Continuous-Discrete Unscented Kalman Filtering

Hermann Singer
FernUniversität in Hagen *

Diskussionsbeitrag
Fachbereich Wirtschaftswissenschaft Nr. 384
Februar 2006

Abstract

The unscented Kalman filter (UKF) is formulated for the continuous-discrete state space model. The exact moment equations are solved approximately by using the unscented transform (UT) and the measurement update is obtained by computing the normal correlation, again using the UT. In contrast to the usual treatment, the system and measurement noise sequences are included from the start and are not treated later by extension of the state vector. The performance of the UKF is compared to Taylor expansions (extended Kalman filter EKF, second and higher order nonlinear filter SNF, HNF), the Gaussian filter, and simulated Monte Carlo filters using a bimodal Ginzburg-Landau model and the chaotic Lorenz model.

Key Words: Stochastic differential equations; Nonlinear systems; Discrete measurements; Continuous-discrete state space model; Unscented Kalman filter; Extended Kalman filter; Gaussian filter; Monte Carlo filter.

1 Introduction

The nonlinear continuous-discrete state space model is most suitable for applications where the system is defined in continuous time, but the measure-

*Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany

ments are available only at certain time points (sampling). This fact has been leading to a strong prevalence of discrete time modeling. However, if the true dynamics are continuous, the sampled (discrete time) dynamics are very complicated because of the accumulated interactions between the measurement times. The continuous-discrete state space model (Jazwinski, 1970), which first appeared in the engineering literature, treats the dynamics between the measurements in continuous time, either by considering the conditional distribution or by using moment equations. At the measurements, the Bayes formula incorporates the new information leading to an updated a posteriori density. In the linear case, one only needs two moments and the involved densities are Gaussian. For nonlinear systems, the equations can be solved only approximately. A classical approach is the extended Kalman filter (EKF) and the second order nonlinear filter (SNF), where the linear formulas are kept by Taylor expansion of the drift and measurement functions around the estimates. On the other hand, the EKF is known to exhibit problems such as filter divergencies and suboptimal performance. The EKF and higher order nonlinear filters (HNF) have the drawback that the usual assumptions of the linear filter get lost (correlation of the error terms with the state and with each other; error terms are nongaussian and have nonzero expectations).
Another approach directs the focus on the approximation of the nongaussian filter densities (see, e.g. Tanizaki and Mariano, 1995). In the Monte Carlo approaches, density approximations are obtained by simulation, whereas the unscented Kalman filter (UKF) utilizes a singular density, concentrated at certain so called sigma points, chosen such, that the density reproduces exactly the first and second moment of the true density. Nonlinear expectation values as required in the filter equations are computed by weighted averaging over the sigma points, so Taylor expansions can be avoided. Alternatively, if the density is approximated by a Gauss distribution, one obtains the Gaussian filter (GF), which can be computed numerically by using Gauss-Hermite quadrature (GHF; cf. Ito and Xiong, 2000).
In the literature (cf. Julier and Uhlmann, 1997, 2004, Julier et al., 2000, Sitz et al., 2002a, b), the UKF is usually developed for deterministic systems and the noise sequences are added by extension of the state vector. However, since the noise cannot be influenced by the measurements, the measurement update must be changed in an ad hoc way. In this paper it is shown that the noise can be included from the start by using stochastic equations and no extension of the state is necessary.
In section 2 the continuous-discrete state space model is defined and the general nonlinear filter equations are stated. Furthermore, the exact moment equations are derived. Section 3 defines the density and the sigma points used in computing expectation values by the unscented transform (UT). The

results are compared to Taylor expansions of the nonlinear functions as utilized in the EKF. Section 4 derives the UKF equations and the corresponding EKF formulae. In section 5 the UKF is compared with several other filter algorithms (EKF, SNF, HNF, GF, functional integral filter FIF). The performance is evaluated by computing the distribution of the filtering error in a simulation study. Furthermore, Bayesian estimates of the parameters of the Lorenz model are computed.

2 State Space Model and Filter Equations

2.1 Nonlinear Continuous-Discrete State Space Model

The nonlinear *continuous-discrete state space model* is defined as (Jazwinski, 1970)

$$dy(t) = f(y(t), t, \psi)dt + g(y(t), t, \psi)dW(t) \tag{1}$$

where discrete measurements z_i are taken at times $\{t_0, t_1, \ldots, t_T\}$ and $t_0 \leq t \leq t_T$ according to the measurement equation

$$z_i = h(y(t_i), t_i, \psi) + \epsilon_i. \tag{2}$$

In state equation (1), $W(t)$ denotes an r-dimensional Wiener process and the state is described by the p-dimensional state vector $y(t)$. It fulfils a system of stochastic differential equations in the sense of Itô (cf. Arnold, 1974) with fixed initial condition $y(t_0)$. The functions $f : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \to \mathbb{R}^p$ and $g : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \to \mathbb{R}^p \times \mathbb{R}^r$ are called drift and diffusion coefficients, respectively. In measurement equation (2), $\epsilon_i \sim N(0, R(t_i, \psi))i.d.$ is a k-dimensional discrete time white noise process (measurement error). Parametric estimation is based on the u-dimensional parameter vector ψ. For notational simplicity, deterministic control variables $x(t)$ are absorbed in the time argument t.

2.2 Exact Continuous-Discrete Filter

The exact time and measurement updates of the continuous-discrete filter are given by the recursive scheme (Jazwinski, 1970) for the conditional density $p(y, t|Z^i)$:

Time update:

$$\begin{aligned} \frac{\partial p(y,t|Z^i)}{\partial t} &= F(y,t)p(y,t|Z^i)\ ; t \in [t_i, t_{i+1}] \\ p(y,t_i|Z^i) &:= p_{i|i} \end{aligned} \tag{3}$$

Measurement update:

$$\begin{aligned} p(y_{i+1}|Z^{i+1}) &= \frac{p(z_{i+1}|y_{i+1},Z^i)p(y_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} \quad (4)\\ &:= p_{i+1|i+1} \\ p(z_{i+1}|Z^i) &= \int p(z_{i+1}|y_{i+1},Z^i)p(y_{i+1}|Z^i)dy_{i+1}, \quad (5) \end{aligned}$$

$i = 0, \ldots, T-1$, where

$$\begin{aligned} Fp &= -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t,\psi)p(y,t|x,s)] \\ &\quad +\tfrac{1}{2}\sum_{ij}\frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t,\psi)p(y,t|x,s)] \quad (6) \end{aligned}$$

is the Fokker-Planck operator, $Z^i = \{z(t)|t \le t_i\}$ are the observations up to time t_i and $p(z_{i+1}|Z^i)$ is the likelihood function of observation z_{i+1}. The first equation describes the time evolution of the conditional density $p(y,t|Z^i)$ given information up to the last measurement and the measurement update is a discontinuous change due to new information using the Bayes formula. The above scheme is exact, but can be solved explicitly only for the linear case where the filter density is Gaussian with conditional moments

$$\begin{aligned} \mu(t|t_i) &= E[y(t)|Z^i] \quad (7)\\ \Sigma(t|t_i) &= \mathrm{Var}[y(t)|Z^i]. \quad (8) \end{aligned}$$

2.3 Moment Equations

In the EKF and UKF algorithms, instead of solving the time update for the conditional density, the moment equations for the first and second moments are solved approximately. Using the Euler approximation for the SDE (1), one obtains in a short time interval δt ($\delta W(t) := W(t+\delta t) - W(t)$)

$$y(t+\delta t) = y(t) + f(y(t),t,\psi)\delta t + g(y(t),t,\psi)\delta W(t). \quad (9)$$

Taking the expectation $E[...|Z^i]$ one gets the moment equation

$$\mu(t+\delta t|t_i) = \mu(t|t_i) + E[f(y(t),t,\psi)|Z^i]\delta t \quad (10)$$

or in the limit $\delta t \to 0$

$$\dot{\mu}(t|t_i) = E[f(y(t),t,\psi)|Z^i]. \quad (11)$$

The conditional variance fulfils

$$\begin{aligned} \mathrm{Var}[y(t+\delta t)|Z^i] &= \mathrm{Var}[y(t) + f(y(t),t,\psi)\delta t|Z^i] + \\ &+ E[\Omega(y(t),t,\psi)|Z^i]\delta t, \end{aligned} \tag{12}$$

since $y(t)$ is independent of $\delta W(t)$. In the limit $\delta t \to 0$ one obtains [1]

$$\dot{\Sigma}(t|t_i) = \mathrm{Cov}[f, y|Z^i] + \mathrm{Cov}[y, f|Z^i] + E[\Omega|Z^i]. \tag{13}$$

The moment equations (11, 13) are not differential equations, since they depend on the conditional density $p(y,t|Z^i)$. Expansion of the drift and diffusion functions leads to approximate differential equations (EKF, SNF), whereas an approximation of the density is used in the UKF (see section 4).

