

# Inference for stochastic differential equations models using filtering algorithms

Saba Infante<sup>a</sup>, Aracelis Hernández<sup>b</sup>, Joan Ordoñez<sup>\*c</sup>, Luis Sánchez<sup>c</sup>

<sup>a</sup>*School of Mathematical Sciences and Information Technology, Yachay Tech University, Ecuador*

<sup>b</sup>*Departamento de Matemáticas, Facultad de Ciencia y Tecnología*

<sup>c</sup>*Instituto de Matemáticas y Cálculo Aplicado (IMYCA), Facultad de Ingeniería, Universidad de Carabobo, Venezuela*

---

## Abstract.-

The practical use of stochastic models in areas of applied research such as: finance, infectious diseases dispersal, telecommunications, audio signal processing, optimum process control, machines of learning, global position systems, and physical systems, among other applications. Five filtering algorithms were implemented: extended Kalman filter (EKF), generic particle filter (GPF), particle filter with resampling (RPF), ensemble Kalman Filter (EnKF), and sigma-point Kalman filter (SPKF). The proposed algorithms allow the approximate reconstruction of the marginal distributions a posteriori of the unknown states of the system. The methodology is illustrated using two stochastic models that come from the world of finance. In the empirical study, it is shown that the filters have good performance in the estimation of the states. No significant differences were observed between the estimated values and the true values. The quality of estimation of the algorithms is evaluated by the goodness of fit measure through the calculation of the root of the mean square error obtaining small estimation errors, in addition we estimated the execution times of the algorithms obtaining satisfactory results.

**Keywords:** filtering algorithms; models space states; equations stochastic differential.

## Inferencia para modelos de ecuaciones diferenciales estocásticas usando algoritmos de filtrado

### Resumen.-

El uso práctico de modelos estocásticos en áreas de investigación aplicada tales como: las finanzas, dispersión de enfermedades infecciosas, telecomunicaciones, procesamiento de señales de audios, control óptimo de procesos, máquinas de aprendizaje, sistemas de posición global, y sistemas físicos, entre otras aplicaciones. Se implementaron cinco algoritmos de filtrado: filtro de Kalman extendido (FKE), filtro de partículas genérico (FPG), filtro de partículas con remuestreo (FPR), filtro Kalman de ensamblajes (FKEN), y el filtro de punto sigmas (FPS). Los algoritmos propuestos permiten la reconstrucción aproximada de las distribuciones marginales a posteriori de los estados desconocidos del sistema. La metodología es ilustrada usando dos modelos estocásticos que provienen del mundo de las finanzas. En el estudio empírico, se demuestra que los filtros tienen buena performance en la estimación de los estados y no se observó diferencias significativas entre los valores estimados y los valores verdaderos. La calidad de estimación de los algoritmos es evaluada por la medida de bondad de ajuste a través del cálculo de la raíz del error cuadrático medio obteniéndose errores de estimación pequeños, además se estimaron los tiempos de ejecución de los algoritmos obteniéndose resultados satisfactorios.

**Palabras clave:** algoritmos de filtrado; modelos espacio estados; ecuaciones diferenciales estocásticas.

---

\*Autor para correspondencia

Correo-e: joanmanuel575@gmail.com (Joan Ordoñez)

Recibido: noviembre 2017

Aceptado: diciembre 2017

## 1. Introducción

Los fenómenos que ocurren en el mundo real por lo general son modelados por proceso de Markov, estos son modelos probabilísticos adecuados para tratar datos que tienen una estructura secuencial. Un proceso de Markov es un proceso de tiempo continuo con trayectorias continuas, pero no diferenciables. El ejemplo más evidente es un sistema dinámico donde los estados evolucionan en el tiempo. En un sistema dinámico los estados son parcialmente observados, solo se tienen mediciones aproximadas del sistema subyacente, las muestras observadas son incompletas, tienen errores y un comportamiento no lineal en la mayoría de los casos. Uno de los objetivos del análisis de estos procesos de difusión es comprender mejor el funcionamiento de los sistemas ocultos para hacer predicción. Por ejemplo en el estudio de un sistema de valoración de activos financieros, un análisis adecuado puede ayudar a mejorar la toma de decisiones, en el estudio de una enfermedad endémica comprender su naturaleza puede contribuir a minimizar el impacto en la sociedad y desarrollar estrategias que permitan erradicarla, o en la elaboración de modelos complejos espacio temporales que permitan comprender la dinámica del sistema del tiempo, pueden ayudar a tomar medidas sobre el cambio climático, lo cual puede tener fuerte impacto sobre el desarrollo de la agricultura y en la contaminación del medio ambiente en las grandes ciudades.

En la inferencia estadística para modelos de ecuaciones diferenciales estocásticas se está interesado en estimar el proceso de Markov (solución de la ecuación) usando toda la trayectoria o el valor esperado de algún funcional del proceso estocástico tales como: los momentos, las distribuciones marginales a posteriori, o las distribuciones a posteriori predictivas, etc.; las cuales generalmente no están disponibles en forma cerrada (analítica), también interesa reducir la varianza de la solución; es decir, se desea determinar soluciones aproximadas a partir de los datos observados con el mínimo costo computacional y que los estimadores tengan

propiedades óptimas. Hay una extensa literatura para tratar estos problemas, ver por ejemplo [1, 2, 3, 4, 5, 6, 7], entre otros.

Las técnicas Bayesianas son señaladas como métodos de estudio para estos sistemas dinámicos. La formulación espacio estado introduce los algoritmos de filtrado que permiten calcular la densidad de probabilidad a posteriori de los estados ocultos del sistema. Los modelos espacio estados, son procesos de Markov, que describen la dependencia probabilista entre dos procesos estocásticos, uno basado en una ecuación de estado y otro basado en una ecuación de observación. Las estructuras de estos modelos son ampliamente utilizadas en muchas aplicaciones tales como: economía y finanzas [8, 9, 10, 11, 12, 13] entre otros, en robótica [14], en biomedicina para caracterización de moléculas de enfermedades [15], para encontrar conexiones en ADN [16, 17], sistemas biológicos [18], epidemiología [19, 20, 21], para datos del servidor de la web [22], para datos de control del tráfico [23], y para modelos de precipitaciones [24, 25, 26, 27, 28], entre otros.

El resto del artículo se como sigue: en la Sección 2 se definen los modelos espacio estado y los sistemas dinámicos, en la Sección 3 se definen los algoritmos a implementar, en la Sección 4 se muestra una aplicación, y en la Sección 5 se establece una discusión y las conclusiones.

## 2. Modelos espacio estado y los sistemas dinámicos

Estamos interesados en dos problemas de inferencia estadística relacionado con los modelos espacio estado (MEE), el primer problema es la inferencia de los estados y el segundo problema esta relacionado con la inferencia de los parámetros. En general los modelos espacio estados se representan por:

$$\begin{aligned} x_0 &\sim p(x_0) \\ x_k &\sim p(x_k|x_{k-1}, \theta) \\ y_k &\sim p(y_k|x_k, \theta) \quad ; \quad k = 1, 2, \dots \end{aligned} \quad (1)$$

donde:

- $x_0$  es un estado inicial conocido.

- $x_k \in \mathbb{R}^n$  es un vector de estados en un tiempo  $k$ , que puede ser continuo o discreto.
- $\theta \in \Theta \subseteq \mathbb{R}^d$  es un vector de parámetros.
- $y_k \in \mathbb{R}^m$  es un vector de observaciones en un tiempo  $k$ .
- $p(x_k|x_{k-1}, \theta)$  es un modelo dinámico el cual describe las dinámicas estocásticas del sistema. El modelo dinámico puede ser una densidad de probabilidad.
- $p(y_k|x_k, \theta)$  es el modelo de medición, que es la distribución de las mediciones dado los estado y parámetros.

En la Figura (1) se muestra una representación gráfica de los modelos espacio estado compuesto por un proceso latente denotado por  $x_t$ , y un proceso observado denotado por  $y_t$ .

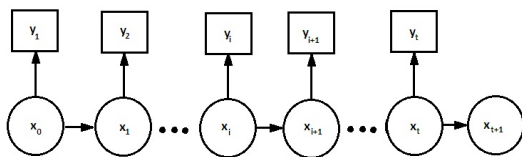


Figura 1: Representación de una cadena de Markov.

Se supone que el modelo dado en (1) es un proceso de Markov, lo que significa que tiene las siguientes dos propiedades:

1. Los estados  $\{x_k : k = 0, 1, \dots\}$  forman una secuencia de Markov de primer orden, esto es:

$$p(x_k|x_{1:k-1}, y_{1:k-1}) = p(x_k|x_{k-1})$$

2. La observación  $y_k$  dado el estado  $x_k$  es condicionalmente independiente de las observaciones y los estados, esto es:

$$p(y_k|x_{1:k-1}, y_{1:k-1}) = p(y_k|x_k)$$

Los MEE discretos implican que los procesos dinámicos y de medición son procesos discretos en el tiempo, mientras que los MEE discretos-continuos implican que el proceso dinámico es un proceso continuo en el tiempo, y el proceso

de medición es un proceso discreto en el tiempo. Los MEE se agrupan en MEE lineales y MEE no lineales, dependiendo de la estructura de los procesos dinámicos y de medición.

