

# **Self-organizing maps of *Kohonen* as a river clustering tool within the methodology for determining regional ecological flows *ELOHA*<sup>1</sup>**

**Mapas auto-organizados de *Kohonen* como una herramienta de agrupación de ríos dentro de la metodología para determinar los caudales ecológicos regionales *ELOHA*<sup>2</sup>**

**Auto-organização de mapas de *Kohonen* como ferramenta *river clustering* na metodologia para determinar fluxos ecológicos regionais *ELOHA*<sup>3</sup>**

SICI: 0123-2126(201307)17:2<311:MAOKHA>2.0.TX;2-X

*Federico González-Cuéllar*<sup>4</sup>  
*Nelson Obregón-Neira*<sup>5</sup>

---

<sup>1</sup> Data Received: January 16, 2012. Accepted: December 12, 2012. This article is derived from the research project entitled “*Ecohydrological grouping of streams in Magdalena-Cauca basin within the framework ELOHA, using self-organizing maps of Kohonen*”. Developed by the research group Ciencia e Ingeniería del Agua y el Ambiente at the Pontificia Universidad Javeriana, Bogotá, Colombia.

<sup>2</sup> Fecha de recepción: 16 de enero de 2012. Aceptado: 12 de diciembre 2012. Este artículo se deriva del proyecto de investigación titulado “*Agrupación ecohidrológica de flujos en la cuenca del Magdalena-Cauca, en el marco ELOHA, utilizando mapas auto-organizados de Kohonen*”. Desarrollado por el grupo de investigación Ciencia e Ingeniería del Agua y el Ambiente de la Pontificia Universidad Javeriana, Bogotá, Colombia.

<sup>3</sup> Data de recebimento: 16 de Janeiro de 2012. Data de aceitação: 12 de dezembro de 2012. Este artigo é derivado do projeto de pesquisa *Agrupamento ecohidrológico de caudais na bacia do Magdalena-Cauca no quadro ELOHA, usando auto-organização de mapas de Kohonen*. Desenvolvido pelo Grupo de Pesquisa Ciencia e Ingeniería del Agua y el Ambiente Bogotá, Colômbia.

<sup>4</sup> Civil Engineer from the Universidad Nacional de Colombia. Graduate of Master in Hydrosystems of the Pontificia Universidad Javeriana. Bogotá, Colombia. E-mail: fgonzalezc@gmail.com.

<sup>5</sup> Civil Engineer from the Universidad Francisco de Paula Santander. Master in Civil Engineering with emphasis in Water Resources and Environment Engineering at the Universidad de los Andes. Ph.D. in Hydrology from the University of California-Davis. Professor of the Civil and Agricultural Engineering Department at the Colombian National University. Bogotá, Colombia. E-mail: nobregonn@unal.edu.co.

### Abstract

This article shows the capacity of the *Kobonen* method in the classification of similar rivers in the light of the determination of regional ecological flows in the Magdalena-Cauca basin, using the methodology of the Ecological Limits of Hydrologic Alteration (ELOHA). For this were used 174 flow stations located throughout the basin, which could be obtained in each case 73 variables for the classification process within ELOHA. Six trials were conducted with different combinations of variables and the criterion for determining the suitability of each combination to classify the streams of the Magdalena-Cauca basin. When the different combinations were estimated with the Kohonen method, we obtained a similar number of groups to the reference number obtained by Ingfofol in 2010, which was also framed within the guidelines of ELOHA, but using other classification methods different from *Kobonen*. It was found that the combinations of variables that best were adjusted to the reference classification were the combinations with lower number of variables which were created in order to characterize in summary form the hydrograph of each analyzed flow station.

### Keywords

ELOHA, Self-organizing Maps and Regional Ecological Flows.

### Resumen

Este artículo presenta la capacidad del método de *Kobonen* para clasificar ríos similares al determinar los flujos ecológicos regionales que se presentan en la cuenca del Magdalena-Cauca, utilizando la metodología de los límites ecológicos de Alteración Hidrológica (ELOHA). Para ello se utilizaron 174 estaciones de flujo las cuales se ubicaron en toda la cuenca, con lo cual se podría obtener en cada caso 73 variables para el proceso de clasificación dentro de ELOHA. Se plantearon 6 ensayos con distintas combinaciones de variables y de criterio para determinar la idoneidad de cada combinación y así poder clasificar las corrientes de la cuenca Magdalena-Cauca. Al determinar mediante el método de *Kobonen* las distintas combinaciones de variables se obtuvo un número similar de grupos al del número de referencia obtenido por Ingfofol en 2010, que también estuvo dentro de las directrices de ELOHA, pero al cual se llegó utilizando otros métodos de clasificación distintas a *Kobonen*. Se encontró que las combinaciones de variables que mejor se ajustaron a la clasificación de referencia fueron las combinaciones con menor número de variables que se crearon, con el fin de caracterizar en forma resumida el hidrograma de cada estación de flujo analizada.

