

INTEGRACIÓN DE DATOS GEOELÉCTRICOS Y GEOQUÍMICOS USANDO MAPAS AUTOORGANIZADOS (SOM) PARA CARACTERIZAR UN VERTEDERO

CAMILA JULIAO^{1,a}, JOHAN DIAZ^{1,b}, YOSMELY BERMÚDEZ^{1,c}, MILAGROSA ALDANA^{2,d}

¹Universidad Simón Bolívar, Coordinación de Ingeniería Geofísica, Caracas, Venezuela.

²Universidad Simón Bolívar, Departamento de Ciencias de la Tierra, Caracas, Venezuela.

e-mail: ^a14-10540@usb.ve, ^b15-10418@usb.ve, ^c13-10143@usb.ve, ^dmaldana@usb.ve

Recibido: febrero de 2021

Aprobado para publicación: abril de 2021

RESUMEN

Los lixiviados, provenientes de los vertederos de basura, pueden comprometer significativamente el área que los rodea. Aunque la distancia entre estos y las zonas pobladas sea considerable, el riesgo de afectar los acuíferos de uso público resulta inminente en la mayoría de los casos. Por esta razón, la delimitación y monitoreo de la pluma de lixiviados es esencial. En el presente estudio se integran y clasifican, usando una Red Neuronal no supervisada, datos geoelectrónicos (resistividad e IP), con mediciones de metano en superficie, para identificar las posibles zonas de riesgo en áreas aledañas a un vertedero. La Red Neuronal utilizada es una red tipo Kohonen, la cual genera como resultado, mapas de clasificación autoorganizados o SOM (Self-Organizing Map). A partir del entrenamiento de la misma, se obtuvieron dos salidas gráficas en las que se seleccionaron grupos de neuronas que presentaban un comportamiento similar. Posteriormente, fueron generados mapas de contornos correspondientes a la localización de dichos grupos y a las variables individuales, con el fin de realizar una comparación entre la clasificación obtenida y las diferentes anomalías asociadas a cada una de estas variables. Dos de los grupos resultantes de la clasificación se relacionan con valores típicos de líquidos percolados en el vertedero para los parámetros evaluados de forma individual. De esta forma, se obtuvo una delimitación precisa de las posibles zonas afectadas en el vertedero estudiado, integrando las variables de entrada vía SOMs. La localización de la zona de estudio no se detalla por razones de confidencialidad.

Palabras claves: Redes Neuronales, Mapas Autoorganizados, Vertedero, Resistividad, Cargabilidad, Metano.

INTEGRATION OF GEOELECTRIC AND GEOCHEMICAL DATA USING SELF-ORGANIZING MAPS (SOM) TO CHARACTERIZE A LANDFILL

ABSTRACT

Leachates from garbage dumps can significantly compromise their surrounding area. Even if the distance between these and the populated areas could be considerable, the risk of affecting the aquifers for public use is imminent in most cases. For this reason, the delimitation and monitoring of the leachate plume are of significant importance. Geoelectric data (resistivity and IP), and surface methane measurements, are integrated and classified using an unsupervised Neural Network to identify possible risk zones in areas surrounding a landfill. The Neural Network used is a Kohonen type, which generates; as a result, Self-Organizing Classification Maps or SOM (Self-Organizing Map). Two graphic outputs were obtained from the training performed in which groups of neurons that presented a similar behaviour were selected. Contour maps corresponding to the location of these groups and the individual variables were generated to compare the classification obtained and the different anomalies associated with each of these variables. Two of the groups resulting from the classification are related to typical values of liquids percolated in the landfill for the parameters evaluated individually. In this way, a precise delimitation of the affected areas in the studied landfill was obtained, integrating the input variables via SOMs. The location of the study area is not detailed for confidentiality reasons.

Keywords: Neural Networks, Self-organizing Maps, Landfill, Resistivity, Chargeability, Methane.

