# Estimating Parameters in Discretely, Partially Observed Stochastic Differential Equations

Jan Nygaard Nielsen,[*] Henrik Madsen,[†] and Henrik Melgaard[‡]

May 10, 2000

**Abstract**

An approximate maximum likelihood method for direct estimation of embedded parameters in nonlinear, multivariate stochastic differential equations using discrete-time input-output data encumbered with additive measurement noise is proposed. The stochastic differential equations act as the system equation of a continuous-discrete time state space model which is introduced to describe nonlinear, multivariate and quasi-stationary systems. The likelihood is formulated as a function of the embedded parameters of the stochastic state space model, and an Iterated Extended Kalman filter is used in evaluating the likelihood function. A transformation is introduced to remove level effects (state-dependent diffusion terms) in some multivariate SDEs such that the filtering problem may be solved using the IEKF. Monte Carlo simulation of a nonlinear predator-prey system is used to study the statistical properties of the proposed method.

KEY WORDS: Stochastic modelling; continuous time systems; Brownian motion; random processes; Extended Kalman filters; Maximum likelihood estimation.

[*]E-mail: jnn@imm.dtu.dk
[†]E-mail: hm@imm.dtu.dk
[‡]E-mail: hmel@novo.dk



# Contents

# 1 Introduction

Models of physical, chemical and biological systems derived from first principles are inherently continuous in time, which implies that continuous-time (CT) models support a better understanding of the actual behavior of the system, see e.g. (Unbehauen and Rao, 1997). Stochastic state space models provide a means of combining the hallmarks of grey box identification (Ljung, 1987; Söderström and Stoica, 1989), namely by combining a priori knowledge about the system and statistical methods for parameter estimation and model validation. However, the formal definition of stochastic state space models in the system identification community, see e.g. (Söderström, Fan, Mossberg and Carlsson, 1997; Haverkamp, Verhaegen, Chou and Johansson, 1997) differs from the definition in other fields (Protter, 1990; Karatzas and Shreve, 1996) in a way that makes genuine probabilistic methods difficult to apply. The latter considers stochastic differential equations in the Itô sense. In (Bohlin and Graebe, 1994), it is argued that a model consisting of a stochastic differential equation (SDE) and a discrete-time measurement equation is a natural framework for modelling real dynamical systems. We confine ourselves to SDEs driven by Wiener processes, although SDEs may be driven by more general classes of stochastic processes (Protter, 1990). Itô SDEs as opposed to Stratonovitch SDEs are particularly well suited for state and parameter estimation purposes. Kloeden and Platen (1995, Chapter 7) provides an extensive list of applications of SDEs in many areas of the technical sciences. Yet it is evident that there are numerous open research problems with respect to model structure identification, choice of sampling time and related topics.

There is increasing evidence of both theoretical and empirical nature that the level of the process noise depends on the state variables. This will be referred to as *level effects*. This multiplicative process noise is an important reason for considering stochastic differential equations more rigorously in the Itô sense. In general, SDEs with level effects necessitate higher order filters (Jazwinski, 1970; Maybeck, 1982) that offers only approximate and very computerintensive solutions to the filtering problem. However, for a limited class of models, a transformation may be used to remove the level effects such that first order filters e.g. the Iterated Extended Kalman Filter (IEKF) is sufficient. The transformation has the additional advantage that the transition probability density function (pdf) of the transformed SDE is "closer" to the normal pdf.

The main contributions are a method for estimation of embedded parameters in Itô stochastic differential equations without level effects using discrete-time measurements and a multivariate generalization of the transformation proposed in (Baadsgaard, Nielsen, Spliid, Madsen and Preisel, 1997).

The remainder of this paper is organized as follows: Section 2 introduces the nonlinear, multivariate and quasi-stationary stochastic state space model and the discrete-time, multivariate measurement equation to be considered. Section 3 introduces a transformation of some SDEs. In Section 4 the IEKF is considered in the continuous-discrete time framework and linear systems are treated as an important special case. Section 5 presents a derivation of the quasi-likelihood function, and Section 6 presents some numerical aspects of a software implementation. In Section 7 Monte Carlo simulation is used to examine the statistical properties of the parameter estimates. Section 8 considers two financial applications where level effects are clearly present. Finally, Section 9 concludes.



# 2 Nonlinear stochastic differential equations

Consider the nonlinear, multivariate, quasi-stationary Stochastic Differential Equation with eXternal inputs (SDEX) in the sense of Itô calculus

$$d\mathbf{X}_t = \mathbf{f}_t(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta})dt + \mathbf{g}_t(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta})d\mathbf{W}_t;\ t_0 \le t \le T, \tag{1}$$

where $\mathbf{X}_t \in \mathbb{R}^n$ is a stochastic state vector, $\mathbf{X}_{t_0}$ is a stochastic initial condition satisfying $E[||\mathbf{X}_{t_0}||^2] < \infty$, and $\mathbf{u}_t \in \mathbb{R}^d$ is a vector of deterministic inputs (e.g. control signals), which is known for all $t$. It is assumed that the drift term $\mathbf{f}\colon [t_0, T] \times \mathbb{R}^n \times \mathbb{R}^d \mapsto \mathbb{R}^n$ and the diffusion term $\mathbf{g}\colon [t_0, T] \times \mathbb{R}^n \times \mathbb{R}^d \mapsto \mathbb{R}^{n\times m}$ satisfy sufficient regularity conditions to ensure the existence of strong solutions to (1), see (Øksendal, 1995). The process noise is modelled as a Wiener process $\mathbf{W}_t = (W_t^1, \dots, W_t^m)'$ with incremental covariance $\mathbf{Q}_t$. For reasons of identifiability it is assumed that $\mathbf{Q}_t$ is the identity matrix. The parameter vector $\boldsymbol{\theta}$ may be restricted to a subset $\Theta \subseteq \mathbb{R}^p$ due to physical considerations. The real-valued discrete-time observations $\{\mathbf{Y}_{t_k}\}$ are obtained at the sampling instants $t_1 < \dots < t_k < \dots < t_N$, where $N$ denotes the number of observations. Let $l$ denote $\dim(\mathbf{Y}_{t_k})$. The measurement equation is

$$\mathbf{Y}_{t_k} = \mathbf{h}_{t_k}(\mathbf{X}_{t_k}, \mathbf{u}_{t_k}; \boldsymbol{\theta}) + \mathbf{e}_{t_k}, \quad k = 1, \dots, N \tag{2}$$

where $\mathbf{h}$ is a nonlinear function, which is assumed to be continuously differentiable with respect to $\mathbf{X}_t$, and $\mathbf{e}_{t_k}$ is a zero mean Gaussian white noise process with covariance $\boldsymbol{\Sigma}_{t_k}$. The stochastic entities $\mathbf{X}_{t_0}$, $\mathbf{W}_t$ and $\mathbf{e}_{t_k}$ are assumed to be mutually independent for all $t$ and $t_k$.

# 3 A multivariate transformation

In this section, a generalization of the transformation proposed by (Baadsgaard et al., 1997) to a special class of multivariate SDEs is introduced. The transformation has been proposed by (Kloeden and Platen, 1995) in order to obtain closed-form solutions to some SDEs and applied by (Aït-Sahalia, 1998) as a means of obtaining a pdf that is closer to the normal pdf, but it also has an interesting application in nonlinear filtering theory, because it alleviates the need for computerintensive higher order filters.

Assume a bijective transformation of $\mathbf{X}_t$ given by

$$\mathbf{Z}_t = \Psi_t(\mathbf{X}_t) \tag{3}$$

where $\Psi_t : [t_0, T] \times \mathbb{R}^n \mapsto \mathbb{R}^n$ and $\Psi_t$ is $C^{1,2}$, i.e. it is continuously differentiable with respect to $t$ and twice continuously differentiable with respect to $\mathbf{X}_t$ such that, by Itô's multivariate formula, see e.g. (Øksendal, 1995), the SDE for $\mathbf{Z}_t$ is given by

$$d\mathbf{Z}_t = \tilde{\mathbf{f}}_t(\mathbf{Z}_t, \mathbf{u}_t; \boldsymbol{\theta})dt + \mathbf{G}_t(\mathbf{u}_t; \boldsymbol{\theta})d\mathbf{W}_t \tag{4}$$

where the diffusion term is independent of the state $\mathbf{Z}_t$. Thus it is assumed that the dimension of the Wiener process $\mathbf{W}_t$ is preserved by the transformation (3). The inverse transformation $\mathbf{X}_t = \Psi^{-1}(\mathbf{Z}_t)$ should be applied to the measurement equation (2).

**REMARK 3.1.** Note that (4) contains the same parameters as (1) and describes a relation between the same input and output variables as the originating continuous-discrete state space model (1)–(2). ▼

**ASSUMPTION 3.1.** Assume that the diffusion terms are strictly nonzero, i.e.

$$g_t^{ij}(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) \neq 0; \quad i = 1, \dots, n, j = 1, \dots, m \tag{5}$$

▲

**ASSUMPTION 3.2.** Assume that for each $i$ there exists only one $g^{ij}$ as a function of one and only one state variable $X_t^{\nu(i)}$, where $\nu(i)$ should be different for each $i$, i.e.

$$g_t^{ij}(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) = g_t^{ij}(X_t^{\nu(i)}, \mathbf{u}_t; \boldsymbol{\theta}); \quad i = 1, \dots, n, j = 1, \dots, m \tag{6}$$



Assume further that $g_t^{ij}(X_t^{\nu(i)}, \mathbf{u}_t; \boldsymbol{\theta})$ is bijective and that the reciprocal function $[g_t^{ij}(x, \mathbf{u}_t; \boldsymbol{\theta})]^{-1}$ is integrable with respect to $x$. ▲

Given these assumptions, we have the following theorem.

**THEOREM 3.1.** Let $\mathbf{X}_t$ be a solution to (1). Then Assumptions 3.1 and 3.2 provide necessary and sufficient conditions for the existence of a transformation (3) given by

$$\psi_t^k(X_t^{\nu(i)}) = \int \frac{dx}{g_t^{ij}(x, \mathbf{u}_t; \boldsymbol{\theta})} \Big|_{x=X_t^{\nu(i)}} ; \quad k, i = 1, \dots, n; j = 1, \dots, m \tag{7}$$

such that (4) is fulfilled.

