Approximations of the solutions of a stochastic differential equation using Dirichlet process mixtures and Gaussian mixtures

Saba Infante [1,*], César Luna [1], Luis Sánchez [2], Aracelis Hernández [1]

[1] *Department of Mathematics, Faculty of Science and Technology, University of Carabobo, Venezuela*
[2] *Department of Mathematics, Faculty of Education. University of Carabobo, Venezuela*

Abstract   Stochastic differential equations arise in a variety of contexts. There are many techniques for approximation of the solutions of these equations that include statistical methods, numerical methods, and other approximations. This article implements a parametric and a nonparametric method to approximate the probability density of the solutions of stochastic differential equation from observations of a discrete dynamic system. To achieve these objectives, mixtures of Dirichlet process and Gaussian mixtures are considered. The methodology uses computational techniques based on Gaussian mixtures filters, nonparametric particle filters and Gaussian particle filters to establish the relationship between the theoretical processes of unobserved and observed states. The approximations obtained by this proposal are attractive because the hierarchical structures used for modeling are flexible, easy to interpret and computationally feasible. The methodology is illustrated by means of two problems: the synthetic Lorenz model and a real model of rainfall. Experiments show that the proposed filters provide satisfactory results, demonstrating that the algorithms work well in the context of processes with nonlinear dynamics which require the joint estimation of states and parameters. The estimated measure of goodness of fit validates the estimation quality of the algorithms.

Keywords   Gaussian mixtures filter; Nonparametric particle filter; Gaussian particle filter

AMS 2010 subject classifications 62L10 , 62L20

DOI: 10.19139/soic.v4i4.242

1. Introduction

Stochastic differential equations are used in many applications related to basic science such as modeling of biological, chemical, physical, environmental, engineering, economics and finance processes among others ([36]). A stochastic differential equation describes the time evolution of the dynamics of a state vector, based on physical observations from a real system obtained with errors. In practice, the interest consists in estimating the states or parameters of the dynamic system. One of the difficulties commonly encountered in the estimation of parameters in a stochastic differential equation is that the transition density of the equation can not be evaluated in closed form. There are many estimation methods that are based on the maximum simulated Likelihood methods ([64]; [65]; [53]; [22]; [1]; [49] and [21]). Another standard method is to approximate the solutions using the Markov Chain Monte Carlo (MCMC) algorithm ([44]; [27]; [13]; [66] and [67]). It is also possible to approximate solutions by using numerical methods ([23] and [26]). There are others methods based on Hermite expansions ([1]); Taylor approximations ([25]); filtering Theory ([25]; [43] and [24]); approaches that use algorithms based on the integration of Gaussian quadratures and sigma point methods ([28]; [29]; [60] and [23]); adaptive MCMC algorithms based on numerical integration methods recently have been used to approximate in the context of

---

nonlinear stochastic differential equations, for example [60] and [5] have used an approach based on a Gaussian process to model the time evolution of the solution of a general stochastic differential equation, the methodology uses data assimilation techniques ([31]). Recently [57] introduced a nonparametric method to estimate the function of drift in a stochastic differential equation where a measure of random probability was considered to model the drift and a Expectation-Maximization (EM) algorithm was developed to try to establish a link between unobserved and observed states.

This paper aims to approximate the solutions of a stochastic differential equation by using the Gaussian mixture distribution model, Dirichlet process mixture models, together with a non-parametric density estimation algorithm and three sequential filters. The most representative references about the Gaussian mixture distribution model are [61]; [3]; [68] and [2]. Dirichlet processes were treated in [16], and they have been identified in the literature as a useful tool for estimating densities in the field of non-parametric models. The idea of using Dirichlet process is not new, and has been considered in previous works such as [15]; [46]; [42]; [45]; [70] among others.

The rest of article is summarized as follows: Section 2 presents the formulation of the problem; Section 3 contains the description of the filtering algorithms defined to approximate a probability density of the solutions; In Section 4, the results obtained for two different examples are shown, and lastly, Section 5 contains a final discussion and conclusions.

2. Formulation of the problem

This article proposes to use approximations, obtained by nonlinear filters, to estimate the solution of an Itô stochastic differential equation (SDE) given by:

$$d\mathbf{x}_k = f(\mathbf{x}_k, k)dk + L(\mathbf{x}_k, k)d\mathbf{B}_k \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is a vector of stochastic states in a continuous-time, $\mathbf{x}_{k_0}$ is a stochastic initial condition satisfying the following condition $\mathbb{E}(|\mathbf{x}_{k_0}^2|) < \infty$; $\{\mathbf{B}_k : k \geq 0\}$ is a $m-$ dimensional vector of a standard Brownian motion; $F : [k_0, T] \times \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ is a known nonlinear drift function, and $L : [k_0, T] \times \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n \times \mathbb{R}^m$ is a diffusion matrix that satisfies sufficient regularity conditions to ensure the existence of strong solutions ([52]). Theoretically the system is imperfectly observed through a process of continuous observation $\mathbf{y}_k \in \mathbb{R}^m$ which is related to the process $\mathbf{x}_k$ by the equation:

$$\mathbf{y}_k = \int_0^k h(\mathbf{x}_s, s, \theta)ds + \sqrt{\mathbf{D}_k}\mathbf{W}_k \tag{2}$$

where $h : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a known function, $\{\mathbf{W}_k : k \geq 0\}$ is a vector of a standard Brownian motion, which is independent of $\mathbf{B}_k$, $\mathbf{D}_k$ is a matrix of diffusion, and the initial state $\mathbf{x}_0$.

When the dynamic system in continuous time is approximated by a discrete system, the derivatives in continuous time are approximated by difference equations in discrete time, which can be expressed in the form of a space state model, as follows:

$$\mathbf{x}_k = \mathbf{M}_k(\mathbf{x}_{k-1}, \mathbf{u}_k) \tag{3}$$

$$\mathbf{y}_k = \mathbf{H}_k(\mathbf{x}_k) + \mathbf{v}_k \tag{4}$$

$$\mathbf{u}_k \sim F(., .) \tag{5}$$

The equation given in (3) represents a dynamic system, where $\mathbf{x}_k$ denotes the vector of unknown states in a time $t = k$, $\mathbf{u}_k$ is a random error state estimation, $\mathbf{M}_k$ is an operator that maps the space transition state within the state space. The equation given in (4) represents the observed system, where $\mathbf{H}_k$ is an operator that maps the states within the space of observations in time $t = k$, $\mathbf{y}_k$ is the vector of observations, and $\mathbf{v}_k$ is a vector of random errors of observation. In the given equation (5), $F(., .)$ is a function of unknown probability distribution. The unobserved solutions x_k are assumed to be modeled by a Markov process of first order, with an initial distribution

function $p(x_0)$, and transition equation $p(x_k|x_{k-1})$ given by the model defined in equation (3). Each observation y_k is conditionally assumed to be independent given the state x_k, where $p(y_k|x_k)$ is obtained from equation (4). The model is summarized in three steps: first an initial seed $x_0 \sim p(x_0)$ is generated; then $x_k \sim p(x_k|x_{k-1})$ is predicted; and finally $y_k \sim p(y_k|x_k)$ is updated. The general objective is to estimate the unknown states $\mathbf{x}_k = x_{0:k} = (x_0, \ldots, x_k)$, based on measurements obtained from the observation process $\mathbf{y}_k = y_{1:k} = (y_1, \ldots, y_k)$; the main interest is in estimating the filtering distribution $p(x_k|y_{1:k})$, the predictive distribution $p(x_{k+1}|y_{1:k})$, and posterior mean and covariance. The recursive estimation in time k, is performed by the following procedure:

- **Step 1:** The filtering distribution:

$$p(x_k|y_{1:k}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \qquad (6)$$

- **Step 2:** The predictive distribution:

$$p(x_{k+1}|y_{1:k}) = \int p(x_{k+1}|x_k)p(x_k|y_{1:k-1})dx_k \qquad (7)$$

- **Step 3:** The posterior predictive expectation:

$$m_{k|k} = \mathbb{E}(x_k|y_{1:k}) = \int x_k p(x_k|y_{1:k})dx_k \qquad (8)$$

- **Step 4:** The posterior predictive covariance:

$$P_{k|k} = \mathbb{E}\left[(x_k - m_{k|k})(x_k - m_{k|k})^T\right] = \int (x_k - m_{k|k})(x_k - m_{k|k})^T p(x_k|y_{1:k})dx_k \qquad (9)$$

Assuming that $\mathbf{v}_k$ is distributed according to a known probability density function $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$ with fixed but unknown parameters, $\mathbf{R}_k$ is to be estimated. If, in addition, the errors $\mathbf{u}_k$ do not have a known distribution, that is $\mathbf{u}_k \sim F(.,.)$, where $F(.,.)$ is an unknown function, then it is necessary to estimate the probability distribution function $F(.,.)$.