**Definición 1.** (*Modelo espacio estado lineal discreto*). En este caso se tiene que la ecuación de observación, y la ecuación de estado son ambas discretas; entonces el modelo definido en (1) se define como:

$$x_0 \sim N(m_0, P_0)$$

$$x_k = M_{k-1}x_{k-1} + u_{k-1} \tag{2}$$

$$y_k = H_k x_k + v_k \tag{3}$$

donde  $M_{k-1}$  es la matriz de transición de la ecuación de estado,  $u_{k-1} \sim N(0, Q_{k-1})$  es un vector de ruidos aleatorios de la ecuación de estado,  $H_k$  es una matriz de la ecuación de observación,  $v_k \sim N(0, R_k)$  es un vector de ruidos aleatorios de la ecuación de observación, y  $x_0$  es un estado inicial.

**Definición 2.** (*Modelo espacio estado lineal continuo-discreto*). En este caso se tiene que la ecuación de observación es discreta, mientras que la ecuación de estado es continua, el modelo definido en (1) se define como:

$$dx_t = M_t x_t dt + L_t dB_t$$

$$y_k = H_k x_k + v_k \tag{4}$$

donde  $M : [0, \infty) \rightarrow \mathbb{R}^{n \times n}$ , y  $L : [0, \infty) \rightarrow \mathbb{R}^{n \times s}$  son funciones de matrices,  $B_t$  es un movimiento Browniano  $s$ -dimensional con densidad espectral  $Q_c \in \mathbb{R}^{s \times s}$ ,  $H_k$  es una matriz que mapea la ecuación de observación, y  $v_k \sim N(0, R_k)$  es un vector de ruidos aleatorios de la ecuación de observación.

En la vida real existen pocos fenómenos con comportamientos lineales y con distribuciones Gaussianas, la mayoría de estos sistemas son no lineales y con errores que tienen comportamiento no Gaussianas. En el caso lineal y distribución Gaussianas se pueden estimar las distribuciones filtradas analíticamente utilizando el algoritmo del filtro de Kalman.

**Definición 3.** (*Modelo espacio estado no lineal discreto*). En este caso se tiene que la ecuación de observación, y la ecuación de estado son ambas discretas, el modelo definido en (1) se define como:

$$x_0 \sim N(m_0, P_0)$$

$$x_k = \mathcal{M}_{k-1}(x_{k-1}) + u_{k-1} \quad (5)$$

$$y_k = \mathcal{H}_k(x_k) + v_k \quad (6)$$

donde:  $\mathcal{M} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  es una función no lineal de las dinámicas de los estados,  $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  es una función no lineal de la ecuación de observación,  $u_{k-1}$  y  $v_k$  son errores que pueden ser Gaussianas o no Gaussianas.

**Definición 4.** (*Modelo espacio estado no lineal continuo-discreto*). En este caso se tiene que la ecuación de observación es discreta, mientras que la ecuación de estado es continua, el modelo definido en (1) se define como: El modelo espacio estado no lineal discreto-continuo se puede escribir como:

$$dx_t = \mathcal{M}(x_t, t)dt + L(x_t, t)dB_t \quad (7)$$

$$y_k = \mathcal{H}(x_{t_k}) + v_k \quad (8)$$

donde  $\mathcal{M} : \mathbb{R}^n \times [0, \infty) \rightarrow \mathbb{R}^n$  es una función no lineal de la ecuación de estado,  $L : \mathbb{R}^n \times [0, \infty) \rightarrow \mathbb{R}^{n \times s}$  es una función de matrices valoradas,  $B_t$  es un movimiento Browniano,  $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  es una función de la ecuación de observación, y  $v_k$  es un vector de ruidos aleatorios de la ecuación de observación que puede o no ser Gaussiana.

Cuando se integra el proceso dinámico discreto-continuo dado en la ecuación (7) desde un estado inicial  $t_0$  hasta  $t$  donde ( $t_0 < t$ ), se obtiene:

$$x_i(t) = x_i(t_0) + \int_{t_0}^t \mathcal{M}_i(x_s, s)ds + \int_{t_0}^t \sum_{j=1}^m L_{ij}(x_s, s)dB_j(s) \quad ; \quad i = 1, \dots, n. \quad (9)$$

donde la primera integral del lado derecho de la ecuación dada en (9) es una integral de Lebesgue, mientras que la segunda integral es una integral estocástica de tipo Ito, (ver Øksendal, (2003) [4],

y Jazwinski, (1970) [3], entre otros). Este proceso puede interpretarse como un sistema determinista gobernado por la parte no aleatoria de la ecuación pero perturbado por un ruido aditivo dado por la integral estocástica. A un proceso estocástico de la forma (9) se le llama proceso de Itô y para que esta ecuación tenga alguna solución se deben imponer condiciones en los coeficientes. De manera análoga al caso determinístico, los teoremas básicos de existencia y unicidad para ecuaciones diferenciales estocásticas establecen que bajo ciertas condiciones de regularidad para los coeficientes  $\mathcal{M}(x_t, t)$  y  $L(x_t, t)$ , la ecuación dada (7) tiene una solución única, por ejemplo, si  $\mathcal{M}$  y  $L$  satisfacen las siguientes condiciones:

- Condición de Lipschitz en la variable  $x$ :

$$|\mathcal{M}(x, t) - \mathcal{M}(y, t)|^2 + |L(x, t) - L(y, t)|^2 \leq K|x - y|^2 \quad (10)$$

- Condición de crecimiento en  $x$ , es dada por:

$$|\mathcal{M}(x, t)|^2 + |L(x, t)|^2 \leq K(1 + |x|^2) \quad (11)$$

- $x(0)$  es independiente de  $(B(t), 0 \leq t \leq T)$ , y  $\mathbb{E}x^2(0) < \infty$ , entonces existe una única solución fuerte de  $x_t$  de la ecuación (7), que tiene trayectorias continuas, más aun:

$$\mathbb{E} \left( \sup_{0 \leq t \leq T} \mathbb{E}x_t^2 \leq C(1 + \mathbb{E}x^2(0)) \right) \quad (12)$$

donde  $C$  es una constante y depende solamente de  $K$  y  $T$ .

Además, la función de densidad de probabilidad de transición de la solución  $x_t$  obedece a la ecuación de Fokker-Planck-Kolmogorov. En este caso a tal solución se le llama solución fuerte. Los teoremas básicos no establecen la forma de encontrar la solución a una ecuación estocástica dada.

La forma clásica de aproximar (7) es suponer una partición de puntos en un intervalo  $[c, d]$ , como sigue:  $c = t_0 < t_1 < \dots < t_n = d$ . La integral de Riemann se aproxima por:

$$\int_c^d f(x)dx = \lim_{\Delta t \rightarrow 0} \sum_{i=1}^n f(t'_i)\Delta t_i \quad (13)$$

donde  $\Delta t_i = t_i - t_{i-1}$ , y  $t_{i-1} < t'_i < t_i$ . En forma similar, la integral de Itô se aproxima por:

$$\int_c^d f(t)dB_t = \sum_{j=1}^m f(t_{i-1})\Delta B_i \quad (14)$$

donde  $\Delta B_i = B_{t_i} - B_{t_{i-1}}$  es un paso del movimiento Browniano en  $[c, d]$ . Para resolver analíticamente la ecuación dada en (9), se necesita introducir la regla de la cadena para diferenciales estocásticas, llamada la fórmula de Itô, que se define como sigue, supóngase que  $Y = u(x(t), t)$ ,  $u : \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$ , es continua, y  $\frac{\partial u}{\partial t}$ ,  $\frac{\partial u}{\partial x}$ ,  $\frac{\partial^2 u}{\partial x^2}$ , existen y son continuas entonces:

$$dY = \frac{\partial u}{\partial t}(x(t), t)dt + \frac{\partial u}{\partial x}(x(t), t)dx + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(x(t), t)(dx)^2 = \left( \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} \mathcal{M} + \frac{1}{2} \frac{\partial^2 u}{\partial x^2} L^2 \right) dt + \frac{\partial u}{\partial x} L dB \quad (15)$$

El término  $(dx)^2$  se interpreta utilizando las siguientes identidades:

- $dt dt = 0$ .
- $dt dB_t = dB_t dt = 0$ .
- $dB_t dB_t = dt$

Una forma muy utilizada en aplicaciones es que en vez de utilizar un método numérico para resolver la ecuación dada en (7), se considera que la solución  $x_t$  es una variable aleatoria que surge cuando se toma el límite y se hace tender a cero el tamaño del paso del método numérico; esto es, para una colección finita de tiempos  $t_0 = \tau_0 < \tau_1 < \dots < \tau_n < \dots < \tau_n = t_j$  denota una discretización de un intervalo  $[t_0, t_j]$ . Supóngase que para  $j = 1, \dots, J$ , existe un entero  $n_j$  que verifica  $t_j = t_{n_j}$ , con  $n_0 = 0$ . Sea  $(\Delta)_{1 \leq n \leq N}$  una secuencia de los tamaños de los pasos definida por  $\Delta_n = \tau_n - \tau_{n-1}$ , y  $\Delta = \max_{1 \leq n \leq N} \Delta_n$  es el tamaño maximal. Entonces la aproximación de Euler-Maruyama de la ecuación (7) define un modelo de series de tiempo de Markov no lineal con dinámicas dadas por:

$$x_n = x_{n-1} + \Delta_n \mathcal{M}(x_{n-1}, \tau_{n-1}) + L(x_{n-1}, \tau_{n-1}) \sqrt{\Delta_n} \zeta_n; \quad \zeta_n \sim N(0, \sigma_\zeta^2) \quad (16)$$

Teóricamente el sistema continuo dado en (7), es observado imperfectamente a través de un proceso de observación  $y_j \in \mathbb{R}$  que está relacionado con el proceso  $x_j$  mediante el siguiente modelo estadístico:

$$y_j = \mathcal{H}(x_{n_j}) + \epsilon_j; \quad \epsilon_j \sim N(0, \sigma_\epsilon^2); \quad 0 \leq j \leq J \quad (17)$$

donde:  $y_{1:T} = (y_1, \dots, y_T)$  es un vector de medidas de observaciones,  $\epsilon = (\epsilon_1, \dots, \epsilon_J)$  representan los errores de medición, el término de la regresión es la realización de un proceso de difusión  $\mathcal{H} : \mathbb{R} \rightarrow \mathbb{R}$ , donde  $\mathcal{H}$  es una función que puede ser lineal o no lineal. El caso más sencillo es la regresión lineal  $\mathcal{H}(x_{n_j}) = x_{n_j}$ .