*Proof.* Applying the multivariate Itô formula to (3) yields a new Itô SDE with the $k$th component $Z_t^k$, $k = 1, \dots, n$, satisfying

$$\begin{aligned} dZ_t^k = & \left( \frac{\partial \psi_t^k(\mathbf{X}_t)}{\partial t} + \sum_{i=1}^{n} \frac{\partial \psi_t^k(\mathbf{X}_t)}{\partial x_i} f_t^i(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) \right. \\ & \left. + \frac{1}{2} \sum_{j=1}^{m} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \frac{\partial^2 \psi_t^k(\mathbf{X}_t)}{\partial x^{i_1} \partial x^{i_2}} g_t^{i_1 j}(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) g_t^{i_2 j}(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) \right) dt \\ & + \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{\partial \psi_t^k(\mathbf{X}_t)}{\partial x_i} g_t^{ij}(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) dW_t^j \end{aligned} \tag{8}$$

To obtain a constant diffusion term (unity for reasons of parameter identifiability), it immediately follows that the following should hold

$$\frac{\partial \psi^k(\mathbf{X}_t)}{\partial x^i} g_t^{\nu(i)j}(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) = 1 \text{ and } \frac{\partial \psi^k(\mathbf{X}_t)}{\partial x^i} g_t^{ij}(\mathbf{X}_t, \mathbf{u}_t; \boldsymbol{\theta}) = 0 \tag{9}$$

for $k, i = 1, \dots, n; j = 1, \dots, m; \nu(i) \neq i$. Assumption 3.2 ensures that there exists only one $X_t^{\nu(i)}$ for each $\psi^k$. Under Assumption 3.1, Eq. (7) follows immediately from (9). ■

**REMARK 3.2.** The drift term $\tilde{\mathbf{f}}_t(\mathbf{Z}_t, \mathbf{u}_t; \boldsymbol{\theta})$ in the transformed system (4) is given by the factor in front of the $dt$ term in (8). Thus the transformation (3) may introduce additional nonlinearities in the drift $\tilde{\mathbf{f}}$. The implications with respect to parameter identifiability must be analyzed in each particular case. An example is given in Section 7. ▼

In the remainder of the paper, only multivariate SDEs without level effects are considered, i.e. the diffusion term is denoted by $\mathbf{G}_t(\mathbf{u}_t;\boldsymbol{\theta})$.

# 4 Nonlinear filtering

Let $\mathcal{Y}^N = [\mathbf{Y}_1,\ldots,\mathbf{Y}_N]$ denote the measurements up to and including time $t_N$. Further, let $p = p(\mathbf{x}_t|\mathcal{Y}^k)$ denote the conditional probability of the process being in state $\mathbf{X}_t = \mathbf{x}_t$ at time $t$ conditioned on the information set $\mathcal{Y}^k$ up to time $t_k$. Assuming that $\mathbf{X}_t$ is given as the solution to (1), the conditional distribution $p(\mathbf{x}_t|\mathcal{Y}^k)$ may be found as the solution to the Kolmogorov partial differential equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n}\frac{\partial(pf^i)}{\partial x^i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2(p(\mathbf{GQG}')^{ij})}{\partial x^i\partial x^j} \quad t\in[t_k,t_{k+1}) \tag{10}$$

From (10), it is possible, given $p(\mathbf{x}_{t_k}|\mathcal{Y}^k)$, which is assumed to exist and to be continuously differentiable with respect to $t$ and twice continuously differentiable with respect to $\mathbf{X}$, to determine the a priori distribution $p(\mathbf{x}_{t_{k+1}}|\mathcal{Y}^k)$ of the state vector at time $t_{k+1}$.

The a posteriori distribution of the state vector $p(\mathbf{x}_{t_{k+1}}|\mathcal{Y}^{k+1})$ at time $t_{k+1}$, when a new observation $\mathbf{y}_{t_{k+1}}$ has been acquired, follows from Bayes' formula,

$$p(\mathbf{x}_{t_{k+1}}|\mathcal{Y}^{k+1}) = \frac{p(\mathbf{y}_{t_{k+1}}|\mathbf{x}_{t_{k+1}},\mathcal{Y}^k)p(\mathbf{x}_{t_{k+1}}|\mathcal{Y}^k)}{p(\mathbf{y}_{t_{k+1}}|\mathcal{Y}^k)} \tag{11}$$

Eqs. (10)-(11) yield the exact solution to the filtering problem provided that $p$ is known. However, closed form solutions to (10)-(11) do rarely exist.

For linear systems driven by Wiener processes both the transition pdf and the stationary pdf are normal provided that the density of the initial value $\mathbf{X}_{t_0}$ is normal. The Kalman filter provides the exact solution to the filtering problem (Jazwinski, 1970), because the mean and covariance of the state estimate $\hat{\mathbf{X}}_{t|t_k}$ at time $t$ given information up to time $t_k$, $t > t_k$,

$$\hat{\mathbf{X}}_{t|t_k} = E\{\mathbf{X}_t|\mathcal{F}_k\} \tag{12}$$

$$\mathbf{P}_{t|t_k} = E\{(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t_k})(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t_k})'|\mathcal{F}_k\} \tag{13}$$

completely characterize the normal distribution. Albeit the infinitesimal transition pdf in the time interval $[t, t+dt]$ is normal due to the properties of the Wiener process, this does not hold for the transition pdfs between sampling instants for nonlinear systems. However, Eqs. (12) and (13) describe the conditional mean of the sample path and the dispersion around it, and by using these quantities to parametrize a normal pdf these moments provide a reasonable approximation to $p$ under the assumptions that $p$ is unimodal and symmetric, i.e. the higher order odd central moments are negligible, and that $p$ is centered around the mean, i.e. higher order even central moments are negligible (Jazwinski, 1970).

## 4.1 The Extended Kalman Filter

For the nonlinear system (1) and (2), the Extended Kalman Filter (EKF) is used as a first order approximative filter, i.e. the nonlinear system is approximated by the first term of a Taylor expansion. No assumptions are made about the distribution of the states and the observations in the following formulation of the EKF besides those stated in the previous section.

**THEOREM 4.1 (CONTINUOUS-DISCRETE EXTENDED KALMAN FILTER).** For the continuous-discrete time state space model (1) and (2) the linearized prediction equations are

$$\frac{d\hat{\mathbf{X}}_{t|t_k}}{dt} = \mathbf{f}_t(\hat{\mathbf{X}}_{t|t_k},\mathbf{u}_t;\boldsymbol{\theta}) \tag{14}$$

$$\frac{d\mathbf{P}_{t|t_k}}{dt} = \mathbf{F}_t(\hat{\mathbf{X}}_{t|t_k},\mathbf{u}_t;\boldsymbol{\theta})\mathbf{P}_{t|t_k} + \mathbf{P}_{t|t_k}\mathbf{F}_t'(\hat{\mathbf{X}}_{t|t_k},\mathbf{u}_t;\boldsymbol{\theta}) + \mathbf{G}_t(\mathbf{u}_t;\boldsymbol{\theta})\mathbf{Q}_t\mathbf{G}_t'(\mathbf{u}_t;\boldsymbol{\theta}) \tag{15}$$

with initial conditions $\hat{\mathbf{X}}_{k|t_k}$ and $\mathbf{P}_{k|t_k}$, respectively, for $t\in[t_k,t_{k+1})$.

When a new observation becomes available at time $t_{k+1}$ the update is given by

$$\hat{\mathbf{X}}_{t_{k+1}|t_{k+1}} = \hat{\mathbf{X}}_{t_{k+1}|t_k} + \mathbf{K}_{t_{k+1}}\boldsymbol{\epsilon}_{t_{k+1}} \tag{16}$$

$$\mathbf{P}_{t_{k+1}|t_{k+1}} = \mathbf{P}_{t_{k+1}|t_k} - \mathbf{K}_{t_{k+1}}\mathbf{H}_{t_{k+1}}(\hat{\mathbf{X}}_{t_{k+1}|t_k},\mathbf{u}_{t_{k+1}};\boldsymbol{\theta})\mathbf{P}_{t_{k+1}|t_k} \tag{17}$$

where the mean and covariance of the one-step prediction errors are given by

$$\begin{aligned} \boldsymbol{\epsilon}_{t_{k+1}} &= \mathbf{Y}_{t_{k+1}} - \mathbf{h}_{t_{k+1}}(\hat{\mathbf{X}}_{t_{k+1}|t_k}, \mathbf{u}_{t_{k+1}}; \boldsymbol{\theta}) & (18)\\ \mathbf{R}_{t_{k+1}|t_k} &= \mathbf{H}_{t_{k+1}}(\hat{\mathbf{X}}_{t_{k+1}|t_k}, \mathbf{u}_{t_{k+1}}; \boldsymbol{\theta})\mathbf{P}_{t_{k+1}|t_k}\mathbf{H}'_{t_{k+1}}(\hat{\mathbf{X}}_{t_{k+1}|t_k}, \mathbf{u}_{t_{k+1}}; \boldsymbol{\theta}) \\ &\quad + \boldsymbol{\Sigma}_{t_{k+1}} & (19) \end{aligned}$$

and $\mathbf{F}_t$ and $\mathbf{H}_{t_{k+1}}$ are given by the linearizations

$$\begin{aligned} \mathbf{F}_t(\hat{\mathbf{X}}_{t|t_k}, \mathbf{u}_t; \boldsymbol{\theta}) &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\hat{\mathbf{X}}_{t|t_k}} & (20)\\ \mathbf{H}_{t_{k+1}}(\hat{\mathbf{X}}_{t_{k+1}|t_k}, \mathbf{u}_{t_{k+1}}; \boldsymbol{\theta}) &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\hat{\mathbf{X}}_{t_{k+1}|t_k}} & (21) \end{aligned}$$

The Kalman gain is given by

$$\mathbf{K}_{t_{k+1}} = \mathbf{P}_{t_{k+1}|t_k}\mathbf{H}'_{t_{k+1}}(\hat{\mathbf{X}}_{t_{k+1}|t_k}, \mathbf{u}_{t_{k+1}}; \boldsymbol{\theta})\mathbf{R}^{-1}_{t_{k+1}|t_k} \qquad (22)$$

*Proof.* See (Jazwinski, 1970). ■

**REMARK 4.1.** If the nonlinearities in (1) vary fast compared to the sampling frequency, the first order Taylor approximation applied in (14) may be too crude. ▼

## 4.2 The Iterated Extended Kalman Filter

In the EKF the evolution of the nonlinear system (1) in each sampling instant is based on the state estimate at time $t_k$, $\hat{\mathbf{X}}_{t_k|t_k}$. A better approximation is provided by the Iterated Extended Kalman Filter (IEKF), which evaluates the linearized version of (1) at $r$ equidistant subsampling instants in each sampling interval $[t_k, t_{k+1})$. In particular, for linear systems (1) with a nonlinear measurement equation (2) the IEKF provides a good approximation (Jazwinski, 1970).

