COMPARISON OF UNIVARIATE AND BIVARIATE APPROACHES TO MAP PRECIPITATION USING GEOSTATISTICS AND THE KALMAN FILTER

ÁVILA-CARRASCO, J. R.[*] – JÚNEZ-FERREIRA, H. E. – GONZÁLEZ-TRINIDAD, J. – VILLALOBOS DE ALBA, A. A. – BAUTISTA-CAPETILLO, C. F.

*Doctorado en Ciencias de la Ingeniería, Universidad Autónoma de Zacatecas,
Av. Ramón López Velarde No. 801, Col. Centro, Zacatecas, Zacatecas, México, C.P. 98000
Tel: +492-924-2432; Fax: +01-492-925-6690 Ext. 1613*

[*]*Corresponding author
e-mail: roberto.avila@yandex.com*

Abstract. Measured rainfall data are very important in agriculture and environmental science. However, in many cases, the information gathered by existing rain gauges is insufficient for characterizing climatic variation within a study area. Thus, the use of interpolation techniques is necessary to predict values to unsampled sites. In this work, the performances of geostatistical algorithms, such as ordinary kriging and ordinary cokriging, and a proposed Kalman filter method were compared for mapping rainfall. The analysis was performed using both univariate and bivariate approaches. Natural terrain elevation was taken as the auxiliary variable for the bivariate case. The analysis was conducted for specific months of the dry and wet seasons in the Santiago River basin in Mexico. After comparison of the statistical errors, it was established that the geostatistical methods provided excellent results (especially cokriging) for the wet season months, with good correlation of 0.7 or above between rainfall and elevation, but not for the dry season months. Nevertheless, good results were achieved for the dry season months using the proposed Kalman filter methodology, due to the high normality and spatial dependence of the sample in this period.
Keywords: *climatic variation, correlation, auxiliary variable, cokriging, Kalman filter*

Introduction

Knowledge of spatiotemporal rainfall variation is essential in hydrology. In most basins in Mexico, the low density of weather stations that constitute the monitoring network is a hindrance to the realization of robust scientific investigation. Additionally, inconsistent field data and missing or short records are other common problems. The consequence of these shortcomings is the lack of sufficient data for the characterization of actual climatic variations, which limits the validity of the conclusions of climatological and hydrological studies (Costa and Soares, 2009).

Some studies require high-resolution data of the spatial variability of rainfall for accurate analysis of extreme hydrological events such as floods and droughts. To achieve this objective, a highly dense network of rain gauges would be required, which is the reason researchers adopt interpolation methods to obtain values at ungauged sites.

Geostatistics represents a set of widely used techniques based on the theory of regionalized variables. It capitalizes on the spatial correlation between neighboring observations to predict values at ungauged locations. Its application has become an increasingly important tool in climatic research, as demonstrated by several authors who have shown that such prediction methods provide better estimates than some conventional techniques (Awadallah, 2012; Béal et al., 2002; Soenario et al., 2010; Moral 2009; Shaghaghian and Abedini, 2012).

To obtain reasonable rainfall estimates, it is necessary to consider all the factors involved in conjunction with techniques that explicitly take into account the error associated with each measurement. One of these techniques is estimation theory, which represents a method for combining measurements and predicted values. The Kalman filter method is a special case of estimation theory that allows the calculation of the uncertainty of the independent variables involved.

The Kalman filter has been applied in many surface hydrology studies. For example, Morales et al. (2014) evaluated the applicability of the discrete Kalman filter algorithm for the prediction of short-term flows in reservoirs with very acceptable results. McMillan et al. (2013) used an ensemble recursive Kalman filter for information assimilation in an operational iterative runoff forecast system. This system improved the generated estimates considerably because of the enhancement of the initial conditions attributable to the Kalman filter. Komma et al. (2008) also used an ensemble Kalman filter for the prediction of real-time flows to update soil moisture data in a rainfall–runoff model, in addition to forecasting floods. Thus, the updating of states helped reduce the estimation errors substantially. Similar studies have been undertaken by Moradkhani et al. (2005), Srikanthan et al. (2015), and Weerts et al. (2006).

This study evaluated univariate and bivariate estimates of rainfall using the Kalman filter. These results were compared with estimates obtained using the geostatistical methods of ordinary kriging (OK) and ordinary cokriging (OCK). For the bivariate case, natural terrain elevation was adopted as the auxiliary variable.

Rain estimation

Different methods have been proposed for the interpolation of rainfall. The simplest method assigns an actual measurement to the nearest ungauged sites. Other methods commonly used include Thiessen polygons and the "inverse distance" method, which was developed by the US National Weather Service in 1972. In this method, ungauged rainfall is calculated based on the weighted average of the surrounding values. Another method, which uses isohyets, has the advantage of allowing the use of values at each measurement position, in conjunction with those factors that affect rainfall measurements, to generate lines of equal precipitation; nevertheless, a dense network of rain gauges is still required in order to obtain reasonable accuracy.

Geostatistics originated as a branch of statistics. Initially, it was used only in the analysis of spatial variability and modeling in earth sciences; however, its scope has been extended to include the analysis and prediction of any phenomena that vary in space and time. Specifically, geostatistics has become an important tool for the interpolation of weather data to unsampled sites (Waylen et al., 1995; Holawe and Dutter, 1999).

Geostatistical prediction methods such as kriging can provide better estimates than conventional techniques, although results have not been favorable for measurements with high dispersion. Dirks et al. (1998) showed that for stations covering areas >35 km^2, the simple kriging method showed no significant improvement over less complex techniques such as the "inverse distance" method. To address this problem, authors such as Hevesi et al. (1991), Borga and Vizzacaro (1997), Bostan and Akyürek (2009), Pardo (1998), Goovaerts (2000), Majani (2007), Murthy and Abbaiah (2007), Huang and Hu (2008), Moral (2009); Volkmann et al. (2010), and Diaz et al. (2002) have all used multivariate methods (e.g., cokriging and/or kriging) with external drift to fuse rainfall data with elevation data and/or radar data for estimations. Such auxiliary information is

readily available and inexpensive and has dense distribution. In addition, these authors tested and compared these methods against univariate methods, and they agreed that multivariate geostatistical estimation methods provide reasonable results as long as the correlation between rainfall and the auxiliary variables is >0.70.

Static Kalman filter

The Kalman filter is a set of mathematical equations from which linear unbiased estimates can be obtained recursively, and the estimations are achieved with minimum variance for the state of a system using noisy data (Briseño et al., 2011). In particular, the static Kalman filter has been used to estimate pollutant concentrations and quality parameters in groundwater (Herrera, 1998; and Júnez, 2005) as well as to estimate the hydraulic head of an aquifer (Briseño et al., 2011; Júnez and Herrera, 2013). In such applications, the Kalman filter uses an initial covariance matrix derived from transport model flows (Herrera, 1998) or by means of geostatistical analysis of the data (Júnez, 2005). The main objective of such work was the optimum design of monitoring networks, based on the reduction of estimation error variance.