3 The Unscented Transform (UT)

The moment equations of the last section require the computation of expectations of the type $E[f(X)]$, where X is random variable with distribution $p(x)$. If $p(x)$ is known, the expectation can be computed using (numerical) integration. On the other hand, if we only know the moments μ and Σ, Taylor expansion of f yields

$$E[f(X)] \approx E[f(\mu) + f'(\mu)(X-\mu) + (1/2)f''(\mu)(X-\mu)^2] \tag{14}$$

$$= f(\mu) + (1/2)f''(\mu)\sigma^2. \tag{15}$$

This formula is the key tool in deriving the EKF and SNF.
In the UT method, the unknown density $p(x)$ is approximated by the singular density

$$p_{UT}(x) = \sum_{l=-n}^{n} \alpha_l \delta(x - x_l) \tag{16}$$

where $\delta(x)$ is the Dirac delta function and x_l are so called *sigma points*, chosen such, that the first and second moment of $p_{UT}(x)$ is correct. According to Julier et al. (2000) one can take the $2n+1$ points

$$x_0 = \mu \tag{17}$$

$$x_l = \mu + \sqrt{n+\kappa}\Gamma_{.l}, l = 1, ..., n \tag{18}$$

$$x_{-l} = \mu - \sqrt{n+\kappa}\Gamma_{.l}, l = 1, ..., n \tag{19}$$

[1] The update (12) is positive semidefinite and numerically more stable than a Euler version of (13). However, it contains terms of order $O(\delta t)^2$.

with weights

$$\alpha_0 = \kappa/(n+\kappa) \quad (20)$$

$$\alpha_l = 1/(2(n+\kappa)) = \alpha_{-l}, \quad (21)$$

where $\Gamma_{.l}$ is the lth column of the Cholesky root $\Sigma = \Gamma\Gamma'$, κ is a scaling factor and n is the dimension of vector X. In the filter, to display the dependence on the moments, the notation $x_l = x_l(\mu, \Sigma)$ will be used.
For example, in the univariate case $n = 1$ one obtains the three points $\mu, \mu \pm \sqrt{1+\kappa}\,\sigma$. Generally, the expectation $E_{UT}[X] = \int x p_{UT}(x)dx = \mu$ and

$$\mathrm{Var}_{UT}(X) = \int (x-\mu)(x-\mu)' p_{UT}(x)dx \quad (22)$$

$$= \sum_{l=-n}^{n} (x_l - \mu)(x_l - \mu)\alpha_l \quad (23)$$

$$= \sum_{l=1}^{n} \Gamma_{.l}(\Gamma_{.l})' = \Gamma\Gamma' = \Sigma \quad (24)$$

yields the correct first and second moment. Nonlinear expectations are easily evaluated as sums

$$E_{UT}[f(X)] = \int f(x) p_{UT}(x)dx \quad (25)$$

$$= \sum_{l} f(x_l)\alpha_l \quad (26)$$

$$= \frac{\kappa}{n+\kappa} f(\mu) + \frac{1}{2(n+\kappa)} \sum_{l=-n, n\neq 0}^{n} f(x_l) \quad (27)$$

Thus, for large κ, the Taylor formula $E[f(X)] \approx f(\mu)$ is retained. Indeed, for large κ, the UKF shows similar behavior as the EKF. The terms outside the mean approximate the higher order corrections in $E[f(X)]$. For example, in the case $n = 1, \kappa = 0$ one obtains

$$E_{UT}(f(X)) = 1/2[f(\mu+\sigma) + f(\mu-\sigma)] \approx f(\mu) + 1/2 f''(\mu)\sigma^2, \quad (28)$$

which coincides with the approximation used in the SNF. It is important, that the UT method does not require the computation of derivatives and the integration over the density is simplified by the delta functions into simple sums. The quality of the approximation depends on the nonlinear function and the unknown density $p(x)$, which may be skewed. The parameter κ also controls the 4th moment of the UT density.
Alternatively, one may assume that the true $p(x)$ is approximated by a Gaussian distribution $\phi(x; \mu, \sigma^2)$ with the same mean μ and variance σ^2. Then,

the Gaussian integral

$$E_\phi[f(X)] = \int f(x)\phi(x;\mu,\sigma^2)dx \tag{29}$$

$$\approx \sum_{l=1}^{m} f(x_l)w_l \tag{30}$$

may be approximated by Gauss-Hermite quadrature (cf. Ito and Xiong, 2000). If such an approximation is used, one obtains the Gauss-Hermite filter (GHF). The authors show, that the choice $\kappa = 2, n = 1$ in the UT corresponds to a Gauss-Hermite rule with $m = 3$ sample points. Generally, filters using Gaussian densities are called Gaussian filters (GF). Simulation of the Gaussian integral leads to the simulated moments Kalman filter (SMKF). Alternatively, one may expand f to higher orders L (higher order nonlinear filter HNF$(2, L)$) and factorize the moments according to the Gaussian assumption $m_l := E[X - \mu]^l = (l-1)!!\sigma^{2l}$ (l even) and $m_l = 0$ (l odd). This leads to an exact computation of (29) for $L \to \infty$. In this limit, the HNF and GF coincide. In the EKF=HNF(2,1) and SNF=HNF(2,2), the higher order corrections are neglected. Also, third and higher order moments could be used (HNF(K, L); cf. Singer, 2006).

4 The Unscented Kalman Filter UKF

4.1 Time Update

Using the moment equations and the formulas for nonlinear expectations of section 3, the time update can be computed step by step in the interval $[t_i, t_{i+1}]$. Chosing a discretization interval of δt and dividing into $J = (t_{i+1} - t_i)/\delta t$ steps, one computes the time updates (dropping ψ)

$$\mu(t+\delta t|t_i) = \mu(t|t_i) + E[f(y(t),t)|Z^i]\delta t \tag{31}$$

$$\begin{aligned}\Sigma(t+\delta t|t_i) &= \Sigma(t|t_i) + \\ &+ \{\mathrm{Cov}[f(y(t),t), y(t)|Z^i] + \\ &+ \mathrm{Cov}[y(t), f(y(t),t)|Z^i] + \\ &+ E[\Omega(y(t),t)|Z^i]\}\delta t.\end{aligned} \tag{32}$$

The expectation values are evaluated using the UT transform with the sigma points computed from the conditional moments $\mu(t|t_i), \Sigma(t|t_i)$. In the next step, the new moments $\mu(t+\delta t|t_i), \Sigma(t+\delta t|t_i)$ are utilized for new sigma points. The iteration starts at $\tau_0 = t_i$ with the a posteriori moments $\mu(t_i|t_i)$, $\Sigma(t_i|t_i)$ and ends at $\tau_J = t_{i+1}$ with the a priori moments $\mu(\tau_J|t_i)$, $\Sigma(\tau_J|t_i)$

where $\tau_j = t_i + j\delta t; j = 0, ..., J$. Since the sigma points are given by $\mu(\tau_j|t_i)$ and the Cholesky roots of $\Sigma(\tau_j|t_i)$, the equations (31–32) are a system of nonlinear differential equations for the sigma points.
In contrast, the EKF and SNF utilize a Taylor expansion of the functions f and Ω around the estimate $\mu(t|t_i)$ yielding another system of nonlinear differential equations. Alternatively, the expectation values could be evaluated by assuming that the filter density is approximately $N(\mu(t|t_i), \Sigma(t|t_i))$. The resulting Gaussian integrals can be computed by numerical integration (Gauss-Hermite filter; Ito and Xiong, 2000), or approximately by simulation, drawing data y_l from this normal distribution. The resulting density $p_{SMKF}(y) = 1/N \sum \delta(y - y_l)$ leads to the formula $E_{SMKF}[f(Y)] = 1/N \sum f(y_l)$. In contrast to the UT method, the weights are equal and the first two moments do not exactly coincide with the moments of the normal distribution. Since the moments are calculated by simulation, I call the approach simulated moments Kalman filter (SMKF).

4.2 Measurement Update

At time t_{i+1} with a priori moments $\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)$, new measurement information $z_{i+1} = z(t_{i+1})$ comes in. Exact incorporation uses the Bayes formula (2.2), which is difficult to implement. The optimal *linear* estimate of $\mu(t_{i+1}|t_{i+1})$, $\Sigma(t_{i+1}|t_{i+1})$ is given by the theorem on normal correlation (Liptser and Shiryayev, 1978, ch. 13–14, theorem 13.1, lemma 14.1)

$$\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \mathrm{Cov}(y_{i+1}, z_{i+1}|Z^i)\mathrm{Var}(z_{i+1}|Z^i)^- \times \\
&\times \quad (z_{i+1} - E[z_{i+1}|Z^i]) \qquad (33) \\
\Sigma(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \mathrm{Cov}(y_{i+1}, z_{i+1}|Z^i)\mathrm{Var}(z_{i+1}|Z^i)^- \times \\
&\times \quad \mathrm{Cov}(z_{i+1}, y_{i+1}|Z^i). \qquad (34)
\end{aligned}$$

Inserting the measurement equation (2) one obtains the a posteriori moments and the likelihood function of measurement z_{i+1} (abbreviating $y_{i+1} := y(t_{i+1}), h_{i+1} := h(y_{i+1}, t_{i+1})$ etc.)

$$\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \quad (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^-(z_{i+1} - E[h_{i+1}|Z^i]) \qquad (35) \\
\Sigma(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \quad (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^-\mathrm{Cov}(h_{i+1}, y_{i+1}|Z^i) \qquad (36) \\
L_{i+1} &= p(z_{i+1}|Z^i) \\
&= \phi(z_{i+1}; E[h_{i+1}|Z^i], \mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1})) \qquad (37)
\end{aligned}$$

where $\phi(.)$ is the Gaussian density. The expectations and covariances can again be computed using the unscented transform with sigma points from

$\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)$. Alternatively, Taylor expansion of h around $\mu(t_{i+1}|t_i)$ yields the usual EKF update. The GHF utilizes Gauss-Hermite integration with the a priori gaussian density as weight function. In the SMKF approach, the expectations are computed by simulation.
The likelihood L_{i+1} can be expressed by the innovation (prediction error) $\nu_{i+1} = z_{i+1} - E[z_{i+1}|Z^i] = z_{i+1} - E[h_{i+1}|Z^i]$ and the prediction error covariance $\Gamma_{i+1} = \mathrm{Var}[z_{i+1}|Z^i] = \mathrm{Var}[h_{i+1}|Z^i] + R(t_{i+1})$.

