

Filtros con Aprendizaje de Parámetros para Optimizar Modelos de Redes Neuronales en la Predicción de Series de Precipitaciones

(Filters with Parameters Learning for Optimizing Neural Network Models in Predicting Rainfall Series)

Saba Infante, Fernando Cedeño, José Ortega

Palabras Clave: Redes Neuronales; Filtro de Kalman Extendido; Filtro de Partículas; Series de Precipitaciones.

Key Words: Neural Networks; Kalman Extended Filter; Particle Filter; Precipitation Series

RESUMEN

En este estudio se usó un modelo de redes neuronales (RN) para predecir los niveles de lluvia en Venezuela. El estudio está basado en series de lluvia total mensual del periodo comprendido entre 1971-2000 de 36 estaciones meteorológicas. Para determinar la dimensión de inmersión de los datos, así como el tiempo de retardo, número de pasos hacia atrás necesarios para predecir un valor futuro, se utilizan dos técnicas estándar de los sistemas dinámicos, la Información Mutua Promedio (AMI) y los Falsos Vecinos más Cercanos (FNN). Una vez determinadas las variables de entrada, se diseña el modelo predictivo formulado en términos de los modelos espacio-estado. Para estimar los pesos de la red se utilizaron dos algoritmos de aprendizaje Filtro de Kalman Extendido (FKE) y Filtro de Partículas o Muestreo de Importancia Secuencial (FP o SIR). Se utilizó el criterio de información Bayesiano (BIC) para seleccionar el número de neuronas de la capa oculta. Se proponen dos métodos de

bondad de ajuste para medir la exactitud del modelo. Los resultados demuestran que el modelo ajustado es útil para predecir niveles de precipitaciones mensuales con alta exactitud y confiabilidad.

ABSTRACT

In this study it was used a neural network model (NN) to predict rainfall levels in Venezuela. The study is based on a series of monthly total rainfall in the period 1971-2000 from 36 meteorological stations. One determine the embedding dimension of the data and the delay time, the number of backward steps necessary to predict a future value, two standard techniques of dynamical systems, the Average Mutual Information (AMI) and the False Neighbors Nearest (FNN). After determining the input variables, were used the predictive model was designed in terms of state-space models. To estimate the weights of the network we used two learning algorithms, the Extended Kalman Filter (FKE) and a Particle Filter or Sequential Importance

Sampling (PF or SIR). We used the Bayesian information criterion (BIC) to select the number of neurons in the hidden layer and we proposed two methods to measure goodness of fit of the model's accuracy. The results show that the fitted model is useful in predicting monthly rainfall levels with high accuracy and reliability.

INTRODUCCIÓN

Durante muchas décadas, físicos, matemáticos, astrónomos, meteorólogos, economista, biólogos e investigadores en las ciencias médicas han tratado de comprender la complejidad de la naturaleza del mundo real. Todos los procesos físicos involucran variabilidad en el espacio y en el tiempo; por ejemplo, los meteorólogos por lo general están interesados en la evolución en el tiempo de ciertos parámetros como: el viento, la temperatura y las precipitaciones sobre un dominio espacial de interés. Por otra parte, las bases de datos de estos fenómenos son típicamente muy grandes, con muchos datos faltantes durante ciertos periodos de tiempo, con comportamientos irregulares y en algunos casos con dinámicas caóticas. La teoría del caos provee herramientas para observar el orden y reglas para determinar el comportamiento del sistema dinámico. Las técnicas de modelación utilizadas por Abarbanel (1996), Fraser y Swinney (1986) permiten caracterizar las series temporales generadas por estos sistemas dinámicos.

En este trabajo se emplean como herramientas matemáticas la función AMI (Información Mutua Promedio o Average Mutual Information) y la función FNN (Falsos vecinos más cercanos o False Nearest Neighbors), con el objetivo de reconstruir la dinámica y el tiempo de retardo del sistema dinámico, y luego sobre la base de esta información se introduce un modelo de redes neuronales que tiene como entrada las miniserias obtenidas a partir de la serie temporal original usando los métodos AMI y FNN. Existe una extensa literatura de modelos estadísticos de RN, (véase Cheng y Titterington, 1994; Bishop, 1995; Ripley, 1994 y 1996; Neal, 1996; Titterington, 2004; Rajunkar et al., 2002; Menchero et al., 2005; Bustami et al., 2007; Saffari et al., 2009).

Entre algunos trabajos recientes relacionados con las funciones AMI y FNN para determinar la dinámica del sistema dinámico y luego utilizar un modelo de RN entrenado por el algoritmo de Backpropagation para estimar los parámetros y predecir, están los de Calderón y Pérez (2006), Lucio et al. (2007) e Infante et al. (2008).

En este trabajo se propone usar una estrategia metodológica diferente para estimar los parámetros del modelo de RN en forma secuencial y en tiempo real usando una formulación de los modelos espacio estado, mediante el algoritmo filtro de Kalman extendido (FKE), De Freitas et al. (1998) y Lary y Mussa

(2004) y el algoritmo de filtro de partículas (FP), Doucet et al. (2000), De Freitas et al. (2000), Morais et al. (2008) e Infante et al. (2009).

El algoritmo de aprendizaje de FKE permite obtener mejores resultados que las técnicas de estimación convencionales tales como el algoritmo de Backpropagation, ya que hace uso de los estadísticos de segundo orden (covarianzas) que son esenciales para propagar el error sobre las predicciones, además muestra excelente convergencia en los parámetros estimados, debido a que se tiene bastante memoria en la base del computador y la precisión de la máquina es alta. El FKE se basa en la expansión analítica de una serie de Taylor de un sistema no lineal dado por la ecuación de observación alrededor de los estimadores máximo verosímil; que en este caso son: la media y covarianza condicional a posteriori. También se propone usar el algoritmo de FP como método para entrenar el modelo de redes neuronales. La estrategia permite que los parámetros de aprendizaje (pesos de la red) se actualicen en forma secuencial utilizando una distribución de probabilidad a posteriori. La técnica es adecuada para modelaje de datos que tengan comportamiento no estacionario, estructuras no lineales, y que se requieran estimar en tiempo real; además, tiene la ventaja que no se requiere de ninguna restricción sobre la distribución y la forma funcional. La idea principal del algoritmo de FP es la

siguiente: el espacio estado es particionado en muchas regiones, las cuales son rellenadas por partículas de acuerdo alguna medida de probabilidad. Las partículas con la mayor probabilidad son seleccionadas. El sistema de partículas se actualiza a lo largo del tiempo de acuerdo a dos ecuaciones, una de estado o transición y la otra de observación o evolución, las cuales involucran funciones de densidad de probabilidad que se estiman recursivamente en dos pasos denominados, predicción y actualización.

El resto del artículo es organizado como sigue: En la primera sección se plantea la metodología a seguir, se define: el modelo de RN en forma espacio estado, y se detallan los algoritmos filtro del Kalman extendido y filtro de partículas; se definen las medidas utilizadas para determinar la exactitud de predicción del modelo de RN y el método de transformación de los datos. En la siguiente Sección se muestra una discusión de los resultados obtenidos; y en la Sección final se establecen las conclusiones obtenidas.