Some authors have used the Kalman filter for the spatial estimation of rainfall. For example, Diaz et al. (2002) obtained the best rainfall estimate with cokriging using radar data as the auxiliary variable. They recommended the consideration of other secondary variables such as topography, and they also proposed to use the Kalman filter in conjunction with a geostatistical approach in their future work. Valdés et al. (2015) compared the results of an application of the Kalman filter technique against kriging in the estimation of rainfall, finding that the Kalman filter provided reasonable estimates in real time and that kriging was suitable for the spatial distribution of rain. Urquiola et al. (2015) also proposed the use of a dynamic time–space model with restricted parameterized covariance matrices to predict rainfall amounts. This was achieved using algorithms such as the Kalman filter and smooth Kalman filter to estimate the unknown states of the system.

Materials and methods

Santiago River watershed case study

The Santiago River basin is located in the central northwestern region of Mexico. It covers an area of 76,274 km^2 with a perimeter of 1923.5 km. The basin partially occupies seven Mexican states: northern parts of Jalisco, southern Zacatecas, Aguascalientes, eastern Nayarit, and small parts of Durango, San Luis Potosi, and Guanajuato (*Fig. 1*). The basin type is exoreic because it discharges into the Pacific Ocean near the town of San Blas, Nayarit. Its highest topographic point is 3130 mamsl. The Santiago River basin belongs to Hydrologic Region number VIII "Lerma-Santiago." This occupies a wide variety of climates and biomes and it is subdivided into the hydrological subregions of Rio Alto Santiago and Rio Bajo Santiago.
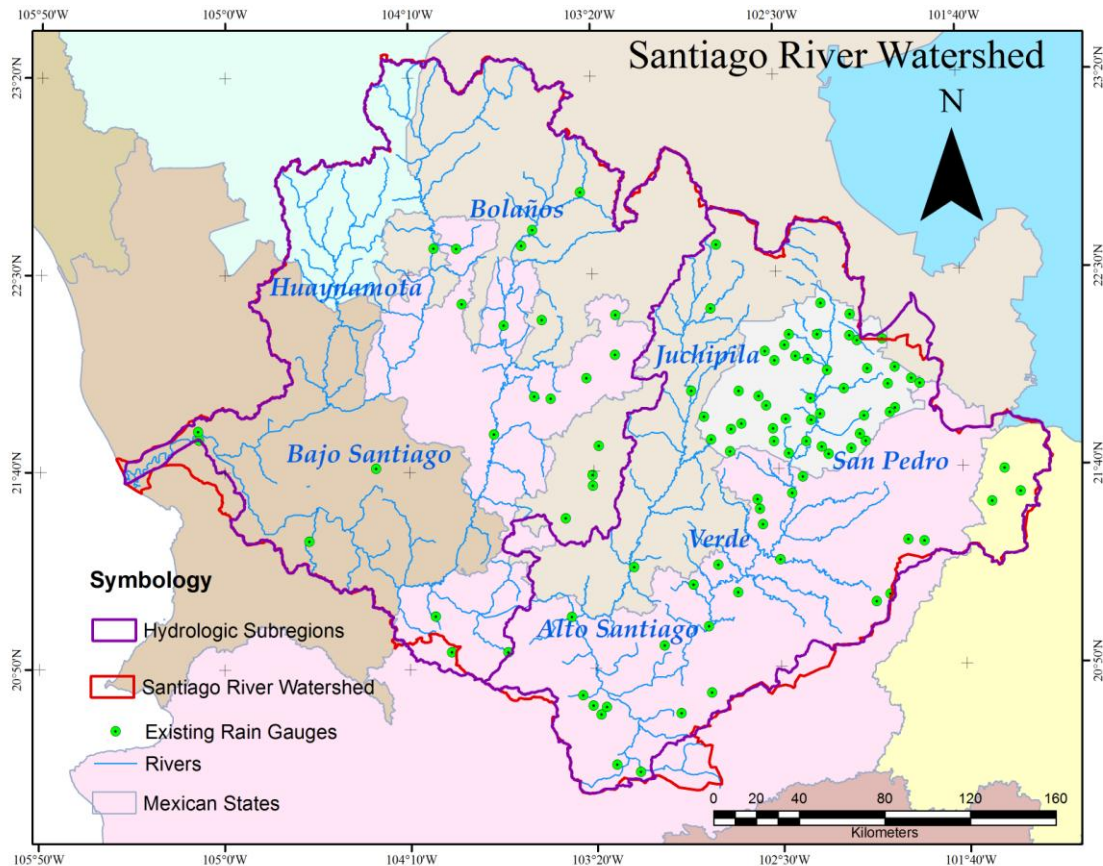
Figure 1. *Santiago River basin and existing rain gauge network*

The Santiago River basin main collector starts from the controlled spills of the Poncitlán dam at the northeastern end of Lake Chapala. It transverses the states of Jalisco, Zacatecas, and Nayarit, to finally discharge into the Pacific Ocean after traveling 524 km. Its most important tributaries are the Verde, Juchipila, Bolaños, and Huaynamota rivers. There are several reservoirs within the basin, which are used primarily for irrigation and energy generation (Gomez et al., 2012).

According to the Mexican National Water Commission (CONAGUA, 2015a), the climate of the region is dry in northern parts of the basin, temperate humid in central regions, and warm and humid in the coastal region. The average annual temperature is 19 °C. The average rainfall is 822 mm/year, 80% of which is concentrated in the months of June–September. The average annual evaporation reaches 1831 mm/year.

Methodology

In the present work, the static Kalman filter was used for univariate and bivariate spatial estimations of rainfall for the Santiago River watershed. The required initial covariance matrix used by the filter was derived by geostatistical analysis. The resulting estimates are compared with univariate and bivariate estimates obtained using the OCK and OK geostatistical techniques, respectively. The methodology for the geostatistical analysis of each parameter followed three steps: exploratory data analysis, structural analysis, and predictions with respective validations.

Exploratory Data Analysis

The principal goal of the exploratory analysis is to characterize the sample of available data, while trying to retain as much information as possible. This analysis is based on conventional statistical techniques to obtain unknown data from a priori information of the sample. This is important in any statistical analysis but particularly in geostatistical analysis. It should also be guaranteed that the sample is not affected by outliers, both distributional and spatial. In other words, the sample should follow a normal function and be trend-free with a homogeneous spatial distribution.