Usando la aproximación del proceso de difusión por el esquema de Euler-Maruyama, se aproxima el modelo estadístico en la forma espacio estado, como sigue:

$$x_n = x_{n-1} + \Delta_n \mathcal{M}(x_{n-1}, \tau_{n-1}) + L(x_{n-1}, \tau_{n-1}) \sqrt{\Delta_n} \zeta_n; \quad \zeta_n \sim N(0, \sigma_\zeta^2) \quad (18)$$

$$y_n = x_n + \epsilon_n; \quad \epsilon_n \sim N(0, \sigma_\epsilon^2) \quad (19)$$

En lo que sigue nos centraremos en el problema de filtrado, y la predicción de procesos que tienen dinámicas continuas, y que son observados en tiempos discretos. Por ejemplo, las dinámicas de las trayectorias que recorren los vehículos en el espacio son continuas, pero las observaciones son tomadas en instantes de tiempos discreto. El objetivo es obtener aproximaciones por filtros lineales y no lineales la solución de una ecuación diferencial estocástica (EDE) tipo Itô dada en (7).

Dada las observaciones  $y_{1:T} = (y_1, \dots, y_T)$ , la estimación de  $x_k$  es llamada filtrada si  $k = T$ , suavizada si  $k < T$ , y predicción si  $k > T$ . En este trabajo se trata el problema del filtrado, donde la solución de la ecuación (7) es una función de densidad de probabilidad condicional de  $x_k$  dada todas las observaciones  $y_{1:k}$ .

La motivación principal es hacer inferencia bayesiana usando una estructura de un modelo de espacio estado. Se asume una distribución a priori para los parámetros  $p(\theta)$ , y se quiere estimar la distribución a posteriori de las siguientes cantidades:

- $p(\theta|y_{1:k})$ .
- $p(x_{0:k}|y_{1:k})$ .
- $p(\theta, x_{0:k}|y_{1:k})$ .

Cuando  $\theta$  es conocido, y  $x_{0:k}$  es desconocido, la distribución conjunta de los estados es dada por:

$$\begin{aligned} p(x_{0:t}) &= p(x_t|x_{t-1})p(x_{t-1}|x_{t-2}) \dots p(x_1|x_0)p(x_0) \\ &= p(x_0) \prod_{i=1}^t p(x_i|x_{i-1}) \end{aligned} \quad (20)$$

donde  $p(x_0)$  es una densidad inicial a priori. La verosimilitud de los datos viene dada por:

$$\begin{aligned} p(y_{1:t}|x_{0:t}) &= p(y_1|x_1)p(y_2|x_2) \dots p(y_t|x_t) \\ &= \prod_{i=1}^t p(y_i|x_i) \end{aligned} \quad (21)$$

La distribución a posteriori viene dada por:

$$p(x_{0:t}|y_{1:t}) = \frac{p(y_{1:t}|x_{0:t})p(x_{0:t})}{\int p(y_{1:t}|x_{0:t})p(x_{0:t})dx_{0:t}} \quad (22)$$

Cuando  $\theta$ , y  $x_{0:t}$  son desconocidos, se puede diseñar un algoritmo Monte Carlo que permite muestrear de las distribuciones a posteriori marginales  $p(\theta|x_{0:k}, y_{1:k})$ , y  $p(x_{0:k}|y_{1:k}, \theta)$ . Entonces, se puede construir una cadena de Markov sobre los estados  $(\theta, x_{1:k})$ , con distribución estacionaria dada por  $p(\theta, x_{1:k}|y_{1:k})$ . La razón de diseñar un algoritmo sobre estos modelos espacio estado es la forma como están definidos (1), lo cual permite escribir la distribución estacionaria del algoritmo hasta una constante de proporcionalidad:

$$\begin{aligned} A &= \left( \prod_{i=1}^k p(x_i|x_{i-1}, \theta) \right) \\ B &= \left( \prod_{i=1}^k p(y_i|x_i, \theta) \right) \end{aligned}$$

$$p(\theta, x_{1:k}|y_{1:k}) \propto p(\theta)p(x_0|\theta) (A) (B) \quad (23)$$

Por lo tanto, es claro como utilizar los pasos de actualización dentro del algoritmo. En la mayoría de las aplicaciones no es complicado implementar el algoritmo (23) dado que su distribución es estacionaria. Una forma natural de implementar el algoritmo es:

- Se muestrea  $x_{0:k}$  dado  $\theta$  y  $y_{1:k}$ .
- Se muestrea  $\theta$  dado  $x_{0:k}$  y  $y_{1:k}$ .

### 3. Algoritmos

Si se considera que  $\theta$  es conocido entonces la tarea se reduce a estimar la distribución marginal  $p(x_k|y_{1:k})$ . El cálculo de  $p(x_k|y_{1:k})$  se realiza recursivamente debido a que las mediciones son recibidas secuencialmente. Este procedimiento recursivo conduce a un algoritmo de filtrado Bayesiano. El proceso recursivo es inicializado usando una densidad de probabilidad asociada con la información a priori de los estados para algunas observaciones dadas. El algoritmo se resume en tres pasos: primero se genera una semilla inicial  $x_0 \sim p(x_0)$ ; luego se predice  $x_k \sim p(x_k|x_{k-1})$ ; y finalmente se actualiza  $y_k \sim p(y_k|x_k)$ . El objetivo general que se persigue es estimar los estados desconocidos  $x_{0:k}$ , basándose en las medidas obtenidas a partir del proceso de observación  $y_{1:k}$ ; en particular interesa estimar la distribución filtrada  $p(x_k|y_{1:k})$ , la distribución predictiva  $p(x_{k+1}|y_{1:k})$ , la esperanza y covarianza a posteriori, las cuales se calculan como sigue:

1. distribución filtrada:

$$p(x_k|y_{1:k}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \quad (24)$$

2. distribución predictiva:

$$p(x_{k+1}|y_{1:k}) = \int p(x_{k+1}|x_k)p(x_k|y_{1:k-1})dx_k \quad (25)$$

3. esperanza a posteriori predicha:

$$m_{k|k} = \mathbb{E}(x_k|y_{1:k}) = \int x_k p(x_k|y_{1:k})dx_k \quad (26)$$

4. covarianza a posteriori predicha:

$$P_{k|k} = \mathbb{E} \left[ (x_k - m_{k|k})(x_k - m_{k|k})^T \right] = \int (x_k - m_{k|k})(x_k - m_{k|k})^T p(x_k|y_{1:k})dx_k \quad (27)$$

Si se supone que  $v_k$  se distribuye de acuerdo a una función de densidad de probabilidad conocida  $v_k \sim N(0, R_k)$  con parámetros fijos pero desconocidos,  $R_k$  se tiene que estimar; además se consideran que los errores  $u_k \sim N(0, Q_k)$ . Entonces la estimación recursiva de  $p(x_k|y_{1:k})$  en el tiempo  $k$ , se realiza mediante el siguiente procedimiento:

*Algoritmo filtro bayesiano óptimo.*

1. Se inicializa con una distribución a priori  $p(x_0)$ .
2. Para  $k = 1, \dots, \dots$  se ejecuta lo siguiente:
  - predicción: se predice el estado  $x_k$ , que puede ser calculado:
    - en el caso discreto por la ecuación de Chapman-Kolmogorov:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \tag{28}$$

- en el caso continuo: se integra la ecuación de Fokker-Planck-Kolmogorov:

$$\frac{\partial p(x, k)}{\partial k} = - \sum_{i=1}^n \frac{\partial}{\partial x_i} [M_i p(x, k)] + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} \{ [LQL^T]_{ij} p(x, k) \} \tag{29}$$

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

- actualización: dada la observación  $y_k$ , la distribución predictiva, es actualizada mediante:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{\int p(y_k|x_k)p(x_k|y_{1:k-1})dx_k} \tag{30}$$

En general no hay solución analítica para calcular las integrales involucradas en (28), y (30), en particular:

1. Si  $\mathcal{M}(\cdot)$  y  $\mathcal{H}(\cdot)$  son lineales,  $u_t$ , y  $v_t$  tienen distribución Gaussiana entonces  $p(x_k|y_{1:k})$  tiene una solución óptima aproximada por el filtro de Kalman.
2. Si  $\mathcal{M}(\cdot)$  y  $\mathcal{H}(\cdot)$  son no lineales,  $u_t$ , y  $v_t$  tienen distribución Gaussiana entonces  $p(x_k|y_{1:k})$  se puede aproximar por:
  - Filtro de Kalman extendido.
  - Filtro de Kalman sin esencia.
  - Filtro Monte Carlo.
  - Filtro de mezclas de Gaussianas.
  - Filtro de la cuadratura Gauss-Hermite.
  - Filtro de los puntos sigmas.
  - Filtro de Kalman de los ensamblés.
3. Si  $\mathcal{M}(\cdot)$  y  $\mathcal{H}(\cdot)$  son no lineales,  $u_t$ , y  $v_t$  tienen distribución no Gaussiana entonces  $p(x_k|y_{1:k})$  se puede aproximar por:
  - Filtro de partículas y todas sus variantes.
  - Filtro de partículas independiente Metropolis-Hastings.
  - Filtro de partículas Gibbs.
  - Filtro de partículas Gibbs con muestreo de sus descendientes.