METODOLOGÍA

En lo que sigue se definen: el modelo de RN en forma de espacio estado y los algoritmos utilizados para estimar los parámetros del modelo. El objetivo de utilizar un modelo RN conjugado con las técnicas de los sistemas dinámicos es que en primer lugar: las precipitaciones

se comportan como un sistema caótico no lineal que es difícil de controlar porque presenta inestabilidad y sensibilidad a las condiciones iniciales; la teoría de caos permite establecer un orden; en segundo lugar, se quiere utilizar la capacidad de las RN para generalizar y en este trabajo se proponen dos estrategias de optimización para estimar los pesos del modelo, que garantizan el aprendizaje de los parámetros en forma eficiente y con mayor control sobre el sistema estudiado.

Los trabajos de Takens (1981) y Casdagli (1989) entre otros, han establecido la metodología necesaria para modelar series de tiempo que tienen un comportamiento no lineal. A partir de una serie de tiempo univariada, $X = \{x_1, x_2, \dots, x_n\}$ se construyen los vectores de estados $x^i(t) = \{x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)\}$; donde: $x_1 = x(t)$, $x_2 = x(t - \tau), \dots, x_m = x(t - (m - 1)\tau)$, $x_1 = x(t)$, $x_2 = x(t - \tau), \dots, x_m = x(t - (m - 1)\tau)$, ($m \leq n$) y m representa la dimensión de inmersión y τ es el tiempo de retardo. Según el teorema de Takens (1981), la trayectoria geométrica de esta secuencia de vectores conforma un objeto multidimensional en \mathcal{R}^m que mantiene inalterada las características del verdadero pero desconocido proceso generado de los datos. Además el teorema garantiza la existencia de una función $g: \mathbb{R}^m \rightarrow \mathbb{R}$ que relaciona el estado actual $x(t)$ y el estado futuro $x(t + \Delta t)$, mediante la ecuación $x(t + \Delta t) =$

$g(x^i(t))$, esto quiere decir que se puede anticipar la dinámica futura de la serie temporal considerando únicamente los valores del pasado. El problema ahora consiste en encontrar un buen predictor $\hat{g}(\cdot)$ que aproxime a $g(\cdot)$.

En este trabajo se utilizará un modelo de RN para aproximar a $g(\cdot)$, formulado en términos de los modelos espacio estado, donde la ecuación de estado (transición) describe la evolución de los pesos de la red, mientras que la ecuación de observación (medida) describe la relación no lineal de las entradas y salidas de un proceso físico que se quiere estudiar.

Las entradas de la red se determinan en función de m y τ mediante las técnicas dadas en Abarbanel (1996), Fraser y Swinney (1986). El modelo en la forma espacio estado se define como:

$$w_{t+1} = Tw_t + u_t \quad (1.a)$$

$$y_t = g_t(w_t, x_t) + v_t \quad (1.b)$$

La ecuación (1.a) se llama ecuación de estado o transición; la ecuación (1.b) se llama ecuación de observación o medida; $y_t \in \mathbb{R}$ es un vector de respuesta conocido (valor supervisado), $x_t = x^i(t) \in \mathbb{R}^m$ es el vector de entrada; $w_t \in \mathbb{R}^q$ es el vector de estado que contiene los pesos de la red neuronal, $g_t(w_t, x_t)$ es una es una función iterativa que describe el mapeo no lineal de la red; $\hat{y}_t = \hat{g}_t(\hat{w}_t, x_t)$ es el valor de la respuesta estimada, $SCE = \frac{1}{N} \sum_{i=1}^N (y_t - \hat{y}_t)^2$ es la suma de cuadrados del error de la función de predicción, μ_t es un

vector de ruido de la ecuación de estado; u_t es un vector de ruido de la ecuación de observación y T es una matriz de constantes conocidas. Es ampliamente conocido que el modelo RN tiene la capacidad de aproximar cualquier función continua, con la precisión deseada independientemente del tamaño de la red, Cybenko (1989).

Los supuestos del modelo son los siguientes: μ_t es un ruido blanco con media cero y matriz de varianza covarianza $E(\mu_t \mu_t^T) = Q_t$, v_t también es un ruido blanco con media cero y matriz de varianza covarianza $E(v_t v_t^T) = R_t$, y $E(\mu_t v_k^T) = 0$, para todo k y t . Los ruidos v_t y μ_t se consideran no correlacionados con los pesos de la red y las condiciones iniciales. La evolución de los estados corresponden a proceso de Markov de primer orden con probabilidad inicial $P(w_0)$ y una probabilidad de transición dada por $P(w_t|w_{t-1})$. Las observaciones se consideran condicionalmente independiente dado los estados, es decir; $P(y_{1:t}|w_{1:t}) = \prod_{i=1}^t P(y_i|w_i)$, donde: $y_{1:t} = \{y_1, \dots, y_t\}$ y $w_{1:t} = \{w_1, \dots, w_t\}$, (véase Harvey, 1970).

En la ecuación (2) se muestra la distribución conjunta de los estados y las observaciones, que se obtiene usando la regla de la cadena de probabilidad, es decir; $P(w_{1:t}, y_{1:t})$ viene dada por:

$$P(w_0) \prod_{i=1}^t P(w_i|w_{i-1}) \prod_{i=1}^t P(y_i|w_i) \quad (2)$$

Entonces usando las técnicas Bayesianas, se puede estimar la densidad a posteriori de $P(w_{1:t}|y_{1:t})$. En muchas aplicaciones, el interés recae en la estimación de las marginales, también llamada densidad filtrada $P(w_t|y_{1:t})$. Conociendo la densidad, se pueden calcular varios estimadores de los pesos de la red, como la media, mediana, varianza e intervalos de confianza. Para obtener el estimador filtrado $P(w_t|y_{1:t})$ se requiere conocer de una distribución a priori $P(w_0)$ y de una distribución a posteriori $P(w_{t-1}|y_{1:t-1})$ en el tiempo $t-1$. Entonces el estimador filtrado se estima recursivamente en dos pasos denominados predicción y actualización:

1. Predicción:

$$P(w_t|y_{1:t-1}) = \int P(w_t|w_{t-1}) * P(w_{t-1}|y_{1:t-1}) dw_{t-1} \quad (3)$$

donde:

$$P(w_t|w_{t-1}) = \int P(w_t|u_{t-1}, w_{t-1}) * P(u_{t-1}|w_{t-1}) du_{t-1} = \int \delta(w_t - u_{t-1} - Tw_{t-1}) * P(u_{t-1}) du_{t-1} \quad (4)$$

$\delta(\cdot)$ denota un punto de masa de la función Delta de Dirac y $P(u_{t-1}|w_{t-1}) = P(u_{t-1})$.