### Palabras clave

ELOHA, mapas de auto-organización y caudales ecológicos regionales.

### Resumo

Este artigo mostra a capacidade do método *Kobonen* na classificação de rios similares à luz da determinação de caudais ecológicos regionais na bacia do Magdalena-Cauca, utilizando a metodologia dos limites ecológicos de Alteração Hidrológica (ELOHA). Para isso foram utilizadas 174 estações de fluxo situadas ao longo da bacia, com o que poderia se obter em cada caso 73 variáveis para o processo de classificação dentro do ELOHA. Foram levantados seis ensaios com diferentes combinações de variáveis e o critério para determinar a adequação de cada combinação na classificação dos riachos da bacia do Magdalena, Cauca, foi determinar, pelo método de Kohonen, as combinações de variáveis que obtiveram um número de grupos similar ao número de referência obtido por Ingfofol em 2010, que também foi enquadrado dentro das diretrizes do ELOHA, ainda que usando outros métodos de classificação diferentes de Kohonen. Verificou-se que as combinações de variáveis que melhor se ajustaram à classificação de referência foram as combinações com menor número de variáveis criadas para caracterizar de forma sumária a hidrografia de cada estação de fluxo analisado.

### Palavras-chave

ELOHA, Auto-organização de Mapas e Fluxos Ecológicos Regionais.

## 1. Introduction

Within the Integrated Water Resources Management (*IWRM*) there are more than 200 methodologies to determine river flow or ecological flow regime of a stream (*Diez*, 2008). However, the vast majority of these can apply only to a specific river stretch, which restricts the possibility of covering large surfaces; aspect that would be ideal for the environmental authorities in Colombia, where resources are limited and jurisdictions are very extensive. The framework *ELOHA* (Ecological Limits of Hydrologic Alteration) is conceptually a methodology that emerged in 2006 precisely in order to cover large territories in regional environmental flow (*Arthington et al.*, 2006).

Despite being relatively new, this methodology has been applied with success in several countries including United States, China, Australia (*Kennard et al.*, 2010) and recently Colombia (*Ingfocol*, 2010), where it was applied to the Magdalena-Cauca basin to obtain the relationship between hydrologic alteration and ecological response for each stream group obtained in the classification process.

The classification of streams is the aspect that gives the regional character to *ELOHA*, since this process it is possible to synthesize in stream groups all the rivers in a large basin, which allows to assign in a preliminary way, environmental flows in all their extension. Normally in hydroclimatology has been used hierarchical and non-hierarchical classification, however, these require a supervised process, i.e. by default assign the number of groups to monitor and calibrate the process of classification (*Lin and Chen*, 2006).

The *Kohonen* Self-Organizing Map Method is a classification method of information that does not need a default number of groups and consequently, this posits it as an ideal methodology to determine the number of groups that in fact underlies the analysis information.

In Colombia, *Kohonen's* networks have been applied successfully in the fields of scientific and technological vigilance of Biogeochemical Cycles (*Gaitán*, 2009) and geospatial data classification (*De Moya*, 2003). On the other hand, exists worldwide literature about application of this method in the hydroclimato-

