J Stat Phys (2009) 134: 365–380
DOI 10.1007/s10955-008-9673-1

The Diffusion Kernel Filter

Paul Krause

Received: 2 July 2008 / Accepted: 19 December 2008 / Published online: 15 January 2009

Abstract A particle filter method is presented for the discrete-time filtering problem with nonlinear Itô stochastic ordinary differential equations (SODE) with additive noise supposed to be analytically integrable as a function of the underlying vector-Wiener process and time. The Diffusion Kernel Filter is arrived at by a parametrization of small noise-driven state fluctuations within branches of prediction and a local use of this parametrization in the Bootstrap Filter. The method applies for small noise and short prediction steps. With explicit numerical integrators, the operations count in the Diffusion Kernel Filter is shown to be smaller than in the Bootstrap Filter whenever the initial state for the prediction step has sufficiently few moments. The established parametrization is a dual-formula for the analysis of sensitivity to gaussian-initial perturbations and the analysis of sensitivity to noise-perturbations, in deterministic models, showing in particular how the stability of a deterministic dynamics is modeled by noise on short times and how the diffusion matrix of an SODE should be modeled (i.e. defined) for a gaussian-initial deterministic problem to be cast into an SODE problem. From it, a novel definition of prediction may be proposed that coincides with the deterministic path within the branch of prediction whose information entropy at the end of the prediction step is closest to the average information entropy over all branches. Tests are made with the Lorenz-63 equations, showing good results both for the filter and the definition of prediction.

Keywords Data assimilation · Filtering · Particle filters · Diffusion kernel filter · Sensitivity analysis · Prediction

1 Introduction

Due to the lack of precise information in some fronts (e.g. physical processes, initial state or parameters) and the loss of information in other fronts (e.g. any form of dimension reduction through modeling and numerical handling), prediction of the future is a stochastic initial

P. Krause (✉)
Department of Atmospheric Sciences,
USP, São Paulo, 05508-090, Brazil
e-mail: krause@model.iag.usp.br

value problem by nature. In some contexts, Data Assimilation aims at turning feasible the sensitive prediction problems (e.g. [14]), at least for short times, by way of providing the initial states and parameters with probability measures and extracting estimates from these [1, 10, 12].

Considered is the nonlinear Itô[1] SODE (stochastic ordinary differential equation) problem with additive noise

$$dx = f(x)dt + g(t, w)dw, \quad t \in [t_0, t_\mathrm{N}], \tag{1.1}$$

$$x(t_0) = x_0, \quad \text{independent of } w(t - t_0) \text{ for all } t \geq t_0, \tag{1.2}$$

and the stochastic process (data function)

$$y_t = h(x(t), t) + z_t, \tag{1.3}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ (drift function) and $g : \mathbb{R}^n \to M(\mathbb{R}, n, p)$ (diffusion matrix) are continuously differentiable and slowly increasing functions; $w : [0, +\infty) \times \mathrm{C}_0([0, +\infty), \mathbb{R}^p) \longrightarrow \mathbb{R}^p$, $w = w(s)$, is a standard vector-Wiener process with independent components, wherein $\mathrm{C}_0([0, +\infty), \mathbb{R}^p)$ is the space of continuous functions from $[0, +\infty)$ to $\mathbb{R}^p$ whose initial value is 0; $h : \mathbb{R}^n \times [t_0, t_\mathrm{N}] \to \mathbb{R}^m$ (observation function) is a continuous function; z_t is a given noise process onto $\mathbb{R}^m$, independent of x_0 and $w(t - t_0)$ for all $t \geq t_0$ [11, 13]. Determining the sequence of random states $\Phi(x_0; t_k) \mid (y_{t_1} = y_1, \ldots, y_{t_k} = y_k)$, $k = 1, \ldots, \mathrm{N}$, where $\Phi(x_0; t)$ is the solution of problem (1.1)–(1.2) and $y_k \in \mathbb{R}^m$ is an outcome of y_t drawn at time t_k, is an instance of Discrete-Time Data Assimilation problem: the Discrete-Time Filtering problem with (1.1). Particle Filters are sample-based numerical methods for the discrete-time filtering problem [4, 5]. These methods suffer from two major shortcomings, reflecting nonlinearity: large operations count and troubles defining prediction[2]. This work introduces a particle filter method for the discrete-time filtering problem with (1.1), whose noise term is supposed to be analytically integrable as a function of the underlying vector-Wiener process and time to keep from sampling its history, along with a suitable definition of prediction. The method, to be called Diffusion Kernel Filter, applies for small noise and short prediction steps. It is arrived at by a parametrization of small noise-driven state fluctuations within "branches of prediction" and a local use of it in the Bootstrap Filter (namely, sampling the end-points of branches of prediction). The established formula reads as the stochastic integral of a diffusion kernel or the accumulated system noise mapped through the deterministic propagator of initial perturbations. It is derived by a reformulation of problem (1.1)–(1.2) into a Liouville SPDE (stochastic partial differential equation) problem, use of Duhamel's principle in weak form, splitting of a term with a projection operator and its complement, restriction of the resulting problem to an open nonlinear SODE problem for the flow of a branch of prediction, and closure of the latter problem. This was inspired from [2, 3], where a similar technique is used to tackle the dimension reduction problem for the random dynamics of a non-gaussian-initial nonlinear ODE (ordinary differential equation).

In Sect. 2, the Bootstrap Filter is described; in Sect. 3, the Diffusion Kernel Filter is derived; in Sect. 4, results obtained with the Lorenz-63 equations are presented. Throughout the text, Fréchet and weak derivatives are handled formally and dw/dt treated like a distribution (i.e. generalized function); the symbol $:=$ stands for a definition and D for a derivative with respect to the underlying independent state variable.

[1] Some colored noise processes can be treated in this framework (cf. [8], Sect. 4.8).

[2] The average estimate is a bad choice when the phase space is an embedded manifold (e.g. a set of shock or thermodynamical profiles).

2 The Bootstrap Filter

Definition 2.1 Let $x_{k,\lambda}$ be the filtered state at time t_k, $k = 1, \dots, \mathrm{N}$, that is,

$$x_{k,\lambda} := \Phi(x_0; t_k) \mid (y_{t_1,\lambda} = y_1, \dots, y_{t_k,\lambda} = y_k),$$

where $\Phi(x_0; t)$ is the solution of problem (1.1)–(1.2).

Consider the nonlinear Itô SODE problem with multiplicative[3] noise

$$dx = f(x)dt + g(x)dw, \quad t \in [t_k, t_{k+1}], \tag{2.1}$$

$$x(t_k) = x_{k,\lambda}, \tag{2.2}$$

and let $\Phi(x_{k,\lambda}; t)$ be its solution.

