Journal of Urban and Environmental Engineering, v.12, n.2, p.219-230

ISSN 1982-3932
doi: 10.4090/juee.2018.v12n2.219230

Journal of Urban and Environmental Engineering

www.journal-uee.org

STOCHASTIC MODELING OF MONTHLY RIVER FLOWS BY SELF-ORGANIZING MAPS

José Adalberto da Silva Filho[1*] and Camilo Allyson Simões de Farias[2]

[1]Department of Rural Technology, Rural Federal University of Pernambuco, Recife, Pernambuco, Brazil
[2]Federal University of Campina Grande, Campina Grande, Paraíba, Brazil

Received 20 February 2018; received in revised form 16 October 2018; accepted 18 October 2018

Abstract: Extreme hydrological conditions and increasing water demands observed in semiarid Brazil have generated conflicts regarding to the best use of existing water resources. Synthetic generation models of river flows are often used as support for the definition of water system operating rules, which allow the establishment of rationing policies before water scarcity spells. This work aims at verifying the applicability of models based on self-organizing maps (SOM) for stochastic modeling of monthly river flows. The basic principle of the study consisted of using SOM models in order to define the deterministic component of river flow series and a density probability function (stochastic component) to represent the resulting residuals. During calibration of all networks, values of *NASH* were above 0.9989 for the applications. The results were promising, indicating that the established models are capable of producing synthetic series of inflows with excellent performance.

Keywords: Artificial neural networks; method of fragments; water resources

* Correspondence to: José Adalberto da Silva Filho, Tel.: +55 81 3631 0221; E-mail: adalbertosilva15@gmail.com

INTRODUCTION

Water scarcity plays a major role in the social and economic development of any region. This factor, associated with growing demands, has allowed the emergence of conflicts among the various users and water resources sectors (Carneiro & Farias, 2013). The proper management of water supply systems, in addition to promoting sustainability, tends to reduce the problem of water crisis.

From planning to completion of projects in engineering, it is necessary to follow a set of technical and operational requirements, which take into account: time, cost, safety, scope, and other variables that influence in its quality (Gomes *et al.*, 2017).

Concerning the safety, for example, the availability of long data records of river flows allows a better management of works related to dam construction, bridges or any hydraulic system. Given this requirement, many localities do not have long and regular measurements of river flows, a condition that could jeopardize the integrity of projects and increase the risk of failure.

The use of modeling processes aims at extending river flow series with excellent performance. Models that are capable of producing new flow rate data, preserving the statistical properties observed in the time series, constitute an important mechanism for engineering. Associated with this requirement, the availability of flow data enables the best planning of the resource in question, since water bodies admit multiplicity of their uses.

In this context, stochastic optimization techniques are important tools for the definition of operational rules of water resources systems. Such techniques try to incorporate hydrological uncertainties and, thereby, support water planning and management (Loucks *et al.*, 1981).

Stochastic methods, compared to deterministic and empirical ones, stand out in this scenario, since they allow a considerable increase of information, which is usually necessary for modeling. These approaches admit that the runoffs follow probabilistic laws of formation, generating synthetic series of flows based on statistical parameters such as mean, standard deviation, skewness, correlation etc. (Farias, 2003). These models usually support in the definition of operating policies, allowing the establishment of rationing rules before users face an actual water crisis.

In recent decades, the appearance of artificial intelligence techniques, such as artificial neural networks (ANN), genetic algorithms and fuzzy logic, favored the development of promising models (Emch & Yeh, 1998; Pulido-Velazquez *et al.*, 2006; Farias, 2009; Chang *et al.*, 2010; Farias *et al.*, 2011; Celeste & Billib, 2012; Kumar *et al.*, 2013).

Self-Organizing Maps (SOM) are unsupervised artificial neural networks that group input data into classes according to their similarities, through competitive training methods (Kohonen, 1982; Haykin, 1999; Silva *et al.*, 2010). The unsupervised process refers to the capacity to learn and cluster information without providing an error indication to estimate the potential solution (Sathya & Abraham, 2013). The simple structure, as well as the dynamics of different trainings, makes the self-organizing maps a promising tool.

Tuevo Kohonen proposed the SOM networks for the first time in 1982 (Kohonen, 1982) and, since then, they have been predominantly applied to data grouping and classification. In addition to the ability to modeling nonlinear relationships, SOM networks are able to reduce a set of multidimensional data into a two-dimensional array that can be used for analysis and prediction.

The lack of indication for the learning algorithm in unsupervised learning has its advantages, since it facilitates and allows the algorithm to look back for similarities that have not been previously considered (Kohonen & Simula, 1996; Farzad & El-Shafie, 2017). In order to find a method for stochastic modeling of monthly river flows, self-organizing maps networks must be investigated as tools that might have the potential to overcome traditional methods of hydrological modeling.

Works on the development and application of SOM in the area of water resources are extremely scarce and even non-existent when it comes to some specific areas. García & Gonzáles (2004), Adeloye *et al.* (2011), Farias & Santos (2014) and Farias *et al.* (2015) are examples of few studies on the applicability of SOM networks in the area of water resources.

Considering the lack of studies based on SOM networks for stochastic modeling of monthly river flows, we compared the performance of different structures of SOM with the purpose of generating new synthetic data. We also compared the results with those obtained by the method of fragments, a traditional process widely used for synthetic generation of river flows.

MATERIALS AND METHOD

Study area

The two data sets used in this study are from *Piancó* and *Emas* stream gauge stations, which were obtained from the database of the Brazilian National Water Agency (ANA, 2016). Both stations are located in *Piancó* River Basin, in a semiarid land of *Paraíba* State, Brazil.

This region is characterized by a semiarid climate, with annual mean rainfall, drainage area, length of Piancó River and annual mean temperature equal to 821 mm, 9.228 km², 208 km and 24°C, respectively. The annual potential evaporation is 2993,4 mm and the natural vegetation is of the xerophytic type, belonging to a biome known in Brazil as *Caatinga* (Scientec, 1997; Lima, 2004; Rodrigues *et al*., 2007; FARIAS *et al*., 2013). Details of the studied stations and their geographic locations are shown in **Fig. 1**.

Data

For calibration of the models in *Piancó* stream gauge station, we used monthly flows from 1965 to 2012. In order to avoid modeling problems, we decided to exclude years with missing data, which were 1969, 1974, 1975, 1983, 1985 1988, 1989, 1990, 1991, 1992, 1995, 1998, 2002, 2005, 2008 and 2009. This resulted in a 32-year long data sample for *Piancó* stream.

As for *Emas* stream gauge station, we used monthly flows between 1964 and 2012. We exclude the incomplete years of 1966, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1993, 2002, 2003, 2004 and 2008.As a result, we used a series of monthly flows with 31 years.

The models and procedures in this study were developed so as to generate 5.000 year-long synthetic scenarios of monthly flows. All models were implemented in Matlab programming language, R2012a version.

We adopted the Gamma probability distribution function for the adjustment of residual series, as it was found to be the best suited for the region's hydrology (Celeste *et al*., 2007).

Method of fragments

The method of fragments (MF) was first proposed by Svanidze (1980) and it has been widely used in literature for synthetic generation of inflows (Celeste *et al*., 2007; Carneiro & Farias, 2013; Silva Filho *et al*., 2015). The MF is a disaggregation model of annual flows into monthly flows. Thus, the first step is to calculate the fragments by dividing the flow rate of each month by the sum of all monthly flows of the current year, as shown in **Eq. (1)**:

$$f(y,m) = \frac{Q(y,m)}{\sum_{m=1}^{12} Q(y,m)} \tag{1}$$

in which $f(y,m)$ and $Q(y,m)$ are, respectively, the fragment and the river flow observed in month m of year y.

