

# Polynomial Chaos based on the parallelized ensemble Kalman filter to estimate precipitation states

Luis Sánchez <sup>1\*</sup>, Saba Infante <sup>2</sup>, José Marcano <sup>2</sup>, and Victor Griffin <sup>2</sup>

<sup>1</sup> *Departamento de Matemáticas, Facultad de Ciencias de la Educación.*

<sup>2</sup> *Departamento de Matemáticas, Facultad de Ciencia y Tecnología. Universidad de Carabobo, Venezuela.*

Received: 26 November 2014; Accepted: 13 February 2015

Editor: Antanas Zilinskas

**Abstract** This article develops a methodology combining methods of numerical analysis and stochastic differential equations with computational algorithms to treat problems which have complex nonlinear dynamics in high dimensions. A method to estimate parameters and states of a dynamic system is proposed inspired by the parallelized ensemble Kalman filter (PEncKF) and the polynomial chaos theory of Wiener-Askey. The main advantage of the proposal is in providing a precise efficient algorithm with low computational cost. For the analysed data, the methods provide good predictions, spatially and temporally, for the unknown precipitation states for the first 24 hours. Two goodness of fit measures provide confidence in the quality of the model predictions. The performance of the parallel algorithm, measured by the acceleration and efficiency factors, shows an increase of 7% in speed with respect to the sequential version and is most efficient for  $P = 2$  threads.

**Keywords** Stochastic differential equations, Numerical methods, Polynomial Chaos, Parallelized ensemble Kalman filter.

**AMS 2010 subject classifications** 62L10, 62L20

**DOI:** 10.19139/soic.v3i1.113

## 1. Introduction

Stochastic differential equations are considered an effective tool to model physical processes in distinct fields of science. The equations are generally expressed in terms of a process in continuous time with random coefficients, random initial conditions or random base functions and the solutions are expressed using random fields. For example if we consider the stochastic differential equation given in [33] of the form

$$dx_t = f(x_t, t, \theta)dt + L(x_t, t, \theta)dW_t \quad (1)$$

where  $x_t$  is a signal in continuous time,  $t$  is a time variable,  $\theta$  is a parameter vector,  $f(\cdot)$  is a nonlinear drift or transfer function,  $L(\cdot)$  is a dispersion matrix, and  $W_t$  is an independent vector whose components represent a Wiener process. Given that the equation (1) is defined in continuous time, the behaviour of the system is observed imperfectly by means of an experimental process  $y_t$  which is related to the process  $x_t$  through the equation

$$y_t = \mathcal{H}(x_t, \theta) + v_t \quad , \quad v_t \sim N(0, R_t(\theta)) \quad (2)$$

where  $y_t$  represents a vector of observations sampled in discrete times  $t_1, \dots, t_k$ ,  $\mathcal{H}$  is a matrix defined in terms of the parameters and states, and  $v_t$  is a random sampling error with Gaussian distribution, mean 0 and covariance

\*Correspondence to: Luis Sánchez (Email: lasa.58@hotmail.com). Departamento de Matemáticas, Universidad de Carabobo, Venezuela.

matrix  $R_t(\theta)$ . Instead of finding the point estimators of the parameters  $\theta = (\theta_1, \dots, \theta_p)$  and the states  $x_t$ , we are interested in studying their joint probability distribution  $P(x_t, t, \theta)$ . In particular, we want to estimate the posterior joint distribution of the states  $P(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ , or the marginal state distribution  $P(x_t|\mathbf{y}_{1:t})$ , and characterize them by their posterior means, modes and covariances where  $x_t$  represents a solution of the equation given in (1),  $\mathbf{x}_{0:t} = (x_0, \dots, x_t)$ , and  $\mathbf{y}_{1:t} = (y_1, \dots, y_t)$ .

The difficulty found in the estimation of the parameters, states, joint and posterior marginal is that the transition density of the equation(1), in general cannot be evaluated in a closed form and techniques are required to estimate the solution. There exist a variety of methods in the literature including the estimation by maximum likelihood ([35], and [20]); the Monte Carlo Markov Chain techniques (MCMC) ([26], [34], [42], [43]); and the sequential Monte Carlo algorithms (SMC). Examples of the latter are the Kalman filter, the extended Kalman filter (EKF), the particle filter (PF) and the unscented particle filter (UPF) (see [2], [22], [16], [23], [24], [4], [44], [18], [25], [29], [9], [30], [36], [41], [14], [8], [15], [17], [28], [39], [40], [6], [21], [3] and [38]).

The main contribution of this study consists in estimating and predicting the rainfall states  $p(x_t|\mathbf{y}_{1:t})$  in ten meteorological stations in Venezuela using the stochastic differential equation proposed by [1] and the polynomial chaos and the parallelized ensemble (PEnKF) methodologies. To achieve this objective it is necessary to solve a stochastic differential equation in terms of the Fokker-Plank equation using a numerical method of finite differences ([7]). The solution is expressed using an expansion of the polynomial chaos. A data assimilation technique is then used to generate large samples of solution ensembles which are used to filter the sampling error in the prediction model given by the equation (2). Two goodness of fit measures are calculated to validate the quality of the model predictions. Finally the acceleration and efficiency factors are calculated to measure the performance of the parallelized algorithm.

In an article related to this proposal, [47], Wuan used numerical methods based on Weiner's chaos expansion to solve stochastic partial differential equations in which the stochastic solution is represented by a spectral expansion with respect to a set of random base functions. It was shown that the methods were more efficient and exact than Monte Carlo methods for short and medium times. In another work, [31], Li and Xiu implemented the EnKF using the expansion of the generalized polynomial chaos to approximate the solution of some known stochastic differential equations obtaining as a result an algorithm with good estimates and low computational cost. Madankan,[32], proposed two recursive algorithms to estimate the posterior moments of higher order for the parameters and states of a stochastic differential equation using the idea of a generalized polynomial chaos. He shows that the advantage of using these methods is that it is possible to obtain confidence intervals as well as point estimates for the parameters and states.

The organization of this article is as follows. Section 2 briefly describes the methodology of the polynomial chaos. Section 3 develops the technical aspects of the stochastic differential Fokker-Plank equation and that of Allen. Section 4 describes the EnKF and PEnKF algorithms. Section 5 defines the goodness of fit measures. Section 6 shows the results and Section 7 summarizes the conclusions and final discussions.

## 2. Polynomial Chaos

The fundamental idea of the polynomial chaos is that the random processes of interest can be approximated by sums of orthogonal polynomials chaos of independent random variables. In this context, the uncertainty about the parameters and states can be considered as a second order random process  $X(\omega)$  with finite variance that can be represented in terms of the polynomial chaos. This random process in the original discrete version of the Weiner

polynomial chaos expansion is defined by

$$\begin{aligned}
 X(\omega) = & a_0 H_0 + \sum_{i_1=1}^{\infty} a_{i_1} H_1(\xi_{i_1}(\omega)) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1 i_2} H_2(\xi_{i_1}(\omega), \xi_{i_2}(\omega)) \\
 & + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3} H_3(\xi_{i_1}(\omega), \xi_{i_2}(\omega), \xi_{i_3}(\omega)) + \dots
 \end{aligned} \tag{3}$$

where  $\omega$  is a random event, and  $H_n(\xi_{i_1}(\omega), \dots, \xi_{i_n}(\omega))$  is a Hermite polynomial of order  $n$ , defined in terms of a multidimensional random variable  $\xi(\omega) = (\xi_{i_1}(\omega), \dots, \xi_{i_n}(\omega))$ . The Hermite polynomials are given in terms of normal random variables with mean zero and variance one. The general expression of the Hermite polynomials is given by

$$H_n(\xi) = (-1)^n \exp\left(\frac{1}{2}\xi^T \xi\right) \frac{\partial^n}{\partial \xi_{i_1} \dots \partial \xi_{i_n}} \exp\left(-\frac{1}{2}\xi^T \xi\right) \tag{4}$$