At some time $t = k$, we can estimate the posterior distribution joint $p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k})$ using the Bayes Theorem, as follows:

$$p(x_{0:k}|y_{1:k}) = \frac{p(y_{1:k}|x_{0:k})p(x_{0:k})}{\int p(y_{1:k}|x_{0:k})p(x_{0:k})dx_{0:k}} \qquad (10)$$

It is possible to obtain a recursive formula to estimate the joint density $p(x_{0:k}|y_{1:k})$:

$$p(x_{0:k+1}|y_{1:k+1}) = p(x_{0:k}|y_{1:k})\frac{p(y_{k+1}|x_{k+1})p(x_{k+1}|x_k)}{p(y_{k+1}|y_{1:k})} \qquad (11)$$

where the filtering distribution $p(x_k|y_{1:k})$ is estimated by the following recursive Bayesian filter:

1. It is initialized with a prior distribution $p(x_0)$.
2. For $k = 1, \ldots, N$ run the following:

    - **Prediction step**: the state x_k is predicted, which can be calculated:
        - In the discrete case: The Chapman-Kolmogorov equation:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \qquad (12)$$

        - In the continuous case: The Fokker-Planck-Kolmogorov equation is integrated:

$$\frac{\partial p(x,k)}{\partial k} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}[f_i p(x,k)] + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial x_i \partial x_j}\left\{\left[LQL^T\right]_{ij} p(x,k)\right\} \qquad (13)$$

    where: $p(x,k) = p(x_k|y_{1:k-1})$, $f_i = f_i(x,k)$, y $L = L(x,k)$.

- **Update step**: given the observation y_k, the predictive distribution is updated by:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{\int p(y_k|x_k)p(x_k|y_{1:k-1})dx_k} \qquad (14)$$

Given the difficulty of calculating the integrals involved in (12) and (14); a way to solve the situation is to use computational techniques to approximate the posterior distribution. This article propose three recursive algorithms to achieve this goal.

A particular problem to be solved in this work is to estimate a probability density using a nonparametric Bayesian method involving an endless mix model as considered in [42] and [56]; but in the context of stochastic differential equations. We propose to use the following hierarchical structure to flexibly approximate the probability density of the solutions of the equation given in (1):

$$x_{k-1}|\theta_k \sim f(x_{k-1}|\theta_k) \quad ; \quad k = 1, \ldots, t$$

$$\theta_k(.,) \sim \mathbb{G}(.,.)$$

$$\mathbb{G}(.,.) \sim DP(\alpha, \mathbb{G}_0)$$

$$x_k = M_k(x_{k-1}, u_k) = \sum_{i=1}^{P} \omega_i f(x_{i-1}|\theta_i) \quad , \quad \sum_{i=1}^{n} \omega_i = 1$$

$$x_{k+1}|x_k \sim N(\mu, \Sigma)$$

$$y_k|x_{k+1} \sim N(H_k(x_{k+1}), R_k)$$

where DP is a Dirichlet processes and:

$$\theta_k = (\mu_k, \Sigma_k) \quad ; \quad f(x_k|\mu_k, \Sigma_k) = N(\mu_k, \Sigma_k)$$

$$\mathbb{G}(\mu_k, \Sigma_k) = \mathbb{G}(\mu_k|\Sigma_k)\mathbb{G}(\Sigma_k) = N\left(\lambda, \frac{\Sigma_k}{\tau}\right) IW(\Sigma_0, \nu)$$

$$\mu = \sum_{i=1}^{n} \omega_i \mu_i \quad ; \quad \Sigma = \sum_{i=1}^{n} \omega_i \left[\Sigma_i + (\mu_i - \mu)(\mu_i - \mu)^T\right]$$

3. Filtering algorithms to approximate a density

Here, three filtering algorithms are proposed to approximate the solutions of stochastic differential equation: Gaussian Mixtures Filter (GMF), Nonparametric Particle Filtering (NPF), and Gaussian Particle Filter (GPF).

3.1. *Nonparametric density estimation*

The problem of estimating a density using a nonparametric method consists of the following, data $x_1, \ldots, x_n$ are assumed to be an interchangeable sequence of independent observations taken according to a unknown probability density function F, that is:

$$x_i \sim F(.,.) \quad ; \quad i = 1, \ldots, n \tag{15}$$

The x_i can be scalars or vectors. Traditional methods of parametric statistics, consider probability models that are indexed by unknown finite dimensional parameters, such as the mean and variance, which must be estimated. In contrast, Bayesian nonparametric methods consider a prior probability model $p(F)$ for the unknown distribution function F, where F is a function in a space of infinite dimension. This requires the definition of a measure of random probability of a collection of distributions functions which should have certain properties such as having a large support and the condition that the resulting posterior distribution must be analytically tractable ([16]).

The DP ([16] and [62]), are an important Bayesian nonparametric modeling tool. A random probability distribution F is generated by a DP if for any partition $A_1, \ldots, A_k$ of sample space Ω, the vector:

$$(F(A_1), \ldots, F(A_k)) \sim Dir(\alpha F_0(A_1), \ldots, \alpha F_0(A_k)) \tag{16}$$

where $Dir(\alpha_1, \ldots, \alpha_k)$ denotes the Dirichlet distributions; $F_0(.)$ is a base measure of random probability defining the expected value, $\mathbb{E}(F) = F_0$ or equivalently $\mathbb{E}(F(A)) = F_0(A)$ for any $A \subset \Omega$, and α is a precision parameter that defines the variance. The motivation to use the DP is that, in this approach, updating the posterior distribution is a simple process. That is:

$$x_1, \ldots, x_n | F \sim F$$
$$F \sim Dir(F_0, \alpha)$$
$$F | x_1, \ldots, x_n \sim Dir(\alpha + n, F_0 + \sum_{i=1}^{n} \delta_{x_i}) \tag{17}$$

where δ_{x_i} denotes the Dirac delta.

In a Dirichlet mixture process model (DPM), samples x_i, for $i = 1, \ldots, n$, are taken from a component of the mixture parametrized by $\theta_i \in \Theta$, where θ_i is a latent variable. This auxiliary variable is generated by an unknown distribution $\mathbb{G}$, which represents a measure of random probability distributed according to a prior distribution (a distribution on the set of probability distributions). That is, the DPM is hierarchically defined as follows:

$$x_i | \theta_i \sim f(x_i | \theta_i) \quad ; \quad i = 1, \ldots, n$$
$$\theta_i | \mathbb{G} \sim \mathbb{G}$$
$$\mathbb{G} \sim DP(\alpha, \mathbb{G}_0) \tag{18}$$

where the mixing distribution $\mathbb{G}$ is a DP with concentration parameter α, and distribution base $\mathbb{G}_0$. If integrated on random measure $\mathbb{G}$, we have a non-parametric model to estimate F, as follows:

$$F(x) = \int_{\theta} f(x | \theta) d\mathbb{G}(\theta) \quad ; \quad \mathbb{G} \sim DP(\alpha, \mathbb{G}_0) \tag{19}$$

There are two basic representations of DP, known as Polya urn scheme ([7]), and a prior Stick-breaking scheme ([62] and [54]). The scheme of Polya urn, assumes that the joint distribution of $\theta_1, \ldots, \theta_n$, can be expressed as follows:

$$\theta_1 \sim \mathbb{G}_0(\theta_1)$$
$$\theta_i | \theta_1, \ldots, \theta_{i-1} \sim \frac{\alpha \mathbb{G}_0 + \sum_{l=1}^{i-1} \delta_{\theta_i | \theta_l}}{\alpha + i - 1} \tag{20}$$

where $\delta_{\theta_i|\theta_l}$ is a point mass θ_l. When $\alpha \to 0$, all the $\theta_i = \theta_1 \sim \mathbb{G}_0(\theta_1)$, and when $\alpha \to \infty$, $\theta_i \sim \mathbb{G}_0(\theta_1)$. Moreover, as the θ_i are exchangeable the Polya urn scheme can be written as:

$$\theta_i|\theta_{-i} \sim \frac{\alpha\mathbb{G}_0 + \sum_{i \neq l}\delta_{\theta_i|\theta_l}}{\alpha + n - 1} \tag{21}$$

where θ_{-i} represents the $\{\theta_l : i \neq l\}$.

