fuzzy c-maens analysis of magnetotelluric and seismic models

Roberto Hugo Melo dos Santos (IFBA)*, Fernando Acácio Monteiro dos Santos (UL)

This paper was prepared for presentation during the 15th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 31 July to 3 August, 2017.

Abstract

Knowing that combining analysis of different kinds of geophysical data has the potential to improve model resolution, since different observations are sensitive to different subsurface features, we examine the compatibility of P-wave seismic velocity and electrical resistivity from magnetotelluric (MT) observations, using fuzzy c-maens method. We propose an approach based on fuzzy c-means cluster analysis for the cooperative inversion of seismic and MT data sets. Thus, joint inversion of these two complementary data sets can be used to construct improved models. Results show the possibility of a successful application to real data.

Introduction

In recent years the use and development of algorithms and methods for inversion and modeling data set has become increasingly popular. The interpretation of geophysical models derived by inversion is a highly subjective part of any geologic study. The incomplete knowledge of the subsurface, the varying spatial resolution of the models, and the non-uniqueness of the geophysical inverse problem make it difficult to interpret geophysical data directly in terms of geologic structure.

It is common to use a combination of geophysical methods to obtain the distribution of independent physical properties over the area of interest in order to discriminate between the different possible data sets. Unfortunately, field and laboratory investigations show that petrophysical relations are often complex, showing non-unique, nonlinear, or site-specific dependencies (Shoen, 1998). Therefore, the most popular approach for inverting disparate data sets is to introduce a link based on common structures (Lines, Schultz, & Treitel, 1988; Santos, Sultan, Represas, & Sorady, 2006).

(Gallardo, 2004; Haber & Oldenburg, 1997) showed more flexible 2D structural joint inversion approaches that minimize data misfits and differences between structures in two different geophysical models. Their approaches result in two (or more) smoothly varying geophysical parameter models, which, in a subsequent step, must be simultaneously interpreted. A manual interpretation of various collocated models is qualitative in nature because the outcome depends on the experience and preconceptions of the interpreter. A more quantitative and objective post-inversion interpretation of different models can be achieved by using statistical techniques such as cluster analyses (Paasche, Tronicke, Holliger, Green, & Maurer, 2006; Tronicke, Holliger, Barrash, & Knoll, 2004).

The magnetotelluric (MT) and seismic methods are the only geophysical exploration techniques that can yield reliable images at depths greater than the km-scale. The MT and seismic tomography techniques provide images of electric resistivity (ρ) and seismic velocity (V_p,V_s), respectively, with similar spatial resolution (Bedrosian, Unsworth, Egbert, & Thurber, 2004; Jones, 1987; M. Unsworth & Bedrosian, 2004) and are often used in combination to derive models of the subsurface (Jones, 1998; Maercklin et al., 2005; Mechie et al., 2004; M. J. Unsworth et al., 2005). Both methods have their characteristic limitations. MT, for example, has an inherent loss of resolution with depth because it is based on diffusive fields. Seismic refraction has trouble imaging vertical contrasts. By looking at both resistivity and velocity at the same time, we can build on the strengths of both methods and mitigate their weaknesses.

Here, we use fuzzy clusters algorithms, the fuzzy c-means (FCM) algorithm, assigning each data point in the multidimensional space to all sunsets with varying degrees of membership. Hence, a data point can be a member of one cluster but also a partial member of other clusters. Thus, fuzzy cluster algorithms provide important additional information about quality and internal consistency of the performed data classification.

Methodology

We assume that subsurface can be described by a zonal model in which each zone is characterized by a set of different geophysical parameter. The method was test on two synthetic data example.

The combined interpretation of different measurement types is a basic principle to confine the ambiguity of the inverse problems in geophysics. (Paasche & Tronicke, 2007) proposed a flexible cooperative inversion approach based on fuzzy c-means (FCM) cluster analysis and conventional single input data set inversion algorithms. This method has been successfully applied by his creators to various set data, but never tried with MT and seismic.

