Application of K-means algorithm to Werner deconvolution solutions for depth and image estimations

Daniel Eshimiakhe*, Kola Lawal

Department of Physics, Ahmadu Bello University, Nigeria

ABSTRACT

One of the most popular techniques for computer-assisted solution estimates for magnetics and gravity field data is Werner deconvolution. The approaches frequently produce erratic results and may not always forecast the maximum number of the geologic entity that produces them due to the intrinsic instability of potential field data. This led to the application of the K-means machine learning algorithm to further enhance the detection of the geologic potential field-generated bodies. Two substances that resembled dikes were combined to form a synthetic magnetic model. Random noise was added to the synthetic data, to make the solutions a bit more complex. Werner deconvolution technique was applied to the synthetic model to generate solutions. K-means unsupervised machine learning algorithm was applied to the generated solutions created by the synthetic data. We further applied this algorithm to real data sets from a mining site. The clustering result shows a good spatial correspondence with the geologic model, and the method was able to estimate the precise location and depth of the dike bodies. The proposed method is entirely data-driven and has proven to work in the presence of noise.

1. Introduction

The Earth System is subject to a variety of geohazards that are not properly understood, but they are having significant impacts as our exposure and insight continue to increase through increased urbanization. Mapping of geological features is an area of research made possible by the development of large-scale, high-resolution geophysical data collection, and one of these geophysical techniques is the potential field. The objective of the magnetic survey is to investigate anomalous geological features, that may alter the observed magnetic field. Because the magnetic properties of the underlying rocks are different, there are fluctuations in the magnetic field. Numerous magnetic interpretation techniques have been created to ascertain the depth of diverse geometric shapes' geological structures.

The strategies are based on: (i) graphical methods utilizing some distinctive features of the magnetic profile (Koulomzine et al., 1970; Am, 1972; Subrahmanyam and Prakasa Rao, 2009); (ii) spectral analysis techniques (Prakasa Rao et al., 1986); and (iii) numerical techniques such as the Werner Deconvolution method (Hartman et al., 1971; Ku & Sharp; McGrath and Hood, 1973; Silva, 1989; Dondurur and Pamukcu, 2003). Large potential field data necessitate automatic interpretation techniques like Euler's algorithm and Werner deconvolution (Thompson, 1982; Werner, 1953). Werner created a technique for analyzing the magnetic anomalies of two-dimensional sheets-like structures that arranges the anomalies and separations from observation points along a profile to create a linear equation with coefficients related to the sheet's characteristics.

From the perspective of mathematical physics, the inversion of potential field data in geophysics belongs to ill-posed problems. The ambiguity of the solution position results from the inherent instability of potential field data inversion. Numerous solutions to this arrogant conundrum have been offered over the past few decades (Blakely, 1995; Mudretsova, 1990). This approach is based on the estimation of the position of the source under the assumption of a simple anomalous body. The so-called deconvolution or semi-automated method category includes this strategy. As an inverse procedure to the direct issue in potential fields theory, which is characterized by a convolution integral, Hartman et al. (1971) coined the term "deconvolution." The deconvolution procedure for a sheet in magnetometry was described by Kilty in 1983. We are emphasizing this information so that you won’t be misled by the deconvolution process utilized in processing seismic data.

The starting values of the parameters are not necessary for Werner's technique. Based on clusters of false positions at the locations of sheets, there has been an emerging trend to use the Werner deconvolution method...
to trace foundation structures (Ku and Sharp, 1983; Malleswara Rao et al., 1983; Thakur et al., 2000). To discover the best answers, Thompson (1982) presented clustering, which tends to group into so-called clusters (grouped arrangements of solutions). However, the number of geologic entities forming such clusters is unknown. The main goal of our contribution is to study the properties of Werner deconvolution and apply the K-means unsupervised machine learning algorithm to introduce an efficient clustering algorithm to detect with precision the exact location and number of the geologic body that generate such solutions, as one can often not determine how many geological bodies are clustered from a potential field data when Werner deconvolution is carried out.