4.3 Continuous-discrete UKF

Summarizing, we obtain the UKF algorithm:

Initial condition: $t = t_0$

$$\begin{aligned}
\mu(t_0|t_0) &= \mu + \mathrm{Cov}(y_0, h_0) \times \\
&\times \ (\mathrm{Var}(h_0) + R(t_0))^-(z_0 - E[h_0]) \\
\Sigma(t_0|t_0) &= \Sigma - \mathrm{Cov}(y_0, h_0) \times \\
&\times \ (\mathrm{Var}(h_0) + R(t_0))^-\mathrm{Cov}(h_0, y_0) \\
L_0 &= \phi(z_0; E[h_0], \mathrm{Var}(h_0) + R(t_0)) \\
\text{sigma points} &: \ y_l = y_l(\mu, \Sigma); \mu = E[y_0], \Sigma = \mathrm{Var}(y_0).
\end{aligned}$$

$i = 0, \ldots, T-1$:
Time update: $t \in [t_i, t_{i+1}]$

$$\begin{aligned}
\tau_j &= t_i + j\delta t; j = 0, ..., J_i - 1 = (t_{i+1} - t_i)/\delta t - 1 \\
\mu(\tau_{j+1}|t_i) &= \mu(\tau_j|t_i) + E[f(y(\tau_j), \tau_j)|Z^i]\delta t \\
\Sigma(\tau_{j+1}|t_i) &= \Sigma(\tau_j|t_i) + \\
&+ \ \{\mathrm{Cov}[f(y(\tau_j), \tau_j), y(\tau_j)|Z^i] + \\
&+ \ \mathrm{Cov}[y(\tau_j), f(y(\tau_j), \tau_j)|Z^i] + \\
&+ \ E[\Omega(y(\tau_j), \tau_j)|Z^i]\}\delta t \\
\text{sigma points} &: \ y_l = y_l(\mu(\tau_j|t_i), \Sigma(\tau_j|t_i))
\end{aligned}$$

Measurement update: $t = t_{i+1}$

$$\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \ (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^-(z_{i+1} - E[h_{i+1}|Z^i]) \\
\Sigma(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \ (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^-\mathrm{Cov}(h_{i+1}, y_{i+1}|Z^i) \\
L_{i+1} &= \phi(z_{i+1}; E[h_{i+1}|Z^i], \mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1})) \\
\text{sigma points} &: \ y_l = y_l(\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)).
\end{aligned}$$

The discretization interval δt is a small value controlling the accuracy of the Euler scheme implicit in the time update. Since the sigma points are functions of the mean and variance, the moment equations (31–32) are a coupled system of nonlinear differential equations for the sigma points. Therefore, other approximation methods such as the Heun scheme or higher order Runge-Kutta schemes could be used.
The several expectation and filtering schemes are summarized in table 1.

4.4 Discussion

It is important to note that no sigma points for the error sequences $g(y(t), t, \psi) * \delta W(t)$ and ϵ_i are necessary, since their statistics are already included in the time and measurement update. This is only required if these errors are treated by extension of the state vector in a deterministic system. More seriously, the error sequences of the extended state are assumed to fulfil the equations (discrete time) $\epsilon_k = \epsilon_{k-1}, \eta_k = \eta_{k-1}$ (Sitz, et al., 2002b, equ. (9), Sitz, et al., 2002a, equ. (27)) This is correct for a constant parameter, but not for the error sequences which are white noise processes.
In this paper it is argued that the state extension is not necessary and avoids ad hoc assumptions regarding the measurement update (cf. Sitz et al., 2002a, sect. III D, Julier and Uhlmann, 1997). The time and measurement updates can be computed recursively using only the moments $\mu(t_i|t_i), \Sigma(t_i|t_i)$ and $\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)$ of the original state. Since no sigma points for the noises must be computed, the algorithm is more efficient.

5 Example: Ginzburg-Landau Model

The several algorithms are compared using a nonlinear system which strongly deviates from Gaussian behavior (cf. Miller et al., 1994, Singer, 2002). The Ginzburg-Landau model is a diffusion process in a double well potential $\Phi(y, \{\alpha, \beta\}) = \frac{\alpha}{2}y^2 + \frac{\beta}{4}y^4$ with vector field $f = -\partial\Phi/\partial y$ and state independent diffusion coefficient $g = \sigma$. The SDE reads explicitly (Ginzburg-Landau equation)

$$dy = -[\alpha y + \beta y^3]dt + \sigma dW(t) \tag{38}$$

with measurement equation

$$z_i = y_i + \epsilon_i, \tag{39}$$

$\mathrm{Var}(\epsilon_i) = R$. A physical picture is the strongly damped random movement of a sphere in a landscape defined by the potential Φ. The potential

Table 1: Overview of filtering schemes and abbreviations. $g(y)$ is an arbitrary function (e.g. $f(y)(y-\mu)$ etc.

Exact moment equations	$\dot{\mu} = E[f(y)]$ $\dot{\Sigma} = 2\mathrm{Cov}[f,y] + E[\Omega]$	scalar notation
method	expectation $E[g(y)]$	remarks
extended Kalman filter EKF	$g(\mu)$	= HNF(2,1)
second order nonlinear filter SNF	$g(\mu) + (1/2)g''(\mu)\Sigma$	= HNF(2,2)
higher order nonlinear filter HNF(K,L)	$\sum_{l=0}^{L}(1/l!)g^{(l)}(\mu)m_l$ $m_l := E[y-\mu]^l$	Gaussian factorization
Gaussian filter GF	$\int g(y)\phi(y;\mu,\Sigma)dy$	Gauss density GF=HNF$(2,\infty)$
Gauss-Hermite filter GHF	$\sum w_l g(y_l)$	Gauss-Hermite quadrature
simulated moments Kalman filter SMKF	$(1/N)\sum g(y_l)$	Monte Carlo
unscented Kalman filter UKF	$\sum \alpha_l g(y_l)$	sigma points $y_l(\mu,\Sigma)$
local linearization filter LLF	$f \approx f_i + f_i'(y-\mu_{i\|i})$ $+(1/2)f_i''\Omega_i(t-t_i)$	Ito formula for drift f $f_i := f(\mu_{i\|i})$
Gaussian sum filter GSF	$p(y) = \sum \alpha_l \phi(y;\mu_l,\Sigma_l)$	Gaussian mixture density
Functional integral filter FIF	$p(y) = \sum \alpha_l \phi(y;y_l,\Sigma_l)$	simulated trajectories y_l

Φ can exhibit a Hopf bifurcation when α becomes negative ($\beta > 0$). In the double wells, the transition density is Gaussian for short times, but for long time intervals a bimodal shape tending to the stationary density $p_0(y) = \lim_{t\to\infty} p(t, y|x, s) \propto \exp(-\frac{2}{\sigma^2}\Phi(y))$ occurs. Thus filters with conditional Gaussian densities will have degraded performance if the measurement intervals are large. Fig. 1 compares several approximate filter algorithms. The parameters are chosen as $\psi = \{\alpha, \beta, \sigma, R\} = \{-1, 0.1, 2, 1\}$. Best known is the EKF (first row, left). The predicted state tends to stay in the potential wells, but the true a priori distribution is bimodal with two maxima at the points $y_s = \pm\sqrt{-\alpha/\beta} = \pm\sqrt{10} = \pm 3.16228$. The EKF is based on linearization, so the state is either located in one of the wells. If the measurement is in the other well, the filter reacts only slowly, since this domain in state space has low weight in the predicted density. In contrast, the UKF prediction always tends to zero, which is the mean of the stationary distribution. Thus the reaction to the measurements is much better. The SNF also tends in prediction to the mean 0, but in the long first sampling interval a filter divergence occurs which is due to the second order correction $(1/2)f''(\mu)\sigma^2$. Higher order expansion in the variance equation avoids this divergence. For example, the HNF(2,3) uses Taylor expansion of f up to 3rd order, which is exact for the drift $-[\alpha y + \beta y^3]$. Thus the result is equal to the Gaussian filter (GF). It was implemented using numerical integration (Gauss-Hermite). The GF(3) is identical with the UKF, $\kappa = 2$, since as noted by Ito and Xiong (2000), the unscented transform coincides with the Gauss-Hermite formula using 3 sample points. Similar in performance to the EKF is the Gaussian sum filter GSF (Sorenson and Alspach, 1971, Anderson and Moore, 1979), where n EKFs run simultanously. [2] In the LLF (local linearization filter), as proposed by Shoji and Ozaki (1997, 1998), the linearization is made in the system equation by expanding the drift, using Ito's lemma (for a detailed comparision see Singer, 2002). The resulting equations are similar to the SNF, but in the present example no divergence occurs. The simulated moments Kalman filter (SMKF) is a variant of the GF, where the expectation values are computed by Monte Carlo samples drawn from $N(\mu, \Sigma)$ using the conditional moments from the filter recursion. Its performance is similar to the GF and the Monte Carlo filter FIF (functional integral filter), where the conditional densities are approximated by simulation of trajectories (cf. Singer, 2002, 2003). From fig. 1 it seems that UKF, HNF(2,3), GF, SMKF and FIF perform better than EKF, SNF, LLF and Gaussian sum filter. This was tested in a simulation study where $M = 100$ trajectories were simulated

