\sigma = 3.8$; matching the autocorrelation at lag δt gives $\sigma = 3.7$. A compromise value of $\sigma = 3.75$ was then adopted.

Figure 13 shows the ln likelihood as a function of both dynamical and observational noise levels. On a mesh of size 0.05 for σ and 0.01 for τ the maximum is at $\sigma = 3.8$ and $\tau = 0.01$. The observational noise level is effectively arbitrarily close to zero. Technically, it must not be set exactly to zero as then the propagated filter covariances may become not positive definite due to rounding errors, which causes the Cholesky decomposition in the UKF to break down. The likelihood is quite flat in the τ direction. But, this is not really a problem here, as even with a value of, say, $\tau = 0.15$ the estimation results from the Kalman filter are virtually the same as with, say, $\tau = 0.0001$ because σ is large and still dominating the uncertainty. The present results are in close agreement with Ref. [5] regarding the value of σ and the fact that τ is virtually zero. The estimated potential coefficients here are $a_4 = 0.16$, $a_3 = -0.37$, $a_2 = -0.85$, and $a_1 = 2.38$, which is very close to those reported in Ref. [5] with $\sigma = 3.75$.

VIII. DISCUSSION

The method can obviously also be used when the dynamical equation is given as a discrete map rather than a continuous-time model. Equation (2) is then replaced with $\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_{k-1}$ and the discrete versions of the Kalman filters are employed. Technically, the calculations are then even easier as the propagation of means and covariances from t_{k-1} to t_k can be done in one single step rather than approximating a continuous propagation with multiple steps.

The present likelihood approach for determining noise parameters is also applicable to the ensemble Kalman filter [3], which is widely used in the weather and climate science community. Moreover, it can be readily extended to non-Gaussian state and parameter estimation frameworks such as Gaussian-sum filters or particle filters. The Gaussian likelihood in Eq. (19) then needs to be replaced with an appropriate non-Gaussian likelihood.

In cases where an imperfect deterministic model is available stochastic terms could be introduced to improve the model and estimated with the present method. This is in line with the idea of stochastic parametrization of model error and uncertainty [21].

The method clearly is only feasible for few noise parameters. In spatially extended systems the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ are large. But typically the entries are not independent and often they can be written as linear combinations of few fixed matrices ($\mathbf{Q} = \sum_i \alpha_i \mathbf{Q}_i$ and $\mathbf{R} = \sum_j \beta_j \mathbf{R}_j$) (cf. [8,9]). Then the coefficients α_i and β_j would be determined via the maximum likelihood principle.

[1] R. Kalman, Trans. ASME J. Basic. Eng. Series **82/D**, 35 (1960).
[2] R. E. Kalman and R. S. Bucy, Trans. ASME J. Basic. Eng. Series **83/D**, 95 (1961).
[3] G. Evensen, J. Geophys. Res. **99**, 10143 (1994).
[4] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, 2002).
[5] F. Kwasniok and G. Lohmann, Phys. Rev. E **80**, 066104 (2009).
[6] H. U. Voss, J. Timmer, and J. Kurths, Int. J. Bif. Chaos **14**, 1905 (2004).
[7] R. K. Mehra, IEEE Trans. Autom. Control **15**, 175 (1970).
[8] P. R. Bélanger, Automatica **10**, 267 (1974).
[9] D. P. Dee, S. E. Cohn, A. Dalcher, and M. Ghil, IEEE Trans. Automatic Control **30**, 1057 (1985).
[10] K. J. Åström, Automatica **16**, 551 (1980).
[11] T. Ozaki, J. C. Jimenez, and V. Haggan-Ozaki, J. Time Ser. Anal. **21**, 363 (2000).
[12] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, IEEE Trans. Autom. Control **45**, 477 (2000).
[13] S. J. Julier and J. K. Uhlmann, Proc. IEEE **92**, 401 (2004).

[14] I. J. Good, J. R. Stat. Soc. B **14**, 107 (1952).
[15] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 1999).
[16] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[17] A. Sutera, J. Atmos. Sci. **37**, 246 (1980).
[18] K. K. Andersen *et al*., Nature (London) **431**, 147 (2004).
[19] P. D. Ditlevsen, Geophys. Res. Lett. **26**, 1441 (1999).
[20] V. N. Livina, F. Kwasniok, and T. M. Lenton, Climate of the Past **6**, 77 (2010).
[21] T. N. Palmer, Q. J. R. Meteorol. Soc. **127**, 279 (2001).