A strategy for representation the DP, using the scheme called Stick-breaking, is to write the measure $\mathbb{G}$ explicitly as an infinite sum of sums of atomic measures [62]. Embodiments of a DP are expressed as follows:

$$\mathbb{G} = \sum_{k=1}^{\infty} \omega_k \delta_{\theta_k} \tag{22}$$

where, δ_{θ_k} denotes the measure of the Dirac delta located at θ_k and

$$\theta_k \sim \mathbb{G}_0$$
$$\omega_k = \beta_k \prod_{j=1}^{k-1}(1 - \beta_j)$$
$$\beta_k \sim Beta(1, \alpha) \tag{23}$$

The underlying random measure $\mathbb{G}$ is discrete with probability one. Using equation (19), a flexible a priori model used to estimate the unknown distribution F is obtained:

$$F(x) = \sum_{k=1}^{\infty} \omega_k f(x|\theta_k) \tag{24}$$

A recursive form to estimate the unknown probability density in nonparametric form is given as follow:

Algorithm 1: Nonparametric algorithm to estimate a density

- **Step 1:** Simulate:

$$x_{k-1}|\mu_k, \Sigma_k \sim N(\mu_k, \Sigma_k) \tag{25}$$

- **Step 2:** Simulate:

$$(\mu_k, \Sigma_k) \sim NIW(\lambda, \nu, \Sigma_0, \gamma_0) \tag{26}$$

  where:

  – The conditional mean is obtained by:

$$\mu_k|\Sigma_k, \lambda, \tau \sim N\left(\lambda, \frac{\Sigma_k}{\tau}\right) \tag{27}$$

  – The conditional variance is obtained by:

$$\Sigma_k|\Sigma_0, \nu \sim IW(\Sigma_0, \nu) \tag{28}$$

- **Step 3:** Calculate:

$$x_k = M_k(x_{k-1}, u_{k-1}) = \sum_{i=1}^{n} \omega_i N(\mu_i, \Sigma_i) \quad ; \quad \sum_{i=1}^{n} \omega_i = 1 \tag{29}$$

- **Step 4:** Conditioned states are generated $x_{k+1}|x_k$, as follows:

$$x_{k+1}|x_k \sim N(\mu, \Sigma) \tag{30}$$

- **Step 5:** The observations of the states are generated $y_k|x_{k+1}$:

$$y_k|x_{k+1} \sim N(\mathcal{H}(x_{k+1}), R) \tag{31}$$

where IW denotes Inverse Wishart distribution, and $NIW(.,.,.,.)$ denotes a Normal Inverse Wishart.

3.2. *Gaussian mixtures Filter (GMF)*

If we assume that the observation model $y_k = H_k(x_k, v_k) = Hx_k + v_k$ is linear, where $v_k \sim N(0, R)$ and also assume that the a priori model of a state x is approximated by $p(x)$, and the likelihood is approximated by $p(y|x)$, where:

$$p(x) = \sum_{i=1}^{n} \omega_i N(\mu, \Sigma) \quad ; \quad \sum_{i=1}^{n} \omega_i = 1 \tag{32}$$

and

$$p(y|x) = \sum_{j=1}^{m} \beta_j N(Hx, R) \quad ; \quad \sum_{j=1}^{m} \beta_j = 1 \tag{33}$$

Then using equations (32) and (33), the posterior distribution is a Gaussian mixture given by:

$$\begin{aligned}
p(x|y) &= \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx} \\
&= \frac{\sum_{j=1}^{m} \beta_j N(y; Hx, R) \sum_{i=1}^{n} \omega_i N(x; \mu, \Sigma)}{\int \sum_{j=1}^{m} \beta_j N(y; Hx, R) \sum_{i=1}^{n} \omega_i N(x; \mu, \Sigma) dx} \\
&= \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} \omega_i \beta_j N(y; Hx, R) N(x; \mu, \Sigma)}{\sum_{j=1}^{m} \sum_{i=1}^{n} \omega_i \beta_j \int N(y; Hx, R) N(x; \mu, \Sigma) dx} \\
&= \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} \omega_i \beta_j N(y; c, D) N(x; a, B)}{\sum_{j=1}^{m} \sum_{i=1}^{n} \omega_i \beta_j N(y; c, D)}
\end{aligned} \tag{34}$$

Note that:

$$N(Hx, R) N(x; \mu, \Sigma) = \frac{1}{\sqrt{|R||\Sigma|}(\sqrt{2\pi})^{n_x + n_y}} \exp\left(-\frac{Q}{2}\right) \tag{35}$$

where:

$$\begin{aligned}
Q &= (y - Hx)^T R^{-1}(y - Hx) + (x - \mu)^T \Sigma^{-1}(x - \mu) \\
&= x^T B^{-1} x - 2bx + y^T R^{-1} y + \mu^T \Sigma^{-1} \mu \\
&= (x - a)^T B^{-1}(x - a) - a^T B^{-1} a + y^T R^{-1} y + \mu^T \Sigma^{-1} \mu \\
&\approx (x - a)^T B^{-1}(x - a) + (y - c)^T D^{-1}(y - c)
\end{aligned} \tag{36}$$

$$B^{-1} = H^T R^{-1} H + \Sigma^{-1} \quad ; \quad b = y^T R^{-1} H + \mu^T \Sigma^{-1} \quad ; \quad a = Bb \tag{37}$$

$$D^{-1} = R^{-1} - R^{-1} H B^{-1} H^T R^{-1} \quad ; \quad d = \mu^T \Sigma^{-1} B^{-1} H^T R^{-1} \quad ; \quad c = Dd \tag{38}$$

so that:

$$\begin{aligned}
N(x; \mu, \Sigma) N(y; Hx, R) &= \frac{1}{\sqrt{|R||\Sigma|}(\sqrt{2\pi})^{n_x + n_y}} \exp\left\{-\frac{1}{2}\left[(x - a)^T B^{-1}(x - a) + (y - c)^T D^{-1}(y - c)\right]\right\} \\
&= N(x; a, B) N(y; c, D)
\end{aligned}$$

For a recursive algorithm, it is assumed that the probability density function of the initial condition $p(x_0)$ is approximated by:

$$p(x_0) \approx \sum_{i=1}^{n} \omega_{i,0} N(x_0; \mu_{i,0|0}, \Sigma_{i,0|0}), \quad ; \quad \sum_{i=1}^{n} \omega_{i,0} = 1 \tag{39}$$

Without loss of generality, it is assumed that in an instant $k-1$ the system states are known $p(x_{k-1}|y_{1:k-1})$, that is:

$$p(x_{k-1}|y_{1:k-1}) \approx \sum_{i=1}^{n} \beta_{i,k-1} N\left(x_{k-1}; \mu_{i,k-1|k-1}, \Sigma_{i,k-1|k-1}\right), \quad ; \quad \sum_{i=1}^{n} \beta_{i,k-1} = 1 \tag{40}$$

If $x_k = M_k(x_{k-1}, u_k) \approx \sum_{i=1}^{n} \alpha_{i,k} N(x_k; \mu_{i,k|k-1}, \Sigma_{i,k|k-1})$, then it is assumed that $x_{k+1}|x_k$ is given by:

$$p(x_{k+1}|x_k) \approx \sum_{i=1}^{n} \alpha_{i,k} N(x_{k+1}; \mu, \Sigma) \quad ; \quad \sum_{i=1}^{n} \alpha_{i,k} = 1 \tag{41}$$

where:

$$\mu = \sum_{i=1}^{n} \alpha_{i,k-1} \mu_{i,k|k-1} \quad ; \quad \Sigma = \sum_{i=1}^{n} \omega_{i,k} \left[\Sigma_{i,k|k-1}) + (\mu_{i,k|k-1} - \mu)(\mu_{i,k|k-1} - \mu)^T \right]$$

Then the conditional state x_k, given the observed data $y_{1:k-1}$, is approximated by:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1}) p(x_{k-1}|y_{1:k-1}) dx_{k-1}$$

$$\approx \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{j,k-1} \beta_{i,k-1} \mathbf{I}_{i,j}(x_k) \tag{42}$$

where:

$$\mathbf{I}_{i,j}(x_k) = \int N\left(x_k; \mu, \Sigma\right) N\left(x_{k-1}; \mu_{i,k|k-1}, \Sigma_{i,k|k-1}\right) dx_{k-1}$$

$$= \int c_{i,k} N\left(x_k; \mu_{i,k|k}^{1b}, \Sigma_{i,k|k}^{1b}\right) dx_{k-1}$$

$$\mu_{i,k|k-1}^{1b} = \left(\Sigma^{-1} + (\Sigma_{i,k|k-1})^{-1}\right)^{-1} \left(\Sigma^{-1}\mu + (\Sigma_{i,k|k-1})^{-1}\mu_{i,k|k-1}\right)$$

$$\Sigma_{i,k|k-1}^{1b} = \left(\Sigma^{-1} + (\Sigma_{i,k|k-1})^{-1}\right)^{-1}$$

and

$$c_{i,k} = \frac{1}{(2\pi)^{\frac{n+m}{2}} |\Sigma^{-1} + (\Sigma_{i,k|k-1})^{-1}|^{\frac{1}{2}}} \exp{-\frac{1}{2}\left[\left(\mu_{i,k|k-1} - \mu\right)^T \left(\Sigma^{-1} + (\Sigma_{i,k|k-1})^{-1}\right)^{-1} \left(\mu_{i,k|k-1} - \mu\right)\right]}.$$