For notation convenience the equation given in (3) can be rewritten in the following way. Let  $(\Omega, \mathcal{F}, \mathcal{P})$  be a complete probability space where  $\Omega$  is a sample space,  $\mathcal{F}$  is the  $\sigma$ -algebra of subsets of  $\Omega$ , and  $\mathcal{P}$  is a probability measure. Then define a second order random process  $X(\omega) \in L_2(\Omega, \mathcal{F}, \mathcal{P})$ , the space of square integrable functions, in terms of a polynomial chaos as

$$X(\omega) = \sum_{i=0}^{\infty} \tilde{a}_i \phi_i(\xi(\omega)) \tag{5}$$

where  $\tilde{a}_i$  are the deterministic coefficients of the polynomial chaos, and  $\phi_i(\xi(\omega))$  is a multidimensional Hermite polynomial. The family  $\{\phi_i\}$  is an orthogonal basis in  $L_2(\Omega, \mathcal{F}, \mathcal{P})$ . There exists a one to one correspondence between the Hermite polynomials  $\{H_n(\xi)\}$  and the base polynomials  $\{\phi_i(\xi)\}$ . In real situations, the series can be truncated to a finite number of  $n_p$  terms so that the general truncated expression of the Hermite chaos is given by

$$X(\omega) = \sum_{i=0}^{n_p} \tilde{a}_i \phi_i(\xi(\omega)) \tag{6}$$

with the number of terms,  $n_p$ , determined by

$$n_p + 1 = \frac{(n + n_p)!}{n! n_p!} \tag{7}$$

where  $n_p$  is the order of the  $y$  expansion and  $n$  is the dimension of the random vector  $\xi(\omega)$  so that  $n_p$  grows rapidly with the dimension and decomposition order and a good approximation is obtained.

The orthogonal relation is expressed in terms of an internal product defined in a Hilbert space as

$$\langle \phi_i, \phi_j \rangle = \int \phi_i(\xi) \phi_j(\xi) \rho(\xi) d\xi = \delta_{ij} \langle \phi_i^2 \rangle \tag{8}$$

where  $\rho(\xi) = \frac{1}{\sqrt{(2\pi)^{n_p}}} \exp\left(-\frac{1}{2}\xi^T \xi\right)$  is the kernel of a standard Gaussian distribution, so that

$$\langle \phi_i, \phi_j \rangle = \delta_{ij} \langle \phi_i^2 \rangle \tag{9}$$

where  $\delta$  is the Kronecker delta.

### 2.1. Polynomial Chaos of Wiener-Askey

The Hermite chaos expansion is performed by resolving stochastic differential equations of Gaussian random variables, but in general when the random variables do not have Gaussian distributions, the exponential convergence

rate is not achieved. Other general polynomials chaos exist and are recommended. Some examples are Laguerre chaos which is associated with Gamma random variables, the Jacobi chaos which is associated with the Beta distribution, the Charlier chaos which is associated with the Poisson distribution, the Meixner chaos which is associated with the negative Binomial distribution, the Krawtchouk chaos which is associated with the Binomial distribution and the Hahn chaos which is associated with the Hypergeometric distribution (see [48] for more details). The polynomial chaos expansion to be used for these random processes is that of Wiener-Askey defined as

$$\begin{aligned} X(\omega) = & a_0 I_0 + \sum_{i_1=1}^{\infty} a_{i_1} I_1(\zeta_{i_1}(\omega)) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1 i_2} I_2(\zeta_{i_1}(\omega), \zeta_{i_2}(\omega)) \\ & + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3} I_3(\zeta_{i_1}(\omega), \zeta_{i_2}(\omega), \zeta_{i_3}(\omega)) + \dots \end{aligned} \quad (10)$$

$I_n(\zeta(\omega))$  denotes the Wiener-Askey polynomial chaos of order  $n$ , and  $\zeta(\omega) = (\zeta_{i_1}(\omega), \dots, \zeta_{i_n}(\omega))$  is a random vector. Again for notation convenience, the equation given in (10) can be rewritten as

$$X(\omega) = \sum_{j=0}^{n_p} \tilde{a}_j \phi_j(\zeta(\omega)) \quad (11)$$

where a one to one correspondence exists between the functions  $I_n(\zeta(\omega))$ , and the polynomials  $\phi_j(\zeta)$ . The Wiener-Askey polynomials chaos form a complete basis in the Hilbert space which guarantees that each chaos expansion converges to some functional in the sense  $L_2(\Omega, \mathcal{F}, \mathcal{P})$ . The relation of orthogonality of the Wiener-Askey polynomial chaos is given in terms of the internal product

$$\langle \phi_i, \phi_j \rangle = \int \phi_i(\zeta(\omega)) \phi_j(\zeta(\omega)) \rho(\zeta(\omega)) d\zeta(\omega) \quad (12)$$

where  $\rho(\zeta(\omega))$  is a probability distribution of the random variables  $\zeta(\omega)$ . A general procedure is now described for applying the Wiener-Askey polynomial chaos to solve a stochastic differential equation. Consider the stochastic differential equation

$$\ell(x_t, t, \zeta(\omega); u) = f(x_t, t; \zeta(\omega)) \quad (13)$$

where  $u = u(x_t, t; \zeta(\omega))$  is the solution of the equation, and  $f(x_t, t; \zeta(\omega))$ , is a source term. The operator  $\ell$  generally involves differentiations in space and time and is possibly nonlinear. Appropriate initial and frontier conditions should be assumed although the presence of the random vector  $\zeta(\omega)$  permits the introduction of uncertainty in the system for different sources of variation including those of initial and frontier conditions. The solution, considered to be a random process, can be expressed by the Wiener-Askey polynomial chaos as

$$u(x_t, t; \zeta(\omega)) = \sum_{i=0}^{n_p} u_i(x_t, t) \phi_i(\zeta(\omega)) \quad (14)$$

where  $u_i(x_t, t)$  is a coefficient to be determined and  $\phi_i(\zeta)$  is a Wiener-Askey polynomial. In particular a Laguerre polynomial is used in this work given that the data used to illustrate the application come from precipitation observations measured in meteorological stations in Venezuela. Finally, substituting (14) in (13), we obtain

$$\ell(x_t, t, \zeta(\omega); \sum_{i=0}^{n_p} u_i(x_t, t) \phi_i(\zeta(\omega))) = f(x_t, t; \zeta(\omega)). \quad (15)$$

The representation given in (14) can be considered as a spectral expansion between the coefficient  $u_i(x_t, t)$  and the vector  $\zeta(\omega)$  evaluated in the orthogonal polynomial  $\phi_i$ . The total number of terms in the expansion is determined by the equation defined in (7).

### 3. The Stochastic Differential Equation of Fokker Planck

When a dynamic system is governed by a stochastic differential equation, the evolution in time of the probability density of the state  $x_t$  satisfies the Fokker-Planck equation. This equation when discretized on a network permits the modeling of physical processes which have nonlinear structures, non-Gaussian distributions and multi-mode distributions ([27]). In order to numerically model the evolution of the states in time, the density can be truncated which permits the use of finite degrees of freedom. One way to do this is to project the density in terms of a mixture of base functions that evolve with time such as convolution or wave methods or an integro difference equation model ([19], [45]). To illustrate the procedure, suppose that it is required to estimate the states of the system given by the simplest stochastic differential equation of an Ito process

$$dx_t = f(x_t, t)dt + L(x_t, t)dW_t \quad (16)$$

where  $x_t$  represents the state vector of the system,  $f(\cdot)$  is a possibly nonlinear vectorial function,  $L(\cdot)$  is a matrix of real values, and  $dW_t$  is a vector of a process describing a Brownian movement with  $E(dW_t dW_t^T) = Q_t dt$ . The function  $f(\cdot)$  represents the deterministic part of the system while  $L(\cdot)$  represents the random part of the dynamics. Given that the model defined in (16) is a continuous process, it can be assumed that the data is obtained through an observation process as described in (2).