**THEOREM 4.2 (CONTINUOUS-DISCRETE ITERATED EKF).** Let $\boldsymbol{\eta}_i$ denote the "pseudo-state" at time $t \in [t_k, t_{k+1})$. The Iterated Extended Kalman Filter consists of Theorem 4.2 with (16) replaced by

$$\begin{aligned} \boldsymbol{\eta}_{i+1} &= \hat{\mathbf{X}}_{t_{k+1}|t_k} + \mathbf{K}_{t_{k+1}}(\boldsymbol{\eta}_{i+1})[(\mathbf{Y}_{t_{k+1}} - \mathbf{h}_{t_{k+1}}(\boldsymbol{\eta}_i, \mathbf{u}_t; \boldsymbol{\theta}) \\ &\quad - \tilde{\mathbf{H}}(\hat{\mathbf{X}}_{t_{k+1}|t_k} - \boldsymbol{\eta}_i)] & (23) \end{aligned}$$



where

$$\tilde{\mathbf{H}} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\boldsymbol{\eta}_i} \qquad (24)$$

and $\mathbf{K}_{t_{k+1}}$ is given by (22). Starting with $\boldsymbol{\eta}_1 = \hat{\mathbf{X}}_{t_{k+1}|t_k}$ and terminating with $\boldsymbol{\eta}_r = \hat{\mathbf{X}}_{t_{k+1}|t_{k+1}}$, Eq. (23) is iterated for $i = 1, \ldots, r$.

*Proof.* See (Jazwinski, 1970). ■

## 4.3 Linear systems

Consider for a moment the linear, multivariate, time-varying stochastic differential equation

$$\begin{aligned} d\mathbf{X}_t &= \mathbf{A}_t\mathbf{X}_t\,dt + \mathbf{B}_t\mathbf{u}_t\,dt + \mathbf{G}_t\,d\mathbf{W}_t; \quad \mathbf{X}_0 = \mathbf{X}_{t_0} & (25)\\ \mathbf{Y}_{t_k} &= \mathbf{C}_{t_k}\mathbf{X}_{t_k} + \mathbf{D}_{t_k}\mathbf{u}_{t_k} + \mathbf{e}_{t_k}, \quad k = 1, \ldots, N & (26) \end{aligned}$$

where the $n \times n$ system matrix $\mathbf{A}_t$, the $n \times d$ input-coupling matrix $\mathbf{B}_t$, the $l \times n$ state measurement matrix $\mathbf{C}_{t_k}$, the $l \times d$ transmission matrix $\mathbf{D}_{t_k}$ and $\mathbf{G}_t$ are assumed to be known up to the parameter vector $\boldsymbol{\theta}$.

**REMARK 4.2.** This system may be obtained from the nonlinear system (1) and the nonlinear measurement equation (2) using the linearizations (20) and (21) at each sampling instant. The system should also be linearized with respect to $\mathbf{u}_t$. ▼

The following well-known result is essential for the proposed estimation method, because it provides a connection between the continuous-time and discrete-time systems.

**LEMMA 4.1.** Let $\tau_k = t_{k+1} - t_k$ denote the sampling time, and let $\hat{\mathbf{X}}_{t_{k+1}|t_k} = E\{\mathbf{X}_{t_{k+1}}|\mathcal{F}_{t_k}\}$ and $\mathbf{P}_{t_{k+1}|t_k} = V\{\mathbf{X}_{t_{k+1}}|\mathcal{F}_{t_k}\}$ denote the mean and covariance of the predicted state estimate, respectively. Assume that $\mathbf{A}_t$, $\mathbf{B}_t$ and $\boldsymbol{\Sigma}_t = \mathbf{G}_t\mathbf{Q}_t\mathbf{G}'_t$ are constant during the sampling interval $[t_k, t_{k+1})$.

Assuming that $\mathbf{u}_t$ is stepwise constant, i.e. $\mathbf{u}_t = \mathbf{u}_{t_k}$, $t \in [t_k, t_{k+1})$, it holds that

$$\begin{aligned} \hat{\mathbf{X}}_{t_{k+1}|t_k} &= \boldsymbol{\Phi}_{t_k}\hat{\mathbf{X}}_{t_k|t_k} + \boldsymbol{\Gamma}_{t_k}\mathbf{u}_{t_k} & (27)\\ \mathbf{P}_{t_{k+1}|t_k} &= \boldsymbol{\Phi}_{t_k}\mathbf{P}_{t_k|t_k}\boldsymbol{\Phi}'_{t_k} + \mathbf{R}^{(1)}_{t_k} & (28) \end{aligned}$$

where

$$\Phi_{t_k} = e^{\mathbf{A}_{t_k}\tau_k} \tag{29}$$

$$\Gamma_{t_k} = \int_{t_k}^{t_{k+1}} e^{\mathbf{A}_{t_k}(t_k+\tau_k-s)}\mathbf{B}_{t_k}ds = \int_0^{\tau_k} e^{\mathbf{A}_{t_k}s}\mathbf{B}_{t_k}ds \tag{30}$$

$$\mathbf{R}^{(1)}_{t_k} = \int_0^{\tau_k} \Phi_s\Sigma_{t_k}\Phi'_s ds; \quad \Phi_s = e^{\mathbf{A}_{t_k}s} \tag{31}$$

Assuming instead that $\mathbf{u}_t$ is linear in the sampling interval, i.e.

$$\mathbf{u}_t = \frac{t-t_k}{\tau_k}(\mathbf{u}_{t_{k+1}} - \mathbf{u}_{t_k}) + \mathbf{u}_{t_k}; \quad t \in [t_k, t_{k+1}[ \tag{32}$$

Then (27) is replaced by

$$\hat{\mathbf{X}}_{t_{k+1}|t_k} = \Phi_{t_k}\hat{\mathbf{X}}_{t_k|t_k} + \Gamma_{t_k}\mathbf{u}_{t_k} + \Upsilon_{t_k}(\mathbf{u}_{t_{k+1}} - \mathbf{u}_{t_k}) \tag{33}$$

where $\Phi_{t_k}$ and $\Gamma_{t_k}$ are given by (30) and

$$\Upsilon_{t_k} = \int_0^{\tau_k} e^{\mathbf{A}_{t_k}s}\mathbf{B}_{t_k}\frac{\tau_k - s}{\tau_k}ds \tag{34}$$

*Proof.* Follows directly by integrating (25) from $t_k$ to $t_{k+1}$ and taking expectations, see e.g. (Åström, 1970) or (Kloeden and Platen, 1995). ■

Clearly the solution to (25) involves the exponential matrix and it may be infeasible to obtain $\mathbf{A}_t$, and henceforth $\boldsymbol{\theta}$ using indirect methods (e.g. by estimating $\Phi$ and taking the natural logarithm). The inherent problems associated with indirect parameter estimation methods are covered in (Haverkamp et al., 1997). In addition these methods do not allow for estimation of unobserved state variables, see Section 5.1.

**REMARK 4.3.** In numerical work the exponential matrix may be computed using the Padé approximation (Moler and van Loan, 1978). It is easily shown that the Tustin approximation $z = -(j\omega\tau_k + 2)/(j\omega\tau_k - 2)$ used in e.g. (Unbehauen and Rao, 1987; Söderström et al., 1997) corresponds to the first order (1,1) Padé approximation of $z = e^{j\omega\tau_k}$. In our work a Padé approximation of higher order with repeated scaling and squaring is used as suggested by (Moler and van Loan, 1978). ▼

**REMARK 4.4.** Assuming that $\mathbf{A}_t$ in (25) is time-invariant and that the system is equidistantly sampled, the exponential matrix in (30) need only to be computed once, see also (Pedersen, 1994). In other cases, but still under the assumption that $\mathbf{A}_t$ is approximately constant during the sample interval $[t_k, t_{k+1})$, it must be computed at each sampling instant $t_k$. ▼

# 5 Maximum likelihood method

The maximum likelihood method is based on an assumption of normality for the innovations (18). The method is similar to the prediction error decomposition method proposed by (Schweppe, 1965).

Given all the observations $\mathcal{Y}^N$, the likelihood function, $\mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}^N)$, is expressed as the joint probability density of $\mathcal{Y}^N$ provided that the parameters are known, i.e.

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}^N) &= p(\mathcal{Y}^N|\boldsymbol{\theta}) = p(\mathbf{y}_N|\mathcal{Y}^{N-1}, \boldsymbol{\theta})p(\mathcal{Y}^{N-1}|\boldsymbol{\theta}) \\ &= \prod_{k=1}^{N} p(\mathbf{y}_k|\mathcal{Y}^{k-1}, \boldsymbol{\theta})p(\mathbf{y}_0|\boldsymbol{\theta}) \\ &= \tilde{\mathcal{L}}(\boldsymbol{\theta}; \mathcal{Y}^N|\mathbf{y}_0)p(\mathbf{y}_0|\boldsymbol{\theta}) \end{aligned} \tag{35}$$

where successive applications of the rule $P(A\cap B) = P(A|B)P(B)$ is used to express the likelihood function as a product of conditional densities. Note that (35) is expressed as the product of the conditional likelihood function $\tilde{\mathcal{L}}(\boldsymbol{\theta}; \mathcal{Y}^N|\mathbf{y}_0)$ and the probability density of the initial observation $\mathbf{Y}_0$, $p(\mathbf{y}_0|\boldsymbol{\theta})$, where the latter is expressed in terms of $p(\mathbf{x}_0)$ such that the initial value of the state $\mathbf{X}_0$ may be estimated.