Structural Analysis

The aim of structural analysis is to characterize the spatial structure of a regionalized variable. It is the process of estimating and modeling the function describing the spatial correlation of the involved variables, commonly called the semivariogram. Geostatistics reliability depends on the semivariogram (Subyani, 2015). It describes the spatial and/or temporal correlation of the studied variable through its components, which are the sill, range, and nugget.

Univariate geostatistics estimates a property using known values obtained from neighboring or nearby positions. The function describing the spatial continuity of the variable is the experimental variogram (Eq. 1), to which a theoretical variogram model is adjusted. There are various theoretical variogram models, the most common of which are the spherical, exponential, and Gaussian (Diaz, 2002).

$$\gamma(\mathbf{h}) = \frac{1}{2\mathbf{N(h)}} \sum_{i=1}^{\mathbf{N(h)}} \{\mathbf{Z}(\mathbf{x}_i) - \mathbf{Z}(\mathbf{x}_i + \mathbf{h})\}^2 \qquad \text{(Eq.1)}$$

where $Z(x_i + h)$ and $Z(x_i)$ are the variable values at points $x_i + h$ and x_i, respectively, and N(h) is the number of pairs of data separated by distance h.

If two regionalized variables $Z_{v1}(x)$ and $Z_{v2}(x)$ are considered, the moment estimator function of cross semivariance is given by the cross variogram equation (Giraldo, 2001).

$$\gamma_{v1v2}(\mathbf{h}) = \frac{1}{2\mathbf{N(h)}} \sum^{\mathbf{N(h)}} \{\mathbf{Z}_{v1}(\mathbf{x}+\mathbf{h}) - \mathbf{Z}_{v1}(\mathbf{x})\}\{\mathbf{Z}_{v2}(\mathbf{x}+\mathbf{h}) - \mathbf{Z}_{v2}(\mathbf{x})\} \qquad \text{(Eq.2)}$$

The linear model of coregionalization assumes that all simple semivariograms (Eqs. 3 and 4) and cross variograms (Eq. 5) can be expressed as a linear combination of these theoretical models (Isaaks and Srivastava, 1989). For a case considering only two variables, the equations are:

$$\gamma_{v1}(h) = \alpha_0 \gamma_0(h) + \cdots + \alpha_m \gamma_m(h) \qquad \text{(Eq.3)}$$

$$\gamma_{v2}(h) = \beta_0 \gamma_0(h) + \cdots + \beta_m \gamma_m(h) \qquad \text{(Eq.4)}$$

$$\gamma_{v1v2}(h) = \delta_0 \gamma_0(h) + \cdots + \delta_m \gamma_m(h) \qquad \text{(Eq.5)}$$

where $\gamma_{v1}(h)$ and $\gamma_{v2}(h)$ are the simple semivariograms, $\gamma_{v1v2}(h)$ is the cross semivariogram, $\gamma_0(h)$, $\gamma_1(h)$, ... , $\gamma_m(h)$ are theoretical models of semivariance, and α_i, β_i, and δ_i with $i = 1, 2, \ldots, m$, are constant values. This is expressed in matrix notation as follows:

$$\Gamma(h) = \begin{pmatrix} \gamma_{\upsilon 1}(h) & \gamma_{\upsilon 1 \upsilon 2}(h) \\ \gamma_{\upsilon 1 \upsilon 2}(h) & \gamma_{\upsilon 2}(h) \end{pmatrix} = \sum_{s=0}^{m} B_s \gamma_s(h) \tag{Eq.6}$$

where

$$B_s = \begin{pmatrix} \alpha_s & \delta_s \\ \delta_s & \beta_s \end{pmatrix} \tag{Eq.7}$$

$$\gamma_s(h) = \begin{pmatrix} \gamma_s(h) & 0 \\ 0 & \gamma_s(h) \end{pmatrix} \tag{Eq.8}$$

where $\Gamma(h)$ in Eq. (6) is called the coregionalization matrix.

As in the case of univariate geostatistics, the key is to have a tool that measures the spatial correlation of the variables involved. The spatial correlation of each of the involved variables is obtained through the covariance function or variogram. The joint spatial correlation, or the relationship, is obtained through the cross-covariance function generated from the measurements and from the marginal and cross variogram parameters of the linear model of coregionalization. The cross-covariance matrix of rainfall and topographic elevation data has the following form:

$$P_0 = \begin{bmatrix} \text{Rain} & \text{Rain} - \text{Elevation} \\ \text{Elevation} - \text{Rain} & \text{Elevation} \end{bmatrix} \tag{Eq.9}$$

Predictions

After the structural analysis has been performed, the prediction of values at the unsampled sites can be undertaken using the techniques for univariate and bivariate cases, respectively. In this work, the parameters of the variograms resulting from the geostatistical analysis were used to generate the a priori covariance matrices for the univariate and bivariate estimates using the Kalman filter.

Kalman filter estimations

The Kalman filter is a set of mathematical equations with which linear unbiased estimates can be obtained recursively. Minimum variance estimates are obtained for the state of a system using noisy data (Briseño et al., 2011). Furthermore, it can be used to update an estimate each time new data of the variable become available, without recourse to previous data. It is a recursive filter because the solution is recalculated when a new observation or measurement is incorporated into the system.

The general form of the Kalman filter is based on two equations: a dynamic equation and a measurement equation. The discrete Kalman filter aims to solve the general problem of estimating the state of a discrete time process, which is represented by a linear stochastic equation as follows:

$$X_{k+1} = AX_k + w_k \tag{Eq.10}$$

The measurement of $Z_k \in R^m$ is related linearly to the system state as follows:

$$Z_k = HX_k + v_k \tag{Eq.11}$$

where X_{k+1} is the system state at time $k + 1$, X_k is the state at time k, A is an n × n dimensional matrix that relates the system at time k with the state at time $k + 1$. The H matrix of dimensions n × m relates the state at time k with the measurement Z_k, while w_k and v_k represent the process error and measurement error, respectively. These errors should be independent of each other and should show a normal probability distribution.

This study used the static Kalman filter used by Herrera (1998), which considers only the measurement equation and incorporates time using space–time vectors (Briseño, 2012).

The linear measurement equation of the discrete Kalman filter relates the state vector h of the variable in the positions and times of the desired estimate with the sampled data z:

$$z_j = H_j h + v_j \qquad (Eq.12)$$

where z_j ($j = 1, 2,…$) is a measurement sequence of the variable of interest, H_j is a matrix of 1 × N dimensions that is not only zero at the position that corresponds to the input of h where sample j is taken, and N is the vector of dimension h. H_j is the sampling matrix of j, $h = \{h_{ip}\}$ is the space–time vector with the estimated values of the variable in the positions and times of interest (where h_{ip} is the variable value at position x_i at time t_p), and $\{v_j, j = 1, 2,…\}$ is a scalar value that represents the measurement error. This is a white Gaussian sequence with zero mean and variance r_j. The measurement error sequence (v_j) and vector h are independent.