On the other hand, by the Theorem of Fokker-Planck, if  $p(x_t, t)$  is a solution to the equation given in (16), then it solves the equation given by

$$\frac{\partial p(x_t, t)}{\partial t} = - \sum_{i=0}^{n_p} \frac{\partial}{\partial x_i} [f_i(x_t, t)p(x_t, t)] + \frac{1}{2} \sum_{i=0}^{n_p} \sum_{j=0}^{n_p} \frac{\partial^2}{\partial x_i \partial x_j} \{ [L(x_t, t)Q_t L^T(x_t, t)]_{ij} p(x_t, t) \}$$

The uncertainty associated with the states  $\mathbf{x}_{0:t}$  is characterized by the marginal probability density function  $p(x_t, t)$ . In this study estimation of the states is performed with filtering techniques to determine the posterior filtered distribution  $p(x_t | \mathbf{y}_{1:t})$ . To implement a recursive filter, the Fokker-Planck equation is replaced by the Chapman-Kolmogorov equation

$$p(x_t | \mathbf{y}_{1:t-1}) = \int p(x_t | \mathbf{x}_{1:t-1}) p(x_{t-1} | \mathbf{y}_{1:t-1}) dx_{t-1} \quad (17)$$

and the update equation is obtained by the Theorem of Bayes

$$p(x_t | \mathbf{y}_{1:t}) = \frac{p(y_t | x_t) p(x_t | \mathbf{y}_{1:t-1})}{\int p(x_t | \mathbf{x}_{1:t-1}) p(x_{t-1} | \mathbf{y}_{1:t-1}) dx_t} \quad (18)$$

In the following section a precipitation model is defined and described in terms of the Fokker-Planck stochastic differential equation. The Wiener-Askey polynomial chaos is then used to expand the solution and finally a finite difference method is used to resolve the resulting system.

#### 3.1. Stochastic Differential Equation for Precipitation

Consider the simple stochastic model for rainfall described in [1], where  $x_t$  denotes the total rainfall at time  $t$  in a locality of  $\mathbf{s} = (x, y) \in \mathbf{R}^2$ . Suppose that the change in total rainfall during a very short time interval  $\Delta t$  and  $(\Delta x)_1 = \gamma$ , and  $(\Delta x)_2 = 0$  have probabilities  $p_1 = \lambda \Delta t$ , and  $p_2 = 1 - \lambda \Delta t$ , respectively. The expected value and the mean squared change in the total precipitation is given by

$$\begin{aligned} E(\Delta x) &= \gamma \lambda \Delta t \\ E((\Delta x)^2) &= \gamma^2 \lambda \Delta t. \end{aligned} \quad (19)$$

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

$$\begin{aligned} dx_t &= \mu dt + \sigma dW_t \\ x_0 &= 0 \end{aligned} \quad (20)$$

where  $dW_t$  is a vector of a Brownian process with mean 0 and covariance matrix  $Q_t$ . Given that the model defined in (20) is a continuous process, the data is assumed to be obtained according to the discrete equation given in (2). Then the samples taken at time  $t_k$  are generated and denoted by  $\mathbf{y}_{1:k}$ . Given that the probability density function represents all the statistical information of the states  $\mathbf{x}_{0:t_k}$  conditioned in the observations  $\mathbf{y}_{1:k}$ , the prior density for the system exists and is continuously differentiable with respect to time  $t_k$  and the state vector  $\mathbf{x}_{0:t_k}$ . The evolution in time of the marginal density of the state is given by the posterior marginal distribution  $p(x_{t_k}|\mathbf{y}_{1:k-1})$  which satisfies the Fokker-Planck equation

$$\frac{\partial}{\partial t} p(x_{t_k}|\mathbf{y}_{1:k-1}) = \mathcal{L}(p(x_{t_k}|\mathbf{y}_{1:k-1})) \quad (21)$$

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

$$\mathcal{L}(p) = - \sum_{i=0}^{n_p} \frac{\partial(\mu p)}{\partial x_i} + \frac{1}{2} \sum_{i=0}^{n_p} \sum_{j=0}^{n_p} \frac{\partial^2[(\sigma Q_t \sigma^T)_{ij} p]}{\partial x_i \partial x_j} \quad (22)$$

with  $p = p(x_{t_k}|\mathbf{y}_{1:k-1})$ . The initial condition is given by  $p(x_{t_{k-1}}|\mathbf{y}_{1:k-1})$ . Applying the Fokker-Planck equation to the equation (20), we obtain

$$\frac{\partial p}{\partial t} = -\mu \frac{\partial p}{\partial x} + \frac{\sigma^2 Q_t}{2} \frac{\partial^2 p}{\partial x^2} \quad (23)$$

The next step is to find a solution to the equation given in (23). To do this, a truncated expansion of the polynomial chaos is used as in the equation (14).

$$p(x_t, t, \xi) = \sum_{i=0}^{n_p} p_i(x_t, t) \phi_i(\xi) \quad (24)$$

where  $p(x_t, t, \xi)$  is a solution of the equation given in (23),  $p_i(x_t, t)$  represents the deterministic coefficients,  $\xi$  is a vector of orthogonal random variables, and  $\phi_i(\xi)$  is a chaotic polynomial of Laguerre. Then, substituting the expansion of the equation (24) in (23) we obtain

$$\sum_{i=0}^{n_p} \frac{\partial p_i \phi_i(\xi)}{\partial t} = \frac{\sigma^2 Q_t}{2} \sum_{i=0}^{n_p} \frac{\partial^2 p_i \phi_i(\xi)}{\partial x^2} - \mu \sum_{i=0}^{n_p} \frac{\partial p_i \phi_i(\xi)}{\partial x} \quad (25)$$

where

$$\sum_{i=0}^{n_p} \frac{\partial p_i \langle \phi_i \phi_l \rangle}{\partial t} = \frac{\sigma^2 Q_t}{2} \sum_{i=0}^{n_p} \frac{\partial^2 p_i \langle \phi_i \phi_l \rangle}{\partial x^2} - \mu \sum_{i=0}^{n_p} \frac{\partial p_i \langle \phi_i \phi_l \rangle}{\partial x} \quad (26)$$

By the orthogonality relations

$$\frac{\partial p_i}{\partial t} - \frac{\sigma^2 Q_t}{2} \frac{\partial^2 p_i}{\partial x^2} + \mu \frac{\partial p_i}{\partial x} = 0 \quad , \quad i = 0, \dots, n_p \quad (27)$$

There are many ways to solve the system. This article proposes a finite difference method to numerically solve the partial differential equations developed by [7]. The result is

$$\begin{aligned} & - \frac{\sigma^2 Q_t}{4(\Delta x)^2} p_{i+1}^{m+1} + \left( \frac{1}{\Delta t} - \frac{\sigma^2 Q_t}{2(\Delta x)^2} + \frac{\mu}{\Delta x} \right) p_i^{m+1} + \frac{\sigma^2 Q_t}{4(\Delta x)^2} p_{i-1}^{m+1} \\ & = - \frac{\sigma^2 Q_t}{4(\Delta x)^2} p_{i+1}^m + \left( \frac{1}{\Delta t} + \frac{\sigma^2 Q_t}{2(\Delta x)^2} + \frac{\mu}{\Delta x} \right) p_i^m - \frac{\sigma^2 Q_t}{4(\Delta x)^2} p_{i-1}^m \end{aligned} \quad (28)$$

where  $i = \{0, \dots, n_p\}$ , and  $m = \{0, \dots, n_p\}$ . The equations form the system defined as  $Ap^{m+1} = Bp^m$ . The elements of the matrices  $A$  and  $B$  are obtained as

$$\begin{aligned} A = \{a_{ij}\}, \quad \text{where} \quad & a_{i,i} = \frac{1}{\Delta t} - \frac{\sigma^2 Q_t}{2(\Delta x)^2} + \frac{\mu}{\Delta x} & a_{i,i+1} = -\frac{\sigma^2 Q_t}{4(\Delta x)^2} & a_{i+1,i} = -\frac{\sigma^2 Q_t}{4(\Delta x)^2} \\ B = \{b_{ij}\}, \quad \text{where} \quad & b_{i,i} = \frac{1}{\Delta t} + \frac{\sigma^2 Q_t}{2(\Delta x)^2} + \frac{\mu}{\Delta x} & b_{i,i+1} = \frac{\sigma^2 Q_t}{4(\Delta x)^2} & b_{i+1,i} = \frac{\sigma^2 Q_t}{4(\Delta x)^2} \end{aligned}$$

and

$$P^{m+1T} = ( p_0 \quad p_1 \quad \dots \quad p_{n_p} )$$

The matrices  $A$  and  $B$  are tridiagonal due to the structure of the base polynomials.  $A^{-1}$  exists for some set of  $n_p + 1$  distinct points. The system  $p^{m+1} = A^{-1}Bp^m$  is then solved after generating an initial state  $p_0^m \sim \mathbf{Unif}(0, 1)$ . The solution is expressed in terms of a spectral expansion of the variables of uncertainty  $p_t(x_t, t, \xi) = \sum_{i=0}^{n_p} p_i(x_t, t) \phi_i(\xi)$ , with  $\phi_i(\xi)$  a polynomial of Laguerre given that the data used for the application is that of precipitations,  $\xi \sim \mathbf{Gamma}(a, b)$ , and  $p_i(x_t, t) = p^{m+1}$ . Once the system of stochastic differential equations is solved, the ensembles of the solution states are generated using a technique of data assimilation. The general ensemble Kalman filter procedure will be now defined and expressed in terms of the polynomial chaos to serve as the basis for the PEnKF algorithm.