Three factors limit the accuracy and utility of the Werner deconvolution method in practice. First, analysis is limited to simple 2-D models such as dipping dikes, interfaces, and horizontal cylinders, with a minimum resolution criterion between neighboring bodies. Because Hartmann et al. (1971), Jain (1976), and Godson (1978) addressed these constraints. They are not the focus of this paper. Second, the matrix of coefficients involved in the system of equations solution is generally-conditioned, making its inversion very susceptible to noise; this is why the K-means machine algorithm was used to solve the problem. According to Radhakrishna Murty et al. (2000), clusters are only restricted to contacts and faults in basements when they are far apart and extremely steep. Werner’s technique does not offer a trustworthy interpretation for structures with smooth edges, modest dipping, proximity, and moderate depth. Although the source of this study is a vertical thin dike, it can also be used for other types of bodies if it is assumed that the body is made up of multiple thin dikes. The procedure is easy, quick, and trustworthy when it is limited to interpreting the anomalies of isolated simple geometric bodies. These methods’ depth estimations could be used as rough starting points for modeling techniques.

Machine learning algorithms have been adopted in geophysical studies, such as exploration geophysics, earthquake localization, aftershock pattern analysis, and earth system analysis (Huang et al., 2017; Helmy et al., 2010; Jia and Ma, 2017; Lim, 2005; Poulton, 2002; Zhang et al., 2014; Mousavi et al., 2016, DeVries et al., 2018, Reichstein et al., 2019). This paper focuses on the application of the K-means clustering algorithm to Werner deconvolution solutions to predict optimal results from solutions generated by Werner deconvolution. This method looks at the possibility of detecting the specific number of geologic bodies that generate the solutions generated by Werner deconvolution.

2. Methodology

We grasp the profile technique under the 2D interpretation method based on the notion of a 2D body. An interpretation equation is converted into a set of linear algebraic equations, and the unknown parameters are then solved. The n-point window traveling along the profile is used to determine the input parameters. The parameters of the specified body type are generated from the linear system solution vector. In the following, we will refer to the spatial position of the computed result as the solution/source position. One method is based on Werner deconvolution and the other is the K-means machine algorithm. The Werner deconvolution method was applied to magnetic anomalies over dike models. The synthetic magnetic data were generated for the tabular model by varying the half-width to depth (w/H) ratio for the dike.

A geological model was first created by generating synthetic magnetic data to test and prove the effectiveness of the K-means unsupervised machine learning algorithm in determining clusters around an anomalous body. The parameters of the model consisted of two bodies (dike-like in form) located at a fixed depth. Werner deconvolution technique was applied to generate the magnetic solutions as clusters. The bodies were dike-like structures in shape and the solutions were plotted.

The most commonly used unsupervised clustering algorithms make use of partitioning the data into a predefined number of clusters (Witten et al., 2017). The k-means (or-means) algorithm separates data into k clusters which are defined by randomly assigning centers. The distance between a sample and a randomly assigned center is calculated and the sample is assigned to the closest center. The algorithm then calculates a new center to the defined cluster and iterates through the same process until the algorithm converges to a minimum (Marsland, 2009).

2.1. Werner deconvolution

Werner (1953) suggested a method for its interpretation after analyzing the magnetic field of random dikes. By solving the systems of linear equations, we can determine the top corners of an object either assuming a thin aisle from the T-curve or assuming a thick aisle from the (T)/x-curve. The approach does offer a wide range of applications, despite the simplifications appearing to be a constraint.

(Hartman et al., 1971; Jain, 1976; Kilty, 1983; Ku and Sharp, 1983; Klingele et al., 1991) conducted additional research on Werner deconvolution. Consider a brief overview of the “Werner” n-point operator’s characteristics. When we use an n-point operator, we are solving n equations with n or fewer unknowns.

Let us write the interpretation equation for an arbitrary spatial position of the dike (x0, h0) on a profile (x, 0) with interference polynomial coefficients c of degree increment.

\[ \Delta T(x, 0) = \frac{A(x - x_0) + B h_0}{(x - x_0) + h_0^2} + c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n \]  

(1)

Or

\[ -b_0 T - b_1 T x + T x^2 = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^{n+2} \]  

(2)

Where:

\[ b_0 = -x_0^2 - h_0^2 \]

\[ b_1 = 2x_0 \]

\[ a_0 = (Bx_0 - Ax_0) + c_0 (x_0^2 + h_0^2) \]

\[ a_1 = A - 2c_0 x_0 + c_1 (x_0^2 + h_0^2) \]

\[ a_2 = c_2 - 2c_1 x_0 + c_2 (x_0^2 + h_0^2) \]

\[ a_3 = c_{k-2} - 2c_{k-1} x_0 + c_k (x_0^2 + h_0^2) \]

\[ a_n = c_{n-2} \]