[2] Ito and Xiong (2000) use a variant of the GSF where the updates are computed using Gaussian densities; the idea of the Gaussian filter is applied to Gaussian sums.

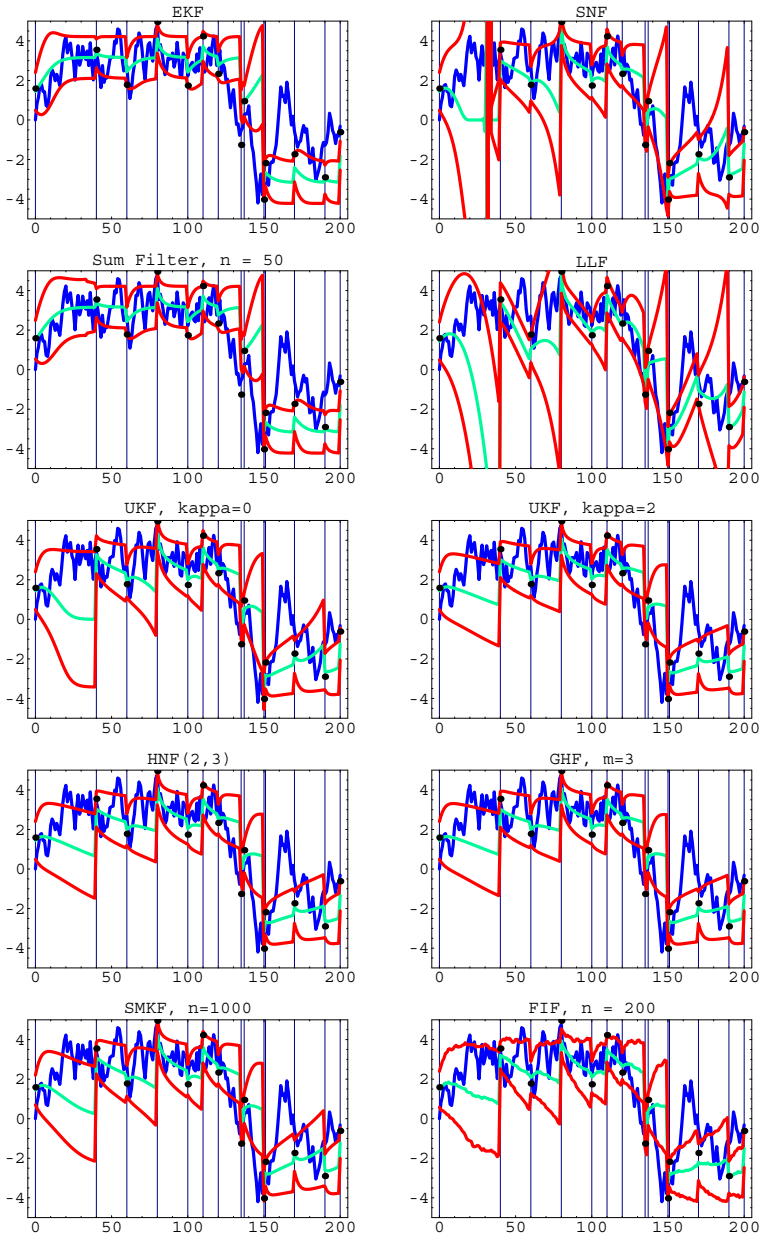

Figure 1: Diffusion process in a bimodal potential. Filtered estimates of $y(t)$ and 67% highest posterior density (HPD) confidence intervals using the EKF, SNF, GSF, LLF, UKF, HNF(2,3), GHF(3) and Monte Carlo filters (SMKF, FIF) (true parameters inserted, irregular sampling interval). See text.

Table 2: Distribution of filter error in $M = 100$ samples. The SNF was omitted because of overflows. $\bar{A}, \mathrm{std}(A)$: mean and standard deviation of squared filter error; $\bar{B}, \mathrm{std}(B)$: mean and standard deviation of averaged filter error (bias); $\bar{C}, \mathrm{std}(C)$: mean and standard deviation of filter error standard deviation;

method	$\bar{A}$	$\mathrm{std}(A)$	$\bar{B}$	$\mathrm{std}(B)$	$\bar{C}$	$\mathrm{std}(C)$
EKF	30.4303	9.29225	-0.0172362	0.851867	1.98594	0.605503
GSF, $n = 50$	29.2217	8.68137	-0.0199376	0.702071	1.94666	0.5835
LLF	30.8888	6.28692	0.0708409	0.844412	2.00486	0.45385
HNF(2,3)	24.4653	4.49457	0.0171205	0.628488	1.60677	0.318386
UKF, $\kappa = 0$	24.3465	3.62642	0.0228194	0.643741	1.58941	0.270704
UKF, $\kappa = 1$	24.3543	4.01249	0.0197056	0.632016	1.59548	0.292646
UKF, $\kappa = 2$	24.4874	4.56068	0.0168696	0.629528	1.60828	0.321831
UKF, $\kappa = 3$	24.7165	5.12368	0.0144765	0.635766	1.62544	0.351386
GHF, $m = 3$	24.4874	4.56068	0.0168696	0.629528	1.60828	0.321831
GHF, $m = 4$	24.4653	4.49457	0.0171205	0.628488	1.60677	0.318386
SMKF, $n = 1000$	24.4115	4.02016	0.0221391	0.651349	1.59226	0.291811
FIF, $n = 50$	24.6415	5.18325	0.126057	0.588081	1.63375	0.351731
FIF, $n = 100$	24.5591	5.03571	0.102591	0.591924	1.62759	0.341957
FIF, $n = 5000$	24.4935	5.14068	0.0172752	0.604748	1.6216	0.347258

and the filter errors were compared. The results are shown in table 2 and figure 2 In table 2, the entries are defined as:

- filter error: $\nu_t = y(t) - \hat{y}(t)$
- squared filter error: $A = \sum_t \nu_t^2$
- error mean : $B = 1/T \sum_t \nu_t$
- error standard deviation: $C = (1/T \sum_t (\nu_t - B)^2)^{0.5}$
- mean and standard deviation in M samples:
  $\bar{X} = 1/M \sum_m X_m$; $\mathrm{std}(X) = (1/M \sum_m (X_m - \bar{X})^2)^{0.5}$; $X = (A, B, C)$

The smallest squared filter error $\bar{A}$ is given by the UKF with $\kappa = 0$. Larger κ values lead to higher errors. The UKF(2) exactly coincides with GHF(3)