From

$$\begin{aligned}\Phi(x_{k,\lambda}; t) &= \Phi(\Phi(x_0; t_k) \mid (y_{t_1,\lambda} = y_1, \dots, y_{t_k,\lambda} = y_k); t)\\ &= \Phi(x_0; t) \mid (y_{t_1,\lambda} = y_1, \dots, y_{t_k,\lambda} = y_k), \quad t \in [t_k, t_{k+1}],\end{aligned}$$

one obtains

$$x_{k+1,\lambda} = \Phi(x_{k,\lambda}; t_{k+1}) \mid (y_{t_{k+1},\lambda} = y_{k+1}).$$

Thus

$$x_{k+1,\lambda} \sim p(x|y_{k+1}) = \frac{p(y_{k+1}|x)p(x)}{p(y_{k+1})}, \tag{2.3}$$

for $(\Phi(x_{k,\lambda}; t_{k+1}), y_{t_{k+1},\lambda}) \sim p(x, y)$.

Algorithm 2.1 (Bootstrap Filter [6]) Let $x_{k,i}$, $i = 1, \dots, I_k$, be distinct samples of $x_{k,\lambda}$; let $J_{k,i}$ be the number of times sample $x_{k,i}$ is repeated, $\sum_{i=1}^{I_k} J_{k,i} = I$.

Consider the nonlinear Itô SODE problem

$$dx = f(x)dt + g(x)dw, \quad t \in [t_k, t_{k+1}], \tag{2.4}$$

$$x(t_k) = x_{k,i} \in \mathbb{R}^n, \tag{2.5}$$

and let $\Phi(x_{k,i}; t)$ be its solution.

Steps:

(1) For each $i = 1, \dots, I_k$, solve problem (2.4)–(2.5) numerically for $J_{k,i}$ paths of $\Phi(x_{k,i}; t)$, up to time $t = t_{k+1}$, so as to sample from $p(x)$. Let $x_{(i,j)}$, $i = 1, \dots, I_k$, $j = 1, \dots, J_{k,i}$, be the samples thus obtained.

(2) Draw I samples from the discrete random variable (i, j) with mass function

$$p_{k+1,(i,j)} := p(y_{k+1}|x_{(i,j)}) \Big/ \sum_{\iota,\jmath} p(y_{k+1}|x_{(\iota,\jmath)}).$$

Let (i', j') be these samples without repetition and $J_{(i',j')}$ the number of times they are repeated. Then $\{x_{(i',j')},$ plus repetitions$\}$ is a set of samples of $x_{k+1,\lambda}$ [15].

[3] The B.F. applies to this framework.

Definition 2.2 In Step (1), call each set of $J_{k,i}$ paths, for each i, a branch of prediction.

By induction on k, upon relabeling the sets $\{x_{(i',j')}\}$ and $\{J_{(i',j')}\}$ as $\{x_{k+1,i}\}$ and $\{J_{k+1,i}\}$, $i = 1, \dots, I_{k+1}$, one establishes a recursive algorithm for sampling

$$\Phi(x_0; t_k) \mid (y_{t_1,\lambda} = y_1, \dots, y_{t_k,\lambda} = y_k), \quad k = 1, \dots, \mathrm{N}.$$

Remark 2.1 Proofs and alternatives to the branching Step (2) are presented in [4].

3 The Diffusion Kernel Filter

Consider the nonlinear Itô SODE with multiplicative[4] noise

$$\frac{d}{dt}x = f(x) + g(x)\zeta, \quad t \in [t_k, t_{k+1}], \tag{3.1}$$

where $\zeta := dw/dt$, and let $\Phi(x_k; t)$, $x_k \in \mathbb{R}^n$, be its dynamics.

3.1 A Linear SPDE Problem for $\Phi(x_k; t)$

Let

$$\begin{aligned} \mathrm{X}(x_k; t) &:= \frac{\partial}{\partial t}\Phi(x_k; t) - ((f(x_k) + g(x_k)\zeta) \cdot \mathrm{D})\Phi(x_k; t), \\ &= \frac{\partial}{\partial t}\Phi(x_k; t) - \mathrm{D}\Phi(x_k; t)(f(x_k) + g(x_k)\zeta). \end{aligned} \tag{3.2}$$

From the definition of $\Phi(x_k; t)$ and (3.2), upon using $dg(t, w)\zeta/dt = 0$ one obtains the linear RPDE (random partial differential equation) problem

$$\begin{aligned} &\frac{\partial}{\partial t}\mathrm{X}(x_k; t) = \mathrm{D}f(x)\big|_{x=\Phi(x_k;t)}\mathrm{X}(x_k; t), \\ &\mathrm{X}(x_k; t_k) = 0. \end{aligned}$$

Thus $\mathrm{X}(x_k; t) = 0$ for all $t \in [t_k, t_{k+1}]$ and $x_k \in \mathbb{R}^n$, that is,

$$\frac{\partial}{\partial t}\Phi(x_k; t) = \mathrm{L}(x_k)\Phi(x_k; t),$$

where

$$\mathrm{L}(x_k) := ((f(x_k) + g(x_k)\zeta) \cdot \mathrm{D}). \tag{3.3}$$

Therefore, the dynamics of (3.1) solves the Liouville SPDE problem

$$\frac{\partial}{\partial t}\Phi(x_k; t) = \mathrm{L}(x_k)\Phi(x_k; t), \tag{3.4}$$

[4]Though only meant for additive noise, the D.K.F. is derived under this framework with the use of semi-groups.

$$\Phi(x_k; t_k) = x_k, \tag{3.5}$$

which may be written

$$\frac{\partial}{\partial t}\Phi(x_k; t) = e^{(t-t_k)\mathrm{L}}\mathrm{L}x_k, \tag{3.6}$$

$$\Phi(x_k; t_k) = x_k, \tag{3.7}$$

where $e^{(t-t_k)\mathrm{L}}$ is the semigroup associated to (3.4).

More generally, writing

$$\frac{\partial}{\partial t}u(\Phi(x_k; t)) = \mathrm{D}u(x)\big|_{x=\Phi(x_k;t)}\frac{\partial}{\partial t}\Phi(x_k; t),$$

one obtains the following.

$\Phi(x_k; t)$ is the dynamics of (3.1) if, and only if, $u(\Phi(x_k; t))$ solves the Liouville SPDE problem

$$\frac{\partial}{\partial t}u(\Phi(x_k; t)) = \mathrm{L}(x_k)u(\Phi(x_k; t)), \tag{3.8}$$

$$u(\Phi(x_k; t_k)) = u(x_k), \tag{3.9}$$

for any $u = u(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ sufficiently smooth.

In particular, for $g = 0$, one obtains that $\phi(x_k; t)$, $x_k \in \mathbb{R}^n$, is the dynamics of the nonlinear ODE

$$\frac{d}{dt}x = f(x), \quad t \in [t_k, t_{k+1}], \tag{3.10}$$

if, and only if, $u(\phi(x_k; t))$ solves the Liouville PDE (partial differential equation) problem

$$\frac{\partial}{\partial t}u(\phi(x_k; t)) = (f(x_k)\cdot\mathrm{D})u(\phi(x_k; t)), \tag{3.11}$$

$$u(\phi(x_k; t_k)) = u(x_k), \tag{3.12}$$

for any $u = u(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ sufficiently smooth.