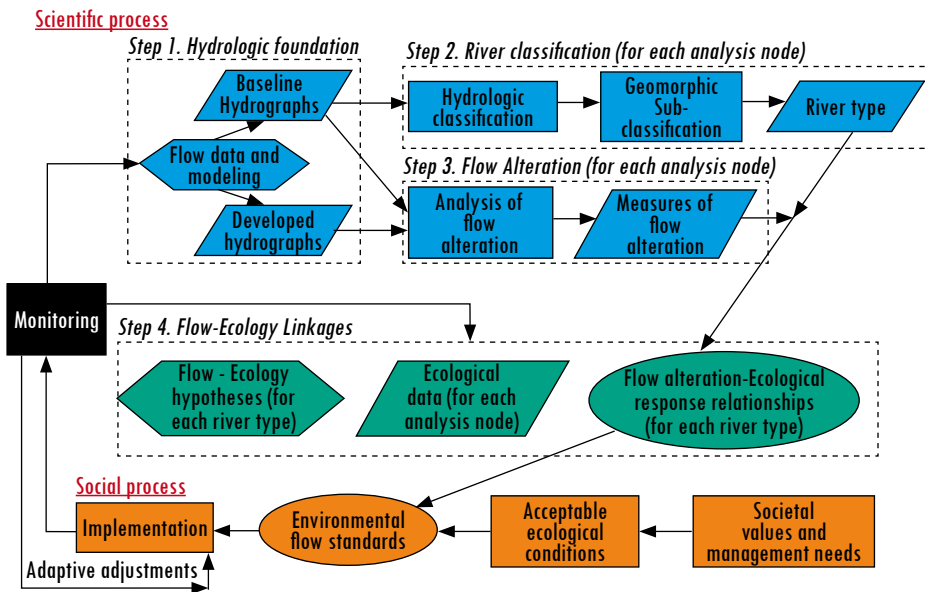
logical classification of regions since that is the case of Lin and Chen (2006), who applied *Kobonen's* networks to classify homogeneous regions of Taiwan for frequency analysis on the basis of effective precipitation data, and Mangiameli *et al.* (1996), who demonstrated that *Kobonen's* method obtained the best results with regard to other seven hierarchical grouping methods.

The follow article intended to show the bounties of *Kobonen's* method in the river classification of Magdalena-Cauca basin within the frame of reference *ELOHA*.

## 2. Ecological limits of hydrologic alteration

*ELOHA* is an adaptative methodology that is divided in three fundamental processes: the scientific, the social, and the monitoring process. In the scientific process are obtained inputs and information necessary to negotiate between the different actors, the environmental flow regime that is agreed in the social process. The monitoring process allow to this methodology be adaptative, since it is possible to evaluate, adapt and to correct the models, the hypotheses and the information used in the scientific and social processes (see Figure 1).

Figure 1. *ELOHA's* methodological scheme



Source: Poff *et al.*, 2010.

In the Figure 1 is possible to observe within the scientific process (process that concern to this article) there exist four steps which goal is to obtain the curves that relate the hydrologic alteration to the ecological response for every stream type. These four steps are:

1. **Hydrological foundation.** It seeks to obtain the hydrograms before and after the development of every analyzed alteration. There are occasions when it is necessary to establish hydrological models that allow to obtain these hydrograms, given the absence of information in the points of interest.
2. **Stream Classification.** This step allows that this methodology should be of regional type since to include big territories it is necessary to group streams with similar characteristics and to obtain the characteristic hydrograms of every stream group to do the subsequent analyses. The streams of study area must be qualified according to their geomorphologic and specially hydrological coincidences. The development of this article focuses concretely on this step.
3. **Flows Alterations.** In order to establish the implications that alterations have or “developments” in every type of river, it is necessary initially to measure the impacts that these generate from the hydrological point of view to be able to relate the degree of hydrologic alteration to the degree of ecological alteration (next step). There are different computer programs like the *IHA* (Indicators of Hydrologic Alteration) that allow to evaluate the hydrologic alteration.
4. **Flow-ecology relationship.** Based on stream classification, curves must be made by the ecological response that every type of river have, regarding the different degrees of hydrologic alteration. These curves will allow to establish in the social process the ecological conditions desired, depending on a hydrologic alteration or a specific hydric utilization associated with every stream group.

For reaching a satisfactory result with *ELOHA*, it is fundamental to adequately classify the streams of the studied region and to obtain the relationships between the hydrologic alteration and the ecological response for every stream group.