The FCM is a partitioning cluster algorithm proposed by (Dunn, 1973) and improved by (Bezdek, 1981) to group n data points in at-dimensional space into a specified number of subsets or clusters c by iteratively minimizing the objective function

$$J = \sum_{i=l}^{c} \sum_{j=1}^{n} \left(u_{ij}\right)^{f} \left\| d_j - v_i \right\|^2, \quad (1)$$

Where u_{ij} denotes the degree of membership of data point d_j to cluster i defined by its center v_i. Memberships are constrained to be positive and to satisfy

$$\sum_{i=l}^{c} u_{ij} = 1, \qquad (2)$$

With $j = 1,...,n$. The weighting exponent f, also referred to as the fuzzification parameter (Fridgen et al., 2004; Güler & Thyne, 2004), controls the degree of "fuzziness" in the resulting memberships and lies in the range of $1 < f > \infty$. For most databases, a selection of f between 1.5 and 2 is regarded as a suitable choice (Hathaway & Bezdek, 2001). As f approaches unity, FCM cluster analysis approximates the crisp *k*-means algorithm while increasing f results in an increased fuzziness of the memberships. In equation 1, the Euclidian distance between the *jth* data point and the *ith* cluster center is calculated in a *t*-dimensional space using

$$\left\|d_j - v_i\right\| = \sqrt{\sum_{\alpha=1}^{t}\left(d_{j\alpha} - v_{i\alpha}\right)^2}, \qquad (3)$$

Where the locations of data points d_j and cluster centers v_i are defined by t attributes. After providing the initial parameters (number of clusters c, fuzzification parameter f, and an initial guess of u_{ij} or v_{ia}), J is minimized with respect to u_{ij} and v_i by iterative alternating optimization (Bezdek & Hathaway, 2002). One iteration consists of updating the membership values u_{ij}

$$u_{ij} = \sum_{k=1}^{c}\left\|\frac{d_j - v_i}{d_j - v_k}\right\|^{\frac{-2}{f-1}}, \qquad (4)$$

and the cluster centers v_i

$$v_i = \frac{\sum_{j=1}^{n}\left(u_{ij}\right)^f d_j}{\sum_{j=1}^{n}\left(u_{ij}\right)^f}, \qquad (5)$$

The order of equations 4 and 5 depends on whether initial memberships or center locations are given. The algorithm terminates after a predefined number of iterations or if the improvement of J falls below a given threshold.

Based on FCM cluster analysis, (Paasche et al., 2006) develop an approach to combine information contained in different collocated physical property models to form a single zoned multiparameter model. In the following, we applied this data-integration approach to a synthetic MT and seismic database, which we will use to test the performance of the cluster algorithms.

The database to be integrated comprises two fully collocated models of 2D spatial distributions of parameters, MT and seismic. Both models are equally discretized. They form a 2D data space, in which the location of each data point is defined by the two parameter values of A and B in each model cell. The data points in this 2D space are now subjected to FCM cluster analysis. Clustering is repeatedly performed, varying the number of clusters c from 2 to 10 and setting the fuzzification parameter to $f = 1.75$.

The membership values obtained from FCM cluster analysis describe the integrated database in a fuzzy sense. However, for most applications there is a need for a crisp solution to emanate from the fuzzy membership information, which can be achieved by defuzzification of the membership information (Lee, 1990; Leekwijck & Kerre, 1999). This is always accompanied by a certain loss of information. Throughout this study, we defuzzify the clustering results by assigning each data point to the cluster for which it has the highest degree of membership. This is a rather simple defuzzification process referred to as "maximum criterion defuzzification" (Lee, 1990).

Synthetic Study

We consider the inversion of synthetic 2D model shown in Figure 1, the MT model comprises a low-resistivity body ($\rho = 10\ \Omega$m) into an uniform $\rho = 100\ \Omega$m medium (Fig. 1a). The seismic model comprises a low-speed body ($v_p = 2000$ m/s) into an uniform $v_p = 3000$ m/s (Fig. 1b) medium. We have use a split-spread geometry with source and the receivers located at the surface of the model. The inversion parameters seismic are the velocity and density of each layer, which is divided into four blocks. We use the 2D pseudo-spectral method (Kosloff and Baysal 1982) to compute seismic traces every.

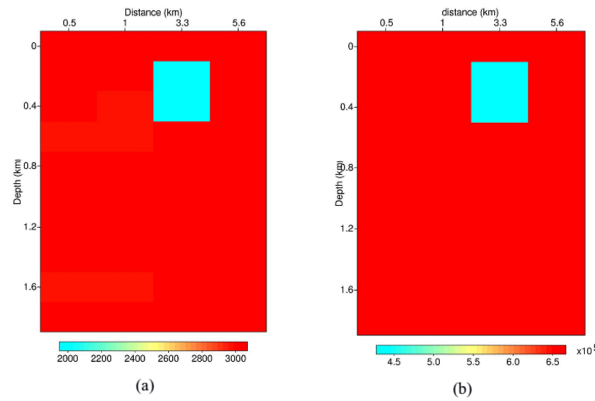

Figure 1 - Structural model MT (a) and Seismic (b) used in the synthetic example.