En este trabajo mostraremos la estimación de estados a través de los algoritmos filtro de Kalman extendido (FKE), filtro de partículas (FP), filtro de Kalman de ensamblés (FKEN), y filtro de Kalman punto sigmas (FKPS), en modelos generados por ecuaciones diferenciales estocásticas aplicadas a las finanzas.

*3.1. Filtro de Kalman extendido (FKE)*

Si en el modelo general dado en las ecuaciones (5) y (6) las funciones  $\mathcal{M}_t(x_{t-1}) = f(x_{t-1})$  y  $\mathcal{H}_t(x_t) = h(x_t)$  son no lineales, entonces se puede hacer una linealización local expandiendo en serie de Taylor de primer orden alrededor de los puntos estimados. El inconveniente con los filtros no lineales es que pueden ser menos precisos que los filtros lineales y pueden implicar más trabajo en la implementación computacional. Igual que en el caso del filtro de Kalman, se puede obtener un algoritmo recursivo para el filtro de Kalman extendido de la siguiente manera:

### Algoritmo filtro de Kalman extendido.

1. Se inicializa el algoritmo con la media  $m_{0|0}$  y la covarianza  $P_{0|0}$ .
2. Para  $t = 1, 2, \dots$ , se ejecuta lo siguiente:

#### Paso de predicción.

- Caso discreto: se hace:

$$m_{t|t-1} = f(m_{t-1|t-1}) \quad (31)$$

y

$$P_{t|t-1} = Q_t + G_x(m_{t-1|t-1})P_{t-1|t-1}G_x(m_{t-1|t-1})^T \quad (32)$$

- Caso discreto continuo, se resuelve:

$$\frac{dm_{t|t-1}(t)}{dt} = f(m_{t|t-1}, t) \quad (33)$$

y

$$\begin{aligned} \frac{dP_{t|t-1}(t)}{dt} = & G_x(m_{t-1|t-1}, t)P_{t-1|t-1} + \\ & + P_{t-1|t-1}G_x(m_{t-1|t-1}, t)^T \\ & + L(m_{t-1|t-1}, t)Q_tL^T(m_{t-1|t-1}, t) \end{aligned} \quad (34)$$

#### Paso de actualización.

$$m_{t|t} = m_{t|t-1} + p_{t|t-1}W_x^T(W_x p_{t|t-1}W_x^T + R_t)^{-1}[y_t - h_t(m_{t|t-1})], \quad (35)$$

$$p_{t|t} = p_{t|t-1} - p_{t|t-1}W_x^T(W_x p_{t|t-1}W_x^T + R_t)^{-1}(p_{t|t-1}W_x^T)^T \quad (36)$$

donde:  $f(\cdot)$  y  $h(\cdot)$  son funciones no lineales y  $G_x$  y  $W_x$  son linealizaciones de estas funciones, que se estiman por:

$$G_x = \left. \frac{df(x_{t-1})}{dx_{t-1}} \right|_{x_{t-1}=m_{t-1|t-1}} \quad (37)$$

y

$$W_x = \left. \frac{dh(x_t)}{dx_t} \right|_{x_t=m_{t|t-1}} \quad (38)$$

El filtro de Kalman extendido utiliza el primer término de la expansión de Taylor de la función no lineal alrededor las medias condicionales.

### 3.2. Filtro de partículas (FP)

Otra estrategia de estimación de los estados desconocidos en el modelo general dado en las ecuaciones (5) y (6), es utilizar el método de Monte Carlo por muestreo secuencial conocido como filtro de partículas. Para desarrollar en detalle el algoritmo del filtro de partículas considérese  $\{x_{1:t}^{(i)}, w_t^{*(i)}\}_{i=1}^N$  una muestra aleatoria que caracteriza a la función de densidad de probabilidad a posteriori conjunta  $p(x_{1:t}|y_{1:t})$ , donde  $\{x_{1:t}^{(i)}, i = 1, \dots, N\}$  es un conjunto de puntos obtenidos con pesos asociados  $w_t^{*(i)}, i = 1, \dots, N$  y  $x_{1:t} = \{x_1, \dots, x_t\}$  es el conjunto de todos los estados hasta el tiempo  $t$ . Los pesos son normalizados de tal manera que  $\sum_{i=1}^N w_t^{*(i)} = 1$ . Entonces la distribución a posteriori llamada también distribución filtrada en el tiempo  $t$  puede ser aproximada por una distribución empírica formada por los puntos de masa o partículas:

$$p_N(x_{1:t}|y_{1:t}) \approx \sum_{i=1}^N w_t^{*(i)} \delta(x_{1:t} - x_{1:t}^{(i)}) \quad (39)$$

donde  $\delta(\cdot)$  es la función delta de Dirac. Dada la aproximación de la distribución a posteriori, se pueden estimar valores esperados de alguna función  $g_n(x_{1:t})$  asociada a la distribución filtrada  $p(x_{1:t}|y_{1:t})$ , esto es:

$$\mathbb{E}[g_n(x_{1:t})] = \int g_n(x_{1:t})p(x_{1:t}|y_{1:t})dx_{1:t}$$

Los pesos  $w_t^{*(i)}$  son elegidos usando el principio de muestreo de importancia. El principio de muestreo de importancia se basa en el siguiente argumento: supóngase que  $p(x) \propto \gamma(x)$  es una densidad de probabilidad de la cual es difícil muestrear, pero  $\gamma(x)$  puede ser evaluada y en consecuencia  $p(x)$  también puede ser evaluada hasta una constante de proporcionalidad. Entonces se procede de la siguiente manera: sea  $x^{(i)} \sim q(x), i = 1, \dots, N$  una muestra generada de una distribución propuesta  $q(\cdot)$ , llamada densidad de importancia. Entonces una aproximación pesada de la densidad  $p(\cdot)$  es dada por:

$$p(x) \approx \sum_{i=1}^N w^{*(i)} \delta(x - x^{(i)}) \quad (40)$$



donde:

$$w^{*(i)} \propto \frac{\gamma(x^{(i)})}{q(x^{(i)})} \quad (41)$$

es el peso normalizado de la  $i$ -ésima partícula. Si las muestras  $\{x_{1:t}^{(i)}\}$  son tomadas usando una densidad de importancia  $q(x_{1:t}|y_{1:t})$  entonces los pesos utilizados para aproximar la ecuación (39) se obtienen utilizando la ecuación (41), esto es:

$$w_t^{*(i)} \propto \frac{p(x_{1:t}^{(i)}|y_{1:t})}{q(x_{1:t}^{(i)}|y_{1:t})} \quad (42)$$

Si la densidad de importancia se puede factorizar de tal manera que:

$$q(x_{1:t}|y_{1:t}) = q(x_t|x_{1:t-1}, y_{1:t})q(x_{1:t-1}|y_{1:t-1}) \quad (43)$$

entonces se pueden obtener las muestras  $x_{1:t}^{(i)}$  a partir de  $q(x_{1:t}|y_{1:t})$  aumentando cada una de las muestras que ya existen  $x_{1:t-1}^{(i)}$  obtenidas de  $q(x_{1:t-1}|y_{1:t-1})$ , generando el nuevo estado  $x_t^{(i)}$  de  $q(x_t|x_{1:t-1}, y_{1:t})$ . Para obtener los pesos actualizados, la distribución filtrada  $p(x_{1:t}|y_{1:t})$  es expresada en términos de  $p(x_{1:t-1}|y_{1:t-1})$ ,  $p(y_t|x_t)$  y  $p(x_t|x_{t-1})$ , es decir:

$$p(x_{1:t}|y_{1:t}) \propto p(y_t|x_t)p(x_t|x_{t-1})p(x_{1:t-1}|y_{1:t-1}) \quad (44)$$

Por otra parte, sustituyendo la ecuación (43) y (44) en la ecuación (42), se encuentra la ecuación para los pesos actualizados:

$$\begin{aligned} w_t^{*(i)} &\propto \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})p(x_{1:t-1}^{(i)}|y_{1:t-1})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, y_{1:t})q(x_{1:t-1}^{(i)}|y_{1:t-1})} \\ &= \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, y_{1:t})} w_{t-1}^{*(i)} \end{aligned}$$

donde:

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

En particular, si se considera que  $q(x_t|x_{1:t-1}, y_{1:t}) = q(x_t|x_{t-1}, y_t)$ , entonces la densidad de importancia depende solamente de  $x_{t-1}$  y  $y_t$ . Esta situación es adecuada cuando se requiere obtener el estimador

filtrado  $p(x_t|y_{1:t})$  en un tiempo real  $t$ . Entonces los pesos modificados quedan de la siguiente manera:

$$w_t^{*(i)} = \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)} w_{t-1}^{*(i)} \quad (45)$$

y la densidad filtrada a posteriori  $p_N(x_t|y_{1:t})$  puede ser aproximada por:

$$p_N(x_t|y_{1:t}) \approx \sum_{i=1}^N w_t^{*(i)} \delta(x_t - x_t^{(i)}) \quad (46)$$

Crisan y Doucet (2002) [29] probaron que cuando  $N \rightarrow \infty$  la ecuación dada en (46) se aproxima a la verdadera distribución a posteriori  $p(x_t|y_{1:t})$ . Para implementar el algoritmo, supóngase que se tienen un conjunto de muestras aleatorias  $\{x_{1:t-1}^{(i)}, i = 1, 2, \dots, N\}$  generadas de la función conocida  $p(x_{1:t-1}|y_{1:t-1})$ . El algoritmo de filtro de partículas es un método utilizado para predecir (ecuación (28)) y para actualizar (ecuación (30)) con base a las muestras conocidas, y obtener un conjunto de nuevos valores  $\{x_{1:t}^{(i)}, i = 1, 2, \dots, N\}$ , los cuales se distribuyen aproximadamente como  $p(x_t|y_{1:t})$ . El algoritmo de filtro de partículas genérico se resume como sigue:

*Algoritmo filtro de partículas genérico (FPG).*

1. Se inicializa el estado inicial  $x_0 \sim p(., .)$ , donde  $p(., .)$  es una distribución a priori conocida, y las muestras  $\{x_{1:t-1}^{(i)}, i = 1, 2, \dots, N\}$  de la distribución a posteriori  $p(x_{t-1}|y_{1:t-1})$  en  $t - 1$ , con  $t = 1, \dots, T$ .
2. Se hace  $i = 1, \dots, N$ , y se genera una muestra de la distribución propuesta:

$$x_t^{(i)} \sim q(x_t|x_{1:t-1}^{(i)}, y_{1:t}) \quad (47)$$

y se construye el conjunto  $x_{1:t}^{(i)} = \{x_{1:t-1}^{(i)}, x_t^{(i)}\}$ .

