



XV Seminario Iberoamericano de Redes de Agua y Drenaje, SEREA2017

## SECTORIZATION OF SUPPLY NETWORKS THROUGH THE IMPLEMENTATION OF SELF ORGANIZED MAPS

## SECTORIZACIÓN DE REDES DE ABASTECIMIENTO MEDIANTE LA IMPLEMENTACIÓN DE MAPAS AUTO ORGANIZADOS

Carlos Martins<sup>a\*</sup>, Enrique Campbell<sup>b</sup>, Amilkar Ilaya-Ayza<sup>c</sup>, Bruno Brentan<sup>d</sup>, Joaquín Izquierdo<sup>a</sup>

<sup>a</sup>FluIng-IMM / Universidad Politécnica de Valencia, Valencia, España; <sup>b</sup>Berliner Wasserbetriebe, Berlín, Alemania; <sup>c</sup>Universidad Técnica de Oruro / Facultad Nacional de Ingeniería, Oruro, Bolivia; <sup>d</sup>Universidade Estadual de Campinas / Laboratorio de Hidráulica Computacional, Campinas, Brasil

---

### Astract

Water supply systems (SAAPs) are critical infrastructures for the operation of any city; hence, that its proper management is an aspect to which particular attention must be paid. The sectorization of the SAAPs has emerged as an operational alternative that helps the water management facilities in this task. The design of sectors involves the solution of an optimization problem with multiple objectives, such as investment (in purchase of valves and meters), leak detection capacity and reliability of network operation, among others. Taking this into account, this paper presents an innovative sectorization methodology based on the use of self-organized maps, heuristic optimization, and operational criteria. Self-organized maps, applied to complex networks, allow grouping by recognizing different patterns. Thus, the methodology proposed is based on the application of self-organized maps to define the sectors based on the information stored in the nodes of the network. Next, the optimization of the inputs of the sectors is executed, counting for this with the configuration of the previously obtained sectors. Finally, for the optimization of the feeding points of the established hydrometric sectors, in this work the minimization of implementation costs and the minimization of the loss of hydraulic capacity of the network are considered as objectives.

*Keywords:* supply networks; sectorization; self-organized maps; heuristic optimization

---

### Resumen

Los sistemas de abastecimiento de agua potable (SAAPs) son infraestructuras críticas para el funcionamiento de cualquier ciudad; de ahí, que su correcta gestión sea un aspecto al que haya que prestar particular atención. La sectorización de los SAAPs ha surgido como una alternativa operacional que ayuda a los organismos gestores de agua en dicha tarea. El diseño de sectores implica la solución de un problema de optimización con múltiples objetivos, tales como la inversión (en compra de válvulas y contadores), la capacidad de detección de fugas y la fiabilidad de la operación de la red, entre otros. Tomando en cuenta esto, en este trabajo se presenta una innovadora metodología de sectorización basada en el uso de mapas auto organizados, optimización heurística, y criterios operacionales. Los mapas auto organizados, aplicados a redes complejas, permiten el agrupamiento mediante el reconocimiento de diferentes patrones. Así, la metodología que se propone se basa en la aplicación de mapas auto organizados para definir los sectores con base en la información almacenada en los nodos de la red. A continuación, se ejecuta la optimización de las entradas de los sectores, contando para esto con la configuración de los sectores anteriormente obtenida. Finalmente, para la optimización final de los puntos de alimentación de los sectores hidrométricos establecidos, en este trabajo se consideran como objetivos la minimización de costos de implementación y la minimización de la pérdida de capacidad hidráulica de la red.

*Palabras clave:* redes de abastecimiento; sectorización; mapas auto organizados; optimización heurística.