**REMARK 5.1.** Prior information about the parameters $\boldsymbol{\theta}$, if any, can be included in (35) by multiplying the likelihood function by the prior distribution $\pi(\boldsymbol{\theta})$ of the parameters, i.e. the parameters are considered as stochastic variables rather than constants. This Bayesian approach gives rise to various possibilities for point estimates of $\boldsymbol{\theta}$ e.g. the Maximum A Posteriori (MAP) estimate proposed by (Goodwin and Payne, 1977), which is obtained as the mode of the posterior pdf. ▼

The innovations (18) are independent stochastic variables with zero mean and pdf $p(\boldsymbol{\epsilon}_{t_k}(\boldsymbol{\theta}))$ such that the conditional likelihood function, $\tilde{\mathcal{L}}$ in (35), may be expressed in terms of the prediction error decomposition, i.e.

$$\tilde{\mathcal{L}}(\boldsymbol{\theta};\mathcal{Y}^N|\mathbf{y}_0) = \prod_{k=1}^{N} p(\boldsymbol{\epsilon}_{t_k}(\boldsymbol{\theta})) \tag{36}$$

Assuming that the innovations are normal with zero mean and covariance matrix $\mathbf{R}_{t_k|t_{k-1}}$, it is convenient to consider the logarithm of (36), i.e.

$$\begin{aligned} l(\boldsymbol{\theta}) &= -\log \tilde{\mathcal{L}}(\boldsymbol{\theta};\mathcal{Y}^N|\mathbf{y}_0) \\ &= \frac{1}{2}\sum_{k=1}^{N}\left(\boldsymbol{\epsilon}'_{t_k}\mathbf{R}^{-1}_{t_k|t_{k-1}}\boldsymbol{\epsilon}_{t_k} + \log\det\mathbf{R}_{t_k|t_{k-1}} + l\log 2\pi\right) \end{aligned} \tag{37}$$

such that the maximum likelihood (ML) estimate is determined by minimizing the negative log-likelihood function, i.e.

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l(\boldsymbol{\theta}) \tag{38}$$

An estimate of the uncertainty of the parameters is obtained using the fact that the ML-estimator is asymptotically normal with mean $\boldsymbol{\theta}$ and covariance $\boldsymbol{\Sigma}$ given by the lower bound of the Cramer-Rao inequality, i.e.

$$\boldsymbol{\Sigma} = \mathbf{H}^{-1} \tag{39}$$

where the Hessian matrix $\mathbf{H} = \{h_{ij}\}$ is given by

$$\{h_{ij}\} = -E\left\{\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta^i\,\partial\theta^j}\right\} \tag{40}$$

An estimate of $\mathbf{H}$ is obtained by equating the observed value with its expectation and applying

$$\{h_{ij}\} \approx -\left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta^i\,\partial\theta^j}\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{41}$$



such that the covariance matrix of the estimated parameter vector is readily available.

For linear systems, $p$ is normal such that an exact maximum likelihood method is readily obtained. For nonlinear systems, the pdf $p$ implied by the model (1) is not normal. However, under the assumption that $p$ is approximately normal, the first order filters described in Section 4 still applies. This assumption may be tested using standard statistical tests for Gaussianity. If these tests are rejected at all reasonable levels of significance, then the method can be considered as a prediction error method (Klimko and Nelson, 1978; Ljung, 1987).

## 5.1 Related work

Extensive literature exists on filtering and estimation in the discrete-discrete time framework, where the model (1) is also formulated in discrete-time, see e.g. (Söderström and Stoica, 1989; Ljung, 1987; Harvey, 1989; Tanizaki, 1996). Similarly for the continuous-continuous framework, where (2) is also formulated as a SDE, see e.g. (Øksendal, 1995; Pugachev and Sinitsyn, 1987). The latter framework is useful for design purposes, but it is argued that for filtering and estimation it is inappropriate.

In recent years, a large variety of indirect orthogonal basis function methods have been proposed for stochastic state space models such that well-known LS-estimation methods may be applied (Young, 1981; Sagara and Zhao, 1990; Unbehauen and Rao, 1997). The Poisson moment functionals (PMF) method retain the physical interpretation of the CT parameters and avoids direct derivative measurements, which tends to accentuate the measurement noise, but it is only applicable for models that are linear in the parameters (Unbehauen and Rao, 1990). This also applies for Laguerre functions and Hartley modulating functions (Unbehauen and Rao, 1997). PMF suppress low levels of measurement noise by low-pass filtering, but it cannot cope with both process and measurement noise. The latter also applies for state-variable filtering methods (Haverkamp et al., 1997). For high levels of noise instrumental variables (IV) may be used to remove the bias of the PMF method (Unbehauen and Rao, 1990).

The Indirect Prediction Error Method (IPEM), proposed by (Söderström, Stoica and Friedlander, 1991), is difficult to analyze with respect to identifiability and infeasible for unevenly sampled data (Bigi, Söderström and Carlsson, 1994). In particular, the concept of embedding, where the CT parameters are obtained by taking the logarithm of the DT transition matrix $\boldsymbol{\Phi}_k$ in (30), poses at least one restriction on the applicability of IPEM, i.e. the eigenvalues of $\boldsymbol{\Phi}_k$ must be positive.

In addition the transition matrix is only an approximation for nonlinear systems.

Forward or backward Euler discretization schemes of the CT model are considered in (Bigi et al., 1994), where they are shown to yield biased estimates, see also (Söderström, 1999). An IV bias-correction method is suggested in (Söderström et al., 1997). Discretization of SDEs is in general complicated due to Itô's Lemma and conventional discretization methods may converge to a SDE in the FS sense as opposed to the appropriate Itô sense (Kloeden and Platen, 1995), so it is difficult to extend classical discretization based methods to cover SDE models. See (Nielsen, Madsen and Young, 1999) for an extensive overview of other estimation methods, which, mostly, considers stochastic state space systems modelled by SDEs.

The Extended Kalman Filter Parameter Estimator (EKFPE) may be used in a continuous-discrete time framework, but it is shown in (Wiberg and DeWolf, 1993) that for CT-systems it fails, with probability one, to converge to the true values of the parameters in a system whose state noise covariance is unknown. For DT systems, a similar result has been shown by (Ljung, 1979).

We argue that the method for direct estimation of embedded parameters of the CT model suggested in the present paper solves most of the problems associated with the other methods in the literature partially due to a proper mathematical specification of the system noise in terms of Wiener processes. The proposed estimation method[1] has been applied to stochastic modelling of environmental systems (Jacobsen and Madsen, 1996; Jacobsen, Madsen and Harremoës, 1996), hydraulic robots (Schmidt, Madsen, Zhou and Hansen, 1997), heat dynamics of buildings (Madsen and Holst, 1995; Nielsen and Madsen, 1996), bond pricing (Nielsen, 1996) and interest rate modelling (Baadsgaard et al., 1997).

# 6 Software implementation

The maximum likelihood procedure described in the previous section is implemented in a software package called CTLSM[2]. The package contains support for preprocessing (or filtering) of the observations and input variables, constraints on the parameter space $\boldsymbol{\theta}$ during optimization and some postprocessing (or model verification) tools.

[1] An implementation of the proposed method written in Fortran 77 may be obtained upon request from the authors.

[2] CTLSM is an acronym for Continuous-Time Linear Stochastic Modelling.



## 6.1 Constrained optimization

A quasi-Newton method based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) updating formula for a secant approximation of the Hessian matrix (40) and soft line search is used for optimizing the likelihood function (37). The first order derivatives of (37) are approximated by finite differences (Dennis and Schnabel, 1983). The minimization algorithm is an extended and modified version of the subroutine VA13CD from the Harwell Library (1989), see (Melgaard and Madsen, 1993).

Although the exact parameter values are not known a priori, the parameter space might be constrained due to prior knowledge of the dynamics of the system. To implement simple constraints on the form $\theta_i^{\text{min}} < \theta_i < \theta_i^{\text{max}}$ the optimization is made with respect to a transformation of the original parameters

$$\theta_i' = \log\left(\frac{\theta_i - \theta_i^{\text{min}}}{\theta_i^{\text{max}} - \theta_i}\right) \qquad i = 1, \ldots, p$$

Furthermore, a penalty function given by the inverse Lagrange relaxation

$$P(\lambda, \boldsymbol{\theta}, \boldsymbol{\theta}^{\text{min}}, \boldsymbol{\theta}^{\text{max}}) = \lambda\left(\sum_{i=1}^{p}\frac{|\theta_i^{\text{min}}|}{\theta_i - \theta_i^{\text{min}}} + \sum_{i=1}^{p}\frac{|\theta_i^{\text{max}}|}{\theta_i^{\text{max}} - \theta_i}\right) \tag{42}$$

is added to the likelihood function (37). For proper choices of $\lambda$, $\theta_i^{\text{min}}$ and $\theta_i^{\text{max}}$, Eq. (42) has little influence on the estimation when $\boldsymbol{\theta}$ is well within the constraints, but it will force the gradient to increase once $\theta_i$ moves towards the imposed constraints.

