

RESEARCH ARTICLE

Open Access



Runoff-erosion modeling at micro-watershed scale: a comparison of self-organizing maps structures

Camilo Allyson Simões de Farias^{*}, Ulisses Alencar Bezerra and José Adalberto da Silva Filho

Abstract

Background: In the last decades, several runoff-erosion models have been proposed to estimate soil erosion, which may lead to loss of fertile land and increase sedimentation and pollution in water bodies. Physically-based erosion models are usually used for such purpose, but a major problem concerning their use is the difficulty to directly measure parameters in the field. This problem can be overcome by exploring empirical models, such as so-called Self-Organizing Maps (SOM). An SOM is a type of Artificial Neural Network (ANN) based on a competitive learning approach for clustering and modeling a variety of databases. Since studies on soil erosion modeling based on SOM are very incipient, we compared some structures of SOM with the purpose of estimating sediment yield based on runoff and climatological data at the micro-watershed scale. The case study was a micro-watershed within the *Sumé* Experimental Basin, which is located in a semiarid region of Brazil. Different from the conventional ANN, SOM-based models represent a multidimensional data set by means of a bidimensional matrix of features, which may be applied for analysis and estimation purposes. In order to calibrate and validate the proposed SOM structures, we used data from 117 rainfall events that occurred between 1985 and 1991.

Results: Analyses of the results indicate that all SOM structures were efficiently calibrated with *NASH* coefficients (Nash & Sutcliffe 1970) varying from 0.88 to 0.90. The SOM structure with 6×8 neurons was the most effective for estimating sediment yields when considering the validation data set ($NASH = 0.73$). The generated maps showed that sediment yields were directly related to runoff and rainfall intensity and inversely correlated to average vegetation heights. The dry period length did not seem to influence the production of sediments.

Conclusions: SOM were shown to be very practical and meant to be applied to specific locations. This type of methodology also demands long term data and dynamic recalibration with up-to-date information in order to account for changes in the watershed.

Keywords: Kohonen maps; Hydro-sedimentology; Sediments; Artificial intelligence; Semiarid

Background

Severe natural conditions together with improper management of intense erosion are sources of numerous environmental impacts. In the Brazilian semiarid, characterized by scarce but intense rainfall events, a large production of sediments occurs only periodically, reducing the storage capacity of rivers and reservoirs as well as negatively affecting the quality of natural resources

and agricultural productivity (Farias & Santos 2014; Farias et al. 2010; Vanmaercke et al. 2010).

Sediment modeling is generally conducted using physically-based erosion models. However, a major problem concerning their use is the difficulty of measuring their parameters in the field (Farias & Santos 2014; Vanmaercke et al. 2010). Empirical models, such as those based on Artificial Neural Networks (ANN), are options to tackle such difficulties.

In the last decade, many authors suggested techniques based on ANN in order to understand hydro-sedimentological processes. Cobaner et al. (2009), for example, proposed an adaptive neuro-fuzzy approach to

^{*} Correspondence: camilo@ccta.ufcg.edu.br
Academic Unit of Environmental Science and Technology, Federal University of Campina Grande, 58840-000 Pombal, Paraíba, Brazil