### **3. *Kohonen* self-organizing maps**

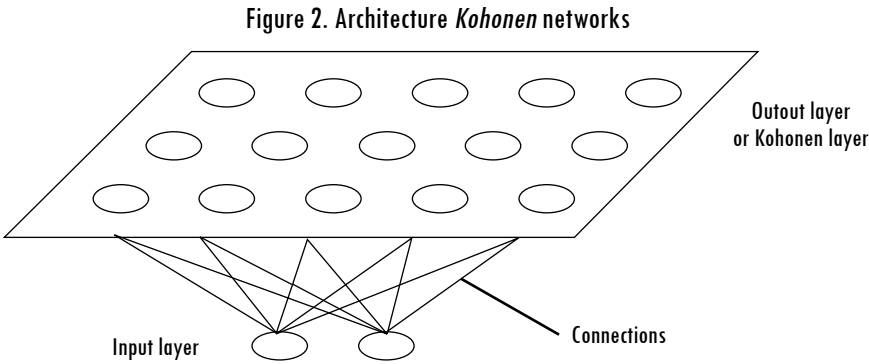
*Kohonen* self-organizing maps (SOM for its initials in english: Self Organizing Maps) are a type of artificial neuronal network (ANN) named Kohonen’s network and usually it is used to classify information and to reduce the number of variables in a specific analysis, no matter how many variables they are, this ANN

allows to visualize the information in two-dimensional maps that preserve and reflect the similarity structure between the incoming information. The aspect of visualization is one of the big advantages of this classification method, so with other methods when there are more than three variables, the visualization of the classification process become enormously complex.

This type of artificial neural networks is characterized because of their competitive learning, that is to say that the neuron models compete between each other in order to know which one is more similar to the presented training pattern, whereupon the weight of the winning neuron are updated in a greater proportion than the weight of the neighboring neurons. The proportion of update in neurons that belong to the neighborhood of the winning neuron diminishes depending on the distance from this. The higher the similarity between two training patterns, the lower will be the distance between their winning neurons. This provides the sense of self-organizing model, because as the network is trained, the winning neurons of similar patterns are forming independent neighborhoods that finally reflect the groups of similar patterns.

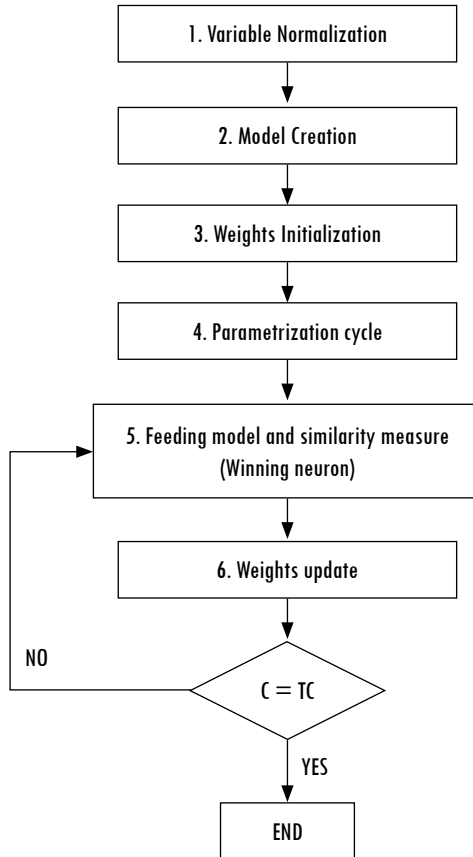
Considering that the training patterns are formed only with the variables of analysis of the classification process and in consequence it is not necessary to include output variables as for example the group to which every pattern belongs, it is said that the training of this type of network classifies as not supervised. This aspect turns out to be another great advantage in opposition to other classification methods of information, which in general need the predefined number of groups.

In the Figure 2 is presented the architecture scheme of this type of artificial neuronal network and in the Figure 3 is shown the flowchart of its training process.



Source: Author's own presentation.

Figure 3. *Kohonen* network operation (C: cycle and TC: total cycles)



Source: Author's own presentation.

In general, the Euclidean distance is used as a similarity metric and the update of the neuron's weight is given according to the equation (1).

$$W_j(t + 1) = W_j(t) + \eta(t) h_j(t) (X - W_j(t)) \tag{1}$$

Where  $W_j$  represents the group of weights of the neuron  $j$ ,  $t$  the corresponding cycle,  $\eta$  the learning rate for the current cycle,  $h_j$  the weighting factor of the neuron  $j$  depending on the established neighborhood for the current cycle and regarding to the winning neuron and  $X$  the set of variables of the pattern presented to the network. The weighting factor of the neuron  $j$  depending on the cycle and neighborhood, regarding to the winning neuron, normally it is decided with a Gaussian function as the one that appears in the equation (2).

