

# Agrupamiento automático de datos magnéticos en prospección geofísica para arqueología

Manuel Ortiz Osio<sup>1</sup>, Erik Molino Minero Re<sup>2</sup>

<sup>1</sup> Universidad Nacional Autónoma de México,  
Posgrado en Ciencia e Ingeniería de la Computación,  
México

<sup>2</sup> Universidad Nacional Autónoma de México,  
Unidad Académica Yucatán,  
México

manuel.ortiz@ingenieria.unam.edu,  
erik.molino@iimas.unam.mx

**Resumen.** Uno de los objetivos de la geofísica es la interpretación de imágenes asociadas con variables físicas de estructuras presentes en el subsuelo. En este trabajo proponemos aplicar a los datos geofísicos una metodología de agrupamiento usando herramientas de aprendizaje computacional, con el fin de brindar una herramienta cuantitativa para la interpretación. Debido a que no siempre es posible conocer la distribución verdadera de las estructuras buscadas, se eligió un enfoque basado en el aprendizaje no supervisado, en donde los algoritmos  $k$ -medias,  $k$ -medianas y mapas auto-organizados fueron aplicados a levantamientos magnetométricos con enfoque arqueológico.

**Palabras clave:** Prospección geofísica, magnetometría, arqueología, interpretación, agrupamiento automático, aprendizaje no supervisado.

## Automatic Clustering of Magnetic Data in Geophysical Prospecting for Archeology

**Abstract.** The interpretation in geophysical prospecting is one of the main objectives, done by analyzing multiple images associated with physical properties of buried structures. In this paper we propose to apply a clustering methodology to geophysical data using machine learning tools, in order to provide a quantitative tool for interpretation. Since it is not always possible to know the true distribution of the buried structures, an approach based on unsupervised learning was chosen, where the algorithms  $k$ -means,  $k$ -medians and self-organizing maps were applied to magnetometric surveys with an archaeological approach.

**Keywords:** Geophysical prospecting, magnetic method, archeology, interpretation, automatic clustering, unsupervised learning.

## 1. Introducción

La interpretación de datos geofísicos provenientes de los métodos prospectivos consiste en el análisis de múltiples imágenes dentro de un dominio restringido. El objetivo es inferir una distribución de propiedades físicas que expliquen las variables registradas, y por ende, la información recolectada debe ser interpretada dentro de un marco delimitado por las características del objetivo y del contexto geológico, hidrogeológico, arqueológico, etc.

Las interpretaciones generalmente se realizan de forma cualitativa, describiendo geometrías y patrones, y las conclusiones dependen de la experiencia del intérprete y su habilidad de analizar datos multidimensionales.

Algunas de las imágenes usadas para realizar las interpretaciones son el resultado de un modelo inverso o de la aplicación de filtros de realce de anomalías geofísicas. El flujo de trabajo que se aplica de forma general a los levantamientos geofísicos se describe en la literatura, como en [7], y puede resumirse en los siguientes pasos:

1. Adquisición: los datos se registran durante un levantamiento geofísico. Deben adquirirse con la mayor calidad posible.
2. Correcciones y preprocesamiento: los datos de algunos métodos requieren ser preprocesados para remover efectos no deseados. Se puede aplicar una metodología de filtrado para atenuar artefactos dentro de las señales registradas.
3. Procesamiento: los datos se procesan para obtener las imágenes a interpretar.

La magnetometría consiste en realizar mediciones del campo geomagnético en la superficie terrestre. Su objetivo es determinar la distribución de las posibles estructuras que son fuente de los datos registrados, cada dato se atribuye a un punto en superficie.

El procesamiento general aplicado a datos de magnetometría puede consultarse en [13, 2, 14]. Algunos detalles relevantes para este trabajo se revisan en la Sección 2 de este documento. En este trabajo proponemos una metodología basada en algoritmos de aprendizaje no supervisado para agrupar datos magnéticos, ya que no siempre es posible conocer la distribución real de las anomalías en el subsuelo.

Las herramientas de aprendizaje no supervisado, dentro del área de aprendizaje computacional, son un conjunto de metodologías en las que los modelos generados se crean a partir de las entradas al sistema, desconociendo total o parcialmente las salidas esperadas.

El objetivo es diseñar una herramienta computacional que apoye las conclusiones de un intérprete, obteniendo a la salida una imagen que muestra la distribución espacial de los grupos o conjuntos resultantes de la aplicación de algoritmos basados en  $k$ -medias,  $k$ -medianas y mapas auto-organizados. Los levantamientos analizados provienen de estudios realizados en dos zonas arqueológicas en México: Xalasco, Tlaxcala y La Ferrería, Durango, Figura 1.

Las bases de datos contienen las variables de algunos de los filtros de realce de anomalías más usados: gradiente vertical, reducción al polo, señal analítica, gradiente horizontal, derivada inclinada y gradiente horizontal de la derivada inclinada, además de la anomalía de campo total.