2. Actualización

$$P(w_t|y_{1:t}) = \frac{P(y_t|w_t)P(w_t|y_{1:t-1})}{P(y_t|y_{1:t-1})} \quad (5)$$

donde:

$$P(y_t|w_t) = \int \delta[y_t - g_t(w_t, x_t) - v_t]$$

$$* P(v_t)dv_t \quad (6)$$

y

$$P(y_t|y_{1:t-1}) = \int P(y_t|w_t) * P(w_t|y_{1:t-1})dw_t \quad (7)$$

Las relaciones dadas por las ecuaciones (3) y (5), constituyen la solución formal del problema de estimación recursiva mediante los métodos Bayesianos.

La solución óptima de la ecuación (5) exige calcular integrales multidimensionales, lo que hace imposible obtener una solución analítica. Entonces para obtener aproximaciones de las integrales se usan métodos numéricos o métodos de simulación Monte Carlo.

En este artículo se realizan dos propuestas alternativas para estimar los pesos del modelo de RN basadas en los algoritmos de aprendizajes: filtro de Kalman extendido y filtro de partículas. El objetivo final es encontrar valores para los pesos del modelo que minimicen la suma de cuadrados del error (SCE). Finalmente se señala, para simplificar el trabajo no se trata el problema de estimar los parámetros de los ruidos blancos Q_t , R_t dadas las observaciones $y_{1:t}$.

1. Algoritmo de Aprendizaje Filtro de Kalman Extendido (FKE)

Considérese el sistema dinámico no lineal en forma de espacio estado, dado en (1.a) y (1.b), con las hipótesis usuales;

esto es, $u_t \sim N(0, Q_t)$, $v_t \sim N(0, R_t)$ y $w_0 \sim N(\mu_0, \Sigma_0)$. El objetivo del FKE es estimar los estados no observados $\hat{w}_{1:t} = \{\hat{w}_1, \dots, \hat{w}_t\}$ basándose en las observaciones $y_{1:t}$. El algoritmo permite obtener una aproximación del estimador óptimo. La no linealidad en la segunda ecuación dada en (1.b) es aproximada por una versión linealizada alrededor del último estado estimado, esto $g_t(w_t, x_t)$ es aproximada por:

$$g(\hat{w}_t, x_t) + \frac{\partial g_t}{\partial w_t} |_{(w_t=\hat{w}_t)}(w_t - \hat{w}_t) + \dots \quad (8)$$

Para obtener las ecuaciones de predicción y actualización del algoritmo, supóngase que $E(w_{t-1}|y_{1:t-1}) = m_{t-1|t-1}$ y $V(w_{t-1}|y_{1:t-1}) = p_{t-1|t-1}$ se conocen en el tiempo $t - 1$. El predictor de w_t y su matriz asociada al error cuadrático medio en el tiempo $t - 1$ están dadas por:

$$\begin{aligned} E(w_t|y_{1:t-1}) &= E(Tw_{t-1} + u_t|y_{1:t-1}) \\ &= TE(w_{t-1}|y_{1:t-1}) \\ &= Tm_{t-1|t-1} \end{aligned} \quad (9)$$

$V(w_t|y_{1:t-1})$ es dada por:

$$\begin{aligned} &= V(Tw_{t-1} + u_{t-1}|y_{1:t-1}) \\ &= TV(w_{t-1}|y_{1:t-1})T^T + Q_t \\ &= Tp_{t-1|t-1}T^T + Q_t \\ &= p_{t|t-1} \end{aligned} \quad (10)$$

$$\begin{aligned} E(y_t|y_{1:t-1}) &= E[g_t(w_t, x_t) + v_t|y_{1:t}] \\ &= g_t(w_t, x_t) \end{aligned} \quad (11)$$

$$V(y_t|y_{1:t-1}) = J_t V(w_t|y_{1:t-1}) J_t^T + R_t$$

$$= J_t p_{t|t-1} J_t^T + R_t \quad (12)$$

donde:

$$J_t = \left[\frac{\partial g_t}{\partial w_t} \Big|_{(w_t = \hat{w}_t)} \right]$$

$C(w_t, y_t | y_{1:t-1})$ es dada por:

$$= C[w_t, E(y_t | w_t) | y_{1:t-1}]$$

$$= C[w_t, g_t(\hat{w}_t, x_t)$$

$$+ J_t (w_t - \hat{w}_t) + \dots | y_{1:t-1}]$$

$$= V(w_t | y_{1:t-1}) J_t$$

$$= p_{t|t-1} J_t \quad (13)$$

Las ecuaciones del ajuste lineal de Bayes, están dadas por:

$$E(w_t | y_{1:t}) = E(w_t | y_{1:t-1})$$

$$+ C(w_t, y_t | y_{1:t-1}) V(y_t | y_{1:t-1})^{-1}$$

$$* [y_t - E(y_t | y_{1:t-1})] \quad (14)$$

$$V(w_t | y_{1:t}) = V(w_t | y_{1:t-1})$$

$$- C(w_t, y_t | y_{1:t-1}) V(y_t | y_{1:t-1})^{-1}$$

$$* C(w_t, y_t | y_{1:t-1})^T \quad (15)$$

Usando las ecuaciones (14) y (15) se obtiene la actualización del estimador óptimo del estado w_t dada la información $y_{1:t}$ en el tiempo t , es decir:

$$m_{t|t} = E(w_t | y_{1:t}) = T m_{t-1|t-1}$$

$$+ p_{t|t-1} J_t [J_t p_{t|t-1} J_t^T + R_t]^{-1}$$

$$* [y_t - g_t(\hat{w}_t, x_t)] \quad (16)$$

$$p_{t|t} = V(w_t | y_{1:t})$$

$$= p_{t|t-1} - p_{t|t-1} J_t [J_t p_{t|t-1} J_t^T + R_t]^{-1}$$

$$* [p_{t|t-1} J_t]^T \quad (17)$$

Bajo los supuestos de normalidad para los errores, algoritmo FKE se resume:

#####

Algoritmo 1: FK E

#####

Paso 1 Se muestrea w_{t-1} de:

$$p(w_{t-1} | y_{1:t-1}) = N(m_{t-1|t-1}, p_{t-1|t-1}) \quad (18)$$

Paso 2 Se muestrea w_t de:

$$p(w_t | y_{1:t-1}) = N(m_{t|t-1}, p_{t|t-1}) \quad (19)$$

Paso 3 Se muestrea w_t de:

$$p(w_t | y_{1:t}) = N(m_{t|t}, p_{t|t}) \quad (20)$$

#####

2. Algoritmo de Aprendizaje Filtro de Partículas (FP) o Muestreo de Importancia Secuencial (SIR)

La estrategia de predicción y actualización dada en las ecuaciones (3) y (5) permite obtener una solución óptima al problema de inferencia, pero las integrales involucradas son difíciles de calcular. En muchas aplicaciones, las soluciones analíticas no son posibles. Se

requieren métodos alternativos, tales como: integración numérica, aproximación gaussiana, y los métodos de simulación Monte Carlo por Cadenas de Markov y las técnicas de Monte Carlo Secuencial. Se puede aproximar la función de distribución a posteriori usando una función sobre un soporte discreto finito mediante la técnica del muestreo de importancia; es decir, se hace uso de la aproximación Monte Carlo, dada:

$$\hat{p}(w_{1:t}|y_{1:t}) = \frac{1}{n} \sum_{i=1}^n \delta(w_{1:t} - w_{1:t}^{(i)}) \quad (21)$$

donde: $w_{1:t}^{(i)}$ representa la muestra usada para aproximar la distribución a posteriori y $\delta(\cdot)$ es un punto de masa de la función delta de Dirac. La ecuación (21) permite aproximar la distribución a posteriori. El valor esperado de algún funcional de los pesos, es dado por la ecuación (22):

$$\begin{aligned} E(h_t(w_{1:t})) &= \int h_t(w_{1:t}) p(w_{1:t} | y_{1:t}) dw_{1:t} \\ &= \int h_t(w_{1:t}) \frac{p((w_{1:t}|y_{1:t}))}{\pi(w_{1:t}|y_{1:t})} \pi(w_{1:t}|y_{1:t}) dw_{1:t} \\ &= \int h_t(w_{1:t}) \frac{p((w_{1:t}|y_{1:t})p(w_{1:t}))}{p(y_{1:t})\pi(w_{1:t}|y_{1:t})} \\ &\quad * \pi(w_{1:t}|y_{1:t}) dw_{1:t} \\ &= \int h_t(w_{1:t}) \frac{q_t(w_{1:t})}{p(y_{1:t})} \pi(w_{1:t}|y_{1:t}) dw_{1:t} \quad (22) \end{aligned}$$

donde: la variable $q_t(w_{1:t})$ se conoce como el cociente de importancia no normalizado, dado por:

$$q_t(w_{1:t}) = \frac{p(y_{1:t}|w_{1:t})p(w_{1:t})}{\pi(w_{1:t}|y_{1:t})} \quad (23)$$

es decir; $E(h_t(w_{1:t}))$ es dada por:

$$\begin{aligned} &= \frac{1}{p(y_{1:t})} \int h_t(w_{1:t}) q_t(w_{1:t}) \pi(w_{1:t}|y_{1:t}) dw_{1:t} \\ &= \frac{\int h_t(w_{1:t}) q_t(w_{1:t}) \pi(w_{1:t}|y_{1:t}) dw_{1:t}}{\int p(y_{1:t}|w_{1:t}) p(w_{1:t}) \frac{\pi(w_{1:t}|y_{1:t})}{\pi(w_{1:t}|y_{1:t})} dw_{1:t}} \\ &= \frac{\int h_t(w_{1:t}) q_t(w_{1:t}) \pi(w_{1:t}|y_{1:t}) dw_{1:t}}{\int q(w_{1:t}) \pi(w_{1:t}|y_{1:t}) dw_{1:t}} \\ &= \frac{E_{\pi}[q_t(w_{1:t})h_t(w_{1:t})]}{E_{\pi}[q_t(w_{1:t})]} \quad (24) \end{aligned}$$

Por lo tanto generando muestras $w_{1:t}^{(i)}$ de la función de distribución propuesta $\pi(\cdot)$, se aproxima el valor esperado mediante el estimador dado:

$$\begin{aligned} E(h_t(w_{1:t})) &\approx \frac{\frac{1}{n} \sum_{i=1}^n h_t(w_{1:t}^{(i)}) q_t(w_{1:t}^{(i)})}{\frac{1}{n} \sum_{i=1}^n q_t(w_{1:t}^{(i)})} = \\ &= \sum_{i=1}^n h_t(w_{1:t}^{(i)}) \tilde{q}_t(w_{1:t}^{(i)}) \quad (25) \end{aligned}$$

donde: $\tilde{q}_t(\cdot)$ es el cociente de importancia normalizado dado por:

$$\tilde{q}_t^{(i)} = \frac{q_t^{(i)}}{\sum_{j=1}^n q_t^{(j)}} \quad (26)$$

Entonces la distribución a posteriori se estima por:

$$p(w_{1:j}|y_{1:t}) = \sum_{i=1}^n \tilde{q}_t^{(i)} \delta(w_{1:t} - w_{1:t}^{(i)}) \quad (27)$$

Para calcular un estimador secuencial de la distribución a posteriori en un tiempo t sin modificar los estados simulados $w_{1:t-1}$ se considera la siguiente densidad de probabilidad propuesta:

$$\pi(w_{1:t}|y_{1:t}) = \pi(w_0|y_0)$$

$$*\prod_{j=1}^t \pi(w_j | w_{1:j-1}, y_{1:j}) \quad (28)$$

Si se considera que los estados corresponden a un proceso de Markov de primer orden, y las observaciones son condicionalmente independiente dado los estados, se tiene que:

$$p(w_{1:t}) = p(w_0) \prod_{j=1}^t p(w_j | w_{j-1}) \quad (29)$$

y

$$p(y_{1:t} | w_{1:t}) = \prod_{j=1}^t p(y_j | w_j) \quad (30)$$

Sustituyendo las ecuaciones (28), (29), y (30) dentro de la ecuación (23), se obtiene el estimador recursivo para los pesos de importancia como sigue:

$$q_t = \frac{p(y_{1:t} | w_{1:t}) p(w_{1:t})}{\pi(w_{1:t-1} | y_{1:t-1}) \pi(w_t | w_{1:t-1}, y_{1:t})} = q_{t-1}^* \frac{p(y_{1:t} | w_{1:t}) p(w_{1:t})}{p(y_{1:t-1} | w_{1:t-1}) p(w_{1:t-1}) \pi(w_t | w_{1:t-1}, y_{1:t})} = q_{t-1} \frac{p(y_t | w_t) p(w_t | w_{t-1})}{\pi(w_t | w_{1:t-1}, y_{1:t})} \quad (31)$$

La ecuación dada en (31) permite actualizar secuencialmente el cociente de importancia. Como se puede muestrear de la distribución propuesta, evaluar la verosimilitud y las probabilidades de transición, se supone la siguiente estructura jerárquica:

1. Se considera una distribución a priori inicial:

$$w_0 \sim N(\mu_0, \sigma_0^2) \quad (32)$$

y

$$q_0^{(i)} = p(y_0 | w_0^{(i)}) \quad (33)$$

2. Función de Importancia viene dada por:

$$\pi(w_t | w_{1:t-1}, y_{1:t}) = P(w_t | w_{t-1}) \quad (34)$$

y

$$w_t | w_{t-1} \sim N(Tw_{t-1}, Q_t) = P(w_t | w_{t-1}) \quad (35)$$

3. La función de verosimilitud viene dada por $p(y_t | w_t)$:

$$\exp\{[y_t - \hat{g}_t(w_t, x_t)]^T R_t^{-1} * [y_t - \hat{g}_t(w_t, x_t)]\} \quad (36)$$

Entonces una vez inicializado los pesos y el cociente de importancia por las distribuciones a priori dadas en (32) y (33), en cada etapa del muestreo se predice un peso usando la ecuación dada en (1.a), luego se evalúa de nuevo el cociente de importancia usando las ecuaciones dadas en (31, 34, 35, 36) y se remuestrea si es necesario.