INTRODUCCIÓN

La agrupación en clústeres constituye una de las herramientas más útiles y es comúnmente usada en el análisis de datos, en particular de datos geofísicos. El tamaño y la dimensionalidad de las bases de datos geoespaciales existentes enfatizan la necesidad de algoritmos de agrupamientos eficientes y robustos. Los mapas autoorganizados ("Self-Organizing Maps" - SOM) son una de estas herramientas de análisis e integración de datos. Los SOM son un tipo de red neuronal artificial, que se entrenan utilizando un aprendizaje competitivo

y no supervisado, para producir una representación discretizada, de baja dimensión, del conjunto de muestras de entrenamiento. Los SOMs se llaman también "mapas de características", ya que conservan las características principales de los datos de entrada (Miljković, 2017). De esta forma, los SOM permiten hacer una integración de distintos parámetros y analizarlos, en forma conjunta, mediante una reducción dimensional de los datos. La salida es un gráfico bidimensional que organiza las muestras a través de un aprendizaje no supervisado (Sakao y Neramballi, 2020). Los mapas autoorganizados son conocidos por sus capacidades de agrupamiento,

visualización y clasificación (e.g. Penn (2005); Dossary, Al-Turki y Badr (2016); Carter-McAuslan y Farquharson (2020)). Strecker y Uden (2002), por ejemplo, utilizan SOMs para clasificar atributos post-apilamiento en un volumen de datos 3D de Lafourche Parish, South Louisiana. El análisis de los mapas autoorganizados les permitió interpretar caracteres estratigráficos que estaban enmascarados en el volumen de amplitud convencional. Kohler, Ohnbergerv y Scherbaum (2010) utilizan este tipo de red neuronal para identificar patrones temporales en registros de vibración sísmica. El objetivo es caracterizar, localmente, un volcán activo en Indonesia. El análisis de SOM les permite visualizar y agrupar características como frecuencia, número de onda, polarización y análisis espectral. Como resultado, logran identificar los campos de ondas que afectan la calidad de las estimaciones de la onda Love.

Dada su utilidad en el reconocimiento de patrones cuando se analizan varias variables, se plantea como objetivo del presente estudio obtener mapas de SOM para caracterizar un área en la que se localiza un vertedero y están presentes acuíferos de uso urbano. Los acuíferos de uso público requieren de un monitoreo constante como consecuencia del crecimiento exponencial de la producción de desechos sólidos que los afectan significativamente. En la actualidad, existen métodos geofísicos no invasivos que permiten llevar a cabo esta tarea. Sin embargo, estos pueden conllevar a ambigüedades en la interpretación de los resultados si se analizan individualmente, los datos derivados de métodos distintos. La falta de exactitud en la interpretación puede ocurrir si las litologías presentes en el área dan respuestas físicas similares a las de los lixiviados; por ejemplo, respuestas resistivas equivalentes (Abdulrahman et al., 2016). En el presente estudio se utilizarán, como datos de entrada para generar los SOM y caracterizar el vertedero, medidas de resistividad, polarización inducida (IP) y mediciones de metano, previamente obtenidas en el área. Se espera generar salidas gráficas que agrupen los datos con mayor semejanza y que ayuden a identificar posibles zonas asociadas a plumas de contaminación.

ASPECTOS TEÓRICOS

Los estudios neurobiológicos indican que distintas entradas, provenientes de los diferentes sentidos del ser humano, es decir, vista, oído, olfato, tacto, movimiento, etc., se asignan a áreas particulares de la corteza cerebral de una manera ordenada. Esto se puede considerar como un mapa topográfico cerebral. Este tipo de mapas corticales tiene, en principio, dos propiedades importantes. Por un lado, en cada estadio de procesamiento de los datos, cada pieza de información entrante es preservada en su propio contexto o vecindad. Adicionalmente, las neuronas que tratan información estrechamente relacionada, se mantienen juntas para que puedan interactuar a través de conexiones sinápticas cortas (Morales, 2018). Con base a este comportamiento, Kohonen (1982) propone un modelo de Red Neuronal Artificial capaz de mapear, efectivamente, patrones semejantes (i.e. vectores de datos de entrada que, por semejanza, estuviesen más cercanos) en localizaciones contiguas en el espacio de salida. Se busca, por tanto,