#### 4. Ensemble Kalman filter (EnKF)

The EnKF is a sequential method that originated in data assimilation techniques used in many areas of applied science such as ocean modelling, weather predicting in meteorology and the development of models in Geoscience. It has been found especially useful in situations involving complex dynamic systems in spaces of high dimension. To fix notation, suppose that the states  $\mathbf{x}_{0:t}$  at a time  $t$  are obtained from the previous states  $\mathbf{x}_{0:t-1}$  generated at time  $t - 1$ . This procedure is called previous knowledge in the assimilation cycle  $t$ , and is denoted by  $\mathbf{x}_t^f$ . Given the experimental observations  $\mathbf{y}_{1:t}$ , the Bayes theorem of linear adjustment is used to update the states. This process, called analysis, is denoted by  $\mathbf{x}_t^a$ . The EnKF developed by [10], [11], [12], [13] is an algorithm based on sequential Monte Carlo methods used to approximate the posterior distribution of the states of high dimensional nonlinear systems. The algorithm starts with an approximation of an ensemble of random state solutions. Let

$$(\mathbf{x}_t^f)_i, \quad i = 1, \dots, N, \quad N > 1 \quad (29)$$

be an ensemble of state forecasts  $\mathbf{x}_t^f$ , where each member of the ensemble is indexed by  $i = 1, \dots, N$ , and is obtained from a solution of equation (20). The analysis step of EnKF is given by

$$(\mathbf{x}_t^a)_i = (\mathbf{x}_t^f)_i + \mathbf{K}_t \left[ (\mathbf{y}_t)_i - \mathcal{H}(\mathbf{x}_t^f)_i \right], \quad i = 1, \dots, N \quad (30)$$

where

$$\mathbf{K}_t = \mathbf{P}_t^f \mathcal{H}^T \left( \mathcal{H} \mathbf{P}_t^f \mathcal{H}^T + \mathbf{R}_t \right)^{-1} \quad (31)$$

$$\mathbf{P}_t^f = E \left[ (\mathbf{x}_t^f - E(\mathbf{x}_t^f)) (\mathbf{x}_t^f - E(\mathbf{x}_t^f))^T \right] \quad (32)$$

and

$$\mathbf{P}_t^a = E \left[ (\mathbf{x}_t^a - E(\mathbf{x}_t^a)) (\mathbf{x}_t^a - E(\mathbf{x}_t^a))^T \right] \quad (33)$$

$\mathbf{P}_t^f$  y  $\mathbf{P}_t^a$  are the approximated covariances of the prediction and analysis respectively, and  $\mathbf{R}_t = E(\mathbf{v}\mathbf{v}^T)$  is the covariance of the observation error.

#### 4.1. Ensemble Kalman Filter based on the Polynomial Chaos

In this section the EnKF based on polynomial chaos methodology is introduced. Consider the equations defined in (2) and (20) and let

$$x_0^f(0) = x_0(\xi), \quad \xi \in \mathbf{R}^n, \quad n \geq 1 \quad (34)$$

called an initial condition, be a solution of the equation (20) where  $\xi$  is a set of random variables parameterized by the initial condition with density function  $\rho(\xi)$ , (see [31] for more details). The prediction states are stochastic variables and can be parameterized by the same set of random variables:

$$x_t^f = p(x_t, t, \xi) \quad (35)$$

The expansion of the polynomial chaos of order  $n_p$  in terms of the solution of the equation given in (20),  $p_{n_p}(x_t, t, \xi)$  can be expressed as

$$p_{n_p}(x_t, t, \xi) = \sum_{|\mathbf{i}|=0}^{n_p} \hat{p}_{\mathbf{i}}(x_t, t) \phi_{\mathbf{i}}(\xi) \quad (36)$$

where  $\mathbf{i} = (i_1, \dots, i_{n_p}) \in N_0^{n_p}$  is a multiple index with  $|\mathbf{i}| = i_1 + \dots + i_{n_p}$ , and

$$\Phi_{\mathbf{i}}(\xi) = \prod_{k=1}^{n_p} \phi_{i_k}(\xi_k), \quad |\mathbf{i}| \leq n_p \quad (37)$$

are orthogonal polynomial base functions constructed as products of single variable polynomials  $\phi_{i_k}(\xi_k)$ . The  $\phi_{i_k}(\xi_k)$  are orthogonal polynomials of order  $i_k$  in the dimension  $\xi_k$  which satisfy

$$E_k [\phi_i(\xi_k) \phi_j(\xi_k)] = \int \phi_i(\xi_k) \phi_j(\xi_k) \rho_k(\xi_k) d\xi_k = \delta_{ij} \quad (38)$$

where  $\delta_{ij}$  is the Kronecker delta, and the polynomials  $\phi_i$  are normalized.

The expansion coefficients  $\hat{p}_{\mathbf{i}}(x_t, t)$  are approximated by an orthogonal projection given by

$$\hat{p}_{\mathbf{i}}(x_t, t) = E [p_{n_p}(x_t, t, \xi) \phi_{\mathbf{i}}(\xi)] = \int p(x_t, t, \xi) \phi_{\mathbf{i}}(\xi) \rho(\xi) d\xi \quad (39)$$

Classical approximation theory guarantees that this is the best approximation in the linear space of polynomials of degree  $n_p$  in the sense of mean squares.

The EnKF algorithm can be summarized as follows. To fix notation let  $\hat{x}_t^b$  and  $\hat{P}_t^b$ , be the mean and the covariance matrix of the posterior distribution filtered at time  $t - 1$ , while  $\hat{x}_t^a$  and  $\hat{P}_t^a$  are the mean and updated covariance matrix of the posterior distribution filtered at time  $t$ . Then perform the following steps:

##### Step 1. Propagation.

Begin by generating  $N$  initial states that are solutions of the equation (20), using a method of finite differences, say  $x_{0,i}^f = p(x_0, 0, \xi)$ , and suppose  $\left\{ x_{0,i}^f \sim N(\hat{x}_0, \hat{P}_0^b), i = 1, \dots, N \right\}$  such that  $E(x_{0,i}^f) = \hat{x}_0$ , and  $Cov(x_{0,i}^f, x_{0,j}^f) = \hat{P}_0^b$ . This is done by factorizing  $\hat{P}_0^b = S_0 S_0^T$ , and defining

$$x_{0,i}^f = \hat{x}_0 + S_0 n_0^i, \quad i = 1, \dots, N, \quad \text{where,} \quad n_0^i \sim N(0, I) \quad (40)$$

The mean of the ensemble is given by

$$\hat{x}_0^b = \frac{1}{N} \sum_{i=1}^N x_{0,i}^f \quad (41)$$



and its covariance by

$$\hat{P}_0^b = \frac{1}{N-1} \sum_{i=1}^N (x_{0,i}^f - \hat{x}_0^b)(x_{0,i}^f - \hat{x}_0^b)^T \quad (42)$$

### Step 2. Predicting the ensemble.

Given  $(\hat{x}_t^b, \hat{P}_t^b)$  in an instant of time  $t$ , let  $\hat{P}_t^b = \hat{S}_t \hat{S}_t^T$  to create an ensemble

$$x_{t,i}^f = \hat{x}_t^b + \hat{S}_t n_t^i, \quad i = 1, \dots, N, \quad \text{where,} \quad n_t^i \sim N(0, I) \quad (43)$$