3. Para  $i = 1, \dots, N$ , se evalúan los pesos de importancia hasta una constante normalizada:

$$w_t^{*(i)} = \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, y_{1:t})} w_{t-1}^{*(i)}$$

4. Para  $i = 1, \dots, N$ , se normalizan los pesos de importancia:

$$\tilde{w}_t^{(i)} = \frac{w_t^{*(i)}}{\sum_{j=1}^N w_t^{*(j)}}, \quad \sum_{i=1}^N \tilde{w}_t^{(i)} = 1$$

El problema que se presenta con el algoritmo de filtro de partículas genérico es el fenómeno de la degeneración. Es conocido que después de unas pocas iteraciones las partículas pueden tener pesos despreciables lo que implica según la Proposición 1 definida en Doucet et al. (2000), que la varianza de los pesos de importancia pueden incrementarse en el tiempo. Para evitar la degeneración del algoritmo, Doucet et al. [30] establecieron la Proposición 2, que permite seleccionar una función de importancia que minimiza la varianza de los pesos sobre las trayectorias de los estados simulados  $\{x_{1:t-1}^{(i)}, i = 1, 2, \dots, N\}$  y las observaciones  $y_{1:t}$ ; en el Apéndice E se definen las proposiciones 1 y 2. La función de importancia sugerida por ellos es  $q_t(x_t|x_{1:t-1}^{(i)}, y_{1:t})$ , con:

$$q_t(x_t|x_{1:t-1}^{(i)}, y_{1:t})_{optimo} = p(x_t|x_{t-1}^{(i)}, y_t) = \frac{p(y_t|x_t, x_{t-1}^{(i)})p(x_t|x_{t-1}^{(i)})}{p(y_t|x_{t-1}^{(i)})} \quad (48)$$

es decir,  $q_t(x_t|x_{1:t-1}^{(i)}, y_{1:t})_{optimo}$  es la función de importancia que minimiza a la varianza de los pesos de importancia  $w_t^{*(i)}$  condicionada sobre los estados  $\{x_{1:t-1}^{(i)}, i = 1, 2, \dots, N\}$  y los datos observados  $y_{1:t}$ . Sustituyendo (48) en (45), se obtiene:

$$w_t^{*(i)} \propto p(y_t|x_{t-1}^{(i)})w_{t-1}^{*(i)} \quad (49)$$

Como la degeneración del algoritmo de filtro de partícula es inevitable, entonces se introduce la idea del remuestreo que consiste en eliminar las trayectorias que tienen pesos de importancias pequeños y concentrarse sobre trayectorias con pesos grandes. Una medida adecuada para solventar el problema es calcular el tamaño de muestra efectiva ( $N_{TME}$ ) introducido en Kong et al. (1994) [31], y Liu (1996) [32], y que se define como:

$$N_{TME} = \frac{N}{1 + Var(w_t^{*(i)})} \quad (50)$$

en la práctica es complicado calcular  $N_{TME}$  exactamente, pero se estima por:

$$\hat{N}_{TME} = \frac{1}{\sum_{i=1}^N [\tilde{w}_t^{(i)}]^2} \quad (51)$$

entonces se compara  $\hat{N}_{TME}$  con un umbral prefijado  $N_U$ . Este procedimiento de filtro de partículas con remuestreo es usado por Rubin (1988) [33]. El algoritmo modificado se resume como sigue:

*Algoritmo filtro de partículas con remuestreo.*

1. Muestreo de Importancia

- Se inicializan el estado inicial  $x_0 \sim p(.,.)$ , donde  $p(.,.)$  es una distribución a priori conocida, y las muestras  $\{x_{1:t-1}^{(i)}, i = 1, 2, \dots, N\}$  de la distribución a posteriori  $p(x_{t-1}|y_{1:t-1})$  en  $t - 1$ , con  $t = 1, \dots, T$ .

- Para  $i = 1, \dots, N$ , se muestrea

$$\tilde{x}_t^{(i)} \sim q(x_t|x_{1:t-1}^{(i)}, y_{1:t})$$

y se construye el conjunto  $\tilde{x}_{1:t}^{(i)} = \{x_{1:t-1}^{(i)}, \tilde{x}_t^{(i)}\}$ .

- Para  $i = 1, \dots, N$ , se evalúan los pesos de importancia:

$$w_t^{*(i)} \propto w_{t-1}^{*(i)} \frac{p(y_t|\tilde{x}_t^{(i)})p(\tilde{x}_t^{(i)}|\tilde{x}_{t-1}^{(i)})}{q(\tilde{x}_t^{(i)}|\tilde{x}_{1:t-1}^{(i)}, y_{1:t})}$$

- Para  $i = 1, \dots, N$ , los pesos de importancia se normalizan:

$$\tilde{w}_t^{(i)} = \frac{w_t^{*(i)}}{\sum_{j=1}^N w_t^{*(j)}}, \quad \sum_{i=1}^N \tilde{w}_t^{(i)} = 1$$

- Se evalúa  $\hat{N}_{TME}$  usando la ecuación (51).

2. Remuestreo

- Si  $\hat{N}_{TME} \geq N_U$ , se hace  $x_{1:t}^{(i)} = \tilde{x}_{1:t}^{(i)}$ , para  $i = 1, \dots, N$ .
- Si  $\hat{N}_{TME} < N_U$ , para  $i = 1, \dots, N$ , se muestrea un índice  $j(i)$  distribuido de acuerdo a la distribución discreta con  $N$  elementos que satisfacen  $P_r\{j(i) = l\} = \tilde{w}_t^{(l)}$  para  $l = 1, \dots, N$ .
- Para  $i = 1, \dots, N$ , se hace  $x_{1:t}^{(i)} = \tilde{x}_{1:t}^{j(i)}$  y  $w_t^{(i)} = \frac{1}{N}$ .

Finalmente, se tiene al cabo del tiempo  $t$  la aproximación:

$$p_N(x_t|y_{1:t}) \approx \frac{1}{N} \sum_{i=1}^N \delta_{x_{1:t}^{(i)}}(x_t) \quad (52)$$

### 3.3. Filtro de Kalman de ensamble (FKEN)

El FKEN fue desarrollado por Evensen (1994) [34], Evensen y Leeuwen (1996) [35]. Es un algoritmo basado en los métodos de Monte Carlo Secuencial utilizado para aproximar la distribución a posteriori de los estados de sistemas no lineales en altas dimensiones.

- Representa los errores usando un ensamble de los estados del modelo.
- Los errores evolucionan a través de la integración de los ensambles.
- Minimiza la varianza usando un paso de predicción y otro de actualización.
- Usa técnicas de Monte Carlo de bajo rango.
- Converge al filtro de Kalman cuando se incrementa el tamaño del ensamble.
- Background: la estimación de un estado  $x_t$ , obtenido a partir de un estado  $x_{t-1}$  se llama conocimiento previo en el ciclo de asimilación  $t$ , y se denota por  $x_t^b$ .
- Análisis: Dado  $x$  el procedimiento de actualizar  $x_t^b$  por el ajuste lineal de Bayes se llama análisis y se denota por  $x_t^a$ .
- La covarianza del E (background):  $P^b = E(x^b - x)(x^b - x)^T$ .
- La covarianza del E (análisis):  $P^a = E(x^a - x)(x^a - x)^T$ .