La etapa de agrupamiento propuesta se añade en la posición 4 al flujo de trabajo geofísico descrito más arriba, después de la generación de las imágenes finales. El agrupamiento resultante muestra una posible distribución de las estructuras que producen la respuesta observada en superficie, a través de las similitudes entre vectores y entre bases de datos.

Los resultados se validan correlacionándolos con los estudios arqueológicos previos liderados por interpretaciones de expertos en el área, mostrando que esta metodología puede aplicarse para apoyar a la interpretación geofísica. La implementación, escrita en scripts en python, se encuentra disponible en [11].

### **1.1. Revisión de literatura**

Con la finalidad de reducir la incertidumbre en la interpretación de datos geofísicos multidimensionales, algunos autores han desarrollado métodos como la denominada inversión conjunta [10]. En este trabajo los autores asumen que dos métodos espacialmente coincidentes deben ser sensibles a la misma geología. Desarrollaron una relación de gradientes cruzados para evaluar características estructurales existentes en ambos métodos.

Por otro lado, se han aplicado técnicas de aprendizaje computacional para la clasificación automática de litología. En [16] se realiza una revisión de herramientas de aprendizaje automático aplicadas a datos de percepción remota. Los autores estudiaron cinco conjuntos de metodologías: reducción de dimensionalidad, clasificación, agrupamiento, regresión y aprendizaje profundo.

Cada conjunto estudiado parece ser una buena alternativa para procesar datos de percepción remota, resolviendo inconvenientes asociados con el mapeo de estructuras geológicas de gran escala y que pueden ser verificadas. En [9], los autores hicieron un mapeo predictivo de enjambres de diques al NE de Brazil, aplicando un algoritmo de SOM a datos de magnetometría aérea y de espectrometría de rayos gamma.

Esta técnica semiautomática permite una clasificación de unidades litológicas, los resultados fueron correlacionados con trabajo de campo. Otros autores han desarrollado herramientas computacionales para automatizar, o semi-automatizar, el mapeo geológico y litológico mediante la aplicación de bosques aleatorios. En [8] se usan estas metodologías para clasificar estructuras de gran escala, cuya existencia puede ser verificada.

## **2. Teoría**

### **2.1. Geofísica**

De acuerdo con [15], la geofísica aplicada consiste en la medición e interpretación de las propiedades físicas de la Tierra para determinar las condiciones del subsuelo. Los métodos geofísicos responden a los contrastes de las propiedades físicas de estructuras enterradas. Ningún método concluye con una interpretación única, por lo que es recomendable aplicar más de un método, o calcular más de un atributo, para facilitar la detección de la anomalía buscada, información al respecto se puede consultar en [3].

Es de suma importancia que los datos sean interpretados dentro de un marco definido por las características de la zona de estudio. La interpretación se lleva a cabo analizando múltiples imágenes que pertenecen al mismo dominio espacial.

Estas imágenes pueden provenir de diferentes métodos o de la aplicación de filtros de realce de anomalías. La interpretación normalmente es un proceso cualitativo, por lo que puede haber múltiples interpretaciones para un mismo levantamiento.

La magnetometría [1] es un método de exploración geofísica basado en la medición del campo geomagnético con el objetivo de inferir la distribución de cuerpos o estructuras magnetizables. Las lecturas se realizan en la superficie terrestre con magnetómetros, donde cada dato se atribuye a un punto en la superficie. Los magnetómetros de campo total miden la magnitud del campo geomagnético, incluyendo las contribuciones de campos internos y externos.

Para estudiar las anomalías locales es necesario retirar las contribuciones del campo principal y de los campos externos. Este proceso se realiza numéricamente en un preprocesamiento. Comúnmente, para interpretar datos provenientes de levantamientos magnéticos, se analizan las imágenes resultantes de la aplicación de los siguientes filtros:

- **Reducción al polo:** Cambia matemáticamente el ángulo de inclinación del campo geomagnético, obteniendo entonces la respuesta de las anomalías a una magnetización vertical.
- **Gradiente vertical:** Define la tasa de cambio vertical del campo geomagnético.
- **Gradiente horizontal:** Es el módulo de las derivadas en dirección  $x$  y en dirección  $y$ .
- **Señal analítica:** Se puede ver como la combinación de los gradientes vertical y horizontal, calculada mediante la transformada de Hilbert.
- **Derivada inclinada:**  $TDR = \tan^{-1} \left( \frac{GV}{GH} \right)$ .
- **Gradiente horizontal de la derivada inclinada:** cuyo cálculo es equivalente al del gradiente horizontal.

Debido a la complejidad de los mapas magnéticos resultantes, la interpretación se lleva a cabo de forma cualitativa analizando patrones, continuidad y geometría de las anomalías resaltadas, características que dependen del filtro aplicado.

Los datos pueden ser interpretados como perfiles individuales (curvas en dos dimensiones) o como mapas (imágenes en dos dimensiones cuyos ejes son las coordenadas y el valor del campo se representa en falso color). Algunos ejemplos sobre la aplicabilidad del método en arqueología se pueden consultar en [5, 6].