From Eq. (2) we have:

\[ x_0 = b_1 / 2 \]

\[ z_0 = \sqrt{-b_0 - k_1^2 / 4} \]

\[ d_0 = a_0 \]

\[ d_{n-1} = a_{n-1} + 2d_n x_0 \]

\[ d_{n-2} = a_{n-2} - 2d_{n-1} x_0 - d_n (x_0^2 - h_0^2) \]

\[ A = d_1 = a_1 - 2d_2 x_0 - d_3 (x_0^2 - h_0^2) \]

\[ B = d_0 / h_0 = \frac{a_0 - d_1 x_0 - d_2 (x_0^2 - h_0^2)}{h_0} \]

Where \( d = d_0, d_1, \ldots, d_n \) is “solution” vector, \( n+1 \) is the number of terms on the right-hand side of Eq. (2). The number of terms on the right-hand side of the equation is \( n+1 \), and the ‘solution’ vector is given
by the notation d = d 0,1,2, …, n (2). From parameters A and B, we can compute dip and susceptibility as well as the parameters of the interference polynomial 0, 1, and 22. Eq. (4) provides the solution for the depth and location of the dike.

From parameters A and B, we can compute dip and susceptibility, and parameters of the interference c0, c1, c2, …, cn, 2. As long as we have computed parameters A, B we can derive susceptibility contrast and the dip of the dike.

\[ k = \frac{M_f^2 + M_l^2}{|T|} \]  
\[ \varnothing = \arctan \frac{M_f}{M_l} + I \]

where I is magnetic inclination of the inducing field T.

We now try to investigate how the estimated parameters affect the solution, taking into account all the equations. All relevant data regarding the dike’s magnetic inclination, declination, strike, dip, and susceptibility (parameters a1) are contained on the right-hand side of Eq. (2). It is evident that modifications to the dike’s a1 parameters, which are related to the right-hand side of Eq. (2), have no effect on the depth and location of a dike for the supplied data. Rarely are we able to determine the precise location of a geological body without knowing any other details about it. On the other hand, as shown in equations, the parameters b1 have an impact on the tilt, strike, and magnetization solutions (3).

2.2. K-means clustering

Our computational strategy is to compute the results for all meaningful ranges of the input parameters (window width, interference polynomial, structural index).

K-means cluster analysis is an example of a hard partitioning algorithm (Hartigan, 1975; Hartigan and Wong, 1979; Jain et al., 1999). Consider a set of P data (x1, x2, …, xn) in d dimensions which are partitioned into K clusters, where each element in the dataset is allocated entirely to a particular cluster.

The chance of recognizing regularities and latent patterns inside any piece of data is what we are interested in whenever we want to cluster it. In practice, we must select how we wish to characterize and search for structures inside a data. The most popular answers to these questions are given by the K-means machine learning algorithm. Structures are represented as subsets of X, and the data is divided into a specified number (k) of clusters. In machine learning, K-means clustering is a fundamental and well-understood technique for splitting data into clusters represented by centroids.

\[ X = x_1, x_2, x_3, …, x_n \]  
\[ S = \{S_1, S_2, …, S_k\} \]

Where;

\[ S_i = \{S_1, S_2, …, S_k\} \]

By minimizing the objective function

\[ J = \sum_{i=1}^{k} \sum_{j=1}^{n} (|x_i - S_j|)^2 \]  

The type of optimization used to minimize J is specific to the k-means cluster implementation.

As a result of reducing Eq. (8), suitable centroids \( S_i \) are determined so that the distances between data points and their closest cluster centroid are as small as possible when the data is partitioned into distinct clusters C.

In a word, the k-means method reduces Eq. (8) to its simplest form by using a greedy iterative update technique. It randomly initializes the parameters \( S_1, S_2, …, S_k \). When it is launched, say at iteration t = 0. The data is then clustered according to this initial guess. This simply means that the program finds k groups in which;

\[ C_i = \{x_j \in X | \|x_j - S_i\|^2 < \|x_j - S_j\|^2 \forall j \} \]  

After clusters have been identified, the algorithm computes to update the current estimations of cluster centroids.

\[ S_i^{(t+1)} = \frac{1}{n_i} \sum_{x_j \in C_i} x_j \]  

The number of items in cluster \( C_i \) is denoted by \( n_i = |C_i| \). The iteration counter is increased t = t + 1 and steps (8) and (9) are repeated until the assignment of data points to each cluster does not change any longer or t exceeds a predefined number of iterations.