For details, see [2], Chap. 6.

3.2 A Nonlinear SODE Open Problem for $\Phi(x_{k,i}; t)$

The use of Duhamel's formula coming next differs considerably from [2, 3].

Definition 3.1 Let P be the conditional expectation $\mathrm{E}[\cdot|x_k]$ and $\mathrm{Q} := \mathrm{I} - \mathrm{P}$ its complement, with x_k (marginally) distributed according to the Dirac measure centered at some sample $x_{k,i}$ of $x_{k,\lambda}$.

Take $\mathrm{A} = \mathrm{PL}$ and $\mathrm{B} = \mathrm{QL}$ in Duhamel's formula

$$e^{(t-t_k)(\mathrm{A}+\mathrm{B})} = e^{(t-t_k)\mathrm{A}} + \int_{t_k}^{t} e^{(t-s)(\mathrm{A}+\mathrm{B})}\mathrm{B}e^{(s-t_k)\mathrm{A}}ds$$

and apply both sides onto $\mathrm{PL}x_k$. One gets

$$e^{(t-t_k)\mathrm{L}}\mathrm{PL}x_k = e^{(t-t_k)\mathrm{PL}}\mathrm{PL}x_k + \int_{t_k}^{t} e^{(t-s)\mathrm{L}}\mathrm{QL}e^{(s-t_k)\mathrm{PL}}\mathrm{PL}x_k ds. \quad (3.13)$$

Remark 3.1 Formula (3.13) is known for $g = 0$ ([9], Chap. 9). Proving its extension to $g \neq 0$ is a good subject for research.

In (3.6), write

$$e^{(t-t_k)\mathrm{L}}\mathrm{L}x_k = e^{(t-t_k)\mathrm{L}}\mathrm{PL}x_k + e^{(t-t_k)\mathrm{L}}\mathrm{QL}x_k. \quad (3.14)$$

Then, upon applying formula (3.13) and Definition 3.1, problem (3.6)–(3.7) reformulates into

$$\frac{d}{dt}\Phi(x_{k,i};t) = e^{(t-t_k)\mathrm{PL}}\mathrm{PL}x_k + \int_{t_k}^{t} e^{(t-s)\mathrm{L}}\mathrm{QL}e^{(s-t_k)\mathrm{PL}}\mathrm{PL}x_k ds + e^{(t-t_k)\mathrm{L}}\mathrm{QL}x_k, \quad (3.15)$$

$$\Phi(x_{k,i};t_k) = x_{k,i}. \quad (3.16)$$

To find how $e^{(t-t_k)\mathrm{PL}}$ acts on $\mathrm{PL}x_k = f(x_{k,i})$, apply P onto each member of (3.8)–(3.9). One obtains

$$\frac{\partial}{\partial t}\mathrm{P}u(\Phi(x_k;t)) = (f(x_{k,i})\cdot \mathrm{D})\mathrm{P}u(\Phi(x_k;t)) = \mathrm{PL}(x_k)\mathrm{P}u(\Phi(x_k;t)), \quad (3.17)$$

$$\mathrm{P}u(\Phi(x_k;t_k)) = u(x_{k,i}). \quad (3.18)$$

Taking $u \equiv f$, one concludes that

$$e^{(t-t_k)\mathrm{PL}}\mathrm{PL}x_k = \mathrm{P}f(\Phi(x_k;t)),$$

if f is sufficiently smooth.

Thus, the nonlinear SODE open problem (3.15)–(3.16) reads

$$\frac{d}{dt}\Phi(x_{k,i};t) = \mathrm{P}f(\Phi(x_k;t)) + \int_{t_k}^{t} e^{(t-s)\mathrm{L}}\mathrm{QLP}f(\Phi(x_k;s))ds + e^{(t-t_k)\mathrm{L}}\mathrm{QL}x_k,$$

$$\Phi(x_{k,i};t_k) = x_{k,i},$$

which, upon substituting L, applying Q, then applying $e^{(t-t_k)\mathrm{L}}$, further reads

$$\frac{d}{dt}\Phi(x_{k,i};t) = \mathrm{P}f(\Phi(x_k;t)) + \int_{t_k}^{t} e^{(t-s)\mathrm{L}}(g(x_{k,i})\zeta\cdot \mathrm{D})\mathrm{P}f(\Phi(x_k;s))ds + g(\Phi(x_{k,i};t))\zeta, \quad (3.19)$$

$$\Phi(x_{k,i};t_k) = x_{k,i}. \quad (3.20)$$

3.3 A Closure for Problem (3.19)–(3.20)

For small fluctuations of $\Phi(x_{k,i};\cdot)$ over $[t_k,t]$ the following applies.

Assumption 3.1

$$\mathrm{P} f(\Phi(x_k;\cdot)) = f(\mathrm{P}\Phi(x_k;\cdot)). \tag{3.21}$$

Assumption 3.1 implies $\mathrm{P}\Phi(x_k;t) = \phi(x_{k,i};t)$ (apply P to the original equation for $\Phi(x_k;t)$). Thus, for small fluctuations of $\Phi(x_{k,i};\cdot)$ over $[t_k,t]$ problem (3.19)–(3.20) may be written

$$\frac{d}{dt}\Phi(x_{k,i};t) = f(\phi(x_{k,i};t)) + \int_{t_k}^{t} e^{(t-s)\mathrm{L}}(g(x_{k,i})\zeta\cdot \mathrm{D}) f(\phi(x_{k,i};s))ds + g(\Phi(x_{k,i};t))\zeta, \tag{3.22}$$

$$\Phi(x_{k,i};t_k) = x_{k,i}. \tag{3.23}$$

Definition 3.2 Let $\mathcal{L}^2_\tau(n,p)$ be the vector space of $M(\mathbb{R},n,p)$-valued stochastic processes $A(s)=(A_{ij}(s))$ over the time interval $\tau := [t_k,t]$ such that, for every i,j, $\mathrm{E}(|A_{ij}(s)|^2) < +\infty$ for all $s\in\tau$, and $\int_{t_k}^{t}\mathrm{E}(|A_{ij}(s)|^2)ds < +\infty$.

Definition 3.3 Call

$$G(x_{k,i};t,s) := \mathrm{D}\phi(x_{k,i};t)g(\Phi(x_{k,i};s)), \quad t\in[t_k,t_{k+1}],\ s\in[t_k,t], \tag{3.24}$$

the diffusion kernel in a branch of prediction of (3.1). This is a $M(\mathbb{R},n,p)$-valued stochastic process in s.