The error covariance matrix for each variable involved in the estimation is:

$$P^n = E\{(h - \hat{h}^2)(h - \hat{h}^2)^T\} \qquad (Eq.13)$$

where $\hat{h}^n = E\{(h/z_1, z_2,…,z_n)^T\}$ is the expected value of h given the measurements z_1, $z_2,…,z_n$, and T stands for transpose. In this notation, the superscript identifies the number n of measurements used for the estimate (Briseño et al., 2011).

For the application of the Kalman filter, it is required to define an estimation mesh. This mesh has two types of nodes: those that correspond to the positions of the sampling points (S) and those points where the estimates of the variable (E) are required. This estimation method requires a priori estimates of the variable in space ($\hat{h}_0$) and the covariance matrix of the estimation error (P_0). Given these a priori estimates, the linear minimum variance estimate of h can be obtained sequentially through the following formulas:

$$\hat{h}^{n+1} = \hat{h}^n + K_{n+1}(z_{n+1} - H_{n+1}\hat{h}^n) \qquad (Eq.14)$$

$$P^{n+1} = P^n - K_{n+1}H_{n+1}P^n \qquad (Eq.15)$$

$$K_{n+1} = P^n H_{n+1}^T (H_{n+1} P^n H_{n+1}^T - r_{n+1})^{-1} \qquad (Eq.16)$$

where superscripts n and $n+1$ represent the current time and forward time respectively (for this work, these formulas were applied only spatially), $\hat{h}$ is the estimated state, K is the Kalman gain, P is the variance state matrix, H is the measurement matrix, z_n stands for measurements, and r_n is the variance.

Validation

In order to compare the estimates obtained by different methods with the real values of sampling data, the errors can be computed. Error e_i is calculated as the difference of data values of measurement *h* and those of the estimated *h* points of the mesh with each of the different discussed methods. The calculated errors are the mean error (ME, mm), mean squared error (MSE, mm^2), and standard mean square error (SMSE, dimensionless). The ME is an indicator of the overestimation or underestimation of the values of *h*. The MSE is a measure of the magnitude of the estimation error. Theory has established that a good fit exists between the measurements of *h* and the estimates of $\hat{h}$ by the different methods the closer the values of ME and MSE are to zero. Conversely, values of SMSE closer to 1 indicate greater agreement between the values of *h* and $\hat{h}$ estimates. The corresponding equations are as below:

$$ME = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{Eq.17}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (e_i)^2 \tag{Eq.18}$$

$$SMSE = \frac{1}{n}\sum_{i=1}^{n} \frac{(e_i)^2}{\sigma_i^2} \tag{Eq.19}$$

where σ_i^2 is the error estimation variance at position *i*.

Results and discussions

Rainfall data were obtained from CONAGUA (2015b), specifically from the project entitled "Program of preventive measures and drought mitigation." In order to obtain consistent and coherent records, data from these stations were checked and only those that did not show inconsistencies or missing data were selected. This resulted into a monitoring network of 89 weather stations distributed throughout the basin (*Fig. 1*). Then, the exploratory analysis of rainfall records (units: mm) was undertaken with the purpose of describing the sample. *Table 1* shows the obtained values of the principal statistics for monthly and annual periods. The correlation values between these rain gauges records and natural terrain elevation (units: mamsl) are also shown. These were obtained by fitting a simple linear model to the rainfall data and digital elevation model values.

Table 1. *Descriptive statistics of rainfall data for 89 selected rain gauges*

Period	Mean	Min.	Max.	Med.	Kurtosis	Skewness	Variance	σ*	Cor.**
January	19.91	11.00	37.56	19.28	5.04	1.12	28.85	5.37	0.0064
February	11.05	2.30	17.48	10.96	3.81	-0.18	7.36	2.71	0.0003
March	2.79	0.75	8.50	2.76	8.02	1.39	1.37	1.17	0.0739
April	5.48	0.96	17.38	5.34	5.41	0.96	7.47	2.76	0.2134
May	18.69	5.92	29.17	18.74	3.36	-0.42	22.07	4.72	0.2491
June	105.1	48.72	208.33	92.35	3.11	1.00	1471.49	38.5	0.2796
July	167.1	55.78	368.96	149.77	4.53	1.16	3704.12	61.1	0.6750
August	145.2	76.69	468.15	133.61	13.58	2.71	4260.49	65.4	0.8321

September	106.0	52.47	380.78	94.19	17.75	3.37	2496.69	50.0	0.7839
October	40.22	24.05	149.96	36.29	23.88	4.14	309.92	17.6	0.7065
November	9.74	4.71	25.19	9.06	8.74	1.75	9.59	3.10	0.3052
December	10.21	4.76	21.78	9.59	4.70	1.37	12.91	3.60	0.1151
Monthly	53.46	28.79	129.84	48.72	7.61	1.81	325.04	18.0	0.7568
Annual	641.5	345.4	1558.09	584.70	7.61	1.81	46806.19	216.9	0.7568

*σ = Standard deviation of the data sample
**Cor = Linear correlation expressed as the Pearson correlation coefficient (R^2) between rainfall and elevation.

Given the statistics shown in *Table 1*, it is easy to describe the histogram frequency of the monthly rainfall data recorded at the 89 analyzed rain gauges. It can be seen that for dry months favorable values in terms of normality were obtained; however, there is a marked difference between the mean and median values for the wet months. *Table 1* also shows the correlation between rainfall and terrain elevation for the different months. There is poor correlation during the dry season in comparison with the wet season. This occurs because the values of measurements of rain in these two periods are small, and thus so is the variation. Generally, a good linear correlation between precipitation and natural terrain elevation exhibits a value of the Pearson correlation coefficient (R^2) close to or greater than 0.7. The highest correlations were obtained for the wet season (July–October). These correlations were negative, indicating that large values of precipitation occur at low natural terrain elevations (*Fig. 2*). Similarly, we obtained negative correlation between terrain and elevation for average monthly and annual rainfall values.
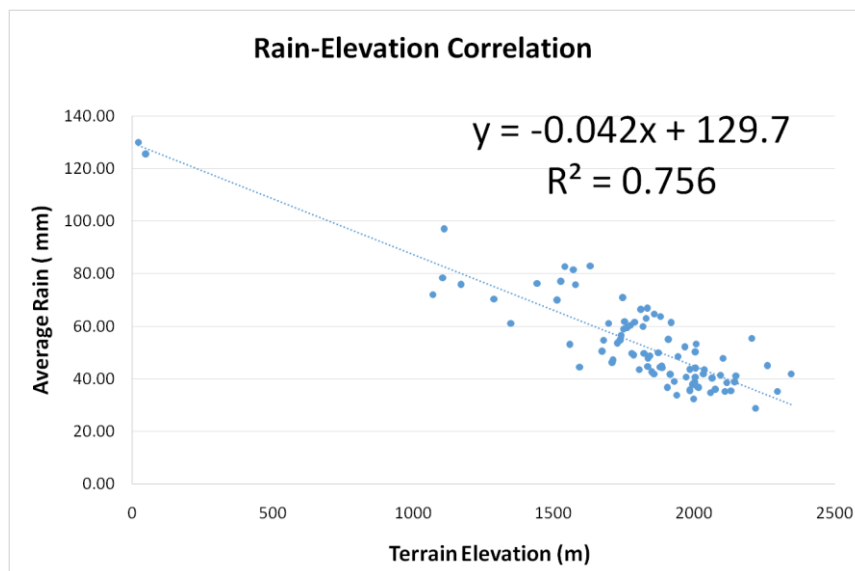