It can be shown that updating each centroid to the sample mean of its cluster will not increase the value of objectives in Eq. (5) or (7). Therefore, each iteration improves on the previous result and k-means clustering usually converges quickly. However, as the algorithm starts from a random initialization, it cannot be guaranteed to find the minimum of its objective function. Running k-means a few times to empirically determine appropriate clusters is critical, as it often converges to a local minimum.

Once the selected Werner solution data set has been divided into coherent and undesirable solutions, the centers x, y, and z of the c clusters within the filtered data set are obtained using the K-means algorithm. The primary geologic structures being mapped are represented by these groupings. The algorithm’s primary flaw is that the number of clusters c must be determined before performing the computation (Balasko et al., 2005). Another option is to perform the clustering for various c values before computing validity indices for each division to evaluate the merits of each.

2.3. Determination of geologic strike of sources mapped

The interpreter must provide a distance threshold around each cluster so that it only tracks points that are close to it. Currently, all clusters share a single threshold. It is possible to set a threshold for each cluster separately, but this would necessitate more human intervention and reduce automation somewhat. The value of this common threshold was determined in our case by using cluster centers.

3. Synthetic magnetic model

The efficiency of this technique is investigated by generating synthetic data by using the method of Barse (Barse et al., 2003) to generate synthetic data. Both simulated and actual data were used to evaluate the algorithm. The synthetic model comprised of two substantially elongated bodies to test the algorithm’s capacity to solve for the strike of geologic structures, which was one of the key objectives (Figure 1 and Table 1).

Both bodies were 30 m deep and 3 m thick. The east-west body was 45 km long while the north-south body was 30 km long. In both instances, the depth to the top was 50 m. The ‘north-south body’ and ‘east-west body’ dip at 45° to the east and south, respectively. The ambient field’s characteristics were declination 0°, inclination 90°, and total intensity 33,000 nT. There was no extra remanent magnetization. Both bodies’ structural magnetic susceptibilities were 0.0035. At a common observation level of 0 m and with north-south lines separated at 500 m, a total magnetic intensity (TMI) grid was generated. The lines were spaced 10 m apart. A common minimal curvature approach was used to grid the forward model at a 100-m grid cell size after it was derived on profiles.
Werner deconvolution applied on $\Delta T$ field of two thin dyke with parameters: inducing field: $T = 33000$ nT, inclination $= 4^\circ$, declination $= 86^\circ$. The total field parameter was chosen for this synthetic model based on the total residual magnetic field in Nigeria, and the inclination and declination were gotten.

The Werner deconvolution technique was applied to the residual field and the result is shown in Figure 2. Only solutions for dyke model were plotted since the bodies are dyke-like in shape. As can be seen, many solutions clustered around the depth to the top of the bodies but it would still have been difficult to ascertain the number of bodies in the model. To concentrate the solutions near the top of the bodies, a window of $100/100$ m was used. The spray effect observed in the depth solutions is what is usually encountered when dealing with noise-free data. Random noise was therefore added to the data to make it more realistic. This noise was created by adding normally distributed random values to the theoretical magnetic anomaly of the dike body.

The theoretical magnetic anomaly of a single dipping dike was added with normally distributed random values to examine the impact of noise on the deconvolution process. The location and depth of the dike were then calculated using Werner deconvolution.

The result of this shown in Figure 3, even though devoid of sprays indicates (based on the clustering observed) that there are most likely three bodies in the model.

Figures 4 and 5 show the initial plot of the distance against depth, before applying the K-means clustering.

### 3.1. Modeling and analysis

Our computational strategy was to compute the results for all meaningful ranges of the input parameters (window width, interference polynomial, structural index). From the huge heap of results, we will pick up those that seem to be the best.