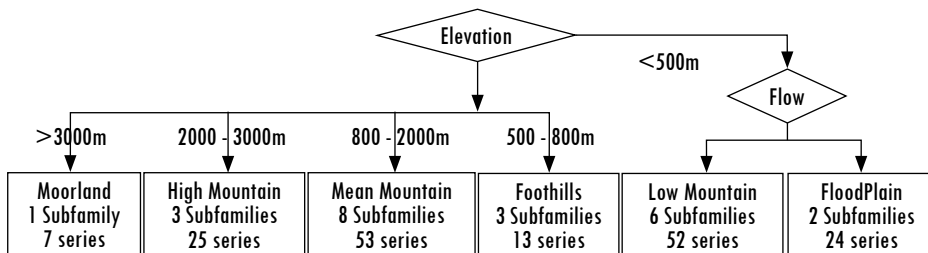
$$h_j = \exp\left(-\frac{\|u_j - u_j^*\|^2}{2\sigma^2}\right) \tag{2}$$

#### 4. Previous experiences of stream classification in the Magdalena-Cauca basin focused on ELOHA

In 2010, Ingfocol Ltda. developed a research project for The Nature Conservancy (TNC) and the Ministry of Environment Housing and Territorial Development (MAVDT), which objective was to determine, under ELOHA's framework, the altered-response curves of the stream groups of Magdalena-Cauca basin. Within this project there were analyzed 174 flow stations distributed in the great basin and 73 variables were used for the stream classification process, of which 68 correspond to the variables given by the computer program IHA, 3 to the spatial coordinates of the stations analyzed and the remaining 2 correspond to the average flow and the respective coefficient of variation of the flow series of each station. The 68 variables given by the computer program IHA, in turn divided in 34 parameters indicated the hydrologic alteration and 34 indicators related to the environmental flow components.

Given the complexity of the streams classification process in this basin, was necessary to perform principal component analysis of, linear correlation and ecological significance of the variables, which were obtained 26 of the 73 variables for the classification process. With these 26 relevant variables were used methods such as *K-means* and cluster with probability distribution. Finally, the stream classification process was guided and supplemented by the help of experts in the basin and the development of several workshops for this purpose. Figure 4 shows the summary of the classification process obtained in this project.

Figure 4. Summary of the stream classification process in the Magdalena-Cauca basin



Source: Ingfocol, 2010.



As can be seen, were obtained in total 23 subfamilies of the 174 stations or series analyzed, which in turn can be classified into 6 geomorphological families shown on top of the rectangles. Given that the methods used for the classification process in this project required the number of default groups, we used the criterion of maximum likelihood for this purpose.

## 5. Methodology

To determine how effective is the *Kohonen* method to classify the streams of Magdalena-Cauca basin framed within the guidelines of *ELOHA*, raised 6 models characterized by different typologies of variables which are presented below in Table 1. The information of flow stations employed in this case was exactly the same as that used for Ingfocol in 2010.

**Table 1. Typologies of variables used.**

No. Typology	Description of variables
1	All 73 variables
2	The 34 variables corresponding to the indicator parameters of hydrologic alteration obtained of IHA program
3	The 34 variables corresponding to parameters related to environmental flow components obtained from IHA program.
4	The 26 variables obtained by Ingfocol in 2010 as relevant variables
5	The 12 average flows of each month and the station elevation
6	The quarter that presents the highest flow, the extremely low average flow, the average flow of large floods, the average flow of the series and the station elevation

Source: Authors' own presentation.

As you can see, the first four typologies contain a lot of variables, while the last two are made by a small number of variables that are directly related to the hydrograph characteristic of each season.

As for the *Kohonen* method for each typology of variable, was tested models from 5 to 900 neurons in the *Kohonen* layer. Models that clearly evidenced the number of groups of the 174 patterns or streams were selected.

The classification prepared by Ingfocol in 2010 was used as validation reference for the classification process of the 6 typologies of variables tested with *Kohonen* method. The criteria to determine which typologies of variables are more appropriate to classify streams of Magdalena-Cauca basin by the *Kohonen* method under *ELOHA*, it is that the number of clusters obtained was similar to the number of clusters obtained by Ingfocol, ie 23 groups.

## 6. Results

One of the main maps for determining the number of groups within the *Kohonen* method is called *U-Matrix* map, which shows in a color scale the magnitude of the Euclidean distance between neurons. The greater the distance between two neurons, the color is darker, while similar neurons or from the same group tend to form colored areas especially yellow. With this map it can quickly view the structure and the number of groups that characterized the classification of the analyzed information.