---

\* Autor de correspondencia: [carlos.martins.a@gmail.com](mailto:carlos.martins.a@gmail.com)

## 1. Introducción

Los sistemas de abastecimiento de agua potable (SAAPs) son infraestructuras críticas para el funcionamiento de cualquier ciudad; de ahí, que su correcta gestión sea un aspecto al que haya que prestar particular atención. La sectorización de los SAAPs ha surgido como una alternativa operacional que ayuda a los organismos gestores de agua en dicha tarea. Esta técnica, surgió en el Reino Unido, a inicios de los años ochenta, con la publicación del reporte 26 de la UK Water Authorities Association [1]. Parte del principio de “divide y vencerás”; en ella, la red es seccionada en varios subsectores con entradas independientes, lo que facilita la operación y monitoreo de los caudales entrantes en los sectores, coadyuvando a la detección de fugas y a su pronta reparación, además de permitir uniformizar las presiones de servicio [2].

La óptima identificación de los subsectores es generalmente una tarea compleja, debido a la cantidad de posibles soluciones de generación de subsectores. Sin embargo, los primeros intentos por sectorizar fueron realizados por ensayo y error; es decir, el profesional encargado evaluaba solo unas pocas alternativas de solución y seleccionaba la que a su parecer era la más adecuada, en función del cumplimiento de unos requisitos, obviamente esto dejaba opciones sin analizar.

En los últimos años, se ha realizado un trabajo importante de investigación con la finalidad de vencer estas dificultades, logrando desarrollar y aplicar técnicas que permiten comparar y probar otras alternativas de generación de subsectores de forma automatizada. Las investigaciones realizadas han sido variadas, pero en cuanto a las técnicas aplicadas para formar o detectar los posibles subsectores, han estado básicamente divididas en dos grandes áreas. La primera de ellas está basada en la partición de grafos [3], [4] y la segunda en clústering jerárquico y la detección de comunidades [5]–[7]. En este trabajo se presenta un enfoque novedoso, aplicando para la determinación de los subsectores una técnica denominada mapas auto organizados (también llamados mapas de Kohonen [8]), combinada con el algoritmo Mean Shift [9] y optimización mediante Particle Swarm Optimization (PSO) [5].

## 2. Metodología

### 2.1. Mapas Auto Organizados

Los mapas auto organizados (SOMs) son un tipo de red neuronal artificial competitiva, de aprendizaje no supervisado, distribuida de forma regular en una rejilla de normalmente dos dimensiones. Estos pueden hacer agrupamientos, clasificaciones, estimaciones, predicciones y minería de datos en una amplia gama de disciplinas [10]. Este algoritmo logra que la red aprenda a agrupar los datos y, de esta forma, reconozca patrones para hallar la soluciones.

Un modelo SOM está compuesto por dos capas de neuronas, tal como se muestra en la figura 1. La capa de entrada (formada por  $N$  neuronas, una por cada variable de entrada) se encarga de recibir y transmitir a la capa de salida la información procedente del exterior. La capa de salida (formada por  $M$  neuronas) es la encargada de procesar la información y formar el mapa de rasgos.

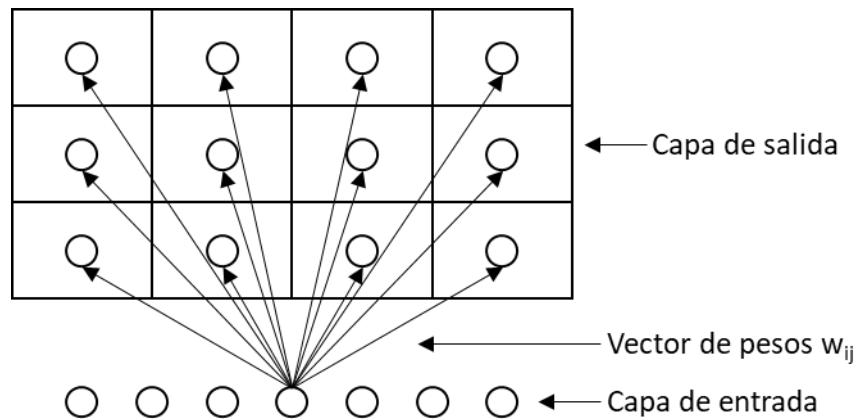


Figura. 1. Esquema de datos de SOM

Normalmente, las neuronas de la capa de salida se organizan en forma de mapa bidimensional, donde las conexiones entre las dos capas que forman la red son siempre hacia delante, es decir, la información se propaga desde la capa de entrada hacia la capa de salida.