The  $N$  members of the ensemble forecast at time  $t+1$  are generated by a solution of equation (20), using a method of finite differences, say  $x_{t+1,i}^f = p(x_{t+1}, t+1, \xi)$ , and suppose  $\{x_{t+1,i}^f \sim N(\hat{x}_{t+1}^b, \hat{P}_{t+1}^b), i = 1, \dots, N\}$  such that  $E(x_{t+1,i}^f) = \hat{x}_{t+1}^b$ , and  $Cov(x_{t+1,i}^f, x_{t+1,j}^f) = \hat{P}_{t+1}^b$ . The unbiased estimator is given by

$$\hat{x}_{t+1}^b = \frac{1}{N} \sum_{i=1}^N x_{t+1,i}^f \quad (44)$$

The forecast error is estimated by

$$e_{t+1}^i = x_{t+1,i}^f - \hat{x}_{t+1}^b \quad (45)$$

The unbiased estimator of the sample covariance is given by

$$\hat{P}_{t+1}^b = \frac{1}{N-1} \sum_{i=1}^N (x_{t+1,i}^f - \hat{x}_{t+1}^b)(x_{t+1,i}^f - \hat{x}_{t+1}^b)^T + Q_{t+1} \quad (46)$$

In practice, it is not common to estimate  $\hat{P}_{t+1}^b$ . Instead calculate

$$\hat{P}_{t+1}^{cr} = \frac{1}{N-1} \sum_{i=1}^N (x_{t+1,i}^f - \hat{x}_{t+1}^b) \left[ \mathcal{H}(x_{t+1,i}^f) - \mathcal{H}(\hat{x}_{t+1}^b) \right]^T \quad (47)$$

and

$$\hat{P}_{t+1}^{pr} = \frac{1}{N-1} \sum_{i=1}^N \left[ \mathcal{H}(x_{t+1,i}^f) - \mathcal{H}(\hat{x}_{t+1}^b) \right] \left[ \mathcal{H}(x_{t+1,i}^f) - \mathcal{H}(\hat{x}_{t+1}^b) \right]^T \quad (48)$$

where  $\hat{P}_{t+1}^{cr}$  is the sample cross covariance of the previous ensemble and its predicted projection in the observation space, while  $\hat{P}_{t+1}^{pr}$  is the sample cross covariance of the predicted projection of the previous ensemble within the observation space.

### Step 3. Data Assimilation.

If no data is available at time  $t$ , the update step is  $\{x_{t,i}^a = x_{t+1,i}^f, \quad i = 1, \dots, N\}$ . If observations are available at time  $t$ , then the update is done using perturbed observations as follows. First generate observations using the observation equation

$$y_{t+1}^i = \mathcal{H}(x_{t+1,i}^f) + v_{t+1}^i, \quad i = 1, \dots, N, \quad v_{t+1}^i \sim N(0, R_{t+1}) \quad (49)$$

This supplies the sample of the joint distribution  $P(x_t, y_t | \mathbf{y}_{1:t-1})$ . The update is performed using the linear adjustment of Bayes

$$\hat{x}_{t+1}^a = \hat{x}_{t+1}^b + \mathbf{K}_{t+1}(y_{t+1} - y_{t+1}^i) \quad (50)$$

and

$$\hat{P}_{t+1}^a = \hat{P}_{t+1}^b - \mathbf{K}_{t+1}(\hat{P}_{t+1}^{cr})^T \quad (51)$$

where

$$\mathbf{K}_{t+1} = \hat{P}_{t+1}^{cr}(\hat{P}_{t+1}^{pr} + R_{t+1})^{-1} \quad (52)$$

is the Kalman gain matrix.

#### 4.2. Parallelized Ensemble Kalman Filter (PEnKF)

In contrast to previous decades, computers now exist which have one or more CPUs with multi-core processors which permit the reprogramming of sequential algorithms in parallel. One of the advantages of using parallel programming is that it is possible to divide large data sets in smaller data sets which can be evaluated separately. The results are then combined by simple summing operations. The paradigm of P-thread programming in multi-core processors plays an important role in understanding and improving software through the help of multitasking. For scientific computing, it is thus possible, in principle, to divide the calculations in groups and each group executed in its own processor.

In particular, the POSIX Pthread library of the ANSI C programming environment provides an interface which permits the interaction of p-threads within the execution of a program [46]. The parallelization with the POSIX library of the EnKF is used in this article to perform the filtering of the large precipitation data sets where the data is divided in small lots and each lot is filtered independently. The approach executes simultaneously the EnKF for several meteorological stations independently. The high level algorithm is:

##### a. Initialize

- Read the data of daily precipitation of the meteorological stations.
- Assign the precipitation data to  $r$  independent variables.
- Define the Fokker-Planck equation for the precipitation model given by the equations (20) and (2).
- Read the parameters  $Q_t$ ,  $\mu$ ,  $\sigma$ ,  $\Delta x$  and  $\Delta t$ , for the system obtained from the Fokker-Planck equation given in (28).

##### b. Create threads

- For each variable, create  $P$  threads.
- The  $P$  threads execute the EnKF as functions.
  - In the prediction ensemble of the EnKF function
    - \* Resolve the system of equations  $p^{m+1} = A^{-1}Bp^m$  given in (28) for each time  $t$  using the scientific numerical library GSL.
    - \* Using the values of  $p_i(x_t, t) = p^{m+1}$  obtained in the solution of the system (28), calculate the equation  $p_{n_p}(x_t, t, \xi) = \sum_{i=0}^{n_p} p_i(x_t, t)\phi_i(\xi)$ , to obtain the approximate value of the solution of the stochastic differential equation given in (20).
- The EnKF calculates the posterior mean and covariance.

##### c. Finalize

- Calculate the execution time of the total program.
- Calculate the criteria for model validation.
- Calculate the efficiency and acceleration factors for the  $P$  threads.

To measure the calculation speed of PEnKF, the acceleration and efficiency factors as defined in [46] were used.

## 5. Criteria for Model Validation

To compare the quality of the predictions and forecasts obtained from the adjusted model, we use the two validation criteria proposed in [5]: the root mean square error (RMSE), and the mean absolute error (MAE). These validation criteria are defined as:

1. Root mean square error

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (\hat{x}_t^j - x_t^j)^2} \quad (53)$$

2. Mean absolute error

$$MAE = \frac{1}{N} \sum_{j=1}^N |(\hat{x}_t^j - x_t^j)| \quad (54)$$

where

- $x_t^j$  is the true state for the  $j$ -th simulation,  $j = 1, \dots, N$ .
- $\hat{x}_t^j$  is the Monte Carlo estimator of  $E(x_t | \mathbf{y}_{1:t})$  for the  $j$ -th signal,  $j = 1, \dots, N$ .

## 6. Results

In this work, a space time model was implemented based on the stochastic differential equation for the prediction of rainfall using the daily precipitation series from January 2011 through May 2012 for ten (10) meteorological stations in Venezuela, three located in Aragua state (Ceniap, Tamarindo and Tucutunemo), two in Portuguesa state (Araure and Turen), one in Merida state (Mucuchies), one in Barinas state (La Planta), one in Guarico state (San Pedro), one in Tachira state (Pueblo Hondo) and another in Yaracuy state (Yaritagua), (see Figure 1). The data is available in <http://agrometeorologia.inia.gob.ve/>. Missing data was estimated by generating independent random variables with the algorithm of Gibbs and sampling from a truncated normal distribution ( see [37] for more details). The EnKF algorithm was programmed in parallel using the POSIX (Pthread) library in the ANSI C programming environment, in an Intel CPU Core i7 3.6GHz with 16GB RAM running 64Bit Debian Linux.