The assessment of $\mathbf{I}_{i,j}(x_k)$ can be treated as an estimation problem of a nonlinear process with non-Gaussian distribution. In the literature, they have used techniques such as Kalman filter assemblies, or Kalman filter sigma points, to approximate $\mathbf{I}_{i,j}(x_k)$ (see [41]). Furthermore, if it is assumed that the likelihood of the data is distributed according a Gaussian:

$$p(y_k|x_k) \sim N\left(y_k; Hx_k, R_k\right) \tag{43}$$

then, the probability density function of the filtered distribution (the probability distribution function modeling the approximate solutions) is approximated by:

$$
\begin{aligned}
p(x_k|y_{1:k}) &\propto p(y_k|x_k)p(x_k|y_{1:k-1}) \\
&= N\left(y_k; Hx_k, R_k\right) \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{j,k-1}\beta_{i,k-1}\mathbf{I}_{i,j}(x_k) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{j,k-1}\beta_{i,k-1} \int c_{i,k} N\left(y_k; Hx_k, R_k\right) N\left(x_k; \mu_{i,k|k-1}^{1b}, \Sigma_{i,k|k-1}^{1b}\right) dx_{k-1} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{j,k}\beta_{i,k-1} \int c_{i,k}c_{i,k}^{*} N\left(x_k; \mu_{k|k}^{2b}, \Sigma_{k|k}^{2b}\right) dx_{k-1}
\end{aligned}
\tag{44}
$$

where:

$$
\mu_{i,k|k}^{2b} = \left(R_k^{-1} + (\Sigma_{i,k|k-1}^{1b})^{-1}\right)^{-1}\left(R_k^{-1}Hx_k + (\Sigma_{i,k|k-1}^{1b})^{-1}\mu_{i,k|k-1}^{1b}\right)
$$

$$
\Sigma_{i,k|k}^{2b} = \left(R_k^{-1} + (\Sigma_{i,k|k-1}^{1b})^{-1}\right)^{-1}
$$

and

$$
c_{i,k}^{*} = \frac{1}{(2\pi)^{\frac{n+m}{2}}|R_k^{-1} + (\Sigma_{i,k|k-1}^{1b})^{-1}|^{\frac{1}{2}}} \exp -\frac{1}{2}\left[\left(\mu_{i,k|k-1}^{1b} - Hx_k\right)^{T}\left(R_k^{-1} + (\Sigma_{i,k|k-1}^{1b})^{-1}\right)^{-1}\left(\mu_{i,k|k-1}^{1b} - Hx_k\right)\right]
$$

A summary of the algorithm used to estimate the density function by GMF is shown in the following procedure:

Algorithm 2: GMF

- **Step 1:** Simulate the initial states $p(x_0)$ as shown in equation (39).
- **Step 2:** Use the Algorithm 1 to estimate the density of solutions.
- **Step 3:** Predict $p(x_k|y_{1:k-1})$, using the equation(42).
- **Step 4:** Update $p(x_k|y_{1:k})$, using the equation (44).

3.3. Nonparametric filter particles (NPF)

Suppose we want to approximate the filtered density $p(x_k|y_{1:k})$ using an importance function in a marginal state space x_k, that is:

$$
\begin{aligned}
p(x_k|y_{1:k}) &\propto p(y_k|x_k)p(x_k|y_{1:k-1}) \\
&= p(y_k|x_k)\int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1}
\end{aligned}
\tag{45}
$$

A strategy for estimating the filtered distribution is to use a Sequential Monte Carlo method known as Particle filters. The particle filters consider $\left\{x_k^{(i)}, w_k^{(i)}\right\}_{i=1}^{n}$ a random sample which characterize the posterior probability density function $p(x_k|y_{1:k})$. The weights are chosen such that $\sum_{i=1}^{n} W_k^{(i)} = 1$. Then the posterior distribution, also called the filtered distribution in time k, can be approximated by an empirical distribution formed by the mass points or particles:

$$
p_n(x_k|y_{1:k}) \approx \sum_{i=1}^{n} w_k^{(i)}\delta(x_k - x_k^{(i)})
\tag{46}
$$

where $\delta(.)$ is the Dirac delta function. Given the posterior distribution, some quantities of interest can be estimated, such as the expected values of a function $g(x_k)$ associated with the filtered distribution $p(x_k|y_{1:k})$, that is:

$$E[g(x_k)] = \int g(x_k)p(x_k|y_{1:k})dx_k$$

Some of the researchers who worked on the subject are: [71]; [20]; [32]; [37]; [11]; [38]; [55]; [10]; [17]; [19]; [39] among others.

The weights $w_k^{(i)}$ are chosen using the principle of importance sampling. The principle of importance sampling is based on the following argument: Suppose that $p(x) \propto \gamma(x)$, is a probability density which is difficult to sample, but $\gamma(x)$ can be evaluated and consequently $p(x)$ also be evaluated to a proportionality constant. Then proceed as follows: Let $x^{(i)} \sim q(x)$, $i = 1, ..., n$ be a generated sample of a proposed distribution $q(.)$, called density of importance. Then a weighted density approximation of $p(.)$ is given by:

$$p(x) \approx \sum_{i=1}^{n} w^{(i)}\delta(x - x^{(i)}) \tag{47}$$

where:

$$w^{(i)} \propto \frac{\gamma(x^{(i)})}{q(x^{(i)})} \tag{48}$$

The $w^{(i)}$ represents a standardized weight of the i-th particle. If samples $\{x_{0:k}^{(i)}\}$ are taken using a density of importance $q(x_k|y_{1:k})$ then the weights used to approximate the equation (46) are obtained by:

$$w_k^{(i)} \propto \frac{p(x_k^{(i)}|y_{1:k})}{q(x_k^{(i)}|y_{1:k})} \tag{49}$$

If the density of importance can be factored so that:

$$q(x_k|y_{1:k}) = q(x_k|x_{k-1}, y_{1:k})q(x_{k-1}|y_{1:k-1}) \tag{50}$$

then we can get the samples $x_{0:k}^{(i)}$ from $q(x_k|y_{1:k})$ by increasing each of the samples $x_{0:k-1}^{(i)}$ that already exist and are obtained from $q(x_{k-1}^{(i)}|y_{1:k-1})$, generating the new state $x_k^{(i)}$ of $q(x_k|x_{k-1}, y_{1:k})$. For the updated weights, the filtered distribution $p(x_k|y_{1:k})$ is expressed in terms of the equation given in (45). Updated weights are given by:

$$w_k^{(i)} \propto \frac{p(y_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})p(x_{k-1}^{(i)}|y_{1:k-1})}{q(x_k^{(i)}|x_{k-1}^{(i)}, y_{1:t})q(x_{k-1}^{(i)}|y_{1:k-1})}$$

$$= \frac{p(y_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, y_{1:k})} w_{k-1}^{(i)}$$

where

$$w_{k-1}^{(i)} = \frac{p(x_{k-1}^{(i)}|y_{1:k-1})}{q(x_{k-1}^{(i)}|y_{1:k-1})}$$

In particular, if $q(x_k|x_{k-1}, y_{1:k}) = q(x_k|x_{k-1}, y_k)$, then the density of importance depends only on x_{k-1}, and y_k. This situation is appropriate when required to obtain the filtered estimator $p(x_k|y_{1:k})$ in real time k. Then the weights are modified as follows:

$$w_k^{(i)} = \frac{p(y_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, y_k)} w_{k-1}^{(i)} \tag{51}$$

The filtered density $p_n(x_k|y_{1:k})$ can be approximated by:

$$p_n(x_k|y_{1:k}) \approx \sum_{i=1}^{n} w_k^{(i)} \delta(x_k - x_k^{(i)}) \tag{52}$$

It was proved in [12] that when $n \to \infty$ the equation given in (52) approaches the true posterior distribution $p(x_k|y_{1:k})$.