## 6.2 Robustness in the estimation

The log-likelihood function (37) is approximately a quadratic loss function in $\boldsymbol{\epsilon}_k$, which implies that the obtained estimates are greatly influenced by outliers. To decrease the sensitivity to outliers (or increase the robustness of the estimation) the quadratic term in $\boldsymbol{\epsilon}_k$, i.e.

$$e_{t_k} = \boldsymbol{\epsilon}_{t_k}' \mathbf{R}_{t_k|t_{k-1}}^{-1} \boldsymbol{\epsilon}_{t_k} \tag{43}$$

is replaced by the function

$$F(e_{t_k}) = \begin{cases} e_{t_k} & \text{for} \quad e_{t_k} < c \\ c(2\sqrt{e_{t_k}} - c) & \text{for} \quad e_{t_k} \geq c \end{cases} \tag{44}$$

which is quadratic in $\epsilon_{t_k}$ for small $e_{t_k}$ and linear in $\epsilon_{t_k}$ for large $e_{t_k}$. Numerical studies show that $c = 3$ is a reasonable choice, but the choice may be case-dependent.

### 6.3 Alternative implementations

It is outside the scope of the present paper to discuss all available software packages, but some comments will be put forth in this Section. Emphasis will be placed on the two software packages, *IdKit* described in (Graebe, 1990*b*), see (Bohlin and Graebe, 1995) for additional references and applications, and *CTLSM*. Both packages may be applied to the transformed model proposed in Section 3.

*IdKit* and *CTLSM* are both based on a Prediction Error Decomposition (PED) that provides the residuals for which a likelihood function is specified assuming that the residuals are Gaussian. The residuals are obtained using the KF and the IEKF, respectively, in the following way: *IdKit* computes a deterministic reference trajectory by numerical integration of (1) under the assumption that the diffusion is zero and applies the KF to a linear perturbation model (linearized about the reference trajectory) to obtain also the covariances, whereas *CTLSM* solves a linearized version of (1) by means of the exponential matrix using the IEKF. The perturbation approach used in *IdKit* is only feasible provided that the level of the process noise is sufficiently "small", cf. (Graebe, 1990*a*). This assumption that may be too restrictive for SDEs with level effects is not made in *CTLSM* which is based entirely on a stochastic model specification and implementation of the filter as argued in (Mortensen, 1969). To decrease the sensitivity of outliers *CTLSM* uses a transformation of the residuals in the optimization.

Both packages allows for imposing constraints on the parameters. In *CTLSM* the mean and covariance of a Gaussian a priori pdf of the parameters may be specified. In the optimization *IdKit* uses a finite difference approximation to the derivative of the residuals wrt. the parameters and uses that to compute the gradient and the Hessian, and finds the optimum by Gauss-Newton iterations. *CTLSM* computes finite difference approximations to the gradient and the Hessian, and a quasi-Newton method based on the BFGS updating formula for a secant approximation of the Hessian and soft-line search to find the optimum. For model validation *IdKit* uses the weighted square of scores, while *CTLSM* computes the portmanteau-lack-of-fit test statistic, the autocorrelation function of the residuals, crosscorrelation functions between the residuals and the input variables and a cumulative residual periodogram.

Both packages rely on a large amount of *provided application-independent* code



and a small amount of *application-dependent* code. However, *IdKit* and the associated user's shell, *IKUS*, constitutes a more modern environment for system identification.

## 7 A simulation study

In this section, the statistical properties of the parameter estimation method is examined using Monte Carlo simulation of a nonlinear system.

Consider a nonlinear system describing the temporal oscillations of a population of predators and their prey in a localized geographic region, where the prey population is denoted by $X_t^1$ and the predator population by $X_t^2$. It is assumed that the growth rate of prey, in the absence of predators, is $\theta_1 X_t^1$, while the predator multiplication rate is $\theta_3 X_t^1 X_t^2$. Further, it is assumed that the loss rate of prey is proportional to the numbers of prey and predators, i.e. loss rate of prey equals $-\theta_2 X_t^1 X_t^2$, while the loss rate of predators equals their death rate $-\theta_4 X_t^2$. These assumptions lead to the deterministic Lotka-Volterra equations, but assuming that the rate of change of each population is further influenced by an additive stochastic term the interaction is described by the stochastic Lotka-Volterra equations, i.e.

$$dX_t^1 = X_t^1(\theta_1 - \theta_2 X_t^2)dt + \sigma_1 dW_t^1 \tag{45}$$

$$dX_t^2 = X_t^2(\theta_3 X_t^1 - \theta_4)dt + \sigma_2 dW_t^2 \tag{46}$$

where $\theta_i$, $i = 1, 2, 3, 4$, and $\sigma_1, \sigma_2$ are positive constants, and $W_t^1$ and $W_t^2$ are assumed to be mutually uncorrelated standard Wiener processes. The diffusion terms represent the effect of other factors not included in the model, such as other predators in the food chain and weather conditions.

It is assumed that the populations are measured at discrete time instants

$$Y_{t_k}^1 = X_{t_k}^1 + e_{t_k}^1 \tag{47}$$

$$Y_{t_k}^2 = X_{t_k}^2 + e_{t_k}^2 \tag{48}$$

where $e_{t_k}^1$ and $e_{t_k}^2$ are mutually uncorrelated zero mean normal white noise processes with variance $\sigma_{1,m}^2$ and $\sigma_{2,m}^2$, respectively.

For the purpose of making a statistical evaluation of the estimation method 50 sequences of stochastic independent realizations of the system were simulated using an Euler discretization scheme (Kloeden and Platen, 1995). Each sequence

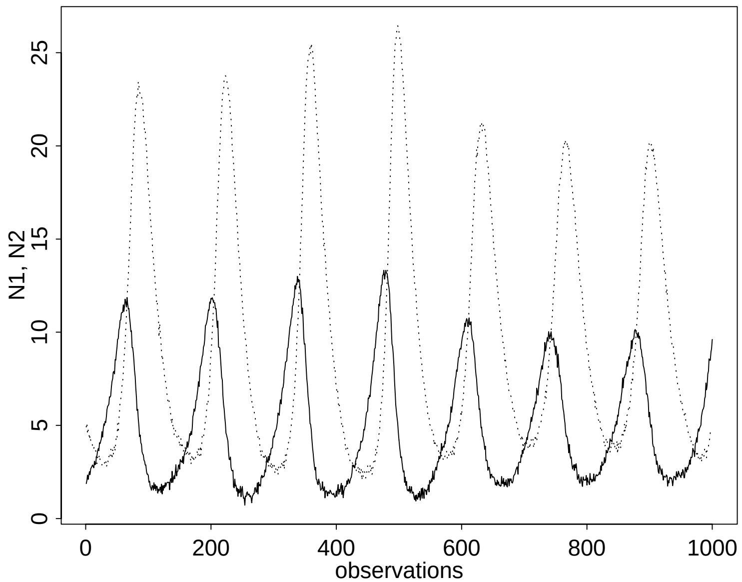


Figure 1: Time series plot of the populations $X_t^1$ and $X_t^2$ for the stochastic Lotka-Volterra system (45)-(46).

consists of 500 observations. The chosen parameters were $\theta_1 = 10$, $\theta_2 = 1$, $\theta_3 = 2$, $\theta_4 = 10$, $\sigma_1^2 = \sigma_2^2 = 0.3$, $\sigma_{1,m}^2 = \sigma_{2,m}^2 = 0.03$, the sampling time, $T_s = 0.005$, and the deterministic initial populations $(x_0^1, x_0^2) = (2, 5)$. The time series of the measured populations of one of these sequences are shown in Figure 1. It is clear that the nonlinear system (45)-(46) exhibits selfsustained oscillations. The estimation results are provided in Table 1.

## 7.1 Statistical tests

The asymptotic normality of the parameter estimates makes it possible to do statistical inference. In particular for simulated data, a statistical test for unbiasedness and a test for correct estimation of the variances of the parameter estimates may be constructed as follows.

In order to verify if the variance of the parameter estimates provided by the pro-



| Parameter | $x_{sim}$ | $\bar{X}$ | $s_x^2$ | $\bar{s^2}$ | $F$-stat. | $\lvert t\rvert$-stat. |
|---|---|---|---|---|---|---|
| $\theta_1$ | 10.0000 | 9.9850 | 8.568e-3 | 9.589e-3 | 0.8936 | 1.151 |
| $\theta_2$ | 1.0000 | 0.9990 | 6.453e-5 | 8.267e-5 | 0.7805 | 0.914 |
| $\theta_3$ | 2.0000 | 2.0012 | 1.253e-4 | 1.095e-4 | 1.1441 | 0.758 |
| $\theta_4$ | 10.0000 | 10.0010 | 4.756e-3 | 4.481e-3 | 1.0614 | 1.446 |
| $x_0^1$ | 2.0000 | 2.0066 | 5.426e-3 | 5.888e-3 | 0.9215 | 0.630 |
| $x_0^2$ | 5.0000 | 5.0175 | 7.893e-3 | 5.902e-3 | 1.3373 | 1.393 |
| $\sigma_1^2$ | 0.3000 | 0.2909 | 2.883e-3 | 3.529e-3 | 0.8169 | 1.196 |
| $\sigma_2^2$ | 0.3000 | 0.2896 | 9.011e-3 | 7.981e-3 | 1.1290 | 0.773 |
| $\sigma_{1,m}^2$ | 0.0300 | 0.02966 | 3.229e-6 | 3.855e-6 | 0.8377 | 1.346 |
| $\sigma_{2,m}^2$ | 0.0300 | 0.03047 | 4.593e-6 | 4.184e-6 | 1.0978 | 1.534 |

Table 1: Results from estimation of the $n_e = 50$ series from the stochastic Lotka-Volterra system. The columns of the table are: $x_{sim}$, the values used for the simulation, $\bar{X}$, the mean of the estimated values, $s_x^2$ is the empirical variance of the estimated parameters, $\bar{s^2}$ is the mean of the estimated variance of the parameters, $F$-stat. is a $F$-distributed statistic given by $Z_F = s_x^2/\bar{s^2}$ and $|t|$-stat. is a $t$-distributed statistic given by $Z_t = |\bar{X} - x_{sim}|/(s_x\sqrt{n_e})$.

posed method is equal to the empirical variance of the estimates, one may wish to test the hypothesis

$$\mathrm{H}_{01}: s_x^2 = \bar{s^2} \quad \text{against} \quad \mathrm{H}_{11}: s_x^2 \neq \bar{s^2}$$

where $s_x^2$ is the empirical variance of the estimated parameters and $\bar{s^2}$ is the mean of the estimated variance of the parameters. The $F$-statistic under the null is given by $z = s_x^2/\bar{s^2}$, and with a 10% level of significance the critical set is $\{z < 0.70 \wedge z > 1.35\}$. It is seen from Table 1 that $H_{01}$ cannot be rejected for any parameter on the chosen level. This indicates that the method provides correct estimates of the variance of the parameter estimates.