construir mapas topográficos artificiales, que aprendan a través de la autoorganización, inspirados en la neurobiología (Kohonen, 2001). Según Vesanto (2000) el SOM es una red neuronal bicapa. Consta de una capa de entrada y una de salida.

Los componentes de un SOM son los nodos o neuronas, que tienen asociados un vector de pesos, de la misma dimensión que la de los vectores de entrada, y una posición en un mapa topográfico. La configuración usual de las neuronas en un espacio regular de dos dimensiones es una rejilla, generalmente hexagonal o rectangular. Así, el modelo está compuesto por un vector de pesos, equivalente a las conexiones sinápticas de una neurona real, una entrada formada por N neuronas a seleccionar y un escaler de salida de las neuronas que procesa la información y forma el mapa de rasgos. Los pesos en cada conexión o enlace se corresponden con la intensidad de la sinapsis equivalente. Estos pesos se determinan durante el entrenamiento o proceso de aprendizaje y, usualmente, se inician aleatoriamente (Torrecilla et al., 2009). Inicializados los pesos, se selecciona un vector del conjunto de datos de entrada, de forma aleatoria o secuencial, y se determina la neurona o nodo más cercano al vector de entrada, de acuerdo a sus pesos. El nodo ganador corresponderá a la Mejor Unidad de Coincidencia o BMU, (por sus siglas en inglés Best Matching Unit). Para determinar el BMU se puede utilizar distintas medidas de distancia (euclideana, coseno, Minkowski, producto interno, entre otras).

Diversas licencias y algunas aplicaciones de plataformas de computación, como el "nctool" de Matlab, usan una distancia euclidiana. Los pesos del BMU, así como los de aquellos nodos cercanos al mismo en la cuadrícula del SOM, se modifican para acercarlos al vector de entrada. El cambio en los nodos vecinos depende de una función de vecindad (FV), de forma que los pesos de las neuronas más cercanas al BMU se modifican más que aquellos de las neuronas más alejadas. El proceso anterior se itera para los datos de entrada, disminuyendo la FV con el número de iteraciones. Diferentes FV pueden utilizarse y la velocidad de convergencia del SOM puede depender de la selección (Vesanto, 2000; Ota et al., 2011; Kohonen, 2013).

Se han implementado SOMs usando una función que asigna un valor de 1 a las neuronas más cercanas al BMU y cero a las más lejanas; funciones gaussianas, utilizadas por aplicaciones como el "nctool" de MATLAB; el negativo de la segunda derivada de una función gaussiana, que es fuertemente atractiva a pequeña escala, algo repulsiva a mediana escala y ligeramente atractiva a gran escala e, incluso, funciones de vecindad asimétricas. De acuerdo a este proceso, entradas sucesivas y diferentes causarán correcciones en distintos subconjuntos o zonas de la configuración inicial de la rejilla. Kohonen (2013) señala que, como toda la vecindad espacial alrededor del BMU en la rejilla se modifica a la vez y esta modificación se propaga de forma gradual sobre la misma, acorde a la FV, el grado de ordenamiento local aumenta con el número de iteraciones. Con este proceso, el SOM va asociando las neuronas de salida con grupos o patrones del conjunto de datos de entrenamiento.

Resumiendo el proceso, cada entrada estará asociada a un vector de peso al que modifica, haciendo que la red aprenda, y las salidas son función de esta modificación. El usar una función de vecindad para preservar las propiedades topológicas del espacio de entrada, hace a los mapas autoorganizados diferentes de otras redes neuronales no supervisadas.