This paper seeks to approximate the filtered density function by using nonparametric statistical techniques. A Dirichlet mixing process is used to estimate the unknown density of states. The algorithm predicts the equation given in (12) and updates the equation given in (14) based on simulated samples. In order to implement the algorithm, it is supposed to start with a set of random samples $\{x_{0:k-1}^{(i)}\}_{i=0}^{n}$ generated by a known distribution function $p(x_{k-1}|y_{1:k-1})$. The procedure is summarized as follows:

Algorithm 3: NPF

- **Step 1:** Initialize with a nonparametric prior distribution of an initial state: (for $k = 0$, and $i = 1, \ldots, n$):

$$\tilde{x}_0^{(i)} \sim p(x_0) \quad ; \quad w_0^{(i)} = 1$$

  where the samples $\{x_{0:k-1}^{(i)}\}$ are considered known from the posterior distribution $p(x_{k-1}|y_{1:k-1})$.
- **Step 2:** Perform the nonparametric importance sampling: (for $k \geq 1$)
    - For $i = 1, \ldots, n$:

$$\tilde{x}_k^{(i)} \sim p(x_k|\tilde{x}_{k-1}^{(i)})$$

    - **Step 3:** Obtain an estimated prediction density:

$$p_n(x_k|y_{1:k-1}) = \sum_{i=1}^{n} w_{k-1}^{(i)} \delta(x_k^{(i)} - \tilde{x}_k^{(i)})$$

      using samples $\left\{\tilde{x}_k^{(i)}, w_{k-1}^{(i)}\right\}_{i=1}^{n}$. For example, the predictive mean and covariance can be estimated as follows:

$$m_{k|k-1} = \sum_{i=1}^{n} w_{k-1}^{(i)} \tilde{x}_k^{(i)}$$

$$P_{k|k-1} = \sum_{i=1}^{n} w_{k-1}^{(i)} (\tilde{x}_k^{(i)} - m_{k|k-1})(\tilde{x}_k^{(i)} - m_{k|k-1})^T$$

    - **Step 4:** For $i = 1, \ldots, n$, the importance density is sampled:

$$\tilde{x}_k^{(i)} \sim q(x_k|y_{1:k})$$

      and the importance weights are calculated:

$$\tilde{w}_k^{(i)} \propto \frac{p(y_k|\tilde{x}_k^{(i)}) p_n(\tilde{x}_k^{(i)}|y_{1:k-1})}{q(\tilde{x}_k^{(i)}|y_{1:k})}$$

    - **Step 5:** Density prediction $p_n(x_k|y_{1:k})$ is updated using the sample $\left\{\tilde{x}_k^{(i)}, \tilde{w}_k^{(i)}\right\}_{i=1}^{n}$.

- **Step 6:** For $i = 1, \ldots, n$, sample:

$$\tilde{x}_k^{(i)} \sim p_n(x_k|y_{1:k})$$

and calculate the importance weights:

$$\tilde{w}_k^{(i)} \propto \frac{p(y_k|\tilde{x}_k^{(i)})p_n(\tilde{x}_k^{(i)}|y_{1:k-1})}{p_n(\tilde{x}_k^{(i)}|y_{1:k})}$$

- **Step 7:** NPF outputs consist of particles $\left\{\tilde{x}_k^{(i)}, \tilde{w}_k^{(i)}\right\}_{i=1}^n$, which approximate the filtered density function:

$$p(x_k, k) = p_n(x_k|y_{1:k}) = \sum_{i=1}^n \tilde{w}_k^{(i)} \delta(x_k^{(i)} - \tilde{x}_k^{(i)})$$

The updated filtered distribution can be characterized by the mean and covariance:

$$m_{k|k} = \sum_{i=1}^n \tilde{w}_k^{(i)} \tilde{x}_k^{(i)}$$

$$P_{k|k} = \sum_{i=1}^n \tilde{w}_k^{(i)} (\tilde{x}_k^{(i)} - m_{k|k})(\tilde{x}_k^{(i)} - m_{k|k})^T$$

3.4. Gaussian particle filter (GPF)

The interest in the following line is to present a method for approximating the solution using a Gaussian Particle Filter (GPF) as proposed in [34]. This algorithm allows the generation of samples that approximate the predictive distribution $p(x_k|y_{1:k-1})$, and the filtered distribution $p(x_k|y_{1:k})$ using Gaussian densities, where the vector of means and the variance-covariance matrix are estimates of the particles generated. If at time $k-1$, the solution is approximated by $N(\mu_{k-1}, \Sigma_{k-1})$, and at time k, it is approximated by $N(\tilde{\mu}_k, \tilde{\Sigma}_k)$, then the steps to implement the GPF consist of the following:

Algorithm 4: GPF

- **Step 1:** Generate $x_{k-1}^{(m)}|\mu_{k-1}, \Sigma_{k-1} \sim N(\mu_{k-1}, \Sigma_{k-1})$.
- **Step 2:** Generate $\mu_k|\Sigma_{k-1}, \lambda, \tau \sim N\left(\lambda, \frac{\Sigma_{k-1}}{\tau}\right)$.
- **Step 3:** Generate $\Sigma_k|\Sigma_0, \nu \sim IW(\Sigma_0, \nu)$, and do:

$$x_k^{(m)} = M_k(x_{k-1}^{(m)}, u_k^{(m)}) = \sum_{k=1}^n \omega_k N(\mu_k, \Sigma_k) \quad ; \quad \sum_{k=1}^n \omega_k = 1 \qquad (53)$$

- **Step 4:** Generate $x_{k+1}^{(m)}|x_k^{(m)} \sim N(\mu, \Sigma)$.
- **Step 5:** Generate $\tilde{w}_k^{(m)} = p(y_k|x_{k+1}^{(m)}) \sim N\left(Hx_{k+1}^{(m)}, R_t\right)$.
- **Step 6:** The weights are normalized $w_k^{(m)} = \frac{\tilde{w}_t^m}{\sum_{k=1}^t \tilde{w}_k^m}$.
- **Step 7:** Estimate $\tilde{\mu}_k$, and $\tilde{\Sigma}_k$ by:

$$\tilde{\mu}_k = \sum_{m=1}^M w_k^{(m)} x_{k+1}^{(m)}$$

$$\tilde{\Sigma}_k = \sum_{m=1}^M w_k^{(m)} (x_{k+1}^{(m)} - \tilde{\mu}_k)(x_{k+1}^{(m)} - \tilde{\mu}_k)^T$$

- **Step 8:** Approximate $p(x_k, k) = p(x_k|y_{1:k}) \sim N(\tilde{\mu}_k, \tilde{\Sigma}_k)$.

4. Results

The first instance consists of a linear case, while the second example consists of a case with a nonlinear structure. To measure the estimation quality of the proposed algorithms we use the square root of the mean square error (RMSE) defined as:

$$RMSE = \frac{1}{n} \sum_{i=1}^{n} \|\hat{x}_k^{(i)} - m_{k|k}\|_2 \tag{54}$$

where $\hat{x}_k^{(i)}$ is an estimator of $\mathbb{E}(x_k|y_{1:k})$, and $m_{k|k} = \sum_{i=1}^{n} w_k^{(i)} x_k^{(i)}$.

4.1. Example 1

As a first example, the Lorenz model is implemented ([40]), representing a coupled system of nonlinear differential equations, given by:

$$\begin{aligned} \frac{dx}{dt} &= s(y - x) \\ \frac{dy}{dt} &= rx - y - xz \\ \frac{dz}{dt} &= xy - bz \end{aligned} \tag{55}$$

where: s, r, b are parameters. The state vector $\mathbf{x} = (x, y, z)^T$ represents a position of the particles in the phase space. The model is discretized using the first order Euler method $x_t = x_{t-1} + hf(x_{t-1})$, with step size h. The discrete evolution equation is given by:

$$\begin{aligned} x_{t+1} &= x_t + h(s(y_t - x_t)) \\ y_{t+1} &= y_t + h(rx_t - y_t - x_t z_t) \\ z_{t+1} &= z_t + h(x_t y_t - bz_t) \end{aligned} \tag{56}$$

where: $s = 10$, $r = 28$, y $b = \frac{8}{3}$. The data was generated by a linear observation equation as proposed in [9]:

$$\gamma_t = \phi_t + \eta_t$$

where: $\gamma_t = (x_t^+, y_t^+, z_t^+)^T$, $\phi_t = (x_t, y_t, z_t)^T$, and $\eta_t = (\eta_{x_t}, \eta_{y_t}, \eta_{x_t})^T$, $\eta_t \sim N(\mathbf{0}, \sigma_\eta^2 \mathbf{I})$, $\mathbf{0}$ is a vector of zeros, and $\mathbf{I}$ is the identity matrix.