For $G(x_{k,i};t,\cdot)\in\mathcal{L}^2_\tau(n,p)$, one has

$$C(x_{k,i};t) = \int_{t_k}^{t}\mathrm{E}(G(x_{k,i};t,s)G^*(x_{k,i};t,s))ds, \tag{3.25}$$

where

$$C(x_{k,i};t) := \mathrm{Cov}\left(\int_{t_k}^{t} G(x_{k,i};t,s)dw(s-t_k)\right) \tag{3.26}$$

and G^* refers to the conjugate transpose of G. This follows from the following properties[5] of an Itô stochastic integral with respect to a standard vector-Wiener process with independent components (cf. [13] or [11], Chap. 3):

$$\mathrm{E}\left(\int_{t_k}^{t} G_{\imath,l}(t,s)dw_l(s-t_k)\right) = 0, \quad \imath=1,\dots,n,\ l=1,\dots,p,$$

$$\mathrm{E}\left(\int_{t_k}^{t} G_{\imath,l_1}(t,s)dw_{l_1}(s-t_k)\int_{t_k}^{t} G_{\jmath,l_2}(t,s)dw_{l_2}(s-t_k)\right) = 0,$$

$$\imath,\jmath=1,\dots,n, \quad l_1,l_2=1,\dots,p,\ l_1\neq l_2,$$

$$\mathrm{E}\left(\int_{t_k}^{t} G_{\imath,l}(t,s)dw_l(s-t_k)\int_{t_k}^{t} G_{\jmath,l}(t,s)dw_l(s-t_k)\right) = \int_{t_k}^{t}\mathrm{E}(G_{\imath,l}(t,s)G_{\jmath,l}(t,s))ds, \quad \imath,\jmath=1,\dots,n,\ l=1,\dots,p,$$

where $G(t,s)$ stands for $G(x_{k,i};t,s)$.

[5] The third one is known as *Itô isometry formula*.

Definition 3.4 Let $\mathcal{L}_\tau^2(1, p)$ be embodied with the inner product $(u, v) := \int_{t_k}^t \mathrm{E}(u(s)v^*(s))ds$ and the norm $\|u\|_2 := \sqrt{(u, u)}$ associated to it. This is a Hilbert space (cf. [11], Chap. 3, Lemma 3.2.1).

Under this notation, (3.25) reads

$$C_{\imath,\jmath}(x_{k,i}; t) = (G_\imath(x_{k,i}; t, \cdot), G_\jmath(x_{k,i}; t, \cdot)), \quad \imath, \jmath = 1, \ldots, n, \tag{3.27}$$

where G_λ refers to a row of G, which implies

$$|C_{\imath,\jmath}(x_{k,i}; t)| \le \|G_\imath(x_{k,i}; t, \cdot)\|_2 \|G_\jmath(x_{k,i}; t, \cdot)\|_2, \quad \imath, \jmath = 1, \ldots, n, \tag{3.28}$$

under the Cauchy-Schwarz inequality, so that

$$\|C(x_{k,i}; t)\|_\infty \le (\max_{\imath=1,\ldots,n} \|G_\imath(x_{k,i}; t, \cdot)\|_2)^2. \tag{3.29}$$

Definition 3.5 Let $\mathcal{L}_\tau^2(n, p)$ be embodied with the norm $\|A\| := \max_{i=1,\ldots,n} \|A_i\|_2$, where A_λ refers to a row of A.

Under this notation, inequality (3.29) reads

$$\|C(x_{k,i}; t)\|_\infty \le \|G(x_{k,i}; t, \cdot)\|^2. \tag{3.30}$$

Therefore, $\mathrm{Cov}(\int_{t_k}^t G(x_{k,i}; t, s)dw(s - t_k))$ is small whenever $G(x_{k,i}; t, \cdot)$ is small in $(\mathcal{L}_\tau^2(n, p), \|\cdot\|)$.

For small fluctuations of $\Phi(x_{k,i}; \cdot)$ over $[t_k, t]$ the following applies.

Assumption 3.2

$$e^{(t-s)\mathrm{L}} f(\phi(x_{k,i}; s)) \simeq f(\phi(x_{k,i}; t)). \tag{3.31}$$

Upon applying (3.31), writing

$$\begin{aligned}(g(\Phi(x_{k,i}; t_k + t - s))\zeta \cdot \mathrm{D}) f(\phi(x_{k,i}; t)) &= \mathrm{D} f(\phi(x_{k,i}; t)) g(\Phi(x_{k,i}; t_k + t - s))\zeta \\ &= \mathrm{D} f(x)|_{x=\phi(x_{k,i};t)} G(x_{k,i}; t, t_k + t - s)\zeta,\end{aligned}$$

then writing

$$\begin{aligned}\int_{t_k}^t G(x_{k,i}; t, t_k + t - s)\zeta ds &= \int_{t_k}^t G(x_{k,i}; t, s)\zeta ds \\ &= \int_{t_k}^t G(x_{k,i}; t, s) dw(s - t_k),\end{aligned}$$

problem (3.22)–(3.23) closes into

$$\begin{aligned}\frac{d}{dt}\Phi(x_{k,i}; t) = f(\phi(x_{k,i}; t)) + \mathrm{D} f(x)\big|_{x=\phi(x_{k,i};t)} \left(\int_{t_k}^t G(x_{k,i}; t, s) dw(s - t_k)\right) \\ + g(\Phi(x_{k,i}; t))\zeta, \quad \Phi(x_{k,i}; t_k) = x_{k,i},\end{aligned} \tag{3.32}$$

$$\frac{d}{dt}\phi(x_{k,i};t) = f(\phi(x_{k,i};t)), \quad \phi(x_{k,i};t_k) = x_{k,i}, \tag{3.33}$$

$$\frac{d}{dt}\mathrm{D}\phi(x_{k,i};t) = \mathrm{D}f(x)\big|_{x=\phi(x_{k,i};t)}\mathrm{D}\phi(x_{k,i};t), \quad \mathrm{D}\phi(x_{k,i};t_k) = \mathrm{I}(n,n), \tag{3.34}$$

where $\mathrm{D}\phi(x_{k,i};t)$ is the deterministic propagator of initial perturbations about $x_{k,i}$.

3.4 The Diffusion Kernel Filter

Problem 3.32 is interpreted as stemming from linearization of the drift function about the deterministic path of a branch of prediction and the parametrization of small state fluctuations about this path into $\int_{t_k}^{t} G(x_{k,i};t,s)dw(s-t_k)$, which may be written

$$\mathrm{D}\phi(x_{k,i};t)\int_{t_k}^{t} g(\Phi(x_{k,i};s))dw(s-t_k). \tag{3.35}$$

See Fig. 1.

Formula (3.35) reads as the accumulated system noise mapped through the deterministic propagator of initial perturbations about $x_{k,i}$. It establishes a duality between the analysis of sensitivity to gaussian-initial perturbations and the analysis of sensitivity to noise-perturbations, in deterministic models. See Fig. 2.