Another test is performed in order to examine if the estimated parameters are unbiased. The following hypotheses is tested

$$\mathrm{H}_{02}: \bar{X} = x_{sim} \quad \text{against} \quad \mathrm{H}_{12}: \bar{X} \neq x_{sim}$$

Under $H_{02}$ the distribution of the test statistic is $Z_t \sim t(49)$. The critical set is $\{z > t(49)_{1-\alpha/2}\}$ on level $\alpha$. For $\alpha = 10\%$, the critical set is $\{|z| > 2.0\}$. Thus, from Table 1, it follows that $H_{02}$ cannot be rejected for any of the parameters on the chosen level. This indicates that the method provides correct estimates of the parameters.

As an illustration of the transformation proposed in Section 3, assume that the growth rates are subject to uncertainty which depends on the current population then level effects are introduced in the model (45)–(46), i.e.

$$dX_t^1 = X_t^1(\theta_1 - \theta_2 X_t^2)dt + \sigma_1\sqrt{X_t^1}dW_t^1 \tag{49}$$

$$dX_t^2 = X_t^2(\theta_3 X_t^1 - \theta_4)dt + \sigma_2\sqrt{X_t^2}dW_t^2 \tag{50}$$

where $\theta_i$, $i = 1,2,3,4$, and $\sigma_1, \sigma_2$ are positive constants, and $W_t^1$ and $W_t^2$ are mutually uncorrelated Wiener processes as before. The measurements are still modelled by (47)–(48).

For this system the transformations

$$\psi_t^i(\mathbf{X}_t) = \psi_t^i(X_t^i) = \int \frac{dx}{\sqrt{x}}\Big|_{x=X_t^i} = 2\sqrt{X_t^i}; \quad i = 1,2$$

yield the transformed model

$$dZ_t^1 = \tfrac{1}{2}\left(\theta_1 Z_t^1 - \tfrac{1}{4}\theta_2 Z_t^1 (Z_t^2)^2 - (Z_t^1)^{-1}\right)dt + dW_t^1 \tag{51}$$

$$dZ_t^2 = \tfrac{1}{2}\left(\tfrac{1}{4}\theta_3 (Z_t^2)^2 Z_t^2 - \theta_4 Z_t^2 - (Z_t^2)^{-1}\right)dt + dW_t^2 \tag{52}$$

$$Y_{t_k}^1 = \tfrac{1}{4}(Z_{t_k}^1)^2 + e_{t_k}^1 \tag{53}$$

$$Y_{t_k}^2 = \tfrac{1}{4}(Z_{t_k}^2)^2 + e_{t_k}^2 \tag{54}$$

Note that the transformed model (51)–(54) without state-dependent diffusion terms contains exactly the same parameters as the original model (47)–(50), and that the transformed model describes also exactly the same input-output relations. For the considered case, however, the input comes only from the Wiener processes.

**REMARK 7.1.** If the diffusion term in e.g. (49) is replaced by e.g. $\sigma_1\sqrt{X_t^1 X_t^2}$ due to uncertainties in e.g. the interaction parameter $\theta_2$, then there does not exist a transformation, because this diffusion term constitutes a violation of Assumption 3.2. In this case higher order filters should be used (Jazwinski, 1970; Maybeck, 1982), see (Nielsen, Vestergaard and Madsen, 2000)(Paper **C** herein) for an example. ▼



# 8 Empirical work

In this section a financial application of the proposed transformation and estimation method is given using simulated and real data. Some comparisons are made with a second order filter (Maybeck, 1982). Finally, a limitation of the transformation method is illustrated.

## 8.1 Second order filters

For easy reference and comparison the derivation of a second order filter is outlined. The presentation follows (Maybeck, 1982; Nielsen, Vestergaard and Madsen, 1999).

Consider a univariate, homogeneous SDE given as the solution to (1) with $n = m = 1$ and no inputs, i.e.

$$dX_t = f(X_t)dt + G(X_t)dW_t \tag{55}$$

Introduce the notation $E_k\{\cdot\} = E\{\cdot|\mathbf{Y}^k\}$ for the conditional mean. The propagation of the conditional mean and variance, respectively, in the time interval $t \in [t_{i-1}, t_i)$ may be shown to satisfy

$$\frac{d\hat{X}_{t|t_{i-1}}}{dt} = E_{i-1}[f(X_t)] \tag{56}$$

$$\frac{dP_{t|t_{i-1}}}{dt} = 2E_{i-1}[f(X_t)X_t] + E_{i-1}[G^2(X_t)] - 2E_{i-1}[f(X_t)]E_{i-1}[X_t] \tag{57}$$

**REMARK 8.1.** Note that these are not ordinary differential equations, because the density $p$ is needed to compute the expected values $E_{i-1}[\cdot]$. Unfortunately a closed form expression for $p$ is rarely available. Informally, the problem is that (55) provides a very local description of the evolution of the state, and that the mapping from (55) to the conditional mean and the conditional variance of the state is not straightforward for nonnormal processes. However, it is possible to derive an approximate set of prediction and updating equations that have a structure similar to the ordinary Kalman filter (Maybeck, 1982). ▼

By performing a second order Taylor expansion about the current estimate $\hat{X}_{t|t_{k-1}}$, the following prediction equations are obtained

$$\frac{d\hat{X}_{t|t_{k-1}}}{dt} = f(\hat{X}_{t|t_{k-1}}) + \frac{P_{t|t_{k-1}}}{2}\frac{\partial^2 f(\hat{X}_{t|t_{k-1}})}{\partial X_t^2} \tag{58}$$

$$\begin{aligned}\frac{dP_{t|t_{k-1}}}{dt} &= 2\frac{\partial f(\hat{X}_{t|t_{k-1}})}{\partial X_t}P_{t|t_{k-1}} + G(\hat{X}_{t|t_{k-1}})^2 \\ &+ \left(\frac{\partial G(\hat{X}_{t|t_{k-1}})}{\partial X_t}\right)^2 P_{t|t_{k-1}} \\ &+ P_{t|t_{k-1}}G(\hat{X}_{t|t_{k-1}})\frac{\partial^2 G(\hat{X}_{t|t_{k-1}})}{\partial X_t^2} \\ &+ \frac{3}{4}P^2_{t|t_{k-1}}\left(\frac{\partial^2 G(\hat{X}_{t|t_{k-1}})}{\partial X_t^2}\right)^2 \end{aligned} \tag{59}$$

where it has been assumed that the transition density is sufficiently close to the normal density to ensure that the third and higher order odd central moments are essentially zero, that the fourth central moment may be expressed in terms of the variance, and that the sixth and higher order even central moments are negligible, see (Maybeck, 1982) for the technical details.

The updating equations are

$$\hat{X}_{t_k|t_k} = \hat{X}_{t_k|t_{k-1}} + K(Z_{t_k} - \hat{X}_{t_k|t_{k-1}}) \tag{60}$$

$$P_{t_k|t_k} = (1-K)P_{t_k|t_{k-1}} \tag{61}$$

and the Kalman gain is

$$K = \frac{P_{t_k|t_{k-1}}}{P_{t_k|t_{k-1}} + \sigma_\varepsilon^2} \tag{62}$$

Equations (58)-(62) constitute the modified truncated second order filter (Maybeck, 1982).

## 8.2 Short term interest rates

To compare the transformation approach using the EKF with a truncated second order filter, a model of short term interest rates is considered. A univariate SDE



is considered for clarity. In financial econometrics, the following model is often considered, see e.g. (Chan, Karolyi, Longstaff and Sanders, 1992),

$$dX_t = \alpha(\theta - X_t)dt + \sigma X_t^\gamma dW_t \tag{63}$$

where $X_t$ is the continuous-time short term interest rate. Many of the term structure models found in the literature may be nested within this model class by imposing appropriate parameter constraints, see (Chan et al., 1992) for a survey. For $\theta \equiv 0$, $\alpha < 0$, Eq. (63) may be used to model biological growth in a single-species population with unlimited ressources.

By inserting the diffusion term from (63) in the transformation (7) and applying Itô's Lemma the following transformed process is obtained

$$dZ_t = \left[\alpha\theta\{(1-\gamma)Z_t\}^{\frac{\gamma}{\gamma-1}} - \alpha(1-\gamma)Z_t - \frac{\gamma\sigma^2}{2(1-\gamma)Z_t}\right]dt + \sigma dW_t \tag{64}$$

Applying the inverse transformation to the measurement equation

$$Y_{t_k} = X_{t_k} + e_{t_k}$$

yields the transformed measurement equation

$$Y_{t_k} = \{(1-\gamma)Z_t\}^{\frac{1}{1-\gamma}} + e_{t_k} \tag{65}$$

For the model (63) the prediction equations (58)–(59) takes the form

$$\frac{d\hat{X}_{t|t_{k-1}}}{dt} = \alpha(\theta - \hat{X}_{t|t_{k-1}}) \tag{66}$$

$$\frac{dP_{t|t_{k-1}}}{dt} = (-2\alpha + \sigma^2\gamma(2\gamma-1)\hat{X}^{2\gamma-2}_{t|t_{k-1}})P_{t|t_{k-1}} + \sigma^2\hat{X}^{2\gamma}_{t|t_{k-1}} \tag{67}$$