## 2.2. $K$ -medias y $K$ -medianas

Los algoritmos de  $k$ -medias y  $k$ -medianas agrupan los datos usando la distancia como métrica para estimar la similitud entre los vectores de entrada. Estos algoritmos iteran para actualizar la posición de centroides considerando la media o mediana de las distancias de sus vectores más cercanos. Los grupos resultantes están definidos por los vectores más cercanos a cada centroide.



(a) Localización de Xalasco.

(b) Localización de La Ferrería.

**Fig. 1.** Ubicación de las zonas de estudio, resaltadas por un cuadro azul.

La base de datos con vectores agrupados muestra entonces los patrones generales de acuerdo a las similitudes inferidas a partir del cálculo de la media o mediana de distancias, en este trabajo usamos la implementación de scikit-learn [12].

### 2.3. Mapas auto-organizados

Los Mapas Auto-Organizados, también conocidos como SOM (por su nombre en inglés), son una herramienta de aprendizaje no supervisado que se usa para mapear y visualizar datos multidimensionales en un espacio comúnmente bidimensional, preservando su estructura topológica.

Cada neurona, elemento del mapa de la SOM, tiene un vector de pesos con la misma dimensión que el espacio de características de los datos. Los pesos de las neuronas se actualizan considerando la distancia entre los vectores de entrada y sus respectivas neuronas ganadoras ( Best Matching Unit - BMU), así como de la vecindad y de una tasa de aprendizaje. La elaboración matemática se puede consultar en [4].

Esta herramienta puede usarse para agrupar datos, donde los grupos resultantes son definidos por las BMU. Comparado con los algoritmos de  $k$ -medias y  $k$ -medianas, este método ofrece un agrupamiento más robusto, añadiendo la sensibilidad a neuronas vecinas. En este trabajo usamos la implementación minisom [18].

## 3. Metodología propuesta

### 3.1. Configuración de los datos

En este paso se eligen las variables que se usarán como entrada. La selección de estas depende de las necesidades del intérprete, pudiendo discriminar entre datos de uno o varios sensores con sus respectivos atributos calculados. Nuestro sistema usa el algoritmo de PCA para reducir la dimensionalidad en las bases de datos. Las variables elegidas para cada conjunto pueden ser aquellas que tengan un comportamiento similar, por ejemplo datos que provienen de sensores independientes. Con este procedimiento se busca disminuir la colinealidad entre variables, de acuerdo con las características

**Tabla 1.** Enumeración de las variables magnéticas usadas para entrenar y agrupar.

Identificador	Variable	Sensor
1	Campo magnético total	Superior
2	Campo magnético total	Inferior
3	Gradiente vertical	N/A
4	Reducción al polo	Superior
5	Reducción al polo	Inferior
6	Gradiente horizontal	Superior
7	Gradiente horizontal	Inferior
8	Señal analítica	Superior
9	Señal analítica	Inferior
10	Derivada inclinada	Superior
11	Derivada inclinada	Inferior
12	Gradiente horizontal de la derivada inclinada	Superior
13	Gradiente horizontal de la derivada inclinada	Inferior

físicas del levantamiento (cuando se usan dos o más sensores para la adquisición) o de las propiedades de los filtros de realce (pudiendo existir una dependencia prácticamente lineal entre sus salidas). Por otro lado, los datos son escalados antes de la aplicación de los algoritmos de aprendizaje computacional.

### 3.2. Entrenamiento y agrupamiento

Con el fin de aumentar el desempeño de las herramientas de aprendizaje no supervisado se aplicó un proceso extra para cada caso.

**Reentrenamiento para  $k$ -medias y  $k$ -medianas.** Estos métodos pueden enmascarar algunas características en los grupos resultantes. Los datos de magnetometría tienden a agruparse en dos conjuntos principales: los puntos con valores positivos y los puntos con valores negativos.

Añadir más centroides al entrenamiento no mejora la resolución en los conjuntos de salida, tendiendo siempre a los mismos dos conjuntos generales y conjuntos adicionales que describen los vectores atípicos. Para evitar el efecto de sobre-ajuste usamos el siguiente algoritmo:

1. Entrenar  $c$  centroides usando el conjunto de entrenamiento completo.
2. Retirar del conjunto de entrenamiento los vectores pertenecientes al centroide más poblado.
3. Entrenar  $c$  nuevos centroides usando el conjunto de entrenamiento filtrado resultante del punto anterior.

Con este algoritmo evitamos que los centroides se ubiquen únicamente en las porciones que definen la anomalía principal, pudiendo ahora discriminar estructuras de transición al efecto regional.