Since taking

$$\Phi(x_{k,i};t) := \phi(x_{k,i};t) + \mathrm{D}\phi(x_{k,i};t)\int_{t_k}^{t} g(\Phi(x_{k,i};s))dw(s-t_k) \tag{3.36}$$

implies

$$\frac{\partial}{\partial t}\Phi(x_{k,i};t) = f(\phi(x_{k,i};t)) + \mathrm{D}f(x)\big|_{x=\phi(x_{k,i};t)}\left(\int_{t_k}^{t} G(x_{k,i};t,s)dw(s-t_k)\right) + G(x_{k,i};t,t)\zeta, \tag{3.37}$$

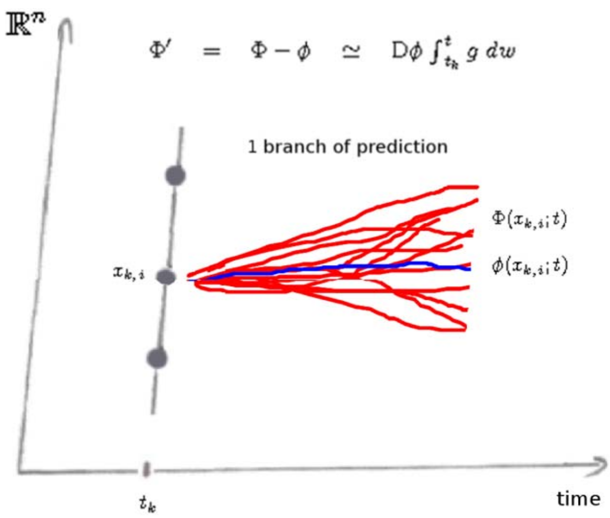

Fig. 1 (Color online) Parametrizing small state fluctuations within branches of prediction with formula (3.35).

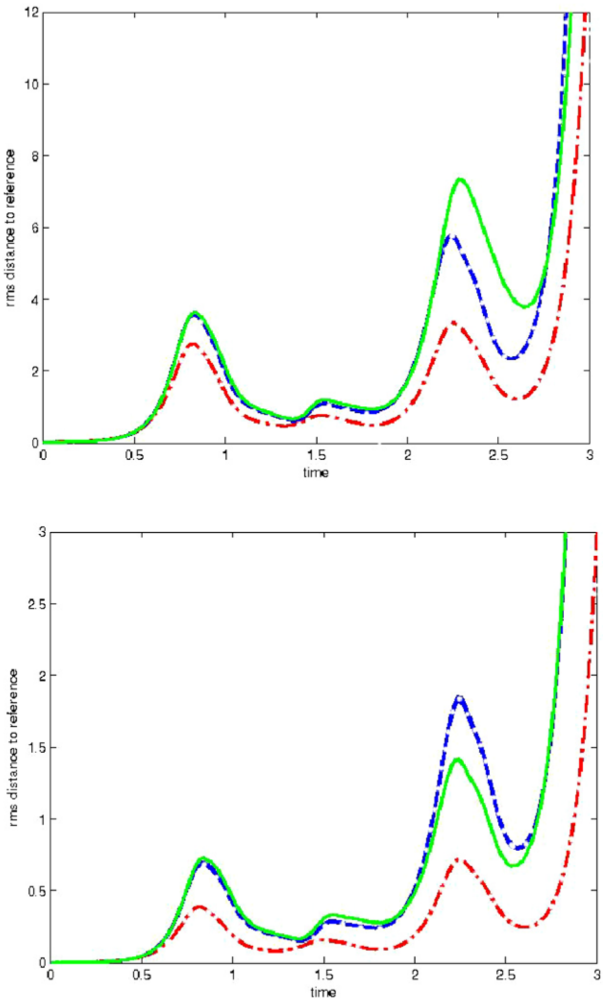

Fig. 2 (Color online) Dynamical sensitivity of the Lorenz-63 equations. *Dash-dot* = rms distance from $\Phi(x_{0,i};t)$ to $\phi(x_{0,i};t)$ (sensitivity to noise perturbations in the equations); *dashed* = rms distance from weakly sampled $D\phi(x_{0,i};t)\int_0^t g\,dw(s)$ to zero (sensitivity to corresponding Gaussian-initial perturbations); *solid* = rms distance from $\phi(x_0;t)$ to $\phi(x_{0,i};t)$, where x_0 is weakly sampled from $x_{0,i}+\int_0^t g\,dw(s)$ for every $t\geq 0$ (sensitivity to corresponding Gaussian-initial perturbations). *Plot* 1: $g=(0.1,0.1,0)$; *Plot* 2: $g=g_0 w(t-t_0)$ with $g_0=(0.03,0.03,0)$

which approximates problem (3.32), the following may be proposed when

$$\int_{t_k}^{t} g(s, w(s-t_k))dw(s-t_k)$$

is analytically integrable as a function of w and t, say into

$$g_0\varphi(w(t-t_k),t,t_k) \tag{3.38}$$

with $g_0 \in M(\mathbb{R},n,p)$.

Algorithm 3.1 (Diffusion Kernel Filter) Apply the Bootstrap Filter with the following change to Step (1): For each $i=1,\dots,I_k$, solve problem (3.33)–(3.34) for the paths of

$\phi(x_{k,i};t)$ and $\mathrm{D}\phi(x_{k,i};t)$, up to time $t=t_{k+1}$; then draw $J_{k,i}$ weak[6] samples from

$$\Phi'(x_{k,i};t_{k+1}) := \mathrm{D}\phi(x_{k,i};t_{k+1}) \int_{t_k}^{t_{k+1}} g(s, w(s-t_k))dw(s-t_k) \tag{3.39}$$

and add them to $\phi(x_{k,i};t_{k+1})$, to obtain $x_{(i,j)}$, $i=1,\dots,I_k$, $j=1,\dots,J_{k,i}$.

When the inequality 3.30 is sharp, the diffusion kernel $\mathcal{L}^2_\tau(n,p)$-norm is expected to be a good measure of the information entropy of a branch of prediction when the latter is close to Gaussian. As such, along with the Diffusion Kernel Filter, the following definition of prediction is proposed.

Definition 3.6 Call average-entropy prediction the deterministic path within the branch of prediction whose diffusion kernel $\mathcal{L}^2_\tau(n,p)$-norm at the end of the prediction time interval is closest to the average norm over all branches, weighed according to their likelihood.

3.4.1 Comments

When g is constant, the parametrization $\Phi'(x_{k,i};t) = G(x_{k,i};t)w(t-t_k)$ stemming from 3.35 for $G(x_{k,i};t) := \mathrm{D}\phi(x_{k,i};t)g$ implies

$$C = (t-t_k)GG^*, \tag{3.40}$$

where $C(x_{k,i};t) := \mathrm{Cov}(\Phi'(x_{k,i};t))$. Therefore $dC/dt = MC + CM^* + GG^*$, where $M := \mathrm{D}f(x)|_{x=\phi(x_{k,i};t)}$. Hence $dC/dt \approx MC + CM^* + gg^*$, for short times, which is consistent with the exact equation for C associated to $dx = Mxdt + gdw$ (cf. [11], Sect. 4.8).