---

**Table. 1** Model parameters assuming an ambient field of 33,000 nT with $4^\circ$ and $86^\circ$ as magnetic inclination and declination respectively.

| Body | Width (m) | Depth to top (m) | Depth to bottom (m) | Horizontal position (at midpoint) (m) | Susceptibility (SI unit) |
|------|-----------|------------------|--------------------|--------------------------------------|-------------------------|
| A    | 3.0       | 5.0              | 27.0               | 31.5                                 | 0.01                    |
| B    | 5.0       | 10.0             | 20.0               | 42.5                                 | 0.01                    |

---

**Figure 1.** Magnetic field intensity produced by the two bodies (A and B) with field intensity of 33000 nT and magnetic inclination and declination of $4^\circ$ and $86^\circ$ respectively.

**Figure 2.** Magnetic solutions generated by the two geologic structures (the blue color represents the solutions), with field intensity of 33000 nT and magnetic inclination and declination of $4^\circ$ and $86^\circ$ respectively.
Figure 3. Magnetic solutions generated by the two geologic structures with random noise added (the blue color represents the solutions).

Figure 4. Initial plotting of the noise-free data before clustering (the blue color represents the solutions).

Figure 5. Initial plotting of the noise-added data before clustering (the blue color represents the solutions).
The K-means machine learning algorithm was applied to the two data sets. The major objective of the K-means is to group similar data points and discover similar patterns. The SciKit-learn library was used to import the necessary libraries for the K-means clustering implementation.

The elbow method was used first to determine the number of clusters before their centers were obtained using the K-means algorithm. **Figures 6 and 7** show the number of clusters determined by the elbow method for the noise-free and noise-induced data. In this case, the K validity was found to be two (2).

The K-means algorithm was then applied to the generated solutions produced after Werner deconvolution was carried out on the synthetic model. This was done to see how effectively it can detect the number of dyke-like bodies that generated the solutions. K-means’ main goal is to group similar data points and discover any similar patterns. K-means clustering will begin by locating k centers to form ‘n’ clusters. A cluster can be formed by aggregating similar data points based on specific functionalities and similarities. Centroids are imaginary or real locations at the center of a cluster identified by k-means clustering and forming k-centers. These k-centers connect the closest data points to the closest clusters, forming ‘n’ clusters while keeping the data centers as small as possible. The magnitude of the data (input and output) was scaled to ensure that the learning algorithm was not biased. After normalization, the data becomes stable within a given range of 0–1.

For the noise-free data, the result of the K-means algorithm shows two clusters with two centroids (color red and blue) as shown in **Figure 8**. **Figure 9** shows the k-means algorithm effect on the noise-enhanced data, with the two centroids also being shown.

The centroids for both the noise-free and noisy data were plotted on the synthetic model, with the red and green dots representing the centroids for noise and noise-free data, respectively (Figure 10). It can be seen that the centers produced by the machine learning program closely approximate the depth to the top of the bodies in the model and the comparison with the true depths is shown in **Table 2**.

Given that we just have two geologic dike shaped bodies from the magnetic synthetic model, we proceeded to perform Werner deconvolution to generate the solutions. K-means clustering algorithm was then applied to the solutions to reveal how effective the algorithm is in detecting the number of magnetic bodies, especially when noise was added. From the results, it was shown how successful the K-means algorithm could cluster magnetic solutions into two (2) clusters, revealing the true number of geologic bodies. After running the K-means algorithm a few times, it was found that the clusters were more or less stable. The centroids of each cluster were the position of the geologic bodies. For the less noise model, the values of the centroids for the two bodies were 6 m and 9 m in depth with a horizontal distance of 34 m and 44 m respectively. Considering the noise model, the values of the centroids were 5 m

**Figure 6.** Elbow point for the noise-free dataset indicating that the optimum number of partition clusters is two (2).

**Figure 7.** Elbow point for the noisy dataset indicating that the optimum number of partition clusters is two (2).
Figure 8. Visualization of K-Means cluster Analysis applied to the noise-free data (the green and red colors represent two partitioned clusters).

Figure 9. Visualization of K-means cluster analysis applied to the noisy data (the green and red colors represent two partitioned clusters).

Figure 10. Visualization of K-Means centroids on the synthetic model (green and red color represents the centroid of both the non-noise and noisy data).
and 9 m in depth, at a horizontal distance of 32 m and 49 m respectively. This shows the effectiveness of the K-means algorithm in determining how many geological bodies created the solutions.