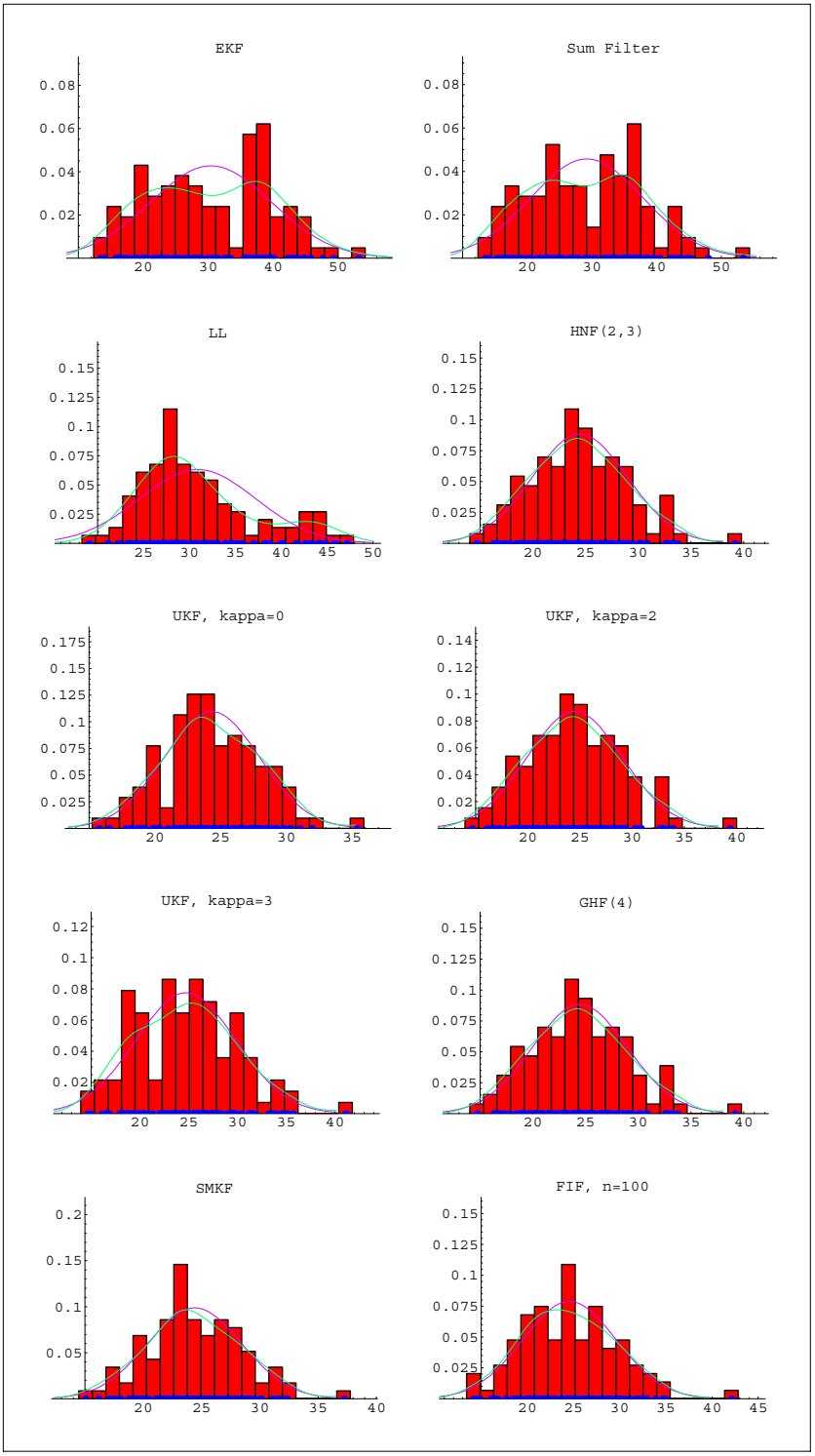

Figure 2: Distribution of squared filter error in $M = 100$ samples.

(cf. sect. 3). The HNF(2,3) coincides with GHF(4) [3] and is comparable to the UKF(2), but has smaller standard error in the error distribution. Generally, UKF, HNF(2,3) and GHF are much better than the EKF, LLF and the Gaussian sum filter. The simulated filters are comparable to the UKF performance. Inspection of the distributions in fig. 2 shows, that the EKF, GSF and LLF yield a bimodal error distribution corresponding to cases with low errors (trajectories stay in potential wells) and high errors (trajectories with frequent change). The UKFs, HNF(2,3), GHF and the Monte Carlo filters give error distributions similar to Gaussian shape.

6 Parameter estimation: the Lorenz model

In most applications, the models contain unknown parameters to be estimated from the available sampled data. The likelihood function, which is computed recursively by the filter, yields maximum likelihood (ML) estimates and asymptotic standard errors by numerically maximizing the likelihood and evaluting the negative Hessian matrix (observed Fisher information). In the context of nonlinear filter theory, a more convenient approach can be taken by considering the (fixed) parameters as latent states with trivial dynamics $d\psi = 0$, i.e. constant. Thus the filter runs with the extended state $\eta(t) = \{y(t), \psi(t)\}$ and yields conditional distributions $p(y(t), \psi(t)|Z^i)$. The filters considered in this paper produce approximate conditional expectations $E(y(t), \psi(t)|Z^i)$ and variances $\mathrm{Var}(y(t), \psi(t)|Z^i)$. Thus we obtain recursive estimates $\hat{\psi}(t) = E(\psi(t)|Z^i)$ of the parameter vector which is updated at the times of measurement. Also, recursive standard errors are obtained. The Bayesian estimates may be called expected a posteriori estimates (EAP). Maximization of the marginal conditional density $p(\psi(t)|Z^i)$ with diffuse a priori distribution $p(\psi(t_0))$ would give the ML estimator.
In this example the chaotic Lorenz model (Lorenz, 1963) is considered. It is the best known simple system exhibiting chaos and has been used for many purposes. Among them are the Navier-Stokes equations in meteorology (the original application), laser theory (Haken, 1977, Graham, 1989), psychology (Singer, 1992) and sociology (Troitsch, 1990). ML estimation has been considered by Ozaki et al. (2000) We assume that a vector of variables $Y(t) = [x, y, z](t)$ evolves according to the system of nonlinear stochastic

[3] HNF(2,3) contains derivatives up to f''' which is exact for the model. The GHF(4) is correct up to $O(y^7)$. Note that $\mathrm{Var}(y + fdt)$ contains terms $O(y^6)$. The UKF(2)=GHF(3) is only of order y^5.

Table 3: Lorenz model: mean and standard deviation of squared filter error $A = \sum_t \nu_t^2$ in $M = 100$ samples ($\Delta t = 0.2$).

method	x	y	z	σ	r	b
EKF	34.4278	49.0568	46.4022	15.968	12.9247	8.18301
(std)	2.35341	4.8163	4.38871	6.61447	4.10521	0.979441
SNF	34.4433	49.2547	46.4675	16.0031	13.7448	8.50263
(std)	2.36142	4.81487	4.41146	6.61724	4.56789	1.1919
HNF(2,2)	34.4388	49.2483	46.4605	15.9944	13.7155	8.50246
(std)	2.35904	4.80811	4.41011	6.59141	4.48563	1.17291
UKF, $\kappa = 0$	34.4355	49.2436	46.4566	15.9841	13.6924	8.50225
(std)	2.35775	4.80203	4.411	6.57232	4.41997	1.15926
FIF, $n = 10000$	35.3055	50.9326	48.7735	22.2944	20.5525	9.43772
(std)	2.81775	5.70727	7.53299	11.3771	10.4972	2.41376

differential equations

$$d\begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} -\sigma x(t) + \sigma y(t) \\ -x(t)z(t) + rx(t) - y(t) \\ x(t)y(t) - bz(t) \end{bmatrix} dt + \begin{bmatrix} g_1 dW_1(t) \\ g_2 dW_2(t) \\ g_3 dW_3(t) \end{bmatrix} \tag{40}$$

The usual choice of parameters $\{\sigma_0, r_0, b_0\} = \{10, 28, 8/3\}$ leading to chaotic trajectories was taken. The system error matrix is $G = \mathrm{diag}(4, 4, 4)$ and the measurement error covariance was taken as $R = \mathrm{diag}(0.1, 0.1, 0.1)$. The data were simulated using a discretization interval of $\delta t = 0.01$ using an Euler-Maruyama scheme and afterwards measured at regular sampling intervals $\Delta t = 0.2, 0.5$. In order to estimate the parameters, the state was extended with $\psi(t) = \{\sigma, r, b\}$ using the dynamics $d\psi = 0$ (constant parameters). The initial state $\eta(0) = \{Y(0), \psi(0)\}$ was assumed to be distributed as $N(\mu, \Sigma)$ with $E[\eta(0)] = \{0, 0, 0, \sigma, r, b\}$ and $\mathrm{Var}(\eta(0)) = \mathrm{diag}(10, 10, 10, 1, 1, 1)$, which reflects moderate a priori knowledge of the unknown parameters. In the filter solutions (figs. 3–4), the initial value $E[\eta(0)] = \{0, 0, 0, \sigma_0 + 1, r_0 + 1, b_0 + 1\}$ was used. As is shown in the figures, the filtered parameters converge to a vicinity of the true value $\{\sigma_0, r_0, b_0\} = \{10, 28, 8/3\}$ and the approximate 95% HPD confidence intervals $E(\psi(t)|Z^i) \pm 1.96[\mathrm{Var}(\psi(t)|Z^i)]^{0.5}; t_i \leq t$ get smaller with increasing sample size. The behavior of the several algorithms is similar. The simulated filter FIF needs many trajectories to fill the 6-dimensional phase space with enough sample points. The performance can be expressed by the squared filtering error $A = \sum_t \nu_t^2$, $\nu_t = y(t) - \hat{y}(t)$. Table 3 reports the means and standard deviations of the squared filter error in a simulation