According to Carneiro & Farias (2013), the next step is to evaluate the linear dependence among historical annual flows. If they are serially dependent, a statistical model should be used to produce a number of independent residues. After this step, the residues of the historical series of annual flows are modeled by a suitable probability density function (PDF).

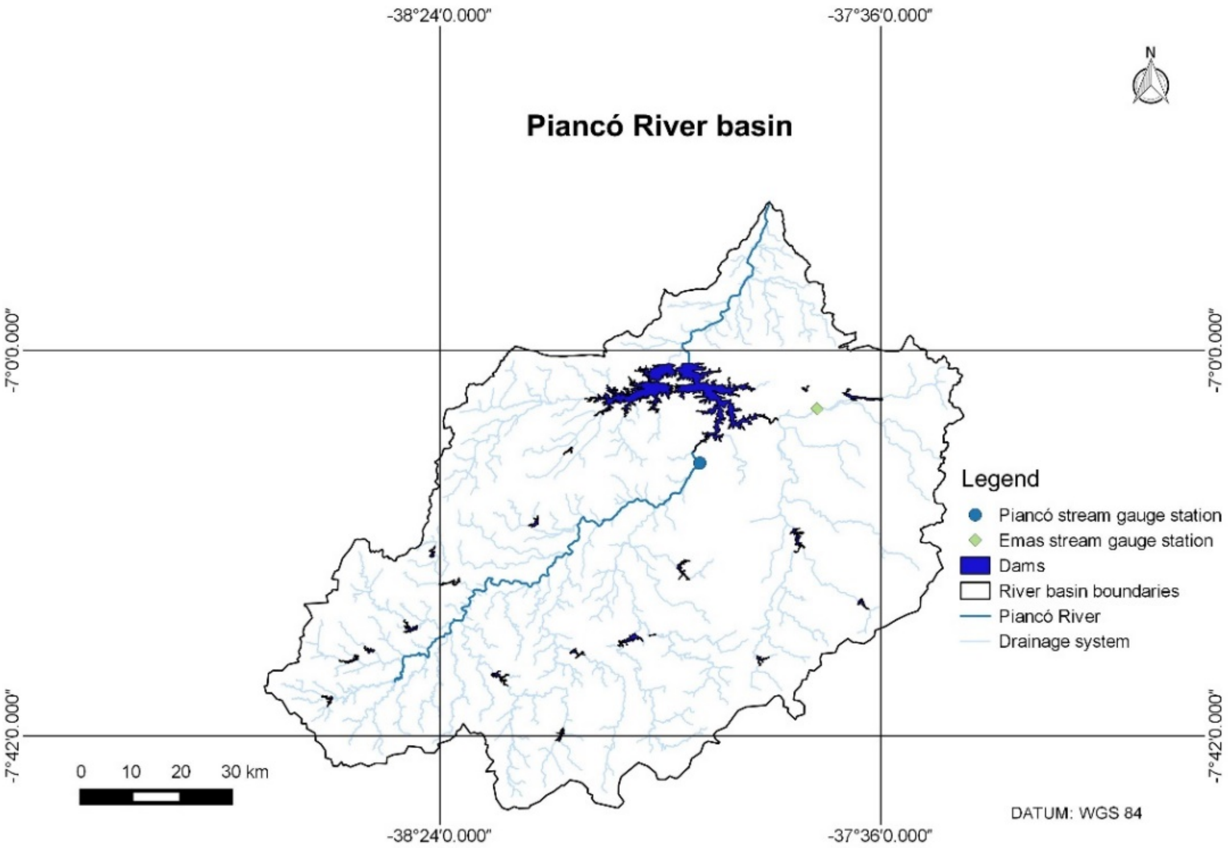

Fig. 1 Location of the stream gauge stations in Piancó River basin, Brazil.

The new stochastic generation of annual flows is carried out by a random simulation of numbers, modeled by the PDF, and subsequent application of inverse functions of the statistical model responsible for the withdrawal of linear dependencies (Carneiro & Farias, 2013). Then, the generated annual flows are disaggregated by following the procedures described in Celeste *et al*. (2007):

(a) annual flows of historical data are arranged in ascending order for establishment of classes;
(b) the first class has zero as inferior limit and the last class has no superior limit, i.e., it has a limit equal to infinity;
(c) intermediate classes are defined by the means between two successive flows;
(d) after the definition of such limits, each generated annual flow will belong to a corresponding class and the monthly flows are obtained by the product of the fragments of that class and synthetic annual value.

Self-Organizing Maps

The modeling using SOM consists of representing multidimensional input vectors by means of one-dimensional or two-dimensional maps, keeping the neighborhood relationship of data (SILVA *et al*., 2010).

In this work, the vectors of the input layer have thirteen components: the twelve fragments derived from each month of the year, obtained by **Eq. (1)**, and the annual flow *Vol* (m³/s), as can be seen in **Fig. 2**.

In the output layer were used hexagonal neurons, whose weights also have thirteen components, which is related to the size of the input vectors. The formulated architectures were based on the number of neurons (*M*), as proposed by García & González (2004), using the heuristic shown in **Eq. (2)**.

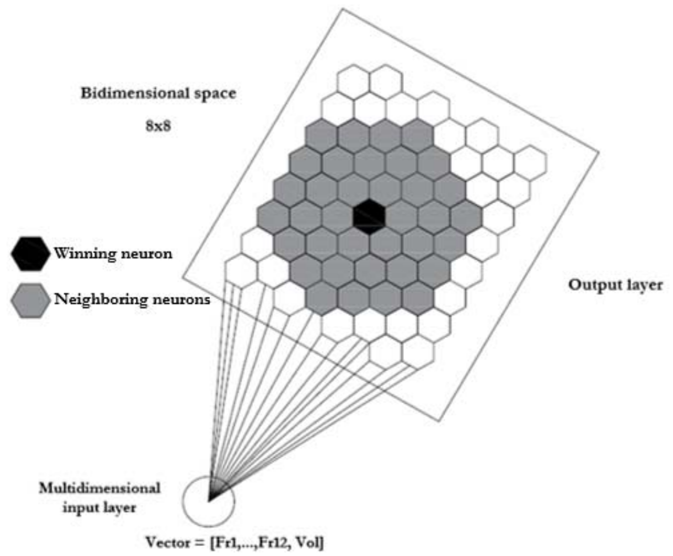

Fig. 2 Structure of SOM model and example with a winner neuron and its neighbors.
Source: Adapted from Farias et al. (2013)

$$M = 5\sqrt{N} \tag{2}$$

in which N is the total number of samples in the calibration data set.

According to Farias *et al*. (2014), in output layer, also known as competitive layer, neurons compete with each other and only one of them is considered to be the winner or the most suitable class for a given input vector **x**. In such networks, each input vector element is connected to all elements of the output layer. The strength of the connection is measured by means of weights w_{ij} between input neurons *j* and output neurons *i* (Beale *et al*., 2012).

During the training of SOM model, Euclidean distances (DIi) between input vectors and weights attached to each one of output neurons are calculated as shown in **Eq. (3)**:

$$DIi = \sqrt{\sum_{j=1}^{J} (xj - wij)^2}\,;\ to\ i = 1,2,\ldots,M. \tag{3}$$

in which *xj* is the *j-th* component of input vector **x**; *J* is the dimension of input vector **x**; and *M* is the total number of neurons in the output layer.