Figure 2. Correlation of monthly precipitation and terrain elevation

Geostatistical estimations

Two geostatistical interpolation methods were first evaluated for the monthly rainfall data: OCK and OK. These algorithms were used to obtain spatial estimates at unmonitored sites, and then the error statistics were evaluated. *Table 2* shows the

adjusted marginal semivariograms and cross covariances for the OCK case, obtained in the structural analysis made for the different algorithms. This was performed using data from 45 of the 89 selected rain gauges. Using these data, rain estimates were obtained using the geostatistical algorithms, which were then evaluated and compared against the values of the remaining 44 rain gauges. The months of February and April were chosen for the dry season, and July and August selected for the wet season. For ease of comparison, maps of the estimates derived using OK and OCK are shown in *Figure 3a–d*, respectively.

Table 2. Adjusted theoretical variograms for different months

METHOD	MONTH	MODEL	NUGGET (m^2)	SILL (m^2)	RANGE (m)
Ordinary Kriging	February	Spherical	3.1495	9.065	201,470
	April	Spherical	0.63267	11.852	145450
	July	Spherical	140.63	14,020	448720
	August	Spherical	16.317	14,574	448720
Ordinary Cokriging	February	Spherical	3.0858	17.56	448720
		Spherical	31374	617220	448720
		CC: Spherical	-	-99.133	448720
	April	Spherical	1.5004	24.162	448720
		Spherical	31374	614220	448720
		CC: Spherical	-	372.28	448720
	July	Gaussian	140.62	14020.29	448718.013
		Gaussian	106889.18	-7682.76	448718.013
		CC: Gaussian	-	671729.35	448718.013
	August	Gaussian	16.317	14574.35	448718.013
		Gaussian	106889.18	-7864.511	448718.013
		CC: Gaussian	-	671729.35	448718.013

CC = Cross covariance

Different patterns were observed in the corresponding variograms for the months in each of the seasons (*Table 2*). Furthermore, the maps of monthly rain also illustrated some differences. In the dry season, which covers the months of February and April, small precipitation values were recorded in the mountainous areas in the northwestern part of the basin (*Fig. 3a* and *b*). However, during the months analyzed in the wet season, the maximum values were recorded in western coastal areas of the basin (*Fig. 3c* and *d*).
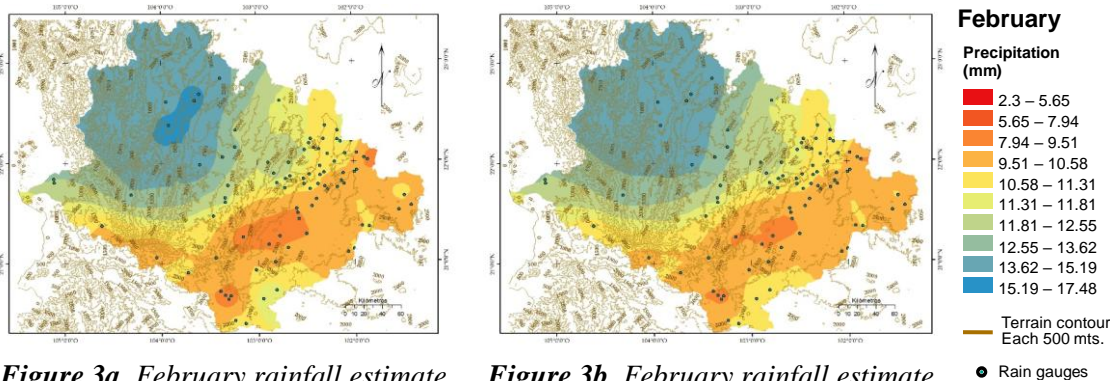

February

Precipitation (mm)

■	2.3 – 5.65
■	5.65 – 7.94
■	7.94 – 9.51
■	9.51 – 10.58
■	10.58 – 11.31
■	11.31 – 11.81
■	11.81 – 12.55
■	12.55 – 13.62
■	13.62 – 15.19
■	15.19 – 17.48

— Terrain contour Each 500 mts.

⊙ Rain gauges

Figure 3a. *February rainfall estimate with OK*   ***Figure 3b.*** *February rainfall estimate with OCK*
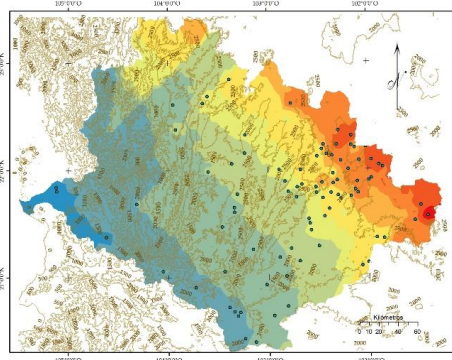
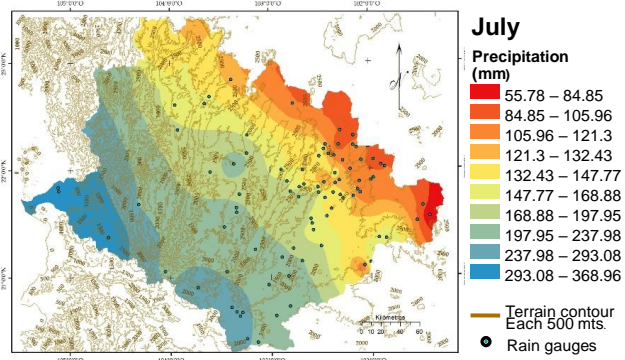**Figure 3c.** *July rainfall estimate with OK*

Figure 3d. *July rainfall estimate with OCK*

The results of the geostatistical methods are similar to those reported by other authors such as Goovaerts (2000), who showed that the contribution of secondary information in the OCK method depends not only on the correlation but also on its pattern of spatial continuity. When the nugget effect of primary information is high, the noise is higher, and therefore it contains less information while the secondary information gains more weight. The benefit of OCK over OK is increased when the spatial dependence between the observations of the primary variable is weak, as evidenced by the relatively high nugget effect in the variogram of rain that is most evident in the dry months (*Table 2*). Furthermore, the correlation is low with respect to the values of terrain elevation. The values of the obtained statistical errors (*Table 3*) show numerically better performance for OCK when compared with OK for humid months. This is because the rain–elevation correlation coefficient is favorable, and additionally, the spatial dependence of the primary variable is lower in the dry months.