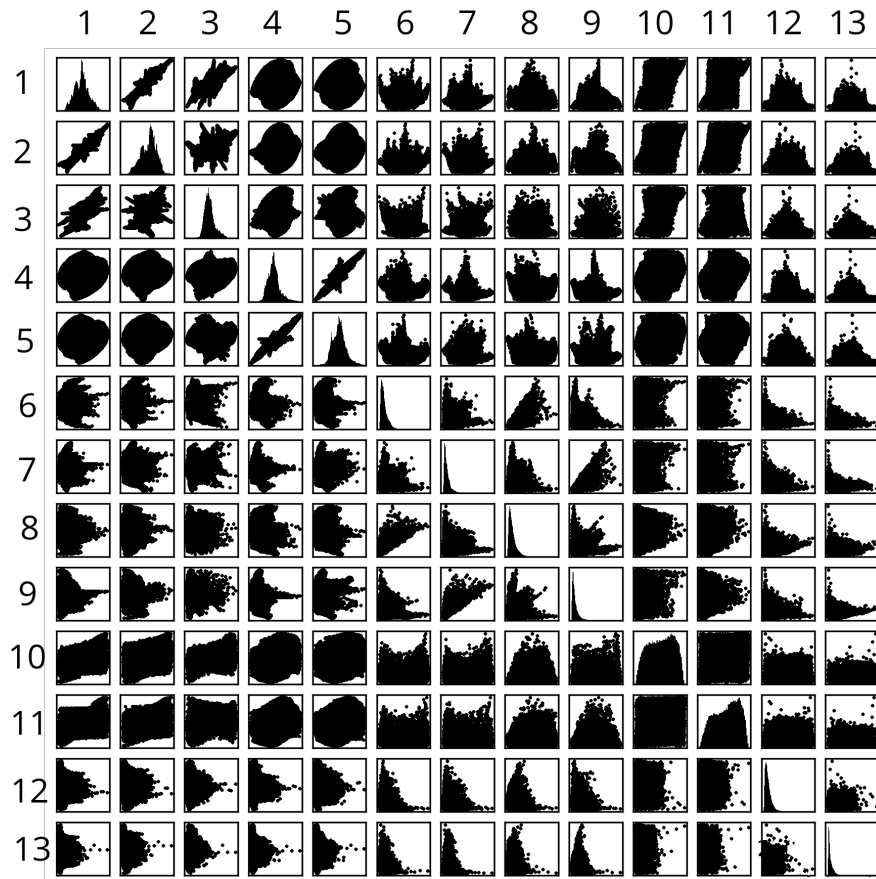
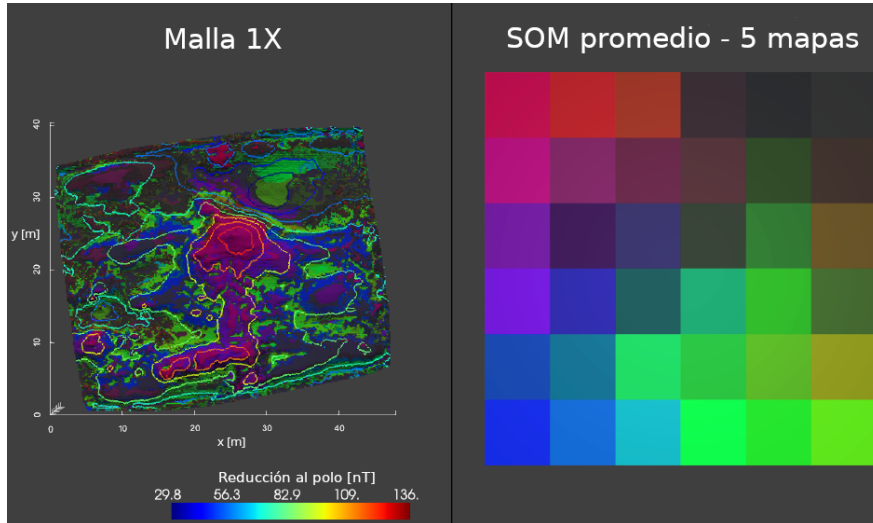


Fig.2. Matriz de dispersión de la base de datos, en la diagonal se muestra el histograma de cada variable.

**Promedio de  $m$  SOMs.** Los SOM dependen de la inicialización de las neuronas, en consecuencia no siempre los mismos vectores se agruparán en la misma neurona en cada nuevo mapa. Para aumentar la sensibilidad de los SOM, buscando distinguir grupos que pueden comportarse como transición a otros grupos, usamos el siguiente algoritmo:

1. Entrenar  $m$  mapas con inicialización aleatoria independiente.
2. Usamos el algoritmo de  $k$ -medoides para agrupar las neuronas en  $c$  grupos de acuerdo con la similitud entre sus vectores de peso, el número de grupos es elegido por el usuario. A cada conjunto de neuronas se le asigna de forma automática un matiz o color, a partir del cual las neuronas que pertenecen al mismo grupo se diferencian entre sí presentando una saturación de color diferente. Cada vector del conjunto de agrupamiento tendrá una BMU caracterizada por un color: los vectores que tengan el mismo matiz tendrán características similares, siendo del mismo grupo aquellos que además presenten la misma saturación.