According to Beale *et al*. (2012), the output neuron *i* that has the smallest Euclidean distance when compared to the input vector is the winner. The weights linked to this winner neuron $i*$ and neurons within a certain neighborhood radius $Vi*$ are then updated by the Kohonen rule (Beale *et al*., 2012), as presented in **Eq. (4)**:

$$\begin{gathered} w_{ij}(n) = w_{ij}\,(n-1) + \alpha\,.\left[x_j\,(n) - w_{ij}\,(n-1)\right]; \\ to\ i\ \in V_{i^*}\ and\ j = 1,2,\ldots,J \end{gathered} \tag{4}$$

in which α is the learning rate, and n is an index that represents the sequence of samples presented to the model.

The Kohonen rule forces weights attached to the winner neuron and its neighbors to move towards the input vector presented to the network, causing the Euclidean distance to become smaller. Thus, these neurons learn how to classify similar vectors (Farias *et al*., 2013).

The presentation of input vectors to the network can also be carried out by using the entire data set before any weight update. This form of presentation is known as batch mode. In this case, the search for winner neurons is performed for each input vector and, then, the weight vector is moved to a specific position calculated by the average of input vectors for which the neuron was the winner or winner's neighbor (Beale *et al*., 2012; Farias *et al*., 2013; Farias & Santos, 2014).

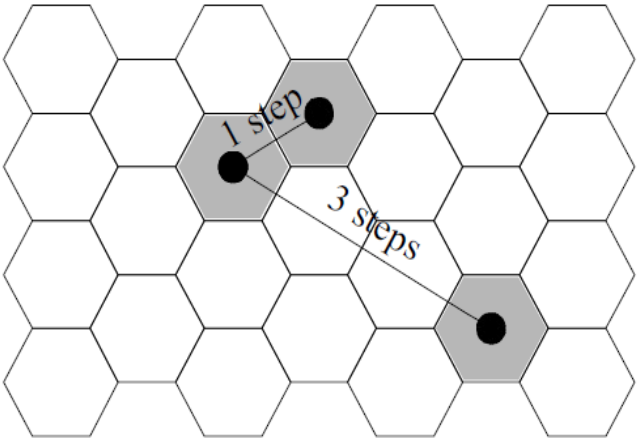

Fig. 3 Distances between neurons of a SOM model for the determination of the neighborhood.
Source: Farias et al. (2013).

After several presentations of the data set, weights tend to stabilize (Farias *et al*., 2015). It is important to emphasize that the training of this network is unsupervised, because there are no desired outputs. **Figure 3** shows how distance between hexagonal neurons are obtained for neighborhood definition.

The network training of this study takes place in two phases: ordering and tuning phases. In first phase, we decided to use 100 presentations of the data set, with the radius of initial neighborhood equal to three steps and final value equal to one. In second phase, the neighborhood radius is below the unit. Thus, there is only update on the weight of the winning neuron. For this phase, we used 900 presentations (Farias *et al*., 2013; Farias & Santos, 2014; Farias *et al*., 2014). The learning rates used were equal to 0,90 and 0,02, in the ordering and tuning phases, respectively. Details about these phases can be seen in Beale *et al*. (2012).

To increase the chances of finding global optimum during calibration of the models, the networks were trained 10 times, choosing the one with the highest efficiency of Nash-Sutcliffe between synthetic and observed river flows (Nash & Sutcliffe, 1970). The number of trainings is among the most scientifically used (Farias *et al*., 2015; Wang *et al*., 2016; Eyduran *et al*., 2017).

Once calibrated, the SOM networks can be used to determine the fragments considering a given synthetic annual flow. Thus, the new scenarios have been generated in accordance with the procedures described by Farias *et al*. (2014); Farias & Santos (2014) and Farias *et al*. (2015):

(a) calculate the Euclidean distances between the input vector and all output neurons of the SOM model, considering the fragment components as non-existent;
(b) determine the winner neuron based on the smallest Euclidean distance;
(c) obtain synthetic monthly flows by multiplying synthetic annual flow and fragments of the winner neuron.

Evaluation of the SOM structures

The performance of hydrological models is commonly measured by statistical indices such as correlation (*r*), relative bias (*RB*) and Nash efficiency (*NASH*) (Nash & Sutcliffe, 1970). Their mathematical equations are described below.

The correlation index refers to the degree of linear dependence between generated scenarios and historical values of flows, expressing a potential good fit of the estimation. The relative bias has the ability to determine if the prediction model tends to underestimate or overestimate the observed flow rates. The *NASH* efficiency, which ranges from -∞ to 1, is traditionally used to express adherence between observed and synthetic flows. This index considers the systematic and random errors, indicating that the adjustment is better as its value approaches one (Asce, 1993; Farias *et al*., 2012).

In order to check the efficiency of SOM and MF models, we used relevant statistical parameters to analyze the generated synthetic series of river flows, including a measure of location (mean) and estimates of second (standard deviation), third (skewness) and fourth (kurtosis) statistical moments.

$$r = \frac{N\sum_{t=1}^{N} Qobs(t)Qcal(t) - \sum_{t=1}^{N} Qobs(t).\sum_{t=1}^{N} Qcal(t)}{\sqrt{N(\sum_{t=1}^{N} Qobs^2(t)) - (\sum_{t=1}^{N} Qobs(t))^2}\sqrt{N(\sum_{t=1}^{N} Qcal^2(t)) - (\sum_{t=1}^{N} Qcal(t))^2}} \tag{5}$$

$$RB = \frac{\overline{Qcal} - \overline{Qobs}}{\overline{Qobs}}\,.100\% \tag{6}$$

$$NASH = 1 - \frac{\sum_{t=1}^{N}(Qobs(t) - Qcal(t))^2}{\sum_{t=1}^{N}(Qobs(t) - \overline{Qobs})^2} \tag{7}$$

in which *Qobs* (*t*) is the flow rate observed at time *t*; *Qcal* (*t*) is the flow rate calculated at time *t*; $\overline{Qobs}$ is the mean of observed flow rates; $\overline{Qcal}$ is the mean of calculated flow rates; and *N* is the size of data.

The comparison consisted of verifying relationships between calculated and observed data. No one expects the synthetic series to be identical to those of observed series. More than that, we desire a model that produce different scenarios in which statistical moments show similarity and proximity with those observed in historical series. Depending on these properties, we can observe a weaker or a stronger stationarity (Wilks, 2006).

RESULTS AND DISCUSSION

In order to determine the SOM architectures, we used networks that resulted in an arrangement of 30 neurons, as proposed by García & González (2004). We also used networks with 60 and 90 neurons for comparison purposes. The size of historical series of both study cases are very similar (31 and 32 year-long series for *Piancó* and *Emas*, respectively). As a result, the proposed structures for both applications were the same.

In order to evaluate the performance of the proposed models, we compared monthly statistical moments (mean, standard deviation, skewness and kurtosis) from January to December of the synthetic scenarios with those observed in historical series. Such results can be seen in **Tables 1** and **2**.

Table 1. Comparison of monthly statistics of synthetic series generated by SOM and MF models with those from historical data in *Piancó* stream gauge station.