Cada neurona de la capa de entrada está conectada con cada una de las neuronas de la capa de salida mediante un peso  $w_{ji}$ . De esta forma, las neuronas de salida también tienen asociado un vector de pesos  $m_j$  llamado vector de referencia (codebook). Así mismo, entre las neuronas de la capa de salida, puede decirse que existen conexiones laterales de excitación e inhibición implícitas, pues, aunque no estén conectadas, cada una de estas neuronas va a tener cierta influencia sobre sus vecinas. Lo cual se consigue a través de un proceso

de competición entre las neuronas y de la aplicación de una función denominada de vecindad, que produce la topología o estructura del mapa; de esta forma, el SOM elabora una proyección desde un espacio multidimensional de datos a un mapa bidimensional de neuronas.

El proceso de aprendizaje cumple con el siguiente procedimiento:

Paso 1. Un vector  $x$  es seleccionado al azar del conjunto de datos y se calcula su distancia (similitud) a los vectores del codebook  $m_j$ :

$$\|x - m_c\| = \min_j \{\|x - m_j\|\}. \quad (1)$$

Paso 2. Una vez que se ha encontrado el vector más próximo o BMU (best matching unit) el resto de vectores del codebook es actualizado. El BMU y sus vecinos (en sentido topológico) se mueven cerca del vector  $x$  en el espacio de datos. La magnitud de dicha atracción está regida por la tasa de aprendizaje.

Mientras se va produciendo el proceso de actualización y nuevos vectores se asignan al mapa, la tasa de aprendizaje decrece gradualmente hacia cero. Junto con ella, también decrece el radio de vecindad,  $\alpha$ . La regla de actualización para el vector de referencia dado  $j$  es la siguiente:

$$m_j(t+1) = \begin{cases} m_j(t) + \alpha(t)(x(t) - m_j(t)), & j \in N_c(t) \\ m_j(t), & j \notin N_c(t) \end{cases} \quad (2)$$

Los pasos 1 y 2 se van repitiendo hasta que el entrenamiento termina. El número de pasos de entrenamiento se debe fijar antes, para calcular la tasa de convergencia de la función de vecindad y de la tasa de aprendizaje. Una vez terminado el entrenamiento, el mapa se ordena preservando la topología original de los datos. Los valores finales de los pesos de las conexiones entre cada neurona de la capa de salida con las de entrada se corresponderán con los valores de los componentes del vector de aprendizaje que consigue activar la neurona correspondiente.

## 2.2. Generación de clústeres mediante Mean Shift

Mean Shift es un algoritmo iterativo no paramétrico, usado con propósitos de búsqueda de modas, formación de clústeres, entre otros, presentado originalmente en 1975 [9]. Básicamente, es un proceso iterativo de punto fijo que converge en un máximo local, donde, en cada iteración, se estima el gradiente normalizado de una función de densidad en el punto correspondiente a cada paso. Aplicado varias veces desde diferentes puntos de partida, se pueden encontrar las zonas donde hay mayor densidad de datos.

Este tipo de algoritmo se basa en asociar cada uno de los datos disponibles a una moda de la función de densidad de probabilidad subyacente. Este simple criterio presenta atractivas ventajas en comparación con otras técnicas de agrupamiento. Además, la estructura de los grupos puede ser arbitraria, sin necesidad de conocer el número de grupos de antemano.