The EnKF was parallelized supposing that the meteorological stations are independent. This supposition is reasonable since the stations are located in different regions distant one from another and permits performing each analysis separately. Even though it is possible to consider a covariance structure among the stations and perform a joint analysis, this is not studied in this article because the objective is to study the prediction capacity of PENKF. To initialize the algorithm in the stations of Aragua state, the following prior distributions were used:

- Ceniap station  $Q_t = 1$ ,  $\sigma_v^2 = 100$ ,  $\mu = 18.56$ ,  $\sigma = 5.10^4$ ,  $\Delta x = 10^{-2}$  and  $\Delta t = 10^{-6}$ .
- Tamarindo station  $Q_t = 1$ ,  $\sigma_v^2 = 10$ ,  $\mu = 18.56$ ,  $\sigma = 10$ ,  $\Delta x = 10^{-1}$  and  $\Delta t = 10^{-1}$ .
- Tucutunemo station  $Q_t = 1$ ,  $\sigma_v^2 = 1000$ ,  $\mu = 18.56$ ,  $\sigma = 10^{-1}$ ,  $\Delta x = 10^{-1}$  and  $\Delta t = 10^{-1}$ .

In Figure (2) the real daily precipitation data is shown for the Ceniap station located in Aragua state, together with the posterior means of the states estimated by PENKF. A good approximation is observed between the real and estimated data. The other stations, Tamarindo and Tucutunemo, have a similar behaviour.

To initialize the PENKF for the stations of Portuguesa state, the following prior distributions were used:

- Araure station  $Q_t = 1$ ,  $\sigma_v^2 = 10^6$ ,  $\mu = 18.56$ ,  $\sigma = 25.10^3$ ,  $\Delta x = 10^{-1}$  and  $\Delta t = 10^{-1}$ .
- Turen station  $Q_t = 1$ ,  $\sigma_v^2 = 1000$ ,  $\mu = 18.56$ ,  $\sigma = 100$ ,  $\Delta x = 10^{-3}$  and  $\Delta t = 10^{-3}$ .

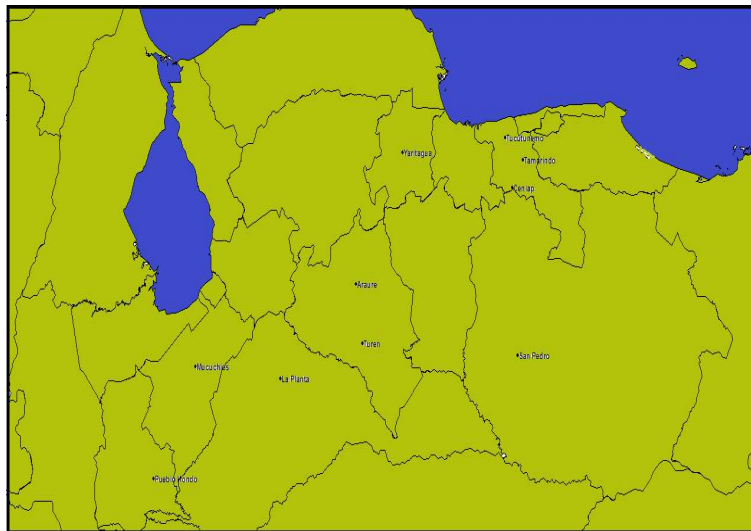


Figure 1. Meteorological stations.

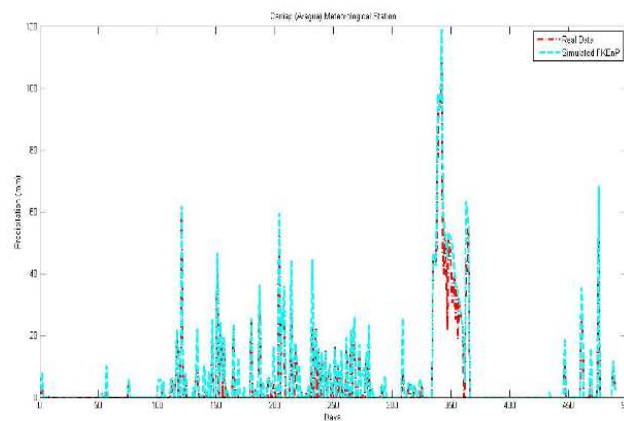


Figure 2. Real data and states estimated by PEnKF for the Ceniap station.

To initialize the PEnKF for the station in Merida state the following prior distributions were used:

- Mucuchies station  $Q_t = 1$ ,  $\sigma_v^2 = 15 \cdot 10^3$ ,  $\mu = 18.56$ ,  $\sigma = 4 \cdot 10^4$ ,  $\Delta x = 10^{-1}$  and  $\Delta t = 10^{-1}$ .

Figure (3) shows the real daily precipitation data for the Araure station located in Portuguesa state, together with the posterior mean of the states estimated by PEnKF. The graph shows a good correspondence between the predicted and real data. The Turen station shows similar behaviour.

Figure (4) shows the real daily precipitation data for the Mucuchies station located in Merida state, together with the posterior states estimated by PEnKF. As in the other cases, the algorithms for the reconstruction of the data perform well.

To initialize the PEnKF for the stations of Barinas state the following prior distributions were used:

- La Planta station  $Q_t = 1$ ,  $\sigma_v^2 = 10^7$ ,  $\mu = 18.56$ ,  $\sigma = 5 \cdot 10^6$ ,  $\Delta x = 10^{-2}$  and  $\Delta t = 10^{-2}$ .
- San Pedro station  $Q_t = 1$ ,  $\sigma_v^2 = 100$ ,  $\mu = 18.56$ ,  $\sigma = 10^6$ ,  $\Delta x = 10^{-4}$  and  $\Delta t = 10^{-4}$ .

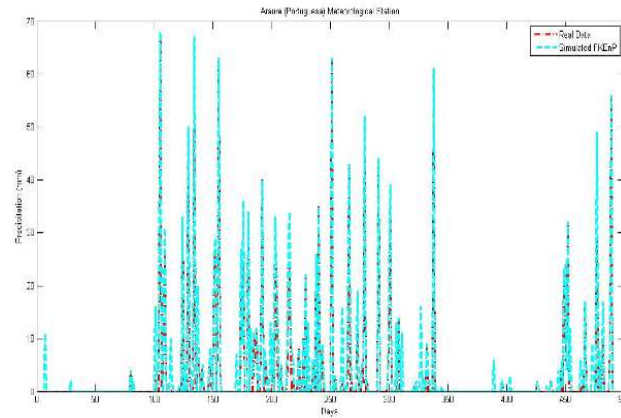


Figure 3. Real data and states estimated by the PEnKF for the Araure station.

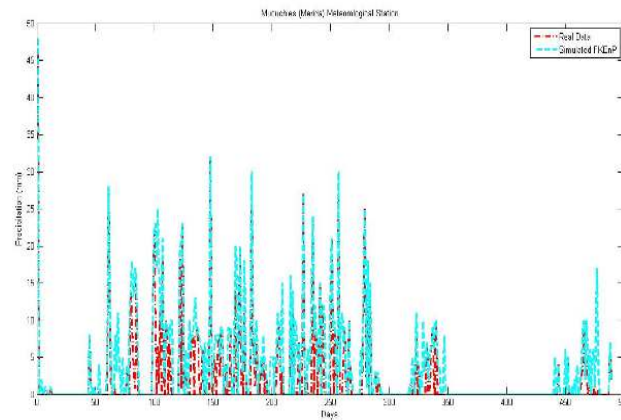


Figure 4. Real data and states estimated by PEnKF for the Mucuchies station.

- Pueblo Hondo station  $Q_t = 1$ ,  $\sigma_v^2 = 5.10^4$ ,  $\mu = 18.56$ ,  $\sigma = 10^5$ ,  $\Delta x = 10^{-1}$  and  $\Delta t = 10^{-1}$ .
- Yaritagua station  $Q_t = 1$ ,  $\sigma_v^2 = 1$ ,  $\mu = 18.56$ ,  $\sigma = 100$ ,  $\Delta x = 10^{-1}$  and  $\Delta t = 10^{-1}$ .

Figure (5) shows the real daily precipitation data for the La Planta, San Pedro, Pueblo Hondo and Yaritagua stations along with the posterior means estimated by PEnKF. Again there is a good correspondence between the real and predicted data.