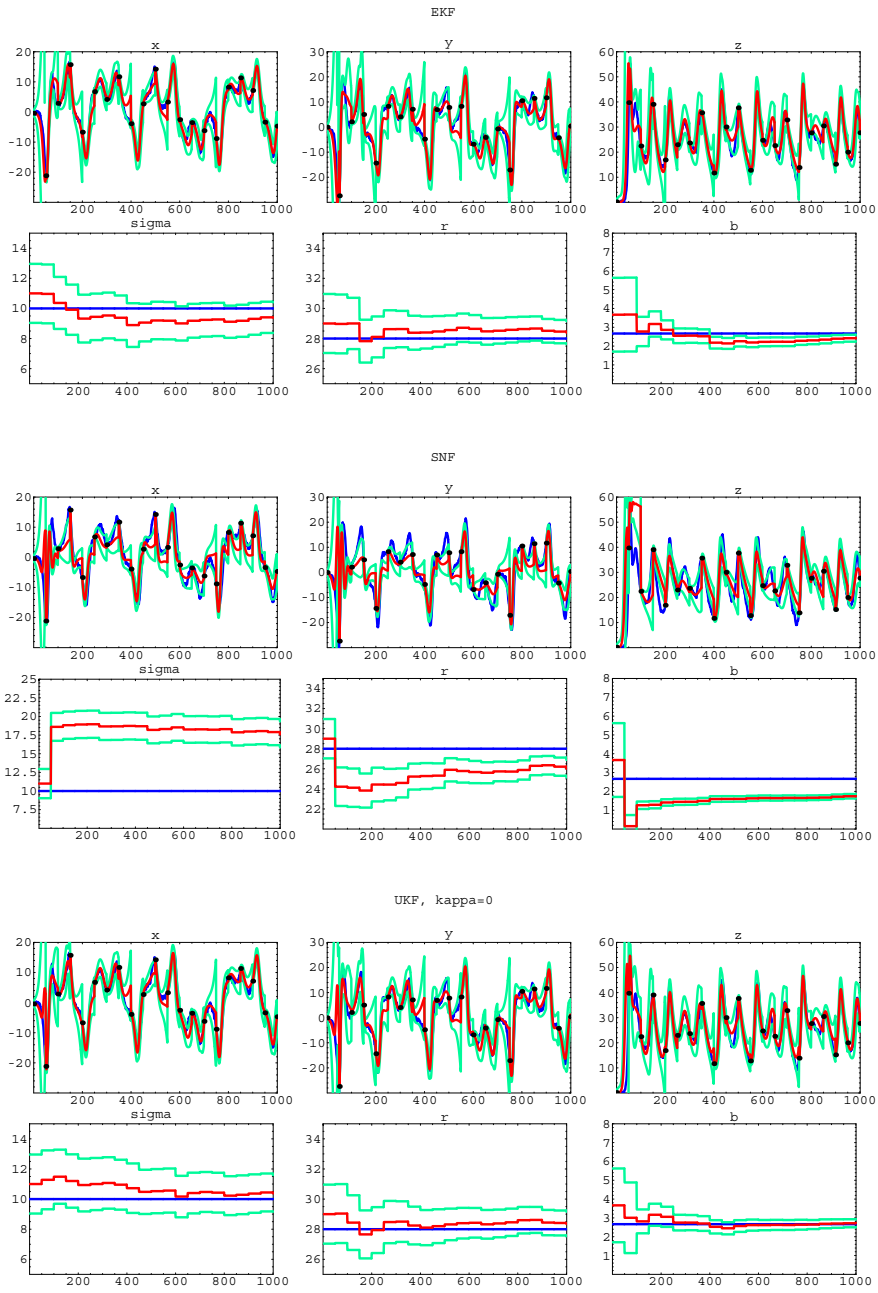

Figure 3: Filtered estimates of the Lorenz model. Comparision of several filtering algorithms. Components 4–6 are the unobserved parameters σ, r, b. Sampling interval $\Delta t = 0.5$.

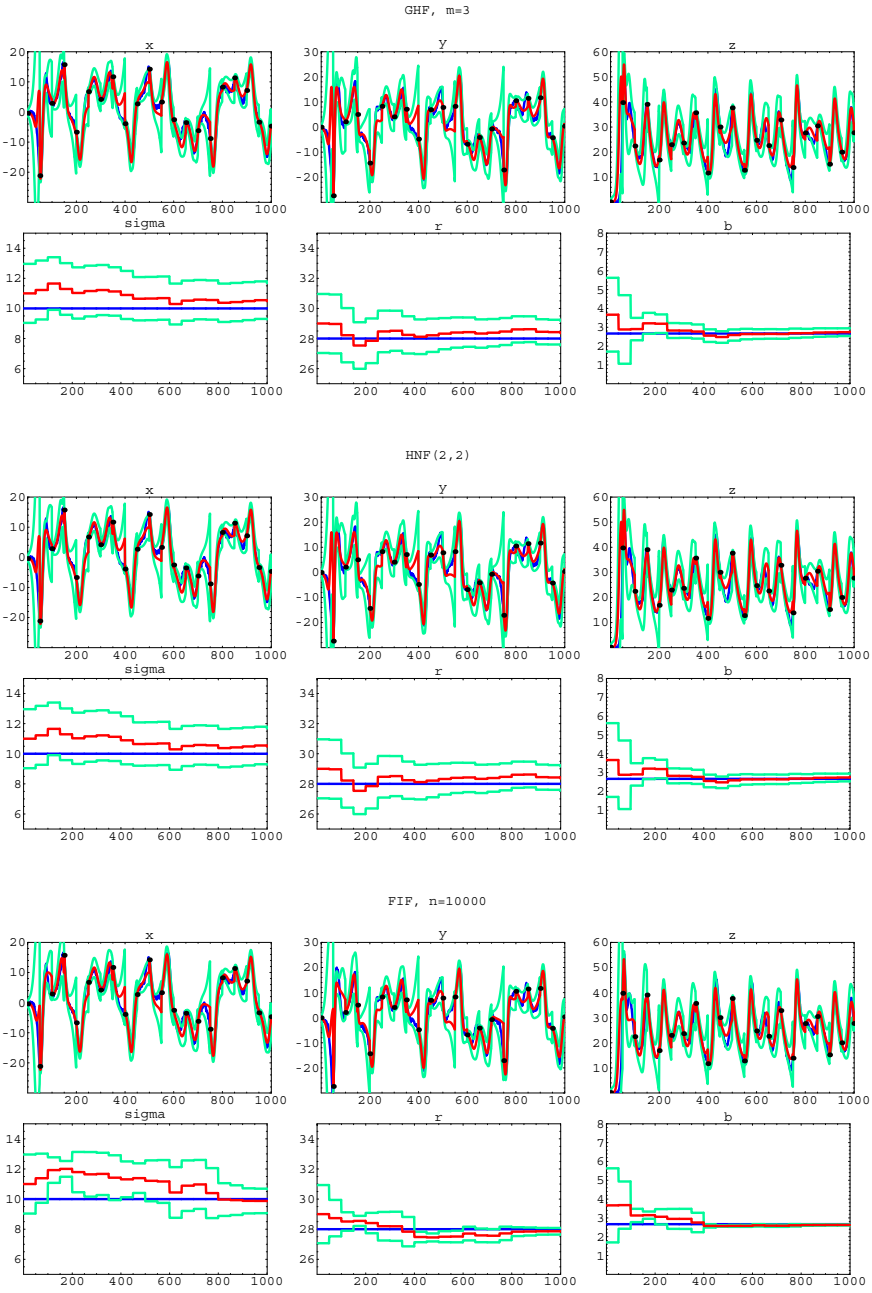

Figure 4: Filtered estimates of the Lorenz model. Comparision of several filtering algorithms. Components 4–6 are the unobserved parameters σ, r, b. Sampling interval $\Delta t = 0.5$.

study with $M = 100$ samples ($\Delta t = 0.2$). Surprisingly, the performance of the EKF is better than the UKF ($\kappa = 0$) and the SNF. The UKF and the SNF are very similar, which corresponds to the discussion at the end of section 3. The FIF ($n = 1000$) shows degraded performance due to simulation errors. Using $n = 10000$ trajectories leads to better results, but still a larger sample size would be necessary. The quality of the parameter estimates for ψ was investigated by comparing their mean and standard deviation as well as bias and root mean square error (RMSE) in $M = 100$ samples. The final estimates $\hat{\psi}(T) = E[\psi(T)|Z^T]; T = 1000$ were extracted from the filter solution and $\bar{\psi} = 1/M \sum \psi_m, s^2 = 1/M \sum (\psi_m - \bar{\psi})^2, \text{bias} = \bar{\psi} - \psi; \text{RMSE} = [s^2 + \text{bias}^2]^{1/2}$ was computed. Table 4 shows the results. In terms of RMSE, SNF, HNF(2,2) and UKF are very similar and slightly better than the EKF. Also, the bias of the estimates is very small. The empirical distributions of the estimates only slightly deviate from normality. The FIF with $n = 1000$ contains considerable statistical error and needs much larger sample size. Using $n = 60000, 100000$ does improve the results, but at huge computational costs.