4. Application: Dangoma mining site, Kaduna, Nigeria

Finally, we applied the algorithm to present a real data set at the Dangoma nickel deposit site in Kaduna, Nigeria. The site is one of the most potential nickel deposit districts in Nigeria. The area has an average elevation of about 700 m above sea level, is of low relief with slightly undulating topography and low-lying outcrops, and a few prominent ridges of weathered granite scattered over the landscape. The lateritic regolith is thick and deeply weathered. The Dangoma area is primarily the Precambrian Basement Complex which is composed of the mostly Archean migmatite-gneiss complex, Proterozoic schist belts, and the Pan-African Older Granites intruded by minor meta-ultramafic bodies and pegmatite dykes (Oyawoye, 1970; Ajibade and Wright, 1988). The Dangoma area lies east of the N-S trending supracrustal schist belts of northwestern Nigeria (Figure 11), it is underlain by migmatites, gneisses, (serpentinite), and granites. An airborne geophysical survey was flown in 2006 to aid in the geologic mapping and exploration of Nigeria. The magnetic sensor was towed at an altitude of 80 m, with a line spacing of 200 m. In each of the four subblocks into which the land was divided, the flight lines were parallel to the principal geologic strike.

Figure 12 displays the Residual Magnetic Intensity data over a window of the region where we evaluated the clustering algorithm and Werner deconvolution. The magnetic data show the contrast between weakly magnetic felsic volcanic units and basalts, gabbros, and strongly magnetic pegmatite dike and intrusive units. The location of the mineral deposit in the area is shown by the black box on the Residual map, for reference. The correlation between the data sets is good, although some areas apparently do not have any associated magnetic signature, like the folded units on the southeast part of the map.

As with the synthetic data set, the Werner deconvolution was applied across a profile drawn on the Residual Magnetic data. The algorithm produced approximately 200 solutions with depths between 0 and 2400 m a window size of 100 m was used. A magnetic susceptibility value of 0.002549 SI was recorded at a dip angle 48.1°/C14, and this occur at the distance of 38000m along the profile, with depth values ranging from 500 m to 2400 m.

The generated Werner solutions are as shown in Figure 13. The elbow method was then used to determine the number of partition clusters before their centers were obtained using the K-means algorithm. The K-means clustering technique was applied to this data set. After experimenting with various cluster numbers and determining the best, an ideal number of two clusters was determined. The algorithm’s strike information closely matches the folded structures in the Dangoma area, and cluster locations are depicted in Figure 14. This data should be useful for any future geophysical modeling efforts in the region.

5. Discussion

This method is far from foolproof for structural mapping from geophysical data. A wise selection of the various parameters required for
Figure 12. Residual magnetic intensity map over the study area (the black box shows the location of the deposits).

Figure 13. Initial plotting of the Werner solutions on the Dangoma data with susceptibility 0.002549 SI set before the application of the K-means algorithm.
The computation is critical to obtaining good results. The results of the clustering algorithm on a high-quality Werner data set, on the other hand, are extremely reasonable and should be useful as input for structural analysis and geologic modeling.

The Werner deconvolution technique is a broadly applicable technique for the analysis of 2-D potential field data. The ease with which it is programmed on a digital computer allows for quick analyses of reconnaissance aeromagnetic and gravity data, and its automatic regional-residual separation. These are its two principal uses. Kitty (1983) showed that its main flaws are sensitivity to geologic noise, field data noise, and the relatively high correlation between the metrics defining interference and those describing the desirable features of the anomaly-causing entity. This served as the bases for applying the K-means machine learning algorithm.

In general, unless one considers that there is no interference, the effect of this correlation is to only provide a certain amount of information about the depth and location of the causative body. Werner deconvolution is less preferable than forward or inverse modeling when more in-depth information is needed due to the amount of scatter in the position and depth estimates. Without a doubt, the Werner deconvolution's chosen algorithm makes or breaks the end product (geological imaging and depth estimations). The data will be filtered using the methods described here even if the Werner solution data set is not properly confined; nevertheless, the strike and depth information gained will only be useful if the data are closely connected to one another and follow real geologic structures. The Werner solutions utilized here were subjected to the machine learning technique, which produced good results in terms of solution coherency; as a result, the strike information in both the synthetic and actual data examples was accurate. Although the methodology was applied mostly to derive the geologic and depth of anomalous bodies, it undoubtedly can be applied to other problems such as model construction.