We now use P-wave velocity model and resistivity model shown in Figure 1 as input the FCM algorithm cluster analysis. Comparable to the synthetic study we choose c = 4 as the optimum number of clusters (Figure 2) to MT and (Figure 3) to seismic.

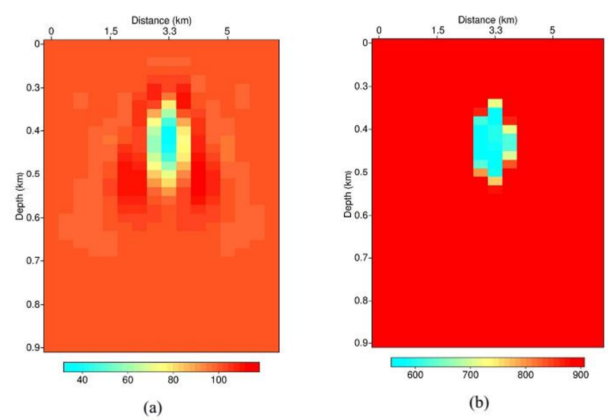

Figure 2 - Structural model MT with low-resistivity body (ρ = 10 Ωm) into a single layer ρ = 100 Ωm(a) and model obtained using the FCM (b)

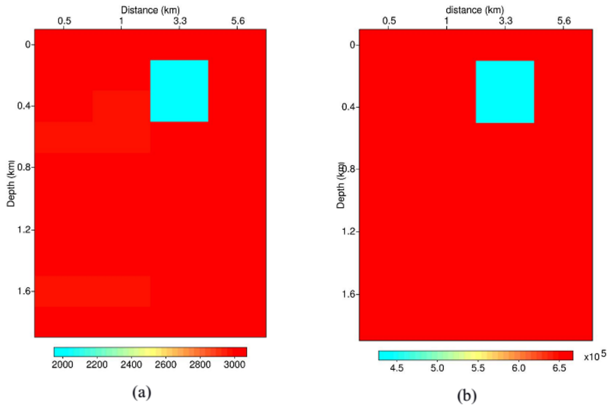

Figure 3 - Structural model seismic with low-velocity body (v_p = 2000 m/s) into a single layer v_p = 3000 m/s (a) and model obtained using the FCM (b).

Discussion and Conclusion

Figure 2 and 3 show a comparison between observed and predicted data, when in overall the general features of both models are similar. We have presented a concept of an analysis FCM method that use diverse data to constrain a common Earth model. For synthetic models we demonstrate that the combination of MT resistivity with seismic velocity data allows visualizing local rock classifications in a manner that can be directly applied to the exploration of deep geophysical reservoirs.
The method is independent of theoretical relations kinking electrical and seismic parameters. It represents an advance from the common qualitative interpretation of multiple physical property models, and is less susceptible to the subjectivity of such a joint interpretation. As shown in synthetic examples, our approach is capable of "breakinf the degeneracy" of overlapping physical parameters, providing structural information not evident in the individual models. Analysis fuzzy c-means of magnetotelluric and seismic models presents a formal method to include diverse data in a common, robuster model and help interpretation by reducing the ambiguity from a range of models derived from single data sets.

Acknowledgements

We would like to acknowledge of the Federal Institute of Science Education and Technology Bahia – IFBA, Lisbon University and Dom Luiz Institute – IDL.

References

Bedrosian, P. A., Unsworth, M. J., Egbert, G. D., & Thurber, C. H. (2004). Geophysical images of the creeping segment of the San Andreas fault: implications for the role of crustal fluids in the earthquake process. *Tectonophysics, 385*(1–4), 137-158. doi: http://dx.doi.org/10.1016/j.tecto.2004.02.010

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*: Kluwer Academic Publishers.

Bezdek, J. C., & Hathaway, R. J. (2002). Some Notes on Alternating Optimization. In N. R. Pal & M. Sugeno (Eds.), *Advances in Soft Computing — AFSS 2002: 2002 AFSS International Conference on Fuzzy Systems Calcutta, India, February 3–6, 2002 Proceedings* (pp. 288-300). Berlin, Heidelberg: Springer Berlin Heidelberg.