The solution to (66) is

$$\hat{X}_{t|t_{k-1}} = \theta + e^{-\alpha t}(\hat{X}_{t_{k-1}|t_{k-1}} - \theta) \tag{68}$$

which shows that, for $\alpha > 0$, $\theta$ is the long-term mean. Inserting this solution into (67) yields a linear first order ODE of the form

$$\frac{dP_{t|t_{k-1}}}{dt} = -\varphi_1(t)P_{t|t_{k-1}} + \varphi_2(t) \tag{69}$$

where

$$\varphi_1(t) = 2\alpha - \sigma^2\gamma(2\gamma-1)(\theta + e^{-\alpha t}(\hat{X}_{t_{k-1}|t_{k-1}} - \theta))^{2\gamma-2} \quad (70)$$
$$\varphi_2(t) = \sigma^2(\theta + e^{-\alpha t}(\hat{X}_{t_{k-1}|t_{k-1}} - \theta))^{2\gamma} \quad (71)$$

which cannot readily be solved in closed-form. However, if $\alpha$ is close to zero, the change in the predicted mean (68) is small, the two ODEs may be uncoupled by replacing the process $\hat{X}_{t|t_{k-1}}$ with a value of the conditional mean in the sampling interval $\bar{t} \in [0, \tau_k]$. For $\bar{t} = \tau_k$, the predicted mean is inserted and the ODE may be solved, i.e.

$$P_{t_k|t_{k-1}} = -\frac{\varphi_2(\tau_k)}{\varphi_1(\tau_k)} + e^{\varphi_1(\tau_k)\tau_k}\left(P_{t_{k-1}|t_{k-1}} + \frac{\varphi_2(\tau_k)}{\varphi_1(\tau_k)}\right) \quad (72)$$

The updating equations are

$$\hat{X}_{t_k|t_k} = \hat{X}_{t_k|t_{k-1}} + \frac{P_{t_k|t_{k-1}}(Y_{t_k} - \hat{X}_{t_k|t_{k-1}})}{P_{t_k|t_{k-1}} + \sigma_e^2} \quad (73)$$
$$P_{t_k|t_k} = \frac{\sigma_e^2}{P_{t_k|t_{k-1}} + \sigma_e^2} \quad (74)$$

A closed-form solution to the prediction equations obviously leads to less computation time, and simulation studies (not reported here) confirms that this approximation does not significantly affect the parameter estimates.

### 8.2.1 A Monte Carlo study

The model (63) is solved numerically using the Euler discretization scheme (Kloeden and Platen, 1995). Each sampling interval $[t_{k-1}, t_k)$ is divided into $S = 200$ small time steps of length $\Delta = 1/S$ and independent $\mathrm{N}(0, \Delta)$ distributed random variables $\Delta W_{t_{k-1}+s\Delta}$, $s = 1, \ldots, S-1$, are simulated. A discrete time approximation to (63) is then generated by the Euler scheme, i.e.

$$\tilde{X}_{t_{k-1}+s\Delta} = \tilde{X}_{t_{k-1}+(s-1)\Delta} + \alpha(\theta - \tilde{X}_{t_{k-1}+(s-1)\Delta})\Delta + \sigma\tilde{X}^{\gamma}_{t_{k-1}+(s-1)\Delta}\Delta W_{t_{k-1}+s\Delta} \quad (75)$$

Using this scheme 50 stochastic independent time series consisting each of $N = 2000$ observations are generated.



| Extended Kalman filter with transformation | | | | |
|---|---|---|---|---|
| Parameter | True values | Mean | $t$-value | $F$-value |
| $\alpha$ | 0.0250 | 0.0261 | 1.4409 | 0.9862 |
| $\theta$ | 10.0000 | 10.1700 | 2.5767 | 0.7881 |
| $\sigma^2$ | 0.0100 | 0.0099 | -0.1314 | 0.9922 |
| $\gamma$ | 0.7500 | 0.7515 | 0.1945 | 0.8150 |
| $\sigma_e^2$ | 0.0500 | 0.0562 | 4.9328 | 1.1299 |

| Truncated Second order filter | | | | |
|---|---|---|---|---|
| Parameter | True values | Mean | $t$-value | $F$-value |
| $\alpha$ | 0.0250 | 0.0263 | 1.7078 | 0.9467 |
| $\theta$ | 10.0000 | 10.1680 | 2.5468 | 1.2039 |
| $\sigma^2$ | 0.0100 | 0.0111 | 1.3518 | 0.6872 |
| $\gamma$ | 0.7500 | 0.7508 | 0.0633 | 0.8984 |
| $\sigma_e^2$ | 0.0500 | 0.0482 | -1.3724 | 0.9624 |

Table 2: Results for 50 sample sequences from the short rate model (63) with large variations.

In Tables 2 and 3 the estimation results for two different parameter sets in (63) for the EKF and the truncated second order filter are shown. Table 2 shows the results for a model with large variations, whereas the parameters in Table 3 represents a model with less variation. In each table the results are listed in three columns. In the first column the mean of the estimated values are given. In the second column the $t$-statistics given by $\sqrt{n}(\bar{x} - x_{\text{sim}})/\sigma_x$ is stated, where $n$ is the number of simulated series, $x_{\text{sim}}$ is the true value and $\sigma_x$ is the empirical standard deviation of the estimated parameters. The $F$-statistics are listed in the third column. The $t$-values in Table 2 show that the four parameters in (63) are estimated consistently except for the long-term mean $\theta$, which is overestimated for both methods. With respect to $\sigma_e^2$ the transformation approach gives a biased estimate, contrary to the second order filter. This is most likely due to the highly nonlinear measurement equation caused by the transformation.

For the estimates in Table 3, the parameter $\alpha$ becomes biased for both methods. Similar simulation studies in (Baillie, 1996) show that this bias frequently occur when the data do not excite the model sufficiently well. However, due to the fact the variations in the data are small, the two diffusion parameters $\sigma^2$ and $\gamma$ become almost perfectly correlated, which obviously gives rise to some estimation

| Extended Kalman filter with transformation | | | | |
|---|---|---|---|---|
| Parameter | True values | Mean | $t$-value | $F$-value |
| $\alpha$ | 0.0250 | 0.0272 | 3.0053 | 0.8707 |
| $\theta$ | 10.0000 | 9.9778 | -0.6862 | 1.4721 |
| $\sigma^2$ | 0.0500 | 0.0712 | 2.1411 | 0.8816 |
| $\gamma$ | 0.5000 | 0.5332 | 0.9866 | 0.8210 |
| $\sigma_e^2$ | 0.0500 | 0.0518 | 4.4624 | 1.0029 |

| Truncated Second order filter | | | | |
|---|---|---|---|---|
| Parameter | True values | Mean | $t$-value | $F$-value |
| $\alpha$ | 0.0250 | 0.0272 | 2.9910 | 1.1234 |
| $\theta$ | 10.0000 | 9.9763 | -0.7212 | 0.6715 |
| $\sigma^2$ | 0.0500 | 0.0296 | 2.3626 | 0.6217 |
| $\gamma$ | 0.5000 | 0.5230 | 0.4348 | 0.7644 |
| $\sigma_e^2$ | 0.0500 | 0.0497 | -0.5570 | 1.0642 |

Table 3: Results for 50 sample sequences from the short rate model (63) with small variations.

problems. Again it is seen that the measurement noise becomes biased when the EKF is applied.

**REMARK 8.2.** The complicated relationship between the choice of sampling time and identification of parameters in the drift term in discretely, partially observed SDEs is largely an unsolved problem, see e.g. (Nielsen, Madsen and Young, 1999). ▼

### 8.2.2 An empirical study

A data set consisting of 2155 weekly observations[3] of the annualised yield of the 3-month US Treasury Bills covering the time period January 1954 to April 1995. No provisions are made for holiday and weekend effects. The sampling time is 1/50 corresponding roughly to weekly observations and annualised model parameters. The results in Table 4 show that the two methods give almost identical parameter estimates except for $\theta$. For $\frac{1}{2} < \gamma < 1$, $\theta$ is the long-term mean of $X_t$.

[3]The data was kindly provided by Jesper Lund. Any errors in the data set are our own responsibility.



| | $\alpha$ | $\theta$ | $\sigma^2$ | $10^4\sigma_e^2$ | $\gamma$ |
|---|---|---|---|---|---|
| EKF | 0.1183 (0.069) | 6.3098 (2.333) | 0.1136 (0.014) | 7.0173 (2.371) | 0.8486 (0.032) |
| TS | 0.1066 (0.070) | 6.9181 (2.955) | 0.1307 (0.013) | 6.8943 (2.290) | 0.8297 (0.025) |

Table 4: Estimation results of the 3-month U.S. Treasury-Bills yield with the Extended Kalman Filter (EKF) and the Truncated Second order filter (TS). The standard errors are shown in parenthesis.

The standard errors of the estimates of $\theta$ are rather large, see Remark 8.2, but the estimate suggests an annual mean interest rate of 6.3-6.9%. The estimates of $\gamma$ clearly suggest the presence of a level effect; a result that is consistent with other empirical findings[4].

It might be argued that a 40 years period covers different kinds of monetary regimes, and hence that the parameters in the model considered should be allowed to be time-dependent. For example in the period from October 1979 to October 1982 the Federal Reserve had a monetary targeting policy as opposed to an interest rate targeting policy, which caused an increase in the volatility. This topic is left for future research.

## 8.3 Stochastic volatility models

Another recent application of the nonlinear filtering approach is univariate *stochastic volatility models*,

$$dX_t = \alpha X_t dt + \sigma_t X_t dW_t^1; \quad X_{t_0} = X_0 \tag{76}$$

where $X_t$ denote the price of a stock at time $t$, $\alpha > 0$ is the *rate-of-return*, $\sigma_t$ is the *stochastic volatility* and $W_t^1$ is a Wiener process. The famous Black-Scholes model (Black and Scholes, 1973) is obtained for $\sigma_t = \sigma$, i.e. constant volatility. A univariate model with $\sigma_t = \sigma X_t^{\gamma-1}$ for which the transformation (63) is applicable is considered in (Nielsen, Vestergaard and Madsen, 1999). This and other studies show that a SDE should be specified for $\sigma_t$ in order to model the dynamics

[4]Recent research (Lind, 1997) has shown that the interpretation of the parameters in (63) is not as clear as suggested in e.g. (Chan et al., 1992), i.e. the interpretation of the parameter $\theta$ as the long-term mean actually depends on $\gamma$ in a complicated way.

of stock prices satisfactorily,

$$d\psi(\sigma_t) = a(\sigma_t)dt + b(\sigma_t)dW_t^2 \tag{77}$$

where $\psi(\sigma_t)$ is some mapping of $\sigma_t$, $a(\cdot)$ and $b^2(\sigma_t)$ is the instantaneous mean and variance, respectively, of the stochastic volatility (these functions should be identified from the data), and $(W_t^1, W_t^2)$ are correlated Wiener processes with correlation coefficient $\rho$. This generalization is considered in (Nielsen et al., 2000), see also the references therein. However the transformation proposed in Section 3 approach is not applicable if $\sigma_t$ is described by the SDE (77), see Remark 7.1.