To initialize the algorithm the following a priori distributions are considered:

- For GMF: $h = 0.0018$, $\mu_{x_{0|0}} = 0.2294$, $\sigma_{x_{0|0}} = 0.001$, $\mu_{y_{0|0}} = 1.636$, $\sigma_{y_{0|0}} = 0.001$, $\mu_{z_{0|0}} = 20.81$, $\sigma_{z_{0|0}} = 0.001$, $wishardf_x = 10$, $wishardf_y = 100$ and $wishardf_z = 10$.
- For NPF: $h = 0.0018$, $\mu_{x_{0|0}} = 0.2294$, $\sigma_{x_{0|0}} = 0.001$, $\mu_{y_{0|0}} = 1.636$, $\sigma_{y_{0|0}} = 0.001$, $\mu_{z_{0|0}} = 20.81$, $\sigma_{z_{0|0}} = 0.001$, $\sigma_{\eta_x} = 0.01$, $\sigma_{\eta_y} = 0.01$, $\sigma_{\eta_z} = 0.01$, $wishardf_x = 10$, $wishardf_y = 100$ and $wishardf_z = 10$.
- For GPF: $h = 0.0018$, $\mu_{x_{0|0}} = 0.2294$, $\sigma_{x_{0|0}} = 0.001$, $\mu_{y_{0|0}} = 1.636$, $\sigma_{y_{0|0}} = 0.001$, $\mu_{z_{0|0}} = 20.81$, $\sigma_{z_{0|0}} = 0.001$, $\sigma_{\eta_x} = 0.01$, $\sigma_{\eta_y} = 0.01$, $\sigma_{\eta_z} = 0.01$, $wishardf_x = 10$, $wishardf_y = 10$ and $wishardf_z = 10$.

In Table I, the results of the goodness of fit measure, used to gauge the estimation quality of the algorithms, are shown. In this case, the RMSE calculations in the table shows no significant differences in the estimated errors and the computing time. Figure 1, shows three graphs depicting the reconstruction of the true Lorenz attractor and the posterior means of the estimated solution by the three algorithms proposed, where it can be seen that all filters fit almost perfectly to the original map of Lorenz.

RMSE	GMF	NPF	GPF
State x	0.0235	0.0458	0.0328
State y	2.2599	2.2421	2.2602
State z	0.0228	0.0237	0.0108
Time (secs.)	2.1326	2.4443	1.9424

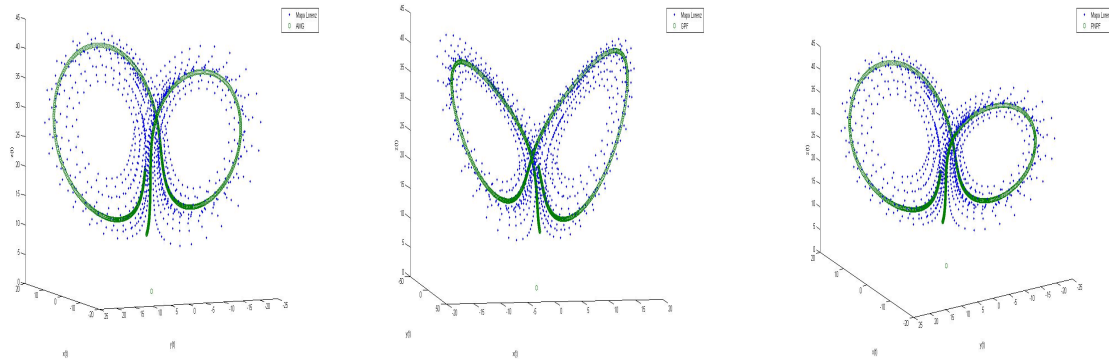Table I. RMSE and computing time of the algorithms GMF, NPF, and GPF.

Figure 1. Lorenz attractor reconstruction using the filters GMF, NPF and GPF.

4.2. Example 2

The second example is a temporary space model based on a stochastic differential equation for predicting rainfall described in [4], where x_t denotes the total rainfall at time t in a locality of $\mathbf{s} = (x, y) \in \mathbf{R}^2$; and x_0 is an initial state. Suppose that the change in total rainfall during a very short time interval Δt and $(\Delta x)_1 = \gamma$, and $(\Delta x)_2 = 0$ have probabilities $p_1 = \lambda \Delta t$, and $p_2 = 1 - \lambda \Delta t$, respectively. The expected value and the mean squared change in the total precipitation is given by:

$$E(\Delta x) = \gamma \lambda \Delta t$$
$$E((\Delta x)^2) = \gamma^2 \lambda \Delta t \tag{57}$$

Letting $\mu = \gamma \lambda \Delta t$, and $\sigma = \sqrt{\gamma^2 \lambda} \Delta t$, a stochastic differential equation to model the total precipitation x_t is obtained:

$$d\mathbf{x}_t = \mu dt + \sigma d\mathbf{W}_t$$
$$x_0 = 0 \tag{58}$$

where $d\mathbf{W}_t$ is a vector of a Brownian process with mean $\mathbf{0}$ and covariance matrix $\mathbf{Q}_t$. The evolution in time of the marginal density of the state is given by the posterior marginal distribution $p(x_k|y_{1:k-1})$, which satisfies the Fokker-Planck equation:

$$\frac{\partial}{\partial t} p(x_k|y_{1:k-1}) = \mathcal{L}(p(x_k|y_{1:k-1}) \tag{59}$$

where $\mathcal{L}$ is a diffusion operator defined by:

$$\mathcal{L}(p) = -\sum_{i=0}^{n_p} \frac{\partial(\mu p)}{\partial x_i} + \frac{1}{2} \sum_{i=0}^{n_p} \sum_{j=0}^{n_p} \frac{\partial^2 [(\sigma Q_t \sigma^T)_{ij} p]}{\partial x_i x_j} \tag{60}$$

with $p = p(x_k|y_{1:k-1})$. The initial condition is given by $p(x_{k-1}|y_{1:k-1})$. Applying the Fokker-Planck equation to the equation (58), we obtain:

$$\frac{\partial p}{\partial t} = -\mu \frac{\partial p}{\partial x} + \frac{\sigma^2 Q_t}{2} \frac{\partial^2 p}{\partial x^2} \qquad (61)$$

The next step is to find a solution to the equation given in (61). The method of discretization of Crank and Nicolson was used to do this. See [59] for detailed. To illustrate the methodology, a data series of daily precipitation was considered. The data included information from January 2011 until May 2012 for three (3) meteorological stations located in Arágua (Ceniap, Tamarindo and Tucutunemo), Venezuela. The data are available in $http://agrometeorologia.inia.gob.ve/$.

To initialize the algorithm GMF, the following a priori distributions were considered:

- Ceniap station $\delta_t = 0.001$, $\delta_x = 0.001$, $\sigma = 100$, $Q_t = 0.0001$, $wishardf = 10$, $\mu = 18.5$, $u = 100$, $v = 100$.
- Tucutunemo station $\delta_t = 0.1$, $\delta_x = 0.1$, $\sigma = 10$, $Q_t = 0.0001$, $wishardf = 10$, $\mu = 18.5$, $u = 10$, $v = 10$.
- Tamarindo station $\delta_t = 0.001$, $\delta_x = 0.001$, $\sigma = 10$, $Q_t = 0.0001$, $wishardf = 10$, $\mu = 18.5$, $u = 100$, $v = 100$.

To initialize the algorithm NPF, the following prior distributions were considered:

- Ceniap station $delta_t = 0.001$, $delta_x = 0.001$, $sigma = 100$, $miu = 18.56$, $q_t = 0.0001$, $wishardf = 10$, $u = 100$ and $v = 100$.
- Tamarindo station $delta_t = 0.001$, $delta_x = 0.001$, $sigma = 100$, $miu = 18.56$, $q_t = 0.0001$, $wishardf = 10$, $u = 100$ and $v = 100$.
- Tucutunemo station $delta_t = 0.1$, $delta_x = 0.1$, $sigma = 100$, $miu = 18.56$, $q_t = 0.0001$, $wishardf = 10$, $u = 10$ and $v = 10$.

To initialize the algorithm GPF, the following prior distributions were considered:

- Ceniap station $delta_t = 0.001$, $delta_x = 0.001$, $sigma = 10$, $miu = 18.56$, $q_t = 0.0001$, $wishardf = 10$, $u = 100$ and $v = 100$.
- Tamarindo station $delta_t = 0.1$, $delta_x = 0.1$, $sigma = 10$, $miu = 18.56$, $q_t = 0.0001$, $wishardf = 10$, $u = 100$ and $v = 100$.
- Tucutunemo station $delta_t = 0.1$, $delta_x = 0.1$, $sigma = 10$, $miu = 18.56$, $q_t = 0.0001$, $wishardf = 10$, $u = 10$ and $v = 10$.