**Fig.3.** Imagen de una aplicación del sistema propuesto a la malla 1 de Xalasco, Tlaxcala. Derecha: SOM obtenida tras la aplicación del algoritmo de entrenamiento propuesto. Izquierda: mapa del levantamiento que muestra las estructuras principales inferidas del agrupamiento aplicado, se sobrepone la variable de reducción al polo como mapa de contornos.

3. Para cada vector del conjunto de agrupamiento se computa el color promedio de sus BMU en los  $m$  mapas. En este punto se diferencian los vectores que tienen BMUs con matices y saturaciones diferentes.
4. Se entrena un nuevo SOM con los colores obtenidos en el paso anterior. El mapa resultante es usado para agregar la etiqueta de color a los vectores del conjunto de agrupamiento. Con este proceso se reduce la paleta de colores y en consecuencia simplificamos las estructuras resaltadas por las similitudes encontradas en la base de datos.

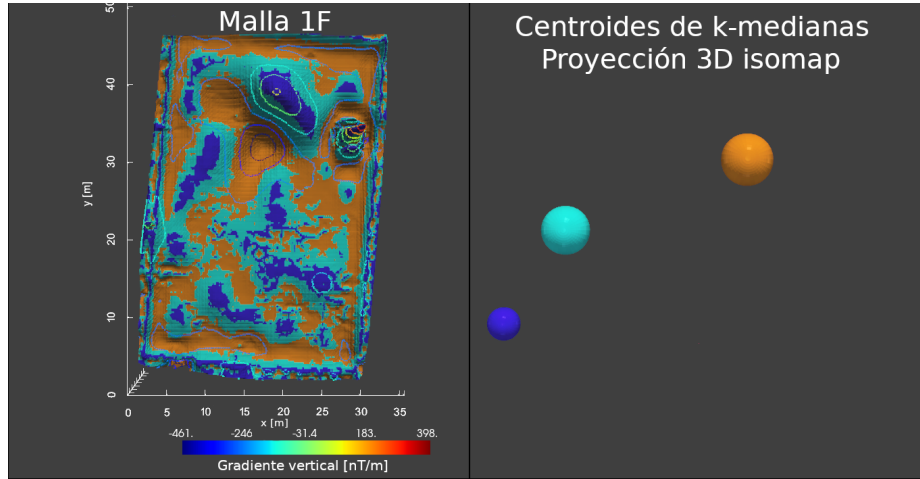
### 3.3. Generación de imágenes

En esta etapa usamos la implementación de pyvista [17] para generar la figura final, esta se divide en dos áreas:

1. Distribución de los grupos o el SOM con etiquetas de color para cada elemento.
2. Un mapa con la distribución espacial del levantamiento de la base de datos agrupada. El color de cada píxel corresponde unívocamente al grupo al que pertenece de acuerdo al punto anterior. La continuidad en el color de los píxeles en esta imagen describe las posibles estructuras, debido a que se puede atribuir como la persistencia espacial de las similitudes en la base de datos.

**Gráfico de distribución de centroides.** El número de centroides es elegido por el usuario en la etapa de entrenamiento, cada uno existe en un espacio de dimensión igual al número de características de la base de datos.





**Fig. 4.** Imagen de una aplicación del sistema propuesto a la malla 1 de La Ferrería, Durango. Derecha: distribución de centroides proyectados de un espacio de siete dimensiones a un espacio de tres dimensiones. Izquierda: mapa del levantamiento que muestra las estructuras principales inferidas del agrupamiento aplicado, se sobrepone la variable de gradiente vertical como mapa de contornos.

Para visualizar la distribución de los centroides resultantes aplicamos el algoritmo de reducción de dimensionalidad isomap, buscando representar a cada centroide como una esfera en un espacio tridimensional, cuyo radio es proporcional al número de vectores que pertenecen al centroide respectivo, el color que define a cada esfera se asigna de forma automática.

El usuario puede cambiar el nivel de acercamiento y el ángulo de visualización de los centroides, con el fin de analizar las posiciones relativas entre ellos y concluir sobre su similitud o cercanía.

**Gráfico del SOM.** El algoritmo implementado construye una malla rectangular o hexagonal de tamaño  $n \times m$ , de acuerdo a la topología ingresada por el usuario. Después de la asignación del color final se construye la malla de forma automática.

**Gráfico del mapa.** Aquí se sustituye el valor de las variables, cuyas coordenadas están definidas por los puntos de adquisición del levantamiento magnetométrico, por un píxel cuyo color corresponde al grupo determinado por los métodos de agrupamiento.

En esta imagen se muestra el resultado del agrupamiento de los datos analizados, la continuidad y geometría de los elementos del mismo color pueden interpretarse como elementos de una misma estructura, ya que estos píxeles tienen un comportamiento similar.