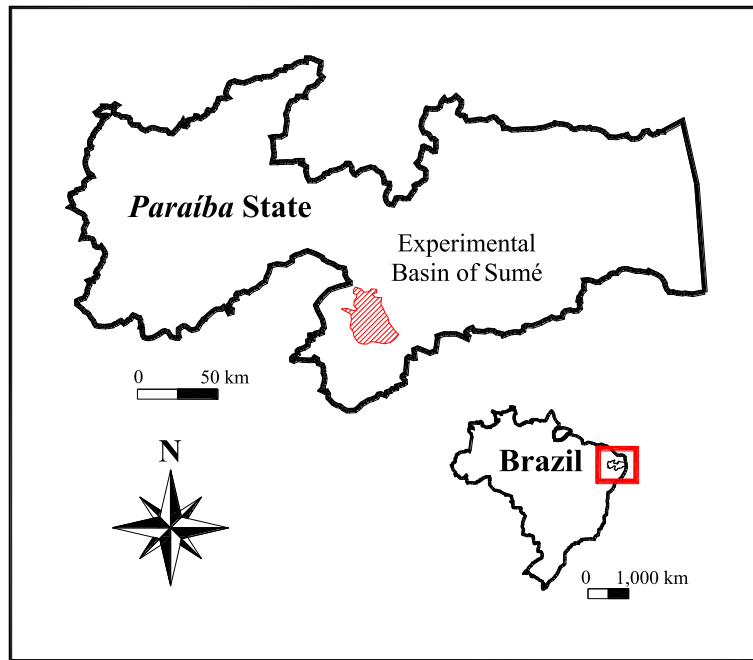


Fig. 1 Location of the Experimental Basin of Sumé, Paraíba State, Brazil

estimate suspended sediment concentration on a daily basis and obtained better performance when compared with sediment rating curves. Márquez & Guevara-Pérez (2010) investigated erosion processes in furrows and found that an ANN-based model outperformed procedures based on physical processes and linear and nonlinear regressions. More recently, Farias & Santos (2014) developed a model based on Self-Organizing Maps

(SOM), which is a type of unsupervised ANN, in order to estimate sediment yields in an erosion plot located in a semiarid land of Brazil and found promising results.

Although there are only a few applications of SOM modeling within the sedimentation engineering field, this technique has been used in other engineering areas. One example is the study of García & González (2004), who applied SOM for understanding the behavior of variables

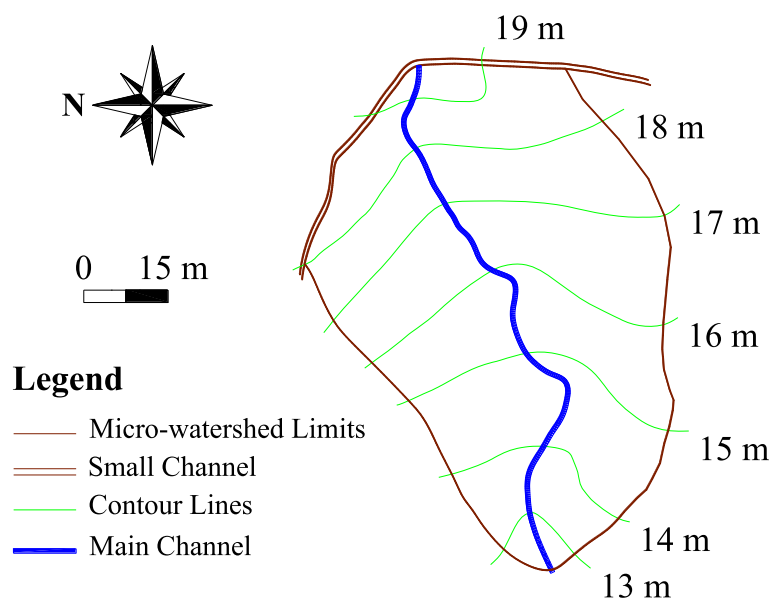


Fig. 2 Contour lines of the micro-watershed of this study. Source: Adapted from Srinivasan & Galvão (2003)

from a wastewater treatment plant. Other application is the work of Adeloje et al. (2011), who proposed an SOM estimator for reference crop evapotranspiration and obtained results superior to those from recommend empirical methods. Farias et al. (2013) developed an SOM-based rainfall-runoff model and verified it was reliable for estimating streamflows in Piancó River, Brazil.

The SOM models represent a multidimensional dataset by means of a bidimensional matrix of features, which can be explored for analysis and estimation purposes. Based on a competitive training approach, this type of model is capable of clustering input data according to their similarities and may be later of use for pattern classifications (Haykin 1999; Kohonen 1982).

Considering the lack of studies based on SOM networks for soil erosion modeling, we compared the performance of five different structures of SOM with the purpose of estimating sediment yield based on runoff and climatological data at the micro-watershed scale.

Materials and methods

Study Area and Data

The selected study area was a micro-watershed within the *Sumé* Experimental Basin, which is located in a semiarid land of *Paraíba* State, Brazil. The Experimental Basin of *Sumé* is placed in the *Paraíba* River Basin, in latitude 7° 40' S e longitude 37° 00' W, and comprises a sub-basin of 10 km²; four micro-watersheds, with areas between 4,800 and 10,000 m²; and nine erosion plots of 100 m². This region is characterized by a semiarid climate, with annual mean rainfall and temperature equal to 590 mm and 24 °C, respectively. The annual potential evaporation is 2,900 mm and the native vegetation is a steppe savannah known as *Caatinga* (Srinivasan & Galvão 2003). Figs. 1 and 2 depict the location of the Experimental Basin of *Sumé* and the chosen micro-watershed for this study, respectively. According to Cadier et al. (1983) *apud* Srinivasan & Galvão (2003), the chosen micro-watershed has an area of 4,800 m², 270 m of perimeter and an average declivity of 6.8 %.