# 9 Conclusion

An approximate maximum likelihood method for direct estimation of embedded parameters in some nonlinear stochastic differential equations from discrete input-output data encumbered with noise have been developed. For a limited class of SDEs with level effects a transformation is proposed such that first order filters like the IEKF may be used to obtain an approximate solution to the continuous-discrete filtering problem. Hence the numerical problems that are most often associated with higher order filters are avoided, but at the cost of a more complicated drift term and measurement equation. For the transformation to exist, some restrictions must be imposed on the diffusion term in the model specification. In addition to these restrictions, it is also recommendable to parameterize the diffusion term such that a likelihood ratio test may be carried out as a means of testing for a statistically significant level effect. The method is evaluated using statistical tests by considering a nonlinear predator-prey system and by using Monte Carlo simulation. It is stressed that the method has already been used for successful identification of a number of physical, financial and technical nonlinear systems.

# References


Aït-Sahalia, Y. (1998), Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach. Princeton University. Department of Economics. Working paper.

Åström, K. J. (1970), *Introduction to Stochastic Control Theory*, Academic Press, New York.





Baadsgaard, M., Nielsen, J. N., Spliid, H., Madsen, H. and Preisel, M. (1997), Estimation in stochastic differential equations with a state dependent diffusion term, *in* Y. Sawaragi and S. Sagara, eds, 'SYSID '97 - 11th IFAC Symposium on System Identification', IFAC.

Baillie, R. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics* **73**, 5–59.

Bigi, S., Söderström, T. and Carlsson, B. (1994), An IV-scheme for estimating continuous-time stochastic models from discrete-time data, *in* M. Blanke and T. Söderström, eds, 'SYSID '94 - 10th IFAC Symposium on System Identification', IFAC.

Black, F. and Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Politicial Economy* **81**, 637–654.

Bohlin, T. and Graebe, S. F. (1994), Issues in nonlinear stochastic grey-box identification, *in* M. Blanke and T. Söderström, eds, 'SYSID '94 - 10th IFAC Symposium on System Identification', IFAC, pp. 213–218.

Bohlin, T. and Graebe, S. F. (1995), 'Issues in nonlinear stochastic grey-box identification', *International Journal of Adaptive Control and Signal Processing* **9**(6), 465–490.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sanders, A. B. (1992), 'An empirical comparison of alternative models of the short-term interest rate', *Journal of Finance* **47**, 1209–1227.

Dennis, J. E. and Schnabel, R. B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, New Jersey.

Goodwin, G. C. and Payne, R. L. (1977), *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press.

Graebe, S. F. (1990*a*), IDKIT: A Software Gray Box Identification: Mathematical Reference, Technical report, TRITA-REG 90/03, Automatic Control, Royal Institute of Technology, Stockholm, Sweden.

Graebe, S. F. (1990*b*), IDKIT: A Software Guide for Gray Box Identification, User's Guide, Version RT-1.0, Technical report, TRITA-REG 90/04, Automatic Control, Royal Institute of Technology, Stockholm, Sweden.

Harvey, A. C. (1989), *Forecasting, Structural Models and the Kalman Filter*, Cambridge University Press, New York.

Haverkamp, B. R. J., Verhaegen, M., Chou, C. T. and Johansson, R. (1997), Continuous-time subspace model identification method using Laguerre filtering, *in* Y. Sawaragi and S. Sagara, eds, 'SYSID' 97 - 11th IFAC Symposium on System Identification', IFAC.

Jacobsen, J. L. and Madsen, H. (1996), 'Grey box modelling of oxygen levels in a small stream', *EnvironMetrics* **7**, 109–121.

Jacobsen, J. L., Madsen, H. and Harremoës, P. (1996), 'Modelling the transient impact of rain events on the oxygen content of a small creek', *Water, Science and Technology* **33**(2), 177–187.

Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Karatzas, I. and Shreve, S. E. (1996), *Brownian Motion and Stochastic Calculus, Second Edition*, Springer-Verlag, New York.

Klimko, L. and Nelson, P. (1978), 'On conditional least square estimation for stochastic processes', *Annals of Statistics* **6**, 629–642.

Kloeden, P. E. and Platen, E. (1995), *Numerical Solutions of Stochastic Differential Equations, Second Edition*, Springer-Verlag, Heidelberg.

Lind, K. H. (1997), Properties and Estimation of the CKLS-model of the Term Structure, Master's thesis, Dept. of Theoretical Statistics, Univ. of Copenhagen.

Ljung, L. (1979), 'Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems', *IEEE Transactions on Automatic Control* **24**(1), 36–50.

Ljung, L. (1987), *System Identification: Theory for the User*, Prentice-Hall, New York.

Madsen, H. and Holst, J. (1995), 'Estimation of continuous-time models for the heat dynamics of a building', *Energy and Buildings* **22**, 67–79.

Maybeck, P. S. (1982), *Stochastic Models, Estimation and Control*, Academic Press, London.



Melgaard, H. and Madsen, H. (1993), CTLSM version 2.6 - a program for parameter estimation in stochastic differential equations, Technical report, IMSOR. No. 1.

Moler, C. and van Loan, C. F. (1978), 'Nineteen dubious ways to compute the exponential of a matrix', *SIAM Review* **20**, 801–836.

Mortensen, R. E. (1969), 'Mathematical problems of modeling stochastic nonlinear dynamic systems', *J. Statist. Phys.* **1**, 271–296.

Nielsen, J. N. (1996), Nonlinear dynamics and time series analysis, Master's thesis, Department of Mathematical Modelling, Lyngby, Denmark.

Nielsen, J. N. and Madsen, H. (1996), Modeling heat dynamics using thermal networks, *in* 'System Identification Competition', Joint Research Centre, Ispra, Italy, chapter 13.

Nielsen, J. N., Madsen, H. and Young, P. (1999), Parameter estimation in stochastic differential equations: An overview, *in* 'Proceedings of The 14th IFAC World Congress', IFAC, Elsevier Science, pp. 289–294.

Nielsen, J. N., Vestergaard, M. and Madsen, H. (1999), Nonlinear filtering of univariate stochastic volatility models, *in* H.-F. Chen, D.-Z. Cheng and J.-F. Zhang, eds, 'Proceedings of the 14th IFAC World Congress', Vol. M, IFAC, Elsevier Science, pp. 123–128.

Nielsen, J. N., Vestergaard, M. and Madsen, H. (2000), 'Estimation in continuous-time stochastic volatility models using nonlinear filters', *International Journal of Theoretical and Applied Finance* **3**(2), 1–30.

Øksendal, B. (1995), *Stochastic Differential Equations, 4th Edition*, Springer-Verlag, Heidelberg.

Pedersen, A. R. (1994), Statistical analysis of gaussian diffusion processes based on incomplete discrete observations, Technical Report 297, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.

Protter, P. (1990), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

Pugachev, V. S. and Sinitsyn, I. N. (1987), *Stochastic Differential Systems – Analysis and Filtering*, John Wiley & Sons, New York.

Sagara, S. and Zhao, Z. Y. (1990), 'Numerical integration approach to on-line identification of continuous-time systems', *Automatica* **26**(1), 63–74.

Schmidt, M. R., Madsen, H., Zhou, J. and Hansen, L. H. (1997), Modelling of hydraulic robot, Technical report, Department of Mathematical Modelling, Technical University of Denmark.

Schweppe, F. (1965), 'Evaluation of likelihood function for Gaussian signals', *IEEE Transactions on Information Theory* **11**, 61–70.

Söderström, T. (1999), On the Cramér-Rao Lower Bound for Estimating Continuous-Time Autoregressive Parameters, *in* H.-F. Chen, D.-Z. Cheng and J.-F. Zhang, eds, 'Proceedings of the 14th IFAC World Congress', IFAC, Elsevier Science, pp. 175–180.

Söderström, T., Fan, H., Mossberg, M. and Carlsson, B. (1997), A bias-compensation scheme for estimating continuous time AR process parameters, *in* Y. Sawaragi and S. Sagara, eds, 'SYSID' 97 - 11th IFAC Symposium on System Identification', IFAC.

Söderström, T. and Stoica, P. (1989), *System Identification*, Prentice Hall, London.

Söderström, T., Stoica, P. and Friedlander, B. (1991), 'An indirect prediction error method for system identification', *Automatica* **27**(1), 183–188.

Tanizaki, H. (1996), *Nonlinear Filters, Second Edition*, Springer Verlag, Heidelberg.

Unbehauen, H. and Rao, G. P. (1987), *Identification of continuous systems*, North-Holland, Amsterdam.

Unbehauen, H. and Rao, G. P. (1990), 'Continuous-time approaches to system identification - a survey', *Automatica* **26**(1), 23–35.

Unbehauen, H. and Rao, G. P. (1997), Identification of continuous-time systems: A tutorial, *in* Y. Sawaragi and S. Sagara, eds, 'SYSID' 97 - 11th IFAC Symposium on System Identification', IFAC.

Wiberg, D. M. and DeWolf, D. G. (1993), 'A Convergent Approximation of the Continuous-Time Optimal Parameter Estimator', *IEEE Transactions on Automatic Control* **38**, 529–545.

Young, P. (1981), 'Parameter estimation for continuous-time models - a survey', *Automatica* **17**, 23–39.