In table II, the RMSE estimated by the proposed algorithms are shown, where we can observe a low variability in the RMSE for the three filters in the Ceniap station. In the Tamarindo station, the lowest RMSE was obtained by the NPF, and the best estimated values in the Tucutunemo station were calculated with the GPF. The computing time for the three filters in all meteorological stations are similar, although the computing time of the NPF is slightly lower than the other two filters. In Figures 2, 3 and 4, the three graphs represent the real states in black, the simulated states using the GMF in red, the NPF in blue and the GPF in green for the three stations. We can see a very similar behavior between the actual values, and simulated values of the states by the three proposed algorithms.

RMSE	GMF	NPF	GPF
Ceniap	0.0018	0.0082	0.0019
Tamarindo	0.1792	0.0790	0.1654
Tucutunemo	0.0540	0.0253	0.0074
Time (secs.)	0.4233	0.3912	0.4826

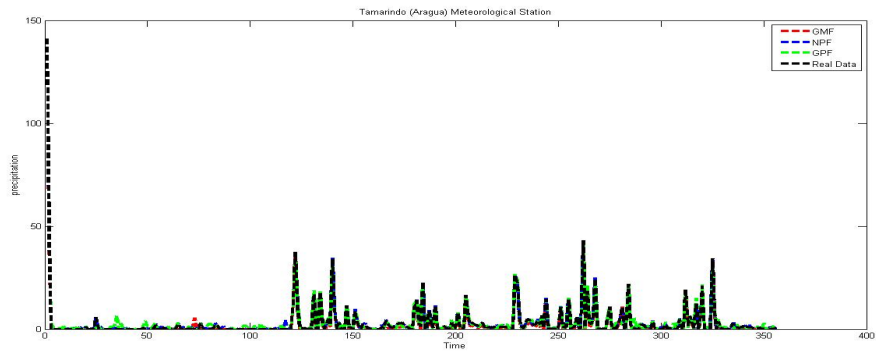Table II. RMSE and computing time of the algorithms GMF, NPF, and GPF.

Figure 2. Real and estimated state by the GMF, NPF and GPF for Tamarindo station.
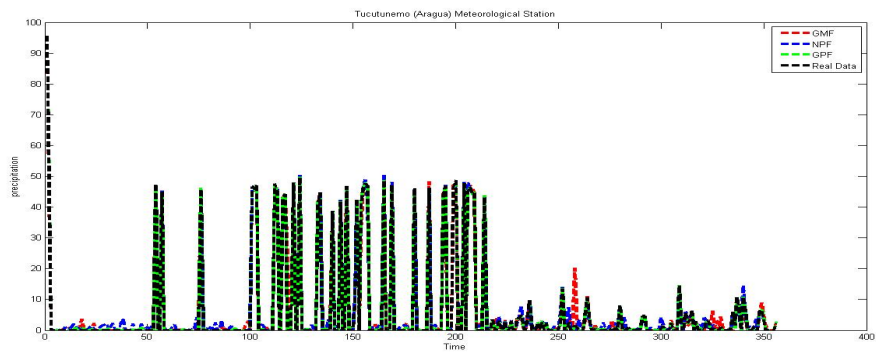
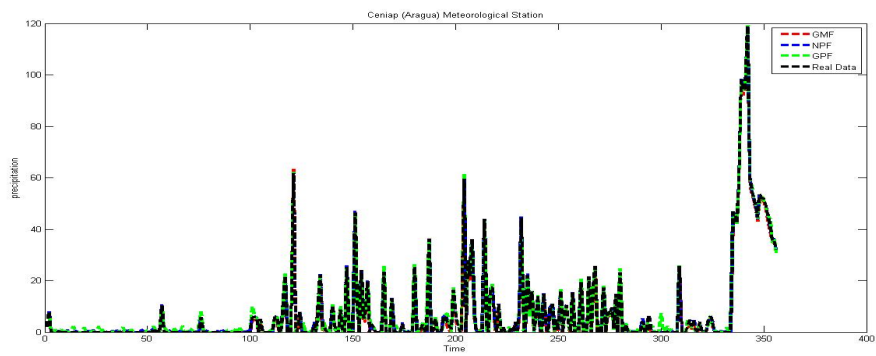Figure 3. Real and estimated state by the GMF, NPF and GPF for Tucutunemo station.

Figure 4. Real and estimated state by the GMF, NPF and GPF for Ceniap station.

5. Conclusions and discussion

This article has introduced a Bayesian nonparametric estimation method to approximate the solutions of a general stochastic differential equation. The main contribution of the work is presenting an approach to combine mixtures of Dirichlet processes and mixtures of Gaussian processes in terms of state space models. The proposed hierarchical structures are extremely flexible with regard to the specifications of the prior assumptions of parameters, are easy to interpret and allow to model a variety of physical phenomena under uncertainty. In addition, we propose an efficient way to predict and update the states of the filtered distribution of the dynamic system through the GMF, NPF and GPF algorithms. The methodology was illustrated by means of two stochastic differential equations, one of them synthetic, and the other obtained from a real process. Since the approximate solutions were obtained from complex models, it is shown that the proposed filters have good performance in the reconstruction of the Lorenz attractor and the states of rainfall model. The computational implementation is also of low cost. It is also shown that the algorithms work well in the context of nonlinear dynamic processes, which require joint estimation of states and distributions of the noise of the equation of state. To measure the quality estimation of the algorithms the square root mean square error was used as a measure of goodness of fit. This measures produced insignificant errors of the estimation.

REFERENCES

1. Y. Ait-Sahalia, *Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach*, Econometrica, vol. 70, no. 1, pp. 223–262, 2002.
2. B. Anderson, and J. Moore, *Optimal Filtering*, Prentice-Hall Information and system sciences series, Englewood Cliffs, New Jersey, 1979.
3. D. Alspach, and H. Sorenson, *Nonlinear Bayesian Estimation Using Gaussian Sum Approximations*, IEEE transactions on automatic control, vol. 17, no 4, pp. 439–448, 1972.
4. E. Allen, *Modeling with Itô Stochastic Differential Equations*, Springer, Texas Tech University, USA, 2007.
5. C. Archambeau, D. Cornford, M. Opper, J. Shawe-Taylor, *Gaussian Process Approximations of Stochastic Differential Equations*, JMLR. Workshop and Conference Proceedings, vol. 1, pp. 1–16, 2007.
6. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, *A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking*, IEEE Transactions on signal processing, vol. 50, no. 2, pp. 174–188, 2002.
7. D. Blackwell, and J. MacQueen, *Ferguson distributions via Polya urn schemes*, The annals of statistics, pp. 353–355, 1973.
8. O. Calin, *An Introduction to Stochastic Calculus with Applications to Finance*, Ann Arbor, 2012.
9. C. Chui, G. Chen, *Kalman Filtering with Real-Time Applications*, Springer Science & Business Media, 2008.
10. A. Doucet, J. De Freitas and N. Gordon, *An introduction to sequential Monte Carlo methods*, Sequential Monte Carlo methods in practice. Springer New York, pp. 3–14, 2001.
11. A. Doucet, S. Godsill and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and computing, vol. 10, no. 3, pp. 197–208, 2000.
12. Dan Crisan and Arnaud Doucet, *A Survey of Convergence Results on Particle Filtering Methods for Practitioners*, IEEE Transactions on signal processing, vol. 50, no. 3, pp. 736–746, 2002.
13. O. Elerian, S. Chib, N. Shephard, *Likelihood inference for discretely observed non-linear diffusions*, Econometrica, vol. 69, no. 4, pp. 959–993, 2001.
14. L. Evans, *An Introduction to Stochastic Differential Equations*, American Mathematical Soc., 2012.
15. M. Escobar, and H. West, *Bayesian density estimation and inference using mixtures*, Journal of the American Statistical Association,vol. 90, no. 430, pp. 577–588, 1994.
16. T. Ferguson, *A Bayesian analysis of some nonparametric problems*, The annals of statistics, vol. 1, pp. 209–230, 1973.
17. W. Fong, S. Godsill, A. Doucet, M. West, *Monte Carlo smoothing with application to audio signal enhancement*, IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 438– 449, 2002.
18. A. Gelfand, A. Kottas, and S. MacEachern, *Bayesian nonparametric spatial modeling with Dirichlet process mixing*, Journal of the American Statistical Association, vol. 100, no. 471, pp. 1021–1035, 2005.
19. S. Godsill, A. Doucet and M. West, *Monte Carlo smoothing for nonlinear Time series*, Journal of the American statistical Association, 2012.
20. N. Gordon, D. Salmond and A. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proceedings F-Radar and Signal Processing. IET, pp. 107-113, 1993.
21. A. Guidoum, and K. Boukhetala, *Quasi Maximum Likelihood Estimation for One Dimensional Diffusion Processes*, In submission (The R Journal), 2014.
22. A. Hurn, K. Lindsay and V. Martin, *On the efficacy of simulated maximum likelihood for estimating the parameters of stochastic differential equations*, Journal of Time Series Analysis, vol. 24, no. 1, pp. 45–63, 2003.
23. A. Hurn, K. Lindsay, and A. McClelland, *A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions*, Journal of Econometric, vol. 172, no. 1, pp. 106–126, 2013.