The data of runoff, vegetation average height, duration and intensity of rainfall, dry period, rainfall amount, and sediment yield of 117 rainfall events between 1985 and 1991 were obtained from the study of Srinivasan & Galvão (2003). Eight and twenty percent of the data, which correspond to 94 and 23 rainfall events, respectively, were chosen for calibrating and validating the proposed models.

Self-Organizing Maps

Structure

The modeling using Self-Organizing Maps (SOM) consists of representing multidimensional input vectors by means of a bidimensional map (Silva et al. 2010). This

map is composed of units known as neurons that are organized in a manner that preserves neighboring relationships. Therefore, the SOM network is composed of a multidimensional input layer and a bidimensional output layer, in which neurons compete in order to define a winner. Each element of the input vector connects all neurons in the output layer.

In this study, the input vectors contains seven components: runoff, vegetation average height, duration and intensity of rainfall, dry period, rainfall amount, and sediment yield. In the output layer, hexagonal neurons with weights of seven components were adopted, each one corresponding to a specific dimension of the input layer.

According to García & González (2004), a satisfactory estimation of the number of neurons M may be carried out by using the heuristic Eq. (1):

$$M = 5\sqrt{N} \tag{1}$$

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

Since 94 samples were used for calibrating the SOM models, the number (M) of neurons was equal to 48. In this study, five structures containing 48 neurons were tested as shown in Table 1. Fig. 3 shows an example of an SOM structure with 6 × 8 neurons and a neighborhood of three steps.

Training

The training of SOM networks is based on the calculation of Euclidian distances E_i between input vectors and weights of each output neuron i , as shown in Eq. (2).

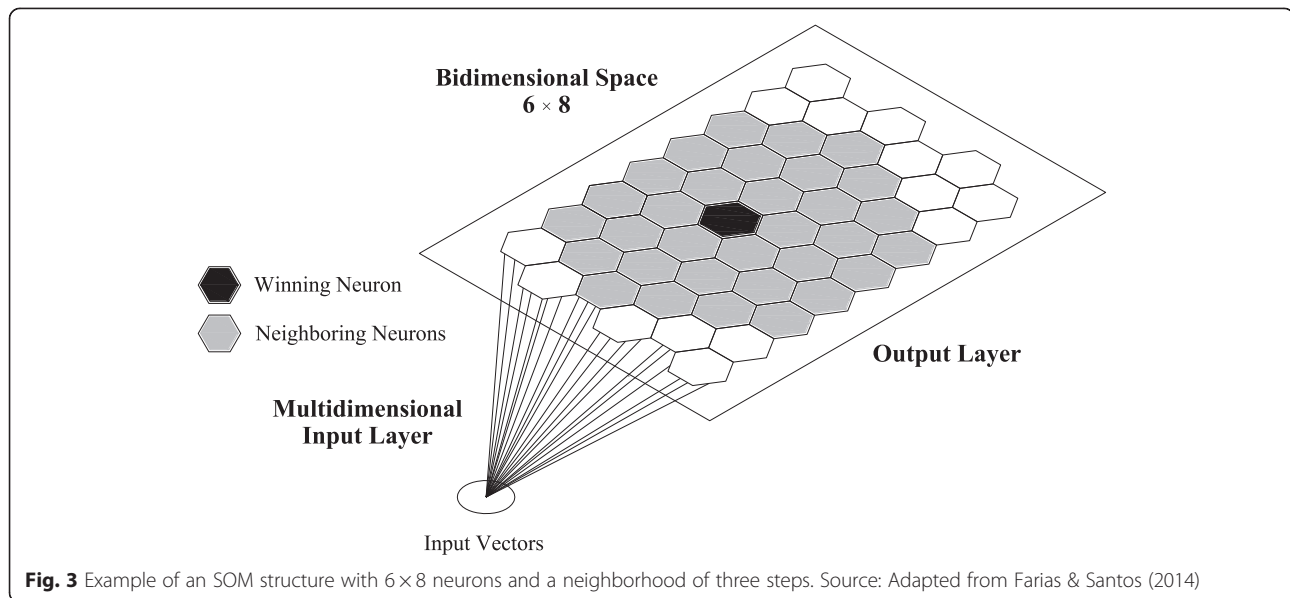
$$E_i = \sqrt{\sum_{j=1}^7 [x(j)-w_i(j)]^2}; \text{ for } i = 1, 2, \dots, 48. \tag{2}$$

in which $x(j)$ is the j th component of the input vector \mathbf{x} and $w_i(j)$ is the j th weight component of the i th output neuron.