In the preceeding example the sampling interval $\Delta t = 0.2$ was fairly small. In this case the deviation of the transition density from the normal distribution is only slight and the performance of the filters is similar. In contrast, table 5 displays results for a larger interval $\Delta t = 0.5$. In this case, filter divergencies occured, especially for the EKF, where only $M^* = 3$ trajectories could be filtered without tuning. In case of very large values (KMAX = 5000) in either the filtered state or covariance, these were reset to 0. In contrast, for the SNF $M^* = 97$ [4] and for the HNF(2,2), UKF and FIF all $M = 100$ trajectories worked without divergencies. Since the vector field is of quadratic order, the HNF(2,2) is equivalent to the GHF. Since the state space is 6-dimensional, the GHF(3) is much slower than HNF(2,2) or UKF. For example, the UKF takes $2p + 1 = 13$ points, whereas the GHF(3) needs $m^p = 3^6 = 729$ sample points. In table 5, the results are compared. For the EKF and SNF, the estimates are strongly biased with large RMSE. In contrast, the HNF(2,2)=GHF(3), UKF and the FIF are well behaved in all cases and show the best results. For the FIF, a sample size of $n = 60000$ or more is necessary to compete with the HNF(2,2) or UKF. However, the algorithm needs much more computing resources.

[4] A version without second order terms $O(\delta t^2)$ in (12) was used, so SNF and HNF(2,2) do not exactly coincide.

Table 4: Lorenz model: conditional mean parameter estimation in $M = 100$ samples. Mean, standard deviation, bias and root mean square error (RMSE) of the final estimate $\hat{\psi}(T) = E[\psi(T)|Z^T]; T = 1000$. Sampling interval $\Delta t = 0.2$.

parameter	true value	mean	std	bias	RMSE
EKF					
r	10.	10.0985	0.341217	0.0984689	0.355141
σ	28.	27.924	0.185297	-0.0759964	0.200276
b	2.66667	2.63736	0.0553276	-0.0293024	0.0626082
SNF					
r	10.	10.1073	0.342805	0.107307	0.359208
σ	28.	28.0051	0.185974	0.00505853	0.186042
b	2.66667	2.67788	0.0551185	0.0112098	0.0562469
HNF(2,2)					
r	10.	10.1069	0.342334	0.106884	0.358632
σ	28.	28.0049	0.18567	0.00487781	0.185734
b	2.66667	2.67796	0.0551133	0.0112907	0.0562579
UKF, $\kappa = 0$					
r	10.	10.1063	0.342057	0.106319	0.358199
σ	28.	28.0047	0.185489	0.00473882	0.18555
b	2.66667	2.67796	0.05511	0.0112894	0.0562544
FIF, $n = 10000$					
r	10.	10.5086	0.647432	0.508642	0.823337
σ	28.	28.4093	0.532197	0.409264	0.671364
b	2.66667	2.70047	0.168087	0.0338062	0.171453
FIF, $n = 60000$					
r	10.	10.1958	0.531189	0.195756	0.566112
σ	28.	28.1318	0.281452	0.131808	0.310787
b	2.66667	2.68683	0.0959574	0.020163	0.0980529
FIF, $n = 100000$					
r	10.,	10.1569	0.472255	0.156942	0.497651
σ	28.	28.0895	0.280081	0.0895371	0.294045
b	2.66667	2.68317	0.0910559	0.0165082	0.0925402

Table 5: Lorenz model: conditional mean parameter estimation in $M = 100$ samples. Mean, standard deviation, bias and root mean square error (RMSE) of the final estimate $\hat{\psi}(T) = E[\psi(T)|Z^T]; T = 1000$. Sampling interval $\Delta t = 0.5$. Last column: Number M^* of used samples without divergencies.

parameter	true value	mean	std	bias	RMSE	M^*
EKF						
r	10.	9.5079	3.95473	-0.492102	3.98523	
σ	28.	26.0415	0.665537	-1.95845	2.06845	3
b	2.66667	1.67255	0.573535	-0.994121	1.1477	
SNF						
r	10.	11.2398	3.62334	1.23982	3.82958	
σ	28.	28.1427	1.90359	0.142697	1.90893	97
b	2.66667	2.82491	0.964601	0.158248	0.977495	
HNF(2,2)						
r	10.	10.3772	0.643647	0.377195	0.746028	
σ	28.	28.2142	0.543558	0.214183	0.584235	100
b	2.66667	2.74046	0.19355	0.0737935	0.207141	
UKF, $\kappa = 0$						
r	10.	10.3315	0.725237	0.331523	0.797419	
σ	28.	28.2332	0.531429	0.233211	0.580349	100
b	2.66667	2.73632	0.181055	0.0696499	0.19399	
FIF, $n = 10000$						
r	10.	10.6329	0.844636	0.632905	1.05545	
σ	28.	28.4276	0.706537	0.427564	0.825836	100
b	2.66667	2.81738	0.318927	0.150716	0.352746	
FIF, $n = 60000$						
r	10.	10.4855	0.645767	0.485523	0.807928	
σ	28.	28.2869	0.530557	0.286886	0.603154	100
b	2.66667	2.74084	0.169687	0.074177	0.185192	
FIF, $n = 100000$						
r	10.	10.4142	0.582437	0.41418	0.714687	
σ	28.	28.1789	0.468461	0.178907	0.501462	100
b	2.66667	2.7399	0.156482	0.0732379	0.172772	

7 Stochastic Volatility

From the preceeding examples it seems, that UKF, HNF or GHF are always the best algorithms, with the FIF as the only competitor (but with high computational demand). The stochastic volatility model is interesting in this respect, since it contains a latent variable acting on the diffusion term. It is an alternative to the well known ARCH specification, where the volatility can be expressed in terms of measured states. Here, the state space model

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sigma(t)S(t)dW(t) && (41)\\ d\sigma(t) &= \lambda[\sigma(t) - \bar{\sigma}]dt + \gamma dV(t) && (42)\\ z_i &= S(t_i), && (43) \end{aligned}$$

$\mathrm{Var}(dW, dV) = 0$ (Scott, 1987, Hull and White, 1987) contains an unobservable volatility process $\sigma(t)$. The data are simulated daily (365 days), but measured on a weekly sampling interval. $\{\mu, \lambda, \bar{\sigma}, \gamma\} = \{.07, -1., .2, .2\}$. The model accounts for the fact, that the returns

$$r(t) = dS/S = \mu dt + \sigma(t)dW(t) \quad (44)$$

on financial time series can exhibit a time dependent variance and leptokurtosis of the return distribution, as is found empirically (cf. Kim et al., 1998, Fridman and Harris, 1998). Fig. 5 shows a comparision of the EKF, SNF, UKF and FIF ($n = 5000$). The SNF uses a second order correction to the diffusion term in the moment equation (13), i.e. $E[\Omega|Z^i]$ is expanded to order $\Omega(y) = \Omega(\mu) + \frac{1}{2}\Omega(\mu)'' * (y-\mu)(y-\mu)'; y = [S, \sigma]'$ (cf. Singer, 2002). Despite the correction, the SNF (and EKF and UKF) are not able to filter the latent volatility correctly, whereas the FIF tracks the volatility fluctuations.

8 Conclusion

For models with parameters in the drift term the unscented Kalman filter UKF is a stable and efficient algorithm both for state and parameter estimation. The Gaussian filter GHF shows similar performance, but is computationally more demanding for higher state space dimensions, since the UKF uses $2p + 1$ sample points, whereas the GHF utilizes m^p points. The higher order nonlinear filter HNF$(2, L)$ is equivalent to the GHF, if the order L of the Taylor series is high enough. The SNF may suffer from filter divergencies, especially if the sampling intervals get large or for strongly nonlinear models. This also applies to the EKF, which diverged for the Lorenz model in most cases ($\Delta t = 0.5$). The FIF is computationally much more demanding