The algorithm showed the ability to detect a number of geologic bodies that produced the solutions. This tool has provided alternative ways of learning from data, and lends itself naturally to different data types, with machine learning being well suited to high-dimensional data. K-means is a good first choice of algorithm that can be applied for geological clustering. Importantly, the K-means machine learning algorithm is relatively straightforward to implement by a non-machine learning expert. Using machine learning before geophysical inversion allows information from a greater variety of data sources to be used to improve detail in 2D Werner deconvolution mapping.

The K-means clustering algorithm is very stable. Both applications (synthetic and real data) gave very good results in terms of partitioning the data into proper clusters and recognizing the major body generating the solutions.

6. Conclusion

In the presented contribution we analyzed solutions for 2D semi-automated inversion methods, namely Werner deconvolution to improve the results. For this purpose, the K-mean machine learning algorithm was applied to the Werner solutions. When good choices are made for the k-mean clustering parameter, it is a useful machine learning tool that can cluster Werner solutions and predict the number of geologic bodies that produced such solutions. This will enhance how subsurface structures are characterized, especially in places of complex geology where human intelligence may be limited.

The algorithm was tested on one synthetic data set and one real data set at the Dagogoma mining site in Kaduna, Nigeria. By using the Werner-generated solutions as input, the K-means algorithm was able to separate the clusters at the mining site and pinpoint the general location of the geologic formations. The best solution set must be found by running the K-means clustering algorithm with various cluster sizes. For large data sets, this is computationally inefficient and can be very slow. Further research is also required in the domain of curvature-based magnetic source filtering to reduce anomalous interference effects prior to Werner deconvolution.

Declarations

Author contribution statement

Daniel Eshimiakhe: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Kola Lawal: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data will be made available on request.
**Declaration of interests statement**

The authors declare no conflict of interest.

**Additional information**

No additional information is available for this paper.

**Acknowledgements**

The first author gratefully acknowledges the Department of Physics, Ahmadu Bello University (ABU), Zaria Nigeria for the platform it created for this research to be carried out.

**References**

Ajibade, A.C., Wright, J.R., 1988. Structural relationships in the schist belts of northwestern Nigeria. In: Oloyide, P.O., Mbounu, W.C., Ogeri, A.E., Ebibuiwe, L.G., Ajibade, A., A. C., Umeji, A.C. (Eds.), Pecencamin Geology of Nigeria. Geological Survey of Nigeria, Kaduna, pp. 103–110, 1988.

Am, K., 1972. The arbitrarily magnetized dike; interpretation by characteristics. Geophysical Prospecting 19 (5), 657–672.

Balsek, B., Abonyi, J., Feil, B., 2005. Fuzzy Clustering and Data Analysis Toolbox: for Use with Matlab. PhD Thesis. University of Veszprem.

Barse, E.L., Kvarnstrom, H., Jonsson, E., 2003. Synthesizing test data for fraud detection systems. In: Proceedings of the 19th Annual Computer Security Applications Conference. IEEE.

Blakely, R.J., 1995. Potential theory in gravity and magnetics applications. Cambridge University Press, New York.

DeVries, P.M.R., Viegas, F., Wattenberg, M., Meade, B.J., 2018. Deep learning of interpolation application. Geophysics 82 (3), V163–V177.

Dondurur, D., Pamukcu, O.A., 2003. Interpretation of magnetic anomalies due to dykes. Geophysical Prospecting 51, 2156–2159.

Dundur, D., Pamulcu, O.A., 2003. Interpretation of magnetic anomalies due to dykes. Geophys. Prospect. 11, 509–522.

Godson, R.H., 1973. Depth to Magnetic Basement System; U.S.G.S Open File Rep. Denver.

Hartigan, J.A., 1974. Clustering Algorithms. John Wiley & Sons, New York, p. 50.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS136: a K-means clustering algorithm. Appl. Stat. 28, 100–108.

Hartman, R.R., Tesky, D.J., Friedberg, J.L., 1971. A system for rapid digital aeromagnetic interpretation. Geophysics 36, 891–918.

Helmy, T., Fatai, A., Faisal, K., 2010. Hybrid computational models for the characterization of oil and gas reservoirs. Expert Syst. Appl. 37 (7), 5353–5363.