Dado un conjunto de datos  $x_i, i = 1 \dots, n$  de un espacio de dimensión  $d, x \in R^d$ , el estimador del kernel de densidad multivariante usando el kernel  $K(x)$  con ancho de banda  $h$ , es:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right). \quad (3)$$

El objetivo de una función kernel es el de expresar en alguna forma la similitud entre dos objetos. Uno de los kernels más utilizados es el kernel gaussiano o normal, que es un kernel radialmente simétrico y se define como:

$$K_N(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\|x\|^2\right). \quad (4)$$

Para todo kernel radialmente simétrico, como es el caso del kernel normal, ha de cumplirse:  $K(x) = c_{k,d} k(\|x\|^2)$ , siendo  $k(x)$  el perfil del kernel, para todo  $x \geq 0$ , y  $c_{k,d}$ , la constante de normalización que hace que la integral del kernel sea la unidad. En el caso del kernel gaussiano o normal, su perfil es:

$$K_N(x) = \exp\left(-\frac{1}{2}x\right) \quad x \geq 0. \quad (5)$$

El primer paso en el análisis de características con densidad  $f(x)$  es encontrar las modas de esta densidad. Estas modas se encuentran en los puntos donde el gradiente de la función de densidad se anula, es decir, cuando  $\nabla f(x) = 0$ . El gradiente del estimador de densidad se define como:

$$\nabla f_{h,k}(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x - x_i) g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) = \frac{2c_{k,d}}{nh^{d+2}} \left[ \sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \right] \left[ \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right], \quad (6)$$

donde  $g(s) = -k(s)$ . El primer término es proporcional al estimador de densidad  $x$  calculado a partir del kernel  $G(x) = c_{g,d}g(\|x\|^2)$  y el segundo término

$$m_h(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad (7)$$

es el vector Mean Shift o el vector de desplazamiento medio. Este vector apunta siempre a la dirección de máximo incremento de la densidad. El método Mean Shift se obtiene a través de la sucesión de los siguientes pasos: 1. Cálculo del Mean Shift vector  $m_h(x^t)$ , 2. Traslación de la ventana de desplazamiento  $x^{t+1} = x^t + m_h(x^t)$ . Así, mediante este procedimiento se garantiza la convergencia a un punto donde el gradiente de la función de densidad es cero, asegurando así la obtención de un punto estacionario de densidad máxima.

### 2.3. El caudal máximo teórico

El caudal máximo teórico de un SAAP, se obtiene a partir de la curva de consigna y esta, a su vez, se obtiene, para cada presión mínima de servicio requerida en la red. Para el cálculo de dicha curva, se requiere de un modelo fiable de la red, con la finalidad de obtener para cada nodo de la red la altura de la fuente de suministro. De esta forma, se puede garantizar el caudal demandado y la presión mínima de servicio [11].

La obtención de la curva de consigna se realiza de la siguiente forma: en primer lugar, se fija una presión mínima de servicio deseada; posteriormente, se fija un estado de carga en los nodos y se modela la red, obteniéndose los valores de presión para cada uno de los puntos de la red. La altura de cabecera requerida para cumplir con la presión mínima se determinará mediante la diferencia entre la cota real en la fuente y la presión real en el nodo menos la presión mínima requerida, según se muestra en la ecuación 8.

$$\text{Altura de Cabecera}_{nodo} = \text{Cota de Fuente} - (\text{Presión}_{nodo} - \text{Presión mínima requerida}) \quad (8)$$

La altura de cabecera para el estado de carga será el mayor valor obtenido de todos los nodos, ya que, de esta forma, al cumplir la presión mínima para el punto más desfavorable, se cumple para todos los demás nodos. Ahora bien, si se generan varios estados de carga manteniendo una misma presión mínima de servicio, y se repite el proceso anteriormente descrito se podrá hallar, entonces, la curva de consigna o curva de comportamiento del sistema. La intersección de esta curva con la altura real de la fuente de suministro permite calcular el caudal máximo teórico con que puede trabajar la red, tal como se muestra en la Figura 2.