For rainfall predictions, the data series from January 2011 to May 2012 were used to predict the precipitation for the first day of June 2012. Figure 6 shows the distribution of the average daily rainfall in millimeters (mm) for the first day of June in the 10 meteorological stations under consideration; the left panel shows a map of the real rainfall and the right panel shows a map of the predicted rainfall. Considering that only a small sample of stations was used, the precipitation model does a good job of reproducing the main spatial features. The range of rainfall greater than 50 mm occurs in San Pedro (Guarico) and is represented by only one station. Analogously, the range of rainfall less than 2 mm occurs in the stations of Yaritagua and Tucutumemo.

Rainfall volume for Guarico state is 60 mm, while the rest of the stations present a volume of less than 15 mm which implies that the rainfall decreases from east to west. The number of stations adequately represent the main

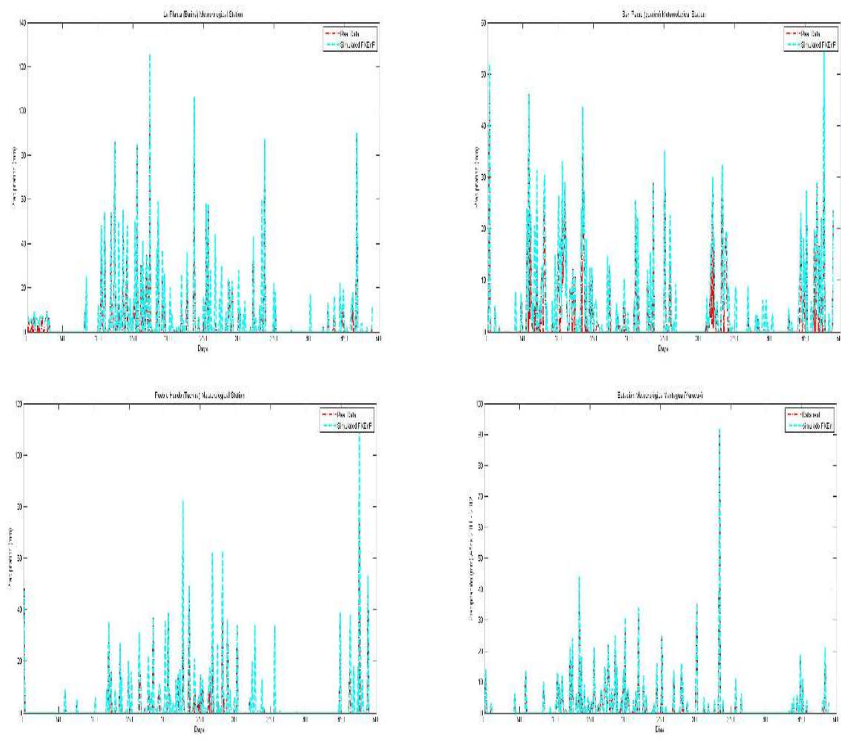


Figure 5. Real data and states estimated by PENKF for the stations La Planta, San Pedro, Pueblo Hondo and Yaritagua.

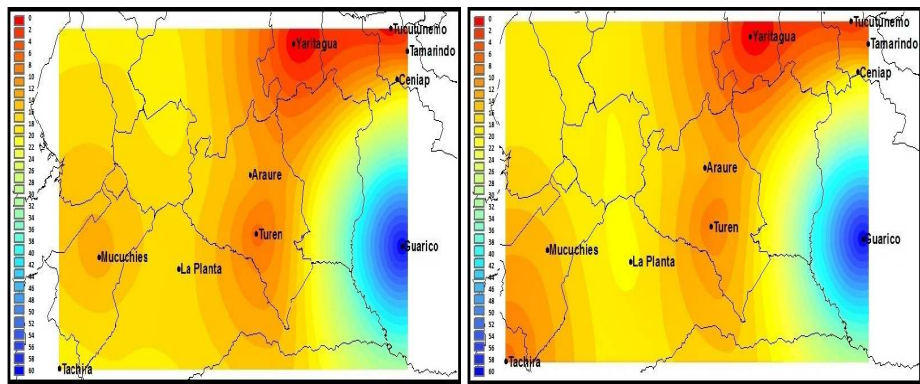


Figure 6. Real and 24 hour predicted precipitation data for the ten stations.

features of the rainfall pattern although the precipitation varies greatly both spatially and temporally in Venezuela. The topography introduces significant variations in the east zone (Guarico state) where an increase in rainfall is seen in contrast to the eastern zone (Tachira state) that presents a decrease in the rainfall.

The rainfall predictions for the Turen, Araure, Mucuchies, la Planta, Tamarindo, and San Pedro stations correspond exactly to the real values whereas the predicted rainfall for the Yaritagua, Tucutunemo, Ceniap, and Pueblo Hondo stations are close to the real values.

Tables (I) and (II) show two goodness of fit criteria for the prediction model applied to the meteorological stations under consideration. Low error levels and little significant difference between the two measures is indicative that the estimator is very close to the true signal.

| Measures | Ceniap   | Tamarindo | Tucutunemo | San Pedro | Pueblo Hondo |
|----------|----------|-----------|------------|-----------|--------------|
| RMSE     | 0.38308  | 0.128569  | 0.124764   | 0.65087   | 0.42141      |
| MAE      | 0.146748 | 0.1652999 | 0.1556601  | 0.423638  | 0.177584     |

Table I. Goodness of fit measures for the considered stations

| Measures | La Planta | Mucuchies | Araure   | Turen   | Yaritagua |
|----------|-----------|-----------|----------|---------|-----------|
| RMSE     | 0.66588   | 0.38381   | 0.34471  | 0.13673 | 0.7596    |
| MAE      | 0.443398  | 0.147308  | 0.118825 | 0.18696 | 0.5770    |

Table II. Goodness of fit measures for the considered stations

Table (III) shows the speed and efficiency factors of the PEnKF for different numbers of  $P$ -threads. To measure the execution time of the algorithm the function `gettimeofday`, of the `sys/time.h` library was used. The execution time of the sequential algorithm EnKF was 1396340128  $\mu$ s for the 10 stations, 520 precipitation states and 100 ensemble members.

| Pthread  | Time ( $\mu$ s) | Speedup | Efficiency |
|----------|-----------------|---------|------------|
| $P = 2$  | 1299109927      | 1.074   | 0.537      |
| $P = 3$  | 1427120646      | 0.978   | 0.326      |
| $P = 4$  | 1330520376      | 1.049   | 0.262      |
| $P = 8$  | 1473391756      | 0.947   | 0.118      |
| $P = 16$ | 1396700306      | 0.999   | 0.062      |
| $P = 32$ | 1312017725      | 1.064   | 0.033      |

Table III. Acceleration and efficiency of PEnKF

Table (III) shows a faster execution of the PEnKF algorithm with respect to the sequential version for  $P = 2$ ,  $P = 4$  and  $P = 32$  threads. As the number of  $P$ -threads increase, the efficiency lowers. Considering this, the ideal case would be  $P = 2$  which provides a 7% increase in speed with respect to the sequential algorithm.

## 7. Conclusions and Discussions

This article presents an implementation of the PEnKF algorithm based on the polynomial chaos theory of Wiener-Askey. After using a finite difference method to determine a solution of the Fokker-Planck stochastic differential equation which models rainfall, the solution is expressed in terms of a truncated expansion of the polynomial chaos



which is then sampled to generate the ensembles. The state estimation uses bayesian inference in a two step data assimilation process. First, the states of the system are predicted using the prior information at time  $t - 1$  and second, the posterior mean and covariance of the predicted states are updated at time  $t$  using the Bayes linear adjustment. A graphical representation of the real data is shown together with the predicted states (see Figures (2), (3), (4), and (5)) showing that the algorithm reconstructs the true states with high precision. Maps are used to show the real rainfall in the ten stations and the posterior means of the estimated rainfall (see Figure (6)). These maps show that the principal spatial features are reproduced exactly for six of the ten stations and almost exactly for the other four stations. To validate the quality of the estimations, the root mean squared error and the mean absolute error of the posterior states were calculated and little variability between the two goodness of fit measures was observed showing that the model predicts adequately. Finally, the performance of the parallel algorithm was measured by determining the acceleration and efficiency factors which showed an optimal performance for  $P = 2$  threads with a speed increase of 7% with respect to the sequential version of the algorithm.