The neuron i whose weight vector most closely matches the input data vector \mathbf{x} (i.e., for which the Euclidean distance is the lowest) is declared the winning neuron. Weights associated with this neuron i^* and neighbor neurons in a certain neighborhood radius V_i^* are then updated by the Kohonen rule shown in Beale

Table 1 Structures of the SOM models of this study

Model	Structure
SOM #1	1 × 48
SOM #2	2 × 24
SOM #3	3 × 16
SOM #4	4 × 12
SOM #5	6 × 8



et al. (2012). With the application of this rule, the weights of the winning neuron and its neighbors are adjusted in order to have smaller Euclidean distances and, therefore, to give the output neurons capacity for classifying similar vectors.

In this research, the trainings occur in *batch mode*, i.e., the search for winning neurons is carried out for each input vector and then the weight vector is moved to the average position of all input vectors for which it is a winner or neighbor of a winner. After several presentations of the data set, the weights tend to stabilize. In this study, the calibration was carried out using 200 presentations of the data set.

The network trainings of this study take place in two phases: ordering– and tuning–phases. In the ordering phase, the trainings are limited by a number of 100 presentations of the whole data set and the radius of the neighborhood starts with three steps that decreases to the unit value. This choice is a way to organize the neuron weights in the input space consistently with neuron positions in the dimensional grids. The tuning–phase uses the last 100 presentations, in which only winner neuron weights are updated. At this phase, it is expected that weights are modified relatively evenly in input space, preserving the topology defined in the ordering–phase (Beale et al. 2012).

The SOM models were implemented in MathWorks' MATLAB R2012a using the Neural Network Toolbox (Beale et al. 2012).

Estimation of Sediment Yields with a Calibrated SOM

After the calibrations, the SOM networks were used to estimate sediment yields. In order to carry this out, we consider the sediment yield as missing in the input

vector and followed three steps: (1) calculate the Euclidean distances between the input vector and weights of output neurons disregarding the sediment yield component; (2) determine the winning neuron based on the lowest Euclidean distance; (3) assume the weight component of the winner neuron connected to the missing value of the input vector as the estimation.

Choice of the Calibrated SOM for each Structure

Ten trainings were set for each network in order to improve chances of achieving the best calibration. The training map chosen for each structure was the one that provided the greatest efficiency of *Nash-Sutcliffe (NASH)* between calculated and observed sediment yields considering the calibration data set (Nash & Sutcliffe 1970).

Evaluation of the SOM Structures

The performance evaluation of the SOM structures for estimating sediment yields considered the indexes of correlation (R), relative bias (VR) and efficiency of *NASH*. For this, the calibration and an independent set of data (validation data set) were investigated. The *NASH* efficiency, which can range from $-\infty$ to 1, indicates a perfect match between calculated and observed data when its value is equal to 1 (Nash & Sutcliffe 1970).

Results and discussion

The performance of all proposed SOM structures for estimating sediment yields considering the calibration and validation data sets are shown in Tables 2 and 3, respectively.

Table 2 Performances of SOM structures for estimating sediment yields considering the calibration data set

	SOM #1	SOM #2	SOM #3	SOM #4	SOM #5
<i>NASH</i>	0.90	0.88	0.89	0.90	0.88
<i>R</i>	0.95	0.94	0.95	0.95	0.94
<i>RB</i>	-1.29 %	-2.54 %	-2.75 %	-0.97 %	-3.60 %

Table 3 Performances of SOM structures for estimating sediment yields considering the validation data set

	SOM #1	SOM #2	SOM #3	SOM #4	SOM #5
<i>NASH</i>	0.71	0.72	0.64	0.67	0.73
<i>R</i>	0.84	0.85	0.81	0.85	0.91
<i>RB</i>	-2.75 %	-3.87 %	-1.91 %	-15.51 %	-8.42 %

Analyzing the results from the calibration data set, it was observed that the correlation, relative bias and efficiency of *NASH* varied from 0.94 to 0.95, -3.60 % to -0.97 % and 0.88 to 0.90, respectively. High correlations and values of *NASH* together with low relative bias indicate that the models were efficiently calibrated. The findings also suggest that all SOM structures presented similar performance in the calibration.