METODOLOGÍA

Para realizar el estudio se utilizaron, como datos de entrada para el SOM, vectores del tipo (Resistividad, Cargabilidad, Metano) medidos en 107 localizaciones en el área. Los datos fueron medidos previamente y proporcionados para este trabajo. Por razones de confidencialidad, no se indica ni el área de estudio ni el grupo de trabajo que los cedió. Los datos se organizaron en una matriz de dimensión 107×3 . Se utilizó el módulo "nctool" de MATLAB para obtener los SOMs. Este módulo permite seleccionar, de forma manual, el número adecuado de neuronas para clasificar las variables del conjunto de datos. En este caso, el proceso de selección del número de neuronas para generar los mapas finales de SOM se llevó a cabo iterando el entrenamiento del mapa autoorganizado con distintos tamaños de red. Estos últimos fueron variados en la interfaz de la herramienta de clasificación con una topología hexagonal y utilizando 200 épocas para el entrenamiento en todos los casos. En este sentido, se probaron redes de 9×9 , 8×8 , 7×7 , 6×6 , 5×5 , 4×4 y 3×3 neuronas, por lo que el número final de estas corresponde a la multiplicación de las dimensiones respectivas. El criterio para seleccionar la dimensión adecuada para este conjunto de datos en particular estuvo basado en la observación de la cantidad de neuronas que resultaban activadas durante el entrenamiento, la cantidad de muestras contenidas en cada una de ellas, y en las distancias entre las vecindades de los pesos de la red. Esto se evaluó a partir de las dos salidas gráficas del SOM: la primera, muestra la disposición de los datos en las neuronas y la segunda la distancia entre las mismas, tal como se ejemplifica en las Figuras 1 y 2, respectivamente, para una red 4×4 . En el mapa de distancias, el color negro representa la distancia más lejana entre las neuronas y, por tanto, entre los datos contenidos en cada grupo asociado a un centro. El color amarillo corresponde a la distancia más cercana.

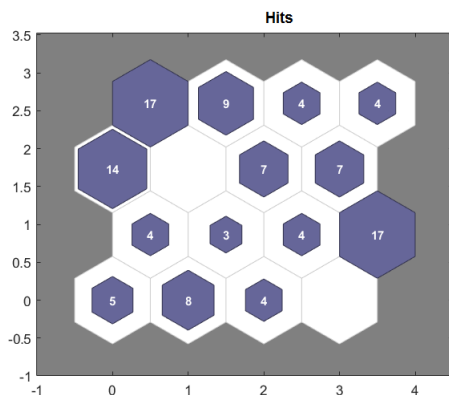


Figura 1. Ejemplo de disposición de los datos de entrada de este estudio, en las neuronas del mapa auto-organizado de una red 4×4 .

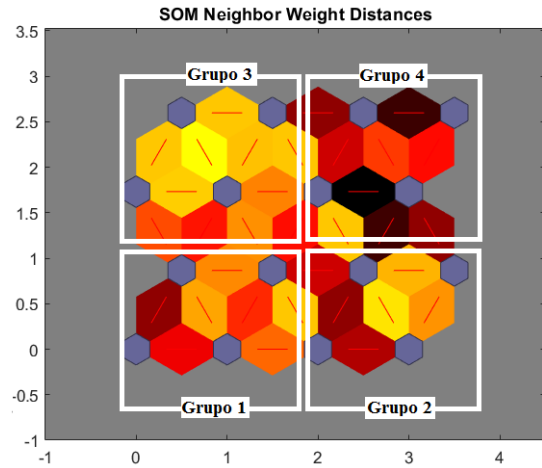


Figura 2. Ejemplo de mapa de distancias entre los pesos de la vecindad del SOM, con la disposición de los grupos interpretados, para una red 4×4 .