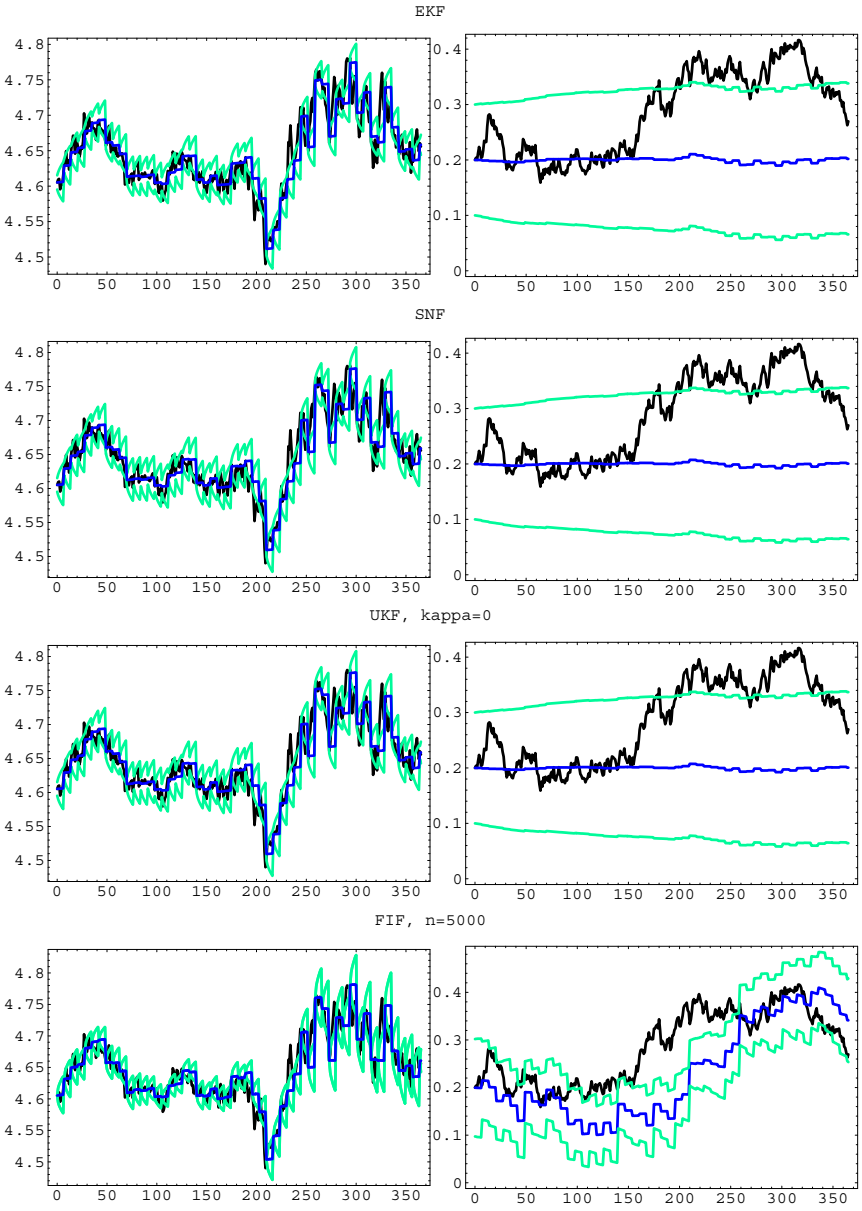

Figure 5: Stochastic volatility model. Left: True and filtered log stock price and HPD confidence interval. Right: True and filtered volatility and HPD interval.

with about equal or better performance for large sample size n, but for the stochastic volatility model, it shows clear advantages over EKF, SNF, GHF and UKF.
In conclusion, for the models investigated, the UKF was a well behaved, efficient algorithm but did not always work satisfactorily.

Appendix: Discrete Time UKF

For reference, the formulas for the discrete time UKF are summarized. Again, the system and measurement error sequences are considered from the start and are not appended later (cf. section 4.4). x_i are deterministic control variables ($x_i = i$ for explicit time dependence).
State space model:

$$y_{i+1} = f(y_i, x_i, \psi) + g(y_i, x_i, \psi)u_i \quad (45)$$
$$z_i = h(y_i, x_i, \psi) + \epsilon_i, \quad (46)$$

$u_i \sim N(0,1)\, i.i.d., \epsilon_i \sim N(0, R_i)\, i.d.$ and independent of u_j.
The UKF is given by the iteration (abbreviation $h_i = h(y_i, x_i, \psi); \mu_{i|i} = E[y_i|Z^i], \Sigma_{i|i} = \mathrm{Var}[y_i|Z^i]$, etc.):

Unscented Kalman Filter (discrete time)
Initial condition: $i = 0$

$$\begin{aligned}
\mu_{0|0} &= \mu + \mathrm{Cov}(y_0, h_0)(\mathrm{Var}(h_0) + R_0)^-(z_0 - E[h_0]) \\
\Sigma_{0|0} &= \Sigma - \mathrm{Cov}(y_0, h_0)(\mathrm{Var}(h_0) + R_0)^-\mathrm{Cov}(h_0, y_0) \\
L_0 &= \phi(z_0; E[h_0], \mathrm{Var}(h_0) + R_0) \\
\text{sigma points} &: y_l = y_l(\mu, \Sigma); \mu = E[y_0], \Sigma = \mathrm{Var}(y_0).
\end{aligned}$$

$i = 0, \ldots, T-1$:
Time update:

$$\begin{aligned}
\mu_{i+1|i} &= E[f(y_i, x_i)|Z^i] \\
\Sigma_{i+1|i} &= \mathrm{Var}(f(y_i, x_i)|Z^i) + E[\Omega(y_i, x_i)|Z^i] \\
\text{sigma points} &: y_l = y_l(\mu_{i|i}, \Sigma_{i|i})
\end{aligned}$$

Measurement update:

$$\begin{aligned}
\mu_{i+1|i+1} &= \mu_{i+1|i} + \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times (\mathrm{Var}(h_{i+1}|Z^i) + R_{i+1})^-(z_{i+1} - E[h_{i+1}|Z^i]) \\
\Sigma_{i+1|i+1} &= \Sigma_{i+1|i} - \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times (\mathrm{Var}(h_{i+1}|Z^i) + R_{i+1})^-\mathrm{Cov}(h_{i+1}, y_{i+1}|Z^i) \\
L_{i+1} &= \phi(z_{i+1}; E[h_{i+1}|Z^i], \mathrm{Var}(h_{i+1}|Z^i) + R_{i+1}) \\
\text{sigma points} &: y_l = y_l(\mu_{i+1|i}, \Sigma_{i+1|i}).
\end{aligned}$$

Setting $f \to y + f\delta t, g \to g\sqrt{\delta t}$ and dropping measurement updates without measurement one obtains the continuous-discrete filter (Euler approximation).

References

[1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, 1979.

[2] L. Arnold. *Stochastic Differential Equations*. John Wiley, New York, 1974.

[3] M. Fridman and L. Harris. A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models. *Journal of the American Statistical Association*, 16, 3:284–291, 1998.

[4] R. Graham. Macroscopic potentials, bifurcations and noise. In F. Moss and P.V.E. McClintock, editors, *Noise in nonlinear dynamical systems, vol. I*. Cambridge University Press, 1989.

[5] H. Haken. *Synergetics*. Springer, Berlin, 1977.

[6] J. Hull and A. White. The Pricing of Options with Stochastic Volatilities. *Journal of Finance*, XLII,2:281–300, 1987.

[7] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45, 5:910–927, 2000.

[8] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[9] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Control*, Orlando, Florida, 1997.

[10] S. Julier and J. Uhlmann. Unscented Kalman Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92, 3:401–422, 2004.

[11] S. Julier, J. Uhlmann, and H. F. Durrant-White. A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators. *IEEE Transactions on Automatic Control*, 45, 3:477–482, 2000.

[12] S. Kim, N. Shephard, and S. Chib. Stochastic Volatility: Likelihood Inference and Comparision with ARCH Models. *Review of Economic Studies*, 45:361–393, 1998.

[13] R.S. Liptser and A.N. Shiryayev. *Statistics of Random Processes, Volumes I and II.* Springer, New York, Heidelberg, Berlin, 1977, 1978.

[14] E. Lorenz. Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, 20:130, 1963.

[15] R.N. Miller, M. Ghil, and F. Gauthiez. Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems. *Journal of the Atmospheric Sciences*, 51, 8:1037–1056, 1994.

[16] T. Ozaki, J.C. Jimenez, and V. Haggan-Ozaki. The Role of the Likelihood Function in the Estimation of Chaos Models. *Journal of Time Series Analysis*, 21, 4:363–387, 2000.

[17] L.O. Scott. Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis*, 22:419–438, 1987.

[18] I. Shoji and T. Ozaki. Comparative Study of Estimation Methods for Continuous Time Stochastic Processes. *Journal of Time Series Analysis*, 18, 5:485–506, 1997.

[19] I. Shoji and T. Ozaki. A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika*, 85, 1:240–243, 1998.

[20] H. Singer. Dynamic structural equations in discrete and continuous time. In G. Haag, U. Mueller, and K.G. Troitzsch, editors, *Economic Evolution and Demographic Change*, pages 306–320. Springer, Berlin, Heidelberg, New York, 1992b.

[21] H. Singer. Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood vs. Extended Kalman Filter and Itô-Taylor Expansion. *Journal of Computational and Graphical Statistics*, 11,4:972–995, 2002.

[22] H. Singer. Simulated Maximum Likelihood in Nonlinear Continuous-Discrete State Space Models: Importance Sampling by Approximate Smoothing. *Computational Statistics*, 18,1:79–106, 2003.

[23] H. Singer. Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models. *Computational Statistics*, 21, 3, 2006. in print.

[24] A. Sitz, U. Schwarz, and J. Kurths. The Unscented Kalman Filter, A Powerful Tool For Data Analysis. *International Journal of Bifurcation and Chaos*, 14, 6:2093–2105, 2002b.

[25] A. Sitz, U. Schwarz, J. Kurths, and H.U. Voss. Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Physical Review E*, 66:016210–1–016210–9, 2002a.

[26] H.W. Sorenson and D.L. Alspach. Recursive Bayesian estimation using Gaussian sums. *Automatica*, 7:465–479, 1971.

[27] H. Tanizaki and R.S. Mariano. Prediction, Filtering and Smoothing in Nonlinear and Non-normal Cases using Monte-Carlo Integration. In H.K. Van Dijk, A. Monfort, and B.W. Brown, editors, *Econometric Inference using Simulation Techniques*, pages 245–261. John Wiley, 1995.

[28] K. G. Troitzsch. *Modellbildung und Simulation in den Sozialwissenschaften*. Westdeutscher Verlag, Opladen, 1990.