The correlation and *NASH* results for the validation data set suggest that SOM #5 outperformed the other

structures, indicating the better match between calculated and observed sediment yields. As for the relative bias, the SOM #3 was the best, but presented the worst values for *NASH* and correlation coefficients. Overall, it can be assumed that the SOM #5 provided better performance than the other structures when a new set of data (validation) was presented.

Figure 4 shows the allocation of calibration data, also known as hits, in the map calibrated for SOM #5. From this map, it is possible to observe that most of the data is concentrated in the upper right corner.

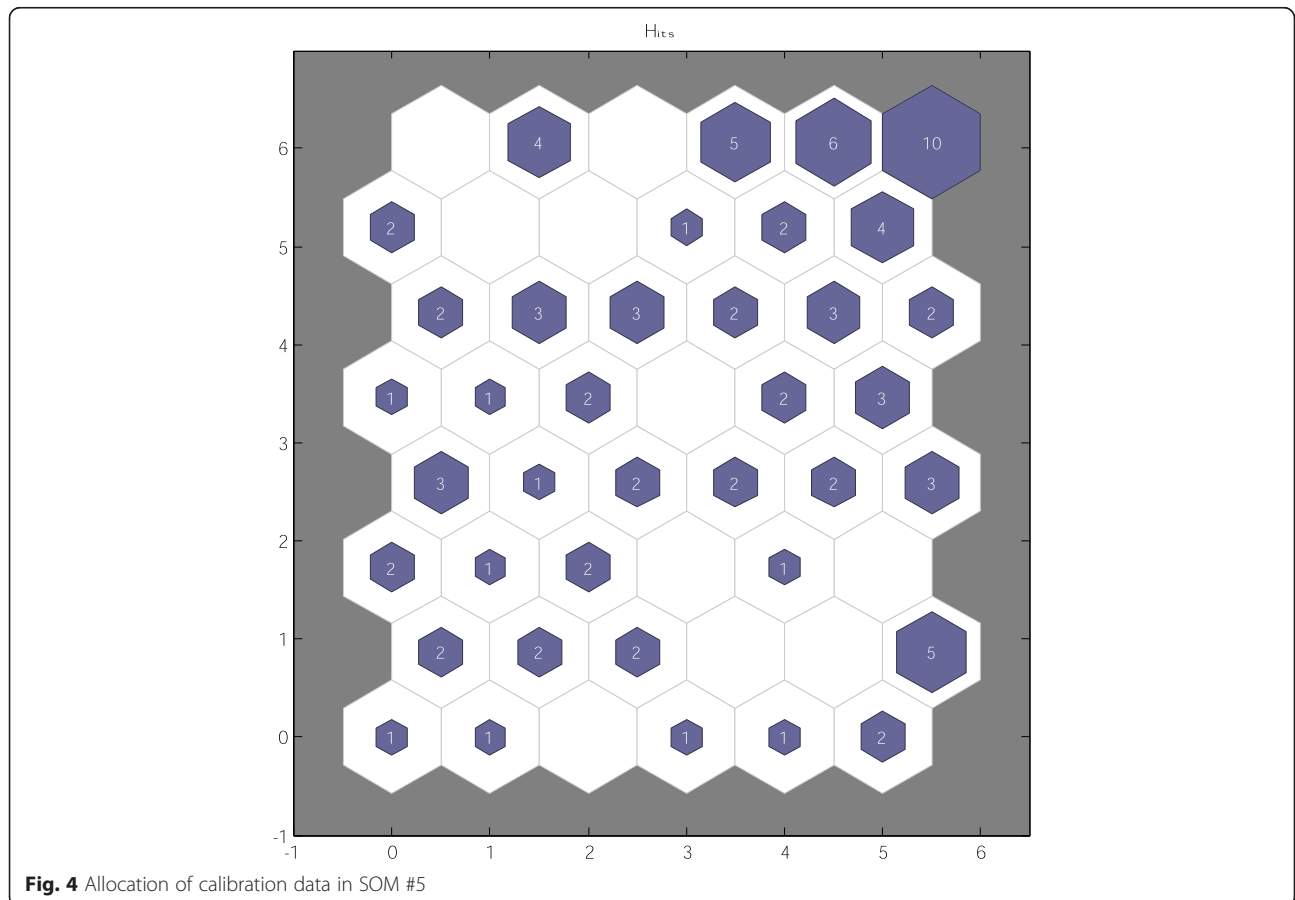


Fig. 4 Allocation of calibration data in SOM #5

Figure 5 illustrates the component planes for SOM structure #5. The color scales represent the values of neurons according to the position in the map. The yellow and black zones correspond to higher and lower values of neurons weights, respectively. Analysis of Fig. 4 and the component plane for sediment yield in Fig. 5 shows that most of the rainfall events resulted in the generation of small amounts of sediments while few produced larger amounts.