Una vez seleccionado el tamaño de red adecuado, se extraen los datos contenidos en cada una de las neuronas con una rutina adicional codificada en MATLAB. Al integrar esta información, se seleccionaron grupos de neuronas que presentaban un comportamiento similar en cuanto a cercanía, para obtener un mapa de contornos que mostrara la localización de dichos grupos, según las muestras contenidas en cada uno.

Además de ello, para poder realizar una comparación entre la clasificación obtenida y las diferentes anomalías asociadas a la resistividad, la polarización inducida y la concentración de metano en superficie, se generaron mapas adicionales para cada variable; estos mapas se muestran en las Figuras 3, 4, y 5, respectivamente.

Es importante resaltar que los puntos negros observados las Figuras 3, 4, y 5, y que se encuentran rodeados por los respectivos contornos, corresponden a los puntos de adquisición en campo.

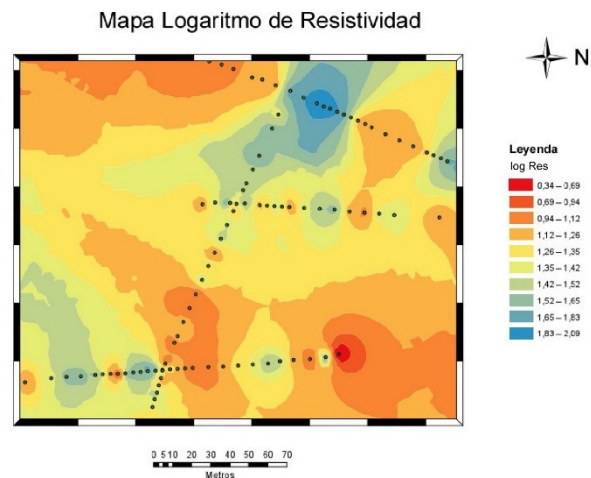


Figura 3. Mapa del logaritmo de la resistividad.

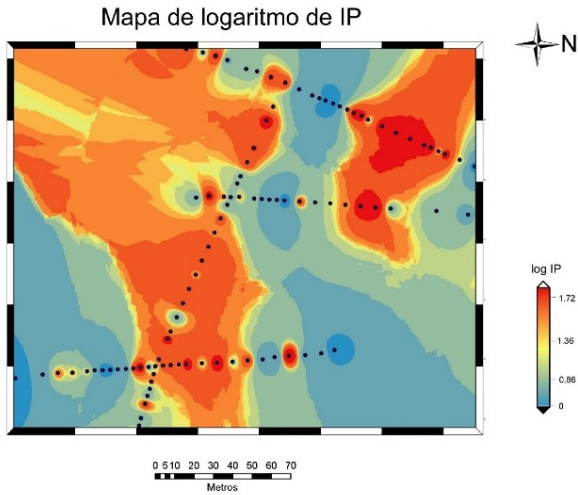


Figura 4. Mapa de logaritmo de polarización inducida.

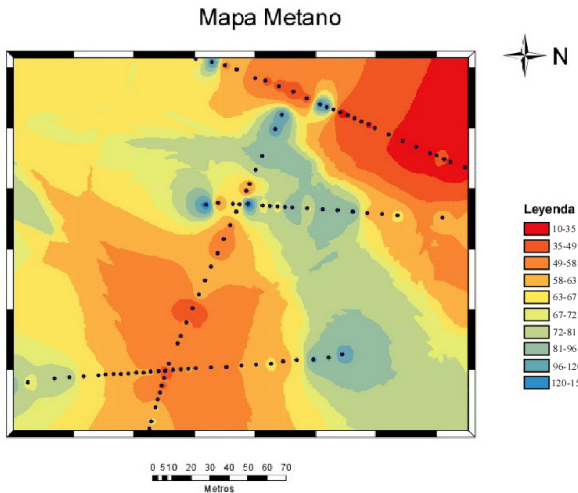


Figura 5. Mapa de logaritmo de concentración de metano, en ppm, en superficie.