Table 3. Error statistics for OK and OCK

Month	Statistic	Ordinary Kriging	Ordinary Cokriging
February	ME	0.385	0.443
	MSE	7.040	6.108
	RMSE	2.653	2.471
	SMSE	0.998	1.009
April	ME	-0.56	-0.526
	MSE	10.136	7.784
	RMSE	3.183	2.790
	SMSE	1.008	1.013
July	ME	0.757	0.027
	MSE	469.120	0.047
	RMSE	21.659	0.218
	SMSE	0.978	0.993
August	ME	1.927	0.022
	MSE	266.196	0.024
	RMSE	16.315	0.155
	SMSE	0.991	0.998

Static Kalman filter estimates

For the Kalman filter estimation method, a mesh consisting of 306 nodes spaced evenly over the entire basin was defined. This mesh, together with the 89 existing rain gauge positions, is shown in *Fig. 4*. Subsequently, the variograms for the univariate and bivariate cases were generated and the linear model of co-regionalization was computed for the bivariate case. Simple and cross-covariance matrices were generated using the variogram parameters, spatial coordinates of the sampling (S), and mesh of the estimation points (E). For the bivariate case, the data of the digital elevation model at the same positions of S and E were also used. This resulted in a square covariance matrix twice as large as in the univariate case.
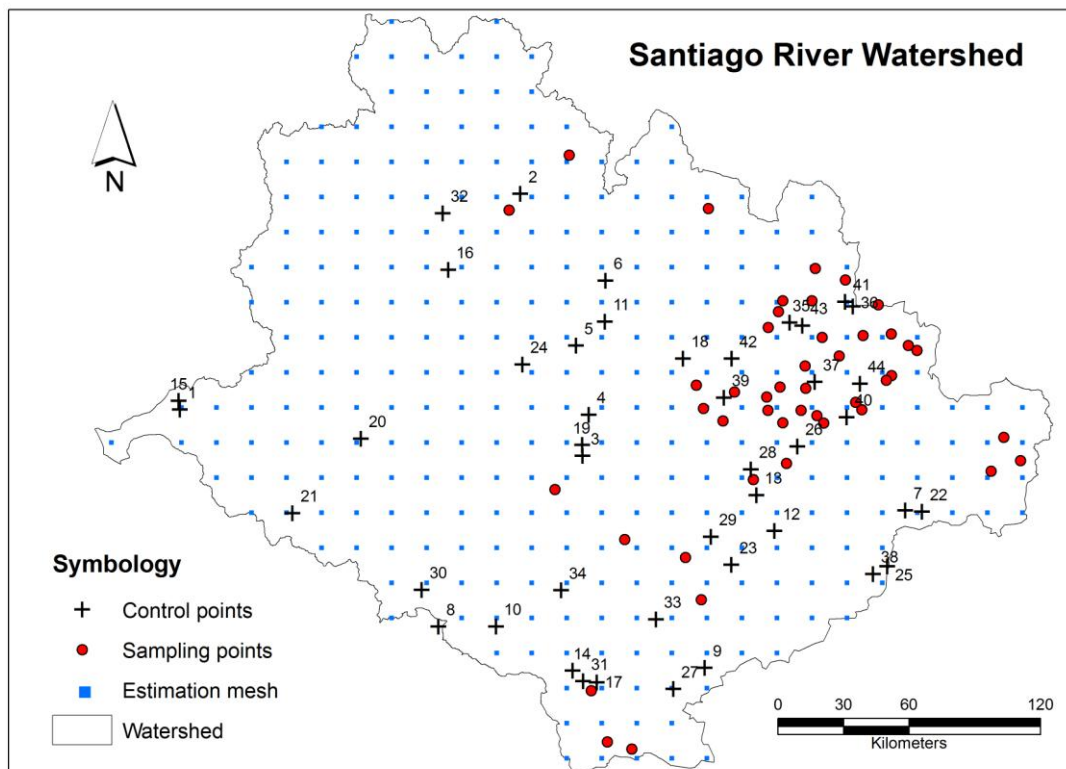

Figure 4. Mesh of estimation points and locations of existing rain gauges

The Kalman filter was applied as a programmed subroutine written in FORTRAN. This used as input data the covariance matrix (and for the bivariate case a cross-covariance matrix), average rain values at the sampled nodes, and UTM coordinates of the sampling and estimation points. As in the geostatistical case, to facilitate comparison, the months of February and April were chosen for the dry season and July and August were selected for the wet season. The results of the univariate and bivariate Kalman filter estimations were compared with the actual values recorded by 44 existing gauges which were left as control points and thus not considered in the estimation process (*Fig. 4*). The results are shown graphically in *Fig. 5a–d*.
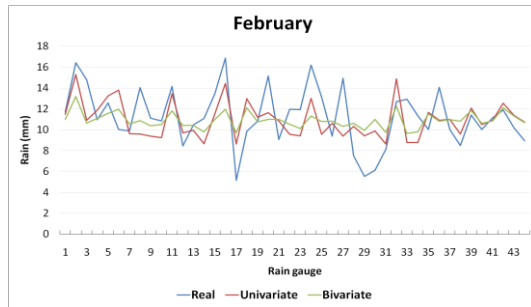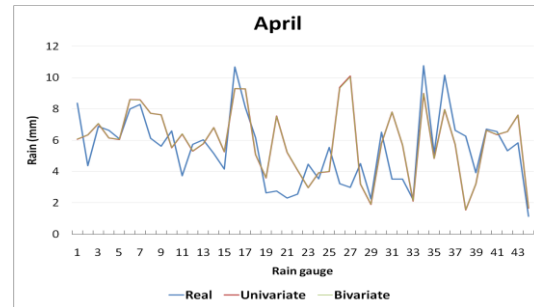
Figure 5(a). *February rainfall estimates*

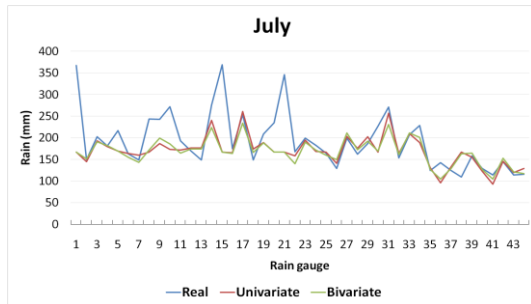***Figure 5(b).*** *April rainfall estimates*