Model/Structure		Correlation				Relative Bias				*Nash*			
		Mean	Standard deviation	Skewness	Kurtosis	Mean %	Standard deviation	Skewness %	Kurtosis %	Mean	Standard deviation	Skewness	Kurtosis
MF		0.9912	0.9948	0.9868	0.9813	1.40	2.68%	18.18	50.86	0.9801	0.9852	0.7324	0.3244
SOM #1	1×30	0.9974	0.9921	0.9701	0.9654	1.40	0.62%	18.02	55.24	0.9945	0.9834	0.6429	0.2009
SOM #2	2×15	0.9969	0.9902	0.9742	0.9794	1.40	4.61%	18.16	60.35	0.9936	0.9788	0.6306	0.1827
SOM #3	3×10	0.9948	0.9768	0.9648	0.9554	1.40	4.51%	6.33	34.50	0.9895	0.9525	0.8113	0.4668
SOM #4	5×6	0.9929	0.9952	0.9609	0.9513	1.40	2.22%	16.29	49.39	0.9843	0.9879	0.6553	0.1652
SOM #5	2×30	0.9921	0.9928	0.9611	0.9510	1.40	0.31%	20.79	60.56	0.9827	0.9838	0.5917	0.1477
SOM #6	**4×15**	**0.9919**	**0.9720**	**0.9700**	**0.9622**	**1.40**	**6.19%**	**6.85**	**34.76**	**0.9836**	**0.9419**	**0.8349**	**0.5231**
SOM #7	6×10	0.9960	0.9944	0.9315	0.9324	1.40	0.07%	19.75	57.79	0.9914	0.9883	0.5250	0.0846
SOM #8	3×30	0.9979	0.9937	0.9818	0.9792	1.40	2.57%	17.43	57.51	0.9954	0.9869	0.6417	0.1207
SOM #9	6×15	0.9900	0.9877	0.9668	0.9545	1.40	0.78%	17.06	50.22	0.9799	0.9743	0.7278	0.3091
SOM #10	9×10	0.9944	0.9902	0.9679	0.9594	1.40	3.96%	16.49	56.28	0.9886	0.9793	0.6275	0.1068

Table 2. Comparison of monthly statistics of synthetic series generated by SOM and MF models with those from historical data in *Emas* stream gauge station.

Model / Structure		Correlation				Relative Bias				*Nash*			
		Mean	Standard deviation	Skewness	Kurtosis	Mean %	Standard deviation	Skewness %	Kurtosis %	Mean	Standard deviation	Skewness	Kurtosis
Method of fragments		0.9978	0.9702	0.9142	0.9342	0.77	3.90%	4.94	15.49	0.9952	0.9393	0.7758	0.6778
SOM #1	1×30	0.9974	0.9973	0.9211	0.9238	0.77	14.78%	7.52	0.72	0.9946	0.9692	0.8239	0.8450
SOM #2	2×15	0.9994	0.9976	0.9401	0.9481	0.77	17.32%	14.94	4.80	0.9983	0.9576	0.7347	0.8201
SOM #3	3×10	0.9990	0.9930	0.9252	0.9191	0.77	19.52%	12.64	3.00	0.9972	0.9457	0.7652	0.7550
SOM #4	5×6	0.9967	0.9975	0.9676	0.9748	0.77	16.50%	6.37	2.75	0.9890	0.9604	0.9142	0.9244
SOM #5	2×30	0.9993	0.9970	0.9602	0.9585	0.77	20.74%	2.18	16.73	0.9983	0.9406	0.9158	0.8599
SOM #6	**4×15**	0.9965	0.9550	0.9279	0.9351	0.77	10.69%	8.09	6.60	0.9929	0.9004	0.6840	0.5541
SOM #7	6×10	0.9901	0.9272	0.9445	0.9041	0.77	13.53%	11.71	4.06	0.9803	0.8472	0.7382	0.4626
SOM #8	3×30	0.9758	0.9490	0.8497	0.8470	0.77	14.64%	10.81	5.14	0.9520	0.8882	0.6597	0.6412
SOM #9	6×15	**0.9992**	**0.9974**	**0.9577**	**0.9750**	**0.77**	**15.92%**	**1.33**	**5.51**	**0.9965**	**0.9651**	**0.9146**	**0.9432**
SOM #10	9×10	0.9984	0.9895	0.9260	0.8819	0.77	20.55%	2.67	20.70	0.9963	0.9344	0.8543	0.6407

For hydrological modeling, values of *NASH* equal to or higher than 0.75 represent accurate models. When its value is between 0.36 and 0.75, we may say that it is acceptable (Collischonn, 2001). According to Farias *et al*. (2012), only high correlation values do not mean high precision forecasts, they must be associated with values of relative bias close to zero. In view of that, the best model was the one that provided values of *NASH*, *r*, and *RB* nearest to 1, 1 and 0, respectively.

During calibration of all networks, we obtained values of *NASH* between 0.9989 and 0.9999 for the application in Piancó stream gauge station, and between 0.9993 and 0.9999 in Emas stream gauge station.

From this perspective, we may say that the results produced by the SOM models were excellent for synthetic generation of monthly flows considering both study cases. According to **Table 1**, in *Piancó* stream gauge station, all networks have produced data with means and standard deviations similar to the monthly statistics observed in historical data, with comparative values of Nash higher than 0.94. The Nash coefficient for the comparison of skewness was considered to be accurate in SOM #3 and #6 networks and acceptable for the others. As for kurtosis, there were also acceptable values of Nash in SOM networks #3 and #6, being the later slightly higher. Thus, we assumed that the SOM #6 model performed better when compared to others structures.

In **Table 2**, we found that the results were also promising for the application in *Emas* stream gauge station. The *Nash* coefficients were accurate for monthly means and standard deviations comparisons, being higher than 0.84. As for skewness, all structures were found to present very good results, with the exception of SOM networks #2, #6, #7 and #8, which presented only acceptable values. Concerning the *NASH* coefficient in kurtosis comparison, we observed accurate values of *Nash* in SOM #1, #2, #3, #4, #5 and #9, being considered acceptable for other structures.

The variability observed among the proposed structures, mainly regarded to skewness and kurtosis, occurred primarily due to the difficulty of the models in stochastically reproducing third and fourth moments. Even so, the developed procedures using self-organizing maps proved to be promising and superior to the MF.

Most of the manuscripts only explore the first two statistical moments (Carneiro & Farias, 2013; Kasiviswanathan & Sudheer, 2013; Chandwani *et al*., 2015), but some researches reinforce the importance of the others (Parmar & Bhardwaj, 2015; Silva Filho *et al*., 2015). According to Wilks (2006), the new series used for analysis and/or prediction purposes should preserve the previously mentioned moments.

In general, the values of relative bias and correlation obtained for the monthly comparisons of all models, in both applications, were close to zero and one, respectively. This indicates that the models performed well. The most distant values expected for these two parameters were found in SOM networks #2 and #5 for *Piancó* stream gauge station; and in the SOM networks #5 and #10 for *Emas* stream gauge station.

Dralle *et al*. (2017) also used the NASH coefficient to evaluate a stochastic model of interannual variation of flow volumes. They obtained values higher than 0.9 for their applications and highlighted this parameter as an excellent performance indicator.

Figures 4 and **5** illustrate, considering both gauge stations, a comparison among the monthly statistics between historical and synthetic data generated by the best SOM structures and MF.