Los colores deben ser interpretados junto con la distribución de los centroides o junto con el SOM: la definición de las estructuras resaltadas depende de la distribución de los centroides o de las neuronas. Para apoyar las conclusiones del intérprete añadimos la opción de añadir un mapa de contornos sobre el mapa con vectores agrupados (e.j. componente total, reducción al polo, gradiente vertical, etc).

## **4. Base de datos**

La metodología descrita anteriormente se aplicó a datos reales, en esta sección se describirá de forma breve la ubicación de los levantamientos y el comportamiento de las variables que se usaron para entrenar y agrupar.

### **4.1. Zonas de prueba**

Las bases de datos usadas en este trabajo provienen de dos sitios arqueológicos: Xalasco, Tlaxcala (figura 1a) y La Ferrería, Durango (figura 1b), en México. Contamos con seis mallas de magnetometría para cada localidad, datos que fueron brindados por el departamento de geofísica de la Facultad de Ingeniería de la UNAM.

En todos los levantamientos se usó un gradiómetro, este instrumento puede definirse como un magnetómetro con dos sensores. Se tienen entonces dos lecturas de componente de campo total, dos componentes resultantes de cada filtro de realce de anomalías (uno para cada sensor) y la lectura directa del gradiente vertical. Cada malla fue corregida y preprocesada apropiadamente.

### **4.2. Configuración de los datos**

La figura 2 muestra los gráficos de dispersión e histogramas de una muestra aleatoria del 60 % de la base de datos completa, las variables se enumeran en el cuadro 1. Los histogramas se encuentran bien distribuidos y los gráficos de dispersión son uniformes sin mostrar dependencias fuertes, excepto para las componentes de campo total y la reducción al polo en ambos sensores.

Al contar con dos versiones para cada variable (excepto para el gradiente vertical), contamos con una base de datos con 13 características y de 1,416,808 vectores considerando las 12 mallas.

## **5. Resultados y discusión**

Se aplicó la metodología propuesta en la sección 3 a la base de datos descrita en la sección 4. El proceso se repitió varias veces con distintas configuraciones de parámetros, se recurrió a un análisis cualitativo para elegir la mejor imagen. Las características consideradas fueron:

1. Detalle: la imagen muestra elementos que pueden ser interpretados como zonas de transición de las estructuras principales, estas últimas definidas por su geometría, relacionándose con el objetivo del estudio.
2. Definición de anomalías geofísicas: la geometría de las estructuras visualizadas tienen una tendencia espacial clara.
3. Paleta de colores: los colores de la imagen resultante tienen un buen contraste que facilita la visualización.

4. Distribución de los centroides o estructura del SOM: para los algoritmos  $k$ -medias y  $k$ -medianas, los centroides resultantes no se enciman; para SOM, la u-matrix (matriz de distancias unificada) y el hit-map (mapa de aciertos) tienen un comportamiento homogéneo.
5. Correlación con evidencia arqueológica: las estructuras definidas por el agrupamiento tienen correlación directa con estudios arqueológicos previos.

### **5.1. Malla 1 Xalasco, Tlaxcala**

La figura 3 muestra el resultado de un agrupamiento, usando el algoritmo SOM, aplicado a la malla 1 de Xalasco, nombrada en este trabajo como 1X. Los parámetros con los que se obtuvo el mejor resultado fueron:

- El entrenamiento se realizó sobre las variables del sensor inferior y el gradiente vertical, ya que los objetivos arqueológicos son someros. Se eligió de forma aleatoria al conjunto de entrenamiento, conformada por el 60 % de la base de datos completa.
- Relativo al SOM:
  - Se usó una red rectangular de  $6 \times 6$ , con la que se obtuvieron errores bajos y u-matrix con grupos claramente definidos. El SOM resultante se obtuvo a partir de cinco SOM con diferentes inicializaciones, en los que sus neuronas fueron respectivamente agrupadas en 3 conjuntos con tonos rojos, verdes y azules.
  - Se realizaron 5000 iteraciones, obteniendo con esto errores topológicos y de cuantización de 25 % y de 0.1 respectivamente.
  - Radio de vecindad del 30 %.
  - Tasa de aprendizaje de 0.1.

Respecto a la validación cualitativa:

- Detalle: se aprecian zonas de transición coherentes que delimitan las estructuras principales.
- Definición de anomalías geofísicas: los grupos resultantes preservan geometrías continuas y patrones estructurados.
- Paleta de colores: los colores principales son azul, lila y verde, la distribución de tonos resulta en una imagen con suficiente contraste para ser analizada.
- Estructura del SOM: los grupos resultantes de la aplicación del SOM se encuentran bien distribuidos, siendo capaz de definir de forma clara cuatro conjuntos principales.
- Correlación arqueológica: de acuerdo con [6] las excavaciones revelan la existencia de una estructura antropogénica al centro de la malla, corresponde con un suelo fragmentado con orientación norte-sur. Esta estructura tiene correlación con los elementos definidos por los tonos lila.