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics, 3*(3), 32-57. doi: 10.1080/01969727308546046

Fridgen, J. J., Kitchen, N. R., Sudduth, K. A., Drummond, S. T., Wiebold, W. J., & Fraisse, C. W. (2004). Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. *AGRONOMY JOURNAL, 96*, 100-108.

Gallardo, L. A. (2004). Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints. *Journal of Geophysical Research, 109*(B3). doi: 10.1029/2003jb002716

Güler, C., & Thyne, G. D. (2004). Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzyc-means clustering. *Water Resources Research, 40*(12), n/a-n/a. doi: 10.1029/2004wr003299

Haber, E., & Oldenburg, D. (1997). Joint inversion: a structural approach. *Inverse Problems, 13*(1), 63-77. doi: 10.1088/0266-5611/13/1/006

Hathaway, R. J., & Bezdek, J. C. (2001). Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 31*(5), 735-744. doi: 10.1109/3477.956035

Jones, A. G. (1987). MT and reflection: an essential combination. *Geophysical Journal International, 89*(1), 7-18. doi: 10.1111/j.1365-246X.1987.tb04380.x

Jones, A. G. (1998). Waves of the future: Superior inferences from collocated seismic and electromagnetic experiments. *Tectonophysics, 286*(1–4), 273-298. doi: http://dx.doi.org/10.1016/S0040-1951(97)00270-9

Lee, C. C. (1990). Fuzzy logic in control systems: fuzzy logic controller. II. *IEEE Transactions on Systems, Man, and Cybernetics, 20*(2), 419-435. doi: 10.1109/21.52552

Leekwijck, W. V., & Kerre, E. E. (1999). Defuzzification: criteria and classification. *Fuzzy Sets and Systems, 108*(2), 159-178. doi: http://dx.doi.org/10.1016/S0165-0114(97)00337-0

Lines, L. R., Schultz, A. K., & Treitel, S. (1988). Cooperative inversion of geophysical data. *Geophysics, 53*(1), 8-20. doi: 10.1190/1.1442403

Maercklin, N., Bedrosian, P. A., Haberland, C., Ritter, O., Ryberg, T., Weber, M., & Weckmann, U. (2005). Characterizing a large shear-zone with seismic and magnetotelluric methods: The case of the Dead Sea Transform. *Geophysical Research Letters, 32*(15), n/a-n/a. doi: 10.1029/2005GL022724

Mechie, J., Sobolev, S. V., Ratschbacher, L., Babeyko, A. Y., Bock, G., Jones, A. G., . . . Zhao, W. (2004). Precise temperature estimation in the Tibetan crust from seismic detection of the α-β quartz transition. *Geology, 32*(7), 601-604. doi: 10.1130/g20367.1

Paasche, H., & Tronicke, J. (2007). Cooperative inversion of 2D geophysical data sets: A zonal approach based on fuzzy c-means cluster analysis. *Geophysics, 72*(3), A35-A39. doi: 10.1190/1.2670341

Paasche, H., Tronicke, J., Holliger, K., Green, A. G., & Maurer, H. (2006). Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy -means cluster analyses. *Geophysics, 71*(3), H33-H44. doi: 10.1190/1.2192927

Santos, F. A. M., Sultan, S. A., Represas, P., & Sorady, A. L. E. (2006). Joint inversion of gravity and geoelectrical data for groundwater and structural investigation: application to the northwestern part of Sinai, Egypt. *Geophysical Journal International, 165*(3), 705-718. doi: 10.1111/j.1365-246X.2006.02923.x

Tronicke, J., Holliger, K., Barrash, W., & Knoll, M. D. (2004). Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation. *Water Resources Research, 40*(1), n/a-n/a. doi: 10.1029/2003wr002031

Unsworth, M., & Bedrosian, P. A. (2004). Electrical resistivity structure at the SAFOD site from magnetotelluric exploration. *Geophysical Research Letters, 31*(12), n/a-n/a. doi: 10.1029/2003GL019405

Unsworth, M. J., Jones, A. G., Wei, W., Marquis, G., Gokarn, S. G., & Spratt, J. E. (2005). Crustal rheology of the Himalaya and Southern Tibet inferred from magnetotelluric data. *Nature, 438*(7064), 78-81. doi: http://www.nature.com/nature/journal/v438/n7064/suppinfo/nature04154_S1.html