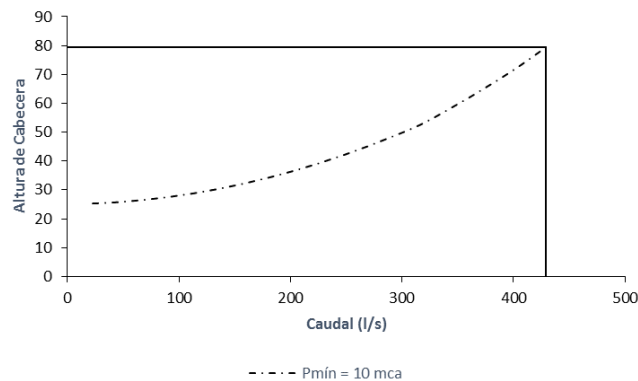


Figura. 2. Esquema Ejemplo de Curva de Consigna

### 2.4. Procedimiento planteado

La generación de clústeres, como analogía a los sectores mediante los SOMs, no siempre resultan en un número manejable o definido de ellos. Por esta razón hay que complementar su funcionamiento con otro método de generación de clúster que simplifique aún más el proceso.

Es importante destacar que los SOM sirven como elemento que permite tomar gran cantidad de información multidimensional y simplificarla en un mapa bidimensional, preservando en todo momento la información topológica de base de la información primaria. Así pues, a partir de los resultados aportados por los SOMs, se aplicará el algoritmo Mean Shift para lograr obtener el número óptimo de sectores de la red de distribución.

El procedimiento de trabajo consistirá, en primer lugar, en indicar al SOM la información resaltante de cada uno de los nodos de la red, para lo cual este algoritmo retornará la información clasificada en una red de neuronas bidimensional, las cuales están organizadas de forma tal, que garantizan la preservación de la información topológica original.

Posteriormente, se le aplica, a la información recibida, el algoritmo Mean Shift, el cual permite agrupar de forma automática (y sin necesidad de indicar el número sectores) la información, quedando de esta forma la red dividida en diferentes sectores.

Para lograr que los sectores queden correctamente definidos, se deben instalar válvulas de cierre y contadores, con la finalidad de lograr el aislamiento correcto de cada uno de los sectores. Pero la selección de qué tubería cerrar y cuál dejar abierta cambia el sentido de circulación de flujo en la red, pudiendo generar una afectación en las presiones y en la capacidad de la red de garantizar el caudal demandado. Debido a esto, se debe hacer un proceso de optimización, esto con la finalidad de seleccionar; así se logrará alcanzar que, con el gasto mínimo de instalación de válvulas de cierre y contadores, se logre garantizar la presión mínima de servicio y la capacidad hidráulica lo más cercana posible al valor inicial.

Para ello se aplica el algoritmo PSO, el cual permite, de forma simple, obtener un vector de solución, que logre alcanzar las metas de optimización anteriormente planteadas. Así pues, la función objetivo a minimizar por el algoritmo estará compuesta por dos términos. En primer lugar, se tendrán los costos de procura e instalación de válvulas y contadores; en este caso, quedará expresada por la ecuación:

$$\text{Costo Valvuleria} = \sum_{i=1}^n c(\text{Val}) + c(\text{Cont}), \quad (9)$$

donde  $n$  es el número de tuberías frontera,  $c(\text{Val})$  es el costo de las válvulas de cierre,  $c(\text{Cont})$  es el costo de los contadores. Así mismo, se deberá garantizar que la disminución en la capacidad de la red (una menor disminución en el caudal máximo teórico de la red) sea la mínima posible, de la siguiente forma:

$$\text{Costo por perdida de capacidad} = \frac{\text{Cap.Ini.} - \text{Cap.Sec.}}{\text{Cap.Ini.}} \times 100.000, \quad (10)$$

donde  $\text{Cap. Ini.}$  es la capacidad inicial de la red o caudal máximo teórico capaz de aportar sin sectorización,  $\text{Cap. Sec.}$  es la capacidad de la red después de sectorizar. Quedando la función objetivo a minimizar de la siguiente forma:

$$\min(\text{Costo Valvuleria} + \text{Costo por perdida de capacidad}). \quad (11)$$

### 3. Resultados y discusión

La metodología planteada fue aplicada en primer lugar a la red de Fossolo [12], la cual es una red pequeña, abastecida desde un embalse, cuenta además con 36 nodos de consumo y 58 tuberías. Esta red resulta ideal, ya que permite verificar el comportamiento del algoritmo de cálculo. Un esquema de la red se puede apreciar en la figura 3.