#### Algoritmo FKEN.

- Creación del ensamble inicial. Dados  $(\hat{x}_0, \hat{P}_0)$ , y  $\hat{P}_0 = \hat{S}_0 \hat{S}_0^T$ .

$$x_{0,i}^a = \hat{x}_0 + \hat{S}_0 n_0^i \quad ; \quad n_0^i \sim N(0, I)$$

$$\hat{x}_0^b = \frac{1}{n} \sum_{i=1}^n x_{0,i}^a$$

$$\hat{P}_0^b = \frac{1}{n-1} \sum_{i=1}^n (x_{0,i}^a - \hat{x}_0^b)(x_{0,i}^a - \hat{x}_0^b)^T$$

- Pronóstico del ensamble. Dadas  $(\hat{x}_t^b, \hat{P}_t^b)$ , y  $\hat{P}_t^b = \hat{S}_t \hat{S}_t^T$ ,

$$x_{t,i}^a = \hat{x}_t^b + \hat{S}_t n_t^i, \quad n_t^i \sim N(0, I)$$

$$x_{t,i}^b = \mathcal{M}_t(x_{t,i}^a) + u_t^i, \quad u_t^i \sim N(0, Q_t).$$

- Media del pronóstico:

$$\hat{x}_t^b = \frac{1}{n} \sum_{i=1}^n x_{t,i}^b$$

- Error del pronóstico:

$$e_t^i = x_{t,i}^b - \hat{x}_t^b$$

- Covarianza del pronóstico:

$$\hat{P}_t^b = \frac{1}{n-1} \sum_{i=1}^n (x_{t,i}^b - \hat{x}_t^b)(x_{t,i}^b - \hat{x}_t^b)^T + Q_t$$

En la práctica, no es común aproximar  $\hat{P}_t^b$  y en su lugar se estima:

$$\hat{P}_t^{cr} = \frac{1}{n-1} \sum_{i=1}^n (x_{t,i}^b - \hat{x}_t^b) [\mathcal{H}_t(x_{t,i}^b) - \mathcal{H}_t(\hat{x}_t^b)]^T$$

$$A = [\mathcal{H}_t(x_{t,i}^b) - \mathcal{H}_t(\hat{x}_t^b)]$$

$$B = [\mathcal{H}_t(x_{t,i}^b) - \mathcal{H}_t(\hat{x}_t^b)]^T$$

$$\hat{P}_t^{pr} = \frac{1}{n-1} \sum_{i=1}^n (A)(B)$$

- Análisis del ensamble. Se generan observaciones sintéticas:

$$w_{t+1}^i = \mathcal{H}_{t+1}(x_{t+1,i}^b) + v_{t+1}^i, \quad v_{t+1}^i \sim N(0, R_{t+1})$$

- Media actualizada:

$$\hat{x}_{t+1}^a = \hat{x}_{t+1}^b + K_{t+1}(w_{t+1} - w_{t+1}^i)$$

- Covarianza actualizada:

$$\hat{P}_{t+1}^a = \hat{P}_{t+1}^b - K_{t+1}(\hat{P}_{t+1}^{cr})^T$$

- Matriz de ganancia:

$$K_{t+1} = \hat{P}_{t+1}^{cr} (\hat{P}_{t+1}^{pr} + R_{t+1})^{-1}$$

### 3.4. Filtro de Kalman de los puntos sigmas (FKPS)

El FKPS utiliza un conjunto de puntos de cuadratura, llamados sigma-puntos, para aproximar la integral de Chapman-Kolmogorov [36]. Estos puntos se generan a partir de una distribución Gaussiana. Se utiliza la parametrización de la media y la covarianza para hacer la aproximación.

La distribución del vector de estados, tiene una distribución Gaussiana multivariada, dada por:

$$p(x_{k-1}|y_{1:k-1}) \sim N(\hat{x}_{k-1}, p_{k-1})$$

donde  $\hat{x}_{k-1}$  es un vector de medias, y  $p_{k-1}$  es una matriz de varianza covarianza.

La integral de Chapman-Kolmogorov es de la forma:

$$\int_{\mathbb{R}^n} \mu(x)\omega(x)dx$$

donde  $\mu(x)$  es una distribución de probabilidad arbitraria,  $\omega(x)$  es la distribución de probabilidad a priori del vector de estados. Entonces:

$$\int_{\mathbb{R}^n} \mu(x)\omega(x)dx = \int_{\mathbb{R}^n} \mu(x)N(\hat{x}_{k-1}, p_{k-1})dx$$

Para aproximar numéricamente la integral, se requiere de una representación discreta de  $p(x_{k-1}|y_{1:k-1})$ , la cual se lleva a cabo generando un conjunto de puntos de una distribución Gaussiana igual que las variables de estados originales [36]:

$$\chi_j = \begin{cases} \sqrt{n + \rho}e_j, & 1 \leq j \leq n \\ -\sqrt{n + \rho}e_{j-n}, & n + 1 \leq j \leq 2n \\ 0, & 2n + 1 \end{cases}$$

y

$$\omega_j = \begin{cases} \frac{\rho}{n+\rho}, & j = 2n + 1 \\ \frac{1}{2(n+\rho)}, & 1 \leq j \leq 2n \end{cases}$$

donde  $\rho$  es una constante, y  $e_j$  es el  $j$ -ésimo vector unitario de  $\mathbb{R}^n$ . Julier et al. (1996) [37] propusieron un FKPS basado en una regla de cuadratura para funciones escalares, que es

idéntica a la regla de la cuadratura Gauss-Hermit, propusieron aproximar:

$$\int_{\mathbb{R}^n} \mu(x)\omega(x)dx \approx \sum_{i=1}^{2n+1} \mu(\chi_j)\omega(\chi_j)$$

Se requieren  $2n + 1$  puntos como mínimo necesario para preservar los primero y segundo momentos de la distribución Gaussiana multivariada. Se supone que  $\omega(\chi_j)$  es un peso de cuadratura constante en el tiempo, mientras que los puntos de cuadratura varían de acuerdo a la covarianza de los estados.

Los puntos sigma se generan de acuerdo al siguiente proceso:

$$\chi_{k,j} = \begin{cases} \hat{x}_{k-1}, & j = 2n + 1 \\ \hat{x}_{k-1} + \psi \sqrt{p_{k-1,j}}, & 1 \leq j \leq n \\ \hat{x}_{k-1} - \psi \sqrt{p_{k-1,j}}, & n + 1 \leq j \leq 2n \end{cases}$$

donde  $\hat{x}_{k-1}$  es el valor esperado de los estados,  $\sqrt{p_{k-1,j}}$  es la raíz cuadrada de la  $j$ -ésima columna asociada a la matriz varianza covarianza de los estados,  $\psi$  es un factor de escala que captura los efectos locales de la no linealidad de la función que se utiliza para aproximar.

#### Algoritmo FKPS.

- se inicializa en un tiempo  $t = 0$ :
 
$$\hat{x}_0 = \mathbb{E}(x_0) \quad ; \quad p_0 = \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T]$$
- predicción:
  - se implementan los puntos sigmas:
 
$$\chi_{k,j}^- = \begin{cases} \hat{x}_{k-1}, & j = 2n + 1 \\ \hat{x}_{k-1} + \psi \sqrt{p_{k-1,j}}, & 1 \leq j \leq n \\ \hat{x}_{k-1} - \psi \sqrt{p_{k-1,j}}, & n + 1 \leq j \leq 2n \end{cases}$$
- evolución de los puntos sigmas:
 
$$\chi_{k,j}^+ = f(\chi_{k,j}^-)$$
- estimación de los estadísticos:

$$x_k^- = \sum_{j=1}^{2n+1} \omega_j \chi_{k,j}^+$$

y

$$p_k^- = \sum_{j=1}^{2n+1} \omega^j (\chi_{k,j}^+ - x_k^-)(\chi_{k,j}^+ - x_k^-)^T$$

■ actualización:

- se calcula la ganancia de Kalman:

$$\mathcal{G}_k = p_k^- \mathcal{H}_k^T (\mathcal{H}_k p_k^- \mathcal{H}_k^T + R_k)^{-1}$$

- se actualizan las predicciones:

$$\hat{x}_k = x_k^- + \mathcal{G}_k (y_k - \mathcal{H}_k x_k^-)$$

y

$$\hat{p}_k = p_k^- - \mathcal{G}_k \mathcal{H}_k p_k^-$$

#### 4. Aplicación en el área de las finanzas

Para mostrar eficiencia de los algoritmos para estimar los estados y parámetros se consideraron dos modelos estocásticos del área financiera; el modelo de Cox Ingersell Ross (1985), y el modelo de Heston (1993) [38].

##### 4.1. El modelo de Cox Ingersell Ross (CIR)

El modelo de Cox, Ingersell y Ross [39], describe la evolución de la tasa de interés mediante la siguiente ecuación diferencial estocástica:

$$dr_t = \beta(\mu - r_t)dt + \sigma \sqrt{r_t}dB_t \quad (53)$$

donde:  $B_t$  es un movimiento Browniano (que modela el factor de riesgo aleatorio del mercado),  $\beta, \mu, y \sigma$  son parámetros. Utilizando la transformación  $x_t = \ln(r_t) - \ln(\sigma^2)$ , y aplicado el Lemma de Itô se tiene:

$$\begin{aligned} dx_t &= \frac{1}{r_t} \left\{ \beta(\mu - r_t)dt + \sigma \sqrt{r_t}dB_t \right\} - \frac{\sigma^2 r_t}{2r_t^2} dt \\ &= \left\{ \beta \left( \frac{\mu}{r_t} - 1 \right) - \frac{\sigma^2}{2r_t} \right\} dt + \sigma^2 \exp\left(-\frac{x_t}{2}\right) dB_t \end{aligned}$$

Usando la versión reparametrizada discutida en Chib et al. (2006), y Poyiadjs et al. (2011), se obtiene:

$$\begin{aligned} x_{t+1} &= \mu + x_t + \beta \exp(-x_t) + \exp\left(-\frac{x_t}{2}\right) \zeta_{t+1} \\ y_t &= \sigma_\zeta \exp\left(-\frac{x_t}{2}\right) \zeta_t \end{aligned} \quad (54)$$

Por facilidad, en este artículo sólo se estiman los estados  $x_{1:t} = (x_1, \dots, x_t)$ , suponemos los parámetros  $\Theta = (\mu, \beta, \sigma_\zeta)$  conocidos; ellos también pueden ser estimados usando la metodología propuesta. Sea, además  $\zeta_{t+1} \sim N(0, \sigma_\zeta^2)$ , y  $\xi_t \sim N(0, \sigma_\xi^2)$ , dada las observaciones  $y_{1:t} = (y_1, \dots, y_t)$ .