De acuerdo con la reducción al polo, los tonos lila y azul corresponden a la anomalía principal, que de acuerdo al SOM son elementos muy diferentes a los elementos definidos por los tonos verdes.

La metodología de agrupamiento distingue dos estructuras principales, definidas por los cambios de lila a azul. Existe correlación con evidencia arqueológica mostrando discontinuidades que pueden atribuirse a la destrucción del suelo como resultado de actividad humana contemporánea.

## 5.2. Malla 1 La Ferrería, Durango

La figura 4 muestra el resultado de un agrupamiento, usando el algoritmo  $k$ -medias, aplicado a la malla 1 de La Ferrería, nombrada en este trabajo como 1F. Los parámetros con los que se obtuvo el mejor resultado fueron:

- El entrenamiento se realizó sobre las variables del sensor inferior y el gradiente vertical, ya que los objetivos arqueológicos son someros. Se eligió de forma aleatoria al conjunto de entrenamiento, conformada por el 60 % de la base de datos completa.
- Relativo al algoritmo de  $k$ -medias:
  - Se entrenó el sistema con tres centroides, esperando un modelo con tres estructuras principales: una anomalía principal, una zona de transición y una estructura con comportamiento regional.
  - Los índices de Davies-Bouldin y de Silhouette resultantes de este procedimiento fueron 1.5 y 0.18 respectivamente.
- Respecto a la validación cualitativa:
  - Detalle: a pesar de que  $k$ -medias es menos robusto que SOM, la imagen resultante muestra estructuras continuas sin perder continuidad por elementos de una clase diferente con comportamiento aparentemente aleatorio. Este comportamiento probablemente se deba a que las estructuras reales en el subsuelo son más profundas.
  - Definición de anomalías geofísicas: el agrupamiento resultante define de forma clara tres estructuras diferentes que corresponden a la anomalía principal, a una anomalía con comportamiento regional y a una estructura de transición entre ambas.
  - Paleta de colores: la imagen agrupada se describe con tres colores: azul, cian y naranja, cuyo contraste es suficiente para analizar las estructuras.
  - Distribución de los centroides: de la proyección con isomap se concluye que los centroides no se superponen entre sí, siendo elementos que pueden describir el fenómeno analizado. Gracias a este entorno interactivo podemos concluir que la inferida estructura de transición corresponde a los colores del centroide cian, ya que se localiza entre los otros dos.
- Correlación arqueológica: aún no hay evidencia en trabajos publicados.

De acuerdo con el gradiente vertical, se puede atribuir que el color azul corresponde a la anomalía principal. La estructura conformada por estos elementos presentan alineaciones con orientación Sur-Norte hacia el oeste, y oeste-este hacia el sur.

Los resultados obtenidos en ambas zonas de estudio muestran que el sistema planteado en este trabajo es capaz de resaltar patrones en las bases de datos, pudiendo correlacionarse con evidencia de excavaciones arqueológicas.

La continuidad en las similitudes de los vectores que conforman las mallas de magnetometría se interpretan como las estructuras que son fuente de este método geofísico, cuyas geometrías y diferencias entre sí (identificadas por los cambios en tono o color) comprenden desde propiedades físicas diferentes (e.j. permeabilidad magnética) hasta la relación con la profundidad.

## **6. Conclusiones**

La metodología propuesta es capaz de generalizar las estructuras y patrones en las bases de datos analizadas. Las imágenes resultantes muestran estructuras que resultan lógicas de acuerdo con el objetivo. Las características resaltadas, producto del agrupamiento realizado con herramientas de aprendizaje no supervisado, pueden ayudar a las conclusiones de un intérprete, quien añade el contexto físico al resultado de la metodología aquí propuesta.

Los resultados del agrupamiento muestran correlación con las excavaciones realizadas en la zona de Xalasco, brindando un resultado cuantitativo para apoyar en la decisión de áreas para excavaciones futuras. Respecto a la zona de La Ferrería, no hay información sobre excavaciones en los sectores estudiados, sin embargo se pueden observar alineaciones que pueden ser lógicas de acuerdo a las estructuras antropogénicas esperadas.

Los algoritmos  $k$ -medias y  $k$ -medianas producen resultados que muestran patrones muy generales, comportamiento que es esperado cuando las estructuras buscadas no son superficiales, estos algoritmos tienden a enmascarar estructuras pequeñas.

Por otro lado, el SOM es capaz de detectar cambios sutiles, comportamiento que es esperado cuando las estructuras buscadas son someras, sin embargo tiende a presentar un comportamiento ruidoso cuando las estructuras son profundas. Ambas metodologías son aplicables para apoyar a la interpretación geofísica, aplicadas de forma conjunta pueden reducir la incertidumbre en la toma de decisiones.

En este trabajo se analizan bases de datos multidimensionales provenientes únicamente de datos magnéticos, sin embargo no se descarta que la metodología descrita tenga un buen desempeño al mezclar datos de distintas prospecciones, siendo cada una sensible a distintas propiedades de la materia.