In the maps shown in Figure 5, only in typologies No. 5 and 6 are identified with the respective number of group, the centroids obtained, since only these two typologies obtained a number of groups of similar streams to the number of groups obtained by Ingfolcol in 2010 (23 groups). The others four typologies clearly provided a much larger number of groups that the reference number. The two typologies of variables that best approached to the classification realized by Ingfolcol in terms of the number of groups correspond to the typologies with fewer variables that in addition also are characterized because those typologies were constructed with a synthetic representation of each stream hydrograph.

Table 2 shows a summary of the classification process by the Kohonen method for each typology of variables.

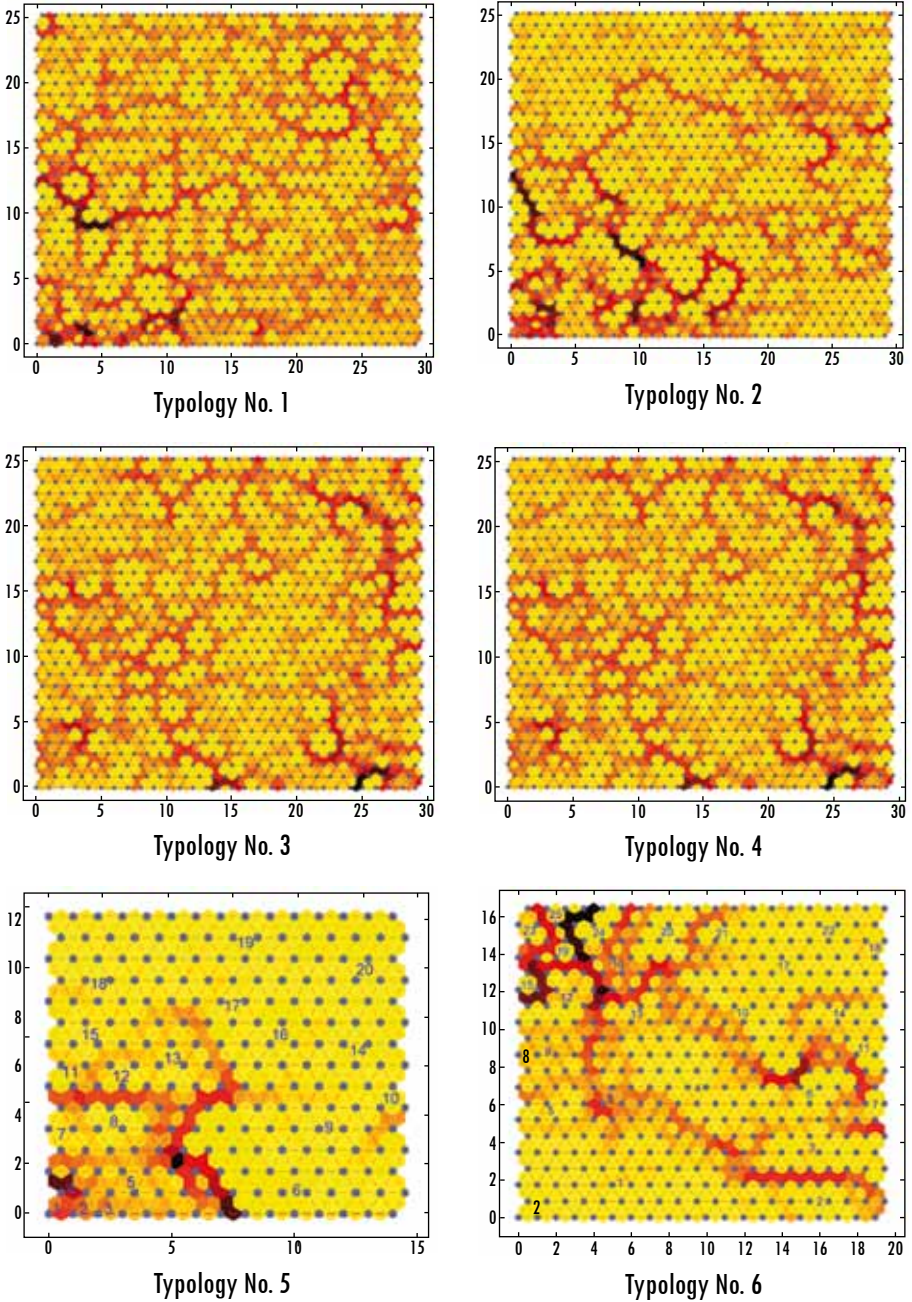
**Table 2. Summary of streams classification process through the *Kohonen* method for different typologies of variables.**

No. Typology	No. Variables	No. neurons	No. groups obtained
1	73	900	>> 23
2	34	900	>> 23
3	34	900	>> 23
4	26	900	>> 23
5	13	225	20
6	5	400	25

Source: Authors' own presentation.