Para implementar los algoritmos, se utilizaron las siguientes distribuciones a priori para inicializar los parámetros en el modelo de CIR:

- FKE se tomó:  $\phi = 0.1; \sigma = 0.1, Q = 0.1, y R = 0.0001$ .
- FPG se tomó:  $\phi = 0.001; \sigma = 0.1, y Q = 0.0001$ .
- FPR se tomó:  $\phi = 0.1; \sigma = 0.1, y Q = 0.0001$ .
- FKEN se tomó:  $\phi = 0.1; \sigma = 0.1, y Q = 0.1; y R = 0.0001$ .
- FKPS se tomó:  $\phi = 0.1; \sigma = 0.1, y Q = 0.3; y R = 0.0001; \beta = 2; y \psi = 1$ .

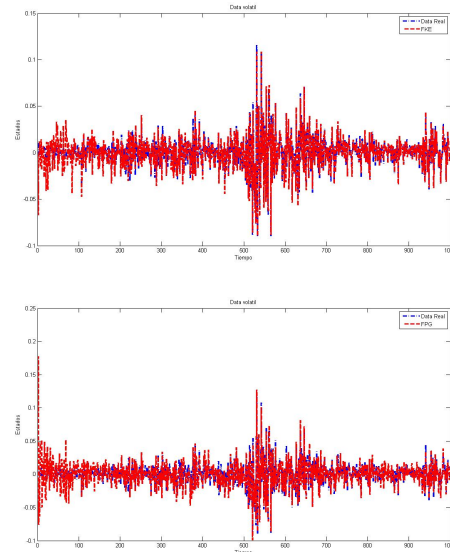


Figura 2: Estimación de los estados usando el FKE y FPG.

En las Figuras (2), (3), y (4), se muestran los estados estimados  $x_t$ , conjuntamente con los estados observados  $y_t$ , mediante los algoritmos: FKE, FPG, FPR, FKEN, y FKPS, respectivamente

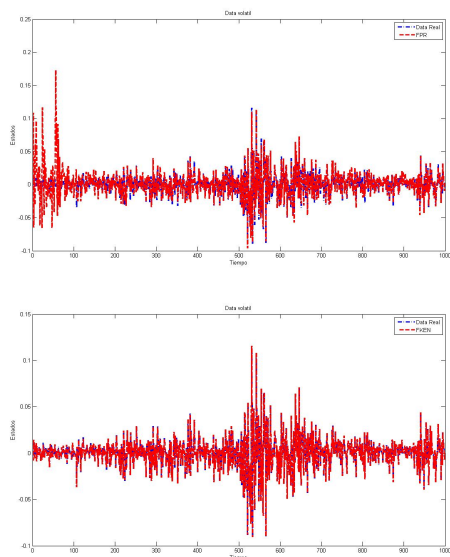


Figura 3: Estimación de los estados usando, el FPR y FKEN.

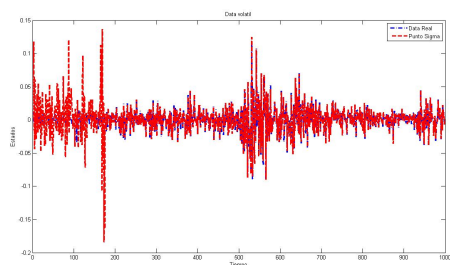


Figura 4: Estimación de los estados usando el FKPS.

Tabla 1: Raíz cuadrada del error cuadrático medio de los algoritmos para el modelo de CIR,

Filtros	FPG	FPR	FKE	FKEN	FKEN
RCECM	0.00016	0.0013	0.0015	0.000934	0.00063
T-CPU (seg)	0.02829	0.04039	0.32471	0.25477	0.10372

en el modelo CIR, observándose pocas diferencias entre los valores estimados y valores verdaderos.

En la Tabla (1) se muestran la raíz cuadrada del error cuadrático medio y los tiempos de ejecución CPU de los filtros implementados en el modelo de CIR, observándose poca variabilidad y tiempos de ejecución rápidos en todos los algoritmos implementados.

### 4.2. El modelo de Heston

El modelo de Heston es un modelo matemático usado para describir el comportamiento de un proceso estocástico bivariado entre el precio de las acciones  $s_t$  y su varianza  $v_t$ . Surgió inicialmente como una generalización del modelo de valoración de opciones Black-Scholes, pero suponiendo que la volatilidad es un proceso estocástico.

El modelo está gobernado por el siguiente sistema de ecuaciones diferenciales estocásticas:

$$ds_t = \mu s_t dt + \sqrt{v_t} s_t dB_{1,t}$$

$$d(\sqrt{v_t}) = -\beta \sqrt{v_t} dt + \delta dB_{2,t}$$

Si se hace  $\delta_t = (\sqrt{v_t})^2$ , y se aplica el Lemma de Itô se obtiene:

$$d\delta_t = 2\sqrt{v_t}d(\sqrt{v_t}) + \frac{1}{2}[2(d\sqrt{v_t})^2]$$

$$= 2\sqrt{v_t}(-\beta\sqrt{v_t}dt + \delta dB_{2,t}) + \delta^2 dt$$

$$= (\delta^2 - 2\beta v_t)dt + 2\delta\sqrt{v_t}dB_{2,t}$$

$$= 2\beta\left(\frac{\delta^2}{2\beta} - v_t\right)dt + 2\delta\sqrt{v_t}dB_{2,t}$$

haciendo  $k = 2\beta$ ;  $\theta = \frac{\delta^2}{2\beta}$ , y  $\sigma = 2\delta$ , se obtiene:  $d\delta_t = dv_t = k(\theta - v_t)dt + \sigma\sqrt{v_t}dB_{2,t}$ , luego el sistema dado previamente se transforma en:

$$ds_t = \mu s_t dt + \sqrt{v_t} s_t dB_{1,t}$$

$$dv_t = k(\theta - v_t)dt + \sigma\sqrt{v_t}dB_{2,t}$$

$$dB_{1,t}dB_{2,t} = \rho dt$$

donde las variables del sistema son definidas como sigue:

- $s_t$  : es el precio del activo,
- $v_t$  : es la volatilidad del activo,
- $\mu$  : es el retorno esperado del activo,
- $\theta$  : es la variación de precios a largo plazo,
- $k$  : es la velocidad a la cual la volatilidad tiende hacia su media de largo plazo,
- $\sigma$  : es la volatilidad de  $\sqrt{v_t}$ ,

- $\rho$  : es la correlación de los movimientos Browniano,
- $dt = t_k - t_{k-1}$  : es un pequeño incremento en el tiempo,
- $dB_{1,t}$  : es un movimiento Browniano estándar unidimensional,
- $dB_{2,t}$  : es un movimiento Browniano estándar unidimensional.

El modelo de Heston versión discretizada es:

$$v_{t+1} = v_t + \kappa(\theta - v_t) + \sqrt{v_t}\sigma(\rho z_1 + \sqrt{1 - \rho^2}z_2) \quad ; \quad (E.E)$$

$$y_{t+1} = y_t + \left(\mu - \frac{1}{2}v_t\right) + \sqrt{v_t}z_1 \quad ; \quad (E.O)$$

donde  $y_{t+1} = \ln(s_{t+1})$ , y los parámetros del modelo son:  $\Theta = (\kappa, \theta, \mu, \sigma, \rho)$ .

Para implementar los algoritmos, se utilizaron las siguientes distribuciones a priori para inicializar los parámetros en el modelo de Heston:

- *FKE* se tomó:
- *FPG* se tomó:  $\sigma = 100$ ;  $\kappa = 3.3065$ ;  $\theta = 0.0215$ ;  $\rho = -0.9108$ ;  $\phi = 0.05$ ;  $\delta = 0.0001$ ; y  $Q = 0.0001$ .
- *FPR* se tomó:  $\sigma = 100$ ;  $\kappa = 3.3065$ ;  $\theta = 0.0215$ ;  $\rho = -0.9108$ ;  $\phi = 0.05$ ;  $\delta = 0.0001$ ; y  $Q = 0.0001$ .
- *FKEN* se tomó:  $\sigma = 100$ ;  $\kappa = 3.3065$ ;  $\theta = 0.0215$ ;  $\rho = -0.9108$ ;  $\phi = 0.05$ ;  $\delta = 0.0001$ ; y  $Q = 0.0001$ .
- *FKPS* se tomó:  $\sigma = 100$ ;  $\kappa = 3.3065$ ;  $\theta = 0.0215$ ;  $\rho = -0.9108$ ;  $\phi = 0.05$ ;  $\delta = 0.0001$ ;  $Q = 0.0001$ ;  $\beta = 2$ ; y  $\psi = 1$ .

En las Figuras (5), (6), y (7), se muestran los estados estimados  $v_t$ , conjuntamente con los estados observados  $y_t$ , mediante los algoritmos: FKE, FPG, FPR, FKEN, y FKPS, respectivamente generados por el modelo de Heston, observándose pocas diferencias entre los valores estimados y valores verdaderos. En la Tabla (2) se muestran

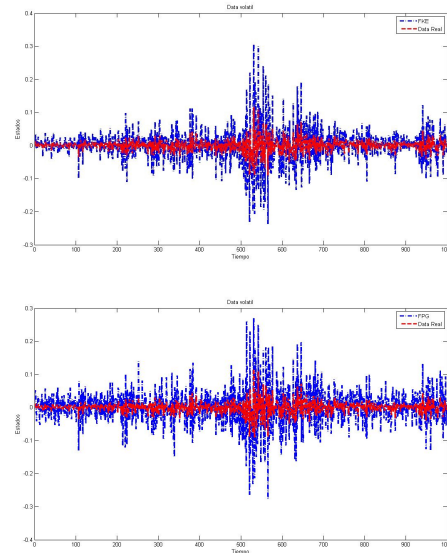


Figura 5: Estimación de los estados usando el FKE y FPG.