Si los distintos levantamientos geofísicos son espacialmente coincidentes, es lógico pensar que la respuesta de cada uno es sensible a la existencia de una distribución de propiedades fija, la metodología de agrupamiento buscaría entonces la similitud entre cada distribución asumiendo que provienen de un mismo fenómeno en el subsuelo, siendo entonces una alternativa a la inversión conjunta.

Como propuesta de trabajo futuro se propone la implementación de este sistema que reciba datos de distintas prospecciones a la entrada, además de la posibilidad de generar modelos a partir de datos sintéticos para añadir una etapa de validación adicional.

**Agradecimientos.** Investigación realizada gracias al Programa PAPIIT - UNAM IG101222.

## Referencias

1. Blakely, R. J.: Potential theory in gravity and magnetic applications, Cambridge university press (1996)
2. Bournas, N., Bake, H. A.: Interpretation of magnetic anomalies using the horizontal gradient analytic signal. *Annals of Geophysics*, vol. 44, no. 3 (2009) doi: 10.4401/ag-3572
3. Butler, D. K., Bennett, H. H., Ballard, J. H.: Overview of multimethod geophysical system development for enhanced near-surface target detection, discrimination, and characterization. *The Leading Edge*, vol. 25, no. 3, pp. 352–356 (2006) doi: 10.1190/1.2184105
4. Davydov, A., Larionov, A., Nagul, N.: The formal description of discrete-event systems using positively constructed formulas. In: 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (2017) doi: 10.23919/mipro.2017.7973599
5. Fassbinder, J. W. E.: Magnetometry for archaeology. *Encyclopedia of Geoarchaeology*, pp. 499–514 (2017) doi: 10.1007/978-1-4020-4409-0\_169
6. Juárez, K., López-García, P., Argote-Espino, D. L., Tejero-Andrade, A., Chávez, R. E., García-Serrano, A.: Magnetic and electrical prospections in the archaeological site of Xalasco northeast of Tlaxcala, Mexico. *Global Journal of Archaeology and Anthropology*, vol. 2, no. 2, pp. 555–581 (2017) doi: 10.19080/GJAA.2017.02.555581
7. Khesin, B. E., Alexeyev, V. G., Eppelbaum, L. V.: Interpretation of geophysical fields in complicated environments. *Springer Science and Business Media*, vol. 14 (2013) doi: 10.1007/978-94-015-8613-9
8. Kuhn, S., Cracknell, M. J., Reading, A. M.: Lithologic mapping using random forests applied to geophysical and remote-sensing data: A demonstration study from the eastern goldfields of Australia. *Geophysics*, vol. 83, no. 4, pp. B183–B193 (2018) doi: 10.1190/geo2017-0590.1
9. Melo, A. C. C., de Castro, D. L., Fraser, S. J., Filho, A. A. M.: Using self-organizing maps in airborne geophysical data for mapping mafic dyke swarms in NE Brazil. *Journal of Applied Geophysics*, vol. 192, pp. 104377 (2021) doi: 10.1016/j.jappgeo.2021.104377
10. Oliver-Ocaño, F., Gallardo, L., Romo-Jones, J., Pérez-Flores, M.: Structure of the cerro prieto pull-apart basin from joint inversion of gravity, magnetic and magnetotelluric data. *Journal of Applied Geophysics*, vol. 170, pp. 103835 (2019) doi: 10.1016/j.jappgeo.2019.103835
11. Ortiz, M.: Interpretar geofísica. (2023) [github.com/CecilRamza/Interprete\\\_geofisica](https://github.com/CecilRamza/Interprete\_geofisica)
12. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, vol. 12, pp. 2825–2830 (2011)
13. Rajagopalan, S.: Analytic signal vs. reduction to pole: Solutions for low magnetic latitudes. *Exploration Geophysics*, vol. 34, no. 4, pp. 257–262 (2003)
14. Salem, A., Williams, S., Fairhead, D., Smith, R., Ravat, D.: Interpretation of magnetic data using tilt-angle derivatives. *Geophysics*, vol. 73, no. 1, pp. L1–L10 (2008) doi: 10.1190/1.2799992
15. Sheriff, R. E.: *Encyclopedic dictionary of applied geophysics*. Society of exploration geophysicists, pp. 442 (2002) doi: 10.1190/1.9781560802969

16. Shirmard, H., Farahbakhsh, E., Müller, R. D., Chandra, R.: A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, vol. 268, pp. 112750 (2022) doi: 10.1016/j.rse.2021.112750
17. Sullivan, C., Kaszynski, A.: Pyvista: 3d plotting and mesh analysis through a streamlined interface for the visualization toolkit (vtk). *Journal of Open Source Software*, vol. 4, no. 37, pp. 1450 (2019) doi: 10.21105/joss.01450
18. Vettigli, G.: MiniSom: Minimalistic and numpy-based implementation of the self organizing map (2018) [github.com/JustGlowing/minisom/](https://github.com/JustGlowing/minisom/)