3.4.2 Operations Count

With an explicit numerical integrator for problem (2.4)–(2.5), the operations count per time step in the Bootstrap Filter (BF) is dominantly $O(Inp)$, which is the cost of computing the noise term gdw. With an explicit numerical integrator for problem (3.33)–(3.34), the operations count per time step in the Diffusion Kernel Filter (DKF) is dominantly $O(I_k n^2 p)$ when (3.38) applies, which is the cost of propagating $\mathrm{D}\phi g_0$ through the equation for initial perturbations. The DKF-to-BF count ratio in a prediction time interval is then $O(\Gamma n(I_k/I))$, where $I_k \le I$ and Γ is the ratio between the number of time steps required by the corresponding integrators to cover that interval. This count ratio is smaller than one whenever $n < O((1/\Gamma)(I/I_k))$, where I_k increases with the number of moments present in the filtered state, reaching I when the filtered state is uniformly distributed over some region of the state space. Therefore, for any moderate dimensional problem and upon clustering the filtered state samples prior to defining the next generation of branches of prediction at filtering times, the operations count in the Diffusion Kernel Filter is expected to be smaller than in the Bootstrap Filter whenever the filtered states have few moments.

[6] The parametrization (3.39) is interpreted in distribution.

4 Tests with Lorenz

The Diffusion Kernel Filter and the average-entropy prediction are to be tested with reference to the Bootstrap Filter and the average estimate[7]. Since there is commonly no separation of time scales between variables in a nonlinear dynamics, tests are made with the Lorenz-63 equations perturbed by additive noise. These are the equations

$$\frac{d}{dt}x_1 = \sigma(x_2 - x_1) + \sum_{l=1}^{p} g_{1,l}(t,w)dw_l, \tag{4.1}$$

$$\frac{d}{dt}x_2 = (\rho x_1 - x_2 - x_1 x_3) + \sum_{l=1}^{p} g_{2,l}(t,w)dw_l, \tag{4.2}$$

$$\frac{d}{dt}x_3 = (x_1 x_2 - \beta x_3) + \sum_{l=1}^{p} g_{3,l}(t,w)dw_l, \tag{4.3}$$

taken with $\sigma = 10.0$, $\rho = 28.0$, $\beta = 8/3$ (chaotic regime). In these tests, the observation function is set to identity and the initial state estimate to $x_0 = (-15, -14, 37)$ almost-surely. The initial real state itself is set to $\xi_0 = (-10, -7, 21)$, a rare event of x_0. The noise process z_t that adds to the observation function is set strictly stationary with independent N(0, 1) components. The equations are solved through the explicit weak order 2.0 Heun method [7], with time step $dt = 10^{-4}$. In order to continuously plot the evolution of moments in the Diffusion Kernel Filter, the samples of (3.39) are also drawn in (every time step of) the prediction steps. The filtering times are determined kernel-adaptively by placing a bound on the maximum value of the diffusion kernel $\mathcal{L}^2_\tau(n,p)$-norm[8] over all branches of prediction, computed in every time step, which adapts the size of the prediction steps to the local stability of the dynamical flow.

The maximum-likelihood prediction is defined to be the deterministic path emanating from the most likely sample. The results obtained for the case of a 3×1 constant diffusion matrix g are presented in Fig. 3. In this setting, one has

$$\int_{t_k}^{t} G(x_{k,i}; t, s)dw(s - t_k) = (\mathrm{D}\phi(x_{k,i}; t)g)w(t - t_k), \tag{4.4}$$

$$\|G(x_{k,i}; t, \cdot)\| = \max_{\iota=1,\ldots,n} \sqrt{t - t_k}\,|(\mathrm{D}\phi(x_{k,i}; t)g)_\iota|. \tag{4.5}$$

With the above parameters in this setting, about 100 K samples were needed for the Bootstrap Filter weak statistics to reach convergence up to third moments over the time interval [0, 100]. Plot 3(1-2) shows that the maximum-likelihood predictor is a bad choice when the probability density of the filtered state is multimodal or uniform, the latter case being common with chaotic systems. The results obtained for the case of a 3×1 random diffusion matrix $g = g_0 w(t - t_k)$ with $g_0 \in M(\mathbb{R}, 3, 1)$ are presented in Fig. 4. In this setting, one has

$$\int_{t_k}^{t} G(x_{k,i}; t, s)dw(s - t_k) = (\mathrm{D}\phi(x_{k,i}; t)g_0)\frac{1}{2}(w^2(t - t_k) - (t - t_k)), \tag{4.6}$$

[7] For short times, the aim is to stay close to the average estimate.

[8] A norm on the covariance matrix, when explicit (e.g. as in (3.40)), would be more appropriate.