Further analysis showed that the results obtained with the *Kohonen* method for the typologies No. 5 and 6 showed a great invariant percentage of similar streams between 80 and 90%, meaning that between 140 and 157 patterns formed groups with at least two invariant streams—which remain in the same group—with respect to the classification made by Ingfolcol in 2010. Specific results shown in this article can be found in detail in the document of degree project of the Master in Hydrosystems of la Pontificia Universidad Javeriana, titled “*Ecohydrological grouping of streams in Magdalena-Cauca basin within the framework ELOHA, using self-organizing maps of Kohonen*”.

Figure 5. U-Matrix maps for the 6 typologies of variables tested.



Source: Authors' own presentation.

## 7. Conclusions

The *Kobonen* method for classification of hydrological information has not been used extensively, however, is clear the great potential of this tool, especially if you consider you do not need by default the number of groups and it is possible to understand and visualize the results easily.

According to the results of the streams classification on basin Magdalena-Cauca with *Kobonen* method and taking as reference the process performed in 2010 for Ingfocol oriented to *ELOHA*, it can be seen that the typologies with fewer variables (typologies No. 5 and 6) were the typologies that provided better results, because the number of groups obtained in these two cases was very similar to that obtained by Ingfocol.

The typologies of variables No. 5 and 6 were raised in order to characterize each stream hydrograph with variables that were relatively easy to obtain. Considering that these yielded the best results, it is clear that for the stream classification within *ELOHA* by *Kobonen* method not only be used a limited number of variables but they should be closely related to a hydrographical characterization summary of each flow series analyzed.

In this respect, it is recommended for stream classification within *ELOHA*, average flows for each month and the station elevation (typology of variables No. 5). These variables are of easy achievement for any flow series and provide good results, at least in a first approximation.

According to the results obtained by Ingfocol in 2010, the Self-organizing maps of *Kobonen* do not achieve a 100% of stream classification within the framework of *ELOHA*. It is necessary to first establish the relevant variables for the classification process and then supplement the classification obtained by the *Kobonen* method using expert knowledge, data mining techniques and/or data clustering methods which may include a second time the *Kobonen* method.

## References

- ARTHINGTON, A. H.; BUNN, S. E.; POFF, N. L., et al. The challenge of providing environmental flow rules to sustain river ecosystems. *Ecological Applications*. 2006, vol. 16, no. 4, pp. 1311-1318.
- DE MOYA, M. E. Representación y clasificación de datos geoespaciales: comparación entre mapas autoorganizados de Kohonen y el método Gas Neuronal. *Revista Ingeniería e Investigación*. 2003, vol. 53, pp. 12 - 22.
- DIEZ, J. M. *Memorias del Curso-Taller Modelamiento Ecobidráulico de Caudales Ecológicos*. Dictado dentro de la materia Ecología y Ecobidráulica de la Maestría en Hidrosistemas de la Pontificia Universidad Javeriana. 2008.

- GAITÁN, C. F. Vigilancia Tecnológica Científica de Ciclos Biogeoquímicos. *Journal of Technology Management & Innovation*. 2009, vol. 4, no. 2.
- INGFOCOL. *Informe Final de Consultoría: Implementación de la metodología ELOHA (Límites ecológicos de la alteración hidrológica) para la determinación de caudales ambientales regionales en la cuenca Magdalena-Cauca*. Convenio entre The Nature Conservancy (TNC) y el Ministerio de Ambiente, Vivienda y Desarrollo Territorial (MAVDT). 2010.
- KENNARD, M. J.; PUSEY, B. J.; OLDEN, J. D., et al. Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology*. 2010, vol. 55, pp. 171-193.
- LIN, G-F. y CHEN, L-H. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*. 2006, vol. 324, pp. 1-9.
- MANGIAMELI, P.; CHEN, S. K. y WEST, D. A comparison of SOM neural network and hierarchical clustering methods. *European Journal of Operational Research*. 1996, vol. 93, pp. 402-417.
- POFF, N. L.; RICHTER, B. D.; ARTHINGTON, A. H.; et al. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. *Freshwater Biology*. 2010, vol 55, pp. 147-170.