El problema que se presenta con el algoritmo de FP, 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 dada en Doucet et al. (2000), que la varianza de los pesos de importancia puede incrementarse en el tiempo. Para evitar la degeneración del algoritmo, Doucet et al. (2000) 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 $\{w_{1:t-1}^{(i)}, i = 1, 2, \dots, n\}$ y las observaciones $y_{1:t}$. La función de importancia sugerida por Doucet et al. (2000) es dada por:

$$\pi(w_t | w_{1:t-1}^{(i)}, y_{1:t-1}) = p(w_t | w_{t-1}^{(i)}, y_t) \quad (37)$$

Es decir, $\pi(w_t | w_{1:t-1}^{(i)}, y_{1:t-1})$ es la función de importancia que minimiza a la varianza de los pesos de importancia $\tilde{q}_t^{(i)}$ condicionada sobre los estados $\{w_{1:t-1}^{(i)}, i = 1, 2, \dots, n\}$ y los datos $y_{1:t}$. La función de importancia sugerida por Doucet et al. (2000) es dada por:

$$\pi(w_t | w_{1:t-1}^{(i)}, y_{1:t}) = p(w_t | w_{t-1}^{(i)}) \quad (38)$$

La ecuación (38) ha sido usada por Avitzour, (1995); Beadle y Djuric, (1997); Gordon et al., (1993); Isard y Blake, (1996); Kitagawa (1996); Berzuini et al., (1997); Doucet (1998); y Liu y Chen (1998), entre otros.

Debido a que la degeneración del algoritmo FP es inevitable, entonces se introduce la idea del remuestreo que consiste en eliminar las trayectorias que tienen pesos de importancia pequeños y se seleccionan las 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) y Liu (1996), y que se define como:

$$N_{TME} = \frac{N}{1 + \text{Var}(\tilde{q}_t^{(i)})} \quad (39)$$

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

$$\hat{N}_{TME} = \frac{1}{\sum_{i=1}^N [\tilde{q}_t^{(i)}]^2} \quad (40)$$

Entonces el estimador de N_{TME} se compara con un umbral prefijado N_U . Este procedimiento de filtro de partículas con remuestreo es usado por Rubin (1988). El algoritmo con remuestreo para un modelo de RN con dos capas ocultas, se resume:

#####

Algoritmo 2: FP o SIR

#####

P 1. Muestreo de importancia

- Se inicializan los pesos de la red en $(t = 0)$. Para $i = 1, \dots, n$ se muestrean los pesos $w_0^{(i)}$ de la primera y segunda capa a partir de las previas $p_1(w_0)$ y $p_2(w_0)$ respectivamente.
- Se evalúa el cociente de importancia:

$$q_0^{(i)} = p(y_0 | w_0^{(i)}) \quad (41)$$

- Se normaliza el cociente de importancia:

$$\tilde{q}_0^{(i)} = \frac{q_0^{(i)}}{\sum_{j=1}^n q_0^{(j)}} \quad (42)$$

- Para $t = 1, \dots, L$ y $i = 1, \dots, n$, se predicen los estados usando la ecuación dinámica:

$$\tilde{w}_{t+1}^{(i)} = Tw_t^{(i)} + u_t^{(i)}; \quad u_t^{(i)} \sim N(0, Q_t) \quad (43)$$

- Se actualiza el cociente de importancia:

$$q_{t+1}^{(i)} = q_t^{(i)} p(y_{t+1} | \tilde{w}_{t+1}^{(i)}) \quad (44)$$

- Se actualiza el cociente de importancia normalizado:

$$\tilde{q}_{t+1}^{(i)} = \sum_{j=1}^n \frac{q_{t+1}^{(j)}}{q_{t+1}^{(j)}} \quad (45)$$

P 2. Etapa de Remuestreo

Para $i = 1, \dots, n$, se evalúa el estimador usando la ecuación (40):

- Si $\hat{N}_{TME} \geq N_U$, entonces se hace:

$$w_{t+1}^{(i)} = \tilde{w}_{t+1}^{(i)} \quad (46)$$

- Si $\hat{N}_{TME} < N_U$, entonces se remuestra un nuevo índice j del conjunto de pesos y estados discretos estimados previamente y almacenados en: $\{\tilde{w}_{t+1}^{(i)}, \tilde{q}_{t+1}^{(i)}\}$.

- Luego se hace que: $w_{t+1}^{(i)} = \tilde{w}_{t+1}^{(j)}$ y $q_{t+1}^{(i)} = \frac{1}{n}$

- Salidas del algoritmo: función de densidad filtrada, esperanza y covarianza a posteriori:

$$\hat{p}(w_{1:t} | y_{1:t}) \approx \sum_{i=1}^n \hat{q}_t^{(i)} \delta(w_{1:t} - w_{1:t}^{(i)})$$

$$\hat{w}_t = E(w_{1:t} | y_{1:t}) \approx \sum_{i=1}^n \hat{q}_t^{(i)} w_{1:t}^{(i)}$$

$$\hat{p}_t \approx \sum_{i=1}^n \hat{q}_t^{(i)} (w_{1:t} - w_{1:t}^{(i)}) (w_{1:t} - w_{1:t}^{(i)})^T$$

#####

3 Determinación de la exactitud.

Para evaluar la calidad de predicción del modelo, se utilizan las siguientes medidas de bondad de ajuste:

1. Estimación de la Exactitud de Predicción (EP). Es una medida propuesta por Mitchell (1997) con el objetivo de estimar la exactitud de predicción de un modelo de RN, el siguiente estimador es usado para tal fin y se define como:

$$EP = \{1 - [\frac{0.5 \sum_{t=1}^N (y_t - \hat{y}_t)^2}{N}]\} * 100 \quad (50)$$

2. La Raíz del Error Cuadrático Medio (RECM), es otra medida utilizada para medir la exactitud del modelo predictor de RN y se define como:

$$RECM = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \quad (51)$$

El objetivo es seleccionar el modelo de RN con menor RECM.

3. El Criterio de Información Bayesiano (BIC). Una práctica difícil en los modelos de RN es seleccionar la arquitectura de la red (capas ocultas de la red). Para tal objetivo se propone usar el criterio de información de Bayes *BIC*

propuesto por Nychka et al. (1992), que se define como sigue :

$$\frac{1}{2} \left\{ 1 + \ln(2\pi) + \ln(SCE) + p \frac{\ln(N)}{N} \right\} \quad (52)$$

donde: y_t es el t -ésimo patrón observado. \hat{y}_t es el t -ésimo valor predicho,

$$SCE = \frac{1}{N} \sum_{i=1}^N (y_t - \hat{y}_t)^2 \quad (53)$$

es la suma de cuadrados del error, p es el número de parámetros (pesos) del modelo y N es la longitud de la serie de tiempo.