Huang, K., You, J., Chen, K., Lai, H., Don, A., 2017. Neural Network for Parameters Determination and Seismic Pattern Detection. SEG Technical Program Expanded Abstracts, pp. 2285–2289.

Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Comput. Surv. 31, 264–322.

Jia, Y., Ma, J., 2017. What can machine learning do for seismic data processing? An interpolation application. Geophysics 82 (3), V163–V177.

Kilty, K.T., 1983. Werner deconvolution of profile potential field data. Geophysics 48, 234–237.

Klinge, E.E., Marson, I., Kable, H.G., 1991. Automatic interpretation of gravity gradiometric data in two dimensions: vertical gradient. Geophysical Prospecting 39, 407–434.

Kolomzern, T., Lamontagne, Y., Nadeau, A., 1970. New methods for the direct interpretation of magnetic anomalies caused by inclined dikes of infinite length. Geophysics 35, 812–830.

Ku, C.C., Sharp, J.A., 1983. Werner deconvolution for automatic magnetic interpretation and its refinement using Marquardt’s inverse modeling. Geophysics 48 (6), 754–774.

Lim, J.S., 2005. Reservoir properties determination using fuzzy logic and neural networks from good data in offshore Korea. J. Petrol. Sci. Eng. 49 (3–4), 182–192.

Malleswara Rao, M.M., Lakshminarayana, S., Murthy, K.S.R., Subrahmanyan, A.S., 1983. Two computer programs for the analysis of marine magnetic data. Comput. Geosci. 19 (5), 657–672.

Marsland, S., 2009. Machine Learning: an Algorithmic Perspective, second ed. Chapman & Hall / CRC / Taylor & Francis Group, pp. 50–70.

McGrah, P.H., Hood, P.J., 1973. An automatic least-squares multi-model method for magnetic interpretation. Geophysics 38, 349–358.

Mousavi, S.M., Horton, S.P., Langston, C.A., Samei, B., 2016. Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. Geophys. J. Int. 207 (1), 29–46.

Gavriometry. In: Modrestova, E.A., Veselov, K.E. (Eds.), 1990. Nedra, Moscow. (in Russian).

Oluwoye, M.O., 1970. The basement complex of Nigeria. In: Dessauvagie, T.F.J., Whiteman, A.J. (Eds.), African Geology. University of Ibadan press, pp. 67–99.

Poulton, M.M., 2002. Neural networks as an intelligence amplification tool: a review of applications. Geophysics 67 (3), 979–993.

Prakasa Rao, T.K.S., Subrahmanyan, M., Sríkrishna, M.A., 1986. Nomograms for the direct interpretation of magnetic anomalies due to long horizontal cylinders. Geophysics 51, 2156–2159.

Rahelkushita Murti, I.V., Swamy, K.V., Rama Rao, P., 2000. Use and abuse of Werner deconvolution technique. J. India Geophys. Union 4 (2), 97–102.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven earth system science. Nature 566 (7743), 195–204.

Silva, J.B.C., 1989. Transformation of nonlinear problems into linear ones applied to the magnetic field of a two-dimensional prism. Geophysics 54, 114–121.

Subrahmanyan, M., Prakasa Rao, T.K.S., 2009. Interpretation of magnetic anomalies using some simple characteristic positions over tabular bodies. Explor. Geophys. 40, 265–276.

Thakur, N.K., Gangadhara Rao, T., Khanna, R., Subrahmanym, C., 2000. Magnetic basement in the Bay of Bengal through Werner deconvolution. Mar. Geol. 162, 2289–2304.

Thompson, D.T., 1982. EULDPH-a technique for making computer-assisted depth estimates from magnetic data. Geophysics 47, 31–37.

Werner, S., 1953. Interpretation of magnetic anomalies of sheet-like bodies. Sveriges Geologer under Ser.C.C, Arsbok 43 (6).

Witten, Ian H., Frank, Eibe, Hall, Mark A., Pal, Christopher J., 2017. Data Mining (Fourth Edition). Morgan Kaufmann, p. iv.

Zhang, C., Fregner, C., Araya-Polo, M., Hohl, D., 2014. Machine-learning-based automated fault detection in seismic tracces. At the. 76th EAGE conference and exhibition 1–5.