24. K. Ito, and K. Xiong, *Gaussian Filters for Nonlinear Filtering Problems*, IEEE Transactions on Automatic Control, vol. 45, no. 5, pp. 910–927, 2000.

25. A. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, Inc. New York, USA, 1970.

26. B. Jensen, and R. Poulsen, *Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques*, Journal of Derivatives, vol. 9, no. 4, pp. 18–32, 2002.

27. C. S. Jones, *A simple Bayesian method for the analysis of diffusion processes*, Available at SSRN 111488, 1998.

28. S. Julier and J. Uhlmann, *A new extension of the Kalman filter to nonlinear systems*, AeroSense'97. International Society for Optics and Photonics, pp. 182–193, 1997.

29. S. Julier and J. Uhlmann, *Unscented Filtering and Nonlinear Estimation*, Proceedings of the IEEE, vol. 92, no. 3, pp. 401-422, 2004.

30. S. Julier, J. Uhlmann, and H. Durrant-White, *A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators*, IEEE Transactions on Automatic Control, vol. 45, pp. 477–478, 2000.

31. E. Kalnay, *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, 2003.

32. G. Kitagawa, *Monte Carlo Filter and Smoother for Nonlinear Non Gaussian State Models*, Journal of computational and graphical statistics, vol. 5, no. 1, pp. 1–25, 1996.

33. K. Kokkala, A. Solin, S. Sarkka, *Sigma-Point Filtering and Smoothing Based Parameter Estimation in Nonlinear Dynamic Systems*, arXiv preprint arXiv:1504.06173, 2015.

34. J. Kotecha, and P. Djuric, *Gaussian sum particle filtering*, IEEE Transactions on signal processing, vol. 51, no. 10, pp. 2602–2612, 2003.

35. P. Kloeden, and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer Verlag, 1995.

36. P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, 1992.

37. J. Liu and R. Chen, *Sequential Monte Carlo Methods for Dynamical Systems*, Journal of the American statistical association, vol. 93, no. 443, pp. 1032–1044, 1998.

38. J. Liu and M. West, *Combine Parameter and State Estimation in Simulation Based Filtering*, Sequential Monte Carlo methods in practice. Springer New York, pp. 197–223, 2001.

39. M. Lin, J. Zhang, Q. Cheng, and R. Chen, *Independent Particle Filters*, Journal of the American Statistical Association, vol. 100, no. 472, pp. 1412–1421, 2005.

40. E. Lorenz, *Deterministic Nonperiodic Flow* , Journal of the atmospheric sciences, vol. 20, no. 2, pp. 130–141, 1963.

41. M. Luo, *State estimation in high dimensional systems: the method of the ensemble unscented Kalman filter*, Inference and Estimation in Probabilistic Time-Series Models, 2008.

42. S. MacEachem, and P. Muller, *Estimating Mixture of Dirichlet Process Models*, Journal of Computational and Graphical Statistics, vol. 7, no. 2 pp. 223-238, 1998.

43. P. Maybeck, *Stochastic models, estimation, and control*, Academic Press, vol. 3, 1982.

44. I. Mbalawata, S. Sarkka, H. Haario, *Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and nonlinear Kalman filtering*, Computational Statistics, vol. 28, no. 3, pp. 1195-1223, 2013.

45. P. Muller, F. Quintana, *Nonparametric Bayesian data analysis*, Statistical science, pp. 95–110, 2004.

46. R. Neal, *Markov chain sampling methods for Dirichlet process mixture models*, Journal of Computational and Graphical Statistics, vol. 9, no. 2, pp. 249–265, 2000.

47. J. Neddermeyer, *Sequential Monte Carlo Methods for General State-Space Models*, Tesis Doctoral. Diplomarbeit (Master thesis), University of Heidelberg, 2007.

48. J. Neddermeyer, *Computationally Efficient Nonparametric Importance Sampling*, Journal of the American Statistical Association, vol. 104, no. 486 788–802, 2009.

49. J. Nicolau, *New Technique for Simulating the Likelihood of Stochastic Differential Equations*, The Econometrics Journal, vol. 5, no. 1, pp. 91–103, 2002.

50. B. Øksendal, *Stochastic Differential Equations, An Introduction with Applications*, Springer, 5th edition, 2000.

51. B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, New York, 6th edition, 2003.

52. B. Øksendal, *An Introduction to Malliavin Calculus with Application to Economics*, Lecture Notes, Norwegian School of Economics and Business Administration, 1997.

53. A. Pedersen, *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*, Scandinavian journal of statistics, vol. 22, no. 1 pp. 55–71, 1995.

54. J. Pitman, M. Yor *The two-parameter Poisson Dirichlet distribution derived from a stable sub-ordinator*, The Annals of Probability, vol. 25, no. 2, pp. 855–900, 1997.

55. M. Pitt and N. Shephard, *Filtering via simulation: Auxiliary particle filters*, Journal of the American statistical association, vol. 94, no. 446, pp. 590-599, 1999.

56. A. Rabaoui, N. Viandier, J. Marais, E. Duos, P. Vanheeghe, *Dirichlet Process Mixtures for Density Estimation in Dynamic Nonlinear Modeling: Application to GPS Positioning in Urban Canyons*, IEEE Transactions on Signal Processing, vol. 60, no. 4, pp. 1638–1655, 2012.

57. A. Ruttor, P. Batz, M. Opper, *Approximate Gaussian process inference for the drift function in stochastic differential equations*, Advances in Neural Information Processing Systems, pp. 2040–2048, 2013.

58. L. Sánchez, S. Infante, V. Griffin and D. Rey, *Spatiotemporal dynamic model and parallelized ensemble Kalman Filter for precipitation data*, Brazilian Journal of Probability and Statistics (in press).

59. L. Sánchez, S. Infante, J. Marcano, V. Griffin *Polinomial Chaos based on the parallelized ensamble Kalman filter to estimate precipitation states*, Statistics, Optimization and Information Computing, vol. 3, no. 1, pp. 79–95, 2015.

60. S. Sarkka, *Bayesian filtering and smoothing*, Cambridge University Press, 2013.

61. H. Sorenson, D. Alspach, *Recursive Bayesian Estimation Using Gaussian Sums*, Automatica, vol. 7, no. 4, pp. 465–479, 1971.

62. J. Sethuraman, *A constructive definition of Dirichlet priors*, Statistica Sinica, pp. 639–650, 1994.

63. Z. Schuss, *Nonlinear Filtering and Optimal Phase Tracking Applied*, Springer Science & Business Media, vol. 180, 2011.

64. M. Stefano, *Simulation and inference for stochastic differential equations: with R examples*, Springer Science & Business Media, 2009.
65. M. Stefano, *sde: Simulation and Inference for Stochastic Differential Equations*, R package version, vol. 2, no. 10, 2009.
66. O. Stramer and M. Bognar, *Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach*, Bayesian Analysis, vol. 6, no. 2, pp. 231–258, 2011.
67. O. Stramer, M. Bognar and P. Schneider, *Bayesian Inference of Discretely Sampled Markov Processes with Closed-Form Likelihood Expansions*, The Journal of Financial Econometrics, vol. 8, no. 4, pp. 450–480, 2010.
68. H. Tanizaki, *Nonlinear Filters Estimation and Applications*, Springer Science & Business Media, 2013.
69. G. Terejanu, T. Singh, P. Scott *A Novel Gaussian Sum Filter Method for Accurate Solution to Nonlinear Filtering Problem*, Information Fusion, 2008 11th International Conference on. IEEE, pp. 1–8, 2008.
70. C. Yau, O. Papaspiliopoulos, G. Roberts, C. Holmes, *Bayesian non-parametric hidden Markov models with applications in genomics*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 73, no. 1, pp. 37–57, 2011.
71. M. West, *Mixtures models, Monte Carlo, Bayesian updating and dynamic models*, Computing Science and Statistics, pp. 325–335, 1993.