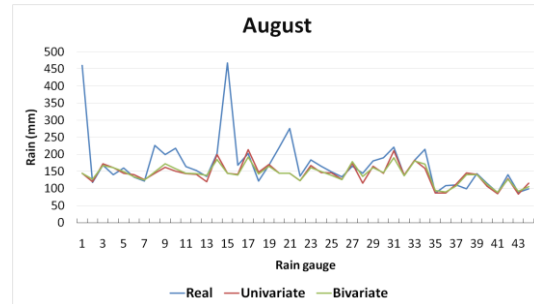***Figure 5(c).*** *July rainfall estimates*

Figure 5(d). *August rainfall estimates*

Both the univariate and the bivariate methods show good performance for the dry months, especially February (*Fig. 5a*). April presents low rainfall values with greater variability, and in this case, the bivariate and univariate estimates appear the same (*Fig. 5b*). However, for the wet months of August and July, extreme values are observed, reflecting high spatial variability in these months (*Fig. 5c* and *d*). In this case, the Kalman filter was unable to model values for the raingauges 1, 8, 11, 15 and 21 which are located near the coast and in a low density sampling area (*Fig. 4*).

Table 4 shows the values of the error statistics for the estimates obtained using the Kalman filter. The estimates of the univariate and bivariate methods were compared with the actual values recorded by the 44 rain gauges not considered in the estimation process. The results show that the bivariate estimation method with the Kalman filter outperformed the univariate method in most cases. This is particularly noticeable when considering the MSE values.

Table 4. *Error statistics for univariate and bivariate Kalman estimates*

Month	Statistic	Univariate Kalman	Bivariate Kalman
February	ME	0.364	0.434
	MSE	7.063	5.634
	RMSE	2.657	2.373
	SMSE	0.995	1.011
April	ME	−0.539	−0.538
	MSE	5.079	5.035
	RMSE	2.254	2.244
	SMSE	1.036	1.037
July	ME	−6.078	2.056
	MSE	4338.875	3046.605
	RMSE	65.870	55.196
	SMSE	0.985	0.978

	ME	2.180	4.792
August	MSE	6165.38	4592.553
	RMSE	78.520	67.768
	SMSE	0.978	0.982

The comparison of the evaluated estimation methods revealed that estimates obtained using the Kalman filter were better for the dry season months of February and April. Conversely, the estimates obtained using the Kalman filter did not improve the performance of the geostatistical methods (OK and OCK) for the wet months (*Table 3*). This is because these months presented high variability.

Estimates made with the OCK method were significantly improved with respect to the OK method, as evidenced by the good correlation obtained between rainfall and terrain elevation in the humid months.

Generally, the Kalman filter estimation was affected by the chosen initial state vector. Thus, the estimates of the univariate and bivariate methods oscillated around the average value of the data, which was selected as the initial system state. Another factor affecting the accuracy of predictions especially for wet months (*Fig. 5c* and *d*), was the low density of sampling data for positions 1, 8, 11, 15 and 21 (*Fig. 4*). The highest errors produced for the lack of surrounding data points, affects the input covariance matrix values. This in turn increases the error statistics for the entire estimate (*Table 4*).

Conclusions

In this work, the performance of geostatistical and Kalman filter estimation methods for both univariate and bivariate cases were compared. The estimates were produced for rainfall data at ungauged sites.

The ME values for the Kalman methods indicate an accuracy advantage for the bivariate Kalman method over the univariate estimation method. This is less evident for the month of April, where a high nugget effect in the cross variogram indicated a low spatial dependence. The dry season presented values of lower spatial dependence than the wet season. In the wet season case, the values of ME and MSE were more favorable for the geostatistical estimation methods.

The error was provided in the first instance by the initial covariance matrix. This was obtained from a geostatistical analysis that should be undertaken carefully for both algorithms because the estimates made using the Kalman filter exhibited no difference with respect to the geostatistical methods in terms of sensitivity to the a priori covariance matrix.

In some cases, when the spatial dependence between data is low, as reflected by a higher relative nugget effect at the variogram of the primary variable, it could be possible to obtain better results with the Kalman filter, as long as there is normality in the sample and the semivariogram adjustment is optimum.

For the estimates made using the geostatistical methods, the correlation between rainfall and terrain elevation, as well as the spatial dependence of the primary variable, played a decisive role. This was particularly true for the OCK estimates, which significantly improved the estimations made using OK for the rainy season.

REFERENCES

[1] Awadallah, A. G. (2012): Selecting Optimum Locations of Rainfall Stations Using Kriging and Entropy. - International Journal of Civil & Environmental Engineering 12: 36-41.

[2] Béal, D., Guillot, G., Courault, D., Bruchou, C. (2002): Interpolation of rainfall at small scale in a Mediterranean region: relation between rainfall and altitude within a simple space-time model. - Geostatistics for environmental applications 13: 379-389, Kluwer Academic Publishers.

[3] Borga, M., Vizzacaro, A. (1997): On interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. - Journal of Hydrology 195 (1-4):160-171.

[4] Bostan, P.A., Akyürek, Z. (2009): Spatio-temporal analysis of precipitation and temperature distribution over Turkey. - The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38: 92-97.

[5] Briseño, J.V. (2012): Método para la calibración de los modelos estocásticos de flujo y transporte en aguas subterráneas para el diseño de redes de monitoreo. - Tesis Doctoral. Recuperado de: http://www.ciencianueva.unam.mx/repositorio/handle/123456789/50

[6] Briseño, J. V., Herrera, G. S., Júnez H. E. (2011): Método para el diseño óptimo de redes de monitoreo de los niveles del agua subterránea. - Tecnología y Ciencias del Agua 2(4): 77-96.

[7] CONAGUA (2015a): Programa de medidas preventivas y de mitigación de la sequía. Consejo de Cuenca Río Santiago. 1a Versión. Recuperado de: pronacose.gob.mx/pronacose14/contenido/documentos/Río%20Santiago.pdf

[8] CONAGUA (2015b): Formulación del Programa Hídrico por Organismo de Cuenca, Visión 2030, de la Región Hidrológico-Administrativa VIII, Lerma-Santiago-Pacifico. INTEGRACIÓN DEL PROGRAMA HÍDRICO REGIONAL. Recuperado de: http://www.conagua.gob.mx/OCLSP07/NotaP/FINALaaResumenEjecutivo.pdf

[9] Costa, A. C., Soares, A. (2009): Homogenization of Climate Data: Review and New Perspectives Using Geostatistics. - International Association for Mathematical Geosciences 41: 291–305

[10] Dirks, K. N., Hay, J. E., Stow, C. D., Harris, D. (1998): High resolution studies of rainfall on Norfolk Island: Part II: Interpolation of rainfall data, - J. Hydrol. 208(3–4): 187–193

[11] Díaz, M. A. (2002): Geoestadística Aplicada, Instituto de Geofísica, UNAM, Instituto de Geofísica y Astronomía, CITMA, Cuba.