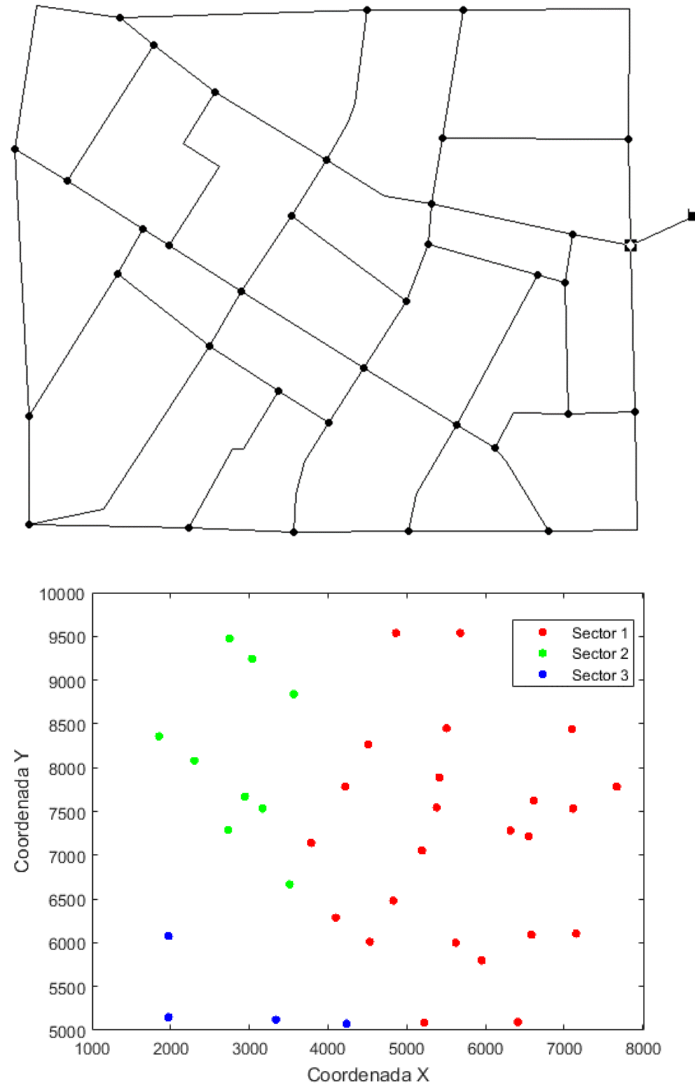


Figura 3. Configuración de la red de Fossolo

El modelo sectorizado aportado por la combinación de SOM y Mean Shift, puede apreciarse en la figura 4. En este se seleccionó una configuración de malla cuadrada de  $7 \times 7$  neuronas para el SOM, con un parámetro de entrenamiento de las neuronas de 4000 pasos, con forma rectangular para las neuronas y con cálculo de distancias entre neuronas basadas en la función euclidiana.

Figura 4. Distribución de los sectores con la aplicación de Mean Shift en la red de Fossolo

Puede apreciarse la conformación de 3 sectores de tamaño diferenciado. Para esta configuración obtenida, se hallaron además 11 tuberías frontera (que limitan los sectores establecidos); a su vez se propone la instalación de cuatro contadores para cuantificar el gasto entrante a los sectores; además de la instalación de siete válvulas de seccionamiento para aislar los sectores formados; todo esto teniendo la red una pérdida de capacidad de 2%.

Posteriormente el algoritmo fue probado en la red denominada Rural [13], la cual posee 476 tuberías y 379 nodos de demanda, alimentados por dos reservorios. En la figura 5 se puede apreciar un esquema de dicha red.