4 Transformación de los datos

Los datos utilizados para mostrar la metodología provienen de las series de lluvia total mensual del período comprendido entre 1971-2000 de 36 estaciones del servicio meteorológico de la Fuerza Aérea de Venezuela, y se le realizó la siguiente transformación:

1. Primero, las bases de datos que presentaron datos faltantes fueron rellenas utilizando la metodología propuesta en Infante et al. (2008). Es decir, dado que en ese trabajo se obtuvo un modelo de RN predictor basado en el algoritmo Backpropagation, entonces se utilizó ese modelo para predecir los datos perdidos y completar la base de datos.
2. Dado que en los modelos de RN se usa una función de

transferencia logística sigmoïdal para las neuronas de la capa oculta y en la capa de salida, la cual está acotada en un rango entre cero y uno, entonces se necesita normalizar o reescalar los datos de tal manera que estén dentro del rango de salida. El aprendizaje de parámetros de la red se hace inefectivo si los valores normalizado están cerca de las cotas (véase, Ooyen y Nichhuis, 1992), entonces se utilizará la siguiente ecuación de normalización:

$$y_{nor} = 0.1 + 0.8 \frac{y_t}{\max(y_t)} \quad (54)$$

donde: y_{norm} es el valor normalizado, y_t es la t -ésima salida observada, y $\max(y_t)$ es el valor máximo observado en la serie. La transformación acota el valor de la salida en el rango [0.1, 0.9].

DISCUSIÓN DE RESULTADO

Una vez que se tienen los datos transformados por estación, se divide la muestra en dos partes: el ochenta por ciento 80% de los datos se utiliza para el entrenamiento de los algoritmos y calibración de los modelos, y el resto veinte por ciento 20% se utiliza para validar la arquitectura de red del modelo seleccionado. El siguiente paso consiste en determinar las dimensiones de inmersión m y el tiempo de retardo τ de las estaciones consideradas,

utilizando los métodos de AMI y los FNN. Conociendo m y τ se determinan las variables de entrada de la red, y finalmente se procede a estimar los parámetros del modelo, usando dos algoritmos de aprendizaje FKE y el FP o SIR. En la Figura 1 se comparan los niveles totales de precipitación mensual de los estados observados, con los niveles totales de precipitación mensual simulados por un modelo de RN ajustado por los algoritmos FKE y FP. Los datos corresponden a la estación 80413 ubicada en Maracay estado Aragua y se utilizaron 278 datos de entrenamiento que representan a la serie de diciembre de 1971 hasta febrero de 1995 para ajustar los parámetros de la red.

Se seleccionó un modelo con 6 variables de entradas (debido a que se determino que $m = 6$ y $\tau = 2$), 4 neuronas de la capa oculta y una salida o respuesta, que representa el nivel total de lluvia mensual predicha. Los valores para inicializar los parámetros (medias y covarianzas) del FKE, son los siguientes: $m_{0|0} = 0.001$ y $P_{0|0} = 0.1$ y se requirió de 1000 iteraciones para obtener los resultados.

Los parámetros del FP se inicializan de la siguiente manera: $w_0 \sim N(0,225)$, $u_t \sim N(0,0.001)$, $v_t \sim N(0,0.1)$, y se eligió $T = I$ donde I es la matriz identidad y se generaron 1000 partículas.

En la Figura 1, se observa que los resultados predichos por los algoritmos propuestos tienen muy buen

desempeño comparado con las observaciones verdaderas de la muestra de entrenamiento.

En la Tabla 1 se muestran los resultados de los estadísticos estimados para medir la exactitud de predicción del modelo de RN. El FKE tiene una exactitud de predicción del 99.727%, mientras que el FP tiene una exactitud de 99.405%. Otra medida de exactitud utilizada es la raíz cuadrada del error cuadrático medio, ambos algoritmos tienen errores pequeños de estimación; el FKE tiene un valor de 0.073, mientras que el FP tiene un valor de 0.126. Para hacer la selección de la arquitectura de la red se utilizó el criterio de información de Bayes BIC, que básicamente consiste en penalizar el error cuadrático medio por los parámetros utilizados en el ajuste y se toma el BIC mínimo. En la serie de entrenamiento analizada el FKE tiene un BIC = -3.621, mientras que el FP tiene un BIC = -2.841, ambos algoritmos seleccionan un modelo con 4 neuronas en la capa oculta. La Figura 2 se obtiene siguiendo el mismo procedimiento que se realizó con la serie de entrenamiento pero este caso se utilizan 70 datos de validación que van desde Marzo de 1995 hasta Diciembre de 2000. Se observa que con el FKE se obtienen los siguientes valores: una exactitud de predicción del 98.638%, un valor para la raíz del error cuadrático medio de 0.1650 y el BIC = -1.668; mientras que cuando se usa el algoritmo FP, se obtienen los valores siguientes: una exactitud de 98.697%, un

valor para la raíz del error cuadrático medio de 0.110, y el BIC =-1.713, ver la Tabla 1, estos resultados son similares a los obtenidos utilizando la serie de entrenamiento. En la Figura 3 se muestra un gráfico similar al obtenido en la Figura 1, la diferencia es la estación que se analiza; en este caso, se estudia la estación 80415 ubicada en Maiquetía

estado Vargas. Se seleccionaron 259 datos para la muestra de entrenamiento que comprende la serie que va desde Diciembre de 1972 hasta Junio 1992; mientras que en la muestra de validación se utilizaron 65 datos que corresponden a la serie que va de Julio 1992 hasta Diciembre

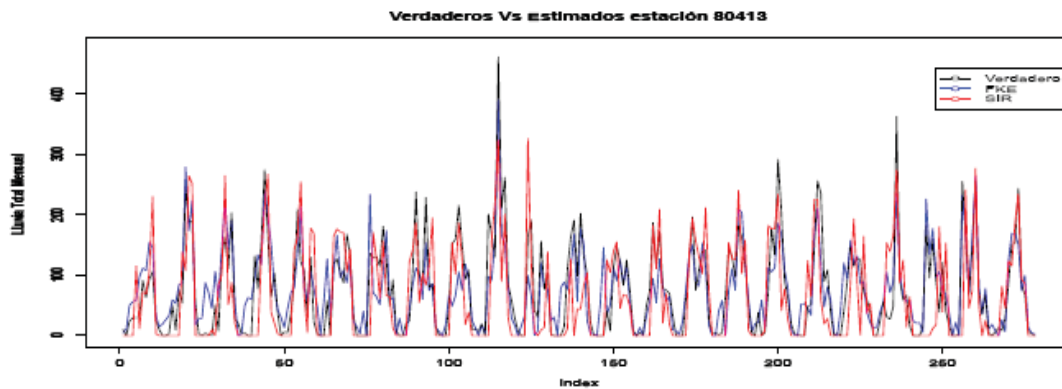


Figura 1: Serie de la estación Maracay, para la muestra de entrenamiento

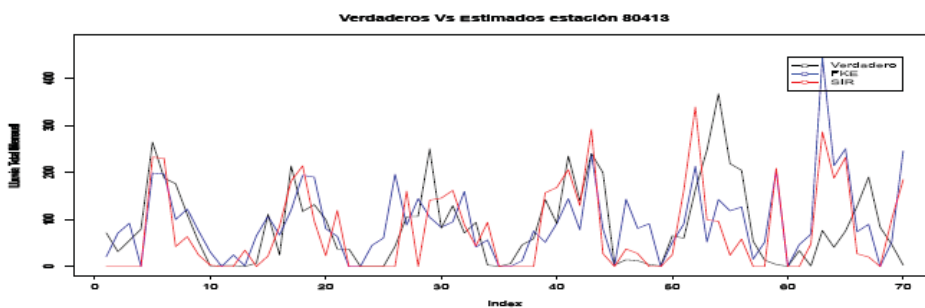


Figura 2: Serie de la estación Maracay, para muestra de validación

Tabla 1: Resumen de test de validación, estación Maracay

TEST	Estadísticos	FKE	SIR
Entrenamiento	EP	99.727%	99.405%
Entrenamiento	RCEM	0.073	0.109
Entrenamiento	BIC	-3.621	-2.841
Validación	EP	98.638%	98.697%
Validación	RCEM	0.165	0.161
Validación	BIC	-1.668	-1.713