[12] Díaz. M., Herrera, G., Valdés, A. (2002): Estimación De La Precipitación En El Valle De México Usando Datos De Pluviómetros Y Radar Meteorológico. III Reunión Nacional de Ciencias de la Tierra, Puerto Vallarta México, Recuperado de: http://mmc2.geofisica.unam.mx/cursos/geoest/Presentaciones/Ejemplo Precipitacion.pdf

[13] Giraldo, H. R. (2001): Propuesta de un indicador como variable auxiliar en el analisis cokriging. - Revista Colombiana de Estadistica. V., 24(1): 1-12

[14] Gómez-Balandra, M. A., Díaz-Pardo, E. , Gutiérrez-Hernández, A. (2012): Composición de la comunidad íctica de la Cuenca del Río Santiago durante su desarrollohidráulico. - Hidrobiológica 22(1): 62-78.

[15] Goovaerts, P. (2000): Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. - Journal of Hydrology 228: 113–129

[16] McMillan, H. K., Hreinsson, E. O., Clark, M. P., Singh, S. K., Zammit, C., Uddstrom, M. J. (2013): Operational hydrological data assimilation with the recursive ensemble Kalman filter. - Hydrol. Earth Syst. Sci. 17: 21–38

[17] Herrera, G. (1998): Cost Effective Groundwater Quality Sampling Network Design, Ph. D. Dissertation, University of Vermont, Burlington, Vermont. Recuperado de: http://mmc2.igeofcu.unam.mx/cursos/gest/Articulos/Geostatistics/Cost Effective Groundwater Quality Sampling Network Design.pdf

[18] Hevesi, J. A., Istok, J. D., Flint, A. L. (1991): Precipitation estimation in mountainous using multivariate geostatistics. Part II: Isohyetal Maps. - J. Appl. Meteor., 31: 677–688.

[19] Holawe, F., Dutter, R. (1999): Geostatistical study of precipitation series in Austria: time and space. - Journal of Hydrology 219: 70–82

[20] Huang, B., Hu, T. (2008): Spatial interpolation of rainfall based on DEM. Advances in Water Resources and Hydraulic Engineering, - Proceedings of 16th IAHR-APD Congress and 3rd Symposium of IAHR-ISHS, 1: 77-81.

[21] Isaaks, E.H., Srivastava, R.M. (1989): An Introduction to Applied Geostatistics. - Oxford University Press, New York, p. 561

[22] Júnez, H. E. (2005): Diseño De Una Red De Monitoreo De La Calidad Del Agua Para El Acuífero Irapuato-Valle, Guanajuato. - Tesis De Grado De Maestría. Universidad Nacional Autónoma De México. Facultad De Ingeniería. DEPFI (Campus Morelos):

[23] Júnez, H.E., Herrera, G.S. (2013): A geostatistical methodology for the optimal design of space–time hydraulic head monitoring networks and its application to the Valle de Querétaro aquifer. - Environmental Monitoring and Assessment 185(4): 3527-3549.

[24] Komma, J., Bloschl, G., Reszler, C. (2008): Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. - Journal of Hydrology 357: 228– 242

[25] Majani, B.S. (2007): Analysis of External Drift Kriging Algorithm with application to precipitation estimation in complex orography. - International Institute For Geo-Information Science And Earth Observation Enschede, The Netherlands. Thesis. Recuperado de: http://www.itc.nl

[26] Moradkhani, H., S., Gupta, H., Houser, P. (2005): Dual State-Parameter Estimation of Hydrological Models using Ensemble Kalman Filter. - Advances in Water Resources 28(2): 135-147.

[27] Moral, F. J. (2009): Comparison of different geostatistical approaches to map climate variables: application to precipitation. - International journal of climatology 30: 620–631.

[28] Morales, M.I., Aparicio, J., Valdés, J. B. (2014): Pronóstico de avenidas utilizando el filtro de Kalman discreto. - Tecnología y Ciencias del Agua 5(2): 85-110

[29] Murthy, K., Abbaiah, G. (2007): Geostatistical analysis for estimation of mean rainfalls in Andhra Pradesh, India. - International Journal of Geology 3: 35-55.

[30] Pardo, E. (1998): Comparison Of Geostatistical Methods For Estimating The Areal Average Climatological Rainfall Mean Using Data On Precipitation And Topography. - International Journal of Climatology 18: 1031–1047

[31] Shaghaghian, M.R., Abedini, M. J. (2012): Rain gauge network design using coupled geostatistical and multivariate techniques. - Transactions A: Civil Engineering, Scientia Iranica A. 20 (2): 259–269

[32] Soenario, I., Plieger, M., Sluiter, R. (2010): Optimization of Rainfall Interpolation - Intern rapport. De Bilt Netherlands: Koninklijk Nederlands Meteorologisch Instituut (KNMI)

[33] Srikanthan, R., Amirthananthan, G., Kuczera, G. (2015): Real-time Flood Forecasting Using Ensemble Kalman Filter. Recuperado de: http://www.mssanz.org.au

[34] Subyani, A. M. (2015): On the Space-Time Variogram Models-An Application to Seasonal Precipitation. http://www.kgs.ku.edu/

[35] Urquiola, A., Infante, S., Cedeño, F. (2015): Modelo Dinámico Espacio Temporal para Predecir Niveles de Precipitaciones en Venezuela. - Departamento de Matemáticas, FaCyT, Universidad de Carabobo, Valencia, Venezuela. Recuperado de: http://www.researchgate.net/publication/236214317_Modelo_Dinmico_Espacio_Tempor al_para_Predecir_Niveles_de_Precipitaciones_en_Venezuela

[36] Valdés, A., Herrera, G., Díaz, M. (2015): Avances En El Monitoreo De La Precipitación En La Cuenca De México Usando Radar Meteorológico Y Pluviómetros. Recuperado de: http://www.inecc.gob.mx/cuencas-eventos/985-cuencas-con-nal2

[37] Volkmann, T. H. M., Lyon, S. W., Gupta, H. V., Troch, P. A. (2010): Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain. - Water Resources Research 46: 1-16.

[38] Waylen, P. R., Quesada, M. E., Caviedes, C. N. (1995): Temporal and spatial variability of annual precipitation in Costa Rica and the southern oscillation. - International Journal of climatology 16: 173-193.

[39] Weerts, A. H., El Serafy, G. Y. H. (2006): Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. - Water Resour. Res., 42, W09403, doi:10.1029/2005WR004093