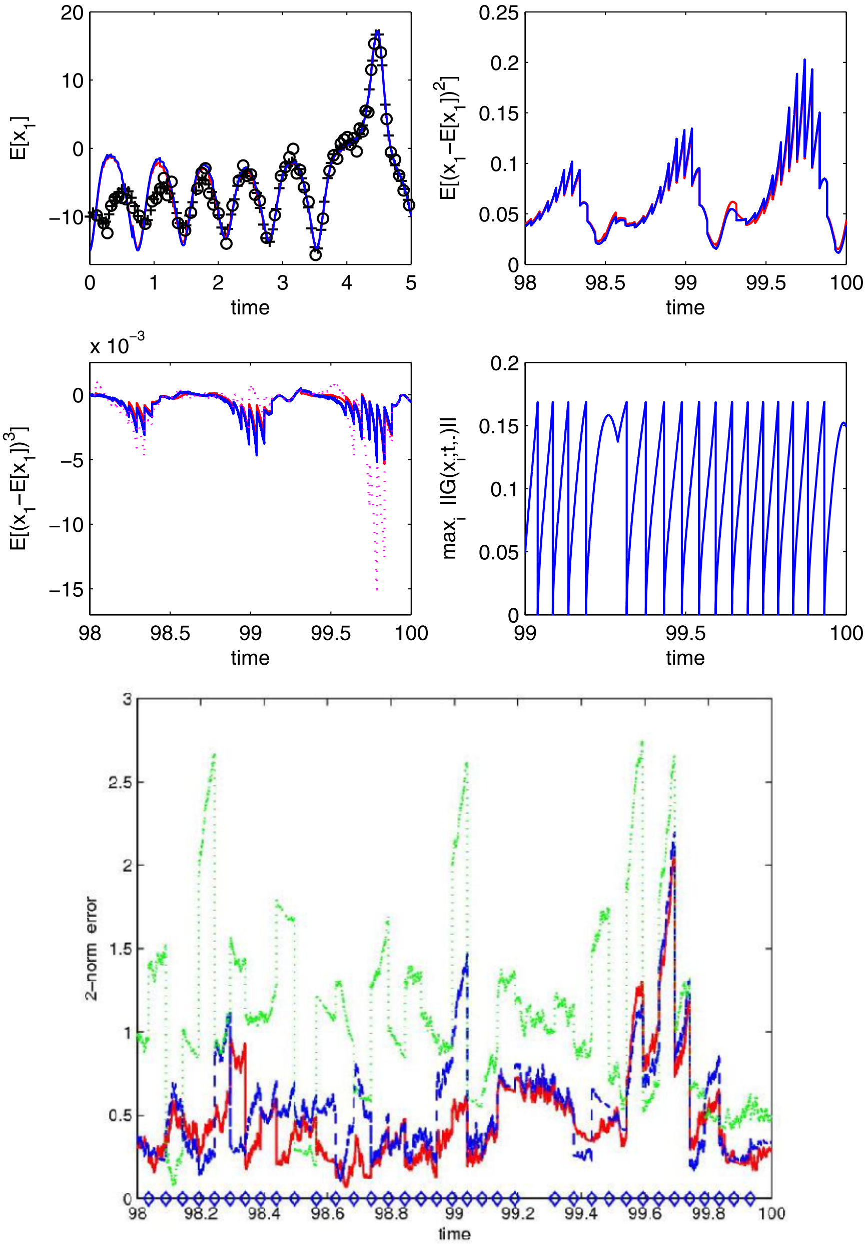

Fig. 3 (Color online) Performance of the Diffusion Kernel Filter and the average-entropy prediction: tests with the Lorenz-63 equations perturbed by additive noise with the constant diffusion matrix $g = (0.4, 0.5, 0.3)$. BF = 100 K samples Bootstrap Filter; DKF = 100 K samples Diffusion Kernel Filter; AV = average estimate; AE = average-entropy prediction; ML = maximum-likelihood prediction. *Plot* 1(1): first moment (*solid* = BF; *solid* = DKF; *cross* = real path); *Plot* 1(2): second central moment (*solid* = BF; *solid* = DKF); *Plot* 2(1): third central moment (*solid* = BF; *solid* = DKF; *dotted* = 10 K samples Bootstrap Filter); *Plot* 2(2): maximum diffusion kernel over all branches of prediction (DKF); *Plot* 3(1-2): prediction error along the filtering process (*solid* = AV; *dashed* = AE; *dotted* = ML; the *diamonds* mark the filtering times)

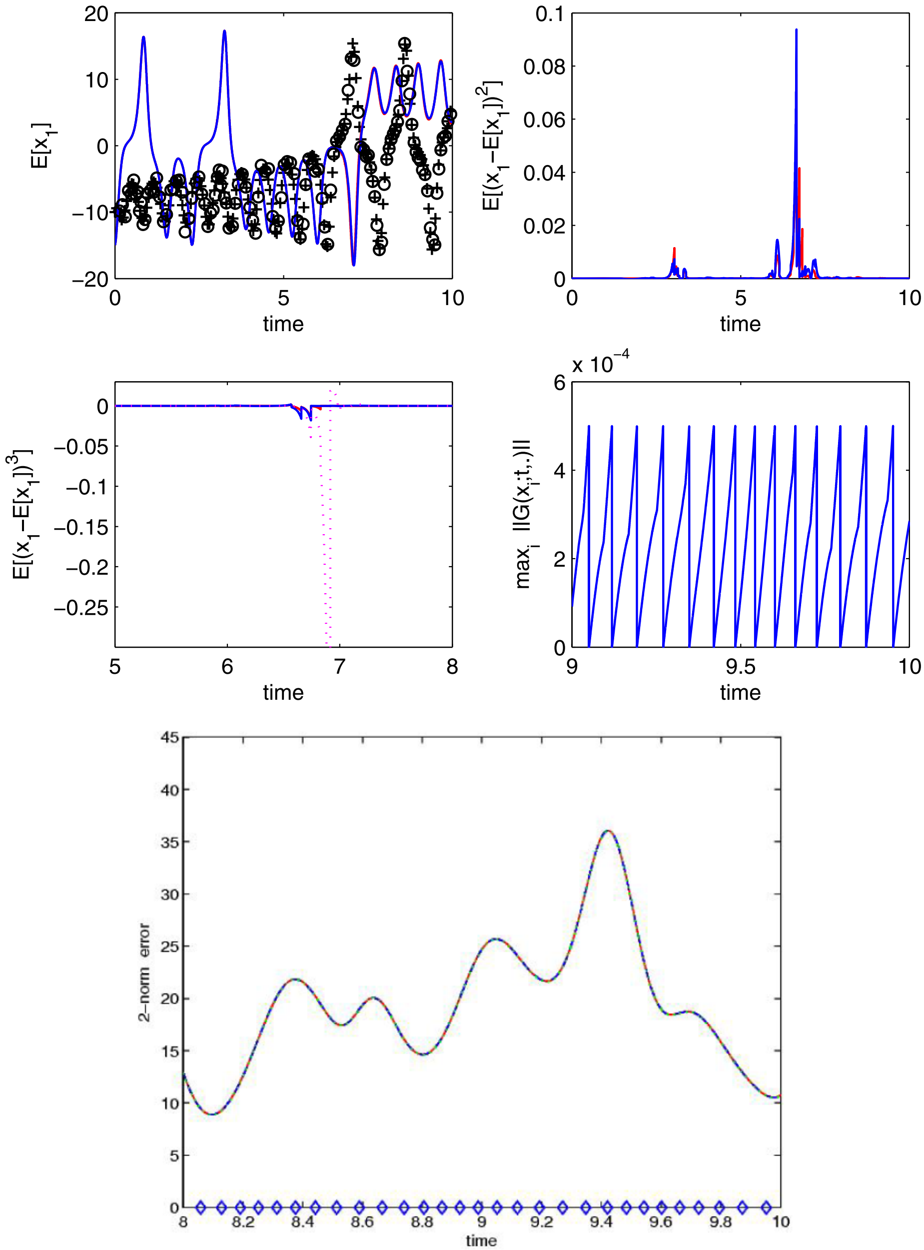

Fig. 4 (Color online) Performance of the Diffusion Kernel Filter and the average-entropy prediction: tests with the Lorenz-63 equations perturbed by additive noise with the random diffusion matrix $g = g_0 w(t - t_k)$ with $g_0 = (0.01, 0, 0.01)$. BF = 500 K samples Bootstrap Filter; DKF = 500 K samples Diffusion Kernel Filter; AV = average estimate; AE = average-entropy prediction; ML = maximum-likelihood prediction. *Plot* 1(1): first moment (*solid* = BF; *solid* = DKF; *cross* = real path); *Plot* 1(2): second central moment (*solid* = BF; *solid* = DKF); *Plot* 2(1): third central moment (*solid* = BF; *solid* = DKF; *dotted* = 100 K samples Bootstrap Filter); *Plot* 2(2): maximum diffusion kernel over all branches of prediction (DKF); *Plot* 3(1-2): prediction error along the filtering process (*solid* = AV; *dashed* = AE; *dotted* = ML; the *diamonds* mark the filtering times)

$$\|G(x_{k,i};t,\cdot)\| = \max_{\iota=1,\dots,n} \frac{\sqrt{2}}{2}(t-t_k)|(\mathrm{D}\phi(x_{k,i};t)g_0)_\iota|. \tag{4.7}$$