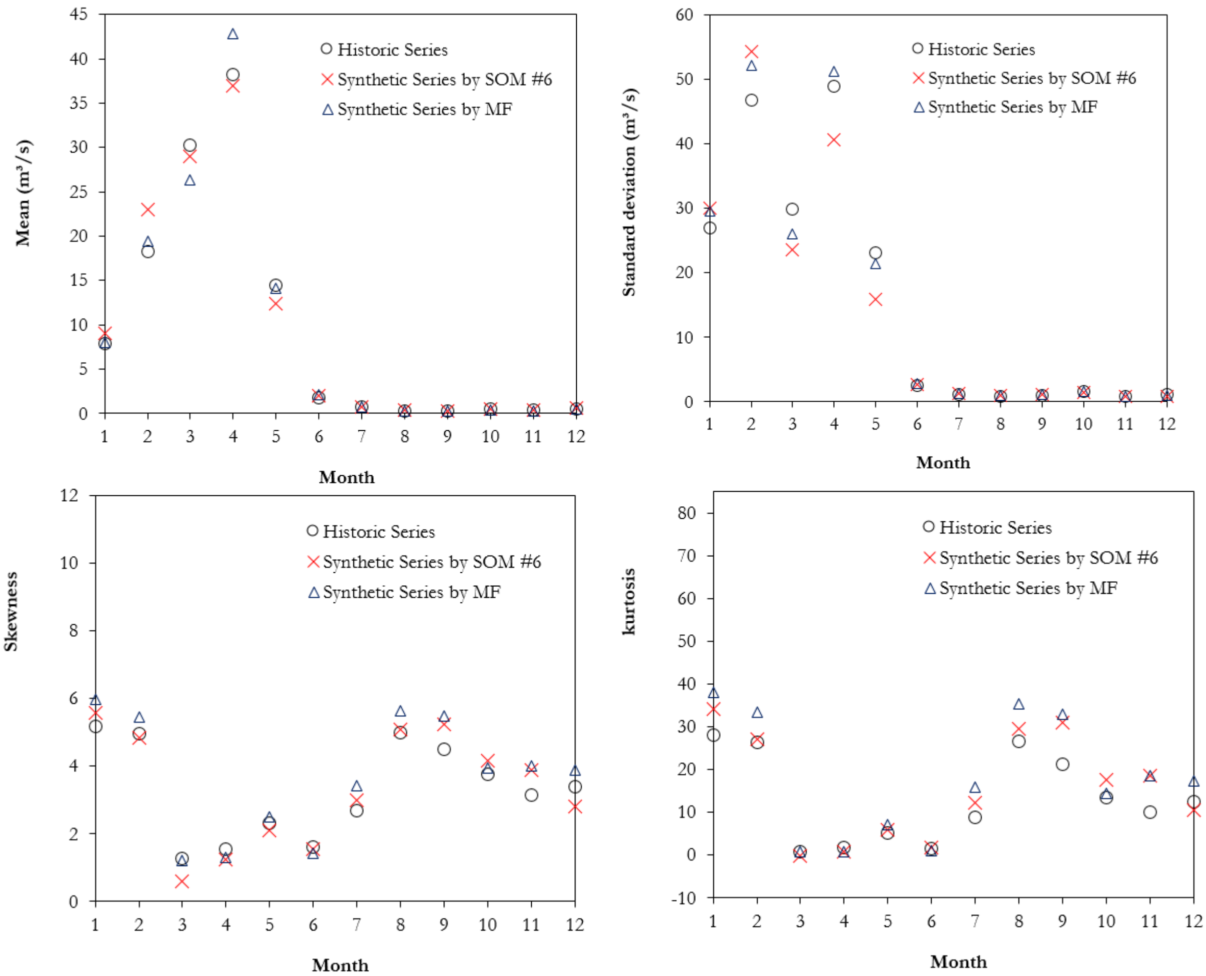

Fig. 4 Comparison of statistical properties between synthetic (models SOM # 6 and MF) and historical flows series in *Piancó* stream gauge station.

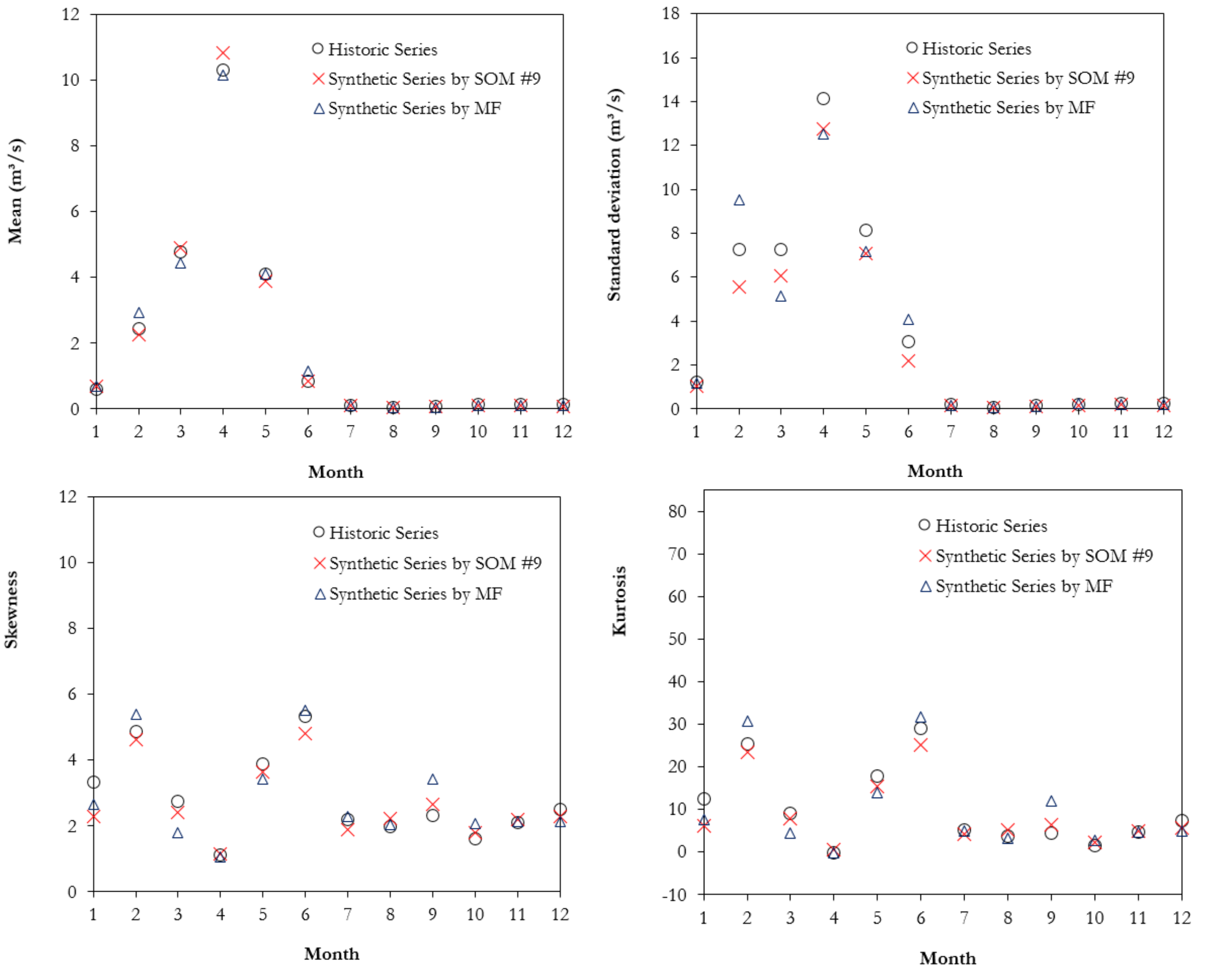

Fig. 5 Comparison of statistical properties between synthetic (models SOM #9 and MF) and historical flows in *Emas* stream gauge station.

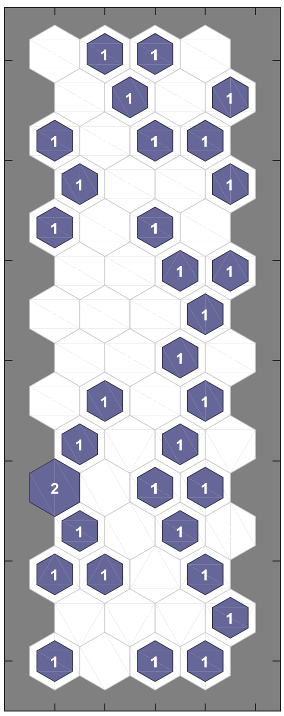

Fig. 6 Allocations of the calibration data by SOM #6 in *Piancó* stream gauge station.