De 1999. La arquitectura de la red seleccionada está compuesta por 6 variables de entradas, 5 neuronas en la capa oculta y una salida. En el gráfico se observa que los totales mensuales de precipitaciones simulados por el modelo de RN ajustado por los algoritmos FKE y FP son comparables con los totales mensuales de precipitaciones observados; adicionalmente se incluyen medidas exactitud de 99.414% y 99.198%, respectivamente cuando se utiliza la muestra de entrenamiento. Para la muestra de validación la exactitud es de 98.844% y 98.945%, respectivamente. También se cálculo la RECM para los dos conjuntos de datos

considerados, observándose errores de estimación pequeños (ver Tabla 2). También se estimó el BIC para la muestra de entrenamiento y validación usando los dos algoritmos propuestos, obteniéndose el BIC mínimo para una red con 5 neuronas ocultas, (ver Tabla 2). Todos los análisis fueron realizados bajo el ambiente de programación del software estadístico **R versión 2. 9. 2**. Finalmente se destaca que se analizaron 36 estaciones, todas tienen un comportamiento muy similar a las dos mostradas en el trabajo.

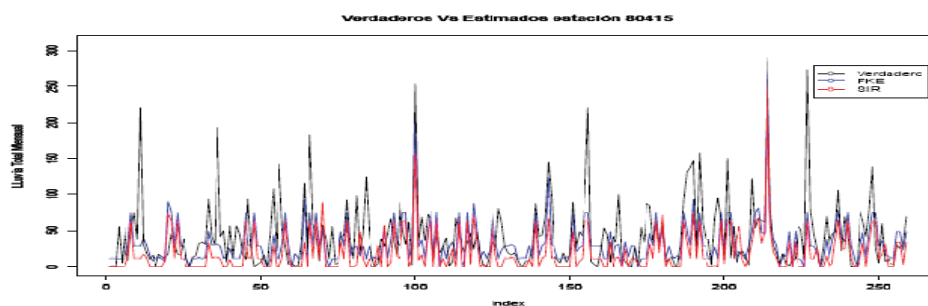


Figura 3: Serie de la estación Maiquetía, para la muestra de entrenamiento

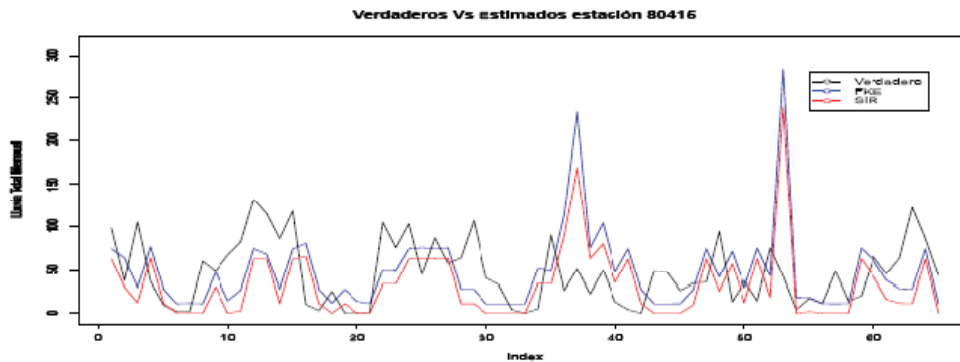


Figura 4: Serie de la estación Maiquetía, para la muestra de validación

Tabla 2: Resumen de test de validación, estación Maiquetía

TEST	Estadísticos	FKE	SIR
Entrenamiento	EP	99.414	99.198
Entrenamiento	RCEM	0.108	0.126
Entrenamiento	BIC	-2.932	-2.617
Validación	EP	98.844	98.945
Validación	RCEM	0.152	0.145
Validación	BIC	-2.059	-2.150

CONCLUSIONES

Las principales contribuciones de este trabajo son:

-Se diseñó un modelo de red neuronal en términos de los modelos espacio estados no lineales para hacer predicciones de series de tiempo de totales mensuales de precipitaciones. Se utilizan las técnicas AMI y FNN de los sistemas dinámicos para definir las entradas del modelo, se demuestra que el uso de patrones de entrada apropiados mejora la performance del modelo. Estos métodos permiten la

separación temporal óptima de los datos, así como ayudan a determinar el número de datos hacia atrás necesarios para pronosticar con exactitud un dato futuro. Para optimizar la selección de la arquitectura de las capas ocultas de la red se utiliza el Criterio de Información de Bayes.

-Se proponen dos algoritmos de aprendizaje alternativos para ajustar los parámetros del modelo de RN distinto al algoritmo Backpropagation clásico. Estos algoritmos pueden ser

considerados como técnicas de optimización global en las cuales se pueden estimar conjuntamente los pesos de la red y la incertidumbre del modelo en tiempo real, mediante la caracterización de distribuciones de probabilidad. El FKE permite obtener mejores resultados que las técnicas de estimación convencionales, debido a que hace uso de los estadísticos de segundo orden (covarianzas) que son esenciales para propagar el error sobre las predicciones; además, muestra excelente convergencia en los parámetros estimados debido a las exigencias de memoria y precisión computacional. El algoritmo FP o SIR permite que los parámetros se actualicen en forma

secuencial de una distribución de probabilidad a posteriori, y tienen la ventaja que no requieren de ninguna restricción sobre la distribución y forma funcional de los datos.

-Se proponen dos métodos para medir la bondad de ajuste del modelo predictor, obteniéndose como resultado que el modelo ajustado por los algoritmos FKE y FP o SIR tienen una exactitud de predicción mayor que 99% y un valor para raíz de error cuadrático medio menor que 0.17, tanto para las muestras de entrenamiento y validación, respectivamente. Los resultados de este estudio muestran que el modelo de RN es útil para predecir precipitaciones mensuales con exactitud y confiabilidad.

Agradecimientos

Agradecemos al servicio de meteorología de la Fuerza Aérea de Venezuela, por proveer los datos que sirvieron de insumo para este trabajo, y a los revisores anónimos por los comentarios y sugerencias que ayudaron a mejorar el manuscrito. Esta investigación fue parcialmente financiada por el Consejo de Desarrollo Científico y Humanístico, proyecto CDCH2006-003 y la partida 407 año 2009 de la Facultad Experimental de Ciencias y Tecnología de la Universidad de Carabobo.