Runoff, rainfall amount and rainfall intensity were found to be directly correlated to sediment yield, as revealed by an examination of Fig. 5. This result was expected and is coherent with those found by Farias & Santos (2014).

The color gradients of rainfall duration and intensity confirmed an inverse correlation between these variables. Analyzing the results from the sediment yield map of components, it is found that high and low sediment yields occur when there were low and high vegetation average heights, respectively. This was also an expected outcome since high vegetation heights tend to reduce the effects of rainfall intensity and, consequently, the production of sediments, and vice-versa.

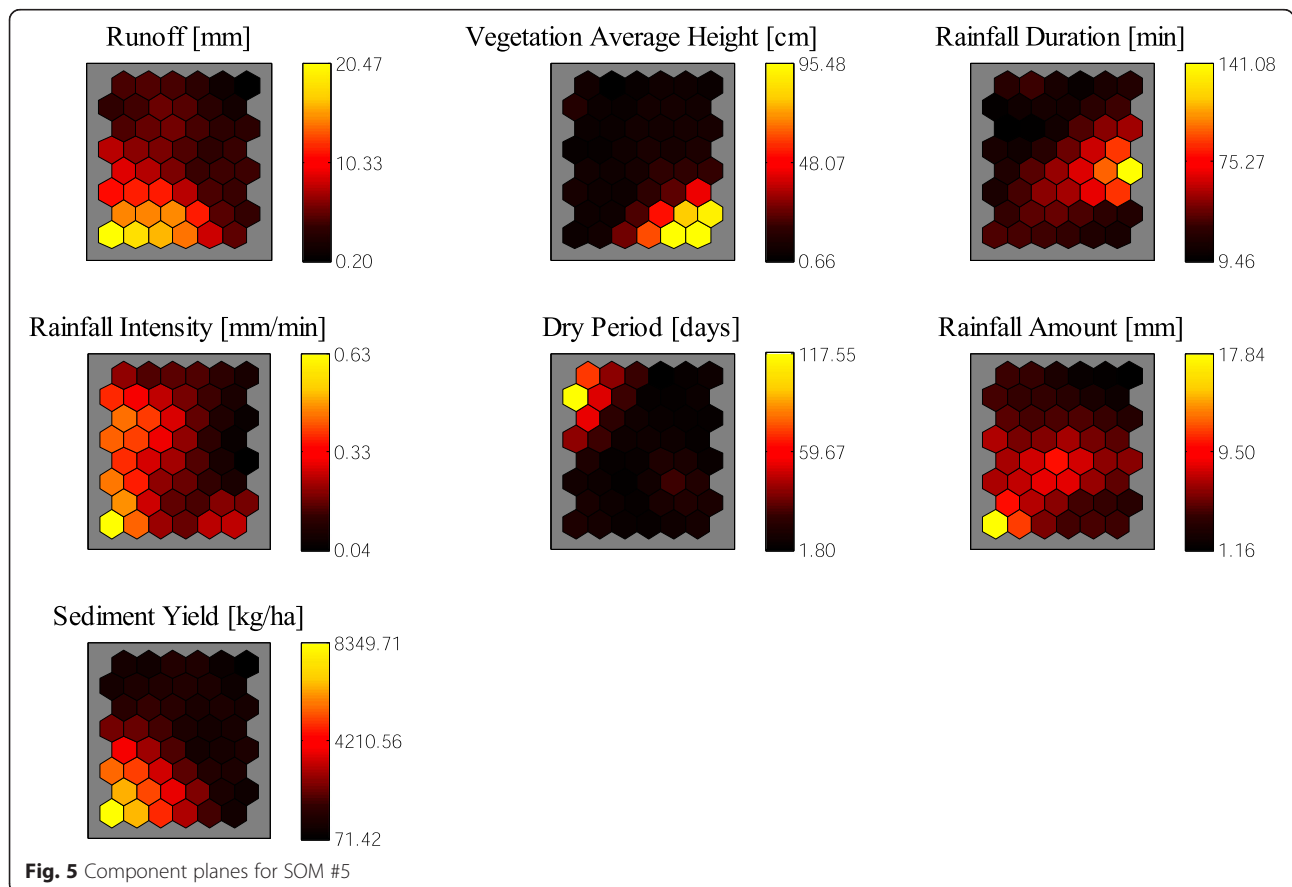
The dry period (i.e. the number of previous days without rainfall events) did not seem to directly influence the production of sediments. This could be justified by

the fact that the longest dry periods coincided with low amounts of rainfall and runoff.