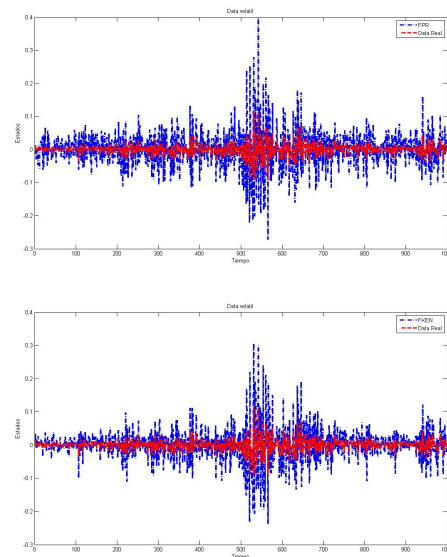


Figura 6: Estimación de los estados usando el FPR y FKEN.

la raíz cuadrada del error cuadrático medio y los tiempos de ejecución CPU de los filtros implementados en el modelo de CIR, observándose poca variabilidad y tiempos de ejecución rápidos para todos los algoritmos probados.

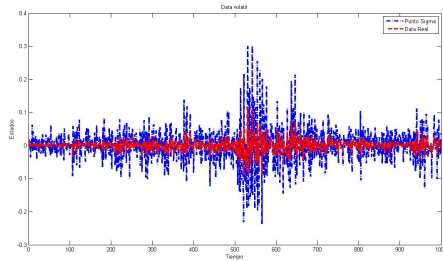


Figura 7: Estimación de los estados usando el FKPS.

Tabla 2: Raíz cuadrada del error cuadrático medio modelo de Heston.

Filtros	FPG	FPR	FKE	FKEN	FKEN
RCECM	0.0019	0.0018	0.0021	0.0012	0.0013
T-CPU (seg)	0.58342	0.23740	0.62397	0.32725	0.49664

## 5. Discusiones y conclusiones

La inferencia estadística sobre procesos estocásticos de tiempo continuo tales como los sistemas dinámicos es un área de investigación activa desde el punto de vista de la matemática pura así como también de la matemática aplicada. Es conocido que un sistema dinámico puede ser caracterizado por un modelo espacio estado. La formulación en términos de los modelos espacio estado permite: cuantificar el error de las ecuaciones de estado y observación; establecer la relación entre el proceso físico y el modelo estadístico, y estimar los estados ocultos a partir de las medidas obtenidas con errores. En este trabajo se proponen cinco algoritmos recursivos: FKE, FPG, FPR, FKEN y FKPS; que permiten hacer inferencia en procesos de tiempo discretos con observaciones tomadas en tiempo continuo, las cuales pueden tener estructuras lineales, no lineales, errores Gaussianos, y no Gaussianos. La metodología propuesta es ilustrada utilizando dos ecuaciones diferenciales estocásticas que provienen de las finanzas. Los algoritmos utilizados permiten generar muestras para la reconstrucción aproximada de las distribuciones marginales a posteriori de los estados desconocidos. En el estudio empírico, se demuestra que los filtros tienen buena performance

en la estimación de los estados; además, no se observó mucha diferencia entre los valores estimados y los valores verdaderos. La calidad de estimación de los algoritmos es valorada a través del cálculo de la raíz del error cuadrático medio obteniéndose errores de estimación pequeños, además se calcularon los tiempos de ejecución de los algoritmos obteniéndose resultados satisfactorios.

## Referencias

- [1] Simo Särkkä and Arno Solin. Lecture notes on applied stochastic differential equations, 2014. Version as of December 4, 2014.
- [2] Simo Särkkä. *Bayesian filtering and smoothing*. Cambridge University Press, Cambridge, 2013.
- [3] Andrew Jazwinski. Stochastic processes and filtering theory. In *Mathematics in Science and Engineering*, volume 64. Elsevier, 1970.
- [4] Bernt Øksendal. *Stochastic Differential Equations. An Introduction with Applications*, chapter 5. Stochastic Differential Equations, pages 65–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
- [5] Stefano M. Iacus. *Simulation and inference for stochastic differential equations. With R examples*. Springer-Verlag, New York, 2008.
- [6] Stefano M. Iacus. *Option pricing and estimation of financial models with R*. John Wiley & Sons, 2011.
- [7] Umberto Picchini. Inference for sde models via approximate bayesian computation. *Journal of Computational and Graphical Statistics*, 23(4):1080–1100, 2014.
- [8] Andrew C Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1990.
- [9] Jesús Fernández-Villaverde and Juan F Rubio-Ramírez. Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *Journal of Applied Econometrics*, 20(7):891–910, 2005.
- [10] Neil Shephard and Michael K. Pitt. Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84(3):653–667, 1997.
- [11] Tsunehiro Ishihara and Yasuhiro Omori. Efficient bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors. *Computational Statistics & Data Analysis*, 56(11):3674–3689, 2012.
- [12] Seung-Ho Yang and Jaewook Lee. Predicting a distribution of implied volatilities for option pricing. *Expert Systems with Applications*, 38(3):1702–1708, 2011.
- [13] Tatiana Miazhynskaia, Sylvia Frühwirth-Schnatter, and Georg Dorffner. Bayesian testing for non-linearity in



- volatility modeling. *Computational statistics & data analysis*, 51(3):2029–2042, 2006.
- [14] Elena Lazkano, Basilio Sierra, Aitzol Astigarraga, and José María Martínez-Otzeta. On the use of bayesian networks to develop behaviours for mobile robots. *Robotics and Autonomous Systems*, 55(3):253–265, 2007.
- [15] Mónica Viviana Alvarado Mora, Camila Malta Romano, Michele Soares Gomes-Gouvêa, Maria Fernanda Gutierrez, Livia Botelho, Flair José Carrilho, and João Renato Rebello Pinho. Molecular characterization of the hepatitis b virus genotypes in colombia: A bayesian inference on the genotype f. *Infection, Genetics and Evolution*, 11(1):103–108, 2011.
- [16] David Allen and Adnan Darwiche. RC.Link: Genetic linkage analysis using Bayesian networks. *International journal of approximate reasoning*, 48(2):499–525, 2008.
- [17] A. Biedermann and F. Taroni. Bayesian networks for evaluating forensic DNA profiling evidence: A review and guide to literature. *Forensic Science International: Genetics*, 6(2):147–157, 2012.
- [18] Andrew Golightly and Darren J. Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*, 1(6):807–820, 2011.
- [19] E. L. Ionides. Inference and filtering for partially observed diffusion processes via sequential Monte Carlo. Technical Report 405, Department of Statistics, The University of Michigan,, Michigan, 2004.
- [20] E. L. Ionides, C. Bretó, and A. A. King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443, 2006.
- [21] Fredrik Lindsten, Michael I. Jordan, and Thomas B. Schön. Particle gibbs with ancestor sampling. *The Journal of Machine Learning Research*, 15(1):2145–2184, 2014.
- [22] R. Vieira and D. J. Wilkinson. Online state and parameter estimation in dynamic generalised linear models. *ArXiv e-prints*, august 2016.
- [23] Jonathan Law and Darren J. Wilkinson. Composable models for online bayesian analysis of streaming data. *Statistics and Computing*, pages 1–19, Oct 2017.
- [24] Fabio Sigrist, Hans R. Künsch, and Werner A. Stahel. A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *The Annals of Applied Statistics*, 6(4):1452–1477, 2012.
- [25] Saba Infante, César Luna, Luis Sánchez, and Aracelis Hernández. Approximations of the solutions of a stochastic differential equation using Dirichlet process mixtures and gaussian mixtures. *Statistics, Optimization & Information Computing*, 4(4):289–307, 2016.
- [26] Saba Infante, Luis Sánchez y Fernando Cedeño. Filtros para predecir incertidumbre de lluvia y clima. *Revista de Climatología*, 12:33–48, 2012.
- [27] Luis Sánchez, Saba Infante, Victor Griffin, and Demetrio Rey. Spatio-temporal dynamic model and parallelized ensemble kalman filter for precipitation data. *Brazilian Journal of Probability and Statistics*, 30(4):653–675, 2016.
- [28] Luis Sánchez, Saba Infante, Jose Marcano, and Victor Griffin. Polynomial chaos based on the parallelized ensemble kalman filter to estimate precipitation states. *Statistics, Optimization & Information Computing*, 3(1):79–95, 2015.
- [29] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, 2002.
- [30] Arnaud Doucet, Nando de Freitas, and Neil Gordon. *An Introduction to Sequential Monte Carlo Methods*, chapter 1, pages 3–14. Springer New York, New York, NY, 2001.
- [31] Augustine Kong, Jun S. Liu, and Wing Hung Wong. Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–289, 1994.
- [32] Jun S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.
- [33] B.D. Rubin. *Bayesian Statistics 3*, chapter Using the SIR Algorithm to Simulate Posterior Distributions, pages 395–402. Oxford University Press, Oxford, 1988.
- [34] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- [35] Geir Evensen and Peter Jan Van Leeuwen. Assimilation of geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Monthly Weather Review*, 124(1):85–96, 1996.
- [36] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, 2000.
- [37] Simon J. Julier and E. L. Uhlman. A general method for approximating nonlinear transformations of probability distributions. Technical Report, Department of Engineering Science, University of Oxford, Oxford, 1996.
- [38] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- [39] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–408, 1985.