REFERENCIAS

- Abarbanel, H. (1996). *Analysis of observed data*. New York: Springer-Verlag.
- Avitzour, D. (1995). A stochastic simulation Bayesian approach to multitarget tracking, *Proc.Inst. Elect. Eng., Radar, Sonar, Navig*, vol. 142, 2, 41 - 44.
- Beadle, E. R. y Djuriac, P. M. (1997). Fast weighted bootstrap filter for non-linear state estimation, *IEEE Trans. Aerosp. Elect. Syst*, vol. 33, 338- 343.
- Bishop, C. M. (1995). *Neural Networks for pattern Recognition*. New York: Oxford University Press.
- Bustami, R., Bessaih, N., Bong, C. y Suhaili, S. (2007). Artificial neural network for precipitation and water level predictions of bedup river. *IAENG International journal of computer science*, 34, 2, 228-233.
- Calderón, G. y Pérez, P. (2006). Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales. *Revista Chilena de Ingeniería*, Vol. 14, 3, 284-290.
- Casdagli, M. (1989). Nonlinear prediction of chaotic time series. *Physica D*, 35, 335- 356.
- Cheng, B. y Titterington, D. (1994). Neural Networks a review from a statistical perspective with discussion. *Statistical Science*, 9, 2- 54.
- De Freitas, J., Niranjan, M., Gee, A. H. y Doucet, A. (1998). Sequential Monte Carlo methods to optimization of Neural Network Models. *Technical report Cued/F-Infeng/Tr 328*.
- De Freitas, J., Niranjan, M., Gee, A. H. y Doucet, A. (2000). Sequential Monte Carlo methods to Train Neural Network Models. *Neural Computation*, 12, 955-993.
- Doucet, A., Godsill, S. y Andrieu, C. (2000). On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Statistic and Computing*, 10, 197-208.
- Doucet, A. (1998). On sequential simulation based methods for Bayesian filtering. *Tech report No. Cued/F-Infeng/tr 310*. Cambridge University.
- Fraser, A. y Swinney, H. (1986). Independent coordinates for strange attractors from mutual information, *Phys Review A* , 33, 1134-1140.
- Gordon, N., Salmond, D. y Smith, A.F.M. (1993). Novel Approach to Nonlinear non Gaussian Bayesian State Estimation, *IEEE Proceedings-F*, 140, 107-113.
- Harvey, A.C. (1970). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Infante, S., Ortega, J. y Cedeño, F. (2008). Estimación de datos faltantes en estaciones meteorológicas de Venezuela vía un modelo de redes neuronales. *Revista de Climatología de España*, 8, 51-70.
- Infante, S., Rojas, J., Hernández, A. y Cartaya, V. (2009). Modelos de espacio estado basados en la distribución normal inversa gaussiana: Una aplicación al análisis de series de tiempo de la economía venezolana. *Revista de Estadística, del Instituto Interamericano de Estadística*, 61, 176 y 177.
- Isard, M. y Blake, A. (1996), Contour tracking by stochastic propagation of conditional density. In *computer vision-ECCV96*, eds. B.

- Buxton and R. Cipolla, New York: Springer Verlag, 343- 356.
- Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Nonlinear Non Gaussian State Models, *Journal of Computation and Graphical Statistics*, 5, 1-25.
- Kong, A., Liu, J. y Wong, W. (1994). Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American Statistical Association*, 89, 278-288.
- Lary, D. J. y Mussa, H. Y. (2004). Using an extended Kalman filter learning algorithm for feed forward neural networks to describe tracer correlations. *Atmospheric Chemistry and Physics Discussions*, 4, 3653- 3667.
- Liu, J. (1996). Metropolized Independent Sampling with Comparison to Rejection Sampling and Importance Sampling, *Statistics and Computing*, No. 6, 113-119.
- Liu, J. y Chen, R. (1998). Sequential Monte Carlo Methods for Dynamical Systems, *Journal of the American Statistical Association*, No. 83, 1032-1044.
- Neal, R. (1996). *Bayesian learning for neural networks*. New York: Springer.
- Nychka, H., Ellner, S., Gallant, A. y McCaffrey, D. (1992). Finding Chaos in noise systems. *Journal Royal of the Statistical Society Series B*, 54, 399-426.
- Menchero, A., Montes Diez, R., Ríos Insua, D. y Muller, P. (2005). Bayesian Analysis of Nonlinear Autoregression Models Based on Neural Networks A. *Neural Computation*, No. 17, 453- 485.
- Mitchell, T. (1997). *Machine Learning*, McGraw hill press international eds.
- Morais, H., Fraga de Campos, H. y Einstein, E. (2008). Data assimilation: particle filter and artificial neural networks. *Journal of Physics: conference Series* 135012073.
- Ooyen, A. y Nichhuis, B. (1992). Improving the convergence of Back Propagation problem. *Neural Networks*, 5, 465- 471.
- Rajunkar, M., Kothyari, U. y Chaube, U. (2002). Artificial neural networks for daily rainfall runoff modelling. *Hydrological Sciences*, Vol. 47, No. 6, 865-877.
- Ripley, B. (1994). Neural networks and related methods for classification (with discussion). *Journal Royal Statistical Society Series B*, No. 56, 409-456.
- Ripley, B. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.
- Rubin, B.D. (1988). Using the SIR Algorithm to Simulate Posterior Distributions, *Bernardo J.M., Degroot M.H., Lindley D.V., and Smith A.F.M. (Eds.), Bayesians Statistics*, 3, Oxford University Press, 395- 402.
- Takens, F. (1981). Detecting strange attractors in turbulence, lecture notes in math, 898, Springer, New York.
- Titterington, D .M. (2004). Bayesian methods for neural networks and related models. *Statistical Science*, Vol. 19, No. 1, 128-139.

Autores

Saba Infante. Profesor Asociado a Dedicación Exclusiva, Departamento de Matemáticas, Facultad Experimental de Ciencias y Tecnología y Centro de Análisis, Modelado y Tratamiento de Datos (CAMYTD), Universidad de Carabobo, Valencia, Venezuela. 2005, Doctor en Estadística, Universidad Simón Bolívar, Venezuela.

Email: sinfante@uc.edu.ve

Fernando Cedeño. Profesor Instructor a Dedicación Exclusiva, Departamento de Matemáticas, Facultad Experimental de Ciencias y Tecnología y Centro de Análisis, Modelado y Tratamiento de Datos (CAMYTD), Universidad de Carabobo, Valencia, Venezuela. 2008, Licenciado en Matemáticas, Universidad Carabobo, Venezuela.

Email: fjcedeno@uc.edu.ve

José Ortega. Profesor Titular a Dedicación Exclusiva, Departamento de Matemáticas Facultad de Ciencia y Tecnología y Centro de Análisis, Modelado y Tratamiento de Datos (CAMYTD), Universidad de Carabobo, Valencia, Venezuela. 2006, Doctor en Ciencias, Universidad Central de Venezuela.

Email: jortega@uc.edu.ve

Recibido: 22/10/2010

Aceptado: 27/12/2010