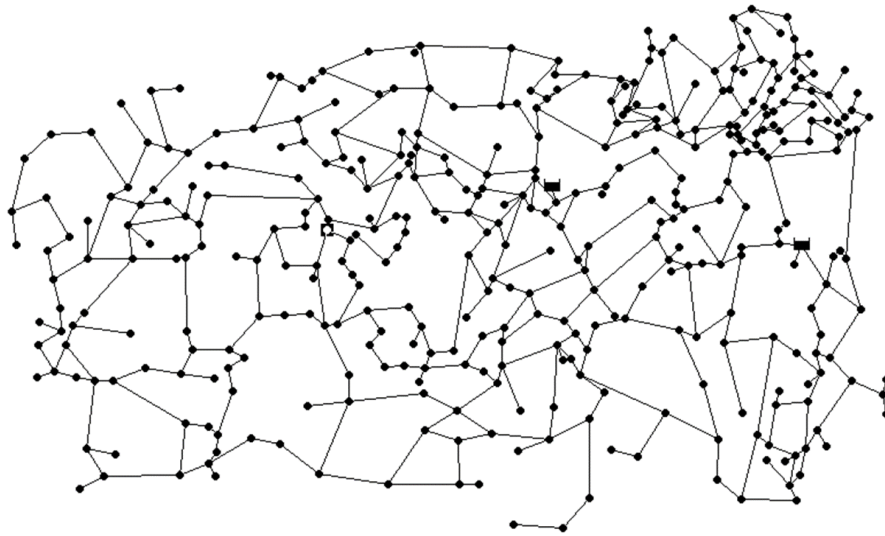
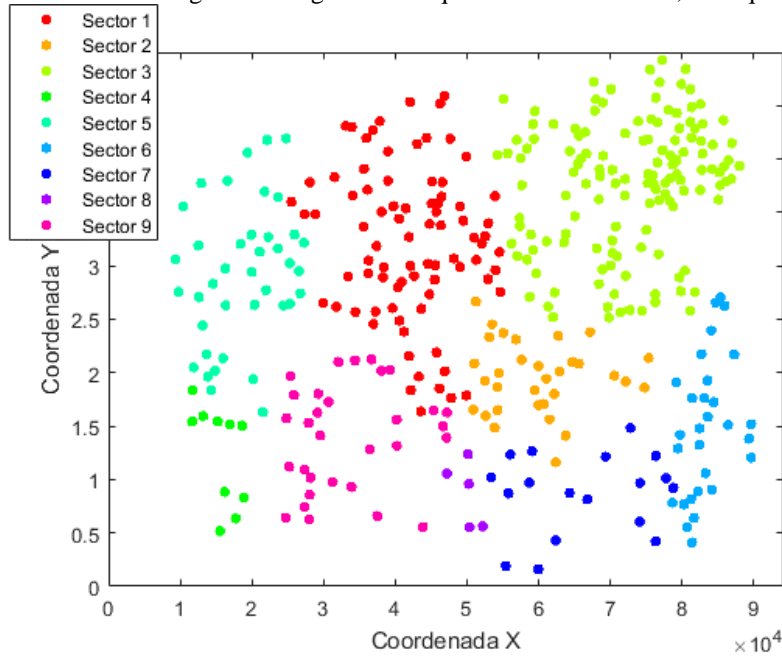


Figura. 4. Arreglo de la red Rural

El modelo introducido en SOM fue configurado de igual forma que en el caso anterior, sólo que se incrementó el tamaño de la red



en  $10 \times 10$  neuronas. El resultado de la sectorización puede apreciarse en la figura 5.

Figura. 5. Distribución de los sectores con la aplicación de Mean Shift en la red Rural

En este caso, se obtuvo 9 sectores de distintas formas. Para esta configuración obtenida, se hallaron además 44 tuberías frontera (que limitan los sectores establecidos), de las cuales en 7 se realizará la instalación de contadores para cuantificar el gasto entrante a los

sectores; además de la instalación de 37 válvulas para aislar los sectores formados, todo esto teniendo la red una pérdida de capacidad de 6 %.

#### 4. Conclusiones

La combinación de los algoritmos utilizados permite la identificación y formación de sectores en los SAAPs, a partir de la información aportada por los nodos. En primer lugar, el algoritmo SOM, permite la simplificación y condensación de toda la información de los nodos en una superficie bidimensional, sobre la que actúa posteriormente el algoritmo Mean Shift, permitiendo de esta forma hallar el agrupamiento natural de los nodos y preservando la información topológica inicial.