It was verified that the component planes can be used as a tool for identifying inverse and direct relationships among the study variables, mainly due to the possibility of visual identification by means of the color gradients.

The comparison between observed and calculated sediment yields generated by the SOM #5 for the calibration and validation data sets can be seen in Fig. 6 and 7, respectively.

For estimation purposes, hydrological models with *NASH* equal to or higher than 0.75 are considered to be accurate, and acceptable when the *NASH* is higher than 0.36 (Collischonn 2001). In the case of the SOM structure #5, the *NASH* coefficients for calibration and validation data sets 0.88 and 0.73, respectively. The correlation values were also considered high for both calibration ($R = 0.94$) and validation ($R = 0.91$). The values for relative bias indicated an underestimate of sediment yields of -3,6 % and -8,42 % for calibration and validation data sets, respectively. These values can be assumed as low, since their absolute magnitudes were less than 10 %, indicating that the proposed model has an acceptable quality.



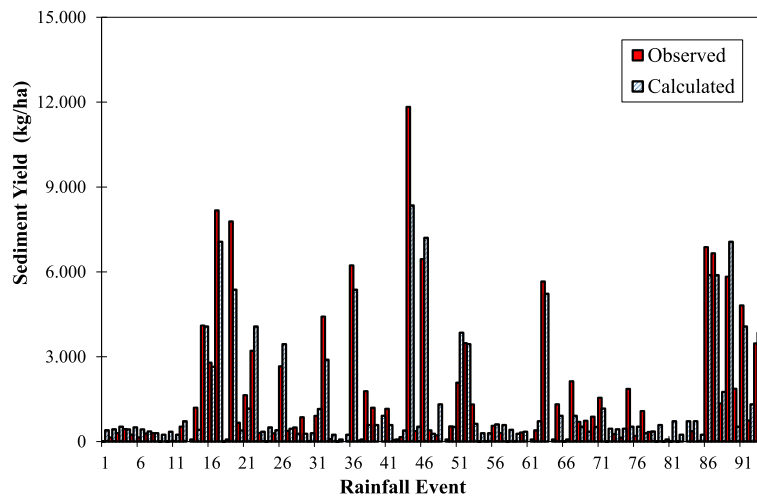


Fig. 6 Comparison between observed and calculated sediment yields generated by the SOM #5 for the calibration data set

The performances of the models were found to be similar to those found by Farias & Santos (2014), who developed a model based on SOM to estimate sediment yield in an erosion plot located in a semiarid land of Brazil.

As opposed to physically-based models, the empirical nature of SOM structures simplifies the soil erosion modeling by not requiring field measurement of parameters for their calibration. Although SOM models are very practical, they are meant to be applied to specific locations. This type of methodology also demands long term data and dynamic recalibration with up-to-date information in order to take into account changes in the watershed, such as climate and land cover changes (Farias & Santos 2014).

Conclusion

In this paper, we compared five structures of Self-Organizing Maps (SOM) with the purpose of estimating

sediment yield based on runoff and climatological data at the micro-watershed scale. The case study was a micro-watershed within the *Sumé* Experimental Basin, which is located in a semiarid region of Brazil.

An analysis of the models indicated that the proposed SOM structures were efficiently calibrated. The SOM structure with 6 × 8 neurons was the most effective for estimating sediment yields when considering the validation data set. The generated maps of components enabled a detailed analysis and understanding of the runoff-erosion process examined in this study.

The empirical nature of the SOM structures simplifies the modeling of erosion processes since it only needs an investigation of the analyzed period. This type of methodology is location specific and demands long term data and dynamic recalibration with up-to-date information in order to take into account changes in the watershed.