The latter expression arises from

$$\|G_\iota\|_2 = \sqrt{\int_{t_k}^{t} \mathrm{E}(w^2(s-t_k))ds}\,|(\mathrm{D}\phi(x_{k,i};t)g_0)_\iota| = \frac{\sqrt{2}}{2}(t-t_k)|(\mathrm{D}\phi(x_{k,i};t)g_0)_\iota|.$$

With the above parameters in this setting, about 500 K samples were needed for the Bootstrap Filter weak statistics to reach convergence up to third moments over the time interval $[0, 20]$. With bound $5 \cdot 10^{-4}$ on the maximum diffusion kernel $\mathcal{L}^2_\tau(n, p)$-norm, agreement between the Bootstrap Filter and the Diffusion Kernel Filter on these moments was found up to time 10. Plot 3(1-2) shows that the predictors and the average estimate coincide in this case. For this to happen, the probability densities of the filtered states must be dominantly unimodal with little skewness.

The results are good in both cases, both for the filter and the definition of prediction. As such the method is shown to be able to weakly sample the filtered states of a chaotic dynamics in a long filtering process with short prediction steps.

5 Conclusion

A particle filter method was presented for the discrete-time filtering problem with nonlinear Itô stochastic ordinary differential equations (SODE) with additive noise supposed to be analytically integrable as a function of the underlying vector-Wiener process and time. The Diffusion Kernel Filter was arrived at by a parametrization of small noise-driven state fluctuations within branches of prediction and a local use of it in the Bootstrap Filter. The referred parametrization was derived by a reformulation of the Itô problem into a Liouville stochastic partial differential equation problem, use of Duhamel's principle in weak form, splitting of a term with the projection $\mathrm{P} := \mathrm{E}[\cdot|x_k]$ and its complement, restriction of the resulting problem to an open nonlinear stochastic ordinary differential equation problem for the flow of a branch of prediction, closure of the latter problem. This was inspired from [2, 3], where a similar technique is used to tackle the dimension reduction problem for the dynamics of a nonlinear ordinary differential equation. For constant diffusion matrices and short times, the established parametrization was shown to be consistent with the Kalman-Bucy Filter prediction of the evolution of the covariance matrix within branches of prediction. It reads as the stochastic integral of a diffusion kernel or the accumulated system noise mapped through the deterministic propagator of initial perturbations. The latter establishes a dual-formula for the analysis of sensitivity to gaussian-initial perturbations and the analysis of sensitivity to noise-perturbations, in deterministic models, showing in particular how the stability of a deterministic dynamics is modeled by noise on short times and how the diffusion matrix of an SODE should be modeled (i.e. defined) for a gaussian-initial deterministic problem to be cast into an SODE problem. With explicit numerical integrators, the Diffusion Kernel Filter was shown to have a smaller count of operations than the Bootstrap Filter whenever the initial state for the prediction step has sufficiently few moments. A norm that estimates the magnitude of the covariance matrix within branches of prediction was provided to the diffusion kernel. From it, a novel definition of prediction was proposed that coincides for constant diffusion matrices with the deterministic path within the branch of prediction whose information entropy at the end of the prediction time interval is closest to the average information entropy over all branches (weighed according to their likelihood). Under this definition, the predictions are expected to stay close to the average estimates for short times. Tests were made with the Lorenz-63 equations, showing good results both for the filter and the definition of prediction. As such, by redefining the branches of prediction at every filtering time, the method was shown to be able to weakly sample the filtered states of a chaotic dynamics in a long filtering process with short prediction steps. If applied with the projection $\mathrm{P} := \mathrm{E}[\cdot|\hat{x}_k]$, where $\hat{x}_k$ stands for a given set of components of x_k, the developed technique produces a marginal fluctuation formula expressed in terms of a reduced diffusion kernel with an effective diffusion matrix, as it will be shown in a succeeding paper.

Acknowledgements The author is grateful to A.J. Chorin, P.L.S. Dias, J.M. Restrepo and A.M. Stuart for support and helpful discussions. This work was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Contract No. 05/56460-3) and partly funded by Instituto do Milênio para o Avanço Global e Integrado da Matemática Brasileira (IM-AGIMB) and the National Science Foundation (NSF, CMG Grant No. 0723765).

References

1. Apte, A., Hairer, M., Stuart, A.M., Voss, J.: Sampling the posterior: an approach to non-Gaussian data assimilation. Physica D **230**, 50–64 (2007)
2. Chorin, A.J., Hald, O.H.: Stochastic Tools in Mathematics and Science. Springer, Berlin (2005)
3. Chorin, A.J., Hald, O.H., Kupferman, R.: Optimal prediction with memory. Physica D **166**, 239–257 (2002)
4. Crisan, D., Del Moral, P., Lyons, T.J.: Discrete filtering using branching and interacting particle systems. Markov Proc. Relat. Fields **5**, 293–331 (1999)
5. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering. Stat. Comput. **10**, 197–208 (2000)
6. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F **140**(2), 107–113 (1993)
7. Higham, D.J., Mao, X., Stuart, A.M.: Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SIAM J. Numer. Anal. **40**(3), 1041–1063 (2002)
8. Jazwinski, A.H.: Stochastic Processes and Filtering Theory. Academic Press, New York (1970)
9. Kato, T.: Perturbation Theory for Linear Operators. Springer, Berlin (1995)
10. Kim, S., Eyink, G.L., Restrepo, J.M., Alexander, F.J., Johnson, G.: Ensemble filtering for nonlinear dynamics. Mon. Wea. Rev. **131**, 2586–2594 (2003)
11. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1999)
12. Miller, R.N., Carter, E.F., Blue, S.T.: Data assimilation into nonlinear stochastic models. Tellus A **51**(2), 167–194 (1999)
13. Öksendal, B.K.: Stochastic Differential Equations: An Introduction with Applications, 6th edn. Springer, Berlin (2003)
14. Raupp, C.F.M., Dias, P.L.S.: Dynamics of resonantly interacting equatorial waves. Tellus A **58**(2), 263–276 (2006)
15. Smith, A.F.M., Gelfand, A.E.: Bayesian statistics without tears: a sampling-resampling perspective. Am. Stat. **46**, 84–88 (1992)