RESULTADOS

La Figura 6 muestra los resultados obtenidos para los tamaños de red 9×9 , 6×6 y 3×3 . Para los tamaños de red entre 9×9 y 5×5 , se observaron varias neuronas desactivadas al agrupar los 107 datos ingresados y no se obtuvieron grupos considerables en el mapa de distancias entre las mismas. En el caso de la red 3×3 , aunque no hubo neuronas desactivadas, no fue posible diferenciar clústeres. La disposición de los datos para cada neurona en la red 4×4 se muestra en la Figura 1. Como se observa al comparar con la Figura 6, esta red es la que menos neuronas inactivas presenta (solo dos neuronas desactivadas). Además, en ella se puede identificar 4 grupos o clústeres de neuronas, utilizando la salida gráfica del SOM que muestra la distancia entre neuronas vecinas (Figura 2). Así, de acuerdo a los criterios previamente establecidos, se seleccionó la red 4×4 para el análisis de los datos.

La disposición de las muestras en grupos de neuronas, de acuerdo con esta selección, se puede ver en el mapa de contornos de la Figura 7. Para este último gráfico, los grupos 1 y 2 corresponden a la zona con los colores rojos a naranja y los grupos 3 y 4 a la zona de los tonos de azules.

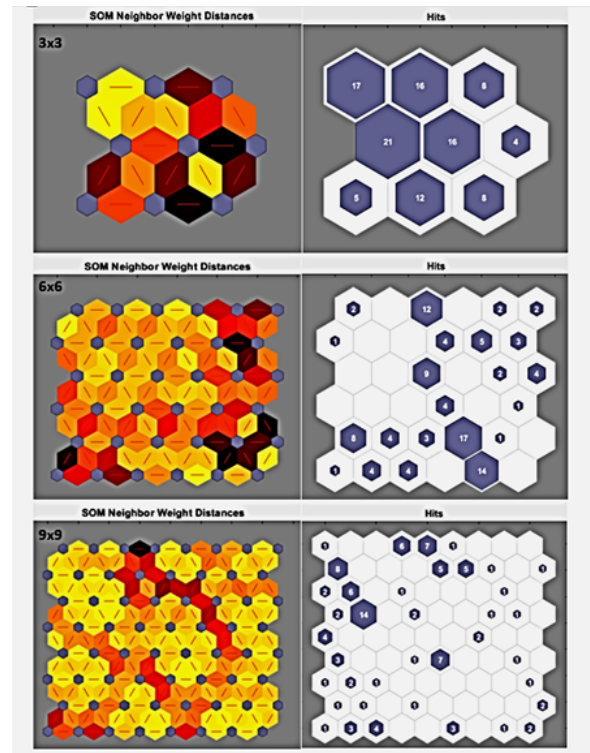


Figura 6. Mapas de distancias entre los pesos de la vecindad del SOM y de disposición de los datos de entrada en las neuronas, para las redes 3×3 , 6×6 y 9×9 , respectivamente.

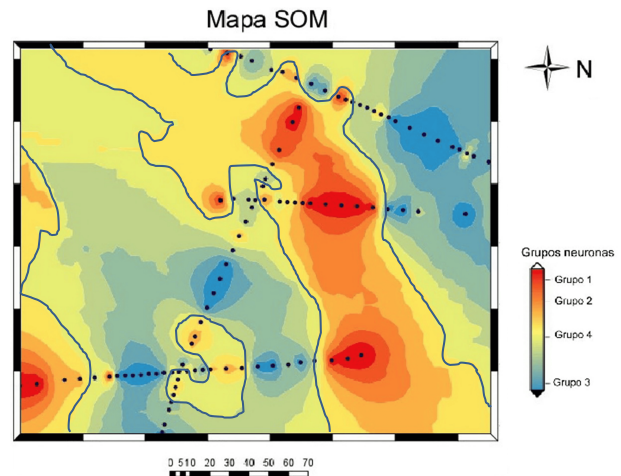


Figura 7. Mapa de contornos correspondientes a los 4 grupos identificados del análisis de SOM. Se encierran, con una línea azul, los grupos 1 y 2, asociados con la posible presencia de lixivados. Los rangos de logaritmo de resistividad, logaritmo de IP y concentración de metano (ppm) asociados a cada grupo, se presentan en la Tabla 1.