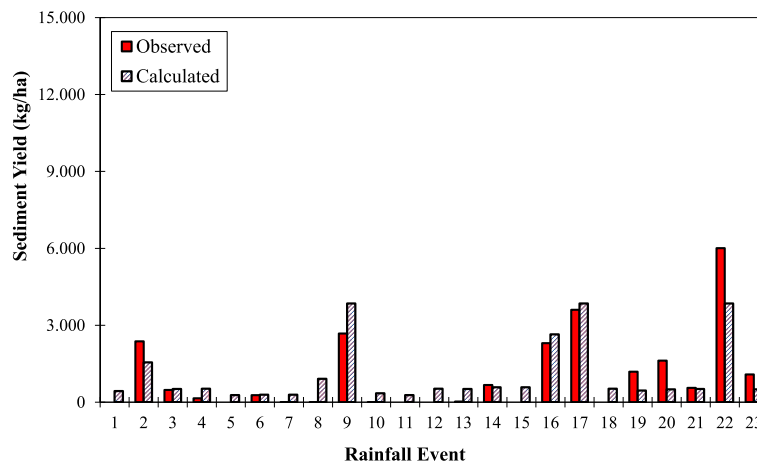


Fig. 7 Comparison between observed and calculated sediment yields generated by the SOM #5 for the validation data set

Competing interests

The authors declare that they have no other competing interests.

Authors' contributions

CASF participated in the analysis of data, carried out the SOM modelings, performed the statistical analysis and drafted the manuscript. Both UAB and JASF organized and analyzed the data, were involved in the design of the study and helped to perform the statistical analysis and draft the manuscript. All authors revised and approved the final version of the manuscript.

Acknowledgments

The authors gratefully acknowledge the Brazilian National Council for Scientific and Technological Development (CNPq) for financial support (Process n.º 477838/2013-8).

Received: 16 December 2014 Accepted: 4 May 2015

Published online: 09 June 2015

References

- Adeloye, AJ, Rustum R, and Kariyama ID (2011) Kohonen Self-Organizing Map Estimator for the Reference Crop Evapotranspiration. *Water Resources Research* 47. doi:10.1029/2011WR010690
- Beale M, Hagan M, Demuth H (2012) *Neural Network Toolbox 7.0.3: User's Guide*, The MathWorks Inc., Natick
- Cadier E, de Freitas BJ, Leprun JC (1983) *Bacia Experimental de Sumé: Instalação e Primeiros Resultados*. SUDENE. Recife - PE, Brazil
- Cobaner M, Unal B, Kisi O (2009) Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. *J Hydrol* 367:52–61
- Collischonn W (2001) *Simulação Hidrológica Em Grandes Bacias*. Universidade Federal do Rio Grande do Sul. Porto Alegre - RS, Brazil
- Farias CAS, Alves FM, Santos CAG, Suzuki K (2010) An ANN-based approach to modelling sediment yield: a case study in a semi-arid area of Brazil. *IAHS Publ* 337:316–321
- Farias CAS, Santos CAG (2014) The Use of kohonen neural networks for runoff-erosion modeling. *J Soils Sediments* 14:1242–1250
- Farias CAS, Santos CAG, Lourenço AMG, Carneiro TC (2013) Kohonen neural networks for rainfall-runoff modeling: case study of piacó river basin. *J Urban Environ Eng* 7:176–182
- García HL, González IM (2004) Self-organizing Map and clustering for wastewater treatment monitoring. *Eng Appl Artif Intel* 17:215–225
- Haykin S (1999) *Neural networks: a comprehensive foundation*. Prentice Hall, Upper Saddle River
- Kohonen T (1982) Self-organized formation of topologically correct feature maps. *Biol Cybern* 43:59–69
- Márquez A, Guevara-Pérez E (2010) Comparative analysis of erosion modeling techniques in a basin of Venezuela. *J Urban Environ Eng* 4:81–104
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models: part I - a discussion of principles. *J Hydrol* 10:282–290
- Silva IN, Spatti DH, Flauzino RA (2010) *Redes neurais artificiais para engenharia e ciências aplicadas*. Artliber, São Paulo - SP
- Srinivasan VS, Galvão CO (2003) *Bacia experimental de sumé: descrição e dados coletados*, UFCCG/CNPq, Campina Grande - PB
- Vanmaercke M, Zenebe A, Poesen J, Nyssen J, Verstraeten G, Deckers J (2010) Sediment dynamics and the role of flash floods in sediment export from medium-sized catchments: a case study from the semi-arid tropical highlands in northern Ethiopia. *J Soils Sediments* 10:611–627

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