## REFERENCES

1. E. Allen, *Modeling with Itô Stochastic Differential Equations*, Published by Springer, Texas Tech University, USA, 2007.
2. B. Anderson and J. Moore, *Optimal Filtering*, Prentice-Hall, 1979.
3. C. Andrieu and J. Thoms, *A tutorial on adaptive MCMC*, Statistics and Computing, Volume 18, Number 4, 343-373, 2008. DOI:10.1007/s11222-008-9110-y
4. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, *A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking*, IEEE Trans. Sig. Proc., Vol. 50(2), PP. 174-188, 2002
5. K. Bakar, *Bayesian Analysis of Daily Maximum Ozone Levels*, Thesis for the degree of Doctor of Philosophy, 2011. <http://eprints.soton.ac.uk>
6. T. Bengtsson, P. Bickel and B. Li, *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, Probability and Statistics: Essays in Honor of David A. Freedman, vol. 2, pp.316-334, 2008.
7. J. Crank and P. Nicolson, *A practical method for numerical evaluation of solutions of partial differential equations of the heat conduction type*, Proc. Camb. Phil. Soc., Vol. 43, pp. 50-67, 1947. doi:10.1007/BF02127704.
8. A. Doucet, J. DE Freitas and N. Gordon, *Sequential Monte Carlo in Practice*, Springer Verlag, New York, 2001.
9. A. Doucet, S. Godsill and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statist. Comput., vol. 10, no. 3, pp. 197-208, 2000.
10. G. Evensen, *Data assimilation: the ensemble Kalman Filter*, Second Edition, Springer, 2009.
11. G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res. Vol. 99, pp. 10143-10162, 1994.
12. G. Evensen, *The ensemble Kalman filter: theoretical formulation and practical implementation*, Ocean dynamics, Vol. 53, pp. 343-367, 2003.
13. G. Evensen and P. van Leeuwen, *Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model*, Mon. Wea. Rev., Vol. 124, pp. 85-96, 1996.
14. P. Fearnhead, *Markov chain Monte Carlo, sufficient statistics, and particle filters*, J. Computat. Graph. Statist., vol. 11, no. 4, pp.848-862, 2002.
15. W. Fong, S. Godsill, A. Doucet and M. West, *Monte Carlo Smoothing With Application to Audio Signal Enhancement*, IEEE Transactions on Signal Processing, Vol. 50, pp. 438-449, 2002.
16. A. Gelb, *Applied Optimal Estimation*, MIT Press, 1974.
17. S. Godsill, A. Doucet and M. West, *Monte Carlo Smoothing for Nonlinear Time Series*, Journal of the American Statistical Association, Vol. 99, pp. 156-167, 2004.
18. N. Gordon, D. Salmond and A. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proceedings F on Radar and Signal Processing, 140 (2): 107-113, 1993. doi:10.1049/ip-f-2.1993.0015.
19. D. Higdon, *A process-convolution approach to modeling temperatures in the north Atlantic Ocean*, Journal of Environmental and Ecological Statistics, 5, 173-190, 1998.
20. A. Hurn, K. Lindsay and V. Martin, *On the efficacy of simulated maximum likelihood for estimating the parameters of stochastic differential equations*, Journal of Time Series Analysis, Vol. 24, pp. 45-63, 2003.
21. S. Infante, J. Rojas, A. Hernández and V. Cartaya, *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, Vol 62, 178, 5-36, 2010.
22. A. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, Inc. New York, USA, 1970.
23. S. Julier and J. Uhlmann, *A new extension of the Kalman filter to nonlinear systems*, In Proc. AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls, pp. 182-193, 1997.
24. S. Julier and J. Uhlmann, *Unscented Filtering and Nonlinear Estimation*, Proceedings of the IEEE 92, 3, 401-422, 2004.
25. G. Kitagawa, *Monte Carlo Filter and Smoother for Nonlinear Non Gaussian State Models*, Journal of Computation and Graphical Statistics. Vol. 5, pp. 1-25, 1996.
26. P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, 1992.



27. D. Lee, *Nonlinear Bayesian Filtering with Applications to Estimation and Navigation*, Thesis for the degree of Doctor of Philosophy, 2005. <http://repository.tamu.edu/bitstream/handle/1969.1/2269/etd-tamu-2005A-AERO-Lee.pdf?sequence=1>
28. M. Lin, J. Zhang, Q. Cheng and R. Chen, *Independent Particle Filters*, Journal of the American Statistical Association, Theory and Methods. 100: 472, 1412-1421, 2005.
29. J. Liu and R. Chen, *Sequential Monte Carlo Methods for Dynamical Systems*, Journal of the American Statistical Association, Vol. 83, pp. 1032-1044, 1998.
30. J. Liu and M. West, *Combine Parameter and State Estimation in Simulation Based Filtering*. In Sequential Monte Carlo Methods in Practice, Eds. A. Doucet, J. F. G. De Freitas, and N. J. Gordon, New York: Springer Verlag, pp. 197-217, 2000.
31. J. Li and D. Xiu, *A Generalized Polynomial Chaos Based Ensemble Kalman Filter with High Accuracy*, Journal of Computational Physics, Vol. 228, No. 15, 2009, pp. 5454-5469, 2009. doi:10.1016/j.jcp.2009.04.029
32. R. Madankan, *Polynomial Chaos Based Method for State and Parameter Estimation*, M.S. Thesis, Univ. at Buffalo, State University of New York, Buffalo, NY, 2011.
33. B. Ø Kesimal, *Stochastic Differential Equations, An Introduction with Applications*, Springer, fifth edition, 2000.
34. E. Platen, *An introduction to numerical methods for stochastic differential equations*, Acta Numerica, pp. 197-246, 1999.
35. A. Pedersen, *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*, Scandinavian Journal of Statistics, 22, 55-71, 1995.
36. M. Pitt and N. Shephard, *Filtering via simulation: Auxiliary particle filters*, J. Amer. Statist. Assoc. , vol. 94, no. 446, pp. 590-599, 1996.
37. L. Sánchez, S. Infante, V. Griffin and D. Rey, *Spatiotemporal dynamic model and parallelized ensemble Kalman Filter for precipitation data*, Reporte técnico, Departamento de Matemáticas, Facultad de Ciencia y Tecnología, Universidad de Carabobo, 2014.
38. L. Sánchez and S. Infante, *Reconstruction of chaotic dynamic systems using non-linear filters*, Chilean Journal of Statistics. Vol. 4, No. 1, 35-54, 2013.
39. D. Simon, *Optimal State Estimation Kalman, H Infinity, and Nonlinear Approaches*, John Wiley and Sons, publication, 2006.
40. C. Snyder, T. Bengtsson, P. Bickel and J. Anderson, *Obstacles to high-dimensional particle filtering*, Monthly Weather Review 136(12): 4629-4640, 2008.
41. G. Storvik, *Particle filters for state-spacemodels with the presence of unknown static parameters*, IEEETrans. Signal Process., vol. 50, no. 2, pp. 281-289, 2002.
42. O. Stramer and M. Bognar, *Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach*, Bayesian Analysis, Volume 6, Number 2, pp. 231-258, 2011.
43. O. Stramer, M. Bognar and P. Schneider, *Bayesian Inference of Discretely Sampled Markov Processes with Closed-Form Likelihood Expansions*, The Journal of Financial Econometrics, 8: 450-480, 2010.
44. M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer, (2nd Ed.) 1997.
45. C. Wikle and S. Holan, *Polynomial nonlinear spatio-temporal integro-difference equation models*, Journal of Time Series Analysis. 32, 339-350, 2011. DOI:10.1111/j.1467-9892.2011.00729.x
46. B. Wilkinson and M. Allen, *Parallel Programming: Techniques and Application Using Networked Workstations and Parallel Computers* 2nd edition Prentice-Hall Inc, 2005.
47. L. Wuan, *Wiener Chaos Expansion and Numerical Solutions of Stochastic Partial Differential Equations*, Thesis for the Degree of Doctor of Philosophy California Institute of Technology Pasadena, California, 2006.
48. D. Xiu and G. Karniadakis, *The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations*, SIAM Journal on Scientific Computing, Vol. 24, No. 2, pp. 619-644, 2002.