Según lo observado en la [Figura 7](#) y como se precisa en la [Tabla 1](#), los grupos 1 y 2 de neuronas coinciden con valores bajos del logaritmo de la resistividad, en un rango de 0,34 a 0,94, y a altos del logaritmo de IP y del logaritmo de metano, en un rango entre 1,36 y 1,72, y entre 81 a 159, respectivamente. Esto se detalla al comparar los mapas de SOM con los mapas de contornos de los parámetros individuales ([Figuras 3, 4 y 5](#)) y constatar con los datos contenidos en las neuronas. Así, los datos en los grupos 1 y 2 corresponden a bajos valores de resistividad y altos valores de cargabilidad y concentración de metano, que es una respuesta típica de líquidos percolados ([Abdulrahman et al., 2016](#)). Esto sugiere la existencia de una pluma de lixiviados en el vertedero en cuestión, cuyos acuíferos se encuentran a una profundidad aproximada de 25 m, y que se asocia a las zonas donde se disponen estos grupos de neuronas.

Tabla 1. Rangos de valores de los parámetros Logaritmo de Resistividad, Logaritmo de IP y concentración de metano (ppm), asociados a cada uno de los 4 grupos identificados en el SOM.

Grupo	Log(Res)	Log(IP)	Metano (ppm)
1	0,34 - 0,69	1,36 - 1,54	81 - 105
2	0,69 - 0,94	1,57 - 1,72	110 - 159
3	1,42 - 2,09	0,91 - 1,35	10 - 49
4	1,12 - 1,42	0,14 - 0,89	49 - 72

En el mapa de la [Figuras 7](#) se delimita la posible pluma de lixiviados encerrando, con una línea azul, los grupos 1 y 2 resultantes del SOM. Es posible observar una morfología alargada, bien definida, de dirección aproximada NO-SE, que correspondería a una pluma de lixiviados a la que designaremos como pluma principal. Hacia el S y SO se observan otras dos zonas posiblemente asociadas con lixiviados. Estas últimas dos zonas podrían ser, en principio, resultado de una de las dos razones siguientes:

Una posible explicación es que se originan por la migración de lixiviados, desde la pluma principal hacia el sur-suroeste del área, y se acumulan en esa zona, debido a las características del terreno; sin embargo, también podrían ser el resultado de la presencia de otra fuente de contaminantes hacia al SO, que da origen a una pluma diferente. No obstante, no se tiene información adicional del área que permita corroborar si estas zonas responden a la existencia de otra pluma o a la migración y acumulación parcial de lixiviados desde la pluma principal identificada.

Es importante notar que, si se analizan las variables de entrada (resistividad, polarización inducida o concentración de metano en superficie) por separado, se podrían delimitar ciertas zonas del vertedero como áreas afectadas por líquidos percolados. Sin embargo, algunas litologías pueden tener respuestas físicas similares a la de los lixiviados, dependiendo del parámetro medido, lo que pudiera conllevar a hacer una interpretación con cierto grado de error al analizar estos parámetros individualmente.

CONCLUSIONES

En el presente trabajo se trató de caracterizar la zona de un vertedero, integrando tres tipos de datos geofísicos (resistividad, cargabilidad y mediciones de concentración de gas metano), integrando la respuesta de estas variables utilizando una red no supervisada, tipo Kohonen, para generar mapas autoorganizados (SOM). Los SOM integraron y clasificaron, de manera efectiva, estos datos permitiendo delimitar posibles zonas afectadas por el vertedero y asociadas a la pluma de migración de los lixiviados. El reconocimiento de patrones con estos mapas, que integran la respuesta de distintas variables, es una ventaja del uso de SOMs para clasificar y reconocer tendencias en este tipo de estudios.