Puede apreciarse que, a medida que se incrementa la amplitud en la red, los algoritmos tienden a incrementar el número de los sectores. Esto tiene sentido, ya que, desde el punto de vista operativo, al incrementarse el número de sectores en redes de gran amplitud, disminuye el área de supervisión, resultando una red más manejable.

La inclusión de la capacidad de la red (caudal máximo teórico) como elemento a introducir en el proceso de selección de entradas en los sectores, permite visualizar de mejor forma el proceso de sectorización, al arrojar información cuantitativa acerca del estado de la red posterior al proceso de sectorización.

El incremento en la disminución de la capacidad de la red en función del incremento del tamaño, guarda estrecha relación con el incremento en el recorrido que tiene que realizar el agua para llegar a los puntos de suministro.

A pesar de que el algoritmo arroja resultados prometedores, es conveniente incluir en los procesos de sectorización no sólo la información de los nodos, sino incluir además información referente a la conectividad de estos y parámetros hidráulicos, a fin de garantizar un modelo de sectorización integral.

#### Referencias

- [1] J. Morrison, "Managing leakage by District Metered Areas: a practical approach," *Water21*, pp. 44–46, 2004.
- [2] R. Sturm and J. Thornton, "Proactive Leakage Management using District Metered Areas (DMA) and Pressure Management—Is it Applicable in North America," in *IWA Leakage 2005 conference proceedings*, 2005, pp. 1–13.
- [3] V. G. Tzatchkov, V. Alcocer-Yamanaka, and V. Bourguett Ortíz, "Graph Theory Based Algorithms for Water Distribution Network Sectorization Projects," in *Water Distribution Systems Analysis Symposium 2006*, 2006, pp. 1–15.
- [4] A. Di Nardo, M. Di Natale, G. Santonastaso, and S. Venticinque, "Graph partitioning for automatic sectorization of a water distribution system," in *Proceedings of CCWI 2011*, 2011, p. 6.
- [5] J. Izquierdo, A. M. Herrera, I. Montalvo, and R. Pérez-García, "Agent-based division of water distribution system into District Metered Areas," in *4th International Conference, ICSOFT*, 2009, pp. 167–180.
- [6] A. M. Herrera, "Improving water network management by efficient division into supply clusters," Tesis Doctoral, Universidad Politécnica de Valencia, 2011.
- [7] K. Diao, Y. Zhou, and W. Rauch, "Automated Creation of District Metered Area Boundaries in Water Distribution Systems," *J. Water Resour. Plan. Manag.*, vol. 139, no. 2, pp. 184–190, 2013.
- [8] T. Kohonen, "Analysis of a Simple Self-Organizing Process," *Biol. Cybern.*, vol. 44, no. 2, pp. 135–140, 1982.
- [9] K. Fukunaga and D. Hostetler, "The estimation of the gradient of a density function. with applications in pattern recognition.," *IEEE Trans. Inf. theory*, vol. 21, no. 1, 1975.
- [10] A. M. Kalteh, P. Hjorth, and R. Berndtsson, "Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application," *Environ. Model. Softw.*, vol. 23, no. 7, pp. 835–845, 2008.
- [11] A. E. Ilaya-Ayza, J. Izquierdo, R. Pérez-García, C. Martins, and D. Ayala-Cabrera, "La curva de consigna , una herramienta útil para la esqueletización , sectorización y determinación de la capacidad de una red de suministro de agua," *Acta Univ. Multidiscip. Sci. J.*, vol. 26, no. 26(NE-2), pp. 49–58, 2016.
- [12] C. Bragalli, C. D. Ambrosio, J. Lee, and A. Paolo, "Water network design by MINLP," *IBM J. Res. Dev.*, vol. RC24495, no. (W0802-056), 2008.
- [13] D. Eliades, M. Kyriakou, S. Vrachimis, and M. Polycarpou, "EPANET-MATLAB Toolkit: An Open-Source Software for Interfacing EPANET with MATLAB," in *14th International CCWI Conference*, 2016.