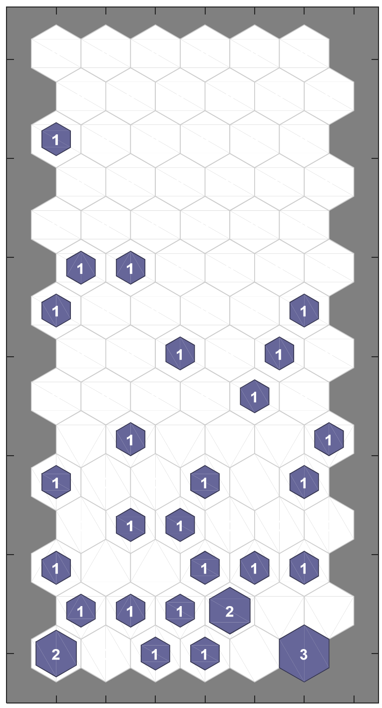

Fig. 7 Allocations of the calibration data by SOM #9 in *Emas* stream gauge station

Through the analysis of **Figs 4** and **5**, it was found that the statistical properties of synthetic monthly flows are similar to those observed in historical data, indicating that the best SOM structure, in both stations, has quality for synthetic generation of stream flows. Moreover, the SOM models were shown to be superior to MF in both study cases, since they could reproduce better the analyzed statistical moments.

Excellent performances of artificial neural networks models applied to water resources field were also found by Carneiro & Farias (2013); Dornelles *et al.* (2013); Farias & Santos (2014); Farias *et al.* (2015); Kasiviswanathan & Sudheer (2016) and Valipour (2016).

Allocations of calibration data in the topological map (hit map), referring to the best architecture of each case study, occurred as shown in **Figs 6** and **7**. We identified that the input data is almost evenly distributed for the application in Piancó stream gauge station. As for Emas stream gauge station, the input data concentrated at the bottom of the map at the end of calibration.

Figures 8 and **9** illustrate the plans of components of the best SOM structures for both study cases. A color scale denotes the values of neuron components. The yellow hexagons correspond to higher values, while hexagons in black mean lower ones.

Comparing the map of annual volumes with the maps of monthly fragments in *Piancó* stream gauge station, we can notice that the highest annual flows are associated with average values in the corresponding fragments of January, February, March, April, May and June. In *Emas* stream gauge station, high annual volumes are associated with medium and high values of fragments in March, April and May. We can also verify through the color gradients that the fragments behavior changes drastically over the months.

Park *et al.* (2014) used plans of components (7×12) with similar geometric structure to those presented in this paper. The authors applied the SOM method to cluster agricultural reservoirs using environmental variables throughout Korea. They found out similarities or discrepancies between water quality variables (chlorophyll-a, total suspended solids, dissolved oxygen, chemical oxygen demand, total nitrogen and total phosphorus) and hydrogeomorphical variables (altitude, bank height, bank width, circumference, reservoir length, surface area, pondage and catchment area).

Nourani *et al.* (2013) developed procedures using the SOM method with feed-forward neural network and wavelet transform to present hybrid black box models for multivariate daily and multi-step ahead rainfall–runoff forecasting. The authors found out high values for the determination coefficient in the calibration (0.94) and validation (0.93) in a structural arrangement of 7×15 for a lead-time of 1-day.

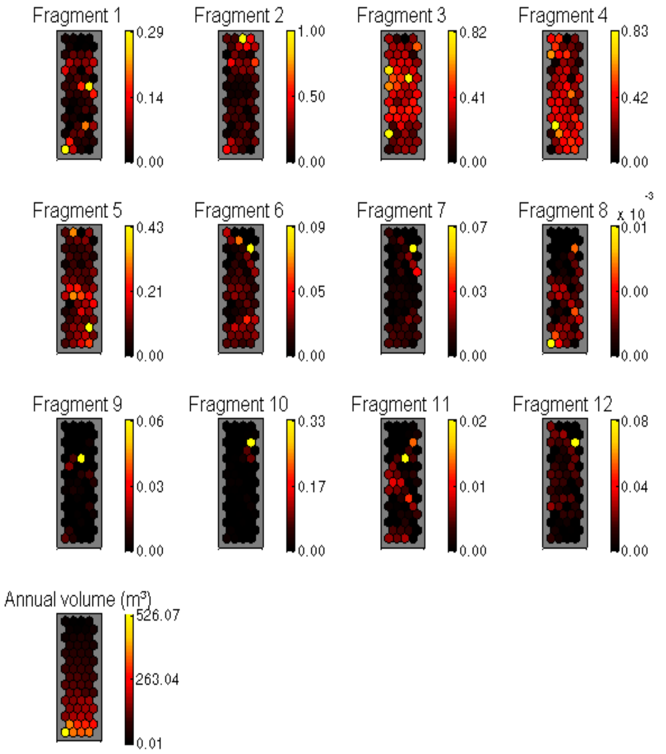

Fig. 8 Plans of the components obtained by the calibration of SOM #6 for Piancó stream gauge station.

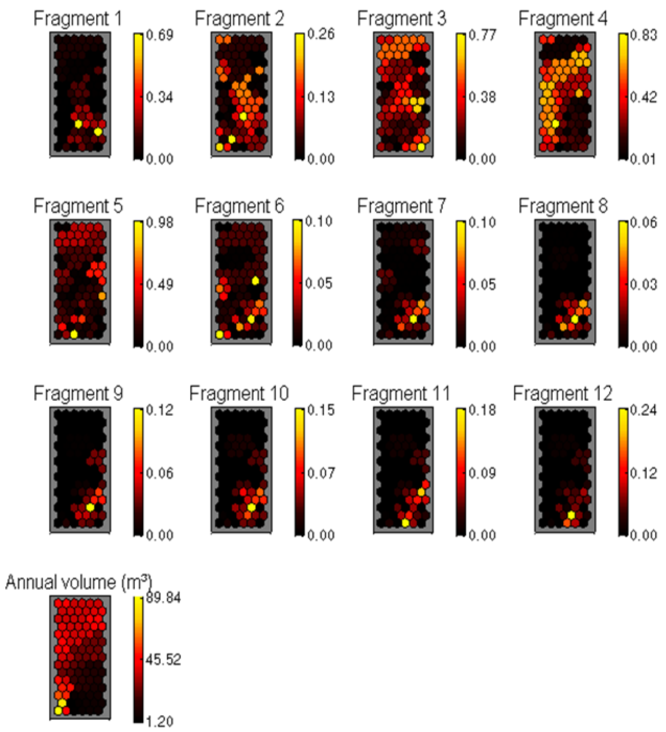

Fig. 9 Plans of the components obtained by the calibration of SOM #9 for Emas stream gauge station.

Figures 10 and **11** show examples of different scenarios of stream flows that can be simulated by using the SOM models in *Piancó* and *Emas* stations, respectively.

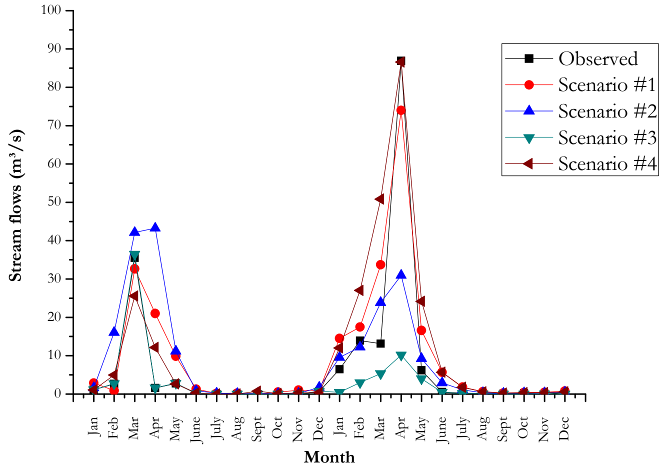

Fig. 10 Comparison of synthetic scenarios (SOM #6) with stream flows observed in 1999 and 2000 in *Piancó* stream gauge station.