AGRADECIMIENTOS

Los autores agradecen a los Drs. María Jácome y Vincenzo Costanzo-Álvarez por el apoyo brindado.

REFERENCIAS

- ABDULRAHMAN, A., NAWAWI, M., SAAD, R., ABU-RIZAIZA, A. S., YUSOFF, M. S., KHALIL, A. E., & ISHOLA, K.S. (2016). Characterization of Active and Closed Landfill Sites Using 2D Resistivity/IP Imaging: Case Studies in Penang, Malaysia. *Environmental Earth Sciences*, 75(4), 1 - 17.
- CARTER-MCAUSLAN, A. & FARQUHARSON, C., (2020). Application of SOMs and K-means Clustering to Geophysical Mapping - Lessons Learned. *SEG Technical Program Expanded Abstracts*, 3843 - 3846.
- DOSSARY, M. AL-TURKI, A. & BADR, H. (2016). Self-Organizing Maps for Regions Exploring and Identification Based on Geological Signatures Similarities and Anomalies. *Memories SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*. Damman, Saudi Arabia.
- KOHONEN, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biol. Cybern.* 43, 59 - 69.
- KOHONEN, T. (2001). *Self-Organizing Maps* (Tercera Ed). Berlin: Springer.
- KOHONEN, T. (2013). Essentials of Self-organizing Map. *Neural Networks*. 37, 52 - 65.
- KOHLER, A., OHNBERGER, M., & SCHERBAUM, F. (2010). Unsupervised Pattern Recognition in Continuous Seismic Waverfield Records Using Self-Organizing Maps. *Geophysical Journal International*, 182(3), 1619 - 1630.

- MILJKOVIĆ, D. (2017). Brief Review of Self-Organizing Maps. *40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, 1061 - 1066.
- MORALES, C. (2018). Self-Organizing Maps: cómo aprovechar las estructuras sinápticas de nuestras neuronas para visualizar datos complejos. Recuperado de <https://www.linkedin.com/pulse/self-organizing-maps-o-cómo-aprovechar-las-sinapticas-morales-boada/?originalSubdomain=es>.
- OTA, K., AOKI, T., KURATA, K., & AOYAGI, T. (2011). Asymmetric neighborhood functions accelerate ordering process of self-organizing maps. *Physical Review E*, 83(2), 021903(1-9). DOI: <https://doi.org/10.1103/PhysRevE.83.021903>
- PENN, B. S. (2005). Using Self-Organizing Maps to Visualize High-Dimensional Data. *Computers and Geosciences*, 31(5), 531 - 544.
- SAKAO, T., & NERAMBALLI, A. (2020). A Product/Service System Design Schema: Application to Big Data Analytics. *Sustainability*, 12(8), 3484. DOI:<https://doi.org/10.3390/su12083484>.
- STRECKER, U. & UDEN R. (2002). Data Mining of 3D Poststack Seismic Attribute Volumes Using Kohonen Self-Organizing Maps. *The Leading Edge*, 21(10), 1032 - 1037.
- TORRECILLA, J., ROJO, E., DOMINGUEZ, J., & RODRIGUEZ, F. (2009). Self-Organizing Maps and Learning Vector Quantization Networks as Tools to Identify Vegetables Oils. *Journal of Agricultural and Food Chemistry*, 57(7), 2763 - 2769.
- VESANTO, J. (2000). Neural Network Tool for Data Mining: SOM Toolbox. In *Proceedings of symposium on tool environments and development methods for intelligent systems (TOOLMET 2000)*, 184 - 196.