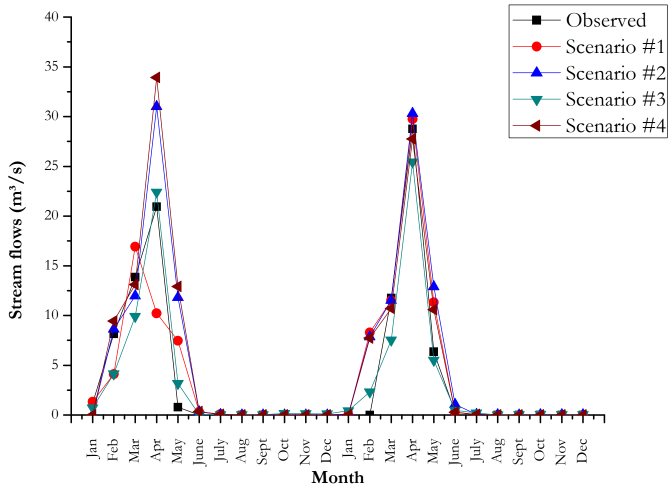

Fig. 11 Comparison of synthetic scenarios (SOM #9) with stream flows observed in 1999 and 2000 in *Emas* stream gauge station

We can observe in **Figs 10** and **11** that the flows have similar behavior when compared four possible scenarios with the observed flows. It is important to observe that the flow simulators provide different possibilities, but maintaining a strong statistical coherence of moments. It was also possible to verify a considerable and expected variation among the stream flow scenarios in the rainiest months.

Stojković *et al.* (2017) developed a procedure with hydrological data from Great Morava River Basin in Serbia that presented: the deterministic component of a monthly flow time-series, which was measured by spectral analysis; and the stochastic component, which was modelled by cross-correlation transfer functions, artificial intelligence (three-layer feedforward neural network) and polynomial regression. Their results statistically showed that the model can be applied to forecast monthly or seasonal flow rates. In this way, we highlight that the self-organizing maps can also be used for the same purpose.

CONCLUSION

We presented models based on self-organizing maps for the synthetic generation of monthly flows considering two stream gauge stations located in *Piancó* River Basin, semiarid Brazil.

For both stream flow stations, the proposed SOM models were able to provide different scenarios of stream flows while preserving the most important statistical moments of the historical series. Moreover, the best-calibrated SOM networks outperformed the traditional Method of Fragments. However, SOM networks will not always be superior to traditional models. In order to overcome this, it is crucial to analyze several options of SOM structures and verify which one performs best.

The main advantage of the proposed model, when compared to other available procedures, is the possibility of better preserving the statistical moments in a modeling using only a process based on artificial intelligence (SOM networks) combined with a traditional method. Thus, this study may support the reconstruction and/or extension of stream flow series, which could help as input data to projects of engineering and support the planning and management of existing water resources.

REFERENCES

Adeloye, A.J., Rustum, R. & Kariyama, D. (2011) Kohonen self-organizing map estimator for the reference crop evapotranspiration. *Water Resources Research*, **47**(8), 1–19. Doi: 10.1029/2011WR010690

Agência Nacional de Águas. (2016) *Banco de dados Hidroweb.* Vazão diária (m³/s). Retrieved from: < http://hidroweb.ana.gov.br/ >. Accessed in: January 2016.

American Society of Civil Engineers. (1993) Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management. Committee Irrigation and Drainage Division. Criteria for evaluation of watershed models. *Journal of Irrigation and Drainage Engineering*, **119**(3), 429-442. Doi: 10.1061/(ASCE)0733-9437(1993)119:3(429)

Beale, M., Hagan, M. & Demuth, H. (2012) *Neural Network Toolbox 7.0.3: User's Guide*. The Math Works Inc, Natick, USA.

Carneiro, T.C. & Farias, C.A.S. (2013) Otimização estocástica implícita e redes neurais artificiais para auxílio na operação mensal dos reservatórios Coremas - Mãe D água. *Revista Brasileira de Recursos Hídricos*, **18**(4), 115-124. Doi: 10.21168/rbrh.v18n4.p115-124

Celeste, A.B. & Billib, M. (2012) Improving Implicit Stochastic Reservoir Optimization Models with Long Term Mean Inflow Forecast. *Water Resources Management*, **26**(9), 2443-2451. Doi: 10.1007/s11269-012-0025-1

Celeste, A.B., Curi, W.F. & Curi, R.C. (2007) Análise de Métodos para Geração Sintética de Vazões em rios intermitentes. In: *XVII Simpósio Brasileiro de Recursos Hídricos*, São Paulo, 1-18. ABRH Publ. 17, ABRH, Porto Alegre, Brazil.

Chandwani, V., Agrawal, V. & Nagar, R. (2015) Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks. *Expert Systems with Applications*, **42**(2), 885–893. Doi: 10.1016/j.eswa.2014.08.048

Chang. L.C., Chang, F.J., Wang, K.W. & Dai, S.Y. (2010) Constrained genetic algorithms for optimizing multi-use reservoir operation. *Journal of Hydrology*, **390**(1), 66-74. Doi: 10.1016/j.jhydrol.2010.06.031

Collischonn, W. (2001) Simulação Hidrológica em Grandes Bacias. Thesis, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

Dornelles, F., Goldenfum, J.A. & Pedrollo, O.C. (2013) Artificial neural networks methods applied to forecasting river levels. *Revista Brasileira de Recursos Hídricos*, **18**(4), 45-54. Doi: 10.21168/rbrh.v18n4.p45-54

Dralle, D.; Karst, N.; Müller, M.; Vico, G. & Thompson, S. E. (2017) Stochastic modeling of interannual variation of hydrologic variables. *Geophysical Research Letters*, **44**(14), 7285–7294. Doi: 10.1002/2017GL074139

Emch, P.G. & Yeh, W.G. (1998) Management model for conjunctive use of coastal surface water and ground water, *Journal of Water Resources Planning and Management*, **124**(3), 129-139. Doi: doi.org/10.1061/(ASCE)0733-9496(1998)124:3(129)

Eyduran, E., Zaborski, D., Waheed, A., Celik, S., Karadas, K. & Grzesiak, W. (2017) Comparison of the Predictive Capabilities of Several Data Mining Algorithms and Multiple Linear Regression in the Prediction of Body Weight by Means of Body Measurements in the Indigenous Beetal Goat of Pakistan. *Pakistan Journal of Zoology*, **49**(1), 257–265. Doi: 10.17582/journal.pjz/2017.49.1.257.265

Farias, C.A.S. (2009) Derivation of hedging rules and forecasts by artificial neural networks for optimal real-time management of an integrated water resources system. Thesis, Ehime University, Matsuyama.

Farias, C.A.S., BEZERRA, U.A. & SILVA FILHO, J. A. (2015) Runoff-erosion modeling at micro-watershed scale: a comparison of self-organizing maps structures. *Geoenvironmental Disasters*, **2**(1), 1–8. Doi: 10.1186/s40677-015-0022-9

Farias, C.A.S., Carneiro, T.C. & Lourenço, A.M.G. (2012) Mapas auto-organizáveis para modelagem chuva-vazão. In: XI Simpósio de recursos hídricos do Nordeste, João Pessoa, 1-14. ABRH Publ. 11, ABRH, Porto Alegre, Brazil.

Farias, C.A.S., Kadota, A., Suzuki, K. & Shigematsu, K. (2011) Stochastic generation of daily groundwater levels by artificial neural networks. *Annual Journal of Hydraulic Engineering*, **55**(4) 55-60. Doi: 10.2208/jscejhe.67.I_55

Farias, C.A.S. & Santos, C.A.G. (2014) The use of Kohonen neural networks for runoff erosion modeling. *Journal of Soils and Sediments*, **14**(7), 1242–1250. Doi: 10.1007/s11368-013-0841-9

Farias, C.A.S., Santos, C.A.G., Lourenço, A.M.G. & Carneiro, T.C. (2013) Kohonen neural networks for rainfall-runoff modeling: case study of Piancó River Basin. *Journal of Urban and Environmental Engineering*, **7**(1), 176–182. Doi: 10.4090/juee.2013.v7n1.157-163

Farias, C.A.S., Silva Filho, J.A. & Bezerra, U.A. (2014) Um modelo de mapas auto-organizáveis para análise e estimativa da produção de sedimentos em uma mircrobacia localizada no semiárido paraibano. In: *XI Encontro Nacional de Engenharia de Sedimentos*, João Pessoa, 1-12. ABRH Publ. 17, ABRH, Porto Alegre, Brazil.

Farias, J.A.M. (2003) Métodos de geração de vazões mensais e suas influências sobre a curva de garantia versus vazão regularizada em reservatórios do Estado do Ceará. Dissertation, Federal University of Ceará, Fortaleza, Ceará, Brazil.

Farzad, F. & El-Shafie, A.H. (2017) Performance Enhancement of Rainfall Pattern – Water Level Prediction Model Utilizing Self-Organizing-Map Clustering Method. *Water Resources Management*, **31**(3), 945–959. Doi: 10.1007/s11269-016-1556-7

García, H.L. & González, I.M. (2004) Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence*, **17**(3), 215–225. Doi: 10.1016/j.engappai.2004.03.004

Gomes, M.B., Pazeto, A.B.O., Tractenberg, L.E.F. & Pinheiro Junior, L.P. (2017) Gestão de stakeholders (GS) no gerenciamento de projetos (GP): casos múltiplos sob a luz do PMBOK. *South American Development Society Journal*, **3**(7), 158–173. Doi: 10.24325/issn.2446-5763.v3i7p158-173

Haykin, S. (1999) *Neural Networks: a Comprehensive Foundation*. New Jersey: Prentice Hall, USA.

Kasiviswanathan, K.S. & Sudheer, K.P. (2013) Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. *Stochastic Environmental Research and Risk Assessment*, **27**(1), 137-146. Doi: 10.1007/s00477-012-0600-2

Kasiviswanathan, K.S. & Sudheer, K.P. (2016) Comparison of methods used for quantifying prediction interval in artificial neural network hydrologic models. *Modeling Earth Systems and Environment*, **2**(22), 1-11. Doi: 10.1007/s40808-016-0079-9

Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**(1), 59–69. Doi: 10.1007/BF00337288

Kohonen, T. & Simula, O. (1996) Engineering Applications of the Self-Organizing Map. *Proceeding of the IEEE*, **84**(10), 1354–1384. Doi: 10.1109/5.537105

Kumar, A.R.S., Goyal, M.K., Ojha, C.S.P., Singh, R.D., Swamee, P.K. & Nema, R.K. (2013) Application of ANN, Fuzzy Logic and Decision Tree Algorithms for the development of reservoir operating rules. *Water Resources Management*, **27**(3), 911-925. Doi: 10.1007/s11269-012-0225-8

Lima, C.A.G. (2004) Análise e sugestões para diretrizes de uso das disponibilidades hídricas superficiais da bacia hidrográfica do rio Piancó, situada no estado da Paraíba. Thesis, Federal University of Campina Grande, Campina Grande, Paraíba, Brazil.

Loucks, D.P., Stedinger, J. R. & Haith, D. A. (1981) *Water Resources Systems Planning and Analysis*. New Jersey: Prentice Hall - Englewood Cliffs, USA.

Nash, J.E. & Sutcliffe, J.V. (1970) River Flow Forecasting through Conceptual Models: Part I - A Discussion of Principles. *Journal of Hydrology*, **10**(3), 282–290. Doi: 10.1016/0022-1694(70)90255-6

Nourani, V., Baghanam, A.H., Adamowski, J. & Gebremichael, M. (2013) Using self-organizing maps and wavelet transforms for space-time pre-processing of satellite precipitation and runoff data in neural network based rainfall-runoff modeling. *Journal of Hydrology*, **476**, 427–445. Doi: 10.1016/j.jhydrol.2012.10.054

Park, Y.S., Kwon, Y.S., Hwang, S.J. & Park, S. (2014) Characterizing effects of landscape and morphometric factors on water quality of reservoirs using a self-organizing map. *Environmental Modelling & Software*, **55**, 214-221. Doi: 10.1016/j.envsoft.2014.01.031

Parmar, K.S. & Bhardwaj, R. (2015) River water prediction modeling using neural networks, fuzzy and wavelet coupled model. *Water Resources Management*, **29**(1), 17–33. Doi: 10.1007/s11269-014-0824-7

Pulito-Velázquez, M., Andreu, J. & Sahuquillo, A. (2006) Economic optimization of conjunctive use of surface water and groundwater at the basin scale. *Journal of Water Resources Planning and Management*, **132**(6), 454-467. Doi: 10.1061/(ASCE)0733-9496(2006)132:6(454)

Rodrigues, A.C.L., Celeste, A.B., Barbosa, R.L., Curi, W.F. & Curi, R.C. (2007) Estudo da influência da transposição do rio São Francisco sobre o processo de outorga de uso da água na bacia hidrográfica do Rio Piancó-PB. In: *XVII Simpósio Brasileiro de*

Recursos Hídricos, São Paulo, 1-17. ABRH Publ. 17, ABRH, Porto Alegre, Brazil.

Sathya, R. & Abraham, A. (2013) Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*, **2**(2), 34–38. Doi: 10.14569/IJARAI.2013.020206

Scientec. (1997) Associação para Desenvolvimento da Ciência e Tecnologia. *Plano Diretor de Recursos Hídricos da Paraíba*: Bacias do Piancó e do Alto Piranhas. SEPLAN, Brasil.

Silva Filho, J.A., Farias, C.A.S., Araújo, S.C. & Bezerra, U.A. (2015) Utilização do método dos fragmentos para geração sintética de vazões: um estudo de caso no semiárido paraibano. In: *XII Congresso Nacional de Meio Ambiente de Poços de Caldas*, Poços de Caldas, 1-8. IFSULDEMINAS Publ. 12, IFSULDEMINAS, Poços de Caldas, Minas Gerais, Brazil.

Silva, I.N., Spatti, D.H. & Flauzino, R.A. (2010) *Redes neurais artificiais para engenharia e ciências aplicadas*. São Paulo: Artliber.

Stojković, M.; Kostić, S.; Plavšić, J. & Prohaska, S. (2017) A joint stochastic-deterministic approach for long-term and short-term modelling of monthly flow rates. *Journal of Hydrology*, **544**, 555–566. Doi: 10.1016/j.jhydrol.2016.11.025

Svanidze, G.G. (1980) Mathematical Modeling of Hydrologic Series (for Hydroeletric and Water Resources Computations). Fort Collins: *Water Resources Publications*.

Valipour, M. (2016) Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms. *Royal Meteorological Society*, **23**(1), 91-100. Doi: 10.1002/met.1533

Wang, P., Gauthier, I. & Cottrell, G. (2016) Are Face and Object Recognition Independent? A Neurocomputational Modeling Exploration. *Journal of Cognitive Neuroscience*, **28**(4), 1–17. Doi: 10.1162/jocn_a_00919

Wilks, D.S. (2006) Statistical Methods in the Atmospheric Sciences. 2d ed. *International Geophysics Series*, Academic Press, London.